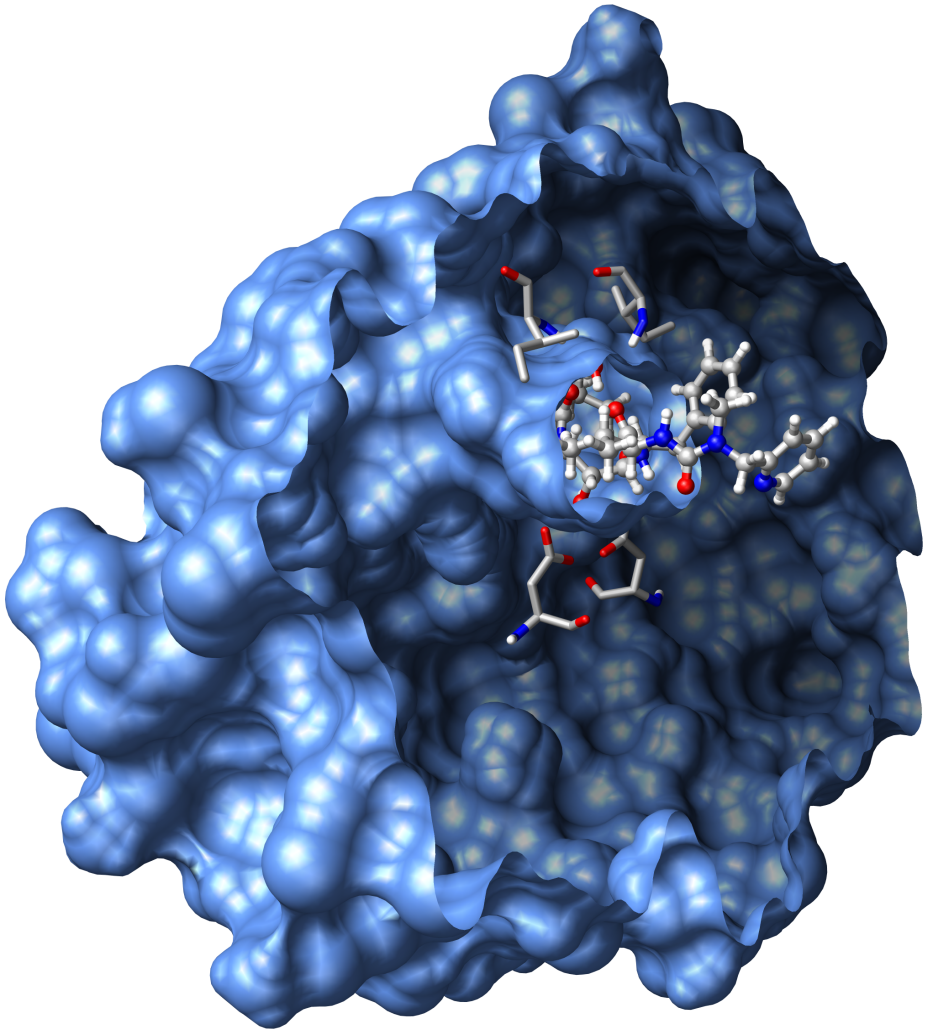
Practice Session 1:
3D structure visualization and high quality imaging using UCSF Chimera



Mathias Ferber, Justyna Iwaszkiewicz, Ute Röhrig, Antoine Daina, Vincent Zoete – 2021
Contacts : vincent.zoete@sib.swiss antoine.daina@sib.swiss

Presentation of Chimera

This introductory course about 3D structure visualization and high quality imaging for publication will make use of the free program Chimera. The official website of chimera can be found at the following address: http://www.cgl.ucsf.edu/chimera

Here is a brief description of Chimera and its features taken from the website:

" UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. Chimera includes complete documentation and several tutorials, and can be downloaded free of charge for academic, government, non-profit, and personal use. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics, funded by the National Institutes of Health (NIGMS P41-GM103311). "

The program can be downloaded at the following address:
http://www.cgl.ucsf.edu/chimera/download.html

When using chimera, one should cite this reference:
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *J. Comput. Chem.* **25**(13):1605-1612 (2004).

The following exercises will be used as tutorials to introduce some basic commands of Chimera. The HIV-1 protease in complex with the A77003 inhibitor (1HVI in PDB) will be used as an example. You are highly encouraged to test the different options that you will encounter in the different menus.

Generally, scientific journals require EPS or TIFF images, with a 300 dots per inch resolution or higher, and with RGB colors. You will see how to obtain such images using Chimera.

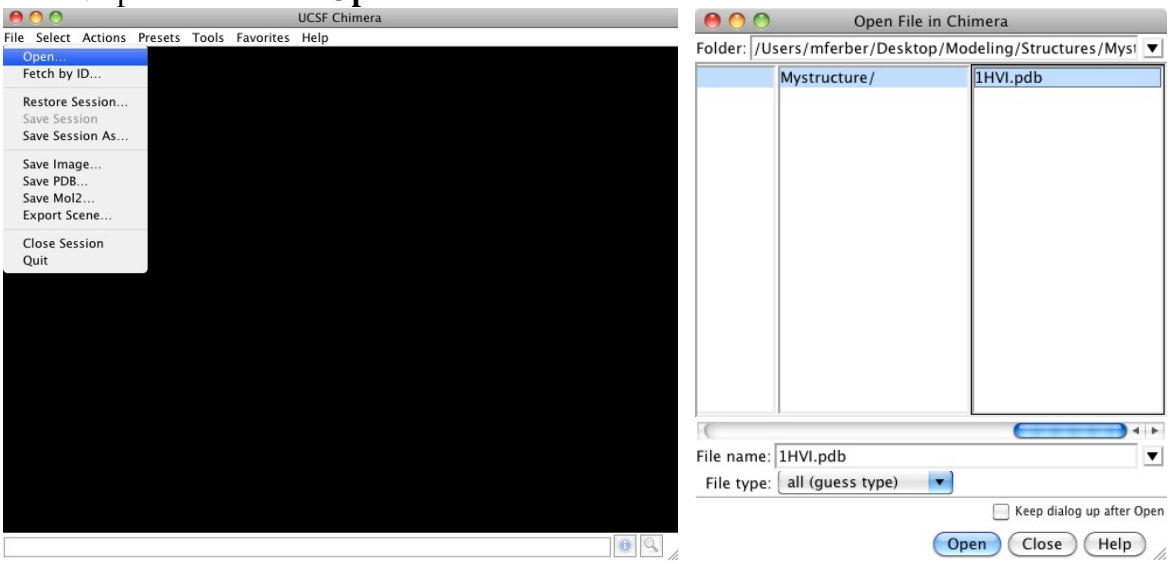
Exercise 1

Important. Before we start. To be able to follow this tutorial and understand the different concepts of molecular visualization, it is necessary to change a recently added option in Chimera. For this, just after opening Chimera for the first time, go to **"Favorites>Preferences…",** then select the category **"New Molecule"**, and change **"smart initial display"** to false. This option allows displaying new molecules with already advanced graphical options. Switching it off will allow to introduce some concepts. After this tutorial, it is recommended to switch it on again, to save time when displaying new molecules.

Loading a structure into Chimera

Two methods can be used to load a structure into Chimera.

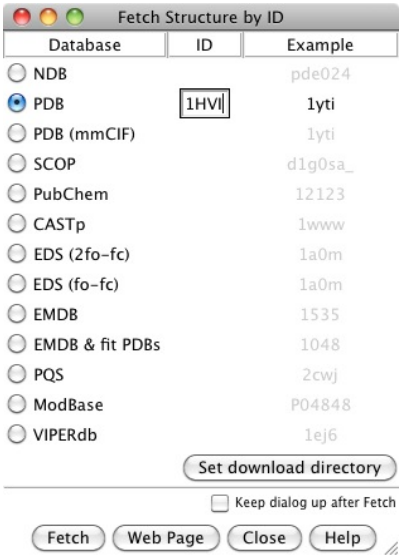
1) If the structure file is present in the user's computer, choose the menu item "**File/Open"**. Then select "1HVI.pdb" and click "**Open**".



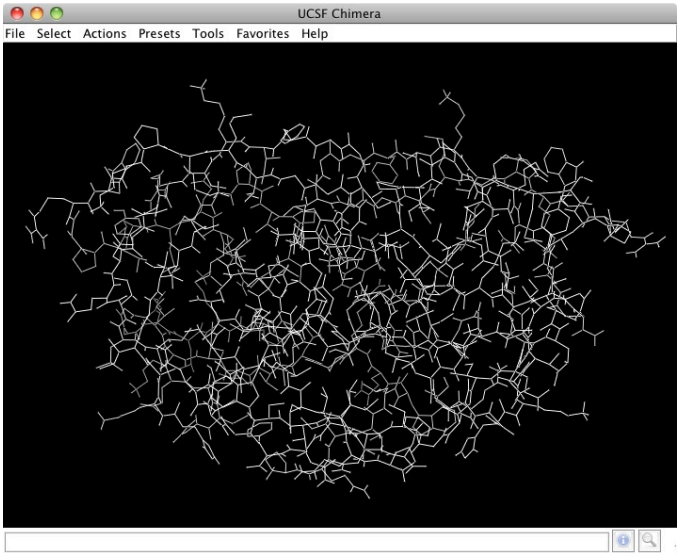
Under Unix or Mac OSX, the structure (for instance 1HVI.pdb) can also be loaded using the following command:

```
> chimera 1HVI.pdb
```

2) The structure might also be fetched from a database, when available. Choose the menu item "**File>Fetch by ID…"**. Choose the PDB databank and type the PDB code "1HVI". Then click "**Fetch**"



Once the structure has been loaded, all macromolecule's bonds should appear in the “wire” representation.



Moving / zooming

The macromolecule can be rotated by clicking the left button and dragging the cursor over Chimera's window. The translation is obtained similarly, but using the central mouse button. Finally, one can zoom in and out using the mouse right button.

Rotate: press left button and drag
Translate: press central button and drag
Zoom: press right button and drag

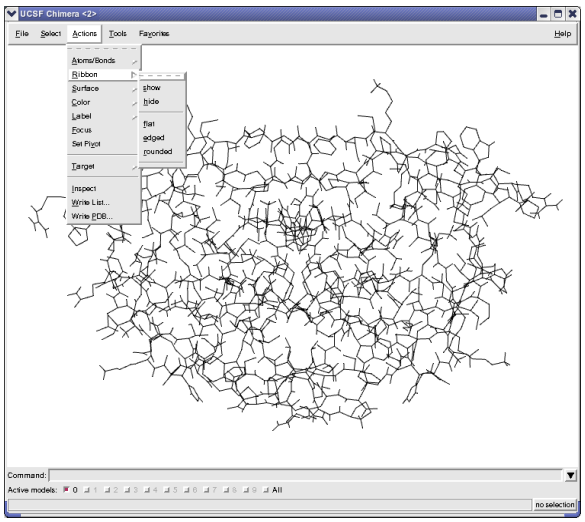
Command line...

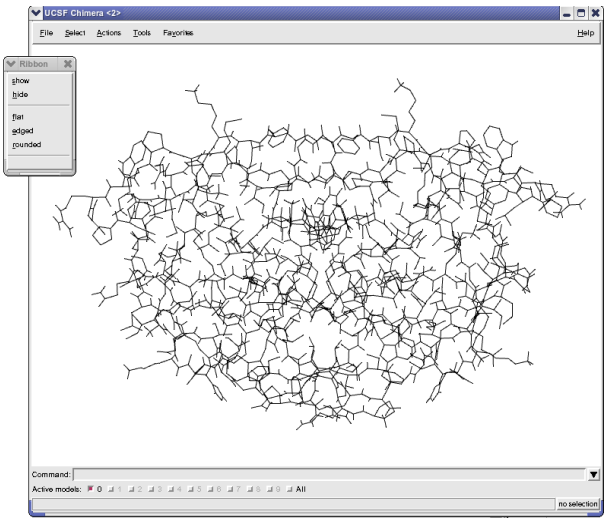
Most of Chimera functionalities can be accessed by typing a given command in the "**Command line**", which is situated at the bottom of the Chimera window. If the command line is not present, display it using "**Favorites>Command Line**". A line preceded by "Command" will appear on the lower part of the principal window. Examples of how to use the command line will be given later.

Ribbon representation

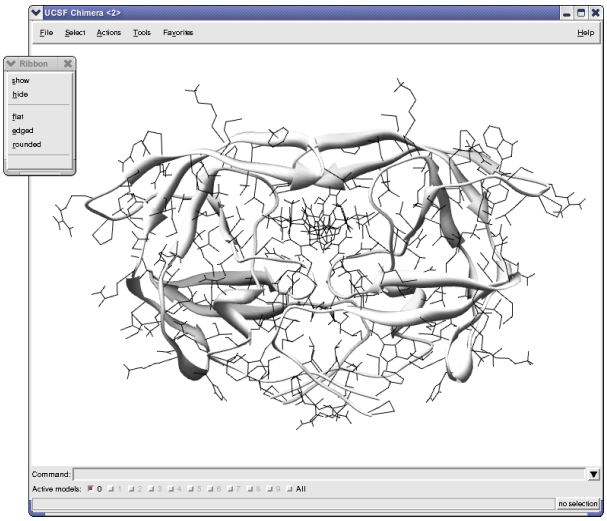
The secondary structure elements can be shown using the ribbon representation. This can be accessed through the "**Actions>Ribbon"** sub-menu.

One convenient feature of Chimera is its ability to detach sub-menus so that one can use them several times without the need to go again through the different menus. For instance, choose the "**Actions>Ribbon"** menu item, and click on the dotted line at the top of this sub-menu. The sub-menu will be detached and freely movable. This option is not available on all operating systems.





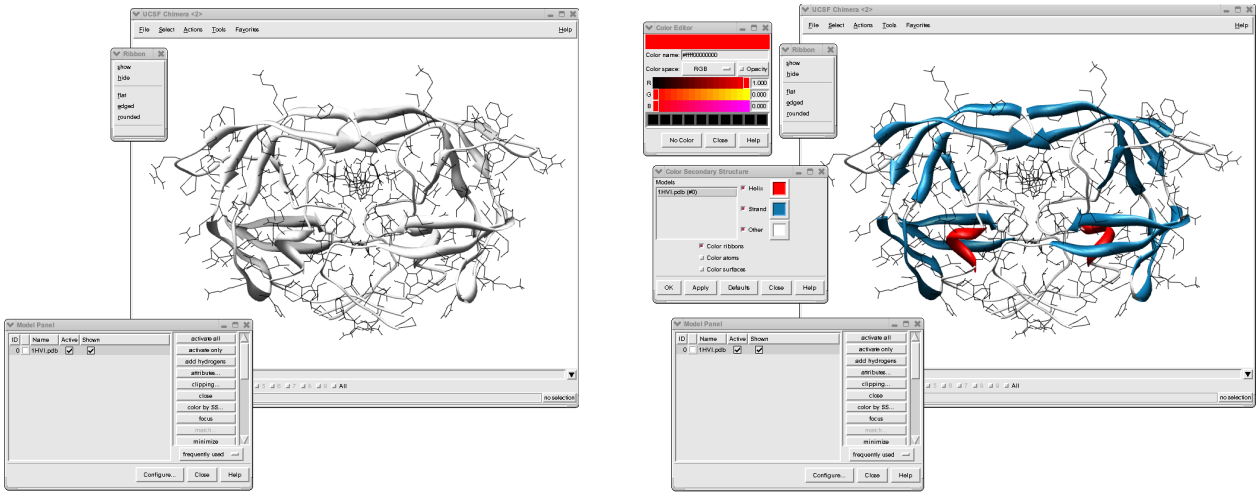
The ribbon representation of the secondary structure elements can be switched on by choosing "**show**" in the "**Ribbon**" menu. Three variants are available: "**flat**", "**edged**" and "**rounded**". You can try them all by clicking the corresponding menu item. Select the representation you prefer.



Important: type "**ksdssp**" in the command line to recalculate the secondary structure elements of the protein. This uses an implementation of the **K**absch and **S**ander algorithm for **d**efining the **s**econdary **s**tructure of **p**roteins.

Coloring the secondary structure elements. The Model Panel

The ribbon can be colored according to the secondary structure element, i.e. strand, helix or loop. Open the Model Panel by choosing the "**Tools>Depiction>Color Secondary Structure"** menu item. This will open the "**Color Secondary Structure**" window. Check the "**Helix**", "**Strand**" and "**Coil**" boxes. For each one, it is possible to modify the default color by clicking on the corresponding colored square to open the "**Color Editor**" window and change the RGB cursors positions. Click on the "**Apply**" button of the "**Color Secondary Structure**" window to apply the coloring.



You can try several color combinations and keep the one you prefer. Finally, you can close the "**Ribbon**", "**Color Secondary Structure**" and "**Color Editor**" windows by clicking the "**Close**" button, or the top right "**X**" icon.

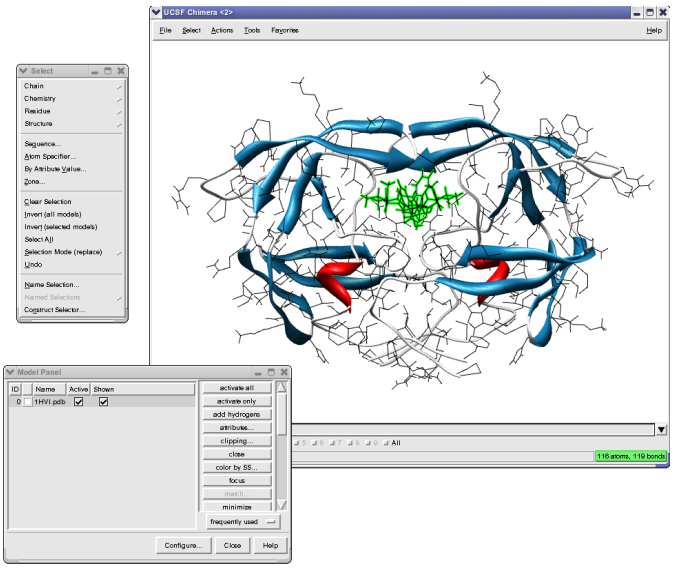
Changing bonds and atoms display. Selections

The objective of this part is to display only the ligand, the active site water molecule and residues 25 and 50 of the HIV-1 protease. Other residues will be hidden to clarify the figure. This will require to select parts of the structure and apply them different representation schemes.

There are three ways of selecting atoms with Chimera: using the "**Select**" menu, using the "**Command Line**", or **using the mouse** to select atoms from the screen. The latter will be described later.

Selections using the select menu

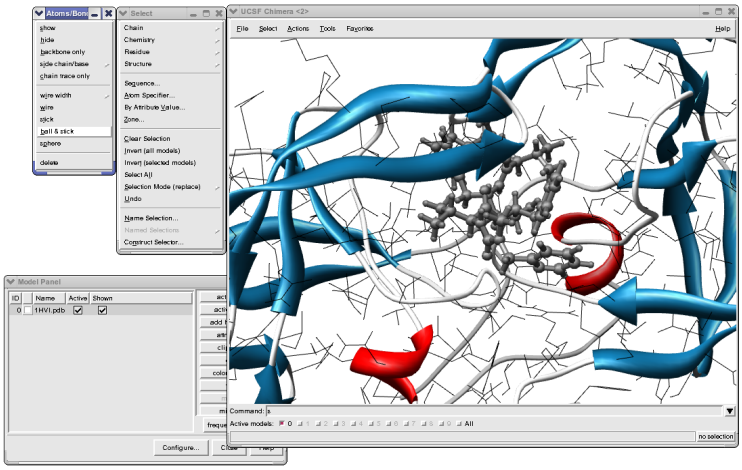
Open the "**Select**" menu and detach it by clicking on the upper dotted line (if this option is available on your operating system). Verify that the "**Selection Mode**" is "**replace**". Otherwise, choose it.
The first four menu items allow to select part of the structure according to the "**Chain**" (protein chains), the "**Chemistry**" (chemical nature of the atoms or functional groups), the "**Residue**" (residue name or type) and the "**Structure**" (ligand, water, ions, backbone, side chains, secondary structure, etc...). You can try to select different parts of the complex. For instance, you can select the ligand by choosing the "**Select>Structure>ligand**" menu item. You will see that the selected parts of the structures are surrounded by green thin lines.
Note that this HIV-1 protease inhibitor is also defined by its "ligand" nature or its residue name (i.e., A77). Therefore, it can be selected also by "**Select>Residue>A77**".



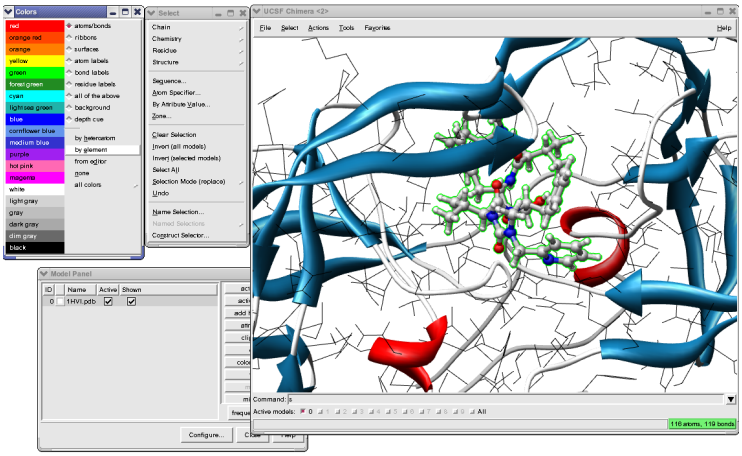
You can try selecting different parts of the complex: the water molecule, chains A or B of the protein, the aspartate residues, the strands, etc...
It is possible to append a selection with a new one. For instance, to select all the protein, you can choose the menu items "**Select>Selection Mode>append**", then "**Select>Chain>A**" and "**Select>Chain>B**".
Finally, you can clear all selections using "**Select>Clear Selection**".

Changing bond display and color

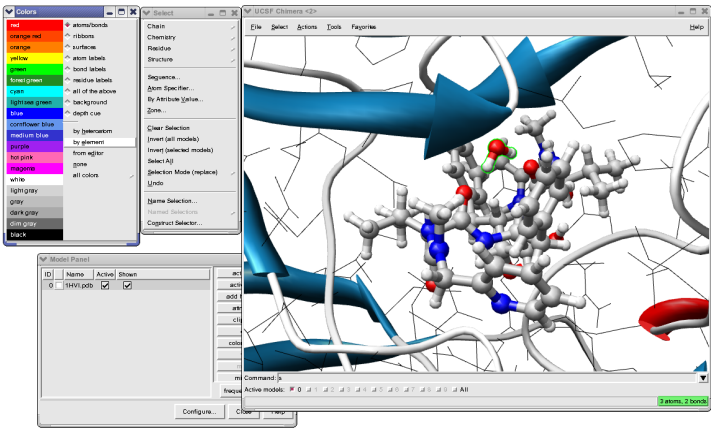
Select the "**replace**" mode for selections using "**Select>Selection Mode>replace**" and select the ligand.
Open the "**Actions>Atoms/Bonds**" menu and detach it by clicking on the upper dotted line. Choose successively the "**stick**", **ball & stick**", "**sphere**" and "**wire**" menu item sand see how the ligand is displayed in each case. Finally, choose the "**ball & stick**" representation.



Open the detachable color action menu using "**Actions>Color>all option…**". You can possibly close some other windows to save space on the screen. Check the "**atoms/bonds**" button so that the color changing will be applied only to atoms and bonds. Then, you can select a color from the left column that will be used for all ligand atoms. Another possibility is to color all atoms according to their atom types. This is obtained by clicking "**by element**".



Similarly, select the water molecule, display it in ball and stick and color it according to the atom types.



Selections using the command line

If it was detached, close the "**Select**" window using the upper right "**X**" icon. If it's not already done, open the command line by choosing "**Favorites>Command line**". A line preceded by "Command" will appear on the lower part of the principal window. This line can be used to type and execute commands relative to selection and display, for instance. To execute a command, one needs to type it in the command line and press Return. The union (logical 'OR') and intersection (logical 'AND') selection keywords are noted "|" and "&", respectively. The negation symbol is "~".
A detailed description of the different selection possibilities using the command line are provided in the quick reference guide. Here is a limited list showing some possible selections based on our particular structure that you could try:

"select" : select everything.
"select :.A" : select chain A of the protein.
"select :.A,.B" : select chains A and B of the protein.
"**select ligand**" : select the ligand.
"**select :A77**": select the residue named A77. Another way to select the ligand.
"**select :HOH**" : select the residue named HOH, i.e. the active site water molecule in our case.
"s**elect :HOH | ligand**" : select both the water molecule and the ligand.
"s**elect :25**" : select all residues that are numbered 25 in the PDB. One in both chains.
"**select :25@CA**": select atom Cα of residues 25.

Selections using the mouse

Using the menus or the command line, display all the atoms in the system, but hide the water molecules.

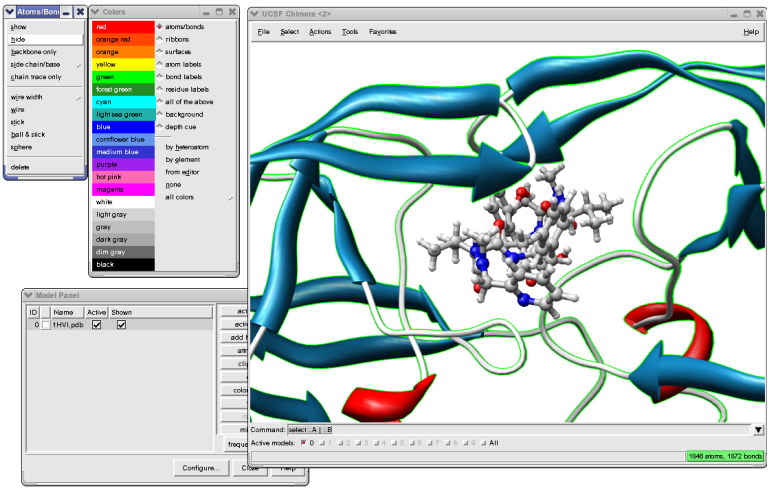
You can select an atom by pressing the "**Ctrl**" key on the keyboard and click on the atom.

Once an atom has been selected, pressing the "**Arrow up**" key of the keyboard will select the residue to which the atoms belong. Pressing a second time the "**Arrow up**" key will select the corresponding protein. You can continue until you select the whole system.

