PREPARATION OF IMAGES OF PROTEIN STRUCTURES USING THE PYMOL SOFTWARE.

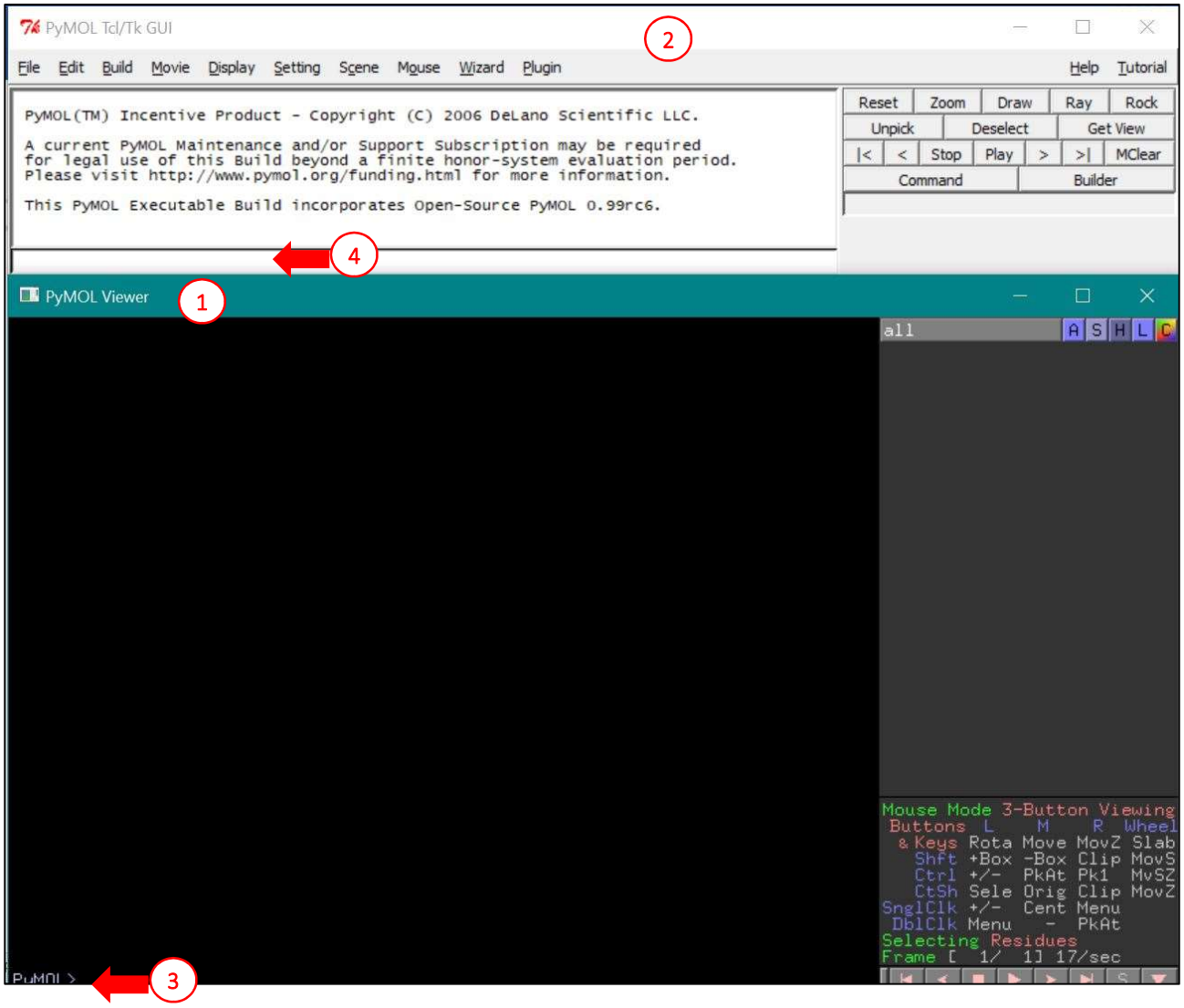
The PyMOL software [1] allows you to visualize protein structures, compare their conformations, and obtain good quality images of the overall structure or its local details. The software requires as input a .pdb file that can be obtained from a previous crystallographic refinement or from the Protein Data Bank [2]. A guide for this software [3], available as a wiki, contains tutorials and examples.

