1

Basic Properties

1.1 INTRODUCTION

Peptides and proteins are essential biological molecules and assemblies, and many structures and functions of organisms are derived by using them. Peptides and proteins comprise chains of amino acids and are produced from DNA in the ribosome via messenger RNA through three-letter codons, i.e. each amino acid is represented by a triplet of nucleotides. Peptides are also known as 'polypeptides' to signify that they are polymers formed of amino acids. In fact, short peptides as considered in this book have the properties of oligomers rather than polymers.

Although there is no rigorous definition of the difference between peptides and proteins in this book, a typical peptide would have up to 100 residues, and longer chains would be considered proteins. In fact, standard synthesis methods lead to peptides up to 50–60 residues in length, which may also be considered a cut-off, with somewhat larger chains being described as 'mini-proteins'.

This chapter is organized as follows. In Section 1.2, the essential properties of the 20 standard natural amino acids that peptides and proteins are built from are first considered, including an analysis of reactivity, charge, and hydrophobicity. When incorporated in a peptide or protein, the term 'residue' is used for the group formed from the corresponding amino acid. Non-natural residues that are found in a few natural and synthetic peptides are also introduced in Section 1.2, especially those which are mentioned in peptide sequences elsewhere in this book. Then, in Section 1.3, the nature and geometry of the peptide bond are considered. The main secondary structures adopted by peptides are discussed in Section 1.4. There follows

Introduction to Peptide Science, First Edition. Ian W. Hamley.

(Section 1.5) a description of characterization methods used to determine peptide structure and conformation. The chapter concludes with a list of useful peptide websites and selected software, including databases and property calculators.

1.2 PROPERTIES OF AMINO ACIDS

Amino acids are chiral molecules, meaning that there is a distinct spatial arrangement of substituents around the central backbone carbon atom, which is termed a C_α atom. This is a stereogenic centre. There are two possible arrangements with different 'handedness', which are mirror images. These forms are termed enantiomers. There are two forms, termed L- and D-amino acids (note the use of a small capital letter). In the L form of a residue in a peptide, the substituents are arranged as shown in Figure 1.1, which can be remembered via the CORN mnemonic, with the CO, R, and N groups in a clockwise configuration from left, when viewed along the bond to the H atom. In nature, peptide and protein structures are almost entirely built from L-amino acids. Molecules that contain amino acids may have more than one stereocentre; for example, Figure 1.2 shows the diastereomers of threonine, L- and D-threonine, in the form of Fischer projections. Figure 1.2 also includes the two *allo*- forms, which are rarely found in natural peptides (isoleucine also has four diastereomers). The *R*,*S* notation for the stereocentres according to the Cahn–Ingold–Prelog rules are also shown. Fischer projections are planar representations of an enantiomer (named after Emil Fischer) in which bonds denoted as horizontal extend above the plane of the paper and vertical bonds denote those which extend below the plane. In contrast to (mono)stereoisomers, diastereoisomers can have distinct chemical and physical properties (for example melting points).

The peptide bond has partial double bond character, so its length 1.33 Å (0.133 nm) is significantly shorter than a usual C–N bond length of 1.45 Å

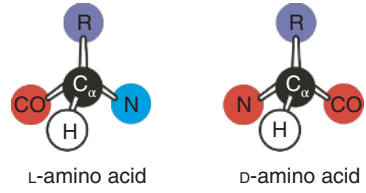


Figure 1.1 Enantiomers of amino acid residues in peptides. Left: L-amino acid, right: D-amino acid.

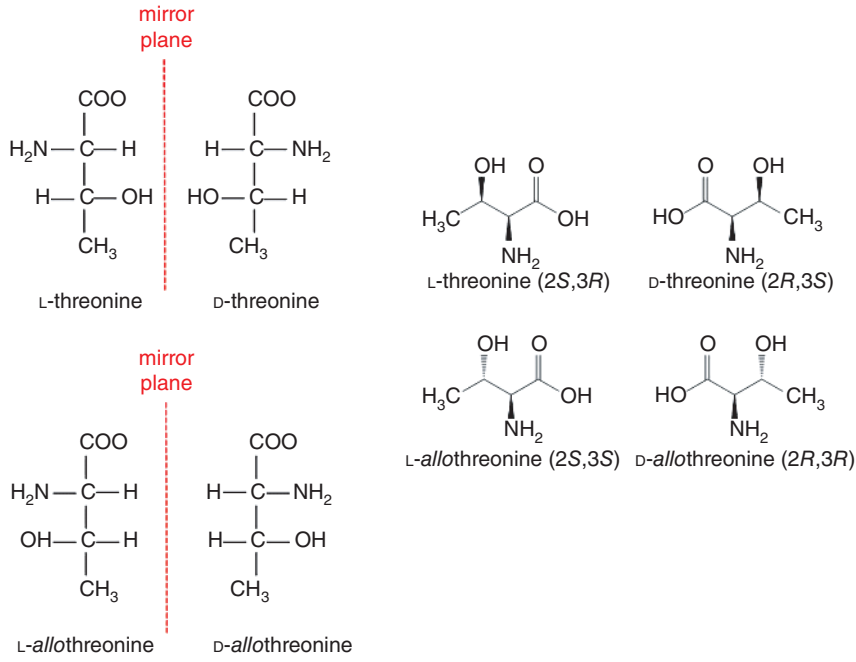


Figure 1.2 Fischer projections and structures of L-threonine, D-threonine, L-*allo*threonine, and D-*allo*threonine.

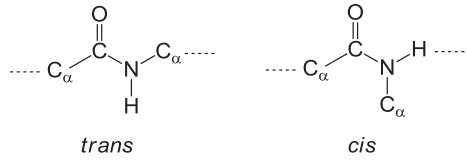
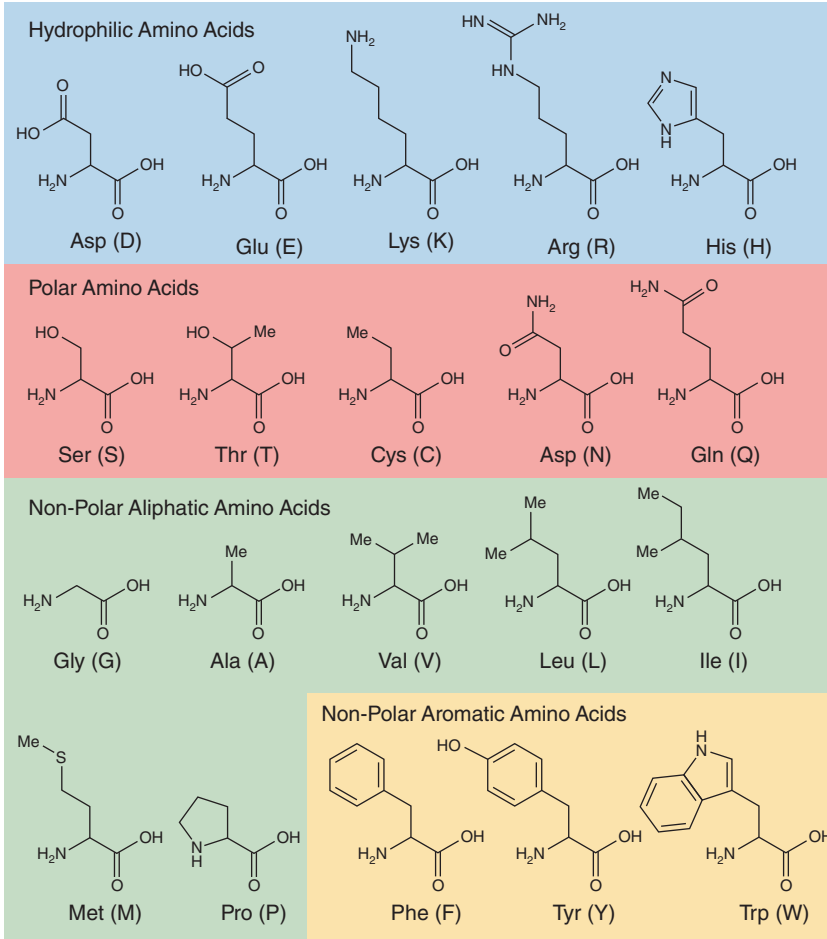


Figure 1.3 Configurations of a peptide bond.

(0.145 nm). The partial double bond character leads to restricted rotation and a preferred co-planar arrangement. Two configurations of the planar peptide bond are possible, called *trans* and *cis*, shown in Figure 1.3. The *trans* form is favoured energetically.

The 20 standard (so-called canonical) amino acids found in nature with different side chains are shown in Figure 1.4. They are grouped according to amino acid polarity or charge and, for the non-polar amino acids, according to whether they have an aliphatic or aromatic substituent. All the amino acids except glycine are chiral, being present as either L- or D-enantiomers. In nature, L-enantiomers are the standard form, although D-amino acids are



Asp	Aspartic acid	Gly	Glycine
Glu	Glutamic acid	Ala	Alanine
Lys	Lysine	Val	Valine
Arg	Arginine	Leu	Leucine
His	Histidine	Ile	Isoleucine
Ser	Serine	Met	Methionine
Thr	Threonine	Pro	Proline
Cys	Cysteine	Phe	Phenylalanine
Asp	Asparagine	Tyr	Tyrosine
Gln	Glutamine	Trp	Tryptophan

Figure 1.4 The standard amino acids.

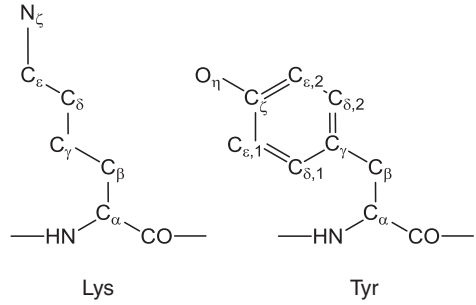


Figure 1.5 Amino acid side chain labelling nomenclature.

present in some natural peptides. Figure 1.4 does not show the enantiomeric forms. Here, and elsewhere, the symbol Me in a chemical formula indicates methyl (CH_3).

The labelling scheme for (non-hydrogen) atoms in amino acid side chains is shown in Figure 1.5, as exemplified for lysine and tyrosine.

The hydrophilic amino acids comprise two that are normally positively charged (except at low pH, see Table 1.2 and associated discussion) – these are the anionic residues Asp (D) and Glu (E). These differ only by the addition of an extra methylene group in the side chain of E. Although the pK_a values of the side chain residues are similar, the length of the side chain (and so the location of the charge) does significantly influence the conformation of the backbone and the reactivity of these groups. Section 2.11.5 discusses some reactions using these residues.

The two basic residues, Lys (K) and Arg (R), are cationic under most conditions, except high pH. Lysine is used for many conjugation reactions, as discussed in Sections 2.11.2 and 2.11.5. Arginine contains a strongly basic guanidino group, which has a resonance structure shown for arginine in the charged state in Figure 1.6, and among its properties it is able to form bidentate hydrogen bonds (see Figure 4.10).

Histidine can also exhibit basic character, although its pK_a is within the range of physiological pH values. It is unique among amino acids in having an imidazole group, and in the charged state the positive charge is shared between the nitrogen atoms by resonance. There are also two tautomeric forms in the uncharged state, depending on which nitrogen has an attached hydrogen atom.

Serine and threonine are related polar amino acids containing hydroxyl groups. They differ only by the addition of a methyl group in serine. Both these residues improve the solubility of peptides. Serine is important in the activity of many enzymes, such as serine proteases. There are two diastereomers of threonine, as discussed above (Figure 1.2).

