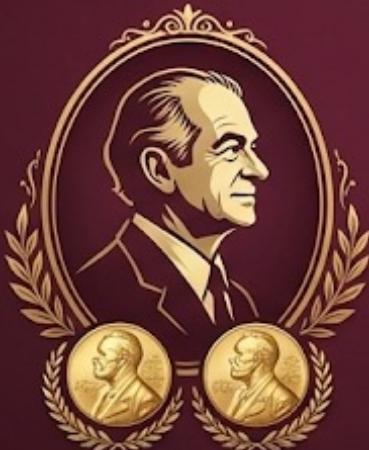


A Double Nobel Lecture



Linus Pauling
(1901-1994)

Nobel Prize in Chemistry, 1954
Nature of the chemical bond



Nobel Peace Prize, 1962
Nuclear test ban advocacy



Key Contributions:
 α -helix, β -sheet,
electronegativity, hybridization

“ The best way to have a good idea is to have lots of ideas. ”



Frederick Sanger
(1918-2013)

Nobel Prize in Chemistry, 1958
Structure of proteins (insulin)



Nobel Prize in Chemistry, 1980
DNA sequencing methods



Key Contributions:
Protein sequencing, Sanger sequencing, dideoxy method

“ I was lucky twice. ”

Only 5 people have won two Nobel Prizes
Today's lecture: Their foundational contributions to protein structure

Chapter 4

Protein Three-Dimensional Structure

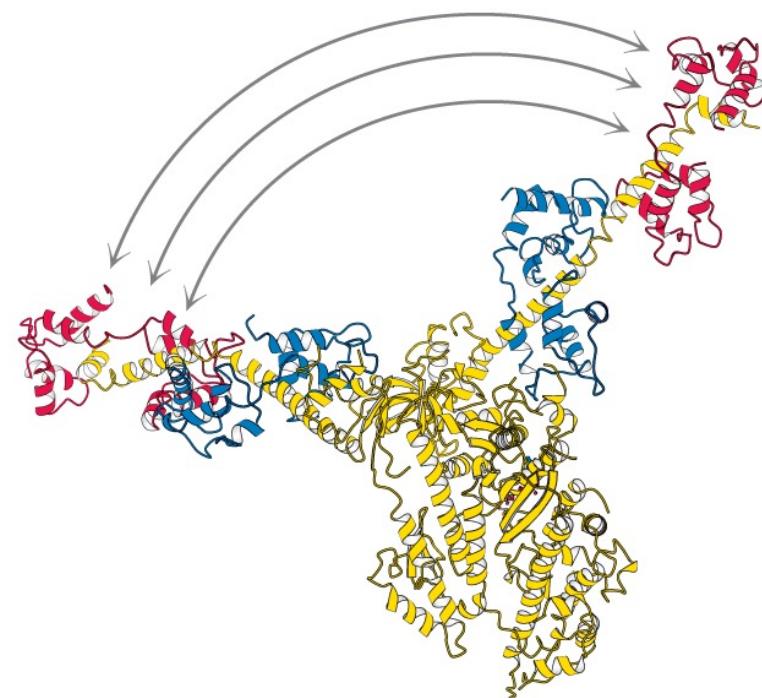


Photo: Royal Collection Trust/© His Majesty King Charles III, 2023/Bridgeman Images
Copyright © Macmillan Learning

By the end of this chapter, you should be able to:

- ✓ Distinguish between primary, secondary, tertiary, and quaternary structures of proteins.
- ✓ Describe the properties of the principal types of secondary structure, including the α helix, the β sheet, and the reverse turn.
- ✓ Describe the biochemical information that determines the final three-dimensional structure of a protein.

Section 4.1 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

Learning objective 2: Compare and contrast the different levels of protein structure and how they relate to one another.



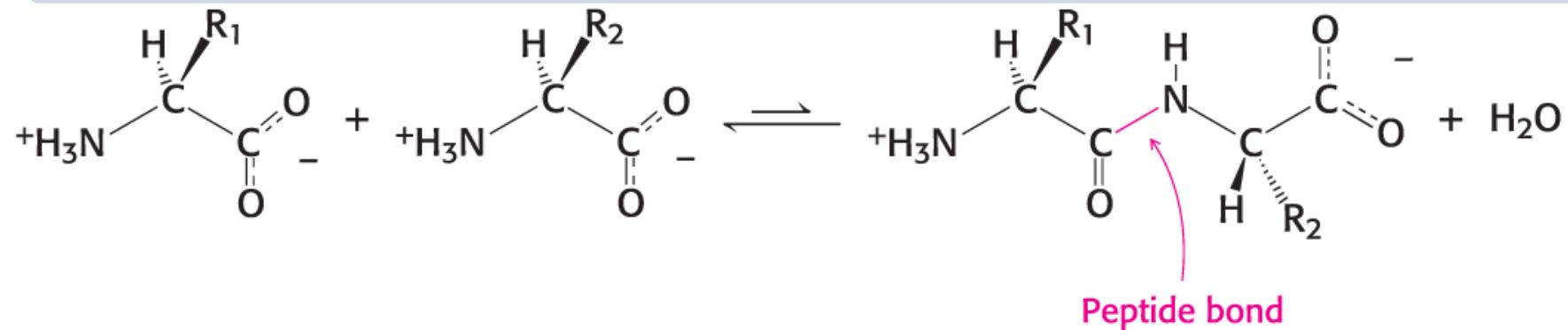
Polypeptides consist of amino acids linked by a peptide bond (also called an amide bond).



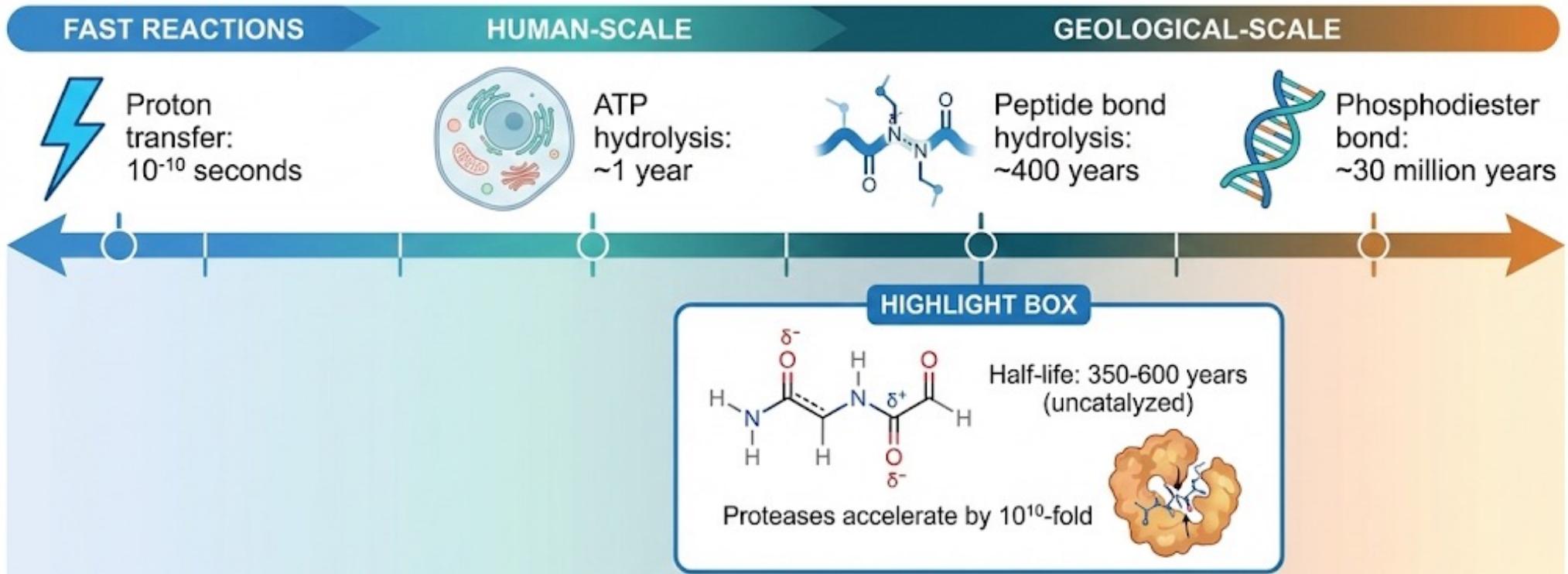
Each amino acid in a protein is called a residue.



Requires energy to make a peptide bond



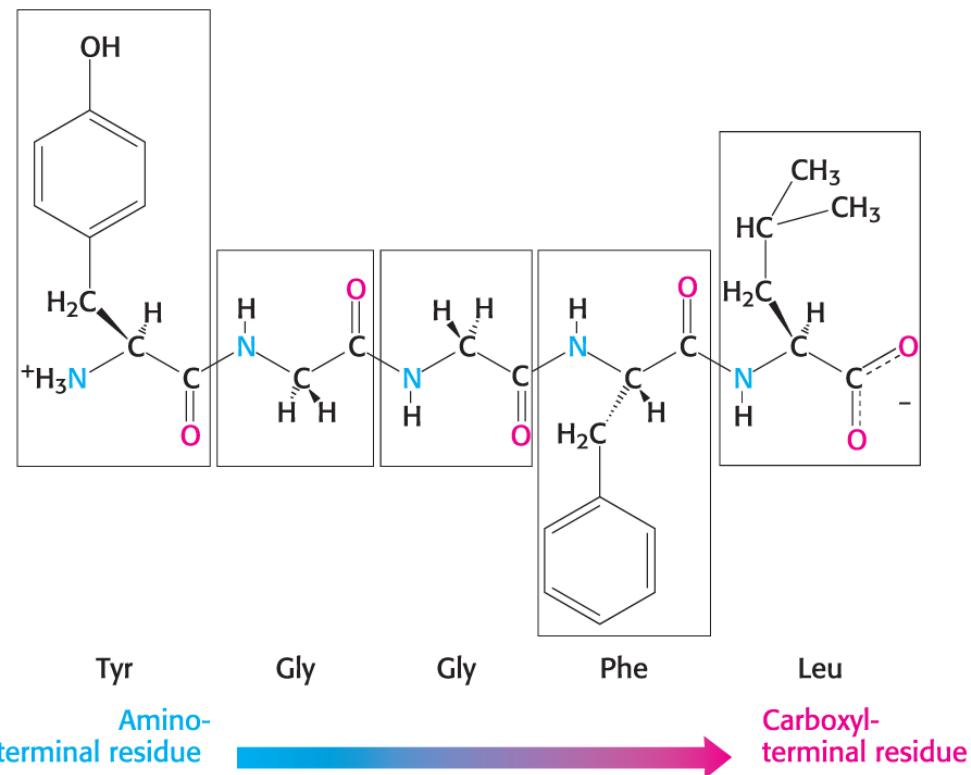
Peptide Bond Stability



Peptide bonds are stable for centuries. So how does your body break down the proteins you eat in minutes?

Polypeptide Chains Have Directionality

- Series of amino acids is a polypeptide
- A polypeptide chain has directionality. The amino terminal end is taken as the beginning of the polypeptide chain.
- The carboxyl terminal end is the end of the polypeptide chain.
- The **primary structure** is always written from the amino terminal to the carboxyl terminal, or left to right.

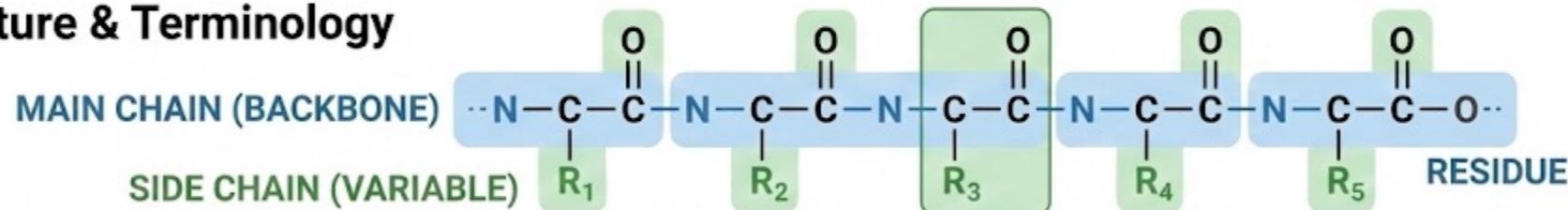


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

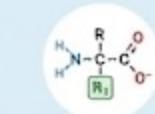
Polypeptide Chains: Structure, Size, and Terminology

From Amino Acids to Proteins

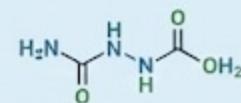
Structure & Terminology



Size & Classification



Amino Acid Residue



OLIGOPEPTIDES
(or peptides)
Small numbers of amino acids

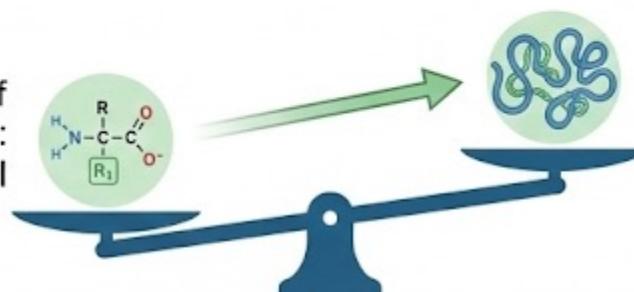


PROTEINS
(Polypeptide Chains)
Typically 50 to 2000 residues

TITIN:
Largest known protein,
~27,000 amino acids

Molecular Weight

Mean Molecular Weight of an Amino Acid Residue:
~110 g/mol



Molecular Weight of Most Proteins:
5500 to 220,000 g/mol

DALTON (Da)

1 Da \approx 1 g/mol \approx Mass of 1 Hydrogen Atom

If the average residue mass is 110 Da, how many amino acids are in a 55 kDa protein?

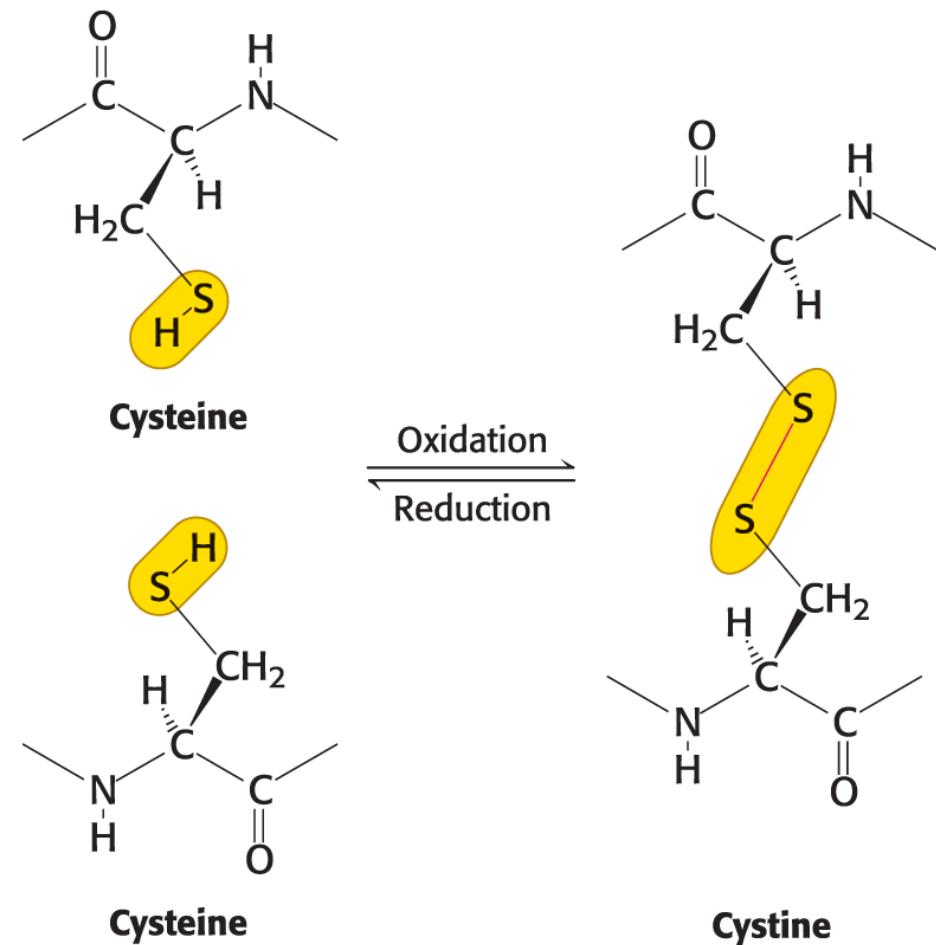
Cross-Linking

In some proteins, the linear polypeptide chain is covalently cross-linked.

The most common cross-links are **disulfide bonds**.

- They are formed by the oxidation of a pair of cysteine residues.
- They can form between cysteine residues in the same polypeptide chain, or they can link two separate chains together.
- The resulting unit of two linked cysteines is called **cystine**.

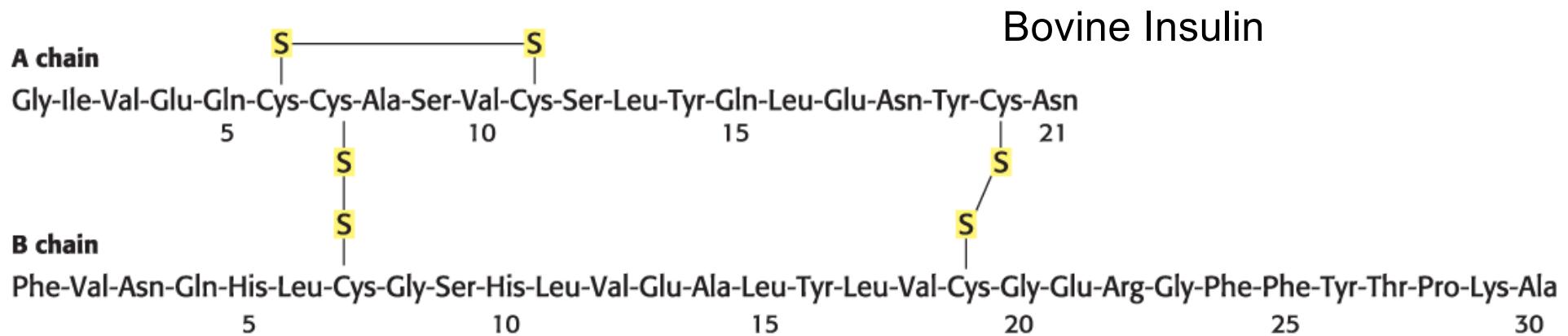
Rarely, nondisulfide cross-links derived from other side chains are present in proteins.



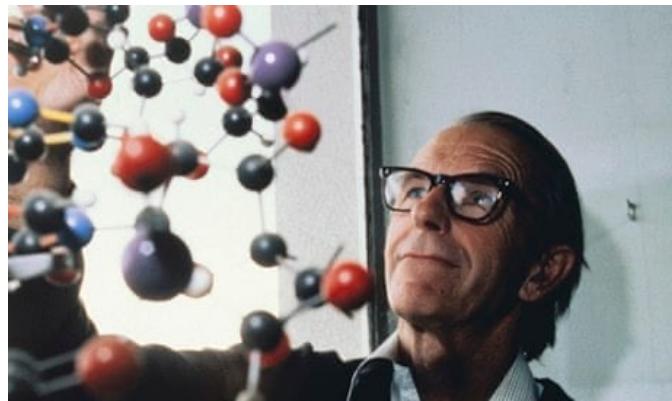
Are disulfide bonds typically found inside the cell or outside?

Proteins Have Unique Amino Acid Sequences Specified by Genes

The complete amino acid sequences of millions of proteins are now known.



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company



Fredrick Sanger

"I was lucky twice."

