Amino Acids & Protein Structure

BCH 100 — Introductory Biochemistry · Dr. Radi

By the end of this unit, you can…

- Identify and draw all 20 standard amino acids and classify each by its side chain
- Recognize the special features of glycine, proline, cysteine, histidine, and the aromatics

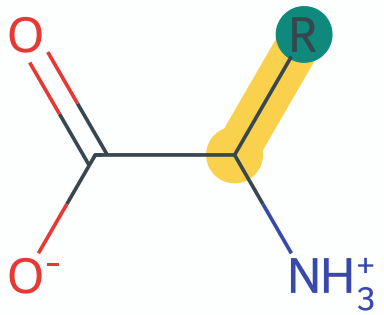
Today's route

1. Amino Acid Structure & the Zwitterion
2. Amino Acid Stereochemistry
3. The 20 Amino Acids, by R Group
4. Amino Acid Ionization & pI
5. The Peptide Bond & Primary Structure
6. Secondary Structure — α-Helix & β-Sheet
7. Tertiary Structure — the 3-D Fold
8. Quaternary Structure · Globular & Fibrous
9. Myoglobin & Hemoglobin
10. Oxygen Binding & Cooperativity
11. Denaturation & Misfolding

1 · Amino Acid Structure & the Zwitterion

"Name the four groups on the α-carbon and explain why an amino acid is a zwitterion at physiological pH."

What is an amino acid?



Twenty little building blocks, one shared design. Each has a central **α-carbon** carrying four things — an **amino group**, a **carboxyl group**, a **hydrogen**, and a **side chain (R)**. String them together and you get a **protein**.

The α-carbon holds four groups

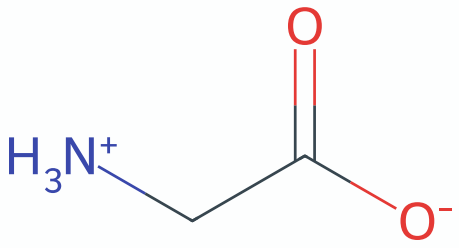
The **α-carbon** is carbon #2 — one carbon over from the carboxyl. Everything hangs off it:

- **–NH_3^+** — the amino group
- **–COO^-** — the carboxyl group
- **–H** — a hydrogen
- **–R** — the side chain that makes each amino acid **unique**

amino group + carboxyl group + H + R, all on one carbon

only R changes from one amino acid to the next

At body pH, it's a zwitterion



Look closely: the amino group is **protonated (+)** and the carboxyl is **deprotonated (−)** at the same time. That double-charged, net-**zero** form is a **zwitterion** — it's how amino acids actually exist in your cells (~pH 7), not the neutral form textbooks sometimes draw.

Twenty, and they share a look

20 standard amino acids · coded by RNA · average MW ≈ 110 g/mol

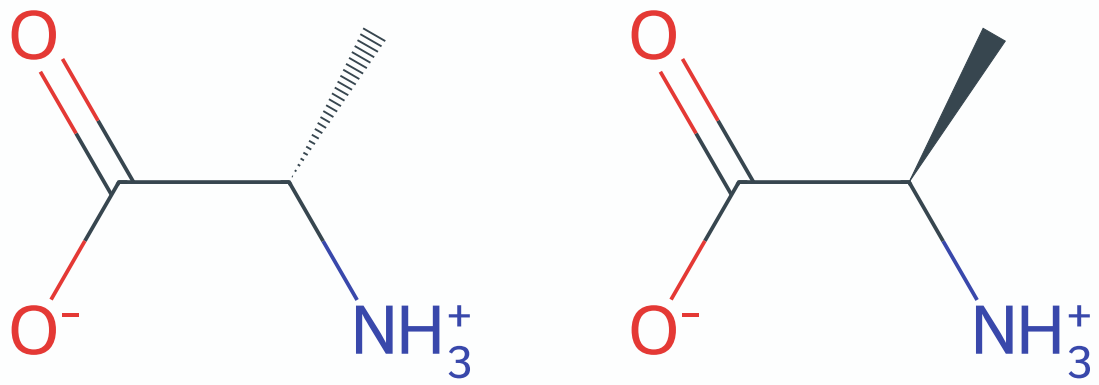
19 of the 20 have a chiral α-carbon — glycine is the exception

Nature actually makes **hundreds** of amino acids — but only these **20** get built into your proteins. Same backbone every time; the twenty different R groups are the whole story from here on.

2 · Amino Acid Stereochemistry

"Explain why the α-carbon is chiral, read a Fischer projection, and know that proteins are built from L-amino acids."

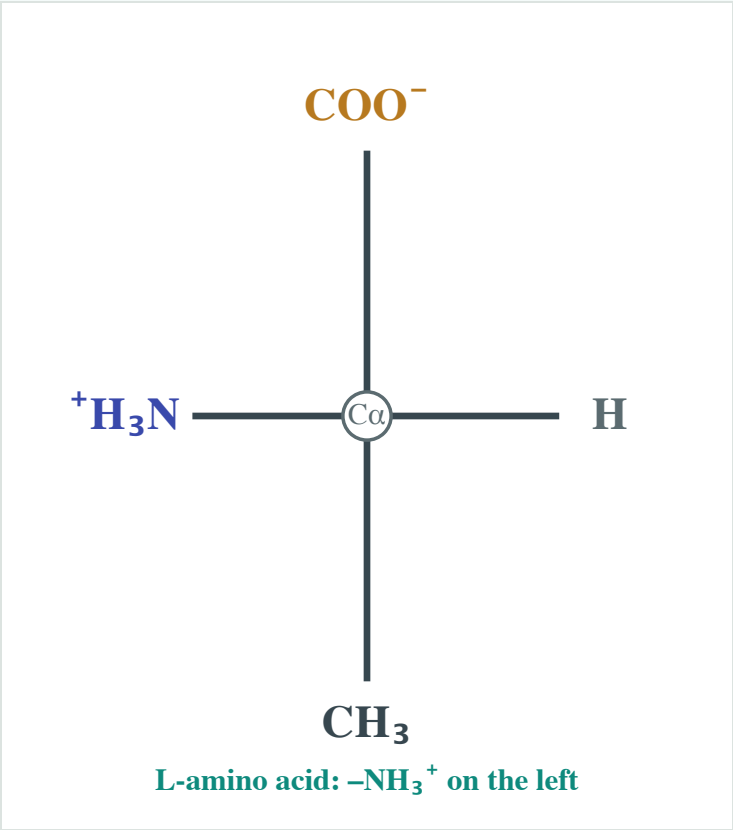
The α-carbon is a mirror trick



Four **different** groups on one carbon means it's a **chiral center** — the molecule and its mirror image can't be superimposed, like your **left and right hands**. These two alanines have the same atoms and bonds, yet they're not the same molecule.

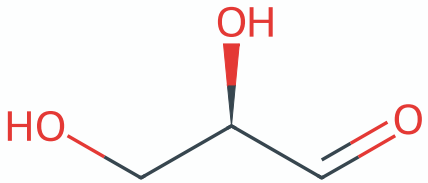
💡 Only **glycine** escapes — its R group is another H, so two of the four groups match and it's **achiral**.

Fischer projections

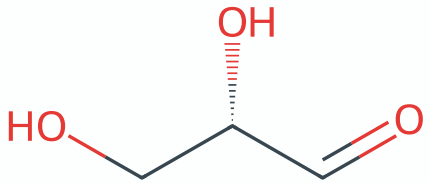


A quick way to draw a chiral center flat. The rule: **horizontal bonds point toward you**, **vertical bonds point away**. For an amino acid drawn this way, **L** means the **–NH_3^+ sits on the left**.

D or L — set by glyceraldehyde



D-glyceraldehyde



L-glyceraldehyde

Chirality is named against one reference molecule, **glyceraldehyde**. From it we get two whole families:

- **sugars** are almost all **D**
- **amino acids** in proteins are almost all **L**

Life is picky — it committed to one mirror form and never looked back.

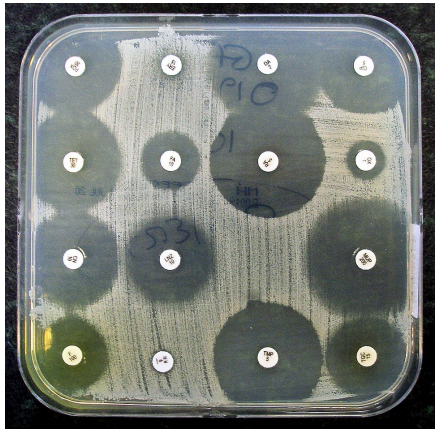
Your proteins are all L

Of the twenty, **19 are chiral** (glycine is the lone exception) — and every single one your ribosome puts into a protein is the **L** form. The enzymes that build and read proteins are themselves L-handed, so they only fit L partners.

19 chiral amino acids · all L in proteins · glycine is achiral

one consistent handedness lets proteins fold and enzymes recognize their substrates

In the body: when D shows up

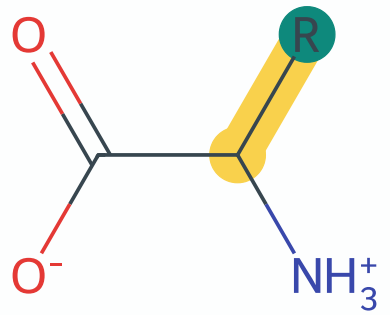


- **Bacterial cell walls** use **D-alanine** cross-links our enzymes can't cut — a perfect drug target. **Penicillin** blocks the enzyme that builds them; each **clear ring** above is bacteria that couldn't grow near the drug.
- A few residues slowly **racemize** L→D over a lifetime (your eye's lens) — a tiny molecular clock.

3 · The 20 Amino Acids, by R Group

"Sort all 20 standard amino acids into five families by their side chain — know what each family does, and where each shows up in the body."

Same backbone, twenty personalities



Every amino acid is built on the **exact** same core — an α-carbon holding an amino group, a carboxyl group, and a hydrogen. Only the **R group** changes. That one swap decides everything: size, charge, and whether the residue **loves water** or **runs from it**.

