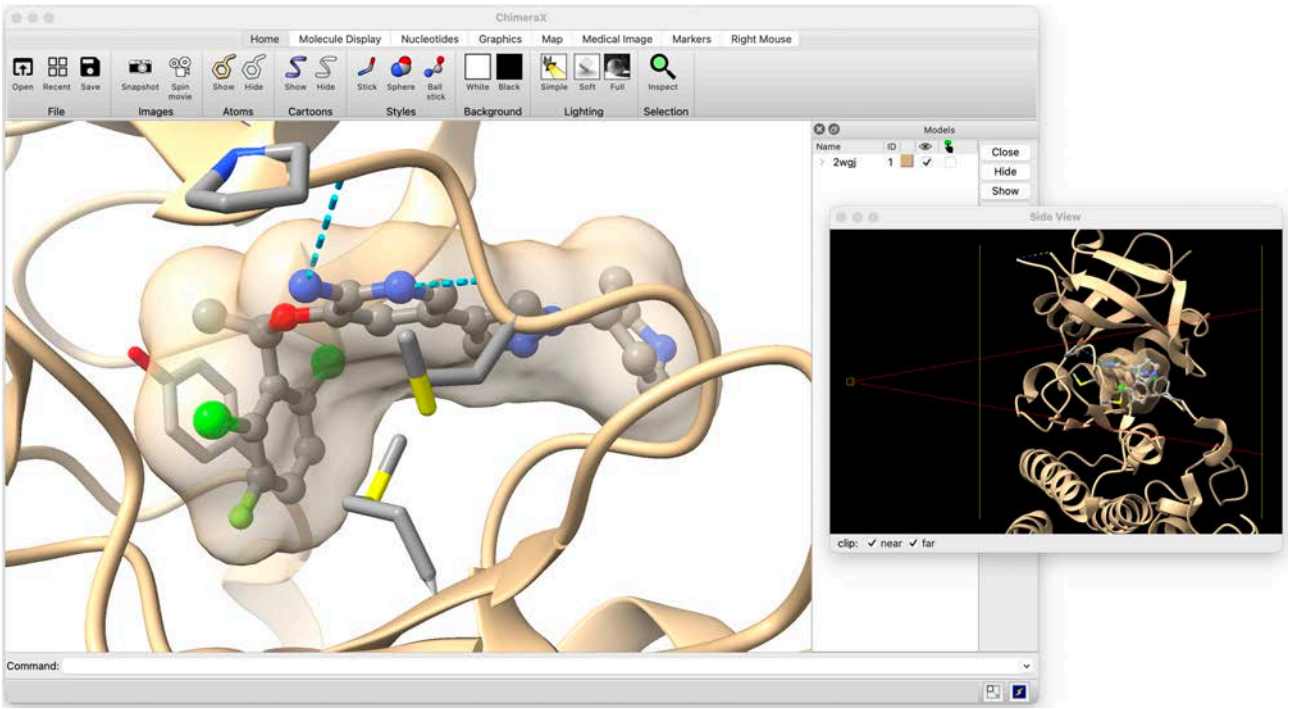
Practice – Session 1
3D structure visualization and high quality imaging using UCSF ChimeraX



Antoine Daina, Ute Röhrig, Vincent Zoete – 2024
Contact : vincent.zoete@unil.ch

Table of Contents

Presentation of UCSF ChimeraX

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

ChimeraX is the state-of-the-art visualization program from **the Resource for Biocomputing Visualization, and Informatics** at UC San Francisco. It is free for academic, government, nonprofit, and personal use.

ChimeraX can be used on Windows, Mac and Linux computers.
It can be downloaded at the following address:
https://www.cgl.ucsf.edu/chimerax/download.html#release

When using ChimeraX in a publication, one should cite at least one of these references

> UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. *Protein Sci.* 2021 Jan;30(1):70-82
>
> UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. *Protein Sci.*2018 Jan;27(1):14-25

and include an acknowledgment like

> "Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases".

The following exercises will be used as tutorials to introduce some basic commands of ChimeraX.

You are highly encouraged to test the different options that you will encounter in the different menus.

Green boxes contain commands to type in the command line interface.

Pink boxes contain very important notes.

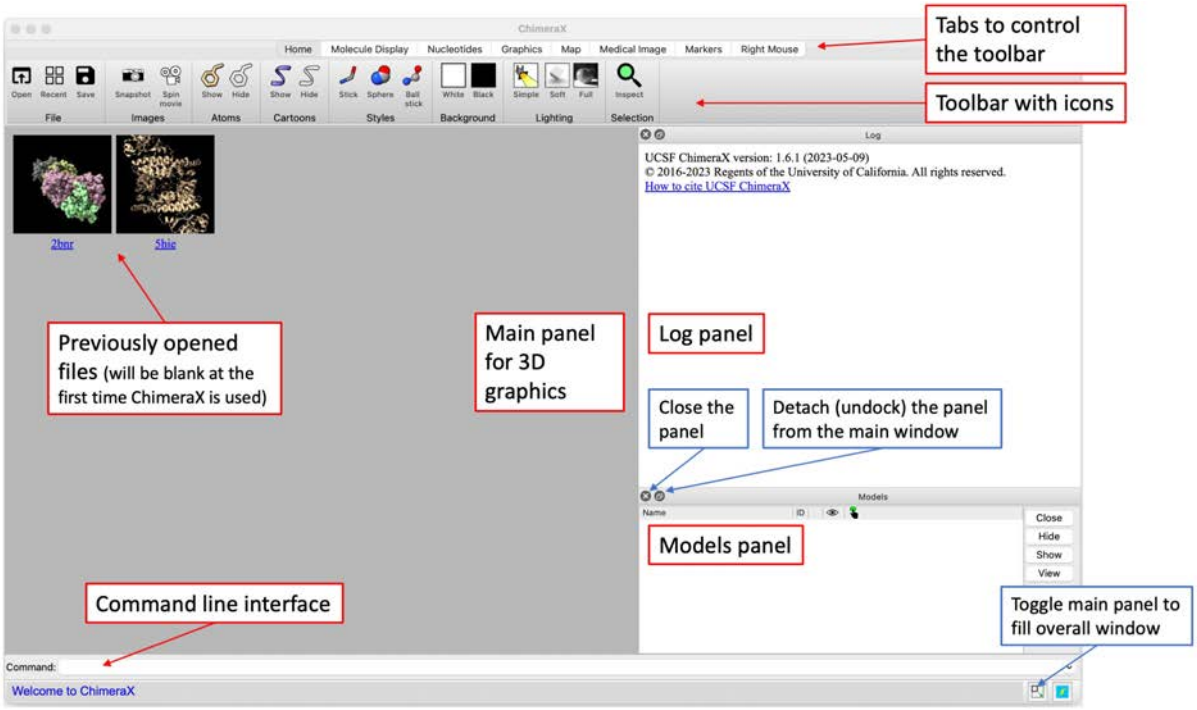
Yellow boxes contain notes that allow users to explore possibilities beyond those exposed in the main practice.

ChimeraX basics

The ChimeraX window

ChimeraX has an all-in-one window interface in which various ***panels*** can be shown, hidden, resized and repositioned.

However, panels can also be detached (undocked) from the main ChimeraX window and subsequently reinserted (docked).

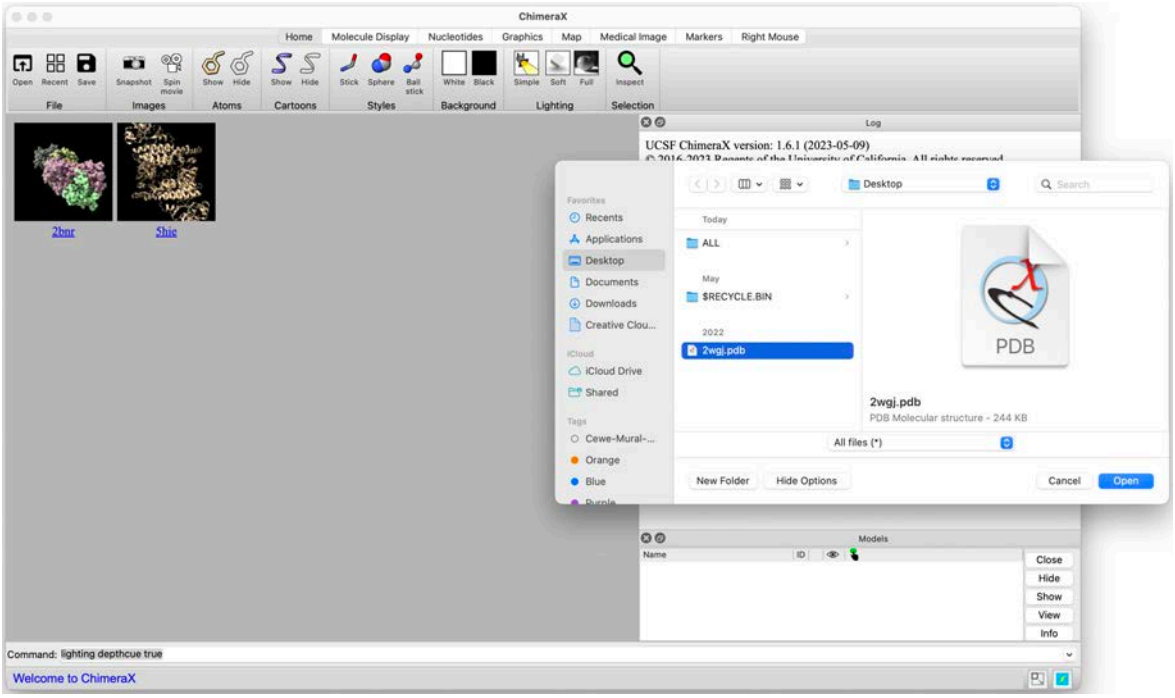


Clicking on the toggle icon  at the bottom right corner allows expanding the main panel (with 3D graphics) to occupy the entire overall window, except the toolbar and command line interface, without losing the overall window configuration. This means that you will find again all the open panels, in their current configuration, when clicking on the  icon.

Loading a structure into ChimeraX

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

1) If the structure file is present in the user's computer, choose the menu item "**File/Open**". Then select "2wgj.pdb" and click "Open".



Note: *Under Linux or Mac OSX, the structure file (for instance 2wgj.pdb) can also be loaded using the following commands in a terminal:*
(standard Linux) > chimerax 2wgj.pdb
or
(standard macOS) > /Applications/ChimeraX-1.6.1.app/Contents/MacOS/ChimeraX 2wgj.pdb

2) The structure can also be fetched directly from a source (database or web service) over the internet, when available, using the **open** *source*:*identifier* **format**:*format-name* command.

The following table lists some of the available sources and recognized formats that can be used with the open command.

database or web service	*source*	description	format-name
RCSB Protein Data Bank	**pdb**	atomic coordinates and associated annotations	**mmcif** (default) or **pdb** or **mmtf**
	rcsb_bio	atomic coordinates for biological assembly or assemblies, see **maxAssemblies**	(no choice, **mmcif** or **pdb**depending on the specific entry)
Protein Data Bank in Europe (PDBe)	**pdbe**	atomic coordinates and associated annotations	**mmcif** (default) or **pdb**
	pdbe_bio	atomic coordinates for biological assembly or assemblies, see **maxAssemblies**	**mmcif**
	pdbe_updated	atomic coordinates and associated annotations, updated to include small-molecule connectivity and binding-site information as described in Velankar *et al.*, *Nucleic Acids Res* **44**:D385 (2016)	**mmcif**
Protein Data Bank Japan (PDBj)	**pdbj**	atomic coordinates and associated annotations	**mmcif**
PDB Chemical Component Dictionary (CCD)	**ccd**	idealized structure of PDB chemical component specified by CCD ID (residue name)	**ccd**

PDBe Electron Density Server (EDS)	**eds** **edsdiff**	(2Fo-Fc) electron density map for a PDB entry; not available for all PDB entries (Fo-Fc) electron density difference map for a PDB entry; not available for all PDB entries	**ccp4**
Electron Microscopy Data Bank (EMDB)	**emdb**	electron density map (from wwpdb.org to .edu/.gov hosts, Chinese mirror site to .cn hosts, Japanese mirror to .jp hosts, otherwise ebi.ac.uk; Japanese site uses https, others ftp)	**ccp4**
EMDB ebi.ac.uk	**emdb_europe**	electron density map (command option **transferMethod** can be **https** or **ftp**)	**ccp4**
AlphaFold	**alphafold**	theoretical protein structure modeled by AlphaFold and specified by UniProt name or accession number (details...)	**mmcif**
Crystallography Open Database	**cod**	crystal structures of small molecules and inorganic compounds	**corecif**
PubChem3D	**pubchem**	modeled atomic coordinates specified by PubChem compound identifier (CID); available for most but not all entries in PubChem Compound (partial charges are assigned as the atom attribute named **charge**)	**sdf**
UniProt	**uniprot**	protein sequence (command option **associate**) (associated tool: **Sequence Viewer**; see UniProt Sequence Features)	**uniprot**

Adapted from https://www.rbvi.ucsf.edu/chimerax/docs/user/commands/open.html

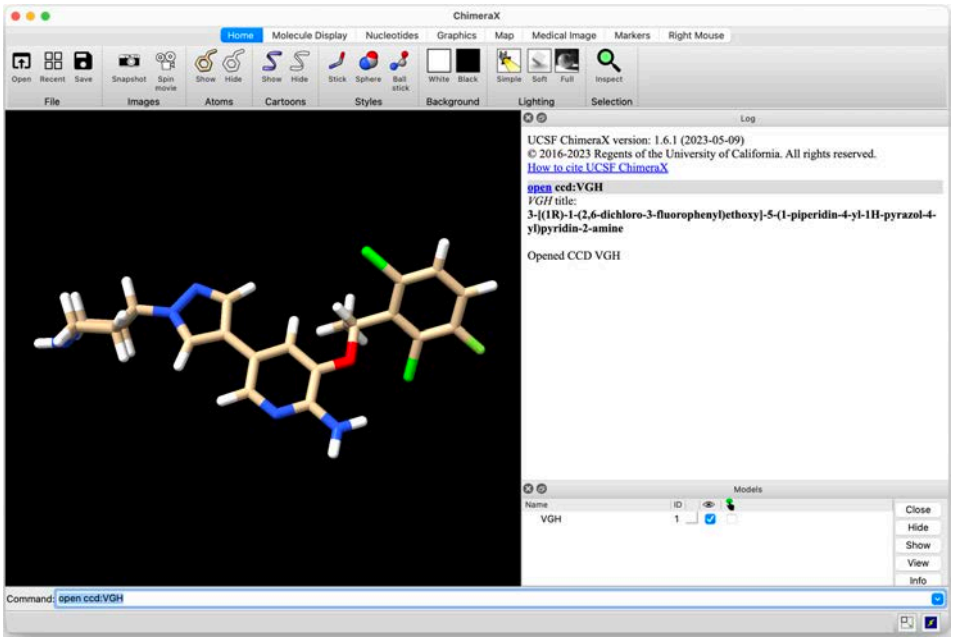
To begin this tutorial, we will load the 3D structure of a single small molecule, crizotinib, which is an FDA approved drug used in the targeted therapy of cancer.

We will fetch this structure from the Chemical Component Dictionary (CCD) of the Protein Data Bank (PDB). This dictionary contains all small molecules (ions, ligands, cofactors, solvents, etc.) that have been crystallized along with macromolecules, and are present in structure files stored in the PDB.

You will learn elsewhere how to find the identifier of a given small molecule in this database. For now, you just need to know that the CCD identifier of crizotinib is **VGH**. Consequently, type the following command in the command line interface, then press Enter.

```
open ccd:VGH
```

Note that we did not specify the format name in this command, since only the ccd format is available for this source.



Moving, rotating and zooming

You can translate and rotate the molecule, as well as control the zoom using the following mouse or trackpad gestures:

	Mouse	Mac trackpad	Windows trackpad
Rotate around X,Y	Left-click-holding *or* Right-click-holding	Click & hold *or* two-finger drag	Click & hold
Rotate around Z	Left-click-holding with cursor near edge of window	Click & hold with cursor near edge of window	Click & hold with cursor near edge of window
Translate	Middle mouse button	Option + click & hold *or* three-finger drag	Ctrl-Alt + click & hold
Zoom	Mouse scroll wheel	Two-finger pinch *or* four-finger drag	Two-finger drag
Select	Ctrl + left mouse button to select one element (e.g. an atom) *and* Shift to toggle or add to selection	Ctrl + click to select one element (e.g. an atom) *and* Ctrl + Shift + click to add to a selection *and* Ctrl + Click & hold to select a region	Ctrl + click to select one element (e.g. an atom) *and* Ctrl + Shift + click to add to a selection *and* Ctrl + Click & hold to select a region

Controlling panels

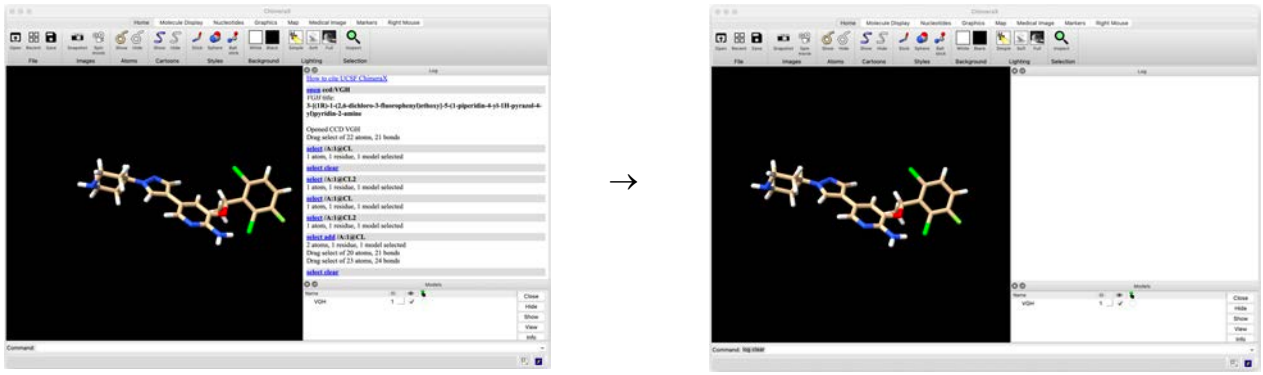
The Log panel can be useful to:

- keep trace of the previous actions and
- translate actions performed using the mouse or menu into commands that can be used in the command line interface.

However, it can also become crowded, or consume space that could be better attributed to another, possibly more useful panel.

To empty the Log panel, type the following command in the command line interface, then press Enter.

```
log clear
```

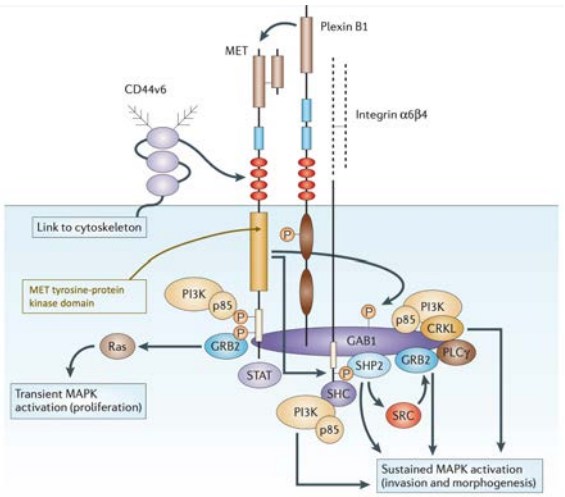


If you want to remove the Log panel, just click on its cross icon, or toggle off the option "**Tools > Log**" in the main menu. To restore a removed Log panel, you can toggle on the option "**Tools > Log**".

These commands can be used for other panels, including the Model panel, the command line interface, or the toolbar.

Exercise 1 – First steps with UCSF ChimeraX

The 3D structure of the tyrosine-protein kinase domain of the Hepatocyte Growth Factor Receptor, a.k.a. HGFR or c-MET, in complex with the FDA-approved drug crizotinib, obtained by X-ray crystallography, will be used as an example (2wgj in PDB).
MET is essential for embryonic development, organogenesis and wound healing. However, abnormal activation of MET or of the MET pathway plays an important role in the development of cancer through activation of downstream key oncogenic pathways (e.g. Ras or PI3K pathways). Consequently, MET constitutes an attractive oncology target for therapeutic intervention. Several MET inhibitors were designed, including crizotinib. The latter was approved by the FDA in 2011.



