I. Protein Structure

Lecture 9 (9/29/25)

A. Primary
   1. Peptide Bond
      a. Planar, strong, ϕ/ψ angles
   2. Determination
      a. Sequence determination; CHEMICAL
         i. aa composition; Divide & conquer; Edman degradation
      b. Sequence determination; PHYSICAL
         i. Tandem Mass Spectrometry for proteins
      c. Sequence determination; BIOLOGICAL
         i. Genome sequenced; need partial sequence
      d. Determination of Disulfide bonds

B. Secondary
   1. Conformational structure; Levinthal paradox
   2. Pauling & Corey's predictions
      a. α-Helix
      b. β-sheets/strands
      c. Connections between β-strands
      d. Connections between α-helices; angle not important
   3. Super secondary structure

C. Tertiary
   1. Picturing and classifications
   2. Topology
   3. Domains
   4. Intrinsically disordered
   5. Stability

- Reading: Ch4; 119-122, 125-127,131-133
  Ch4; 114-115, 120-121, 123-124
- Homework #9

NEXT
- Reading: Ch4; 116-117, 126-128
- Homework #10

Determination of primary structure

THREE basic ways to know the primary structure:

CHEMICAL Edman Degradation requires >100 pmole (1-5 µg)
PHYSICAL MS/MS requires >1-10 pmole (100-500 ng)
BIOINFOMATICAL

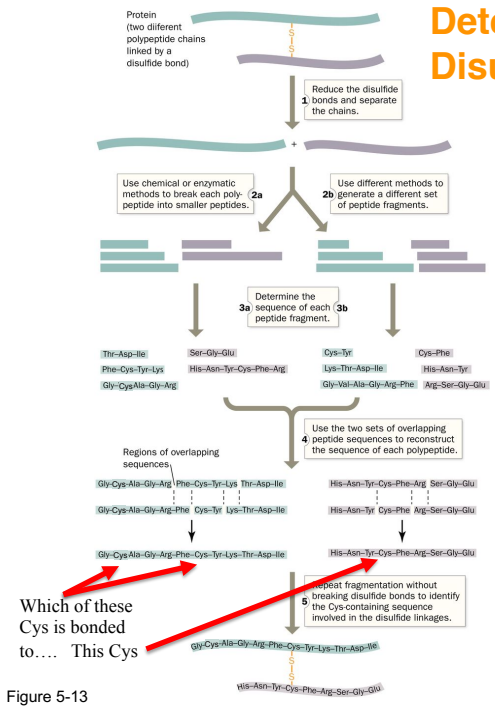
We just went through the CHEMICAL and PHYSICAL.
The BIOINFORMATICAL method requires information from chemical or physical, but only a limited amount of sequence.

- Example: a sequence of 6 AA is only possible as one of 20^6 possible hexa-peptide sequences (1 of 64×10^6).
- There are no more than 50,000 protein-coding genes with ≤400 AA on average. This is ~20×10^6 possible unique sequences.
- So, a hexamer is not likely to appear more than once.
- Once you have at least 6 AA sequence, you can compare that to all possible proteins encoded in the entirety of the gene sequences (genome) for a species for which the genome is known using appropriate bioinformatic tools. This will then give you the entire protein sequence.

There is one remaining issue: Where are the Disulfides, if any?

…..This requires chemical and/or physical methods

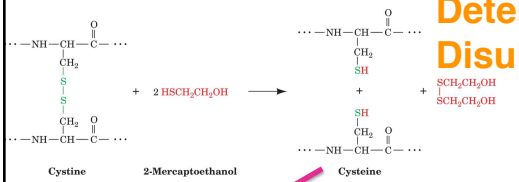
Determine which Cys are in Disulfide bonds



Which of these Cys is bonded to…. This Cys

Figure 5-13

Determine which Cys are in Disulfide bonds

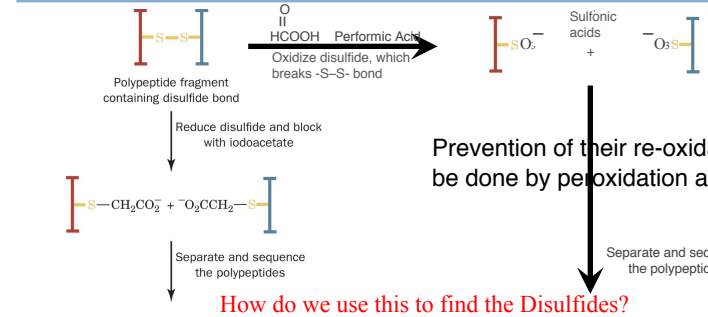


Recall that to separate polypeptides for sequence analysis, the disruption of –S-S- bonds is needed

And, that prevention of their re-oxidation can be done by alkylation or oxidation.

$$\text{Cys}-\text{CH}_2-\text{SH} + \text{ICH}_2\text{COO}^- \longrightarrow \text{Cys}-\text{CH}_2-\text{S}-\text{CH}_2\text{COO}^- + \text{HI}$$

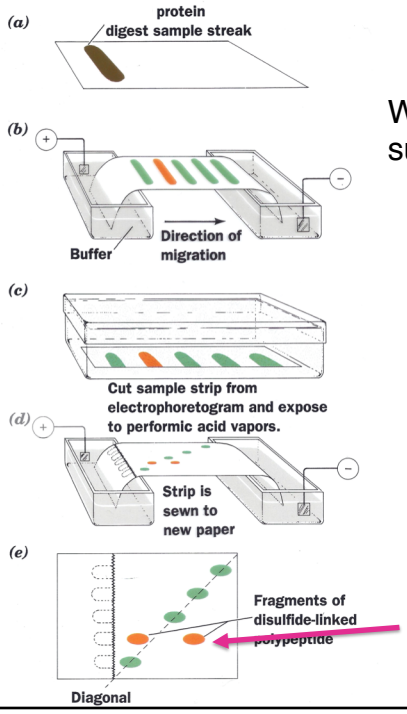
Cysteine **Iodoacetate** ***S*-Carboxymethylcysteine (CM-Cys)**



Prevention of their re-oxidation can be done by peroxidation as well.

Separate and sequence the polypeptides

How do we use this to find the Disulfides?

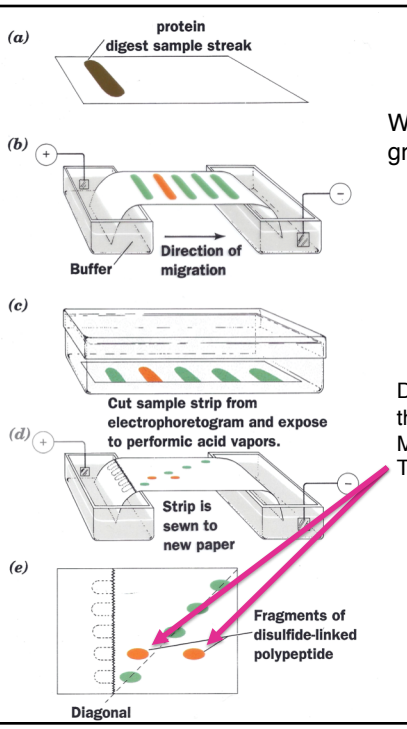


Determine which Cys are in Disulfide bonds

We change the protection step for sulfhydryl groups.

- Cleave/Protect (reduction/alkylation or oxidation) AFTER fragmentation
- Separate fragments as before, but any linked by –S–S- will not separate and remain together (e.g., orange peptide).
- THEN break –S–S- bonds, and re-separate.

Determine the sequence of those peptides that fall off the diagonal by either Edman degradation or tandem MS/MS.
Technique is called ”2D-diagonal electrophoresis.”

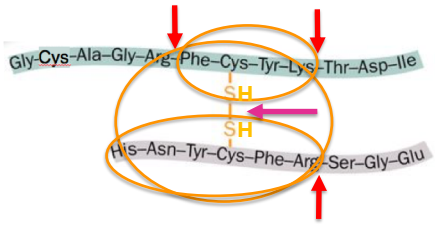


Determine which Cys are in Disulfide bonds

We change the protection step for sulfhydryl groups.

- Cleave/Protect (reduction/alkylation or oxidation) AFTER fragmentation
- Separate fragments as before, but any linked by –S–S- will not separate and remain together (e.g., orange peptide).
- THEN break –S–S- bonds, and re-separate.

