Computer Lab 5 - Protein Simulation in Explicit Water

Part B: Analyze Trajectory

I. Introduction. At the end of the last lab you started the production phase of your protein simulation in explicit water. You should now have a trajectory called *prod.dcd* in your *protsim/* subdirectory on the cluster. In today's lab we will analyze the trajectory for the following characteristics:

- Energy distribution
- Polypeptide backbone conformational fluctuation
- Overall conformational fluctuation

Answer the questions highlighted in yellow and send the answers to **achin14@jhu.edu** either in the body of an email or as an attached file with your JHEDID in the name (e.g. *JHEDID_lab5b.txt* if made with vi or *JHEDID_lab5b.docx* if made with MSWord).

II. Energy Distribution. The immediate goal of any molecular simulation is to achieve equilibrium and generate a Boltzmann distribution of energy states. Did your simulation achieve this goal?

1. While in the *protsim/* directory on the Mac fetch the following files from the cluster,
   prod.dcd
   production.log
   back to the Mac.
2. Gather the total energy data from the log file for plotting.

```
grep ^ENERGY production.log | awk '{print $2, $12}' > prod.dat
```

3. Plot these data in **xmgrace** (or your favorite plotting app) and calculate a linear regression on the data. (**Data | Transformations | Regressions, Load: Fitted Values**)
   1. Is the slope of the total energy close to zero? Are the energy fluctuations around this mean value random (qualitative inspection)?

4. Now make a histogram of the energy data. Use your favorite code or the same methods used in Lab 1C (**Data | Transformations | Histograms**; **Destination | Create New**; Select **G1** in the right hand **Graph** box and then select **G0** in the left hand **Set** box; **Start at:** [something less than your lowest value, e.g. -1.85e4 when I did it] | **Stop at:** [something greater than your largest value, e.g. -1.75e4] | **# of bins 100** | **Accept; Edit | Arrange Graphs** then make sure both **G0** and **G1** are highlighted (hold down shift while clicking each one) | Number of **columns = 1** | number of **rows = 2** | **Accept**; Click on the bottom plot to activate it, then click **Edit | Auto Scale Graph.**)

   2. Are the energies distributed in approximately a Boltzmann fashion?

(I.e., is the distribution pseudo-bell shaped?)
3. How could you obtain a distribution with less noise?

III. Backbone Conformational Fluctuation. It is difficult to quantitate conformational flexibility with a single metric. A common method to compare protein backbone dynamics is to compare the root mean square deviations (RMSD) of backbone or Cα (CA) atoms in Cartesian space.

Launch VMD with the *ionized.psf* and *prod.dcd* files. Calculate the Cα ("Trace", in VMDspeak) RMSD using a VMD analysis plugin,

Extensions | Analysis | RMSD Trajectory Tool
Options | Plotting Program | Xmgrace
Click Trace (Leave Top as the reference)
Click ALIGN
Click RMSD

5. What are the **average** and **S.D.** RMSD values?
Make a note of the average Cα RMSD. This value is the fluctuation *within* your trajectory it is not the RMSD difference of the simulated protein compared to the X-ray structure.

Leave the VMD session open and close the RMSD Trajectory Tool.

V. Overall Conformational Fluctuation. To obtain an overall qualitative sense of the protein conformational fluctuations during your simulation we will observe a movie of the protein flexibility during simulation. An all-atom representation can provide an idea of the sidechain flexibility and a Cα backbone representation can provide a visual comparison to the RMSD values calculated above. For the all atom movie enter the following in the VMD menus,

Graphics | Representations
Selected Atoms | **`protein`** (Enter return)
Drawing Method | VDW
Put the speed slider at about 2/3 maximum speed and click the movie arrow (▶).
Enlarge the graphics window and view the movie. Qualitatively estimate the approximate Å range of sidechain conformational flexibility.
6. Remembering that a carbon atom has a radius of ~2Å what would you estimate the range of side chain conformational fluctuations?
7. Is this more or less than the backbone CA RMSD you calculated above?
8. Do you observe a change in any side chain rotamers during this time range? 9. How many degrees of bond rotation is required for a rotamer change?
10. Would you *expect* to see a rotamer change in 5 ns?

Stop the movie by clicking the movie arrow (▶).

Now look at the CA trace RMSD. In the Graphics | Representations window,
Selected Atoms | **protein** (Enter return)
Drawing Method | Trace
Create Rep
Selected Atoms | protein and name CA
Drawing Method | VDW
Play the movie again.
11. Remembering that a carbon atom has a radius of ~2Å what would you estimate the range of backbone CA atom conformational fluctuations?
12. Is this more or less than the backbone CA RMSD you calculated above?

Stop the movie but leave the VMD session open on the Mac.

VI. Local Backbone Conformational Fluctuation. In this section, you will calculate the mean and S.D. of ϕ, ψ torsion angles along the protein backbone and compare these results to the trajectory movie.

First you will write out all the frames in your trajectory into a new directory. On the cluster in your *protsim/* directory check that you have the script *write_frames.tcl*. Make a new directory for the frames,

```
mkdir frames
```

VMD has the ability to load your trajectory and write out separate frames but you can't just launch VMD on the cluster because it can't display back on the Mac. But you can launch VMD in the "no-display" mode on the cluster and use its internal capabilities to run scripts on your trajectory. Inspect the code in *write_frames.tcl*. Then enter the following on the cluster in the *protsim/* directory where you just made the frames directory,

```
vmd -dispdev none -e write_frames.tcl
```

Quit the no-display session of VMD on the cluster. List the contents of the frames directory; you should see 1000 PDB files there. Copy the script *phi_psi_mean_SD_frames.py* to the *frames/* directory. Also, copy a *normal-python2* file there. Change to the *frames/* directory and edit the *normal-python2* script to have the following command,

```
$PYTHONBIN phi_psi_mean_SD_frames.py frame > phi_psi.stats
```

Then submit the job to the queue (**./normal-python2**). After a few minutes you should see a table of ϕ, ψ torsion angles and S.D. appear in a file called *phi_psi.stats*. Inspect the S.D. values and note that the peptide bond between residues 13-14 has large ψ and ϕ torsion angle S.D. values.

In the VMD session on the Mac where you have the trajectory displayed use the

Graphics | Representations window to change the selection to **`protein`**, drawing method to New Cartoon, create a new representation (Create Rep), select some residues in the segment with high ϕ, ψ torsion angle S.D. in the new representation,

`protein and (resid 13 to 15)` (enter return)

color them (ColorID | 1 red) and play the movie. 13. What kind of secondary structure (helix, strand, loop) are these two residues with **high** backbone mobility in?

Now find the six residue segment with the lowest ϕ, ψ torsion angle S.D in your *phi_psi.stats* file. Create a new rep in VMD, select these seven residues in the new rep; color them green and play the movie. 14. What kind of secondary structure (helix, strand, loop) are these seven consecutive residues with **low** backbone mobility in?

You should now have an intuitive understanding of the extent of side chain rotation that occurs in 5 ns; what a backbone fluctuation of approximately ±1 Å (RMSD) correlates to visually; and the fact that different secondary structural segments have different degrees of conformational dynamics.

In the next lab, you will investigate the average structure of the protein during your simulation and compare it to the crystal structure using probability density maps.

