

**BT351**

**Protein Structure**

# Protein Structure

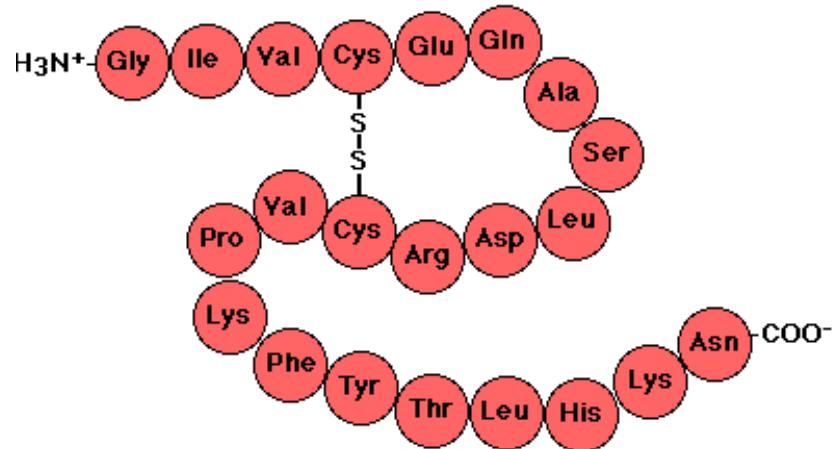
- **Proteins are polymers consisting of amino acids linked by peptide bonds**
- Many different conformation (three dimensional) are possible for proteins
- At least one major structure level has biological activity called (native conformation)

# Levels of Protein Structure

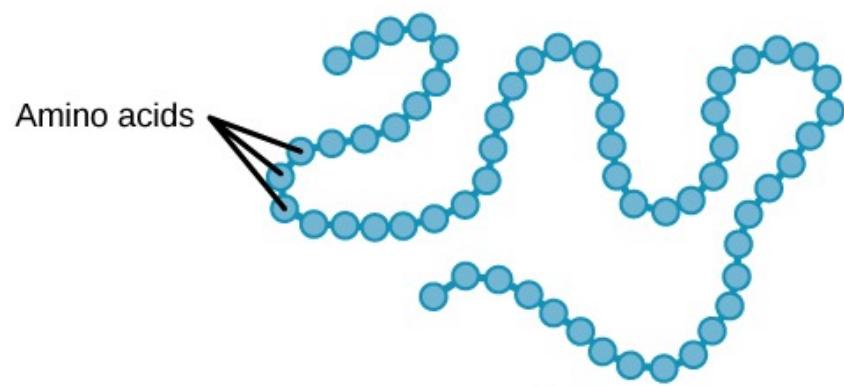
- **Primary structure:** is the order of amino acids in a polypeptide chain, read from the N-terminal end to the C-terminal end
- **Secondary structure:** the repetitive arrangement of the atom in the peptide backbone of protein. ( $\alpha$ -helix and  $\beta$ -pleated sheet). Resulted from hydrogen bonding between the amide N-H and the carbonyl group of the peptide.
- **Tertiary structure:** 3-D arrangement of all atoms including those in the side chains and any prosthetic groups
- **Quaternary structure:** arrangement of subunits with respect to each other.
  - Subunits are multiple polypeptide chains

# 1° Structure (one dimensional)

Is the linear sequence (order) of amino acids from the amino to carboxyl end of protein and the location of disulfide (-S-S-) bridges or bonds.

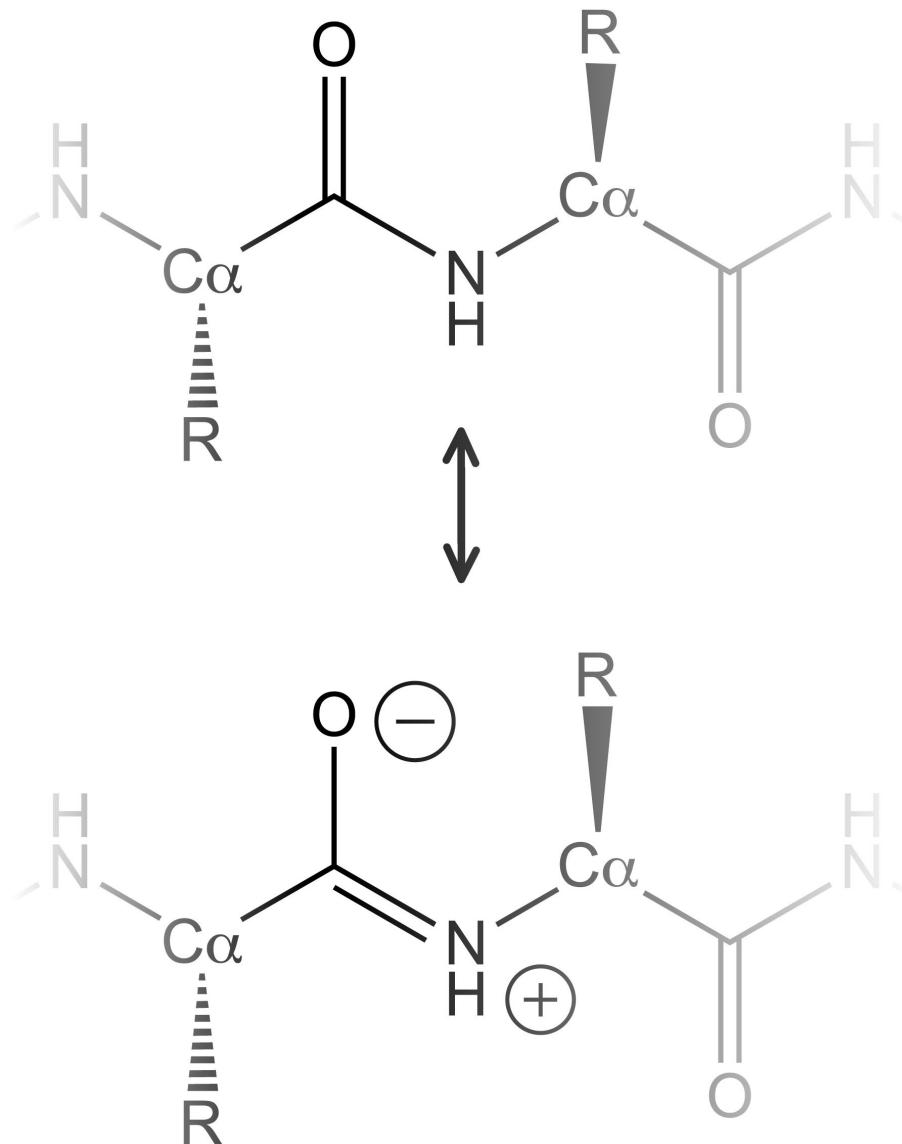


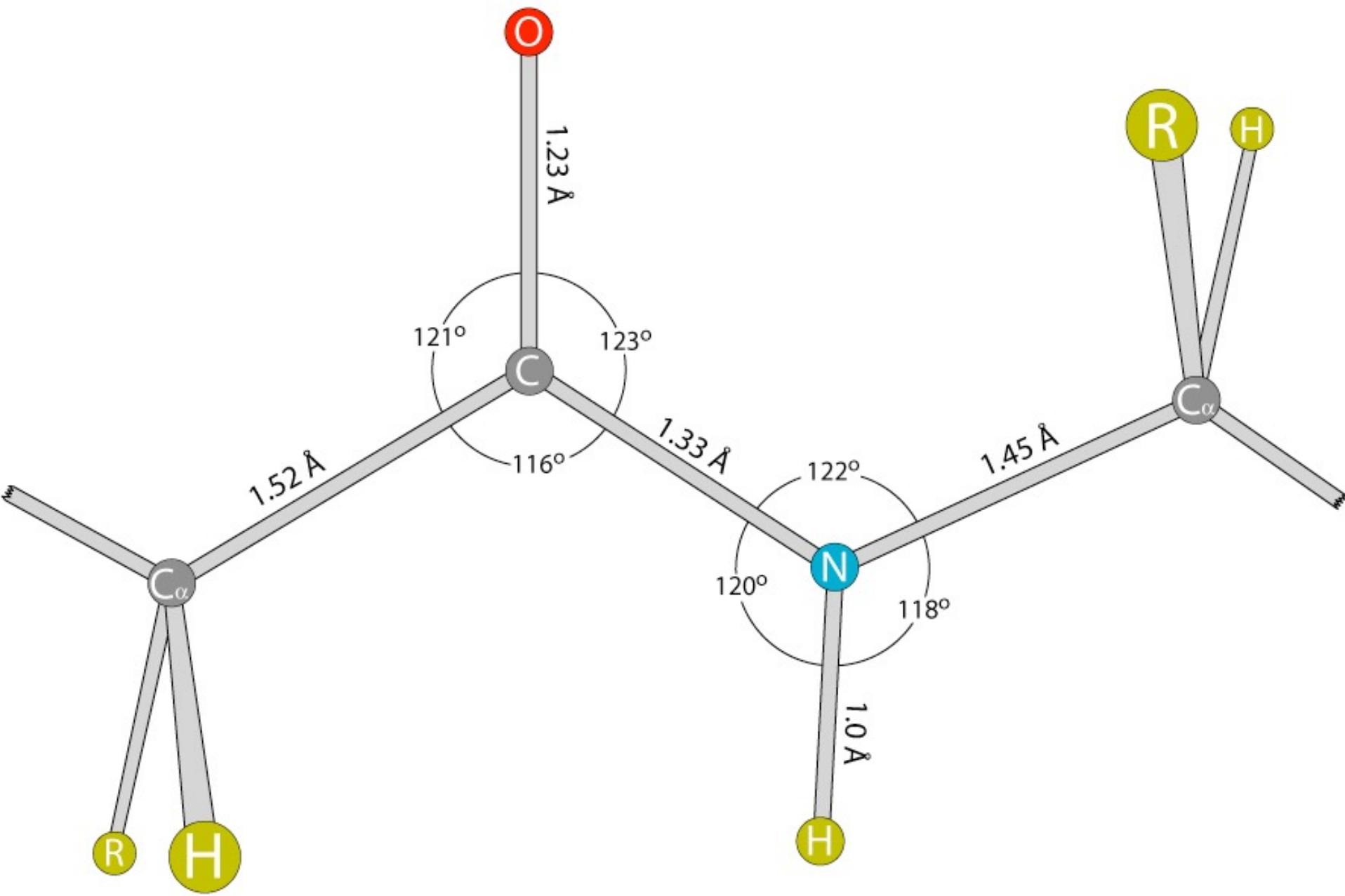
- The 1° sequence of proteins determines its 3-D conformation
- Amino acid substitution ranges from negligible effect to a complete loss of activity and it depends on the nature of altered residue.
- Changes in just one amino acid in a sequence of protein can alter biological function, e.g. hemoglobin associated with sickle-cell anemia
- Determination of 1° sequence is routine biochemistry lab work.

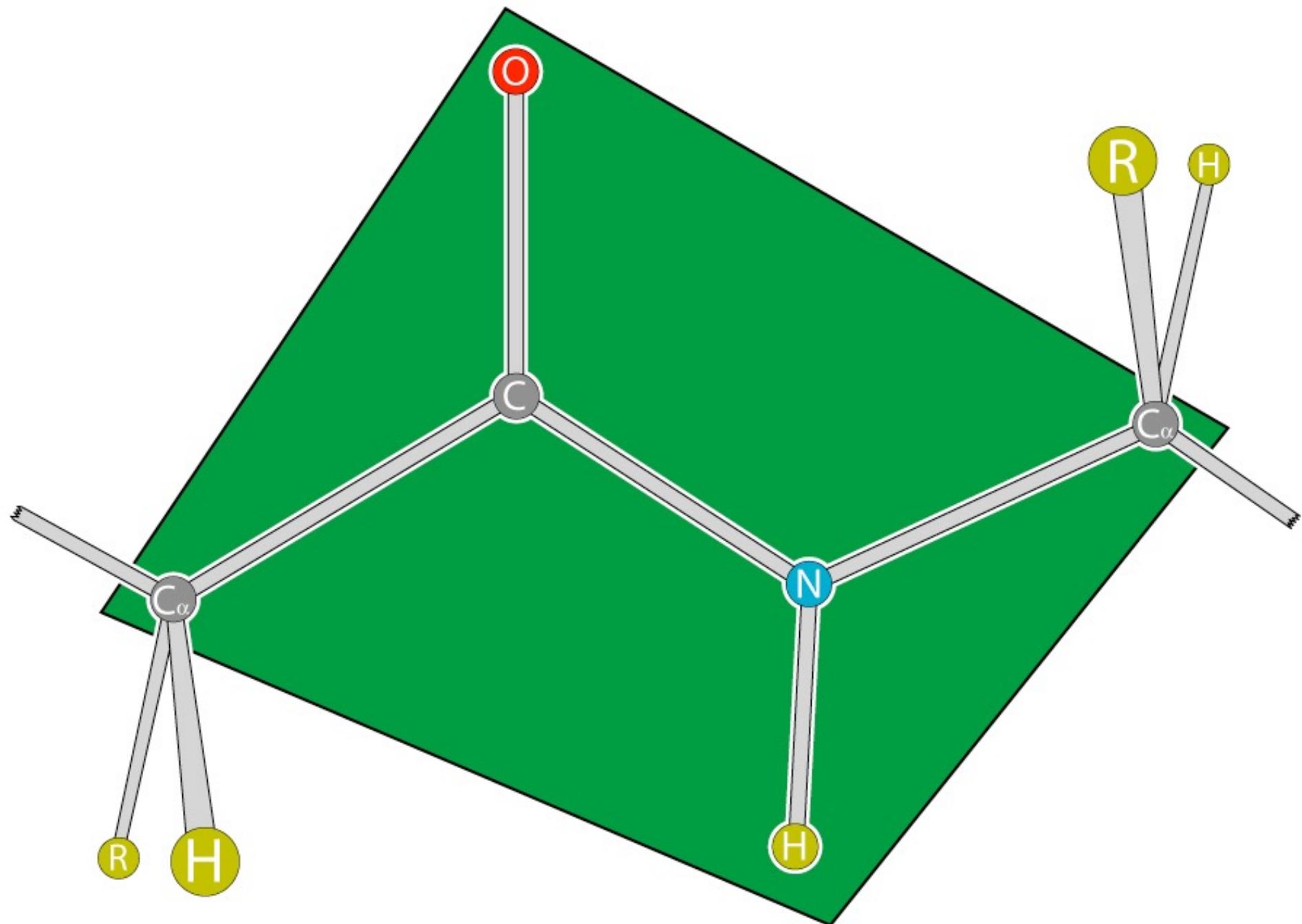


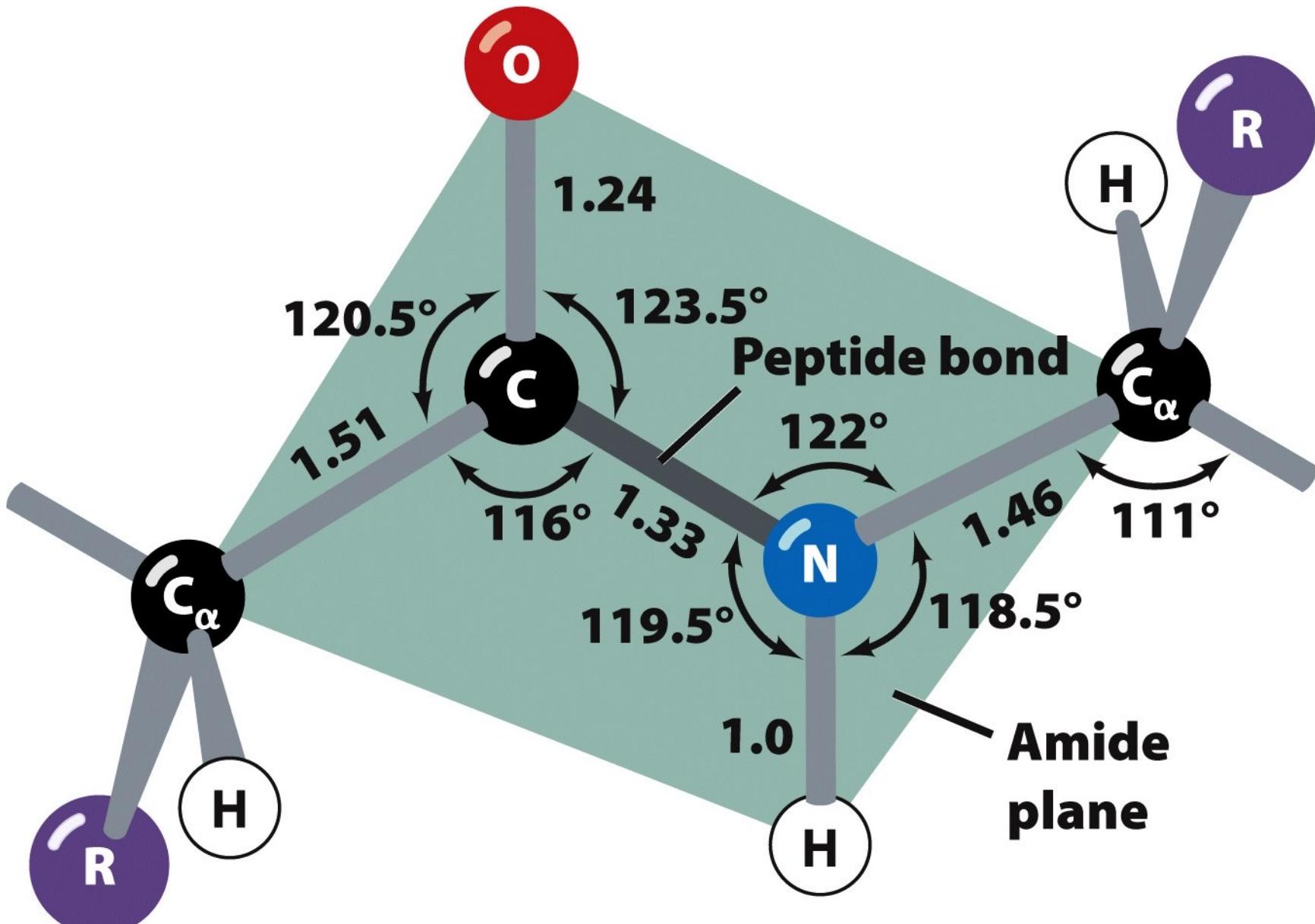
**Primary Protein structure**  
sequence of a chain of  
amino acids

# Peptide bond



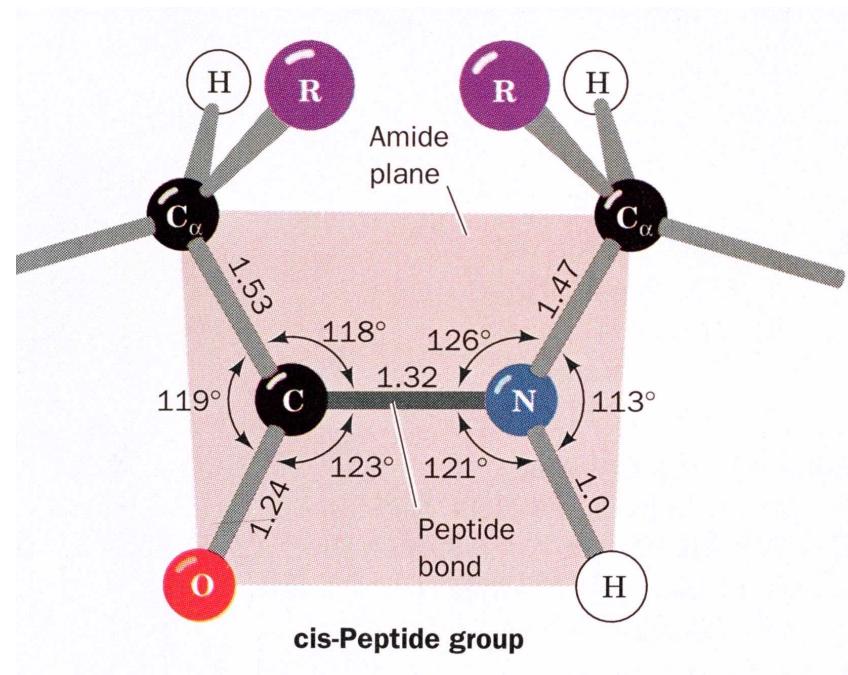
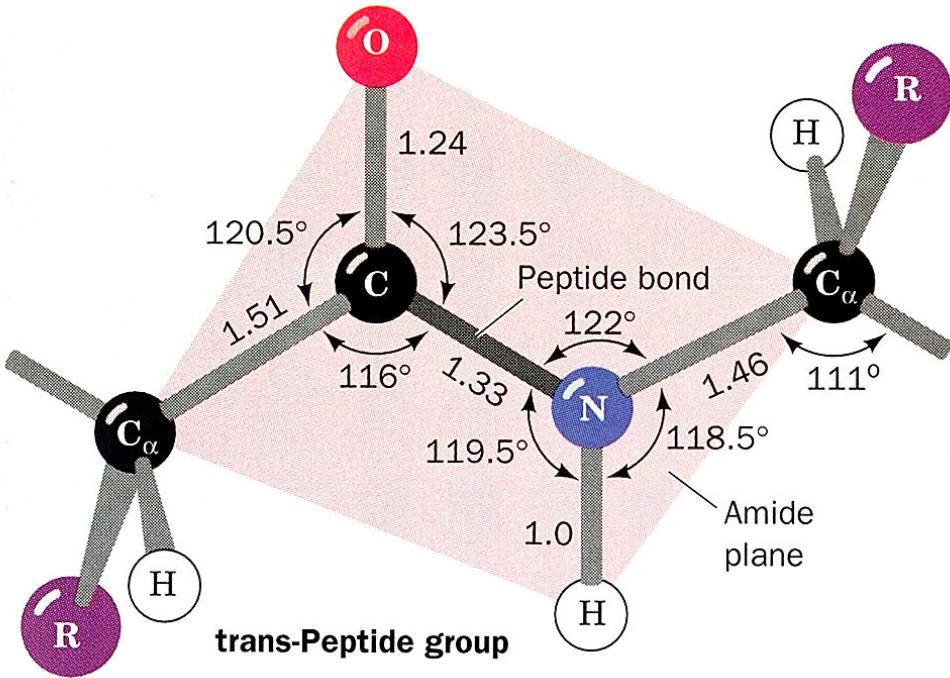




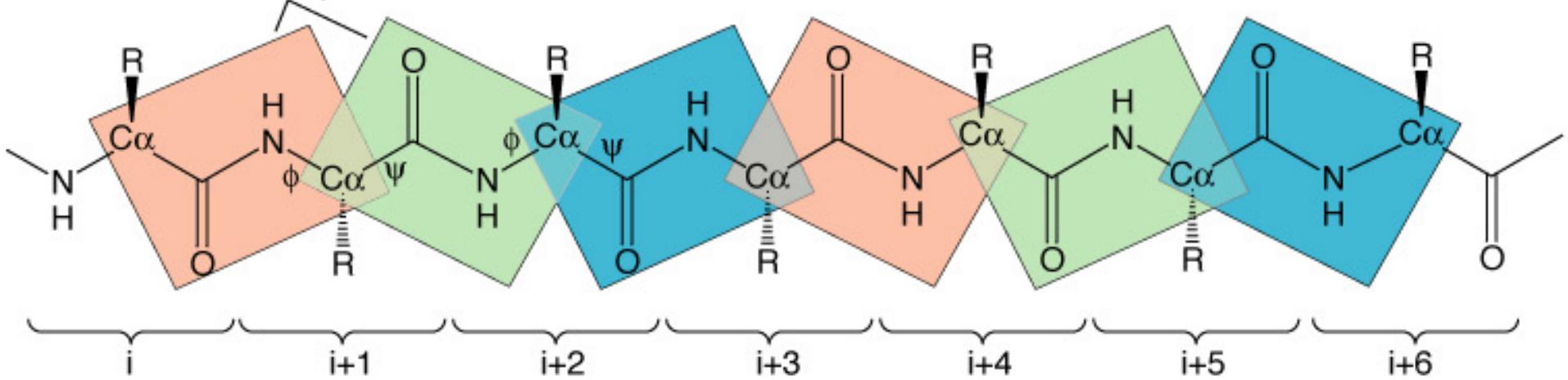


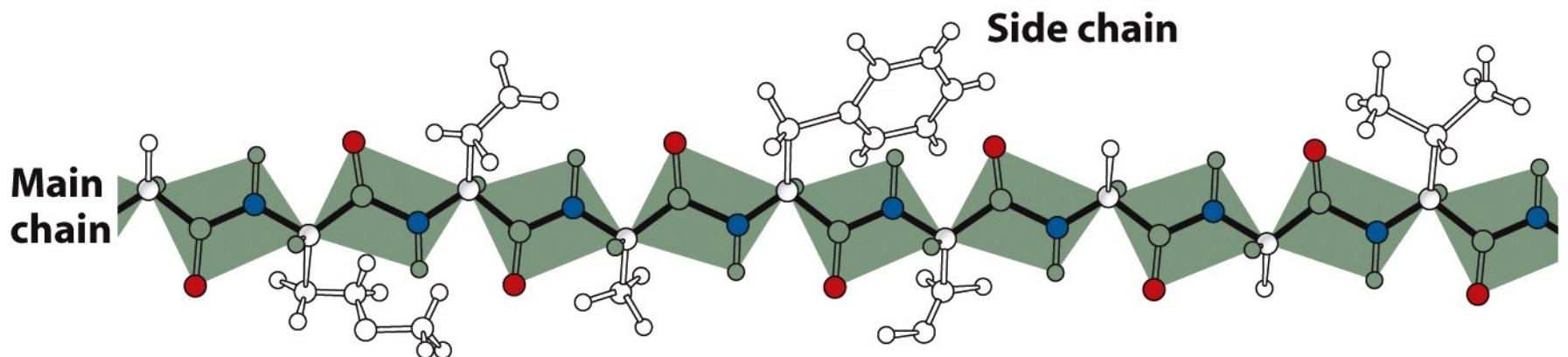
# Peptide bonds are planar

- Resonance energy depends on dihedral/torsional angle (Ca-C-N-Ca)
- For peptides, this is the angle between the Ca-C and N-Ca bonds
- For a **trans** peptide bond, the dihedral angle is  $180^\circ$  by definition.
- In a **cis** peptide bond, the dihedral angle is  $0^\circ$  by definition.
- Most peptide bonds are trans, 10% that follow proline may be cis
- Note: differences between bond angles and bond lengths comparing **cis** and **trans** forms of a generic dipeptide.



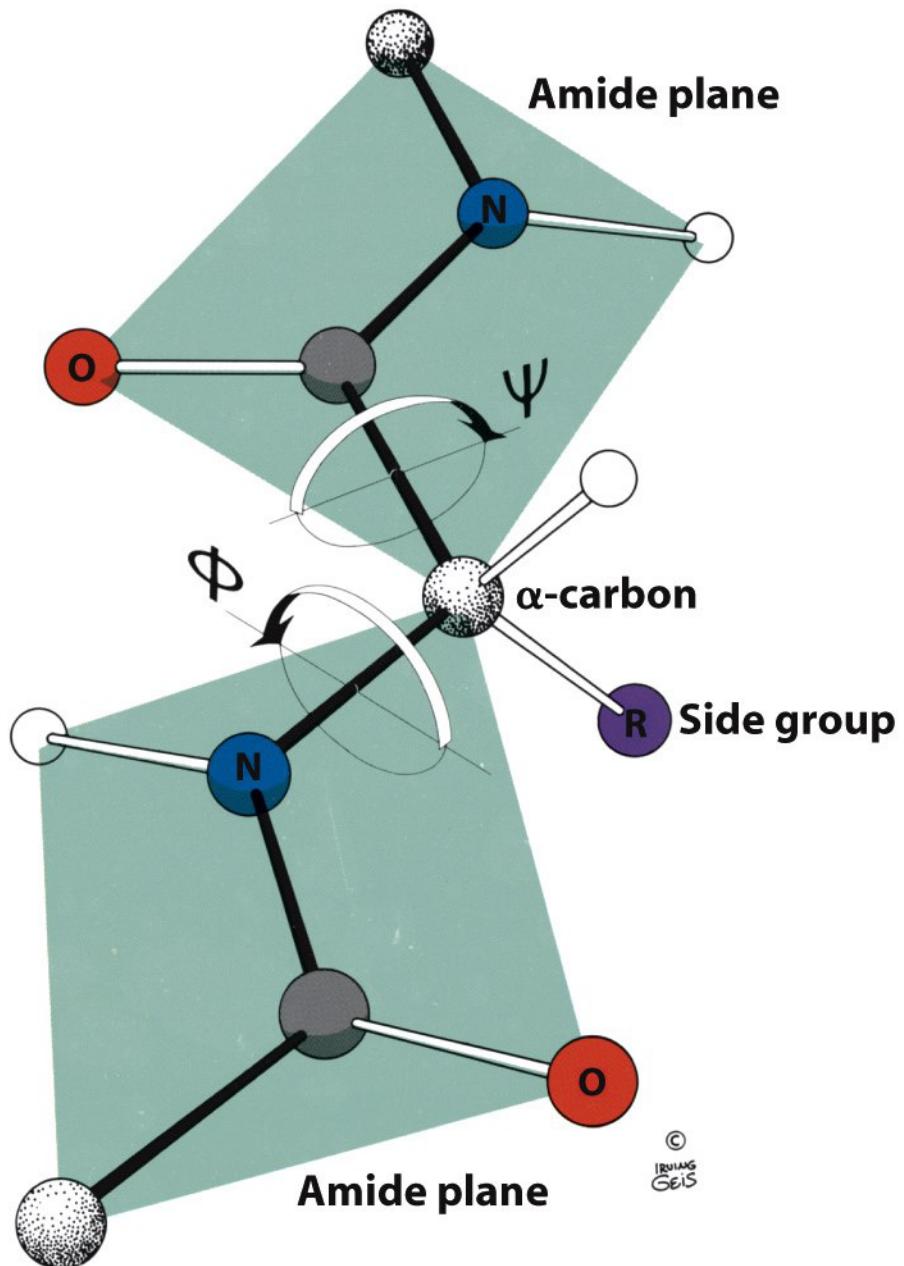
sets of 6 coplaner atoms





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

Figure 6-3

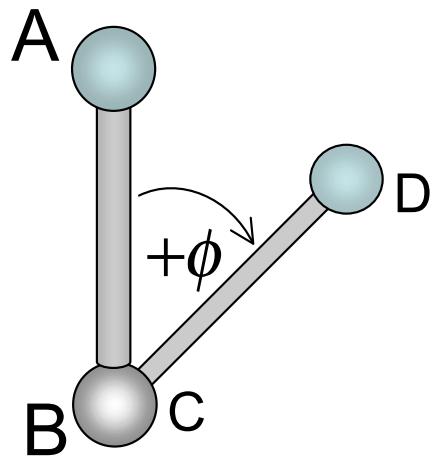
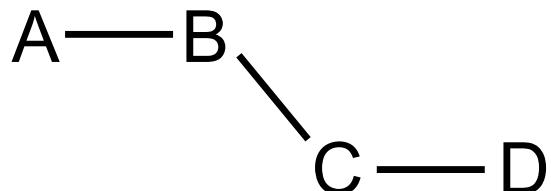


Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Figure 6-4

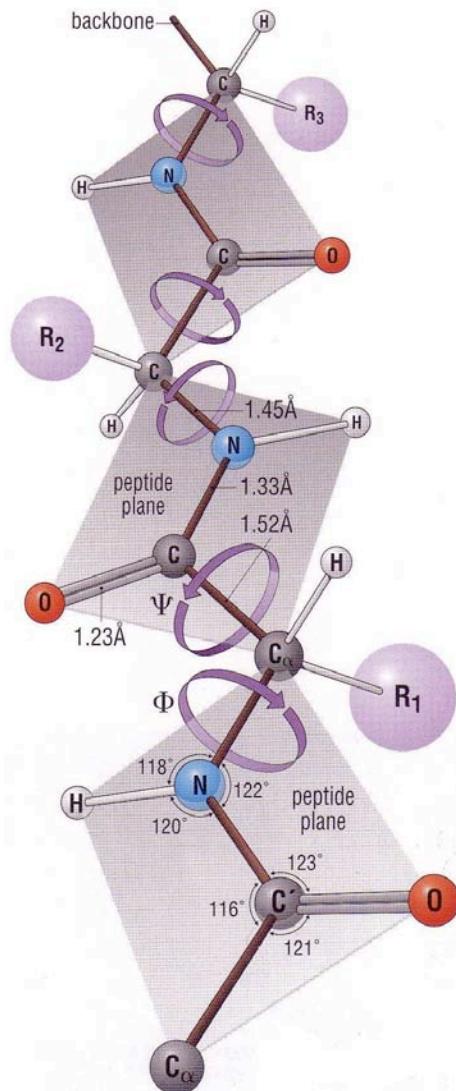
# Dihedral angles

Definition of  $\phi$  (AB-CD) dihedral angle



$\phi$  (AB-CD) is  
+ if clockwise  
- if counterclockwise

# Polypeptide structure can be described by backbone dihedral angles



Variable backbone dihedrals

$$\phi_i \quad \text{C}'_{i-1}-\text{N}_i-\text{C}_i^\alpha-\text{C}'_i$$

$$\psi_i \quad \text{N}_i-\text{C}_i^\alpha-\text{C}'_i-\text{N}_{i+1}$$

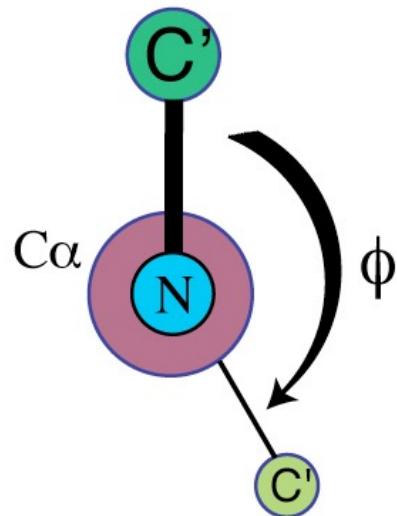
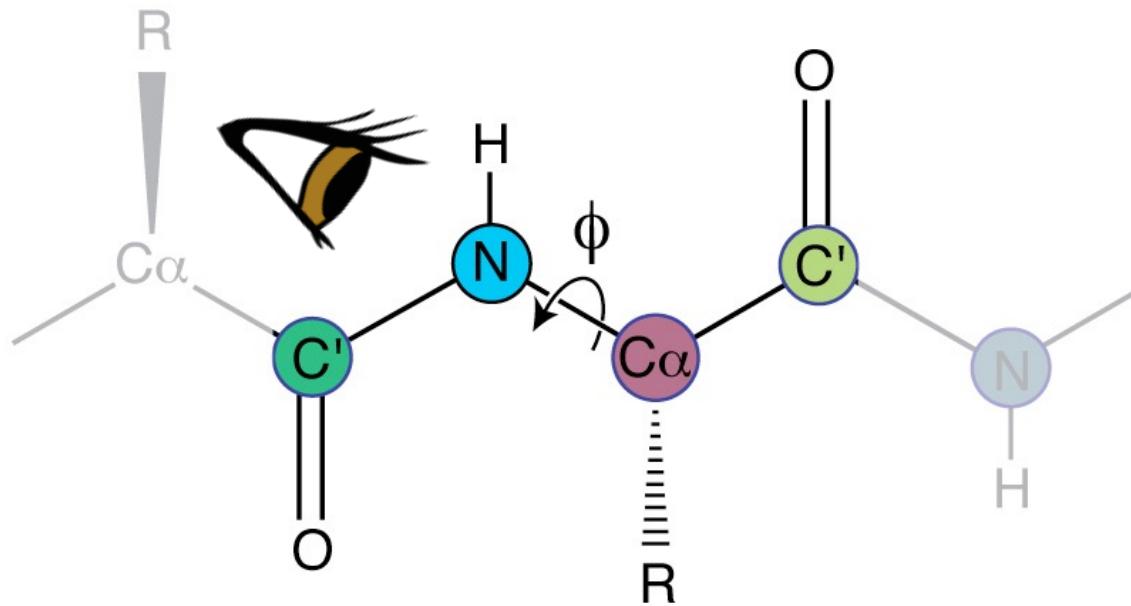
Peptide plane dihedral

$$\omega_{i,i+1} \quad \text{C}_i^\alpha-\text{C}'_i-\text{N}_{i+1}-\text{C}_{i+1}^\alpha$$

