Lab exercise 8: Gromacs Simulation - Biophysical Chemistry

April 23th 2018

1 Introduction

In this lab exercise we will attempt to do a simple molecular dynamics (MD) simulation. We will however not bore you with all the complexities that goes into a proper MD simulation, since the this would make up a course in itself, but rather let you dive right into using it.

The system that we will simulate is a small box made up of a bilipid membrane, the Trp-cage and water. The Trp-cage is a simple protein system that is considered to be the smallest protein. It consists of a single alpha helix and a tail that together form a cage around a hydrophobic tryptophan side-chain, hence the name. It is a designed sequence that is non-functional, but is commonly used as a test case for simulations due to its size.

In this exercise we will pull the Trp-cage from one side of the membrane through to the other side, by applying an external pulling force.

2 Login & Setup

To get long enough simulations within reasonable time we will be using the computational resources of the cryo-EM infrastructure located at Scilifelab in Stockholm. However, you are free to do this from home on you own computer (if you for some reason like fan noise), but you'll likely will have to wait 10 times longer for each simulation to finish. Also you will have to install Gromacs on your own.

Login to the access-node by running `ssh -X <user>@login.cryoem.se`, where you replace `<user>` with your provided username and password.

You will need to download the simulation topology file to give as input to the simulation engine. The name of the topology file is `topol.tpr`. If you have a file available online you can easily download it by running `wget http://<url>`, where you replace `<url>` with the address to the file. You can also find the files on the access node through the path:

`/nethome/projects/BioPhysLabs/lab8.tar.gz`

Create a project directory inside your user directory and place all the files for this exercise in there.

We will be using a queuing system for dispatching computational jobs to the cluster. This will require a job-script that contains all the instructions for running the job, which is called `submit.job`. You should also have this file in your project directory.

3 VMD

We will be using a different visualization tool that is more specialized for simulations. It is called VMD and is freely available for download from www.ks.uiuc.edu/Research/vmd/. If you prefer a smooth interaction with the program we recommend that you install it on your local computer, but you can also run it on the access node by running `/nethome/dari/vmd/bin/vmd`. Note that for this to work you will have to have initiated the ssh-session with xForwarding, i.e. using the flag -X. If you prefer to run it on your own computer you will have to download the needed files from the node.

Try visualizing the initial conformation of our system by loading the `conf.gro`-file into VMD. This will show the starting conformation of the simulation box before applying any dynamics.

4 Dispatching the Job

When all of your files are setup properly you can, inside the project directory, submit a simulation job by dispatching the job-script to the queuing system. Do this by running `sbatch submit.job`.

You can monitor running jobs but issuing the command `squeue`. This will show you all the running jobs, the total time they have been running and the owner of the job.

While the job is running you can follow the log that is outputted by the simulation by running `tail -f md.log`. This will continuously output the content of `md.log` as it gets updated.

5 Analyzing the Simulation Results

The simulation will have created a number of files inside you project directory. Here are the ones you need to keep track of:

- `traj_comp.xtc` - The simulation trajectory, i.e. structure of the conformations of the system thoughout the simulation, we output frames every 5 000 MD step.
- `md.log` - Simulation log output
- `ener.edr` - File containing statistical numbers about the simulation at different steps.

Visualize the simulation trajectory in VMD by first loading the gro-file that you've been provided and then loading the trajectory file. You can play through the trajectory with the play button. However, as you will notice, it looks ugly. This is because of the periodic boundary condition (PBC). We will fix this using a Gromacs tool called `trjconv`. But first you need to load the Gromacs module. Run `module load gromacs`. Now you can access all the tools through the command `gmx`. Try running it and look at the output. There are a set of operation that can be executed with this command. To show them all you can run `gmx help commands`. You can read the manual for each operation by running `gmx <command> -h`, where you replace `<command>` with the operation name. Now try `gmx trjconv -h`.

To fix the PBC issue run:

```
gmx trjconv -s topol.tpr -f traj_comp.xtc -o traj_comp_pbc.xtc -pbc mol -ur compact
```

You will have to input a number for selecting the part of the system that you would like to output to the new file. We want everything, so input `0` and hit Enter. Now visualizing the output file `traj_comp_pbc.xtc`.

Tasks

1. Look at the trajectory before and after fixing the PBC issue, what is the cause of the ugliness?

2. Explain in a few sentences what a periodic boundary condition means when doing simulation. Use Google if you're unsure.

6 Running a Longer Simulation

The provided submit.job will run the simulation for 20 000 MD steps. You can control the number of steps that is run by modifying it. Open it up in a text-editor and change `nsteps=20000` to `nsteps=350000`. Now submit the new job-script. This time it will take more time, so you'll have to wait.

When the simulation is finished preform the same command as above to fix the PBC issue, and visualize the results.

Tasks

1. Knowing that the box size is approximately 9 nm along the Z-axis, that frames are written out every 5000 MD step and that each MD step is 4 fs long, what is the approximate average velocity of the Trp-cage throughout the simulation in meters/second?
2. How would you, using the information you have about the position of the Trp-cage, approximate the free energy difference between being inside the membrane (surrounded by lipids) and outside (surrounded by water)?
3. What sources of error do you have in this approximation? Consider for instance the shape of the membrane.
4. How would you do the pulling differently to improve the approximation, without going all the way to Umbrella sampling?

7 Analyzing the Statistics

Use the command `gmx energy -f ener.edr` and select the temperature, by inputting the index for it and pressing Enter. You will have to input also a zero and press Enter for the program to exit and output the data. Use the provided `plot.py` python-tool to plot the output file, named `energy.xvg`. You can issue `-h` as usual to show help for the tool.

Tasks

1. Include a plot of the temperature in your report. What is the average temperature of you system?
2. Do the same thing as above but for the option `Entalpy`. This will show you the sum of all enthalpies in your system. Explain the bump in enthalpy that you should see. Remember that an increase in enthalpy means less interaction.
3. Should you see a similar effect in any of the other statistical values that you can select from in the options list, which one?