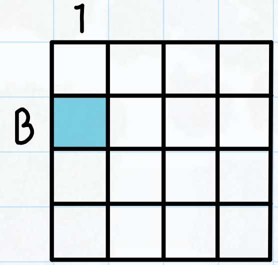
B1.2 Proteins

Ver. 2

Guiding Questions
What is the relationship between amino acid sequence and the diversity in form and function of proteins?
How are protein molecules affected by their chemical and physical environments?

Linking Questions
How do abiotic factors influence the form of molecules?
What is the relationship between the genome and the proteome of an organism?



Theme: Form and Function

Level of Organization: Molecules

Written and drawn by:

Peter Marier

https://pdb101.rcsb.org/motm/motm-about

SL Learning Outcomes

B1.2.1	Generalized structure of an amino acid	Students should be able to recognize and interpret the generalized structure of an amino acid, including the alpha carbon atom with an amine group, a carboxyl group, an R-group and a hydrogen atom. The general structure of an amino acid is provided in the data booklet.
B1.2.2	Condensation reactions forming dipeptides and longer chains of amino acids	Students should be able to write the word equation for this reaction and draw a generalized dipeptide after modelling the reaction with molecular models.
B1.2.3	Dietary requirements for amino acids	Essential amino acids cannot be synthesized and must be obtained from food. Non-essential amino acids can be made from other amino acids. Students are not required to give examples of essential and nonessential amino acids. Vegan diets require attention to ensure essential amino acids are consumed.
B1.2.4	Infinite variety of possible peptide chains	Include the ideas that 20 amino acids are coded for in the genetic code, that peptide chains can have any number of amino acids, from a few to thousands, and that amino acids can be in any order. Students should be familiar with examples of polypeptides.
B1.2.5	Effect of pH and temperature on protein structure	Include the term “denaturation”.

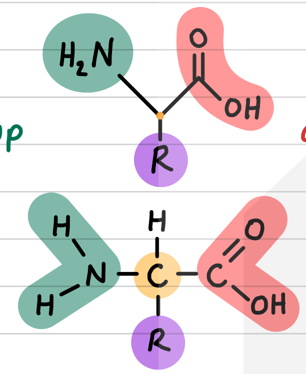
HL Learning Outcomes

B1.2.6	Chemical diversity in the R-groups of amino acids as a basis for the immense diversity in protein form and function	Students are not required to give specific examples of R-groups. However, students should understand that R-groups determine the properties of assembled polypeptides. Students should appreciate that R-groups are hydrophobic or hydrophilic and that hydrophilic R groups are polar or charged, acidic or basic.
B1.2.7	Impact of primary structure on the conformation of proteins	Students should understand that the sequence of amino acids and the precise position of each amino acid within a structure determines the three-dimensional shape of proteins. Proteins therefore have precise, predictable and repeatable structures, despite their complexity.
B1.2.8	Pleating and coiling of secondary structure of proteins	Include hydrogen bonding in regular positions to stabilize alpha helices and beta-pleated sheets.
B1.2.9	Dependence of tertiary structure on hydrogen bonds, ionic bonds, disulfide covalent bonds and hydrophobic interactions	Students are not required to name examples of amino acids that participate in these types of bonding, apart from pairs of cysteines forming disulfide bonds. Students should understand that amine and carboxyl groups in R-groups can become positively or negatively charged by binding or dissociation of hydrogen ions and that they can then participate in ionic bonding.
B1.2.10	Effect of polar and non-polar amino acids on tertiary structure of proteins	In proteins that are soluble in water, hydrophobic amino acids are clustered in the core of globular proteins. Integral proteins have regions with hydrophobic amino acids, helping them to embed in membranes.
B1.2.11	Quaternary structure of non-conjugated and conjugated proteins	Include insulin and collagen as examples of non-conjugated proteins and haemoglobin as an example of a conjugated protein. **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.
B1.2.12	Relationship of form and function in globular and fibrous proteins	Students should know the difference in shape between globular and fibrous proteins and understand that their shapes make them suitable for specific functions. Use insulin and collagen to exemplify how form and function are related.

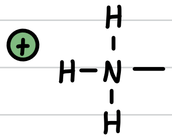
Composition: all proteins contain carbon, hydrogen, oxygen, nitrogen, and sometimes sulfur (R group dependent)

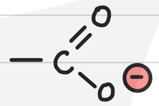
Structure: polymers made from amino acids all composed of an α carbon bound to functional groups:

acts like a base, accepting a H^+ becoming positive — amine group



carboxyl group — acts like an acid donating a H^+ becoming negative



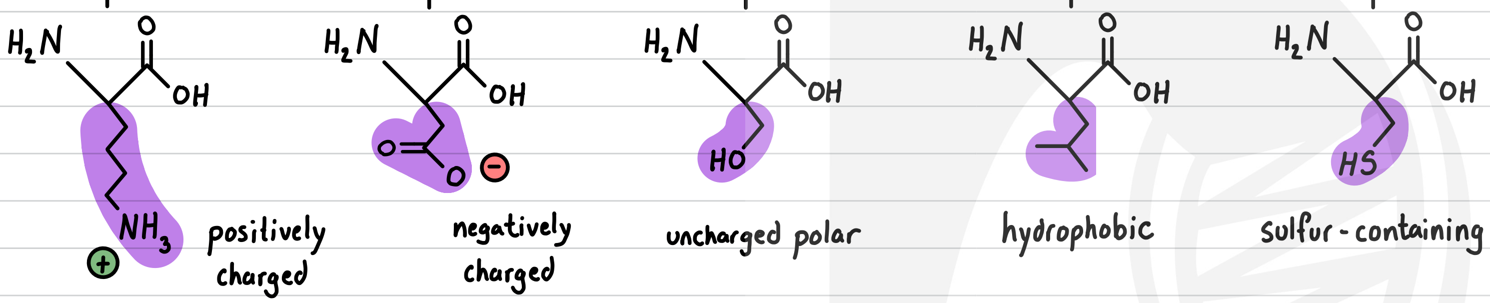


※ found in data booklet

※ amino acids are amphiprotic as they can accept and donate H^+

variable R side-chain group "residue"