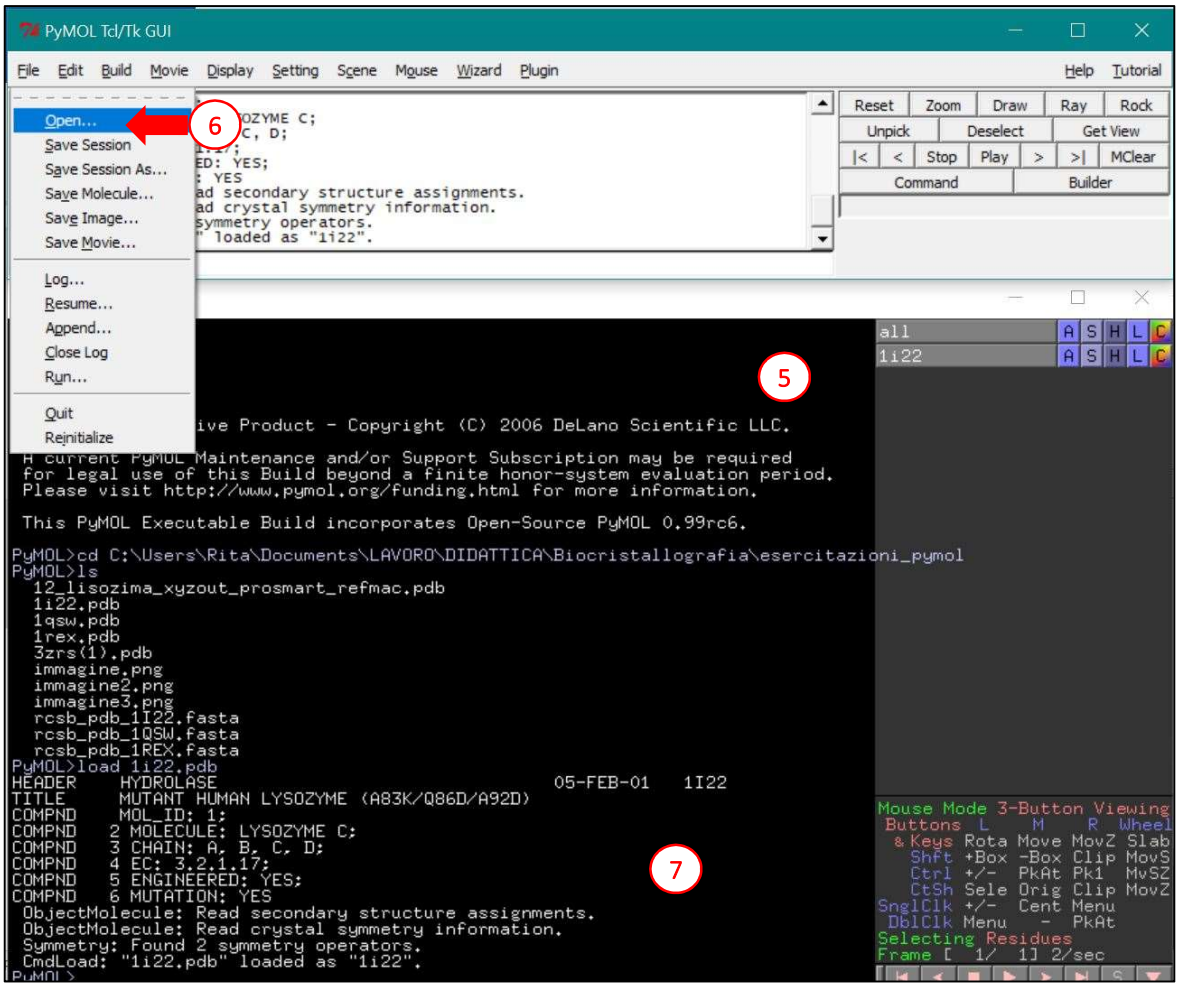
Viewing a pdb file

When the program is started, 2 different windows open (or 2 panels on the same window, depending on the software version): a display window (**1**) where the structure is displayed, and a command window (**2**). To work with PyMOL, you can select commands from the menus of the command window, or type commands in the prompt of the display (**3**) or the command window (**4**). The list of previous commands (**5**) can be visualized by pressing the *Esc* key on the keyboard while working in the display window.

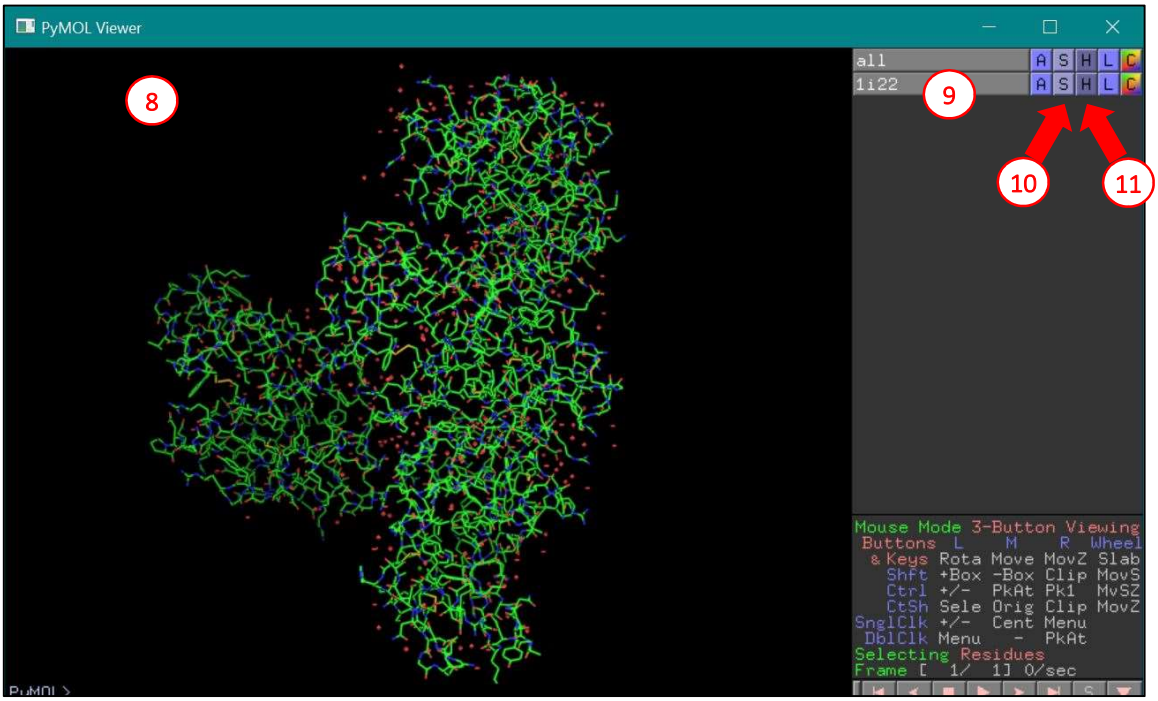
To open a pdb file containing the coordinates of a protein structure, select *File* > *Open* from the top menu in the command window (**6**), or type in the command line:

```
> cd C:\<working directory>
> load <filename>.pdb, <(optional) object's name>
```





When the structure is loaded, the software automatically reads the crystal symmetry (**7**), from the CRYST1 line of the pdb file and the unit cell parameters. By default, the protein is represented with lines (**8**). By clicking on the loaded objects in the menu on the right in the display window (**9**), you can select which ones to display.



