Exploring DNA structure and DNA-protein binding

This laboratory will be held on Monday, October 7, in Wilcox 073. There are 15 workstations in Wilcox. All students should log in with:

```
username – cheme355
password – giagyu
```

The tutorial introduces the molecular modeling software VMD through simple tasks about base pair interaction and DNA-protein binding. The first part is a VMD tutorial, and the second part is the actual assignment. There are five questions in the assignment.

Part 1: VMD Tutorial

This lab should allow you to become familiar with the Protein Data Bank, visualizing structures and comparing them first visually and then mathematically.

1. Go to http://www.rcsb.org
2. Do a search (keyword) for "myoglobin"
3. Download a PDB file for horse hemoglobin (PDB identifier 1AZI) and save it with no compression
4. Launch VMD by clicking on the icon on the desktop
5. VMD opens a graphical window and a command prompt window. Here is what appears in both:
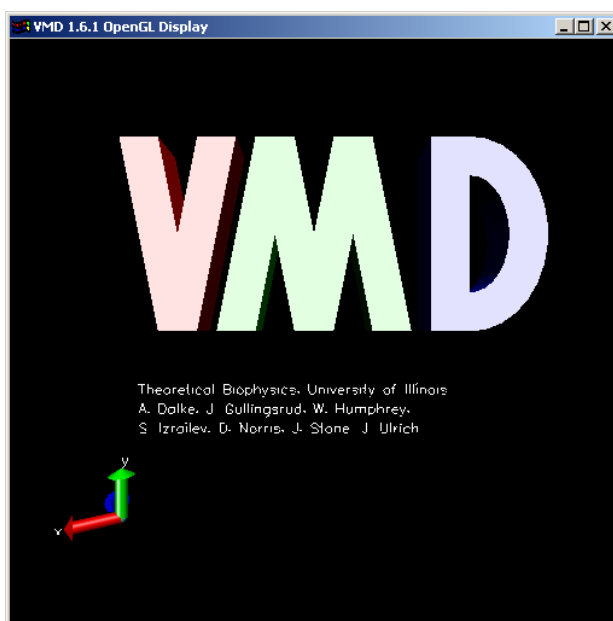
```
Info) vmd for WIN32, version 1.6.1 (April 20, 2001)
Info) VMD WWW Page:  http://www.ks.uiuc.edu/Research/vmd/
Info) Please email all comments and bug reports to vmd@ks.uiuc.edu
Info) The authors request that any published work which utilizes