Why do we care about knowing amino acid sequences

1. The sequence dictates three dimensional structure
2. This structure dictates function
3. Alterations in aa sequence underly many many diseases: CF, Sickle cell, etc.
4. Sequence of protein reveals evolutionary history.
 1. There is resemblance between same protein across different species
 2. The 3d structure and function is more highly conserved than the aa sequence

Protein mass: residues → daltons → kilodaltons



Residues

"Mass \approx (# residues) \times 110 Da"

- Avg residue mass \approx 110 Da (good rule of thumb)



Daltons & Kilodaltons

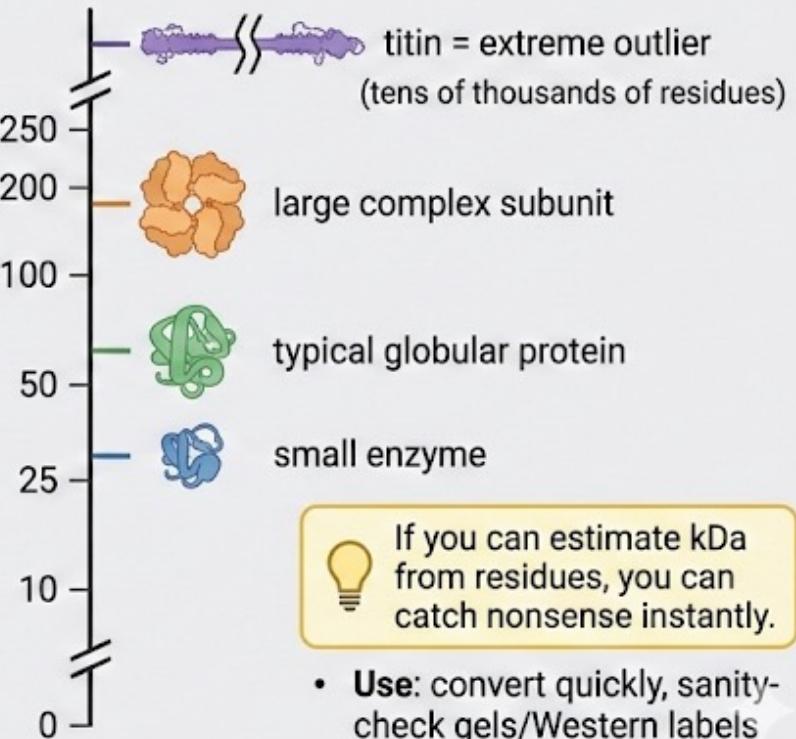
"1 kDa = 1000 Da"

- 1 Da \approx 1 g/mol
- kDa is the common unit for proteins



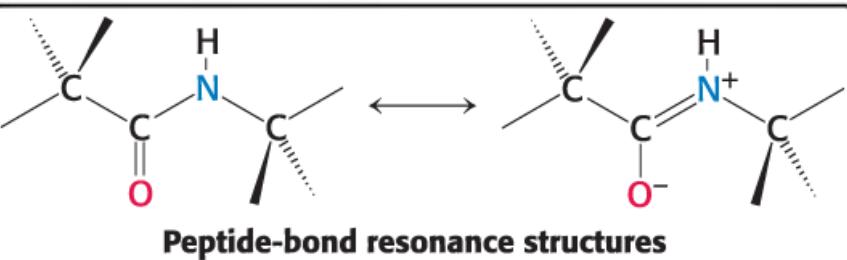
Protein Size Scale

Typical proteins: 5–200 kDa (rough range)



How do we figure out the relationship between amino acid sequence and 3d structure?

The first thing we need to know is something about rotation around the bonds.



Peptide bond is planar and rigid

1) Planar peptide unit

Six atoms lie in one plane:

- C_{α} (residue 1)
- N and H (amide)
- C and O (carbonyl)
- C_{α} (residue 2)

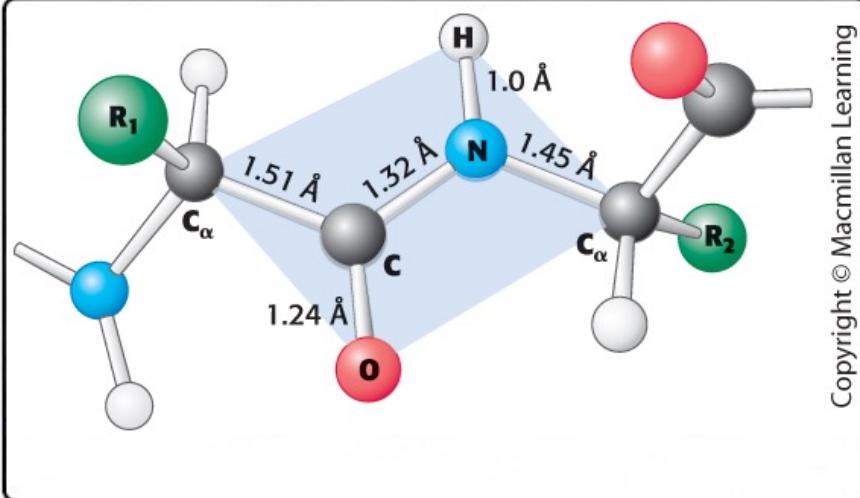
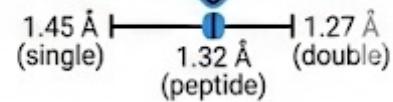
2) Resonance gives partial double-bond character

- Electrons are delocalized between $C-N$ and $C=O$
- $C-N$ behaves partly like a double bond

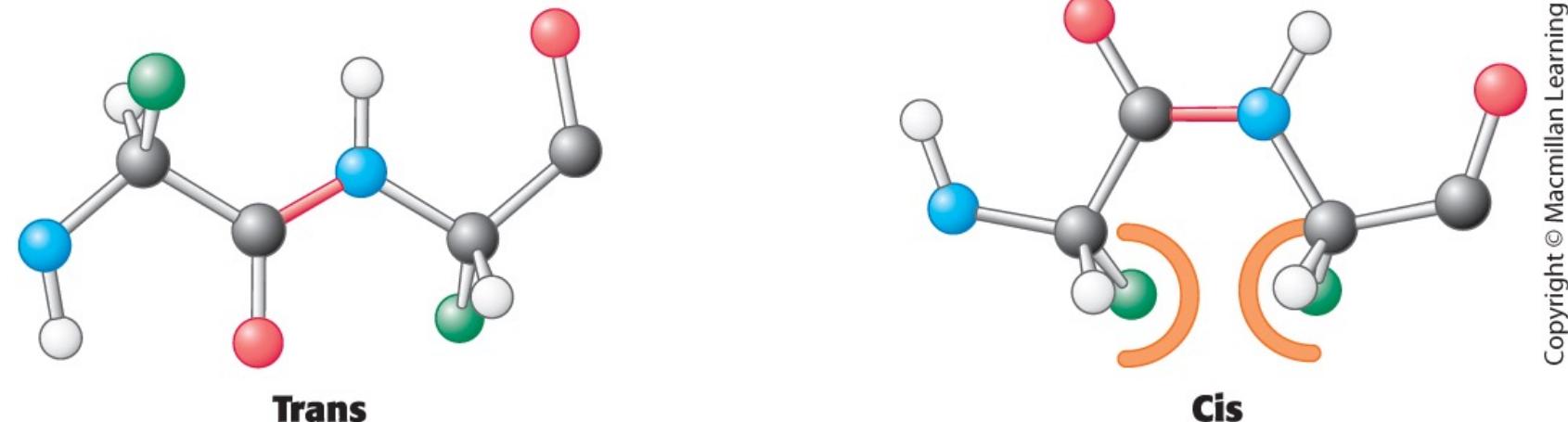
3) Consequences

- Rotation about $C-N$ is restricted
- Backbone conformation is constrained

- Peptide $C-N \approx 1.32 \text{ \AA}$
- Between single (1.45 \AA) and double (1.27 \AA)



Two Configurations of the Peptide Bond



Copyright © Macmillan Learning

In the *trans configuration*, the two α -carbon atoms are on opposite sides of the peptide bond.

In the *cis configuration*, the two α -carbon atoms are on the same side of the peptide bond.

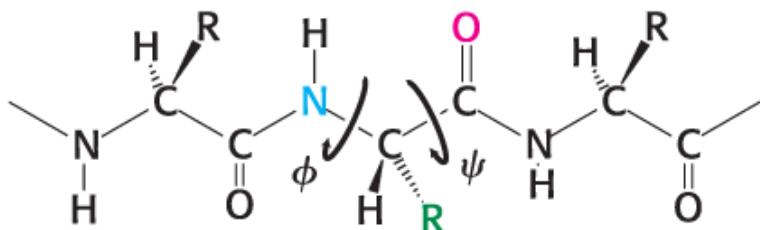
Almost all peptide bonds in proteins are trans.

- This preference can be explained by the fact that there are steric clashes between R groups in the cis configuration but not in the trans configuration.

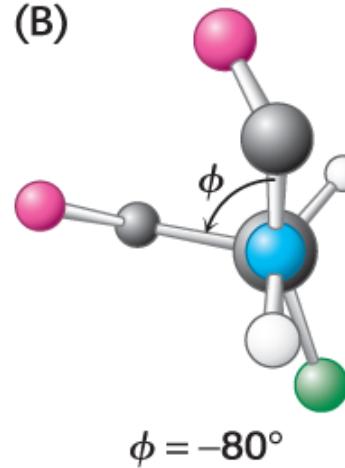
Polypeptide Chains Are Flexible Yet Conformationally Restricted (3/3)

- Rotation is permitted about the N-C_α bond [the phi (Φ) bond] and about the C_α- carbonyl bond (the psi (ψ) bond).
- The rotation about the Φ and ψ bonds, called the torsion angle, determines the path of the polypeptide chain. Not all torsion angles are permitted.

(A)



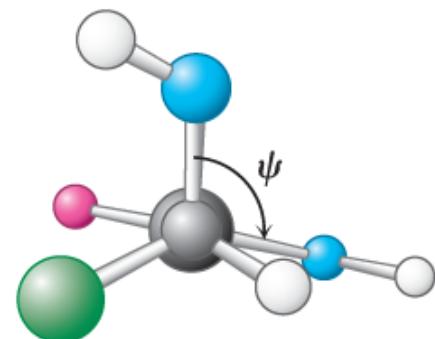
(B)



$$\phi = -80^\circ$$

View down
the N-C_α bond

(C)

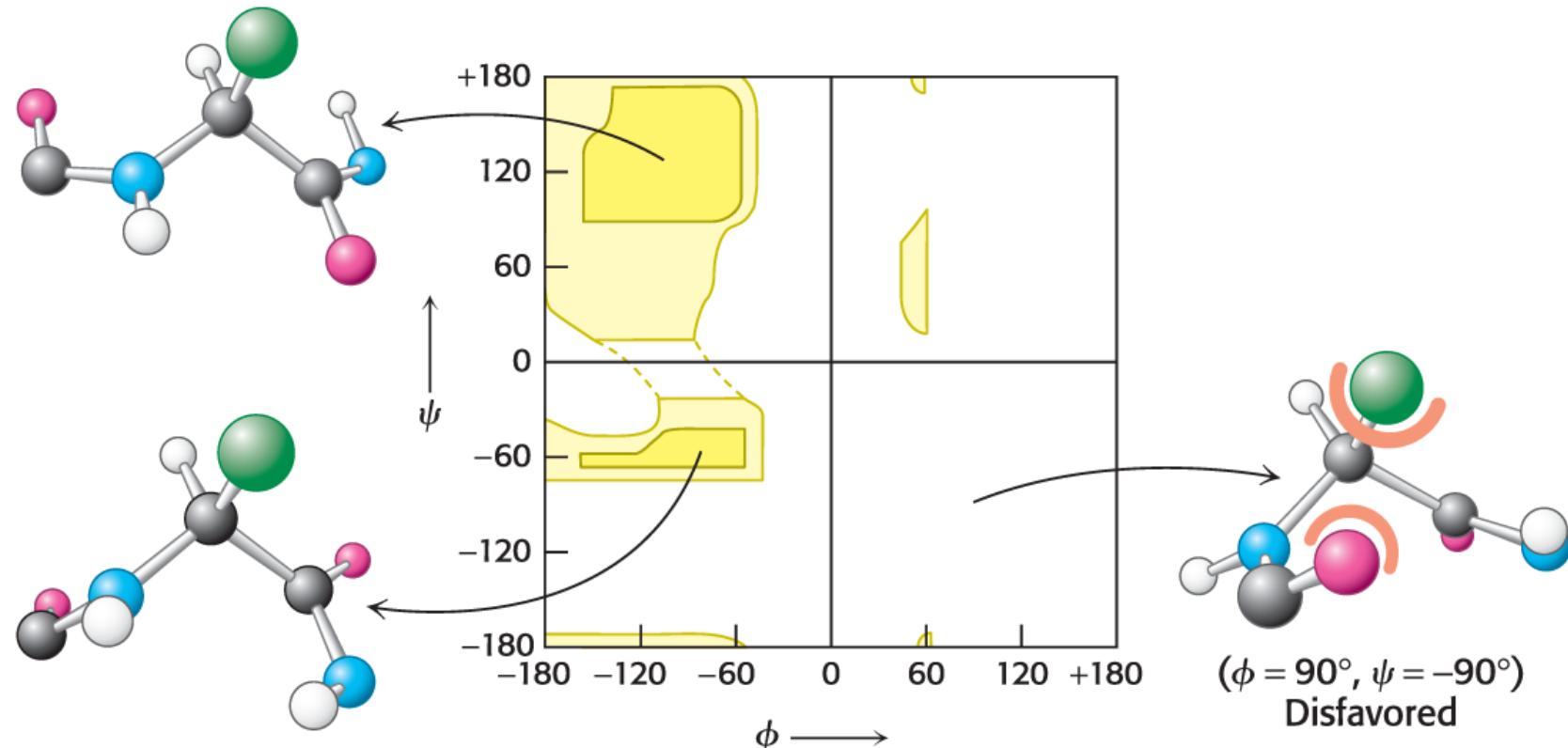


$$\psi = -85^\circ$$

View down
the CO-C_α bond

If phi and psi each range from -180° to $+180^\circ$, that's $360 \times 360 = \sim 130,000$ possible combinations per residue. Are they all equally likely?

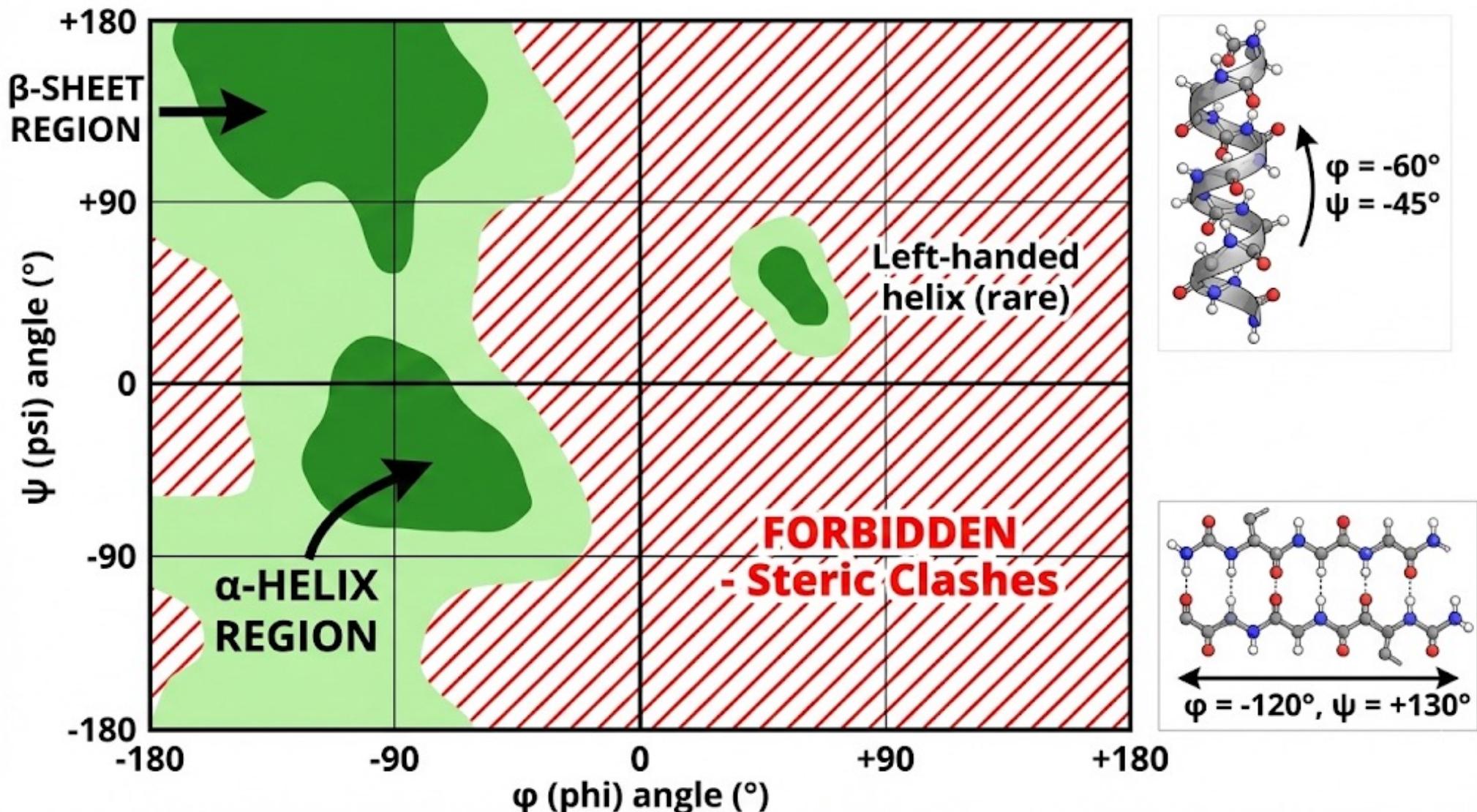
A Ramachandran Diagram Showing the Values of Φ and Ψ .



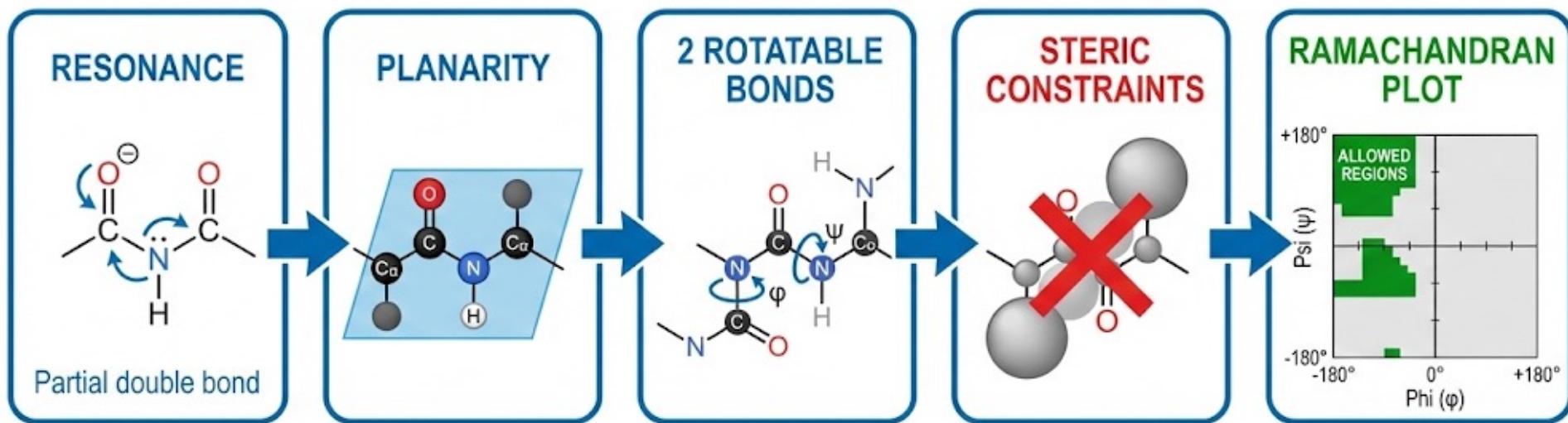
Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

Do these favored confirmational neighborhoods tell us something else about the protein structure?

Reading the Ramachandran plot.

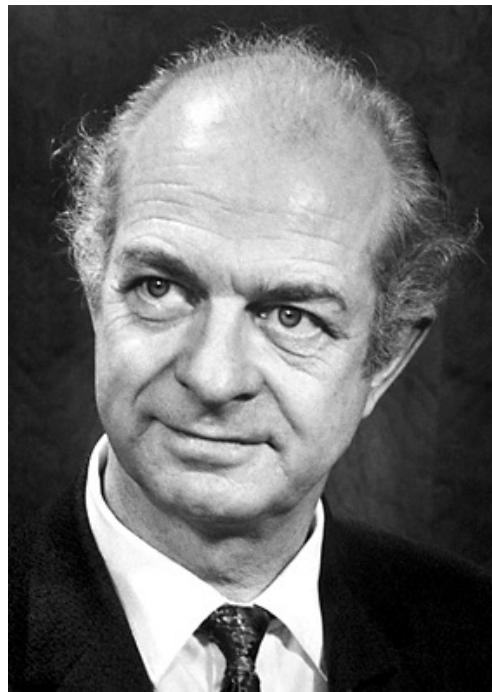


FROM PEPTIDE BOND CHEMISTRY TO ALLOWED CONFORMATIONS



- The peptide bond is locked (planar, partial double-bond)
- Trans configuration dominates (less steric clash)
- Only TWO bonds in the backbone CAN rotate: phi (ϕ) and psi (ψ)
- But even these aren't free—steric clashes eliminate most combinations
- Result: Only ~15% of the phi-psi space is allowed

Secondary structure: Linus Paulina

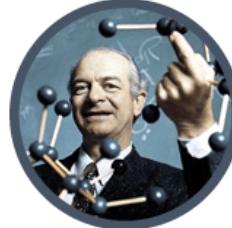


Linus Pauling
1954 Nobel in Chemistry
1962 Nobel Peace prize

They described him as "Linus Carl Pauling, who ever since 1946 has campaigned ceaselessly, not only against nuclear weapons tests, not only against the spread of these armaments, not only against their very use, but against all warfare as a means of solving international conflicts."[\[108\]](#)

TODAY IN CHEMISTRY HISTORY

28TH FEBRUARY - LINUS PAULING'S BIRTHDAY



LINUS PAULING
BORN
28 February 1901
DIED
19 August 1994

Pauling is famed for his work on the nature of chemical bonding, for which he received a Nobel Prize in Chemistry. He proposed the Pauling electronegativity scale in 1932. He also carried out research on biological molecules.

