Biochemistry

Unit 2 · Amino Acids and Proteins

Twenty small molecules, joined in different orders, make every enzyme, antibody, hormone and muscle fibre in your body. After this unit you can classify any amino acid, calculate its isoelectric point, draw a peptide bond, explain the four levels of protein structure and read a Ramachandran plot, and plan how to purify and measure a protein.

Taught at 8 of 10 | Usually Sem 3 | 8 chapters

About this unit

This unit is about the molecules that do most of the work in a cell. It starts with the 20 amino acids, how they are built and grouped (Chapter 2.1), and how they behave in water, with acids, bases and reagents (2.2). Chapter 2.3 joins them through the peptide bond. Chapters 2.4 and 2.5 show how a chain of residues turns into a shape: the four levels of protein structure, and the Ramachandran plot that says which backbone angles are allowed. Chapter 2.6 explains how a chain finds its shape and how heat, urea or acid destroys it. Chapter 2.7 sorts proteins by shape, composition and job, with the colour tests used in the lab, and Chapter 2.8 shows how a single protein is pulled out of a cell and measured.

Chapters

1. Amino acids: structure and classification
2. Physical, chemical and biological properties of amino acids ADVANCED
3. Peptide bond and peptides
4. Protein structure: primary, secondary, tertiary, quaternary
5. Ramachandran plot ADVANCED
6. Protein folding and denaturation
7. Classification and properties of proteins
8. Protein purification and quantification ADVANCED

How to use this book

- Read the lead paragraph and the figure of each chapter first; they carry the one idea.
- Cover the solution of the worked example and do it yourself with a pencil, then compare step by step.
- Answer each "Check yourself" question in your head or on paper before you open it.

2.1 Amino acids: structure and classification

An amino acid is a small molecule with an amino group and a carboxyl group on the same carbon. Twenty of them, the standard amino acids, are coded by DNA and joined into proteins. They all share one backbone and differ only in one side chain, called R. The side chain decides whether the amino acid likes water, carries a charge or can form a cross-link, so learning the 20 side chains in five groups is the key to everything later in this unit: folding, protein properties and purification all follow from it.

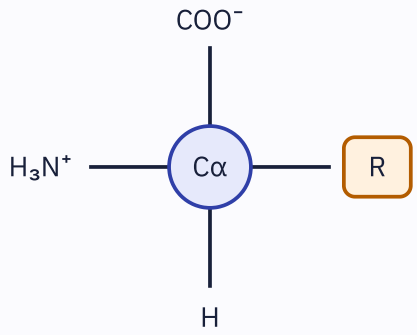
You need: pH, pKa and weak interactions from Unit 1 (Water, weak interactions, pH and biological buffers).

The common structure

Every standard amino acid is an α-amino acid. The central carbon is the α-carbon (Cα). Four groups hang on it: a carboxyl group (–COOH), an amino group ($–NH_2$), a hydrogen atom and the side chain R. In water at pH 7, the carboxyl group has lost its proton ($–COO^-$) and the amino group has gained one ($–NH_3^+$). A molecule with both a positive and a negative charge but no net charge is a **zwitterion** (German for "hybrid ion"). Chapter 2.2 uses this to explain the isoelectric point.

Proline is the exception in shape. Its side chain loops back and bonds to its own amino nitrogen, making a five-membered ring. Its nitrogen is therefore a secondary amine, and older books call proline an **imino acid**. The ring makes proline rigid, which matters for the α-helix (Chapter 2.4).

Common structure at pH 7



Same backbone in all 20;
only the side chain R differs.

Grouped by side chain (R)

Nonpolar, aliphatic
Gly, Ala, Pro, Val, Leu, Ile, Met

Aromatic
Phe, Tyr, Trp

Polar, uncharged
Ser, Thr, Cys, Asn, Gln

Positively charged (basic)
Lys, Arg, His

Negatively charged (acidic)
Asp, Glu

Top two groups: hydrophobic. Bottom three: hydrophilic.

Figure 2.1.1 Left: every standard amino acid has the same four groups on its α-carbon; at pH 7 it is a zwitterion. Right: the five side-chain groups used by Lehninger and most syllabi. Notice that Gly (R = H) has no chiral centre and that Tyr and Trp are aromatic but still slightly polar.

Classification by side chain

The table gives the three-letter and one-letter codes, which you must know for sequences.

GROUP	AMINO ACIDS (3-LETTER, 1-LETTER)	WHAT THE SIDE CHAIN DOES
Nonpolar, aliphatic	`Gly G, Ala A, Pro P, Val V, Leu L, Ile I, Met M`	Hydrocarbon chains; pack together in the core of a protein. Met carries a thioether sulfur.
Aromatic	`Phe F, Tyr Y, Trp W`	Ring side chains; absorb UV light near 280 nm (Tyr, Trp). Tyr has an –OH, Trp an indole N–H.
Polar, uncharged	`Ser S, Thr T, Cys C, Asn N, Gln Q`	Form hydrogen bonds with water. Cys has a thiol (–SH) that forms disulfide bridges. Asn and Gln are the amides of Asp and Glu.
Positively charged	`Lys K, Arg R, His H`	Extra amino (Lys), guanidino (Arg) or imidazole (His) group; His side chain has pKa near 6, so it can gain or lose a proton near pH 7.
Negatively charged	`Asp D, Glu E`	Extra carboxyl group, ionised ($–COO^-$) at pH 7.

Older Indian textbooks also group amino acids by the number of amino and carboxyl groups: **neutral** (one of each, such as Gly, Ala), **acidic** (an extra carboxyl: Asp, Glu), and **basic** (an extra basic group: Lys, Arg, His). Some add structural groups: sulfur-containing (Cys, Met), hydroxy

(Ser, Thr), branched-chain (Val, Leu, Ile). These are the same 20 sorted a different way; learn the five-group table and derive the rest.

Stereochemistry: L and D

Because Cα carries four different groups, it is a chiral centre and the amino acid has two mirror-image forms, called enantiomers. Glycine is the only standard amino acid that is not chiral, since its R is a second hydrogen. Proteins use only the **L** form. In a Fischer projection drawn with $-COO^-$ at the top and R at the bottom, the L form has the amino group on the left. L and D describe relation to L-glyceraldehyde, not the direction a solution rotates light; L-alanine is dextrorotatory, for example. D-amino acids do occur in nature: D-alanine and D-glutamate in the peptidoglycan of bacterial cell walls (Unit 5) and in some peptide antibiotics.

Threonine and isoleucine each have a second chiral carbon in the side chain, so each has four stereoisomers. Only one of the four is used in proteins.

Nutritional, metabolic and other groupings

Nutritional. Humans cannot make nine amino acids fast enough and must eat them: His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Val. These are the **essential** (indispensable) amino acids. Many Indian textbooks call Arg and His semi-essential, because a growing child needs more than the body makes. Cereals such as rice and wheat are low in lysine, and pulses such as dal are low in methionine; that is why rice with dal gives better protein than either alone.

Metabolic. When the carbon skeleton is broken down, it can feed glucose synthesis (glucogenic), ketone body synthesis (ketogenic) or both. Leu and Lys are purely ketogenic; Ile, Phe, Trp and Tyr are both; the rest are glucogenic. The pathways are in Unit 6 (Amino acid catabolism).

Beyond the 20. Selenocysteine (coded by a UGA codon with a special signal) and pyrrolysine (UAG, in some archaea and bacteria) are sometimes called the 21st and 22nd amino acids. Other residues are made by changing a standard one after the protein is built: 4-hydroxyproline and 5-hydroxylysine in collagen, phosphoserine in signalling proteins. Some amino acids are never in proteins at all: ornithine and citrulline (urea cycle, Unit 6) and GABA, a brain neurotransmitter.

WORKED EXAMPLE

Question (6 marks). For Val, Ser, Lys, Glu, Tyr and His: (a) give the side-chain group, (b) state the charge on the side chain at pH 7.4, (c) mark which are essential in humans. (d) How many stereoisomers does threonine have?

1. **Val**: nonpolar aliphatic (branched chain); side chain uncharged; essential.
2. **Ser**: polar uncharged ($-CH_2OH$); uncharged; not essential (made from a glycolysis intermediate).
3. **Lys**: positively charged; side-chain amino group (pKa about 10.5) is protonated at pH 7.4, so +1; essential.
4. **Glu**: negatively charged; side-chain carboxyl (pKa about 4.3) is ionised, so −1; not essential.
5. **Tyr**: aromatic; phenol –OH (pKa about 10) stays protonated, so uncharged; not essential (made from Phe).
6. **His**: positively charged group, but its imidazole pKa is about 6.0. At pH 7.4, which is 1.4 units above the pKa, only about 1 in 26 molecules carries the proton (ratio $10^{1.4} \approx 25$ to 1). So the side chain is mostly neutral, with a small positive fraction. Essential.
7. **Threonine** has two chiral centres (Cα and Cβ), so $2^2 = 4$ stereoisomers; proteins use only L-threonine.

EXAM TRAPS

- Glycine is not chiral; do not draw L- and D-glycine.
- L does not mean levorotatory. The L/D label is about configuration, not the sign of optical rotation.
- Histidine is grouped as "basic", but at pH 7 its side chain is mostly uncharged. Examiners like asking why His is common in enzyme active sites: because it can both give and take a proton near pH 7.
- Asn and Gln are polar but not acidic. They are amides, not carboxylic acids.
- Write "essential" as nutritionally essential. All 20 are needed in proteins; "essential" only means the diet must supply it.

CHECK YOURSELF

– REMEMBER **List the nine essential amino acids for adults.** 2 marks

His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Val. (Arg is called semi-essential in many Indian textbooks.)

– REMEMBER **Which standard amino acid is not optically active, and why?** 2 marks

Glycine; its α-carbon carries two hydrogen atoms, so it has no four different groups and no chiral centre.

– UNDERSTAND **Why is proline called an imino acid?** 2 marks

Its side chain is bonded to the α-amino nitrogen, forming a ring, so the nitrogen carries only one hydrogen (a secondary amine, –NH–) instead of a free $–NH_2$.

– UNDERSTAND **Explain the difference between glucogenic and ketogenic amino acids with one example of each.** 4 marks

Glucogenic amino acids give pyruvate or citric acid cycle intermediates that can be turned into glucose (Ala, Ser). Ketogenic ones give acetyl-CoA or acetoacetate, which form ketone bodies but not glucose (Leu, Lys).

– APPLY **Write the one-letter sequence for Met-Lys-Trp-Glu-Asn-Gln.** 2 marks

M K W E N Q, written MKWENQ.

– APPLY **At pH 7, what is the net charge of free lysine and of free aspartate?** 3 marks

Lysine: α-COO^- (–1), α-NH_3^+ (+1), side-chain NH_3^+ (+1), net +1. Aspartate: α-COO^- (–1), α-NH_3^+ (+1), side-chain COO^- (–1), net –1.

– ANALYZE **A protein has many Leu, Val and Phe residues inside and many Lys and Glu on its surface. Explain this pattern.** 4 marks

Nonpolar side chains (Leu, Val, Phe) avoid water and cluster in the core (the hydrophobic effect). Charged side chains (Lys, Glu) interact with water and with ions, so they sit on the surface, where they also keep the protein soluble.

BEYOND THE SYLLABUS

Chemists can now add amino acids that nature does not use. In genetic code expansion, a cell is given an extra tRNA and enzyme pair that puts a chosen unnatural amino acid at a stop codon, for example to attach a fluorescent label or a drug to an antibody. D-amino acids are used to make peptide drugs that the body's proteases cannot easily cut. Search "genetic code expansion" and "noncanonical amino acids". This is not for the exam.

2.2 Physical, chemical and biological properties of amino acids

Amino acids behave the way they do because they are zwitterions: each carries a positive and a negative charge at once. That explains their high melting points, water solubility and acid–base behaviour. Their amino and carboxyl groups also react with lab reagents in ways used to detect and sequence them, and inside the body they are raw material for hormones, neurotransmitters and haem.

