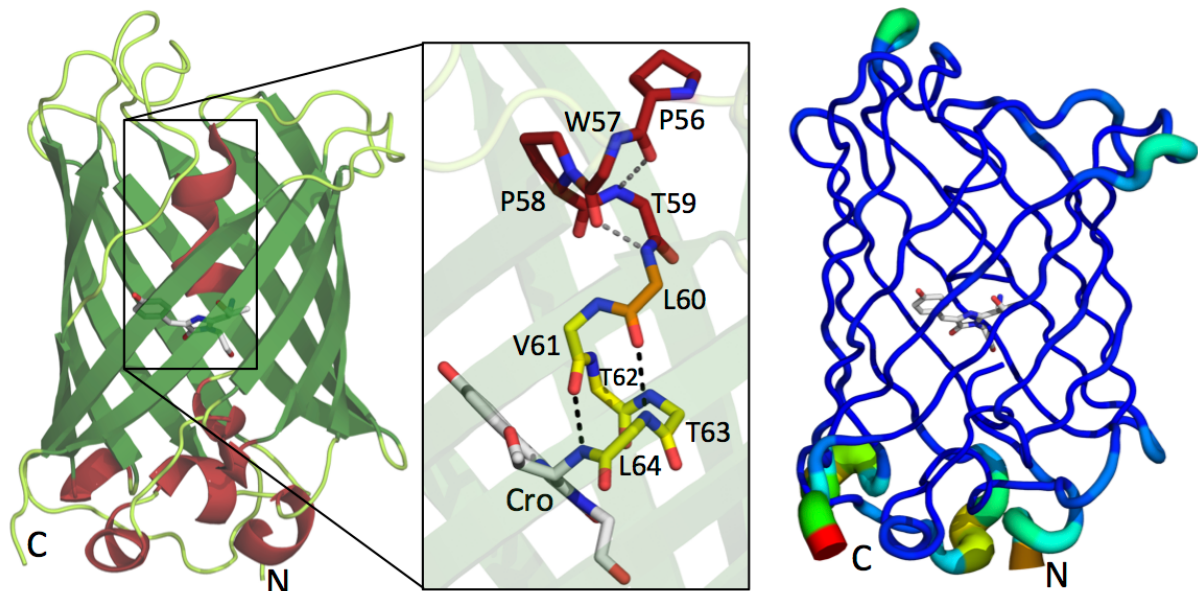


A Simple Tutorial for PyMOL: Visualising Proteins Using Molecular Graphics

by Dr Dafydd Jones
School of Biosciences, Cardiff University



Above figure taken from Arpino, Rizkallah, Jones (2012). PLoS One, e47132.

Overview

This tutorial provides an introduction to the powerful molecular graphics program known as **PyMOL**. This program is now used routinely to generate pretty but useful and informative pictures of protein structures present in many journal publications and in lectures. But PyMOL is more than that – it is a very useful tool to interact with, inspect and analyse the 3D structure of a protein, which is difficult to do in lectures. Although PyMOL has a powerful and flexible interface, it is complex, and can appear daunting to new users. This guide is intended to introduce you to the PyMOL interface and basic tasks. You will first learn how to obtain structure files for viewing, and then carry out exercises in manipulating and analysing protein structures. You can download the program for educational purposes from the sources listed below to use on your own computer. The tutorial is not meant to be exhaustive but get you started using the program and some of its features. The full power of PyMOL comes when you use the command line but we will not be going into that today.

The main thing is just to have a go and experiment. The more you use PyMOL the easier it will be to use and the more you will learn about proteins and the program

PyMOL is available to from the following sources:

Download (available for Windows, Apple Mac and Linux):

We will be using the free Educational version of PyMOL. Please take a note of the restrictions of use for this version of PyMOL (<https://pymol.org/edu/faq.php>)

You will need to register for Educational-Use-Only PyMOL version:

<https://pymol.org/edu/?q=educational/>

You will then get an email through with a username and password to input into:

<http://pymol.org/ep>

There are two components to download and install. The **first** is the program itself. Click on the “**Installers:** PyMOL download page” link. Then download the appropriate version. The **second** is the **license file** and install in PyMOL under “Help” pull down menu and “Install new license file”. Remember the folder you downloaded the license file to.

Further information:

<http://www.pymolwiki.org/index.php/Category:Tutorials>

<http://sourceforge.net/projects/pymol/>

<http://www.youtube.com/watch?v=vDlyfk2zC-k>

1. Getting Started

ADVICE (no charge): Take time to **PLAY** with features of the program. By playing, by trying out various functions, you will discover additional features, and you will develop a personal style of using the program that will make it a more powerful tool in your hands. PyMOL also has a text interface option whereby you can input commands via text descriptions or even whole scripts. We will not go into this approach during the tutorial but if interested, you can look at some of the links above concerning advanced text command options.

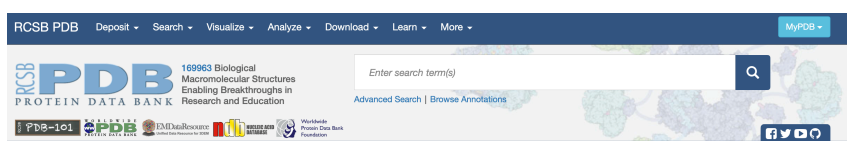
PDB File Names

PDB files are essentially a set of atomic coordinates or “atom placements” read by PyMOL in a defined format to allow a molecular structure to be built. PDB file names consist of a number followed by three characters -- for example, 1HEW. Although it does not follow common convention, in this tutorial capitals are used for the letters of PDB file codes, primarily because many file names contain the number 1 or the letter L, which are identical in most fonts if not capitalized. (The number zero sometimes occurs in PDB file names, so watch out for the subtle difference between 0 [zero] and O [the letter O]. Depending on your browser font, these might be completely indistinguishable.)

