Development of Complex Curricula for Molecular Bionics and Infobionics Programs within a consortial* framework**

Consortium leader

PETER PAZMANY CATHOLIC UNIVERSITY

Consortium members

SEMMELWEIS UNIVERSITY, DIALOG CAMPUS PUBLISHER

The Project has been realised with the support of the European Union and has been co-financed by the European Social Fund ***

**Molekuláris bionika és Infobionika Szakok tananyagának komplex fejlesztése konzorciumi keretben

***A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

INTRODUCTION TO BIOPHYSICS

(Bevezetés a biofizikába)

STRUCTURE OF PROTEINS

(A fehérjék szerkezete)

GYÖRFFY DÁNIEL, ZÁVODSZKY PÉTER

Introduction

- Proteins are linear polymers of amino acids
- Four main levels of the structure of proteins are discerned
  - Sequence of amino acids
  - Local conformational elements
  - Global spatial arrangement of the whole protein
  - Subunit structure of proteins consisting of two or more chains
- Several secondary structural elements can form motives
- Several motives can form domains

- Covalent and non-covalent interactions determine the structure and drive the folding of proteins
- An entropic force, called hydrophobic effect is the main contributor to the stability of proteins
- The structure of proteins can be described and studied by means of statistical physics
- The native state of proteins is usually the state with the lowest free energy
- The folding of proteins is often a cooperative process

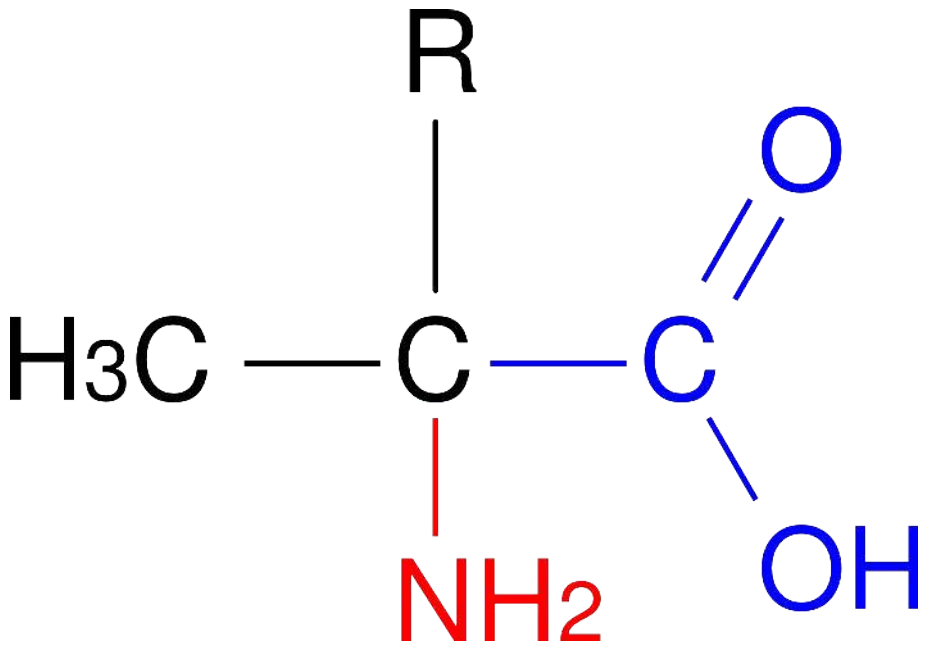
- A funnel-shaped free energy landscape describes the folding and association of proteins
- Dynamic properties of proteins are important for the function

Amino acids

- Amino acids are compounds containing both a carboxyl and an amino group
- Amino acids forming proteins are amino acids in which the amino group is connected to the α-carbon
- Only twenty of the α-amino acids take part in building up protein chains
- Protein forming amino acids can be grouped based on their chemical properties such as hydrophobicity or acid-base properties

- Different kinds of amino acids have different side chains while their backbone atoms are the same
- A general amino acid can be seen in the following figure
- The α-amino group is coloured red and the α-carboxyl group is coloured blue
- R represents the side chain

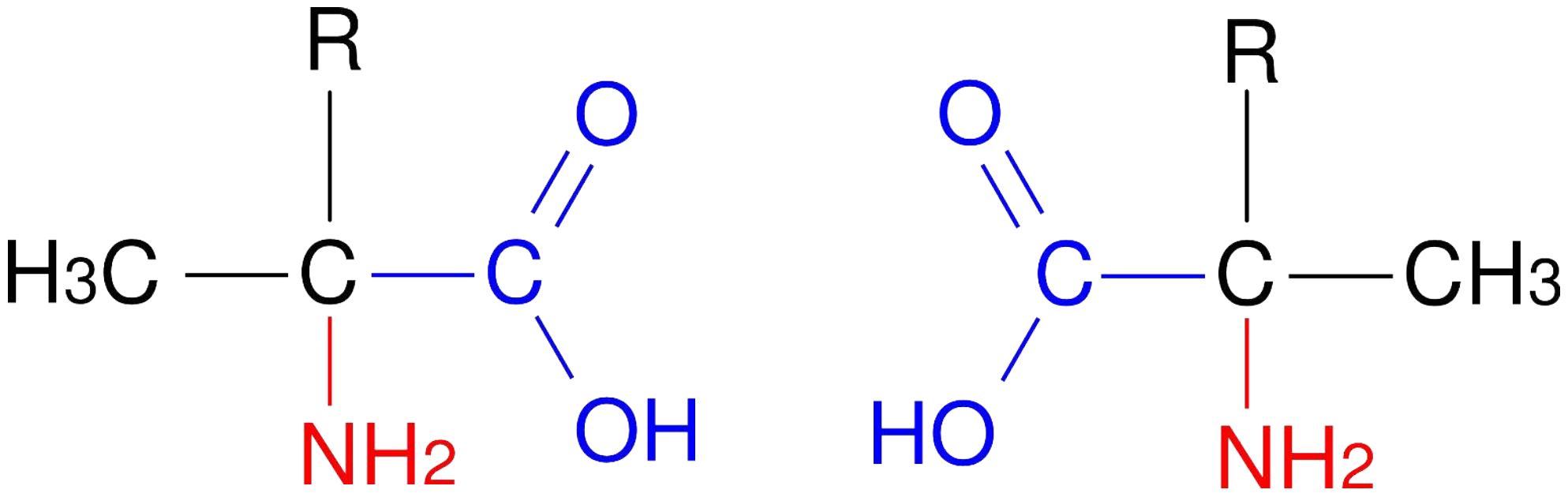
General forms of amino acids



Chirality of amino acids

- Amino acids - except glycine - have asymmetric carbon atoms
- Due to asymmetric α-carbon atoms, optical isomerism occurs, so two isomers of such amino acids exist which are mirror images of each other (this phenomenon is called chirality)
- In proteins, only the L-conformers of amino acids are found (Figure 2.)
- Amino acid derivatives forming the protein chains are often referred to as 'residues'

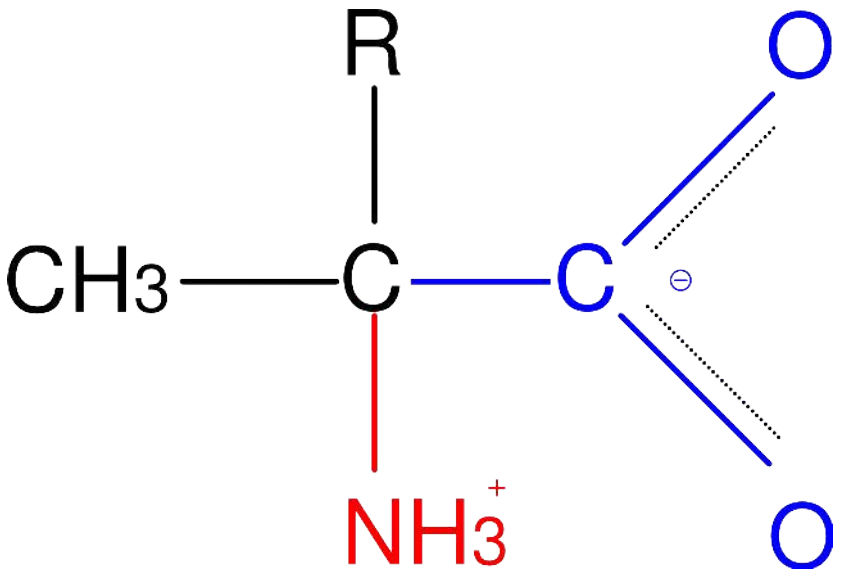
L and D-amino acids



Electrical properties of amino acids

- In some pH range, amino acids can be electrically neutral but they have dissociable groups such as carboxyl group or amino group
- At low pH, the carboxyl group is protonated while at high pH, the H^+ ion dissociates from it
- At high pH, the amino group is deprotonated while at small pH, it carries one more H^+ ion
- There exists a pH range where both the α-carboxyl and the α-amino groups are charged which is called a zwitter-ion (Figure 3.)