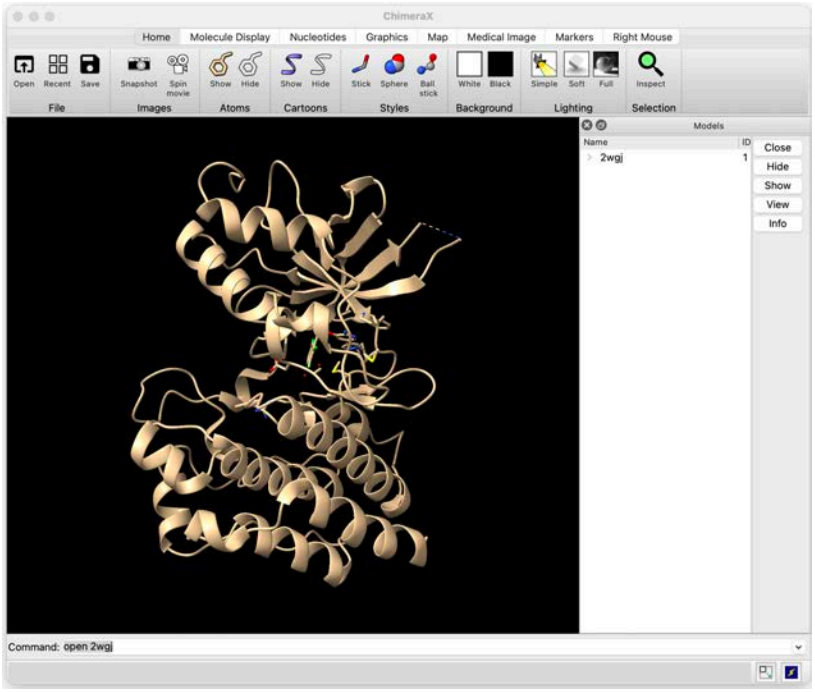
MET pathway. Adapted from *Nature Reviews Cancer* volume 6, pages637–645 (2006)

Loading a 3D structure from the Protein Data Bank

As we have seen in the ChimeraX basics section, structures can be fetched directly from a database over the internet, when available. To load the crizotinib:MET complex of interest, whose ID is 2wgj in the PDB, just execute the following command using the command line interface:

```
open 2wgj
```

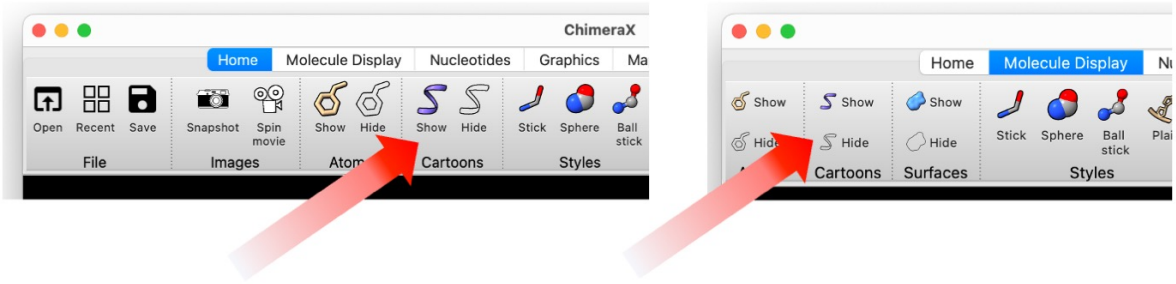
There is no need to specify the source database, since the PDB is used by default when nothing is specified.
Once the structure has been loaded, the complex between MET and crizotinib should appear in a representation automatically chosen by ChimeraX:



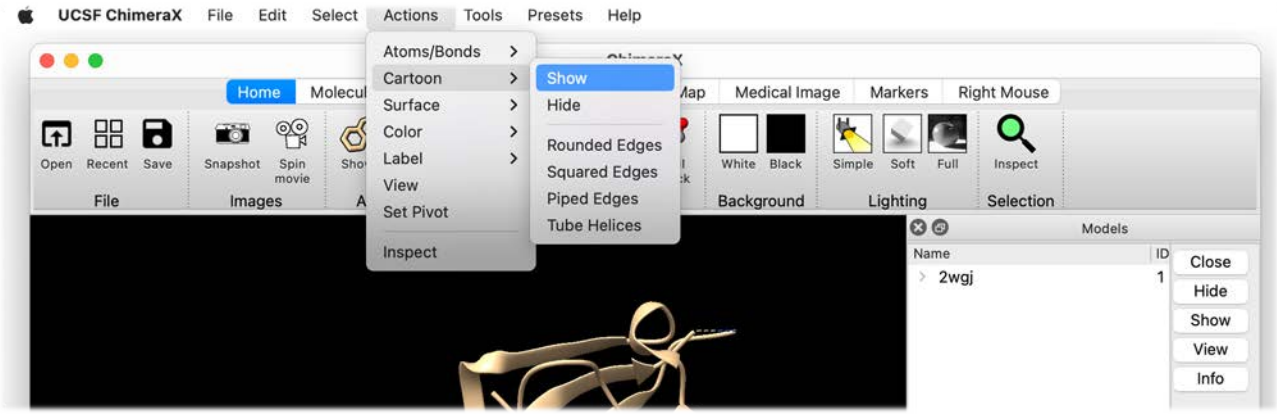
Cartoon representation

The secondary structure elements can be shown using the so-called *cartoon* representation, in which protein helices and beta-strands are represented by ribbons taking the shape of helices and arrows, respectively.

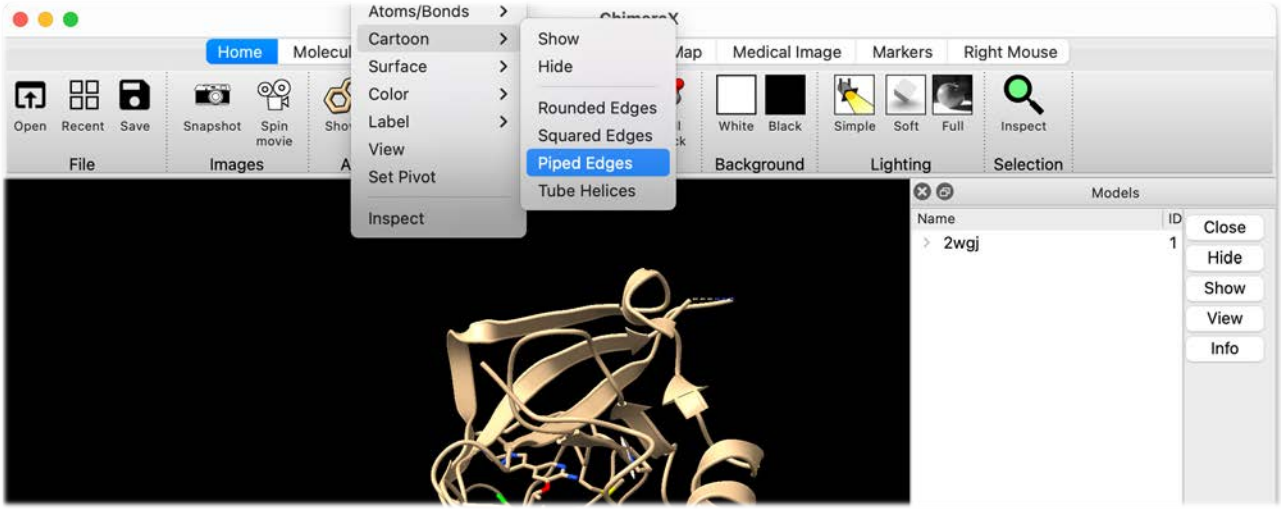
This cartoon representation of the secondary structure elements can be switched on or off by clicking on the “Show” or “Hide” icons of the “Cartoons” section of the toolbar. This section is present in the toolbar of the “Home” and “Molecule Display” tabs.



It is also possible to switch on or off the cartoon representation by choosing “Show” or “Hide” in the “**Actions>Cartoon**” menu, respectively.



The “**Actions>Cartoon**” menu can also be used to select one of the variants available for the cartoon representation: “**Rounded Edges**”, “**Squared Edges**”, “**Piped Edges**” and “**Tube Helices**”. You can try them all by clicking the corresponding menu item.



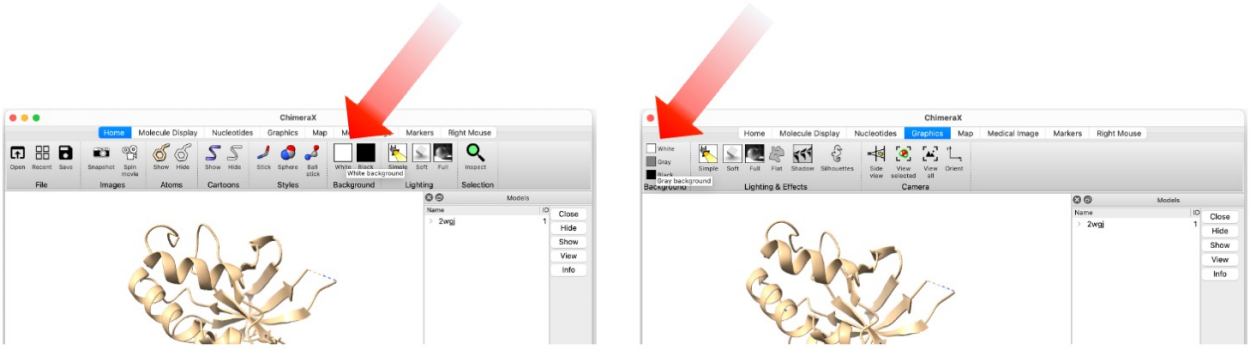
Select the “**Rounded Edges**” to continue this tutorial. This corresponds to the default cartoon representation.

Note: when necessary, you can type “**dssp**” in the command line interface to recalculate the secondary structure elements of the protein. This uses an implementation of the Kabsch and Sander algorithm for **d**efining the **s**econdary **s**tructure of **p**roteins. Biopolymers 22:2577 (1983).

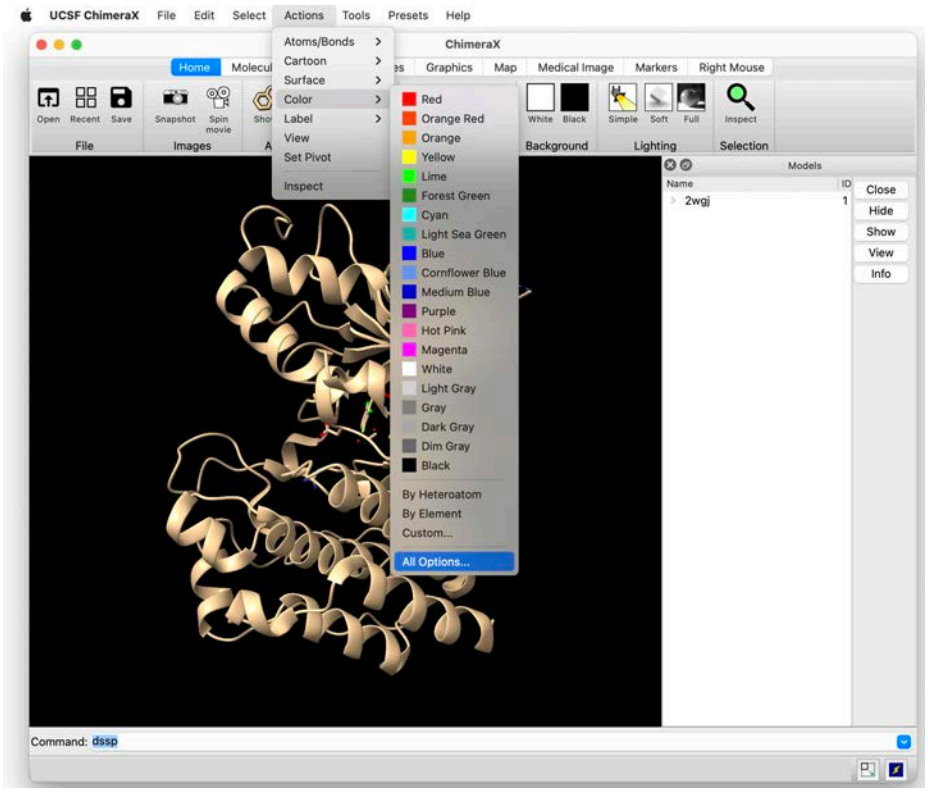
Changing background color

By default, ChimeraX colors the background in black. This generally provides a better contrast to images. However, it also consumes more ink when printing.

To change the color of the background, it is possible to click on the “White” or “Black” squares of the “Background” section of the toolbar, which is available when the “Home” tab is selected, or alternatively on the “White”, “Gray” or “Black” squares when the “Graphics” tab is selected.

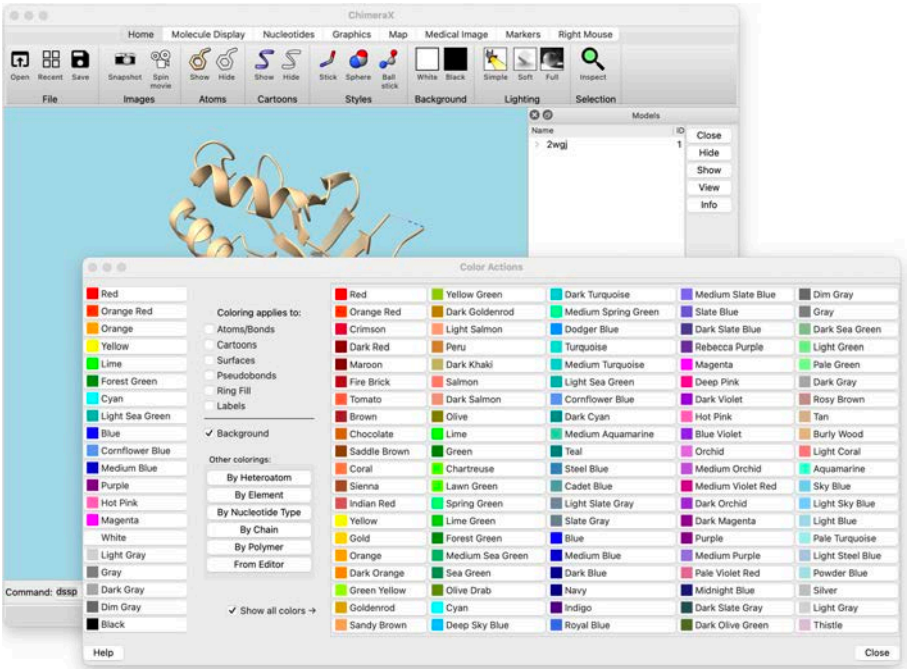


The above option allows to rapidly switch between a white, gray or a black background only. However, it is possible to apply whatever color to the background. For this, click on “**Actions>Color>All Options…**” to open the full “Color Actions” panel.

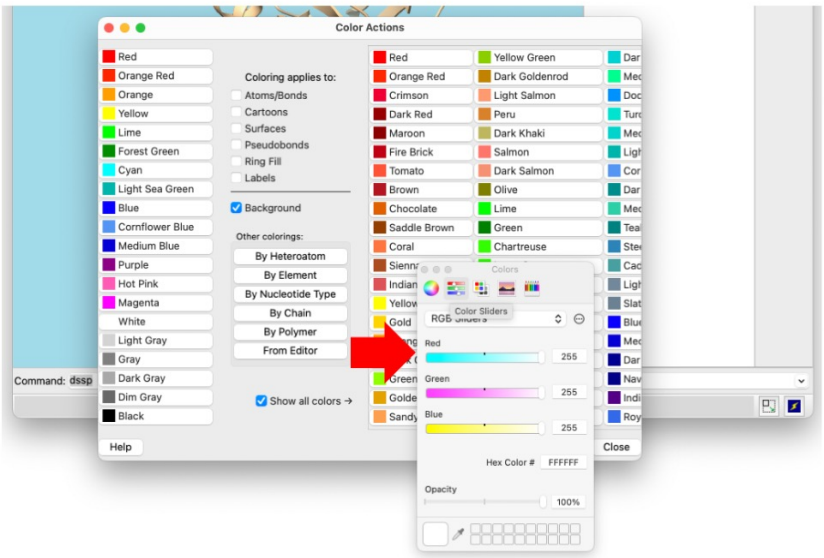


The “Color Actions” panel provides users with a sample of predefined colors on the left. This set can be extended by checking the “Show all colors” box.

This panel also provides the possibility to select which item or property the coloring should be applied to, i.e. atoms, bonds, cartoons, surfaces, … To change the color of the background, check the “Background” box (this will uncheck all the other possibilities), then click on a predefined color.



Coloring can also be done using a user-defined color. For this, click on “From Editor” in the “Color Actions” panel. This opens a new “Colors” panel, which allows user to define a color using a “Color Wheel”, “Color Sliders”, “Color Palettes”, “Image Palette” or “Pencils”.

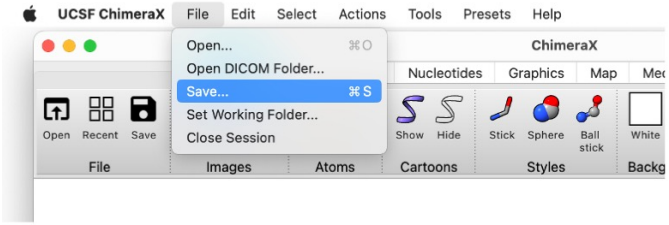


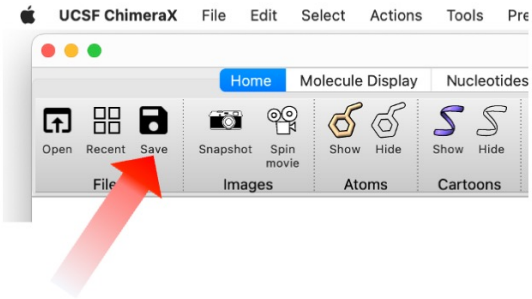
The “Color Sliders” also offers the possibility to choose a color using a Hex color code (see here for more explanations and examples: https://en.wikipedia.org/wiki/Web_colors).

You can try several colors. Finally, apply the white color to the background to continue the tutorial.

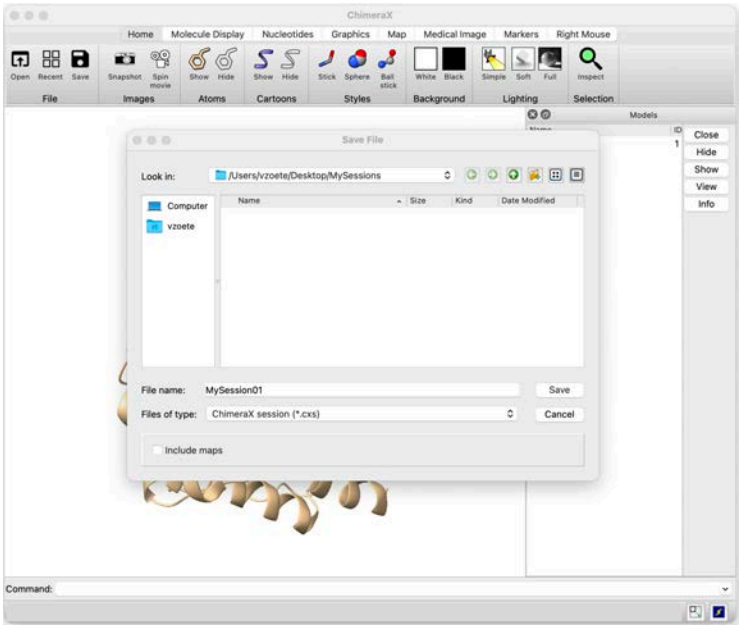
Saving the session status

The ChimeraX session (the actual state and representation) can be saved for future use or modifications, or to share it with a collaborator. To do so, open the “Save File” panel by clicking on “**File>Save...**” in the main menu or click on the “Save” button in the toolbar, when the “Home” tab is selected.





In the “Save File” panel, select a location on your hard drive, choose “ChimeraX session (*.csx)” as file type, provide or select a file name, and click “Save”.



Note: Altering the structure, for instance by deleting some atoms, cannot be undone in ChimeraX. It is thus highly recommended to save a session before any such action, to be able to restore the previous state in case of error.

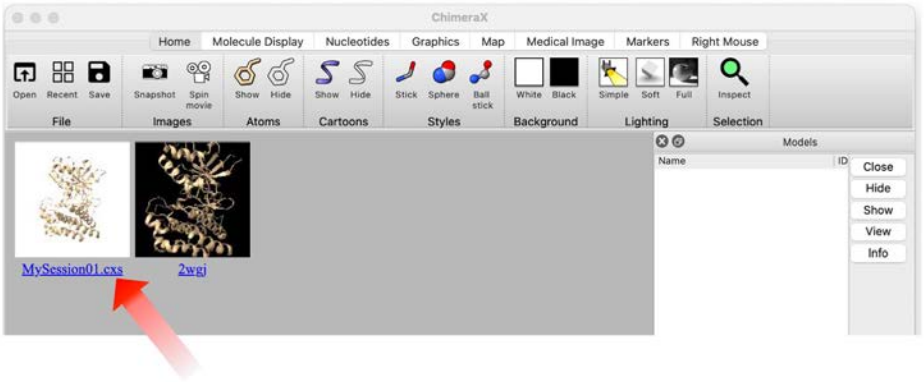
Closing the session.

The session can be closed using the “**File>Close Session**” item in the main menu.