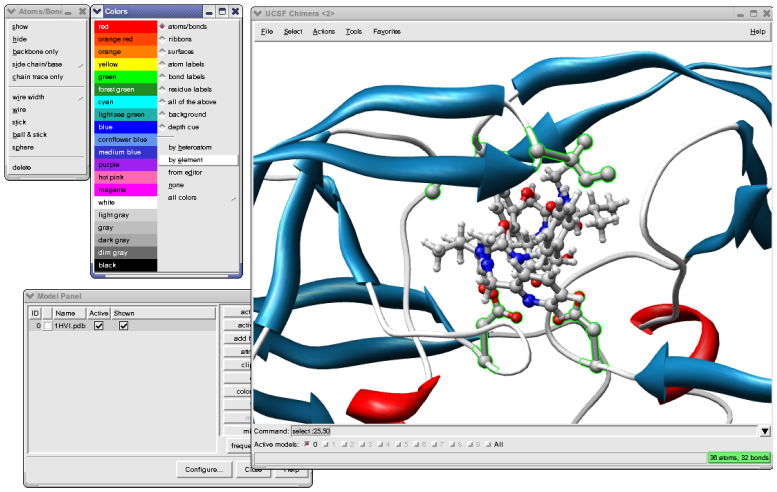
Pressing the "**Arrow down**" key of the keyboard will successively narrow down the selection till the initially selected atom.

You can select several atoms this way: select the first atom by clicking on it while pressing the "**Ctrl**" key on the keyboard. Then press both the "**Ctrl**" and "**Shift**" keys on the keyboard and click on the second atom.
You can select as many atoms as you wish.

Select all the protein by clicking on "**Select>Residue>standard amino acids**", and hide the atoms and bonds using "**Action>Atoms/bonds>hide**".

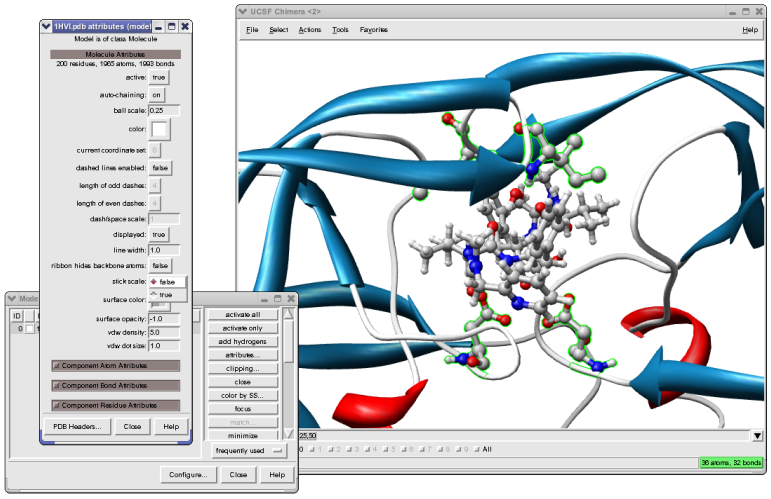


Select residues 25 and 50 by typing "**select :25,50**" in the command line, and show them in ball and stick representation, colored according to the atom type: "**Action>Atoms/Bonds>show**", "**Action>Atoms/Bonds>ball & stick**", "**Actions>Colors>by element**".



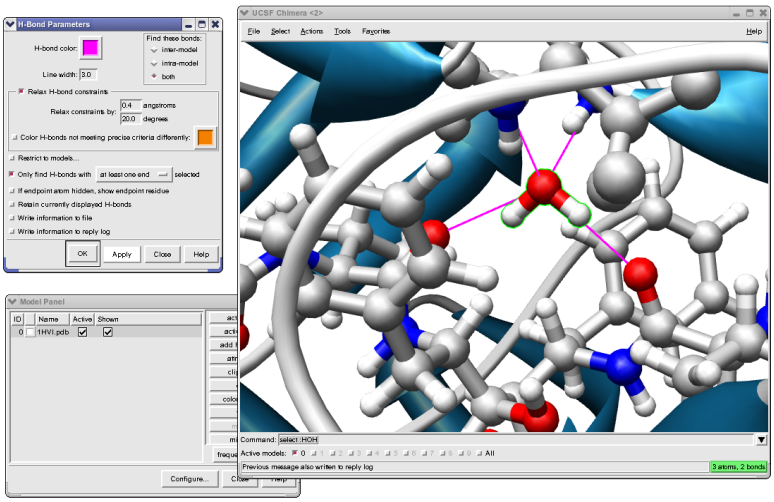
Changing ribbon attributes

As can be seen, the backbone atoms of the protein residues are not shown, while the bond lengths between the backbone (displayed as a ribbon) and C_B atoms are artificially too long. This is due to the default behavior of Chimera that hides backbone atoms when ribbons are displayed. This can be modified by changing some "attributes" of the structure. From the "**Favorites>Model Panel**" menu, open the "**attributes**" window, and choose the "**false**" value for "**ribbon hides backbone atoms**".



Calculating and showing hydrogen bonds

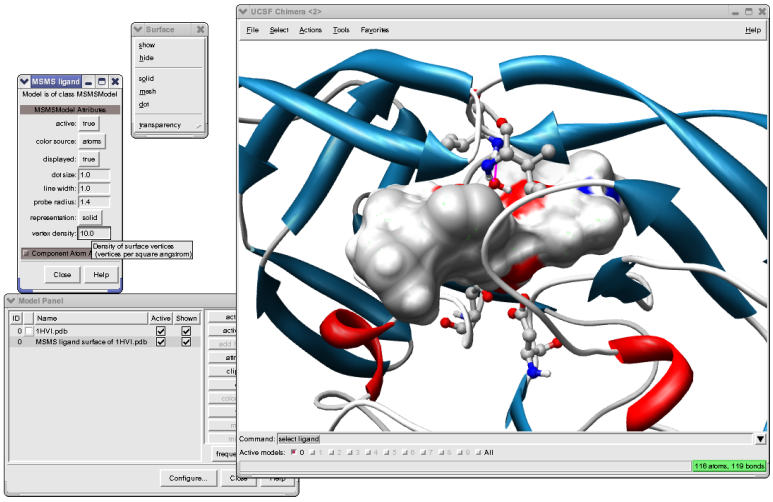
Chimera can calculate and display the hydrogen bonds between selected atoms. To show the hydrogen bond network between the ligand, the active site water molecule and the Ile50 backbone atoms, the following sequence of actions can be used. Select the active site water molecule (using the command line or the "**Select**" menu). Open the "**Tools>Surface/Binding Analysis>FindHBond**" window. Check the "**Only find H-bonds with**" box, so that it will calculate only hydrogen bonds involving the active site water molecule. Finally, click on the "**Apply**" button.



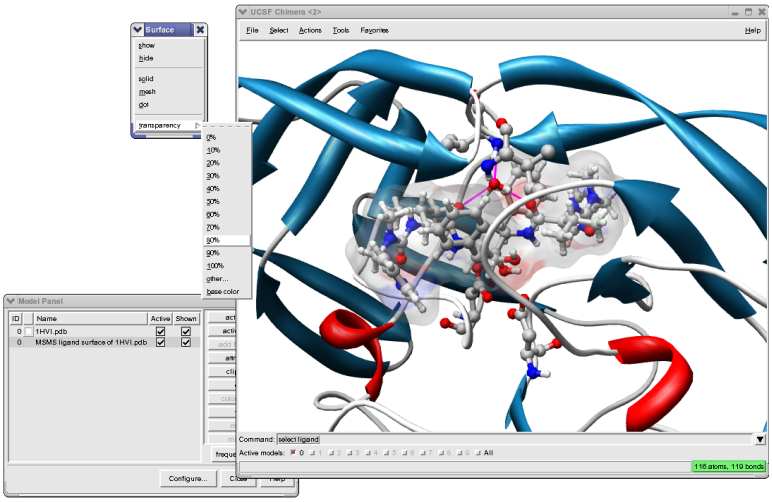
You can change the hydrogen bond color and line width from the "**H-Bond Parameters**" window, and click "**Apply**" to apply them.

Showing the ligand surface

Select the ligand. Then, click "**Actions>Surface>show**". You can try modifying the surface representation to "**mesh**" and "**dot**". Then, go back to the "**solid**" representation.
The surface quality may be increased by changing the corresponding attribute. In the "**Favorites>Model Panel**" window, select the MSMS ligand surface in the left list. Then click "**attributes**" and change the "**vertex density**" attribute to 10.0. Press Enter to apply the change.

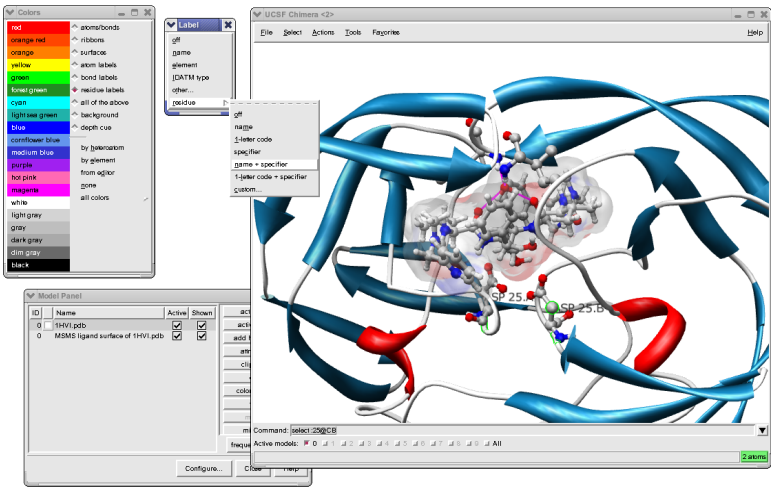


This surface gives a good idea of the volume occupied by the ligand. However, it also hides the molecule. To correct this, it can be made transparent. In the "**Action>Surface>transparency**" menu, select 70%.



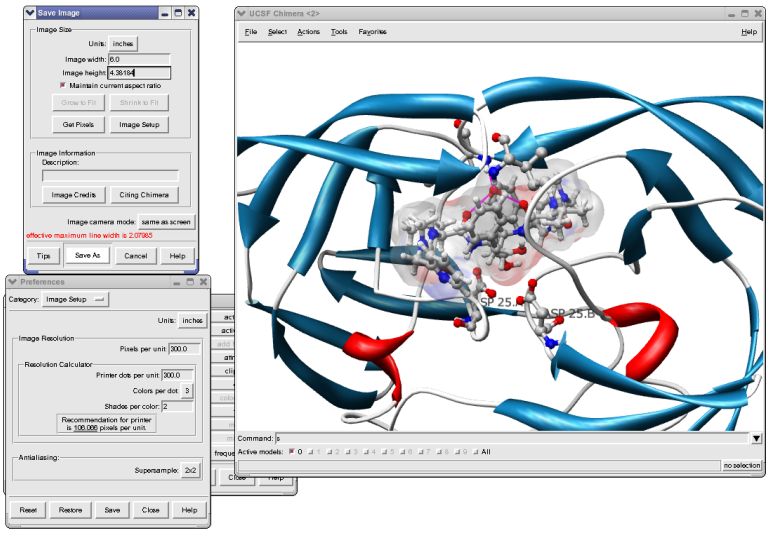
Label residues

Select atoms Cß of residues 25 by typing "**select :25@CB**" in the command line. Choose "**Actions>Label>residue>name + specifier**". You can try other types of labeling. Possibly, you can change the label color using the "**Color**" window, checking "**residue labels**" and choosing a color.



Saving image

Clear all selections using "**Select>Clear Selection**". Choose an orientation and a zoom that provides a satisfying point of view. Then, select the "**File>Save Image…**" menu item. In the new window, choose an image resolution of 300 dots per unit (dpi). Select "**Maintain current aspect ratio**", and enter an "**Image width**" of 20 centimeters. Select the format of the file that will be saved in the "**File Type**" menu. Click "**Save As**". Wait while the image is calculated. Scientific journals generally accept TIFF, JPEG and EPS files. Select TIFF in this case. Finally, choose and "**File name**" and click "**Save**".



Images are saved as RGB figures.

> **Please note that in some bugs were reported for the generation of images with the Windows version of UCSF Chimera. If the picture generated is all black, please try to set "print unit" to "points" and adapt the "Image width" to 225.**
> **You may also try with another output format (PNG, for instance).**

Saving the session status

The Chimera session (the actual representation) can be saved for future use or modifications. This can be made using the "**File>Save Session As...**" menu. Select a file name and click "**Save**". The saved file is actually a python script and will have the ".py" extension.

Closing the session. Quitting Chimera

The session can be closed using the "**File>Close Session**" menu item. One can quit Chimera with "**File/Quit**".

Restoring a previous session

A previously saved session can be restored using the "**File>Restore Session**" menu item. Select the file (with a .py or .pyc extension) and click "Open".

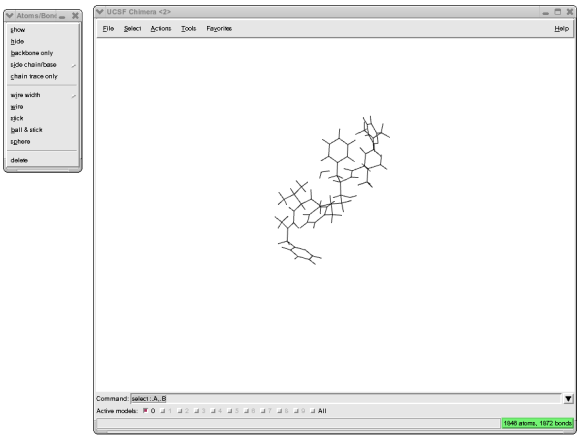
Exercise 2

Loading the macromolecular structure

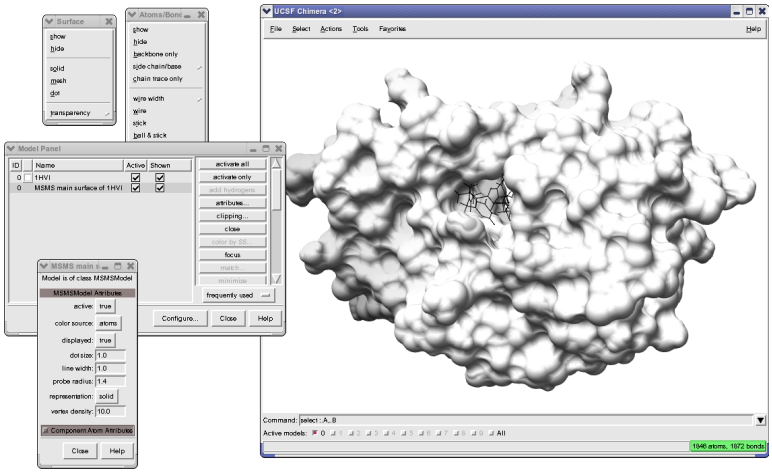
Load the 1HVI structure file into Chimera using one of the two methods seen in exercise.

Showing the molecular surface of the protein

First, hide the bonds and atoms of the protein chains and keep only the ligand and active site water molecule visible. To do so, select the two protein chains, for example by choosing "**Select>Residues>standard amino acids**". Then choose "**Action>Atoms/Bonds>hide**".

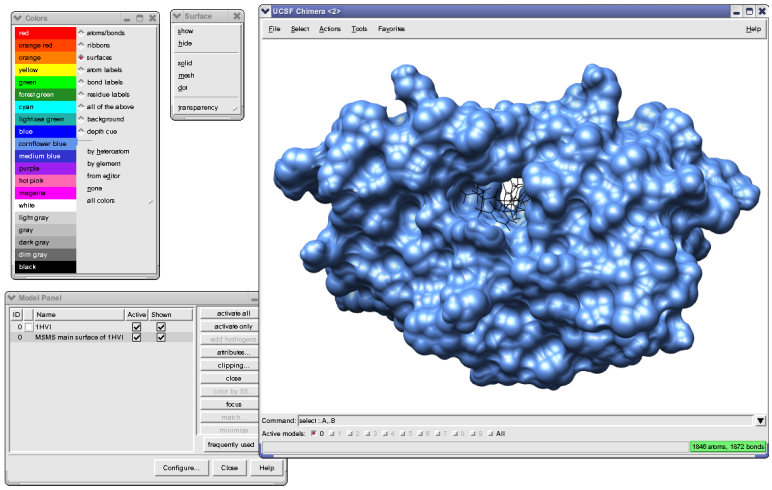


Be sure that the protein is still selected. If necessary, or in case of doubt, select it again as described above. Then, choose "**Action>Surface>show**". Open the "**Favorites>Model panel**" menu, select the "**MSMS main surface**" in the left list, then click "**attributes**" and change the "**vertex density**" to 10. If your computer is too slow, it might be necessary to reduce the vertex density to 5.



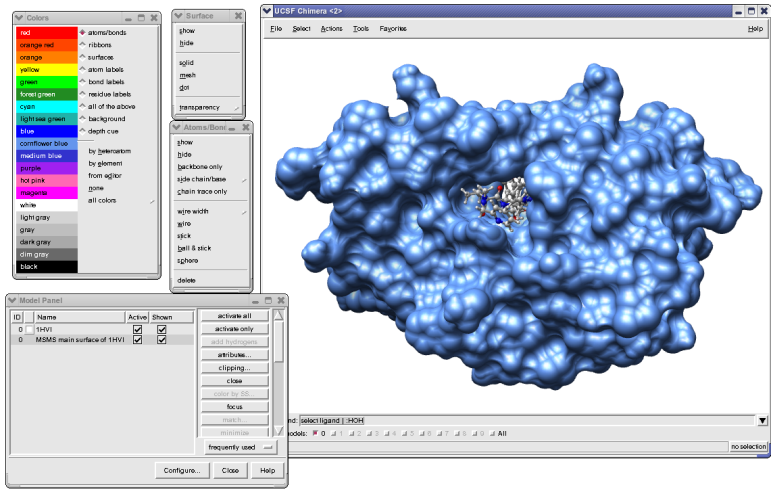
Change the surface color

Be sure that the protein is still selected. Open and detach the "**Actions>Colors>all options…**" menu. Check the "**surfaces**" button, and click on the color you would like to apply to the protein surface. The following image has been obtained using the cornflower blue color.



Changing bonds and atoms display

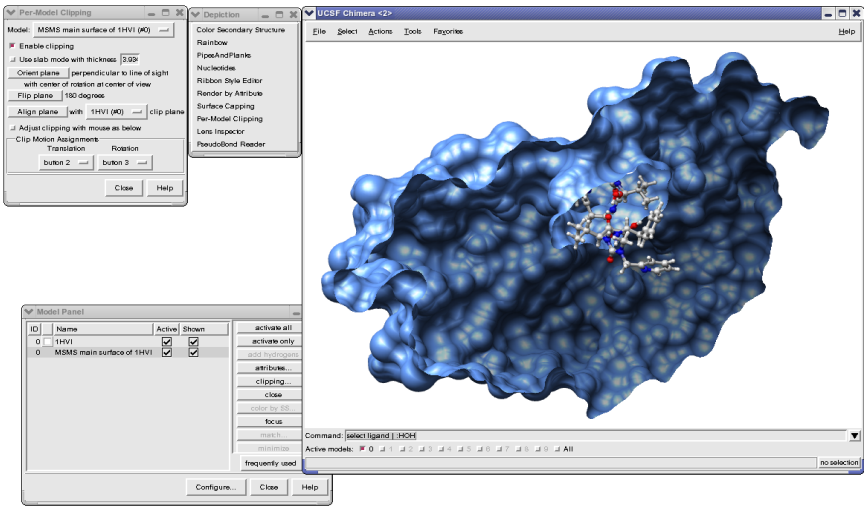
Select the ligand and water molecules, then display them in ball and stick representation, colored according to the atom type.



Clipping the protein

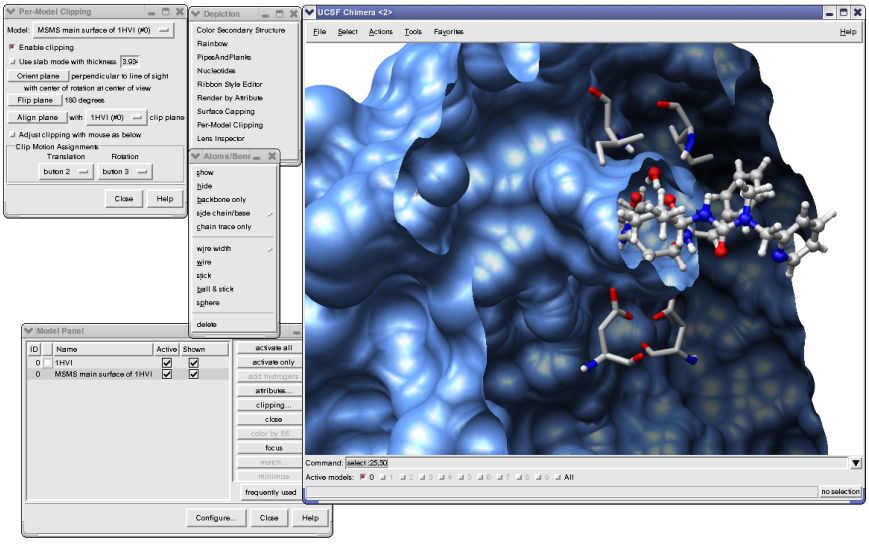
It is possible to cut the protein surface in to get a better view of the ligand binding mode inside the binding site. To do so, open the "**Tools>Depiction>Per-model Clipping**" menu. In "**Model**", select the "**MSMS main surface**" and check the "**Enable clipping**" box.
To begin with, we will remove the capping of the surface. For this, click on the **"Surface capping…"**

button, and uncheck the **"Cap surface at clip planes"** box in the new **"Surface capping"** menu. We will reactivate the capping in the next section.
Rotate the structure to get a better view of the clipping effect.

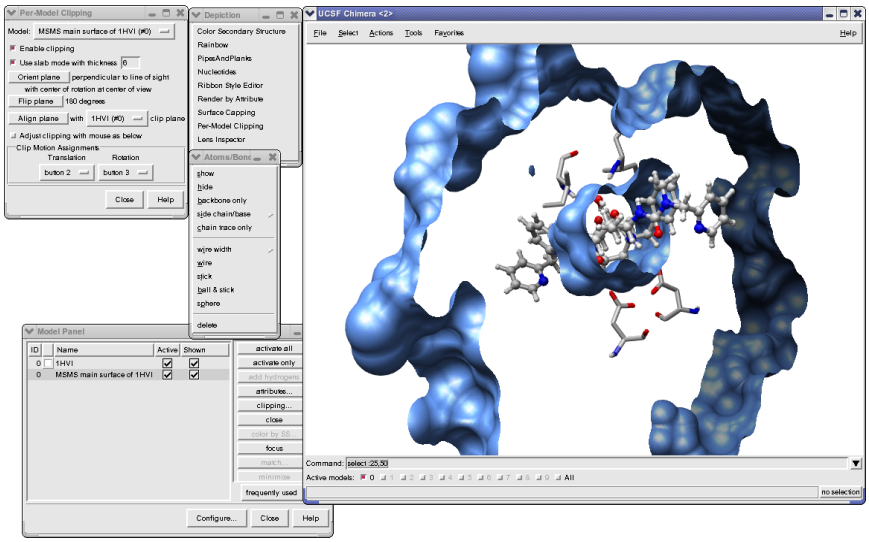


It is possible to manipulate the clipping position using the mouse. Check the "**Adjust clipping with mouse**" box. A click/drag of the mouse central button will control the translation of the clipping plane. A click/drag of the mouse right button will modify its orientation. Note that this cancels any possibility of protein translation or zoom using the mouse. You can get back to the default behavior of the mouse by unchecking the "**Adjust clipping with mouse**" box.

You can select and display residues 25 and 50 of the HIV-1 protease by typing "**select :25,50**" in the command line, and then display them in the stick representation. Then change the position and orientation of the clipping plane to have a good view of the interactions between these residues and the ligand.



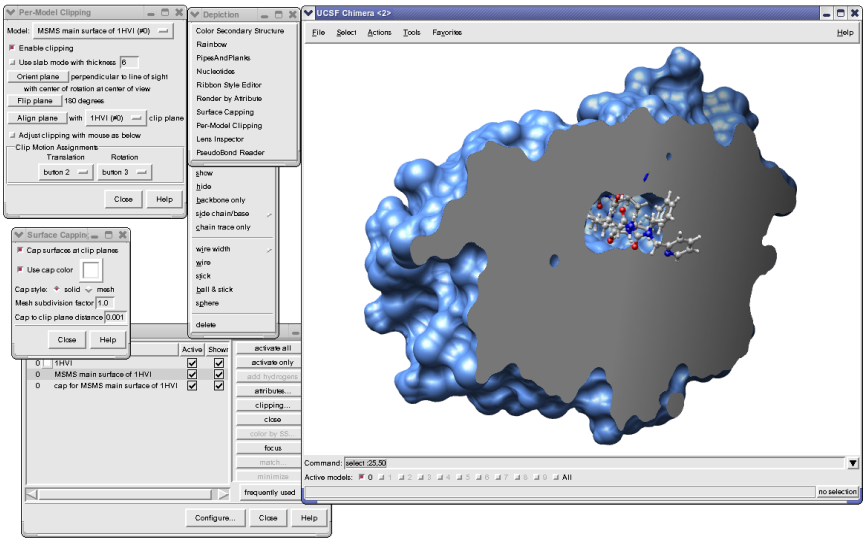
You can also get a good view of the shape of a buried binding site using the slab mode. In the "**Per-Model Clipping**" menu, check the "**Use slab mode with thickness**" and choose 6 for the thickness. This creates a protein slab that you can manipulate using the mouse by checking the "**Adjust clipping with mouse**" box.



Uncheck the "**Use slab mode with thickness**" box.

Surface capping

A clipped surface may be capped. To do so, choose the "**Surface Capping**" menu item in the "**Depiction"** window, or click on "Surface capping…" in the "**Per-Model Clipping**" menu **. In the "Surface Capping**" window, check the "**Cap surface at clip planes**" box. You can change the color of the capping plane by checking the "**Use cap color**" box and choose the color by clicking the colored square next to it.



This capping can also be applied with the slab mode.

Save an image as described in the previous exercise, then save the session and close it.

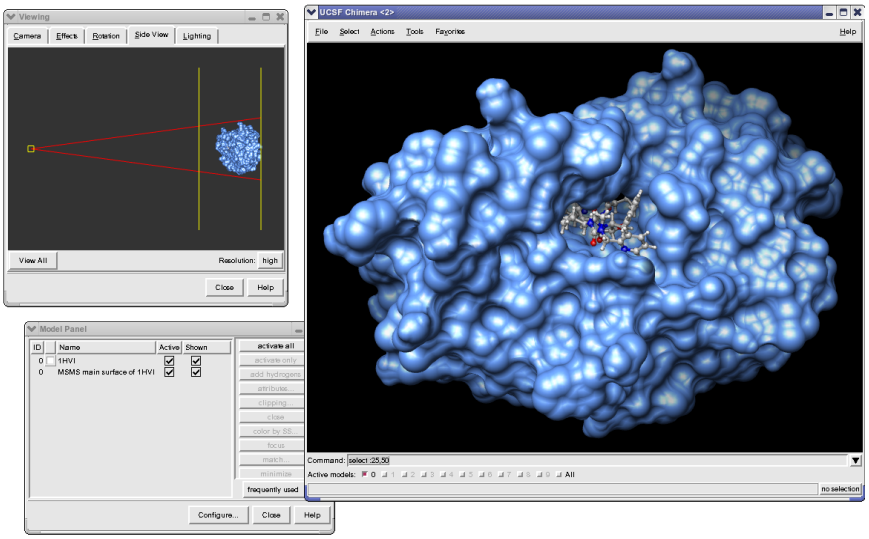
Exercise 3 (optional)

Open a previous session

Open the final state of exercise 2 using the "**File>Restore Session**" menu. Removing all clipping and capping, and change the background color to black if necessary.