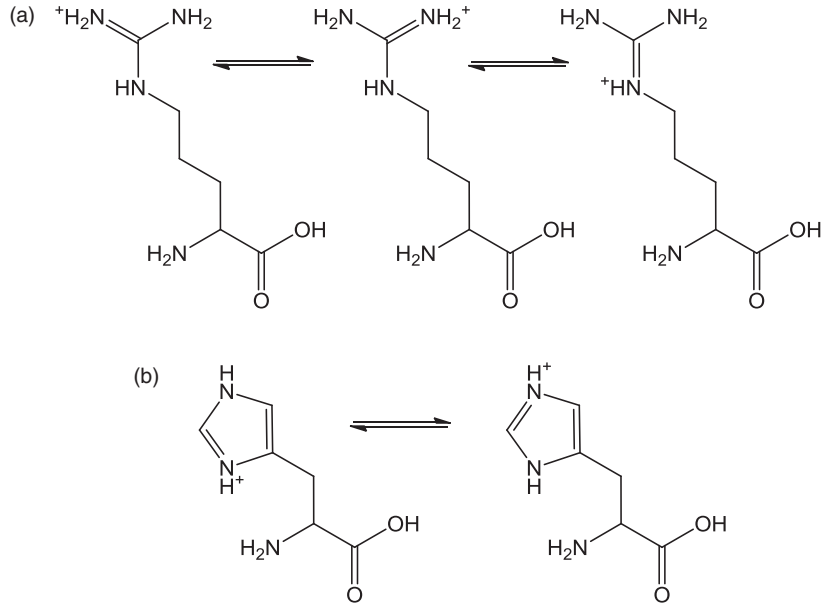


Figure 1.6 Resonance structures of (a) arginine and (b) histidine.

Cysteine is one of the two residues that contain sulfur in the thiol (sulfhydryl) group. This is very commonly used as a tag in bioconjugation reactions, as described further in Sections 2.11.2 and 2.11.5. The thiol group in cysteine can also form disulfide links, which are common stabilizing elements in proteins, and are also used in natural and synthetic cyclic peptides. Disulfide bridges are discussed further in Section 1.4 and the synthesis of cyclic peptides including disulfide bridging is discussed in Section 2.6. The sulfhydryl group ionizes at high pH to give a $–CH_2–S^–$ side chain.

The two residues Asn (N) and Gln (Q) with side chain amide groups differ only in side chain length. They are the amide forms of the acidic residues Asp (D) and Glu (E) respectively. These residues are polar, but do not ionize. The amide groups act as both hydrogen donor and hydrogen acceptor. The amide groups are labile at very high or low pH values and high temperatures and these residues can undergo deamidation to form Asp and Glu. N-Terminal glutamine residues can spontaneously cyclize to form pyroglutamic acid, which is shown in Table 1.3.

Glycine is the simplest amino acid, and the only one that is non-chiral because it only has a second hydrogen atom as its ‘side group’. The amino acids Ala (A), Val (V), Leu (L), and Ile (I) constitute hydrophobic aliphatic

residues. The hydrophobicity of these residues increases with the side-chain size. These residues are inert although they tend to associate through hydrophobic interactions, this being important in the stabilization of β-sheet structures for example. Ile contains an additional asymmetric centre (C_β) but only one diastereomer (L-isoleucine) occurs naturally. Methionine is grouped with the non-polar aliphatic residues in Figure 1.4, although because it contains a sulfur atom it is related to cysteine. However, the thiol group in cysteine is methylated, leading to a non-polar character and a lack of ionizability. The sulfur atom in methionine is susceptible to oxidation, forming first a sulfoxide, then a sulfone. Proline is a unique residue in that it does not contain an amide hydrogen able to participate in hydrogen bonds. In addition, the ring in proline leads to a conformational constraint that increases the fraction of *cis*- configured peptide chain before the Pro residue.

The three aromatic amino acids, Phe (F), Tyr (Y), and Trp (W), are hydrophobic. Phe and Trp are not very reactive under most conditions. The –OH group in Tyr is reactive and ionizes at high pH, under which conditions dityrosine formation is possible. The indole group in Trp leads to the intrinsic fluorescence of this residue, which is a useful property in aggregation assays since the fluorescence is sensitive to the local microenvironment. The aromatic groups in Phe, Tyr, and Trp can lead to π-stacking interactions which can stabilize some secondary and aggregate structures (Section 1.4). In addition, the aromatic groups have characteristic peaks in their UV spectra, which can be used in peptide/protein concentration measurements (Section 1.5.1).

Important properties of the amino acids, including the pK_a values of the side chains as well as hydrophobicity, are listed in Table 1.1. Other properties of amino acids are included, such as their typical abundance in proteins (from database analysis), van der Waals volumes, and hydrophobicity. There are a considerable number of published hydrophobicity scales, examples of the most well-known ones are included in Table 1.1. These are obtained from measurements of free energies of partition. Kyte and Doolittle presented a hydropathy scale based on water-vapour transfer free energy and the tendency of amino acids to be on the exterior or interior of proteins. This scale correlates to a good degree to the water-vapour ΔG(transfer). White and Wimley present tables of similarly determined ΔG(partition) values, ΔG(water/lipid interface), and ΔG(water/octanol). The trend in the Wimley–White ΔG values is clear when plotted (Figure 1.7). Other hydrophobicity scales have been proposed.

Since many peptides are charged due to the presence of residues with charged side chains and/or charged termini, analysis of pK_a values enables

Table 1.1 Properties of amino acids.

Symbol (one letter code)	Abundance in proteins (%)	Side chain pK_a	van der Waals volume (Å^3)	Hydrophobicity: Kyte–Doolittle hydropathy index	Hydrophobicity: Wimley–White interface scale, ΔG (kcal mol^{-1})	Hydrophobicity: Wimley–White octanol scale, ΔG (kcal mol^{-1})
Gly (G)	7.2		48	−0.4	0.01	1.15
Ala (A)	7.8		67	1.8	0.17	0.50
Val (V)	6.6		105	4.2	0.07	−0.46
Leu (L)	9.1		124	3.8	−0.56	−1.25
Ile (I)	5.3		124	4.5	−0.31	−1.12
Met (M)	2.2		124	1.9	−0.23	−0.67
Pro (P)	5.2		90	−1.6	0.45	0.14
Phe (F)	3.9		135	2.8	−1.13	−1.71
Tyr (Y)	3.2	10.5	141	−1.3	−0.94	−0.71
Trp (W)	1.4		163	−0.9	−1.85	−2.09
Ser (S)	6.8		73	−0.8	0.13	0.46
Thr (T)	5.9		93	−0.7	0.14	0.25
Cys (C)	1.9	8.4	86	2.5	−0.24	−0.02
Asn (N)	4.3		96	−3.5	0.42	0.85
Gln (Q)	4.3		114	−3.5	0.58	0.77
Asp (D)	5.3	3.9	91	−3.5	1.23	3.64
Glu (E)	6.3	4.0	109	−3.5	2.02	3.63
Lys (K)	5.9	10.5	135	−3.9	0.99	2.80
Arg (R)	5.1	12.5	148	−4.5	0.81	1.81
His (H)	2.3	6.0	118	−3.2	0.96	2.33

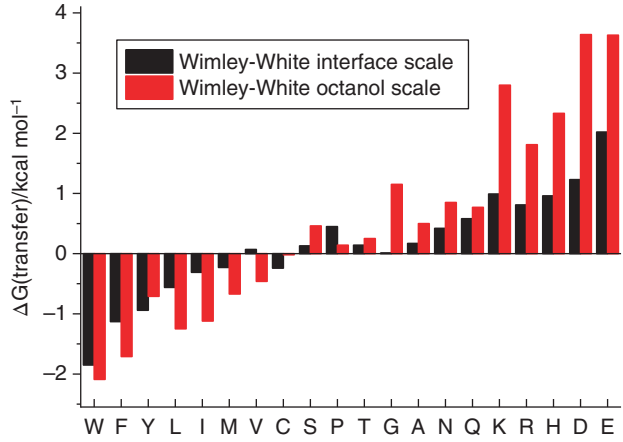


Figure 1.7 Hydrophobicity of amino acids, according to the Wimley–White scale. The hydrophobicity decreases left-to-right.

estimation of the overall net charge on the peptides (see Eq. (1.4) below). The isoelectric point (pI) is another useful characteristic of peptides; this corresponds to the pH at which the net charge is zero. Table 1.2 summarizes the state of charged amino acids below and above the respective side chain pK_a values. For the termini, typical pK_a values are pK_a = 9–10 for the N-terminal NH_3 and pK_a = 2–3 for the C-terminal COOH. It can be assumed that when the solution pH = pK_a, equal numbers of the charged and uncharged species will be present. It should be noted that pK_a values of residues in peptides (such as those quoted in Tables 1.1 and 1.2) are typical values; the pK_a of a particular residue will depend on its local environment (for example it will be modified by the presence of other nearby charged residues). Acid or base titrations may be used to experimentally determine pK_a values for residues in short charged peptides. Figure 1.8 shows a schematic titration curve.

Close to the pK_a, the two forms of the peptide can be represented as an acid–base equilibrium. Using the Henderson–Hasselbalch equation, the charge can be obtained.

$$HA \rightleftharpoons H^+ + A^-$$

The acid dissociation constant is given by

$$K_a = \frac{[A^-][H^+]}{[HA]} \tag{1.1}$$

The Henderson–Hasselbalch equation is

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right) \tag{1.2}$$

Table 1.2 Side chain protonation depending on pH.

Amino acid	$pH < pK_a$	pK_a	$pH > pK_a$
Asp (D)	–COOH	3.9	$–COO^-$
Glu (E)	–COOH	4.0	$–COO^-$
Lys	$–NH_3^+$	10.5	$–NH_2$
Arg (R)	$–C(NH_2)_2^+$	12.5	$–C(NH_2)(NH)$
His (H)	imH^+	6.0	imH
Tyr (Y)	–PheOH	10.5	$–PheO^-$
Cys (C)	–SH	8.4	$–S^-$

im = imidazole.

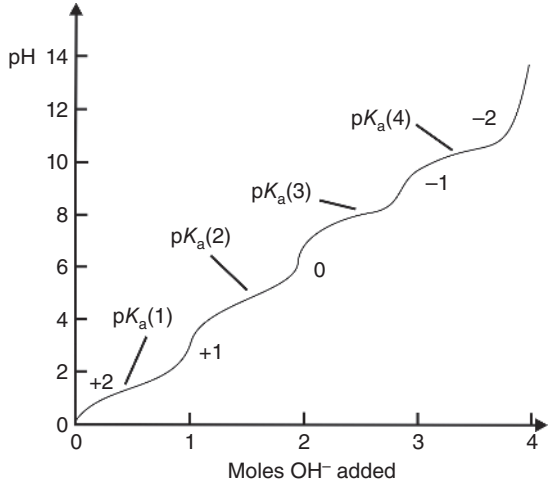


Figure 1.8 Schematic titration curve for a tetrapeptide NH_2–EGAK–COOH, plotting pH versus molar concentration of OH^- per mol of tetrapeptide. The peptide net charge is indicated between pK_a values (defined as the mid-point of the pseudo-plateaus) as shown. Here, $pK_a(1)$ is that for the C terminus, $pK_a(2)$ is for the E residue, $pK_a(3)$ is for the N terminus and $pK_a(4)$ is for the K residue. Net charge +2 corresponds to NH_3^+–(E)GA(K^+)–COOH, +1 to NH_3^+–(E)GA(K^+)–COO^-, 0 to NH_3^+–(E^-)GA(K^+)–COO^-, −1 to NH_2–(E^-)GA(K^+)–COO^-, and −2 to NH_2–(E^-)GA(K)–COO^-, where E^- denotes charged glutamic acid (with COO^- side chain terminus) and K^+ denotes charged lysine with NH_3^+ side chain terminus. The pI of this peptide is at pH = 7.

