

Chapter 1

BIOMOLECULES

In this chapter, you will learn about . . .

- ◆ . . . the biomolecules that are most commonly analysed in bioanalytical chemistry: amino acids, proteins and nucleic acids.
 - ◆ . . . the structure of these biomolecules and their physical and chemical characteristics.
 - ◆ . . . some of the functions of these biomolecules and how they interact with each other in the cell.
-

Chemists are likely to be familiar with certain biomolecules such as carbohydrates and lipids from their organic chemistry lectures. However, many do not have a clear understanding of the composition and function of other biomolecules such as proteins and DNA. This chapter introduces the biomolecules, which are the target of the analytical methods described in the following chapters.

1.1 Amino Acids, Peptides and Proteins

Amino acids are the building blocks for peptides and proteins and play an important part in metabolism. 20 different amino acids are found in living organisms. They can connect to each other via peptide bonds to form long chains. Proteins may consist of thousands of amino acids and can have molecular weights of up to several million Dalton (Da). Shorter chains of up to a few hundred amino acids are referred to as peptides. The sequence of the amino acids within the molecule is essential for the structure and function of proteins and peptides in biological processes.

1.1.1 Amino Acids

The general structure of an amino acid is shown in Fig. 1.1. It consists of a tetrahedral carbon atom (C-alpha) connected to four groups: a basic amino group ($-\text{NH}_2$), an acidic carboxyl group ($-\text{COOH}$), a hydrogen atom ($-\text{H}$) and a substituent group ($-\text{R}$), which varies from one amino acid to another. The amino group is in the alpha position relative to the carboxyl group, hence the name α -amino acids. Amino acids are chiral with the exception of glycine, where the R substituent is a hydrogen atom. All natural amino acids have the same absolute configuration: the L-form in the Fischer convention or the S-form according to the Cahn-Ingold-Prelog rules, with the exception of cysteine, which has the R-configuration.

Amino acids can be classified according to their substituent R groups (Fig. 1.2 to Fig. 1.8): in *basic amino acids*, R contains a further amino group, whereas in *acidic amino acids*, R contains a further carboxyl group. In addition, there are *aliphatic*, *aromatic*, *hydroxyl containing* and *sulfur containing amino acids* according to the nature of the substituent, as well as a *secondary amino acid*.

For convenience, the names for amino acids are often abbreviated to either a *three symbol* or a *one symbol short form*. For example, Arginine can be referred

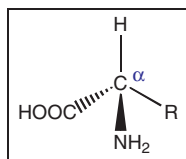


Fig. 1.1. General structure of an α -L-amino acid.

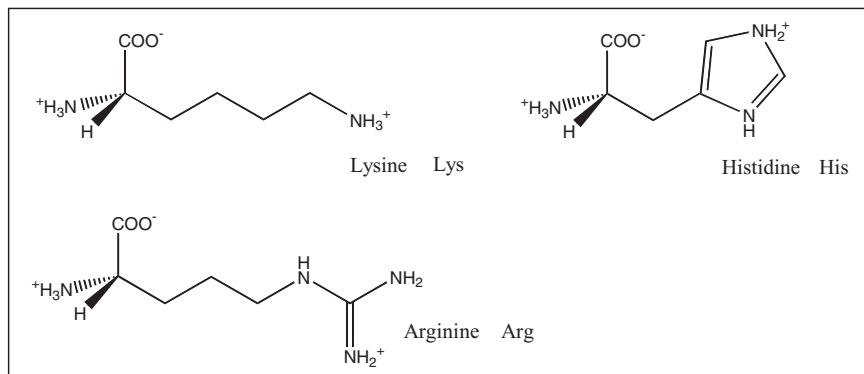


Fig. 1.2. Basic amino acids.

