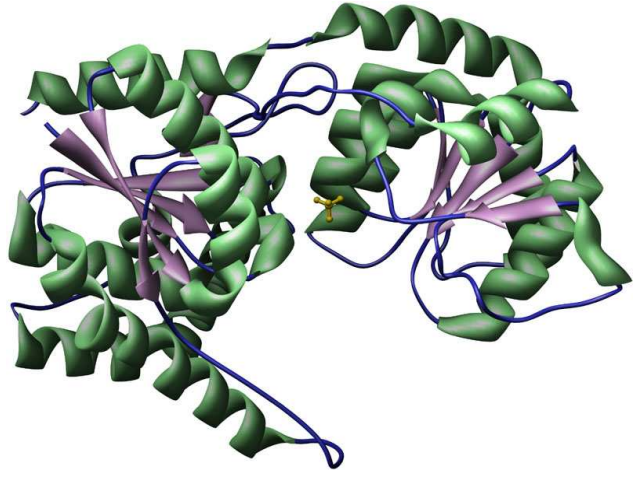
Modelli semplici di interesse biologico: dalle proteine ai neuroni

L. Bongini[(2)], Th. Kreuz[(1)], A. Politi[(1)], R. Zillmer[(3)] & A. Torcini[(1)]

(1) Istituto dei Sistemi Complessi - CNR - Firenze - Italy

(2) Dipartimento di Fisica - Firenze

(3) INFN - Firenze



Programma

ore 14:30-16:30 - Aula 2 - Area CNR

- mer 26/4 Introduzione su proteine e ripiegamento delle proteine (Torcini)
- ven 28/4 Modelli semplici (Bongini)
- mar 02/05 Introduzione sui neuroni e loro caratteristiche (Torcini) ***
- mer 03/05 Modelli di singolo neurone quali paradigmi di sistemi eccitabili: Fitz-Hugh Nagumo e Hindmarsh-Rose (Zillmer)
- ven 05/05 Effetti di coerenza indotti dal rumore (Kreuz)
- lun 08/05 Una introduzione alle reti neuronali (Politi)

[*** ore 16:30-18:30 - Aula 2 - Area CNR]

Il ripiegamento delle proteine

- La proteina in breve (vista da un fisico)
- Struttura primaria, secondaria, terziaria
- Termodinamica del ripiegamento
- Il dogma di Anfisen
- Il paradosso di Levinthal
- L'imbuto energetico

Copia di questa e delle seguente lezioni:

www.fi.isc.cnr.it/∼torcini

Le proteine . . .

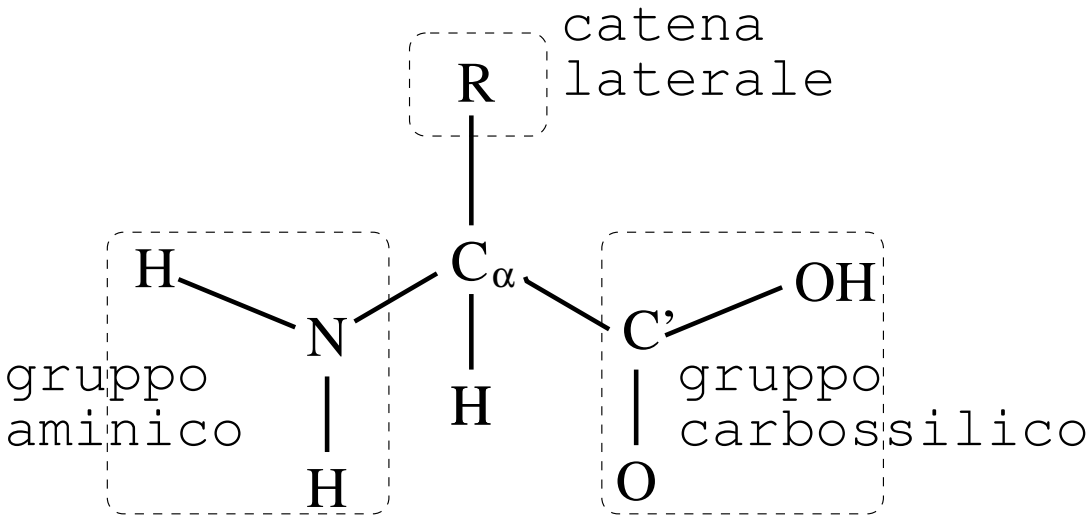
Le proteine sono macromolecole costituite da una sequenza di aminoacidi , la cui funzionalitá biologica é strettamente legata alla loro struttura tridimensionale (alla configurazione che assumono). Tipicamente presentano un sito attivo che interagisce con specifiche molecole o particelle e ne determina la funzione biologica specifica.
Tutte le proprietá legate ai processi vitali sono influenzate dalle proteine.

- Le proteine immagazzinano e trasportano una varietá di particelle (dagli elettroni a macromolecole), ad esempio controllano il passaggio ionico attraverso la membrana cellulare;
- Sotto forma di ormoni trasmettono le informazioni tra cellule specifiche ed organi piú complessi;
- Le proteine guidano il flusso di elettroni nel processo della fotosintesi;
- Gli anticorpi sono proteine.

É quindi di primario interesse capire come certe strutture tridimensionali si creino a partire dalla sequenza degli aminoacidi che codifica ogni proteina.