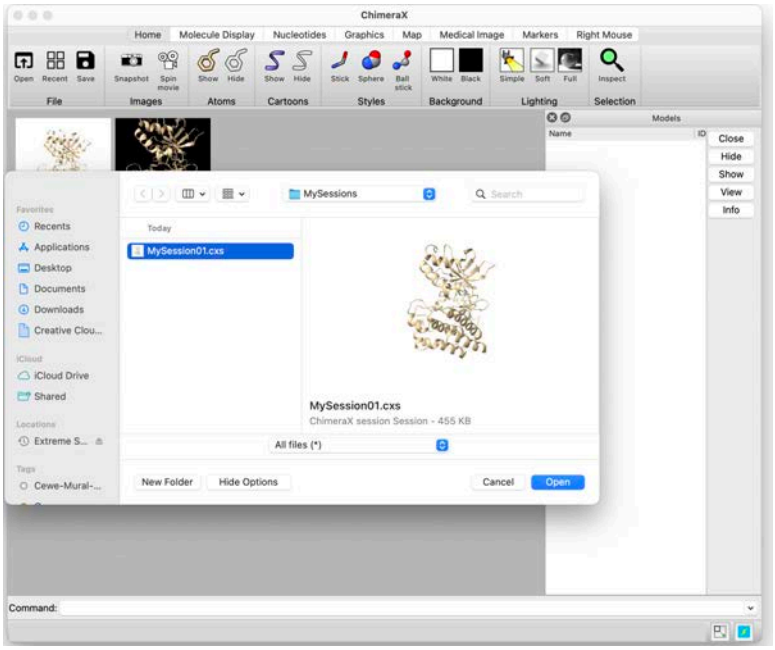
Restoring a previous session

There are two ways to restore a saved session.

First, if the session file was freshly created on your computer, it will appear on the left-hand gray panel. In this case, you can simply click on the corresponding session to restore it exactly in the state it was at the last “Save session” action.



The second method is useful if the session file does not appear on the left-hand panel with other previous structures and sessions, because it is too old or it was sent by a collaborator. In these cases, select "**File>Open…**" in the main menu or click on the "Open" button in the toolbar when the "Home" tab is selected. Then, in the new panel, select the file (with a *.cxs extension) and click "Open".



Note: **Session files are system independent.** It is therefore possible to share a session file with a collaborator whose computer is equipped with a different Operating System (Mac OSX, Windows, Linux…). The collaborator will see exactly the presentation you prepared on his/her own screen. However, these session files are dependent on the version of ChimeraX used: a session file created with a recent version of ChimeraX might not be correctly restored by an older version of the software.

Changing the display of bonds and atoms. Selections

The objective of this part is to display the global structure of MET, as well as crizotinib and some of the surrounding residues of MET. Other residues will be hidden to simplify the representation. This will require to select parts of the structure and to apply different actions and representation schemes to them.

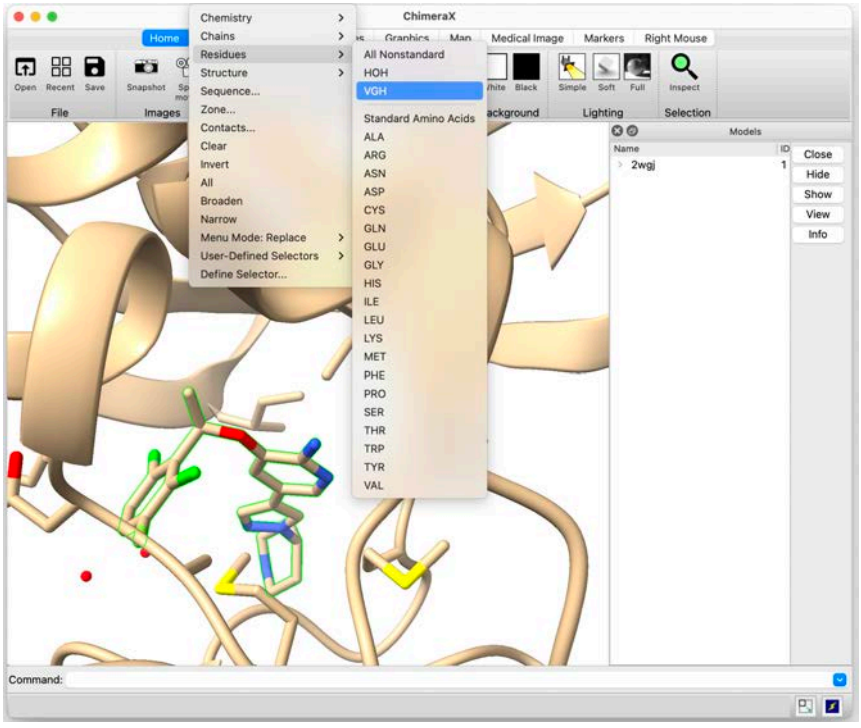
Note: ChimeraX allows to display differently each atom, bond, secondary structure element, etc. according to the user's needs to obtain the final desired representation. To achieve this, it is possible to select each of these elements independently to apply a different representation scheme to it. Item selection is therefore an essential function in ChimeraX, and several alternatives exist.

There are three ways of selecting atoms with ChimeraX: using the "Select" menu, using the command line interface, or using the mouse to select atoms from the screen.

Selecting using the "Select" menu

Click on "**Select**" in the main menu and verify that the "**Menu Mode**" is "**Replace**". Otherwise, choose it.

The first four menu items allow to select part of the structure according to the "**Chemistry**" (chemical nature of the atoms or functional groups), the "**Chains**" (i.e. a protein chain), the "**Residues**" (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 crizotinib by choosing the "**Select>Residues>VGH**" menu item. You will see that the selected parts of the structure are surrounded by thin green lines.



You can try selecting different parts of the complex: the water molecule, chain A of the system (i.e. in this case, the protein, the ligand and the water molecules), the aspartate residues, the strands, etc...

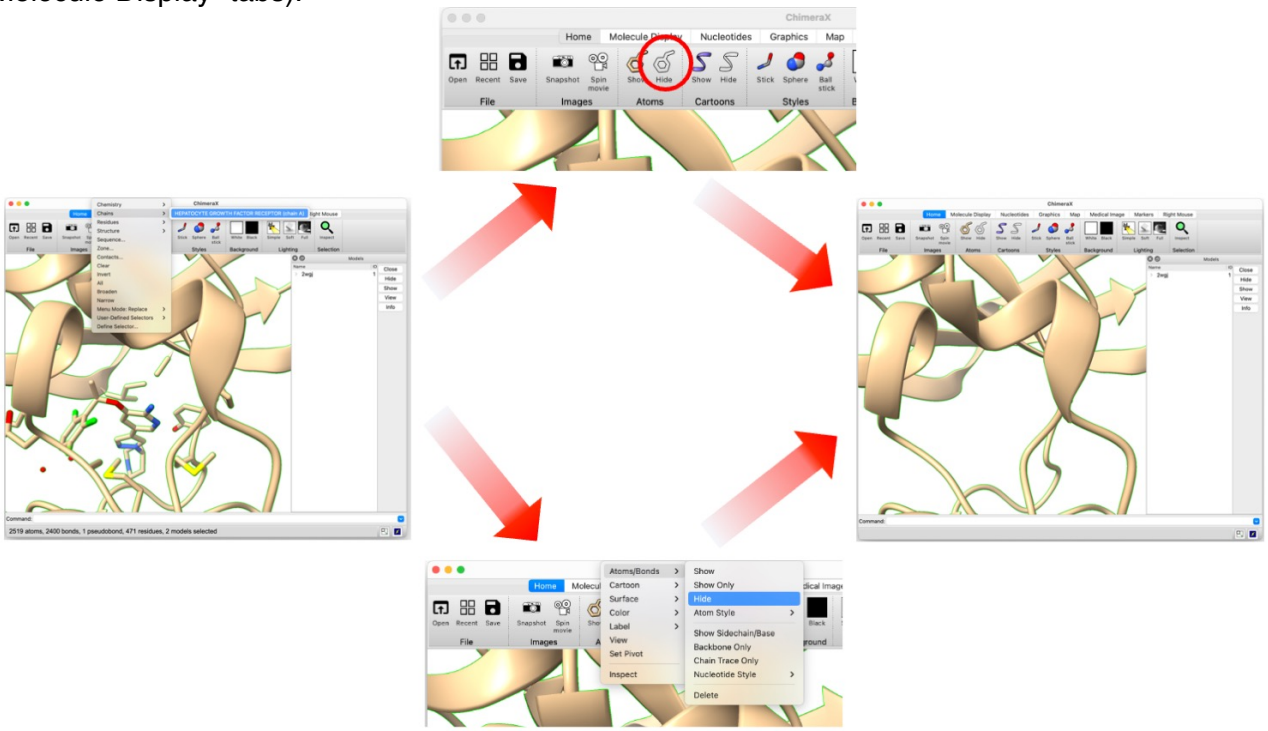
Finally, you can clear all selections using "**Select>Clear**".

Note: Most of the options of the "**Select**" menu, like the list of available protein chains or the list of residues present in the structure, are dynamically created by ChimeraX upon analysis of the content of the loaded PDB file. Therefore, they will differ between PDB files.

Note: VGH is the three-character code of crizotinib in the PDB, as can be seen in https://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/VGH or http://ligand-expo.rcsb.org/pyapps/ldHandler.py?formid=cc-index-search&target=VGH&operation=ccid.

Deleting and hiding atoms

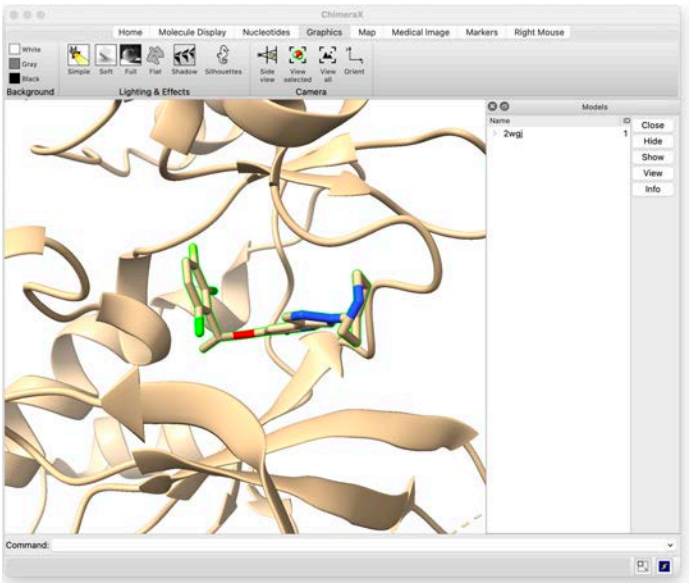
Select chain A (which groups the entire complex in this particular case) by selecting "**Select>Chain>A**". Then, hide all the selected atoms using "**Actions>Atoms/Bonds>hide**", or simply on clicking on the "Hide" button of the "Atoms" section of the toolbar (for the "Home" or "Molecule Display" tabs).



Note: There is a fundamental difference between deleting and hiding atoms in ChimeraX.
- **"delete" definitively suppresses the atoms from the system**. It will not be possible to show them again, unless the user re-loads their coordinates into ChimeraX. The "Undo" action (see below) does not work when part of the system has been deleted.
- **"hide" undisplays the atoms from the current representation but keeps them in the system**. Consequently, these atoms can still be selected and shown again.

Note: When no structural element is selected, commands are applied to all atoms.

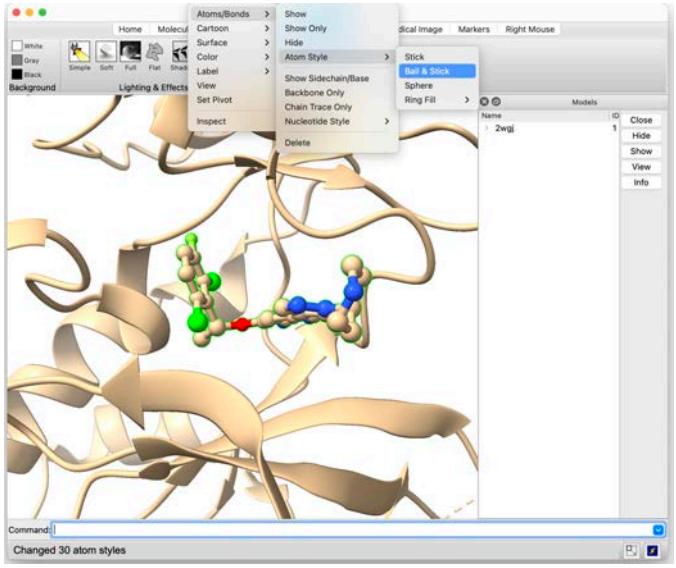
Now that the representation has been cleared and simplified, we can display crizotinib.
For this, select the residue VGH using "**Select>Residues>VGH**" and display it using "**Actions>Atoms/Bonds>Show**" or the "Show" button of the "Atoms" section of the Toolbar.



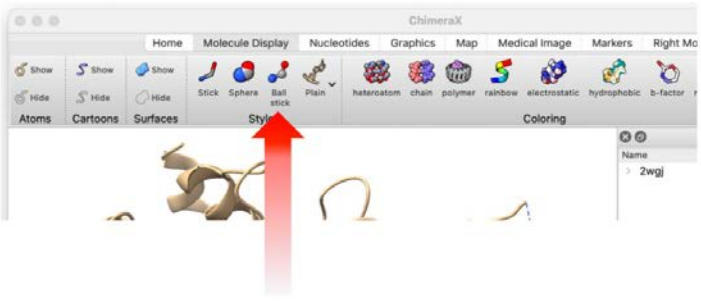
Changing atom and bond display

Verify that crizotinib (residue VGH) is still selected. If so, it should be surrounded by a thin green line. If needed, select it again using "**Select>Residues>VGH**".

In the "**Actions>Atoms/Bonds/Atom Style**" menu, choose successively the "**Stick**", **Ball & Stick**" and "**Sphere**" menu items and see how the ligand is displayed in each case.



It is also possible to change the atom and bond display by clicking on the "Stick", "Sphere" and "Ball stick" buttons of the "Styles" section of the toolbar when the "Molecule Display" tab is selected.



Finally, choose the ball & stick representation.

Undo and Redo

In ChimeraX, it is possible to undo the last action (in case of error, for instance). It is also possible to restore the last undone action using the redo option.

	Mac	Windows
Undo	Command + z	Ctrl + z
Redo	Shift + Command + z	Shift + Ctrl + z

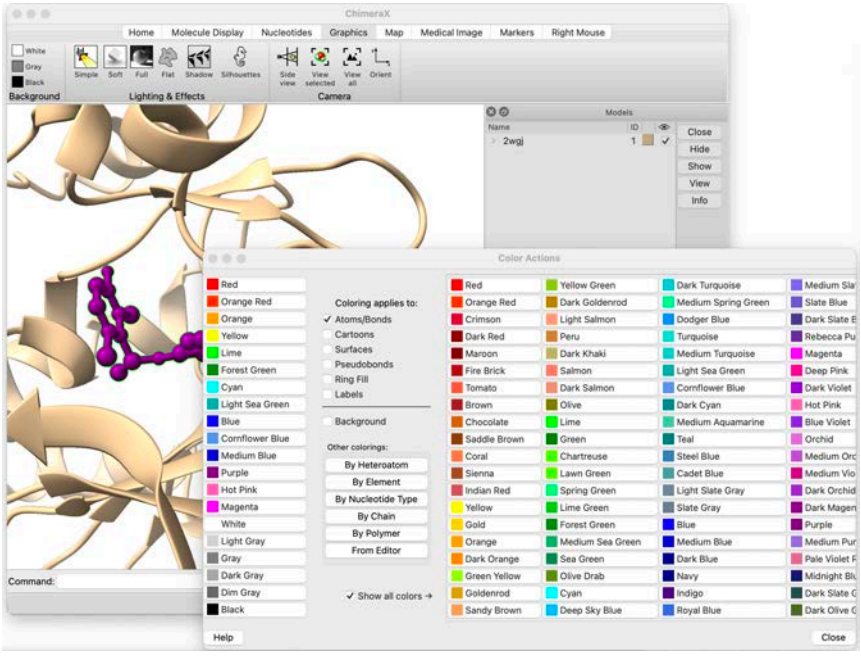
You can test these commands now. You will see that you can re-change the bond display of crizotinib in this example.

Display crizotinib again in ball and stick representation before continuing the tutorial.

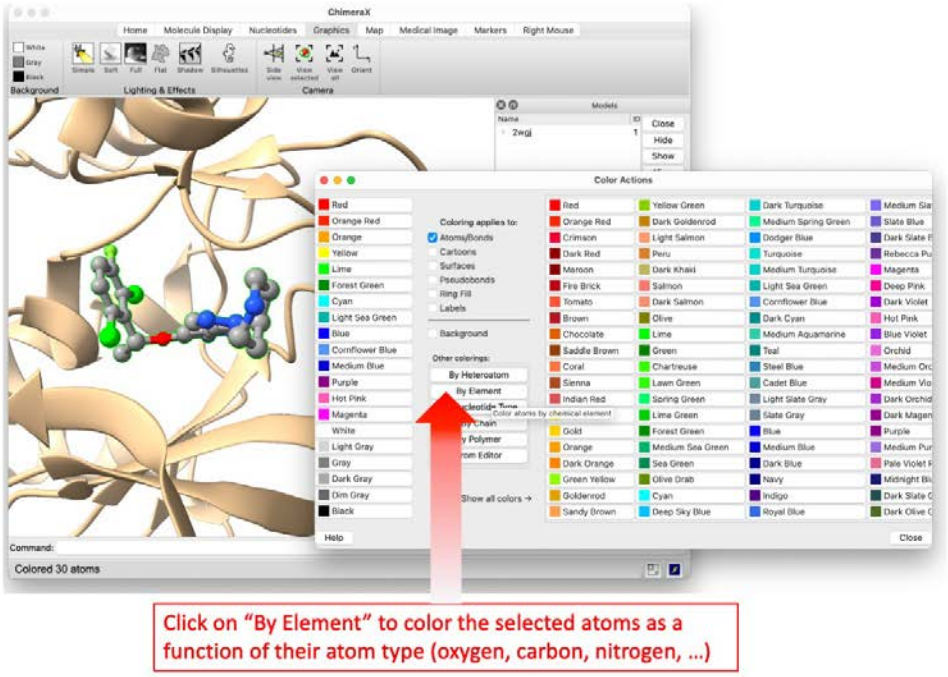
Note: As mentioned previously, altering the system, by deleting atoms, cannot be undone. It is therefore recommended to save the status in a session file before such an action.

Changing bond color

Open the detachable color action panel using "**Actions>Color>All Options…**". Check "Atoms/Bonds" 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.



Once you have tested different possibilities, color all atoms according to their atom types. This is obtained by clicking "By element".



You can close the "Color Actions" menu.

Selecting using the command line

Fine selections, which are difficult to achieve using the simple "**Select**" menu, can be obtained via the command line interface, using the **select** command.
This is achieved by typing select *spec* in the command line interface, where *spec* specifies the items to be selected. The following symbols can be used to create the specifiers:

Symbol	Reference level	Definition	Examples
#	Model	model number assigned to the data in ChimeraX (hierarchical, with positive integers separated by dots: N, N.N, N.N.N, *etc.*)	#1 *or* #2.1
/	Chain	chain identifier (case-insensitive unless both upper- and lowercase chain IDs are present)	/A *or* /B
:	Residue	residue number *OR* residue name (case-insensitive)	:1230 or :glu
@	Atom	atom name (case-insensitive)	@ca or @OH

Adapted from https://www.rbvi.ucsf.edu/chimerax/docs

Note: In the command line, capitalization of chain IDs, residue names, and atom names is not important. *The only exception is when a model contains both uppercase and lowercase chain identifiers; case matters for chain specification in that model only.*

It is possible to specify more than a single item, by providing lists and ranges:

- Multiple items can be entered as a comma-separated list. E.g. **:1230,1245** defines residues 1230 and 1245, while **:@ca,oh** defines atoms ca (Cα of every amino acid) and oh (oxygen atom of the side chain of the tyrosine residues).

- Ranges are entered in the form *start-end*. E.g. **:1230-1245** specifies all residues between 1230 and 1245 included, while **#1-3** specifies models 1, 2 and 3.

Precise specifiers can be obtained by combining the simple ones defined above. For instance:

- **:asp,leu** specifies all aspartate and leucine residues.
- **:asp,leu@cb** specifies all Cβ atoms of aspartate and leucine residues.
- **/A:1230,1245@n,ca,c,o** specifies all N, Cα, C and O atoms (i.e. the backbone) of residues 1230 and1245, belonging to chain A.
- **#1/A:1230@ca** specifies the Cα atom of residue 1230 or chain A in model 1.
- **#1,2:1230,1250-1255@ca** specifies all Cα atoms of residues 1230 and 1250 to 1255, belonging to models 1 and 2.

Users can select all atoms belonging to a first pre-established list *pre-spec* that are within a given distance *cutoff* from a specified reference *spec-ref* using **select zone** *ref-spec cutoff pre-spec* . It is possible to add the keywords *residues true* to include the entire residues within the *cutoff* distance instead of just some of their atoms. By default, the reference is not selected. For instance:

- **select zone /A 6** will selects all atoms with 6 Å from chain A, excluding chain A.
- **select zone :1230 10** selects all atoms within 10 Å of residue(s) 1230, excluding residue(s) 1230.
- **select zone :1230 10 :1220-1270** selects all atoms from residues 1220 to 1270 that are within 10 Å of residue(s) 1230, excluding residue(s) 1230.

It is possible to select items around the currently selected ones by using the keyword *sel* for the *ref-spec*. For instance:

- **select zone sel 5** selects all atoms within 5 Å of the elements that are selected at the moment this command is typed.

To select the reference too, it is possible to add the *extend true* keywords. For instance:

- **select zone :1230,1235 6 extend true** selects all atoms within 6 Å of residues 1230 and 1235, including residues 1230 and 1235 themselves.

Finally, it is possible to select the entire residues and not just atoms by adding the residues true keywords. For instance:

- **select zone :1230 10 residues true** selects all residues (instead of just some of their atoms) that are within 10 Å of residue(s) 1230.