You will start learning about PyMOL by looking at the enzyme Hen Egg lysozyme in complex with the trisaccharide inhibitor tri-(N-acetylglucosamine) or tri-NAG (see a biochemistry text if you want to know more about lysozyme). Atomic coordinates for this model are contained in the Protein Data Bank file with code 1HEW.

You can obtain this file from the Protein Data Bank using your web browser by following these instructions:

1. Go to www.rcsb.org and open a new window for the PDB Home Page. You may wish to add this site to your



menu of bookmarks or favourites.

2. A search box with a *Enter search terms(s)* is near the top of the screen, in the middle. Type **1HEW** (the first character is the number one) in the box. Then click the **Site Search (magnifying glass)** button. The **Structure** page for 1HEW appears.

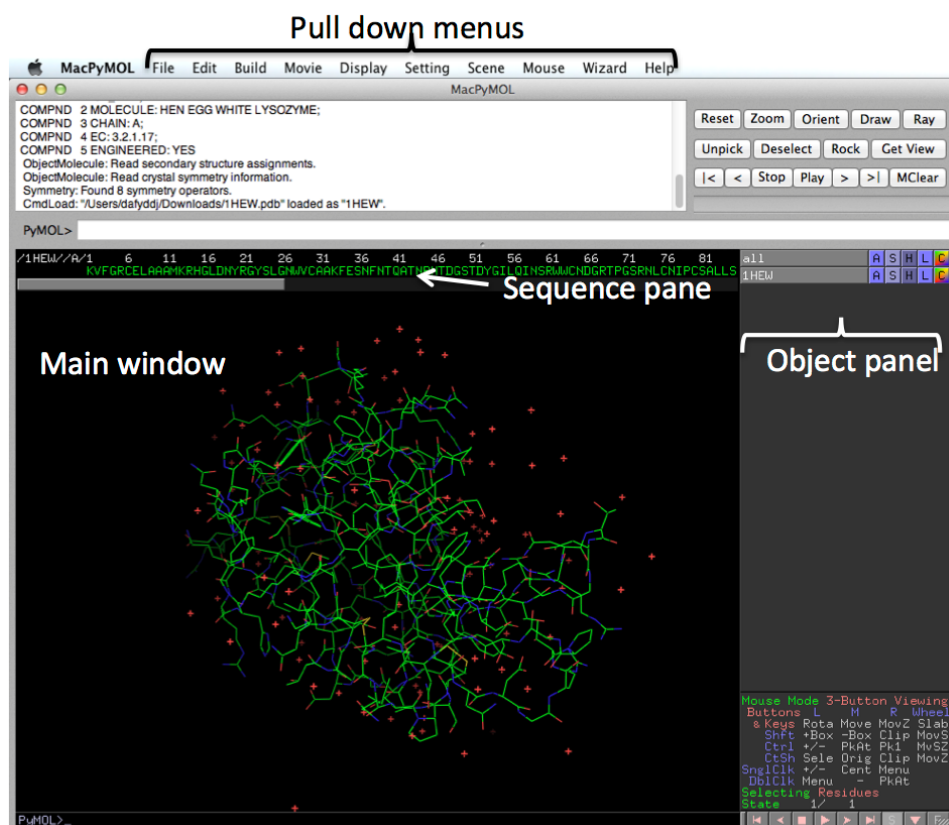
3. At the right side of the page, click **Download Files** to reveal a drop-down list, and then click **PDB format**.

4. Your browser should download the file, perhaps asking you where to save it under the file name **1HEW.pdb**. Remember where you saved the 1HEW.pdb file as you may have to open it from within PyMOL using the **File > Open** option from the pull-down menu.

2. Starting PyMOL and loading files.

Install PyMOL on your own personal computer as outlined above.

See below an overview of the main PyMOL viewing window and the main features to note. The diagrams are taken from the Mac PyMOL version, but the PC version is very similar with a few stylistic changes.



Conventions used in this tutorial

In this tutorial, instructions for giving a command or using a menu appear in consistent formats. Usually, such commands appear at the beginning of a new line or in brackets. Here are some sample instructions and their meanings:

File > Open

Means the **File** tab in the **pull down menu** and select **Open**.

all 'S' > cartoon

refers to the menus in the **object panel** to the right hand side of the main viewing window and means from the object tab **all** chose option (pull down menu) **S** and select the command **cartoon**.

In the object panel on right hand side, selections are different from full objects.

Selections have brackets e.g (**sele**)

Objects, such as PDB files, do not have brackets e.g **1HEW**.

The first steps....

The first step is to load your PDB file: **File > Open** and choose **1HEW**.

HINT. You can download a PDB straight into PyMOL by going to **File>Get PDB...**

You should now be looking at a green, blue and red line model, on a black background. Such a raw representation is of little use as it is difficult to ascertain important regions of a structure. See if you can spot the inhibitor, which is composed of three chair-form rings of N-acetylglucose. It's not easy but as we progress through the tutorial, you will learn how to clearly present information.

Try to rotate and zoom in:

Rotate: press and hold left mouse button and move mouse

Zoom in and out: press right and hold mouse button and move mouse

Mouse wheel: scroll to see the "slab" view of the molecule change. The "slab" view allows you to look inside a protein molecule.

Alternatively look in the lower right hand box of the Object panel regarding mouse-based commands. Here are described additional options and commands available when using the mouse in conjunction with the keyboard.