Determine the sequence of those peptides that fall off the diagonal by either Edman degradation or tandem MS/MS.
Technique is called ”2D-diagonal electrophoresis.”



Protein Structure

Conformational Structure
How does the polypeptide chain fold?

Protein Structure

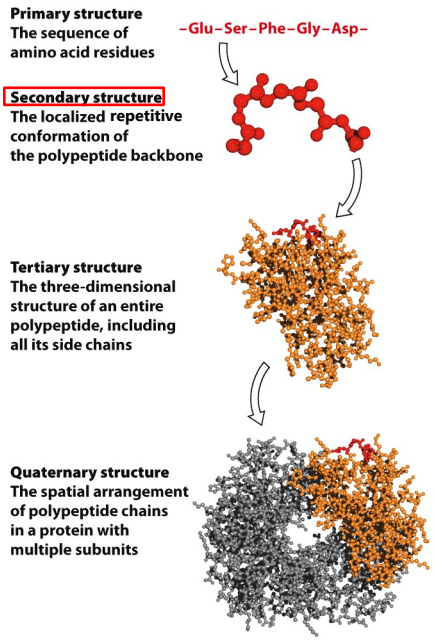
Conformational Structure
How does the polypeptide chain fold?

1) **primary structure** – sequence of amino acids
2) **secondary structure** – small units of repetitive structure
3) **tertiary structure** – overall 3D shape
4) **quaternary structure** – shape of ≥2 chains

4 levels of protein structure

In order to understand these levels of structure, you need to understand the nature of the polymer first.

In other words, the linkage or PEPTIDE BOND



Protein Structure-Secondary

The 4 S's for secondary structure:

Size	-dependent on number of amino acids
Solubility	-dependent on AA composition and shape
Stability	-complex and not well understood
Shape	

Why is there Secondary Structure?

Protein Structure-Secondary

The Levinthal Paradox (1969):

Theoretical calculation:

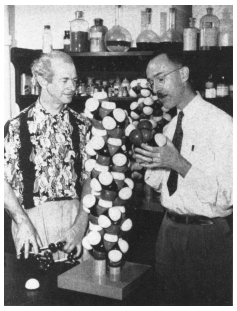
- Consider just the α-carbon backbone…..
- If there are 4 clearly different angles allowed of all the angles at the α-carbon (ϕ and ψ), then each residue has 2x4=8 degrees of freedom.
- For a protein of 100 residues, there are 8^{100} possible conformations to "test" for optimal energetics
- $8^{100} = 2 \times 10^{90}$ different conformations
- At 1000 billion "tests" per second (1/psec), this is 2×10^{78} seconds to find the best.
- $\Longmapsto 7 \times 10^{70}$ years
- Well …... The age of the universe is 14×10^{9} years

The shortcut proteins use to fold is the use of 2° structure where most of these degrees of freedom are prescribed by a regular structure.

What are these "regular structures?"

Secondary Structure

Protein Structure-Secondary



In the early 1950's, Linus Pauling and Robert Corey predicted some rules that proteins should follow to find the lowest energy conformation.

1) The peptide bond must be planar without free rotation
2) The degree of H-bonding should be maximized to achieve the lowest energy state
3) The best H-bonds are linear
4) There should be repeating units of conformation (same) as you go from one residue to the next

[consider energetic consequences in the $(\text{unfolded})^{\text{water}} \rightleftharpoons (\text{folded})^{\text{water}}$ transition]

Using these rules they predicted two basic structures:

α-helix

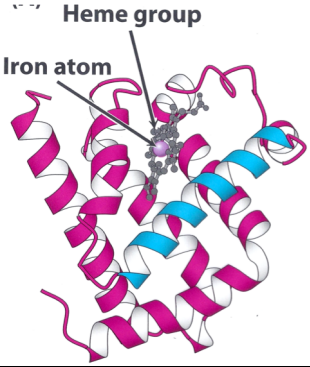
β-sheet

Protein Structure-Secondary

John Cowdery Kendrew (1917 - 1997)

There were no known protein structures until 1957, when Kendrew solved the structure of myoglobin:

Imagine the excitement when indeed there were the very α-helices Pauling predicted!

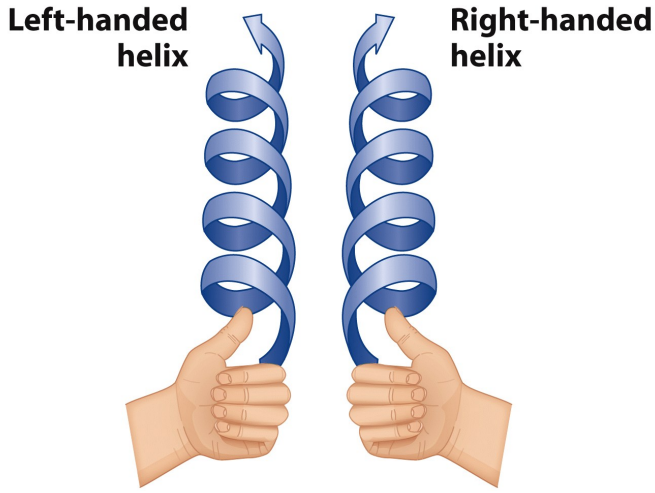


Protein Structure-Secondary

The α Helix

1

Right handed



Protein Structure-Secondary

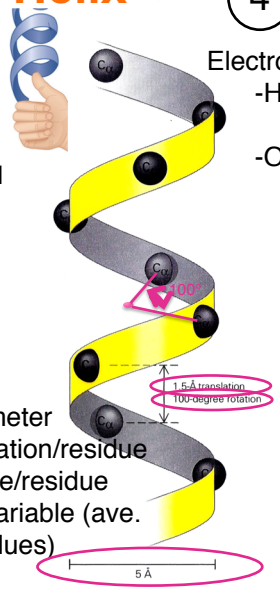
The α Helix

1

Right handed

2

Dimensions
-5 Å diameter
-100° rotation/residue
-1.5 Å rise/residue
-length variable (ave. ~45Å (8 residues)



3

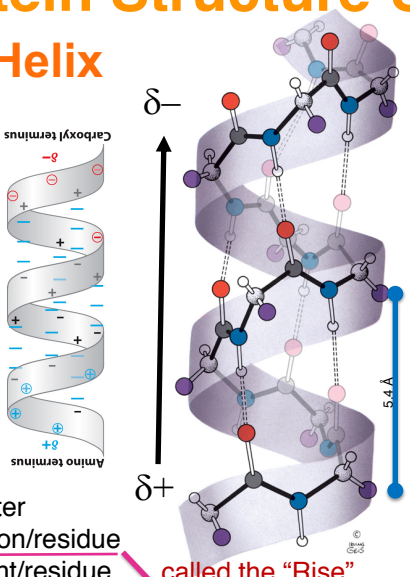
Conformational angles
- $\phi = -57°$
- $\psi = -47°$

4

Electrostatics
-H-bonds all satisfied (as predicted)
-Overall polarity of helix

Protein Structure-Secondary

The α Helix



1

Right handed

2

Dimensions
- -5 Å diameter
- -100° rotation/residue
- -1.5 Å height/residue ….called the "Rise"
- -length variable (ave ~45Å (8 residues)

3.6 residues/turn
5.4 Å/turn ….called the "Pitch"

3

Conformational angles
- ϕ = – 57°
- ψ = – 47°

4

Electrostatics
- -H-bonds all satisfied (as predicted)
- -Overall polarity of helix

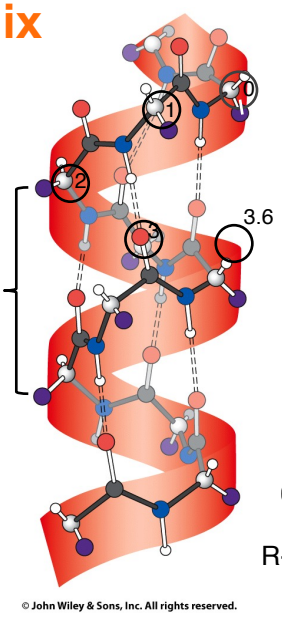
Protein Structure-Secondary

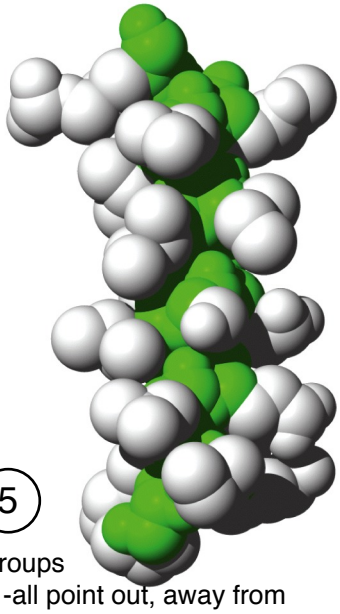
The α Helix

3.6 residues per turn (repeat)

