Introduction

In the practical, you learned how to set up a solvated system containing a simple protein, MD-2, a lipid recognition protein that forms a complex with Toll-like receptor 4 (TLR) involved in activating pathways in the innate immune system (http://www.rcsb.org/pdb/101/motm.do?momID=143). You then prepared the system for a molecular dynamics (MD) simulation. Now you are supplied with the resultant 100 ns simulation trajectory, and you must utilize VMD to analyze it (**Part A**). You will also do the same for another 100 ns simulation trajectory in the presence of a bound ligand (**Part B**).

The files are:

- For the unbound MD-2 system, the pdb file and corresponding trajectory file: '**apo_md.pdb**' and '**apo_md.xtc**'

- For the ligand-bound MD-2 system, again the pdb and trajectory: '**bound_md.pdb**' and '**bound_md.xtc**'

VMD may be downloaded here (and runs on Linux, Mac, or Windows):
http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD

You have already started to use VMD to do some simple displaying of molecular graphics. The online tutorial (particularly the 'Working with a Single Molecule' and 'Trajectories and Movie Making' sections) will be helpful:
http://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/

Note 1: If you find that VMD is too slow, particularly when trying to display the trajectory of the whole system, try reducing the value for Resolution below each Rep in the Graphical Representations window, and/or only displaying the many water molecules in a simple representation, such as Lines.

Note 2: In general when viewing trajectories, you may find it easier to switch from Perspective to Orthographic mode (do this from the Main Menu: Display->Orthographic), and possibly to remove the Depth Cueing (Display->Depth Cueing).

Part A: Analysis of the Apo (Unbound) Protein System

(i) The first analysis we usually do is simply visual inspection, viewing the trajectory as a 'movie' in VMD ('**vmd apo_md.pdb apo_md.xtc**'). Display the starting structure of the protein on its own, highlighting the secondary structure elements. (You may render the image with File->Render from the main menu). How would you describe its structure?

(ii) Now scan through the trajectory. The protein's behaviour will be hard to see as it is free to diffuse in the unit cell – remember, it is no longer position-restrained! Fortunately, you can use the **'RMSD Trajectory Tool'** (**Extensions→Analysis→RMSD Trajectory Tool**) - in the RMSD Trajectory Tool window, make sure Trajectory Frame ref = 0, and the top-left box says **'protein'**. Click the **Align** button and this should align all frames to the starting structure (which is very close to the X-ray structure). Now scan through the trajectory again and things should be a lot easier to see!

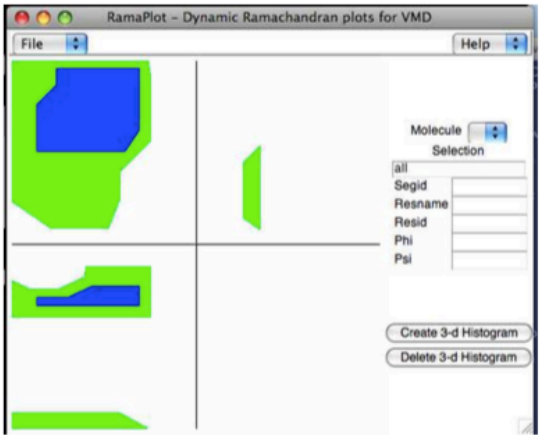
(iii) In the **Graphical Representations** window, create a second Rep of the protein, but for this one, click the **Trajectory** tab and change **'now'** to **0** so that the starting structure is shown. By suitably representing this state, you should be able to see any conformational changes in the protein visually, relative to the starting structure. Align the protein so that you're 'looking into' the binding cavity at the start of the simulation - do you see any changes as you scan through the trajectory? How would you describe them?

