DP IB Biology: SL

Your notes

2.3 Proteins

Contents

Your notes

2.3.1 Proteins

Amino Acids & Polypeptides

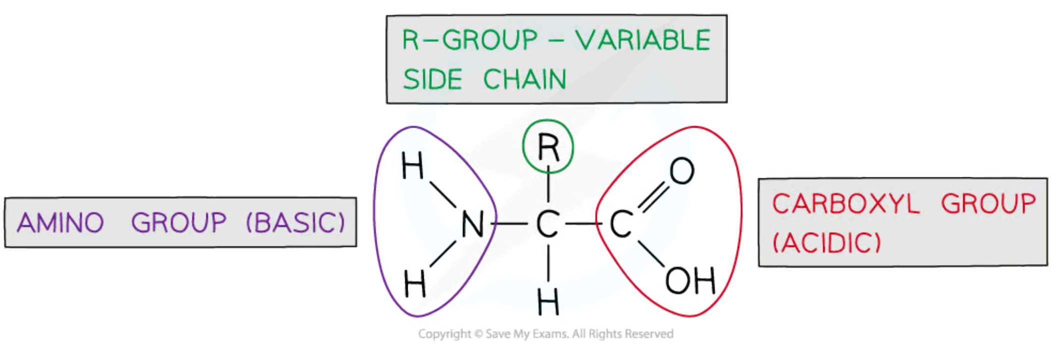
Proteins

- Proteins are polymers (and macromolecules) made of monomers called **amino acids**
- The **sequence, type** and **number** of the amino acids within a protein determines its shape and therefore its function
- Proteins **are extremely important in cells** because they form all of the following:
  - **Enzymes**
  - Cell membrane proteins (eg. carrier)
  - **Hormones**
  - Immunoproteins (eg. immunoglobulins)
  - **Transport** proteins (eg. haemoglobin)
  - **Structural** proteins (eg. keratin, collagen)
  - **Contractile** proteins (eg. myosin)

- Because all genes code for proteins, **all of the reactions necessary for life** are dependent on the function of proteins

Amino acids

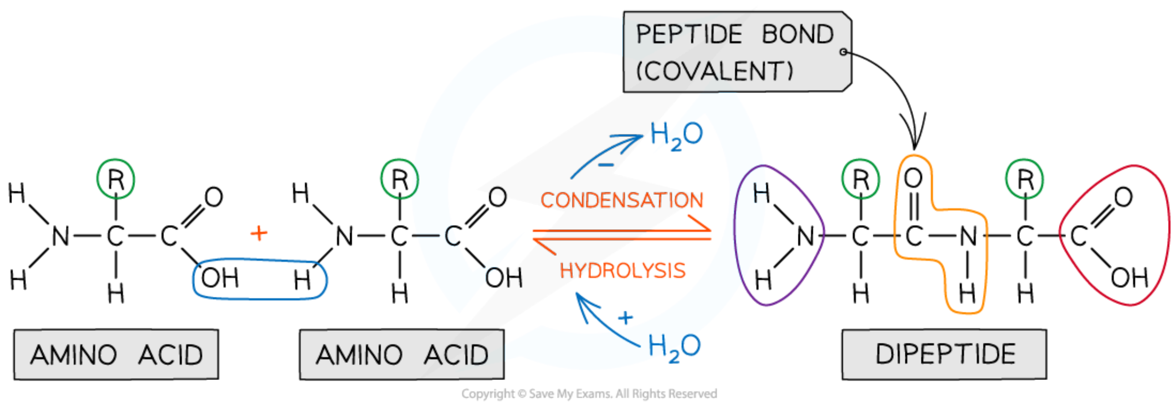
- Amino acids are the **monomers** of polypeptides
- There are **20 amino acids** found in polypeptides common to all living organisms
- The general structure of all amino acids is a central carbon atom bonded to:
  - An **amine** group $-NH_2$
  - A **carboxylic acid** group -COOH
  - A **hydrogen** atom
  - An **R** group (which is how each amino acid differs and why amino acid properties differ e.g. whether they are acidic or basic or whether they are polar or non-polar)
  - The **R** group can be as simple as another hydrogen atom (glycine), right through to complex aromatic ring structures (eg. phenylalanine)



The generalised structure of an amino acid

Peptide bond

- In order to form a **peptide bond** a hydroxyl group (-OH) is lost from the carboxylic group (-COOH) of one amino acid and a hydrogen atom is lost from the amine group (-NH2) of another amino acid
- The remaining carbon atom (with the double-bonded oxygen) from the first amino acid bonds to the nitrogen atom of the second amino acid
- This is a **condensation** reaction so water is released
- **Dipeptides** are formed by the condensation of **two** amino acids
- **Polypeptides** are formed by the condensation of **many** (3 or more) amino acids
- A protein may have only one polypeptide chain or it may have multiple chains interacting with each other
- During **hydrolysis** reactions, the addition of water **breaks the peptide bonds** resulting in polypeptides being broken down into amino acids



Amino acids are bonded together by covalent peptide bonds to form a dipeptide in a condensation reaction

Examiner Tip

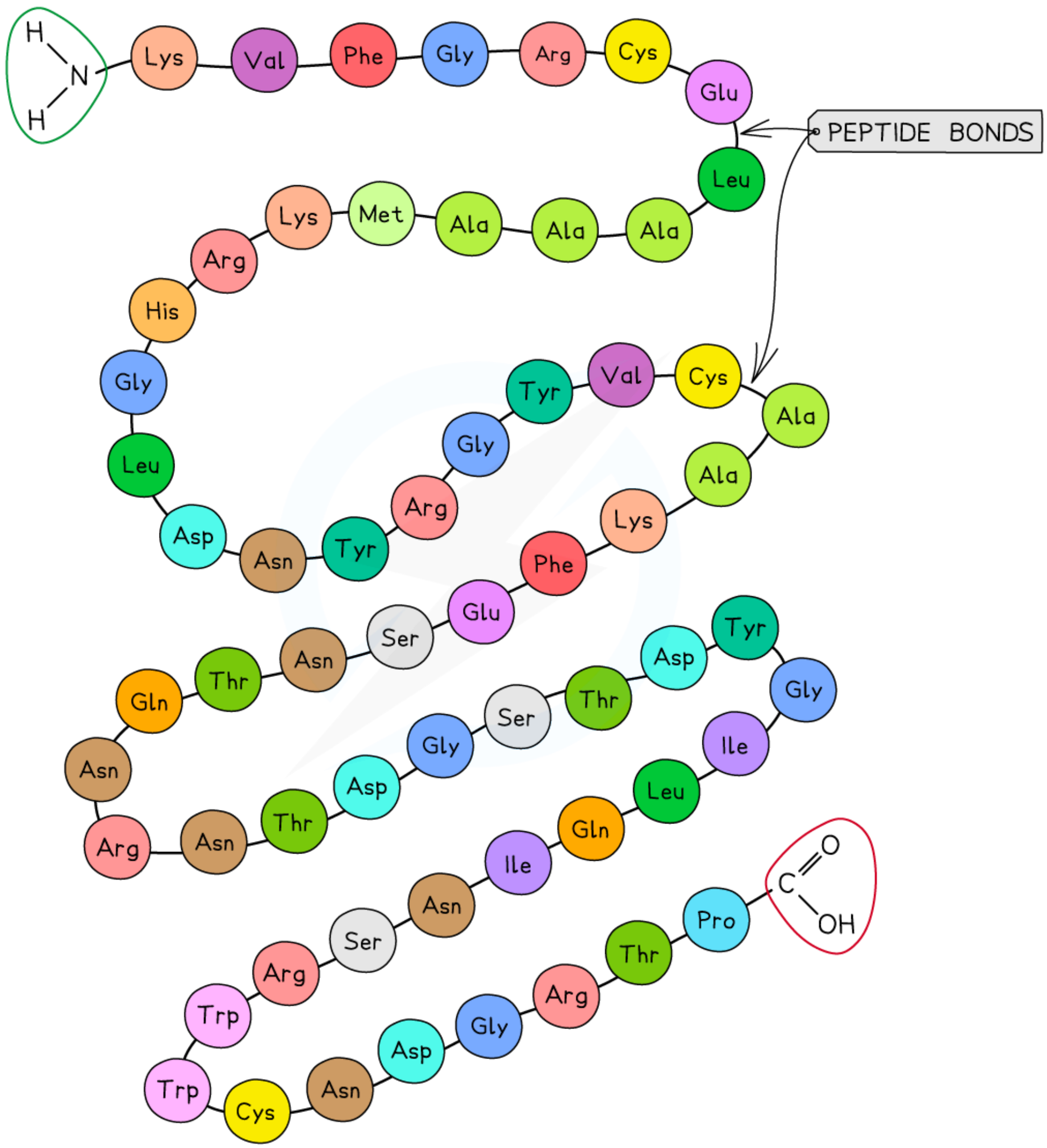
You will be expected to recognise whether an unfamiliar molecule is an amino acid or polypeptide so look for the functional groups (amine and carboxyl). When asked to identify the location of the peptide bond, look for where nitrogen is bonded to a carbon that has a double bond with an oxygen atom, note the R group is not involved in the formation of a peptide bond.

