

Biochemistry is the study of all the chemical processes that take place within living organisms and is concerned primarily with what is happening on a molecular level with cells. Therefore, the interactions and reactions of the major macromolecules, proteins, DNA, carbohydrates and lipids are central to the study of biochemistry.

Proteins are important because they provide the molecular machinery of the cell; proteins act as enzymes, hormones and structural components and regulate the movement of other substances in and out of cells. In this chapter, you will learn about the structure of amino acids and how they react together to form proteins. You will also examine the interactions that control the folding of proteins into their three-dimensional structure and enable them to perform their functional role within the body.

Science understanding

- α -amino acids can be represented using a generalised structure
- the characteristic properties of α -amino acids include the formation of zwitterions and the ability to form amide (peptide) bonds through condensation reactions
- α -amino acids undergo condensation reactions to form polypeptides (proteins) in which the α -amino acid monomers are joined by peptide bonds
- the sequence of α -amino acids in a protein is called its primary structure
- secondary structures of proteins (α -helix and β -pleated sheets) result from hydrogen bonding between amide and carbonyl functional groups; hydrogen bonding between amide and carbonyl functional groups within a peptide chain leads to α -helix structures while hydrogen bonding between adjacent polypeptide chains leads to β -pleated sheets
- the tertiary structure of a protein (the overall three dimensional shape) is a result of folding due to interactions between the side chains of the α -amino acid in the polypeptide, including disulfide bridges, hydrogen bonding, dipole–dipole interactions, dispersion forces and ionic interactions

17.1 Amino acids



FIGURE 17.1.1 Spiders can produce silk using a strong protein called fibroin to form webs and protect their eggs.

i The prefix ‘mono’ means one, only or single. Therefore, the term ‘monomer’ means one unit. When monomers join other monomers, a polymer is formed (the prefix ‘poly’ means many).

In this section, you will investigate the structure and properties of amino acids. **Amino acids** are the **monomers** that make the thousands of different proteins that are essential to life.

Proteins are organic biopolymers that have many important functions in living things. These include:

- enzymes that catalyse specific biochemical reactions
- hormones that control biological processes
- structural components in cell membranes, muscles, hair, feathers and spider silk (Figure 17.1.1)
- transport of substances across cell membranes or around the body; for example, many proteins form channels in the cell membrane through which specific substances may travel (Figure 17.1.2)
- antibody molecules of the immune system.

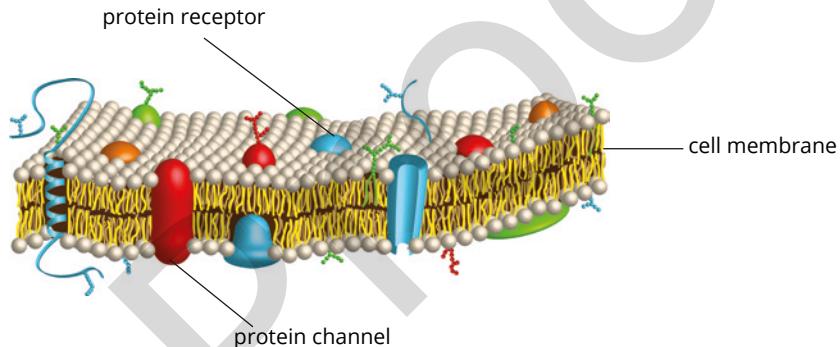


FIGURE 17.1.2 Many proteins are associated with the phospholipid membrane surrounding each cell. They act as channels, allowing other molecules to pass through. They also act as receptors in signalling pathways, carrying signals or messages into the cell from the external cell environment.

Plants can manufacture all of the amino acids required to make proteins. Plants synthesise amino acids from simple inorganic ingredients—such as carbon dioxide, water and nitrates—obtained from the atmosphere and soil. However, animals cannot do this. Although animals can synthesise some of the amino acids they require, they must obtain the other amino acids they need from the plant and animal proteins in the food they eat.

STRUCTURE OF AMINO ACIDS

There are 20 different amino acids that make up all of the proteins in the human body. Every amino acid has an **amino functional group** ($-\text{NH}_2$), a **carboxyl functional group** ($-\text{COOH}$), and a hydrogen atom attached to a central carbon, called the α -carbon (alpha-carbon).

Two of the simplest amino acids are glycine and alanine. The structures of these two molecules are shown in Figure 17.1.3.

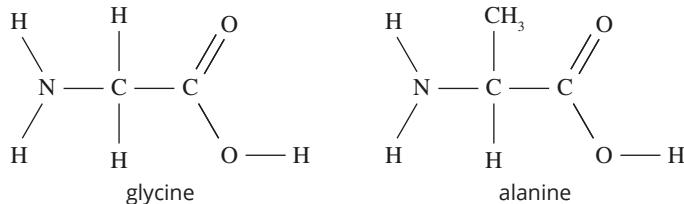


FIGURE 17.1.3 The structures of amino acids glycine and alanine

General formula of amino acids

The amino acids used to produce the proteins in the human body have the general formula $\text{H}_2\text{N}-\text{CH}(\text{R})-\text{COOH}$. The general structural formula of an amino acid is shown in Figure 17.1.4.

Amino acids such as those shown in Figure 17.1.4 are known as **α -amino acids**. This is because the amino group and the carbon atom of the carboxyl group are attached to the same carbon atom. This atom is known as the α -carbon. α -Amino acids are also sometimes called 2-amino acids.

The 20 α -amino acids used to make human proteins are listed in Table 17.1.1 on page 460, along with their three-letter abbreviations.

Properties of side chains in amino acids

The major difference between one amino acid and another is the group of atoms that make up the **side chain** (sometimes called R groups). The side chain may be:

- non-polar (e.g. $-\text{CH}_3$ in alanine and $-\text{CH}(\text{CH}_3)_2$ in valine)
- polar (e.g. $-\text{CH}_2\text{COOH}$ in aspartic acid and $-\text{CH}_2\text{OH}$ in serine).

The side chain may also include functional groups that can behave as:

- proton donors (e.g. $-\text{CH}_2\text{COOH}$ in aspartic acid contains an acidic carboxyl group)
- proton acceptors (e.g. $-\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2$ in lysine contains a basic amino group).

i Amino acids are commonly given three-letter abbreviations. For example, alanine is represented by Ala and glycine is represented by Gly.

ESSENTIAL AND NON-ESSENTIAL AMINO ACIDS

Within the cells in their bodies, humans can synthesise 11 of the 20 amino acids required to make proteins. These proteins sustain bodily functions. The other nine amino acids (isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine and histidine (in infants)) must be provided directly through the proteins you eat. This is because human cells do not have the capacity to produce them.

These nine amino acids are known as **essential amino acids**. A healthy diet should include proteins that contain these essential amino acids. The human body does not store amino acids, so a balanced intake of protein is required each day. Lack of amino acids in the diet can cause serious diseases, such as kwashiorkor, which you can see in Figure 17.1.5. Although this disease is characterised by low overall protein intake, it is thought that the lack of the sulfur-containing amino acids cysteine and methionine are particularly important in the disease development. ‘Kwashiorkor’ means ‘rejected one’ in the Ga language of coastal Ghana where the disease is common. It reflects the incidence of the disease in children who have been weaned from protein-rich breast milk to a diet that is high in carbohydrates and lacking in protein.

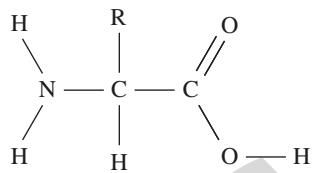


FIGURE 17.1.4 General structural formula of an amino acid. Different amino acids have different groups of atoms as the R group.



FIGURE 17.1.5 This child is suffering from kwashiorkor, which is a form of malnutrition caused by a protein-deficient diet.

TABLE 17.1.1 Amino acids found in the human body, with their three-letter symbols and structure

Name	Symbol	Structure
Alanine	Ala	$\begin{array}{c} \text{CH}_3 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Arginine	Arg	$\begin{array}{c} & \text{NH} \\ & \\ \text{CH}_2—\text{CH}_2—\text{CH}_2—\text{NH}—\text{C}—\text{NH}_2 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Asparagine	Asn	$\begin{array}{c} \text{O} \\ \\ \text{CH}_2—\text{C}—\text{NH}_2 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Aspartic acid	Asp	$\begin{array}{c} \text{CH}_2—\text{COOH} \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Cysteine	Cys	$\begin{array}{c} \text{CH}_2—\text{SH} \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Glutamine	Gln	$\begin{array}{c} \text{CH}_2—\text{CH}_2—\text{C}—\text{NH}_2 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Glutamic acid	Glu	$\begin{array}{c} \text{CH}_2—\text{CH}_2—\text{COOH} \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Glycine	Gly	$\text{H}_2\text{N}—\text{CH}_2—\text{COOH}$
Histidine	His	$\begin{array}{c} \text{CH}_2—\text{C}_6\text{H}_4—\text{NH}_2 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$
Isoleucine	Ile	$\begin{array}{c} \text{CH}_3—\text{CH}—\text{CH}_2—\text{CH}_3 \\ \\ \text{H}_2\text{N}—\text{CH}—\text{COOH} \end{array}$

Leucine	Leu	$\begin{array}{c} \text{CH}_3 — \text{CH} — \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Lysine	Lys	$\begin{array}{c} \text{CH}_2 — \text{CH}_2 — \text{CH}_2 — \text{CH}_2 — \text{NH}_2 \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Methionine	Met	$\begin{array}{c} \text{CH}_2 — \text{CH}_2 — \text{S} — \text{CH}_3 \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Phenylalanine	Phe	$\begin{array}{c} \text{CH}_2 — \text{C}_6\text{H}_5 \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Proline	Pro	$\begin{array}{c} \text{H} \\ \\ \text{N} \\ \\ \text{CH}_2 — \text{COOH} \end{array}$
Serine	Ser	$\begin{array}{c} \text{CH}_2 — \text{OH} \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Threonine	Thr	$\begin{array}{c} \text{CH}_3 — \text{CH} — \text{OH} \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Tryptophan	Trp	$\begin{array}{c} \text{H} \\ \\ \text{N} \\ \\ \text{CH}_2 — \text{C}_6\text{H}_4 — \text{CH} — \text{COOH} \end{array}$
Tyrosine	Tyr	$\begin{array}{c} \text{CH}_2 — \text{C}_6\text{H}_4 — \text{OH} \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$
Valine	Val	$\begin{array}{c} \text{CH}_3 — \text{CH} — \text{CH}_3 \\ \\ \text{H}_2\text{N} — \text{CH} — \text{COOH} \end{array}$

CHEMFILE

Essential amino acids—the work of William C. Rose

William Cumming Rose (1887–1985) was an American biochemist who devoted his professional life to the study of amino acids and their role in metabolism. He conducted many experiments on rats and discovered that rats were unable to survive on the 19 known amino acids as their sole nitrogen source. This led Rose to isolate and identify the final amino acid—threonine. This was a laborious task given the techniques available at the time.

