Amino acids link together through condensation reactions

Amino acids are able to react together in a condensation reaction in which a molecule of water is eliminated and a new bond is formed between the acid group of one amino acid and the amino group of the other. This bond is a substituted amide link known as a **peptide bond** and two amino acids linked in this way are known as a **dipeptide**. By convention the free $-NH_2$ group (known as the N-terminal) is put on the left of the sequence and the free $-COOH$ group (C-terminal) on the right.

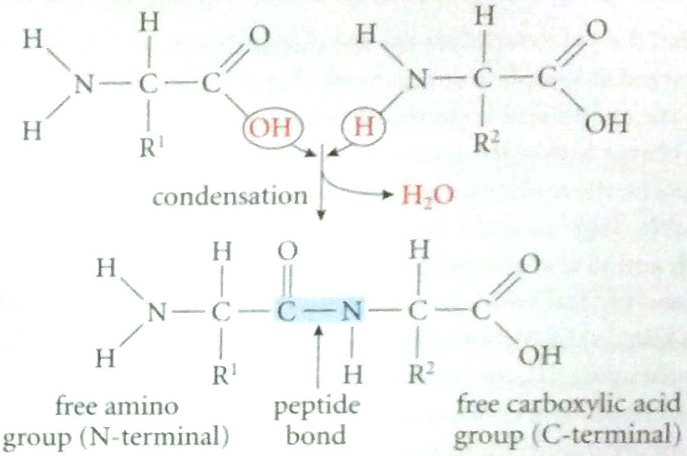


Figure 13.3 Formation of a dipeptide by condensation of two amino acids.

We can see that the dipeptide still has a functional group at each end of the molecule $-NH_2$ at one end and $-COOH$ at the other – so it can react again by condensation reactions, forming a **tripeptide** and eventually a chain of linked amino acids known as a **polypeptide**.

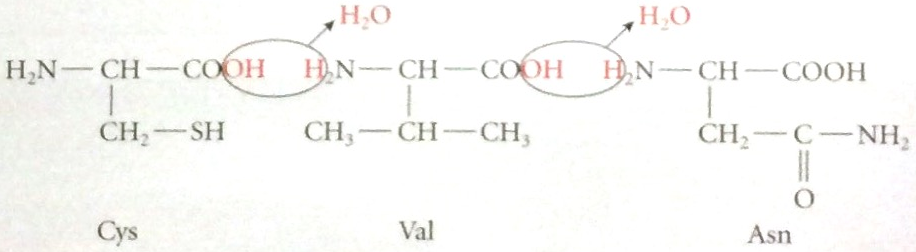
Worked example

Draw a tripeptide with the following sequence:

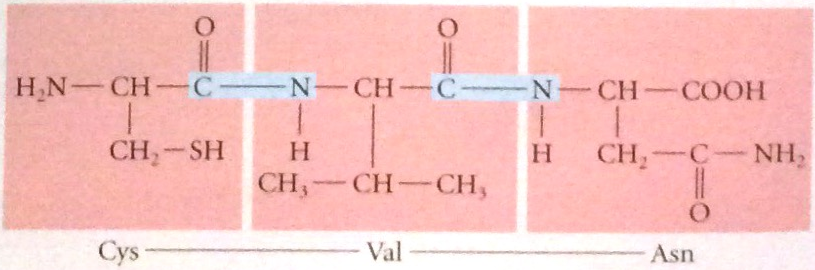
Cys–Val–Asn

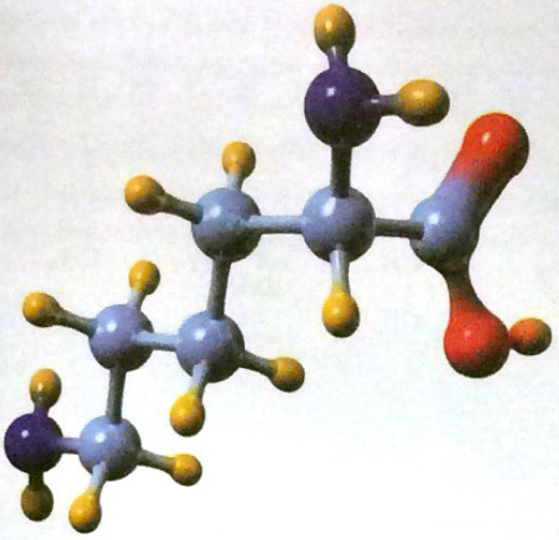
Solution

Look up the structures (the R groups) of the amino acids in Table 19 of the IB Data booklet and draw them out in the same order as given in the question.