You need: the side-chain groups from Chapter 2.1 and the Henderson–Hasselbalch equation from Unit 1.

Physical properties

- **State and melting point.** Colourless crystalline solids that melt (usually with decomposition) above 200 °C, because the crystal is held by ionic attraction between zwitterions, like a salt.
- **Solubility.** Soluble in water, poorly soluble in nonpolar solvents such as ether. Least soluble at the isoelectric point.
- **Optical activity.** All except glycine rotate plane-polarised light (Chapter 2.1).
- **UV absorption.** Tryptophan and tyrosine absorb strongly near 280 nm; this is why protein solutions are measured at 280 nm (Chapter 2.8).

Amphoteric behaviour and the isoelectric point

An amino acid can give a proton (from $-NH_3^+$) or take one (on $-COO^-$), so it is **amphoteric**. Titrating glycine with NaOH shows two buffering plateaus. Near $pKa_1 = 2.34$ the carboxyl loses its proton; near $pKa_2 = 9.60$ the amino group does. Between them sits the **isoelectric point (pI)**, the pH at which the average net charge is zero and the molecule does not move in an electric field.

```
For an amino acid with no ionisable side chain:
  pI = (pKa₁ + pKa₂) / 2
For an acidic or basic amino acid:
  pI = average of the two pKa values on either side of the neutral (zwitterion) form
```

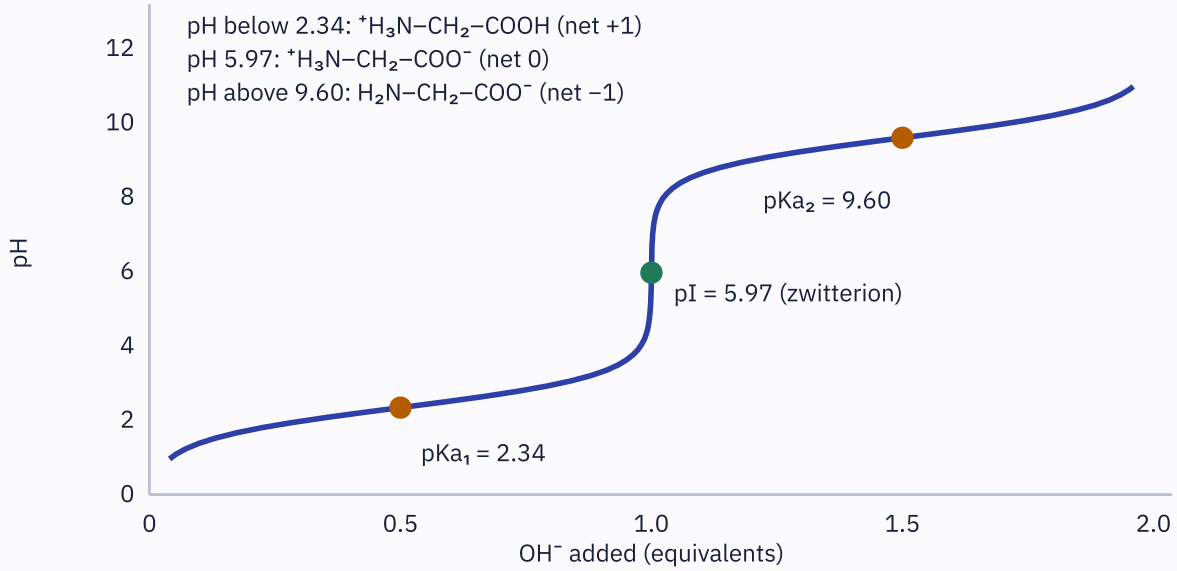


Figure 2.2.1 Titration curve of glycine, computed from $pKa_1 = 2.34$ and $pKa_2 = 9.60$ with the Henderson–Hasselbalch equation (water's own buffering ignored). The flat parts are where glycine buffers best; the steep part at 1 equivalent is the isoelectric point.

Chemical properties

- **Ninhydrin.** Heated with ninhydrin, free α-amino acids give a purple colour (Ruhemann's purple). Proline, with its ring nitrogen, gives yellow. Used to detect amino acids on paper chromatograms and in amino acid analysers.
- **Sanger's reagent (FDNB)** and **Edman's reagent (phenyl isothiocyanate)** attach to a free amino group; they identify the N-terminal residue of a peptide (Chapter 2.3).
- **Decarboxylation** removes CO_2 and gives an amine: histidine → histamine, tryptophan (via 5-hydroxytryptophan) → serotonin, glutamate → GABA.
- **Thiol oxidation.** Two cysteine molecules oxidise to cystine, joined by a disulfide (–S–S–) bond.

Biological roles, and glycine and cysteine

Besides building proteins, amino acids are precursors: tyrosine gives adrenaline, thyroxine and melanin; tryptophan gives serotonin and niacin; arginine gives nitric oxide. **Glycine** is the smallest amino acid; it is used to make haem, purines and creatine, conjugates bile acids (glycocholic acid) and is an inhibitory neurotransmitter in the spinal cord. **Cysteine** forms

disulfide bridges that hold proteins such as insulin and keratin together, and supplies the –SH of glutathione (γ-Glu-Cys-Gly), the cell's main antioxidant, and of coenzyme A.

WORKED EXAMPLE

Question (5 marks). (a) Calculate the pI of glycine (pKa 2.34 and 9.60). (b) Aspartate has pKa values 1.88 (α-COOH), 3.65 (side-chain COOH) and 9.60 (α-NH_3^+). Find its pI. (c) In electrophoresis at pH 6.0, which way does each move?

1. Glycine: pI = (2.34 + 9.60) / 2 = 11.94 / 2 = **5.97**.
2. Aspartate: list forms from low pH. Net +1 (all protonated) → lose α-COOH proton at 1.88 → net 0 → lose side-chain proton at 3.65 → net −1 → lose amino proton at 9.60 → net −2.
3. The neutral form lies between pKa 1.88 and 3.65, so pI = (1.88 + 3.65) / 2 = **2.77**.
4. At pH 6.0 glycine is very slightly above its pI, so it is almost neutral and hardly moves. Aspartate is far above its pI, so it is negative and moves toward the anode (+).

EXAM TRAPS

- For acidic and basic amino acids, do not average all three pKa values. Average only the two around the neutral form.
- At the pI, the molecule is not uncharged; it carries + and − charges that cancel.
- Ninhydrin with proline is yellow, not purple. Examiners ask why: proline has a secondary amino group.
- Cystine (two residues, S–S) and cysteine (one residue, SH) are different molecules.

CHECK YOURSELF

– REMEMBER **Define isoelectric point.** 1 mark

The pH at which an amino acid or protein has zero net charge and does not move in an electric field.

– UNDERSTAND **Why do amino acids have high melting points compared with organic molecules of similar size?** 2 marks

In the solid they are zwitterions held by strong ionic attractions, like a salt crystal, not by weak van der Waals forces.

– UNDERSTAND **Why does ninhydrin give a yellow colour with proline but purple with other amino acids?** 2 marks

Proline has a secondary amino group in a ring, so it forms a different (yellow) product; primary α-amino groups give Ruhemann's purple.

– APPLY **Lysine has pKa values 2.18, 8.95 and 10.53. Find its pI.** 3 marks

Forms: +2 → +1 (after 2.18) → 0 (after 8.95) → −1 (after 10.53). Neutral form lies between 8.95 and 10.53, so pI = (8.95 + 10.53)/2 = 9.74.

– ANALYZE **A mixture of Asp, Gly and Lys is run in electrophoresis at pH 6. Predict the order of movement.** 4 marks

Asp (pI 2.77) is negative and moves to the anode; Gly (pI 5.97) is about neutral and stays near the start; Lys (pI 9.74) is positive and moves to the cathode.

BEYOND THE SYLLABUS

Modern amino acid analysis hydrolyses a protein in 6 M HCl, separates the free amino acids by HPLC and detects them after a labelling reaction, or by mass spectrometry. Glutamate and lysine for food and animal feed are made on an industrial scale by fermentation with the bacterium Corynebacterium glutamicum. Search "amino acid fermentation" and "pKa shifts in proteins". This is not for the exam.

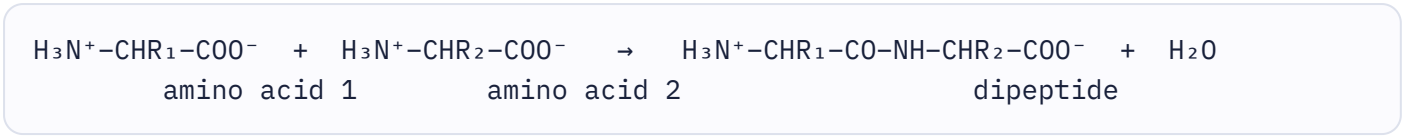
2.3 Peptide bond and peptides

A peptide bond is the amide bond that joins the carboxyl group of one amino acid to the amino group of the next, with the loss of one water molecule. Repeating it gives a chain called a peptide, and long chains are proteins. The peptide bond is not an ordinary single bond: it is flat and rigid, and that one fact limits how a protein chain can bend. Many small peptides are also important on their own, as hormones, antioxidants and sweeteners.

You need: amino acid structure (Chapter 2.1) and the reactions of the amino group (Chapter 2.2).

How the bond forms

The α-carboxyl of amino acid 1 reacts with the α-amino group of amino acid 2. An –OH leaves from the carboxyl and an –H from the amine; together they leave as H_2O. This is a **condensation** (dehydration) reaction, and the product –CO–NH– is an amide.

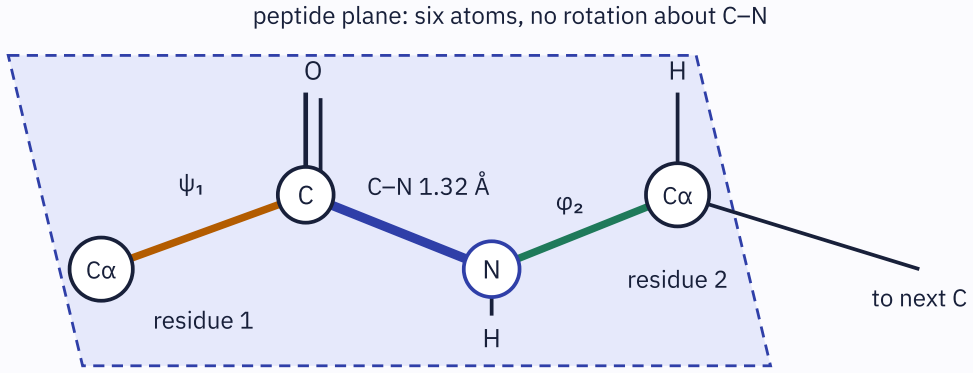


In water, the reverse reaction (hydrolysis) is favoured by free energy, but it is very slow without a catalyst, so proteins are stable for years. Cells do not make peptide bonds by simply mixing amino acids. The ribosome joins amino acids that have first been activated by attachment to tRNA, a step that uses ATP. Breaking peptide bonds needs proteases, or boiling in strong acid (6 M HCl) for many hours, as in amino acid analysis.

The character of the peptide bond

The lone pair of electrons on the amide nitrogen is shared with the carbonyl group, so the bond is a resonance hybrid of two forms: C(=O)–N and C(–O^-)=N^+. This gives the C–N bond **partial double-bond character**. Its length, about 1.32 Å, lies between a C–N single bond (about 1.47 Å) and a C=N double bond (about 1.27 Å). Four results follow.

