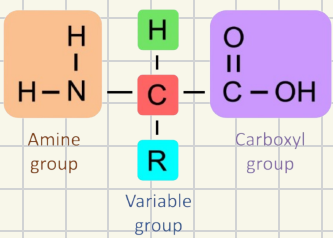
Proteins - B1.2

B1.2.1 - generalized structure of an amino acid

Amino acids are the building blocks of proteins. Each amino acid molecule has a central carbon atom called the alpha carbon, with single covalent bonds to four other atoms. One of these is the nitrogen atom of an amine group and another is the carbon atom of a carboxyl group. The carboxyl group (-COOH) is acidic because it can donate a proton and the amine group is basic because it can accept one, so amino acids are amphiprotic.



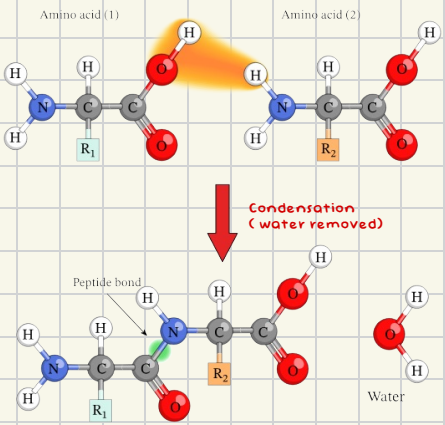
The alpha carbon atom also has a single covalent bond to a hydrogen atom. The other covalent bond links the alpha carbon to a side chain, called the R-group. The R-group can be any one of a wide range of possibilities.

B1.2.2 - condensation reactions forming dipeptides and long chains of amino acids

To form a dipeptide, two amino acids are linked by a condensation reaction. More amino acids can be linked by further condensation reactions to create a longer chain. Polypeptides can contain any number of amino acids, though chains of fewer than 20 amino acids are usually referred to as oligopeptides rather than polypeptides. Polypeptides are the main component of proteins.

Amino acids are linked with peptide bonds. These are C-N bonds formed by a condensation reaction between the amine group (NH_2) of one amino acid and the carboxyl group (COOH) of another.

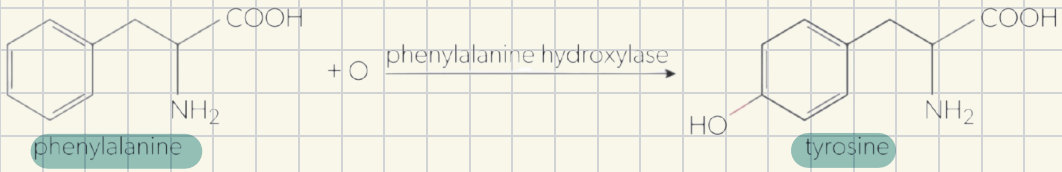
The reaction is catalysed in cells by ribosomes. It is a directional process: the amine group of a free amino acid is linked to the carboxyl group at the end of the growing chain. Because peptide bonds are made using groups that are part of all amino acids, the bond is the same, whatever the R-groups of the amino acids are.



B1.2.3 - dietary requirements for amino acids

Twenty different amino acids are used by ribosomes to make polypeptides. Plants can make all of these by photosynthesis. Animals need to obtain essential amino acids from their food. A non-essential amino acid can be synthesized by an animal using metabolic pathways that transform one amino acid to another.

Nine of the twenty amino acids are essential in humans. The others are non-essential, though several become essential in special circumstances. For example, the amino acid phenylalanine is essential because it cannot be synthesized by the human body; tyrosine is non-essential because it can be made from phenylalanine. *



Foods vary in their amino acid content. It is possible to eat a protein-rich diet and still be deficient in an essential amino acid. Animal-based foods (fish, meat, milk, eggs) have a balance of amino acids that is similar to what is needed in the human diet. Plant-based foods have a different balance and some are deficient in specific amino acids. For example, cereals such as wheat have a low lysine content, and peas and beans are low in methionine.

Both lysine and methionine are essential amino acids for humans. So, people eating a vegan diet must ensure enough of each essential amino acid is consumed.

B1.2.4 - infinite variety of possible peptide chains

Ribosomes link amino acids together one at a time, until a polypeptide is fully formed. The ribosome can make peptide bonds between any pair of amino acids, so all sequences are possible. Ribosomes do not make random sequences of amino acids. They receive instructions in the form of genetic code. Twenty different amino acids are included in the code.

