Assignment 3

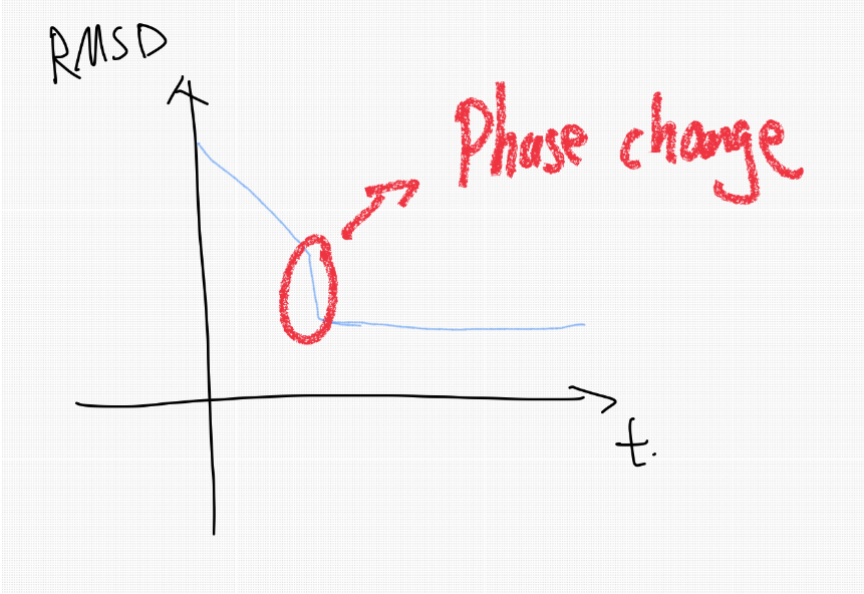
Yusheng Zhao

April 29, 2023

1 Problem 1

From statistical mechanics, we know the following rule. When there is a phase change, some function related to the structure of the system will change abruptly. More specifically, you could calculate **RMSD** for the system under investigation. Over time, you should see a discontinuity in the plot of time versus **RMSD**. That's when a phase change occurred.

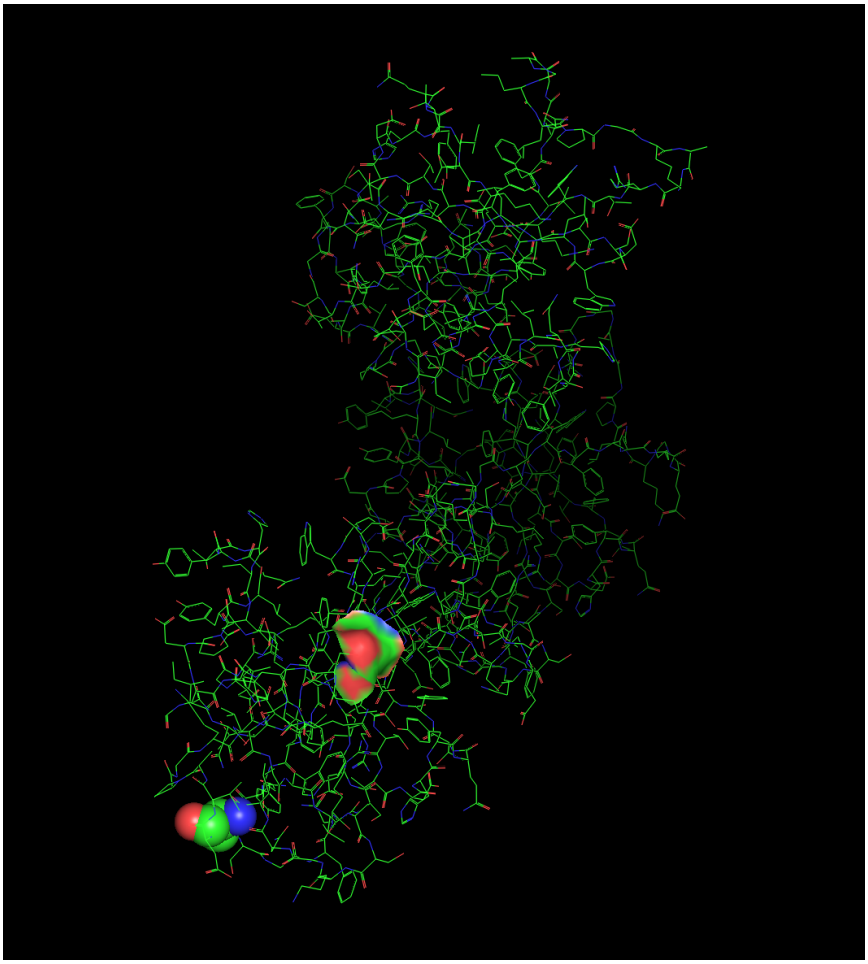
A plot may look something like this.



