

Science of Living System (BS20001)

- **SOUMYA DE**

- **Office:** 2nd Floor, Diamond Jubilee Building
- **Phone:** 03222-2284552
- **E-mail:** somde@iitkgp.ac.in
- **Website:** <http://iitkgpbioscience.weebly.com/soumya-de.html>
- **Research interests:**
 - Biophysics
 - Nuclear Magnetic Resonance (NMR) Spectroscopy
 - Protein Engineering
 - Signal Transduction and Gene Expression
 - Enzymology

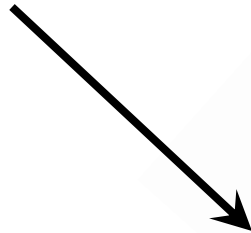
Central Dogma of molecular biology



DNA: Storage Medium

Polymer of nucleotides

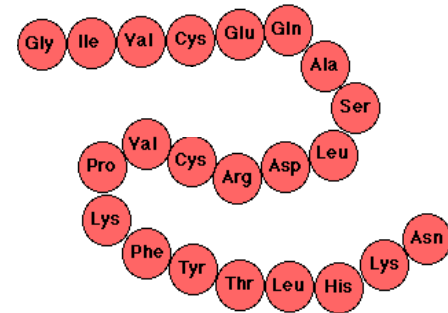
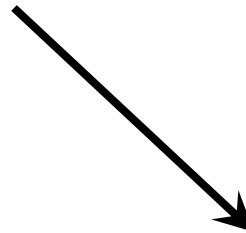
CCTGAGCCAACTATTGATGAA



RNA: Transmission Medium

Polymer of nucleotides

CCUGAGCCAAACUAUUGAUGAA



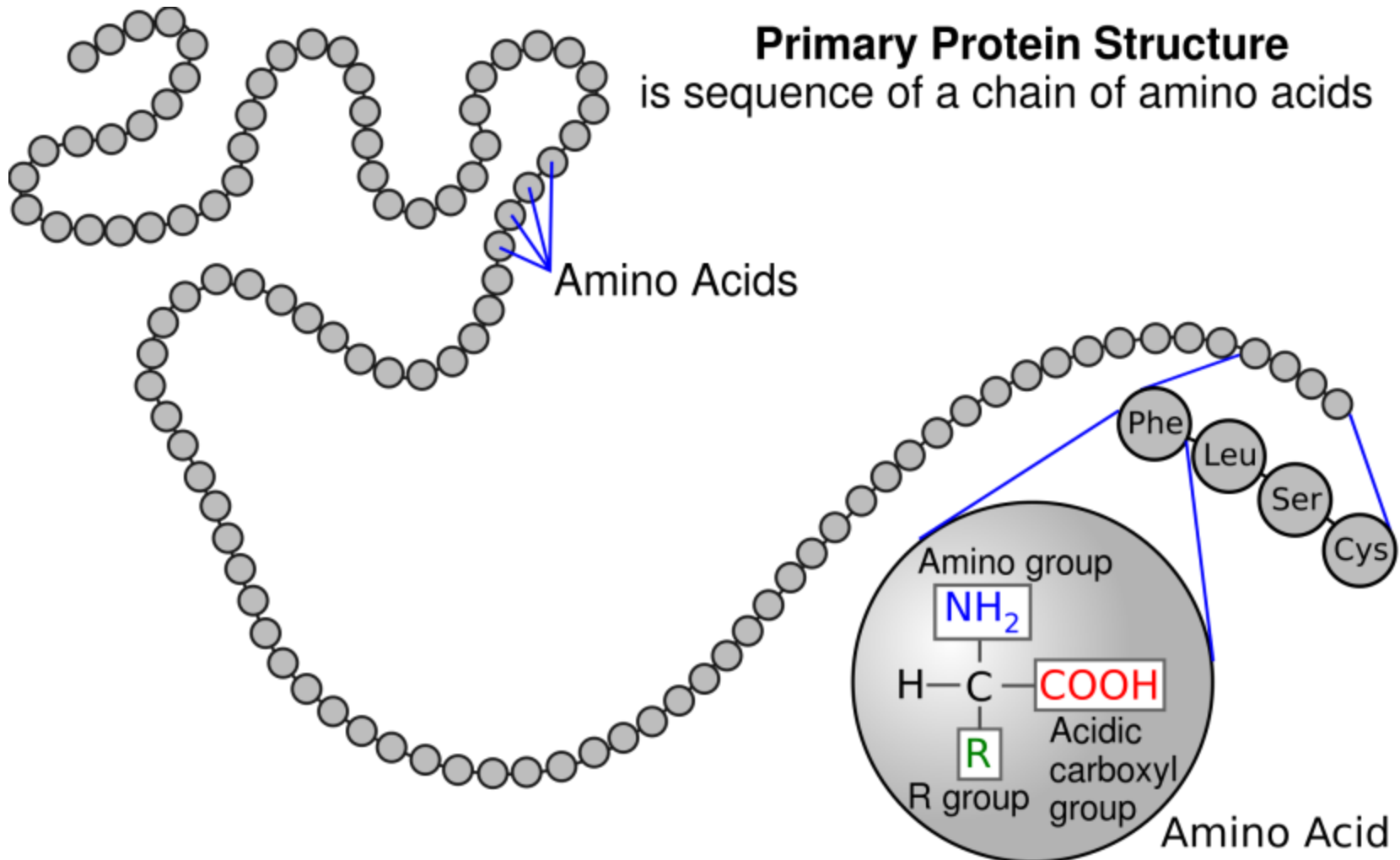
Protein: Molecular Machines

Polymer of amino acids

PEPTIDE

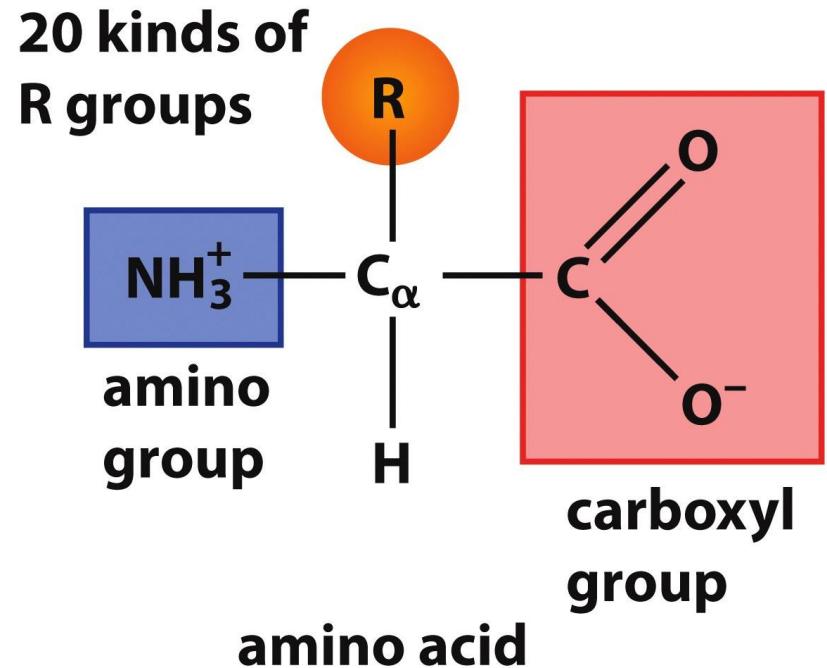
Primary Structure of Proteins

The primary structure of a protein is its amino acid sequence



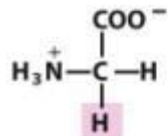
Amino acids: Building blocks of Proteins

- Protein is a polymer of amino acids.
- There are 20 common amino acids.
- Amino acids have a common chemical structure - A tetrahedral sp^3 carbon (C_α) with four different functional groups:
 1. Amino group
 2. Carboxyl group
 3. H-atom
 4. Side chain (R) with distinct chemical property

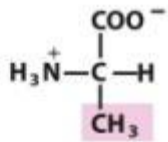


The 20 Common Amino Acids of Proteins

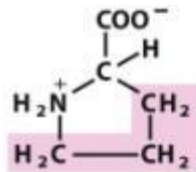
Nonpolar, aliphatic R groups



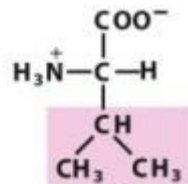
Glycine



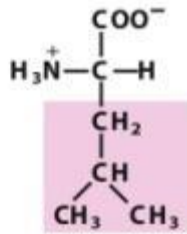
Alanine



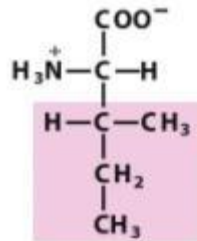
Proline



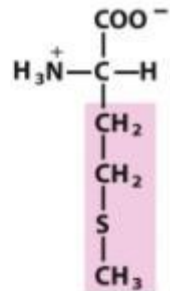
Valine



Leucine

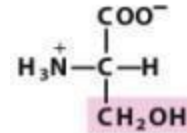


Isoleucine

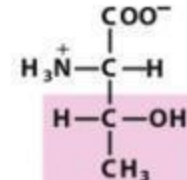


Methionine

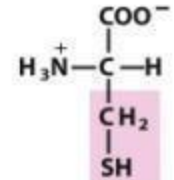
Polar, uncharged R groups



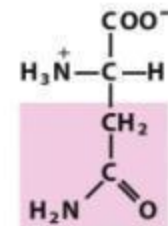
Serine



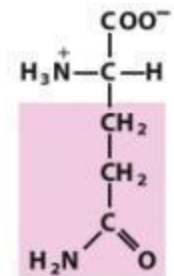
Threonine



Cysteine

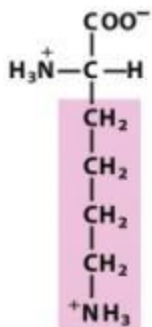


Asparagine

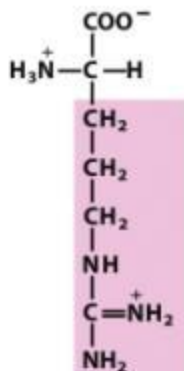


Glutamine

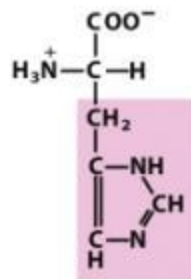
Positively charged R groups



Lysine

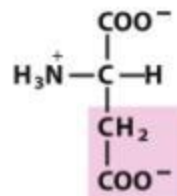


Arginine

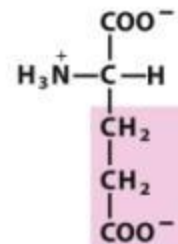


Histidine

Negatively charged R groups

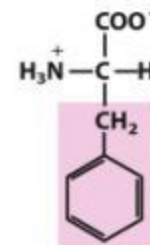


Aspartate

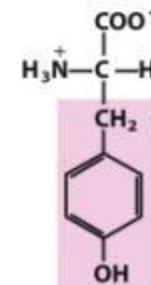


Glutamate

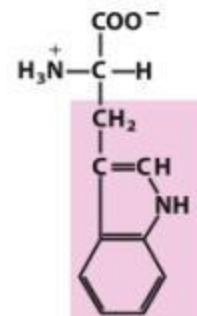
Aromatic R groups



Phenylalanine



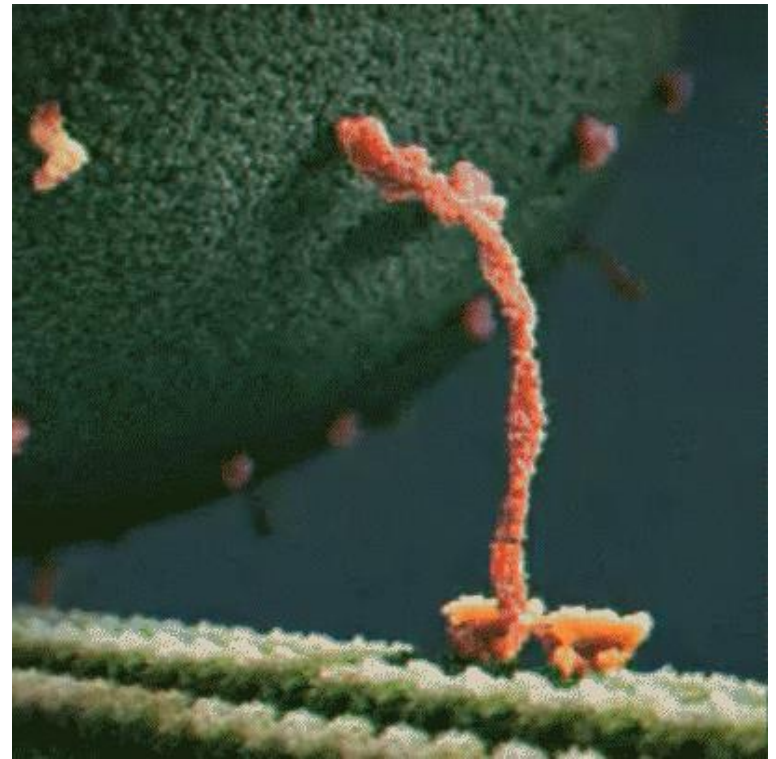
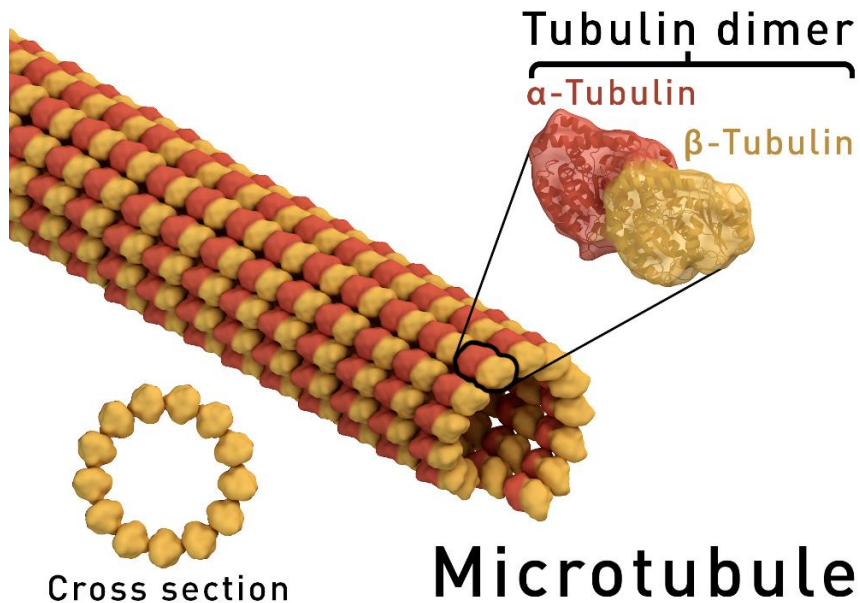
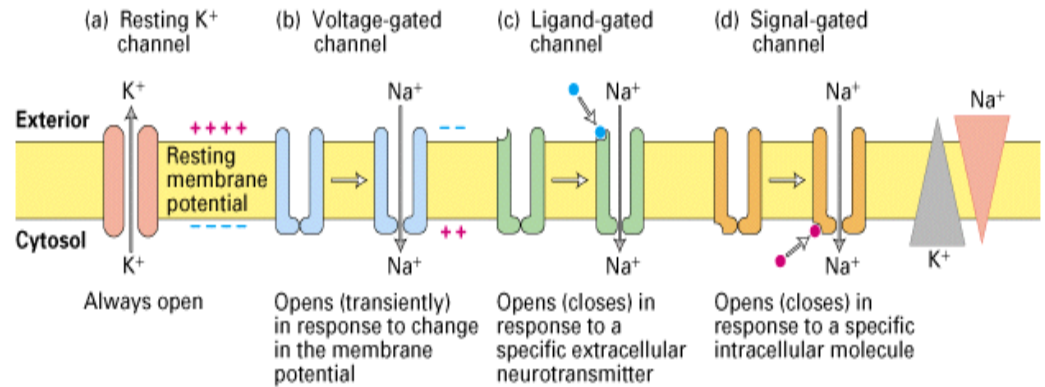
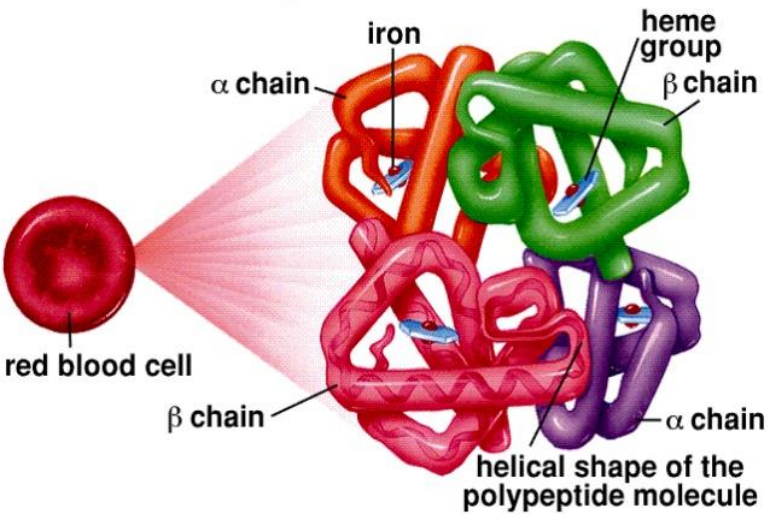
Tyrosine



Tryptophan

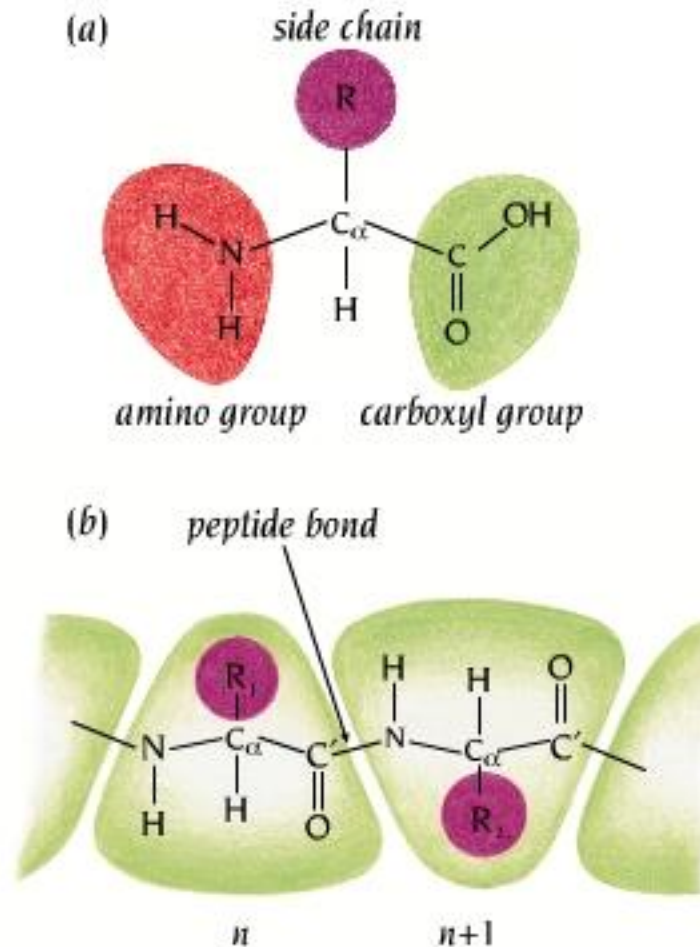
Proteins come in various shapes and sizes