Your notes

Amino Acid Diversity

Your notes

- The same 20 amino acids make up most of the proteins found on Earth
- Around 500 amino acids have been found in nature, but only **20 are commonly found in proteins**
- **Eleven** of these can be naturally synthesised within cells by humans
- The other nine amino acids are **essential** (have to be in the human diet)
- **You don't need to remember the names of the amino acids**, but it's useful to see their names, which are usually **abbreviated to three letters**
  - Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His*, Ile*, Leu*, Lys*, Met*, Phe*, Pro, Ser, Thr*, Trp*, Tyr, Val*
  - * indicates the essential amino acids
- Because the R groups vary so much between the 20 amino acids, there is a **lot of chemical diversity** between the amino acids



An amino acid sequence of a short polypeptide. The three-letter abbreviations indicate the specific amino acid (there are 20 commonly found in cells of living organisms).

NOS: Looking for patterns, trends and discrepancies; most (but not all) organisms build proteins from the same amino acids.

- **All life** except for a few species use these 20 amino acids
- The reason why only 20 amino acids are used has been the subject of a lot of differing hypotheses
  - That only these 20 were available at the origins of life, so have remained ever since, **OR**

Your notes

- That these 20 amino acids are diverse enough to give the wide range of functions that proteins possess, **OR**
- Because of the theory that all organisms share a common ancestor, the link between the genetic code and amino acid sequence is fixed and is not easily altered, even by mutations
- The almost **infinite number of amino acid combinations** make polypeptides suitable to determine **all the characteristics of life**
- Only a few primitive, single-celled organisms use other amino acids
- One unusual amino acid includes the trace element **selenium** and is found in many polypeptides, though at **very low frequencies**
    - A discrepancy is that in some organisms, the stop codon UAG codes for this unusual amino acid containing selenium

- All life goes by the **Central Dogma** that **all genes code for proteins** and the actions of proteins determine all of an organism's characteristics

Polypeptide Diversity

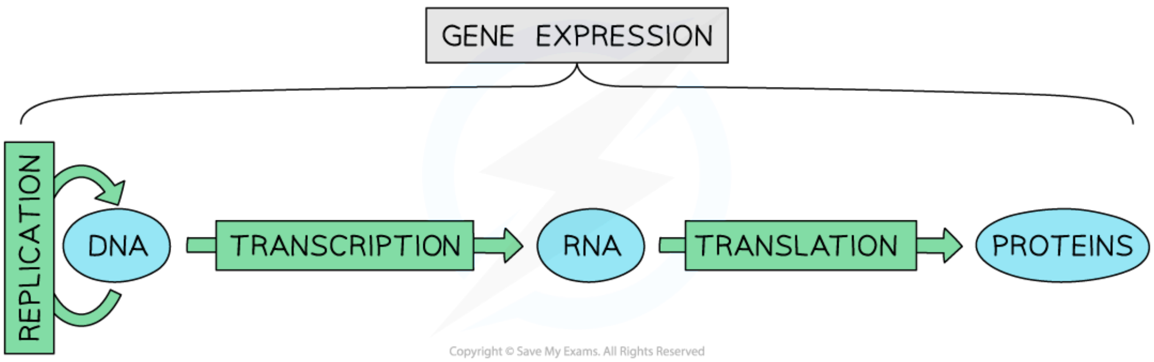
- 20 amino acids can give an almost infinite number of polypeptides
- Polypeptides are assembled at a ribosome by condensing **individual amino acids** onto a growing chain, **one by one**
- This allows a **choice of 20 amino acids** each time one is added
- The **mRNA codon** determines which amino acid is added
- For a polypeptide chain of 50 amino acids in length (considered a very **short protein**), there would be **20^{50}** possible combinations of amino acids
    - This gives 1.13×10^{65} combinations
    - **Standard form** is preferable for showing such a large number, but writing it out in full shows its size, which is
    - 113,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000 combinations!

- Given that the average length of a protein is **300 amino acids**, the number of possible combinations is so large, we can consider it to be **infinite**

Your notes

Genes & Polypeptides

- The amino acid sequence of polypeptides is coded for by genes
- Despite the huge number of amino acid sequences that could be produced, only a **small fraction** of these are produced in nature
- Nevertheless, **many thousands of different polypeptide sequences** are synthesised
- The code for the sequence in which amino acids are joined together is the **genetic code**, held in a sequence of DNA bases in the **genome**
- The **expression** of a gene always results in the production of a polypeptide
- **Three consecutive DNA bases** are required to code for **each amino acid** in a polypeptide



The central dogma of gene expression. All genes code for proteins; proteins carry out the genes' instructions.

