Conceptual toolkit: the molecular principles for understanding proteins

1

Introductory note

This chapter is intended to provide a concise summary of the principles with which you should be familiar in order to understand the structures and functions of proteins. If you have already studied chemistry in some depth, you may well not need to study this chapter in detail; a good test would be to see how well you can tackle the problems at the end of the chapter. If you are in any doubt, however, it is probably worthwhile taking the time to read it carefully and master its contents!

If you wish to read more about the molecular principles outlined in this chapter you will find that there is good coverage of these in most of the standard textbooks on biochemistry or introductory chemistry, such as Berg *et al.* (2007), Crowe *et al.* (2006), Jones (2005), Mathews *et al.* (2000), Price *et al.* (2001), and Voet *et al.* (2006)

Aims of this chapter

Proteins are complex three-dimensional structures with a very wide variety of functions in nature. This chapter is designed to help you acquire or consolidate your understanding of the structure and function of molecules, with a particular focus on those relevant to the study of proteins. It is not meant to be a 'small-scale' chemistry course, but to build on a basic knowledge of the Periodic Table of the elements and the ways in which ionic and covalent compounds can be formed. It starts with the amino acids as the basic building blocks of proteins and then deals with the various levels of protein structure. The forces involved in stabilizing the three-dimensional structures of proteins and in the interactions between proteins and other binding partners are described. At the end of this chapter, there is a set of chemical structures that you should aim to learn, as they will help to deepen your understanding of the ways in which proteins behave. There is also a set of problems that you can use to check your understanding of the principles in this chapter.

1.1 The different aspects of protein structure

KEY CONCEPT

- Understanding the four levels at which protein structure can be defined

To understand how proteins function it is necessary to know their structure. Whereas for small molecules such as ethanol, benzene, or glucose we can gain very useful insights into the properties from simple two-dimensional representations

of their covalent structure, the same cannot be said for the much more complex molecules of proteins, which contain many thousands of atoms. In this case we have to extend the description of structure into three dimensions.

It is convenient to discuss the structure of proteins at four levels:

A. *Primary structure*, which refers to the *sequence of amino acids* in the polypeptide chain, i.e. the covalent structure of the protein. This includes any post-translational modifications such as glycosylation, phosphorylation, etc.

B. *Secondary structure*, which refers to the *local folding* of the polypeptide chain, such that segments of the chain may form helices, strands of sheet or turns.

C. *Tertiary structure*, which refers to the *long-range folding* of the polypeptide chain so that portions of the chain that are remote in terms of sequence are brought close together in space.

D. *Quaternary structure*, which refers to the association of the individual polypeptide chains (or subunits) in a multi-subunit protein.

Before we discuss these aspects of structure we should review the properties of the amino acids, which represent the building blocks of proteins.

1.2 The constituents of proteins, the amino acids

KEY CONCEPTS

- Knowing the general structure of an amino acid and understanding its stereochemistry
- Knowing the structures of the 20 common amino acids found in proteins
- Understanding the basis of amino acid classifications
- Explaining the behaviour of amino acids in chemical terms
- Understanding the properties of water as a solvent

There are 20 different amino acids commonly found in proteins. Of these one (proline) is actually a secondary amino acid (sometimes termed an imino acid). The general structure of an amino acid is shown in Fig. 1.1, where NH_2– is the amino group, –CO_2H is the acid group, and R is the side chain. The central C atom to which R is attached is referred to as the α-carbon atom (denoted C_α). If the side chain consists of a chain of C atoms, these are usually referred to by Greek letters

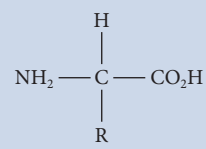


Fig. 1.1 General structure of an amino acid.

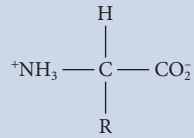


Fig. 1.2 The zwitterionic form of an amino acid in which the carboxyl group has lost a proton and the amino group has gained a proton.

going away from the α-carbon atom, i.e. β, γ, δ, ε. Thus the side chain of glutamic acid (Glu) (see section 1.2.1) contains a γ-carboxyl group and of lysine (Lys) contains an ε-amino group.

In reality, in aqueous solution at pH values around neutrality, the amino group gains a proton and the acid group loses a proton, giving rise to the so-called zwitterionic form, shown in Fig. 1.2.

The zwitterionic (from the German *zwitter*, meaning hybrid) form of an amino acid is one which is neutral overall but carries both a positive and a negative charge.

It will greatly help your understanding of proteins if you can learn the names, abbreviations, and chemical structures of the amino acids, i.e. of their side chains, because then you will be able to account for the structure and properties of these proteins in molecular terms. A list of the full structures of the amino acids is included in the Compendium of Structures in section 1.8.

1.2.1 The variety of amino acids

The amino acids are listed in alphabetical order in Table 1.1, with details of some of their important properties.

Many of the amino acids were named for what now seem rather obscure reasons, e.g. glycine comes from the Greek *glykeros* (sweet) because pure glycine was found to have a sweet taste. Leucine is named from the Greek *leukos* (white) because it was first isolated as a white crystalline solid; at the time whiteness was considered to be the defining property of a pure compound. Serine is derived from the Greek *serikon* and Latin *sericum* (silk) because it is found in significant quantities in the hydrolysis products of silk.

1.2.2 Classification of the amino acids in terms of polarity

The different amino acids can be classified in a number of ways, but probably the most useful in terms of understanding protein structure is that based on the polarity of the side chain. This can be thought of as a measure of whether the side chain promotes (polar) or discourages (non-polar) solubility in water. The polar amino acids are often further sub-divided into charged and uncharged side chains.

It should be noted that the classification of a few of the amino acids (particularly Cys and Gly) is difficult; they may be placed in different categories in other books.

Non-polar side chain

Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val

Polar, uncharged side chain

Asn, Cys, Gln, Ser, Thr, Tyr

Polar, charged side chain

Arg, Asp, Glu, His, Lys

Table 1.1 Properties of the amino acids commonly found in proteins

Name	Abbreviation		Mass[a]	pK_a of side chain[b]	Frequency of occurrence (%)[c]
	Three-letter	One-letter			
Alanine	Ala	A	71.08	–	7.83
Arginine	Arg	R	156.19	12.5	5.35
Asparagine	Asn	N	114.10	–	4.18
Aspartic acid	Asp	D	115.09	3.9	5.32
Cysteine	Cys	C	103.14	8.4	1.52
Glutamine	Gln	Q	128.13	–	3.95
Glutamic acid	Glu	E	129.12	4.1	6.64
Glycine	Gly	G	57.05	–	6.93
Histidine	His	H	137.14	6.0	2.29
Isoleucine	Ile	I	113.16	–	5.91
Leucine	Leu	L	113.16	–	9.64
Lysine	Lys	K	128.17	10.5	5.93
Methionine	Met	M	131.20	–	2.38
Phenylalanine	Phe	F	147.18	–	4.00
Proline	Pro	P	97.12	–	4.83
Serine	Ser	S	87.08	–	6.86
Threonine	Thr	T	101.11	–	5.42
Tryptophan	Trp	W	186.21	–	1.15
Tyrosine	Tyr	Y	163.18	10.5	3.06
Valine	Val	V	99.13	–	6.71

[a]The mass in Da (Daltons) is given minus that of water. The molecular mass of a protein can be obtained by adding the values shown for the masses of the constituent amino acids plus that of one molecule of H_2O (18.02). This mass would be adjusted if necessary for the effect of any post-translational modifications, e.g. the addition of one phosphate group would add 79.98 Da.

[b]The pK_a value is that of the side chain in the free amino acids; in a protein the value of the pK_a can be markedly influenced by the precise environment of the side chain. The pK_a values of the carboxyl and α-amino groups for most amino acids are typically about 2.2 and 9.5, respectively.

[c]The frequency is taken from the occurrence of the amino acids in all sequences deposited in the Swiss-Prot database (release 49.0 7 Feb 2006; representing 75438310 amino acids in 207132 sequences).

The sign and magnitude of the free energy change indicates the tendency of a process or reaction to occur (see Chapter 4, section 4.1). A process with a large positive free energy change will have little tendency to proceed; conversely a process with a large negative free energy change will have a high tendency to proceed.

A quantitative measure of polarity can be provided by a so-called hydrophobicity scale. There are several ways in which this can be set up; that due to Engelman *et al.* (1986) corresponds to the free energy change for transfer of a given amino acid in an α-helix from the interior of a membrane to an aqueous medium (Table 1.2). In the table, the amino acids are ranked in order from most non-polar to most polar.

1.2.3 General properties of the amino acids

In this section we shall explore two general properties of amino acids before looking at specific chemical characteristics in the next section.

Table 1.2 A hydrophobicity scale for amino acids

Amino acid	Transfer free energy (kJ mol^{-1})
Phe (F)	15.5
Met (M)	14.2
Ile (I)	13.0
Leu (L)	11.7
Val (V)	10.9
Cys (C)	8.4
Trp (W)	7.9
Ala (A)	6.7
Thr (T)	5.0
Gly (G)	4.2
Ser (S)	2.5
Pro (P)	−0.8
Tyr (Y)	−2.9
His (H)	−12.5
Gln (Q)	−17.1
Asn (N)	−20.1
Glu (E)	−34.3
Lys (K)	−36.8
Asp (D)	−38.5
Arg (R)	−51.4

1.2.3.1 Stereochemistry

Since the α-carbon atom has four different substituents (except in the case of Gly for which R = H), all amino acids, except Gly, will display chirality. Only one of the enantiomers (mirror image forms) occurs in proteins; this is the L form of the amino acids (see Fig. 1.3). The D/L system for designating the stereochemistry of the amino acids is based on the D and L forms of glyceraldehyde as a reference compound, i.e. related to other compounds by a series of reactions of known stereochemistry. The more recent and absolute *R*/*S* system for describing the stereochemistry of chiral compounds is based on ranking the atoms or groups attached to the asymmetric carbon atom primarily in terms of atomic number and then establishing the direction of rotation of the ranked groups when looking along the bond from the carbon atom to the lowest ranked group (see Fig. 1.4). Almost all the L-amino acids have the *S* configuration at the α-carbon atom.

The side chains of Ile and Thr each contain a second chiral centre; these can be readily designated in the *R*/*S* system as *S* and *R* for Ile and Thr, respectively.

D-amino acids are found occasionally in nature, for example D-Ala and D-Gln in the peptidoglycan component of some bacterial cell walls, and D-Phe in gramicidin S, a cyclic decapeptide synthesized by the bacterium *Bacillus brevis*. Peptides containing these amino acids are made by special enzyme systems distinct from the normal ribosome-based protein synthesizing machinery.

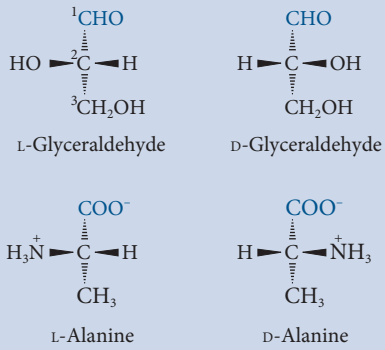


Fig. 1.3 The D/L system for denoting the stereochemistry of amino acids. The upper panels show the structures of L- and D-glyceraldehyde, which is the reference compound; the lower panels show the L and D forms of alanine. The solid arrows show atoms pointing towards the reader; the dotted arrows show atoms pointing away from the reader. The L and D forms (enantiomers) of each compound are mirror images of each other and are non-superimposable.

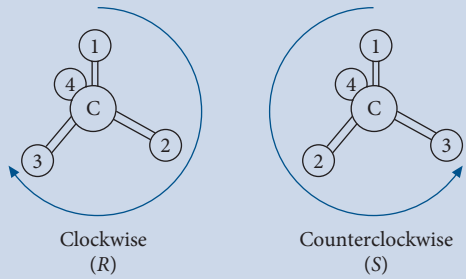


Fig. 1.4 The *R*/*S* system for denoting the stereochemistry at a chiral carbon atom. In this system the atoms linked to the carbon atom are ranked in terms of priority (1 being the highest ranked and 4 the lowest). When viewed looking from the carbon atom to atom 4, the direction in which the other atoms are ranked 1→2→3 is either clockwise (denoted *R*) or counterclockwise (denoted *S*). The priorities of some common substituent groups are $-OCH_2X > -OH > -NH_2 > -COOH > -CHO > -CH_2OH > -CH_3 > -H$.

1.2.3.2 Ionization

In Table 1.1, the average pK_a values of the α-amino and carboxyl groups of amino acids are given (9.5 and 2.2, respectively). These allow us to see why at neutral pH amino acids will exist in the zwitterionic form (NH_3^+–CHR–CO_2^-), since at pH 7 the amino group is below its pK_a and will thus be protonated, and the carboxyl group will be above its pK_a and thus be deprotonated. The actual charge carried by an amino acid at any particular pH will also depend on the nature and state of ionization of the side chain (see Table 1.1).

The isoelectric point (denoted by pI) of an amino acid is the pH at which it carries no net charge, i.e. where there is an exact balance of positive and negative charges. At the pI, the amino acid will not move in an electric field. For an amino acid with

no ionizable side chain, the pI is equal to the average of the pK_a values of the carboxyl and α-amino groups (e.g. for Gly, pI = (2.4 + 9.8)/2 = 6.1). For an amino acid with an ionizable side chain, the pI is equal to the average of the two pK_a values on either side of the zwitterionic form. Thus for Asp, where the pK_a values for the α-carboxyl, the side chain carboxyl, and the α-amino group are 2.0, 3.9, and 9.9, respectively, the pI = (2.0 + 3.9)/2 = 2.95. For a protein, the actual pI will depend on the balance between the numbers of the acidic and basic amino acids in that protein.

