CS444(001): Homework3

George Mason University
Instructor: Amarda Shehu

Out: October 16, 2012
Due: November 13, 2012

Note: You may need to refer to the VMD user's guide for this homework. This guide can be found at *http://web.mit.edu/vmd_v1.8/distrib/vmd/doc/ug.pdf*. You can work in 2-person teams on this homework.

Problem 1. Visualizing an Ensemble of Conformations (20 points)

The 1D3Z PDB id corresponds to the NMR ensemble of the ubiquitin protein. Download this file from the PDB. You can now visualize the 10 conformations in this file with VMD.

One way to do so after you upload the 1D3Z.pdb file into VMD is to draw the first conformation of the file in opaque and superimpose the rest in transparent. The graphics/representation menu in VMD allows you to choose whether you want to draw different frames in the trajectory (corresponding to the different conformations) with different representations, colors, materials. The 'Draw Multiple Frames' field under the 'Trajectory' submenu allows you to draw either one conformation or multiple in the form "b:e", where b corresponds to the beginning conformation and e to the end conformation.

For instance, to view only the first conformation, you can write 0:0. Choose the 'New Cartoon' Drawing Method for this conformation, and the 'Structure' Coloring Method. Note that the default value for the 'Material' is opaque. To view other conformations on the pdb file simultaneously, you need to create a new representation first, by clicking on the 'Create Rep' tab. To view all the other 9 conformations remaining in the PDB file, you can write 1 : 9 in the 'Draw Multiple Frames' field. Choose the New Cartoon' Drawing Method for these conformations, and the 'Structure' Coloring Method. Change the 'Material' from opaque to transparent.

You should now be able to see all conformations in the 1D3Z.pdb file, the first one in opaque, and the rest superimposed in transparent. Render the scene in an image file, as you did on hw 1. Remember to change the background of the drawing window from black to white so that you do not waste ink when printing the image file. Submit a hard-copy of the image file you have created.

Problem 2. Diversity of Conformations in an Ensemble (20 points)

The ubiquitin conformations you superimposed on one another have some source of variability. One way to quantify the geometric difference or similarity between two conformations is through lRMSD, which stands for least root mean squared deviation. lRMSD is a measure of the average of the distance each atom would have to move to convert one conformation to the other. Writing code to implement the lRMSD measure can be done, but the first step requires aligning all conformations in the trajectory with respect to some reference conformation. This reference conformation can be the first one in the trajectory or ensemble. The purpose of the alignment is to remove any source of variability that comes from an overall translation or rotation of a conformation.

For this part of the assignment, you will use a script that already computes the lRMSD of each conformation in a trajectory from the first conformation. The script is written in Tcl, and you can download it from the syllabus, alongside with the homework. Note that the script already sets the mol variable to 0. If you load and delete the 1D3Z many times, the mol id may not be 0. You have to change that properly. In addition, the script will write the lRMSD values to the screen. You can execute the script by copying it on the TKConsole that appears under the 'Extensions' menu. You can copy the lRMSD values that the script outputs on the TkConsole to a file of your choice.

(a) Submit a hard-copy of this file.

(b) Which conformation has the largest lRMSD from the reference conformation (which the script sets to be the first conformation in the trajectory - frame 0)?

(c) Plot the lRMSD values that you obtain on the Y-axis and the conformation id (starting from 0 to 9 - which is the id of the last conformation in the 1D3Z trajectory) on the X-axis. You can use Excel, Gnuplot, SM, Matlab, or any other plotting software. Please submit this plot as well.

Problem 3. Ranking Conformations in an Ensemble (30 points)

Protein conformations are not only geometric objects but are characterized by energy, as presented in class. You will implement a very simple energy function that only penalizes collisions between atoms. That is, your energy function will consider only unfavorable interactions due to collisions between atoms, also referred to as steric clashes.

This function should report high energies for conformations with collisions and low energies (the lowest value being 0) for collision-free conformations. You can model each atom as a sphere with a certain radius known as the van der Waals (vdw) radius. Even though different atoms have different vdw radii, you can assume that all atoms have the same radius of 1.7 Å. One way to make sure that there are no pairwise collisions is through the following:

$$(C_i - C_j)^2 > (2 \cdot R)^2$$

That is, for each atomic pair i, j in a conformation, the distance between their centers C_i and C_j needs to be more than twice their radii. Recall that you can set R to 1.7 Å. The center coordinates you will need to read from the atom coordinates associated with a conformation. Your energy function should use this proper atomic separation to associate a positive value for atom pairs whose centers are closer than twice the radii.

(a) Submit the code of your energy function.

Now apply it to rank the conformations in the 1D3Z.pdb file. To be able to do so, you need to output the coordinates associated with each conformation in the 1D3Z.pdb file. You can spit out these coordinates with vmd if you prefer not to write your own code to parse a PDB file. Under 'File/Save Coordinates' you can now select the *all* default representation under 'Selected Atoms' and type 0 under the 'Last:' box to save the atomic coordinates of the first conformation in the 1D3Z ensemble.