2.3.2 Protein Structure & Function

Your notes

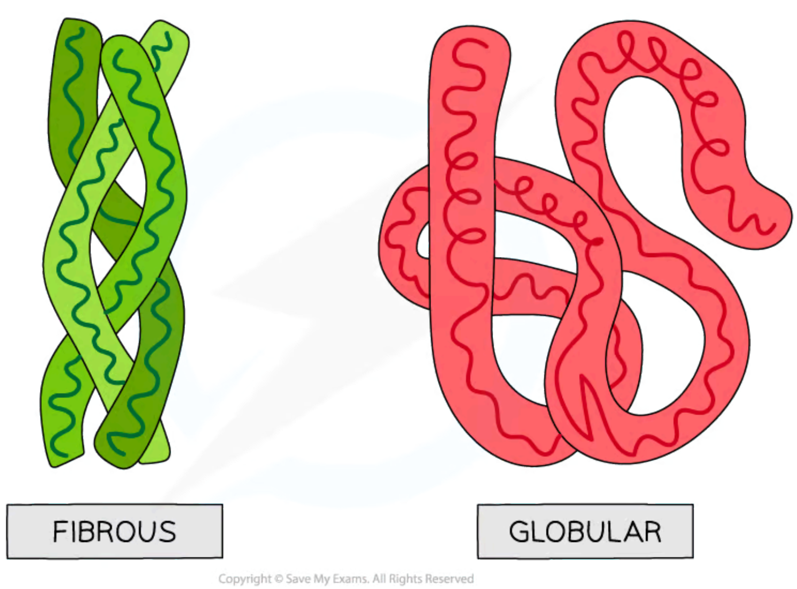
Protein structure

- A protein may consist of a single polypeptide or more than one polypeptide linked together
- Some proteins exist as a **single polypeptide chain** (of amino acids)
- Other proteins are made up of **two or more polypeptide chains** joined together
- **Single polypeptide chain** proteins include **lysozyme**, an enzyme present in mucus secretions and tears, that **kills bacteria** as part of our primary defences against pathogens
- Proteins with **two polypeptide chains** include
  - **insulin**, a hormone responsible for regulating blood glucose levels
  - **integrins**, a group of membrane proteins that span a **phospholipid bilayer** and act as a receptor
    - integrins' two polypeptide chains each have a **hydrophobic section** that sits in the membrane bilayer
- Proteins with **three polypeptide chains** include **collagen**, the main structural protein in skin, tendons, ligaments and the walls of blood vessels
- Proteins with **four polypeptide chains** include **haemoglobin**, which binds oxygen in red blood cells and delivers it from the lungs to respiring tissues
- Each polypeptide chain in a multi-polypeptide protein is referred to as a **subunit** of the protein

3D Structure of Proteins

Your notes

- The amino acid sequence determines the three-dimensional conformation of a protein
- Proteins perform their diverse roles because of their 3-D shape and structure
- This is known as the protein's **conformation**
- The **precise sequence of amino acids** determines how the protein **folds** and aligns itself **as the individual amino acids are being added** at the ribosome
  - Amino acids are always added in the same sequence so a **protein can start to form its shape even** before it is fully formed
  - **Bonds form** between parts of amino acids that can cause a **bridge** to form between one part of the chain and another
    - this creates **loops**, **sheets**, **helices** and **folds**
  - Many of the bonds that hold the protein's shape form between the various **R groups** of individual amino acids
    - If an amino acid is not present in its usual place in the chain due to **mutation**, this can drastically alter the protein's 3-D shape, and affect its function
- Haemoglobin is a **globular** protein (forms a globe-shaped protein)
  - Some of haemoglobin's outer parts are **hydrophilic** to be in contact with water whilst its inner parts are made up of amino acids with **hydrophobic** R groups
- Collagen is a **fibrous** protein (forms a rope-like protein for tensile strength)
  - It has a **repeating sequence of amino acids** to create a helical structure
  - The chain of amino acids **remains in an elongated conformation** to give fibrous strength



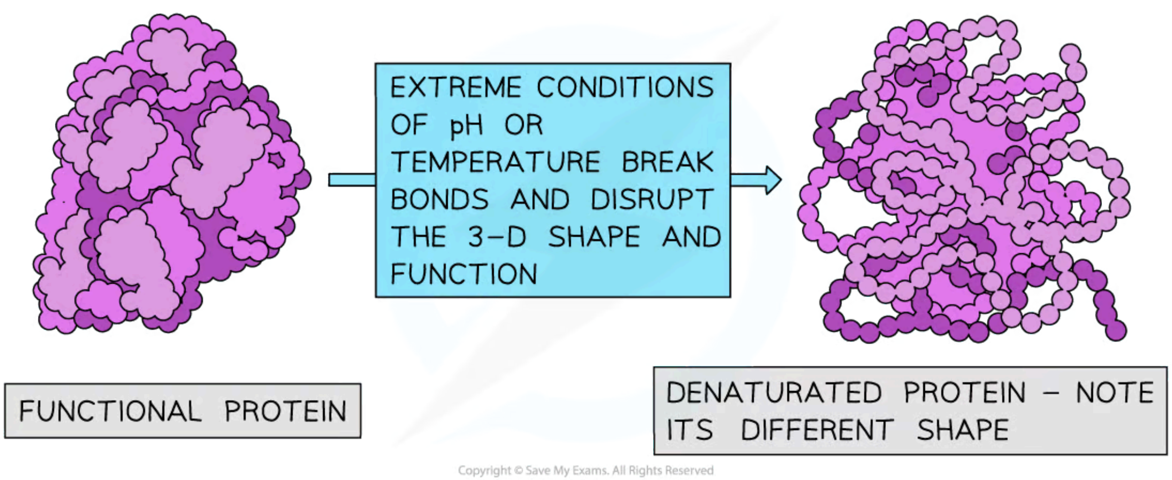
Globular and fibrous protein models illustrating the roughly spherical shape of globular proteins and the long, stranded shape of fibrous proteins

Your notes

Denaturation: Proteins

- Denaturation is the irreversible change of protein conformation caused by temperature and pH extremes
- The bonds that form **between different R groups** are **relatively weak** (compared to the peptide bonds that hold the amino acids in sequence)
- These bonds can be **broken easily,** which can cause the **conformation** of the protein to change
- The **altered protein shape** may affect its **function**, **physical state** and general usefulness in its original role
- This is called **denaturation**
- **Heat** and **extremes of pH** are the most common causes of denaturation
  - Both cause **breaking of the weak bonds** between R groups

- A certain pH is considered as an **optimum** for a particular protein, because at that pH, the protein's 3-D structure is not denatured
- Denaturation is almost always **irreversible**
  - The protein **cannot be re-formed** in its original conformation by reversing the change in conditions
  - However, **small denaturations** and **renaturations** are possible in certain proteins to respond to small fluctuations in pH eg. haemoglobin



The effect of heat and pH on the shape and function of a globular protein

Denaturation in action

- Denaturation can be seen most easily by looking at the **changes in an egg white** as the egg is fried or poached
- Egg white is mainly the protein **albumin**
- The **hydrophobic amino acids** in albumin are at the centre of the molecule in its normal state, so albumin is soluble
- Heating causes the hydrophobic amino acids to appear **at the edges**, where they cause the protein to become **insoluble**

- A harder, solid layer forms, which is the **cooked white**
- Similar events occur in the proteins of the **egg yolk** as it cooks
- Denaturation also occurs in the **stomach**, where the low pH (pH2) causes **proteins in the diet to become denatured** on their way to being fully hydrolysed further down the digestive system
- The stomach enzyme **pepsin**, a protein-digesting enzyme has an optimum pH of 2 for this reason
- Certain **extremophiles** have evolved to have proteins that are stable even at extreme pH or temperature
  - eg. *Thermus aquaticus*, a **bacteria that lives in hot springs** at 80°C
  - This temperature would denature most other proteins

- **Denaturation of enzymes** can be used as part of experiments to measure enzyme activity
  - For example, **an experiment to establish the optimum pH or temperature** of an enzyme eg. pepsin or lipase

- Many drugs are proteins that **cannot be taken by mouth**, because the protein will be **denatured by stomach acid**
  - These drugs should be **delivered in another way** eg. by **direct injection** into the blood

Examiner Tip

Remember to avoid confusing the bonds that hold a protein's shape together with the peptide bonds that attach each amino acid in sequence. Picture the peptide bonds holding the amino acids in a straight chain, then the other bonds holding the chain in its folded, 3-D structure.

