Biomolecules

Adolfo Bastida

Universidad de Murcia (Spain)

Computational Biochemistry

Master in Theoretical Chemistry and Computational Modelling

TCCM

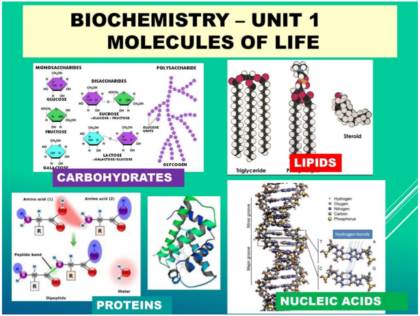
Index

Definition and classification

Biomolecule or biological molecule ⇒ molecules present in organisms that are essential to biological processes

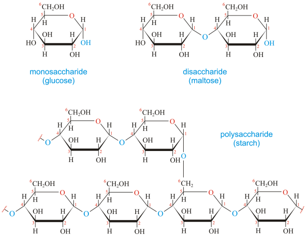
- Inorganic (H_2O, CO_2, NH_3, ...)
- Organic (based on C)

  - Carbohydrates
  - Lipids
  - Nucleic acids
  - Proteins
  - others



Definition and classification

- Carbohydrates



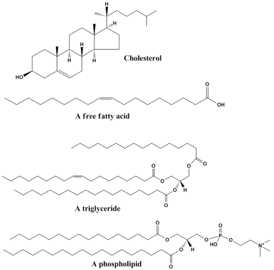
- Energy storage

  starch, glycogen,...

- Structure

  cellulose,...

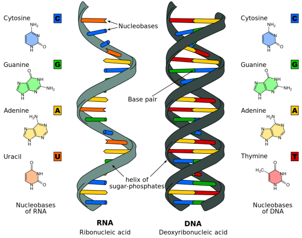
- Lipids



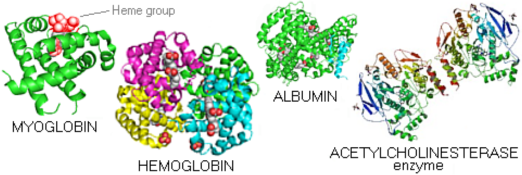
- Energy storage
- Structure
- Signaling

Definition and classification

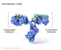
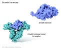
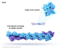
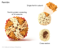
- Nucleic acids



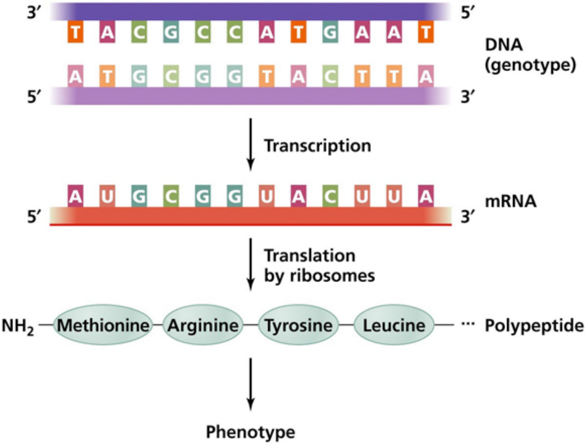
- Proteins



Definition and classification

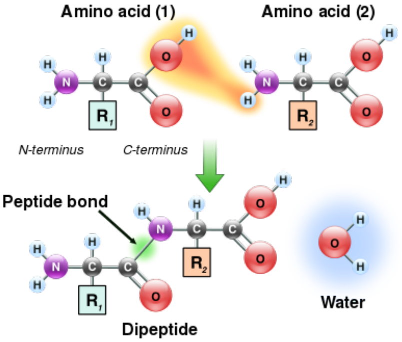
Function	Description	Example
Antibody	Antibodies bind to specific foreign particles, such as viruses and bacteria, to help protect the body.	Immunoglobulin G (IgG) 
Enzyme	Enzymes carry out almost all of the thousands of chemical reactions that take place in cells.	Phenylalanine hydroxylase 
Messenger	Messenger proteins, such as some types of hormones, transmit signals to coordinate biological processes between different cells, tissues, and organs.	Growth hormone 
Structure	These proteins provide structure and support for cells. On a larger scale, they also allow the body to move.	Actin 
Transport/storage	These proteins bind and carry atoms and small molecules within cells and throughout the body.	Ferritin 