It is important that you become familiar with using the main command buttons in the **object panel** as these are used to access contextual pull down menus. These are labelled

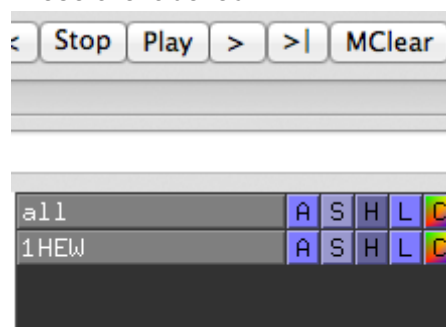
A – Action. Allows you to perform various actions, analyses and calculations

S – Show. Allows you to apply a particular graphical effect to a selection of residues/molecules

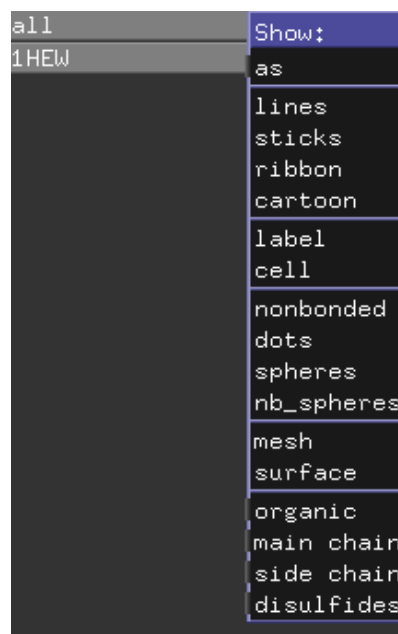
H – Hide. Same as **S** but hides the graphical effect

L – Label. Allows you to add a label.

C – Colour. Alters the colour scheme of particular selection or object.



By using your left mouse button, you can access the pull down menus. Press each of the buttons and see the next level of commands. See below what you get when you press 'A' or 'S'



Try the following

- 1HEW 'S' > cartoon (now see green cartoon representation)
- 1HEW 'H' > lines (the lines now disappear)
- 1HEW 'H' > waters (the red crosses, which are water molecules, disappear)

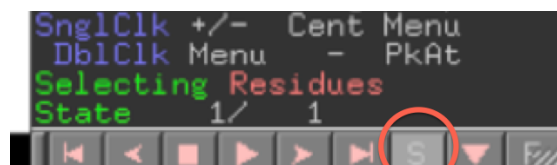
It is important to note that each of the buttons can be applied to various aspects of the pdb(s) files depending the regions/chains/residues that form a particular selection. The defaults tabs are

- 'all': refers to all atoms, including multiple PDB files if present
- '1HEW': refers to all the atoms in a particular PDB file, namely the PDB loaded
- '(sele)': refers to the atoms in a particular selection.

You can also use the individual tabs to toggle the view of your structure. Try pressing the greyed **all** tab. The tab will turn black and the structure will disappear. Press the tab once again and the box will turn grey and your structure will reappear. Remember that these tabs are hierarchical so if you have a particular selection of residues within a PDB file, altering the main PDB file will effect the selections related to that PDB file. We will look at this later.

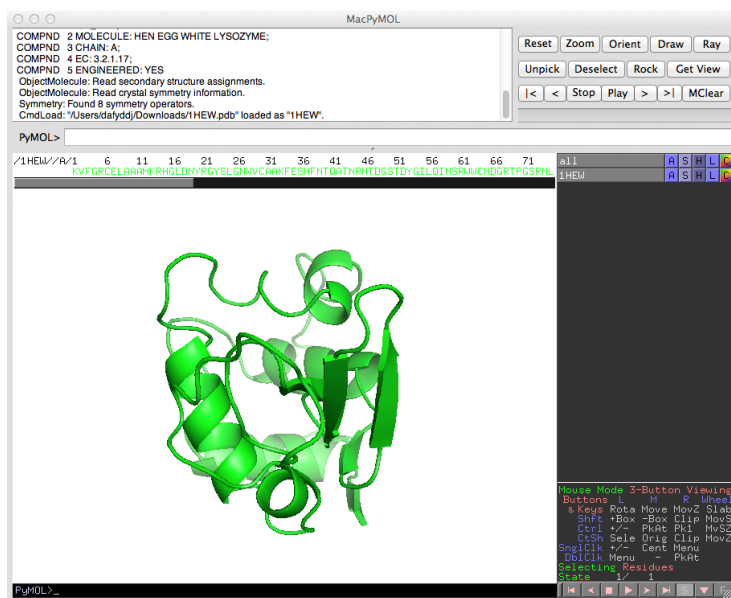
Sometimes it is easier to see the protein on a white background. This is especially important when preparing images for publications. So, to change background to white: **Display > Background > White**

The next step is to display the sequence in the **sequence pane**. To do this press the **S** button at the bottom right of the screen (not the 'S' button described above). Above the main protein structure window, text will appear – this is sequence of the proteins plus other "letters". Using the scroll bar, scroll to the right to see the rest of the protein sequence and the other letters. The **NAG** is the individual units of the trisaccharide



inhibitor and the **red O** are the water molecules, commonly present in structures determined by X-ray crystallography. While waters can play an important role in protein structure and function, we may wish to remove them if required. We can either hide them (see above) or remove them altogether. We will remove them so as we can generate images easily.

all 'A' > remove waters



One very simple way to generate a useful figure for publication is to try the following:

1HEW 'A' > preset > publication .

If you try this, see how the trisaccharide inhibitor becomes visible. To reset back to the previous setting **1HEW 'A' > preset > default** and repeat the steps above.

Important. Saving sessions

So that you don't lose any information/selections etc you have generated I strongly suggest you save you session at regular intervals. To do this, go to **File > Save Session**. This is also useful if you want to save a particular analysis of your protein structure through using **Save Session As...** to generate a new filename. To start the session up at any time in the future, you can go to **File > Open** and chose the file with the **.pse** extension.

3. Selecting and displaying

To understand and convey the essential elements of the protein structure-function relationship, it is important that individual residue/regions are selected and highlighted. There are various defaults open to you. We have already seen one preset (**publication**) above. Try some other possible presentation options for your protein:

1HEW 'C' > by ss (secondary structure colour) select one of the options.

1HEW 'S' > spheres (view atoms according to the van der Waals radius)

1HEW 'S' > surface (view the surface representation of the protein – note how the NAG units look.

At each stage, remember to move your structure around. For understanding and presentation purposes, it is important you select the right viewing aspect.