ELECTRONEGATIVITY AND THE PAULING SCALE

X:X	δ^- :Y	-X: δ^+ Z
COVALENT BOND	POLAR COVALENT	IONIC BOND
0.0–0.4	0.4–1.7	>1.7

DIFFERENCE IN ELECTRONEGATIVITY VALUES

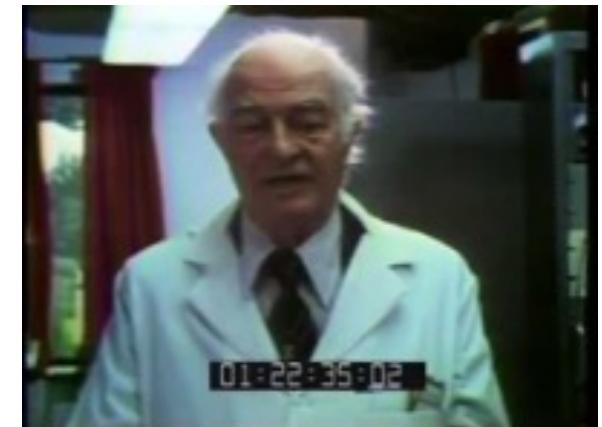
Electronegativity is a measure of the tendency of an atom to attract electrons when it is part of a compound. Generally, electronegativity increases moving towards the top right of the Periodic Table. The difference in electronegativity between two bonded atoms gives information on the nature of the chemical bond between them.

THE PAULING ELECTRONEGATIVITY SCALE

UNKNOWN	1.25-1.50	2.00-2.50
0.75-1.00	1.50-1.75	2.50-3.00
1.00-1.25	1.75-2.00	3.00-4.00

© COMPOUND INTEREST 2017 - WWW.COMPOUNDCHEM.COM | @COMPOUNDCHEM
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.

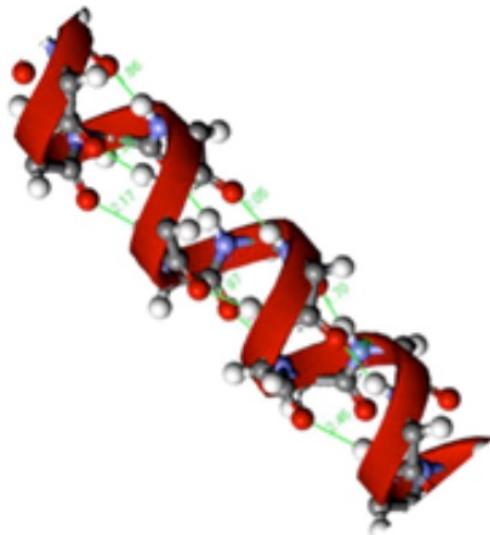
CC BY NC ND



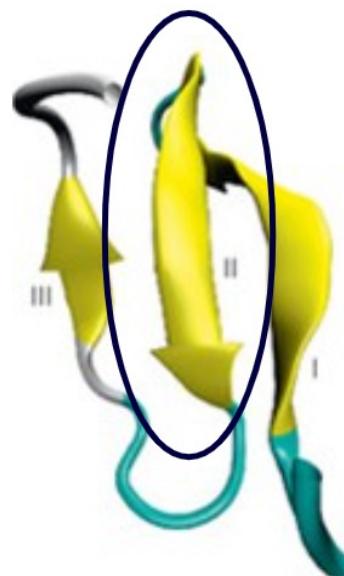
Section 4.2

Secondary Structure: Polypeptide Chains Can Fold into Regular Structures

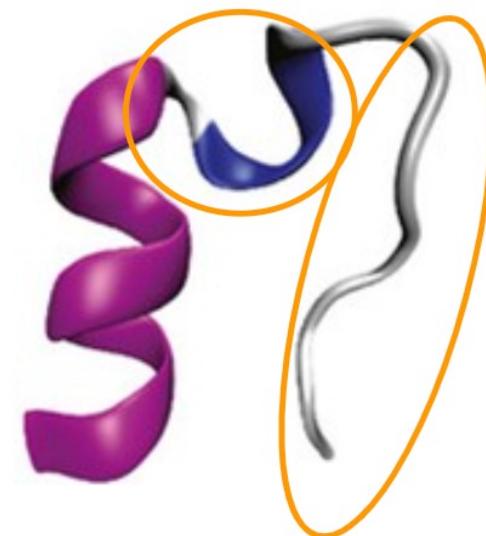
- Secondary structure is the three-dimensional structure formed by hydrogen bonds between peptide NH and CO groups of amino acids that are near one another in the primary structure.
- The α helix, β sheets, and turns are prominent examples of secondary structure.



Alpha helix



Beta strand (sheet)

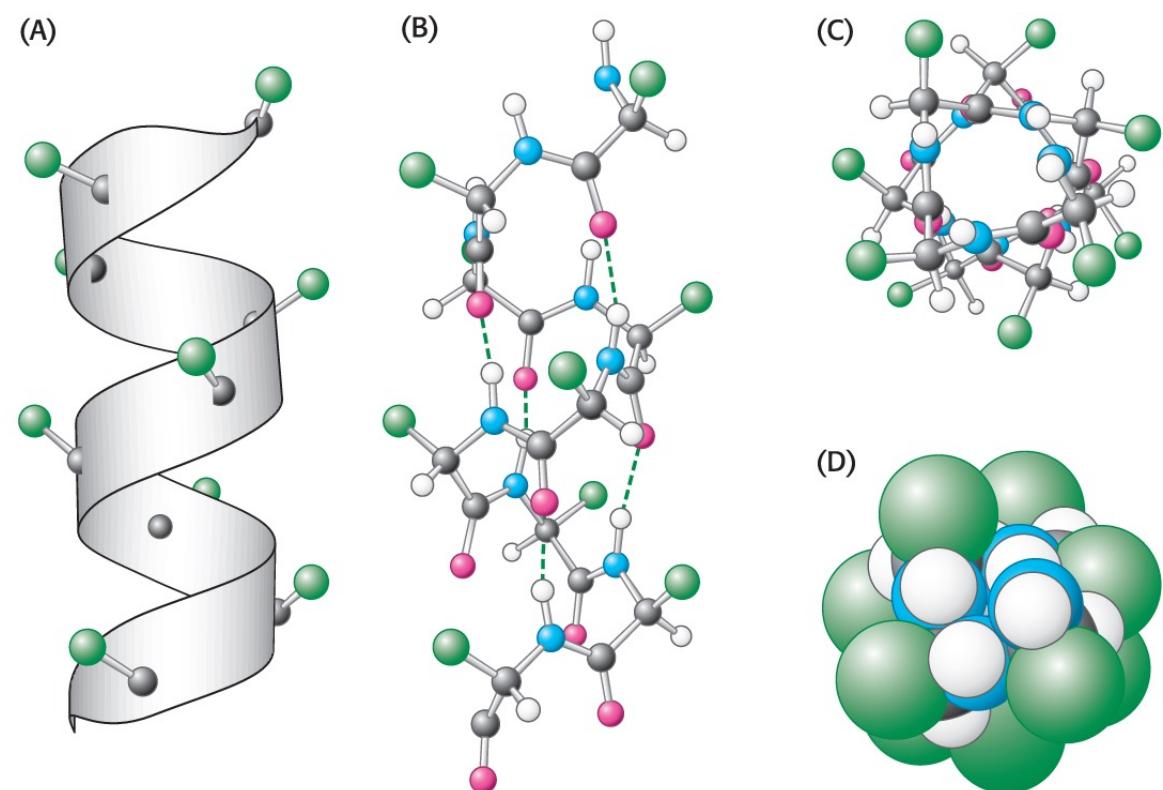


– Anything else
turn/loop

Structures of the Alpha Helix

The Alpha Helix Is a Coiled Structure Stabilized by Intrachain Hydrogen Bonds

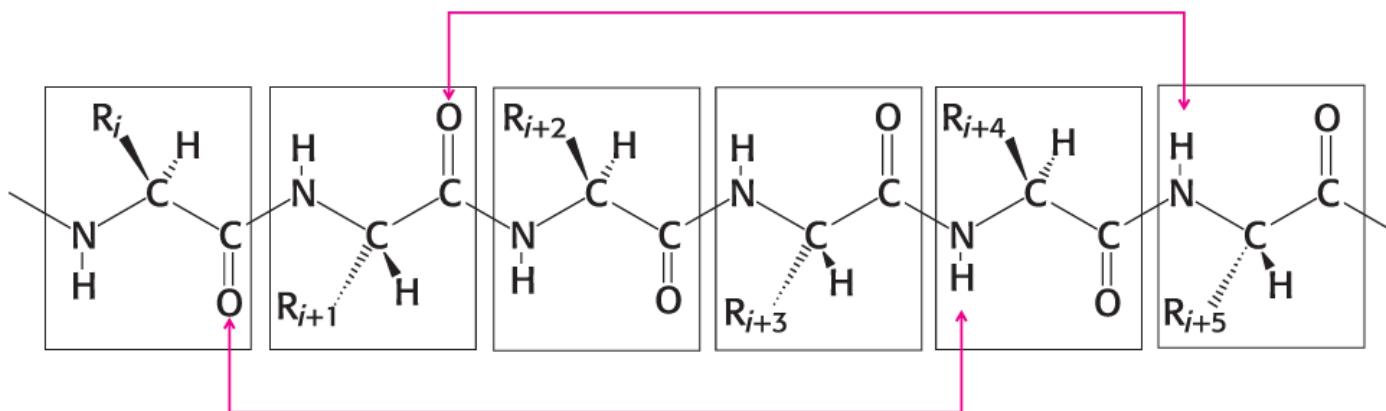
- The α helix is a tightly coiled rodlike structure, with the R groups bristling out from the axis of the helix.
- The CO group of each amino acid forms a hydrogen bond with the NH group of the amino acid that is situated four residues ahead in the sequence. All of the backbone CO and NH groups form hydrogen bonds except those at the end of the helix.
- Essentially all α helices found in proteins are right-handed.



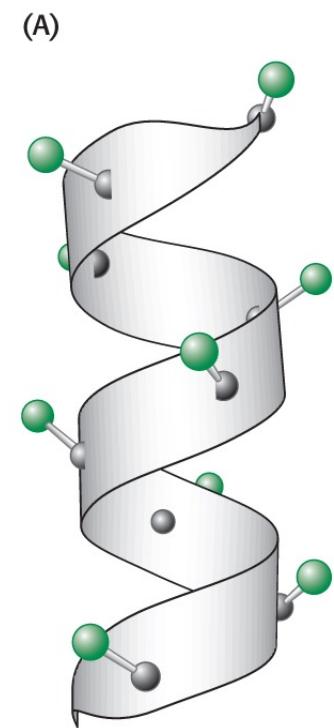
Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

The Hydrogen-Bonding Scheme for an Alpha Helix

5.4 Angstroms per turn



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company



Tymoczko et al., *Biochemi*:

Schematic views

(A)



(B)

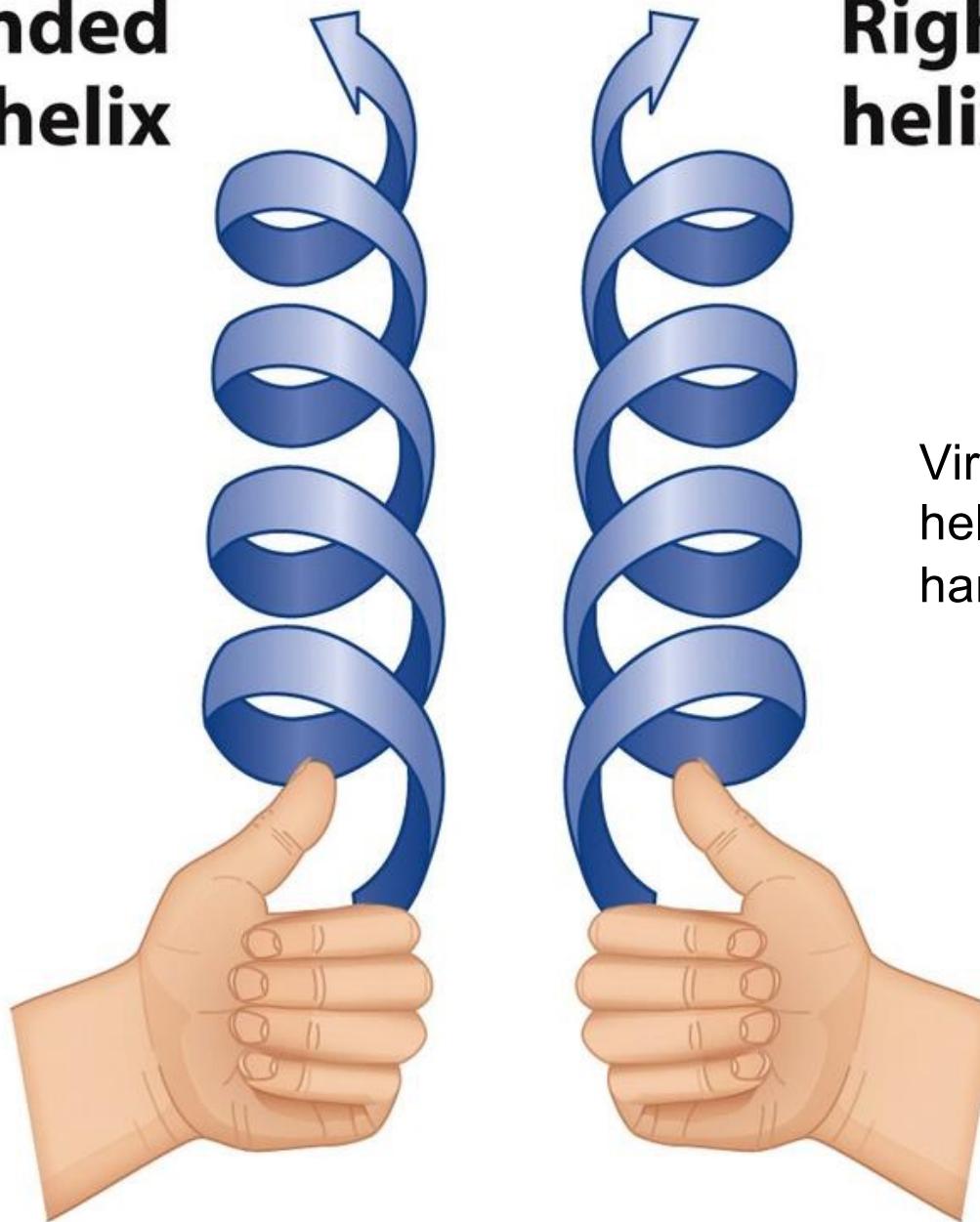


3.6 amino acids per turn

In an alpha helix, do the R-groups point INTO the helix core or OUT?

Screw Sense

Left-handed helix



Right-handed helix

Virtually all alpha helices are right handed

Box 4-1

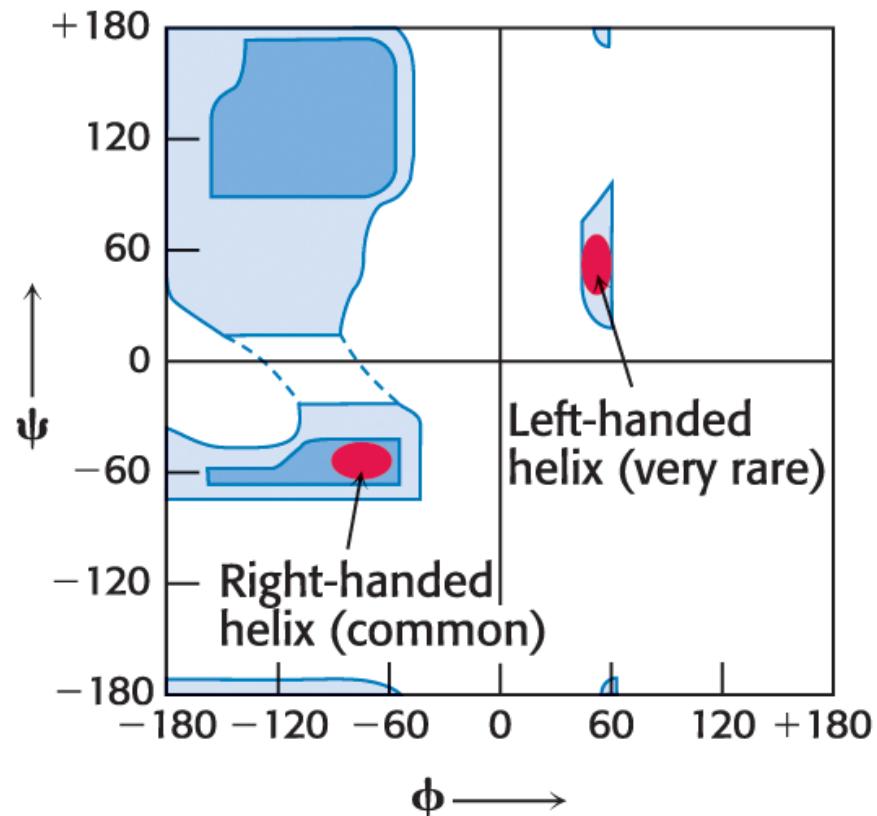
Lehniger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

The α helix (Alpha Helix)

The *screw sense* of a helix refers to the direction in which a helical structure rotates with respect to its axis. If viewed down the axis of a helix (N terminus to C terminus) and the chain turns

- clockwise, it has a right-handed screw sense.
- counterclockwise, the screw sense is left-handed.

Right-handed helices are energetically more favorable because there are fewer steric clashes between the side chains and the backbone.

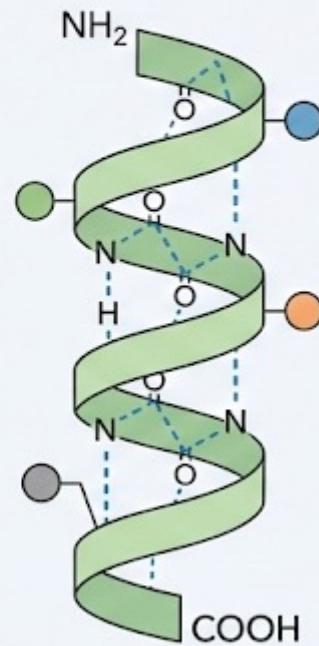


Essentially all α helices found in proteins are right-handed.

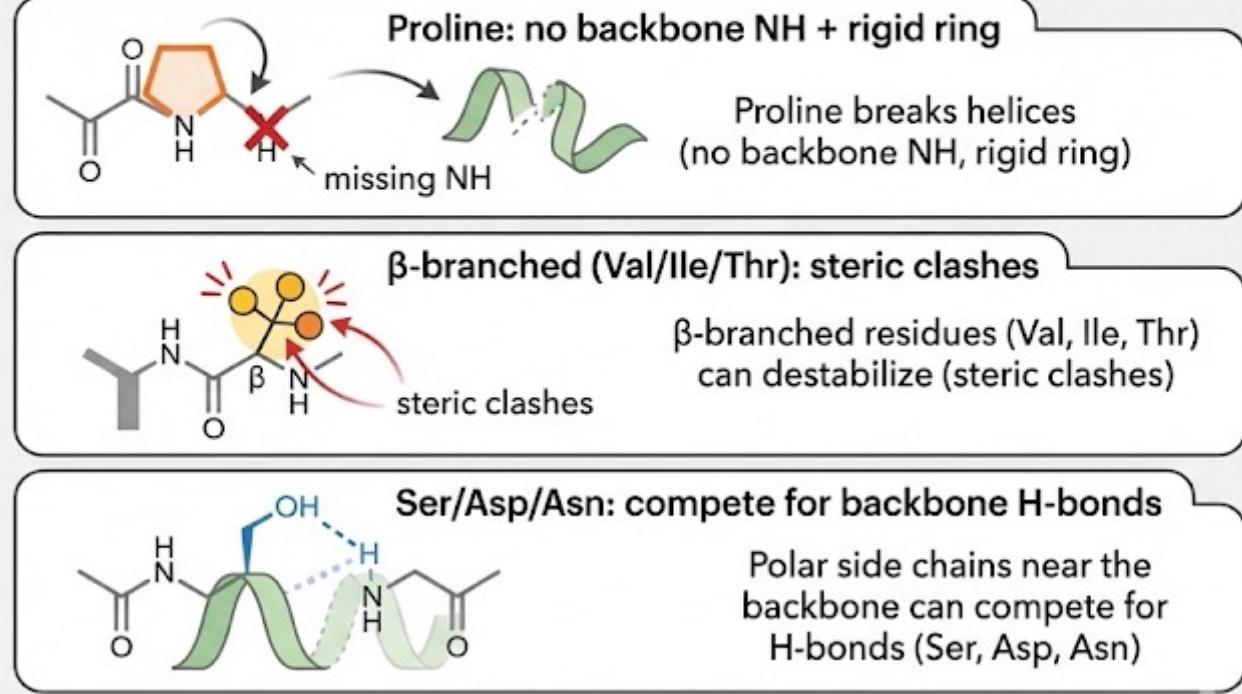
Not every sequence likes an α helix

Primary structure biases what secondary structure is even possible.