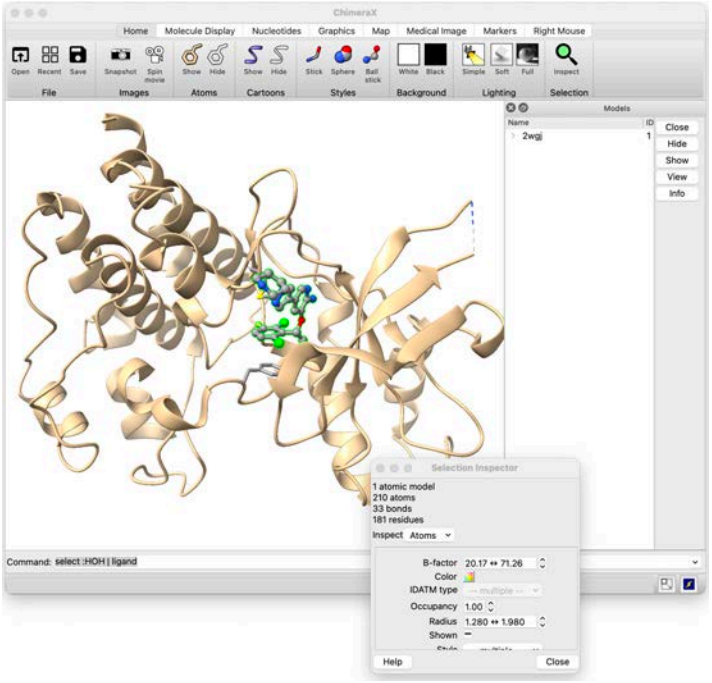
Explicit keywords can also be used for *spec-ref* and *pre-ref*, such as *protein, nucleic, sidechain, mainchain (*or *backbone), helix, strand, coil, ligand, ions, water.*

It is possible to use logical operators like

- **&** for intersection (AND)
- **|** for union (OR)
- **~** for negation (NOT)

Interestingly, a “Selection Inspector” panel can be used to verify the number of models, atoms, bonds and residues present in the current selection. This tool is useful to check the results of complex specifiers.

To open the “Selection Inspector” panel, click on the “Inspect” button in the “Selection” section of the toolbar (in the “Home” tab). Alternatively, you can use “**Tools>General>Selection Inspector**”.



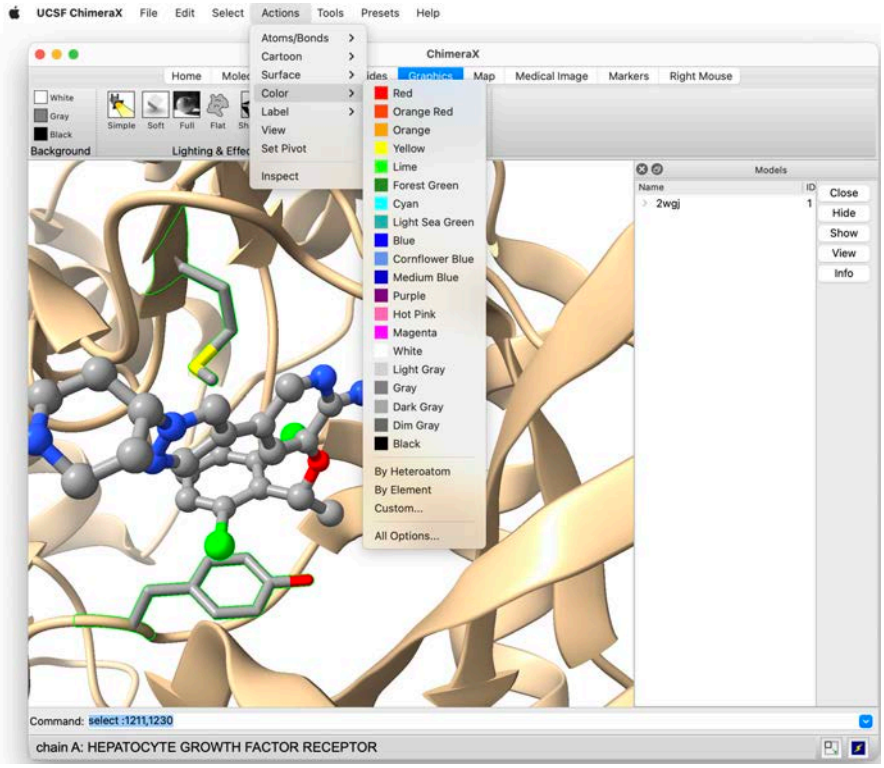
Please, type the following commands and verify the selection they provide:

Commands	Results
select	Selects everything
select /A	Selects chain A **Of note, this command will select all chains A in case multiple models are present**
select /A,B	Selects chains A and B
select ligand	Selects all ligands
select protein	Select all natural amino-acids, so all proteins
select :VGH	Selects all residues whose PDB identifier is VGH
select :HOH	Selects all residues whose PDB identifier is HOH (i.e. water molecules)
select :HOH | ligand	Select all residues whose PDB identifier is HOH as well as all ligands
select :1230	Selects residue 1230. **Of note, this command will select all residues numbered 1230 if there is more than one**
select :1230@ca	Selects Cα atoms of residue 1230
select /A:1230	Selects residue 1230 of chain A

select :1211,1230	Selects residues 1211 and 1230
select strand	Selects all β strands
select helix	Selects all helices
select :VGH @<5	Select all atoms within 5 Å of residues whose PDB identifier is VGH. This command is an alternative to the command **select zone :VHG 5**
select solvent	Selects all water molecules, as well as all other small molecules with a maximum of 10 atoms and present in at least 10 copies in this structure

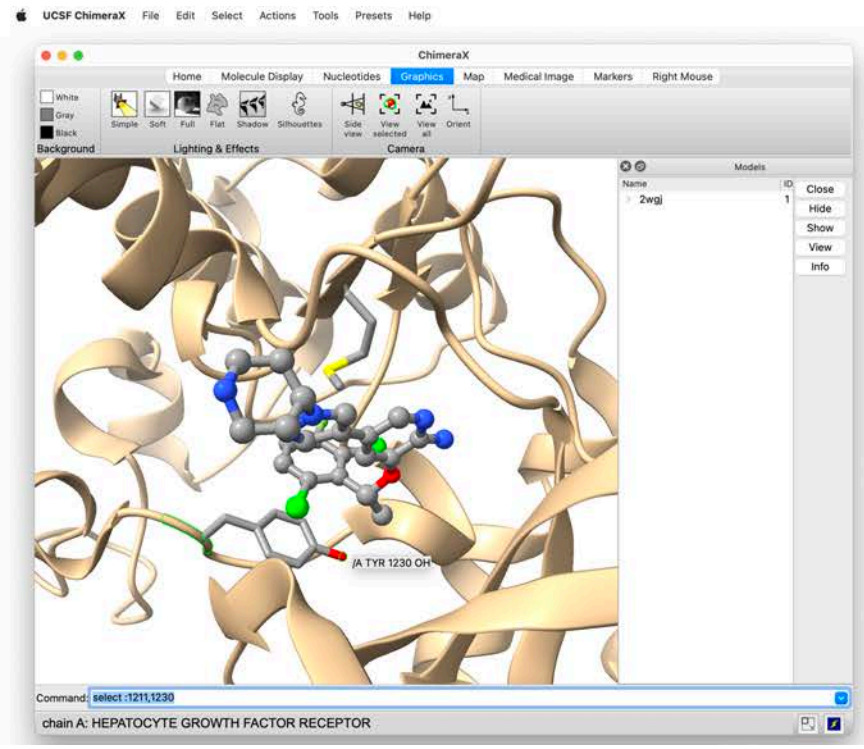
More information and examples are available on pages
https://www.rbvi.ucsf.edu/chimerax/docs/user/commands/select.html and
https://www.rbvi.ucsf.edu/chimerax/docs/user/commands/atomspec.html

To continue this tutorial, select residues 1211 and 1230 by typing select :1211,1230 in the command line interface, and show them in stick representation, colored according to the atom type: "**Actions>Atoms/Bonds>Show**", "**Actions>Atoms/Bonds>Atom Style>Stick**", "**Actions>Color>by element**".



Selecting using the mouse and the keyboard

To select one atom that is currently displayed on the screen, it is possible to press the **Ctrl** key on the keyboard while performing a **left-click** with the mouse on the atom of interest. Use this technique to select the oxygen atom of the Tyr1230 side chain of MET.



This oxygen atom (named "/A TYR 1230 OH") is now surrounded by a thin green line.
Now that this atom has been selected, **put the mouse cursor over the 3D panel**, and perform the following commands to obtain different selection schemes:

- Pressing the "**arrow up**" key on the keyboard will select the entire residue Tyr1230
- Pressing again "**arrow up**" key will select the secondary structure element to which Tyr1230 belongs (a loop in this case)
- Pressing again "**arrow up**" key will select the entire protein chain to which Tyr1230 belongs (in this case chain A, excluding ligand)
- Pressing again "**arrow up**" key, if necessary several times, will select the entire system (including proteins, ligands, etc…)
- Pressing the "**arrow down**" key will change the selection down to the protein chain to which Tyr1230 belongs
- Pressing "**arrow down**" key again will change the selection down to the secondary structure element to which Tyr1230 belongs
- Pressing the "**arrow down**" key again will select down residue Tyr1230
- Finally, pressing the "**arrow down**" key again will select down the atom OH of Tyr1230

Note: It is possible to select several atoms using the keyboard and mouse. Select the first atom using the "**Control**" key and the **left mouse click** as described above, then press **simultaneously the "Control" and "Shift" keys** of the keyboard and perform a **left click with the mouse** on the second atom. You can select as many atoms as you wish with this procedure.

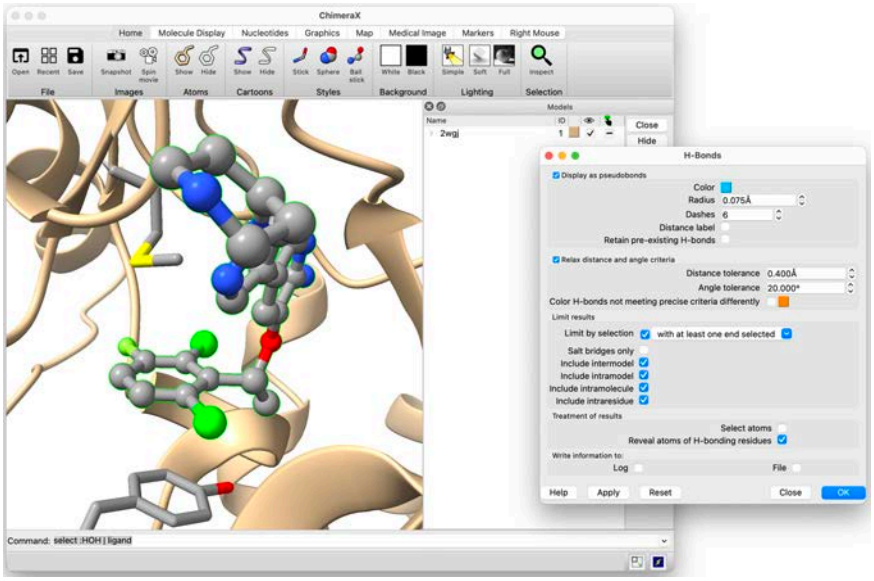
It is also possible to clear the selection by using the **Ctrl** key and clicking with the left mouse button in a void space (in the background).

Note: Pressing the **Ctrl** key on the keyboard while holding the **left button of the mouse and dragging it** on the ChimeraX window will select all atoms within the rectangle defined by the dragging of the pointer. Note that the selection is done in 3D.

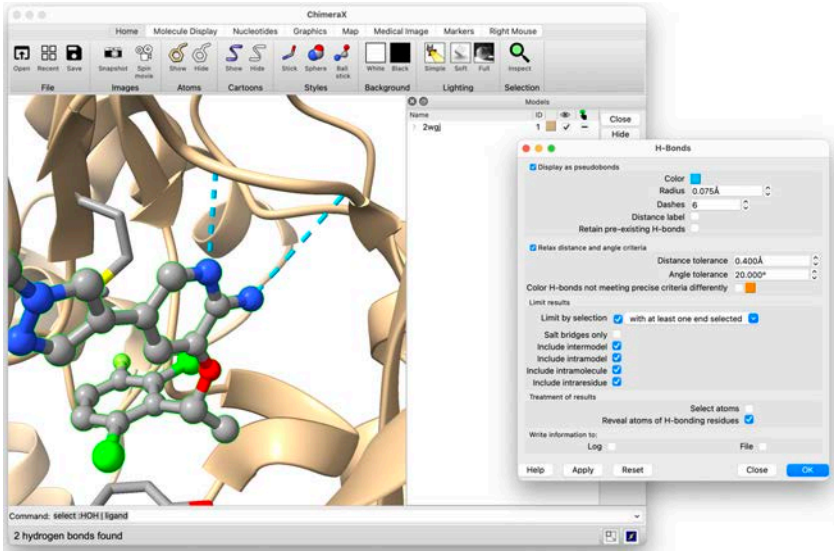
Calculating and showing hydrogen bonds

ChimeraX can calculate and display hydrogen bonds between given atoms.

To show the hydrogen bond network between the ligand and the MET protein, the following sequence of actions can be performed. Select crizotinib using the command line, the selection menu or the keyboard/mouse technique. Open the "**Tools>Surface/Structure Analysis>H-Bonds**" panel. Check the "Limit by selection with at least one end selected" box, so that it will calculate only hydrogen bonds involving crizotinib. Also check the box "Reveal atoms of H-bonding residues" to display all residue side chains that are making a hydrogen bond with crizotinib, even if they are currently hidden (this is very useful not to miss any hydrogen bond).



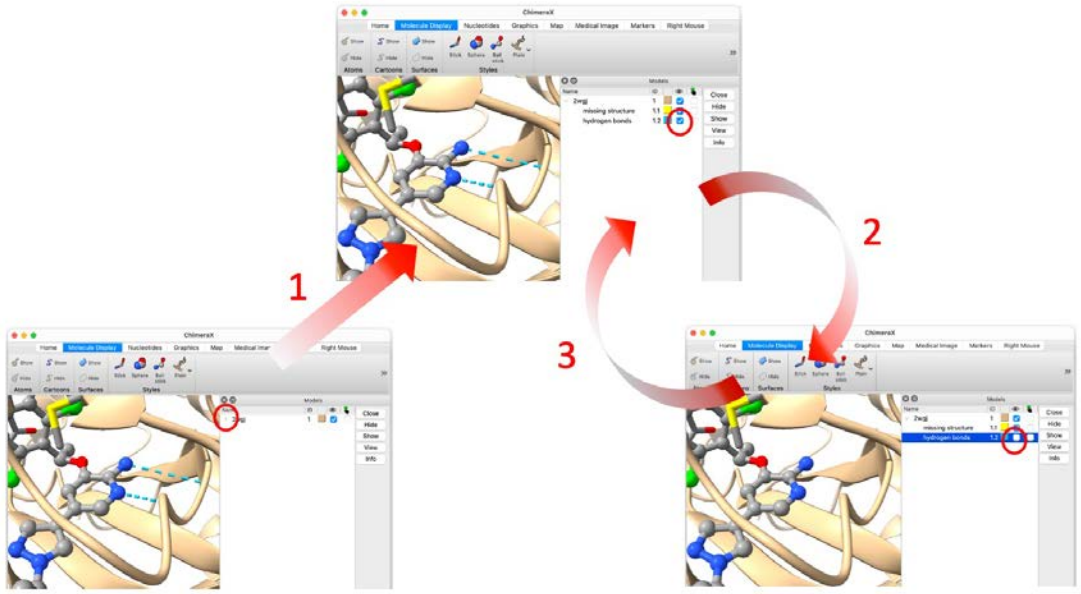
Finally, click on the "Apply" button. Hydrogen bonds will be displayed by thick blue dashed lines.



Note: You can change the hydrogen bond color, radius and dashes using the options available in the top of the "H-Bonds" panel, from the "**H-Bond Parameters**" window and click "Apply" to apply them.

Hiding and displaying hydrogen bonds with the "Models" panel

Clicking on the right arrow ">" in front of the name of a model in the "Model" panel displays a list of items related to that model. This list includes previously calculated hydrogen bonds:



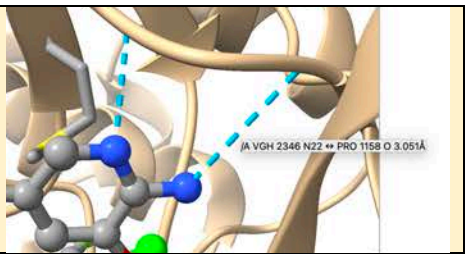
You can uncheck the “Shown” box symbolized by an *eye* in the “hydrogen bonds” line, to hide the hydrogen bonds (2 in the image above) or check it to display them again (3 in the image above).

Finally, display the hydrogen bonds before continuing the tutorial.

Hiding and displaying ribbon representation for selected residues

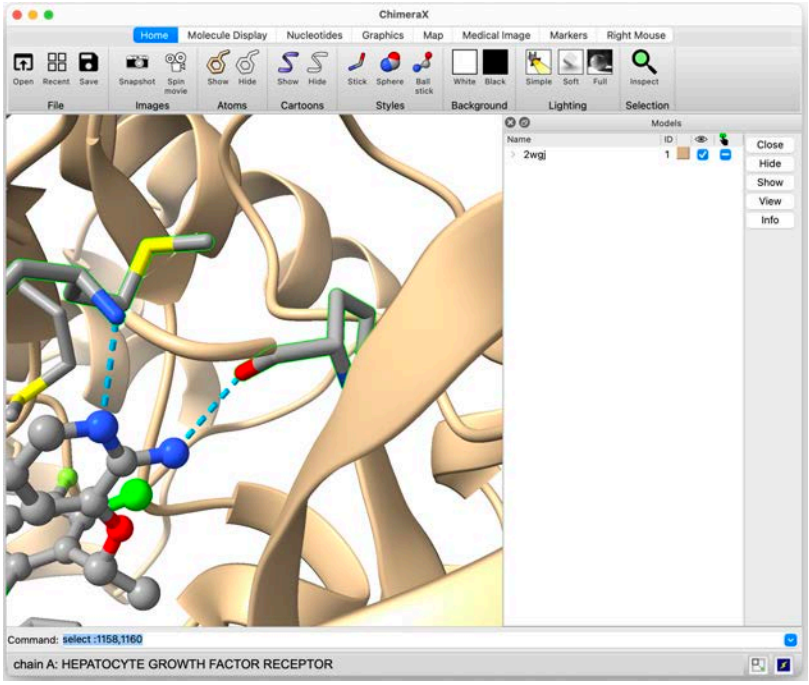
ChimeraX has found hydrogen bonds between crizotinib and Pro1158 of MET as well as between crizotinib and Met1160 of MET.

Note: If you leave the pointer of the mouse over the dashed line of a hydrogen bond, information regarding the atoms involved will be displayed:



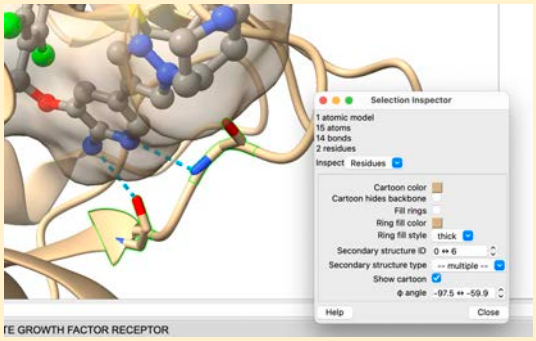
You can notice that the side chains of these residues have not been displayed automatically when calculating hydrogen bonds, even though the “Reveal atoms of H-bonding residues” option was selected in the “H-Bonds” panel. This implies that crizotinib makes hydrogen bonds with the backbone atoms of these residues, and not with their side chains. However, since the cartoon representation is currently applied to all protein residues, the backbone atoms are hidden. To better see these hydrogen bonds, it is therefore necessary to switch off the cartoon representation for these two residues and display their atoms.

For showing backbone atoms, select these two residues. This can be done using the mouse as seen above, or by typing `select :1158,1160` in the command line. Once the residues are selected, hide their cartoon representation by selecting “**Actions>Cartoon>Hide**”, or by clicking the corresponding button in the toolbar. Keep these two residues selected and display their atoms using “**Actions>Atoms/Bonds>Show**” or by clicking the corresponding button in the toolbar. Finally, color them according to their elements. Now you can see the exact position of the MET backbone atoms that make hydrogen bonds with crizotinib.



To show again the ribbon representation for these two residues, select them (if not already done) and click on "**Actions>Cartoon>Show**".

Note: Alternatively, it is possible to display the backbone atoms on top of the Cartoon representation by unchecking the "Cartoon hides backbone" box in the "Selection Inspector" panel, after selecting "Residues" for the "Inspect" option (See page 21 to open the "Selection Inspector" panel).

