2

Amino acids: the building blocks of proteins

Despite enormous functional diversity all proteins consist of a linear arrangement of amino acid residues assembled together into a polypeptide chain. Amino acids are the 'building blocks' of proteins and in order to understand the properties of proteins we must first describe the properties of the constituent 20 amino acids. All amino acids contain carbon, hydrogen, nitrogen and oxygen with two of the 20 amino acids also containing sulfur. Throughout this book a colour scheme based on the CPK model (after Corey, Pauling and Kultun, pioneers of 'space-filling' representations of molecules) is used. This colouring scheme shows nitrogen atoms in blue, oxygen atoms in red, carbon atoms are shown in light grey (occasionally black), sulfur is shown in yellow, and hydrogen, when shown, is either white, or to enhance viewing on a white background, a lighter shade of grey. To avoid unnecessary complexity 'ball and stick' representations of molecular structures are often shown instead of space-filling models. In other instances cartoon representations of structure are shown since they enhance visualization of organization whilst maintaining clarity of presentation.

The 20 amino acids found in proteins

In their isolated state amino acids are white crystalline solids. It is surprising that crystalline materials form the building blocks for proteins since these latter molecules are generally viewed as 'organic'. The crystalline nature of amino acids is further emphasized by their high melting and boiling points and together these properties are atypical of most organic molecules. Organic molecules are not commonly crystalline nor do they have high melting and boiling points. Compare, for example, alanine and propionic acid – the former is a crystalline amino acid and the other is a volatile organic acid. Despite similar molecular weights (89 and 74) their respective melting points are 314 °C and −20.8 °C. The origin of these differences and the unique properties of amino acids resides in their ionic and dipolar nature.

Amino acids are held together in a crystalline lattice by charged interactions and these relatively strong forces contribute to high melting and boiling points. Charge groups are also responsible for electrical conductivity in aqueous solutions (amino acids are electrolytes), their relatively high solubility in water and the large dipole moment associated with crystalline material. Consequently amino acids are best viewed as charged molecules that crystallize from solutions containing dipolar ions. These dipolar ions are called zwitterions. A proper representation of amino acids reflects amphoteric behaviour and amino acids are always represented as the zwitterionic state in this

Proteins: Structure and Function by David Whitford

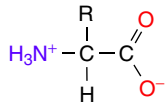


Figure 2.1 A skeletal model of a generalized amino acid showing the amino (blue) carboxyl (red) and R groups attached to a central or α carbon

textbook as opposed to the undissociated form. For 19 of the twenty amino acids commonly found in proteins a general structure for the zwitterionic state has charged amino (NH_3^+) and carboxyl (COO^-) groups attached to a central carbon atom called the α carbon. The remaining atoms connected to the α carbon are a single hydrogen atom and the R group or side chain (Figure 2.1).

The acid–base properties of amino acids

At pH 7 the amino and carboxyl groups are charged but over a pH range from 1 to 14 these groups exhibit a series of equilibria involving binding and dissociation of a proton. The binding and dissociation of a proton reflects the role of these groups as weak acids or weak bases. The acid–base behaviour of amino acids is important since it influences the eventual properties of proteins, permits methods of identification for different amino acids and dictates their reactivity. The amino group, characterized by a basic pK value of approximately 9, is a weak base. Whilst the amino group ionizes around pH 9.0 the carboxyl group remains charged until a pH of ~2.0 is reached. At this pH a proton binds neutralizing the charge of the carboxyl group. In each case the carboxyl and amino groups ionize according to the equilibrium

$$HA + H_2O \longrightarrow H_3O^+ + A^- \qquad (2.1)$$

where HA, the proton donor, is either –COOH or $-NH_3^+$ and A^- the proton acceptor is either $-COO^-$ or $-NH_2$. The extent of ionization depends on the equilibrium constant

$$K = [H^+][A^-]/[HA] \qquad (2.2)$$

and it becomes straightforward to derive the relationship

$$pH = pK + \log[A^-]/[HA] \qquad (2.3)$$

known as the Henderson–Hasselbalch equation (see appendix). For a simple amino acid such as alanine a biphasic titration curve is observed when a solution of the amino acid (a weak acid) is titrated with sodium hydroxide (a strong base). The titration curve shows two zones where the pH changes very slowly after additions of small amounts of acid or alkali (Figure 2.2). Each phase reflects different pK values associated with ionizable groups.

During the titration of alanine different ionic species predominate in solution (Figure 2.3). At low pH (<2.0) the equilibrium lies in favour of the positively charged form of the amino acid. This species contains a charged amino group and an uncharged carboxyl group leading to the overall or *net* charge of +1. Increasing the pH will lead to a point where the concentration of each species is equal. This pH is equivalent to the first pK value (~pH 2.3) and further increases in pH lead to point of inflection, where the dominant

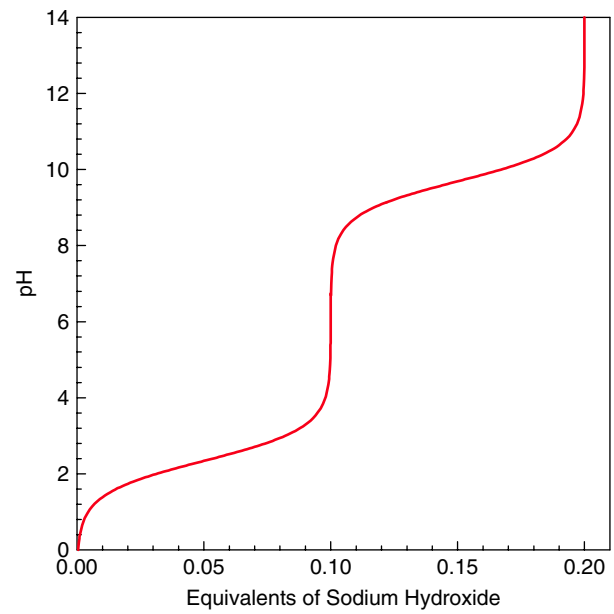


Figure 2.2 Titration curve for alanine showing changes in pH with addition of sodium hydroxide

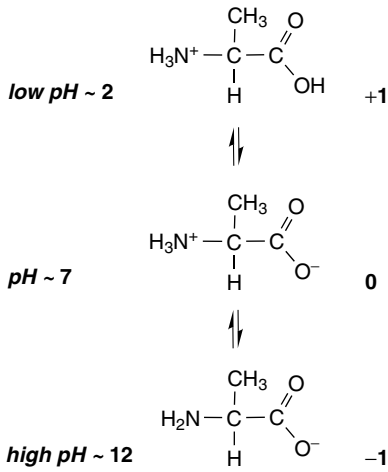


Figure 2.3 The three major forms of alanine occurring in titrations between pH 1 and 14

species in solution is the zwitterion. The zwitterion, although dipolar, has no overall charge and at this pH the amino acid will not migrate towards either the anode or cathode when placed in an electric field. This pH is called the isoelectric point or pI and for alanine reflects the arithmetic mean of the two pK values pI = (pK_1 + pK_2)/2. Continuing the pH titration still further into alkaline conditions leads to the loss of a proton from the amino group and the formation of a species containing an overall charge of −1. The R group may contain functional groups that donate or accept protons and this leads to more complex titration curves. Amino acids showing additional pK values include aspartate, glutamate, histidine, argininine, lysine, cysteine and tyrosine (see Table 2.1).

Amino acids lacking charged side chains show similar values for pK_1 of about 2.3 that are significantly lower than the corresponding values seen in simple organic acids such as acetic acid (pK_1 ~4.7). Amino acids are stronger acids than acetic acid as a result of the electrophilic properties of the α amino group that increase the tendency for the carboxyl hydrogen to dissociate.

Stereochemical representations of amino acids

Although an amino acid is represented by the skeletal diagram of Figure 2.1 it is more revealing, and certainly more informative, to impose a stereochemical view on the arrangement of atoms. In these views an attempt is made to represent the positions in space of each atom. The amino, carboxyl, hydrogen and R groups are arranged tetrahedrally around the central α carbon (Figure 2.4).

Table 2.1 The pK values for the α-carboxyl, α-amino groups and side chains found in the individual amino acids

Amino acid	pK_1	pK_2	pK_R	Amino acid	pK_1	pK_2	pK_R
Alanine	2.4	9.9	–	Leucine	2.3	9.7	–
Arginine	1.8	9.0	12.5	Lysine	2.2	9.1	10.5
Asparagine	2.1	8.7	–	Methionine	2.1	9.3	–
Aspartic Acid	2.0	9.9	3.9	Phenylalanine	2.2	9.3	–
Cysteine	1.9	10.7	8.4	Proline	2.0	10.6	–
Glutamic Acid	2.1	9.5	4.1	Serine	2.2	9.2	–
Glutamine	2.2	9.1	–	Threonine	2.1	9.1	–
Glycine	2.4	9.8	–	Tyrosine	2.2	9.2	10.5
Histidine	1.8	9.3	6.0	Tryptophan	2.5	9.4	–
Isoleucine	2.3	9.8	–	Valine	2.3	9.7	–

Adapted from Dawson, R.M.C, Elliot, W.H., & Jones, K.M. 1986 *Data for Biochemical Research*, 3rd edn. Clarendon Press Oxford.