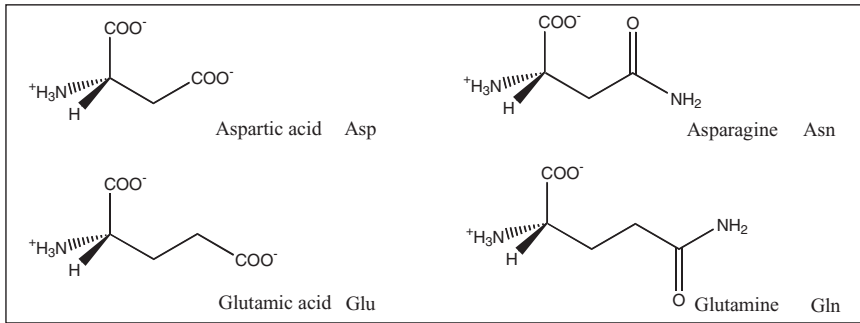


Fig. 1.3. Acidic amino acids.

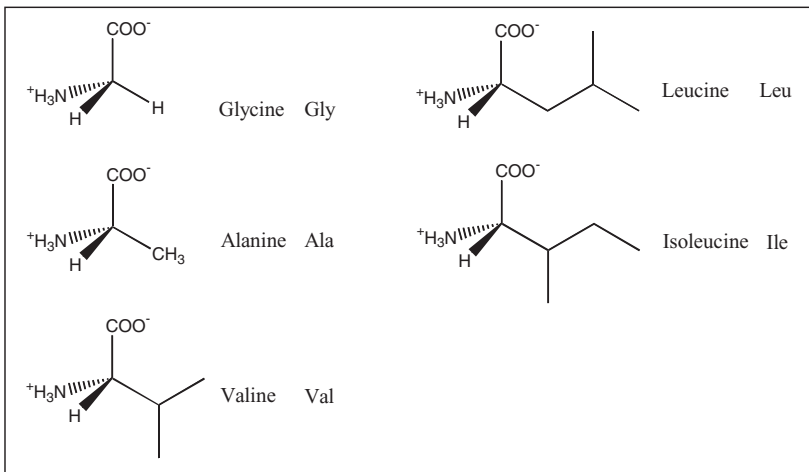


Fig. 1.4. Aliphatic amino acids.

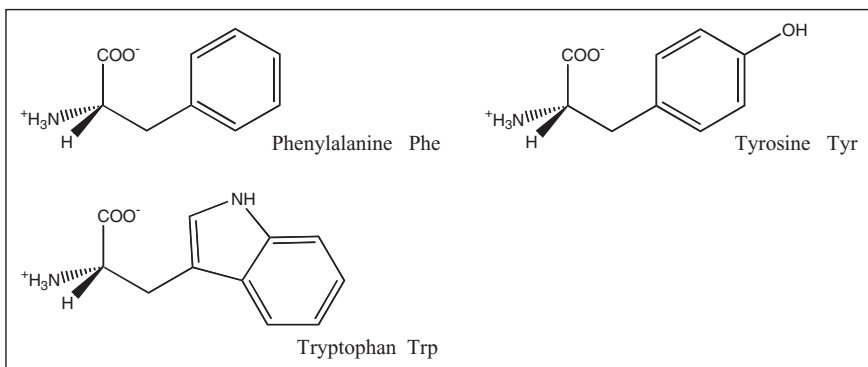


Fig. 1.5. Aromatic amino acids.

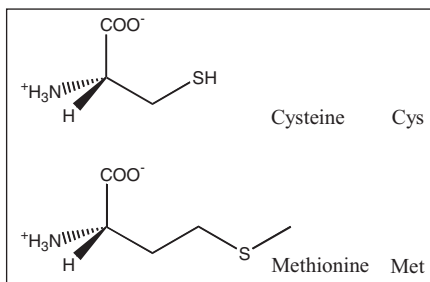


Fig. 1.6. Sulfur containing amino acids.

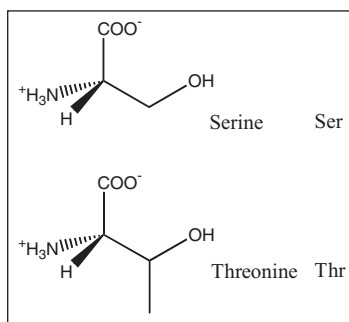


Fig. 1.7. Amino acids with an alcoholic hydroxyl group.

