Protein structure

Four levels of protein structure
1. Primary structure
2. Secondary structure
3. Tertiary structure
4. Quaternary structure

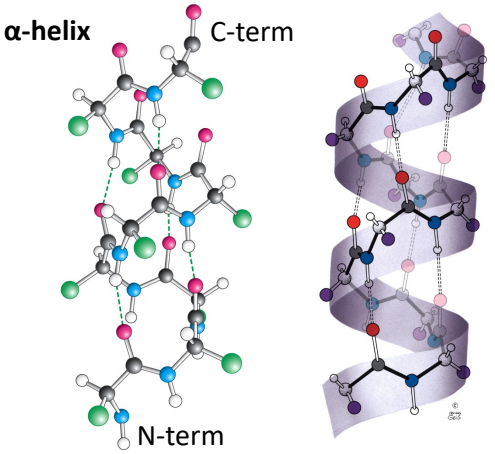
1. Primary structure

Protein structure

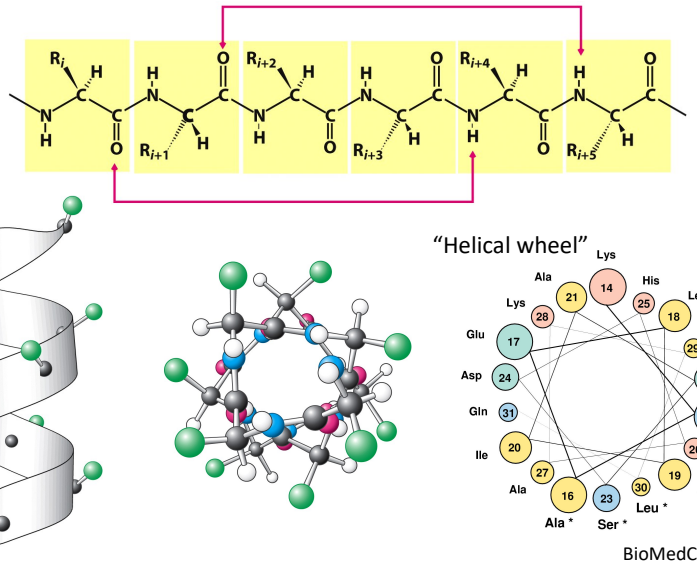
2. Secondary structure

"Local structure", H-bonds *within the protein backbone*

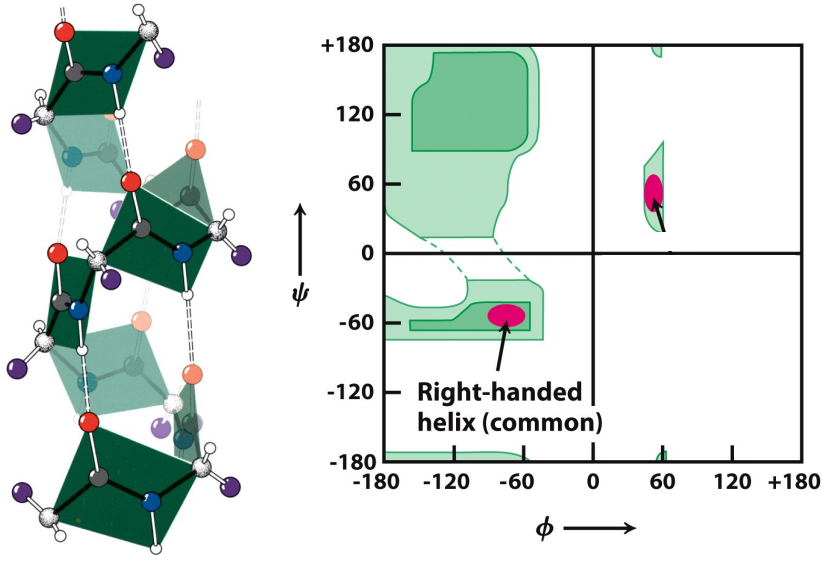
- Two main types: **α-helix** and **β-sheet**



