Veronica	*Bedin*	*2097013*

Midterm test No. 1

29 / 03 / 2023

Please answer all questions below and submit this document in **PDF format** by **12:30 - 12 April 2022** (two weeks after) to **damiano.piovesan@unipd.it**.

Each student is assigned a different **protein structure (PDB ID)**. The entire exercise is based on the analysis of that structure. Please add your **name**, **surname**, **university ID** and **email** in the **assignment file** here which contains a list of PDB IDs. **Assignment file here**

Questions

Answer the following questions concisely (**max 500 words in total**).

1. What is the difference between an ionic bond and a covalent bond?

The ionic bond is between two ions with opposite charge -one metal and one non metal- while the covalent bond is between two atoms -two nonmetals or a nonmetal and a metalloid- that share one or more electron pairs.

The difference in electronegativity between the atoms must be greater than 1.9 for the bond to be considered ionic, less than 1.9 to be considered covalent.

There is a significant difference also in the energy of the bond: for the ionic is around 170-1500 kJ/mol, for the covalent 50-110 kJ/mol.

2. What is the difference between a weak and strong acid?

A strong acid reacts completely with water to form H_3O^+, causing a complete ionization, while a weak acid ionizes only slightly in aqueous solutions.

This reflects in the numerical value of $K_a = \frac{[H^+][A^-]}{[HA]}$, the acid ionization constant, that measures the fraction of the original acid that has been ionized in solution. For strong acids, the component $[HA]$ tends to zero so the K_a tends to infinity.

3. What is the isoelectric point of a protein?

The net charge of an amino acid changes based on the pH of the solution in which it is dissolved. The isoelectric point (pI) is the pH value at which an amino acid is found in its Zwitterionic form, meaning its net charge is null. pI is computed as the mean of the two pK_a from the two different reactions happening when changing the pH of the solution.

4. What is the “native conformation” of a protein?

The native conformation of a protein is the folded structure that reaches the minimum of the Gibbs energy, while still being operative and functional. Natural selection requires proteins to reach the native state in a biologically reasonable time in order to exist. That is because shape changing in proteins are the primary cause of several neurodegenerative diseases.

5. What is the difference between the asymmetric unit and the biological entity of a PDB structure?

The asymmetric unit is the smallest portion of crystal to which symmetry operations can be applied in order to generate the complete unit cell.

The biological entity instead, is the macromolecular assembly that has either been shown to be or is believed to be the functional form of the molecule.

There is not a 1:1 map from asymmetric unit to biological entity, as one asymmetric unit can contain:

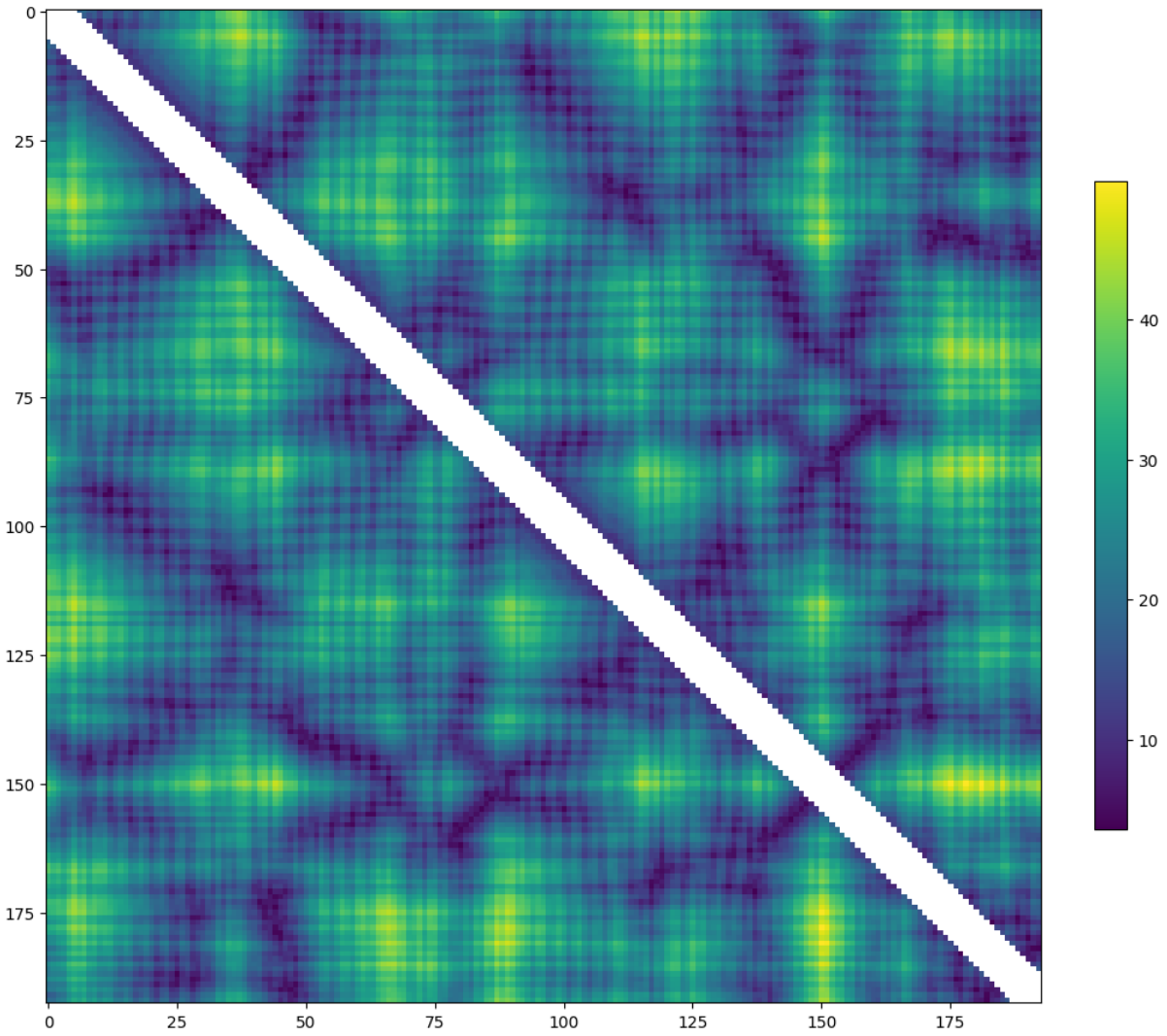
- One biological entity
- A portion of a biological entity
- Multiple biological entities