Protein structure display options

The software allows several display options, including:

- lines: bonds are represented as lines
- sticks: bonds are represented as sticks (thicker than lines and with a cylindrical shape)
- cartoon: secondary structure is represented as helices or arrows (for β-sheets), without side chains
- ribbon: only the α-carbon structure is represented
- spheres: each atom is represented with the corresponding van der Waals sphere.

Display options can be selected from the "S" menu (S = show, **10**) on the right of the display window, or by typing the command:

```
> show <lines/sticks/cartoon/ribbon/spheres>, <object name>
```

On the contrary, graphical representations can be hidden from the "H" menu (H = hide, **11**) on the right, or by typing:

```
> hide <lines/sticks/cartoon/ribbon/spheres>, <object name>
```

Commands from the command line can be used also to apply a specific representation mode (or hide it) to a specific part of the models, by use the selection commands (see below).

Selection commands and syntax

A specific selection of atoms can be described using the following identifiers:

- *object*: identifies the specific object (in this case the identifier can be omitted, directly typing the object name)
- *chain*: using the letter that identifies the chain in the pdb file
- *resi*: selection of specific residues using their numbers (e.g., *resi 4-14* identifies all residues between 4 and 14, included)
- *resn*: identifies a residue type (e.g., *resn trp* selects all tryptophan residues)
- *name*: identifies an atom name (e.g., *name CA* selects all atoms whose name is CA in the pdb file)

The selection uses logical operators, among which *and* and *or*.

Some examples:

1) if you want to select residues 4 to 14 of the A chain of the object 1i22, and also all the proline residues present in the same A chain, you type the command:

```
> sele object 1i22 and chain A and (resi 4-14 or resn pro)
```

2) if you are interested in selecting residues 4 to 14 of chain A of object 1i22, but all proline residues present in the object, the command becomes:

```
> sele object 1i22 and ((chain A and resi 4-14) or resn pro)
```

Once selected, these residues can be represented as stick by typing:

```
> show sticks, sele
```

Or directly, without the selection command:

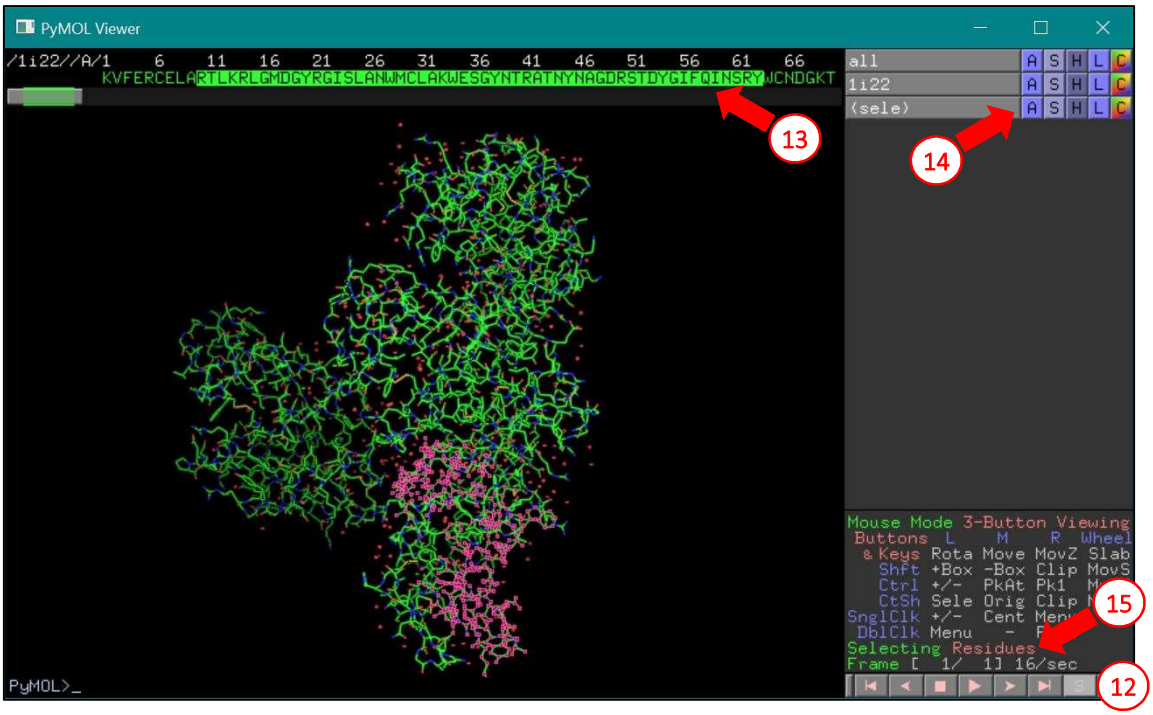
```
> show sticks, object 1i22 and ((chain A and resi 4-14) or resn pro)
```

You can select the same portion of the object using clicking on one residue at a time with the mouse left button. Alternatively, residues can be selected from the sequence, which can be displayed clicking on the "S" button at the bottom right (**12**). The protein sequence is visualized at the top (**13**). Once the selected portion is active, a new *(sele)* object appears in the menu on the right and display options for this selection can be modified using the "S" and "H" menus of the *(sele)* object (**14**).

In addition to selecting individual residues, it is also possible to select entire objects, chains, or individual atoms, either by typing the suitable commands or by clicking on the specific portions. In the latter case, you need to change first the selection method by clicking on the *Selecting* option at the bottom right (**15**).