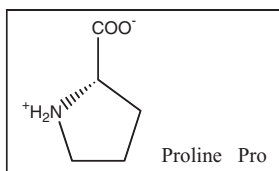


Fig. 1.8. Secondary amino acid.

to as Arg or R and Glycine can be shortened to Gly or G. The abbreviations for the 20 natural amino acids are listed in Table 1.1. These naturally occurring amino acids are the building blocks of peptides and proteins. Any particular amino acid is not likely to exceed 10 % of the total composition of a protein (see Table 1.1).

Amino acids can also be classified according to their polarity and charge at pH 6 to 7, which corresponds to the pH range found in most biological systems. This is often referred to as the *physiological pH*. *Non-polar amino acids* with no

Table 1.1. Natural amino acids.

Name	Three and one letter symbols		M _r (Da)	found ⁽¹⁾ (%)	pK ₁ ⁽²⁾ α-COOH	pK ₂ ⁽²⁾ α-NH ₃ ⁺	pK _R ⁽²⁾ side-chain
<i>basic amino acids</i>							
Lysine	Lys	K	146.2	5.9	2.16	9.06	10.54 ε-NH ₃ ⁺
Histidine	His	H	155.2	2.3	1.8	9.33	6.04 imidazole
Arginine	Arg	R	174.2	5.1	1.82	8.99	12.48 guanidino
<i>acidic amino acids</i>							
Aspartic acid	Asp	D	133.1	5.3	1.99	9.90	3.90 β-COOH
Glutamic acid	Glu	E	147.1	6.3	2.10	9.47	4.07 γ-COOH
Asparagine	Asn	N	132.1	4.3	2.14	8.72	
Glutamine	Gln	Q	146.2	4.3	2.17	9.13	
<i>aliphatic amino acids</i>							
Glycine	Gly	G	75.1	7.2	2.35	9.78	
Alanine	Ala	A	89.1	7.8	2.35	9.87	
Valine	Val	V	117.2	6.6	2.29	9.74	
Leucine	Leu	L	131.2	9.1	2.33	9.74	
Isoleucine	Ile	I	131.2	5.3	2.32	9.76	
<i>aromatic amino acids</i>							
Phenylalanine	Phe	F	165.2	3.9	2.20	9.31	10.46 phenol
Tyrosine	Tyr	Y	181.2	3.2	2.20	9.21	
Tryptophan	Trp	W	204.2	1.4	2.46	9.41	
<i>sulfur containing amino acids</i>							
Cysteine	Cys	C	121.2	1.9	1.92	10.70	8.37 sulfhydryl
Methionine	Mel	M	149.2	2.2	2.31	9.28	
<i>amino acids with alcoholic hydroxyl groups</i>							
Serine	Ser	S	105.1	6.8	2.19	9.21	
Threonine	Thr	T	119.1	5.9	2.09	9.10	
<i>amino acid with secondary amino group</i>							
Proline	Pro	P	115.1	5.2	1.95	10.64	

Sources:

(1) R. F. Doolittle, *Database of nonredundant proteins*, in G. D. Fasman (Ed.), *Predictions of Protein Structure and the Principles of Protein Conformation*, Plenum Press, 1989.

(2) R. M. C. Dawson, D. C. Elliott, W. H. Elliott, K. M. Jones, *Data for Biochemical Research*, 3rd edition, Oxford Science Publications, 1986.

net charge are Alanine, Valine, Leucine, Isoleucine, Phenylalanine, Tryptophan, Methionine and Proline. *Polar amino acids* have no net charge but carry a polar group in the substituent R. Glycine, Asparagine, Glutamine, Tyrosine, Cysteine, Serine and Threonine fall into this category. *Positively charged amino acids* at physiological pH are Lysine, Histidine and Arginine; whereas *negatively charged amino acids* are Aspartic acid and Glutamic acid.

In addition to the 20 natural amino acids, there are other amino acids, which occur in biologically active peptides and as constituents of proteins. These will not be covered in this textbook.