The number of possible amino acid sequences starts with dipeptides, where both amino acids can be one of 20, resulting in 20 × 20 combinations. For tripeptides, it's 20 × 20 × 20, and for a polypeptide of n amino acids, there are 20^n possible sequences. A polypeptide with 400 amino acids, for example, has 20^400 possible sequences, a number so large it's often considered infinite. However, only a tiny fraction of these sequences are made by organisms, known as the proteome.

Examples of polypeptides

- Beta-endorphin is a natural pain killer secreted by the pituitary gland that is a polypeptide of 31 amino acids

- Insulin is a small protein that contains two short polypeptides, one with 21 amino acids and the other with 30 amino acids.

- Alpha amylase is the enzyme in saliva that starts the digestion of starch. It is a single polypeptide of 496 amino acids, with one chloride ion and one calcium ion associated.

- Titin is the largest polypeptide discovered so far. It is part of the structure of muscle. In humans, titin is a polypeptide of 34,350 amino acids, but in mice it is even longer with 35,213 amino acids.

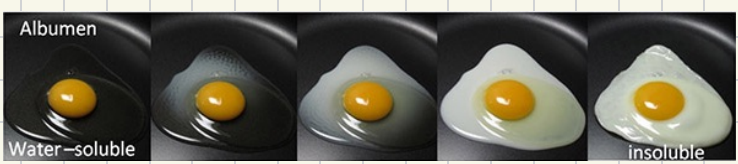
Cysteine and methionine are the only amino acids with sulfur

B1.2.5- effect of pH and temperature on protein structure

The three-dimensional conformation of proteins is stabilized by bonds or interactions between the R-groups of amino acids within the molecule. Most of these bonds and interactions are relatively weak and they can be disrupted or broken. This results in a change to the conformation of the protein and is called denaturation.

A denatured protein does not normally return to its former structure, the denaturation is permanent. Soluble proteins often become insoluble and form a precipitate. This is due to the hydrophobic R-groups in the centre of the molecule becoming exposed to water by the change in conformation.

Heat can cause denaturation because it causes vibrations within the molecule that can break intermolecular bonds or interactions. Proteins vary in their heat tolerance. Some microorganisms that live in volcanic springs or in hot water near geothermal vents have proteins that are not denatured by temperatures of 80°C or higher. Nevertheless, heat causes denaturation of most proteins at much lower temperatures.



Extremes of pH, both acidic and alkaline, can cause denaturation. This is because positive and negative charges on R-groups are changed, breaking ionic bonds within the protein or causing new ionic bonds to form. As with heat, the three-dimensional structure of the protein is altered and proteins that have been dissolved in water often become insoluble. There are exceptions: the contents of the stomach are normally acidic, with a pH as low as 1.5, but this is the optimum pH for the protein-digesting enzyme pepsin that works in the stomach.

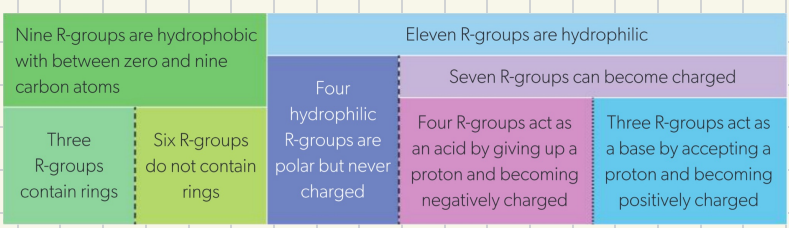
When a fluid environment such as cytoplasm or blood plasma is flooded with either H+ ions (an acid) or -OH ions a(base), the extra charges can prevent normal hydrogen bonding. The protein wil not take on its "normal" shape and wil not function normally.

B1.2.6 - Chemical diversity in the r-groups of amino acids

The 20 amino acids that ribosomes use to make polypeptides are very varied in the chemical nature of their R-groups. The elements present in the R-groups:

Elements in R-group	Number of amino acids
H only	1
C and H only	5
C, H and S only	2
C, H and N only	5
C, H and O only	5
C, H, N and O	2

When amino acids are linked up into a polypeptide, their amine and carboxyl groups are used to make peptide bonds. This leaves an amine group ($-NH_2$) at one end of the chain and a carboxyl group (-COOH) at the other end. The hydrogen atom attached to the alpha carbon atom of each amino acid has little effect on the properties of the polypeptide; it is the R-groups that determine the chemical characteristics. Some of the R-groups are hydrophobic and some hydrophilic. Of the hydrophilic R-groups, some are polar and others become charged (+ or -) by acting as an acid or a base. This broad diversity of R-groups allows living organisms to make and use an amazingly wide range of proteins. Some of the differences between R-groups are show:



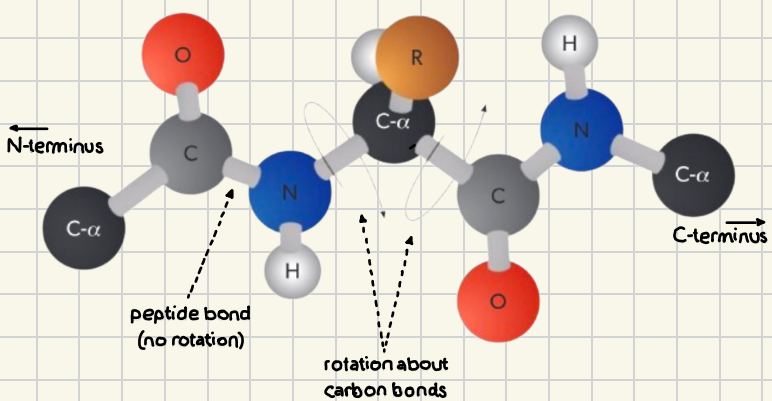
- Proteins are generally made from 20 standard amino acids.
- Some proteins have amino acids that are modified after the protein is synthesized (called post-translational modification).
- Collagen is an example of a structural protein that undergoes such modifications.
- In collagen, the amino acid proline is present in many places along its polypeptide chain.
- After synthesis, some proline residues are converted to hydroxyproline.

- This modification strengthens the collagen, making it more stable and durable.
- Collagen's enhanced stability provides tensile strength in tendons, ligaments, skin, and blood vessels.

B1.2.7 - Impact of primary structure on the conformation of proteins

The structure of proteins has four levels of complexity: primary, secondary, tertiary and quaternary. Primary structure is the primary sequence of amino acids in a polypeptide.

The backbone of a polypeptide is the main chain of atoms that make up a protein, and it's made up of a repeating pattern of carbon (C), nitrogen (N), and more carbon atoms. These atoms are held together by strong covalent bonds, and the angles between them are arranged in a specific way (called tetrahedral). Importantly, the bonds between the alpha carbon (a central carbon in the chain) and the nearby nitrogen and carbon atoms can rotate, giving the chain flexibility. Because of this rotation, the polypeptide chain can bend and fold into many different shapes. This ability to fold is crucial because the specific shape of a protein determines how it works in the body.



The three-dimensional arrangement of atoms in a polypeptide or protein is its conformation. Most polypeptides self-assemble into a specific conformation determined by the sequence of amino acids and their R-groups.

The conformation of proteins determines their functions and through this the behaviour of cells

B1.2.8 - Pleating and coiling of secondary structure of proteins

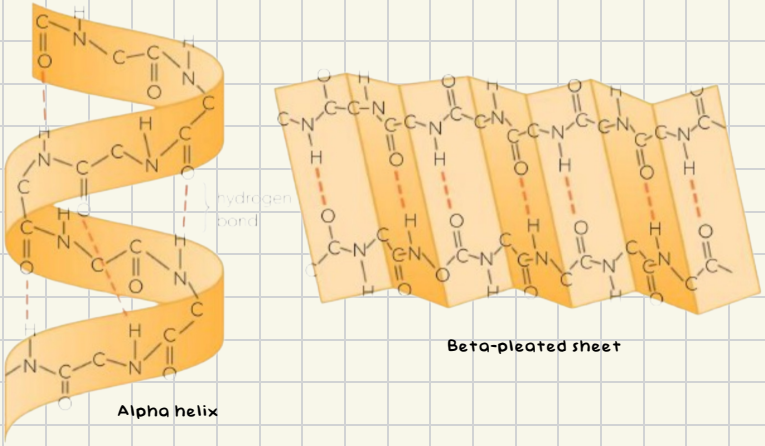
At regular intervals along a polypeptide chain there are C=O and N-H groups. They are what remains of carboxyl and amine groups after they have been used to make peptide bonds. Both C=O and N-H are polar, with the oxygen having a slight negative charge and the hydrogen a slight positive charge. Due to this polarity, hydrogen bonds can form between these groups. Although hydrogen bonds are individually weak, the frequency of C=O and N-H groups along polypeptide chains allows many of them to form and collectively they are strong enough to stabilize distinctive conformational structures within protein molecules.