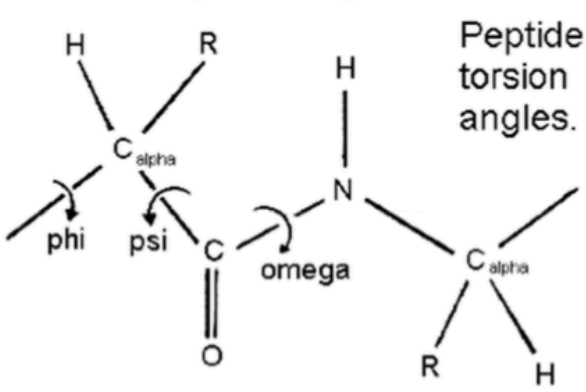
(iv) After aligning to the protein, tick the **Plot** button, and then click the **RMSD** button to calculate the root-mean-square deviation of the protein, relative to the initial structure. Tick the '**Save to file**' button and click **RMSD** to save the data (which you may then plot using e.g. R or Excel). For more details of this tool, see http://www.ks.uiuc.edu/Research/vmd/plugins/rmsdtt/

The RMSD of the protein is very large as it incorporates the motions of the amino acid side chains. To remove this effect, try calculating the RMSD of the protein for its peptide backbone atoms only (i.e. for '**protein and name N CA C**'). You can also select certain regions or structural features of the protein to calculate the RMSD for, to give insight into specific conformational changes. For example, you could type '**protein and name N CA C and (sheet or helix)**' into the top-left box of the **RMSD Trajectory Tool** window to calculate the RMSD for all regions of ordered secondary structure.

Use similar approaches to calculate the RMSD for the beta-sheet backbone atoms of the protein only, and similar for the alpha-helix backbone atoms. Overlay all of these onto a single graph, and also ns of the simulation. How does this help to rationalize which parts of the protein exhibit the most and the least structural drift?

(v) You can check the structural integrity of the protein over the course of the simulation by calculating Ramachandran plots, which display the psi and phi backbone torsional angles for each residue in the protein. Only certain values for these angles are 'allowed'. To calculate this, from the **Main Menu** go to **Extensions→Analysis→Ramachandran Plot**. Select **'Molecule 0'** – which is the system you're looking at – and by scanning through the trajectory, you can see values for each residue. You could select individual residues if you wish. You can also create a histogram of the distribution – click on the **'Create 3-d Histogram'** button. When you're done, click **'Delete 3-d Histogram'**.





(vi) You can assess how stable the secondary structure of the protein is using the **Timeline** tool (selected from the **Main Menu** with **Analysis→Timeline**). You can get more information on the **Timeline** plugin here: http://www.ks.uiuc.edu/Research/vmd/plugins/timeline/