1. **Planar.** Six atoms lie in one plane: the Cα of residue 1, its C and O, the N and H of residue 2, and the Cα of residue 2. This flat unit is the **peptide plane** (amide plane).
2. **Rigid.** There is no free rotation about the C–N bond. The chain can rotate only about the two single bonds on either side of each Cα: N–Cα (angle φ, phi) and Cα–C (angle ψ, psi). These two angles are the subject of the Ramachandran plot (Chapter 2.5).
3. **Trans.** The two Cα atoms usually sit on opposite sides of the C–N bond (trans), which avoids a clash between side chains. The cis form is seen mainly in bonds to proline, whose ring makes cis and trans closer in energy.
4. **Polar but uncharged.** The C=O can accept a hydrogen bond and the N–H can donate one. These backbone hydrogen bonds build the α-helix and β-sheet (Chapter 2.4). Unlike free amino acids, the peptide bond itself does not ionise between pH 0 and 14.



Orange: Cα–C bond, rotates (angle ψ). Green: N–Cα bond, rotates (angle φ).
Blue: peptide C–N bond, partial double bond, fixed. Cα atoms are trans.

Figure 2.3.1 The peptide unit. The six atoms from one Cα to the next lie in a flat plane because the C–N bond is partly double. Rotation happens only at the bonds that touch each Cα, so a protein chain is like a row of flat cards hinged at the α-carbons.

Naming and writing peptides

A chain is always written from the **N-terminus** (free α-amino group) on the left to the **C-terminus** (free α-carboxyl) on the right. Each amino acid in a chain is a **residue**, because it has lost the parts of water. Name a peptide by changing the ending of every residue except the last to "-yl": Gly-Ala-Ser is glycylalanylserine. Two residues make a dipeptide, a few make an oligopeptide, and many make a polypeptide; chains of more than about 50 residues are usually called proteins, though the line is not sharp. With 20 choices at each position, there are 20^n possible sequences of length n: 400 dipeptides and 8,000 tripeptides.

A chain of n residues has n − 1 peptide bonds. Its charge comes from the two ends and from ionisable side chains only.

Biologically important peptides

PEPTIDE	SIZE AND MAKE-UP	ROLE
Glutathione	3 residues: γ-Glu-Cys-Gly; the Glu is linked through its side-chain (γ) carboxyl	Main antioxidant in cells; keeps proteins' –SH groups reduced
Oxytocin, vasopressin	9 residues each, with a disulfide ring; they differ at two positions	Hormones of the posterior pituitary: uterine contraction and milk release; water retention by the kidney
Insulin	51 residues in two chains (A 21, B 30) held by two disulfide bridges, with a third inside chain A	Lowers blood glucose; first protein to be sequenced (Sanger)

PEPTIDE	SIZE AND MAKE-UP	ROLE
Enkephalins	5 residues (Met-enkephalin: Tyr-Gly-Gly-Phe-Met)	Natural pain-relieving peptides of the brain
Aspartame	Asp-Phe methyl ester (a dipeptide)	Artificial sweetener; labels warn people with phenylketonuria

Finding the ends and making peptides in the lab

End-group analysis. Sanger's reagent (1-fluoro-2,4-dinitrobenzene) labels the free N-terminal amino group; after acid hydrolysis, the yellow DNP-amino acid shows which residue was first. Edman degradation uses phenyl isothiocyanate to remove the N-terminal residue one at a time without breaking the rest of the chain, so repeated cycles read the sequence.

To join a chosen carboxyl to a chosen amine, every other reactive group must be blocked by a **protecting group**. The α-amino group is protected by Boc (removed by trifluoroacetic acid) or Fmoc (removed by the base piperidine). Using groups that come off under different conditions, so one can be removed while the others stay, is called **orthogonal protection**. The carboxyl is activated by a **coupling reagent**: a carbodiimide such as DCC or DIC, often with HOBt added to reduce racemisation, or a phosphonium reagent such as PyBOP.

In **solid-phase peptide synthesis** (SPPS), introduced by R. B. Merrifield, the C-terminal residue is fixed to an insoluble resin bead. Each cycle removes the amino protection, couples the next protected amino acid and washes away excess reagent, so no purification is needed between steps. At the end, acid cleaves the finished peptide from the resin. Chemical synthesis builds the chain from the C-terminus to the N-terminus, the opposite of the ribosome.

WORKED EXAMPLE

Question (6 marks). For the tripeptide Gly-Ala-Ser: (a) name it and mark its N- and C-terminal residues; (b) how many peptide bonds does it have? (c) Calculate its molar mass, given Gly 75.07, Ala 89.09, Ser 105.09 and H_2O 18.02 g/mol. (d) What is its net charge at pH 7? (e) How many different tripeptides can be made using Gly, Ala and Ser once each?

1. Name: glycylalanylserine. N-terminal residue Gly, C-terminal residue Ser.
2. Peptide bonds = n − 1 = 3 − 1 = 2.
3. Sum of the free amino acids = 75.07 + 89.09 + 105.09 = 269.25 g/mol. Two bonds release 2 × 18.02 = 36.04 g/mol of water. Molar mass = 269.25 − 36.04 = **233.21 g/mol**.
4. At pH 7: N-terminal $-NH_3^+$ (+1), C-terminal $-COO^-$ (−1), and no ionisable side chains (Ser's –OH does not ionise). Net charge **0**.
5. Three different residues in three positions: 3! = 3 × 2 × 1 = **6** tripeptides (GAS, GSA, AGS, ASG, SGA, SAG). They have the same mass but different properties.

EXAM TRAPS

- A peptide bond is between the α-carboxyl and the α-amino group. Glutathione's Glu–Cys link uses the γ-carboxyl, so it is not a normal α-peptide bond; examiners ask about this.
- Do not say the peptide bond rotates freely. Rotation is about the N–Cα and Cα–C bonds, not the C–N bond.
- Gly-Ala and Ala-Gly are different dipeptides. The N-terminus is always written first.
- When calculating the mass of a peptide, subtract one water per peptide bond, not per residue.

CHECK YOURSELF

– REMEMBER **Which two groups react to form a peptide bond, and what is released?** 2 marks

The α-carboxyl group of one amino acid and the α-amino group of the next; one molecule of water is released.

– REMEMBER **Name two biologically active peptides and their functions.** 2 marks

Glutathione, an antioxidant; oxytocin, which causes uterine contraction and milk release. (Vasopressin, insulin or enkephalin are also correct.)

– UNDERSTAND **Why is the peptide bond planar?** 3 marks

Resonance shares the nitrogen lone pair with the carbonyl, giving the C–N bond partial double-bond character. Like a double bond, it cannot rotate, so the six atoms of the peptide unit stay in one plane.

– APPLY **How many peptide bonds and how many possible sequences are there for a pentapeptide made from the 20 standard amino acids?** 2 marks

4 peptide bonds; 20^5 = 3,200,000 possible sequences.

– APPLY **Find the net charge of Lys-Gly-Asp at pH 7.** 3 marks

N-terminal NH_3^+ +1, Lys side chain +1, Asp side chain –1, C-terminal COO^- –1. Net 0.

– ANALYZE **Why must the side-chain amino group of lysine be protected during peptide synthesis, and why are Fmoc and a tBu-type side-chain group called orthogonal?** 4 marks

Otherwise the coupling reagent would join the next amino acid to the side-chain amine as well, giving a branched product. Fmoc comes off with base (piperidine) while acid-labile side-chain groups stay; these come off later with acid (TFA). Because each is removed under conditions that leave the other intact, they are orthogonal.

BEYOND THE SYLLABUS

Peptide drugs are a fast-growing class of medicines; semaglutide, a modified 31-residue peptide used for diabetes and weight loss, is one example. Automated SPPS machines now make peptides of 50 or more residues, and longer proteins are built by joining synthetic pieces with native chemical ligation. Mass spectrometry has largely replaced Edman degradation for sequencing. Search "native chemical ligation" and "peptide therapeutics". This is not for the exam.

2.4 Protein structure: primary, secondary, tertiary, quaternary

A protein works only when its chain is folded into the right shape, and that shape is described at four levels. Primary structure is the sequence of amino acids. Secondary structure is the local folding of the backbone into helices and sheets. Tertiary structure is the full three-dimensional fold of one chain, and quaternary structure is how several folded

chains fit together. Each level is held by particular bonds, and each is decided, in the end, by the sequence. This is the one topic every university in this unit teaches.

You need: side-chain groups (Chapter 2.1) and the planar peptide bond (Chapter 2.3).

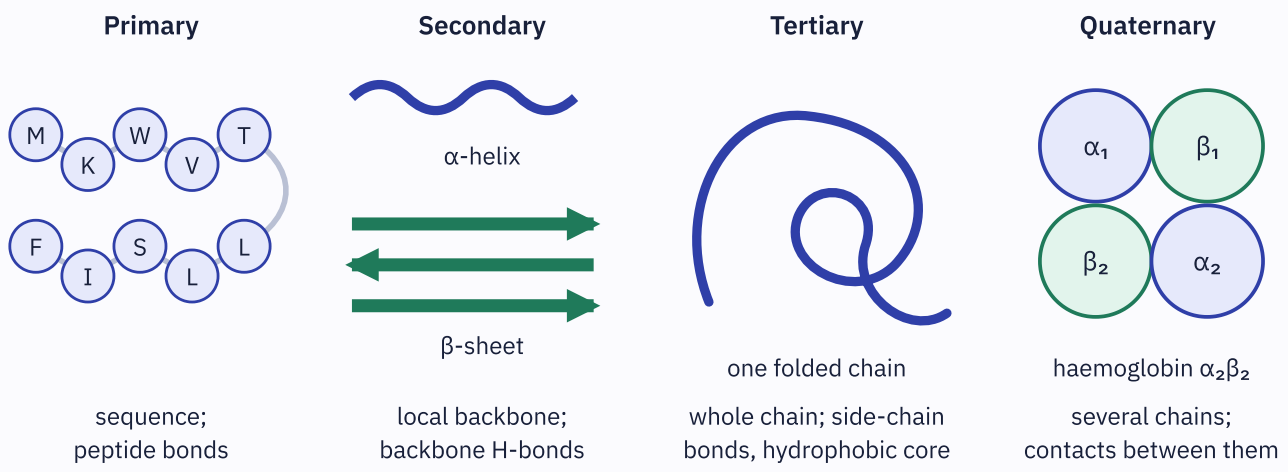


Figure 2.4.1 The four levels of protein structure and the bonds that hold each. Notice that secondary structure is held by hydrogen bonds between backbone atoms only, while tertiary and quaternary structure depend on side chains.

Primary structure and the bonds that hold proteins

The **primary structure** is the order of amino acids from N- to C-terminus, plus the positions of any disulfide bridges. It is coded by the gene, and it contains all the information needed for the higher levels (Chapter 2.6 shows the proof). One change can matter: in sickle-cell haemoglobin, glutamate at position 6 of the β chain is replaced by valine. The new hydrophobic patch makes deoxygenated haemoglobin molecules stick together into fibres that bend red cells into a sickle shape.

BOND	BETWEEN	MAIN LEVEL
Peptide bond (covalent)	α-COOH and α-NH_2 of adjacent residues	Primary
Disulfide bridge (covalent)	–SH groups of two cysteines	Tertiary, sometimes quaternary
Hydrogen bond	Backbone C=O and N–H; polar side chains	Secondary and tertiary
Ionic bond (salt bridge)	Opposite charges, such as Lys and Asp	Tertiary and quaternary

BOND	BETWEEN	MAIN LEVEL
Hydrophobic interaction	Nonpolar side chains clustering away from water	Tertiary and quaternary (the main driving force of folding)
Van der Waals forces	All atoms in close contact	All levels, weak but numerous

Secondary structure

Secondary structure is a regular, repeating fold of the backbone, held by hydrogen bonds between C=O and N–H groups of the backbone. Linus Pauling and Robert Corey predicted the two main kinds in 1951 by asking which shapes keep every peptide bond planar and every C=O paired with an N–H.

α-helix. The chain winds into a right-handed spiral. There are 3.6 residues per turn, each residue rises 1.5 Å along the axis, and one turn (the pitch) is 3.6 × 1.5 = 5.4 Å. The C=O of residue n bonds to the N–H of residue n + 4, so every backbone group in the helix is hydrogen-bonded. Side chains point outward. Proline breaks a helix because its nitrogen has no H to donate and its ring cannot fit the turn; glycine is too flexible; a run of like charges (several Glu or Lys) repels itself. Keratin of hair is mainly α-helix.