Defined selections and generating new tabs in the object panel

To select residue/regions, we will interact with the **sequence pane** above the main protein structure. First, change the colour of the whole sequence back to green (**1HEW 'C' > greens > green**) and show only the **cartoon** (**1HEW 'H' > everything** followed by **1HEW 'S' > cartoon**)

Select a residue in the sequence panel, e.g. **cys30** (C before 31). If you cannot see the numbers above the amino acid sequence go to **Display > Sequence Mode > All residue numbers**.

A new tab appears below **1HEW** in the object panel called **(sele)**, the default name for a new selection. **Note** that the selected residue is highlighted in the sequence pane and red dots appear at the position occupied by the residue in the protein structure. You can toggle between selecting and deselecting the residue by pressing the left mouse button while the cursor is over the residue in the sequence pane.

Change the name of **sele** to **Cys30** by choosing **(sele) 'A' > rename selection**. A new text panel appears in the main window. Delete the 2nd sele and replace it with **Cys30**. The selection below **1HEW** is now **(Cys30)**.

NOTE.

*You need to change the default **(sele)** selection if you want to restrict a selection to a certain residue range otherwise any further selections in the sequence pane will by default be added to a **(sele)** tab.*

Also, if the tab is "active" (in other words the tab background is grey) any additional residue selections will be added to that particular tab, whether intended or not. Deselect a tab if you have finished making sequence additions/removals.

*You can also directly select residue/atoms from the structure by clicking on them in the structure itself. Any selection will be added to the selection tab "in use" at that time or a new **(sele)** tab. A successful selection on the structure will be indicated by the appearance of the red and white dots and selection of the residue in the sequence pane. If you want to deselect a residue you can still use sequence pane.*

While making sure the **1HEW** object is highlighted, show **Cys30** as spheres and colour it by atom colour:

Cys30 'S' > Spheres then **(Cys30) 'C' > by element**.

Toggle both the **(Cys30)** and **1HEW** tab in the object panel.

To see only the **Cys30** residue, hide **1HEW** cartoon representation (**1HEW 'H' > cartoon**). All you should see is **Cys30**.

Centre on **Cys30**: **(Cys30) 'A' > center**. Zoom in and move around. The focal point for the structure is now centred on **Cys30**.

To see this clearly within the context of the whole protein structure, show the rest of **1HEW**: **1HEW 'S' > cartoon**. The cartoon representation should now reappear. Repeat the **center** procedure but on the **1HEW** tab and zoom out to see the whole protein. Now hide the spheres of **(Cys30)**: **(Cys30) 'H' > spheres**

Black out the **Cys30** tab – this is important otherwise any additional selections will be added to this panel. Practice selecting residues and renaming the default **(sele)** name. **Exercise:** Highlight a helix (**NOTE** if you colour the protein by secondary structure, the **sequence pane** will change colour to reflect their colour in the structure. Residues in a helix will be coloured red or cyan). You can select (or deselect) an individual residue or a run of residues by keeping your finger on the left mouse button and dragging. Once you make a new selection, notice that a new **(sele)** tab appears under **(Cys30)**.

There are a lot of options available – I suggested you try them out.

Remember: Save your session. Try saving as a separate session after each section.

4. Displaying interactions

How particular residues interact with other residues within a protein or with other proteins or molecules are important in defining both the structure and function. To demonstrate how PyMOL can be used to analyse and illustrate such interactions we are going to look at selecting the trisaccharide inhibitor and show how it interacts with lysozyme. This is an example of how we can glean molecular information on interactions that define functional effects.

One easy way to manipulate the inhibitor segment independent of the protein is to separate it out as a new **object** (as opposed to a residue selection described in section 3). This will allow it to be manipulated independent of **1HEW**.

First, delete any **(sele)** selections in the right hand panel (**(sele) 'A' > delete selection**). Scroll to the right of the sequence panel and select the 3 NAG residues. These should form the new **(sele)** selection. Rename to **Inhib** as described above.

Extract the **(Inhib)** selection to a new object as followed: **(Inhib) 'A' > extract object**. A new **obj01** should be generated below **(Inhib)**. Also, there will be new sequence below **1HEW** in the sequence pane.

Rename the **obj01** to **Inhib** (**obj1 'A' > rename object**).

Let's now look at the inhibitor itself. Colour **1HEW** to a uniform colour of choice (this will also affect **(Cys30)** as it is part of **1HEW**).

Show **Inhib** as **sticks** (**Inhib 'S' > sticks**). Colour according to **element** (there are various options for the colour of carbon but grey is usually the default). Make sure it is different to that chosen for **1HEW** so that the two objects are easily identified. **Inhib 'C' > by element**.

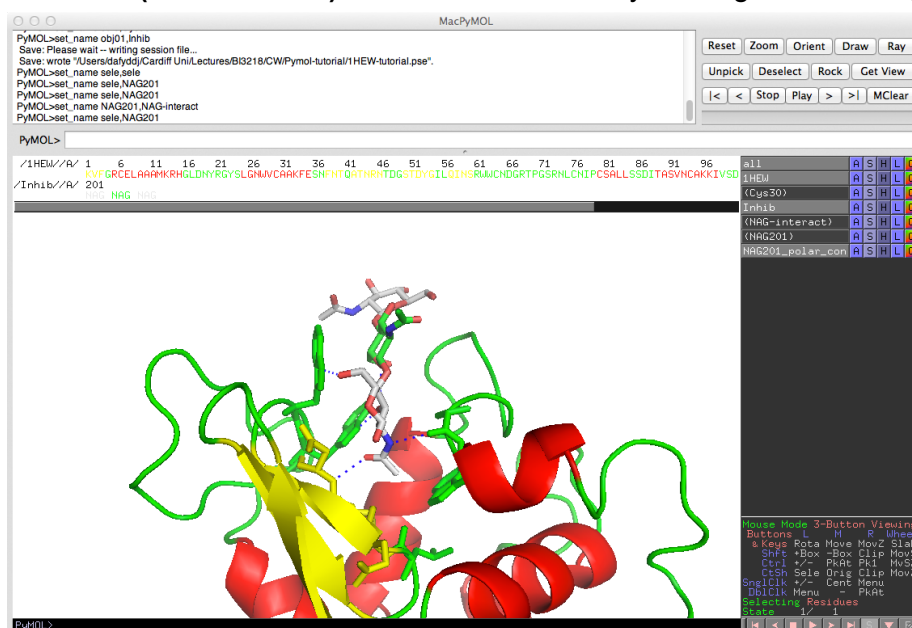
Rotate and move the structure around

We are now going to analyse the inhibitor-protein complex to determine the residues in the protein that interact with a unit of the trisaccharide inhibitor. First select **NAG201** (the first NAG unit) from the sequence pane and rename it **NAG201** (**(sele) 'A' > rename selection**).