Now draw peptide bonds between the carbon of the $-COOH$ group and the nitrogen of the $-NH_2$ group, ensuring that H_2O is released and that each atom has the correct number of bonds in the final structure.





Molecular model of the amino acid lysine $NH_2(CH_2)_4CH(NH_2)COOH$. Lysine is known as an 'essential' amino acid because it cannot be synthesized by the body and so must be obtained in the diet. Carbon atoms are shown in blue, hydrogen in gold, oxygen in red and nitrogen in dark blue.

In just the same way as letters linked together in different orders make different words (e.g. eat, ate, tea), so amino acids linked together in different sequences make different peptides. The three amino acids above could also have been linked for example as: Asn—Cys—Val or Cys—Asn—Val, all of which would be different tripeptides with different properties.

The sequence in which the amino acids are linked to form the chain is of tremendous significance as it will determine the exact nature of the polypeptide. This is where the extraordinary variety of protein structures comes from.

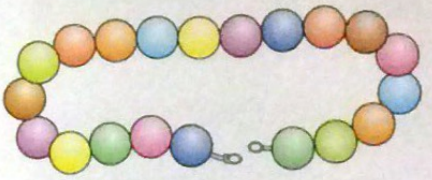


Figure 13.4 Different coloured beads can be strung together in different combinations to make a wide variety of different necklaces. In a similar way, different sequences of amino acids link together to give rise to an almost infinite variety of proteins.

Consider making a necklace by joining together 20 different colours of beads, with each colour being used as many times as you like. In every position you would have a choice of 20 different possibilities, so just imagine how many different combinations of beads you could have! Building polypeptides from amino acids presents a similar situation – at any point in the chain there are 20 different possibilities. If it was only a tripeptide there would be $20 \times 20 \times 20 = 8000$ different possible combinations. But proteins are typically made from polypeptide chains with 50 or more amino acids, so the number of possible structures becomes enormous, in this case 20^{50} – literally millions.

We will now move on to look at how the sequence of amino acids in the polypeptide uniquely determines the structure and therefore the function of the protein. Understanding protein structure can seem quite complex, so for convenience it is divided into four levels of organization.

The primary structure of a protein is the amino acid sequence

The primary structure of a protein refers to the number and sequence of amino acids in its polypeptide chain. Held together by peptide bonds, this forms the covalent backbone of the molecule. Interestingly, once the primary structure has been determined, all the other levels of protein structure will follow – so it really does dictate the entire structure and function of the protein. So crucial is the primary structure that the alteration of just one amino acid can completely change the functioning of the protein, as is the case in the disease sickle-cell anaemia. This condition occurs when the protein haemoglobin is not able to carry oxygen efficiently, the result of a single amino acid change in its chain of 146 amino acids. The primary structure is of key interest to biochemists and is now routinely analysed in laboratories. It is also the aspect of protein structure used in studies of biochemical evolution exploring the relationships between organisms.