The life code



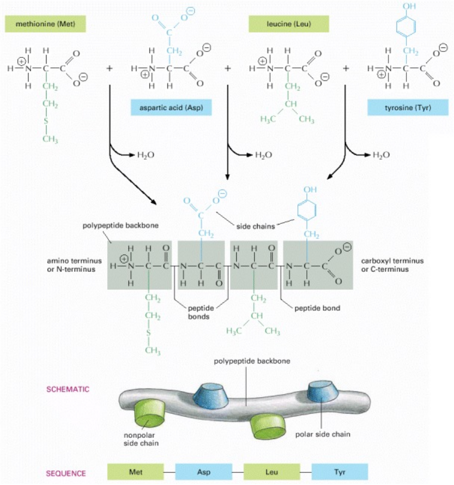
Molecular structure

- Carbohydrates, lipids and Nucleic acids ⇒ well-defined structures
- Proteins ⇒ Flexibility



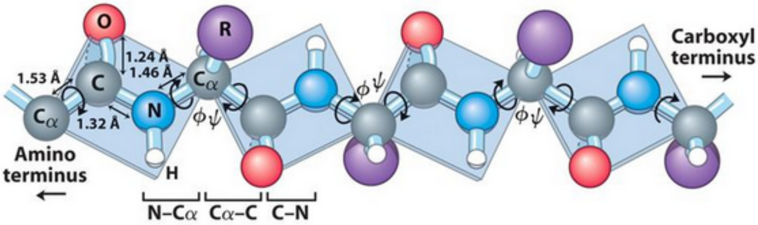
- aa ⇒ NH_2-C_αHR-COOH
- 22 aa ⇒ R=H, Glycine, R=CH_3, Alanine,....
- Why only 22? ⇒ Limit = 4^3=64
- Peptide bond ⇒ Long chains
- Proteins ⇒ up to ∼ 30.000 aa
- Residue ⇒ aa inside a peptide chain

Protein backbone



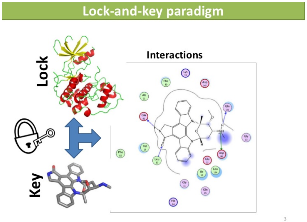
Flexibility

- Every aa $\Rightarrow$ 2 dihedral angles (ϕ_i, ψ_i)
- Low torsional energy barriers $(\sim k_\mathrm{B} T)$
- Flexibility

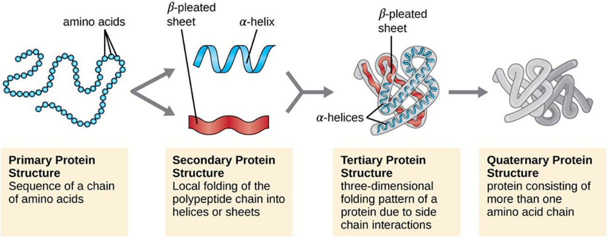


Native struture

- Conformation ⇔ Functionality
  - Enzymes ⇒ Binding to a specific substrate before they can catalyze a chemical reaction ⇒ Selectivity

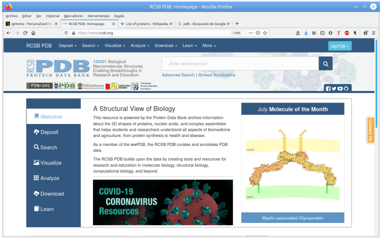


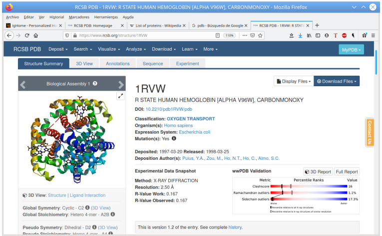
- Native structure

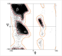


Structural databases

- Protein Data Bank (www.rcsb.org)
- *This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers*
- X-ray structures
- Letter code ⇒ Hemoglobin (1RVW)
- Papers, experimental data, sequence, pdb files, ...