Using the side view

Open the "**Viewing**" menu by choosing the "**Favorites>Side view**" menu item. The "**Side View**" tab of the "**Viewing**" window should be active. Otherwise, click on the "**Side View**" tab. You will see a reduced view of the structure appearing in the "**Viewing**" window. The vertical lines show the clipping planes that define the region of space displayed on the screen. The square gives the viewer's eye position. The red lines show the field of vision.
Clicking the "**View All**" button adjusts the scale and clipping plane positions so that the view will include everything that is displayed. It is possible to move the position of the viewer's eye and clipping planes using the mouse.

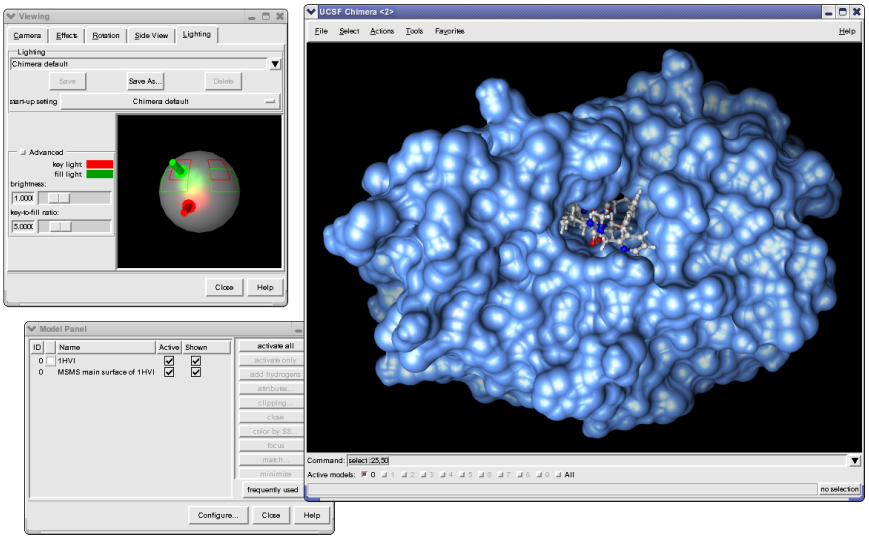


If you drag the rear clipping plane closer to the protein, you will see the effect of the depth cueing, which causes regions farther from the viewer to be shaded. The depth cueing parameters can be changed in the "**Effects**" tab of the "**Viewing**" window.

Changing the Lighting

It might be useful to change the lighting (intensity and direction) to get a better view of the system. Select the "**Lighting**" tab in the "**Viewing**" window. This tab displays the light sources and parameters. The key light is the dominant brighter source of light. The fill light gives a secondary source. The solid arrows in the right view allow manipulating the lighting directions with the mouse.

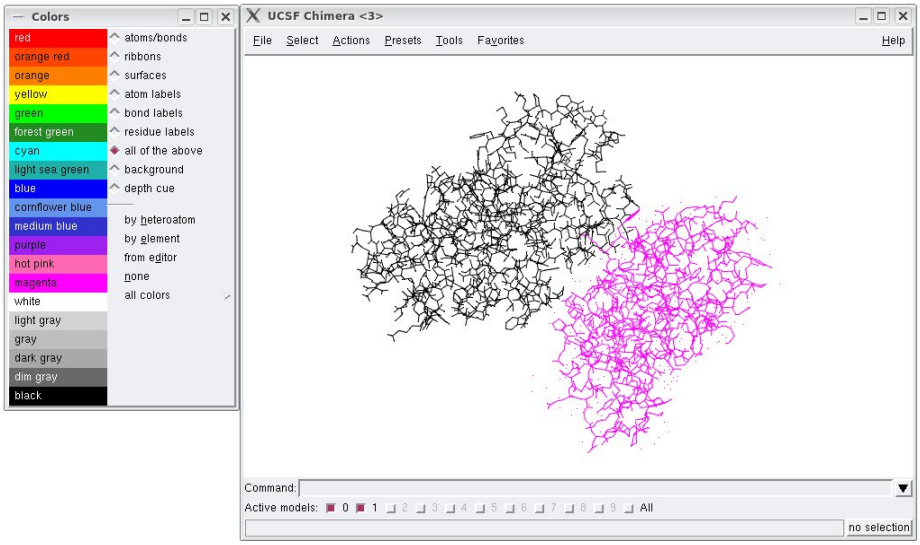
You can try different lighting directions to see their effect.



Exercise 4

Loading the macromolecular structures

Load the 1HVI structure file into Chimera using one of the two methods seen in exercise 1. Then open the 1HII structure file. These structure files correspond to the HIV-1 and HIV-2 proteases, respectively.



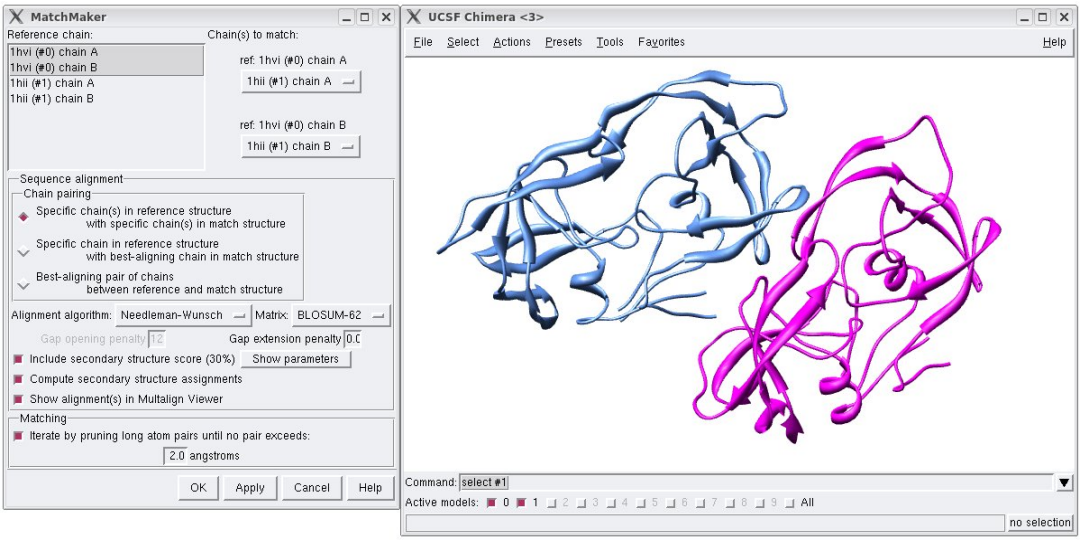
They are now two active models mentioned in the lower line of the main Chimera window. Model 0 is HIV-1 protease (1HVI), whereas model 1 is HIV-2 protease (1HII). Display both molecules using different colors. For this, each model can be selected subsequently by typing "**select #0**" and "**select #1**" in the command line.

Showing the proteins as ribbons

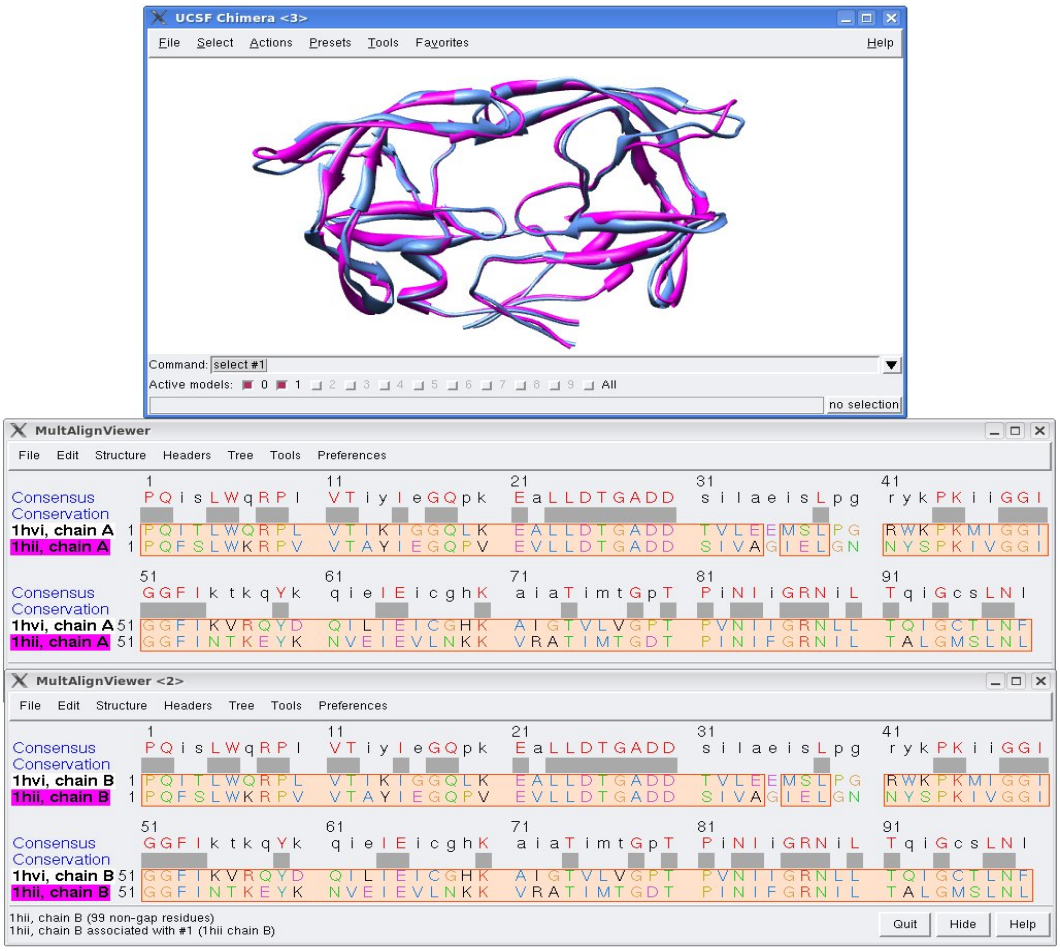
Select all heteroatoms and water molecules by typing "**select :.het | :HOH**" in the command line, and delete the corresponding atoms using "**Action>Atoms/bonds>delete**". Only the two proteins are remaining in Chimera. Display them in ribbon representation, as explained in exercise 1. Hide bonds for both molecules. Recalculate the secondary structure elements by typing "**ksdssp**" in the command line.

Match the two molecules. Obtain a structural alignment

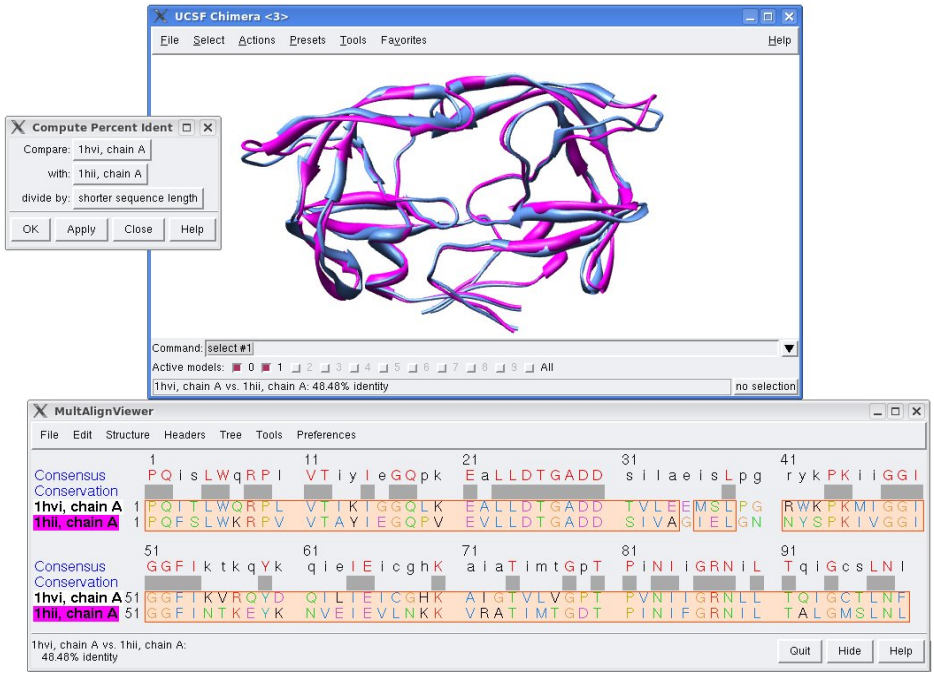
Open the structural alignment tool using "**Tools>Structure Comparison>MatchMaker**". Click on "**Specific chain(s) in reference structure with specific chain(s) in match structure**". Then, subsequently, select chains A and B of 1HVI as reference structures, and match them with chains A and B of 1HII, respectively. Verify that the "**Show pairwise alignment**" box is selected. Finally, select the "**Needleman-Wunsch**" algorithm and click "**OK**". This will perform a matching of the two molecules, based on their sequence and 3D structure similarities.



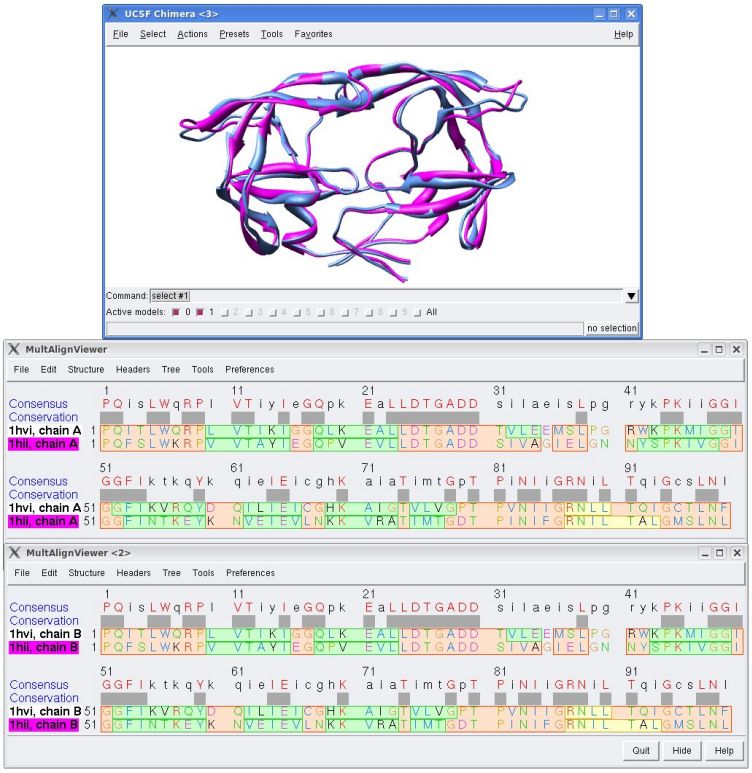
After a short moment, the two molecules should be superimposed. A new window, called "**MultiAlignViewer**" will also appear, showing the corresponding sequence alignments.



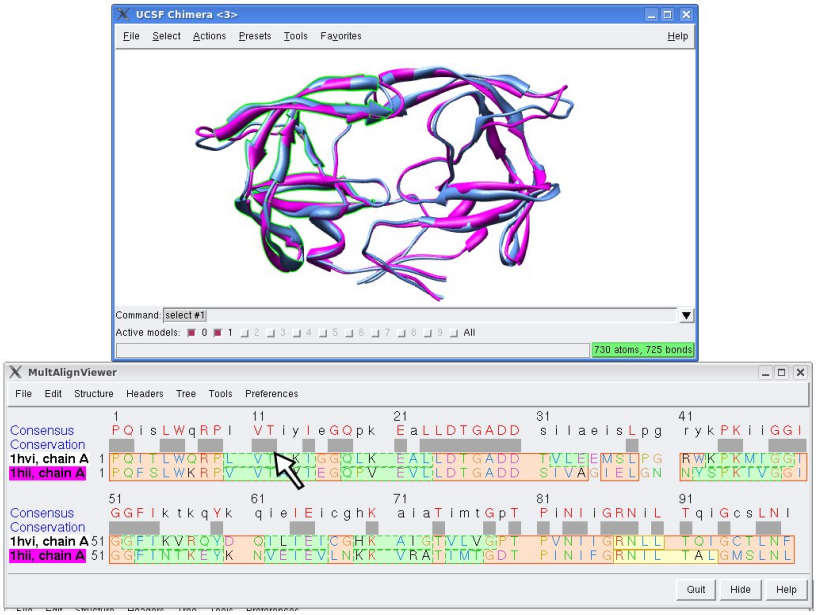
The percentage of sequence identity between the two molecules can be obtained using the menu "**Info>Percent identity...**" of the MultiAlignViewer window. It is possible to show the conservations between the two sequences by selecting "**Headers>Conservation**" in the MultiAlignViewer window.



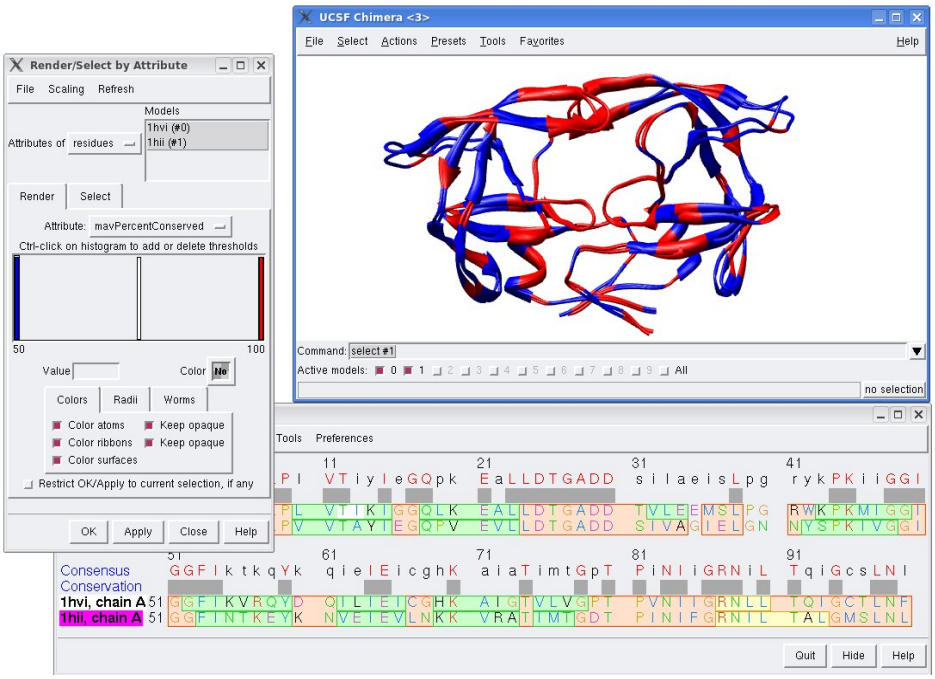
The secondary structure elements can be highlighted in the sequence using the MultiAlignViewer command: "**Structure>Secondary Structure>show actual**". Secondary structure elements appear in green for ß-strands and yellow for -helices.



Isolated residues or groups of residues can be selected by clicking over the sequence or in the "**Conservation**" line of the MultiAlignViewer, allowing the analysis of the protein's structural differences or analogies.



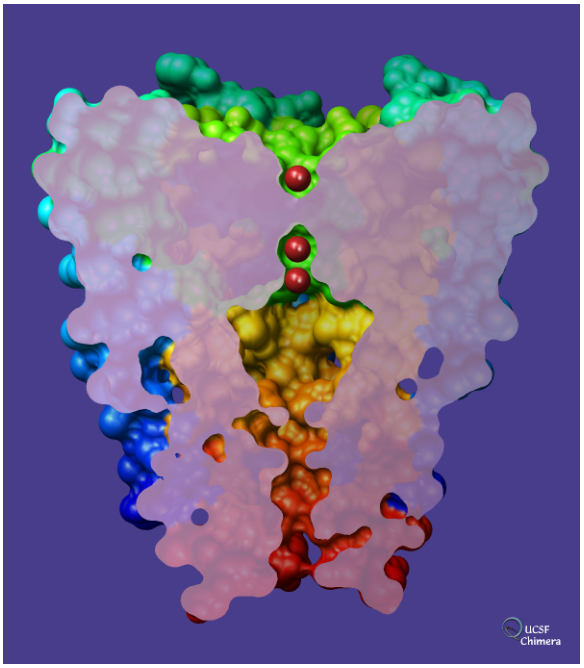
It is possible to color the structures according to the percentage of conservation in the sequence alignment. For this, open the "**Structure>Render by Conservation**" menu of the MultiAlignViewer, then click "**Apply**" using the default values for the different options. In the present case, the most conserved regions are colored in red and the less conserved ones in blue.



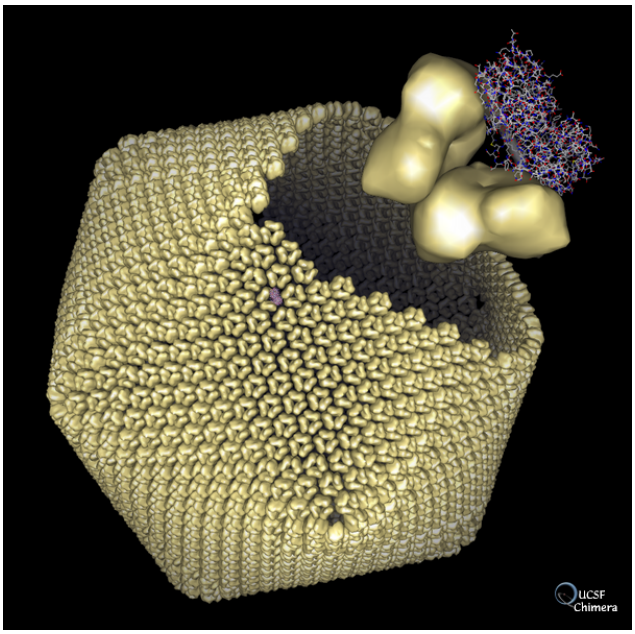
Going further with UCSF Chimera

These exercises give only a very limited overview of what Chimera is able. You can find a detailed documentation, as well as some tutorials, at the following address:
http://www.cgl.ucsf.edu/chimera/docindex.html
Here are some examples of images produced using Chimera that were taken from the official Website.

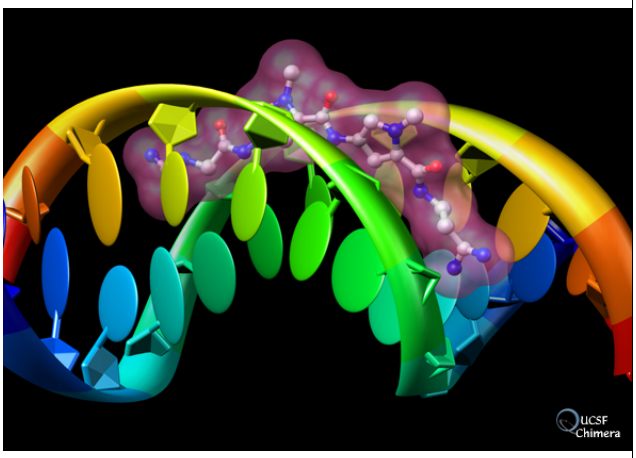
Sliced potassium channel



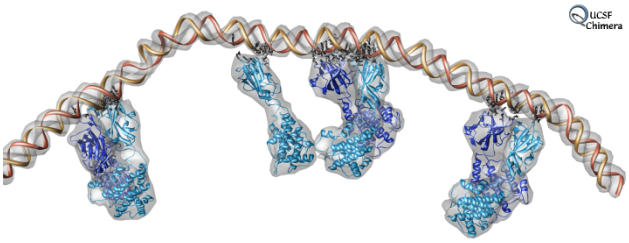
Paramecium Bursaria Chlorella Virus



DNA and Netropsin



Bluetongue Virus and Viral RNA



Practice Session 2:
Ligand-protein docking

Antoine Daina, Ute Röhrig, Vincent Zoete – 2021
Contact : vincent.zoete@sib.swiss or antoine.daina@sib.swiss

Exercise 5. Docking of anti-inflammatory drug Celecoxib into protein COX2.

✉*Don't forget to save* ***Chimera sessions*** *from time to time for easy recovering in case of mistake!*

First save: **>File >Save Session As…** Give a name to the .py file and click the **"Save"** button**.**
Save frequently: **>File >Save Session.**
To get the last session back: **>File >Restore Session…**

✉*For this exercise* ***Smart Display*** *option should be switched on:*

>Favorites >Preferences…
then in the Cartgory: **New Molecule**, **smart initial display: "true"** and click on **"Close".**

✉*For this exercise, you should have access to* ***Command Line***.

>Favorites >Command Line

- *Download the crystal structure of COX2 in complex with inhibitor Naproxen from the Protein Databank into Chimera*

>File >Fetch by ID…
In the pop-up window, **select PDB** and type "**3q7d"** in the **text field**
click the "Fetch" button.