β-sheet. The chain is stretched almost fully, about 3.3 to 3.5 Å per residue, into a **β-strand**. Two or more strands lie side by side and hydrogen-bond to each other, forming a pleated sheet. Strands running the same direction (N→C) form a **parallel** sheet; alternating directions form an **antiparallel** sheet, which has straighter, stronger hydrogen bonds. Side chains point alternately above and below the sheet. Silk fibroin is mostly antiparallel β-sheet rich in small Gly and Ala, which lets the sheets stack closely.

β-turn. A tight turn of four residues, where the C=O of residue 1 bonds to the N–H of residue 4, reverses the chain's direction between strands. Gly and Pro are common in turns.

Collagen triple helix. Collagen, the main protein of skin, tendon and bone, has a repeating Gly-X-Y sequence, where X is often proline and Y often 4-hydroxyproline. Three left-handed chains twist round each other into a right-handed superhelix. Glycine is needed at every third position because only its tiny side chain fits in the crowded centre. Making hydroxyproline needs vitamin C, which is why its lack (scurvy) weakens gums and skin.

Each amino acid has a measured tendency, or propensity, to be found in each structure. In the **Chou–Fasman method**, Glu, Met, Ala and Leu are strong helix formers; Val, Ile, Tyr, Phe and Trp favour β-strands; Pro and Gly break both and are common in turns. A stretch of a few residues rich in helix formers is predicted to start a helix, which is then extended until breakers appear. Such methods are about 50 to 60 percent accurate; modern programs that use many related sequences and machine learning do much better.

Tertiary and quaternary structure

Tertiary structure is the complete three-dimensional arrangement of one polypeptide, including how its helices and sheets pack against each other. In water-soluble proteins, nonpolar side chains are buried in a core and polar ones cover the surface. Large proteins fold into separate compact units called **domains**, each often with its own job. **Fibrous proteins** (keratin, collagen, silk) have long, simple shapes and are insoluble; **globular proteins** (enzymes, myoglobin, antibodies) are compact and soluble (Chapter 2.7). Myoglobin, the oxygen store of muscle, was the first protein whose tertiary structure was solved by X-ray crystallography (John Kendrew); it is one chain of 153 residues, mostly α-helix, wrapped round a haem group.

Quaternary structure exists only in proteins with more than one chain (subunit). Haemoglobin has four: two α and two β chains ($\alpha_2\beta_2$), each with a haem. The subunits are held by hydrophobic contacts, hydrogen bonds and salt bridges. Because they touch, oxygen binding to one subunit changes the others and makes them bind oxygen more easily; this **cooperative binding** lets haemoglobin load oxygen in the lungs and release it in tissues. Myoglobin, with one chain, cannot do this.

WORKED EXAMPLE

Question (5 marks). A 36-residue segment of a protein forms an α-helix. (a) How many turns does it make? (b) How long is the helix? (c) If the same segment were a fully extended β-strand at 3.5 Å per residue, how long would it be? (d) Which form would fit across a cell membrane about 3 nm thick?

1. Turns = residues ÷ residues per turn = 36 ÷ 3.6 = **10 turns**.
2. Length = residues × rise per residue = 36 × 1.5 Å = 54 Å = **5.4 nm**. (Check: 10 turns × 5.4 Å pitch = 54 Å.)
3. β-strand length = 36 × 3.5 Å = 126 Å = **12.6 nm**, more than twice as long.
4. A membrane about 3 nm (30 Å) thick needs 30 ÷ 1.5 = 20 residues of α-helix. So a 20-residue helix of hydrophobic residues spans it; that is why membrane-crossing segments are usually helices of about 20 residues.

EXAM TRAPS

- Secondary structure uses backbone hydrogen bonds only. Do not say that side-chain bonds make the α-helix.
- Rise per residue (1.5 Å) and pitch (5.4 Å) are different numbers; pitch = 3.6 × 1.5.
- Quaternary structure needs more than one folded chain. Myoglobin (one chain) has none; haemoglobin (four chains) has it. Do not confuse it with tertiary structure, which is the fold of one chain.
- Disulfide bridges are covalent, so they survive urea; they are broken only by reducing agents (Chapter 2.6).
- The examiner's favourite: "Describe the α-helix with a neat diagram." Give residues per turn, rise, pitch, the n to n + 4 hydrogen bond, right-handedness and the role of proline.

CHECK YOURSELF

– REMEMBER **State the residues per turn, rise per residue and pitch of the α-helix.** 3 marks

3.6 residues per turn; 1.5 Å rise per residue; pitch 5.4 Å.

– REMEMBER **What is the subunit composition of adult haemoglobin?** 1 mark

Two α and two β chains, $\alpha_2\beta_2$, each carrying one haem.

– UNDERSTAND **Why is proline called a helix breaker?** 2 marks

Its nitrogen has no hydrogen to give for the n to n + 4 hydrogen bond, and its ring fixes the φ angle and clashes with the turn of the helix.

– UNDERSTAND **Compare parallel and antiparallel β-sheets.** 4 marks

In parallel sheets all strands run N→C in the same direction and hydrogen bonds are slanted; in antiparallel sheets neighbouring strands run in opposite directions and hydrogen bonds are straight, making the sheet more stable. Both have side chains alternating above and below.

– APPLY **How many residues are needed for an α-helix 4.5 nm long, and how many turns is that?** 2 marks

45 Å ÷ 1.5 Å = 30 residues; 30 ÷ 3.6 ≈ 8.3 turns.

– APPLY **Using Chou–Fasman propensities, which segment is more likely to be a helix: E-A-L-M-K-A or V-I-Y-T-V-G?** 2 marks

E-A-L-M-K-A, which is rich in the helix formers Glu, Ala, Leu and Met. V-I-Y-T-V-G is rich in β-formers and ends with a breaker.

– ANALYZE **Explain why replacing Glu6 by Val in the β chain of haemoglobin causes disease, though it changes only one residue of 146.** 5 marks

Position 6 is on the surface. Glu there is charged and faces water; Val is hydrophobic and creates a sticky patch. In deoxygenated haemoglobin this patch fits a hydrophobic pocket on a neighbouring molecule, so molecules polymerise into long fibres that distort red cells into sickles. A change in primary structure alters the surface and how molecules associate.

BEYOND THE SYLLABUS

Predicting the full three-dimensional structure from sequence was a 50-year problem until deep-learning systems such as AlphaFold reached near-experimental accuracy for many proteins around 2020. The AlphaFold Protein Structure Database now offers predicted structures for a very large number of proteins, free to all. Experimental structures are still solved by X-ray crystallography, NMR and, more and more, cryo-electron microscopy, and are stored in the Protein Data Bank. Search "AlphaFold database" and "Protein Data Bank". This is not for the exam.

2.5 Ramachandran plot

Because the peptide bond is flat, each residue of a protein chain can turn only about two bonds, and the angles of those two bonds decide the whole backbone shape. The Ramachandran plot is a map of these two angles that shows which pairs are possible and which would make atoms crash into each other. It was worked out at the University of Madras by G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, published in 1963, and it is still the first check made on any new protein structure.

You need: the planar peptide unit (Chapter 2.3) and the α-helix and β-sheet (Chapter 2.4).

The three backbone angles

A torsion (dihedral) angle measures the twist about a bond, from −180° to +180°. Each residue has three:

- **φ (phi)**: rotation about the N–Cα bond.
- **ψ (psi)**: rotation about the Cα–C bond.

- **ω (omega)**: rotation about the peptide C–N bond. It is fixed near 180° (trans) because the bond is partly double, so it is left out of the plot.

So the conformation of a backbone is fully described by one (φ, ψ) pair per residue.

How the plot was made

Ramachandran's group treated atoms as hard spheres and, for every (φ, ψ) pair, checked whether any two atoms of neighbouring peptide units came closer than their normal contact distance. Pairs with no clash are **fully allowed**; pairs that are allowed only if atoms come slightly closer are **partially allowed** (the outer limit); the rest are **disallowed**. Most of the map is disallowed: the backbone oxygen, the amide hydrogen and the side-chain Cβ get in each other's way.

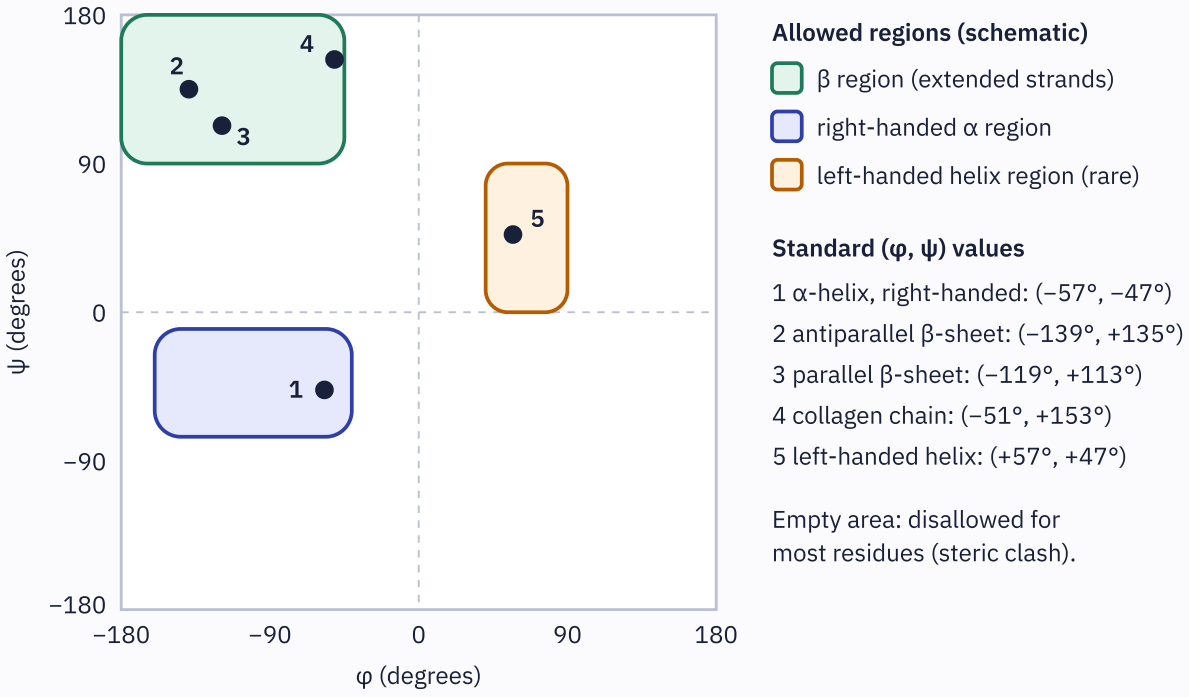


Figure 2.5.1 A schematic Ramachandran plot. The shaded boxes show roughly where the allowed regions lie for a residue with a side chain (real boundaries are curved); the dots are the standard angles of each secondary structure. Notice that both the α-helix and the β-sheet lie at negative φ, on the left half of the map.

Reading the plot

The top-left region holds extended chains: β-strands and the collagen helix. The lower-left region holds the right-handed α-helix. A small region at positive φ holds the left-handed helix, which is rare because the side chain clashes with the backbone there.

Glycine, with only H as its side chain, has a much larger allowed area, including large parts of the right half; this is why glycine is common in tight turns and where a chain must bend sharply. **Proline** is the opposite: its ring locks φ near −60°, so it fits only a narrow vertical band. These two are usually shown on separate plots.

When a new crystal or cryo-EM structure is deposited, programs plot every residue. In a good structure nearly all residues fall in the favoured regions; a residue in a disallowed area points to an error in the model, or to a strained residue that may matter for function.