5.4 Å per turn (repeat)

"Pitch"



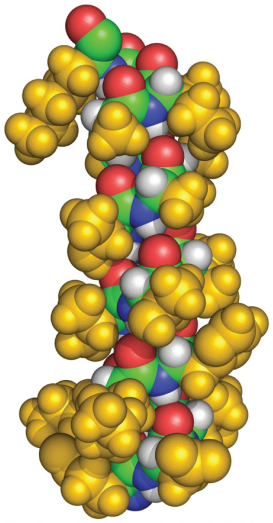


5

R-groups
- -all point out, away from the helix

Protein Structure-Secondary

The α Helix: Space Filling Model



Based on an X-ray structure by Ilme Schlichting, Max Planck Institut für Molekulare Physiologie, Dortmund, Germany. PDBid 1A6M (for the definition of "PDBid" see Section 6-2E).

Helices can be hydrophobic

Helices can be hydrophilic

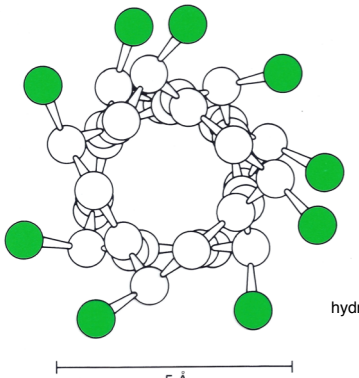
Helices can be mixed: amphipathic

Oxy-Myoglobin
PDBid 1A6M

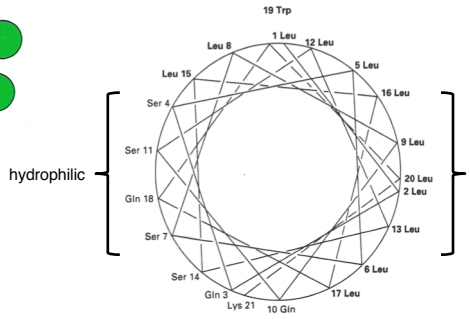
Protein Structure-Secondary

The α Helix

Helices can be mixed: amphipathic

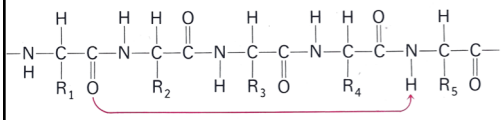


LLQSLLSLLQSLLSLLLQW



hydrophilic

hydrophobic



Protein Structure-Secondary

The α Helix: Propensities

Table 6-1 Propensities of Amino Acid Residues for α Helical Conformations

Residue	P_α	
Ala	1.42	⇦
Arg	0.98	
Asn	0.67	
Asp	1.01	
Cys	0.70	
Gln	1.11	⇦
Glu	1.51	⇦
Gly	0.57	
His	1.00	
Ile	1.08	
Leu	1.21	⇦
Lys	1.16	⇦
Met	1.45	⇦
Phe	1.13	⇦
Pro	0.57	
Ser	0.77	
Thr	0.83	
Trp	1.08	
Tyr	0.69	
Val	1.06	

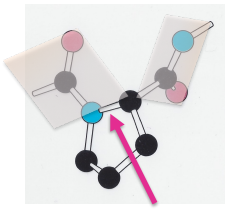
Source: Chou, P.Y. and Fasman, G.D., *Annu. Rev. Biochem.* **47,** 258 (1978).

Like:

-prefer small, medium, hydrophobic/charged
-no steric hindrance at C_β
-Glu, Met, Ala, Leu, Lys, Phe, Gln

Don't Like:

-Pro
-Gly



Pro has fixed ϕ = -65°

Gly is opposite: has too many degrees of freedom

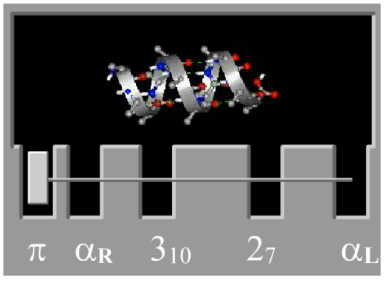
22

Protein Structure-Secondary

The α Helix: other helices

- The α-helix is a common secondary structure in proteins.
- The α-helix is a spiral structure with defined dimensions, rise and pitch.
- The α-helix is stabilized by hydrogen bonds between residues in the helix.
- Several types of helices (such as the π, α_R, 3_{10}, 2.2_7 and α_L) can form intramolecular hydrogen bonds.

Increasing pitch (distance per turn) & rise (distance per residue) →



4.1_{16} 3.0_{10}

Repeat units/turn → 3.6_{13} 2.2_7

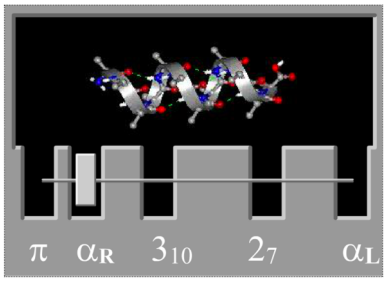
# of atoms in one turn →

Protein Structure-Secondary

The α Helix: other helices

Increasing pitch (distance per turn) & rise (distance per residue) →

- The α-helix is a common secondary structure in proteins.
- The α-helix is a spiral structure with defined dimensions, rise and pitch.
- The α-helix is stabilized by hydrogen bonds between residues in the helix.
- Several types of helices (such as the π, α_R, 3_{10}, 2.2_7 and α_L) can form intramolecular hydrogen bonds.



4.1_{16} 3.0_{10}

Repeat units/turn → 3.6_{13} 2.2_7

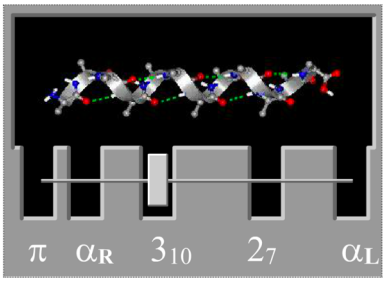
# of atoms in one turn →

Protein Structure-Secondary

The α Helix: other helices

Increasing pitch (distance per turn) & rise (distance per residue) →

- The α-helix is a common secondary structure in proteins.
- The α-helix is a spiral structure with defined dimensions, rise and pitch.
- The α-helix is stabilized by hydrogen bonds between residues in the helix.
- Several types of helices (such as the π, α_R, 3_{10}, 2.2_7 and α_L) can form intramolecular hydrogen bonds.



4.1_{16} 3.0_{10}

Repeat units/turn → 3.6_{13} 2.2_7

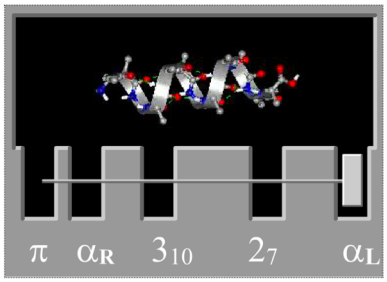
# of atoms in one turn →

Protein Structure-Secondary

The α Helix: other helices

- The α-helix is a common secondary structure in proteins.
- The α-helix is a spiral structure with defined dimensions, rise and pitch.
- The α-helix is stabilized by hydrogen bonds between residues in the helix.
- Several types of helices (such as the π, α_R, 3_{10}, 2.2_7 and α_L) can form intramolecular hydrogen bonds.