So the fractional charge

$$\frac{[A^-]}{[A^-] + [HA]} = \frac{10^{(pH - pK_a)}}{1 + 10^{(pH - pK_a)}} \tag{1.3}$$

For example, for lysine at pH 10, using the side chain $pK_a = 10.5$ (Table 1.2),

$$\frac{[A^-]}{[A^-] + [HA]} = \frac{10^{(pH - pK_a)}}{1 + 10^{(pH - pK_a)}} = \frac{10^{-0.5}}{1 + 10^{-0.5}} = 0.24$$

So approximately one quarter of the lysine residues are uncharged (in the form $[A^-]$, i.e. NH_2) under these conditions, which means for example for a peptide containing four charged lysine residues, three of them will be charged. The same result can be obtained from the formula for the net charge, q:

$$q = \sum_i N_i \frac{+1}{1 + 10^{+(pH - pK_{a,i})}} + \sum_j N_j \frac{-1}{1 + 10^{-(pH - pK_{a,j})}} \tag{1.4}$$

where the index 'i' labels basic residues (total N_i of each type with $pK_{a,i}$) and 'j' labels acidic residues (N_j of each type with $pK_{a,j}$).

Several websites (listed in Table 1.6) provide convenient calculators for peptide net charge, pI, and other properties (molar mass) based on input sequences.

The amphiphilicity of α-helices can be quantified via the hydrophobic moment, which provides a useful measure of the propensity of a α-helical peptide to interact with membranes; this is relevant, for example, to antimicrobial activity. The hydrophobic moment for a peptide with N residues is defined as

$$\mu_H = \left\{ \left[\sum_{n=1}^{N} H_n \sin(\delta n) \right]^2 + \left[\sum_{n=1}^{N} H_n \cos(\delta n) \right]^2 \right\}^{1/2} \tag{1.5}$$

Here H_n is the hydrophobicity index (chosen from a suitable scale such as the Kyte–Doolittle or Wimley–White scales, see Table 1.1) for residue n, and δ is the angle of rotation of side chains around the backbone ($\delta = 100°$ for an α-helix).

Table 1.1 shows that some amino acids have similar properties. This can be further quantified in terms of physicochemical parameters used to assess amino acid replacement similarity. These are typically used to assess protein substitutions. Scales include the Grantham distance, Sneath's index, and others. These scales are based on parameterization of amino acid characteristics such as molar volume and polarities.

The tendency for residues to form α-helices or β-sheets has also been quantified, based on analysis of protein databases. A variety of scales have been proposed, which can be used in the design of peptides targeting these secondary structures. Figure 1.9 shows an example of one scale. Other scales

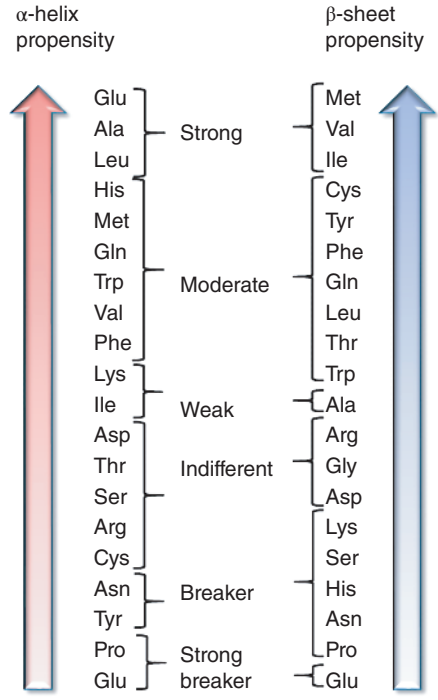


Figure 1.9 Chou–Fasman scale of α-helix and β-sheet propensity for protein and peptide residues.

to assess aggregation propensity for peptide sequences are available on the web and are listed in Table 1.6.

Amino acids that bear functional groups which can be removed or functionalized are substrates for enzymes. Residue-specific enzymes have names based on the amino acid appended with '-ase', for example arginase, asparaginase, and tyrosinase. Another class of residue-specific enzymes are kinases and phosphatases that respectively add or remove phosphoryl groups to tyrosine, serine, or threonine. Transferases have activity, as the name suggests, in transferring functional groups such as methyl groups, on to specific residues in proteins and peptides. Other examples of residue-specific enzymes include amino acid hydrolases and dehydroxylases. Enzymes such as proteases recognize specific configurations of amino acids in peptide/protein substrates. An important class, for example, are the serine proteases, which contain catalytic triads of residues for example Ser-His-Asp (these are not sequential but the notation means that these residues are present in specific configurations at the catalytic site)

in chymotrypsin. Other reactions that exploit the functional groups on charged, hydrophilic, and tyrosine residues are discussed in Section 2.11.

Table 1.3 lists some non-standard (mostly non-proteinogenic) amino acids found in natural products and mentioned elsewhere in the book, among others. An extra $–CH_2–$ group between the carbonyl and amine groups in the backbone is characteristic of β-amino acids such as β-alanine. Two $–CH_2–$ groups separate the carbonyl and amine groups in γ-amino acids such as γ-aminobutyric acid, GABA (which is an important neurotransmitter) or statine (present in pepstatin, see Section 5.2). δ-Amino acids have three added $–CH_2–$ groups. Aminohexanoic acid (also known as aminocaproic acid) is an analogue of lysine with five $–CH_2–$ groups in the backbone instead of the side chain. Aminohexanoic acid is an important intermediate in the synthesis of nylon-6. The homo-amino acids are analogues of amino acids with additional $–CH_2–$ groups in the side chain.

Other residues in Table 1.3 are designed specifically to reduce the susceptibility of the peptide in which they are incorporated to proteolysis. Examples, other than the β-, γ-, and δ-amino acids, include the dehydro amino acids and amino acids containing unsaturated rings, as well as amino acids with blocked functional groups such as *N*-ε-methyllysine, *N*-methylglycine (sarcosine), which is strictly a peptoid residue (see Section 2.7).

Phosphorylation and dephosphorylation of tyrosine and serine are important in many biochemical reactions and these processes are achieved via enzymes. Kinases perform phosphorylation and are important targets for cancer therapies since kinases are involved in cell signalling, metabolism, and protein regulation. Activity of the enzyme tyrosine hydroxylase on tyrosine produces L-DOPA (L-3,4-dihydroxyphenylalanine) (see Table 1.3), which is used as a treatment for Parkinson's disease. This amino acid is an important precursor of the catecholamine neurotransmitters including dopamine (which results from removal of the carboxyl group from L-DOPA), norepinephrine (noradrenaline), and epinephrine (adrenaline). The enzyme tyrosinase acting on tyrosine also produces L-DOPA as an intermediate oxidant on the pathway to dopaquinone and its polymerization to produce melanin (Section 3.7). L-Ornithine (Table 1.3) is an analogue of lysine, having one extra methyl group in the side chain. It is produced by the activity of arginase on L-arginine.

There are only a few amino acids that contain elements other than the standard C, H, N, O, and S (or P in the case of the phosphorylated amino acids). Selenocysteine is the analogue of cysteine in which selenium replaces sulfur. It incorporates selenium into proteins *in vivo* and is considered to be the 21st proteinogenic amino acid.

Table 1.3 Some non-standard amino acids.

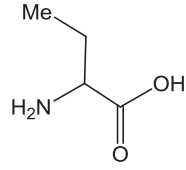
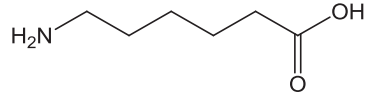
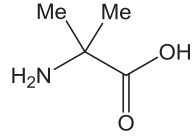
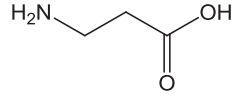
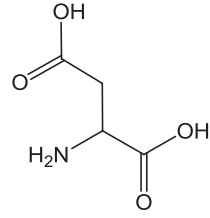
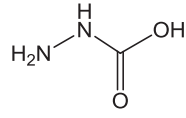
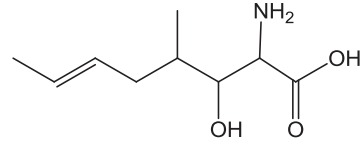
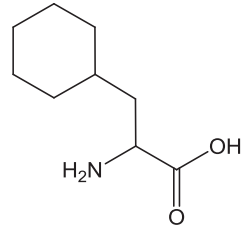
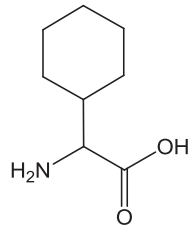
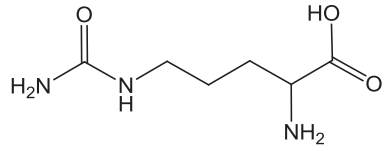
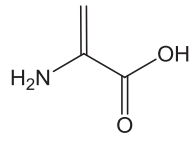
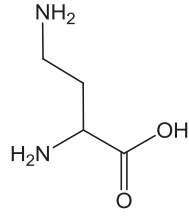
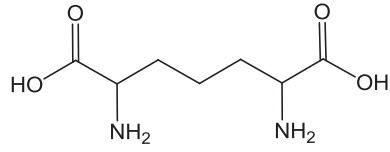
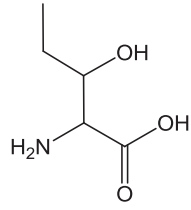
Symbol	Name	Structure
Abu	1-aminobutyric acid	
Ahx	6-aminohexanoic acid (6-aminocaproic acid)	
Aib	α-aminoisobutyric acid	
β-Ala	β-alanine	
β-Asp	β-aspartic acid	
AzGly	aza-glycine	
2-Bmt	(*E*)-2-butenyl-4-methylthreonine	
Cha	β-cyclohexylalanine	

Table 1.3 (*continued*)

Symbol	Name	Structure
Chg	α-cyclohexylglycine	
Cit	citrulline	
Dha	dehydroalanine	
Dab	diaminobutyric acid	
Dap	diaminopimelic acid	
Dhb	2,3-dehydroaminobutyric acid	

(*continued*)

Table 1.3 (*continued*)

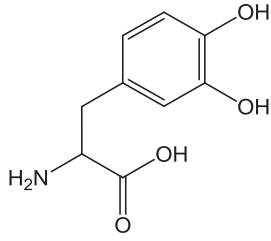
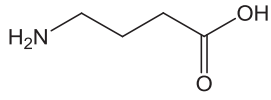
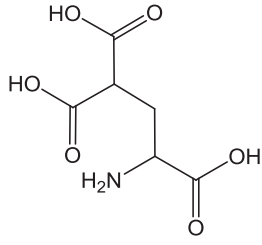
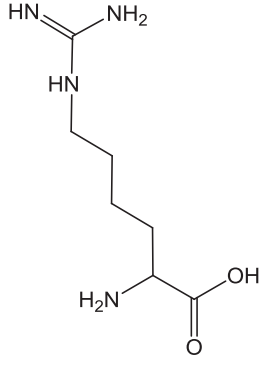
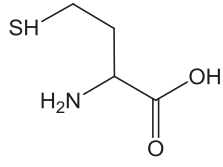
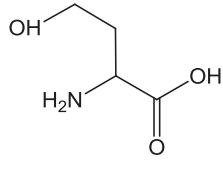
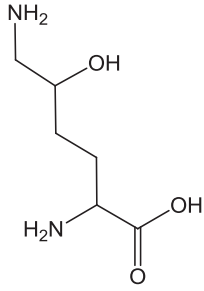
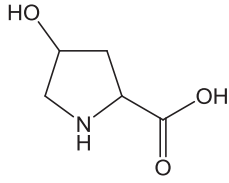
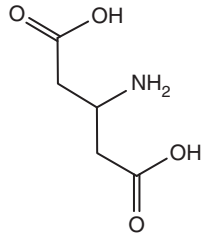
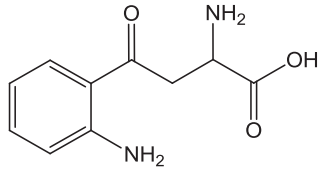
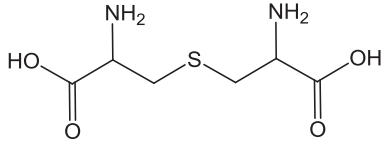
Symbol	Name	Structure
DOPA	3,4-dihydroxyphenylalanine	
GABA	γ-aminobutyric acid	
Gla	γ-carboxyglutamic acid	
HArg	homoarginine	
HCy	homocysteine	
HSe	homoserine	

