Exam Papers Practice

Proteins

IB Biology - Revision Notes

www.exampaperspractice.co.uk

Formation of Proteins

Amino Acid Structure

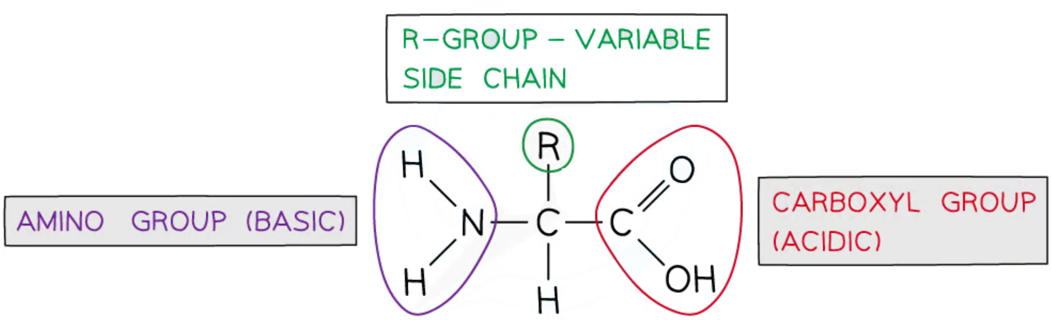
Proteins

- 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 (e.g. carrier)
  - **Hormones**
  - Immunoproteins (e.g. immunoglobulins)
  - **Transport** proteins (e.g. haemoglobin)
  - **Structural** proteins (e.g. keratin, collagen)
  - **Contractile** proteins (e.g. 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, the alpha carbon, bonded to:
  - An **amine**/amino group $-NH_2$
  - A carboxylic acid/**carboxyl** 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 (e.g. phenylalanine)

Structure of an amino acid diagram



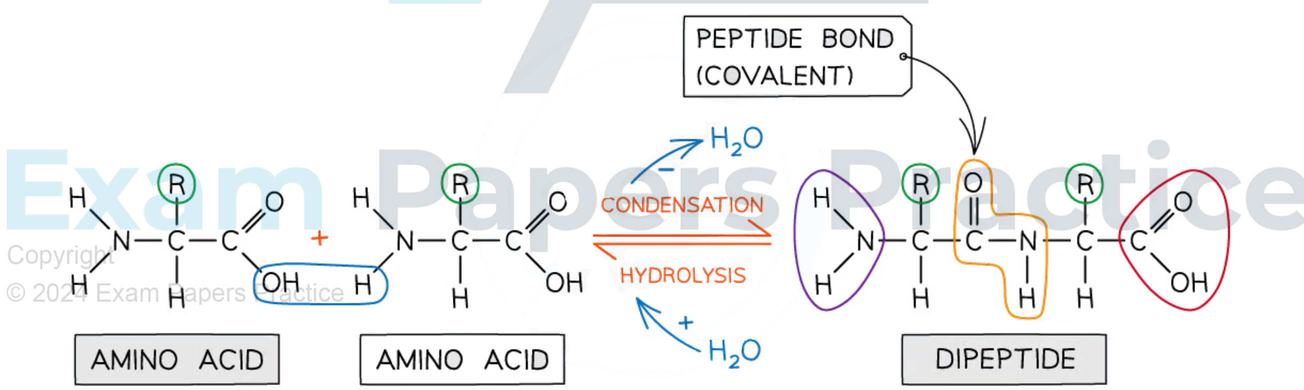
The generalised structure of an amino acid

The Peptide Bond

Peptide bond

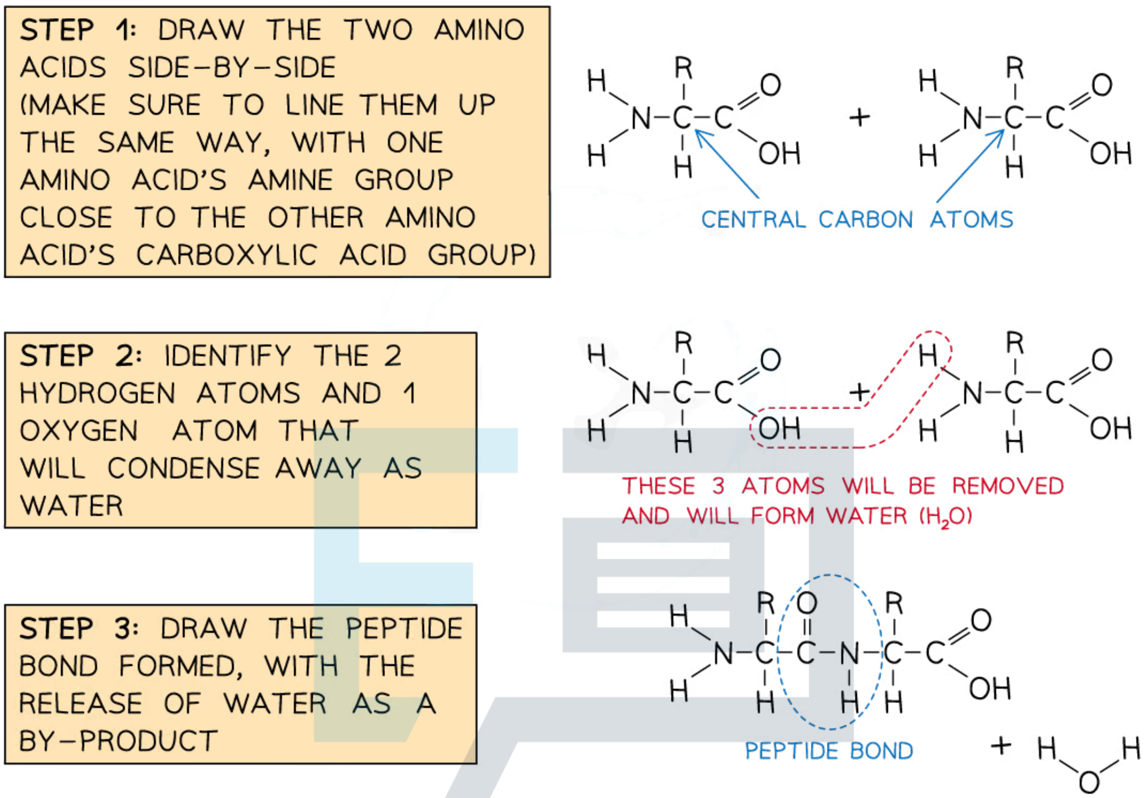
- In order to form a **peptide bond**, a hydroxyl group (-OH) is lost from the carboxylic group (-COOH) of one amino acid and a hydrogen atom is lost from the amine group (-NH2) of another amino acid
- The remaining carbon atom (with the double-bonded oxygen) from the first amino acid bonds to the nitrogen 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
  - The word equation for this reaction is **amino acid + amino acid → dipeptide**
- **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
- **Molecular modelling** kits can be used to build physical models that demonstrate peptide bond formation between different types of amino acids

