

**PyMOL: A Brief Introduction**

****

**Dr. Daniel Feder**

 **daniel.feder@uq.net.au**

**Part 1: An introduction to the PDB and crystal structures**

Navigate to <https://www.rcsb.org/>

* Welcome to the Protein Data Bank (PDB), a global resource for the structure of biological macromolecules *i.e.* DNA, RNA and proteins. In order to appreciate the vastness of available structural data, locate the “PDB History” selection under the “More” tab in the homepage, and scroll down to “Timeline”. What kind of experimental method has the most entries in the PDB? This is the method we will focus on in this workshop.
* Next, Locate “October 2011 Molecule of the Month” and scroll through the various examples of protein structures that were solved and deposited in the PDB throughout history. Do you recognize any of these macromolecules?
* Return to the main page of the PDB and locate the PDB search function at the top of the page. Add a search query for the X-ray crystal structure of the influenza A Matrix protein 2 transmembrane domain. Using the “Refinements” tab to the left, find a way to narrow down your search to yield the four influenza M2 structures published in *J Am Chem Soc* (2018) **140** 15219-15226.
* Click on the PDB code that corresponds to the M2 transmembrane domain in complex with Amantadine and open the page for this entry. In this panel, there is an abundance of information about the structure you are about to view. Scroll through the tabs from “Structure Summary” to “Experiment”. In “Experiment”, you can view the crystallization conditions, X-ray source and methods used to collect and process the crystallographic data. The M2 transmembrane domain forms a membrane-spanning homotetramer, can you identify the building block for this tetramer using the resources outlined so far?
* Now, familiarize yourself with some new terms. By definition, a crystal is a solid with constituents that are arranged in a highly ordered and repeating fashion- extending in all directions and forming a *crystal lattice.* The smallest repeating molecular unit in a *crystal structure* is set by the crystallographer and is called the *asymmetric unit.* The asymmetric unit lies in part of a repeating geometric arrangement, termed the *unit cell*. The unit cell defines the simplest repeating geometric arrangement of molecules in the crystal lattice from which an entire crystal can be built by repetition in three dimensions. Once the geometric parameters of the unit cell and the contents of the asymmetric unit for a crystal are set, computer software is used to repeat these parameters *ad infinitum* to generate a crystallographic model. The model is then refined and edited by the crystallographer until simulated structural data is in agreement with experimental data. At this point, the model may be published and/or deposited to the PDB.



In the figure above, imagine that the *b* axis (going into and out of the page) is very small, giving an almost two-dimensional rectangle. Can you identify the unit cell and the asymmetric unit? Could there be more than one possible asymmetric unit? What symmetry operators govern the two molecules above? Use simple words to describe this.

* Going back to our M2 entry, to what resolution was this crystal structure solved? In structure determination studies, the resolution (in Ångström units; 1 Å= 0.1 nm) corresponds to the distance of the smallest observable feature, for example: if two objects are closer than this distance, they are depicted as one combined feature as opposed to two distinct, separate features. For a short but powerful illustration of this concept, navigate to <https://proteopedia.org/wiki/index.php/Resolution> and watch the video uploaded by Joel Sussman that demonstrates how an electron density map changes as a function of the resolution of the structure. What kind of atomic features should one be able to observe at resolutions under 1 Å?
* Next, click on “Download Files” and choose PDB format. Open PyMOL and navigate to File🡪Open, locate and open your .pdb file. At the bottom right of your screen, locate the S button and click on it, this will enable the PyMOL sequence viewer above the display and allow us to make atom/residue selections directly from the sequence of the protein.

**Part 2: A brief introduction to PyMOL**

* In PyMOL, navigate to “Display”🡪”Background” and select “White”. Now, hold the left click of your mouse to rotate the model and the right click to zoom in/out, use the middle (scroll) button to drag the display to another center of rotation. In order to select residues in PyMOL we must first left-click on them, either directly on the display, from the sequence viewer or from the selection panel at the right-hand side of the screen. Clicking on any feature of this structure will highlight a selection in the sequence viewer above and in the selection panel to the right. Left-click on one of the two large spheres present in the structure, can you identify the type of atom bound to the M2 transporter here?
* Next, let’s familiarize ourselves with some basic functions in the selection panel. Select the two ions you identified in the last step, notice how “sele” becomes activated in the selection panel. Click on “C” in the selection panel and choose a color for these ions, then deactivate this selection by left-clicking on “sele” again. “sele” changes every time you click on something- but if you wish to keep a selection defined, click on “A” for the corresponding selection and choose “duplicate”. This will create a separate entry for your selection in the selection panel and “sele” can then be used for other purposes.
* Now, find the acetate ions (termed ACE) either in the model or from the sequence viewer, and use the “H” function in the selection panel to hide these molecules. Locate the atoms that correspond to water molecules (in red) and find a convenient way to highlight all of them- then use the selection panel (or right click) to show them as non-bonded spheres (nb\_spheres). Next, find one of the Amantadine molecules in one of the two transporters in the asymmetric unit and set the center of rotation upon it, then zoom in on it.
* Find a suitable color for the carbons of Amantadine, ideally this should be a color that is in contrast with existing colors in the model. Right-click on Amantadine and color it by element. Notice that the color of the non-carbon atoms remains similar, but the color of the carbon atoms is changed. Alternatively, you may choose to color all of Amantadine with a single color. Now, find a way to color the protein model “by chain”.
* Locate residue Ser31 in the sequence viewer for each helix that forms the tetramer around your Amantadine molecule and display it by showing sticks. This residue tetrad allows for the formation of a shapely interface that is capable of binding Amantadine, which is an old anti-influenza drug that is no longer in clinical use. In order to get a better grasp of this shapely fit, highlight three of the four tetrameric helices, and show their surface using the tools you’ve learned so far. Hide the helix that does not form part of the surface, find a suitable color for your surface and scroll with the middle mouse button to toggle the fog level around the model. Go to “Setting”🡪”Transparency”🡪”Surface” and pick a transparency level for your surface. In the command line above the sequence viewer, type: set specular, x (where x is a number from 1 to 10), this will include specular lighting around the model.
* Find an orientation (and/or set of colors of your choice) that best illustrates the binding of Amantadine in the pocket of the M2 transporter, hit “Draw/Ray” at the top-right of your PyMOL window, select 300 DPI, untick “transparent background” and hit “Ray (slow)”. This will generate a highly rendered publication (or thesis)-quality figure. Displaying molecular information with the right objects, colors and a suitable orientation could mean the difference between an accepted or rejected manuscript for publication, an excellent or O.K. thesis or good/bad marks in an assignment. Thus, molecular visualization and presentation is both a science and an art in itself.

