workshop6: Protein Secondary Structure, Comparison and Classification

Anahita behbood

professor : Dr.Reza Rezazadegan

Protein Secondary Structure, Comparison and Classification

Chapter 8 – Comparative Structural Description of Proteins

Goals of this talk

- Connect Chapter 8 with real protein structures.
- Visually compare proteins that are all-α, all-β, and α/β.
- Contrast soluble, membrane, and fibrous proteins.
- Show how similar functions can emerge from different folds (and vice versa).
- Introduce structural classification (SCOP) and the idea of folds and superfamilies.

Quick reminder: levels of protein structure

- Primary: amino-acid sequence.
- Secondary: local regular structures – α-helices, β-sheets, turns, loops.
- Tertiary: overall 3D fold of a single polypeptide chain.
- Quaternary: assembly of several chains into a larger complex.

Secondary structure – the focus of Chapter 8

- α-helix: right-handed coil stabilized by i→i+4 hydrogen bonds.
- β-sheet: extended strands connected by hydrogen bonds between chains.
- Turns and loops: non-regular segments connecting helices and strands.
- Different combinations of these elements give rise to characteristic folds.

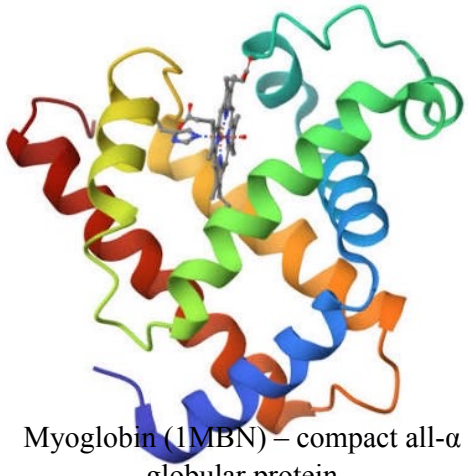
Structural classes in Chapter 8

- All-α proteins – mostly α-helices, very few β-strands.
- All-β proteins – mostly β-sheets with connecting loops.
- α/β proteins – mixed architectures (e.g. TIM barrel).
- We will now look at concrete examples of each class.

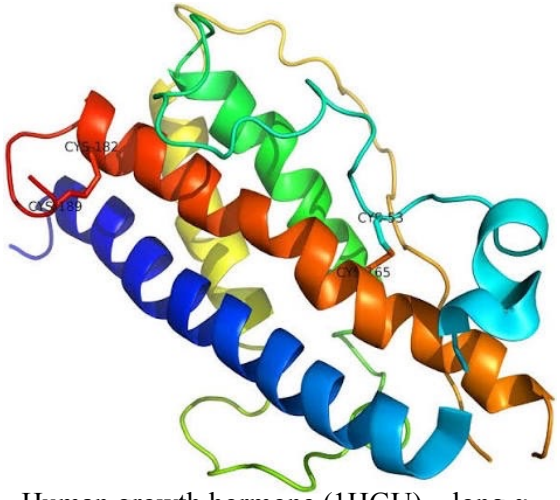
Example 1 – Two mostly α globular proteins

- Myoglobin: oxygen-storage protein in muscle; compact, mostly α-helical.
- Human Growth Hormone (hGH): secreted hormone with long α-helices.
- Both belong to the all-α class, but have different shapes and biological roles.