6. What are “missing residues” in a PDB structure?

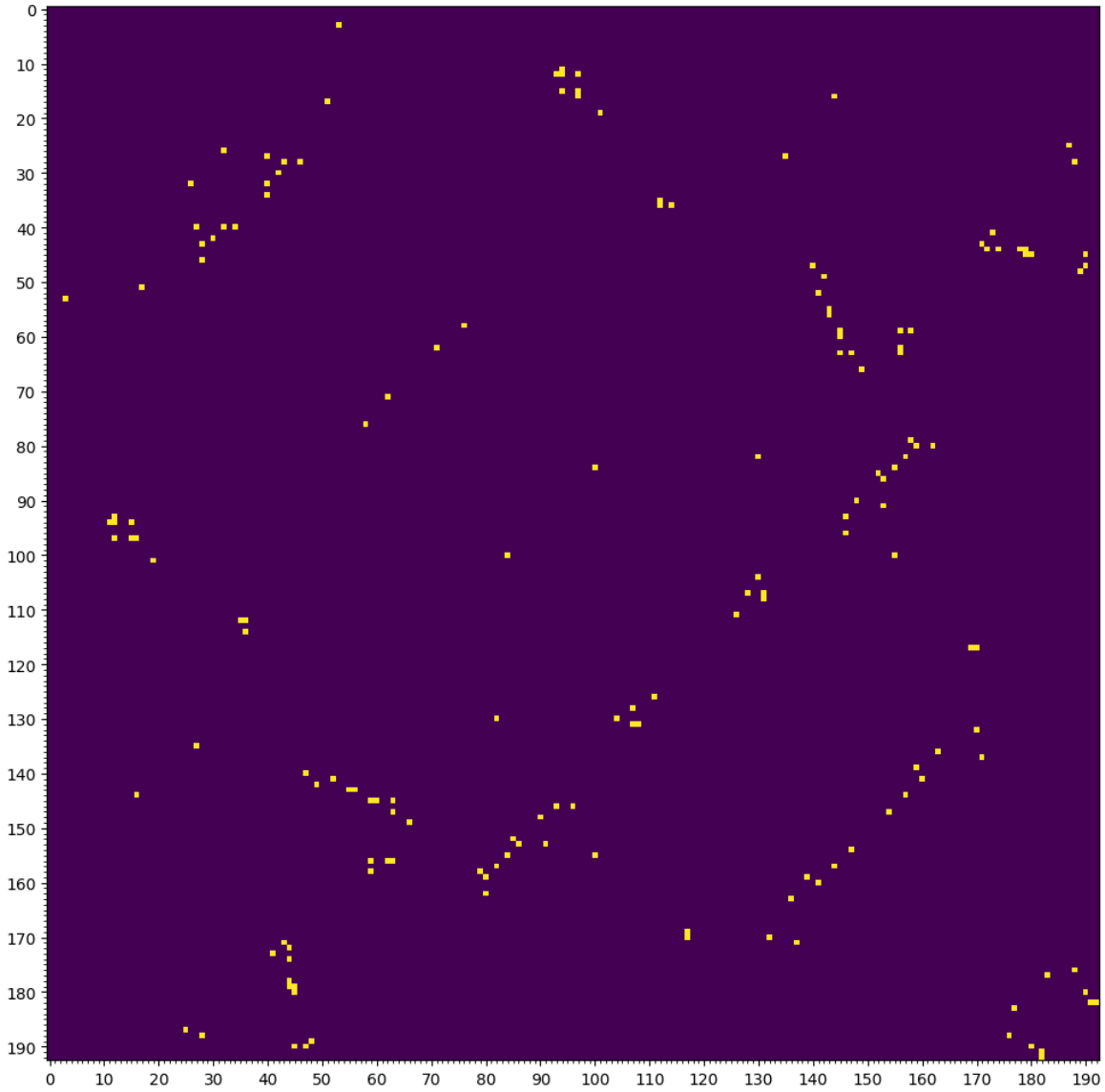
They are residues not included in the PDB file, their absence mostly caused by experiment limits. For example, in crystallography, parts of the protein that are too mobile will not be observed even if they exist in the protein.

Download the assigned PDB structure and create a new file removing non relevant chains (if it is a complex), solvent molecules and cofactors. For each question **concisely explain all passages (max 5 rows)** necessary to reproduce the results (e.g. parameters, database queries, algorithms, etc.). Optionally, if relevant, you can provide source code (not necessary).

1. Plot a heatmap representing distances between beta-carbons (CB), consider alpha-carbons when the CB is missing, i.e. for proline. Provide residue indexes along the axes.



