Workshop #1: PyMOL

PyMOL is a molecular visualization tool. There are many such tools available — both commercially and publicly available:

- PyMOL (www.pymol.org)
- Swiss-PdbViewer (spdbv.vital-it.ch)
- RasMol (www.umass.edu/microbio/rasmol)
- VMD (www.ks.uiuc.edu/Research/vmd)
- MolSim (www.cchem.berkeley.edu/molsim)
- DS Visualizer (accelrys.com/products/discovery-studio/)
- *etc.*

PyMOL is particularly attractive to us, since it has excellent features for viewing, it is fast and the display quality is superb, it can handle multiple molecules at once, and it is easy to define custom objects such as complexes or sets of atoms. It is also open-source and extensible, so the expert user can create new functions such as colors and measurements related to protein design specifications. The goals of this workshop are to have you become familiar with (1) the basic operation of the software, (2) the tools for analyzing protein structures and for creating high-quality pictures, and (3) the ability to create and save scripts for repeated use.

Suggested Readings

1. Introductory: Chapters 1 and 2 of Brandon and Tooze, *Introduction to Protein Structure*, Garland Publishing (1999).
2. Advanced: J. S. Richardson, "The Anatomy and Taxonomy of Protein Structure," 1981 (updated 2000-2007), available at http://kinemage.biochem.duke.edu/teaching/anatax.

Basic Operations

Download PDB file 1YY8.pdb from www.rcsb.org.

Open PyMOL. (If asked, use the PyMOL + Tcl/Tk mode.) Load 1YY8.pdb. (From the menu bar at the top of the upper window, select File→Open→(select your file).

Use the mouse and mouse buttons to rotate, translate, and zoom the molecule:

On Linux or PC:
left button = rotate, middle button = translate, right button = zoom

On a Mac:
left button = rotate; alt/option+click = translate; right (secondary) click = zoom

The buttons at the top right can set the viewing parameters.
A = Actions, S = Show, H = Hide, L = Label, C = Color.

In the line for 1YY8, select Hide→Everything, then Show→Cartoon, then Color→By Chain→By Chain (e. c), then Label→Chains.

You should see two copies of the antibody fragment, since there were two copies of the antibody in the unit cell of the crystal that was measured to determine the structure. You should be able to see two separate chains in each antibody fragment.

If you click on any atom or residue, the viewer window will display information identifying that atom/residue. Confirm that the four chains are identified as chains A, B, C, and D.

Let's focus in on one fragment. In the command-line window (Depending on your PyMOL version, Windows labels this Tcl/Tk GUI or The PyMOL Molecular Graphics System), type the following commands:

```
select AB, chain A+B
hide all
show cartoon, AB
orient AB
```

Note that in the panel at the top right, you now can operate on the subset AB using the buttons.

You can use the mouse or the `select` command with other protein descriptors (*e.g.*, `name ca+cb+cg+cd`, `symbol o+n`, `resn lys`, `resi 100-150`, `ss h+s+l`, `hydro`, or `hetatm`) to create objects for various subsets of the molecule, and there are a variety of operations you can perform on those subsets.

You can also combine descriptors (`chain A and hydro`) and assign it a selection name as in the following:

```
select linkerA, chain A and resi 107-112
color red, linkerA
select linkerB, chain B and resi 117-122
color orange, linkerB
```

(Note that there is a distinction in PyMOL — as in many scripting/coding languages — between the addition operator `+` and the Boolean operator `and`. Make sure you understand the difference.)

Type `help select` and `help selections` for full details. (Hit the Esc key to exit the help screen.) Test out the mouse operations and various colorings and display options to get a feel for the general operation of the molecular visualization.

Note that you can use File→Save Session at any time. This will store all your objects, selections, and views.

Structural Analysis

The structure you have downloaded is cetuximab, a therapeutic antibody in development for cancer treatment. Antibodies are composed of two heavy chains and two light chains; the particular construct of its upper part is known as a Fab fragment and contains one full light chain (chain A) and the N-terminal half of one heavy chain (chain B). At one end of the Fab fragment are six loops known as the "complementarity determining regions" (CDR), which bind a particular antigen. For the following four problems, we will examine the N-terminal domain of chain A (the light chain). To make this easier, type `select L, chain A and resi 1-107`, then from the right panel controls, hide everything but for selection L and click Color→Spectrum→Rainbow.

1. Looking down the direction of the first strand, which way does it twist? _____
   Do all strands twist the same direction? _____

2. Next, let's analyze a couple strands in the N-terminal domain. Zoom in on strands 3 and 8, (which should be adjacent and colored cyan and marigold, respectively.) What are the residue number ranges for these two strands? (Click on the strand ends and look in the console window for the residue numbers.)

   strand 3: ____–____ strand 8: ____–____

   Create a new object for these two strands with `select` and hide the rest of the molecule. Display the atoms of the amino acid and color them by their element (Show→Sticks and Color→ByElement). What color are oxygens? _____ What color are nitrogens? _____ By looking at the side chains, identify the amino-acid sequence (using 1-letter abbreviations) of these two strands:

   strand 3: _____ strand 8: _____

   What is the pattern in these sequences, and why does it occur?

