

B.1.2 · Proteins

International Baccalaureate · IB Diploma · Biology · HL

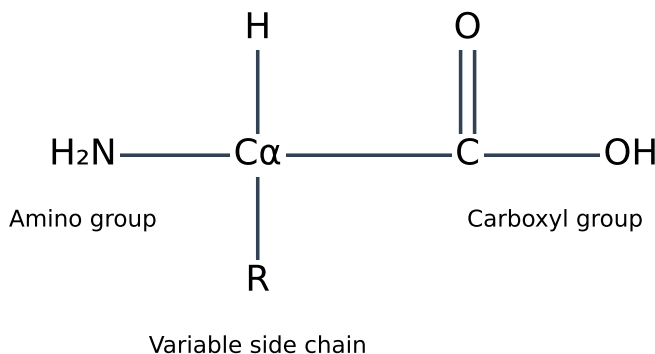
Scope and prerequisites

An egg white becomes opaque and firm when heated. Cooling it does not restore the original liquid. Yet digestion can still break its proteins into smaller molecules. Why are changing a protein's shape and cutting its chain different processes?

This unit covers B1.2 SL understandings plus the additional higher-level content (marked AHL): amino-acid structure, peptide formation, dietary amino acids, sequence variety, denaturation, side-chain chemistry, protein structure and structure–function examples. Prerequisites are covalent bonds, condensation, hydrolysis and interactions with water. Numerical datasets and short letter sequences are invented teaching examples. Sections marked **AHL** are for Higher Level only.

Amino acids have a common framework and a variable side chain

A **protein** contains one or more folded polypeptide chains. The amino acids used to build these chains share a central alpha carbon attached to four groups: an amino group, a carboxyl group, a hydrogen atom and a variable R-group or side chain. In the neutral drawing convention these groups can be shown as $-\text{NH}_2$ and $-\text{COOH}$. In aqueous solution their charge states depend on pH; a common form has $-\text{NH}_3^+$ and $-\text{COO}^-$.



Neutral representation; charge states depend on pH.

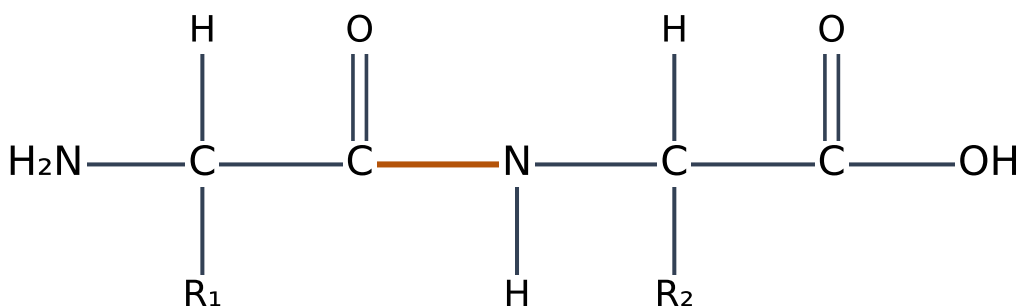
The R-group varies among amino acids. This neutral structural convention makes the four attachments clear; the amino and carboxyl groups can gain or lose hydrogen ions in solution.

The R-group distinguishes one amino acid from another. The R is a placeholder for a chemical group, not an extra element called “R”. Glycine has H as its side chain, so its alpha carbon is attached to two hydrogens. Most other standard amino acids have a more complex side chain. Counting the amino group and carboxyl group as part of R would erase the useful distinction between the shared framework and the variable part.

A good generalized drawing shows all four attachments. For the carboxyl group, show the carbonyl C=O and the hydroxyl –OH attached to the same carbon. An oxygen atom cannot silently replace that carbon. Follow the bonds rather than copying a decorative shape.

Peptide bonds join residues into a chain

In the simple condensation model, the carboxyl group of one amino acid reacts with the amino group of another. A **peptide bond** forms between the carboxyl carbon and amino nitrogen. OH from the carboxyl group and H from the amino group account for one released water molecule. The products are a dipeptide and water:



Orange C–N link: peptide bond.

