PyMOL tutorial:
protein structure visualization

By Luigi Scietti

luigiangelo.scietti@ieo.it   🐦·Ⓧ **luigi_scietti**

European Institute of Oncology (IEO, Milan - IT)

Version: P4EU structure workshop - November 2023

1. Basic requirements to run PyMOL

In order to properly perform this tutorial, you are required to:

1. **Install PyMOL** on your PC. In the folder "INSTALLATION" you can find PyMOL for MAC and PyMOL for windows. For windows users go to PyMOL_WIN>PyMOLWin.exe. Follow the procedure and click "precompile" when asked.

If you want, you can instead use the educational version of PyMOL. You will need to register here https://pymol.org/edu/ . You will get an email with instructions to download the program and insert the license file. It has more advanced features, but most importantly it removes the annoying red alert "no license" on top of the screen.

2. Have a **mouse with a scroll wheel**.
3. Have a **second monitor** to be able to follow in teams and perform the tutorial

2. Getting started

PyMOL may look quite complicated at the very beginning, but I will guide you through it. I will show you how to load a structure, deal with different visualizations, and understand more advanced commands through the command line.

Please, take your time to have some fun. The more you play with PyMOL, the more you will learn how to deal with structure visualization.

In case you need a support, an online wiki is accessible here:

https://pymolwiki.org/index.php/Main_Page

This is a specific tutorial built for you, but you can find others on more advanced features here:

https://pymolwiki.org/index.php/Category:Tutorials

We will be using command line to type in commands for protein visualization. They are a lot, and you might forget them. Don't worry, here you have a full list of command to browse:

https://pymolwiki.org/index.php/Category:Commands

3. Structure repository and loading

If it exists, you can download the structure of your favorite protein target in the Protein Data Bank, the database collecting all the protein structures. It's available here: https://www.rcsb.org/ .
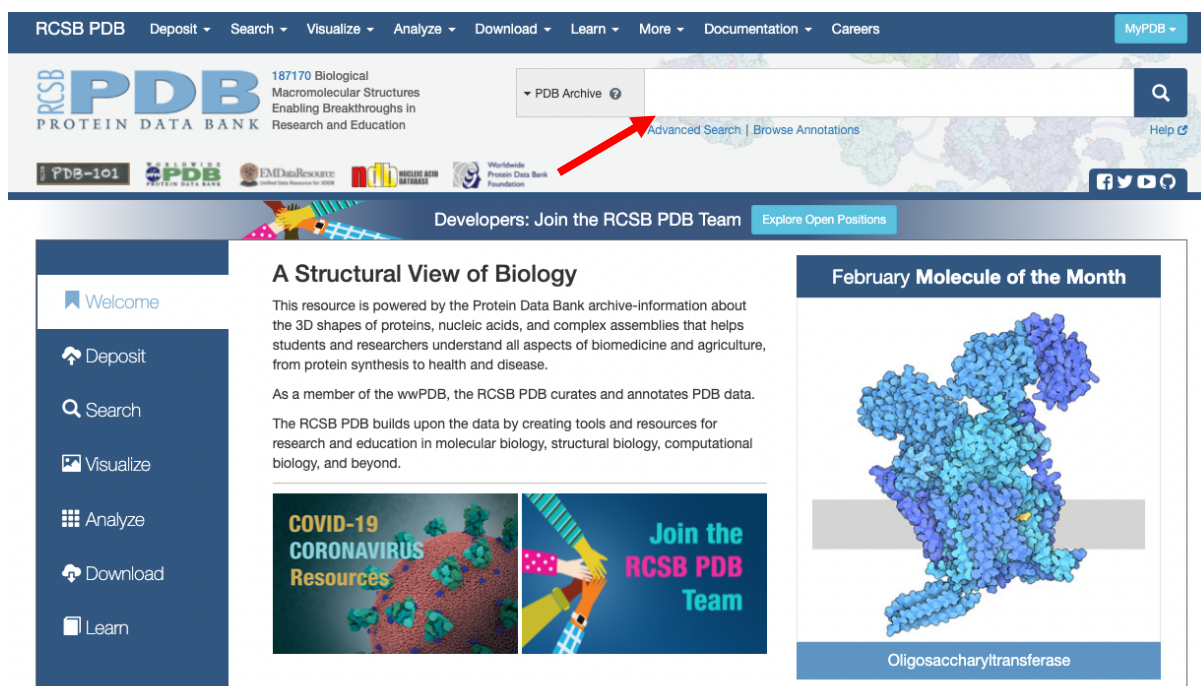