WORKED EXAMPLE

Question (4 marks). Four residues of a non-glycine, non-proline protein have (φ, ψ) of A (−60°, −45°), B (−130°, +140°), C (+60°, +45°), D (0°, 0°). Assign each to a region and comment.

1. **A**: φ negative, ψ negative and close to (−57°, −47°): right-handed α-helix.
2. **B**: φ negative, ψ large and positive, near (−139°, +135°): β region, probably an antiparallel β-strand.
3. **C**: φ positive, near (+57°, +47°): left-handed helix region. Allowed but rare for a residue with a side chain; check whether it is really glycine, or whether the model is wrong.
4. **D**: (0°, 0°) puts backbone atoms of neighbouring peptide units on top of each other: disallowed. The model must be corrected.

EXAM TRAPS

- ω is not plotted because it is fixed near 180°; do not label the axes ω.
- φ is about N–Cα and ψ about Cα–C; students often swap them.
- The right-handed α-helix has negative φ and ψ. A positive-φ helix is left-handed.
- The examiner's favourite: "Why does glycine have a larger allowed region?" Because its side chain is only H, so there is less to clash.

CHECK YOURSELF

– REMEMBER **Who proposed the Ramachandran plot, and where?** 1 mark

G. N. Ramachandran with C. Ramakrishnan and V. Sasisekharan, University of Madras, 1963.

– REMEMBER **Give the approximate (φ, ψ) of a right-handed α-helix.** 1 mark

About (−57°, −47°).

– UNDERSTAND **Why are large parts of the plot disallowed?** 2 marks

At those angles, atoms of neighbouring peptide units or the Cβ come closer than their van der Waals contact distance, which is not physically possible (steric clash).

– APPLY **A residue has φ = −65°, ψ = +145°. Which region is it in?** 1 mark

The upper-left (β) region, close to the collagen and polyproline value; an extended conformation.

– ANALYZE **Explain why glycine and proline have very different Ramachandran plots, and how this links to their roles in proteins.** 4 marks

Glycine has no Cβ, so few clashes and a wide allowed area: it allows sharp turns and flexible loops. Proline's ring fixes φ near −60°, giving a narrow allowed band: it is rigid, breaks helices and is common in turns and in collagen.

BEYOND THE SYLLABUS

Structure-checking tools such as MolProbity use Ramachandran plots built from many high-quality structures, rather than the hard-sphere model, to flag outliers. Protein design and structure-prediction programs use the same angle preferences. Search "Ramachandran outliers" and "MolProbity". This is not for the exam.

2.6 Protein folding and denaturation

A new protein leaves the ribosome as a floppy chain, and within milliseconds to seconds it folds into one exact shape, its native structure. Folding is the process that gets it there; denaturation is the loss of that shape when heat, acid, urea or detergent breaks the weak bonds that hold it. Only the folded protein works, so these two ideas explain why fever is dangerous, why boiled eggs go solid, how paneer is made, and why some diseases come from proteins that fold wrongly.

You need: the bonds that hold proteins and the four levels of structure (Chapter 2.4).

Sequence decides shape: Anfinsen's experiment

Christian Anfinsen worked with ribonuclease A, a small enzyme of one chain with four disulfide bonds. He used it to show that a protein's sequence is enough to fold it.

1. He treated the active enzyme with 8 M urea (which breaks non-covalent bonds) and 2-mercaptoethanol (which reduces the disulfides to –SH). The chain unfolded and lost all activity.
2. When he removed both reagents slowly by dialysis and let air re-oxidise the –SH groups, the enzyme regained nearly full activity and the correct disulfides re-formed.
3. When instead he let the –SH groups oxidise while urea was still present, the eight cysteines paired at random. Of the 105 possible pairings only one is correct, and this "scrambled" ribonuclease had about 1 percent of normal activity.
4. Adding a trace of 2-mercaptoethanol to the scrambled enzyme without urea let the disulfides swap partners, and activity returned as the chain found its native fold.

Conclusion, the **thermodynamic hypothesis**: the native structure is the most stable shape available to the sequence under normal conditions, so the information for folding is in the primary structure. Anfinsen shared the 1972 Nobel Prize in Chemistry for this work.

How folding happens

Driving forces. The main force is the **hydrophobic effect**: nonpolar side chains are buried, which frees the ordered water molecules that surrounded them and raises the entropy of the water. Hydrogen bonds, salt bridges and van der Waals contacts then fix the details. The folded state wins only narrowly; its free energy is lower than the unfolded state's by just a few tens of kJ/mol, small compared with the energy of one covalent bond. That is why a modest change in temperature or pH can unfold a protein.

Levinthal's paradox. If a chain tried every possible shape at random, folding would take longer than the age of the universe (see the worked example). Real proteins fold fast, so folding cannot be a random search. The current picture is a **folding funnel**: the energy landscape slopes toward the native state, so each small step that forms local structure (a helix, a hydrophobic cluster) lowers the energy and narrows the choices. A partly folded intermediate with most secondary structure but loose packing, the **molten globule**, is often seen on the way.

Helpers in the cell. Inside a crowded cell, hydrophobic patches of half-folded chains can stick to each other. **Molecular chaperones** prevent this. Hsp70 proteins bind exposed hydrophobic segments; **chaperonins** such as GroEL with its lid GroES in *E. coli* give a single chain a closed chamber to fold in, using ATP. Two enzymes speed slow steps: **protein disulfide isomerase**

shuffles wrong disulfides, and **peptidyl-prolyl cis–trans isomerase** flips bonds before proline. Chaperones do not carry the shape information; they only stop mistakes.

Misfolding diseases. Some proteins misfold into long, insoluble fibres rich in β-sheet called **amyloid**. Amyloid-β plaques are found in Alzheimer's disease and α-synuclein deposits in Parkinson's disease. In prion diseases such as Creutzfeldt–Jakob disease, a misfolded prion protein converts the normal form into more of itself. In the commonest form of cystic fibrosis, the CFTR protein lacks one phenylalanine, misfolds and is destroyed before it reaches the cell membrane.

Denaturation

Denaturation is the loss of the native secondary, tertiary and quaternary structure, and with it biological activity, without breaking peptide bonds. The primary structure stays intact; breaking it is hydrolysis, not denaturation.

AGENT	WHAT IT BREAKS
Heat	Hydrogen bonds and other weak interactions, by increasing molecular motion
Extreme pH (strong acid or alkali)	Changes side-chain charges; breaks salt bridges and adds charge repulsion
Urea (about 8 M), guanidinium chloride (about 6 M)	Hydrogen bonds; also make it easier for nonpolar groups to be in water
Detergents such as SDS	Bind hydrophobic parts and coat the chain with negative charge
Organic solvents (ethanol, acetone)	Hydrophobic interactions
Heavy metal ions (Hg^{2+}, Pb^{2+}, Ag^{+})	Bind –SH and carboxyl groups
Reducing agents (2-mercaptoethanol, DTT)	Disulfide bridges
Mechanical stress (whipping, shaking)	Unfold proteins at air–water surfaces, as in whipped egg white

Effects: loss of biological activity; lower solubility, often with precipitation or coagulation; higher viscosity; and easier digestion by proteases, which is one reason cooked food is easier to digest. Denaturation can be **reversible** (ribonuclease after urea is removed) or **irreversible**, when unfolded chains tangle and aggregate, as when an egg is boiled. In the kitchen, lemon juice added to hot milk lowers the pH toward the isoelectric point of casein, about 4.6; casein loses its charge and coagulates into paneer.

Heat denaturation of a small protein is usually **cooperative**: it goes from folded to unfolded over a narrow range of temperature, because the bonds support each other. The **melting temperature (Tm)** is the temperature at which half the molecules are unfolded.

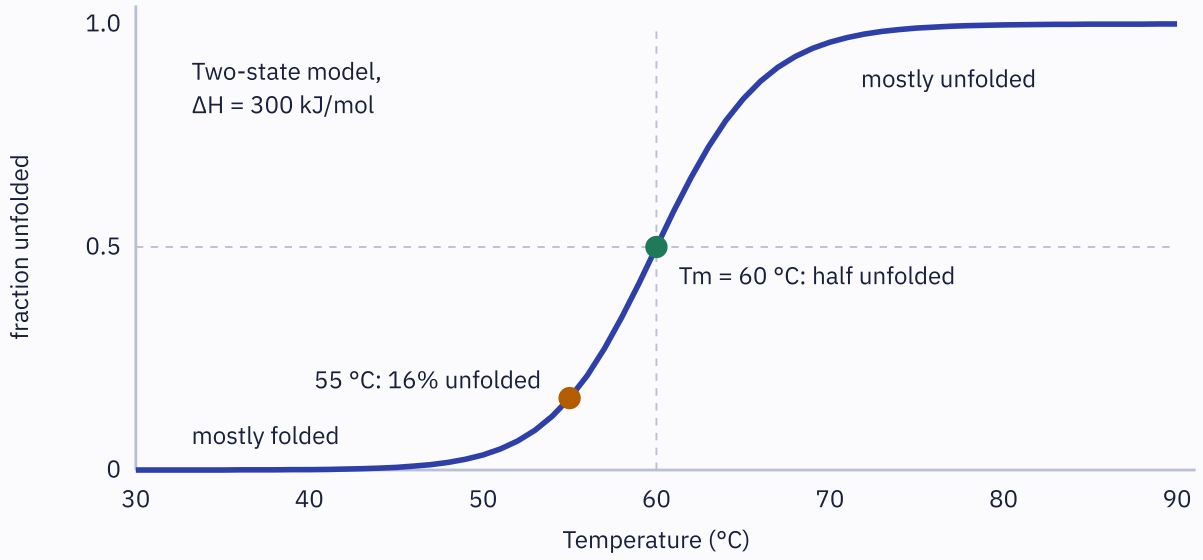


Figure 2.6.1 Computed heat-denaturation curve for a model protein that is either fully folded or fully unfolded (two-state), with Tm = 60 °C and unfolding enthalpy ΔH = 300 kJ/mol. Notice how sharp the change is: most of it happens within about 10 °C of Tm.

WORKED EXAMPLE

Question (6 marks). (a) A protein of 100 residues; assume each residue can take 3 backbone conformations and the chain tries a new shape every 10^{-13} s. How long would a random search of all shapes take? (b) For the protein in Figure 2.6.1, what fraction is unfolded at 55 °C? Use the two-state relation $K = \exp[-(\Delta H/R)(1/T - 1/Tm)]$ and fraction unfolded $= K/(1 + K)$, with $R = 8.314$ J mol^{-1} K^{-1}.

1. Number of shapes $= 3^{100} \approx 5.15 \times 10^{47}$.
2. Time $= 5.15 \times 10^{47} \times 10^{-13}$ s $= 5.15 \times 10^{34}$ s. One year $\approx 3.16 \times 10^{7}$ s, so time $\approx$ **1.6×10^{27} years**, far longer than the age of the universe (about 1.4×10^{10} years). Real folding takes milliseconds to seconds, so it cannot be a random search.
3. Convert temperatures to kelvin: $T = 55 + 273.15 = 328.15$ K; $Tm = 60 + 273.15 = 333.15$ K.
4. $1/T - 1/Tm = 0.0030474 - 0.0030017 = 4.57 \times 10^{-5}$ K^{-1}. $\Delta H/R = 300{,}000 / 8.314 = 36{,}084$ K.
5. Exponent $= -36{,}084 \times 4.57 \times 10^{-5} = -1.65$, so $K = e^{-1.65} = 0.192$.
6. Fraction unfolded $= 0.192 / 1.192 =$ **0.16**, so 16 percent is unfolded and 84 percent still folded, 5 °C below Tm.

EXAM TRAPS

- Denaturation does not break peptide bonds. If you write that the primary structure is destroyed, you lose the mark.
- Urea alone does not break disulfide bonds; a reducing agent is needed. Anfinsen used both.
- Chaperones help folding but do not decide the final shape; the sequence does.
- Use kelvin, not °C, in any thermodynamic formula.
- The examiner's favourite: "Describe Anfinsen's experiment and its conclusion." Give the reagents, the two refolding conditions, the 1 percent activity of scrambled enzyme and the thermodynamic hypothesis.