Protein structure



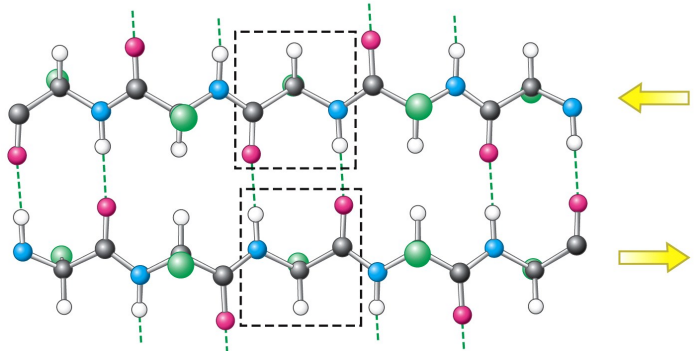
Ramachandran plot



Protein structure

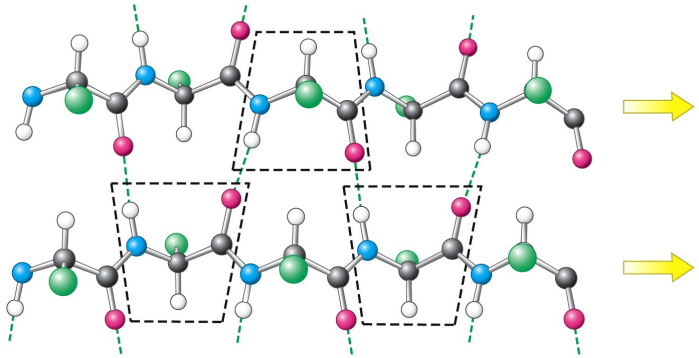
β-sheet
Composed of multiple **β-strands** that are parallel or antiparallel

Antiparallel β-sheet
- H-bonds between strands are linear
- Side chains alternating above and below the plane of the sheet



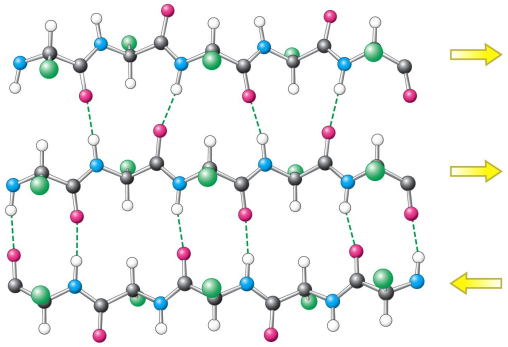
Protein structure

Parallel β-sheet
- H-bonds between strands are NOT exactly linear
- Side chains alternating above and below the plane of the sheet