All the protein structures exist as ".pdb" file. In the top right part of the PDB main page, you can search for protein name or PDBid. The PDB file name consists of a number followed by three other characters (e.g. 6FXR) and is a text file containing the coordinates of atoms in the 3D space. PyMOL and other visualization software "read" this list by placing atoms in the space, allowing us to visualize them.

First of all, start PyMOL.
The main window is composed in this way:
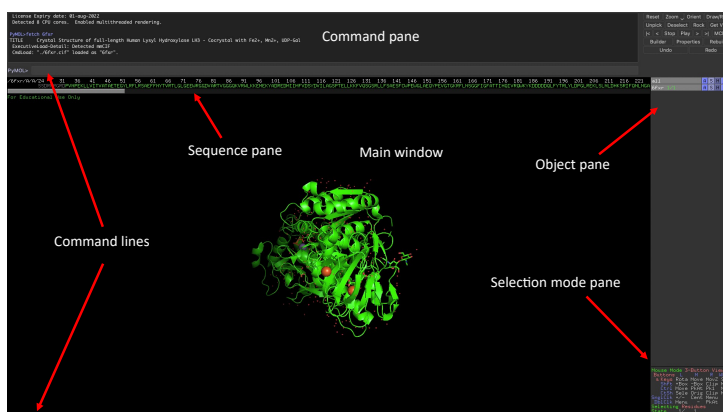
To load a structure, you have 2 ways:

A. You go on the PDB site, on the structure page (where you can also see several details on the deposited structure) and download the PDB code

B. You can go directly in the PyMOL command line and type: fetch 6FXK. The structure will be directly fetched from the pdb (you must be connected to the net!)

This tutorial will focus on the collagen lysyl hydroxylase and glycosyltransferase LH3/PLOD3 enzyme. It's a multidomain enzyme that carries two catalytic activities. If you type PLOD3 in the PDB, you will find many structures.

Now load in pymol the LH3/PLOD3 structure in complex with cofactors Fe^{2+}, Mn^{2+}, 2-oxoglutarate and UDP-Glc. The PDB id is **6FXK.** To do this, go in the command line and type:

> fetch 6FXK

Once the molecule is loaded, try to rotate, zoom in and center using the mouse:

• Rotate: click and hold the left mouse button and move

• Zoom: click and hold the right mouse button and move

• Center: click with the scroll wheel on the atom where you want to center slabbing. If you scroll the rotating wheel, you will change the "slab".

4. Structure visualization

In the Object pane on the right, you have different letters:

• **A**: action pane to perform action on the object

• **S**: show pane to show representations of the structure

• **H**: hide pane to hide representations of the structure

• **L**: label pane to put a label

• **C**: color pane to color the structure

The representation you see is called cartoon. There are different types of representations: **cartoon, sticks, lines and surface**. Alternatively, you can also type in the command line. For simplicity, we will let you use the command line.

You can also switch on the **sequence** view by clicking "**S**" at the bottom right of the PyMOL main window. There, you can see the aminoacid sequence of the proteins in the PDB and the **chains** ID (usually letters A, B, C etc) that corresponds to the different proteins in the PDB under investigation.

LH3 is a dimer, and by loading 6fxt we get only the monomer. This is not good for our purposes, so reinitialize the program by typing:
> reinitialize

Then, to load the so called "**biological assembly**" and therefore the protein as a dimer, you will need to type this command:
fetch 6fxk, type=pdb1, multiplex=1, async=0
You will see that 2 chains will appear.