Rose then turned his attention to human metabolism and tried to determine which amino acids are essential in our diets. Graduate students were paid \$1 a day to eat mixtures of corn starch, sucrose, butter fat, corn oil, vitamins and mixtures of purified amino acids as their only food source. The students' total nitrogen excretion was measured by analysing their faeces and urine, which they had to deliver to the laboratory each day. Rose observed that when particular amino acids were missing in their diets, subjects experienced dizziness, exhaustion and nervousness. In this way, he concluded that there were eight essential amino acids for humans. Histidine was later added to the list as it was found to be essential for infants.

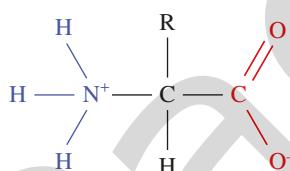


FIGURE 17.1.6 Structure of a zwitterion of an amino acid molecule

A balanced diet

Proteins from animal sources contain all the essential amino acids. As Table 17.1.2 shows, vegetable proteins are often deficient in one or more essential amino acids. For example, corn is too low in lysine and tryptophan to support normal growth in young children.

TABLE 17.1.2 Foods deficient in one or more of the nine essential amino acids

Food	Amino acid deficiency
Legumes	Methionine
Soyabean	Methionine
Corn	Lysine, tryptophan
Nuts	Methionine
Wheat	Lysine

For this reason, vegetarians must ensure that their diets contain complementary proteins. This means that instead of eating plant protein from a single source, vegetarians should eat a variety of plant foods. Deficiencies of amino acids in one plant can then be supplemented by the amino acids present in another. For example, rice and lentils contain complementary proteins. In Asia, rice and lentils have been part of the staple diet for thousands of years. In developed countries, people usually get all the amino acids they require by eating complementary proteins.

ACID-BASE PROPERTIES OF AMINO ACIDS

Amino acids as zwitterions

Amino acids contain polar amino and carboxyl functional groups. Therefore, amino acids can form hydrogen bonds with water molecules and are soluble in water. In solution, the:

- $-\text{NH}_2$ group can act as a base, accepting a proton to become a $-\text{NH}_3^+$ group
- $-\text{COOH}$ group can act as an acid, donating a proton to become a $-\text{COO}^-$ group.

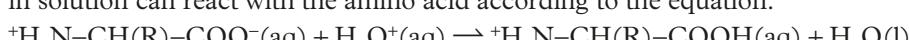
As a result, an amino acid molecule in an aqueous solution may be in the form $^+\text{H}_3\text{N}-\text{CH}(\text{R})-\text{COO}^-$. Such a molecule is shown in Figure 17.1.6 and is called a **zwitterion** or dipolar ion. A zwitterion is a molecule that contains positive and negative charges but has no charge overall.

The relatively high melting point of pure crystalline amino acids is due to the zwitterion being present in the solid state, meaning that the amino acids are held together by ionic bonds in the solid state. These ionic bonds require a large amount of energy to break, resulting in a high melting point.

Acid-base properties of amino acids

The dual acidic and basic nature of amino acids means that in a solution, different chemical forms of an amino acid can be in equilibrium. The predominant form depends on the pH of the solution and the particular amino acid concerned.

- At intermediate pH (typically pH 5–7), the zwitterion $^+\text{H}_3\text{N}-\text{CH}(\text{R})-\text{COO}^-$ is most abundant.
- At low pH, the cation $^+\text{H}_3\text{N}-\text{CH}(\text{R})-\text{COOH}$ is most abundant. The H_3O^+ ions in solution can react with the amino acid according to the equation:

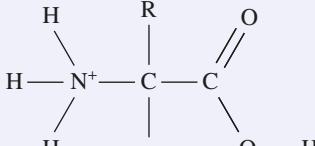
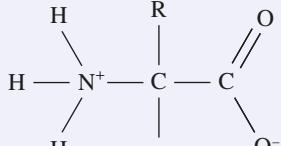
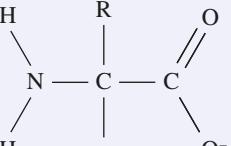


If the concentration of $\text{H}_3\text{O}^+(\text{aq})$ is very high (as it is in a solution of low pH), the position of this equilibrium lies well to the right.

- At high pH, the anion $\text{H}_2\text{N}-\text{CH}(\text{R})-\text{COO}^-$ is most abundant. The OH^- ions in solution can react with the amino acid according to the equation:

$$^+\text{H}_3\text{N}-\text{CH}(\text{R})-\text{COO}^-(\text{aq}) + \text{OH}^-(\text{aq}) \rightleftharpoons \text{H}_2\text{N}-\text{CH}(\text{R})-\text{COO}^-(\text{aq}) + \text{H}_2\text{O}(\text{l})$$
If the concentration of $\text{OH}^-(\text{aq})$ is very high (as it is in a solution of high pH), the position of this equilibrium lies well to the right.
As you can see in Table 17.1.3, the charge on the predominant form of the amino acid depends on the pH of the solution.

TABLE 17.1.3 Formation of cations, anions and zwitterions when amino acids are in solutions of various pH

Low pH	Intermediate pH	High pH
 cation	 zwitterion (no overall charge)	 anion

Effect of the side chain on the acid–base properties of amino acids

If the side chain contains a functional group with acid–base properties, it is possible for other charged forms of the amino acid to form. For example, at a pH of 12, the predominant form of aspartic acid is an ion with a charge of 2– (see Figure 17.1.7). On the other hand, at a pH of 2, lysine exists predominantly as an ion with a charge of 2+ (see Figure 17.1.8).

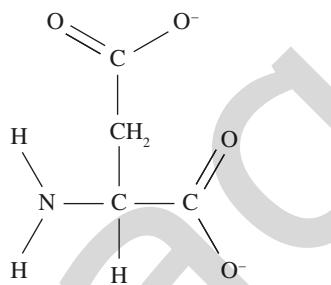


FIGURE 17.1.7 At pH 12, aspartic acid has a charge of 2–.

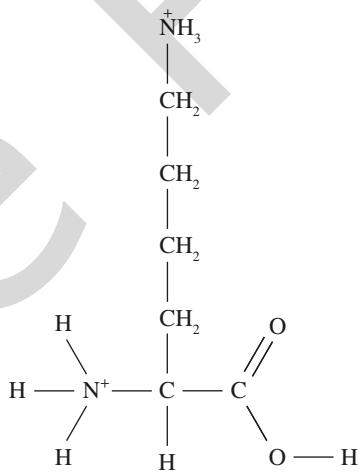


FIGURE 17.1.8 At pH 2, lysine has a charge of 2+.

17.1 Review

SUMMARY

- α -Amino acids have the structure shown in Figure 17.1.9.

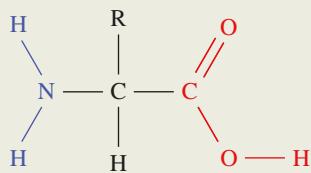


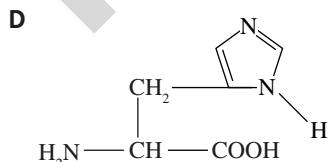
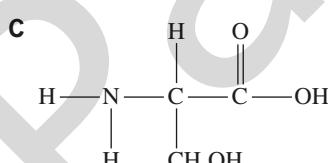
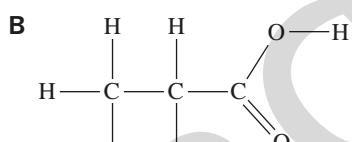
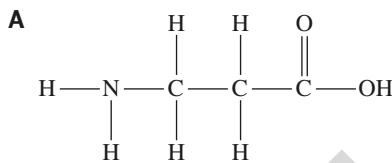
FIGURE 17.1.9 General structure of an α -amino acid.

- Amino acids contain a carboxyl group (highlighted in red in Figure 17.1.9), an amino group (highlighted in blue in Figure 17.1.9), a hydrogen atom and a side chain, all bonded to the same carbon atom.
- There are 20 different amino acids found in human proteins.
- Essential amino acids must be provided directly through the proteins in the diet, as human cells cannot produce them.

- Non-essential amino acids can be manufactured in the body from other dietary components.
- Three-letter abbreviations are commonly used to represent amino acids.
- The side chain may be a non-polar group of atoms or a group of atoms that includes a polar functional group. The side chain may also exhibit acid–base properties.
- Zwitterions are present in crystalline salts of amino acids and also in amino acid solutions of intermediate pH.
- An amino acid is present in cationic form in solutions of low pH and in anionic form in solutions of high pH.

KEY QUESTIONS

- 1 Which one of these molecules is not an α -amino acid?



- 2 Why should essential amino acids be included in a healthy diet?

- 3 Draw the structures of the following amino acids at the specified pH.

- a Serine at low pH
- b Alanine at high pH
- c Glutamic acid at high pH

- 4 Use your understanding of bond polarity and the acid–base properties of functional groups to complete the following table that categorises properties of an amino acid's R group.

Structure of R group	Is the R group polar or non-polar?	Is the R group acidic, basic or neutral?
$-\text{CH}(\text{CH}_3)_2$		
$-\text{CH}_2\text{COOH}$		
$-\text{CH}_2\text{C}_6\text{H}_5$		
$-(\text{CH}_2)_4\text{NH}_2$		

17.2 The formation of proteins

In this section, you will learn about the chemical reactions of amino acids that produce larger molecules called **polypeptides** and proteins. These reactions are examples of **condensation polymerisation** reactions.

The monomers that make up a polymer by a condensation polymerisation reaction have a pair of functional groups that can react together. When they react, a new functional group is produced that links the monomers. At the same time, a small molecule, such as water, is also produced. This type of polymerisation reaction is explored in detail in section 15.3.

Many biologically important molecules are produced by this type of reaction. These molecules include proteins, DNA, cellulose and starch. Insulin (see Figure 17.2.1) is a small protein that is made from 51 amino acids. It allows your body to use glucose from carbohydrates in your food, either for energy or for storing glucose for future use.



FIGURE 17.2.1 This young girl is injecting herself with insulin to control diabetes. Diabetes is a group of diseases associated with abnormally high levels of sugar in the blood.

Dipeptides

As you learnt in Chapter 15, a **condensation reaction** can occur between a molecule that contains a carboxyl group ($-COOH$) and a molecule that contains an amino group ($-NH_2$). An **amide** functional group ($-CONH-$) is formed that links the two molecules. A molecule of water is also produced as shown in Figure 17.2.2.

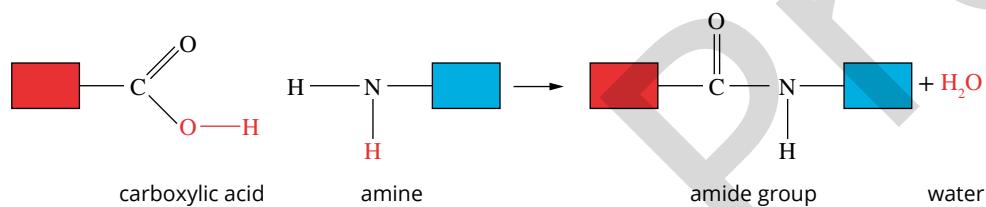


FIGURE 17.2.2 An amide group forms when a carboxyl group reacts with an amino group. A water molecule is also produced. The red and blue boxes represent the remaining atoms in the amino acids involved in this reaction.

