Computational biophysics

Giacomo Fantoni
telegram: @GiacomoFantoni
Github: https://github.com/giacThePhantom/mathematical-modelling-in-biology

November 4, 2022

Contents

Chapter 1

Introduction and proteins

1.1 Introduction

Biomolecular modelling has seen a recent increase in its use in the recent years, with a field still destined to expand. Most of this models take a top-down approach, starting from the macroscopic rather than to build simulation from the fundamental and quantistic concepts. Examples of systems studied through biomolecular modelling are:

- Channels.
- Photosynthetic systems.
- Viruses.
- DNA/RNA interactions.
- Inorganic systems.

Through biomolecular modelling it is possible to obtain:

- Molecular rationale for biological processes like proteins' function or its misfolding.
- A quantitative evaluation of molecular driving forces.
- A prediction of properties of macromolecular structures and architectures.
- A comparative assessment of molecular affinities through the binding free energy.

1.2 Proteins

Proteins have different functions within a cell:

- Give structure.
- Provide exchange of materials.
- Code for messages.
- Transport ions.
- Catalytic.
- Movement.
- Storage.
- Act as toxins.

Proteins are a polymer of amino-acids and occupy a space-scale of $10nm$. The amino-acids are in the range of $1nm$. They are built through a polymerization reaction as chain of amino-acids coded through a degenerate code of RNA nucleotides. Three bases of RNA code for an amino-acid.

1.2.1 Amino-acids

Amino-acids are the monomers of a protein. They have a general structure with an amino and a carboxyl terminal group for all of them. They are distinguished by a residue on the α-carbon which gives them different chemical and physical properties.

1.2.2 Structure

There are four level of a protein structure.

- Primary structure: the amino-acid sequence.
- Secondary structure: here α-helices and β-sheet can be distinguished.
- Tertiary structure: the spatial, 3D dynamic configuration of a protein which arise during protein folding.
- Quaternary structure: the interaction of multiple correctly-folded proteins.

Chapter 2

Proteins' geometry

2.1 Introduction

The study of the geometry of proteins involve what can be learned from protein coordinates.

2.2 The peptide bond

A protein is a collection of amino acids linked together by a peptide bond. A carboxylic end and an amino end of two amino acid react together losing a water molecule and forming a peptide bond. Beside the α-carbon there is another one bonded to the oxygen in the carboxylic group and a nitrogen bond in the amino group. The α carbon is linked to the nitrogen in the amino group of another amino acid. The carbon and nitrogen display $sp2$ hybridization, the central atom and the 3 that form a bond with it form a plane, so the peptide bond is planar. A plane of the peptide bond is formed and rotation of the plane is allowed only around one axis.

2.2.1 Trans and cis

Looking at the peptide bond the carbon atom of the carboxylic group C' and the nitrogen N are each bonded to a different α-C and a trans or cis conformation can happen. Trying to visualize the atoms that belong to the molecules these repel through the Van der Waals interactions, that can be computed through the Lennard-Jones potential:

$$U_{Lj}(r) = E_0\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right]$$

Where:

- r_0 is the distance where the energy is minimum.
- r is the distance between two atoms.
- r_{min} is the distance at which the energy becomes high.

With atoms most of the time the distance between them will be close to r_0. When decreasing the distance a lot of energy is needed and strain is introduced in the molecule. Plotting the values

for the energy, r_0 and r_{min} the expected distance for each couple of atoms can be seen. Focusing on the C-C interaction:

- $r_0 = 3.4$Å.
- $r_{min} = 3.0$Å.

When two carbons atoms are below the minimum value the conformation is strained. Looking back at the conformation of the peptide bond it can be seen that the cis conformation creates a distance of 2.8Å between the two αC, so it is not favourable. So the trans conformation is the least energy-hungry and the most present.

2.3 The Ramachandran angles

The planes formed by the peptide bonds can rotate with respect to each other. So the Ramachandran angles ϕ and ψ can be defined between these planes. For each αC:

- ϕ describes the rotation around its bond with the nitrogen.
- ψ describes the rotation around its bond with the carboxylic group.

These are the angles between the subsequent planes. Some of the angles will require more energy.

2.3.1 Difficulty of rotation

It can be seen how a rotation of the ϕ angle could cause the two C' to come at a distance of 2.9Å (where $r_{min} = 3.0$Å. On the other hand a rotation of the ψ angle could cause the two N to come at the same distance, but in this case $r_{min}(N \dots N) = 2.7$Å. In the case of carbon atoms the distance is less than the minimum distance, while in the case of nitrogen it is greater than the minimum allowed value. Looking at this it can be seen how the ψ rotation is easier.

2.3.2 Ramachandran plot

A Ramachandran plot is a map with the ϕ angle on the x axis and the ψ angle on the y axis. Because a rotation along the ϕ angle is highly disfavoured the angle 0 is strongly disfavoured and is represented like a black stripe (disallowed region). If the amino acids where composed only by carbon and nitrogen atom the Ramachandran map would be **??**, where:

- A forbidden region in the middle.
- Some strained region like for $\psi = 0$.

Looking at a real protein the complexity is increased and the other oxygen and nitrogen atoms are included **??** and other regions become disallowed due to steady clashes. It can be seen how the regions are quite complex. Looking at a glycine and alanine complex it can be seen in **??** the space becomes even more complex. In this case the white regions is very small and a strained region can be seen and the black one. Including other residues the allowed region reduces **??**. This is due to the presence of larger residues.

2.3.2.1 Observed Ramachandran plot

Trying to plot for each amino acid its angles an amino acid is represented as a dot. Most of the points fall inside of the allowed regions but there are some outliers. In some conformation the protein forces the amino acid to assume strange conformations. This is done to check if the structure places the amino acids in a proper way.

2.4 Contact map of proteins

Starting from the coordinates a contact can be built. It is a matrix that map all the contact between the amino acids. A primary structure can be represented as a collection of beads which will be in contact in the 3D structure. A square matrix can be built such that each entry in the matrix will determine whether there is a contact or not. This matrix will be symmetric with diagonal elements with value 1 and two parallel diagonals for the neighbouring amino acids. Secondary structures will have specific signatures:

- α-helices: is usually represented by a line parallel to the diagonal. This is because the amino acids i is interacting with $i + 4$.
- β-strands: the situation is complicated. For parallel β sheets can be parallel to the diagonal. For anti-parallel it can be anti-parallel to the diagonal.

2.4.1 Defining a contact

The contact between two amino acids needs to be defined. To do so the distance between α-C or the distance between the tail of the residue and an α-C. There is also the need to make a trade-off between computational speed and cost. Also the dimension of the protein need to be considered when choosing the distance.

2.5 Topology diagram

Having found the secondary structures with a contact map a topology diagram help to understand how those interact with each other. In a topology diagram the start is the N terminus and the end the C terminus. β-strands are represented as arrows. If the strands always change direction they will form an anti-parallel β-sheet. α-helices are represented as small cylinder. Usually color codes represent the nature of the structure. This helps with numbering of the secondary structures.

Restart at 41:42 of the video.

Chapter 3

Force fields

Chapter 4

Classical mechanics

Chapter 5

Foundations of statistical mechanics

Chapter 6

Microcanonical ensemble

Chapter 7

Molecular dynamics

Chapter 8

Direct translation

Chapter 9

Evaluation of energies and forces

Chapter 10

Canonical ensemble

Chapter 11

Thermostats