The calculation of the pI value for amino acids with ionizing side chains is discussed in Problems 1.6 and 1.7 at the end of this chapter.

KEY INFO !

It should be remembered that the ionization states of the side chains of amino acids in proteins can be significantly different from those of the free amino acids. For example, the pK_a of Glu 35 in lysozyme is raised from 4.1 in the free amino acid to about 6.0 because the side chain is in a rather non-polar environment. This has the effect of favouring the neutral (protonated) form rather than the negative (deprotonated) form. Conversely, the pK_a of Tyr 28 in type II dehydroquinase is lowered from 10.5 in the free amino acid to about 8.0 by the presence of a neighbouring Arg side chain. The positive charge on the Arg will favour the deprotonated form of the Tyr side chain relative to the protonated form. In the case of Ser 195 of chymotrypsin, the pK_a is lowered dramatically (by over 7 units) due to the charge relay system. It is therefore always important to examine the actual environment of an amino side chain in a protein before assessing its likely function.

1.2.4 Chemical characteristics of the amino acids

It is convenient to discuss the various amino acids as groups which display similar chemical features. Again, knowing the structures of the side chains involved will help to understand the molecular basis of the property.

1.2.4.1 Aliphatic side chains (Ala, Gly, Ile, Leu, Val)

These side chains are chemically unreactive. They are important in hydrophobic interactions (section 1.7.4) with other non-polar side chains or parts of other molecules.

Because of its small side chain, Gly has a large degree of conformational flexibility, which is important when the three-dimensional structures of proteins are considered (section 1.4).

Aliphatic compounds are those in which carbon atoms are linked in straight or branched chains, rather than rings.

1.2.4.2 Aromatic side chains (Phe, Tyr, Trp)

These side chains absorb radiation in the near-ultraviolet (the absorbance maxima for Phe, Tyr, and Trp are 258 nm, 275 nm, and 280 nm, respectively).

The side chain of Phe is chemically unreactive but can participate in hydrophobic interactions (section 1.7.4).

Aromatic compounds are those which contain benzene-type rings. Their enhanced stability arises from the delocalized π-electron systems.

An electrophile is a species or part of a molecule with a positive or partial positive charge which will seek an electron-rich species or centre. Examples include the Zn^{2+} ion or the C atom in the >C=O bond. In the case of tyrosine, reagents that bring about electrophilic substitution include I_2 and the NO_2^+ (nitronium) ion.

Thyroxine, an iodinated derivative of tyrosine, is a hormone which is produced in the thyroid gland. It stimulates the synthesis of specific proteins and is essential for normal growth and development. A number of isotopes of iodine (particularly ^{129}I and ^{131}I) are radioactive and can be dangerous because they will be concentrated in the thyroid gland, leading to thyroid cancer.

A nucleophile is a species or a centre in a molecule with a negative or partial negative charge which will seek an electron-deficient species or centre. Examples include OH^- ions and the side chains of Cys ($-CH_2-S^-$), Lys ($-NH_2$), and Ser ($-CH_2-O^-$) with one or more lone pairs of electrons.

The side chain of Tyr is more reactive (due to the hydroxyl group) and will undergo electrophilic substitution reactions. One such example is iodination, which occurs naturally to give the hormone thyroxine. The hydroxyl group has a high pK_a (typically about 10.5) so will not be significantly ionized under normal physiological circumstances; it can, however, participate in hydrogen bonding (section 1.7.2) and be phosphorylated by the action of kinases (section 1.2.4.5).

The side chain of Trp is of limited chemical reactivity, but the N–H group can participate in hydrogen bonding (section 1.7.2).

1.2.4.3 Basic side chains (Arg, Lys)

These side chains are grouped together because at neutral pH they carry a positive charge and can therefore be involved in ionic interactions with negatively charged side chains (Asp and Glu) or with negatively charged parts of other molecules, such as phosphate groups.

The pK_a of Lys is around 10.5 for the free amino acid. The $-NH_3^+$ group is unreactive, but the deprotonated form ($-NH_2$) can act as a powerful nucleophile.

The pK_a of Arg is around 12.5, and the guanidine part of the side chain is therefore essentially positively charged under all physiological conditions.

1.2.4.4 Acidic side chains (Asp, Glu)

For both free amino acids, the pK_a values of the side chains are around 4.0, and hence would generally be negatively charged at neutral pH. These side chains can be involved in ionic interactions with positively charged side chains, or in binding to metal ions such as Ca^{2+} or Zn^{2+}.

1.2.4.5 Hydroxyl side chains (Ser, Thr)

The –OH group of the side chain of both amino acids can take part in hydrogen bonding (section 1.7.2). The pK_a of the side chains is very high (about 15), so they are in the protonated state under physiological conditions and act as very weak nucleophiles. However, in certain cases, such as the Ser proteases (see Chapter 9, section 9.9.3), the Ser can lose its proton by the charge relay system and become a very powerful nucleophile. Ser and Thr side chains (along with Tyr side chains, section 1.2.4.2) in a number of proteins can become phosphorylated by the action of specific enzymes (kinases); such processes are often involved in regulation of the activity of these proteins.

1.2.4.6 Amide side chains (Asn, Gln)

The amide group has only very weak acidic and basic properties so the side chains of these amino acids are always uncharged. They can participate in hydrogen bonding (section 1.7.2) as either donors or acceptors.

1.2.4.7 Sulphur-containing side chains (Cys, Met)

The sulphur atom of the Cys side chain is a highly reactive nucleophile, especially in the deprotonated state ($-CH_2S^-$). The pK_a of the side chain (8.4 for the free amino acid) is sufficiently low that at neutral pH there will be a small, but significant, fraction in the deprotonated form.

Under oxidizing conditions, two appropriately positioned Cys side chains can form a disulphide bond ($-CH_2-S-S-CH_2-$), which constitutes the amino acid cystine. Disulphide bonds are found in extracellular, secreted proteins such as antibodies, digestive enzymes, etc.

Disulphide bonds are only found in secreted proteins. In intracellular proteins (i.e. those which remain inside cells) the cysteine side chains remain in the reduced ($-CH_2-SH$) form. The oxidation of secreted proteins occurs in the lumen (interior compartment) of the endoplasmic reticulum in eukaryotic cells and in the periplasmic space of prokaryotic cells.

In the presence of oxidizing agents, the Cys side chain can be oxidized to a sulphonic acid ($-CH_2-SO_3H$) or the intermediate sulphenic ($-CH_2-SOH$) and sulphinic ($-CH_2-SO_2H$) acid forms. The sulphur atom in Cys can act as a ligand to a number of metal ions, e.g. Zn^{2+} or Fe^{2+}, and is a major site of inhibition of proteins by heavy metals such as Hg, Cd, or Pb.

The sulphur atom of Met is much less reactive, although it can function as a nucleophile towards adenosine-5′-triphosphate (ATP) to form *S*-adenosylMet (adoMet) in a reaction involving the release of the three phosphate groups, which is catalysed by methionine adenosyl transferase. The positive charge on the sulphur atom of adoMet makes it susceptible to nucleophilic attack, and hence adoMet is a powerful methylating agent, e.g. of cytosine bases in DNA. The side chain of Met can also be modified by strong oxidizing agents to give Met sulphoxide.

Oxidative damage to proteins, particularly to the sulphur-containing amino acids, is thought to be an important factor in ageing-related diseases. Reaction of Cys side chains in proteins with reactive oxygen species such as H_2O_2 or the superoxide anion can yield the sulphenic acid form. This reaction is usually reversible, but further oxidation cannot generally be reversed.

1.2.4.8 Proline

Proline is a special amino acid since its side chain is a cyclic ring. This structural feature means that there are rigid geometrical constraints on the peptide bonds to Pro (section 1.3.1). The side chain of Pro is chemically rather unreactive, although it can be hydroxylated in the 4 position by the action of prolyl-4-hydroxylase. 4-Hydroxyproline is found in the protein collagen, which occurs extensively in bone, skin, and connective tissue.

Prolyl-4-hydroxylase requires ascorbic acid (vitamin C) as a cofactor. A deficiency of this vitamin leads to the condition of scurvy, a disease of the skin, joints, etc. caused by disruption to the correct synthesis and assembly of collagen fibres.

1.2.4.9 Histidine

The pK_a of the imidazole ring is about 6.0 in the free amino acid. Thus, at near-neutral pH, there is balance of the protonated (positively charged) and deprotonated (neutral) forms. His side chains are often found at the active sites of enzymes as they can play a key role in acid–base catalysis. The neutral form of His is a powerful nucleophile and can participate in hydrogen bonding (section 1.7.2); in addition it can act as a ligand to metal ions such as Zn^{2+} or Fe^{2+}.

1.2.5 The structure of water

Since the majority of biological processes take place in an aqueous medium, it is appropriate to review the structure of liquid water. The geometry of the water molecule is well known, with the H–O–H bond angle = 104.5° and the O–H bond

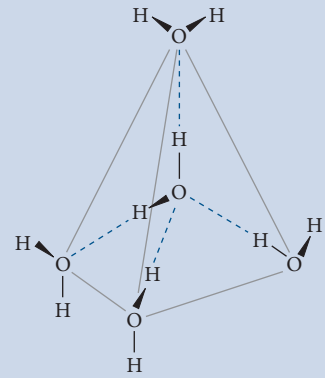


Fig. 1.5 A water molecule hydrogen bonded to its four nearest neighbours. The covalent O–H bonds are shown as solid lines; hydrogen bonds are shown as dotted lines. The four nearest-neighbour water molecules are arranged in a tetrahedral fashion. The arrangement is found in one of the most common structures of ice, where there are effectively layers of water molecules.

length = 0.096 nm. There is a separation of charge due to the higher electronegativity (electron attracting power) of the O atoms, leading to partial positive charges on the hydrogens and a partial negative charge on the oxygen (about 33% ionic character). The two lone pairs of electrons on the O atom can each act as acceptors in hydrogen bonds and the two hydrogen atoms can act as donors (section 1.7.2). Thus, each water molecule can have up to four nearest-neighbour hydrogen-bonded water molecules (Fig. 1.5). This network is found in ice, which has a significant amount of empty space and thus a relatively low density (ice floats on water).

Water is unusual in that its solid form (ice) is less dense than its liquid form. This is of considerable importance for aquatic organisms in the winter!

In liquid water, this regular hydrogen-bonded network is broken down and instead there are fluctuating clusters of water molecules with, on average, each water having about 3.5 hydrogen-bonded nearest-neighbour molecules. Because of the polar nature of the water molecule there will be strong interactions with charged or other polar species, making water an excellent solvent for these types of molecules. However, water is a very poor solvent for non-polar molecules; this is discussed further in the context of hydrophobic interactions in section 1.7.4.

1.3 The primary structure of proteins

KEY CONCEPTS

- Drawing the structure of a peptide chain
- Understanding the *cis*/*trans* isomerism of the peptide bond unit
- Understanding the information available from the primary structure of a protein, including molecular mass, extinction coefficient, isoelectric point, and any post-translational modifications

The term ‘primary structure’ refers to the sequence of amino acids in a protein, and is dictated by the nucleotide sequence of the gene encoding the protein, taking into

account (a) the splicing out of any intervening sequences (introns) to produce the mature messenger RNA and (b) any post-translational modifications which might remove amino acids by proteolysis or modify amino acids, e.g. by addition of carbohydrate or phosphoryl groups.

1.3.1 The peptide bond

The peptide bond is formed between the carboxyl group of one amino acid and the amino group of a second amino acid, with the loss of a molecule of H_2O. By convention, a peptide chain is written such that the amino acid with the free amino group (the amino (or N)-terminal acid) is at the left-hand end of the chain and the amino acid with the free carboxyl group (the carboxyl (or C)-terminal acid) is at the right-hand end (Fig. 1.6). Peptide chains in proteins are unbranched.

The formation of a dipeptide from two amino acids is thermodynamically unfavourable. In nature, ATP is used as an energy source to activate one of the amino acids and drive the reaction to the side of synthesis.

The peptide bond itself is not adequately represented by the structure shown in Fig. 1.7, as shown by the fact that the C–N bond distance (0.132 nm) is between

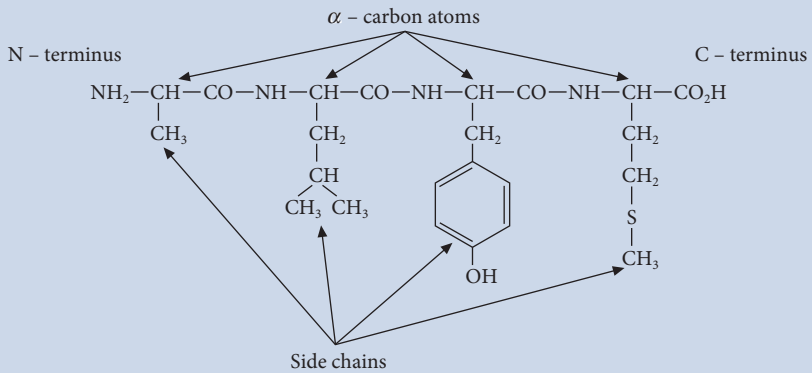


Fig. 1.6 Structure of a tetrapeptide showing the α-carbon atoms, side chains, and N- and C-termini. The tetrapeptide is written as Ala–Leu–Phe–Met (ALFM in the one-letter code). In this diagram the terminal amino and carboxyl groups are shown in the uncharged forms; in practice they would carry a positive charge and negative charge from the gain and loss of a proton, respectively.