Five families, twenty members

We sort all twenty by what the side chain is like:

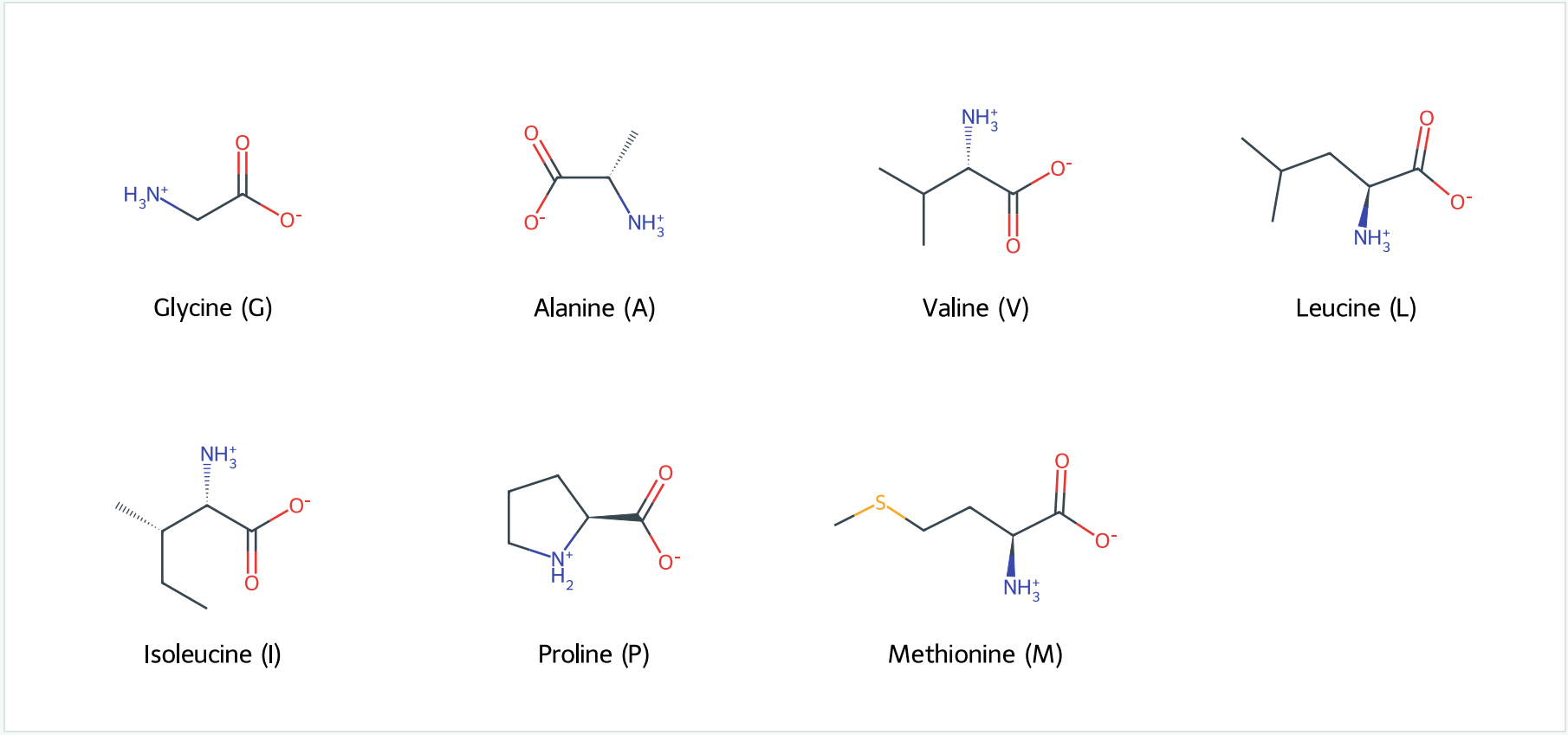
- **Nonpolar, aliphatic** (7) – **greasy, water-fearing**
- **Aromatic** (3) – flat rings, mostly **hydrophobic**
- **Polar, uncharged** (5) – **H-bond happily with water**
- **Acidic** (2) – **negative** at pH 7
- **Basic** (3) – **positive** at pH 7

7 + 3 + 5 + 2 + 3 = 20

every standard amino acid lands in exactly one family

💡 Your job: **KNOW HOW TO DRAW ALL 20** and which family each belongs to. Name → structure → category.

Nonpolar, aliphatic

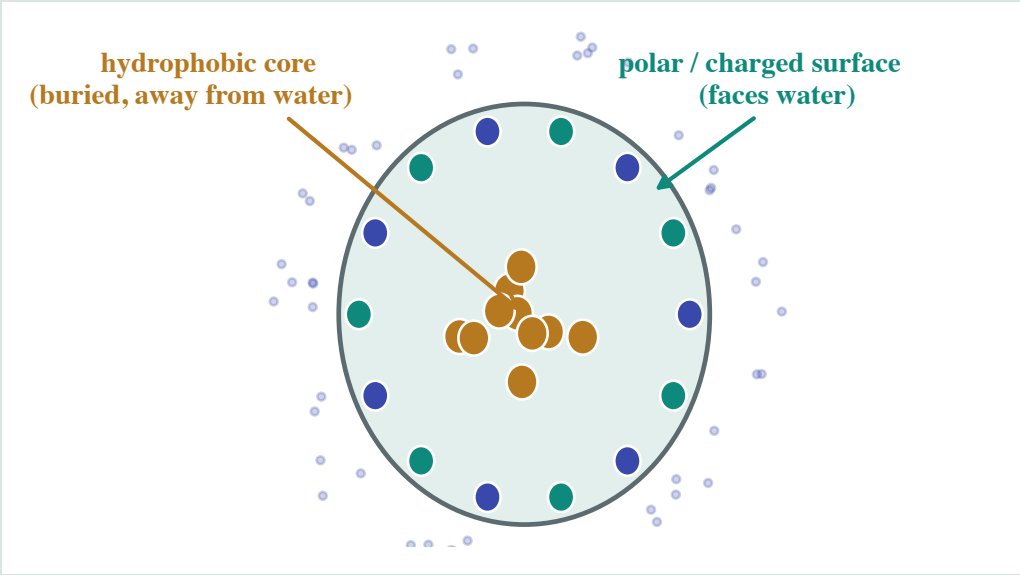


Plain hydrocarbon side chains — **greasy and water-avoiding**. They bury themselves in the core of a folded protein.

Nonpolar, aliphatic – features

- **Glycine (G):** the R group is just an H. The smallest, the most flexible – and the **only achiral** amino acid!
- **Ala → Val → Leu → Ile:** more –CH_2– = **more hydrophobic**! (Isoleucine even sneaks in a second chiral carbon.)
- **Proline (P): so weird!** Its side chain loops back onto the nitrogen – the α-"amino" group is really a 2° amine. That ring **kinks the backbone**.
- **Methionine (M):** carries a sulfur, and it's the **start** amino acid for every new protein. (Scientists named it: **methyl + thio**.)

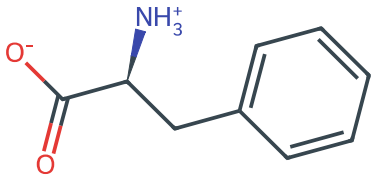
Where they live: the hydrophobic core



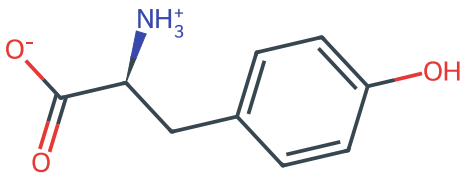
Greasy side chains hide in the **buried core** of a folded protein; the polar and charged ones sit on the **surface** facing water. That sorting is a big part of what makes a protein **fold**.

Glycine, so tiny, is the only residue that fits inside the tightly packed triple helix of **collagen**. Swap it for anything bigger and the helix can't close — that's **osteogenesis imperfecta**, brittle-bone disease.

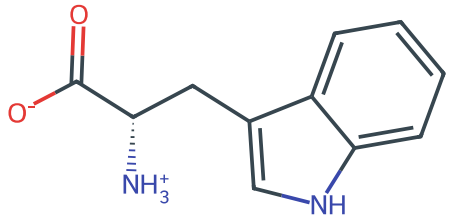
Aromatic



Phenylalanine (F)



Tyrosine (Y)



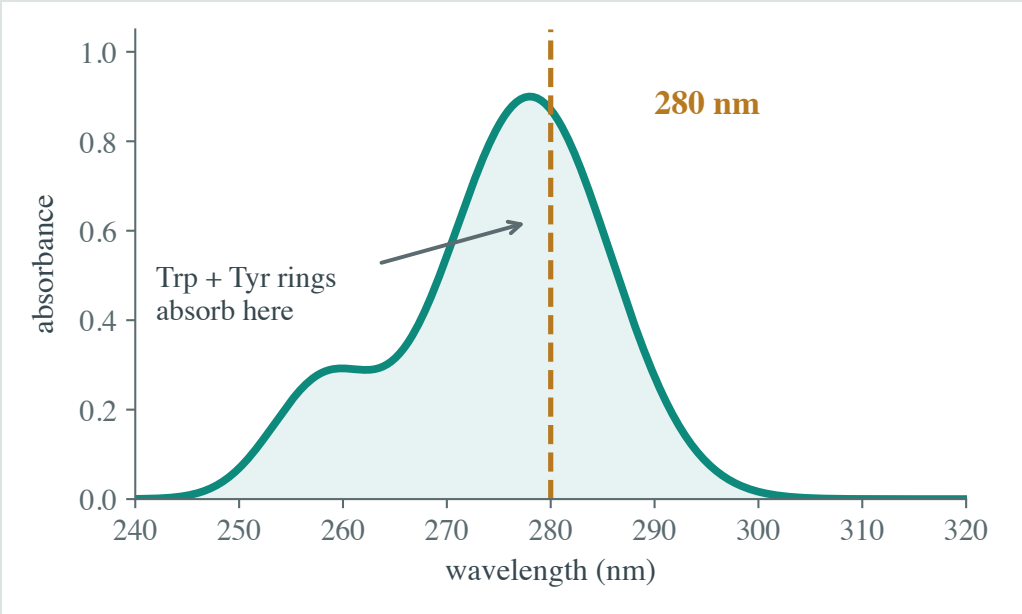
Tryptophan (W)

Big, flat ring systems — bulky and, mostly, **hydrophobic**.

Aromatic – features

- **Phenylalanine (F):** a plain benzene ring, very **hydrophobic**. (Scientists named it: **phenyl + alanine**.)
- **Tryptophan (W):** the bulkiest of all twenty. Hydrophobic – though **less** than Phe. (Why? Its ring nitrogen can hydrogen-bond a little.)
- **Tyrosine (Y):** the ring **–OH** makes it **polar** – and that –OH gets **phosphorylated**, ionized, even radicalized.

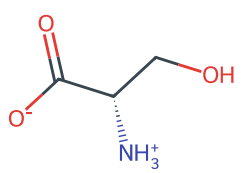
In the body: reading protein by UV



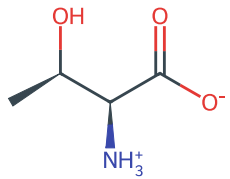
Those rings soak up UV at **~280 nm** — so one quick A_{280} reading tells you how much protein is in a tube. Every biochemist uses it daily.