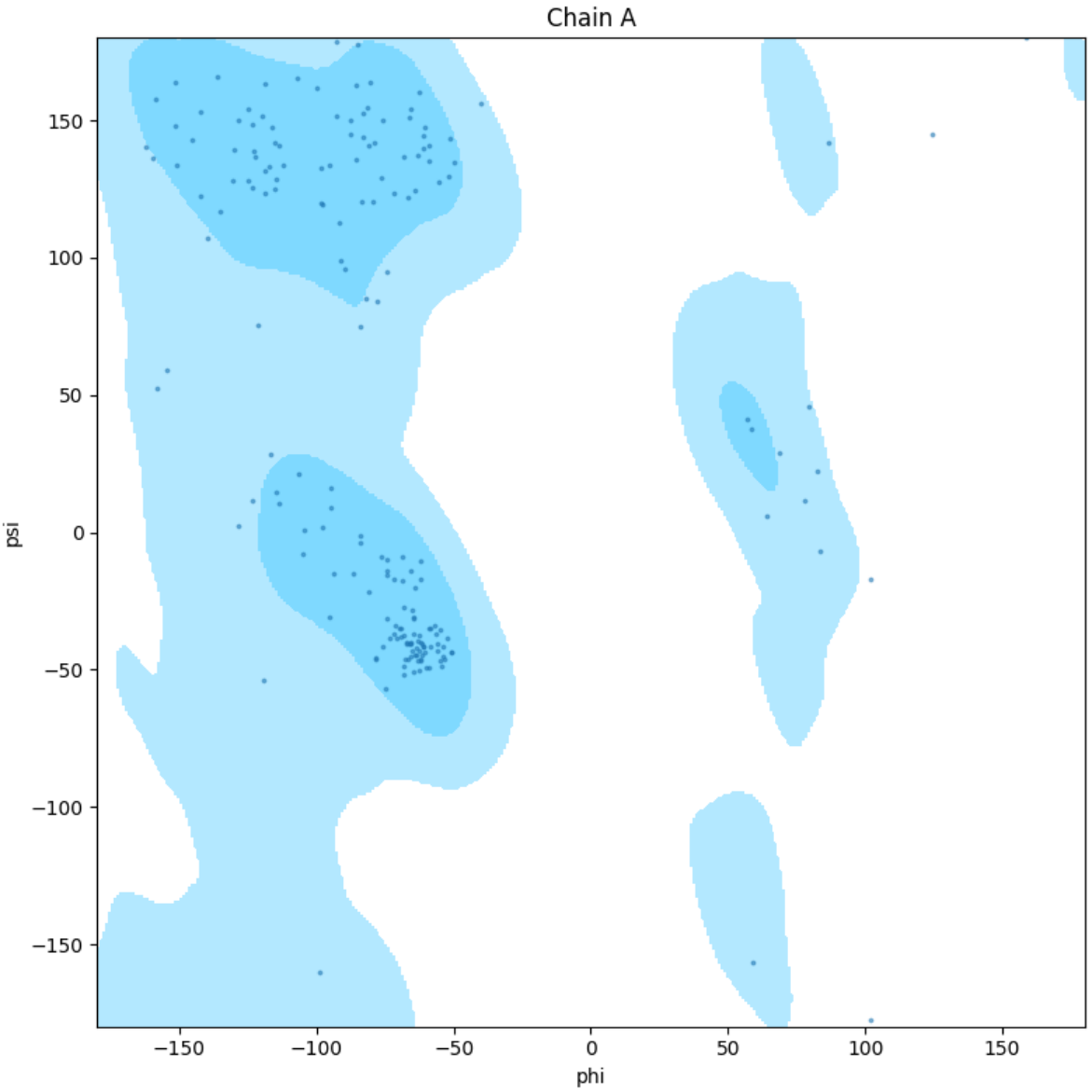
2. Plot the contacts map from the matrix above considering a distance threshold of 5 Å.



3. Report the number of residues in contacts for different ranges of sequence separation. Consider the following intervals [0,6], [7,12] and [13,24] and [25,∞].

```
Sequence separation [0,6]:  593
Sequence separation [7,12]:  53
Sequence separation [13,24]:  69
Sequence separation [25, ]:  349
```

4. Generate the Ramachandran plot of your protein.



5. How many residues are Ramachandran outliers? Consider the Ramachandran regions as shown during the practicals.

```
for chain_id in rama:
  count=0
  int_phi = [round(element)+ int(len(np.array(regions_matrix))/2) -1 for element in rama[chain_id][1]]
  int_psi = [round(element)+ int(len(np.array(regions_matrix))/2) -1 for element in rama[chain_id][2]]
  for phi, psi in zip(int_phi, int_psi):
    if np.array(regions_matrix)[::-1][psi,phi]==0:
      count+=1
print("The number of outliers is:", count)
```

```
The number of outliers is: 4
```