Info) VMD please include the following reference:
Info)    Humphrey, W., Dalke, A. and Schulten, K., `VMD - Visual
Info)    Molecular Dynamics', J. Molec. Graphics 1996, 14.1, 33-38.
Info) -------------------------------------------------------------
```
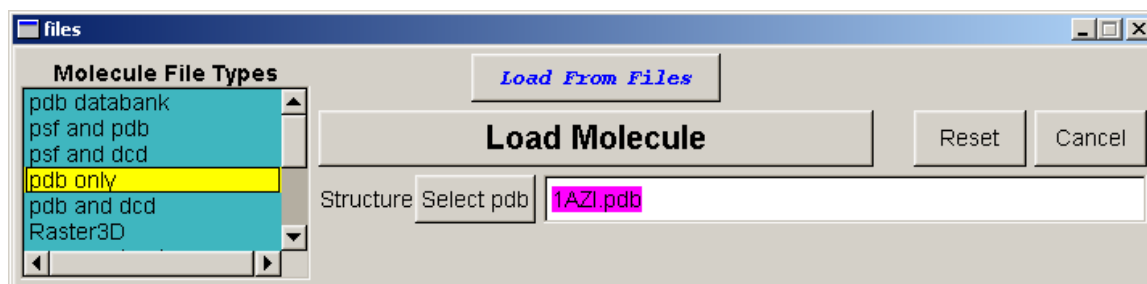
6. Load the molecule into VMD: to do that, select `Molecule` from the Main Menu and `Load from Files`, and select PDB only and the PDB file. Then, select `Load Molecule`. The red dots which are not connected to the molecule are oxygen from water: to not see those, select `Graphics` from the main menu, and under `Selected atoms`, type `protein`, or `backbone`, to visualize different subsets of the atoms in the molecule (there is also a name list one can use). Note that you can animate the molecule by dragging the mouse, or by pressing the left button and then dragging the mouse out of the screen, you can get the molecule to spin continuously.
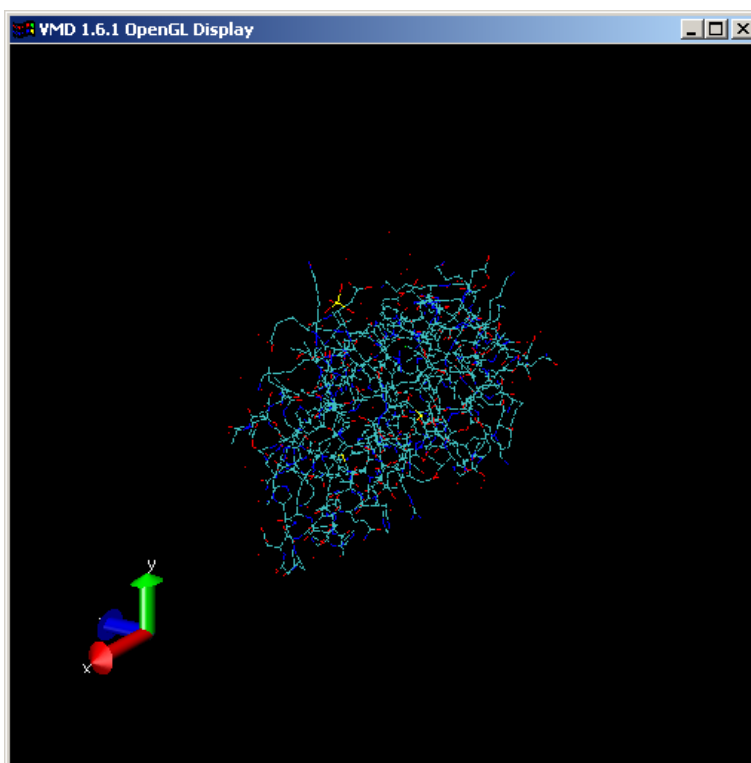


```
vmd > Info) 1AZI.pdb      2          6
Info)    Determining bond structure from distance search ...
Info) Loading new molecule ...
Info) Analyzing structure ...
Info)    Atoms: 1304
Info)    Residues: 211
Info)    Waters: 54
Info)    Segments: 1
Info)    Fragments: 59    Protein: 1    Nucleic: 0
```