Because α -amino acids contain both an amino functional group and a carboxyl functional group, they can undergo condensation reactions with each other. When two amino acids react, an amide group is formed that links the molecules together. This amide link is sometimes also called a **peptide link**. The bond between the carbon and the nitrogen atoms in the amide group is called a **peptide bond**.

Molecules made from amino acids are often called peptides. When two amino acid molecules react together as shown in Figure 17.2.3, the product is referred to as a **dipeptide**.

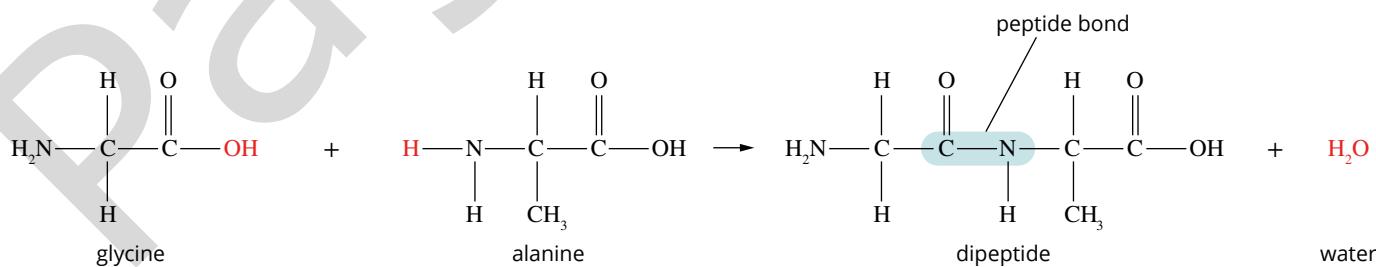


FIGURE 17.2.3 A condensation reaction between two amino acids, glycine and alanine to form the dipeptide Gly-Ala. Note how the carboxyl and amino groups react to form the dipeptide and water.

Each time a pair of different amino acids reacts in this way, there are two possible product molecules, depending on which ends of the two molecules react together (Figure 17.2.4).

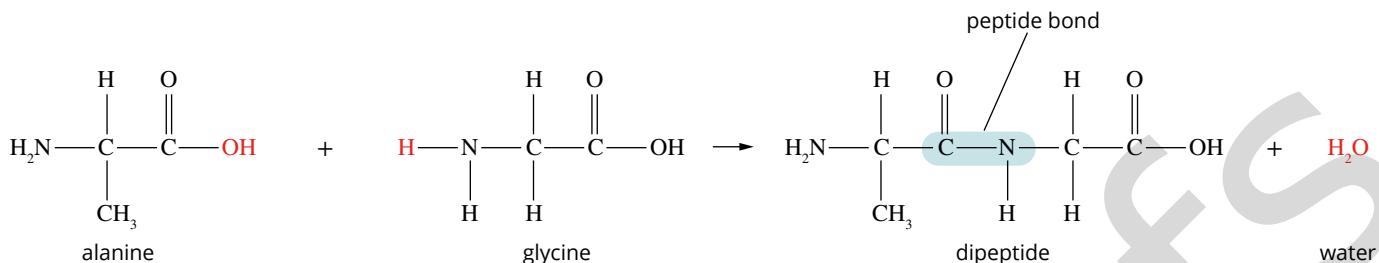


FIGURE 17.2.4 Another dipeptide molecule (Ala–Gly) formed from the reaction between glycine and alanine

POLYPEPTIDES

Polypeptides are polymers formed by condensation polymerisation of amino acids. During these reactions, the amino acids can form long chains. When three amino acid molecules react together, a **tripeptide** is formed. A polymer made from many amino acids is known as a polypeptide.

Naming polypeptides

A shorthand notation is often used to describe the amino acid sequence in a polypeptide, using the three-letter abbreviations listed in the table of α -amino acid structures (Table 17.1.1 on page 460). By convention, the structure is drawn so the free amino group is on the left and the free carboxyl group is on the right. When naming a polypeptide or protein, the order of the amino acids is also given in this direction, from the free amino group to the carboxyl group. A polypeptide consisting of six amino acids is shown in Figure 17.2.5. The polypeptide would be named as Ala–Glu–Gly–Cys–Val–Lys. Note that the alanine amino acid has a free amino group, and is called the **N-terminal** amino acid. The lysine has a free carboxyl group and is called the **C-terminal** amino acid.

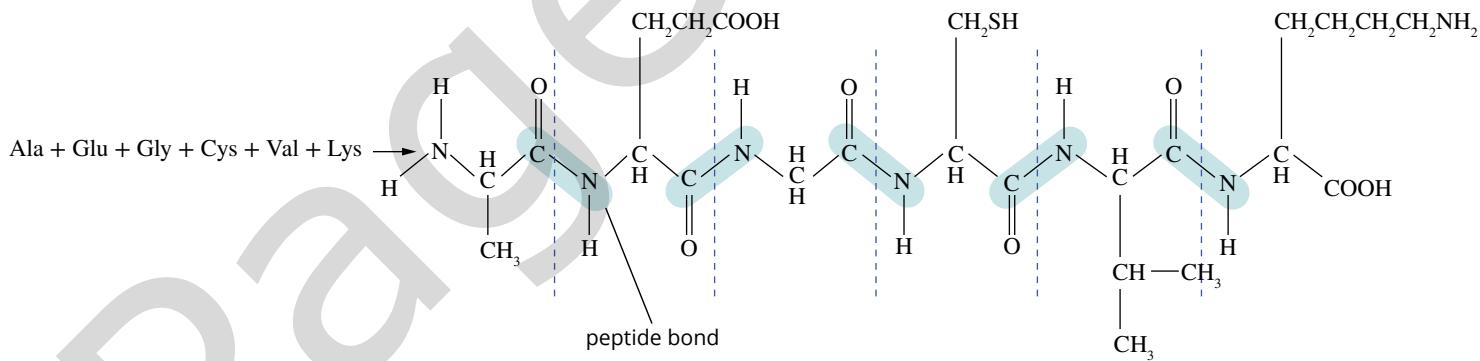


FIGURE 17.2.5 The polypeptide Ala–Glu–Gly–Cys–Val–Lys. Note that the amino acids are named in order from the N-terminal to the C-terminal. The peptide bonds between each amino acid are shaded.

A polypeptide constructed from more than 50 amino acids is usually called a protein. As longer molecules, proteins generally have a more complex three-dimensional structure than polypeptides. This will be discussed in sections 17.3 and 17.4.

The hormone insulin is a protein that regulates the metabolism of sugars and fats in the human body. Insulin is one of the smallest proteins in the human body. It is made up of two linked chains, containing a total of 51 amino acids. The sequence of amino acids in the two chains is shown in Figure 17.2.6. You can also see that the chains are linked by covalent bonds between sulfur atoms from the side chains of the cysteine units.

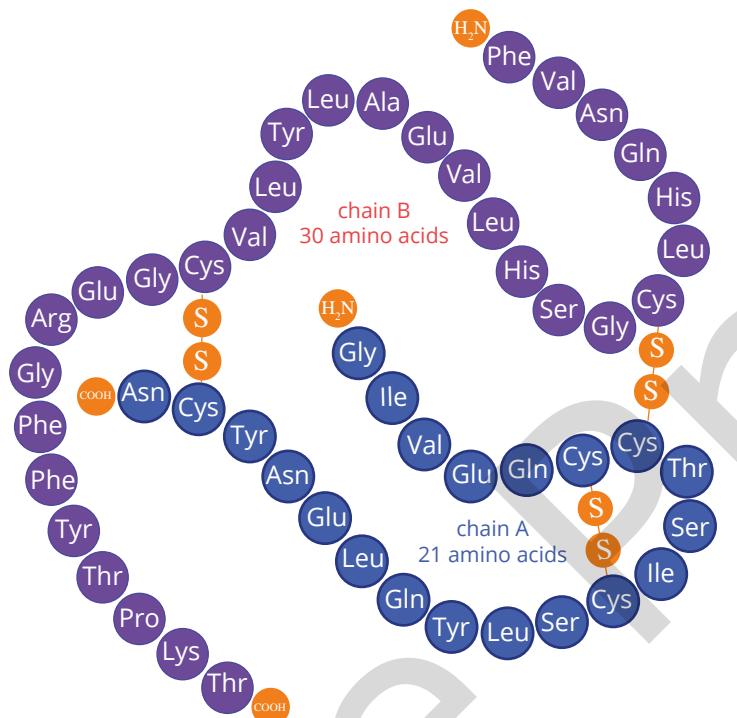


FIGURE 17.2.6 The amino acid sequence in the human insulin molecule. The polymer contains 51 amino acid units.

i When amino acids react with each other, peptide bonds form. Small molecules, such as water, are also released in this process, making this a condensation polymerisation reaction.

You have seen that when polypeptides are formed, water is produced as a by-product. When proteins are broken down in the body (through digestion), water molecules are required to break the peptide bonds. This type of reaction is known as **hydrolysis** and, in this case, hydrolysis can be regarded as the reverse reaction to the condensation reactions that formed the protein.

CHEMFILE

Fred Sanger—a pioneer of protein sequencing

Fred Sanger (Figure 17.2.7) was awarded two Nobel Prizes in Chemistry. The first was for his work on protein sequencing. Sanger determined the order of different amino acids in the two chains of the insulin protein. Sanger's second Nobel Prize was for his work on DNA sequencing.



FIGURE 17.2.7 Fred Sanger (1918–2013) did pioneering work on protein and DNA sequencing.

Born in England in 1918, Sanger grew up very interested in nature and science. Sanger decided not to follow in his father's footsteps and study medicine because he felt that a career in science would give him a better chance to become a problem solver.

Sanger completed his science degree at the University of Cambridge in England in 1939. He then studied amino acid metabolism for his PhD before starting work on identifying the amino acid sequence of (bovine) insulin extracted from cattle (Figure 17.2.8). Sanger was the first person to determine a protein sequence and was awarded his first Nobel Prize in 1958, in recognition of his important discovery.

By this stage, Sanger had moved on to study DNA-related problems, collaborating with other eminent scientists, including Francis Crick and John Kendrew. Solving the problem of DNA sequencing became a natural extension of his work in protein sequencing. In 1980, Sanger shared the Nobel Prize with Paul Berg and Walter Gilbert for his contribution to the determination of the nucleotide sequence in DNA.