CHECK YOURSELF

– REMEMBER **Define denaturation.** 2 marks

Loss of a protein's native secondary, tertiary and quaternary structure, and so its activity, without breaking peptide bonds.

– REMEMBER **Name two molecular chaperones.** 1 mark

Hsp70 and the chaperonin GroEL–GroES (also acceptable: Hsp90, protein disulfide isomerase as a folding enzyme).

– UNDERSTAND **Why does a scrambled ribonuclease regain activity when a trace of 2-mercaptoethanol is added after urea is removed?** 3 marks

The trace of thiol lets wrong disulfides break and re-form with new partners. Without urea, the chain can fold, and the native pairing, which belongs to the most stable fold, accumulates.

– UNDERSTAND **What is the hydrophobic effect, and why does it drive folding?** 3 marks

Nonpolar groups in water force nearby water molecules into ordered cages. Burying these groups inside the protein releases that water, raising its entropy and lowering the free energy of the folded state.

– APPLY **Why does adding lemon juice to hot milk give paneer?** 2 marks

The acid lowers the pH to near casein's isoelectric point (about 4.6), where it has no net charge and so stops repelling itself; heat helps the chains unfold and coagulate into curds.

– APPLY **Using the model of Figure 2.6.1, what fraction is unfolded at 65 °C?** 3 marks

$1/338.15 - 1/333.15 = -4.44 \times 10^{-5}\ K^{-1}$; exponent = +1.60; $K \approx 4.96$; fraction unfolded = $4.96/5.96 \approx 0.83$.

– ANALYZE **Compare reversible and irreversible denaturation with an example of each, and explain what decides which occurs.** 4 marks

Reversible: ribonuclease in urea refolds when urea is removed. Irreversible: egg albumin on boiling. If unfolded chains stay separate, they can refold; if exposed hydrophobic groups make chains stick together (aggregation), or covalent changes occur, they cannot return to the native state.

BEYOND THE SYLLABUS

Biopharma companies, including Indian biosimilar makers, spend great effort keeping therapeutic proteins such as insulin and monoclonal antibodies folded during production, storage and transport, since aggregates lower activity and can cause immune reactions. Recombinant proteins made in *E. coli* often form insoluble inclusion bodies that must be solubilised in urea and refolded, the Anfinsen experiment at industrial scale. Search "inclusion body refolding" and "differential scanning fluorimetry". This is not for the exam.

2.7 Classification and properties of proteins

There are many thousands of different proteins, so they are sorted in four ways: by shape, by chemical composition, by solubility and by biological function. Their properties, such as solubility, the isoelectric point, salting out and the colour reactions, all follow from the amino acids they contain and from their large size. These properties are used every day in labs to detect proteins, to tell one kind from another and to separate them.

You need: the isoelectric point (Chapter 2.2) and denaturation (Chapter 2.6).

Classification by shape, function and nutrition

Shape. Fibrous proteins are long and thread-like, insoluble in water and structural: keratin (hair, nails), collagen (tendon, skin), elastin, silk fibroin. **Globular proteins** are compact, roughly spherical, mostly water-soluble and do the cell's active work: enzymes, haemoglobin, myoglobin, albumin, antibodies, insulin.

FUNCTION	EXAMPLES
Catalytic (enzymes)	Amylase, trypsin, DNA polymerase
Transport	Haemoglobin (O_2), serum albumin (fatty acids, drugs), transferrin (iron)
Storage	Ferritin (iron), casein (milk), ovalbumin (egg)
Structural	Collagen, keratin, elastin
Contractile and motile	Actin, myosin, tubulin
Defence	Antibodies (immunoglobulins), fibrinogen and thrombin (clotting)
Regulatory (hormones)	Insulin, glucagon, growth hormone
Receptors and signalling	Insulin receptor, rhodopsin

Nutritional value. A **complete** protein supplies all essential amino acids in good amounts (egg, milk). An **incomplete** protein lacks one or more: gelatin lacks tryptophan; zein of maize is poor in lysine and tryptophan.

Classification by composition and solubility

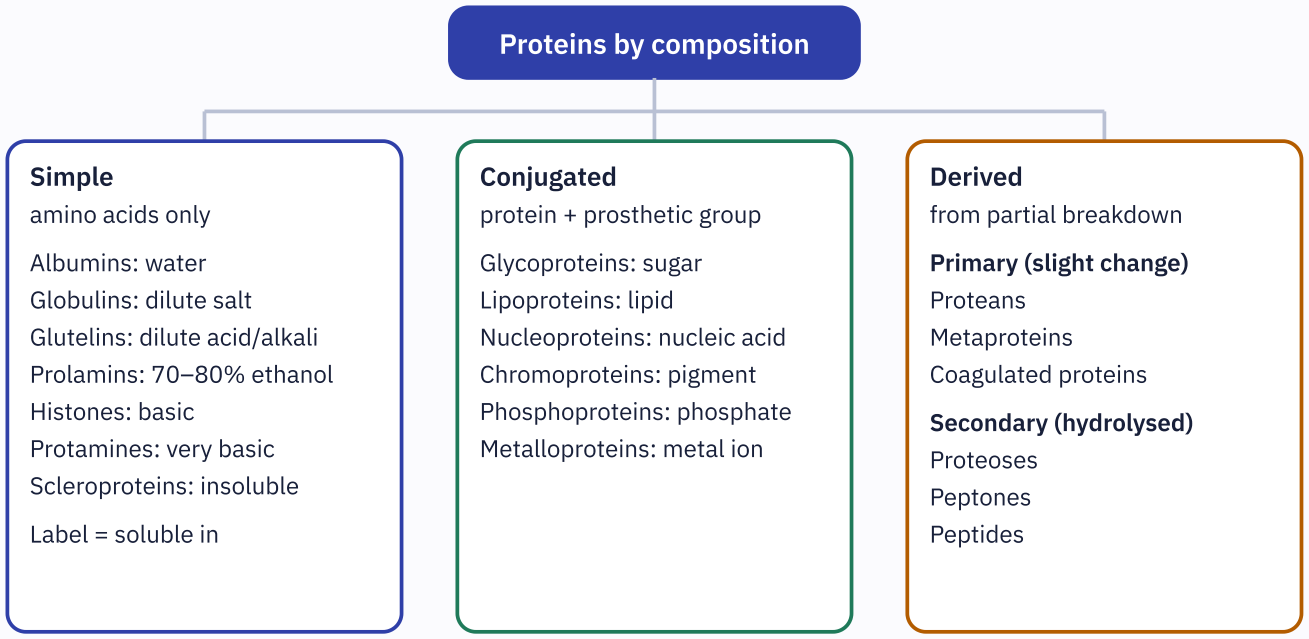


Figure 2.7.1 Classification of proteins by composition, the scheme used in most Indian biochemistry textbooks. Simple proteins are further sorted by the solvent that dissolves them; conjugated proteins are named after their non-protein part (prosthetic group); derived proteins are products of denaturation or partial hydrolysis.

Simple proteins give only amino acids on hydrolysis. Examples: **albumins** (serum albumin, ovalbumin; also coagulated by heat), **globulins** (serum globulins), **glutelins** (glutenin of wheat, oryzenin of rice), **prolamins** (gliadin of wheat, zein of maize), **histones** (bind DNA in chromatin), **protamines** (in sperm) and **scleroproteins** or albuminoids (keratin, collagen, elastin). Gliadin and glutenin together make gluten, which gives chapati dough its stretch.

Conjugated proteins carry a non-protein **prosthetic group**. **Glycoproteins** carry sugars: mucins of saliva, immunoglobulins, many cell-surface proteins (sugar chemistry is in Unit 1). **Lipoproteins** carry lipids in blood: chylomicrons, VLDL, LDL and HDL (lipids are in Unit 3). **Nucleoproteins**: chromatin, ribosomes. **Chromoproteins** have a coloured group: haemoglobin (haem), flavoproteins (FAD), rhodopsin (retinal). **Phosphoproteins**: casein of milk, vitellin of egg yolk. **Metalloproteins**: ferritin (iron), ceruloplasmin (copper), carbonic anhydrase (zinc).

Derived proteins come from denaturation or partial hydrolysis of the above. Primary derived proteins are only slightly changed (metaproteins, coagulated proteins); secondary derived proteins are hydrolysed into smaller pieces (proteoses, peptones, peptides).

Physical properties

- **Size and colloidal nature.** Molar masses range from a few thousand to millions of g/mol. Proteins form colloids, do not pass through semipermeable membranes (the basis of dialysis) and scatter light.
- **Amphoteric; isoelectric point.** Their many ionisable side chains make proteins buffers. At the pI, net charge is zero, molecules stop repelling each other and solubility is lowest; they do not move in an electric field.
- **Salting in and salting out.** A little salt shields charges and increases solubility (salting in). A high concentration of a salt such as ammonium sulfate takes away the water the protein needs and precipitates it (salting out). Different proteins come out at different salt levels; this is the first step of many purifications (Chapter 2.8).
- **Optical and spectral.** Proteins are optically active and absorb UV light at 280 nm because of Trp and Tyr.
- **Denaturation and coagulation** by heat and other agents (Chapter 2.6).

Chemical properties: colour reactions and precipitation

TEST	REAGENT	POSITIVE RESULT	DETECTS
Biuret	Alkaline $CuSO_4$	Violet	Two or more peptide bonds (tripeptides and larger)
Ninhydrin	Ninhydrin, heat	Purple	Free α-amino groups
Xanthoproteic	Concentrated HNO_3, heat; then alkali	Yellow, turning orange	Aromatic rings (Tyr, Trp)
Millon's	Mercuric nitrate in HNO_3, heat	Red	Phenol group (Tyr)
Hopkins–Cole	Glyoxylic acid, then concentrated H_2SO_4	Violet ring	Indole group (Trp)
Sakaguchi	α-Naphthol, alkaline hypobromite	Red	Guanidino group (Arg)
Lead acetate (sulfur)	Boil with NaOH, add lead acetate	Black precipitate (PbS)	–SH and S–S sulfur (Cys, cystine); not Met

Precipitation. Above their pI, proteins are negative and are precipitated by heavy metal cations (Hg^{2+}, Pb^{2+}, Ag^{+}); this is why raw egg white or milk is given as a first-aid antidote in heavy-metal poisoning. Below their pI, they are positive and are precipitated by acid reagents such as trichloroacetic acid (TCA), picric acid and tungstic acid, used to remove proteins from blood

samples before analysis. Ethanol and acetone also precipitate proteins. **Hydrolysis** by acid, alkali or proteases breaks peptide bonds and gives free amino acids.

WORKED EXAMPLE

Question (6 marks). Three white protein samples are tested. **A** dissolves in water, coagulates on boiling, and gives positive biuret, Millon's and lead acetate tests. **B** is insoluble in water but dissolves in 70 percent ethanol; biuret positive. **C** dissolves in hot water and sets to a jelly on cooling; biuret positive, Hopkins–Cole negative. Identify the type of each and give an example.

1. All three are biuret positive, so all contain many peptide bonds: they are proteins, not free amino acids.
2. **A**: water-soluble and heat-coagulable is the definition of an albumin. Millon's positive shows Tyr; lead acetate positive shows Cys or cystine. Example: egg albumin (ovalbumin) or serum albumin.
3. **B**: soluble in 70–80 percent ethanol is the definition of a prolamin. Example: zein of maize or gliadin of wheat.
4. **C**: dissolves hot and gels on cooling, so it is gelatin, a derived protein made by boiling collagen. Hopkins–Cole negative fits: gelatin lacks tryptophan, which is also why it is a nutritionally incomplete protein.