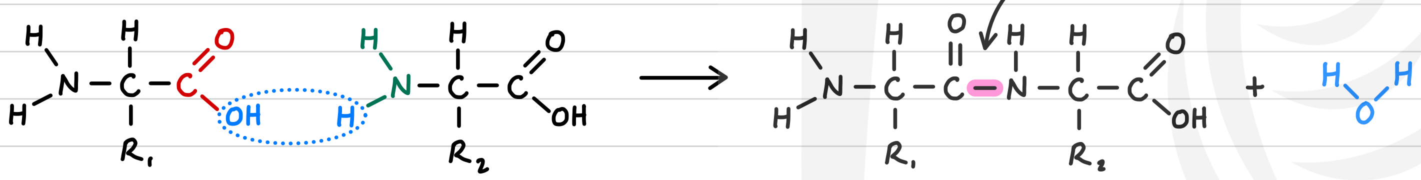
20 different side-chains or R groups | with different properties and compositions (extended in HL)



amino acids join together in condensation reactions to form dipeptides (carbohydrates B1.1)

※ this occurs within ribosomes during translation (protein synthesis D1.2)

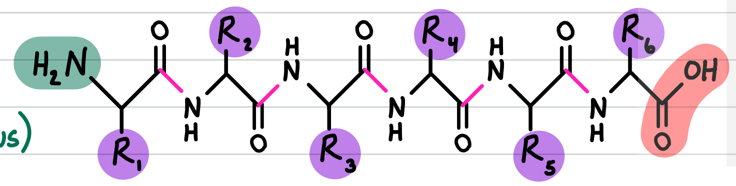
↳ carboxyl group of 1 amino acid reacts with the amine of another forming a peptide bond and water



amino acid 1 + amino acid 2 ⟶ dipeptide + water

↳ many amino acids join together to form polypeptides (named after many peptide bonds)

amino terminus (N-terminus)



carboxyl terminus (C-terminus)

explore using molecular models:
↳ https://chemagic.org/molecules/amini.html
↳ https://molview.org/?cid=2519

ex: polypeptide above composed of 6 amino acids ∴ 6 R groups, 5 peptide bonds, 5 water molecules involved

Proteins are synthesized from a code of DNA nucleotides

※ proteome: all proteins made by a cell / tissue / organism

DNA gene —transcription→ mRNA —translation→ polypeptide

(protein synthesis D1.2)

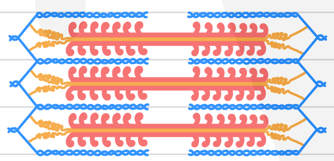
There is an ~infinite variety of possible peptide chains — Why?
↳ 20 different amino acids that can combine and be arranged in any order
↳ polypeptides vary in length and can have any number of amino acids from a few to tens of thousands long

$$P = A^n$$

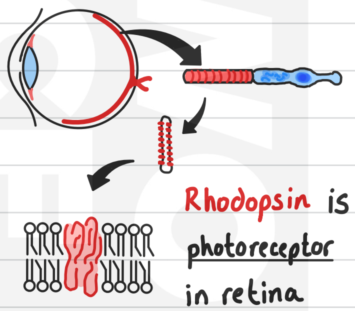
P: number of possible polypeptides
n: number of amino acids in polypeptide
A: number of different types of amino acids

ex: how many different polypeptides 10 amino acids long are possible? $P = 20^{10}$ = over 10 trillion!

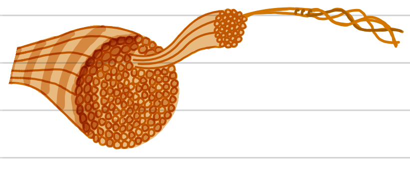
Proteins display a variety of functions – examples:



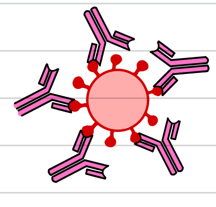
Titin, Actin and Myosin cause movement via contraction in muscles



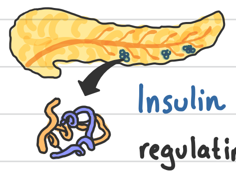
Rhodopsin is photoreceptor in retina



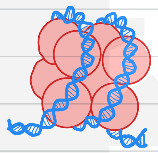
Collagen provides tensile strength and structure in connective tissues (skin, ligaments, tendons)



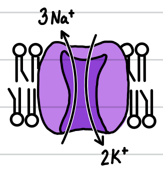
Immunoglobulin (i.e. antibodies) aid in immune response



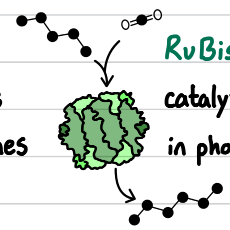
Insulin hormone regulating blood sugar



Histones aid in DNA packing and gene expression



Na^+/K^+ pump transports ions across membranes



RuBisCo enzyme catalyzes C fixation in photosynthesis

Sources of amino acids

Non-essential amino acids: amino acids that can be synthesized by the body
↳ not an essential component of a diet as if missing, can be made from other amino acids

Essential amino acids: amino acids that cannot be synthesized by the body (9 out of 20 are essential)
↳ essential component of a diet as if missing, cannot be made, potentially causing protein deficiency malnutrition; where the body is unable to create sufficient proteins it requires

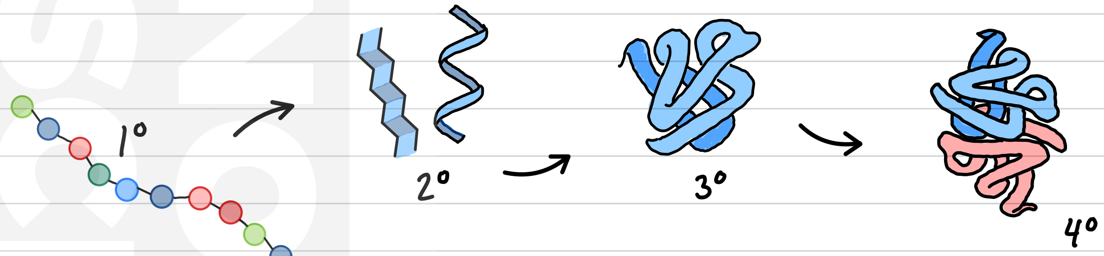
↳ foods vary in their amino acid content; it is possible to eat a protein-rich diet and still be deficient
∴ a balanced diet which contain a variety of essential amino acids is key (such as fish, meat, milk, eggs)

