Biomolecular visualization with PyMOL

Structural bioinformatics in practice (Færøyene 2016)

Lars Skjærven, University of Bergen

Overview

This tutorial provides a basic introduction to molecular visualisation with PyMOL. It covers aspects of the program useful for structural biology and medicinal chemistry such as exploring protein-ligand interactions, comparing protein structures of different species as well as making figures for use in publications.

The graphical user interface (GUI)

The PyMOL GUI consists of multiple panels as shown in **Figure 1**. The main **display** panel provides the visualization of the loaded molecules while the **object menu** panel on the right hand side provides a menu of options for the loaded molecules. While the object menu provides various operations by clicking the same tasks can be achieved using the **command line** located directly below the display panel. The **upper control window** contains a console (also with a command line) as well ass a set of handy buttons for quick access to various functions. Note also the **mouse** and **movie** controls panel in the lower right corner.

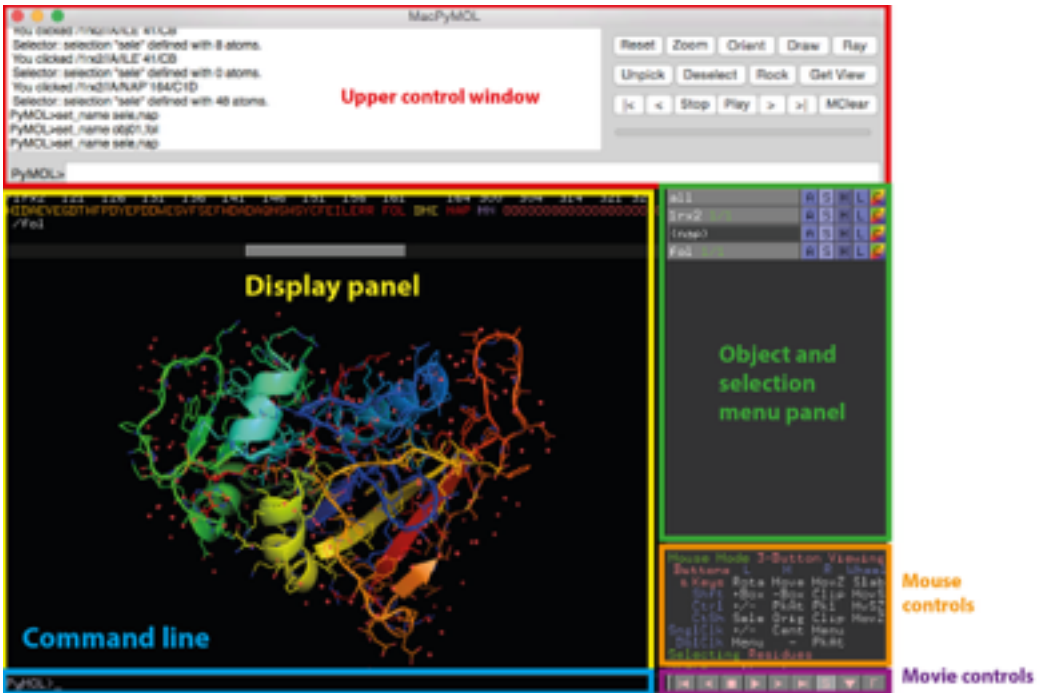
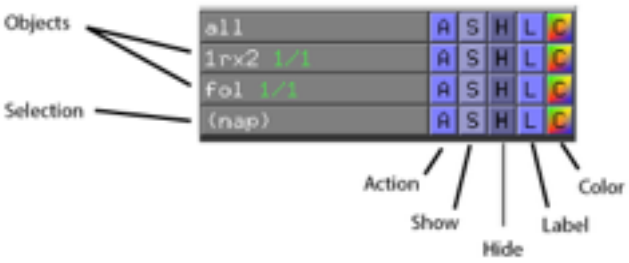


Figure 1: The PyMOL graphical user interface with PDB ID **1rx2** loaded.

Most tasks in PyMOL can be achieved using the object menu on the right hand side of the display panel. Here, each line (or entry) in the menu correspond to the objects you have loaded into PyMOL (e.g. molecules). In the figure below, **1rx2** is one such object, and **fol** represents another object. Note that the entries in the menu in a parenthesis (**(nap)** in the figure) represents an *atom selection* and not an object (important difference).

Next to the object names the menu items are located. These are labelled **A**, **S**, **H**, **L**, **C**. These letters stand for **A**ction, **S**how, **H**ide, **L**abel, and **C**olor. Go ahead and click a letter and observe the menu appearing. We will explore these menus in more detail later.



Molecular representations

PyMOL provides multiple molecular representations for visualization. For a quick overview of these use the main menu and click

Wizard → Demo → Representation

This will display the 8 different molecular representations available in PyMOL. These are lines, sticks, spheres, surface, mesh, dots, ribbon, and cartoon. Of these, the lines, sticks and cartoon representation are probably the most commonly used to display bimolecular structures. Go ahead and select various other demos from the menu on the right hand side (e.g. **Cartoon Ribbons**). Click the **End Demonstration** button when you are done.