Peptide bond diagram



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

Drawing a peptide bond diagram



These steps can be followed to draw a peptide bond and a generalised dipeptide

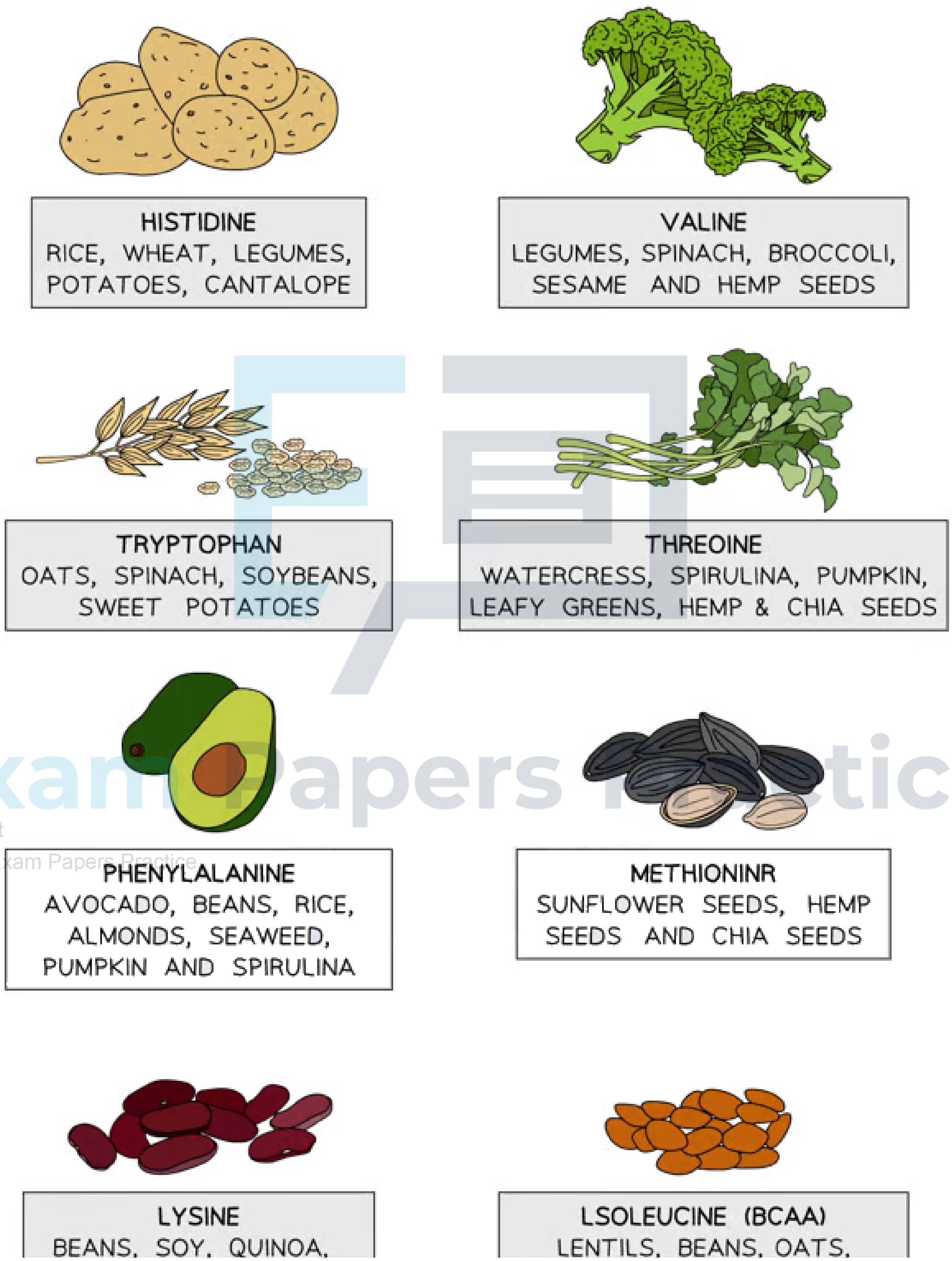
Exam Tip

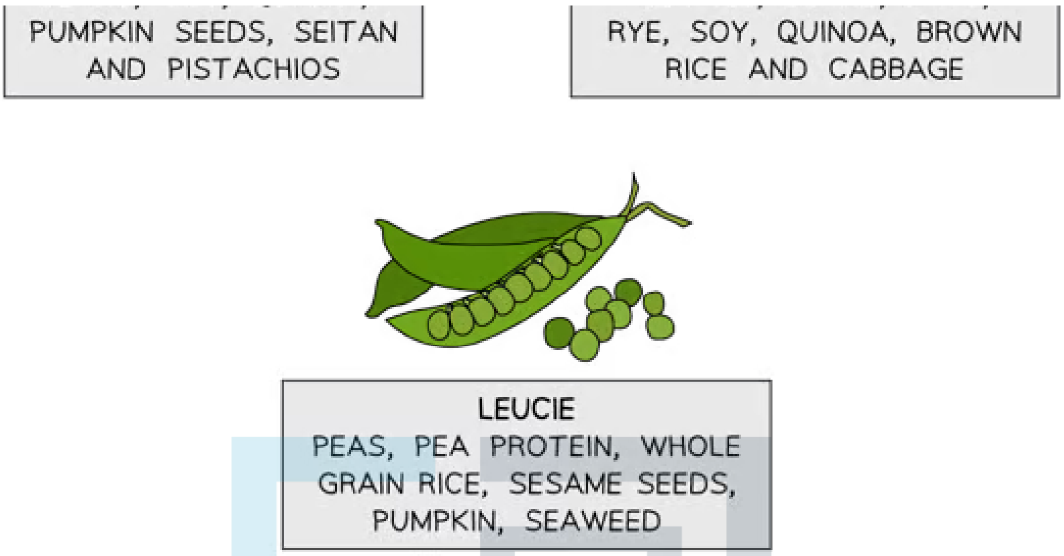
You will be expected to recognise whether an unfamiliar molecule is an amino acid or polypeptide so look for the functional groups (amine and carboxyl). When asked to identify the location of the peptide bond, look for where nitrogen is bonded to a carbon that has a double bond with an oxygen atom, note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

