800 ANNI 1222·2022

Università
degli Studi
di Padova

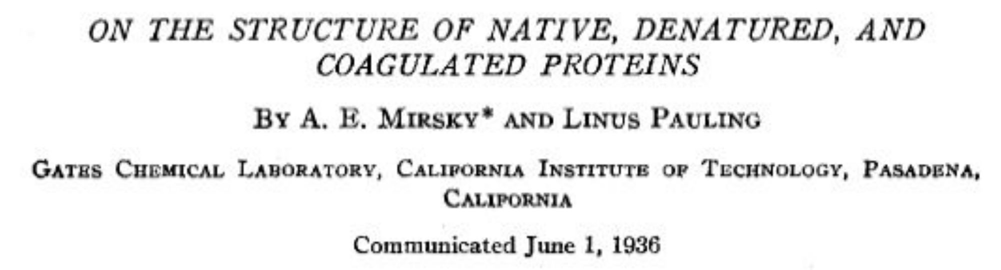
DIPARTIMENTO
MATEMATICA

DATA SCIENCE
UNIVERSITY OF PADOVA

PROTEIN FOLDING PROBLEM

Master of Science in Data Science

Damiano Piovesan

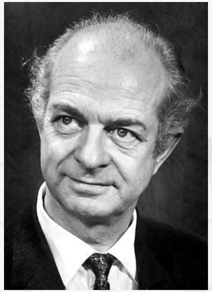


ON THE STRUCTURE OF NATIVE, DENATURED, AND COAGULATED PROTEINS

By A. E. Mirsky* and Linus Pauling

Gates Chemical Laboratory, California Institute of Technology, Pasadena, California

Communicated June 1, 1936

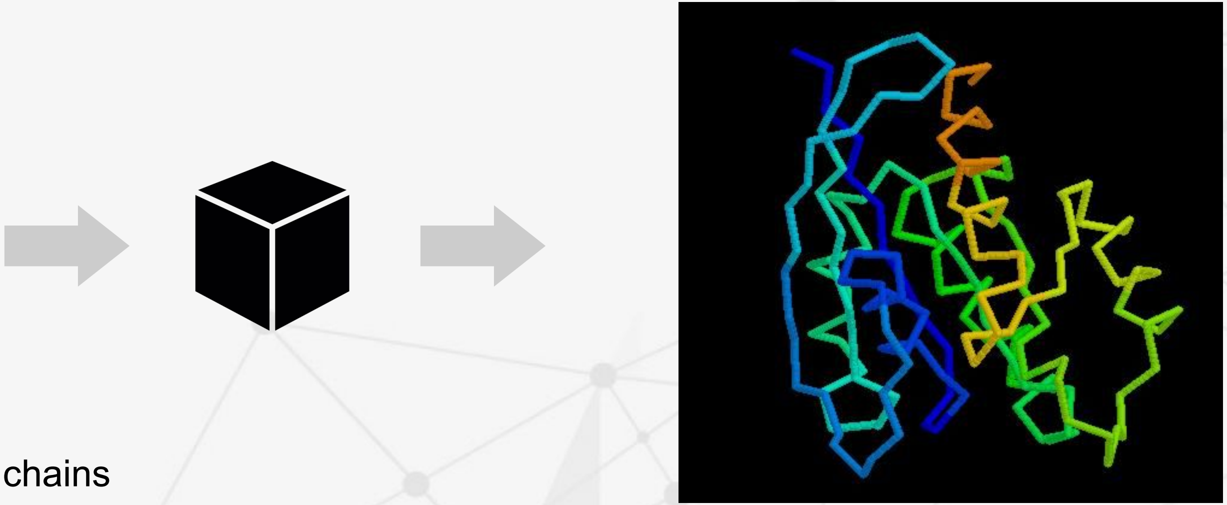


“Our conception of a ***native protein*** *molecule (showing specific properties) is the following. The molecule consists of one polypeptide chain which continues without interruption throughout the molecule (or, in certain cases, of two or more such chains), this chain is* ***folded into a uniquely defined configuration****”*

Linus Pauling, 1904 - 1994

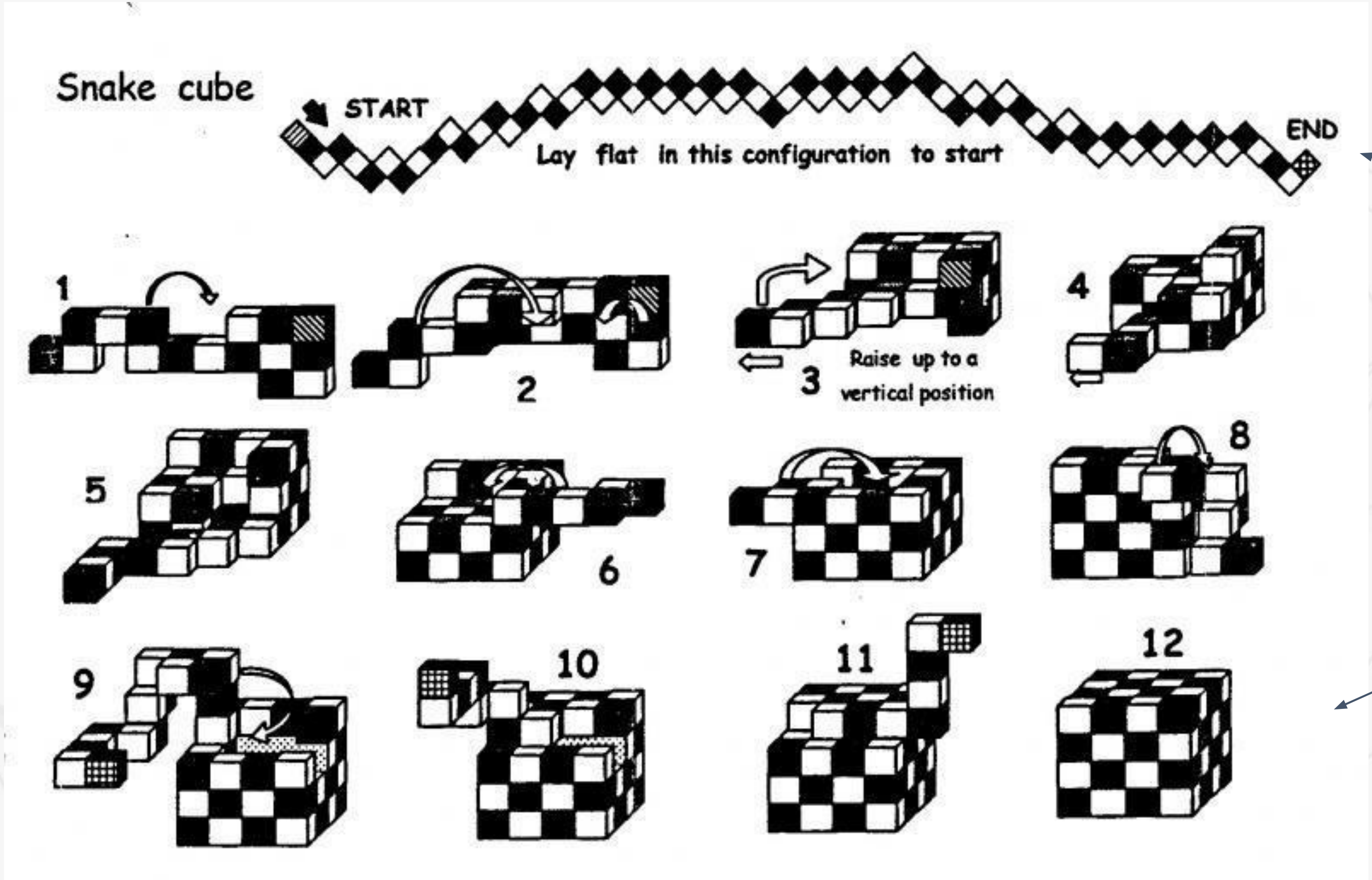
Protein folding problem

MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKE
GVLYVGSKTKEGVVHGVATVAEKTKEQVTNVGGAV
VTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEE
GAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPEA



