BT5340: Protein Folding and Stability

Assignment 1

February 9, 2025

Deadline for Submission: 4 pm on March 2, 2025

Instructions The code employed to calculate the phi-psi angles, details of the calculations and a write-up about the results along with figures/results should be sent to my email athi.naganathan@gmail.com (**as one zipped file with the title Name1_Name2.zip or Name1_Name2_Name3.zip; first names suffice**) (the database need not be attached as it will be too large a file). This will account for 30% of the total points. Late submissions or request for deadline extensions will NOT be entertained.

The submission will not be evaluated without the code employed – each group should write their own code. **Any kind of plagiarism, copying and use of chatbots/AI servers will not be accepted and will be reported to the Disciplinary Committee followed by a U grade**.

Question

Folding of specific protein segments (short sequences of 4-10 residues) are intrinsically sensitive to phi-psi preferences of amino acids within that segment. It is potentially possible to extract some signal by looking at phi-psi preferences of amino acids in the context of their immediate neighborhood.

To explore this question, download protein structures from the Dunbrack database with the following modifications to the default parameters:

- Set 'Maximum chain length' to 1000
- For 'Include NMR entries?', click on 'Yes'

On downloading the protein structures, write a code (whichever coding language you prefer to use) to extract the phi-psi angles of specific residues ('X') according to the list below.

- How does the distribution of phi-psi of residue X appear in the Ramachandran plot? In other words, bin the occurrences in 3° x 3° cells and plot the distribution.
- How different is the distribution from a random distribution?
- Explain the rationale for deviations, if any, you observe from a conventional Ramachandran plot seen in Biochemistry textbooks.
- Consider a 7-residue segment of sequence 'a-b-c-X-d-e-f' with X being at the center of this segment. Extract all such sequences from the database of protein structures you have created. Do you see any differences in the phi-psi preference of X if there is *at least* one charged residue either side of the central residue X? If so, rationalize the answer.

- Do you see differences in the phi-psi distribution of residue X (compared to the distribution generated from the first question) if the 7-residue segment is present in a helix or a strand?

Group	Amino Acid
1	G
2	S
3	W
4	P
5	I
6	A
7	T
8	Y
9	N