Zwitter-ion



- Several amino acids have multiple dissociable groups (see the following table)

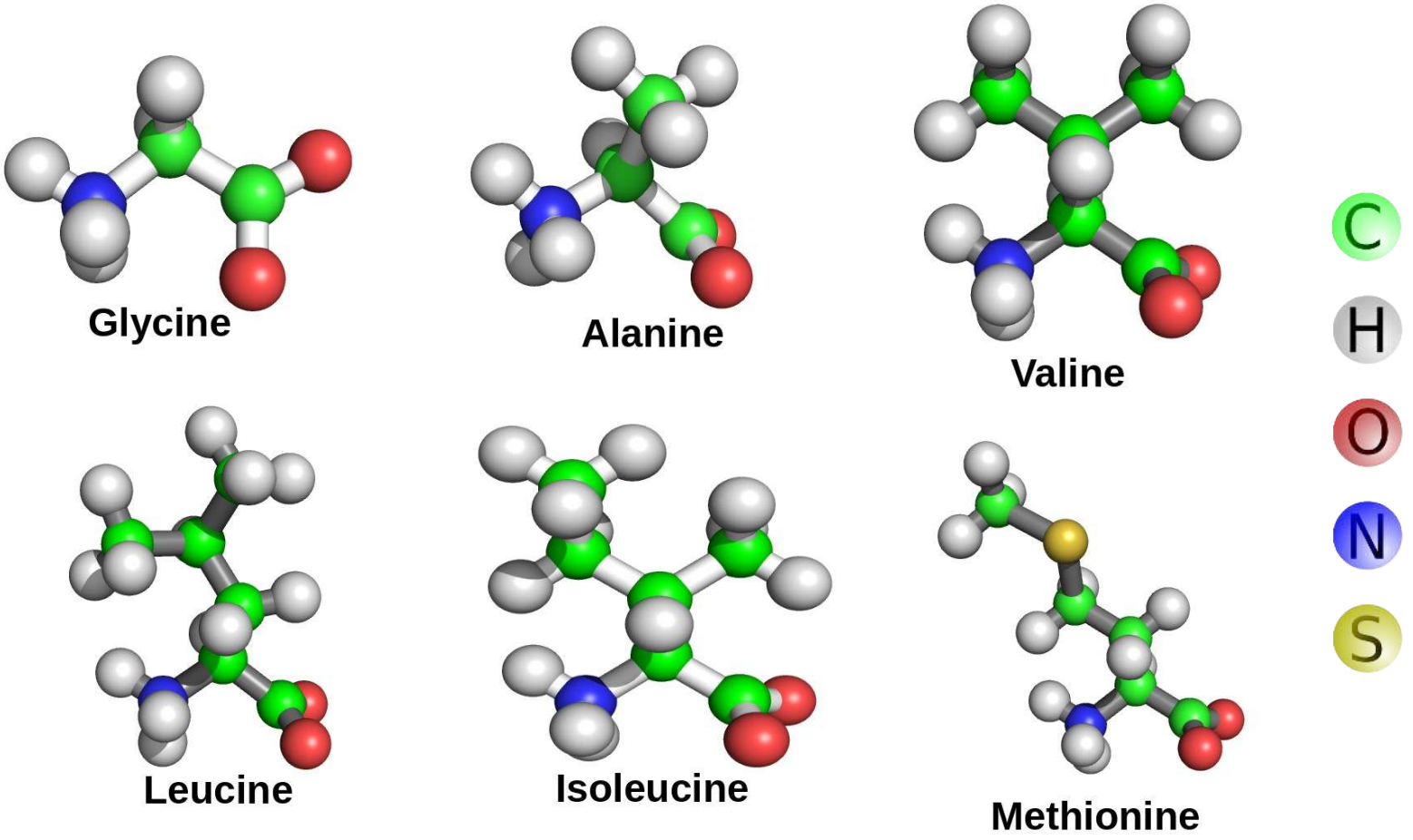
Dissociable groups

Group	pK
α-carboxyl	1.88-2.36
α-amino	8.8-10.96
Tyrosine -OH	10.07
Cysteine -SH	8.18
Lysine ε-amino	10.53
Histidine Imidazole-NH	6.00
Arginine guanidino	12.48
Aspartate β-carboxyl	3.65
Glutamate γ-carboxyl	4.25

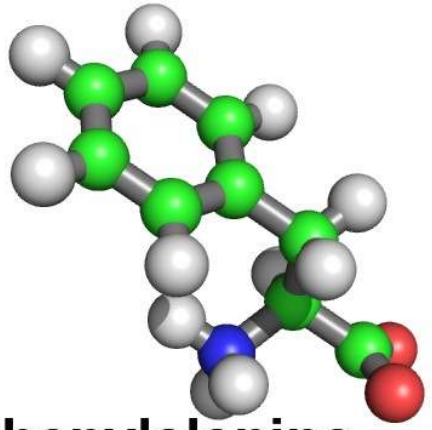
Groups of proteinogenic amino acids

- Amino acids can be clustered into several groups based on their chemical properties
  - Hydrophophobic, aliphatic
  - Aromatic
  - Uncharged polar
  - Positively charged
  - Negatively charged