Example 1 – Myoglobin vs hGH



Myoglobin (1MBN) – compact all-α globular protein

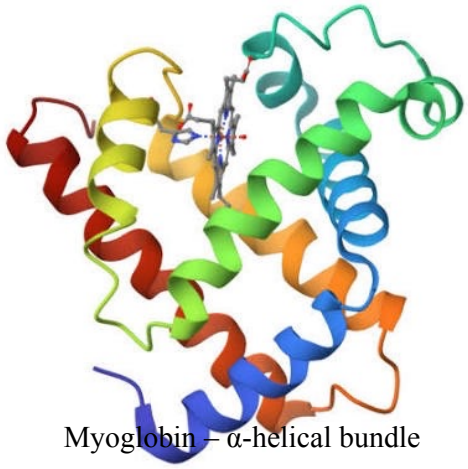


Human growth hormone (1HGU) – long α-helices, signaling protein

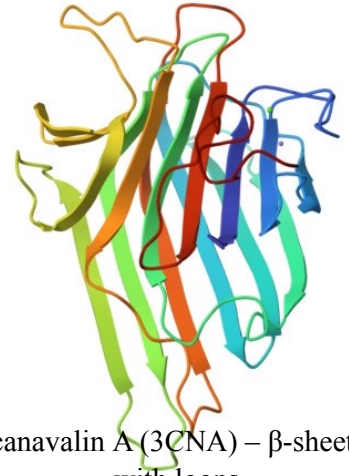
Example 2 – Mostly α vs mostly β

- Myoglobin: dense bundle of α-helices, no large β-sheet core.
- Concanavalin A (ConA): plant lectin with a large β-sheet core and surface loops.
- This pair nicely illustrates the contrast between all-α and all-β proteins.

Example 2 – Myoglobin vs Concanavalin A



Myoglobin – α-helical bundle

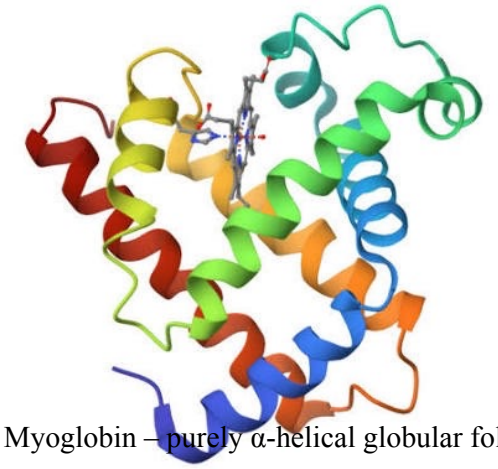


Concanavalin A (3CNA) – β-sheet core with loops

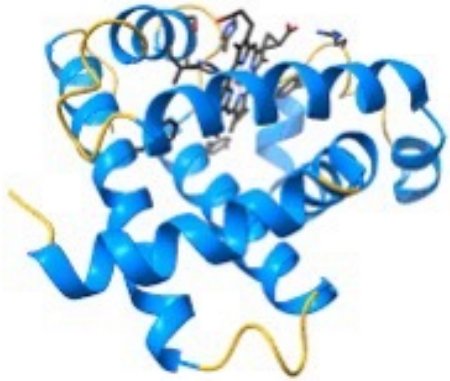
Example 3 – TIM barrel vs all-α fold

- Triosephosphate isomerase (TIM): classic TIM-barrel fold (β-α-β-α… in a ring).
- Central β-barrel surrounded by α-helices – typical α/β architecture.
- Myoglobin lacks a central β-barrel and consists almost entirely of α-helices.

Example 3 – TIM (α/β) vs Myoglobin (all-α)