Two residues; one peptide bond; one water released in the model.

The highlighted carbon–nitrogen link is the peptide bond. A carbonyl C=O remains beside it. The R-groups remain attached to their own alpha carbons.

A residue is the part of an amino acid incorporated into the chain. Repeated linkage produces a polypeptide with a repeating backbone and side chains projecting from it. The amino end and carboxyl end are different, so sequence order has a direction. Hydrolysis uses water to break peptide bonds, producing shorter peptides or free amino acids depending on how many links are cut.

The condensation equation is an accounting model. Protein synthesis in cells uses ribosomes, tRNAs, activated amino acids and energy-transfer reactions; it is not simply free amino acids spontaneously releasing water in a beaker. The model is useful for identifying the linkage, counting residues and conserving atoms.

■ Worked example: residue count is not bond count

One linear chain of 35 residues has $35 - 1 = 34$ peptide bonds. In the simple model, making the chain from free amino acids releases 34 water molecules; complete peptide-bond hydrolysis uses 34. A 120-residue chain has 119 peptide bonds, not 120.

Two separate chains of 12 and 18 residues contain $11 + 17 = 28$ peptide bonds. Joining their ends by one additional peptide bond gives one 30-residue chain with 29. Merely holding two chains together by other interactions does not create that extra backbone bond. Cross-links between side chains must not be counted as peptide bonds.

Dietary amino acids supply building material

Digestion breaks dietary proteins into smaller products, including amino acids that can be absorbed and used to synthesize the body's own proteins. Eating a particular protein does not mean that intact protein is placed directly into a matching body tissue. Cells use amino-acid supplies according to their own gene expression and metabolism.

An essential amino acid cannot be synthesized by the body in sufficient amounts and therefore must be obtained from food. A non-essential amino acid can be synthesized, for example through reactions involving other amino acids. “Non-essential” does not mean unnecessary: both types are needed for appropriate protein synthesis.

Different foods provide different proportions of amino acids. A high total protein mass alone does not prove that every essential amino acid is supplied adequately. A vegan diet needs attention to the availability and balance of essential amino acids as well as total protein. Diverse plant sources can contribute different amino-acid profiles; “plant proteins contain no essential amino acids” is false.

■ Worked example: identify the limiting supply

In an invented synthesis model, each copy of a short peptide requires two residues of amino acid X and three of Y. A cell has 50 available X residues and 60 Y residues, with every other required resource abundant. X permits $50/2 = 25$ copies; Y permits $60/3 = 20$ copies. Y limits production to 20 complete copies, leaving $50 - 2(20) = 10$ X residues. This is a simplified resource calculation, not a recommended diet or a claim about a real cell.

Length and sequence create enormous variety

The standard genetic code specifies 20 common amino acids used in polypeptide synthesis. Chains can be short or contain hundreds to thousands of residues. Both the number of residues and their order can vary. The **primary structure** is the amino-acid sequence, read in a specified direction. Two chains can have the same length and amino-acid composition but different sequences.

For example, the invented three-residue sequences A–B–C and A–C–B use the same three types but place them differently. They are not the same primary structure. With three available types, repetition allowed and three positions, there are $3^3 = 27$ possible sequences. With 20 types and four positions, there are $20^4 = 160000$ possibilities. “20 amino acids” does not mean “only 20 proteins”. For any fixed length the number is finite; allowing different lengths produces an enormous range of possible chains.

Different sequences can support different structures and functions. Insulin is a protein hormone, collagen is a structural protein, and many enzymes are proteins that catalyse reactions. Not every protein is an enzyme; some catalysts are RNA molecules rather than proteins. A functional protein also depends on appropriate folding and conditions, not just on having a long chain.

Temperature and pH can alter protein structure

Proteins interact with surrounding water and with groups within their own chains. A specific folded arrangement helps create a binding surface, active site or structural role. **Denaturation** is disruption of the normal functional conformation, often with loss of activity. Heating can disturb interactions stabilizing a fold and increase unfolding or aggregation. A pH change can alter charges and interactions. The effect depends on the protein, conditions and duration of exposure.