Amino Acids: Dietary Requirements

- There are 20 naturally occurring amino acids
- Our cells can synthesise 11 of these from other amnio acids
  - These are termed **non-essential amino acids**
- The remaining nine we need to consume via our diets
  - These are called **essential amino acids**
- A healthy, varied, well balanced diet will contain all the nine essential amino acids required
- Diets that restrict certain foods may require supplementation
  - Meat contains all nine essential amino acids so a vegetarian or vegan diet needs to be well balanced and varied to ensure all essential amino acids are consumed regularly

Essential amino acid sources diagram


HISTIDINE
RICE, WHEAT, LEGUMES, POTATOES, CANTALOPE
VALINE
LEGUMES, SPINACH, BROCCOLI, SESAME AND HEMP SEEDS
TRYPTOPHAN
OATS, SPINACH, SOYBEANS, SWEET POTATOES
THREOINE
WATERCRESS, SPIRULINA, PUMPKIN, LEAFY GREENS, HEMP & CHIA SEEDS
PHENYLALANINE
AVOCADO, BEANS, RICE, ALMONDS, SEAWEED, PUMPKIN AND SPIRULINA
METHIONINR
SUNFLOWER SEEDS, HEMP SEEDS AND CHIA SEEDS
LYSINE
BEANS, SOY, QUINOA,
LSOLEUCINE (BCAA)
LENTILS, BEANS, OATS,



Plant based (vegan) sources of the nine essential amino acids

Exam Tip

You are not required to remember or give examples of non-essential and essential amino acids.

The Variety of Proteins

Peptide Chain Diversity

- There is a large variety of proteins available to living organisms
- This is because:
  - There are 20 naturally occurring amino acids that form the basic structure of a polypeptide chain
  - Polypeptides can vary in length from a few to thousands
  - The structure and amino acid sequence can also vary
  - The genetic code, meaning DNA base sequence, codes for the number and order of amino acids in a polypeptide, and there is a huge variety of options for DNA base sequence
- 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
- 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**

Role of proteins

- The range of proteins available means that they are very **versatile** so that they have many different roles in cells, tissues and organs, such as:
  - 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 e.g. **collagen, keratin**
  - **Transport** of vital metabolites e.g. 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

Examples of polypeptides

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** - approximately 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

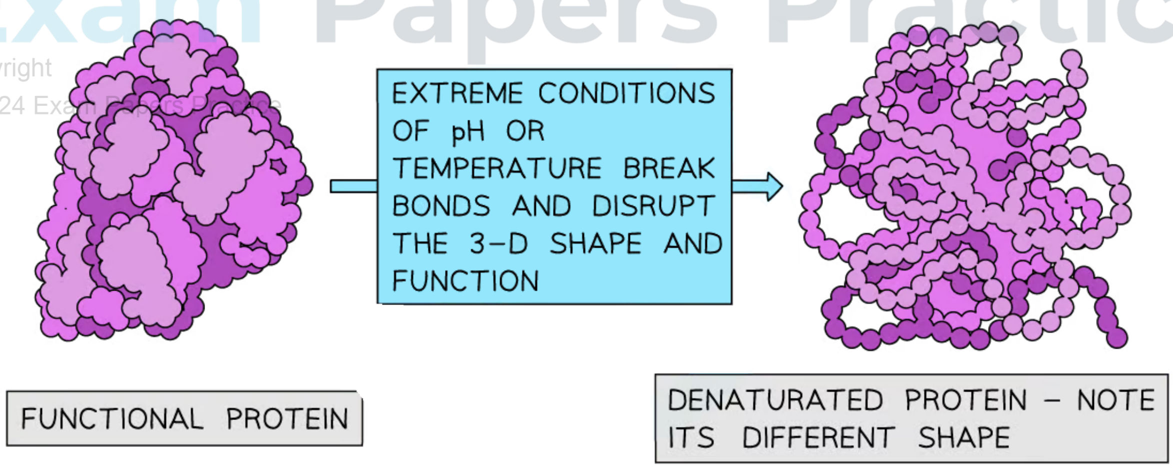
- The silk used by spiders to suspend themselves and create the spokes of their webs is as **strong as steel wire** though considerably lighter
- Contains **rope-like, fibrous parts** but also **coiled parts** that stretch when under tension, helping to **cause extension** and **resist breaking**
- Does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

Protein Structure: Effect of pH & Temperature

Protein Structure: Effect of pH & Temperature

- Proteins structure is sensitive to changes in the environment, particularly **temperature** and **pH changes**
- The **precise structure** of a protein is dependent on the ionic interactions, hydrogen bonds and other intermolecular forces between polypeptide chains being intact
- **Denaturation** may occur by temperature and pH extremes that interfere with these bonds
  - Denaturation is the irreversible change of protein conformation
- 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 and denaturation
- The **altered protein shape** may affect its **function**, **physical state** and general usefulness in its original role
- A certain pH is considered as an optimum for a particular protein, because at that pH, the protein's 3D 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 e.g. haemoglobin

Denaturation of a protein diagram



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 e.g. 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** e.g. by **direct injection** into the blood