And **PKU:** some people can't break phenylalanine down, so it piles up and harms the developing brain. It's why a diet soda warns, **"Phenylketonurics: contains phenylalanine."**

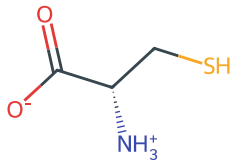
Polar, uncharged



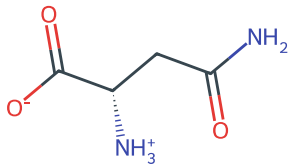
Serine (S)



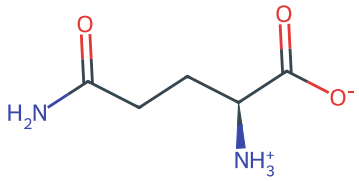
Threonine (T)



Cysteine (C)



Asparagine (N)



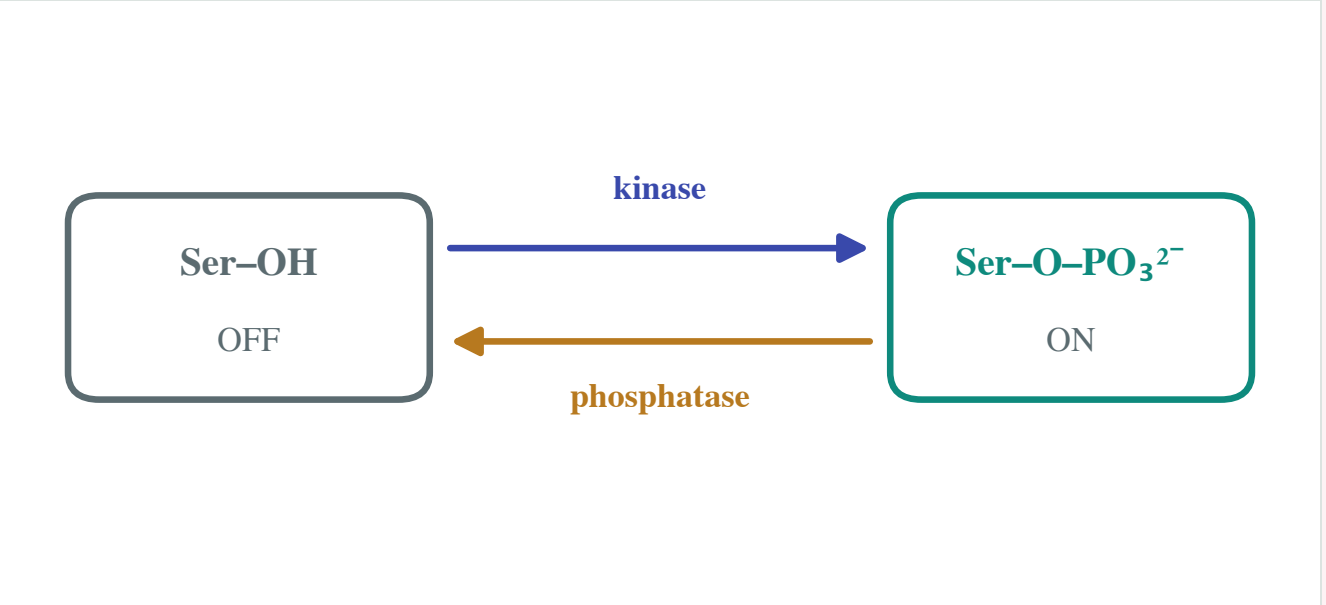
Glutamine (Q)

No net charge at pH 7 – but these side chains **hydrogen-bond happily with water**.

Polar, uncharged – features

- **Serine (S) & Threonine (T):** –OH side chains that H-bond – and get **phosphorylated**. **Huge** in signaling!
- **Cysteine (C):** the –SH can bond another Cys → a **disulfide bridge** that staples a protein together. (It's also what a perm does to your hair!)
- **Asparagine (N):** named after a veggie – **asparagus**! Its amide side chain H-bonds but is **not ionizable**.
- **Glutamine (Q):** one of the body's main nitrogen carriers. Also an amide – no charge, ever.

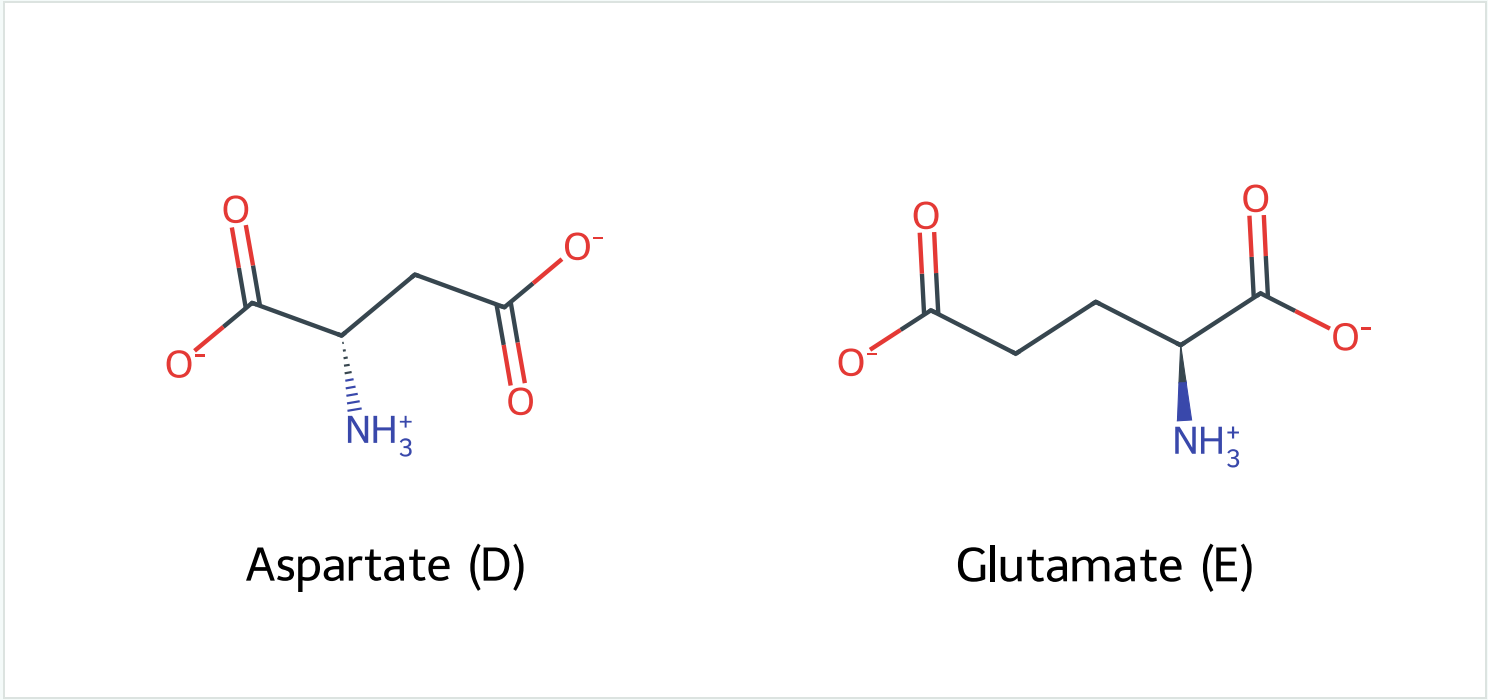
In the body: the on/off switch



Adding a phosphate to a **Ser, Thr, or Tyr** flips a protein **ON** or **OFF** — a **kinase** puts it on, a **phosphatase** takes it off.

This single trick runs most of cell signaling — and it's why a huge share of cancer drugs are **kinase inhibitors** aimed right at this switch.

Negatively charged (acidic)

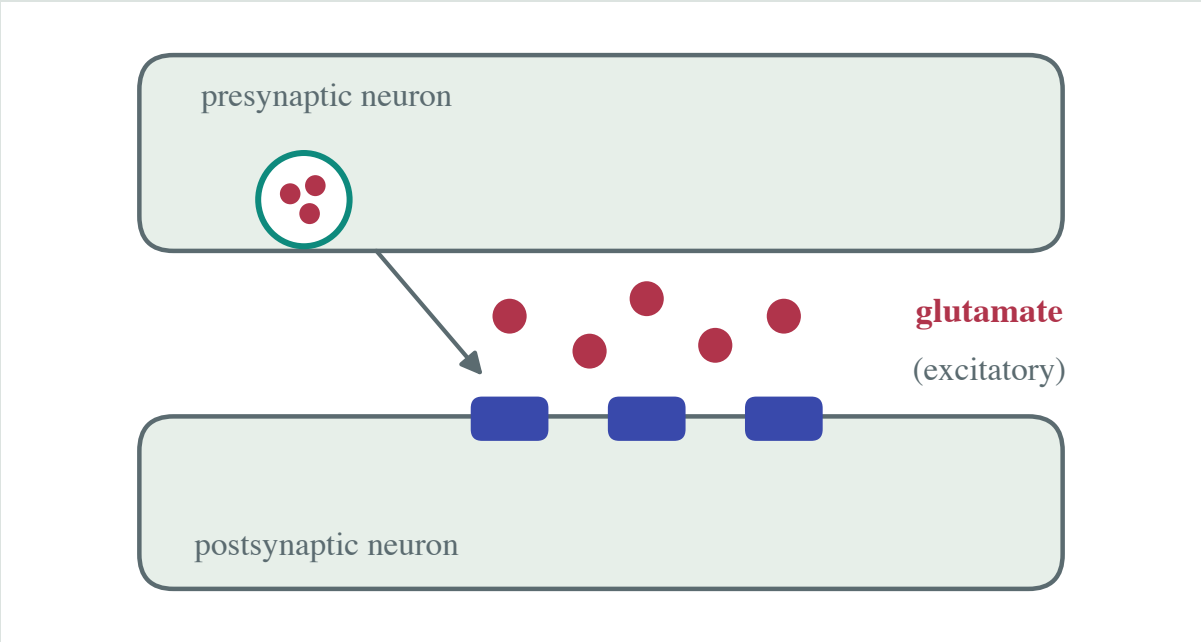


Carboxylic-acid side chains – at pH 7 they've given up a proton and carry a **negative charge**.

Acidic – features

- **Aspartate (D) & Glutamate (E):** the **"-ate"** ending is the tell – the side-chain –COOH is deprotonated to **$-COO^-$** at pH 7.
- Two **negatives** to place on a protein's surface, or inside an active site to grab a **positive** substrate.
- And **glutamate** moonlights as your brain's main excitatory signal.