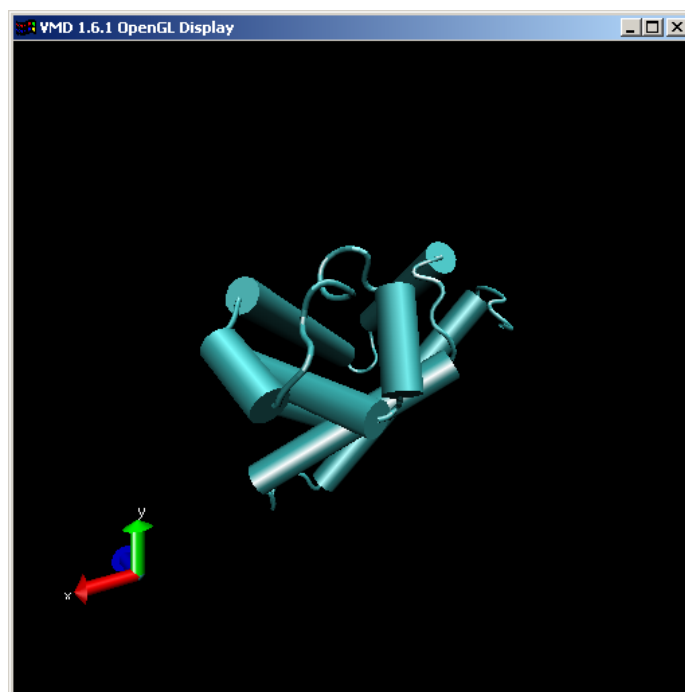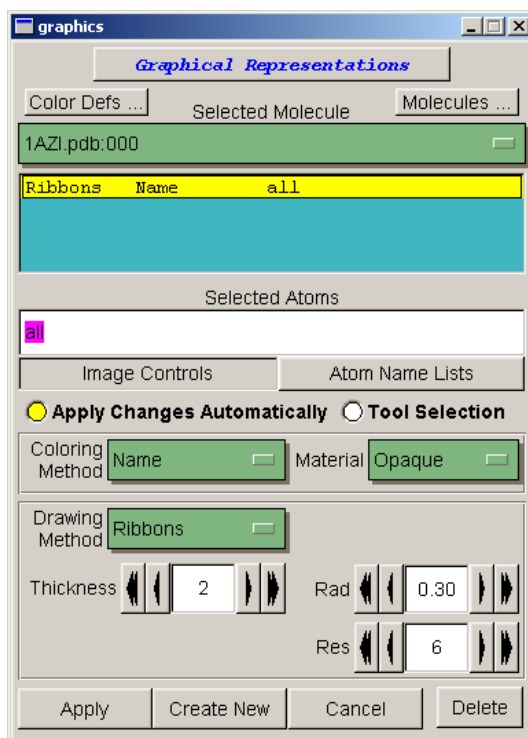
7. To visualize the molecule differently, use `Graphics` and select `Ribbon` under `Drawing Method`. `Cartoon` is another useful way to visualize the atoms and the bonds. Now, repeat the PDB search with another myoglobin, e.g. sperm whale (PDB identifier `11OM`) and again save the file in PDB format (no compression). Load the molecule into VMD:

```
Info) 110M.pdb  2        6
Info)     Determining bond structure from distance search ...
Info) Loading new molecule ...
Info) Analyzing structure ...
Info)     Atoms: 1455
Info)     Residues: 337
Info)     Waters: 179
Info)     Segments: 1
Info)     Fragments: 184   Protein: 1   Nucleic: 0
```
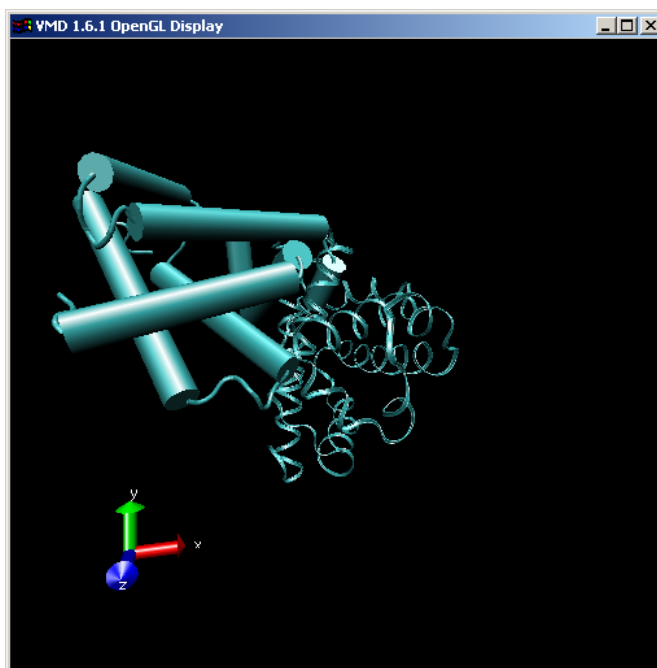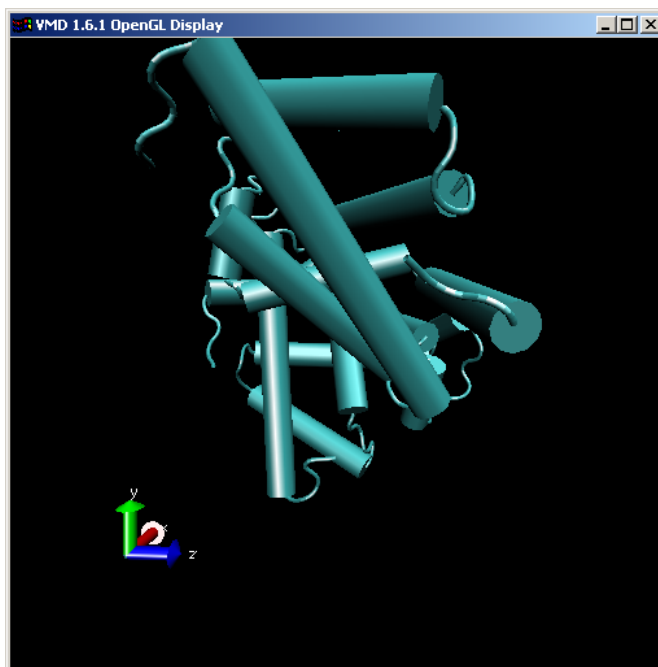
8. Plot both the molecules using `Cartoon` as a graphics choice. To color them differently, use `Color` from the main menu and then under `Category` choose `Molecule` (in the `Graphics` main menu, you need also to specify that you are coloring by `Molecule`). To find if the two molecules match by structure, i.e. to calculate the distance between the two molecules, we type:


```
vmd > set sel1 [atomselect 0 "backbone"]
atomselect0
vmd > set sel2 [atomselect 1 "backbone"]
atomselect2
```