Myoglobin – purely α-helical globular fold

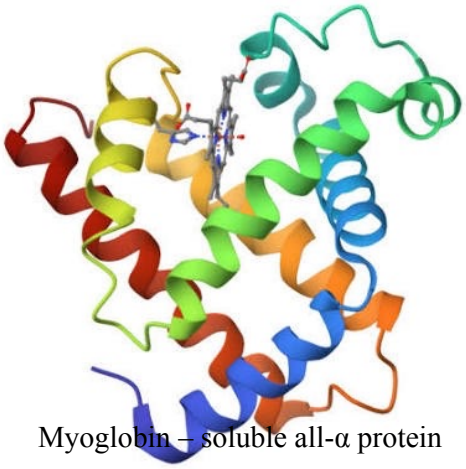


TIM (1TIM) – α/β fold with central β-barrel

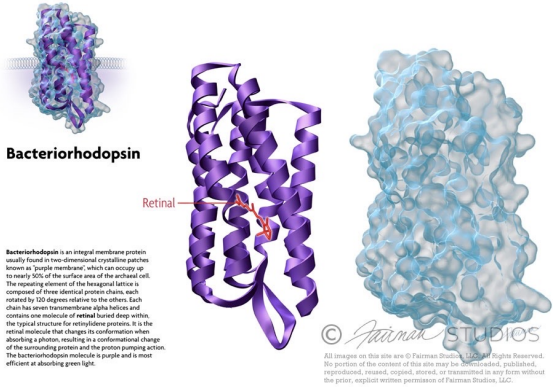
Example 4 – Soluble vs membrane all-α proteins

- Myoglobin: soluble α-helical protein in an aqueous environment.
- Bacteriorhodopsin: membrane protein with multiple transmembrane α-helices.
- Same secondary structure element (α-helix), but different surface chemistry and environment.