Hemoglobin Molecule

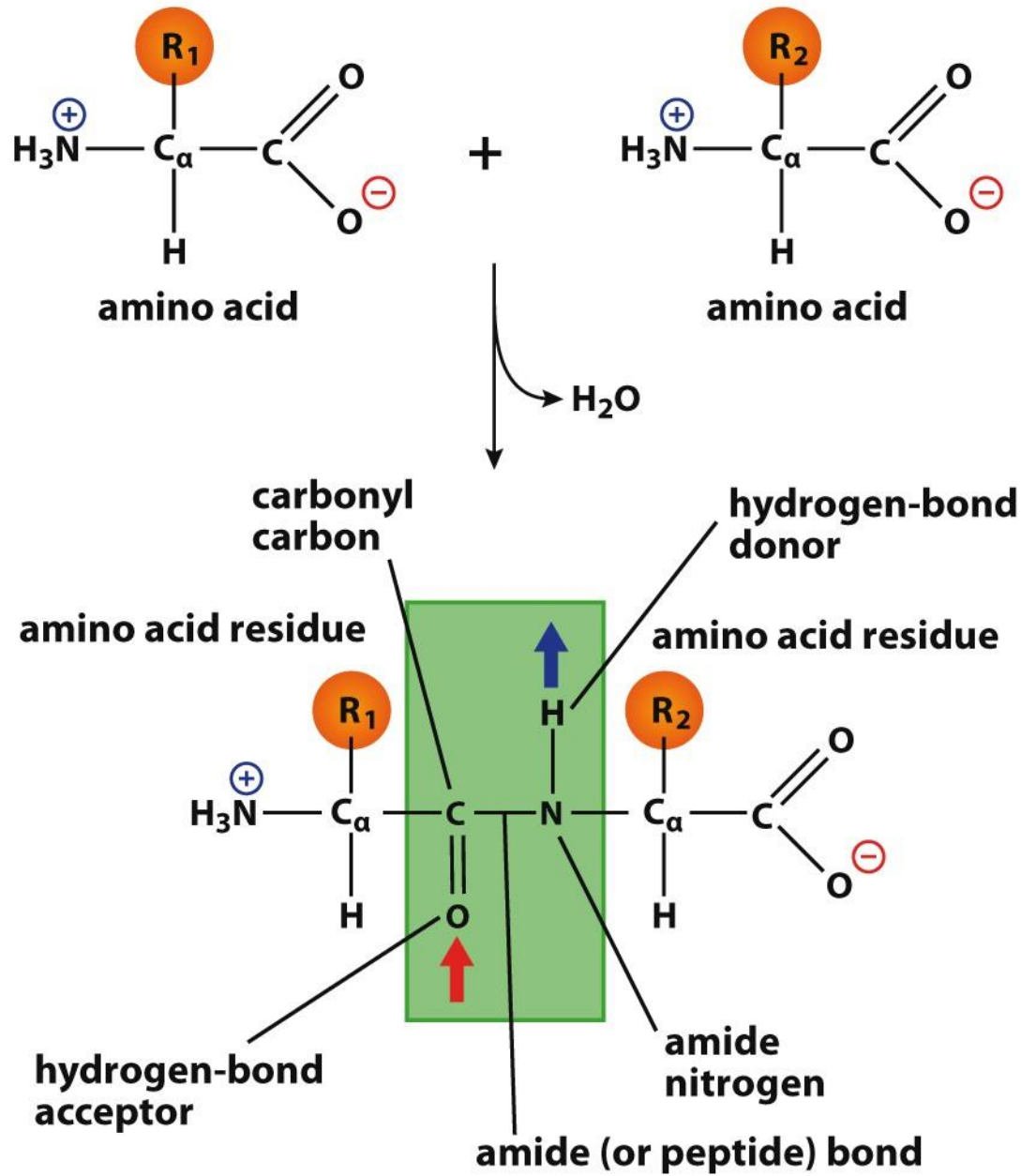


Proteins are polypeptide chains

Successive polypeptide bonds: main chain or backbone



Formation of the peptide bond



The amide plane: partial double bond character of the peptide bond

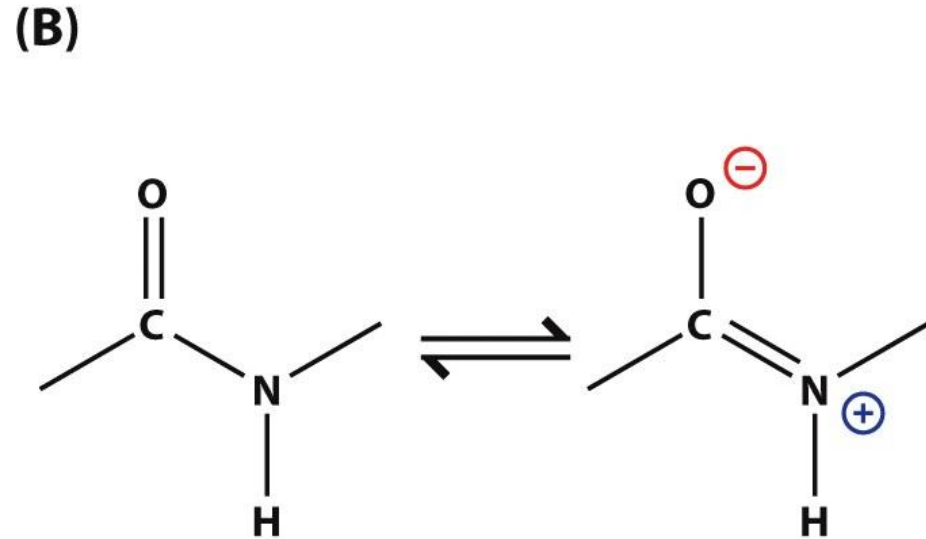
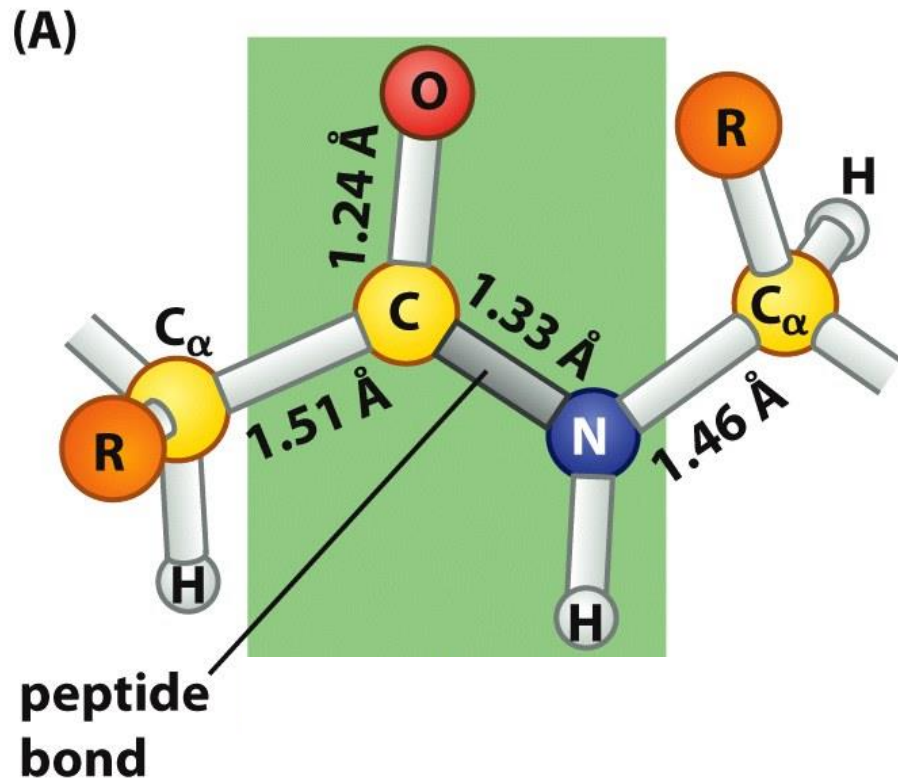
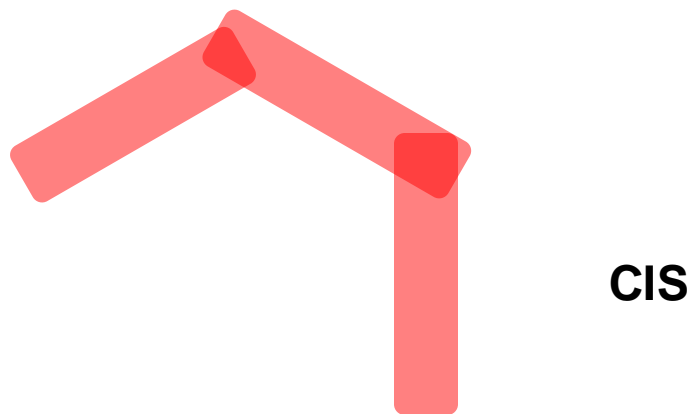
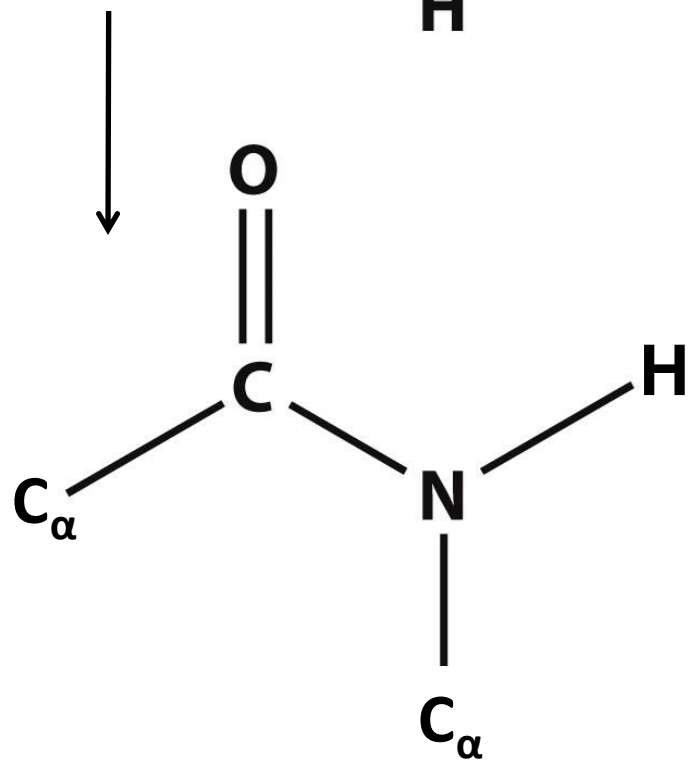
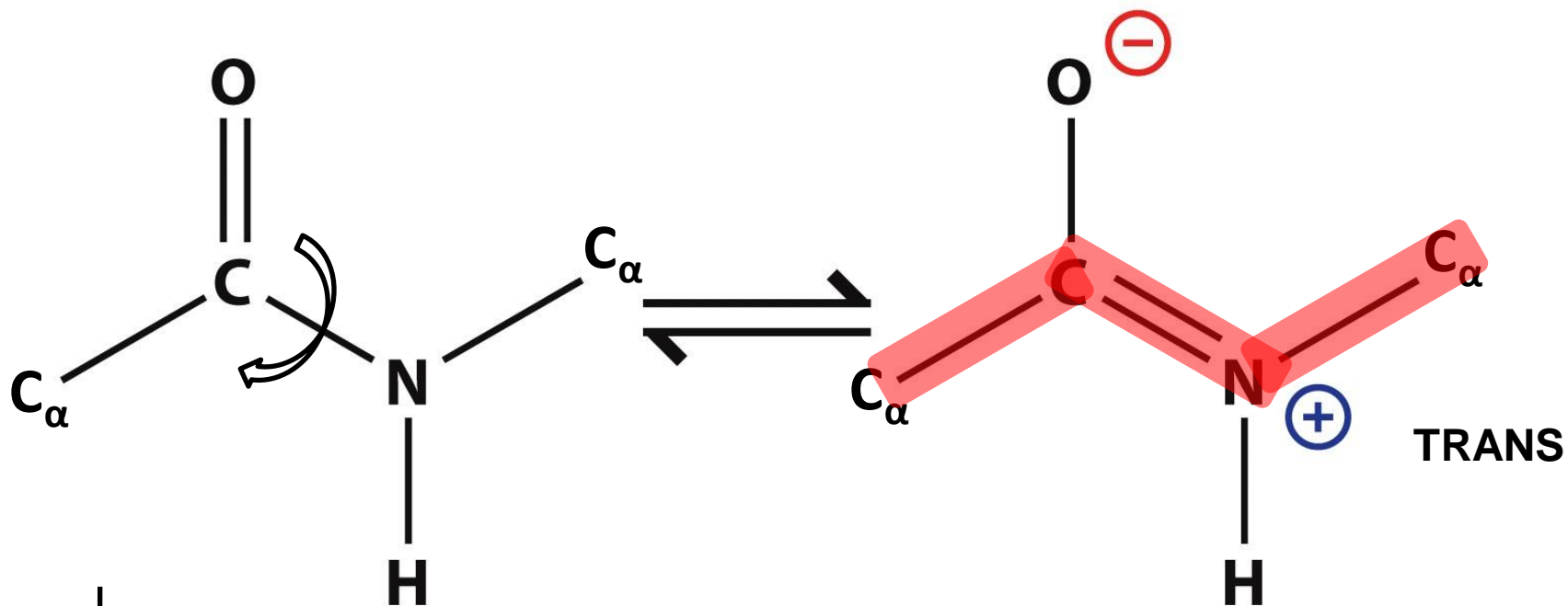


Figure 4.14 The Molecules of Life (© Garland Science 2013)



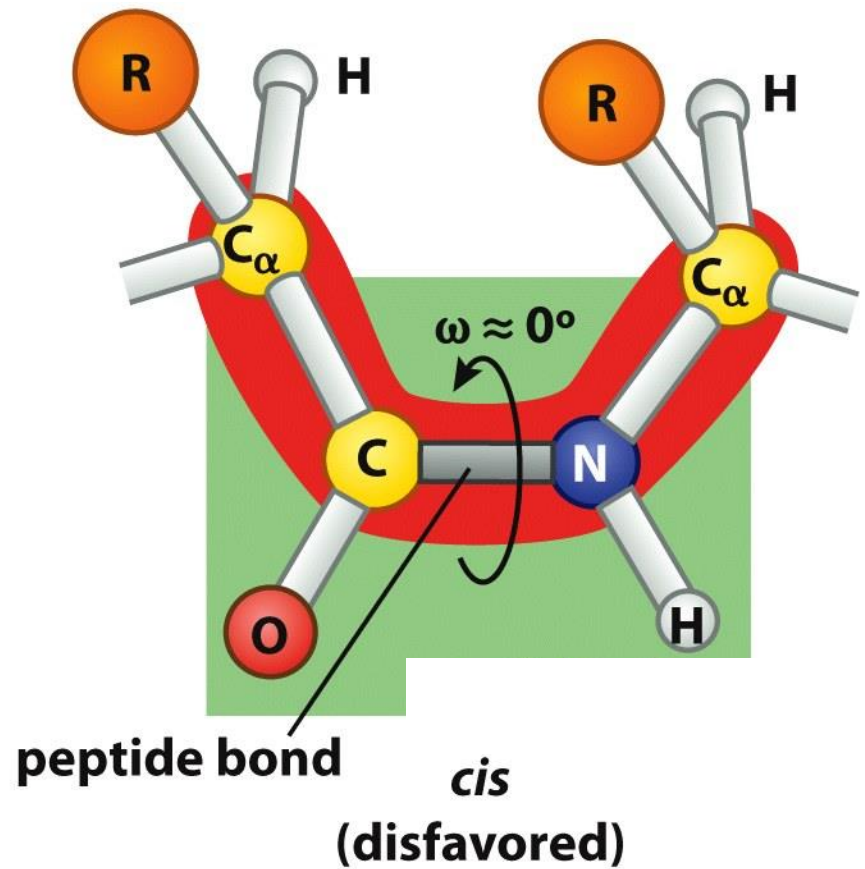
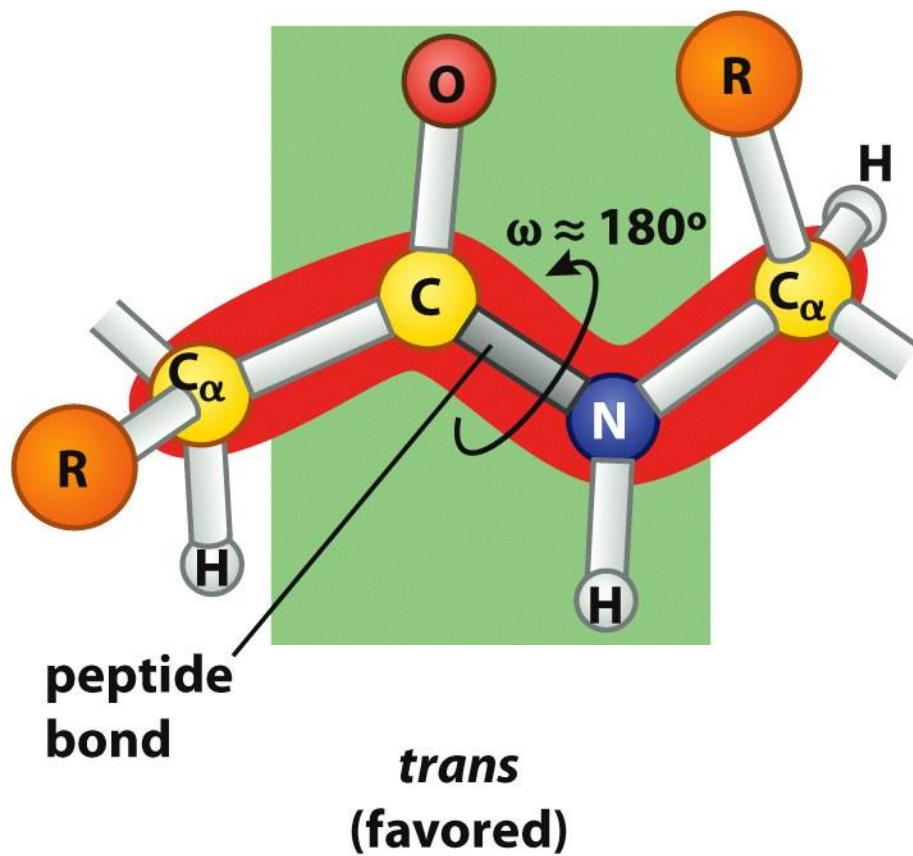
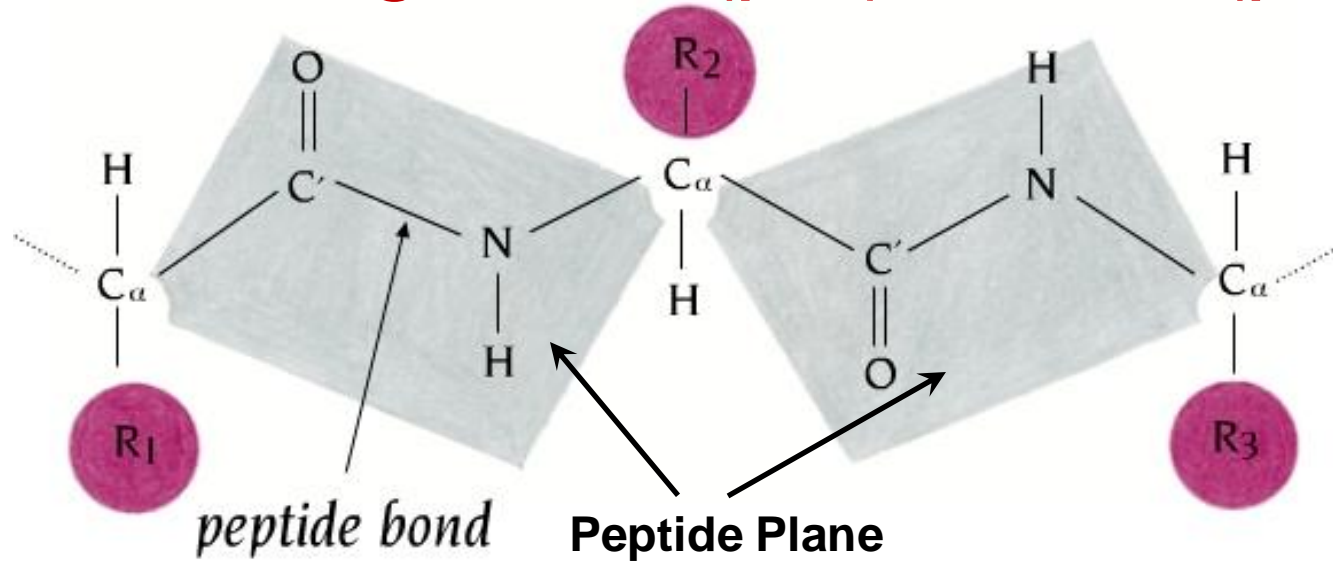
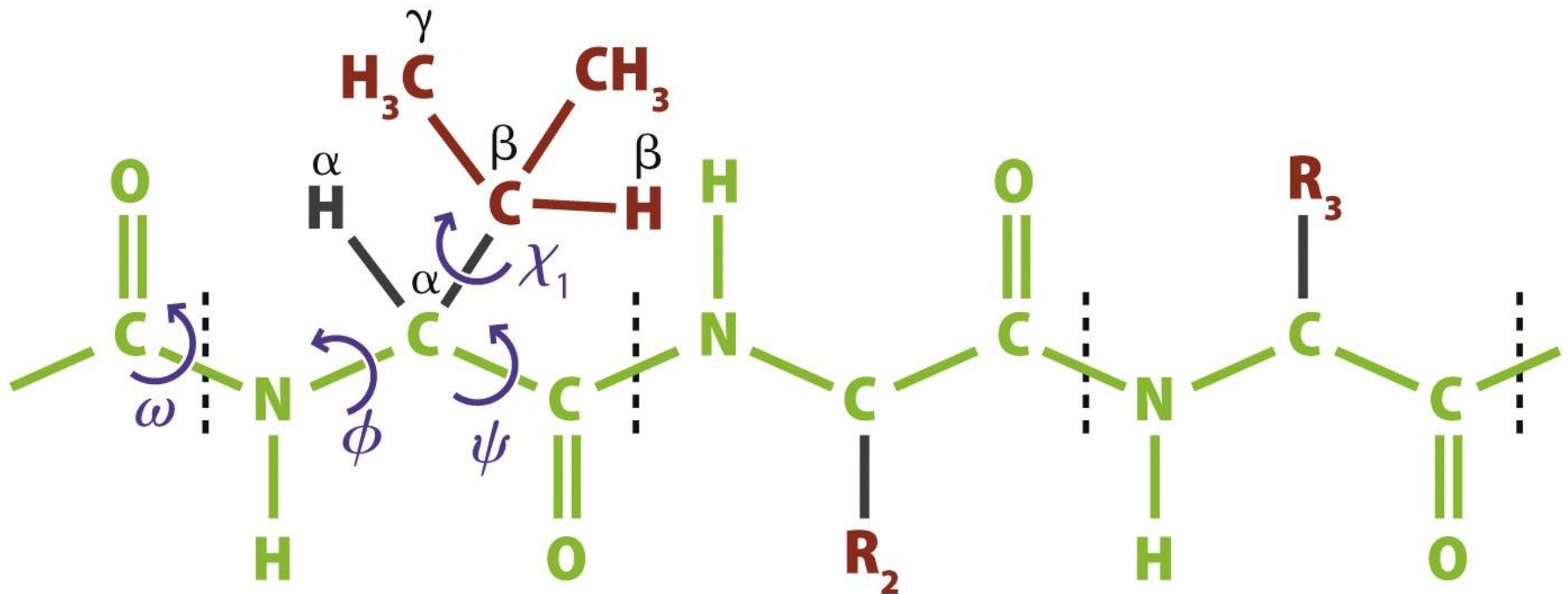