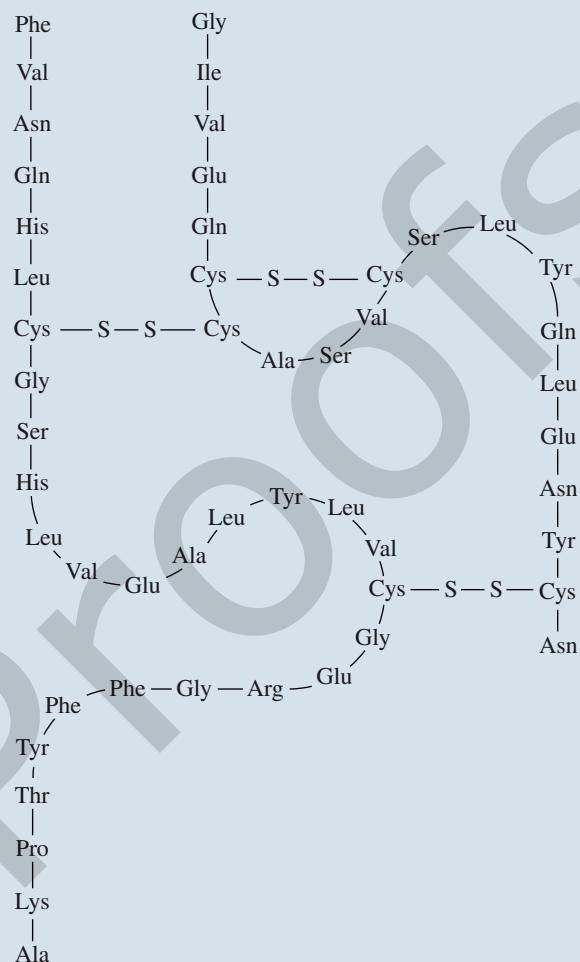


FIGURE 17.2.8 The amino acid sequence of bovine insulin. Its sequence of amino acids differs from human insulin by just three amino acids.

17.2 Review

SUMMARY

- The carboxyl group and amino group of two α -amino acids can take part in a condensation reaction that links them through a peptide bond.
- Water is also produced as a by-product, making this reaction an example of a condensation polymerisation reaction (see Figure 17.2.9).
- In a similar manner, several amino acids combine to produce a polypeptide.
- Proteins are large polypeptide chains containing approximately 50 or more amino acids.

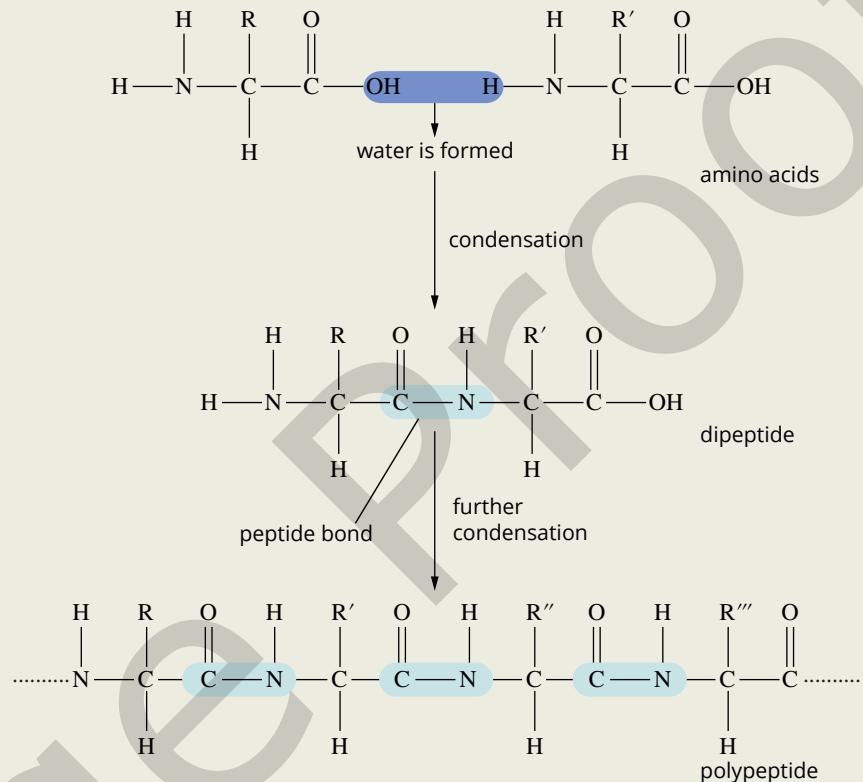


FIGURE 17.2.9 Condensation polymerisation of amino acids.

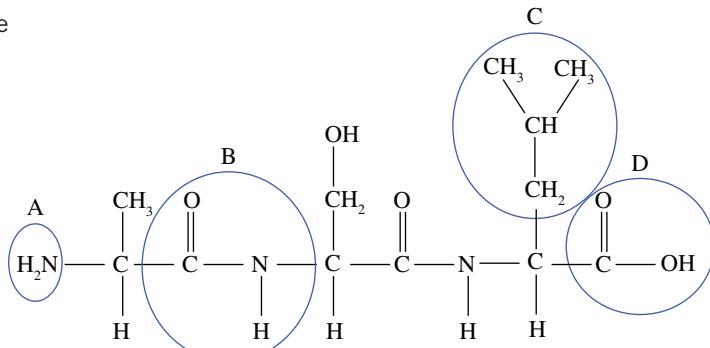
KEY QUESTIONS

- 1 Match the terms on the left with their descriptions on the right.

Tripeptide	By-product of the reaction that produces polypeptides
Side chain	Three amino acids joined by peptide links in a polypeptide chain
Polypeptide	Variable part of an amino acid
Amino group	Several amino acids that are joined by peptide links
Water	Functional group present in all amino acids

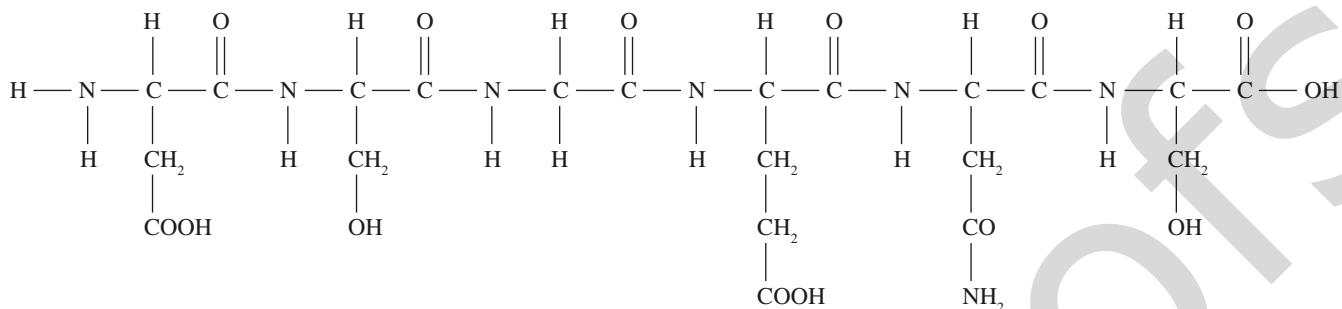
- 2 A tripeptide is shown below.

- a Identify the parts labelled A–D.
b Give the names of the three amino acids in the tripeptide molecule.



17.2 Review *continued*

- 3 Which one of the following statements about the polypeptide shown below is correct?



- A This polypeptide contains only two carboxyl groups.
B Six water molecules were produced in this reaction.
C Six different amino acids were used to make this polypeptide.
D This polypeptide was produced by an additional polymerisation reaction.
- 4 a Draw structural formulas of serine and cysteine.
b Write an equation to show the formation of a dipeptide from these amino acids.
c Write the formula of another dipeptide that could be formed from these two amino acids.
d Name the type of reaction in part b.

17.3 Primary and secondary structures of proteins

Proteins differ from one another in the number, type and sequence of their constituent amino acids. Each protein has a precise chemical composition and amino acid sequence, which leads to it having a unique three-dimensional shape. There may be more than 500 amino acid units in a large protein. Determining the structure of these complex materials has proved challenging for chemists. Myoglobin (see Figure 17.3.1) was one of the first proteins to have its amino acid structure determined. It is made up of 153 amino acids.

The role of any protein in an organism depends on the protein's shape. The structure of a protein is usually considered at several levels of organisation within the molecule. All proteins are regarded as being made up of a primary, secondary and tertiary structure, while some of them also display a quaternary level of organisation.

In this section, you will investigate the primary and secondary structures of proteins.

PRIMARY STRUCTURE OF PROTEINS

The number, type and sequence of the amino acid units in a protein is known as the protein's **primary structure**. As you can see in Figure 17.3.2, the primary structure may be represented by the three-letter abbreviations for the amino acids or by a structural formula of the protein. By convention, the sequence is written from left to right, starting from the N-terminal amino acid and ending with the C-terminal amino acid.

- i** The function of a protein is a consequence of its shape, which in turn is determined by the order in which amino acids are joined together, the primary structure.



FIGURE 17.3.1 A representation of myoglobin. Myoglobin is a protein consisting of 153 amino acids in one polypeptide chain. This protein is found in muscle tissue where it binds to oxygen.

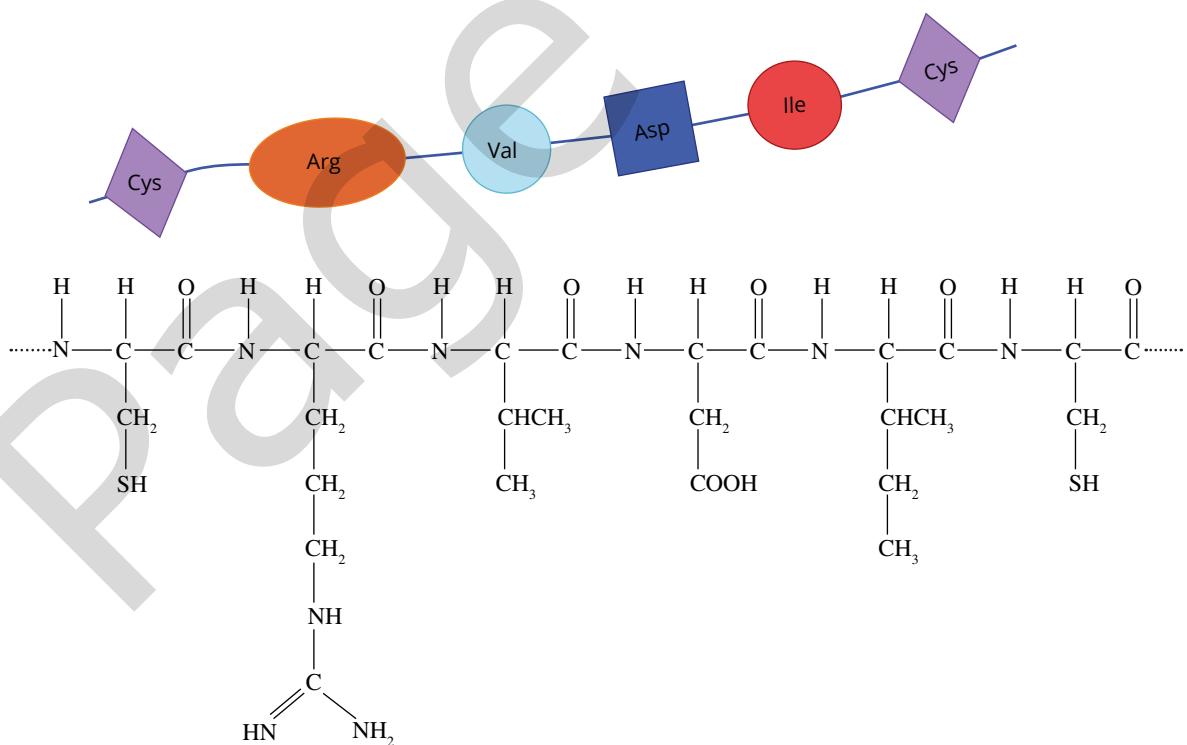


FIGURE 17.3.2 A small section of a protein molecule showing the primary structure represented by three-letter abbreviations and its structural formula

The three-dimensional shape of a protein is determined by the precise order in which its amino acids are joined together. This is due to the many interactions between nearby amino acids, leading to the folding of the protein chain. One of the smallest proteins in the human body, insulin, has a specific sequence of 51 amino acids. Many proteins contain hundreds of amino acids in a specific sequence.