9. and then the RMSD (root mean square error of the distance) is computed as:

```
vmd > measure rmsd $sel1 $sel2
measure: rmsd: two selections don't have the same number of atoms
vmd > measure rmsd $sel1 $sel1
0.0
vmd > measure rmsd $sel2 $sel2
0.0
```

10. We get an error saying that the two sequences do not have the same number of atoms (strange, as we would expect the two backbones to indeed have the same number of atoms). Note that a sequence is identical to itself in terms of RMSD. We can repeat the analysis by selecting a different number of residues so that the number of atoms matches:

```
vmd > set sel1 [atomselect 0 "backbone and resid 1 to 20"]
atomselect3
```

```
vmd > set sel2 [atomselect 1 "backbone and resid 1 to 20"]
atomselect4
vmd > measure rmsd $sel1 $sel2
36.6138114929
```

11. One can align the two molecules and then compute the RMSD again (the RMSD depends on the alignment). To align, one defines a transformation matrix (rotation + translation) which is a 4 by 4 matrix:

```
vmd > set tmat [measure fit $sel1 $sel2]
{0.474566906691      0.603444337845       0.640813231468       7.79903411865}
{0.394691437483 -0.
796608269215      0.457857787609       20.2783298492}      {0.786768615246
0.0356394760311 -0.61
6218149662 0.765040159225} {0.0 0.0 0.0 1.0}
```

12. Then, the transformation matrix can be applied to the two sequences:

```
vmd > set movsel [atomselect 0 "all"]
atomselect5
vmd > $movsel move $tmat
vmd > measure rmsd $sel1 $sel2
0.599867045879
```
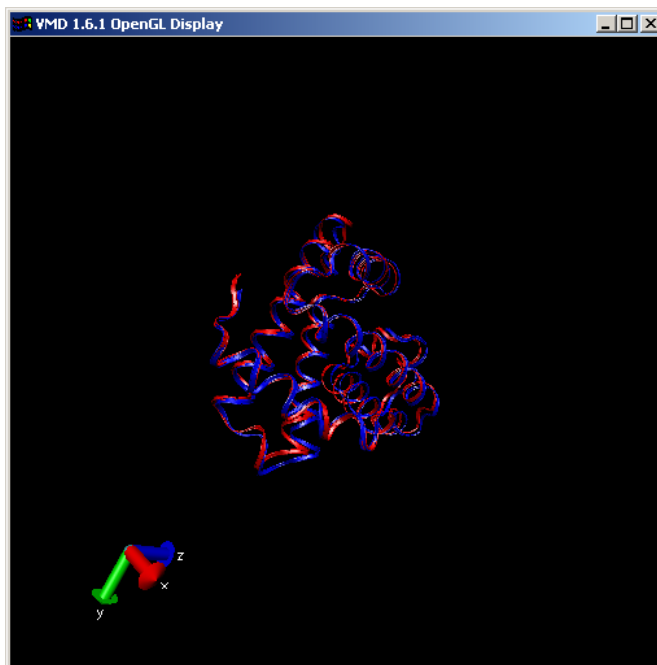

13. When the two molecules are aligned, then the match improves significantly, in terms of RMSD.

Part 2: DNA structure and DNA protein binding

This lab assignment is focused on exploring structural properties of the DNA molecule. For further reading, and as a reference for completing the exercises and answering the questions, students are encouraged to consult a basic textbook in biochemistry (e.g., L. Styer *Biochemistry, fourth edition.* Freeman and Co., 1995). In this lab, window and file names are in `Courier New` and button and option names are in *Italic*. Hypertext links are underlined.

1. **Running VMD:** Start VMD, a molecular visualization program (details on the program are available at http://www.ks.uiuc.edu/Research/vmd/). Use the Windows Start Menu and select VMD. Three windows will open: `main`, `VMD OpenGL Display` and `winvmd`. You will mainly use the `main` and `OpenGL Display` panels.