Your notes

2.3.3 The Variety of Proteins

Your notes

Functions of Proteins

- Living organisms synthesize many different proteins with a wide range of functions
- Proteins are so **versatile** that they have many different roles in cells, tissues and organs
- All of the following functions are performed by proteins:
  - Speeding up cellular reactions, or **catalysis**, is performed by **enzymes**
  - **Blood clotting**, where blood proteins interact with oxygen to form a gel-like scab across a wound
  - **Strengthening** fibres in skin, hair, tendons, blood vessels eg. **collagen, keratin**
  - **Transport** of vital metabolites eg. oxygen which is carried by **haemoglobin**
  - Formation of the **cytoskeleton**, a network of tubules within a cell that cause chromosomes to move during the cell cycle
  - **Cell adhesion**, where cells in the same tissue stick together
  - **Hormones**, chemical messengers that are secreted in one part of the body to have an effect elsewhere
  - **Compaction of DNA** in chromosomes for storage, caused by **histone** proteins
  - The immune response produces **antibodies**, the most diverse group of proteins
  - Membrane transport **channel and carrier proteins** that determine which substances can pass across a membrane
  - **Cell receptors**, which are binding sites for hormones, chemical stimuli such as tastes, and for other stimuli such as light and sound

Examiner Tip

Many exam questions focus on enzymes but don't forget all the other types of protein when discussing protein functions.

Your notes

Examples of Proteins

Rubisco

- **R**ib**u**lose **Bis**phosphate **C**arb**o**xylase
- An enzyme that catalyses the **fixing of CO_2 from the atmosphere** during photosynthesis
- Composed of **16 polypeptide chains** as a **globular** protein
- This is **the source of all organic carbon**, so Rubisco is arguably the most important enzyme in Nature!
- The **most abundant enzyme on Earth** as it's present in all leaves
- Rubisco is **a very slow catalyst**, but it's the most effective to have evolved so far to fulfil this vital function

Insulin

- A **hormone** produced and secreted by β-cells in the **pancreas**
- Binds to insulin receptors (on liver, fat and muscle cells) reversibly, causing **absorption of glucose from the blood**
- Composed of **2 polypeptide chains** as a **short, globular protein**

Immunoglobulins

- Also known as **antibodies**
- They have a **generic 'Y' shape**, with specific binding sites at the two tips of the 'Y'
- They bind to specific **antigens**
- The binding areas of immunoglobulins are **highly variable**, meaning that antibodies can be produced **against millions of different antigens**
- Immunoglobulins (as the name suggests) are **globular** and are the **most diverse range of proteins**

Rhodopsin

- A **pigment in the retina** of the eye
- A **membrane protein** that is expressed in rod cells
- Contains a light-sensitive part, **retinal**, which is derived from **Vitamin A**
- A photon of **light causes a conformational change** in rhodopsin, which sends a nerve impulse along the optic nerve to the **central nervous system**

Collagen

- A **fibrous protein** made of **three separate polypeptide chains**
- The **most abundant protein in the human body** - approx 25%
- Fibres **form a network** in skin, blood vessel walls and connective tissue that can **resist tearing forces**
- Plays a role in **teeth** and **bones**, helping to **reduce their brittleness**

Spider Silk

- The silk used by spiders to suspend themselves and create the spokes of their webs is as **strong as steel wire** though considerably lighter
- Contains **rope-like, fibrous parts** but also **coiled parts** that stretch when under tension, helping to **cause extension** and **resist breaking**

Your notes

- Does not denature easily at extremes of temperature
- Has many attractive aspects for **engineering** and **textile product design** thanks to its **strength** and **low weight**
- Can be **genetically engineered** to be **expressed in goats' milk** as spiders can't be farmed on a large enough scale
- Other kinds of spider silk protein are **tougher** though lack the tensile strength, eg. the silk they use to encase their prey after capture

Proteome

- The **proteome** is the full range of **proteins** that a cell or organism is able to produce
- By contrast, a **genome** is the complete set of **genes** present in a cell/organism
  - The full genome is present within every cell of an organism, but not every gene is **expressed** in every cell. Which genes are expressed, **depends on the cell type**
- The proteome is usually **larger** than the genome of an organism
- Every individual has a **different proteome**
  - Because of small differences in the amino acid sequence of proteins
- The **proteome varies during an organism's lifetime** as certain proteins are not needed throughout the organism's life
    - An example is **fetal haemoglobin**. The gene for that protein is not expressed after the baby is around 3 months old, as the baby expresses **adult forms of haemoglobin**, which are encoded by **separate genes**
  - This is also due to a large amount of **modification of proteins** that can take place after synthesis (often in the **Golgi** apparatus)
    - For example, adding a carbohydrate part to form **glycoproteins**, which are important in cell signalling
- **Splicing** of RNA during transcription can allow one gene to code for many proteins

Examiner Tip

You don't need to know the details of splicing for Standard Level but it accounts for several proteins being produced from just one gene. Even though a lot of genes do not code for proteins, the proteome is larger than the genome because of the sheer range of proteins that can be produced from the DNA code.

2.3.4 Skills: Molecules

Your notes

Molecular Diagrams: Drawing

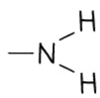
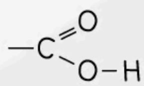
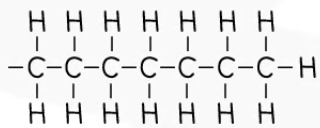
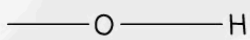
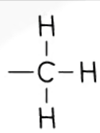
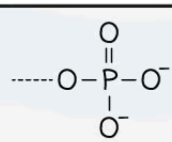
Drawing biological molecules

- It is important to be able to **draw a few key molecules**
  - There is a huge variety of biological molecules, but **only the most important ones** are required

- **Element symbols** from the **Periodic Table** are used
- A **short, straight line** is used for a covalent bond, with **two lines** for a double bond
- Some chemical groups may be denoted by a **symbol** such as Ⓟ for a phosphate group
- The symbol **R** represents a **variable chemical group**, such as the variable side groups of amino acids
- An exam question may require you to **draw various molecules**
- It's advisable to break the task down **into stages**

