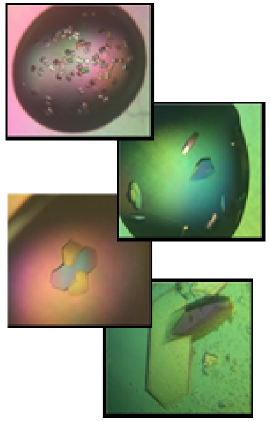
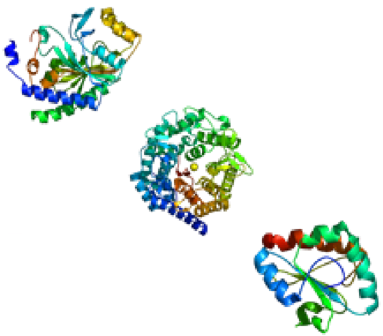
Introduction to Macromolecular Structures



Zhi-Jie Liu

Institute of Biophysics
Chinese Academy of Sciences
Beijing, China



SFU, Krasnoyarsk, March 1, 2012

Outline

1. Varieties of macromolecules
2. Macromolecular structures
3. Structure determination by X-ray crystallography
4. Structure validation and deposition.

Varieties of macromolecules

1. Proteins
2. DNA
3. RNA
4. Complexes: protein-protein, protein-DNA/RNA

Lipids, peptides, sugars, etc are categorized as non macromolecules

Our discussion is more focused on protein molecules

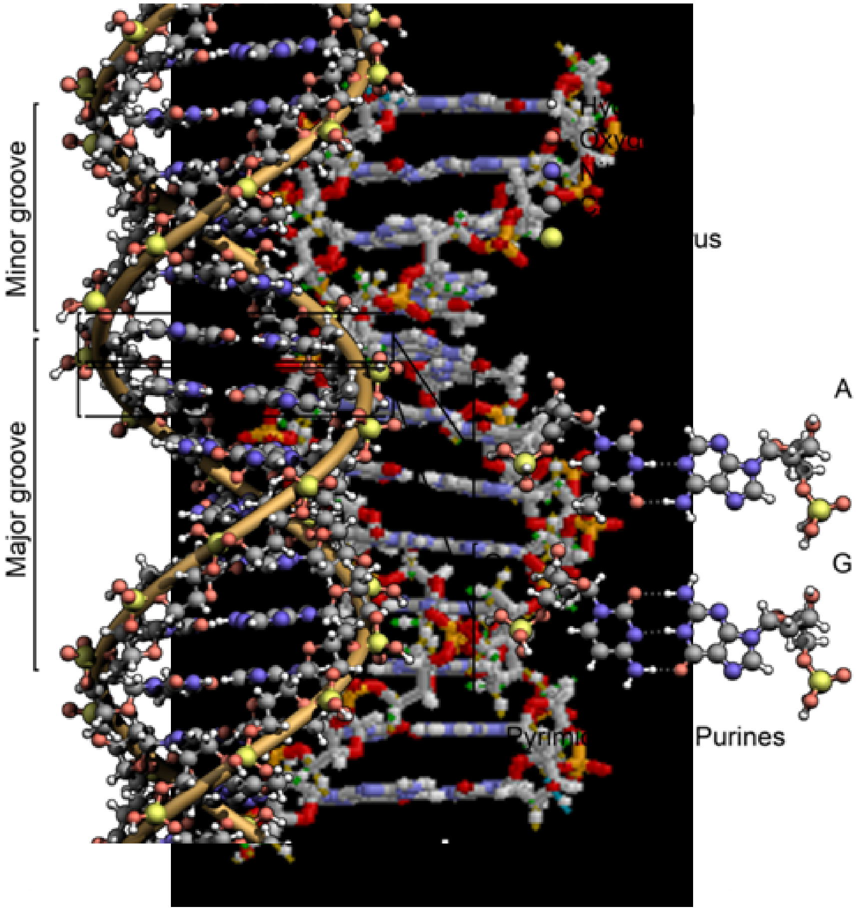
DNA/RNA

Deoxyribonucleic acid, **DNA**:

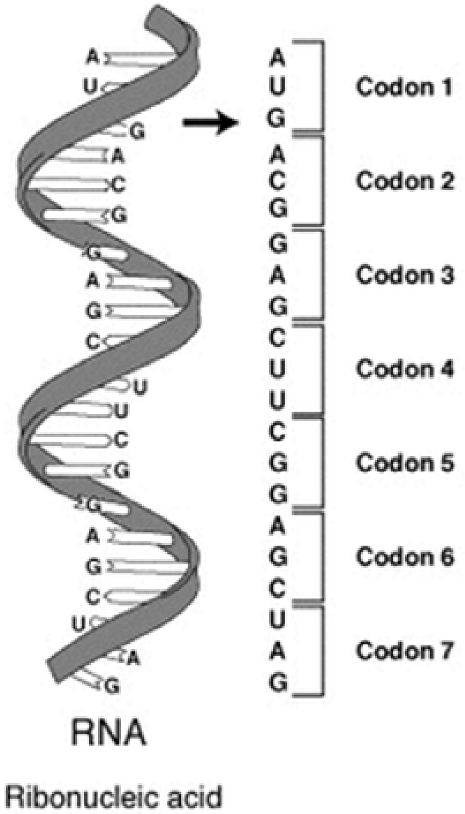
consists of two long polymers of simple units called nucleotides, **C**ytosine, **G**uanine, **A**denine and **T**hymine.

The sequence of these four bases along the backbone encodes information, or the genetic code.

RNA has the same nucleotides except that **T**hymine is replaced by **U**racil.



Genetic code



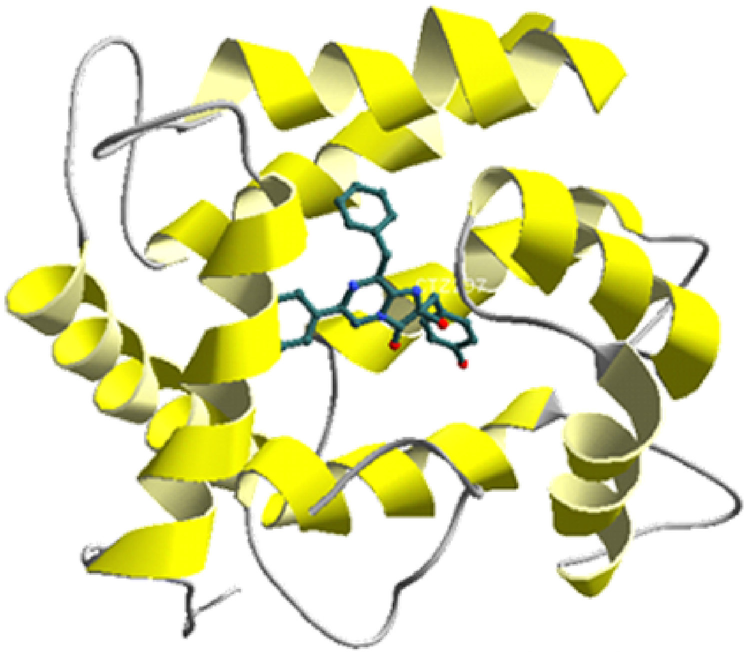
A series of codons in part of a mRNA molecule. Each codon consists of three nucleotides, usually representing a single amino acid.

Macromolecular structures

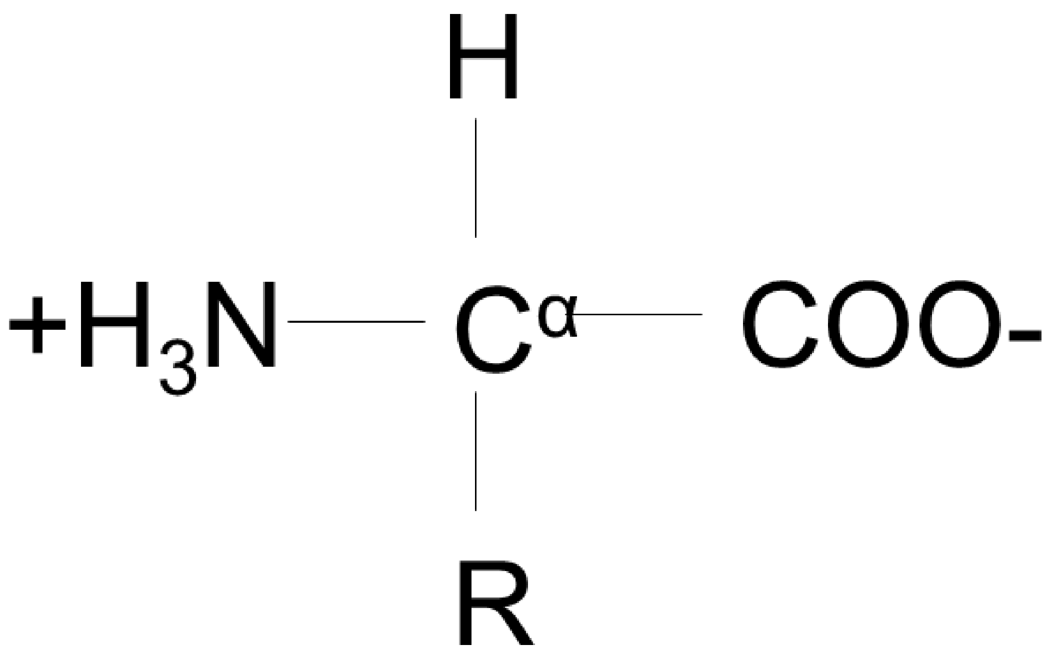
Proteins

Composed of one or more polypeptides which is a single linear polymer chain of amino acids. The sequence of amino acids in a protein is defined by the sequence of a gene, which is encoded in the genetic code.