Exam Tip

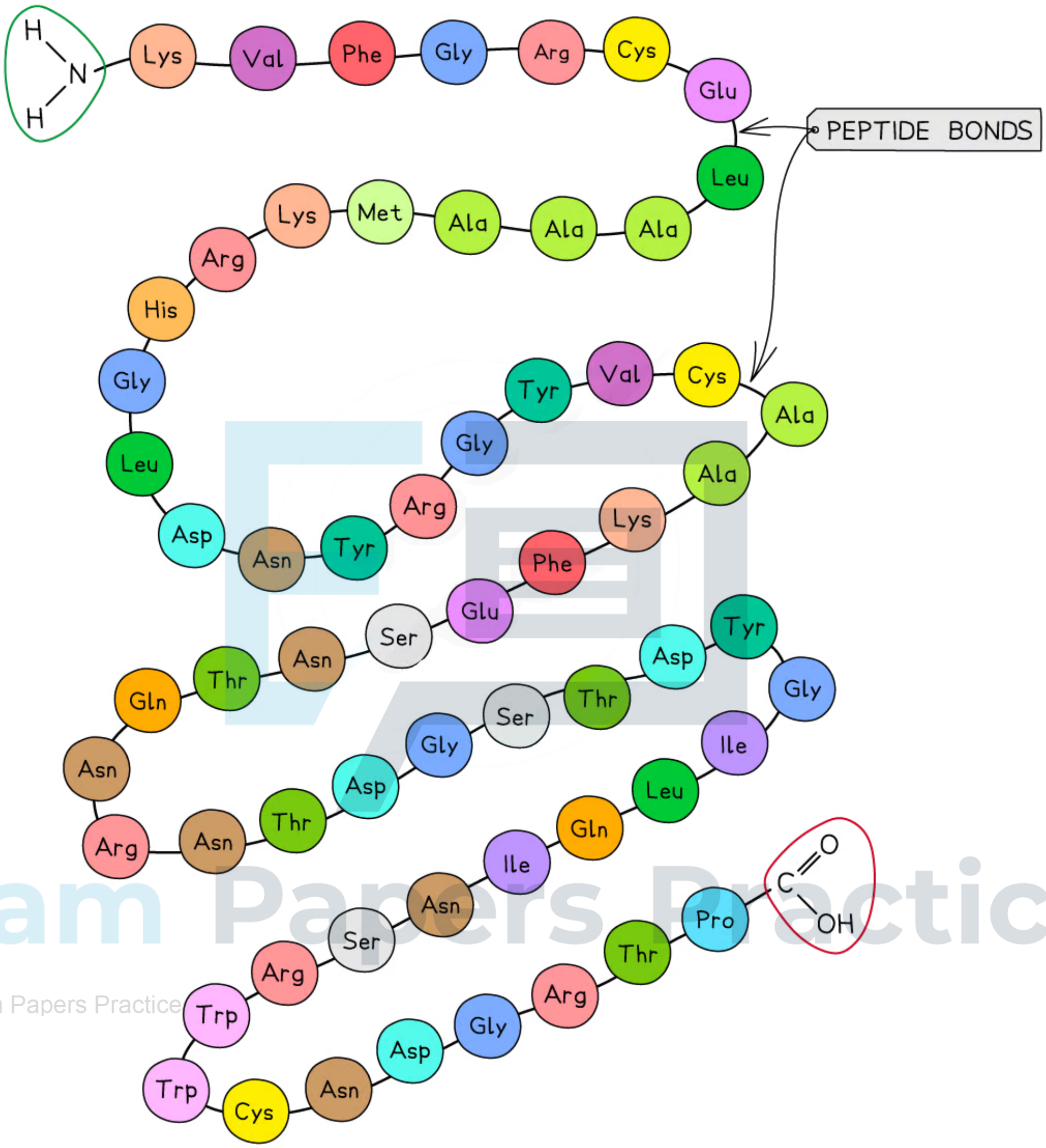
Remember to avoid confusing the bonds that hold a protein's shape together with the peptide bonds that attach each amino acid in sequence. Picture the peptide bonds holding the amino acids in a straight chain, then the other bonds and forces holding the chain in its folded, 3D structure.

Amino Acid Diversity (HL)

Amino Acid Diversity

- 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

Amino acid diversity diagram



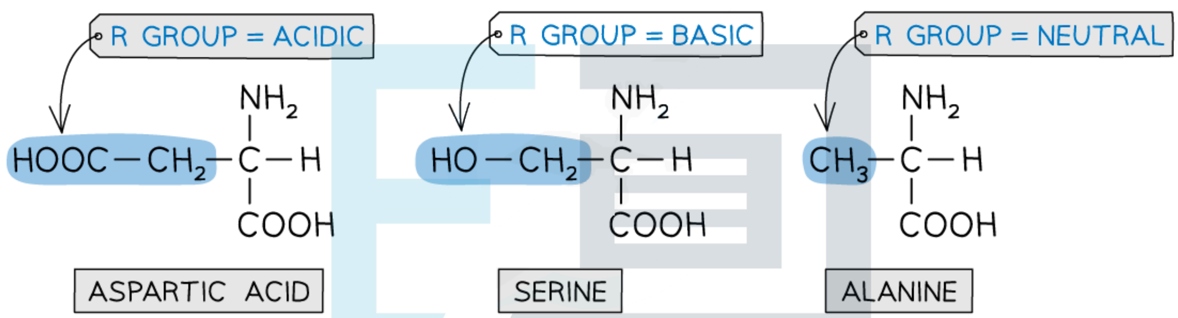
An amino acid sequence of a short polypeptide. The three-letter abbreviations indicate the specific amino acid (there are 20 commonly found in cells of living organisms).

R-groups

- An **R-group**, or variable group, 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 (forming the amino acid glycine), right through to complex aromatic ring structures (which forms phenylalanine)
- It is the R-group that gives the basis of diversity in protein function and form
- This occurs because:
  - All amino acids have a the same **carboxyl** and **amine group,** bonded by a central carbon atom, which **ionise** when in a neutral environment
    - The carboxyl group donates a hydrogen ion whilst the amine group accepts it so that the carboxyl group is left with a negative charge and the amine group left with a positive charge
  - The R-group is either hydrophobic or hydrophilic
    - Hydrophilic R-groups are polar and can be either acidic or basic
    - Hydrophobic R-groups are non-polar

R-Group Diagram



The variety of R-groups of the amino acids determine the properties of polypeptide chains and therefore overall protein

Levels of Protein Structure (HL)