↳ plant-based diets (such as vegans) have fewer protein options so extra care needs to be taken to ensure all essential amino acids are being consumed. Like meat, different plants (ex: beans, lentils, nuts, seeds, tofu) have different essential amino acids so a varied, balanced diet is important

Once polypeptides are synthesized, they will fold into specific structures, based on the chemical properties of their R-groups

↳ the sequence of amino acids will determine the arrangement of R groups and thus interactions and shape

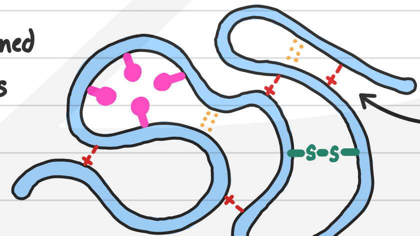
↳ the shape of the protein will determine its function (such as an enzyme's active site) enzymes C1.1



proteins can fold in increasingly complex structures and can be composed of multiple polypeptides

extended in HL →

↳ the 3-dimensional conformation of proteins is determined and stabilized by intramolecular forces between R-groups
↳ within molecule



intramolecular forces such as ionic bonds, hydrogen bonds, hydrophobic interactions and covalent bonds

↳ if these forces are disrupted, the 3-dimensional conformation will be altered, impacting the protein's function, i.e. denaturation

denaturation: structural change in a protein that results in a loss (typically permanent) of its biological properties
↳ particularly important for enzyme's whose shape (active site) is specific to particular substrates enzymes C1.1

Effect of temperature

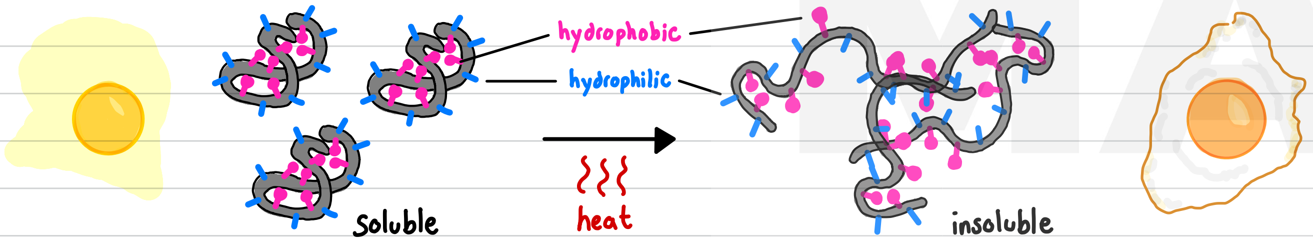
Temperature: average kinetic energy of a substance measured in K or °C
↳ higher temperature means faster movement of atoms and thus energy

↳ Proteins each have temperature ranges where they function optimally which will depend on their amino acid sequence and the intramolecular forces present

↳ Above this optimal temperature, the increased energy will cause the polypeptide to vibrate/move so much that weak intramolecular bonds (i.e. non-covalent) are stressed and can break, altering its 3-D shape and causing it to denature - leading to a loss of biological function

※ denaturation does not typically disrupt peptide bonds so the polypeptide chain is intact
∴ in some cases, returning to optimal temperature can renature the protein, re-establishing intramolecular forces (although these cases are more rare and often denaturation via heat is permanent)

ex: when cooking eggs, it turns from a clear liquid to solid white as the main protein, albumin was originally soluble but as it denatured, the hydrophobic regions became exposed and different chains formed new bonds, changing its structure

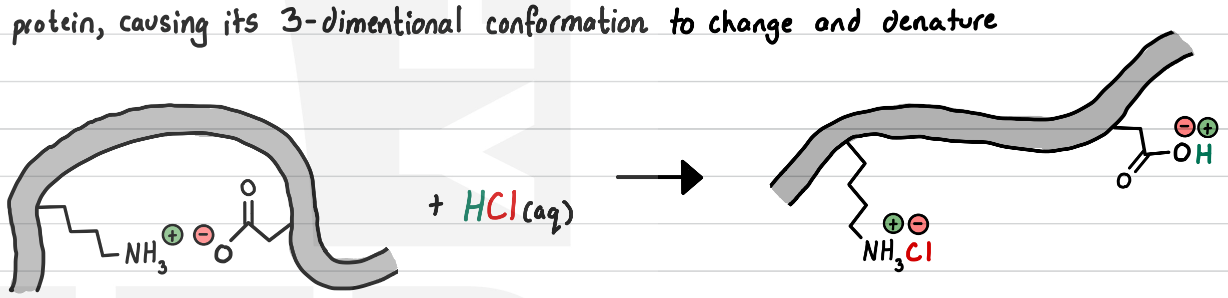


Effect of pH

pH: a measure of acidity/alkalinity of a substance measured on a logarithmic scale (0 - 14)
↳ 0 - <7 = acidic, i.e. more concentration of H^+
↳ >7 - 14 = alkaline/basic, i.e. more concentration of OH^- (less concentration of H^+)

↳ Proteins each have pH ranges where they function optimally which will depend on their amino acid sequence and the intramolecular forces present such as ionic bonds between positive and negatively charged R-groups as well as hydrogen bonds between polar R groups

↳ a change in this pH can alter chemical properties of the R-groups, namely their charge. a change in charge (ex: positive to neutral) can disrupt and break ionic bonds within the protein, causing its 3-dimentional conformation to change and denature



※ unless exposed to very strong acids and bases, returning a protein to its optimum pH should restore charges and thus intramolecular forces, causing it to renature