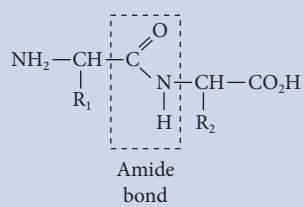


Fig. 1.7 Structure of a peptide bond with the C–N bond shown as a single bond.

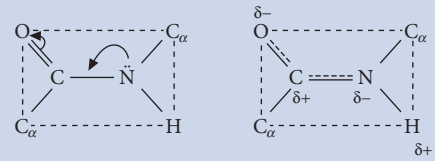


Fig. 1.8 Resonance stabilization of the peptide bond. The left-hand panel shows delocalization of the lone pair of electrons on the N atom. The right-hand panel shows the partial double bonds in the peptide bond unit and the partial charges on the C, O, N, and H atoms.

that of a C–N single bond (0.145 nm) and a C=N double bond (0.125 nm). This is usually explained by resonance stabilization, with electron density in the lone pairs on the O atom being transferred to the N atom, giving rise to partial double bond character (Fig. 1.8).

Since the N atom carries a net negative charge, there must be extensive back-donation of σ electrons from C to N to compensate for the π electron movement.

This resonance has two consequences. Firstly, it establishes a permanent dipole (separation of charge) associated with the peptide bond. The H and N atoms carry positive and negative charges of about 0.28 electron charges respectively and the C and O atoms carry positive and negative charges of about 0.39 electron charges respectively (Fig. 1.8).

Secondly, it means that peptide bonds are planar units in which the two α-carbon atoms are on the same side of the C–N bond (*cis* form) or on opposite sides of the C–N bond (*trans* form) (Fig. 1.9).

The repulsion between non-bonded atoms linked to the two α-carbon atoms is greater in the *cis* form than in the *trans* form. This means that the *trans* form

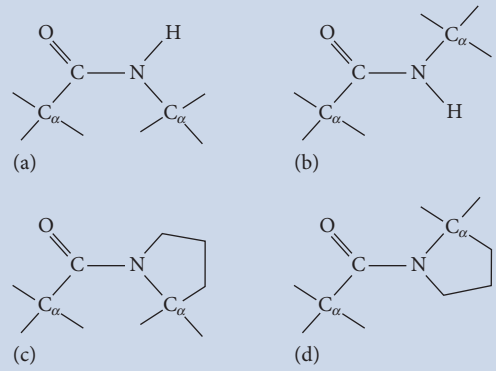


Fig. 1.9 The *cis* and *trans* forms of the peptide bond. (a) and (b) show the *cis* and *trans* forms, respectively, of the peptide bond (Xaa–Zaa), where Xaa is any amino acid and Zaa is any amino acid other than proline. (c) and (d) show the *cis* and *trans* forms, respectively, of the peptide bond (Xaa–Pro).

is more stable than the *cis*; in proteins it can be estimated that this is by a margin of about 20 kJ mol^{-1}.

An exception to this statement is in the case when the peptide bond is to proline (Fig. 1.9), which as we have seen is an unusual amino acid with the side chain forming a cyclic ring. This has the effect that the repulsion between non-bonded atoms is considerably reduced and the balance between *cis* and *trans* is much closer. The *trans*/*cis* ratio is now much lower (about 20), i.e. about 5% of Xaa–Pro bonds in proteins are in the *cis* form (Jabs *et al.*, 1999).

It should be noted that there is a considerable activation energy barrier (of the order of 80 kJ mol^{-1}) to rotation about the peptide bond because of its partial double bond character. This isomerization of the peptide bonds to Pro can represent one of the slow steps in the folding of proteins and there is an enzyme, peptidyl-prolyl isomerase, which can catalyse this process.

1.3.2 Information available from the amino acid sequence of a protein

Knowledge of a protein sequence gives us a good deal of information, as set out below. Further details are given in the section on bioinformatics (see Chapter 5, section 5.8).

1.3.2.1 Exact molecular mass

From the numbers of each amino acid in a protein, we can calculate the exact mass using the values given in Table 1.1 for the masses of each amino acid and adding the mass of H_2O. For example, the calculated molecular mass of the polypeptide chain of type II dehydroquinase from *Streptomyces coelicolor* is 16550.6 Da. This can be compared with the very precise measurements of mass available from mass spectrometry to confirm the identity and integrity of the protein. In this case the measured mass of the purified enzyme is 16549.7 ± 1.2 Da, in excellent agreement with the predicted value.

1.3.2.2 Isoelectric point (pI)

As mentioned in section 2.3, the pI of a protein is the pH where it carries no net charge. The way that the charge on the protein will change with pH is illustrated in Fig. 1.10.

The pI of a protein will reflect a balance between the numbers of positively charged groups (N-terminal amino group and the side chains of Arg, His, and Lys) and negatively charged groups (C-terminal carboxyl group and the side chains of Asp and Glu). The predicted pI is based on the numbers of these charged groups and their average pK_a values in proteins. If the positive groups are predominant, the pI will be high (e.g. for pig lysozyme, a value of 9.06 is predicted); if the negative groups predominate, the pI will be low (e.g. for pig pepsin, a value of 3.24 is predicted). It should be noted that these are only rough estimates of the pI value since the pK_a values of side chains in proteins can depend markedly on their environment. Nevertheless, the predicted value is a useful guide to the behaviour of the

Electrophoresis refers to the movement of a charged molecule in an electric field. Electrophoresis of proteins is often performed in a cross-linked polyacrylamide gel (see Chapter 6, section 6.2, and Chapter 8, sections 8.2.1 and 8.2.2). Ion-exchange chromatography is a widely used technique for separating proteins on the basis of the charge that they carry (see Chapter 7, section 7.2.2).

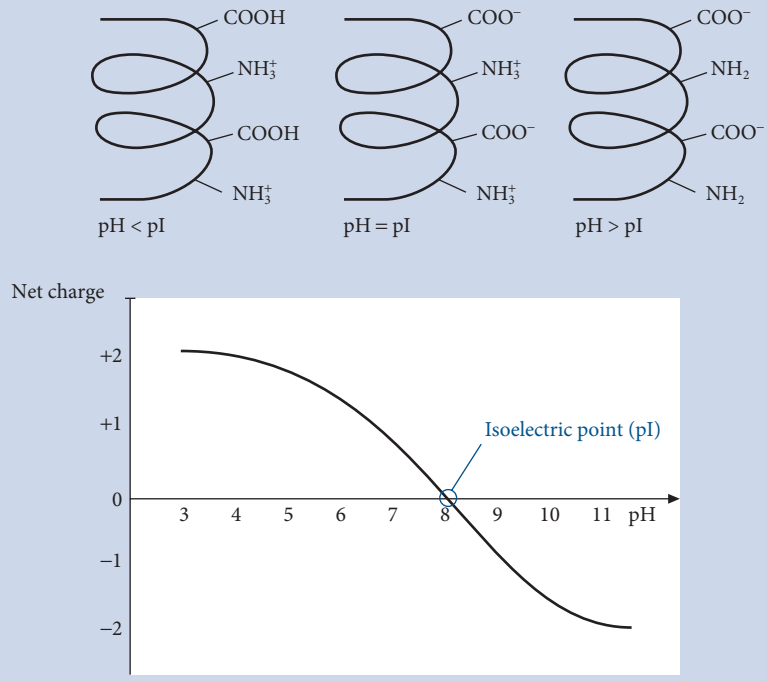


Fig. 1.10 The effect of pH on the charge of a protein. At low pH, the protein has an overall positive charge. As the pH is raised, acidic side chains (Asp and Glu) become deprotonated. When the number of negatively charged side chains equals the number of positively charged side chains, the protein has no net charge; this pH equals the isoelectric point (pI). As the pH is raised further, the protein has an overall negative charge as a result of deprotonation of basic side chains.

protein on electrophoresis (see Chapter 6, section 6.2.3) or ion-exchange chromatography (see Chapter 7, section 7.2.2), for example.

1.3.2.3 *Absorption coefficient*

The calculation of the absorption coefficient depends on the assumption that there is no other chromophore (absorbing group) present, such as the haem group in haemoglobin. If such a group is present, its contribution to the absorption at 280 nm must be added.

The absorption of radiation at 280 nm depends on the aromatic amino acids Tyr and Trp; there is also a small contribution from any disulphide bonds which may be present. The molar absorption coefficient (also known as the extinction coefficient) of the protein at 280 nm can be derived by adding together the contributions from the numbers of these different amino acids; these are based on their numbers in the protein and their known absorption coefficients at 280 nm. Division by the molecular mass of the protein gives the absorption coefficient for a 1 mg mL^{-1} solution of the protein (see Chapter 6, section 6.1.5).

1.3.2.4 *Hydrophobicity*

The balance between polar and non-polar amino acids in the protein will give a guide to the overall polarity of the protein and can point to whether it may be membrane-associated, for example. The aliphatic index is a measure of the frequency of occurrence of the bulky aliphatic side chains (Ala, Ile, Leu, and Val).

Table 1.3 Some important post-translational modifications of proteins

Type of modification	Possible effects on function
Proteolysis	Removal of targetting sequences Generation of several new products, e.g. hormones Activation of proteins, e.g. enzymes
Disulphide bond formation	Stabilization of structure of secreted proteins
Hydroxylation	Formation of hydroxy–Lys or –Pro increases the stability of the triple helix of collagen
Glycosylation	Many cell surface proteins are involved in cell–cell recognition Attachment of glycosyl-phosphatidyinositol groups anchors proteins to membranes The polar nature of proteins can be enhanced
Phosphorylation	Phosphorylation of Ser, Thr or Tyr side chains can regulate the activity of proteins, especially in signalling pathways
N-terminal acylation	Attachment of C_{14} (myristoylation) or C_{16} (palmitoylation) chains will enhance the association of the protein with membranes

1.3.2.5 Post-translational modifications

Many proteins undergo changes in covalent structure after translation of the mRNA on the ribosome. These changes are collectively termed post-translational modifications and some of the more important of these are listed in Table 1.3.

The occurrence of post-translational modifications can be deduced either by direct analysis, e.g. testing for the presence of saccharide units in the case of glycosylation, or by careful measurements of the molecular mass of the protein using mass spectrometry. If the measured molecular mass of a protein does not agree with the value calculated on the basis of the amino acid sequence, then it can be concluded that either the sequence is incorrect or (more likely) post-translational modification has occurred. From the observed mass differences it should be possible to draw firm conclusions about the types of modifications which have occurred. In the case of horse lysozyme the measured mass is 14644 Da, indicating a loss of 8 Da from the predicted value (14652 Da); this can be accounted for by the formation of four disulphide bonds from the eight Cys residues in the sequence, which corresponds to the loss of eight hydrogen atoms.

In a number of cases, extensive post-translational modification (e.g. glycosylation) may hinder the analysis of a protein by mass spectrometry, as the protein cannot readily form gaseous phase ions.

1.3.2.6 Structural and functional motifs

Analysis of the sequence can reveal the way certain parts of the protein may play distinct roles. For example, a stretch of 20 amino acids which is very likely to form a membrane-spanning (or transmembrane) α-helix (see section 1.4) can be

identified by constructing a hydropathy plot in which the free energy for transferring this stretch of amino acids from the membrane to water (obtained by adding the appropriate values of the amino acids as listed in Table 1.2) is plotted as a function of the first residue in the stretch. If the free energy is above a certain threshold value (+84 kJ mol^{-1}) this indicates that a transmembrane helix is likely. Protein sequences can be scanned automatically to locate such elements, which often occur many times in a single chain, for example the G-protein coupled receptors have seven transmembrane helices and the family of glucose transporters in mammalian tissues have 12 transmembrane helices.

It should be noted that the choice of the free energy threshold for identification of a membrane-spanning element of a protein is somewhat arbitrary; the quoted value has been found to predict such elements with reasonable accuracy.

Other types of motifs that can be revealed by sequence analysis are listed below (Xaa equals any amino acid):

Targetting sequences, e.g. –Ser–Lys–Leu at the carboxyl terminus and –Lys–Asp–Glu–Leu at the carboxyl terminus act as sequences which direct proteins to the peroxisome or to be retained in the endoplasmic reticulum respectively.

Metal binding, e.g. –Cys–Xaa_4–Cys–Xaa_2–Cys– is a consensus site for Fe binding in the 2Fe–2S cluster iron sulphur proteins.

Phosphorylation sites, e.g. –Arg–Xaa_{1-2}–Ser/Thr– or –Arg–Arg–Xaa–Ser/Thr– are consensus sites for phosphorylation at the Ser or Thr by protein kinase A.

Glycosylation sites, e.g. –Asn–Xaa–Ser/Thr– (where Xaa is any amino acid but is rarely Pro or Asp) is a consensus sequence for N-glycosylation at Ser/Thr.

A commonly used tool for identifying structural motifs is Prosite (http://www.expasy.org/prosite). For further details see Chapter 5, section 5.8.5.