Primary Structure

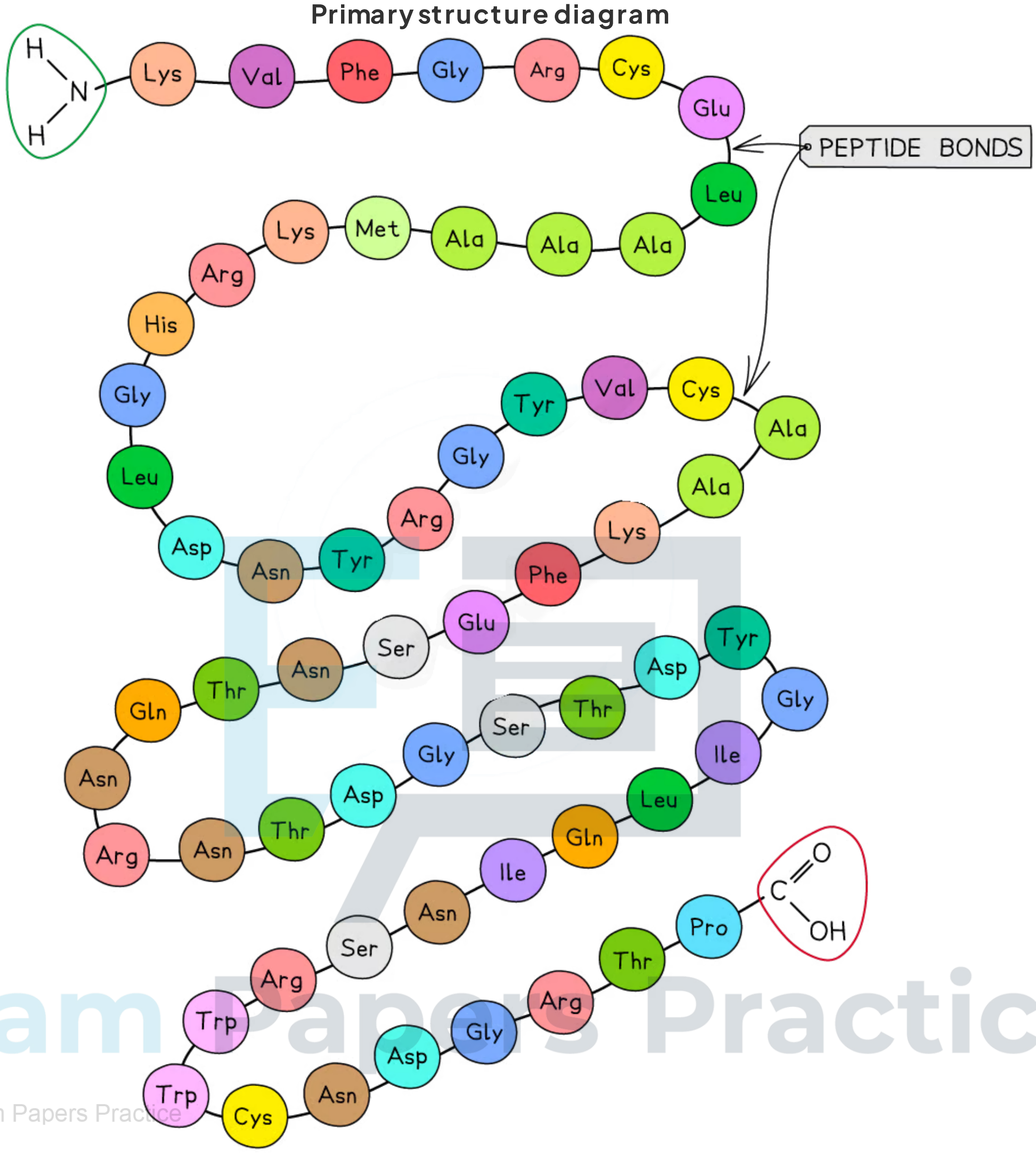
Levels of Protein Structure

- Proteins are relatively **large**, **complex molecules** that contain **one or more chains** of amino acids known as **polypeptides**
- The **three-dimensional arrangement** of polypeptide chains dictates a protein's structure and function
- There are **four levels of structure in proteins**
  - Three levels are structural aspects of a single polypeptide chain
  - The fourth level relates to a protein that has more than one polypeptide chain

Primary structure

- The **sequence** of amino acids bonded by covalent peptide bonds is the **primary structure** of a protein
- The **DNA** of a cell **determines** the primary structure of a protein by instructing the cell to add certain amino acids in specific quantities in a specific, ordered sequence
- This affects the **shape,** and therefore the **function,** of the protein
- The primary structure is **specific** for each protein
- Some mutations can lead to the incorrect amino acid being incorporated into the polypeptide chain which can affect the function of the protein