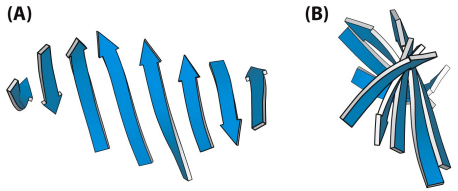


Protein structure

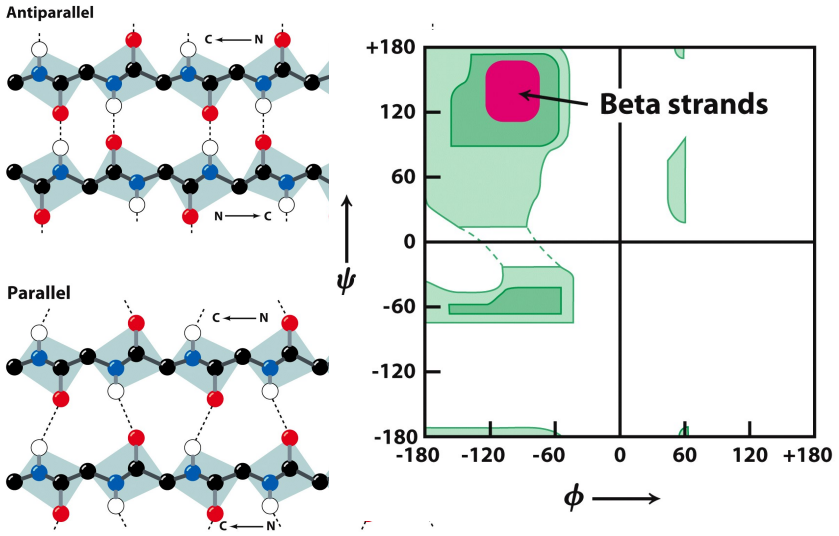
Mixed β-sheet



Twist in β-sheet

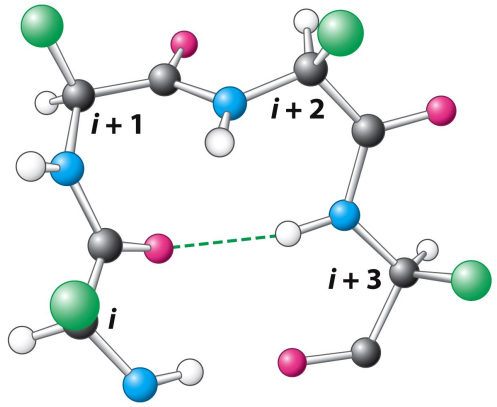


Ramachandran plot



Protein structure

β-turns



- H-bond between *i* and *i* + 3
- Glycine and proline are common. **Why?**

Protein structure

Amino acids have varying tendencies to form secondary structures

<u>In general (these are not firm rules!):</u>

- **α-helices**
  - Common: Many/most!
  - Slightly less common: branched β-carbon (V, T, I), small side chain H-bonding (S, D, N)
  - Rare: Proline! ***Why??***
- **β-strands**
  - Common: Many/most
  - Rare: Proline!
- **β-turns**
  - Common: G, N, P, others

Protein structure

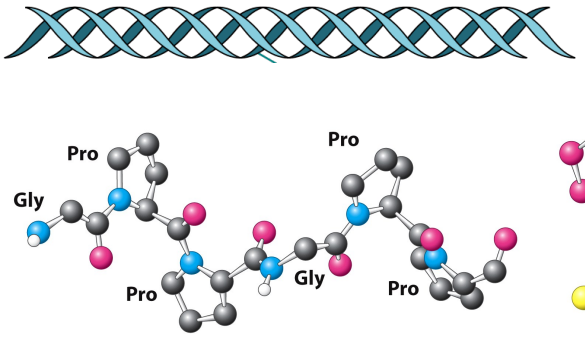
Table 2.3 Relative frequencies of amino acid residues in secondary structures

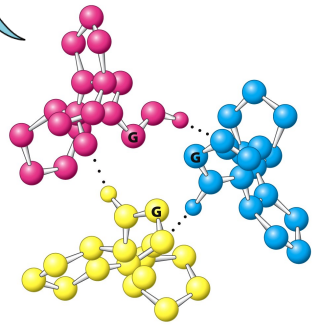
Amino acid	α helix	β sheet	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.96
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (middle group), or turns (bottom group).
Source: T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.

Collagen

Collagen, a triple helix
The most abundant protein in mammals, structural role
Fibrous component of skin, bone, tendon, cartilage, teeth, etc.
Glycine every third residue, proline (and hydroxyproline) are frequent





Collagen

Osteogenesis imperfecta (OI, "brittle bone disease")
- Improper collagen folding and structure
- One common cause: mutation of Gly to a bulkier residue

Affects 1 in 20,000 live births

Symptoms
- Severe bone fragility
- Short stature
- Respiratory problems
- Blue sclera (white parts of eyes)

Treatment
- Bisphosphonates: increase bone mass by killing osteoclasts
- Surgery to strengthen bones, physiotherapy, physical aids

Protein structure

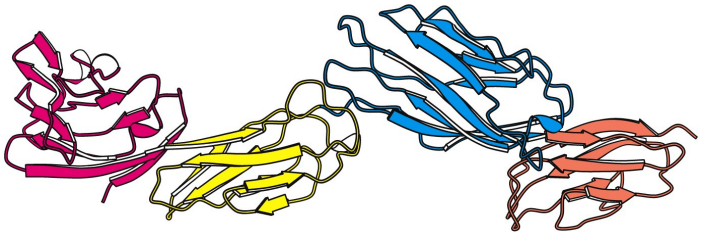
3. Tertiary structure
Overall, three-dimensional fold
Intermolecular forces, often involving side chains

Protein structure

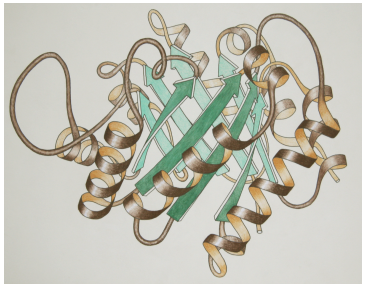
Protein domains

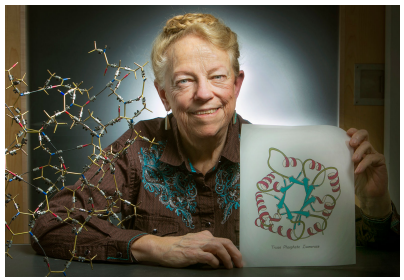
- Many protein regions fold into discrete, compact "domains"
- Each folds independently, has its own tertiary structure
- Can be strung together like beads on a string