Ordinary denaturation does not require hydrolysis of the peptide backbone. A protein can lose its working shape while retaining its amino-acid sequence. Digestion, by contrast, cuts peptide bonds. This distinction does not mean every bond survives every extreme chemical treatment: severe conditions can cause additional damage. Explain the process actually being tested.

Some proteins can regain a functional conformation after suitable conditions are restored. Others aggregate or undergo changes that prevent recovery. Heating egg-white proteins causes unfolding and association into an opaque network; ordinary cooling does not reverse the cooked structure. “Every denaturation is reversible” and “no protein can ever refold” are both too strong.

■ Worked example: test recovery under the same assay conditions

An invented experiment exposes equal enzyme samples to different temperatures for five minutes. Every sample is then assayed at 25°C, at the same pH and substrate concentration. Relative activities are:

Pre-treatment	Activity after return to 25°C
Kept at 25°C	100 units
Heated to 45°C	87 units
Heated to 65°C	20 units
Kept cold at 5°C	98 units

The 65°C sample retains $20/100 \times 100 = 20\%$ of control activity, an 80% loss. Because all assays take place at the same temperature, the result is not explained simply by particles moving faster or slower during the assay. It supports lasting damage to function after the hot treatment. Activity alone does not reveal the complete molecular structure or prove which interaction changed. Repeats, an unheated control and checks for evaporation or changed concentration strengthen the comparison.

A low reaction rate while an enzyme is cold would not by itself establish denaturation. The recovery test asks whether function returns after the original conditions are restored. Keep treatment conditions separate from measurement conditions.

AHL: R-groups influence folding and interactions

Side chains differ in polarity, charge, size and bonding possibilities. Hydrophobic side chains are largely non-polar. Hydrophilic side chains can be polar without a net charge or carry a positive or negative charge. Acidic side-chain groups can lose a hydrogen ion and become negatively charged; basic groups can bind a hydrogen ion and become positively charged. Their charge states depend on pH, so a change in pH can alter interactions without changing the amino-acid sequence.

The sequence fixes which side chains occur at which positions. Interactions among them, the backbone and the environment constrain the conformation. Many functional proteins adopt reproducible structures under appropriate conditions, but they are not rigid statues: movement and conformational changes can be part of function. Folding may require cellular assistance, and an isolated chain is not guaranteed to fold correctly in every environment.

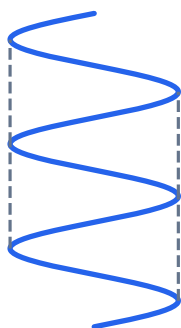
A substitution replacing one side chain with another can change packing, charge, binding or stability. Its effect depends on the position and chemical difference. One substitution can have a large effect, a small effect or little measurable effect in a chosen assay; it is not valid to predict complete loss of function from “one amino acid changed” alone.

AHL: Secondary structure uses regular backbone hydrogen bonds

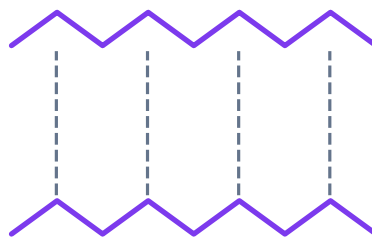
Alpha helices and beta-pleated sheets are recurring local arrangements of a polypeptide backbone. Hydrogen bonds between backbone carbonyl oxygens and amide hydrogens stabilize these structures. An alpha helix coils one segment of chain, with regular hydrogen bonding within that helix. A beta sheet contains extended strands held

together by hydrogen bonds; the strands can come from different regions of one chain or from different chains.

Alpha helix



Beta sheet



Solid: backbone paths. Dashed: hydrogen bonds.
Schematic organization; backbone atoms omitted.

Solid curves represent backbone paths; dashed lines represent stabilizing hydrogen bonds. This is an organizational cartoon, not an atom-by-atom bond diagram.

These hydrogen bonds are not the covalent peptide bonds that connect successive residues. Nor are they simply a line of R-group-to-R-group bonds. The sequence and side chains influence which secondary arrangements are favoured, but the regular stabilizing pattern shown here concerns backbone groups.