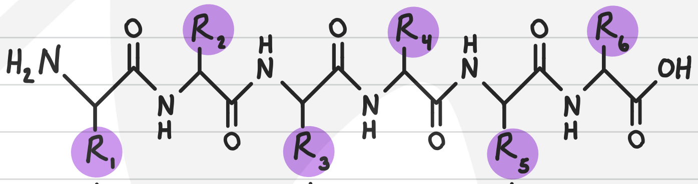
ex: exposing milk to lemon juice will cause it to curdle. The protein within, casein, denatures as lemon juice is acidic and its structure is altered

HL B1.2.6—Chemical diversity in the R-groups of amino acids as a basis for the immense diversity in protein form and function

The immense diversity of proteins form (and thus function) is due to the chemical diversity of R-groups

note: do not need to memorize all amino acids and R-groups

↳ all polypeptides will have an amino and carboxyl terminal and be joined by peptide bonds. Where they vary are R-groups

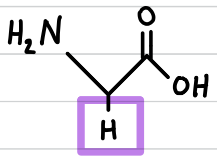


↳ R-groups determine properties of assembled polypeptides as they have different compositions and chemical characteristics

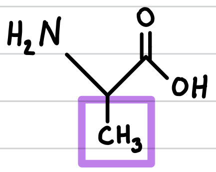
※ the following R group structures are shown as they would exist at the pH inside a cell

hydrophobic groups

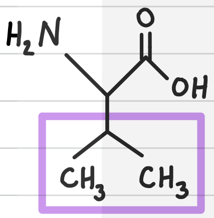
9 groups are non-polar and uncharged, causing them to repel and be insoluble in water. Groups form hydrophobic interactions with other non-polar uncharged groups



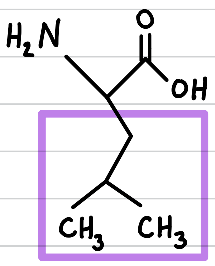
(Gly)



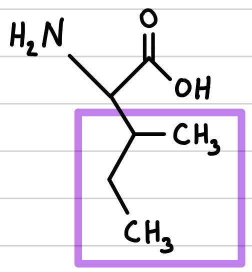
(Ala)



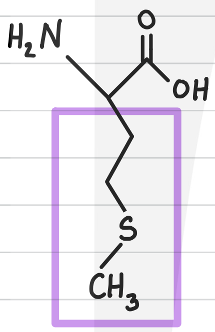
(Val)



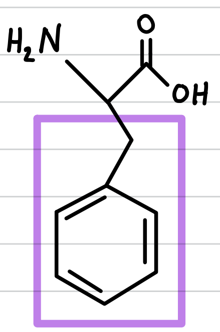
(Leu)



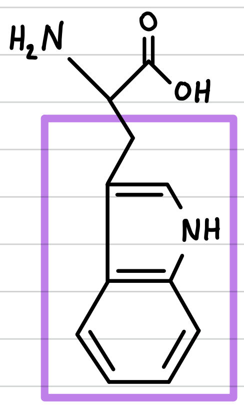
(Ile)



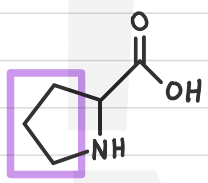
(Met)



(Phe)



(Trp)



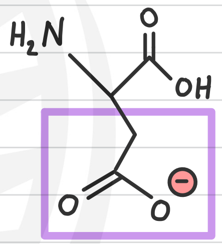
(Pro)

hydrophilic groups

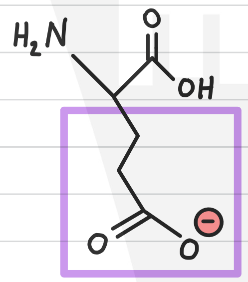
charged

acidic (negatively charged)

2 groups act as acids in aqueous solutions, donating a H^+ becoming negatively charged, allowing the formation of ion-dipole interactions with water (soluble) and ionic bonds with positively-charged groups



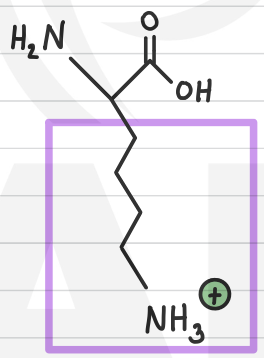
(Asp)



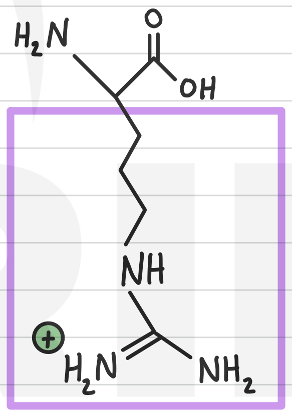
(Glu)

basic (positively charged)

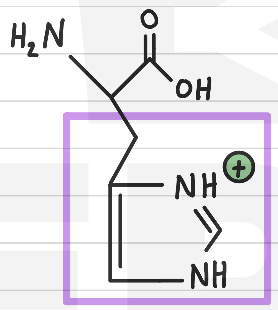
3 groups act as bases in aqueous solutions (i.e. alkalis), accepting a H^+ becoming positively charged allowing ion-dipole interactions and ionic bonds



(Lys)



(Arg)

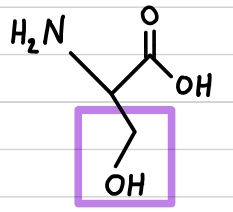


(His)

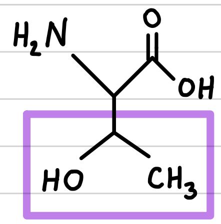
* partly protonated

polar

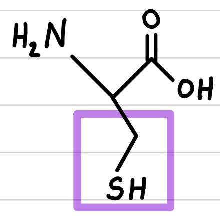
6 groups have a δ+ and δ- charge due to their polarity, allowing the formation of Hydrogen bonds making them soluble in water