1.3.2.7 Sequence relationships between proteins

Analysis of the huge number of sequences in the databases has provided powerful insights into many aspects of proteins. For example, comparisons between the analogous proteins from different organisms has given clues about evolutionary relationships and helped in classification of species. When many such proteins are compared, it is seen that some amino acids are totally conserved between species, some are partially conserved (i.e. are of similar chemical types) and others are freely variable. The totally conserved residues are likely to play essential roles in the structure and/or the function of the protein. The comparisons between related, but distinct, proteins within the same organism, for example haemoglobin and myoglobin, trypsin and chymotrypsin, point to gene duplication followed by independent (divergent) evolution as the likely mechanism. Comparisons between sequences have also been used to look for clues to possible functions of proteins, or of segments of proteins. For example, analysis of the sequences of a number of steroid hormone receptor proteins shows that they consist of three domains (independent folded and functional units), namely a transactivation domain which regulates transcription, a DNA binding domain, and a hormone binding domain. For further details about the use of bioinformatics-based approaches, see Chapter 5, section 5.8.

1.4 The secondary structure of proteins

KEY CONCEPTS

- Defining the torsional (dihedral) angles in a peptide bond
- Drawing the hydrogen bonding patterns in the α-helix, β-sheet, and β-turn structures
- Explaining why proline cannot be accommodated in regular α-helix and β-sheet structures

Because of the partial double bond character of the peptide bond, the six atoms of C_α–CO–NH–C_α can be considered to lie in a plane (Fig. 1.11). Rotations around the bonds to the linking C_α atoms between peptide units will define the path traced and hence the structure adopted by the polypeptide chain. The two angles (known as torsion or dihedral angles) are defined as ϕ (phi) and ψ (psi) for the bonds between N and C_α and between C_α and C, respectively. Values of 180° refer to the fully extended polypeptide chain. Positive values of ϕ and ψ refer to clockwise rotation, negative values to counter-clockwise rotation.

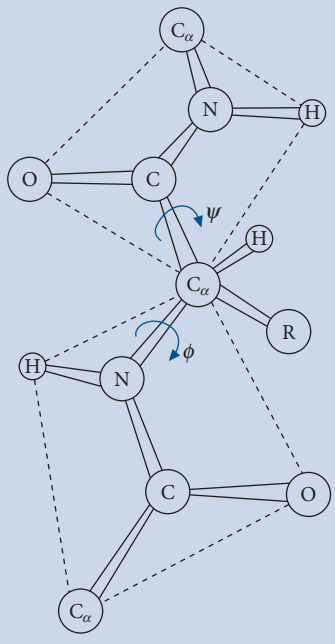


Fig. 1.11 Structure of a dipeptide element of a protein showing the planar peptide bond units. The angles of rotation (ϕ and ψ) about the bonds to the central α-carbon atom are shown. Values of ϕ = ψ = 180° correspond to the fully extended polypeptide chain.

Rotations about the bonds to the C_α atoms will alter the distances between non-bonded atoms of the side chains and hence the energy of the system. Calculations of the energy as a function of the angles ϕ and ψ have been made and the resulting plot (known as a Ramachandran diagram, Fig. 1.12) shows that there are certain

The precise allowed areas of the Ramachandran plot will depend on the nature of the side chains involved. Thus glycine, with its very small side chain, can occupy a considerably greater fraction of the total Ramachandran plot than can an amino acid with a bulky side chain such as valine or tryptophan.

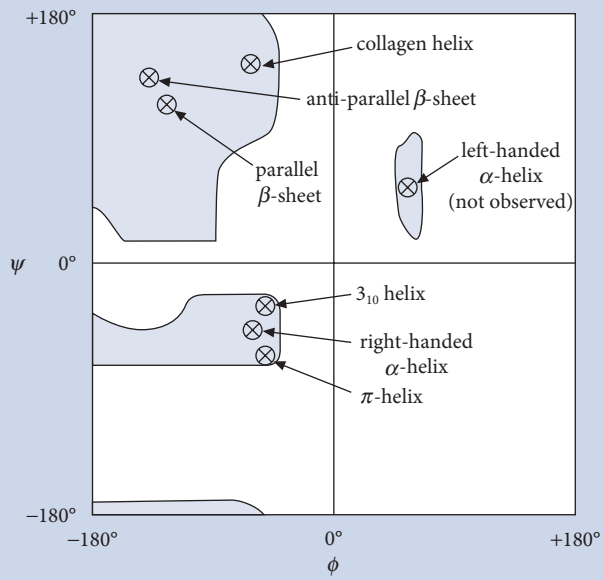


Fig. 1.12 The Ramachandran diagram showing the secondary structures of proteins formed by successive residues with identical conformations. The allowed values of the angles (ϕ and ψ) for a polypeptide chain consisting of alanine residues are shown by the shaded areas. The positions of a number of secondary structures are indicated.

well-defined structures adopted by peptide chains that are of greatest stability and are termed 'allowed'.

Some of the more common structural elements adopted by proteins are listed below.

1.4.1 The α-helix

Electrostatic interactions between the carbonyl group of amino acid *n* and the carbonyl groups of amino acids *n* + 3 and *n* + 4 also contribute to the stability of the α-helix. For L-amino acids, the right-handed α-helix is more stable than a left-handed helix.

The α-helix is a right-handed helix (i.e. the direction of twist from the N- to the C-terminus is that of a right-handed corkscrew). Hydrogen bonds are formed between the carbonyl group of the amide bond between amino acids *n* and *n* + 1, and the amino group of the amide bond between amino acids *n* + 3 and *n* + 4 (Fig. 1.13).

For model peptides of regular structure (polyamino acids) the dimensions of the α-helix are well characterized. There are 3.6 amino acids per turn, the pitch of the helix is 0.54 nm, and the angles ϕ and ψ are −57° and −47°, respectively. The hydrogen bonds are parallel to the axis of the helix and there is a linear arrangement of nuclei with the O⇌N distance 0.286 nm. There is a helix dipole with the N-terminal and C-terminal ends of the helix having positive and negative charges, respectively; this dipole can be important in stabilizing interactions between adjacent helices. Since there are 13 atoms in the ring containing the hydrogen bond (Fig. 1.13) and 3.6 amino acids per turn, an α-helix is sometimes referred to as a 3.6_{13} helix.

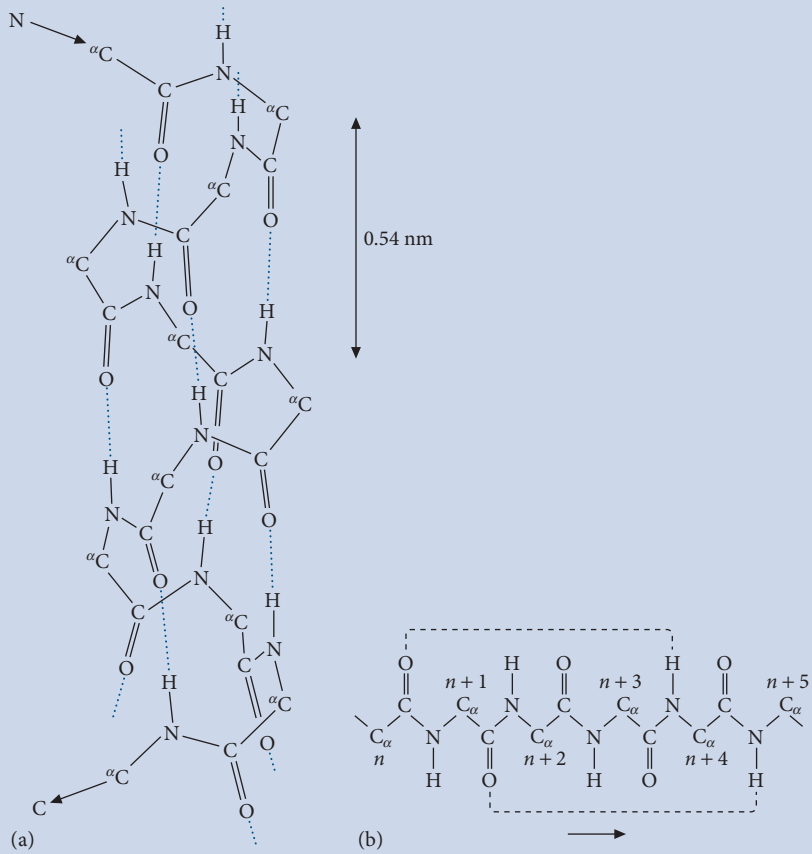


Fig. 1.13 The structure of the α-helix showing the arrangement of hydrogen bonds. (a) The path of the polypeptide chain with the α-carbon atoms and hydrogen bonds labelled. The arrows show the direction of the chain (N→C). (b) The hydrogen bonding arrangement; there are 13 atoms in the ring formed by the hydrogen bond.

Because of the steric constraints imposed by its cyclic side chain, Pro cannot be accommodated in the regular α-helix structure and will therefore either act as a helix breaker or will cause a major kink in the helix.

In proteins the α-helices are often considerably distorted from the idealized structures seen in model polypeptides with, for example, the hydrogen bond geometry less than ideal. The average length of α-helices in proteins is quite short, about 12 amino acids or 3.5 turns.

The representation of the α-helix in Fig. 1.13 does not show the side chains; these project out from the helix into the solvent. For some purposes it is useful to depict the arrangement of side chains in a projection known as the helical wheel (Fig. 1.14).

Note that with 3.6 amino acids per turn of the α-helix, the angle between successive α-carbon atoms in the helical wheel plot is 360/3.6 degrees, i.e. 100 degrees.

This might show, for example, that one face of an α-helix is polar and the other non-polar, in which case the helix is said to be amphipathic; such a helix might be involved in anchoring the protein to a membrane. If both sides of the helix were

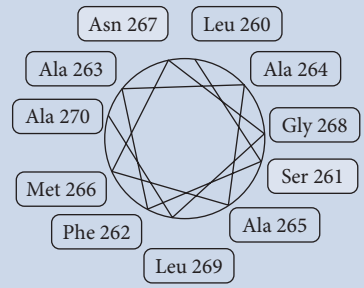


Fig. 1.14 The helical wheel projection of helix structure. The positions of 11 amino acid residues are plotted at 100° intervals in a clockwise direction looking down the helix from the N-terminus. In this sequence all the amino acids except Ser261 and Asn267 are non-polar. This helix would be expected to be located in the non-polar interior of a protein.

The α-helix is not the only secondary structure found in membrane proteins; several membrane-spanning proteins are β-sheet in character. A good example is the family of porins found in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts, which form channels for small polar molecules. Porins consist of 16 strands of β-sheet arranged in an anti-parallel fashion to form a pore through the membrane.

The π-helix has been found very occasionally in proteins (Weaver, 2000).

non-polar, the helix would be very likely to be buried in the interior of a protein or might be a membrane-spanning element of a protein. From the dimensions of the α-helix, 6 turns would be $6 \times 3.6 = 21.6$ amino acids in length and would span a distance of 6×0.54 nm = 3.24 nm, which is a typical value for the thickness of a phospholipid bilayer.

Other types of helix include the 3_{10} helix (three amino acids per turn, ten atoms in the hydrogen bond ring), in which the hydrogen bond is between the carbonyl group of the amide bond between amino acids n and $n+1$, and the amino group of the amide bond between amino acids $n+2$ and $n+3$. This more tightly folded helix is sometimes found in short stretches in proteins, particularly near the ends of helical segments or of polypeptide chains. Another possible type is the π-helix, which is more loosely coiled than the α-helix, with the hydrogen bond between the carbonyl group of the amide bond between amino acids n and $n+1$, and the amino group of the amide bond between amino acids $n+4$ and $n+5$.

1.4.2 The β-sheet

Electrostatic interactions between the carbonyl groups of adjacent chains contribute significantly to the stability of β-sheet structures. In actual protein structures (rather than model compounds) there appears to be little difference between the dihedral angles for the parallel and anti-parallel sheet patterns.