Example 4 – Soluble α vs membrane α



Myoglobin – soluble all-α protein

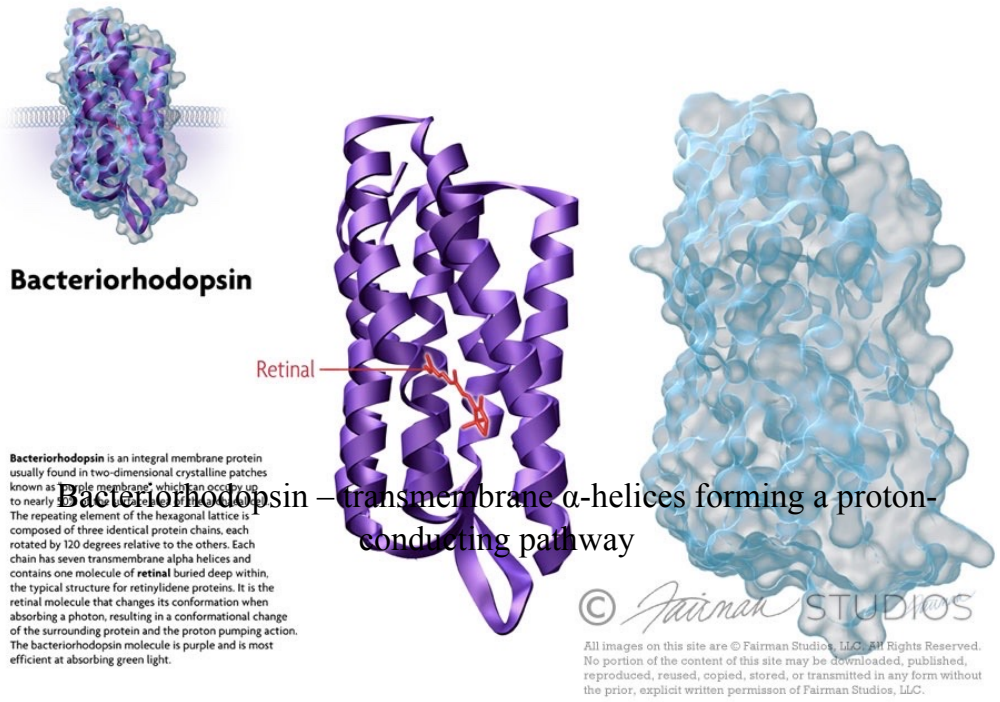


Bacteriorhodopsin – transmembrane α-helical bundle with retinal

Example 5 – Two membrane channels: α vs β-barrel

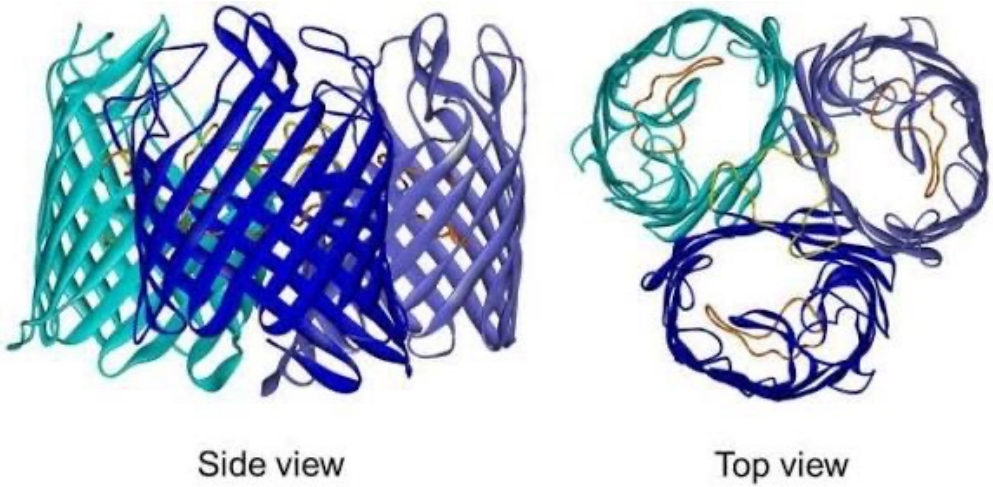
- Bacteriorhodopsin: proton pump built from α-helices in the membrane.
- OmpF porin: outer-membrane channel built from a large β-barrel.
- Both span the membrane, but their secondary structures and architectures are very different.

Example 5 – Bacteriorhodopsin as an α-helical channel



Bacteriorhodopsin – transmembrane α-helices forming a proton-conducting pathway

Example 5 – OmpF as a β-barrel channel

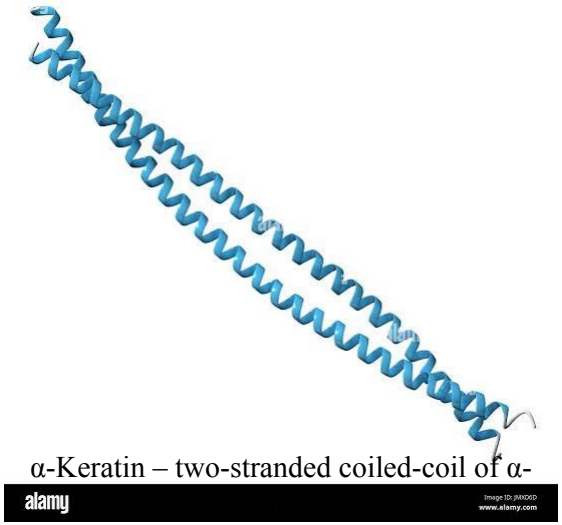


OmpF porin – β-barrel viewed from the side and from the top (central pore)

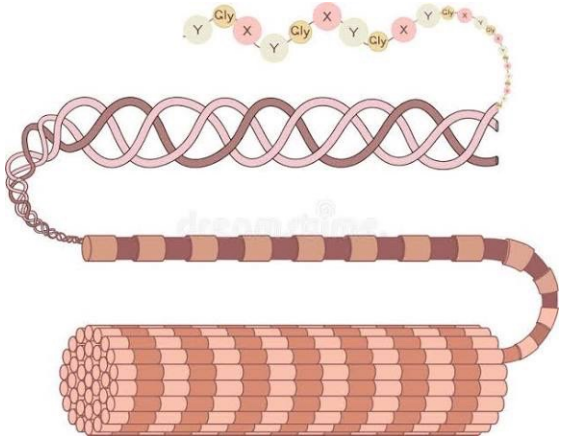
Example 6 – Fibrous proteins: α-keratin vs collagen

- α-Keratin: two α-helices form a coiled-coil; found in hair, nails, etc.
- Collagen: three chains with repeating Gly–X–Y sequence form a triple helix; pack into thick fibers.
- Both are fibrous and mechanically strong, but their packing and detailed geometry differ.