- Single polypeptide chains
- 20 L-amino acids, no modifications
- Water solvent, no reagents

Conformations / Configurations



Denatured conformation

Native conformation (folded)

Levinthal’s paradox

Assumptions (wrong)

- A protein sample all possible conformations (random walk)
- The conformation of a residue is independent of the rest

Statement

- The protein will never fold to its native structure

Example

- 6 possible conformations (type of secondary structure) x 100 residues
- $6^{100} \simeq 10^{78}$ conformations
- 10^{58} years to fold. 1 picosecond (10^{-12} seconds) for a single molecular vibration

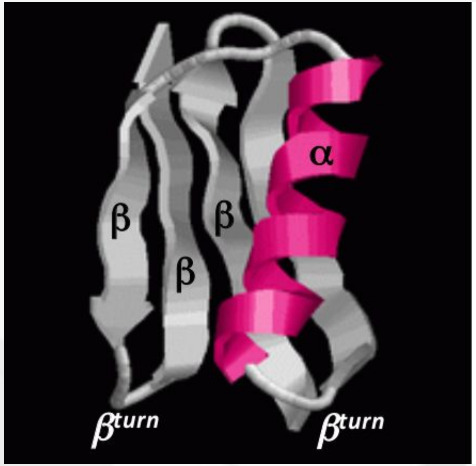
Local conformation

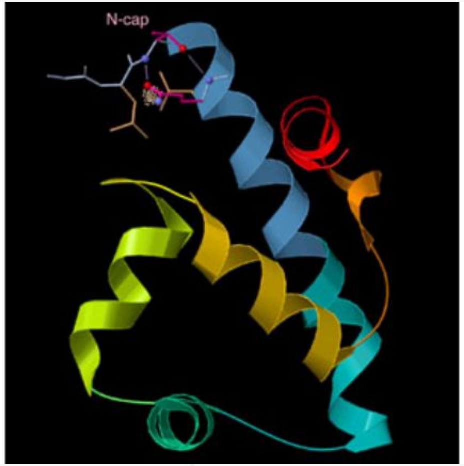
Master of Science in Data Science

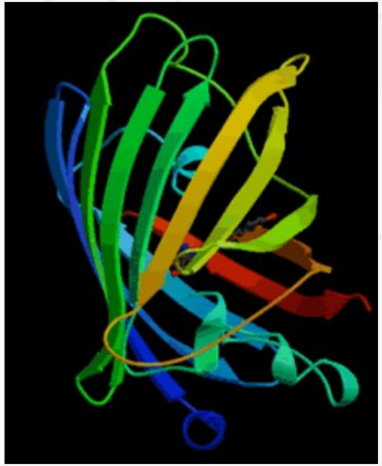
Damiano Piovesan

Secondary structures

- **α-helix** and **β-sheet** are regular structures, stable and frequent in proteins. They minimize steric repulsion and maximize H bonds
- **Random coil**, apparently not regular







Secondary structure

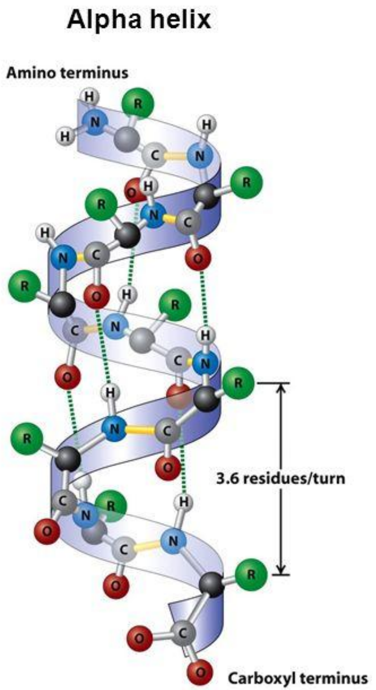


Figure 3-4
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

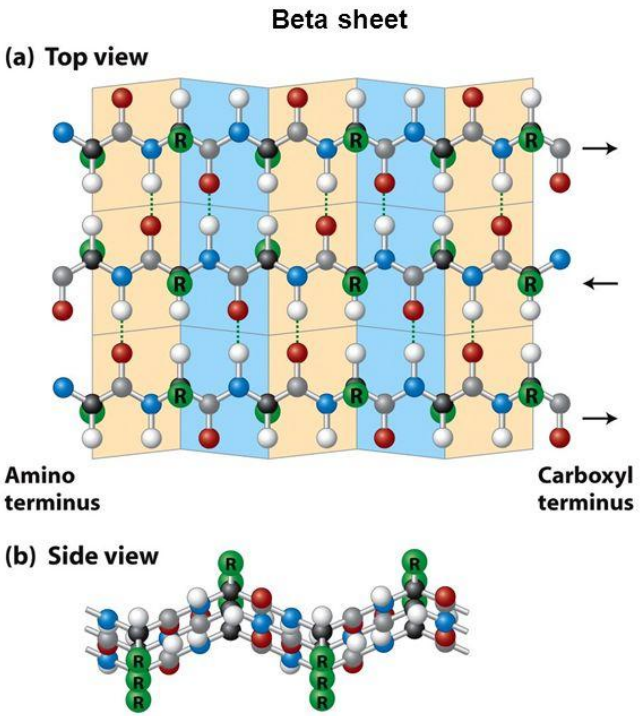
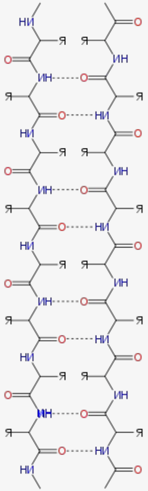
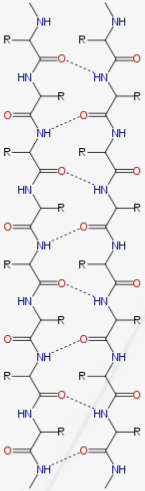