Table 1.3 (*continued*)

Symbol	Name	Structure
Hylys	hydroxylysine	
Hyp	hydroxyproline	
IsoGlu	isoglutamic acid (amino acid isomer)	
Kyn	kynurenine	
Lan	lanthionine	

(*continued*)

Table 1.3 (*continued*)

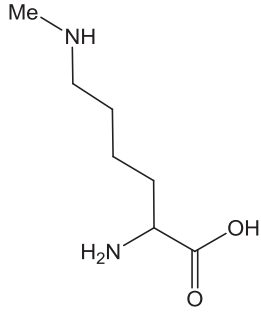
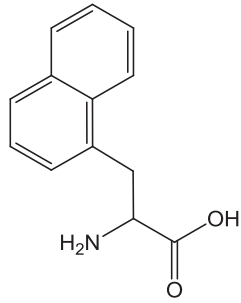
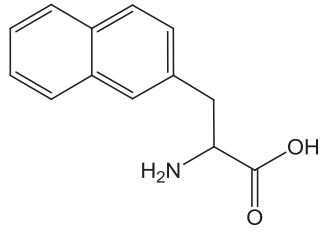
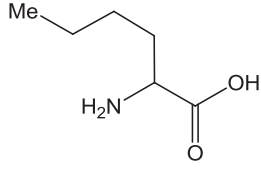
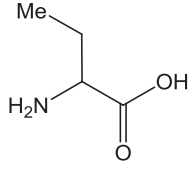
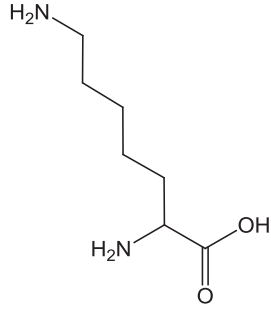
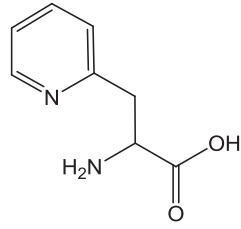
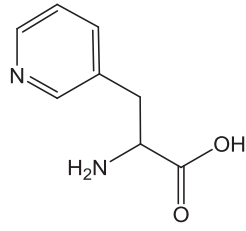
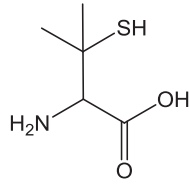
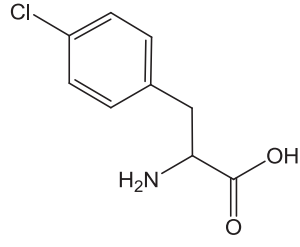
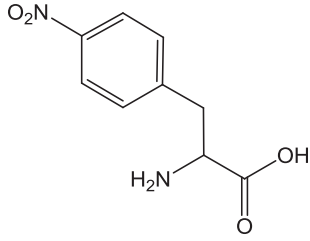
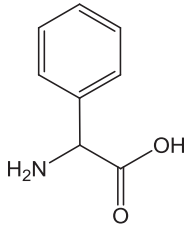
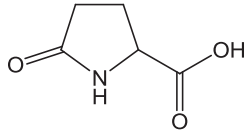
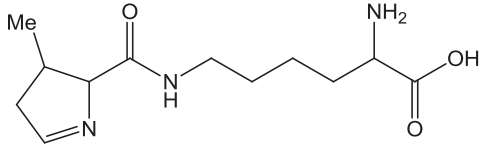
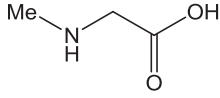
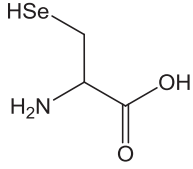
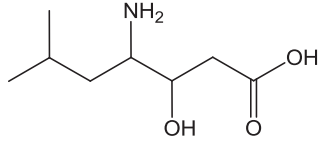
Symbol	Name	Structure
Lys(Me)	*N*-ε-methyllysine	
1-Nal	1-naphthylalanine	
2-Nal	2-naphthylalanine	
Nle	norleucine	
Nva	norvaline	

Table 1.3 (*continued*)

Symbol	Name	Structure
Orn	ornithine	
2-Pal	3-(2-pyridyl)alanine	
3-Pal	3-(3-pyridyl)alanine	
Pen	penicillamine	
4-Cl-Phe	4-chlorophenylalanine	

(*continued*)

Table 1.3 (*continued*)

Symbol	Name	Structure
4-NO_2-Phe	4-nitrophenylalanine	
Phg	phenylglycine	
pGlu or Pyr	pyroglutamic acid	
Pyl	pyrrolysine	
Sar	sarcosine	
Sec	selenocysteine	
Sta	statine	

(*continued*)

Table 1.3 (*continued*)

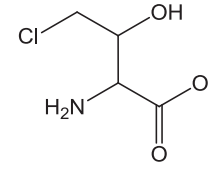
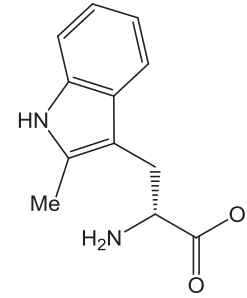
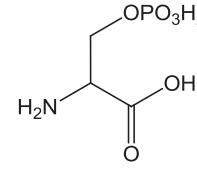
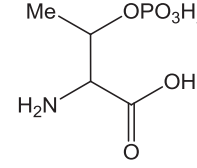
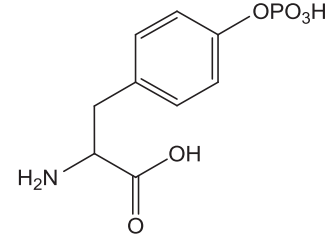
Symbol	Name	Structure
4-ClThr	4-chlorothreonine	
D-2-Me-Trp	D-2-methyltryptophan	
pSer	phosphoserine	
pThr	phosphothreonine	
pTyr	phosphotyrosine	

Table 1.3 includes one example (dehydroalanine) of a dehydroamino acid (one which contains a double bond in the side chain) although others have been used as non-natural residues resistant to proteolysis. Dehydroamino acids also occur naturally, being (rare) post-translational modifications. Analogues of aromatic amino acids with bulkier side chains, such as naphthylalanine, are included in Table 1.3, as are saturated variants such as α-cyclohexylalanine, an analogue of phenylalanine. Phenylglycine is a variant of Phe which lacks a spacer in the side chain. (There are no natural aromatic residues without a methyl spacer before the aromatic group.) Other entries in Table 1.3 include analogues of aromatic residues with substituents on the aromatic group, e.g. chloro- or nitro-substituted phenylalanine, or variants with other aromatic moieties, such as pyridyl groups or methyl-substituted tryptophan.

Norleucine and norvaline, shown in Table 1.3, are isomers of the natural amino acids leucine and valine respectively, lacking the side group branches present in the natural residues.

Among the other non-natural residues in Table 1.3, lanthionine is the product of the addition of cysteine to dehydroalanine. It is a key residue in the lantibiotics discussed in Section 4.6.1. Other amino acids in Table 1.3 are encountered rarely, and are included for completeness; in particular they are present in specific peptide therapeutic molecules mentioned in Chapter 5, where their contextual usage is explained.

1.3 THE PEPTIDE BOND

Nature produces peptides by linking amino acids; this takes place at the ribosome. Amino acids are coded by triplets of nucleic acids (codons), and considering that mRNA (messenger RNA) is translated, triplets of RNA bases specify particular amino acids. This is the standard genetic code. Of course, peptides can now easily be synthesized in the laboratory using methods discussed in Chapter 2.

The peptide bond is formed in a condensation (dehydration) reaction between the carboxyl group of one amino acid and the amine group of another (Figure 1.10). The residues in a peptide are linked through amide units, which are planar.

The geometry around a peptide bond is shown in Figure 1.11. An important fact about peptide bonds is the planarity of the amide C=O–NH unit, which constrains the possible conformations that the chains can adopt. On average, The C–N bond length is 1.33 Å, the N–C_α bond length is 1.45 Å,

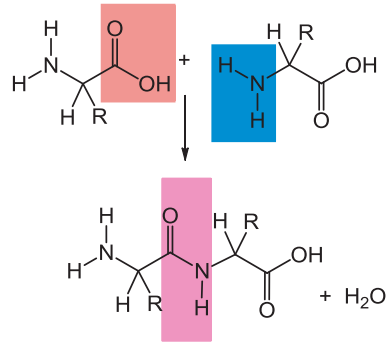


Figure 1.10 Formation of a peptide bond (amide group highlighted in purple) from condensation reaction between the carboxyl group (red) of one amino acid and the amino group (blue) of another.

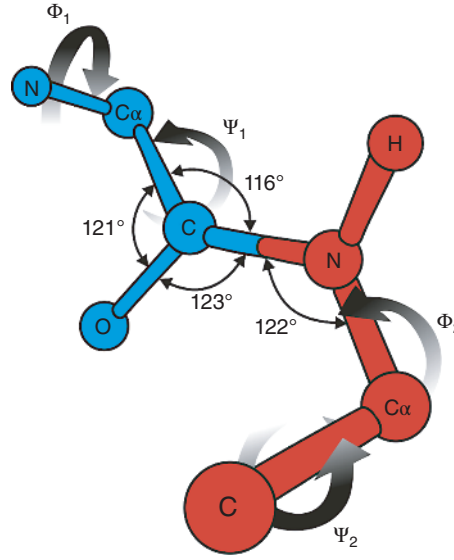


Figure 1.11 Peptide bond geometry with typical angles shown.

and C–C_α is 1.52 Å. The angles Φ are defined by rotation about the N–C_α bonds while Ψ is defined by rotation around C_α–C. The case $\Phi = 0$ defines the case where the C_α–C bond is *trans* to the N–H bond, whereas $\Psi = 0$ corresponds to C_α–N *trans* to C=O. Not shown in Figure 1.11 for simplicity is the angle **Ω**, the rotation angle about the N–CO bond; this angle is usually 180°.

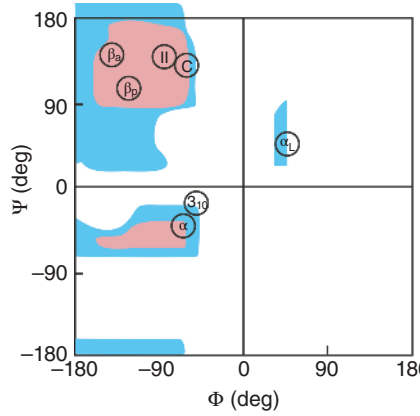


Figure 1.12 A Ramachandran plot. The shaded areas show the 'normally allowed conformations' in pink and the 'outer limit' structures in blue. Symbols as follows: β_a is antiparallel β-sheet, β_p is parallel β-sheet, II is polyproline II/polyglycine II, C is collagen (a triple helical), 3_{10} is a 3_{10}-helix, α is a (normal right-handed) α-helix, and α_L is a left-handed α-helix.