Primary structure diagram

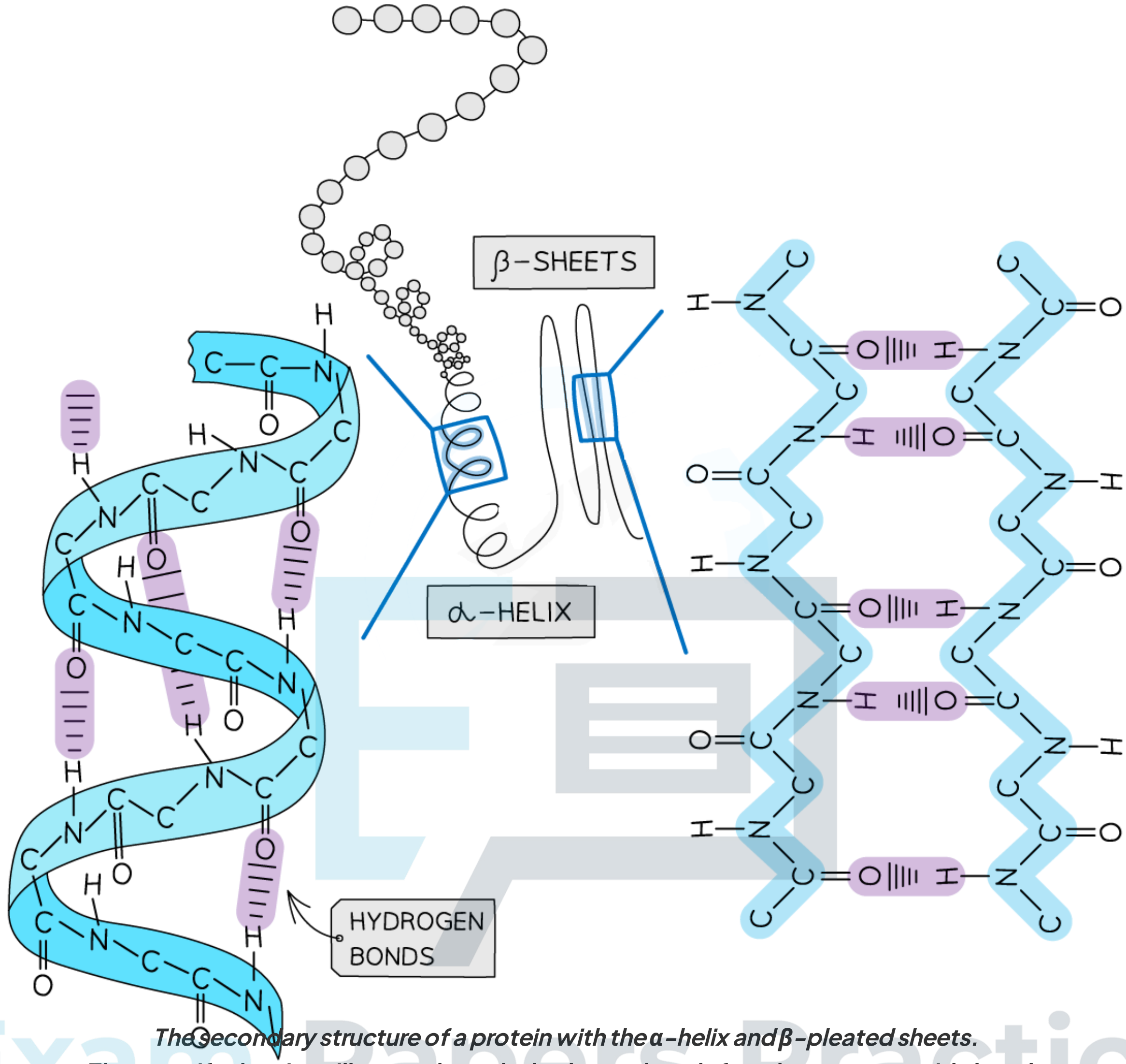


The primary structure of a protein. The three-letter abbreviations indicate the specific amino acid (there are 20 commonly found in cells of living organisms).

Secondary Structure

- Secondary structure is the **formation of complex shapes** within the polypeptide chain
- Secondary structure of a protein occurs due to **weak hydrogen bonds**
  - Hydrogen bonds form between **carboxyl** (C=O) groups and **amino** (N-H) groups
  - The bonds usually form **between non-adjacent amino acids** resulting in a change in shape of the linear polypeptide chain
- There are **two shapes** that can form within proteins due to the hydrogen bonds:
  - Alpha-helix (or **α-helix**)
  - Beta-pleated sheet (or **β-pleated sheet**)

Protein secondary structure diagram



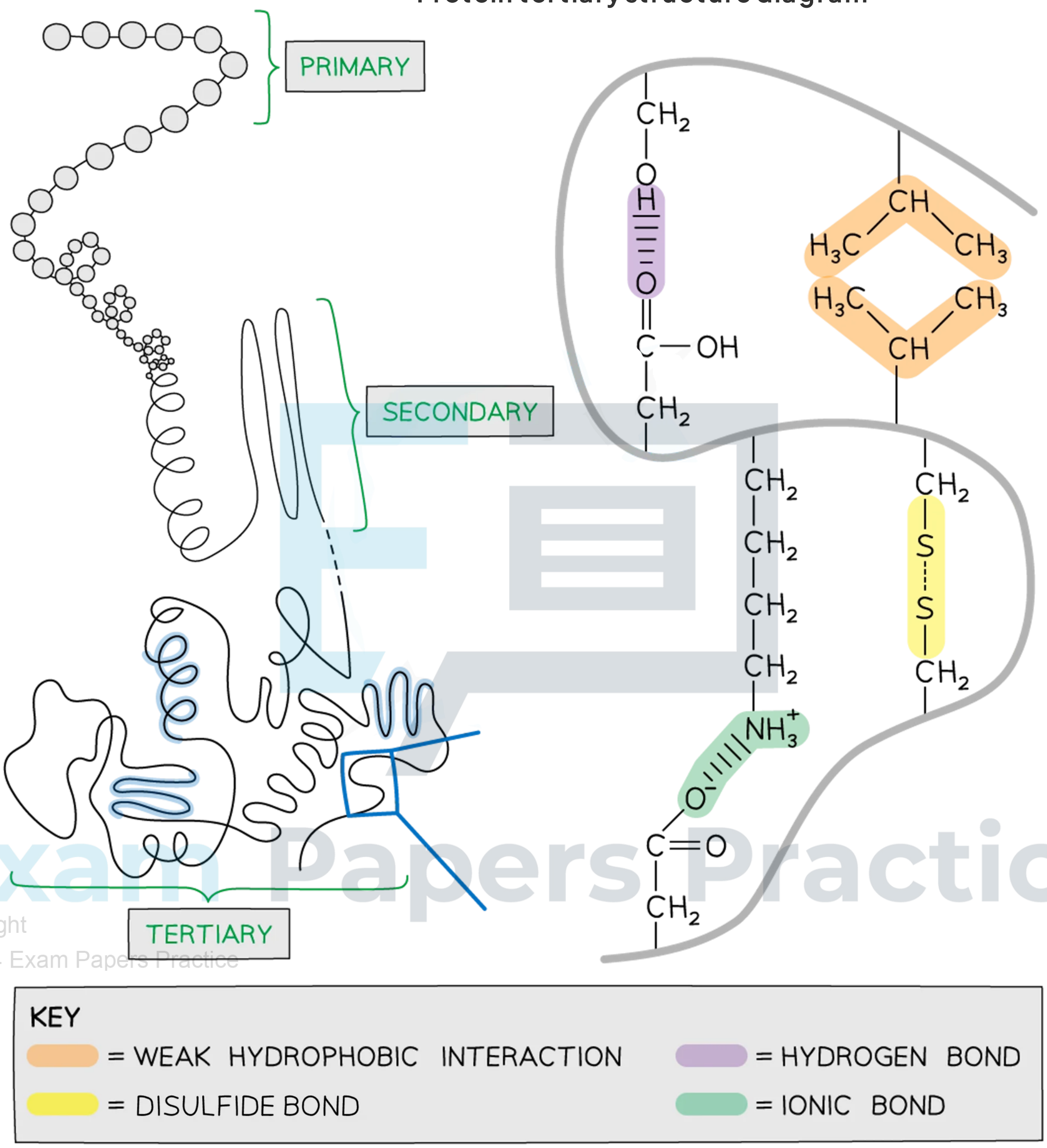
The secondary structure of a protein with the α-helix and β-pleated sheets.
The magnified regions illustrate how the hydrogen bonds form between peptide bonds.