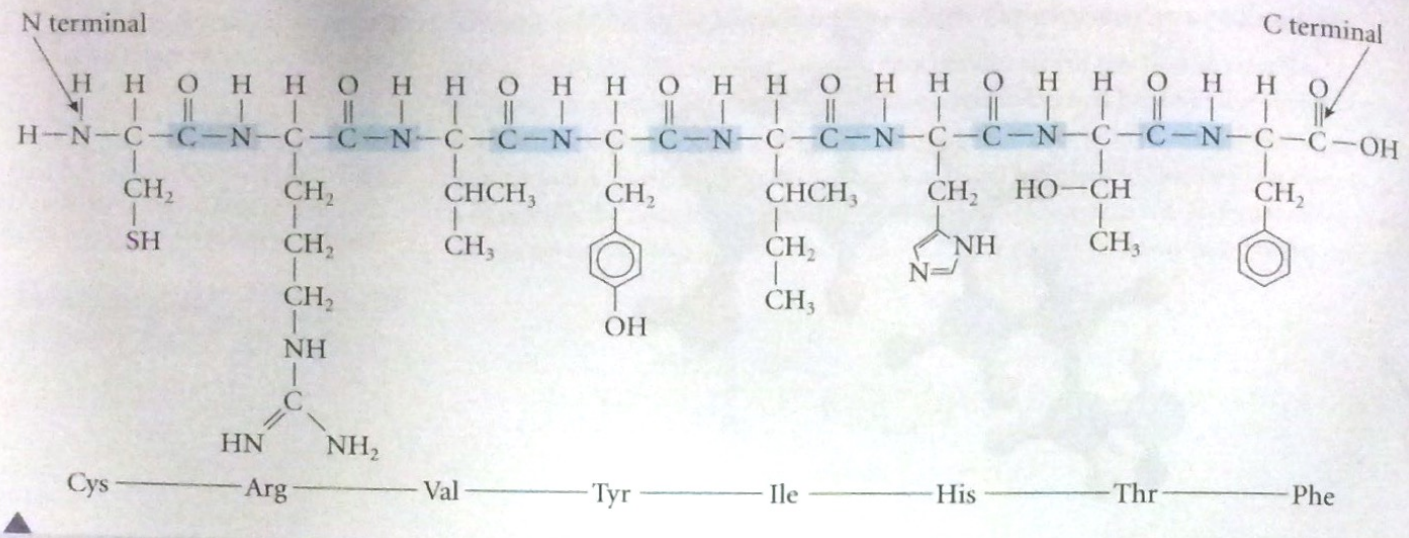


Figure 13.5 Primary structure of a small polypeptide. Peptide bonds are highlighted in blue.

So what determines the primary structure of a protein? The answer to this question when it came in the early 1960s was considered to be one of the most important discoveries of molecular biology. DNA (deoxyribose nucleic acid), which determines genetic information, acts by dictating to cells the primary structure of their proteins. Simply put, the expression of our genes is through the proteins that we synthesize. So it is the primary structure of our proteins that gives each of us our unique genetic characteristics.

The secondary structure of proteins is regular hydrogen bonding

The secondary structure refers to folding of the polypeptide chain as a result of hydrogen bonding between peptide groups along its length. Hydrogen bonds can form between the —C=O group of one peptide bond and the —N—H group of another peptide bond further along the chain which will cause the chain to fold. The exact configuration of this will be influenced by the R groups along the chain and so differs in different proteins and even in different sections of the same protein. We can distinguish two main types of secondary structure, the **α-helix** and the **β-pleated sheet**.

The α-helix is a regular coiled configuration of the polypeptide chain resulting from hydrogen bonds forming between two peptide bonds four amino acid units apart. This twists the chain into a tightly coiled helix, much like a spiral staircase, with 3.6 amino acids per turn.

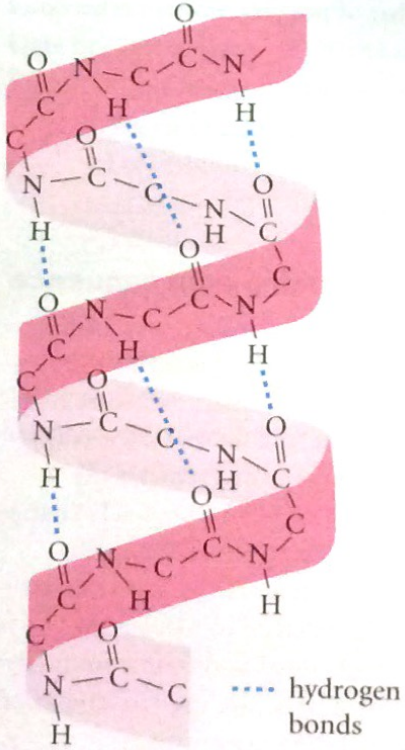


Figure 13.6 α-helical secondary structure of keratin. The amino acid backbone winds in a spiral, held by hydrogen bonds shown in blue.