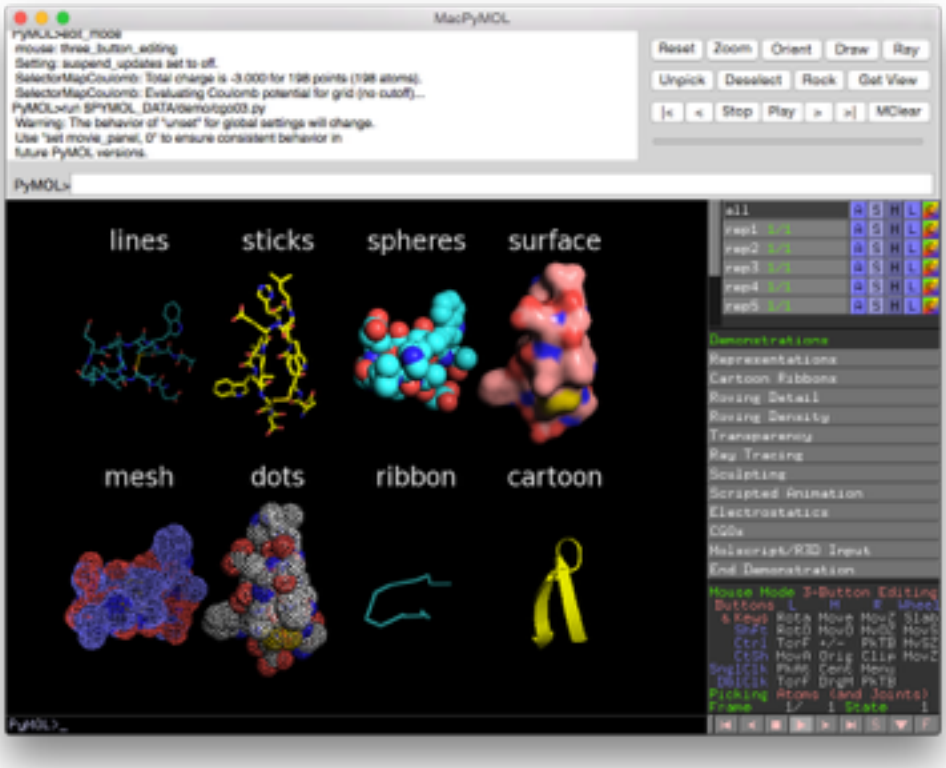


Figure 2: Available Molecule representation in PyMOL.

Exploring *E. coli* dihydrofolate reductase

Loading your first protein structure

In this tutorial we will first explore the structure of *E. coli* dihydrofolate reductase (DHFR) and investigate the molecular interactions with Methotrexate (MTX) – a highly potent drug used in the treatment of cancer and autoimmune diseases.

To load the DHFR PDB structure into PyMOL navigate in the main menu system:

Plugin → **PDB** → [enter the 4 character PDB code **1rg7**]

or use the **fetch** command in the command line:

```
fetch 1rg7
```

Side note: To load a PDB file from your own file system use the menu system

File → **Open** → [Select 1rg7.pdb]

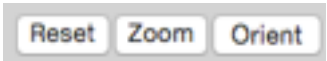
or simply write the following into the command line:

```
load 1rg7.pdb
```

The PyMOL Viewer Window will now display the 3D structure of *E. coli* DHFR in the default line representation colored by atom type. Note that the Object menu panel now contains an entry named **1rg7** representing our newly created object (**Figure 3**).

Familiarize yourself with the mouse controls before we move on:

- **Rotate**: Press and hold the left mouse button while moving the mouse.
- **Zoom**: Press and hold the right mouse button while moving the mouse.
- **Translate**: Press and hold the middle mouse button while moving the mouse.
- **Clip the view**: Scroll.



Click the **Reset** button in the upper control window any time to come back to the initial view, and the **Zoom** button to zoom on all loaded molecules. The **Orient** button will orient the view to display the molecule along its longest axis aligned with the x-axis. (Try it out).

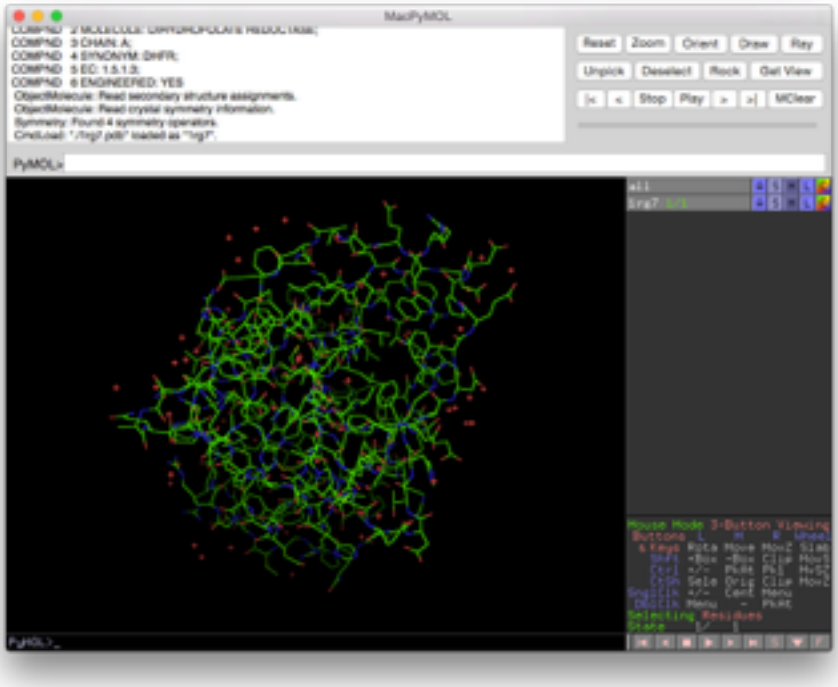


Figure 3: The loaded 3D structure of E. coli DHFR (PDB ID 1rg7) shown in lines representation.