**Part 3: Homology modeling and protein engineering**

* Navigate back to the PDB and find the S31N mutant structure of the influenza A virus M2 transporter in two distinct conformational states. The S31N is a mutant whereby Ser31 (which we modeled earlier) is substituted with an Asn31 residue, resulting in an Asn tetrad at the Amantadine binding site. Download the PDB file for the S31N structure and open it in PyMOL. Now, remove one M2 transporter and all of its associated solvent molecules/ligands from the asymmetric unit for the wild-type (WT) structure, *i.e.* that which has Ser at position 31. To achieve this, make a large selection by holding shift and the left-click while dragging, and click remove (using right-click and “remove”). Navigate back to the mutant (S31N) structure by middle-clicking on one of its residues in the sequence viewer, notice that here there are two M2 transporters in two different conformations within the asymmetric unit. Identify the open-state transporter and remove it and all of its associated molecules, leaving behind the closed-state transporter.
* In the command prompt, align the WT and S31N structures by typing “align 6mjh, 6bkk” and note the RMSD value output in the black window above. The RMSD represents the root-mean-squared-deviation of backbone atomic positions, which is a measure of the average distance between atoms. The lower this value, the more similar the conformation of the structures. Navigate to the two aligned transporters and display sticks for relevant residues surrounding Amantadine. Identify the residues in the mutant structure that clash with Amantadine, *i.e.* residues that are too close to Amantadine, you may also utilize a surface to identify these residues. You have now identified a mechanism of viral resistance to a drug, the next step is to make a drug that would bind to both structures! But that is somewhat beyond the scope of this workshop…
* Navigate to “Wizard”🡪”Mutagenesis”🡪”Protein”, on the bottom right-hand side of the screen, click on “No Mutation” and select an amino acid substitution that you think would restore the ability of Amantadine to bind in the same site, click on the residue to mutate and hit “Apply”. You have now hypothetically engineered a protein to serve a biological purpose.
* Next, Navigate back to the PDB and download the crystal structure of the open M2 transporter in complex with Amantadine published in *Nature* (2008) **451**: 596-600, open it and align it with the closed WT transporter, note the RMSD. Is the binding-mode of Amantadine similar between the open and closed channels? Is there anything notably different in how Amantadine binds when the channel opens? How is this drug able to bind to both open and closed transporters?
* Delete the entry 6MJH from the workspace using the drop-down menu under “A” and re-open it in PyMOL. This time, remove the closed-state transporter in the asymmetric unit, align the open-state S31N transporter with the open-state WT transporter in complex with Amantadine and show sticks for both complexes. Why is Amantadine unable to bind to the open-state S31N transporter? Rationalize your answer, and as you did previously, design a mutant that would restore Amantadine’s ability to bind in the same site for the open transporter.

**Part 4: Mechanistic studies**

#### High resolution structures often enable mechanistic studies of proteins due to the depiction of very precise atomic coordinates, sometimes down to the single hydrogen level. Furthermore, the higher the resolution- the more waters (per residue) in the model. In the PDB, locate the Influenza A M2 wild type TM domain at high pH in the lipidic cubic phase under cryo diffraction conditions solved to 1.1 Å resolution, this time- download “Biological Assembly 1”, extract the .gz file and open the .pdb1 file in PyMOL. In the command line, type “split\_states 4qk7”. Show water molecules as spheres.

* Navigate to water number 223 in the sequence, and center on this water molecule, toggle the fog level and rotate the picture around this water molecule until you notice a spiral geometry about the waters that emerge from the opening of the M2 pore. Choose your preferred colors, viewing orientation, and raytrace this image. Then change the size of the water molecules by typing: set sphere\_scale=0.1 in the command prompt.
* This geometry about the water molecules allows protons to “bounce” from water to water effectively until they reach one of the His37 residues, which then conveys these protons to the inside of the virus. Once enough protons enter the virus, this results in unpacking of the viral particle and release of viral ribonucleoproteins into the endosome of the host cell, and is a vital part of the virus’ replication cycle. Use “Wizard”🡪”Measurement” and draw a “water wire” of hydrogen bonds from the center of the pore of the transporter to one of the His37 residues and from the His37 residue to the waters on the “inside” of the transporter.

**Part 5: Use the above tools to model a protein structure and/or process of your choice**