1.1.1.1 Zwitterionic character, pK and pI

As amino acids contain a basic and an acidic functional group, they are *amphoteric*. The carboxyl group of an amino acid has a pK between 1.8 and 2.5, the amino group has a pK between 8.7 and 10.7 (see Table 1.1). At the pH found under physiological conditions, pH 6 to 7, the amino group is ionised to $-\text{NH}_3^+$ and the carboxyl group is ionised to $-\text{COO}^-$. Hence, at physiological pH amino acids are *zwitterionic*. At low pH values, the carboxyl group is protonated to $-\text{COOH}$ and the amino acid becomes positively charged. At high pH values, the amino group is deprotonated to $-\text{NH}_2$ and the amino acid becomes negatively charged (Fig. 1.9). Functional groups in the substituents may have different pK values as well (see Table 1.1).

For every amino acid, there is a specific pH value at which it exhibits no net charge. This is called the *isoelectric point*, pI . At its isoelectric point, an amino acid remains stationary in an applied electric field, i.e. it does not move to the positive or negative pole. The isoelectric point can be estimated via the *Henderson-Hasselbalch equation*:

$$pI = \frac{1}{2}(pK_i + pK_j) \quad (\text{equation 1.1})$$

where pK_i and pK_j are the dissociation constants of the ionisation steps involved. This calculation is straightforward for mono-amino and mono-carboxylic acids, where pK_i and pK_j are the pK values of the amino group and the carboxylic group, respectively. For amino acids with ionisable side chains, the calculation of the pI value is more complex. The pI values for the natural amino acids are listed in Table 1.2, and in Table 1.3 pI values are given for some proteins. Differences in pI can be utilised to separate amino acids or proteins in an electric field. This technique is called isoelectric focussing and will be discussed in detail in sections 3.2.4 and 3.3.3.

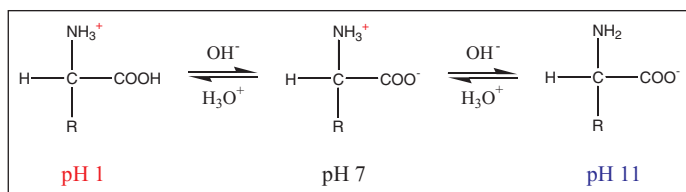


Fig. 1.9. Charge of an amino acid at different pH values: zwitterionic character at pH 7, positive charge at low pH and negative charge at high pH.

Table 1.2. pI values of natural amino acids.

Amino acids Non-polar chain	pI	Amino acids Polar chain	pI	Amino acids Charged chain	pI
Alanine	6.02	Glycine	5.97	Lysine	9.74
Valine	5.97	Asparagine	5.41	Histidine	7.58
Leucine	5.98	Glutamine	5.65	Arginine	10.76
Isoleucine	6.02	Tyrosine	5.65	Aspartic acid	2.87
Phenylalanine	5.98	Cysteine	5.02	Glutamic acid	3.22
Tryptophan	5.88	Serine	5.68		
Methionine	5.75	Threonine	6.53		
Proline	6.10				

Table 1.3. pI values of some proteins.

Protein	pI	Protein	pI
Pepsin	<1.0	Myoglobin (horse)	7.0
Ovalbumin (hen)	4.6	Hemoglobin (human)	7.1
Serum albumin (human)	4.0	Ribonuclease A (bovine)	7.8
Tropomyosin	5.1	Cytochrome c (horse)	10.6
Insulin (bovine)	5.4	Histone (bovine)	10.8
Fibrinogen (human)	5.8	Lysozyme (hen)	11.0
γ -Globuline (human)	6.6	Salmine (salmon)	12.1
Collagen	6.6		

1.1.2 Peptides and Proteins

Peptides and proteins are *macromolecules* made up from long chains of amino acids joined head-to-tail via *peptide bonds*. The three-dimensional structure of a protein is very well defined and is essential for it to function. Proteins are found

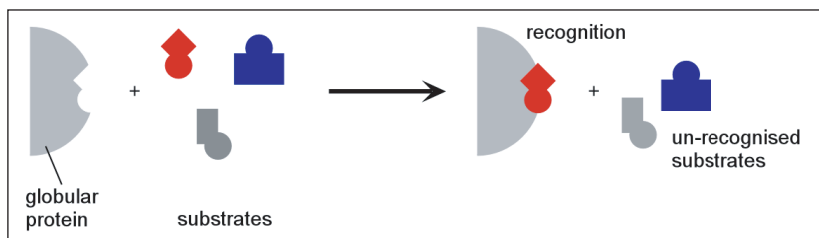


Fig. 1.10. Globular proteins like enzymes and antibodies have a specific surface that recognises only specific substrates.

in all forms of living organisms and perform a wide variety of tasks. The function and structure of proteins are outlined in the following sections.