Two commonly occurring types of structure are stabilized by hydrogen bonding.

- The a-helix: the polypeptide is wound into a helical shape, with hydrogen bonds between adjacent turns of the helix.

- The b-pleated sheet: two or more sections of polypeptide are arranged in parallel with hydrogen bonds between them. The sections of polypeptide run in opposite directions, forming a sheet that is pleated because of the tetrahedral bond angles.

Regular structures stabilized by hydrogen bonding within polypeptides are the secondary structure of a protein.



B1.2.9 - Dependence of tertiary structure on hydrogen bonds, ionic bonds, disulfide covalent bonds and hydrophobic interactions

Tertiary structure is the folding of a whole polypeptide chain into a three-dimensional structure. This structure is stabilized by interactions between R-groups. There are four main types of interaction.

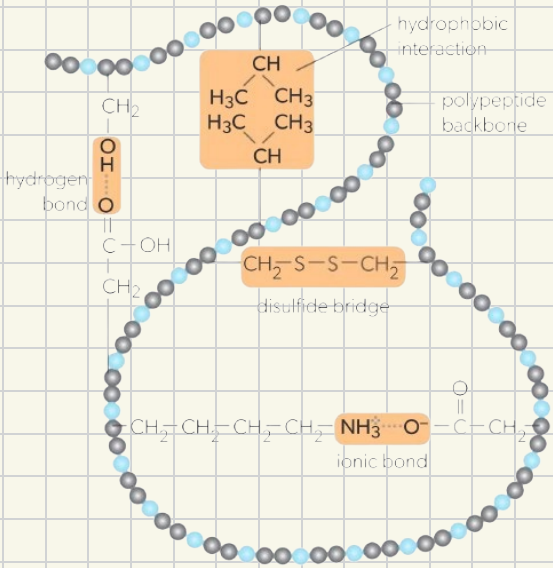
Ionic bonds between positively charged and negatively charged R-groups. Amine groups become positively charged by accepting a proton

$-NH^2 + H^+ \rightarrow -NH^{3+}$

Carboxyl groups become positively charged by donating a proton

$-COOH \rightarrow -COO + H^+$

Because of the involvement of protons (hydrogen ions), ionic bonds in proteins are sensitive to pH changes.



Hydrogen bonds between polar R-groups. A hydrogen atom forms a link between two electronegative atoms such as O or N. It is covalently bonded to one of them, which results in the hydrogen having a slight positive charge, making it attractive to the other, which has a slight negative charge.

Disulfide bonds between pairs of cysteines. This is a covalent bond and the strongest of all the interactions.

Hydrophobic interactions between any of the non-polar R-groups.

Tertiary structure develops as a polypeptide is synthesized by the ribosome. In some cases, a chaperone protein helps with this process to ensure that it results in a correctly folded and fully functional protein.

A wide range of three-dimensional shapes is produced, most of which are globular. Within these tertiary structures there are often parts with secondary structure: α-helices and/or β-pleated sheets. Some polypeptides do not become folded and instead remain elongated: they do not have tertiary structure. These are fibrous proteins and have structural roles.

B1.2.10 Effect of polar and non-polar amino acids on tertiary structure of proteins

Amino acids in proteins can be divided into two broad categories:
- non-polar and therefore hydrophobic
- polar or charged and therefore hydrophilic.

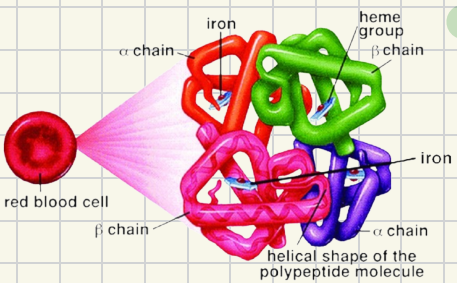
Many globular proteins need to be soluble in water because they carry out their function in the cytoplasm or in an aqueous solution outside the cell. These proteins have hydrophilic amino acids on their surface where they are in contact with water and hydrophobic amino acids clustered in the centre where water is excluded. This arrangement stabilizes the tertiary structure of the protein because it maximizes hydrophobic interactions between amino acids in the centre and hydrogen bonding between amino acids on the surface and the water around the protein.