Increasing pitch (distance per turn) & rise (distance per residue) →



4.1_{16} 3.0_{10}

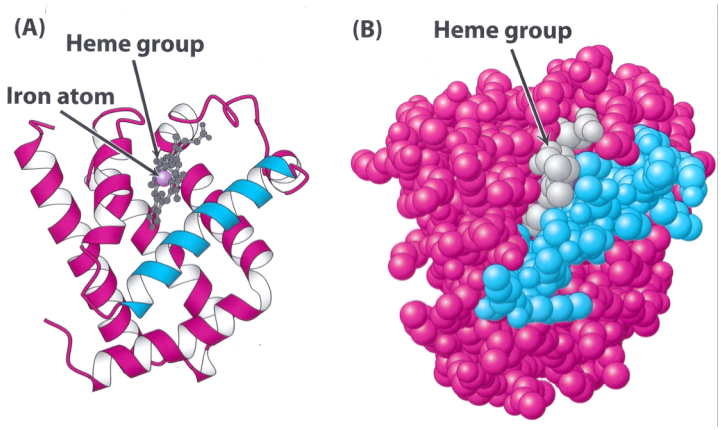
Repeat units/turn → 3.6_{13} 2.2_7

# of atoms in one turn →

Protein Structure-Secondary

The α Helix: examples

Globular Proteins



Myoglobin (Mb)

Protein Structure-Secondary

The α Helix: examples

Fibrous Proteins: the α-keratin helix

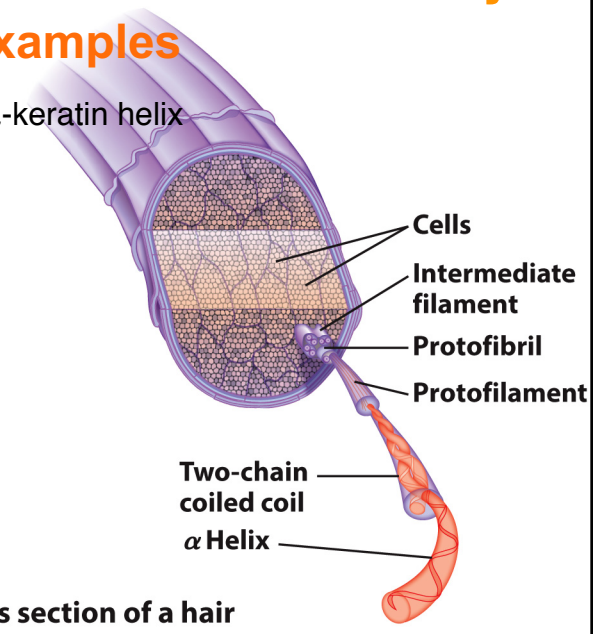
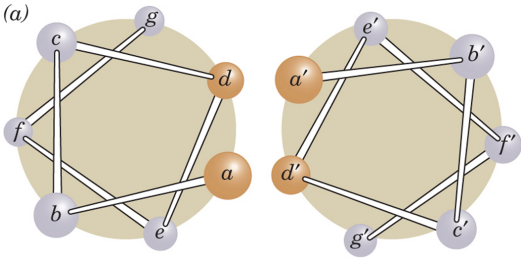


Figure 4-11b

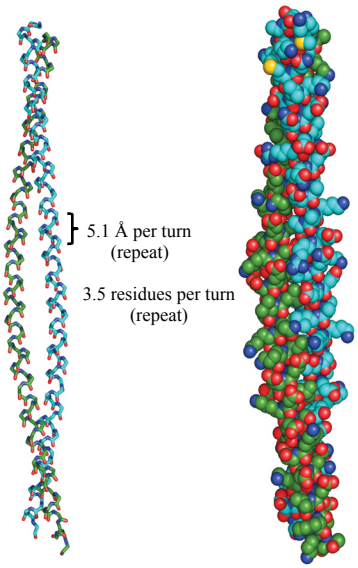
Protein Structure-Secondary

The α Helix: examples (b)

Fibrous Proteins: the α-keratin helix



Based on an X-ray structure by Carolyn Cohen, Brandeis University. PDBid 1IC2.



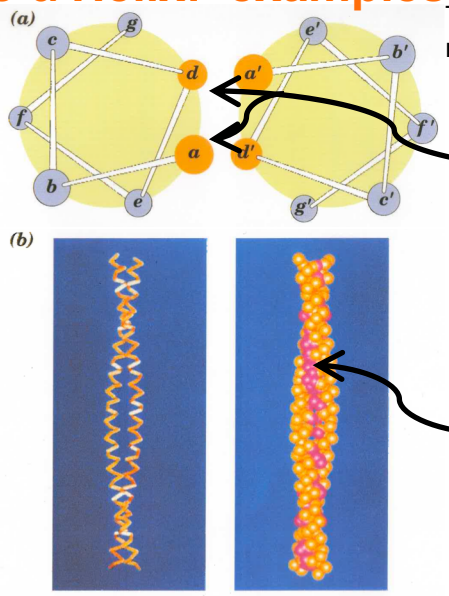
Based on an X-ray structure by Carolyn Cohen, Brandeis University. PDBid 1IC2.

A Coiled Coil

Protein Structure-Secondary

The α Helix: examples

α-keratin packing



Two α-helices are coiled together to make a coiled-coil, which is straight

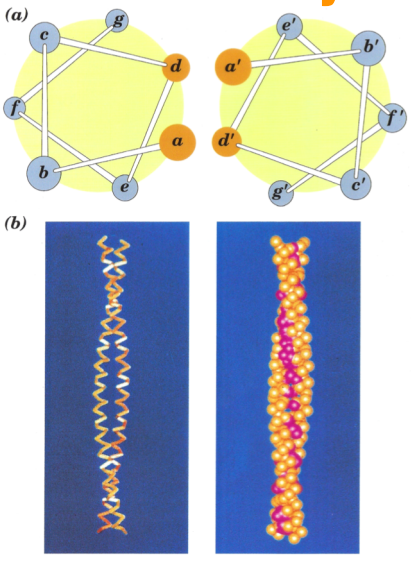
Residues at interface are hydrophobic to stabilize the coiled coil.

Residues at interface (pink) must be small to accommodate the packing

Protein Structure-Secondary

The α Helix: examples

Fibrous Proteins: the α-keratin helix



a-b-c-d-e-f-g

hϕ hϕ +/– +/–

(L,V,A,G) (E,D,R)

b-c-&-f tend to be Ser, Gln, Cys

A Coiled Coil

Protein Structure-Secondary

Amino Acid	α-Keratin (Wool)
Gly	8.1
Ala	5.0
Ser	10.2
[illegible] + Gln	12.1
Cys	11.2
Pro	7.5
Arg	7.2
Leu	6.9
Thr	6.5
[illegible]	6.0
Val	5.1
Tyr	4.2
Ile	2.8
Phe	2.5
Lys	2.3
Trp	1.2
His	0.7
Met	0.5

Note: The three most abundant amino acids in each protein are ~~indicated in red~~ boxed. Values given are in mole percent.