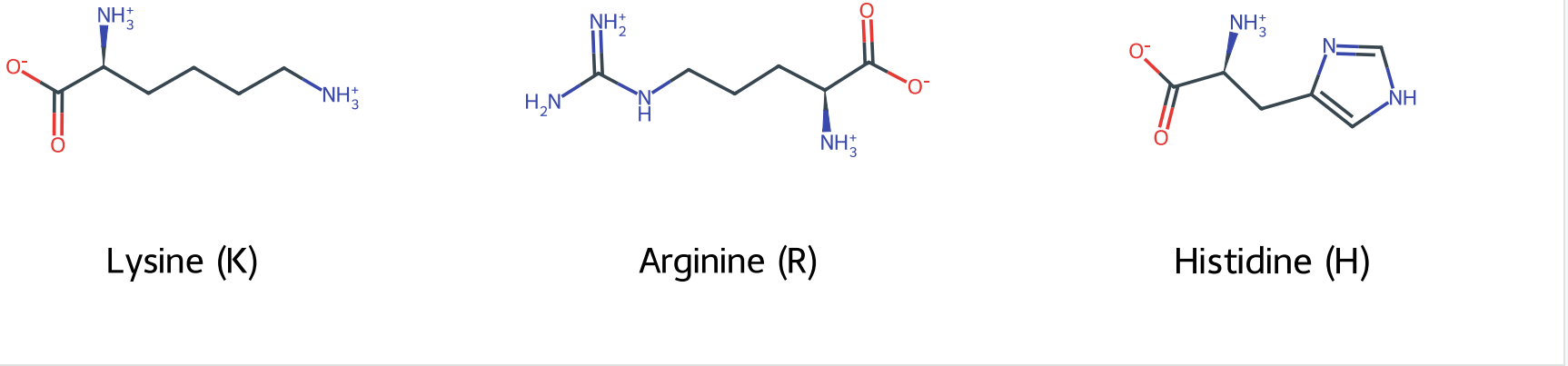
In the body: glutamate – taste & brain



That extra **–COO⁻** is exactly what your tongue reads as **umami** – yes, this is **MSG**.

In your brain, **glutamate** is the main **excitatory** neurotransmitter. Useful in small doses – but **too much** is **excitotoxic**, part of how neurons die after a stroke.

Positively charged (basic)



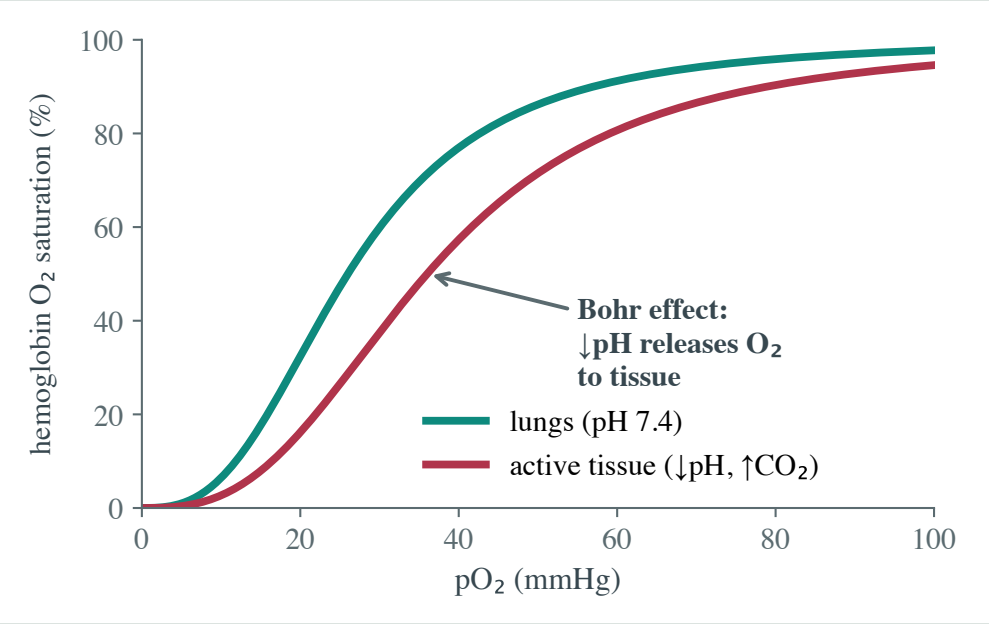
Lysine (K) Arginine (R) Histidine (H)

Nitrogen-rich side chains that grab a proton and carry a **positive charge** at pH 7.

Basic – features

- **Lysine (K) & Arginine (R):** protonated and **positive** at pH 7 – perfect for gripping the **negative** backbone of DNA.
- **Histidine (H):** its imidazole pKa sits right around **6–7**, so it flips between charged and neutral near physiological pH.
- That flip makes histidine the **proton shuttle** of choice in enzyme active sites.

In the body: histidine, the proton broker

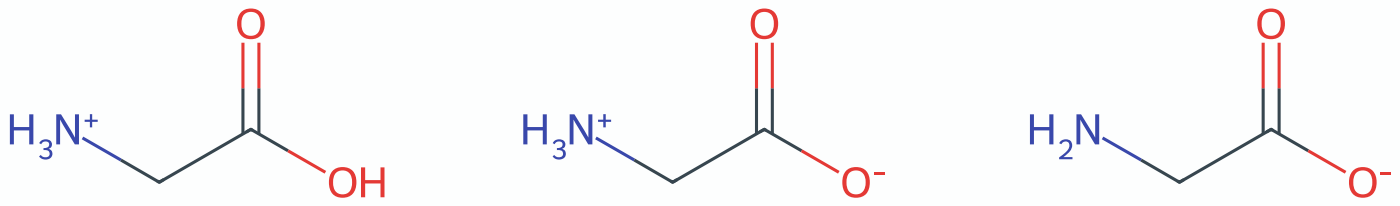


Histidine's pKa sits **right at** physiological pH, so it grabs and releases protons on demand. That powers hemoglobin's **Bohr effect**: in acidic, hard-working tissue, His picks up protons and hemoglobin **dumps its O_2** right where it's needed — and it's why His is the go-to residue in enzyme **active sites**.

4 · Amino Acid Ionization & pI

"Track how an amino acid's charge changes with pH, read its titration curve, and calculate its isoelectric point."

Amino acids are acids AND bases



+1 (low pH) 0 (zwitterion) −1 (high pH)

Every amino acid has at least two ionizable groups — an **α-carboxyl** and an **α-amino** — so as pH changes it gains or loses protons. At physiological pH it's a **zwitterion**: **+** on one end, **−** on the other, net **zero**.

Each group has its own pKa

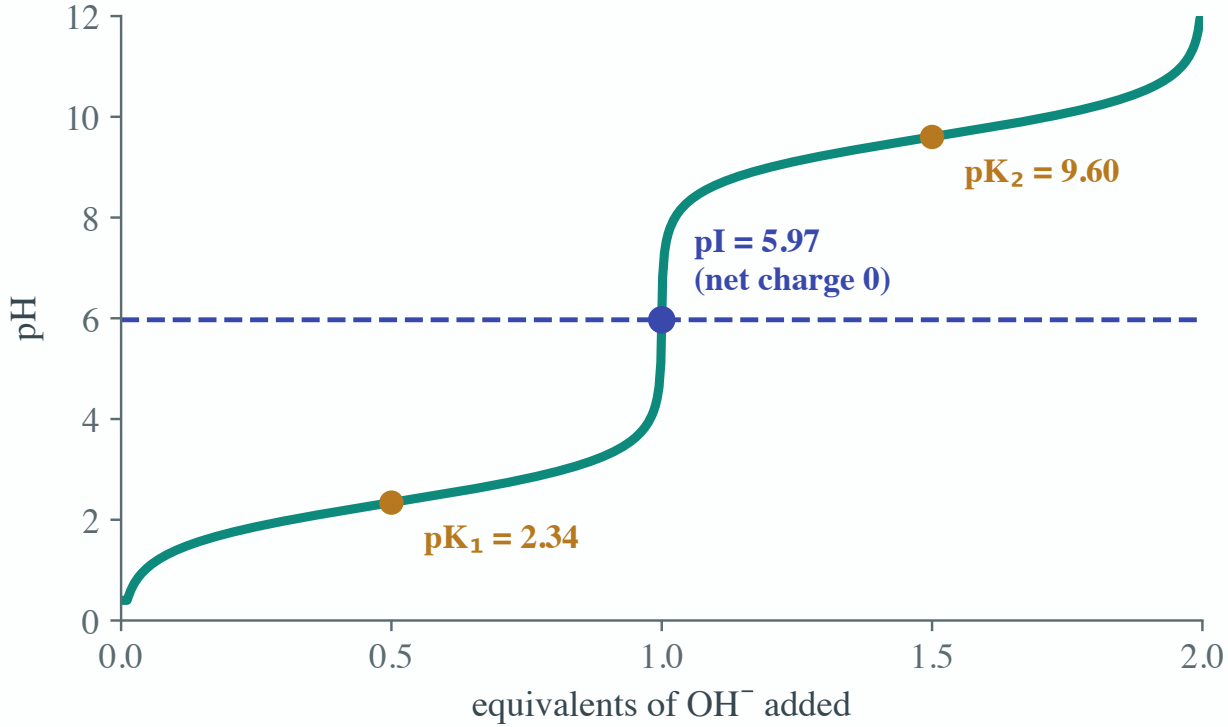
An ionizable group flips at its **pKa**: **below** the pKa it holds its proton, **above** the pKa it lets go.

- **α-COOH:** pKa ≈ **2** — lets go early, at low pH
- **α-NH_3^+:** pKa ≈ **9–10** — holds on until high pH
- **some R groups** ionize too — Asp, Glu, Lys, Arg, His, Cys, Tyr

below pKa → protonated · above pKa → deprotonated

at pH = pKa, the group is half protonated, half not

Reading a titration curve



Add base and the pH climbs in two steps — one **buffering plateau** per ionizable group, each centered on its **pKa**. Halfway between them the amino acid is a pure zwitterion: that pH is the **pI**.