EXAM TRAPS

- Biuret needs at least two peptide bonds; a dipeptide and free amino acids give a negative test.
- The lead acetate test detects Cys and cystine but not methionine, whose sulfur is held too tightly.
- Albumins dissolve in water; globulins need dilute salt. Swapping these is a common slip.
- Salting out is not denaturation: a salted-out enzyme redissolves with its activity intact.
- Heavy metals precipitate proteins above the pI (protein negative); acid reagents below the pI (protein positive).

CHECK YOURSELF

– REMEMBER **Classify conjugated proteins with one example each.** 4 marks

Glycoproteins (immunoglobulins), lipoproteins (LDL), nucleoproteins (chromatin), chromoproteins (haemoglobin), phosphoproteins (casein), metalloproteins (ferritin).

– REMEMBER **Which colour test is specific for tryptophan, and what is seen?** 2 marks

Hopkins–Cole test: a violet ring at the junction of the two layers.

– UNDERSTAND **Explain salting in and salting out.** 3 marks

At low salt, ions shield charged groups and reduce attraction between protein molecules, raising solubility (salting in). At high salt, the salt ions take up the water that hydrates the protein, so protein molecules associate and precipitate (salting out).

– UNDERSTAND **Why are proteins least soluble at their isoelectric point?** 2 marks

With zero net charge, molecules no longer repel each other electrically, so they come together and precipitate.

– APPLY **Why is raw egg white given in mercury poisoning?** 2 marks

Its proteins bind Hg^{2+} and precipitate it in the stomach, so less metal reaches the body's own proteins; the precipitate is then removed by vomiting or stomach wash.

– ANALYZE **A solution gives a positive ninhydrin test but a negative biuret test. What can you conclude?** 3 marks

Free amino groups are present, but there are fewer than two peptide bonds. So the solution contains free amino acids or dipeptides, not proteins.

BEYOND THE SYLLABUS

Classical colour tests are now mainly teaching tools. Labs identify proteins by mass spectrometry and classify them by sequence families and structural folds in databases such as UniProt, Pfam and CATH, and by function using Gene Ontology terms. Search "UniProt" and "protein families". This is not for the exam.

2.8 Protein purification and quantification

To study an enzyme or make a protein drug, you must separate one protein from the thousands of others in a cell and know how much you have. Purification uses differences in solubility, size, charge and binding to separate proteins step by step; quantification measures protein concentration by UV absorbance or a colour reaction. At other universities they usually appear in a bioanalytical or downstream processing course.

You need: isoelectric point and salting out (Chapter 2.7) and denaturation (Chapter 2.6).

Purification step by step

1. **Release.** Break cells (sonication, a homogeniser or lysozyme for bacteria) and spin off debris. The clear liquid is the **crude extract**. Work cold, with buffer, so the protein does not denature.
2. **Fractionation by solubility.** Add ammonium sulfate in steps; collect the precipitate that contains the target (salting out). Remove the salt by **dialysis** in a membrane bag that lets small ions through but keeps the protein.
3. **Ion-exchange chromatography** separates by charge. Above its pI a protein is negative and binds an anion exchanger (such as DEAE-cellulose); below its pI it binds a cation exchanger (such as CM-cellulose). Bound proteins are released by increasing salt or changing pH.
4. **Size-exclusion (gel filtration) chromatography** separates by size. Beads have pores; small proteins enter them and take a longer path, so **large proteins come out first**.
5. **Affinity chromatography** uses a ligand that binds only the target, such as a substrate analogue or, for recombinant proteins, nickel ions that bind a His-tag. Wash, then release with free ligand. It often gives the largest single step of purification.
6. **Check purity** by SDS-PAGE: SDS denatures and coats proteins with negative charge, so they move through a gel by size alone; a pure protein gives one band. Its position against standards gives the subunit mass.

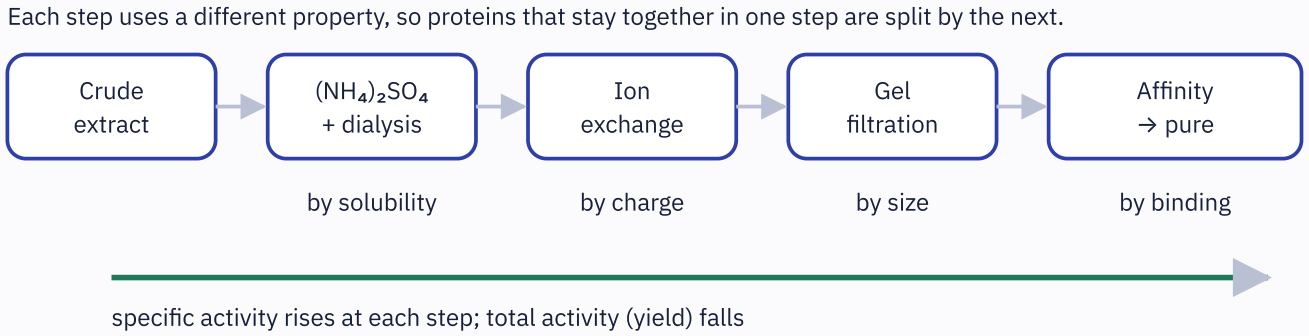


Figure 2.8.1 A typical purification. Notice that each method separates by a different property, and that the order usually goes from cheap, high-capacity steps (salting out) to selective, low-capacity ones (affinity).

Measuring progress and quantity

A purification is tracked with a **purification table**. **Specific activity** = total activity ÷ total protein (units per mg); it rises as other proteins are removed. **Yield** = activity at a step ÷ activity in the crude extract × 100. **Fold purification** = specific activity at a step ÷ specific activity of the crude extract.

Protein quantification methods:

- **UV absorbance at 280 nm**: fast and non-destructive, using the Beer–Lambert law $A = \varepsilon c l$ (A absorbance, ε molar absorption coefficient in $M^{-1}\,cm^{-1}$, c concentration in M, l path length in cm). Depends on Trp and Tyr content; nucleic acids interfere.
- **Biuret**: violet Cu^{2+}–peptide complex; simple but needs milligram amounts.
- **Lowry**: biuret reaction followed by the Folin–Ciocalteu reagent; about a hundred times more sensitive than biuret.
- **Bradford**: Coomassie Brilliant Blue G-250 binds protein and its absorption peak shifts to 595 nm; quick and sensitive.
- **BCA**: protein reduces Cu^{2+} to Cu^{+}, which forms a purple complex with bicinchoninic acid, read at 562 nm.

Colour methods need a **standard curve**, usually made with bovine serum albumin.

WORKED EXAMPLE

Question (6 marks). Complete the table and comment. Then find the concentration of the pure protein if A_{280} = 0.86 in a 1 cm cell, ε_{280} = 43,000 M^{-1} cm^{-1} and molar mass 66,000 g/mol.

STEP	PROTEIN (MG)	ACTIVITY (U)	SPECIFIC ACTIVITY (U/MG)	YIELD (%)	FOLD
Crude extract	10,000	200,000	20	100	1
$(NH_4)_2SO_4$	2,500	160,000	64	80	3.2
Ion exchange	400	120,000	300	60	15
Gel filtration	120	90,000	750	45	37.5
Affinity	20	70,000	3,500	35	175

1. Specific activity, for example ion exchange: 120,000 ÷ 400 = 300 U/mg. Yield: 120,000 ÷ 200,000 × 100 = 60 %. Fold: 300 ÷ 20 = 15.
2. Comment: the affinity step gives the largest jump (3,500 ÷ 750 ≈ 4.7-fold) with little loss. Overall the protein is purified 175-fold with 35 % yield.
3. Concentration: c = A ÷ (ε l) = 0.86 ÷ (43,000 × 1) = 2.0 × 10^{-5} M = 20 µM.
4. In mass units: 2.0 × 10^{-5} mol/L × 66,000 g/mol = 1.32 g/L = **1.32 mg/mL**.

EXAM TRAPS

- In gel filtration, large proteins come out first; in SDS-PAGE, small proteins move furthest. Do not mix these up.
- Specific activity, not total activity, shows purity.
- SDS-PAGE gives subunit mass, not the mass of the native multi-subunit protein.

CHECK YOURSELF

– REMEMBER **Define specific activity.** 1 mark

Enzyme activity per mg of total protein (units/mg).

– UNDERSTAND **Why do large proteins elute first in gel filtration?** 2 marks

They cannot enter the pores of the beads, so they travel only through the space between beads, a shorter path.

– APPLY **A protein has pI 5.0. Which ion exchanger binds it at pH 7.5?** 2 marks

At pH 7.5 (above pI) it is negative, so an anion exchanger such as DEAE-cellulose.

– APPLY **A step recovers 30,000 U in 5 mg from a crude extract of 150,000 U in 3,000 mg. Find yield and fold.** 3 marks

Yield = 30,000/150,000 × 100 = 20 %. Specific activities: 6,000 and 50 U/mg, so 120-fold.

– ANALYZE **Bradford gives 1.0 mg/mL but A_{280} suggests 3.0 mg/mL for the same sample. Suggest reasons.** 3 marks

The sample may contain nucleic acid, which also absorbs near 280 nm; or the ε used is wrong for this protein (unusually Trp-rich); or the Bradford standard (BSA) binds the dye differently from this protein.

BEYOND THE SYLLABUS

In industry these steps form downstream processing: monoclonal antibodies are captured on Protein A affinity columns and polished by ion exchange, often costing more than growing the cells. Search "downstream processing monoclonal antibodies". This is not for the exam.

Unit summary

TERM	MEANING OR FORMULA	CHAPTER
α-Amino acid	Cα carrying $–COO^-$, $–NH_3^+$, H and side chain R; L form in proteins	2.1
Essential amino acids	His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Val	2.1
Isoelectric point	$pI = (pKa_1 + pKa_2)/2$, averaging the two pKa values around the neutral form	2.2
Henderson–Hasselbalch	$pH = pKa + \log([A^-]/[HA])$	2.2
Peptide mass	Σ(amino acid masses) − (n − 1) × 18.02 g/mol for n residues	2.3
Peptide bond	Planar, rigid, trans; C–N about 1.32 Å; partial double bond	2.3
α-Helix	3.6 residues/turn, 1.5 Å rise, 5.4 Å pitch; H-bond n → n + 4; right-handed	2.4
β-Sheet	Extended strands, about 3.3–3.5 Å per residue; parallel or antiparallel	2.4
Ramachandran angles	φ about N–Cα, ψ about Cα–C; α-helix about (−57°, −47°), β about (−139°, +135°)	2.5
Two-state unfolding	$K = \exp[-(\Delta H/R)(1/T - 1/Tm)]$; fraction unfolded = K/(1 + K)	2.6
Biuret test	Violet with alkaline Cu^{2+}; needs two or more peptide bonds	2.7
Specific activity, fold, yield	Units/mg; SA(step)/SA(crude); activity(step)/activity(crude) × 100	2.8
Beer–Lambert law	$A = \varepsilon c l$	2.8

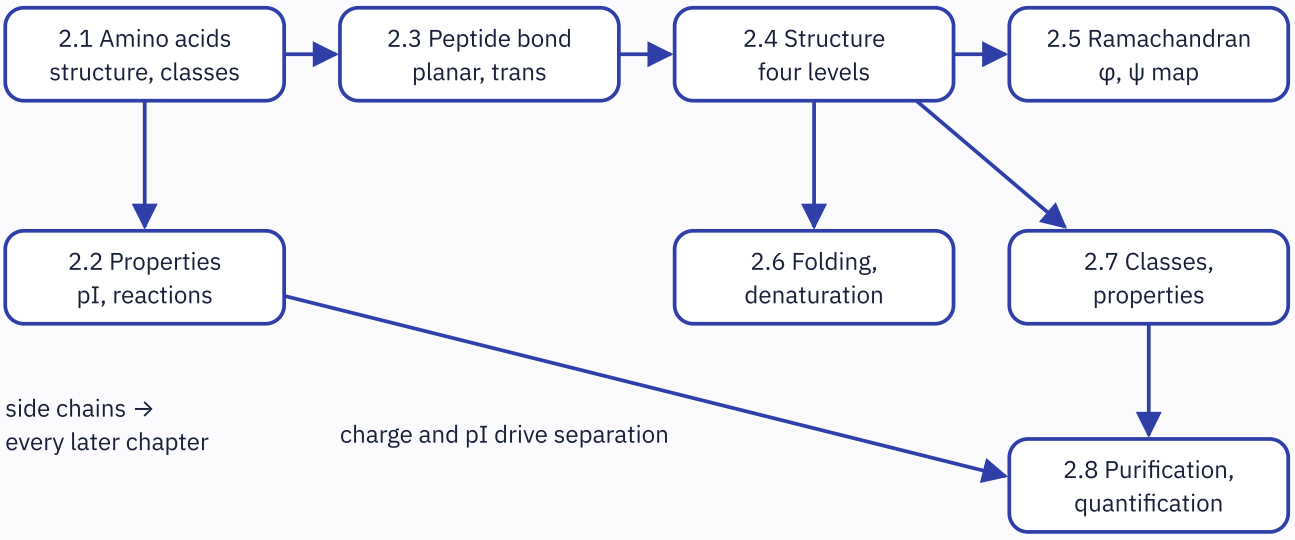


Figure 2.S.1 How the chapters connect. Side-chain chemistry (2.1, 2.2) feeds everything; the peptide bond (2.3) limits backbone shapes (2.4, 2.5); structure explains folding (2.6) and protein properties (2.7), and charge and size are what purification uses (2.8).