$\omega$  tends to be planar due to delocalization of the  $\text{C}'$   $\pi$  electron and the nitrogen lone pair.

$$\omega = 180^\circ \pm 10^\circ$$

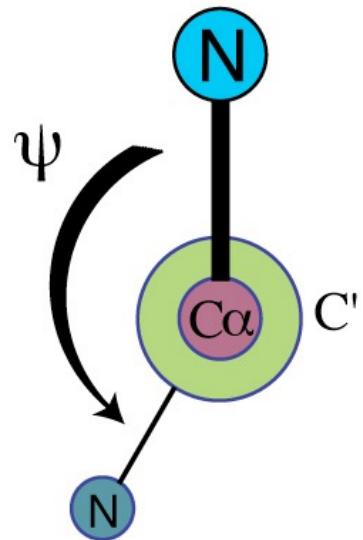
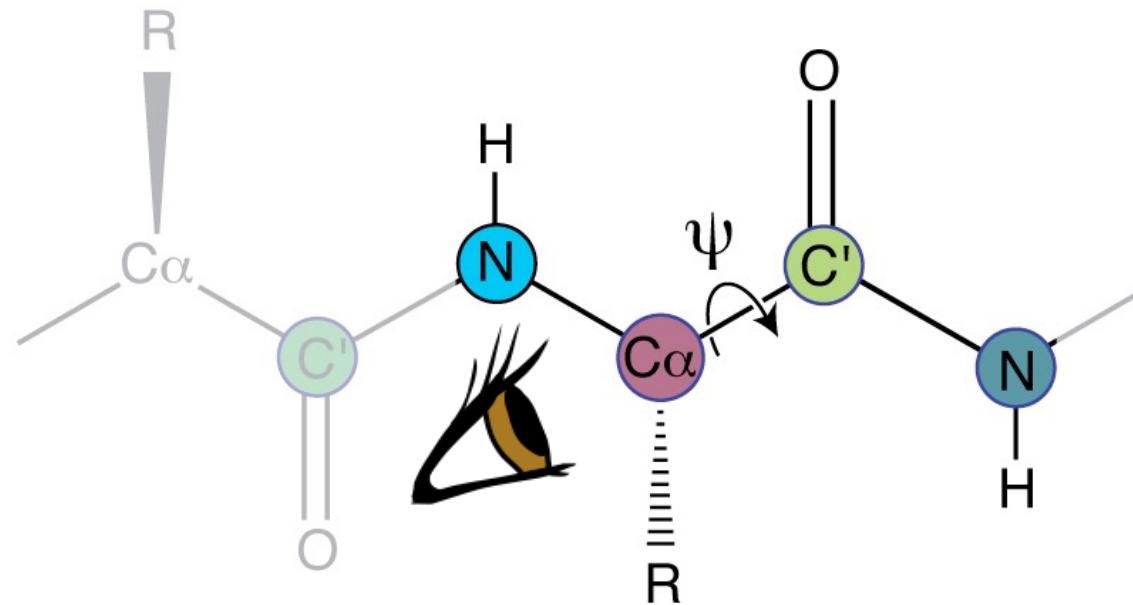
# Torsion Angle Phi ( $\phi$ )



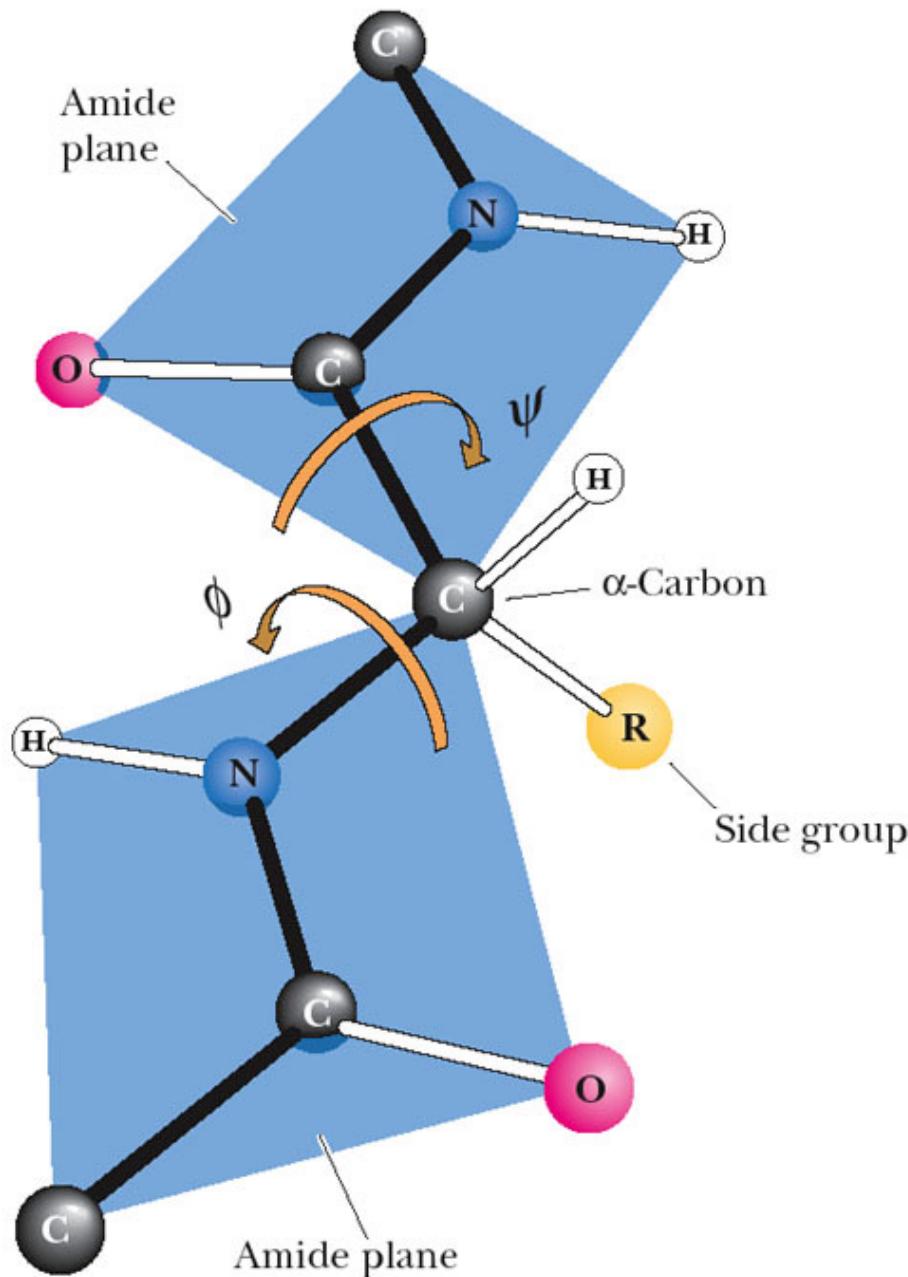
This torsion angle  $\phi$  is right handed and  
is therefore positive (+130)

S

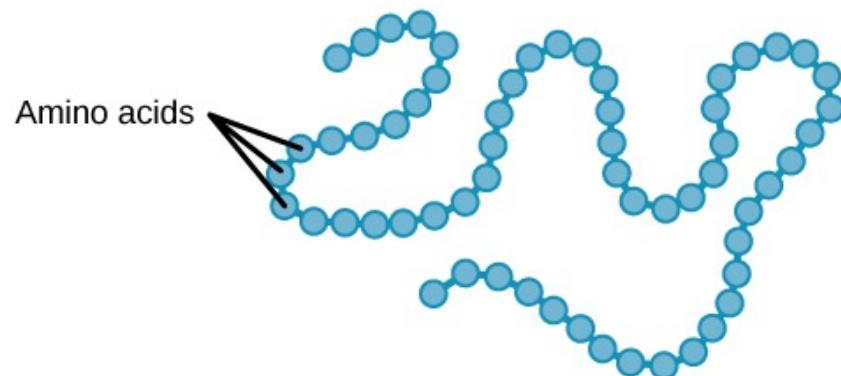
# Torsion Angle Psi ( $\psi$ )



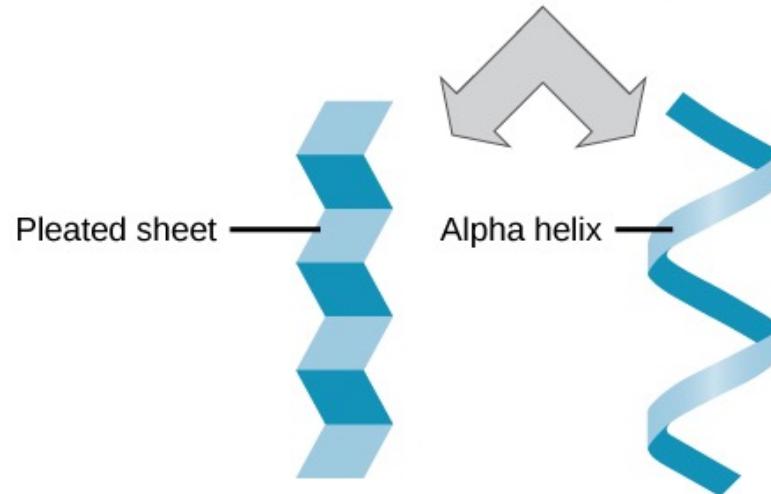
This torsion angle  $\psi$  is left-handed and is therefore negative (-130)



$$\phi = 180^\circ, \psi = 180^\circ$$



**Primary Protein structure**  
sequence of a chain of  
amino acids



**Secondary Protein structure**  
hydrogen bonding of the peptide  
backbone causes the amino  
acids to fold into a repeating  
pattern

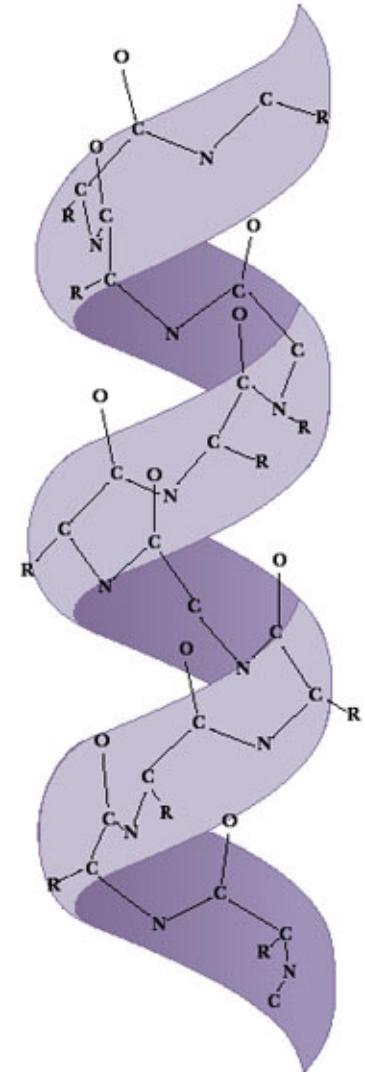
# Secondary Structure

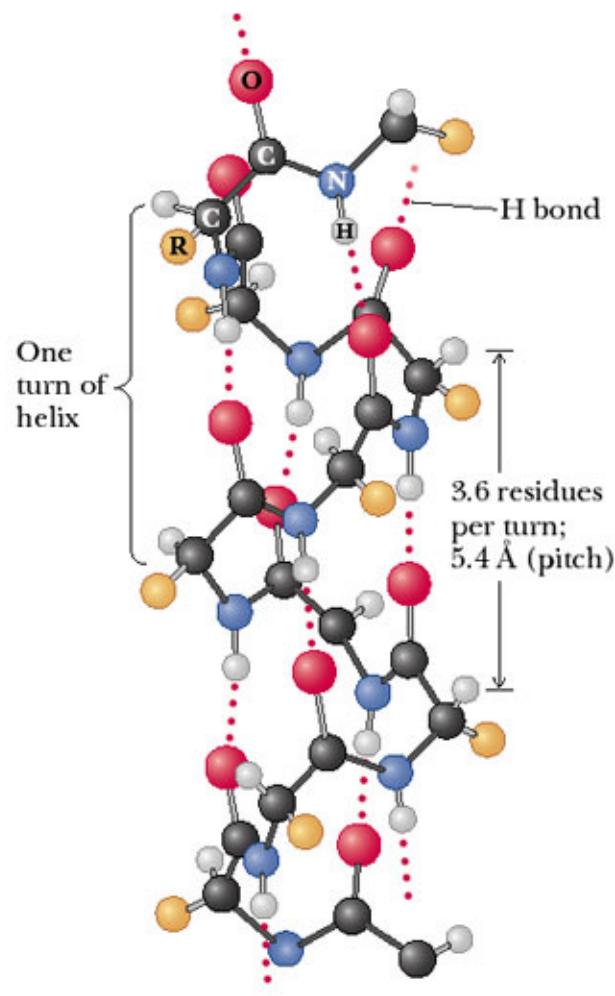
- Hydrogen-bonded arrangement of backbone of the protein
- Within each amino acids residue, two bonds have reasonably free rotating ability:
  - The Bond between a carbon and amino nitrogen of that residue residue (C-N) called phi ( $\phi$ )
  - The Bond between the a carbon and carboxyl carbon of that residue (C-C) called psi ( $\Psi$ )
  - The peptide bond is rigid and planar

Most common secondary structures are: The repeating  $\alpha$ -helix and  $\beta$ -pleated Sheets

# $\alpha$ -helix

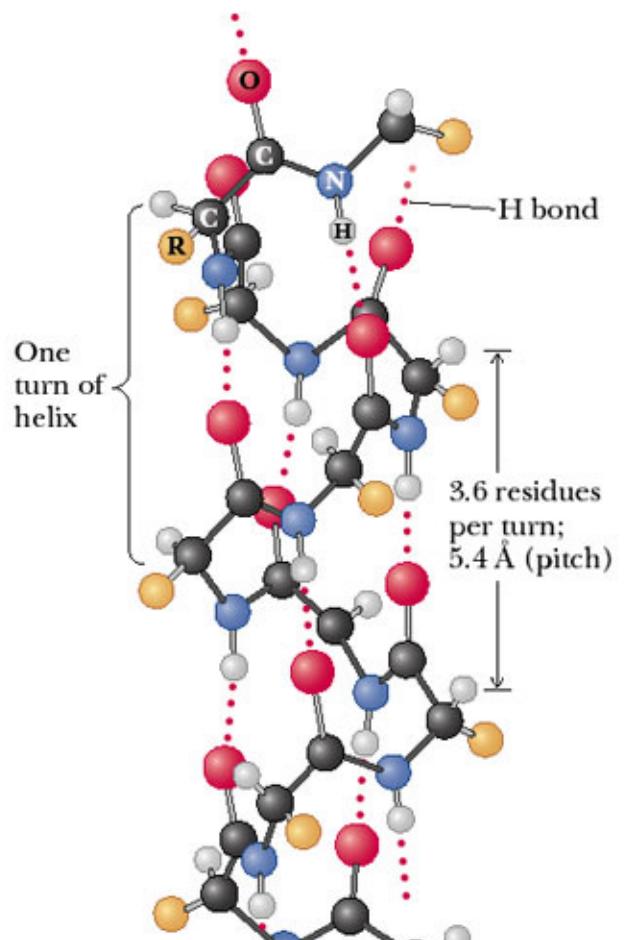
- This structure is stabilized by hydrogen bonds parallel to the helix axis within the backbone.
- From the N-terminal of a protein the C=O group of each amino acid residue is hydrogen bonded to N-H group of the amino acid 4 residues away
  - C=O on aa 1 is hydrogen bonded to N-H on 5
  - C=O on aa 2 is hydrogen bonded to N-H on 6
  - There are 3.6 aa/ turn
  - Repeat distance or the linear distance between corresponding points on successive turns (pitch of helix) is 5.4 Å





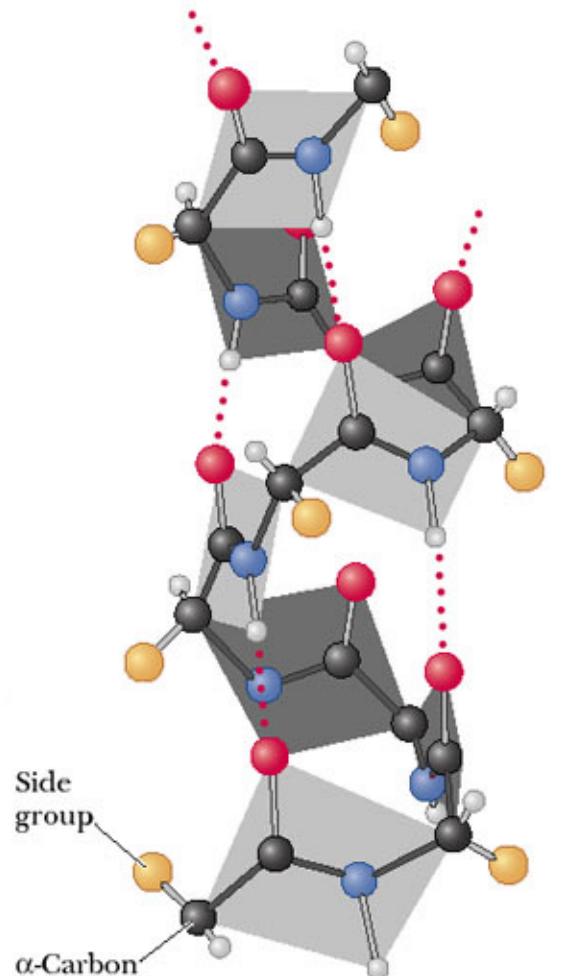
Hydrogen bonds stabilize  
the helix structure.

(a)



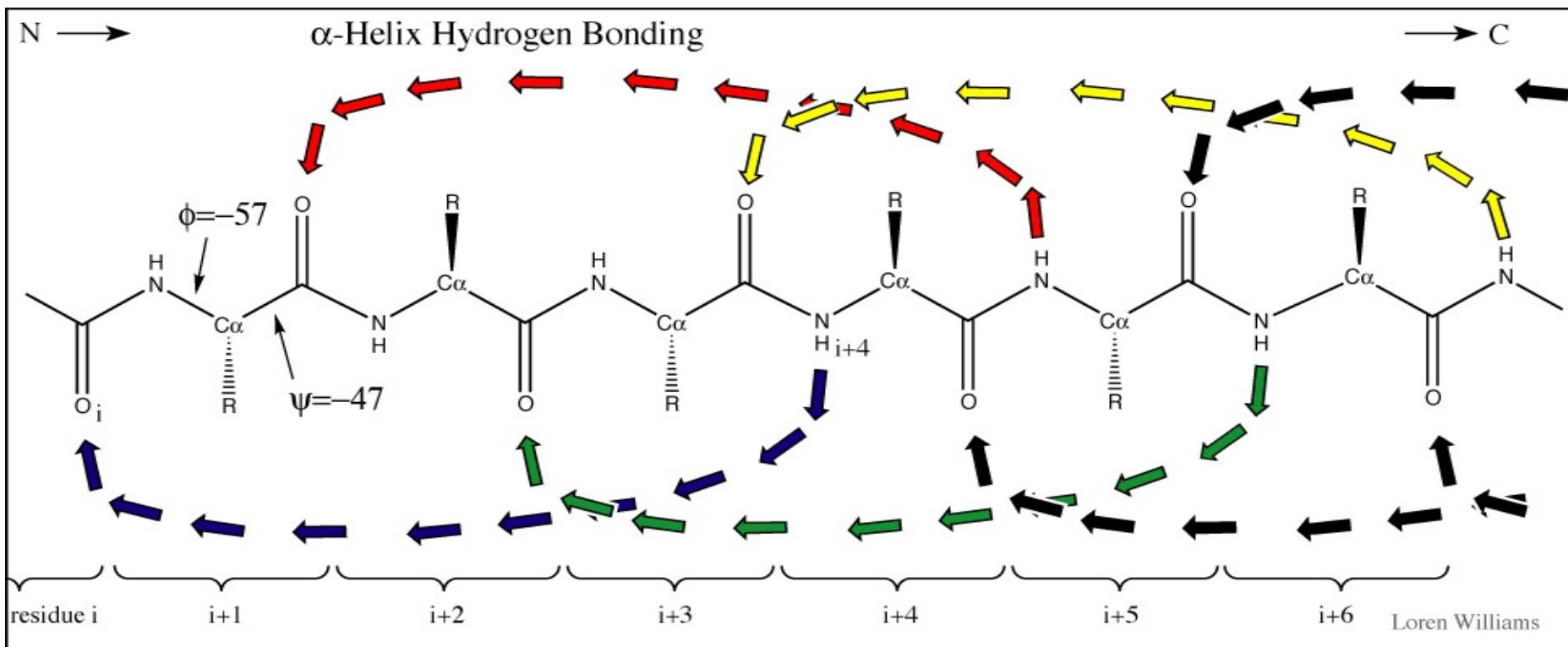
Hydrogen bonds stabilize the helix structure.

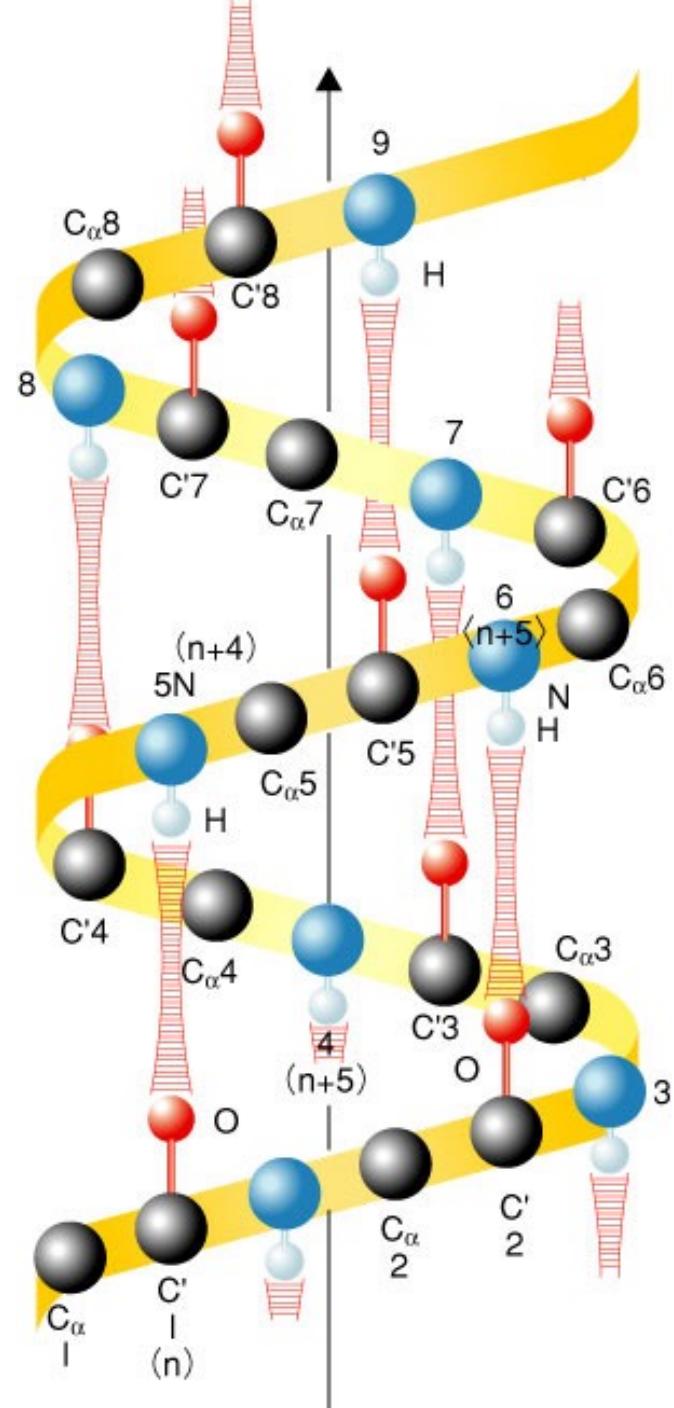
(a)



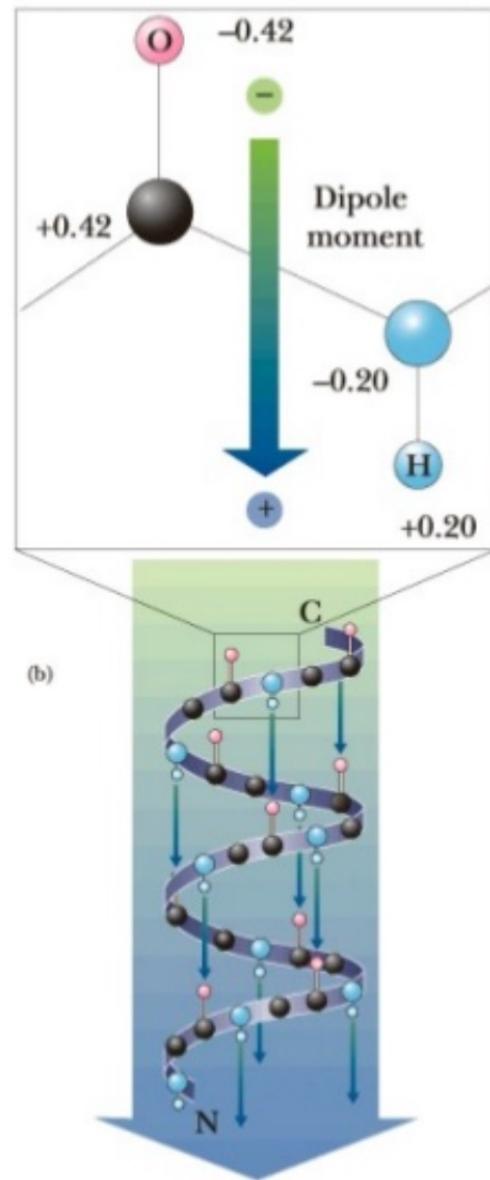
The helix can be viewed as a stacked array of peptide planes hinged at the  $\alpha$ -carbons and approximately parallel to the helix.

# **Secondary Structure - $\alpha$ -helix**





# *$\alpha$ -helix dipole moment*

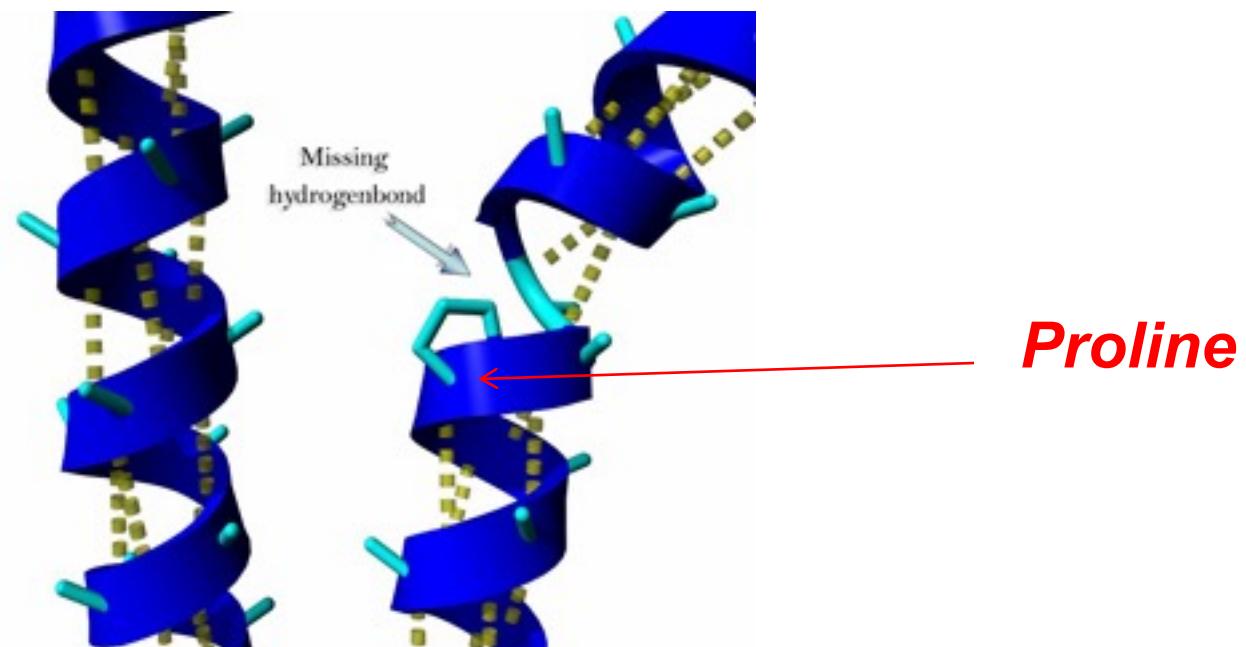


# **α-helix disruptors**

- **Several factors can disturb an α-helix**
  1. Proline creates a bend because of
    - the restricted rotation due to its cyclic structure
    - its α-amino group has no N-H for hydrogen bonding
      - cannot participate in intrachain hydrogen bonding.

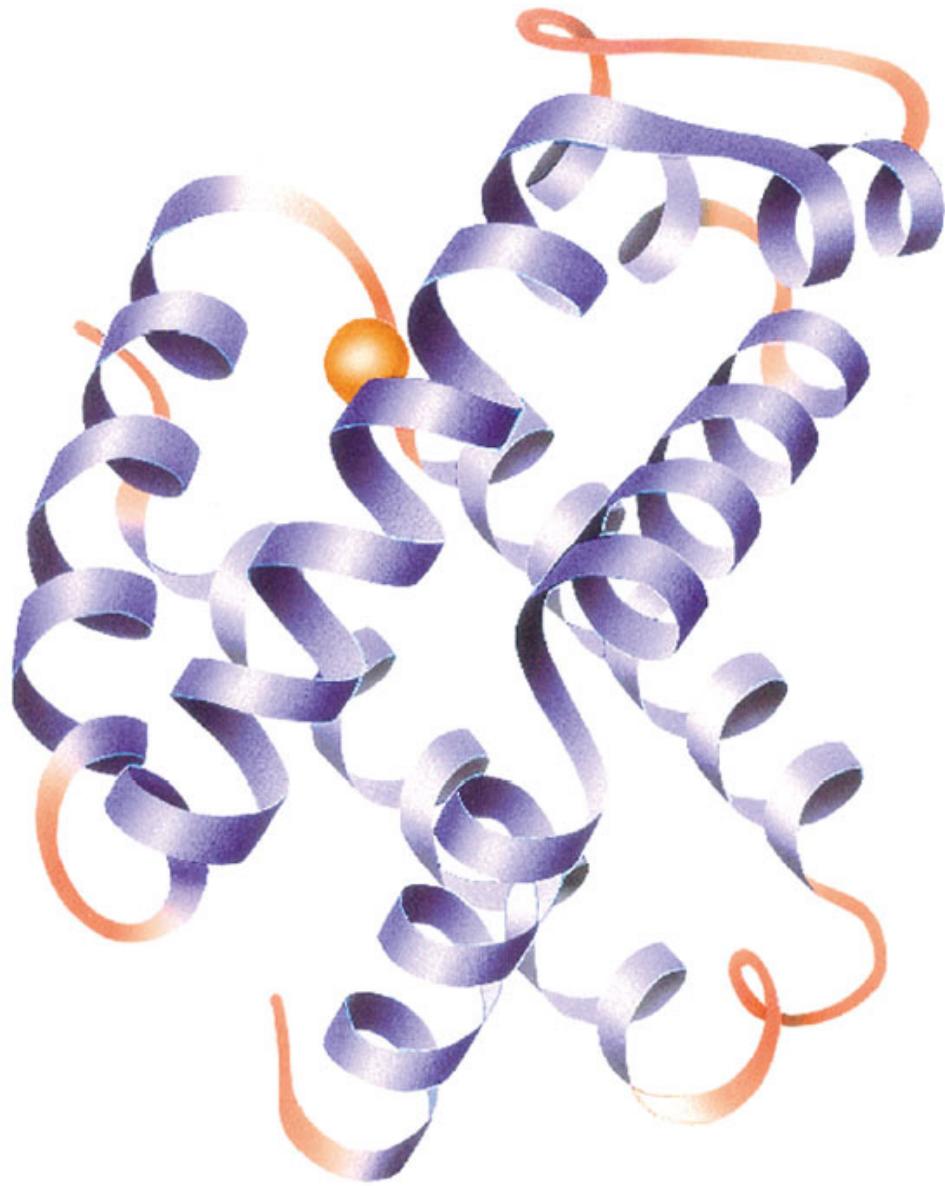
# $\alpha$ -helix disruptors

- Several factors can disturb an  $\alpha$ -helix
  - 1. Proline creates a bend because of
    - the restricted rotation due to its cyclic structure
    - its  $\alpha$ -amino group has no N-H for hydrogen bonding
      - cannot participate in intrachain hydrogen bonding.