Example 6 – Keratin and collagen



α-Keratin – two-stranded coiled-coil of α-

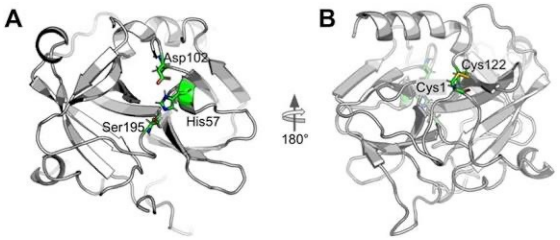


Collagen – triple helix and higher-order collagen fibril

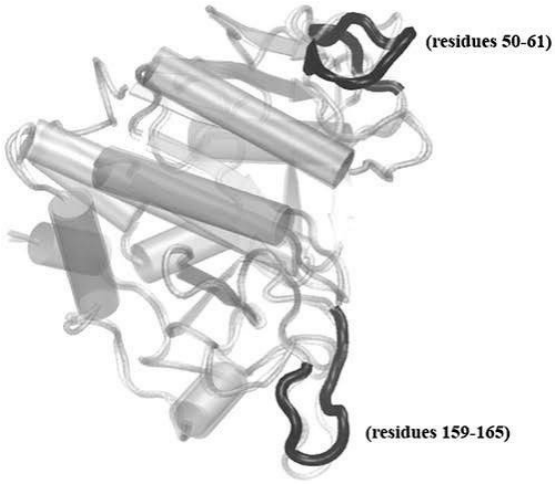
Example 7 – Convergent evolution in serine proteases

- Chymotrypsin: serine protease with the classic trypsin/chymotrypsin fold.
- Subtilisin: serine protease with a completely different fold.
- Both use the same catalytic triad (Ser–His–Asp), but evolved independently – convergent evolution.

Example 7 – Two serine protease folds



Chymotrypsin – trypsin-like serine protease fold

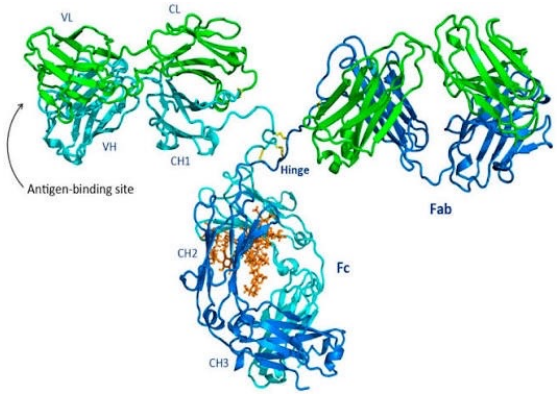


Subtilisin – unrelated fold with the same catalytic triad

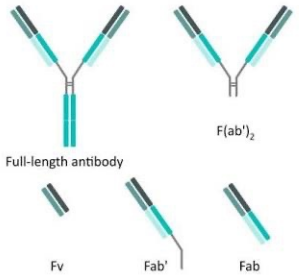
Example 8 – Antibody architecture and domains

- IgG antibodies have a Y-shaped architecture with two heavy and two light chains.
- Repeated β-sandwich domains form the constant and variable regions.
- The variable domains at the tips of the Y contain the antigen-binding site.