Takeaways

1. All 20 standard amino acids share one backbone; the side chain R decides polarity, charge and reactivity. Learn them in five groups.
2. Amino acids are zwitterions at pH 7, amphoteric, and have zero net charge at their pI.
3. The peptide bond is planar and trans because of resonance; the chain can rotate only at φ and ψ.
4. Primary structure fixes all higher levels; secondary structure uses backbone H-bonds; tertiary and quaternary use side-chain interactions, mainly the hydrophobic effect.
5. The Ramachandran plot shows which (φ, ψ) pairs avoid steric clashes; helices and sheets sit in the allowed regions.
6. Anfinsen showed that sequence alone decides the fold; chaperones only prevent mistakes.
7. Denaturation destroys shape and activity but not peptide bonds; it may be reversible or irreversible.
8. Proteins are classified by shape, composition (simple, conjugated, derived), solubility and function, and detected by colour tests.
9. Purification combines steps that use different properties; specific activity tracks purity, and A_{280} or dye methods measure amount.

Exam question bank

Remember and understand

- REMEMBER **Classify amino acids on the basis of side chain, nutrition and metabolic fate, with examples.** 10 marks

  Outline: general structure with Cα; five side-chain groups with all members (nonpolar aliphatic, aromatic, polar uncharged, positive, negative); essential nine and semi-essential Arg; glucogenic, ketogenic (Leu, Lys) and both; mention non-standard (selenocysteine, hydroxyproline) for full marks.

- UNDERSTAND **Describe the structure and characteristics of the peptide bond. Why is it planar and trans?** 8 marks

  Outline: condensation reaction with equation; resonance and partial double bond; bond length about 1.32 Å compared with single and double; six atoms in plane; trans preferred to avoid clash, cis before Pro; polar, H-bond donor and acceptor; rotation only at φ and ψ; diagram.

- UNDERSTAND **Explain the four levels of protein structure and the bonds that stabilise each, with examples.** 16 marks

  Outline: primary (sequence, sickle cell); secondary (α-helix with numbers, β-sheet parallel and antiparallel, β-turn, collagen); tertiary (fold of one chain, hydrophobic core, domains, myoglobin); quaternary (subunits, haemoglobin $\alpha_2\beta_2$, cooperativity); table of bonds (peptide, disulfide, H-bond, ionic, hydrophobic, van der Waals); a labelled diagram.

- UNDERSTAND **Classify proteins on the basis of composition, shape, solubility and function.** 12 marks

  Outline: simple (seven classes by solubility with examples), conjugated (six classes by prosthetic group, including glycoproteins and lipoproteins), derived (primary, secondary); fibrous versus globular; functional table; complete and incomplete proteins.

- UNDERSTAND **What is the Ramachandran plot? Explain its basis and the allowed regions, with a sketch.** 10 marks

  Outline: φ, ψ, ω defined; hard-sphere contact distances; fully allowed, partially allowed, disallowed; regions for α-helix, β-sheet, left-handed helix with angles; glycine and proline; use in structure validation; sketch with labelled axes.

Apply

- APPLY **Draw the titration curve of glycine and calculate its pI. Find the pI of aspartate (pKa 1.88, 3.65, 9.60) and lysine (pKa 2.18, 8.95, 10.53), and predict their direction of movement at pH 6.** 12 marks

  Outline: curve with two plateaus, pKa 2.34 and 9.60 marked, ionic forms; pI Gly 5.97; Asp 2.77 (moves to anode); Lys 9.74 (moves to cathode); Gly near pI barely moves; rule for averaging the two pKa values around the neutral form.

- APPLY **For the peptide Lys-Ala-Glu-Cys-Gly: name it, count peptide bonds, give its net charge at pH 7 and list the colour tests it would answer.** 8 marks

  Outline: lysylalanylglutamylcysteinylglycine; 4 peptide bonds; charges +1 (N-term) +1 (Lys) −1 (Glu) −1 (C-term), Cys mostly neutral, net 0; biuret positive (≥2 bonds), ninhydrin positive (free N-terminal amino group), lead acetate positive (Cys); xanthoproteic, Millon's, Hopkins–Cole negative (no aromatic residues).

- APPLY **A protein has a 25-residue α-helix and a 7-residue β-strand. Calculate the length of each and the number of helical turns; explain which forms a membrane-spanning segment.** 8 marks

  Outline: helix 25 × 1.5 = 37.5 Å, 25/3.6 ≈ 6.9 turns; strand about 7 × 3.5 = 24.5 Å; membrane core about 30 Å needs about 20 helical residues of hydrophobic side chains; helices satisfy their backbone H-bonds internally, so they are stable inside the lipid layer.

- APPLY **Prepare a purification table from given data and calculate specific activity, yield and fold purification; suggest the best step and an order of methods.** 10 marks

  Outline: formulas; worked values (as in Chapter 2.8); identify the step with the biggest fold increase per loss; order from salting out to ion exchange, gel filtration, affinity; check by SDS-PAGE; concentration by A_{280} with Beer–Lambert.

Analyze and evaluate

- ANALYZE **Describe Anfinsen's experiment on ribonuclease and evaluate what it proves about protein folding. Why do cells still need chaperones?** 12 marks

  Outline: 8 M urea and 2-mercaptoethanol; slow removal gives full activity; oxidation in urea gives scrambled enzyme with about 1 % activity (1 of 105 pairings correct); trace thiol restores activity; thermodynamic hypothesis; crowded cell, exposed hydrophobic patches, aggregation; Hsp70, GroEL–GroES, PDI, prolyl isomerase; Levinthal paradox and funnel.

– ANALYZE **Explain, using side-chain chemistry, why a single substitution Glu → Val in haemoglobin causes disease while many other substitutions are harmless.** 10 marks

Outline (spans 2.1 and 2.4): charged to hydrophobic on the surface; sticky patch binds a neighbouring deoxy-haemoglobin; polymer fibres; sickled cells; contrast with conservative substitutions (Leu → Ile, Asp → Glu) that keep size and polarity; position matters (surface versus core, active site).

– ANALYZE **Explain how the isoelectric point is used in three different laboratory or industrial situations.** 10 marks

Outline (spans 2.2, 2.7 and 2.8): minimum solubility at pI (isoelectric precipitation, casein and paneer); choice of anion or cation exchanger from pH relative to pI; isoelectric focusing separates proteins by pI; electrophoresis direction; buffering near pKa.

– EVALUATE **Compare UV absorbance at 280 nm, biuret, Lowry, Bradford and BCA methods for protein estimation. Which would you choose for a dilute crude extract rich in nucleic acids, and why?** 10 marks

Outline: principle, wavelength and sensitivity of each; standard curve with BSA; interferences (nucleic acids absorb near 280 nm; detergents affect Bradford; reducing agents affect Lowry and BCA); choose Bradford or BCA for dilute samples with nucleic acid; justify.

– ANALYZE **Outline solid-phase peptide synthesis of a tripeptide Ala-Gly-Phe and explain the role of protecting groups and coupling reagents.** 10 marks

Outline: attach Fmoc-Phe to resin; remove Fmoc with piperidine; couple Fmoc-Gly with DCC or DIC plus HOBt; deprotect; couple Fmoc-Ala; deprotect; cleave with TFA; direction C → N; orthogonal protection; washing replaces purification between steps.

Glossary

Affinity chromatography	Separation by specific binding of the target protein to a fixed ligand.
Amphoteric	Able to act as both an acid and a base.
Amyloid	Insoluble, β-sheet-rich fibres formed by misfolded proteins.
Chaperone	A protein that prevents wrong folding and aggregation of other proteins.
Conjugated protein	A protein with a non-protein prosthetic group, such as sugar, lipid or haem.
Denaturation	Loss of native shape and activity without breaking peptide bonds.

Dialysis	Removal of small molecules through a semipermeable membrane that keeps proteins.
Disulfide bridge	A covalent –S–S– bond between two cysteine side chains.
Domain	A compact, independently folded part of a larger protein.
Essential amino acid	One the body cannot make in enough amount; it must come from food.
Fibrous protein	A long, insoluble, structural protein such as keratin or collagen.
Globular protein	A compact, usually soluble protein such as an enzyme or haemoglobin.
Hydrophobic effect	The tendency of nonpolar groups to cluster away from water, which drives folding.
Isoelectric point (pI)	The pH at which a molecule has zero net charge.
Levinthal's paradox	A random search of all shapes would take far too long, so folding must follow guided paths.
Molten globule	A partly folded state with secondary structure but loose packing.
Peptide bond	The amide bond between the α-carboxyl of one amino acid and the α-amino group of the next.
φ and ψ	Torsion angles about the N–Cα and Cα–C bonds of each residue.
Prosthetic group	A non-protein part tightly bound to a protein and needed for its function.
Residue	An amino acid unit within a peptide chain.
Salting out	Precipitation of proteins by high salt concentration.
SDS-PAGE	Gel electrophoresis of SDS-coated proteins, separating them by subunit size.
Specific activity	Enzyme activity per mg of protein; a measure of purity.
Tm (melting temperature)	Temperature at which half the protein molecules are unfolded.
Zwitterion	A molecule with both positive and negative charges and no net charge.

Sources and further reading

Textbooks

- David L. Nelson and Michael M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman; chapters on amino acids, peptides and proteins, and on protein structure (from memory; check). Source of the pKa values used here.
- Jeremy M. Berg, John L. Tymoczko, Gregory J. Gatto Jr. and Lubert Stryer, *Biochemistry*, W. H. Freeman; chapters on protein composition and structure and on exploring proteins (from

memory; check).

- Donald Voet and Judith G. Voet, *Biochemistry*, Wiley (from memory; check).
- U. Satyanarayana and U. Chakrapani, *Biochemistry*, Elsevier India; widely prescribed in Indian universities for the classification of proteins and colour reactions (from memory; check).

Open resources opened for this book

- G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations", *Journal of Molecular Biology* 7 (1963) 95–99, record at the Indian Academy of Sciences repository: repository.ias.ac.in/49564.
- C. B. Anfinsen, "Principles that govern the folding of protein chains", *Science* 181 (1973) 223–229 (the Nobel lecture): PDF copy. Source of the ribonuclease details in Chapter 2.6.
- Nobel Prize page for Anfinsen's lecture, 1972: nobelprize.org.

Syllabus: the University free n easy universal syllabus, built from top universities' syllabi.

Built for the University free n easy YouTube channel from the universal syllabus of Biochemistry (Oct 2026). Written with top universities' syllabi and standard textbooks; check a formula against your prescribed book before an exam, and report mistakes on the channel.