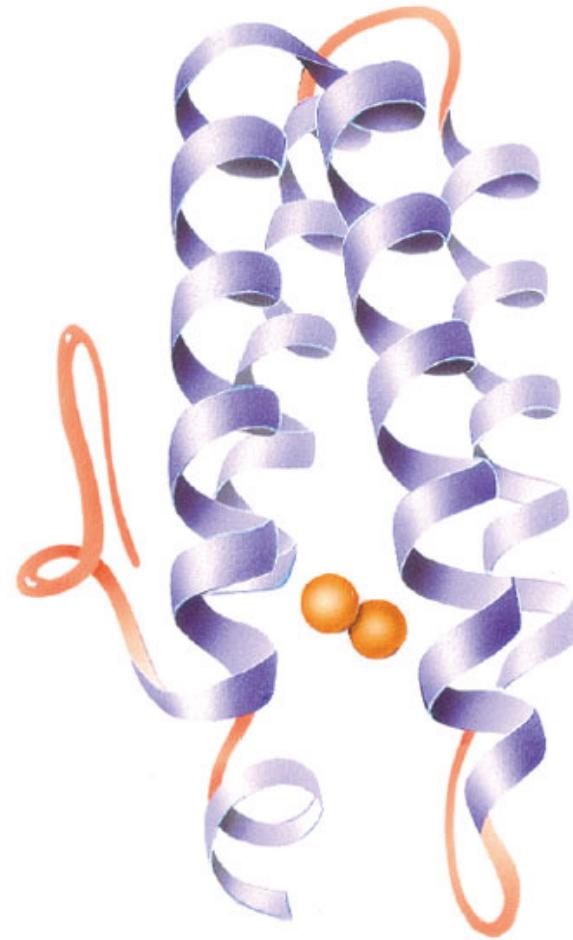


# **α-helix disruptors**

- **Several factors can disturb an α-helix**
  1. Proline creates a bend because of
    - the restricted rotation due to its cyclic structure
    - its  $\alpha$ -amino group has no N-H for hydrogen bonding
      - cannot participate in intrachain hydrogen bonding.
  2. Strong electrostatic repulsion caused by the proximity of several side chains of like charge, e.g.,
    - Lys and Arg (+ charges) or Glu and Asp (- charges)
  3. Steric repulsion (Crowding) caused by the proximity of bulky side chains, e.g., Val, Ile, Thr
    - The R groups of amino acids extend to the outside (No enough room for them in the interior)



$\beta$ -Hemoglobin subunit

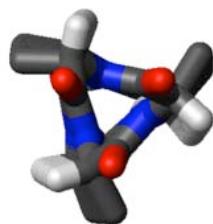
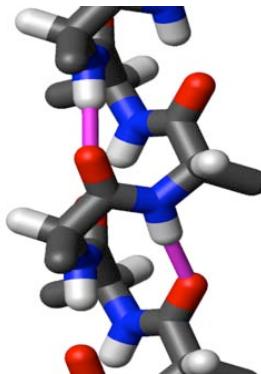


Myohemerythrin

## $\text{3}_{10}$ helix

3 residues per turn  
H-bond (i,i+3)

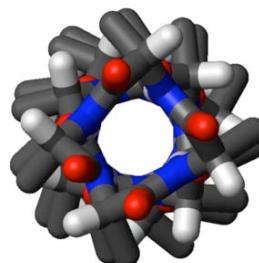
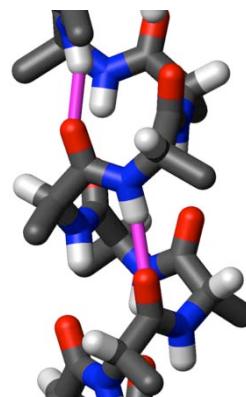
$$\phi, \psi = -49, -26$$



## $\alpha$ helix

3.6 residues per turn  
H-bond (i,i+4)

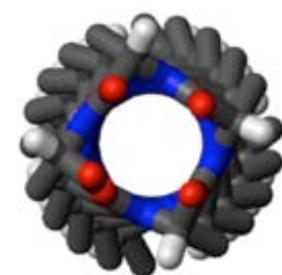
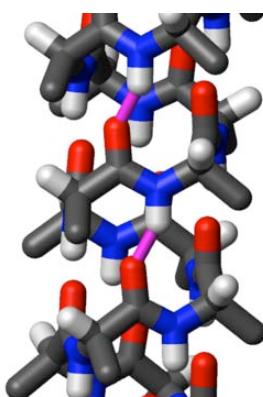
$$\phi, \psi = -60, -40$$

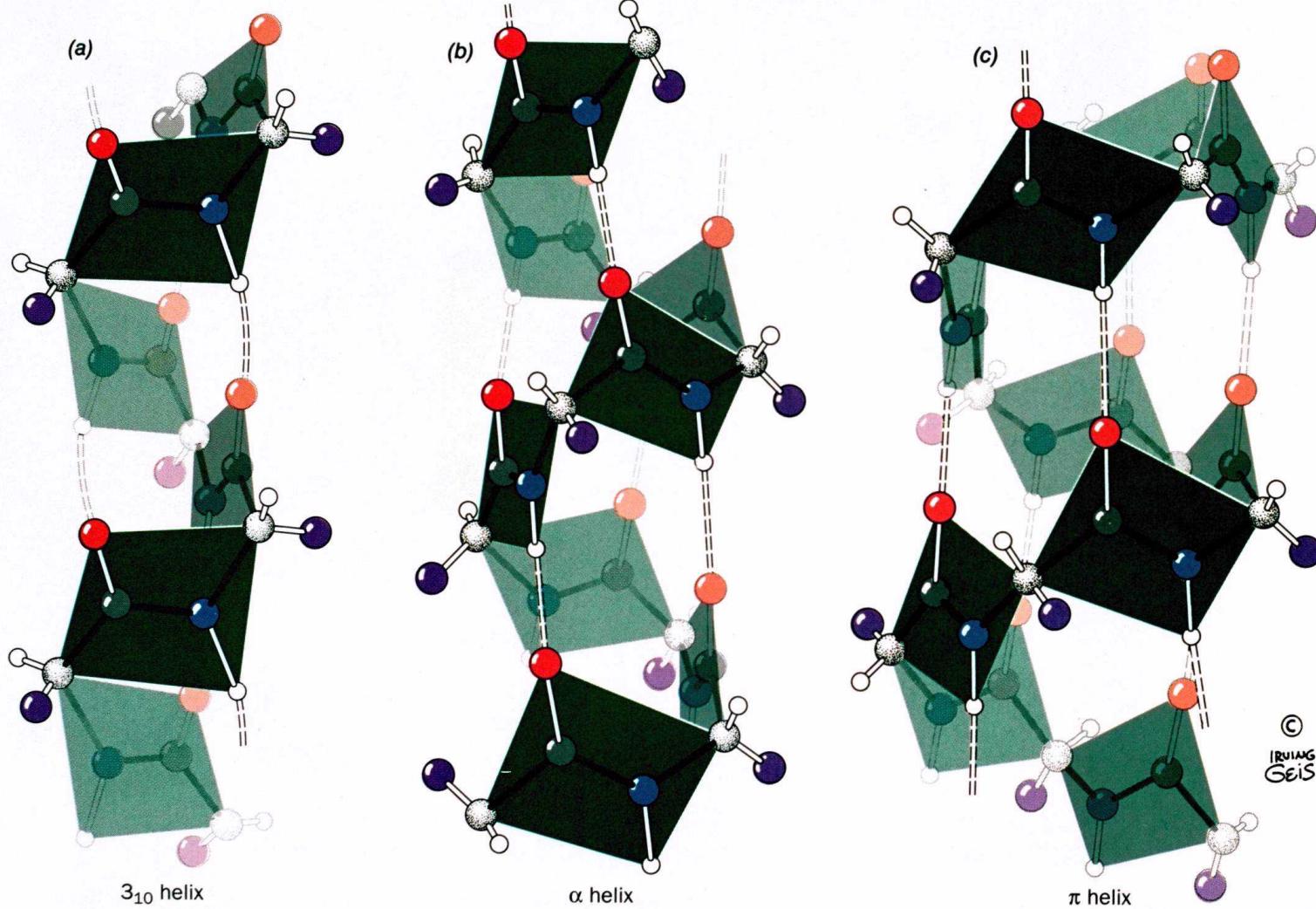


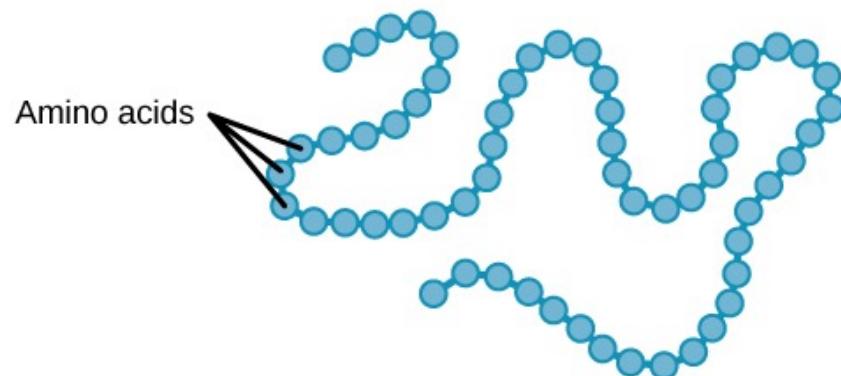
## $\pi$ helix

4.1 residues per turn  
H-bond (i,i+5)

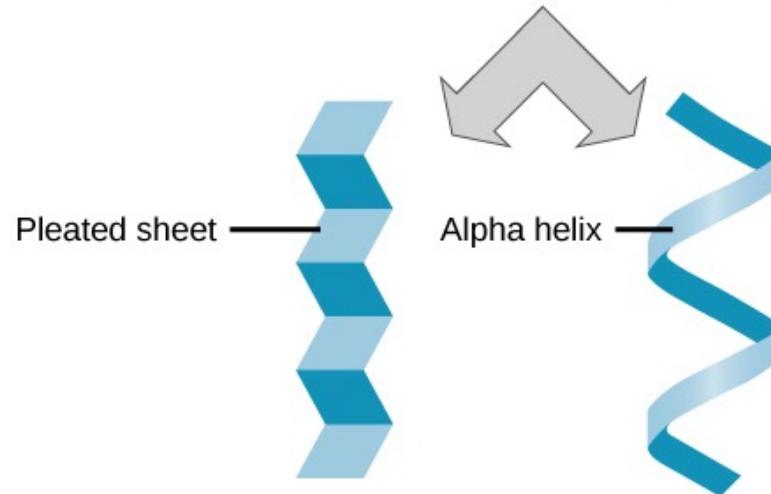
$$\phi, \psi = -55, -70$$





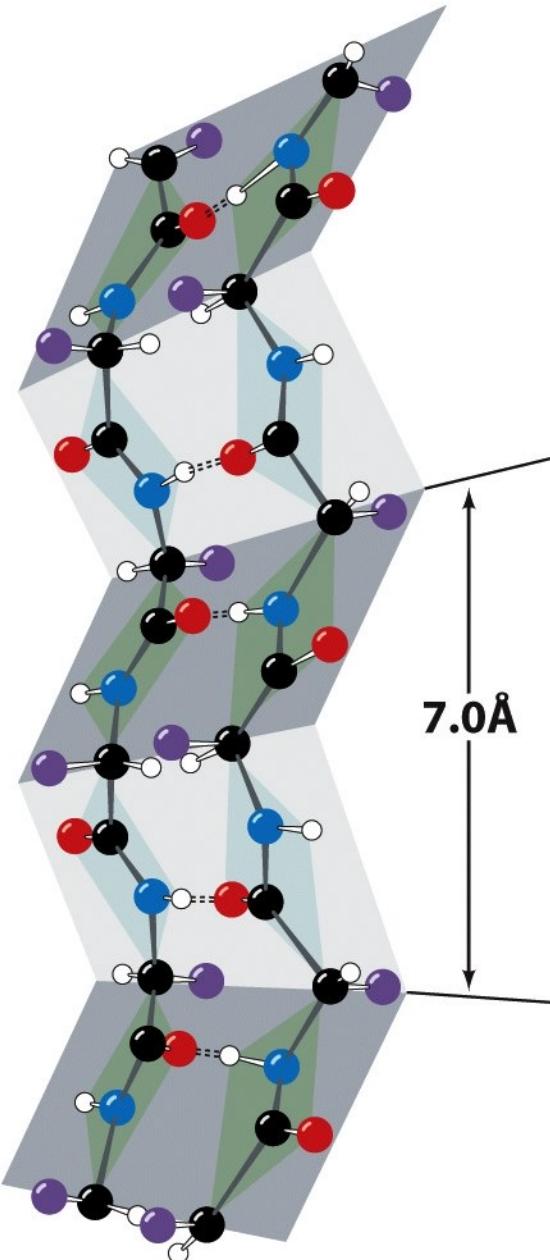


**Primary Protein structure**  
sequence of a chain of  
amino acids



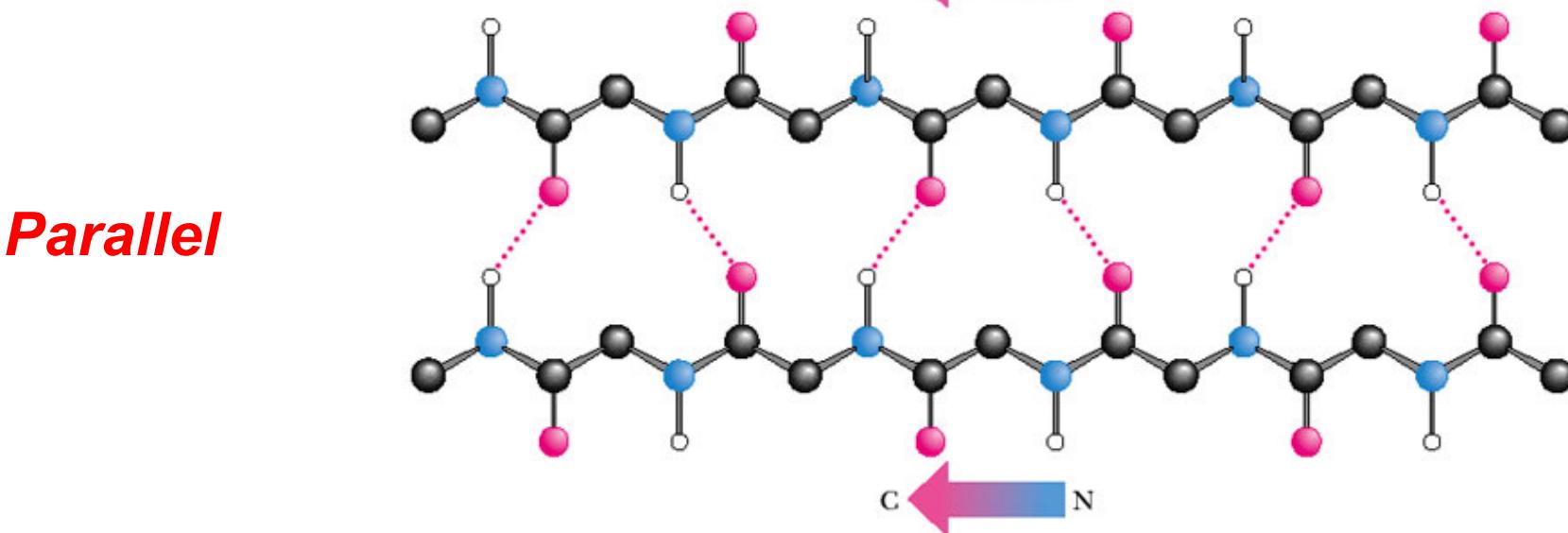
**Secondary Protein structure**  
hydrogen bonding of the peptide  
backbone causes the amino  
acids to fold into a repeating  
pattern

# $\beta$ sheet



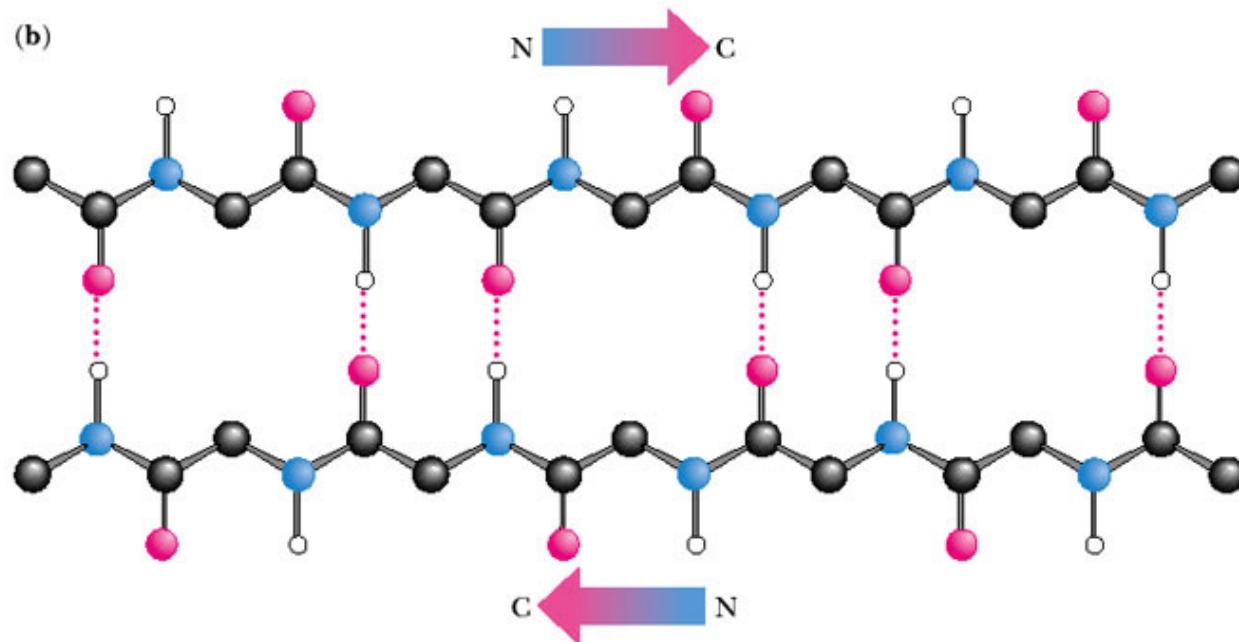
- $\beta$  strand: a single continuous stretch of amino acids involved in backbone hydrogen bonds another strand
- $\beta$  sheet is an assembly of at least two  $\beta$  strands.
- Backbone N-H groups of one strand form hydrogen bonds with backbone C=O groups of adjacent strands.
- Sidechain disposition: Successive side chains are above, then below the plane of the sheet.
- Parallel and Antiparallel: Antiparallel sheets have the strongest inter-strand H-honds, which are linear, which is their preferred geometry.
- $\phi = -110^\circ$  to  $-140^\circ$  ,  $\psi = +110^\circ$  to  $+135^\circ$

(a)



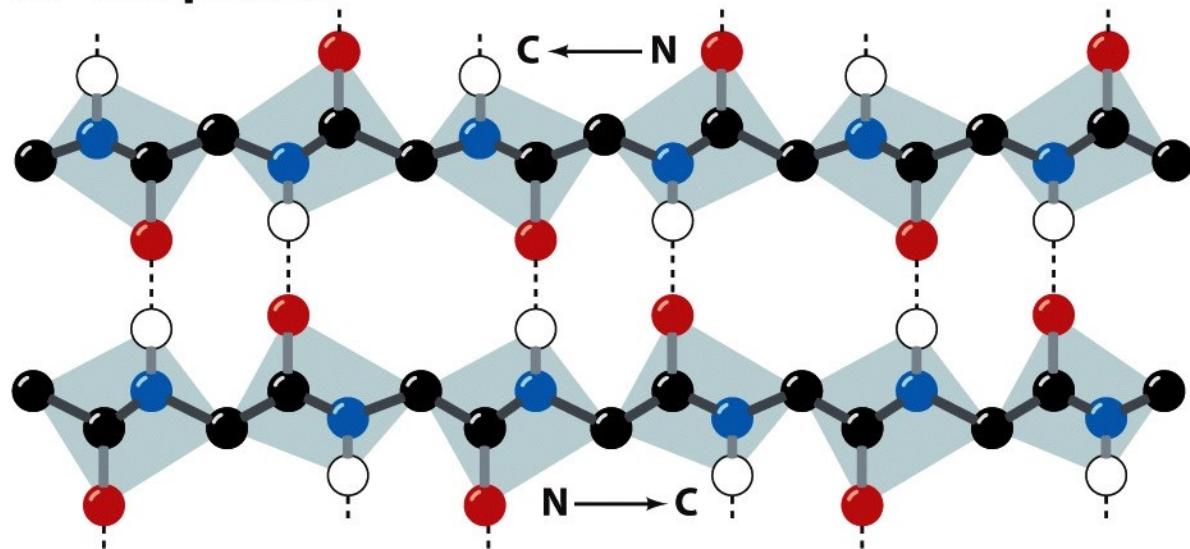
*Parallel*

(b)

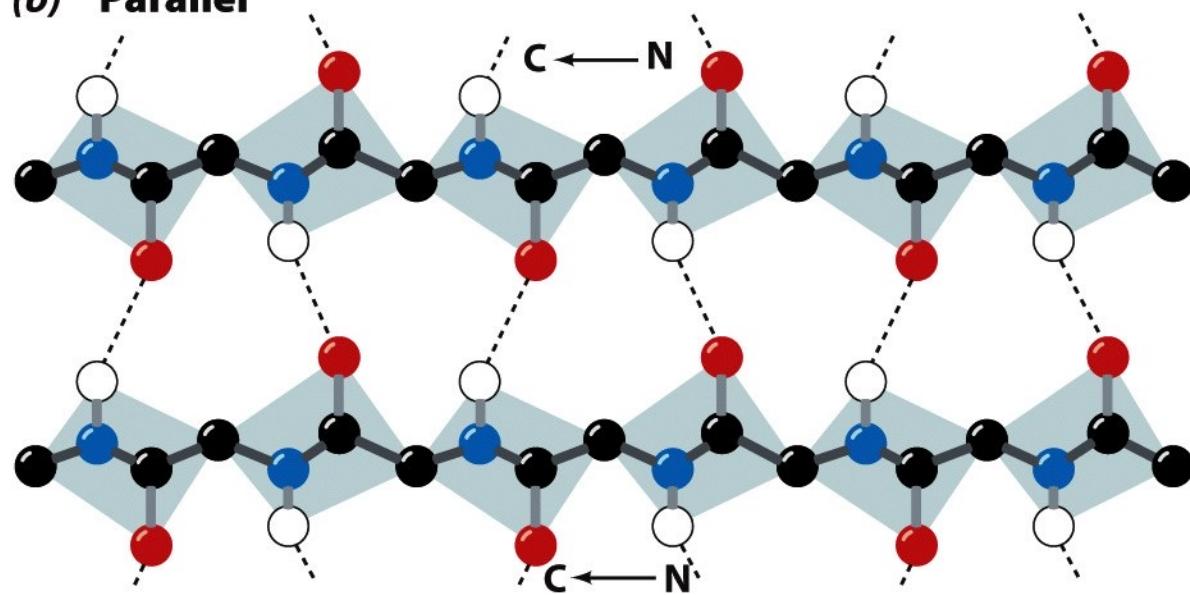


*Anti-parallel*

**(a) Antiparallel**



**(b) Parallel**



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

# $\beta$ -Pleated Sheet

***The backbone in the  $\beta$ -pleated sheet is extended***

- Polypeptide chains lie adjacent to one another; may be **parallel** (same direction) or **antiparallel** (opposite direction)
- Hydrogen bonds run perpendicular to the direction of the protein chain not parallel as in a helix
- Backbone H-bonding between C=O and N-H on adjacent strands can be formed between different parts of a single chain that is double backed on it self  $\Rightarrow$  **intrachain bonds**
- different chains  $\Rightarrow$  **interchain bonds**

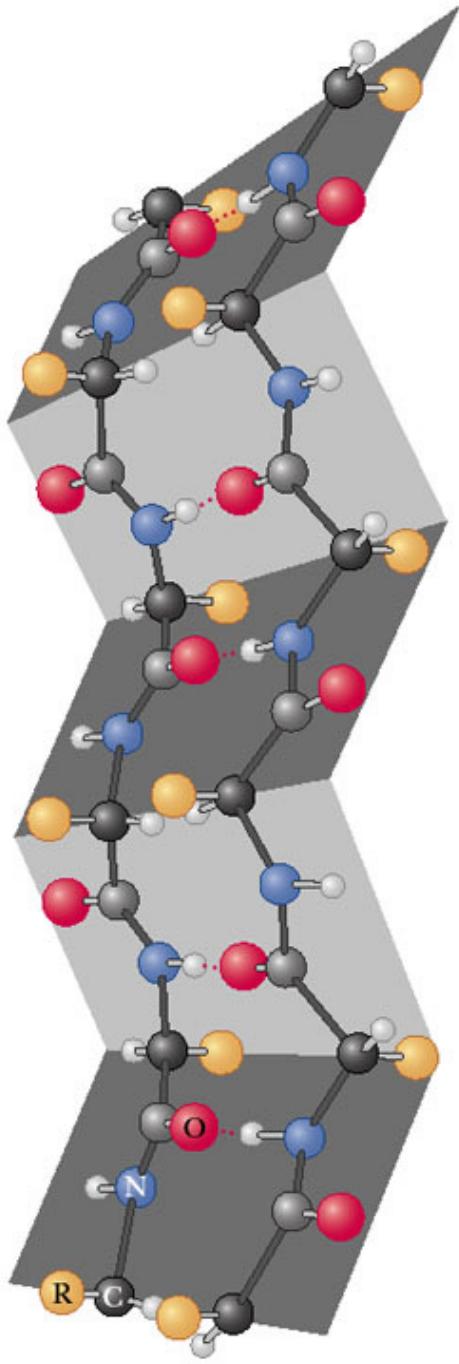
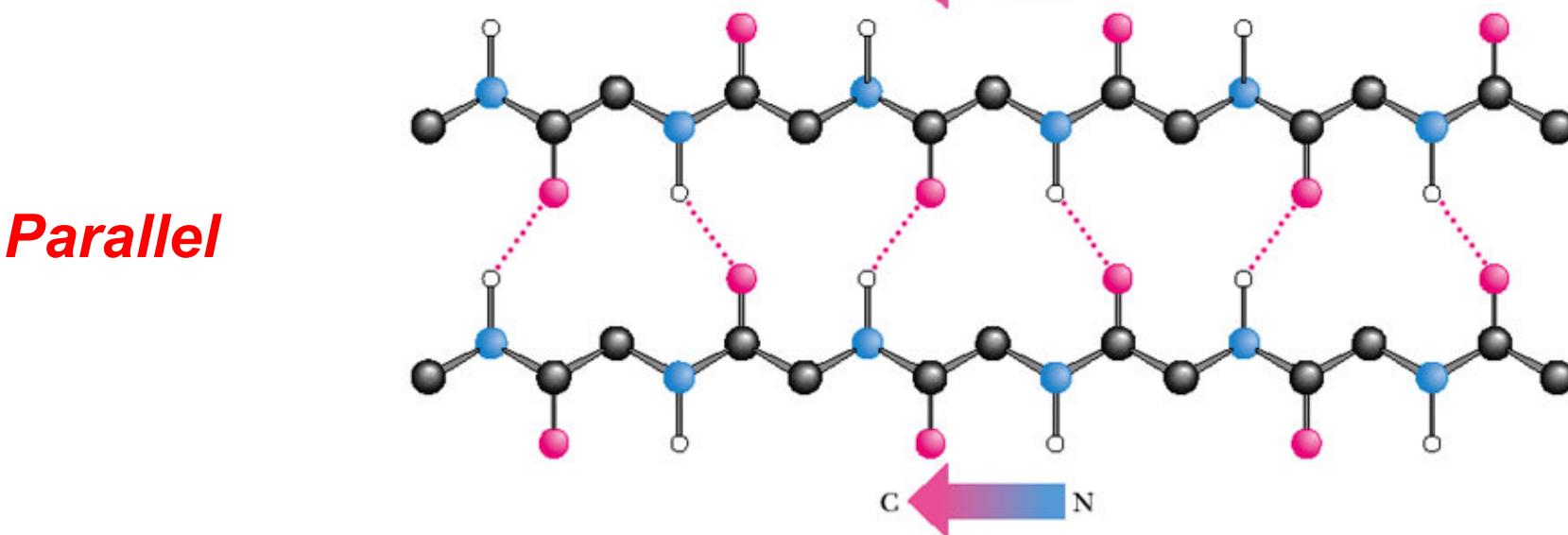


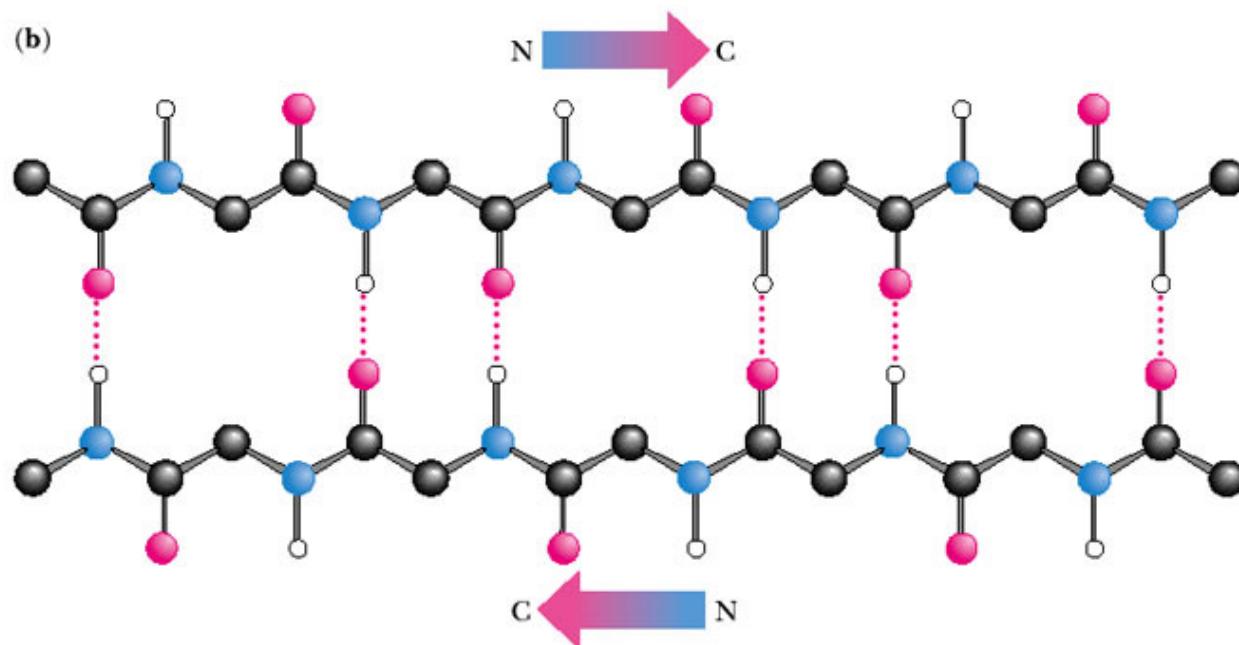
Fig. 4-5, p.87

(a)



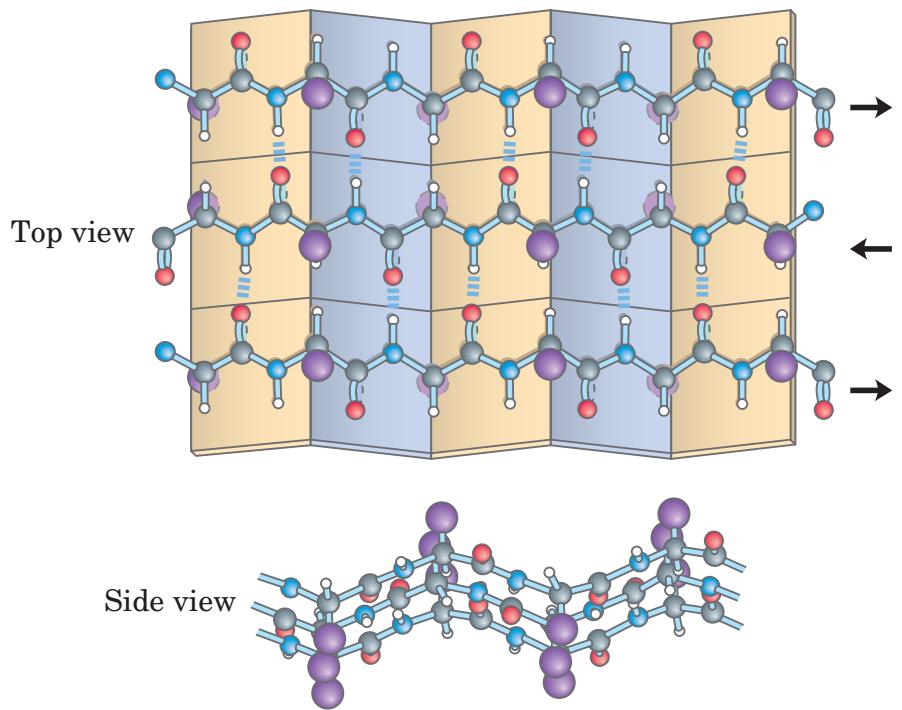
*Parallel*

(b)