Glu

a&d

~25-30%

b,c,&f

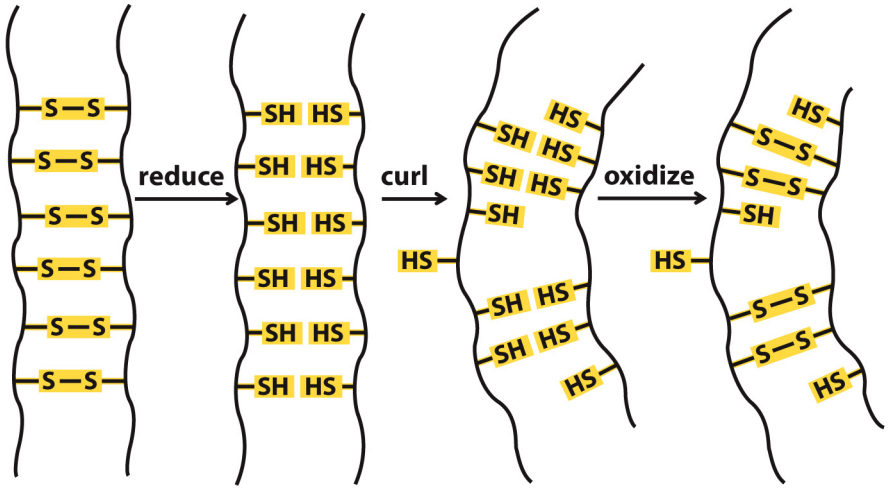
~33%

e&g

~15-20%

Notice the >10% Cys

Chemistry of Permanent Waving

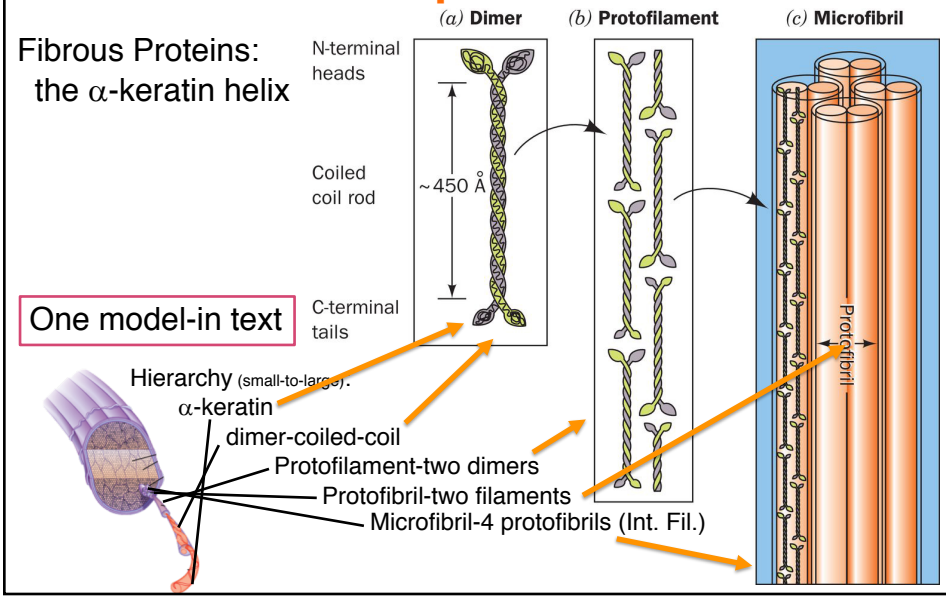


Box 4-2
Lehninger Principles of Biochemistry, Seventh Edition

Protein Structure-Secondary

The α Helix: examples

Fibrous Proteins:
the α-keratin helix



One model-in text

Hierarchy (small-to-large):
α-keratin
dimer-coiled-coil
Protofilament-two dimers
Protofibril-two filaments
Microfibril-4 protofibrils (Int. Fil.)

Protein Structure-Secondary

The α Helix: examples

Fibrous Proteins:
the α-keratin helix

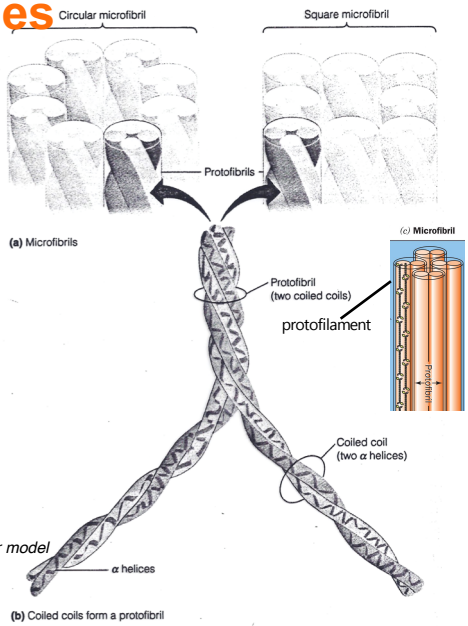


FIGURE 6.11
The coiled-coil structure of α-keratin. **(a)** Structure of microfibrils formed by two different modes of interaction of protofibrils. **(b)** Formation of a protofibril from two coiled coils.

Another model

Hierarchy:
α-keratin
dimer-coiled-coil
Protofibril-coiled two dimers
(same as Protofilament), Protofibril in other model is 2 protofibrils in this one)
Microfibril-8 protofibrils

Protein Structure-Secondary

The α Helix: Summary

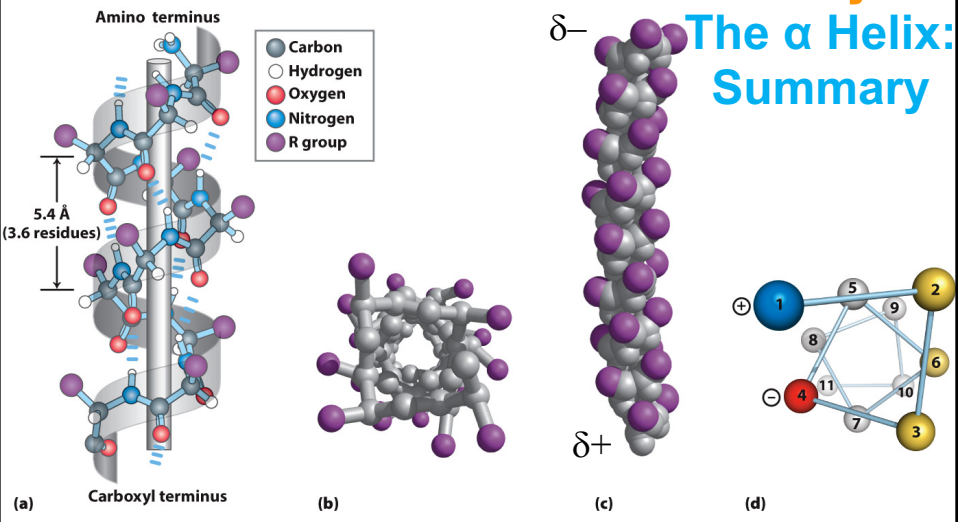


Figure 4-4

- Right handed helix
- 3.6 aa per turn
- 5.4 Å rise per turn

- Carbonyl of residue "n" is H-bonded to NH of n+4 residue
- Has tightly packed core of main-chain atoms
- R-groups project outward
- Has overall macro-dipole (N-term +; C-term –)
- Can be amphipathic

Secondary Structure

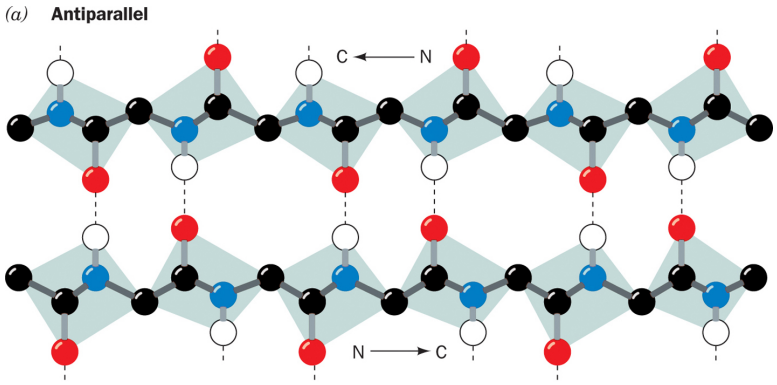
Protein Structure-Secondary

β Sheets- antiparallel

Using his rules, Pauling predicted two basic structures:

α-helix

β-sheet, which he called a "back-and-forth" structure

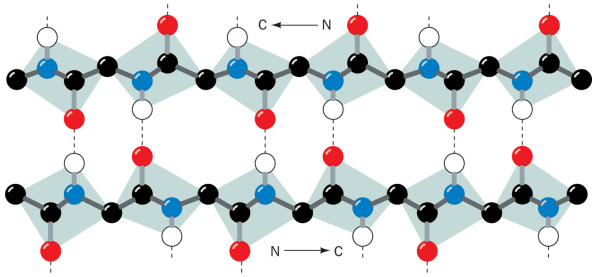


Protein Structure-Secondary

β Sheets- antiparallel

1

Right handed
(2.0_7)



2

Dimensions
- it's a sheet

Almost fully extended:
- 3.4 Å rise/residue
- 2 residues/repeat → Pitch is 6.8 Å
- length variable

