

Secondary and Higher Order Structures of Proteins

詹迺立 (生物化學暨分子生物學研究所)

Office: 基礎醫學大樓 9F 912

Office hours: by appointment

Phone: 02-23562214 (院內: 262214)

Email: nlchan@ntu.edu.tw

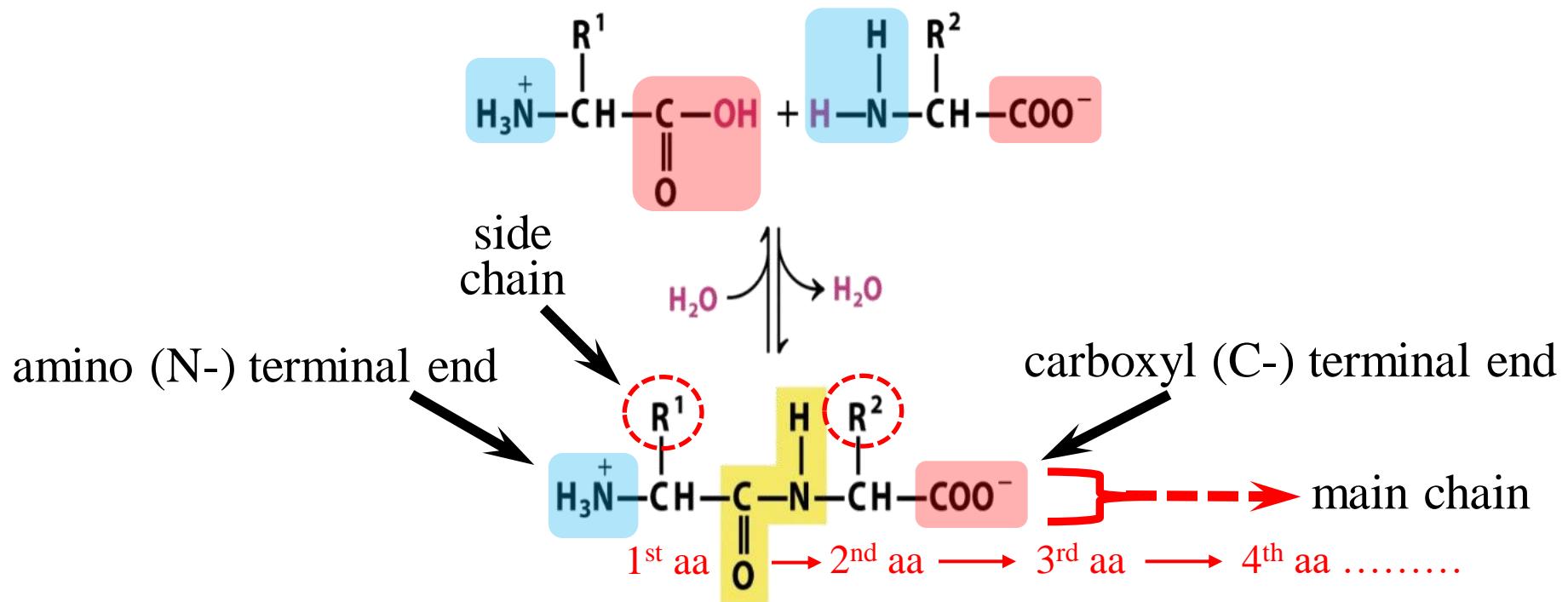
Structure Biology

Structural biology is a key branch of molecular biology, biochemistry, and biophysics, dedicated to determining and interpreting the three-dimensional structures of biological macromolecules, such as proteins and nucleic acids, as well as the complexes they form.

The goal of structural biology is to understand how the three-dimensional structures of biological macromolecules relate to their specific biological functions, that is, to elucidate the structure-function relationships.

Structural biology provides essential insights into cellular processes, disease mechanisms, drug resistance, drug design, and protein/enzyme engineering.

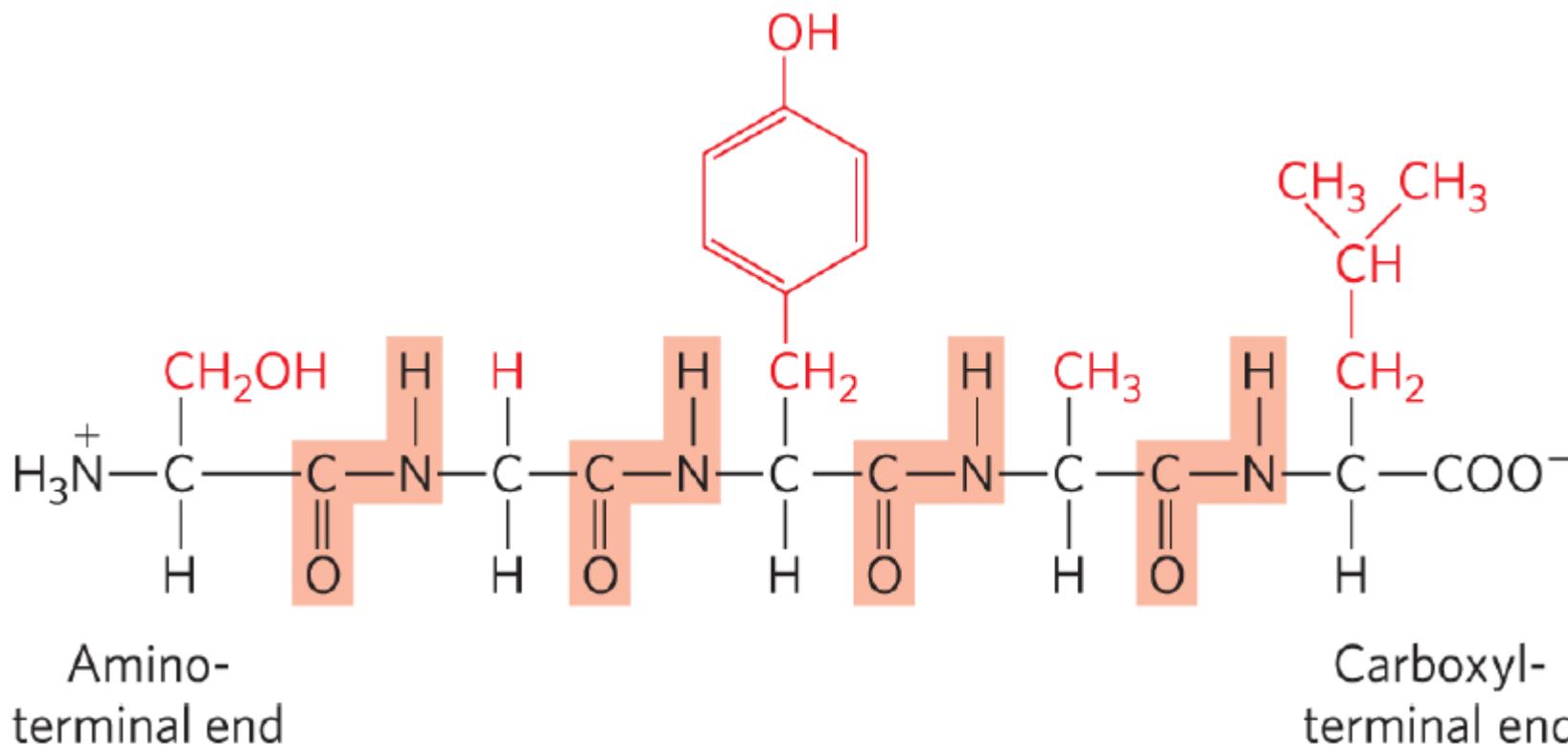
The most important reaction of amino acids is the formation of a “peptide bond”



The peptide bond is polar (hydrophilic, capable of forming hydrogen bonds).

The functional group (the so-called “R-group” or “side chain”) of an amino acid does not participate in peptide bond formation, therefore can exhibit all of its characteristic chemical properties. (eg., $-\text{SH}$ group of cysteine can be oxidized to form disulfide bond; $-\text{OH}$ groups from serine, threonine, and tyrosine can be phosphorylated; etc.)

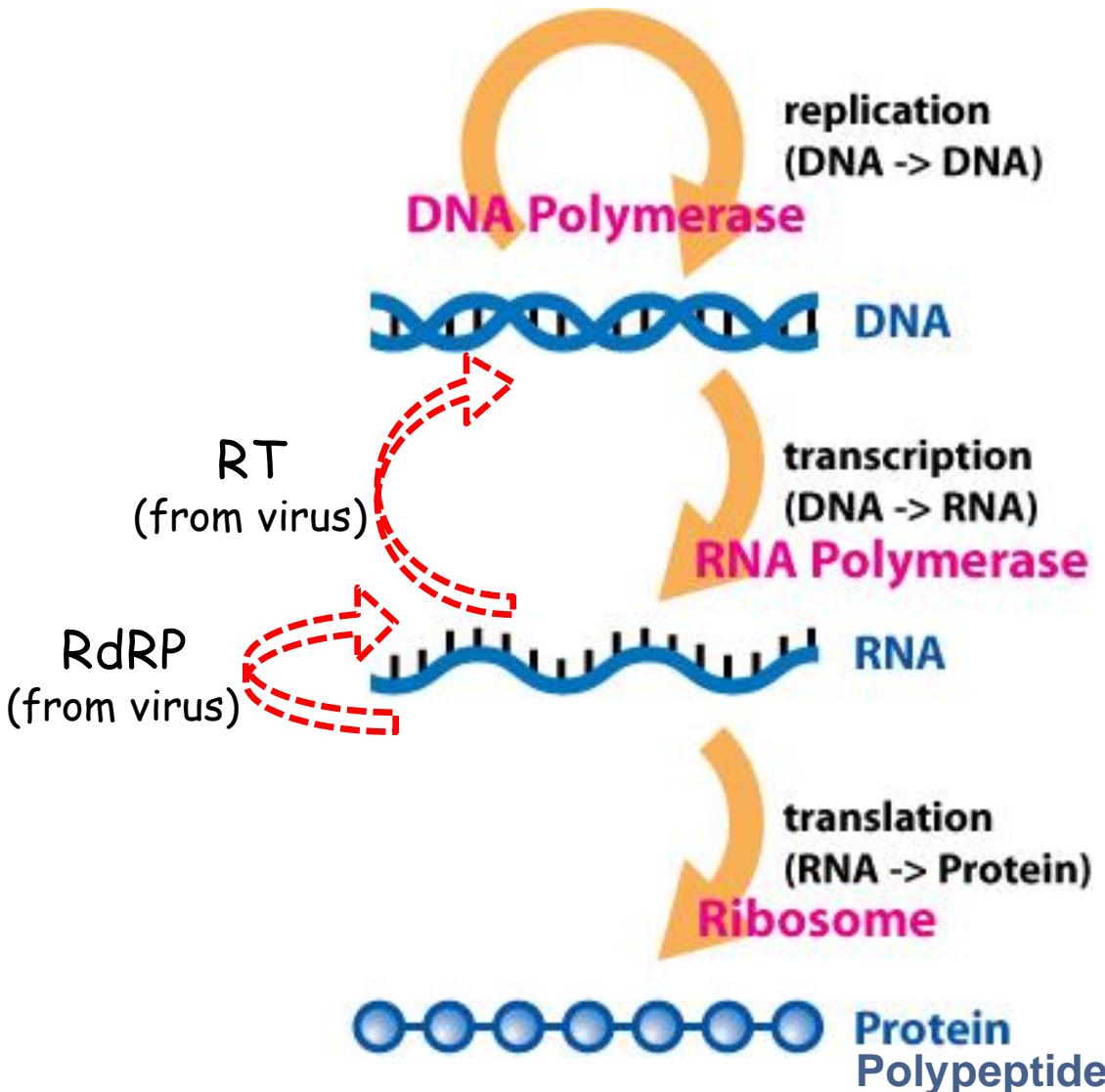
The pentapeptide serylglycyltyrosylalanylleucine, Ser-Gly-Tyr-Ala-Leu, or SGYAL



Lehninger, Fig. 3-14

Biosynthesis of proteins

The "Central Dogma of Molecular Biology": DNA makes RNA, and RNA makes proteins (or simply **DNA → RNA → protein**).



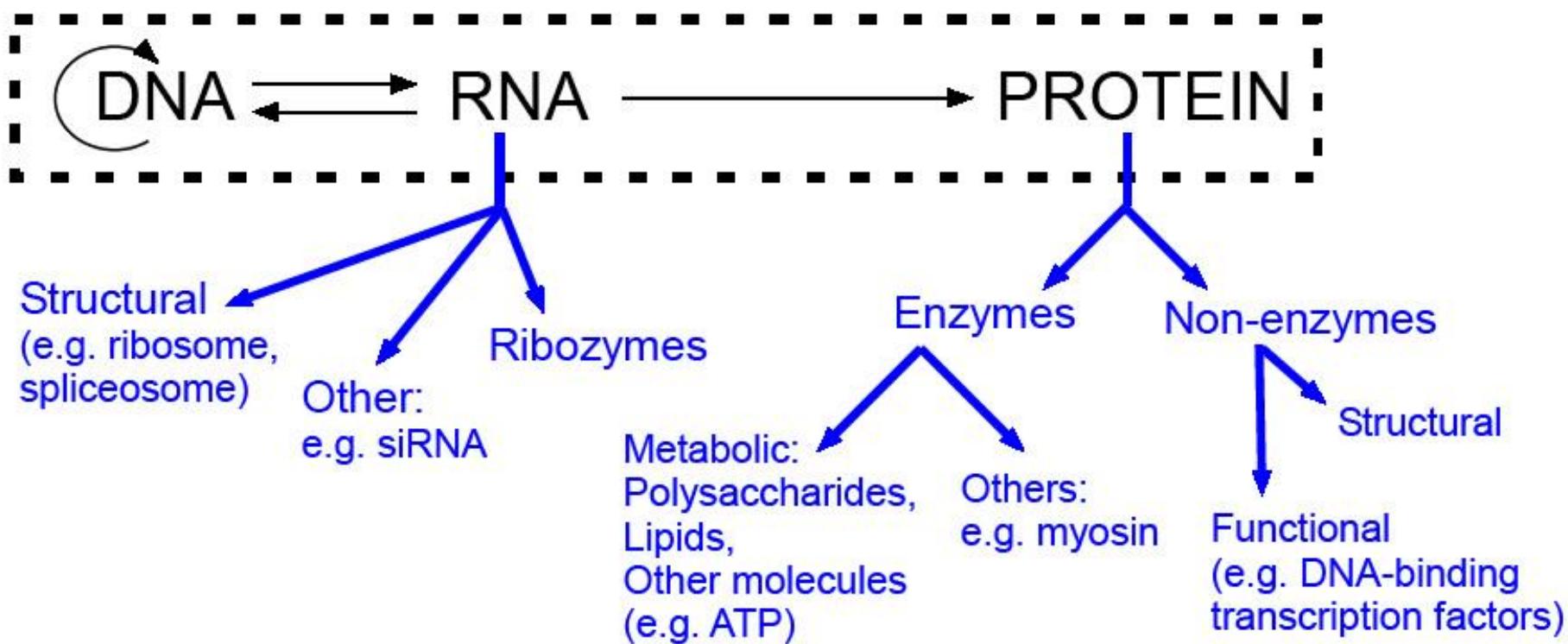
In general, the central dogma is unidirectional and NOT reversible, but there are exceptions (shown in **dashed red arrows**) due to the presence of RNA-dependent RNA polymerase (RdRP) and reverse transcriptase (RT) in some viruses.

RdRP: using RNA as the template to synthesize a new RNA molecule.

RT: using RNA as the template to synthesize a new DNA molecule.

Biological significance of the “Central Dogma”

Central Dogma of Molecular Biology:



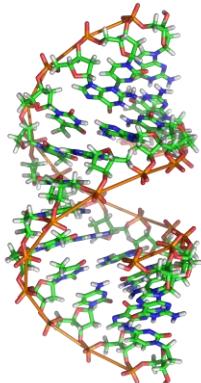
DNA: very few distinct 3D structures

RNA: some but still limited 3D structural diversity

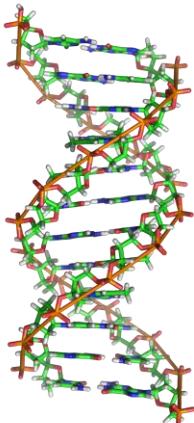
Proteins: highly diversified 3D structures → (functional expansion)

Structures of nucleic acids

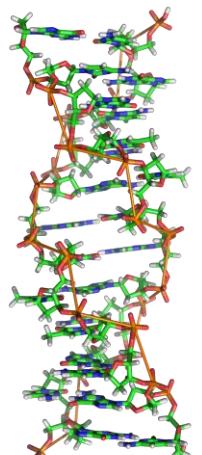
DNA



A form

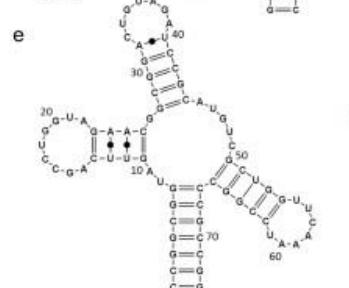
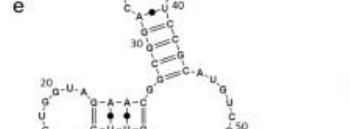
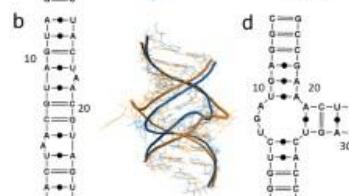
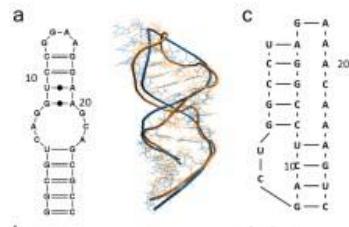


B form

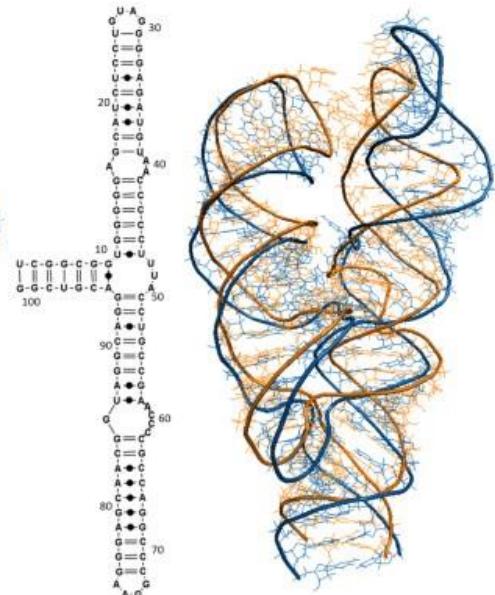


Z form

RNA

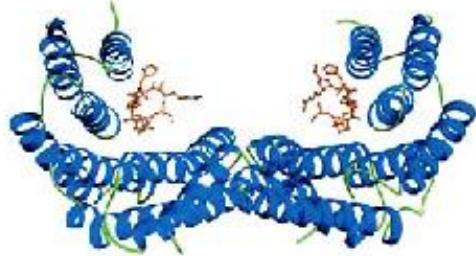


f



Examples of protein tertiary/quaternary structures - 1

14-3-3

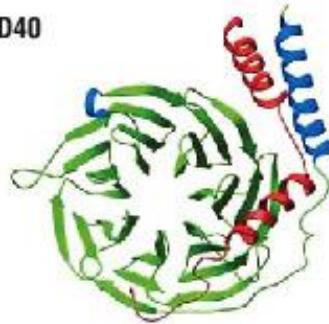


Example: 14-3-3

Function: protein–protein interactions

Specificity: phosphotyrosine

WD40

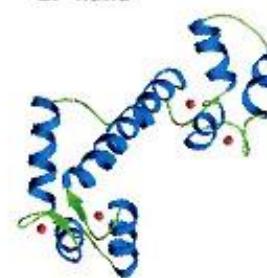


Example: G protein beta subunit

Function: protein–protein interactions; a stable propeller-like platform to which proteins bind either stably or reversibly

Specificity: various

EF-hand

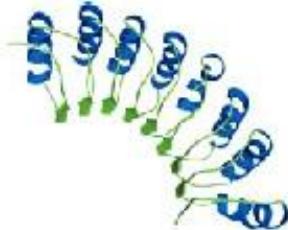


Example: Calmodulin

Function: calcium binding

Specificity: Ca²⁺

LRR

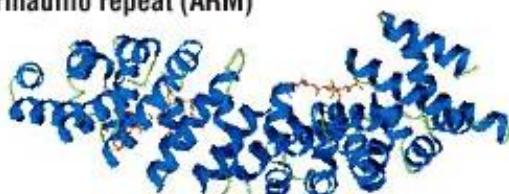


Example: Rpn1

Function: protein–protein interactions

Specificity: various

Armadillo repeat (ARM)

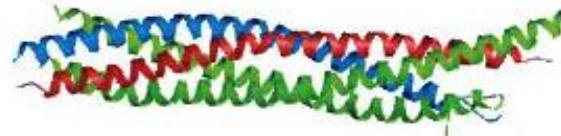


Example: Importin alpha

Function: protein–protein interactions

Specificity: various

SNARE

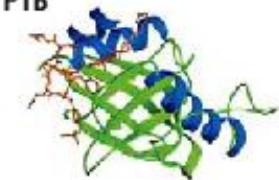


Example: SNAP-25B

Function: protein–protein interactions in intracellular membrane fusion

Specificity: other SNARE domains

PTB



Example: Shc

Function: protein–protein interactions

Specificity: phosphotyrosine

Death domain (DD)



Example: FADD

Function: protein–protein interactions in pathway that triggers apoptosis

Specificity: other DD domains through heterodimers

ANK (ankyrin repeat)

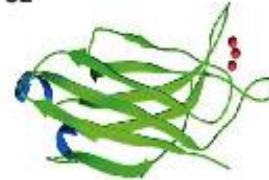


Example: Swi6

Function: protein–protein interactions

Specificity: various

C2

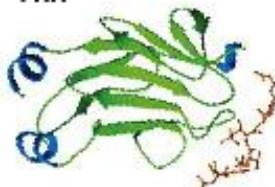


Example: PKC

Function: electrostatic switch

Specificity: phospholipids

FHA

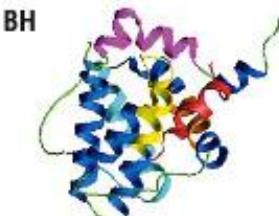


Example: Rad53

Function: protein–protein interactions

Specificity: phosphotyrosine

BH



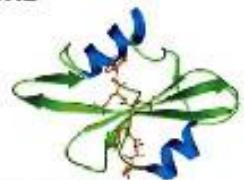
Example: Bcl-XI

Function: protein–protein interactions

Specificity: Other BH domains through heterodimers

Examples of protein tertiary/quaternary structures - 2

SH2



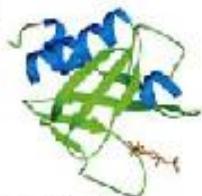
Example: Src
Function: protein–protein interactions
Specificity: phosphotyrosine

SH3



Example: Sem5
Function: protein–protein interactions
Specificity: proline-rich sequences

PH



Example: PLC- δ
Function: recruitment of proteins to the membrane
Specificity: phosphoinositides

SAM



Example: EphA4
Function: protein–protein interactions via homo- and heterodimers
Specificity: other SAM domains

Bromo



Example: P/CAF
Function: protein–protein interactions in chromatin remodeling
Specificity: acetylated lysine

PDZ



Example: PSD-95
Function: protein–protein interactions, often involving transmembrane proteins or ion channels
Specificity: -XXXV/I-COOH

GYF



Example: CD2
Function: protein–protein interactions
Specificity: proline-rich sequences

Chromo



Example: Mouse modifier protein 1
Function: protein–protein interactions in chromatin remodeling
Specificity: methylated lysine

FYVE



Example: Vps27p
Function: Regulation of signaling
Specificity: phosphatidyl-inositol-3-phosphate

RING finger



Example: c-Cbl
Function: protein–protein interactions in ubiquitin-dependent degradation and transcription regulation
Specificity: various

WW



Example: Pin1
Function: protein–protein interactions
Specificity: proline-rich sequences

LIM



Example: CRP2
Function: protein–protein interactions, usually in transcription regulation
Specificity: various

F-box



Example: Skp2
Function: protein–protein interactions in ubiquitin-dependent protein degradation
Specificity: various

C1



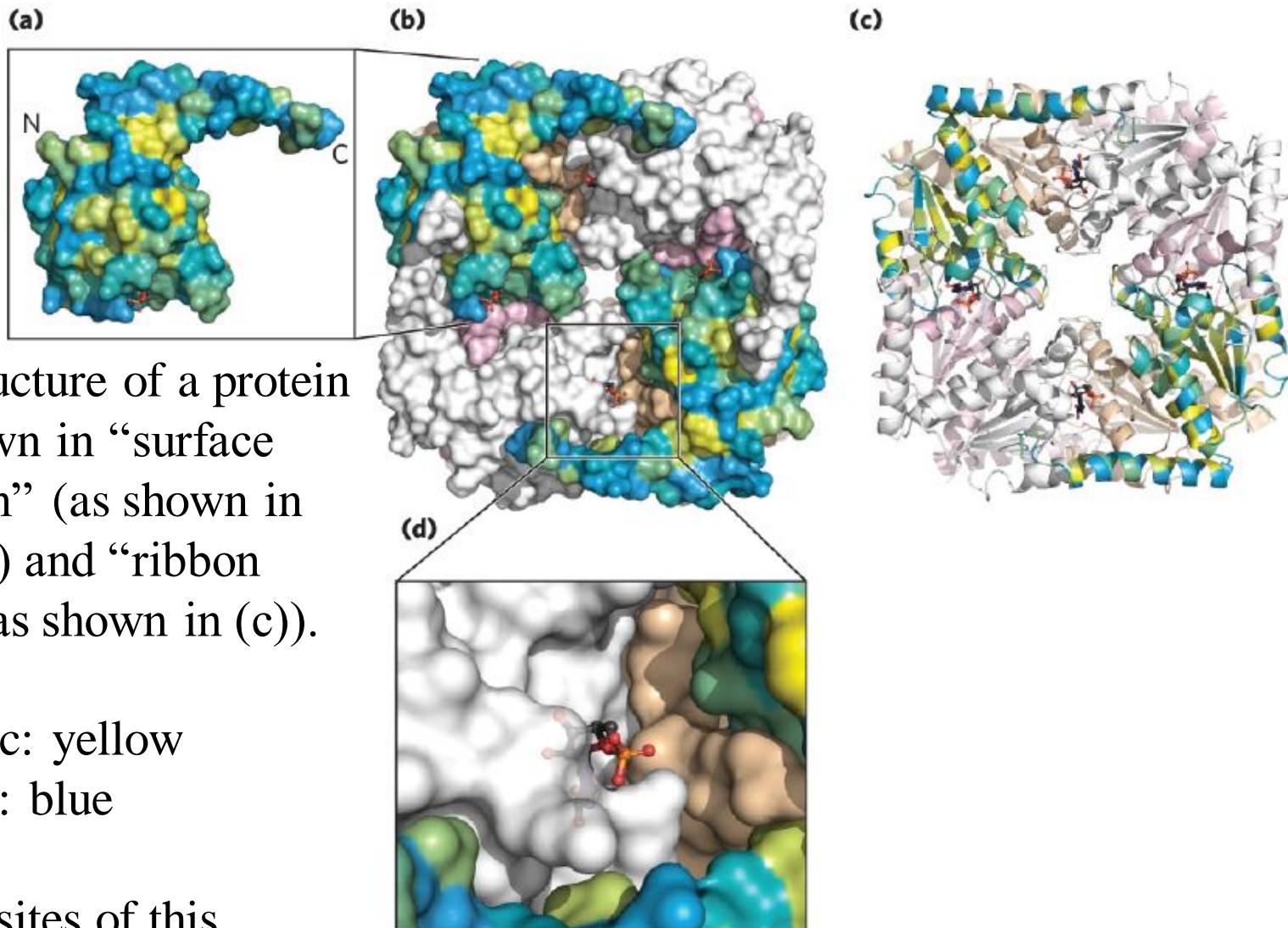
Example: PKC
Function: recruitment of proteins to the membrane
Specificity: phospholipids

Fibronectin



Example: Fibronectin III
Function: protein–protein interactions in cell adhesion to surfaces
Specificity: RGD motif of integrins

Relationship between protein structure and function



The 3D structure of a protein can be shown in “surface presentation” (as shown in (a), (b), (d)) and “ribbon diagram” (as shown in (c)).