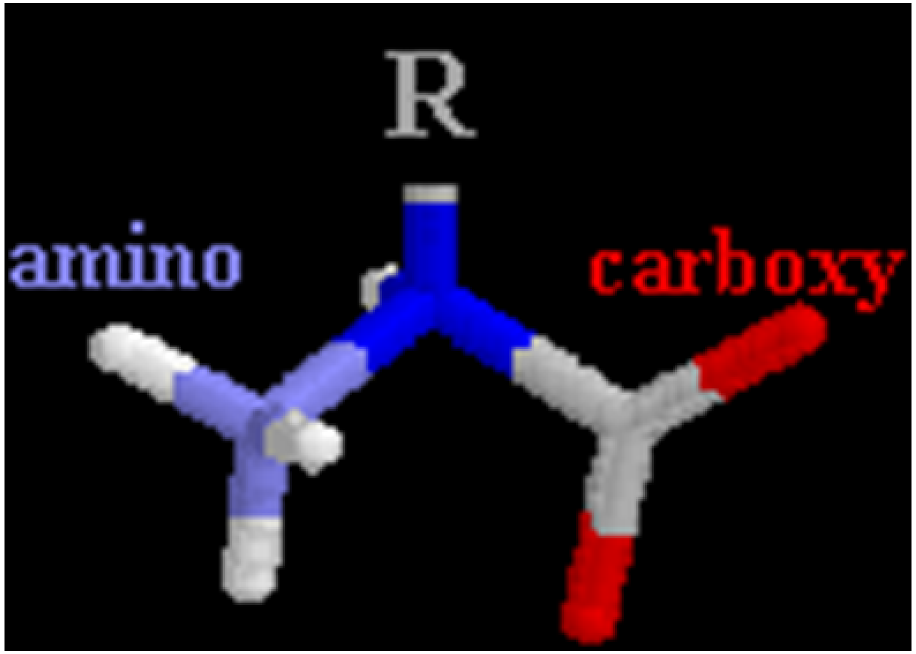
Proteins are the molecular building block of life. Protein molecules are three-dimensional, so is life.



General Amino Acid Structure At pH 7.0



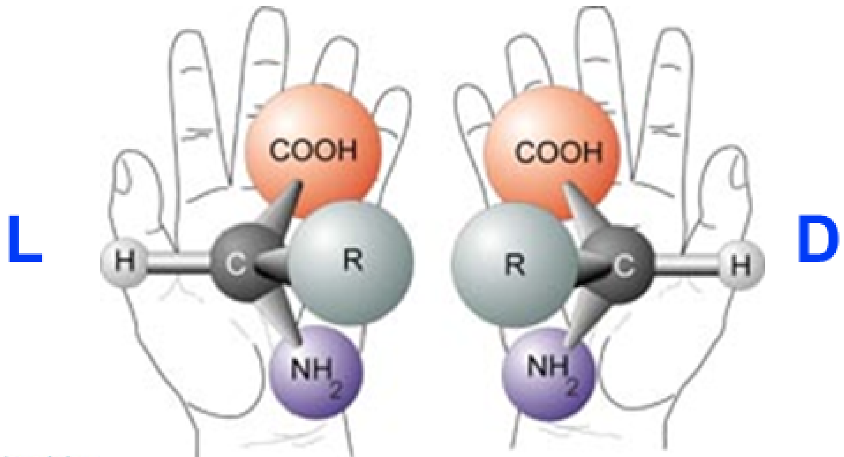
General Amino Acid Structure



Chirality of amino acids

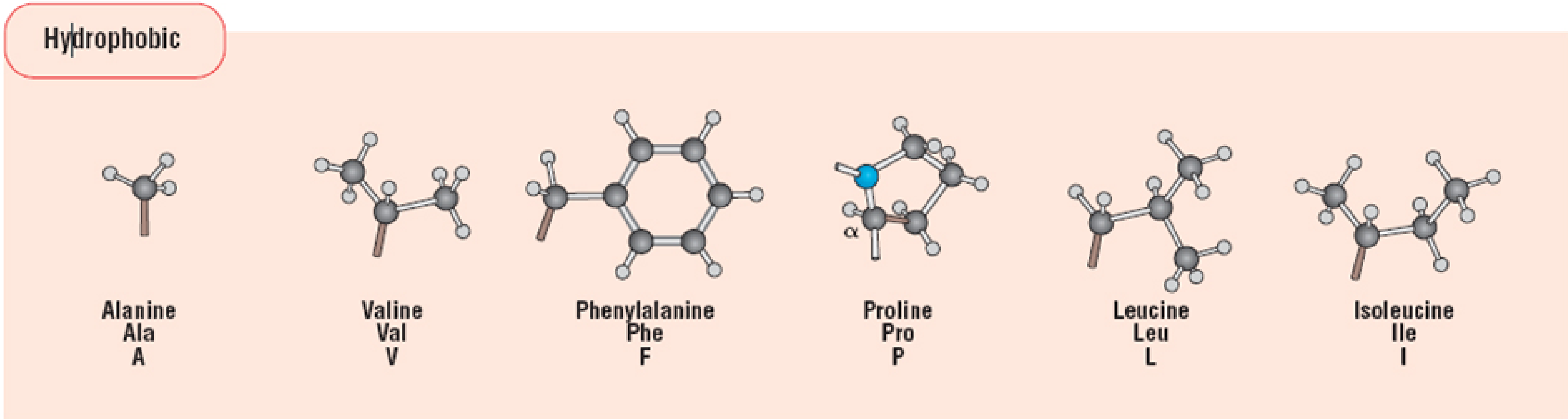
The "**CORN**" rule for determining the D/L isomeric form of an amino acid :

COOH, R, NH2 and H (where R is a variant carbon chain) are arranged around the chiral center C atom. Starting with the hydrogen atom away from the viewer, if these groups are arranged clockwise around the carbon atom, then it is the D-form. If counter-clockwise, it is the L-form.



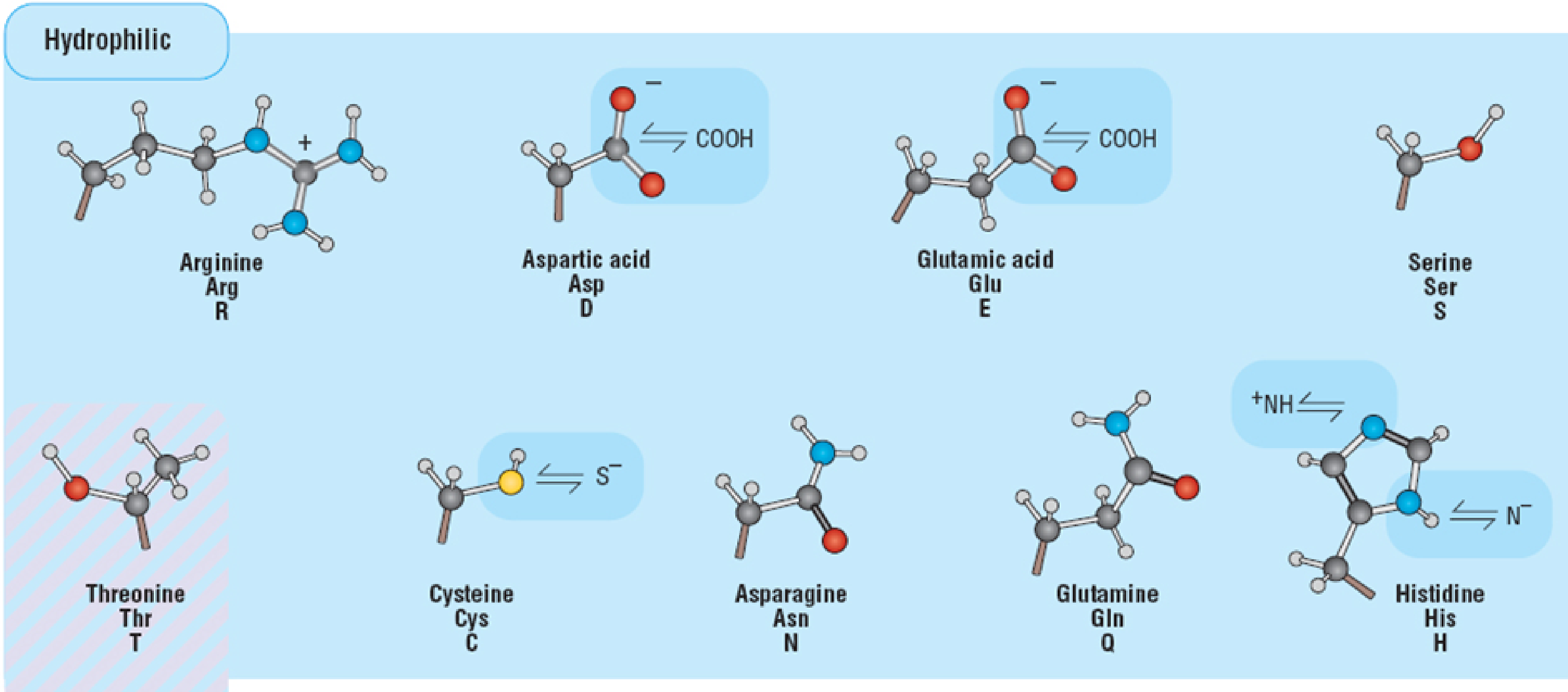
Varieties of amino acids

tending to avoid water, nonpolar and uncharged, relatively insoluble in water. **Side chains tend to associate with each other** to minimize their contact with water or polar side chains.