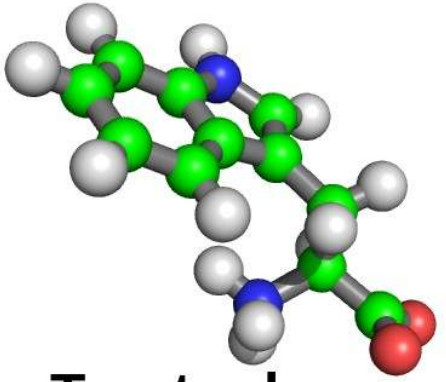
Hydrophobic, aliphatic amino acids



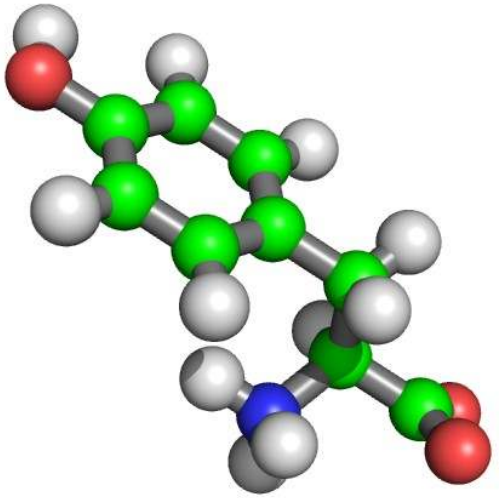
Aromatic amino acids



Phenylalanine

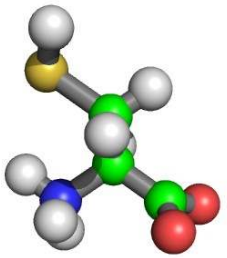


Tryptophan

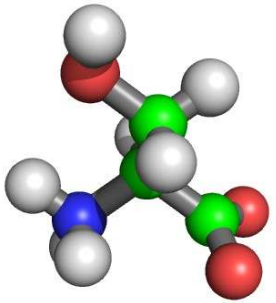


Tyrosine

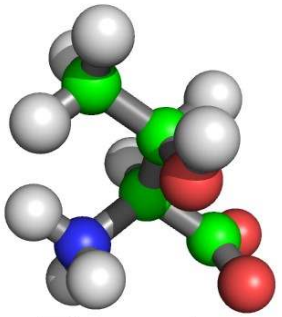
Uncharged polar amino acids



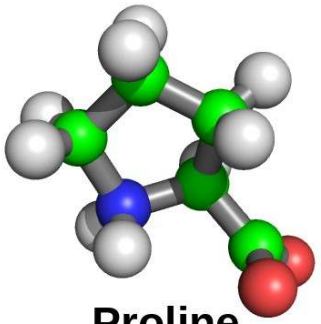
Cysteine



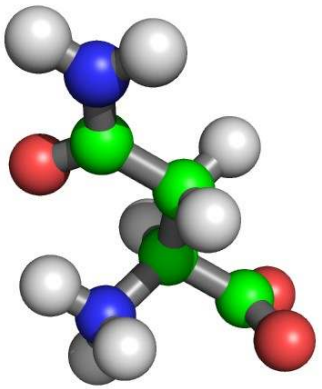
Serine



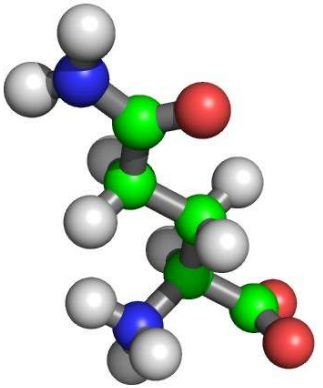
Threonine



Proline

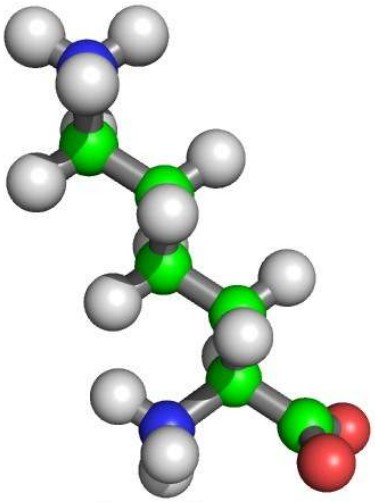


Asparagine

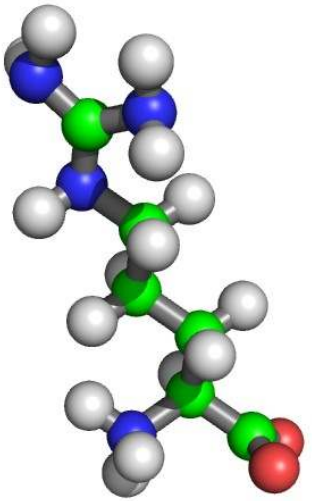


Glutamine

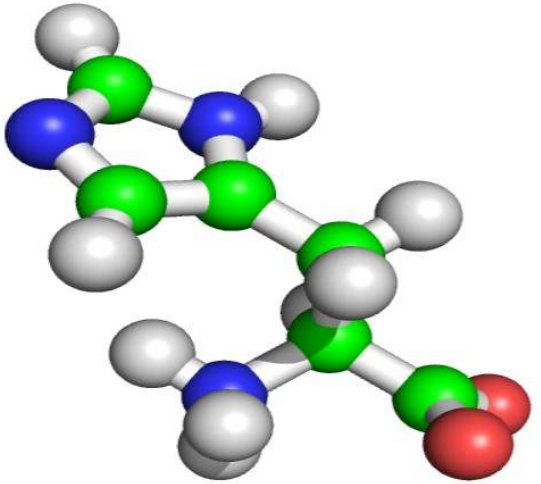
Positively charged amino acids



Lysine

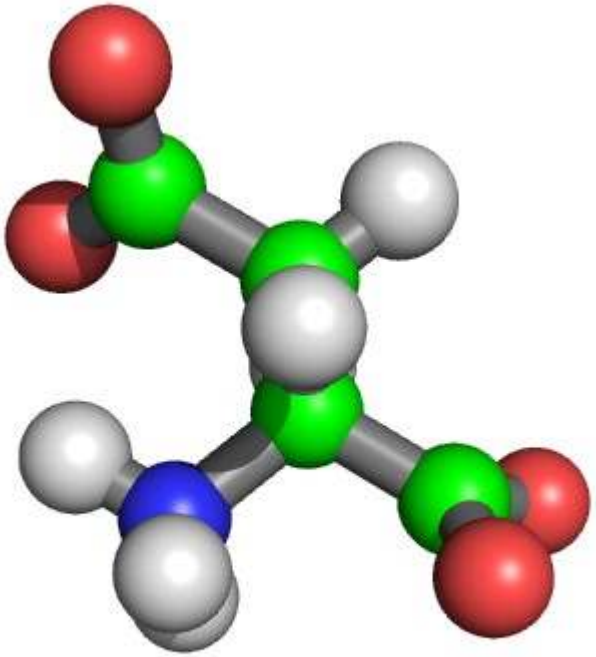


Arginine

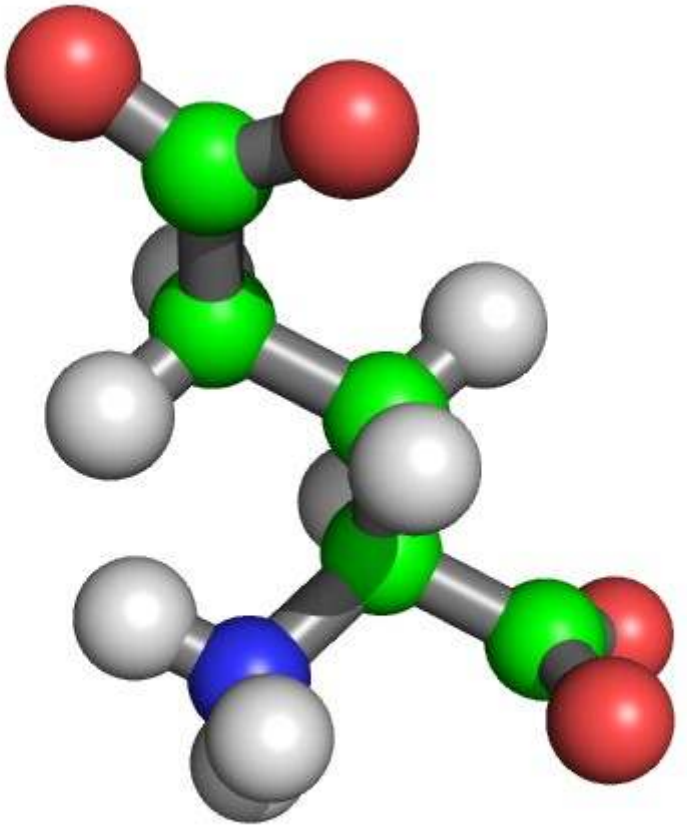


Histidine

Negatively charged amino acids



Aspartate



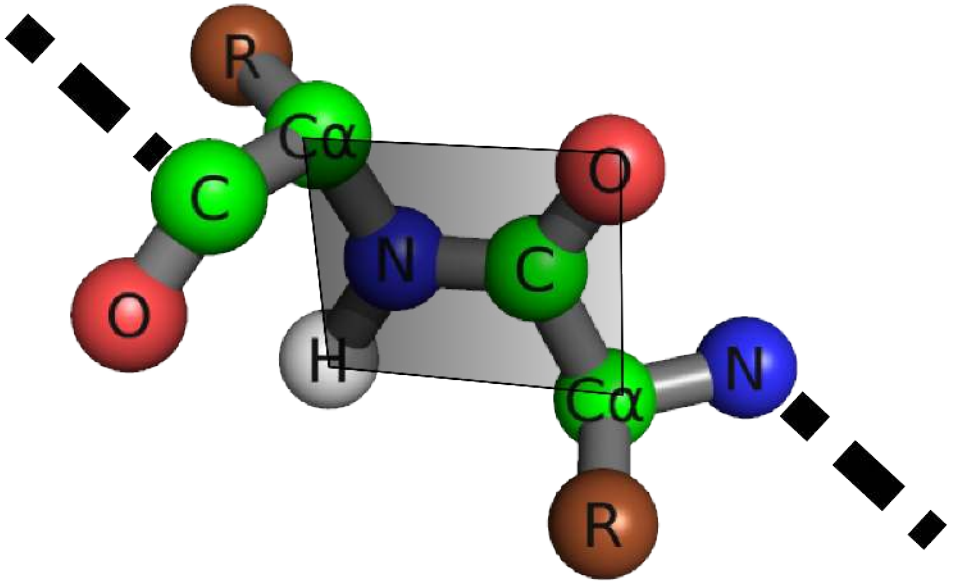
Glutamate

Peptide bond

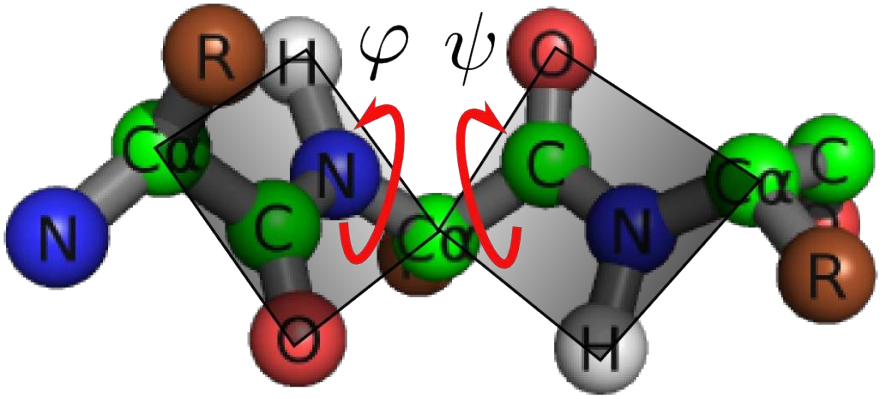
- Amino acids are connected to each other by the peptide bond
- Peptide bond is an ester-bond between the α-carboxyl group of an amino acid and the α-amino group of another one
- The peptide bond is approximately planar due to delocalization of electrons between single and double covalent bonds (see the following figure)

- Due to the planarity, only two bonds can rotate
  - The bond between C_α and amino-N by the φ angle
  - The bond between C_α and the carbon atom in the carbonyl group by the ψ angle
- φ and ψ angles are called torsion angles
- The energy of a given conformation is a function of the two torsion angles
- The plot depicting the energy versus torsion angles is called a Ramachandran plot
- On the Ramachandran plot, there are distinct areas corresponding to special local structural elements

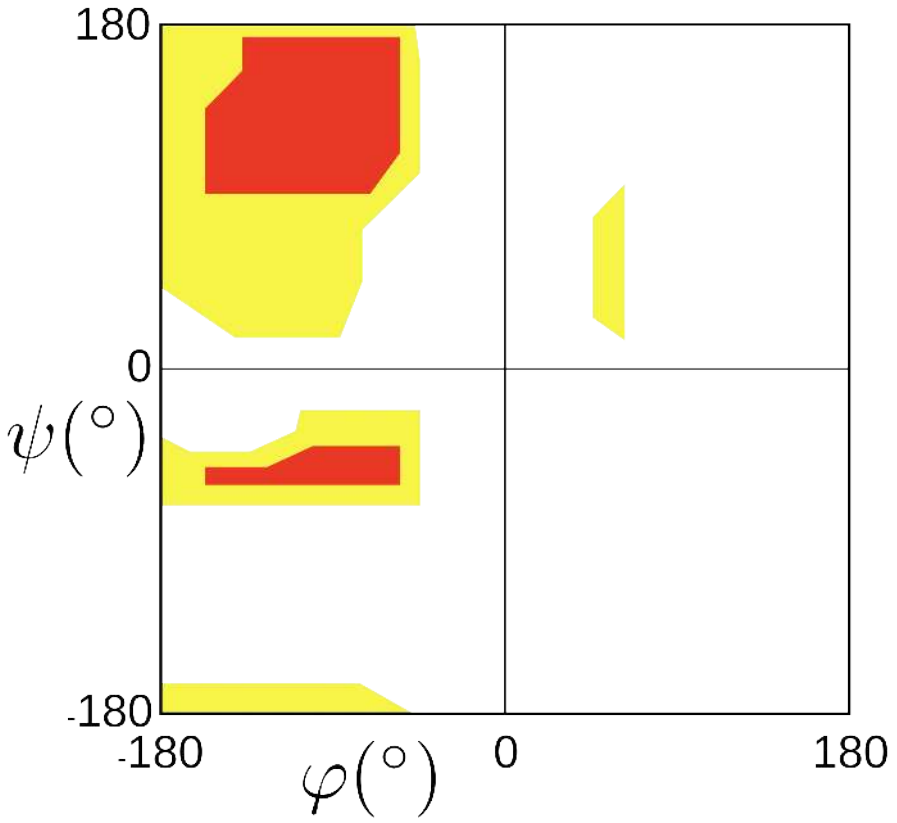
The planar peptide bond



The torsion angles



The Ramachandran plot



Levels of protein structure

- The primary structure of proteins is the linear sequence of amino acids forming the polypeptide chain
- The secondary structure consists of local, often regular and periodic structural elements
- The tertiary structure of a protein is the global spatial arrangement of its atoms
- Quaternary structure is the subunit structure of proteins consisting of more than one chains

Elements of secondary structure

- Secondary structure consists of local structural elements stabilized by interactions between atoms being close to each other
- The elements are determined according to the permitted areas on the Ramachandran plot and the hydrogen bonds possible to be formed
- More such secondary structural elements can form structural motifs and supersecondary elements

Main types of secondary structural elements

Type	Example
Helix	α-helix
Extended	β-sheet
Turn	β-turn
Loop	ω-loop

Helices

- Helices are produced by a translation followed by a rotation
- Several helices are found in the structures of proteins, each of which is characterized by the torsion angles of the part of the chain forming the helix
- Most of the helices are right-handed except the π-helix of which there exists a left-handed variation
- Helices are stabilized by hydrogen bonds within the backbone