Simplified α -helix Structure



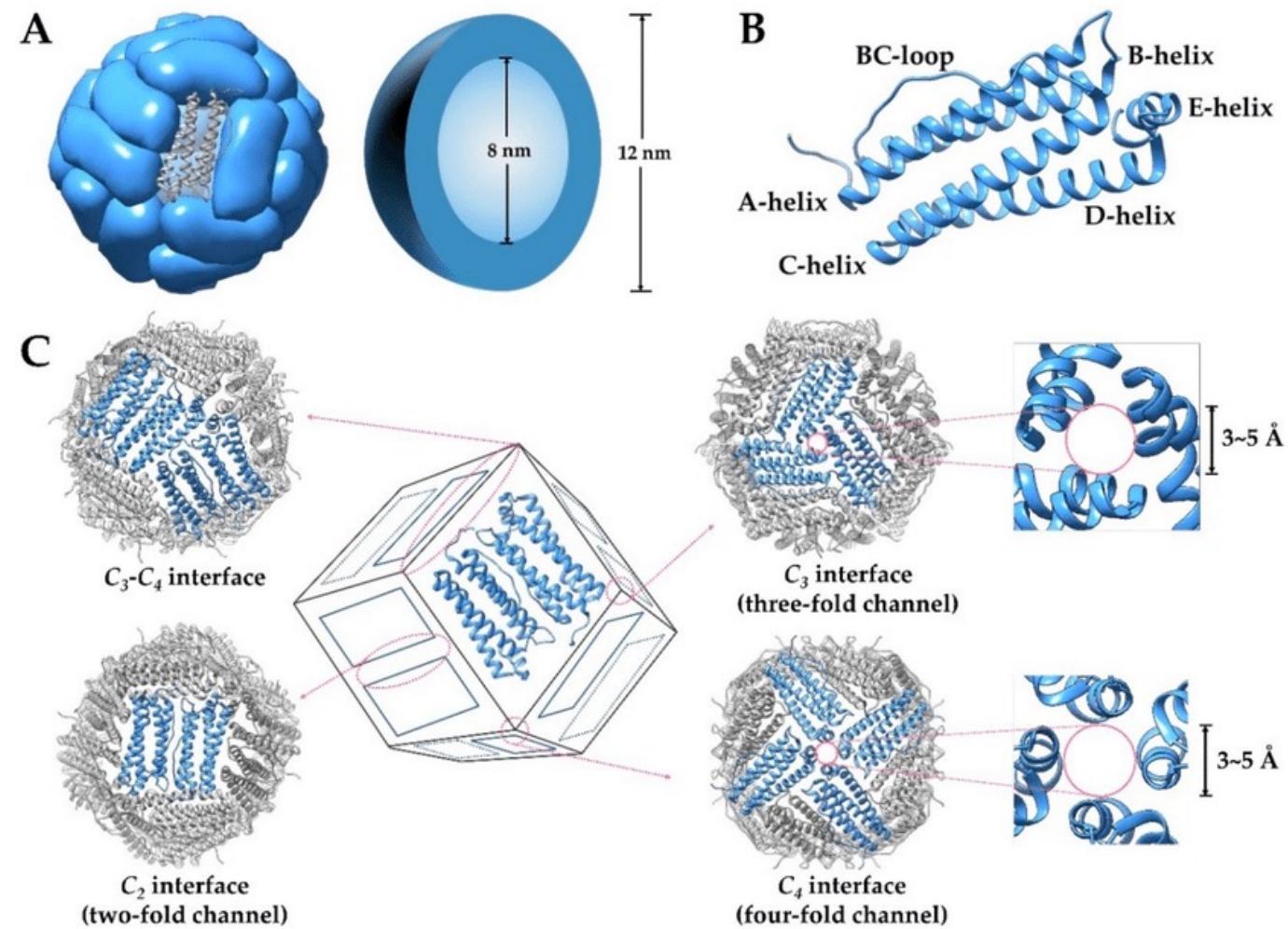
Mechanisms of Helix Disruption



Helices tolerate many residues, but sequence biases shape secondary structure

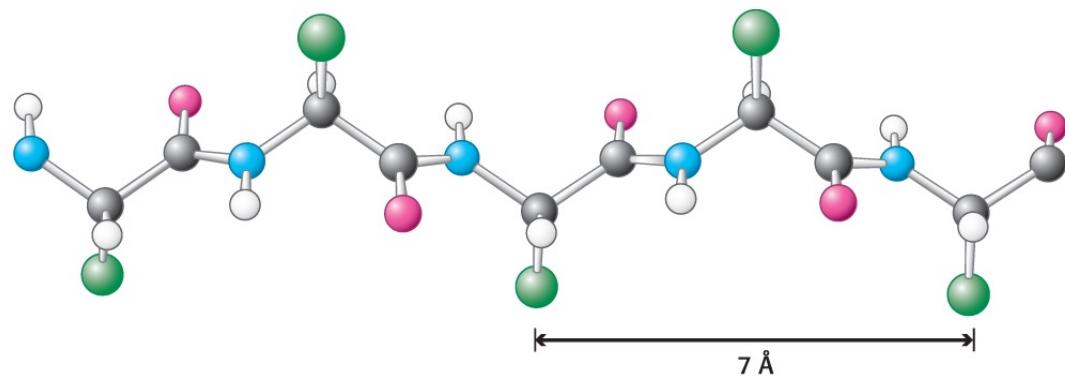
Why does proline break alpha helices?

Model of Ferritin, a Largely Alpha-Helical Protein



Beta Sheets Are Stabilized by Hydrogen Bonding Between Polypeptide Strands (1/2)

- The β sheet is another common form of secondary structure.
- Beta sheets are formed by adjacent β strands.
- In contrast to an α helix, the polypeptide in a β strand is fully extended.

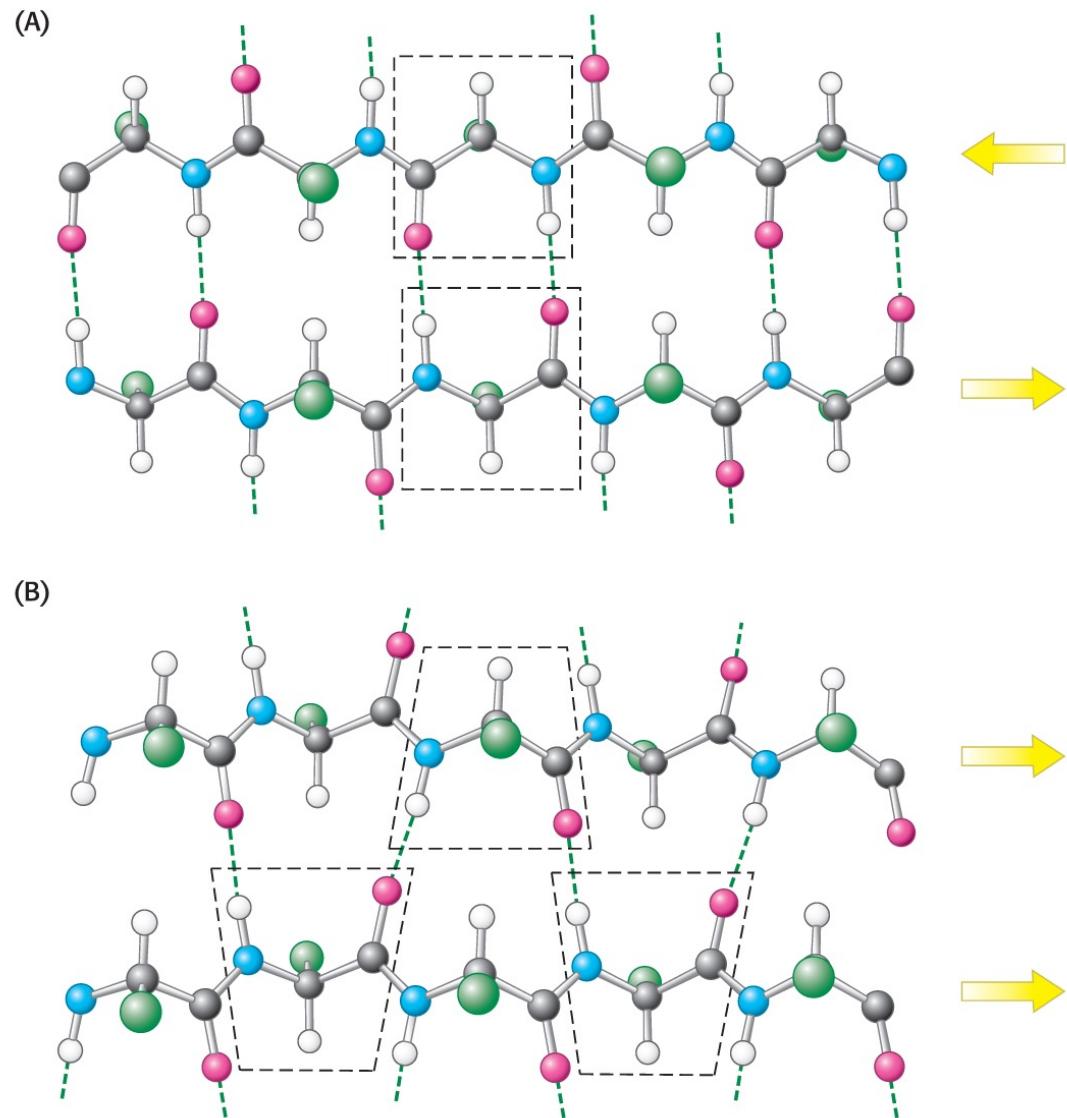


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

Structure of Antiparallel and Parallel Beta Sheets

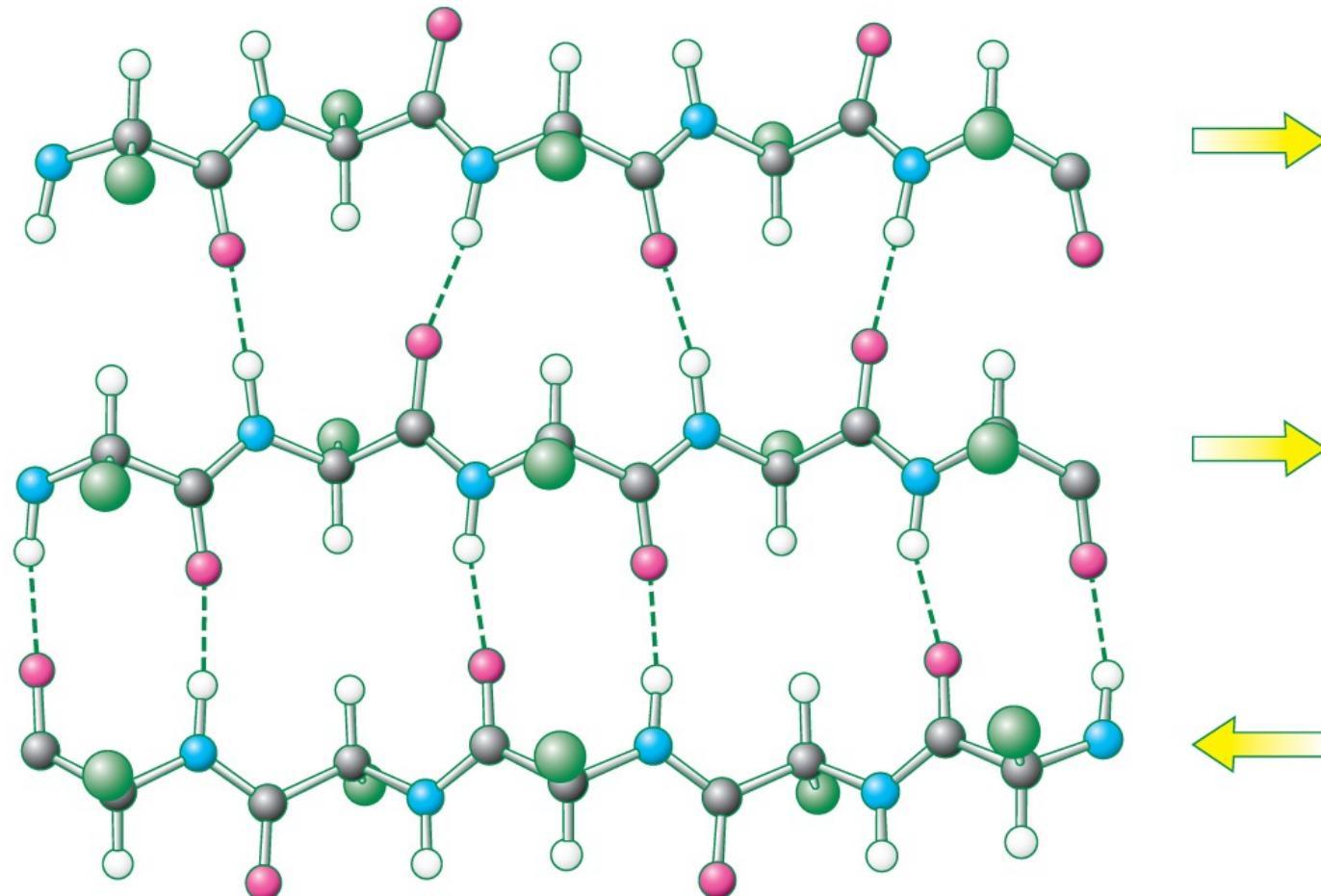
Beta Sheets Are Stabilized by Hydrogen Bonding Between Polypeptide Strands (2/2)

- Hydrogen bonds link the strands in a β sheet.
- The strands of a β sheet may be parallel, antiparallel, or mixed.
- β sheets may be almost flat or adopt a twisted conformation.



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

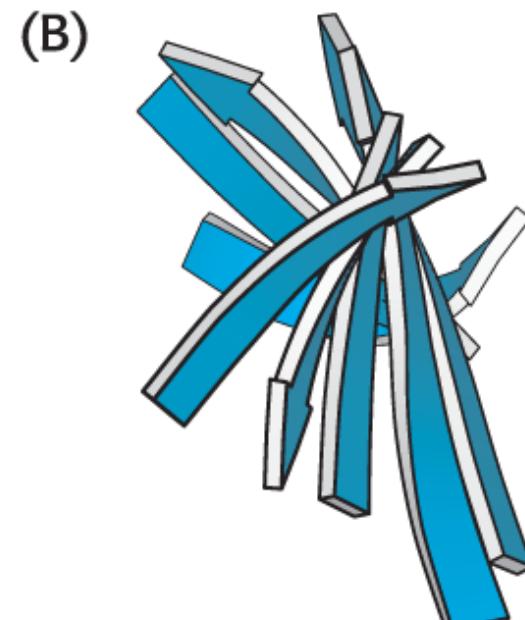
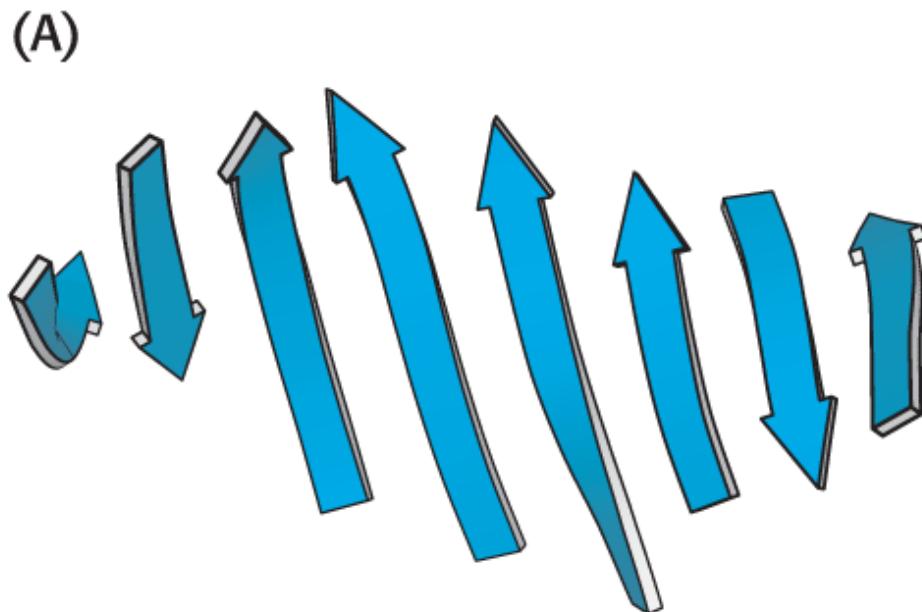
Sometimes We Have a Mixed Beta Sheet



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

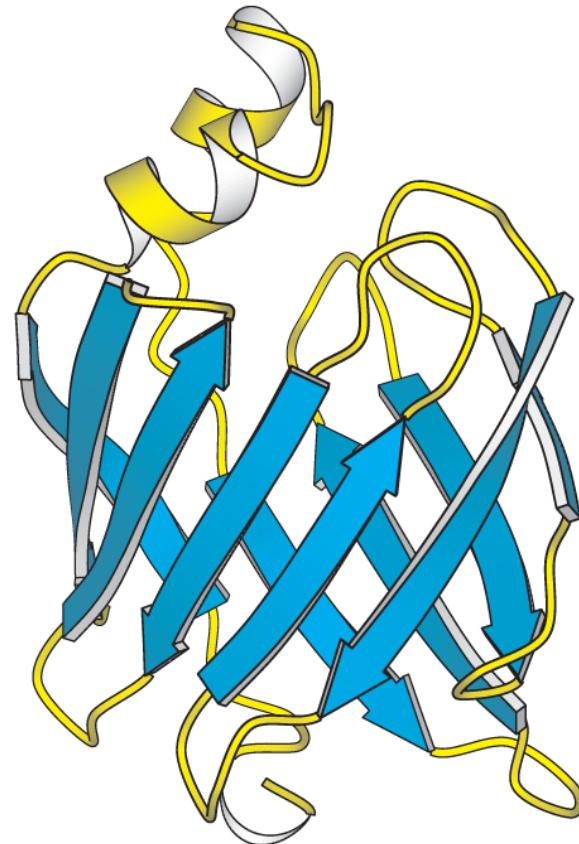
Model of a Twisted Beta Sheet

- **Twist and Turns**
- Unlike alpha-helices, which are coiled, beta sheets often feature twists and turns. These twists are not random but rather introduce a slight 'pleated' appearance, maximizing hydrogen-bonding efficiency and structural stability.



Model of a Protein Rich in Beta Sheets

- Imagine a barrel made entirely of beta sheets. Sounds like a science fiction concept, right? Wrong! This is the 'beta barrel,' a cylindrical structure comprising beta sheets, and it's commonly found in outer membrane proteins of gram-negative bacteria. This barrel-like formation allows for the passage of nutrients while maintaining structural integrity."

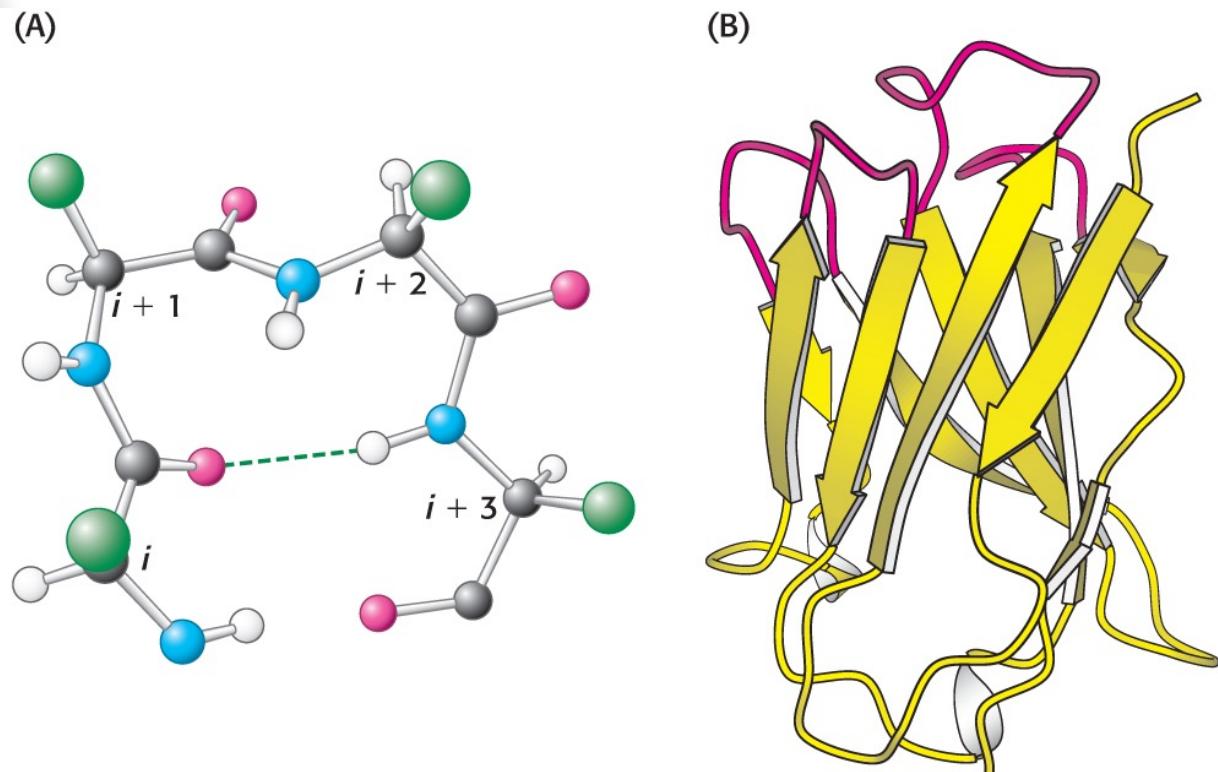


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

Structure of a Reverse Turn

Polypeptide Chains Can Change Direction by Making Reverse Turns and Loops

- Turns and loops invariably lie on the surfaces of proteins and thus often participate in interactions between other proteins and the environment.



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

Which statement is correct?

- A. The distance between adjacent amino acids along a β strand is approximately 7 Å.
- B. The rise of an α helix is equal to 3.6 Å.
- C. The number of residues per turn of an α helix is 1.5 Å.
- D. The pitch of an α helix is equal to the product of the rise and the number of residues per turn, or 5.4 Å.

