DESKTOP MOLECULAR GRAPHICS in PyMol
[Adapted from:
https://bioquest.org/nimbios2010/wp-content/blogs.dir/files/2010/07/pymol_tutorial3.pdf**]**

Pymol Installation:

Go to: https://pymol.org/edu/?q=educational to fill out the form - strongly suggest you specify an end date of 12/31/20 as you will be using pymol in some of the Quantitative Biology Labs

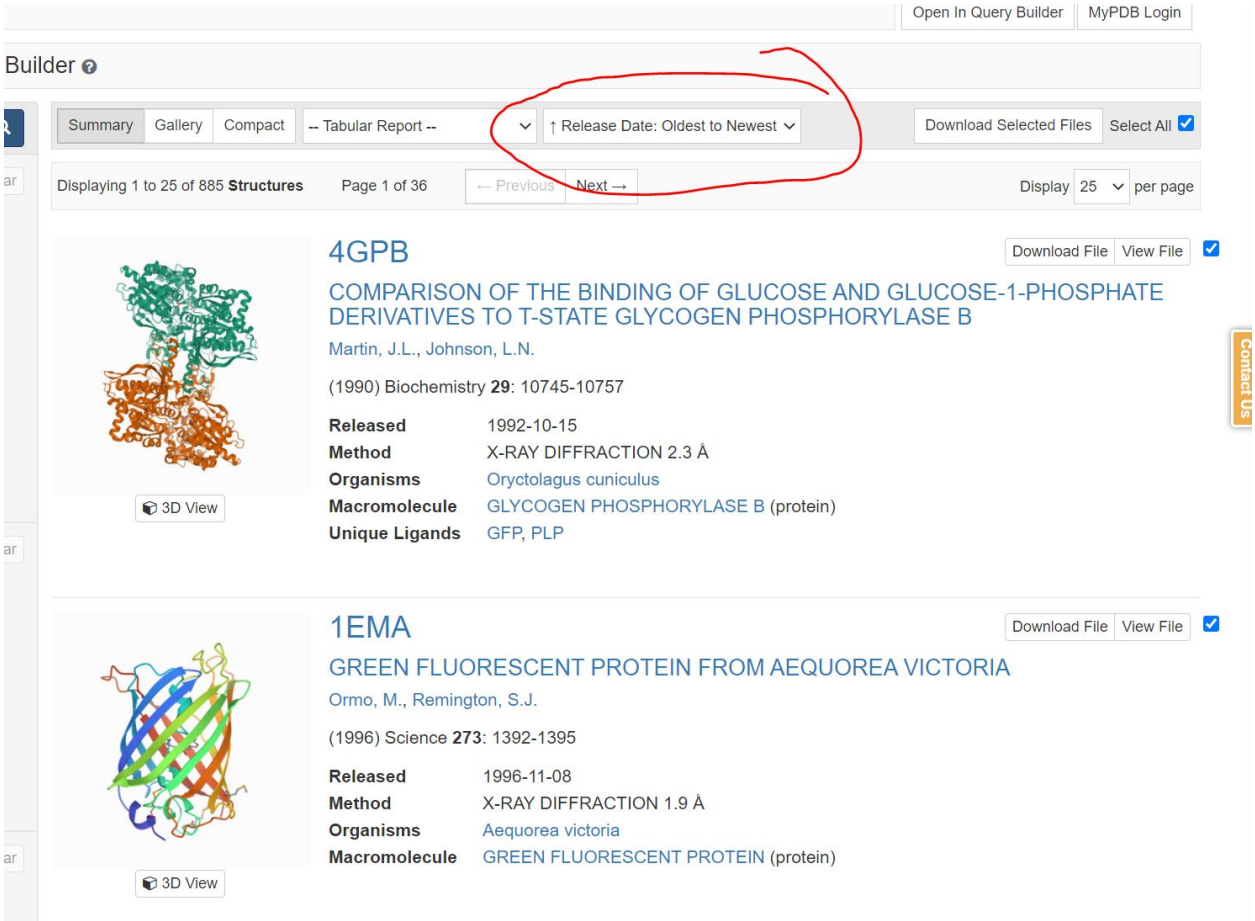
You will receive a link in your email with a ".lic" file, as well as a link to the installation site for pymol (https://pymol.org/2/) - after downloading (either EXE or DMG as appropriate for your computer) you should install the program. Once installed, run it. It will request the license file - point it at the ".lic" file you downloaded.