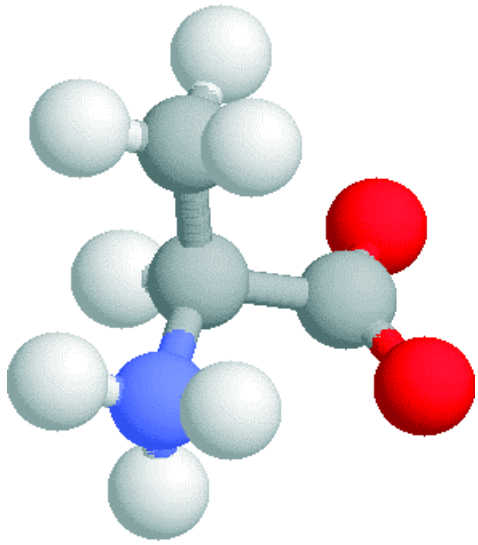


Figure 2.4 The spatial arrangement of atoms in the amino acid alanine

The nitrogen atom (blue) is part of the amino ($-NH_3^+$) group, the oxygen atoms (red) are part of the carboxyl ($-COO^-$) group. The remaining groups joined to the α carbon are one hydrogen atom and the R group.

The R group is responsible for the different properties of individual amino acids. As amino acids make up proteins the properties of the R group contribute considerably to the physical properties of proteins. Nineteen of the 20 amino acids found in proteins have the arrangement shown by Figure 2.4 but for the remaining amino acid, proline, an unusual cyclic ring is formed by the side chain bonding directly to the amide nitrogen (Figure 2.5).

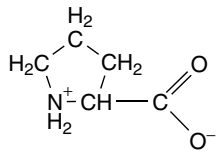


Figure 2.5 The structure of proline – an unusual amino acid containing a five-membered pyrrolidine ring

A glance at the structures of the 20 different side chains reveals major differences in, for example, size, charge and hydrophobicity although the R group is always attached to the α carbon (C2 carbon). From the α carbon subsequent carbon atoms in the side chains are designated as β, γ, δ, ε and ζ. In some databases of protein structures the C_β is written as CB, the C_δ as CD, C_ζ as CZ, etc. Both nomenclatures are widely used. The nomenclature is generally unambiguous but care needs to be exercised when describing the atoms of the side chain of isoleucine. Isoleucine has a branched side chain in which the C_γ or CG is either a methyl group or a methylene group. In this instance the two groups are distinguished by the use of a subscript 1 and 2, i.e. CG1 and CG2. A similar line of reasoning applies to the carbon atoms of aromatic rings. In phenylalanine, for example, the aromatic ring is linked to the C_β atom by the C_γ atom and contains two C_δ and C_ε atoms ($C_{\delta1}$ and $C_{\delta2}$, $C_{\varepsilon1}$ and $C_{\varepsilon2}$) before completing ring at the C_ζ (or CZ) atom.

Peptide bonds

Amino acids are joined together by the formation of a peptide bond where the amino group of one molecule reacts with the carboxyl group of the other. The reaction is described as a condensation resulting in the elimination of water and the formation of a dipeptide (Figure 2.6).

Three amino acids are joined together by two peptide bonds to form a tripeptide and the sequence

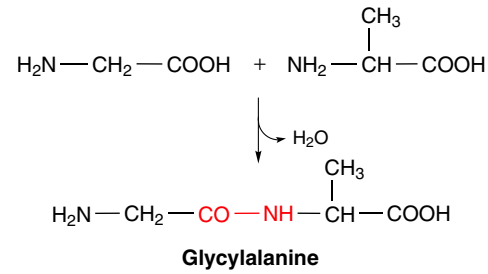


Figure 2.6 Glycine and alanine react together to form the dipeptide glycylalanine. The important peptide bond is shown in red

continues with the formation of tetrapeptides, pentapeptides, and so on. When joined in a series of peptide bonds amino acids are called residues to distinguish between the free form and the form found in proteins. A short sequence of residues is a peptide with the term polypeptide applied to longer chains of residues usually of known sequence and length. Within the cell protein synthesis occurs on the ribosome but today peptide synthesis is possible *in vitro* via complex organic chemistry. However, whilst the organic chemist struggles to synthesize a peptide containing more than 50 residues the ribosome routinely makes proteins with over 1000 residues.

All proteins are made up of amino acid residues linked together in an order that is ultimately derived from the information residing within our genes. Some proteins are clearly related to each other in that they have similar sequences whilst most proteins exhibit a very different composition of residues and a very different order of residues along the polypeptide chain. From the variety of side chains a single amino acid can link to 19 others to create a total of 39 different dipeptides. Repeating this for the other residues leads to a total of 780 possible dipeptide permutations. If tripeptides and tetrapeptides are considered the number of possible combinations rapidly reaches a very large figure. However, when databases of protein sequences are studied it is clear that amino acid residues do not occur with equal frequency in proteins and sequences do not reflect even a small percentage of all possible combinations. Tryptophan and cysteine are rare residues (less than 2 percent of all residues) in proteins whilst alanine, glycine and leucine occur with frequencies between 7 and 9 percent (see Table 2.2).

Amino acid sequences of proteins are read from left to right. This is from the amino or N terminal to the carboxyl or C terminal. The individual amino acids have three-letter codes, but increasingly, in order to save space in the presentation of long protein sequences, a single-letter code is used for each amino acid residue. Both single- and three-letter codes are shown alongside the R groups in Table 2.2 together with some of the relevant properties of each side chain. Where possible the three-letter codes for amino acids will be used but it should be stressed that single letter codes avoid potential confusion. For example Gly, Glu and Gln are easily mistaken when rapidly reading protein sequences but their single letter codes of G, E and Q are less likely to be misunderstood.

Joining together residues establishes a protein sequence that is conveniently divided into main chain and side chain components. The main chain, or polypeptide backbone, has the same composition in all proteins although it may differ in extent – that is the number of residues found in the polypeptide chain. The backbone represents the effective repetition of peptide bonds made up of the N, C_α and C atoms, with proteins such as insulin having approximately 50 residues whilst other proteins contain over 1000 residues and more than one polypeptide chain (Figure 2.7). Whilst all proteins link atoms of the polypeptide backbone similarly the side chains present a variable component in each protein.

Properties of the peptide bond

The main chain or backbone of the polypeptide chain is established by the formation of peptide bonds between amino acids. The backbone consists of the amide N, the α-carbon and the carbonyl C linked together (Figure 2.8).

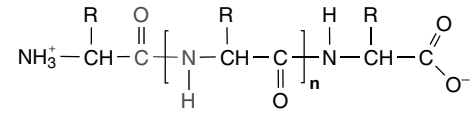


Figure 2.7 Part of a polypeptide chain formed by the covalent bonding of amino acids where *n* is often 50–300, although values above and below these limits are known.

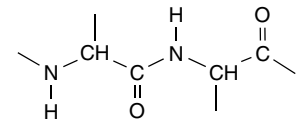


Figure 2.8 The polypeptide backbone showing arrangement of i, i + 1 residues within a chain.

Table 2.2 The frequencies with which amino acid residues occur in proteins

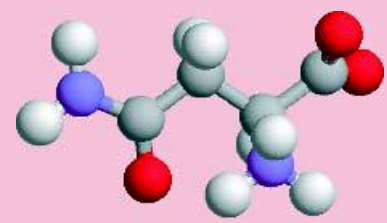
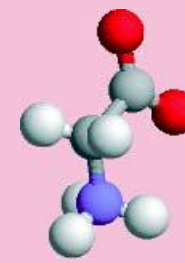
Amino acid	Property of individual amino acid residues	Ball and stick representation of each amino acid
Alanine A Ala M_r 71.09	Non-polar side chain. Small side chain volume. Van der Waals volume = 67 Å^{3*} Frequency in proteins = 7.7 % Surface area = 115 Å^2 Unreactive side chain	
Arginine R Arg M_r 156.19	Positively charged side chain at pH 7.0. pK for guanidino group in proteins ~12.0 Van der Waals volume = 167 Å^3 Frequency in proteins = 5.1 % Surface area = 225 Å^2 Participates in ionic interactions with negatively charged groups	
Asparagine N Asn M_r 114.11	Polar, but uncharged, side chain Van der Waals volume = 148 Å^3 Frequency in proteins = 4.3 % Surface area = 160 Å^2 Polar side chain will hydrogen bond Relatively small side chain volume leads to this residue being found relatively frequently in turns	
Aspartate D Asp M_r 115.09	Negatively charged side chain pK for side chain of ~4.0 Van der Waals volume = 67 Å^3 Frequency in proteins = 5.2 % Surface area = 150 Å^2 Charged side chain exhibits electrostatic interactions with positively charged groups.	

Table 2.2 (*continued*)

Amino acid	Property of individual amino acid residues	Ball and stick representation of each amino acid
Cysteine C Cys M_r 103.15	Side chain contains thiol (SH) group. Van der Waals volume = 86 Å^3 Frequency in proteins = 2.0 % Surface area = 135 Å^2 Thiol side chain has pK in isolated amino acid of ~8.5 but in proteins varies 5–10 Thiol group is very reactive	
Glutamine Q Gln M_r 128.12	Polar but uncharged side chain Van der Waals volume = 114 Å^3 Frequency in proteins = 4.1 % Surface area = 180 Å^2 Polar side chain can hydrogen bond	
Glutamate E Glu M_r 129.12	Negatively charged side chain. Van der Waals volume = 109 Å^3 Frequency in proteins = 6.2 % Surface area = 190 Å^2 Side chain has pK of ~4.5.	
Glycine G Gly M_r 57.05	Uncharged, small side chain. Often found in turn regions of proteins or regions of conformational flexibility No chiral centre; due to two hydrogens attached to C_α centre Van der Waals volume = 48 Å^3 Frequency in proteins = 7.4 % Surface area = 75 Å^2	

(*continued overleaf*)

Table 2.2 (*continued*)