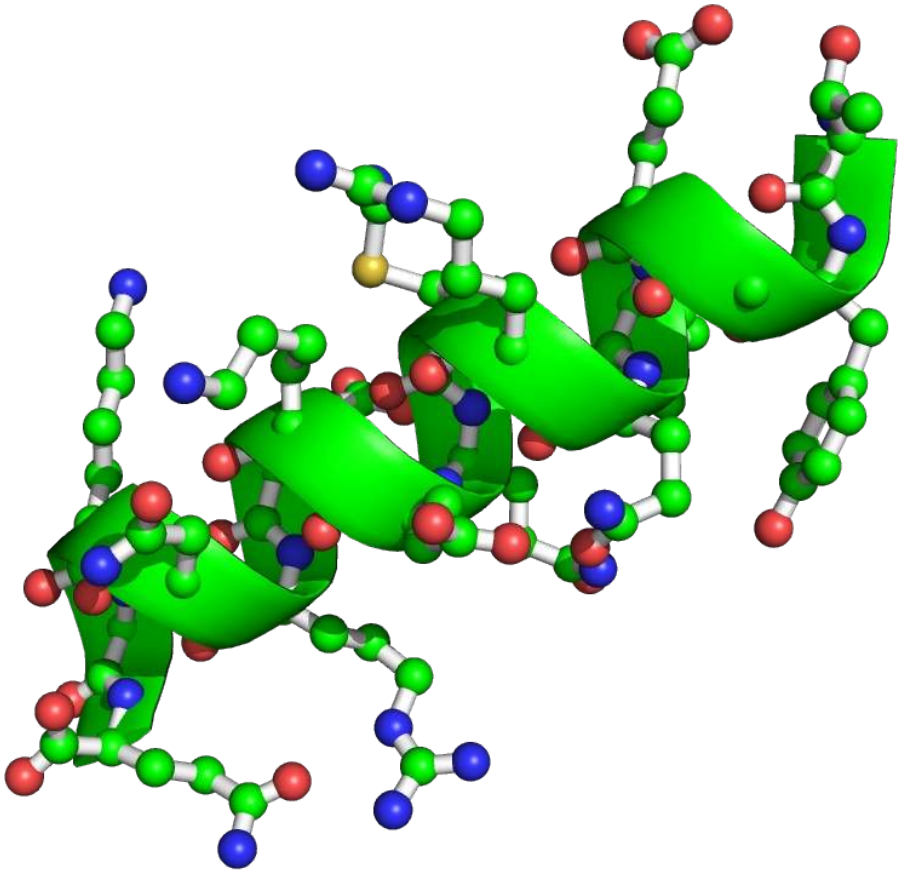
- Due to the electrostatic properties of amino acids, helices have a dipole moment pointing from the C towards the N-terminus

Helix types

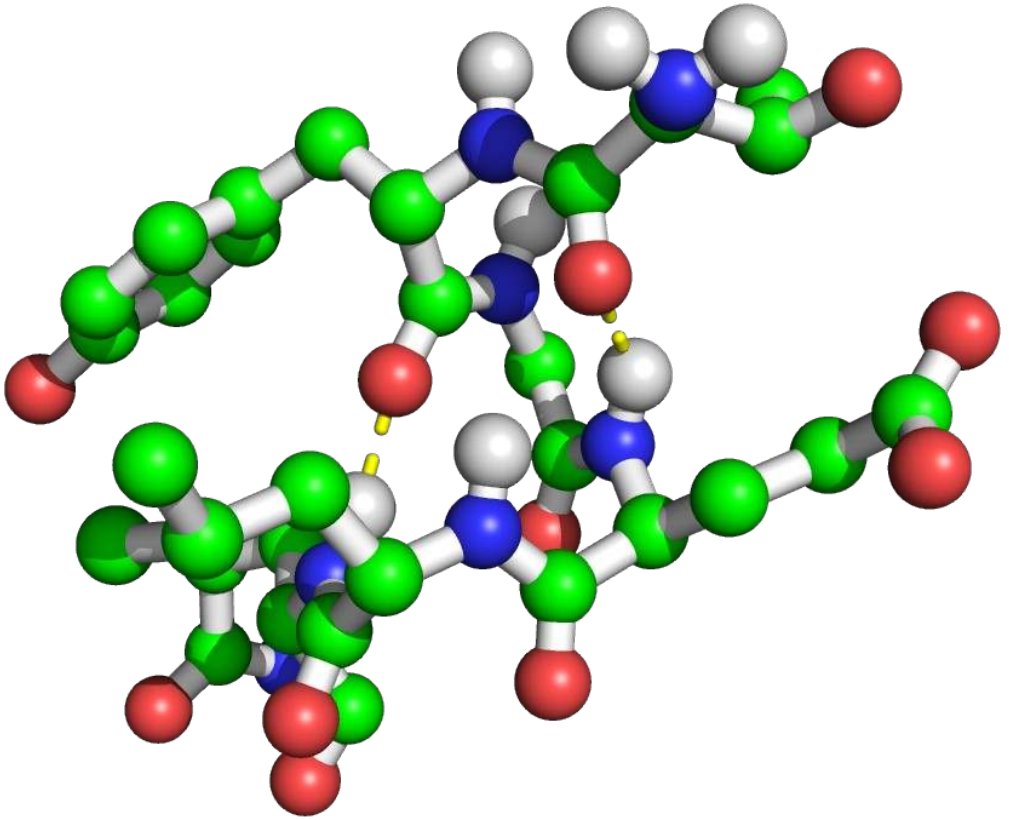
Structure	Amino acid / turn	Pitch (Å)	Donor-acceptor distance *
α-helix	3.6	5.4	13
3_{10}-helix	3	2	10
π-helix	4.4	1.15	16

* The number of atoms in the ring formed by making the hydrogen bond

α-helix



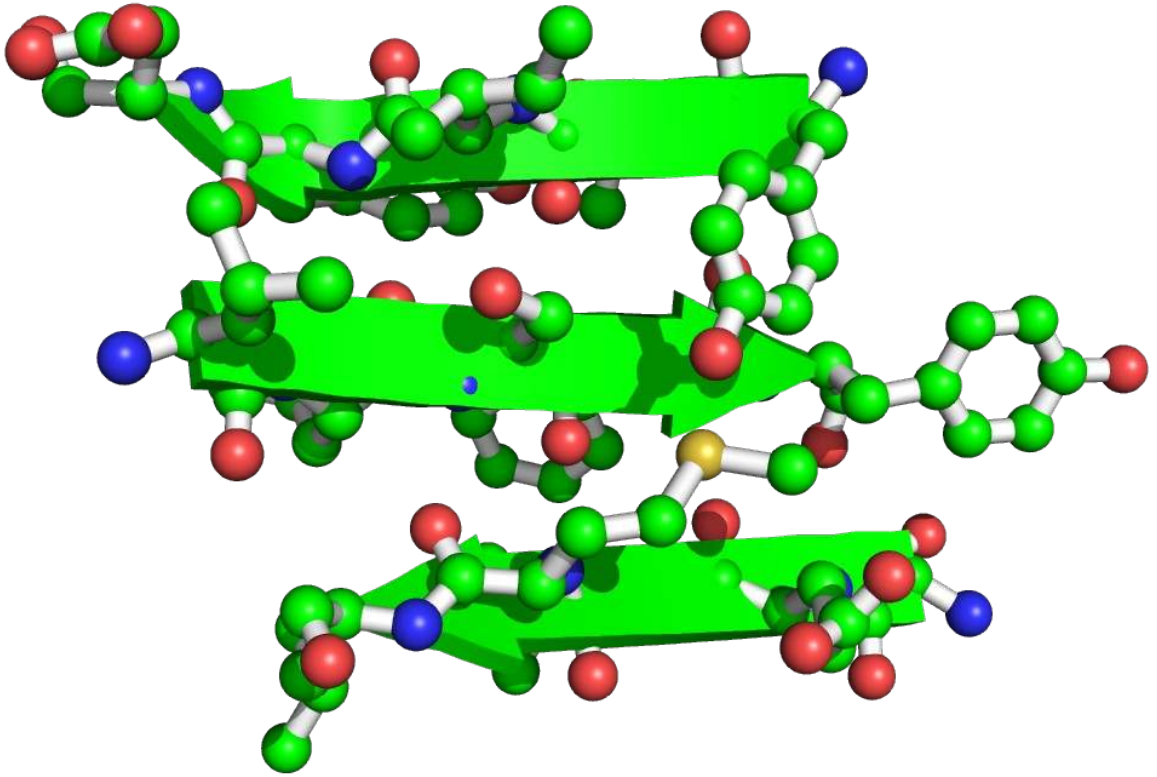
Hydrogen bonds within α-helix



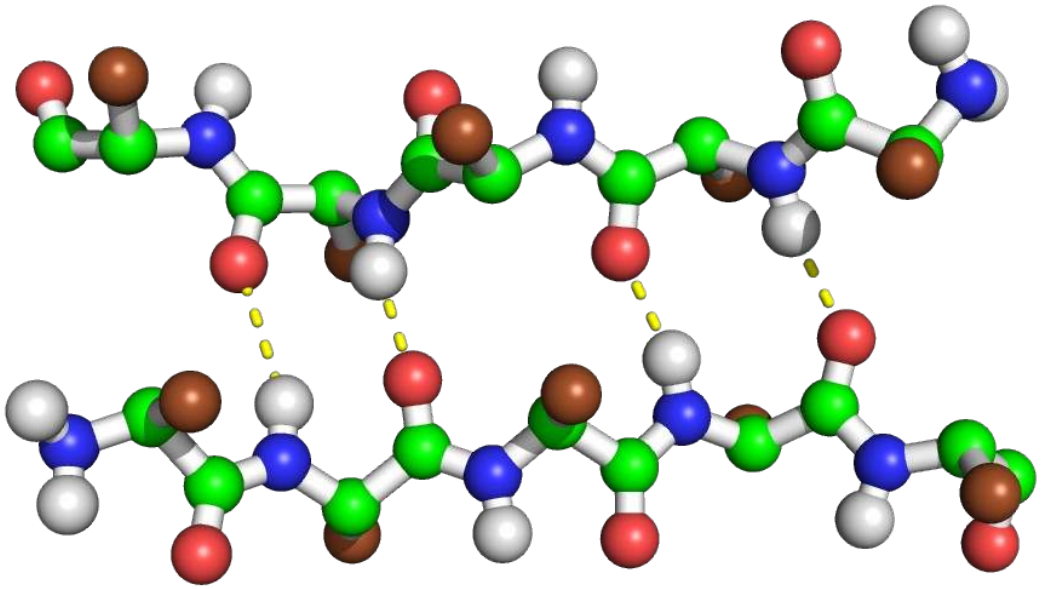
Extended structures

- Extended structures can also be considered as helices because they can be produced by applying translations and rotations
- Extended structures are the most stretched conformations permitted by excluded volume constraints
- The most frequent type of extended structures is the β-strand
- Several β-strands can form a β-sheet within which the strands can be in parallel or antiparallel orientation