Symbols Used in Biological Molecule Drawings Table

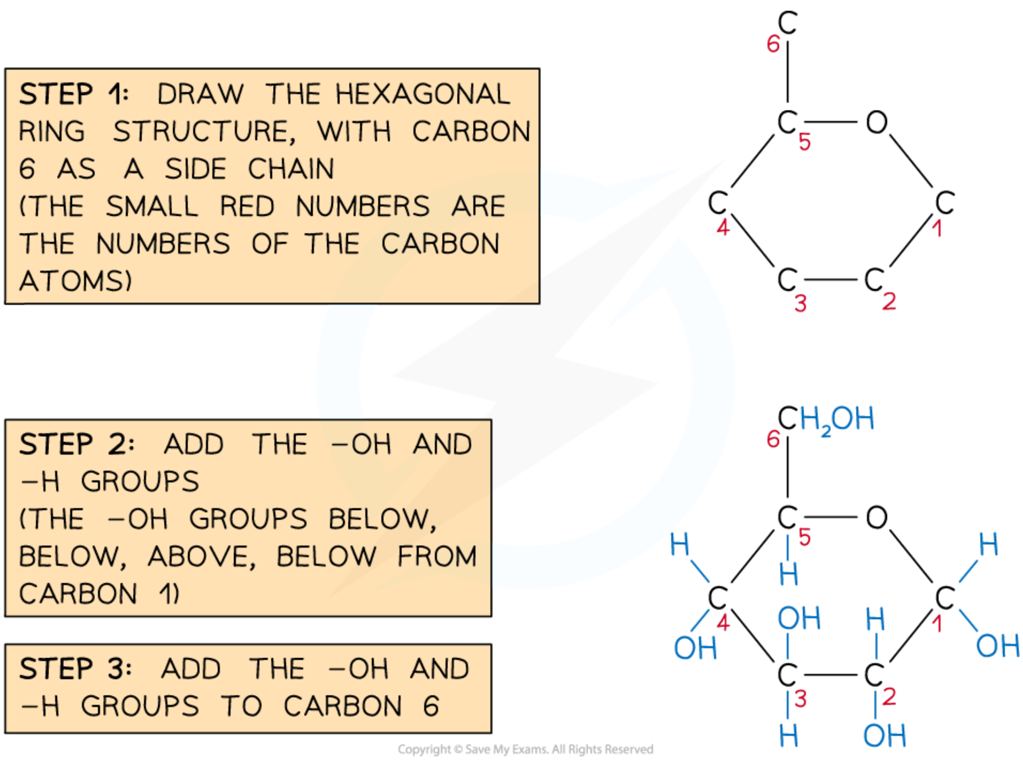
Your notes

Name of Group	Group Structure	Abbreviated Notation
Amine		$-NH_2$ or H_2N-
Carboxylic Acid		$-COOH$ or $HOOC-$
Hydrocarbon Chain		
Hydroxyl		$-OH$ or $HO-$
Methyl		$-CH_3$ or H_3C-
Phosphate		Ⓟ

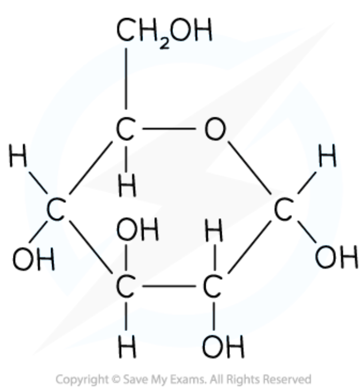
Drawing α-D-glucose

- Aspects to remember
  - Glucose has the formula $C_6H_{12}O_6$
  - In solid form, glucose has a linear structure
  - It forms a **hexagonal** ring in an aqueous solution
    - As aqueous glucose is **the only state** that glucose exists in biology, it's the **ring structure** that should be learned
  - **One of the corners** of the ring (draw this in the top-right) is occupied by an **oxygen** atom
  - The 6th carbon occupies a **side chain** (top-left)
  - The **carbon atoms are numbered 1 to 6** starting on the right and working clockwise
  - The hydroxyl groups occupy positions above or below the ring as follows
    - Carbon atom 1 - below
    - Carbon atom 2 - below

    - Carbon atom 3 - above
    - Carbon atom 4 - below
- You can **ignore the 'D'** in the names alpha-D-glucose or beta-D-glucose
    - The only other version is L-glucose which plays **no significant role in biology**



Recommended steps to draw a molecule of α-D-glucose



Structure of α-D-glucose

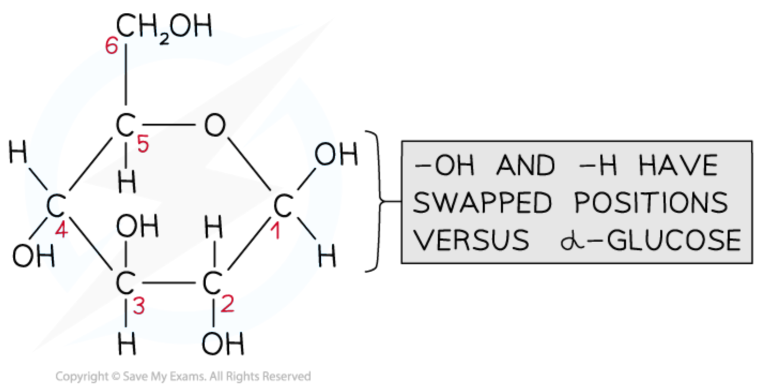
β-D-glucose is very slightly different in structure

Your notes

Your notes

- Beta-glucose (β-glucose) has **a small, subtle difference** to α-glucose
- The hydroxyl group on carbon atom 1 **sits ABOVE the ring**, rather than below
- This sugar is the **monomer** of **cellulose**
- This example of two different **isomers** changes the properties of the polysaccharide formed from these monomers drastically
  - It accounts for all the many differences between starch and cellulose

- The hydroxyl groups occupy positions above or below the ring as follows
  - Carbon atom 1 - **above**
  - Carbon atom 2 - below
  - Carbon atom 3 - above
  - Carbon atom 4 - below



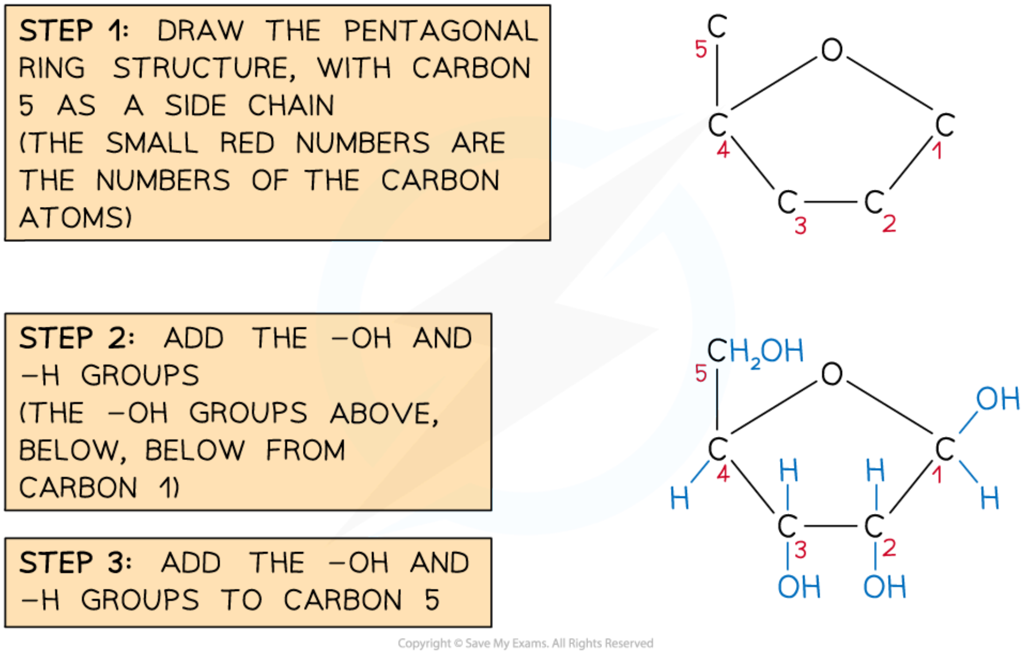
Structure of β-D-glucose

Drawing a ribose sugar

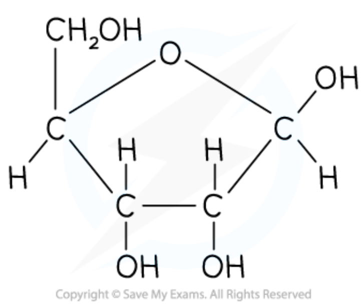
- This **family of sugars** play a role in DNA and RNA structure, as well as ATP
- Ribose is a form of **pentose** sugar (5 carbon atoms)
- Like glucose, ribose has a **ring** structure
- Aspects to remember
  - Ribose has the formula $\mathbf{C_5H_{10}O_5}$
  - It forms a **pentagonal ring** in an aqueous solution
  - **One of the corners** of the ring (draw this in the top) is occupied by an **oxygen** atom
  - The 5th carbon occupies a **side chain** (top-left)
  - The **carbon atoms are numbered 1 to 5** starting on the right and working clockwise
  - The hydroxyl groups occupy positions above or below the ring as follows
    - Carbon atom 1 - above
    - Carbon atom 2 - below
    - Carbon atom 3 - below