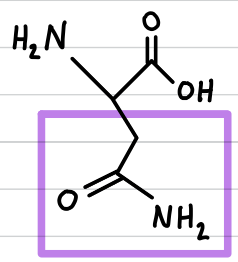
(Ser)



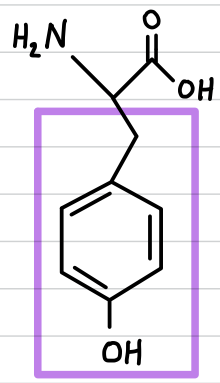
(Thr)



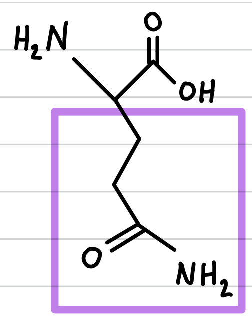
(Cys)



(Asn)



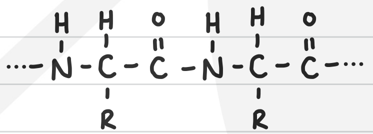
(Tyr)



(Gln)

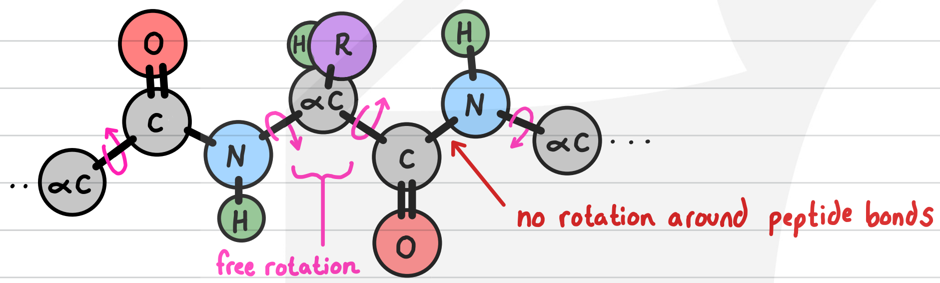
Primary protein structure: linear sequence of amino acids in a polypeptide linked together by covalent peptide bonds

↳ the primary structure is a sequence of amino acids in a repeating



pattern

↳ the bond angles are tetrahedral and their is free rotation about the α-C and adjacent N and C allows the polypeptide to fold in many 3-D shapes

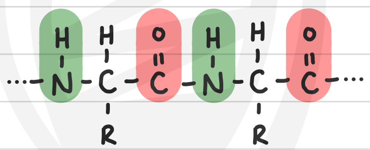


↳ a DNA gene provides the instructions for a polypeptide sequence including its length (how many amino acids), composition (which amino acids, i.e. R-groups) and placement (the order of amino acids in the chain) giving proteins a precise, predictable and repeatable structure → not random

∴ DNA dictates the primary structure which dictates the 3-D conformation and function of a protein

Secondary protein structure: pleating and coiling of a polypeptide into alpha helices and beta pleated sheets due to hydrogen bonding between backbone amide N-H and carbonyl C=O

↳ a polypeptide backbone has repeating N-H and C=O groups



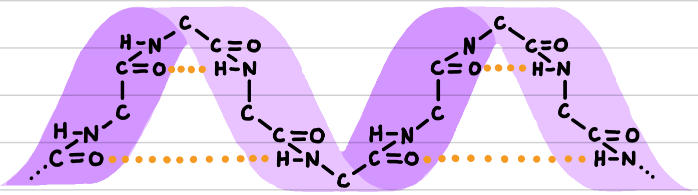
↳ these groups are polar due to uneven sharing of electrons   δ+ δ−   −N−H·········O=C−

∴ these groups will form a hydrogen bond with each other, stabilizing the structure

↳ these interactions cause the polypeptide to fold into different shapes:

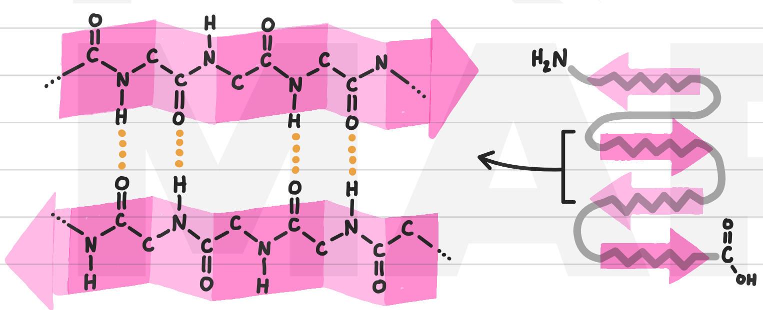
alpha (α) helix

↳ polypeptide wound into right-handed helix with hydrogen bonds between adjacent turns of the helix