PyMol - Exercise A: Download a PDB from the repository

First - let's download a "PDB" File. PDB or "Protein Data Bank" files have contain coordinates of the different molecules within either a crystal or a cryo-EM structure (depending on how the structure was solved. This is an extremely valuable database whenever you want to look at different structures of proteins, often even combined with their ligands or other molecules. The biological functional entity can be either a multimer of the deposited structure, or just one of multiple copies within the file. In the following example we will download one functional biological subunit, in this case a monomer.

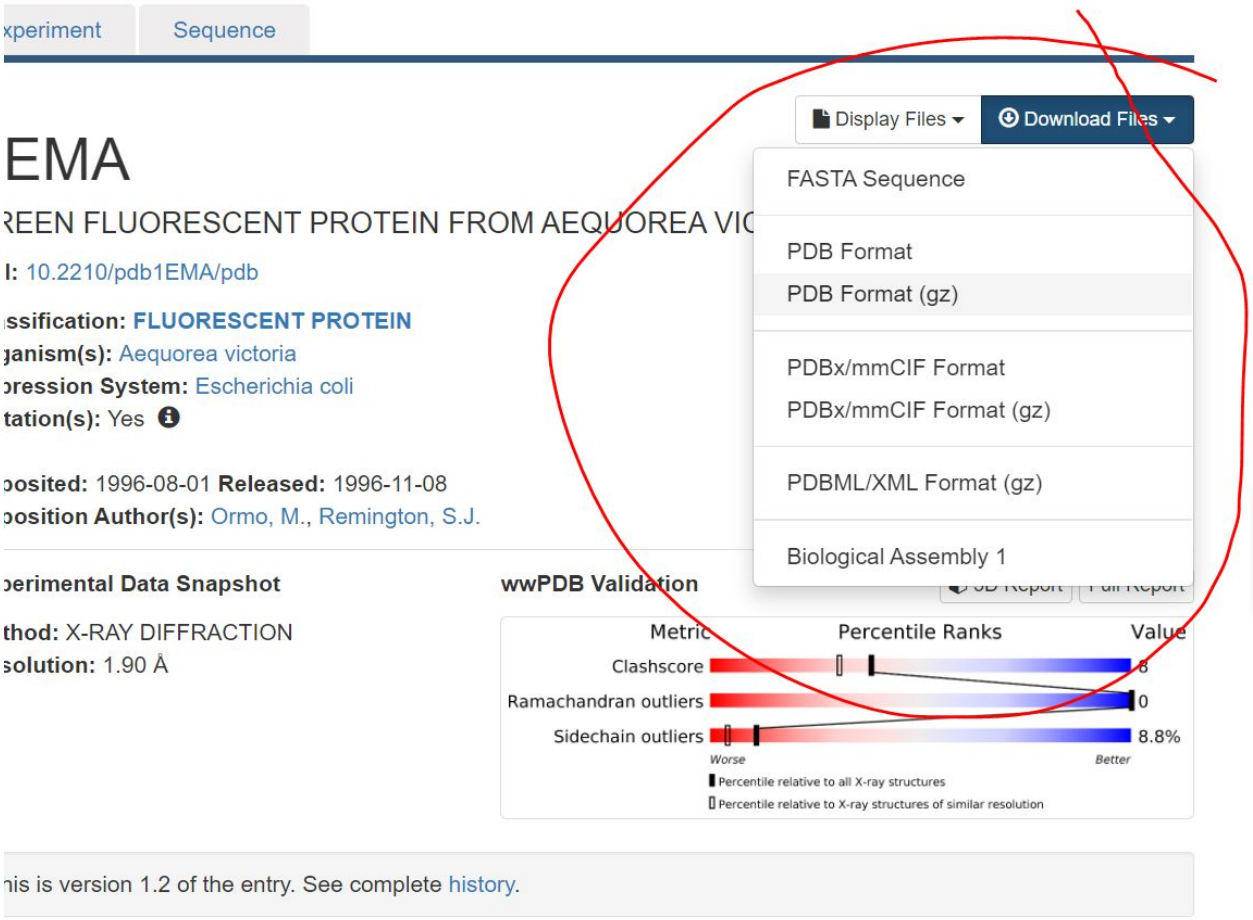
Structures have a PDB ID code made of 4 letters and numbers.

