

Molecular Graphics Using PyMOL

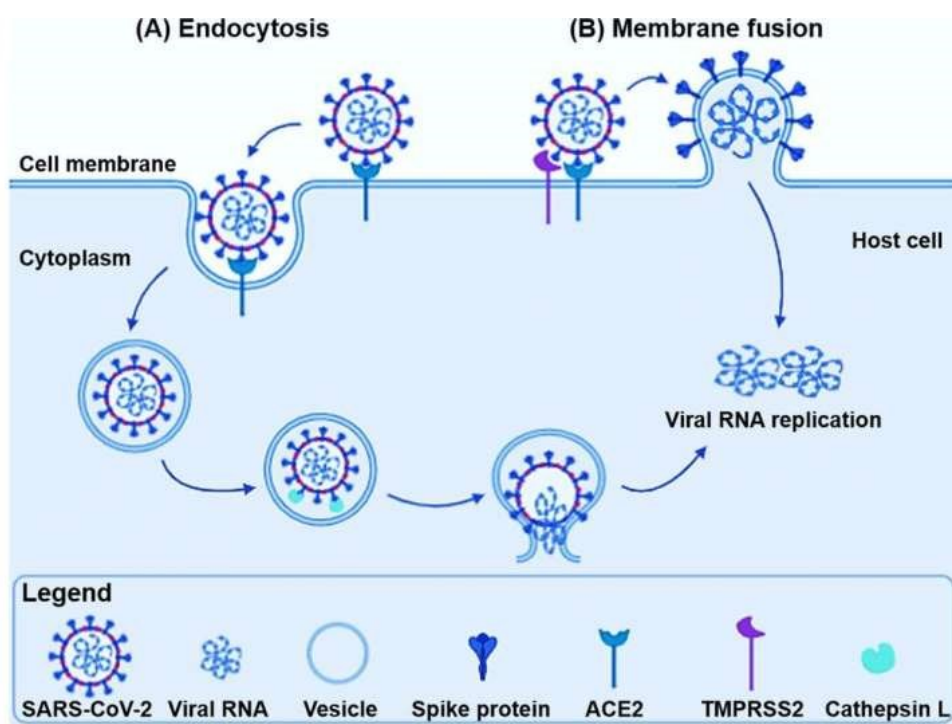
This tutorial was initially prepared by Fasih Rehman.

Introduction:

Part 1: Background on the Coronaviruses

The coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) emerged in December 2019. The rapid spread of the disease and its fatality rate led to the ongoing global pandemic and made it a research priority to scientists all over the world [1]. Genome sequence analysis has revealed that SARS-CoV-2 belongs to the coronaviridae family with its sequence being closely related to the severe acute respiratory syndrome coronavirus (SARS-CoV), which caused the 2002–2004 SARS outbreak.

SARS-CoV-2 virus invades the target host cells via its spike glycoprotein. The spike protein is composed of two subunits, S1 and S2. The S1 facilitates viral attachment to the host cell receptor, angiotensin-converting enzyme 2 (ACE2), while the S2 enables viral-membrane fusion, which allows the virus to enter the host [2]. The spike protein is the major focus to develop neutralizing antibodies and mRNA vaccines.



The current mRNA vaccines for SARS-CoV-2 allow our body to utilize the mRNA to synthesize the spike protein domain responsible for ACE2 binding shortly after immunization. Expression of the spike protein in our cells initiates the process of protective antibody and T cell production [3]. A very

concerning feature of the SARS-CoV-2 spike protein is its ability to change over time as the virus evolves. Mutation of the protein can result in a new variant of the virus that is more transmissible or infectious.

Part 2: Molecular Graphics

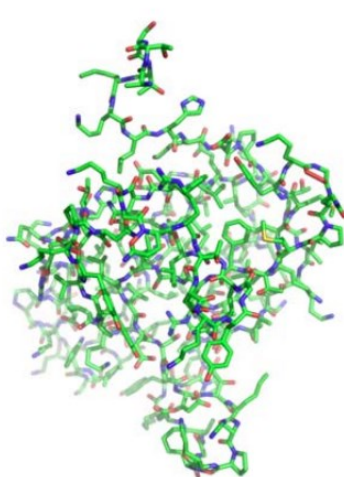
Molecular graphics is the study of molecules through the visualization of the molecules and their components. It depicts the three-dimensional structures of molecules and their conformational changes upon interaction with ligands or in comparison with homologous molecules. Molecular graphics is used as a key component in many scientific studies, especially in structural-based drug design. For this lab, we'll use PyMOL to visualize and analyze protein structures. PyMOL is a powerful tool to study proteins, DNA, and other biological molecules with a structural model. It is the most widely used tool to prepare high-quality molecular graphics for publication.