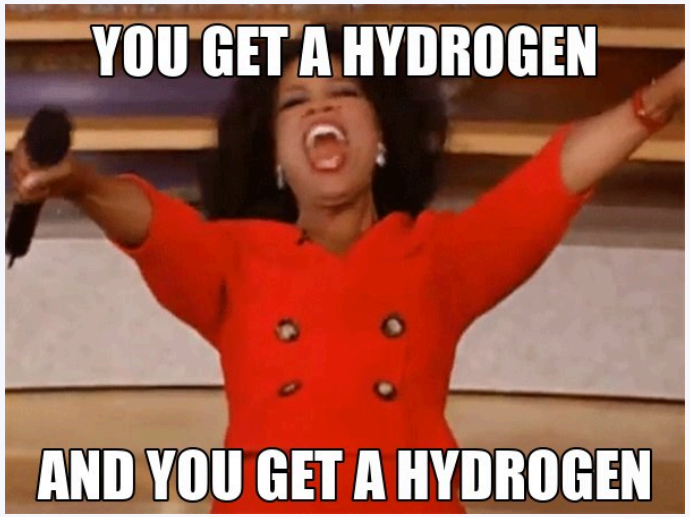
The isoelectric point, pI

The **pI** is the pH where the net charge is **exactly zero**. There the amino acid won't move in an electric field, and it's least soluble in water.

pI = average of the two pKa's flanking the net-zero form

for a simple amino acid: $pI = (pK_1 + pK_2) / 2 = (2.34 + 9.60) / 2 = 5.97$ for glycine

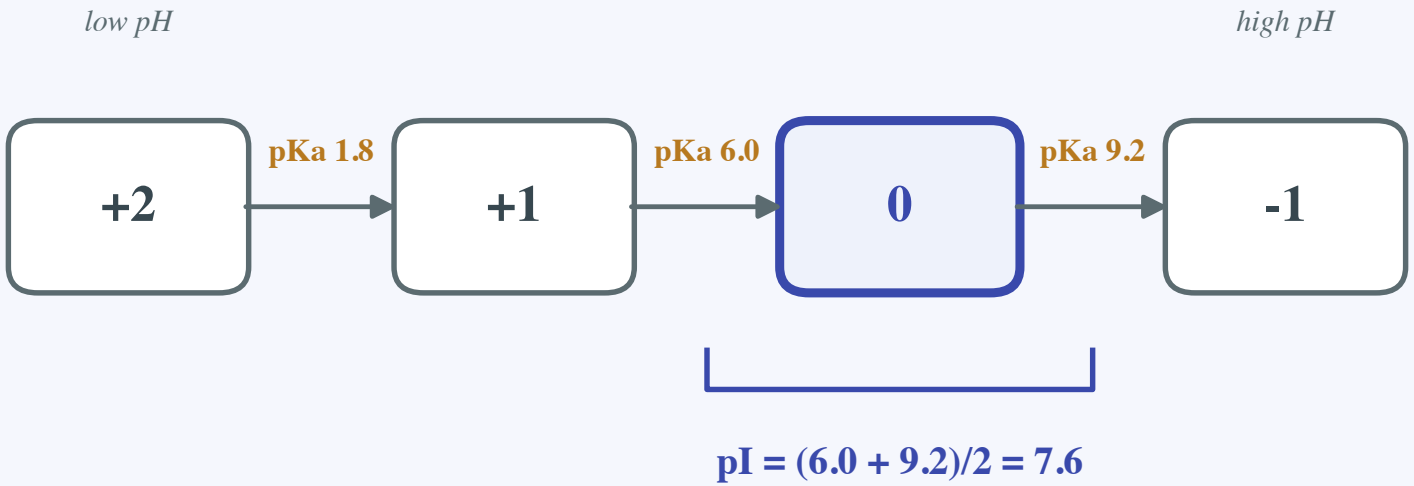
Steps to calculate pI



1. **Draw it fully protonated** — every ionizable group starts with its proton.
2. **List the ionizable groups + pKa's:** α-COOH (~2), α-NH_3^+ (~9–10), and the R group if it has one.
3. **Find the net-zero form.** Peel protons off in pKa order until you land on it.
4. **Average the two pKa's flanking** that form. That's your pI.

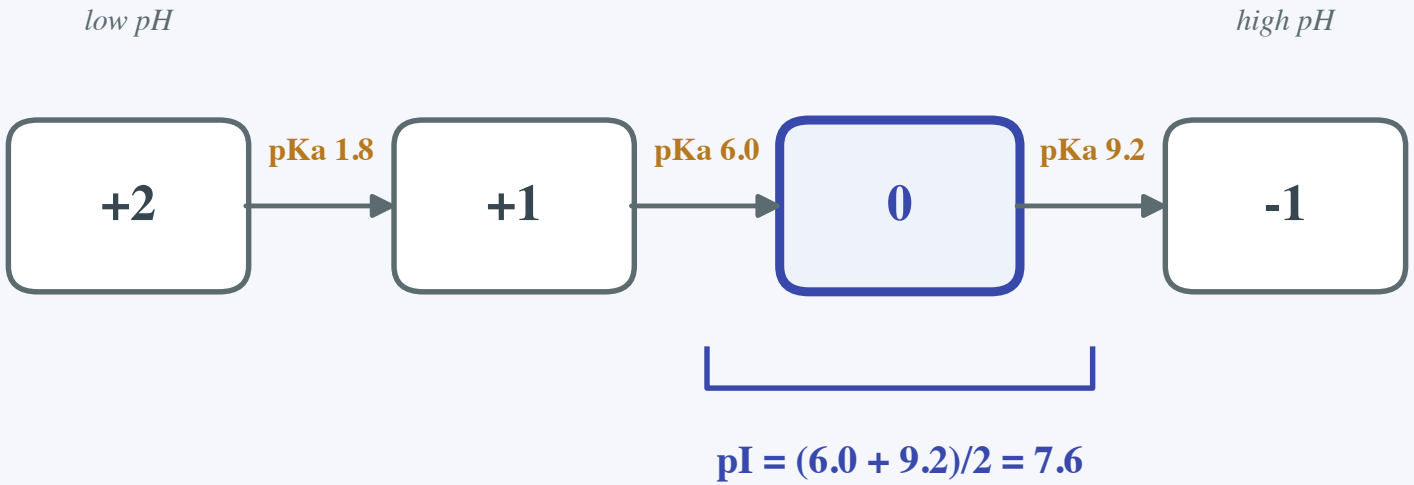
Calculating pI of histidine

Histidine has **three** ionizable groups, so we watch the net charge fall as pH climbs.



Calculating pI of histidine

Histidine has **three** ionizable groups, so we watch the net charge fall as pH climbs.



The net-zero form sits between the **imidazole (6.0)** and the **α-amino (9.2)** – so those are the two we average, and the ladder reads out **pI = 7.6**.

Try this one…

Glutamate — an acidic amino acid. Its pKa's: α-COOH **2.2**, side-chain COOH **4.2**, α-NH_3^+ **9.7**.

Your turn — find its pI.

Try this one…

Glutamate — an acidic amino acid. Its pKa's: α-COOH **2.2**, side-chain COOH **4.2**, α-NH_3^+ **9.7**.

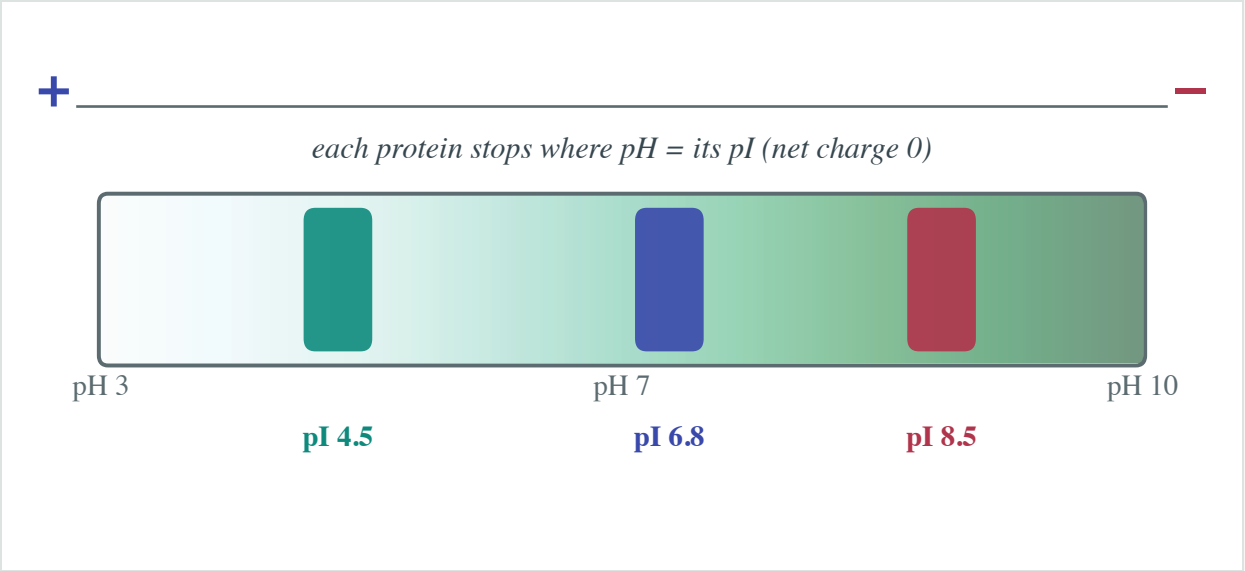
Your turn — find its pI.

Fully protonated, Glu is **+1**. Drop the α-COOH proton (2.2) and it's at net **0** — so the zero form sits between the **two carboxyls**.

$$pI = (2.2 + 4.2) / 2 = 3.2$$

an acidic amino acid has a low pI — it's **negative** at body pH

In the body: pI puts proteins to work

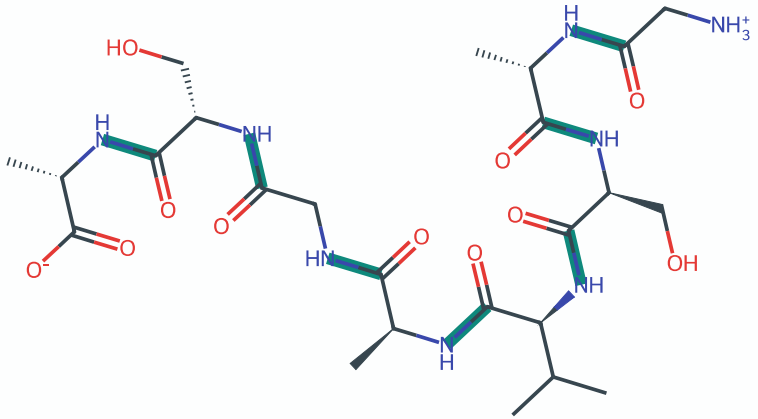


Give a mixture of proteins a **pH gradient** and an electric field, and each one **stops dead at its own pI** — net charge **zero**, so it quits moving. That's **isoelectric focusing**, how we separate proteins that differ by a single charge — and the same reason a protein is least soluble, **crashing out of solution**, right at its pI.