Figure 4.15 The Molecules of Life (© Garland Science 2013)

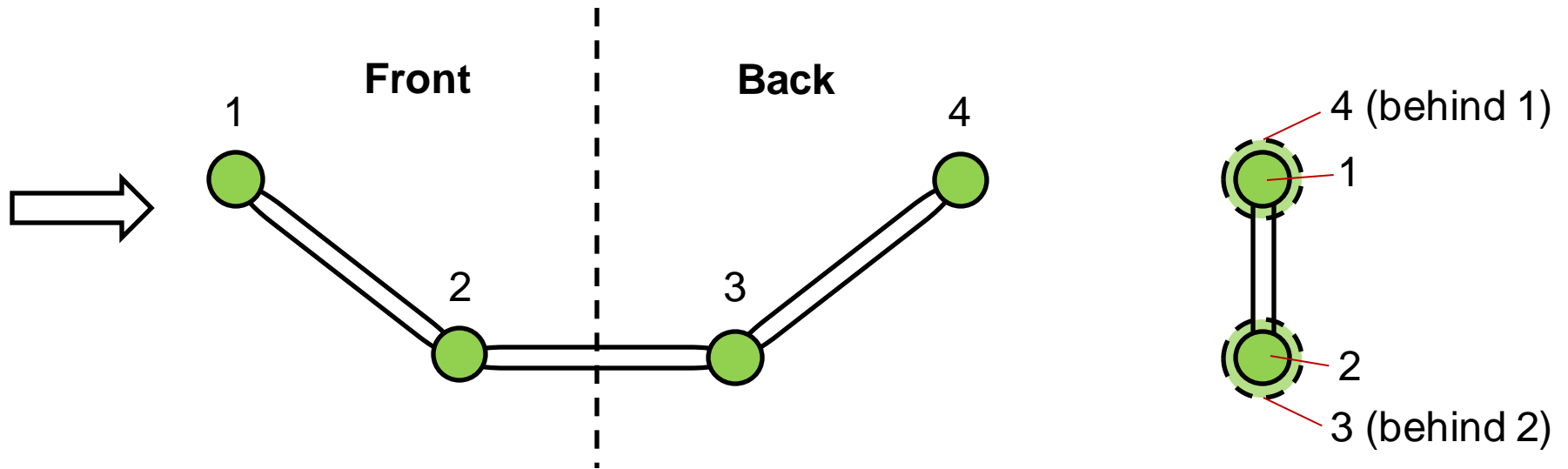
Torsion angles: Φ (phi) and Ψ (psi)



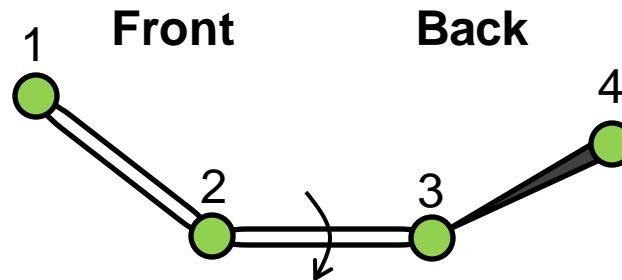
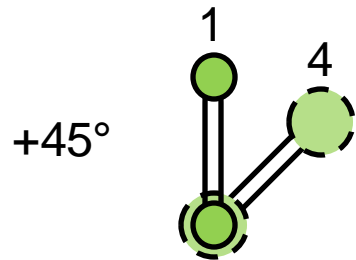
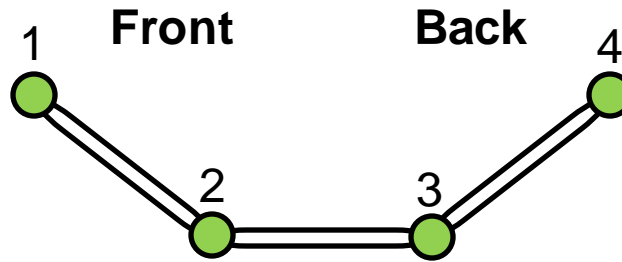
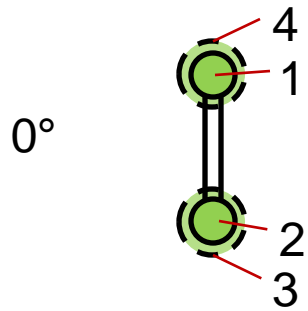
©1999 GARLAND PUBLISHING INC.



Visualizing a few torsion angles

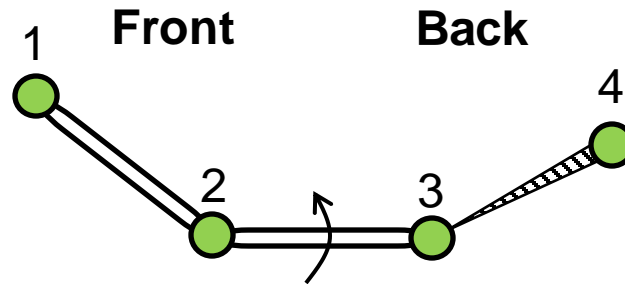
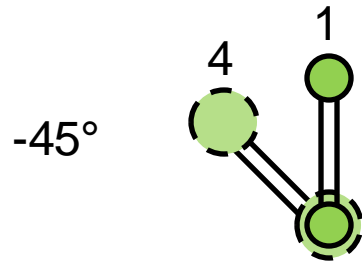


Visualizing a few torsion angles

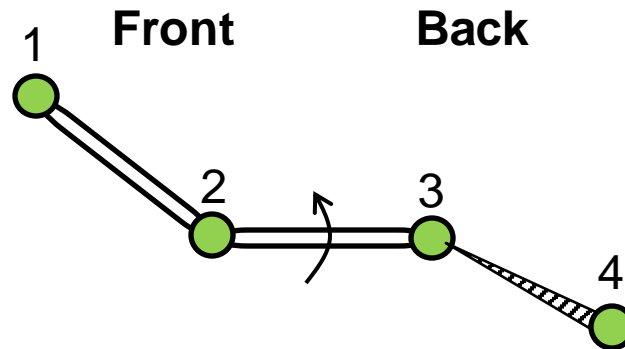
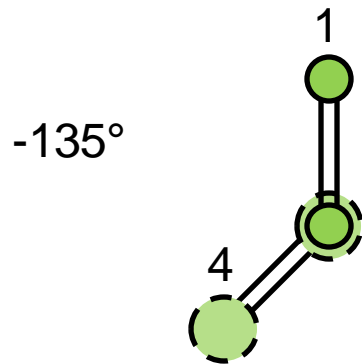


Atom 4 is above
the plane of the
board

Visualizing a few torsion angles



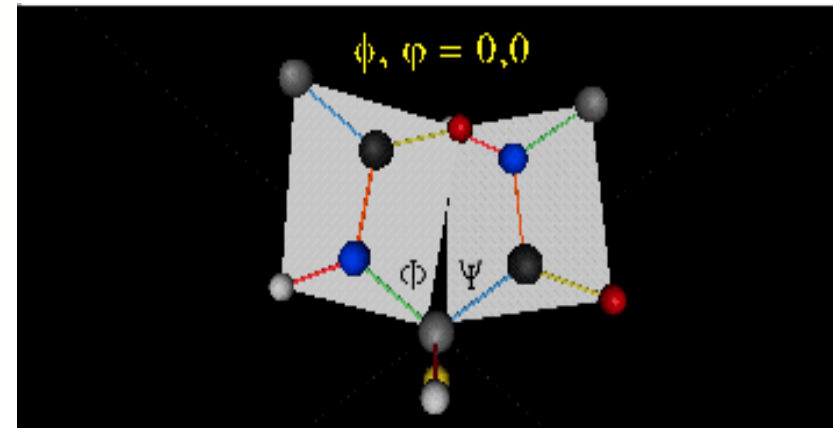
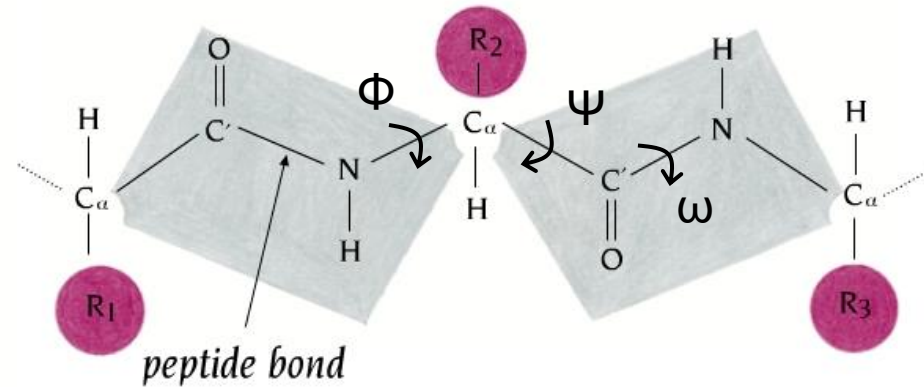
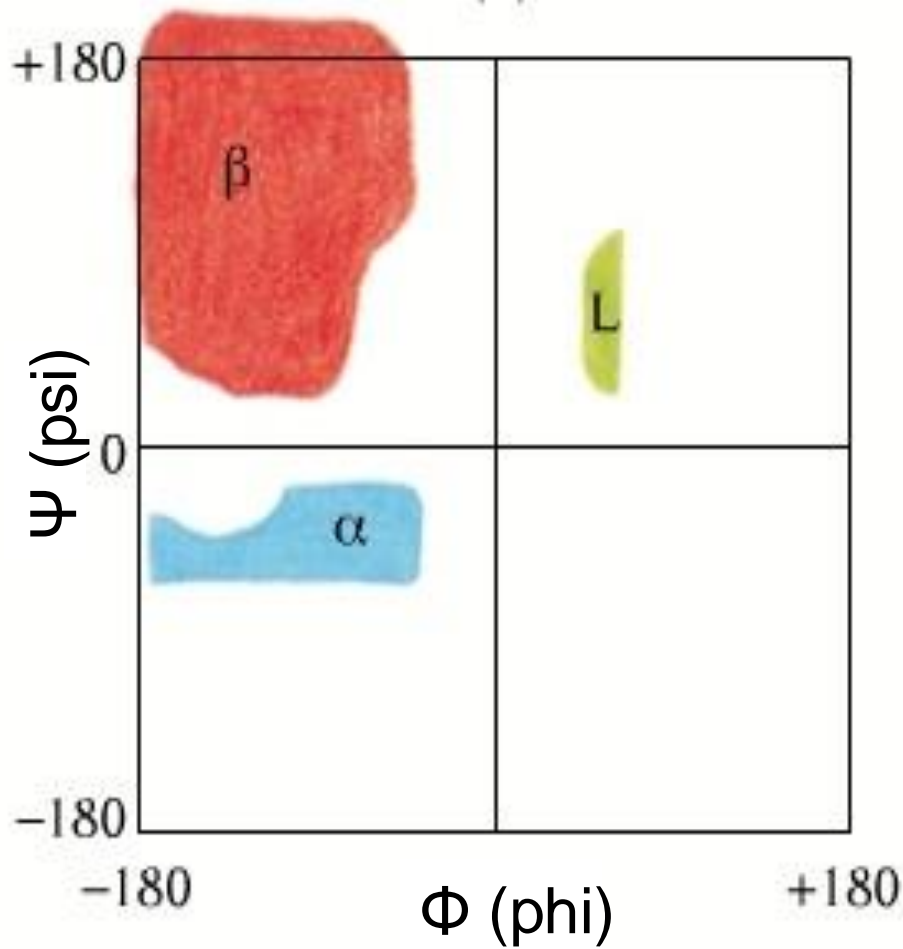
Atom 4 is below the plane of the board



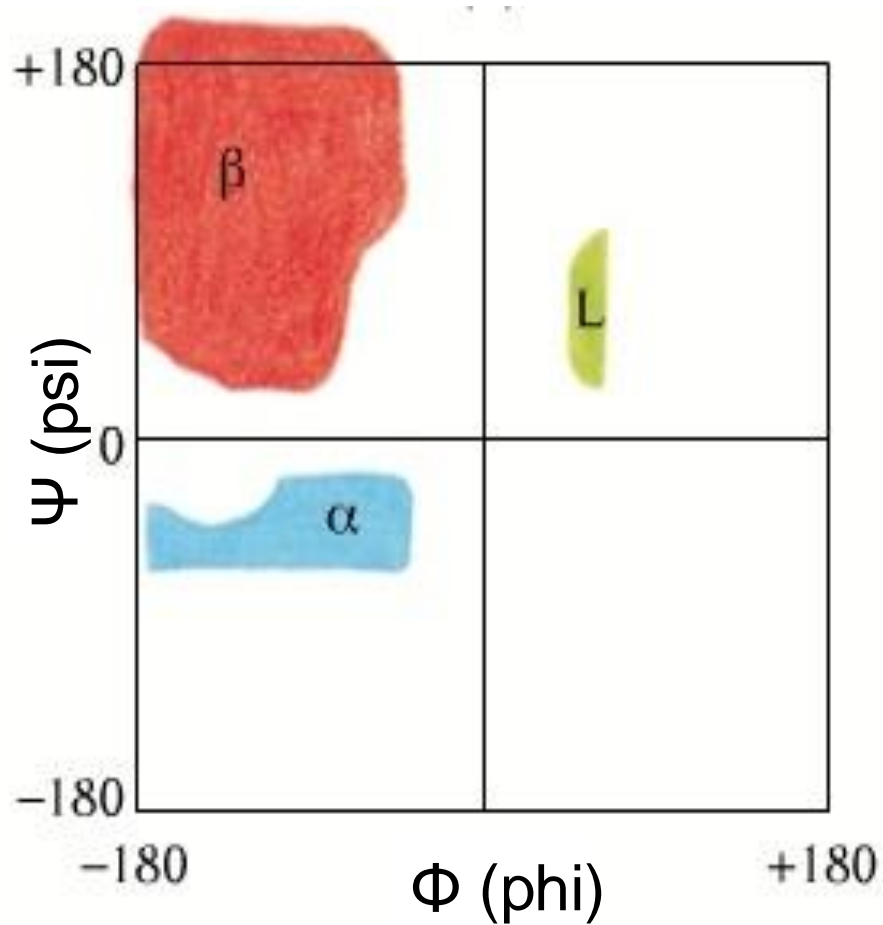
Atom 4 is below the plane of the board

Ramachandran Plot

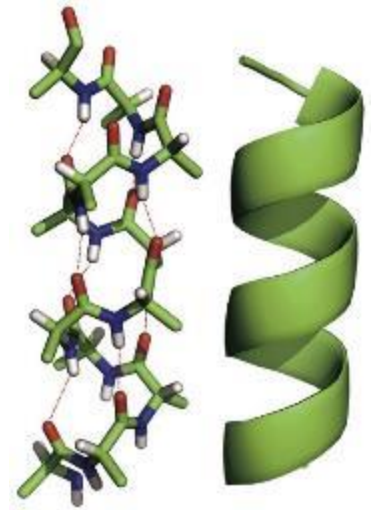
G. N. Ramachandran



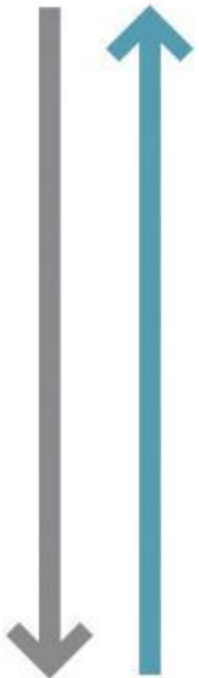
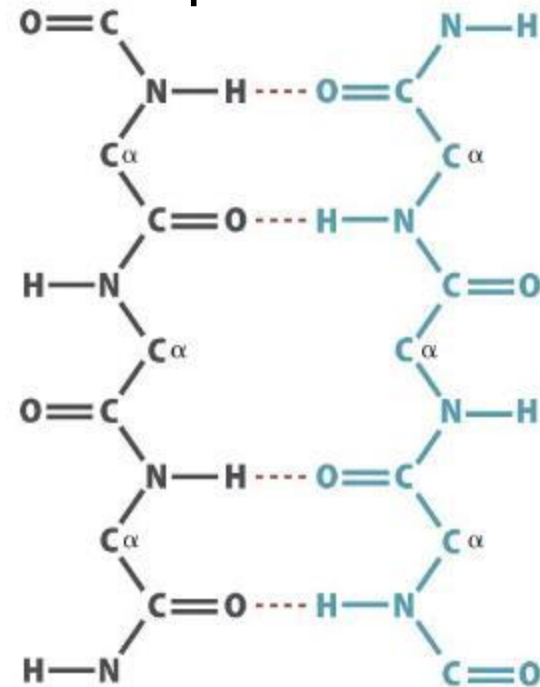
Ramachandran Plot