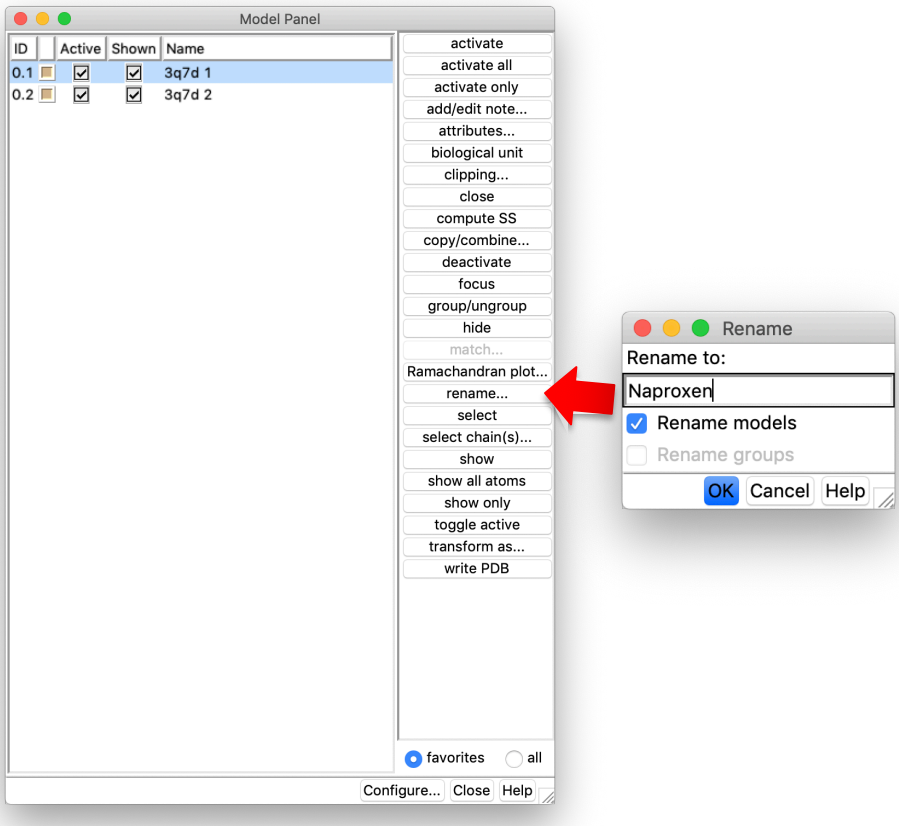
- *Remove the chains B of the protein*

>Select >Chain >B
>Actions >Atoms/Bonds >delete

- *Spot non-protein parts of the structure*
  >Select >Residue >all nonstandard
  Fly-over all highlightened items to capture the name of residues

- *Remove all crystallization adjuvants*
  In the Command Line, type `sel :BOG,NAG`
  >Actions >Atoms/Bonds >delete

- *Color, display and extract the molecule NPX (Naproxen) from the cocrystal and rename the protein (COX2) and the ligand (Naproxen).*
  In the Command Line, type: `sel :NPX`
  >Actions >Color > by element
  >Actions >Atoms/Bonds > ball & stick
  In the Command Line, type: `split atoms :NPX`
  In the **Model Panel (>Favorites >Model Panel),** select on the first line **(0.1)** and click on the **"rename…"** button on the left**.**
  In the Rename window**,** type **"Naproxen"** and then click **"OK".**
  In the **Model Panel** select on the second line **(0.2),** click on the **"rename…"** button and type **"COX2"** and **"OK".**



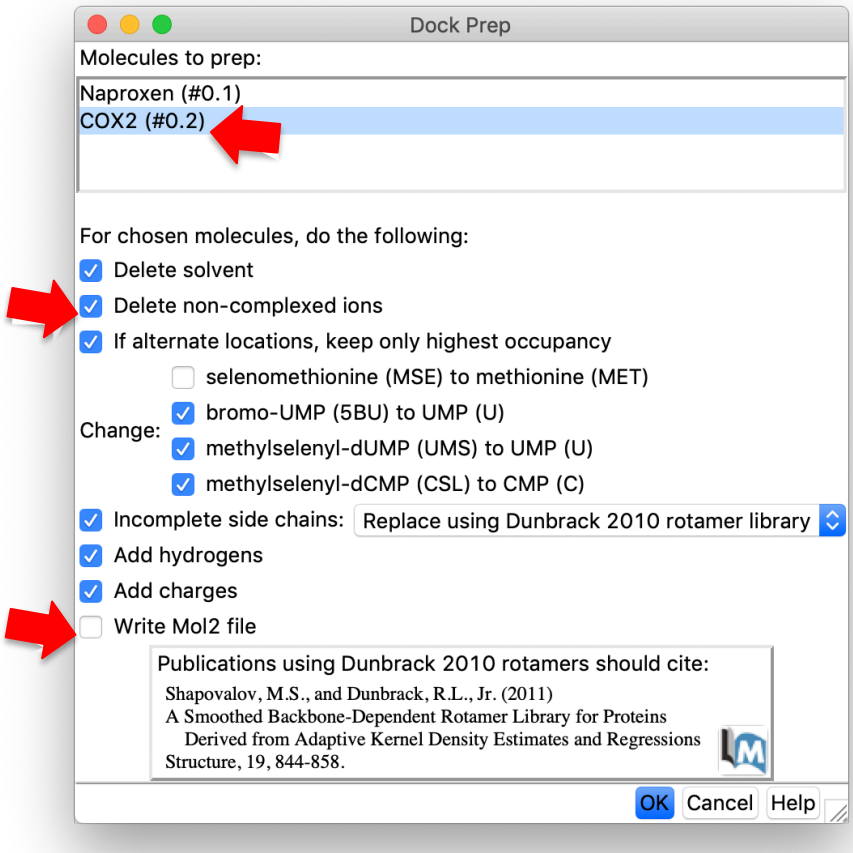
- *Generate the protein surface*
  >Select >Residue >standard amino acids
  >Actions >Surface >Show

  ✉At this point, one should have the protein COX2 in entry 0.2, its surface (MSMS) in 0.2 (on another line in the Model Panel) and the cocrystallized inhibitor Naproxen in entry 0.1.

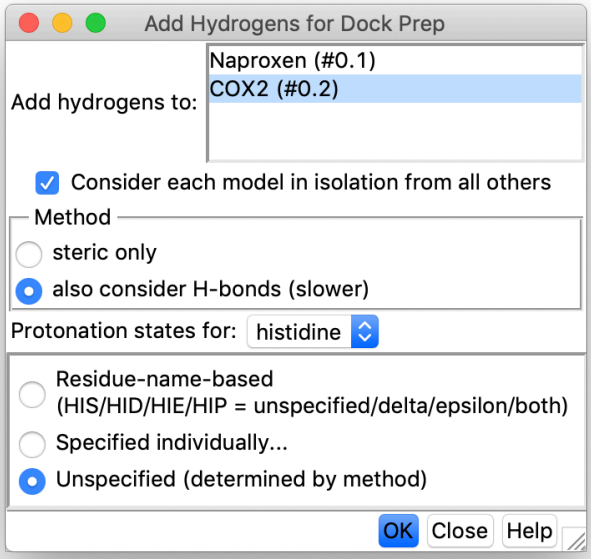
  You can now visualize easily the binding pocket of the heme. Try to visualize the more buried binding site of the inhibitors by locating the Naproxen.

- *Prepare the COX2 structure for docking*
  >Tools >Surface/Binding Analysis >Dock Prep

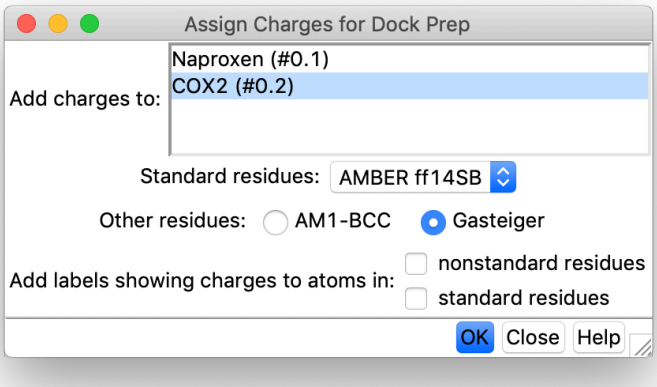
  - In the **first panel "Dock Prep":**
    Select **only** the protein entry **COX2 (#0.2)** to be prepared in the **"Molecules to prep" list.**
    Tick all boxes except "Write Mol2 file".
    Make sure that the **Incomplete side chains will be replaced** using the **"Dunbrack 2010 rotamer library".**
    Click **"OK"**



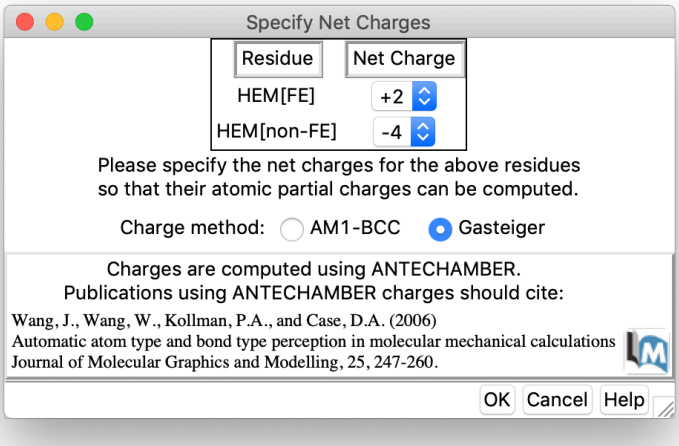
- In the **second panel "Add Hydrogens for Dock Prep":**
  Select the protein to be prepared in the **"Add hydrogens to" list**
  Add hydrogen atoms **also considering H-bonds**
  Treat histidine **unspecified (determined by method)**
  Click **"OK".**



- In the **third panel "Assign Charges for Dock Prep":**
  Select the protein to be prepared in the **"Add charges to" list**
  Use **AMBER ff14SB** charges for standard residues
  Use **Gasteiger** charges for non-standard residues
  Let both label boxes unchecked.
  Click **"OK".**



- In the **fourth panel "Specify Net Charges":**
  Assign a charge of **+2 to HEM[Fe]**
  Assign a charge of **-4 to HEM[non-Fe]**
  Use **Gasteiger** method of calculation.
  Click **"OK".**



- *Download Celecoxib from the PubChem database*

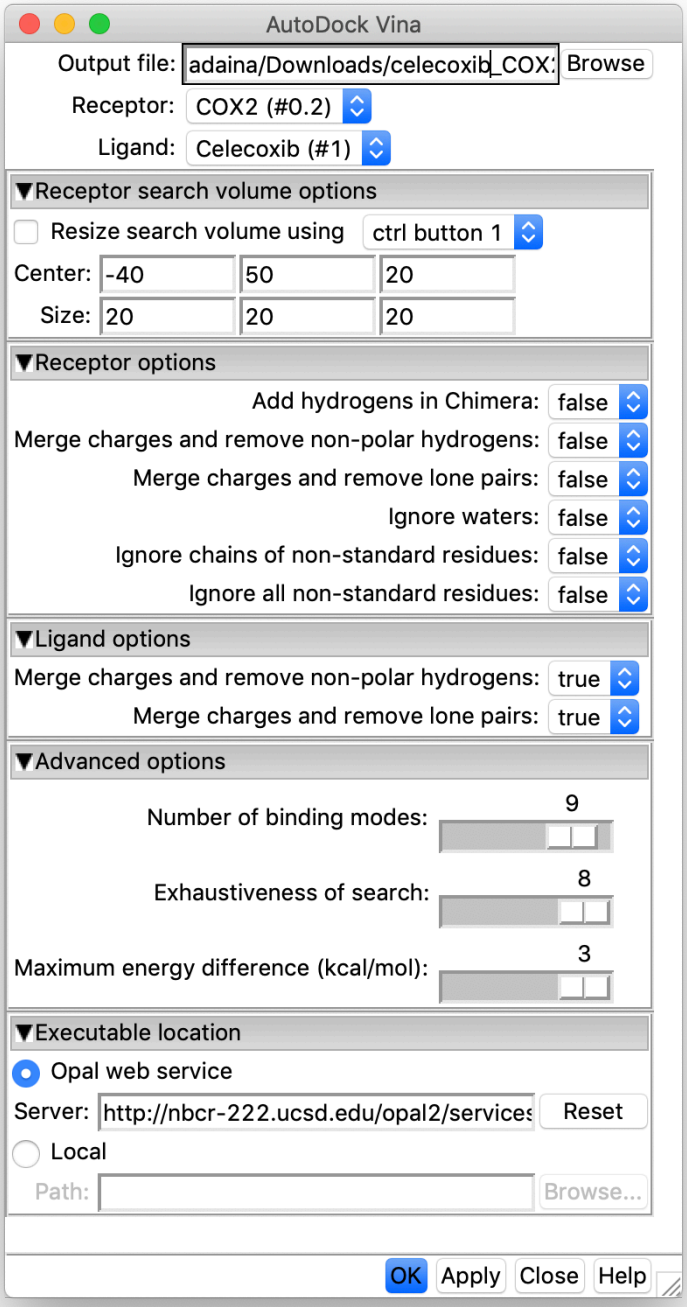
One needs a reliable structure of Celecoxib with a realistic 3D geometry in a manageable file format. This can be found in the **PubChem database**.

>File >Fetch by ID…
In the pop-up window, **select PubChem** and type Celecoxib ID code "**2662**" in the **text field click the "Fetch" button.**
In the **Model Panel (>Favorites >Model Panel),** select on the first line **(1)** and click on the **"rename…"** button on the left**.**
In the Rename window**,** type **"Celecoxib"** and then click **"OK".**

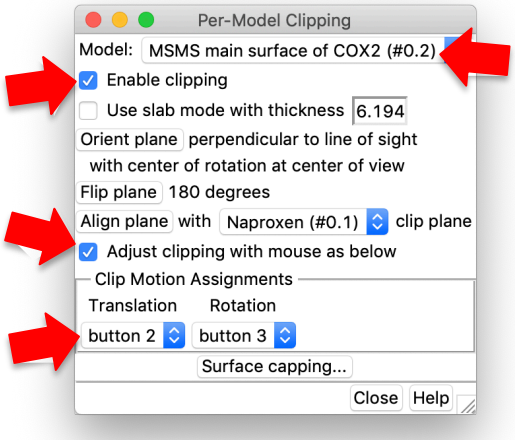
✉At this point, one should have the cocrystallized ligand Naproxen in **#0.1**, the target COX2 prepared for docking and its MSMS in **#0.2**, as well the Celecoxib to be docked in **#1**.

- *Setup and run docking (with VINA)*
  >Tools >Surface/Binding Analysis >AutoDock Vina

  - In the AutoDock Vina window, click the **"Browse"** button, navigate to a convenient location on your hard drive (you have the possibility to create a new folder), enter **"celecoxib-in-cox2.pdbqt"** in the **file name: text box** and click on **"Set Output Location"**.
  - In the AutoDock Vina window, select **COX2 (#0.2)** as **Receptor** and **Celecoxib (#1)** as **Ligand.**



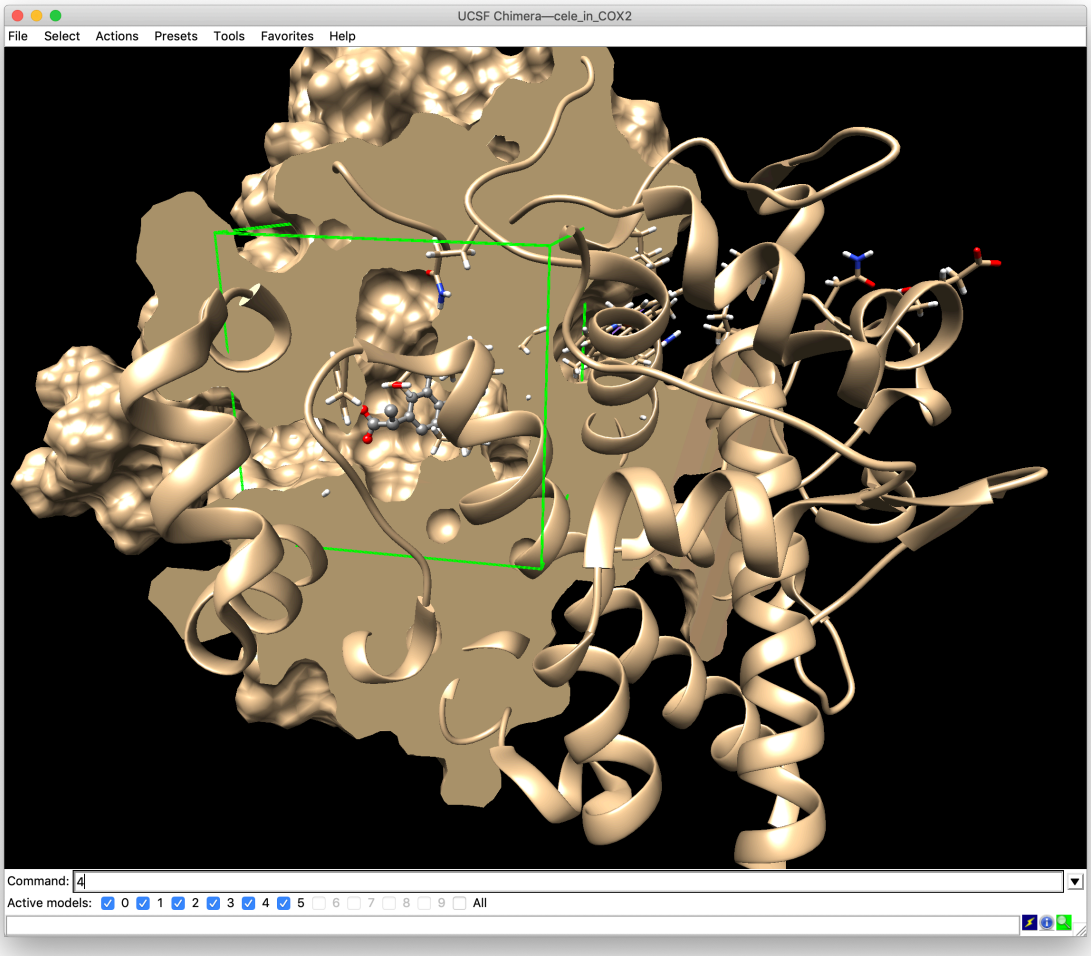
- **the Receptor search volume options** panel to define a box encompassing the binding site, where to focus the docking search.
  Enter random numbers in the **Center** boxes (x,y,z coordinates, eg. **1 1 1**).
  Enter **20 20 20** the in the **Size** boxes (side length in Å);
  Visualization of the pocket is easier by cropping the surface of COX2: **Tools >Depiction >Per Model Clipping**.
  Select the protein surface,
  tick "Enable clipping",
  tick "Adjust clipping with mouse as below"
  Select "**button 2**" or another.



- Play with mouse with e.g. **button 2 pressed and hold** to find the best cutting plan to visualized the binding site.
  Untick "Adjust clipping with mouse as below".

- Tick the box **"Resize search volume using** (eg.) **button 2**" in the **AutoDock Vina** Panel.
  By **clicking and holding** mouse button 2, one can **translate** the green box in the space. Place the cubic search region so that it includes the volume of the guessed binding pocket entirely

- Something like:



- Develop the **Receptor options** panel
  Set **all options** to "**false**" (because the receptor was already prepared through the Dock Prep procedure).
- Develop the **Ligand options** panel
  Set **both options** to "**true**"
-
- Develop the **Advanced options** panel
  Verify that the Number of binding modes is 9, the Exhaustiveness of search is 8 and the Maximum energy difference is 3 kcal/mol.

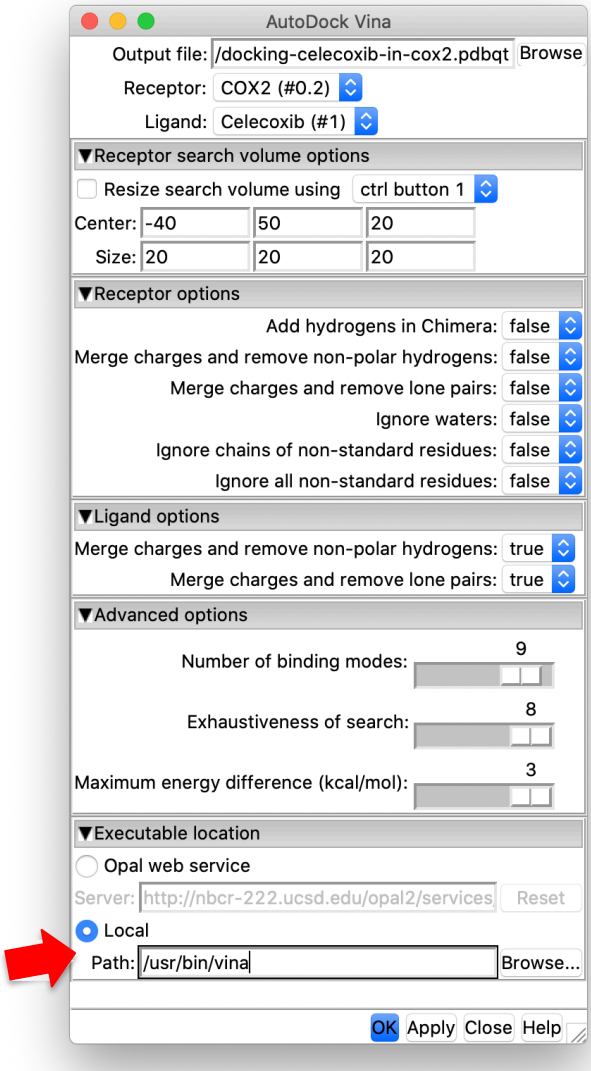
- Develop the **Executable location** panel
  The docking computation will be run *remotely* on an **Opal web service**. Make sure that the Server address is
  http://nbcr-222.ucsd.edu/opal2/services/vina_1.1.2

- Launch your docking! Click on **"Apply"**.
  An informative disclaimer pops-up, click "**Yes**".

- You'll most probably receive **the following error message** because the NCBR Opal web server was recently put out of service

- We can propose three workarounds (not shown in the video):

1. Run docking through a **local executable**, if you have installed the AutodockVINA on your computer.
   Enter the path to your local installation at the bottom of the last panel (instead of the Opal web service)

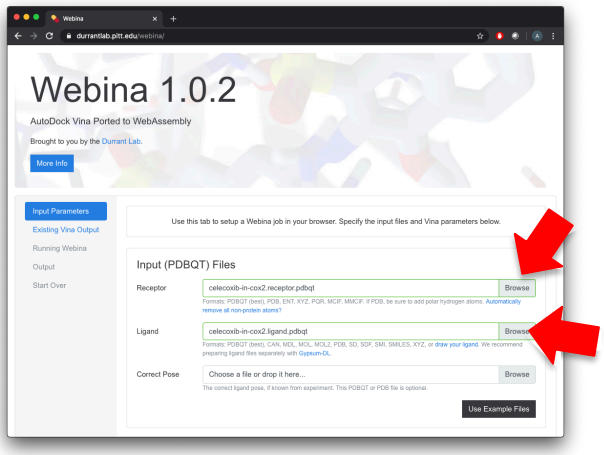


   - Launch your docking! Click on **"Apply"**.

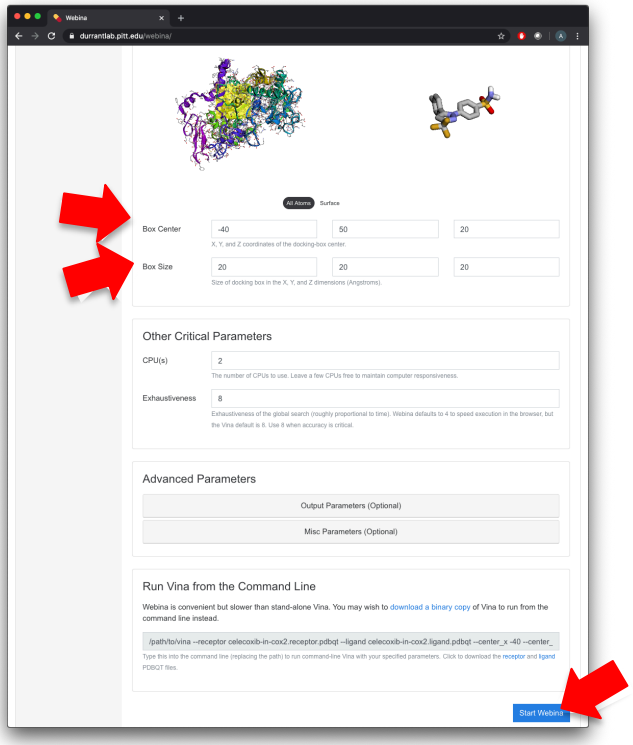
2. Download an **already calculated** VINA docking at http://drug-design-teaching.ch/download/docking-celecoxib-in-cox2.pdbqt Depending on your Web browser, the download is not automatic and the full text is displayed. If so, save the file on your hard-drive and keep the .pdbqt extension (not .txt).

3. Use **Webina**, a Web interface that performs local Vina docking without installation. Go to: https://durrantlab.pitt.edu/webina/

   - You have to **input the prepared receptor and ligand as PDBQT** files. Both files were created on your hard drive, when you have **"Set Output Location"** in Chimera.

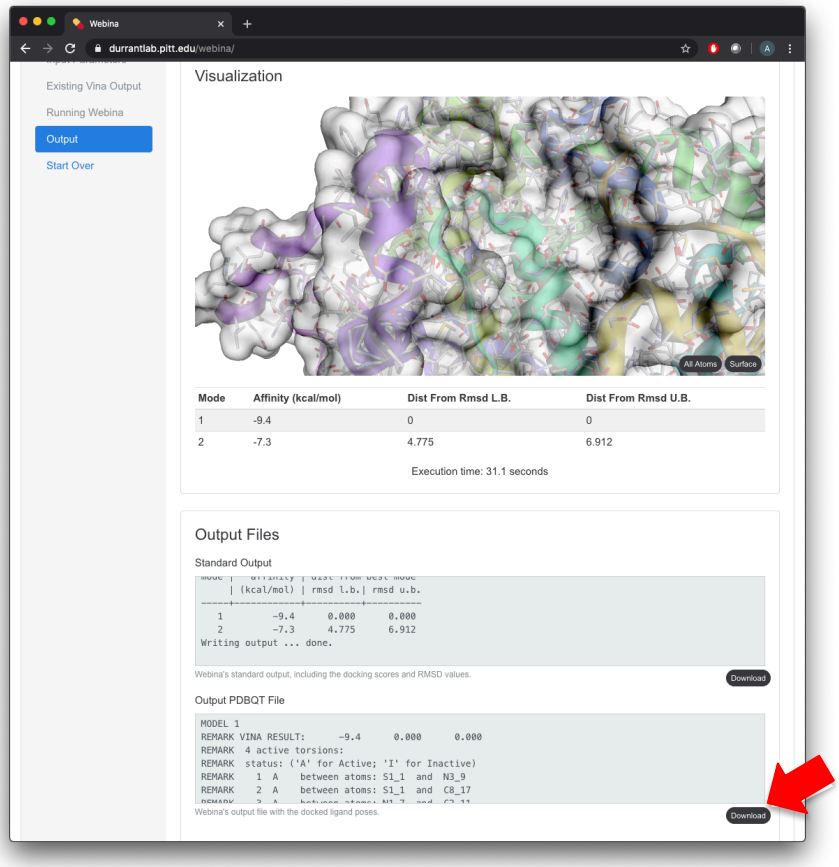


   - Define the Search Volume, as **Box center** and **Box Size**, the same way as set before in Chimera. Let all other parameters unchanged.