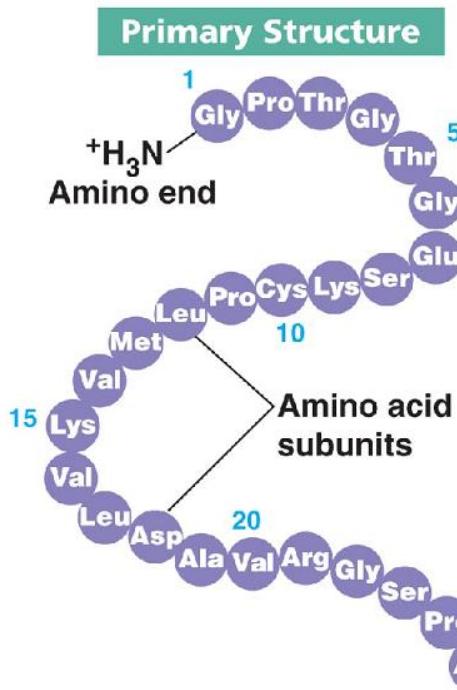
hydrophobic: yellow

hydrophilic: blue

The active sites of this enzyme (PurE) is located at subunit interface.

Lehninger, Fig. 4-1

The process of forming protein's 3D structure is called "protein folding"

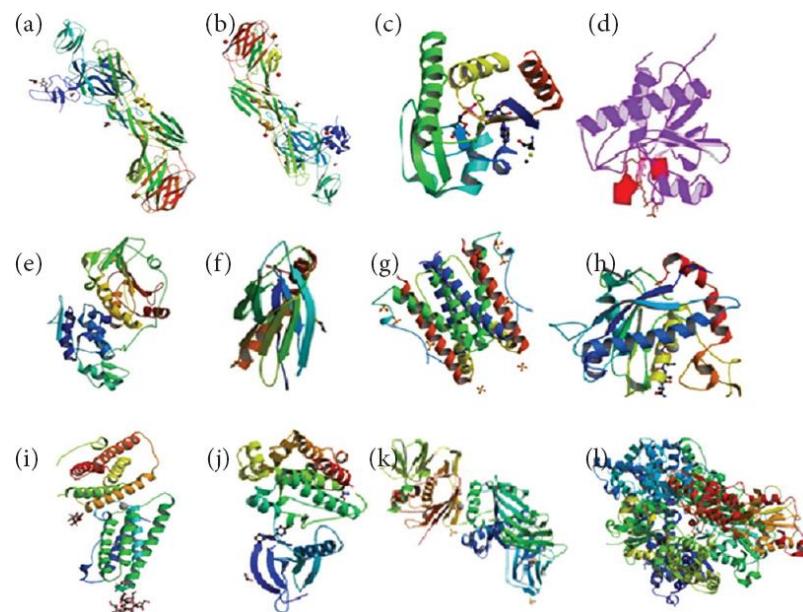
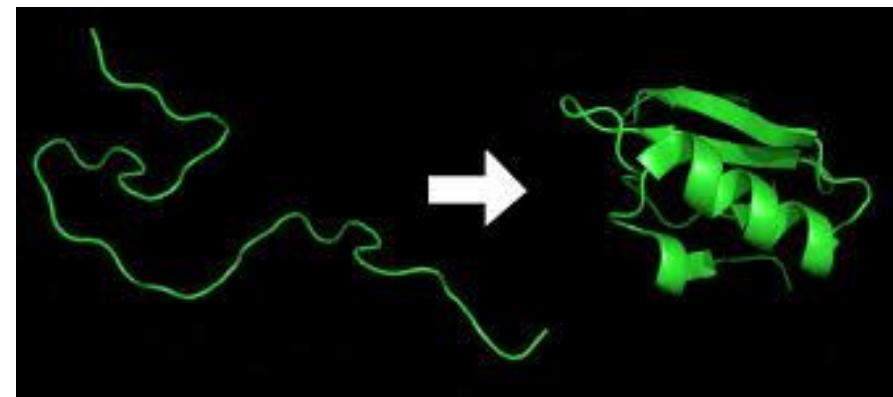


Amino Acid	Abbreviations
Alanine	Ala; A
Arginine	Arg; R
Asparagine	Asn; N
Aspartic acid	Asp; D
Cysteine	Cys; C
Glutamic acid	Glu; E
Glutamine	Gln; Q
Glycine	Gly; G
Histidine	His; H
Isoleucine	Ile; I
Leucine	Leu; L
Lysine	Lys; K
Methionine	Met; M
Phenylalanine	Phe; F
Proline	Pro; P
Serine	Ser; S
Threonine	Thr; T
Tyrosine	Tyr; Y
Tryptophan	Trp; W
Valine	Val; V

different sequences



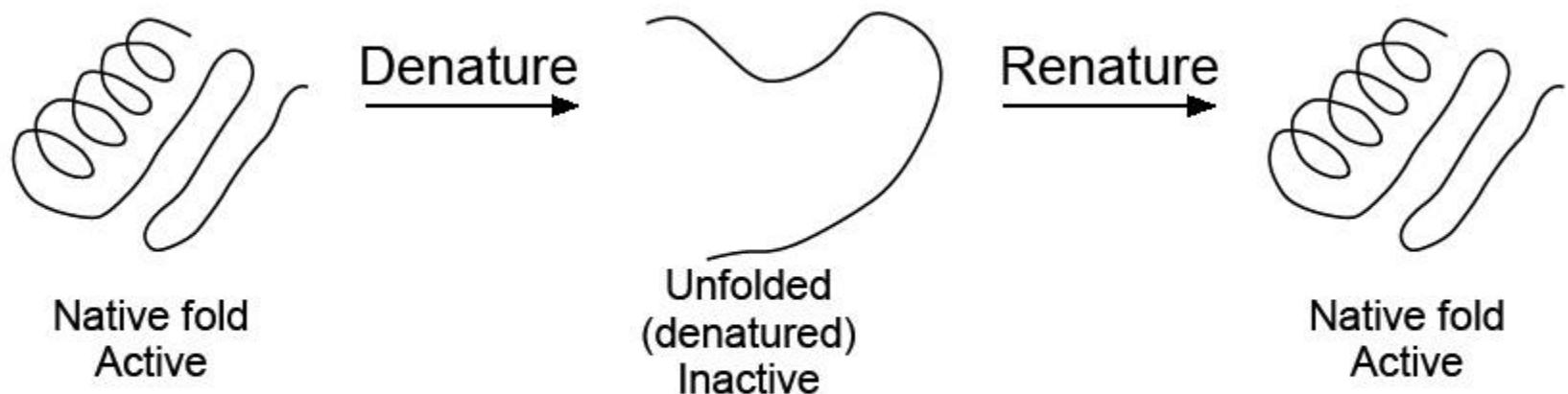
different 3D structures



Structure-Function Relationships

The functional properties of a protein is depended upon its three-dimensional structure(s).

Anfinsen expt. (1955): Ribonuclease



- 1) Molecular sequence sufficient to determine 3D-fold (structure)
- 2) Activity (interactions) depends on correctly folded structure

Proteins in their **functional, folded conformations** are called “**native**” proteins. (“**Denatured**” proteins lose their 3D structure and therefore cannot function.)

“New Central Dogma of Biology”

To emphasize how the three-dimensional structure of a protein may lead to better understanding of protein function, Petsko and Ringe proposed the “**New Central Dogma of Biology**”.

“sequences → 3D structures → functions”

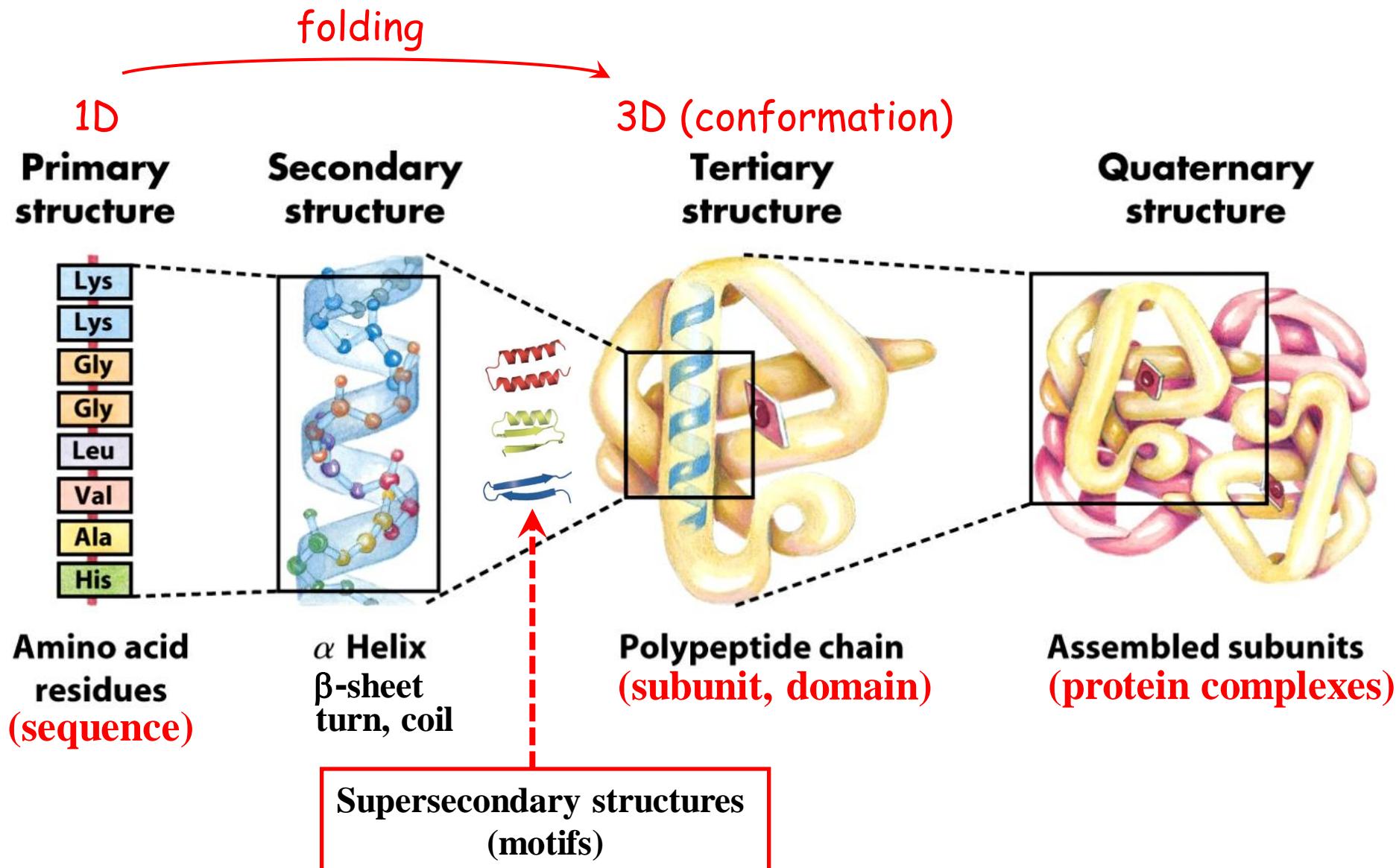
"I'm sorry, but I just don't understand anything in biology unless I know what the players (biomacromolecules) look like."

~**Don Wiley** 1944-2001 (Harvard)

“It is very easy to answer many fundamental biological questions; you just look at the players (biomacromolecules).”

~**Richard Feynman** 1918-1988 (Caltech)

Four levels of protein structure

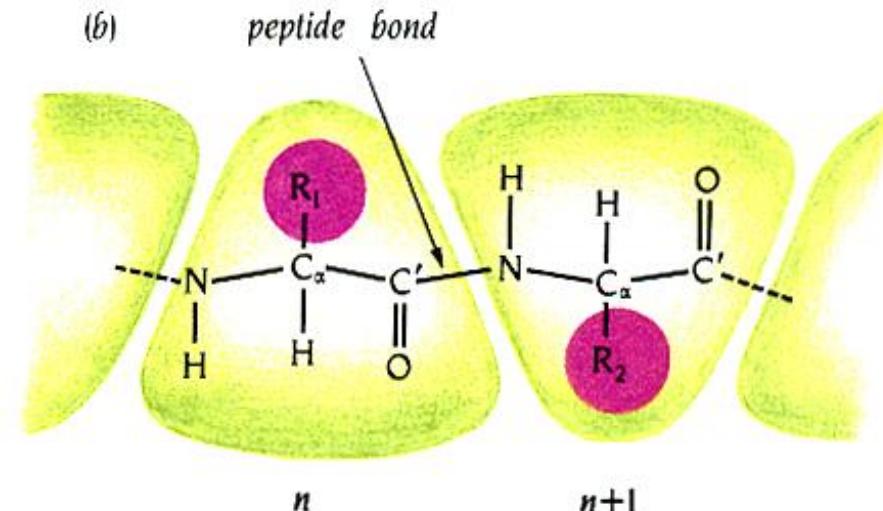


Protein molecules can be described in terms ***of four levels*** of structural organization, as defined below:

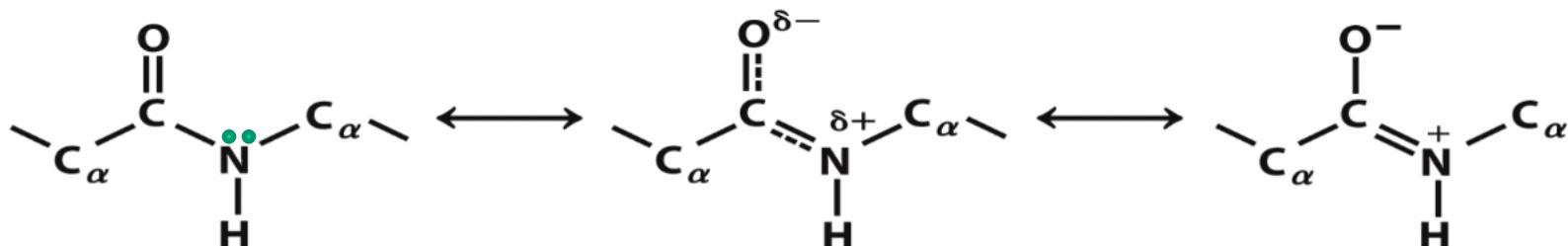
- **Primary structure** - the amino acid sequence
- **Secondary structure** - local spatial arrangement of the polypeptide main chain, forming common patterns like α helices, β sheets, and connecting elements such as turns and loops.
- **Tertiary structure** - overall folding of a monomeric protein or subunit
- **Quaternary structure** - subunit association (**protein complex made of ≥ 2 subunits**)

Amino acids are linked together by peptide bond to form proteins or polypeptides

Peptide bond (an amide bond):

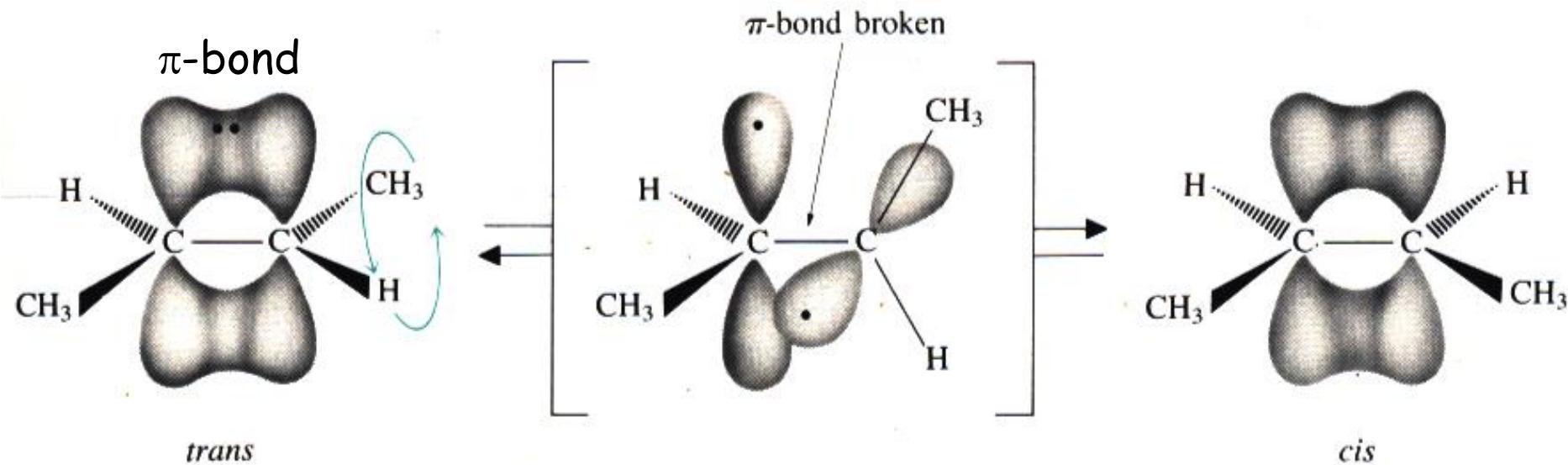
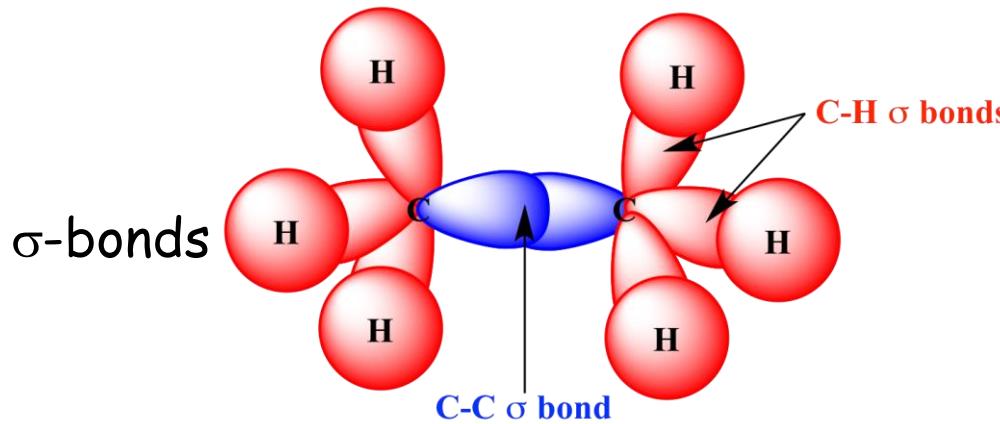


Peptide bonds are rigid and planar, due to electron sharing between the carboxyl carbon and the amide nitrogen which give it a partial double-bond character.

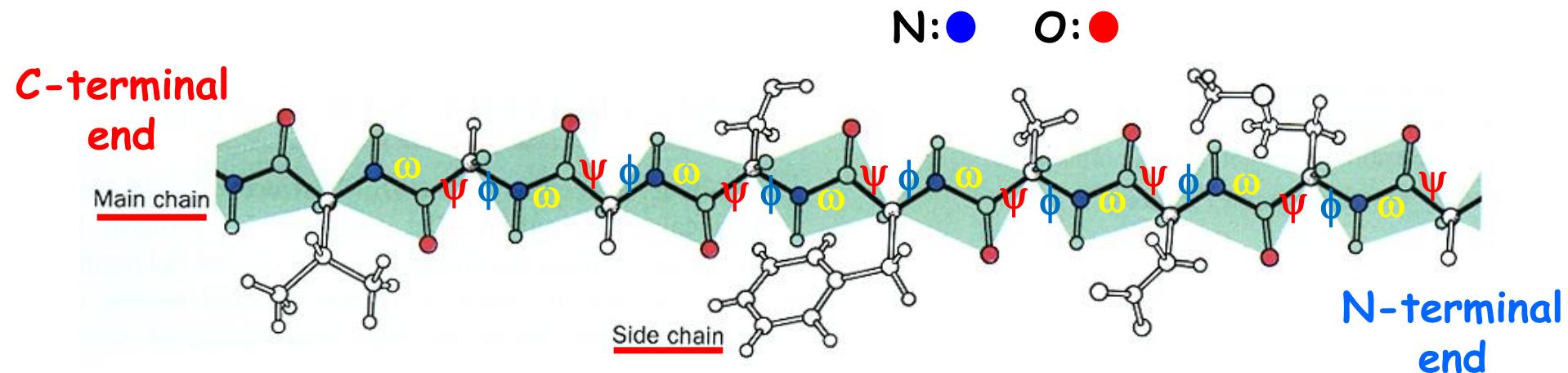


The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive¹⁶ charge, setting up a small electric dipole.

Whereas a single bond (σ -bond) can rotate freely, rotation of a double bond (π -bond) is usually disfavored, and only two conformations (**cis** and **trans**) are allowed.



The conformation of a polypeptide chain is defined by the three main-chain dihedral (conformational) angles



ϕ : rotation around the $\text{C}\alpha\text{-N}$ bond (single bond)

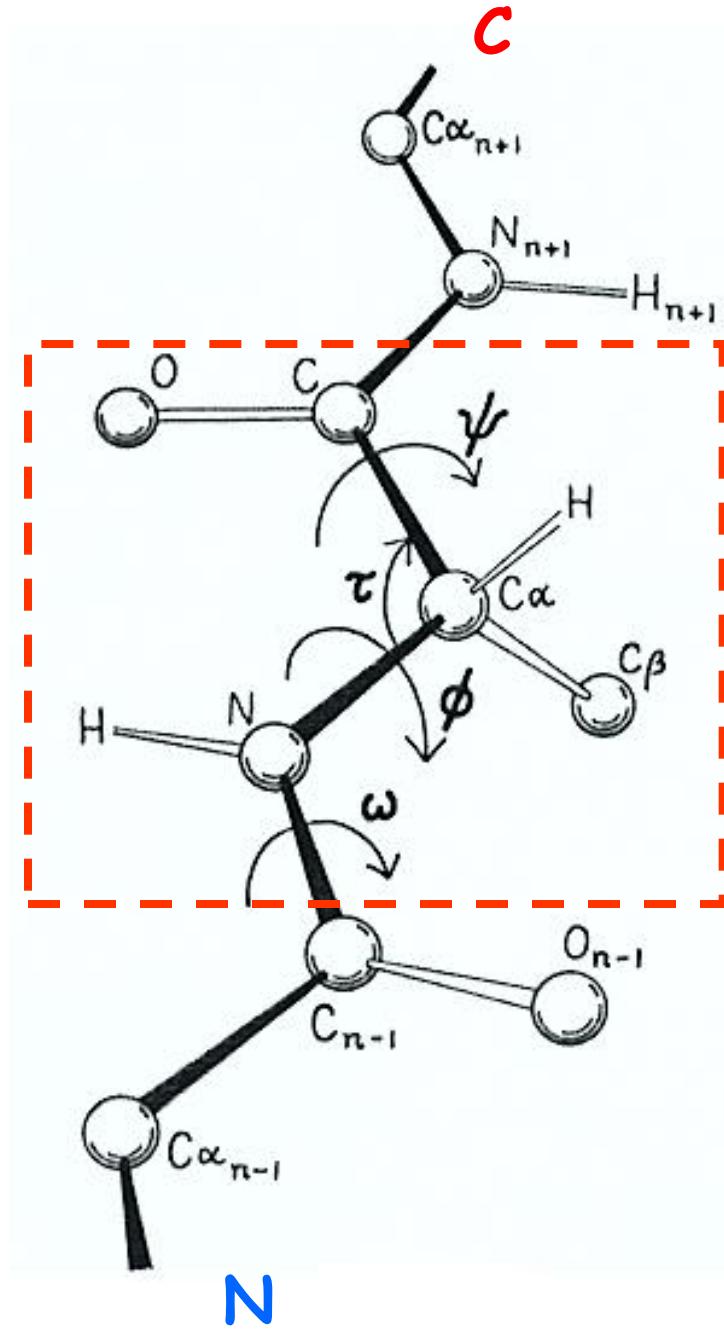
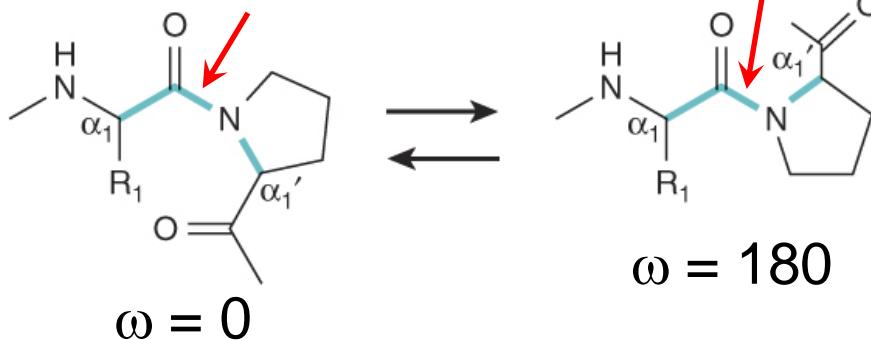
ψ : rotation around the $\text{C}\alpha\text{-C}$ bond (single bond)

ω : rotation around the peptide bond (partially double bond, so usually 180°, except for proline)

The backbone conformation of a polypeptide can be defined by the rotation of three dihedral angles ω , ϕ , and ψ .