Using the visualization presets

PyMOL comes with multiple built-in visualisation preset views (Table 1). These can be accessed through the Action (A) button(s) in the object panel. Click the A next to the **1rg7** object. From the menu that appears click preset → ligands. The program will now automatically zoom into the ligand region of the protein and change the representation of the protein to ribbon, the residues close to the ligand will show in lines, and the ligand is show as sticks. Also notice the yellow dashed lines appearing between polar groups between the protein and ligand (Figure 4).

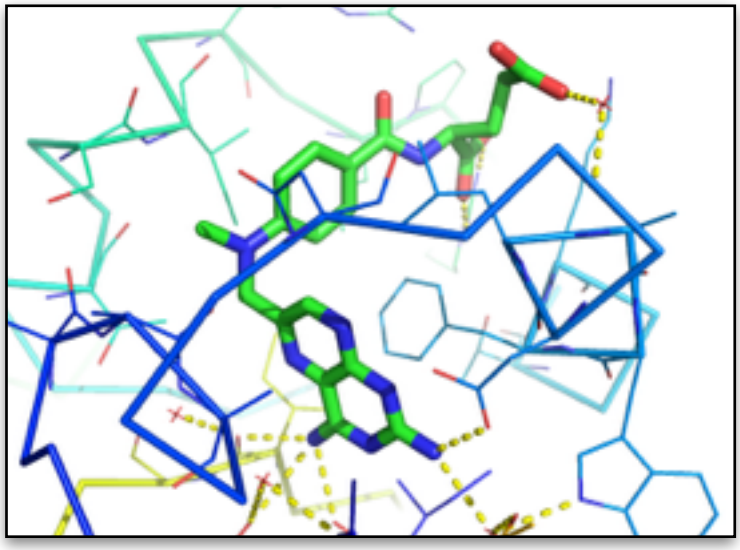


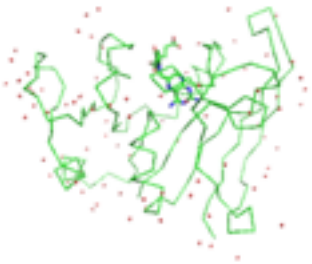
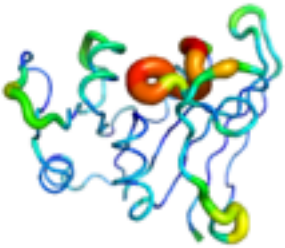
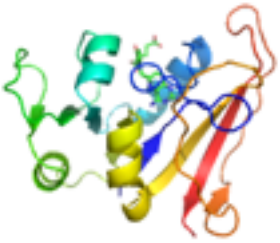
Figure 4: Methotrexate bound to *E. coli* DHFR. The visualization preset **ligands** is used to automatically obtain this view.

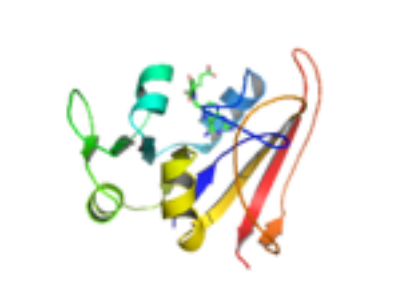
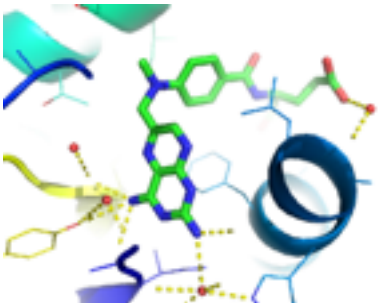
PyMOL also offers various similar visualization presets for the protein-ligand interactions. Explore these through navigating in the object menu (**A** → **preset** → **ligand sites**). See in particular the following:

- **cartoon**: provides the same view as **ligands** but with the protein in cartoon representation,
- **solid surface**: draws the binding pocket in solid surface,
- **transparent**: draws the binding pocket in transparent surface.

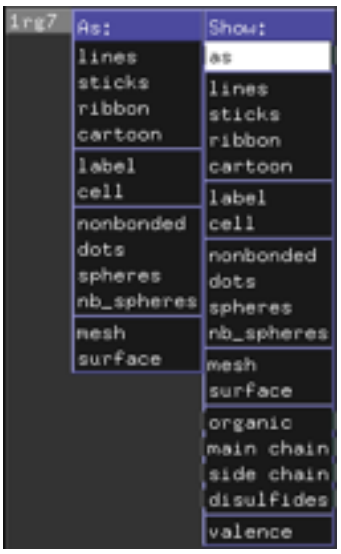
Task: Explore the other built-in visualisation presets – in particular those listed in in **Table 1**.