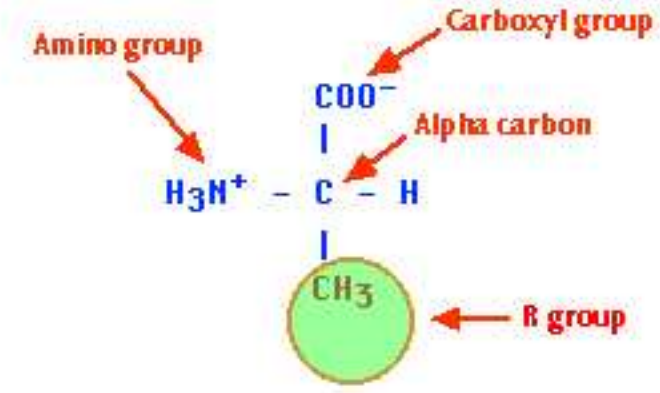
Building blocks of proteins

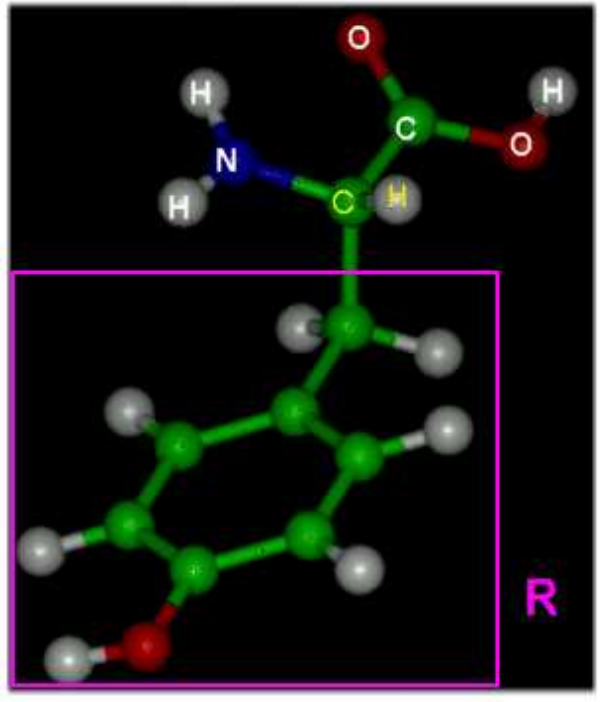
Proteins are heteropolymeric chains made up of monomers, called amino acids.



There are over 300 naturally occurring amino acids on earth, but the number of different amino acids in proteins is only 20 .

Building blocks of proteins



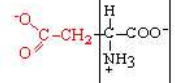
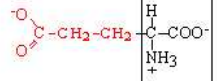
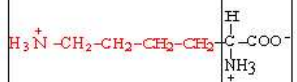
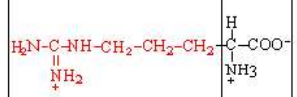
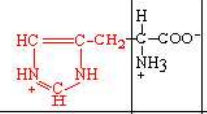
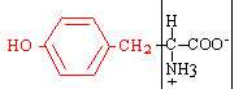
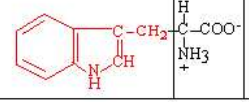


Alanine (Ala) - $m = 71$ UMA - $V = 67$ A^3 – Tyrosine (Tyr) - $m = 163$ UMA - $V = 141$ A^3 -
(1 UMA= 1.6 10^{-24} gr)

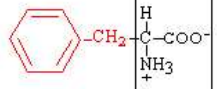
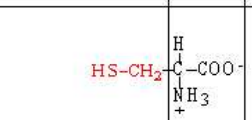
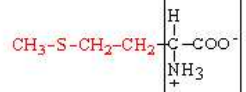
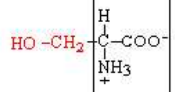
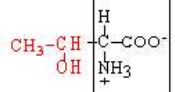
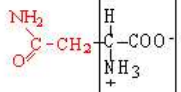
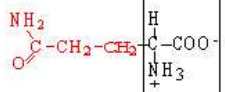
R = Side Chain: characterizes the different amino acids (20 types)
3 groups: Hydrophobics (H) , Polars (P) and Charged Polars
Their frequency in the proteins is 40%, 30 % and 22 %, respectively.

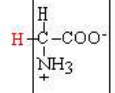
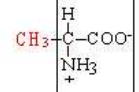
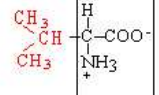
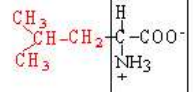
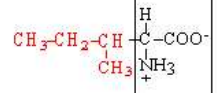
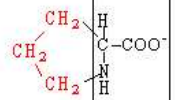
Building blocks of proteins

- The hydrophobicity index tells the relative hydrophobicity among amino acids.
- Strongly positive values indicate hydrophobic (H) amino acids;
- Polars (P) amino acids have negative values;
- Neutral (N) amino acids are characterized by small values of the index.
- In a protein, H amino acids are more likely to be located in the protein interior , whereas P amino acids are more likely to face the aqueous environment .

Name	Symbol		R group	Hydro-phobicity
	3 Lett.	1 Lett.		
Aspartate	Asp	D		-3.5
Glutamate	Glu	E		-3.5
Lysine	Lys	K		-3.9
Arginine	Arg	R		-4.5
Histidine	His	H		-3.2
Tyrosine	Tyr	Y		-1.3
Tryptophan	Trp	W		-0.9

Building blocks of proteins

Phenylalanine	Phe	F		2.8
Cysteine	Cys	C		2.5
Methionine	Met	M		1.9
Serine	Ser	S		-0.8
Threonine	Thr	T		-0.7
Asparagine	Asn	N		-3.5
Glutamine	Gln	Q		-3.5