Example: CD4, recognized by HIV

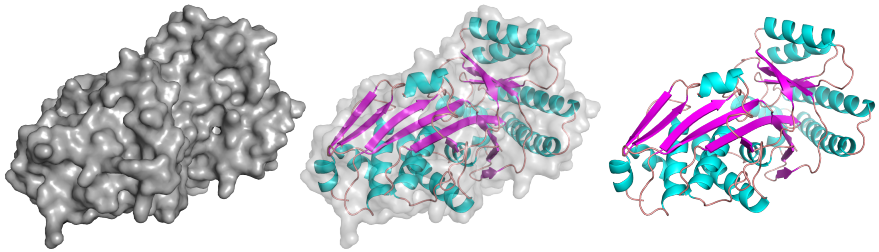


Protein structure - Ribbon diagrams





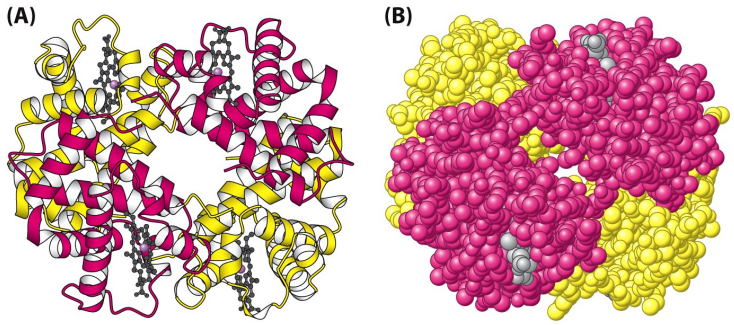
Dr. Jane Richardson (Duke University)



Protein structure

4. Quaternary structure
3D arrangement of *multiple* interacting polypeptide chains
Each chain is called a "subunit" (α, β, γ, …)

Example: hemoglobin ($\alpha_2\beta_2$)



Sequence dictates structure

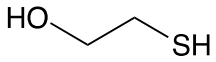
Primary structure (sequence) dictates the protein structure

- Theoretically possible to unfold protein, and it can re-fold!
  -

How to unfold?

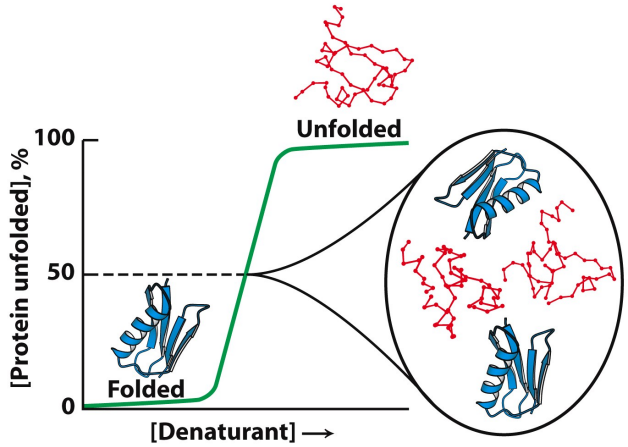
- Treat with 8 M urea or 6 M guanidinium hydrochloride
  - Disrupt tertiary structure

- Add β-mercaptoethanol (BME), a reducing agent



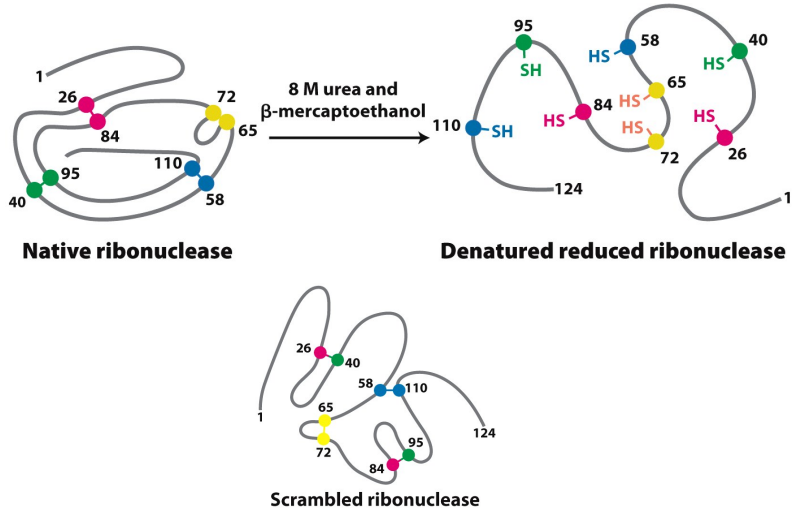
Sequence dictates structure

Sharp transition between folding and refolding



Sequence dictates structure

Christian Anfinsen, 1961: **Anfinsen's dogma**



Sequence dictates structure

Christian Anfinsen, 1961: **Anfinsen's dogma**

A protein's structure is determined only by its sequence!