α -Helix



β -Strand



ϕ and ψ torsion angles are the only degrees of freedom for the backbone

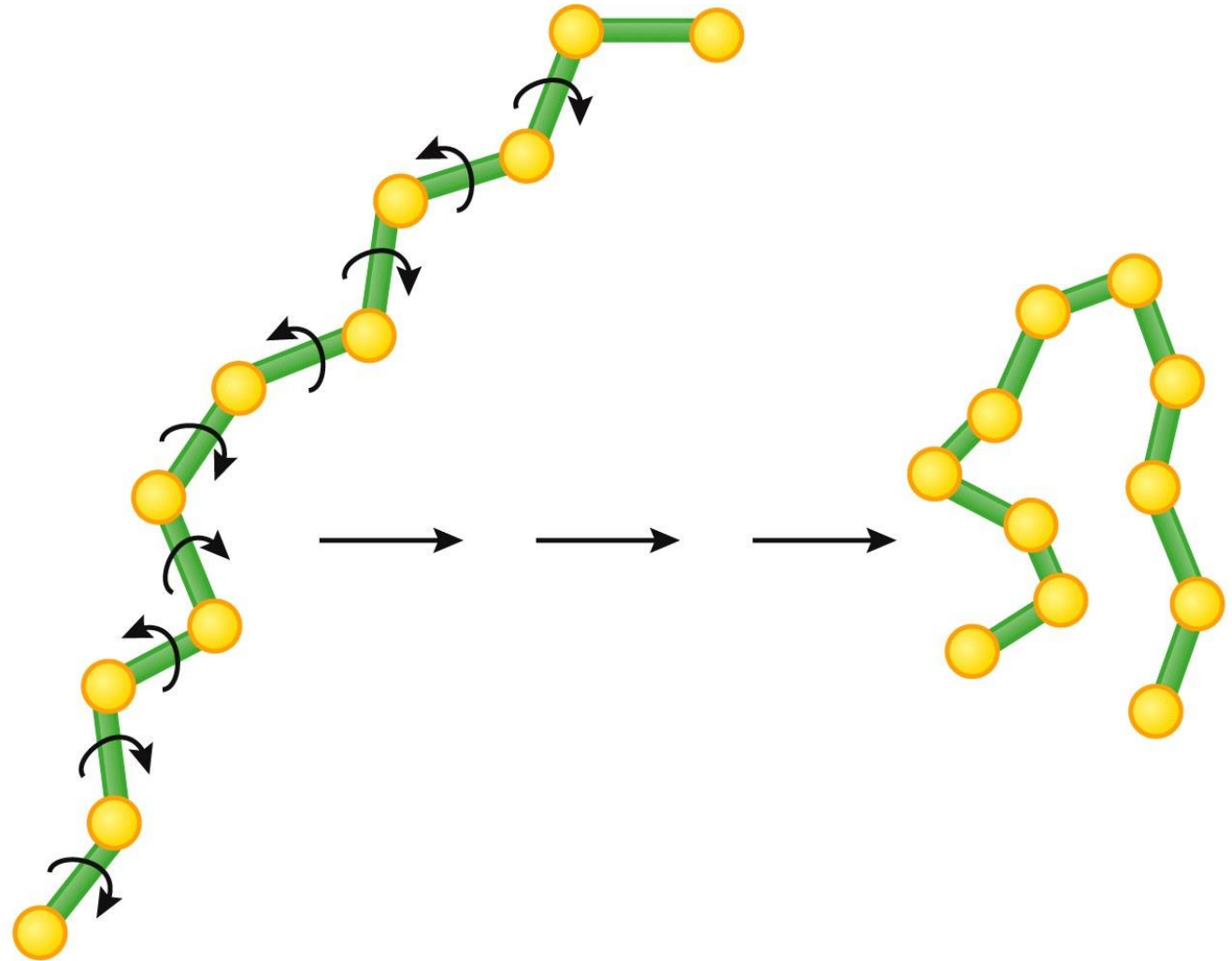
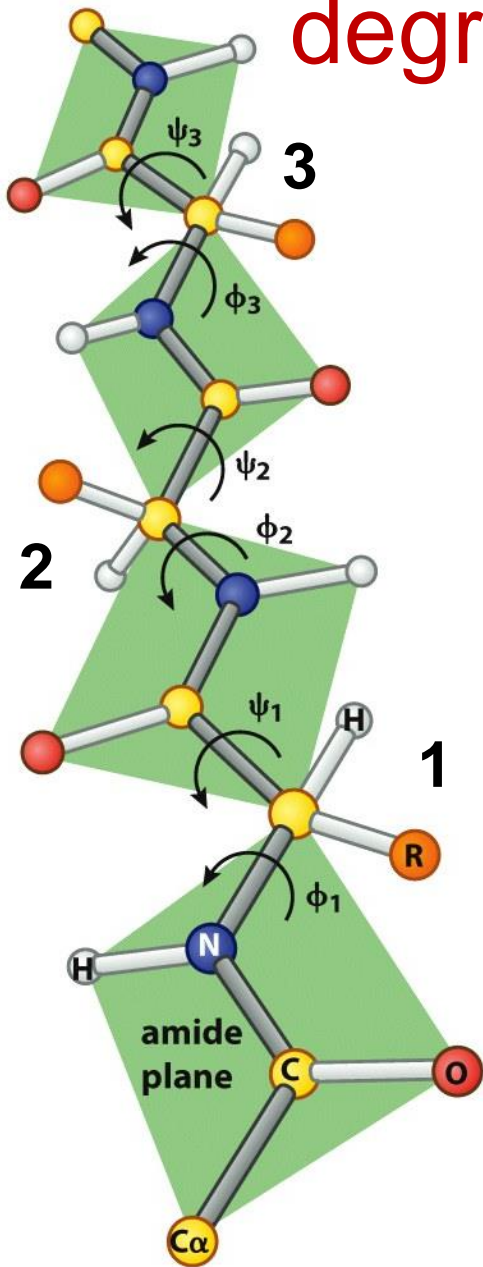
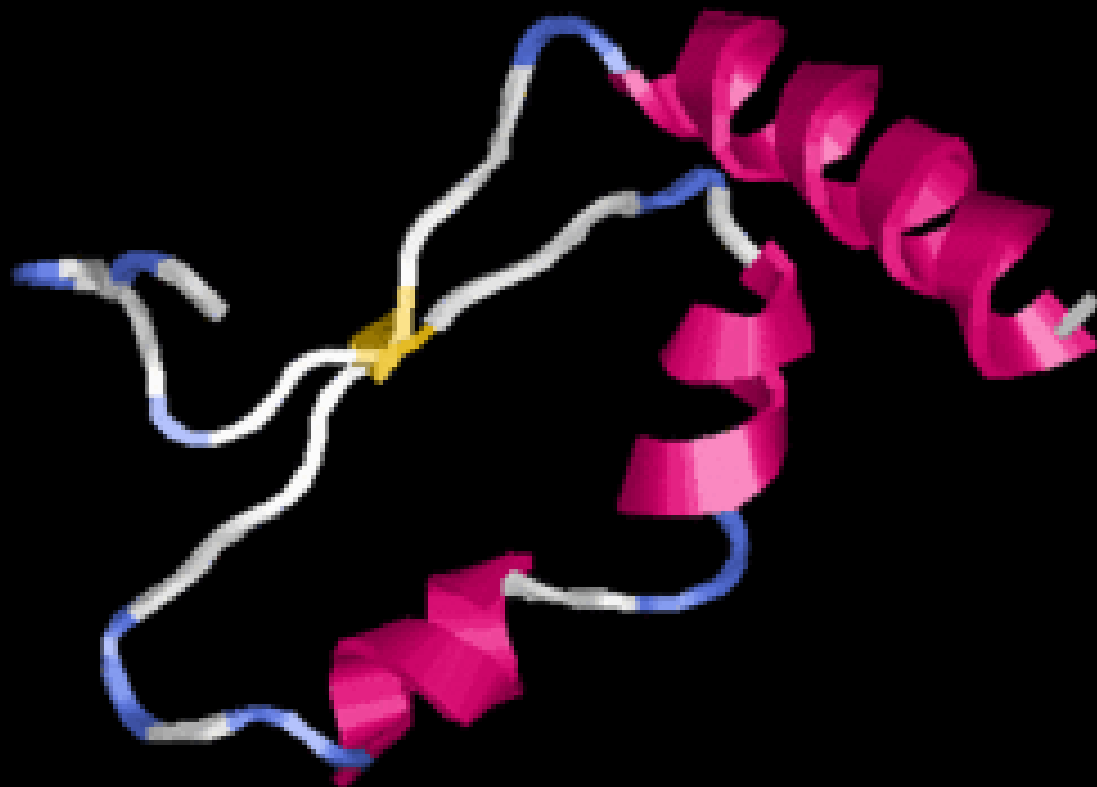
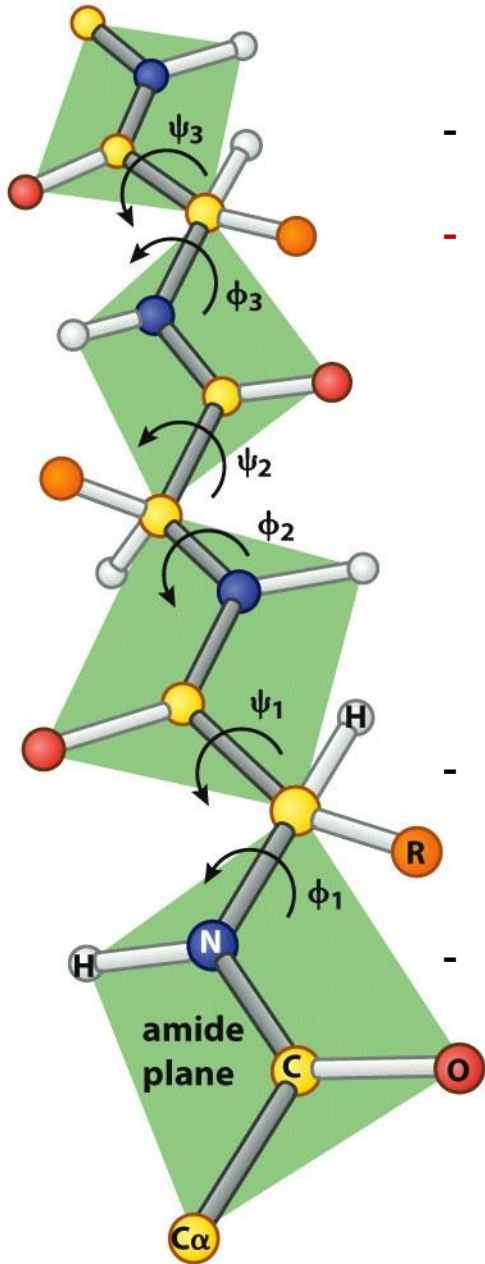


Figure 4.11b The Molecules of Life (© Garland Science 2013)



The protein folding problem



- Consider a small protein with 100 residues.
- **Cyrus Levinthal** calculated that, if each residue can assume three different conformations, the total number of structures would be 3^{100} , which is equal to 5×10^{47} . If it takes 10^{-13} s to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or **10^{27} years** i.e. longer than the age of the universe!
- Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations.
- The enormous difference between calculated and actual folding times is called ***Levinthal's paradox***.

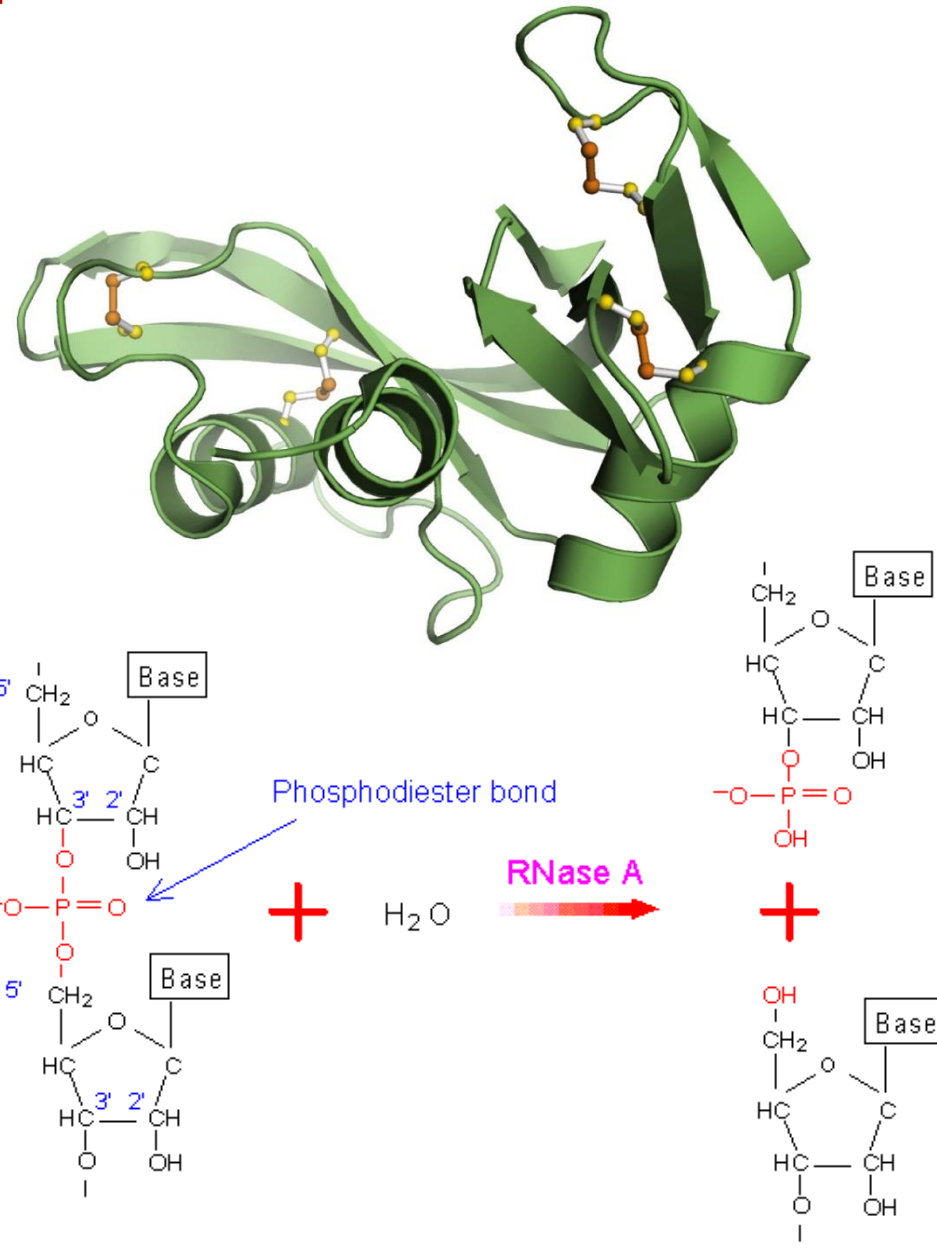
Important Questions on Protein Folding

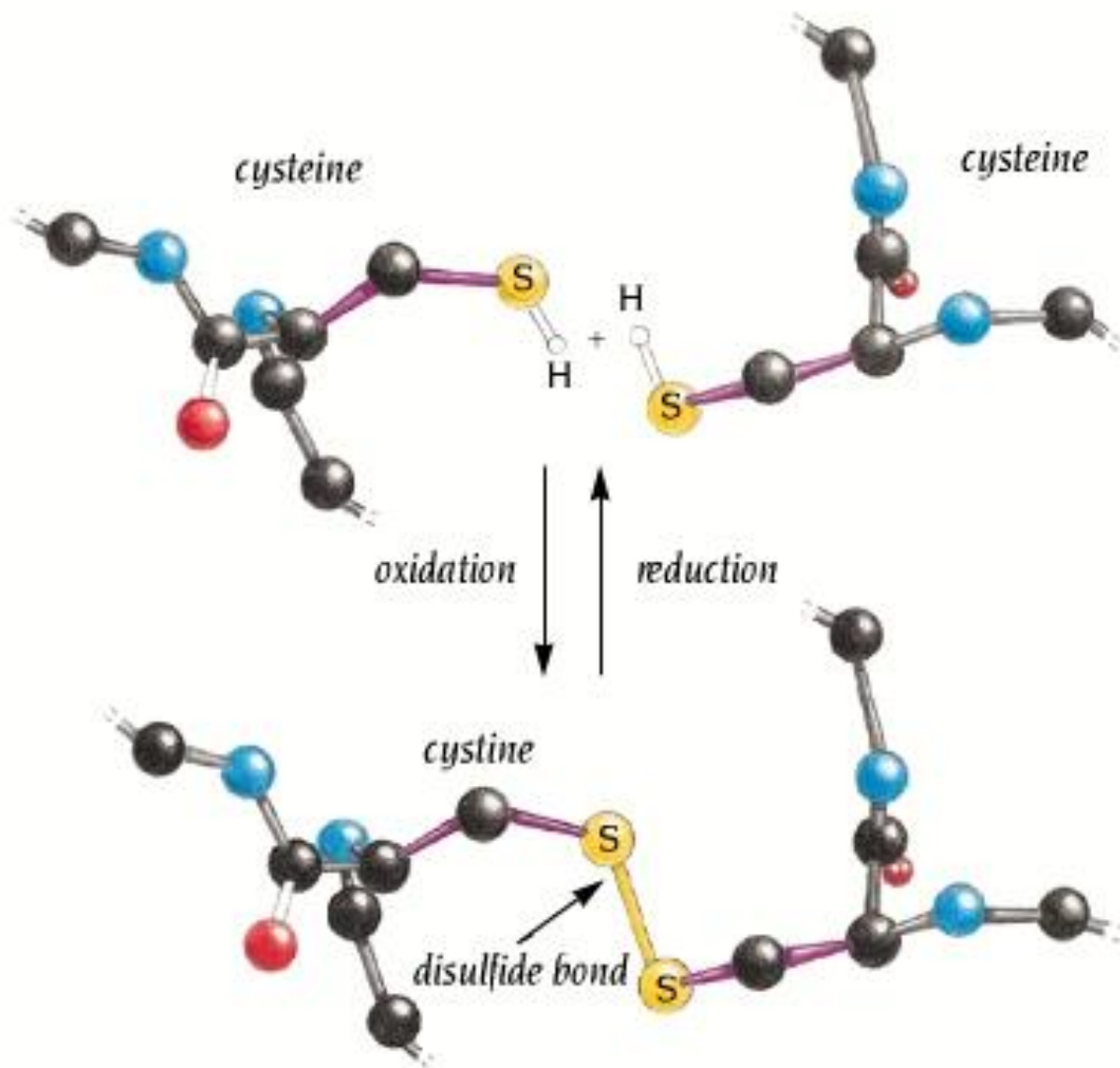
1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.
3. Can we predict protein structures without experimentally solving them?
4. Can we design artificial proteins with unique functions to solve some of our problems?

The 3D structure of a protein is encoded in its primary sequence: Anfinsen's Experiment

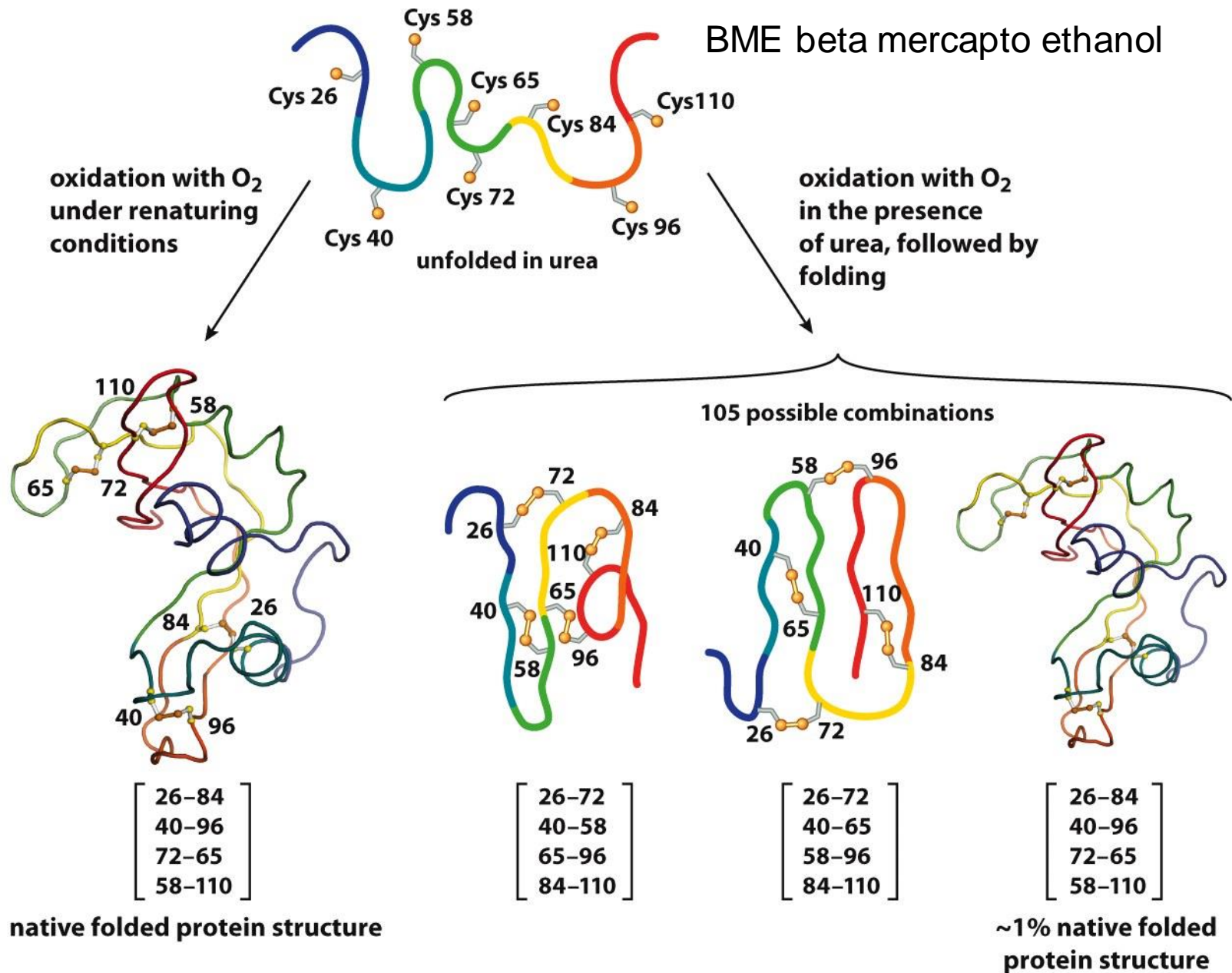
Thermodynamic hypothesis of Protein Folding: The interactions between the atoms in a protein control the folding of the protein molecule into a well-defined three-dimensional structure.