Many of the techniques we will learn are explored in greater detail in *PyMOL User's Guide*. Even though it is somewhat dated, the User's Guide has very useful information and is worth reading if you are looking to learn more. You can download the guide at <http://pymol.sourceforge.net/newman/userman.pdf>.

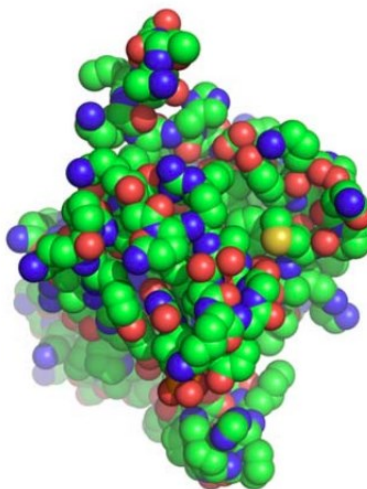
There have been multiple schemes that have been developed over time by scientists to allow for the visualization of proteins and nucleic acids. Some examples are:

- **Sticks:** These are similar to a "line" representation, but they are thicker, like the molecular models used in organic chemistry.
- **Spheres:** In this representation, all (or selected) atoms are drawn as spheres, with radii that are characteristic of their electron orbitals. This is also called CPK representation, after Corey, Pauling, and Branson, the originators of such models.
- **Cartoons:** In this representation, the side-chain atoms are ignored, and smooth line is drawn through the backbone alone. Alpha helices and beta strands are drawn as coils and arrows, respectively.

These representations in PyMOL look something like the picture shown below:



Sticks



Spheres (CPK)



Cartoons

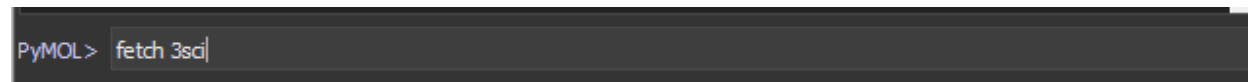
Part 3: Get familiar with PyMOL

Launch PyMOL by searching “PyMol” in the Linux Mint search bar and clicking on “PyMOL Molecular Graphics System” or launch it by typing “pymol” in a terminal.

When you launch PyMOL, your screen should look similar to this:



In the command line on PyMOL, type “fetch 3sci” to have PyMOL download the crystal structure of the spike protein receptor-binding domain from a SARS coronavirus strain (2003). Similarly, you can load just about any other structure that is on the Protein Data Base if you know the 4-digit PDB ID.



The protein should now be on your screen.

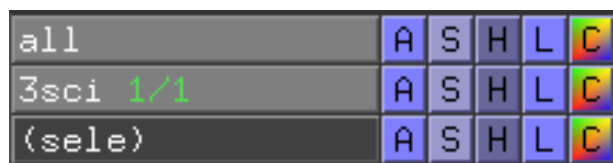
Using your mouse, the following actions can be completed:

- **Rotation:** Clicking on the background and moving the mouse will rotate the molecule around the X and Y axes (clicking at the edge of the screen will rotate around Z axis).
- **Selection:** Clicking on the molecule will select residues (or chains, or atoms, depending on the current mode, see below). Holding "Shift" while left click mouse button and moving the mouse will draw a rectangle and select all atoms (or residues, chains, depending on the current mode) in the box.
- The mode that you are in can be viewed and changed in the lower right-hand corner of the window. By default, you will be in mouse mode "3-Button Viewing". If you click on that (either the text that says Mouse Mode or 3-Button Viewing) you will enter "3-Button Editing" mode. You can also change the selection mode (possible options are: Objects, Segments, Chains, Molecules, Residues, Atoms, and C-alpha atoms).
- **Zoom:** Right-click and hold the mouse button, then move the mouse, will zoom in or out on the molecule (move the mouse up and down).
- **Move:** Click-hold the middle button/wheel and move the mouse will allow you to move the structure sideways (translation).
- **Slicing:** Rolling the mouse wheel will slice away Z-planes (you'll see only a slice of the protein), or change the display depth along the Z-axis, a good way to focus only on some details of the protein.
- **Center:** Double click the mouse wheel when the pointer is pointed at an atom (or residue) can be used to move the residue to the center of the display.