- Ribose sugars have an important close relative - **deoxyribose** sugars
  - Both are key components of **RNA** and **DNA** respectively
  - The 'R' and 'D' of **R**NA and **D**NA comes from the sugar in the structure, ribose or deoxyribose

Your notes



Recommended steps to draw a molecule of ribose

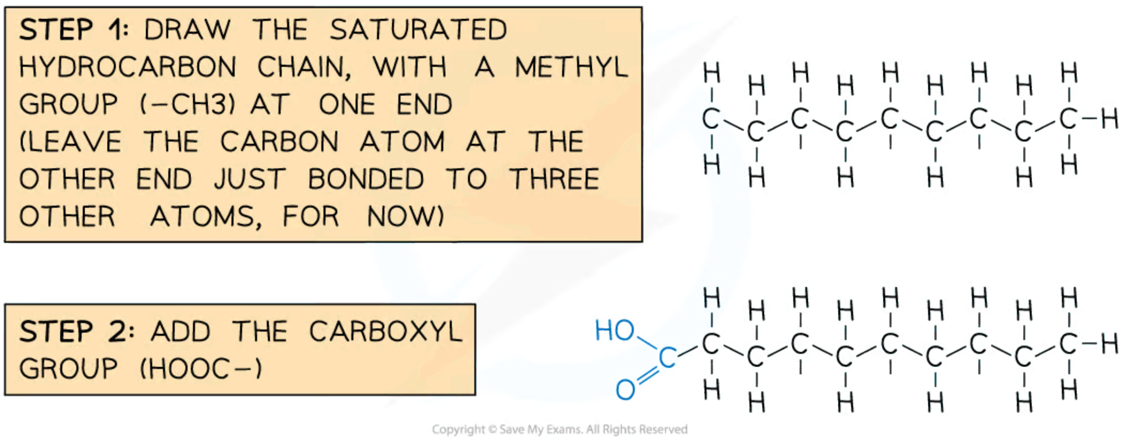


Structure of ribose

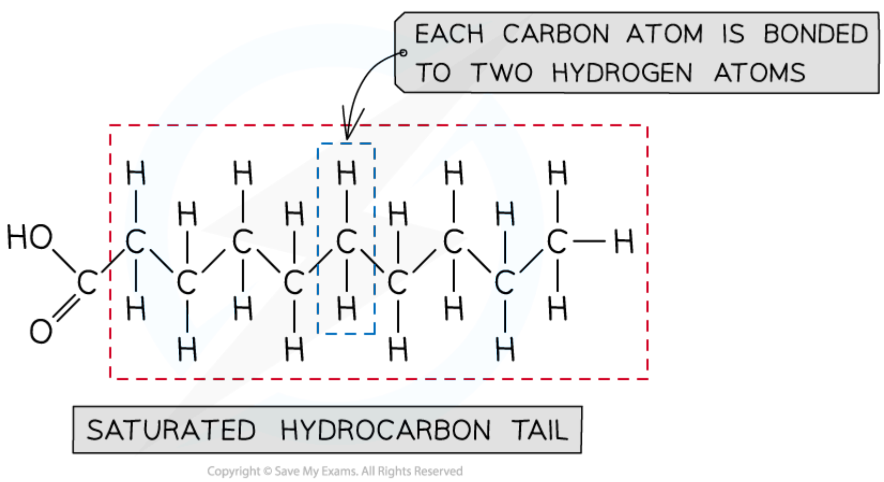
Drawing a saturated fatty acid

- There are two aspects to a saturated fatty acid
  - A **saturated** hydrocarbon chain
    - Contains **only C–C single bonds**
    - Each internal carbon atom is bonded to **2 hydrogen atoms**
  - A **carboxylic acid group** at one end
- You don't need to memorise any names of saturated fatty acids
  - The number of carbon atoms in your chain is also not important, but greater than around 8 is advised.