Figure 3-5
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

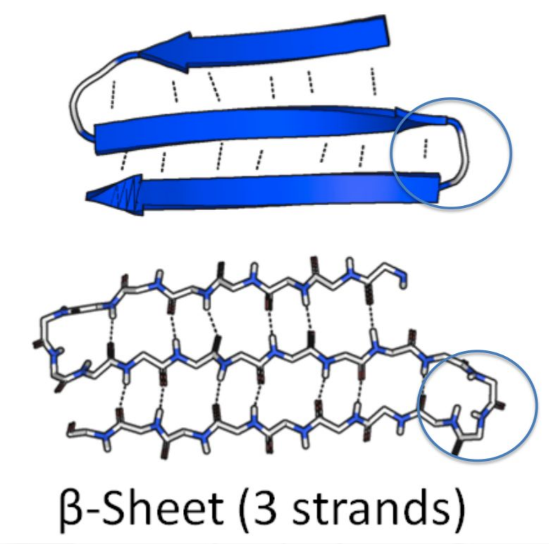
β-sheets



Anti-parallel

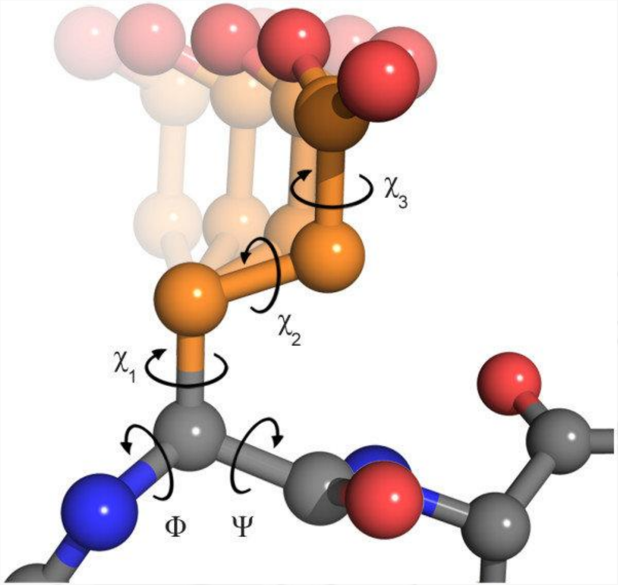


Parallel



β-Sheet (3 strands)

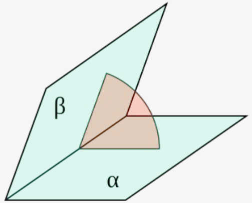
Amino acid rotamers (degree of freedom)

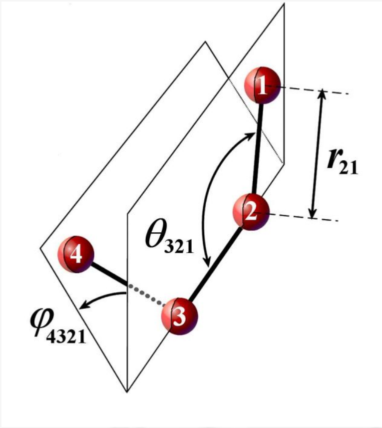


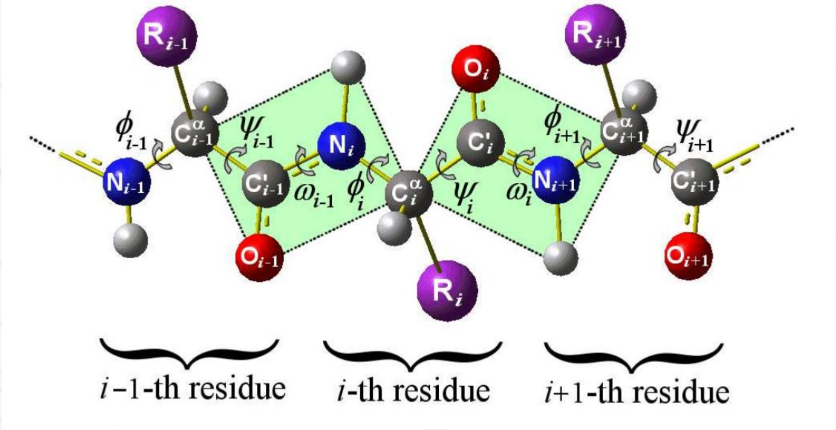
Beyond rotamers: A generative, probabilistic model of side chains in proteins.
Harder at al. 2010, BMC Bioinformatics, 11(1):306

Dihedral angles (backbone)

- The **peptide bond** is **rigid** and **planar** bond because it has a **partial double bond** character
- It is **0.13 Angstrom shorter** than the C-N single bond yet not as short as a double bond







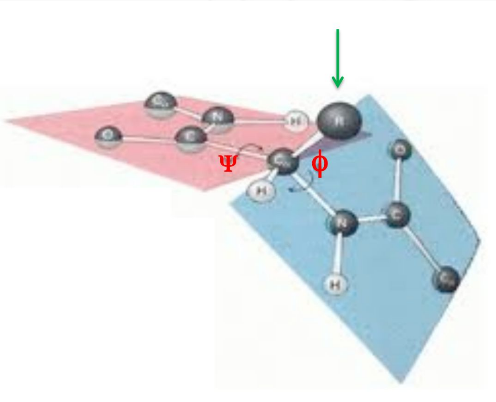
Constraints & parameters

Hard (no freedom)

- Bond lengths
- Bond angles
- Peptide bond
    - Main chain dihedral angle→ ω

Soft (rotations)

