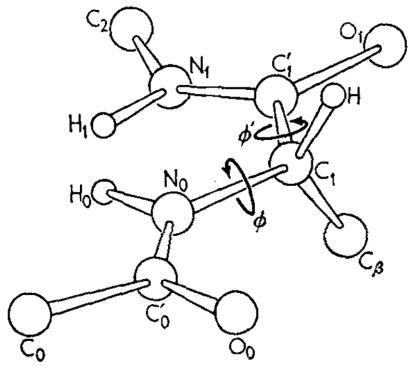
The Ramachandran Plot

1.0 Introduction:

The Ramachandran plots show the distribution of the torsion angles Φ and Ψ of the protein backbone. Ramachandran plots were initially researched and published in 1960s by scientists, Ramachandran, Sasisekharan, Ramakrishnan, and Thathachari, and these plots became foundational tools to understand structural biology.

In this project, we will be comparing and analysing two groups of protein data sets, with Ramachandran plots. One of the datasets is with high-resolution protein structure PDB files, and the other one is with a spread of high to low resolutions. We will extract Φ and Ψ angles from PDB data files, using Python code. For this purpose, we use the **gemmi** Python library.

From the extracted Φ and Ψ angle data, we make Ramachandran plots for both the groups of protein data. Filtering data of the second dataset, for the outliers we try to match the plot for the second dataset.



When two residues are linked at the αC atom, the rotations between the $\mathrm{N} - \alpha\mathrm{C}$ bond and $\alpha\mathrm{C} - \mathrm{C}'$ are possible. The torsion angle formed due to the rotation between $\mathrm{N} - \alpha\mathrm{C}$ is denoted as Φ and that formed due to the rotation between $\alpha\mathrm{C} - \mathrm{C}'$ is denoted as Ψ. When the two carbonyl carbon atoms are eclipsed, the $\Phi = 0°$, and when they are staggered the $\Phi = 180°$. Similarly, when two nitrogen atoms are eclipsed the $\Psi = 0°$ and when they are staggered $\Psi = 180°$. Due to the interactions between atoms, a set of discrete values for Φ and Ψ are possible, for any given protein structures. When these points are plotted, the point of the plot covers a specific region of the plot, and these regions correspond to specific configurations of dipeptide and tripeptide structures as well as polypeptide chains. This means if we know Φ and Ψ values of a residue, we can say the peptide chain configuration that the residue is part of.

2.0 Methodology:

In this project to extract Φ and Ψ angles, as well as the resolution and the average B-factor values, the **gemmi** Python library has been used.

To extract Φ and Ψ angles as well as resolution values,

- We run a loop over each file of the dataset using **gemmi.CoorFilewalk** function to get the file path, and load the model using **gemmi.read_structure** function.
- From each structure, we use **structure.resolution** function to extract resolutions values.
- We run a loop over each chain of the structure, each residue of the chain to extract Φ and Ψ angles.

```
# Iterate over each PDB file in dataset_1
for path in tqdm(gemmi.CoorFileWalk("download/dataset_1"), total=len(pdb_list_1)) :

    # Read the structure from the PDB file
    structure = gemmi.read_structure(path)
    model = structure[0]

    # Extract and store resolution value for each structure
    high_resolution_values.append(structure.resolution)

    # Iterate over each chain in the structure
    for chain in model:

        # Iterate over each residue in the chain
        for residue in chain :

            # Get the next and previous residues in the chain
            next_res = chain.next_residue(residue)
            prev_res = chain.previous_residue(residue)

            if next_res:
                phi, psi = gemmi.calculate_phi_psi(prev_res, residue, next_res)

                # Check if the calculated angles are valid (not NaN)
                if not isnan(phi) and not isnan(psi):

                    # Convert angles from radians to degrees and append to lists
                    phi_angles.append(degrees(phi))
                    psi_angles.append(degrees(psi))
```