1.1.2.1 *The biological function of proteins*

In general, there are two types of protein structures: (1) *fibrous*, elongated proteins which are not soluble in water and provide structural support and (2) *globular* spherical proteins which are water soluble and have specific functions in the immune system and metabolism.

Globular proteins have a compact, spherical structure with very characteristic grooves and peaks on their surface. Analogous to a key fitting into a lock, other molecules fit into these grooves and peaks. This makes globular proteins *specific* when it comes to interacting with or recognising other molecules (Fig. 1.10). *Enzymes* are an example of such specific proteins. They are biochemical catalysts, which lower the activation energy and, thus, accelerate immensely the reaction rate of biological reactions. An enzyme can only react with a substrate if the location of its functional groups and hydrogen bonds as well as its shape matches the active site of the enzyme. Ribonuclease for example is an enzyme secreted by the pancreas to specifically digest ribonucleic acid (RNA). *Antibodies* are another example of highly specific globular proteins. They can recognise intruders, *antigens*, and bind to them in a *key-lock mechanism*. Enzymes and antibodies are used as molecular recognition elements in bioassays (section 5.1) and biosensors (section 5.2).

In the body, proteins also function as *transport* and *storage media*. For example, haemoglobin is responsible for the transport of oxygen in the blood stream, transferrin for the transport of iron. Ferritin is an example of a protein with a storage function, which can be found in the liver. It forms a complex with iron, and thus binds and stores the metal. In the form of *hormones*, polypeptides can also act as *chemical messengers*. By interacting with a matching receptor, usually found in the cell membrane, they regulate a wide variety of tasks in metabolism. For

example, three hormones found in the pancreas, glucagon, insulin and somatostatin, regulate the storage and release of glucose and fatty acids. Other hormones control digestion, growth and cell differentiation. Hormones form a large class of chemical substances. Most hormones are polypeptides, however, some are amino acid derivatives or steroids.

Fibrous proteins have a high tensile strength and mechanical stability. Their function is to provide *structural support* to tissues. Collagen, for example, gives connective strength to skin, bones, teeth and tendons. Ceratin is the major component of hair and nails.

1.1.2.2 The structure of proteins

Proteins are not just randomly coiled chains of amino acids. A variety of intramolecular interactions enables the amino acid chain to fold in a specific way to give the protein a three-dimensional structure and shape. This structure is critical for its activity and function. Several amino acid strings can be entangled and connected to each other via *disulfide bridges*. Parts of the amino acid chain can be organised into helices or sheets. Globular proteins like enzymes and antibodies are more folded and coiled whereas fibrous proteins are more filamentous and elongated. To describe the complex structure of proteins, four levels of organisation are distinguished: *primary*, *secondary*, *tertiary* and *quaternary* structures.

Primary structure

The sequence of amino acids determines the primary structure of a protein. Changing just a single amino acid in a critical position of the protein can significantly alter its activity and function and be the cause of disease and disorders. The amino acids are connected to each other in a head-to-tail fashion by formation of a *peptide bond* (Fig. 1.11), the condensation of a carboxylic and an amino group with the elimination of water.

Two amino acids connected via a peptide bond are called a *dipeptide*, three acids a *tripeptide* and so on. With an increasing number of acids in the sequence, the molecules are referred to as *oligopeptides* and *polypeptides*. The C—N bond cannot

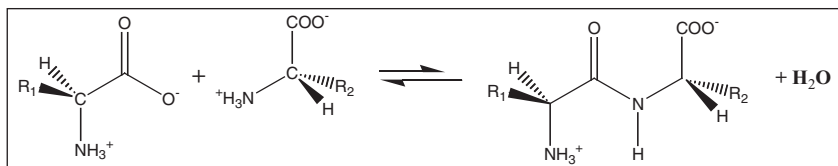


Fig. 1.11. Peptide bond formation from two amino acids.