Table 1

Representation	Description	Illustration
simple [A → preset → simple]	a simple ribbon representation with ligands highlighted in sticks	
b factor putty [A → preset → b factor putty]	a sausage representation highlighting areas with high temperature factors	
pretty [A → preset → pretty]	cartoon representation colored in a rainbow spectrum with ligands in sticks	

Representation	Description	Illustration
publication [A → preset → publication]	similar to “pretty” but with smooth loops and “fancy” helices	
ligand sites [A → preset → ligand sites → cartoon]	detailed view of the protein-ligand interactions	

Displaying molecular representations



Besides the various automatic representation presets shown above you can also set the molecular representation to be shown more manually. Click the **S** (show) button from the object menu next to the **1rg7** entry. From the menu appearing you can choose to show various molecular representations, such as **lines**, **sticks**, **ribbon**, etc. From the first menu entry (**as**) you can select the representation in which you want the protein to be shown as.

Try the following and observe how the molecular representation changes in the viewer window:

- **S** → **as** → **spheres**
- **S** → **as** → **surface**
- **S** → **as** → **cartoon**
- **S** → **as** → **lines**
- **S** → **cartoon**
- **S** → **ribbon**

The various representations can also be hidden using the **H** (hide) button from the object menu. e.g. try the following:

- **H** → **ribbon**
- **H** → **lines**

Explore also the various other options in the Show menu (such as **organic**, **main chain**, etc).

Make sure you use **S** → **as** → **lines** before continuing to the next step.

Introducing the command line

All commands through the menu system can also be provided through the PyMOL command line console. You might find this complicated at first, but once you master it, you will find this very useful (if not essential).

Locate the command line (see **Figure 3**; its above and below the visualisation window). To show a particular representation, use the **show** command followed the by representation name:

```
show cartoon          [shows cartoon for all objects]
show lines            [shows lines for all objects]
```

To hide a representation use

```
hide lines
```

You can also choose to show only one particular representation with the **as** command:

```
as cartoon
```

```
as ribbon
```

The zoom and orient commands are the equivalent to the zoom and orient buttons, respectively:

```
zoom
```

```
orient
```

We will come back to more useful commands through the tutorial. Make sure you use

```
as lines
```

before you continue to the next section.

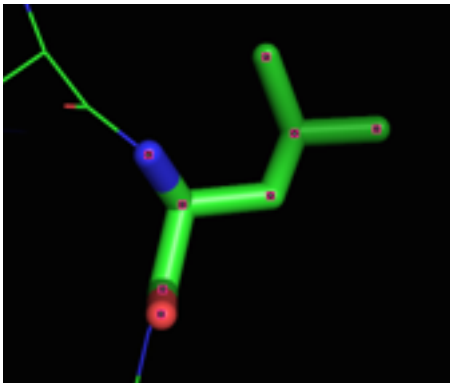
Working with atom selections

More customized visualisation and investigations of biomolecular structures often requires functionality to select individual or group of atoms. In PyMOL the simplest approach for atom selection is by simply clicking the particular atom in the visualisation window.

Click any atom in the protein. Note the new entry in the object menu on the right hand side called **(sele)**, and the selected atoms are shown in the viewer window with small pink squares.

To zoom in on your selection simply do A → zoom on the **(sele)** entry. This is equivalent to

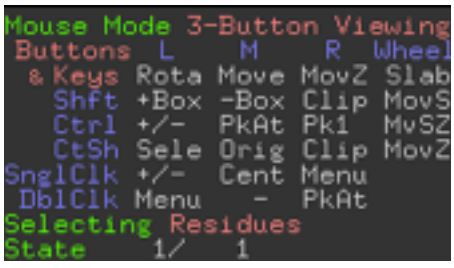
```
zoom sele
```



in the command line. You can now assign a different molecular representation to the selection e.g. with S → sticks on the **(sele)** entry. The selected atoms should now appear as sticks in view window. The equivalent in the command line is

```
show sticks, sele
```

Side note: Even though we initially only clicked at one atom in the protein all atoms in the particular residue is selected by default. This behaviour can be changed by clicking on the **Selecting** in the mouse controls panel (lower right). Click it once, and notice that **Selecting** now changes from **Residues** to **Chains**. Click now once at the protein, and notice that the entire protein is selected (depicted with pink boxes). Click back to **Residues** in the **Selecting** field.



Side note: The parenthesis here means that this entry is a *selection*. Conversely, the **1rg7** entry is an object, and **(sele)** is an atom selection of this object. The selection entry has much the same menu system attached to it as an object entry, but there are important differences as we will see.

Another way of selecting residues is through the sequence viewer which you can toggle through the **S** (for sequence) in the lower right corner of the screen (below the object menu). When clicking the **S** once you will see the sequence browser on the top of the

viewer window. To select a residue from here, simply click (and drag) the relevant residue codes. To disable the selection, click on an empty void in the viewer window, or click the **(sele)** entry in the object menu.

Side note: Selections can be deleted by clicking **A** → **delete selection**, and renamed with **A** → **rename selection.**

Click the residue of name **MTX** (at position 161 in the sequence). This will select the methotrexate (MTX) atoms bound to the DHFR structure. Rename the selection to (mtx) by clicking **A** → **rename selection** any type in “mtx” in the dialog.

Task: Zoom on the selected atoms, and view the MTX molecule in sticks.