3.0 Results:

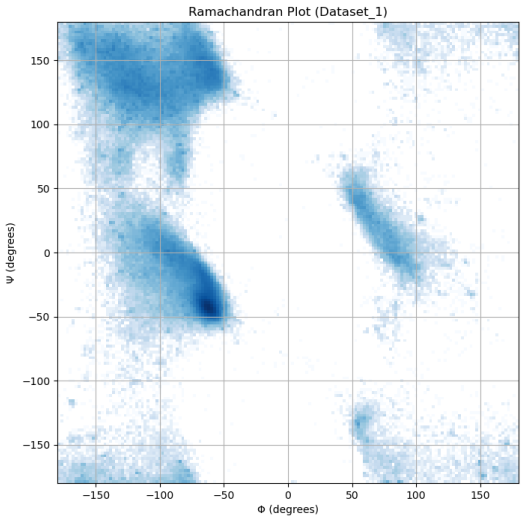


Figure 1

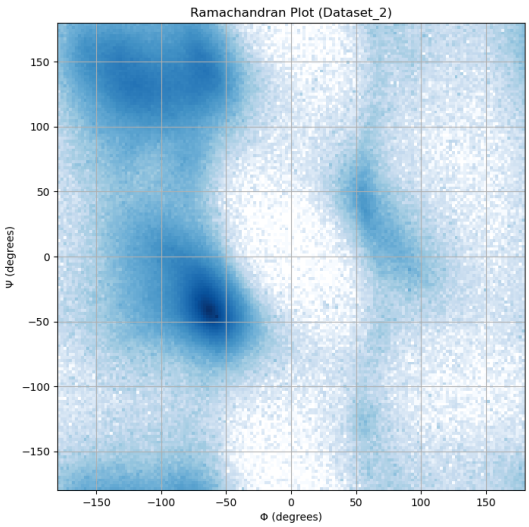


Figure 2

Ramachandran plot gives possible torsions angles Φ and Ψ for all residues of peptide chains. Due to chemical and ionic interactions, atoms of a residue orient in certain range of shapes. This mean, even though the torsion angles Φ and Ψ of residues range is $(-180°, +180°)$ they only take a certain allowed combinations. Ramachandra plots give these boundaries of values Φ and Ψ can take. We can see the clearly defined boundaries from Figure 1, Ramachandran plot for Dataset1.

At lower resolutions, two distinct objects are observed as one, hence lower resolutions are prone to give higher uncertainty values of positions and orientations. Therefore the low-resolution structures will produce noise in the Ramachandran plot. Even though the number of structures taken to plot Figure 1 and Figure 2 are approximately the same, Figure 2 is noisier and does not have well-defined boundaries.

4.0 Data Filtering:

The code shown in section 2.0 extracts Resolution values. Being able to detect tiny details is a hallmark of high resolution. Hence the smaller given length numbers correspond to better resolutions. The upper cap for resolution values of Dataset2 has been set to 3.4, which is the mean as well as mode of the resolutions of dataset_2. The plot after filtering the resolution values can be seen in Figure 3.

B-factor (atomic displacement parameter) describes the attenuation of X-ray scattering or coherent neutron scattering caused by thermal motion. To extract average B-factor values, a loop has been run for each chain in the model, each residue in the chain and each atom in the residue. The .**b_iso** function has been used to extract B-factor values. The average has been taken over the all atoms of the residue, and mean B-factor of the residue is calculated in each step.

```
# Iterate over each PDB file in dataset_2
for path in tqdm(gemmi.CoorFileWalk("download/dataset_2"), total=len(pdb_list_2)) :
    structure = gemmi.read_structure(path)
    model = structure[0]

    # Check if structure resolution is below the specified filter
    if structure.resolution < resolution_filter:

        # Iterate over each chain in the structure
        for chain in model:

            # Iterate over each residue in the chain
            for residue in chain :
                monomer_bfactor = [ ] # one for main-chains, one for side-chains
                next_res = chain.next_residue(residue)
                prev_res = chain.previous_residue(residue)

                for atom in residue:
                    monomer_bfactor.append(atom.b_iso)

                # Calculate phi and psi angles if next residue is present and
                # if average B-factor of the residue is less than filter value
                if next_res and np.mean(monomer_bfactor) <= bfactor_filter:
                    phi, psi = gemmi.calculate_phi_psi(prev_res, residue, next_res)

                    # Check if the calculated angles are valid (not NaN)
                    if not isnan(phi) and not isnan(psi):
                        # Convert angles from radians to degrees and append to filtered lists
                        filtered_phi_angles_2.append(degrees(phi))
                        filtered_psi_angles_2.append(degrees(psi))
```