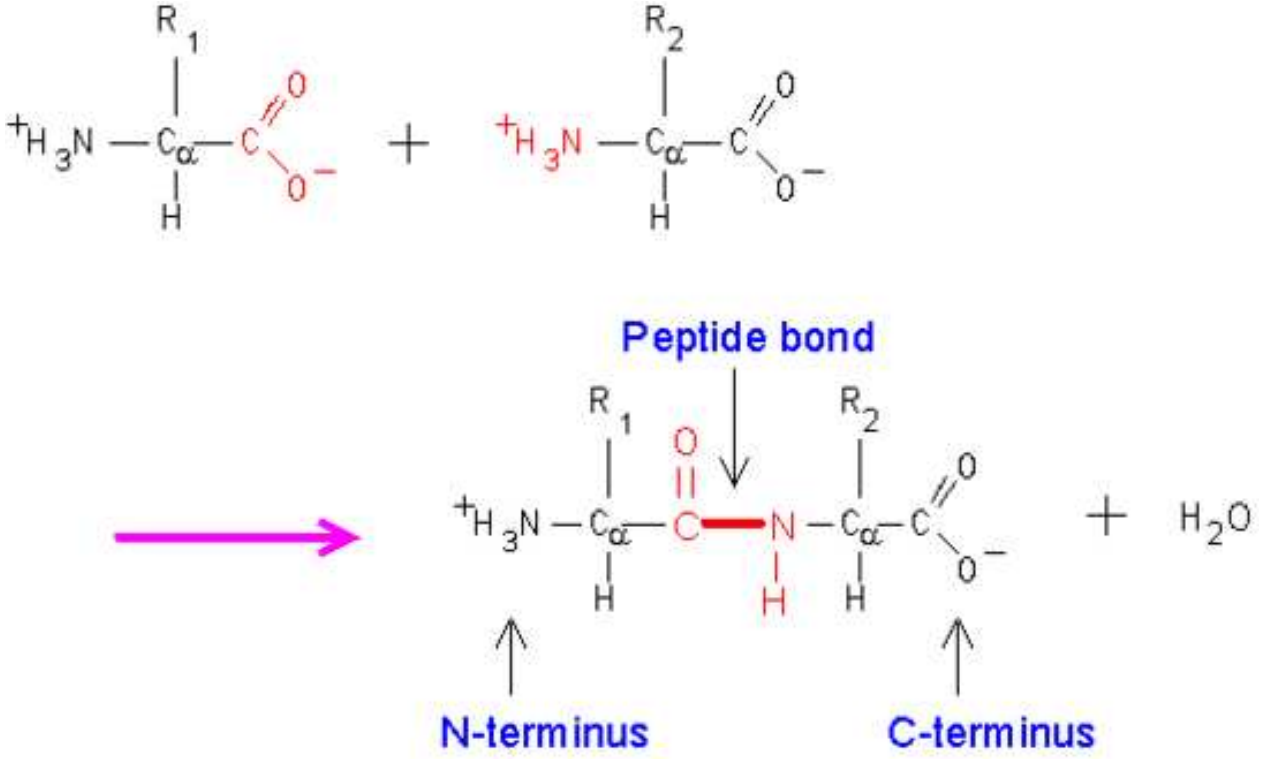
Glycine	Gly	G		-0.4
Alanine	Ala	A		1.8
Valine	Val	V		4.2
Leucine	Leu	L		3.8
Isoleucine	Ile	I		4.5
Proline	Pro	P		-1.6

In most amino acids, the R group and the amino group are not directly connected.

Proline is the only exception, due to this special feature, proline is often located at the turn of a peptide chain in the three-dimensional structure of a protein.

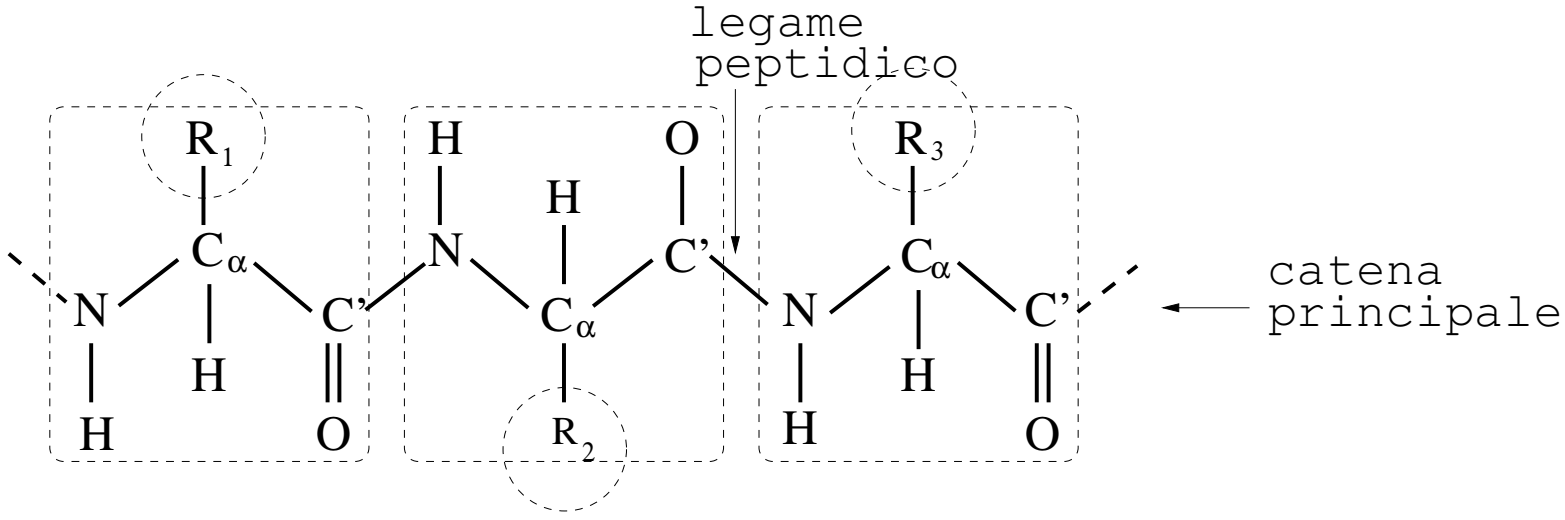
The primary structure

The peptide bond



The primary structure

The (poly)-peptide



- The peptide is a chain of amino acids linked together by peptide bonds .
- Polypeptides usually refer to long peptides whereas oligopeptides are short peptides (< 10 amino acids).
- Proteins are polypeptides with a well defined 3D structure under physiological conditions.
- Their size can vary from 50 to 25,000 residues (amino acids), the average size is 250 residues .