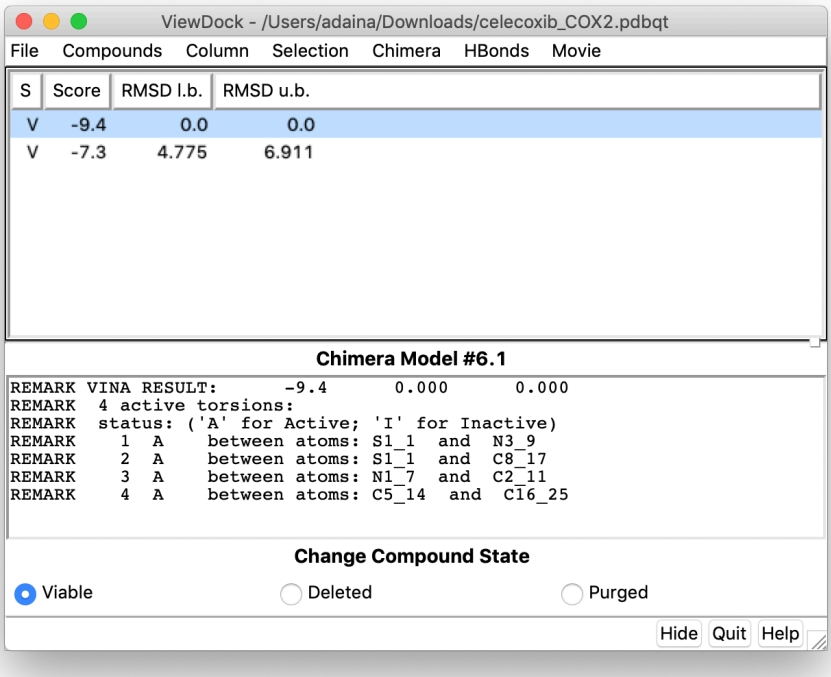
   - Click "**Start Webina**"

- After one minute or so (depending on your computer), you can **download the Output PDBQT file** on your hard-drive.



- *Analyze docking results*

If you have performed a Vina docking locally using the Chimera interface (workaround 1), the **ViewDock** window should pop up automatically upon docking completion. If you have downloaded the pre-calculated results (workaround 2) or you have used Webina (workaround 3), you can open it with **>Tools >Surface/Binding Analysis >ViewDock** and by retrieving the file **"celecoxib-in-cox2.pdbqt"** from your hard drive.



- In the **ViewDock** window, the lines starting with a "V" are individual docking poses. Go through every docking solution by **clicking the first line** and pressing the ↓ **key**. It is also possible to display multiple poses by holding the **ctrl key.**
-
    - *How many docking poses are proposed?*
    - *Are all poses located inside the binding site as defined?*

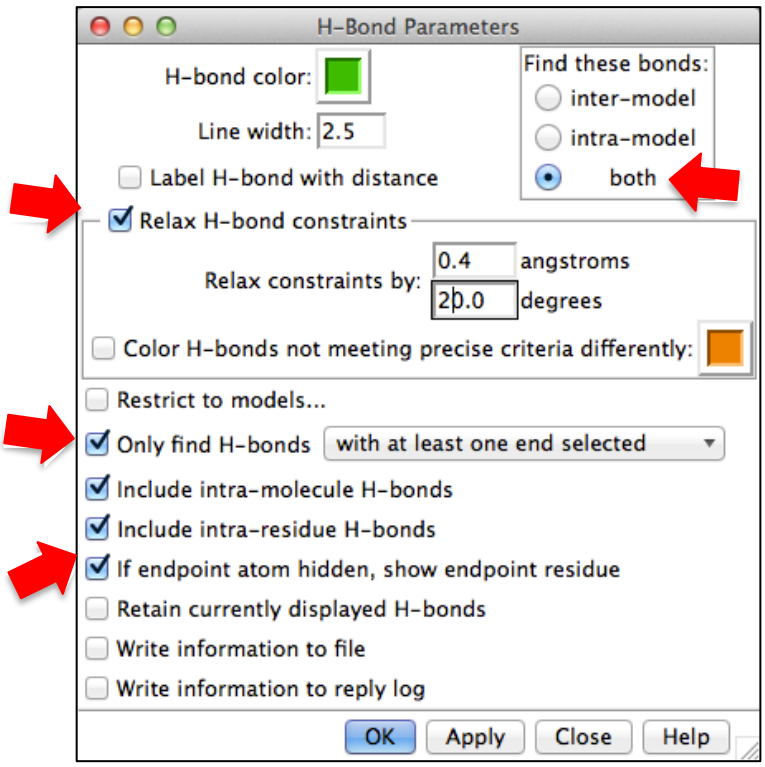
- If cocrystallized Naproxen is hidden, display it:
  In the **Model Panel (>Favories >Model Panel**):
  Tick the cocrystalized ligand Naproxen (#0.1)
  Untick the surface: MSMS main surface of COX2 (#0.2)
  Untick the AutopDock Vina Search Volume

    - *Qualitatively, is one Celecoxib docking pose better superimposed over the crystallographic pose of Naproxen?*
    - *Is it the solution evaluated as the most favorable by the scoring function? Note the score. In what range is the predicted K_i of Celecoxib inhibiting COX2?*

- Evaluate the intermolecular **hydrogen-bonds** explaining (at least partly) the recognition of the Celecoxib by COX2.

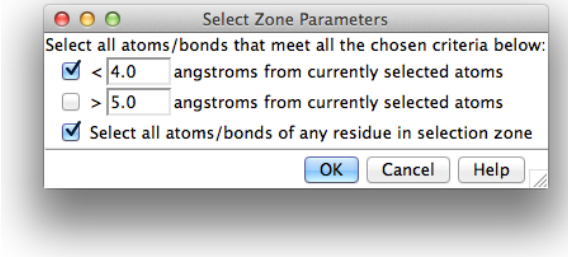
  Hide the protein surface and the Naproxen (from the **Model Panel**). Display COX2.
  Display and select what you consider the best docking pose (for instance, **select #3.1**),
  change the rendering: **Actions >Atoms/bonds > ball & stick**.
  and then **>Tools >Surface/Binding Analysis >FindHbond.**
  In the **H-Bond Parameters** window, make sure that "**both**" (intra- and intermolecular hydrogen-bonds will be searched) and that **Relax H-bond constraints** will be applied.
  Select "**Only find H-bonds with at least one end selected**", and

Include intra-molecule H-bonds and **Include intra-residue H-bonds**, and **If enpoint atom hidden, show endpoint residue.** Finally click **"Apply".**



- *How many possible intermolecular hydrogen-bonds were found (in thin lines)?*
- *Which amino acid(s) is (are) involved? What atoms are involved? Are they part of backbone or side chain?*

Display the amino acids close to the docked Celecoxib: Select the best pose (for instance **select #3.1**), then **Select > Zone…** In the Select Zone Parameters window, **Select all atoms/bonds of any residue in 4.0 Å** of the binding solution. Click **"OK"**.



Then **Actions >Atoms/Bonds >Show only.**

For clarity, hide non-polar hydrogens: **>Select >Chemistry >IDATM type >HC** and **>Actions >Atoms/Bonds >hide**.

- *Give two examples of hydrophobic interactions.*
- *Can you spot aromatic interactions?*
- *Propose one targeted mutation that could validate the predicted binding mode of Celecoxib in COX2.*

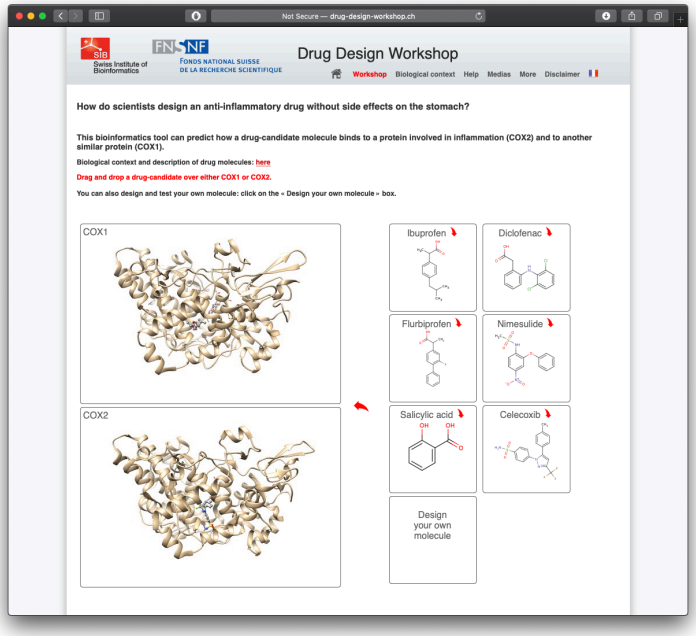
- Make the best possible JPEG image that illustrates how Celecoxib is predicted to bind to COX2. Emphasize, select and label (once selected **>Actions >Label >residue >name + specifier**) the amino acids involved in intermolecular hydrogen-bonds.
  >File >Save Image…

  📣 **Please note that in some bugs were reported for the generation of images with the Windows version of UCSF Chimera. If the picture generated is all black, please try to set "print unit" to "points" and adapt the "Image width" to 225.**
  You may also try with another output format (PNG, for instance).

  - *Perform structure-based optimization yourself! Are you able to design, build and dock a chemically modified Celecoxib? Or a totally different putative COX2 inhibitor? If not, you can use the simplified (educational) web-based procedure "Drug-Design-Workshop" described in the next section.*

Exercise 6. Structure-based optimization of COX2 inhibitors

- *Retrieve precalculated docking pose of Celecoxib*
  In a web browser, go to the website **http://www.drug-design-workshop.ch/cox.php**
  Drag-and-drop Celecoxib structure over the picture of "COX2".



Almost immediately, the **pre-calculated docking** results appear in the web page.

- *Can you recognize the binding mode you predicted in the part A of the session?*

- *Dock interactively your own molecules into COX2*

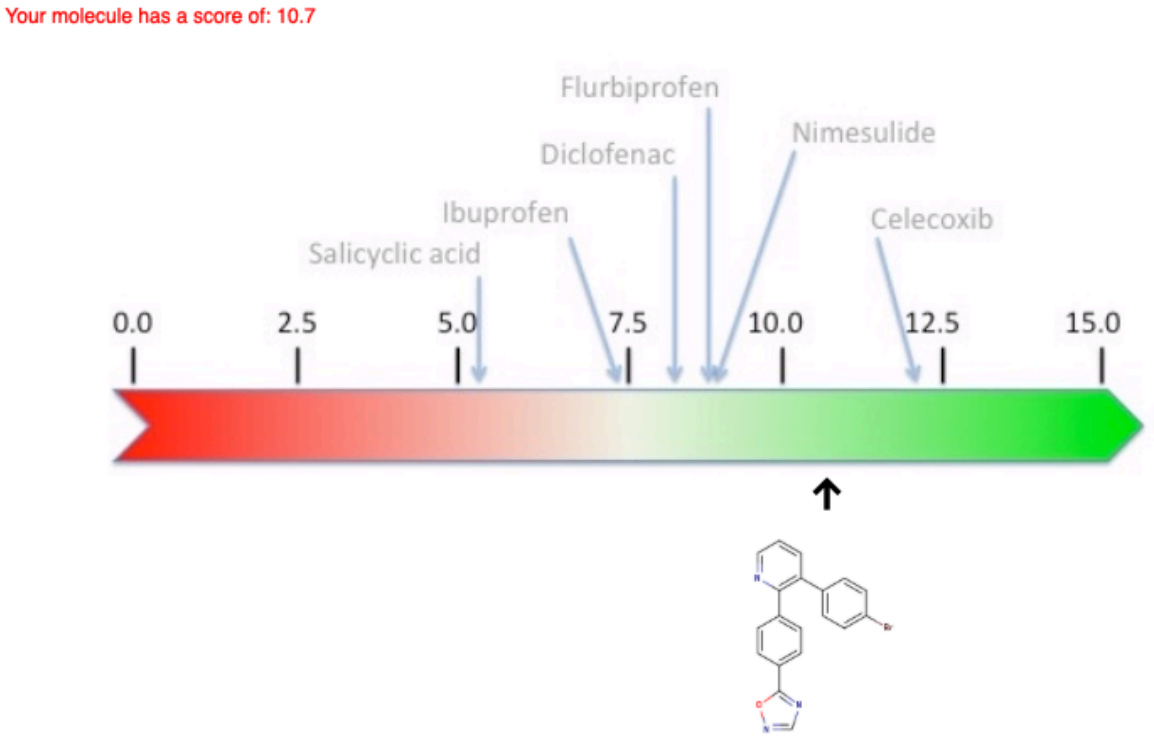
Click on the "**Design your own molecule**" box.
A **sketcher** appears at the bottom of the page.
Draw the molecule to dock and click on "**Done**". The molecule structure appears in the box.

Drag-and-drop it over the "COX2". The docking is prepared and starts automatically, when a processor on the server is free.

Upon completion (less than a minute, you can monitor your docking on a progression bar), the message "Docking completed!" appears.
Click on the link: "**The results are available here**".

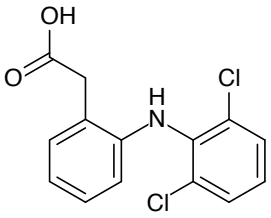
The highest ranked docking pose is displayed together with its score on a scale.

- *With this information will you be able to optimize rationally high affinity molecules iteratively?* (Return to the **first tab** of your browser, click on **"Design your own molecule" box** and optimize the structure or restart from scratch!)



Practice Session 3. Ligand-based virtual screening with SwissSimilarity

Let's start with the simple virtual screening of the cyclooxygenase (COX) inhibitor, anti-inflammatory molecule, **Diclofenac**, against all drugs approved by the FDA. This will illustrate the concept, show how methods can influence the outcome and how to calculate enrichment. Here is the 2D structure and SMILES of Diclofenac:

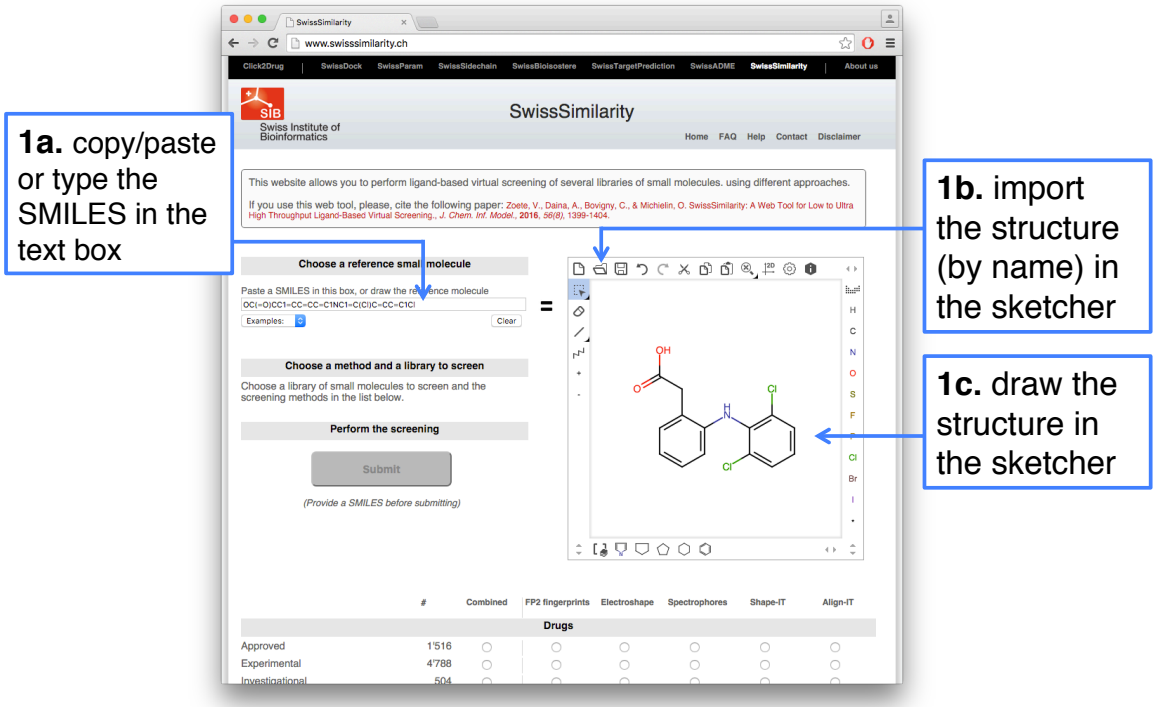


SMILES: OC(=O)CC1=CC=CC=C1NC1=C(Cl)C=CC=C1Cl

- Access the web interface of SwissSimilarity in a web browser by typing the following URL http://www.SwissSimilarity.ch (better with *Google Chrome* and *Mozilla Firefox*).

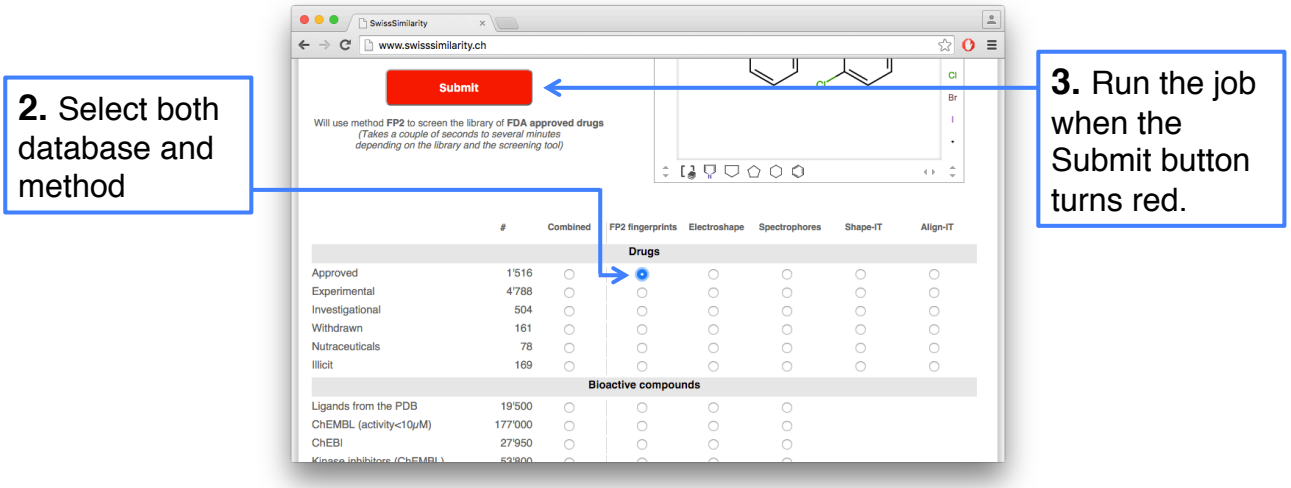
1. Use one of the following options to **input the molecule** to be screened against (i.e. Diclofenac):
   a. copy/paste or type the SMILES in the text box, or
   b. import the structure (by name) in the sketcher, or
   c. draw the structure in the sketcher.

   Please note that the SMILES in the text field and the 2D structure are synchronized.

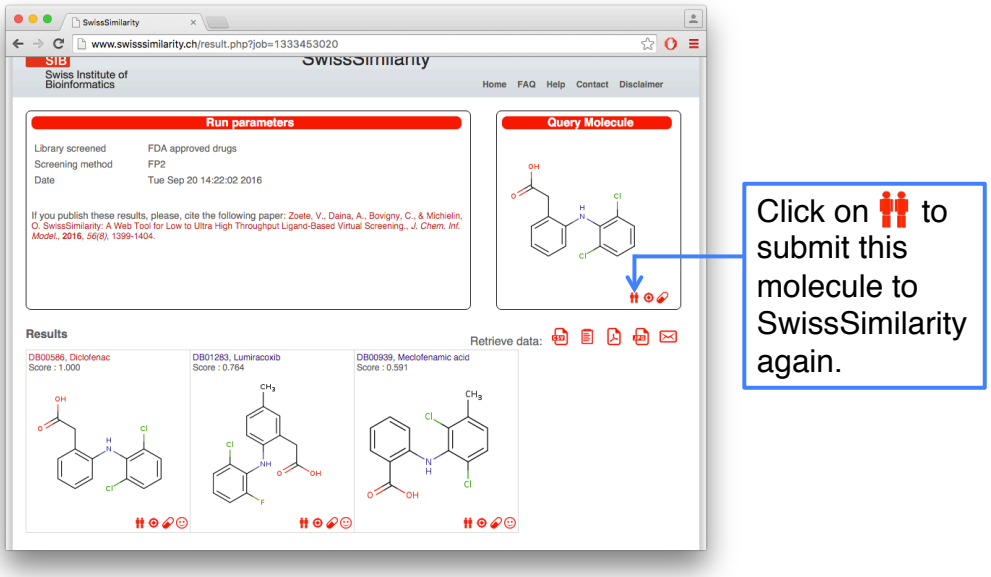


2. Click on the radio button to **select both the screening method and the library** to screen. Here it is proposed to evaluate the "**Approved**" drugs for similarity with Diclofenac using "**FP2 Fingerprints**".

3. Run the screening by clicking the **"Submit" button**, which turns red only upon the input of molecule and selection of database and method. A text describing the job to be launched appears below the button as well.



After a few seconds, the FDA approved drugs most similar to Diclofenac, from a chemical point of view, are shown.



📣 Looking at the output page, let's answer these questions:

- *How many marketed drugs have been found, excluding Diclofenac itself?*
- *Are these structures chemically similar to that of Diclofenac?*

Let's try to re-submit the FDA-approved drugs screening against Diclofenac but with another methods.

- Click on the **"Twins"** icon below the Diclofenac in the output page to open a new SwissSimilarity submission with the molecule already input.
- Click on radio button corresponding to "**Approved**" drugs and "**Electroshape**".

After some time, the drug molecules that show similarity in 3D with Diclofenac are displayed. This run is a bit longer than FP2. Please note that the duration depends on the choice of the database and method. A time estimate pops-up by leaving the pointer over the corresponding radio button for a second.

📣 Looking at the output page, let's try to answer these questions:

- *How many marketed drugs have been found, excluding Diclofenac itself?*
- *Are these structures chemically similar to that of Diclofenac?*

Exercise 7. Enrichment of anti-inflammatory molecules in FDA-approved drugs.

There are **1516 FDA approved drugs**, among which **39 COX inhibitors**:

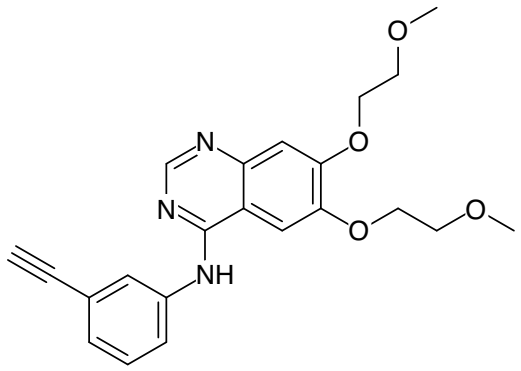
Acetaminophen, Acetylsalicylic acid, Antipyrine, Antrafenine, Balsalazide, Bromfenac, Carprofen, Diclofenac, Diethylcarbamazine, Diflunisal, Dihomo-gamma-linolenic acid, Etodolac, Fenoprofen, Flurbiprofen, Ibuprofen, Icosapent, Indomethacin, Ketoprofen, Ketorolac, Lornoxicam, Lumiracoxib, Meclofenamic acid, Mefenamic acid, Meloxicam, Mesalazine, Nabumetone, Nepafenac, Oxaprozin, Phenylbutazone, Piroxicam, Resveratrol, Salsalate, Sulfasalazine, Sulindac, Suprofen, Tenoxicam, Tiaprofenic acid, Tolmetin, Trisalicylate-choline

- *Calculate the enrichment at 1.3% (= top 20)*
  - *How many of these COX inhibitors can you find in the top 20 molecules provided by Electroshape screening* ($\#COX_{screening}$)*?*
  - *How many COX inhibitors are in total in the FDA library* ($\#COX_{library}$)*?*

$$\boldsymbol{Enrichment\ (at\ 1.3\%)} = \frac{\frac{\#COX_{screening}}{\mathbf{20}}}{\frac{\#COX_{library}}{\mathbf{1516}}}$$

This first exercise is finished. You can close all windows and tabs.

Erlotinib (Tarceva®) is an inhibitor of the epidermal growth factor receptor (EGFR, erbB1, uniprotID P00533) tyrosine kinase. It has been approved by the FDA as an anti-cancer drug used to treat non-small cell lung cancer, pancreatic cancer and several other tumor types. The 2D structure and SMILES of Erlotinib are displayed below:



SMILES: COCCOC1=CC2=C(C=C1OCCOC)C(NC1=CC=CC(=C1)C#C)=NC=N2

The objective of this exercise is to use the tool **SwissSimilarity** to screen the entire collection of active molecules from the ChEMBL database in order to retrieve compounds similar to Erlotinib in terms of chemical structures (2D) and in terms of shape (3D).

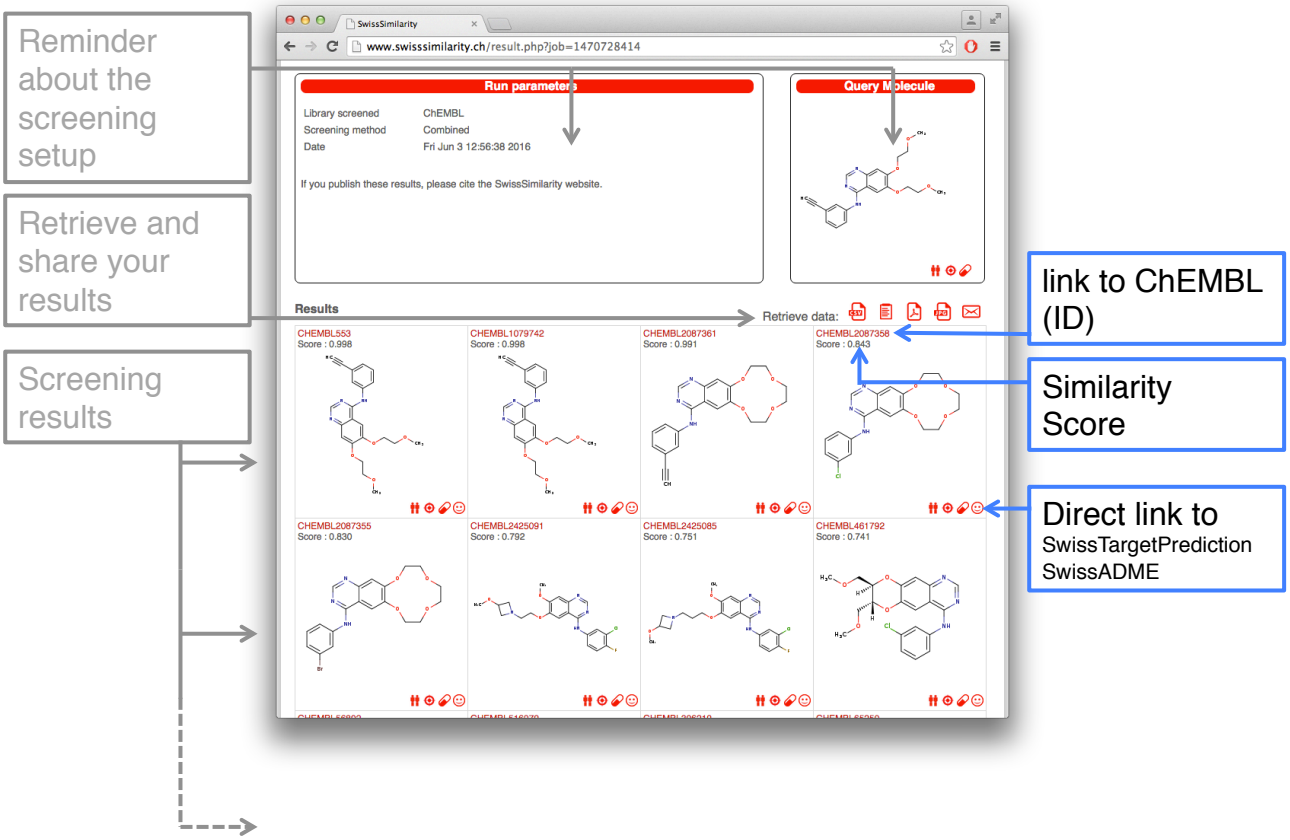
- Access the web interface of SwissSimilarity in a web browser by typing the following URL http://www.SwissSimilarity.ch.