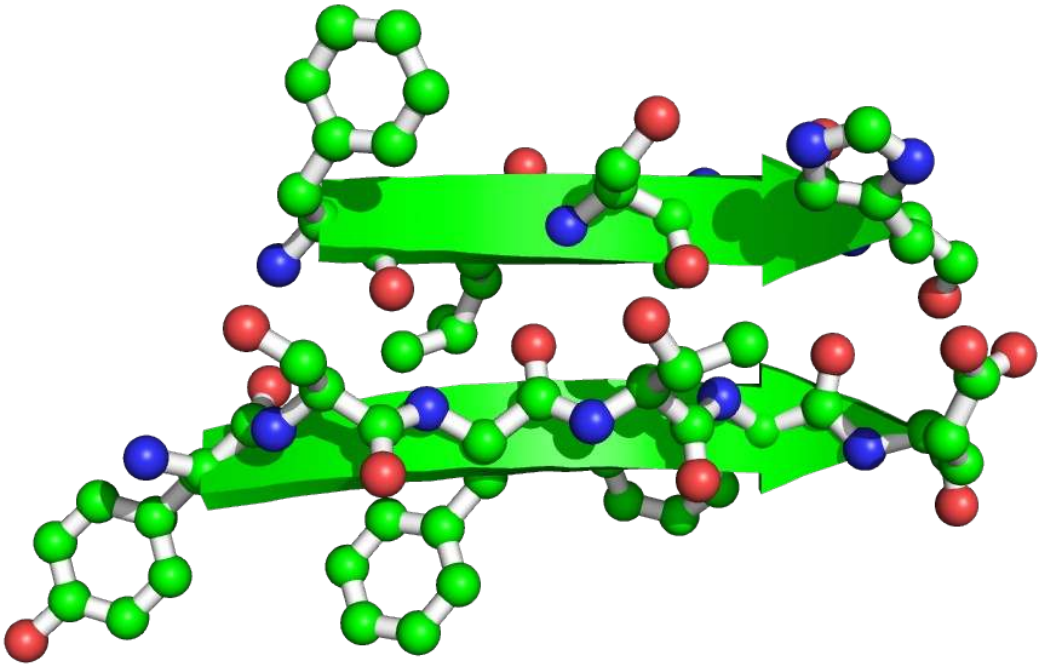
Antiparallel β-sheet



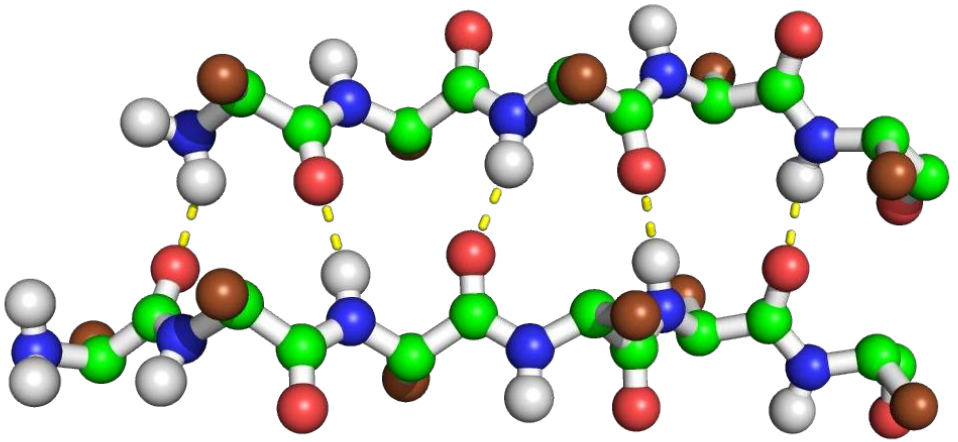
Hydrogen bonds in antiparallel β-sheets



Parallel beta sheet



Hydrogen bonds in parallel β-sheets



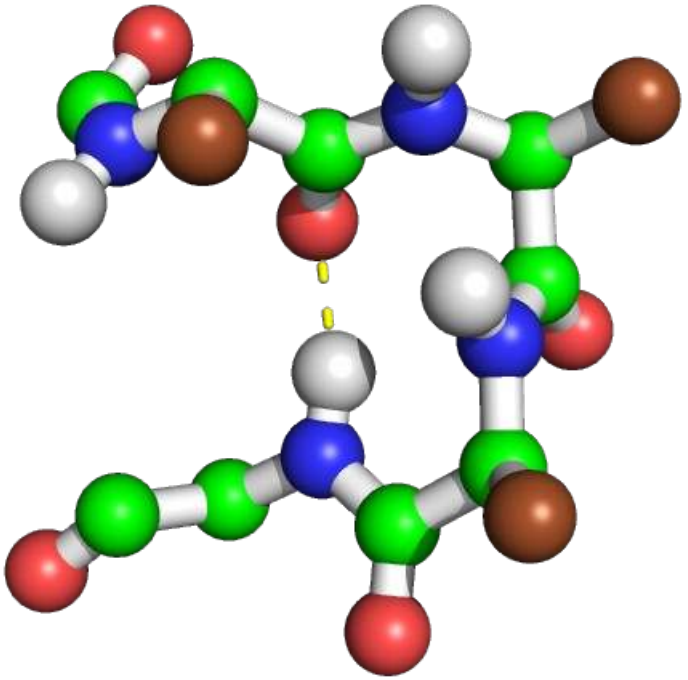
Non-periodic elements: turns and loops

- Turns are short elements usually consisting of only a few residues
- Turns are stabilized by a hydrogen bond formed within the turn
- Loops are longer, extended elements but they do not have a regular structure
- Usually loops are the most flexible parts of proteins

Types of turns

Type	Hydrogen bond
α-turn	between i and i+4
β-turn	between i and i+3
γ-turn	between i and i+2
π-turn	between i and i+5

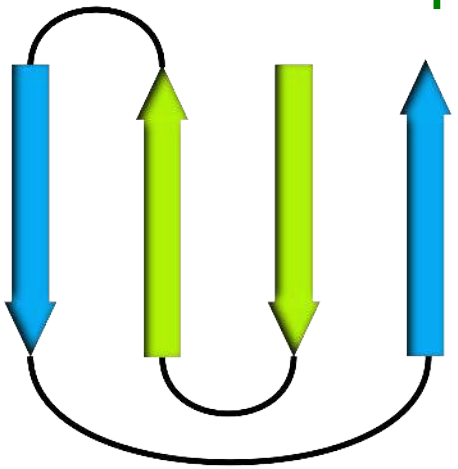
β-turn



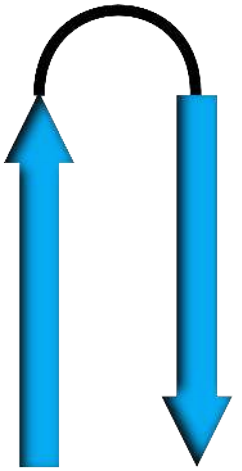
Structural motifs or supersecondary elements

- Secondary structural elements can assemble to bigger units called motifs or supersecondary elements
- Only a small number of such motifs are known but they can be found in a huge number of structures
- Supersecondary elements together can form structural domains

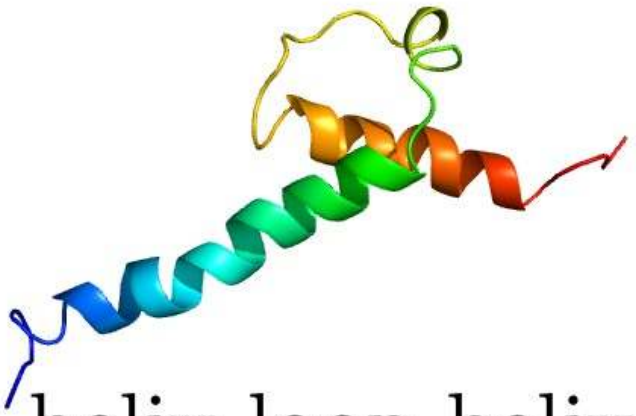
Supersecondary elements



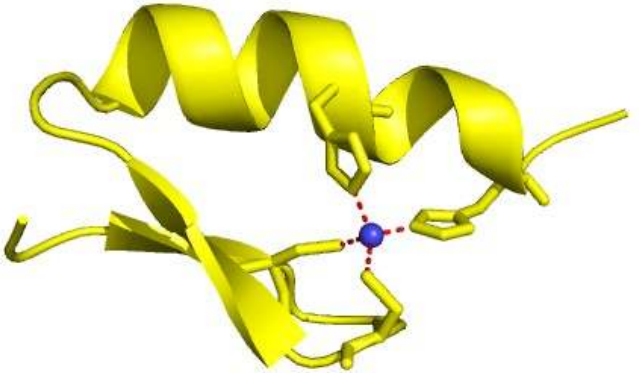
greek key



β hairpin



helix-loop-helix



Zn-finger

Tertiary structure

- The tertiary structure of proteins is the global arrangement of their secondary structure elements, stabilized by interactions between elements far away from each other along the sequence
- By tertiary structure, the main classes of proteins are: globular, fibrillar and membrane proteins
- These types of structures are stabilized by the same covalent bonds and non-covalent interactions

Structural domains

- The most important structural unit of proteins is the domain
- Domains are defined as portions of structure within which there are more interactions than between them
- Structural domains usually correspond to folding units and even functional units
- Structural classifications of proteins are based on domains

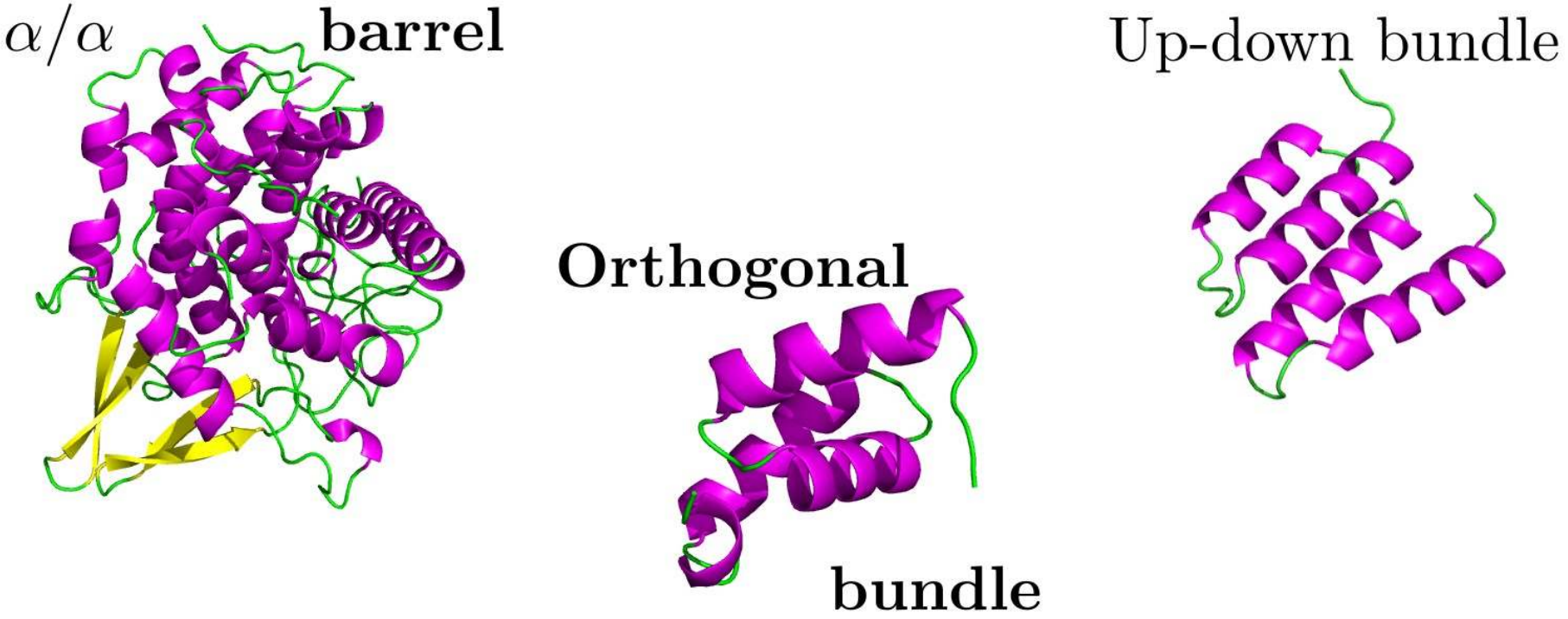
Structural classification of proteins in the CATH database