The peptide bond links amino acids in polypeptide chains, which nature has evolved to exploit as scaffolds that present sequences of residues with the right degree of flexibility to adopt configurations corresponding to different secondary structures that provide the structure and function of proteins and peptides.

Peptide secondary structures (discussed further in Section 1.4) are characterized by particular favoured torsional angles, as summarized in a Ramachandran plot of conformational space (Figure 1.12). Such a plot is also widely used for protein conformational analysis.

Conventionally (as in this book), peptide sequences are written or drawn from the N-terminus on the left to the C-terminus on the right.

1.4 SECONDARY STRUCTURES

The most common ordered peptide secondary structures are the α-helix and β-sheet. These are also common elements of protein tertiary structure. Figure 1.13 shows an α-helix and the two types of β-sheet structure, with β-strands (peptide chains within β-sheet structures) aligned parallel or antiparallel to one another. The α-helix is stabilized by intramolecular hydrogen bonding, in contrast to the intermolecular hydrogen bonding of β-sheets. The standard α-helix is a right-handed structure, although the

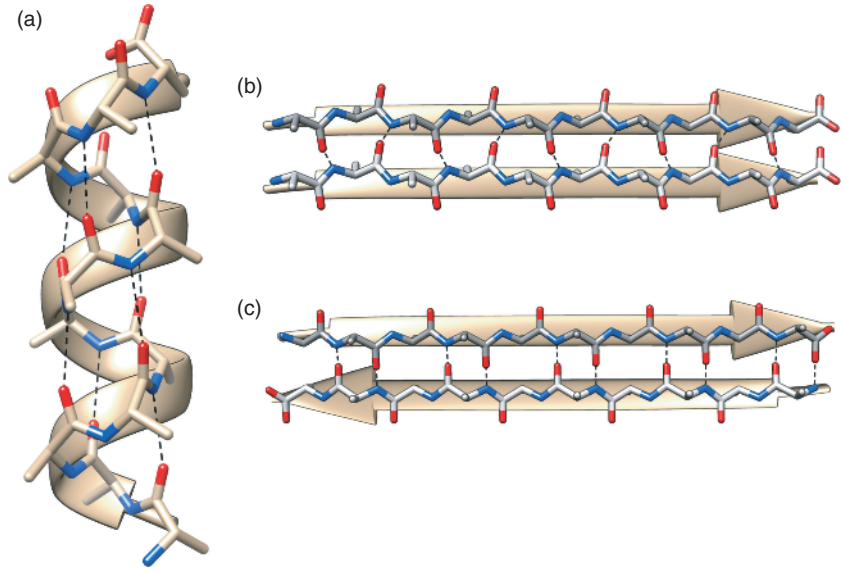


Figure 1.13 Principle regular peptide secondary structures: (a) α-helix (shown as a chain representation superposed on a ribbon representation) and (b, c) β-sheet, including (b) parallel and (c) antiparallel β-sheet structures, superposed on ribbon representations. Hydrogen bonds are shown as dashed lines.

non-standard left-handed α_L structure is also known. Figure 1.12 shows the corresponding region of the Ramachandran diagram. The α-helix may also be represented as a projection along the axis, in a so-called helical wheel diagram. An example is shown in Figure 1.14. Due to the patterning of the carbonyl groups, the α-helix structure has a net dipole moment. Typical structural parameters for regular peptide conformations are listed in Table 1.4. As an example, a decameric (10-mer) peptide adopting an antiparallel β-sheet structure would be calculated to be $10 \times 3.4 = 34$ Å long, while the same peptide forming a parallel β-sheet would be slightly shorter (32 Å). In a fully extended peptide chain, $\Phi = \Psi = \Omega = 180°$.

As well as the ordered secondary structures, peptides can also adopt a disordered structure or coil structure, which is less rigorously known as a random coil structure. The term 'random coil structure' is not recommended since peptides will not adopt a truly random coil due to the finite volume of the chain and steric interactions between side chains.

The transition from α-helix to disordered (or vice versa) can be measured as a function of temperature (or chain length). The transition typically shows a sigmoidal-shaped profile (Figure 1.15) which can be described using the Zimm–Bragg model. This allows for the cooperativity of the transition

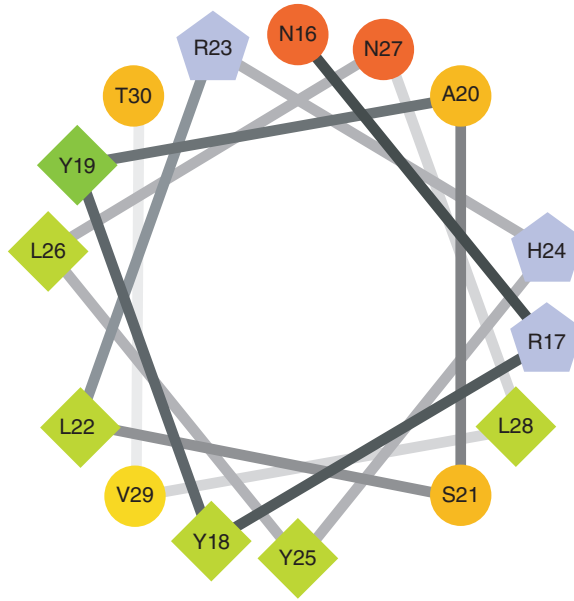


Figure 1.14 Example of a helical wheel diagram of a α-helical peptide sequence. This is a representation of the α-helical domain of peptide $PYY_{3–36}$ (discussed further in Section 5.3.5). The hydrophilic residues are represented as circles, hydrophobic residues as diamonds, and positively charged residues as pentagons. Hydrophobicity is colour coded in shades of green. The positively charged residues are blue. Hydrophilic uncharged residues are coded red-yellow, with pure red being the most hydrophilic residue, and the amount of red decreasing proportionally to the hydrophilicity. This figure was generated using one of several web servers able to generate these diagrams.

Table 1.4 Bond angles and residue spacings for regular peptide conformations.

Structure	Bond angle (deg)			Residues per turn	Translation per residue (Å)
	Φ	**Ψ**	**Ω**		
Antiparallel β-sheet	−139	135	−178		3.4
Parallel β-sheet	−119	113	180		3.2
Right-handed α-helix	−57	−47	180	3.6	1.50
3_{10}-helix	−49	−26	180	3.0	2.00
π-helix	−57	−70	180	4.4	1.15
Polyproline II	−78	149	180	3.0	3.12
Polyglycine II	−80	150	180	3	3.1

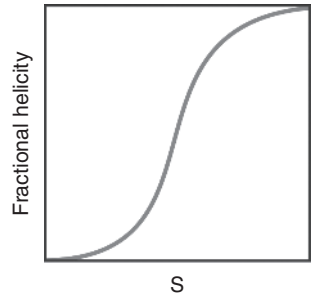


Figure 1.15 The coil–helix transition is cooperative, leading to a sigmoidal shaped profile.

between the two states at the level of each residue which is described by the probability $\sigma \times s$ that a disordered residue (rather confusingly termed a 'coil residue', referring to random coil) is followed by a helical residue (or vice versa). Here $s = [H]/[C]$ is an equilibrium constant expressed in terms of helix and coil concentrations, and σ is a nucleation probability (for the coil state). This leads to a 2×2 matrix of conditional probabilities describing the probabilities for pairs of residues. The model (which is a one-dimensional lattice model) can be computed analytically. The Lifson–Roig model is related to the Zimm–Bragg model but allows for the fact that in α-helices the hydrogen bonds occur after every three residues. The conditional probabilities for residue pairs are represented in a 4×4 transfer matrix. These models can be extended to describe transitions to other helical states (described below), other than α-helical. The Zimm-Bragg model can quantify the average degree of helicity but shows that this can comprise a significant fraction of partially helical chains even if (as in the case of short peptides) the fraction of fully helical sequences is very small.

The β-sheet structure is said to be 'pleated' since C_α atoms are successively above and below the plane of the β-sheet. Most β-sheet structures in proteins and many β-sheet amyloid structures of peptides are twisted. The twist is usually right-handed and arises from relative rotation of residues (the degree of rotation depends on the side chains). Parallel β-sheets are less twisted than antiparallel ones. Antiparallel β-sheet structures are considered more stable for this reason and due to the more regular hydrogen bond geometry (see Figure 1.13).

A strategy to design peptides that adopt β-sheet structures uses alternating residues, especially alternating aliphatic residues and charged residues, e.g. AKAKAK or ELELEL. In some cases, for example silk (discussed further in Section 3.7), alternating residues without charges, e.g. GAGAGA, can favour β-sheet structures.

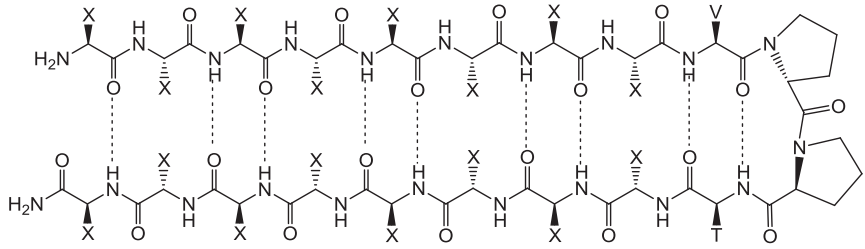


Figure 1.16 Example of a β-hairpin structure.

Where antiparallel β-sheets are required by design, then it is possible to use a β-hairpin motif. An example is that shown in Figure 1.16, which contains intramolecular hydrogen bonds as well as intermolecular hydrogen bonds. β-hairpins are structures with short 2–5 residue loops. A common motif employs the $V^{D}PPT$ motif, as in Figure 1.16. Other methods use non-peptide linkers to control the alignment of strands within molecules to provide a constrained intramolecular antiparallel structure. Proline and glycine are common residues in β-hairpin structures (and also turn structures) since they allow the chain to take up this unusual conformation.

In addition to the α-helix and β-sheet, a number of other regular structures may be adopted by peptides, including helical structures. Examples are shown in Figure 1.17, such as polyproline II (PPII), polyglycine II (PGII), 3_{10} and π helices, and collagen (triple helical) structures. Despite its name, the helical PPII structure can be observed for peptides that do not have a polyproline backbone.

A variety of turn structures are known for peptides and are common at or near the surface of proteins where the peptide chain needs to fold. Turn structures are those in which the backbone changes direction, nearly reversing direction in the case of the main types of turn structures, β- and γ-turns. Many types of turn structure have been identified, including α-, δ-, and π-turns as well as at least nine types of β-turns and two types of γ-turns. These have different backbone torsion angles. Figure 1.18 shows some examples of β- and γ-turn structures; β-turn structures have two non-hydrogen bonded residues in the turn, whereas γ-turns have only one non-hydrogen bonded residue.

The unusual α-sheet structure has recently been reported for several peptides. The structure is related to that of the β-sheet, but all carbonyl groups in a strand are oriented on one side of the sheet, while all the amino groups are oriented on the other side.