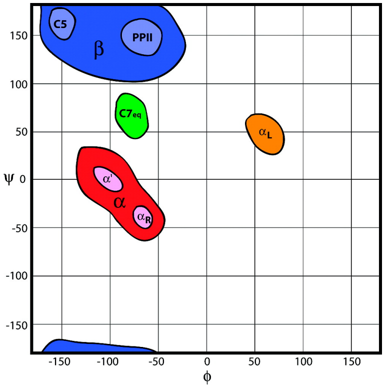






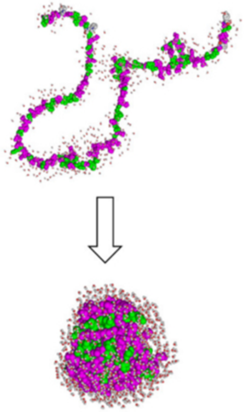
Ramanchandran plot

- Native structure ⇒ minimun of the free energy
- Water solution at room temperature ⇒ thermal fluctuations
- Each residue may explore different conformations
- There are preferential conformations corresponding to local minima of the free energy
- Ramachandran plots for every residue



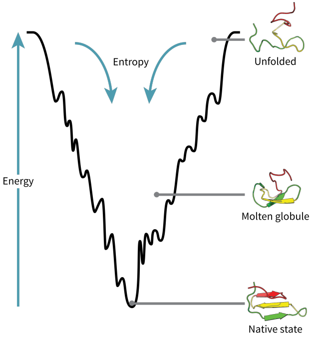
Levinthal's Paradox

- Proteins fold spontaneously and on short timescales (ns-μs)
- 100 aa protein $\Rightarrow$ 200 dihedrals $\Rightarrow$ 3 conformations for angle $\Rightarrow$ 3^{200} conformations $\Rightarrow$ 1 native structure
- Random sampling conformational space $\Rightarrow$ time longer than the age of the universe
- Folding pathways!
- Target $\Rightarrow$ Sequence-Native structure



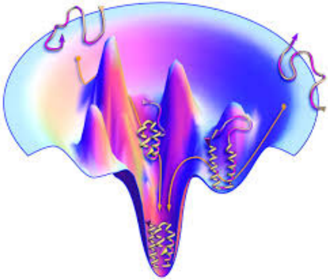
Folding funnel

- During the folding process, the conformational entropy of the protein decreases since the formation of native contacts reduces the accessible conformational space.
- This entropic reduction must be compensated by the remaining contributions to the free energy, which are the energy resulting from the intra- and intermolecular interactions and the entropy of the solvent, for the total free energy to decrease during the folding process.
- Folding funnel



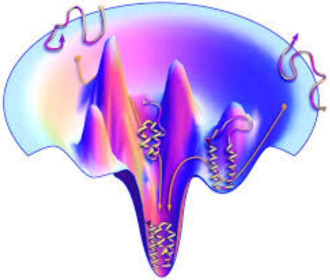
Folding funnel

- Folding funnel $\Rightarrow$ Energy $\Leftrightarrow$ Entropy
- $NVT \rightarrow \Delta F_{\mathrm{f}} = \Delta E_{\mathrm{f}} - T\Delta S_{\mathrm{f}}$
- Different energetic and entropic contributions find a delicate balance that minimizes the total free energy along the folding process. Indeed, as folding progresses, both the potential energy and the entropy of the protein decreases, although resulting in opposite effects on the free energy, whereas the entropy of the solvent plays a much more uncertain role.
- Biological design $\Rightarrow$ avoid too stable proteins

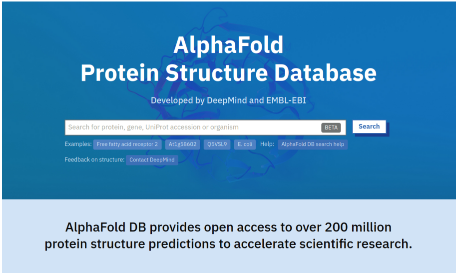


Folding funnel

- The free energy landscapes are intrinsically rugged, and typically include different local minima which act as traps during folding, with the free energy barriers being of the same order of magnitude as the thermal energy, so the molecule may overpass them during the folding process ⇒ Roughness



Alphafold



- In CASP14, AlphaFold was the top-ranked protein structure prediction method by a large margin, producing predictions with high accuracy.
- AlphaFold DB provides open access to over 200 million protein structure predictions.