- Use one of the following options to **input the molecule** to be screened against (i.e. Erlotinib):
    a. copy/paste or type the SMILES in the text box, or
    b. import the structure (by name) in the sketcher, or
    c. draw the structure in the sketcher.

- Click on a radio button in the list to **select both the screening method and the library** to screen. Here it is proposed to evaluate the entire "**ChEMBL database (activity < 10 μM)**" for similarity with Erlotinib using a "**Combined**" score, which makes the process screen the library with both 2D-fingerprint and 3D-electroshape to return a consensus value, corresponding to a combined score obtain by logistic regression of both individual (2D and 3D) similarity measurements.

- Start the screening by clicking the **"Submit" button**, which turns red only upon the input of molecule and selection of database/method. A text describing the job to be launched appears below the button as well.

- After less than one minute, your **screening results** should appear on your web browser. In the meantime, you can follow the advancement of the job thanks to the grey bar.



📣 With that results let's try to answer the following questions about compound CHEMBL461792:

- *What is the similarity score and ranking of compound CHEMBL461792?*
- *What are the two structural differences between CHEMBL461792 and Erlotinib?*
- *Which of these chemical modifications makes CHEMBL461792 more rigid than Erlotinib?*
- *Any clue about the potential benefit to test a more rigid ligand?*

Please, KEEP this SwissSimilarity result page OPEN as it is the starting point for the following sessions.

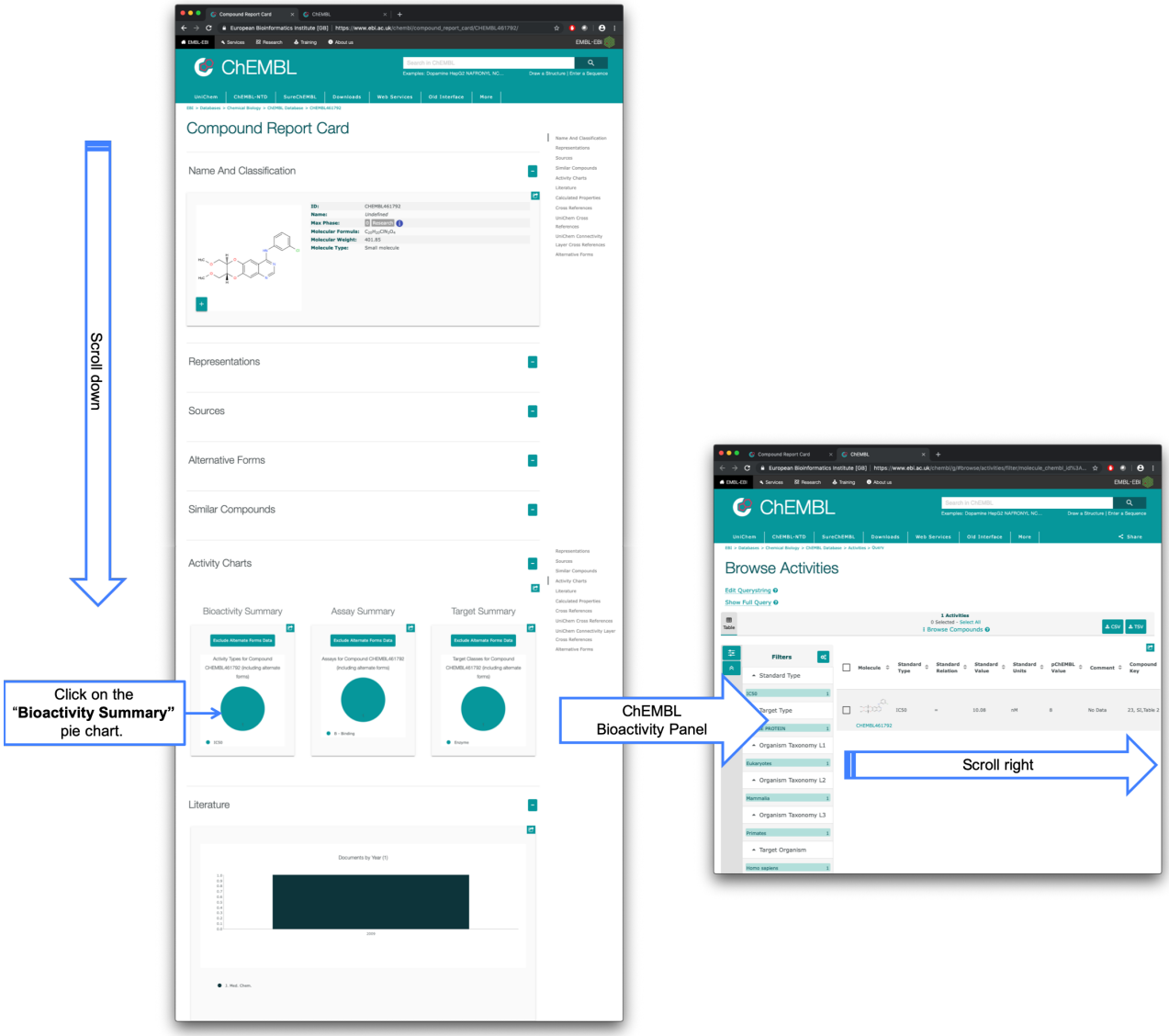
You can copy, bookmark or email the URL of your SwissSimilarity result page.

Practice Session 4. Reverse screening with SwissTargetPrediction

The purpose of this session is to analyze the biological activity of compound CHEMBL461792, which shows significant similarity with Erlotinib. The first actions illustrate the interoperability of the different SwissDrugDesign tools and the link with external resources.

1. Let's go back to the **SwissSimilarity result page** in your web browser and access to ChEMBL database for entry CHEMBL461792 (by **clicking the ID link**).

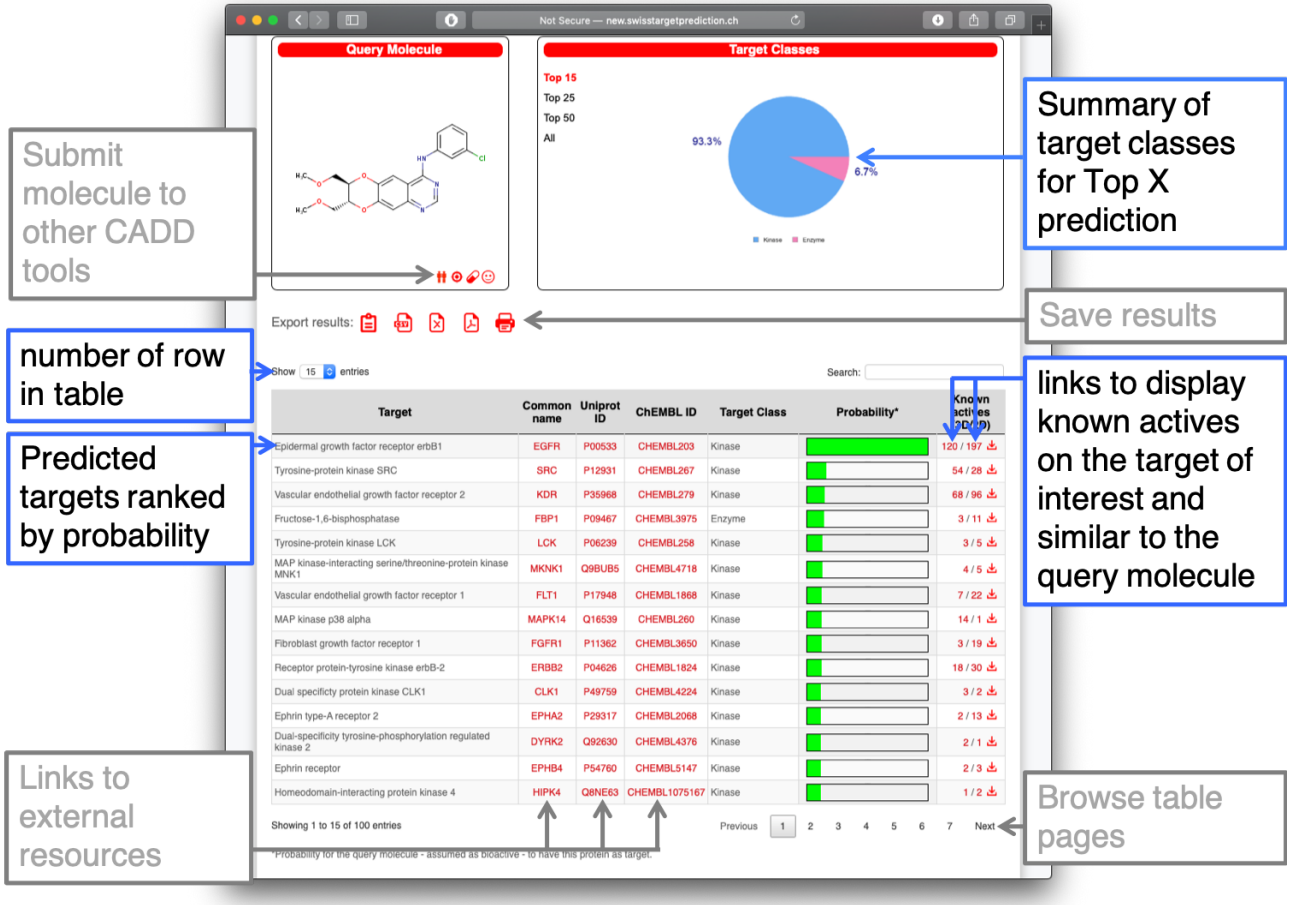
2. A first **ChEMBL** panel corresponding the entry CHEMBL461792 opens, scroll down to the "Activity Charts" section and click on the "**Bioactivity Summary"** pie chart. This brings you to a second "ChEMBL Bioactivity" panel.



Confirm that compound CHEMBL461792 has been tested on EGFR and note the IC_{50}.

3. Go back to the SwissSimilarity result page in your web browser. Submit compound CHEMBL461792 to **SwissTargetPrediction for reverse screening to predict biotargets**, either by
    a. clicking the corresponding "**target" icon** from the SwissSimilarity results page, or
    b. by going directly to http://www.swisstargetprediction.ch in a new tab or window. In this case, you will need to draw the chemical structure in the sketcher or copy/paste the SMILES in the dedicated text box, and finally click on the "Submit" button.

- After less than one minute, the **target prediction results** should appear on your web browser.



📣 With that results let's try to answer the following questions about compound CHEMBL461792:

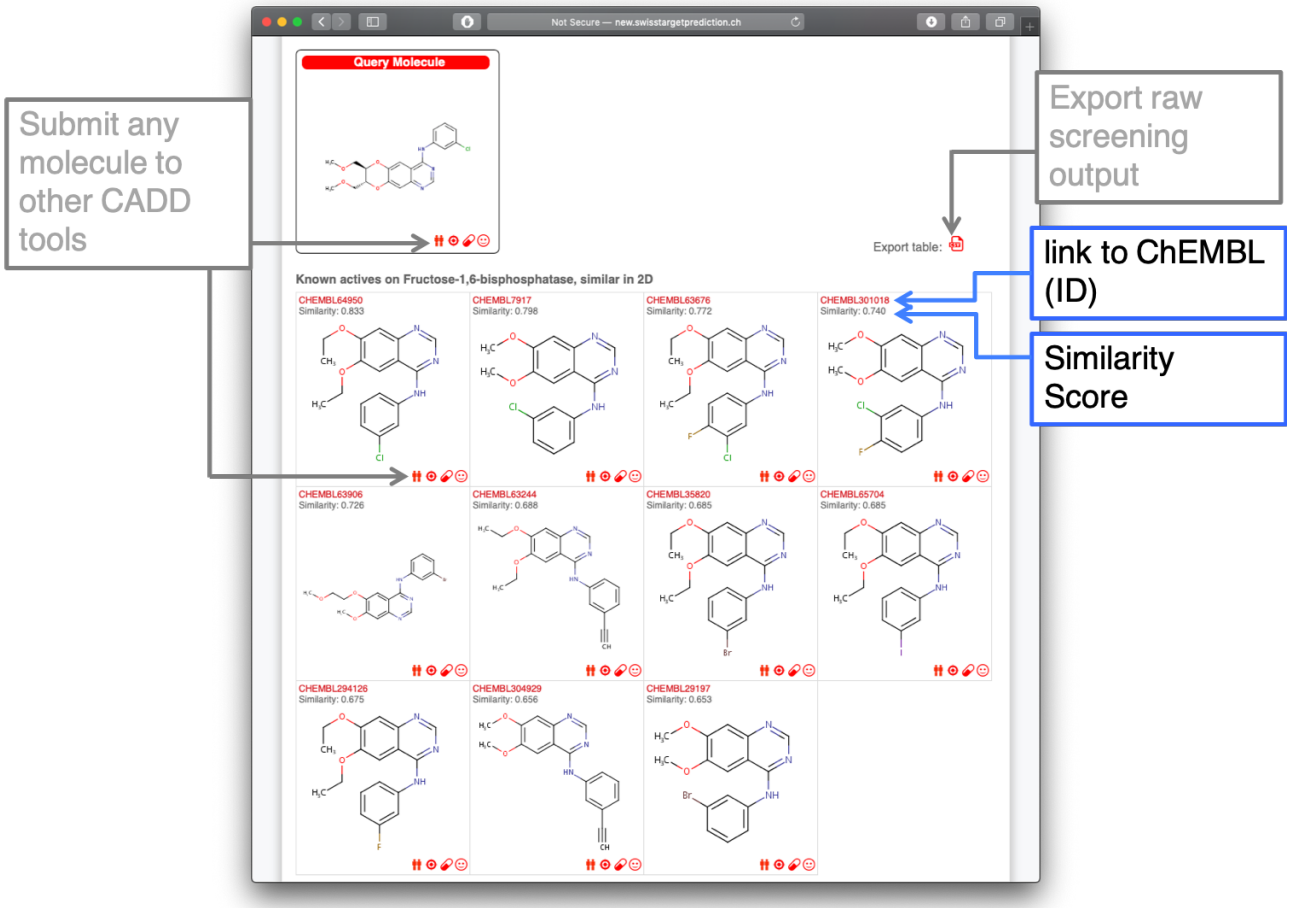
- *How many protein targets are predicted in total (select* Show "**All**" entries *on the menu above the table). What is the proportion of kinases (click on "**All**" left to the pie-chart)?*

- *What is the most probable protein target for this compound? Can we consider this result as an actual prediction? Why?*

- *Looking at the ranking, what is the most probable non-kinase target?*

- *How many known actives of this non-kinase protein are similar to molecule CHEMBL461792 based on 2D chemical structure similarity? Same question for 3D shape similarity?*

4. By **clicking the number of similar molecules** for a given target (either from 2D or 3D screenings), a second window opens with a complete description of the outcome of the reverse screening. The structure of similar molecules active on the target of interest (those having driven the prediction) is displayed.



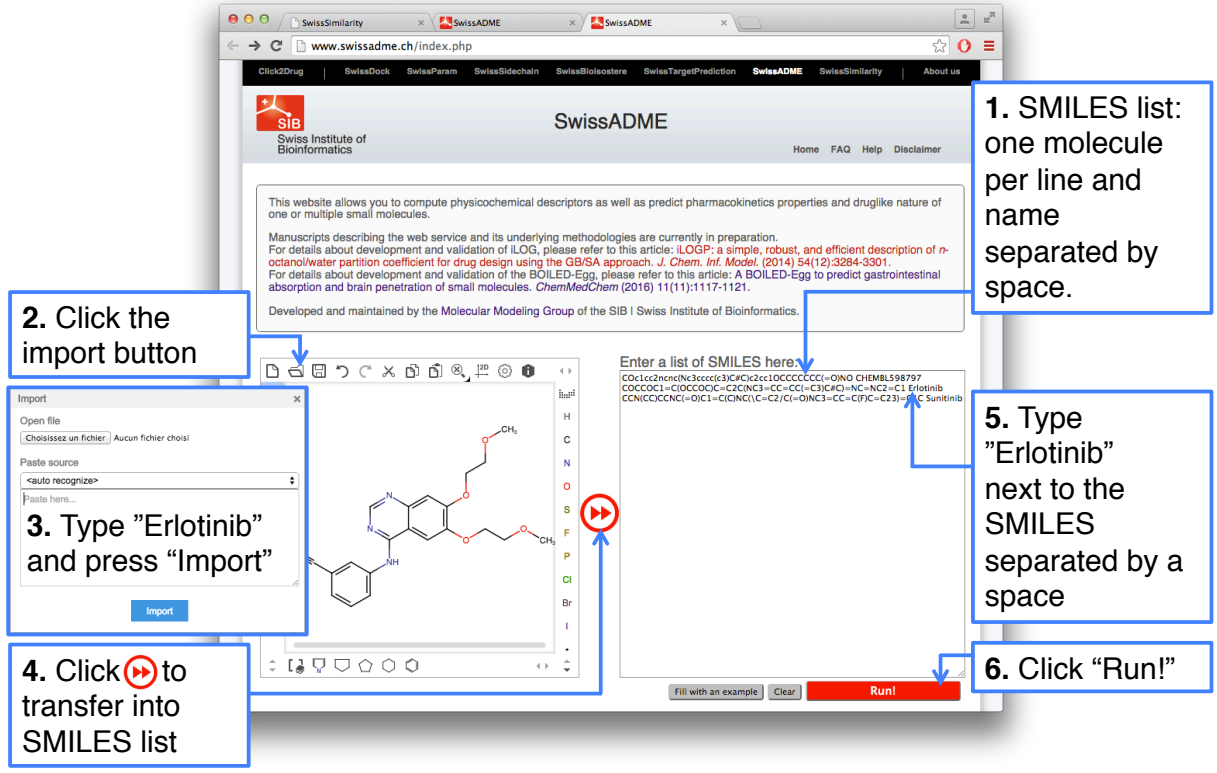
- *Are the actives most similar in 2D and in 3D to CHEMBL461792 the same compound?*

Practice Session 5. Pharmacokinetics evaluation with SwissADME

During this session, it is proposed to estimate the druglikeness and two important ADME behaviors regarding absorption and side effects: the gastrointestinal passive absorption and the brain penetration, respectively. To this end, the SwissADME web tool will be used to analyze Erlotinib and an analogue molecule (CHEMBL598797) together with another kinase inhibitor: the anti-cancer drug Sunitinib.

- Let's go back to the **SwissSimilarity result page** in your web browser. Locate compound CHEMBL598797 (score: 0.520) and click on the corresponding "**pill" icon** to launch directly a SwissADME calculation.

- A new tab opens with SwissADME calculating for about 3 to 10 seconds. Upon completion the output panel related to CHEMBL598797 is displayed in the same page. Now let's add two anti-cancer drugs Erlotinib and Sunitinib to the list and let's rerun the calculation.



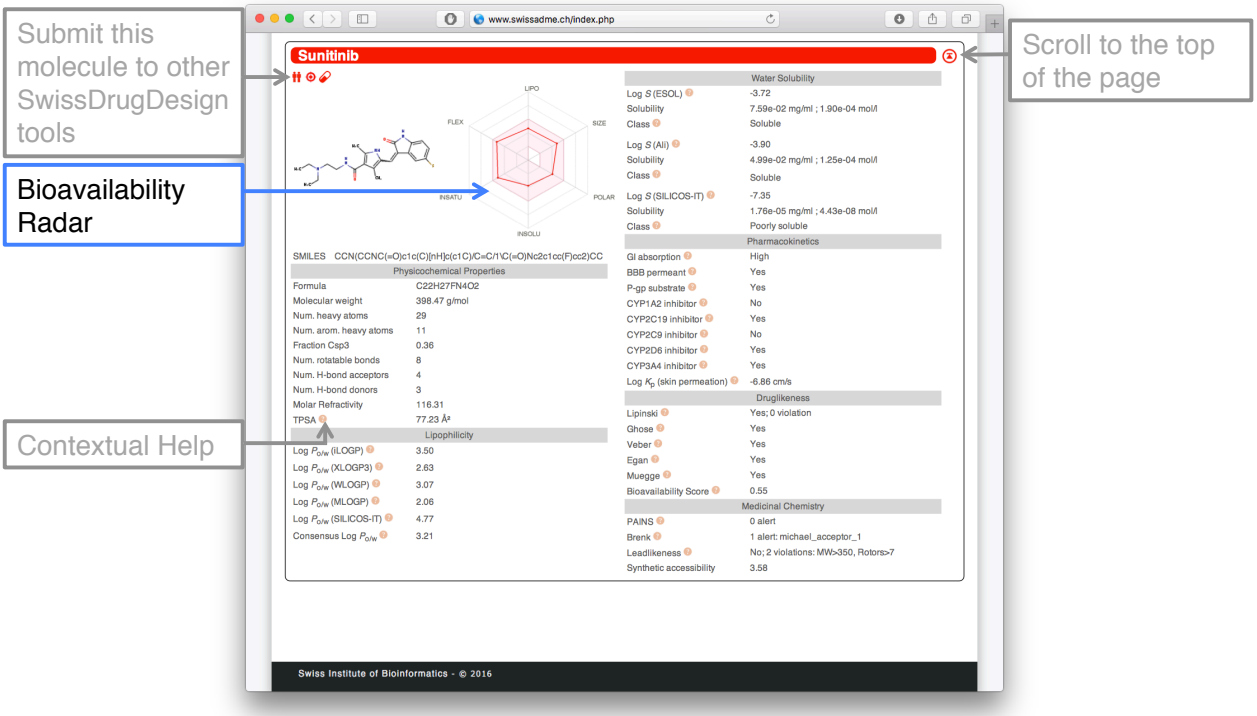
1. Scroll to the top of the SwissADME page and in the SMILES list box, type the name of the compound "CHEMBL598797" next to the SMILES, separated by a **space**. Then press **enter** to go to the next line.
2. Click on the "**Import**" button of the sketcher (second button top-left).
3. Type "Erlotinib" in the pop-up "Paste source" box and click the "**Import**" blue button.
4. The structure of Erlotinib appears in the sketcher. Click on the **double-arrow button** in between the sketcher and the SMILES list to transfer the molecule into the list.

5. Type “Erlotinib” next to the SMILES on the second line, separated by a **space**.

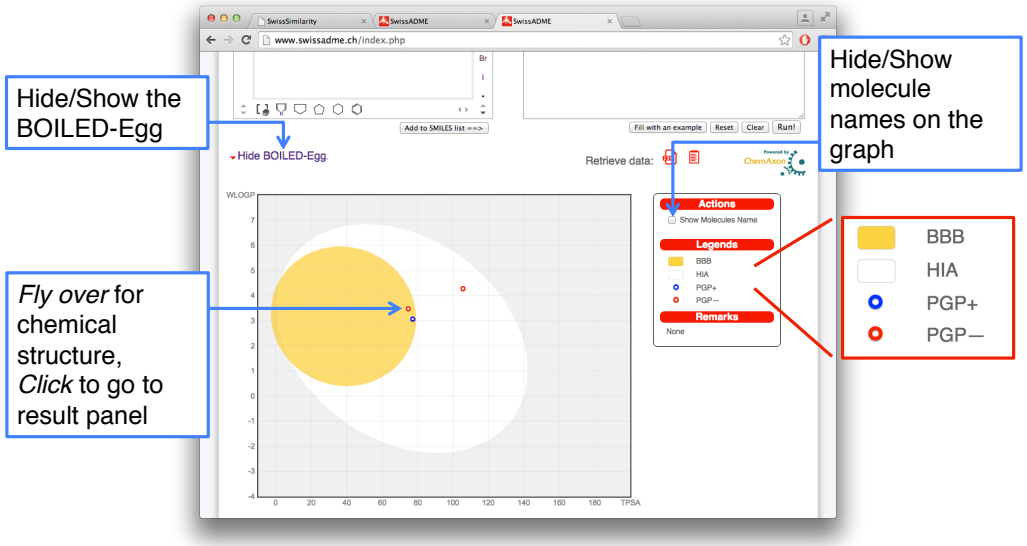
- Repeat steps 2, 3, 4 and 5 for Sunitinib.

6. Click on the “**Run!**” button at the bottom of the list.

- Upon calculation completion (3 to 10 seconds per molecules), the results appear in the same page; three panels per molecule, each summarizing the values for a given compound.



- By clicking on the “**Show BOILED-Egg**” red button (below the sketcher), the graphical output is displayed on the same page



📣 With those results let’s try to answer the following questions about the ADMET of those three molecules:

- *One of these compounds is predicted toxic, can you point out which one and the alert related to this prediction?*
- *Between both marketed drugs, which one of Sunitinib or Erlotinib is more prone to create drug-drug interactions linked with metabolism?*
- *Which of these molecules is the less druglike? What is the molecular property responsible for that?*
- *Are all three compounds predicted as well-absorbed by the gastrointestinal tract when administered orally?*
- *Qualitatively, what is the propensity for each compound to passively cross the blood-brain barrier?*
- *Which is the physicochemical property mostly explaining the difference in passive brain permeation behavior?*
- *Which compound(s) is (are) predicted actively pumped out from the central nervous system? Why?*
- *Finally, which compound has the highest probability to be in significant concentration in the brain?*

Exercise 8. Pharmacokinetics optimization of EGFR inhibitor.