In each iteration of **structure** in **Dataset2**, code will check for resolutions value and if it is less than 3.4 it will go further. Similarly in each iteration of **residue** in **chain** code will check for the mean B-factor of the reside. The limit has been defined as 34.8, the mode

value of all mean B-factors of the resides. Ramachandran plot for dataset_2 after filtering both resolution and mean B-factor values of residues is as shown in figure 4.

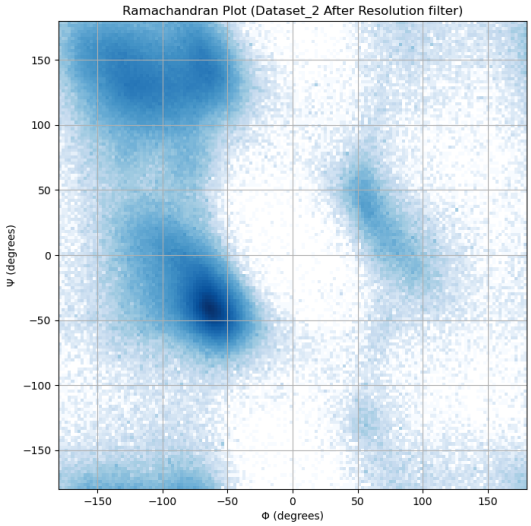


Figure 3

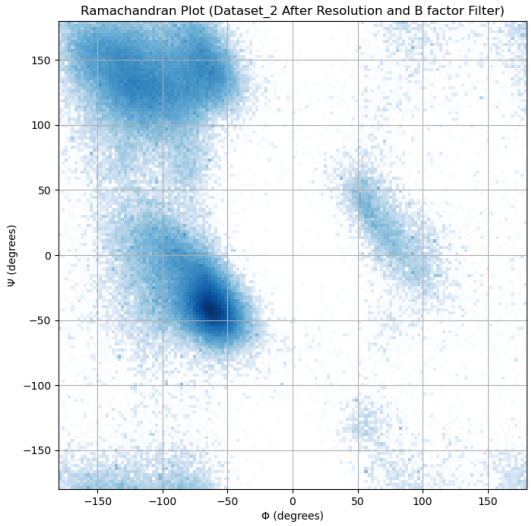


Figure 4

5.0 References:

1. Ramachandran, G.N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1), 95-99. https://doi.org/10.1016/S0022-2836(63)80023-6.
2. Title: Understanding PHI and PSI Angles, Ramachandran Plots, & Newman Projections.
   URL: https://www.youtube.com/watch?v=JyUMLSsbecI
3. Title: Tutorial: Ramachandran principle and phi psi angles
   URL:https://proteopedia.org/wiki/index.php/Tutorial:Ramachandran_principle_and_phi_psi_angles
4. Title: Resolution
   URL: https://proteopedia.org/wiki/index.php/Resolution
5. Title: Resolution
   URL:https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/resolution#:~:text=So%20resolution%20is%20a%20measure,in%20the%20electron%20density%20map
6. Sun Z, Liu Q, Qu G, Feng Y, Reetz MT. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. Chem Rev. 2019 Feb 13;119(3):1626-1665. doi: 10.1021/acs.chemrev.8b00290. Epub 2019 Jan 30. PMID: 30698416.