2 Problem 2

In a folded protein, I expect to find hydrophobic amino acids in the **innerds** of the structure. Conversely, I expect to find hydrophilic amino acids on the **outside** of the folded protein. This is basically due to the need to minimize energy in the folded state of a protein.

- Hydrophobic Amino Acid: Alanine
- Hydrophilic Amino Acid: Aspartic Acid
- I have illustrated the protein: BARNASE MUTANT WITH ILE 88 REPLACED BY ALA (1BRJ). In addition, I illustrated the hydrophobic amino acid, Alanine, as balls. I also illustrated the hydrophilic amino acid, Aspartic Acid, as surface.



3 Problem 3

Professor Chu gave a very good flow chart on 27 of Week 10 slides. I could not have done better and don't feel like robbing him of his work. Therefore, the flow chart will be omitted. But the following steps are summarized from that lecture note.

3.1 Structure Conversion and Topology

- Load PDB file
- Choose a certain force field

- Generate topology file with `pdb2gmx` command

3.2 Define Periodic Boundary Condition

- Define the box for PBC
- Limit minimal interaction distance

3.3 Add Solvent and Ion

- Add solvent explicitly with `solvate` command
- Replace some solvent molecule with ions to make system charge neutral.

3.4 Energy Minimization

- Equilabrate the system by performing energy minimization on the system.
- It's necessary because the added solvent might have created a large repulsion on the system that will ruin the MD simulation process.

3.5 NVT Ensemble Equilibration

- Couple system to heat bath and equilibrate the temperature of the system to desired value.
- Run simulation to allow for equilibration.

3.6 NPT Ensemble Equilibration

- Turn on the pressure coupling. Allow for the system to equilibrate.
- The end of simulation. Get ready to analyze result.

4 Problem 4

4.1 A

In general, we cannot **accurately** estimate the binding affinity. By definition, binding affinity is the concentration of ligand where half of protein is bounded with the ligand. It is statistical average value. Therefore, we need

a statistical ensemble to accurately estimate it. For a single simulation we might be able to rely on ergodicity. However, we don't know how long the simulation ran. So ergodicity condition may not apply. Single trial may not represent all possible starting configuration of the drug. Furthermore, random fluctuation during the simulation may render the simulation result not representative.

4.2 B

According to the lecture note, I propose to calculate the binding affinity from the free energy calculation. Free energy calculation will be carried out using the Free Energy Perturbation (FEP) method and alchemical method to speed things up.

Firstly, we add a non-existing force onto the drug molecule to slowly drive it to bind with the target protein. Then, we divide the process of evolving from the un-binded state to the binded state into many small steps where the initial and final configuration of the drug and protein within each step is not too different. This is a perturbation. Free energy difference of between the perturbed state and un-perturbed state is easily calculable.

In case of a driving process being extremely long with the added force, we could deploy alchemical method to directly induce a mutation to speed up the process.

Lastly, the free energy difference along the entire process is accumulated to get the free energy difference between the un-binded and binded state. The binding affinity could be derived from the free energy difference.