Tertiary Structure: Chemical Bonds

Polar and non-polar amino acids are relevant to the bonds formed between R groups

- Tertiary structure refers to how the polypeptide chain folds to form a complex, **three-dimensional shape**
- Tertiary structure gives proteins a **very specific shape** that is important for function
  - Such as **receptor sites** on cell membranes and **active sites** in enzymes
- Folding results from **interactions between R groups** (side chains) of the amino acids and the surrounding environment
- A number of different interactions between R-groups contribute to the tertiary structure
  - **Hydrogen bonds** form between polar R-groups
  - **Hydrophobic interactions** form between the R-groups of non-polar amino acids **within the interior of proteins** to avoid contact with water
  - **Covalent bonds** form between the R-groups of cysteine amino acids to form **disulfide bridges**
  - **Ionic bonds** form between positively and negatively charged R-groups
    - R-groups can become positively or negatively charged by the dissociation or binding of **hydrogen ions**

Protein tertiary structure diagram



The interactions that occur between the R groups of amino acids determine the tertiary structure and function of a protein

Summary of bonds in proteins table

Bonds	

	Level		
	Primary	Secondary	Tertiary
Peptide	✓	✓	✓
Hydrogen		✓(only between the amino and carboxyl groups)	✓(R groups + amino and carboxyl groups)
Disulfide			✓
Ionic			✓
Hydrophobic interactions			✓

Tertiary Structure: Amino Acids

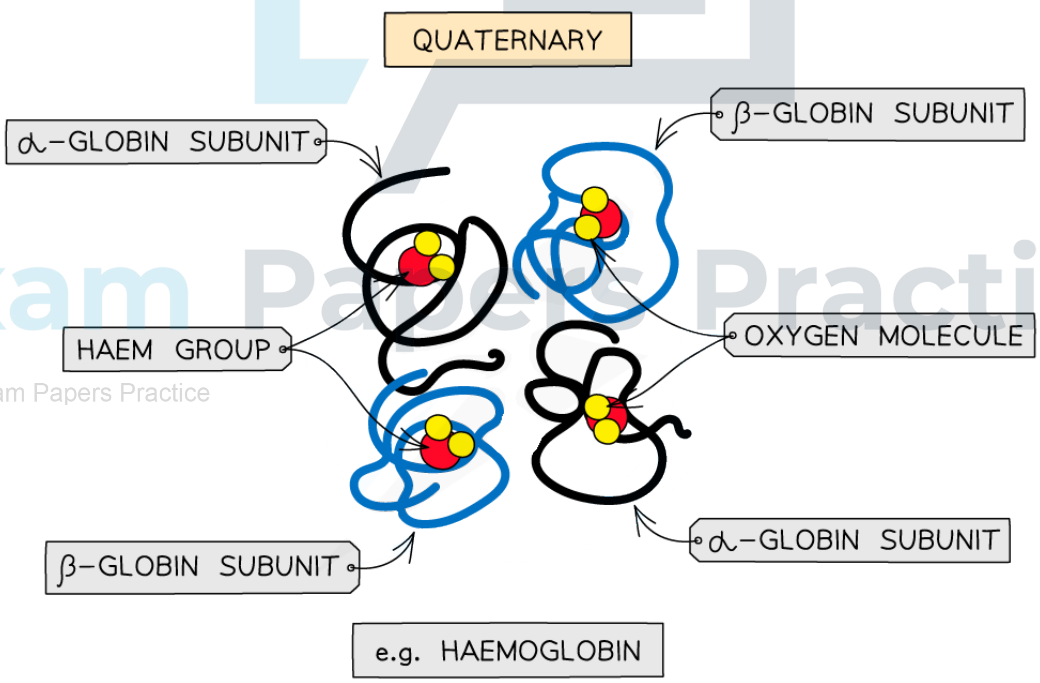
- Amino acids are either **polar** or **non-polar** depending on their R-groups
- Proteins composed of **non-polar amino acids** are less soluble in aqueous solutions such as the cytoplasm
- Therefore these proteins are generally used for structural purposes and are stationary so are not required to be soluble
    - They are found in the centre of a protein helping to stabilise the structure
    - They can help form the active site of lipase enzymes to allow interaction with lipid substrates
    - They tend to be localised on the surface of a cell so are in contact with the membrane, such as glycoproteins
- Proteins with **polar amino acids** are soluble and are found in a variety of places within the cell
    - They can be found on the surface of a membrane as they are capable of interacting with water molecules
    - They can line interior pores within the membrane, which creates hydrophilic channels for transport of polar molecules into and out of a cell
    - They are found on the outside of enzymes so that enzymes are soluble in aqueous environments

Quaternary Structure

Quaternary structure

- Large proteins often consist of **multiple polypeptide chains** functioning together as a larger biologically active macromolecule
  - Each polypeptide chain is referred to as a **subunit** of the protein
- Many proteins also contain non-polypeptide components (**prosthetic groups**) and are classed as **conjugated proteins**
- **Quaternary structure** refers to how polypeptides and other components are arranged
  - This relates closely to function
  - Proteins with only one polypeptide chain do not have a quaternary structure
- **Haemoglobin** is a **conjugated protein**, having quaternary structure, as it consists of **multiple polypeptide chains** (making **four subunits**) each with a prosthetic group
  - There are two pairs of identical polypeptide chains (**α-globins** and **β-globins**)
  - Each subunit has a prosthetic **haem** group which contains an **iron** ion (Fe^{2+})

Haemoglobin structure diagram



The quaternary structure of haemoglobin.
Four subunits (polypeptide chains) and prosthetic haem groups work together to carry oxygen.