Find residues that are likely to make contact with **NAG201** by **NAG201 'A' > modify > expand > by 4 Å, residues**. This will add to the **NAG201** selection, any residues within 4 Å. Within this distance range, these residues are essentially in contact with **NAG201**.

The newly selected residues are added to the original (**NAG201**) selection. Therefore, rename this as **NAG-interact** (**(NAG201) 'A' > rename selection**). You may wish to regenerate (**NAG201**) by generating a new selection with just **NAG201** selected (see above). This is important for calculating interactions (see below)

Show the new (**NAG-interact**) selection as **sticks**. By rotating and zooming, you should get be



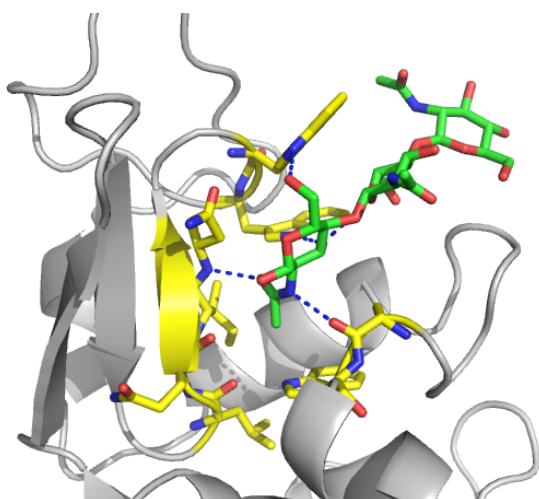
able to improve the view. **Note** that the residues with 4 Å are also displayed in the sequence pane so you can note which residues are forming the interacting interface.

Label the selected residues. **NAG-interact 'L' > Residues**. The residues selected are now labelled on the structure of the protein itself. **Note**. You can change the size and font of the labels through **Setting > Label**.

Now we will select and display polar interactions involving NAG201. Select polar interactions from NAG201: (**NAG201**) '**A**' > **Find > Polar contacts > to any atoms**. A new panel will appear below **NAG201** called **NAG201_polar_con**. In this panel select a colour (e.g. blue) and the interactions will show as dashed lines. The end result of this procedure can be seen below.

Rendering: that publication quality image!

To make a structural image of the required quality suitable for publication, you will need to optimise the view and the residues that interact (e.g. refine your selection) by selecting only those in which polar contacts have been calculated. You can then **RayTrace** (or render) your image to generate a high quality version. For example, to remove the emphasis on the majority of the redundant Lysozyme structure colour **1HEW** grey. To emphasise the residues in lysozyme interacting with the inhibitors, colour elements in the **NAG-interact** tab yellow by carbon atom. Then colour the elements in **Inhib** green by colour carbon atom.



Setting > Rendering.

You can then save the image as it appears by selecting the **Save Image to File** option after the rendering process is complete. It will automatically save the image as PNG format, which is compatible with many common software packages. **NOTE**: DO NOT move the structure prior to

saving or the rendering will be removed from the object. An example of a final ray traced image is shown above.

Advice. The label option in PyMOL is not the best of quality but serves a purpose when undertaking analysis. It is better to annotate your structure by importing the figure into another program (e.g. Powerpoint) and adding the labels. It is essential that important regions/residues of the structure are labelled clearly and unambiguously.

Remember: Save your session. Try saving as a separate session after each section.

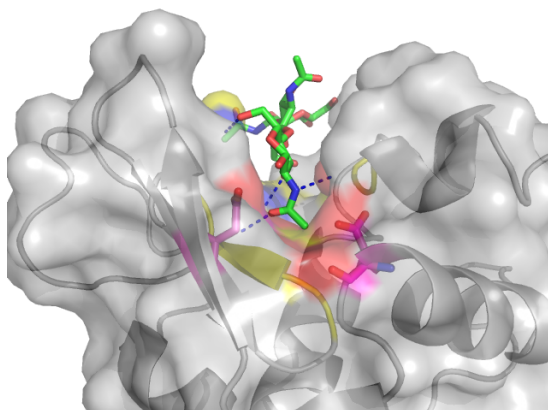
Alternatively, try using the “scene” option. When you have a particular view and selection you want to quickly revisit, select the menu tab **Scene > Store > F1**. F1 will appear in the bottom left corner of main window. Clicking on the F1 will automatically return you to stored configuration. You add additional configurations such as F2, F3 etc.

5. Surface of a protein.

The surface representation of a protein can be quite informative, highlighting the general shape of a protein as might be encountered by a solvent (i.e. water) molecule. PyMOL is very at good calculating surfaces and allows for the surface to be transparent. Using the surface representation, we will see how lysozyme interacts with the inhibitor.

First, go back to the default cartoon for **1HEW** (**1HEW 'S' > Cartoon** followed by **1HEW 'H' > sticks**). Keep the existing colour schemes that differentiate **Inhib** from **NAG-interact** and the rest of **1HEW**.

Add a new **sele** selection tab in the **object panel** by selecting the two catalytic residues, Glu35 (E35) and Asp52 (D52) from the sequence pane.



Rename this selection to **cat-res** ((**sele**) '**A**' > **rename selection**). Show them as **sticks** ((**cat-res**) '**S**' > **sticks**) and colour them by element with the carbon set to magenta (purple) (**cat-res 'C' > colour > by element**).