AHL: Tertiary structure depends on several interactions

Tertiary structure is the overall three-dimensional arrangement of one polypeptide chain, including how its helices, sheets and other regions pack together. Several interactions contribute:

Interaction	Structural contribution
Hydrogen bonds	Attractions involving suitable polar groups
Ionic interactions	Attractions between oppositely charged groups
Hydrophobic interactions	Non-polar groups tend to cluster away from water
Disulfide bonds	Covalent S–S cross-links formed between suitable cysteine side chains

A disulfide bond is covalent, not a weak hydrogen bond. It can connect different regions of one chain or connect different chains. Side-chain cross-linking does not turn the ordinary polypeptide backbone into a branching tree of peptide bonds. Keep the

bond type and the structural level distinct.

In a water-soluble globular protein, many hydrophobic side chains cluster in the interior while hydrophilic groups interact with water at the surface. This is a broad pattern, not a claim that every surface residue is charged or every internal atom non-polar. A membrane-spanning region of an integral protein presents hydrophobic side chains toward the lipid tails. A water-filled pore can instead have hydrophilic groups facing its lumen. The same “hydrophobic groups always face inward” rule cannot describe both environments.

AHL: Quaternary structure and protein examples

Quaternary structure is the arrangement and association of two or more polypeptide chains. Each chain has its own sequence and conformation. A protein consisting of one folded chain can have tertiary structure without quaternary structure. A non-protein group does not count as an extra polypeptide chain.

A conjugated protein includes a non-protein component needed for its normal function, such as a prosthetic group. A non-conjugated protein lacks such an associated prosthetic component in this classification. “Non-conjugated” does not mean “one chain”, “unfolded” or “contains no covalent cross-links”.

■ Insulin: a compact signal with two chains

Mature human insulin contains an A chain of 21 residues and a B chain of 30 residues. Disulfide bonds stabilize the structure, including links between the chains. Its compact globular form suits transport in aqueous blood and specific binding to receptors. It is a non-conjugated protein in this syllabus comparison: the mature hormone does not require a haem-like prosthetic group to make its two-chain structure. Zinc-associated storage assemblies are a separate level of organization, not a reason to call each A–B pair a six-chain hormone.

The two chains contain $(21 - 1) + (30 - 1) = 49$ peptide bonds. They do not contain 50 merely because $21 + 30 = 51$ residues are present. The peptide-bond rule must be applied to each separate chain. The RCSB insulin structure resource shows the two-chain organization and distinguishes it from the precursor.

■ Collagen: a fibrous structure that resists tension

A collagen molecule contains three polypeptide chains wound into a triple helix. Repeating sequence features, including glycine at every third position in the triple-helical region, allow close packing. Associations and cross-links in collagen fibrils help create strong structures that resist pulling forces. This suits collagen’s role in tendons, skin and other supporting tissues. A collagen triple helix is not simply an alpha helix renamed because it contains three chains.

Collagen is the fibrous, non-conjugated example in the syllabus comparison. Its elongated structure serves a mechanical role; it does not need to be a compact soluble messenger like insulin. Chemical modifications of residues, such as hydroxylation,

must not automatically be confused with adding a separate prosthetic group. The RCSB collagen account explains the three-chain structure.

■ Haemoglobin: globular chains plus haem

The main adult human haemoglobin has two alpha and two beta globin chains. Each chain carries a non-protein haem group containing iron that can bind one oxygen molecule reversibly. Four haem groups therefore give a maximum of four bound oxygen molecules per haemoglobin molecule. The protein is conjugated, and its four-chain association is quaternary structure. The four haem groups do not make it an eight-polypeptide protein.

Its soluble globular form supports transport in the red-cell cytoplasm, while interactions between subunits help relate oxygen binding at one site to behaviour at others. Haem provides a binding component that the amino-acid chain alone does not supply. The RCSB haemoglobin resource connects structural observations with oxygen binding.