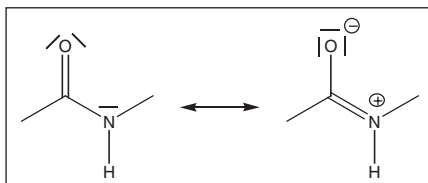


Fig. 1.12. Double bond character of the C–N bond in a peptide.

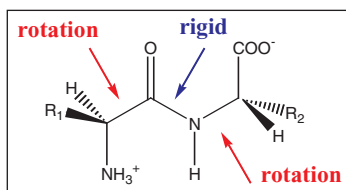


Fig. 1.13. The C–N bond is rigid due to the partial double bond character, rotation is possible within steric constraints around the bonds to the α C-atoms.

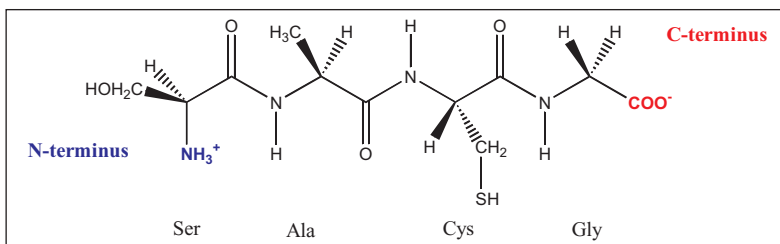


Fig. 1.14. A peptide with the amino acid sequence Ser-Ala-Cys-Gly showing N-terminus and C-terminus.

rotate due to its partial double bond character (Fig. 1.12). Hence, the peptide unit NH–CO is rigid. The bonds to the neighbouring alpha C-atoms can rotate within steric constraints (Fig. 1.13) and play an important part in folding of the protein. The peptide units together with the tetrahedral C-atoms form the *backbone* of a protein, while the R substituents are referred to as *side chains*.

An example of a peptide consisting of four amino acid residues (Ser-Ala-Cys-Gly) is given in Fig. 1.14. To be unambiguous about start and end of a sequence, the first amino acid residue is always the one with the free amino group, the *N-terminus*, and is written to the left. The last amino acid in the chain is the *C-terminus* with the free carboxyl group and is written to the right.

Peptides can also have a circular structure, i.e. they “bite their own tail”. An example of such a peptide is the potassium carrier, valinomycin.

With the 20 naturally occurring L-amino acids, it is possible to form an immense number of combinations and permutations. For a dipeptide there are already $20^2 = 400$ possible arrangements, for a tripeptide $20^3 = 8,000$. A relatively small protein with 100 amino acid residues can be arranged in $20^{100} = 1.27 \times 10^{130}$ different ways, an enormous number, especially when bearing in mind that there are “only” 10^{78} atoms in the whole universe. The bioanalytical chemist has to face a difficult task, if he wants to determine the exact sequence of amino acids in a protein. Nonetheless, their analysis has become commonplace and the methods involved are discussed in chapter 7.

Secondary structure

Secondary structures are regular elements such as α -*helices* and β -*pleated sheets*, which are formed between relatively small parts of the protein sequence. These structural domains are determined by the conformation of the peptide backbone, the influence of side-chains is not taken into account for secondary structures.

An α -*helix* (Fig. 1.15) is a right-handed coil, which is held together by hydrogen bonding between a $-\text{CO}$ group of an n^{th} amino acid residue in the sequence and the $-\text{NH}$ group of the $n + 4^{\text{th}}$ amino acid residue. The coiling is such that the $-\text{R}$ groups