In other words, the protein sequence contains enough information required for the proper folding of the protein into its functional three-dimensional structure.

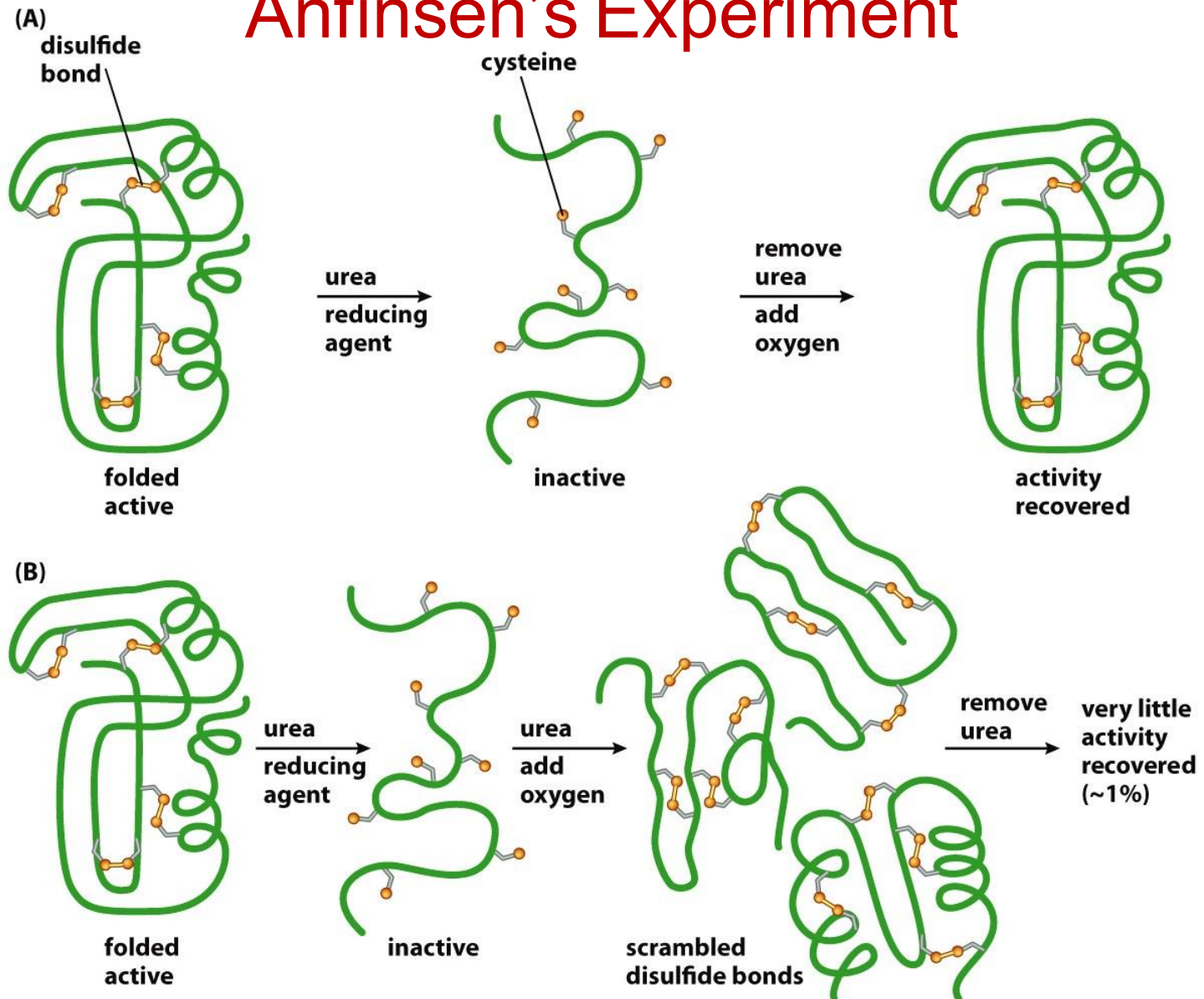




Anfinsen's Experiment



Anfinsen's Experiment



If we understand HOW PROTEINS FOLD, we can predict their structure from sequence! Then we can design proteins with novel functions.

8 Cys

Select two at a time to form a disulphide bond

$${}^8C_2 * {}^6C_2 * {}^4C_2 * {}^2C_2 / 4! = 2520/4! = 105$$



of ways to choose the
first disulfide bond



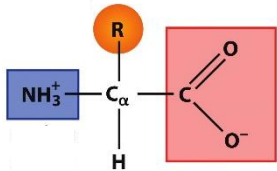
It does not matter in which order
the FOUR disulfide bonds are
formed. # of permutations of
FOUR disulfide bonds.

Important questions on Protein Folding

3. Can we predict protein structure from sequence

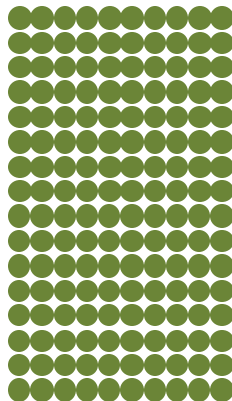
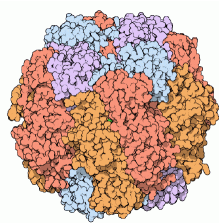
- Anfinsen's experiment demonstrates that a protein sequence encodes its structure.
- Can we decipher this code? I.e. can we predict the structure of a protein from its primary sequence?

Sequence $\xrightarrow{\text{Prediction}}$ Structure

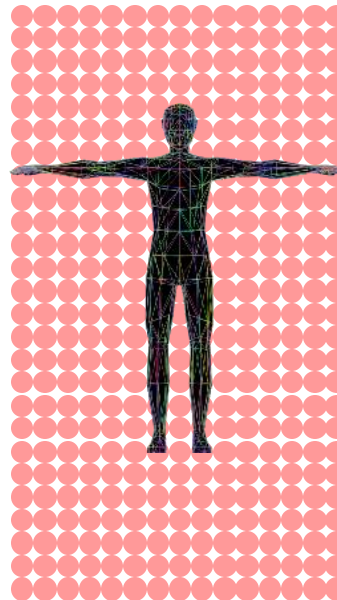


A	C	D	E
F	G	H	I
K	L	M	N
P	Q	R	S
T	V	W	Y

20 Amino Acids



300 AA/Protein



21,000 Proteins
in Human body



200,000,000 Proteins
in the world

Important questions on Protein Folding

3. Can we predict protein structure from sequence

- Anfinsen's experiment demonstrates that a protein sequence encodes its structure.
- Can we decipher this code? I.e. can we predict the structure of a protein from its primary sequence?



NEWS • 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

 Microsoft Azure

[Overview](#) [Solutions](#) [Products](#) [Documentation](#) [Pricing](#) [Training](#) [Marketplace](#) [Partners](#)

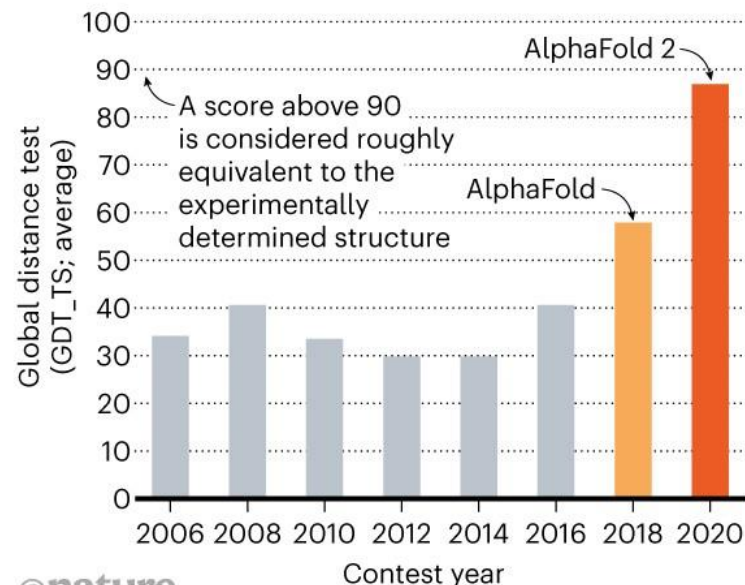
[Blog](#) / [Cloud Strategy](#)

Windows Azure Helps Scientists Unfold Protein Mystery and Fight Disease

Posted on 16 June, 2011

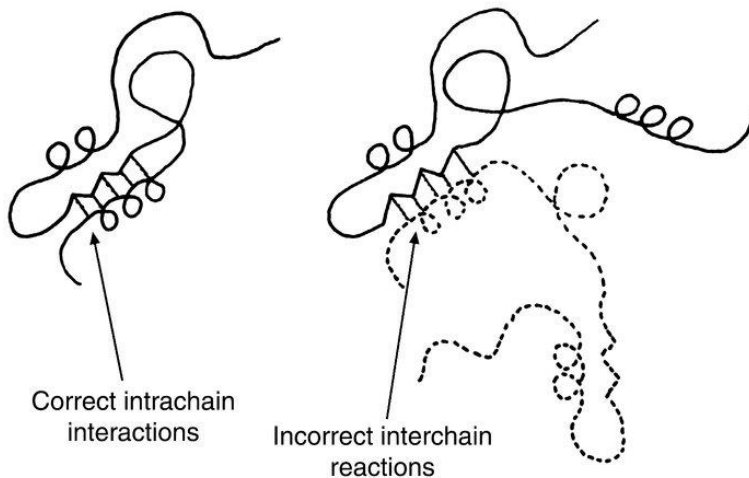
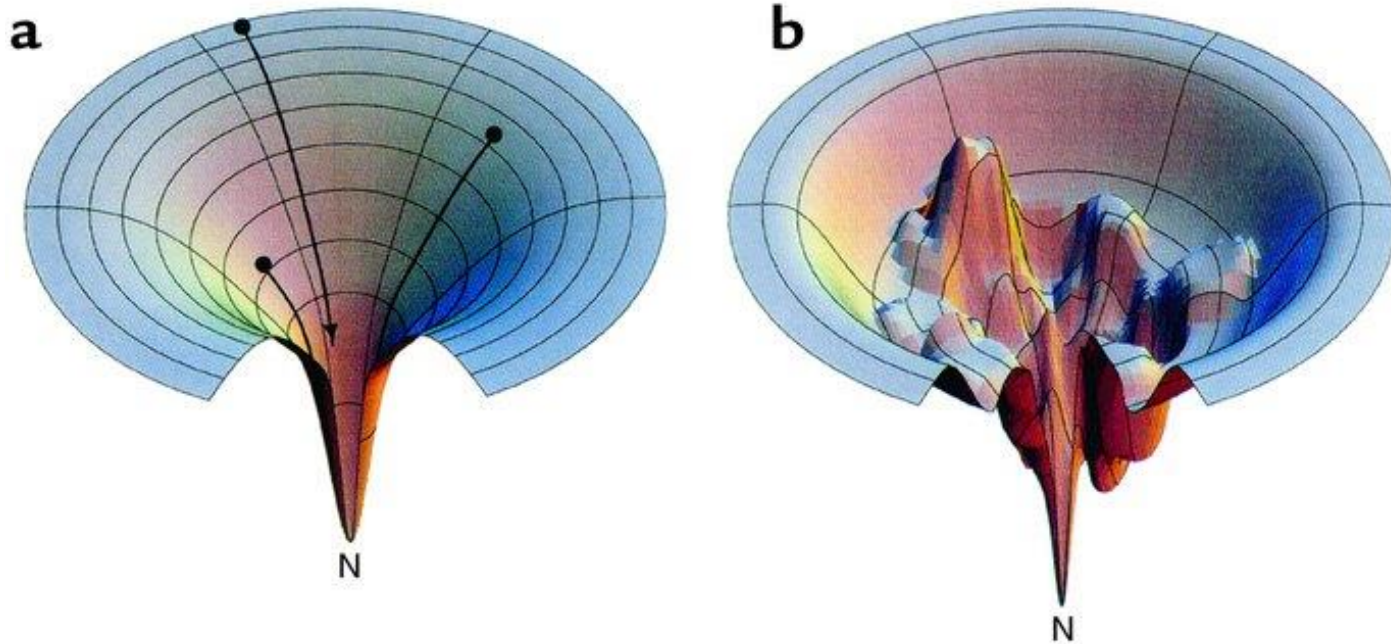
STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



How Proteins Fold



Important interactions between amino acids:

- Hydrophobic interactions
- Hydrogen bonding interactions
- Electrostatic interactions

How Proteins Fold

1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
 2. How do proteins fold so fast? Most proteins fold within milliseconds.
- Protein Folding is a stochastic process i.e. not all conformations are sampled.
 - Proteins fold in small segments (~20 amino acids) independent of the rest.

These folded segments or FOLDONS collapse to give the final structure.

Molecular Dynamics (MD) Simulations use physical laws to study protein folding

Animation of Protein Folding Funnel

<https://www.youtube.com/watch?v=YANAso8Jxrk>

The protein folding game - Foldit

<https://fold.it/portal/>

**foldit**BETA

Solve Puzzles
for Science

03:53:48 GMT

PUZZLES  **CATEGORIES** **GROUPS** **PLAYERS** **RECIPES** **CONTESTS**
BLOG  **FEEDBACK** **FORUM** **WIKI** **FAQ** **ABOUT** **CREDITS**



Click to learn how you
contribute to science by
playing Foldit.

**NANOCRAFTER** Try our new scientific discovery game!
Be creative and build extraordinary tiny machines!

What's New

Super Snowflakes

The results are in. Thanks to everyone who submitted a design, and even those who put in more than one!

GET STARTED: DOWNLOAD

**Win Beta**
Windows
(XP/Vista/7/8)

**Mac Beta**
OSX
(10.7 or later)

**Linux Beta**
Linux
(64-bit)

Are you new to Foldit? Click here.

Are you an educator? Click here.

SEARCH

☒ Only search fold.it

RECOMMEND FOLDIT

USER LOGIN

Username: *

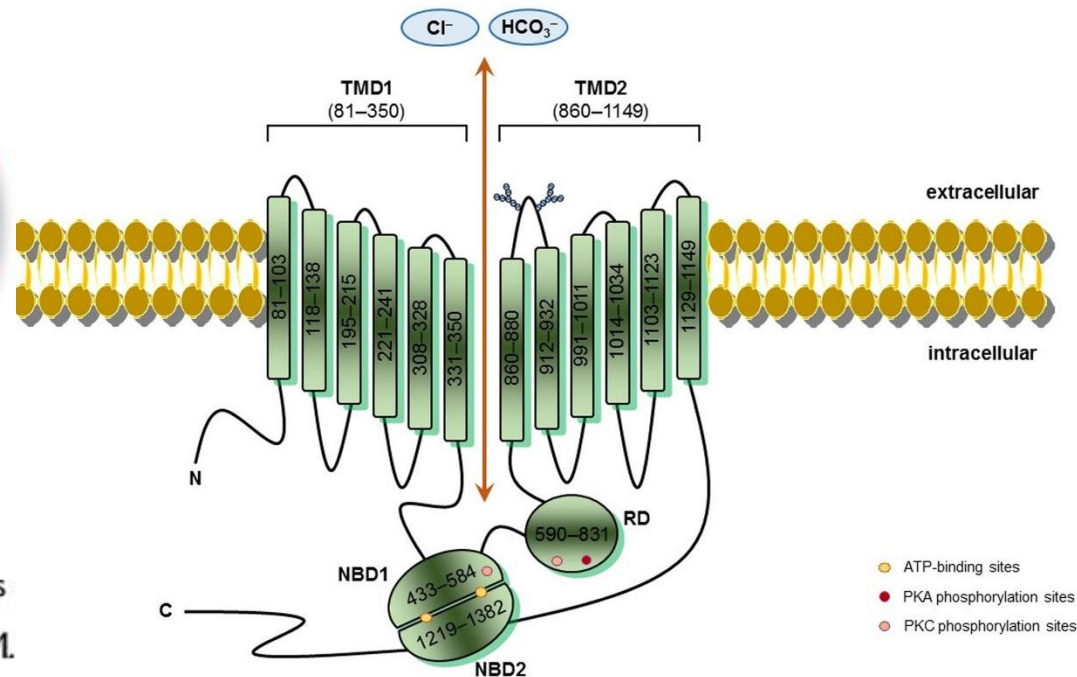
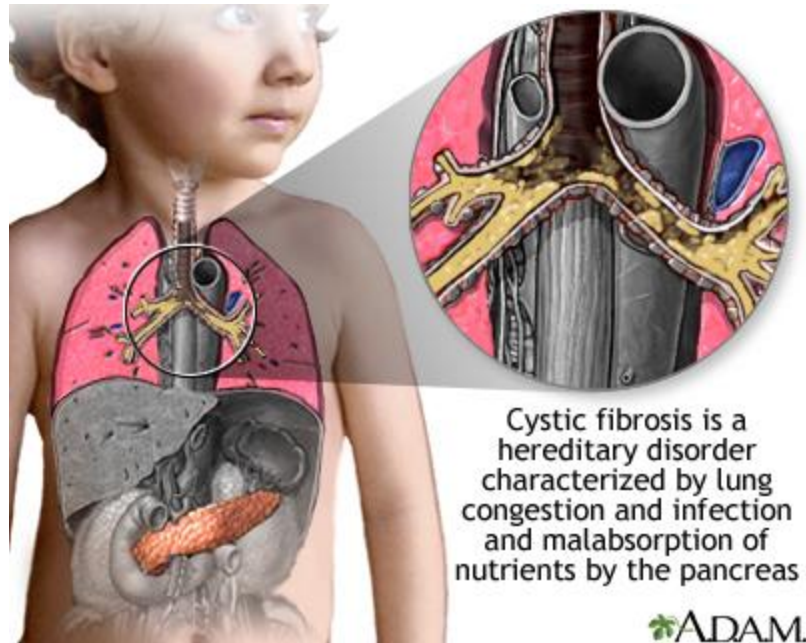
Password: *

• [Create new account](#)

Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

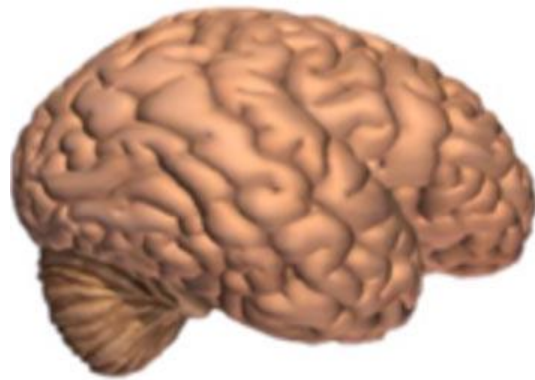
- 1) **Cystic Fibrosis:** It results from the misfolding of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), a protein functioning as a chloride (Cl^-) ion channel. The loss of CFTR function interferes with the body's ability to efficiently secrete fluids and salts. It damages the lungs and digestive systems.



