Exercise 05

Simulation data analysis with GROMACS

Deadline: Please hand in your protocol in pdf format by Thursday, the 30th of Mai 2019, 10 am to jan.joswig@fu-berlin.de or marco.mannig@fu-berlin.de. The protocol should contain analytical solutions, short discussions, Python-code and plots.

Two simulated protein trajectories (A and B) to use in this exercise can be found at `login.bcp.fu-berlin.de:/home/janjoswig/MD18/Ex06/`.

6.1 Processing simulation output

The provided trajectories belong to two different systems (A: apo-langerin, B: holo-langerin, protonated at H294 and H229 sidechains) and contain only coordinates of protein-atoms. Solvent and ions have been removed. Find out how many atoms belong to each system, how many samples (time-frames) are in the trajectories and how big the time-step between the frames is (*hint:* `gmx check`).

When you look at the data in VMD, you will notice, that the molecule moves around in the box and may also appear to be broken, when its leaving the box. For the further analysis, periodic boundary conditions have to be fixed (*hint:* `gmx trjconv -pbc mol -s .tpr`) and the trajectory should be fitted to a reference structure (*hint:* `gmx trjconv -fit rot+trans -s reference.gro`). Choose "Protein" as group, when your asked for the output, and "Backbone" as group for the fit. In VMD the molecule should now stay always in the centre of the box.

6.2 Root mean square fluctuation (RMSF)

The RMSF of an atomic coordinate x_i is defined as its time-averaged displacement (euclidean distance) from a reference or the mean value $\mu(x_i)$.

$$\sigma_i = \sqrt{\frac{1}{t_n}\sum_{t=0}^{t_n}\left(x_i(t) - \mu(x_i)\right)^2} \tag{1}$$

Calculate the RMSF for all C_α-atoms in the given trajectories (*hint:* `gmx rmsf -s reference.gro`). Show the result (RMSF vs. residue-index) in one plot to see the difference between the two systems (Note that there is an offset of 198 in the numbering of protein residues, that is C_α-atom 1 belongs to G198).

6.3 Root mean square deviation (RMSD)

In contrast, the RMSD of a set of atomic coordinates x_i to x_n is defined as the averaged displacement from the reference or the mean values at specific time-steps.

$$\tilde{r}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i(t) - \mu(x_i))^2} \tag{2}$$

Collect the RMSD for all atoms of the residues K257 to W264, using the provided index files (*hint:* `gmx rms -s reference.gro -n index.ndx`). Show the result (RMSD vs. time) in one plot two see the difference between the two systems.

It can be useful to illustrate a time-series of data-points in a histogram instead. Calculate normalized histograms for the two systems (*hint:* numpy.histogram, density=True. You probably need to specify a reasonable range, e.g. range=[0,1], and the number of bins, e.g. bins=100) and plot them in one figure.

6.4 Dihedral angles

Extract dihedral angles for the backbone and sidechains of all residues (*hint:* `gmx chi -maxchi 4 -phi -psi`). Plot the χ_1- and χ_2-angle distribution for H294 as normalised histogram in the angle-range of -180 to 180.

Backbone dihedral distributions are ideally visualised as Ramachandran-plot, which corresponds to a 2D histogram of the $\phi-$ and ψ-angles (*hint:* numpy.histogram2d, normed=True). To obtain a pseudo free energy surface from the data, take the negative logarithm of the histogram. For plotting you can use matplotlib.pyplot.imshow. Add a colorbar for the value range between 0 and 6. Compare the plots for the residues A289, P286 and K257. Which regions (α, β, ...) are populated?

6.5 Atomic distances

Collect the distance between the C_α-atoms of M260 and G290 for the given trajectories (*hint:* `gmx mindist -s reference.gro -n index.ndx`). Create the necessary index file yourself (*hint:* `gmx make_ndx -f reference.gro` or just pick the atomic-indices from the .gro-file manually). The Index file should have the form:

[*name atom1*]
index atom1
[*name atom2*]
index atom2

Plot the distance distributions as normalised histogram in one figure. Render a figure in VMD that shows aligned example structures corresponding to the maxima in this distributions. What low energetic (highly probable) conformations are possible?