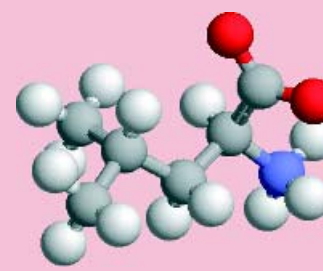
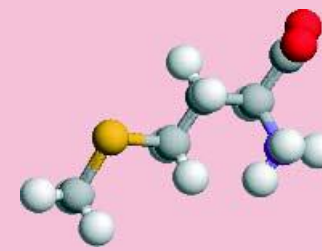
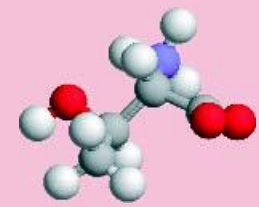
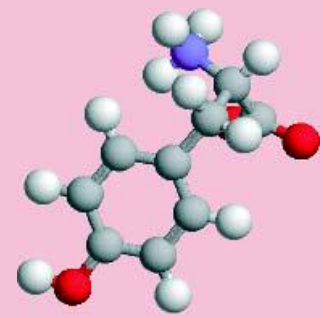
Amino acid	Property of individual amino acid residues	Ball and stick representation of each amino acid
Histidine H His M_r 137.14	Imidazole side chain Van der Waals volume = 118 Å^3 Frequency in proteins = 2.3 % Surface area = 195 Å^2 The side chain exhibits a pK ~ 6.0 in model peptides but in proteins can vary from 4–10	
Isoleucine I Ile M_r 113.16	Hydrophobic side chain exhibiting non-polar based interactions but generally unreactive Van der Waals volume = 124 Å^3 Frequency in proteins = 5.3 % Surface area = 175 Å^2	
Leucine L Leu M_r 113.16	Hydrophobic side chain Van der Waals volume = 124 Å^3 Frequency in proteins = 8.5 % Surface area = 170 Å^2	
Lysine K Lys M_r 128.17	Positively charged side chain Van der Waals volume = 135 Å^3 Frequency in proteins = 5.9 % Surface area = 200 Å^2 Side chain is basic with pK of ~10.5. Shows ionic interactions	

Table 2.2 (*continued*)

Amino acid	Property of individual amino acid residues	Ball and stick representation of each amino acid
Methionine M Met M_r 131.19	Sulfur containing hydrophobic side chain The sulfur is unreactive especially when compared with thiol group of cysteine Van der Waals volume = 124 Å^3 Frequency in proteins = 2.4 % Surface area = 185 Å^2	
Phenylalanine F Phe M_r 147.18	Hydrophobic, aromatic side chain Phenyl ring is chemically unreactive in proteins. Exhibits weak optical absorbance around 280 nm Van der Waals volume = 135 Å^3 Frequency in proteins = 4.0 % Surface area = 210 Å^2	
Proline P Pro M_r 97.12	Cyclic ring forming hydrophobic side chain The cyclic ring limits conformational flexibility around N-C_α bond In a polypeptide chain lacks amide hydrogen and cannot form backbone hydrogen bonds Van der Waals volume = 90 Å^3 Frequency in proteins = 5.1 % Surface area = 145 Å^2	
Serine S Ser M_r 87.08	Polar but uncharged side chain. Contains hydroxyl group (–OH) that hydrogen bonds Oxygen atom can act as potent nucleophile in some enzymes Van der Waals volume = 73 Å^3 Frequency in proteins = 6.9 % Surface area = 115 Å^2	

(*continued overleaf*)

Table 2.2 (*continued*)

Amino acid	Property of individual amino acid residues	Ball and stick representation of each amino acid
Threonine T Thr M_r 101.11	Polar but uncharged side chain. Contains hydroxyl group (–OH) Hydrogen bonding side chain Van der Waals volume = 93 Å^3 Frequency in proteins = 5.9 % Surface area = 140 Å^2	
Tryptophan W Trp M_r 186.21	Large, hydrophobic and aromatic side chain Almost all reactivity is based around the indole ring nitrogen Responsible for majority of near uv absorbance in proteins at 280 nm Van der Waals volume = 163 Å^3 Frequency in proteins = 1.4 % Surface area = 255 Å^2	
Tyrosine Y Tyr M_r 163.18	Aromatic side chain Van der Waals volume = 141 Å^3 Frequency in proteins = 3.2 % Surface area = 230 Å^2 Phenolic hydroxyl group ionizes at pH values around pH 10 Aromatic ring more easily substituted than that of phenylalanine	
Valine V Val M_r 99.14	Hydrophobic side chain Van der Waals volume = 105 Å^3 Frequency in proteins = 6.6 % Surface area = 155 Å^2	

From Jones, D.T. Taylor, W.R. & Thornton, J.M. (1991) *CABIOS* **8**, 275–282. Databases of protein sequences are weighted towards globular proteins but with the addition of membrane proteins to databases a gradual increase in the relative abundance of hydrophobic residues such as Leu, Val, Ile, Phe, Trp is expected. The surface area was calculated for an accessible surface of residue X in the tripeptide G-X-G (Chothia, C. (1975) *J. Mol. Biol.*, **105**, 1–14). Volumes enclosed by the van der Waals radii of atoms as described by Richards, F.M. (1974) *J. Mol. Biol.* **82**, 1–14.

*1 Å = 0.1 nm.

The linear representation of the polypeptide chain does not convey the intricacy associated with the bond lengths and angles of the atoms making up the peptide bond. The peptide bond formed between the carboxyl and amino groups of two amino acids is a unique bond that possesses little intrinsic mobility. This occurs because of the partial double bond character (Figure 2.9)–a feature associated with the peptide bond and resonance between two closely related states.

One of the most important consequences of resonance is that the peptide bond length is shorter than expected for a simple C–N bond. On average a peptide bond length is 1.32 Å compared to 1.45 Å for an ordinary C–N bond. In comparison the average bond length associated with a C=N double bond is 1.25 Å, emphasizing the intermediate character of the peptide bond. More importantly the partial double bond between carbon and nitrogen atoms restricts rotation about this bond. This leads to the six atoms shown in Figure 2.9 being coplanar; that is all six atoms are found within a single imaginary plane. For any polypeptide backbone represented by the sequence –N–C_α–C–N–C_α–C–only the C_α–C and N–C_α bonds exhibit rotational mobility. As a result of restricted motion about the peptide bond two conformations related by an angle of 180° are possible. The first occurs when the C_α atoms are *trans* to the peptide bond whilst the second and less favourable orientation occurs when the C_α atoms are *cis* (Figure 2.10).

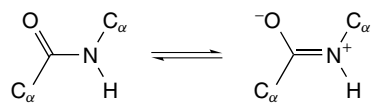


Figure 2.9 The peptide bond may be viewed as a partial double bond as a result of resonance

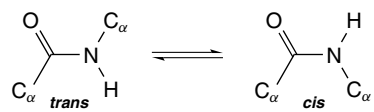


Figure 2.10 *Cis* and *trans* configurations are possible about the rigid peptide bond

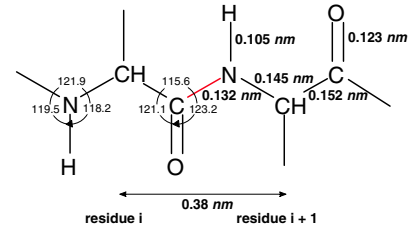


Figure 2.11 Detailed bond lengths and angles for atoms of the polypeptide backbone

The *trans* form is the more favoured state because in this arrangement repulsion between non-bonded atoms connected to the C_α centre are minimized. For most peptide bonds the ratio of *cis* to *trans* configurations is approximately 1:1000. However, one exception to this rule is found in peptide bonds where the following residue is proline. Proline, unusual in having a cyclic side chain that bonds to the backbone amide nitrogen, has less repulsion between side chain atoms. This leads to an increase in the relative stability of the *cis* peptide bond when compared with the *trans* state, and for peptide bonds formed between Xaa and proline (Xaa is any amino acid) the *cis* to *trans* ratio is 1:4.

The dimensions associated with the peptide bond have generally been obtained from crystallographic studies of small peptides and besides the peptide bond length of 0.132 nm other characteristic bond lengths and angles have been identified (Figure 2.11). One of the most important dimensions is the maximum distance between corresponding atoms in sequential residues. In the *trans* peptide bond and a fully extended conformation this distance is maximally 0.38 nm although in proteins it is often much less.

The chemical and physical properties of amino acids

Chemical reactions exhibited by the 20 amino acids are extensive and revolve around the reactivity of amino and carboxyl groups. In proteins, however, these groups are involved in peptide bonds and the defining properties of amino acids are those associated with side chains.

Glycine

Glycine is the simplest amino acid containing two hydrogen atoms attached to the central α carbon. Consequently it lacks an asymmetric centre and does not occur as R/S isomers (see below). More importantly the absence of any significant functional group means that glycine possesses little intrinsic chemical reactivity. However, the absence of a large side chain results in conformational flexibility about the N–C_α and C_α–C bonds in a polypeptide chain. This fact alone has important consequences for the overall structure of proteins containing significant numbers of glycine residues.

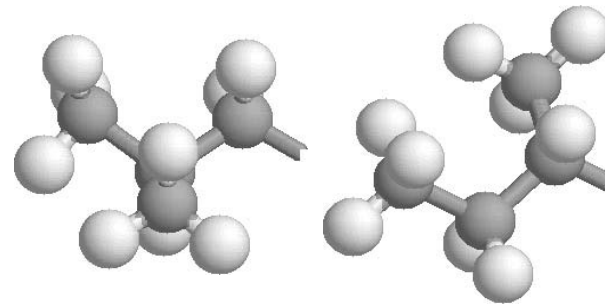


Figure 2.13 The structures of the side chains of leucine (left) and isoleucine (right)

Aliphatic side chains: alanine, valine, isoleucine and leucine

Alanine, like glycine, has a very simple side chain – a single methyl group (CH_3) that is chemically inert and inactive (Figure 2.12). It is joined in this remarkable non-reactivity by the side chains of valine, leucine and isoleucine (Figure 2.13). The side chains of these amino acids are related in containing methine (CH) or methylene ($–CH_2$) groups in branched aliphatic chains terminated by methyl groups. However, one very important property possessed by these side chains is their unwillingness to interact with water – their hydrophobicity. The side chains interact far more readily with each other and the other non-polar side chains of amino acids such as tryptophan or phenylalanine. In later chapters on tertiary structure, protein stability and folding the important role of the weak hydrophobic interaction in maintaining the native state of proteins is described. Alanine, valine, leucine and isoleucine will all exhibit hydrophobic interactions. Isoleucine and leucine as their names suggest are isomers differing in the arrangement of methylene and methyl groups. The first carbon atom in the side chain (CB) is a methylene (CH_2) group in leucine whilst in isoleucine it is a methine (–CH–) moiety. The CB of leucine is linked to a methine group (CG) whilst in isoleucine the CB atom bonds to a methyl group *and* a methylene group. The side chain of valine (Figure 2.14) has one methylene group less than leucine but is otherwise very similar in property.