Your notes



Recommended steps to draw a molecule of a saturated fatty acid

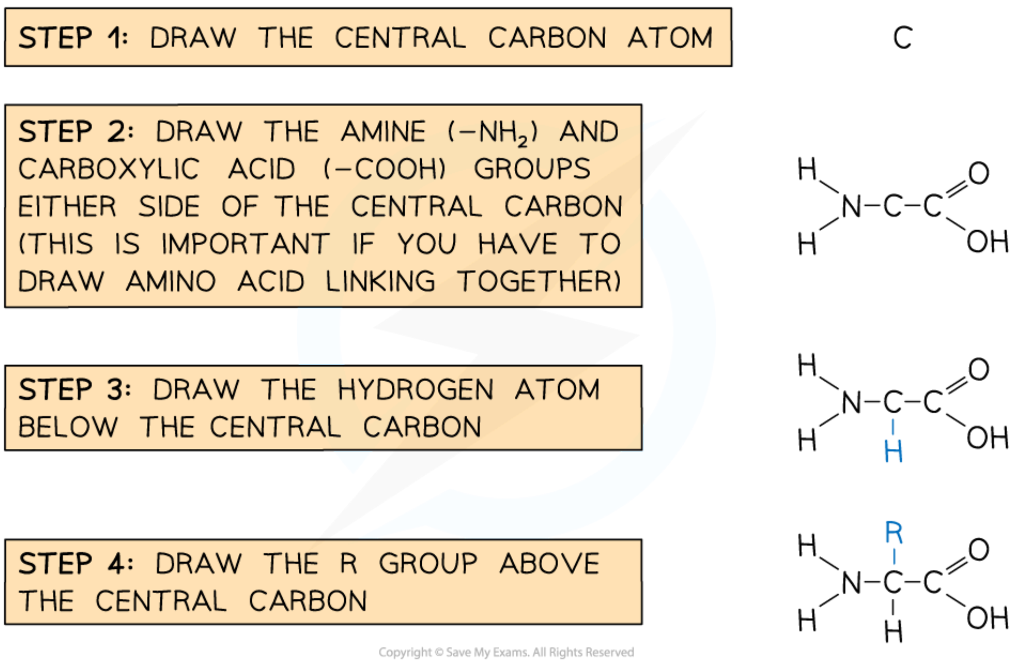


A saturated fatty acid

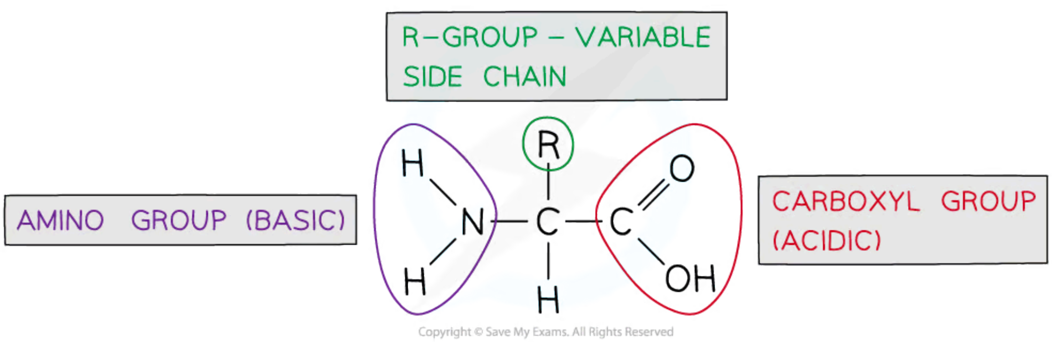
Drawing a generalised amino acid

- Each amino acid has **central carbon atom**
- Three of the bonds from the central carbon atom are occupied as follows
  - a **hydrogen atom**
  - a **carboxylic acid** group
  - an **amine** group
- The fourth bond attaches the central carbon to the **R group**
  - The R group is **variable** and determines the **identity of the amino acid**
  - You won't need to remember any of the R groups or amino acid names
- Drawing the 4 groups surrounding the central carbon in a **flat structure** is acceptable, although the real arrangement of bonds around a carbon atom is in a **tetrahedral shape**

Your notes



Recommended steps to draw a molecule of generalised amino acid



The generalised structure of an amino acid

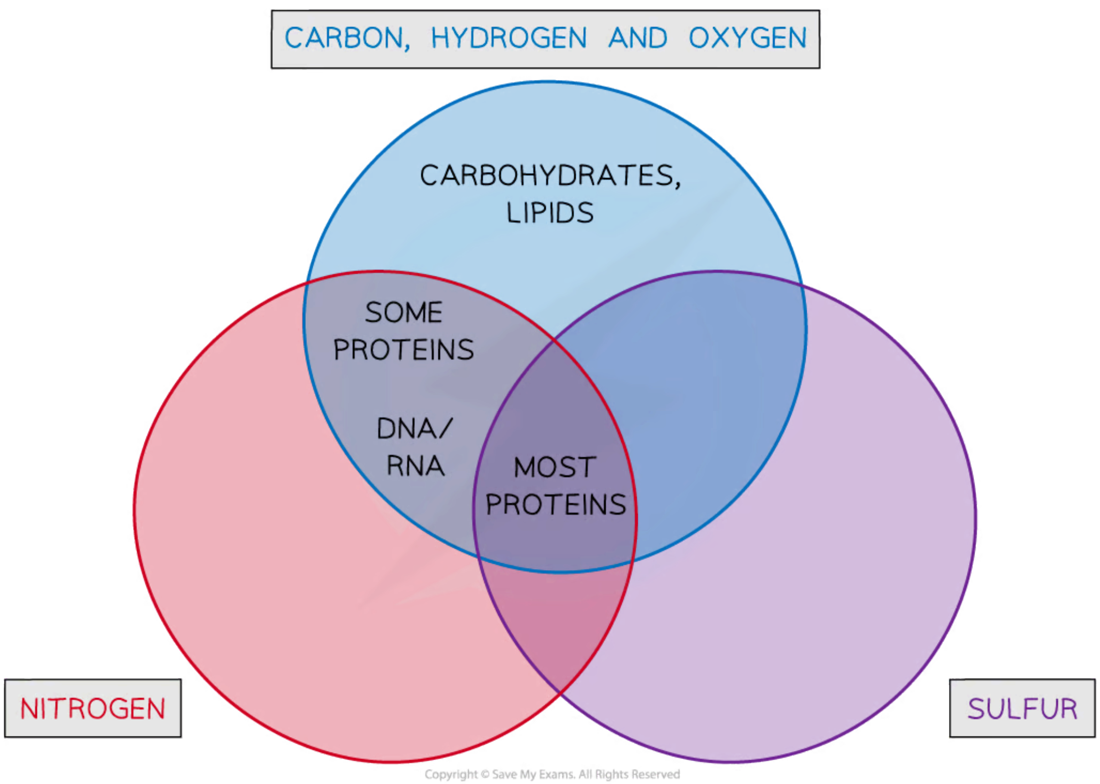
Examiner Tip

The specification is very clear about what you should remember and equally importantly, what you **don't** have to memorise in this topic.With all these drawings, practise with a pencil and plenty of paper until you get it right. It WILL stick! Once you've memorised it, 'draw' questions should be easy marks!

Molecular Diagrams: Identification

Your notes

- As well as being able to draw certain molecules, an important skill is being able to **recognise certain biochemicals** from molecular diagrams
- There are **several features** that help to identify molecules
- The presence of **carbon, hydrogen, oxygen, nitrogen, sulfur** and **phosphorus** can help in the identification
- All biological macromolecules contain **carbon**, **hydrogen** and **oxygen**
- **Nitrogen** and **sulfur** are present in **proteins**
- **Nitrogen** is present in nucleic acids (DNA, RNA)
- **Phosphorus** is also present in certain molecules (DNA, RNA and phospholipids)
- The presence of **ring structures**, **hydrocarbon chains**, **carbon-to-carbon double bonds**, **double-stranded** areas and the **ratio of carbon to oxygen** in a molecule all give clues about the molecule's identity



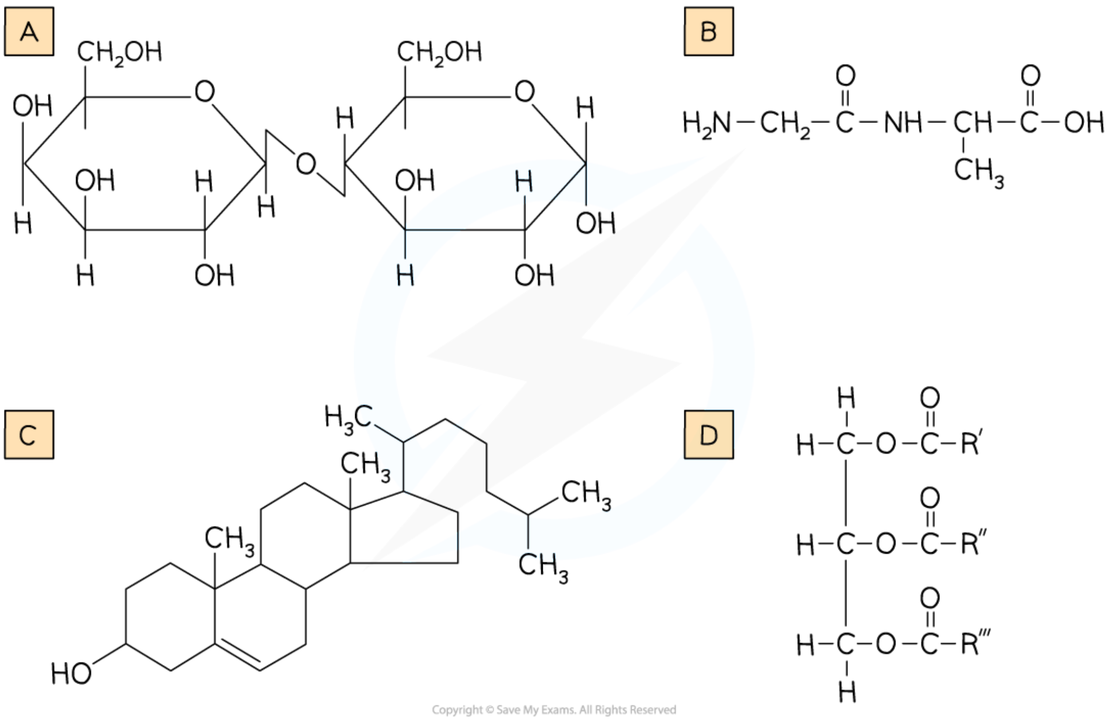
Using the Presence of Various Atoms to Identify Biochemicals