Usually, the angle ω is constrained to 180 (trans) degree.

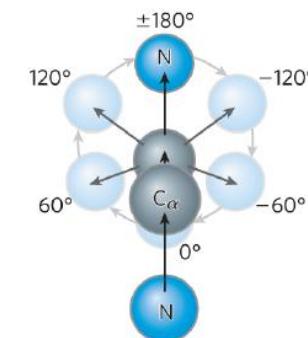
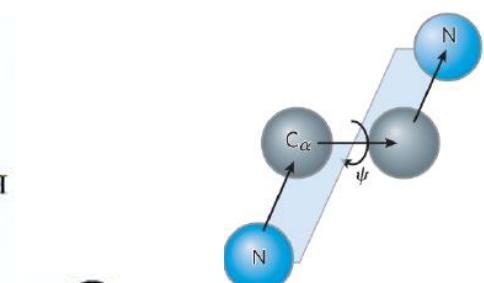
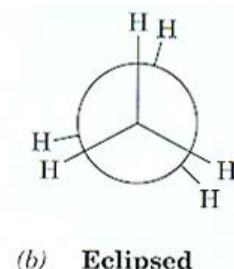
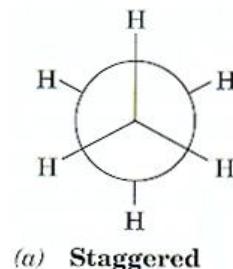
However, the peptide bond formed at the N-terminus of proline can be either 180 (trans) or 0 (cis) degree.



Steric constraints define the observed dihedral angle

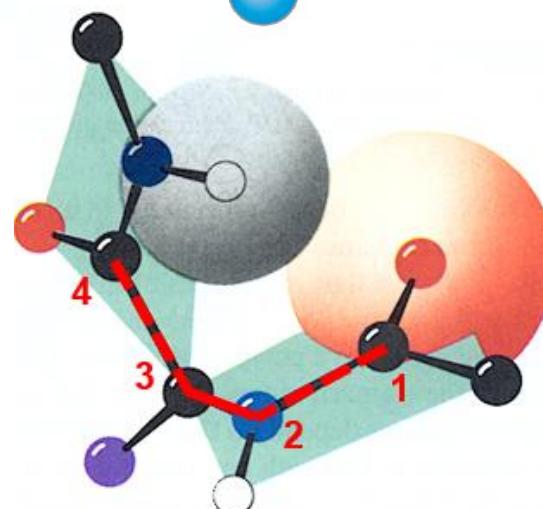
There are several steric constraints on the torsion angles ϕ and ψ of a polypeptide backbone that limit its conformational range.

Staggered conformation is more stable than the eclipsed form because of smaller steric repulsion.

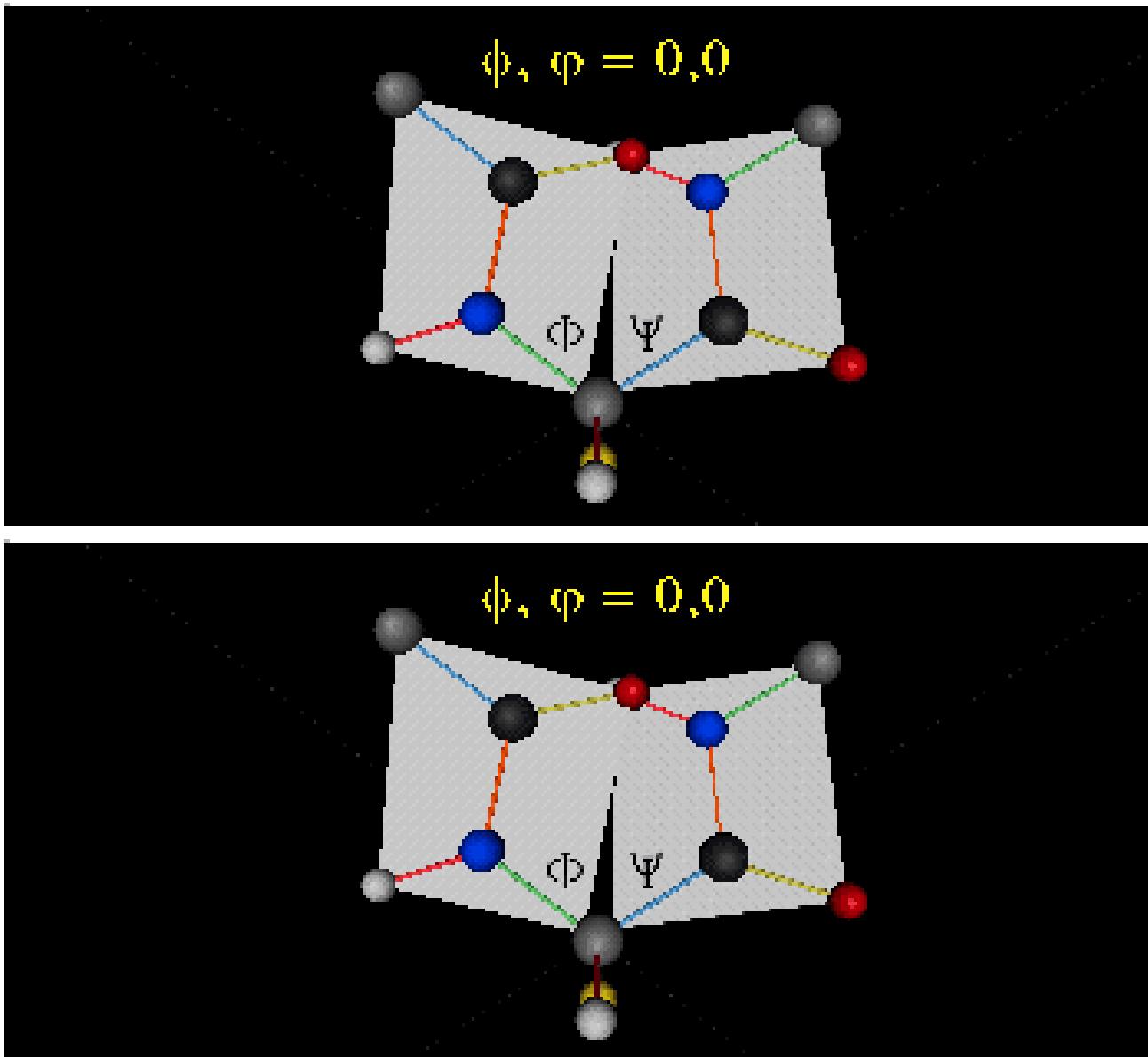


Lehninger, Fig. 4-2c,d

Some conformations are sterically forbidden due to steric clashes between atoms..

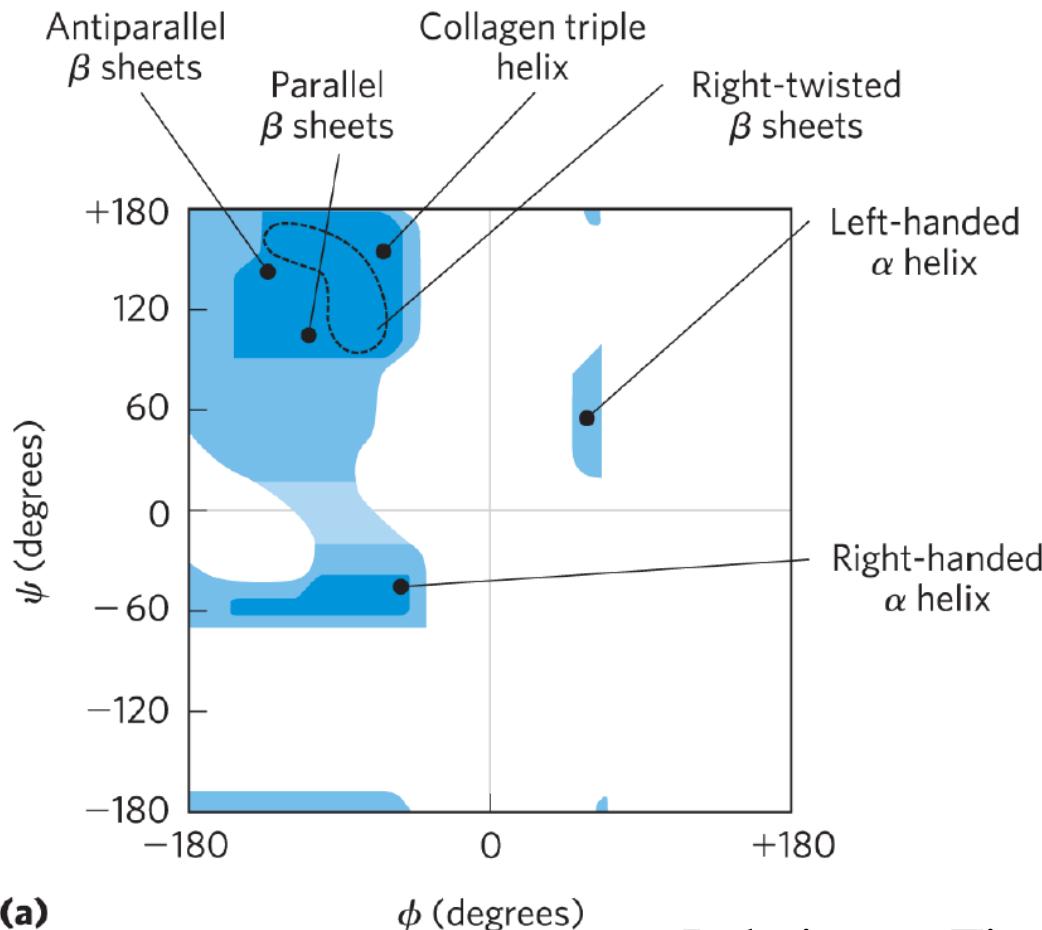


Construction of Ramachandran Plot (define the allowed ϕ and φ)



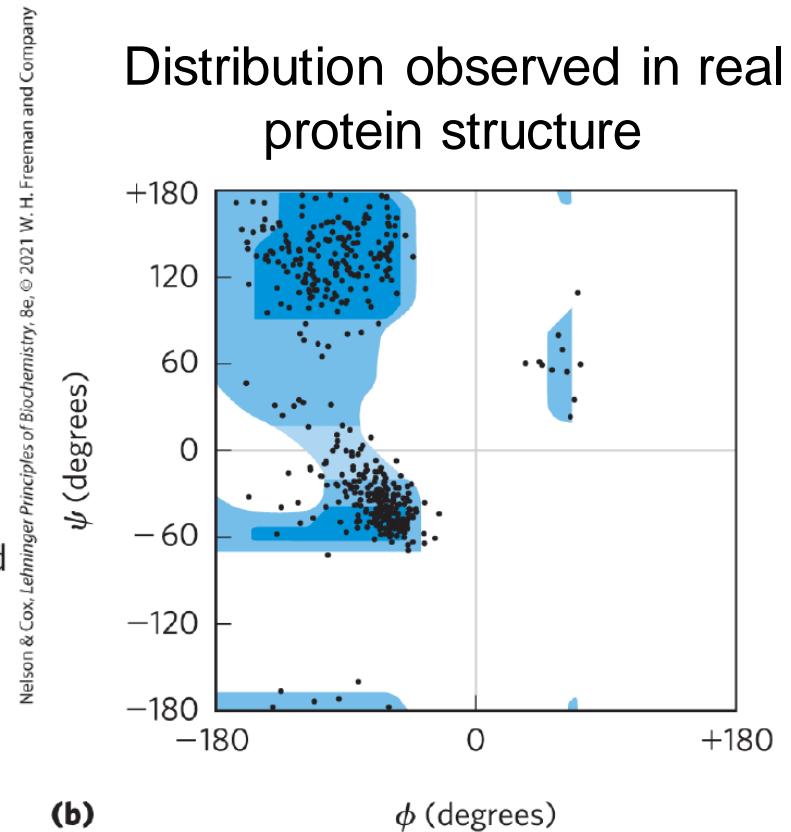
Allowed conformations of polypeptides are indicated by the Ramachandran plot

Theoretically derived diagram showing which atomic collision produce the restriction on main chain dihedral angles ϕ and ψ .



(a)

Lehninger, Fig. 4-8

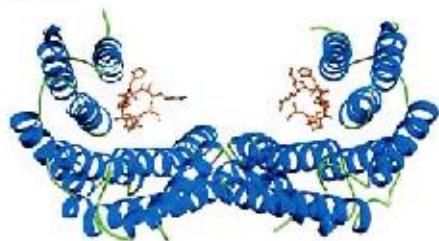


(b)

Note that the distribution of ϕ and ψ is NOT random!²²

Examples of protein tertiary/quaternary structures - 1

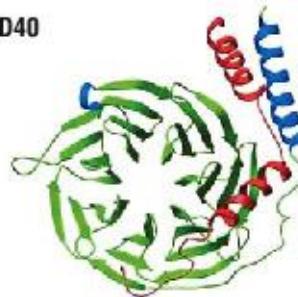
14-3-3



Example: 14-3-3

Function: protein–protein interactions
Specificity: phosphotyrosine

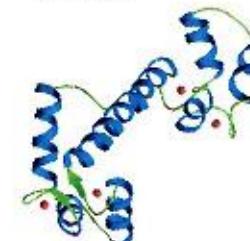
WD40



Example: G protein beta subunit

Function: protein–protein interactions;
a stable propeller-like platform to which
proteins bind either stably or reversibly
Specificity: various

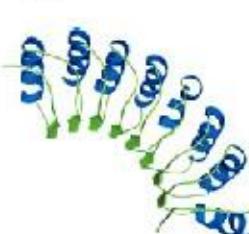
EF-hand



Example: Calmodulin

Function: calcium binding
Specificity: Ca²⁺

LRR



Example: Rpn1

Function: protein–protein
interactions
Specificity: various

Armadillo repeat (ARM)



Example: Importin alpha

Function: protein–protein interactions
Specificity: various

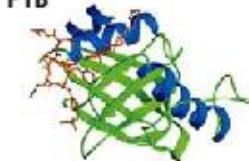
SNARE



Example: SNAP-25B

Function: protein–protein interactions
in intracellular membrane fusion
Specificity: other SNARE domains

PTB



Example: Shc

Function: protein–protein
interactions
Specificity: phosphotyrosine

Death domain (DD)



Example: FADD

Function: protein–protein
interactions in pathway that
triggers apoptosis
Specificity: other DD domains
through heterodimers

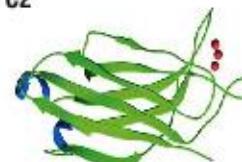
ANK (ankyrin repeat)



Example: Swi6

Function: protein–protein
interactions
Specificity: various

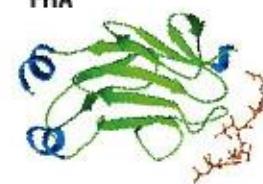
C2



Example: PKC

Function: electrostatic switch
Specificity: phospholipids

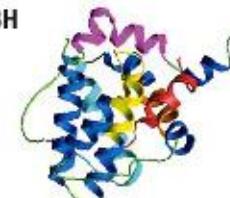
FHA



Example: Rad53

Function: protein–protein
interactions
Specificity: phosphotyrosine

BH



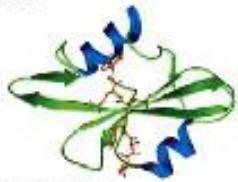
Example: Bcl-XI

Function: protein–protein
interactions
Specificity: Other BH domains
through heterodimers

The overall characteristics of “soluble” proteins are usually classified into two types: **globular** vs **fibrous** (based on the “axial ratio”: globular (<1:3), fibrous (>1:10)).

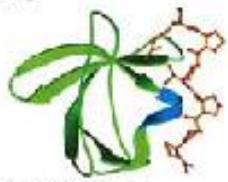
Examples of protein tertiary/quaternary structures - 2

SH2



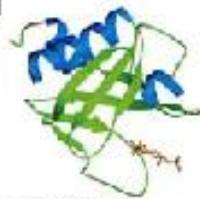
Example: Src
Function: protein–protein interactions
Specificity: phosphotyrosine

SH3



Example: Sem5
Function: protein–protein interactions
Specificity: proline-rich sequences

PH



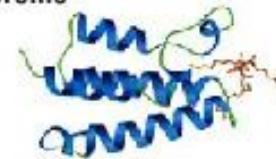
Example: PLC- δ
Function: recruitment of proteins to the membrane
Specificity: phosphoinositides

SAM



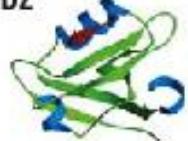
Example: EphA4
Function: protein–protein interactions via homo- and heterodimers
Specificity: other SAM domains

Bromo



Example: P/CAF
Function: protein–protein interactions in chromatin remodeling
Specificity: acetylated lysine

PDZ



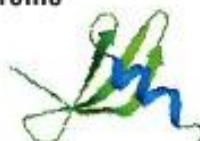
Example: PSD-95
Function: protein–protein interactions, often involving transmembrane proteins or ion channels
Specificity: -XXXV/I-COOH

GYF



Example: CD2
Function: protein–protein interactions
Specificity: proline-rich sequences

Chromo



Example: Mouse modifier protein 1
Function: protein–protein interactions in chromatin remodeling
Specificity: methylated lysine

FYVE



Example: Vps27p
Function: Regulation of signaling
Specificity: phosphatidyl-inositol-3-phosphate

RING finger



Example: c-Cbl
Function: protein–protein interactions in ubiquitin-dependent degradation and transcription regulation
Specificity: various

WW



Example: Pin1
Function: protein–protein interactions
Specificity: proline-rich sequences

LIM



Example: CRP2
Function: protein–protein interactions, usually in transcription regulation
Specificity: various

F-box



Example: Skp2
Function: protein–protein interactions in ubiquitin-dependent protein degradation
Specificity: various

C1



Example: PKC
Function: recruitment of proteins to the membrane
Specificity: phospholipids

Fibronectin

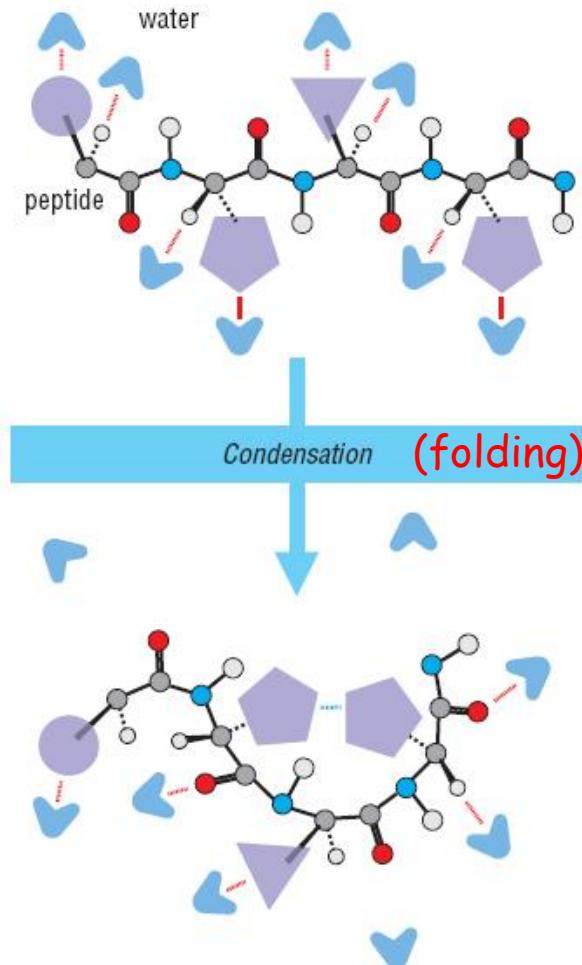


Example: Fibronectin III
Function: protein–protein interactions in cell adhesion to surfaces
Specificity: RGD motif of integrins

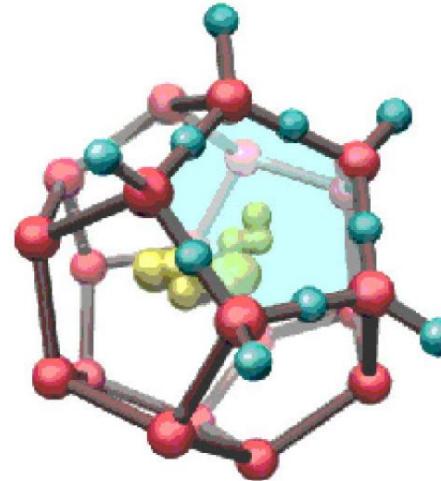
Regions of ordered secondary structure arise when a series of amino acids adopt similar ϕ and ψ values.

Primary structure determines how a protein folds into a unique 3D structure

At physiological condition, the polypeptide chains of soluble proteins have a tendency to fold into their native conformations.



Entropic costly “Clathrate”-type cage structure of H_2O molecules will surround a hydrophobic molecule.

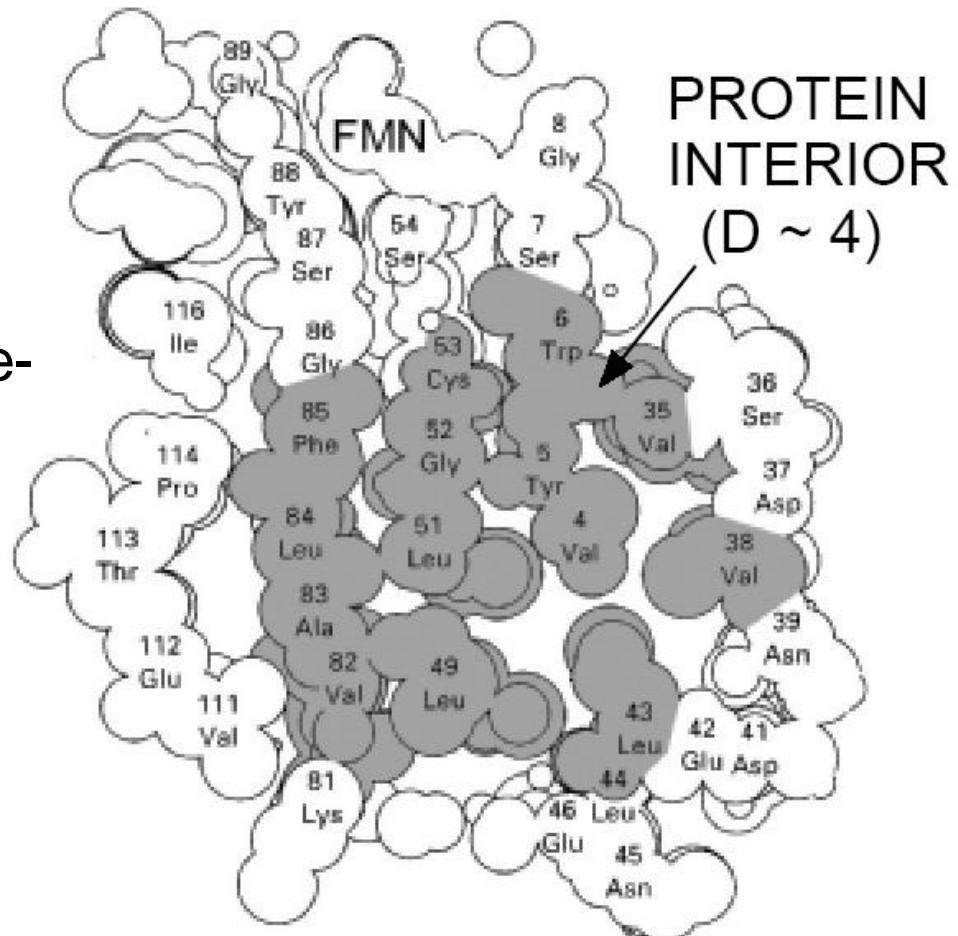


Due to “**hydrophobic effect**”, the interior of all soluble proteins are hydrophobic.