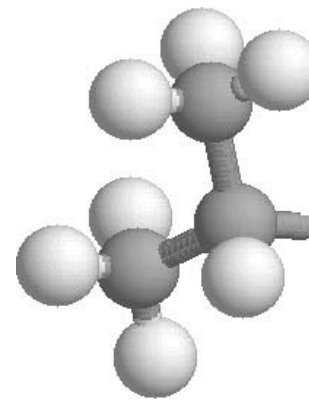


Figure 2.14 The side chain of valine is branched at the CB (–CH or methine) group and leads to two methyl groups

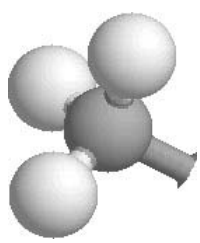


Figure 2.12 The side chain of alanine is the unreactive methyl group ($–CH_3$)

The hydroxyl-containing residues: serine and threonine

The side chains of serine and threonine are characterized by the presence of a polar hydroxyl (–OH) group (Figure 2.15). Their side chains are generally small (Table 2.2). Although small the side groups of both serine and threonine hydrogen bond with other residues in proteins whilst in an isolated state their reactivity is confined to that expected for a primary alcohol, with esterification representing a common reaction with organic acids. One of the most important reactions to a serine or threonine side chain is the addition of a phosphate group to create phosphoserine or phosphothreonine. This occurs as a post-translational modification in the cell after protein synthesis on the ribosome and is important in protein–protein interactions and intracellular signalling pathways.

In one group of enzymes – the serine proteases – the activity of a single specific serine side chain is enhanced by its proximity to histidine and aspartyl side chains. By itself the side chain of Ser is a weak nucleophile but combined with His and Asp this triad of residues becomes a potent catalytic group capable of splitting peptide bonds.

The acidic residues: aspartate and glutamate

Aspartic acid and glutamic acid have side chains with a carboxyl group (Figure 2.16). This leads to the side chain having a negative charge under physiological conditions and hence to their description as acidic side chains. For this reason the residues are normally referred to as aspartate and glutamate reflecting the ionized and charged status under most cellular conditions. The two side chains differ by one methylene group (CH_2) but are characterized by similar pK values (range 3.8–4.5).

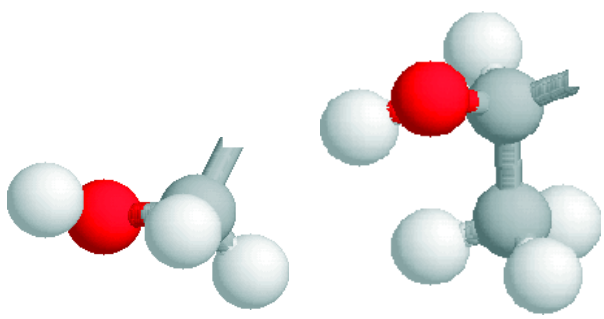


Figure 2.15 The side chains of serine and threonine are aliphatic containing hydroxyl groups

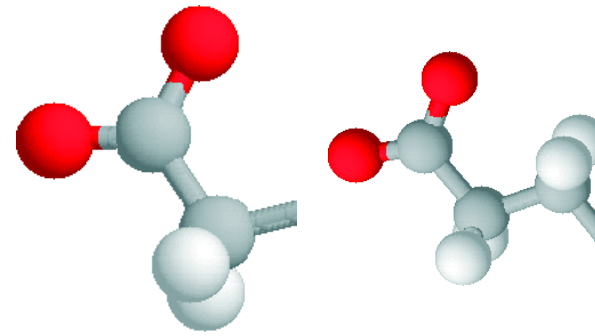


Figure 2.16 The acidic side chains of aspartate and glutamate

The side chains behave as typical organic acids and exhibit a wide range of chemical reactions including esterification with alcohols or coupling with amines. Both of these reactions have been exploited in chemical modification studies of proteins to alter one or more aspartate/glutamate side chains with a loss of the negatively charged carboxyl group. As expected the side chains of aspartate and glutamate are potent chelators of divalent metal ions and biology exploits this property to bind important ions such as Ca^{2+} in proteins such as calmodulin or Zn^{2+} in enzymes such as carboxypeptidase.

The amide-containing residues: asparagine and glutamine

Asparagine and glutamine residues are often confused with their acidic counterparts, the side chains of aspartate and glutamate. Unlike the acidic side chains the functional group is an amide – a generally unreactive group that is polar and acts as hydrogen bond donor and acceptor (see Figure 2.17). The amide group is labile at alkaline pH values, or extremes of temperature, being deamidated to form the corresponding acidic side

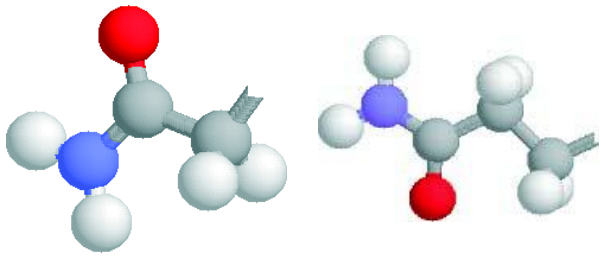


Figure 2.17 The uncharged side chains of asparagine and glutamine

chain. This reaction can occur during protein isolation and occasionally leads to protein sequences containing the three letter code Glx or Asx where the identity of a residue, either Gln or Glu and Asp or Asn, is unclear.

The sulfur-containing residues: cysteine and methionine

Sulfur occurs in two of the 20 amino acids: cysteine and methionine. In cysteine the sulfur is part of a reactive thiol group whilst in methionine sulfur is found as part of a long, generally unreactive, side chain. The side chain of methionine is non-polar (Figure 2.18) and is larger than that of valine or leucine although it is unbranched. Its properties are dominated by the presence of the sulfur atom, a potent nucleophile under certain conditions.

The sulfur atom is readily methylated using methyl iodide in a reaction that is often used to introduce a 'label' onto methionine residue via the use of ^{13}C labelled reactant. In addition the sulfur of methionine interacts with heavy metal complexes particularly those involving mercury and platinum such as K_2PtCl_4 or

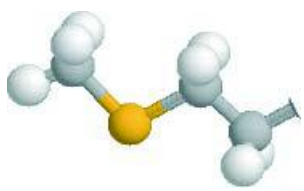


Figure 2.18 The side chain of methionine

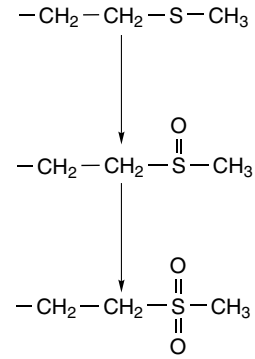


Figure 2.19 Oxidation of methionine side chains by strong oxidizing agents

$HgCl_2$ and these have proved extremely useful in the formation of isomorphous heavy atom derivatives in protein crystallography (see Chapter 10). The sulfur atom of methionine can be oxidized to form first a sulfoxide and finally a sulfone derivative (Figure 2.19). This form of oxidative damage is known to occur in proteins and the reaction scheme involves progressive addition of oxygen atoms.

One of the most important reactions of methionine involves cyanogen bromide – a reagent that breaks the polypeptide chain on the C-terminal side of methionine residues by sequestering the carbonyl group of the next peptide bond in a reaction involving water and leading to formation of a homoserine lactone (Figure 2.20). This reaction is used to split polypeptide chains into smaller fragments for protein sequencing. When compared with cysteine, however, the side chain of methionine is less reactive undergoing comparatively few important chemical reactions.

The functional group of cysteine is the thiol (–SH) group, sometimes called the sulfhydryl or mercapto group. It is the most reactive side chain found amongst the 20 naturally occurring amino acids undergoing many chemical reactions with diverse reagents (Figure 2.21). Some enzymes exploit reactivity by using a conserved cysteine residue at their active sites that participates directly in enzyme-catalysed reactions.