- **Glycosidic** bonds, **ester** bonds and **peptide** bonds all have a distinctive appearance in molecular drawings and will immediately **identify carbohydrates**, **lipids** and **proteins** respectively

Your notes

Worked example

Identify the Diagram (A, B, C or D) which Shows a Triglyceride Structure



Step 1: Look at the elements present in all the diagrams

We're looking for a **triglyceride**, a type of lipid. Lipids contain C, H and O only. **B** also contains nitrogen (**B** is a dipeptide)

Eliminate Answer **B**

Step 2: Look for lipid structures

A contains ring structures, so is likely a carbohydrate (**A** is a disaccharide)

Eliminate Answer **A**

Step 3: Look for three hydrocarbon chains

D contains three hydrocarbon chains (attached by ester bonds to glycerol). **C** is cholesterol, which *is* a lipid, but not a triglyceride lipid.

Select Answer D

The ratio of hydrogen to oxygen

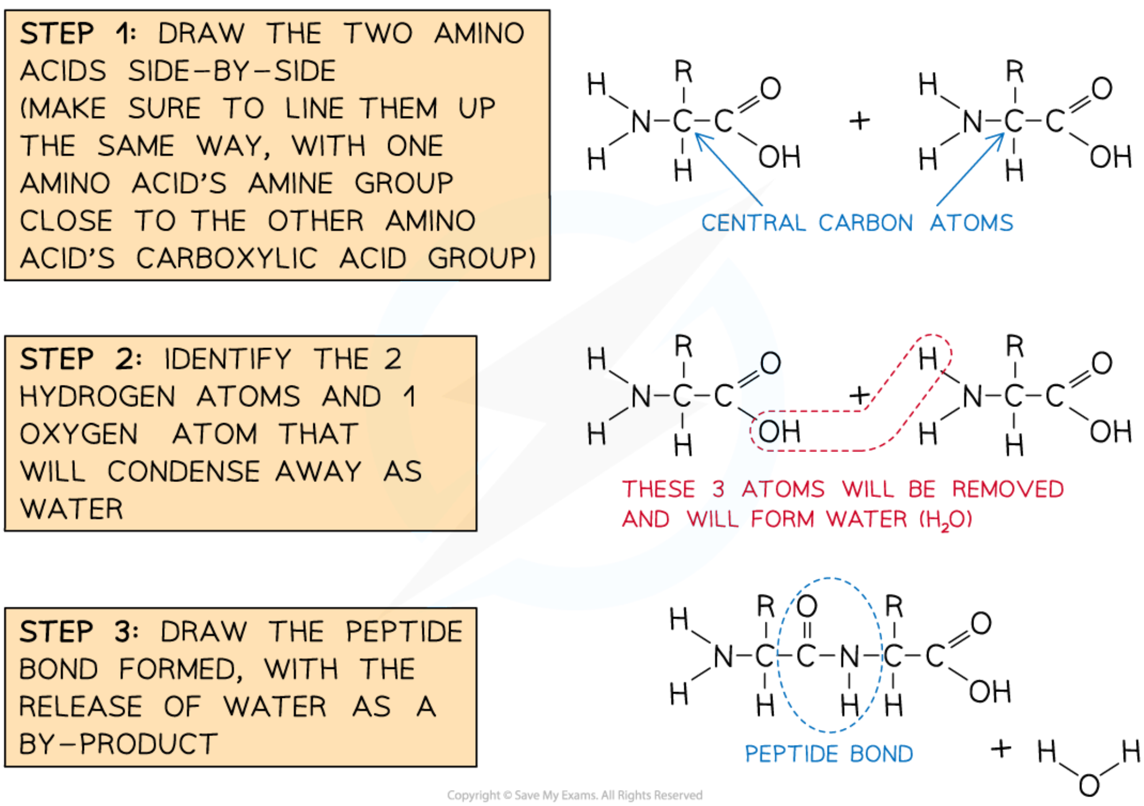
- The **numbers of hydrogen and oxygen atoms in a molecule** can help to identify it
- Carbohydrates contain **hydrogen and oxygen** in a **2:1 ratio**
  - Think of water, formula H_2O, and where the '-**hydrates**' part of the word 'carbo**hydrates**' comes from
- Lipids **contain a much lower proportion of oxygen** than carbohydrates eg. $C_{18}H_{34}O_2$, where the hydrogen to oxygen ratio is 17:1

Your notes

Molecular Diagrams: Peptide Bond Formation

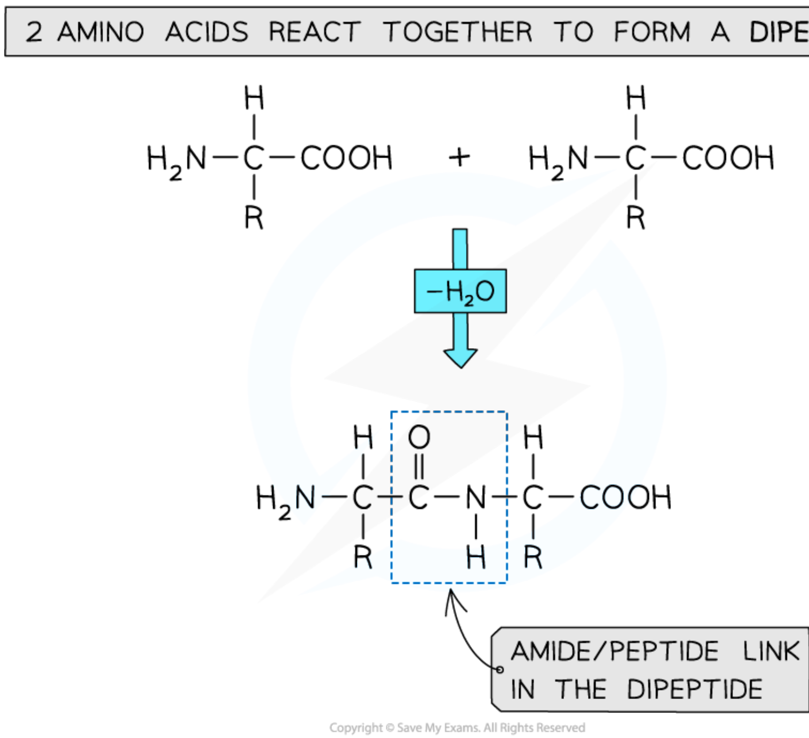
Your notes

- Having learned to draw the structure of a **generalised amino acid**, two or more of these can be joined together to **show how peptide bonds form** during protein synthesis
- Amino acid monomers link together via a **condensation** reaction
  - This releases a molecule of water (H_2O)
  - One **H** atom (of the released water) comes from **one amino acid's amine** group
  - The other **H** atom and an **O** atom come from the **other amino acid's carboxylic acid** group
- This knowledge can be **useful when drawing how two amino acids condense** to form a peptide bond



The recommended steps in drawing a peptide bond formation

Your notes



Formation of a dipeptide