Often considered the father of modern molecular biology, Fred Sanger of Cambridge, England, is one of the few people to have been awarded two Nobel Prizes. In 1958 he was awarded the prize for establishing the sequence of the 51 amino acids in insulin chain B. This was the first protein to have its primary structure elucidated in this way and it was the culmination of 12 years of work. In 1980 Sanger shared the Nobel Prize for similar work on the base sequencing of nucleic acids.

You can watch an interview with Fred Sanger reflecting on his achievements in research and on his life philosophy.

Now go to www.heinemann.co.uk/hotlinks, insert the express code 4259P and click on this activity.

The α-helix is flexible and elastic because the *intra-chain* hydrogen bonds easily break and reform as the molecule is stretched. A good example of the α-helix is found in keratins – structural proteins found in hair, skin and nails.

The β-pleated sheet is a structure composed of 'side by side' polypeptides which are in extended form, that is, not tightly coiled as in the α-helix. The polypeptides are arranged in pleated sheets that are cross-linked by *inter-chain* hydrogen bonds.

Human hair (made of the protein keratin) grows approximately 15 cm in one year, which means that 9.5 turns of the α-helix must be produced every second.

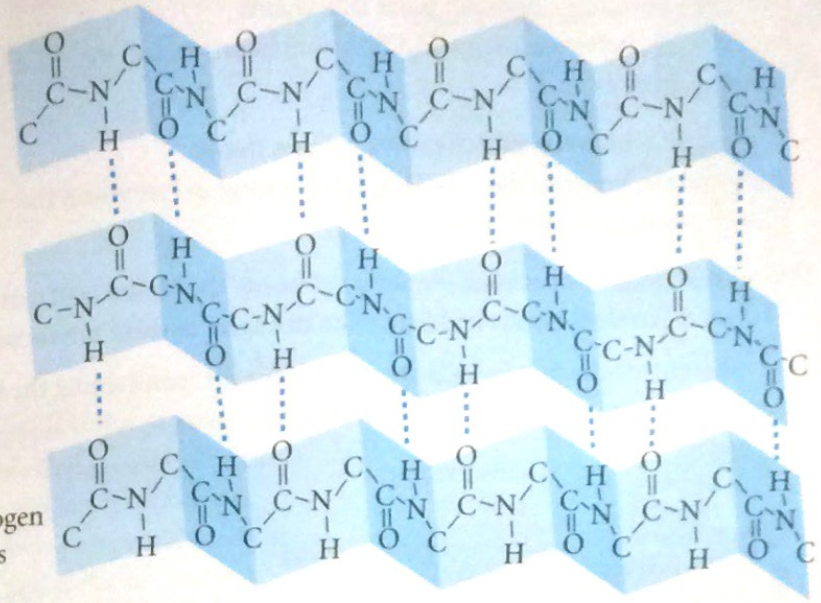


Figure 13.7 β-pleated sheet secondary structure of silk fibroin. The polypeptides run parallel to each other, held in place by hydrogen bonds.

Hair stretches to almost double its length when exposed to moist heat but contracts to its normal length on cooling. This is why hair is often much curlier in humid conditions.

The β-pleated sheet is flexible but inelastic. It is found in the fibres spun by spiders and silkworms and in the beaks and claws of birds.

Spider's web with drops of morning dew. The web fibres are made of the protein fibroin containing a β-pleated sheet secondary structure and spun from special secreting glands. A typical web may contain 20 m of fibroin.

Proteins that have a well-defined secondary structure such as those described here are known as **fibrous** proteins. They are physically tough and insoluble in water.