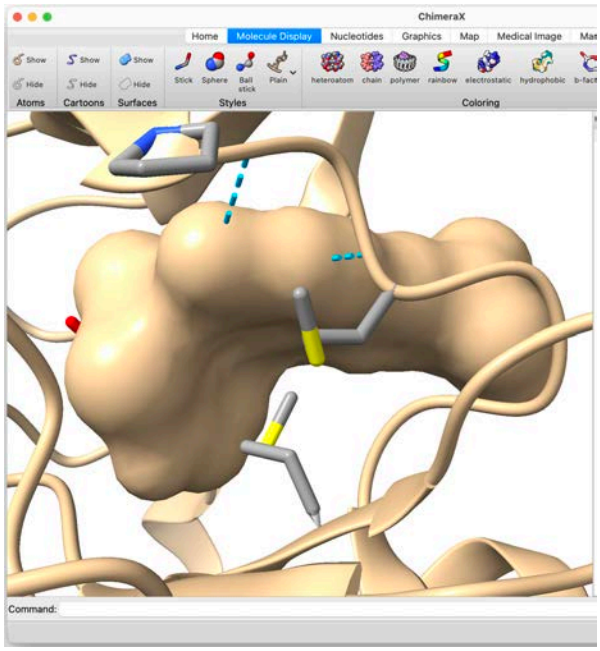


Calculating and displaying the ligand surface

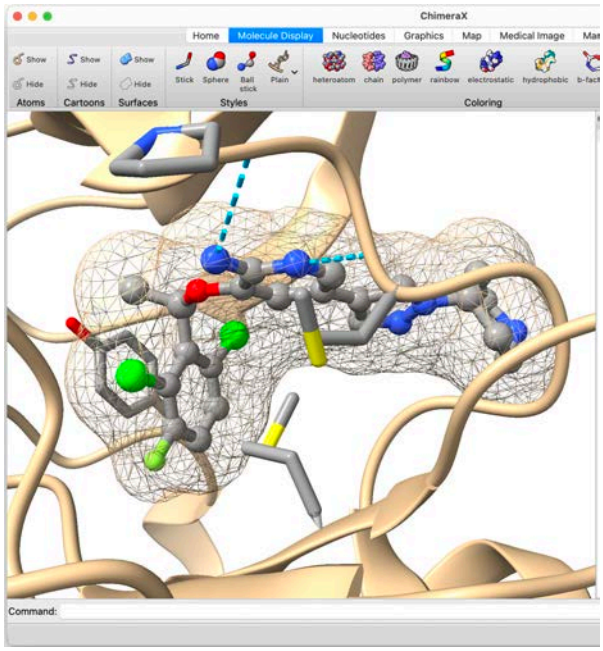
Select crizotinib, and then click "**Actions>Surface>Show**".

Note: It is possible to change the color of the surface. Open the "**Actions>Color>All Options…**" menu and check the "surfaces" button so that the color changing will be applied only to surfaces. Surfaces can be colored using a selected color, or by elements.

Through the "**Actions>Surface**" menu, you can try to modify the surface representation to "**Mesh**" and "**Dot**". Then, go back to the "**Solid**" representation.

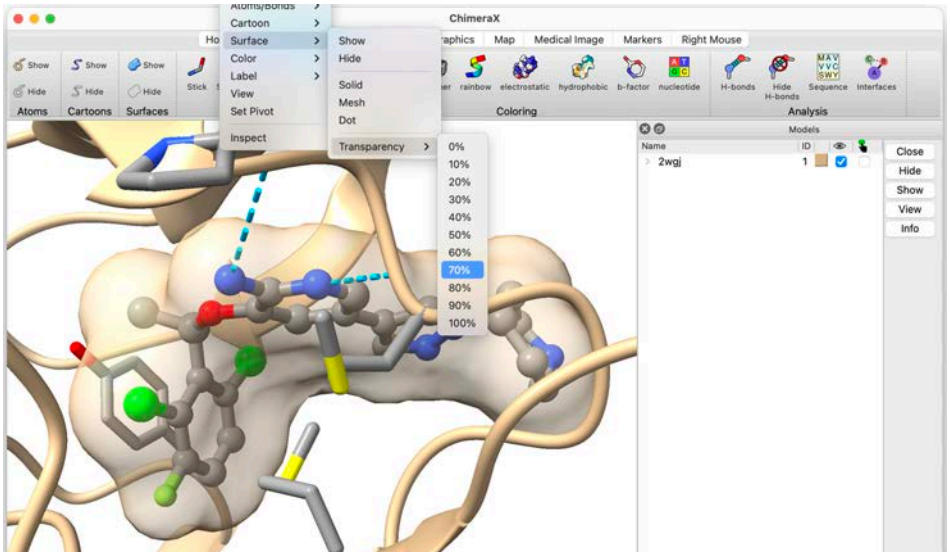


Surface in solid representation



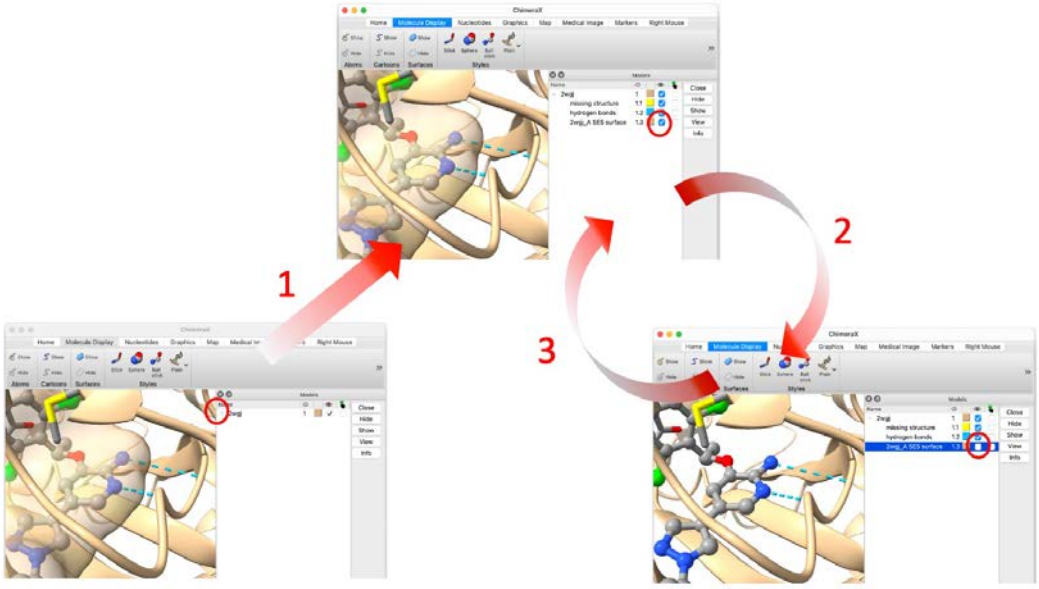
Surface in mesh representation

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



Hiding, displaying and closing a surface using the "Models" panel

As mentioned previously, clicking on the right arrow ">" in front of the name of a model in the "Model" panel displays a list of items related to that model. This list includes previously calculated surfaces:



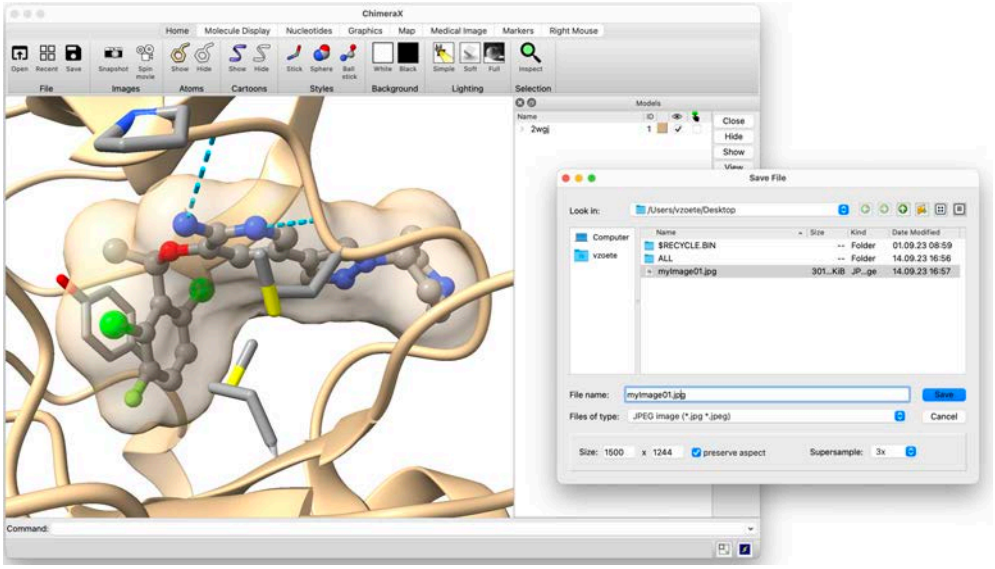
You can uncheck the “Shown” box symbolized by an *eye* in the “surface” line, to hide the surface of the ligand (2 in the image above) or check it to display them again (3 in the image above).

Finally, display the surface before continuing the tutorial.

Note: It is also possible to delete the surface from the system by selecting the corresponding line in the “Models” panel (it will be highlighted in blue) and click on “Close” on the right-hand list. Note that this permanently removes the surface from the system. To display it again, you'll need to recalculate it.

Saving images

Clear all selections using “**Select>Clear**”. Choose an orientation and a zoom that provides a satisfying point of view. Then, select the “**File>Save…**” menu item, or click on the “Save” button of the toolbar. In the new panel, select the format of the file that will be saved in the “Files of Type” option. Select JPEG in this case. Select “preserve aspect” and enter an image “Size” of 1500 pixels for the width. The height will be adjusted automatically. Finally, choose a “File name” and click “Save”.

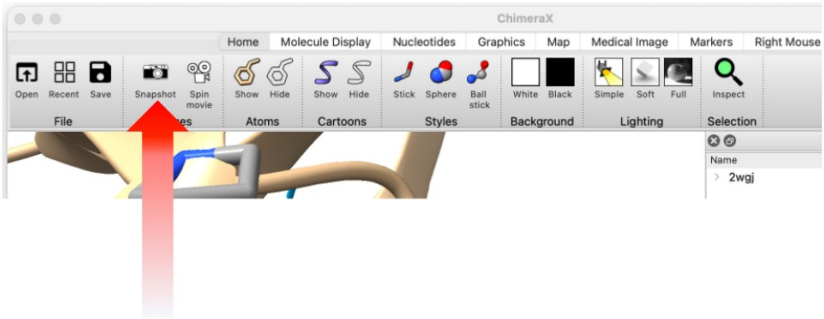


Images are saved as RGB figures.

Note: Generally, scientific journals require JPEG, EPS or TIFF images, with a 300 dots per inch resolution or higher, and with RGB colors.

Note: TIFF and PNG formats allow to save an image with a transparent background.

To rapidly generate an image, it is possible to click on the "Snapshot" button of the toolbar ("Home" Tab).



Saving the session status

Save the last state of this exercise as a ChimeraX session (as seen previously). This session will be used in exercise 2.

Closing the session. Quitting ChimeraX

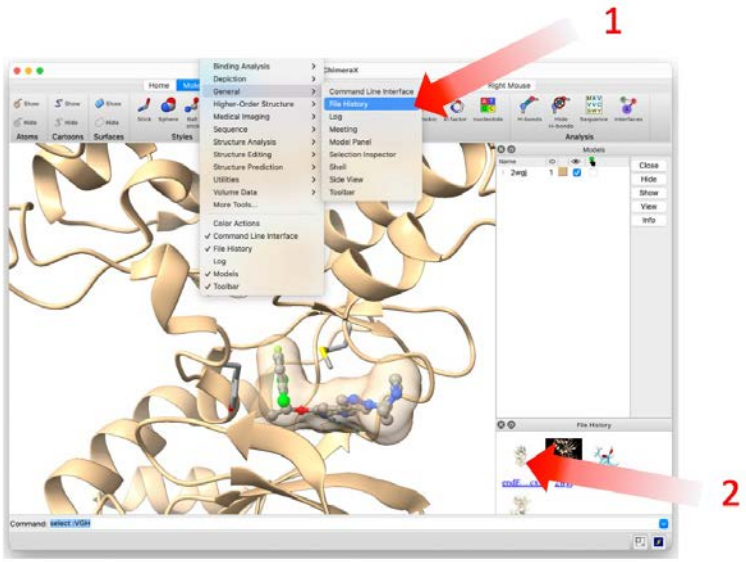
The session can be closed using the "**File>Close Session**" menu item.

Exercise 2 – Advanced surface options

Restoring the last session of exercise 1

Restore the session saved at the end of exercise 1 by clicking on its corresponding image in the list of previously opened files (see page 4).

Of note, this list can also be accessed as a “File History” panel, if you click on “**Tools>General>File History**”:



Alternatively, it is possible to use the “**File>Open…**” menu item.

Hiding and deleting a molecular surface using the command line interface

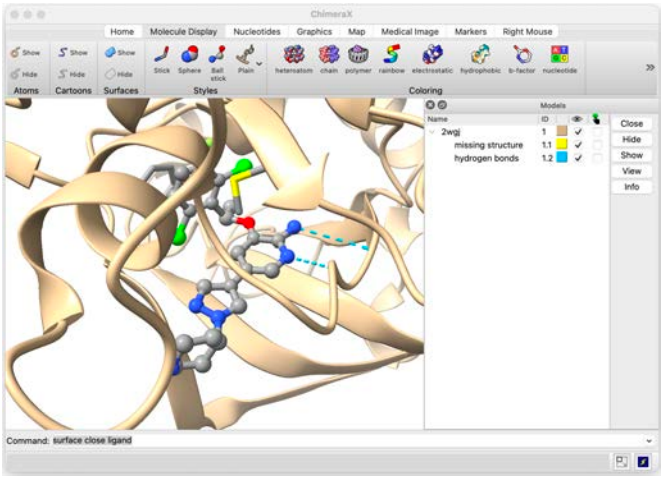
It is possible to hide an existing surface using the command surface hide *spec*, where *spec* specifies the surface to be hidden.

To hide the surface of the ligand, type:

```
surface hide ligand
```

Of note, the surface of the ligand still exists in the system. To permanently delete it, type

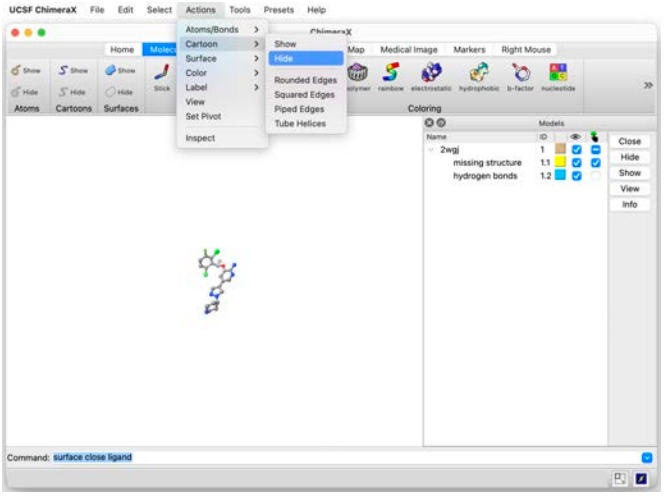
```
surface close ligand
```



This will also update the “Models” panel, by deleting the surface entry.

Showing the molecular surface of the protein

Hide the bonds, atoms and cartoon of MET and keep only crizotinib visible. To do so, select the protein, for example by choosing “**Select>Residues>Standard Amino Acids**”. Then choose “**Actions>Atoms/Bonds>Hide**” and “**Actions>Cartoon>Hide**”. Alternatively, it is possible to achieve the same results by clicking on the “Hide” buttons of the “Atoms” and “Cartoons” sections of the toolbar.

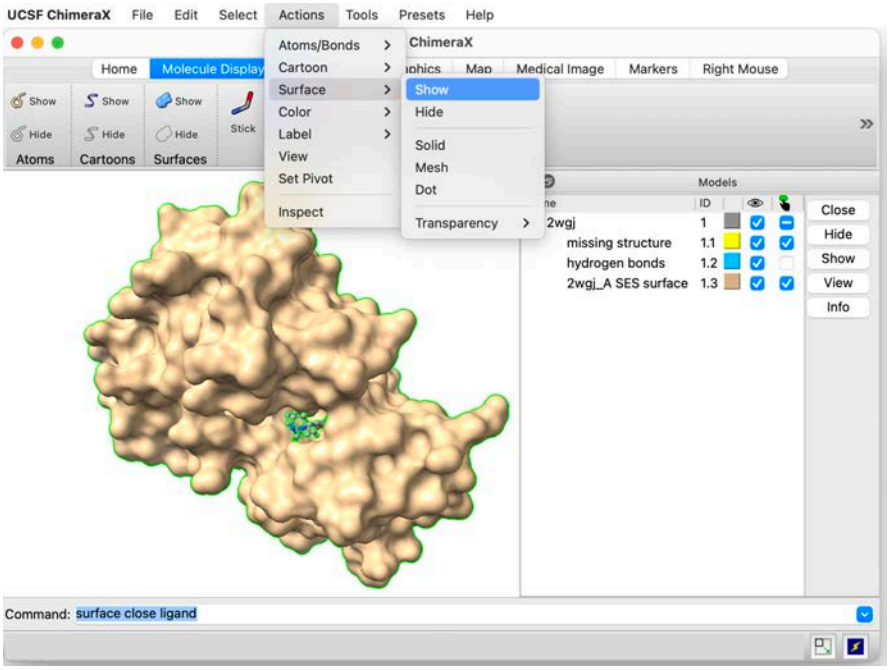


Be sure that the protein is still selected. If necessary, or in case of doubt, select it again as described above. Then, choose “**Action>Surface>Show**”. This can also be achieved by selecting the protein and clicking on the “Show” button of the “Surfaces” section of the toolbar.

Another option to calculate and display a surface is to use the command surface *spec*, where *spec* specifies the atoms for which the surface should be calculated. In our case, we can type:

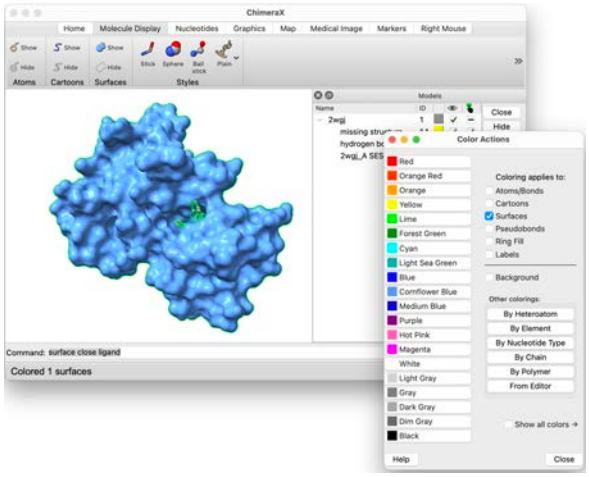
```
surface protein
```

Note that the protein surface now appears in the “Models” panel.

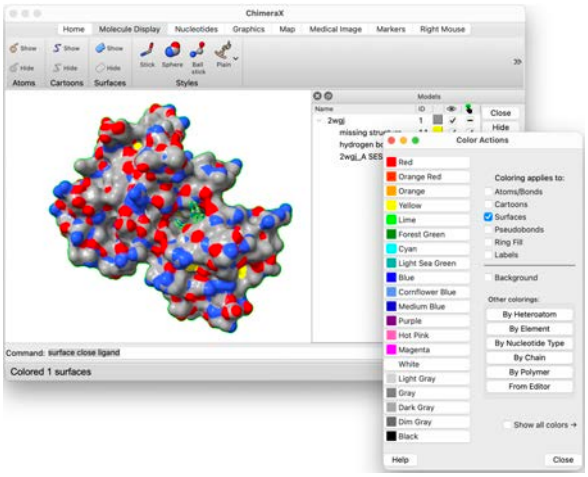


Changing the surface color

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 wish to apply to the entire protein surface.
It is also possible to color the surface according to the type of the underlining atoms by clicking on "By Elements".



Single color applied to the entire surface



Color surface by atom type

Finally, apply the cornflower blue color to the entire surface and close the "Color Actions" panel.

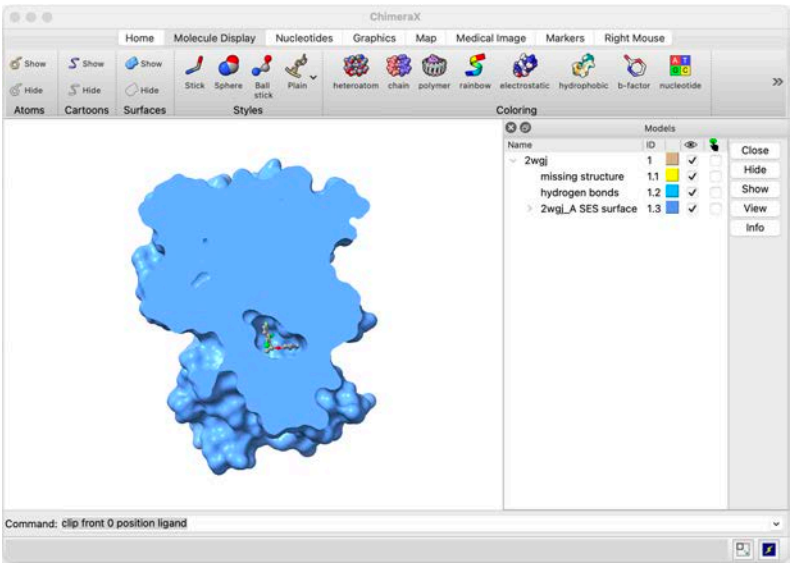
Note: The molecular surface can also be colored according to other properties like the electrostatic potential, the hydrophobicity, the B-factors, etc. We will address these in separate exercises.