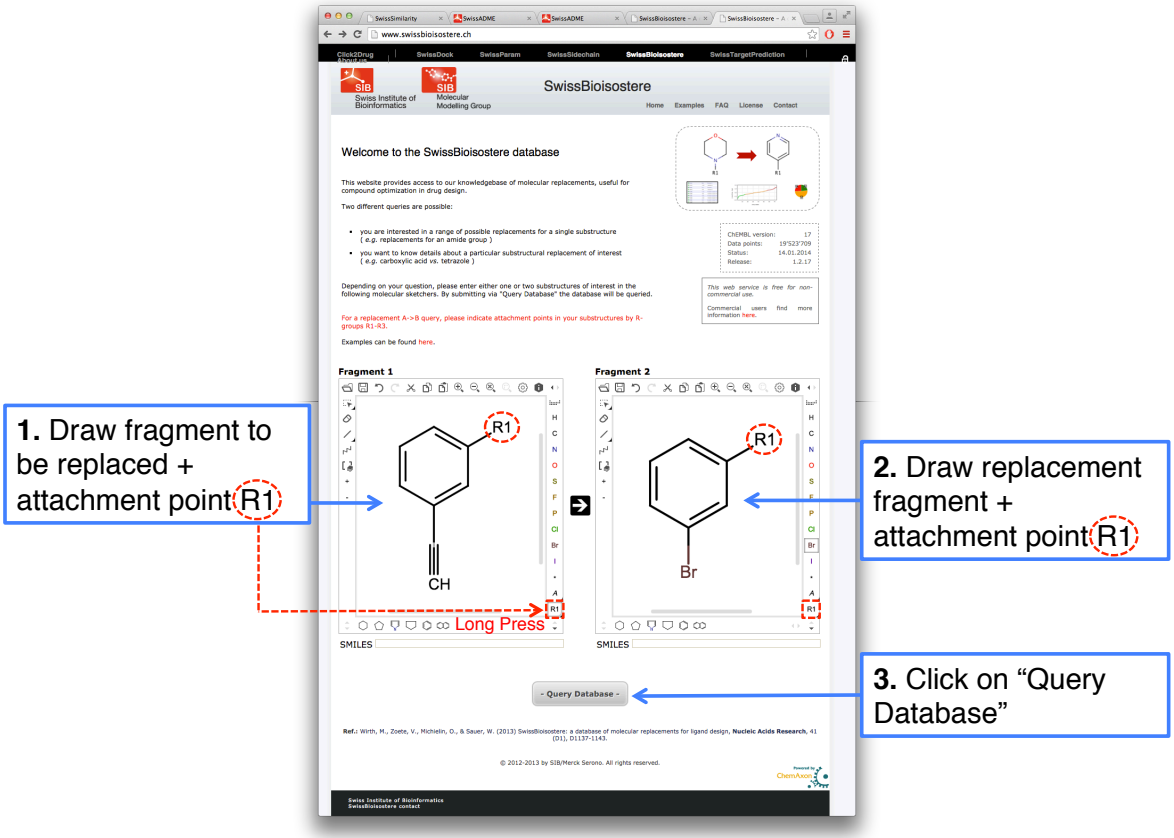
- Imagine that your endeavor consists in optimizing the properties of CHEMBL598797, which has to inhibit a kinase in the CNS. Have some tries of small chemical modifications (e.g. copy/paste SMILES in the sketcher, apply chemical modifications and transfer multiple entry lines to the SMILES list). You have so initiated an iterative optimization process. Once you are happy with the ADMET properties, click on the target icon to submit your optimized molecule to SwissTargetPrediction.
  - *Describe your optimization strategy.*
  - *What are your conclusions regarding pharmacokinetics and pharmacodynamics?*

Practice Session 6. Bioisosteric design with SwissBioisostere

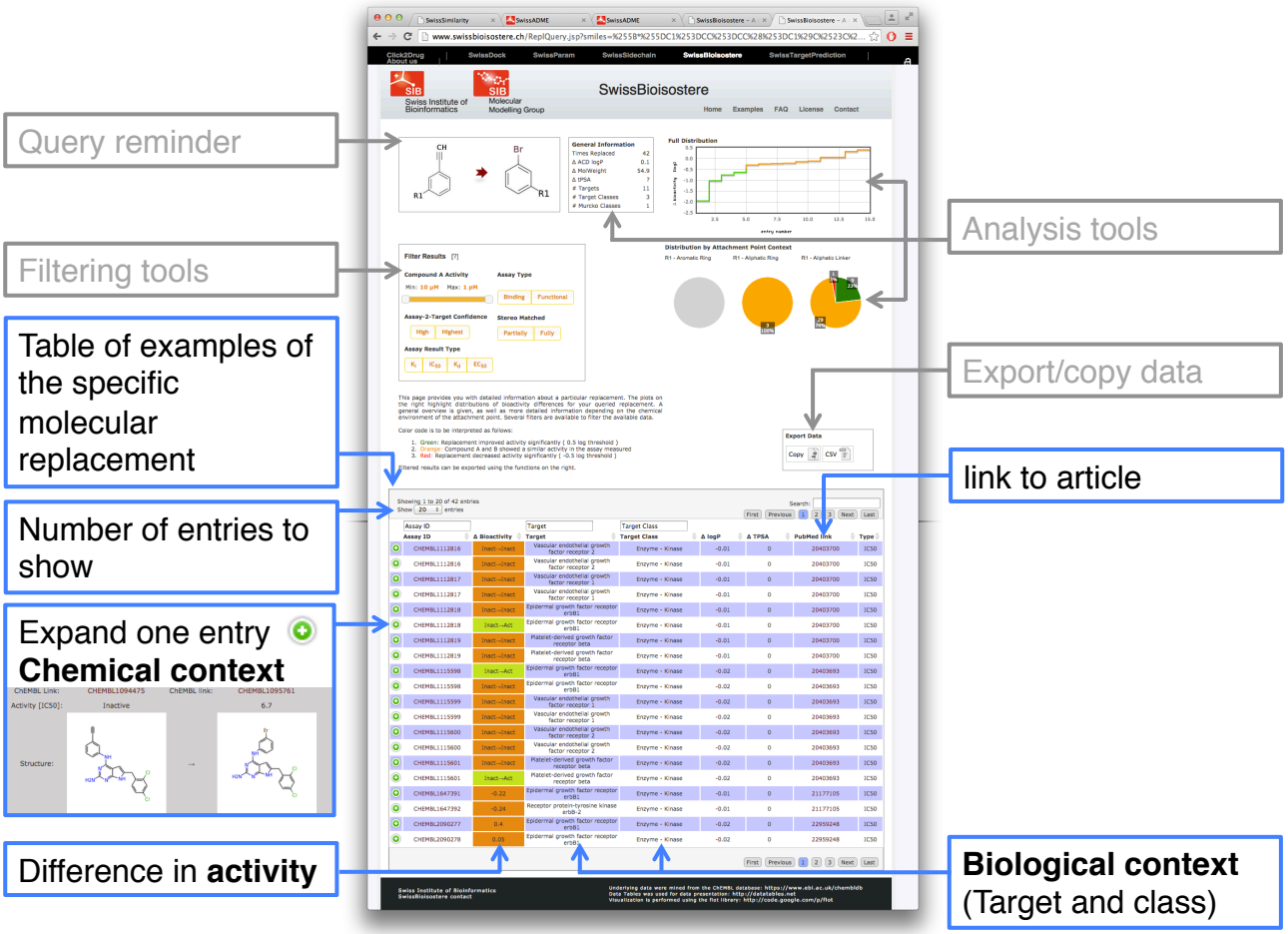
The purpose of this session is twofold. First it is proposed to analyze a specific replacement found in the virtual screening. Then we will suggest other possible bioisosteric replacements for ligand design.

- Go back to the **SwissSimilarity result page** in your web browser, locate compound CHEMBL2087361 (rank 3) and compound CHEMBL2087355 (rank 5). *Can you point out what is the chemical difference between those two molecules?*
- Let's investigate this specific molecular replacement. Type the following URL in a new tab or window of your web browser **http://www.SwissBioisostere.ch** to access the submission page of SwissBioisostere. Alternatively, you can click on the link in the black toolbar at the top of any SwissDrugDesign Website.

1. Draw in the left sketcher the fragment in CHEMBL2087361 that is replaced. To add the attachment point **long press** the R_ button (bottom-right) select **R1** and click at the correct place in the structure.
2. Draw the replacing fragment in CHEMBL2087355 in the sketcher on the right. Pay attention to add the attachment point with **R1** the same as in point 1.
3. Start the search by clicking the button "**Query Database**" at the bottom of the page.



- After a few seconds is displayed the SwissBiosisostere output page compiling all examples found in the literature about the replacement of *m*-ethynylbenzene fragment by *m*-bromobenzene.

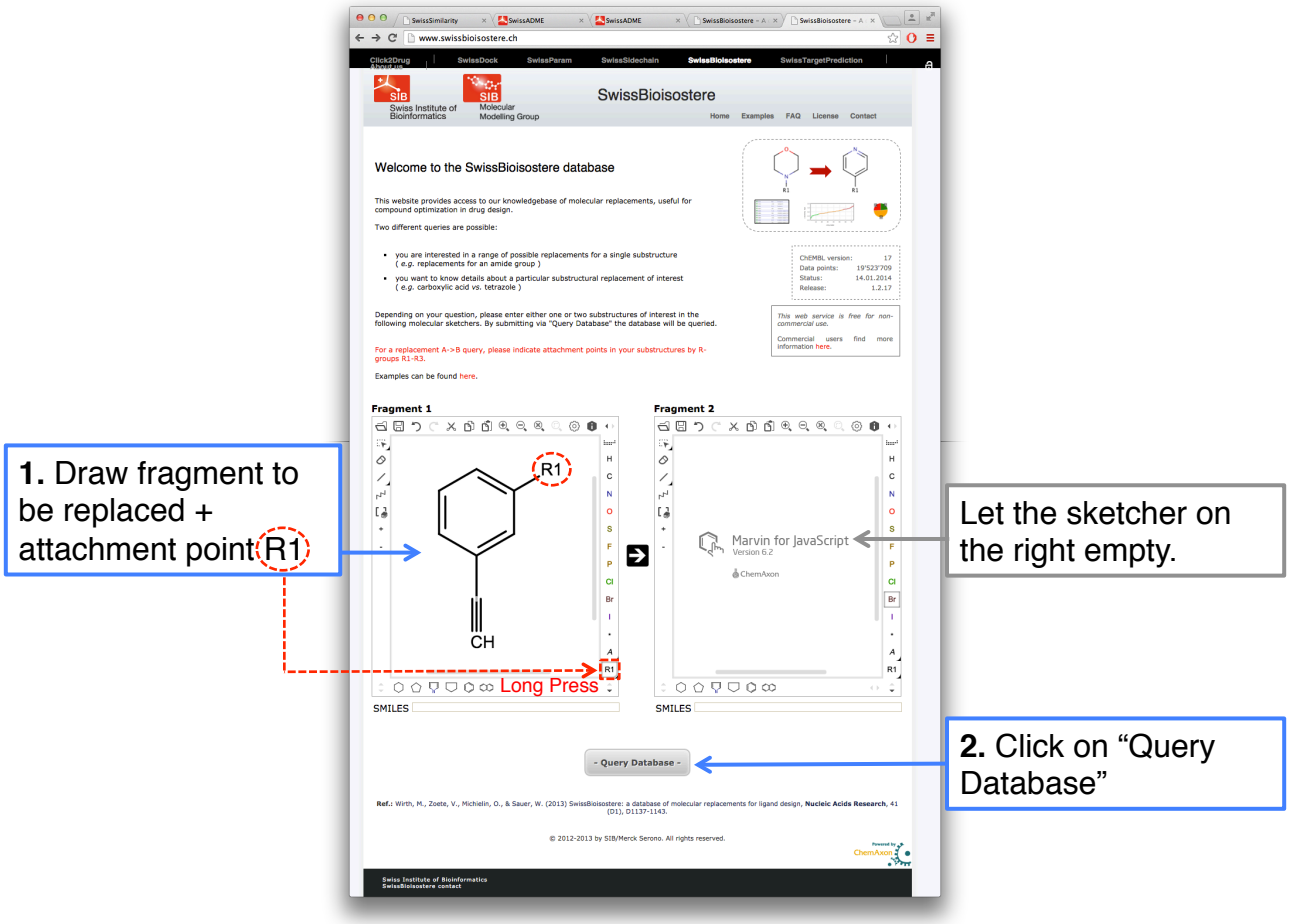


📣 Display all lines by setting "**50**" in the number of **entries to show** on the upper left corner of the result table. Let's try to answer the following questions:

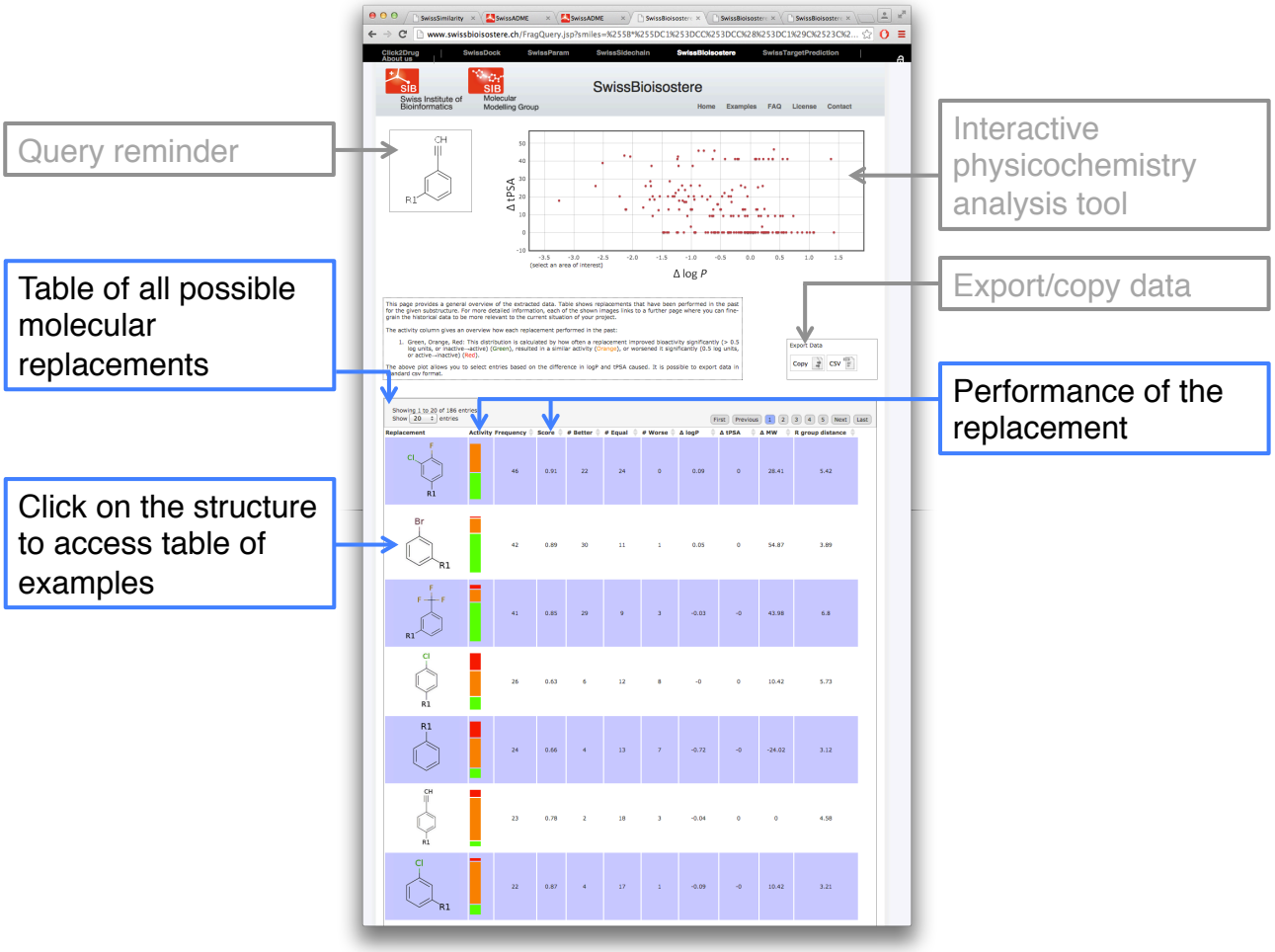
- *Overall, how many times this specific replacement was found in the literature?*
- *Generally speaking, what is the trend: increasing or decreasing the biological activity?*
- *In what biological context this replacement was mainly tried?*
- *How many times this replacement was found for compounds tested on our target of interest (erbB1)?*
- *What is the trend for activity on this specific protein?*
- *Can you find the entry corresponding to our case (CHEMBL2087361 to CHEMBL2087355)? Expand the corresponding entries to see the molecules.*
- *Which molecules is the most potent on erbB1 between CHEMBL2087361 and CHEMBL2087355? Note that the activity is given as pIC50.*
- *Any idea why is the entry seems duplicated? Click on the PubMed link to get more info from the abstract.*

- Let's try to find **other** relevant **replacements** for *m*-ethynylbenzene. Click on the "**home**" link to go back to the SwissBioisostere submission page.

1. Draw the fragment to be replaced in the sketcher on the left (*m*-ethynylbenzene). Leave the sketcher on the right **empty**.
2. Click on "**Query Database**" for SwissBiosiostere to search for all possible molecular replacements.



- After a few seconds the SwissBiosisostere output page listing all possible replacements of *m*-ethynylbenzene found in the literature is displayed.

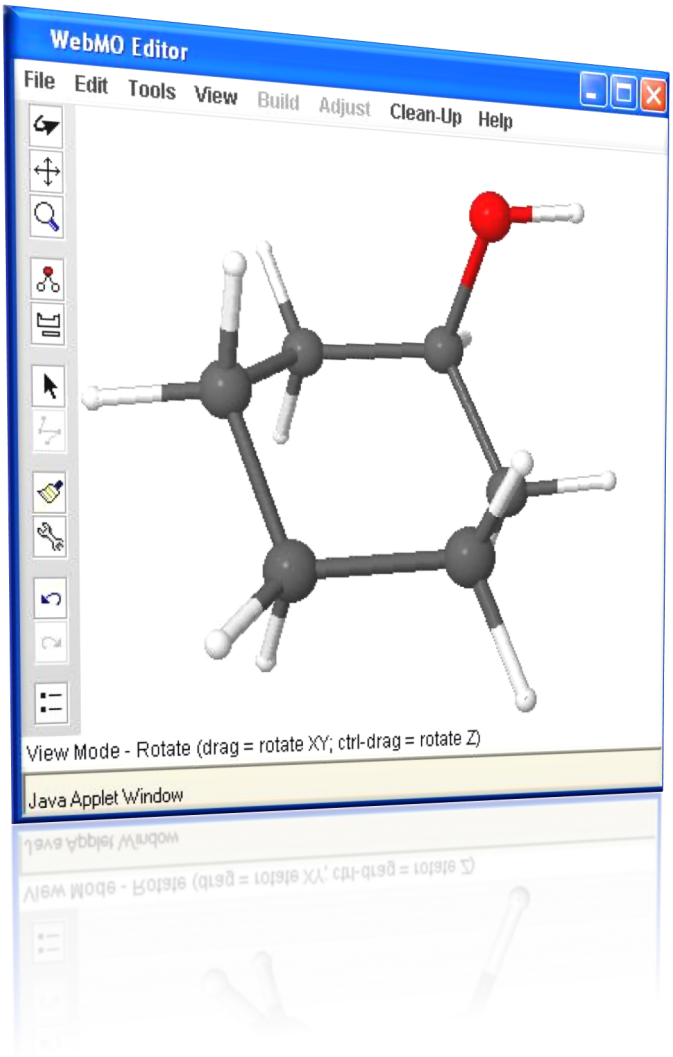


📣 With that results let's try to answer the following questions:

- *Can you find the replacement that we studied in details in the previous section (CHEMBL2087361 to CHEMBL2087355)?*

- *Propose two other sensible replacements for m-*ethynylbenzene *in our biological/chemical contexts and explain how you went to that.*

Practical Session 7: Quantum-Chemical Calculations Using WebMO



UNIL | Université de Lausanne

Swiss Institute of Bioinformatics

Ute Röhrig, Antoine Daina, Vincent Zoete – 2021

Contact: ute.roehrig@sib.swiss

In this exercise, we will use a WebMO, a free World Wide Web-based interface to computational chemistry packages (Schmidt, J.R.; Polik, W.F. WebMO, version 20.0.012; WebMO LLC: Holland, MI, USA, 2020; https://www.webmo.net, accessed May, 2021).

Some key features of WebMO:

- Support for all modern web browsers, iOS, and Android (via WebMO App)
- Support for Gamess, Gaussian, MolPro, Mopac 7 & 20XX, NWChem, Orca, PQS 3.3, PSI 4, QChem, Tinker, PWSCF (Quantum Expresso), and VASP
- Users draw structures in a 3D editor, run calculations, and view results, all from their web browser
- No software to install on the client. WebMO is installed on the single web server that runs or accesses the computational chemistry programs
- All administrative tasks (user accounts, program configuration, job time limits, etc) are performed through a web browser interface
- Simple enough for undergraduate computational chemistry curriculum: reasonable default values and results are presented on web pages
- Flexible enough for computational chemistry research: full access to input and output files
- Free academic license

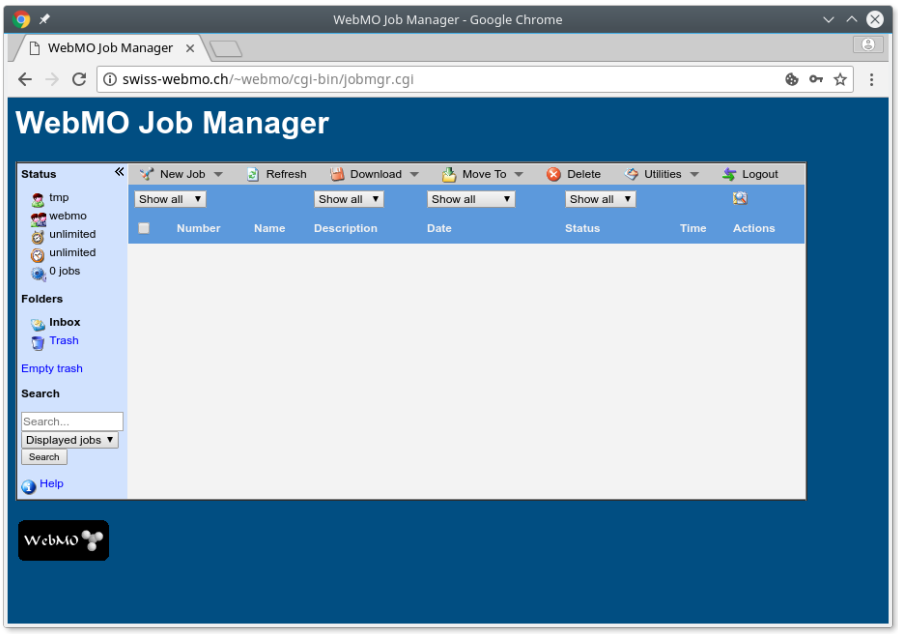
You can access our installation of WebMO under the following link: http://swiss-webmo.ch. Your login is *givenname.familyname* (as in your UNIL e-mail address, all in lower case letters). The password is *cadd-qm@fbm-unil*.

Please also download the necessary files from https://drive.switch.ch/index.php/s/PUcbHVa0gL6RfBu and save them in a dedicated folder.

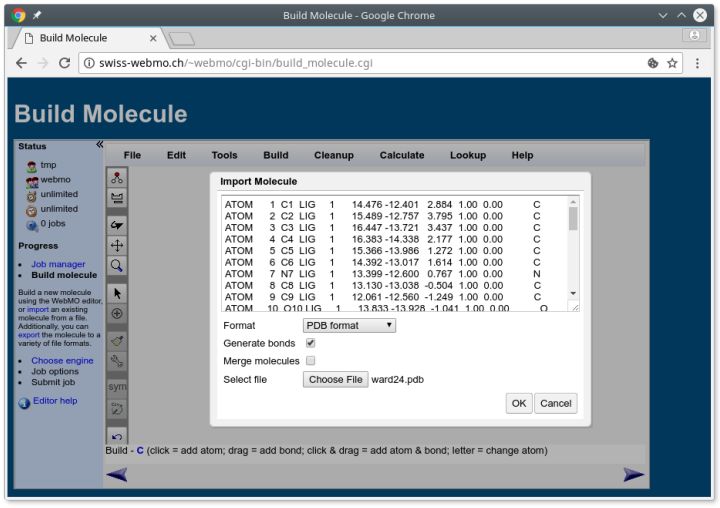
As an example, we will calculate some properties of model compounds of covalent EGFR inhibitors. The compounds are taken from R. A. Ward et al., *Structure- and Reactivity-Based Development of Covalent Inhibitors of the Activating and Gatekeeper Mutant Forms of the Epidermal Growth Factor Receptor (EGFR),* J. Med. Chem. **56**, 7025 (2013).

Task 1: Optimizing the geometry of a small organic molecule

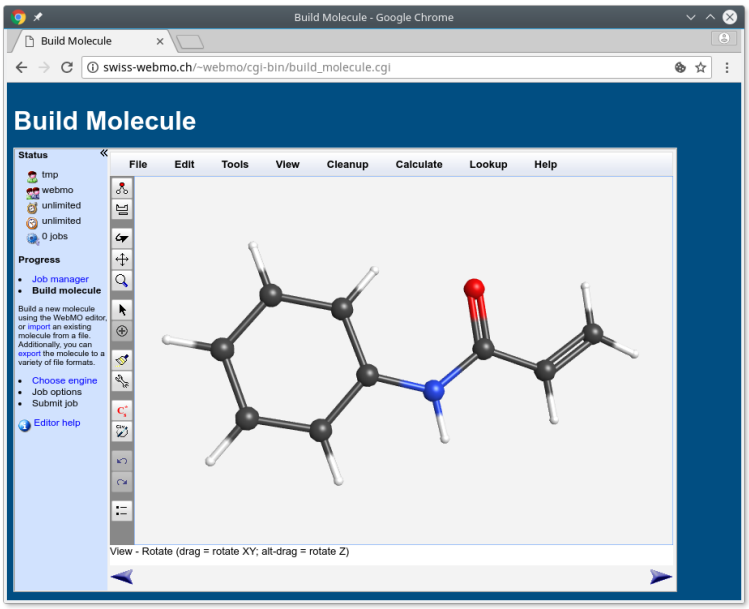
Download the files for this exercise from the link given on page 2 and store them in a dedicated directory on your computer. Then log into the http://swiss-webmo.ch website with your personal login and password (you must allow the site to store cookies to do so). This should take you to the WebMO Job Manager.