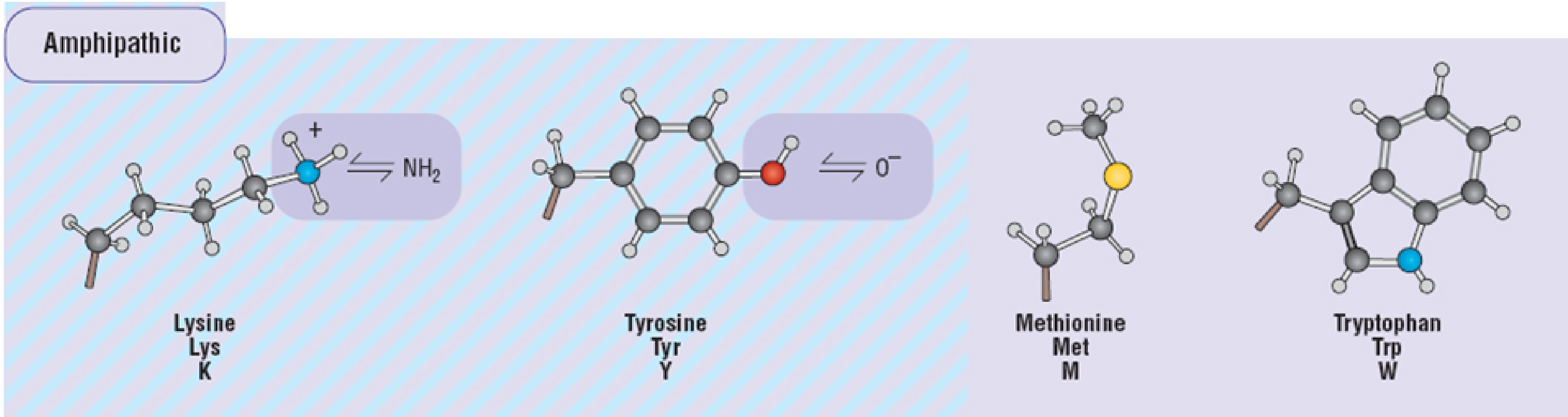
Varieties of amino acids

Interact with water, polar or charged,very soluble in water. **side chains tend to associate with other hydrophilic side** chains, or with water molecules, usually by means of hydrogen bonds.

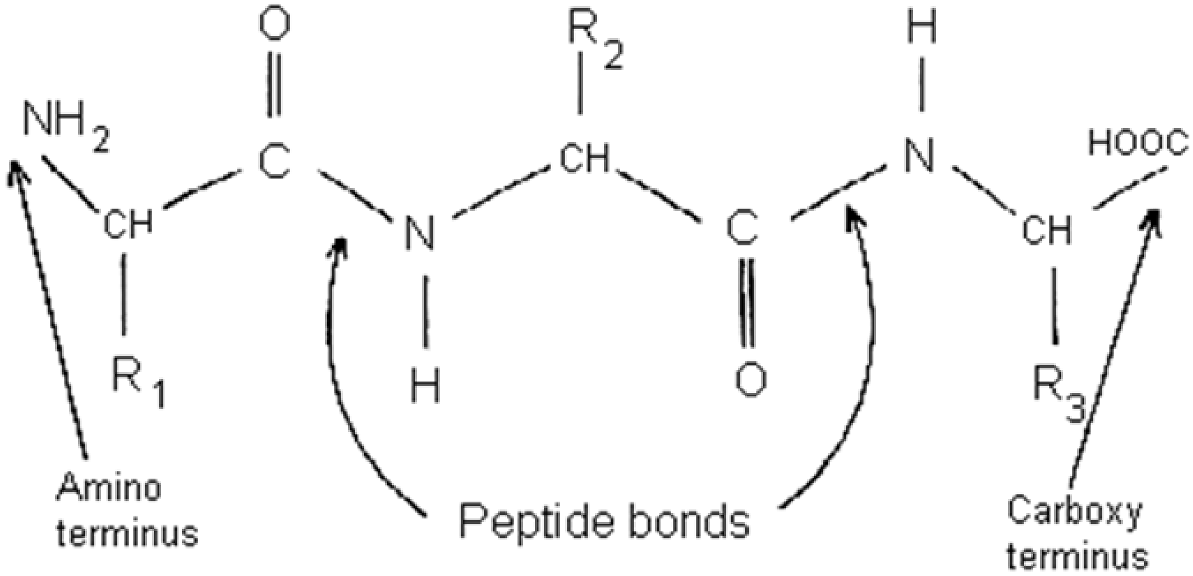


Varieties of amino acids

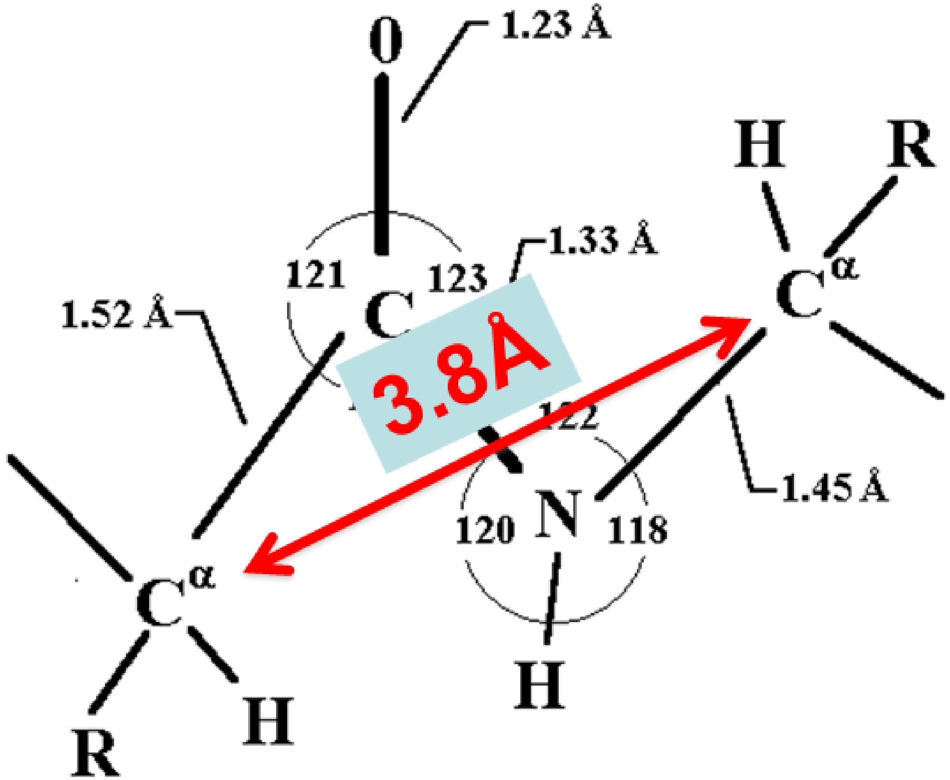
having both polar and nonpolar character
and therefore a tendency to form interfaces between
hydrophobic and hydrophilic molecules.