A β-sheet is made up from strands of polypeptide chain in an extended structure, with the side chains of the amino acids projecting alternately above and below the plane of the sheet. The strands can run parallel or anti-parallel to each other with hydrogen bonds forming between them; the bonds are more distorted in the case of parallel strands (Fig. 1.15). The values of the angles ϕ and ψ for idealized versions of (a) parallel β-sheet are −119° and +113°, respectively, and (b) anti-parallel β-sheet −139° and +135°, respectively. In both cases, the span of the sheet is about 0.33 nm per amino acid, i.e. more than twice that of the α-helix. When the structures of proteins are analysed, it is found that sheets composed of anti-parallel strands are more common than those formed from parallel strands, with generally at least four of the latter needed to form a sheet, compared with only two of the former. There

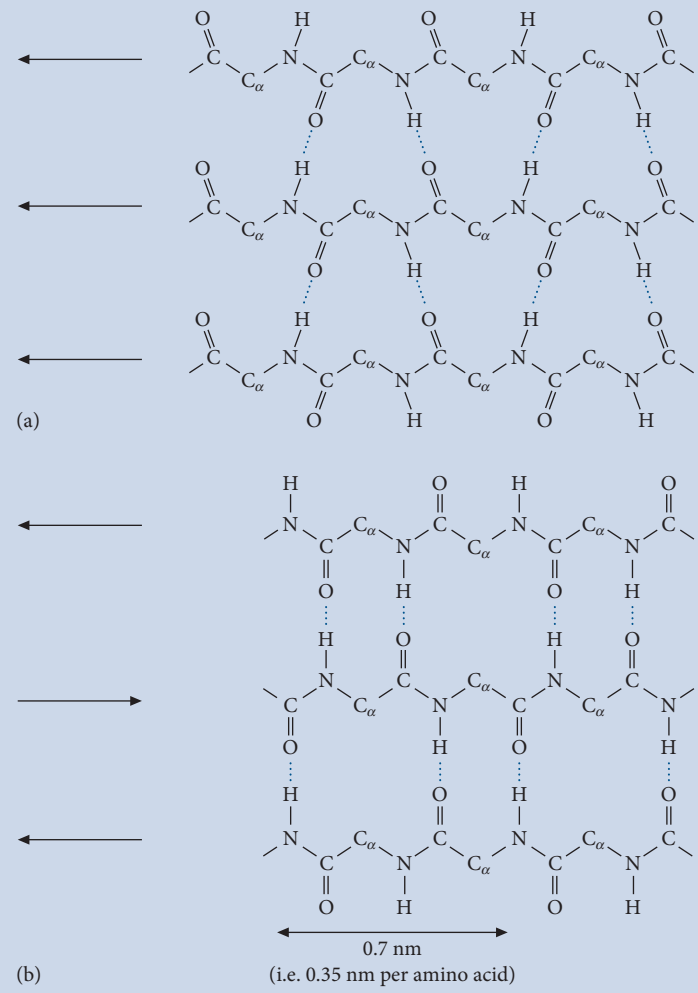


Fig. 1.15 Structure of β-sheets showing the arrangement of hydrogen bonds. (a) Hydrogen bonding in the parallel β-sheet; (b) hydrogen bonding in the anti-parallel β-sheet. In three dimensions, the sheet structures are in fact corrugated with the amino acid side chains projecting perpendicularly to the plane of the sheet.

is a right-handed twist between the strands, which appears to be a result of the electrostatic interactions between the carbonyl groups of adjacent peptide bonds in each strand; these are much more favourable with a right-handed than a left-handed twist. The average length of β-strands in proteins is about six amino acids.

1.4.3 Other structural features in proteins

In addition to α-helices and β-sheets, the most abundant structural element found in proteins is the β-turn. A β-turn is a sequence of amino acids in which the polypeptide chain folds back on itself by nearly 180°, allowing proteins to adopt globular, rather than extended, shapes. There are at least eight different types of

About 15% of residues in proteins not in α-helices or β-sheets can be in β-turns (Panasik *et al.*, 2005).

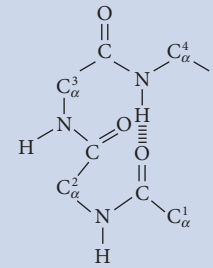


Fig. 1.16 Structure of a typical β-turn. Turns allow the formation of an anti-parallel sheet structure; in this turn there are ten atoms in the ring formed by the hydrogen bond.

turn which differ in the number of amino acids in the loop, the types of hydrogen bonds formed, and the values of the angles ϕ and ψ for the various amino acids. The most frequently occurring β-turns consist of four amino acids (Fig. 1.16).

An analysis of the crystal structures of a number of proteins of different structural types showed that a significant number (approximately 5%) of the amino acid residues could be assigned to poly-Pro II structures which were at least three amino acids in length (Sreerama and Woody, 1994).

As previously noted, Pro cannot be accommodated in regular secondary structural elements such as the α-helix and β-sheet. Pro residues are often found in β-turns or at the ends of helices or strands of sheets. Chains of Pro residues can also adopt regular structures, thus poly-Pro can exist in two structures poly-Pro I (all peptide bonds *cis*) and poly-Pro II (all peptide bonds *trans*). The angles ϕ and ψ for poly-Pro I are −83° and +158°, respectively, and for poly-Pro II are −78° and +149°, respectively. Poly-Pro I is a right-handed helix with 3.3 residues per turn and a span of 0.19 nm per residue, whereas poly-Pro II is a highly extended left-handed helix with three residues per turn and a span of 0.31 nm per residue.

1.4.4 Structural preferences of the different amino acids

The 20 amino acids have side chains which differ in terms of their bulkiness, polarity, and charge properties. These will lead in turn to different preferences for the various types of secondary structures in proteins. These preferences have been quantified in two main ways, firstly from an analysis of the structures of actual proteins and secondly from studies of the effect of incorporating different amino acids into regular polymers (e.g. introducing a Val residue into poly-Ala, which has a strong tendency to form an α-helix). From analysis of the structures of proteins, tables of preferences for the amino acids have been compiled. Met, Glu, Leu, and Ala have the highest tendency to be found in α-helices, whereas Pro, Gly, and Tyr have the lowest. Val, Ile, and Phe have the highest tendencies to be found in β-sheets; Pro and Asp the lowest. Pro, Gly, and Asp have the highest tendencies to be found in β-turns; Met, Val, and Ile the lowest. While detailed structural explanations for these preferences are not always clear, such data are useful in formulating 'rules' for predicting secondary structural features from protein sequences. It should be remembered that these rules are only guidelines; in general helices can be predicted with more confidence than sheets or turns, but even so the accuracy is relatively modest (of the order of 70%).

A commonly used tool for predicting helix, strand, and loop regions of a protein from the amino acid sequence is PHDsec (PredictProtein@EMBL-Heidelberg.de), which is stated to achieve 72% accuracy when applied to water-soluble, globular proteins.

1.5 The tertiary structure of proteins

KEY CONCEPTS

- Understanding the general principles governing the formation of the tertiary structure of proteins
- Describing how protein structures can be classified

The tertiary structure of a protein refers to its long-range folding so that parts of the polypeptide chain remote in sequence are brought into close proximity.

1.5.1 General principles

Although proteins adopt a very wide of three-dimensional structures, a number of general principles have emerged from the analysis of the many thousands of known structures of proteins. These are summarized below.

- *Close packing* Proteins generally adopt very closely packed globular structures with only a small number of internal cavities, which are generally filled by water molecules
- *Elements of secondary structure* The structural elements (helices, strands, turns etc.) are generally similar to those seen in small model compounds, but there can be deviations. For example, there is a tendency for α-helices to be distorted towards 3_{10} helices, and for strands to adopt a twist in forming sheets
- *Distributions of side chains* Non-polar side chains tend to be buried in the interior of the protein away from the solvent, and polar side chains exposed to the solvent. Polar side chains which are buried will often play key functional roles. This tendency would, of course, be reversed for membrane-spanning elements of a protein, where the non-polar side chains of a helix will project into the non-polar interior of the membrane
- *Pairing of polar groups* Polar groups which are internal are paired in hydrogen bonds; this can include carbonyl and amino groups of the main chain and pairing with internal water molecules
- *Formation of domains* Larger protein molecules tend to form structural domains, which are independent folded globular units, typically of the order of 100–150 amino acids in size. These domains are usually associated with particular functions, e.g. binding or catalysis, so that it is possible to assemble a multifunctional protein in this modular fashion.

In many cases it is possible to prepare the individual domains either by controlled proteolysis of the multi-domain protein or by expression of the appropriate part of the gene coding for that domain.

1.5.2 Classification of protein structures

A number of classification schemes have been devised to characterize the tertiary structures of proteins. These generally place the structures into one of three main categories, namely all-α, all-β, and αβ. (In some schemes, the αβ class is further divided into α/β and α + β, depending whether the helix and sheet elements

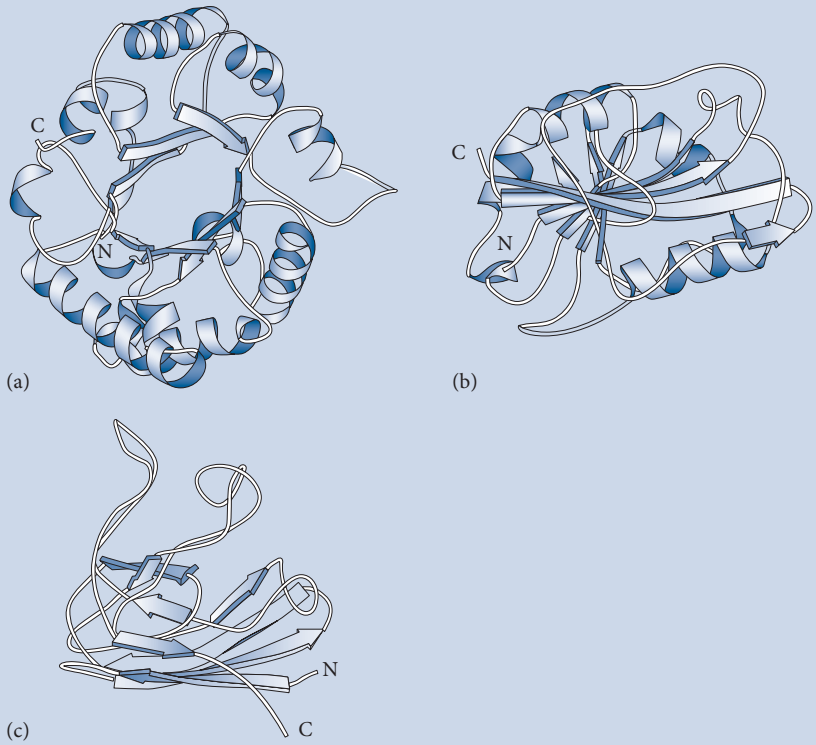


Fig. 1.17 Examples of protein superfolds. (a), (b), and (c) show examples of the TIM barrel, α/β doubly wound, and Greek key structures respectively; these are three of the most commonly occurring superfolds. α-Helices and β-sheets are shown as ribbons and arrows respectively.

alternate along the chain or are in separate regions of the sequence). One of the most useful tools for the classification of protein is the SCOP (Structural Classification Of Proteins) database, which classifies proteins at three levels: (a) protein folds, where there are similar arrangements of secondary structural elements, but no particularly close relationship in terms of function or overall structure, (b) protein superfamilies, where there are probable relationships between the proteins, as shown by similar functions, such as binding nucleotides, and (c) protein families, where there is very strong evidence for close relationships with a high (>30%) level of amino acid sequence identity between members.

The TIM barrel is named after the enzyme triosephosphate isomerase in the glycolytic pathway, the first protein structure shown to possess the repeating $(\alpha\beta)_8$ pattern.

The number of different protein folds has been estimated to be of the order of 1000. Analysis shows that only nine fold families contain proteins with no sequence or functional similarity to each other. These so-called superfolds account for over 30% of all known structures; they include the TIM barrel, α/β doubly wound, and Greek key structures (see Fig. 1.17).

1.5.3 Forces involved in stabilizing tertiary structures

The forces involved in stabilizing tertiary structures of proteins are (apart from any disulphide bonds which may be formed in extracellular proteins) of the weak non-covalent type and are discussed in section 1.7. The same types of forces are involved in the interaction of proteins with each other or with other ligands.

1.6 The quaternary structure of proteins

KEY CONCEPT

- Understanding the significance of association of polypeptide chains (subunits) in terms of the function of proteins

As noted in section 1.5, there is a tendency in polypeptide chains of more than 100–150 amino acids to form domains. For larger proteins, there is a tendency to exist as multiple subunits (polypeptide chains); this generates the quaternary structure of proteins. Proteins of molecular mass <30 kDa tend to be monomers (single subunits) and those of molecular mass >50 kDa tend to be oligomers (several subunits), although these figures should be viewed as general guidelines only. In general the arrangement of the subunits maximizes the number of inter-subunit contacts, for example in a tetrameric protein (four subunits) the preferred geometrical arrangement will be tetrahedral rather than square planar or linear. The subunits in a multi-subunit protein are generally held together by the weak non-covalent forces of the type described in section 1.7, although it should be noted that in a few cases, such as immunoglobins (Chapter 7, section 7.3.6) and bovine seminal ribonuclease, a disulphide bond is involved in holding the subunits together.

The association of several subunits in a protein can offer possibilities of novel properties to a protein. These include (a) regulation of activity, by communication between the binding sites or active sites on different subunits (e.g. in the case of O_2 binding to haemoglobin), (b) increased stability as a result of favourable packing of subunits which might be prone to unfolding as monomers, and (c) generation of novel structures such as found in chaperone proteins and proteasomes, where internal 'cavities' can be created to allow protein folding or protein degradation, respectively.

A protein consisting only of multiple copies of the same polypeptide chain is termed a 'homooligomer', e.g. lactate dehydrogenase is a homotetramer (denoted α_4). If there are different types of polypeptide chain present the protein is termed a heterooligomer, e.g. the G-proteins involved in signal transduction are heterotrimers (denoted $\alpha\beta\gamma$). Haemoglobin is a heterotetramer ($\alpha_2\beta_2$), where the α and β chains are very similar to each other. The Greek letters are used in a generic sense to distinguish different types of polypeptide chains in a protein, rather than to denote specific sequences of amino acids.

The proteasome is a large complex (1.4 MDa) that contains a barrel-shaped proteolytic unit (700 kDa) consisting of 14 copies of each of two types of subunit. The proteasome degrades proteins which have been tagged with the small protein ubiquitin for destruction.

Chaperones are proteins which assist the folding of proteins in the cell by preventing unwanted interactions between hydrophobic surfaces, which would otherwise lead to the formation of protein aggregates. In the cpn60 class of chaperones, which contains 14 copies of a 57 kDa subunit, there is a large internal cavity that allows the correct folding of polypeptide chains to occur.