The tertiary structure of proteins is the result of interactions between the R groups

The tertiary structure refers to the further twisting, folding and coiling of the polypeptide chain as a result of interactions between the R groups, known as **side chains**. The structure that results will be a very specific compact three-dimensional structure, known as the protein's **conformation**. It is the most stable arrangement of the protein, taking into account all the possible interactions along the entire length of the polypeptide. Note that the interactions between the R groups are all *intra*-molecular forces, as they occur within the one polypeptide chain.

The conformation is particularly important in the so-called **globular** proteins, which include all the enzymes and protein hormones. They are water soluble because their structure positions nearly all of the polar (or hydrophilic) R groups on the outer surface of the molecules where they can interact with water and most

Computer artwork showing the structure of the protein albumen from human blood. The spiral regions of α-helical structure can be clearly seen as well as the overall three-dimensional conformation.

of the non-polar (or hydrophobic) R groups in the interior out of contact with water. The interactions that stabilize this conformation are of the following types.

(a) Hydrophobic interactions – between non-polar side chains
For example, between two alkyl side chains in valine; these weak interactions, based on van der Waals' forces between induced dipoles, produce non-polar regions in the interior of the protein.

(b) Hydrogen bonding – between polar side chains
For example, between the $-CH_2OH$ group in serine and the $-CH_2COOH$ group in aspartic acid.

(c) Ionic bond – between side chains carrying a charge
For example, between the $-(CH_2)_4NH_3^+$ group in lysine and the $-CH_2COO^-$ group in aspartic acid.

(d) Disulfide bridges – between the sulfur atoms in the amino acid cysteine
These are covalent bonds and hence the strongest of these interactions.

These different interactions which are responsible for maintaining the tertiary structure are summarized in the diagram below.

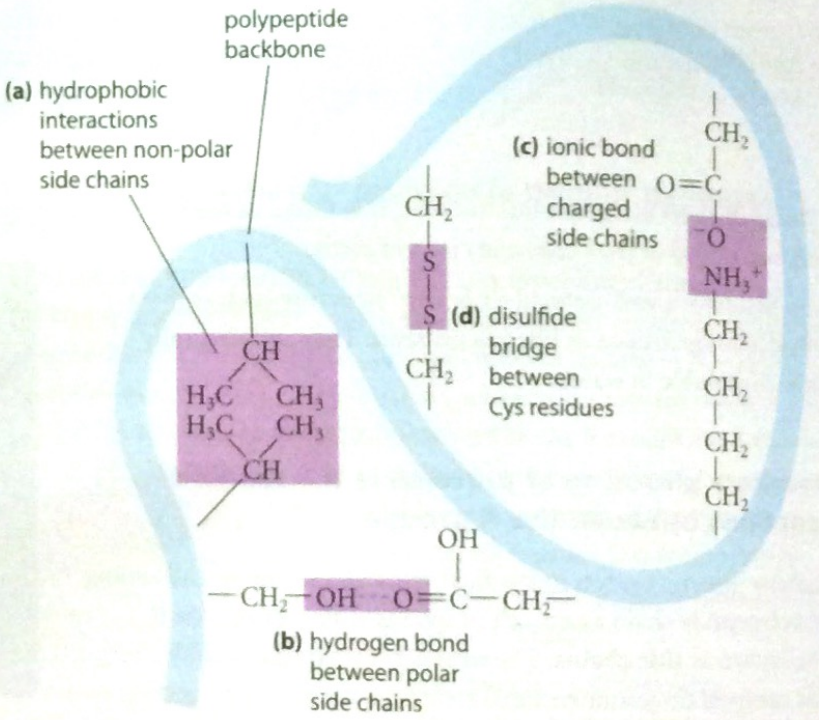


Figure 13.8 Summary of interactions contributing to tertiary structure.

The process of 'perming' which introduces more curls into hair involves first breaking the disulfide bridges between the cysteine residues in the keratin using a reducing agent. New disulfide bridges are formed when the hair is chemically re-oxidized while it is twisted around rollers. The size of the roller determines the position of the new disulfide bridges that form. As these are covalent bonds they do not break on normal treatments like washing and combing, so are said to be 'permanent'. Similar processes are used in straightening curly hair.