Peptide Chain

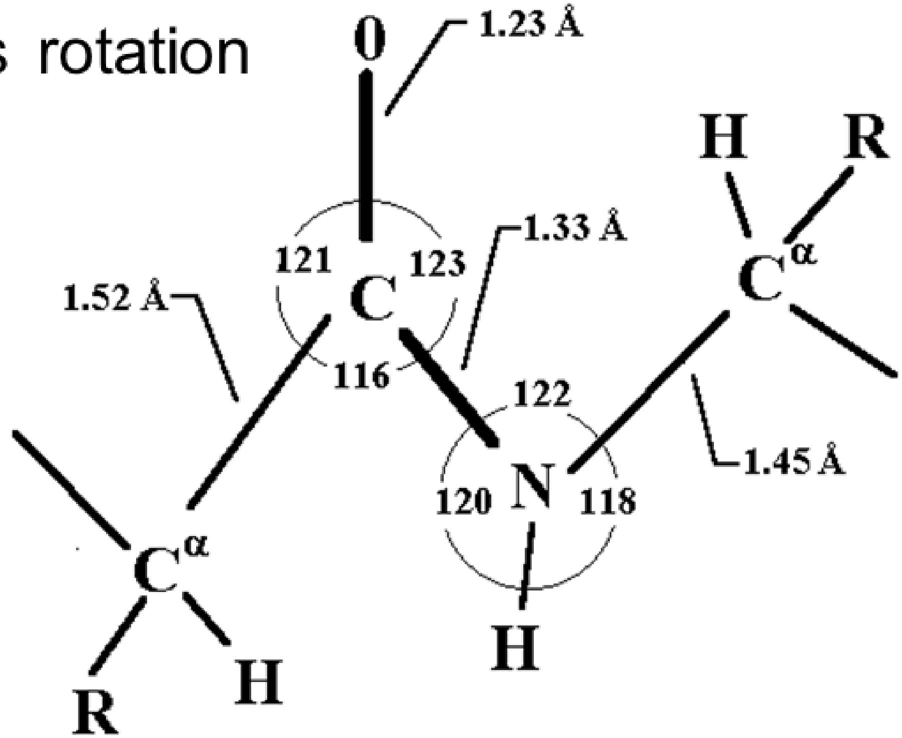


Peptide Bond Lengths



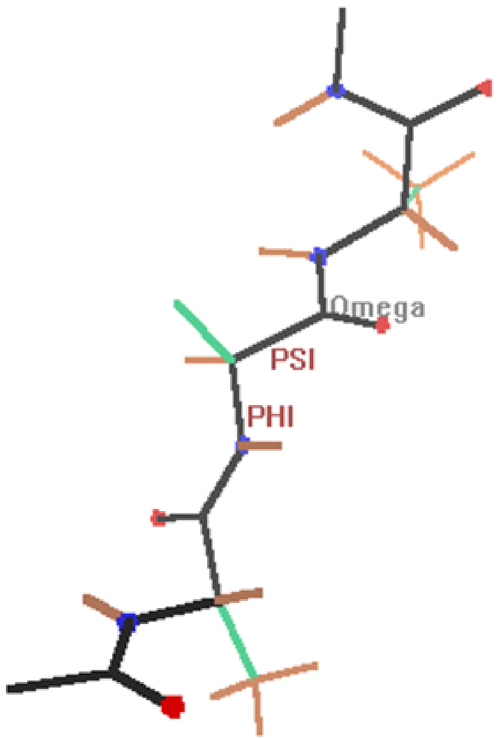
Protein Conformation Framework

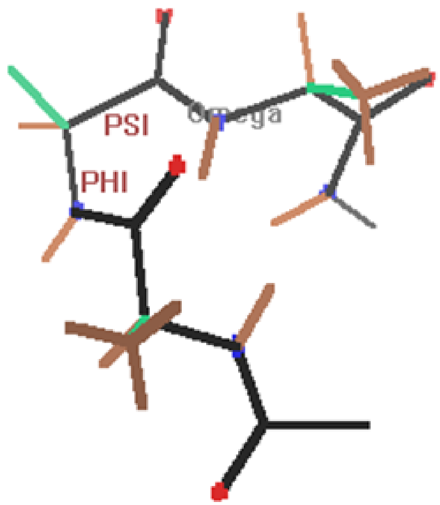
- Bond rotation determines protein folding, 3D structure
- Double bond disallows rotation



中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

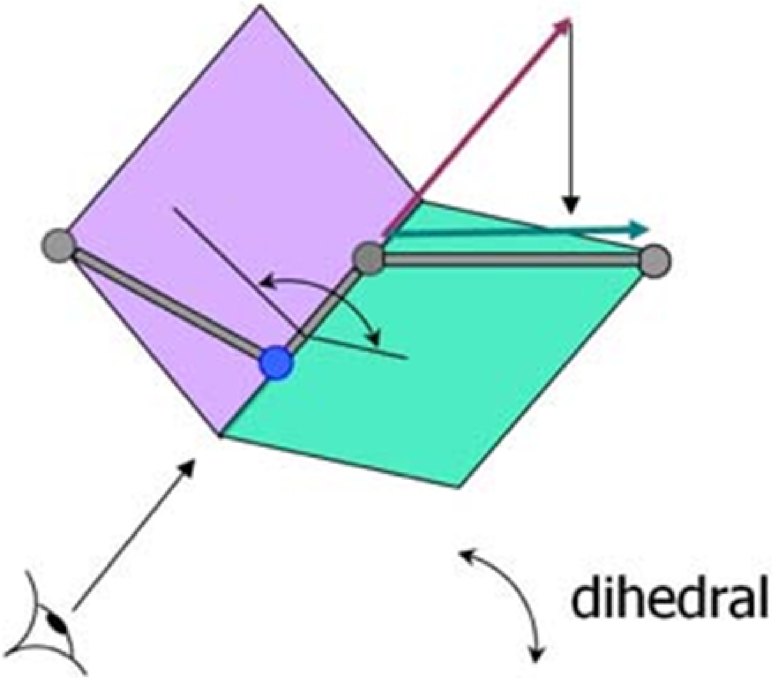
Bond Rotation Determines Protein Folding



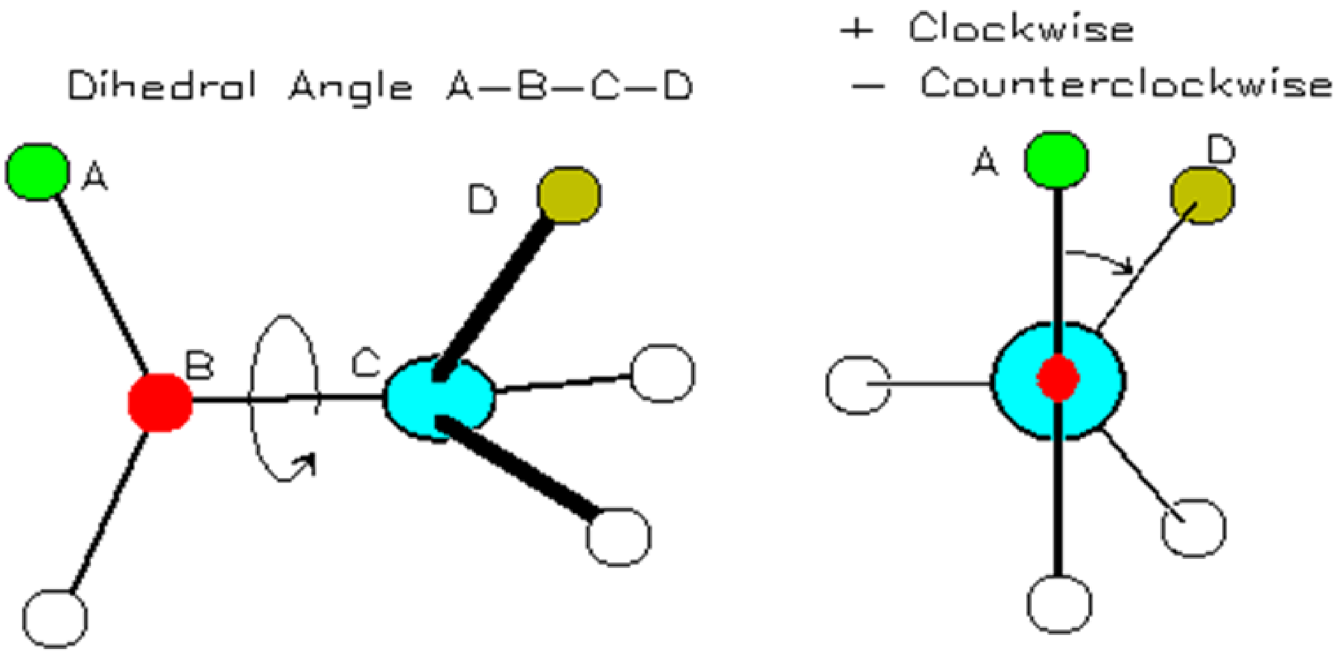


Protein Conformation Framework

- Torsion angle (dihedral angle)
  - Measures orientation of four linked atoms in a molecule: A, B, C, D

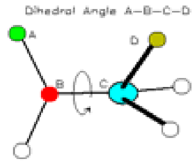


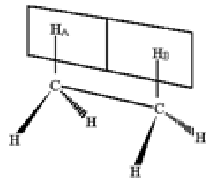
Dihedral angle



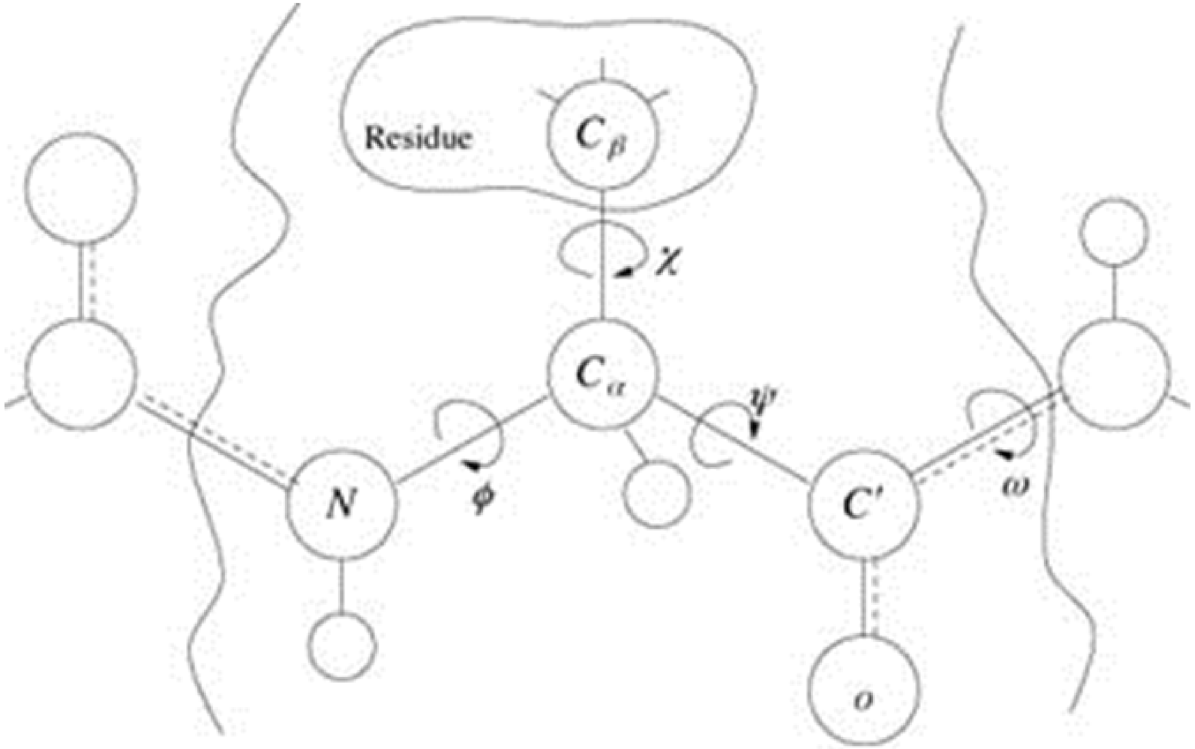
Protein Conformation Framework

- Torsion angle (dihedral angle)
  - Measures orientation of four linked atoms in a molecule: A, B, C, D
  - τ_{ABCD} defined as the angle between the normal to the plane of atoms A-B-C and normal to the plane of atoms B-C-D
  - Three repeating torsion angles along protein backbone: ω, φ, ψ