Protein structures can be studied with methods including X-ray crystallography and cryogenic electron microscopy. These techniques use measurements and reconstruction to reveal details invisible to the unaided eye. A ribbon cartoon summarizes a structural model; it is not an ordinary light-microscope photograph, and a two-dimensional drawing does not mean the molecule is flat.

Practice

Numerical supplies, activity data and short letter sequences below are invented. Use the simple condensation-accounting model where specified.

1. Draw a generalized amino acid. Label the alpha carbon, hydrogen, amino group, carboxyl group and R-group. Explain what R represents and why pH can change the charge of the amino and carboxyl groups.
2. Write the word equation for forming a dipeptide. Identify the peptide bond, the groups supplying the atoms of water, and the groups that remain attached to each alpha carbon.
3. A linear polypeptide contains 120 residues. Calculate its peptide bonds and the water molecules released in the simple formation model. How many waters are needed for complete peptide-bond hydrolysis?
4. Two linear chains contain 12 and 18 residues. Count their peptide bonds before and after joining their ends by one peptide bond. Distinguish this from holding the two chains together without a new backbone link.
5. Define essential and non-essential amino acids. Explain why total protein mass alone does not establish the adequacy of every essential amino acid, including in a vegan diet.

6. A synthesis model requires two X and three Y residues per peptide. Available supplies are 50 X and 60 Y. Calculate the maximum complete peptides and unused X. State one limitation of this model.
7. Using three permitted amino-acid types, repetition allowed, calculate the number of four-position sequences. Explain why same length and composition do not necessarily mean same primary structure.
8. Distinguish denaturation from peptide-bond hydrolysis. Explain why a cooked egg does not regain its initial liquid structure merely on cooling, without claiming that all proteins can never refold.
9. Samples assayed at 25°C after treatment give activities 100, 87, 20 and 98 units after 25°C, 45°C, 65°C and 5°C treatments respectively. Calculate retained activity and percentage loss for the 65°C sample. Explain why assay temperature matters.
10. Design a comparison to test whether low activity after heating persists on cooling. Give four controlled variables and one source of uncertainty or alternative explanation.
11. Correct these claims: every protein is an enzyme; all catalysts are proteins; a low rate while an enzyme is cold proves denaturation; non-essential amino acids are unnecessary.
12. **AHL** Compare alpha helices, beta sheets and tertiary structure. Locate the regular hydrogen bonds in secondary structure. Explain how pH, ionic interactions and cysteine disulfide bonds can affect a fold.
13. **AHL** Compare the likely locations of hydrophobic and hydrophilic side chains in a water-soluble globular protein and a membrane-spanning protein with a water-filled pore. Explain why a substitution's effect depends on position and chemistry.
14. **AHL** Compare insulin, collagen and adult human haemoglobin in shape, chain organization, conjugation and function. Calculate the peptide bonds in insulin's 21- and 30-residue chains and the maximum oxygen molecules bound by one four-haem haemoglobin. Explain why neither haem groups nor a ribbon drawing should be counted as extra chains.

■ Worked answers

1. The alpha carbon has four attachments: H, R, an amino group and a carboxyl group. The carboxyl carbon is bonded to O by a double bond and to OH by a single bond in the neutral representation. R is the variable side chain, not an element. Binding or loss of hydrogen ions changes charge; a common aqueous form has an ammonium and a carboxylate group.
2. Amino acid + amino acid → dipeptide + water. The new C–N link connects a carboxyl carbon with an amino nitrogen. OH from the first carboxyl and H from the second amino group account for water. Each residue retains its own side chain on its alpha carbon. The product still contains a carbonyl beside the peptide C–N link.

