0.1 Extra Credit

Part A:

Download a PDB file for a protein and

1. Compute the distance distribution between sequential C_α carbons. What is the mean of the distribution?

2. Compute the distance distribution between C_α carbons separated in sequence by k. That is, the sequential neighbors have $k = 1$. How does the mean distance vary as a function of k? Compare the distributions for $k = 3$ and $k = 4$; which has C_α carbons closer together?

3. Compute the N-O distance distribution between all pairs of carbonil and amide groups in the peptide bonds. What is the part of the distribution that corresponds to ones forming a hydrogen bond? (Hint: exclude the N's and O's that are in the same peptide bond as well as those in sequence neighbors.)

Part B:

Proteins are *oriented*: there is a C-terminal end and an N-terminal end. Determine whether there is a bias in α-helices in proteins with regard to their *macrodipole* μ which is defined as follows. Suppose that a helix consists of the sequence $p_i, p_{i+1}, \ldots, p_{i+\ell}$ where each p_j denotes an amino-acid sidechain. Let $\mathcal{C}(p)$ denote the charge of the sidechain p, that is, $\mathcal{C}(D) = \mathcal{C}(E) = -1$ and $\mathcal{C}(K) = \mathcal{C}(R) = \mathcal{C}(H) = +1$, with $\mathcal{C}(p) = 0$ for all other p. Define

$$\mu(p_i, p_{i+1}, \ldots, p_{i_\ell}) = \sum_{j=0}^{\ell} \mathcal{C}(p_{i+j}) \left(j - \frac{1}{2}\ell \right) \tag{1}$$

Plot the distribution of μ over a set of proteins. Compare with the peptide dipole, which can be modelled as a charge of $+0.5$ at the N-terminus of the helix and a charge of -0.5 at the C-terminus of the helix. How does this differ for left-handed helices versus right-handed helices? (Hint: the PDB identifies helical regions of protein sequences. The peptide dipole in our simplification is just ℓ, so μ/ℓ provides a direct comparison.)