Clipping the protein surface & Surface capping

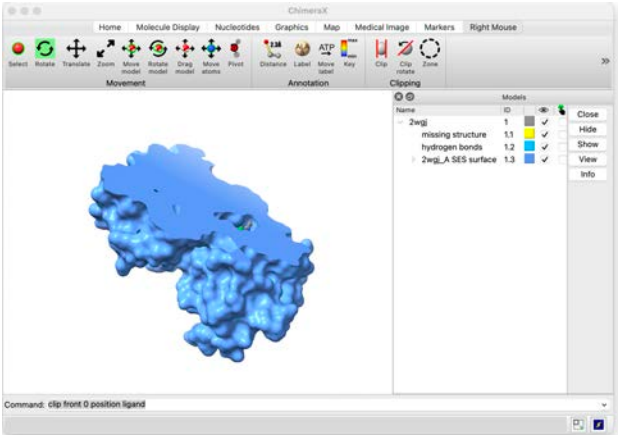
It is possible to cut the protein surface to get a better view of the ligand binding mode inside the binding site. This is called clipping. To activate it, type the following command:

```
clip front 0 position ligand
```

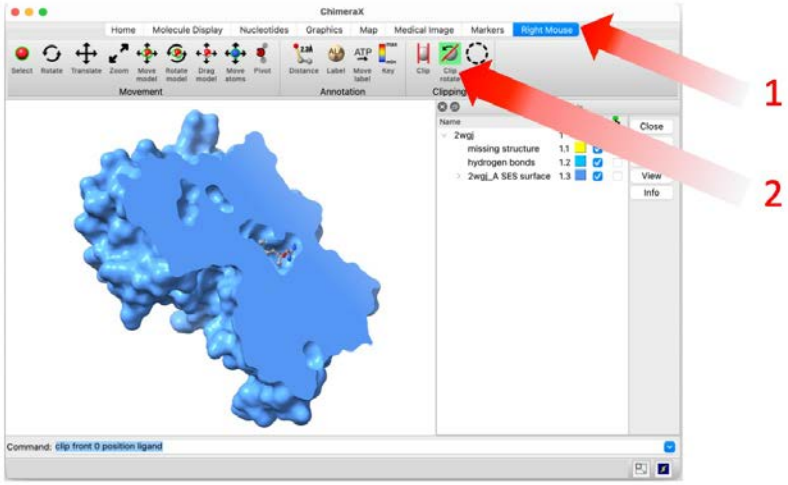
This command will cut the system with a plane perpendicular to the line of sight (so, parallel to the computer screen, at first), positioned on the center of the ligand, and will hide everything that is in front of it. You should see something like this:



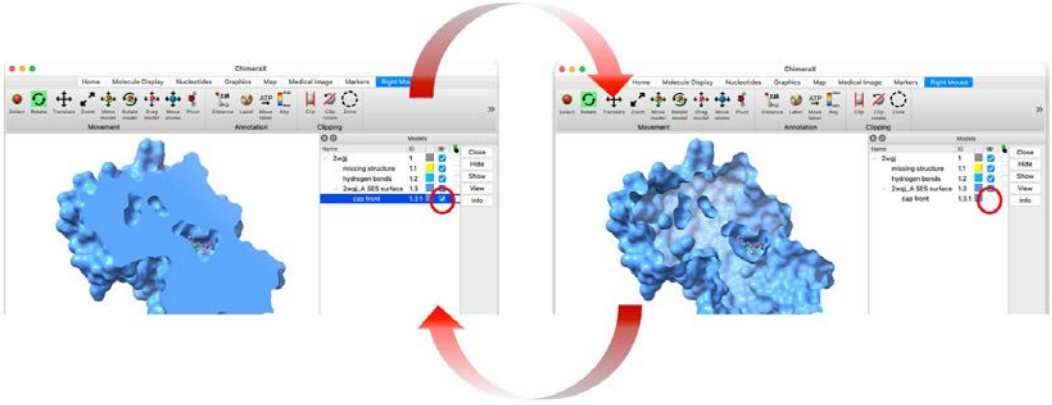
It is possible to translate, rotate and zoom on the protein to better appreciate the position of the clipping plane and its effect on the display.



You can change the position and orientation of the clipping plane with the right button of the mouse. To do so, select the “Right Mouse” tab on the toolbar.
By default, the right button of the mouse controls the rotation of the system. To change this, click on the “Clip” button (the selected option is highlighted in green). Now, the right button will allow translating the position of the clipping plane.
If you click on the “Clip rotate” button, you will control the orientation of the clipping plane by rotating it.
If you are satisfied with the position of the clipping plane, you can click on the “Rotate” button of the toolbar to go back to the default behavior of the mouse right button.



Note that now the protein surface is preceded by a “>” sign in the “Models” panel. Clicking on it reveals a new entry called “cap front”. Checking or unchecking the corresponding “Shown” box will remove the capping of the clipping plane. This allows you to see the ‘inside’ of the protein.

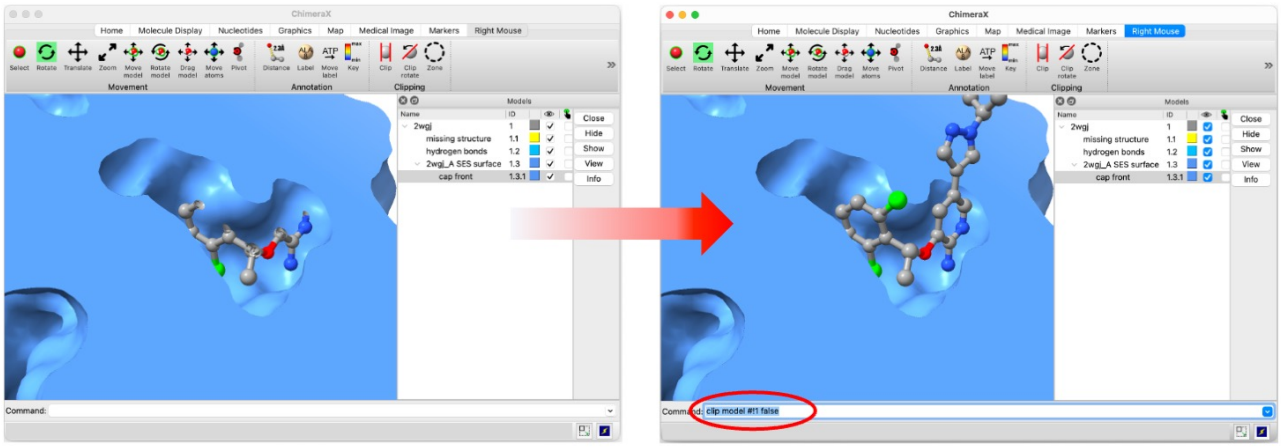


Display the capping of the clipping plane before continuing.

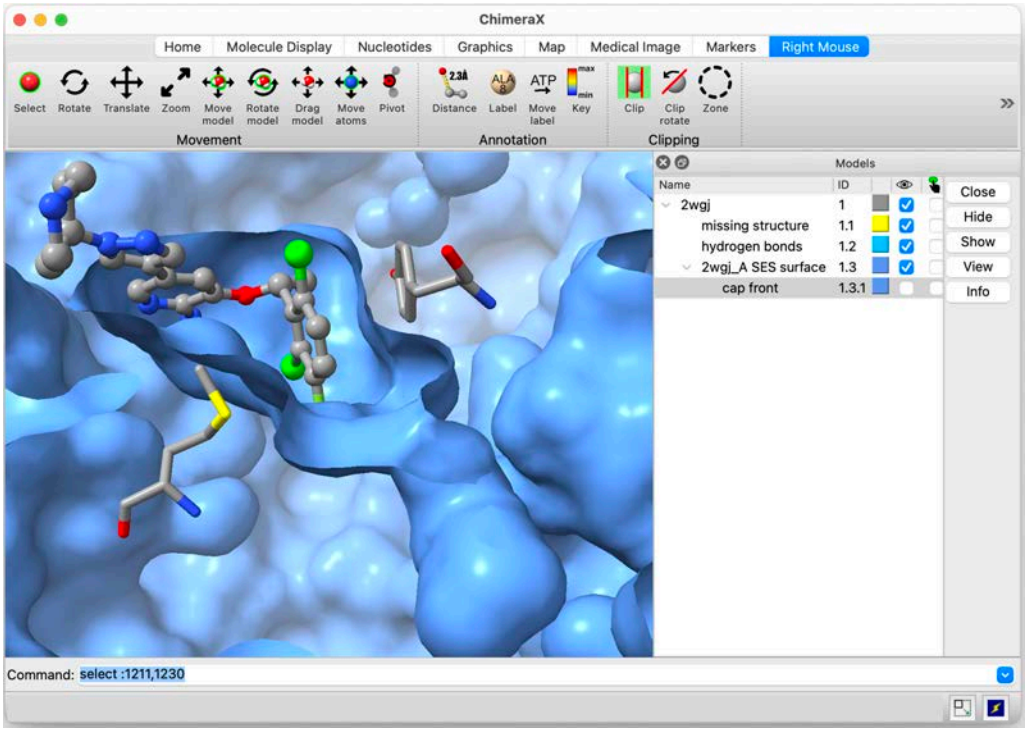
By default, the previous command clipped the entire system, i.e. not only the protein surface but also all the atoms, bonds and cartoons. For a better representation of the system, it can be convenient to clip only the protein surface, while displaying the rest. This can be achieved by typing the following command:

```
clip model #!1 false
```

The “clip model” and “false” keywords indicate that clipping will not affect a given selection. The latter is specified by “#!1”, which means ‘all models except #1’. Since all atoms are included in #1, while the surface is in #1.3, this command will continue applying the clipping plane to the protein surface, but not to the ligand or protein atoms.



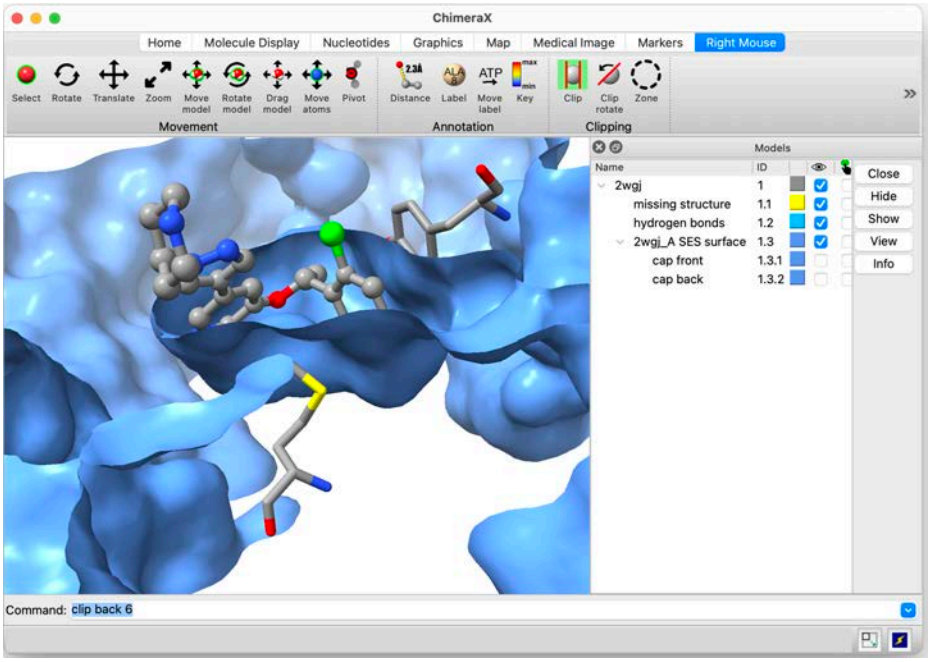
Select and display residues 1211 and 1230 by typing “**select :1211,1230**” in the command line (followed by **Return**), and then display them in the stick representation. Change the position and orientation of the clipping plane to have a good view of the interactions between these residues and the ligand. Remove the capping of the clipping plane to obtain a view similar to this:



You can also get a good view of the shape of a buried binding site using the slab mode. To activate the latter, type the following command:

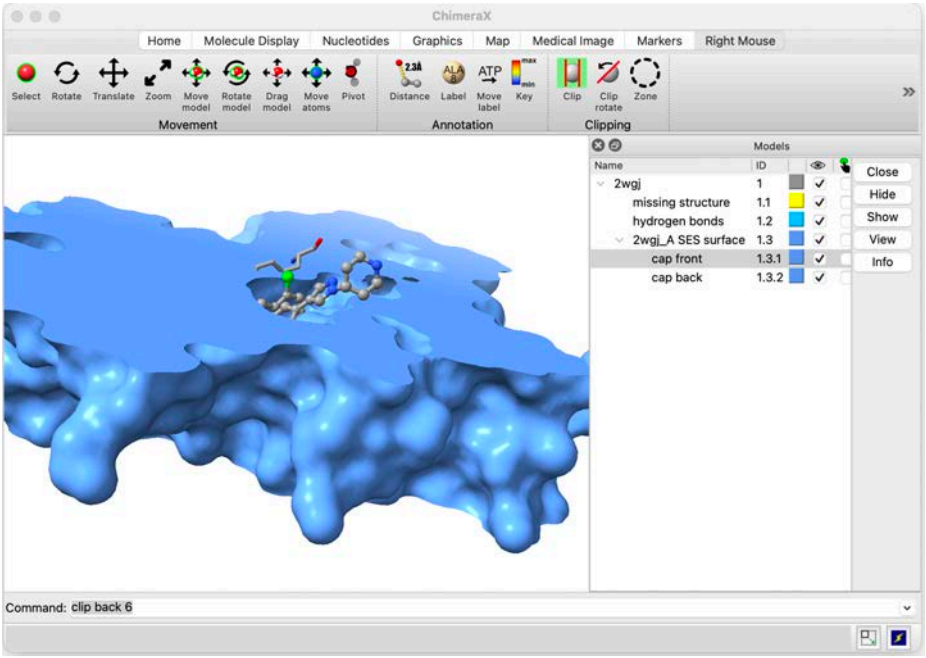
```
clip back 6
```

This will create a second clipping plane, parallel to the first one. All the surface present behind this so-called back clipping plane will not be displayed. Consequently, only the section of the surface present between the first (front) and the second (back) clipping planes will be displayed. As for the front clipping plane, this back clipping plane will appear in the “Models” panel, allowing the user to switch on or off its capping.

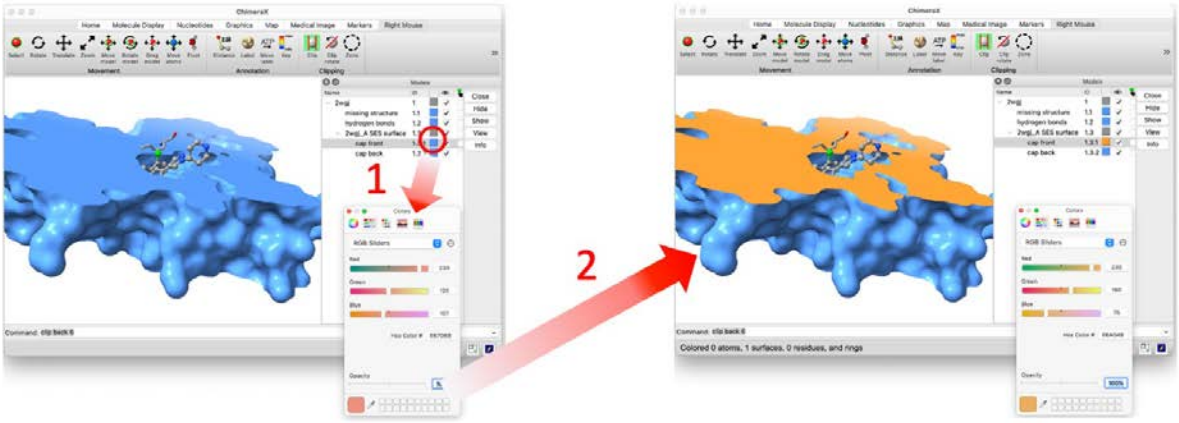


Changing the aspect of the surface capping

Display again the surface capping of both the front and back clipping planes by checking the corresponding “Shown” boxes in the “Models” panel.



Clicking on the blue square related to the font capping plane in the “Models” panels opens a “Color” panel that can be used to change the color of the front capping plane only.

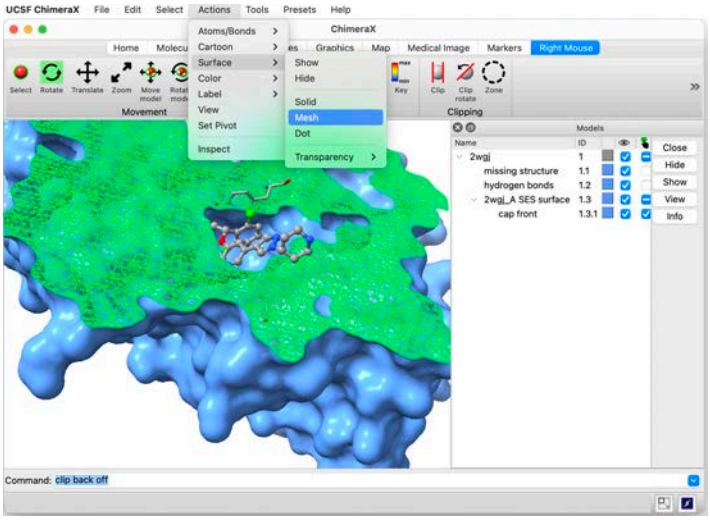


It is possible to do the same for the back clipping plane.

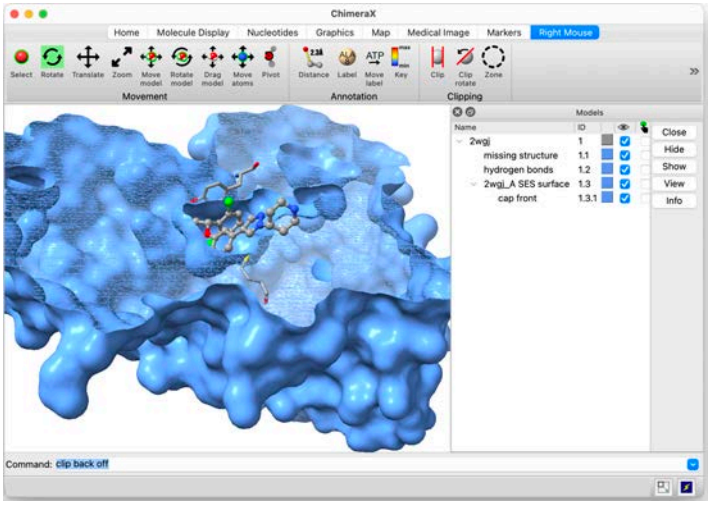
Turn off the back clipping plane by typing the command:

```
clip back off
```

In the “Models” panel, select the front capping plane by checking the corresponding “Checking” box. Then, apply “**Actions>Surface>Mesh**” to replace the solid capping surface by a mesh.



Clear the selection for a better representation.