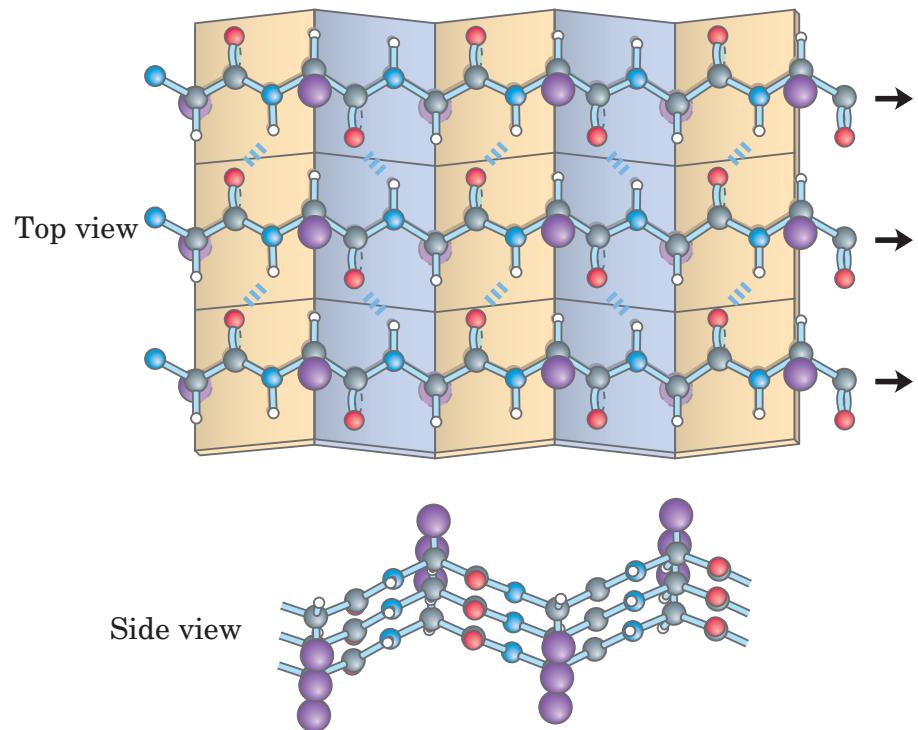


*Anti-parallel*

**(a) Antiparallel**



**(b) Parallel**



**Antiparallel sheets often use  $\beta$  turns  
Parallel sheets use  $\beta\alpha\beta$**

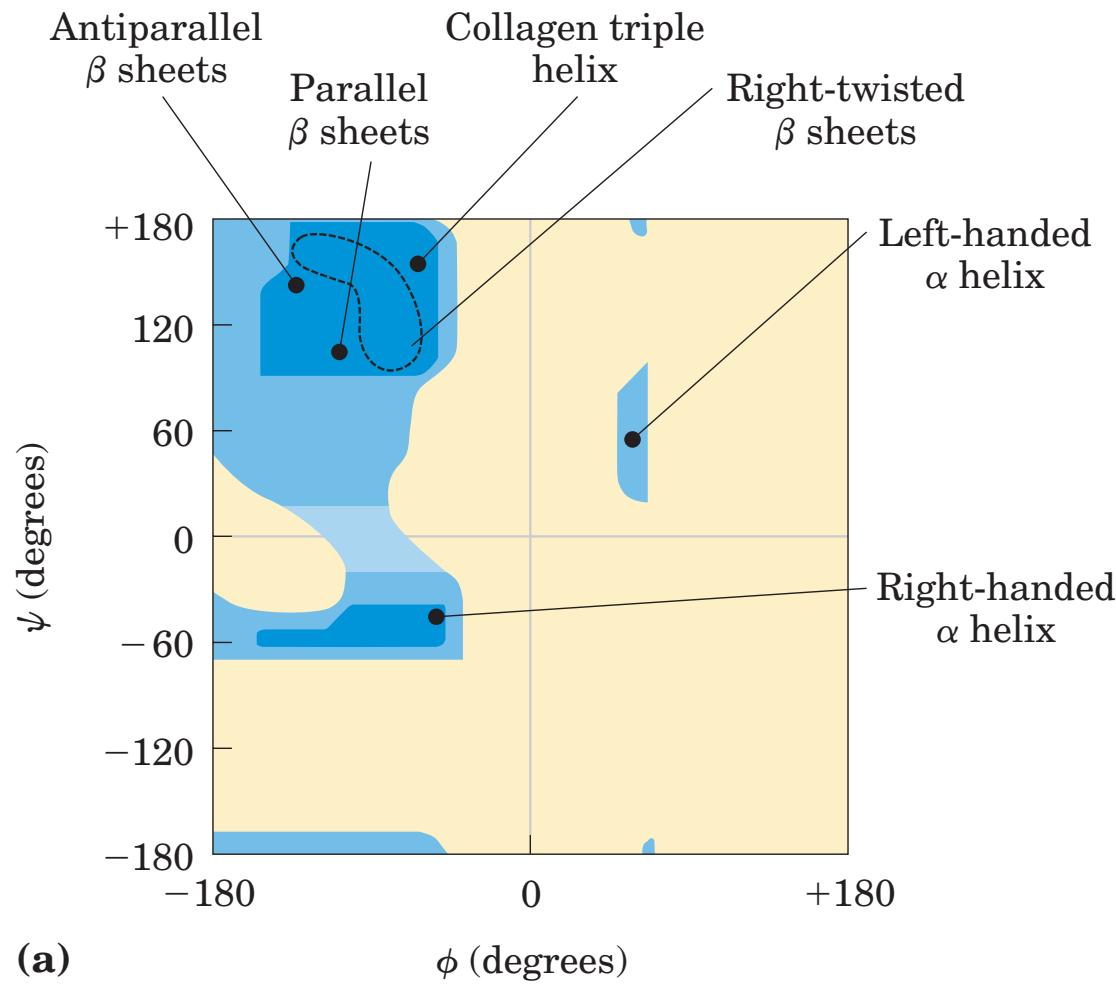
**(a)**



**(b)**

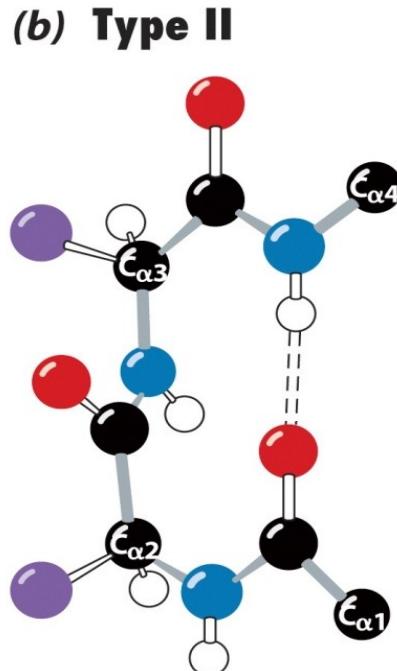
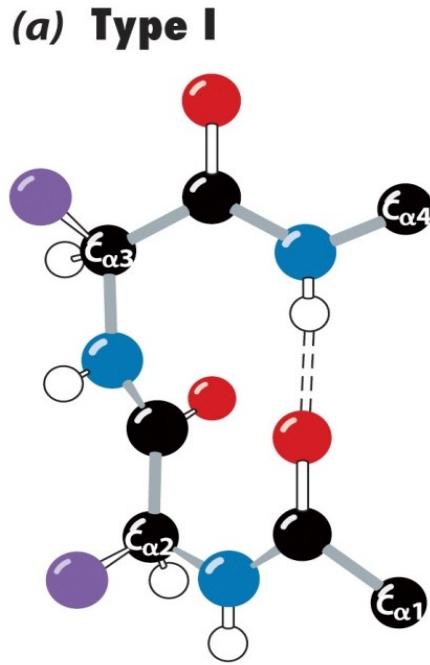


# Ramachandran plot



# Non-repetitive regions

- **Turns - coils or loops**
  - 50% of structure of globular proteins are not repeating structures
- **b-bends**
  - Type I and type II: hairpin turn between anti-parallel sheets



Type I

$$\phi_2 = -60^\circ, \psi_2 = -30^\circ$$

$$\phi_3 = -90^\circ, \psi_3 = 0^\circ$$

Type II

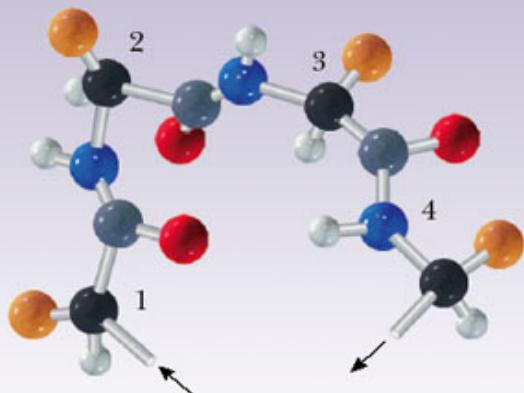
$$\phi_2 = -60^\circ, \psi_2 = 120^\circ$$

$$\phi_3 = 90^\circ, \psi_3 = 0^\circ$$

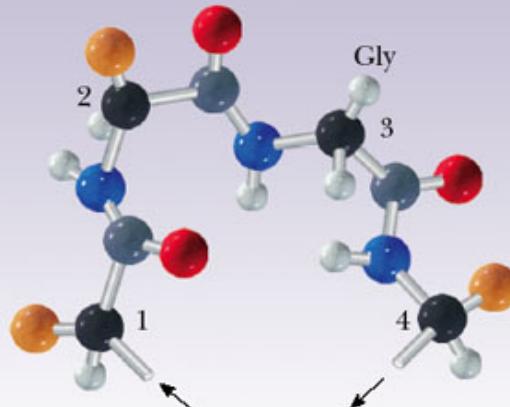
# Structures of Reverse Turns

- Protein folding required that peptide backbones structures be able to change direction.
- Reverse turns mark a transition between one secondary structure to another
- Glycine and proline are involved in the reverse turns
- Glycine a single side chain which prevents crowding
- Proline has a cyclic structure and therefore, has a correct geometry for reverse turn

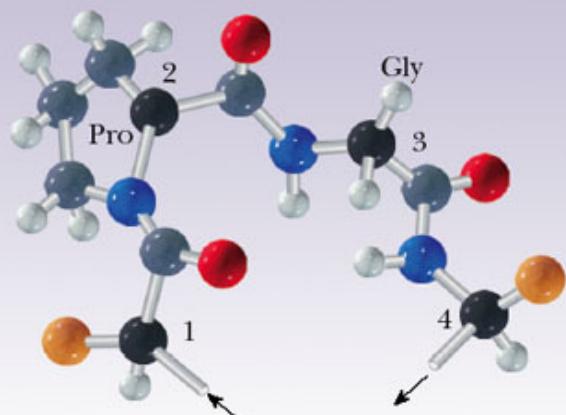
(a) Type I



(b) Type II



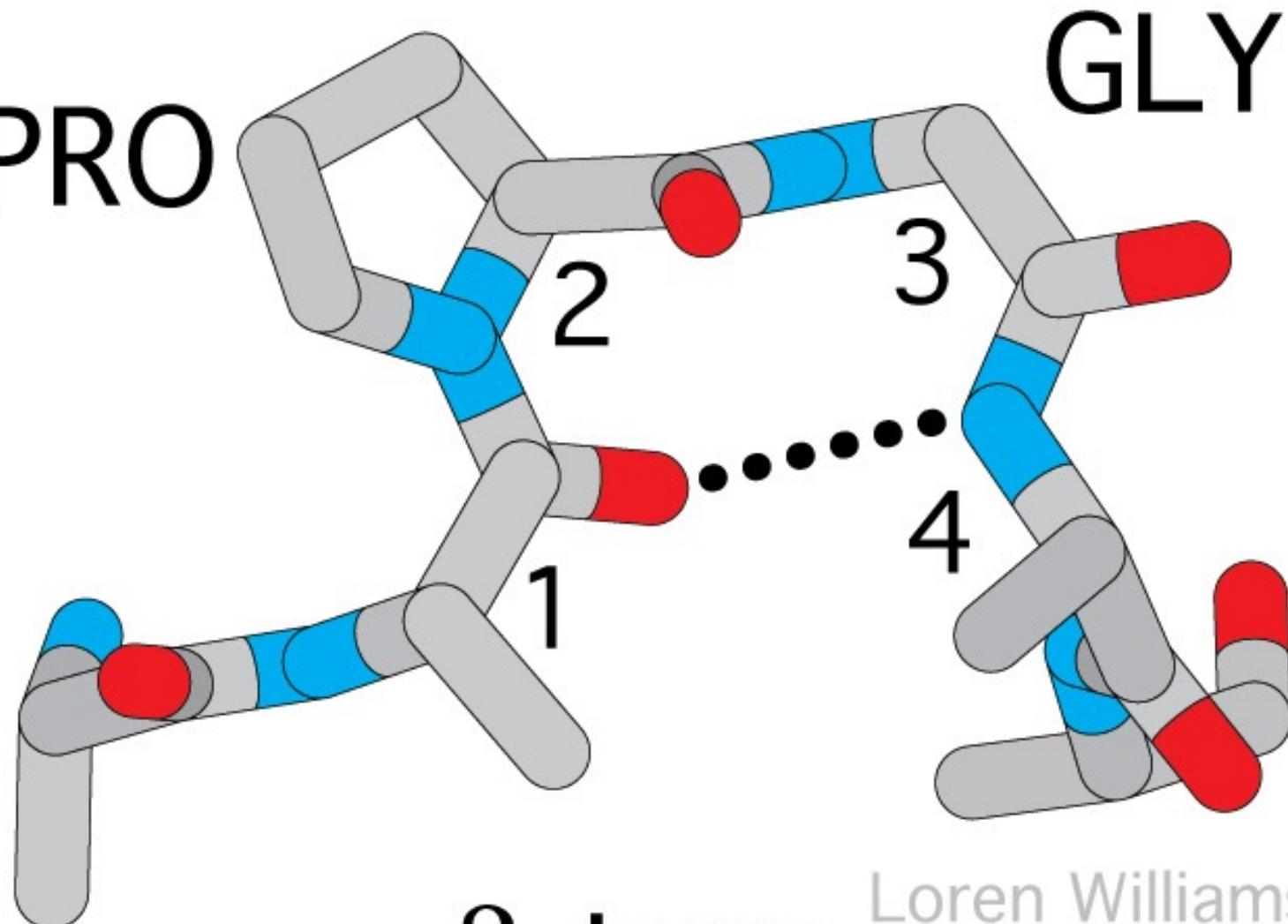
(c) Type II (proline-containing)



# The $\beta$ -turn

- A simple protein structural motif
- Generally found on protein surfaces
- Turns the backbone around by 180 degrees to fold back on itself
- Links two anti-parallel two beta strands
  - That are adjacent in primary structure,
  - That are adjacent in secondary structure
- Is a short loop of two to five amino acids
- Frequently contains glycine at position 3, (small H side chain allows extreme phi-psi).
- Frequently contains proline at position 2, (p forces the backbone into the appropriate conformation)

PRO



$\beta$ -turn

Loren Williams

# *Sidechain propensities*

**Table 6-1**

Propensities of Amino Acid Residues for  $\alpha$  Helical and  $\beta$  Sheet Conformations

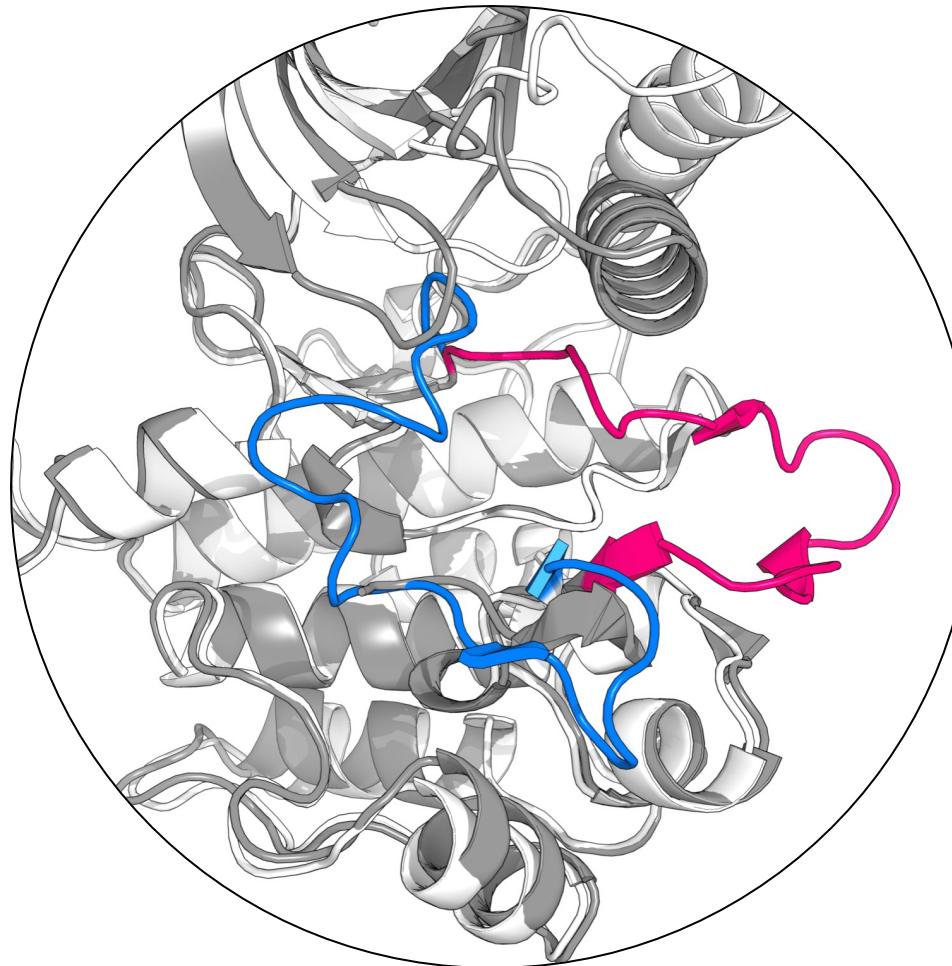
Residue	$P_\alpha$	$P_\beta$
Ala	1.42	0.83
Arg	0.98	0.93
Asn	0.67	0.89
Asp	1.01	0.54
Cys	0.70	1.19
Gln	1.11	1.10
Glu	1.51	0.37
Gly	0.57	0.75
His	1.00	0.87
Ile	1.08	1.60
Leu	1.21	1.30
Lys	1.16	0.74
Met	1.45	1.05
Phe	1.13	1.38
Pro	0.57	0.55
Ser	0.77	0.75
Thr	0.83	1.19
Trp	1.08	1.37
Tyr	0.69	1.47
Val	1.06	1.70

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

# Loops

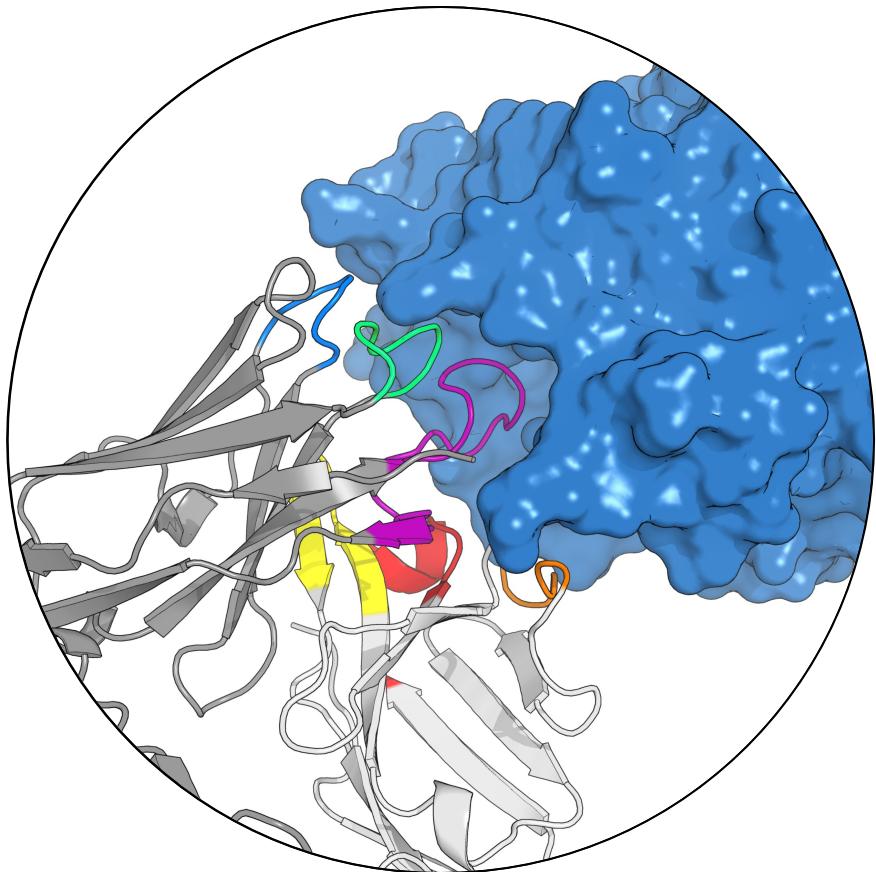
- Most proteins consist of several segments of  $\alpha$ -helix and/or  $\beta$  strands separated from each other by various loop regions.
- In general, loops are present on the surface vary in length and shape, and allow the overall polypeptide to fold into a compact tertiary structure.
- Loops are rich in polar/charged amino acid residues that, along with the N—H and C=O groups of their associated peptide bonds, hydrogen bond with the surrounding water molecules.
- Loop regions are flexible and participate or contribute directly to the polypeptide's biological function. For example, the antigen-binding region of antibodies and the active site of enzymes.

## Functional roles of loops

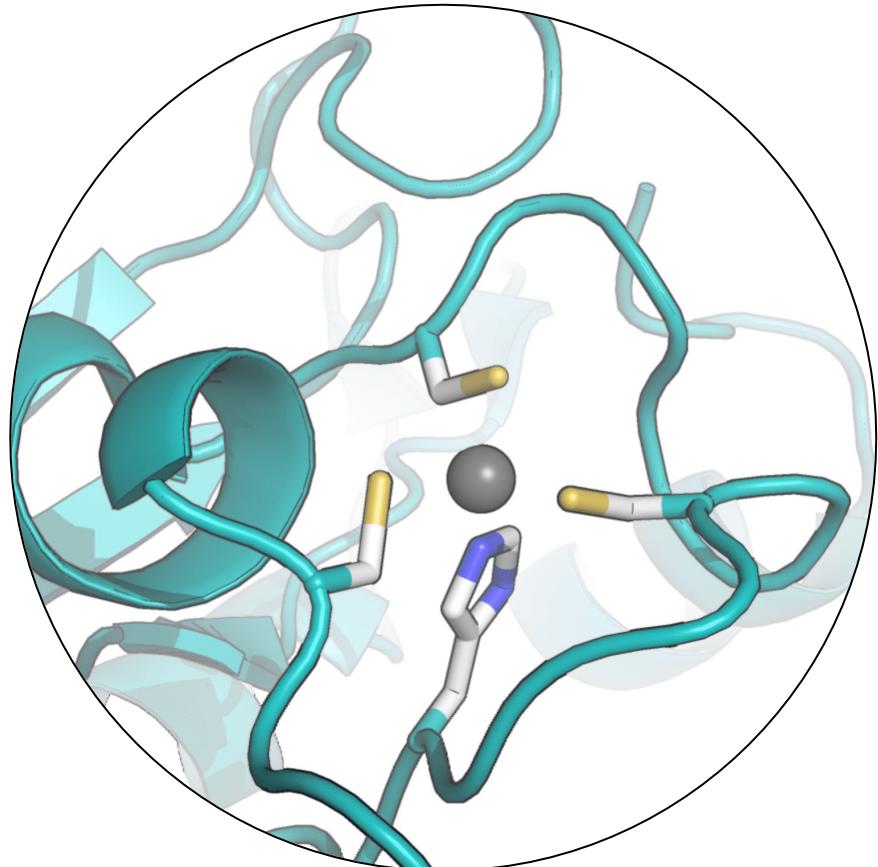


The activation loop of a tyrosine kinase has a different conformation in the active (pink) and inactive (blue) forms

## Functional roles of loops

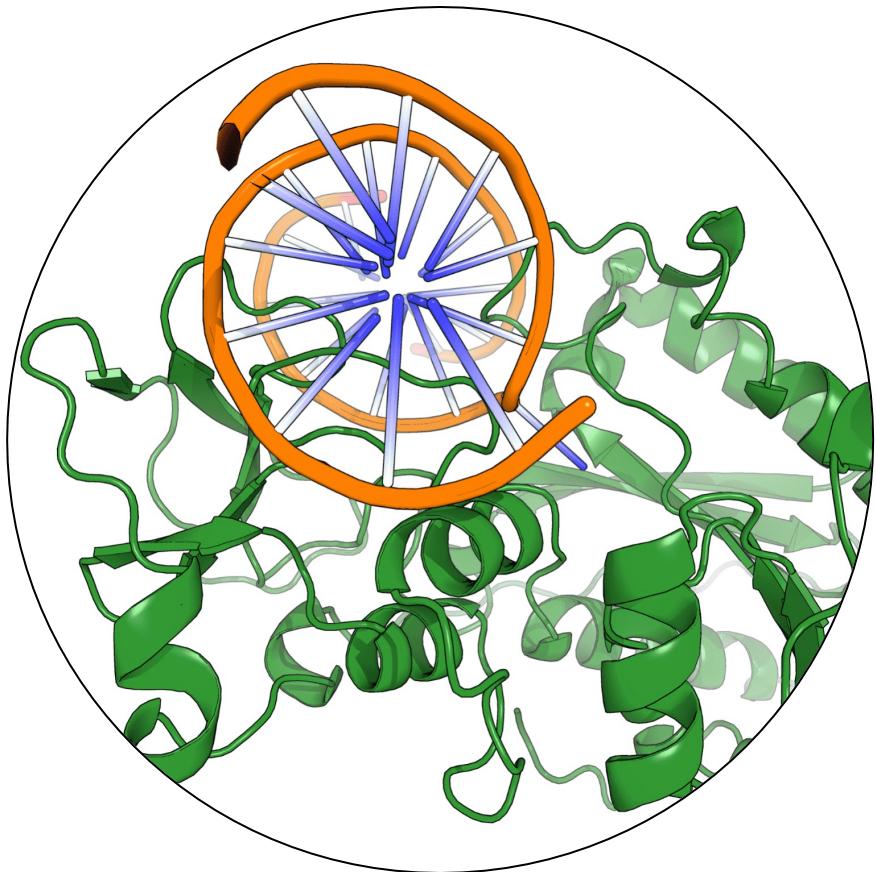


An antibody binding  
to its antigen

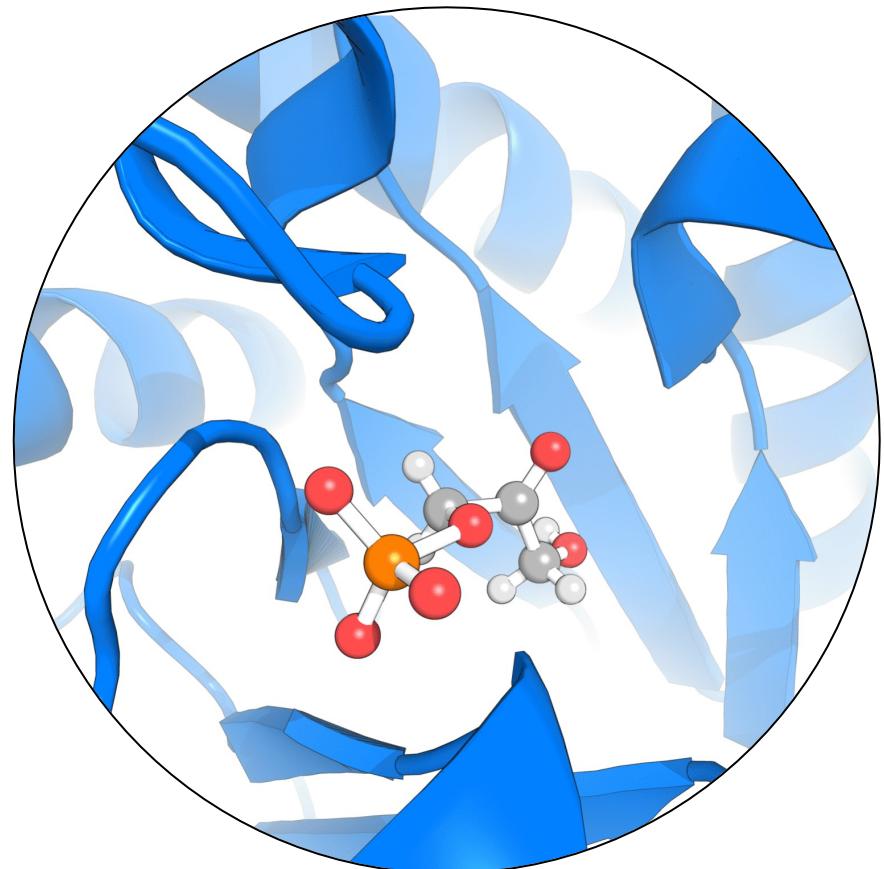


A zinc ion is coordinated by  
the sidechain atoms of a loop

## Functional roles of loops



Methyltransferase  
binding to DNA



The active site of a  
triosephosphate isomerase

# Super secondary structures (motifs)

- Super secondary structures as combinations of alpha-helices and beta-structures connected through loops, that form patterns that are present in many different protein structures.
- These folding patterns are stabilized through the same kind of linkages than the tertiary level.
- Sometimes the term “motif” is used to describe these super secondary structures.