3

Conformational angles
- ϕ = – 139°
- ψ = + 135°

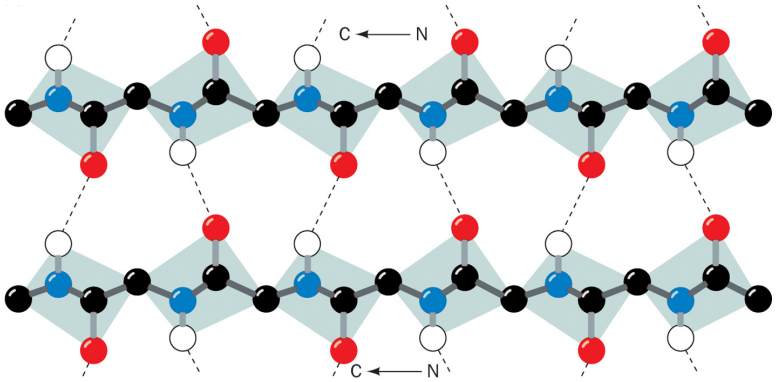
4

Electrostatics
- H-bonds all satisfied (as predicted)
- no polarity

Protein Structure-Secondary

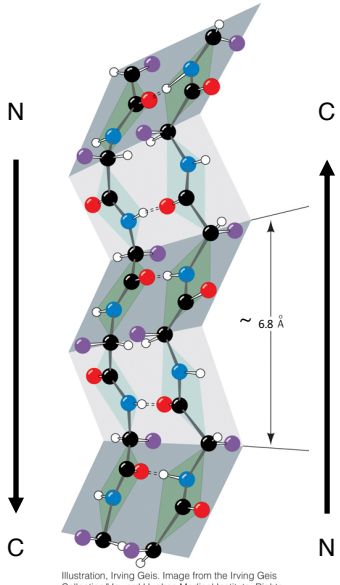
β Sheets- parallel

Pauling did not predict a β-sheet made of β-strands going the same direction.



Protein Structure-Secondary

β Sheets

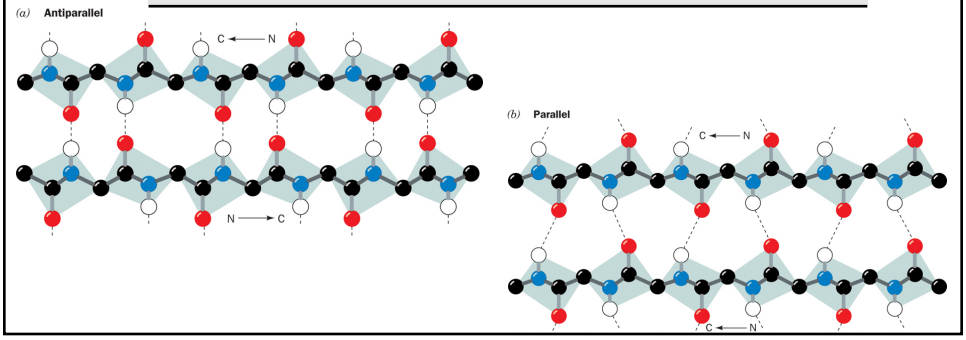


Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Protein Structure-Secondary

β Sheets

Structure	ϕ	ψ	Rise (Dist/residue) (Å)	Residues/ Repeat	Pitch (Distance/ repeat) (Å)	Diameter (Å)
α-helix	-57	-47	1.5	3.6	5.4	5.0
Anti- ⇌ β-sheet	-139	+135	3.4	2	6.8	-
Parallel ⇒ β-sheet	-119	+113	3.2	2	6.4	-



Protein Structure-Secondary

β Sheets

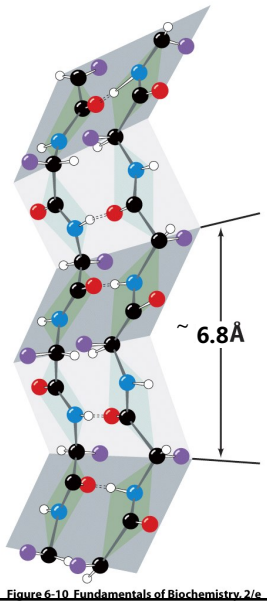


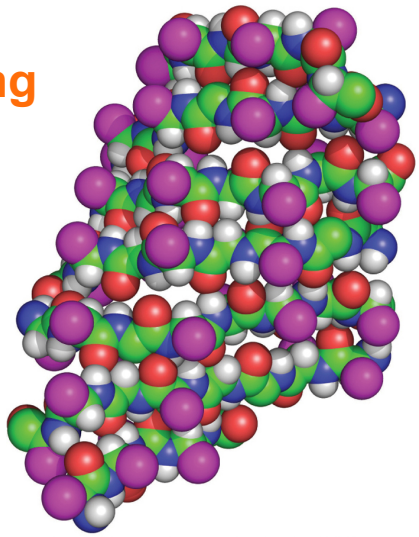
Figure 6-10 Fundamentals of Biochemistry, 2/e

5

R-groups
-alternate and point away from sheet

Protein Structure-Secondary

β Sheet: Space Filling Model



Based on an X-ray structure by Gerald Edelman, The Rockefeller University. PDBid 2CNA.

Concanavalin A
PDBid 2CNA

Protein Structure-Secondary

The β Sheet: Propensities

Table 6-1 Propensities of Amino Acid and β Sheet Conformations

Residue	P_β	
Ala	0.83	
Arg	0.93	
Asn	0.89	
Asp	0.54	
Cys	1.19	
Gln	1.10	
Glu	0.37	
Gly	0.75	
His	0.87	
Ile	1.60	⇦
Leu	1.30	⇦
Lys	0.74	
Met	1.05	
Phe	1.38	⇦
Pro	0.55	
Ser	0.75	
Thr	1.19	
Trp	1.37	⇦
Tyr	1.47	⇦
Val	1.70	⇦

Like:
- -prefer large, bulky groups
- -Val, Ile, Leu, Tyr, Trp, Phe

Don't Like:
- -Pro (same reason)
- -Glu/Asp/Lys (full charges too close)
- -Gly (same reason)

Protein Structure-Secondary

The β Sheet: examples

Globular Proteins

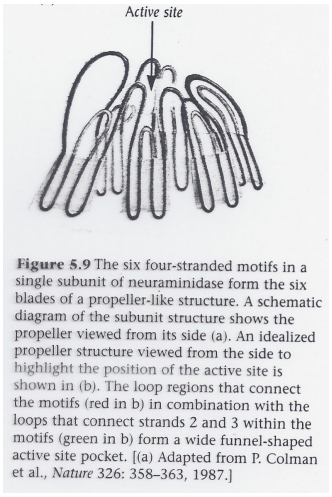
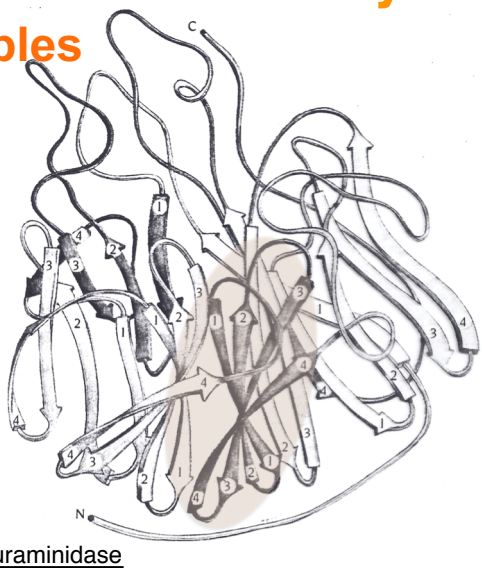


Figure 5.9 The six four-stranded motifs in a single subunit of neuraminidase form the six blades of a propeller-like structure. A schematic diagram of the subunit structure shows the propeller viewed from its side (a). An idealized propeller structure viewed from the side to highlight the position of the active site is shown in (b). The loop regions that connect the motifs (red in b) in combination with the loops that connect strands 2 and 3 within the motifs (green in b) form a wide funnel-shaped active site pocket. [(a) Adapted from P. Colman et al., *Nature* 326: 358–363, 1987.]