The primary structure

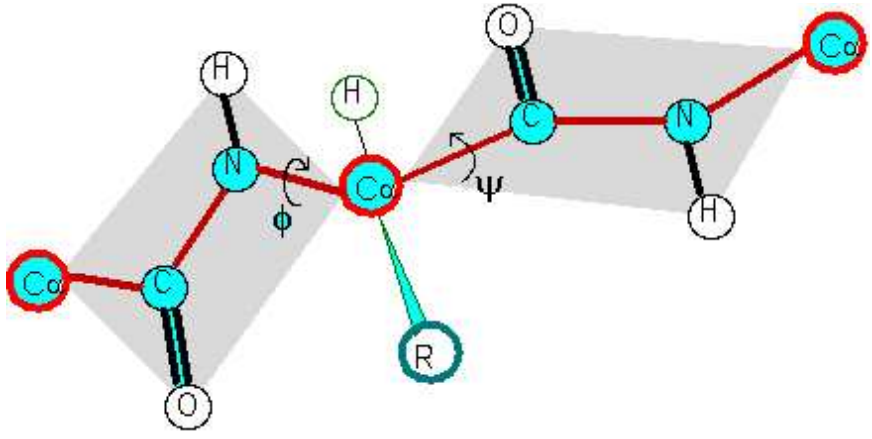
[Ribonuclease A (RNase A), which is an enzyme acting on RNA]

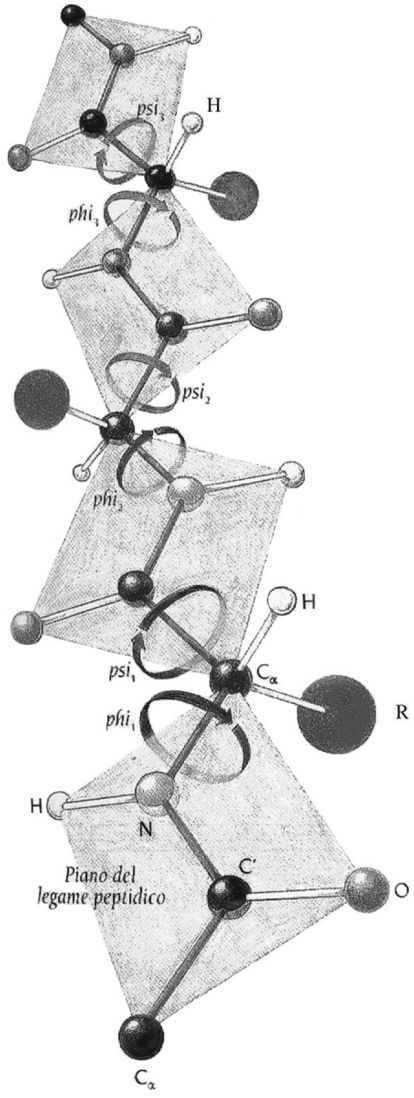
- The primary structure of a protein is synthetized by following the instruction encoded in the genetic material of the chromosomes (DNA or even RNA (viruses)).
- The specific part of DNA (or RNA) that codes for the primary structure of a certain protein is called gene.
- For every protein there is a gene coding for its synthesis, and most genes are unique; only rarely one ore more genes produce the same protein.

The secondary structure

The dihedral angles

- Due to the specific electronic structure of the peptide bond, the atoms on its two ends cannot rotate around the bond.
- Hence, the atoms of the group, O=C-N-H, are fixed on the same plane, known as the peptide plane .
- The whole plane may rotate around the N-C_α bond (ϕ angle) or C-C_α bond (ψ angle).
- C_α is the carbon atom connected to the residue, these atoms constitutes the protein backbone .
- (ϕ,ψ) = Dihedral Angles : identify the structure of the backbone