# Super secondary structures

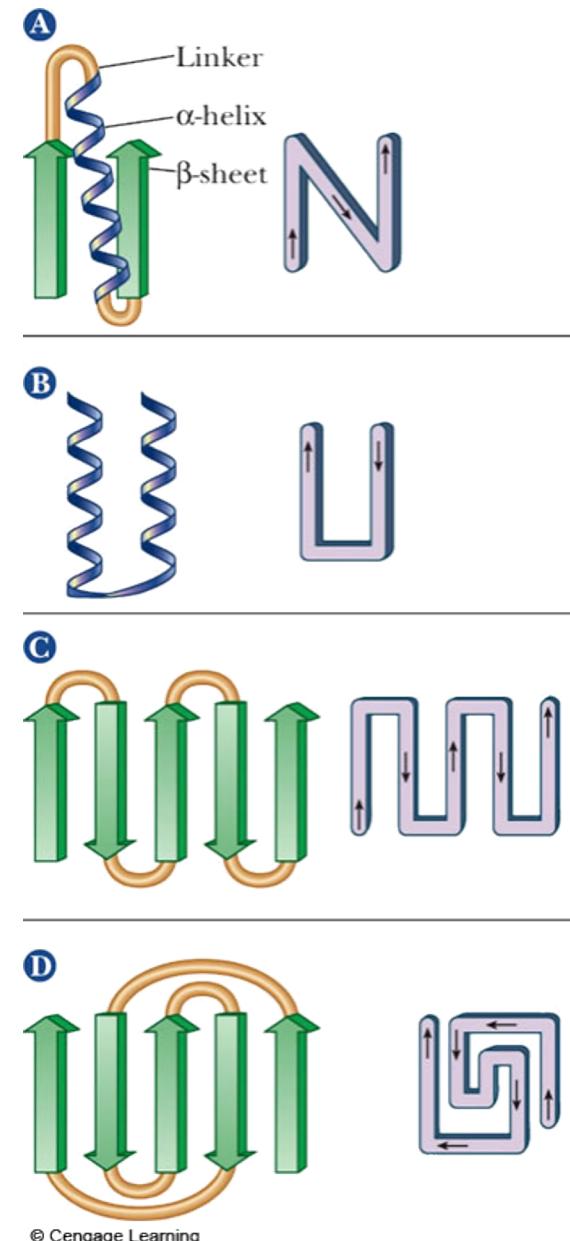
A combination of  $\alpha$ - and  $\beta$ -pleated sheets of protein as it is folded back on itself:

$\beta\alpha\beta$ : two parallel strands of  $\beta$ -sheet connected by a stretch of  $\alpha$ -helix (A)

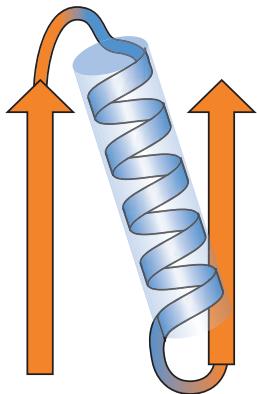
$\alpha\alpha$  unit: two antiparallel  $\alpha$ -helices (Helix-turn-helix) (B)

$\beta$ -meander: an antiparallel sheet formed by a series of tight reverse turns connecting stretches of a polypeptide chain (C)

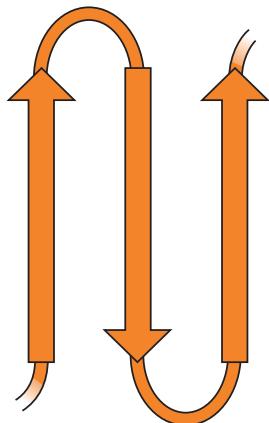
Greek key: a repetitive supersecondary structure formed when an antiparallel sheet doubles back on itself



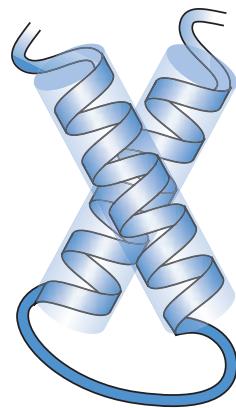
# Super secondary structures



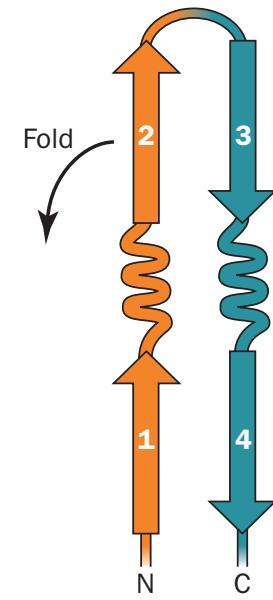
$\beta\alpha\beta$



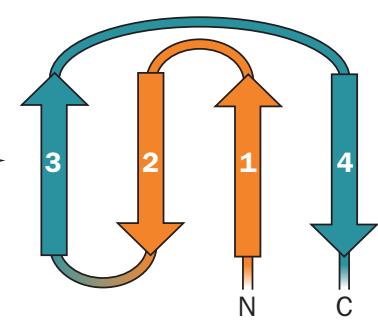
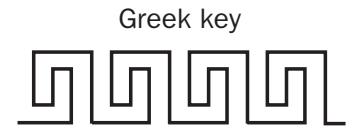
$\beta$  hairpin



$\alpha\alpha$



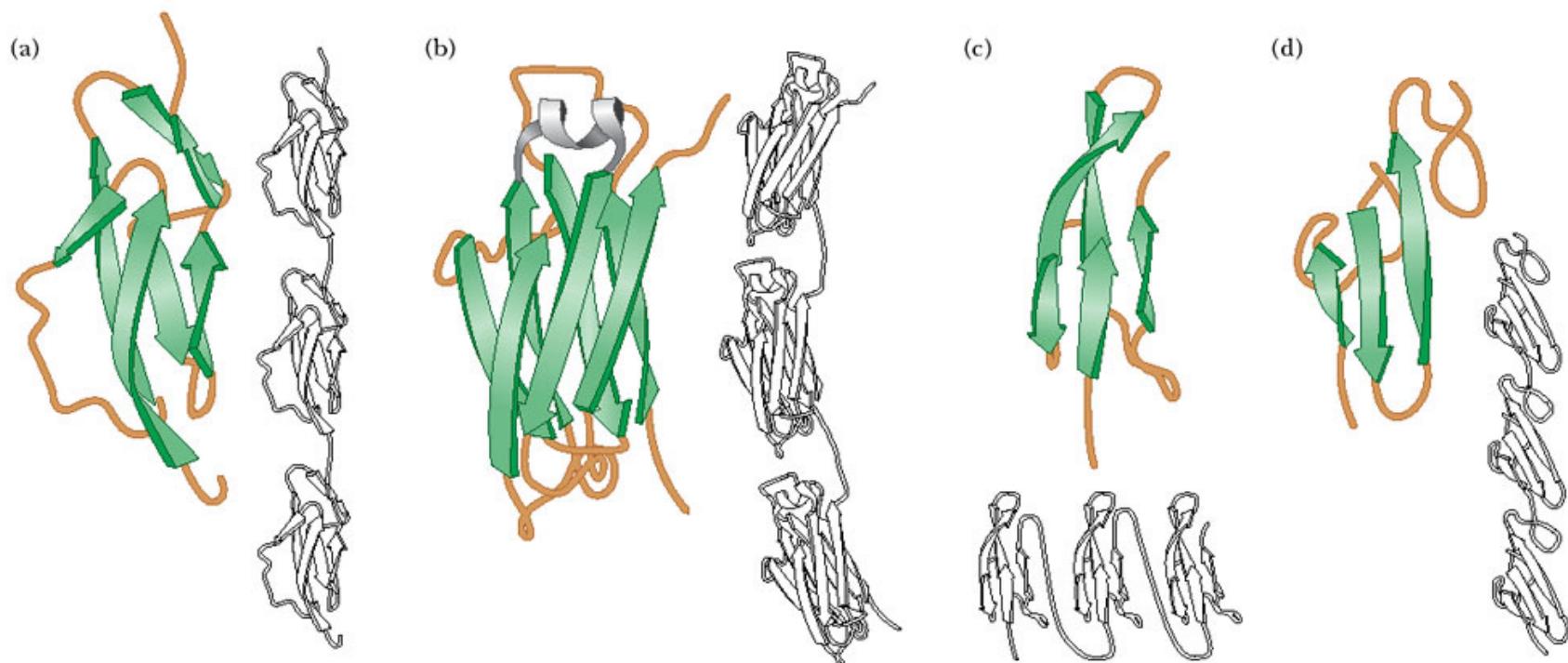
Greek key



# Super secondary structures

**Motifs:** a particular secondary structure that is repeated in the a protein

Give information about the folding of protein but do not allow for the prediction of the biological activity.



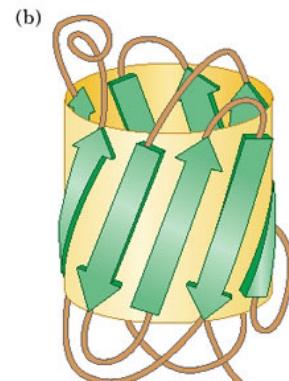
# Some $\beta$ -barrel arrangements

- When  $\beta$ -sheets are extensive enough they can fold back on themselves, forming a  $\beta$ -barrel

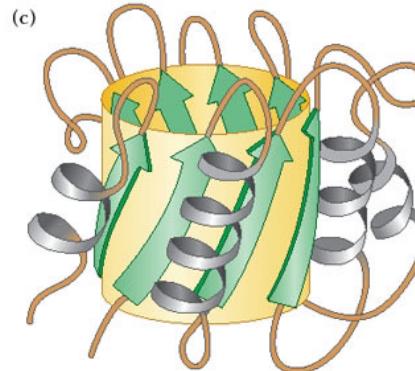
$\beta$ -meanders



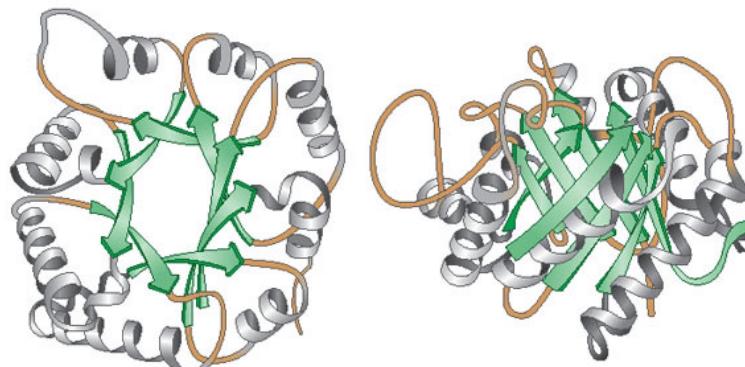
Greek key



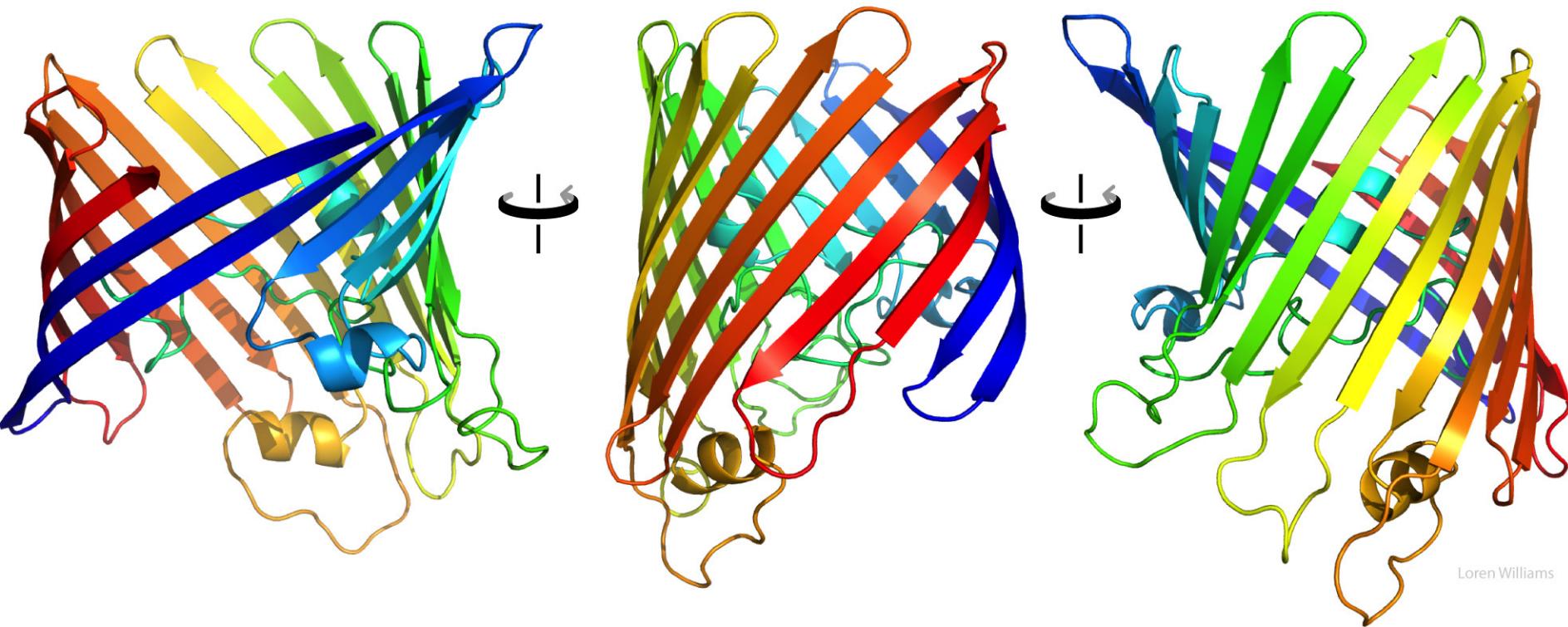
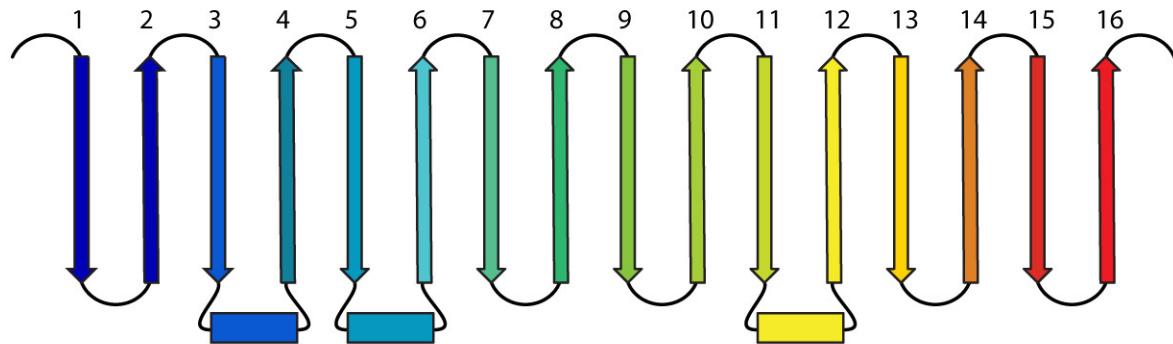
$\beta\alpha\beta$  units



(d)

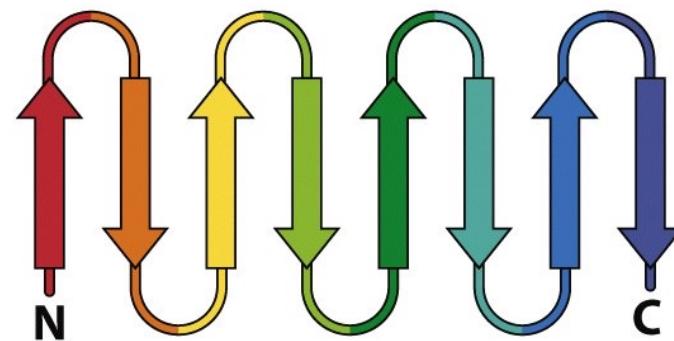
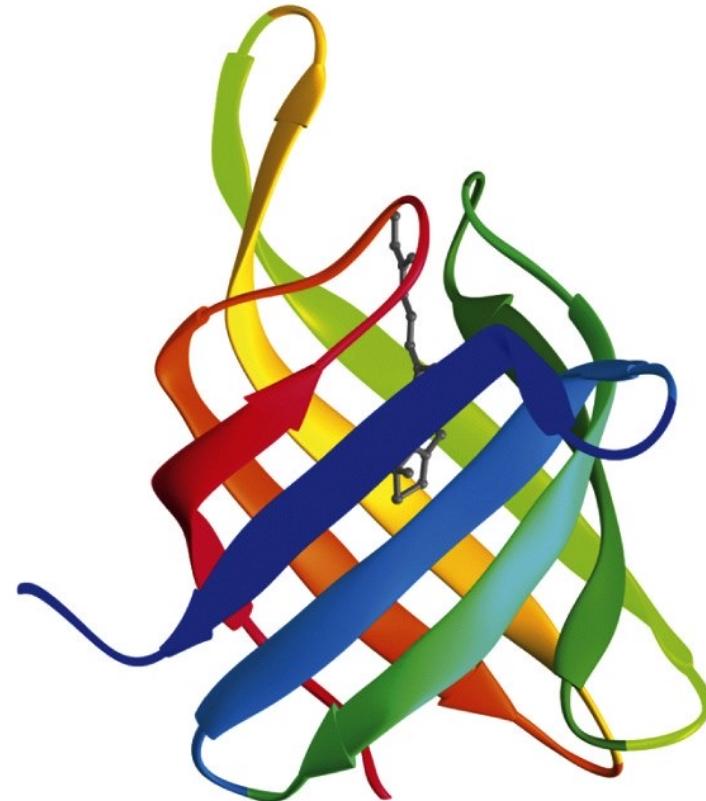
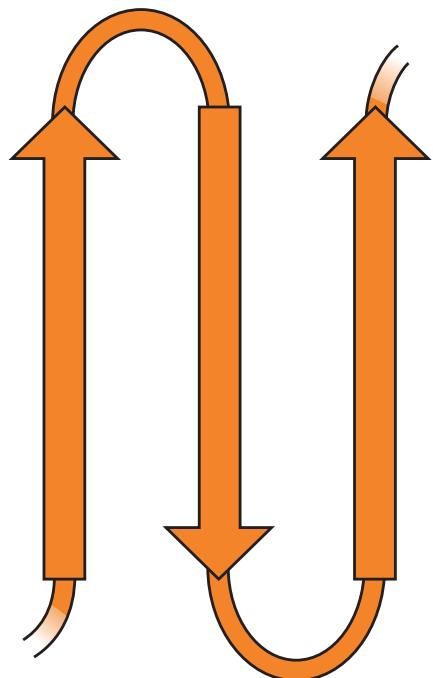


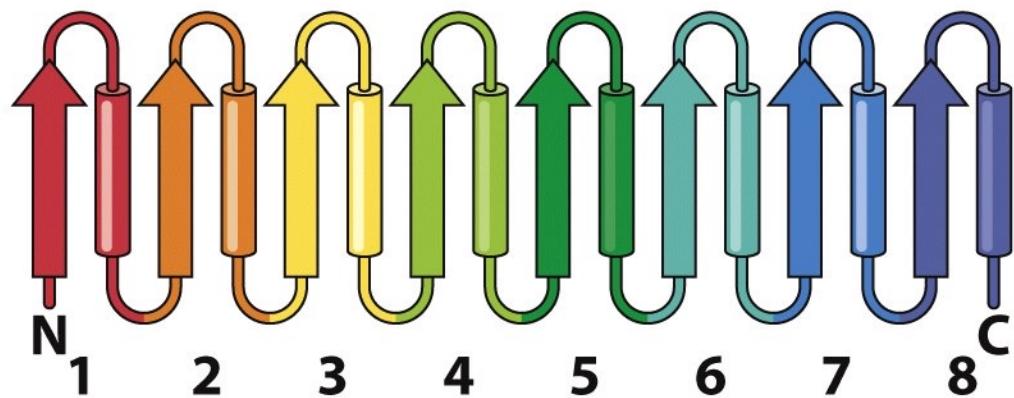
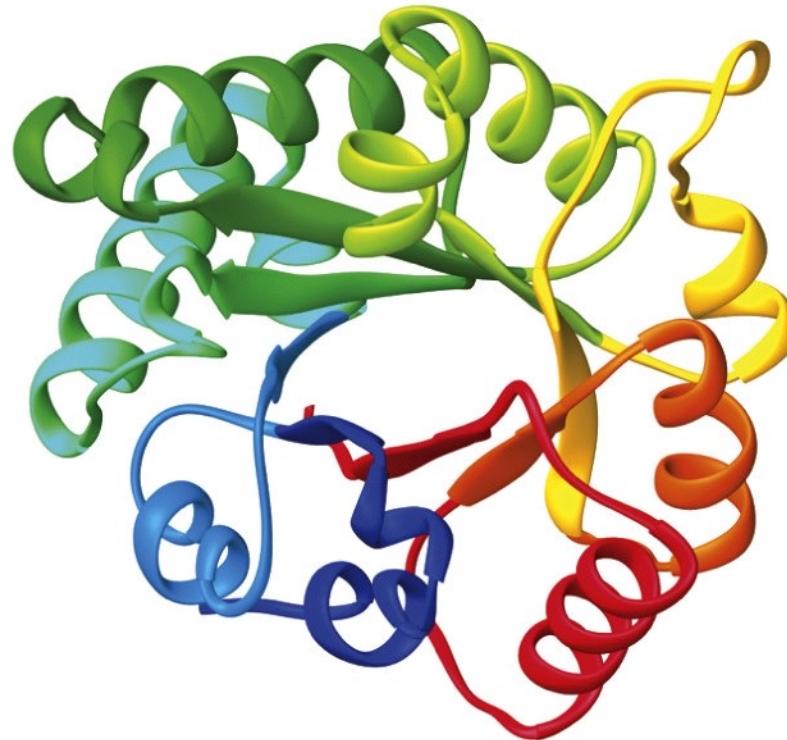
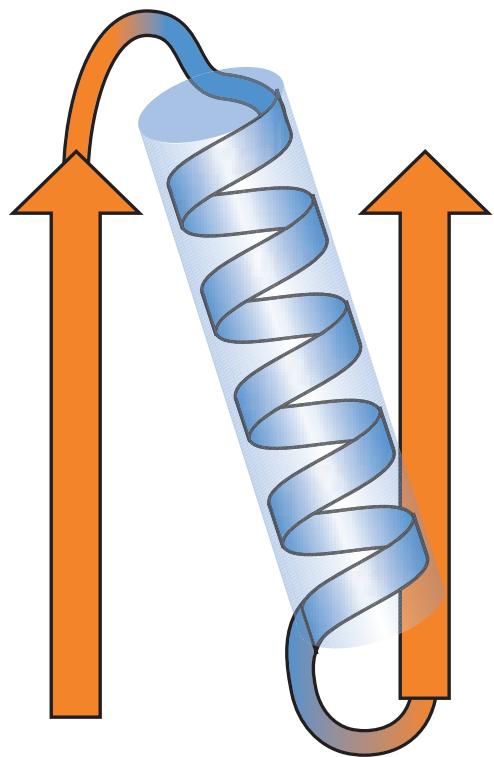
Topology of Porin



Loren Williams

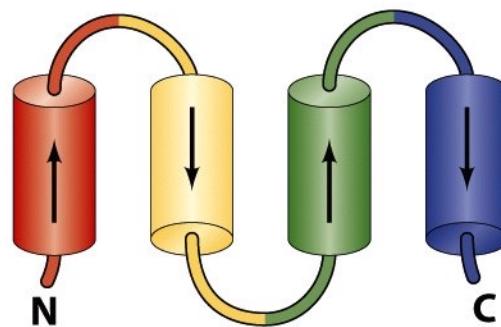
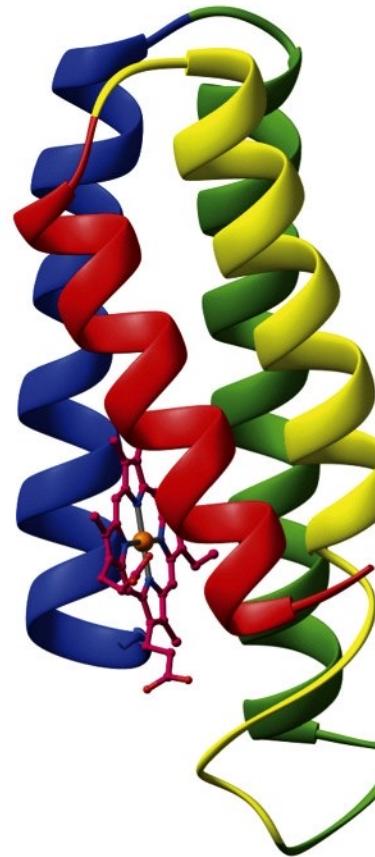
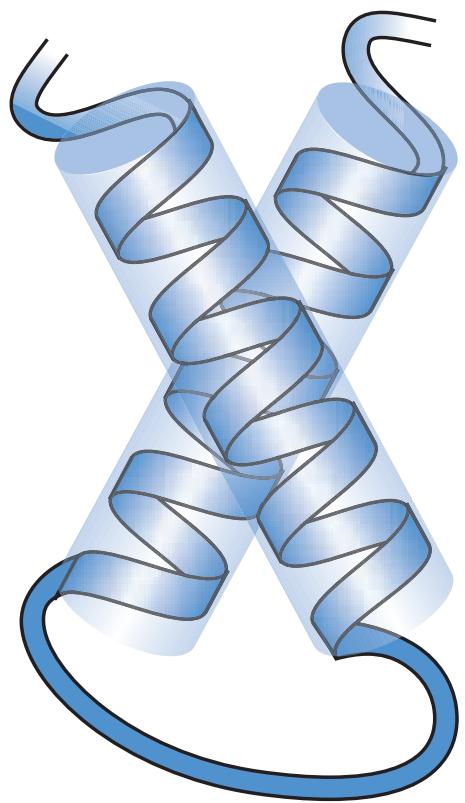
$\beta$  hairpin





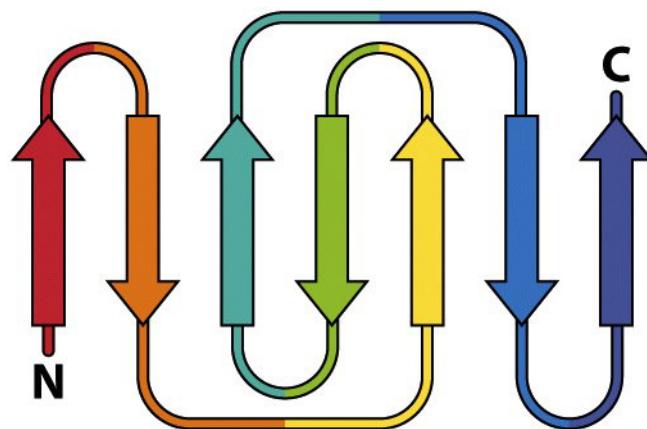
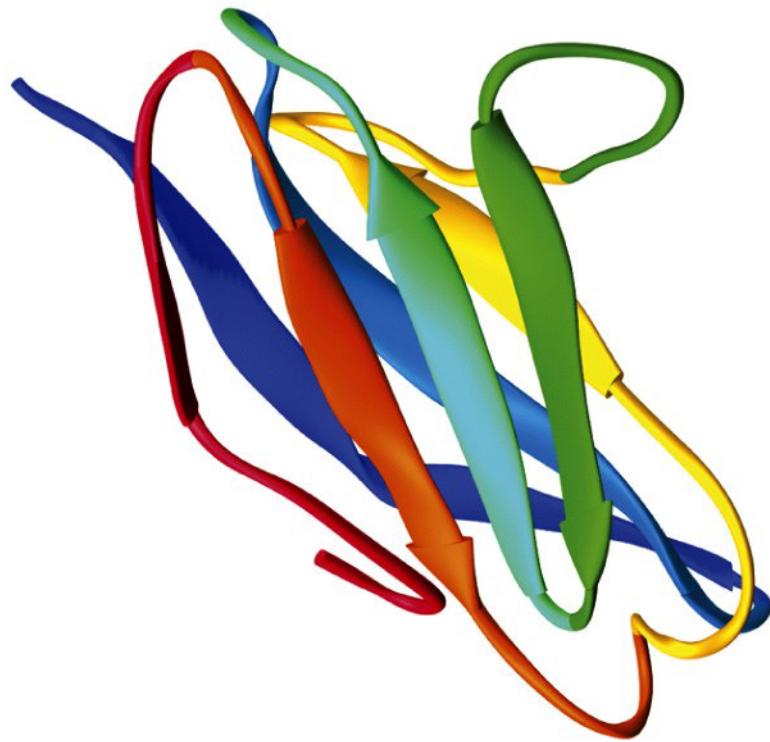
© 2008 John Wiley & Sons, Inc. All rights reserved.

Figure 6-30c



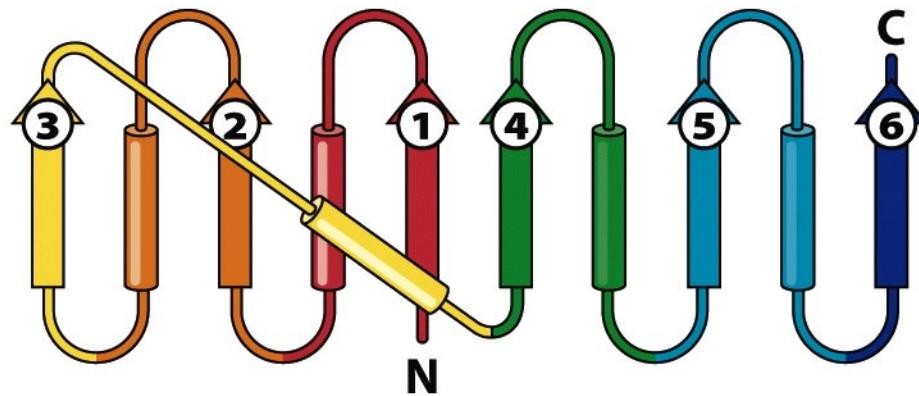
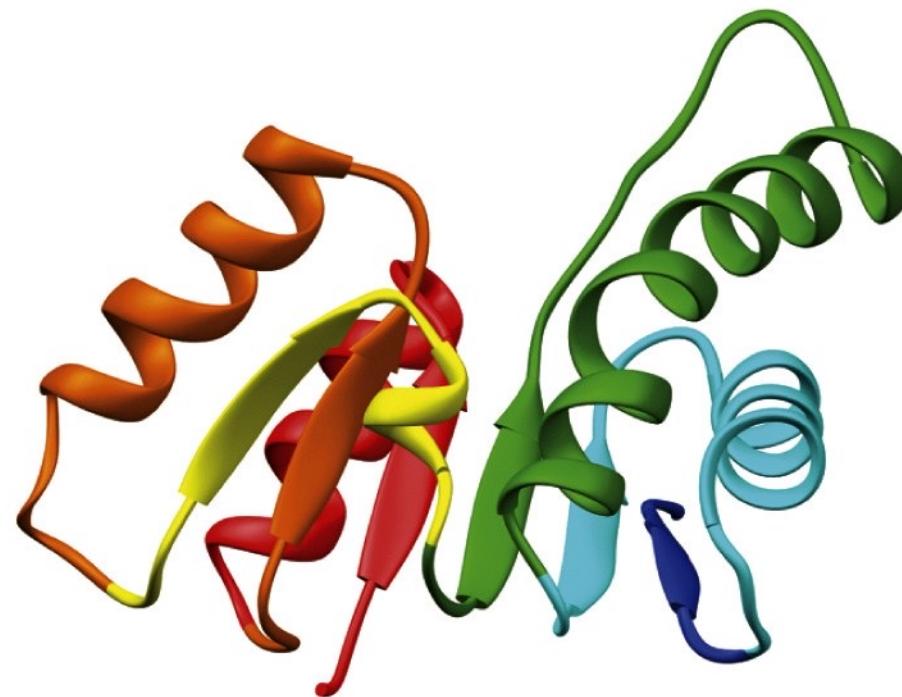
© 2008 John Wiley & Sons, Inc. All rights reserved.

Figure 6-29a



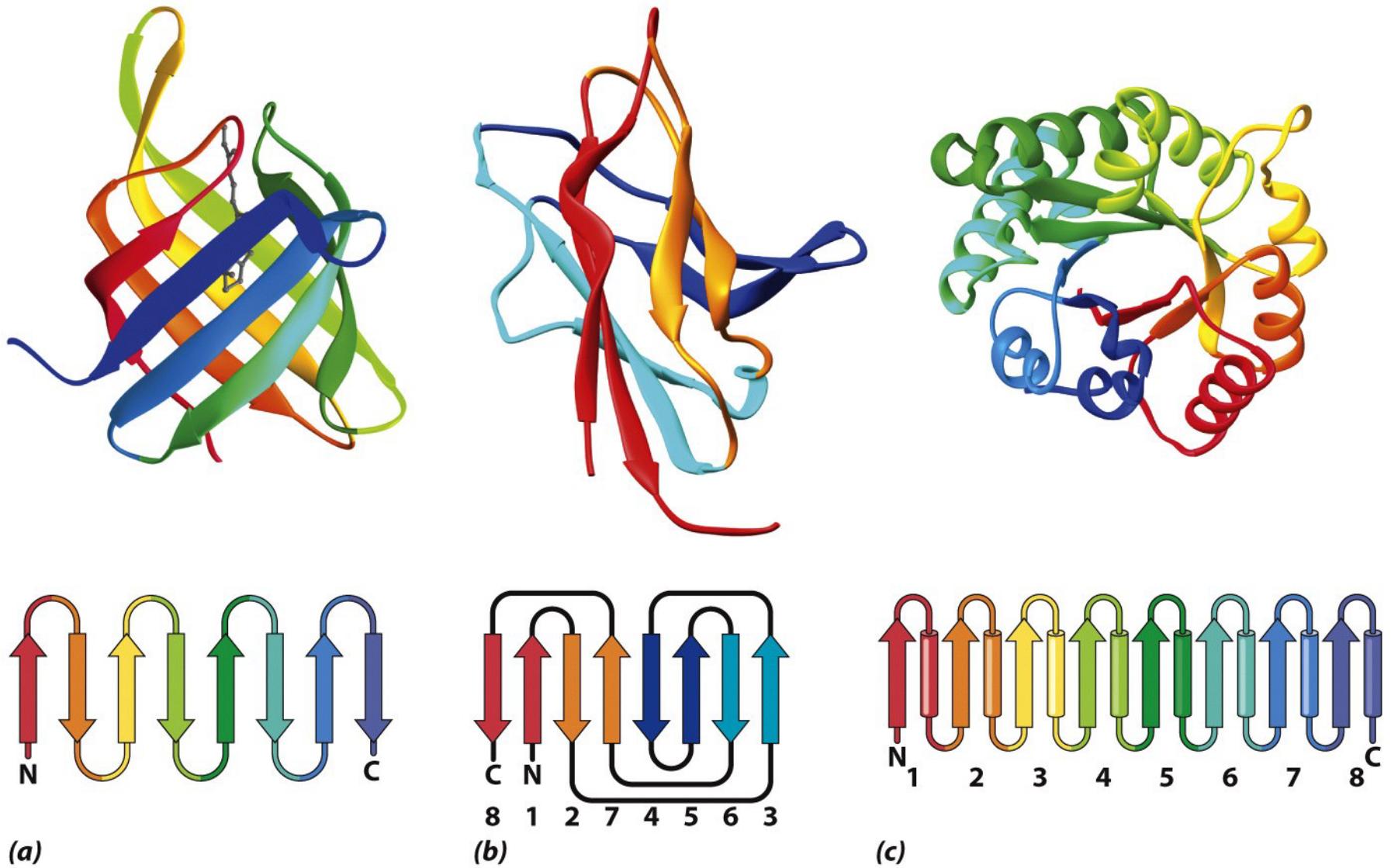
© 2008 John Wiley & Sons, Inc. All rights reserved.

Figure 6-29b



© 2008 John Wiley & Sons, Inc. All rights reserved.

Figure 6-29c

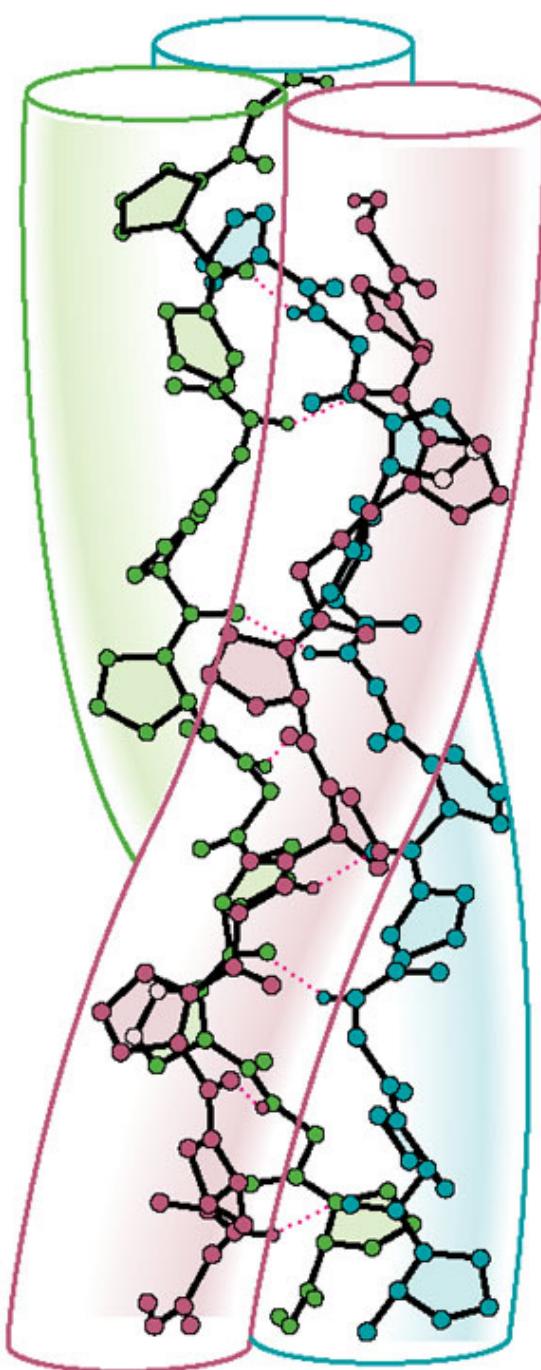


© 2008 John Wiley & Sons, Inc. All rights reserved.