The interior of all soluble proteins is hydrophobic (hydrophobic core)

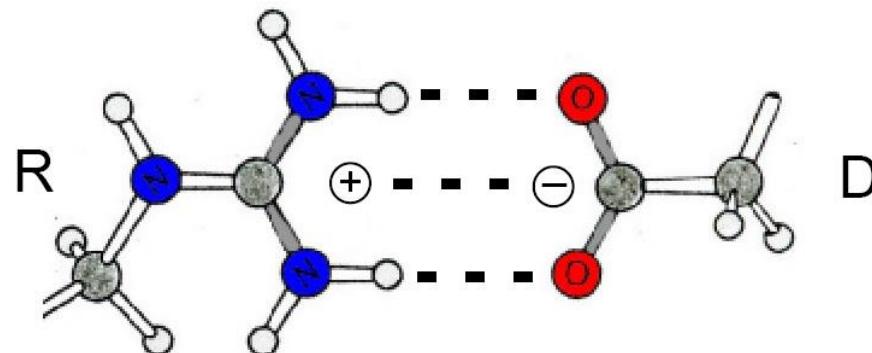
- A compact core formed by hydrophobic residues. (Buried polar or charged residues have bonded partners of opposite charge.)
- Protein surface is composed of mostly polar or charged residues.

Cross section of the space-filling model of a flavo-mononucleotide (FMN) binding protein.

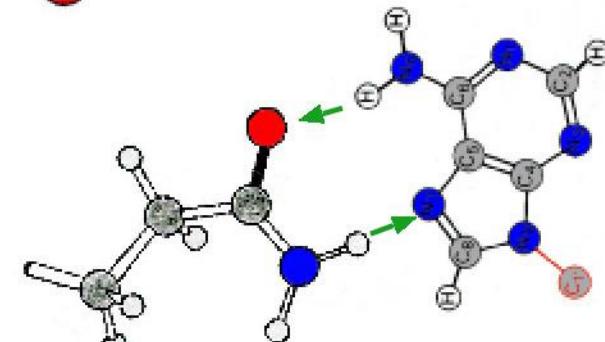
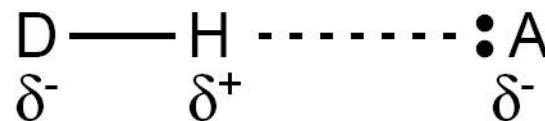


In addition to the hydrophobic effect, three more non-covalent forces (ionic interactions (or salt bridges), hydrogen bonds, and van der Waals forces) are involved in the folding of polypeptides to form the native structure of a protein. A protein's conformation is usually stabilized by numerous non-covalent interactions.

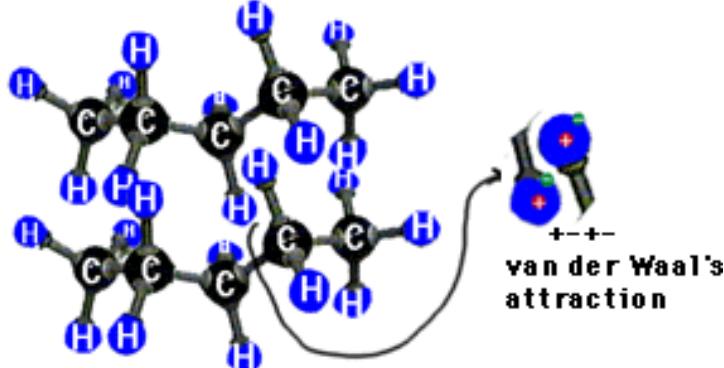
ionic interaction
(salt bridge)



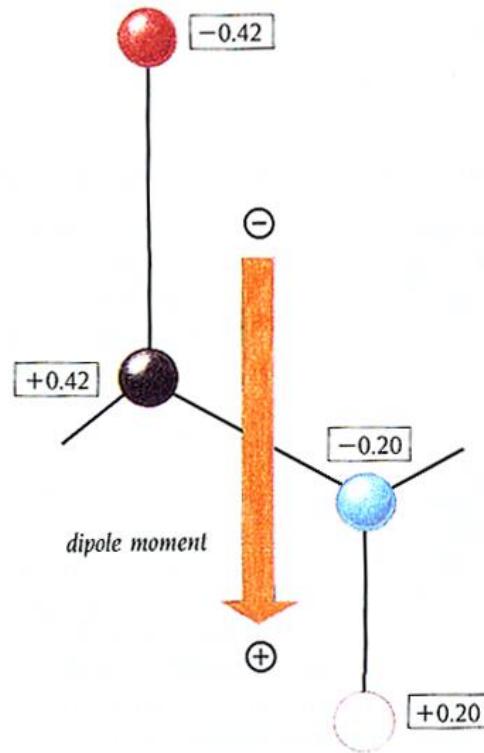
hydrogen bond



van der Waals interaction:
interactions between non-polar groups.



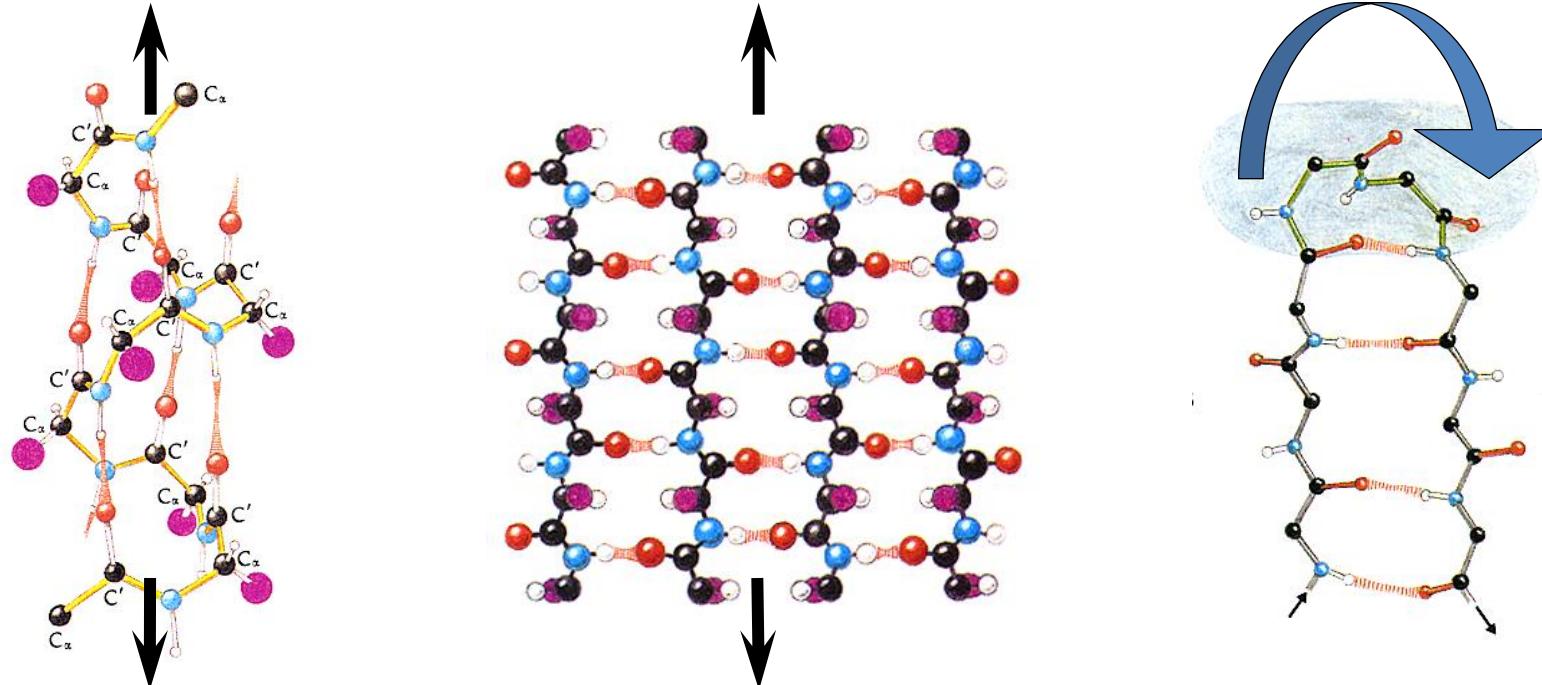
How to fold the polar and hydrophilic main chain into the hydrophobic interior of soluble globular proteins??



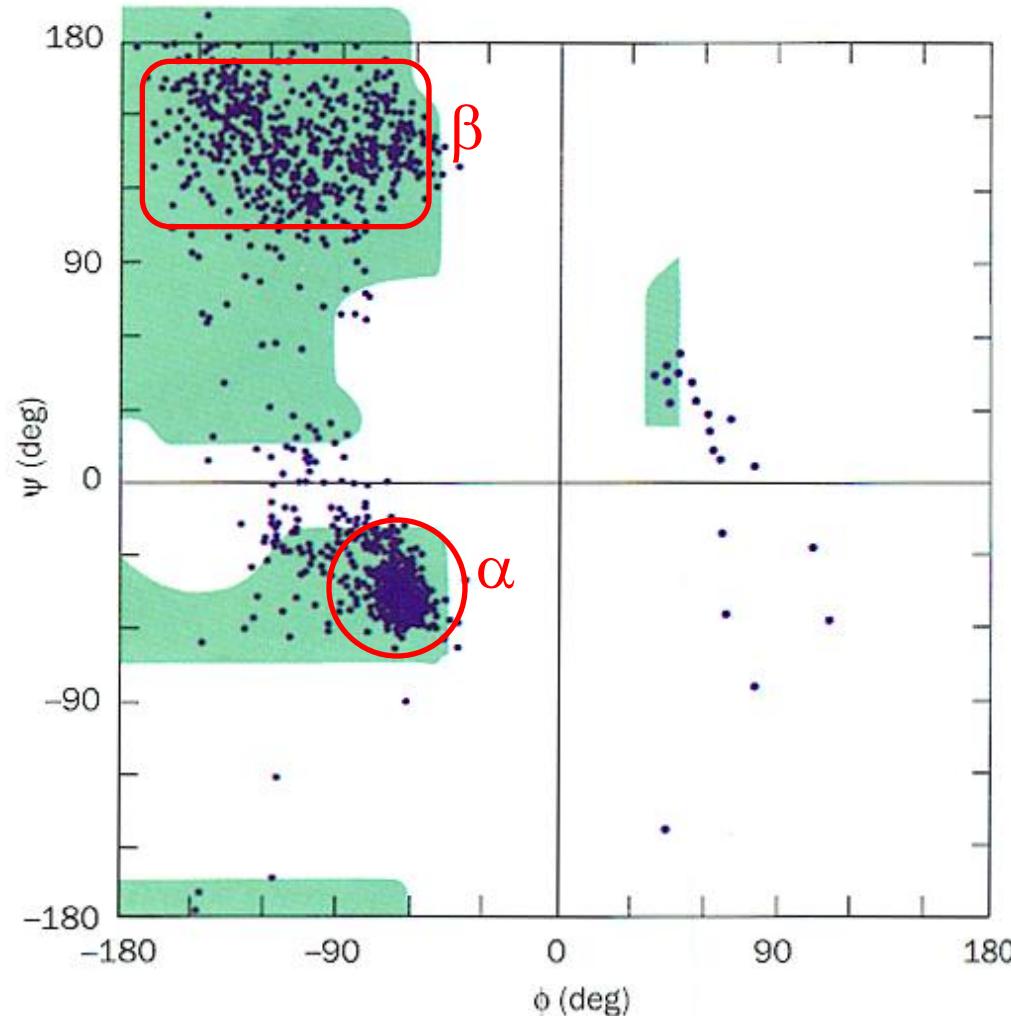
Nature's solution: formation of the so-called "repeating" secondary structures: **helices** and **sheets**.

Secondary Structures

Secondary structure refers to folded segments of a polypeptide chain (with characteristic ϕ and ψ backbone torsion angles) that are usually stabilized by a regular pattern of H-bonds between the peptide NH and CO groups from different residues. There are four types of secondary structures: [1] (α) helices, [2] β sheets, [3] turns, and [4] loops. Helices and sheets are “repeating” secondary structures, whereas turns and loops are “non-repeating” (in terms of the backbone dihedral angles ϕ and ψ).

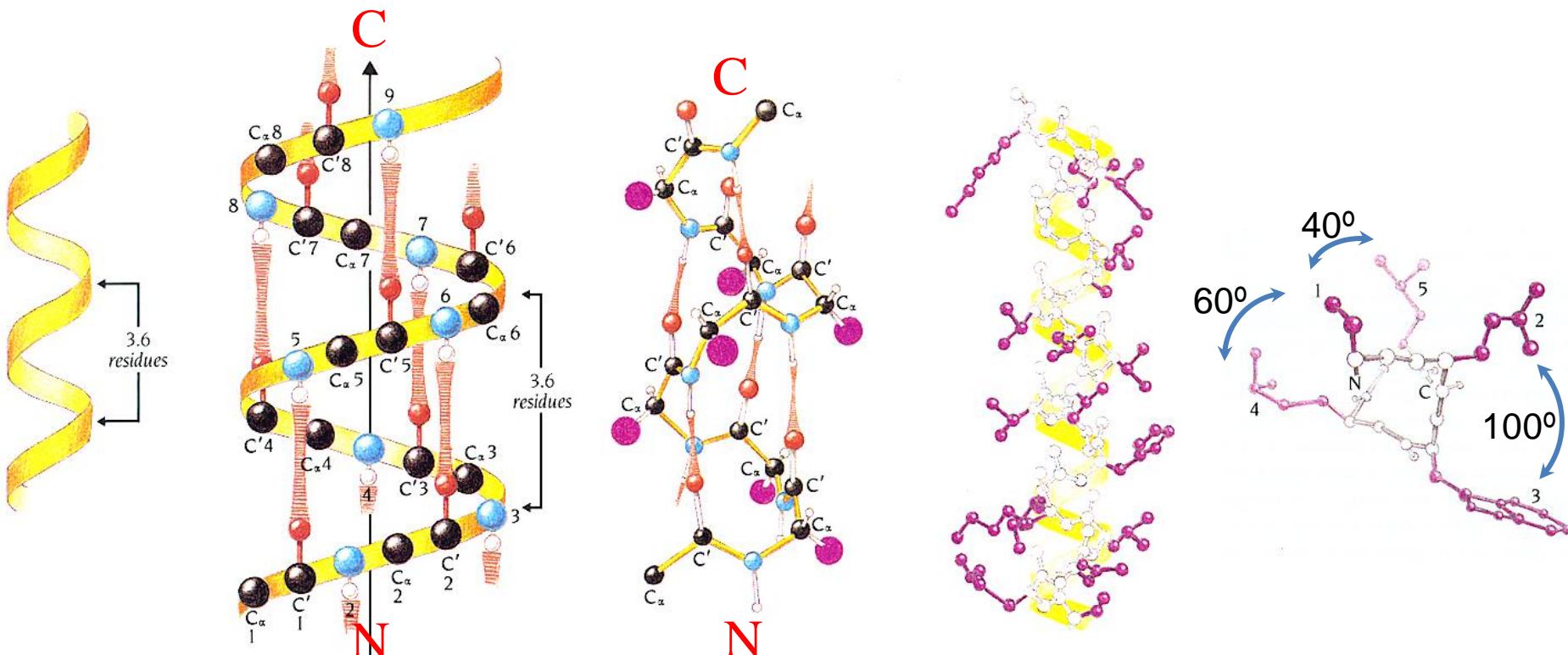


Observed conformational angle distribution of all residues except gly and pro in 12 precisely determined high-resolution X-ray structure. Note that certain regions in the Ramachandran Plot are densely populated with residues.



Secondary Structure: α helix

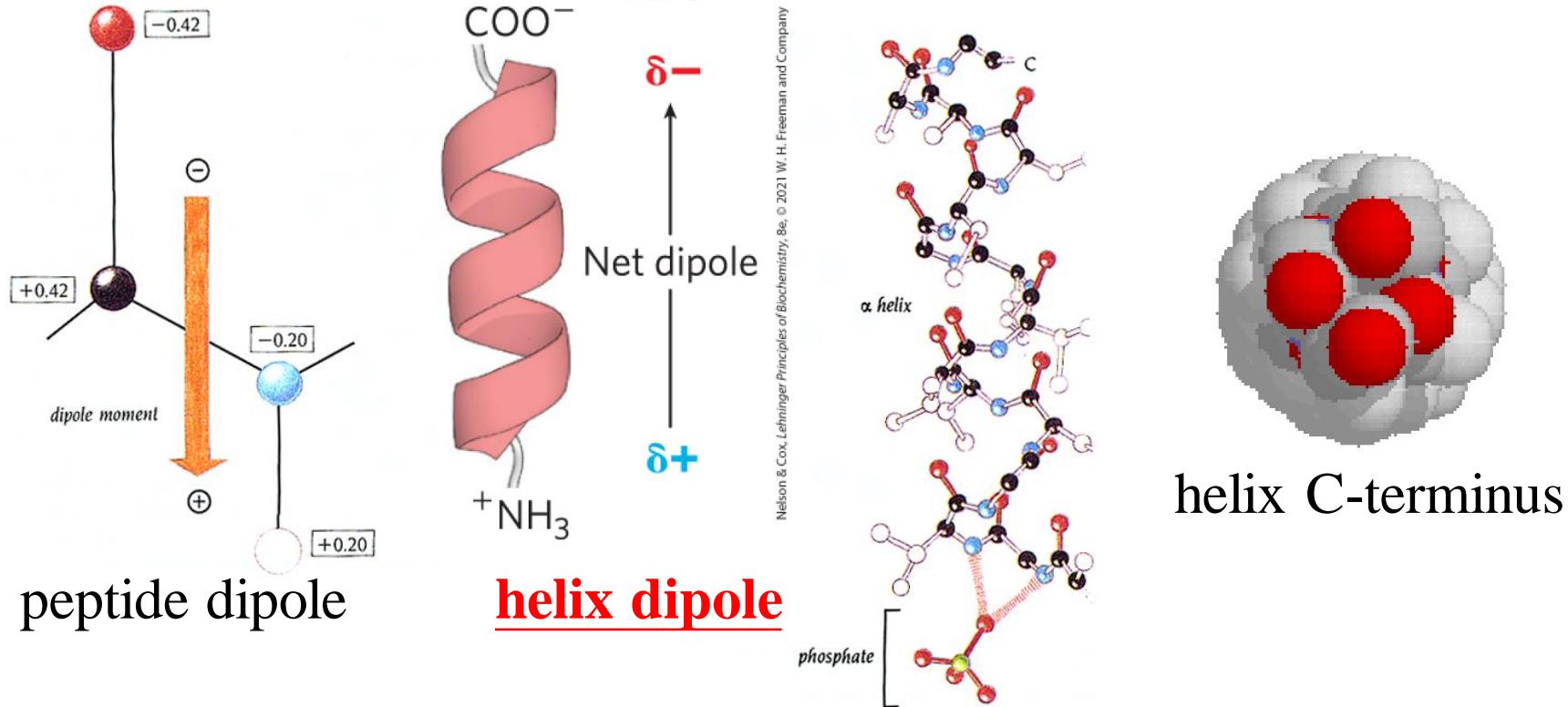
The α helix is right handed with $\phi = -57^\circ$ and $\varphi = -47^\circ$, 3.6 residues per turn, and a pitch of 5.4 \AA (1.5 \AA per residue along the helical axis). In an α helix, the CO of residue i is hydrogen bonded to the NH of residue $i+4$ ($\text{N} \cdots \text{O}$ distance $\sim 2.8 \text{ \AA}$; $\text{N} \cdots \text{H} \cdots \text{O}$ close to 180°). Thus all NH and CO groups are joined with hydrogen bonds except the first 4 NH groups and the last 4 CO groups. The α helix is also called 3.6_{13} helix.



With optimized hydrogen bonds and van der Waals interactions at its core, the α helix is considered fairly stable.

The CO groups point toward the C-terminal end of the α helix.

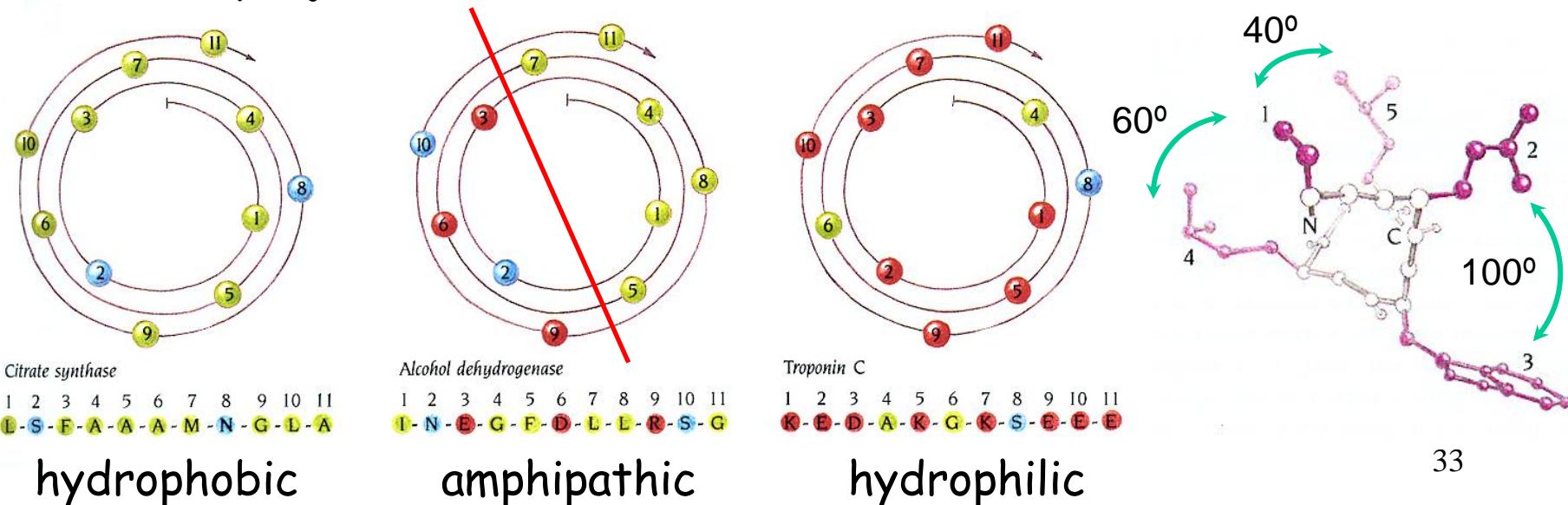
The regular, repeating conformation in the α helix makes all of the charge dipoles of the peptides pointing in the same direction, roughly parallel to the helical axis.



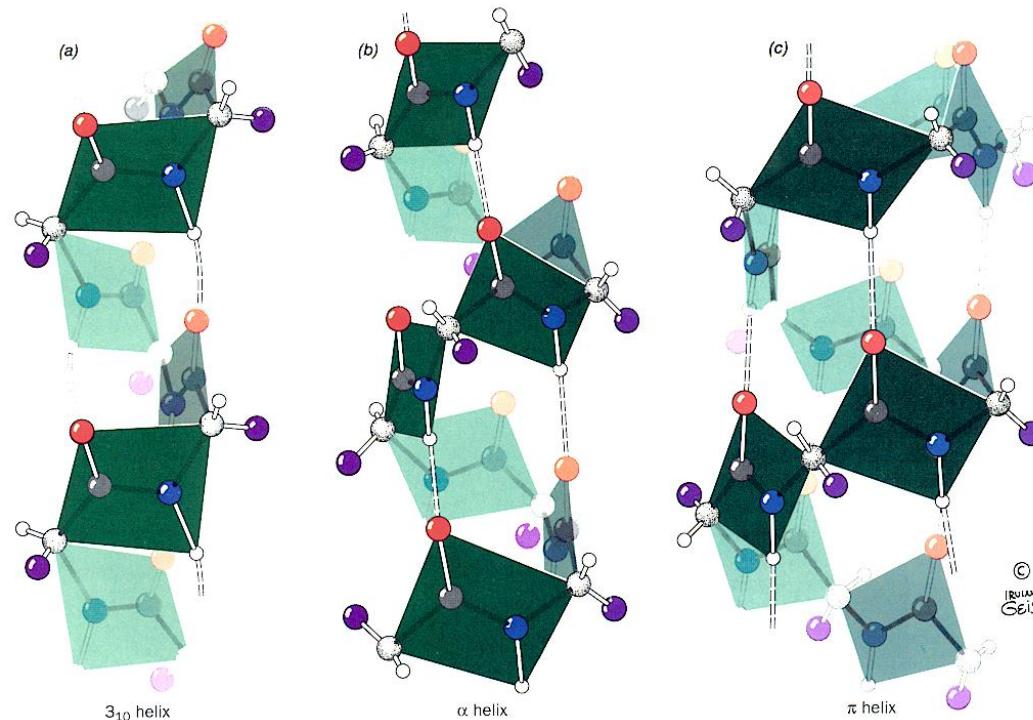
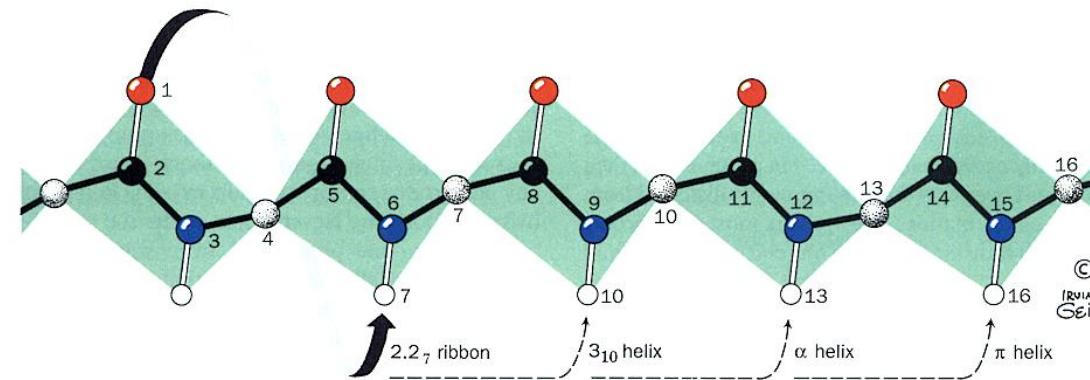
The helix dipole may contribute to the binding of charged species to the protein. (Can a helix end in the interior of a protein??)