Atom selections with the command line

Using the command line enables more versatile and specific atom selections. The “select” command enables atom selection through the command line. The command can be combined with various keywords speciyfing which atoms, residues, or chains to select. Here are some examples:

select resname selects the residue(s) by name. e.g. select residue with name MTX:

```
select resname MTX
```

select resid selects the residue(s) by identifier. e.g. select residue with ID 161:

```
select resid 161
```

select chain selects atoms in the specified chain. e.g. select chain with ID A:

```
select chain A
```

select name selects atoms with the specified name. e.g. select all CA atoms:

```
select name CA
```

The selection commands above will generate a selection with the default name **(sele)**. You can specify another name by adding it to the command like this:

Create a new selection called **(mtx)**:

```
select mtx, resname MTX
```

Create a new selection called **(protein)**:

```
select protein, resid 1-160
```

Similarly to using the in range (-) symbol, you can use the addition (+) syntax. The following command will select the residues in rage 15 to 20 as well as residue number 25:

```
select resid 15-20+25
```

You can also combine selection statements using the **AND** and **OR** operator. AND will select the atoms present in both the first and the second statement (intersection) while OR will select the atoms present in the first or the second statement (union). The following example will select 5 atoms corresponding to the CA (c-alph) atoms in the residues 15-20:

```
select resid 15-20 and name CA
```

Read more about atom selection in PyMOL here.

Task: Create a selection with name "bsite" consisting of residues with IDs 27, 31, and 94. Display the selection sticks representation on zoom on it.

Colour the atom selections

Now that we have a few selection entries (protein, mtx, bsite) we can easily color these to contrast the different elements in the structure. As an example, use the **C** button (for color) next to the **(mtx)** selection to change the color e.g. from green to blue.

Coloring can also be done using the command line. Below we color the mtx selection red and the protein blue:

```
color red, mtx

color white, protein
```

Note that it is often useful to color by atom elements, i.e. to keep oxygens red, nitrogens blue, sulphurs yellow etc, and only adjust the colors of carbon atoms. This colouring option is available through **C** → **by element** → **CHNOS** (try it on the mtx selection). In the command line this is less trivial but can be carried out with combining the color command with selection statements:

```
color green, mtx and elem C

color red, mtx and elem O

color blue, mtx and elem N
```

Now we have coloured all carbon elements of the mtx selection white; oxygens red; and nitrogens blue.

Overview of DHFR

Cartoon representation

With our selection entries and skills from above we can now easily create a nice overview figure of *E. coli* DHFR in complex with MTX. Make sure you focus on the entire protein by clicking on the **Orient** button.

Task: Show the ligand (mtx) as spheres, and the protein in cartoon representation. Try different colouring options for the two selections and find a suitable orientation.

Residues 9-24 in DHFR have been termed the "Met20" loop and is an important contributor in the reduction of dihydrofolate to tetrahydrofolate.

Task: Make a selection called met20 consisting of residues 9 to 24. Color the protein white, and the loop red. Here is my version:

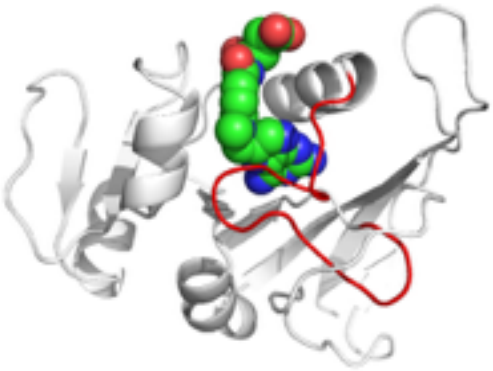
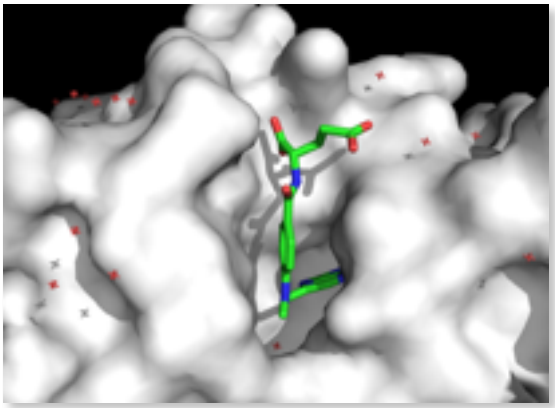


Figure 5: Overview of DHFR in complex with MTX where the "Met20" loop is coloured red.

Task: Hide the ligand and color the protein by secondary structure elements (hint: **C → by ss**) and rainbow (hint: **C → spectrum → rainbow**).

Save this session by navigating in the main menu before we continue: **File → Save session**.



Surface representation

While the cartoon representation display how the protein chain is folded the surface representation are better at displaying cavities and other surface properties. From the show (**S**) menu of the protein selection entry select **surface**, and show the ligand as sticks. Observe the cavity in which MTX binds.

Verify that you still have the selections **(protein)** and **(mtx)** in the object menu on the right. Show the protein in cartoon representation either through using the menu system, or the command line:

```
as cartoon, protein
```

```
as sticks, mtx
```

Exploring the binding site of DHFR

To explore the binding site in more detail it can be useful to only select and show the residues in the immediate vicinity of the ligand. This can be done by using the *modify* function from the action (**A**) menu:

From the selection entry (e.g. **(mtx)**) click **A** → **modify** → **around** → **residues within 5 A**. This approach will modify the selection to contain the atoms around the ligand (and not the ligand itself). Rename this selection to **near** using **A** → **rename selection**. You will now have to re-select MTX again.

The command line can also be used to select the residues within a certain radius of another selection using the byres and around keywords (read more here):

```
select near, byres mtx around 5
```

Our new new **(near)** selection contains the atoms within 5 Å of the ligand. Show these residues as lines and zoom at the ligand:

```
show lines, near
```

```
zoom mtx
```

Now that we see all residues close to the ligand we can more easily pick out residues particularly important for the strong binding of MTX to DHFR.