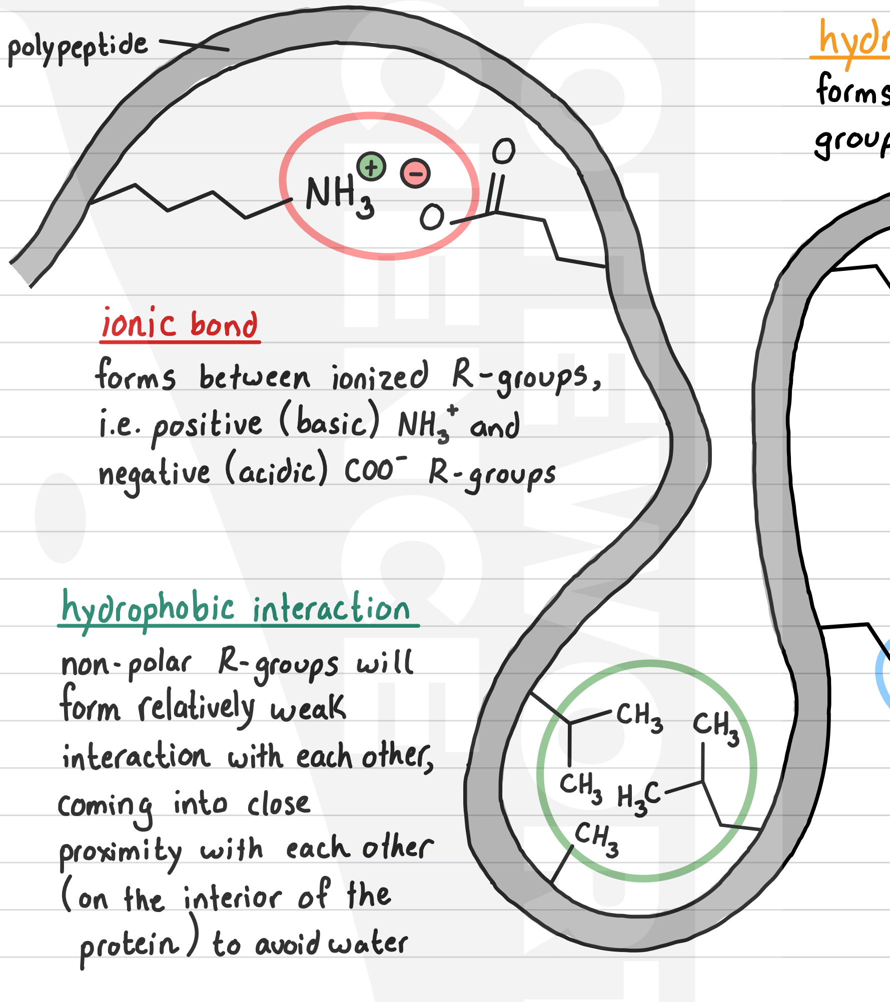
beta (β) pleated sheet

↳ two or more sections of the polypeptide arranged either in parallel (same direction) or antiparallel (opposite directions) with hydrogen bonds between them



※ a single polypeptide can have only α-helices, only β-sheets or both, depending on its sequence

Tertiary protein structure: overall 3-dimensional shape of the protein due to intramolecular interactions / bonds between R-groups



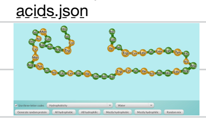
ionic bond
forms between ionized R-groups, i.e. positive (basic) NH_3^+ and negative (acidic) COO^- R-groups

hydrogen bond
forms between the δ+ and δ− atoms of polar R groups. Weak and most common of interactions

hydrophobic interaction
non-polar R-groups will form relatively weak interaction with each other, coming into close proximity with each other (on the interior of the protein) to avoid water

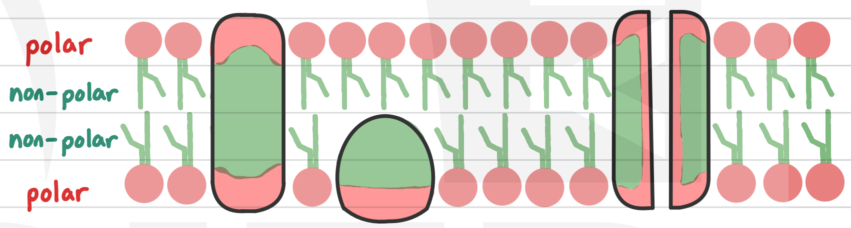
disulfide covalent bond
the sulfur atoms from two cysteine R-groups will form a covalent disulfide bridge / linkage (strongest interaction)

explore folding using simulation:
↳ https://lab.concord.org/embeddable.html#interactives/samples/5-amino-acids.json



As some R-groups are polar (hydrophilic) and non-polar (hydrophobic) their position in the polypeptide will impact the structure, properties, and function of the protein in the cell (chemical signalling C2.1)

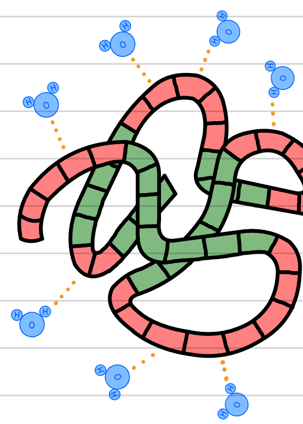
↳ **Integral proteins** (imbedded in the plasma membrane)



Proteins have hydrophilic regions facing intracellular / extracellular space, allowing interactions with water while the core is hydrophobic, allowing it to embed with non-polar tails

channel proteins have hydrophilic regions in a tunnel / pore, allowing hydrophilic substances (ions, polar molecules) to pass through hydrophobic core

↳ tertiary globular proteins can be soluble in water, despite having many non-polar amino acids by folding:



hydrophilic polar amino acids are on the outside of the protein, allowing it to form hydrogen bonds with water and dissolve

hydrophobic non-polar amino acids are in the core, avoiding contact with water

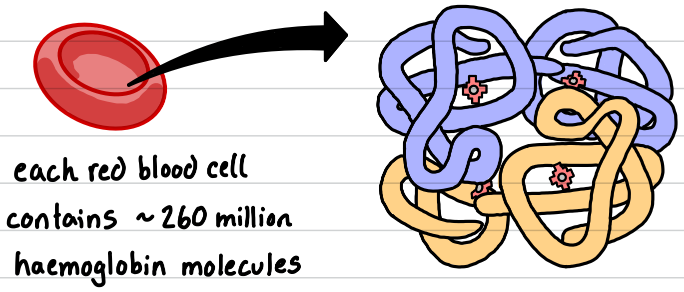
HL

B1.2.11—Quaternary structure of non-conjugated and conjugated proteins.
B1.2.12—Relationship of form and function in globular and fibrous proteins

Quaternary protein structure: protein complex composed of 2 or more polypeptides

↳ **Conjugated proteins**: composed of both protein and non-protein prosthetic group(s) such as carbohydrates, lipids, metal ions, and other organic groups

ex: Haemoglobin - quaternary protein made of 4 subunits.



each red blood cell contains ~260 million haemoglobin molecules

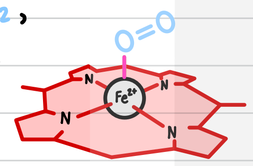
↳ each subunit is a conjugated protein made of globin and a prosthetic haem group