With 3.6 residues per turn, side chains protrude from the α helix every 100° . This periodicity means that residues 3~4 amino acids apart in the linear sequence will project from the same face of an α helix. In many α helices, polar and hydrophobic residues are distributed 3~4 residues apart in the sequence to produce an helix with one hydrophobic face and one hydrophilic face; such a helix is known as an **amphipathic α helix**. Helices with this character frequently occur on the surfaces of proteins. This trend can sometimes be seen in the amino acid sequence.

Axial projection of α helix: **helical wheel**



Other helical structures: variations on the α helix are possible. The polypeptide chain can pack tighter to form 3_{10} helix or more loosely to form an π helix.



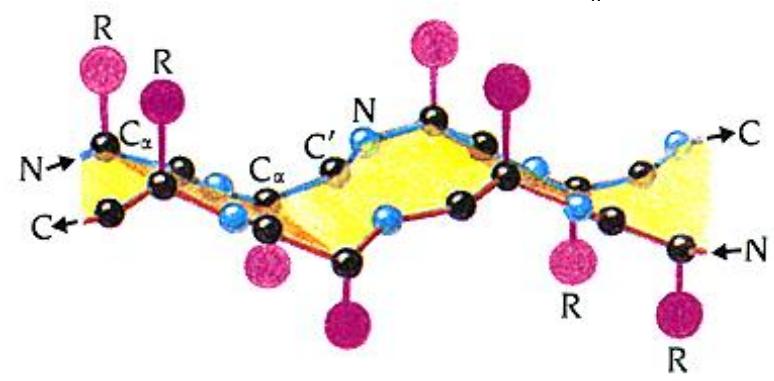
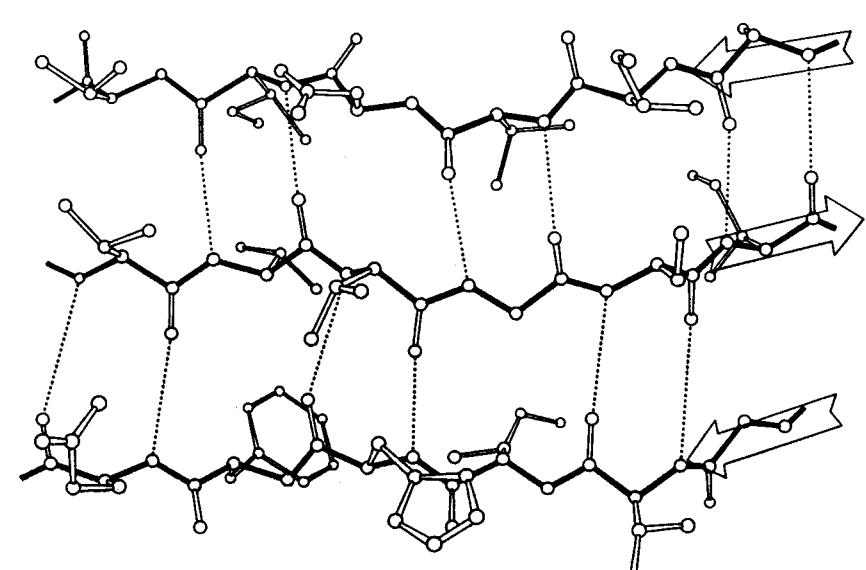
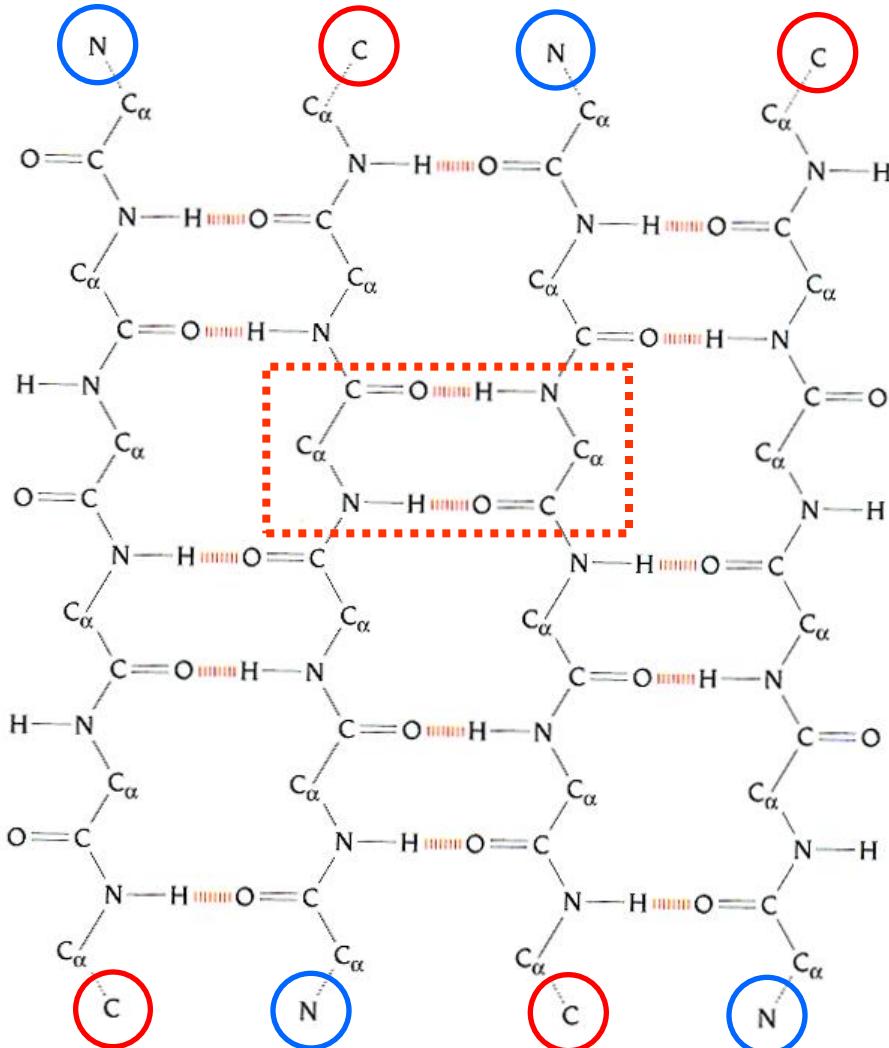
Secondary Structure: β Sheet

β strands are aligned parallel to each other such that hydrogen bonds can form between CO (or NH) groups of one β strand and NH (or CO) groups on an adjacent β strand. All possible backbone hydrogen bonds are formed, except for the two end strands.

β strands can interact in either parallel or antiparallel orientation, and each form has a distinctive hydrogen bonding pattern. Both types of β sheets are pleated, with the successive $C\alpha$ and R groups alternating above and below the plane of the sheet. Side chains on neighboring strands extend to the same side of the sheet and pack against each other. These side chain pairs on neighboring strands show preferences for having hydrophobic groups together or opposite charges together.

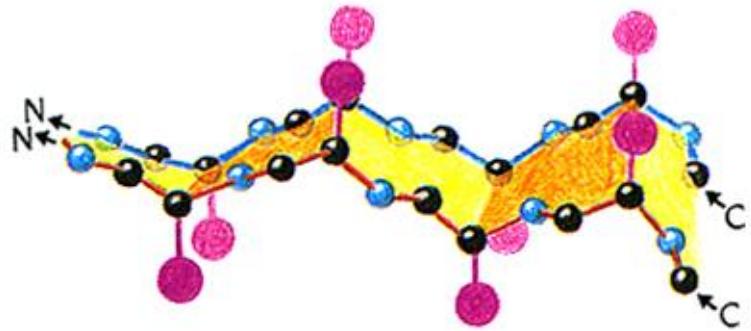
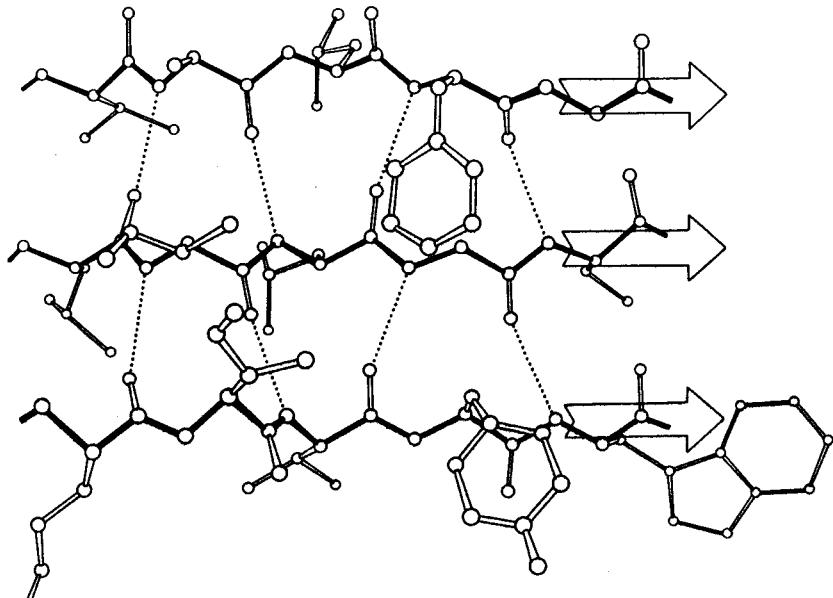
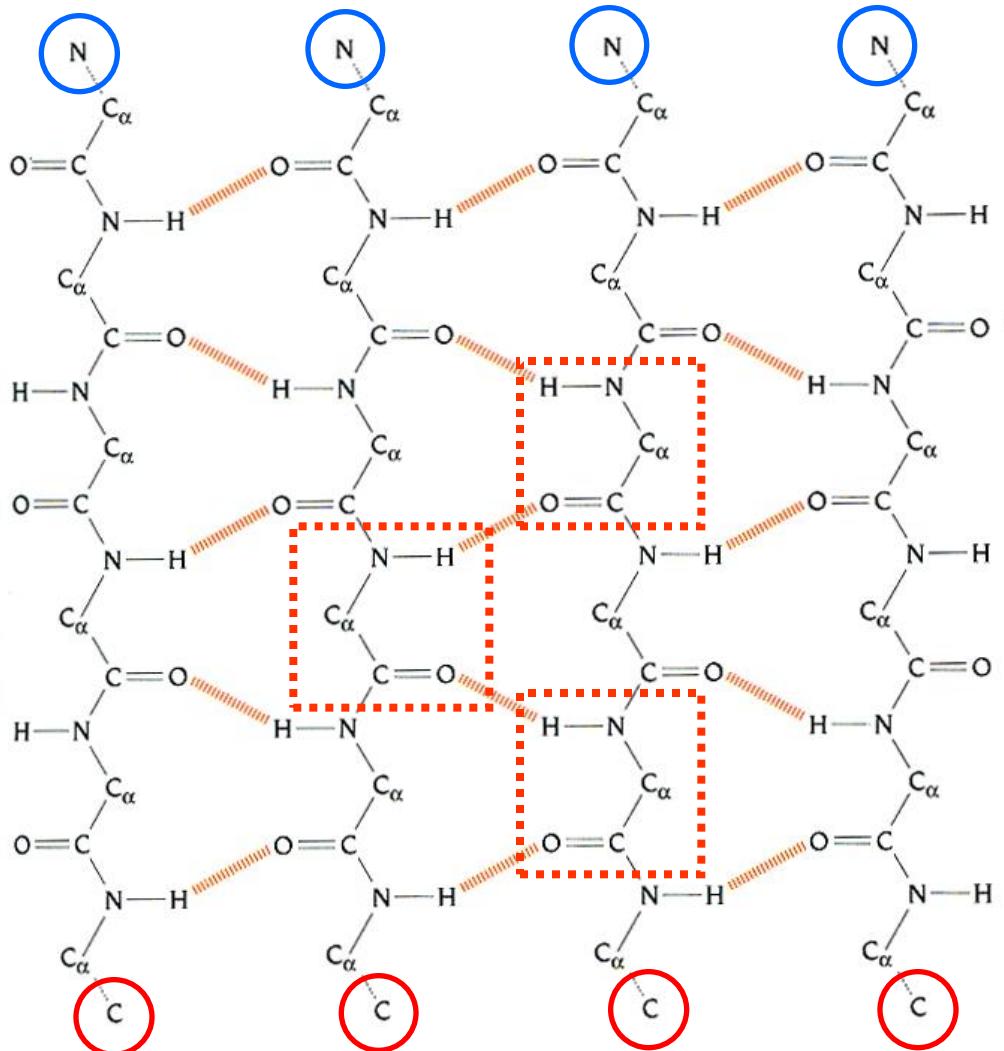
Secondary Structure: β Sheet

Antiparallel β sheet



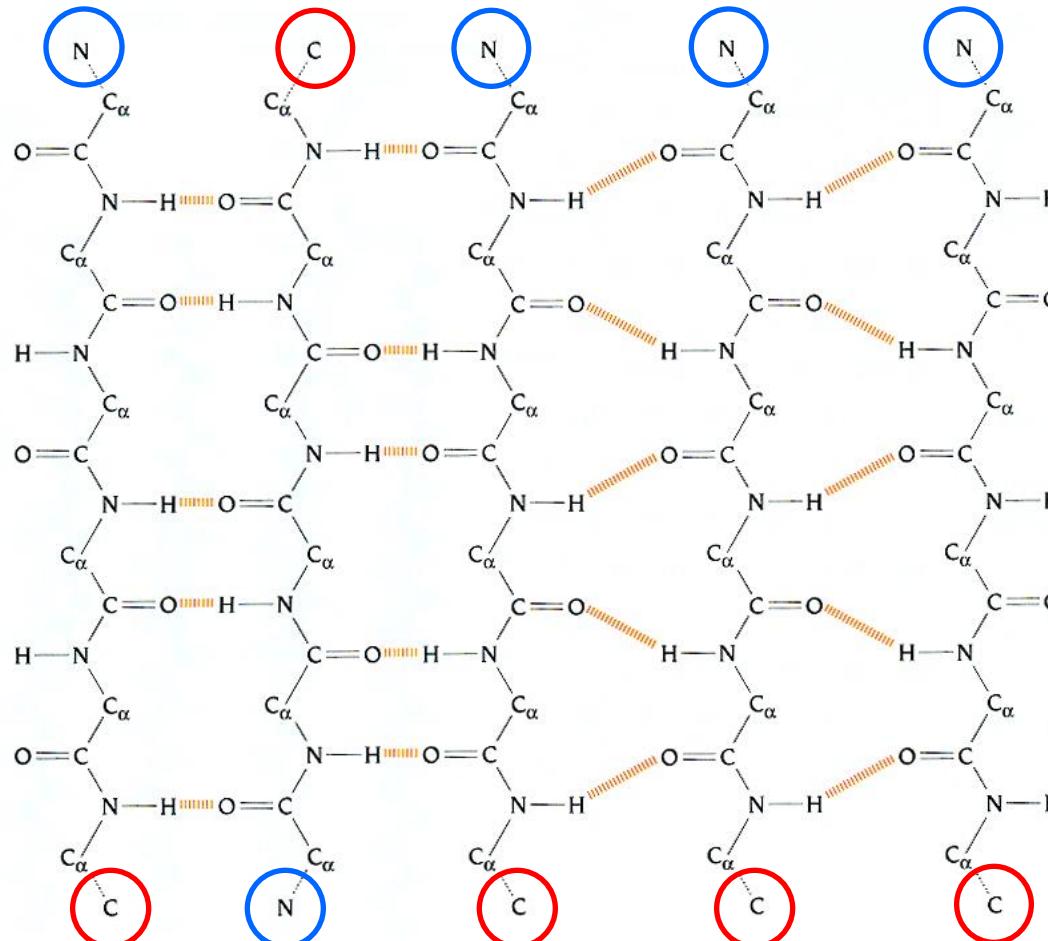
Secondary Structure: β Sheet

Parallel β sheet



Mixed β Sheet

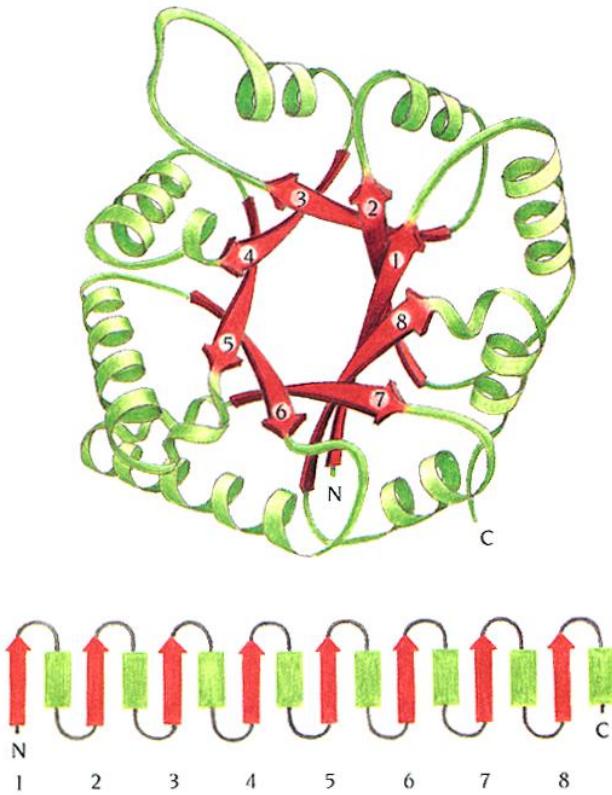
β strands can form either a pure parallel sheet, a pure antiparallel sheet, or a mixed sheet with some strand pairs parallel and some antiparallel.



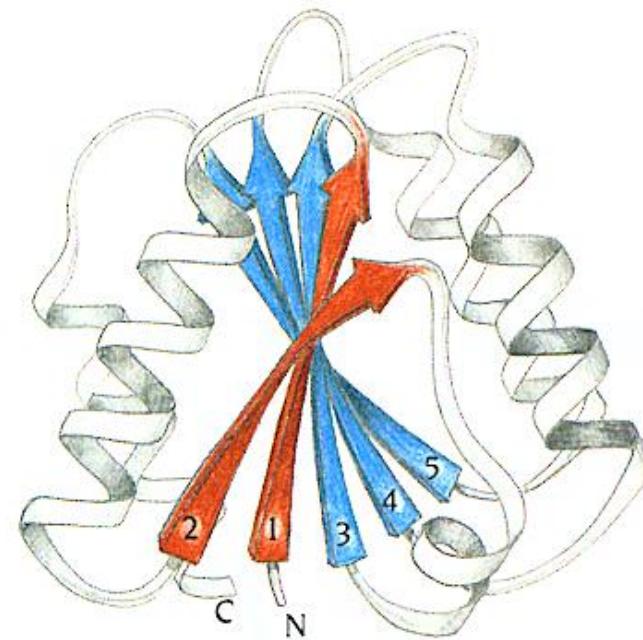
In general, antiparallel β -sheets are more stable than parallel ones.

Parallel β sheets are usually protected by α helices

(TIM barrel)



Rossmann fold



Secondary Structure: β Turn

The shape of globular proteins depends on the presence of sharp bends in the direction of the polypeptide chain. This is often accomplished by the formation of a structural element called a β -turn. **The β -turns almost always occur at protein surfaces.**

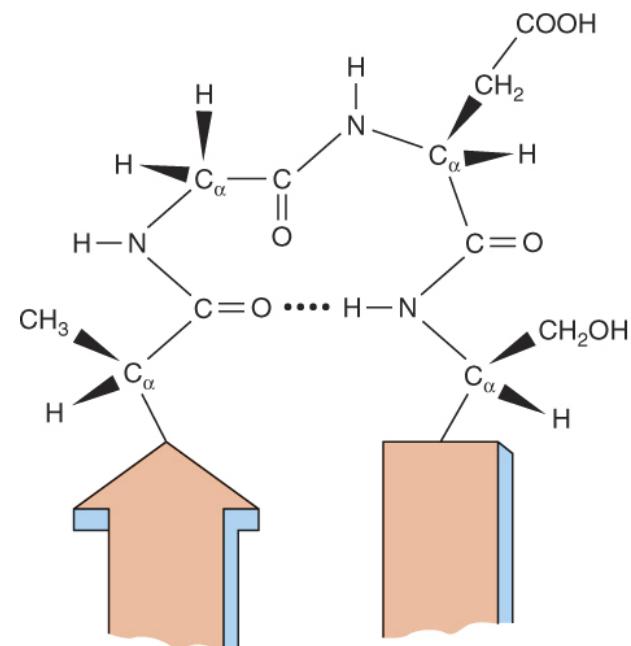
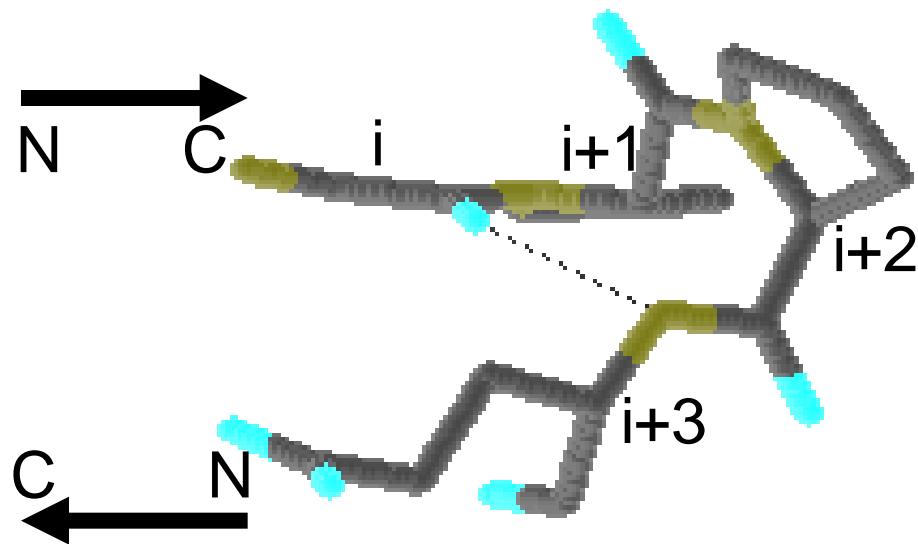
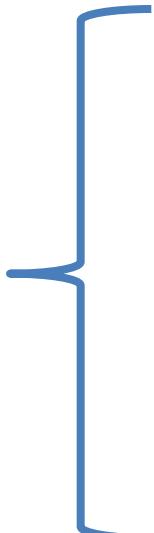


TABLE 4-1 Idealized ϕ and ψ Angles for Common Secondary Structures in Proteins

Structure	ϕ	ψ
α Helix	-57°	-47°
β Conformation		
Antiparallel	-139°	$+135^\circ$
Parallel	-119°	$+113^\circ$
Collagen triple helix	-51°	$+153^\circ$
β Turn type I		
$i + 1^a$	-60°	-30°
$i + 2^a$	-90°	0°
β Turn type II		
$i + 1$	-60°	$+120^\circ$
$i + 2$	$+80^\circ$	0°

Note that the ϕ and ψ angles of residues i and $i+3$ are irrelevant in β turns, meaning that....



Note: In real proteins, dihedral angles often vary somewhat from these idealized values.

^aThe $i + 1$ and $i + 2$ angles are those for the second and third amino acid residues in the β turn, respectively.

Secondary Structure: β Turn

Type I β -turn and type II β -turn differ by a 180° flip of the peptide unit linking residues 2 and 3.

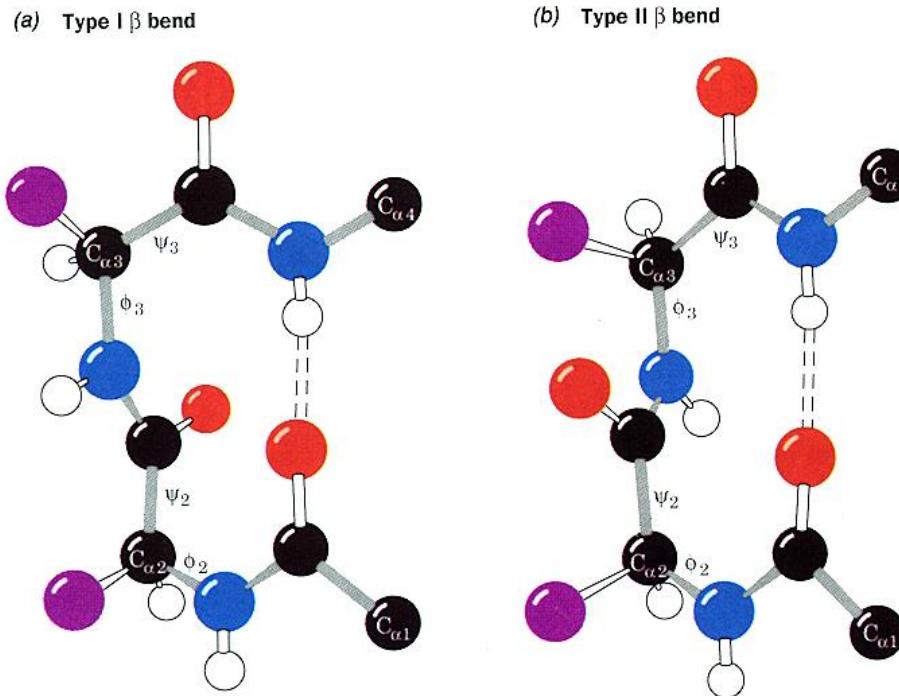


FIGURE 7-22. Reverse turns in polypeptide chains: (a) A Type I β bend, which has the following torsion angles:

$$\phi_2 = -60^\circ, \quad \psi_2 = -30^\circ, \\ \phi_3 = -90^\circ, \quad \psi_3 = 0^\circ.$$

(b) A Type II β bend, which has the following torsion angles:

$$\phi_2 = -60^\circ, \quad \psi_2 = 120^\circ, \\ \phi_3 = 90^\circ, \quad \psi_3 = 0^\circ.$$

Variations from these ideal conformation angles by as much as 30° are common. Hydrogen bonds are represented by dashed lines. [Figure copyrighted © by Irving Geis.]