The importance of the primary structure to the function of a protein can be seen in genetic conditions such as cystic fibrosis. The protein affected by cystic fibrosis functions in healthy individuals allows chloride ions to move across cell membranes. It is a large protein, consisting of 1480 amino acids. The disease is caused by genetic mutations that involve either deleting or swapping just one amino acid in the sequence. This tiny change in the primary structure affects the overall three-dimensional shape of the protein and prevents it from working normally, causing mucus to build up in the lungs and pancreas.

CHEMISTRY IN ACTION

Protein sequencing

The primary structure of a protein is found experimentally by the process of protein sequencing (Figure 17.3.3).

The amino acids in a polypeptide can be identified by heating the polypeptide in concentrated (6 M) HCl. The polypeptide undergoes hydrolysis, forming the individual amino acid monomers. The amino acid mixture is separated by column chromatography, and then each amino acid is identified by its retention time. However, this technique does not indicate the sequence of amino acid units in a polypeptide.

One method of determining the sequence of a polypeptide begins with using different enzymes to hydrolyse specific peptide bonds within the molecule. For example, trypsin is an enzyme that hydrolyses the peptide link at the carboxyl side of arginine and lysine. This process is repeated with different enzymes to produce shorter peptides of varying lengths. These peptides can then be sequenced through a process called Edman degradation or by mass spectrometry. Careful analysis of how the sequences of the peptides overlap allows the overall sequence of the polypeptide to be determined.



FIGURE 17.3.3 This scientist is preparing to use an automated microsequencer machine to analyse the amino acid sequence of samples of protein.

SECONDARY STRUCTURE OF PROTEINS

Coiling and pleating of sections of a protein molecule produce a secondary level of structure in a protein. Hydrogen bonds between the polar $-\text{NH}$ group in one peptide link and the polar $-\text{C=O}$ group in another peptide link can form at regular intervals (Figure 17.3.4).

This creates regions in which the molecule coils into a spiral shape called an **α -helix**, or where sections line up parallel to each other, forming a **β** -pleated sheet.

Such highly ordered segments, stabilised by hydrogen bonds, are referred to as the **secondary structure** of the protein.

α -Helices

Keratin is a protein found in the fibres of hair and wool. The helical structure of keratin results from extensive hydrogen bonding between peptide links in the polypeptide chain. The hydrogen bonds result from the attraction between the partial positive charge on the H of a $-\text{NH}$ group in a peptide link with the partial negative charge on the O of a $-\text{C=O}$ group of a peptide link four amino acid units along the chain.

The hydrogen bonds make the molecule coil into the shape of an α -helix, the same shape as a spring (see Figure 17.3.5).

Some amino acids are more likely to be found in an α -helix than others. Both proline and glycine, for example, are unlikely to exist in an α -helix structure. Due to its ring structure, proline's structure is too rigid to allow it to form part of a helix. Conversely, the fact that glycine only has a hydrogen atom for its R group means that it introduces too much flexibility and reduces the stability of the α -helix structure.

β -Pleated sheets

Hydrogen bonds can also form between peptide links to produce regions where two or more parts of the polypeptide chains line up parallel to each other. The repeating structure of the backbone of the protein chain ($-\text{N}-\text{C}-\text{C}-\text{N}-\text{C}-\text{C}-\text{N}-\text{C}-\text{C}-$) allows these hydrogen bonds to form at regular intervals. This stabilises the protein structure. This type of secondary structure in the protein is called a β -pleated sheet (Figure 17.3.6).

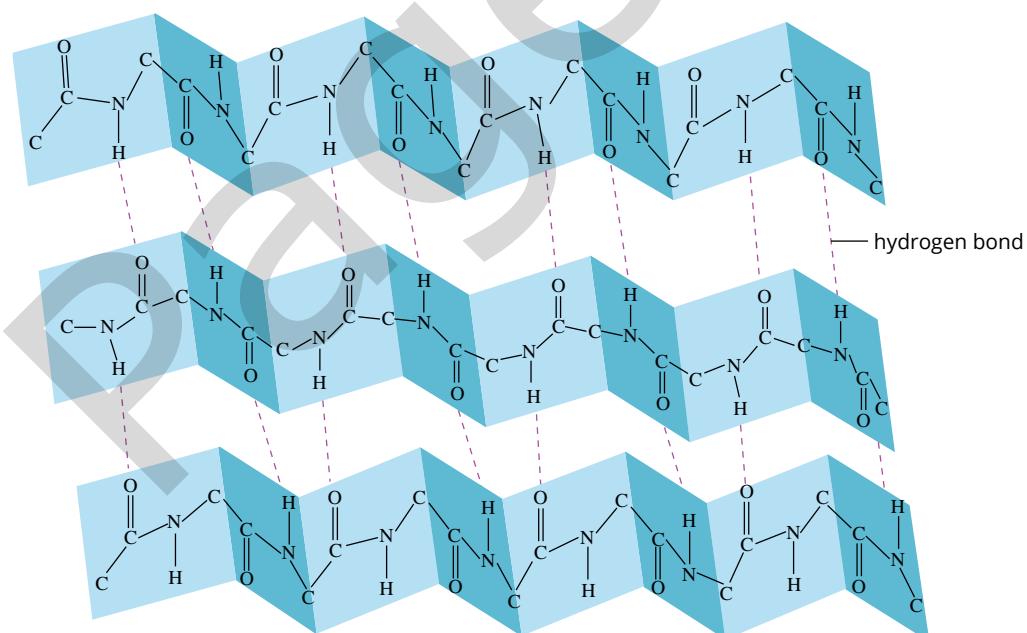


FIGURE 17.3.6 β -Pleated sheet. Hydrogen bonds between the sheet are shown as dashed lines. The side chains (R groups) and hydrogen atoms on the α carbons have been omitted for clarity.

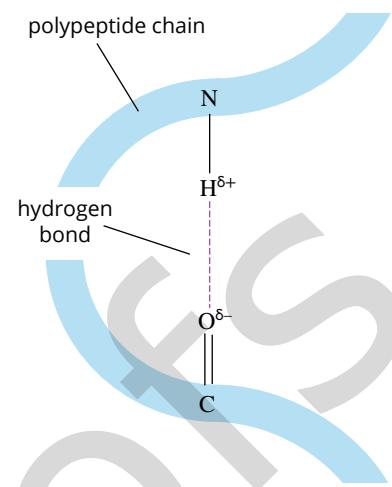


FIGURE 17.3.4 Hydrogen bonds can form between the polar $-\text{NH}$ group in one peptide link and the polar $-\text{C=O}$ group in another peptide link further along the polypeptide chain.

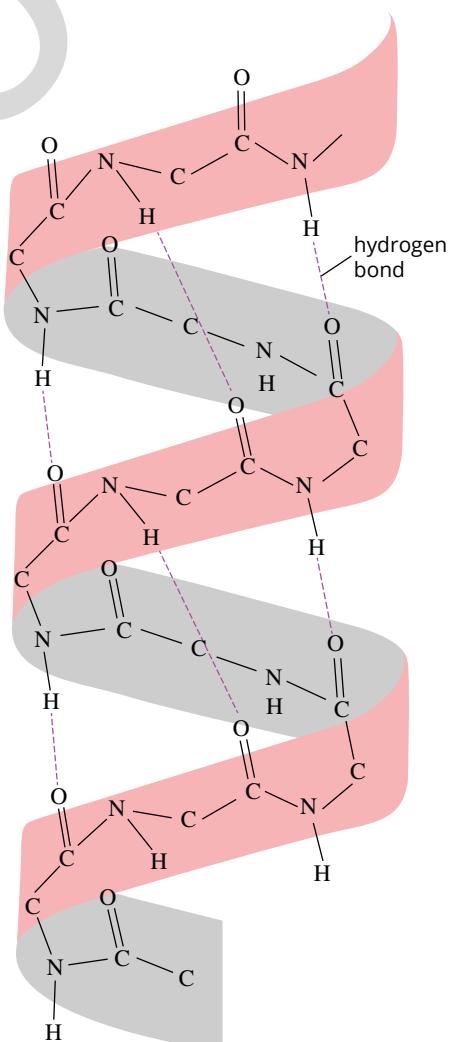


FIGURE 17.3.5 A polypeptide chain coils in an α -helix due to hydrogen bonds. The side chains (R groups) and hydrogen atoms on the α carbons have been omitted for clarity.

Silk is a protein with a β -pleated sheet structure. The polypeptide chains involved mainly contain the amino acids glycine, alanine and serine as shown in Figure 17.3.7. Notice that every second R group is H. With only these small side groups, sections of the protein molecule in silk can line up closely, enabling strong hydrogen bonds to form between these adjacent sections and producing β -pleated sheets.

- i** The secondary structure of a protein results from hydrogen bonding between the C=O and N–H groups in the peptide links along the protein backbone. This leads to the formation of α -helices or β -pleated sheets.

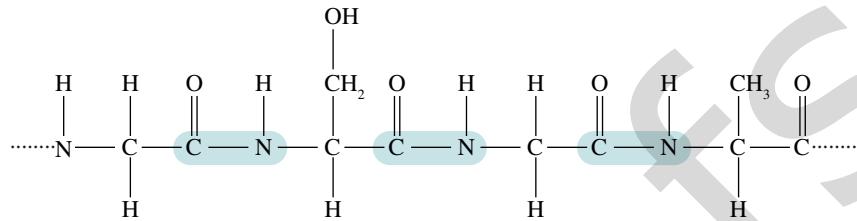


FIGURE 17.3.7 The amino acid sequence that is found in large parts of the β -pleated sheets in the silk protein. Peptide bonds are shaded blue.

CHEMFILE

Modelling protein structure

In computer models of protein structure, a segment twisted into an α -helix is often represented as a twisted ribbon and sections of β -pleated sheets are represented by a set of wide parallel ribbons with arrows.

A computer-generated model of a protein is shown in Figure 17.3.8.

There are wide variations in the relative amounts of α -helix and β -sheet structures in individual proteins, with many proteins containing several regions of both.