3. Let's analyze some geometry. From the main menu, select Wizard→Measurement. You should see a panel on the right in which you can select distances, angles, dihedrals, and neighbors, and PyMOL will prompt you to select the atoms for measurement. Label→Residues from the right-side panel might also be helpful.

   What is the distance between the N of L73 and the O of F21 (*i.e.*, the hydrogen bonding distance across the β chain)? _____

   Measure all of the backbone hydrogen bonding distances between these two strands. What is the range of distances you observe? _____–_____

   On residue F21, what is the bond angle around the C_β? _____

   On residue F21, measure φ, ψ, and χ_1:

   φ: _____ ψ: _____ χ_1: _____

Confirm that these values are within the β-sheet region of the Ramachandran plot.

Type `h_add chain A and resi 73` to place hydrogen atoms on residue 73. (Hydrogens are usually too small to see by crystallography, so PyMOL must calculate the theoretical positions.) What is the H–O–C bond angle for the backbone hydrogen bond between residues L73 and F21? _____

4. Restore the view of `L` and sketch a "Brandon & Tooze style" topology diagram (*not* a 3-D sketch!) showing the β-sheet strand arrangement for the light chain. Hint: Begin by drawing strand 1 on the bottom left side of the paper pointing up. There are two separate sheets; draw one atop the other; otherwise, it implies a single sheet. Draw the secondary structure first, and then connect the linkers. Do not try to approximate the size of linkers between secondary structure. (p. 62 of Brandon & Tooze has an example of such a two-row topology diagram.)

Comparing Molecules

5. From www.rcsb.org, find a second PDB file of cetuximab, this time bound to its antigen.

   What is the antigen? _____

   Clear your current PyMOL session (All→Actions→Delete everything") and load your new PDB file. Use the cartoon view and color and label by chain to see an overview of the structures. You should see the antibody Fab fragment and the antigen. The antigen also has several post-translational glycosylation modifications.

   Load 1YY8 into the same session. As you did before, create an object for chains A and B and hide chains C and D. (You will now need to specify the molecule: `select unboundFab, 1YY8 and chain A+B`.) Similarly, create an object (Call it `boundFab`,) for the Fab fragment of the bound complex. (Be careful to specify the correct chain identifiers; they are arbitrary and can vary between PDB files.) Now, superimpose the two structures using `align unboundFab, boundFab`.

The structural match between the two molecules is measured by the root-mean-squared (RMS) distance of the aligned atoms:

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|\mathbf{x}_i - \mathbf{y}_i|^2}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the vector coordinates (displacement vectors) of the *n* atoms in the two structures. The `align` command automatically generates a sequence alignment to pick the right atoms to compare and then solves for and executes the coordinate transformation that yields the minimal RMS deviation between the structures.

6. In the command window, there should be a few lines describing the alignment process. What is the RMS error calculated for this structural alignment (include units)? _____ Over how many atoms? _____

7. Is there much difference between the bound and unbound forms of the antibody? In particular, are there differences in the six complementarity-determining loops at the far end of the N-terminal domains?

High-Quality Visualization and Scripting

Your commands can be saved to a file or read in from a file. Use the File→Log option to record your steps and create a script. You can edit this script using a text editor such as Notepad, WordPad, jEdit, vi, Emacs, or IDLE. You can then read in the script using File→Run or simply with the command `run myfile.pml`. (UNIX commands, such as `pwd`, `ls`, and `cd`, can be typed in PyMOL to assist in locating any scripts you wish to run.) The script will record all of your settings, but not necessarily the transformations you make by reorienting the molecule with the mouse. To record the screen orientation matrices in your script, type `get_view`.

The command `ray` will use a ray-tracing algorithm to compute the lighting on the molecule. (`ray 800,800` will set the image size to 800 × 800 pixels.) Use this before saving an image using File→Save Image to create publication-quality results. Since the natural background color on a piece of paper is white, use the command `bg white` to change the background color (and use less ink!). Other options are under the Display menu; some options that may help include Display→Color Space→CMYK and Display→Depth Cue→On. The menu command Setting→Transparency can also help show depth and occluded molecules, but it is most important to orient the molecule carefully to show features and to hide all but the most relevant parts of the molecule. Finally, you might also try some of the preset settings from the right-side menu under Actions→Preset.

8. For your last task, choose an interesting feature of cetuximab (β-sheet structure, the antibody complementarity-determining regions, a comparison of bound and unbound antibody loops or the CDR H3 loop in detail, the glycosylation on one of the EGFR side chains, *etc.*) and create a beautiful, ray-traced, white-background, publication-quality figure. Color and label protein features and measurements as you feel appropriate. Use the script feature to gather the list of commands that you find optimal for viewing your object. Edit the script to eliminate the non-essential pieces and make the script clean, concise, and comprehensible.

   If you are doing this exercise for a class, submit the figure printed in color, the script that can re-create the figure, and a brief statement of which structural feature your figure is designed to show.

   If you work in a research lab, you are encouraged to create a new figure for a protein relevant to your research.