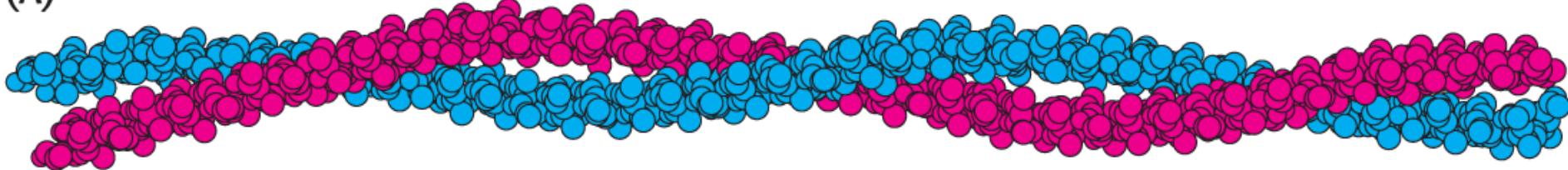
Break.



Fibrous Proteins Provide Structural Support for Cells and Tissues (1/3)

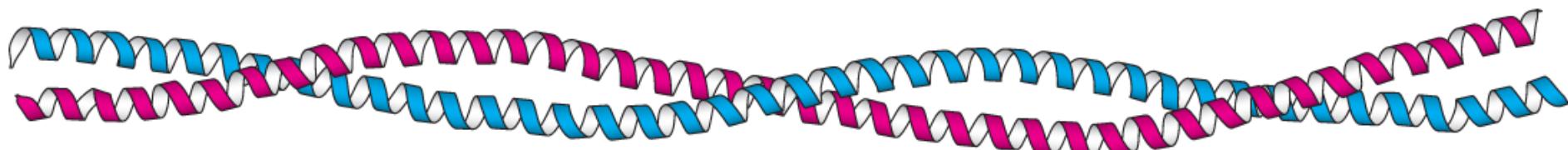
- α -Keratin, a structural protein found in wool and hair, is composed of two right-handed α helices intertwined to form a left-handed superhelix called a coiled coil. The helices interact with ionic interactions or van der Waals forces.
- α -Keratin is a member of a superfamily of structural proteins called coiled-coil proteins.
- Whether it's keratin in your hair, collagen in your tendons, or fibroin in spider silk, fibrous proteins are versatile multitaskers. They demonstrate how diverse structural elements can be employed to meet a variety of functional demands, from mechanical strength to elasticity.

(A)



Structure of an Alpha-Helical Coiled Coil

(B)

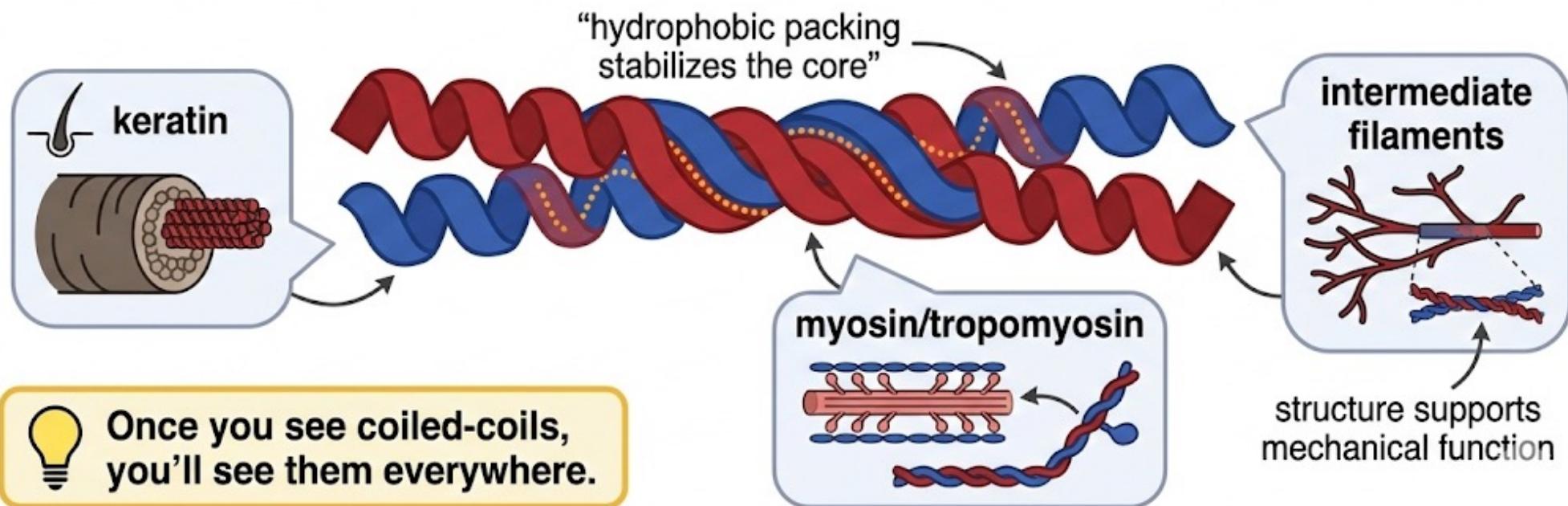


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

These two helices may be linked by disulfides

Coiled-coils are a common structural “cable” motif

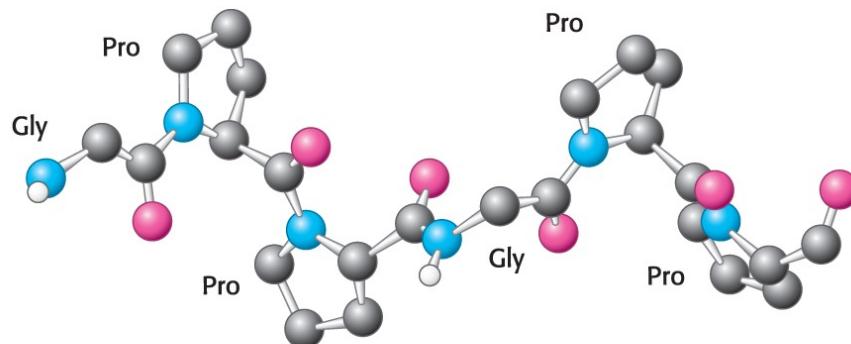
- ✓ Coiled-coil = 2+ α helices wrapped together (superhelix)
- ✓ Stabilized by hydrophobic packing and complementary charges
- ✓ Used for strength, scaffolding, movement (keratin, myosin, tropomyosin)
- ✓ Same motif shows up across many proteins (pattern recognition payoff)



Fibrous Proteins Provide Structural Support for Cells and Tissues (2/3)

- Collagen is a structural protein that is a component of skin, bone, tendons, cartilage, and teeth.
- Collagen consists of three intertwined helical polypeptide chains that form a superhelical cable. The helical polypeptide chains of collagen are not α helices.
- Glycine appears at every third residue, and the sequence Gly-Pro-Pro is common.

The Amino Acid Sequence of a Part of a Collagen Chain

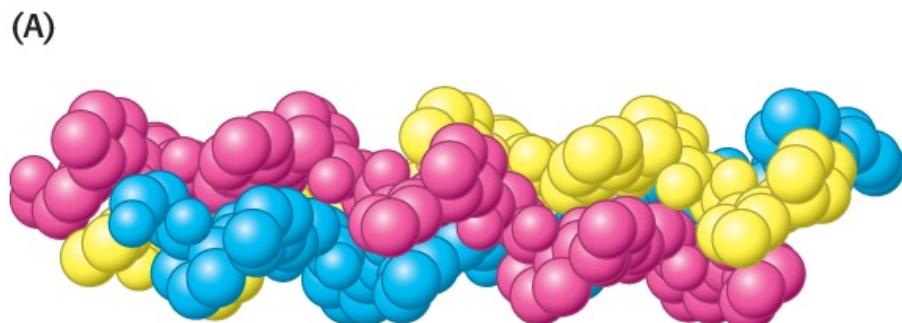


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

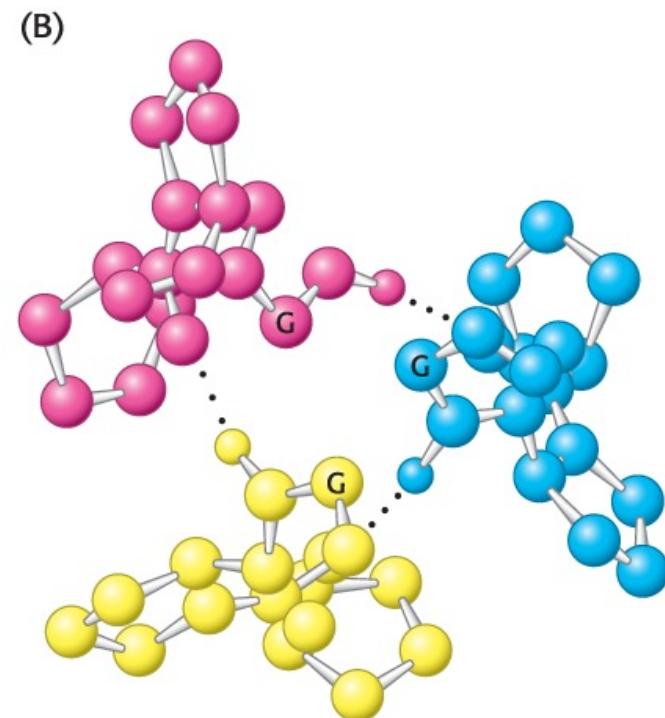
13
-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-
22
-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Ala-Hyp-
31
-Gly-Pro-Gln-Gly-Phe-Gln-Gly-Pro-Hyp-
40
-Gly-Glu-Hyp-Gly-Glu-Hyp-Gly-Ala-Ser-
49
-Gly-Pro-Met-Gly-Pro-Arg-Gly-Pro-Hyp-
58
-Gly-Pro-Hyp-Gly-Lys-Asn-Gly-Asp-Asp-

Fibrous Proteins Provide Structural Support for Cells and Tissues (3/3)

- The helices in collagen are not stabilized by hydrogen bonds. Rather, they are stabilized by steric repulsion of the pyrrolidine rings of proline. The three intertwined chains interact with one another with hydrogen bonds.
- The interior of the superhelical cable is crowded, and only glycine can fit in the interior.

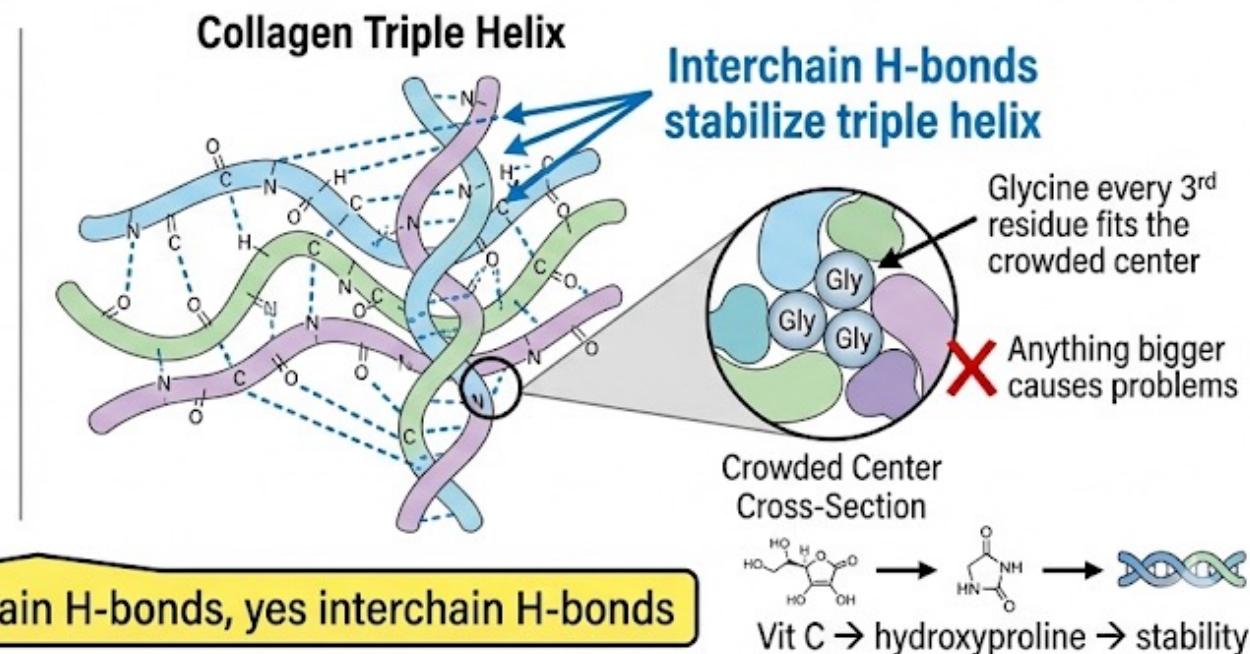
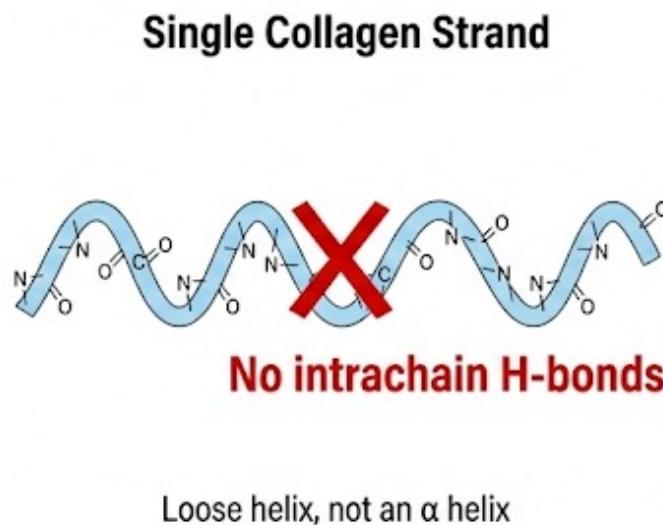


The Structure of the Protein Collagen



Collagen triple helix: what is and is not H-bond stabilized

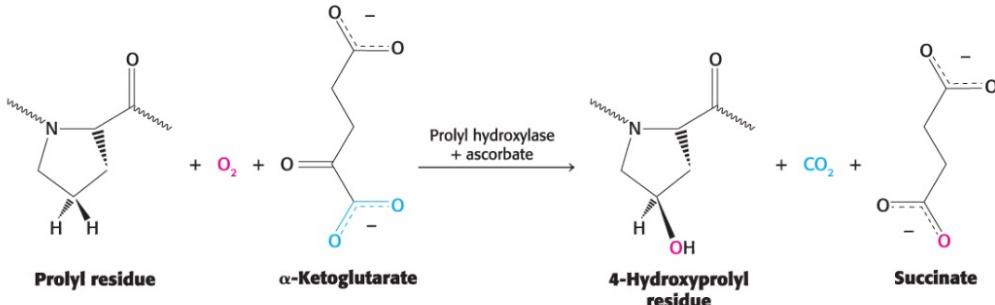
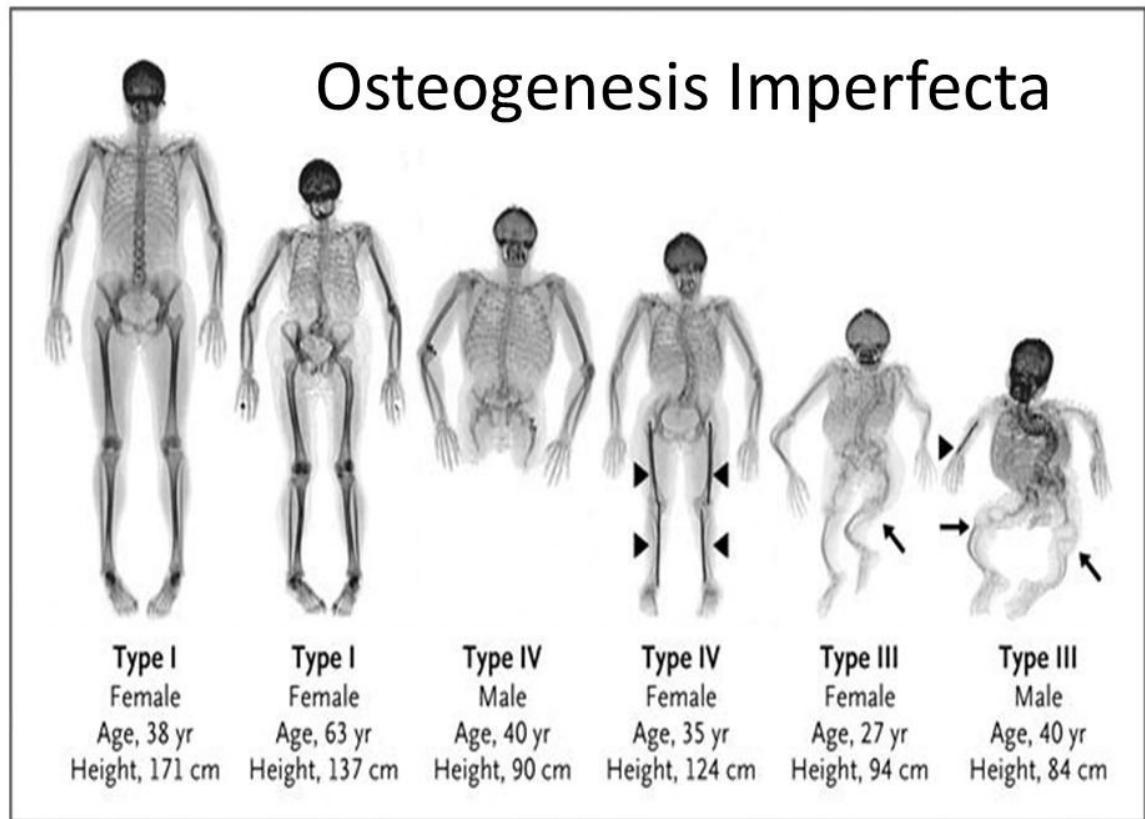
- Each collagen strand is not an α helix and has no intrachain H-bonds
- The triple helix is stabilized by interchain H-bonds (between strands)
- Gly every 3rd fits the crowded center (anything bigger causes problems)
- Hydroxyproline increases triple-helix stability (vitamin C dependent)



CLINICAL INSIGHT

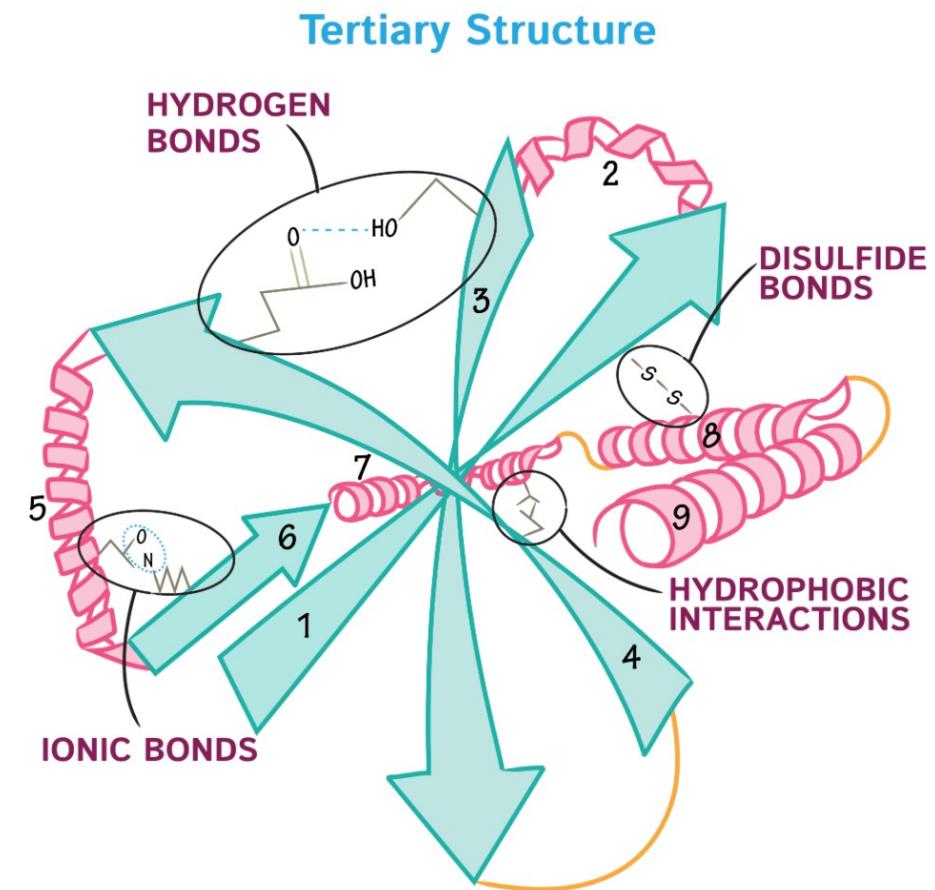
Defects in Collagen Structure Result in Pathological Conditions

- Osteogenesis imperfecta, or brittle bone disease, occurs if a mutation results in the substitution of another amino acid in place of glycine
- Hydroxyproline, a modified version of proline in which a hydroxyl group replaces a hydrogen, is important for the stabilization of collagen.



