MOLECULAR VISUALIZATION LAB USING PYMOL – a supplement to Chapter 11

Please complete this tutorial before coming to your lab section

(Adapted from Dr. Vardar-Ulu Fall 2015)

Before coming to your lab section during the week of Oct 18-Oct24th, make sure that you have a working PyMOL program on your laptop computer. If you don't have a laptop you can bring to class, email me for special arrangement. Also, please make sure to bring a three-button mouse if you have one.

To download PyMOL:

DOWNLOAD URL: http://pymol.org/ep
USERNAME: dec2017
PASSWORD: ubiquitin

To check that the program is working:

Open up the program and make sure that you get two windows (one black display and one white text window) On the text window type:

fetch 1IHY (the second character is capital letter i)

If you see a molecule appearing on the display window, your PYMOL is working.

For Mac Users:

To use the pdb-loader service plug-in, they simply need to rename the executable, and they need to have X-Quartz installed.

http://xquartz.macosforge.org/landing/

Using MacPyMOL

The standard OS X Pymol application, MacPyMOL.app does not run with the Tcl/Tk interface which is required for plugins to work. However, a quick renaming of the program from MacPyMOL.app to PyMOLX11Hybrid.app makes the application run as an X11 application, and plugins are now available.

To rename the executable, right click (or control click) on MacPyMOL and choose "Get Info" in the Panel. Change the Name & Extension to PyMOLX11Hybrid.app. This name can also be changed using the mv command in Terminal.app.

WEEK6: Examination of GAPDH Structure using PYMOL

INTRODUCTION: (Adapted from CH11 from the Lab Manual)

Molecular modeling and database searching have become essential parts of modern molecular sciences. It is important for anyone working in biochemistry and molecular biology today to be familiar with the many databases with molecular information. Access to the tremendous amounts of data compiled is essential before, during, and after any scientific project. After all, you should know what has already been done before attempting any new project or experiment. As a biochemistry student, you should become familiar with retrieving and manipulating molecular data using the computer. In a recent article in USA Today entitled, "Databases May Put Drugs on Shelves Years Faster" they cited the increasing efforts by thousands of biotechnology firms to use available DNA sequence and protein structure data to supplement, or even replace, the high cost trial-and-error work of drug design. Scientists are manipulating almost every gene in the human body and other organisms, and are developing huge databases to house the knowledge gleaned from their studies. This information is helping design new therapeutics at an unforeseen rate. In particular, protein structures, especially likely targets for therapeutics, as well as molecular modeling have led to a field called "rational drug design," which is revolutionizing the pharmaceutical industry. This revolution's consequences are already being felt throughout medicine and society.

Forty-five years ago at the height of the cold war, the United States Defense Department sponsored a small computer communication project designed for wartime emergencies. Several universities such as the University of California, MIT, and University of Michigan further developed this project. Later, scientists at Stanford University and the European Organization Nuclear Research (CERN) standardized these network projects for companies like Hewlett-Packard and Digital. The Internet went worldwide in 1989 and is now a thriving worldwide commercial platform in the public domain.

The Human Genome Project was independently begun in 1984 with the goal of obtaining the complete DNA sequence of humans, as well as other organisms. Furthermore, the advent of more rapid methods to determine the three-dimensional structures of proteins has led to an explosion in the database that collects all these structures. This database, The Protein Database (PDB), was first housed at the Brookhaven National Laboratory in New York, but is currently curated by Rutgers University in New Jersey. With the growth and popularity of the Internet, access to all the information in these databases is now readily available.

TUTORIAL:

PYMOL program can be run via command lines as well as clickable buttons. In this tutorial you will have instructions to do it either way. Command line instructions will be in "black" and clickable button instructions in blue. You can find a much more extensive tutorial and help at the PYMOL WIKI page: **http://pymolwiki.org/index.php/Main_Page**

Start by opening PyMOL and loading your protein of interest (human LDH chain A) by typing in the PyMOL Molecular Graphics System window command line in lowercase letters the pdb ID number. The PDB ID we will use is 1IHY.

fetch 1IHY

(You can also load your structure through the PDB Loader service found under the "Plug-in" Tab on the right hand top of your Command/text window next to the tab Wizard – Please note if you have a Mac and you don't see this tab please go back to the setup instructions and install the plug-in). Click: Plugin→ PDB Loader Service and type 4OKN in the "code" box leaving "chain box" empty.