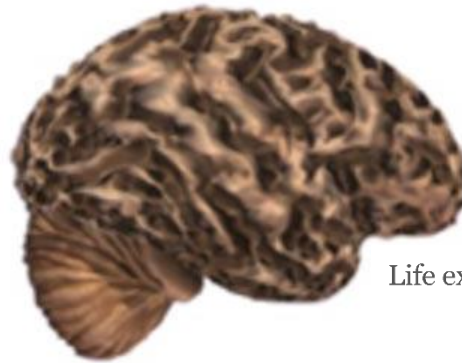
Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

2) Alzheimer's Disease (AD): It results from the aggregation and precipitation of a peptide called amyloid- β ($A\beta$). AD is the most common form of progressive dementia in the elderly, and of neuro-degenerative diseases in general.



Normal Brain

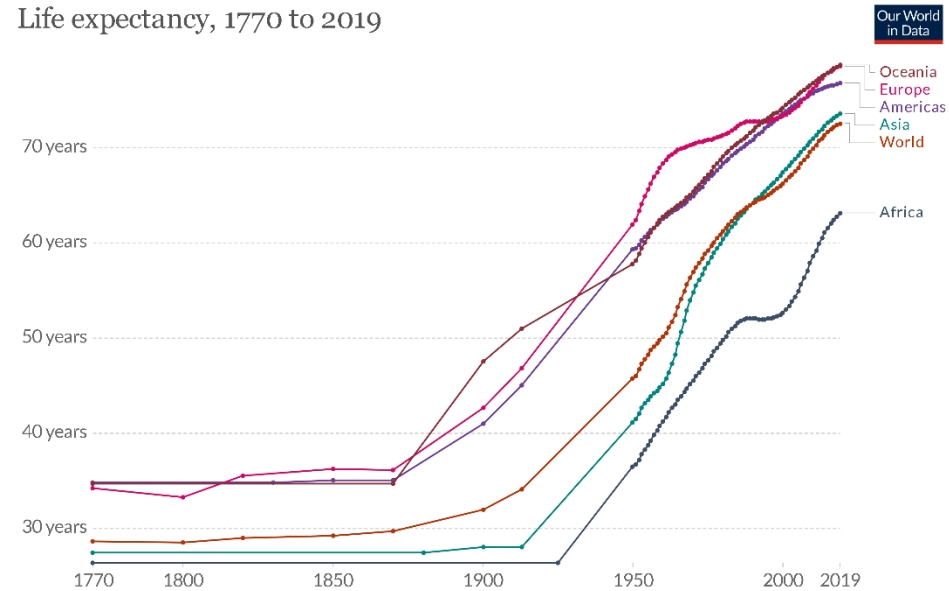


Brain with AD



Comparison of the
two Brains

Life expectancy, 1770 to 2019



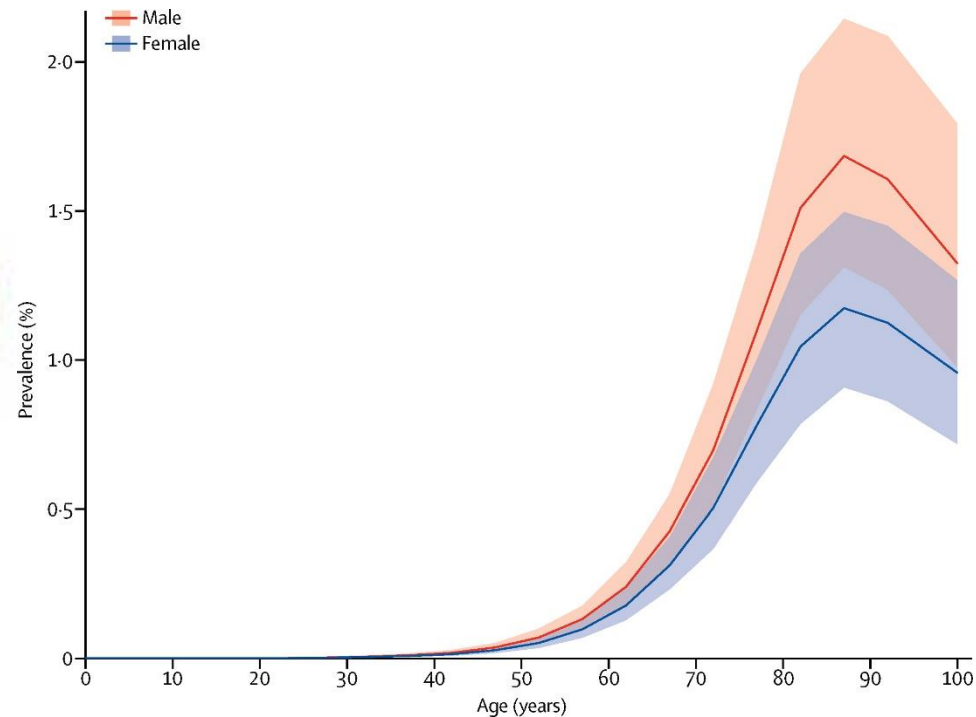
Source: Riley (2005), Clio Infra (2015), and UN Population Division (2019)
Note: Shown is period life expectancy at birth, the average number of years a newborn would live if the pattern of mortality in the given year were to stay the same throughout its life.

OurWorldInData.org/life-expectancy • CC BY

Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

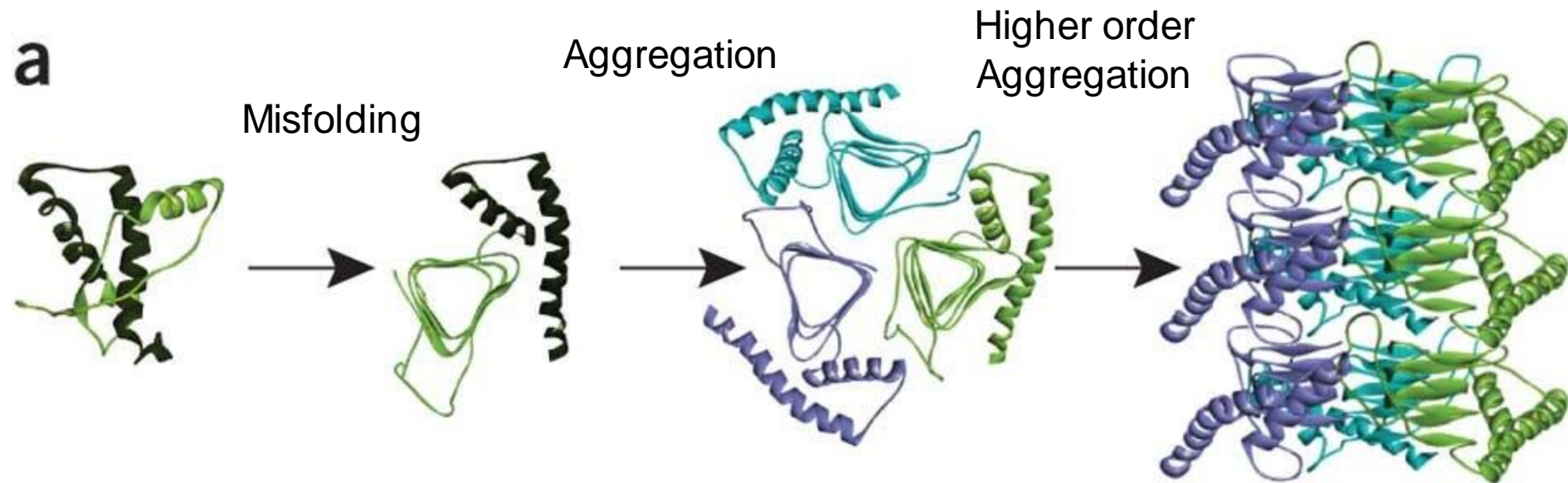
3) Parkinson's Disease (PD): It results from aggregation and precipitation of the protein α -synuclein. PD is a motor disorder common among the elderly (but can also hurt young people). It leads to shaking, stiffness, and difficulty with walking, balance, and coordination.



Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

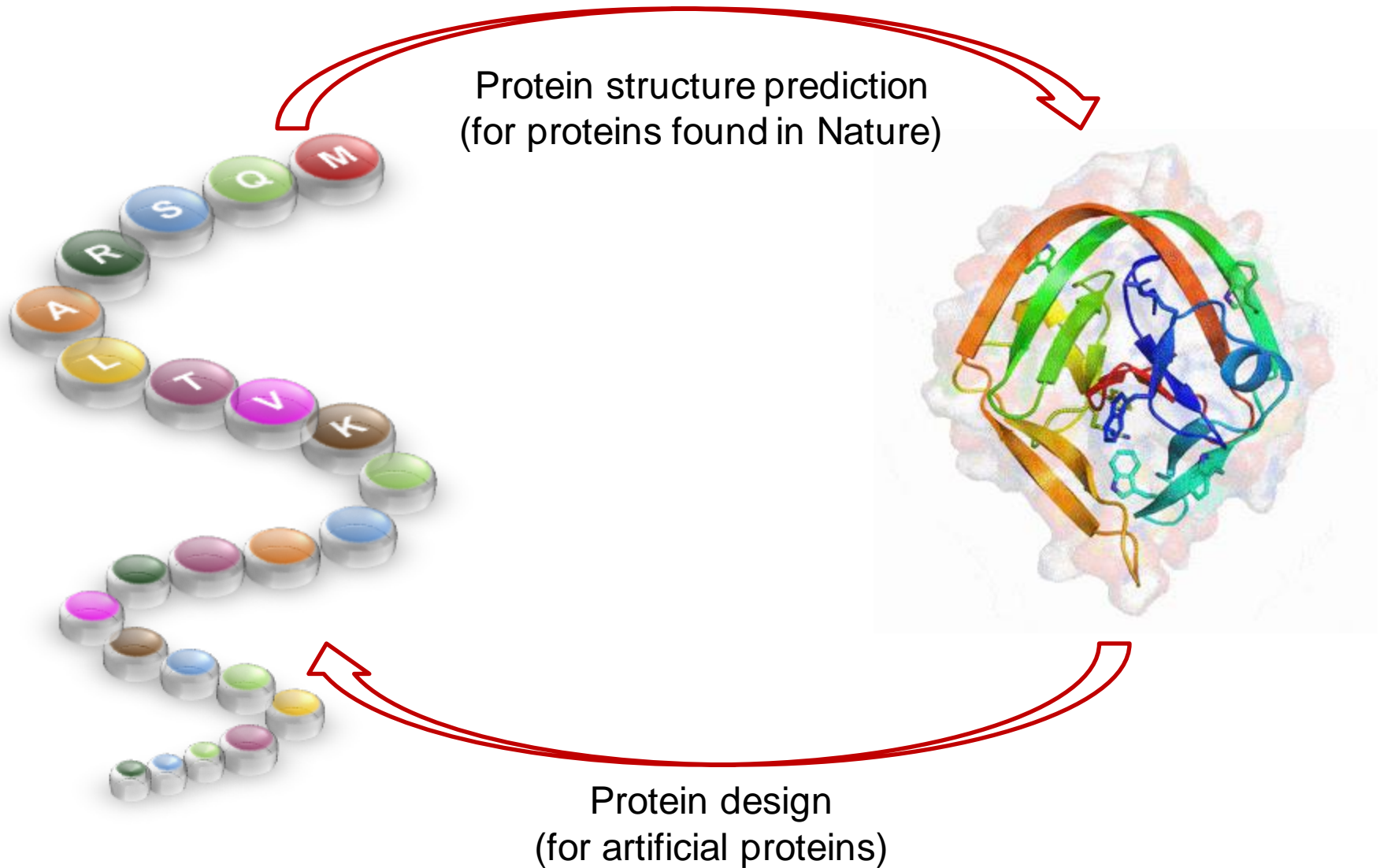
- 4) **Creutzfeldt-Jakob Disease (CJD)**: It is caused by the aggregation and precipitation of the protein prion. It results in progressive motor dysfunction, cognitive impairment, and cerebral ataxia.



Important Questions on Protein Folding

1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.
3. Can we predict protein structures without experimentally solving them?
4. Can we design artificial proteins with unique functions to solve some of our problems?

Protein Design Problem



Influenza A virus

NA
M2
HA

Adhesion

Endocytosis

Fusion and
uncoating

Budding

Assembly

Release

NA inhibitors

Protein
synthesis

RNP formation

mRNA

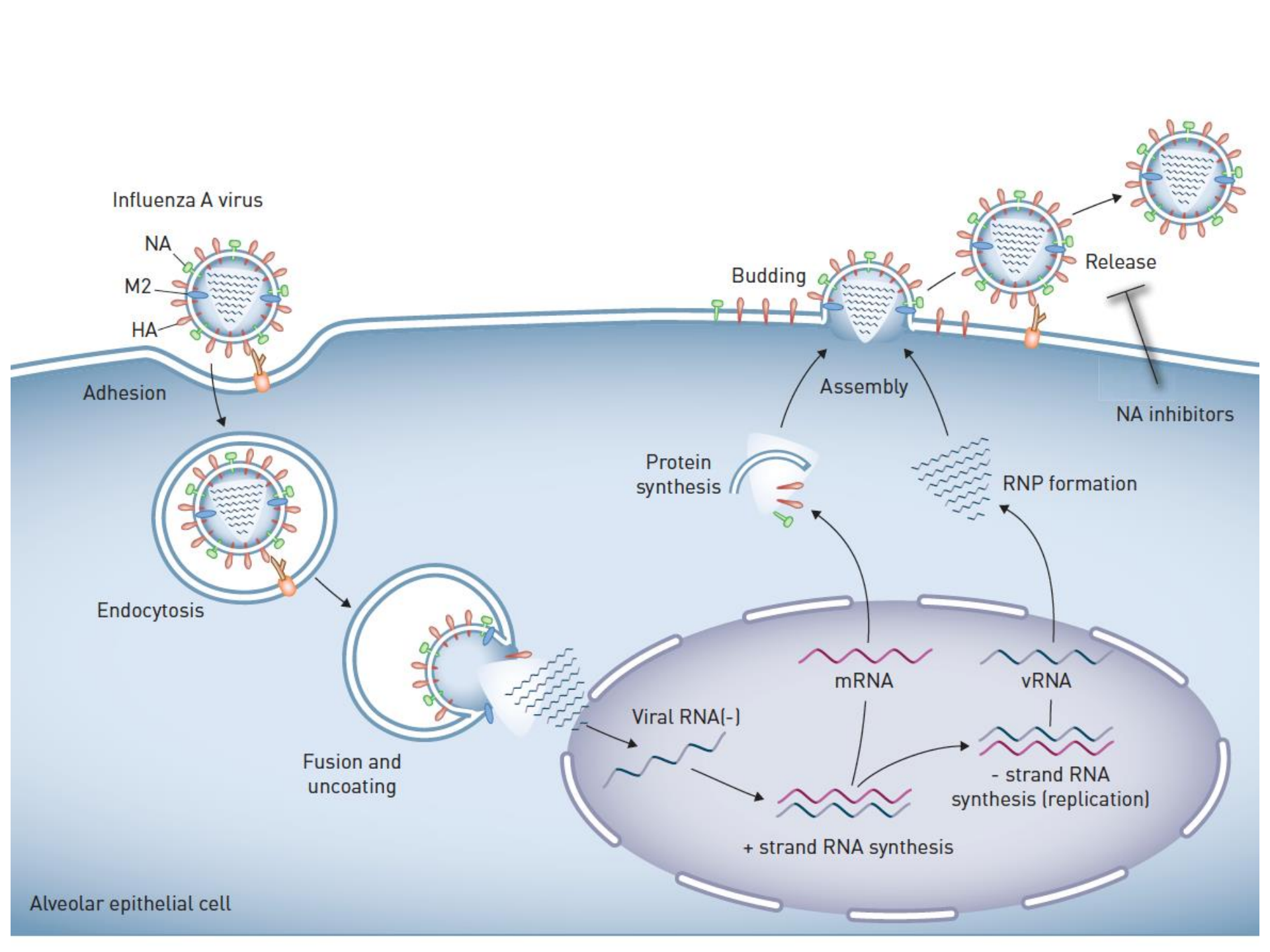
vRNA

Viral RNA(-)

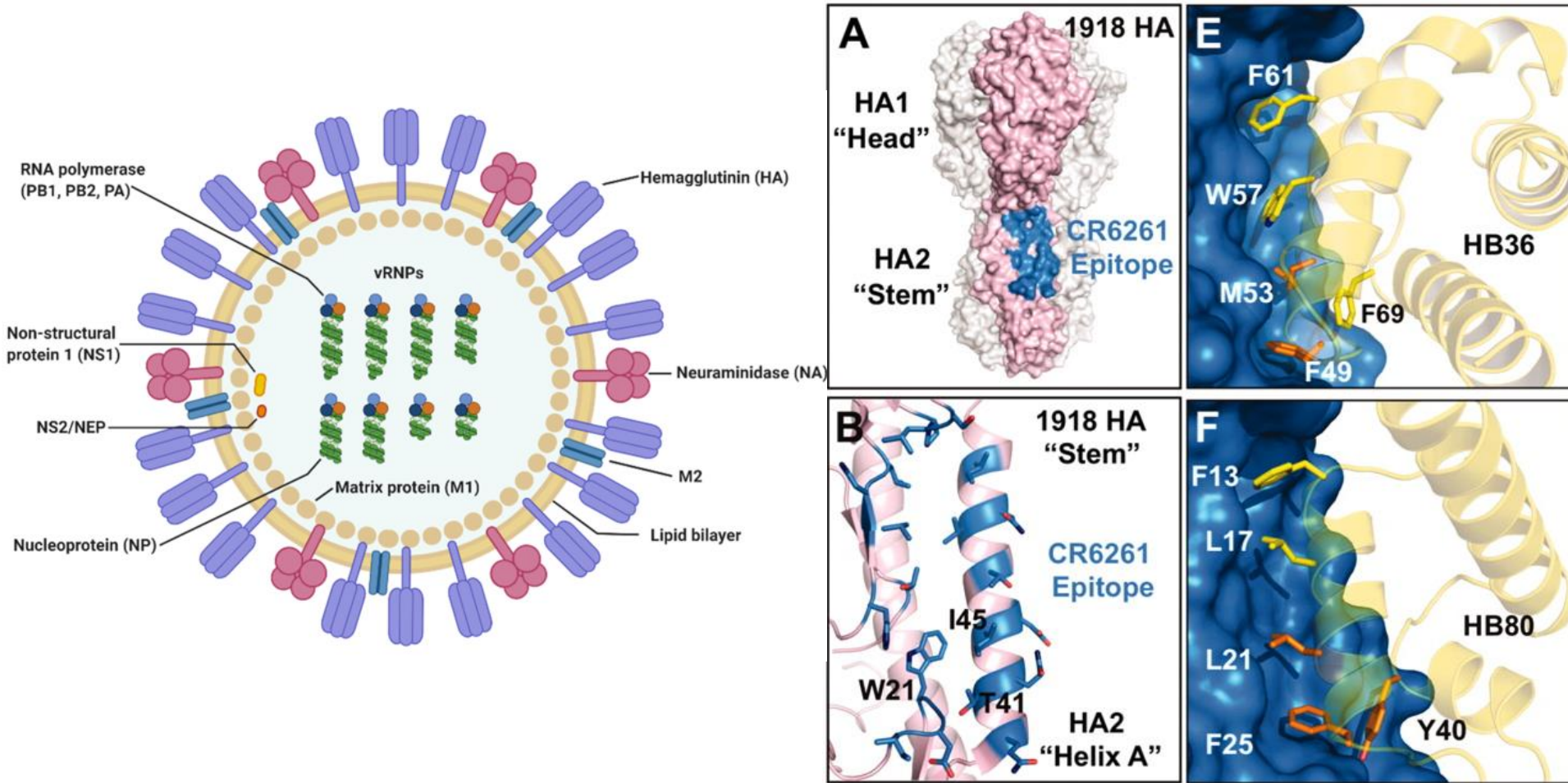
- strand RNA
synthesis (replication)