1.7 Forces contributing to the structures and interactions of proteins

KEY CONCEPTS

- Understanding the contribution of weak forces to the structures and interactions of proteins
- Describing the basis of the principal types of weak forces (ionic interactions, hydrogen bonds, van der Waals interactions, and hydrophobic interactions)
- Knowing the range of energies involved in protein interactions

The amino acids in a protein are linked together by strong covalent peptide bonds (section 1.3.1). In some cases the three-dimensional structures will also be stabilized by covalent bonds, for example disulphide bonds, which can form between pairs of Cys side chains, and the pairs of Lys side chains, which (after oxidative

deamination) can form strong cross-links in collagen. These covalent bonds have energies of the order of several hundred kJ mol^{-1}. However, the forces which stabilize the folded structures of the vast majority of proteins, and which underpin the interactions between proteins and other molecules, are of a non-covalent type and hence much weaker and usually short lasting. These forces can be described under the following categories.

1.7.1 Ionic (electrostatic) interactions

The effective dielectric constant (D) in the interior of proteins has been estimated from the effects of mutations on the pKa values of amino acid side chains. In the case of proteins such as cytochrome c and subtilisin, it has been estimated that D is in the range 30 to 60, but in staphylococcal nuclease, D is estimated to be 12. The value of D appears to depend on the extent of water penetration into the interior of the protein. In a truly non-polar region in the protein interior, D may be as low as about 4 (Dwyer *et al.*, 2000).

These refer to the electrostatic forces between oppositely charged groups, e.g. between negatively charged side chains (Asp, Glu) and positively charged side chains (Arg, Lys, and (under certain circumstances) His). The energy of interaction (E) between charges of magnitude q_1 and q_2, separated by a distance r is given by $E = q_1q_2/Dr$, where D is the dielectric constant. By definition, $D = 1$ for a vacuum; the value of D for water = 78.5. Non-polar solvents have low values of D (e.g. benzene, $D = 2.3$; ethanol, $D = 25$). Since the majority of charged amino acid side chains in a protein are on the surface and therefore significantly exposed to the aqueous medium, ionic interactions are generally weak (a high value of D means a low value for the energy), typically of the order of 5 kJ mol^{-1}. However, if ionic interactions occur in the much less polar interior of a protein (as in the case of the α-amino group of Ile 16 and the side chain of Asp 194 in chymotrypsin), the forces will be considerably stronger (possibly up to about 20 kJ mol^{-1}).

1.7.2 Hydrogen bonds

A hydrogen bond is essentially a form of electrostatic interaction involving partial charges. It occurs when an H attached to an electronegative atom (almost always O or N in biological systems) interacts with a second electronegative atom. For example, if we call these atoms A (donor) and B (acceptor), we would have the arrangement shown in Fig. 1.18, where the δ indicates a partial charge. There is an attractive force between the partial charges on the H and on the acceptor atom. This is a weak interaction (typically of the order of 5–10 kJ mol^{-1}), but when a number of hydrogen bonds are involved they can contribute significantly to local stability, as is the case in secondary structural features such as the α-helix or β-sheet (section 1.4). The particular importance of hydrogen bonds in terms of the structure and function of proteins is that they impose geometrical constraints, and thus specificity, on interactions. For maximum stability of the hydrogen bond, the three nuclei (A, H, and B) should be in a straight line and the separation between the nuclei A and B should be within fairly narrow limits (0.30 ± 0.05 nm depending on the atoms concerned). (This is how hydrogen bonds are identified in X-ray crystal protein structures, since the H atom cannot be 'seen' by X-rays as it has so little electron density compared with the heavier atoms C, N, and O). In proteins,

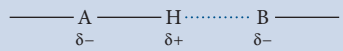


Fig. 1.18 Formation of a hydrogen bond. Atoms A and B represent electronegative atoms (generally O and/or N in biological systems). The partial charges on the atoms are shown.

hydrogen bonds can involve interactions between particular side chains (see section 1.2.4) or main chain atoms (the N–H or C=O groups of the peptide bond). It should also be remembered there is often a compensating mechanism whereby groups on the protein may equally well form hydrogen bonds with water, rather than with other parts of the protein, thereby reducing the energetic contribution of the interaction to the structural or binding properties of the protein.

The strengths of individual hydrogen bonds vary considerably; they are stronger if a charged group is involved, e.g. if $-NH_3^+$ acts as a donor.

1.7.3 van der Waals' interactions

This term is applied collectively to various weak forces involving interactions between dipoles (separations of charge) in molecules or parts of molecules. In addition to permanent dipoles reflecting differences in electronegativity (e.g. between C and O in the C=O bond), there are transient dipoles which arise from the fact that electrons occupy fluctuating positions in space. These dipoles can in turn induce dipoles in other molecules or groups. The van der Waals' forces include dipole⇌dipole, dipole⇌induced dipole, and induced dipole⇌induced dipole interactions. Although the forces are weak individually (of the order of 5 kJ mol^{-1}), there are usually many such interactions and thus collectively they can contribute to the stability of a protein structure or the strength of a protein–ligand interaction.

The van der Waals' forces between atoms or molecules represent a balance between the longer range attractive forces (which show a $1/r^6$ dependence, where r is a measure of internuclear distance) and the shorter range repulsive forces (which show a $1/r^{12}$ dependence).

1.7.4 Hydrophobic interactions

Water is a uniquely poor solvent for non-polar molecules or groups such as the aliphatic or aromatic side chains in proteins. As mentioned in section 1.2.5, liquid water consists of fluctuating clusters of hydrogen-bonded water molecules. When a non-polar group is introduced into this system, the water molecules tend to form more ordered hydrogen bonded 'cage' structures around the non-polar group. This process has a favourable (negative) change in enthalpy, but there is a highly unfavourable (negative) change in entropy due to the ordering of the solvent, thus the overall free energy change is unfavourable (positive) (see Chapter 4, section 4.1). The term 'hydrophobic interactions' is used to describe the tendency of non-polar molecules to be excluded from water and associate with each other, although it should be noted that such interactions are in fact very weak, being largely due to forces of the van der Waals type. For this reason, some workers prefer the term 'hydrophobic effect'.

It should be pointed out that the term 'hydrophobic' is something of a misnomer since it implies that the non-polar groups 'hate water', whereas in fact it is the water which hates them!

The hydrophobic interactions make the most important contribution in energy terms to the stability of the folded structure of a protein in water. Since there is a positive change in enthalpy associated with the exclusion of non-polar groups from water, hydrophobic interactions become weaker at lower temperatures (at least over a restricted range); this can account for the cold-induced denaturation of some proteins.

The strength of hydrophobic interactions can be altered by various agents. For example, urea, and guanidinium chloride (GdmCl) decrease the hydrophobic interactions; since they can participate in hydrogen bonds, they presumably disrupt the hydrogen-bonded network of water molecules. By contrast, the hydrophobic interactions can be strengthened by the addition of various salts; the order of effectiveness follows the Hofmeister series:

The Hofmeister series was derived in the 1880s from studies of the effectiveness of different salts in precipitating proteins from solution.

Cations $Mg^{2+} > Li^{+} > Na^{+} > NH_4^{+}$

Anions $SO_4^{2-} > HPO_4^{2-} >$ acetate $> Cl^{-} > ClO_4^{-} > SCN^{-}$

1.7.5 Balance of energy contributions

Many of the forces involved in stabilizing the three-dimensional structures of proteins and the interactions between proteins and other molecules are relatively weak and represent a balance between competing processes. Thus, for example, hydrophobic interactions represent a balance between an unfavourable enthalpy term but a favourable entropy term, and an α-helix is stabilized by hydrogen bonding between the main chain –N–H and C=O groups, but these are formed at the expense of hydrogen bonds that might be formed to the solvent water if the protein were unfolded. The stability of the folded structure of a protein represents a very delicate balance between two opposing tendencies, namely the large (unfavourable) negative change in entropy when a compact globular structure is formed from a disordered polypeptide chain and the large (favourable) negative change in enthalpy due to the favourable interactions which occur within the folded state of the protein. For most proteins the (thermodynamic) stability of the folded state is very small, in the range 20 – 60 kJ mol^{-1} (see Chapter 8, section 8.6), compared with the entropy and enthalpy terms, each of which is of the order of several hundred kJ mol^{-1}. Of course, there may be a large kinetic barrier (activation energy) to unfolding of the folded state of a protein, but it is important to realize that, in thermodynamic terms at least, life hangs by a slender thread!

The balance of the forces can be altered by addition of certain agents. As indicated in section 1.7.4, urea and GdmCl weaken hydrophobic interactions. This is shown by the fact that different amino acid side chains are much more nearly equally soluble in 8 M urea or 6 M GdmCl than would be the case in water. Since the hydrophobic interactions make the largest contribution in energy terms to the stability of the folded states of proteins, it is easy to understand how these agents promote the unfolding of proteins.

In concentrated solutions of urea and GdmCl, proteins unfold because the entropy gain in unfolding is dominant.

Proteins can also be unfolded by exposure to extremes of pH (reflecting changes in the ionization states of amino acid side chains and hence disruption of

electrostatic forces and hydrogen bonds) or temperature (reflecting alterations in the balance between polar and non-polar interactions); see also Chapter 5, section 5.6.1.

1.7.6 The range of energies involved in protein interactions

The function of the vast majority of proteins is to undertake specific interactions with other molecules (ligands). It is instructive to examine the strengths of these interactions in the context of the discussion of the forces involved. Some representative types of interactions are listed in Table 1.4, together with a note of the corresponding standard free energy change at 37°C ($\Delta G°_{310}$), derived using the relationship $-\Delta G° = RT \ln K_{eq}$ (eqn. 4.4; see Chapter 4, section 4.1).

The magnitudes of the free energy changes are such that only a relatively small number of the weak interactions (hydrogen bonds, van der Waals' forces, etc.) are required to account for them. It should be remembered, however, that many interactions involve a balance between forces, as described in section 1.7.5, thus a protein and a ligand may well each form favourable interactions with the solvent that are lost when the protein and ligand form a complex. The net free energy change can therefore be relatively small.

The values of the dissociation constants can be correlated with the functions of the complexes. For example, the extremely tight interaction between the vitamin biotin (which is produced by bacteria) and the egg white protein avidin may play a role in anti-bacterial defence for the developing chick embryo. The interactions involved in forming oligomeric proteins have to be very favourable in order to generate the stability required for a protein to survive and function under cellular conditions. Antibodies and receptors must bind their respective partners sufficiently tightly to allow subsequent events such as complement activation or signal transduction to occur. On the other hand, it is important that enzyme–substrate (and enzyme–product) interactions are not too strong since otherwise the catalytic cycle (consisting of binding of substrate(s), structural changes in the enzyme, conversion of substrate(s) to product(s), and release of product(s)) would be slowed and the enzyme would become an unacceptably inefficient catalyst.

The dissociation constant K_d for an interaction represents the ratio of the rate constants k_{off}/k_{on} (see Chapter 4, section 4.1). Since there is an upper limit for the association rate constant k_{on} (set by the rate of diffusion) of about 10^9 M^{-1} s^{-1}, an estimate can be made of the dissociation rate constant, k_{off}, and hence the life-time of the complex.

Table 1.4 Typical dissociation constants for some interactions involving proteins

Interaction	Typical dissociation constant (K_d) (M)	ΔG^0_{310} (kJ mol^{-1})*
Avidin–biotin	10^{-15}	89
Protein–protein	10^{-10}	59
Antibody–antigen	10^{-9}	53
Receptor–hormone	10^{-7}	42
Enzyme–substrate	10^{-5}	30

*The free energy values are positive as they refer to the dissociation of the complex ($PL \rightleftharpoons P + L$), which would be unfavourable under standard state conditions (1 M concentrations of PL, P, and L). Dissociation would become progressively more favourable as the concentration of PL is lowered. In the cell the concentrations of many complexes are likely to be in the micromolar range.

1.8 Compendium of chemical structures

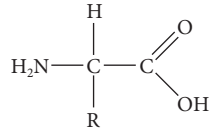
It is a good idea to try to learn the structures of a number of important molecules that are connected with the study of proteins; a list is given below. Once you know the structure of a compound, it will help you to understand its properties. For example, you should be able to recognize the following features of a molecule from its structure:

- Which parts are polar and which are non-polar
- Which parts may be involved in weak interactions such as hydrogen bonds, ionic bonds or hydrophobic interactions
- Which parts may be involved in ionization processes and over which pH range different charged forms may predominate
- Which parts may have important chemical roles, such as acting as a nucleophile or coordinating a metal ion.

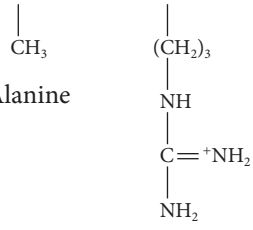
When looking at these structures, bear in mind the following:

- Chains of carbon atoms are often represented as zig-zag lines, rather than each individual carbon atom being displayed
- In complex structures, hydrogen atoms are often omitted for the sake of clarity
- In ring structures the ring atoms are assumed to be C unless otherwise indicated (e.g. N, O, S, etc.)
- The actual charged states of compounds will depend on the prevailing pH and the pK_as of ionizing groups such as $-NH_2$ and $-CO_2H$ (see Chapter 3, section 3.7). They may not be in exactly the forms shown below.