Section 4.3 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures

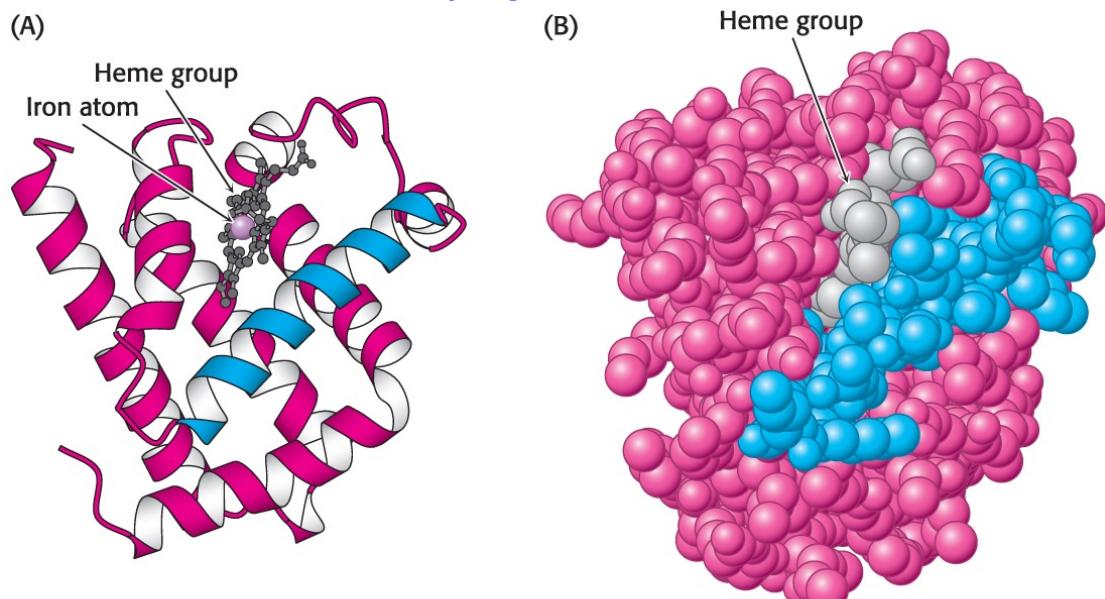
- Tertiary structure refers to the spatial arrangement of amino acids that are far apart in the primary structure and to the pattern of disulfide bond formation.
- This level of structure is the result of interactions between the R groups of the peptide chain.



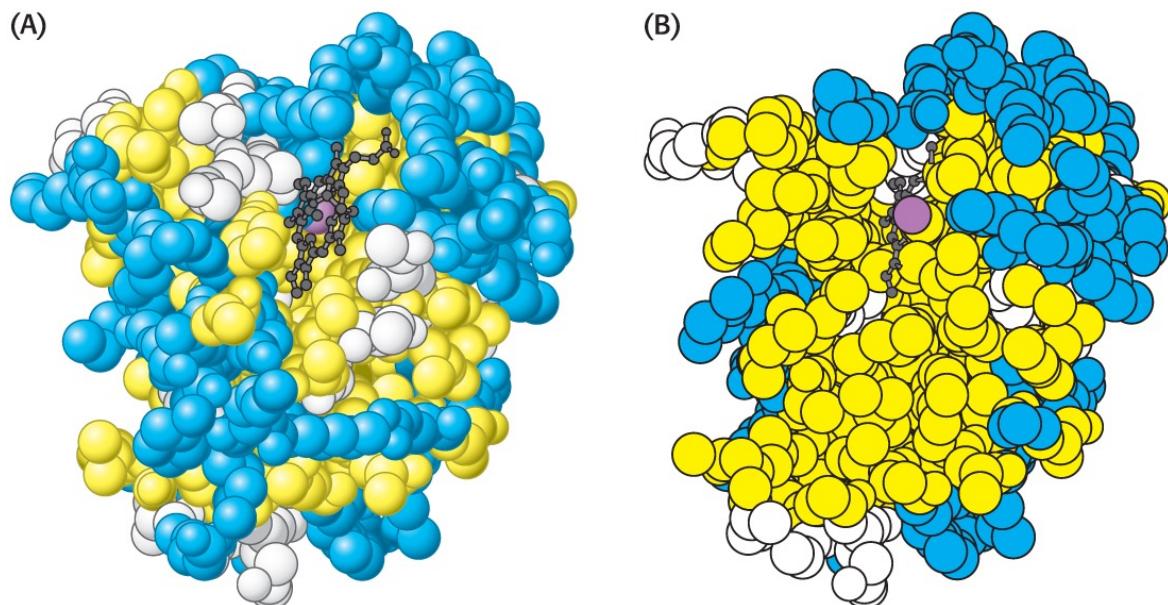
Myoglobin Illustrates the Principles of Tertiary Structure

- Globular proteins such as myoglobin form complicated three-dimensional structures.
- Globular proteins are very compact. There is little or no empty space in the interior of globular proteins.
- The interior of globular proteins consists mainly of hydrophobic amino acids.
- The exterior of globular proteins consists of charged and polar amino acids.

The Three-Dimensional Structure of Myoglobin



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

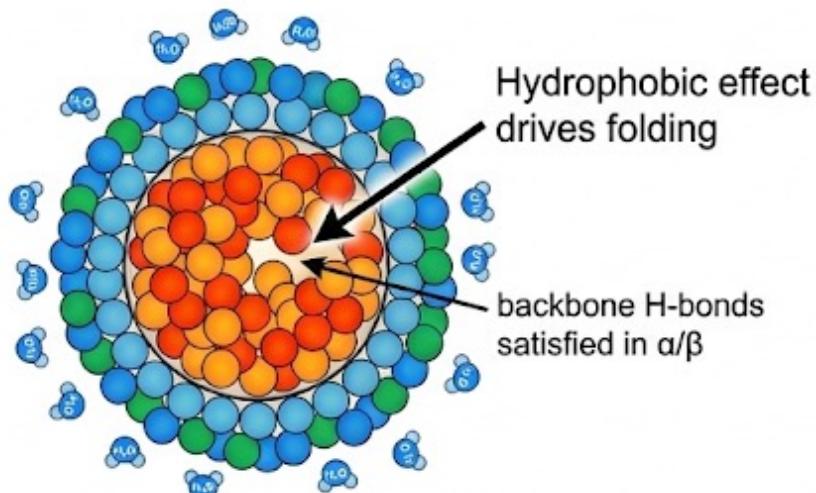


Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

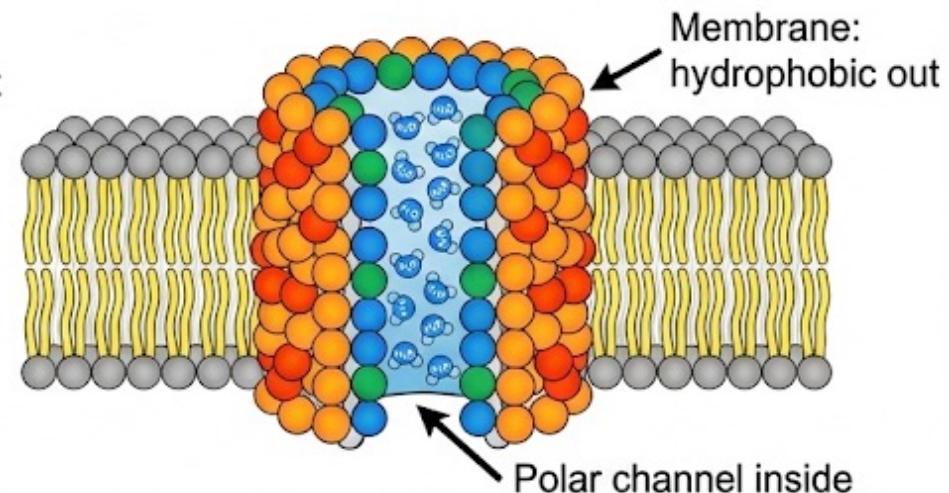
The hydrophobic effect is the engine of folding

- ✓ Water “pushes” hydrophobics together: bury nonpolar side chains
- ✓ Polar/charged residues favor the surface in soluble proteins
- ✓ Buried backbone must satisfy H-bonds, so α/β help pack safely
- ✓ Membrane porins are “inside-out”: hydrophobic outside, polar channel inside

Soluble proteins: hydrophobic core, polar surface



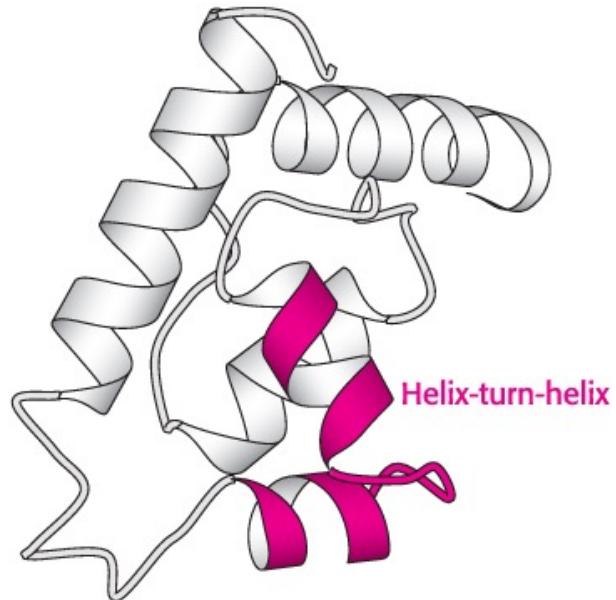
Porins: inside-out (hydrophobic outside, polar channel)



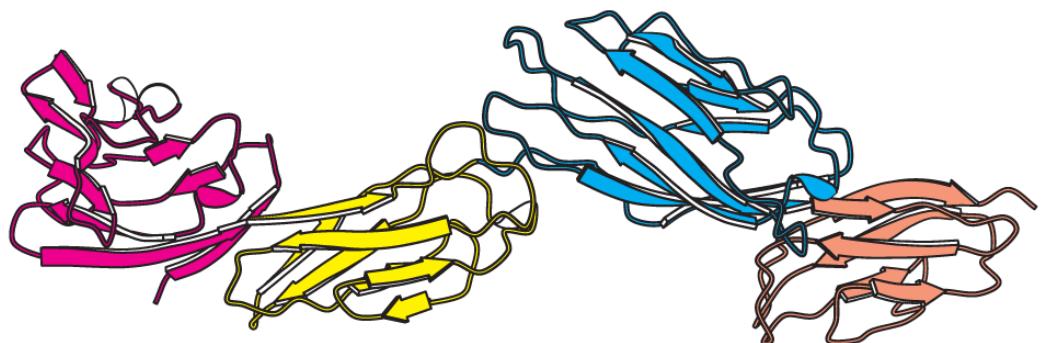
Folded proteins are a negotiation between water and chemistry.

The Tertiary Structure of Many Proteins Can Be Divided into Structural and Functional Units

- Motifs, or supersecondary structures, are combinations of secondary structure that are found in many proteins.
- Some proteins have two or more similar or identical compact structures called domains.



DNA binding protein



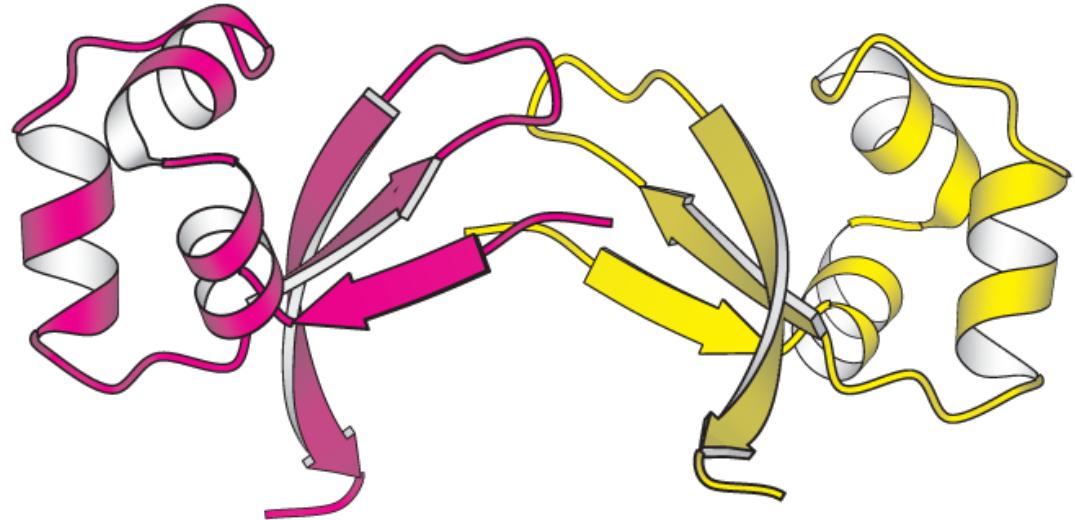
Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

Cell surface CD4 protein

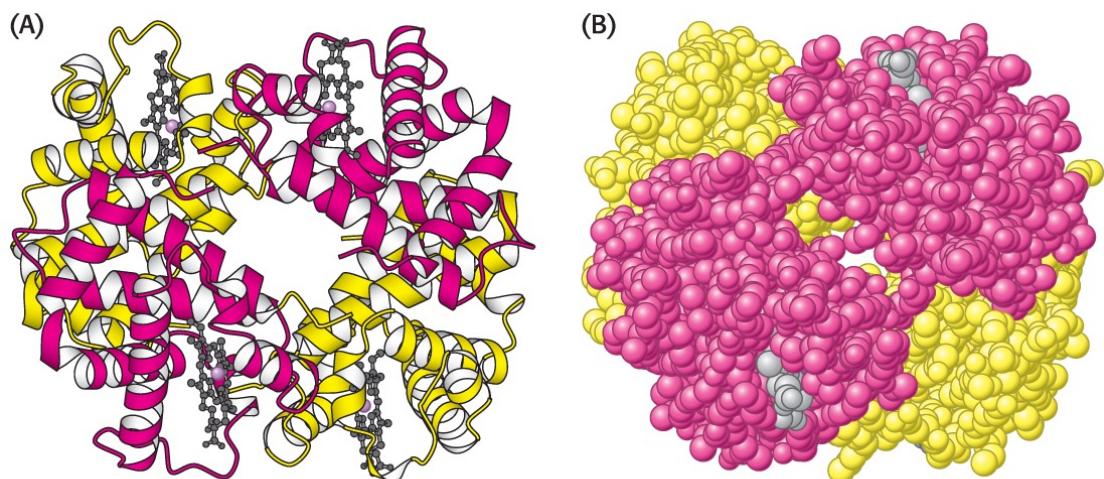
Section 4.4 Quaternary Structure Multiple Polypeptide Chains Can Assemble into a Single Protein

- Many proteins are composed of multiple polypeptide chains called subunits. Such proteins are said to display quaternary structure.
- Quaternary structure can be as simple as two identical polypeptide chains or as complex as dozens of different polypeptide chains.

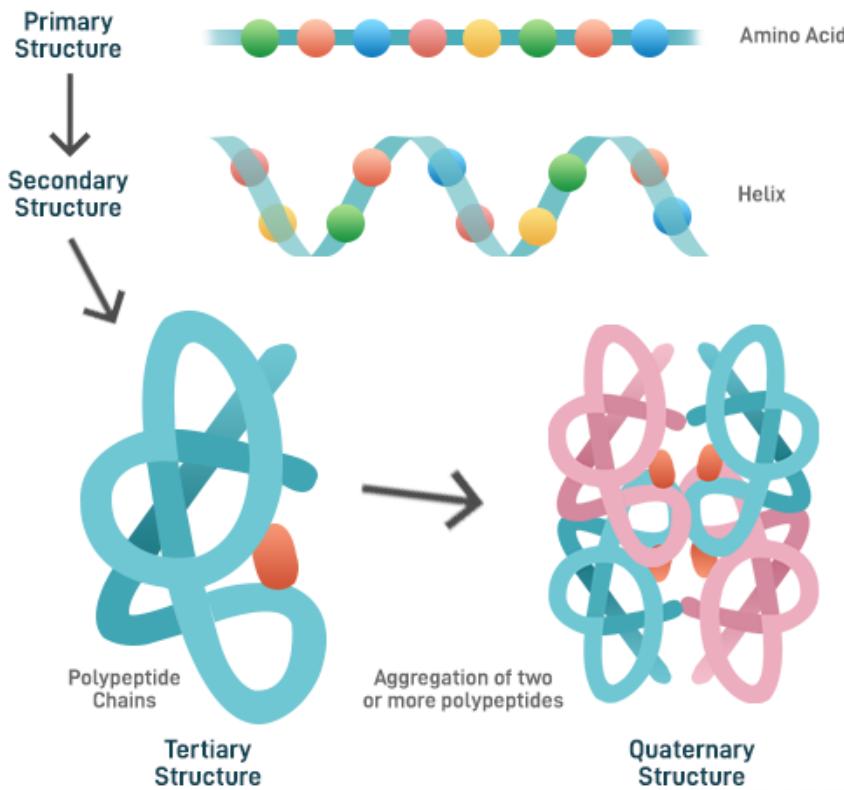
Model of Quaternary Structure



Tymoczko et al., *Biochemistry: A Short Course*, 4e,
© 2019 W. H. Freeman and Company



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company



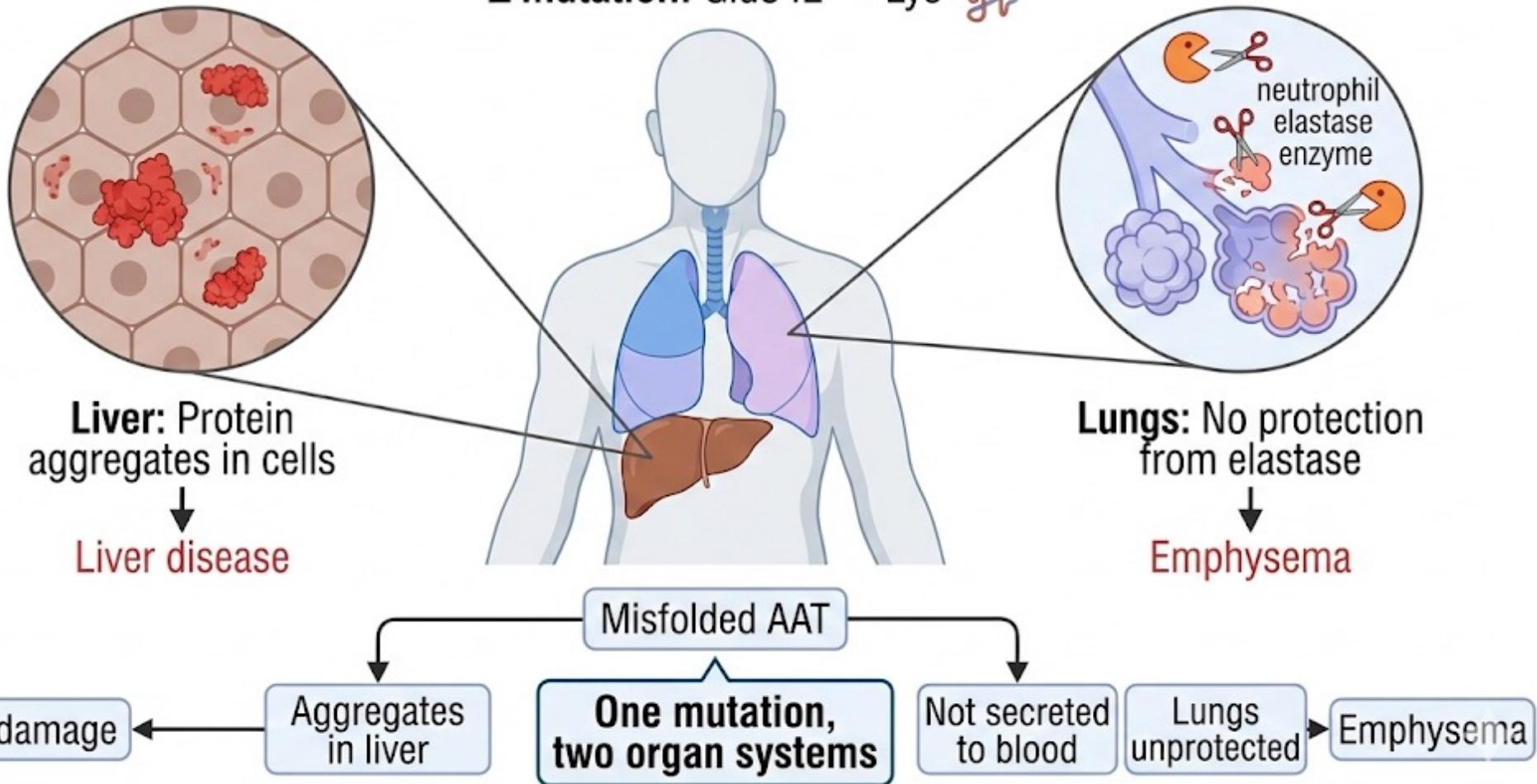
Putting it all together

Consequences of misfolding?