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

Exercise 3 – Lighting, cueing and shadow effects

Opening a previous session

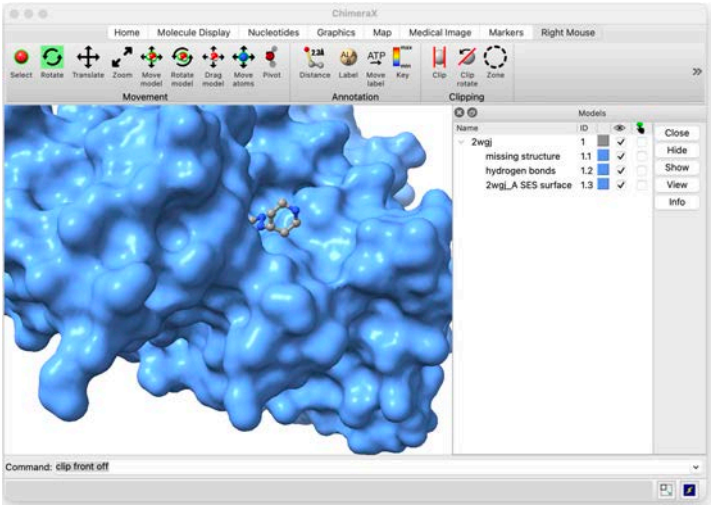
Restore the session saved at the end of exercise 2 using the "**File>Open…**" menu item

Showing the molecular surface of the protein

Type the command

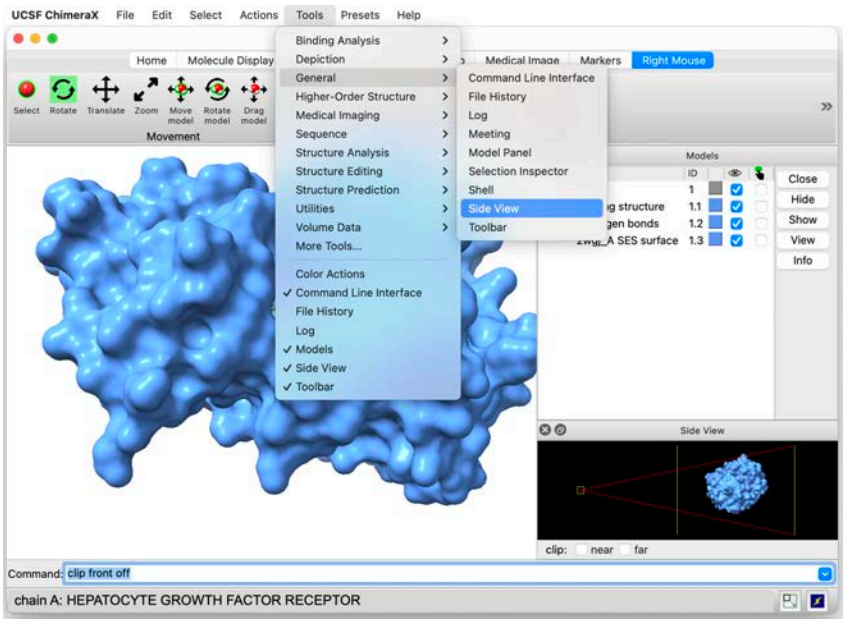
```
clip front off
```

to display the surface of the protein this way:



Using the side view

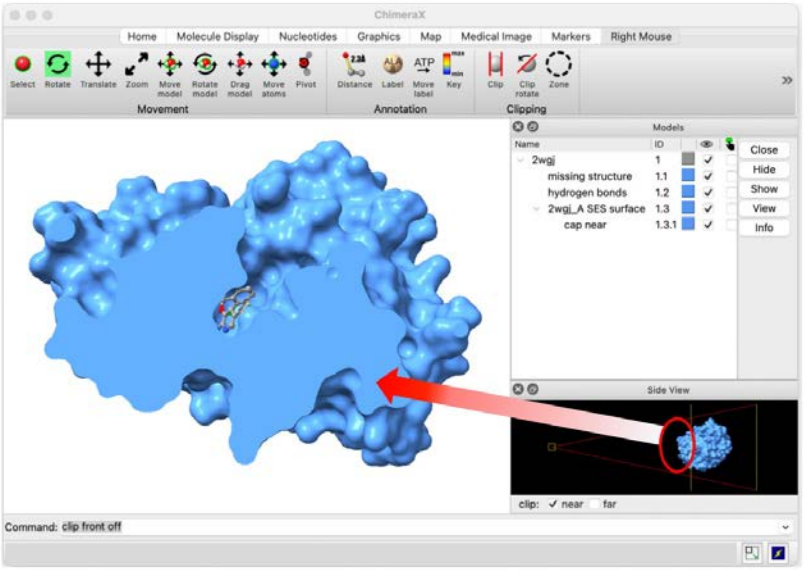
Open the "**Side View**" panel by choosing the "**Tools>General>Side View**" menu item, or by clicking on the "Side view" button in the toolbar, after selecting the "Graphics" tab.



You will see a reduced view of the structure appearing in the "**Side View**" panel. The vertical lines show the "near" and "far" 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.

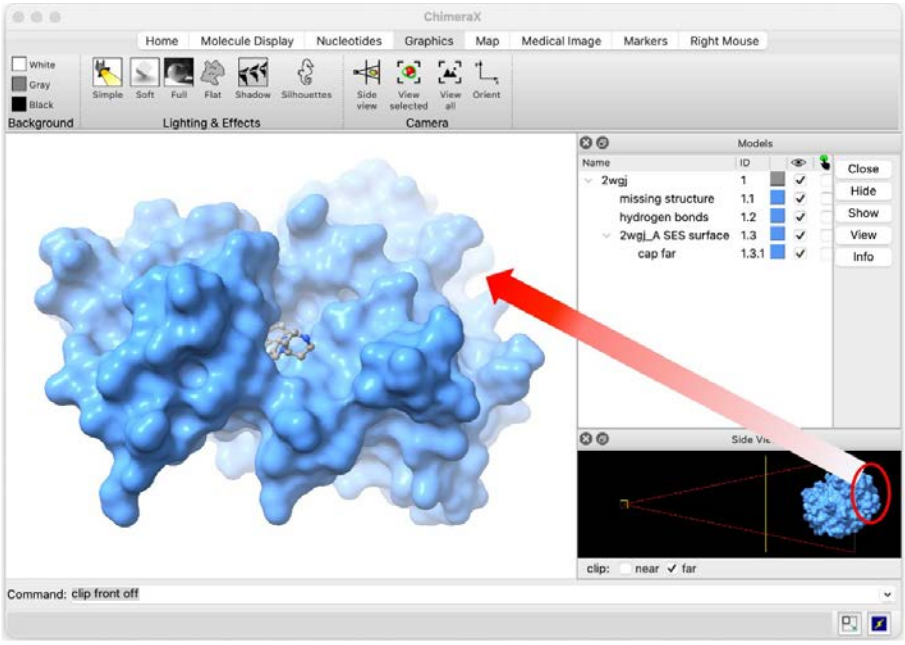
Note: Contrarily to the “front” and “back” clipping planes that we have seen above, the “near” and “far” clipping planes are always perpendicular to the line of sight and do not move with the scene. They can only be moved via the “Side View” panel.

Clicking on the “near” clipping plane (corresponding to the **left** vertical yellow line in the “Side View” panel) and dragging it changes its position. If you drag it to the right, all items (atoms, surfaces, etc.) situated between the eye position and the “near” clipping plane will be clipped and made invisible.



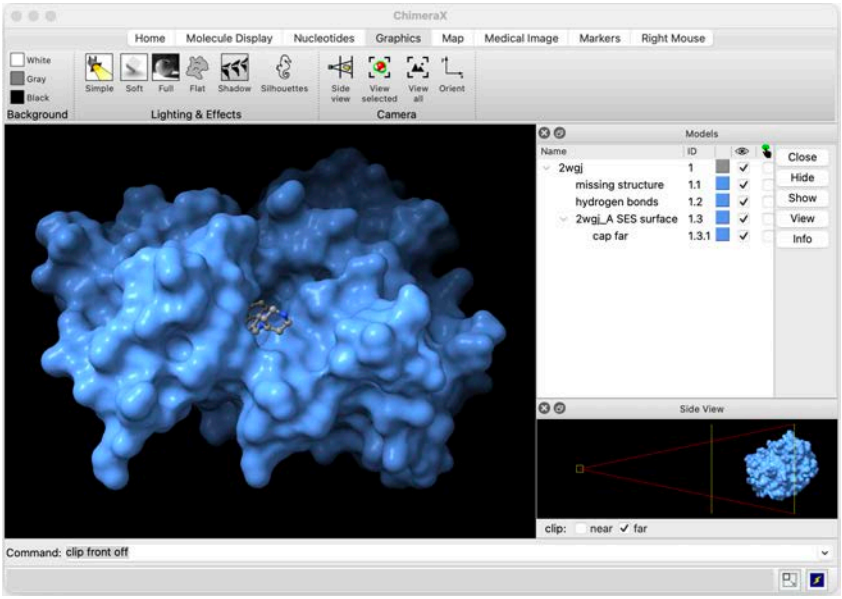
Uncheck the “near” box of the “Side View” to go back to the default situation and see the entire system.

Now, click on the “far” clipping plane (corresponding to the **right** vertical yellow line in the “Side View” panel) and dragging it changes its position. If you drag it to the left, all items (atoms, surfaces, etc.) situated after the “far” clipping plane will be clipped and will disappear *in the fog*.



Note: This option, known as **depth-cueing or front-to-back shading,** is very useful to provide a **3D effect** to a 2D image.

Changing the background color to black will make the depth-cueing hide remote part of the system *into the night* rather than *in the fog*. Try it using the "**Action>Color>All Options…**" menu.



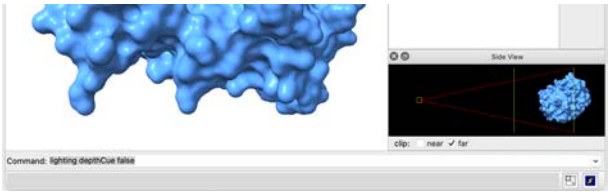
Then, color again the background in white before continuing the exercise.

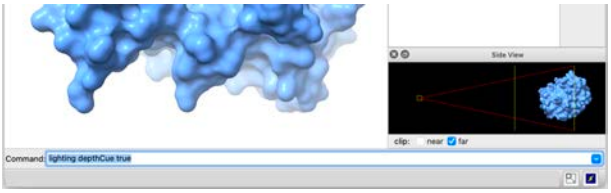
The depth-cueing can be switched off using the command

```
lighting depthCue false
```

and switched on again using

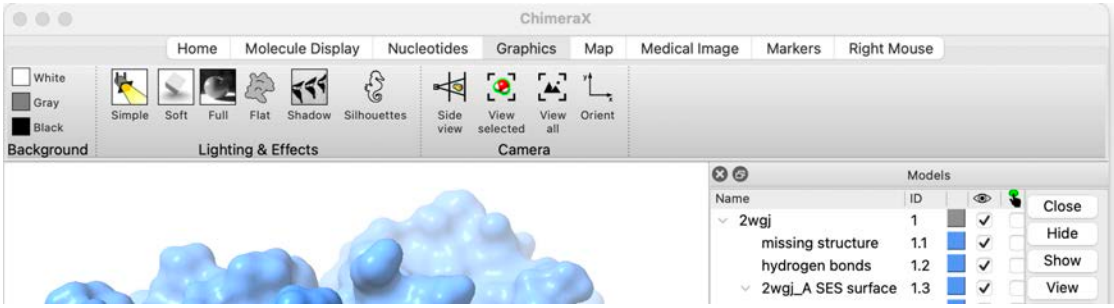
```
lighting depthCue true
```





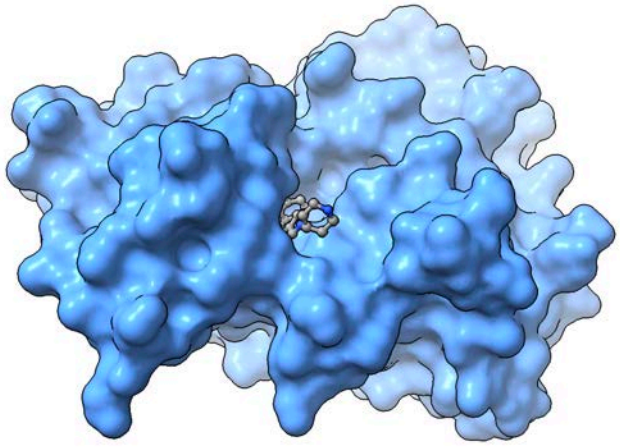
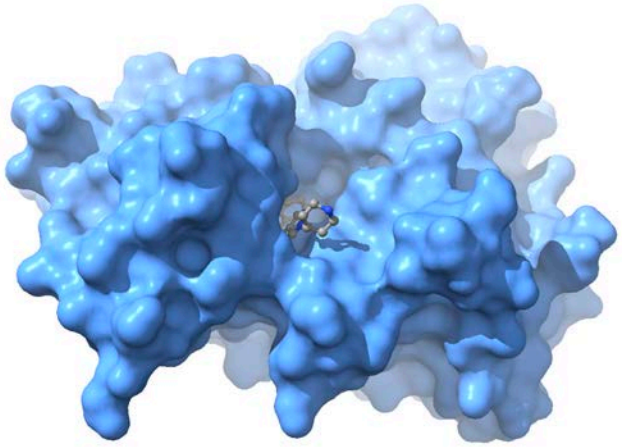
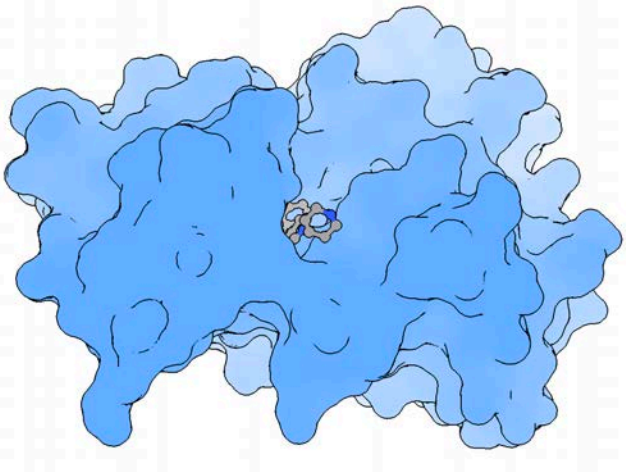
Changing the lighting and effects

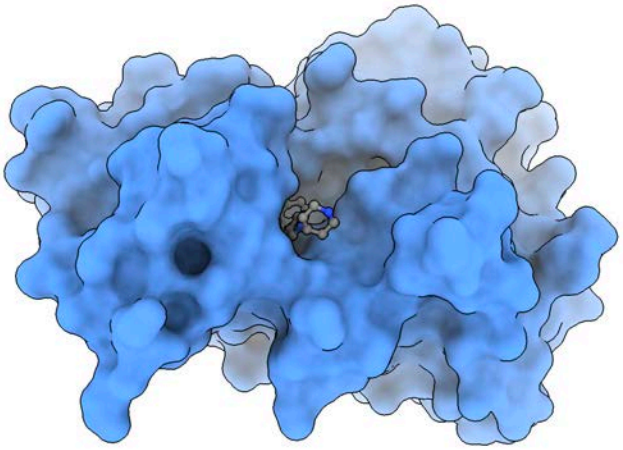
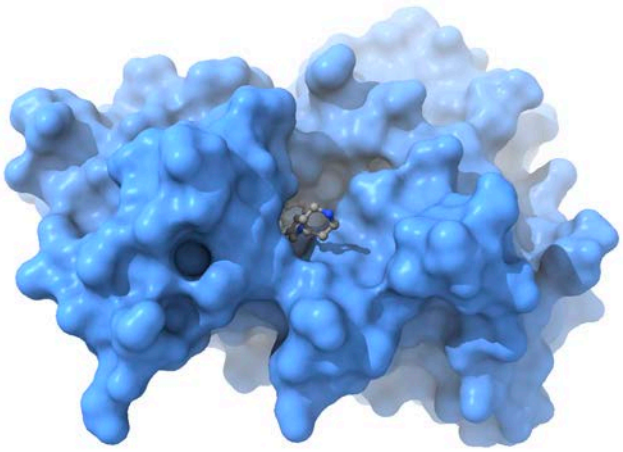
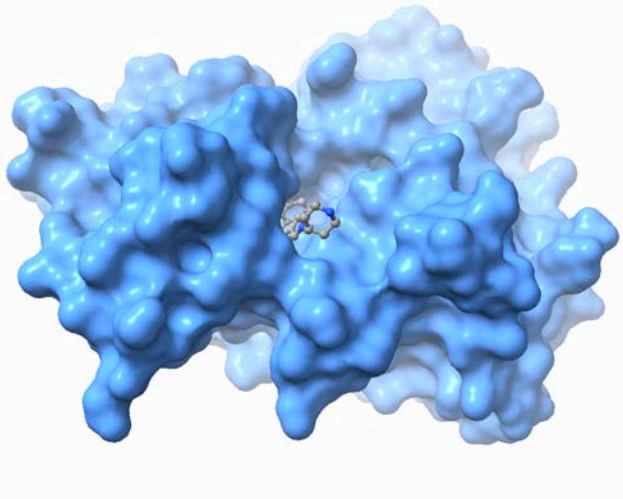
Several options can be used to change the lighting and the effects applied on the display of the system. They are accessible in the "Lighting & Effects" section of the "Graphics" tab.



There are three sources of light in ChimeraX:

- the ***key light*** is often the dominant source and can cast the main shadows
- the ***fill light*** serves as a secondary directional source to improve the visibility of regions that would otherwise be dark. It can cast secondary shadows
- the ***ambient light*** approximates omnidirectional illumination. Shadows cast from multiple uniformly distributed directions can produce ambient shadowing

Lighting & effect	Example
Silhouettes Outlines that emphasize borders and discontinuities	
Shadows Whether the key light should cast shadows	
Flat Ambient lighting only, without shadows but with silhouettes	

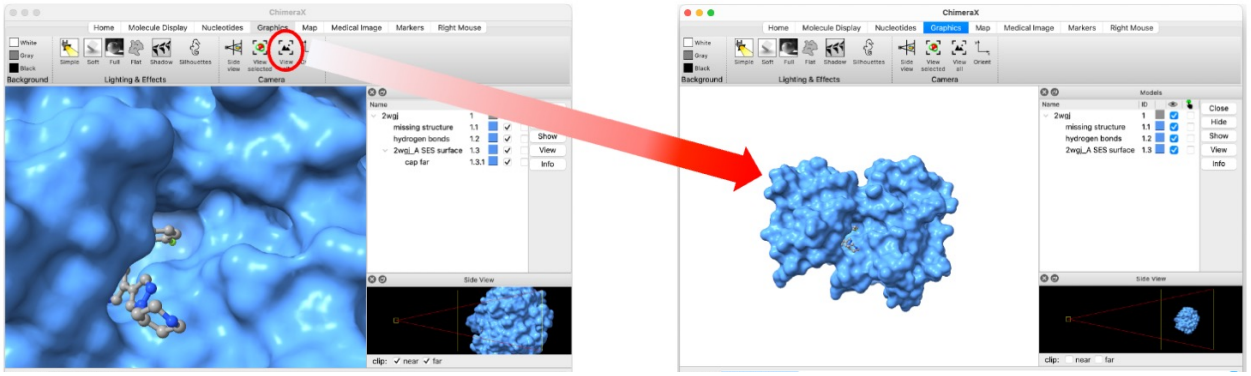
Soft Ambient lighting only, with ambient shadowing and silhouettes	
Full All kind of lights (key and ambient), with key shadows and ambient shadowing	
Simple (default) All kind of lights (key and ambient), without shadows	

A detailed description of all lighting options can be found here:
https://www.cgl.ucsf.edu/chimerax/docs/user/commands/lighting.html

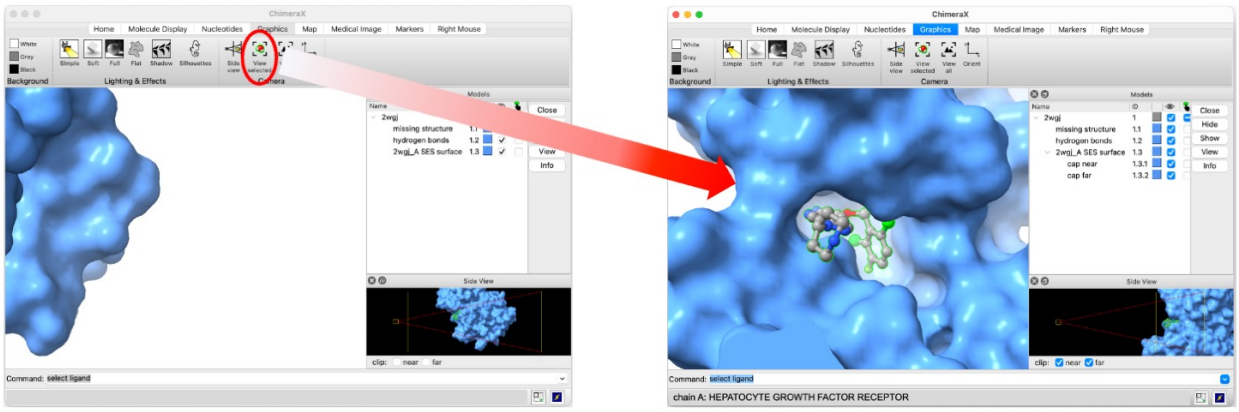
Other Graphics options

Additional options from the “Graphics” tab can be useful.

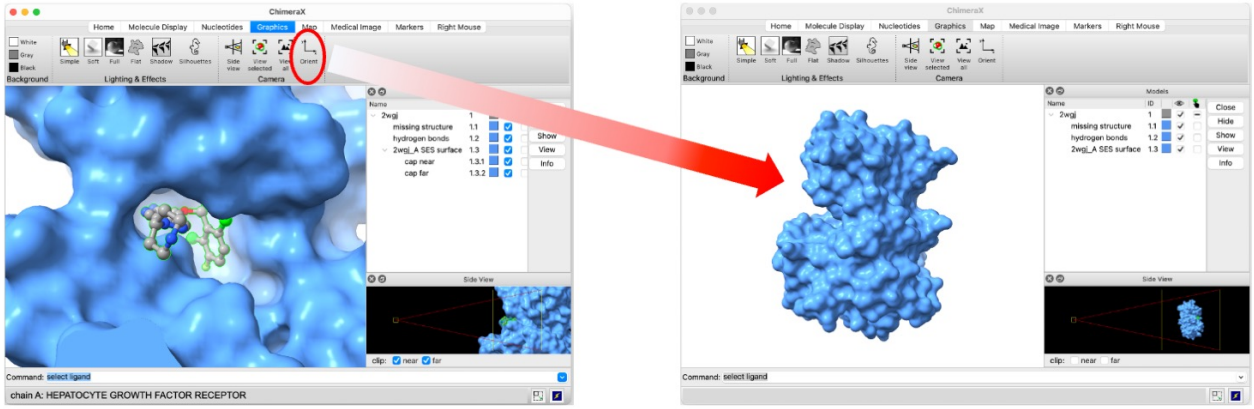
The “View all” button switch off the near and far clipping planes, and zoom in or out so that the entire system is visible and centered on the screen:



The “View selected” button centers the view on the selected items, zoom on it, and switch on the near and far clipping planes to focus the attention on this selection. The following was obtained by selecting the ligand and clicking on the “View selected” button:



The “orient” button rotates the scene to a standard orientation with X-axis horizontal increasing rightward, Y-axis vertical increasing upward, and Z-axis perpendicular to the screen increasing toward the viewer. In other words, it aligns scene coordinates with screen coordinates.