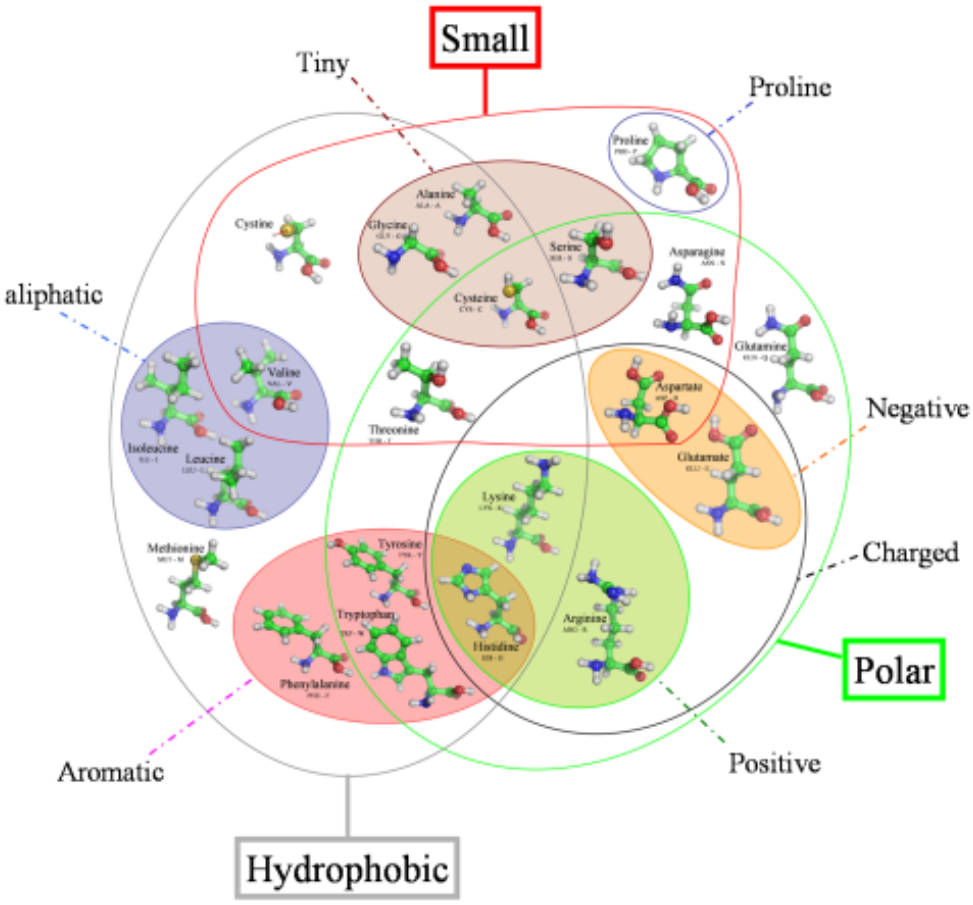
In the **Timeline** window, click '**Calculate→Calc. Sec. Struct.**' You'll have to wait for the tool to run through each frame of the trajectory, and a little time at the end for it to 'think'. When it's done, you should see a color-coded 'map' - the Y-axis represents each protein residue, and the X-axis represents the frame number. The "sliders" on the left of the window allow you to play around with the scale etc. Compare the colors you see on the map to those you see in the graphical representation of your protein (**colored by Structure**) to decipher the map. In particular, you should be able to distinguish beta-sheet, alpha-helix, and 'other' structural states (such as loops and turns). Is the secondary structure stable over the course of the simulation? Are there specific regions that change? If so, why do you think this is? Use '**File→Print to File**' to save the plot in postscript format.

(vii) If the protein undergoes changes in secondary structure, VMD does not by default update this in the graphical representation. If you wish to observe any changes graphically, copy and paste the content of this script into the vmd text window: http://www.ks.uiuc.edu/Research/vmd/script_library/scripts/sscache/sscache.tcl and then press **enter**, type **'start_sscache'** and press enter again.

You should now see these changes as you scan through the trajectory. If you observe such changes, render a suitably displayed image.

(viii) You can characterize distances between regions of the protein that change. Display the C-alpha atoms of the protein beta-sheet ('**name CA and beta**') with VDW representation, and label '**bonds**' (not chemical ones necessarily!) between pairs of interesting C-alpha atoms. To do this, from the **Main Menu** choose the **Mouse→Label→Bonds** menu item, and then click on pairs of C-alpha atoms. When you scan through the trajectory, the distance between these pairs of atoms will be displayed and will update at each frame. To display graphs of any interesting distances, choose **Graphics→Labels** from the **Main Menu**, and in the **Labels** window, select **Bonds**, switch to the **Graph** tab and select different pairs of bonds. Press the **Graph** button to view it, and press **Save** to save the data in order to plot it. Do you see any interesting changes in distances?

(ix) Why do you think you observe any conformational changes in this protein? To answer this, you should work out what the characteristic distribution of amino acids in this protein is. For example, you could create a Rep corresponding to hydrophobic-aliphatic groups by typing '**protein and resname ILE VAL LEU**' in the **Graphical Representations** window. ***For a reminder of the amino acid properties, see the figure below.*** Referring to this, make additional Reps for other groups of amino acid, and display them in different colours.