The **controls** can also be viewed on the bottom right of PyMOL:



Objects manager: In the right panel you will see two or more objects: *all*, one or more *<pdb code>* (i.e., *3SCI*), and possibly some selections '*(sele)*'. You can click on these buttons to display or hide molecules without deleting them (light grey buttons are displayed, dark grey buttons are hidden).



Next to them there are five buttons: **A, S, H, L, C**

- **A** stands for *action*; you can zoom, center, rename, delete, and duplicate objects from this menu.
- **S** stands for *show*; here you can change the representation of the molecules: e.g., line, cartoon, or stick.
- **H** stands for *hide*; it performs the same functions as S except that it hides whatever selection you make.
- **L** stands for *label*.
- **C** stands for *color*: here you change to a specific color, color by chain, by secondary structure, or color by rainbow. Coloring by spectrum/rainbow highlights visually the N- and C-termini of the protein. If you use color by chain/rainbow each chain will have its own rainbow spectrum and it will be easy to see where they start and end.

PyMol also offers an internal summary of all the commands derived from the manual. These can be called from the line command `help xxx` where xxx is the name of the command, e.g., `help align` or `help color`.

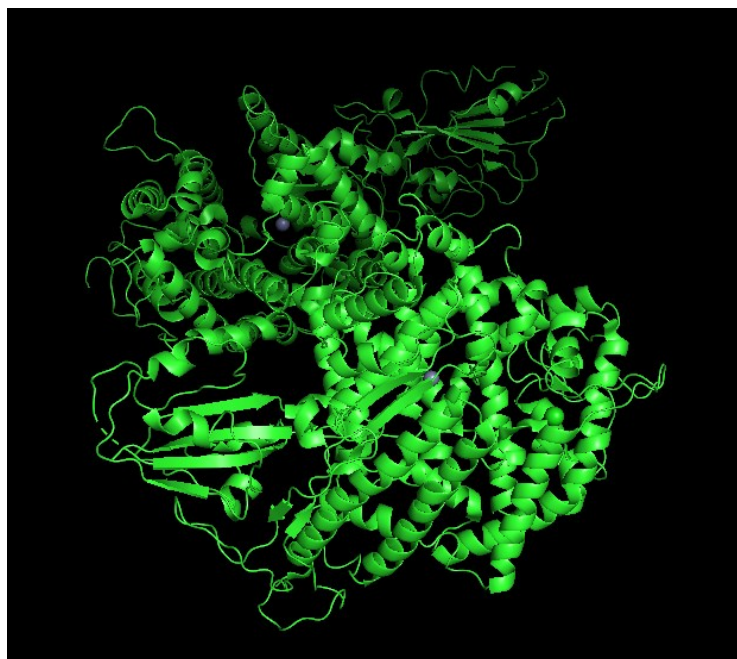
Part 4: Looking at basic operations that can be done with PyMOL

Sequence Viewer:

We will now try to visualize the protein by cartoons.

Go to **H** → everything. This will hide all the representations and allow you to start from scratch.

Go to **S** → Cartoon on the 3sci line. You should now see this:

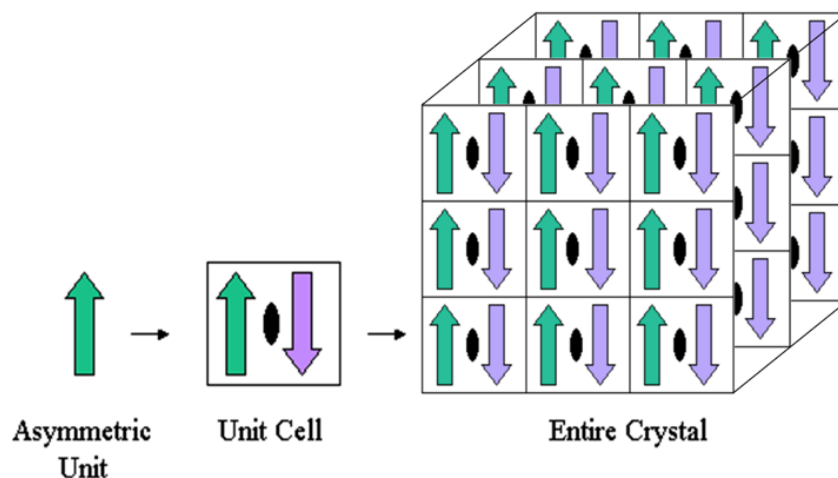


C → by chain → chainbows will colour your structure nicely.

You can also play around with the different colour and show/ hide settings for the protein.

In many PDB entries, you might notice files for “Biological Assembly” and the “Asymmetric Unit”. In many cases, they are the same but for some entries (usually in structures solved by X-ray crystallography) there is a difference.

