CHEM 498Q / CHEM 630Q: Molecular Modelling of Proteins
TUTORIAL #4b: Molecular dynamics: Analysis

INTRODUCTION

The purpose of this tutorial is to get familiarized with the trajectory analysis tools of VMD and the various issues related to thermal fluctuations in proteins.

Step 9 will be performed using the "RMSD Trajectory Tool". Steps 10 to 12 will be performed using the "Timeline", "Salt Bridges", and "Hydrogen Bonds" extensions of VMD.

PRE-LAB REPORT

None

PROCEDURE

STEP 7: Monitor the dimensions of the simulation cell

Retrieve all files from the simulation from STEP 6. You should have the following:

- `step5_production.log`
- `step5_production.coor  step5_production.vel  step5_production.xsc`
- `step5_production.dcd  step5_production.xst`

✦ Plot the dimensions of the simulation cell as a function of time. (This information is written in columns #2, #6, and #10 of file "`step5_production.xst`". The time step number is in column #1.) What is the average volume of the cell?

Note:
You can plot the data using `gnuplot`:

```
$ gnuplot
gnuplot> plot 'step5_production.xst' using 1:2, '' using 1:6, '' using 1:10
```

You can find the average dimensions of the simulation cell by fitting constant functions "A", "B", and "C" to the data for $t > 10.0$ ps:

```
gnuplot> fit [5000:] A 'step5_production.xst' using 1:2 via A
gnuplot> fit [5000:] B 'step5_production.xst' using 1:6 via B
gnuplot> fit [5000:] C 'step5_production.xst' using 1:10 via C
```

To produce an image file of the graph, type the following additional commands:

```
gnuplot> set terminal png
gnuplot> set output 'plot_ABC.png'
gnuplot> replot
```

STEP 8: Visualize the MD trajectory

Load your simulation system into VMD:

```
$ vmd ../step3_pbcsetup.xplor.ext.psf step5_production.dcd
```

Use an orthographic view of the coordinates ("Display > Orthographic" menu). Create four separate representations ("Graphics > Representations…" menu):

- Selected Atoms = "water" and Drawing Method = "Lines"
- Selected Atoms = "resname POT CLA ZN2 CAL MAG" and Drawing Method = "VDW"
- Selected Atoms = "protein" and Drawing Method = "Lines"
- Selected Atoms = "not water and not protein" and Drawing Method = "Licorice"

If the protein is not fully inside the primary water box, draw the periodic images as well (otherwise you will not see all contacts).

STEP 9: Calculate the RMSD of the alpha-carbons

Use "Extensions > Analysis > RMSD Trajectory Tool" to create a plot of the root-mean-square deviation (RMSD) of the protein. Proceed as follows:

- Make sure the selection is "protein" and check "Trace" (which means that only the alpha-carbons will be considered in the calculation).
- Align all frames on frame "0" of your trajectory. (Just click on the "Align" button.)
- Check "Plot" and click on the "RMSD" button. This will produce a graph of the RMSD as a function of the frame number. (To save the data to a text file, check the box "Save to file" and perform the RMSD calculation again.)

✦ Include the graph of the alpha-carbon RMSD as a function of time to your report. Has the RMSD value stabilized after 1 ns?

STEP 10: Visualize the fluctuations in the secondary structure

Open the "Extensions > Analysis > Timeline" window and select the menu "Calculate > Calc. Sec. Struct.". This will produce a "timeline" showing the secondary structure adopted by each amino acid of your sequence, color-coded as following:

- Cyan: Turn ("T")
- Yellow: Extended conformation ("E")
- Brown: Isolated bridge ("B")
- Magenta: Alpha helix ("H")
- Blue: 3_{10} helix ("G")
- Red: Pi helix ("I")
- White: Random coil ("C")

Note:
The secondary structure is computed using the STRIDE algorithm (http://webclu.bio.wzw.tum.de/stride).

✦ Save the data as a PostScript file (menu "File > Print to file…") and put in appendix to your report. Report any change in the alpha helix or beta strand structures. Are any of them breaking apart of forming? Are any of them getting longer of shorter?

Note:
The size of the PostScript file may be quite large, so please send it to the instructor in a compressed form:

```
$ gzip timeline.ps
```

STEP 11: Analyze the stability of salt bridges

From the “Timeline” window, select menu “Appearance > Set scaling…” and use Bottom value = 0 and Top value = 1.