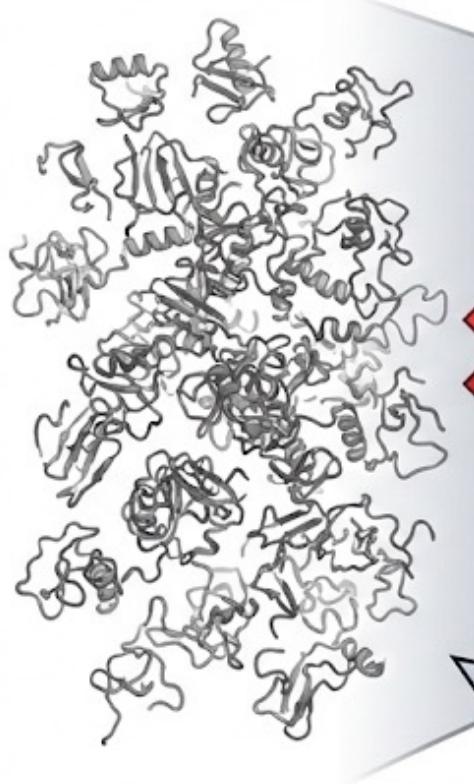
Structure Level	Definition	Primary Stabilizing Forces
PRIMARY (1°)	Linear amino acid sequence	Covalent peptide bonds $\text{C=O}-\text{N}(\text{H})-\text{H}$ (Disulfides as cross-links)
SECONDARY (2°)	Local folding: α -helix, β -sheet, turns	Backbone H-bonds ($\text{C=O}\cdots\text{H-N}$) $\text{O}=\text{C}-\text{H}\cdots\text{N}-\text{H}$
TERTIARY (3°)	Global 3D structure of single chain	Hydrophobic effect (primary driver) Side chain H-bonds, salt bridges, disulfides
QUATERNARY (4°)	Assembly of multiple subunits	Same as tertiary: hydrophobic effect, H-bonds, salt bridges, disulfides

α 1-Antitrypsin Deficiency

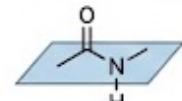
Z mutation: Glu342 → Lys 



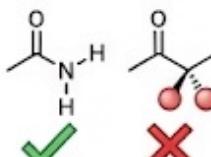
10^{100}
POSSIBLE
CONFORMATIONS



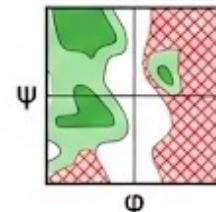
PEPTIDE BOND
PLANARITY



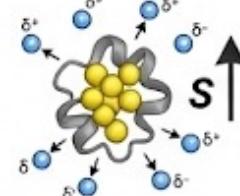
TRANS > CIS
(steric)



RAMACHANDRAN
CONSTRAINTS
(ϕ - ψ limited)



HYDROPHOBIC
EFFECT
(entropy)



ONE
CORRECT
FOLD

10^{90}
remaining

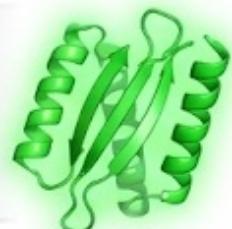
10^{90}
remaining

10^{80}
remaining

10^{80}
remaining

10^{60}
remaining

10^{60}
remaining

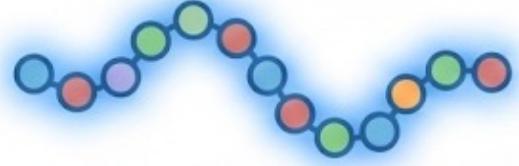


ONE
CORRECT
FOLD

CONSTRAINTS DON'T LIMIT FUNCTION—THEY CREATE IT

The \$64,000 question: Where does folding information live?

THREE HYPOTHESES



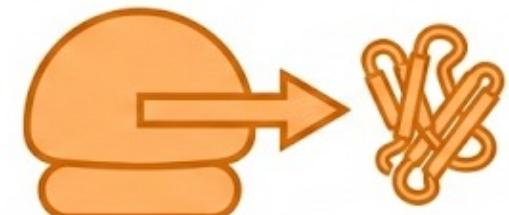
THE SEQUENCE

Information is intrinsic



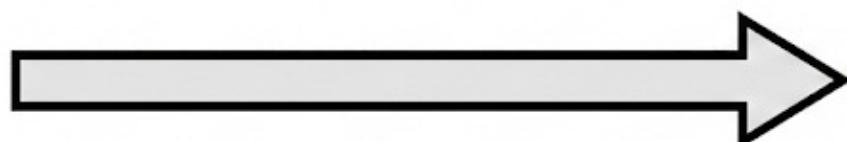
THE CELL

Cellular factors guide folding



THE RIBOSOME

Folding is co-translational



Which one? → Anfinsen's Experiment

Section 4.5 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

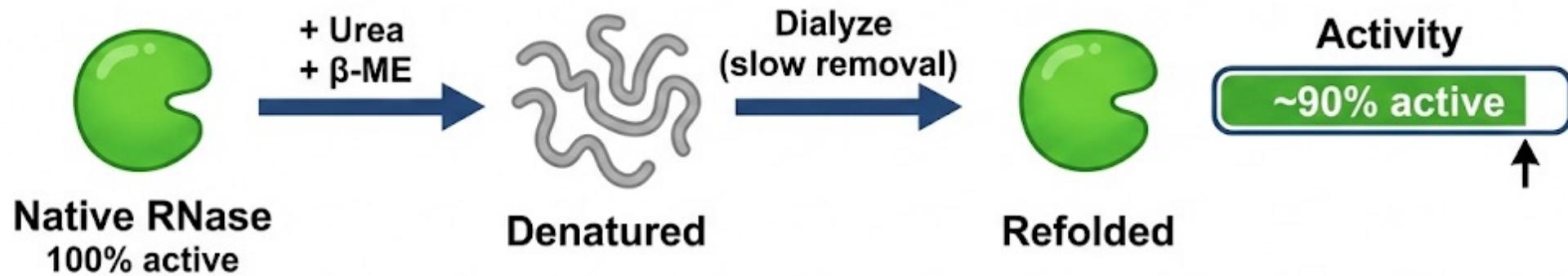
- Christian B. Anfinsen was an American biochemist known for his work on the structure and function of enzymes and proteins, especially his study on ribonuclease and how protein folding is guided by the amino acid sequence, which won him a share of the Nobel Prize in Chemistry in 1972.



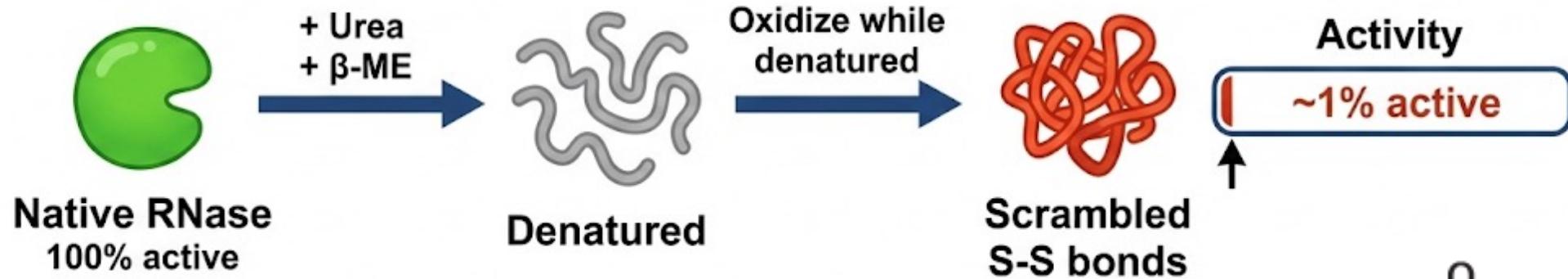
Section 4.5 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

THE MAIN EXPERIMENT

THE EXPERIMENT



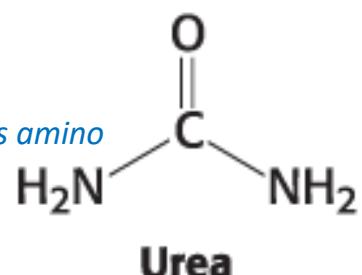
THE CONTROL



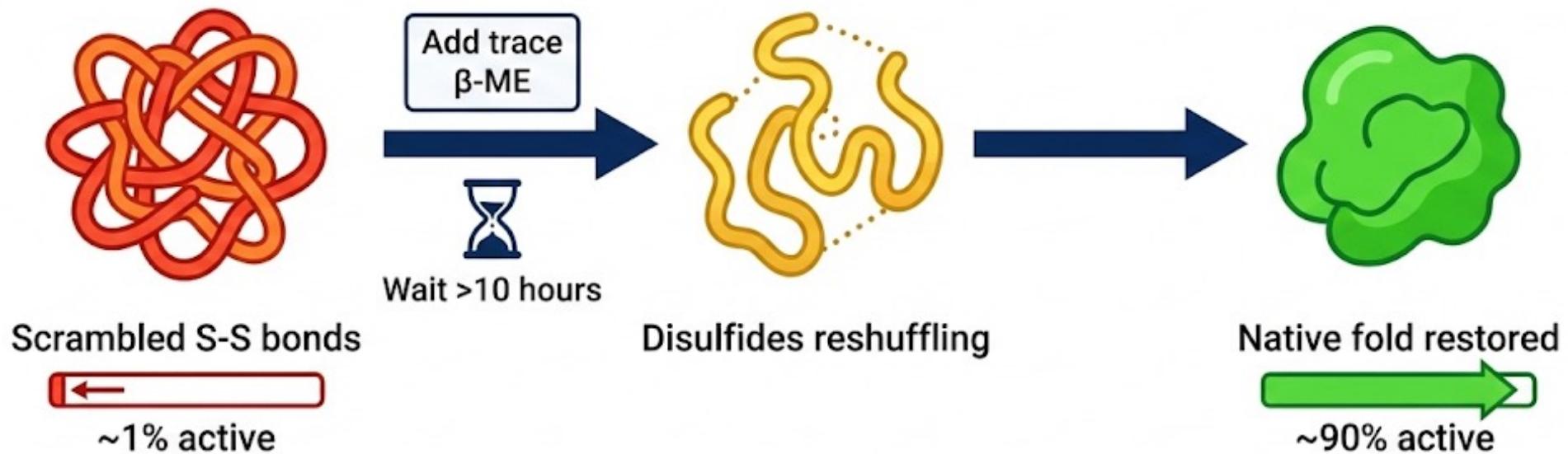
Anfinsen's Hypothesis

The information needed to specify the catalytically active three-dimensional structure of ribonuclease is contained in its amino acid sequence. In other words, the native structure is the thermodynamically most stable structure.

The dependence of conformation on sequence is especially significant because conformation determines function.



THE RESCUE—SEQUENCE FINDS THE RIGHT FOLD



THE SEQUENCE CONTAINS ALL INFORMATION NEEDED TO FOLD

Given the opportunity to find thermodynamic minimum, it will.

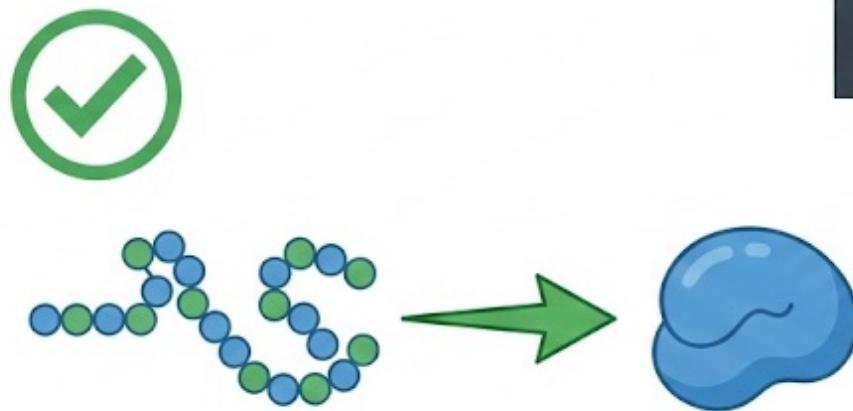
Anfinsen's Hypothesis

The information needed to specify the catalytically active three-dimensional structure of ribonuclease is contained in its amino acid sequence. In other words, the native structure is the thermodynamically most stable structure.

The dependence of conformation on sequence is especially significant because conformation determines function.

Anfinsen proved sequence contains the answer. But there's a problem...

WHAT ANFINSEN PROVED

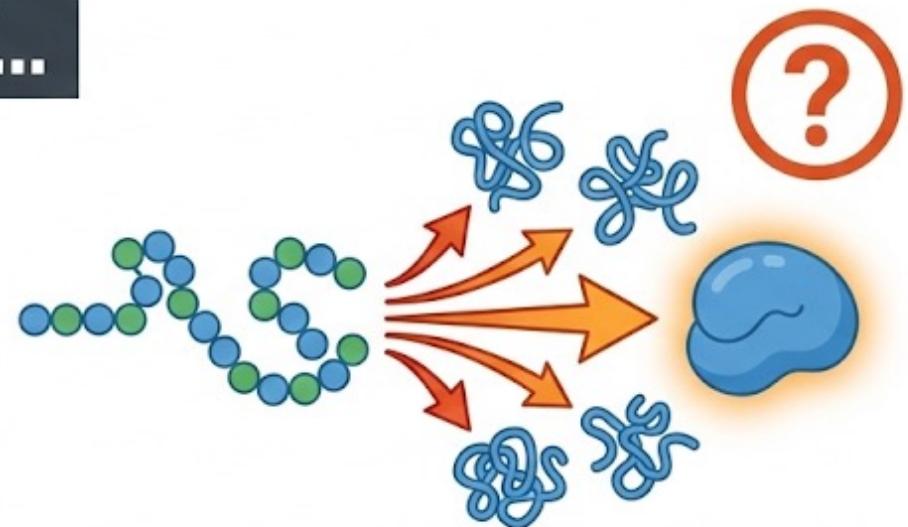


Sequence → Structure

All information is in the sequence

BUT...

THE PROBLEM THAT REMAINS



But HOW does it find the right one?

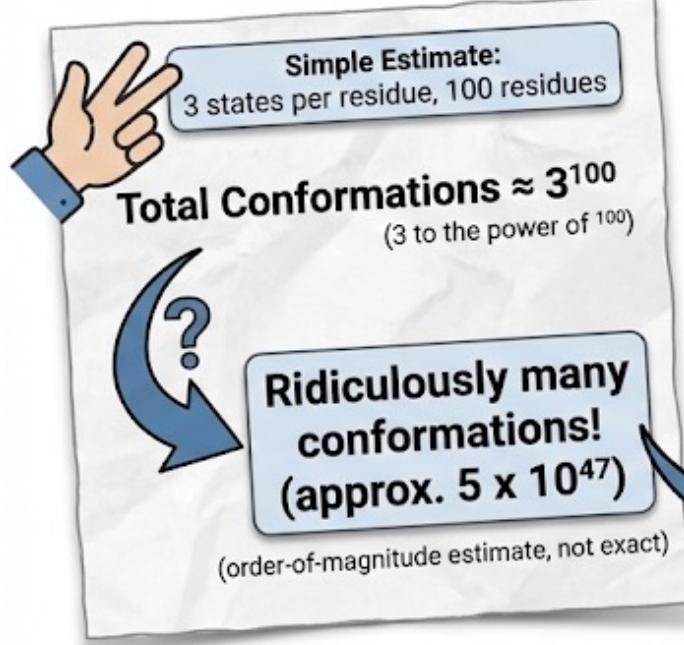
10^{100} Possible confirmations

Levinthal's paradox: folding cannot be random search

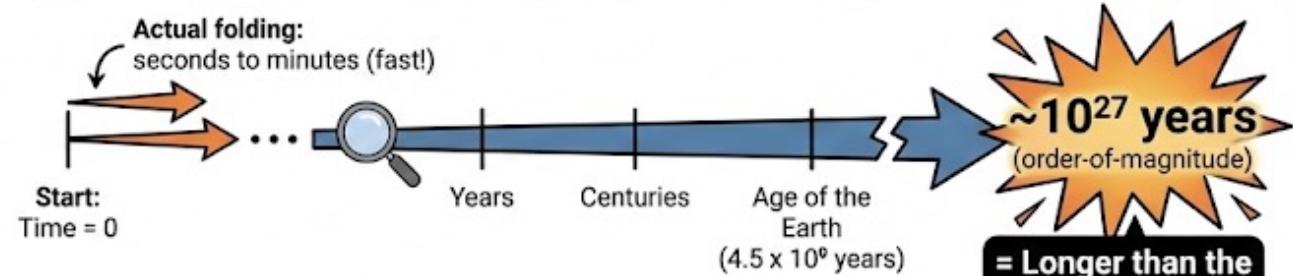
- Even a small protein has an astronomically large conformational space
- Randomly sampling conformations would take longer than the age of the universe
- Conclusion:** folding follows biased pathways and stabilizes intermediates



The Math: Conformational Space



The Time: Random Search (if every move takes a tiny fraction of a second)



Levinthal's paradox: Random search is impossible

Therefore: folding follows biased pathways and stabilizes intermediates



Proteins do not “try everything,” they “fall” downhill energetically.

Random search would take 10^{27} years
Solution is progressive stabilization

Proteins Fold by the Progressive Stabilization of Intermediates Rather Than by Random Search (1/2)

- A monkey randomly poking at a keyboard could type a sentence from Shakespeare in a few thousand keystrokes if the correct letters are retained, a process called cumulative selection.
- Protein folding also occurs by cumulative selection. Partly correct folding intermediates are retained because they are slightly more stable than unfolded regions.

Computer Simulations for Typing-Monkey Analogy

```
200 ?T(\G{+s x[A.N5~, #ATxSGpn`e@  
400 oDr'Jh7s DFR:W41'u+^v6zpJse0i  
600 e2ih'8zs n527x8l8d_ih=Hldseb.  
800 S#dh>)/s ]tZqC%P%DK<|!^aseZ.  
1000 v0th>nLs ut/isjl_kwojjwMasef.  
1200 juth+hvs it is[lukh?SCw=ase5.  
1400 Iithdn4s it is0l/ks/IxwLase~.  
1600 M?thinrs it is lXk?T" _woasel.  
1800 MSthinWs it is lwkN7OKw(asel.  
2000 Mhthin's it is likv,aww_asel.  
2200 MMthinn's it is lik+5avwlasel.  
2400 MethinXs it is likydaqw)asel.  
2600 Methin4s it is lik2dasweasel.  
2800 MethinHs it is like@Tweasel.  
2883 Methinks it is like a weasel.
```

```
200 )z~hg)W4{{cu!kO{d6jS!N1EyUx}p  
400 "W hi\kR.<&CfA%4-Y1G!iT$6({|6  
600 .L=hinkm4(uMGP^IAWoE6klwW=yiS  
800 AthinkapPa_vYH liR\Hb,Uo4\-(  
1000 OFthinksP)@fZO li8v] /+Eln26B  
1200 6ithinksMvt -V likm+g1#K~)BFk  
1400 vxthinksEt @w like.S1Geutks.  
1600 :0thinks<it MC likesN2[eaVe4.  
1800 uxthinksqt Or likeQh)weaoew.  
2000 Y/thinks it id like7alwea)e&.  
2200 Methinks it iw like a[weaWel.  
2400 Methinks it is like a;weasel.  
2431 Methinks it is like a weasel.
```

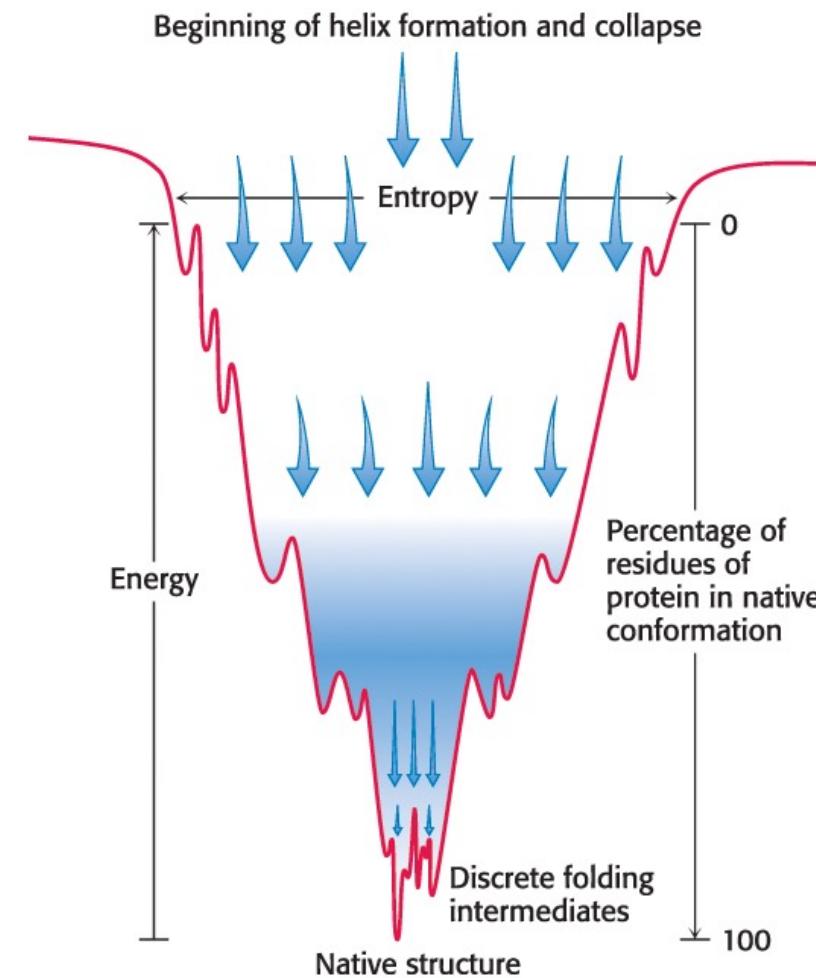
Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019
W. H. Freeman and Company

