Exercise 06

Analysis of MD Simulations II

Deadline: Please hand in Exercise 06 in **pdf format** by **Thursday, 22th 10.15 a.m.** to stevan.aleksic@fu-berlin.de

6.1 Normalized Euclidean distance (10)

The Euclidean distance or Euclidean metric is the "ordinary" (i.e straight line) distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. Euclidean distance can be used to measure the difference between the histograms of χ-1 angles of a pair of amino acids.

Write a Python script, which calculates the normalized Euclidean distance between the χ-1 histograms of RES_1 and RES_2, according to the following equation:

$$\delta = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(RES1_i - RES2_i)^2} \tag{1}$$

where δ is a normalized Euclidean distance for the given pair of amino acids, N=360 (number of bins), $RES1_i$ the probability of finding the χ-1 of RES1 in i-th bin, and $RES2_i$ the probability of finding χ-1 of RES2 in i-th bin.

6.2 Kullback-Leibler divergence (10)

In probability theory, and information theory, the Kullback-Leibler divergence is a non-symmetric measure of the difference between two probability distributions P and Q. Similarly to Euclidian distance,Kullback-Leibler divergence can be used to estimate the difference between the histograms of χ-1 angles of a pair of amino acids. For discrete probability distributions P and Q, the Kullback-Leibler divergence of Q from P is defined to be:

$$\delta_{PQ} = \sum_{i=1}^{N} P(i) ln \frac{P(i)}{Q(i)} \tag{2}$$

By implementing equation (2), write a matlab script which computes Kullback-Leibler divergence of discretized χ-1 histograms for RES1 and RES2 (N=360).

6.3 Mutual information (50)

In probability theory and information theory, the mutual information (MI) of two random variables is a measure of the variables' mutual dependence. The mutual information of two discrete random variables X and Y can be defined as:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \tag{3}$$

where p(x,y) is the joint probability distribution function of X and Y, and p(x) and p(y) are the marginal probability distribution functions of X and Y respectively. The values of MI are not confined to a certain interval. Therefore, in practice, one uses the normalized mutual information NMI, which is confined to [0,1], with NMI=0 corresponding to absence of mutual dependence.The normalized mutual information is given as:

$$NMI(X,Y) = \frac{MI(X,Y)}{min(H(X),H(Y))} \tag{4}$$

where H(X) is the informational entropy of the marginal probability distribution of variable X, given as:

$$H(X) = -\sum_{x \in X} p(X) log(p(X)) \quad (5)$$

Based on exations 3 to 5, write a matlab script, which calculates informational entropies, mutual information, and normalized mutual information of χ-1 angles of RES1 and RES2.

6.4 Computing ^{3}J-coupling constants from MD data (30)

^{3}J coupling constants can be calculated from the MD simulations by implementing Karplus equation:

$$^3J = A\cos^2(\phi + \theta) + B\cos(\phi + \theta) + C \quad (6)$$

where A,B,C are empirically calibrated parameters, ϕ is a backbone torsion angle, and θ is a correction angle.

Write a Python script and calculate the ^{3}J-coupling constants for the five residues (files 1-5.ascii). You should convert ϕ trajectories from degrees into radians, since np.cos works with radians. After the ^3J constants are calculated for each time step, an average value should be taken, and standard deviation of the Karplus curve should be added (see parameteres below). Produce errorbar plots by comparing the experimental values with the computed values of ^3J-coupling constants, and decide which parameters predict coupling constants more accurately.

Empirical Parameters accoridng to Bax: A=7.09 B=-1.42 C=1.65 θ=-60 STD=0.39

Empirical Parameters accoridng to Schmidt: A=7.9 B=-1.05 C=1.55 θ=-60 STD=0.78

6.5 Files

https://www.dropbox.com/sh/88lydj4g4dubz30/AABHNfP1G9k2Fxj50Yo_2MLaa?dl=0