Task: Search for those residues you assume can facilitate hydrogen bonding with the ligand, and select those by clicking on them one at the time. Rename the selection to **hbondres.** Use your skills from above to display the selected residues as well as the ligand in sticks and the protein in cartoon representation. (You should be able to identify Asp27, Arg57, Ile94, and Tyr100, and rename the selection to **hbondres**.)

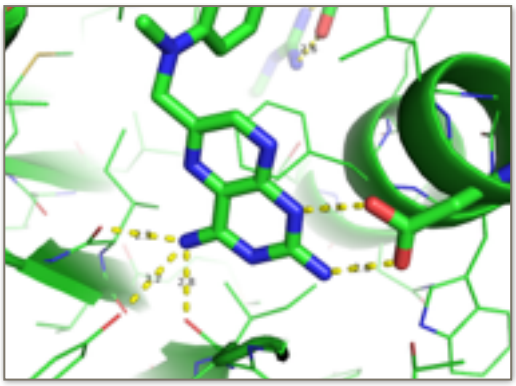
(**IMPORTANT:** make sure you do not have another selection enabled prior to clicking. Click first at an empty void to disable the current selection.)

Side note: When clicking an atom in the structure the console will output details on the selected atom(s). This will look something like

```
You clicked /1rg7//A/ASP`27/CG
Selector: selection "sele" defined with 8 atoms.
```

The first line specifies the *selection macro*. This represents another way of writing the selection. The logic is:

```
/object-name/segi-identifier/chain-identifier/resi-
identifier/name-identifier
```



To add dashed lines between selected atoms we can use the measurement tool. Find this by navigating in the main menu through **Wizard** → **Measurement**. Follow the instructions on the screen (**click first atom, click second atom**). Notice the dashed line appearing between the two atoms you clicked together with a label of the distance (in Å), as well as a new **measure1** object appearing in the object menu.

Task: Measure the distances between a few of the potential hydrogen bonds between MTX and the protein residues (in particular the charged residues) as shown in the figure.

Publication quality figures

Ray tracing for better resolution

Now we're getting close to a nice representation of the DHFR:MTX complex. To use the figure in a publication (or equivalent) it is good practice to enhance the quality and resolution by ray tracing. Type

```
ray
```

in the command line and observe the better quality when the ray tracing has completed. Ray tracing can be done in four different modes (see figures below) and can be set with the **set ray_trace_mode** command.

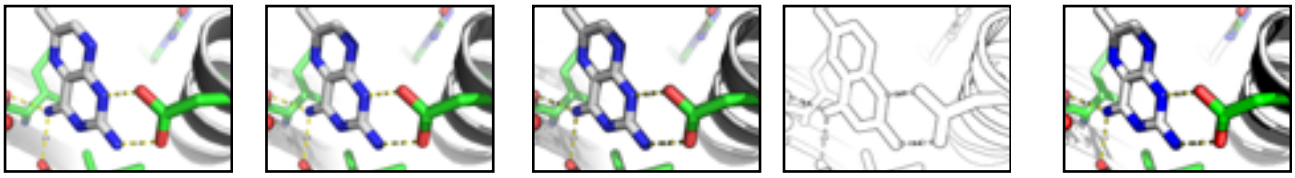


Figure 6: Ray trace modes. A) image not ray traced, B) ray trace mode 0, C) mode 1, D) mode 2, E) mode 3.

You can change the ray trace mode to 1 with the command

```
set ray_trace_mode, 1
```

Adjusting the default settings

For publications in particular it's good practice to use white background. Set this with by clicking **Display → Background → White**. Or alternatively use the command line:

```
bg white
```

Personal preference obviously plays a role, but here are some additional settings that can be used to make a nice publication quality figure.

```
set cartoon_fancy_helices, on

set cartoon_highlight_color, grey50

set ray_trace_mode, 1
```

Note that you can also ray trace specifying the width of the figure:

```
ray 1000
```

Task: Use your new skills to color the protein, ligand and a few selected residues and display them in favourable representations. Use your favorite ray trace mode and ray trace your view. Here is my version:

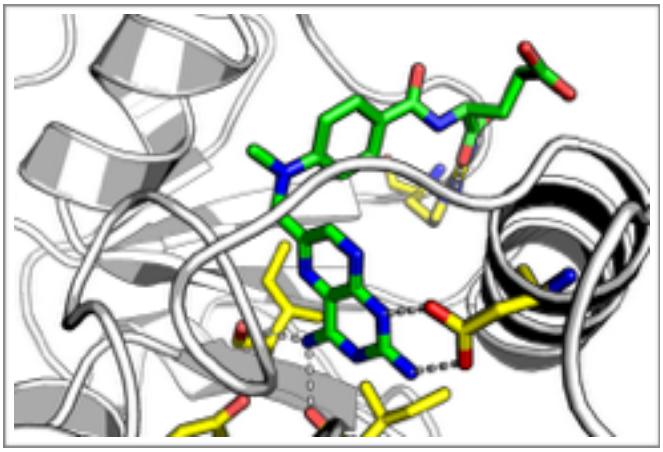


Figure 7: Ray traced model of the DHFR:MTX complex.