Peptide secondary structures are stabilized primarily by hydrogen bonds. Other intermolecular interactions play a role in specific structures and in

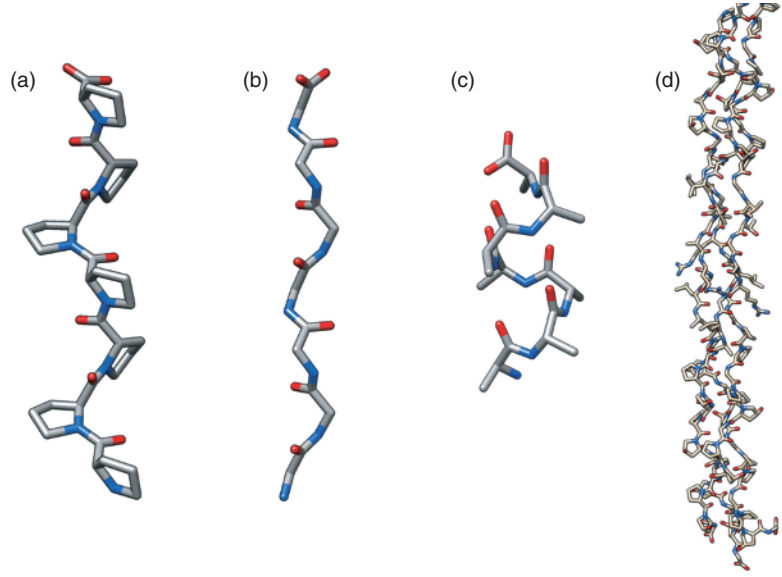


Figure 1.17 Helical peptide secondary structures. (a) PPII structure, (b) PGII structure, (c) 3_{10}-helix (for illustrative purposes only, shown with alanine residues), (d) collagen triple helix.

stabilizing protein tertiary structures. Figure 1.19 shows key interactions that stabilize secondary and tertiary structures. Disulfide bonds are present in many proteins, providing thermal conformational stability. Other amino acids, including tyrosine, can also form cross-links; two tyrosine residues can form dityrosine at high pH. Disulfide bridges are formed oxidatively and be broken by using reducing agents such as DTT (dithiothreitol). Salt bridges stabilize folded protein conformations and coiled coil structures. Hydrophobic interactions stabilize folded tertiary structures and are important in β-sheet aggregation. Aromatic π-stacking interactions are also important in the stabilization of β-sheet amyloid structures, discussed in Chapter 3.

Coiled coil peptides are aggregates of α-helices with facial amphiphilicity. The peptide sequence in each coil in many designed coiled coil peptides may be represented as a heptad abcdefg. For example, the peptide represented by the helical wheel diagram in Figure 1.20a has a hydrophobic face and a partly charged face, and these helices are likely to aggregate into dimeric structures. The packing of α-helices into dimeric coiled coils can also be driven by purely side chain interactions, as in the 'knobs-in-holes' model, proposed by Crick and shown in Figure 1.20b. This model is also

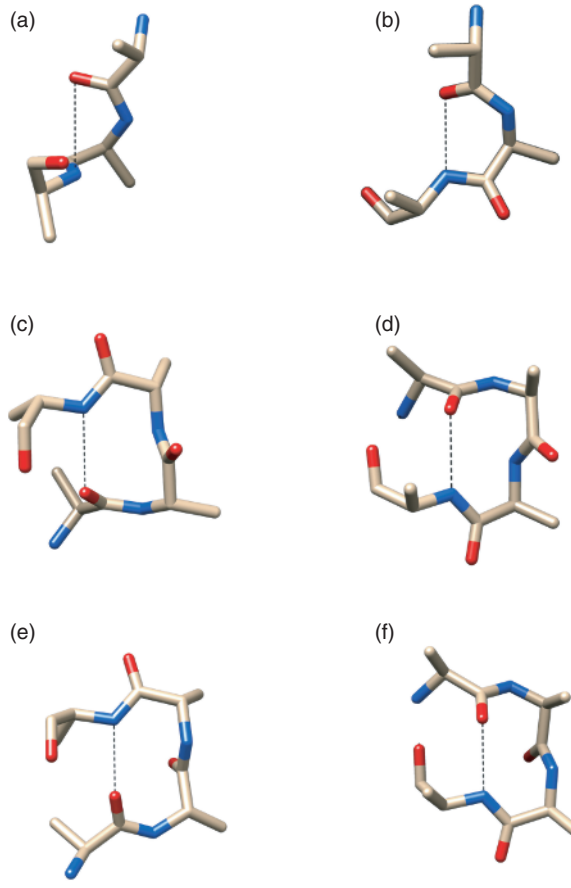


Figure 1.18 Examples of β- and γ-turn structures. (a) Classical γ-turn, (b) inverse γ-turn, (c) type I β-turn, (d) type I′ β-turn, (e) type II β-turn, (f) type II′ β-turn. The dashed lines indicate hydrogen bonds.

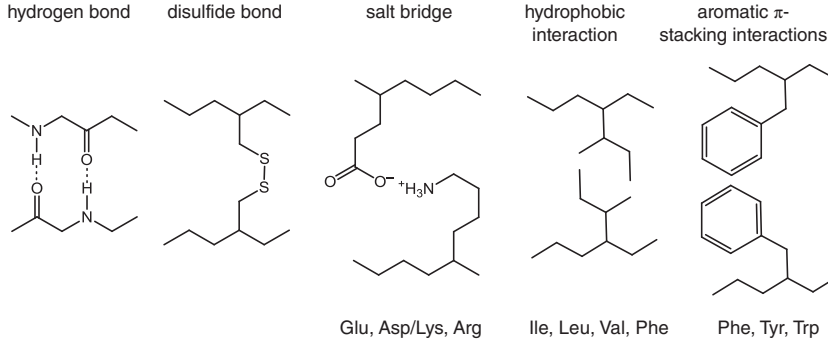


Figure 1.19 Backbone and side chain interactions that stabilize peptide structures.

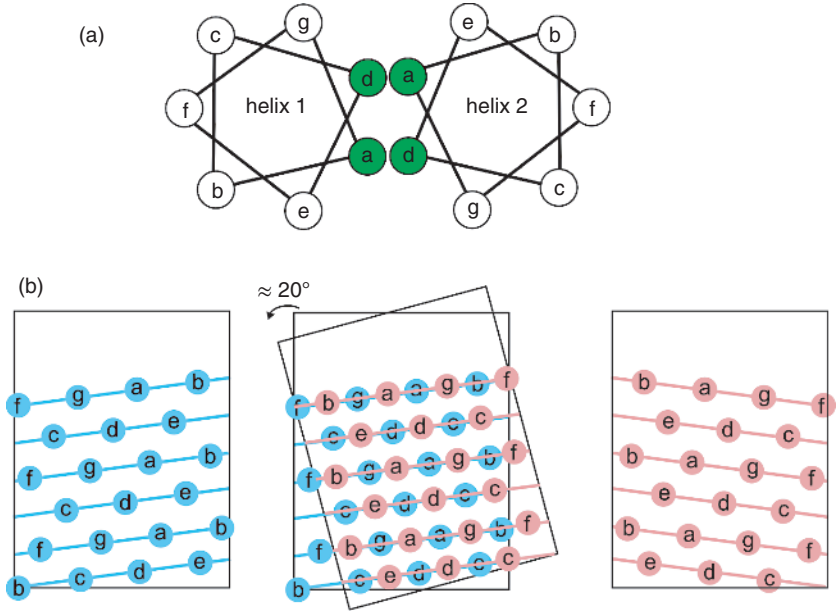


Figure 1.20 (a) A dimeric coiled coil structure based on the abcdefg heptad in α-helical peptides is favoured, for example, by E and K residues in positions d and a (green circles) while the other residues may be hydrophobic. (b) 'Knobs in holes' packing of side chains in a coiled coil dimer.

important in interpreting helix bundle structures in proteins. Figure 1.20b shows projections of the side chains of two helical peptides onto a plane parallel to the helix axes – a so-called helical net diagram. Crossing of the side chains at an angle of about 20° leads to efficient packing of 'knobs in holes'.

A special class of coiled coil stabilized by regular leucine side chain packing are the leucine zipper structures, which are important structural motifs in proteins. Every seventh residue is leucine. Figure 1.21 shows a leucine zipper structure called bZip (basic-region leucine Zipper) which templates (anionic) DNA during transcription in eukaryotes through cationic lysine and arginine residues in the N-terminal region.

Higher-order coiled coil structures are possible (helix bundles) and the sequence can of course be designed to favour such structure by substitution of residues in suitable positions, to favour salt bridging (E–K interactions) and/or hydrophobic interactions. The engineering of coiled coil peptide structures is a fascinating topic that is the subject of ongoing research.

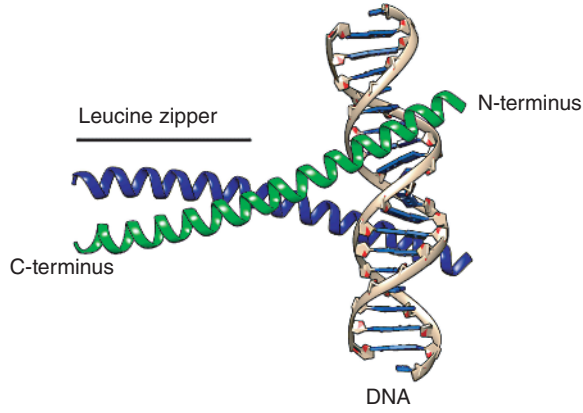


Figure 1.21 Leucine zipper (bZip) structure involved in DNA transcription.

1.5 PEPTIDE STRUCTURE AND CONFORMATION CHARACTERIZATION METHODS

1.5.1 UV/vis Absorbance

For peptides containing tyrosine and/or tryptophan and/or cysteine, the concentration can be obtained from UV absorbance measurements based on the known molar absorptivities (extinction coefficients, ε) of the three aromatic residues. Measurement of UV absorbance at 280 nm enables the extinction coefficient calculated from

$$\varepsilon = (m\mathrm{W} \times 5500) + (n\mathrm{Y} \times 1490) + (p\mathrm{C} \times 125) \tag{1.6}$$

where m, n, and p are the numbers of tryptophan, tyrosine, and cysteine residues, respectively, and 5500, 1490, and 125 are residue absorptivities in units of $\mathrm{M}^{-1}\ \mathrm{cm}^{-1}$. Concentration can then be obtained via the Beer–Lambert law.

For peptides lacking Y or W residues, a common method to determine concentration is to measure UV absorbance at 205 nm, and use an average $\varepsilon(205\ \mathrm{nm}) = 31\ \mathrm{ml\ mg}^{-1}\ \mathrm{cm}^{-1}$ (for example). This method (or using the above formula at 280 nm) is used in commercial UV concentration measurement instruments.

1.5.2 Circular Dichroism Spectroscopy (CD)

Circular dichroism (CD) is a technique to probe the secondary structure of chiral biomolecules, including peptides. The method relies on the differential

absorption of right and left circularly polarized light. The usual method is based on 'fingerprinting' of spectral features in the 190–250 nm far-UV region. Data in the near-UV region can provide information on the conformation of peptides containing aromatic residues with absorption features in the 250–310 nm range. In the far-UV region, characteristic minima are observed in the absorption spectra at (approximately) 208 nm and 222 nm (α-helix) or 216–220 nm (β-sheet), as shown in Figure 1.18. On the other hand, a broad weak minimum in the range 195–200 nm is characteristic of disordered, sometimes known as random coil, conformation. This can be contrasted with the spectra for the polyproline II (PPII) conformation, which is characterized by a minimum in the CD spectrum at around 190–205 nm, along with a broad positive maximum at around 215–225 nm (Figure 1.22).