- Single bonds (dihedral angles)
    - main chain → Φ, Ψ
    - sidechain → X

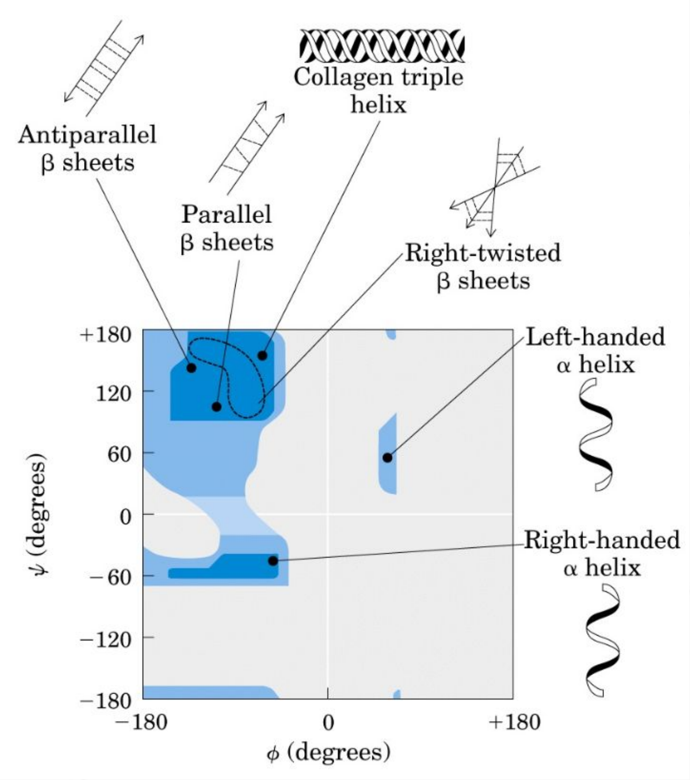


Ramachandran plot

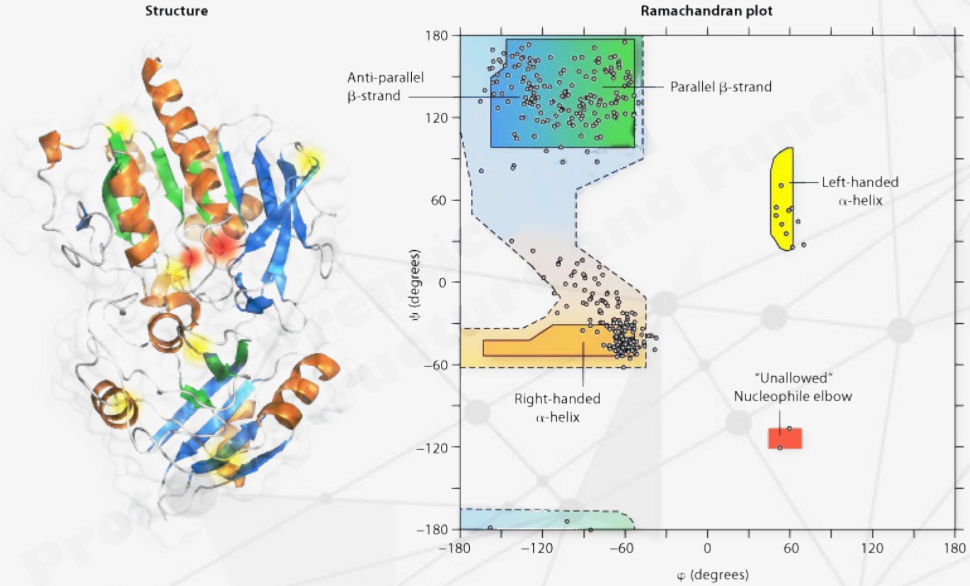
Patterns of **hydrogen bonds**

⇕

Specific **Φ, Ψ** angles



Ramachandran plot



Folding driving forces

Master of Science in Data Science

Damiano Piovesan

Thermochemistry

It studies the **transfers and transformations** (variation of state functions) of energy in chemical processes.

- **Energy (heat, work)** involved in a transformation **(First Law)**
- **Spontaneity** of the transformation
- **Equilibrium** of the system **(Second Law)**

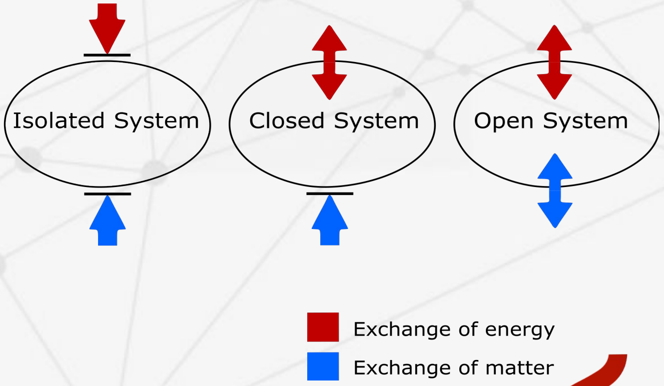
State Functions → define the state of a system and **do not depend on the path taken** to reach the final state from the initial state.

- **Examples**: Pressure (P); Temperature (T); Concentration ©; Internal Energy (U)

First law - Conservation of energy

The **total energy** of the system and its surroundings is **constant**

- **System** → A portion of the Universe (a body or specific region of interest)
- **Surroundings** → The external environment (or another body)
- **Universe** → System + Surroundings



Enthalpy (H) - Internal energy

The energy of a system at constant pressure is called enthalpy (H)

- **Ideal gas** → Kinetic Energy
- **Solution** → Kinetic Energy + Bond Energy (Intra- / Inter-molecular)

Biological Reactions take place in solution where volume and pressure (atmospheric) are constant

$$\Delta H = H_{final} - H_{initial} = q_p$$

Unit of measurement → Kj mol^{-1}

The enthalpy of a substance cannot be measured in an absolute sense; only its change (ΔH) during a reaction can be measured

Reaction enthalpy

$$Q_{reaction} = -\Delta H_{reaction}$$

$\Delta H_r > 0$ **Endothermic Reaction** → The products have more energy than the reactants
$q < 0$ Heat is absorbed from the surroundings

$\Delta H_r < 0$ **Exothermic Reaction** → The products have less energy than the reactants
$q > 0$ Heat is released to the surroundings (the system has produced heat)

Hess's Law → The energy variation of an overall process is the sum of the enthalpy changes of the individual steps.

Enthalpy of Formation → The heat involved in the formation of one mole of a compound from its constituent elements
Enthalpy of Combustion → The heat developed when one mole of a "fuel" (combustible substance) burns

Entropy

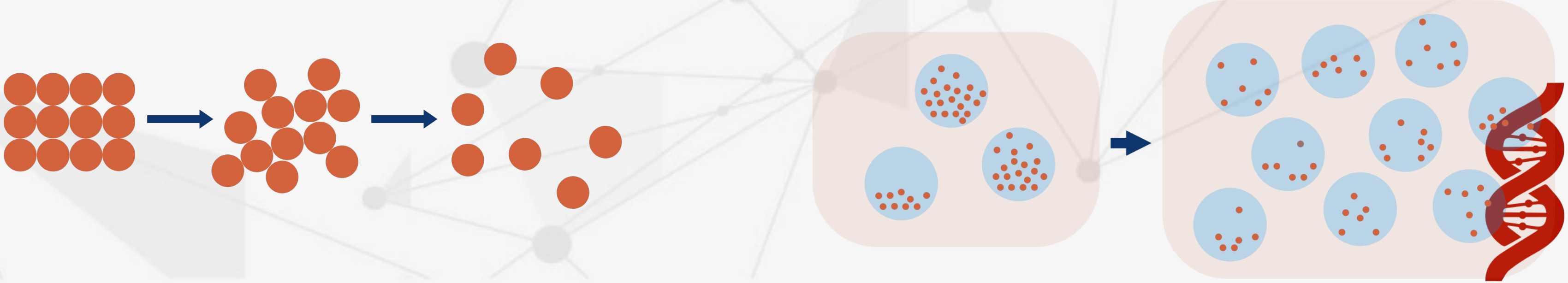
The energy of a system that cannot be exchanged as heat (**"dispersed" energy**)

The higher the **number of microstates** in a system, the higher the entropy

$$\mathbf{S = k \ln W}$$

k → Boltzmann constant ($R / N_{avogadro}$)
W → Number of possible microstates

Unit of measurement → $\mathbf{J\ K^{-1}\ mol^{-1}}$. The more particles a system has, the more microstates it can have. This is why "moles" are included in its unit of measurement



II law of thermodynamics

Natural transformations are **irreversible** and **spontaneous** (the entropy of the universe is constantly increasing)

$\mathbf{\Delta S_{total}} = \Delta S_{system} + \Delta S_{surrounding} \mathbf{> 0}$ **Irreversible and spontaneous** (Natural transformations)

$\mathbf{\Delta S_{total}} = \Delta S_{system} + \Delta S_{surrounding} \mathbf{< 0}$ **Irreversible and non-spontaneous**