The large sulfur atom as part of the thiol group influences side chain properties significantly, with disulfide bonds forming between cysteine residues that are close

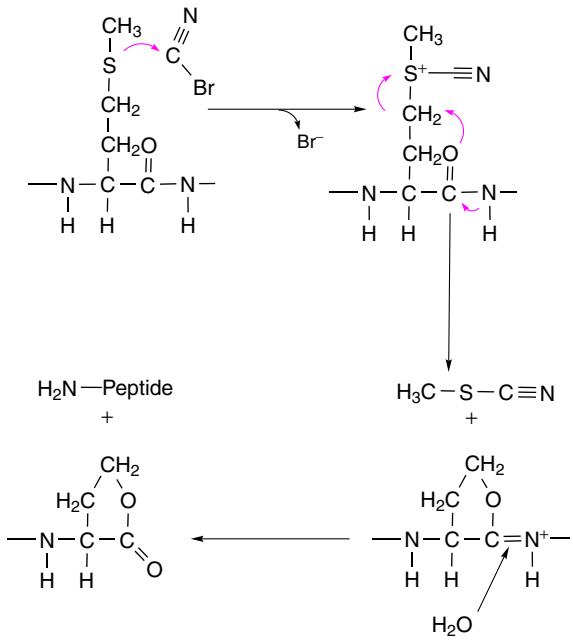


Figure 2.20 The reaction of cyanogen bromide with methionine residues

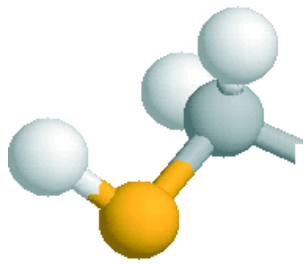


Figure 2.21 The thiol group in the side chain of cysteine is the most reactive functional group found in amino acid residues

together in space. This forms a strong covalent bond and exercises considerable conformational restraint on the structure adopted in solution by polypeptides. Formation of a disulfide bridge is another example of post-translational modification (Figure 2.22).

The thiol group ionizes at alkaline pH values (~pH 8.5) to form a reactive thiolate anion (S^-). The

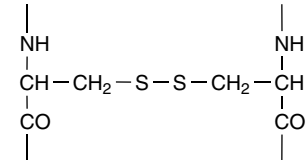


Figure 2.22 Formation of a disulfide bridge between two thiol side chains

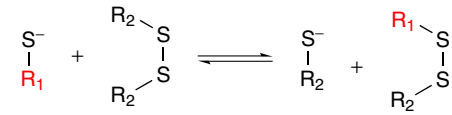


Figure 2.23 The formation of mixed disulfides via the interaction of thiolate groups. A thiolate anion reacts with other symmetrical disulfides to form a mixed disulfide formed between R_1 and R_2

thiolate anion reacts rapidly with many compounds but the most important includes other thiols or disulfides in exchange type reactions occurring at neutral to alkaline pH values. A general reaction scheme for thiol-disulfide exchange is given in Figure 2.23.

A common reaction of this type is the reaction between cysteine and Ellman's reagent (dithionitrobenzoic acid, DTNB; Figure 2.24). The aromatic disulfide undergoes exchange with reactive thiolate anions forming a coloured aromatic thiol – nitrothiobenzoate. The benzoate anion absorbs intensely at 416 nm allowing the concentration of free thiol groups to be accurately estimated in biological systems.

Thiols are also oxidized by molecular oxygen in reactions catalysed by trace amounts of transition metals, including Cu and Fe. More potent oxidants such as performic acid oxidize the thiol group to a sulfonate (SO_3^{2-}) and this reaction has been exploited as a method of irreversibly breaking disulfide bridges to form two cysteic acid residues. More frequently the disulfide bridge between cysteine residues is broken by reducing agents that include other thiols such as mercaptoethanol or dithiothreitol (Figure 2.25) as well as more conventional reductants such as sodium borohydride or molecular hydrogen. Dithiothreitol,

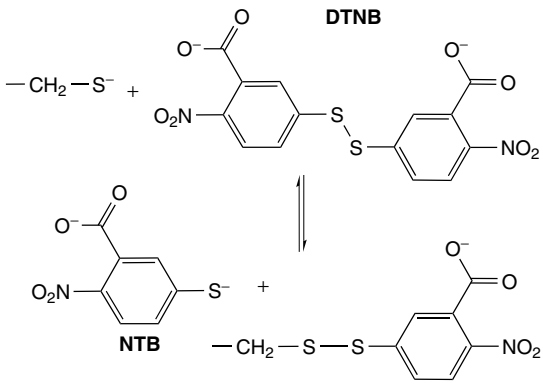


Figure 2.24 The reaction between the thiolate anion of cysteine and Ellman's reagent

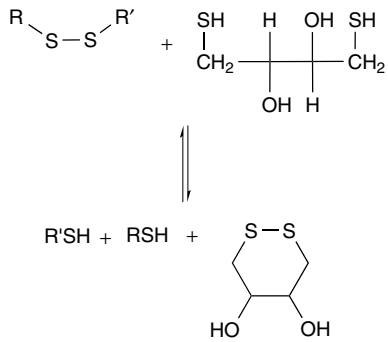


Figure 2.25 The reaction between dithiothreitol and disulfide groups leads to reduction of the disulfide and the formation of two thiols

sometimes called Cleland's reagent, reacts with disulfide groups to form initially a mixed disulfide intermediate but this rapidly rearranges to yield a stable six-membered ring and free thiol groups (–SH). The equilibrium constant for the reaction lies over to the right and is largely driven by the rapid formation of cyclic disulfide and its inherent stability.

The basic residues: lysine and arginine

The arginine side chain (Figure 2.26) contains three methylene groups followed by the basic guanidino (sometimes called guanadinium) group, which is usually protonated, planar and with the carbon atom exhibiting sp^2 hybridization.

The guanidino group is the most basic of the side chains found in amino acids with a pK of 12 and under almost all conditions the side chain retains a net positive charge (Figure 2.27). The positive charge is distributed over the entire guanidino group as a result of resonance between related structures

Lysine possesses a long side chain of four methylene groups terminated by a single ε amino group (Figure 2.28). The amino group ionizes with a pK of approximately 10.5–11.0 and is very basic. As expected the side chain interacts strongly in proteins with oppositely charged side chains but will also undergo methylation, acetylation, arylation and acylation. Many of these reactions are performed at high pH (above pH 9.0) since the unprotonated nitrogen is a potent nucleophile reacting rapidly with suitable reagents. One of the most popular lysine modifications

Figure 2.26 The side chain of arginine is very basic and contains a guanidino group

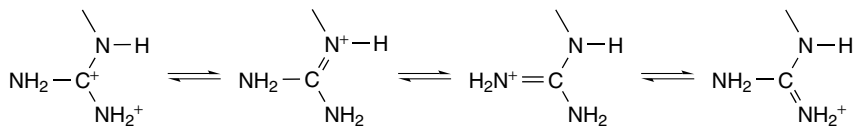


Figure 2.27 Charge delocalization and isomerization within the guanidino group

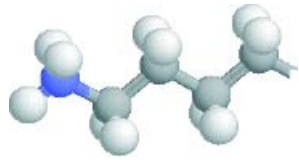


Figure 2.28 The lysine side chain is basic

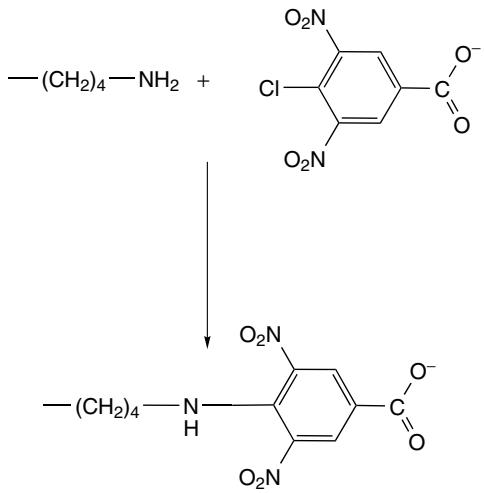


Figure 2.29 Reaction of lysine side chains with dinitrobenzene derivatives

involves adding a nitrobenzene derivative to the ε-amino group. The nitrobenzene group is coloured and the rate of the reaction is therefore comparatively easily followed spectrophotometrically. In the example shown in Figure 2.29 4-chloro-3, 5-dinitrobenzoic acid reacts with lysine side chains to form a negatively charged dinitrophenol derivative. Methylation, unlike the arylation reaction described above, preserves the positive charge on the side chain and in some systems, particularly fungi, trimethylated lysine residues are found as natural components of proteins.

One of the most important reactions occurring with lysine side chains is the reaction with aldehydes to form a Schiff base (see Figure 2.30). The reaction is important within the cell because pyridoxal phosphate, a co-factor derived from vitamin B_6, reacts with the ε amino group of lysine and is found in many enzyme

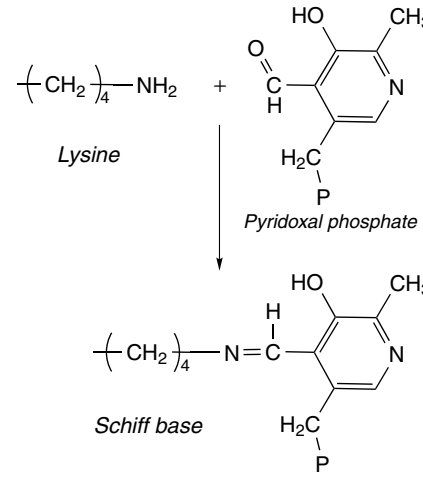


Figure 2.30 Reaction between pyridoxal phosphate and lysine side chains results in formation of a covalent Schiff base intermediate. The aldehyde group of pyridoxal phosphate links with the ε amino group of a specific lysine residue at the active site of many enzymes

active sites. Pyridoxal phosphate is related to vitamin B_6, pyridoxine.

Proline

The side chain of proline is unique in possessing a side chain that covalently bonds with the backbone nitrogen atom to form a cyclic pyrrolidine ring with groups lacking reactivity. One of the few reactions involving prolyl side chains is enzyme-catalysed hydroxylation. The cyclic ring imposes rigid constraints on the N-C_α bond leading to pronounced effects on the configuration of peptide bonds preceding proline. In addition the ring is puckered with the C_γ atom displaced by 0.5 Å from the remaining atoms of the ring which show approximate co-planarity.