You need to specify a file name after you click on the 'Save' tab where VMD will write out the main chain coordinates of the first conformation. Select the crd type for the output file so that only raw atomic coordinates are written out. To write out the atomic coordinates for each conformation in the 1D3Z file, you can put values 0 through 9 in the 'Last:' box and separately write out the conformations in crd files, which you can name 'conformation0.crd', conformation1.crd', and so on down to 'conformation9.crd'

(b) Plot the energy values that you obtain on the Y-axis versus the conformation id on the X-axis. Again, you can use any graph plotting software. Please submit a hard-copy of this graph.

(c) Plot the energy values on the Y-axis versus the lRMSD values of each conformation on the X-axis (you obtained lRMSD values on the problem above). Please submit a hard-copy of this graph.

Problem 4. Manipulating a Conformation - Bond Rotation (30 points)

In this problem you will only focus on the main chain of a protein conformation. The first step will be to extract the x,y,z coordinates of the main-chain atoms of the first conformation in the 1D3Z NMR ensemble. You can either write your own script to strip the coordinates of the N, CA, C atoms that constitute the main chain, or use VMD to do so. The 'Selected Atoms' box under Graphics/Representation in VMD allows to display only a selection of atoms on the screen. You can narrow the selection from the 'all' default to one that specifies the main chain. One way to do so is to type *backbone and not oxygen* in the 'Selected Atoms' text box.

You should now see on the vmd screen the chain that only connects the N to the CA to the C of each of the 76 amino acids of ubiquitin.

(a) Render this image and submit as part of the homework. Do not change the Drawing Method under the Draw Style menu from the default 'Lines.'

Under 'File/Save Coordinates' you can now select the *backbone and not oxygen* representation under 'Selected Atoms' and type 0 under the 'Last:' box to only save the main chain coordinates of the first conformation in the 1D3Z ensemble. You need to specify a file name after you click on the 'Save' tab where VMD will write out the main chain coordinates of the first conformation. Remember to choose the pdb type for the output file so that you can then upload this file to properly define and carry out the transformations asked in the next part.

(b) Submit a hard-copy of the output file that VMD has written out (or that you have extracted yourself).

The final part of this exercise now involves modifying the main chain coordinates of the first conformation that you have written out in (b). Your modification will be a rotation by 30 degrees of the dihedral bond that connects the N and CA atoms of the previous to last amino acid in the outputted conformation. Rotation of this bond should affect only the location in space of the atoms following the N-CA bond of the previous to last amino acid.

You can use any software to define your own rotation matrix about a bond. You can even use Matlab. However, there is a also a simple way to obtain and apply a rotation matrix in VMD.

First start a new session of vmd and upload the pdb file you outputted above. First you have to make a selection that shows vmd which is the left index of the bond and which is the right index. Assuming that you are rotating the bond between atoms at atom indices x and y:
set bsel1 [atomselect 0 "index x" frame 0]
set bsel2 [atomselect 0 "index y" frame 0]

Then, if you want to rotate by 30 degrees around the bond defined with the two above statements, you have to define the transformation matrix:
set mat [trans bond [lindex [$bsel1 get {x y z}] 0] [lindex [$bsel2 get {x y z}] 0] 30 deg]

(c) Submit the rotation matrix that you get by defining it properly as the rotation by 30 degrees around the bond that connects the N and CA atoms of the previous to last amino acid.

(d) You now need to change the location in space of the atoms that follow the N-CA bond of the previous to last amino acid. You need to apply the transformation matrix you obtain in step (c) to the atoms that will be affected by the rotation.

Again, you can use your own code to do so or use the functionality that VMD provides. You first need to define a selection that has the atoms you want to rotate. If you need to define a selection with atoms in some residue x, you can use the following:
set sel [atomselect 0 "resid x" frame 0]

If you need to add atoms from different residues, you can use a different selection (vmd manual has more details on how to define selections). Here is a hint that, if you want to combine a carbon atom of residue x with all atoms of a residue y, you can use the following selection:
set sel [atomselect 0 "(name C and resid x) or resid y" frame 0]

Alternatively, you can define a selection that has all the atoms following the N-CA bond of the previous to last residue, by including all atoms with indices larger than the index of the CA atom in the N-CA bond. For instance, if this CA has index x, you can define a selection:
set sel [atomselect 0 "index > x" frame 0]

Whichever way you use to define a selection, to transform it, you need to do:
$sel move $mat.

Submit the coordinates of the atoms affected by the rotation matrix before and after the rotation (so I can see that the rotation was carried out).

(e) Finally, prepare an image (with VMD) that clearly shows that the location of the C atom has changed in space after the application of the transformation matrix. You can superimpose the new coordinates over the old coordinates of the C atom and show that the rest of the amino acid to which C belongs has not changed in space. Submit a hard-copy of this image.