Under **1HEW**, select surface. You should now see a surface with **Inhib** still as **sticks** and lysozyme now shown as surface, with the different selections as different colours. Rotate and zoom in to look at the groove into which the inhibitor binds. Remember, if you

want to centre on a particular selection, go to one of the selection tabs you generated on the right hand side and select '**A**' > **center**.

You can also make the surface **transparent** to different degrees to see some of the underlying detail. To do this, go to **Setting > Transparency > Surface** in the pull down menu and chose a value. Your selection will be presented on the screen in real time. You can play with the setting for surfaces under the **Setting** menu.

By rendering (using the **Ray** button) the picture as described above, you can generate a smooth publication quality figure of your protein. The one above, clearly highlights key features of lysozyme, namely the binding cavity of enzyme, inhibitor binding, residues/regions involved in binding the inhibitor and the structural relationship of the protein with respect to the active site residues. This provides an essential aid to any description of the protein structure-function relationship.

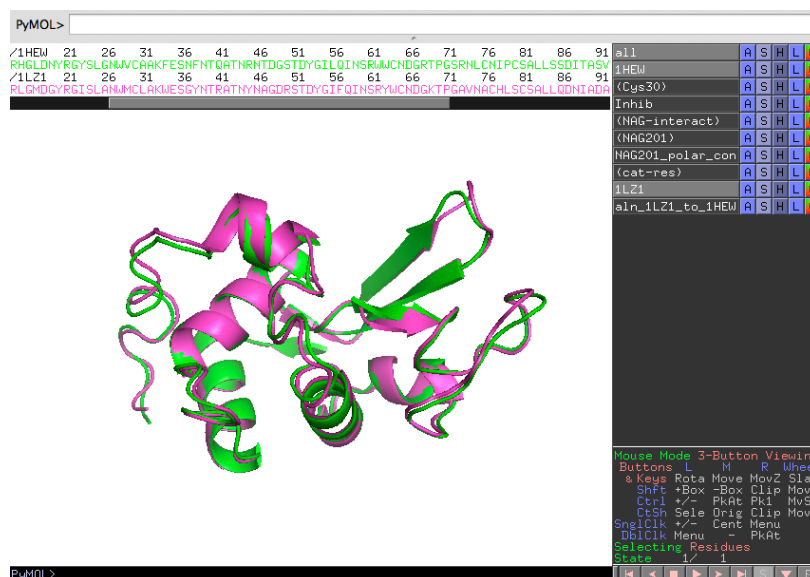
Remember: Save your session. Try saving as a separate session after each section.

6. Comparing and contrasting protein structures: alignment

One of the most important aspects of analysing the protein structure-function relationship is to understand how different amino acid sequences translate to changes in structure and function or the requirement of residue conservation to preserve required functional features (e.g. catalytic mechanism). This is achieved by comparing the structures of structural and functional homologues and/or mutants of a parent protein. We will go through the structural alignment process by comparing 1HEW, the Hen Egg Lysozyme, with the human equivalent 1LZ1.

Set the original **1HEW** molecule back to the green cartoon as described above. You will need to use a combination of the 'H', 'S' and 'C' options.

Download the PDB file **1LZ1** from www.rcsb.org as outlined in section 1. Open within PyMOL (**File > Open > 1LZ1.pdb**). You should now see a new object tab in the **object panel** labelled **1LZ1**. The first thing you will see is that the two structures occupy different regions in the window.



The next step is to align the Human enzyme with the Hen lysozyme. Select **1LZ1 'A' > align > to molecule > 1HEW**. The structures should now overlap and a new tab appears in the object panel (**aln_1LZ1_to_1HEW**), as shown in the image below. The new **aln_1LZ1_to_1HEW** tab is a program generated tab and aids analysis by showing via lines the major differences between the 2 structures.

Center the view on **1HEW** (**1HEW 'A' > center**). Display **1LZ1** as **cartoon** and look at the general similarities between them.

We will now look at 1 set of residues that are similar and one set that differ. **Remember**. There are now 2 sets of protein sequences. They are easily differentiated by the object name (e.g. **/1LZ1**) above the sequence.

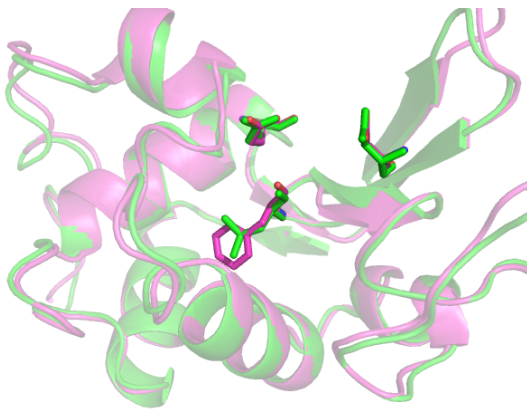
Select the catalytic residues for **1LZ1** (residues E35 and D53) from the sequence pane. Make sure they are added to the (**cat-res**) tab by highlighting that tab (switched to grey from black) before selecting residues.

Show the residues as **sticks** and look at the similarities. You will notice that they are virtually fully superimposable.

Now deselect (**cat-res**) (switched to black from grey) and in the sequence pane select residue 56 in **1HEW** and 57 **1LZ1**. If you remember, residue 56 in **1HEW** comes into close contact with

Inhib. Show both residues as **sticks** and see how they are similar (backbones) and different (side chain).