Histidine

The imidazole side chain of histidine (Figure 2.31) is unusual and alone amongst the side chains of amino acids in exhibiting a pK around 7.0. In its ionized

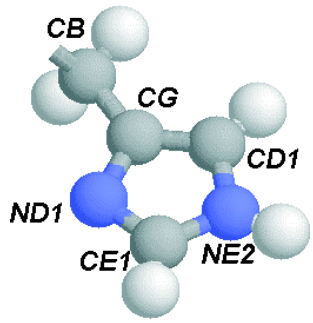


Figure 2.31 The side chain of histidine showing nomenclature for side chain atoms. The IUPAC scheme encourages the use of *pros* (meaning 'near' and abbreviated with the symbol π) and *tele* (meaning 'far', abbreviated with the symbol τ) to show their position relative to the main chain

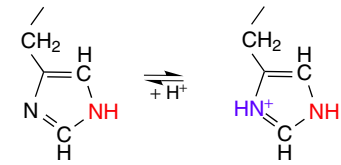


Figure 2.32 Protonation of the imidazole side chain leads to a positively charge

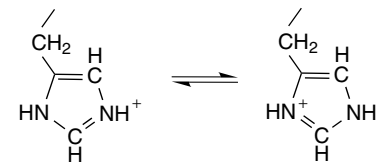


Figure 2.33 Delocalization of charge in protonated imidazole side chains of histidine is possible

state the side chain has a positive charge whilst in the unionized state the side chain remains neutral. However, protonation not only modulates the charge and acid–base behaviour of the side chain but also alters nucleophilic and electrophilic properties of the ring (Figure 2.32).

The protonated nitrogen, shown in red in Figure 2.32, is called the NE2 or ε2. Experimental evidence suggests that the hydrogen atom is usually located on the NE2 nitrogen but upon further protonation the structure on the right is formed with the ND1 nitrogen now binding a proton. The charge cannot be accurately assigned to one of the nitrogen atoms and resonance structures exist (Figure 2.33).

The unprotonated nitrogen of the uncharged imidazole ring is a potent nucleophile and has a capacity for hydrogen bonding.

The aromatic residues: phenylalanine, tyrosine and tryptophan

The aromatic side chains have a common property of absorbing in the ultraviolet region of the electromagnetic spectrum. Table 2.3 shows the spectroscopic properties of Phe, Tyr and Trp. As a result Phe, Tyr and Trp are responsible for the absorbance and fluorescence of proteins frequently measured between

Table 2.3 Spectroscopic properties of the aromatic amino acids

Amino acid	Absorbance		Fluorescence	
	λ_{max} (nm)	ε (M^{-1} cm^{-1})	λ_{max} (nm)	Quantum yield
Phenylalanine	257.4	197	282	0.04
Tyrosine	274.6	1420	303	0.21
Tryptophan	279.8	5600	348	0.20

λ, wavelength; ε, molar absorptivity coefficient.

250 and 350 nm. In this region the molar extinction coefficients of Phe, Tyr and Trp, an indication of how much light is absorbed at a given wavelength, are not equal. Tryptophan exhibits a molar extinction coefficient approximately four times that of tyrosine and over 28 times greater than phenylalanine. Almost all spectrophotometric measurements of a protein's absorbance at 280 nm reflect the intrinsic Trp content of that protein. At equivalent molar concentrations proteins with a high number of tryptophan residues will give a much larger absorbance at 280 nm when compared with proteins possessing a lower Trp content.

Phenylalanine

The aromatic ring of phenylalanine is chemically inert and has generally proved resistant to chemical modification (Figure 2.34). Only with genetic modification of proteins have Phe residues become routinely altered. The numbering scheme of atoms around the aromatic benzene ring is shown for carbon and hydrogens. Alternative nomenclatures exist for the ring proteins with labels of HD, HE or HZ (H_δ, H_ε or H_ζ in some schemes) being the most common. In many proteins rotation of the aromatic ring around the CB–CG axis leads to the protons at the 2,6 positions (HD) and the 3,5 positions (HE) being indistinguishable.

Although the aromatic ring is inert chemically it remains hydrophobic and prefers interactions in proteins with other non-polar residues. With other aromatic rings Phe forms important $\pi-\pi$ interactions derived from the delocalized array of electrons that arises from bonding between the ring carbon atoms. In a simple alkene such as ethene each carbon links to two hydrogens and a single carbon. This is achieved by sp^2 hybridization where the normal ground state of carbon $1s^2 2s^2 2p_x^{\,1} 2p_y^{\,1}$ is transformed by elevating one of the 2s electrons into the vacant p_z orbital yielding $1s^2 2s^1 2p_x^{\,1} 2p_y^{\,1} 2p_z^{\,1}$. The sp^2 hybrids, made from rearrangement of the 2s orbital and two of the three 2p orbitals, are orientated at an angle of 120° to each other within a plane. The remaining $2p_z$ orbital is at right angles to these hybrids. Formation of three bonds leaves the p_z orbital vacant and a similar situation exists in benzene or the aromatic ring of phenylalanine; each sp^2 hybridized carbon is bonded to two carbon atoms and a single hydrogen and has a single unpaired electron. This electron lies in the vacant p_z orbital proximal to and overlapping with another carbon atom possessing an identical configuration. The extensive overlap produces a system of π bonds where the electrons are not confined between two carbon atoms but delocalize around the entire ring. Six delocalized electrons go into three π molecular orbitals – two in each. The interaction of these π molecular orbitals by the close approach of two aromatic rings leads to further delocalization. The coplanar orientation is the basis for $\pi-\pi$ interactions and may account for the observation that aromatic side chains are found in proteins in close proximity to other aromatic groups. $\pi-\pi$ interactions are expected to make favourable and significant contributions to the overall protein stability. Studies with model compounds suggest that an optimal geometry exists with the aromatic rings perpendicular to each other and leading to positively charged hydrogens on the edge of one ring interacting favourably with the π electrons and partially negatively charged carbons of the other. For proteins, as opposed to model compounds, this type of interaction is less common than a co-planar orientation of aromatic rings.

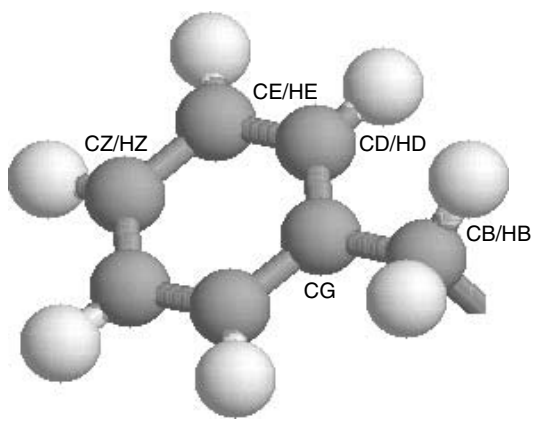


Figure 2.34 The aromatic side chain of phenylalanine

Tyrosine

Tyrosine is more reactive than phenylalanine due to the hydroxyl group (Figure 2.35) substituted at the fourth (C_Z) aromatic carbon, often called the *para*

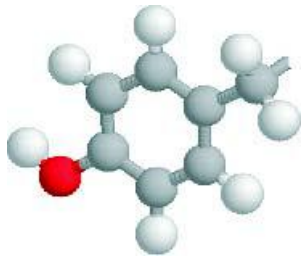


Figure 2.35 The side chain of tyrosine contains a reactive hydroxyl group

position. The hydroxyl group leaves the aromatic ring susceptible to substitution reactions.

Nucleophiles such as nitrating agents or activated forms of iodide react with tyrosine side chains in proteins and change the acid–base properties of the ring. Reaction with nitrating agents yields mono nitrotyrosine derivatives at the 3,5 position followed by dinitro derivatives. Normally the hydroxyl group is a weak acid with a pK $\sim$ 11.0 but the addition of one or more nitro groups causes the pK to drop by up to 4 pH units. The tyrosine interacts with other aromatic rings but the presence of the OH group also allows hydrogen bonding.

Tryptophan

The indole side chain (Figure 2.36) is the largest side chain occurring in proteins and is responsible for most of the intrinsic absorbance and fluorescence. As a crude approximation the molar extinction coefficient of a protein at 280 nm may be estimated by adding up the number of Trp residues found in the sequence and multiplying by 5800 (see Table 2.3). However, the relatively low frequency of Trp residues in proteins means that this approach is not always accurate and in some cases (for proteins lacking Trp) will be impossible.

The side chain is hydrophobic and does not undergo extensive chemical reactions. Exceptions to the rule of unreactivity include modifying reagents such as iodine, *N*-bromosuccinimide and ozone although their use in chemical modification studies of proteins is limited. Universally the weakest or most sensitive part of the indole ring is the pyrrole nitrogen atom.

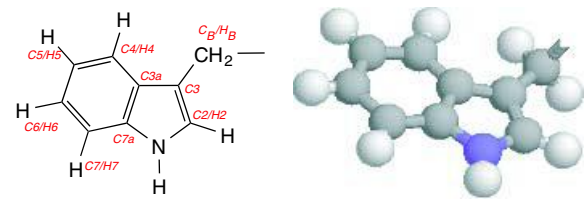


Figure 2.36 The indole ring of tryptophan. The nomenclature of the ring protons names the ring protons as H2, H4, H5, H6 and H7. The carbons with protons attached are similarly named i.e C2, C3 . . . C7. The carbon centres lacking protons are C3, C3a and C7a. The nitrogen is properly termed the N1 centre but is frequently called the indole, or NE1, nitrogen. In addition Greek symbols are often applied to these carbon and protons with the following symbols used $C_{\xi 2}$ = C7, $C_{\eta 2}$ = C6, $C_{\xi 3}$ = C5, $C_{\eta 3}$ = C4

Detection, identification and quantification of amino acids and proteins

To quantify protein isolation requires a means of estimating protein concentration and most methods use the properties of amino acid side chains. Proteins absorb around 280 nm principally because of the relative contributions of tryptophan, phenylalanine and tyrosine. If their occurrence in proteins is known then, in theory at least, the concentration of a protein solution is calculated using the data of Table 2.4 and Beer–Lamberts law where

$$A_{280} = \varepsilon_{280} c\ l \qquad (2.4)$$

(where A is absorbance, ε is molar absorptivity coefficient, c is the concentration in moles dm^{-3} and l is the light path length normally 1 cm). In practice this turns out to be a very crude method of measuring protein concentration and of limited accuracy. The number of Trp, Tyr or Phe residues is not always known and more importantly the absorbance of proteins in their native state does not usually equate with a simple summation of their individual contributions. Moreover the presence of any additional components absorbing at 280 nm will lead to inaccuracies and this could include the presence of impurities, disulfide

bridges, or additional co-factors such as heme or flavin. Performing spectrophotometric measurements under denaturing conditions allows improved estimation of protein concentration from the sum of the Trp, Tyr and Phe components. The contribution of disulfide bridges, although weak around 280 nm, can be eliminated by the addition of reducing agents, and under denaturing conditions co-factors are frequently lost.