These interactions can all be upset by changes in the medium, such as changes in temperature or pH. When a protein loses its specific tertiary structure as a result of such disruptions, it is said to be **denatured**. The familiar sight of the white of an egg solidifying on heating is an example of this. Denaturation of enzymes renders them biologically inactive, which is one of the reasons why intracellular conditions must be tightly controlled.

The quaternary structure of a protein is the association between different polypeptides

Some proteins comprise more than one polypeptide chain and in these cases the association between these chains is known as the quaternary structure. For example, the protein collagen, which is found in skin and tendons and is actually the most abundant protein in the human body, is a triple helix of three

polypeptide chains, with inter-chain hydrogen bonds between them. This helps to give it a stable rope-like structure that is resistant to stretching.

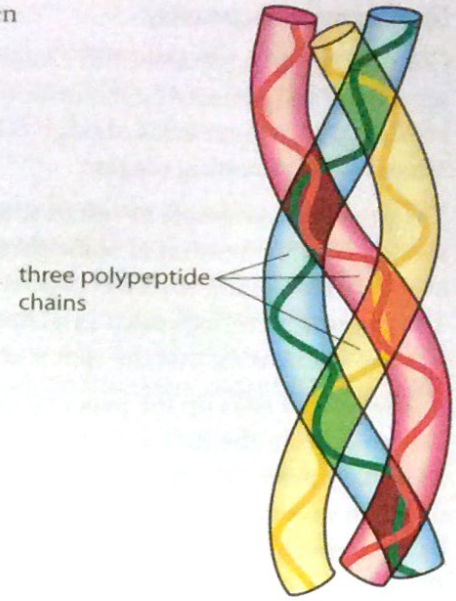


Figure 13.9 Triple helical quaternary structure of collagen.

Another example is the protein haemoglobin that is responsible for carrying oxygen in the blood. It is made up of four polypeptide chains, known as α and β chains, which fit together tightly in the protein assembly.

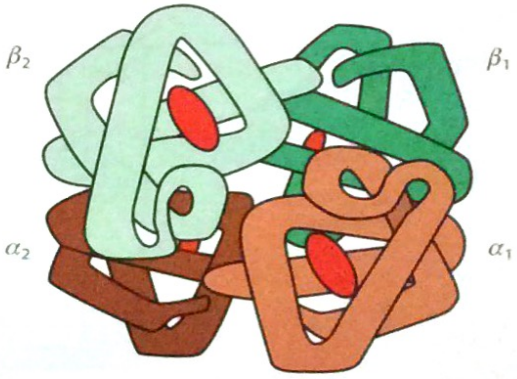


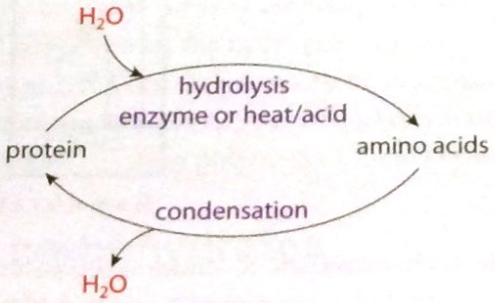
Figure 13.10 Quaternary structure of haemoglobin, an assembly of four polypeptides.

The four polypeptide chains in adult haemoglobin are two α chains and two β chains. Before birth, haemoglobin in the foetus has a different structure with two α chains and two γ chains. This form has a higher affinity for oxygen and so is able to extract it from the maternal blood. After birth, the foetal haemoglobin levels decline and by six months adult haemoglobin becomes the predominant form.

Many proteins consist of only one polypeptide chain and so do not have a quaternary structure.

Analysis of proteins

The analysis of a protein is typically begun by determining its amino acid composition. (Note that this is not the same as its primary structure, as the *sequence* of the amino acids will not be known). This involves first chemically separating the amino acids from each other by breaking the peptide bonds between them through **hydrolysis** reactions usually using acid. These reactions reverse the condensation reactions discussed earlier and occur in the body during enzyme-catalysed protein digestion in the intestine.



You can follow an illustrated summary of the four levels of protein structure.