Asymmetric unit (AU) is the smallest portion of a crystal structure to which symmetry operations can be applied to generate the complete unit cell (the crystal repeating unit). Symmetry operations most common to crystals of biological macromolecules are rotations, translations, and screw axes (combinations of rotation and translation). Asymmetric unit yields one unit cell that when translated in three dimensions, makes up the entire crystal.



The biological assembly (also known as the biological unit) is the macromolecular assembly that is believed to be the functional form of the molecule. Depending on the crystal structure, symmetry operations consisting of rotations, translations or their combinations may need to be performed in order to obtain the complete biological assembly.

In the ASU, there are two copies of the complex. We want to show only one of them so the following steps must be done:

- select ace2, chain A and polymer
- select spike, chain E and polymer

To then look at only one biological assembly, you can click “H” on 3sci on the control panel to hide everything and then “S” on both ace2 and spike and show them as cartoons. This polymer will now be visible, you may change the colours on them to make them more distinct.

Displaying Surfaces:

Hide everything by H → everything. To show use the function S → surface to display the solvent-accessible surface of the protein.

Try colouring by the element, pick your favourite combination for your figure.

You can use the following commands in the PyMOL command line to select hydrophobic, charged, polar or nonpolar residues:

- select hydrophobes, resn ala+val+ile+leu+phe+met+pro+tyr+trp
- select charged, resn arg+glu+lys+asp+his

You can colour the different selections preferable according to standard colouring procedure (Ex. Hydrophobic/polar - Grey, Neg – Blue, and Pos – red).

Sequence View:

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

Alternatively, you can click on the “S” in the bottom right corner of the control panel, and it will do the same.

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 but you can also change this to the three-letter code, which might be helpful. Do this with the menu cascade: Display > Sequence Mode > Residue Names

You can slide the sequence cursor to see the different residues and clicking on one of them will allow you to visualize it. When you click on a side chain, you can transform it to a thicker stick in the menu: (sele) > S > sticks

Measuring distances:

Distances are measured between two atoms and are expressed in the same unit as the XYZ coordinates within the PDB file: Angstroms ($1\text{\AA} = 10^{-10}\text{ m}$).

As an example, we will measure the distance between two atoms. Make sure you go back to the “sticks-mode” first. Within the top menu select Wizard > Measurement (Note: in older PyMOL versions Measurement was called Distance.) 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 of your choosing that you would like to measure.

Click on the second atom of your choosing. 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

We will measure the distance of the hydrogen bond between ace2 Tyr83 and spike Asn473.

- select ace2inter, byres (spike around 4 and chain A)
- select spikeinter, byres (ace2 around 4 and chain E)
- select interface, ace2inter or spikeinter
- show sticks, ace2inter
- show sticks, spikeinter

We can now colour them from the objects panel and hide everything except for the ace2inter and spikeinter.

If you click on the “A” for the “**interface**” object and scroll down to “find”, you can have PyMOL identify any polar contacts, or residues within certain distance between two chains. And you can determine the distance of the bonds by “S” to show “label”.

Task: Find Tyr83 on the ace2 and the Asn473 on the spike, and measure how close they are.

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 or just simply hide the measurements by clicking on them in the object panel.

Labels:

A label containing the residue type and sequence number (e.g., ALA 36) can be added to a selected residue with the menu cascade (sele) > L > residues.

You can explore the L menu, knowing that “clear” will remove the mess that may occur!

You can also select specific parts of the protein if you wish to do so and then label them.

- To select a chain type in the command box ‘sele chain A’
- To select an amino acid name ‘resn’, type ‘sele resn LEU’
- To select a residue by number (‘resi’) and chain, type ‘sele chain A and resi 74’
- You can also use the ‘or’ operator for complex logic: ‘sele resi 74 or resn LEU’ will select residue number 74 and all the Leu in the protein
- Once you enter a selection, the corresponding residues become highlighted, and you can perform an action with the **ASHLC** buttons next to ‘(sele)’ in the object box.

Electrostatic potentials:

PyMol offers an approximate map nicknamed "charge smoothed" surface representing approximate charge distribution on the protein surface.

In order to generate an electrostatic potential map, you first need to do the following:

A > copy to object > new, this will generate a new object that is the copy of the selection in PyMOL.

Do this for both spike and ace2. Rename both of the new “obj” to appropriate names.

The potential surface is calculated and displayed with the following menu cascade on the Names Panel: (ace2) > A > generate > vacuum electrostatics > protein contact potential (local).