*Quick quiz. Why does the residue number shift by one for D53 in **1LZ1** compared to **1HEW**? How is this “insertion” incorporated as part of the overall protein fold?*

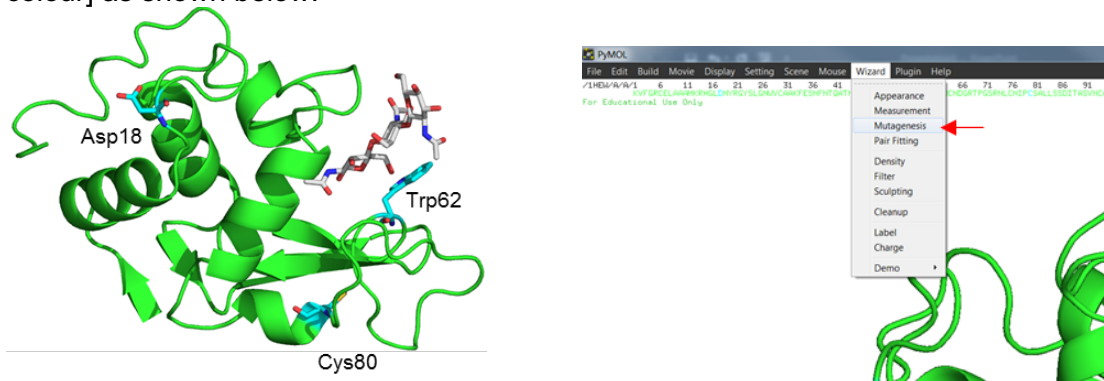


We can now make the **cartoon** transparent to highlight the similarities and differences in these key residues. **Setting > Transparency > Cartoon > .** By rendering (**ray**) the figure, you can now have a publication quality figure concerning key residues involved in function, their position with respect to the whole protein and how sequence differences between homologues generated structure/functional differences.

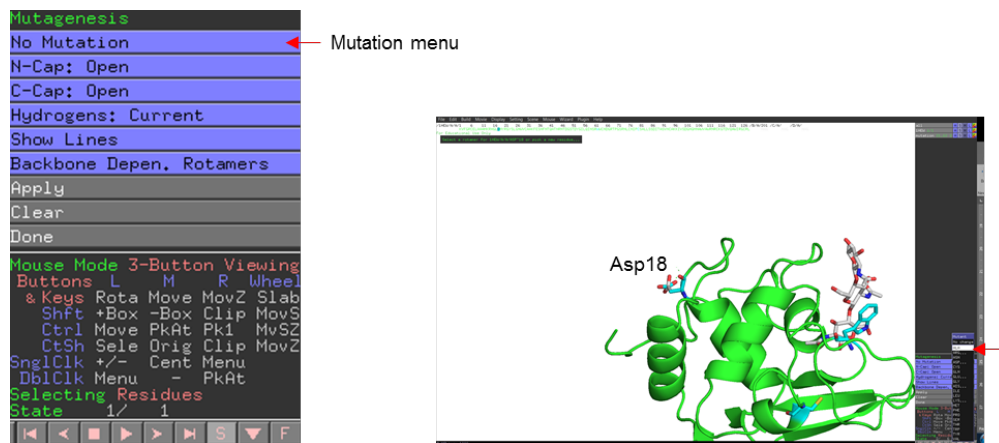
7. *In silico* mutagenesis.

PyMOL also provides with the opportunity to perform mutagenesis in which you can change one amino acid to another and assess the impact on protein structure. The mutagenesis tool can be found under the “Wizard” pull down menu at the top of the screen. Using the 1HEW structure, change the following residues: **Asp 18**, **Trp62**, and **Cys80**. **You may wish to start a fresh session, so save your previous session (File > Save Session As...)** and close then reopen the 1HEW structure. *What impact do you expect each mutation to have on the enzyme’s structure and function?*

First of all, select the three residues. These will be highlighted in the (**sele**) in the object panel. Rename the selection if you like. Show the all three residues as sticks (**sele**) > **S** > **Sticks** and colour differently to the rest of the structure (**sele**) > **C** > **by element** > [choose preferred Carbon colour] as shown below.



Next open the mutagenesis plugin by going to **Wizard > Mutagenesis**. At the bottom of the right hand object panel you will see the Mutagenesis menu. There are several options in this menu, but we will focus on **Mutation**, **Apply**, **Clear** and **Done**. The later 3 buttons are fairly self-explanatory; **Apply** applies the selected mutation, **Clear** deletes the current mutation, and **Done** closes the mutagenesis menu. The mutation menu (first menu) allows you to choose between the 20 canonical amino acids plus selected ionisation states for selected amino acids.

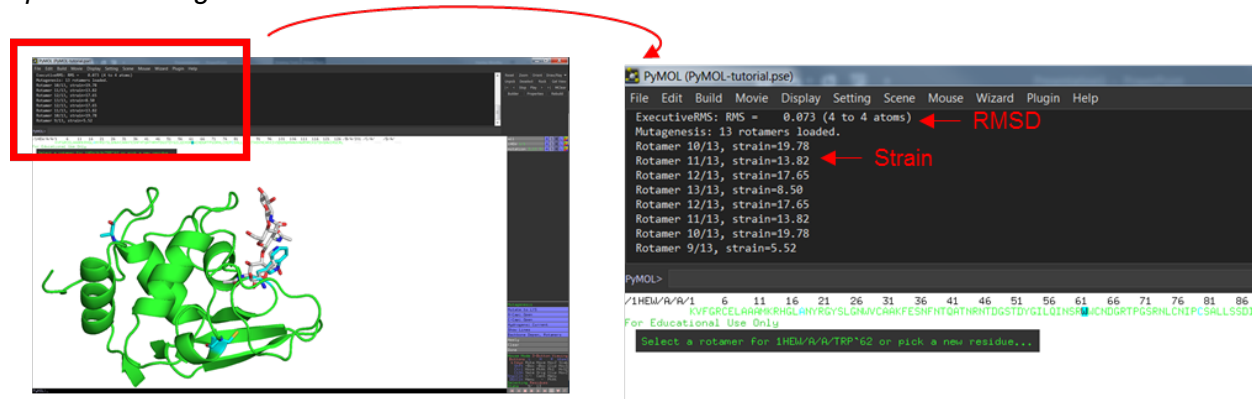


First select Asp18 by clicking on the residue then mutate to alanine by selecting **Mutation > Ala** and press apply.