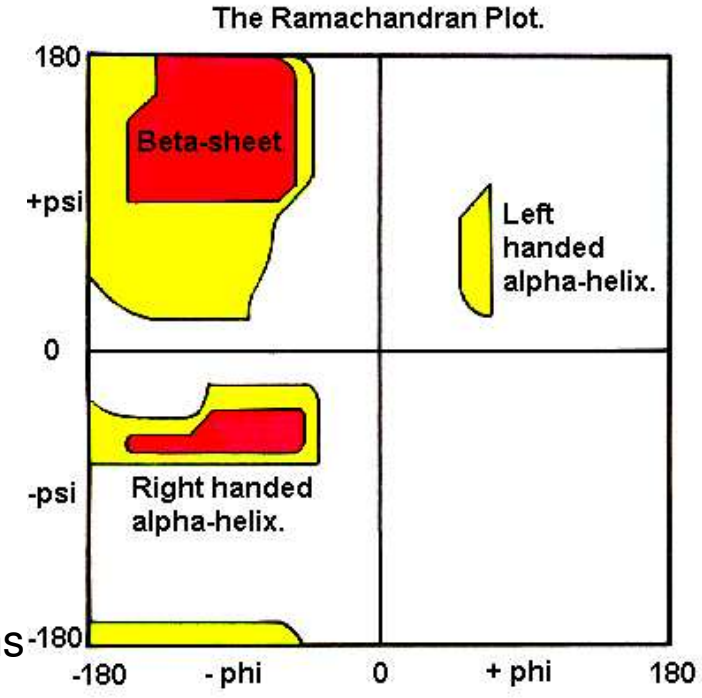




The secondary structure

The Ramachandran plot

- Not all dihedral angles are allowed;
- due to steric clashes some configuration is prohibited;
- Repeating values of ϕ and ψ along the chain result in regular structure.
- For example, repeating values of $\phi \sim -57$ and $\psi \sim -47$ give a right-handed helical fold (the alpha-helix).
- Repetitive values in the region of $\phi = -110$ to -140 and $\psi = +110$ to +135 give extended chains with conformations that allow interactions between closely folded parallel segments (beta sheet structures).



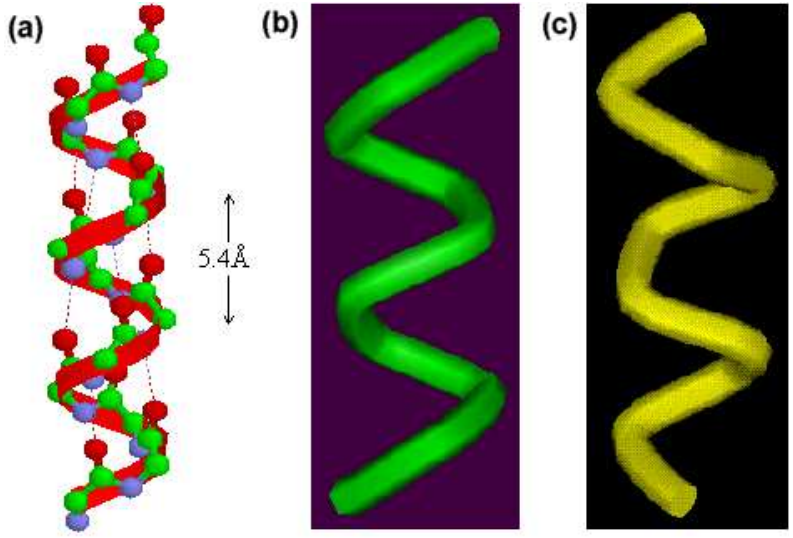
The secondary structure

The Alpha-Helix

A alpha-helix has the following features:

- every 3.6 residues make one turn,
- the distance between two turns is 5.4 A,
- the C=O (or N-H) of one turn is hydrogen bonded to N-H (or C=O) of the neighboring turn.

An a helix can be either right-handed or left-handed, as defined in the following figure.

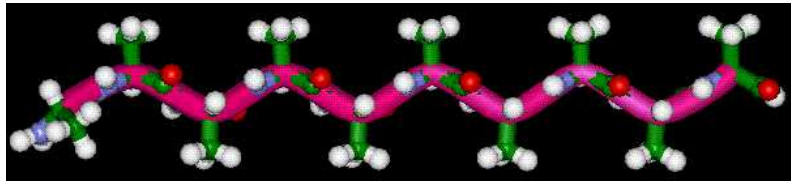


The secondary structure

The Beta Sheet

Beta strand

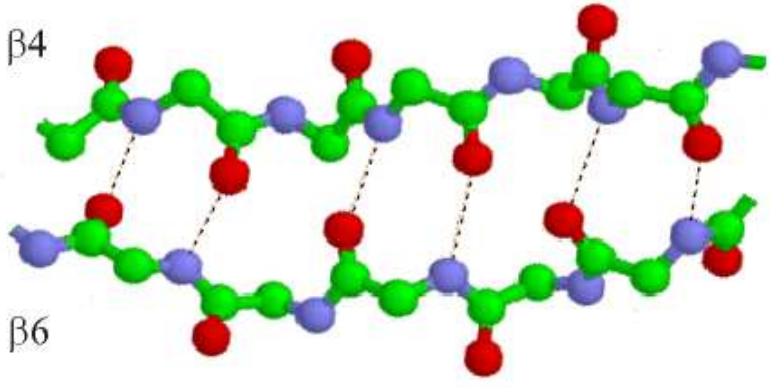
The torsion angle of N-C$_\alpha$-C-N in the backbone is about 120 degrees.



The sidechains of two neighboring residues project in the opposite direction from the backbone.

Beta sheet

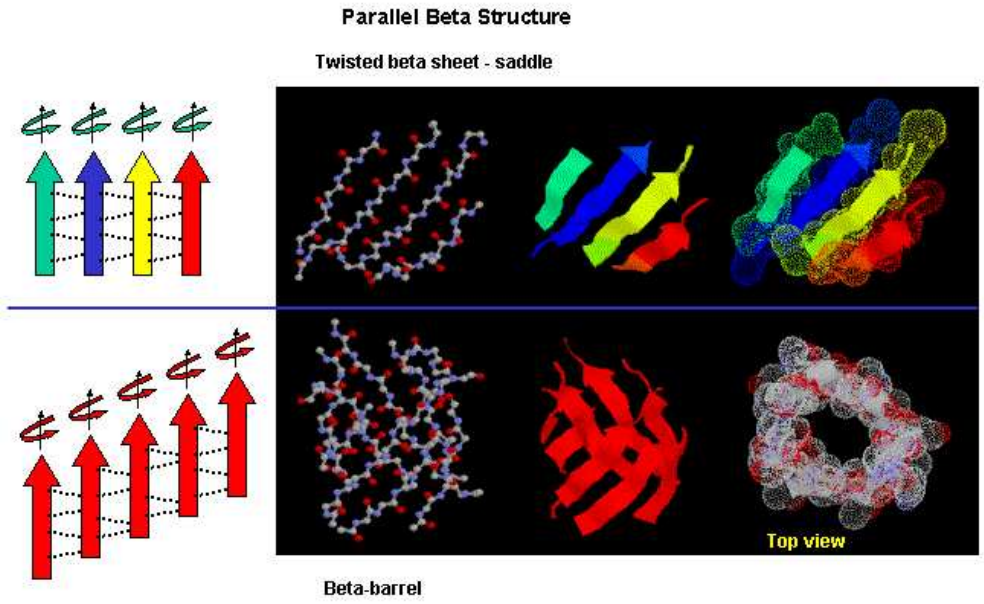
A β-sheet consists of two or more hydrogen bonded β-strands.