Now go to www.heinemann.co.uk/hotlinks, insert the express code 4259P and click on this activity.

Separation of the resulting amino acid mixture into its components can then be achieved in two ways.

(i) Chromatography

Chromatography (described in Chapter 12, page 245) is a useful technique for separating and identifying the components of a mixture, particularly when they are coloured. Amino acids, though colourless in solution, take on colour when treated with a **locating reagent**.

The procedure is simple to run. A small sample of the amino acid mixture is spotted near the bottom of some chromatographic paper and this position, known as the **origin**, clearly marked (in pencil so as not to interfere with the experiment). The paper is then suspended in a chromatographic tank containing a small volume of solvent, ensuring that the spot is above the level of the solvent.

As the solvent rises up the paper by capillary action it will pass over the spot. Amino acids in the spot will distribute themselves between two phases – the stationary phase (the water in the paper) and the mobile phase (the solvent) to different extents and so move up the paper at different speeds. They will therefore become spread out according to their different solubilities.

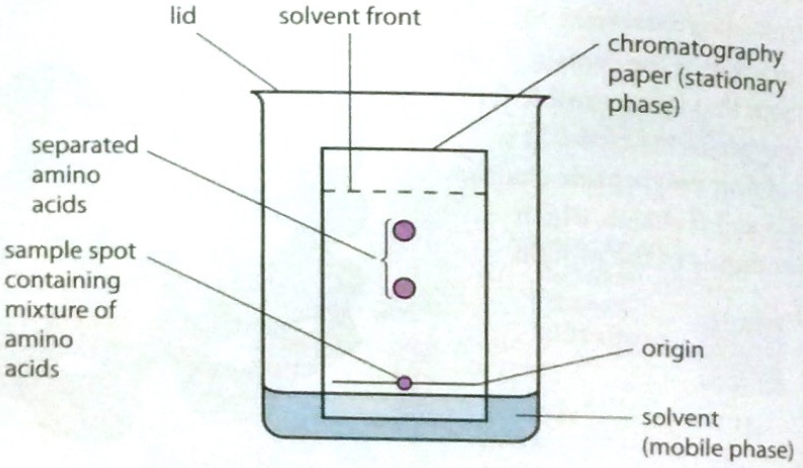
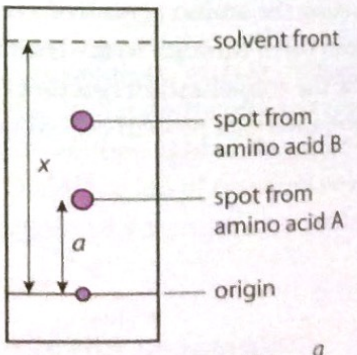


Figure 13.11 Apparatus used to separate amino acids by paper chromatography.

When the solvent reaches almost to the top of the paper, its final position is marked and is known as the **solvent front**. The paper is removed from the tank and developed by spraying it with the locating reagent **ninhydrin**. Most amino acids will now take a purple colour and can be distinguished as separate isolated spots up the length of the paper.

The position of each amino acid can be expressed as an R_f value (retention factor), calculated as follows:

$$R_f = \frac{\text{distance moved by amino acid}}{\text{distance moved by solvent}}$$



R_f value for amino acid A = $\frac{a}{x}$

Figure 13.12 Calculation of R_f values in chromatography.