Class	Number of architectures
Mainly α	5
Mainly β	20
Mixed α-β	14
Few secondary structures	1

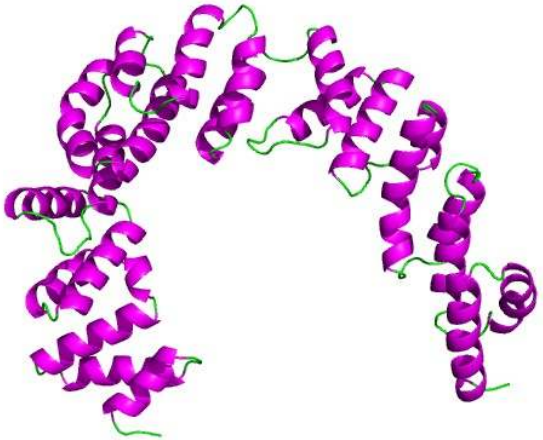
Mainly α architecture types

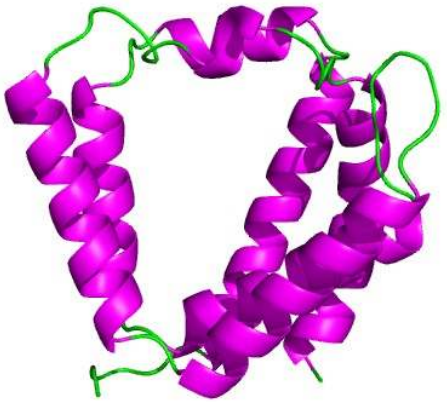
Class	Architecture	Representative domain
Mainly α	Orthogonal bundle	1oai chain A
	Up-down bundle	1mz9 chain A
	α horseshoe	1jdh chain A
	α solenoid	1ppr chain M
	α/α barrel	1h12 chain A

Mainly α architecture types



α horseshoe





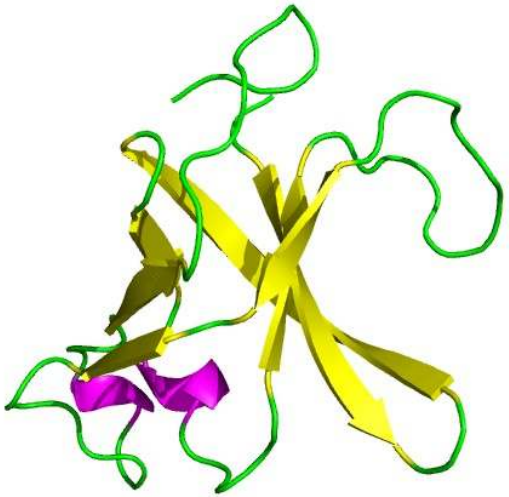
α solenoid

Class	Architecture	Representative domain
Mainly β	Ribbon	1h8p chain A
	Single sheet	1bds chain A
	Roll	1nh2 chain D
	β Barrel	1gvk chain B
	Clam	4bcl chain A
	Sandwich	2hnu chain A
	Distorted sandwich	1m3y chain A
	Trefoil	1rg8 chain A
	Orthogonal prism	2dpf chain A
	Aligned prism	1i5p chain A

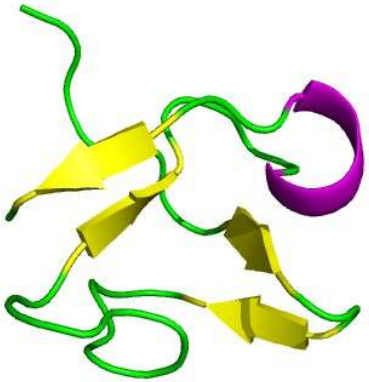
Architecture	Representative domain
3-layer sandwich	2bmo chain A
3 propellor	1n7v chain A
4 propellor	3c7x chain A
5 propellor	1tl2 chain A
6 propellor	3sil chain A
7 propellor	2bbk chain H
8 propellor	1w6s chain A
2 solenoid	1k7i chain A
3 solenoid	3ftt chain A
β complex	1ylh chain A

Mainly β architecture types

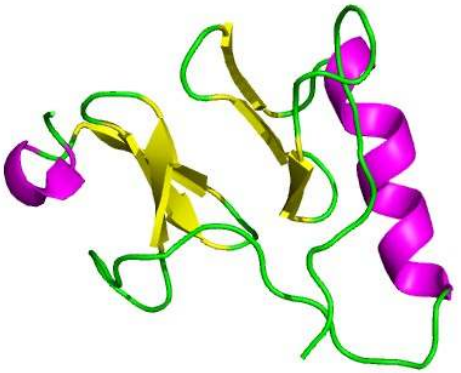
Mainly β architecture types



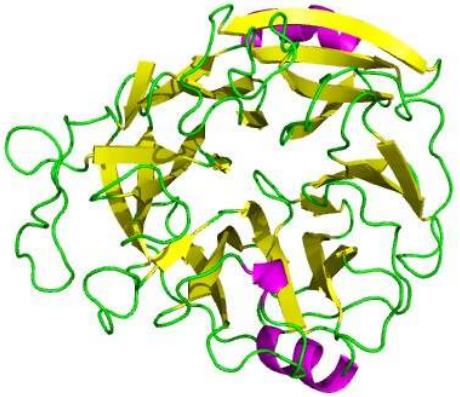
β barrel



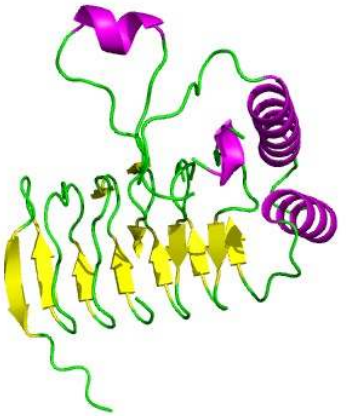
Ribbon



Sandwich



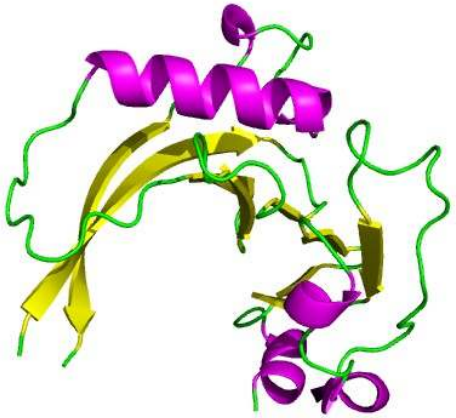
6 propellor



3 solenoid

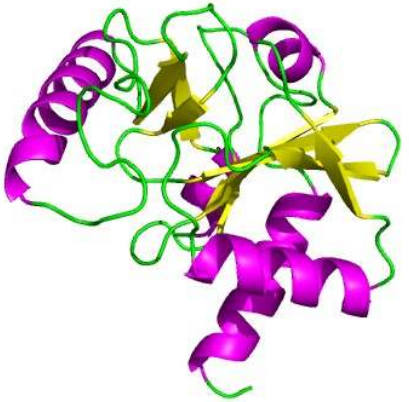
Class	Architecture	Representative domain
Mixed α-β	Roll	3dlk chain B
	Super roll	1ewf chain A
	α-β barrel	2eiy chain B
	2-layer sandwich	1c0p chain A
	3-layer(αβα) sandwich	2hba chain A
	3-layer(ββα) sandwich	2qj2 chain A
	3-layer(βαβ) sandwich	1j5u chain A
	4-layer sandwich	1b25 chain A
	α-β prism	1g6s chain A
	Box	1t6l chain A
	5-stranded propeller	1xkn chain A
	α-β horseshoe	1ozn chain A
	α-β complex	1j0p chain A
	Ribosomal protein L15, chain K, domain 2	1vq8 chain O

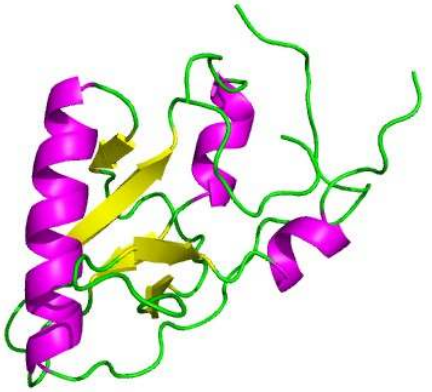
Table 8.