If you feel advanced: Unfortunately when colouring the individual residues one color (e.g. yellow) and the cartoon another color (e.g. white) this causes a small problem in PyMOL: we have now also coloured the cartoon representation of the hbondres residues yellow. Thats doesn't look any good so we need to tell PyMOL to maintain white on the cartoon representation:

```
set cartoon_color, white, protein
```

After adjusting this setting the regular colouring of the cartoon representation in this object will have no effect. To enable regular colouring again:

```
set cartoon_color, default, protein
```

Some useful links

The internet is full of tricks on how to make nice figures with PyMOL. Here are some links that you might want to check out:

http://www.pymolwiki.org/index.php/PLoS

http://www.pymolwiki.org/index.php/Gallery

http://www-cryst.bioc.cam.ac.uk/members/zbyszek/figures_pymol

Comparing multiple species of DHFR

PyMOL is also an excellent tool to compare and align protein structures with unequal sequences. Below we fetch the structure of DHFR from two species: *Escherichia coli* and *homo sapiens*. Open a new PyMOL session by navigating to **File → reinitialize** (or restart the program). Now fetch the three structures using the fetch command, and zoom on all structures:

```
fetch 1rg7 1dhf

zoom
```

Use your skills in PyMOL to view these in cartoon representation (hint: **`as cartoon`**, or **S → as → cartoon)**. Note that PyMOL automatically colors the various objects in different colours.

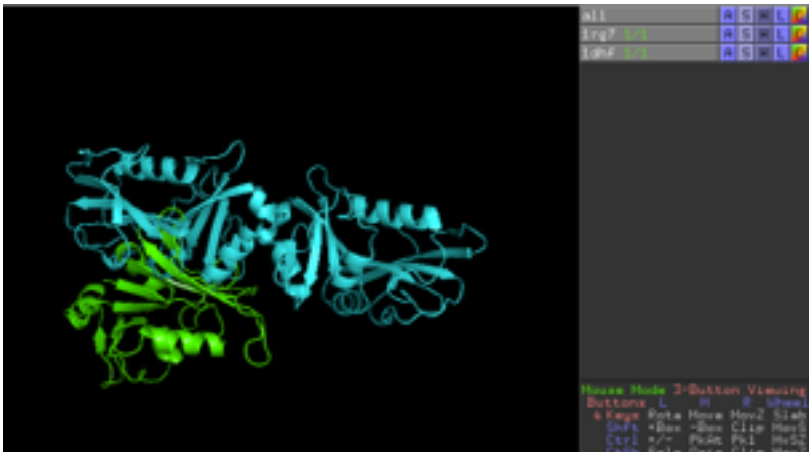


Figure 8: Loaded PDBs **1rg7** and **1dhf**. By default multiple loaded objects are coloured differently. Note that 1dhf contains two chains.

Notice that the structure from humans (1dhf; cyan) contains two chains. Here we will only consider the momomers and we can go ahead and remove of the chains. This can be done by selecting the atoms in the chain and remove these atoms. Below we select chain B of the 1dhf structure, and use the **`remove`** command to delete the selected atoms:

```
select 1dhf and chain B

remove sele
```

(you can remove atoms with A → remove atoms). We can now easily align these structures using the align tool. From the 1rg7 object click A → align → all to this. Click the zoom button, or type **`zoom`** in the command line.

Note that aligning can also be performed through the command line using the **`align`** command:

```
align 1dhf, 1rg7
```

Open the sequence viewer by clicking the **S** in the bottom right corner. This shows the sequence alignment of the two structures with dashes (–) representing gaps in the alignment. Investigate the sequence

alignment. Note that the amino acids colored grey depict areas with structural deviations, while coloured residues depict areas where the two structures obtain a similar structure/ conformation.

Task: Select the first gap containing columns, zoom at the selected residues and colour these residues red.

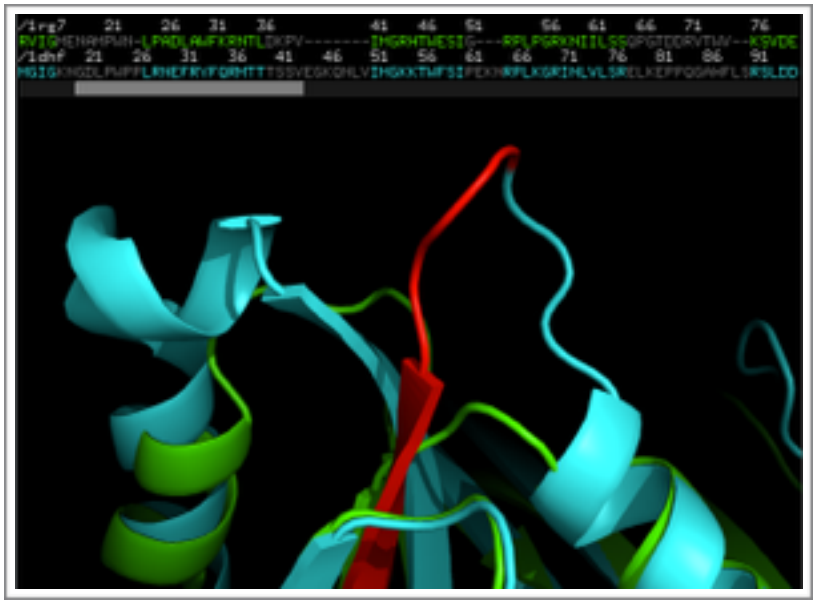
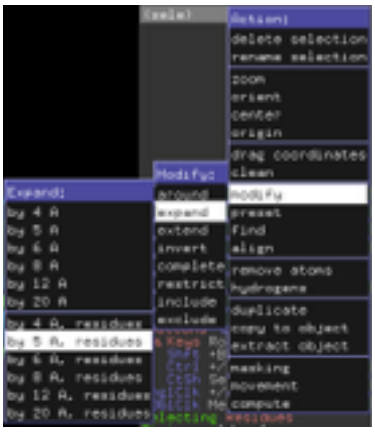


