Name	*Surname*	*ID*

Midterm test No. 2

03 / 05 / 2023

Please complete the following tasks and submit the document in **PDF format** to damiano.piovesan@unipd.it by **12:30 PM on May 17, 2022**, two weeks after the deadline. Please include your **surname** in the **file name**.

Each student has been assigned a different **protein structure (<PDB ID>_<chain ID>)**. The assessment consists of two parts: open questions and an analysis of the assigned structure. You can find the **assignment file here**.

Please answer the following questions concisely, with a maximum of **500 words** in total:

1. What is the relationship between sequence similarity and structure similarity in biological proteins?
2. What are the main steps involved in homology modelling?
3. How can you measure the quality of a structural alignment?
4. What are the differences, in terms of amino acid composition, between globular and intrinsically disordered proteins?

Download the assigned PDB structure and consider only **standard (non-hetero) residues** of the specified chain (<PDB ID>_<chain ID>). Calculate the contact map (question 1) and the conformational energy (questions 2 and 3) as described in the IUPRED paper. The M and P matrices are available from the ***iupred_data.py***. The smoothed energy is the moving average of the raw energy over a window of 21 residues (±10 residues around the current position).

Complete the following tasks:

5. Calculate and plot the contact map of your chain. Use the ***NeighborSearch*** module and the ***search_all(3.5, level="R")*** method. Consider only contacts between positions with a **sequence separation ≥ 2**.
6. Calculate the **exact energy** of each residue based on the weighted contribution of its **contacts** (as calculated above) and plot the raw and smoothed energy for each residue on the same figure. Use the ***M matrix*** to calculate the contact energy.
7. Calculate the **estimated energy** of each residue based on the weighted contribution of the **frequency of neighboring amino acids** in the sequence and plot the raw and smoothed energy for each residue on the same figure. Use the ***P matrix*** to calculate the estimated energy. Neighboring residues are those 2-100 residues apart from the current position.

8. Report the **disorder content** for the two different calculations. Disorder content can be calculated as the fraction of **residues with positive energy** (≥ 0) over the length of the sequence. Please report both the fraction and the raw count of residues with positive energy.