Neuraminidase

- 6 x 4-stranded β Sheets: make a funnel to active site
- right-handed twist

Protein Structure-Secondary

The β Sheet: examples

Fibrous Proteins: Silk fibroin (β-keratin)

Silk

all parallel β-sheet

Sequence repeats:

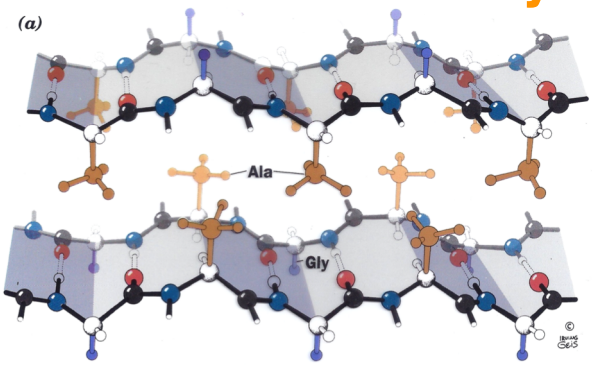
$(GAGAGSGAAG(SGAGAG)_8Y)_X$ $(x>10)$

- Gly is every-other residue, and Ala as well
- Recall the alternating R-groups
- Therefore, Gly is all on one side of sheet, and Ala on the other side of sheet

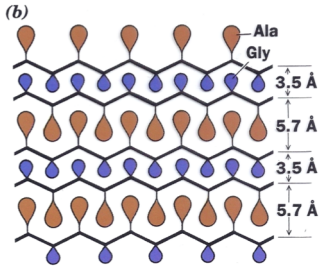
Protein Structure-Secondary

The β Sheet: examples

Silk fibroin (β-keratin)



Silk's tensile strength is comparable to that of steel and about half as strong as Kevlar



Protein Structure-Secondary

β Sheets

Amino Acid	α-Keratin (Wool)	Fibroin (Silk)
Gly	8.1	44.6
Ala	5.0	29.4
Ser	10.2	12.2
Glu + Gln	12.1	1.0
Cys	11.2	0
Pro	7.5	0.3
Arg	7.2	0.5
Leu	6.9	0.5
Thr	6.5	0.9
Asp + Asn	6.0	1.3
Val	5.1	2.2
Tyr	4.2	5.2
Ile	2.8	0.7
Phe	2.5	0.5
Lys	2.3	0.3
Trp	1.2	0.2
His	0.7	0.2
Met	0.5	0

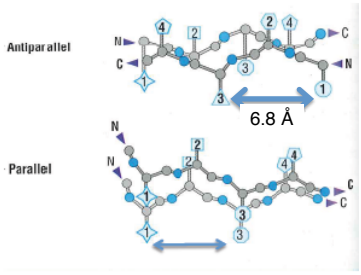
Note: The three most abundant amino acids in each protein are ~~indicated in red~~ boxed. Values given are in mole percent.

85% G, A, S

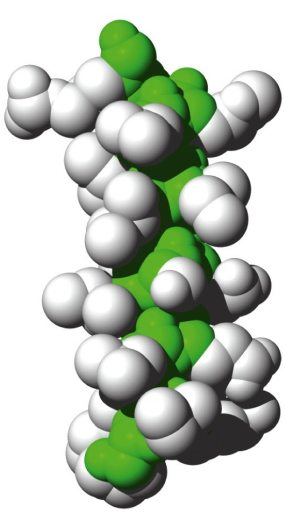
Protein Structure-Secondary

Summary of β Sheet

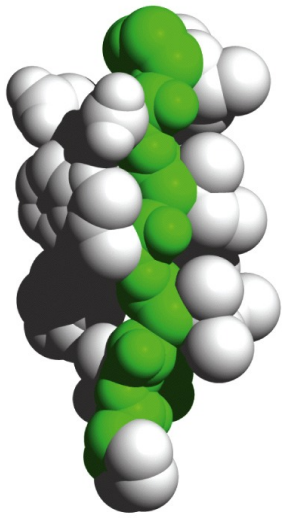
- Parallel or antiparallel
- ~7.0 Å between R groups
- ~3.5 Å between alpha carbons
- Alternating residues face opposite sides
- Extended structure
- Right-handed twist



Protein Structure-Secondary

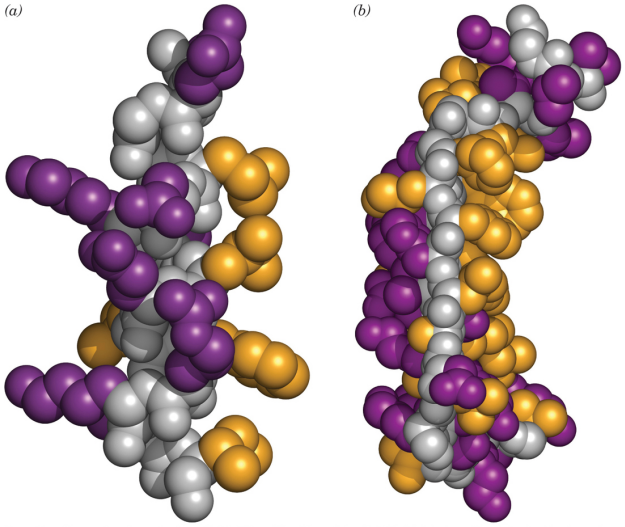


α-helix



β-sheet

Protein Structure-Secondary



Purple=polar
Gold=hydrophobic

Recall that β-sheets require strands from different parts of the same polymer, so….
How are the connected?

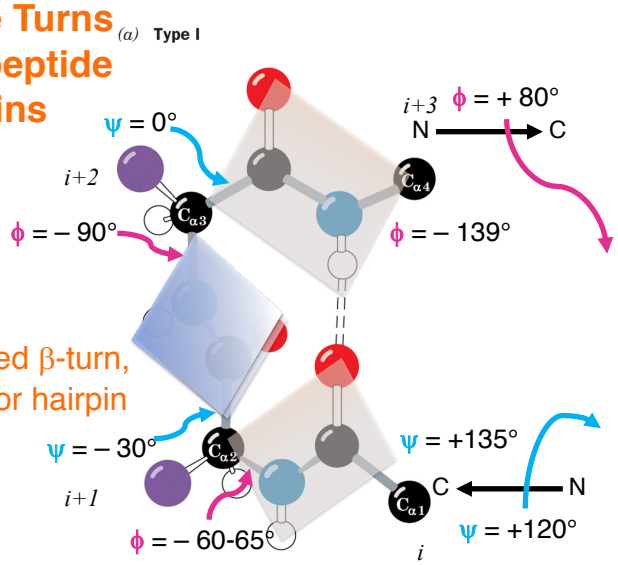
Based on X-ray structures by Ilme Schlichting, Max Planck Institut für Molekulare Physiologie, Dortmund, Germany, and Gerald Edelman, The Rockefeller University. PDBids 1A6M and 2CNA.

Oxy-Myoglobin and Concanavalin A
PDBids 1A6M and 2CNA

See "Protein Structure Videos" (First link)

Protein Structure-Secondary

Reverse Turns in Polypeptide Chains

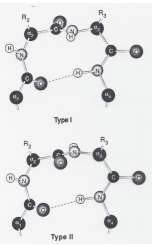


Also called β-turn, β-bend, or hairpin

Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Protein Structure-Secondary

Structure	ϕ	ψ	Rise (Dist/residue) (Å)	Residues/Repeat	Pitch (Distance/repeat) (Å)	Diameter (Å)
α-helix	-57	-47	1.5	3.6	5.4	5.0
Anti- ⇌ β-sheet	-139	+135	3.4	2	6.8	-
Parallel ⇒ β-sheet	-119	+113	3.2	2	6.4	-
β-turn-Type I				4	0	-
i + 1	-60	-30	-			
i + 2	-90	0	-			
β-turn-Type II				4	0	-
i + 1	-60	120	-			
i + 2	80	0	-			