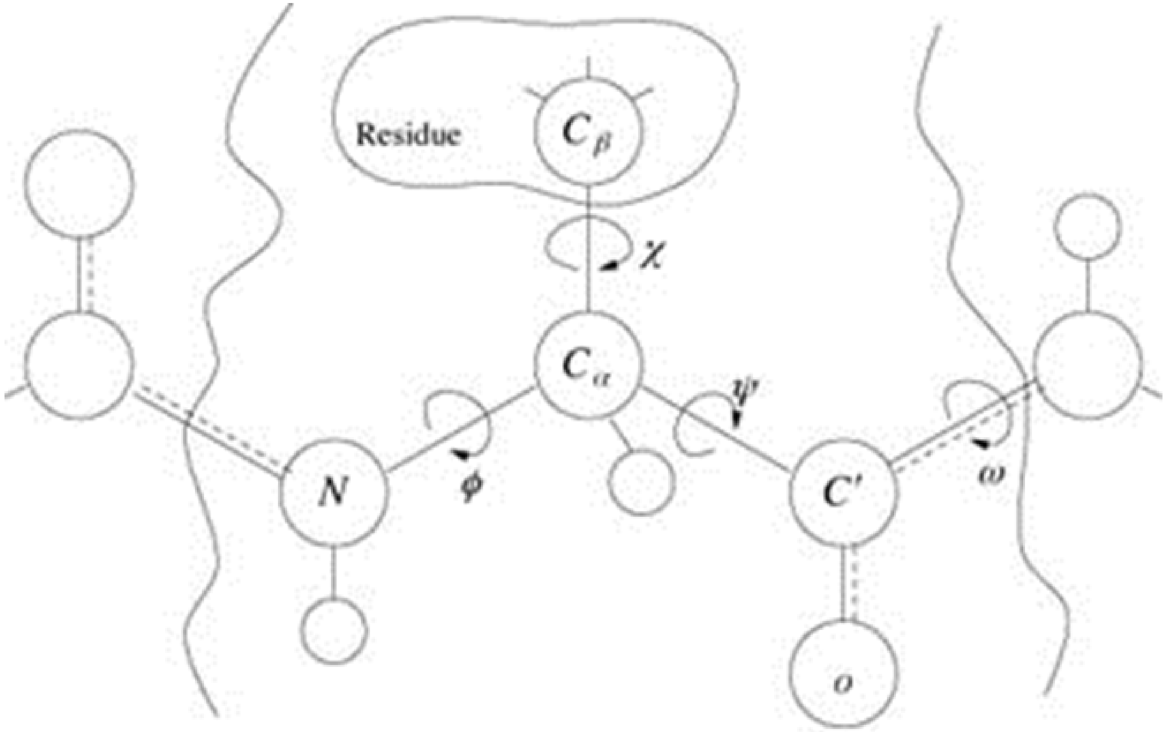
Backbone Torsion Angles



Backbone torsion angles of a protein

Backbone Torsion Angles

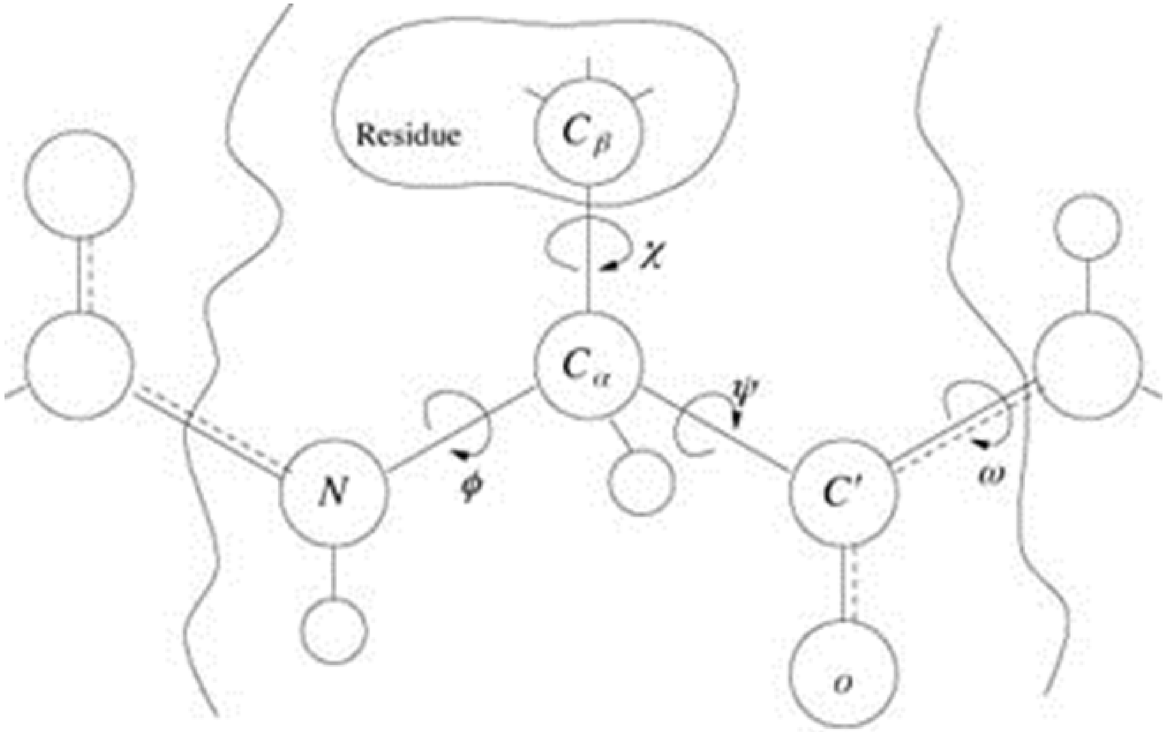
- Dihedral angle ω : rotation about the peptide bond, namely C^{α}_{1}-{C-N}- C^{α}_{2}



Backbone torsion angles of a protein

Backbone Torsion Angles

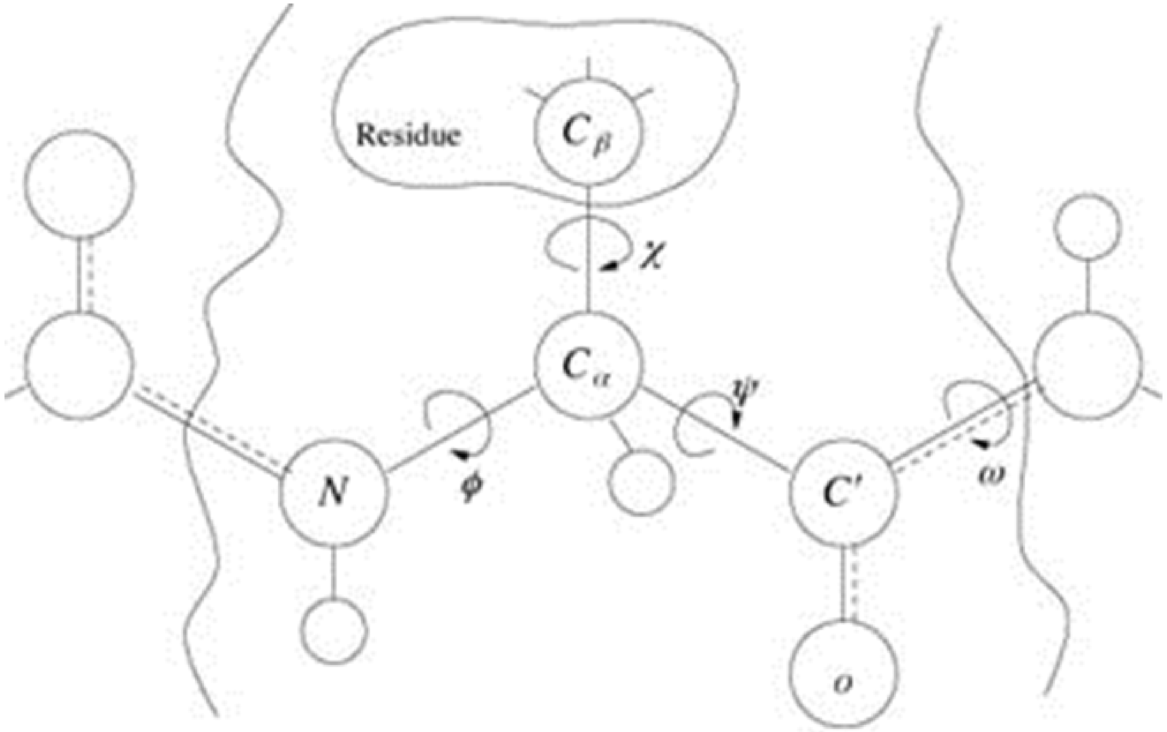
- Dihedral angle φ : rotation about the bond between N and C^{α}