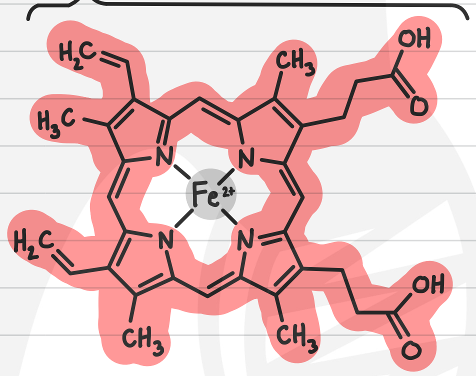
↳ 2 α globins, 2 β globins: each a polypeptide, folded into many helices and bound to a haem prosthetic group

↳ function: the haem group contains an Fe^{2+} which binds reversibly to O_2, allowing it to collect O_2 in the lungs and deliver it to cells around the body

🔗 transport B3.2

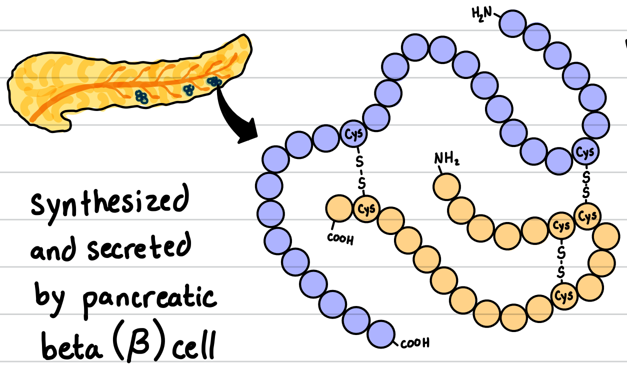
🔗 gas exchange B3.1





↳ **Non-conjugated protein**: composed of proteins only

ex: Insulin - quaternary protein made of 2 polypeptides (A chain and B chain)



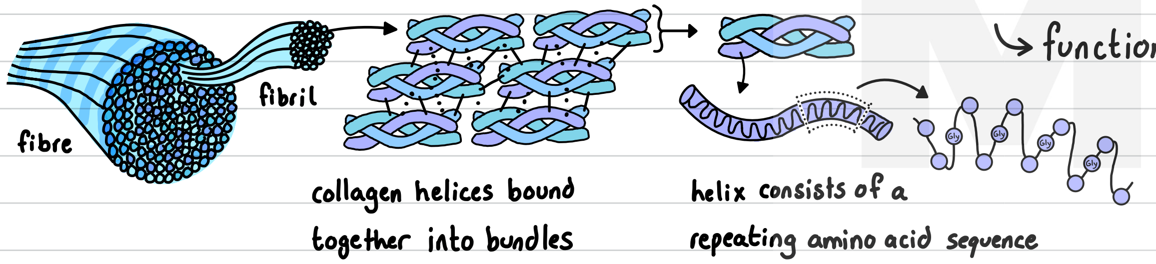
Synthesized and secreted by pancreatic beta (β) cell

↳ after initial synthesis (translation) it is a single polypeptide where it is then modified: splitting it into two chains linked by covalent disulfide bridges

🔗 protein synthesis D1.2

↳ function: hormone which promotes synthesis and storage of glycogen (glycogenesis) in the liver and muscle cells, reducing blood glucose levels

ex: Collagen - quaternary protein made of 3 left-handed helices wound together into a tight right-handed triple helix. These associate in groups (fibrils) to form tough, inextensible fibres



↳ function: structural protein making up connective tissue, giving tensile strength and structure to tendons, ligaments, and skin

NOS: Technology allows imaging of structures that would be impossible to observe with unaided senses

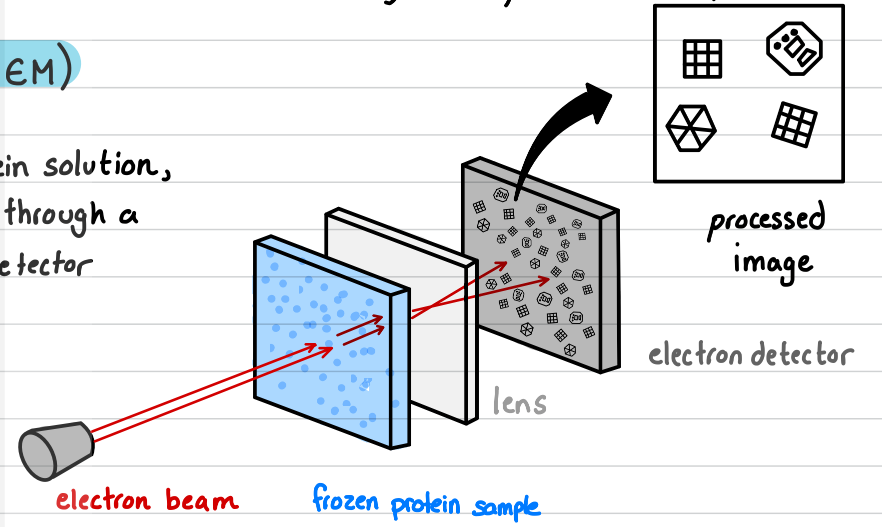
🔗 cell structure A2.2

↳ haemoglobin has a diameter of ~5nm, too small to be observed clearly even by most microscopes

solution: **cryogenic electron microscopy (cryo EM)**

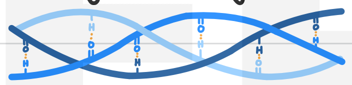
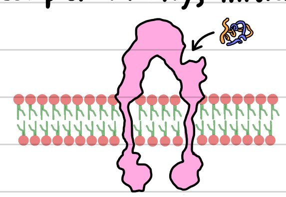
↳ Electron beams are fired at a frozen protein solution, scattering them. Scattered electrons focused through a lens, producing a magnified image on a detector where the structure can be worked out

↳ technique allows images to be produced to incredible resolution (0.12 nm) allowing atom positions to be seen



↳ freezing technique allows conformational changes to be seen as protein carries out task, allowing not just form but function to be determined as well as interactions with other molecules

