Activity 3: MD simulation and trajectory analysis of BPTI

Overview

In this exercise we will carry out an MD simulation of the protein bovine pancreatic trypsin inhibitor (BPTI). This was one of the first proteins to ever be simulated in its entirety, and the results of this simulation – that at its time was a computational tour de force – can be found in a Nature article from 1977 by Andy McCammon, Bruce Gelin, and Martin Karplus (co-recipient of the 2013 Nobel Prize in Chemistry). Today we'll carry out a very similar simulation in just a few minutes!

Starting structure and simulation preliminaries

The X-ray crystal structures (and NMR-derived structures) of many folded proteins can be found at the Protein Data Bank (http://www.rcsb.org). Every structure deposited in the PDB has a four character code; for this exercise we will use the structure 4PTI. This structure was deposited in 1982 (!), but has excellent resolution (1.5 Å) and provides a good starting point for our simulation.

Taking a structure from the PDB and getting it ready for simulation is not a trivial task. For us to carry out a gas phase simulation of this protein, crystallographic waters must be removed and disulfide bonds between various cysteine residues must be specified. Thankfully there are applications that help to automate this task. In this particular case, we used the application *pdb4amber* (part of the AmberTools distribution; see http://www.ambermd.org) to generate an appropriate pdb file. We then used another AmberTools application, *tleap*, to create parameter/topology (*BPTI_gas.prmtop*) and starting coordinate (*BPTI_gas.inpcrd*) files that can be understood by Amber or OpenMM. For this exercise the force field we will be using the Amber ff14SB protein force field.

MD simulation protocol

The notebook *BPTI_OpenMM.ipynb* contains all of the necessary commands to use OpenMM to carry out a gas phase simulation of BPTI. As written, it requires the two aforementioned files, *BPTI_gas.prmtop* and *BPTI_gas.inpcrd*, to be in the same directory as the notebook.

This notebook will carry out a simulation protocol very similar to that of McCammon et al. (as well as our earlier exercises), with *some* modernized aspects to it:

1. Up to 100 steps of energy minimization using the L-BFGS algorithm. [In the original study the authors performed 100 steps of MD with initial velocities set to zero/a starting temperature of 0 K.]

2. A 1.0 ps MD simulation to bring the BPTI protein to an equilibrium temperature of 298 K (25 °C). [Original study: velocities from previous phase are multiplied by 1.5 and then 0.25 ps of equilibration in the NVE ensemble are performed, after which the system is at 285 K. This means the simulation was carried out without thermostatting and instead the total energy of the system – kinetic + potential – was conserved.]
3. A 9.0 ps MD simulation at 298 K in which structures are saved every step (1 fs) into a file called *BPTI_sim.dcd*. [This is basically identical to the original study except that the original was again performed without thermostatting. Additional note: These days we usually do not save protein coordinates more frequently than once every 1 ps, but our simulations are typically 5-6 orders of magnitude longer than 40 years ago!]

Download the necessary files and run the notebook. These simulations take around three minutes to complete on a 2012 MacBook Air.

Trajectory analysis

For this exercise we are going to provide a bit less guidance than before. Download and open the notebook *BPTI_MDTraj.ipynb*. This notebook expects to be in the same directory as *BPTI_gas.prmtop* and *BPTI_sim.dcd*. Our overall goal is to reproduce – in a rough way – Figure 3a from McCammon et al., as shown below.

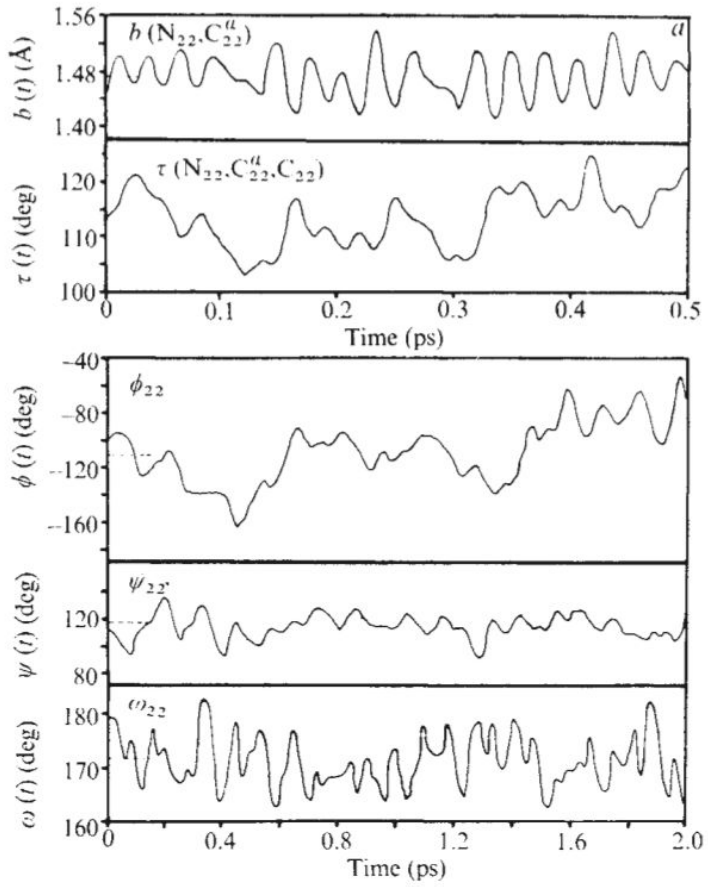

Figure 3a from McCammon, Gelin, and Karplus, *Nature* 1977.

One thing we need to do is figure out all of the atoms that will be involved in the bond length, angle, and torsions we want to analyze. The notebook contains code to do this for you. The notebook also contains code to plot the N-Cα bond length vs. time for residue Phe22 for 0.5 ps, beginning with step 5000. You should write code to:

1. Convert the N-H bond length from nm to Å and then plot the bond length vs. time over the same time interval as the N-Cα bond. How do the dynamics of this bond differ from those of the N-Cα bond? [Just by eyeballing the range of N-H bond lengths in this plot, the distribution of the N-H bond length should look *very* different from the C-H bond lengths in the butane simulation. What's different about this simulation?]
2. Convert the N-Cα-C bond angle from radians to degrees and then plot the bond angle vs. time over the same time interval.
3. Convert the φ, ψ, and ω torsions from radians to degrees and then plot them vs. time for 2.0 ps, also beginning with step 5000. You might notice that the ω torsion plot looks suboptimal because this torsion angle oscillates around ±180°. How might we fix it to look more human-readable?

...and that's a wrap!

Appendix (for those interested in how we prepared the prmtop/inpcrd files)

Assuming you have AmberTools18 installed on your computer, you can readily reproduce our simulation preparation process:

1. Download the file *4PTI.pdb* from the PDB website into the directory of your choice.
2. In a terminal window, change to that directory and type the following command:

```
pdb4amber -i 4PTI.pdb -o 4PTI_withCYX.pdb --dry --nohyd
```

3. Create a text file *setup_BPTI.leap* with the following content:

```
source leaprc.protein.ff14SB
bpti = loadpdb 4PTI_withCYX.pdb
bond bpti.5.SG bpti.55.SG
bond bpti.14.SG bpti.38.SG
bond bpti.30.SG bpti.51.SG
saveamberparm bpti bpti_gas.prmtop bpti_gas.inpcrd
savepdb bpti bpti_gas.pdb
quit
```

4. Type the following command:

```
tleap -f setup_BPTI.leap
```

You should now have Amber-format prmtop and inpcrd files for BPTI in the gas phase, as well as an Amber-friendly pdb file.