The CD spectra for the main secondary structures are characterized by typical values of the molar (or alternatively mean residue) ellipticity, as well as the position of the maxima/minima. In fact, CD spectra should be normalized (Eq. (1.7)) for this reason, and also to facilitate comparison between

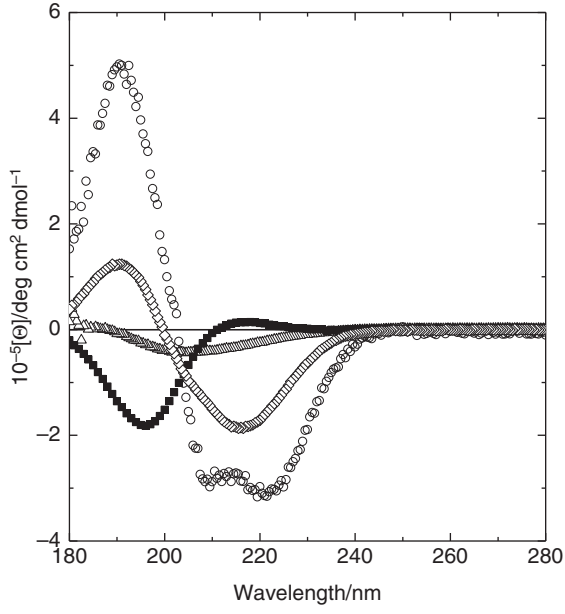


Figure 1.22 Measured CD spectra for peptides. Open circles: Derivative of peptide $PYY_{3–36}$ (discussed in detail in Section 5.3.5) α-helix. Source: Taken from V. Castelletto et al. 2018, diamonds: β-sheet spectrum of lipopeptide C_{16}-KTTβAH Source: Taken from V. Castelletto et al. 2019, closed squares: PPII structure of RAAAAAAAAAR Source: Taken from C. J. C. Edwards-Gayle et al. 2019, open triangles: disordered structure of IKPEAP Source: Taken from J. A. Hutchinson et al. 2019.

samples at different concentrations or measurements in different path length cells. The molar ellipticity is given by

$$[\theta] = \frac{\theta}{10cl} \tag{1.7}$$

Here $[\theta]$ is the molar ellipticity (in units deg cm^2 dmol^{-1}), θ is the measured CD signal amplitude in millidegrees (mdeg), c is the molar concentration, and l is the path-length of the cell in cm. The magnitude of the mean residue ellipticity (MRE, molar ellipticity divided by number of residues) at 222 nm can be used to determine the α-helical content, f_α, of coil peptides via the equation:

$$f_\alpha = 100\,[\theta]_{222}/[\theta]^{\text{ex}}_{222} \tag{1.8}$$

$[\theta]^{\text{ex}}_{222}$ in Eq. (1.3) is the extrapolated value for the molar ellipticity:

$$[\theta]^{\text{ex}}_{222} = [\theta]^{\infty}\left(1 - \frac{k}{n}\right) \tag{1.9}$$

where $[\theta]^{\text{ex}}_{222} = -31\,556\,\text{deg cm}^2\ \text{dmol}^{-1}$, $[\theta]^{\infty} = -37\,400\,\text{deg cm}^2\ \text{dmol}^{-1}$ is the maximum MRE at 222 nm of a peptide with infinite length and 100% helix content, n is the number of residues/helix, and k is a wavelength-dependent constant (2.5 at 222 nm). Comparison of the ratio of the molar ellipticity at 222 and 208 nm gives information on the coiled coil content. Values $[\theta]_{222}/[\theta]_{208} > 0.9$ typically indicate coiled coil formation.

CD spectra for proteins are usually analysed using algorithms based on databases compiled for proteins for which the X-ray crystal structure is known. A range of software is available based on various databases. Most consider only larger proteins although there are limited reference data sets (and curve-fitting programs) for shorter peptides. This type of analysis is of little use for peptides (especially short ones) for which individual residues (especially aromatic residues) and specific conformations (e.g. turns) can dominate the CD spectrum. There is as yet not a comprehensive theory to describe these effects. Theoretical interpretation of CD spectra involves complicated quantum mechanical analysis of electronic states, specifically of electric dipole and magnetic dipole transition moments.

Linear dichroism (LD) is a more specialized method, which is less commonly used than CD. LD refers to the differential absorbance of plane-polarized UV radiation, and it gives information on the alignment of extended objects resulting from peptide self-assembly. Since amyloid fibrils and nanotapes are highly anisotropic; they can align under flow or in other fields and this can be probed using LD, which in particular provides information on the orientation of the peptide backbone and of chromophores such as aromatic residues.

1.5.3 FTIR and Raman Spectroscopy

FTIR (Fourier transform infrared) spectra of peptides may be measured in transmission mode from solutions in a closed cell or from a drop deposited on a crystal in ATR (attenuated total reflectance) mode or from dried films (less preferred since the spectrum does not correspond to a peptide with conformation under the solution conditions). FTIR is sensitive to the vibrational modes of bonds within peptides. Specific regions of the spectrum are the focus of particular analysis.

The amide I region in the range 1600–1700 cm^{-1} is sensitive to the modes of CO, CN, and NH groups, which are influenced by H-bonding. A prime is added to the name of the region of the spectrum, for example to give the term amide I′ for spectra measured in D_2O. All band positions mentioned above are slightly downshifted if measurements are performed in D_2O. In fact, measurements in D_2O are beneficial since water absorption features around 1650 cm^{-1} can be avoided. Characteristic FTIR bands of peptides and proteins are listed in Table 1.5.

The amide I region gives information on secondary structure, via 'fingerprinting' or peak-fitting methods. These are prone to uncertainty associated with the overlap of peaks in the spectra, although clear features of secondary structures such as β-sheets can be resolved by performing measurements in transmission mode on a modern FTIR instrument, with sufficiently concentrated samples in D_2O in narrow path-length cells. A peak typically in the 1620–1640 cm^{-1} range is associated with β-sheet structures. A peak in the typical range 1648–1657 cm^{-1} is associated with α-helix structure, whilst disordered peptides give a peak in the typical range 1642–1650 cm^{-1}. The

Table 1.5 Typical FTIR band ranges for peptides.

Band	Approximate wavenumber (range)	Assignment
Amide A	3270–3310	N–H stretch
Amide B	3030–3100	N–H stretch
Amide I	1600–1700	Mainly C=O stretch with contribution from C–N stretch and N–H bend
Amide II	1500–1600	Combination of N–H in-plane bend and C–N stretch with minor contributions from other modes
Amide III	1200–1400	N–H bend and C–H stretch, C=O stretch, OCN bend, and others

narrow intense band observed for some peptides at 1675–1695 cm^{-1} is usually ascribed to antiparallel β-sheet structure. Caution is required since a peak in the amide I region of an FTIR spectrum close to 1673 cm^{-1} is due to residual trifluoroacetic acid from the peptide synthesis, bound to peptide cations (unless this is removed by ion-exchange methods). Tables of FTIR peaks associated with side groups in some residues are available. A peak near 1705 cm^{-1} may be assigned to the carbonyl stretch, e.g. in acidic side residues or from the C terminus.

The amide II band around 1550 cm^{-1} mainly results from the N–H bending vibrations which are highly sensitive to deuteration (the deuterium from D_2O exchanges positions with hydrogen from the N–H bond). As a consequence the amide II′ band is shifted by approximately 100 cm^{-1} to 1450 cm^{-1} in D_2O.

Raman spectra provide similar information on peptide secondary structure and side chain deformation to FTIR, since the vibrational transition selection rules are generally satisfied for peptide deformation modes for both Raman and FTIR spectroscopy. As it is a scattering-based method, Raman spectroscopy also typically requires a higher sample concentration than FTIR to get a useful spectrum. However, measurement of Raman spectra can be recorded in H_2O, in contrast to FTIR where the water absorption band prevents measurements in the vital amide I region. Raman microscopy enables the chemical mapping of peptide samples scanning across the sample at micron resolution and measuring Raman spectra.

Isotope labelling of peptides enables specific features in the FTIR spectrum to be resolved, as bands associated with deformation modes of labelled residues will be shifted. Typically, isotopic substitution using a ^{13}C-labelled peptide splits amide I bands for β-sheets into higher and lower frequency peaks. The magnitude of the ^{12}C band wavelength shift depends on the extent of perturbation of the ^{12}C carbonyl coupling induced by ^{13}C substitution and strongly coupled β-sheets are generally sensitive to such isotopic substitution. This can be used to infer information on the registry of β-strands.

FTIR can be extended to study linear dichroism (polarized FTIR) on aligned samples with isotope labelling and to vibrational circular dichroism (VCD). VCD is analogous to (UV) electronic CD, but extended to the IR range of the spectrum, sensitive to bond vibrations. The analogous technique using Raman scattering is Raman optical activity (ROA). Polarized Raman spectroscopy can provide information on the orientation of specific features such as aromatic residues.

Recently, two-dimensional (2D) IR methods have attracted great interest as it is possible to correlate vibrational modes, providing information on mode coupling. When combined with isotope editing of peptides, 2D IR can provide detailed information on β-sheet conformation and the pathway of amyloid formation (a topic discussed further in Sections 3.2 and 3.4).

1.5.4 NMR Spectroscopy

This section does not consider the routine use of NMR (nuclear magnetic resonance) to analyse synthesized peptide structures, but is rather focused on specialized NMR techniques to probe peptide aggregates.

The secondary structure of a peptide or protein can be analysed via the chemical shift index (CSI), which compares the chemical shift of $^{1}H_{\alpha}$ or ^{13}C backbone units to those for the corresponding residue in an ideal disordered peptide. Figure 1.23 shows an example for the single immunoglobulin binding domain, which comprises 56 residues (including the NH_2-terminal Met) of protein G from group G *Streptococcus*. The CSI is consistent with the structure for this small protein which comprises four β-strands and an α-helix. Subsequent to the development of the CSI, other methods such as probability-based secondary structure identification (PSSI) have been introduced to analyse peptide chemical shift data.

Information on secondary structure can also be obtained from NOEs (nuclear Overhauser effects), since these depend on the separation between nuclei.

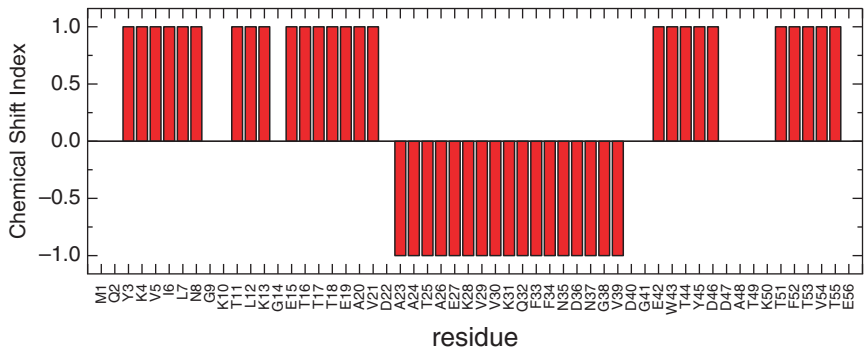


Figure 1.23 Chemical shift index plot (for $^{1}H\alpha$) for 56-residue sequence from protein G. Strands correspond to +1 states, helices to −1 and coil (disordered) to 0.

Two-dimensional NMR has become a standard method for more detailed analysis of peptides via correlation or exchange spectroscopic methods. The most commonly employed for peptide structure analysis are NOESY (nuclear Overhauser effect spectroscopy), TOCSY (total correlation spectroscopy), and DQF-COSY (double quantum filtered correlation spectroscopy) for homonuclear measurements and HSQC (heteronuclear single-quantum correlation spectroscopy) for heteronuclear studies (e.g. ^{1}H–^{15}N, ^{1}H–^{13}C). For peptides with fewer than 50 residues, peak assignments are possible using homonuclear 2D NMR methods; for longer peptides heteronuclear experiments using isotopically labelled samples can be used. These can also be used to probe strand alignment in β-sheet structures (composed of identical peptides) which are otherwise impossible to distinguish.

Amide proton exchange rates also provide information on secondary structure, since these are influenced by the hydrogen bonding that is characteristic of secondary structure formation.

Solution NMR can be used to provide high-resolution peptide structures, such as those of helical peptides, via 2D correlation methods – as described above. 2D correlation methods are used to provide distance and backbone torsional angle information that can then be used, along with hydrogen bond and side chain restraints, as constraints in molecular models of the conformation.