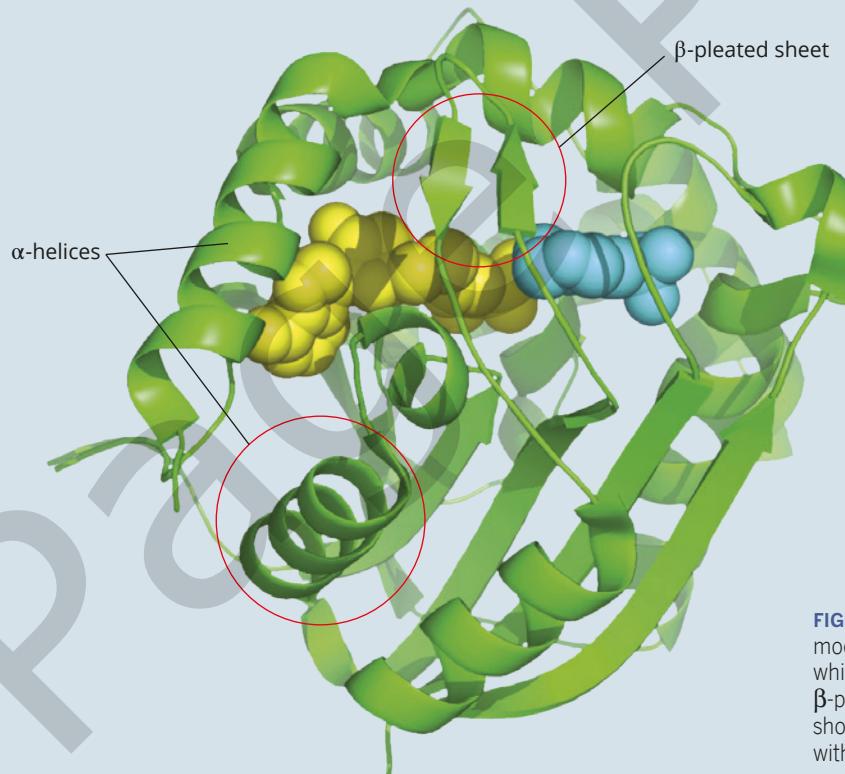


FIGURE 17.3.8 Computer-generated model of a protein (green and yellow) in which several regions of α -helices and β -pleated sheets are seen. The blue part shows drug-like molecules interacting with the protein.

17.3 Review

SUMMARY

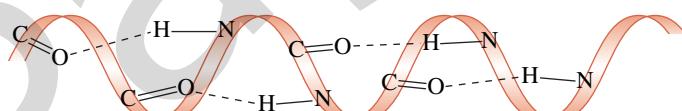
- There are 20 different α -amino acids from which tens of thousands of different protein molecules are synthesised by cells. Each protein has a unique structure and function in the organism.
- The complex structure of these large molecules is often considered in distinct levels. The primary and secondary levels are summarised in Table 17.3.1.

TABLE 17.3.1 Primary and secondary structures of proteins

Structural level	Description	Bonds involved
Primary	The sequence of amino acids in the polypeptide chain	Only covalent bonds are responsible for joining the monomer units together in the polymer. Monomers are linked by $-\text{C}-\text{N}$ bonds, known as peptide bonds in proteins.
Secondary	The shape of the polypeptide as it is either twisted into an α -helix or bent back on itself to produce regions of β -pleated sheets	Hydrogen bonding between $-\text{NH}$ and $-\text{C}=\text{O}$ bonds in peptide links introduces a secondary structure to protein molecules, producing regions of α -helices or β -pleated sheets.

KEY QUESTIONS

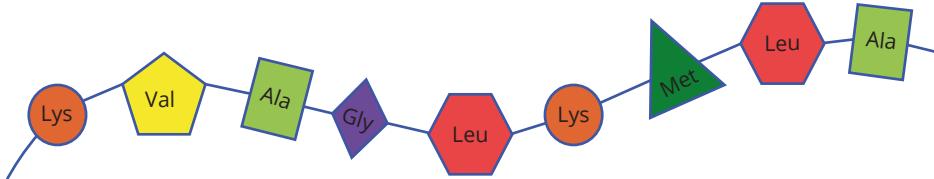
- 1 The following diagram shows a section of a protein.



- a What secondary structure is represented by the diagram?
 b How does this type of secondary structure form? Refer to particular atoms and functional groups associated with the bonding that forms this structure.

17.3 Review *continued*

- 2 Why is it important to be able to determine the specific sequence of amino acids in a protein?
- 3 The following diagram represents a section of a polypeptide molecule. Complete the paragraph by filling in the blanks. You may need to refer to the Table 17.1.1 on pages 460 to answer this question.



The image shows a section of the primary structure of part of a polypeptide. It consists of _____ amino acid units that are linked together by _____ bonds. It was produced by a _____ reaction. In this polypeptide there are two units of leucine, two units of _____, two units of _____ and one unit each of _____, _____, _____.

- 4 How do α -helices and β -pleated sheets differ?

Page Proofs

17.4 Tertiary structure of proteins

In the previous section, you learnt that the sequence and order of amino acid units in a protein chain is called the primary structure of the protein. As a consequence of hydrogen bonding between $-C=O$ and $-N-H$ groups along the chain, the protein chain can be twisted and folded into arrangements such as α -helices and β -pleated sheets, which are described as the secondary structure of the protein.

In this section, you will examine the overall three-dimensional shapes of proteins and the types of bonds that cause proteins to adopt their different shapes. These bonds are due to forces between the different functional groups in the side chain of amino acid units.

OVERALL SHAPE OF A PROTEIN

The overall three-dimensional shape adopted by a protein molecule is called its **tertiary structure**. A tertiary structure is produced by the three-dimensional folding of its secondary structures (α -helices and β -pleated sheets). The protein can twist back over itself to create a unique shape, which is responsible for the protein's function (see Figure 17.4.1).

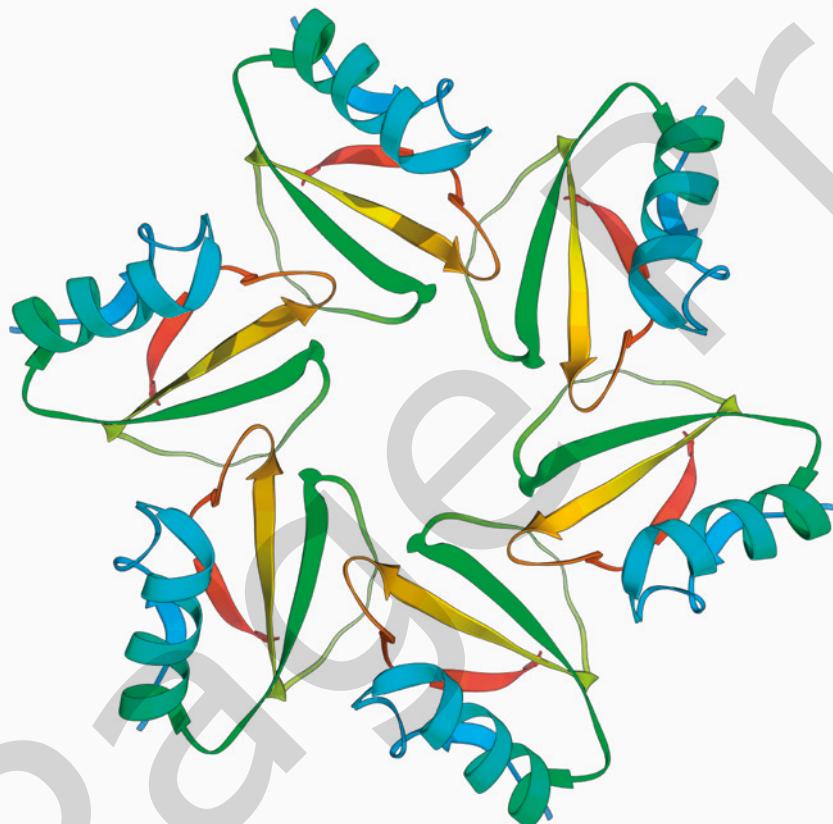


FIGURE 17.4.1 A model representing the tertiary structure of the protein CI2 (Chymotrypsin inhibitor 2) showing how the α -helices and β -pleated sheets are folded to produce a three-dimensional structure

The side chains of the amino acid units making up the polypeptide chain influence the overall three-dimensional shape of the molecule. These side chains can interact with each other in a number of ways, causing the protein to fold into its three-dimensional shape. Side chains can be relatively large (such as in phenylalanine), contain polar functional groups, or become charged depending on the pH of their surroundings. In addition, some amino acids have **hydrophobic** (non-polar) chains, which tend to fold towards the interior of protein molecules, away from contact with water molecules.

i Side chains of amino acids influence the properties of the amino acid. There are four main groups of amino acids, which can be classified as polar, non-polar, acidic or basic. These were discussed in section 17.1 (page XXX).

Bonding in the tertiary structure of proteins

Five types of attractions that are important in protein folding are summarised in Table 17.4.1, which also identifies the features of the side chains that are involved.

TABLE 17.4.1 Types of bonds formed from interactions of specific side chains in different regions of the protein

Bond type	Required components in side chain	Visual representation
Hydrogen bonds	Contains $-O-H$, $-N-H$ or $-C=O$	<p>A hydrogen bond linking two parts of a polypeptide chain</p>
Dipole-dipole interactions	Any polar group such as those containing $-S-H$, $-O-H$ or $-N-H$	<p>A dipole-dipole interaction linking two parts of a polypeptide chain</p>

Bond type	Required components in side chain	Visual representation
Ionic interactions	Contains NH_3^+ and another group that contains COO^-	$\cdots \text{--- NH---CH---C=O ---\cdots}$ $ $ CH_2 $ $ CH_2 $ $ COO^- $ $ NH_3^+ $ $ $(\text{CH}_2)_4$ $ $ $\text{C=O --- CH --- NH ---\cdots}$ <p>An ionic interaction linking two parts of a polypeptide chain</p>
Covalent cross-links	Cysteine side groups react to form a disulfide bridge (---S---S---)	$\cdots \text{--- NH---CH---C=O ---\cdots}$ $ $ CH_2 $ $ S $ $ S $ $ CH_2 $ $ $\cdots \text{--- C---CH---NH ---\cdots}$ $ $ O <p>A disulfide bridge linking two parts of a polypeptide chain</p>
Dispersion forces	Any non-polar group	$\cdots \text{--- NH---CH---C=O ---\cdots}$ $ $ CH $ $ $\text{H}_3\text{C} \quad \text{CH}_2$ $ $ CH_2 $ $ CH_2 $ $ CH_2 $ $ CH_2 $ $ $\text{C=O --- CH --- NH ---\cdots}$ $ $ O <p>Dispersion forces can link two parts of a polypeptide chain</p>

CHEMISTRY IN ACTION

X-ray crystallography

Scientists use a technique called X-ray crystallography to determine the three-dimensional structures of large molecules such as proteins. The first step in this process is producing large enough quantities of the protein to enable it to be purified. This is usually achieved using cells that have been genetically modified to produce large quantities of the target protein.

A single protein molecule is not large enough to be analysed by this technique, so scientists produce protein crystals, in which the molecules are arranged in a regular pattern, in order to magnify the signal. This can be a difficult technique and requires a long experimental process to determine the conditions (temperature, pH, salt concentration, time) under which the protein will form suitable crystals.

The crystal is then placed in front of an X-ray source and the pattern of X-rays that passes through is recorded (Figure 17.4.2). The pattern formed and the relative brightness of the dots is determined by the arrangement of the electrons in the crystal. Computer programs then use this map of electron density to determine the position of each atom in three-dimensional space; that is, the tertiary structure of the protein.