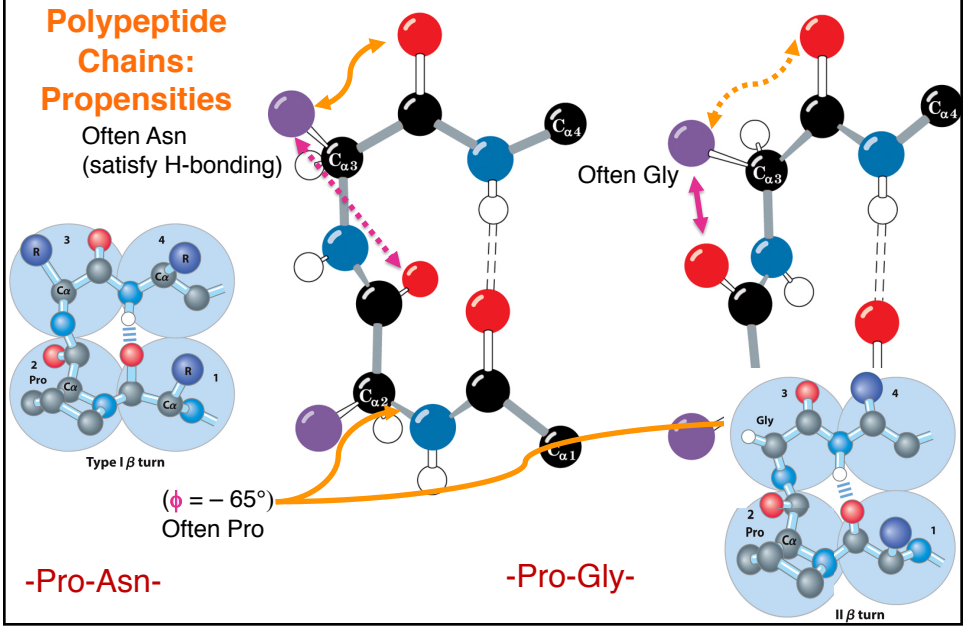


Start and stop with same angles

Protein Structure-Secondary

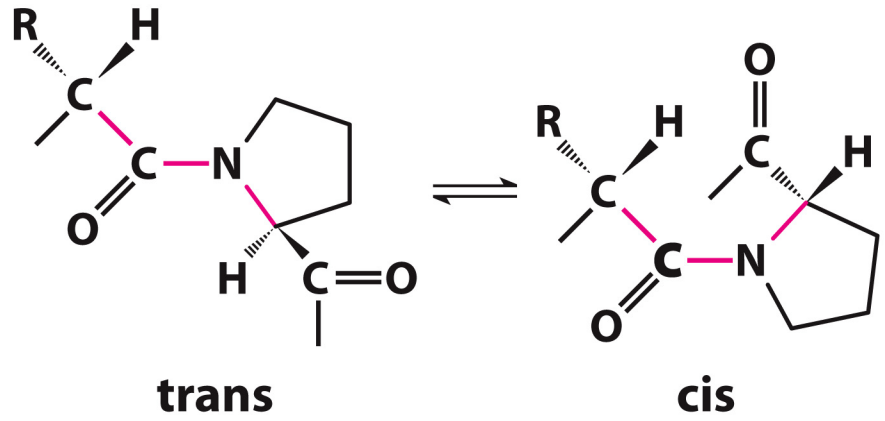
Reverse Turns in Polypeptide Chains: Propensities

Often Asn
(satisfy H-bonding)



Protein Structure-Secondary

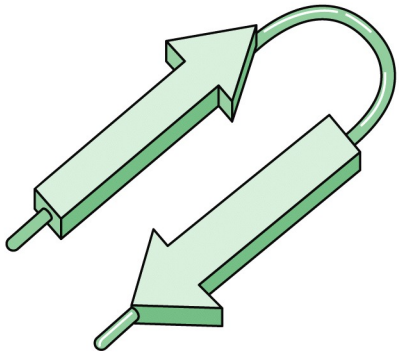
Turns in Polypeptide Chains often use *cis*-Pro



Proline isomers

Figure 4-8

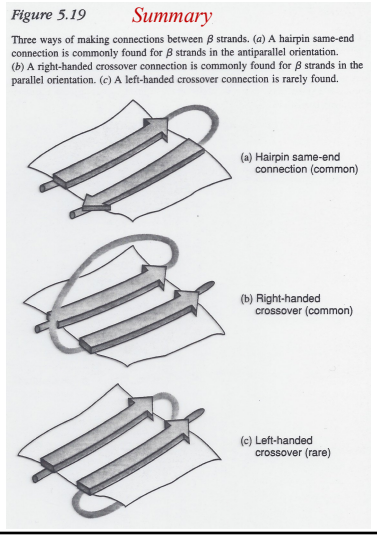
Protein Structure-Secondary

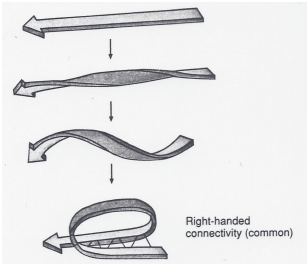


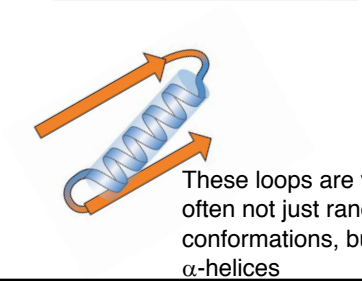
Reverse Turns are used for anti-parallel sheets, but how do parallel strands find each other to make a sheet?

Protein Structure-Secondary

These strands are not straight as we saw, and this twist helps the conformation of these loops.







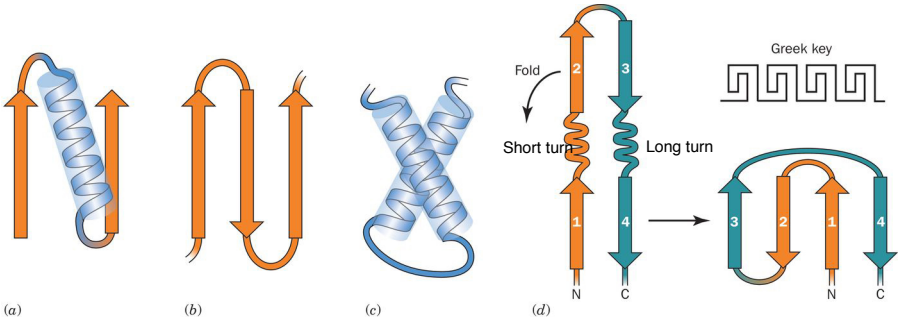
These loops are very often not just random conformations, but form α-helices

Protein Structure-Secondary

What is happening?

- Different pieces of 2° structure are mixing together.
- These are called “Motifs” or Super-secondary Structures
- What are the structures and names of some of the most common motifs?
  - βαβ
  - Rossmann Fold
  - β-hairpin
  - αα
  - Greek key
  - β-meander
  - β-barrel
  - αβ–barrel

Protein Structure-Supersecondary

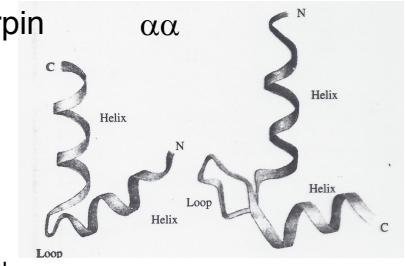


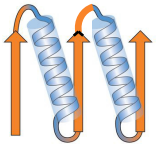
βαβ

β-hairpin

Take a β-hairpin, which is an antiparallel β–sheet, but the **topology** of the strands is not sequential

αα





βαβαβ=Rossmann Fold

αα with different sized loops

Protein Structure-Supersecondary

Due to the right-handed twist in the β-strands, as you add more strands the structure comes back on itself to form "barrels"

β-meander $(\beta)_8$

Greek Key $(\beta)_8$ (β-barrel)

$(\alpha\beta)_8$-Barrel

4 (βαβ) motifs connected by 4 α-helices

FIGURE 7-49. Comparisons of the backbone folding patterns of protein β barrels (*right*) with geometric motifs commonly used to decorate Native American and Greek weaving and pottery (*left*). (*a*) Native American polychrome cane basket and the polypeptide backbone of **rubredoxin** from *Clostridium pasteurianum* showing its linked β meanders. [Museum of the American Indian, Heye Foundation.] (*b*) Red figured Greek amphora with its Greek key border area showing Cassandra and Ajax (about 450 B.C.) and the polypeptide backbone of human **prealbumin** with its "Greek key" pattern. [The Metropolitan Museum of Art, Fletcher Fund, 1956.] (*c*) Early Anasazi redware pitcher from New Mexico and the polypeptide backbone of chicken muscle triose phosphate isomerase showing its "lightning" pattern of overlapping βαβ units. This so-called α/β barrel is also diagrammed in Fig. 7-19*b*. [Museum of the American Indian, Heye Foundation.] [After Richardson, J.S., *Nature* **268**, 498 (1977).]