Biomedical Data Science | HW 2
Spring 2020

Programming students are responsible for Problems 1, 2, 4-11.
Non-Programming students: 3, 4-11.

For Problems 1 and 2, submit two programming files, one for each problem. You must also submit two figures, one of the Ramachandran plot for threonine and one showing the observed χ_1 side chain dihedral angle distribution for threonine.

For all other questions, typed or legibly scanned handwritten answers are accepted.

Part 1 – Programming Students

1. Ramachandran plots allow us to investigate the sterically allowed and disallowed backbone dihedral angle combinations φ and ψ in proteins. Using the file *core_THR_residues.txt* provided, produce a Ramachandran plot for threonine residues. The file *core_THR_residues.txt* contains 1000 threonine dipeptides taken from a database of high-resolution protein crystal structures. The C_α, caroboxyl carbon, and oxygen atoms on the prior amino acid are labelled pCa, pC, and pO. The N, C_α and H atoms on the subsequent amino acid are labeled: nN, nCa and nH. Using this file, calculate φ and ψ for each residue and produce a Ramachandran plot similar to that shown in Figure 1. See the lecture notes for definitions of φ and ψ. See Table 1 for a description of the data structure of *core_THR_residues.txt.*

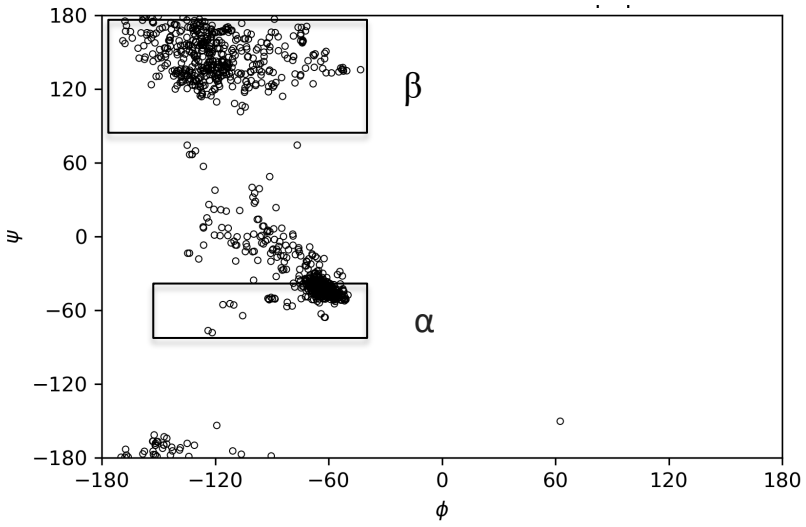

Figure 1: Ramachandran plot for 1000 threonine residues from high-resolution protein crystal structures. The strict Ramachandran limits for α-helices and β-sheets are shown in black rectangles.

2. In the lecture notes, we not only discussed backbone dihedral angles φ and ψ, but we also discussed sidechain dihedral angles. As the sidechains have different numbers of atoms, they can have different numbers of sidechain dihedral angles. In the case of threonine, there is only one sidechain dihedral angle, χ_1. Generate the observed side chain dihedral angle distribution from *core_THR_residues.txt* discussed in Problem 1. The observed distribution should be similar to that shown in Figure 2. See the lecture notes for the definition of χ_1 in threonine.