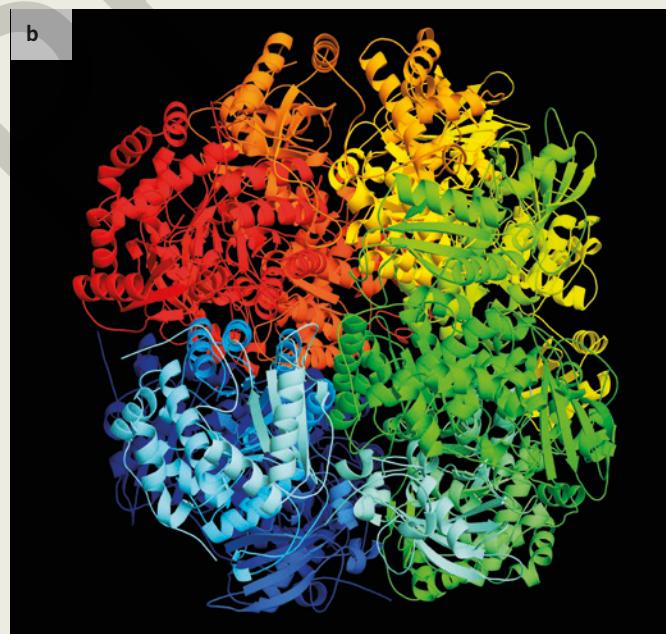
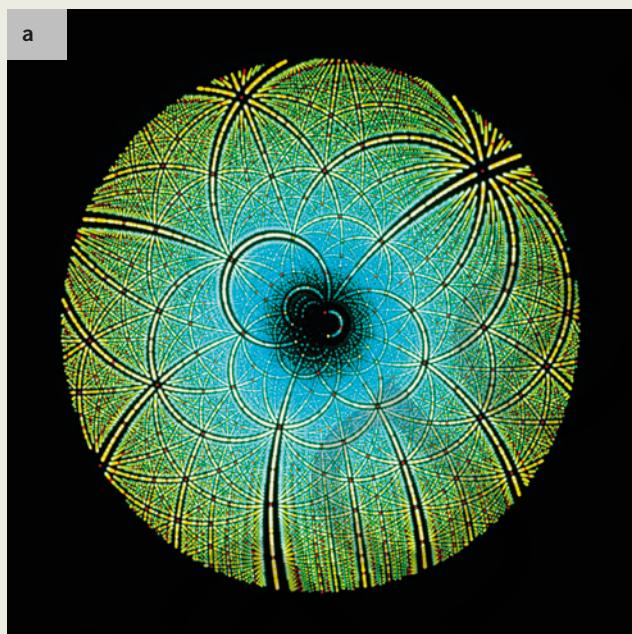


FIGURE 17.4.2 Rubisco is the most common protein in nature. Found in all plants, rubisco fixes carbon dioxide during photosynthesis. (a) The X-ray diffraction pattern produced when rubisco crystals are bombarded by X-rays. (b) The three-dimensional structure of rubisco, determined from the X-ray diffraction pattern.

An enormous variety of protein shapes exist as a result of the different types of bonds that can fold a polypeptide into a three-dimensional shape. Some proteins resemble flat sheets, others are long and helical, and others are compact and globular.

CHEMFILE

Alzheimer's disease

The three-dimensional shape of a protein is critical to its function. Several diseases are linked to proteins that have not folded properly into their three-dimensional shape; these are called protein misfolding diseases and include Alzheimer's disease. Although the causes of Alzheimer's disease are not well understood, the disease is characterised by the accumulation of dense plaques in the brains of sufferers (Figure 17.4.3). These plaques eventually lead to the degeneration of affected neurons.

These plaques consist of aggregations of the short protein amyloid beta. In healthy people, this protein is soluble. However, in Alzheimer's patients, the protein changes shape and forms a crossed structure rich in β -sheets. Many similarly misfolded amyloid beta protein chains then combine to form long fibrils. In this form, the amyloid beta protein is insoluble and forms the plaques on neurons.

Many other proteins can form amyloid fibrils when they misfold. These proteins are generally also rich in β -sheets and have been linked to diseases such as type 2 diabetes, Huntington's disease, Parkinson's disease, atherosclerosis and bovine spongiform encephalopathy (mad cow disease). When this process occurs in the heart (cardiac amyloidosis), the heart tissue becomes stiff and thickened. This is often a fatal condition, commonly known as 'stiff heart syndrome'. It is not known what causes these proteins to misfold.

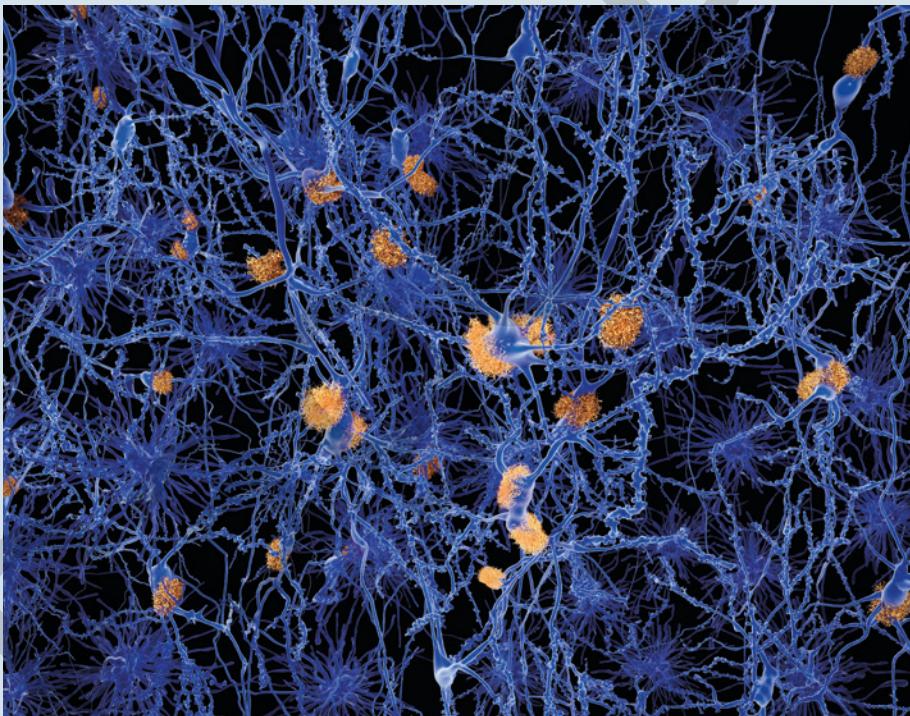


FIGURE 17.4.3 An illustration showing the formation of amyloid plaques (orange) among neurons (blue) in the brain of an Alzheimer's patient. The plaques lead to degeneration of affected neurons.

EXTENSION

Quaternary structure

Some proteins are composed of two or more polypeptide chains, and may even interact with non-protein molecules to produce a larger, more complex functional unit, known as the **quaternary structure**. Haemoglobin is an example of a protein with a quaternary structure.

Red blood cells are manufactured by the bone marrow at a rate of about 2 million per second. Red blood cells are functional for about 3 months in the human body. Each red blood cell contains approximately 250 million molecules of haemoglobin.

Figure 17.4.4 is a diagrammatic representation of haemoglobin. The four distinct subunits that make up its quaternary structure are highlighted in colour. Each subunit contains a polypeptide chain (collectively called globin) that is coiled into α -helices and then folded into a tertiary structure. Within each subunit, there is also an oxygen-binding site, or haem group (light blue). Each haem group contains one atom of iron (green). Haemoglobin transports oxygen around the body by binding one oxygen molecule to each haem group.

The iron of the haem group can bond to one oxygen molecule, so each haemoglobin molecule is able to carry four oxygen molecules. In this form, the molecule causes the characteristic bright red colour of oxygenated blood and it is called oxyhaemoglobin.

Proteins that function in water, such as haemoglobin, tend to fold so that their hydrophobic non-polar side chains are least exposed to water molecules. The polar side chains are hydrophilic and are therefore located near the outer surface.

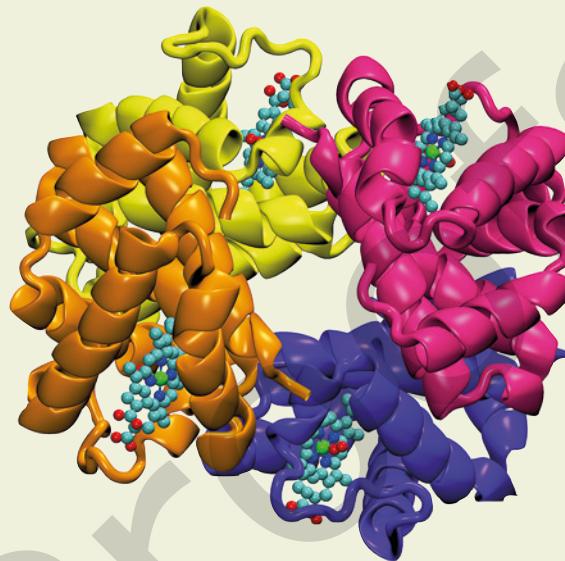


FIGURE 17.4.4 A representation of the structure of haemoglobin. Within each subunit is a haem group (light blue). Each group contains one atom of iron (green). In this model, the haem groups are oxygenated and the oxygen molecules are shown as paired red spheres.

Two or more separate polypeptide chains can assemble together so that hydrophobic sections in the chains are in contact with each other, which minimises the contact of these sections of the chains with water. Haemoglobin has four chains that join together in this way. Dispersion forces are the major force of attraction holding the separate polypeptide chains together. However, hydrogen bonding, dipole–dipole attractions and ionic interactions may also be present between the chains.

17.4 Review

SUMMARY

- The complex structure of large molecules such as proteins is often considered in distinct levels. The overall three-dimensional shape adopted by a protein molecule is called its tertiary structure (Figure 17.4.5).
- Many bond types produce the three-dimensional shape of proteins:
 - dispersion forces, such as between two non-polar side chains, e.g. $-\text{CH}(\text{CH}_3)_2$
 - dipole-dipole attractions
 - hydrogen bonds between polar functional groups in side chains, such as $-\text{OH}$ and $-\text{C=O}$
 - covalent bonds in disulfide bridges, such as between cysteine's R groups
 - ionic interactions between charged R groups.
- The three-dimensional shape of a protein is critical to its function.