What effect is any will this have on the protein? Consider the location of the mutation and what properties are changing i.e. charge, surface area, polar/non-polar.

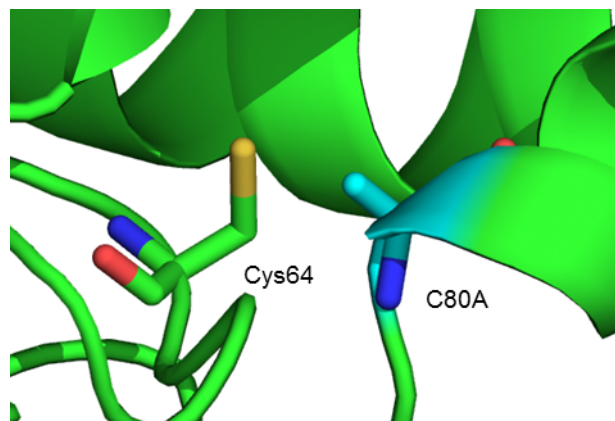
Next select Trp62 and mutate to lysine **Mutation > LYS... > LYS**. By using the left and right arrow keys. Notice the command line prompt as you cycle through the possible rotamers. Once you are happy with the rotamer press **Apply**. Note the red disks show areas of steric clashes.

What do you think 'strain' refers to and how it could be used to determine the appropriate rotamer? What do you think this mutation would have on substrate binding? Why wasn't there an option to change rotarmers for the D18A mutation?



Finally, select Cys80 and mutate it to alanine (**Mutation > ALA**) and click **Apply** then **Done** to close the mutagenesis menu. Select Cys64 and show as sticks (**sele**) > **S** > **Sticks**.

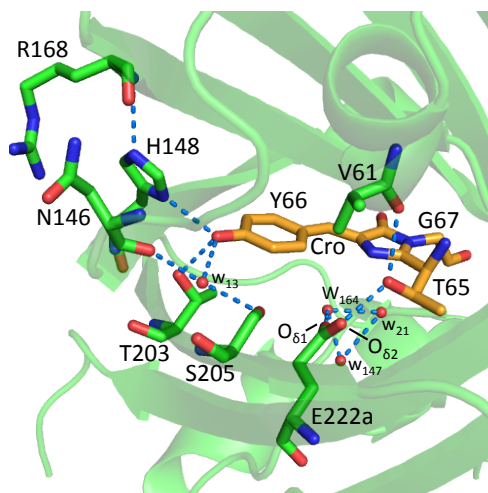
What immediate structural change has occurred? How would this mutation affect the protein structure/function?



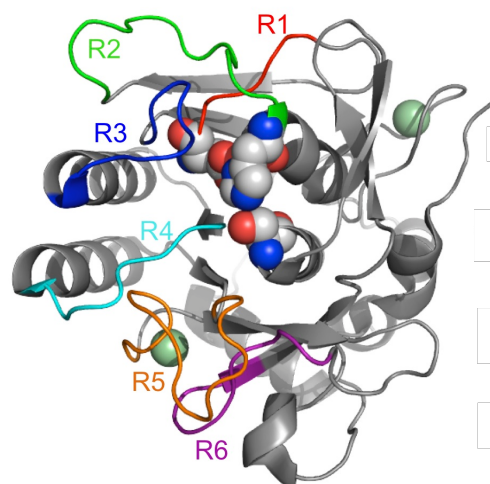
Try reversing the mutation C80A mutation back to cystine (**Wizard > Mutagenesis; Mutation > Cys**; select residue 80). *What happens? Are you able to reverse the structural change?*

This tutorial has provided you, the student, with a basic introduction to using PyMOL and should now allow you to interact with, analyse, compare and contrast protein structure. There are many other features available that you may want to play with. These include the **measurement** tool present in the Wizard pull down menu that allows you to measure distances between two defined points of your choosing. You may want to change some of the settings – feel free to experiment with the **Settings** menu.

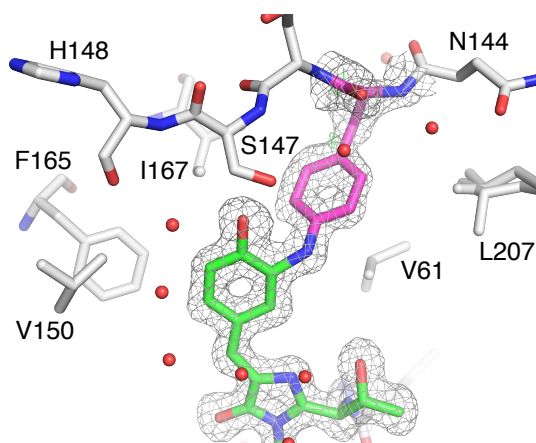
Below are a selection of images that can be generated by PyMOL that have been published in papers originating from my lab.



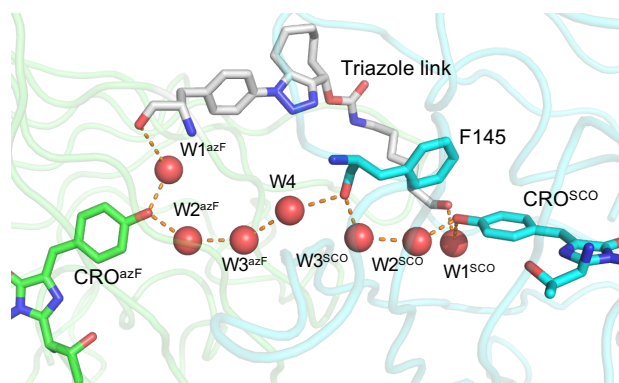
PLoS One (2012):
doi:10.1371/journal.pone.0047132.



PLoS One (2011):
doi:10.1371/journal.pone.0024319



Angewandte Chemie Int Ed. (2013)
DOI: 10.1002/anie.201301490



Communications Chemistry (2019)
DOI: 10.1038/s42004-019-0185-5