PROTEIN FOLDING PROBLEM

Master of Science in Data Science
Damiano Piovesan
“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

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

Snake cube

START

Lay flat in this configuration to start

END

1

2

3

Raise up to a vertical position

4

5

6

7

8

9

10

11

12

Native conformation
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
- $6^{100} \approx 10^{78}$ conformations
- $10^{58}$ years to fold. 1 picosecond ($10^{-12}$ seconds) for a single molecular vibration
Local conformation
Amino acid rotamers (degree of freedom)
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.
Degrees of freedom

**Hard** (no freedom)
- Bond lengths
- Bond angles
- Peptide bond
  - Main chain dihedral angle $\rightarrow \omega$

**Soft** (rotations)
- Single bonds (dihedral angles)
  - main chain $\rightarrow \Phi, \Psi$
  - sidechain $\rightarrow \chi$
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

Alpha helix
- Amino terminus
- Carboxyl terminus
- 3.6 residues/turn

Beta sheet
- (a) Top view
- (b) Side view
- Amino terminus
- Carboxyl terminus
β-sheets

- Anti-parallel
- Parallel
- β-Sheet (3 strands)
Patterns of hydrogen bonds

Specific $\Phi$, $\Psi$ angles
Ramachandran plot
Folding driving forces

Master of Science in Data Science

Damiano Piovesan
Folding energy

\[ \Delta G_{\text{fold}} = G_{\text{native}} - G_{\text{unfold}} \]

- \( G \), energy of Gibbs
- **Spontaneous processes** have **negative** \( G \)
- Proteins are **marginally stable** ca. -5 / -15 kcal/mol
### Folding energy

\[ \Delta G_{\text{fold}} = \Delta H_{\text{fold}} - T\Delta S_{\text{fold}} \]

- **\( \Delta H_{\text{fold}} \rightarrow \text{enthalpy gain} \)**, the contribution of novel interactions formed in the folded configuration
- **\(-T\Delta S_{\text{conf}} \rightarrow \text{entropy loss} \)**, the cost of reducing the degree of freedom generated by adopting a fixed conformation
- **\(-T\Delta S_{\text{hydrophobic}} \rightarrow \text{entropy gain} \)** (hydrophobic effect), solvent molecules are less ordered when hydrophobic residues are excluded from the solvent (buried in the protein core)
Hydrophobic effect

- Water molecules form a cage-like structure around the non-polar molecule.
- Positive $\Delta H \rightarrow$ the cage has to be broken to transfer the nonpolar molecule.
- Positive $\Delta S \rightarrow$ 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
Folding pathway
Energy landscape

In simple chemical reactions there are steep well defined energy paths

- Spontaneous processes have negative $G$
- The transition state is reached when substrate molecules collide with enough kinetic energy
- $E \rightarrow$ free energy of 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](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 explanation is a smooth “funneled” energy landscape

• The “old view” is a particular case of the “new view”
Rugged funnel - Realistic
Protein “frustration”

A single conformation that optimizes all the interactions at the same time does not exist

- Degrees of freedom
  - Rotamers
- Constraints
  - Chain connectivity
  - Different affinities of the residues for their neighbours and the environment
In every point of the conformational space it is more stabilizing (less energy) to form “native contacts”

- 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 **score function** is the ability to fold into a **native structure** in a biologically **reasonable time**
Protein folding - Recap

• Process driven by **non-covalent interactions** (low energy, many interactions)

• The **energy landscape** of natural sequences is **funneled** → random movements (trajectories) have high probability to make stabilizing contacts

• Random sequences will never fold → natural sequences have been selected to satisfy the **principle of minimal frustration**

• The **native conformation** is at the **global minimum**, but proteins are marginally stable
Folding models
Chaperones are proteins that help other proteins to fold properly and prevent errors.

Pathological conditions (e.g. Alzheimer)
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