$\mathbf{\Delta S_{total}} = \Delta S_{system} + \Delta S_{surrounding} \mathbf{= 0}$ **Reversible** (Equilibrium)

Predicting Entropy Changes (ΔS)

$Cl_2(g) \rightarrow Cl_2(l)$

- **Decreases** because the gas becomes a liquid (liquids have fewer microstates than gases)

$CaCO_3(s) \rightarrow CaO(s) + CO_2(g)$

- **Increases** because gas is produced and the number of molecules increases (Solid → Gas is a huge jump in entropy)

$N_2(g) + 3H_2(g) \rightarrow 2NH_3(g)$

- **Decreases** because 4 molecules of reactants form 2 molecules of products. All are gases (Fewer gas molecules mean fewer ways to arrange them in space)

$AgNO_3(aq) + NaCl(aq) \rightarrow NaNO_3(aq) + AgCl(s)$

- **Decreases** because a solid (precipitate) is formed from aqueous reactants (Dissolved ions move freely; ions locked in a crystal lattice do not)

$H_2(g) + Cl_2(g) \rightarrow 2HCl(g)$

- **Unknown (not zero).** There is the same number of molecules on both sides, and all are gases
- While the volume/pressure might stay the same, the identity of the molecules changes. HCl might have slightly different rotational or vibrational energy levels than H_2 or Cl_2, so ΔS will be close to zero but not exactly zero

Gibbs Free Energy (ΔG)

If pressure and temperature are constant → It is possible to evaluate the spontaneity of a process by considering only the properties of the system.

$$\Delta G = \Delta H - T\Delta S$$

ΔG < 0 → Spontaneous reaction (Exergonic)
ΔG > 0 → Non-spontaneous reaction (Endergonic); the reverse reaction is spontaneous
ΔG = 0 → State of equilibrium → The entropy of the universe does not increase

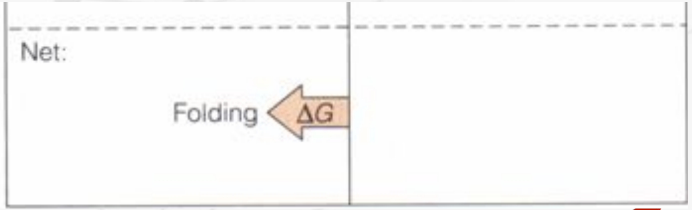
ΔH	ΔS	Sign ΔG	Spontaneous
Exothermic (-)	Increase (+)	-	Yes
Exothermic (-)	Decrease (-)	- or +	Yes, at low temperature
Endothermic (+)	Increase (+)	- or +	Yes, at high temperature
Endothermic (+)	Decrease (-)	+	No

Folding energy

$$\Delta G_{fold} = G_{native} - G_{unfold}$$

Proteins are **marginally stable**

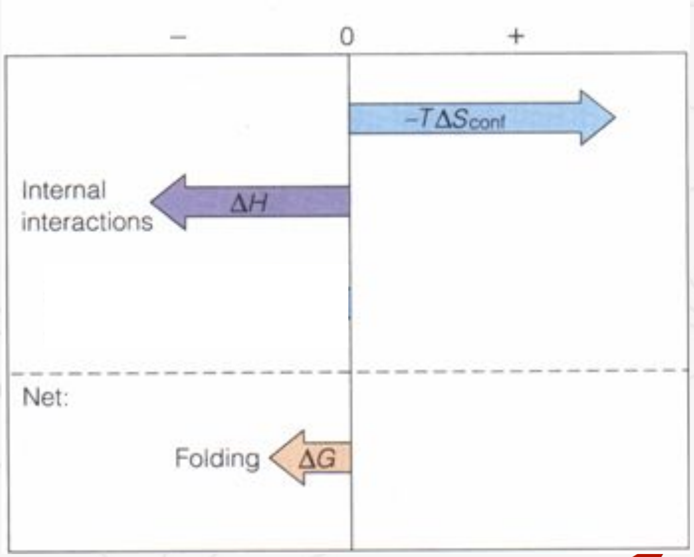
ΔG ≈ -5 / -15 kcal/mol



Folding energy

$$\Delta G_{fold} = \Delta H_{fold} - T\Delta S_{fold}$$

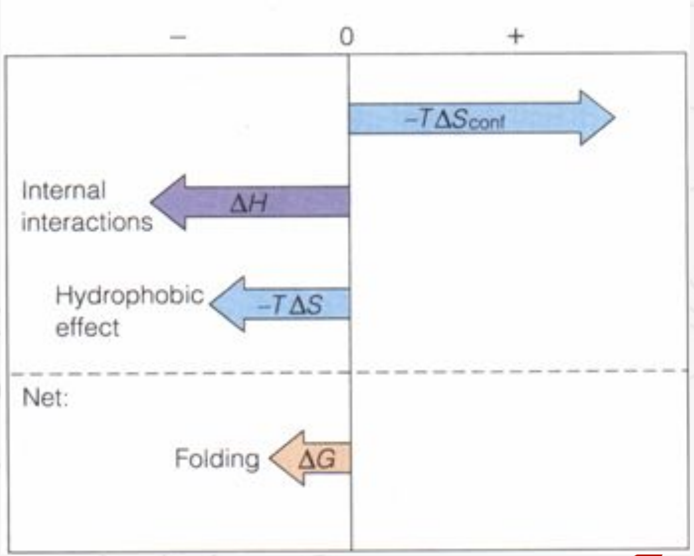
- $\boldsymbol{\Delta H_{fold}}$ → **enthalpy gain**, the contribution of novel interactions formed in the folded configuration (endothermic reaction)
- $\boldsymbol{-T\Delta S_{conf}}$ → **entropy loss**, the cost of reducing the degree of freedom generated by adopting a fixed conformation (negative ΔS_{system})



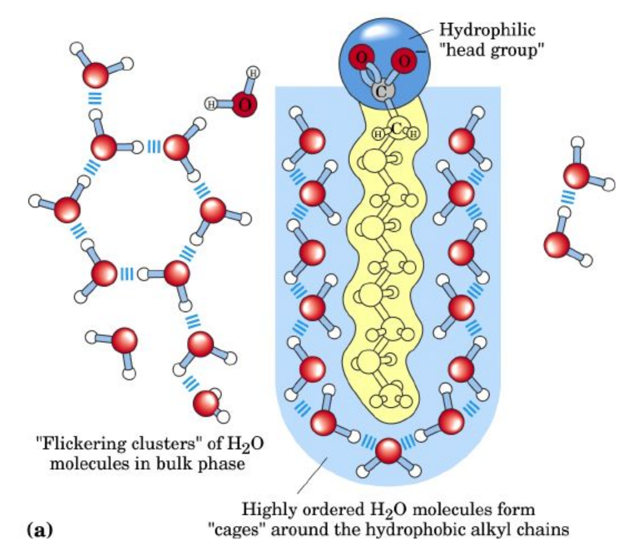
Folding energy

$$\Delta G_{fold} = \Delta H_{fold} - T\Delta S_{fold}$$

- $\boldsymbol{-T\Delta S_{hydrophobic}}$ → **entropy gain** (hydrophobic effect), solvent molecules are less ordered when hydrophobic residues are excluded from the solvent (buried in the protein core) (positive $\Delta S_{surrounding}$)