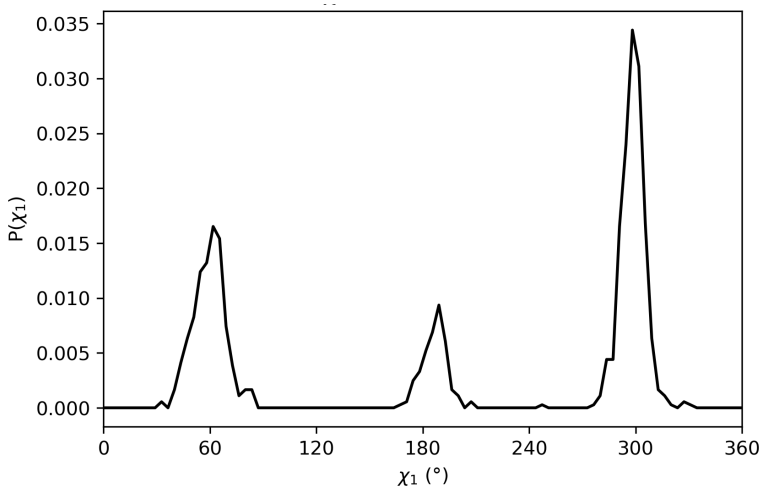

Figure 2: The side chain dihedral angle distribution P(χ_1) for threonine observed in high-resolution protein crystal structures.

Residue#	atom type	X	Y	Z
1	pCA	0.796000	-4.634000	-35.564999
1	pC	1.707000	-5.823000	-35.827999
1	pO	1.546000	-6.572000	-36.794998
1	N	2.657000	-6.020000	-34.912998
1	CA	3.666000	-7.064000	-35.043999
1	C	4.084000	-7.527000	-33.657001
1	O	4.127000	-6.739000	-32.709000
1	CB	4.865000	-6.547000	-35.873001
1	OG1	5.740000	-7.628000	-36.230999
1	CG2	5.685000	-5.477000	-35.141998
1	H	2.732000	-5.549000	-34.198002
1	HA	3.300000	-7.826000	-35.520000
1	HB	4.482000	-6.142000	-36.667000
1	HG1	5.315000	-8.195000	-36.681999
1	HG21	6.420000	-5.191000	-35.706001
1	HG22	5.118000	-4.716000	-34.939999
1	HG23	6.036000	-5.846000	-34.317001
1	nN	4.370000	-8.826000	-33.562000
1	nCA	5.080000	-9.401000	-32.428001
1	nH	4.153000	-9.402000	-34.162998

Table 1. An example of a THR residue in core_THR_residues.txt. The first column is the residue number, the second column is the atom type, and the last three columns are the x-, y-, z-coordinates.

Part 1 – Non-Programming Students

3. Derive the expressions for the x-, y-, and z-components of the force $\vec{F}_j$ on atom j=i+1 from the previous atom i and successive atom k=i+2 using the bond angle potential, $V_{ba} = \frac{k_\theta}{2}\left(\theta_{ijk} - \theta_0\right)^2$, where k_θ is the constant bond stiffness, $\theta_{ijk} = \cos^{-1}\left(\frac{\vec{r}_{ij} \cdot \vec{r}_{kj}}{r_{ij} r_{kj}}\right)$ is the bond angle between bonded atoms i, j, and k, $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$, and θ_0 is the preferred bond angle. Note that $\vec{F}_j = \frac{-dV_{ba}}{dx_j}\hat{x}\frac{-dV_{ba}}{dy_j}\hat{y}\frac{-dV_{ba}}{dz_j}\hat{z}$.

Part 2 – Programming and Non-Programming Students

4. Suppose a unbranched, simple polymer has N monomers in three spatial dimensions. How many dihedral angles does it have? How many bond angles (between three adjacent monomers) does it have?

5. Name three highly hydrophobic amino acids.

6. Name three types of terms in typical potential energy functions for proteins.

7. How does the radius of gyration R_g scale with the number of amino acids N for a Gaussian chain and a fully extended chain?

8. What is the difference between a deterministic and stochastic model? Which term applies to Ordinary Differential Equation (ODE) models?

9. The SIR model can be used to predict the course of an epidemic in a large population. In the model equations given below, describe the meaning of each state variable (S, I and R) and each parameter (β and μ):

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \mu I$$

$$\frac{dR}{dt} = \mu I$$

10. In the context of mathematical modeling, what is an F test used for? What does it mean when $F >> 1$?

11. What does it mean for a model parameter to be identifiable?