Example 8 – IgG and derived fragments

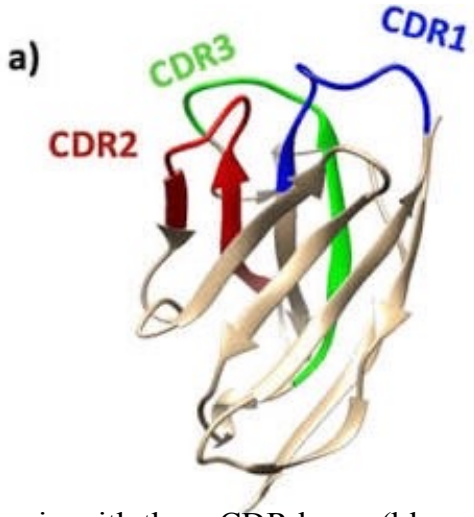


IgG molecule – Fab arms and Fc stem

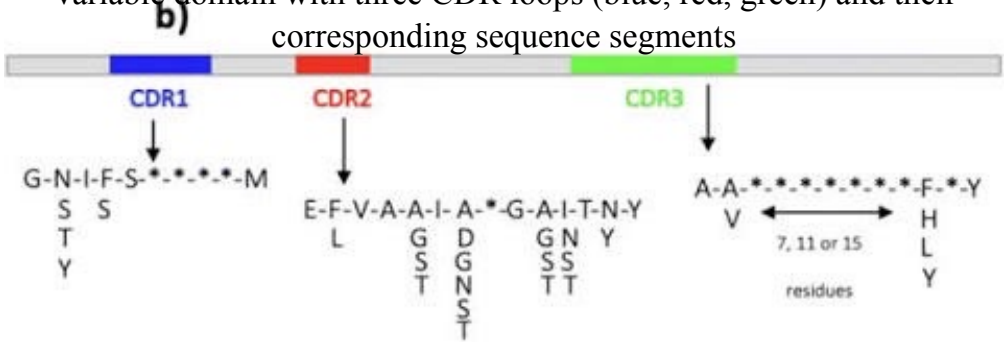


Different antibody formats – full IgG, Fab, $F(ab')_2$, Fv, etc.

Example 8 – CDR loops and diversity



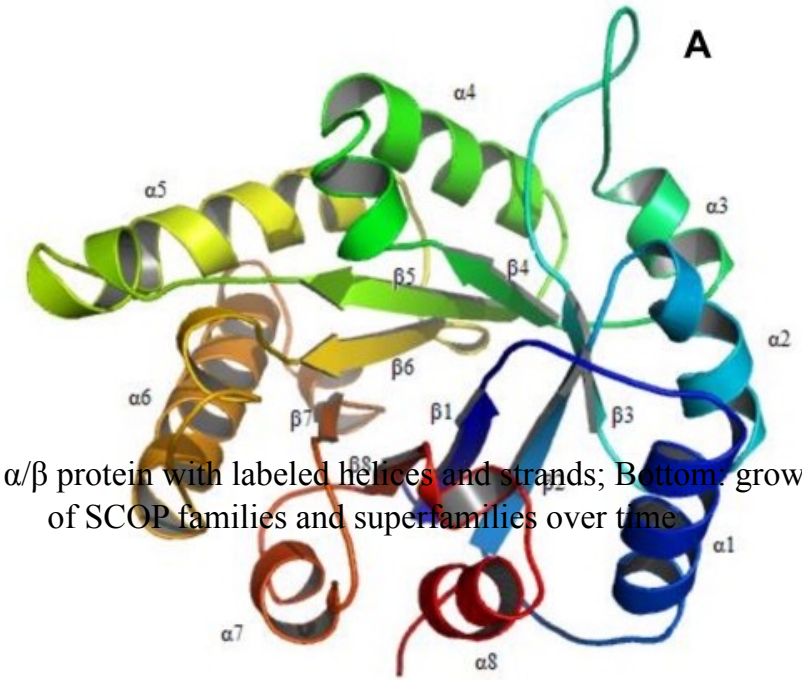
Variable domain with three CDR loops (blue, red, green) and their corresponding sequence segments



Example 9 – Structural classification (SCOP)

- SCOP organizes protein structures into classes, folds, superfamilies, and families.
- Classes: all-α, all-β, α/β, α+β, etc.
- Folds group proteins with the same major secondary-structure arrangement and topology.
- Superfamilies and families track evolutionary relationships and sequence similarity.