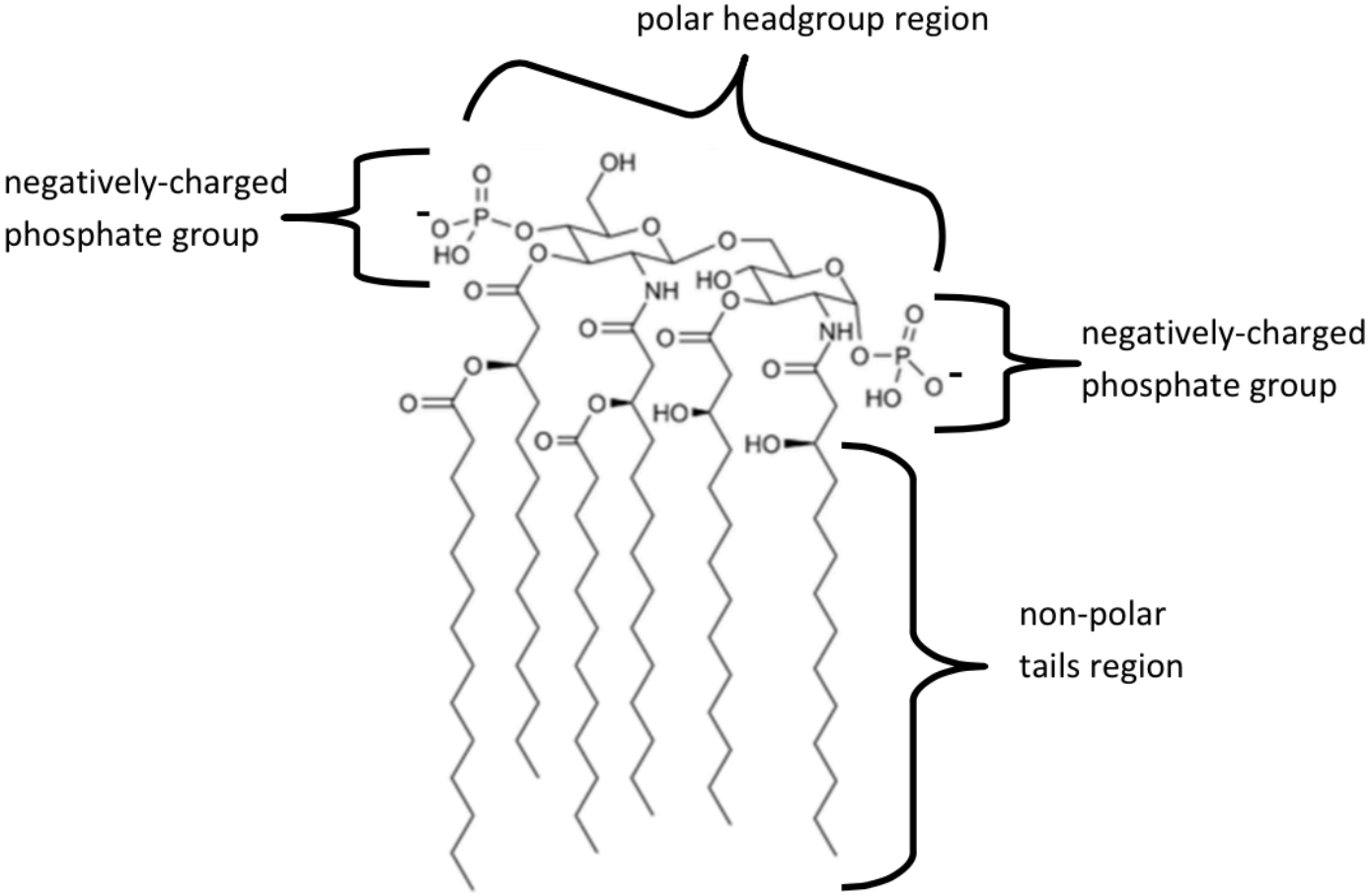
(x) You may also try visualizing the water molecules in the system. Try creating a **Rep** in the **Graphical Representations** window called '**water**', drawn with **Lines** and colored by **name** (you may have to reduce the Resolution.) You should see all waters in the system (shaped in the periodic box used for the simulation, namely a truncated octahedron) - there are a lot! To make things easier to see, create a Rep in which only waters within a certain distance of the protein are displayed.