The selected portion can be colored using the "C" menu (C = color, **16**) at the top right, or with the command:

```
> color <color name>, <selection>
```



Available color names are reported in the webpage [4].

Further visualization options

There are many more options for displaying structures (check the online wiki [3]). A list of display options is displayed in another dialog window by selecting the *Settings* > *Edit All* menu of the command window. Variables present in this list can be changed either from the dialog window or from the command line, using the `set` command:

> `set <variable>, <value>`

Some of the display options present in the list:

- *sphere_scale*: changes the representation of atoms as spheres, reducing or increasing the scale with respect to the van der Waals radius of the atom. The default value is 1, corresponding to the van der Waals radius. Accepts a number as a value.
- *stick_radius*: changes the representation of the bonds as sticks by varying their thickness. Accepts a number as a value.
- *ortho*: switches from the perspective view (default when opening the software) to the orthoscopic view. It accepts *1/0* (or *on/off*) values.
- *depth_cue*: the software may represent the third dimension, i.e. depth, using a transparency gradient, with dimming for objects farther from the observer. This option can be removed by setting the variable to a value of *0*. It accepts *1/0* (or *on/off*) values.
- *ray_shadow*: By default, when the image is prepared for saving with the `ray` command (see below), the software adds the shadows of the chains, considering the illumination source(s). It accepts *1/0* (or *on/off*) values.
- *ray_opaque_background*: by default, the image background is a solid black fill (the color can be changed, see below). With this setting the background can be made transparent, a very useful option for preparing images for presentations. It accepts *1/0* (or *on/off*) values.
- *bg_rgb*: changes the background color of the figure. It accepts 3 values, which must be separated by a space, corresponding respectively to the red, green and blue components of the new background color. For example, the white color is obtained with [1 1 1], black is [0 0 0].

The latter visualization option can be obtained from the *Display* > *Background* menu or with the command:

> `bg_color <color name>`

Displaying symmetry-related molecules

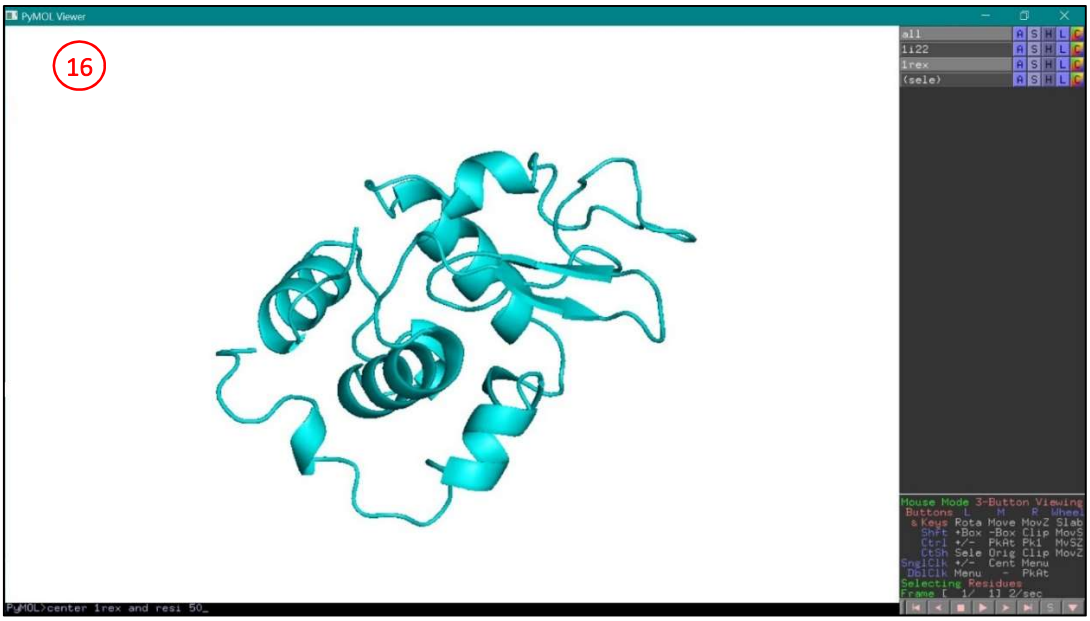
When a crystallographic structure is loaded, the software displays only the content of the pdb file, corresponding to the asymmetric unit. However, the software can display also symmetry related molecules taking advantage of the unit cell and symmetry information reported in the CRYST1 line of the pdb file corresponding to the object (`<object>`). With the `symexp` command, the software displays symmetry related molecules around a center (`<selection>`) and within a defined radius (`<radius>`). The `symexp` command syntax is:

> `symexp <prefix>, <object>, <selection>, <radius>`

The software creates new objects containing the symmetry related molecules, naming them using the defined prefix.

In the example below (**16-18**), the molecule is moved so that the center of the image is on residue 50 and the residue is represented as sticks:

> `center 1rex and resi 50`
> `show stick, 1rex and resi 50`



The next command is used to create the symmetry related objects (in this case only one, 19) named with the specified prefix:

```
> symexp sym, 1rex, 1rex and resi 50, 4
```

This PyMOL command is very useful to analyze protein-protein contacts in the crystal and evaluate their influence on side chain conformations. Furthermore, when the protein of interest is an oligomer located on a crystallographic symmetry element, the `symexp` command is necessary to show the functional protein, with the whole quaternary structure.