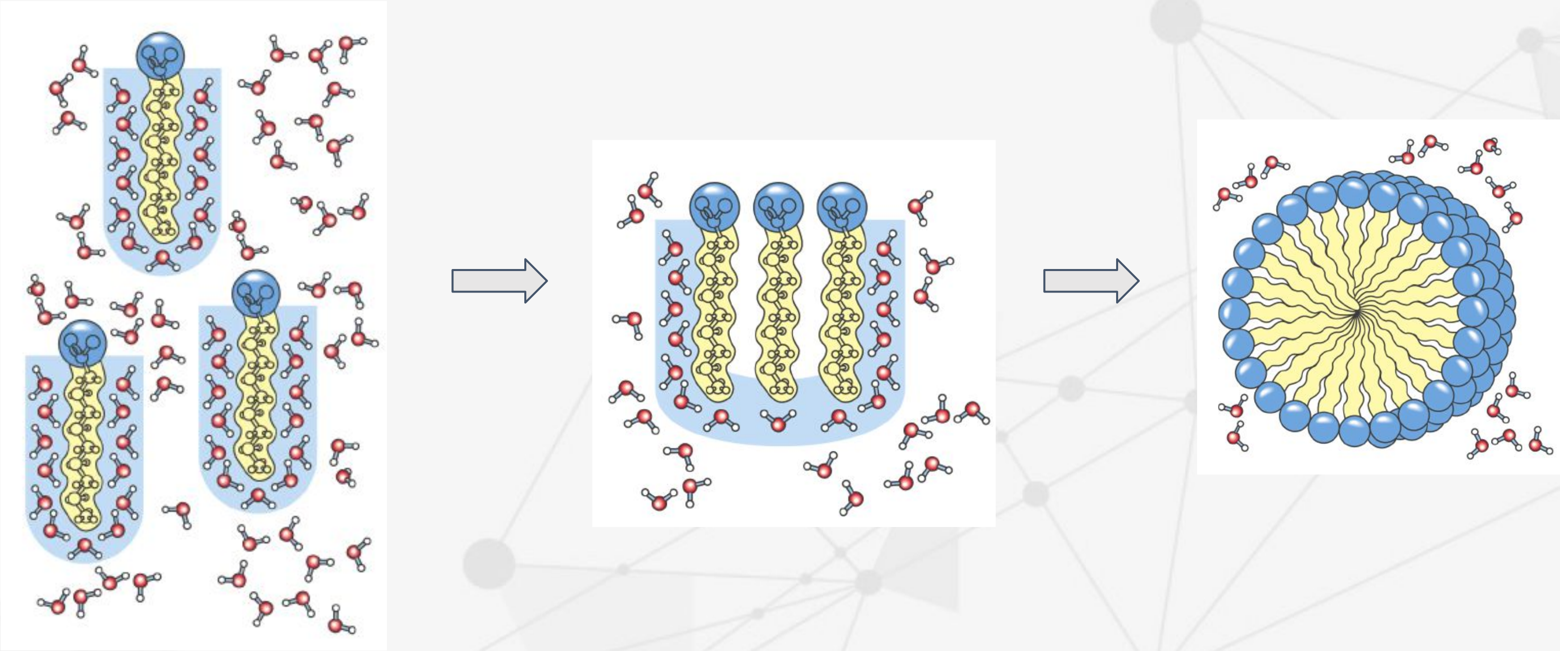
Hydrophobic effect



Water molecules form a cage-like structure around the non-polar molecule

- Positive *ΔH* → the cage has to be broken to transfer the nonpolar molecule
- Positive *ΔS* → water molecules are less ordered when the cage is broken

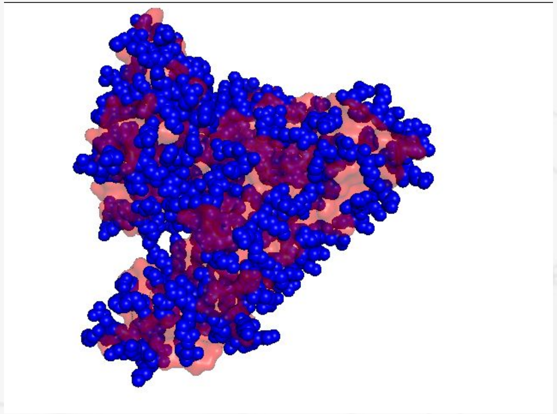
Burial of hydrophobic tails

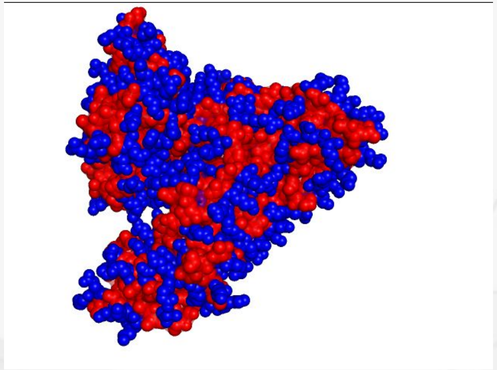


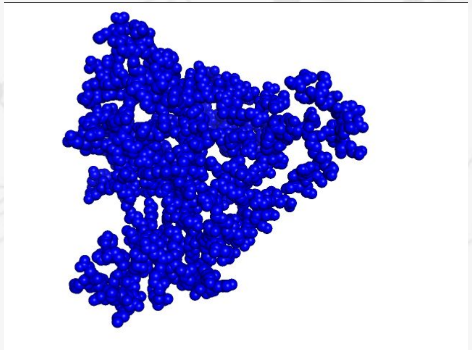
Hydrophobic core

Hydrophobic residues

(cys, ala, gly, val, ile, leu, phe, met, thr, ser, trp, tyr, pro)







PDB 1AO6
HUMAN SERUM ALBUMIN

DIPARTIMENTO
MATEMATICA

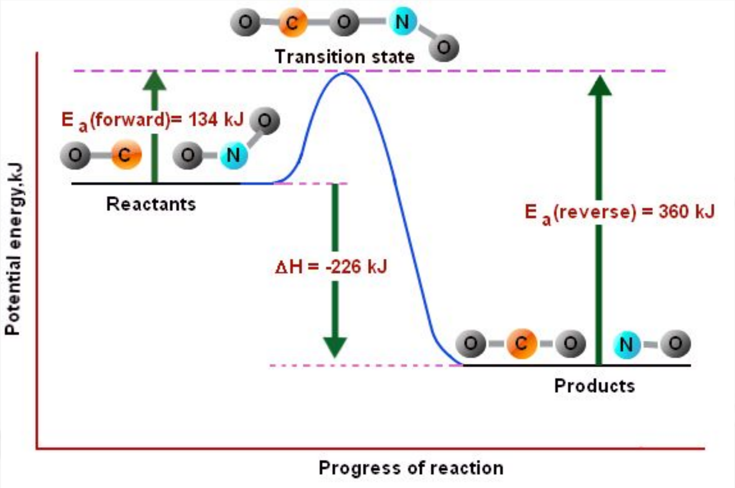
Folding pathway

Master of Science in Data Science

Damiano Piovesan

Thermodynamics vs. Kinetics of Reactions

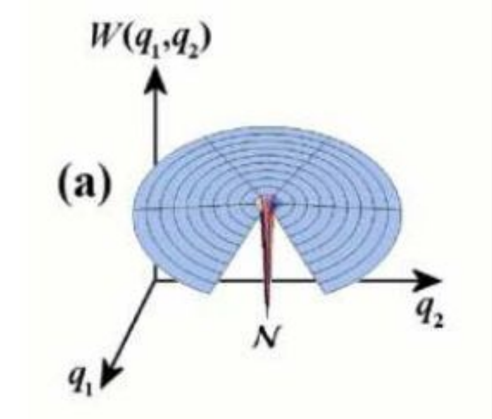
Spontaneity ($\Delta G < 0$) → A negative Gibbs Free Energy indicates a reaction is energetically favorable, but does not dictate its speed