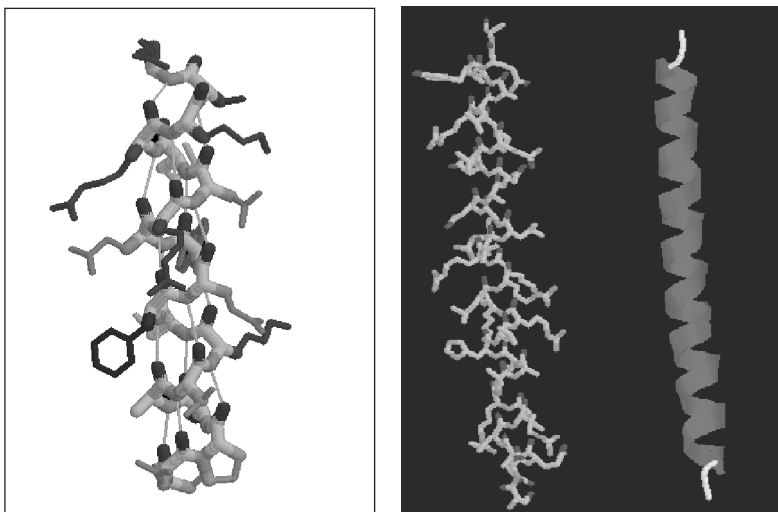


Fig. 1.15. Left: Structure of an α -helix with the $-\text{R}$ substituents pointing outwards. Right: Schematic drawing of an α -helix as commonly used in drawings of proteins.

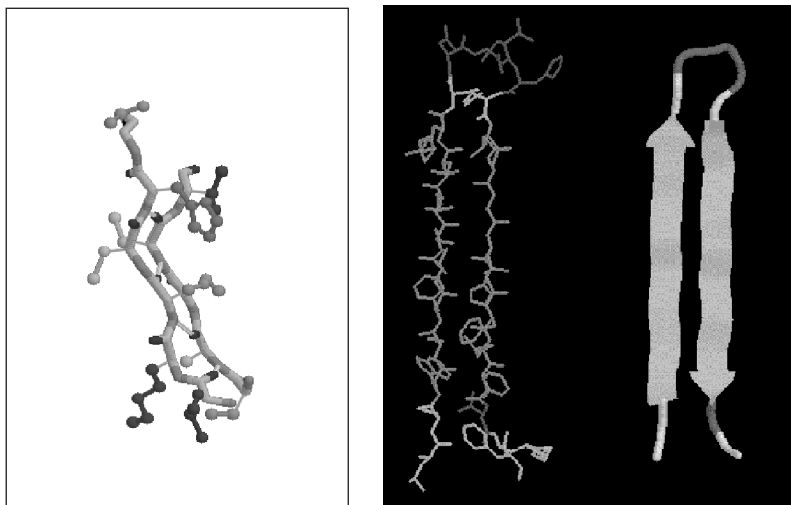


Fig. 1.16. Left: Structure of a β -pleated sheet with the $-R$ substituents pointing outwards. Right: Schematic drawing of a β -pleated sheet as commonly used in drawings of proteins.

are pointing outwards perpendicular to the axis of the coil. α -helices are important in structural proteins like keratin. Not all amino acids favour α -helix formation due to steric hindrance. The secondary amino acid proline, for example, is likely to disrupt the formation of a helix.

In a β -pleated sheet (Fig. 1.16), two polypeptide backbones are folded and aligned next to each other. They are connected via hydrogen bonds. The amino acid substituents R are pointing outwards to the top or bottom of the sheet. Adjacent chains can be aligned either in the same direction (parallel β -folding) or in opposite directions (antiparallel β -folding), as shown in Fig. 1.16. β -folding often occurs with amino acids carrying small non-charged side chains.

Tertiary structure

The tertiary structure describes the complete three-dimensional structure of the whole polypeptide chain. It includes the relationship of different domains formed by the protein's secondary structure and the interactions of the amino acid substituent $-R$ groups. An example of a protein chain with α -helices and β -folding, the enzyme ribonuclease, is shown in Fig. 1.17. The specific folding of a protein is only thermodynamically stable within a restricted range of environmental parameters, i.e. the right temperature, pH and ionic strength. Outside of this range, the protein could unfold and lose its activity.



Fig. 1.17. 3D-structure of ribonuclease H from *Escherichia coli* with α -helices and β -folding.