Thermodynamics of Protein Folding

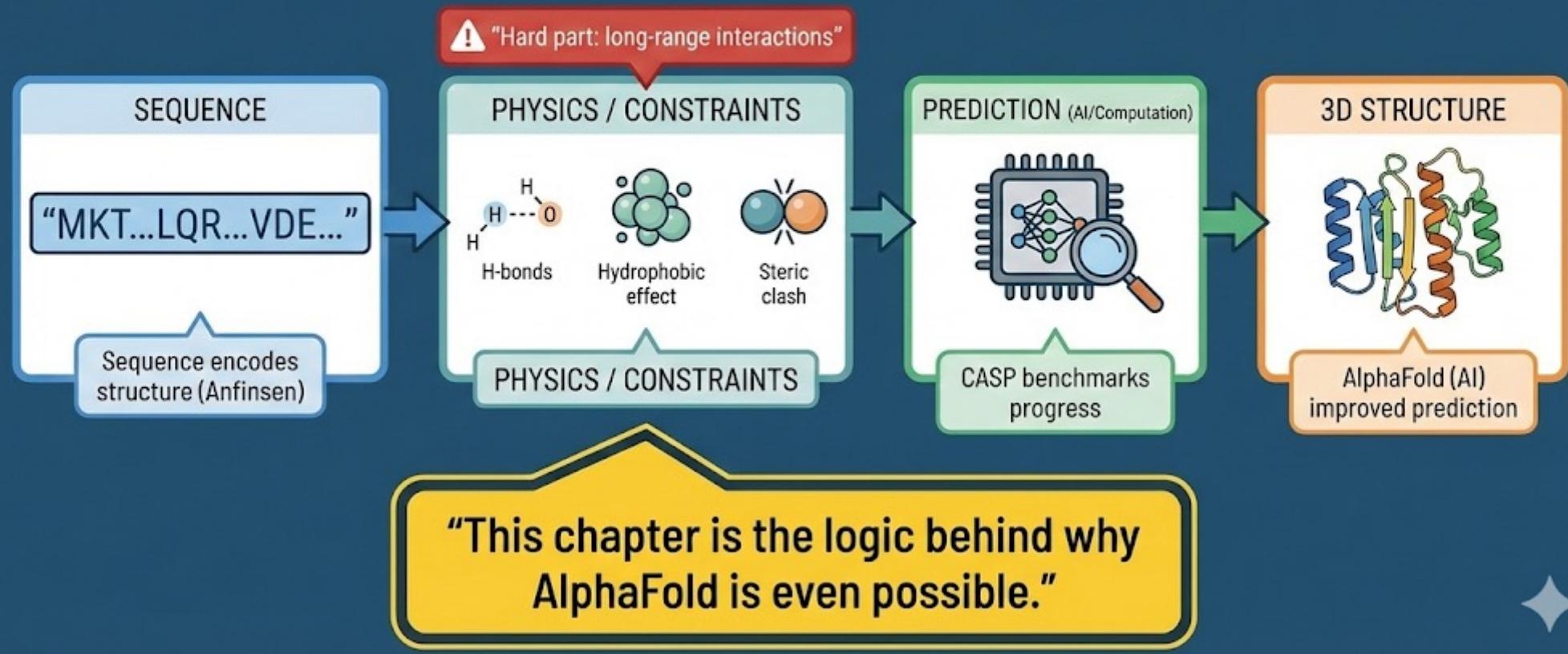
The folding of proteins is sometimes visualized as a folding funnel, or energy landscape.

- The breadth of the funnel represents all possible conformations of the unfolded protein.
- The depth of the funnel represents the energy difference between the unfolded and the native protein.
- Each point on the surface represents a possible three-dimensional structure and its energy value.

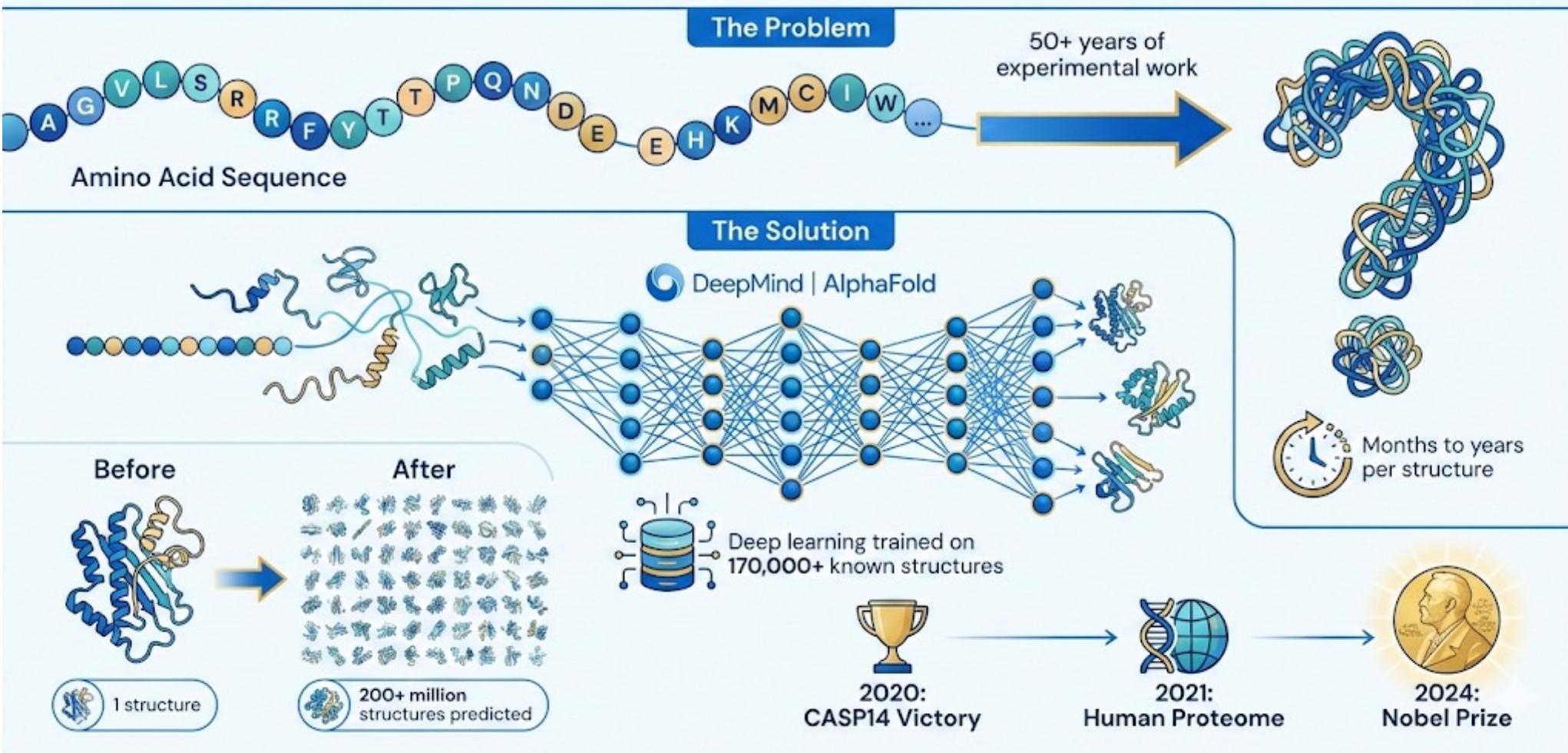
The funnel suggests that there are alternative pathways to the native, or most stable, structure.



Predicting structure from sequence: why it's hard, why it's changing



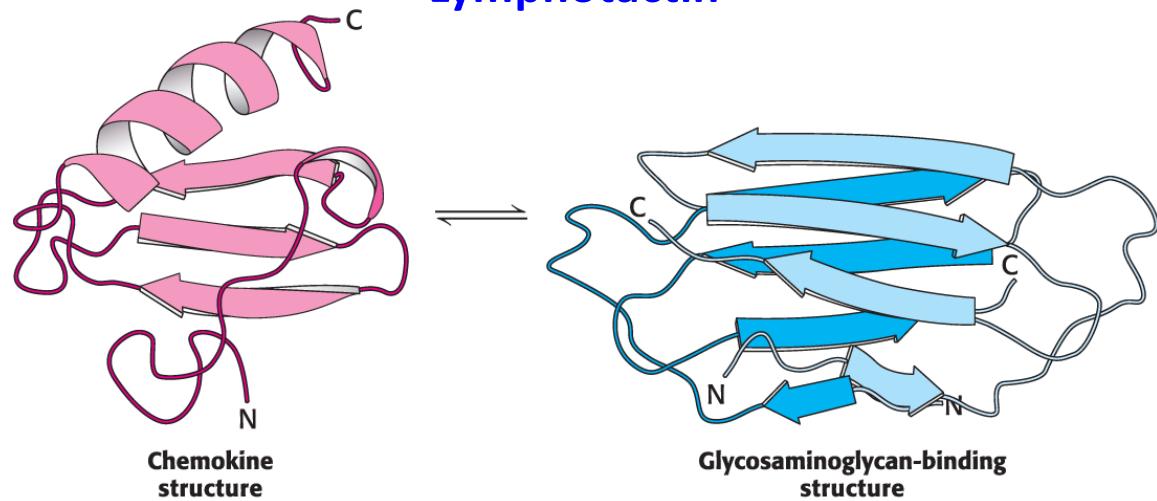
AlphaFold: From Sequence to Structure in Minutes



Some Proteins Are Intrinsically Disordered and Can Exist in Multiple Conformations

- Intrinsically disordered proteins (IDPs) do not have a defined structure under physiological conditions until they interact with other molecules.
- Metamorphic proteins exist in an ensemble of structures of approximately equal energies that are in equilibrium.

Structures of Two Conformations of Lymphotactin



Tymoczko et al., *Biochemistry: A Short Course*, 4e, © 2019 W. H. Freeman and Company

The same sequence be found in shifting conformations

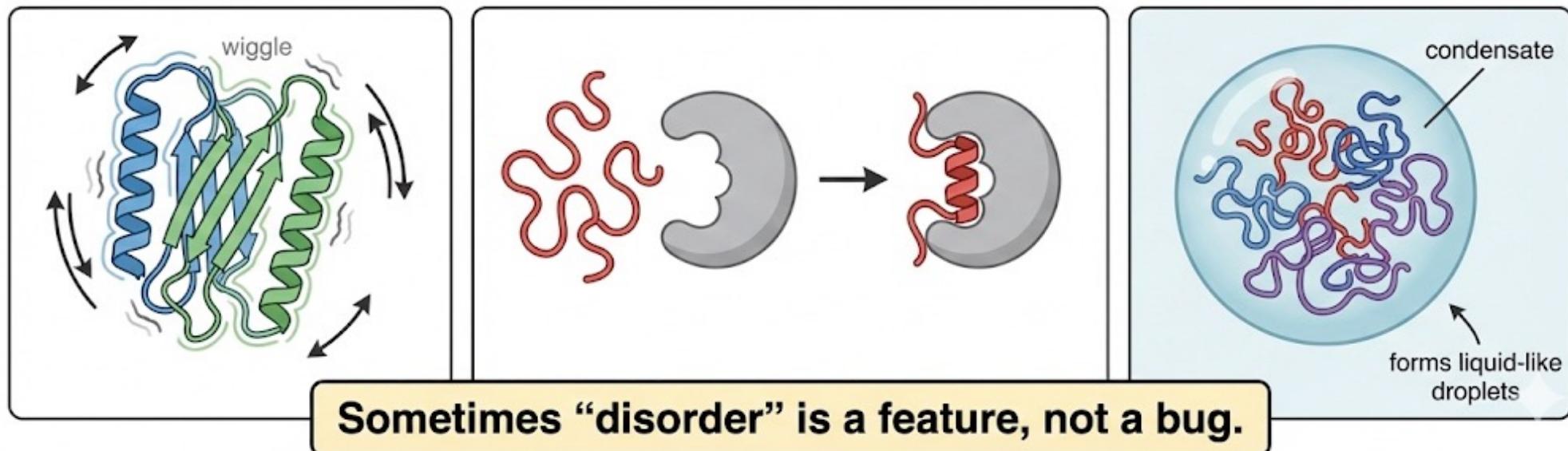
IDPs in real life: dynamics and phase separation

- Proteins are dynamic (structures fluctuate, “breathing”)
- IDPs often function in signaling/regulation via binding-induced folding
- Many IDPs participate in liquid phase separation to form condensates
- Same sequence can support multiple functional states depending on partners

Protein dynamics
('breathing')

Binding-induced folding

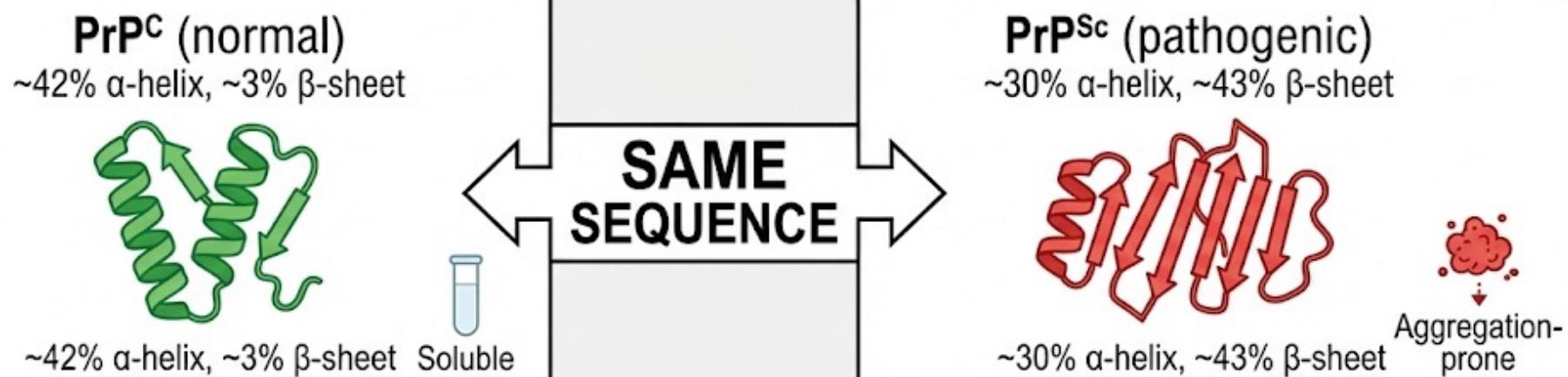
Liquid phase separation
(condensates)



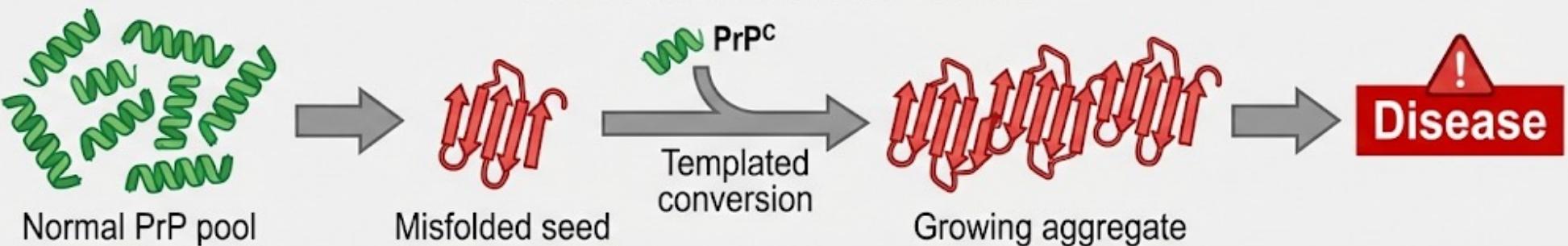
Clinical Insight: Protein Misfolding and Aggregation Are Associated with Some Neurological Diseases

- Some infectious neurological diseases are caused by infectious proteins called prions. Prions exist in two states, one α -helix-rich (PrP^{C}) and the other β -sheet-rich (PrP^{Sc}).
- PrP^{Sc} forms aggregates that disrupt cell function.

SIMPLIFIED PRION MECHANISM



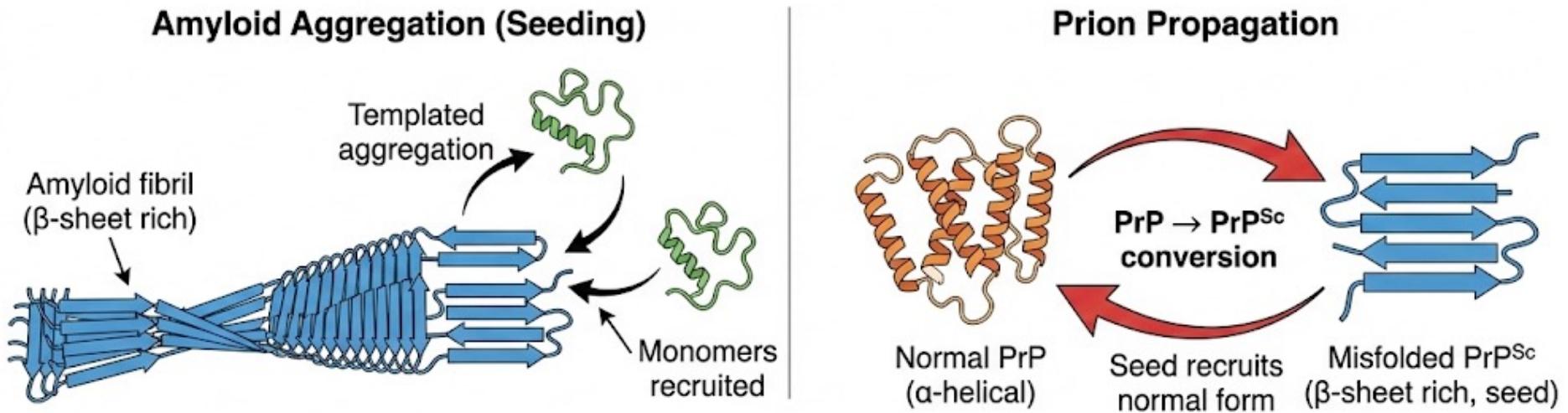
CONVERSION MECHANISM



The misfolded form TEMPLATES conversion of normal protein
This is why prions are 'infectious' without DNA/RNA

Misfolding can propagate: amyloids and prions

- ✓ Many amyloids are β -sheet rich fibrils that template further aggregation
- ✓ Alzheimer's involves aggregation of A β into plaques (big-picture)
- ✓ Prions: misfolded PrP can seed conversion of normal PrP
- ✓ Named transmissible spongiform encephalopathies: CJD, mad cow, scrapie



Aggregation spreads by seeding: “bad fold recruits good fold.”

- CJD • mad cow • scrapie

In closing: Proteins don't choose to fold correctly. They have no choice. The sequence encodes constraints, and constraints eliminate every wrong fold. When that system fails—when one residue changes, or one kinetic trap opens—you get sickle cell, you get Alzheimer's, you get prions. Constraints create life. And when constraints fail, disease follows.

Prion Diseases: When Proteins Become Infectious



SPORADIC (largest spoke, ~85%)

Creutzfeldt-Jakob Disease (sCJD)

- 1 in 1 million/year, average age 60+
- Rapid dementia, death within 1 year



ANIMAL PRIONS

Scrapie (sheep), BSE (cattle), CWD (deer)

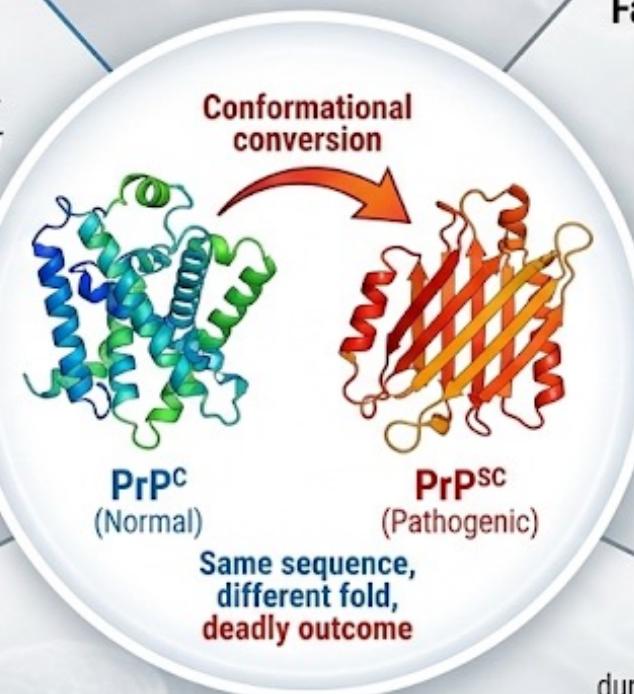
- Species barrier but can be crossed



ACQUIRED - VARIANT

vCJD (Mad Cow related)

- BSE → humans, UK 1990s
- Younger patients, longer course



INHERITED (~10-15%)

Fatal Familial Insomnia (FFI) Gerstmann-Sträussler-Scheinker (GSS)



ACQUIRED - DIETARY

Kuru

- Ritualistic cannibalism - Fore people
- Discovered by Gajdusek (Nobel 1976)



ACQUIRED - IATROGENIC

iCJD

- Contaminated growth hormone, dura mater grafts, surgical instruments



KEY CHARACTERISTICS



No immune response



No nucleic acid



Cannot be sterilized
by autoclaving



Incubation:
years to decades