The two neighboring β strands may be parallel if they are aligned in the same direction from one terminus (N or C) to the other, or anti-parallel if they are aligned in the opposite direction.

The secondary structure

Beta Motifs



In a protein with parallel strand in phase, and given the inherent twist in the stands, the strands arrange in a twisted saddle shape (top structure above).

Twisted beta sheet from arabinose binding protein

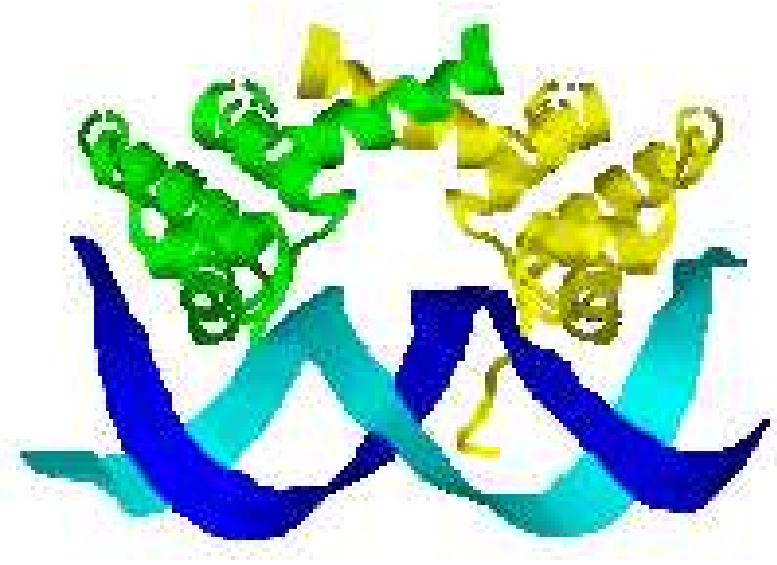
In a protein with parallel strand out of phase, and given the inherent twist in the stands, the strands arrange in a beta barrel (bottom structure above).

Beta barrel from triose phosphate isomerase

The secondary structure

Motifs

A motifs is a combination of secondary structures, that is found in several different proteins. Its identification is quite importante because they are associated with particular function.

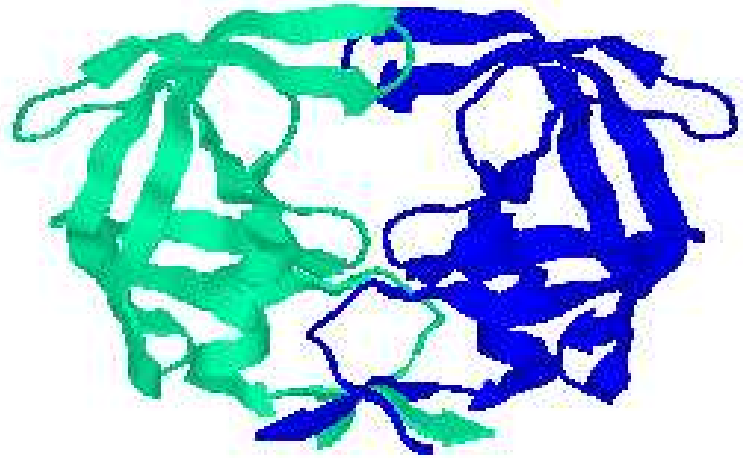


The most famous is the helix-turn-helix motif, that is associated with binding to calcium or DNA.

Tertiary structure

The tertiary structure is given by the combination of several motifs forming or or more domains . Domains are functional subunit of globular form.

Many proteins have a globular form and are formed by a single domain, however a certain number is constitued by more domains defining a quaternary structure .



Some proteins are composed by identical domains, a simple example is the dimer of HIV Protease.

Protein Folding in a Nutshell

Anfinsen's thermodynamic principle

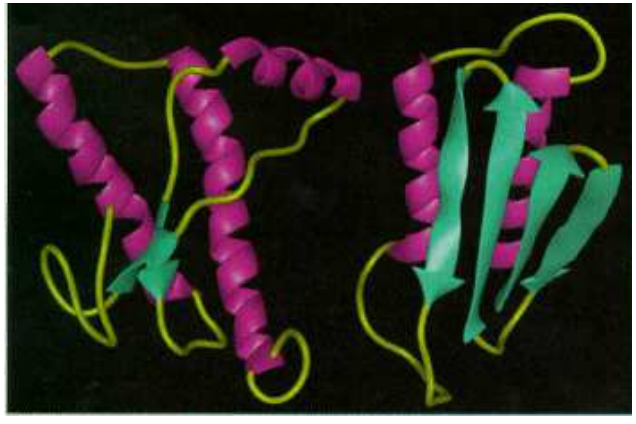
- The primary sequence is produced by the RNA messenger as a linear sequence of amino acids. All the information concerning the folding process are already presnt in the primary structure.
- After the synthesis this sequence folds in an unique tertiary structure the Native Conformation under physiological conditions (acqueous solution, 37 C, pH 7, atmospheric pressure).
- The sequence folds in its native configuration in a relatively short time (10^{-3}-100 sec).
- The Native Conformation is a stable tertiary structure that determines the protein's biological function This state corresponds to a minimum in the free energy of the system, it is favourite in a thermodynamics sense .
- The protein can denaturate (loose its specific 3d structure and unfolds) due to changes in the enviroment conditions: e.g. temperature and pH However, if the physiological conditions are newly established the chain will fold again the the same structure with the same functionality. The Native Conformation is marginally stable.