+ strand RNA synthesis

Alveolar epithelial cell



Designed Protein Targets the Conserved Stem Region of Influenza Hemagglutinin



Hierarchy of Protein Structure

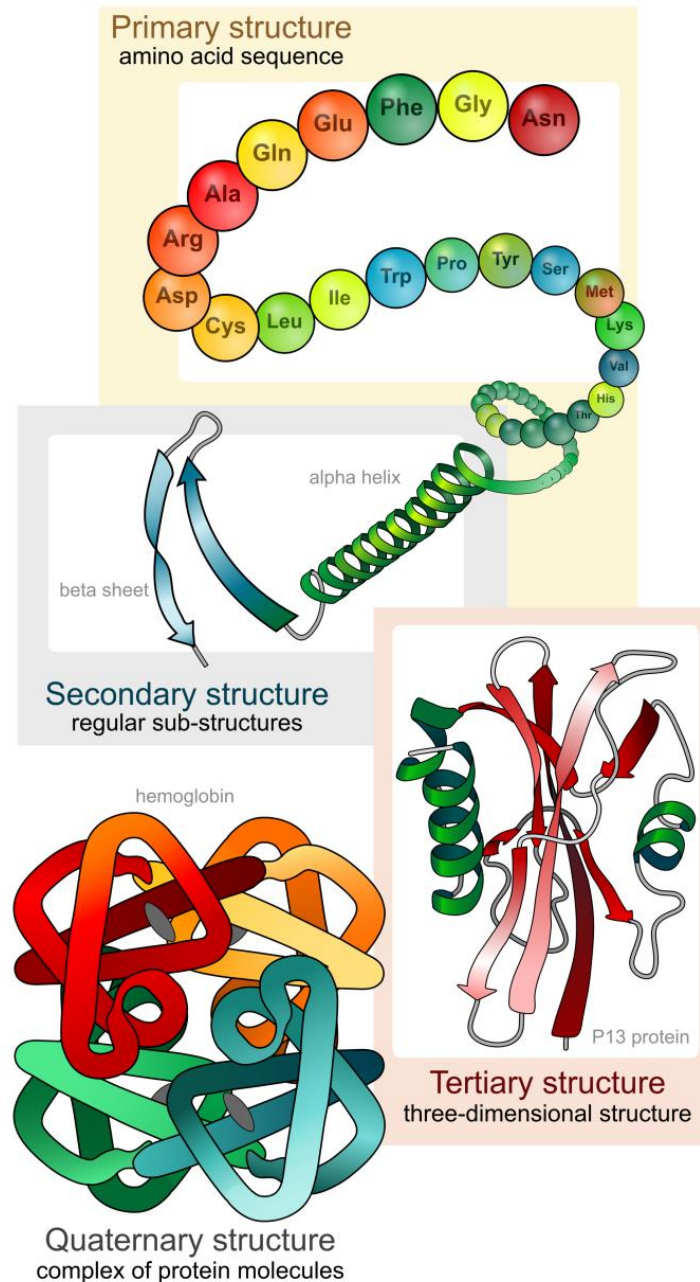
Protein Molecules are Organized in a Structural Hierarchy

Secondary

Primary

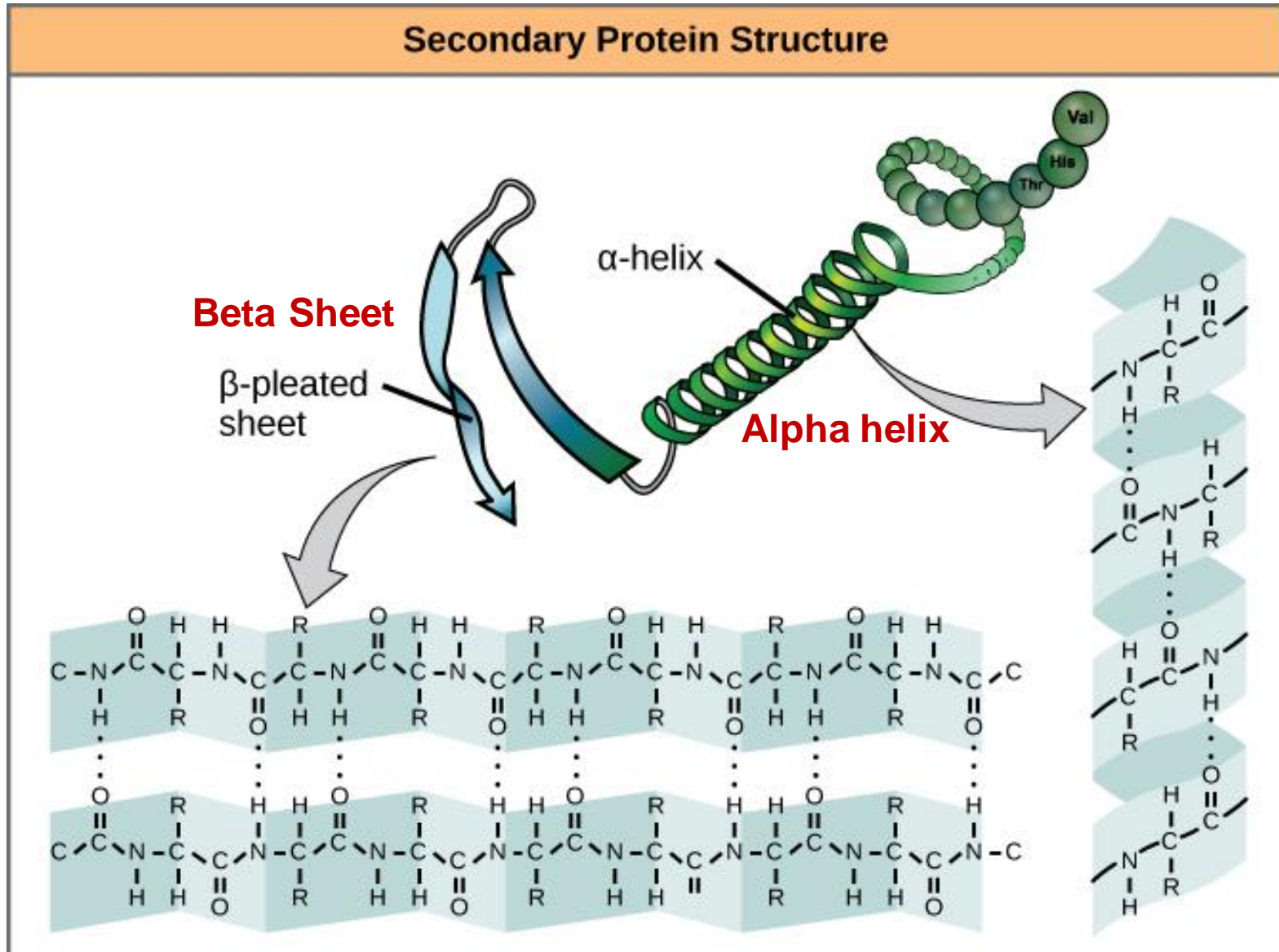
Quaternary

Tertiary

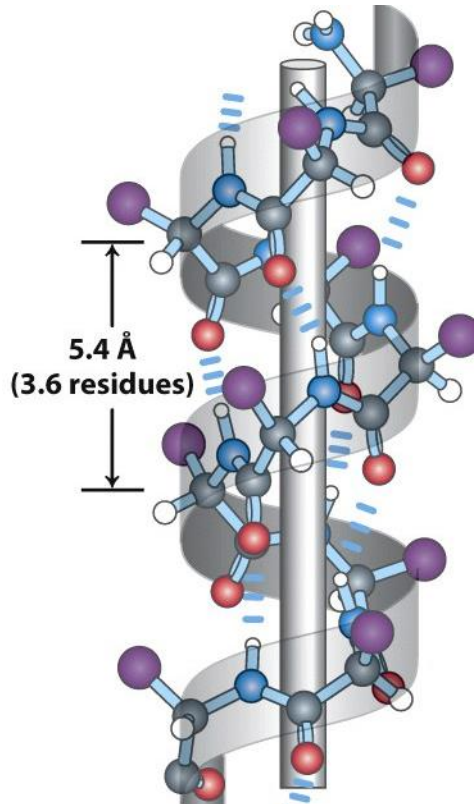


Secondary Protein Structure

Characterized by main chain NH and CO groups participating in H-bonds



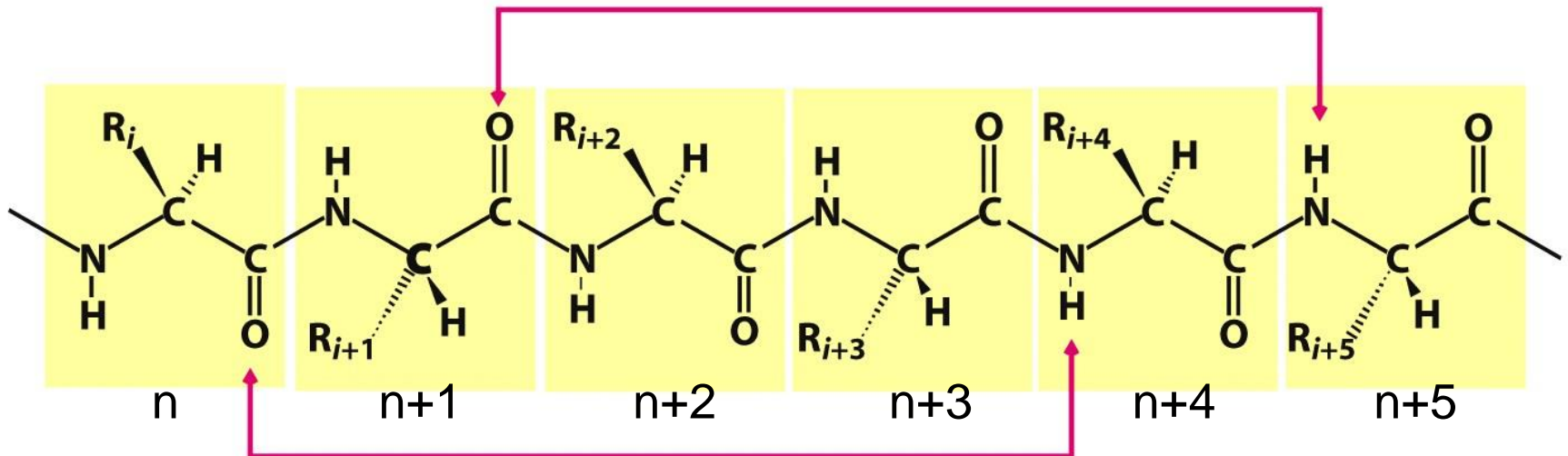
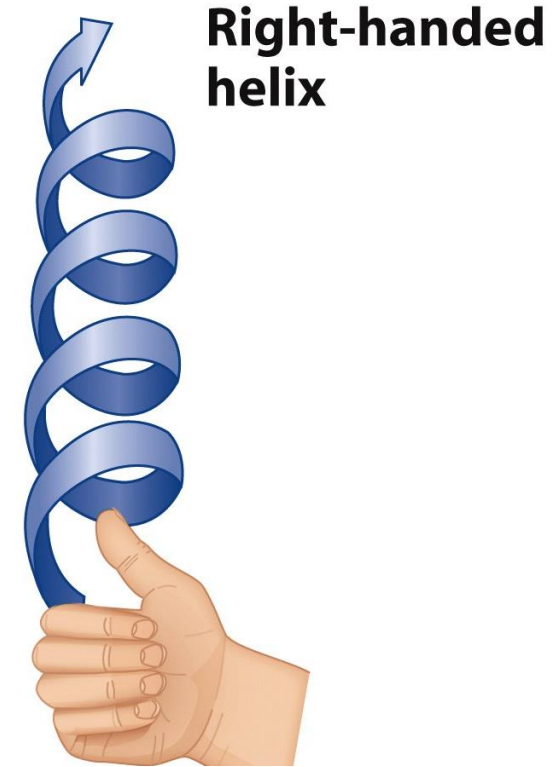
Alpha Helix



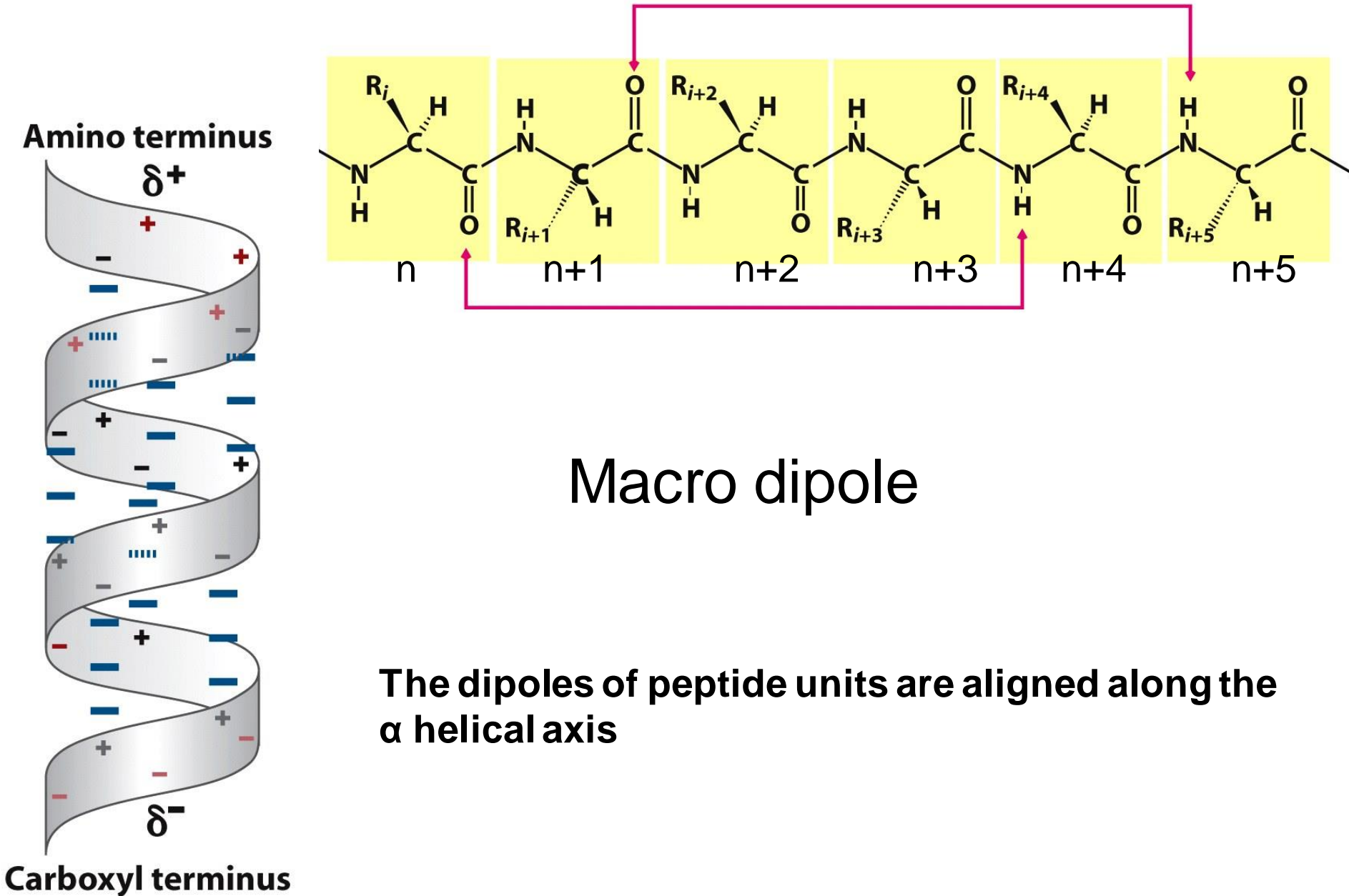
Every 3.6 residues make one turn

The distance (pitch of helix) between two turns is 5.4 Å

The C=O of residue 'n' is hydrogen bonded to N-H of residue 'n+4'



The Alpha-Helix has a Dipole Moment

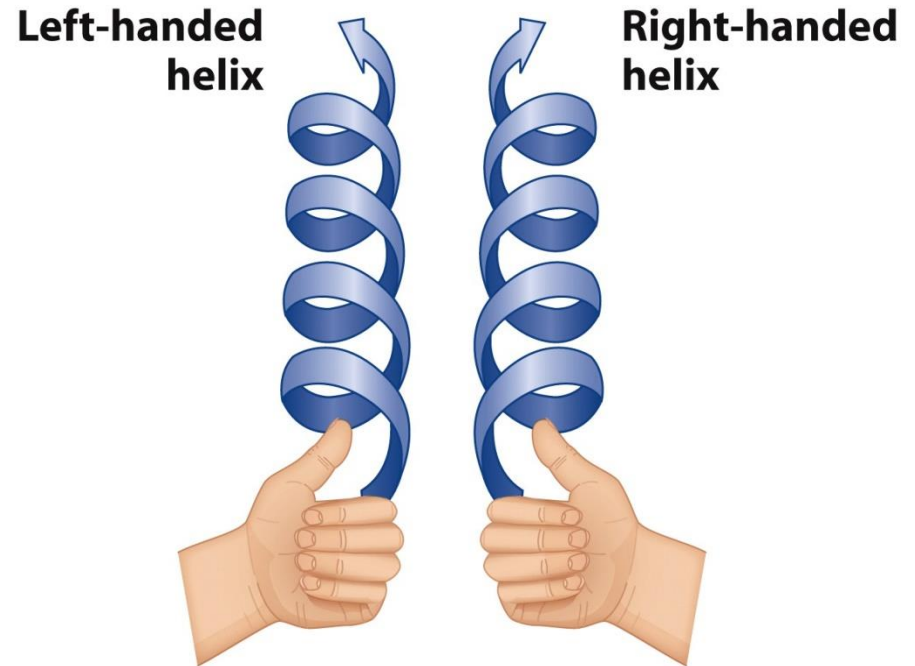


Alpha Helix: Right-handed or Left-handed?

Alpha helix can be – Right-handed or Left handed

BUT, left handed helix is not possible for L-amino acids due to close approach of the side chains and CO group.

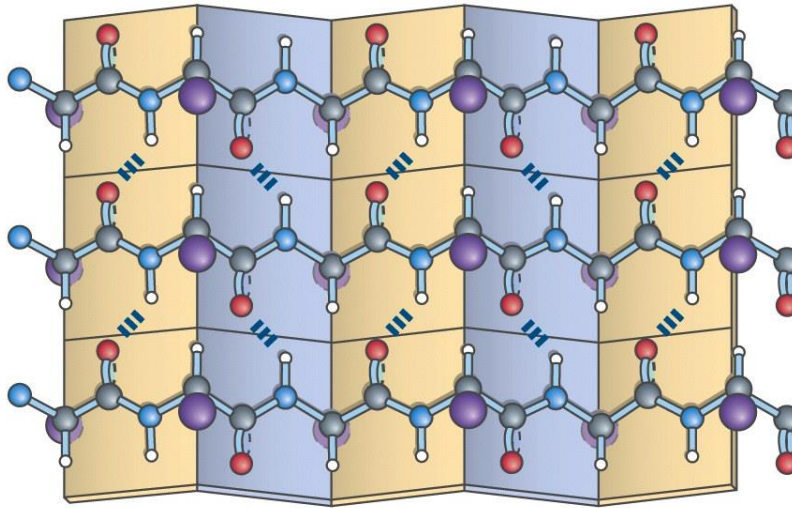
Right handed – most commonly observed in proteins.