Backbone torsion angles of a protein

Backbone Torsion Angles

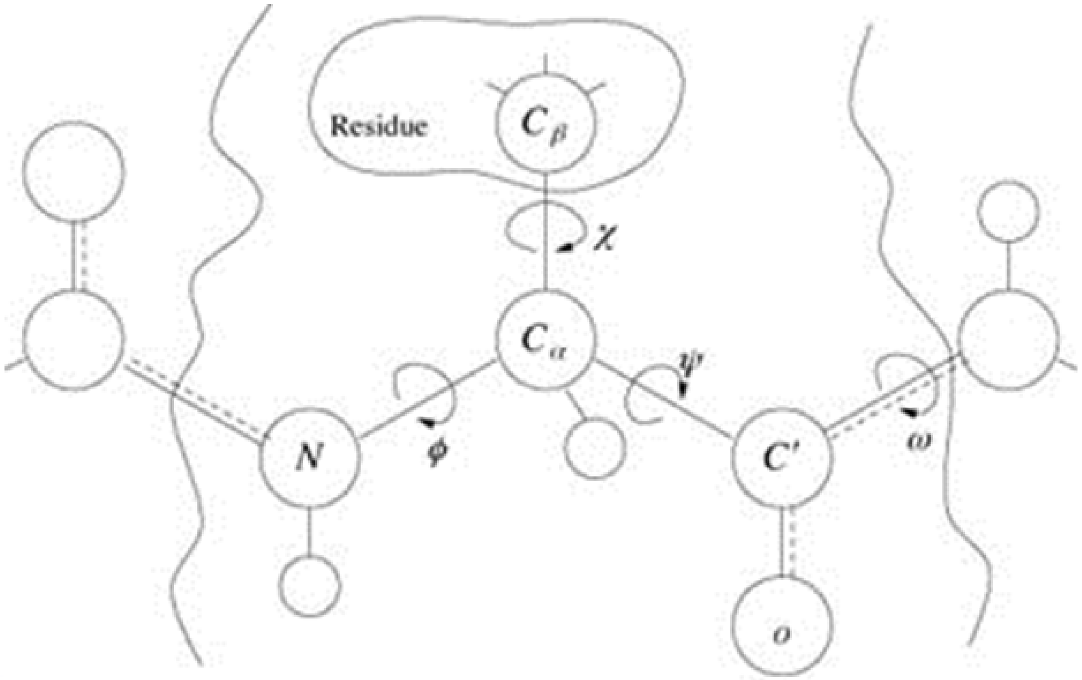
- Dihedral angle ψ : rotation about the bond between C^{α} and the carbonyl carbon



Backbone torsion angles of a protein

Backbone Torsion Angles

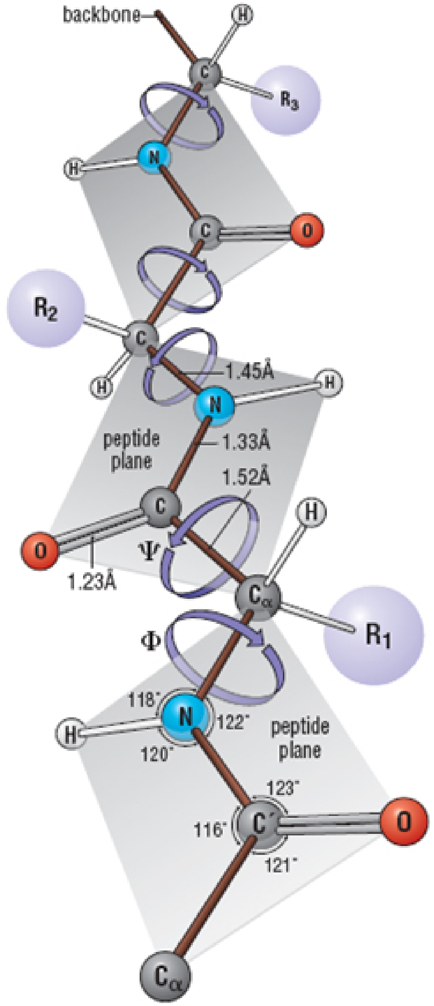
- ω angle tends to be planar (0º - cis, or 180 º - trans)



Backbone torsion angles of a protein

Backbone Torsion Angles

- φ and ψ are flexible, therefore rotation occurs here
- However, φ and ψ of a given amino acid residue are limited due to steric hindrance



Steric Hindrance

- Interference to rotation caused by spatial arrangement of atoms within molecule
- Atoms cannot overlap
- Atom size defined by van der Waals radii
- Electron clouds repel each other

G.N. Ramachandran

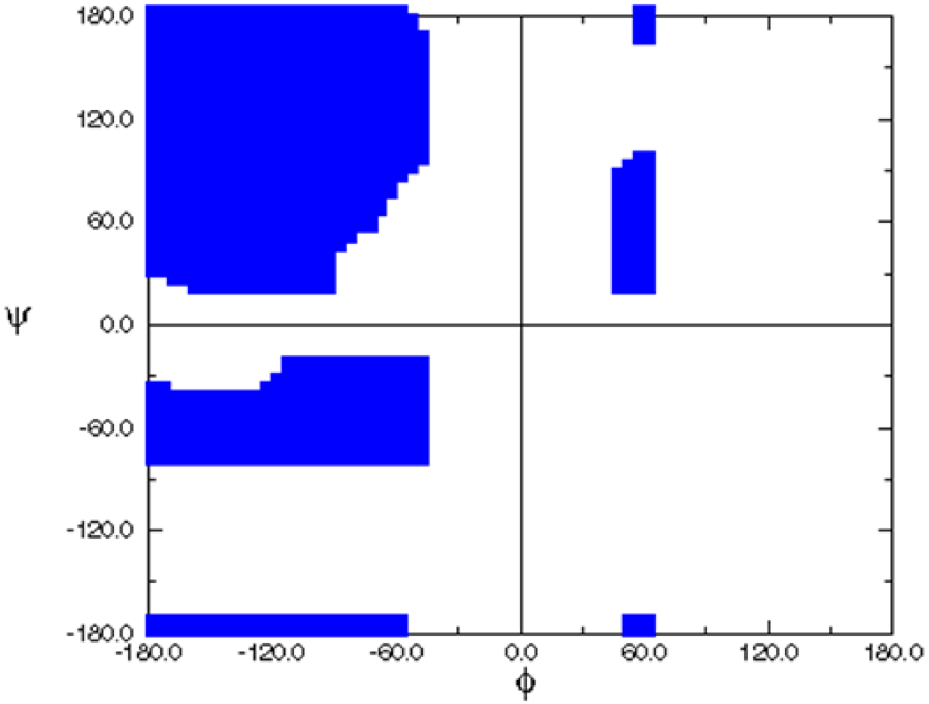
- Used computer models of small polypeptides to systematically vary φ and ψ with the objective of finding stable conformations
- For each conformation, the structure was examined for close contacts between atoms
- Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii
- Therefore, φ and ψ angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone
- Only 10% of the {φ, ψ} combinations are generally observed for proteins
- First noticed by G.N. Ramachandran

Ramachandran Plot

- Plot of φ vs. ψ
- The computed angles which are sterically allowed fall on certain regions of plot

Computed Ramachandran Plot

Hard Sphere Ramachandran Map

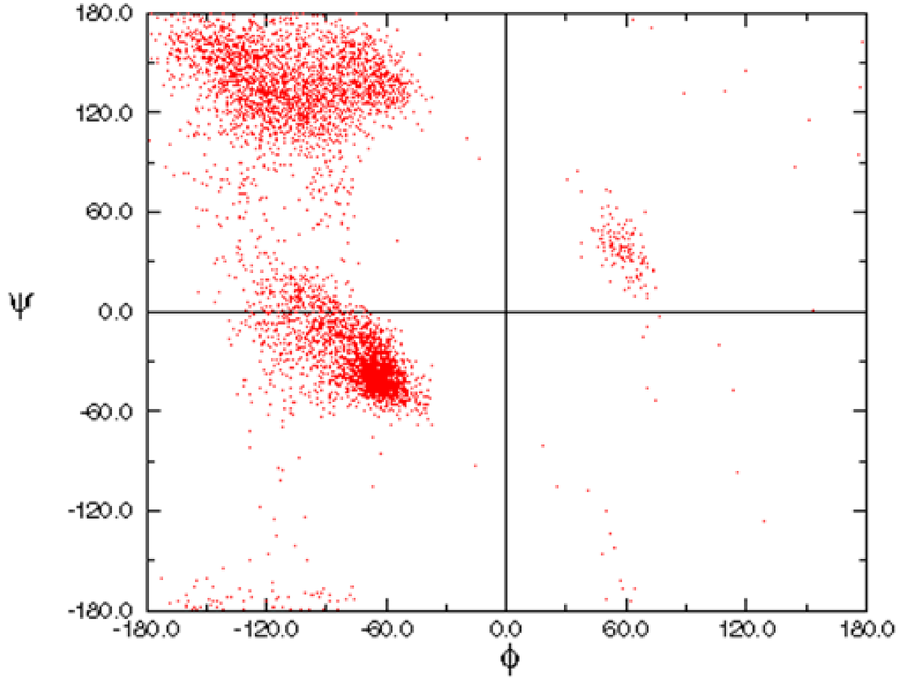


White = sterically disallowed conformations (atoms come closer than sum of van der Waals radii)