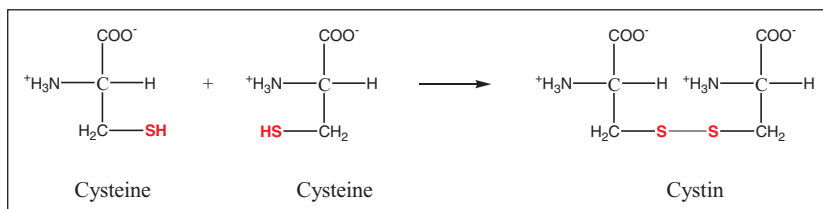


Fig. 1.18. Formation of a disulfide bridge.

Quaternary structure

A protein can consist of two or more separate polypeptide chains linked together. Other, non-amino acid components such as minerals, lipids and carbohydrates can also be part of a protein. The *quaternary structure* describes how these different chains and components interact and connect to each other by hydrogen bonding, electrostatic attraction and *sulfide bridges*. Such sulfide bridges are formed by *oxidation* of the $-\text{SH}$ groups of Cysteine (Fig. 1.18). The product of this reaction is a covalently bonded dipeptide called *Cystin*.

The hormone insulin, which is produced in the pancreas, contains two different polypeptide chains, *A* and *B*. Sulfide bridges can occur within a chain as well as between the two chains (Fig. 1.19).

The separate polypeptide chains forming a protein can be identical (*homogenic* protein) or, as in the case of insulin, different (*heterogenic* protein).

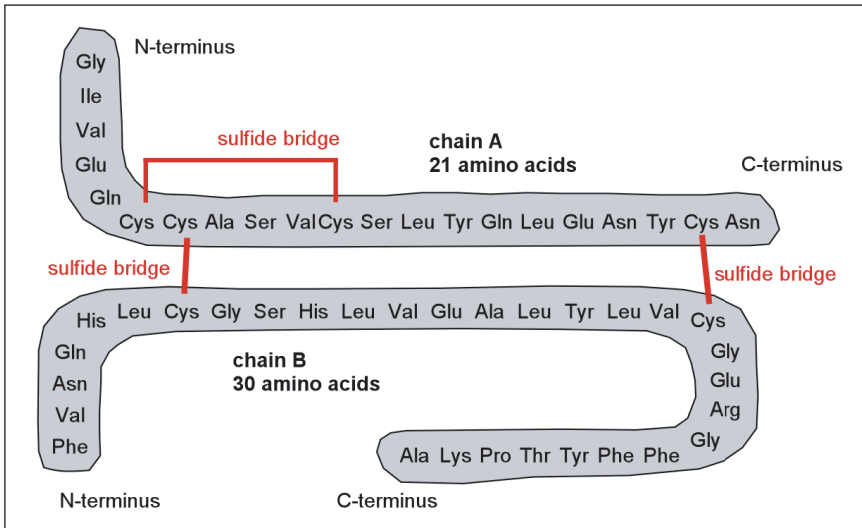


Fig. 1.19. Insulin has one sulfide bridge within chain A and two sulfide bridges between chain A and chain B.

1.1.2.3 Degradation of proteins

The three-dimensional structure of a protein which is held together by hydrogen bonding, electrostatic attraction and sulfide bridges is very sensitive to its chemical and physical environment. A change in pH, temperature or ionic strength disrupts these interactions and causes the protein to unfold; this process is called *denaturation*. The protein loses activity once its normal shape is lost. In some cases, this denaturation is reversible and the protein can *renature*, although in most cases the activity loss is permanent.

1.2 Nucleic Acids

Nucleic acids are long, linear biomolecules that can have molecular weights of several million Da. There are two classes of nucleic acids, *deoxyribonucleic acid (DNA)* and *ribonucleic acid (RNA)*.

DNA contains the “code of life.” It is the hereditary molecule in all cellular life forms as it is used by cells to store and transmit genetic information. During cell division, exact copies of DNA are made. Cells use DNA to determine and control the synthesis of proteins with the help of messenger RNA (mRNA).

RNA is essential for the synthesis of proteins in the cells. Messenger RNA (mRNA) is synthesised in the cell nucleus as a transcript of a specific part of DNA.