(Anfinsen - Nobel Laureate in Chemistry - 1972)

Protein Folding in a Nutshell

Anfinsen was not completely right (wrong)

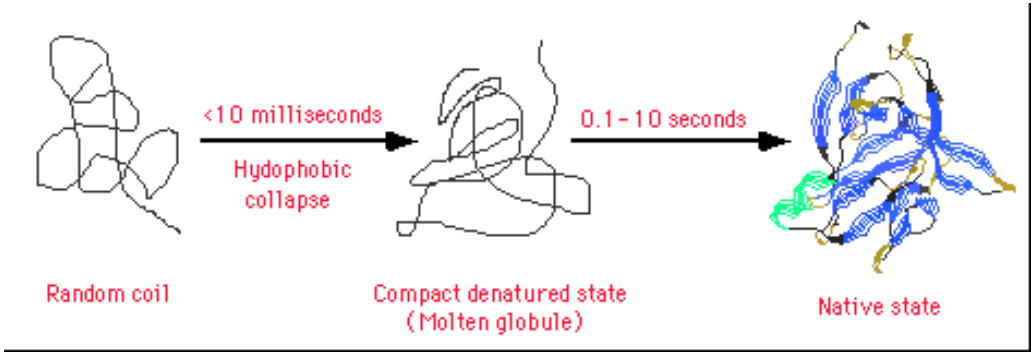
- Prion diseases (like Creutzfeld-Jakob diseases and bovine spongiform encephalopathy) are related to misfolded configurations of the prion protein Prp. The prion propagation is related to a change of PrP from a structure dominated by α-helices to one where β-sheets are predominant.



- For small (water soluble) globular proteins (< 300 residues) it is indeed common to observe unasstisted refolding ; while for larger proteins, spontaneuous refolding does not occur *in vitro*. Between 10 % to 15 % of proteins need the help of molecular chaperons to stabilize their structure in the folded state . They achieve this task by binding sequences of amino acids with hydrophobic side chains, which are no longer exposed when the protein has folded correctly.

Protein Folding in a Nutshell

Folding Phenomenology for globular (small) proteins



- Thermodynamical Aspects The folding process is highly cooperative
- Typical Folding Times $10^{-2} - 10^{2}$ sec
- Current Opinions
  - the folding process is driven and stabilized by the Hydrophobic Interactions
  - Native State
    - Absolute Minimum of the Free Energy $\rightarrow$ Statistical mechanics Approach
    - Dynamically Accessible $\rightarrow$ Study of the Dynamics

Interactions within the protein

Many different interactions are present between atoms in the protein, and also between atoms and the solvent (water), which are the important ones for folding ?

Covalent interactions

The peptide bond is a covalent interaction and it is reasonably strong, this interaction is not modified by the temperature (at the considered temperatures) and therefore is not relevant for folding. ($E \simeq 2.5$ eV)

Non-covalent interactions

- Hydrogen bonds (dipole-dipole interactions arising when two electronegative atoms compete for the same hydorgen atom);
- Electrostatic interactions among charged atoms follows the Coulomb law, long range interactions;
- van der Waals interactions among non charged atoms, repulsive at short distances and attractive at long distances.

These interactions are relevant for folding $E \simeq 0.01$ - 0.3 eV and are thermically excitable at room temperature.

Interactions within the protein

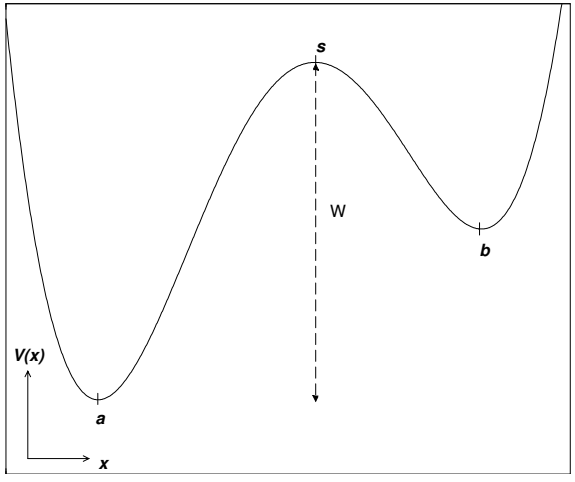
Hydrophobic interactions

- The non-covalent interactions are not particularly favoured in water, since there are comparable competing interactions among atoms (molecules) and the water surrounding them.
- However nonpolar molecules (hydrophobic ones) cannot partecipate to the formation of hydrogen bonds (extremely important in liquid water), therefore water molecules tend to avoid to have bonds with them and as a final effect the nonpolar molecules are confined among them far from water. This give rises to an effective long range hydrophobic attraction among these molecules ($E \simeq 0.08$ eV)
- The hydrophobic interactions are considered to be the responsible for the stability of proteins and for their folding. For sure they are at the basis of the fast hydrophobic collapse taking place in the first phase of folding and leading to the formation of a hydrophobic core of the protein.
- The single non covalent and hydrophobic interactions are weak , but the simultaneous presence of many interactions of this type in a protein can lead to a stabilizing effect if they do cooperate .