3. $120 - 1 = 119$ peptide bonds; formation from free amino acids releases 119 waters in the simple model. Complete hydrolysis uses 119 waters. Cutting only some bonds produces shorter peptides, not necessarily free monomers.
4. Initially $(12 - 1) + (18 - 1) = 11 + 17 = 28$ bonds. One new end-to-end peptide bond gives 29 in the 30-residue chain and releases one additional water in the model. Association without that new backbone bond leaves two chains and 28 peptide bonds.
5. Essential amino acids must come from food because synthesis is insufficient; non-essential ones can be synthesized but are still needed. Amino-acid proportions differ among foods. Total protein mass does not identify the supply of each required essential type. Vegan diets therefore need attention to amino-acid balance and supply; plant sources do contain essential amino acids.
6. X permits 25 copies; Y permits 20, so the maximum is 20. X used = $2(20) = 40$; X left = $50 - 40 = 10$. Real synthesis also depends on other amino acids, energy, enzymes and regulation, which the model treats as non-limiting.
7. $3^4 = 81$ possible sequences. With the same composition, A-B-C-A and A-A-B-C have different orders and thus different primary structures when read in the same direction. Neither length nor composition alone specifies all positions.
8. Denaturation changes functional conformation and can leave the peptide backbone intact. Hydrolysis cuts peptide bonds using water. Heated egg-white proteins unfold and associate into an altered network that ordinary cooling does not undo. Some other proteins can refold under suitable conditions, so irreversibility must be tied to the example and conditions.
9. Retained activity = $20/100 \times 100 = 20\%$; loss = $(100 - 20)/100 \times 100 = 80\%$. All assays at 25°C allow comparison after treatment instead of confusing pretreatment with an immediate temperature effect during measurement. The result supports persistent functional damage but does not identify a complete molecular mechanism alone.
10. Heat equal samples for a fixed time, return them and an unheated control to the same assay conditions, then compare activity with repeats. Control enzyme concentration/volume, treatment duration, assay temperature, pH, substrate concentration and measurement interval; any four appropriate controls suffice. Evaporation, concentration errors or unequal timing could alter results and should be checked.
11. Proteins include structural proteins and hormones as well as enzymes. Some catalytic RNAs are not proteins. Cold conditions can reversibly slow a reaction; test recovery after restoring conditions. Non-essential means synthesizable, not dispensable.
12. **AHL** An alpha helix is a coiled local backbone arrangement; a beta sheet contains extended strands. Regular backbone $\text{C}=\text{O} \cdots \text{H}-\text{N}$ hydrogen bonding stabilizes both. Tertiary structure describes the overall fold of one chain. A pH change can alter side-chain protonation and ionic attractions; hydrogen bonding and hy-

drophobic interactions also contribute. A disulfide bond is a covalent S–S link between cysteine side chains, not a peptide bond or a weak hydrogen bond.

13. **AHL** A water-soluble globular protein often has hydrophobic groups toward its core and hydrophilic groups exposed to water. A membrane-spanning surface facing lipid tails is hydrophobic, while a water-filled pore can present hydrophilic groups inward. Thus environment determines which surface contacts water. A substitution at a buried packing site, binding site or lipid-facing region can have different effects; not every substitution destroys function.
14. **AHL** Insulin: compact globular hormone, two chains, non-conjugated in the stated comparison, specific receptor binding. Collagen: three-chain triple helix, fibrous, non-conjugated, suited to resisting tension in supporting tissues. Adult haemoglobin: four globin chains and four haem groups, globular and conjugated, reversible oxygen binding. Insulin has $(21 - 1) + (30 - 1) = 49$ peptide bonds, not 50. Four haem groups bind at most four oxygen molecules, not eight oxygen molecules because oxygen is diatomic. Haem is a non-protein component; a ribbon is a way of displaying a structural model, not an extra molecule or a flat physical chain.

Exam tips

Identify which carbon and nitrogen make a peptide link, and retain the carbonyl oxygen in a dipeptide drawing. State the number of separate chains before using residues minus chains to count peptide bonds. Separate treatment conditions from assay conditions, and distinguish loss of a functional fold from cutting a backbone. In a structure–function explanation, connect a specific feature to the interaction or mechanical property it permits.

- **AHL:** Distinguish backbone hydrogen bonds, side-chain interactions and covalent disulfide bonds. A hydrophobic surface can face membrane lipids rather than a protein's core. Count polypeptides separately from prosthetic groups and use the appropriate example for each level of structure.

Scope reference: IB Biology guide, B1.2. Structural examples are linked in their teaching sections; invented data are not measurements from those sources.