Amino acids

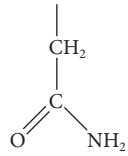


The general structure of an amino acid

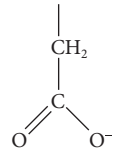


Alanine

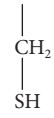
Arginine



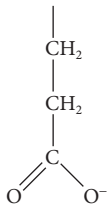
Asparagine



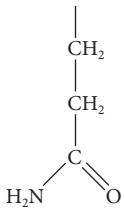
Aspartic acid



Cysteine

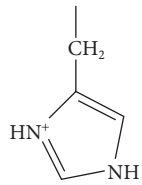


Glutamic acid

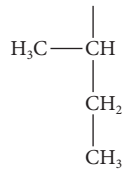


Glutamine

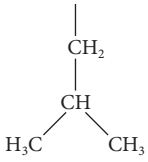
Glycine



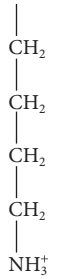
Histidine



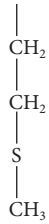
Isoleucine



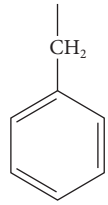
Leucine



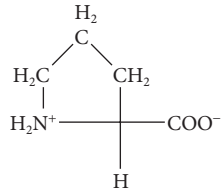
Lysine



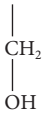
Methionine



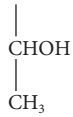
Phenylalanine



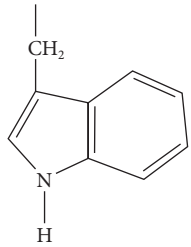
Proline



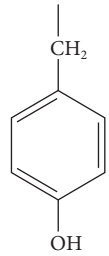
Serine



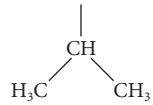
Threonine



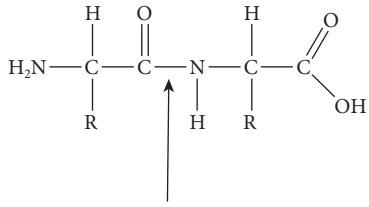
Tryptophan



Tyrosine

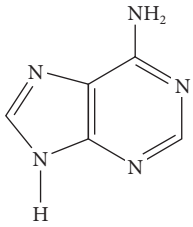


Valine

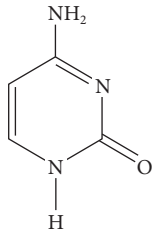


The peptide bond

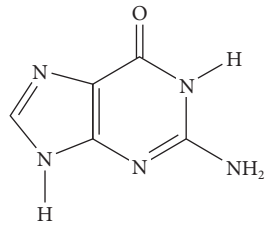
Bases, nucleotides, and related compounds



