Computer Lab 5 - Protein Simulation in Explicit Water

Part C: Probability Density Maps

I. Introduction. The last steps in this lab are to take multiple protein configurations from your production trajectory and make a probability density map from these configurations. The multiple configurations and the probability density map will be viewed on the Mac. These results will be compared to the X-ray crystal electron density map for this protein and the static structure that was fit to that map. **The point is to observe the conformational diversity that may be represented by probability density maps and how this observation can help one interpret experimentally determined X-ray crystal structures.**

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_lab5c.txt* if made with vi or *JHEDID_lab5c.docx* if made with MSWord).

II. Probability Density Map. You will use VMD to create a map (3D grid containing a numerical value at each grid point or vertex) based on the positions of atoms in your protein trajectory. We will ignore the water and salt atoms. The VMD plugin, **volmap**, creates a map of the weighted atomic density at each grid point. This is done by replacing each atom in the selection with a normalized Gaussian distribution of width (standard deviation) equal to its atomic radius. So the position of each atom is spread over several grid points based on a weighted distribution.

In your *protsim/* directory on the Mac fetch the file *bfactor.tcl* from the /home/*compbio2*/*Shared/* directory on the cluster. Then enter the following commands,

`vmd ionized.psf prod.dcd`
Graphics | Representations...
  Selected Atoms | **`protein and not hydrogen`**
  Drawing Method | Licorice
Extensions | Analysis | RMSD Trajectory Tool
  Selection Modifiers | Backbone
  Reference mol | Top
  Click ALIGN (This superpositions the protein in all frames to position in the first frame.)
Close RMSD Trajectory Tool
Extensions | Analysis | VolMap Tool
  Click compute for all frames... button
  Leave other options as default
  Click Create Map
  When done close VolMap Tool
Graphics | Representations...
  For the Isosurface Rep
  Draw | Wireframe
  Isovalue | **`1.0`** (Enter number or use the slider)
Play the movie (use slider to slow it down)

Notice that some of the sidechains spend some time outside the density blobs at this isovalue (contour level) and that some sidechains don't even have densities. One of the goals of this lab is to have you appreciate the conformational diversity that underlies a probability density map at a particular contour level so take a good look.

Stop the movie. With the cursor in the VMD graphics window enter "=" to reset the orientation and then zoom in to fill the window. Now ARG_1 should be at the top and LYS_15 at the bottom of the structure and neither should have any density for the sidechain. In the Graphical Representations widget use the isovalue slider to decrease the isovalue so that density for all conformations of ARG_1 and LYS_15 appears (or at least 90% of the sidechain conformations). 1. What is the probability density value (isovalue) necessary to encompass all conformers?

Now let's calculate B factors and create a PDB file of the average structure from the dynamics of your trajectory and view this information. Stop the movie and enter the following commands at the VMD prompt (you may have to hit return to get the vmd> prompt),

vmd> **source bfactor.tcl**
vmd> **bfactor "protein" 0 1000 aveB.pdb**
Wait until the calculation finishes. This reads all 1000 frames of the trajectory and calculates an average structure and writes out a PDB file with calculated B factors in the B factor column.
Now load the average structure into VMD.
vmd> **mol load pdb aveB.pdb**
In the Graphical Representations widget
Selected Molecule | 0: ionized.psf and double click the Licorice rep to turn it off
Selected Molecule | 0: ionized.psf, single click the Isosurface and set Isovalue to **1.5**
Selected Molecule | 1: aveB.pdb
Selected Atoms | **protein and not hydrogen**
Drawing Method | Licorice
Coloring Method | Beta
In the VMD Main window
Graphics | Colors...
Color Controls | Color Scale | Method | BGR | Offset **0.00** | Midpoint **0.10**

You should see the atoms with large B factors as green-to-red and the atoms with small B factors as blue. 2. Which atoms tend not to have defined densities at this isovalue, high or low B factor atoms? (Notice that the licorice structure is an *average structure* with atoms in their average x,y,z positions and therefore they may not follow normal stereochemistry. There may be strange bonds present.)

Set the isovalue to 1.0 for the Selected Molecule | 0: ionized.psf and leave the VMD session open.

III. Electron Density Map. When a crystallographer deposits a new structural model in the PDB s/he now has to also deposit the experimental data in the form of a file

containing the structure factors of the diffraction data. One can create an electron density map from these structure factors and you can download the map for a crystal structure from the PDBe (www.ebi.ac.uk/pdbe/). Go to this site, search for the 2PTC page, and click on the Downloads link on the right side menu list. Then download the EDS map to your *protsim/* directory on the Mac (It should be called *2ptc.ccp4*). First, we will separate the inhibitor (I) and enzyme (E) chains present in the *2PTC.pdb* file using the following commands in a separate window open to your *protsim/* directory,

```
grep '^ATOM' 2PTC.pdb | grep ' I ' > inh.pdb
grep '^ATOM' 2PTC.pdb | grep ' E ' > enz.pdb
```

(Note: The caret, ^, indicates that the grep command should select only those lines which have the word ATOM at the *beginning* of the line.)