5 · The Peptide Bond & Primary Structure

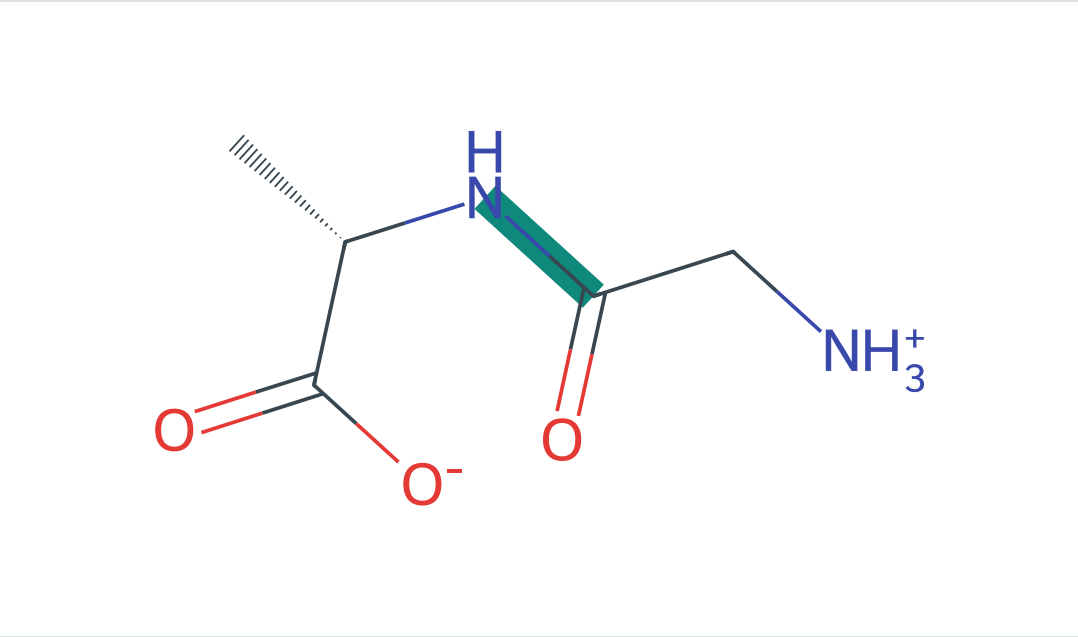
"Link amino acids into a chain through peptide bonds, read a sequence N→C, and see why the primary sequence dictates everything."

Amino acids link into a chain



A protein is just amino acids joined **head to tail** — here are eight. Each link is a **peptide bond** (highlighted in teal). Real proteins run **dozens to thousands** of residues long; that whole chain is a **polypeptide**, the raw material of every protein.

The peptide bond



The **carboxyl** of one amino acid and the **amino** of the next join and kick out a molecule of **water** (a **condensation**). What's left is an **amide** – flat and rigid, so the backbone can only swivel **between** the peptide bonds, not through them.

Direction matters: N → C

Every chain has a free **amino end** (the **N-terminus**) and a free **carboxyl end** (the **C-terminus**). By convention we always read and write a sequence from N to C.

H_3N^+–Met–Val–His–Leu–Thr– –COO^-

this ordered list of residues, read N → C, is the primary structure

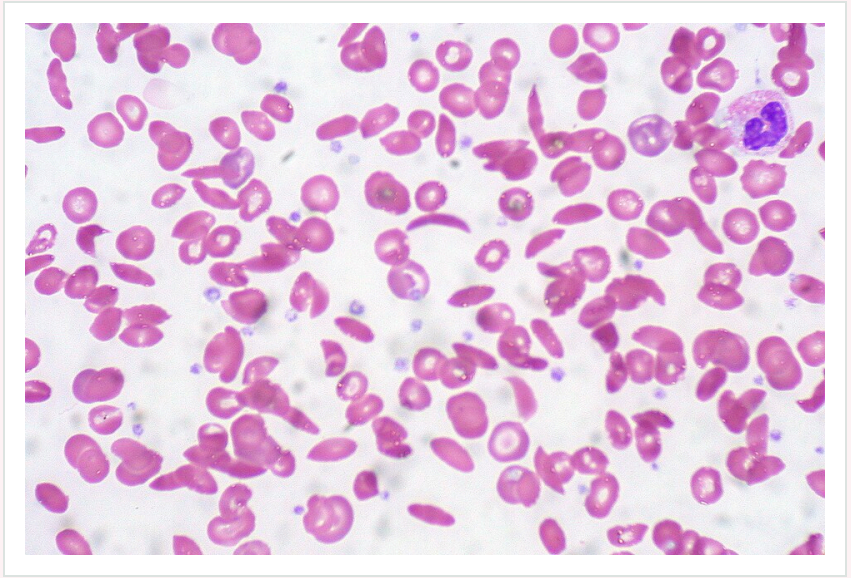
Primary structure runs the show

The sequence is not just a label — it **contains all the folding instructions**. Those side chains, in that order, decide which parts bury, which twist into helices, which pair into sheets. Get the sequence, and (in principle) you get the whole protein.

sequence → fold → function

change the sequence and you can change everything downstream

In the body: one letter — sickle cell

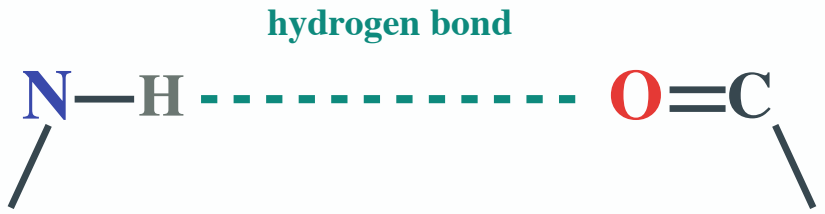


Swap a **single** amino acid — position 6 of hemoglobin's β-chain, **glutamate** → **valine** — and everything breaks. The new greasy patch makes deoxygenated hemoglobin **clump into fibers**, warping round red cells into rigid **sickles** that jam capillaries. One residue out of ~150 — a whole disease.

6 · Secondary Structure — α-Helix & β-Sheet

"Recognize the two secondary-structure motifs and see that both are stitched together by backbone hydrogen bonds."

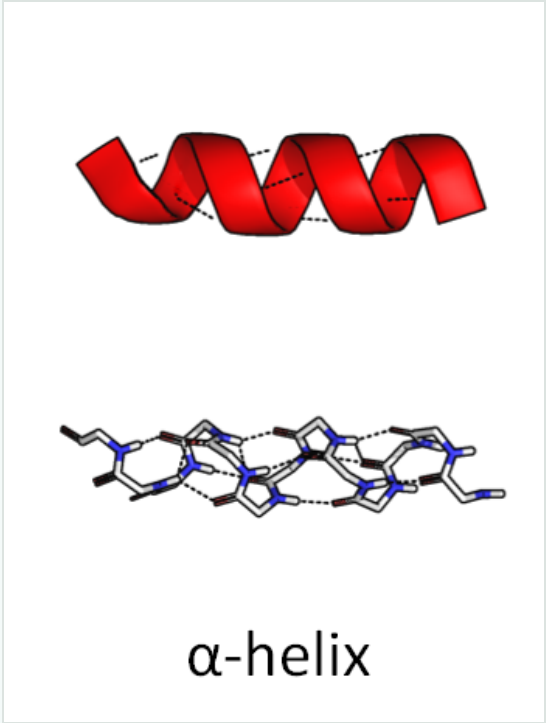
The backbone starts to fold



every α-helix and β-sheet is stitched together by these backbone N–H ⋯ O=C bonds

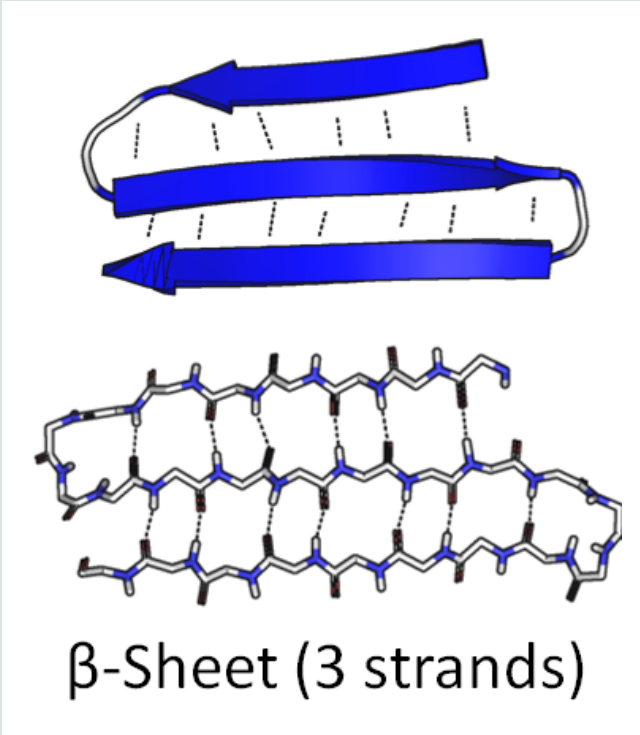
Once the chain exists, the **backbone can't help itself** — it folds into regular, repeating local shapes, each one zipped up by **hydrogen bonds** between a backbone **N–H** and a backbone **C=O**. Two shapes dominate every protein: the **α-helix** and the **β-sheet**.

The α-helix



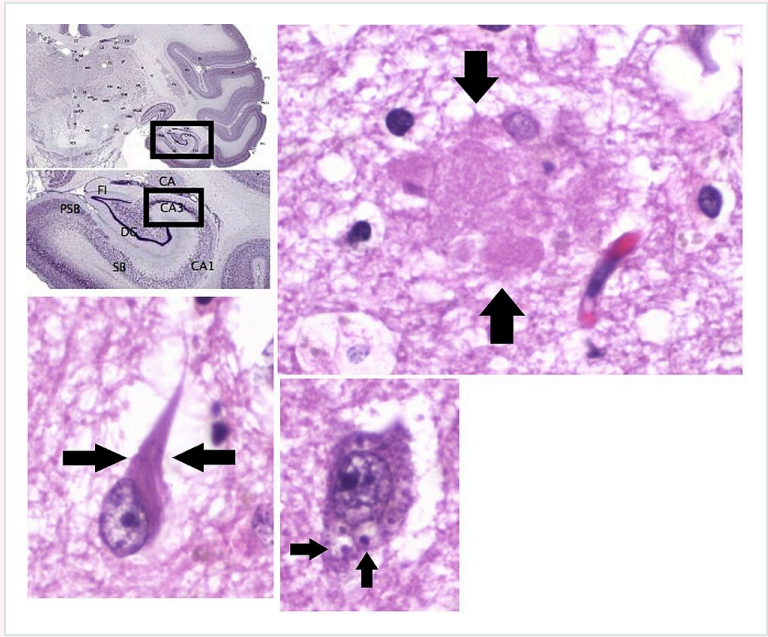
The backbone coils into a tight **right-handed spiral** (the red ribbon). Every **N–H** hydrogen-bonds to the **C=O four residues ahead**, locking the coil in place. The **side chains point outward**, away from the axis — so a helix can sit happily in water **or** buried in the core, depending on which R groups it carries.