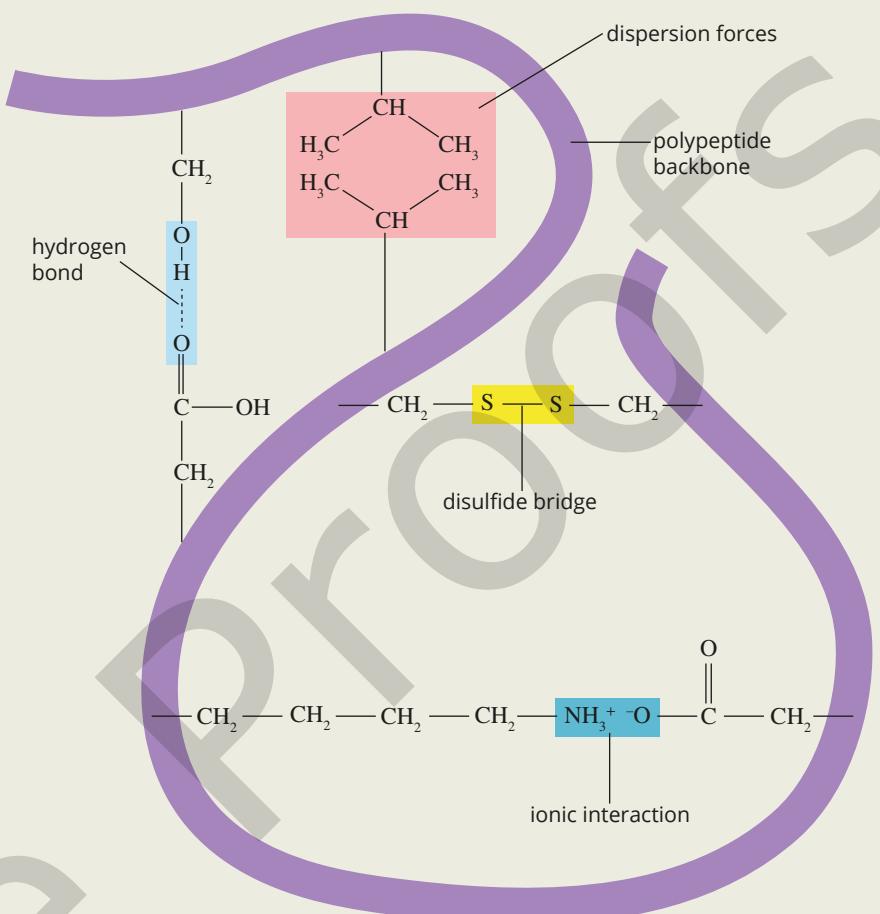


FIGURE 17.4.5 Protein tertiary structure

KEY QUESTIONS

- Identify the levels of protein structure in which hydrogen bonds may contribute to the structure. More than one option may be correct.
 - Primary structure
 - Secondary structure
 - Tertiary structure
- List the five main bond interactions that occur when side chains in different regions of a polypeptide chain interact with each other. Give examples of the required components in side chains involved in such interactions, and an example of an amino acid that has a side chain of this type.
- Identify the major type of bonding that would exist between the side chains of the following pairs of amino acids. You can assume that the conditions are such that any amino side group is positively charged and any carboxyl side group is negatively charged.
 - Proline and leucine
 - Glutamic acid and lysine
 - Glutamine and serine
 - Cysteine and cysteine

Chapter review

KEY TERMS

α -amino acid
 α -helix
amide
amino functional group
amino acid
 β -pleated sheet
C-terminal
carboxyl functional group

condensation
polymerisation
condensation reaction
dipeptide
essential amino acid
hydrolysis
hydrophobic
monomer

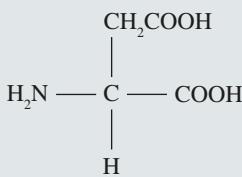
N-terminal
peptide bond
peptide link
polypeptide
primary structure
protein
quaternary structure
secondary structure

17

side chain (R group)
tertiary structure
tripeptide
zwitterion

Amino acids

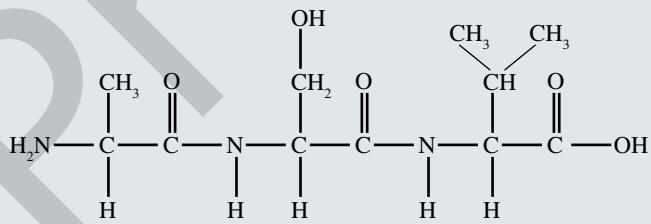
- 1 Select the correct statement about amino acids.
 - A Amino acids contain only the elements C, H, O and N.
 - B All 20 amino acids needed by humans are known as essential amino acids.
 - C Amino acids may contain more than one nitrogen atom.
 - D The functional group on an amino acid side chain is always non-polar.
- 2 What two functional groups are present in all amino acids?
- 3 Draw the general structure of an amino acid. Colour the amino group in blue and the carboxyl group in red.
- 4 Aspartic acid, shown below, is one of the non-essential amino acids. It exists mainly as a zwitterion at pH 2.8.



- a Name the:
 - i acidic functional group
 - ii basic functional group.
- b Explain what is meant by the term 'zwitterion'.
- c Draw the structure of the molecule as it is most likely to exist in a solution at pH:
 - i much greater than 2.8
 - ii much less than 2.8
 - iii 2.8.

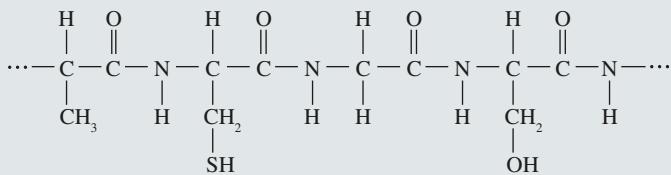
The formation of proteins

- 5 Which two functional groups react when a protein such as insulin is formed from its constituent amino acids?
- 6 The following diagram shows a tripeptide molecule. Select the correct statement about this tripeptide.



- a Three water molecules were released when this molecule was produced.
 - b This molecule can undergo hydrolysis to produce three different amino acids.
 - c There are four amino acids in this molecule.
 - d There are three carboxyl groups in this molecule.
- 7 When alanine and glycine react, two different dipeptides can be formed.
 - a Write the condensed structural formula of each dipeptide at pH 7.
 - b How many tripeptides can be formed from three amino acids?
 - c Proteins are formed using 20 different amino acids and can be hundreds of amino acid units in length. On the basis of your answers to parts a and b, what can you say about the number of different proteins possible?

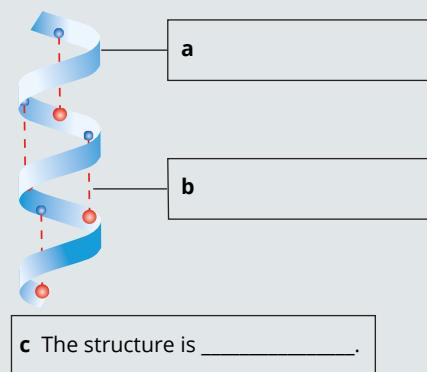
- d Circle the peptide bonds in the section of the protein chain shown below.



- e Name the amino acids that make up this section of the chain.

Primary and secondary structures of proteins

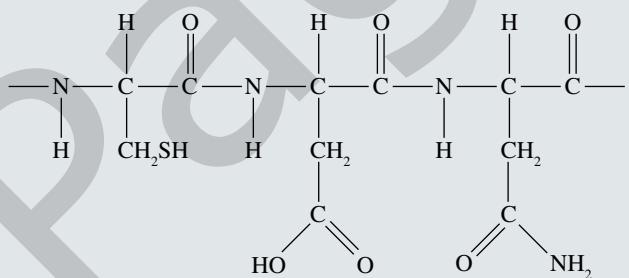
- 8** Identify the type of bonding represented by labels **a** and **b** of the following diagram. Identify the type of secondary structure in **c**.



- 9** How does a β -pleated sheet structure form in a protein?

Tertiary structure of proteins

- 10** Circle the parts of the polypeptide segment that might be involved in forming bonds responsible for the tertiary structures of the protein molecule.



- 11** Mutations can cause the wrong amino acid to be inserted into a protein chain when it is being assembled. In some instances, such a mistake has no effect on the function of the protein but in others it disables the protein. Explain how a single change in a protein's primary structure could cause such dramatically different results.

- 12** Explain what is meant by the primary, secondary and tertiary structures of a protein.

- 13** Name the type of interaction most likely to occur in the following pairs of amino acids, which are part of the same polypeptide chain. Assume that the amino groups are in the NH_3^+ form and the carboxyl groups are in the COO^- form.

- a** Aspartic acid and lysine
 - b** Cysteine and tyrosine
 - c** Leucine and isoleucine

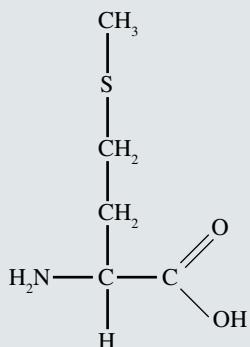
- 14** Keratin is the protein that makes up our hair and nails and the horns and claws of other mammals. Keratin is approximately 14% cysteine. This is why hair produces a pungent smell when burnt—the sulfur in the cysteine amino acids produces sulfur-containing gases. The cysteine amino acids in keratin are also responsible for much of the protein's strength and toughness. Explain this, using your knowledge of the interactions that are responsible for a protein's tertiary structure.

- 15** There are two possible mutations in a protein that result in the wrong amino acid being inserted in the protein chain in the place of leucine. These mutations result in either an isoleucine or a serine being inserted instead of leucine. Which mutation would have the greatest effect on the tertiary structure? Explain your answer.

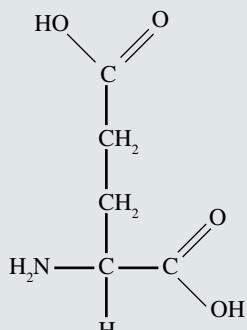
CHAPTER REVIEW CONTINUED

Connecting the main ideas

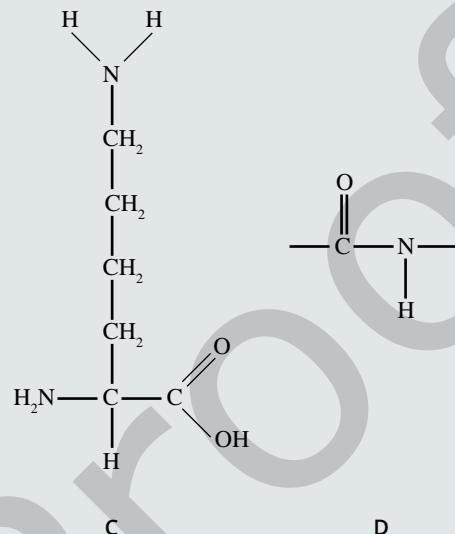
- 16 Use the structures of the three amino acids and a peptide group labelled A–D below to complete the statements that follow.



A



B



C

D

The side chain of molecule _____ might become positively charged if the pH is sufficiently _____. If this molecule were present in a protein, the side chain could take part in ionic interactions that contribute to the protein's tertiary structure.

The polar bonds that are present in _____ are responsible for the hydrogen bonds in an α -helix, contributing to the secondary structure of the protein. If there are two molecules of _____ in close proximity within a protein, they can form covalent bonds with each other. These covalent interactions contribute to the tertiary structure of the protein.

- 17 In a water-soluble protein, are the following amino acids most likely to be present on the surface of the protein, or hidden within the protein structure?

- a Leucine
- b Lysine
- c Phenylalanine
- d Aspartic acid

- 18 The amide links that form between amino acids are integral to both the primary and secondary structure of a protein. Explain this statement.