Some proteins are routinely in contact with non-polar substances over some or all their surface. Such proteins have hydrophobic amino acids on parts of their surface. Integral proteins embedded in membranes have hydrophobic amino acids where they contact the non-polar hydrocarbon core of the membrane. In transmembrane proteins this hydrophobic region is a belt, with hydrophilic regions inside and outside that are in contact with aqueous solutions inside and outside the cell. This arrangement both stabilizes the tertiary structure of the protein and ensures that it remains positioned correctly in the membrane where its function can be performed.

Channel proteins in membranes allow hydrophilic solutes or water to diffuse across the hydrophobic core of the membrane. They have hydrophilic regions with a hydrophobic region between, which holds them in a transmembrane position. In addition, they have a tunnel lined with hydrophilic amino acids through the centre of the protein. The width and charge distribution of this channel allows specific hydrophilic ions or molecules to pass through.

B1.2.11 - Quaternary structure of non-conjugated and conjugated proteins

All proteins have at least one polypeptide, but many consist of two or more polypeptides linked together and some have one or more non-polypeptide components. In proteins that consist of more than a single polypeptide, the three-dimensional arrangement of subunits is the quaternary structure.



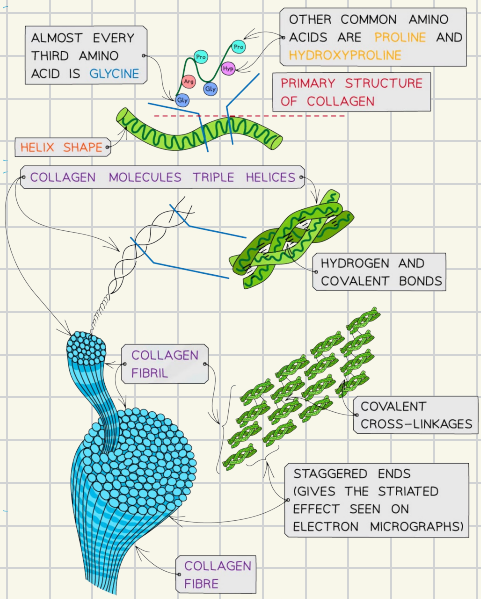
In a non- conjugated protein there are only polypeptide subunits. To form the quaternary structure the polypeptides are linked by the same types of interaction as in tertiary structure. For example, insulin has two polypeptides, linked by disulfide bonds. Collagen is another non-conjugated protein. It consists of three polypeptides wound together to form a rope-like structure with high tensile strength.

Conjugated proteins have one or more non-polypeptide subunits in addition to their polypeptides. For example, the haemoglobin molecule consists of four polypeptide chains, each associated with a haem group. The inclusion of non-polypeptide components increases the chemical and functional diversity of proteins. The haem group of haemoglobin binds oxygen, allowing this protein to transport oxygen. Many enzymes have a non-polypeptide component that contributes to the catalytic activity of their active site.

B1.2.12 - Relationship of form and function in globular and fibrous proteins

The function of a protein depends on its form. This can be illustrated by considering the difference between **fibrous and globular proteins**. Fibrous proteins consist of **elongated** polypeptides that lack the folding of typical tertiary structure. Also, the polypeptides in fibrous proteins do not develop secondary structures such as **alpha helices**. Their quaternary structure is developed by linking together polypeptide chains into **narrow fibers or filaments**, with **hydrogen bonds** between the chains.

Collagen is an example of a fibrous protein. The quaternary structure is three polypeptides, wound together into a triple helix. The primary structure of the polypeptides is a repeating sequence of three amino acids: P G X. The P in this sequence is proline or hydroxyproline, which has the special property of preventing formation of an a-helix. The winding together of the three polypeptides would be impossible if they were a-helices. The R-group of every third amino acid faces inwards towards the centre of the triple helix and glycine is the only amino acid with an R-group small enough to fit: it is a single hydrogen atom.



The rope-like structure of collagen gives it **very high tensile strength.** The R-group of amino acid X faces outwards and is variable, allowing many variations of collagen to be produced for use in skin, **tendons, ligaments, cartilage**, basement membranes of epithelia and the tough outer coat of the eye (visible at the front as the white of the eye).

Globular proteins have a rounded shape, formed by the folding up of polypeptides. The shape is very intricate and is stabilized by bonds between the R-groups of the amino acids that have been brought together by the folding. There are many examples of the precise position of each atom in a globular protein, known as the conformation, being critical to the protein's function. The active site of enzymes and the ligand-binding site of receptors show this relationship. Insulin is another example. Only an insulin molecule has the conformation needed to bind to a specific site on the insulin receptor. This allows a specific and unambiguous signal to be sent to body cells when blood sugar concentration is too high.