Example 9 – SCOP example



Top: an α/β protein with labeled helices and strands; Bottom: growth of SCOP families and superfamilies over time

Summary – Structural comparison of proteins

- Secondary structure (α, β, loops) combines into recognizable architectures and folds.
- All-α, all-β, and α/β proteins can be readily distinguished visually.
- Soluble, membrane, and fibrous proteins use different secondary-structure patterns.

Summary – Function, evolution, and classification

- The same fold can support many different functions (e.g. TIM barrel).
- The same function can arise from different folds (convergent evolution in serine proteases).
- Classification schemes like SCOP help us map the "universe of folds".
- Chapter 8 provides the conceptual basis; real structures make these ideas concrete.

Proteins as data objects

- Sequence: a 1D string over a 20-letter alphabet (amino acids).
- Structure: a 3D set of atomic coordinates for each atom in the protein.
- Files: stored as PDB/mmCIF text files with tables of atoms, residues and chains.
- From a CS perspective: we can parse, store, index, search and visualize these data.
- ML view: learn mappings sequence → structure → function.

Where do structures come from?

- Experimental methods: X-ray crystallography, NMR spectroscopy, cryo-EM.
- Structures are deposited in the Protein Data Bank (PDB).
- PDB currently contains on the order of ~240,000 experimentally determined structures.
- Predicted structures: AlphaFold Protein Structure Database with >200 million models.
- Together, these are huge datasets for computational and ML approaches.

Online tools for structural exploration

- RCSB PDB (rcsb.org) – main portal for PDB entries with Mol* 3D viewer.
- PDBe (pdbe.org) – European PDB portal with Molstar visualizations and analysis tools.
- AlphaFold DB (alphafold.ebi.ac.uk) – predicted structures with per-residue confidence (pLDDT).
- UniProt (uniprot.org) – sequence and functional annotation, with links to structure viewers.
- In class, you can open any of these and rotate the structures live in the browser.

Live demo plan (web tools)

- Step 1: Open rcsb.org and load 1MBN (myoglobin) – view in 3D (Mol*).
- Step 2: Color by secondary structure to highlight α-helices and β-sheets.
- Step 3: Load 1HGU (hGH), 1TIM (TIM barrel), and a bacteriorhodopsin and OmpF structure.
- Step 4: Show how to switch between models and compare different folds interactively.
- Step 5: Optionally, show an AlphaFold DB entry and its confidence coloring.

Representing 3D structure for algorithms

- Atomic coordinates can be stored as an (N atoms × 3) matrix of (x, y, z).
- Residue-level graphs: nodes = residues, edges = spatial proximity (distance below a cutoff).
- Node features: residue type, secondary structure, solvent exposure, conservation, etc.
- Edge features: inter-residue distance, hydrogen bonds, contacts, interaction type.
- These representations are used in structure prediction, docking and protein–protein interaction studies.

Protein structure problems for CS/ML

- Structure prediction: mapping sequence → 3D (e.g. AlphaFold, RoseTTAFold).
- Structure alignment and search: compare folds and detect structural neighbors.
- Classification: assign new structures to existing folds and superfamilies (e.g. SCOP, CATH).
- Learning on structures: graph neural networks and transformers over atoms/residues.
- Protein structure is a rich benchmark for geometric and graph-based deep learning.

How to explore structures yourself

- Pick a protein in UniProt and open the “Structure” section to find links to PDBe/RCSB.
- Open the corresponding PDB entry in RCSB PDB and launch the Mol* 3D viewer.
- Play with coloring options: by chain, by secondary structure, by residue type or by domain.
- Try to classify each protein visually as all-α, all-β, α/β, membrane or fibrous.
- This hands-on exploration makes the ideas from Chapter 8 much easier to remember.