(reaction coordinates)

- **Energy Barrier** → Spontaneous reactions may occur slowly if the **Activation Energy (E_a)** is too high for molecules to overcome at ambient temperatures
- **Transition State** → Reached only when reactants collide with sufficient kinetic energy to break existing bonds
- **Reaction Velocity** → Determined by the relationship between the system's thermal energy (Temperature) and the height of the energy barrier (E_a)

Flat golf course - Impossible



q_1 and q_2 →configuration coordinates

$W(q_1,q_2)$ →potential energy (Gibbs)

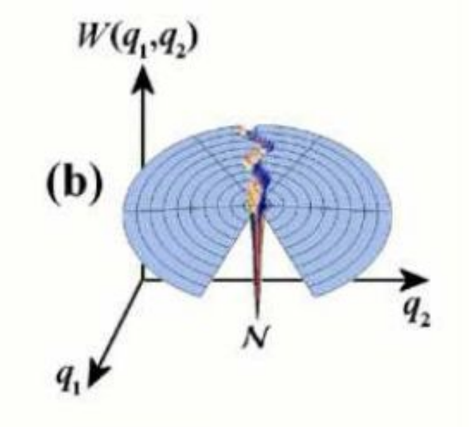
Levinthal's paradox

- A protein sample all possible conformations (random walk)
- The conformation of a residue is independent of the rest
- How it is possible that proteins fold in milliseconds / seconds range?

Example →NTL9 https://www.youtube.com/watch?v=gFcp2Xpd29I

Ant trail - Old view

Hypothesis (1969) →As for simple chemical reactions, there are steep well defined energy paths leading to the native conformation

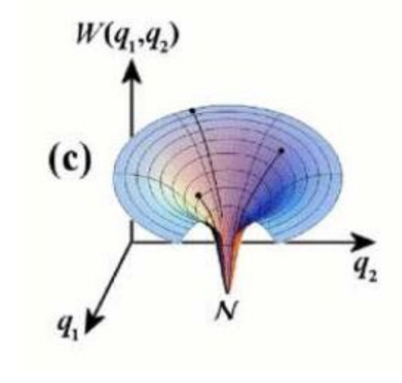


However, in protein folding...

- Driving **forces** are **weaker**
- Short-lived **transient interactions** form randomly and the system describes stochastic trajectories that are never the same
- The native state may be reached in **many ways**, there is not a single minimum energy path dominating over the others

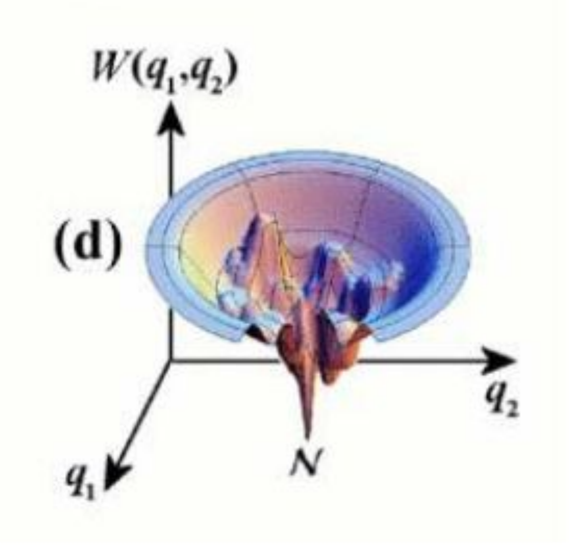
Smooth funnel - New view

Hypothesis (late 80s) →Statistical treatment in which folding is a **heterogeneous reaction** involving broad **ensembles of structures**, where each molecule follows a partially **stochastic trajectory** determined by the intrinsic energetics of the system



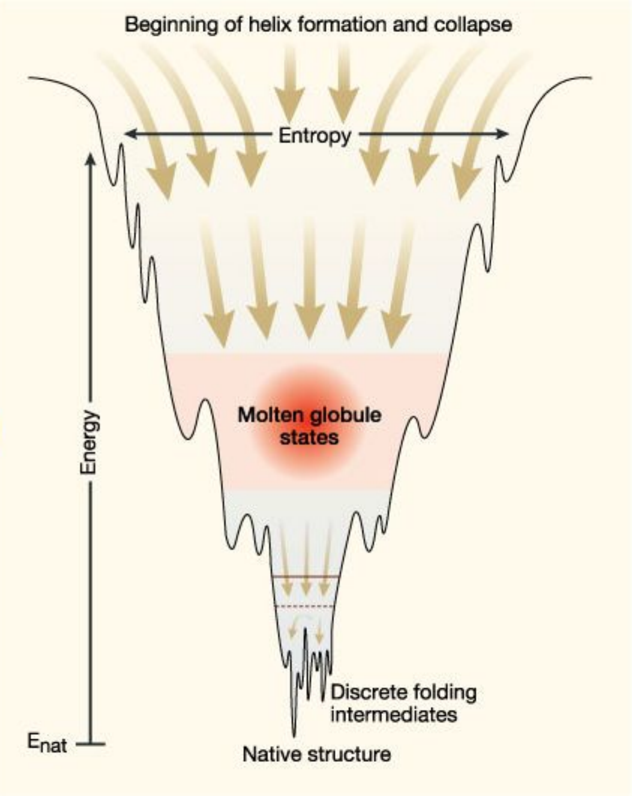
- The **probability** of going towards the native basin is very high (>99%) → the only possible explanation is a **smooth "funneled" energy landscape**
- The "old view" is a particular case of the "new view"