Blue = sterically allowed conformations

Experimental Ramachandran Plot

φ, ψ distribution in 42 high-resolution protein structures (x-ray crystallography)



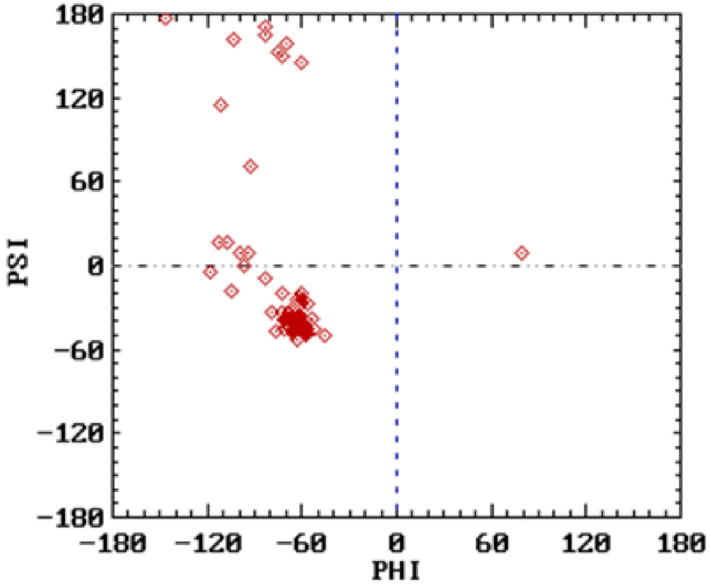
中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

Ramachandran Plot And Secondary Structure

- Repeating values of φ and ψ along the chain result in regular structure
- For example, repeating values of φ ~ - 57° and ψ ~ - 47° give a right-handed helical fold (the alpha-helix)

The structure of cytochrome C shows many segments of helix and the Ramachandran plot shows a tight grouping of φ, ψ angles near -50,-50

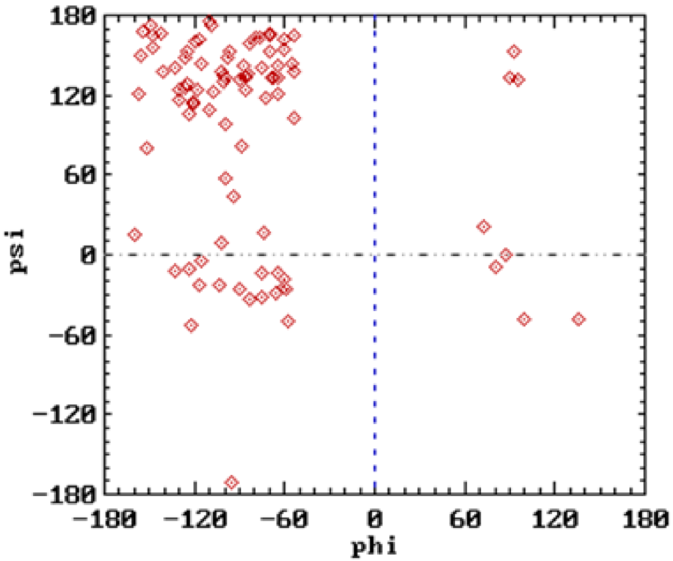
alpha-helix



cytochrome C
Ramachandran plot

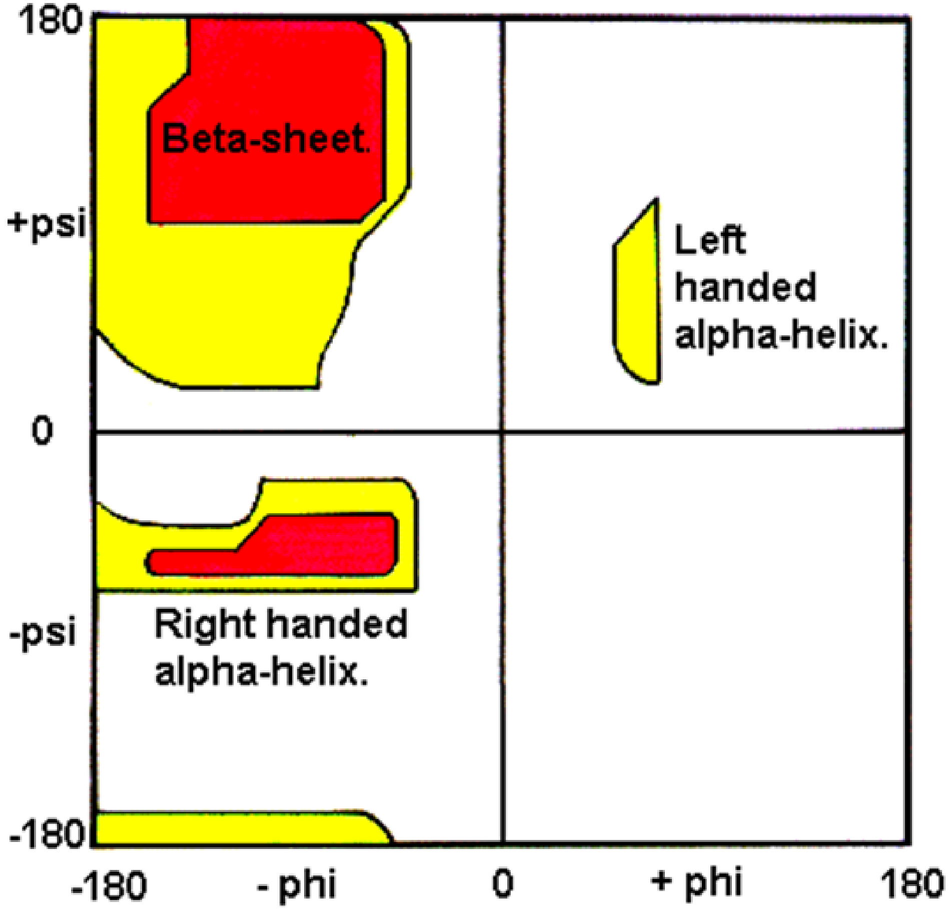
Similarly, repetitive values in the region of φ = -110 to –140 and ψ = +110 to +135 give beta sheets. The structure of plastocyanin is composed mostly of beta sheets; the Ramachandran plot shows values in the –110, +130 region:

beta-sheet



plastocyanin
Ramachandran plot

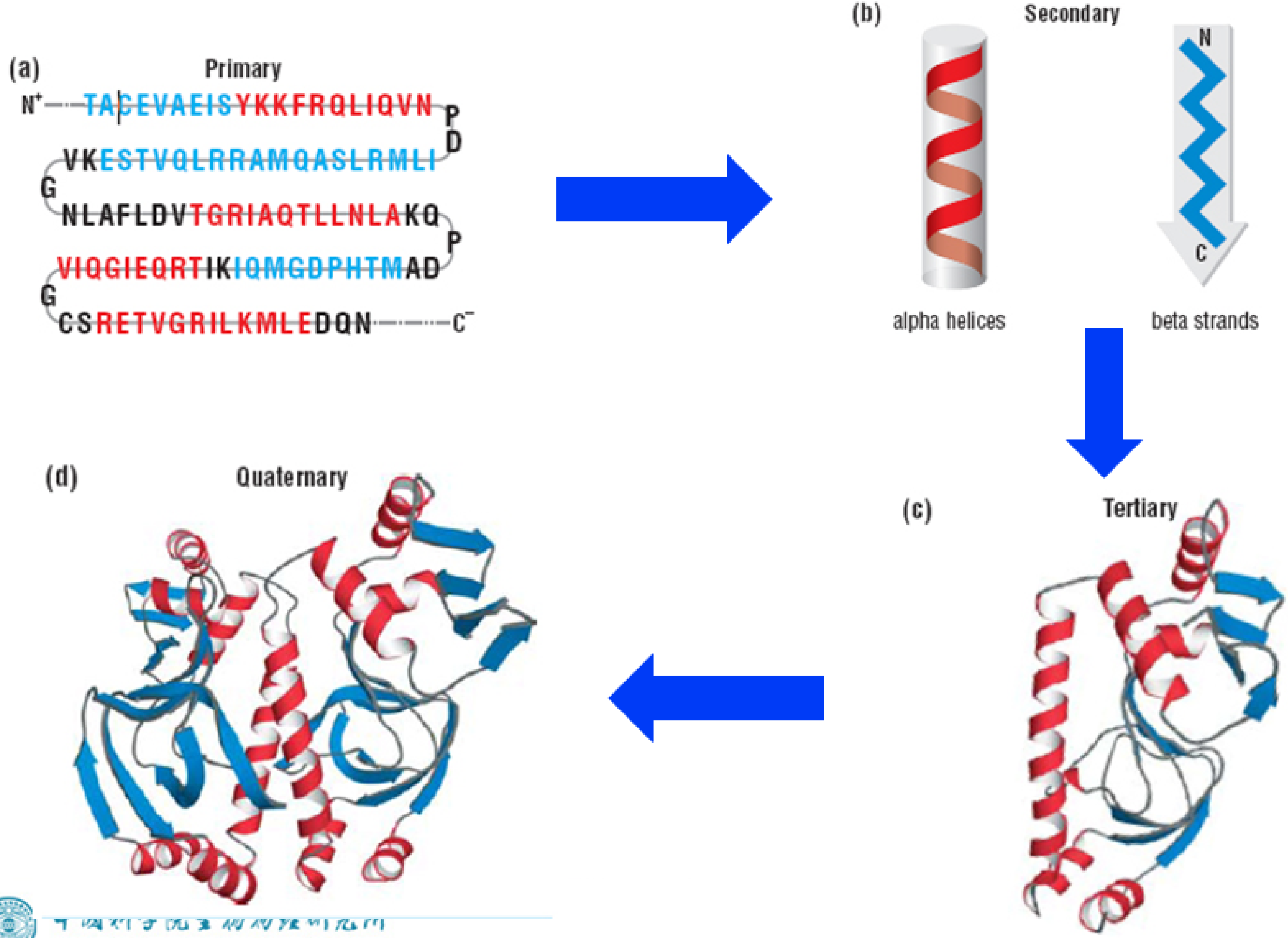
The Ramachandran Plot.