Distances, angles and torsional angles

PyMOL can be used to measure and represent distances, angles and torsional angles. Atoms defining the geometric measurement must be selected: 2 atoms for a distance, 3 atoms for an angle, or 4 atoms for a torsion angle. In this case, atom selection must be performed with the *Editing* mode of the mouse, which is set by clicking on *Mouse mode* on the bottom right (20). At this point atoms can be selected with the left button of the mouse (21). Finally, the suitable command should be typed in the command line: `distance`, `angle`, or `dihedral` commands are used to measure a distance, an angle, or a torsion angle, respectively. For example, to measure a distance (22) type:

```
> distance d1, pk1, pk2
> color black, d1
```

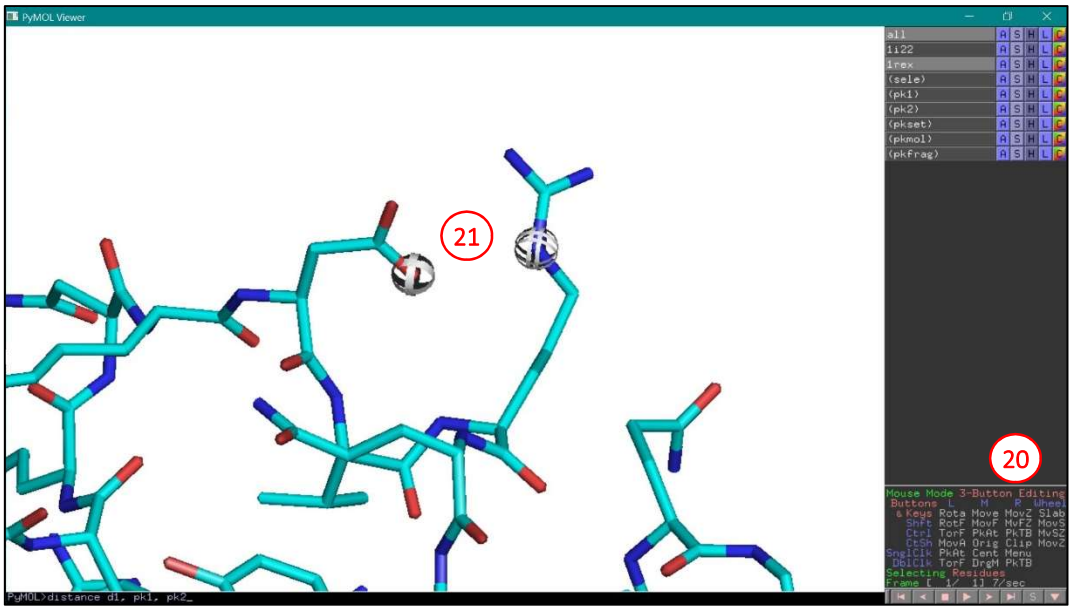
For every geometric measurement, the software creates a new object, whose visualization options can be modified from the menu on the right. In the example, the distance object is given the name “d1”. The following variables typed in the command, “pk1” and “pk2”, represent selected atoms (displayed also in the list on the right, 23). (The second command, which is optional, allows coloring the distance in the visualization).

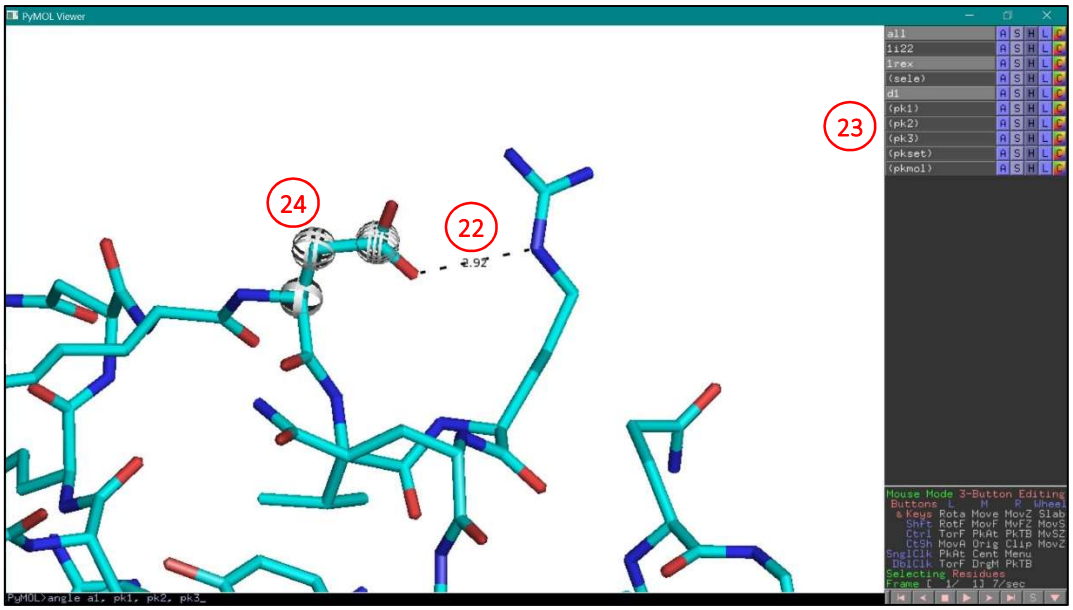
To measure an angle, select 3 atoms (24) and type:

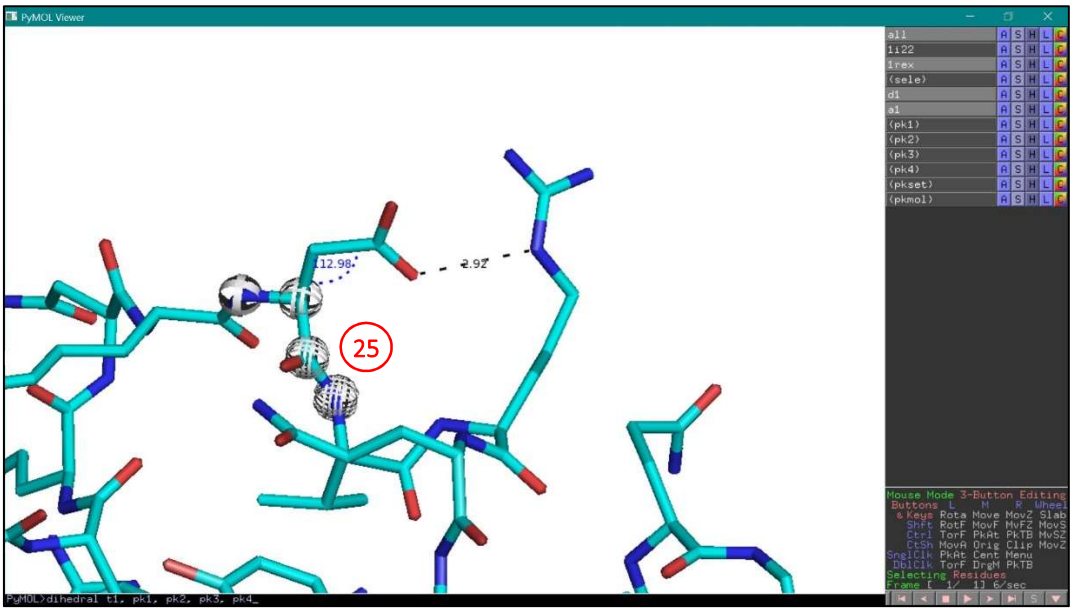
```
> angle a1, pk1, pk2, pk3
> color blue, a1
```

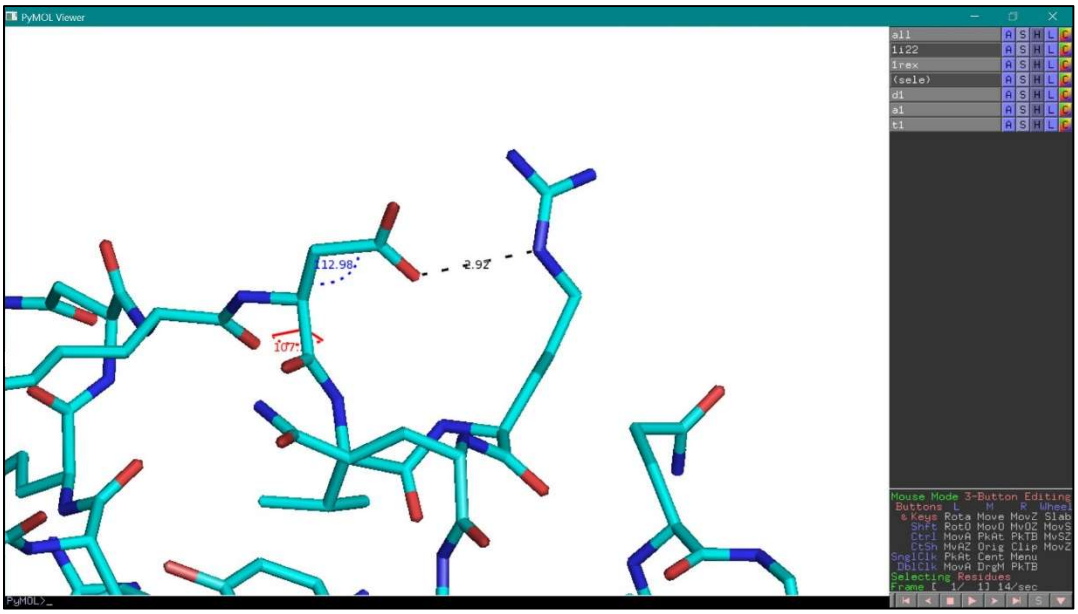
To measure a torsion angle, select 4 atoms (25) and type:

```
> dihedral t1, pk1, pk2, pk3, pk4
> color red, t1
```


PyMOL Viewer
21
20
PyMOL>distance d1, pk1, pk2_


PyMOL Viewer
23
24
22
2.97
PyMOL>angle a1, pk1, pk2, pk3_


PyMOL Viewer
112.98
2.97
25
PyMOL>dihedral t1, pk1, pk2, pk3, pk4_



Overlaying homologous proteins to compare their structures

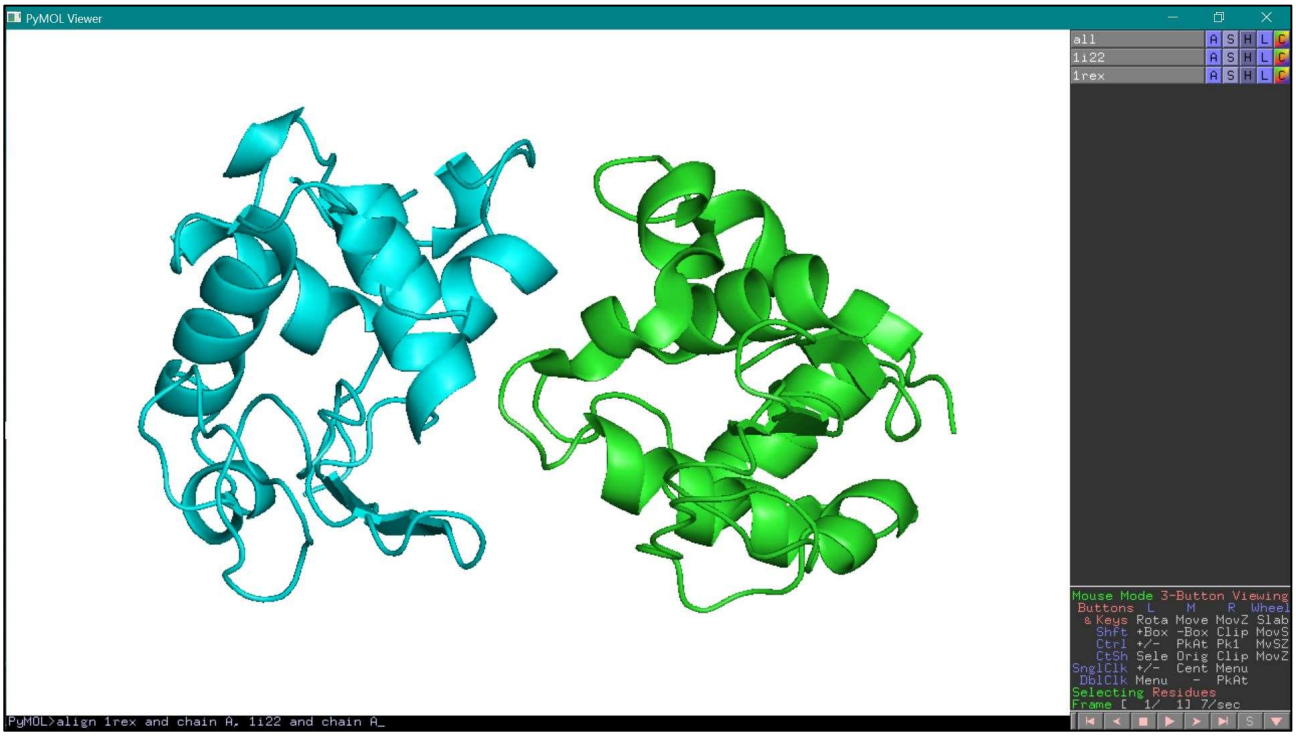
When comparing different structures of proteins from the same family, it can be useful to overlay their structures. The `align` command of PyMOL moves one molecule (`<selection 1>`) to overlay it with the other (`<selection 2>`), minimizing the distance between corresponding atoms of the two models. The syntax of the command is as follows:

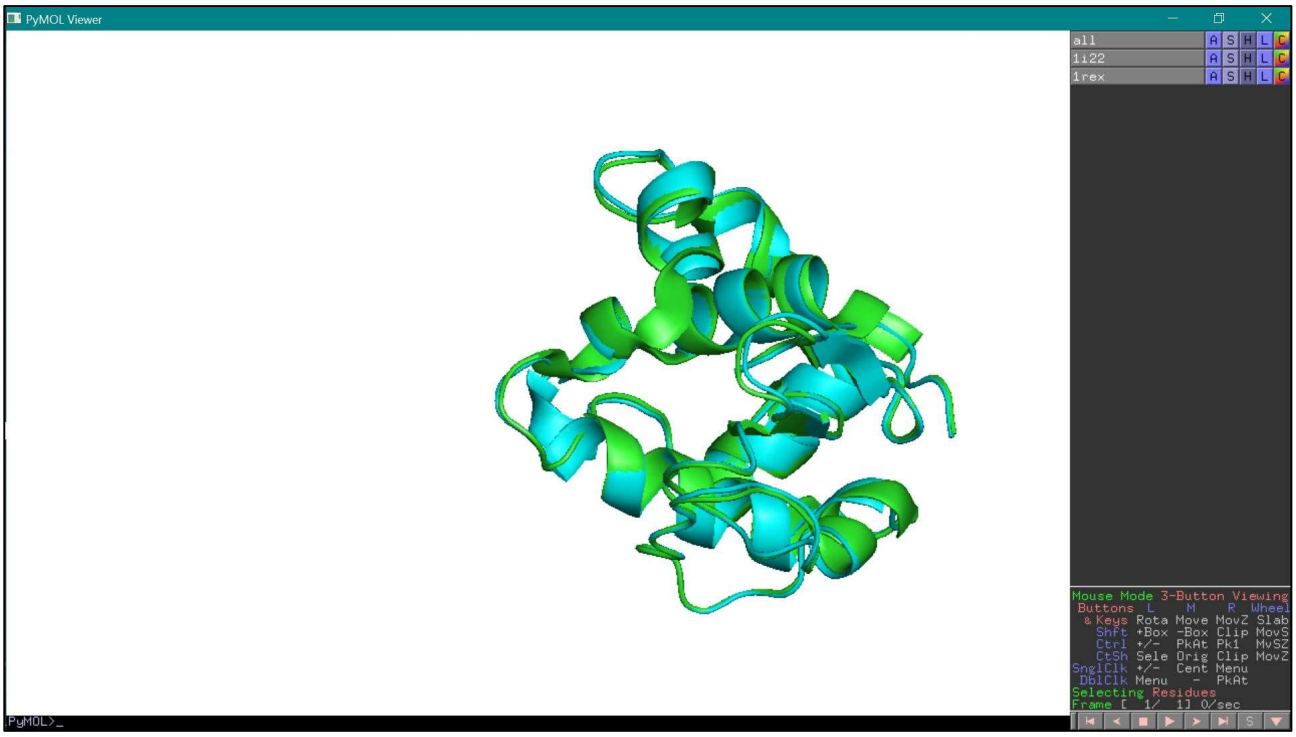
```
> align <selection 1>, <selection 2>
```

In using this command, you must remember that the two selections must belong to different objects.

For example, to align lysozyme structures corresponding to PDB codes 1i22 (green model) and 1rex (cyan model), type:

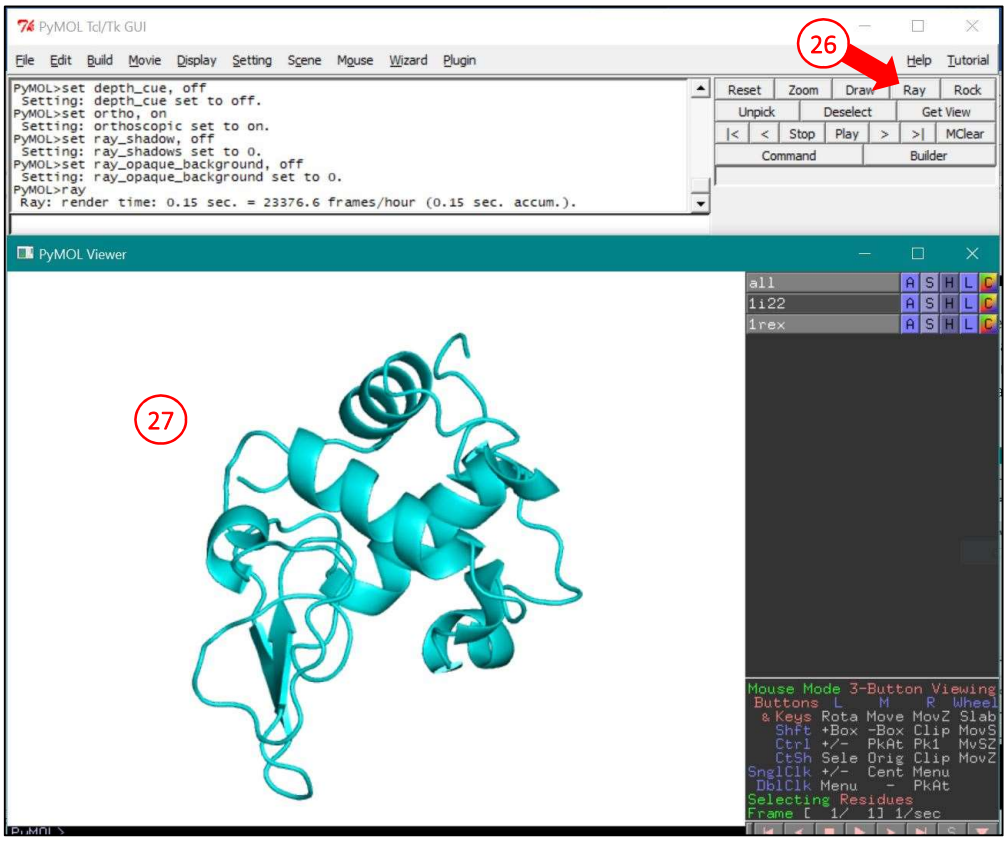
```
> align 1rex and chain A, 1i22 and chain A
```





Preparing images of the protein structure

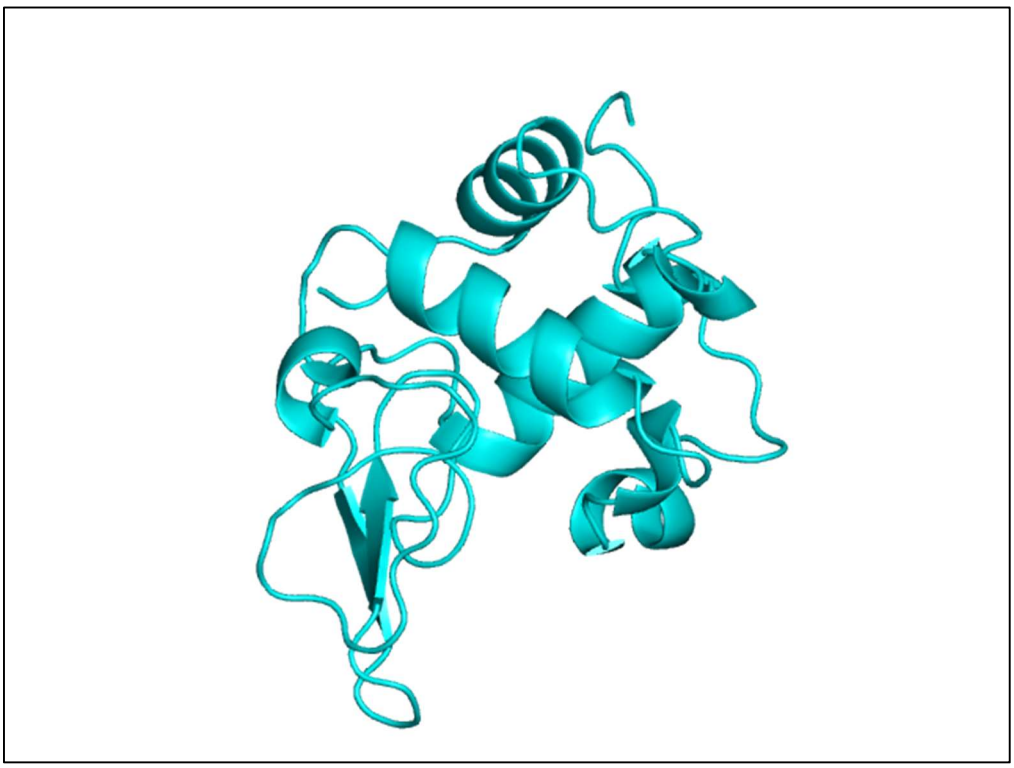
PyMOL allows you to prepare high-quality images of the structure of interest, with a high resolution compared to the display window. Before saving the image, the `ray` command improves the quality of the visualization. The `ray` command can be typed in the command line or selected from the command window (**26**). The resulting image has a higher quality, a higher resolution, and details as the depth cue or the shadowing are added (**27**).



At this point the structure should not be modified or rotated, otherwise the visualization optimization will be lost (but it can be easily restored with the `ray` command). The next command is used to save the image in png format:

> `png <nome file>.png`

Warning: the file is saved in the working directory. If the pdb file was loaded by clicking on the file icon, PyMOL will save the resulting image files in a specific folder inside the program directory. You can see in which directory files are saved by typing the command *pwd* and pressing the *Esc* key on the keyboard.



References

[1] The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

[2] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. *"The Protein Data Bank"* Nucleic Acids Research. 2000; 28, 235-242 (https://www.rcsb.org/)

[3] Community-run support site for the PyMOL molecular viewer (https://pymolwiki.org/)

[4] https://pymolwiki.org/index.php/Color_Values