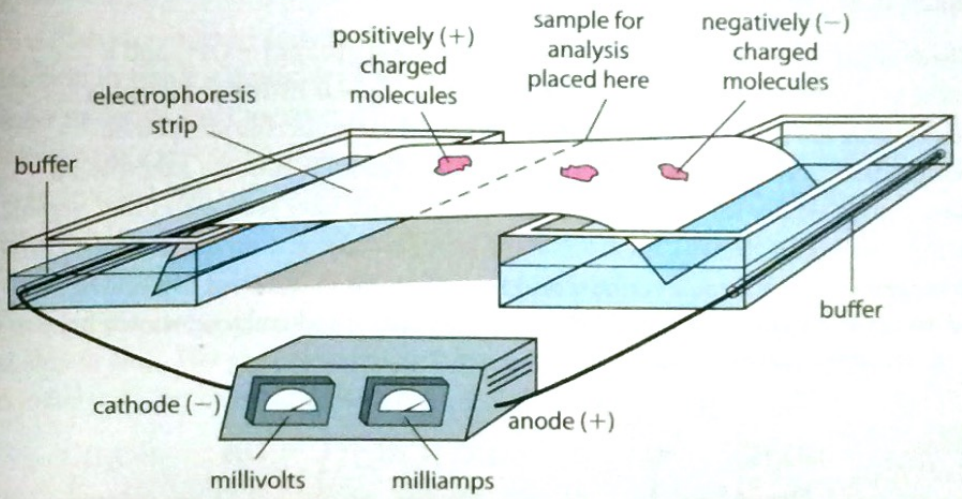
Specific amino acids have characteristic R_f values when measured under the same conditions, so can be identified by comparing the values obtained with data tables. It is helpful to spot known amino acids alongside the mixture to act as markers for the experiment.

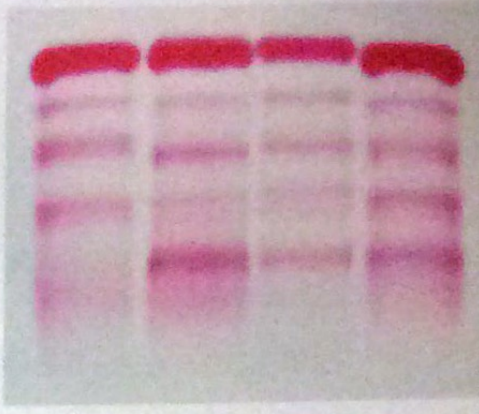
(ii) Electrophoresis

Electrophoresis is a technique for the analysis and separation of a mixture based on the movement of charged particles in an electric field. As we learned earlier, amino acids carry different charges depending on the pH and so can be separated by this means when placed in a buffered solution.

In gel electrophoresis, the medium is a gel, typically made of polyacrylamide. The amino acid mixture is placed in wells in the centre of the gel and an electric field is applied. Depending on the pH of the buffer used, different amino acids will move at different rates towards the oppositely charged electrodes. At their isoelectric point, amino acids will not move as they carry no net charge. When separation is complete they can be detected by a stain or made to fluoresce under UV light and identified from their position using data tables.

Electrophoresis can also be used to separate and identify intact proteins according to their different rates of migration towards the poles.





▲ Electrophoresis of human blood serum. The charged protein molecules in a buffer solution have separated according to their different attractions to opposite electrical poles across the polyacrylamide gel (Polyacryl Amide Gel Electrophoresis or PAGE). The bands have been stained, so the separated proteins can be identified from their position. This technique is used to detect whether proteins associated with particular diseases are found in the blood.

◀ **Figure 13.13** Electrophoresis apparatus for amino acid or protein separation.

Exercises

Make reference to Table 19 in the IB Data booklet to answer these questions.

2 Explain why in gel electrophoresis the amino acid isoleucine migrates towards the anode at high pH and towards the cathode at low pH.

3 You are attempting to separate a mixture of glutamic acid and histidine by gel electrophoresis. Give a suggested pH for an appropriate buffer solution to use and say in which direction each amino acid would migrate.

W You can do a virtual experiment on gel electrophoresis.

Now go to www.heinemann.co.uk/hotlinks, insert the express code 4259P and click on this activity.

B.3 Carbohydrates

Carbohydrates (literally *hydrated-carbon*) are composed of the three elements carbon, hydrogen and oxygen, with the hydrogen and oxygen always in the same ratio as in water i.e. 2:1. They therefore can be expressed by the general formula $C_x(H_2O)_y$. There are two main types of carbohydrate – simple sugars or **monosaccharides** and condensation polymers of these known as **polysaccharides**.

Functions of carbohydrates

The monosaccharides, for example **glucose** and **fructose**, are readily soluble in water and are mostly taken up by cells quite rapidly. They are used as the main substrate for respiration, releasing energy for all cell processes. They also act as