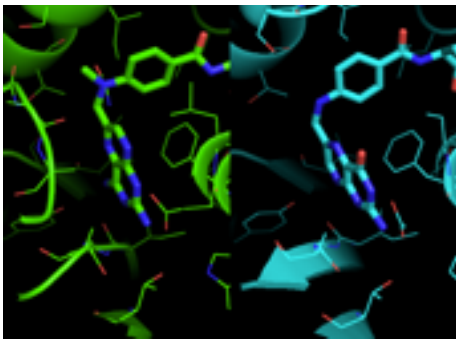
Figure 9: Alignment of *E.coli* (green) and human (cyan) DHFR. Residues not present in *E-coli* DHFR are coloured red.

Task: Select the two ligands present in these structure (residue names MTX and FOL). Zoom at the selection, and display them as sticks. From the selection entry, go to **A → modify → expand → by 5 A, residues**. And show lines (**S → lines)**.



Explore the similarities and differences in the binding site. **Can you find which residues differ?** Appreciate also the similarities and differences of the two ligands (folate and methotrexate).

You can view the two structures side by side by enabling grid viewing (**Display → Grid → By Object**).



Various

Point mutations

Methotrexate is a potent drug used in the treatment of cancer. However, drug exposure can render the enzyme resistant. The mechanisms involved in such resistance are generally mutations in the *dhfr* gene resulting in a weaker binding of MTX to the DHFR enzyme. Leu22Arg is one such mutant variant giving a 1300-fold decrease in MTX affinity.

PyMOL facilitates generating and visualising such point mutations of the amino acid residues in the loaded structure. Open the human DHFR:MTX complex with PDB ID 1U72 and show the protein in cartoon and MTX residue 22 in sticks. To perform the point mutation navigate in the main menu to the Mutagenesis wizard: **Wizard → Mutagenesis**. Follow the instructions on the screen by click at the Leu22 residue (which you display in sticks). Now hit the **No mutation** button on the right hand side, and select **ARG**, then hit the **Apply** and finally the **Done** buttons.

We have now changed a highly hydrophobic residue to a charged and bulky residue close to the MTX binding site.

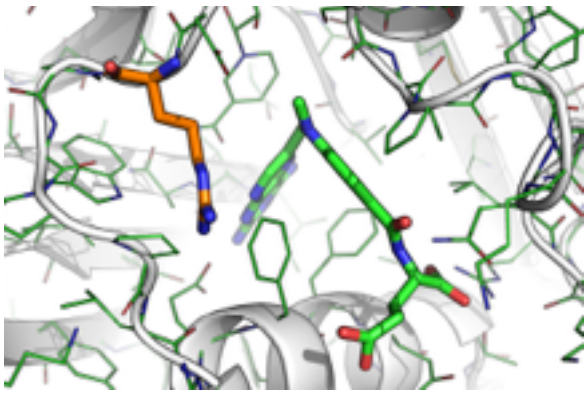


Figure 10: Mutant variant L22R of human DHFR. R22 is shown in orange sticks while MTX is shown in green sticks.

Generating morphs

PyMOL can also be used to generate morphs been two related structures with different conformations. We will illustrate this by looking at two conformations of *E. coli* DHFR. Load the structures with PDB IDs **1rx2** and **1rx5**. Make sure the structures are aligned and show them in cartoon representation. From the menu of 1rx5: **A → generate → morph → to molecule → 1rx2**. PyMOL will now search for a path between the two structures in a iterative approach. When the calculation is complete you have a new object with name **morph01** in your object menu. Click the play button (on the bottom of the window) to show the morph movie. Show the movie both with and without lines.

Binding pockets

The surface representation in PyMOL can aid in identifying cavities and pockets in the structure. Open the PDB with ID **1rx2**. From the main menu, navigate to **Setting → surface → Cavities and pockets only**. Now toggle the surface representation of the protein.

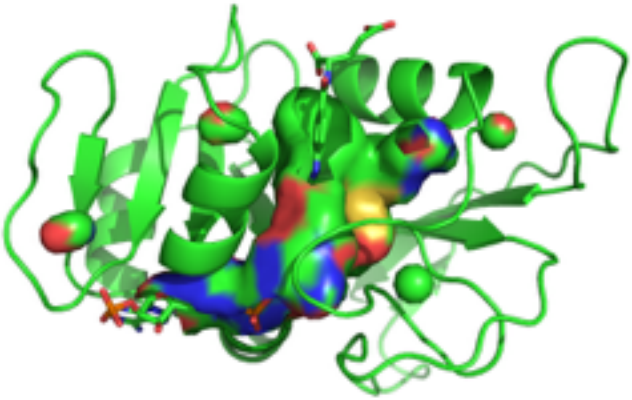


Figure 11: View of the cavities and binding pockets in DHFR.

Done!