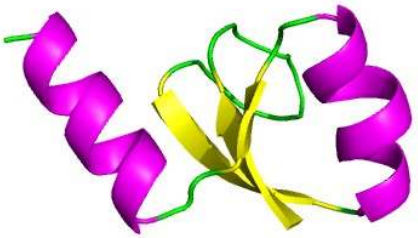
2-layer sandwich

α-β barrel

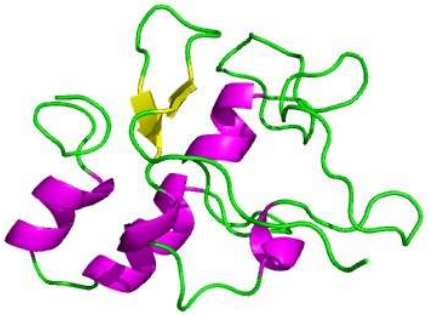




Roll



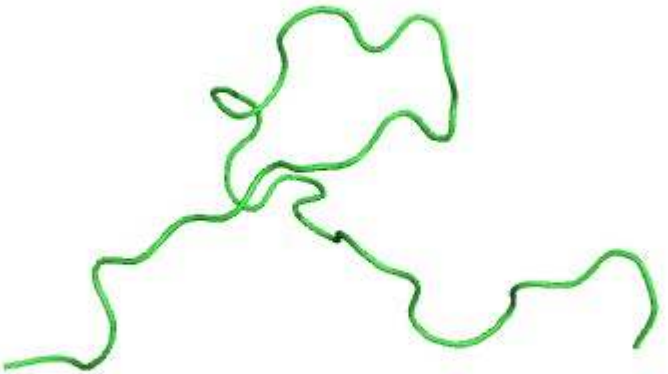
$\alpha\beta\alpha$ sandwich



α-β complex

Few secondary structures

Few secondary

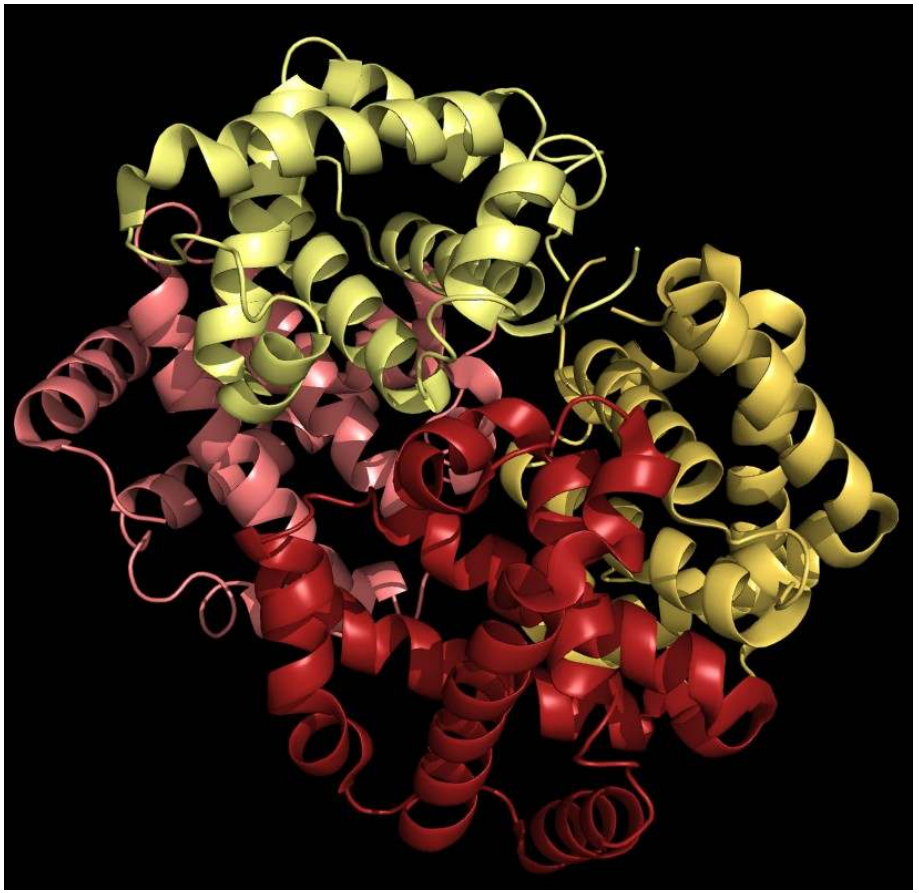


structures

Quaternary structure

- Several protein chains together can form together a non-covalent complex, called an oligomer
- The arrangement of such chains defines the quaternary structure of such proteins
- The sequence of chains can be identical (homooligomers) or different (heterooligomers)

Quaternary structure of hemoglobin



Cofactors

- Some non-protein molecules can be attached to the proteins
- Co-enzymes bind to enzymes via non-covalent interactions
- Prosthetic groups bind tightly or covalently to the protein chain
- In enzymes, the cofactors are responsible for the reaction type and the protein – the apoenzyme – is responsible for the substrate specificity

Interactions stabilizing the structure

Short-range repulsion
Van der Waals interaction
Electrostatic interaction
Hydrogen bond
Hydrophobic effect
Disulfide bond

Table 9.

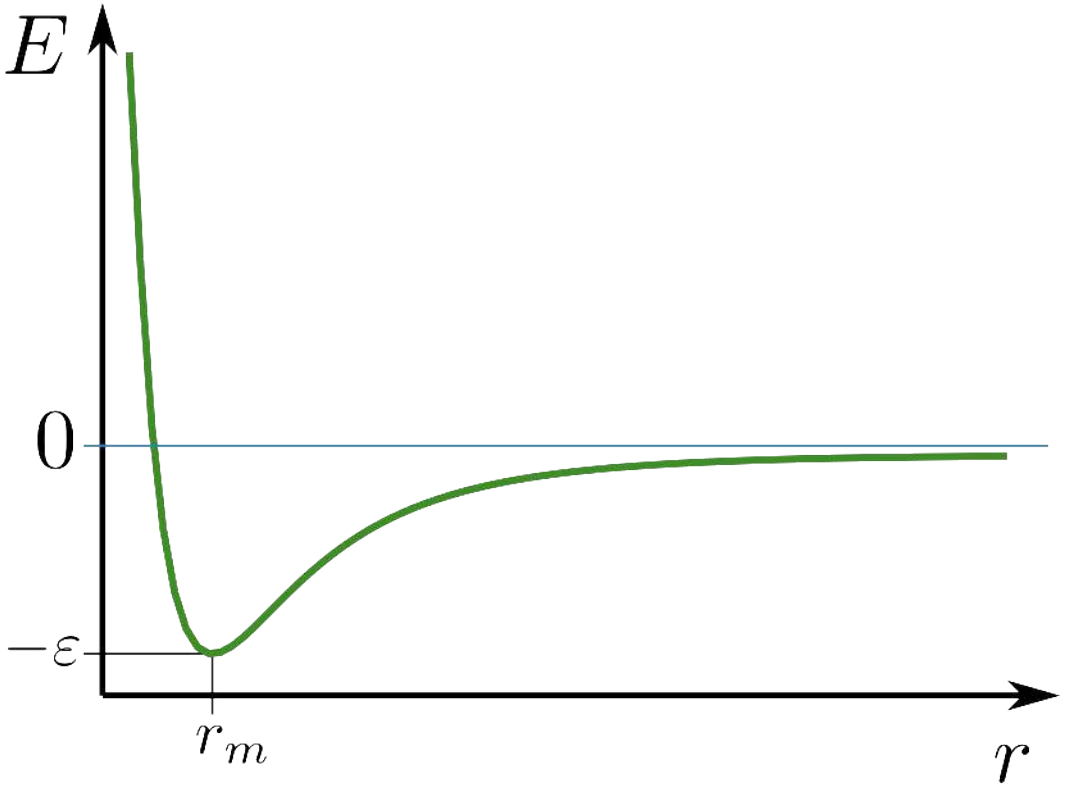
- Short-range repulsion is due to the repulsion of the orbitals of electrons
- Van der Waals interactions are attractive forces between induced dipole moments
- Short-range repulsion and van der Waals attraction can be treated in one expression, the Lennard-Jones potential
- Lennard-Jones potential contains a term corresponding to the r^{-12} repulsion and a term corresponding to the r^{-6} van der Waals attraction

- The Lennard-Jones potential is:

$$V_{LJ} = \varepsilon \left(\left(\frac{r_m}{r} \right)^{12} - 2 \cdot \left(\frac{r_m}{r} \right)^{6} \right)$$

where ε is the depth of the potential well, r is the distance of the two particles and r_m is the distance where the potential reaches its minimum

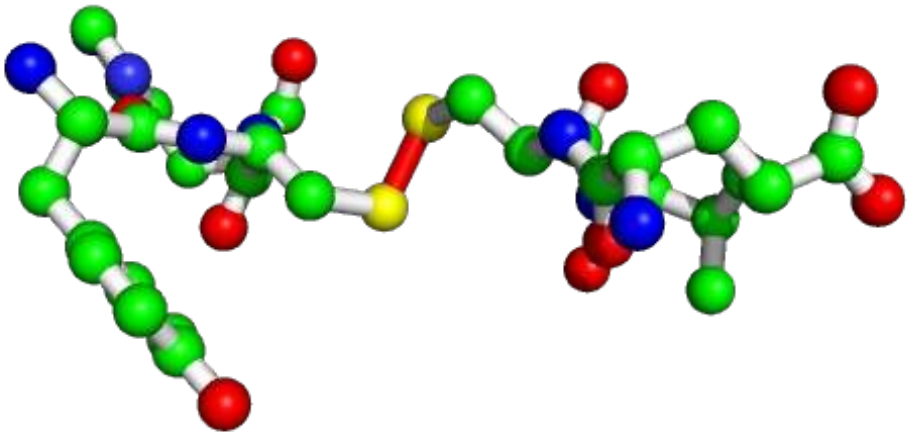
The Lennard-Jones potential



- Electrostatic interactions are formed between a positively charged – for example lysine – and a negatively charged – for example aspartate – residue
- Due to the screening effect of water, the influence of these interactions is restricted to short distances

- The disulfide bond is a covalent bond formed by two SH-groups of cysteine residues
- Because the formation of disulfide bonds requires oxidative conditions, only extracellular proteins have disulfide bonds
- Disulfide bonds have an important role in the stabilization of small proteins