Free energy

Free energy represents the energy stored in a certain system and retrievable in the form of work.
Free energy V at constant temperature T is a combination of Entropy S (related to the number of different configurations of the system) and Potential Energy U

$$V = U - TS$$



Minima (a) and (b) in this function represents (meta)-stable state of the protein, x is a reaction coordinate. Under physiological conditions the folded state is stable and the unfolded one unstable .

Levinthal Paradox

C. Levinthal, J. Chim. Phys. 65, 44 (1968)

Theoretical estimation of the folding times assuming all the configurations as equiprobable .

1. The number of possible different configurations for a protein made up of 100 amino acids, each of which can adopt 2 stable configurations , is $\sim 2^{100} = 10^{30}$
2. Let us suppose that the protein can test 10^{12} structures per second (by considering as characteristic time the vibrational period)
3. The protein would need 10^{18} sec $\approx 3 \times 10^{10}$ years to visit all the possible configurations.
4. Instead the protein folds in much shorter time periods.

WHY ?

Levinthal Paradox

The blindfolded golf player

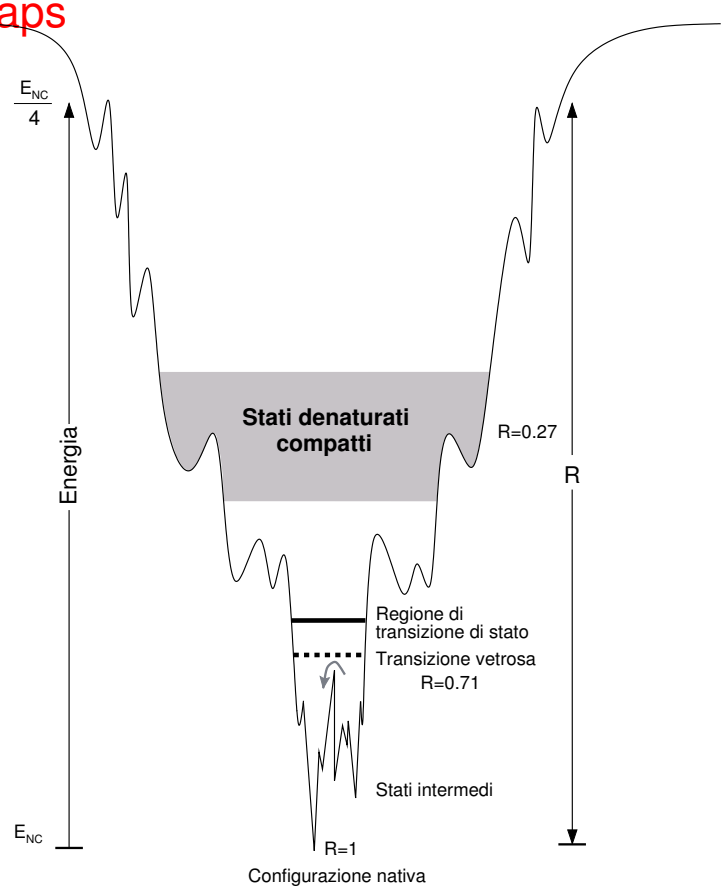
Hypothesis of the Energetic Funnel

J.N. Onuchic et al. PNAS 92, 3626 (1995)

- The Native State is a low-lying free energy minimum
- The minimum should be kinetically accessible,all the paths converge towards the native state.
- Efficient folding require the absence of kinetics traps

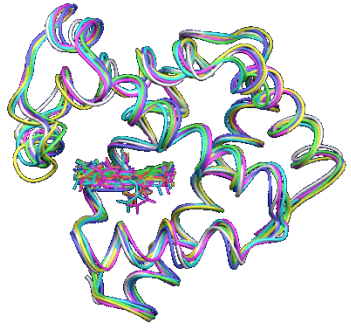
Kinetics and thermodynamics reconciliated

- Collapse Θ, folding T_f and glassy T_g temperatures
- The protein folds rapidly for $T \simeq T_f$
- A good folder has a $T_f/T_g >> 1$, the native state is kinetically accessible around T_f



Folding and Unfolding

- In a liquid solvent (water), under physiological condition, (T,ph,...) the protein folds assuming always the same tridimensional shape (tertiary structure) that is completely determined by the sequence of the amino acids (Folding Transition).
- By increasing the temperature or varying the ph or adding chemical agents, the protein can loose its shape together with its biological functionality. This denaturation process is reversible. (Unfolding Transition).



Problem of the direct and inverse folding

- Given the sequence $\longrightarrow$ to obtain the structure
- Given the structure $\longrightarrow$ to obtain the sequence

$50,000$ sequences and $11,000$ structures have been identified

Protein structure recognition

- Threading uses a database of known structures to match new sequences with protein folds, this is accomplished via a scoring function that assesses the fit of a sequence to agiven fold. The fold with the best score is assumed to be the native conformation of the studied sequence. Often more scoring functions are used at the same time.
- Ab initio method are based on Molecular Dynamics (MD) simulations of proteins, Monte Carlo (MC) simulations that do not use forces among atoms but instead compare energies and Genetic Algorithms, which try to improve convergence and sampling efficiency of MC schemes. These methods try to recover the Native Configuration starting from the sequence by sampling the space of possible conformations of the polypeptide and determining the most probable and/or stable configuration at a certain temperature.
- A combination of MD and MC methods with threading has been also recently employed.