1) Go to rcsb.org
2) Let's search for an oldie but a goodie - GFP. Type GFP into the search box.
3) Lots of hits! Let's look at the older (and less tailored) structures - sort from oldest to newest:



4) Click on the 1EMA link. Let's look around - lots of information here, including a link to the original paper, how the data was generated (X-Ray) the purported resolution of the structure, and even a 3D view of the classic "barrel" structure of GFP

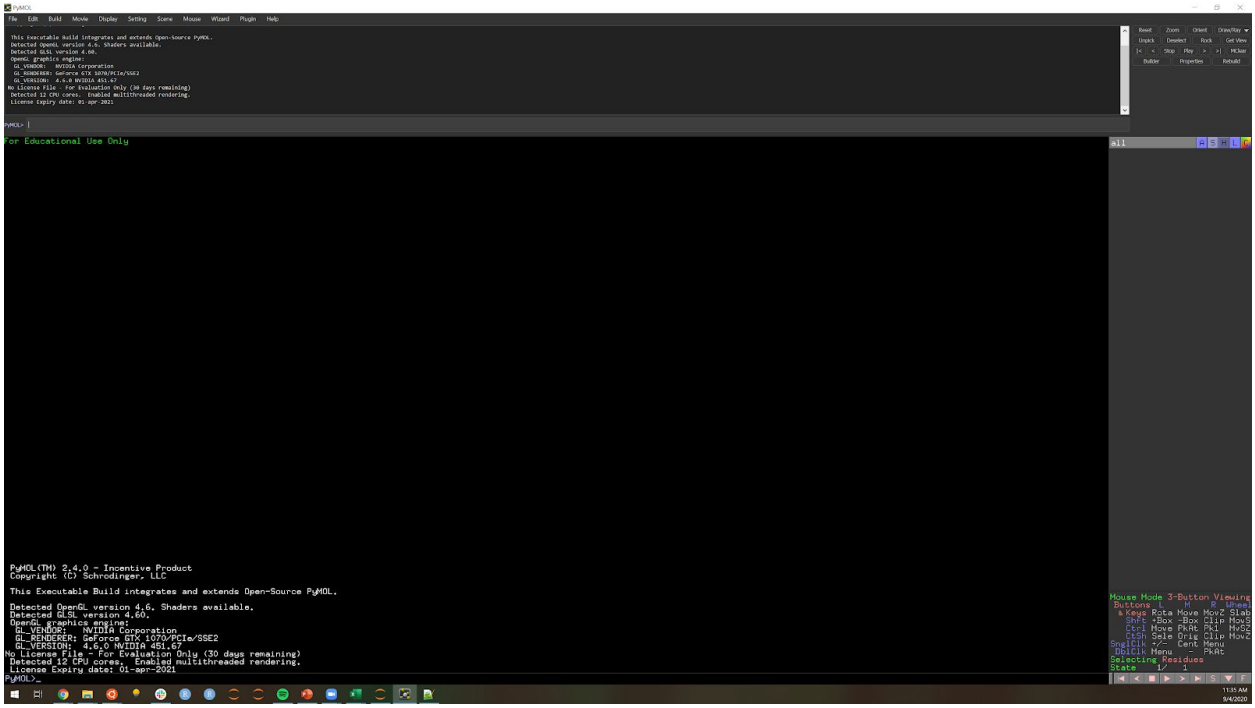
5) Let's download the PDB file to your local computer:



6) Open the PDB file in Notepad or whatever local text browser tool you have - you'll see a lot of "header" lines of text at the top, then a series of coordinates for the different parts of the protein, even down to the individual atom.