φ, ψ and Secondary Structure

Name	φ	ψ	Structure
alpha-L	57	47	left-handed alpha helix
3-10 Helix	-49	-26	right-handed.
π helix	-57	-80	right-handed.
Type II helices	-79	150	left-handed helices formed by polyglycine and polyproline.
Collagen	-51	153	right-handed coil formed of three left handed helicies.

Four levels of protein structure



The Universe of Protein Structures

How many proteins in the universe?

The smallest archaea genome encodes above 600 ORFs

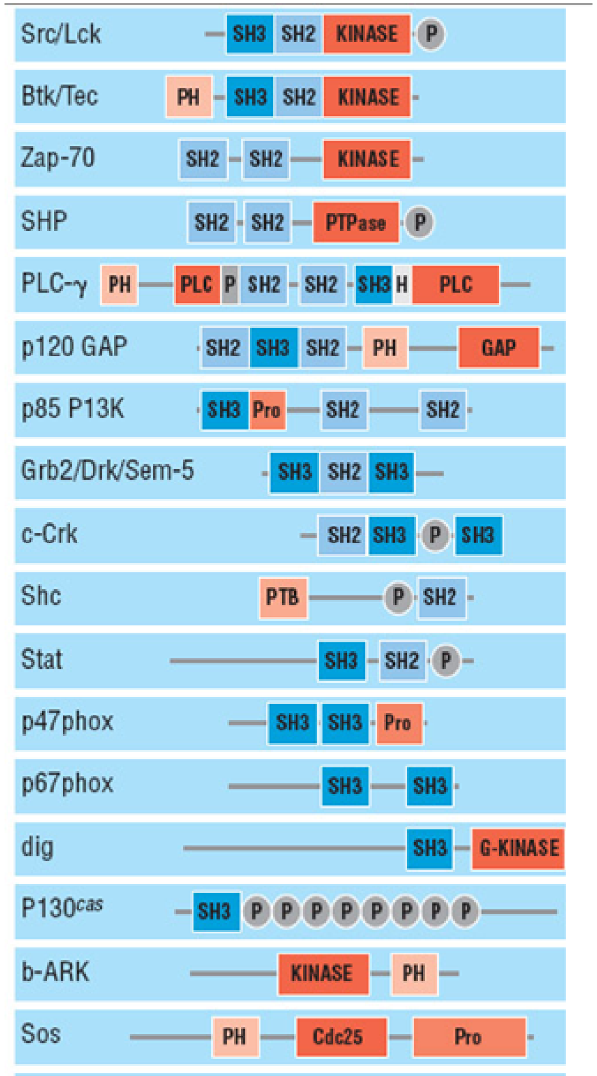
Pyrococcus furiosus encodes 2200 ORFs

Homo sapiens encodes around 30,000 ORFS

The facts:
The number of protein folds is large but limited. the number of different protein folds in nature is limited. They are used repeatedly in different combinations to create the diversity of proteins found in living organisms.

The Universe of Protein Structures

Protein structures are modular and proteins can be grouped into families on the basis of the domains they contain



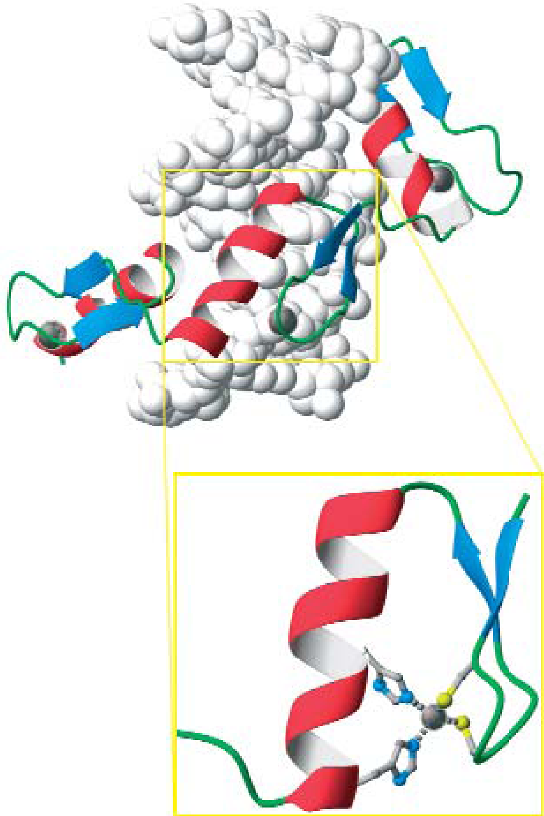
There are around 1000 different protein folds

中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

The Universe of Protein Structures

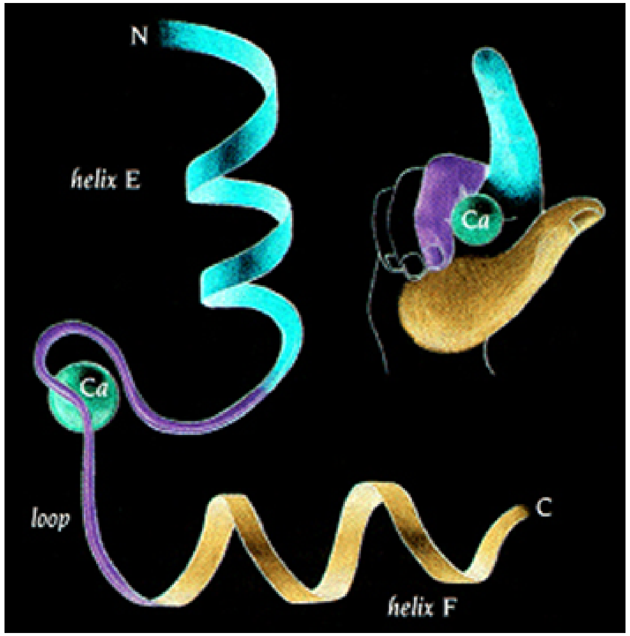
Protein motifs may be defined by their primary sequence or by the arrangement of secondary structure elements

Zinc finger motif

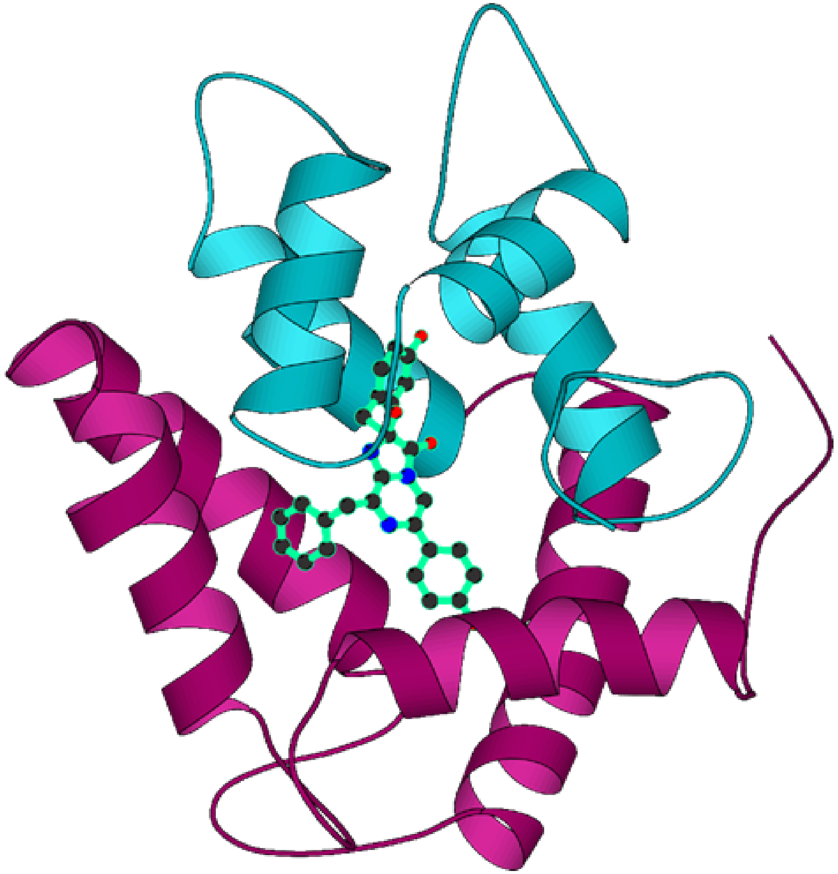


中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

The Universe of Protein Structures



EF-hand motif



Protein Function in Cell

1. Enzymes
   - Catalyze biological reactions
2. Structural role
   - Cell wall
   - Cell membrane
   - Cytoplasm