Both types of β -turns are stabilized by hydrogen bonds. Residue 2 of both types of β -turns is often Pro since it can easily assume the required conformation. In type II turns, the CO of residue 2 bumps into the R group of residue 3, so residue 3 is usually Gly.

Amino acid propensities for forming secondary structures

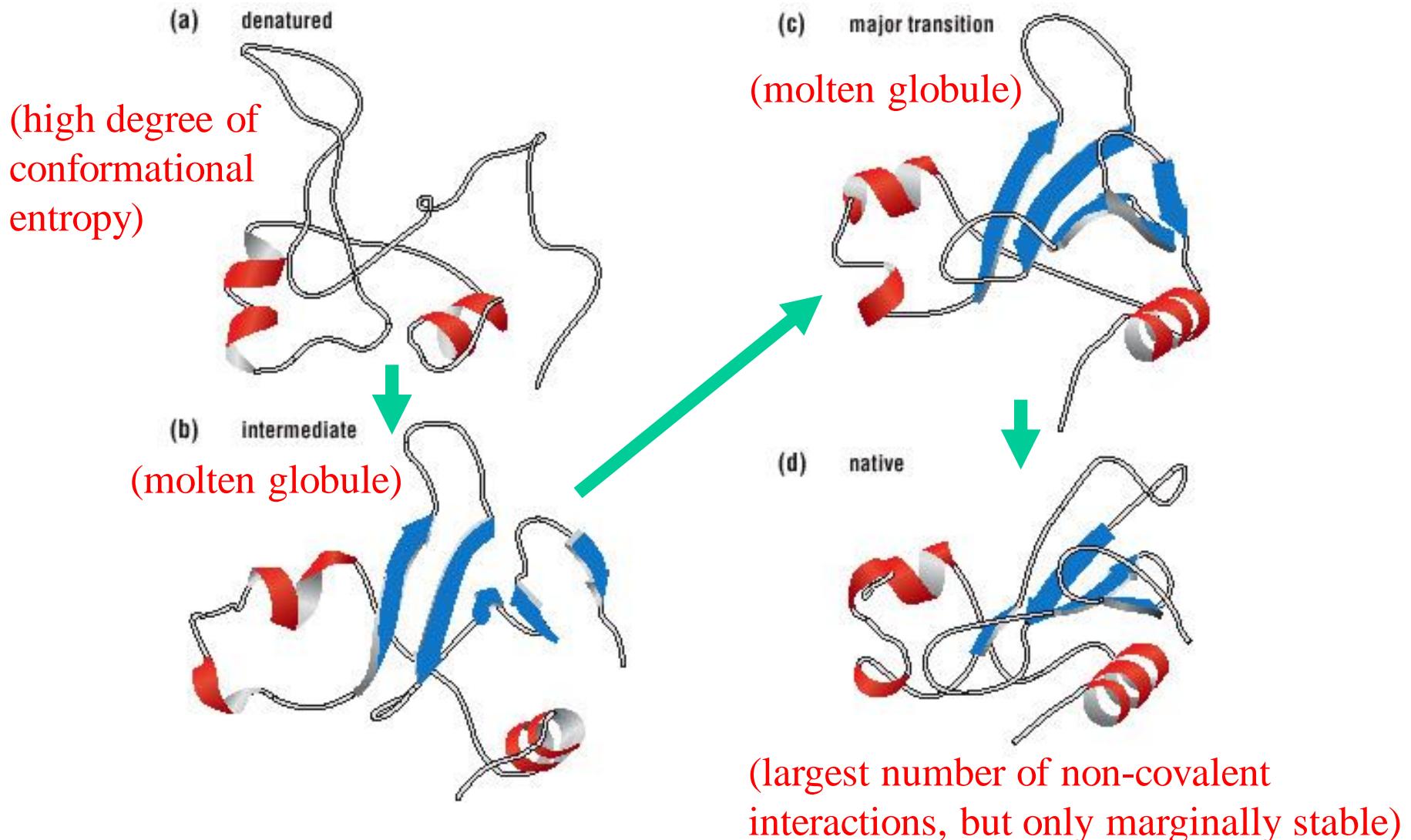
High α helix propensity: **Ala, Leu, Met, Phe, Glu, Gln, His, Lys, Arg.**
OK in α helix: **Met, Phe, Tyr.**

More likely to be observed in β sheet and less favored in α helix: (1) just large: **Trp**, (2) bulkier due to branched C^β : **Ile, Val, Thr**, (3) large S atom on C^β : **Cys**.

The side chains of some amino acids may disrupt repetitive secondary structures (known as secondary structure breakers): (1) **Gly**: too flexible to be constrained in the highly ordered secondary structure, (2) **Pro**: side chain linked to alpha N, has no N-H to participate in H-bond; more commonly seen in the first turn of an α helix; (3) **Asp, Asn, Ser**: side chains compete directly with backbone H-bonds.

In addition to the individual propensity, consecutive sequences of residues of like charge (eg. Glu-Glu; Arg-Arg) or bulky residues (eg. Ile-Trp) tend to destabilize an α helix.

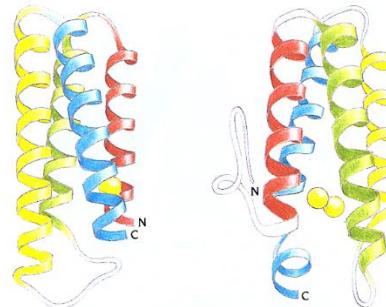
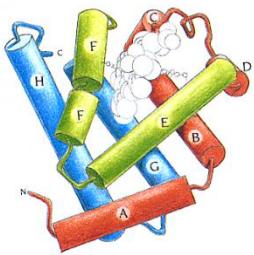
Illustration of a possible pathway for protein folding



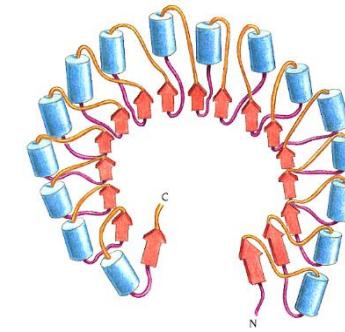
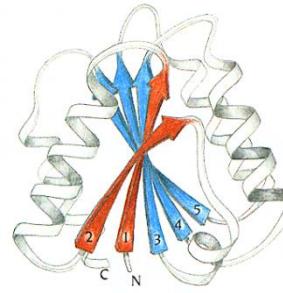
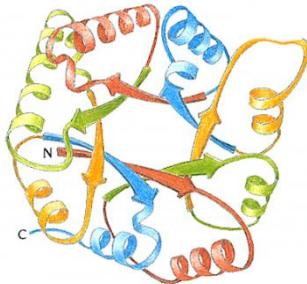
The native conformations of proteins are stabilized mainly by non-covalent interactions, so they are constantly undergoing local unfolding and in equilibrium with the molten globule states.

Structural Classification of Proteins (SCOP) database: 4 classes of protein **tertiary structures**

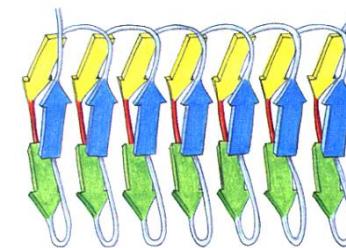
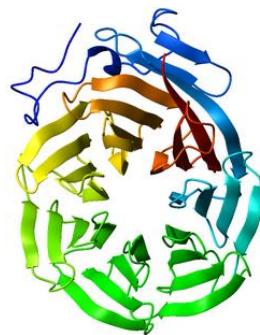
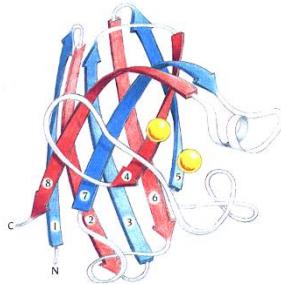
All α



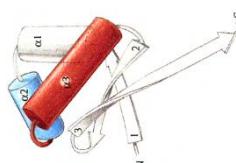
α/β



All β



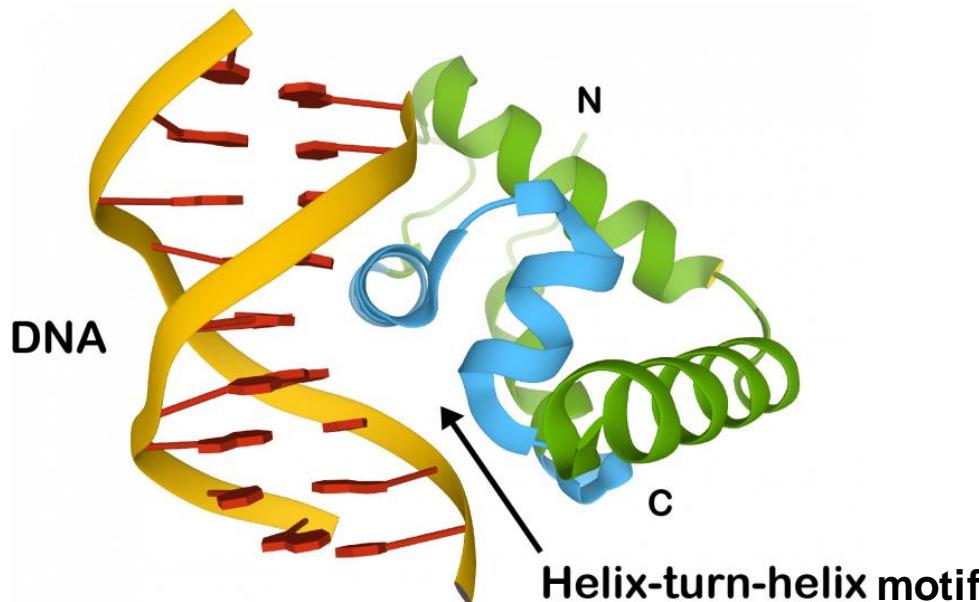
$\alpha+\beta$



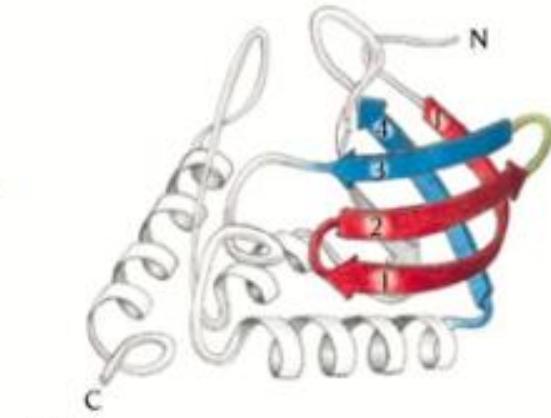
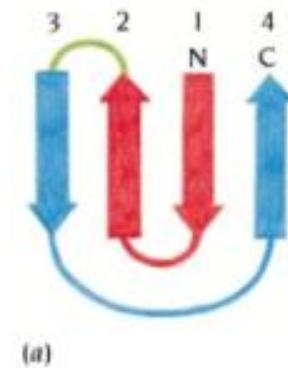
Domains, motifs, supersecondary structures

Domains: an independent regions of a protein that **can fold and function on their own**, often corresponding to a particular function or structural feature within the protein.

Motifs/supersecondary structures: small, recurring structural patterns that can be found in many proteins. Unlike domains, **motifs and supersecondary structures can NOT fold into a stable tertiary structure.**

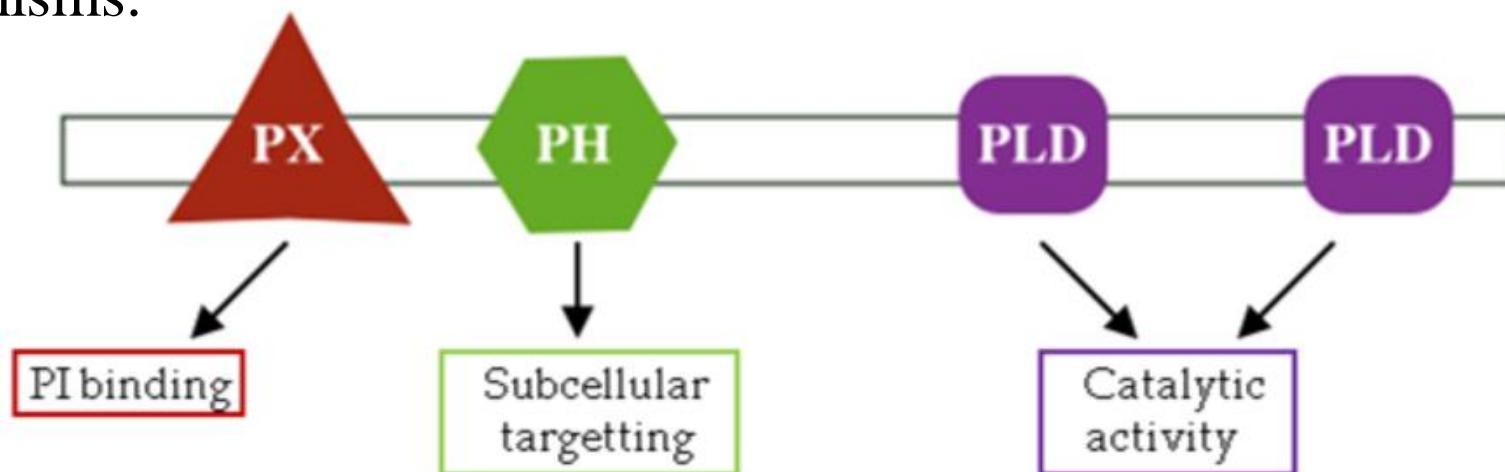


Greek Key Motif



Many eukaryotic proteins are multi-domain proteins

The importance of this observation lies in the versatility and complexity that multi-domain proteins offer. Having multiple domains allows proteins to perform a variety of functions within a single molecule, such as binding to different molecules, catalyzing reactions, or undergoing structural changes. This modularity enables the evolution of new functions by combining existing domains in novel ways, which is crucial for the diversity and adaptability of eukaryotic organisms.



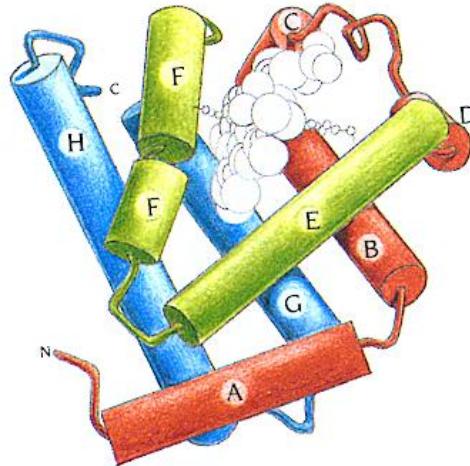
Domain composition of phospholipase D1, an enzyme that breaks down phosphatidylcholine.

Packing of α Helices

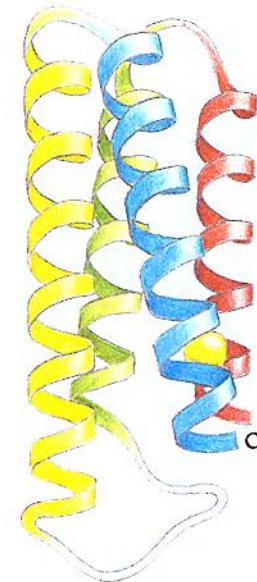
In solution, an isolated α helix sometimes is not very stable, however, α helices can be greatly stabilized by packing together to form a three-dimensional structures of proteins.

All- α proteins:

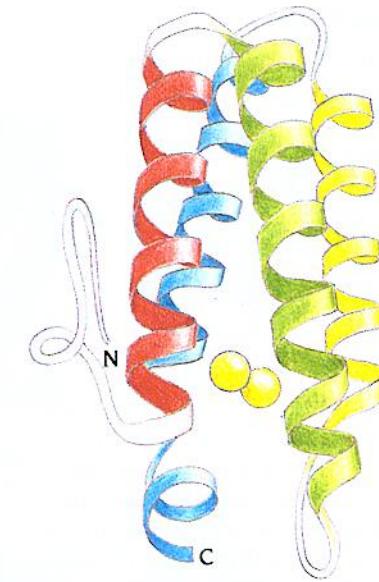
see Structural Classification of Proteins (SCOP) database.



globins



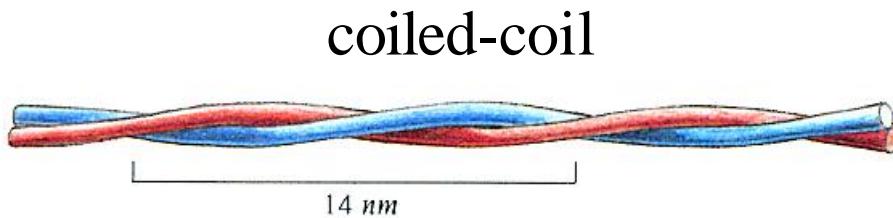
helix bundles



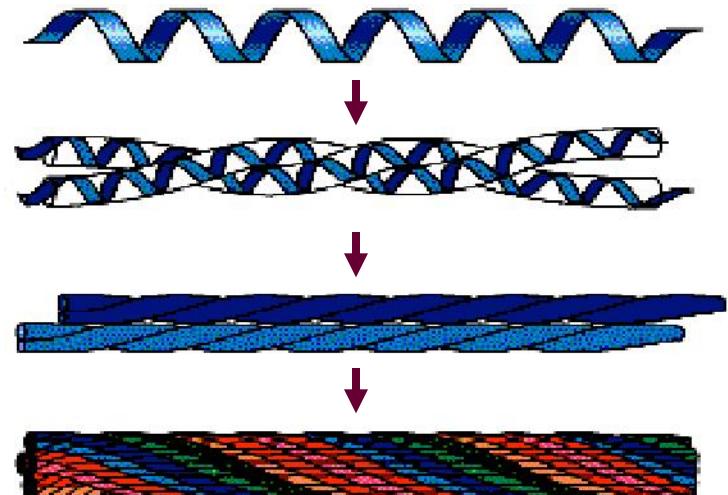
Two simple rules governing the association of α helices: (1) maximized VDW packing for hydrophobic side chains; (2) upon packing, helices retain a conformation close to the minimum free energy conformation of the isolated helix (no serious distortion to the helix).

Packing of α Helices

Coiled-coil arrangement of two α helices: a natural structural module for dimerization. To maximize the interactions between hydrophobic side chains, the two α helices are not straight rods but are wound around each other to form a left-handed supercoil.



Assembly of α -keratin filament

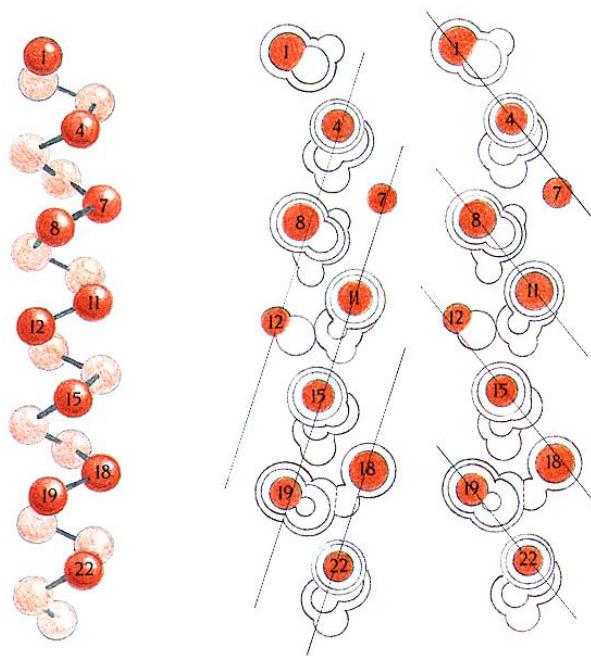
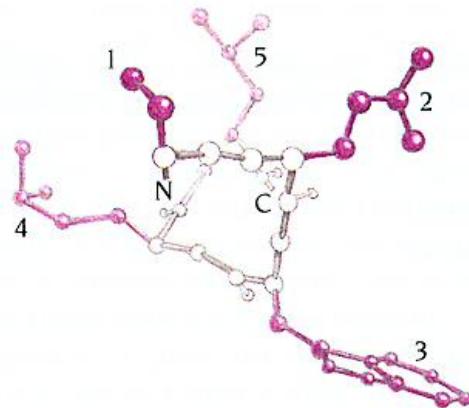


Coiled-coils are the basis of some of the fibrous proteins such as **α -keratin** (hair, wool, nails, claws, rhinoceros horns, ...).

α -Keratin is an structural protein that provides support, shape, and external protection to vertebrates. Shorter coiled-coils are seen in lots of proteins (e.g., transcription factors) to promote dimerization.

Ridge-in-Groove model of α helix packing

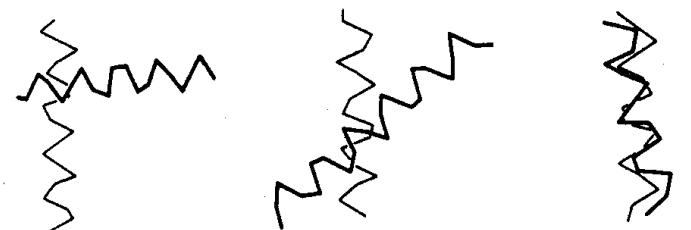
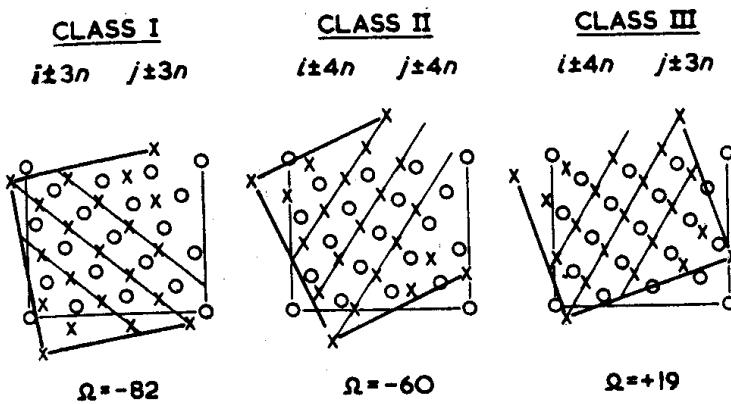
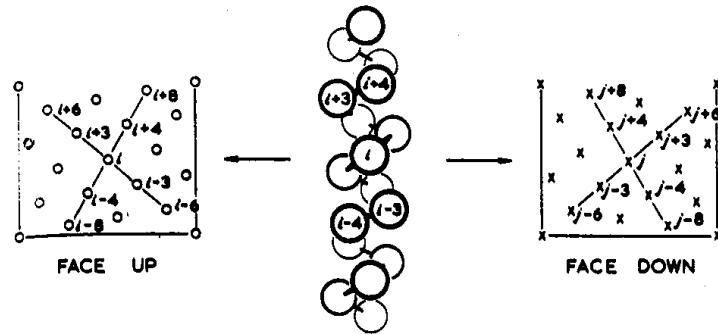
The surface of an α helix can be described in terms of rows of adjacent side chains. For residue i in a helix, it has two neighbors ($i+3$ and $i+4$) above it and two neighbors ($i-3$ and $i-4$) below it. Due to the regularity of the α helix, a ridge is formed by residues i , $i\pm 3$, $i\pm 6$, ... $i\pm 3n$ ($i\pm 3n$ ridge) and another ridge by i , $i\pm 4$, $i\pm 8$, ... $i\pm 4n$ ($i\pm 4n$ ridge).



Ridge-in-Groove model of α helix packing

Two types of grooves can also be recognized on the surface of the α helix: ($i \pm 3n$ groove) and ($i \pm 4n$ groove).

The helix-helix packing can be viewed as the surface ridges in the first helix pack into the grooves in the second helix and vice versa.

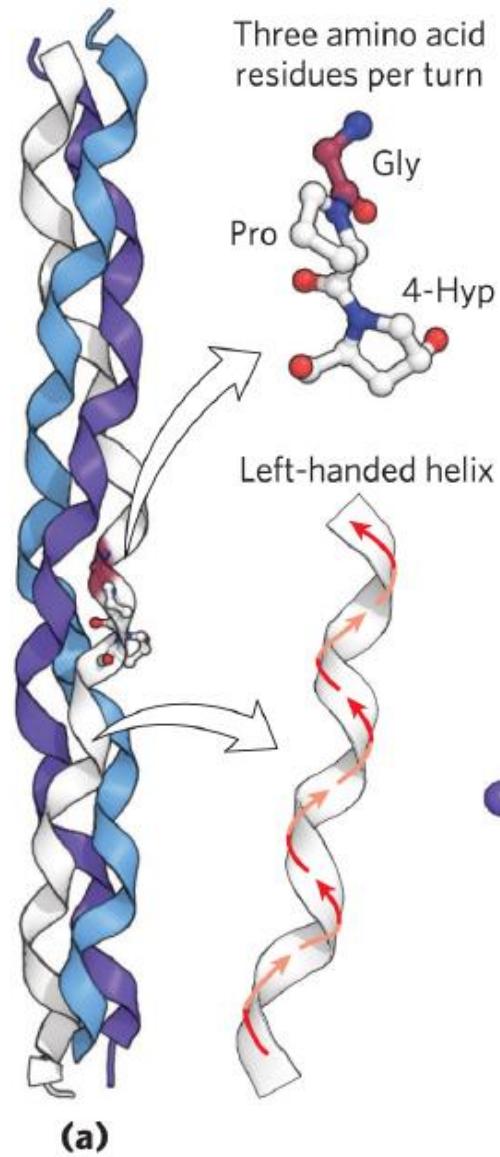


Collagen - 1

Collagen has evolved to provide strength. It is found in connective tissue such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye.