The β-sheet



Here the strands stretch out nearly **flat** and lie side by side (the blue arrows point N→C), hydrogen-bonding **across** to their neighbors. Neighboring strands can run the **same way (parallel)** or in **opposite directions (antiparallel)**. The surface **pleats** like a folded fan, side chains alternating above and below.

In the body: when sheets go rogue

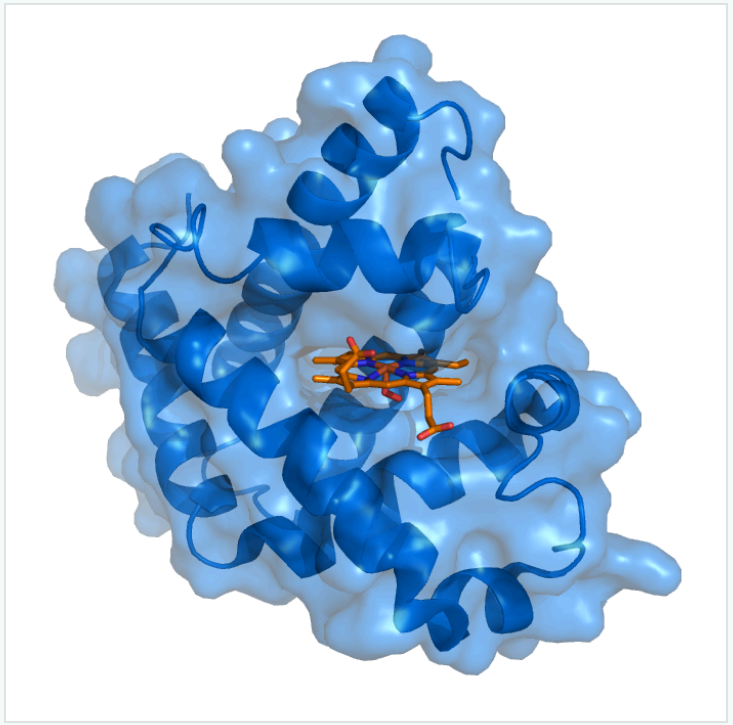


β-sheets are usually a good thing — but proteins that misfold can stack their strands into stubborn, insoluble **amyloid fibers**. In the brain those pile up into the **plaques of Alzheimer's** (arrows) and prion disease. Same H-bonds, same motif — just built where it shouldn't be.

7 · Tertiary Structure — the 3-D Fold

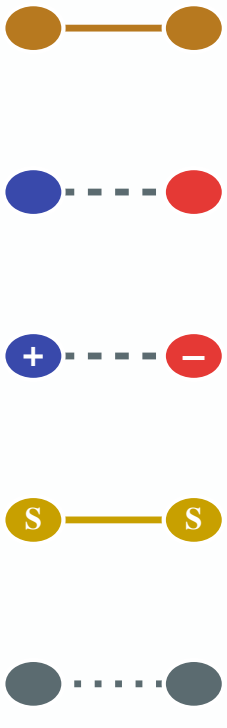
"See the full 3-D fold of a single chain, the forces that hold it, and how big proteins split into domains."

Tertiary structure — the 3-D fold



Now the **whole chain** collapses on itself — helices, sheets, and loops all packing into one compact, **specific** three-dimensional shape. This is **myoglobin**: a bundle of α-helices wrapped around a **heme** so it can grab oxygen. The fold **is** the function.

What holds the fold together



Hydrophobic core

greasy side chains huddle away from water — the main driver

Hydrogen bonds

N–H ⋯ O between side chains and backbone

Ionic (salt bridge)

a + Lys/Arg meets a – Asp/Glu

Disulfide bond

the one covalent clip: two cysteines, –S–S–

Van der Waals

weak close-packing contacts, everywhere at once

Five interactions lock the 3-D shape in place — the buried **hydrophobic core** does most of the work, and the other four hold the line.

Domains — proteins built from modules

A big protein rarely folds as one blob. It folds into **domains** — semi-independent modules, each ~100–200 residues, each with its own job (one binds DNA, another does the chemistry).

one chain → several domains → one multi-talented protein

evolution mixes and matches domains like LEGO to build new proteins

8 · Quaternary Structure · Globular & Fibrous

"Assemble several chains into one machine (quaternary), and contrast compact globular proteins with long fibrous ones."

Quaternary structure — subunits assemble

The top level: several **separate chains** lock together into one working machine. This is **hemoglobin** — **four** subunits (2 α, 2 β), each a myoglobin-like fold cradling its own **heme**. Working as a team lets them do something one chain can't: pass oxygen cooperatively.

Hemoglobin (PDB 1GZX): RCSB PDB, rcsb.org

Two lifestyles: globular vs fibrous

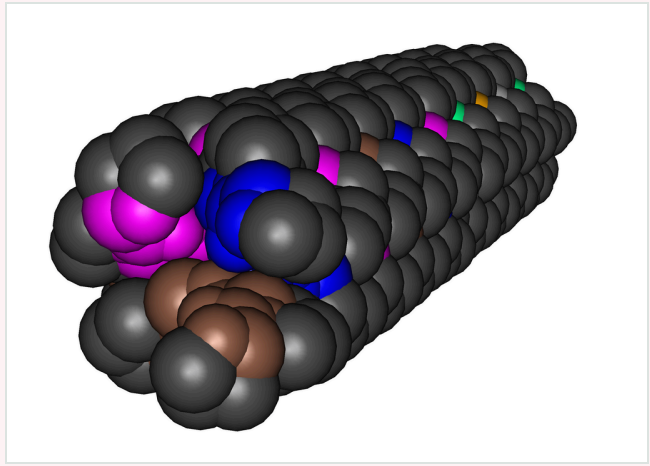
Most proteins fold into one of two shapes for two very different jobs:

- **Globular** — compact, water-soluble balls that **do the chemistry**: enzymes, hemoglobin, antibodies.
- **Fibrous** — long, repetitive, insoluble strands that **hold you together**: collagen, keratin, silk.

globular = the workers · fibrous = the scaffolding

shape follows job: a ball to react, a rope to build

In the body: collagen, the scaffold



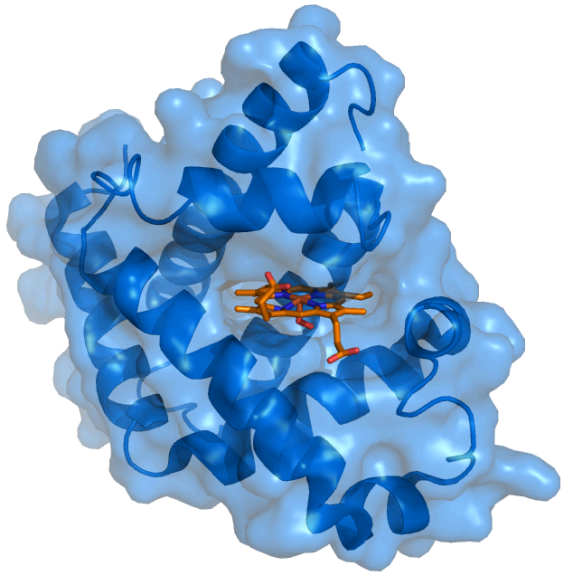
Collagen is the fibrous workhorse — about **a third of all the protein in you**, wound into a rope-like **triple helix** in skin, bone, and tendon. Building it needs **vitamin C** to cross-link the strands. Run out and the scaffold can't set — that's **scurvy**: loose teeth, bruising, wounds that won't heal.

Collagen triple helix (PDB 1K6F): Nevit Dilmen, Wikimedia Commons, CC BY-SA 3.0

9 · Myoglobin & Hemoglobin

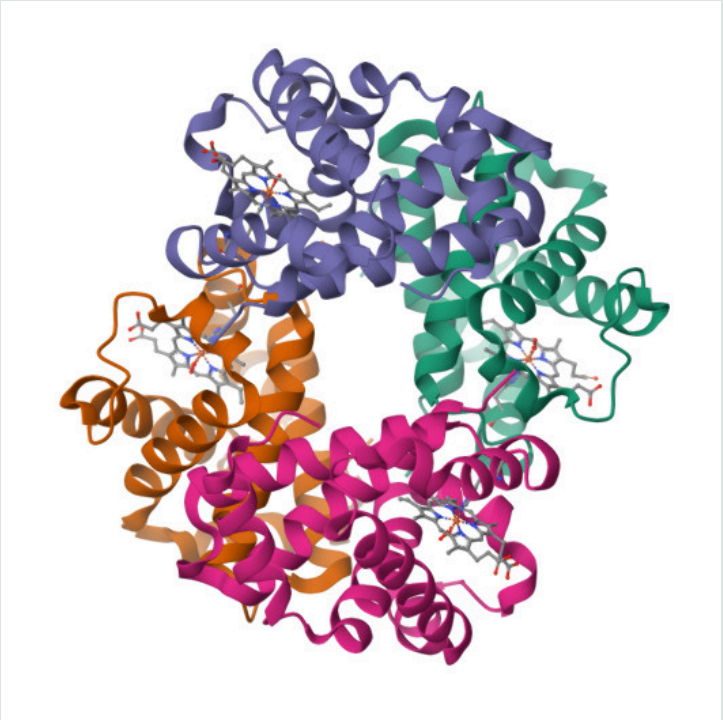
"Meet the two oxygen-carrying globins and the heme group where oxygen actually binds."

Two oxygen carriers



Your body handles oxygen with two related globins. **Myoglobin** sits in **muscle** and **stores** O_2 — one chain, one **heme**, one O_2. **Hemoglobin** rides in the **blood** and **transports** O_2 — four chains, four hemes, four O_2. Each hemoglobin subunit is basically a myoglobin.

Hemoglobin — a team of four



Four subunits (2 α, 2 β), each cradling its own **heme** — so one hemoglobin carries **four O_2** at once. Riding as a team of four is the whole trick: it lets the subunits **cooperate**, grabbing O_2 where it's plentiful and letting go where it's scarce.

Hemoglobin (PDB 1GZX): RCSB PDB, rcsb.org

Heme – where oxygen binds

The real O_2 grabber is **heme**: a flat ring (a **porphyrin**) with an **iron (Fe^{2+})** at its center. The iron holds one O_2 – **reversibly**, so it can let go again in the tissues.