Rename the two PDB with a more "standard" name by typing:
> set_name 6fxk_0001, LH3_1
> set_name 6fxk_0002, LH3_2

Let's remove water molecules (the red dots all around the protein) by typing
> remove resn HOH
> remove resn GOL

Then, let's make some **selections** that will help us in structure depiction.
As you can see, the protein is composed by three domains, an N-terminal glucosyltrasferase (GT) domain, a central accessory domain (AC) and a C-terminal lysyl hydroxylase (LH) domain. Select and name each of them by typing:
> select GT, resi 33-281 and LH3_1
> select AC, resi 282-526 and LH3_1
> select LH, resi 527-738 and LH3_1

You can see that in the object panel appeared the buttons with names corresponding to the selection we made. If you click on them, you will automatically select the corresponding part.

By showing the sequence with the "S" button at the bottom right you can display the proteins sequence. Usually at its end you will find the cofactors or ligands or inhibitors.
Select all the cofactors (2OG, Fe^{2+}) cofactors by typing in the command line:
> select FE, resn FE2 and LH3_1

> select 2OG, resn AKG and LH3_1

Let's select also the glycans attached to the enzyme (it's glycosylated):

> select glycans, resn NAG and LH3_1

Now let's put some **colors** (you can find all the colors by clicking "C" in the object pane)

> color marine, GT

> color brightorange, AC

> color splitpea, LH

> color aquamarine, 2OG

> color black, glycans

NOTE1: if you use the "tab" button on your keyboard, PyMOL will "suggest" possible autocompile options.

NOTE2: when you color atoms, also the sequence gets colored in the same way

NOTE3: if you click the "up" arrow when you are in the command line, the previous command will show-up

As you can see, the cofactors are show as sticks or spheres. They are also colored as we decided, but we also want to show atom **color depending on their chemical nature**.

If you want to color atom in this way you can use the color button in the object pane (C→by element) or type:

> util.cnc

As you can see, C atoms remained yellow, O atoms are red, N atoms are blue, P atoms are orange etc.

Now color the second monomer in gray80 by typing

> color gray80, LH3_2

To avoid loss of work, I suggest you to save the session. Go to "File>Save session as" and save it.

If you want to export the image, we will need a white background and high-resolution rendering. Type

> bg_color white

Now, center the image with the mouse wheel (click), find the right "slab" (rotate the wheel) find a proper orientation of the image and render the image (with transparent background) by typing
>ray

NOTE1: rendering may take a while, especially for surface representation!

NOTE2: Don't touch the image anymore, otherwise you will need to render it again.

To save the image go to **File > Export image as > PNG**. Call it "Figure1".

5. Structure analysis

EXERCISE 1

The 6fxk structure is an apo structure although it retains the iron and 2OG. Now let's see how the protein behaves if cofactors of the glycosyltransferase site are added.

1. Open by fetching the 6fxt structure, which is the structure in complex with UDP and Mn^{2+}
2. Superpose the 6fxt with the LH3_1 monomer (super 6fxt, LH3_1)
3. Have a look at the GT (blue) domain, can you spot any difference? Help yourself by switching on and off the two structures.
4. Make a selection and call it "glycoloop" (resi 72-87)
5. Color the 6fxt in gray80 and the glycoloop in magenta
6. Select and color also the cofactors

> select MN, resn MN

> select UDP, resn UDP

> color yellow, UDP

> util.cnc

> set sphere_scale, 0.5

7. Render and save the image

EXERCISE 2

Now let's have a closer look to the glycosyltransferase active site, where the UDP is bound to the protein.

1. Center on the UDP and select a good "slab" in which you can see the active site
2. How is the UDP coordinated in the active site?
3. **Show as sticks the most important residues** involved in the coordination of the cofactors. (suggestion: show as lines all the residues, select the important one, show

6

the selection as stick, rename the selection, hide all the lines). You may want to color them in a different color, or just by chemistry.

4. Select a good position and make a high-resolution figure. Call it "**Figure 2".**