Figure 6-30

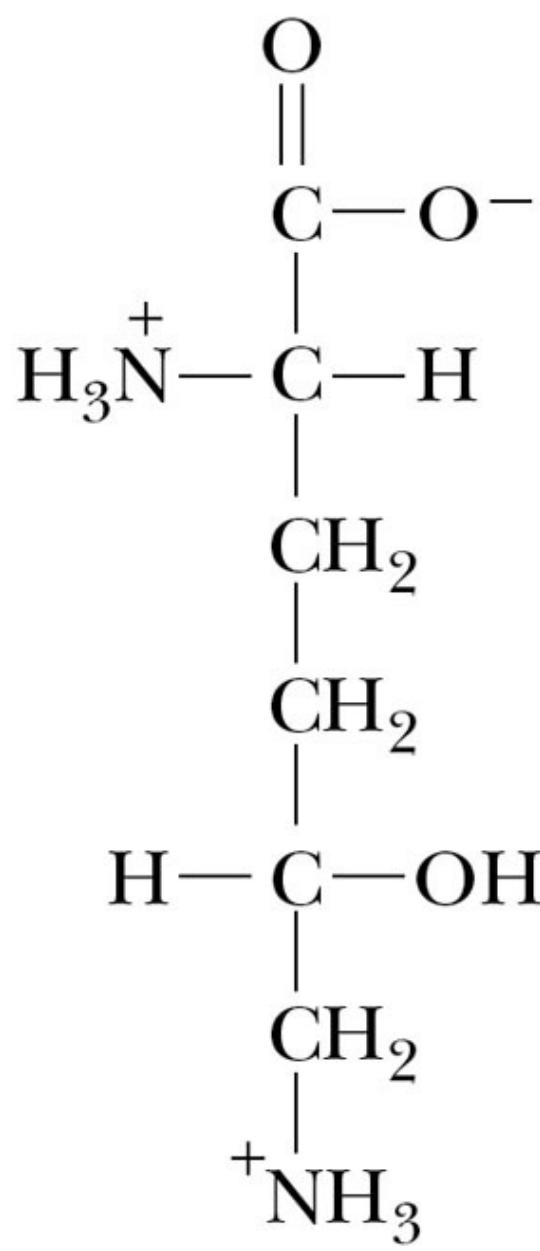
# Collagen Triple Helix

- Collagen is a component of bone and connective tissue
- Consists of three polypeptide chains wrapped around each other in a ropelike twist to form a triple helix called tropocollagen fiber
- Each polypeptide chain is a helix but not an  $\alpha$ -helix
- 30% of amino acids in each chain are Proline and (hydroxyproline); hydroxylysine also occurs
- Every third position is Gly and repeating sequences are X-Pro-Gly or X-Hyp-Gly
- Glycine is small enough to fit inside the helix

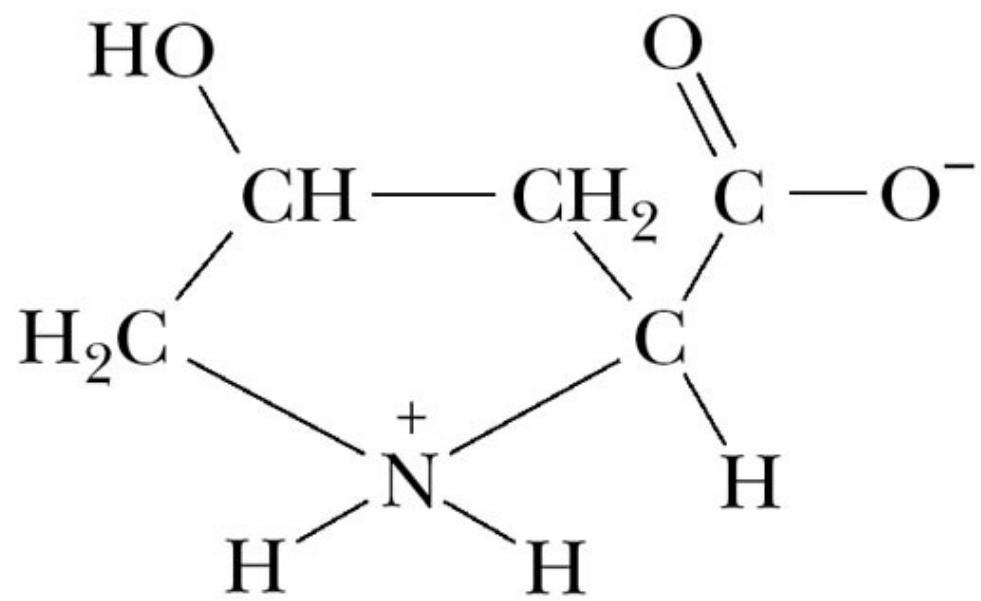


# Collagen Triple Helix

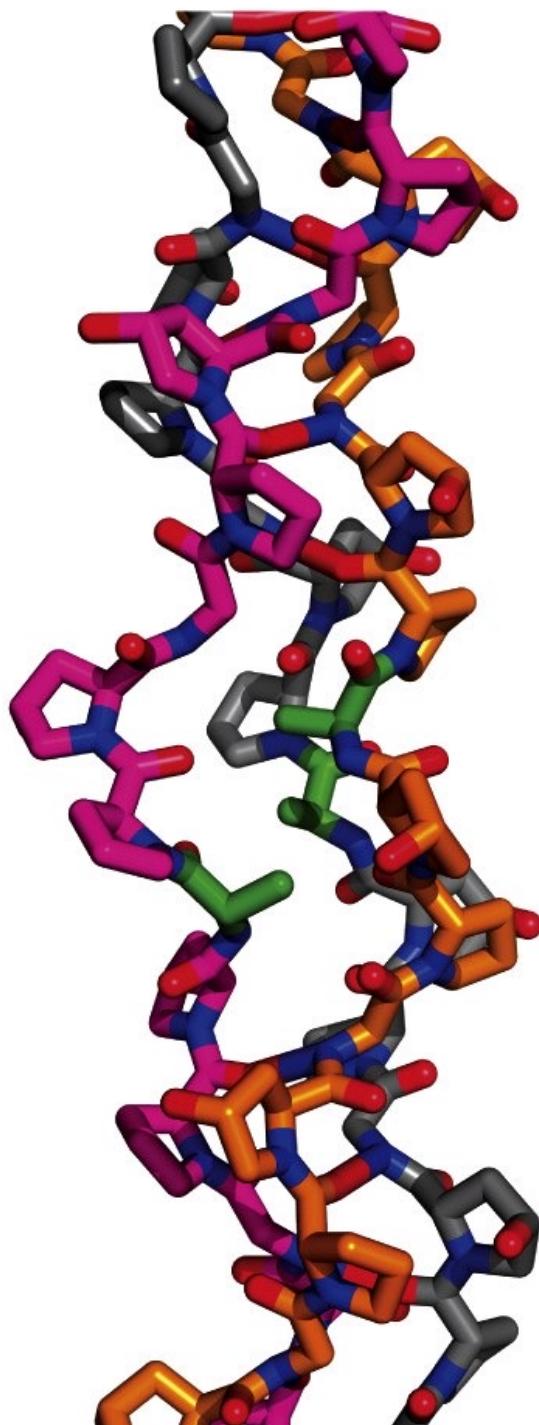
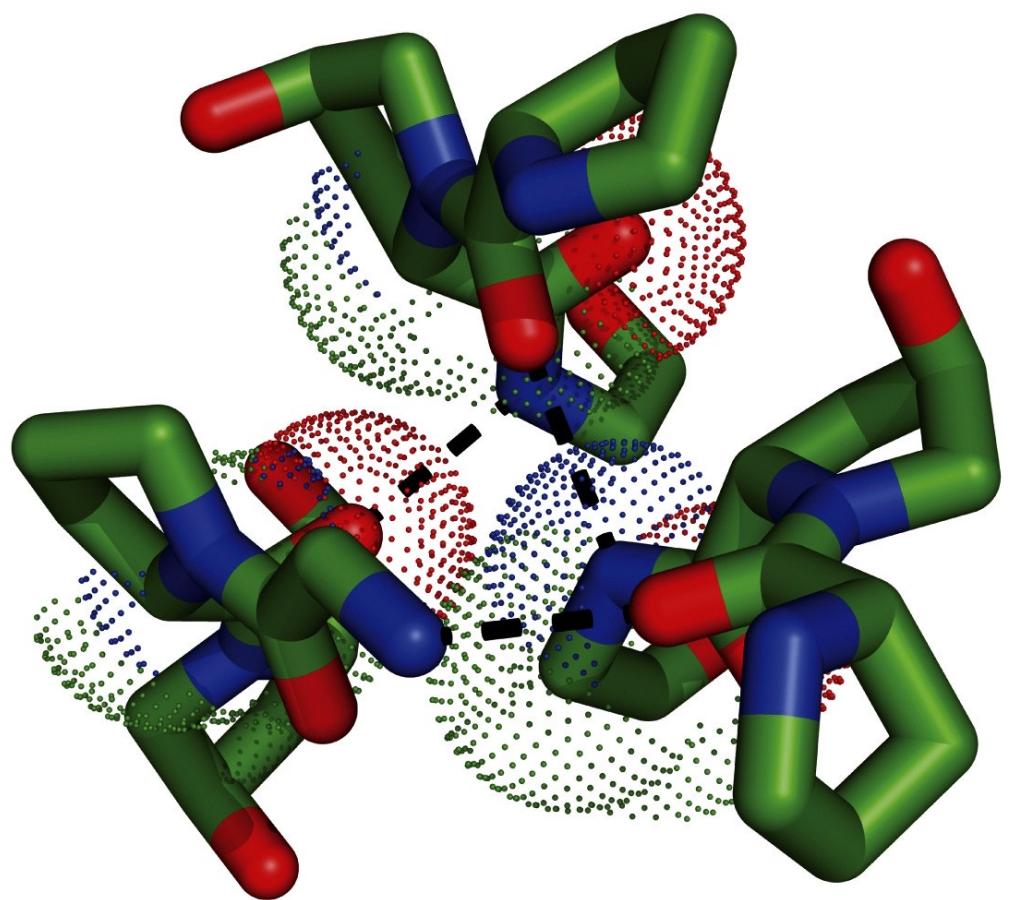
- The three strands are held together by hydrogen bonding involving hydroxyproline or hydroxylysine
- Proline get hydroxylated by specific enzymes (requires Vit C)
- Non hydroxylated proline is less stable (diseases - scurvy)
- Scurvy-skin lesions, broken blood vessels, wounds don't heal, teeth fall out, one cannot stand.
- Collagen helices are cross linked by covalent bonds formed between Lys and His residues
- This cross linking become increasing with age and explain why meat from older animals is tougher than younger animals

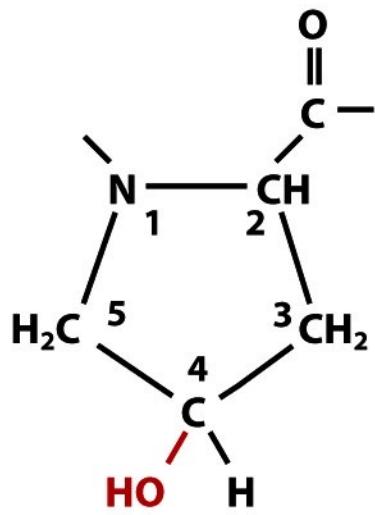


**Hydroxylysine**

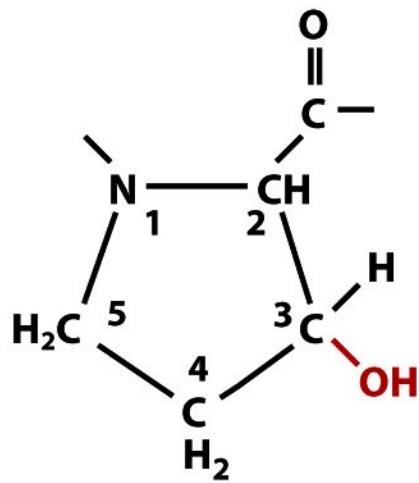


**Hydroxyproline**

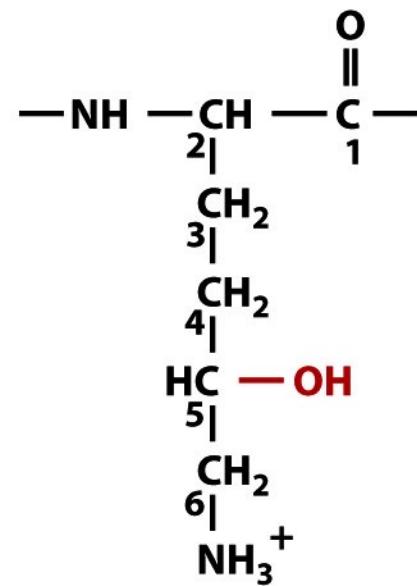




**4-Hydroxyprolyl residue  
(Hyp)**



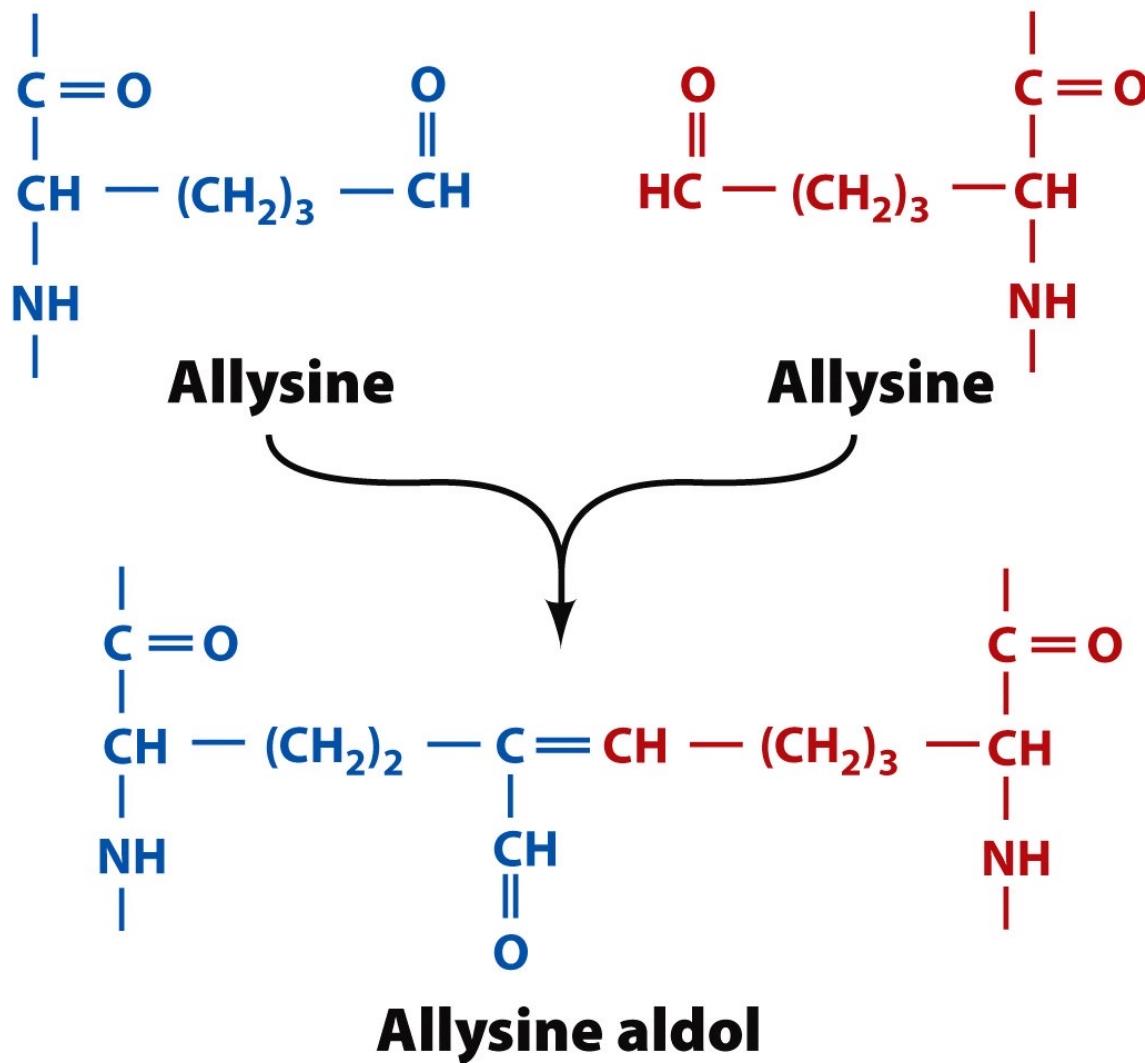
**3-Hydroxyprolyl  
residue**



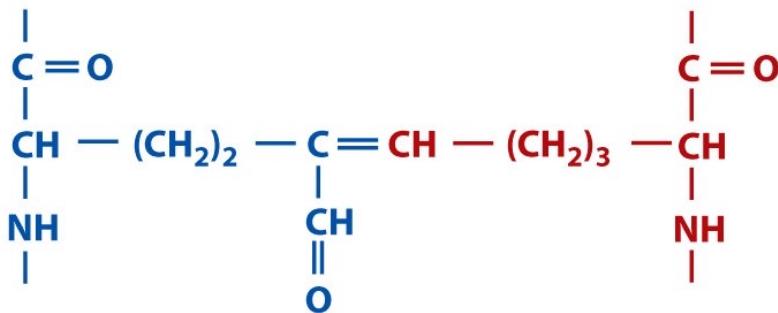
**5-Hydroxylysyl  
residue (Hyl)**

© 2008 John Wiley & Sons, Inc. All rights reserved.

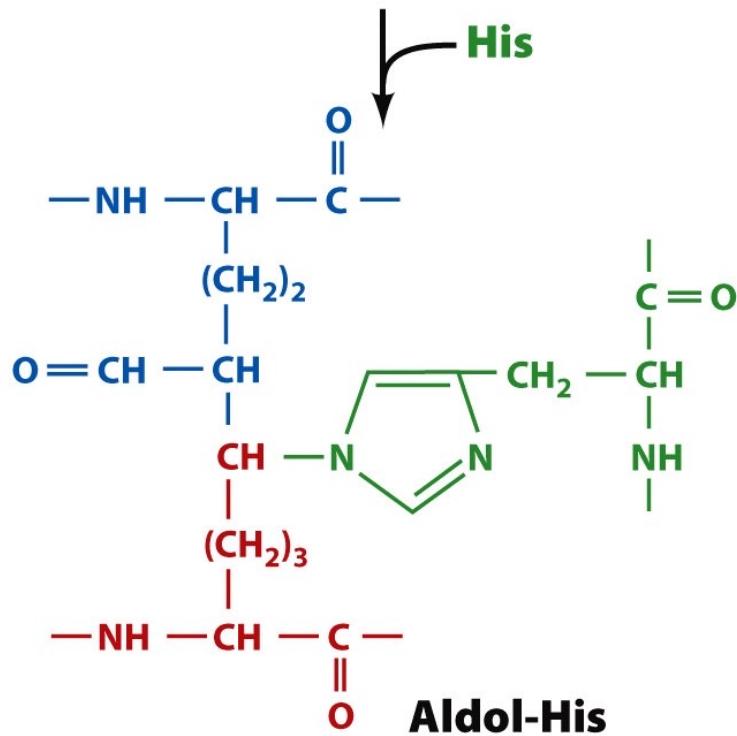
## Collagen is also covalently cross-linked



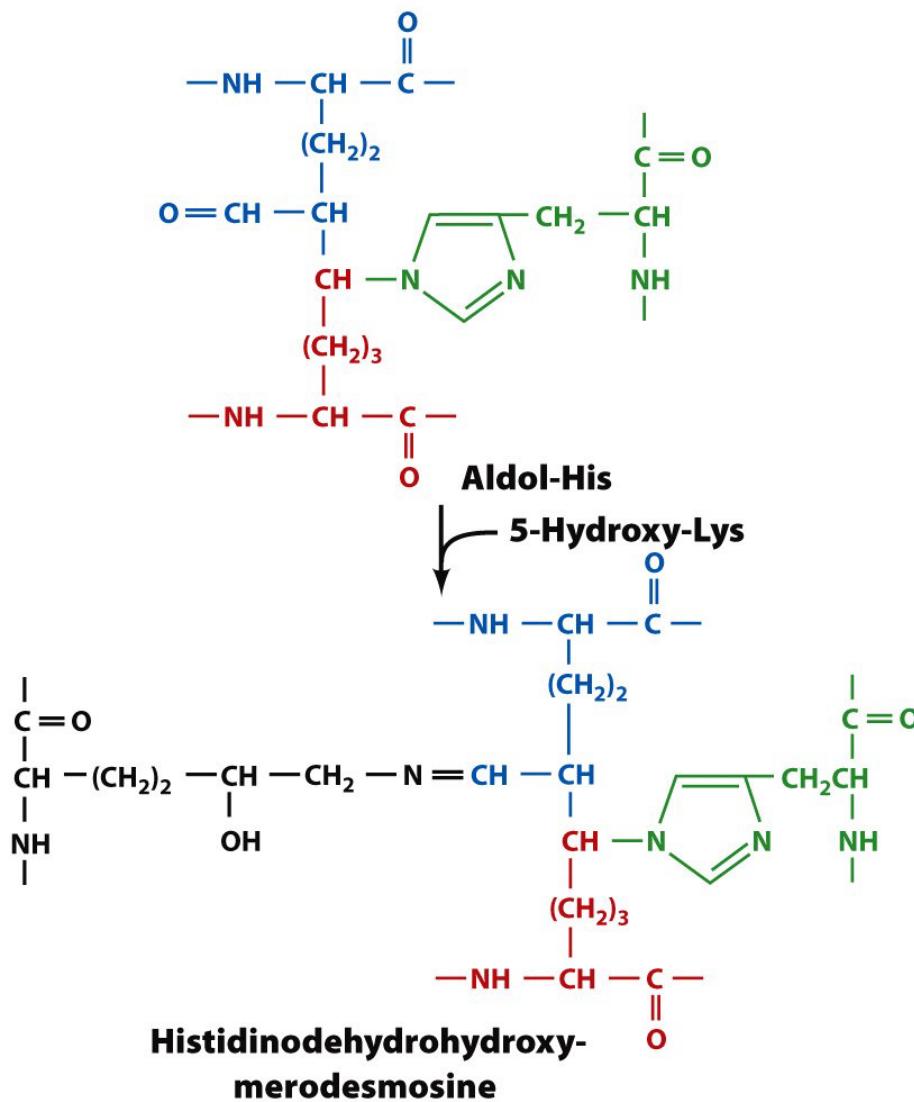
# Collagen is also covalently cross-linked

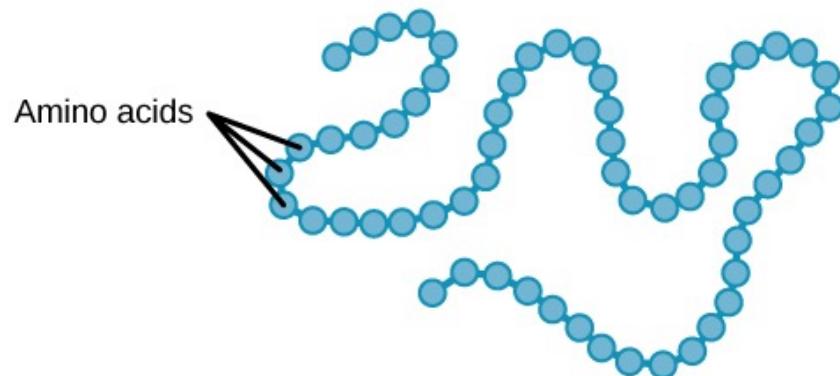


Allysine aldol

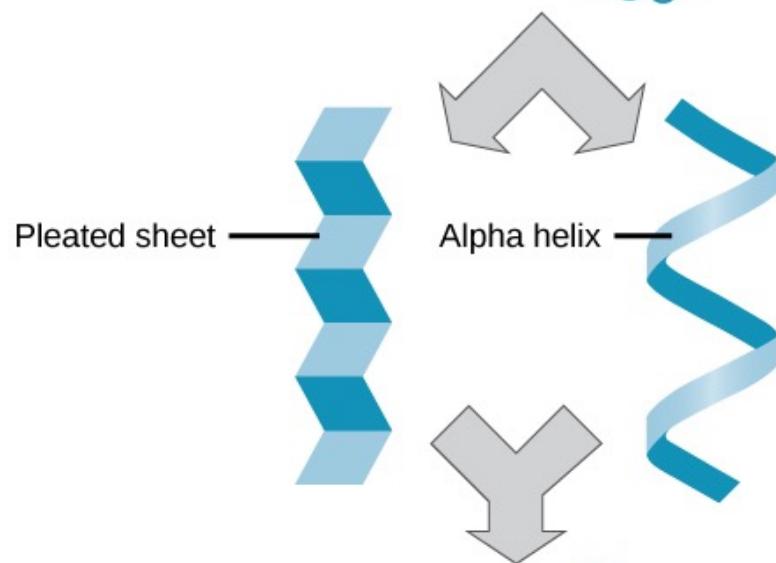


# Collagen is also covalently cross-linked

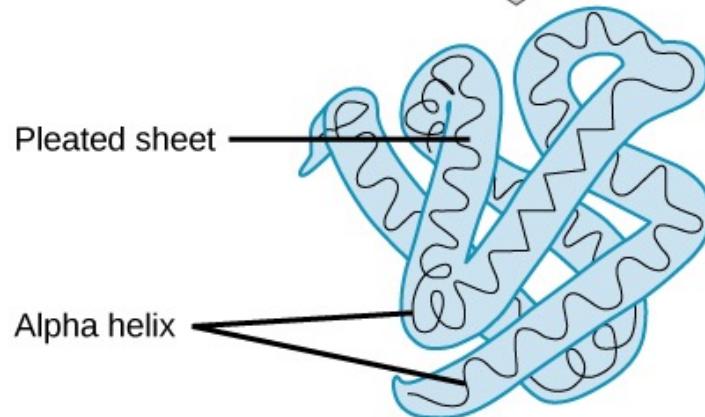




**Primary Protein structure**  
sequence of a chain of amino acids



**Secondary Protein structure**  
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern



**Tertiary protein structure**  
three-dimensional folding pattern of a protein due to side chain interactions

# Tertiary protein structure

- Describes the folding of its secondary structural elements
- Determined mainly via nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and for some protein by cryogenic electron microscopy (cryoEM)

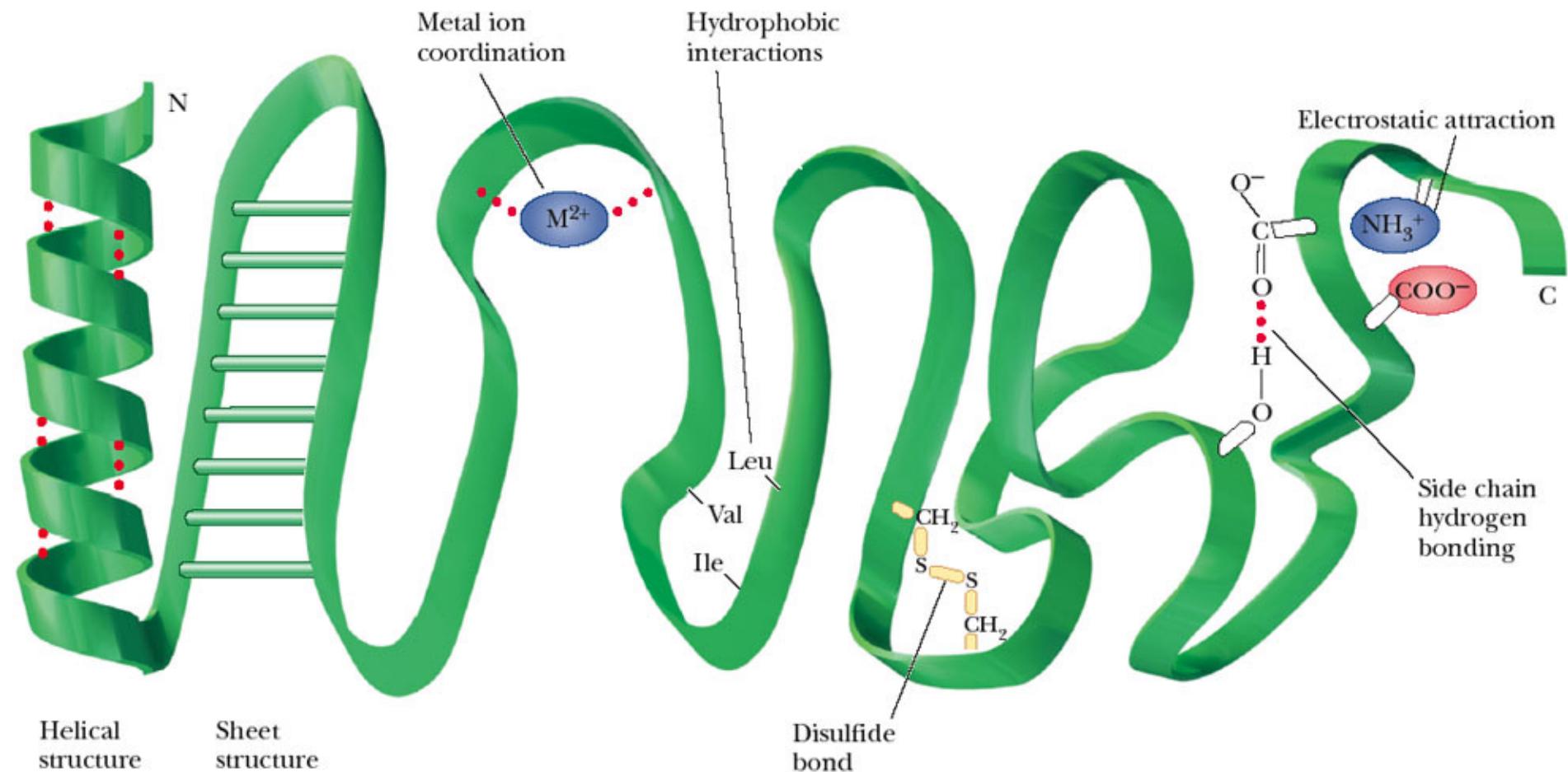
# Tertiary structure

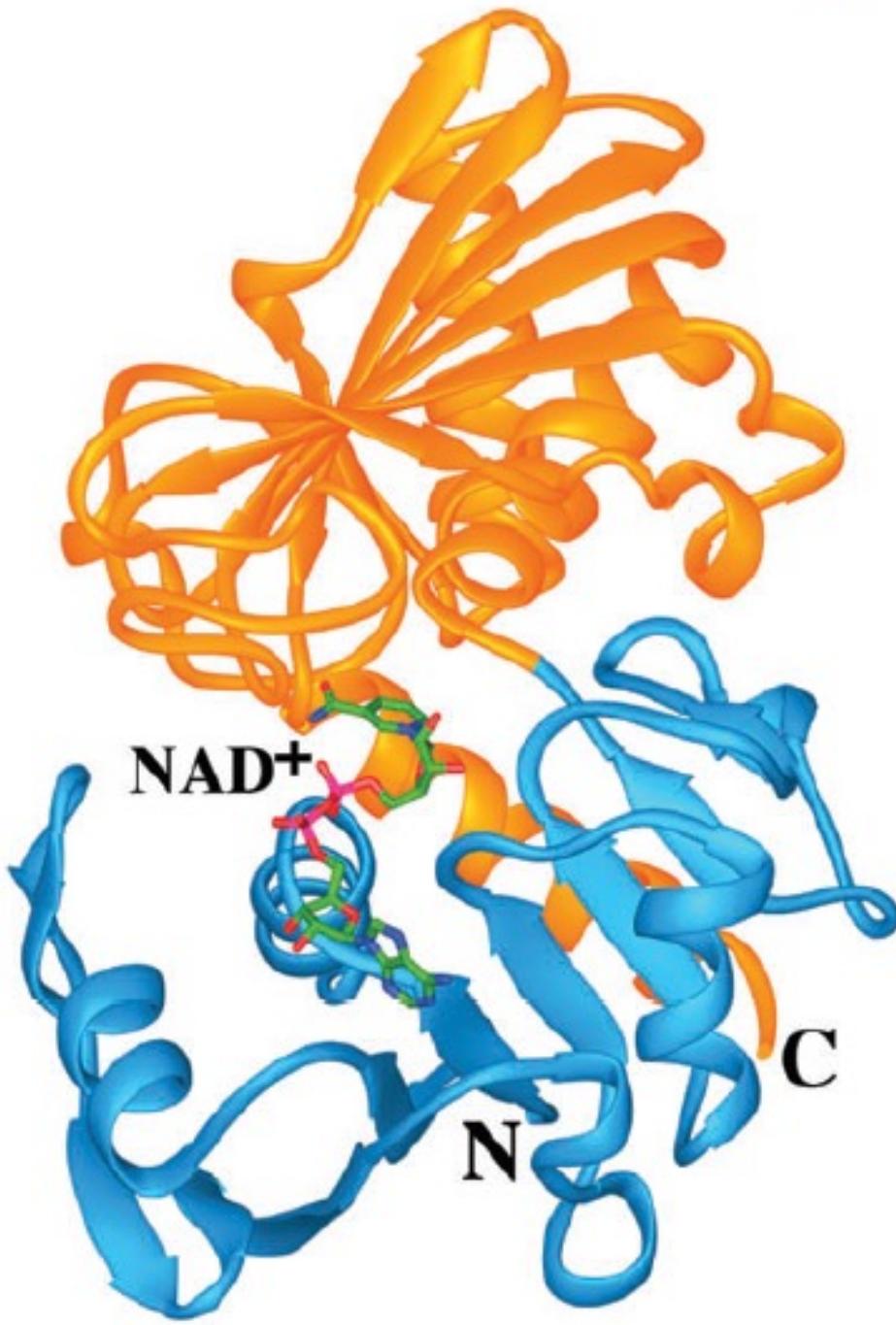
- The tertiary structure of small polypeptides (approximately 200 amino acid residues or less) usually forms a single discrete structural unit.
- Larger polypeptides contain two or more structural subunits within the polypeptide termed domains.
- Domains are tightly folded sub regions of a single polypeptide, connected to each other by more flexible or extended regions.
- Domains are structurally distinct, and often serve as independent units of function

# Forces stabilizing tertiary structures

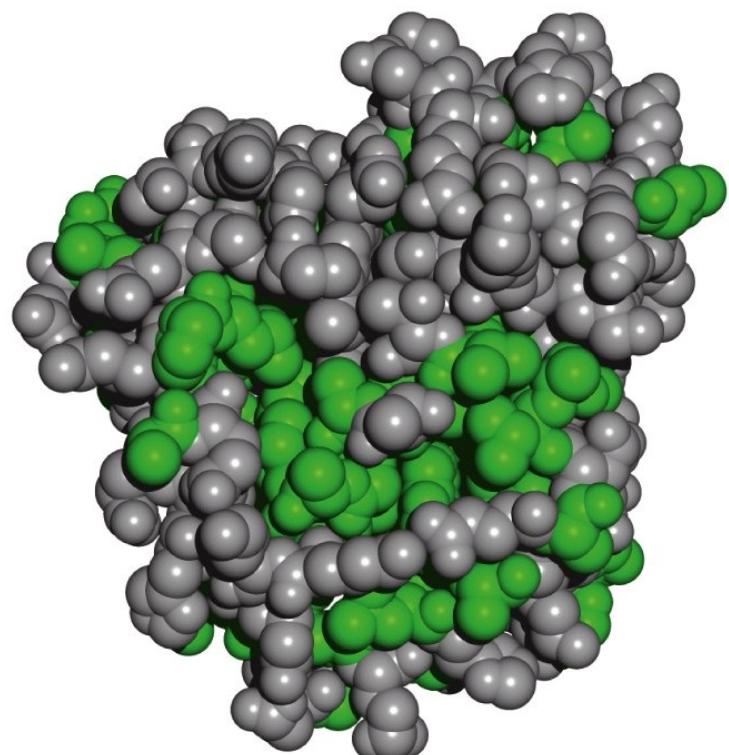
1. Hydrogen bonds (backbone and side chains)
2. **Hydrophobic interaction** ⇒ nonpolar residues tend to cluster together in the interior of protein molecules.
3. **Electrostatic attraction** between oppositely charged groups ⇒ occur in the surface of the molecule.
4. **Metal ion coordination**: several side chains can be complexed to a single metal ion.
5. Covalent bond such as disulfide bonds.