PyMol - Exercise B: Open PyMol and load a PDB file

1) Start PyMol. It should look something like this:



2) There are a couple of ways to load our GFP PDB file.
    a) You can also use the GUI and go to File->Open.
    b) Pymol also knows about the RCSB - so you can just type in the console `fetch 1ema` and it will download and load the file for you.

PyMol - Exercise C: PyMol interface

By default PyMol will display the molecule(s) contained within the PDB file as a "cartoon" representation. The 3D molecule is represented within a virtual 3D world. The flat surface of the screen represents the X and Y axes while the Z, depth axis is perpendicular to the screen.

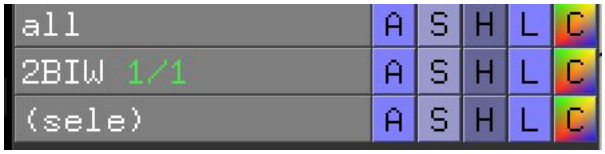
1) Test mouse control of the 3d representation

PyMol is optimized for a 3-button mouse but most basic functions can still be achieved by a one-button mouse or trackpad, in particular the rotations around X, Y, and Z.
Rotation: (left) click and drag.
Translate (move sideways) X or Y: click middle button and drag.
Zoom (move along Z axis): click right button and drag up or down.
Slice (cutaway mode along Z-axis): mouse wheel.
Practice manipulating the structure and zooming in and out on a point of interest.

2) Changing the representation of the molecule

PyMol can open more than one molecule at a time, or separate complex PDB files into individual components. Each opened or loaded molecule is given a name within the "Names Panel" (see picture above). The first name is always "all." Clicking on the name itself will undisplay the corresponding molecule(s) (temporarily invisible).

The ASHLC menu (below) is abbreviated for Action, Show, Hide Label and Color. Some menu items have submenu components. Selections made under the "all" line will affect all the opened molecules.



Rule: Once a selection is shown (S) it must be selectively hidden (H) as it is not removed when another selection (S) is made. Selections are therefore additive, which allows for the creation of images with mixed graphical representations.

Let's make a wireframe representation of this protein:

1) First go to 1ema->H-> all to clear all the representation
2) Then go to 1ema->S->Wire
3) Ok - let's go back to cartoon - 1ema->S->cartoon and 1ema->H->Wire

Options: most options can be set within the "Setting" menu within the top menu bar. For example, it is possible to change the way all alpha helices are rendered.
Go to the following menu cascade:

Setting > Cartoon > Cylindrical Helices

Select this option again to remove its effect and do the following:

Setting > Cartoon > Fancy Helices

Testing other cartoon settings:
Engaging the option Smooth Loops will simplify the drawing.

Changing the background to white is usually very useful: The "Display" menu within the top menu bar contains options for most options pertinent to displaying the image within the PyMol viewer. To change the background color to white follow this menu cascade:

Display > Background > White

Note that there is "fog" within the back of the molecule, which can be toggled on and off with the Depth Cue menu item within this same menu list.
Changing the color of the ribbon is easy with the following cascade menu within the PyMol "Names Panel" of the "Internal GUI" under the C menu as shown in the following menu cascade:

1EMA > C > oranges > orange

You may also choose another option which is to color by secondary structure by following this menu cascade instead:

1EMA > C > by ss > Helix Sheet Loop

Changing the representation of the molecule: adding ligand
When we opted to show the molecule as a cartoon above, one thing happened: the protein was shown as the familiar cartoon representation, but if any ligand is present (which is the true for this file) it simply disappears as it is not part of the cartoon representation of the protein. Here we will "rescue" the ligand!