$Fe^{2+} + O_2 \rightleftharpoons Fe^{2+}–O_2$ · one heme, one O_2

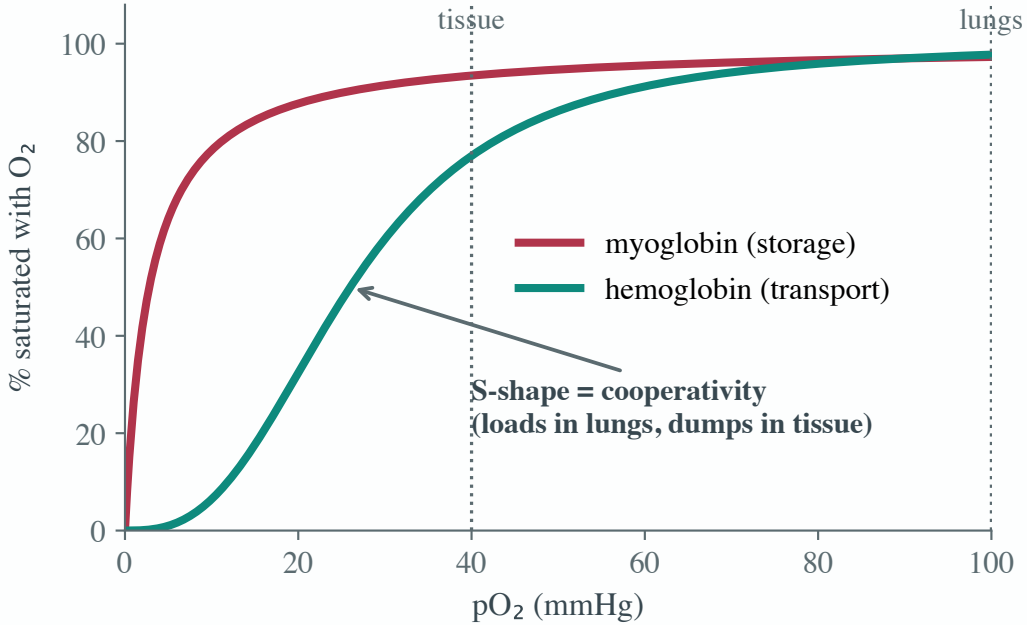
the iron must stay Fe^{2+} – oxidize it to Fe^{3+} and it can't carry oxygen

💡 Watch out: **carbon monoxide** grabs that same iron ~200× tighter and won't let go – which is exactly what makes it so deadly.

10 · Oxygen Binding & Cooperativity

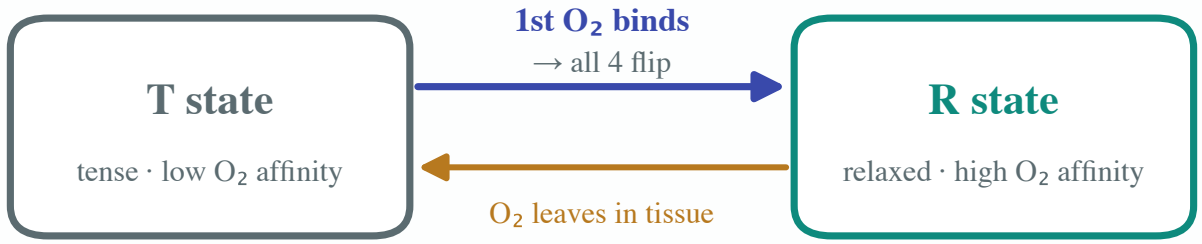
"Read the myoglobin vs hemoglobin binding curves, explain cooperativity through T→R states, and see what shifts the curve."

The oxygen-binding curves



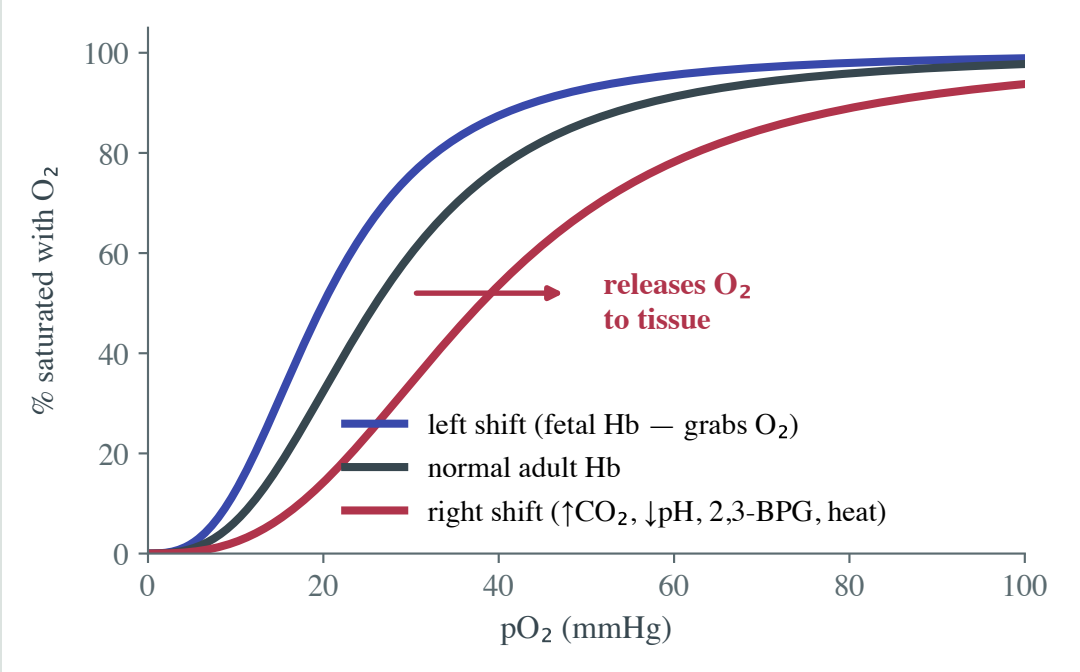
Plot how **full** each protein gets against the O_2 around it. **Myoglobin** traces a plain **hyperbola** – it grabs O_2 hard and won't let go (perfect for **storing**). **Hemoglobin** traces an **S-curve** – it loads up in the O_2-rich **lungs** and dumps its cargo in the O_2-poor **tissues**. That S-shape is **cooperativity**.

Cooperativity – teamwork



The four subunits **talk to each other**. Empty hemoglobin sits in a **tense (T)** state that binds O_2 poorly. But the moment **one** O_2 binds, all four subunits **flip to the relaxed (R)** state – and now the rest bind **easily**. One yes makes the next yes easier – that's what bends a plain curve into an **S**.

In the body: shifting the curve

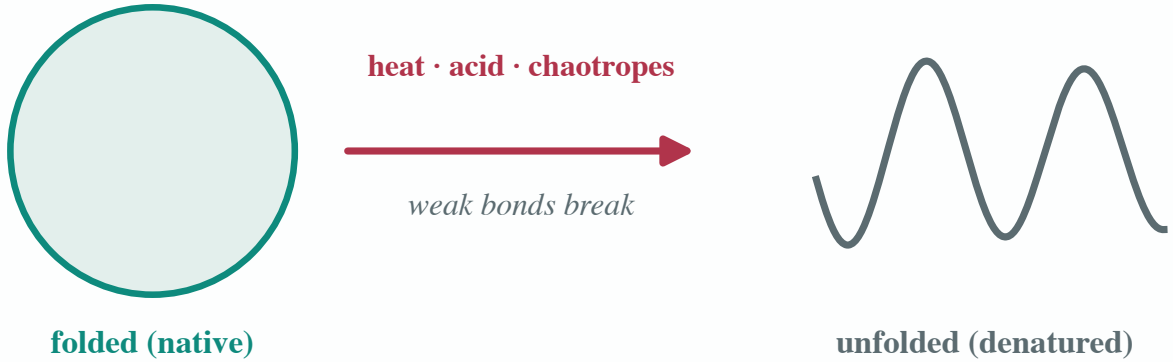


Real tissues re-tune this on the fly. Hard-working muscle is **acidic, warm, and CO_2-rich**, which **right-shifts** the curve — the **Bohr effect** — so hemoglobin **dumps more O_2** exactly where it's needed. **2,3-BPG** shifts it the same way and rises at **high altitude**. And **fetal hemoglobin** is **left-shifted** — it pulls O_2 **out of** the mother's blood, straight across the placenta.

11 · Denaturation & Misfolding

"Unfold a protein by breaking its weak bonds, and see what happens when folding goes wrong."

Denaturation — losing the shape



The fold is held by **weak** interactions, so it's fragile. **Heat**, a big swing in **pH**, or a chaotrope like **urea** snaps those bonds and the chain **unravels** — this is **denaturation**. The peptide backbone survives; the 3-D shape (and the function) is gone. That's a fried egg: heat unfolds its proteins for good.

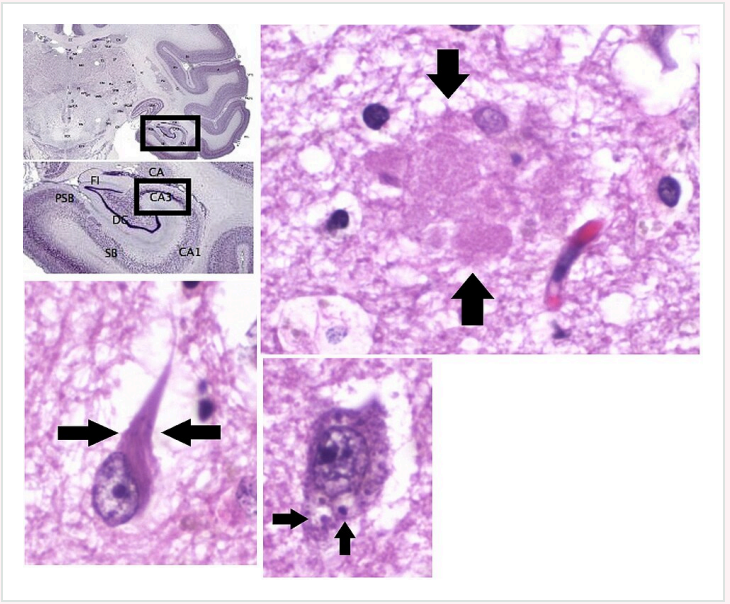
Reversible — or not

Take the stress away and some small proteins **refold on their own** — proof that the **sequence already holds all the folding instructions** (Anfinsen's classic result).

sequence → fold → function · remove the stress → sometimes it snaps back

but most proteins tangle and aggregate irreversibly — you can't un-fry an egg

In the body: when folding goes wrong



A protein that misfolds and **clumps** is more than useless — it's dangerous. Sticky aggregates build up as the **plaques of Alzheimer's** (shown) and Parkinson's. **Prions** are the eeriest: a single misfolded protein that forces its neighbors to misfold too, spreading like an infection — that's **mad cow** disease and CJD.

Can you…?

- ☐ identify and draw all 20 standard amino acids and classify each by its side chain?
- ☐ recognize the special features of glycine, proline, cysteine, histidine, and the aromatics?

If any box stays empty, the practice site has a drill for it.
🧪