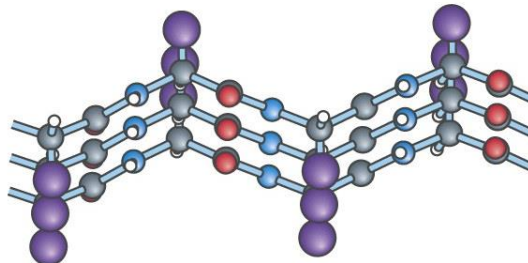
β -sheet

(Number of β -Strands are Involved)

Top view



Side view



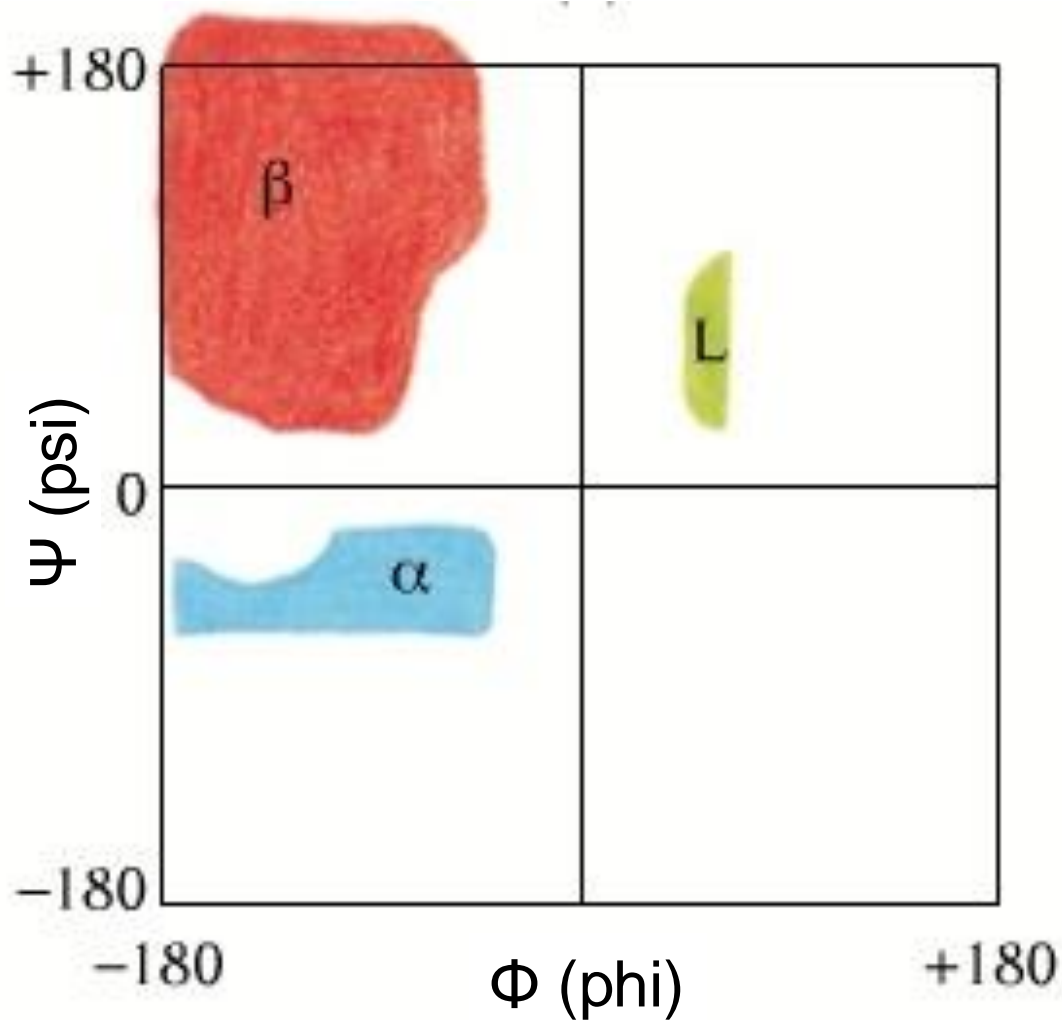
β -sheet from several regions of the chain; Each β -strand, typically 5-10 residues long

H-bonds are perpendicular to strands

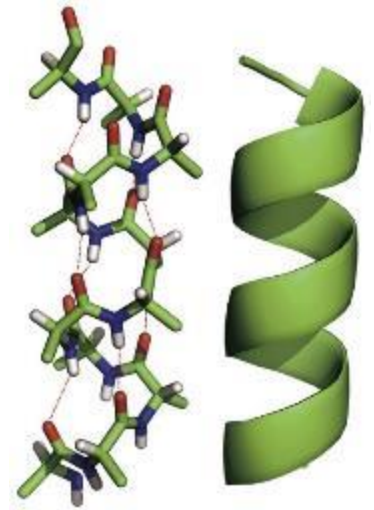


α -helix: from one continuous region

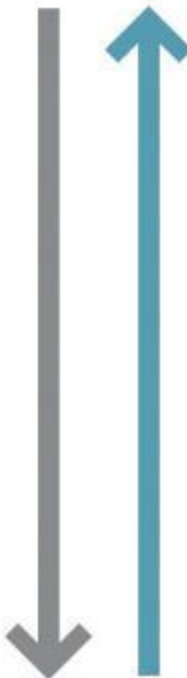
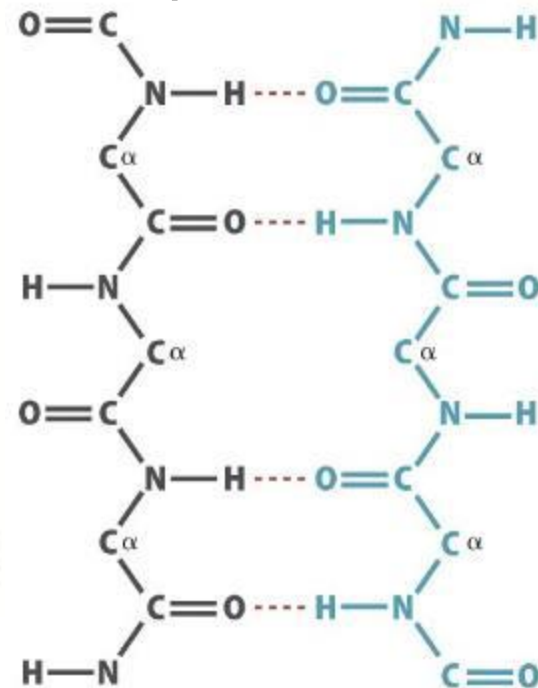
Ramachandran Plot



α -Helix



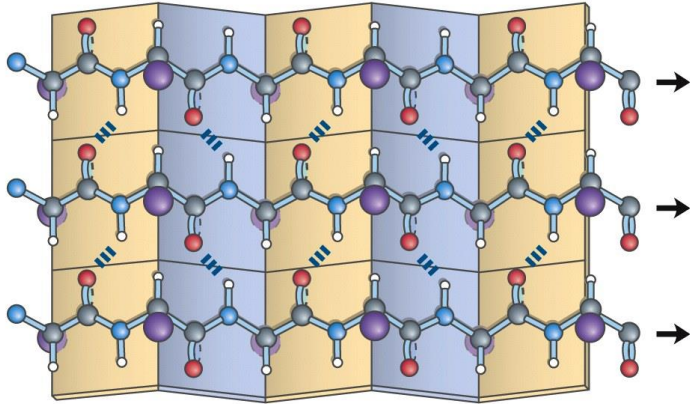
β -Strand



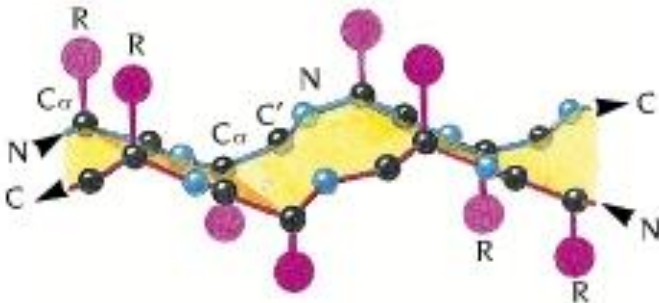
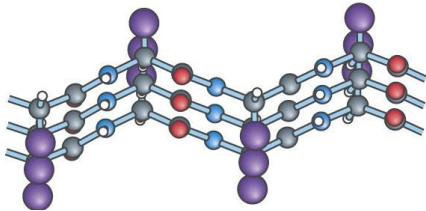
Parallel and Antiparallel β -sheet

Parallel

Top view

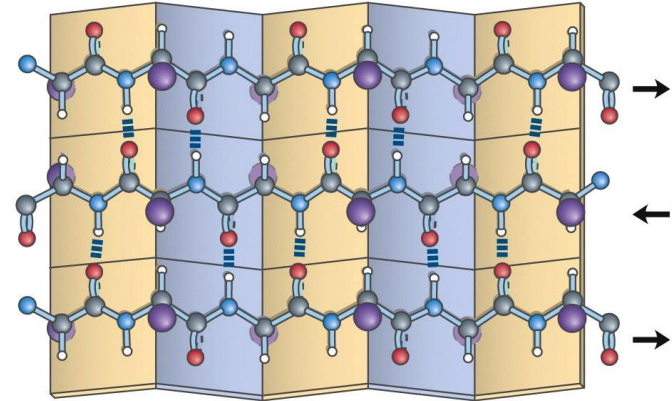


Side view

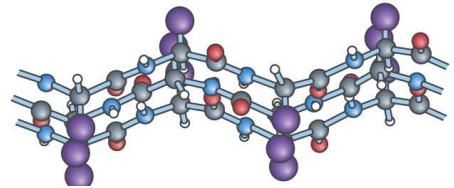


Antiparallel

Top view

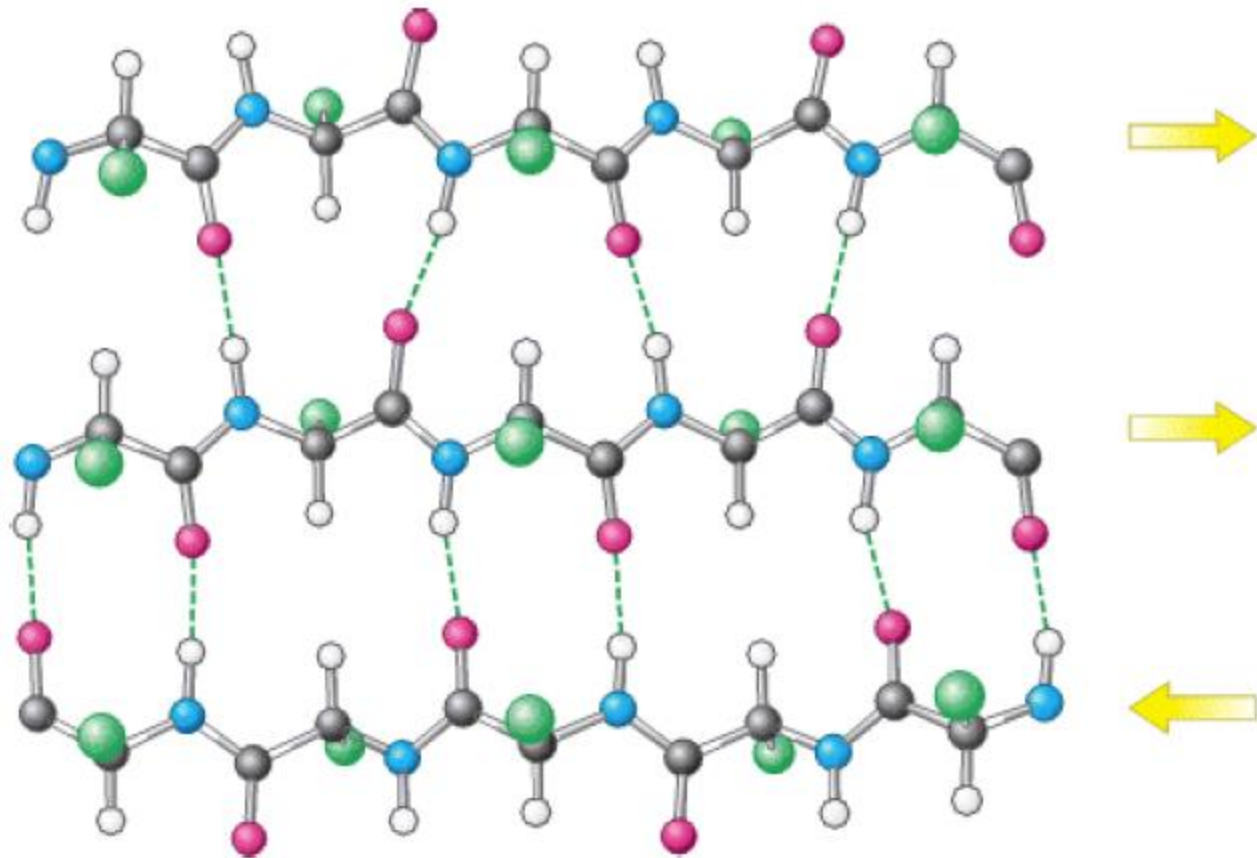


Side view



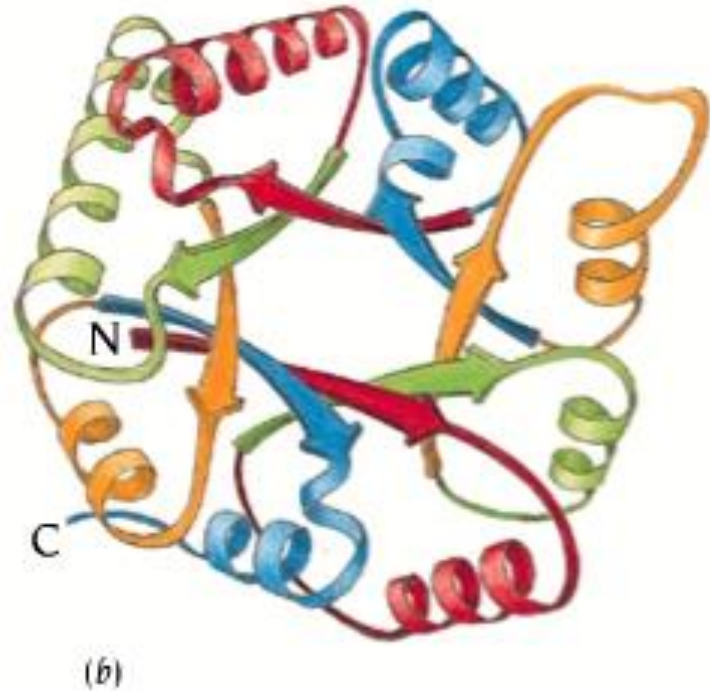
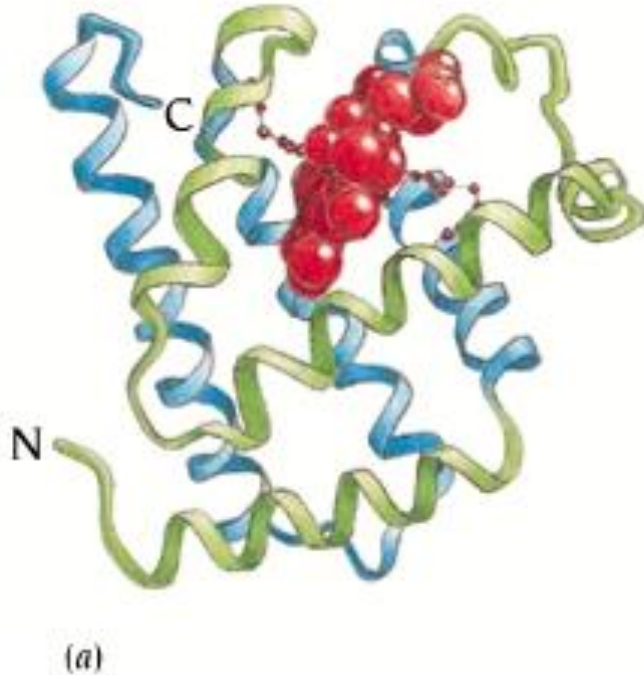
β -pleated sheet: 'pleated' because side chains point up and down alternatively

Mixed β -sheet

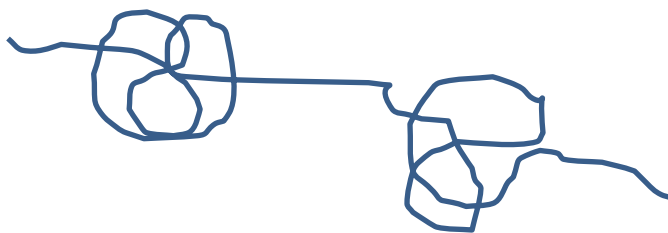
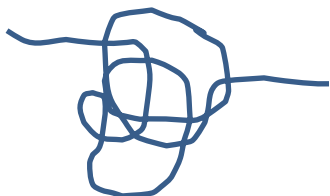


Polypeptide Chains Fold into Several Domains

- Fundamental unit of **tertiary structure** – DOMAIN
- **Domain**: polypeptide chain or a part of polypeptide chain that can independently fold into a stable tertiary structure
- Domains are also units of function



Tertiary structure refers to the spatial arrangement of amino acid residues that are far apart in the sequence and to the pattern of disulfide bonds.



Quaternary Structure

Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a **subunit**. **Quaternary structure** refers to the spatial arrangement of subunits and the nature of their interactions.

The simplest quaternary structure is a dimer, consisting of two identical subunits.

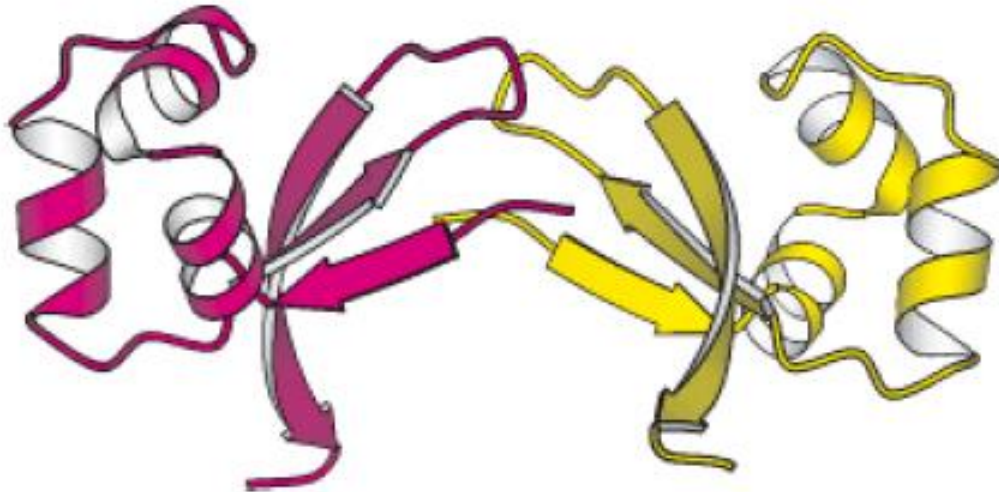
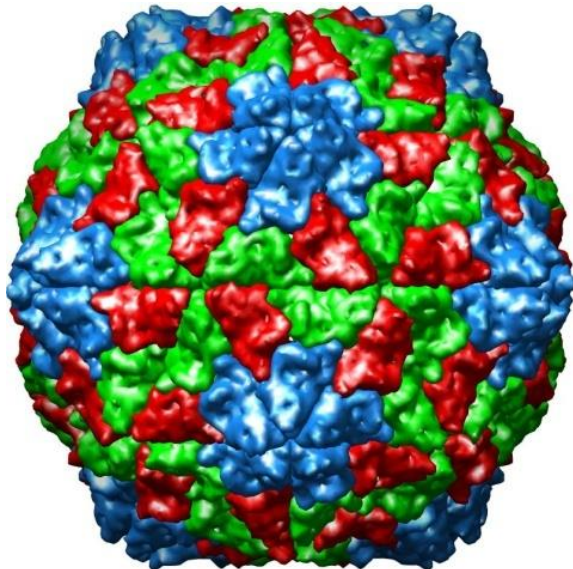


Figure 3.48. Quaternary Structure. The Cro protein of bacteriophage λ is a dimer of identical subunits.

Quaternary Structure (higher order)



The $\alpha_2\beta_2$ tetramer of human haemoglobin. The structure of the two identical α subunits (red) is similar to but not identical with that of the two β subunits (yellow).



Complex Quaternary Structure. The coat of rhinovirus comprises 60 copies of each subunits

Methods to study protein structures

X-ray crystallography

- Prof. Amit K Das

(<http://www.iitkgp.ac.in/departement/BT/faculty/bt-amitk>)

Structure-guided protein engineering

- Prof. Dibyendu Samanta

(<http://iitkgp.ac.in/departement/BS/faculty/bs-dibyendu.samanta>)

NMR spectroscopy

- Nuclear magnetic resonance (NMR) spectroscopy

- Prof. Soumya De

(<http://iitkgp.ac.in/departement/BS/faculty/bs-somde>)

Protein Structure, Function, Kinetics and Energetics

Books Followed:

- How Proteins Work (Mike Williamson)
- Introduction to protein structure (Carl Branden & John Tooze)
- Biochemistry (Lubert Stryer)