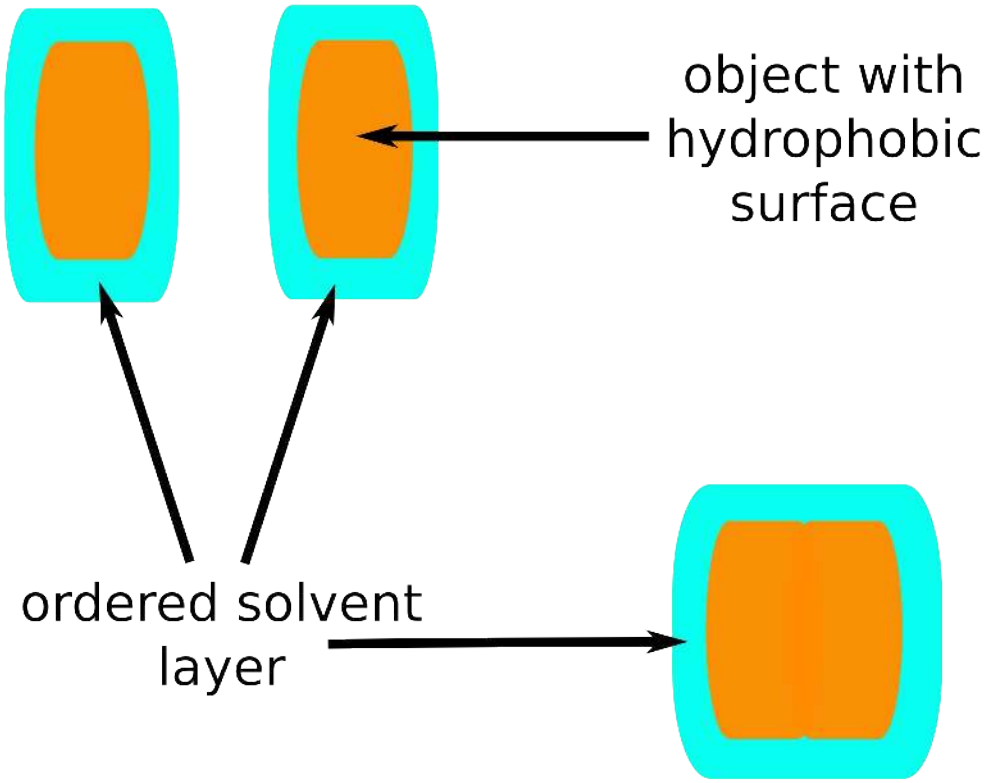
Disulfide bond



- Protein stability and folding are strongly determined by the hydrophobic effect
- The hydrophobic effect is based on an entropy increase upon the association of hydrophobic groups
- Around hydrophobic surfaces, water molecules adopt the ordered arrangement which has a low entropy
- Reducing hydrophobic surface allows the water molecules to be released, accompanied by an entropy increase

- The increase in solvent compensates for the decrease in the entropy of the protein chain
- Increasing the entropy decreases the free energy
- Due to hydrophobic effect, residues with hydrophobic side chains collapse to a core of the structure of protein
- So water-soluble proteins have a hydrophobic core and a polar surface
- Proteins with hydrophobic side chains on their surface have an intrinsic propensity for aggregation

The hydrophobic effect



Protein folding

- It is known since the famous experiment of *Anfinsen* that the primary structure of proteins determines the spatial structure under the given conditions
- The structure in which proteins can perform their physiological function is called the native state
- The native state corresponds to the global free energy minimum
- The process through which protein chains reach their native state is called folding

Christian B. Anfinsen 1916-1995)

Levinthal's paradox

- Native state of proteins corresponds to the global free energy minimum
- Assuming only three distinct conformational states per residue, and time of 10^{-13} seconds to switch between states, it would take $1.6 \cdot 10^{27}$ years for an only 100 residue long peptide chain to reach its state with minimal free energy
- In real proteins, this time would be far longer because of the practically infinite number of conformers per residue

- Contrary to this, proteins in nature reach their native state in at most a few seconds
- Based on his calculation, Levinthal postulated paths by which a protein folds and assumed that not only the final structure but also the path to it is encoded in the primary structure
- The funnel-shaped landscape of proteins will resolve this paradox

Statistical mechanical description of protein structure and folding

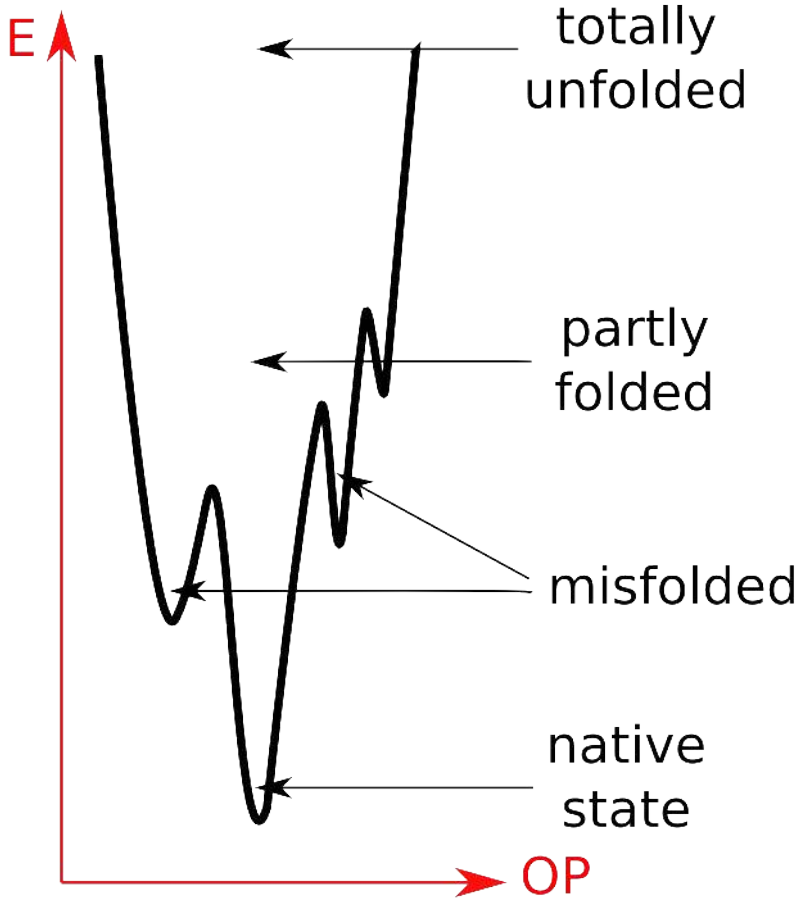
- The protein chain and the solvent surrounding it can be considered as a subsystem which is in thermal equilibrium with the environment
- Thus, the Boltzmann distribution is valid for the microstates of the subsystem consisting of the protein chain and the solvent
-

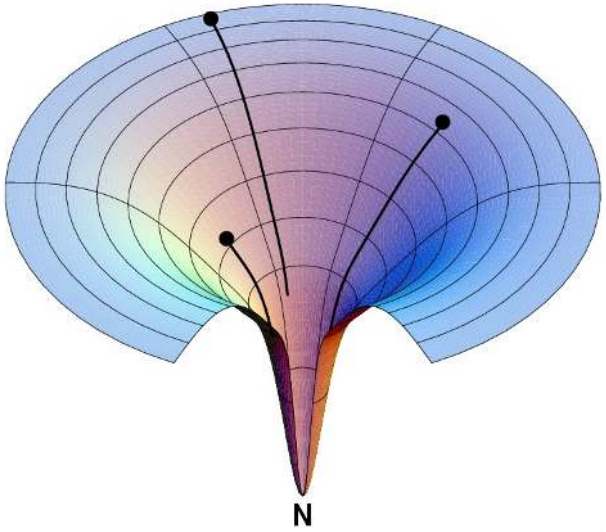
- Although the native state contains only one or a few chain conformations and thus relatively few microstates correspond to it, it can be the most favourable state because of its low energy
- The protein chain-solvent system has many degrees of freedom so we are forced to make the model of it simpler
- Chain conformation can be described by the positions of its atoms
- Many solvent arrangements belong to a single chain conformation

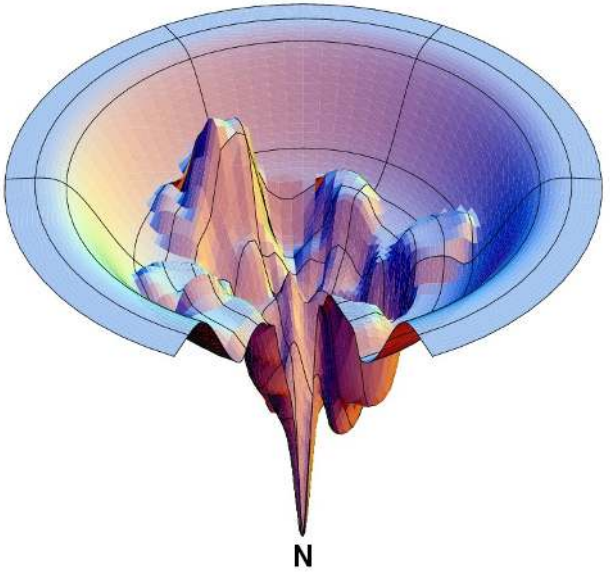
- Potential energy functions can be constructed to determine the energy of a given conformation so the energy is the function of degrees of freedom
- If we average the energy over the solvent arrangements, we can consider only the degrees of freedom of the chain itself
- Degrees of freedom can be merged to a few order parameters
- Energy can be plotted as a function of degrees of freedom or as a function of order parameters

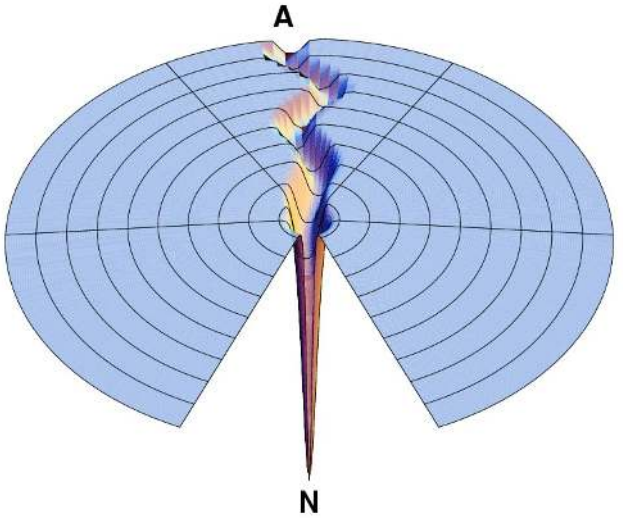
- The shape of such surfaces resembles a funnel
- Levinthal's paradox can be resolved based on the funnel-shaped energy landscape
- Folding can proceed through several paths but every path has its end on the bottom of the funnel

2D folding funnel









Dill KA, Chan HS. 1997. From Levinthal to pathways to funnel. *Nat. Struct. Biol.* 4(1):10-9.

- Some mesostates, containing one or more microstates can be defined and a *partition function* can be calculated for it
- The free energy of mesostates can be calculated from the partition function by

$$F = -kT \ln Q$$

where Q is the partition function and F is the free energy

- If we plot energy as a function of degrees of freedom we obtain a free energy surface

Free energy profile of two-state folding