Collagen is the most abundant protein in mammals (25% to 35% of total protein content).

The collagen helix is a unique type of secondary structure: left-handed and 3 residues per helical turn. Three separate polypeptides twisted about each other to form the three-stranded collagen superhelix.



Repeating tripeptide unit of collagen:
Gly-X-Y, where X is often Pro and Y is often 4-Hyp (4-hydroxyproline).

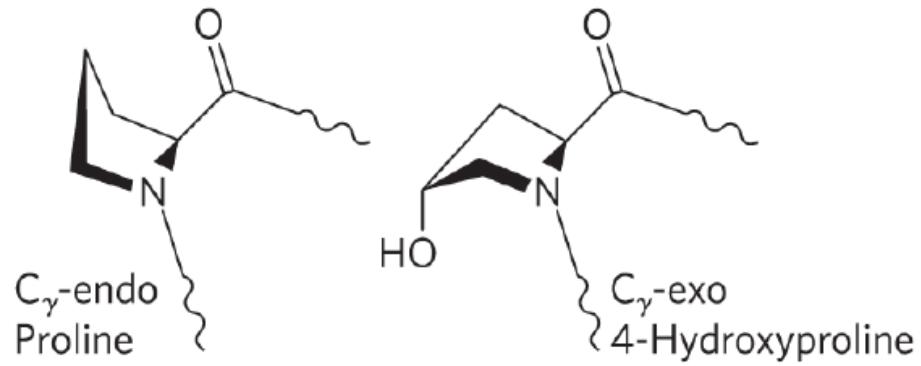
Collagen - 2

Why Gly-X-Y:

Only Gly can fit within the tight junctions of the collagen triple helix. Mutations that replace Gly with larger amino acids disrupt this precise packing, destabilizing the helix. This disruption underlies several collagen-related diseases, including osteogenesis imperfecta and Ehlers-Danlos syndrome, where collagen's structural integrity is compromised.

The Pro and 4-Hyp residues permit the sharp twisting of the collagen helix. structure becomes unstable.

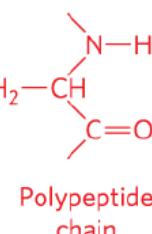
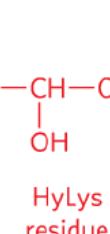
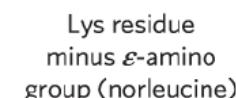
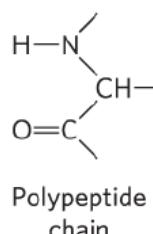
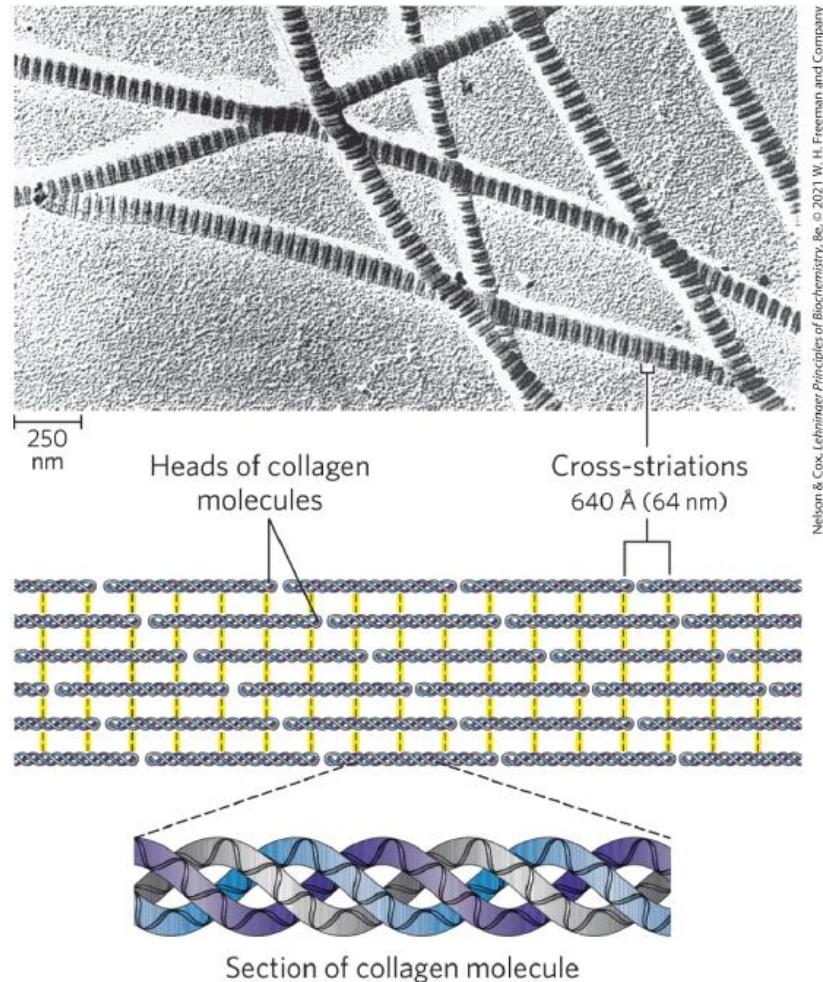
Moreover, the correct collagen structure requires the Pro/4-Hyp residues in the Y positions to be in the C_y-exo conformation. Otherwise, the collagen structure becomes unstable.



Collagen - 3

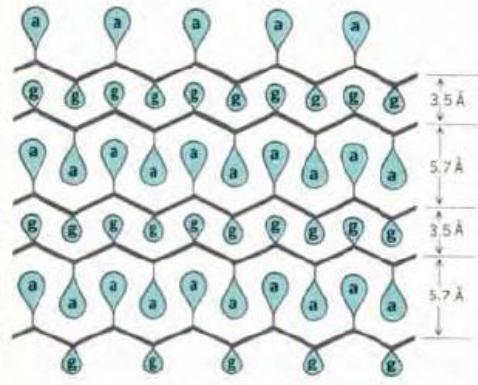
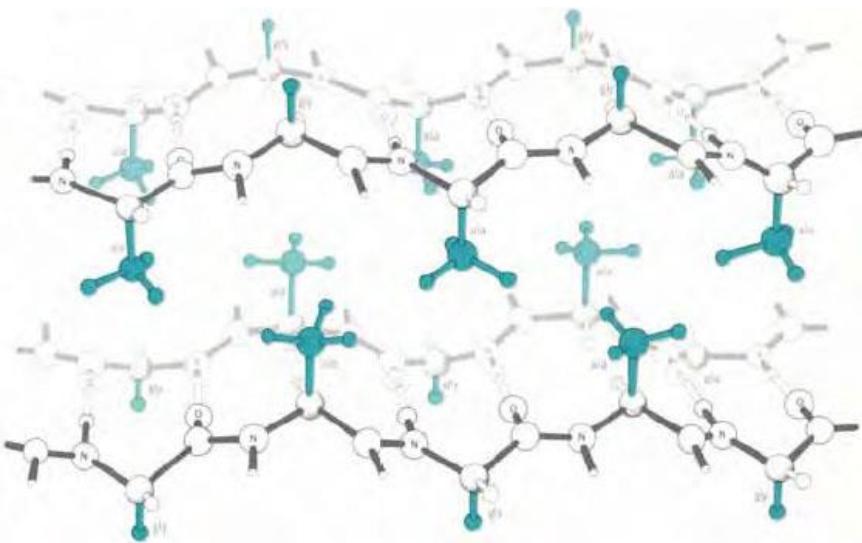
Each collagen triple helix is called a tropocollagen. These tropocollagen molecules can be cross-linked in a variety of ways to form collagen fibrils with different degrees of tensile strength. Cross-linking of tropocollagen molecules are achieved by unusual types of covalent bonds involving Lys, HyLys (5-hydroxylysine), or His residues that are present at a few of the X and Y positions.

The increasing rigid and brittle character of aging connective tissue results from accumulated covalent cross-links in collagen fibrils.



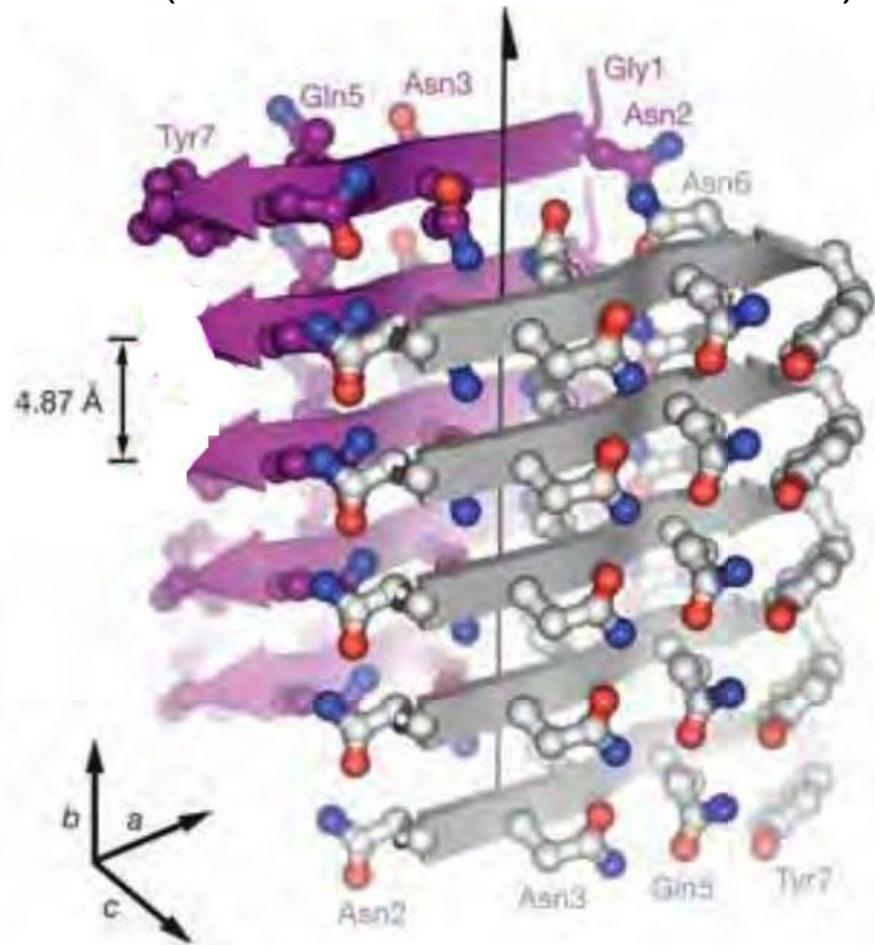
Some β sheet structures are extremely stable

Silk



Amyloid fibril

(mad cow disease, Alzheimer....)



Physical factors that affect protein stability may induce protein unfolding (denaturation) or misfolding

Temperature: thermal denaturation (heat breaks non-covalent interactions); cold denaturation (water molecules become more structured, increased structuring of water reduces the entropic gain that typically drives hydrophobic interactions, making it less energetically favorable for hydrophobic residues to cluster together)

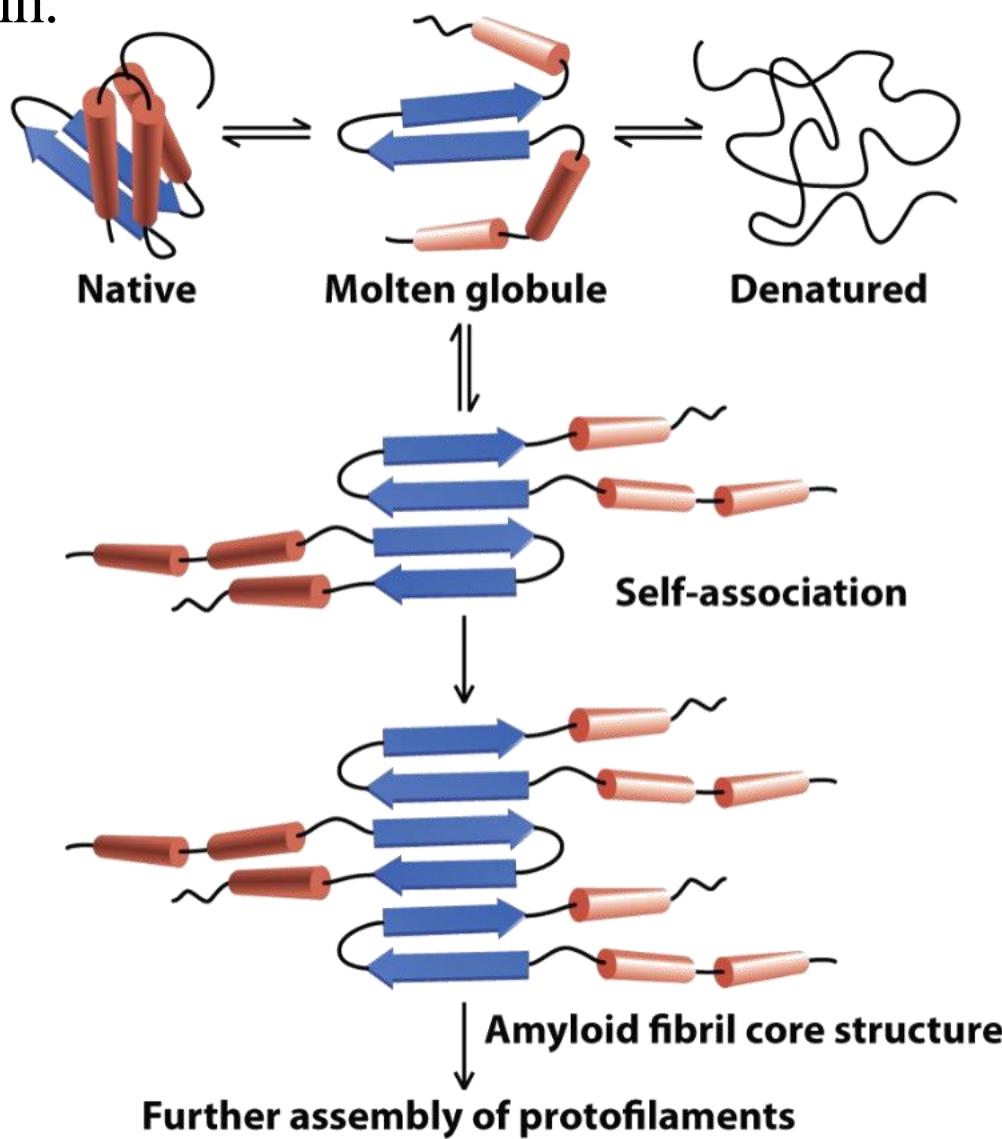
pH: charges on acidic or basic side chains can change, disrupting hydrogen bonds, ionic bonds, and other interactions that maintain the protein's three-dimensional structure.

Redox state: oxidation of intracellular proteins; reduction of extracellular proteins.

Additional factors: ionic strength; pressure, solvent environment; presence of other proteins...

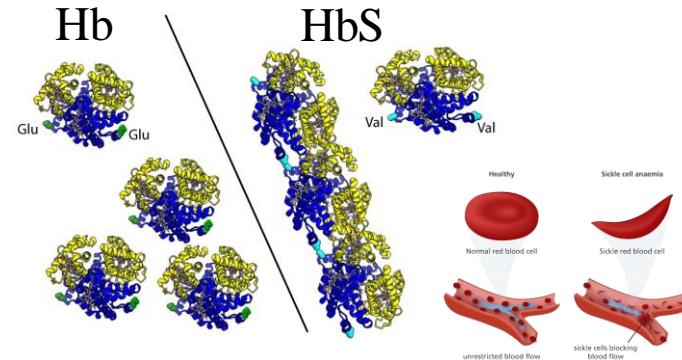
Protein unfolding/misfolding may have deleterious consequences

A possible mechanism for the formation of **β amyloid fibrils** by a globular protein.

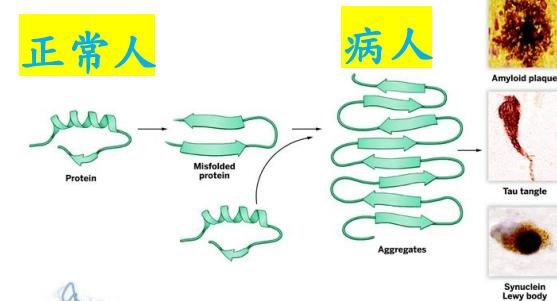


Diseases related to aberrant protein structure

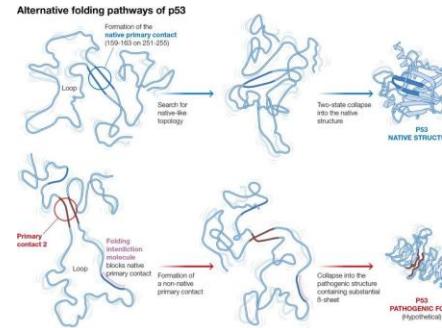
Sickle-cell anemia: mutant hemoglobin



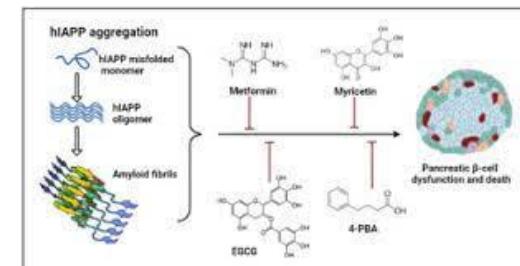
Neurodegenerative diseases: Alzheimer's disease, Parkinson's disease, Huntington's disease, the transmissible spongiform encephalopathies.....



Cancers: p53 (a tumor suppressor)
misfolding.....



Type II diabetes: accumulation of misfolded aggregates composed of the islet amyloid polypeptide (IAPP)

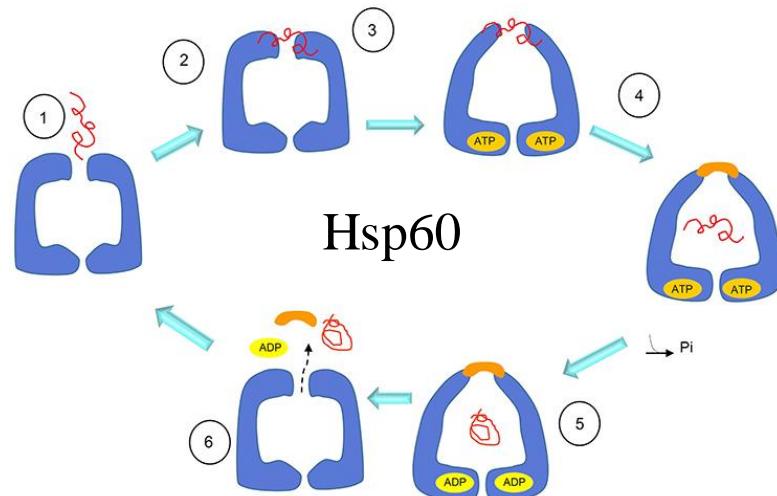
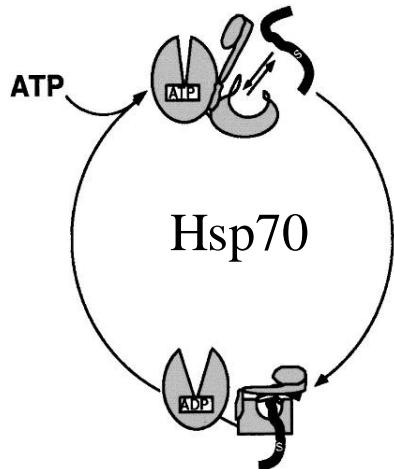


Cells may employ auxiliary proteins to assist protein folding

Chaperones:

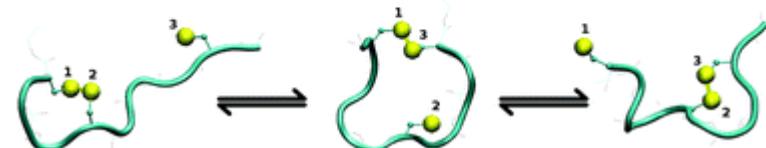
Heat shock protein 70 (Hsp70): binds short sequences of hydrophobic amino acids.

Heat shock protein 60 (Hsp60): provides an sheltered environment to assist folding.



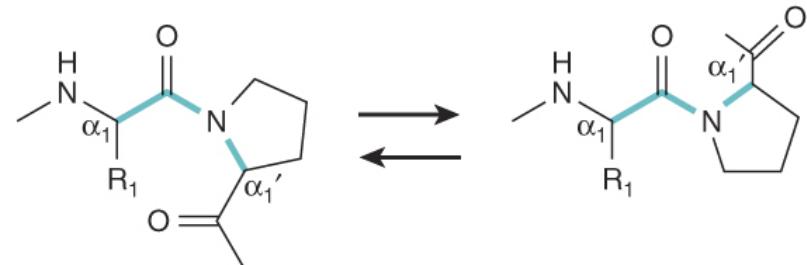
Protein disulfide isomerases:

Facilitates the formation of disulfide bonds that stabilize native conformation.



Proline-cis, trans-isomerase (cyclophilins):

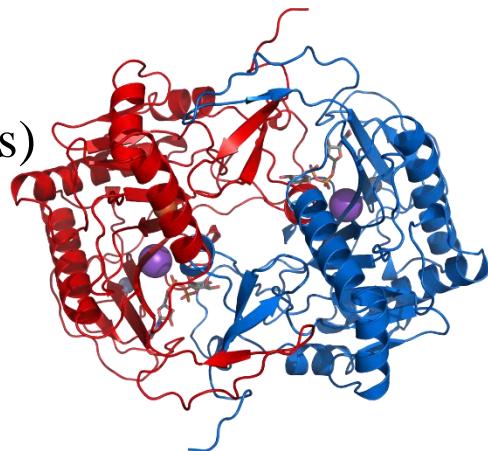
Catalyzes the cis-trans isomerization of the X-Pro peptide bond.



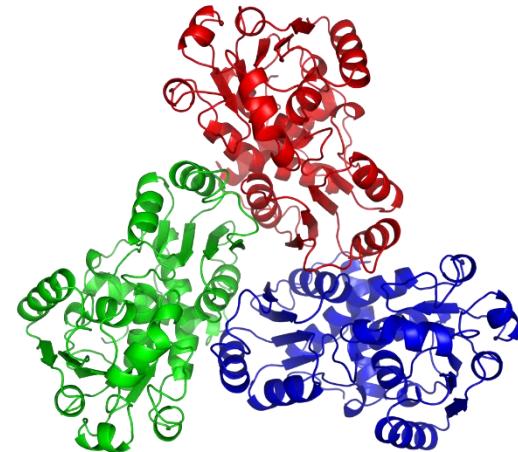
Examples of protein quaternary structures (protein **oligomers/complexes**)

dimer

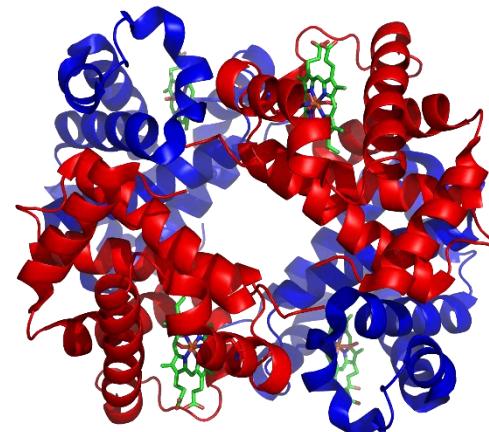
(two monomers)



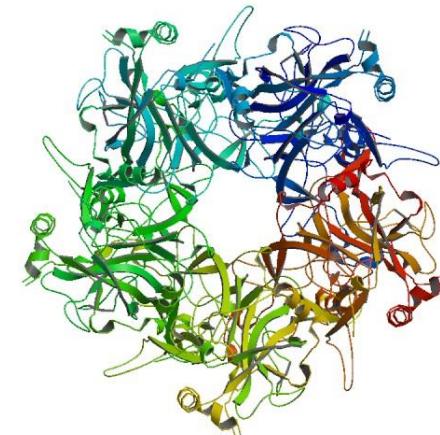
trimer



tetramer

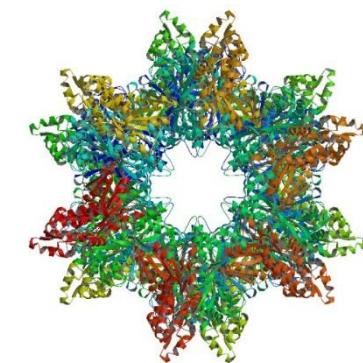


others...



Homooligomers are complexes formed by multiple identical protein subunits (monomers, protomers).

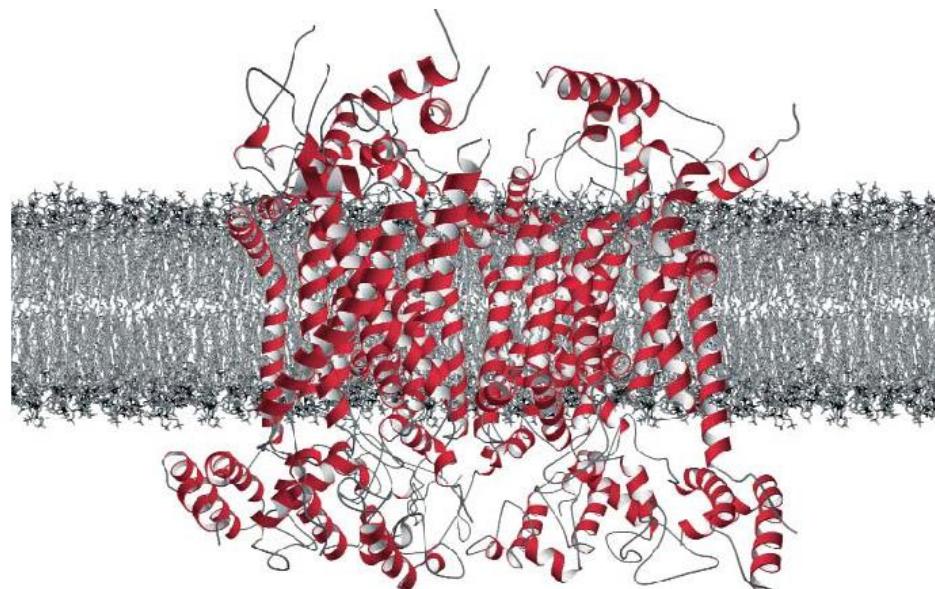
Heterooligomers are complexes made up of different protein subunits.