An alternative method of determining protein concentration exploits the reaction between the thiol group of cysteine and Ellman's reagent. The reaction produces the nitrothiobenzoate anion and since the molar absorptivity coefficient for this product at 410 nm is accurately known (13 600 $M^{-1}\,cm^{-1}$) the reagent offers one route of determining protein concentration if the number of thiol groups is known beforehand. In general, protein concentration may be determined by any reaction with side chains that lead to coloured products that can be quantified by independent methods. Other reagents with applications in estimating protein, peptide or amino acid concentration include ninhydrin, fluorescamine, dansyl chloride, nitrophenols and fluorodinitrobenzene. They have the common theme of possessing a reactive group towards certain side chains combined with a strong chromophore. A common weakness is that all rely on reactions with a restricted number of side chains to provide estimates of total protein concentration. A better approach is to devise assays based on the contribution of all amino acids.

Historically the method of choice has been the biuret reaction (Figure 2.37) where a solution of copper(II) sulfate in alkaline tartrate solution reacts with peptide bonds to form a coloured (purple) complex absorbing around 540 nm. The structure of the complex formed is unclear but involves the coordination of copper ions by the amide hydrogen of peptide bonds. An important aspect of this assay is that it is based around the properties of the polypeptide backbone. Reduction of Cu(II) to Cu(I) is accompanied by a colour change and if the absorbance of the unknown protein is compared with a calibration curve then concentration is accurately determined. Ideally the standard protein used to construct a calibration curve will have similar properties, although frequently bovine serum albumin (BSA) is used.

Another commonly used method of estimating protein concentration is the Folin–Lowry method. The colour reaction is enhanced by the addition of Folin–Ciocalteau's reagent to a protein solution containing copper ions. The active ingredient of Folin–Ciocalteau's reagent is a complex mixture of phosphotungstic and phosphomolybdic acids. The reduced Cu(I) generated in the biuret reaction forms a number of reduced acid species in solution each absorbing at a wavelength maximum between 720 and 750 nm and leading to an obvious blue colour. Estimation of protein concentration arises by comparing the intensity of this blue colour to a 'standard' protein of known concentration at wavelengths between 650 and 750 nm.

Coomassie Brilliant Blue (Figure 2.38) does not undergo chemical reactions with proteins but forms

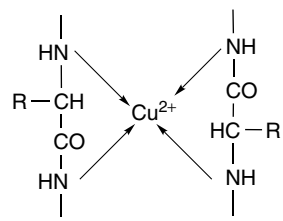


Figure 2.37 The complexation of cupric ions by peptide bonds in the biuret method

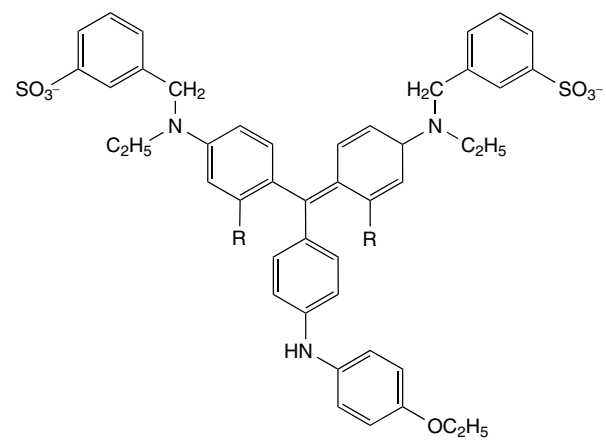


Figure 2.38 Coomassie blue dyes are commonly used to estimate protein concentration (R250 = H; G250 = CH_3)

stable complexes via non-covalent interactions. Coomassie blue (G250) is used to estimate the amount of protein in solution because complex formation is accompanied by a shift in absorbance maxima from 465 to 595 nm. The mechanism of complex formation between dye and protein is unclear but from the structure of the dye involves non-polar interactions.

Stereoisomerism

Up to this point the α carbon of amino acids has been described as an asymmetric centre and little attention has been paid to this characteristic property of tetrahedral carbon centres. One of the most important consequences of the asymmetric α carbon is that it gives rise to a chiral centre and the presence of two isomers. The two potential arrangements of atoms about the central carbon are shown for alanine (Figure 2.39) and whilst superficially appearing identical a closer inspection of the *arrangement* of atoms and the relative positions of the amino, carboxyl, hydrogen and methyl groups reveals that the two molecules can never be exactly superimposed.

The molecules are mirror images of each other or stereoisomers. Most textbooks state that these isomers are termed the L and D isomers without resorting to further explanation. All naturally occurring amino acids found in proteins belong to the L absolute configuration. However, at the same time the asymmetric or chiral centre is also described as an optically active centre. An optically active centre is one that rotates plane-polarized light. Amino acids derived from the hydrolysis of proteins under *mild* conditions will, with the exception of glycine, rotate plane-polarized light in a single direction. If this direction is to the right the amino acids are termed *dextrorotatory* (+) whilst rotation to the left is *laevorotatory* (−). Values for isolated amino acids are known, with L-amino acids rotating plane-polarized light in different directions and to differing extents (Table 2.4).

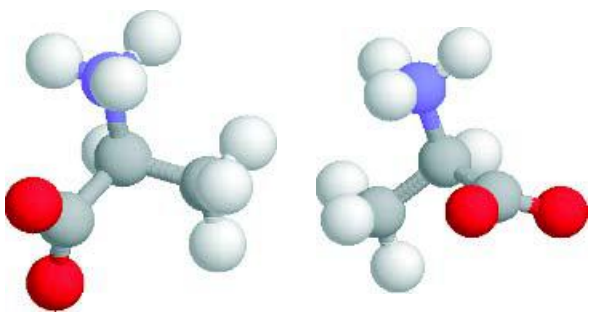


Figure 2.39 Two stereoisomers of alanine. As viewed the C–H bond of the α-carbon is pointing away from the viewer (down into the page) and superposition of the remaining groups attached this chiral centre is impossible. Each molecule is a mirror image of the other known variously as enantiomers, optical isomers or stereoisomers.

Table 2.4 The specific optical rotation of selected amino acids

L-Amino Acid	$[\alpha]_D(H_2O)$	L-Amino Acid	$[\alpha]_D(H_2O)$
Alanine	+1.8	Isoleucine	+12.4
Arginine	+12.5	Leucine	−11.0
Cysteine	−16.5	Phenylalanine	−34.5
Glutamic Acid	+12.0	Threonine	−28.5
Histidine	−38.5	Tryptophan	−33.7

The terms dextro- and levorotatary must not be confused with the L and D isomers, which for historical reasons are related to the arrangement of atoms in L- and D-glyceraldehyde. L-Trp is laevorotatory whilst L-Glu is dextrorotatory. The D and L stereoisomers of any amino acid have identical properties with two exceptions: they rotate plane-polarized light in opposite directions and they exhibit different reactivity with asymmetric reagents. This latter point is important in protein synthesis where D amino acids are effective inhibitors and in enzymes where asymmetric active sites discriminate effectively between stereoisomers.

The Cahn–Ingold–Prelog scheme has been devised to avoid confusion and provides an unambiguous method of assigning absolute configurations particularly for molecules containing more than one asymmetric centre. The amino acids isoleucine and threonine each contain two asymmetric centres, four stereoisomers, and potential confusion in naming isomers using the DL system. Using threonine as an example we can see

that the α carbon and the β carbon are each attached to four different substituents. For the β carbon these are the hydroxyl group, the methyl group, the hydrogen atom and the α carbon. The RS system is based on ranking the substituents attached to each chiral centre according to atomic number. The smallest or lowest ranked group is arranged to point away from the viewer. This is always the C–H bond for the C_α carbon and the remaining three groups are then viewed in the direction of decreasing priority. For biomolecules the functional groups commonly found in proteins are ranked in order of decreasing priority:

$$SH > P > OR > OH > NH_3 > COO > CHO > CH_2OH > C_6H_5 > CH_3 > H$$

The priority is established initially on the basis of atomic number so H (atomic no. = 1) has the lowest priority whilst groups containing sulfur (atomic no. = 16) are amongst the highest. To identify RS isomers one first identifies the asymmetric or stereogenic centre and in proteins this involves an sp^3 carbon centre with a tetrahedral arrangement containing four different groups. With the C–H bond (lowest priority) pointing away from the viewer the three remaining groups are ranked according to the scheme shown above. In the case where two carbon atoms are attached to an asymmetric centre then one moves out to the next atom and applies the same selection rule at this point. This leads to CH_3 being of lower priority than C_2H_5 which in turn is lower than COO^-. Where two isotopes are present and this is most commonly deuterium and hydrogen the atomic number rule is applied leading to D having a greater priority than H. Having ranked the groups we now assess if the priority (from high to low) is in a clockwise or anticlockwise direction. A clockwise direction leads to the R isomer (Latin: *rectus* = right) whilst an anticlockwise direction leads to the S isomer (Latin: *sinister* = left).