- **Insulin** and **collagen** are **non-conjugated proteins** meaning they have no other non-protein components
- Insulin:
  - Consists of two chains of amino acids, one being 21 amino acids long, the other 30 amino acids in length; the chains are joined by disulfide bridges
  - It forms two quaternary different structures called dimers and hexamers which act as storage molecules of insulin
- Collagen:
  - It is a fibrous protein consisting of three polypeptide chains wound together in a helix shape
  - It is the arrangement of the helix shape that gives collagen its quaternary structure

Exam Tip

Familiarise yourself with the difference between the four structural levels found in proteins, noting which bonds are found at which level. Remember that the hydrogen bonds in tertiary structures are between the R groups whereas in secondary structures the hydrogen bonds form between the amino and carboxyl groups.

NOS: Technology allows imaging of structures that would be impossible to observe with the unaided senses. For example, cryogenic electron microscopy has allowed imaging of single-protein molecules and their interactions with other molecules

- The technique of cryogenic electron microscopy (cryo-EM) involves **rapid freezing** of protein solutions and then exposing them to many **electrons** to produce a microscopic image
- The images can be used to **recreate 3D shape** or structure of proteins allowing us to visualise how they **interact** with other molecules within a cellular environment
- Cryo-EM has different applications depending on the type of protein or molecule being studied so observations can be extremely purposeful and exact
- Until recently, proteins had to be crystallised to reconstruct and visualize them with **X-ray crystallography** which posed many problems such as:
  - Crystallisation is time consuming and can only work on single purified protein
  - Some proteins do not crystallise
  - The structure has to be visualised outside of the cellular environment which removes contextual information and interactions with other molecules

Globular & Fibrous Proteins (HL)

Globular & Fibrous Proteins

Globular

- **Globular proteins** are **compact**, roughly **spherical** (circular) in shape and **soluble** in water
- Globular proteins form a spherical shape when folding into their tertiary structure because:
  - Their **non-polar hydrophobic R-groups** are orientated towards the **centre** of the protein away from the aqueous surroundings and
  - Their **polar hydrophilic R-groups** orientate themselves on the **outside** of the protein
- This orientation enables globular proteins to be (generally) **soluble** in water as the water molecules can surround the **polar hydrophilic R-groups**
- The **solubility** of globular proteins in water means they play important **physiological** roles as they can be easily **transported** around organisms and be involved in **metabolic reactions**
- The folding of the protein due to the interactions between the R-groups results in globular proteins having **specific shapes**. This also enables globular proteins to play physiological roles, for example, **enzymes** can catalyse specific reactions and **immunoglobulins** can respond to specific antigens
- Some globular proteins are conjugated proteins that contain a prosthetic group eg. haemoglobin which contains the prosthetic group called haem

Insulin

- The first protein to have its sequence determined by scientists was the **hormone** insulin
- Insulin is a globular protein produced in the pancreas. It plays an important role in the **control of blood glucose concentration**
- It consists of **two polypeptide chains**
  - Polypeptide A has 21 amino acid residues
  - Polypeptide B has 30 amino acid residues
- The two polypeptide chains are held together by **three disulfide bridges**

Fibrous

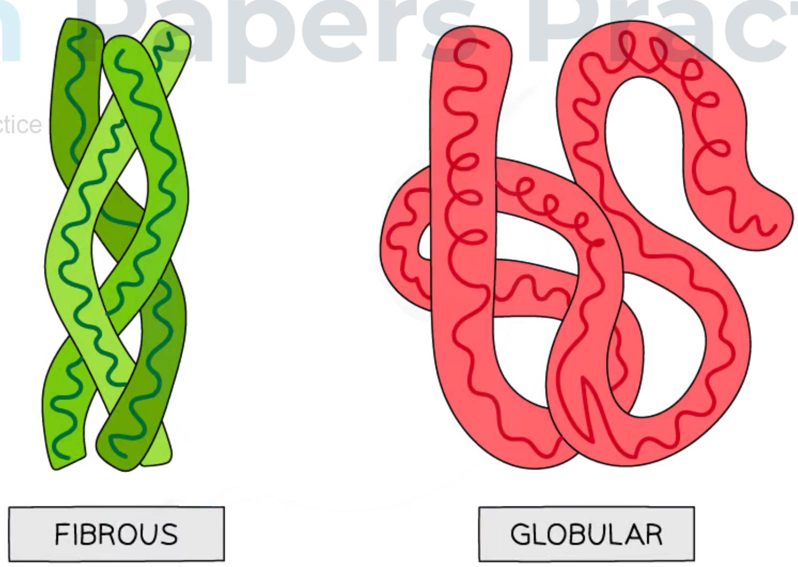
- **Fibrous proteins** are long strands of polypeptide chains that have cross-linkages due to hydrogen bonds
- They have little or no tertiary structure
- Due to the large number of **hydrophobic R-groups** fibrous proteins are **insoluble** in water
- Fibrous proteins have a limited number of amino acids with the sequence usually being highly repetitive
- The highly repetitive sequence creates very organised structures that are strong and this along with their insolubility property, makes fibrous proteins very suitable for structural roles, for

example, keratin that makes up hair, nails, horns and feathers and collagen which is a connective tissue found in skin, tendons and ligaments

Collagen

- **Collagen** is the most common **structural** protein found in vertebrates
- It has a **flexible structure**, forming **connective tissues:**
  - Tendons
  - Cartilage
  - Ligaments
  - Bones
  - Teeth
  - Skin
  - Walls of blood vessels
  - Cornea of the eye
- Collagen is an **insoluble fibrous** protein
- Collagen is formed from **three polypeptide chains** closely held together by **hydrogen bonds** to form a **triple helix;** the hydrogen bonds give **great tensile strength**
- Each polypeptide chain is a helix shape and contains about 1000 amino acids with glycine, proline and hydroxyproline being the most common
- Along with hydrogen bonds forming between the three chains there are also covalent bonds present
- Covalent bonds also form **cross-links** between R-groups of amino acids in interacting **triple helices** when they are arranged parallel to each other. The cross-links hold the collagen molecules together to form **fibrils**
- The collagen molecules are positioned in the **fibrils** so that there are **staggered ends** which provide **strength**
- When many fibrils are arranged together they form collagen **fibres**
- Collagen fibres are positioned so that they are lined up with the forces they are withstanding

Globular and fibrous proteins diagram



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