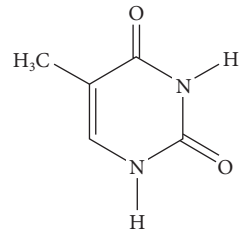
Adenine



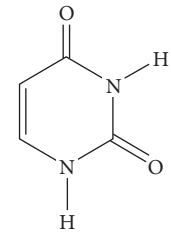
Cytosine



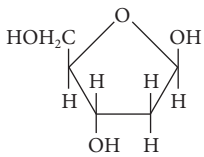
Guanine



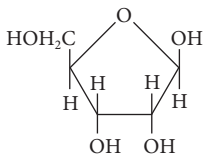
Thymine



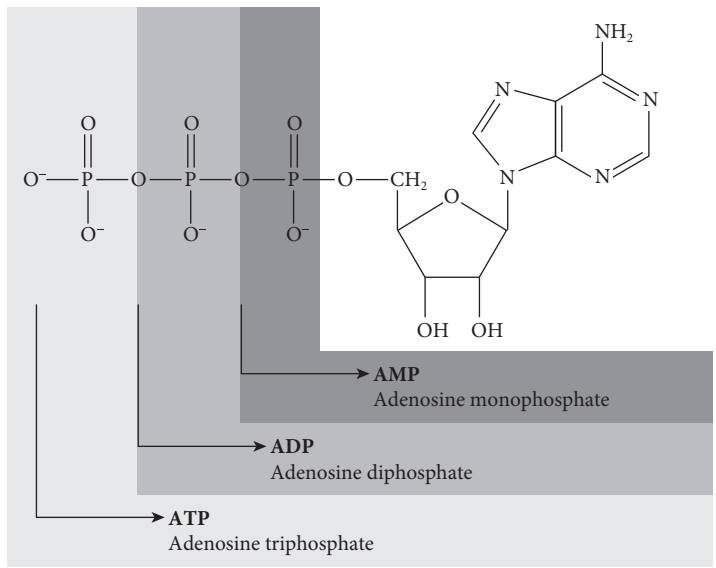
Uracil

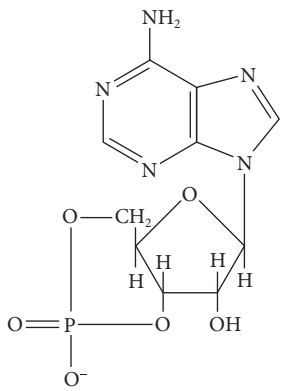


2-deoxyribose

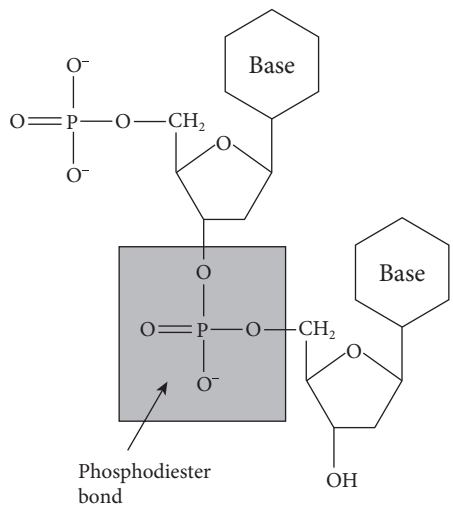


Ribose



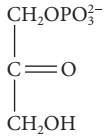


Cyclic AMP

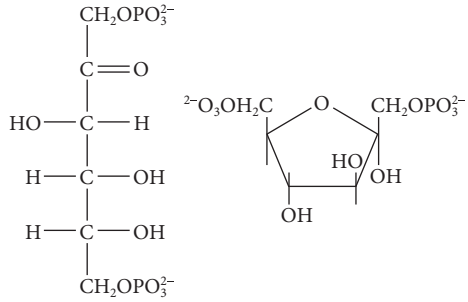


The 5′–3′ bond between nucleotides

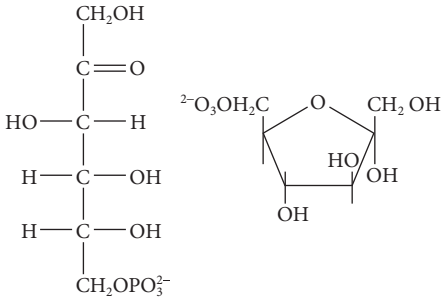
Glycolytic intermediates



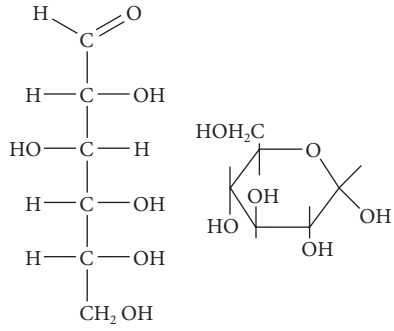
Dihydroxyacetone phosphate



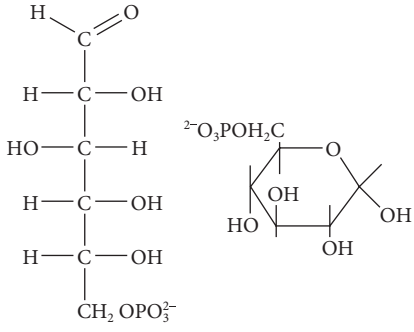
Fructose-1,6-bisphosphate



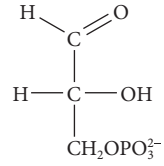
Fructose-6-phosphate



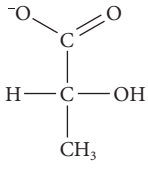
Glucose



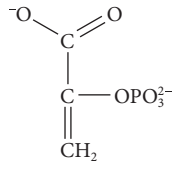
Glucose-6-phosphate



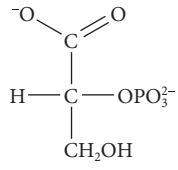
Glyceraldehyde-3-phosphate



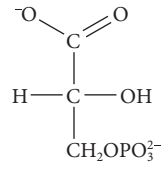
Lactate



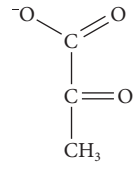
Phosphoenolpyruvate



2-phosphoglycerate

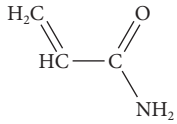


3-phosphoglycerate

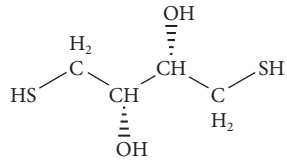


Pyruvate

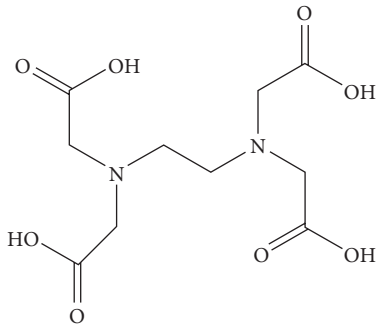
Reagents



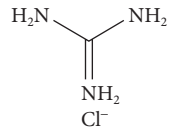
Acrylamide



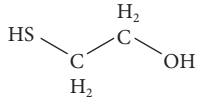
Dithiothreitol



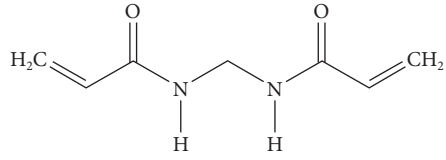
EDTA



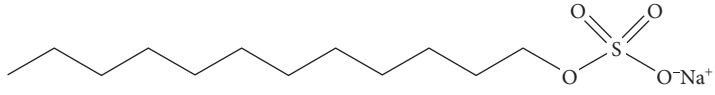
Guanidinium chloride



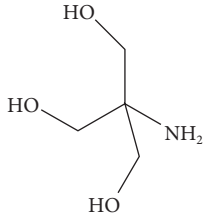
2-mercaptoethanol



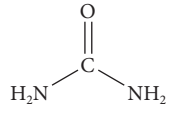
Methylene *bis*-acrylamide



SDS

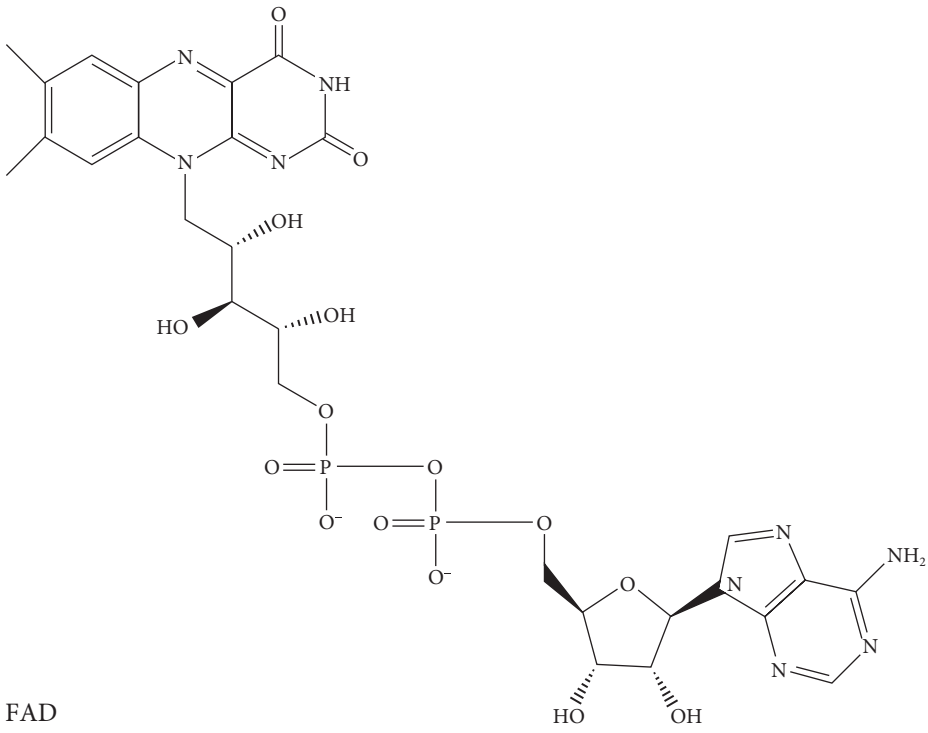


Tris

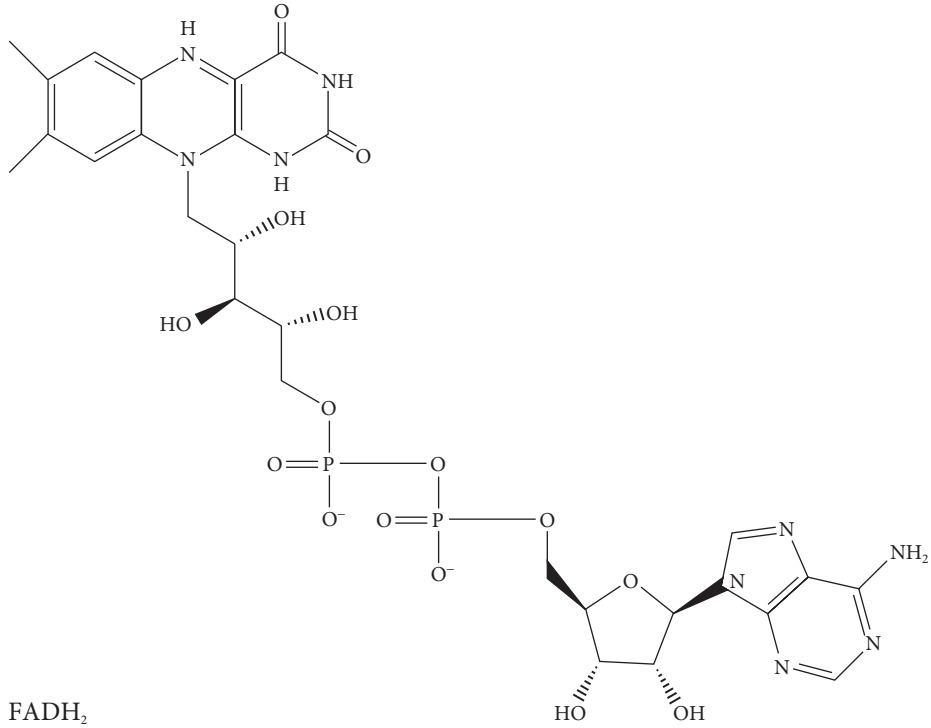


Urea

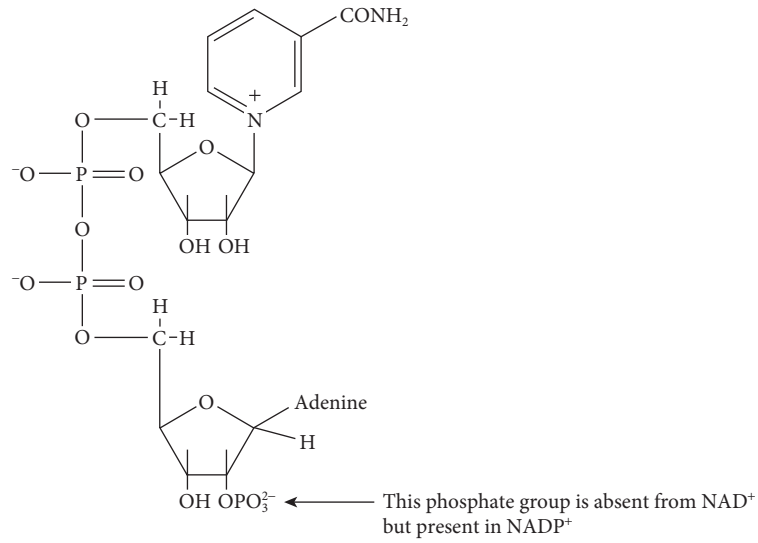
Redox cofactors



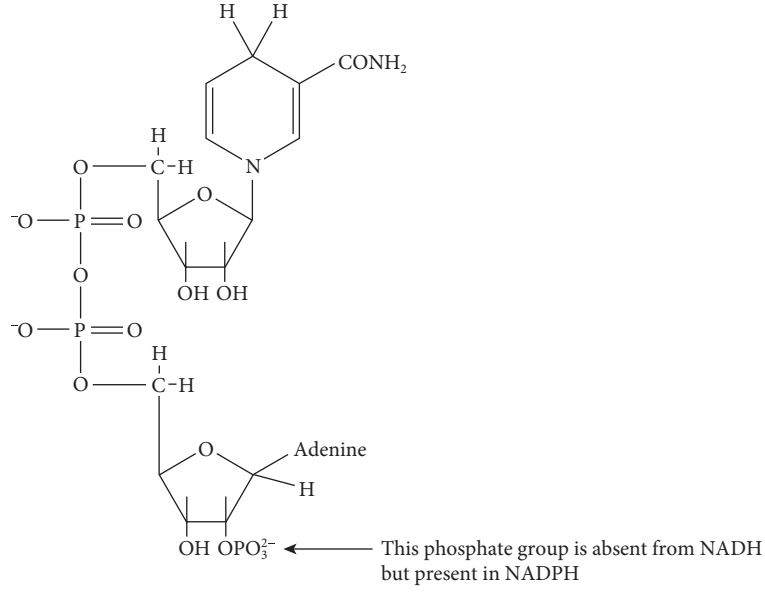
FAD



$FADH_2$

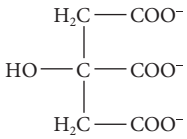


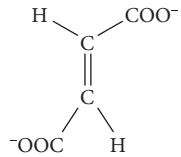
NAD(P)$^+$

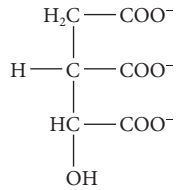


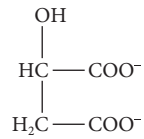
NAD(P)H

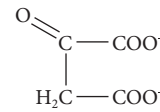
TCA cycle intermediates

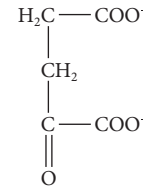

Citrate

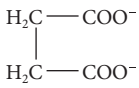

Fumarate


Isocitrate


Malate


Oxaloacetate


2-oxoglutarate


Succinate

1.9 Problems

Full solutions to odd-numbered problems are available to all in the student section of the Online Resource Centre at www.oxfordtextbooks.co.uk/orc/price/. Full solutions to even-numbered problems are available to lecturers only in the lecturer section of the Online Resource Centre.

1.1 The amino acid ornithine (which is an intermediate in the synthesis of arginine) has the side chain $-(CH_2)_3-NH_2$. What would you expect the principal chemical characteristics of this amino acid to be?

1.2 The synthetic amino acid norleucine has the side chain $-(CH_2)_3-CH_3$. How would you expect incorporation of this amino acid in place of leucine to affect the properties of a protein?

1.3 The sequence of the tripeptide glutathione is usually depicted as (γ)Glu–Cys–Gly. Draw the full chemical structure of glutathione. What might happen to glutathione under oxidizing conditions?

1.4 Calmodulin is a member of a family of small (17 kDa) proteins that can bind Ca^{2+} ions tightly. What types of amino acids might allow such proteins to achieve this?

1.5 Compared with Ca^{2+} ions, Zn^{2+} ions have a preference to coordinate with nitrogen and sulphur as compared with oxygen atoms. Which amino acids might you expect to find in Zn-binding sites in proteins?

1.6 For aspartic acid, the pK_a values for the α-carboxyl, the side chain carboxyl and the α-amino group are 2.0, 3.9, and 9.9, respectively. Draw the structures of the predominant charged forms of the amino acid at pH values 1, 3, 6, and 11. Explain why the pI is obtained by averaging the pK_a values of the α-carboxyl and the side chain carboxyl groups.

1.7 For lysine, the pK_a values of the α-carboxyl, the α-amino, and the side chain amino groups are 2.2, 9.1, and 10.5, respectively. What is the pI for lysine?

1.8 Iodoacetamide ($I-CH_2-CO-NH_2$) is a commonly used reagent to react with nucleophilic side chains in proteins. Using structural formulae, show how it reacts with the side chains of cysteine and lysine. How might you monitor the extent of modification of a protein by the reagent?

1.9 The pK_a of the side chain amino group of lysine is 10.5 in the free amino acid. Explain how the pK_a of a lysine side chain in a protein could be affected by the presence of (a) a neighbouring Arg side chain and (b) a neighbouring Asp side chain.

1.10 The following sequences of 20 amino acids occur in the protein glycophorin, which is found in the membrane of red blood cells: (a) YPPEEETGERVQLAHHFSEP and (b) EITLIIFGVMAGVIGTILLI. What would you predict about where these parts of the protein would be located?

1.11 The free energy of the *trans* form of a peptide bond is estimated to be 20 kJ mol^{-1} lower than that of the *cis* form. If a peptide bond could freely interconvert between *trans* and *cis* forms, what would be the equilibrium constant for the *trans* $\rightleftharpoons$ *cis* equilibrium at 37°C? (The gas constant $R = 8.31$ J K^{-1} mol^{-1}).

1.12 In a study of protein structures, it is found that approximately 5% of the peptide bonds to proline (Xaa–Pro) occur in the *cis* form. What is the free energy difference at 37°C between the *trans* and *cis* forms in the case of the Xaa–Pro bond?

1.13 The molar absorption coefficients at 280 nm for small model compounds of Trp and Tyr are 5690 and 1280 M^{-1} cm^{-1} respectively. The polypeptide chain of yeast alcohol dehydrogenase has a molecular mass of 36712 Da and contains 5 Trp and 14 Tyr. Calculate the molar absorption coefficient of alcohol dehydrogenase at 280 nm. (You can assume that the values for the model compounds apply to these amino acids in the protein). What would be the absorbance at 280 nm of a 0.32 mg mL^{-1} solution of the protein in a cuvette of 1-cm pathlength?

1.14 The polypeptide chain of the chaperone protein GroEL from *Escherichia coli* has a molecular mass of 57200 Da and does not contain tryptophan. The A_{280} of a 1 mg mL^{-1} solution of the purified protein in a 1-cm pathlength cuvette is 0.160. Assuming that the molar absorption coefficient at 280 nm for a small model compound of Tyr is 1280 M^{-1} cm^{-1}, calculate the number of Tyr in the polypeptide chain.

1.15 The following amino acid sequences are found in α-helical structures in three folded proteins. Use the helical wheel projection to predict where each of these helical segments may be located in the structure of the protein: (a) LSFAAAMIGLA (citrate synthase), (b) INEGFDLLRSG (alcohol dehydrogenase), and (c) KEDNKGKSEEE (troponin C).

1.16 The calculated molecular mass of the polypeptide chain of type II dehydroquinase from *Streptomyces coelicolor* is 16550.6 Da. What is the expected mass of a mutant form of the enzyme in which Tyr 28 has been replaced by Phe? Compare your answer with the observed mass (16535.1 ± 1.4 Da). In a preparation of a second mutant in which Ser 108 is replaced by Ala, there is a significant amount of material (30% of the total) with a mass some 227 Da lower than the expected value. How could this lower molecular mass material have arisen?

1.17 Draw chemical structures to show the hydrogen bonding which could arise in each case between a Tyr side chain and (a) a second Tyr side chain, (b) the carbonyl group of a peptide bond, and (c) the α-amino group of an N-terminal amino acid.

1.18 From the X-ray structure of the complex between two proteins (X and Y), the following interactions have been identified: (a) Arg 45 of X with Glu 13 of Y, (b) Ile 53 of X with both Val 84 and Phe 85 of Y. Explain the probable basis of these interactions.

1.19 Explain how the strength of ionic interactions and hydrophobic interactions would be affected by changes in the ionic strength of the solution. Use your answers to explain how changes in ionic strength can be used to elute adsorbed proteins in ion-exchange chromatography and hydrophobic interaction chromatography.

1.20 The folded form (F) of ribonuclease T1 is estimated to be 31.4 kJ mol^{-1} more stable than the unfolded form (U) at 25°C. Calculate the equilibrium constant for the F $\rightleftharpoons$ U equilibrium. In the presence of 5.6 M urea, 58.4% of the protein is unfolded; what is the free energy difference between the two forms under these conditions?

References for Chapter 1

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2007) *Biochemistry*, 6th edn. Freeman, New York, 1026 pp.

Crowe, J., Bradshaw, A., and Monk, P. (2006) *Chemistry for the Biosciences: The Essential Concepts*. Oxford University Press, Oxford, 496 pp.

Dwyer, J.J., Gittis, A.G., Karp, D.A., Lattman, E.E., Spencer, D.S., Stites, W.E., and Garcia-Moreno, E.B. (2000) *Biophys. J.* **79**, 1610–20.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 321–53.

Jabs, A., Weiss, M.S., and Hilgenfeld, R. (1999) *J. Mol. Biol.* **286**, 291–304.

Jones, A. (2005) *Chemistry: An Introduction for Medical and Health Sciences.* John Wiley and Sons, Chichester, 260 pp.

Mathews, C.K., van Holde, K.E., and Ahern, K.G. (2000) *Biochemistry*, 3rd edn. Benjamin/Cummings, San Francisco, California, 1186 pp.

Panasik, N., Jr., Fleming, P.J., and Rose, G.D. (2005) *Protein Sci.* **14**, 2910–14.

Price, N.C., Dwek, R.A., Ratcliffe, R.G., and Wormald, M.R. (2001) *Principles and Problems in Physical Chemistry for Biochemists*, 3rd edn. Oxford University Press, Oxford, 401 pp.

Sreerama, N. and Woody, R.W. (1994) *Biochemistry* **33**, 10022–25.

Voet, D., Voet, J.G., and Pratt, C.W. (2006) *Fundamentals of Biochemistry*, 2nd edn. John Wiley and Sons, Hoboken, New Jersey, 1130 pp.

Weaver, T.M. (2000) *Protein Sci.* **9**, 201–6.