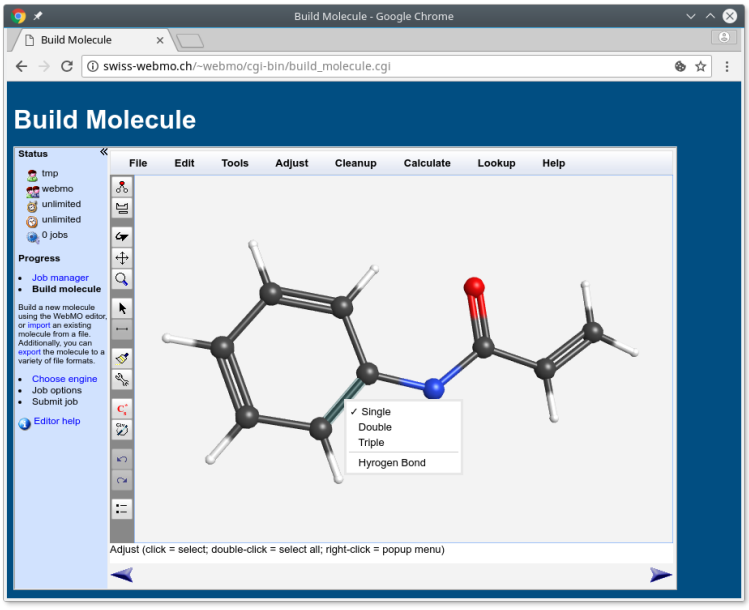
On the upper toolbar, click on "**New Job>Create New Job**". This will take you to the molecular builder, which allows you to create and to modify molecular structures. As its use is not very intuitive, we will import the 3D structure of the first molecule by choosing "**File>Import Molecule…**" from the upper toolbar. Choose **Format: PDB format** and check the "**Generate bonds**" box, then upload the file *ward24.pdb*.



You should now see the 3D structure of a small aromatic Michael acceptor (acrylamide) in the main window. By hoovering over the symbols on the left toolbar, you will see a short description of what mode the respective button will activate. First, try the "**rotate**" (3rd from top), **translate** (4th), and **zoom** (5th) buttons to move and inspect the molecular structure.



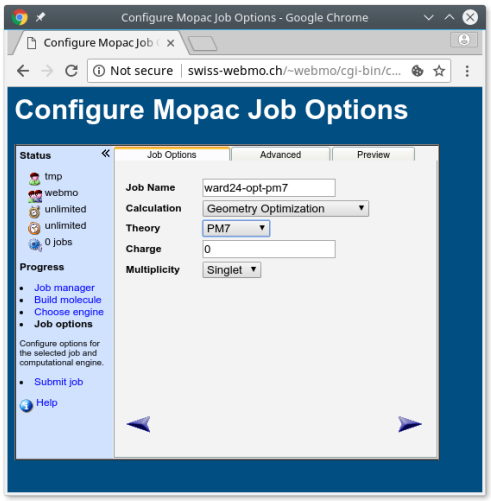
You will notice that all aromatic bonds are displayed as single bonds. To correct this, chose the 6th button ("**Adjust**") from the left toolbar, which displays an arrow, left-click on one of the aromatic bonds to select it, then right-click on it to open a popup menu, which lets you define the type of bond. Chose "**Double**", then do the same for two more alternating aromatic bonds.



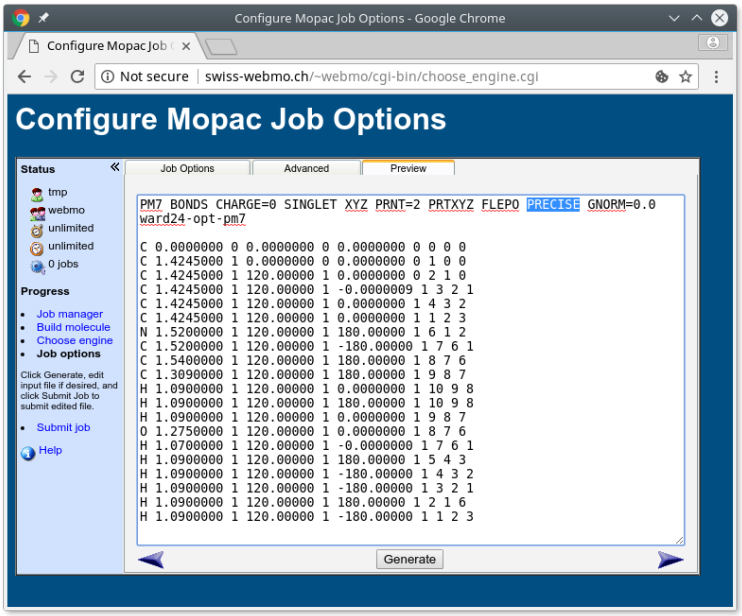
Now you can try one of the "Cleanup" procedures, either from the left (paintbrush or wrench) or the upper (Cleanup) toolbar, using either an idealized or a molecular mechanics procedure.

When you are happy with the initial structure, you can proceed by pushing the **purple arrow** in the lower right corner of the window. In the next pane, you can "**Choose the Computational Engine**". To optimize the molecular geometry, we will use the very efficient **Mopac** code, which gives access to a multitude of semi-empirical QM methods. Clicking again on the purple arrow in the lower right corner of the window will open the "**Configure Mopac Job Options**" pane.

Here, you can choose the type of calculation, the level of theory, the charge, the multiplicity, and a name for your job. Please choose a name, Geometry Optimization, PM7, 0, and Singlet (you can try other values later if you wish).

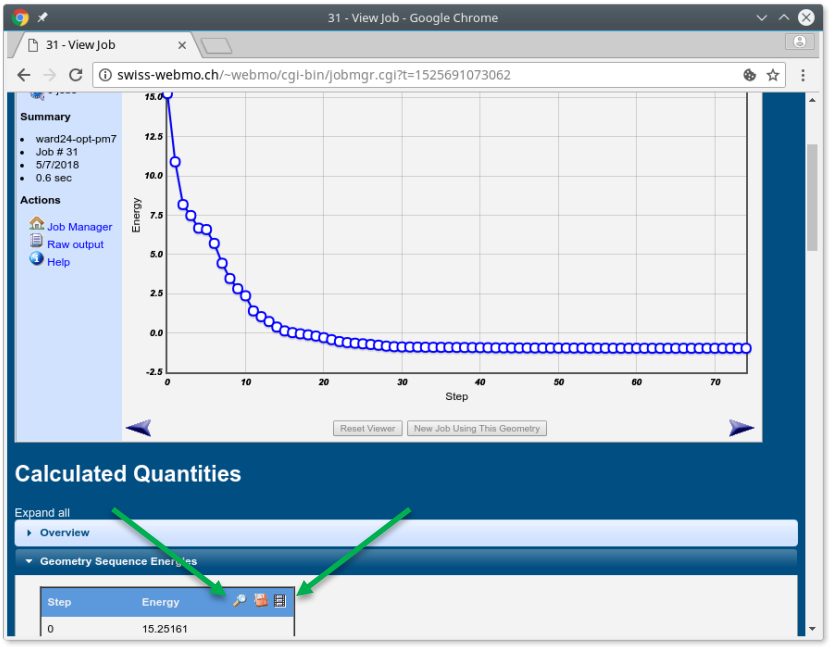


In the "**Advanced**" tab you can leave the default values (Precise Optimization checked). When you click on "**Generate**" in the "**Preview**" tab, you can see the text file that serves as an input to Mopac for this job.



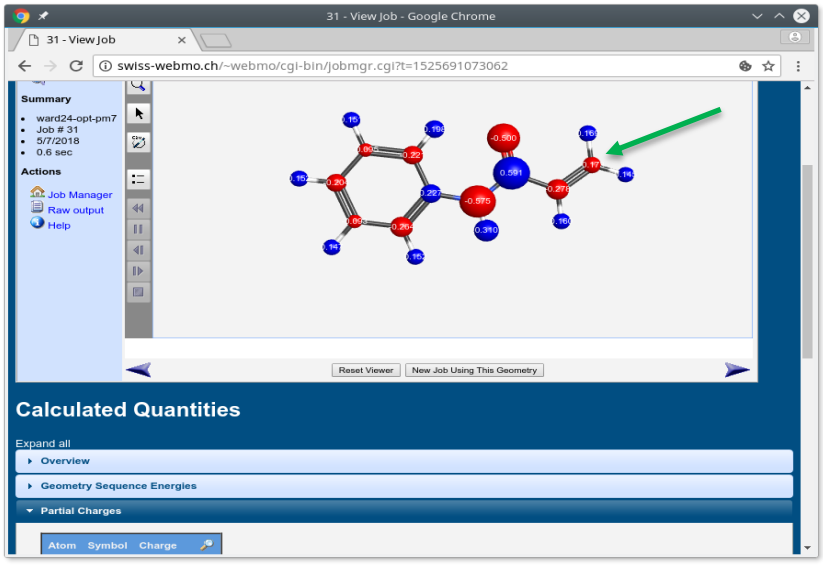
The first line contains the keywords, the second line the name, and the following lines, after a blank, contain the starting geometry of the molecule in internal coordinates. It is possible to introduce modifications to the file at this point.

By clicking on the **purple arrow** in the lower right corner of the window you can submit the calculation (choose "**OK**" in the popup window) and go back to the WebMO Job Manager. The calculations should only take a few seconds to complete. By clicking on the **magnifying glass** on the right side, you can now see the results of the calculation. In the upper part of the window you have the molecule viewer, and by scrolling down you can access some Calculated Quantities.



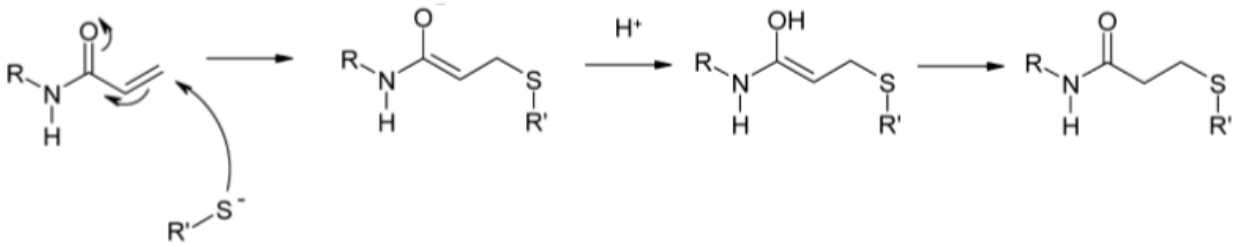
Clicking on the magnifying glass in the "**Geometry Sequence Energies**" section will display the energy (in kcal/mol) of each structure, while clicking on the negative will play an animation of each 3D structure during the optimization routine.

Clicking on the magnifying glass in the "**Partial Charges**" section will display atoms according to their net Mulliken charge, red for negative charges and blue for positive charges. Please note the charge of the reactive terminal carbon atom in the table on page 7.



Task 2: Comparing three small organic molecules

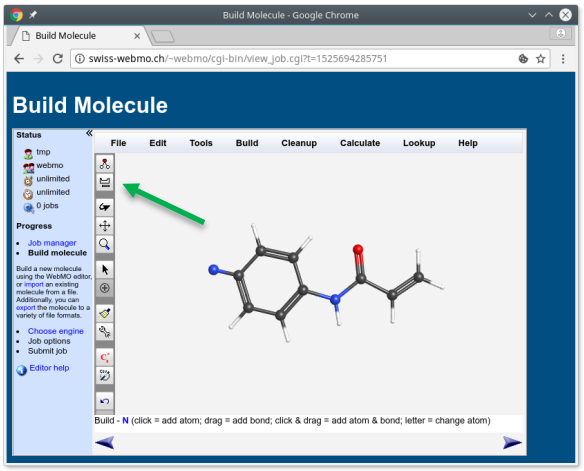
We will now compare the quantum chemical properties of three acrylamides with different substituents on the phenyl ring. The first molecule is the one studied in the previous exercise (ward24), which does not carry any substitution on the phenyl ring. The other two compounds are ward20 with a 4-amino substitution and ward31 with a 4-cyano substitution. Their respective reactivities for addition of glutathione (GSH), a model compound for cysteine, from the reference cited on page 2 are given below.



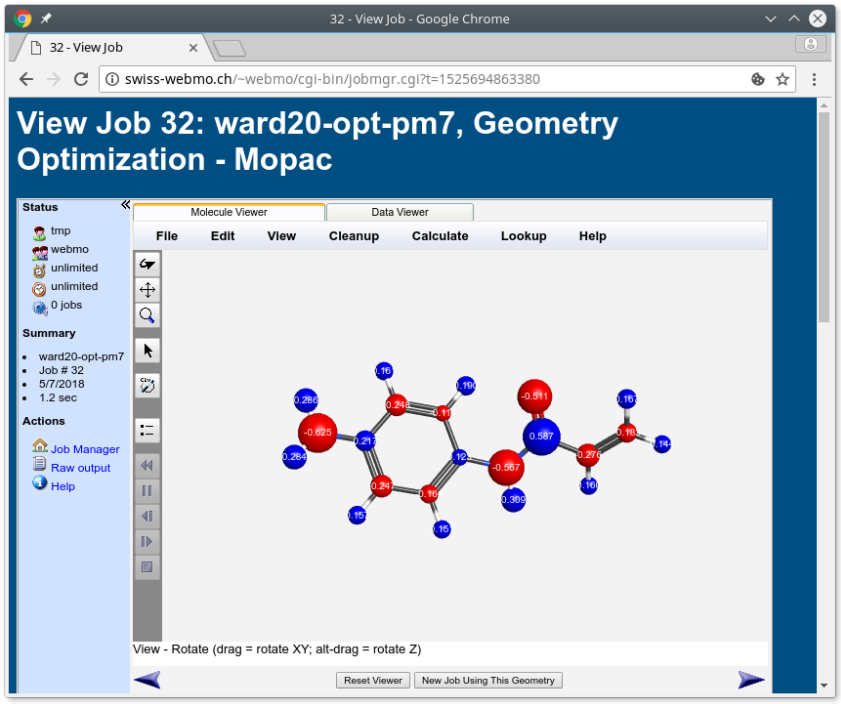
Compound	Substituent	GSH $t_{1/2}$ (min)	Log k_{GSH} ($M^{-1}s^{-1}$)	Charge	E_{LUMO} (eV)
20	4-amino	816	-2.51		
24	-	299	-2.08		
31	4-cyano	32.9	-1.12		

First, we will carry out the same geometry optimizations for compounds ward20 and ward31 as we have done for ward24. To this end, we can start from the optimized structure of ward24 and choose "**New Job Using This Geometry**" form the lower part of the Molecule Viewer. This will bring you back to the "**Build Molecule**" pane.

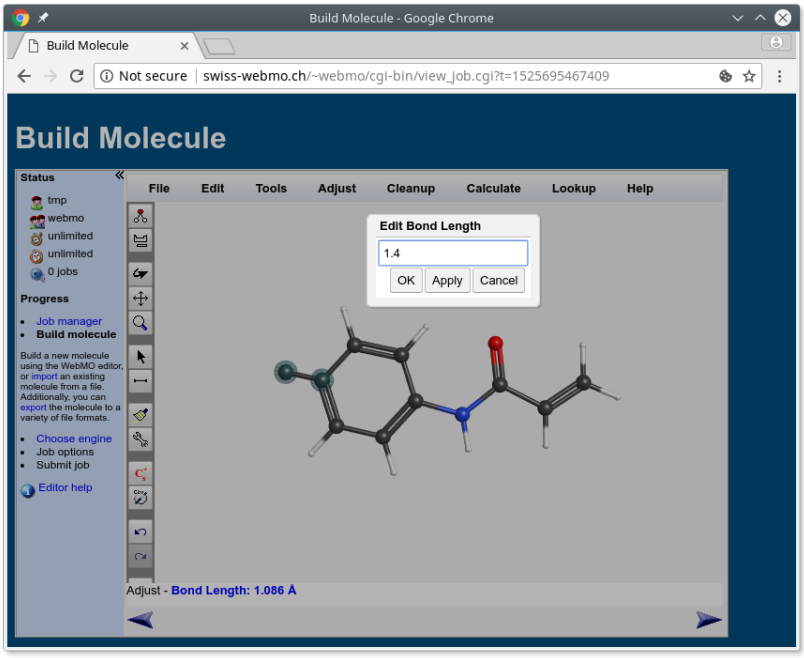
To add a 4-amino group to the compound, click on the **periodic table symbol** on the left tool bar (2nd from top) and choose the nitrogen (N), then click on the phenyl hydrogen in *para*-position. This will replace the hydrogen by a nitrogen.



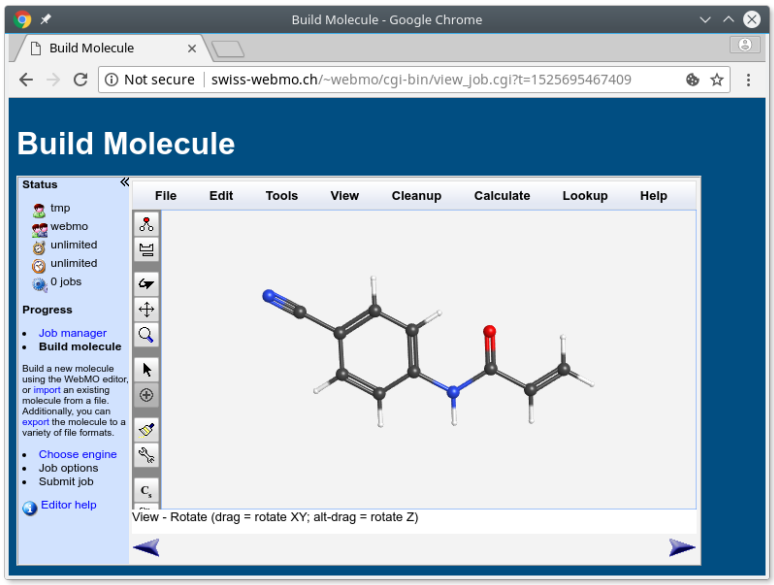
Using the menu “**Cleanup>Add Hydrogens**”, add hydrogen atoms to the group, then clean up the structure as shown before (idealized or molecular mechanics). Choosing the same parameters as before (except for the name), carry out a Geometry Optimization with PM7. Again, check the energy convergence and note the charge of the terminal carbon atom in the table on page 7.



In the Job manager, go back to the structure of the unsubstituted compound ward24 and choose “**New Job Using This Geometry**”. Now, place a carbon atom in place of the *para* hydrogen. Then choose the “**Adjust**” tool (arrow, 6th button from top), select the newly placed carbon atom and the one it is bound to, and select the “**Adjust Bond Length**” button (line, 7th button from top). Apply a bond length of 1.4 Å.



Now place a nitrogen atom in the proper position, and adjust its distance to the previously placed carbon atom to 1.1 Å. Under "**Cleanup>Generate Bonds**", generate the CN triple bond (*attention*: for some reason this seems to remove the alternating aromatic double bonds, please add them back). You can clean up the final structure as before.

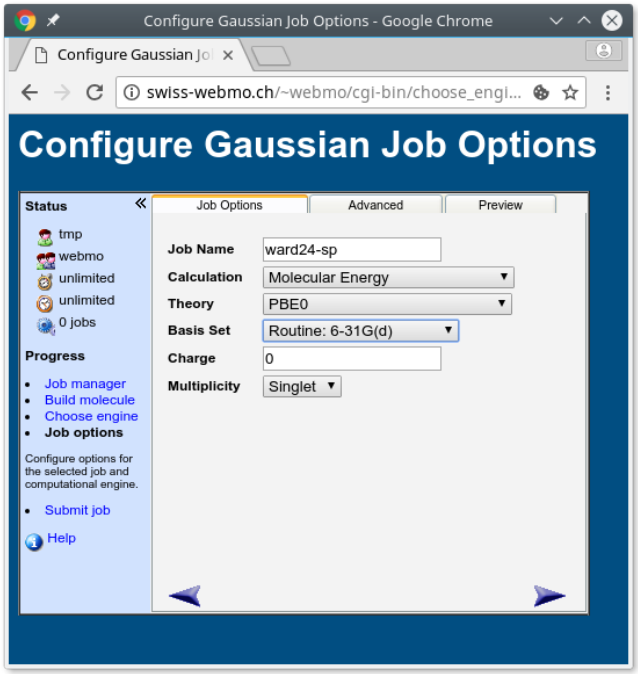


Optimize the geometry of this compound, check the energy convergence and note the charge of the terminal carbon atom.

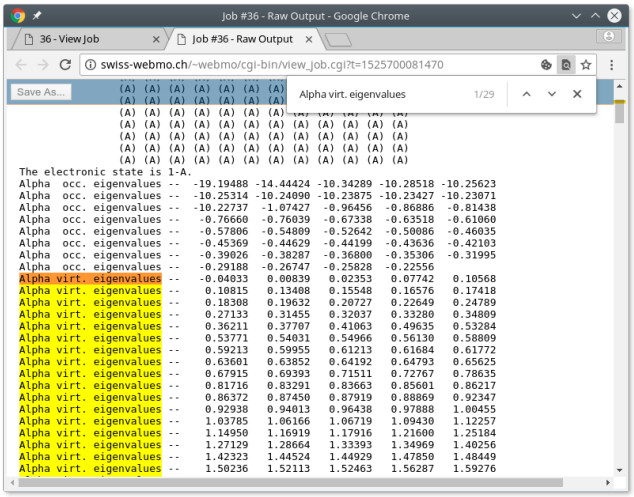
Task 3: Calculating Orbital Energies

The lowest unoccupied molecular orbitals (LUMO) of the acrylamides play an important role for the addition reaction. In order to get insight into their shape and energies at reasonable computational cost, carry out single-point DFT calculations at the PM7-optimized geometries.

In the WebMO Job Manager, choose one of the previously optimized compounds and click on "**New Job Using This Geometry**". Without modifying the geometry, choose this time **Gaussian** as the Computational Engine. In the "**Configure Gaussian Job Options**", choose a job name, Molecular Energy as Calculation, PBE0 as Theory, and the Routine Basis Set (6-31G(d)). The calculation should not take more than a few minutes.



When the calculation is finished, click on “**Raw output**” in the left pane of the results window. This will open a new tab or window in your browser, displaying the output text file of Gaussian09. With the search function of your browser, search for “**Alpha virt. eigenvalues**”. The first value in this line corresponds to the energy of the LUMO.



Note this energy and repeat the same calculations for the other two molecules. Is there a relationship between the experimentally measured GSH reactivity and the energy of the LUMO? Why?

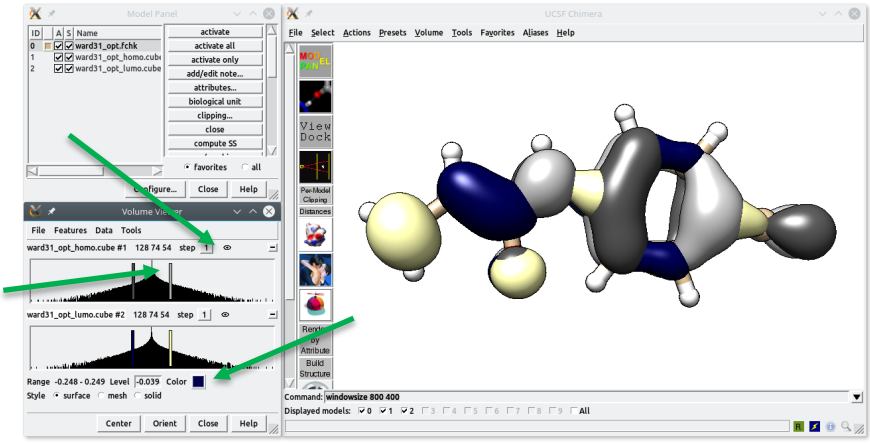
- *Complete the table from page 7 with your calculated values and explain how the calculated charges and orbital energies relate to the experimentally measured reactivities.*

Unfortunately, the free version of WebMO does not allow to display orbitals, therefore we use UCSF Chimera for this task.

Optional Task: Displaying and comparing molecular orbitals

Among the files provided for this exercise, there are three Gaussian09 formatted checkpoint files (*.fchk). The geometries in these files should correspond closely to the PM7-optimized geometries that you generated. These as well as the provided *.cube files were generated with the Gaussian09 utilities *formcheck* and *cubegen*.

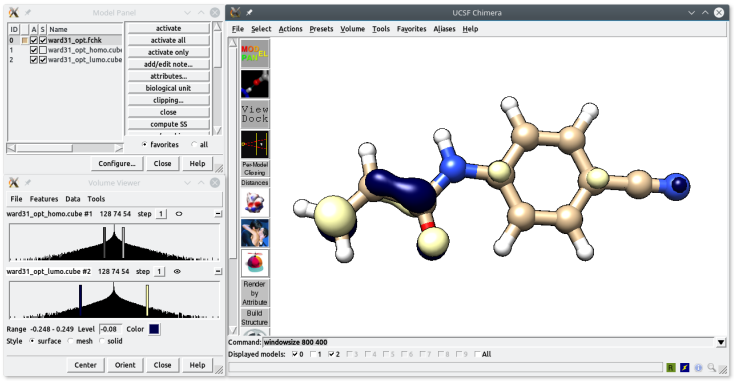
Open the molecular structure from the *ward31_opt.fchk* file and display the molecule in the way you prefer. Then open the files containing the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), *ward31_opt_homo.cube* and *ward31_opt_lumo.cube*. By default, Chimera recognizes them as volume files and displays a positive and a negative isosurface for each orbital.



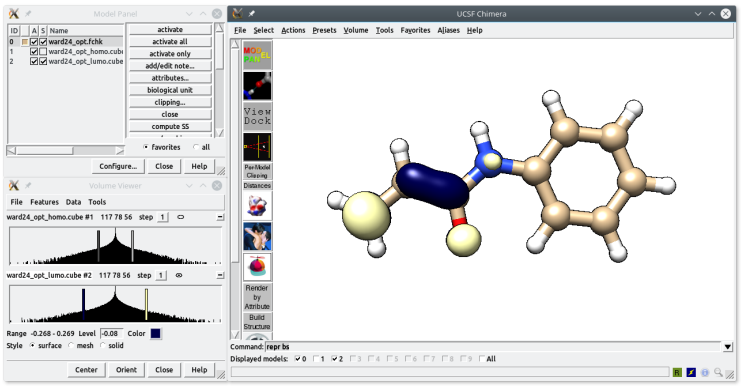
By clicking on the eye in the Volume Viewer window on the right side of each orbital, you can display/hide each orbital. By sliding the color bars in the histograms, you can change the isosurface values and by clicking on the “Color” button you can change the color of each lobe.

- *How many nodal surfaces (surfaces with zero probability) do you see in the HOMO and in the LUMO? Which orbital has more nodal surfaces? In order to answer, set the isosurface values to very low numbers, for example to ±0.001.*

- *Now display only the LUMO and set its isosurface values to -0.08 and 0.08. On which atoms does it have important contributions in each of the 3 molecules?*



Save and close this Chimera session, then open the files *ward24_opt.fchk*, *ward24_opt_homo.cube* and *ward24_opt_lumo.cube*. Display only the LUMO and chose isosurface values of -0.08 and 0.08 as done for compound ward31. What is the difference? What is the relation to the reactivity of the two compounds?



Do the same visualization for the third compound (ward20).

- *What is the difference in the LUMOs between the three compounds? Can you relate this to their reactivity?*