This will bring up the crystal structure with the polypeptide chain represented in lines. Carbon will be colored green, Nitrogen will be blue, and Oxygen will be red. Hydrogen atoms are not shown. The red X marks around your protein are water molecules whose electron density was seen in the crystal structure. Also there is a ligand in this structure. ARP (recorded as residue 335)

- Clicking with the left mouse button will select an amino acid residue, clicking anywhere in the black background around the molecule will de-select it.
- Holding down the left mouse button and moving the mouse will rotate the structure.
- Holding down the right mouse button and moving the mouse will zoom in and out on the structure.
- Holding down the middle button and moving the mouse will shift the whole structure to a new center point (translate).
- At any point you can "save" the current state of your work as a session by go to: File→ Save session as→ name. This will create a PYMOL specific file with an extension .pse which you can always reopen (File→open) within PYMOL, or by double clicking it within a folder. Please save your work often as going back (UNDO) doesn't work very well in PYMOL.
- To save a figure you have created in a format that you can use to insert in the word document, go to: File→ Save image as→ png and save on your computer. Then insert this picture in the requested space for any Question on the worksheet following this tutorial.

1. With the line representation it is difficult to see structural features of the protein. It helps to switch to another representation. On the right-hand panel of your display window, you should now be seeing an object name "1IHY". You can use the menu items ASHLC next to the any object in the right-hand part of the viewer window to generate various actions (A), show various

representations (S), hide various representations (H), manipulate labels (L) or change the colors (C).

Hide the complicated default line representation by clicking on H (hide) next to your object name and choosing "hide everything". Make a cartoon representation of your protein by choosing S (show) and submenu "cartoon". You should now see the molecule represented as a cartoon of secondary structure elements (helices, strands, and coil). You can change the look of the cartoon by choosing Setting (top menu in the small text window) and then Cartoon. For example, you can select the Highlight Color. This makes the edges of the strands and the inside surface of the helices a second color (grey by default).

Click on C (color) and then select “by chain” and then again “by chain”. Rotate your molecule around to get a better feel for what you are viewing.

Finally if you prefer a different background color for your display screen click on the “Display” tab at the top, select “Background” and your desired color.

Click (A) and “delete the object”. Then reload the pdb file using the plug-in PDB Loader service and typing in 1IHY for “code and A for chain.

2. Often when working in PyMOL, you’ll want to make some comparisons between different structures. To keep everything easy to understand, it’s useful to re-name your structure. Type in the PyMOL window:

copy GAPDH, 1IHYA
delete 1IHYA

OR click (A) → rename object

Now your protein structure will be displayed as the object GAPDH in the list to the right of the PyMOL Viewer window. PyMOL is picky about file names—keep things simple and short with no spaces or characters other than letters and numbers.

3. To display the primary structure (sequence) of your loaded structure above the structure type:

set seq_view, 1 (to turn it off set seq_view, 0)

Or click (S) found at the bottom of the right hand panel to the right of the arrow buttons (not the S next to the object name)

You can also display the sequence by clicking on Display then Sequence in the top PyMOL window.

By pulling the slider under the sequence, you can see the sequence for the entire structure including water molecules (all the 0’s at the end), and ligands (if present they are listed right before or after the protein chain before the waters).

The sequence viewer is a very handy tool. When the sequence viewer is turned on, you can select individual residues or multiple residues by selecting residues, using the mouse, on the sequence viewer.

4. Hide the complicated default line representation by clicking on H (hide) next to your object name and choosing "hide everything". Make a cartoon representation of your protein by choosing S (show) and submenu "cartoon". Try various representations of the molecule by clicking on S (show) choosing a representation (lines, sticks, ribbons, surface etc). Note that one representation does not replace another but is added on top of another. Representations are only removed by using the corresponding entry in the H (hide) menu. For the surface representation it can take a few seconds to display, depending on how fast your machine is. To color the protein by secondary structure, click the C (color) menu and choose by ss (secondary structure) and any of the coloring options.