This menu cascade is also within the "ASHLC" menus of the PyMol "Internal GUI" "Names Panel." Since ligands are usually small, organic molecules, the following cascade within the Show menu for line 1ema will show the ligand. Perform the following cascade:

1ema > S > organic > spheres

This cascade will select the chromophore present within the PDB file.

5) Mouse selection
Now that we have made the ligand visible, it becomes easier to select it with the mouse to make further changes. Click on one of the spheres of the chromophore. This causes you to select the ligand - and a resulting new row appears on the right side of the GUI. The name of the atom that was clicked appears within the top text window of the "external GUI."

Now that the ligand atoms are segregated within the name of (sele) we can change its color if we want by the following cascade within the Names Panel:

(sele) > by element > CHNOS….

6) Final image(s)

To export the image - you have a couple of options. You want to go to File->Export Image As -> PNG. This will open a dialog box with a couple of options of varying levels of quality - ray-traced is the nicest with shadows, etc, but takes the most time to render.

We can also make *movies*. Go to Movie-> Program -> Camera Loop -> Y-Rock -> 180 deg over 24s. Then go to File->Export Movie->MPG - and save out the movie. This will take quite a bit longer, as it has to render each screen carefully (if you set Ray), but is beautiful for presentations.

PyMol - Exercise D: Useful commands to analyze structure and create images

Let's work with a different molecule for now - 2biw. Fetch it or download it from RSC.

1) Sequence viewer

The molecule sequence can be shown at the top of the graphical area. To do this follow the top menu:

Display > Sequence

The sequence appears just below the PyMOL> line command at the top of the Viewer. The slide cursor underneath can be used to move the sequence viewed further. By default the one-letter code is displayed. Since we are looking at a protein, changing the display to the three-letter code will be useful within the next step. Do this with the menu cascade:

Display > Sequence Mode > Residue Names

Navigate to Tyr 322 on the sequence panel and select it with your mouse. Change its aspect to stick and its color to red with the following menu cascade:

(sele) > S > sticks

(sele) > C > red

Now select residues PHE 303 and HIS 304. With the mouse click on residue 303 and residue 304 within the sequence line. Note that they are selected within the graphical window. The arrows within the following image show where you should look. As before, transform those side chains to a thicker stick with the menu cascade:

(sele) > S > sticks

(sele) > C > red

Consider the orientation, proximity and identity of these three selected amino acids and what their potential functionality within the protein might be.

2) Measuring distances

Distances are measured between two atoms and are expressed in the same unit as the XYZ coordinates within the PDB file: Angstroms (1Å = 10–10 m). As an example, we will measure the distance between two atoms within the carotenoid ligand. Within the top menu select

Wizard > Measurement

This will create a prompt within the Viewer: " Please click on the first atom…" Within the "Internal GUI" a "Measurement" table also appears. It can also be used to remove measurement objects after they are no longer needed.

Click on the first atom on the ligand.

Click on the second atom on the ligand, or an alternate atom. The default color of the measurement object is yellow. This can be easily changed with the familiar ASHLC menu for each measurement, as a new name is entered within the "Names Panel" for each distance, e.g. "measure01." Change the color to white with the following menu cascade:

measure01 > C > grays > white

Measure the distance between the OH at the tip of TYR 322 and the nearest ligand atom (C30).

When you are done using the "Measurement" panel on the bottom right click Done.

If you no longer need to display the distance object, click Delete All Measurements. Alternatively, you can use the corresponding “A” menu and select the delete option.

3) Labels

A label containing the residue type and sequence number (e.g. TYR 322) can be added to a selected residue with the menu cascade

(sele) > L > residues

For publications or slide making, labels could be added with e.g. Adobe (PhotoShop, Illustrator) or Microsoft (PowerPoint) graphic editing software. At this point you can explore the L menu, knowing that “clear” will remove the mess that may occur!