The aggregation state of many peptides has been probed using solution ^{1}H NMR. The method can also be used to determine the degree of ionization of termini or side chains, for example as a function of concentration. Likewise, ^{1}H NMR solubility measurements have been used to obtain solubility phase diagrams and related diffusion measurements have enabled the determination of critical aggregation concentrations (Section 3.12.1). Specific chemical information, for example on metal ion binding to peptides, can also be obtained from solution ^{1}H NMR.

Solid state NMR (ss-NMR) has provided much detail on amyloid fibril structure. A number of high-resolution experiments employing magic angle spinning (MAS) may be performed using isotopically labelled ^{13}C or ^{15}N peptides. MAS is a technique to improve the spectral resolution by averaging orientation-dependent dipolar interactions. Using this method, information on strand registry can be obtained. Homonuclear and heteronuclear 2D and 3D NMR spectroscopy enable information on interatomic distances and torsion angles to be obtained for isotopically labelled peptides in the dried state.

Hydrogen/deuterium exchange techniques have also been employed to probe structural changes during amyloid aggregation. Information on structural rearrangements of subsegments of a protein or peptide during folding, unfolding, or fibrillization can be obtained from exchange rate measurements. This method has also been used, for example, to provide a 3D structure for the amyloid β peptide Aβ42 (Section 3.3).

1.5.5 X-Ray Diffraction

Single-crystal X-ray diffraction (XRD) can be used to obtain the crystal structure of peptides, although there are relatively few reports on this for aggregating peptides due to the difficulty in growing a crystal from a sample in which self-assembly occurs (in preference to crystal formation). In particular, with a very few exceptions, amyloid peptides do not crystallize, prohibiting single-crystal XRD. This is because fibrillar assemblies are by their nature non-crystalline, one- or two-dimensional arrays with a low degree of molecular ordering. Dehydration of the sample to prepare a crystal can disrupt the hydrogen bonding of both α-helical and β-sheet structures. However, with careful preparation, it is possible to obtain single-crystal XRD patterns from small amyloid peptide fragments (Section 3.6).

Fibre diffraction is an alternative to single-crystal XRD that is particularly suited to analyse the structure of peptides which form fibrillar assemblies, including amyloid peptides, coiled coil peptides, lipopeptide fibrils, and peptide nanotubes. Fibre XRD is performed on dried samples, in the form of films or 'stalks', the latter being dried threads of solution. Other methods of alignment include the use of stretch frames or cryo-loops, the latter producing a dried flat film or 'mat'.

1.6 PEPTIDE DATABASES AND WEB SOFTWARE

Many web-based servers are available to calculate simple properties of peptides such as molar mass and pI value, given the input sequence. There are too many of these (and website operability varies) to list. Table 1.6 lists examples of freely downloadable software useful for more detailed peptide property calculation, along with web resources such as databases of different classes of peptides, aggregation propensity calculators, and others.

Table 1.6 Examples of useful peptide software and websites.

Web-server (WS) /stand-alone software (SAS)	Purpose	Website
molinspiration	Useful to calculate molar mass/volume and rule-of-5 parameters, e.g. log*P* for structures (e.g. peptide conjugates) drawn on web applet	https://www.molinspiration.com/cgi-bin/properties
Pymol (SAS)	Graphics for pdb files and built sequences, H-bond analysis, etc.	https://pymol.org/edu/?q=educational A fuller featured commercial version is also available
Chimera (SAS)	Powerful graphics capability for pdb files and built sequences and various analysis tools including H-bond analysis, solvation, Ramachandran plot generator, and others	https://www.cgl.ucsf.edu/chimera
Uppsala Ramachandran Server	Ramachandran plots (from pdb files)	http://eds.bmc.uu.se/ramachan.html
Peptides package for R (SAS, R is a package for statistical computing)	Various properties including hydropathy analysis, hydrophobic moment calculation	https://cran.r-project.org/web/packages/Peptides/index.html
Transporter classification database	Hydropathy analysis, hydrophobic moment calculation, helical wheel plots, sequence alignment	http://www.tcdb.org/analyze.php
Pepwheel	Helical wheel projection plots	http://www.bioinformatics.nl/cgi-bin/emboss/pepwheel
heliquest	Properties of different helical structures including hydrophobicity, hydrophobic moment, and helical wheel representations	http://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParams.py

Table 1.6 (*continued*)

Web-server (WS) /stand-alone software (SAS)	Purpose	Website
ISAMBARD	Biomolecular design and analysis, particular useful for coiled coils	https://github.com/isambard-uob/isambard
CC Builder	Coiled coil peptide modelling and design	http://coiledcoils.chm.bris.ac.uk/ccbuilder2/builder
LOGICOIL	Predicts oligomerization state of coiled coils	http://coiledcoils.chm.bris.ac.uk/LOGICOIL
ExPASy	Huge range of tools for coiled coil calculations/M_w and pI calculations, pdb viewing, sequencing, and others (mostly for proteins)	https://www.expasy.org/structural_bioinformatics
PepFold	Secondary structure prediction from sequence	http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3
SOPMA	Secondary structure prediction from sequence	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html
BLAST	Sequence database for alignment checking	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&
Biophython	Python processing for BLAST, ExPASy searches, and analysis, etc.	https://biopython.org
FASTA	Sequence similarity search	www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=fasta-I20191004-135659-0538-29792313-p2m
TANGO	Aggregation domain (amyloid) prediction	http://tango.crg.es
Waltz	Aggregation domain (amyloid) prediction	http://waltz.switchlab.org
Camsol and S2D	Aggregation (amyloid) propensity prediction	http://www-mvsoftware.ch.cam.ac.uk

(*continued*)

Table 1.6 (*continued*)

Web-server (WS) /stand-alone software (SAS)	Purpose	Website
Dichroweb	Secondary structure estimation from CD spectra	http://dichroweb.cryst.bbk.ac.uk/html/home.shtml
Bestsel	Secondary structure estimation from CD spectra	http://bestsel.elte.hu/index.php
δ2D and camcoil	Predict secondary structures from chemical shifts and calculate random coil chemical shifts from sequence	http://www-mvsoftware.ch.cam.ac.uk
Biological Magnetic Resonance Databank	Protein and peptide chemical shift data	http://bmrb.wisc.edu
Bayreuth University TOCSY and COSY amino acid spectra	Amino acid 2D TOCSY and COSY spectra	http://www.bp.uni-bayreuth.de/NMR/nmr_aminotocsy.html
Protein Data Bank	Peptide structures from XRD, NMR, and electron microscopy	http://rcsb.org
APD2/3	Natural antimicrobial peptide database	http://aps.unmc.edu/AP
CAMP	Antimicrobial peptide database	http://www.camp.bicnirrh.res.in
YADAMP	Antimicrobial peptide database	http://yadamp.unisa.it
DBAASP	Database of antimicrobial peptides and structure prediction (hydrophobicity, tilt angle, pI, etc.)	https://dbaasp.org/home
MilkAMP Database	Milk antimicrobial peptide database	http://milkampdb.org/home.php
THpdb	Database of FDA-approved therapeutic peptides and proteins	http://crdd.osdd.net/raghava/thpdb

BIBLIOGRAPHY

Barth, A. (2007). Infrared spectroscopy of proteins. *Biochimica Et Biophysica Acta-Bioenergetics* 1767: 1073–1101.

Branden, C. and Tooze, J. (1999). *Introduction to Protein Structure*. New York: Garland Publishing.

Castelletto, V. and Hamley, I.W. (2018). Methods to characterize the nanostructure and molecular organization of amphiphilic peptide assemblies. In: *Peptide Self-Assembly: Methods and Protocols* (eds. B.L. Nilsson and T.M. Doran), 3–21. Totowa: Humana Press Inc.

Castelletto, V., Hamley, I.W., Seitsonen, J. et al. (2018). Conformation and aggregation of selectively PEGylated and lipidated gastric peptide hormone human PYY_{3-36}. *Biomacromolecules* 19: 4320–4332.

Castelletto, V., Edwards-Gayle, C.J.C., Greco, F. et al. (2019). Self-assembly, tunable hydrogel properties and selective anti-cancer activity of a carnosine-derived lipidated peptide. *ACS Applied Materials & Interfaces* 11: 33573–33580.

Chou, P.Y. and Fasman, G.D. (1974). Prediction of protein conformation. *Biochemistry* 13: 222–245.

Creighton, T.E. (1993). *Proteins. Structures and Molecular Properties*. New York: W.H. Freeman.

Crick, F.H.C. (1953). The packing of α-helices: simple coiled coils. *Acta Crystallographica* 6: 689–697.

Edwards-Gayle, C.J.C., Castelletto, V., Hamley, I.W. et al. (2019). Self-assembly, antimicrobial activity and membrane interactions of arginine-capped peptide bola-amphiphiles. *ACS Applied Bio Materials* 2: 2208–2218.

Eisenberg, D., Weiss, R.M., Terwilliger, T.C., and Wilcox, W. (1982). Hydrophobic moments and protein structure. *Faraday Symposia of the Chemical Society* 17: 109–120.

Hamley, I.W. (2007). Peptide fibrillisation. *Angewandte Chemie, International Edition in English* 46: 8128–8147.

Hutchinson, J.A., Hamley, I.W., Torras, J. et al. (2019). Self-assembly of lipopeptides containing short peptide fragments derived from the gastrointestinal hormone PYY_{3-36}: from micelles to amyloid fibrils. *Journal of Physical Chemistry B* 123: 614–621.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta* 1751: 119–139.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105–132.

Langel, U., Cravatt, B.F., Gräslund, A. et al. (2010). *Introduction to Peptides and Proteins*. Boca Raton: CRC Press.

Mielke, S.P. and Krishnan, V.V. (2009). Characterization of protein secondary structure from NMR chemical shifts. *Progress in Nuclear Magnetic Resonance Spectroscopy* 54: 141–165.

Moore, D.S. (1985). Amino acid and peptide net charges: a simple calculational procedure. *Biochemical Education* 13: 10–11.

Nordén, B., Rodger, A., and Dafforn, T.R. (2010). *Linear Dichroism and Circular Dichroism: A Textbook on Polarized-Light Spectroscopy*. Cambridge: RSC.

Phoenix, D.A., Dennison, S.R., and Harris, F. (2013). *Antimicrobial Peptides*. Weinheim, Germany: Wiley-VCH.

Schneider, J.P., Pochan, D.J., Ozbas, B. et al. (2002). Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. *Journal of the American Chemical Society* 124: 15030–15037.

Sewald, N. and Jakubke, H.-D. (2002). *Peptides: Chemistry and Biology*. Weinheim: Wiley-VCH.

Stuart, B. (1997). *Biological Applications of Infrared Spectroscopy*. Chichester: Wiley.

Su, J.Y., Hodges, R.S., and Kay, C.M. (1994). Effect of chain length on the formation and stability of synthetic α-helical coiled coils. *Biochemistry* 33: 15501–15510.

Surewicz, W.K. and Mantsch, H.H. (1988). New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochemica et Biophysica Acta* 952: 115–130.

Van Vranken, D. and Weiss, G. (2013). *Introduction to Bioorganic Chemistry and Chemical Biology*. New York: Garland Science.

Voet, D. and Voet, J.G. (1995). *Biochemistry*. New York: Wiley.

Westermann, J.-C. and Craik, D.J. (2008). NMR in peptide drug development. In: *Peptide-Based Drug Design* (ed. L. Otvos). Totowa, New Jersey: Humana Press.

White, S.H. and Wimley, W.C. (1999). Membrane protein folding and stability: physical principles. *Annual Review of Biophysics and Biomolecular Structure* 28: 319–365.

Wishart, D.S. (2011). Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy* 58: 62–87.