For example, you could create a Rep called '**water and same residue as within ??? of protein**' (where ??? is a distance chosen by you in Angstroms), and then click '**Update Selection Every Frame**' in the **Trajectory** tab in the **Graphical Representations** window. This would show all water residues within a certain number of Angstroms of the protein, and VMD will update the selection every time you change the frame of the simulation. Try to find a suitable distance cutoff. Things may still be hard to see, so you *could* also try changing the selection to view water within a certain distance of the different amino acid distributions of the protein. How do waters in the vicinity of the protein behave over the course of the simulation?

(xi) Experimentally, it has been shown that residue 126 (a Phenylalanine) is critical to the function of MD-2 in signalling to TLR4 in the active macromolecular complex. Use the Graphical Representations window to display this residue. Does it remain in the same orientation over the entire simulation, compared to the initial structure?

Part B: Analysis of the Ligand-Bound Protein System

A simulation trajectory has also been provided for MD-2 in complex with lipid A. Lipid A is the bioactive component of bacterial lipopolysaccharide (LPS), which means its structure is sufficient to bind to and active the TLR4/MD-2 pathway – see: http://www.rcsb.org/pdb/101/motm.do?momID=143 and references therein for further details. Lipid A contains a number of hydrophobic (non-polar) fatty acid "tails", capped by a polar sugar-containing 'headgroup', including two attached, negatively-charged phosphate groups.



(i) Load the ligand-bound trajectory into VMD, and observe the behaviour of the lipid A molecule in its binding site. Use approaches similar to those used when analyzing the apo state to follow the stability of the protein, and then also create suitably-displayed Reps for the lipid molecule ('**resname LPS**'). Do the hydrophobic lipid tails of the ligand remain stable, relative to their starting positions? What are the main interactions they make with the protein? i.e. what kind of amino acid residues do they seem to interact with, and why?

(ii) Do the other parts (i.e. polar headgroup region) of the ligand also remain close to their starting positions? If they are more - or less - mobile than the lipid tails, why do you think this is? To answer this, try to characterize the interactions between amino acid sidechains and the headgroup. In particular, do you see any interactions with the negatively-charged phosphate groups? Remember, electrostatic interactions can be energetically very strong. To find amino acids close to the phosphate groups, you could create Reps for residues within a certain distance, e.g. **'protein and same residue as within ??? of name P1 P2'** (and don't forget to click on '**Update Selection Every Frame'** in the **Trajectory** tab for this Rep). Do you see any charged amino acids forming stable interactions with the lipid A phosphate groups?

(iii) You would normally expect oppositely-charged (i.e. positive) amino acids (such as Lysine or Arginine) to interact with the negatively-charged phosphates. Is this all you see, or do you see nearby negative amino acids too? If so, it is possible that charged ions in the system may bridge two like-charged groups. Create a Rep which allows you to test this. What do you see?

(iv) Perform the same kind of protein analysis relevant for this system as for the unbound MD-2 simulation. How do the protein structure and dynamics compare over the course of the trajectory? Explain this in the context of the stability of the lipid within the binding cavity.

(v) In the X-ray structure of the entire lipid A bound complex [pdb=3FXI] Phe126 forms a key interaction with the activated MD-2/TLR4 dimer that stabilizes the signalling-competent complex. Whilst lipid A is an agonist of (i.e. activates) the TLR4/MD-2 complex, activation of the complex becomes less pronounced as the number of acyl tails of the bound lipid is reduced (or in the complete absence of lipid). Does a comparison of the protein dynamics (and Phe126) in the unbound and bound simulations help to provide a 'molecular clue' for the mechanism of TLR4 regulation?