The process creates 3 new entries within the Names Panel:

ace2_e_chg, ace2_e_map, and ace2_e_pot

X_e_chg is an object containing the surface, colored red/white/blue and is an approximate map.

X_e_map is usually not shown (click on name to show) and displays the volume boundaries as e.g., a big cube. X_e_pot is the object representing the color ramp and value at the bottom of the display.

If you do this for both ace2 and spike, you should have two electrostatic potential maps. You can click to disable when done.

Making mutations:

The menu cascade Wizard > Mutagenesis > protein opens a new panel below the Names Panel and above the mouse control reminder. Directions will be prompted with text overlaid on the Viewer: “Pick a residue” and “Select a conformational state, or pick a new residue...”

Steps to mutate one amino acid on a protein structure:

- Open from the menu Wizard > Mutagenesis > Protein
- Click on the residue you wish to mutate.
- Select a conformational state in the new mutagenesis panel menu (bottom right) (options are backbone dependent or independent)
- Click on the No Mutation button and select a new amino acid (e.g., Pro)
- Click Apply
- Repeat process for mutating more residues.
- Click Done when finished with the Mutagenesis Wizard.

Aligning structures:

PyMol will have no problem aligning 2 similar structures. The following are just 3 of the methods that can be used to align structures.

Align: The “align” performs a sequence alignment followed by a structural superposition, and then carries out zero or more cycles of refinement in order to reject structural outliers that are found. Align works okay on proteins with >30% identity sequence similarity.

Super: The “super” command does a sequence-independent structure based dynamic programming alignment, followed by a series of refinement cycles that are intended to improve the fit by eliminating pairing with high relative variability. “super” is more robust than “align” for proteins with low sequence similarity.

CeAlign: The “cealign” command aligns two proteins using the CE (combinatorial extension) algorithm. Robust for proteins with little to no sequence similarity.

PyMol firsts creates a sequence alignment and then tries to align the structures accordingly. If 2 proteins are named struct1 and struct2 within the Names Panel, the simple line command will align them:

```
align struct1, struct2
```

```
super struct1, struct2
```

```
cealign struct1, struct2
```

struct1 and struct2 being the selections in this case

We can look at structures 4Z30 (Crystal structure of the ROQ domain of human Roquin-2) and 4Z31 (Crystal structure of the RC3H2 ROQ domain in complex with stem-loop and double-stranded forms of RNA) to align. Unfortunately, neither of the previous 3 commands would make a good alignment for these structures. We must align the subdomain.

The following command will give you an idea of how alignment works in PyMOL:

```
- align 4z30 and chain A, 4z31 and chain A
```

We now want to align the roq0 and roq1 subdomains:

```
- select roq0, 4z30 and chain A and i. 171-324
- select roq1, 4z31 and chain A and i. 171-324
- align roq0, roq1
```

Note: "i." (with a dot) here is short for residue identifier, more short form commands can be seen in the PyMOL guide.

After aligning the 2 subdomains, you should be able to clearly identify the differences between the original alignment and the new one.

Making pictures from PyMol

For publications, it's good to use a white background instead of black. You can change this by selecting "Display > Background > White" in the external GUI window. Since black is easier to view on a screen, you can change back to black by selecting "Black" from the menu.

When happy with how you have presented your structure or part of the structure, click on Draw/Ray in the upper right corner of the PyMol window and select Ray (slow). This will make your picture less pixelated. You should then be able to save your picture. Upload the saved picture to your OneDrive using a web browser.

References:

1. Fu L, Ye F, Feng Y, Yu F, Wang Q, Wu Y, Zhao C, Sun H, Huang B, Niu P, Song H, Shi Y, Li X, Tan W, Qi J, Gao GF. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat Commun.* 2020 Sep 4;11(1):4417. doi: 10.1038/s41467-020-18233-x. PMID: 32887884; PMCID: PMC7474075.
2. Wu K, Peng G, Wilken M, Geraghty RJ, Li F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem.* 2012 Mar 16;287(12):8904-11. doi: 10.1074/jbc.M111.325803. Epub 2012 Jan 30. PMID: 22291007; PMCID: PMC3308800.
3. Ho TY, Wu SL, Chen JC, Li CC, Hsiang CY. Emodin blocks the SARS coronavirus spike protein and angiotensin-converting enzyme 2 interaction. *Antiviral Res.* 2007 May;74(2):92-101. doi: 10.1016/j.antiviral.2006.04.014. Epub 2006 May 15. PMID: 16730806; PMCID: PMC7114332.