The different types of secondary structure are stabilized by different hydrogen-bonding networks between the peptide bond amine and carboxyl groups in the backbone of the protein as well as interactions between amino acid side chains. Alpha helices have the following hydrogen-bonding pattern:

H - bonds in α - helix

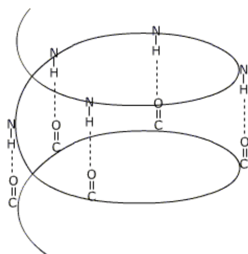


Image downloaded from TutorVista.com

Beta sheets have a different hydrogen bonding pattern:

Anti-parallel

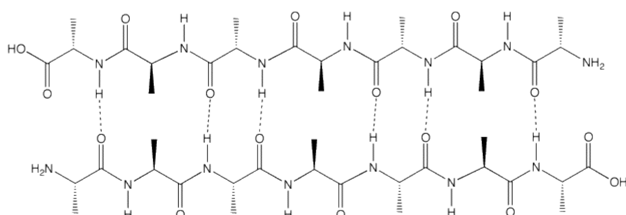
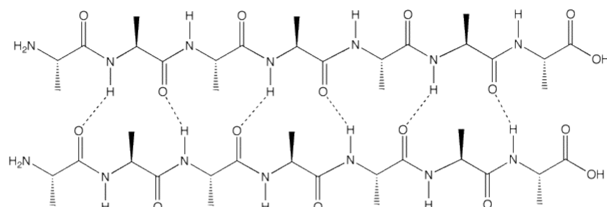


Image from:
http://www.vanderbilt.edu/chemistry/Rizzo/Chem220b/Q_A.html

Parallel



The arrow on the beta strand cartoon shows the orientation of the polypeptide chain from N to C terminus. Traditionally helices are labeled by letters (A through Z) and beta strands by numbers (1 through 10 or more)

To show the identity of an atom, double click with the left button on any point in the molecule. This brings up a menu that allows you to do various actions to the atom, residue, chain or molecule. Anytime you click on a residue to select it, the text window will tell you what you have selected, e.g. "You clicked /GAPDH//A/ARG`10/CB" means that you have clicked on the beta carbon of the Arginine residue at position 10 on polypeptide chain A of GAPDH.

When you have familiarized yourself with your protein's basic structure, it is time to start describing it.

4. To make descriptions of a protein clearer, people often number the β-strands and letter the α-helices in order, beginning at the N-terminus. Changing the coloring to "by secondary structure" "ss" from the Color menu may help you to identify the number and orientation of the secondary structural elements. Additionally turning on the Sequence viewer will help you to identify the start and end of each of the secondary structures in your structure.

5. If your protein contains a cofactor or bound ligand in the crystal structure, you can display it by choosing "organic" from the show menu (S) and then selecting the stick or sphere representation for the ligand or cofactor. This GAPDH structure contains one ligand: APR (ADENOSINE-5-DIPHOSPHORIBOSE)

During your work, it will be important to freely manipulate the color appearance of your enzyme and ligand separately. To create a separate ligand entry, clear your selection by clicking on the background of the graphics window, scroll through the structure line at the top until you find the ligand "APR". Click on APR and you will have selected on the ligand in its entirety. In the updated (sele) entry to the right, click on "A" -> "duplicate", and a new entry will appear. On this new entry, click on "A" -> "rename selection" and type "LIGAND". Now in the **all** entry, click on "H" -> "hide everything", then on **GAPDH**, click on "S" -> "cartoon" and on **LIGAND**, click on "S" -> "sticks" and then click on "C" -> "by element" and select the first set of atoms.

After you focus in on a characteristic that you think is particularly important for the structure or function of your protein, it may also be useful to measure distances (in Ångstroms) between atoms in the protein. To do this type

distance atom#1, atom#2

Ex: distance 49/CA, 236/CA (distance between residue 49 alpha carbon and residue 236 alpha carbon). This will now be marked on your graphics interface with a dotted line and the measurement "7.2" angstroms

OR select "Measurement" from the Wizard drop down menu on the top PyMOL screen and then identify the first and second atom for the distance measurement by clicking on the atoms in the structure.

Residues in the protein that participate in ligand interactions can be viewed in the "2D Diagram & Interactions" section of the "Small molecules" part of the original PDB entry site for your molecule.