# Tertiary structure

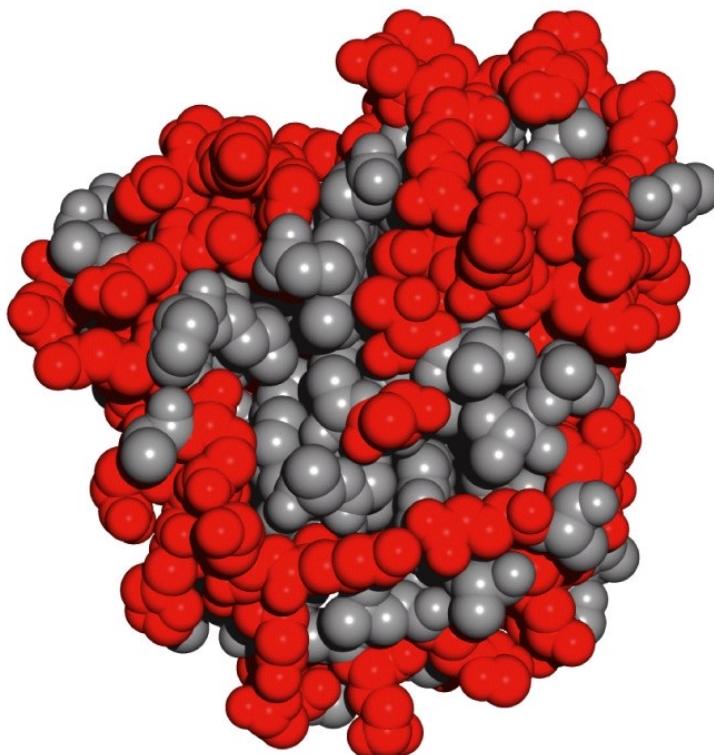




# A folded polypeptide assumes a shape with a **hydrophilic surface** and a **hydrophobic core**



(a)

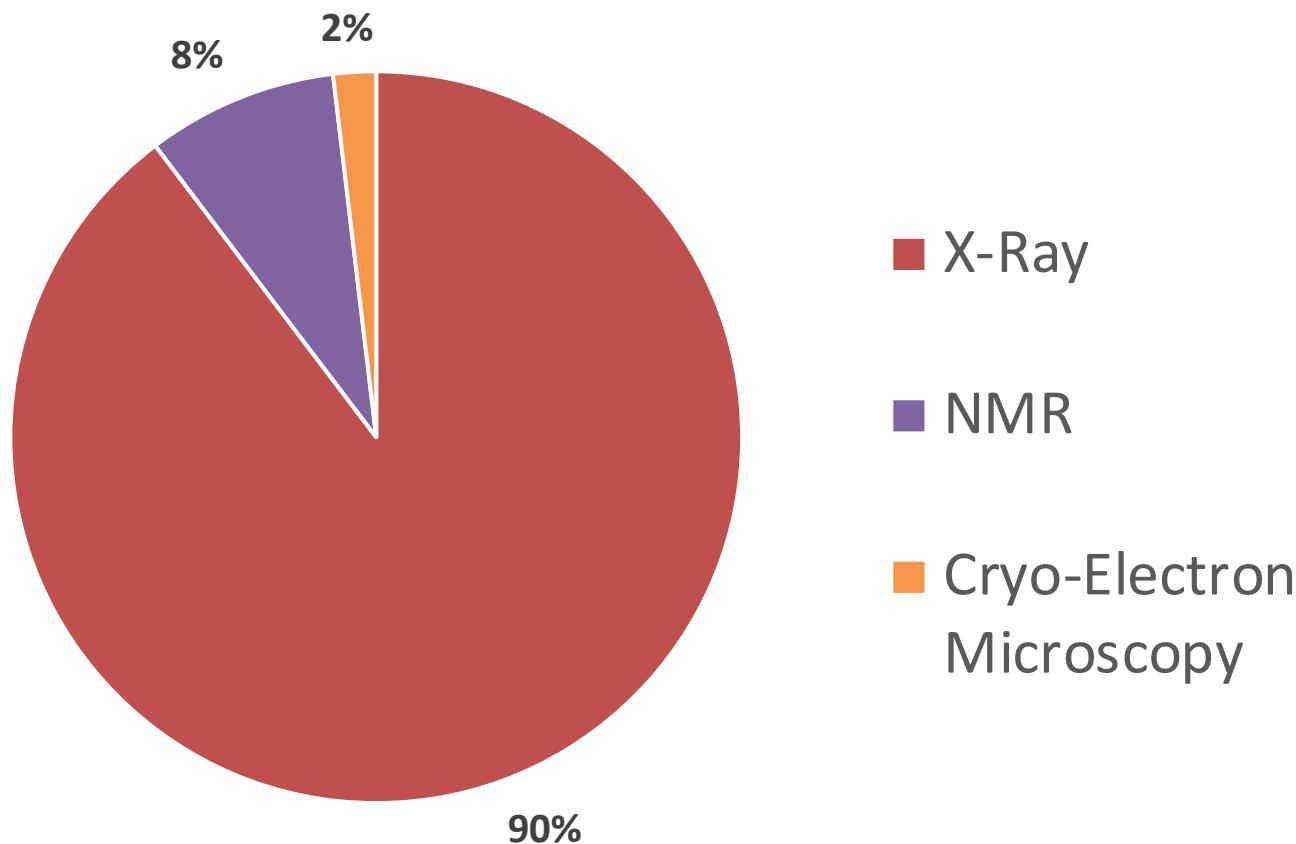


(b)

# Higher structure determination

- Structural determination techniques:
  - X-ray crystallography
  - Nuclear Magnetic Resonance
  - Cryogenic Electron Microscopy
- Structural prediction:
  - Bioinformatics

# Higher structure determination



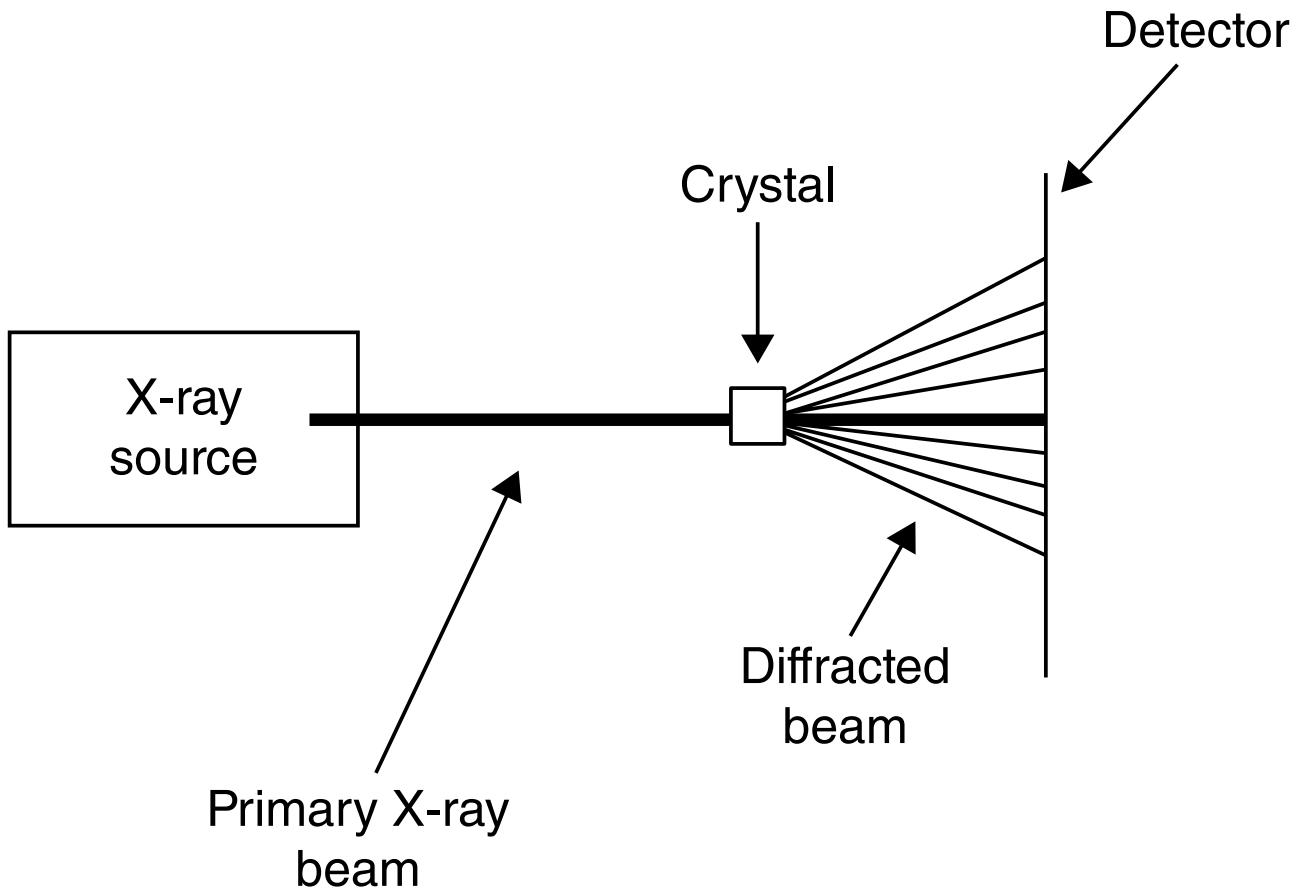
# Higher structure determination

	Advantages	Disadvantages	Objects	Resolution
<b>X-ray crystallography</b>	<ul style="list-style-type: none"><li>Well developed</li><li>High resolution</li><li>Broad molecular weight range</li><li>Easy for model building</li></ul>	<ul style="list-style-type: none"><li>Difficult for crystallization</li><li>Difficult for diffraction</li><li>Solid structure preferred</li><li>Static crystalline state structure</li></ul>	<ul style="list-style-type: none"><li>Crystallizable samples</li><li>Soluble proteins, membrane proteins, ribosomes, DNA/RNA and protein complexes</li></ul>	High
<b>NMR</b>	<ul style="list-style-type: none"><li>High resolution</li><li>3D structure in solution</li><li>Good for dynamic study</li></ul>	<ul style="list-style-type: none"><li>Need for high sample purity</li><li>Difficult for sample preparation</li><li>Difficult for computational simulation</li></ul>	<ul style="list-style-type: none"><li>MWs below 40–50 kDa</li><li>Water soluble samples</li></ul>	High
<b>Cryo-EM</b>	<ul style="list-style-type: none"><li>Easy sample preparation</li><li>Structure in native state</li><li>Small sample size</li></ul>	<ul style="list-style-type: none"><li>Relatively low resolution</li><li>Applicable to samples of high molecular weights only</li><li>Highly dependent on EM techniques</li><li>Costly EM equipment</li></ul>	<ul style="list-style-type: none"><li>&gt;150 kDa</li><li>Virions, membrane proteins, large proteins, ribosomes, complex compounds</li></ul>	Relatively Low ( $<3.5 \text{ \AA}$ )

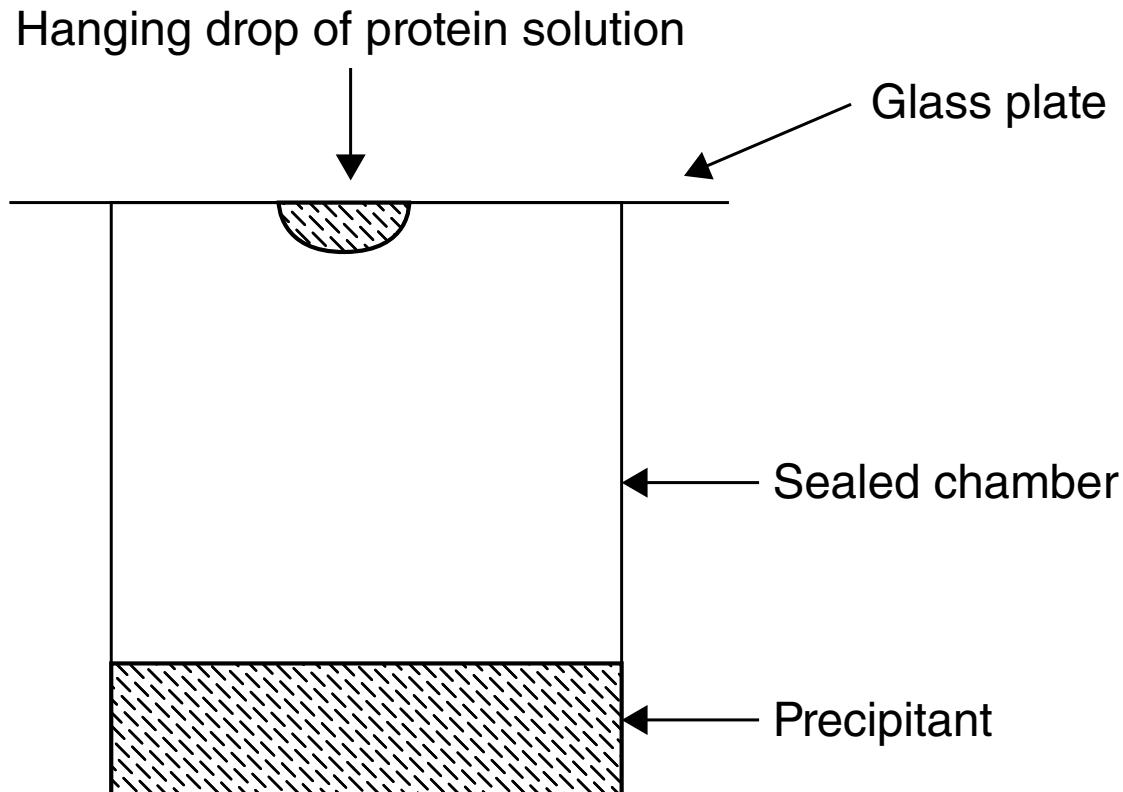
# X-ray crystallography

- The wavelength- in ångströms ( $10^{-10}$  m, 0.1nm) - of X-rays approximates the distances between atoms in proteins.
- Proteins in crystalline form are bombarded with a beam of X-rays
- Most of these X-rays pass straight through the crystal but some are diffracted by the electrons of the atoms in the crystal.
- The resultant diffraction pattern, recorded on a detector, is a reflection of the three-dimensional structure of the protein molecules present in the crystal.
- Individual atoms are normally distinguishable with a resolution of 1–1.5 Å (0.1–0.15 nm)

# X-ray crystallography

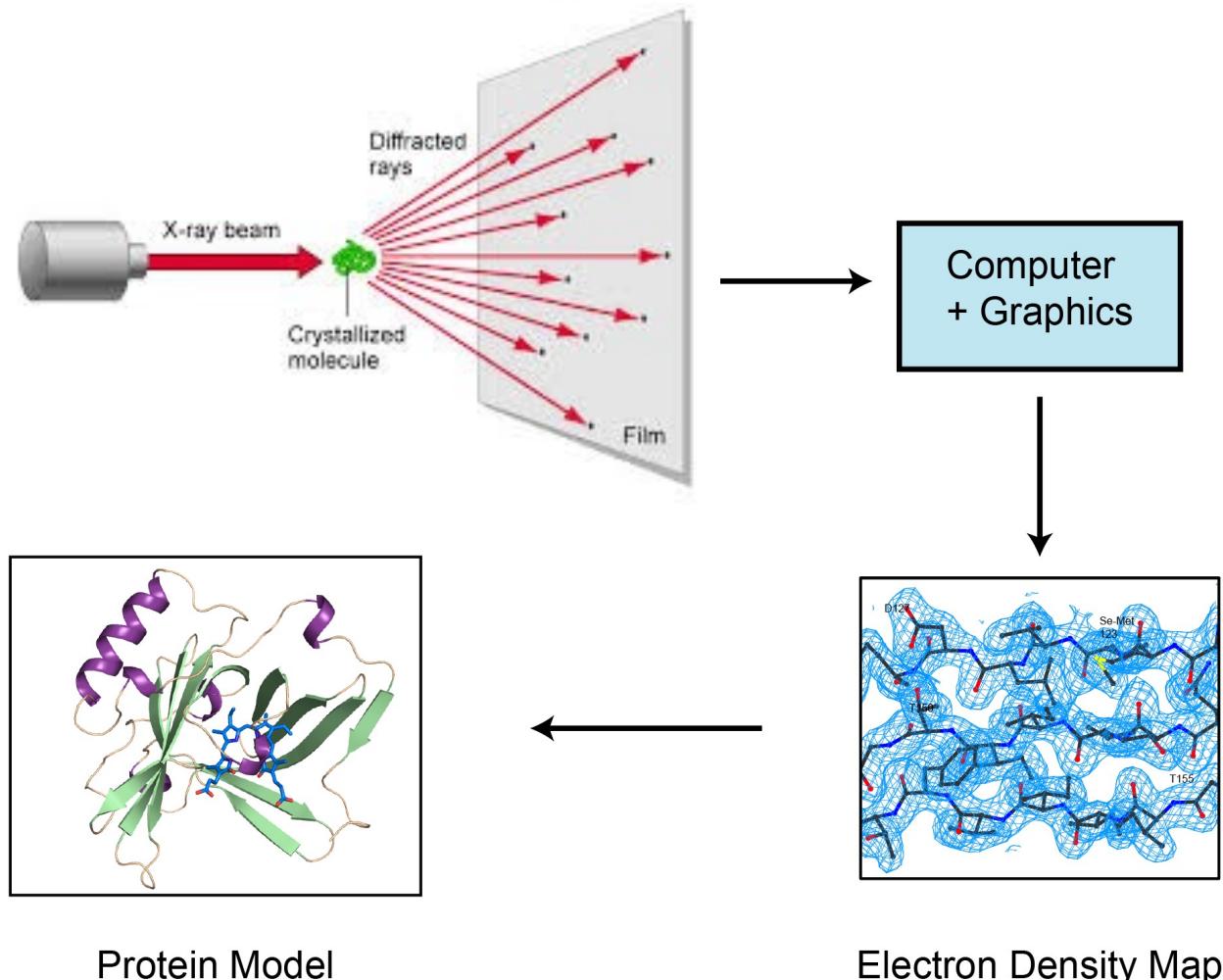


# X-ray crystallography



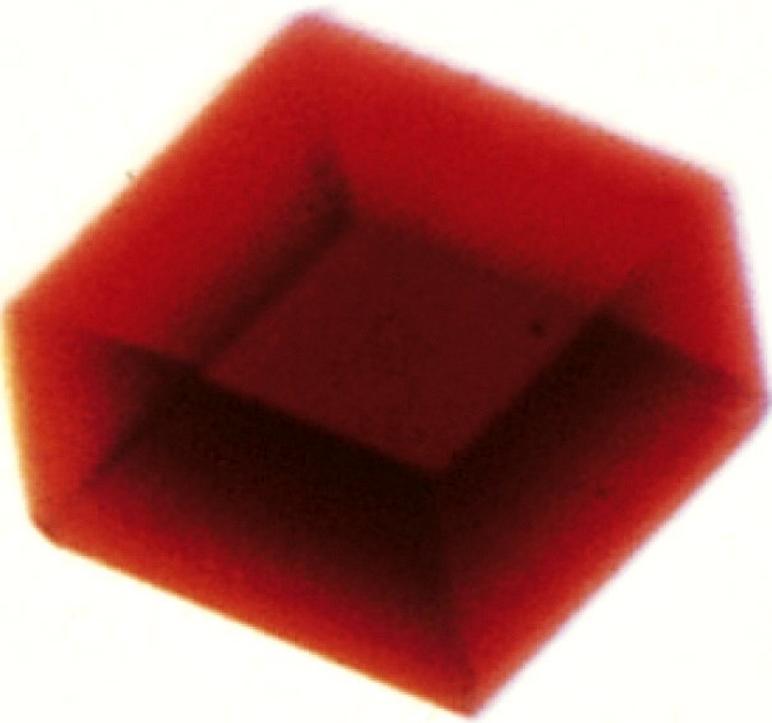
# Determination of tertiary Structure

## X-ray crystallography



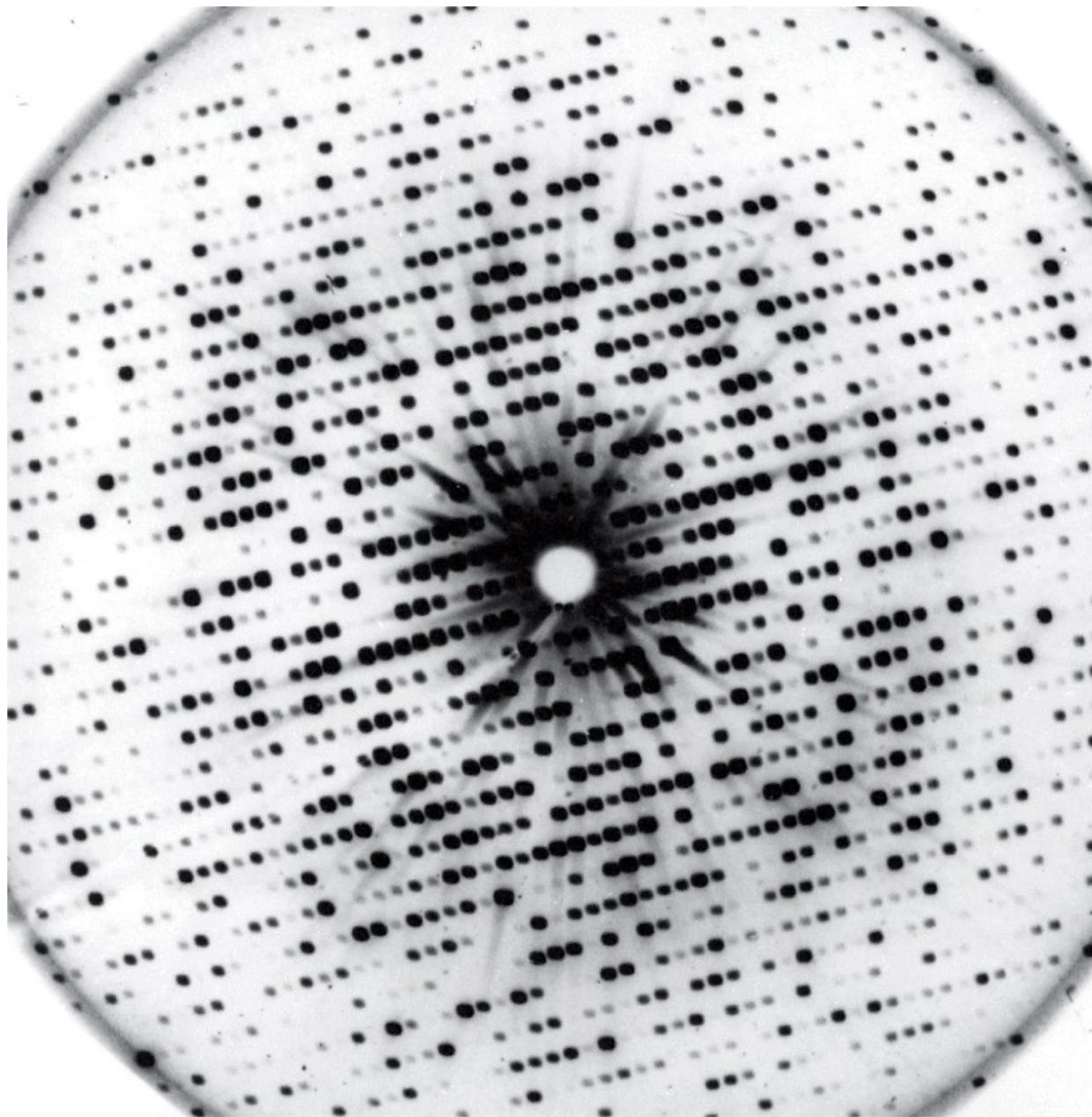
Protein Model

Electron Density Map

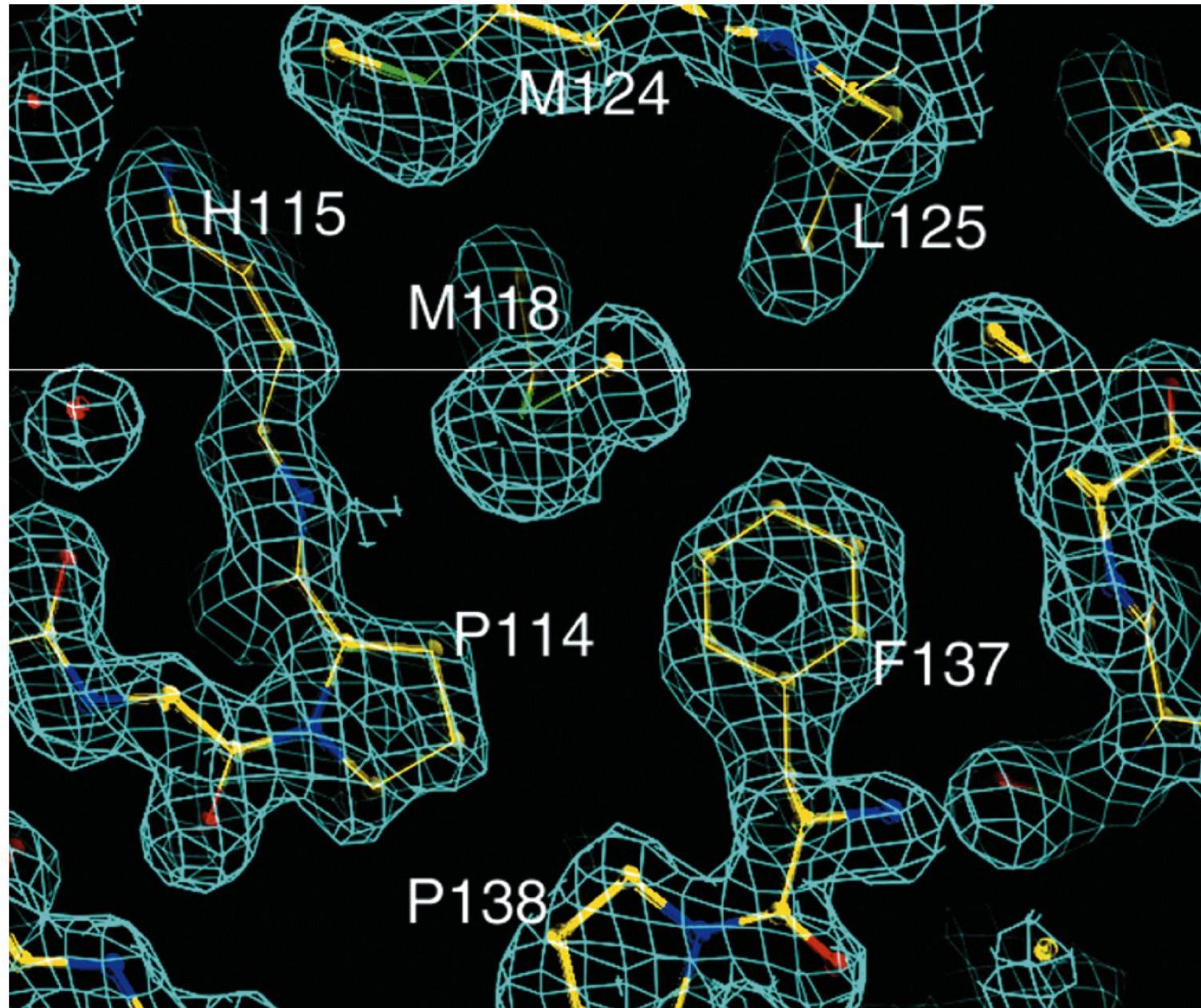


Courtesy of Larry Siecker, University of Washington

Figure 6-20c

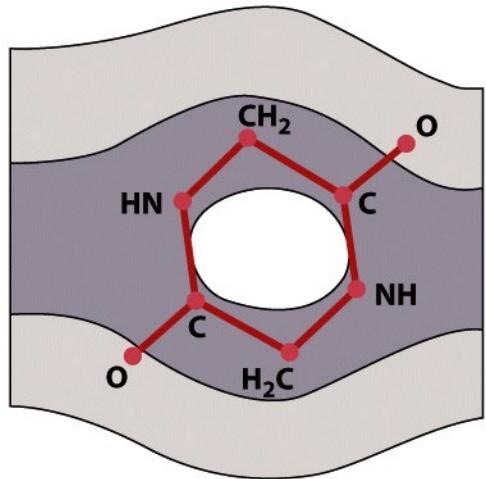


Courtesy of John Kendrew, Cambridge University, U.K.

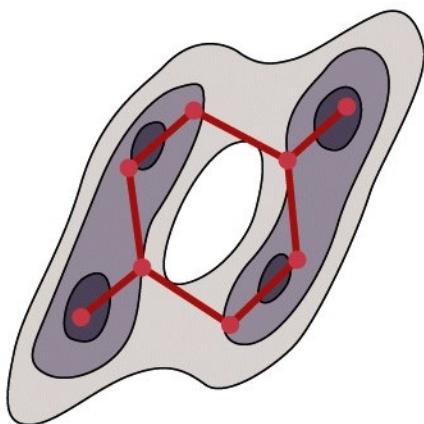


Courtesy of Xinhua Ji, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland

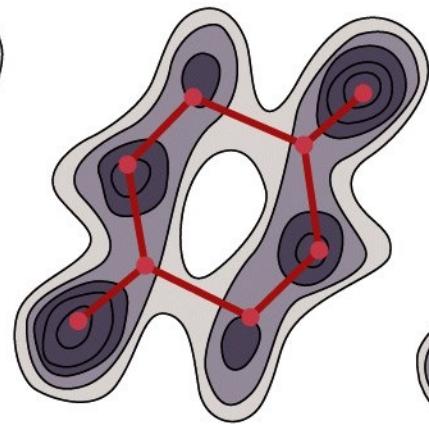
**(a) 6.0-Å resolution**



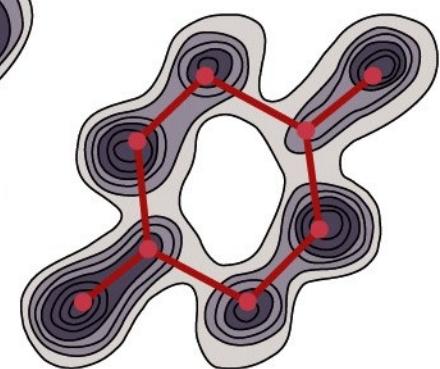
**(b) 2.0-Å resolution**



**(c) 1.5-Å resolution**



**(d) 1.1-Å resolution**



# Nuclear Magnetic Resonance

**A molecule**  
is made up of atoms.

**An atom**

Atoms

● Protons (+)  
● Electrons (-)  
● Neutrons (no charge)

Think of the nucleus as a bar magnet...

In the absence of an external magnetic field:

Randomly oriented nuclei

But when the nuclei are inside a very strong magnetic field (million times that of the Earth), they get oriented along or against the magnetic field.

Electromagnet

Direction of magnetic field

High energy  
S S S  
N N N  
Low energy  
N N N N  
S S S S

When energy is added using radio-frequency pulses, more of the 'bar magnets' move to the higher energy state.

Radio-frequency (r.f) pulse

When the pulse is turned off, the system returns back to the original 'equilibrium' state by releasing energy, which is detected as the NMR signal.