4) Saving a PyMol session

A session file is a binary file containing all the information and graphical displays currently within PyMol. It is a way to save the “current state” of the software with all that it contains. Later, the file can be opened and everything is restored.

a. Saving the session file: From the top menu follow the menu cascade: File > Save Session In the next window enter a name: MySession and save it on the Desktop to easily find it again. The successful writing of the file will be reported within the ext window of the “External GUI” and a file called Mysession.pse on the desktop (.pse is added automatically). Now quit PyMol: File > Quit

==========

PyMol - Exercise E: Harnessing the power of PyMol: introducing scripts

GUI interfaces are nice, but they are slow and cumbersome. PyMol also has a line-command interface that we have already used to type specific commands. These commands are simply lines of text, and therefore can be placed sequentially within a plain text file called a script. When the script is invoked from the PyMol line command, the commands on each line of the script are executed, and the final image is shown within the Viewer.

1) Set your working directory. On the PyMol line command, type “pwd” (without quotations). You can either choose to save your new script to this location or you can change your working directory to be wherever you wish to save the script. You can change the default directory using “cd”, for example “cd Desktop “, then “pwd” will show that you’ve altered your working directory to the location Desktop.

2. Use a text editor to create a new script called abc.pml. Save abc.pml to your working directory. Within abc.pml file, type the following lines (comment lines which start with #)

```
# this is my first script:
fetch 2BIW
bg_color white
hide lines
show cartoon
color red, ss h
color cyan, ss s
color yellow, not ss h and not ss s
```

[Note the use of the comma (,) after the name of the color. ss signifies secondary structure h signifies helix. (Note that the word “helix” would not be recognized.) s signifies sheet. (Note that the word “sheet” would not be recognized.) not h and not s are be the loops or turns.]

7) Invoking a script: the @ command Rule: a script is invoked with the @ command after its name within the line command: example: @abc.pml.
Type @abc.pml within the line command.

==========

PyMol - Exercise F: Select command, parameters, scripting, and subsets.

The PyMol "select" command has the special property to create atom selections with names appearing within the Names Panel. This is a nice feature, but some confusion can arise depending on the words that are used. For example, in the abc.pml script we have just used the command "color red, ss h" to colors red only the secondary structures (ss) that are in helix (h) form. However, the command "color red, ss helix" would not at this time have any effect, and PyMol might give us an error message. The reason is that "helix" has no meaning at this point. Interestingly this command CAN work if "helix" is defined first. "helix" can be defined when an atom selection is created with the "select" command. Any subsequent command can then contain the chosen word, and the commands will only apply to that atom selection, which can be a subset of a larger selection.

1) Defining by selection

First let's assume that you just ran the abc.pml script above on a fresh start of PyMol. If not, simply quit and restart PyMol, then in the line-command area type cd desktop, press return and then type @abc.pml to be in the same state as if you just had done the previous exercise. In the script the helices are made red. Here we will make them green. Within the line command type:

color green, helix

Note the error echoed within the PyMol text window

Now within the line command type:

select helix, ss h

Two things will happen: a new name "(helix)" will appear within the Names Panel, and all the atoms participating in a helix structure will be selected within the Viewer (pink squares):

Within the line command type:

color green, helix

Note: this time it worked and the helices were colored green. The (helix) selection can be deleted with the command: delete helix, however the name within the Names Panel only disappears when the mouse is clicked within it. Alternatively the selection can be deleted with the "A" Actions menu "delete selection." The next menu allows to rename the selection.

Important lesson: the word "helix" here could have been any other word such as "MyHelixSelection." Using a "My" prefix before some of the selections might help avoid confusing words that are commands and words that are the names of created selections. This can be important when writing or deciphering other people's scripts!

For Fun at home!! (Nothing to Turn in)

PyMol – Practical exercises: Structural analysis

Fetch 1YY8. 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: _____

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? _____–_____