2. **Accessing Protein and DNA Structure Files:** Go to http://www.rcsb.org/pdb/. Under *Search the Archive* enter the PDB identifier: `1ZQA` and click *Find a Structure*. Information will be displayed about the molecule. From the left-hand column, select *Download/Display file* and for a file format choose PDB, complete with coordinates, and TEXT. After looking at the file, select *Save full entry to disk* and save it in a place of your own liking under the name (default) `1ZQA.pdb` (remember the location!). Now go to http://chemistry.gsu.edu/glactone/PDB/pdb.html (this is an interesting site to bookmark), click on the link *Nucleic Acids* and download the files under *DNA Double Helixes* (`adna.pdb`, `bdna.pdb` and `zdna.pdb`) and under *DNA Base Pairs* download those under *B-DNA* (`at-bdna.pdb` and `gc-bdna.pdb`). You should now have a grand total of six PDB files.

3. **Base Pair Interaction:** Press the button *Molecule* on the `main` panel. This action brings up the `molecule` window with a series of options. Select the *Load from Files* option. Under *Molecule File Types*, select `pdb only`. Go to the place where you saved your PDB files, select the file `at-bdna.pdb` and load the structure. Click on *Load Molecule*. This action will load and display the A-T (adenine-thymine) base step of the canonical B-DNA helix. Select the *Graphics* button on the main panel to open a new window called `graphics` that can be used to adjust the display. In the `graphics` window, in the panel labeled *Drawing Method* (lower left side of the window) select *Licorice*. Next, load another copy of the molecule using the `molecule` panel. In the `graphics` window you can toggle between the two copies of the molecule using the panel labeled *Select Molecule*. You will not be able to see a difference as the two molecules will be exactly superimposed. Choosing one of the copies, select the Drawing Method *HBonds*, which displays the hydrogen bonds between the base pairs. The structure can now be rotated and translated using the mouse: in the `OpenGL Display` panel, for translation, type "t" on the keyboard and move the structure using the mouse with the left button engaged; for rotation, type "r"; for sizing, type "s". Familiarize yourself with these commands.

   - *Exercise 1*: Select the `Render` window from the `main` panel: a new window labeled `render` will open. Output an image of the base pair in the file type of your choice (use `snap.bmp` for this tutorial). Before printing you should select the *Color Defs* button on the `graphic` window and select the following options: *Display* $\Rightarrow$ *Background* $\Rightarrow$ *White*. Open the image and print it. Identify on the image the adenine and the thymine bases. Identify the atoms involved in the hydrogen bonding.

- *Exercise 2*: Repeat above analysis for the G-C (guanine-cytosine) base step in the file `gc-bdna.pdb`. (You can delete molecules from the session using the `molecule` window, selecting the molecule to clear – it will be highlighted in yellow – and clicking on *Delete*).

4. **B-DNA Structure:** Load and display the molecule `bdna.pdb`. Explore the different display options available in the `graphics` window under *Drawing Methods* such as *Ribbons*, *Tube*, *CPK*. The *Tube* option traces out the curve defined by the positions of the phosphate groups along the backbone. Notice that the backbones are not equally spaced along the helical axis: there is a major (wide) groove and a minor (narrow) groove.

- *Exercise 3*: How many base pairs are in the structure? As you move along the main axis of the helix, is the rotation from one base pair to the next right-handed or left-handed? Approximately how many base pair steps make one full turn of the helix? What is the structural origin of the major and minor grooves?

5. **Other DNA Forms:** Explore the A form and Z form DNA helices found in the files `adna.pdb` and `zdna.pdb`.

- *Exercise 4*: Which helices are right-handed? Which are left-handed? In what other ways do the structures differ?

6. **Protein-DNA Binding:** Load the structure `1ZQA.pdb`, which is an X-Ray crystallographically-obtained structure of DNA polymerase beta complexed with a seven-base pair DNA. Using the *Ribbons* drawing mode, observe that the polymerase is a largely alpha-helical protein. The *Cartoon* drawing option renders the helices in the protein as solid cylinders (note that the DNA is not visible under *Cartoon*). The pdb file labels four distinct chains in the molecule as chains A, P, T, and X. The DNA chains can be selected and visualized by choosing *Licorice* as a drawing method and deleting the selection "all" and replacing the text with "chain P T" under *Selected Atoms* in the `graphic` window. Translate and zoom the DNA structure so that it is maximized on the screen.

- *Exercise 5*: Compare the base-pair interactions in the complexed DNA with the canonical B-form DNA. How do they differ?