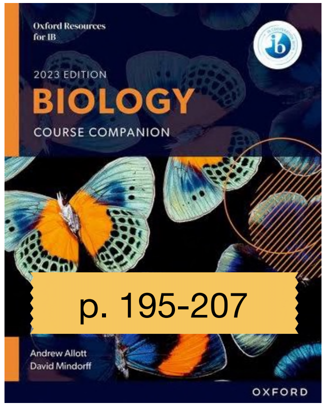
※ the function of a protein depends on its form/structure, i.e. "form follows function"

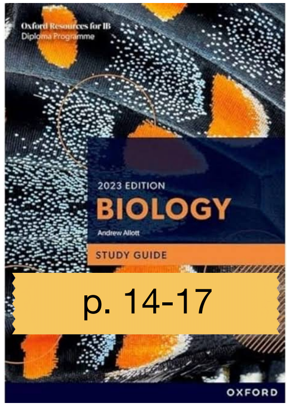
	Fibrous proteins	Globular proteins
structure	long and narrow, typically composed of repeating amino acid sequences	round/spherical, typically composed of variable, irregular amino acid sequences
properties	generally insoluble in water stable in a large range of conditions	generally soluble in water sensitive to temperature and pH changes
function	structural role (strength and support)	physiological / functional / specialized role
examples	keratin, fibrin, elastin	haemoglobin, enzymes, immunoglobulin
Key example	Collagen is a triple helix, each strand composed of repeating 3 amino acids, giving it a regular and geometric fibrous shape  → bonds hold helices together Uniformity allows it to form rope-like fibres with high tensile strength: making it an excellent structural support material around the body	Insulin is a small, globular protein, allowing it to quickly move through blood. Its specific shape allows it to bind to an insulin receptor complementarily, initiating a cellular response  The specific structure or conformation is key to the specific roles of globular proteins

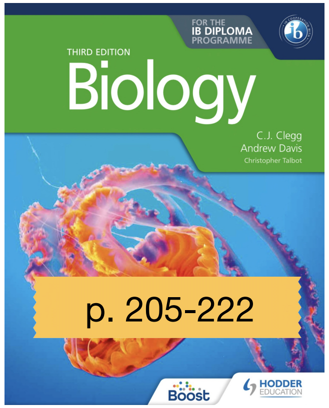
Resource Links

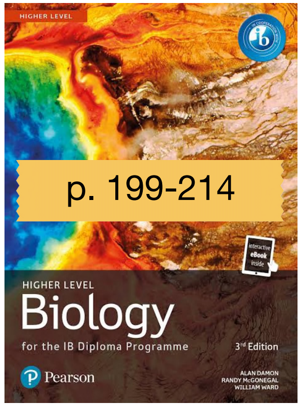
each resource is hyperlinked

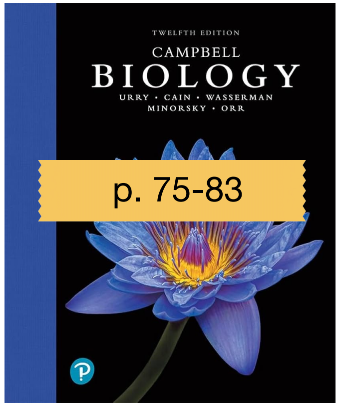
Textbooks



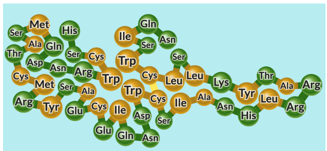




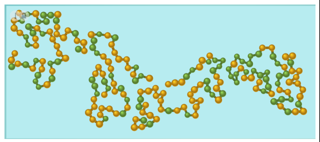




Simulators / Interactives



Protein folding



Protein denaturation

LabXchange

Protein folding

EzMol 2.1

EzMol

foldit

Foldit

CheMagic

Chemagic

Molview

Articles

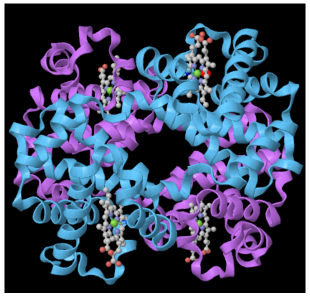
Database, A. P. S. (n.d.). AlphaFold Protein Structure Database. https://alphafold.ebi.ac.uk/?utm_source=deepmind.google&utm_medium=referral&utm_campaign=gdm&utm_content=

Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

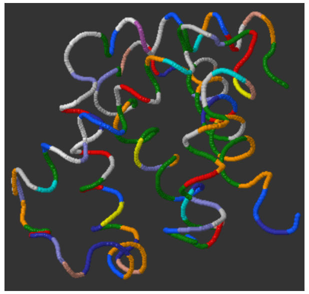
PDB101: Home page. (n.d.). RCSB: PDB-101. https://pdb101.rcsb.org/

Rd, K. M. M. (2024, May 28). 13 complete protein sources for vegetarians and vegans. Healthline. https://www.healthline.com/nutrition/complete-protein-for-vegans

3D models

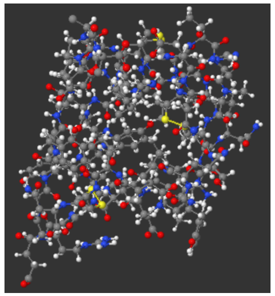


Haemoglobin

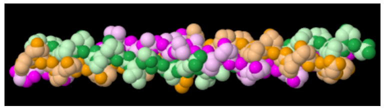


Haemoglobin

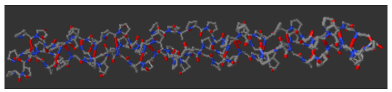
Insulin



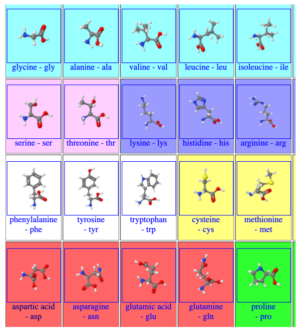
Insulin



Collagen



Collagen



All 20 amino acids