Demis Hassabis

"for protein structure prediction"

Demis Hassabis. Ill. Niklas Elmehed © Nobel Prize Outreach

John Jumper

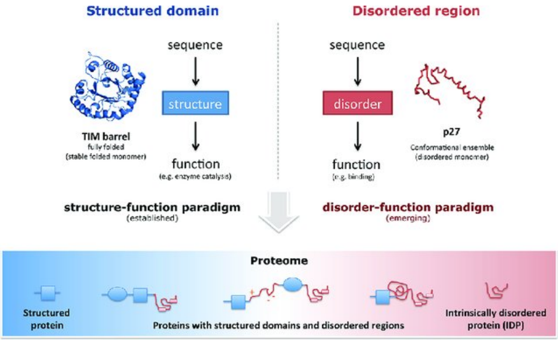
"for protein structure prediction"

John Jumper. Ill. Niklas Elmehed © Nobel Prize Outreach

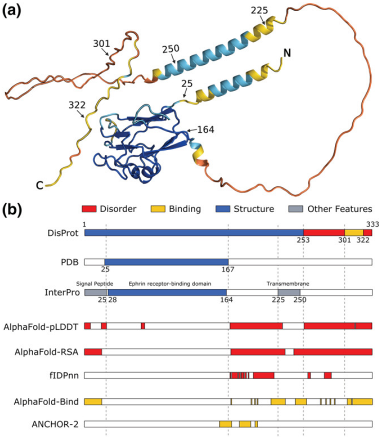
- Nobel Prize in Chemistry 2024

Intrinsically disordered proteins (IDPs)

- >40 % of eukaryotic proteins are disordered.
- Database of intrinsically disordered proteins (https://disprot.org/)
- Critical Assessment of Protein Intrinsic Disorder (CAID) experiment



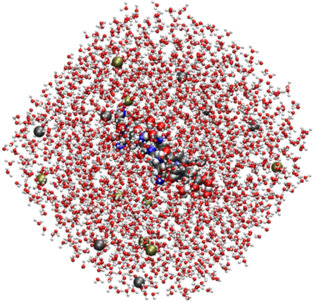
Intrinsically disordered proteins (IDPs)



The human Ephrin-B2 protein (UniProt accession: P52799) is shown as a representative example to illustrate the overlap between AlphaFold predictions and various sequence features. Panel a, the structure of the protein predicted by AlphaFold and colored by pLDDT score (<50 orange, <70 yellow, <90 light blue, >90 dark blue). (Protein Sci. 2022 Nov; 31(11): e4466.)

Complexity

- Biophysical problem ⇒ no formation/break of chemical bonds
- Complexity
  - high flexibility of the peptide chains
  - weak noncovalent interactions
  - water solvent
- Economical interest ⇒ Pharmacy



- Biology ⇔ Biochemistry ⇔ Physics ⇔ Maths/Computing (⇔ Economy?)

in vitro vs *in silico*

- *in vitro* studies ⇒ difficult, slow and expensive.

- *in silico*
  - Computer simulations at atomistic level
  - High computational cost ⇒ but cheaper than lab experiments
  - Agreement with experimental data? ⇒ filter candidates

Understanding computer simulations?

- Lack of first principles governing the folding process

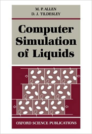
To milliseconds and beyond: challenges in the simulation of protein folding
Thomas J Lane[1], Diwakar Shukla[1,2], Kyle A Beauchamp[3] and Vijay S Pande[1,3,4]

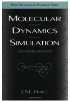
Current Opinion in Structural Biology 2012, **23**:1–8

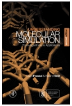
While challenging, generating enough sampling in an accurate forcefield does not constitute the end of the road for a folding simulation. Another major challenge is gaining scientific insight from the simulation — turning data into knowledge.

- Paradox: Being able to simulate a process does not mean that we understand it.

Bibliography

- 

Computer simulation of liquids by M.P. Allen and D.J. Tildesley. Oxford University Press. 1987.
- 

Molecular dynamics simulations. Elementary methods. by J.M. Haile. John Wiley and sons. 1997.
- 

Understanding molecular simulations. by D. Frenkel and B. Smit. Academic press. 2002.