In L-threonine at the first asymmetric centre (C_α) when looking down the C_α–H bond (pointing away from the viewer in the arrangement shown in Figure 2.40) the direction of decreasing order of priority (NH_3^+, COO^- and CH–OH–CH_3) is counter clockwise. The C_α, the second carbon, is therefore configuration S. A similar line of reasoning for the asymmetric centre located at the third carbon (C_G) yields

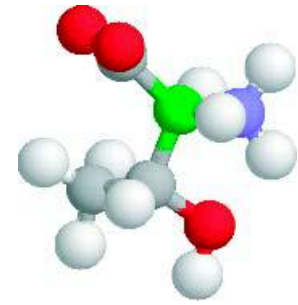


Figure 2.40 The asymmetric C_α centre of Thr is in the S configuration. The C_α is shown on green whilst the other atoms have their normal CPK colours

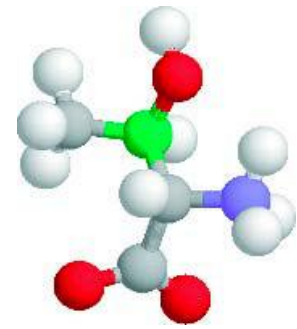


Figure 2.41 The asymmetric C_G centre of threonine is in the R configuration. The C_G atom is shown in green and the C–H bond is pointing away from the viewer

a decreasing order of priority in a clockwise direction and hence the R configuration. Combining these two schemes yields for L-threonine the nomenclature [2S-3R] threonine (Figure 2.41).

Non-standard amino acids

The 20 amino acids are the building blocks of proteins but are also precursors for further reactions that produce additional amino acids. Amino acids with unusual stereochemistry about the C_α carbon often called D amino acids are comparatively common in micro-organisms. Prominent examples are D-alanine and D-isoglutamate in the cell wall of the Gram-positive bacterium *Staphylococcus aureus*. In other bacteria

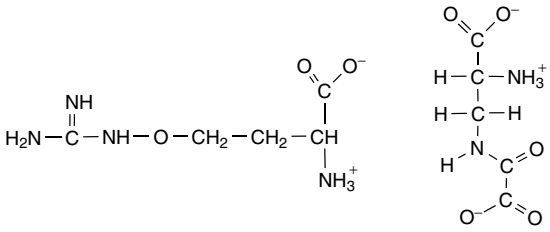


Figure 2.42 Canavanine and β-*N*-oxalyl L-α, β-diamino propionic acid (ODAP)

small peptide molecules known as ionophores form channels in membranes through the use of proteins containing D amino acid residues. A well-known example is the 15 residue peptide Gramicidin A containing a series of alternating D and L amino acids. In each case the use of these unusual or non-standard amino acids may be viewed as a defensive mechanism. Plants and microorganism use non-standard amino acids as protective weapons. Canavanine, a homologue of arginine containing an oxygen instead of a methylene group at the δ position, accumulates as a storage protein in alfafa seeds where it acts as a natural defence against insect predators (Figure 2.42).

A serious occurrence of an unusual amino acid is the presence of β-*N*-oxalyl L-α, β-diamino propionic acid (ODAP) in the legume *Lathyrus sativus*. Ingestion of seeds containing this amino acid is harmless to humans in small quantities but increased consumption leads to a neurological disorder (lathyrism) resulting in irreversible paralysis of leg muscles. ODAP is the culprit and acts by mimicking glutamate in biological systems.

Amino acids undergo a wide range of metabolic conversions as part of the cell's normal synthetic and degradative pathways. However, in a few cases amino acids are converted to specific products. Instead of undergoing further metabolism these modified amino acids are used to elicit biological response. The most important examples of the conversion of amino acids into other modified molecules include histamine, dopamine, thyroxine and γ-amino butyric acid (GABA) (Figure 2.43). GABA is formed from the amino acid glutamate as a result of decarboxylation and is an important neurotransmitter; thyroxine is an iodine-containing hormone stimulating basal metabolic rate via increases in carbohydrate metabolism in vertebrates, whilst in amphibians it plays a role in metamorphosis; histamine is a mediator of allergic response reactions and is produced by mast cells as part of the body's normal response to allergens.

Summary

Twenty different amino acids act as the building blocks of proteins. Amino acids are found as dipolar ions in solutions. The charged properties result from the presence of amino and carboxyl groups and lead to solubility in water, an ability to act as electrolytes, a crystalline appearance and high melting points.

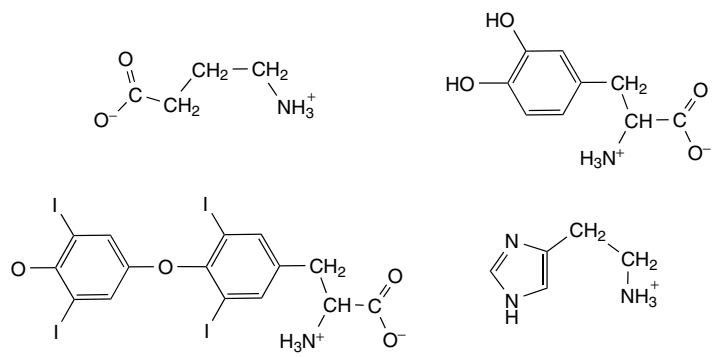


Figure 2.43 Biologically active amino acid derivatives. Clockwise from top left are GABA, dopamine, histamine and thyroxine

Of the 20 amino acids found in proteins 19 have a common structure based around a central carbon, the C_α carbon, in which the amino, carboxyl, hydrogen and R group are arranged tetrahedrally. An exception to this arrangement of atoms is proline. The C_α carbon is asymmetric with the exception of glycine and leads to at least two stereoisomeric forms.

Amino acids form peptide bonds via a condensation reaction and the elimination of water in a process that normally occurs on the ribosomes found in cells. The formation of one peptide bond covalently links two amino acids forming a dipeptide. Polypeptides or proteins are built up by the repetitive formation of peptide bonds and an average sized protein may contain 1000 peptide bonds.

The peptide bond possesses hybrid characteristics with properties between that of a C–N single bond and those of a C=N double bond. These properties result in decreased peptide bond lengths compared to a C–N single bond, a lack of rotation about the peptide bond and a preferred orientation of atoms in a *trans* configuration for most peptide bonds. One exception to this rule is the peptide bond preceding proline residues where the *cis* configuration is increased in stability relative to the *trans* configuration.

The side chains dictate the chemical and physical properties of proteins. Side chain properties include charge, hydrophobicity and polarity and underpin many aspects of the structure and function of proteins.

Problems

1. Why are amino acids white crystalline solids and how does this account for their physicochemical properties?
2. Using Figures 2.13 and 2.14 identify methine, methene and methyl groups in the side chains of leucine, isoleucine and valine. Label each carbon atom according to the usual nomenclature. Repeat the exercise of nomenclature for Figure 2.18.
3. Translate the following sequence into a sequence based on the single letter codes.
   Ala-Phe-Phe-Lys-Arg-Ser-Ser-Ser-Ala-Thr-Leu-Ile-Val-Thr-Lys-Lys-Gln-Gln-Phe-Asn-Gly-Gly-Pro-Asp-Glu-Val-Leu-Arg-Thr-Ala-Ser-Thr-Lys-Ala-Thr-Asp.
4. What is the average mass of an amino acid residue? Why is such information useful yet at the same time limited?
5. Which of the following peptides might be expected to be positively charged, which are negatively charged and which carry no net charge?
   Ala-Phe-Phe-Lys-Arg-Ser-Ser-Ser-Ala-Thr-Leu-Ile-Val-
   Ala-Phe-Phe-Lys-Arg-Ser-Glu-Asp-Ala-Thr-Leu-Ile-Val-
   Ala-Phe-Phe-Lys-Asp-Ser-Ser-Asp-Ala-Thr-Leu-Ile-Val-
   Ala-Phe-Phe-Lys-Asp-Ser-Ser-His-Ala-Thr-Leu-Ile-Val-
   Ala-Phe-Phe-Lys-Asp-Ser-Glu-His-Ala-Thr-Leu-Ile-Val-
   Are there any contentious issues?
6. The reagent 1-fluoro 2,4-dinitrobenzene has been used to identify amino acids. Describe the groups and residues you would expect this reagent to react with preferentially. Under what conditions would you perform the reactions.
7. Histidine has three ionizable functional groups. Draw the structures of the major ionized forms of histidine at pH 1.0, 5.0, 8.0 and 13.0. What are the respective charges at each pH?
8. The pK_1 and pK_2 values for alanine are 2.34 and 9.69. In the dipeptide Ala-Ala these values are 3.12 and 8.30 whilst the tripeptide Ala-Ala-Ala has values of 3.39 and 8.00. Explain the trends for the values of pK_1 and pK_2 in each peptide.
9. A peptide contains the following aromatic residues 3 Trp, 6 Tyr and 1 Phe and gives an absorbance at 280 nm of 0.8. Having added 1 ml of a solution of this peptide to a cuvette of path length 1 cm light calculate from the data provided in Table 2.4 the approximate concentration of the peptide (you can assume the differences in ε between 280 nm and those reported in Table 2.4 are negligible). The molecular mass of the peptide is 2670 what was the concentration in mg ml^{-1}.

10. Citrulline is an amino acid first isolated from the watermelon (*Citrullus vulgaris*). Determine the absolute (RS) configuration of citrulline for any asymmetric centre. Isoleucine has two asymmetric centres. Identify these centres and determine the configuration at each centre. How many potential optical isomers exist for isoleucine?