Calculate the time evolution of all salt bridges (Asp^- or Glu^- interacting with Arg^+ or Lys^+) formed during the simulation by selecting menu “Calculate > Calc. Salt Bridges”. You can view each one individually by clicking on the button “every residue”.

Note:
The same pair of amino acids may be found multiple times in the timeline data. Each one corresponds to a specific oxygen atom (OD1 or OD2 for Asp, OE1 or OE2 for Glu) binding to a specific nitrogen atom (NE, NH1, or NH2 for Arg).

- ✦ Use “Extensions > Analysis > Salt Bridges” (from the “Main” window) and write a file with the distances for each salt bridge. This will create a set of files “`saltbr-*-*.dat`”, containing the shortest of all O-N distances as a function of the frame number. Plot all data using gnuplot:

```
gnuplot> plot 'saltbr-ASP93-ARG58.dat', 'saltbr-ASP22-LYS98.dat', ...
```

- ✦ Identify the pairs of residues involved in the most stable salt bridges and see if they have a high score in the PSSM of the conserved domain corresponding to your sequence. (This will require you to go back to Tutorial #1a.) Based on your literature search, are any of these residues known to be functionally important?
- ✦ Your report should describe how a salt bridge is mathematically defined in VMD.

STEP 12: Analyze the stability of hydrogen bonds

From the “Timeline” window, use “Calculate > Calc. H-bonds” to get an overview of the stability of hydrogen bonds in the protein. This will provide a picture of all hydrogen bonds present in the protein (backbone–backbone, backbone–sidechain, and sidechain–sidechain).

Use “Extensions > Analysis > Hydrogen Bonds” (from the “Main” window) and set the following input options:

- Selection 1: “protein and not backbone”
- Calculate detailed info for: Residue pairs
- Ouput options: Write output to file.

- ✦ Examine the content of file “hbonds-details.dat” and identify the amino acid pairs involved in the most stable hydrogen bonds (those with the highest occupancy). Classify them according to the charge of the side chains (charged–charged, charged–neutral, and neutral–neutral) and compare their relative occupancies. Do not report the salt bridges found in STEP 11.
- ✦ Do any of these pairs have a high score in the PSSM of the conserved domain? Based on your literature search, are any of these residues known to be functionally important?
- ✦ Your report should describe how a hydrogen bond is mathematically defined in VMD.

STEP 13: Examine the metal coordination

If your protein binds zinc atoms, create a new representation with Selected Atoms = “same residue as within 2.49 of resname ZN2” and Drawing Method = “Licorice”. To ensure that any ligand entering

the coordination sphere of the metal atom is displayed, check the box “Update Selection Every Frame” from the “Trajectory” tab.

- ✦ Examine and report any change in the coordination structure of the zinc atoms. Are new protein ligands or water molecules coming into the coordination sphere? Is the coordination of Asp or Glu ligands changing from monodentate to bidentate (or vice versa)? For each ion, present a figure of the coordinating geometry (ion + ligand) from the last frame of the simulation.
- ✦ Repeat the same analysis for calcium ions (“resname CAL”) using a cutoff distance of 2.70 Å, and for magnesium ions (“resname MAG”) using a cutoff distance of 2.95 Å.

Note:
Zinc is usually either tetra- or penta-coordinated. Any coordination number larger than that is most probably an artifact of the force field. For calcium and magnesium, the maximum coordination number is usually 8.

STEP 14: Examine the active site

Since the protein is simulated without its substrate (and without an inhibitor), we expect the active site to change its shape (to get wider or narrower).

- ✦ Discuss the presence of water molecules in the site and their interaction with catalytically important residues.
- ✦ Report any motion that may obstruct the active site of the protein, such as protein loops or amino acid side chains moving into the pocket. Report these by measuring the distance between various residues across the active site at different points in time during the simulation.

Note:
Once you have labelled a distance between two atoms, you can monitor it over time by using the “Graphics > Labels > Bonds” menu.

- ✦ Include in your report a figure of the protein at a point in time at which its active site is the widest open. To provide a better view of the pocket, change the drawing method of the protein to “Surf”. Report the MD frame it corresponds to and save the coordinates of that frame into a PDB file (using the “File > Save Coordinates...” menu). This PDB file will be used in Tutorial #5.