Transmembrane helices

Some proteins are embedded in the hydrophobic interior of the membranes that form cell surfaces, organelles, and vesicles.

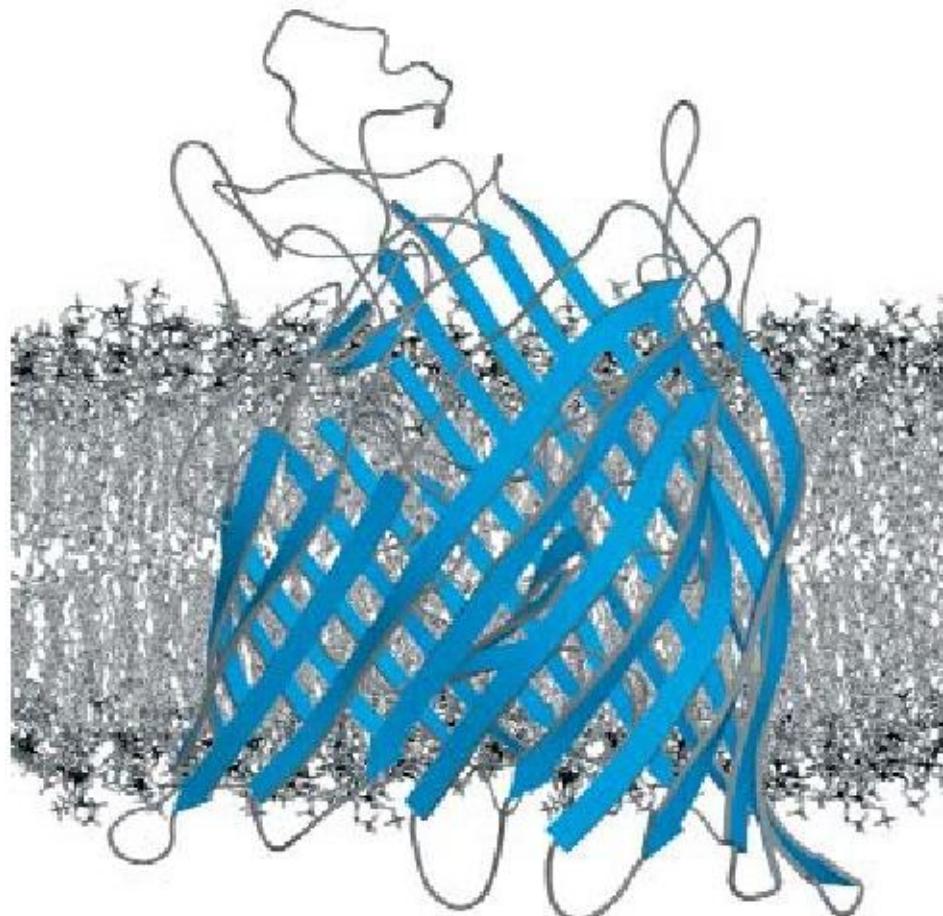
The non-polar interior of the membrane is approximately 30 Å across; the head-group layers contribute an additional 5~10 Å on each side of the membrane. α helices are common in the transmembrane portions of membrane-bound proteins, it would take an α helix about 20-residue long to span the thickness of the hydrophobic portion of a lipid bilayer.



Transmembrane β structures

Porins are proteins which cross a cellular membrane by forming a large cylindrical transmembrane up-down β -barrel. These proteins act as a pore through which molecules can diffuse.

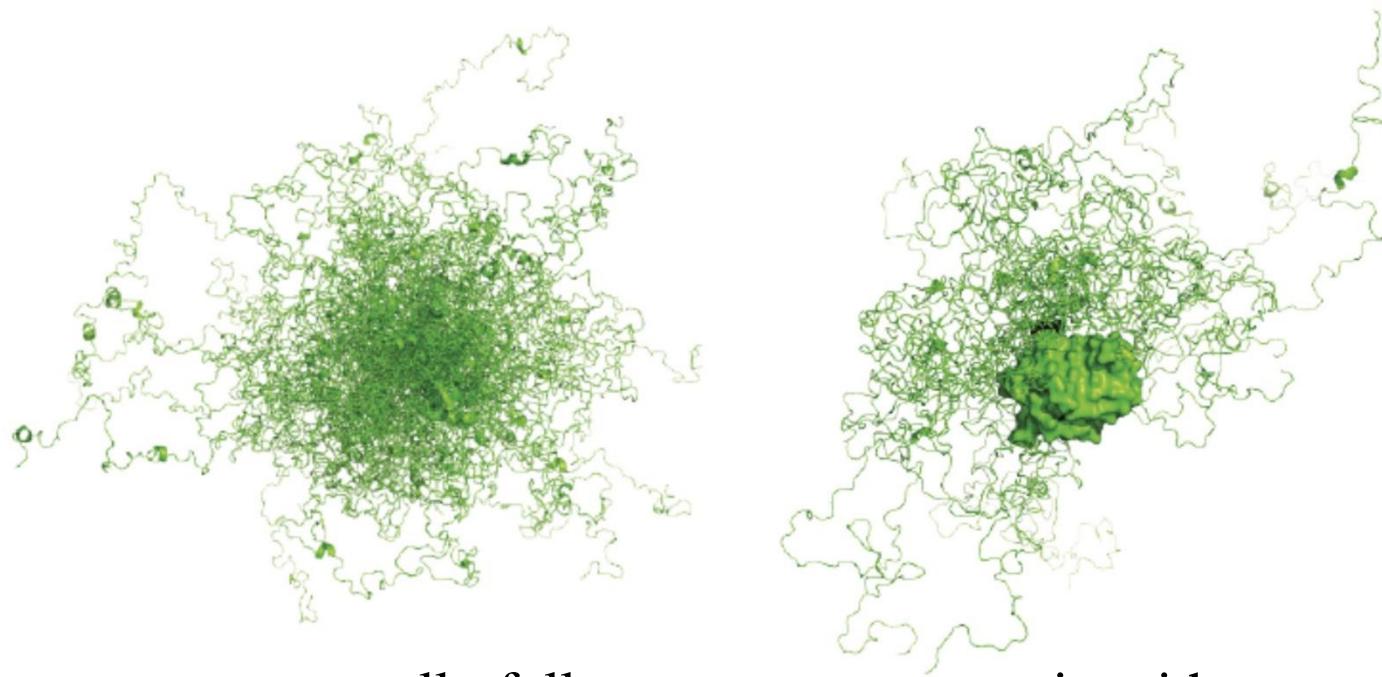
Structure of a porin



Intrinsically disordered proteins - 1

Recent studies suggest that protein can have a function in the absence of well-defined three-dimensional structure.

Approximately 50% of mammalian proteins contain long disordered regions (more than 30 residues), and approximately 25% of their proteins are expected to be fully disordered under physiological conditions.



a structurally fully
disordered protein

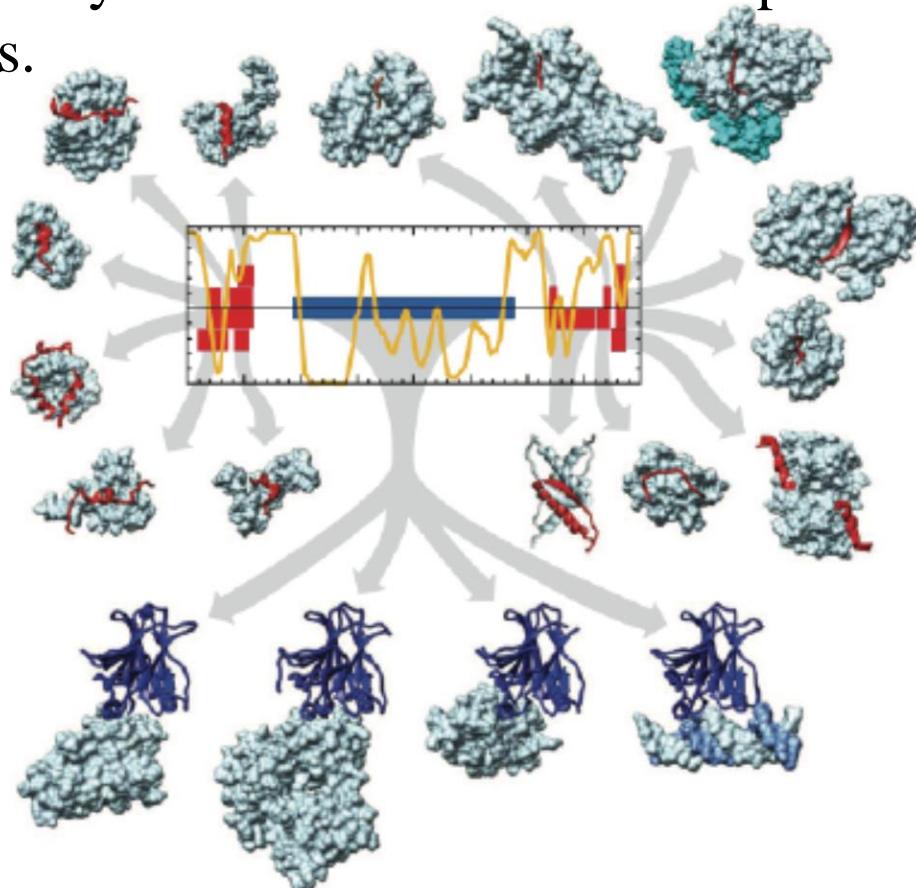
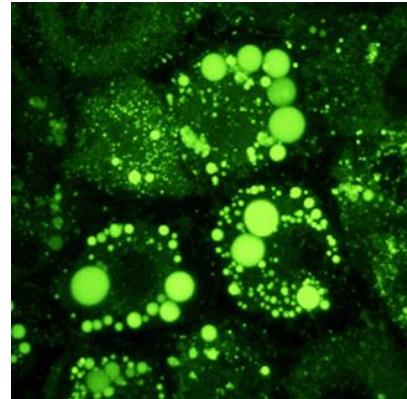
a protein with
disordered regions

Intrinsically disordered proteins - 2

Functional advantages of intrinsically disordered proteins: (1) they have a larger solvent exposed surface to enhance the chance to find their binding partners or being post-translationally modified; (2) they allow themselves to act as scaffolds to interact with different proteins (**usually via the formation of liquid-liquid phase separation (LLPS); lipid droplet**) and adapting different conformations; (3) they act as a flexible linker to provide the dynamics of multidomain proteins.

The intrinsically disordered regions of p53 protein may act as a ‘hub’ to interact with different binding partners.

LLPS/lipid droplets formed inside cells.



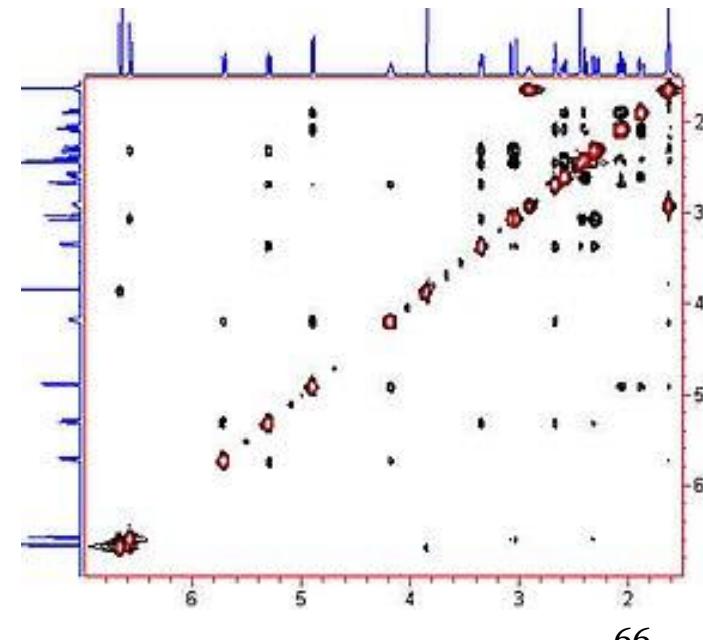
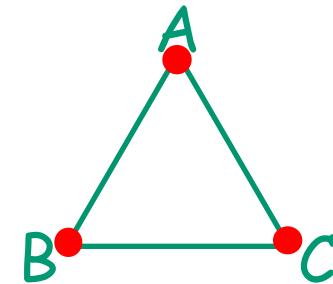
解析生物巨分子立體結構的工具

- 核磁共振 (Nuclear Magnetic Resonance, NMR)
- X-射線結晶學 (X-ray crystallography)
- 冷凍電子顯微鏡 (Cryo Electron Microscopy, cryo-EM)

核磁共振 (NMR)

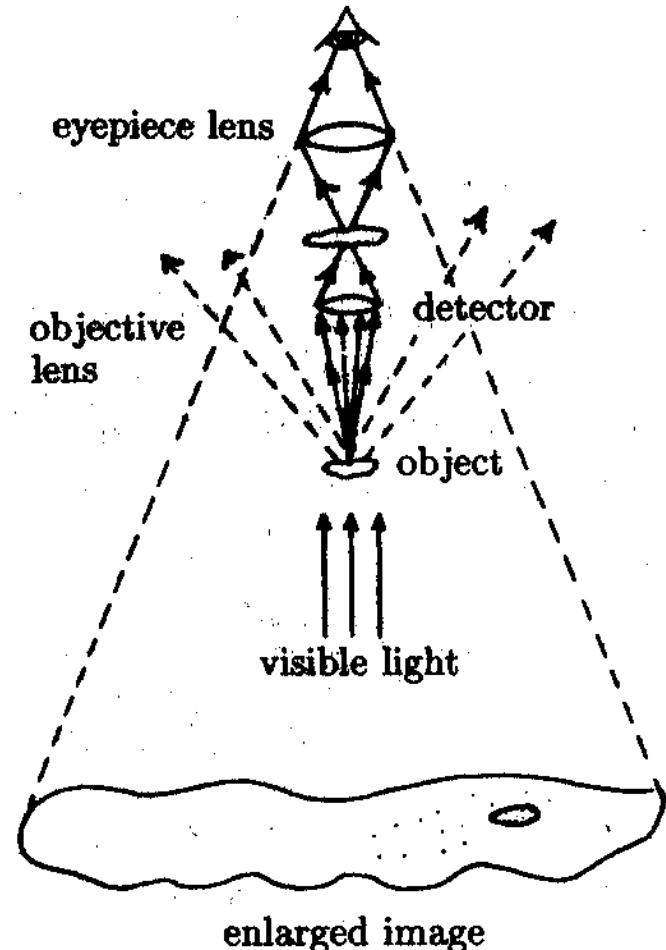
Using the distances between spin centers to determine structure.

Example: A molecule is composed of three atoms, A, B, and C. The distance between A and B is 1.5 Å, between B and C is 1.5 Å, and between C and A is also 1.5 Å. What is the structure of this molecule?

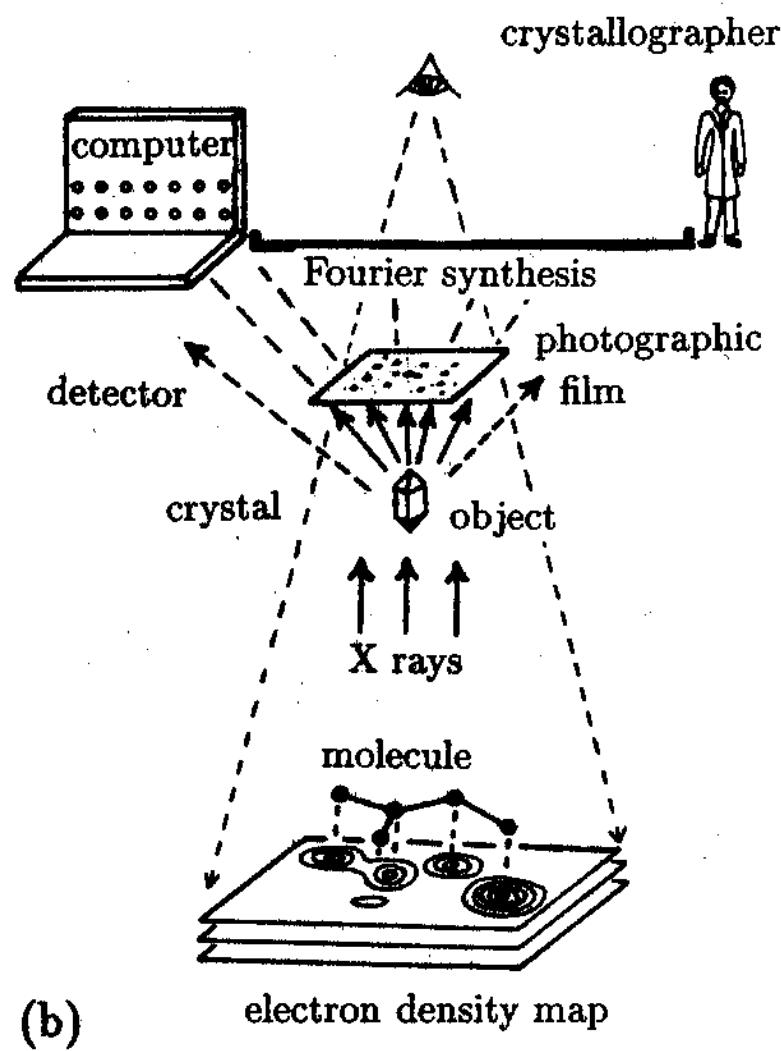


X-ray crystallography

Analogies between light microscopy and X-ray crystallography.

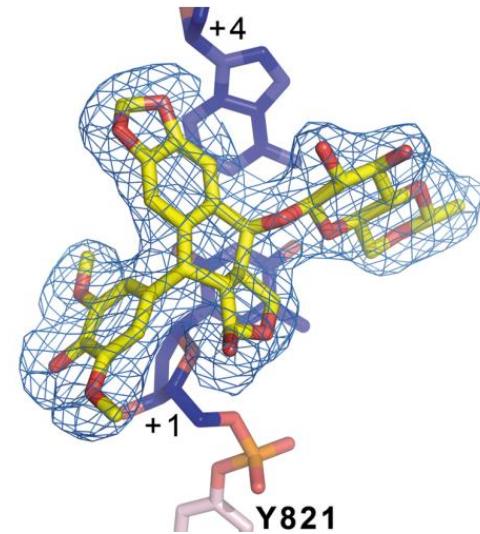
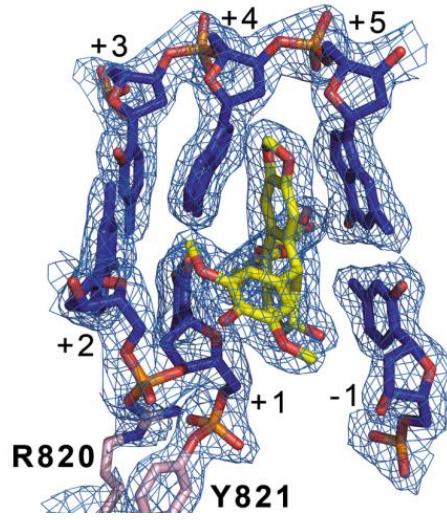
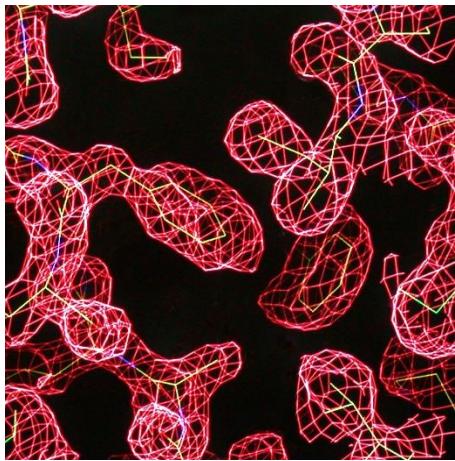
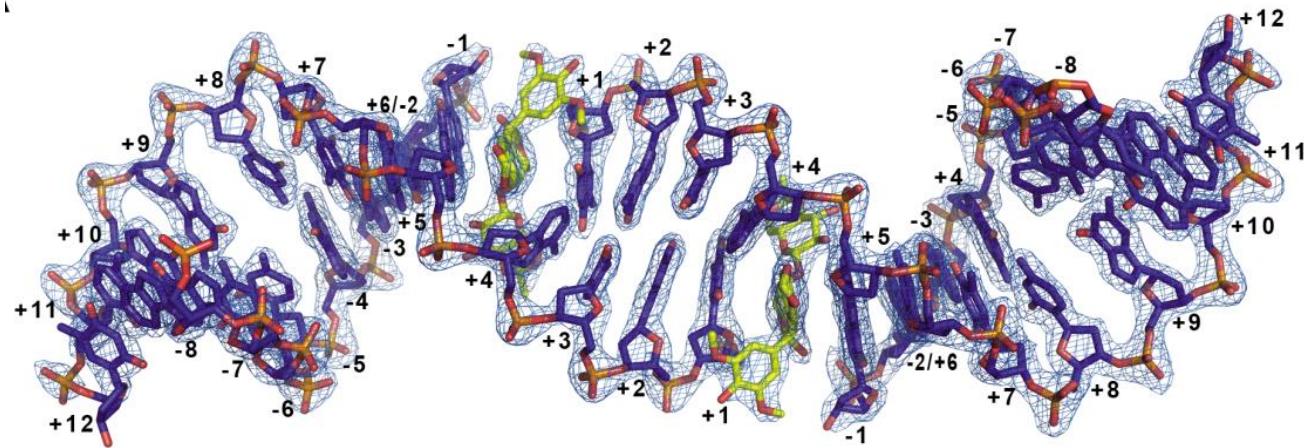


(a)



(b)

X-ray crystallography can provide the electron density map of biomacromolecules



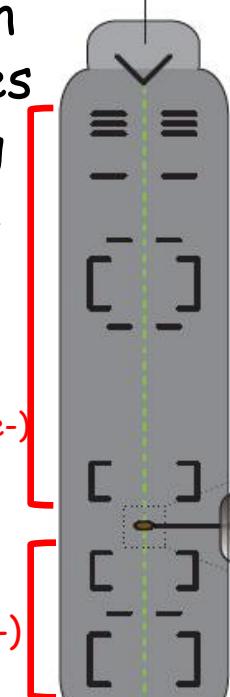
Cryo Electron Microscopy (cryo-EM) - 1

Structure determination through cryo-EM involves several stages: including sample grid preparation, data collection, data processing, and 3D reconstruction.

objective lens
(for producing e-)

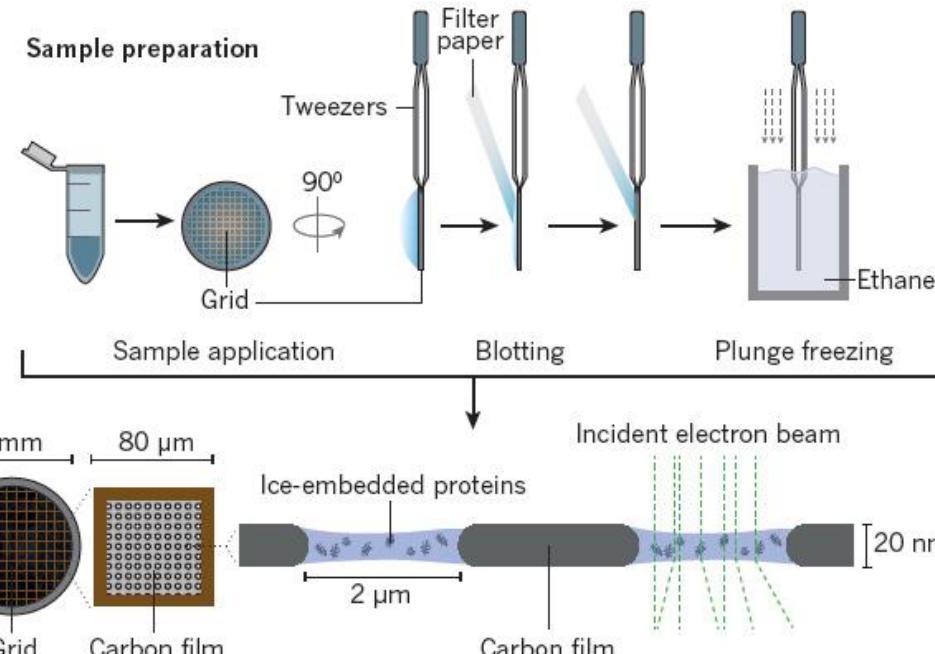
image lens (for focusing e-)

Transmission electron microscope

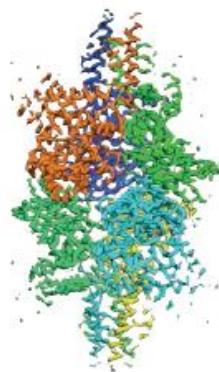
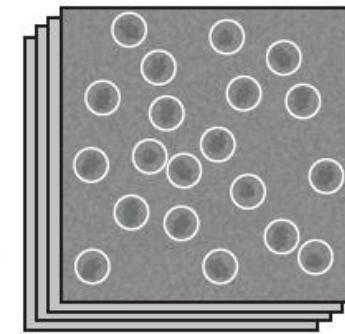


Detector

Sample preparation



Data processing and 3D reconstruction



Glutamate dehydrogenase

Cryo Electron Microscopy (cryo-EM) - 2

Illustration of data collection, data processing, and 3D reconstruction:

Many images of the same complex are obtained. The images are separated by "boxing". All the boxed images are then clustered. The images in each cluster are averaged to improve the S/N ratio. The orientation of the averaged images is then found and back-projected to produce a 3D density map.