Energy released

NMR signal

High energy = oriented **against the direction** of the magnetic field

Low energy = oriented **in the direction** of the magnetic field

3.1 3.0 2.9 2.8 2.7 2.6 2.5 2.4 2.3 2.2 2.1 2.0 1.9 1.8 1.7 1.6 1.5 1.4 1.3 1.2 1.1 1.0 0.9 0.8

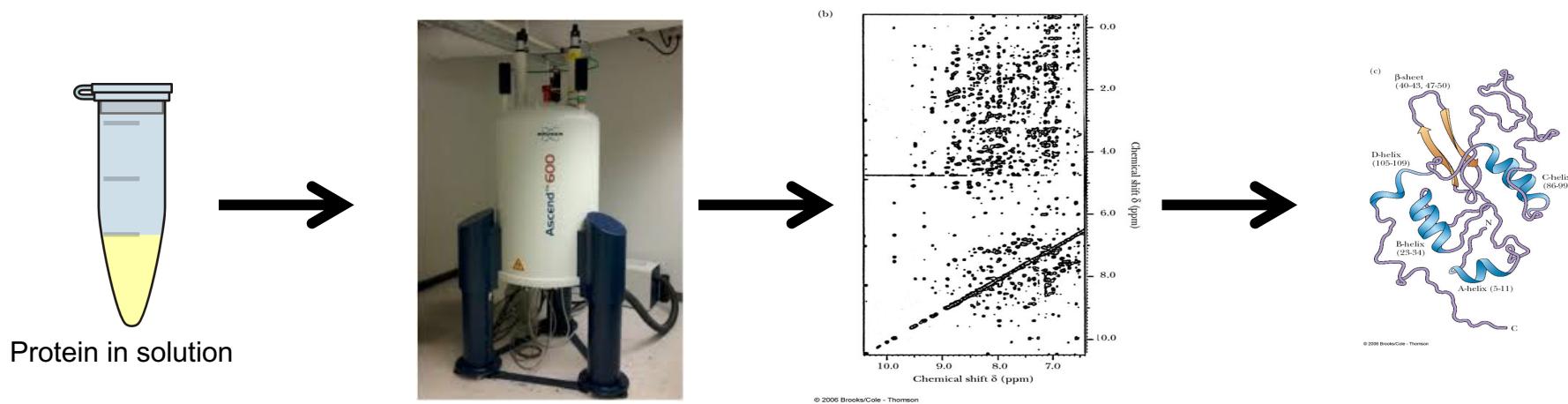
# NMR

- NMR may be used to determine the structure of proteins in free solution.
- The technique generates a range of closely related conformational structures, which largely reflects the fact that protein conformation can flex or ‘breathe’ in solution.
- NMR analysis involves applying a strong magnetic field to a sample of the protein of interest.
- Electromagnetic radiation in the radio- frequency range is then applied.
- NMR analysis is based on the fact that a number of atomic nuclei display a magnetic moment. These include  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{31}\text{P}$ .

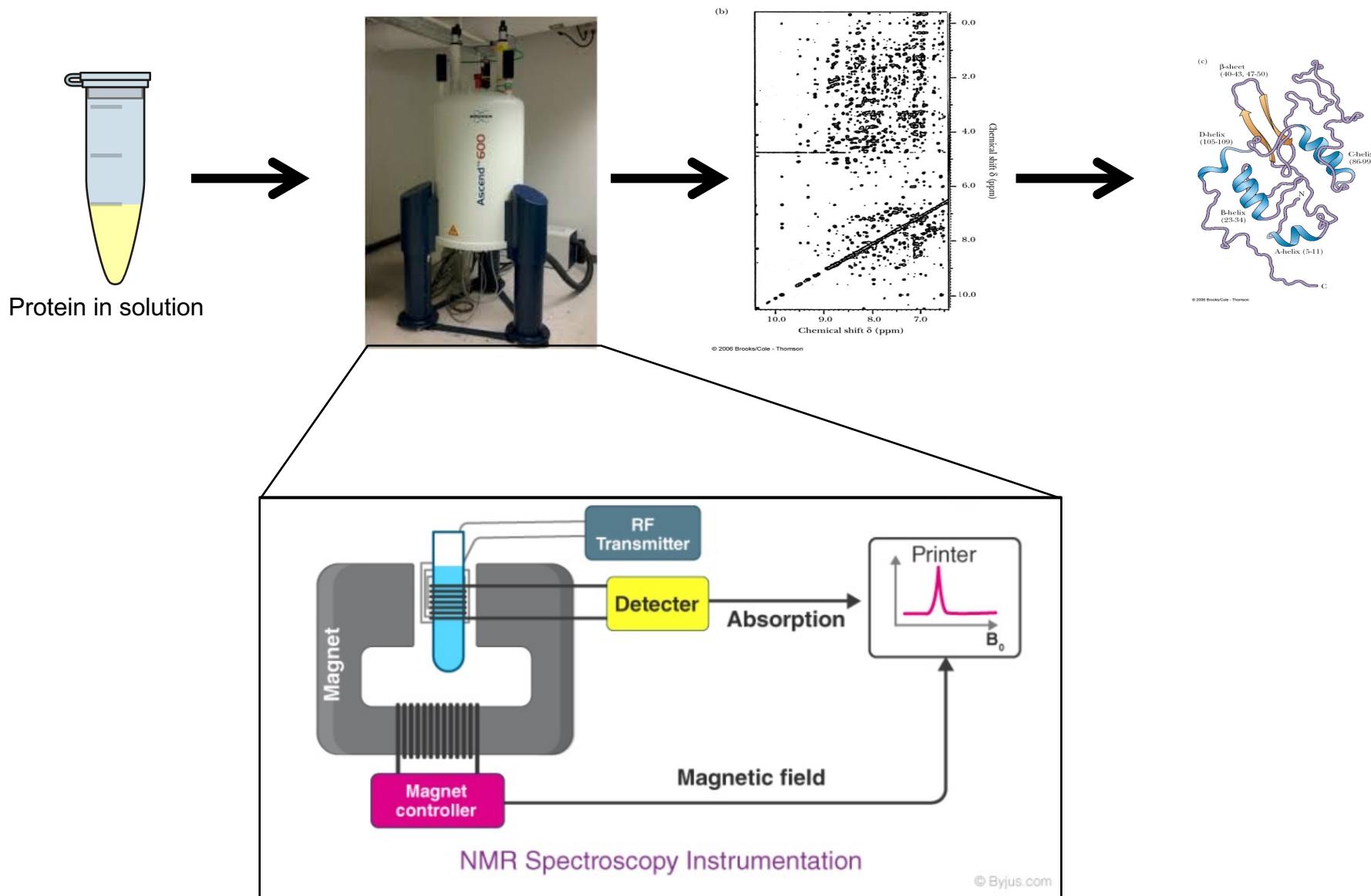
# NMR

- Nuclei spin about an axis and act like tiny magnets and will therefore interact with an applied magnetic field.
- If a protein is placed in a strong magnetic field, the spin on such nuclei aligns along this field. This alignment can be converted to an excited state if radiofrequency energy of the appropriate frequency is applied.
- When nuclei revert to their unexcited state, they emit radiofrequency radiation, which may be detected and measured.
- The exact frequency emitted by any given nucleus is influenced by its molecular environment, and these shifts in frequency emitted can be used to provide three-dimensional structural information about the protein.

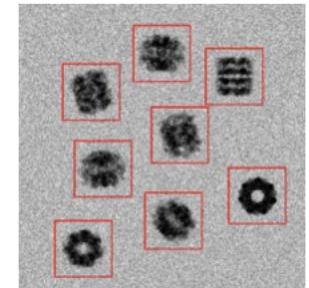
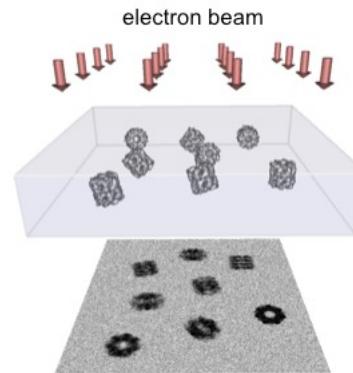
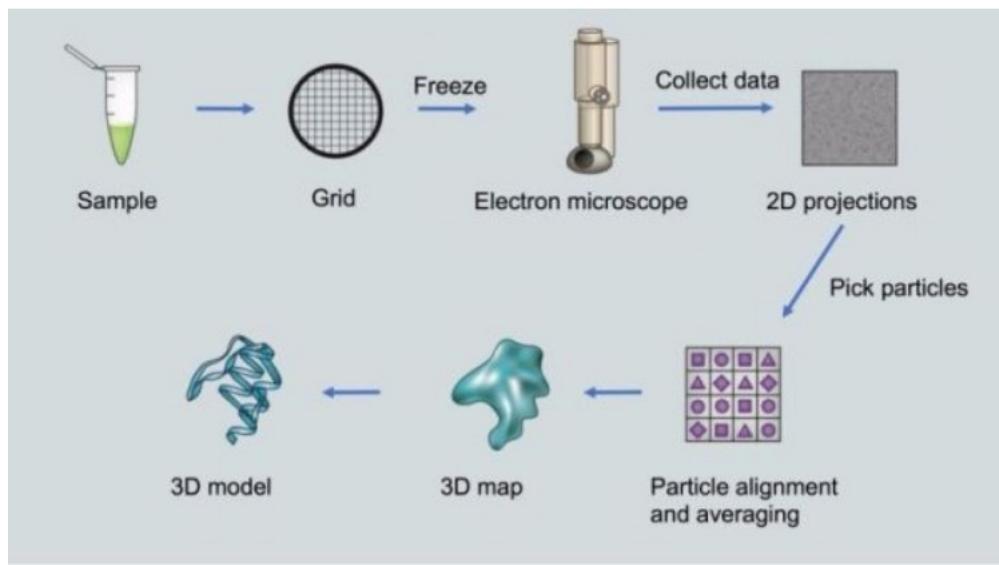
# Determination of tertiary Structure by NMR



# Determination of tertiary Structure by NMR



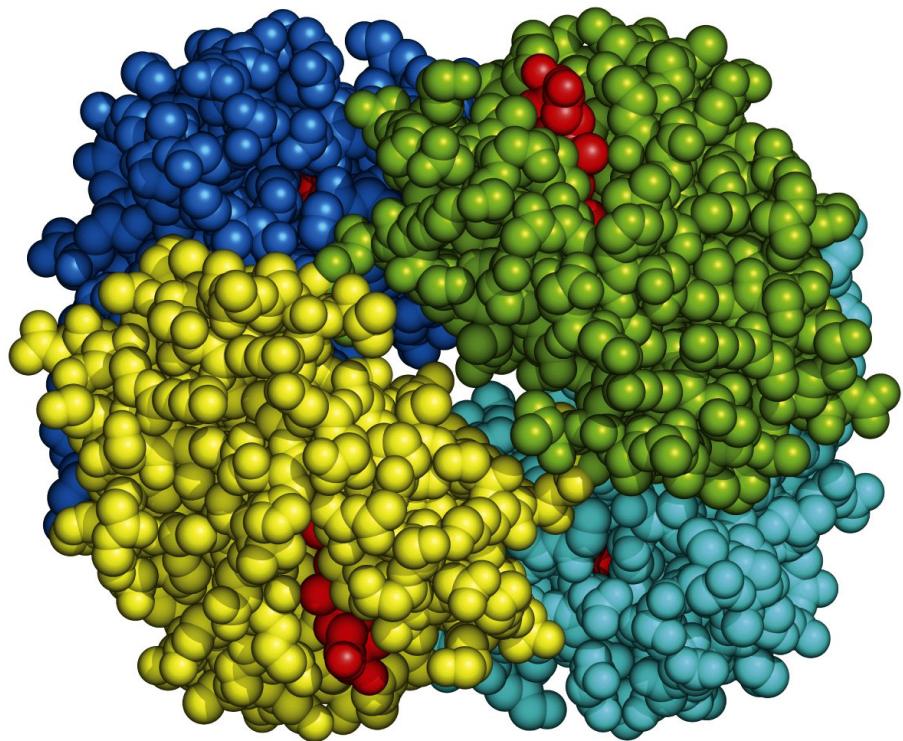
# Cryogenic Electron Microscopy



Creative Biostructure

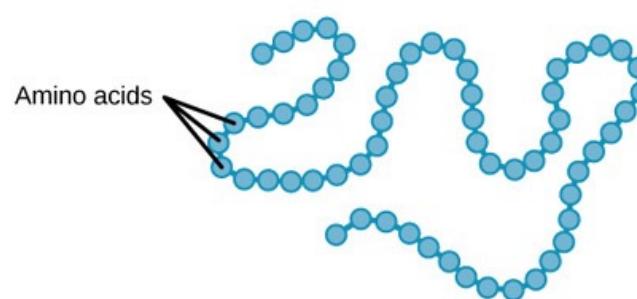
# Quaternary Structure of Proteins

- Protein that consists of more than one polypeptide chain (**called subunits**) (e.g. dimers, trimers, tetramers).
- The chains (subunits) interact with one another noncovalently via **electrostatics, hydrogen bonds, hydrophobic**



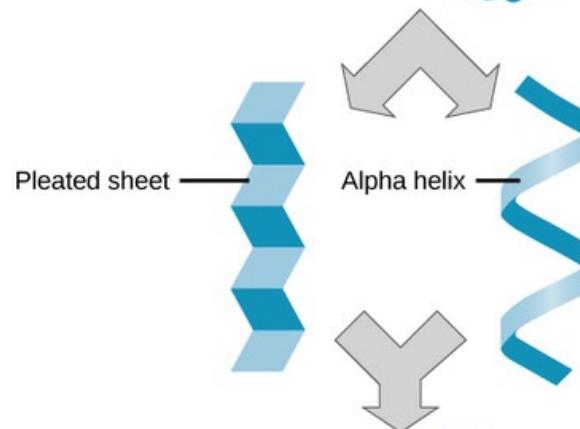
© 2008 John Wiley & Sons, Inc. All rights reserved.

## 1. Peptide bonds

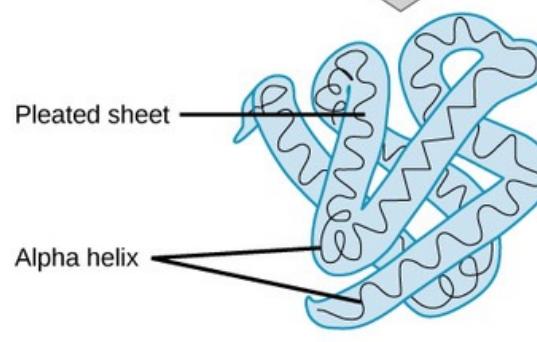


**Primary Protein structure**  
sequence of a chain of amino acids

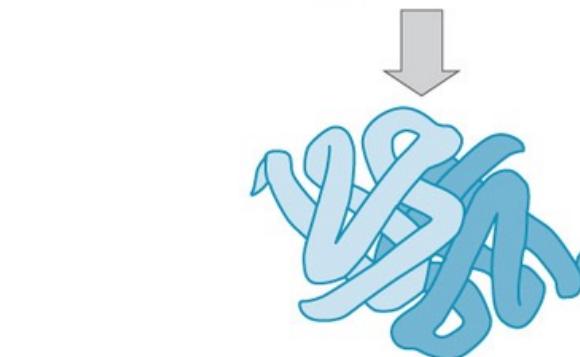
## 1. Hydrogen bonds



**Secondary Protein structure**  
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern



**Tertiary protein structure**  
three-dimensional folding pattern of a protein due to side chain interactions



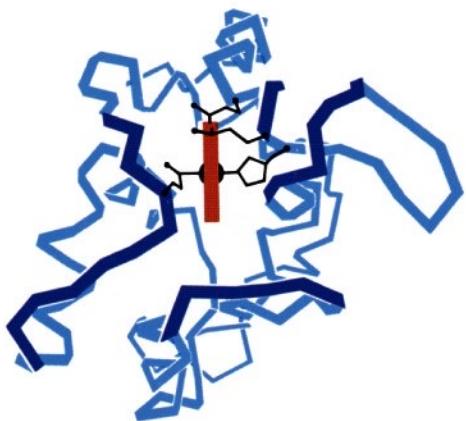
**Quaternary protein structure**  
protein consisting of more than one amino acid chain

- 1. Hydrogen bonds
- 2. Hydrophobic interaction
- 3. Electrostatic attraction
- 4. Metal ion coordination
- 5. Disulfide bonds.

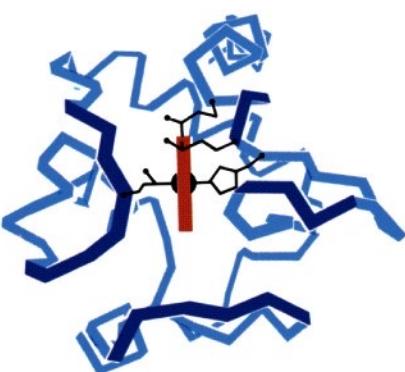
- 1. Hydrogen bonds
- 2. Hydrophobic interaction
- 3. Electrostatic attraction

# Structure is more conserved than sequence

(a) *Paracoccus c550*  
134 amino acid residues



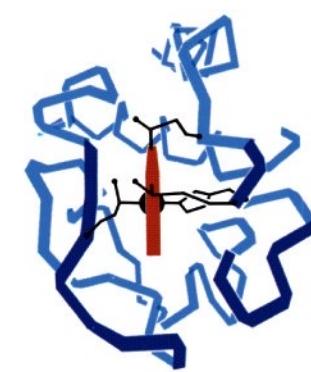
(b) *Rhodospirillum c<sub>2</sub>*  
112 amino acid residues



(a) Tuna c  
103 amino acid residues



(a) *Chlorobium c555*  
86 amino acid residues



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

# Fibrous Proteins

- **Fibrous proteins:** contain polypeptide chains organized approximately parallel along a single axis. They
  - consist of long fibers or large sheets
  - tend to be mechanically strong
  - are insoluble in water
  - play important structural roles in nature
- Examples are
  - **Keratin** of hair and wool
  - **Elastin** of connective tissue including cartilage, bones, teeth, skin, and blood vessels



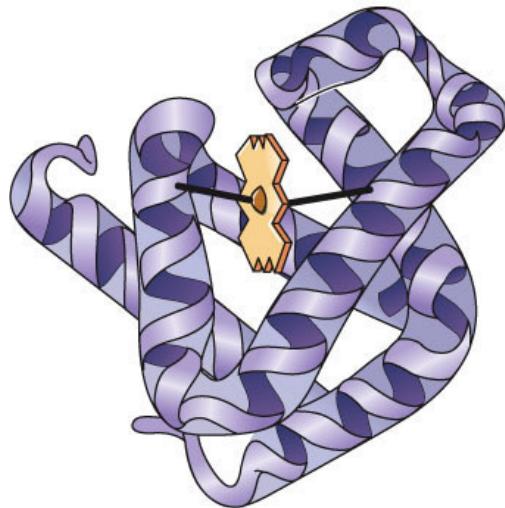
Filament  
(four right-hand  
twisted protofilaments)

# Domains

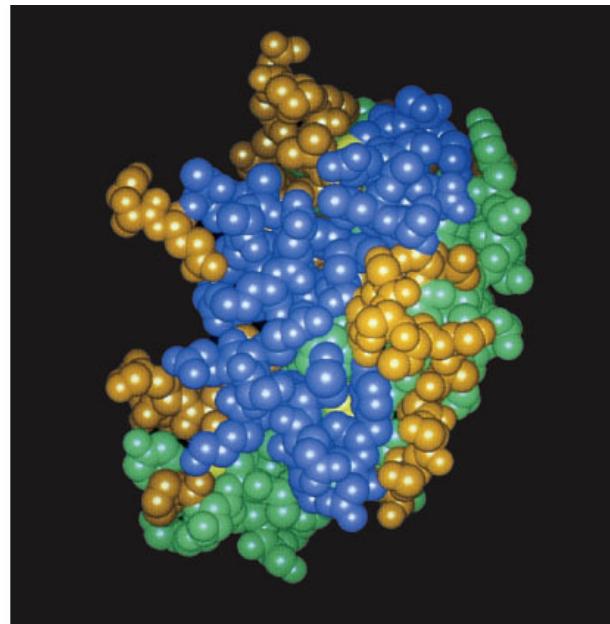
- Stable units of Protein.
- Conserved part of a given protein sequence
- Decides structure and function of Protein
- Folded 3-D compact structure with stability
- Length vary between 25 to 500 amino acid.

# Globular Proteins

- **Globular proteins:** proteins which are folded to a more or less spherical shape (compact shape)
  - They tend to be soluble in water
  - Most of their polar side chains are on the outside and interact with the aqueous environment by hydrogen bonding and ion-dipole interactions

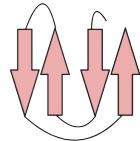


Myoglobin, a globular protein

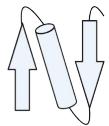


# Motifs VS Domains VS Subunit

## MOTIFS



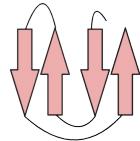
Greek key  
motif



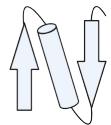
$\beta\alpha\beta$  motif

# Motifs VS Domains VS Subunit

## MOTIFS

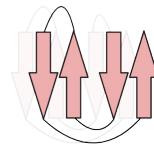


Greek key motif

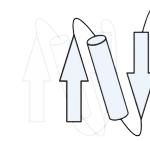


$\beta\alpha\beta$  motif

## DOMAINS



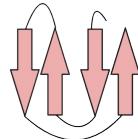
A domain made of  
Greek key motifs



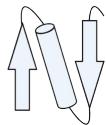
A domain made  
of  $\beta\alpha\beta$  motifs

# Motifs VS Domains VS Subunit

## MOTIFS

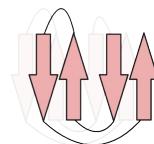


Greek key motif

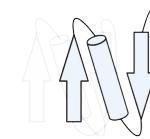


$\beta\alpha\beta$  motif

## DOMAINS

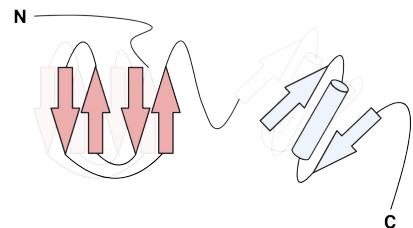


A domain made of Greek key motifs



A domain made of  $\beta\alpha\beta$  motifs

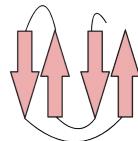
## POLYPEPTIDE (SUBUNIT)



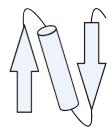
A polypeptide consisting of TWO domains connected with each other on the same polypeptide

# Motifs VS Domains VS Subunit

## MOTIFS

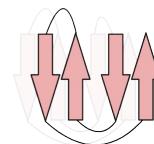


Greek key motif

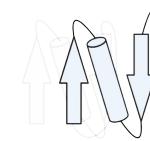


$\beta\alpha\beta$  motif

## DOMAINS

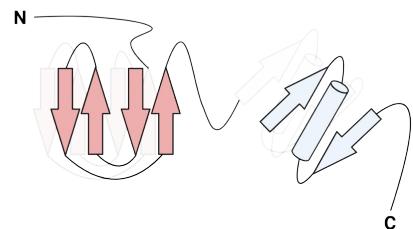


A domain made of Greek key motifs



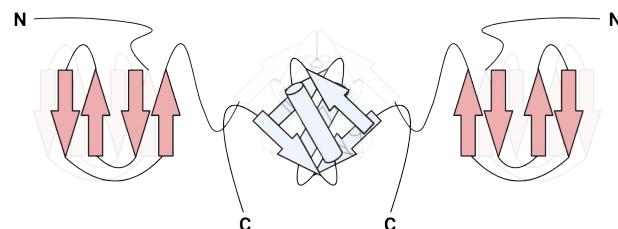
A domain made of  $\beta\alpha\beta$  motifs

## POLYPEPTIDE (SUBUNIT)



A polypeptide consisting of TWO domains connected with each other on the same polypeptide

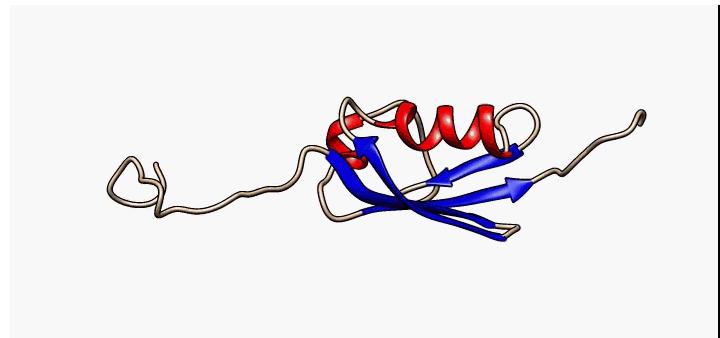
## PROTEIN



A protein consisting of TWO identical subunits

# Intrinsically disordered proteins (IDPs)

- Are proteins that do not adopt a defined three-dimensional structure but are nevertheless important for cellular function.
- Some proteins display complete or near-complete intrinsic structural disorder
- A significant number of proteins display long stretches (>30 amino acid residues) that are intrinsically disordered.



# Intrinsically disordered proteins (IDPs)

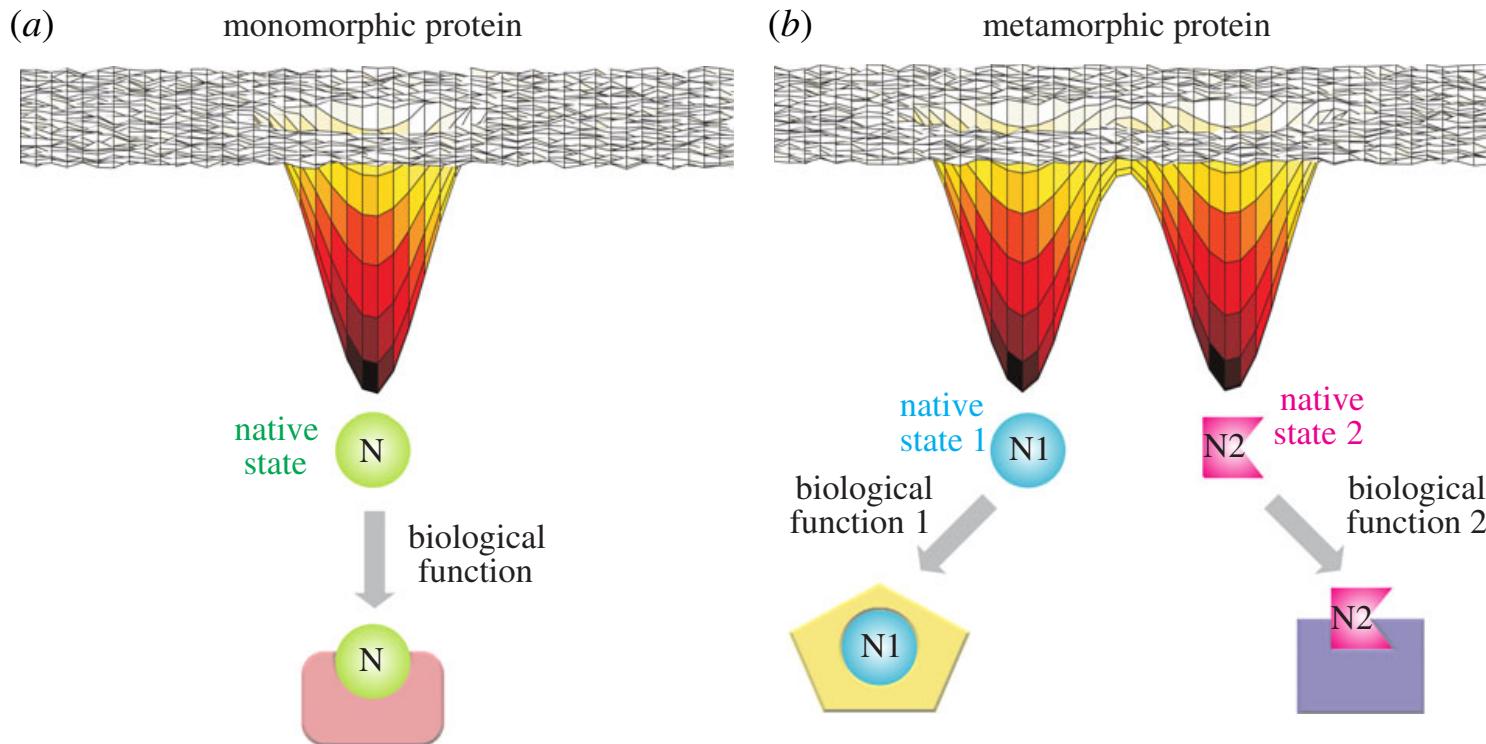
- IDPs are investigated using NMR (which evaluates protein structure in solution, providing not a single defined structure but a range of structural variants).
- Bioinformatic analysis also reveals that intrinsically disordered regions have characteristic sequence signatures, including the presence of:
  1. Low sequence complexity
  2. Low content of bulky hydrophobic amino acid residues
  3. A high content of polar and charged amino acid residues.

# Intrinsically disordered proteins (IDPs)

- Many disordered regions display a high degree of conservation, so several bioinformatic tools were developed to predict based on primary sequence the intrinsically disordered regions or proteins
- IDPs common functions include:
  1. The regulation of transcription
  2. Signal transduction
  3. Protein phosphorylation
- The binding of IDPs to one or more ligands can induce their adoption of a specific conformation.

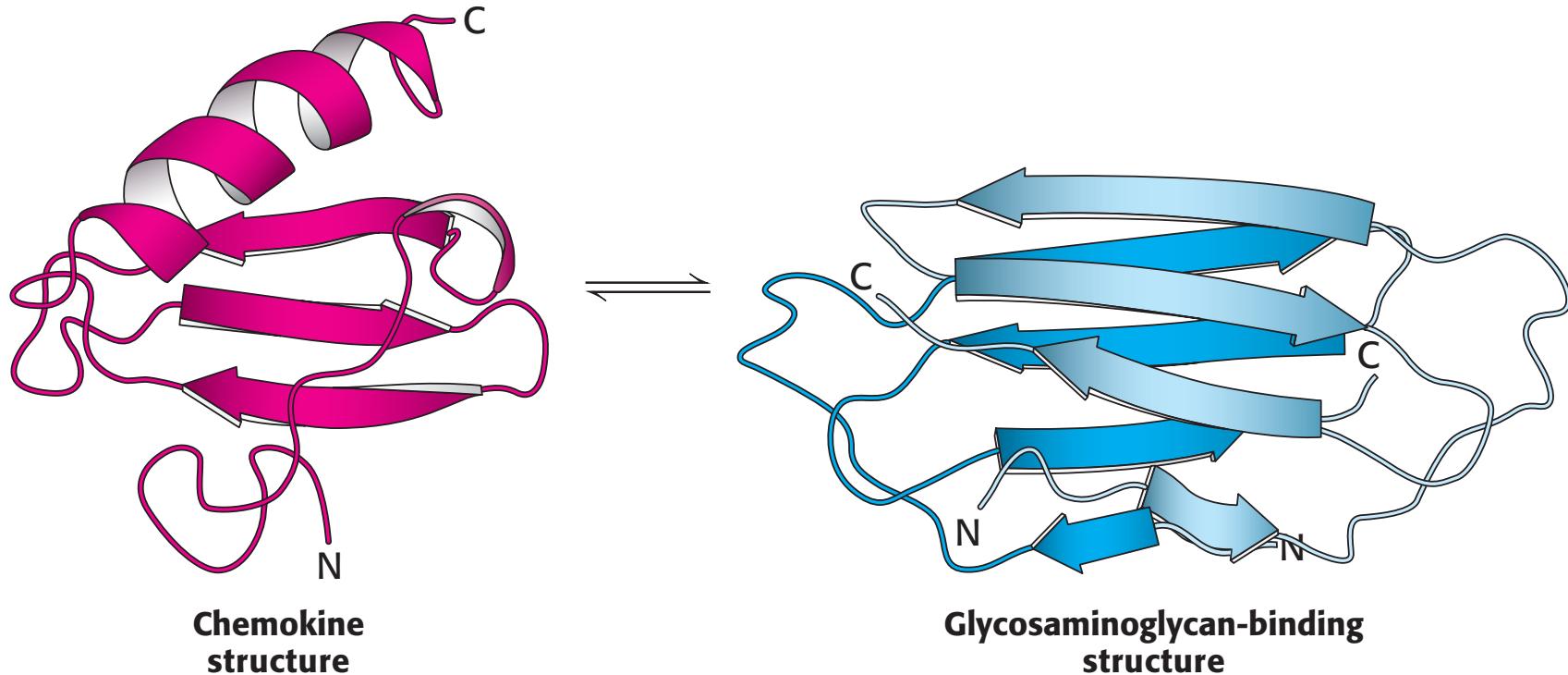
# Metamorphic proteins

- Metamorphic proteins exist in two or more well-defined structures in the absence of ligands or cofactors



# Cytokine lymphotactin

- Metamorphic proteins exist in two or more well-defined structures in the absence of ligands or cofactors



# Protein post-translational modification

- Polypeptides undergo covalent modification, either during or after their ribosomal synthesis is a process called **post-translational modification, PMT**).
- PTMs are characteristic particularly of eukaryotic proteins and are introduced by specific enzymes or enzyme pathways.
- Many occur at the site of a specific characteristic protein sequence (signature sequence) within the protein backbone.

# Protein post-translational modification

**Table 2.7** The more common forms of post-translational modifications that polypeptides may undergo. Refer to text for additional details.

Modification	Comment
Glycosylation	For some proteins glycosylation can increase solubility, influence biological half-life and/or biological activity
Proteolytic processing	Various proteins become biologically active only on their proteolytic cleavage (e.g. some blood factors)
Phosphorylation	Influences/regulates biological activity of various regulatory proteins including polypeptide hormones
Acetylation	Modulation of target protein activity
Acylation	May help some polypeptides interact with/anchor in biological membranes
Amidation	Influences biological activity/stability of some polypeptides
Sulfation	Influences biological activity of some neuropeptides and the proteolytic processing of some polypeptides
Hydroxylation	Important to the structural assembly of certain proteins
$\gamma$ -Carboxyglutamate formation	Important in allowing some blood proteins to bind calcium
ADP-ribosylation	Regulates biological activity of various proteins
Disulfide bond formation	Helps stabilize conformation of some proteins