So… Can we predict protein structure based on sequence?
Until 2021: Not very well… ***2021:* AlphaFold2 (DeepMind)**

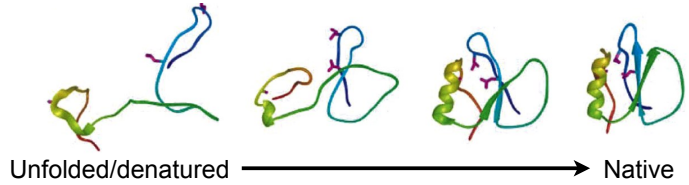
Challenges predicting structure:

- Amino acids don't greatly prefer any one structure
- Same/similar sequences in one protein may have a different structure in another protein
- Long-range tertiary interactions difficult to predict
    - They can be important to secondary structure too!

Protein folding

Of the seemingly infinite possibilities a protein could fold into, usually ***only one*** is observed! The "native" state: most IMFs, most stable.

How do proteins fold? How do they "know" how to fold correctly?



Folding is not random…

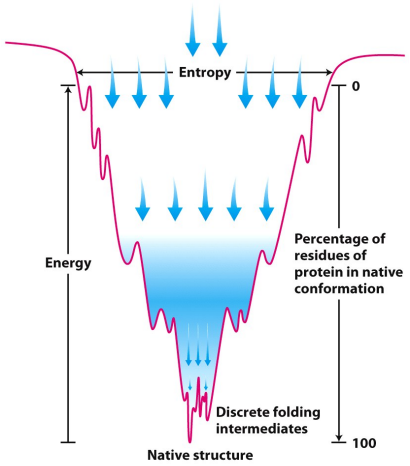
- Partially correct structures are retained
- *Cooperativity*: local regions adopt energetically favored structure, inducing other regions to fold, steadily increasing stabilization

Protein folding

- Folding does not necessarily follow one specific pathway
- Multiple paths lead to the same energy minimum

Folding funnel

1.



Protein folding

Chaperones
Proteins that "assist" in folding/unfolding

- Some act ***while*** protein is made by the ribosome
    -
- Other act ***after*** the whole protein has been made
    -

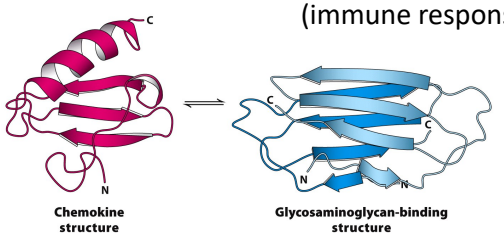
How? Different strategies…

- "Shield" regions to prevent "incorrect" interactions
- Prevent aggregation
- Some: sequester protein
- Some: ATP dependent, use energy input to prevent getting stuck in local energy minima

Exceptions to Anfinsen's dogma!

Sometimes sequence *doesn't* specify one specific structure!

- Alternate structures for the same protein
  - **Metamorphic proteins,** e.g. lymphotactin, a cytokine (immune response)



- Estimated: ~50% of eukaryotic proteins have at least one unstructured region of >30 amino acids!
  - Few hydrophobic, many charged/polar amino acids
- Intrinsically disordered proteins

Protein misfolding diseases

Examples:
Alzheimer disease
Parkinson disease
Huntington disease
Bovine spongiform encephalopathy (BSE, "mad cow disease")
- Analogous diseases in other organisms

Protein aggregation: **amyloids** or **plaques**

Prion
Disease causing agent in BSE and related diseases
A misfolded protein, PrP (native function is unknown)

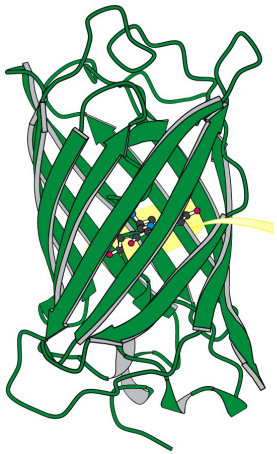
Posttranslational modification (PTM)

Extend the chemical repertoire of proteins, beyond 20 amino acids

Deleted groups
- Hydrolysis: proteolysis (proteases)

Added groups
- Acetyl groups (acetylation)
- Carboxylate groups (carboxylation)
- Phosphate groups (phosphorylation)
- Sugars (glycosylation)
- Lipids (lipidation)

Rearrangements
- e.g. green fluorescent protein (GFP)



Protein structure determination

Some common methods used to study protein structure

-
-
-
-