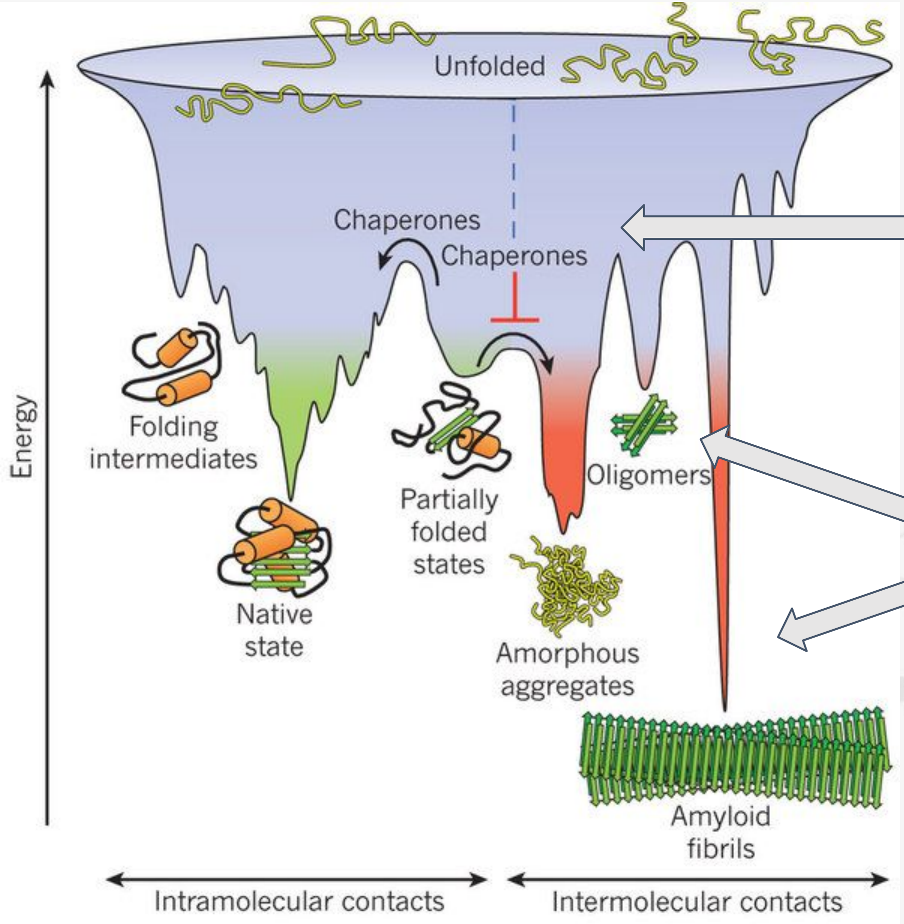
Rugged funnel - Realistic



Principle of minimal frustration

- Proteins are **not random polymers,** their sequences satisfy the principle of **minimal frustration**
- They are **selected** and improved by **natural selection**. Random sequences will never fold. The **scoring function** is the ability to fold into a **native structure** in a biologically **reasonable time**
- **Minimal Frustration** → Natural proteins have evolved to minimize conflicts. Most interactions work together toward the same goal: reaching the bottom of the funnel
- **Residual Frustration** → Proteins aren't perfectly smooth. Some "roughness" or "bumps" in the funnel exist. This "leftover" frustration is often functional—it allows proteins to be flexible, to change shape when binding a drug, or to be "switched" on and off.

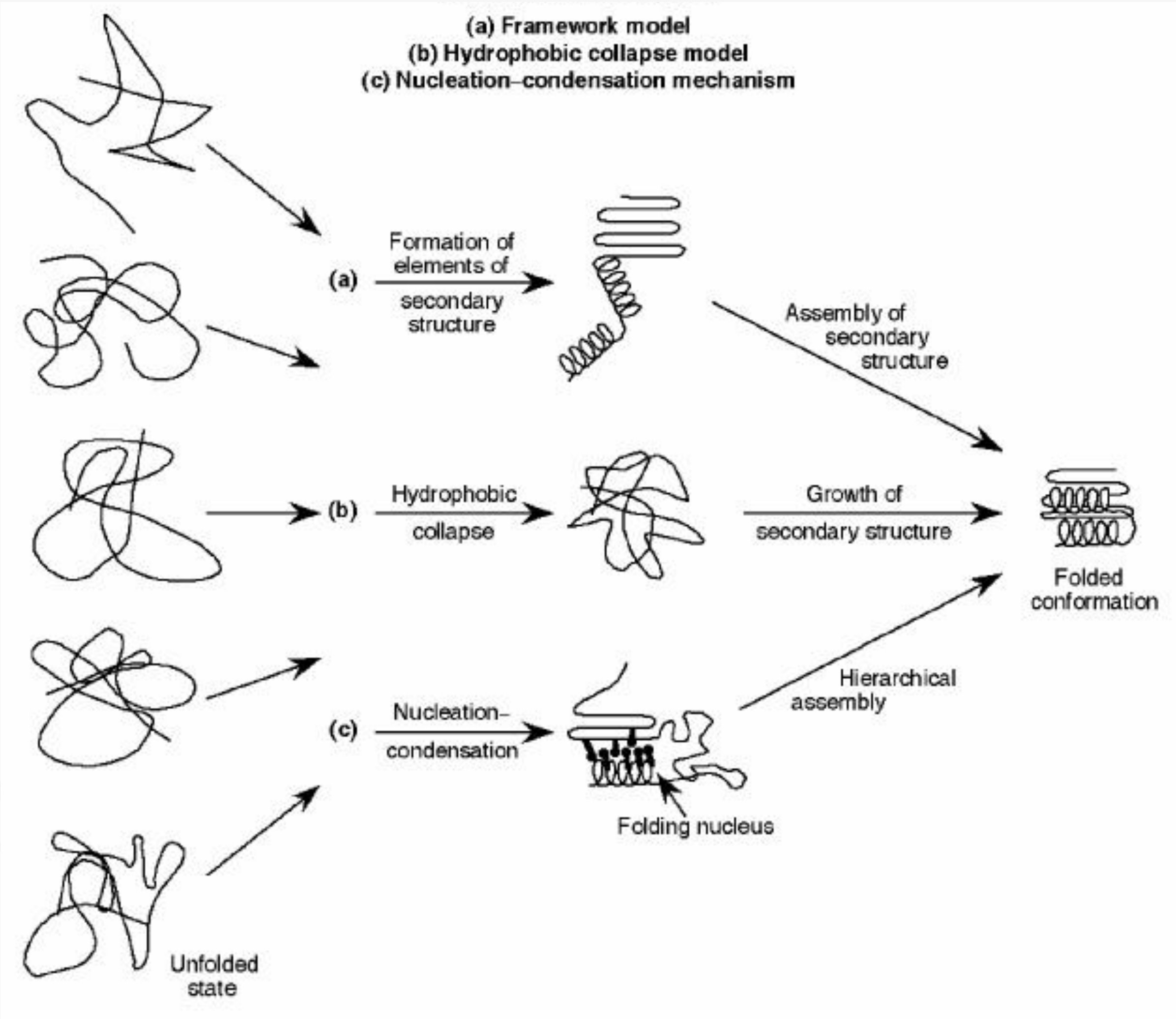




Chaperones are proteins that help other proteins to fold properly and prevent errors

Pathological conditions (e.g. Alzheimer)

Folding models



References & Links: Protein folding

Introduction to protein folding for physicists

Pablo Echenique

2007, arxiv.org

https://arxiv.org/abs/0705.1845

TMP Chem (Trent Parker's YouTube channel)

https://www.youtube.com/user/TMPChem

PlayLists: PChem Math, Chemical thermodynamics, Computational Chemistry