Inspect the B-factors in the newly created file *inh.pdb*. The first thing you may notice is that the first two residues have occupancies and B-factors of zero. The reason for this is given in the header of *2PTC.pdb* the file as follows:

```
REMARK   4 AN OCCUPANCY OF 0.0 INDICATES THAT NO SIGNIFICANT ELECTRON   2PTC  86
REMARK   4 DENSITY WAS FOUND IN THE FINAL FOURIER MAP.                  2PTC  87
```

3. How can there be a residue in the PDB file if there was no electron density seen?

Let's look at the actual X-ray crystallographic electron density map of the inhibitor protein and the static model built into this map by the crystallographers.

```
pymol inh.pdb
load 2ptc.ccp4, map
isomesh msh,map,1.0,inh,carve=3.0
```

This will display the new map only around the inhibitor protein out to a distance of 3 Å.

Rotate the molecule so that the orientation is similar to that in your VMD session with ARG_1 at the top and LYS_15 at the bottom. (The command **reset** in PyMOL and **=** in VMD should show the default X-ray orientation.) Compare this map visually with your probability density map calculated from the trajectory of conformations?

Set the contour density higher or lower and try to make the electron density map in PyMOL look like your probability density map in VMD. For example,

msh | A | delete
isomesh msh,map,1.5,inh,carve=3.0

4. What is the difference between the two maps for LYS_15?
Hint: For PyMOL> **select inh and resid 15** and color the (sele) magenta.
For VMD:

In the Graphical Representations widget
Selected Molecule | 1: aveB.pdb
Selected Atoms | **protein and not hydrogen and resid 15**
Coloring Method | ColorID | 27

To understand the difference remember that the inhibitor protein you are viewing in PyMOL was part of a complex with the enzyme, trypsin. Load the enzyme now into PyMOL,

PyMOL> **load enz.pdb**

Note the position of LYS_15 relative to the enzyme molecule.
5. Write an explanation as to why LYS_15 is less mobile in the X-ray structure than in your simulation.

Congratulations, you have finished the most complicated lab of the course. You should be convinced that an electron density map from which a static model is built actually represents a very dynamic ensemble of structures. **In the future when you see an electron density map, or a static protein model, your mind should visualize the movie giving rise to that map or structure.**

[For lab 6 you will need a trajectory of the protein dihydrofolate reductase. Since today's lab is relatively short you should have time to get started making this trajectory. The instructions for starting your dihydrofolate reductase trajectory are reproduced below. It would be good to get started during class so that you can ask any questions that come up – the instructions below are cryptic on purpose. You will need this trajectory on November 16 so budget your time.]

MD trajectory of apo-dihydrofolate reductase. In its functional form dihydrofolate reductase (DHFR) contains two prosthetic groups, NADPH and folate. We will create trajectory of only the apoenzyme without the prosthetic groups.

The DHFR PDB file *1RA2.pdb* is available in the */home/compbio2/Shared/* directory on the cluster or you can download it from the PDB (www.rcsb.org). On a Mac, make a *network/* directory under your permanent home directory, change there and put the file *1RA2.pdb* there. To remove the prosthetic groups and waters you can grep out only the ATOM records to create a "clean" polypeptide PDB file. For example,

```
grep ^ATOM 1RA2.pdb > 1ra2_cln.pdb
```

This would be equivalent to the *inh.pdb* file in your previous simulation.

Copy over the *prot.pgn* and *top_all27_prot_lipid.inp* files from your *protsim/* directory to your *network/* directory. Edit the *prot.pgn* file appropriately (just change the file names

and delete references to disulfides) and make both the *start.psf* and *start.pdb* files as you did for lab 5.

Solvate, ionize and center the system as previously. Write down the minimum and maximum system coordinates so you can calculate the primary cell size for use in the **.conf* files.

On the cluster you should create a new directory called *network/* under your *compbio2/[JHEDID]* directory. Then gather the files needed to do an MD simulation exactly similar to the trypsin inhibitor simulation you previously carried out. You will need the **.conf*, **.inp* and *run.namd* files from that exercise. These should be edited for the appropriate file names, primary cell sizes, and PME grid sizes.

After the minimization and equilibration run, make a trajectory of 1000 frames as you did before. Fetch the final *prod.dcd* trajectory back to the Mac in the *network/* directory there.

Hint: After submitting your job to the queue watch the log file grow for about 5 minutes to make sure the job doesn't crash right away. This will give you time to make corrections and restart the simulation (and you will have a trajectory to analyze!). You can watch the log file grow with the following command,

```
tail –f min_equil.log
```

or

```
tail –f production.log
```

You can stop the **tail** session with ctrl-**c**.