A detailed description of the view commands can be found here:
https://www.cgl.ucsf.edu/chimerax/docs/user/commands/view.html

Exercise 4 - Comparing structures

Loading macromolecular structures

The objective of this exercise is to learn how to compare two structures. For this, we will compare the kinase domain of MET, which we used during the previous exercises, and the experimental structure of the kinase domain of B-Raf, another therapeutic target in oncology.

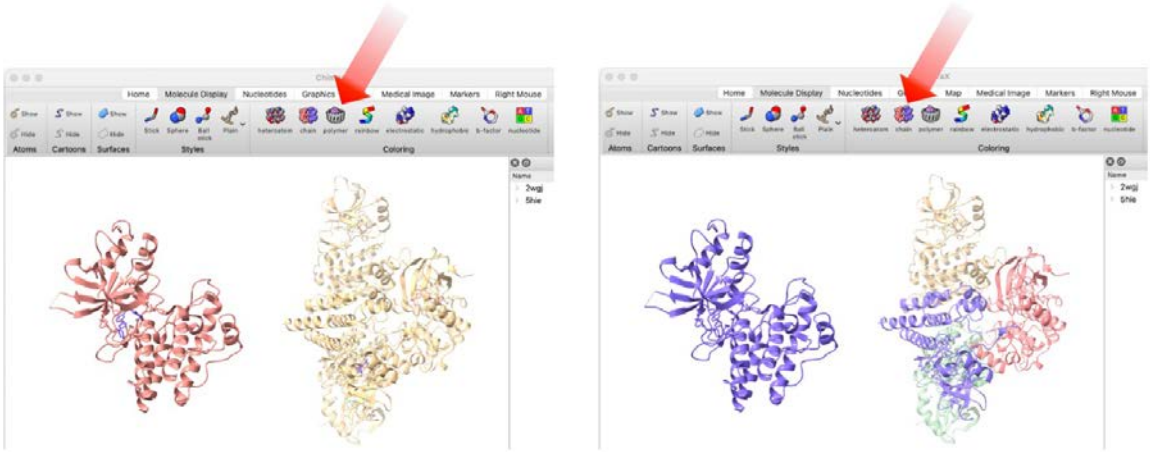
Fetch or load successively the PDB files:

- 2wgj, which contains the MET kinase domain binding crizotinib that we already used
- 5hie, which contains the B-Raf experimental structure, in complex with the FDA-approved drug dabrafenib

There are now two active models shown in the “Models” panel. Model #1 is MET (2wgj), whereas model #2 is B-Raf (5hie). Zoom out with the mouse to see both proteins in the 3D panel. You can also click on the “View all” button of the toolbar after selecting the “Graphics” tab.

Click on the “polymer” button of the “Molecule Display" tab of the toolbar. This applies a different color to each unique biopolymer (macromolecular entity). Non-biopolymer residues are not affected. This coloring is useful for easily differentiating several biopolymers, especially if they are superimposed on the screen.

Then, click on the “chain” button of the “Molecule Display” tab of the toolbar. This applies a different color to each biopolymer chain. Again, non-biopolymer residues are not affected. This coloring is useful for easily differentiating several chains in the same biopolymer, e.g. the monomers present in an experimental 3D structure of a protein complex. A similar coloring can be achieved by typing rainbow chains in the command line interface.

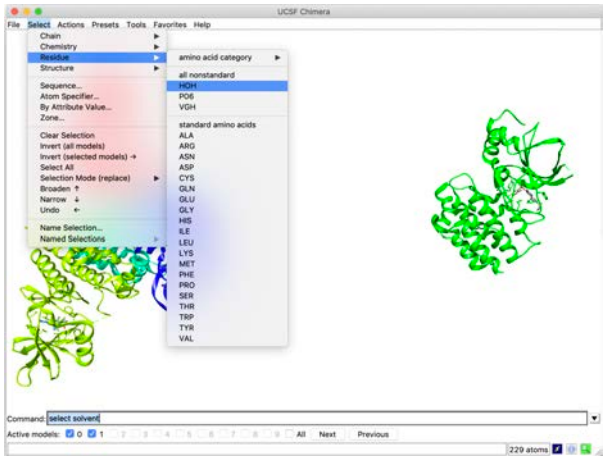


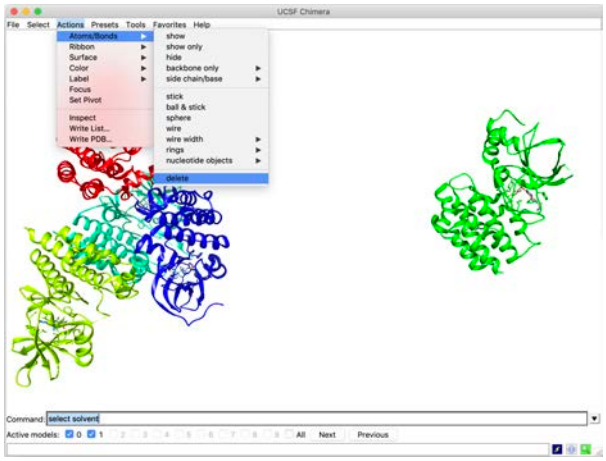
Note that the “Models” panel lists the two systems: 2wgj and 5hie.

Delete atoms

To simplify the analysis, we will delete the atoms that are not of interest.

For this, you can select all water molecules, by typing select solvent in the command line or by selecting “**Select>Residues>HOH**”. Then, choose “**Actions>Atoms/Bonds>Delete**”. Now, the system contains only natural amino-acids (i.e. the proteins) and some copies of the ligands.



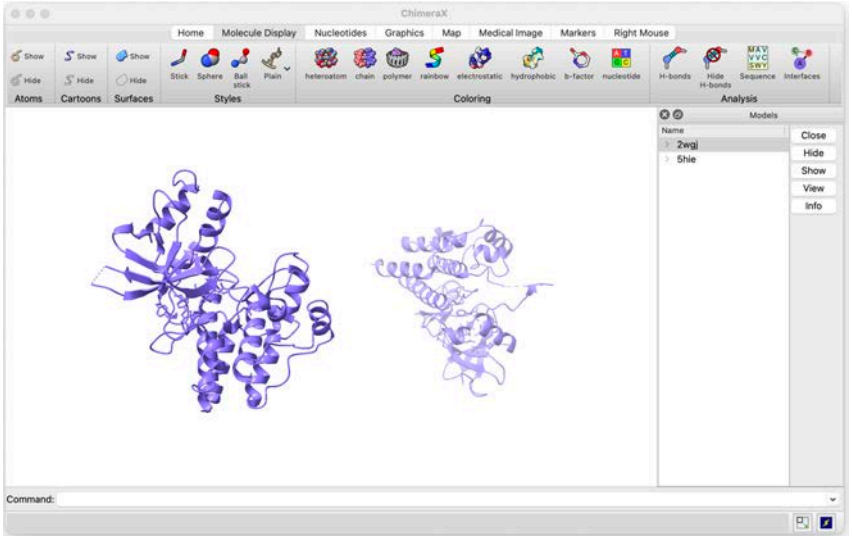


Four copies of B-Raf, corresponding to chains A, B, C and D are present in the 5hie structure. Since we only need one copy, we will delete chains B, C and D. For this, you can select one chain at a time in the selection menu, e.g. "**Select>Chains>Serine/threonine[…]B-Raf>B**" and then delete the corresponding atoms using "**Actions>Atoms/Bonds>Delete**".

Alternatively, you can type "**select #2/B,C,D**" in the command line interface to select all three chains at the same time, and then delete the atoms, again using "**Actions>Atoms/Bonds>Delete**".

Note: The same result could have been obtained in one action, by typing delete #2/B,C,D in the command line

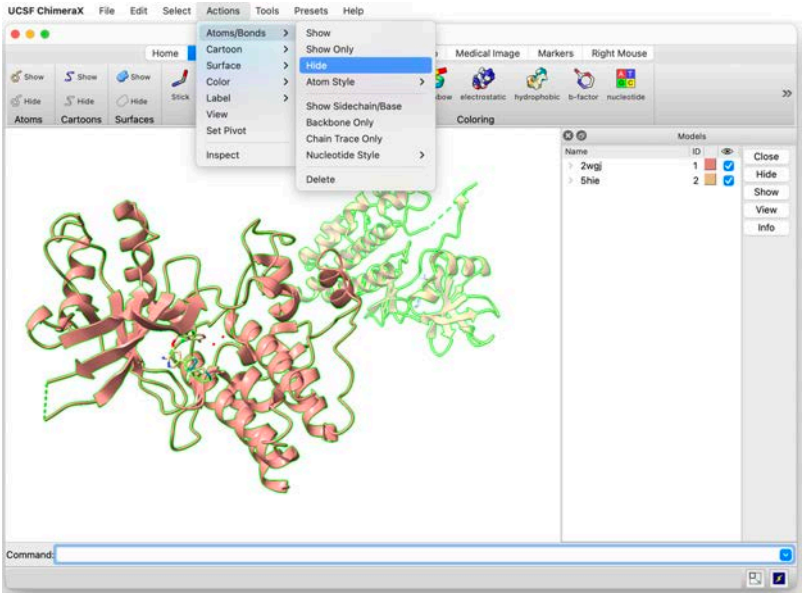
Now, the system contains only one chain of each protein.



Apply different color schemes

To better differentiate the two proteins, apply the polymer-based coloring mentioned previously, byt clicking on the “polymer” button of the “Molecule Display”’ tab of the toolbar. You can see that the color coding is also visible in the “Models” panel.

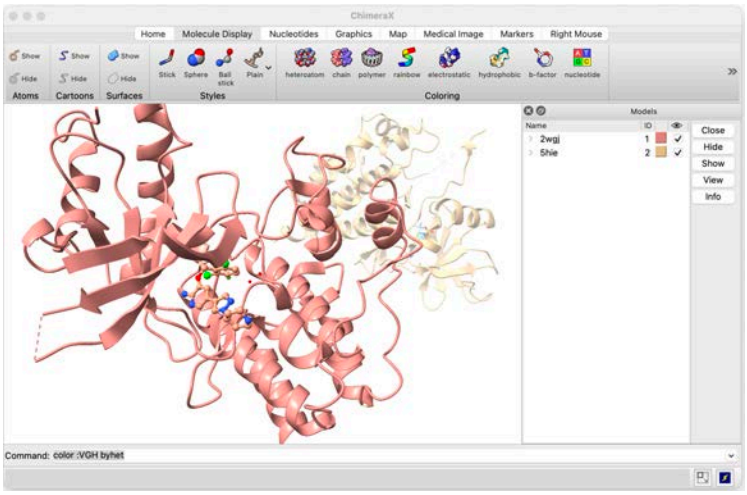
For clarity, we will hide all atoms and bonds, except those belonging to the ligands. For this, select the natural amino acids with “**Select>Residues>Standard Amino Acids**” and hide them with “**Actions>Atoms/bonds>Hide**”.



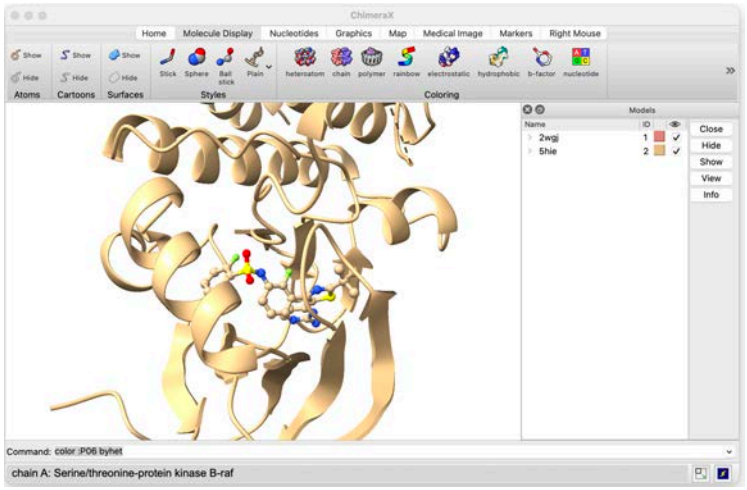
Now, we will color the ligand of MET (i.e. crizotinib, with residue ID **VGH**) to make it easily recognizable after the superimposition.
First, select it using the command `select :VGH`.
Then, display it in ball and stick representation using “**Actions>Atoms/Bonds>Atom Style>Ball & Stick**”.
In the “Color Actions” panel, check the “Atoms/Bonds” option, and apply a color identical or similar to that of the MET backbone. Now all, atoms are colored the same way.
Finally, with the ligand (VGH) still selected, color the non-carbon atoms as a function of their atom types by clicking on the “By Heteroatom” button of the “Color Actions” panel.

Alternatively, this result could have been obtained using the following commands:

```
style :VGH ball
color :VGH dark salmon
color :VGH byhet
```



Now, do the same actions to display the ligand of B-raf (i.e. dabrafenib, with residue ID **P06**) in ball and stick representation, with carbon atoms colored like the protein cartoon, and with heteroatoms colored according to their atom types.



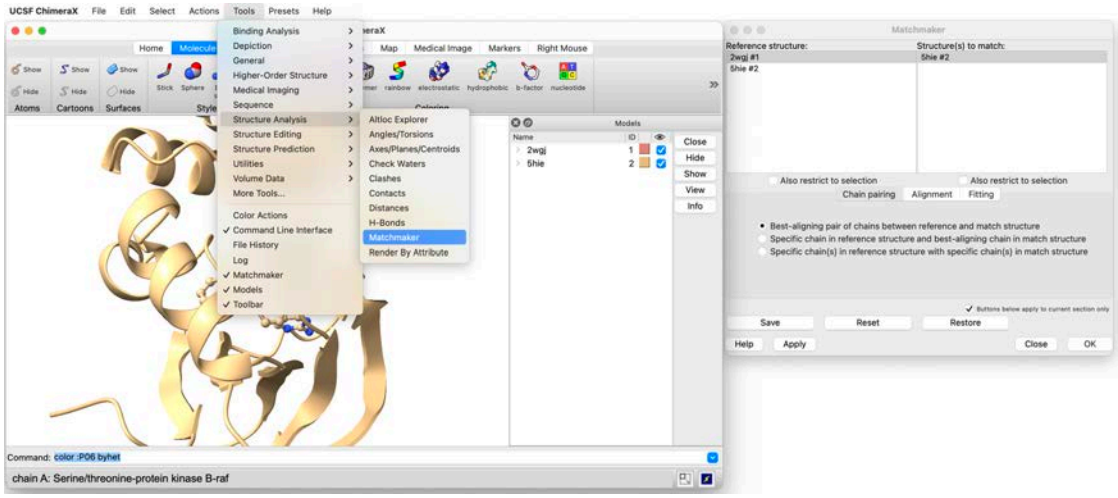
Match two proteins. Obtain a structural alignment

Open the structural alignment tool using "**Tools>Structure Analysis>Matchmaker**". Select 2wgj as "Reference structure" and 5hie as "Structure(s) to match".
In the "Chain pairing" tab of this panel, select "**Best aligning pair of chains […]**", so ChimeraX will try to superimpose each possible chain of the first protein to each possible one of the second protein.

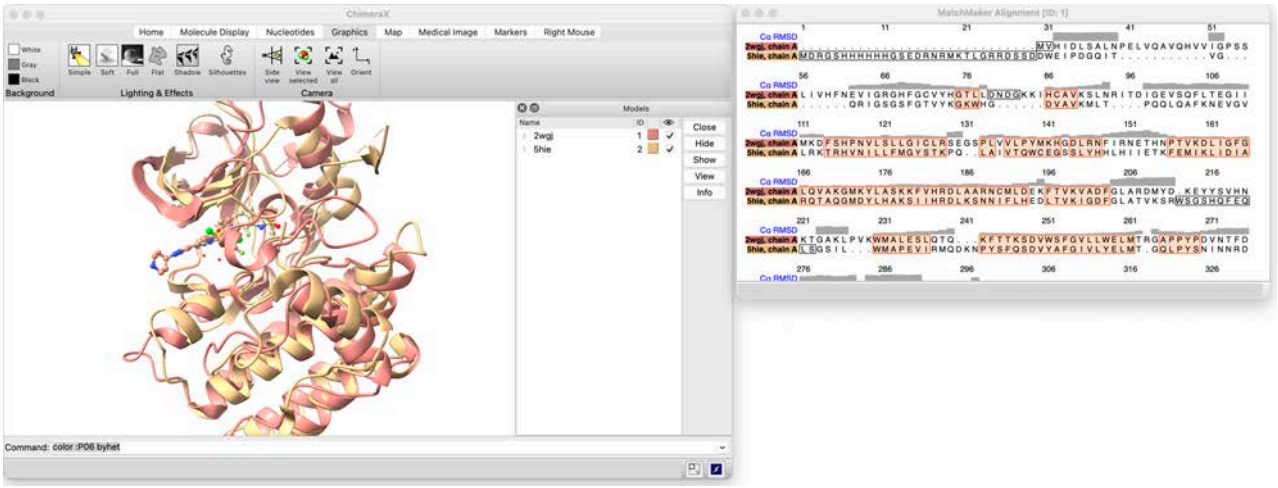
Note: Here, there is no need to select one particular option, since each protein has only one chain. However, these 3 "**Chain pairing**" options can be useful for instance to superimpose systems composed of different chains, by allowing to specify which chains to use as reference for the structural alignment.

In the "Alignment" tab, verify that the "Show pairwise sequence alignment(s)" box is selected. Finally, select "Needleman-Wunsch" as the "Sequence alignment algorithm" and click "OK". This

will perform a matching of the two molecules, based on their sequence and 3D structure similarities.



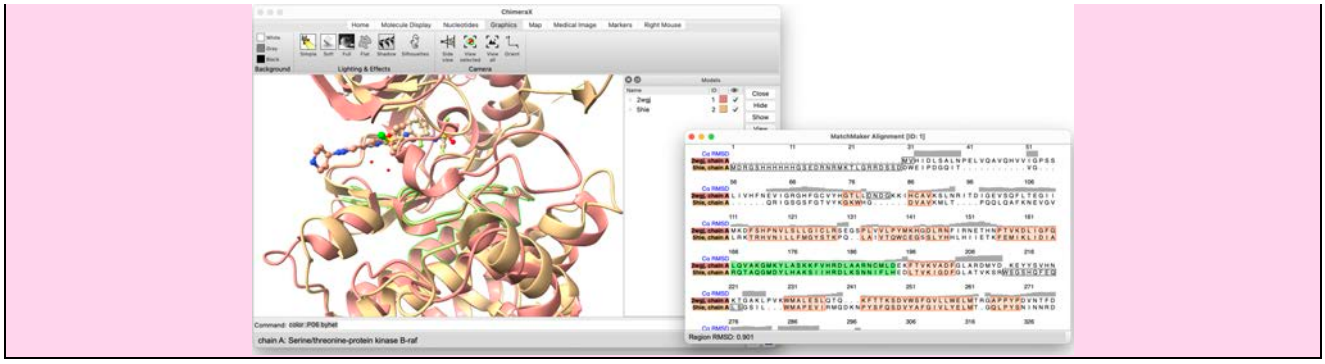
After a short moment, the two molecules should be superimposed. A new panel, called “MatchMaker Alignment” will also appear, showing the corresponding sequence alignments.



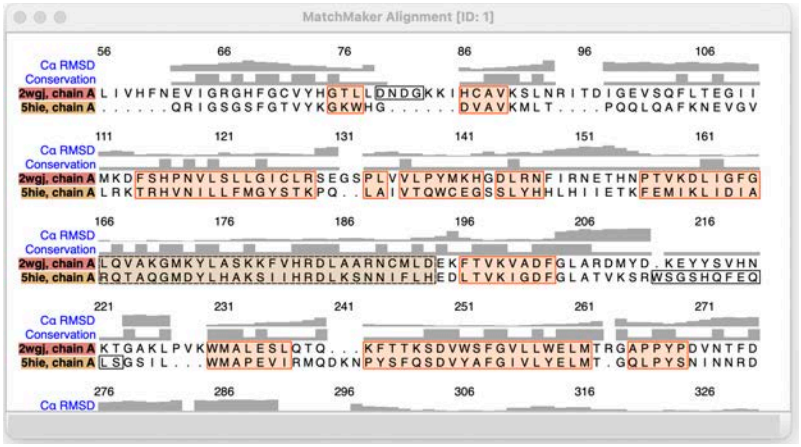
In the sequence alignment, secondary structure elements are present in colored boxes.

Above the sequence alignment there is a “Cα RMSD” line. The latter provides a histogram of the root mean square distance between the Cα atoms of aligned residues in the superimposed 3D structures. The higher this value, the more distant the corresponding Cα atoms (i.e., the less similar the two structures are in this region). As can be seen, the secondary structure elements are among the best superimposed regions, while loops are generally more different conformationally.

Note: It is possible to click-and-drag over some residues in the sequence alignment to select them. These residues will also be selected in the 3D panel.
Reversely, selecting a residue using the “Select” menu or directly in the 3D panel, will also select residues in the sequence alignment.



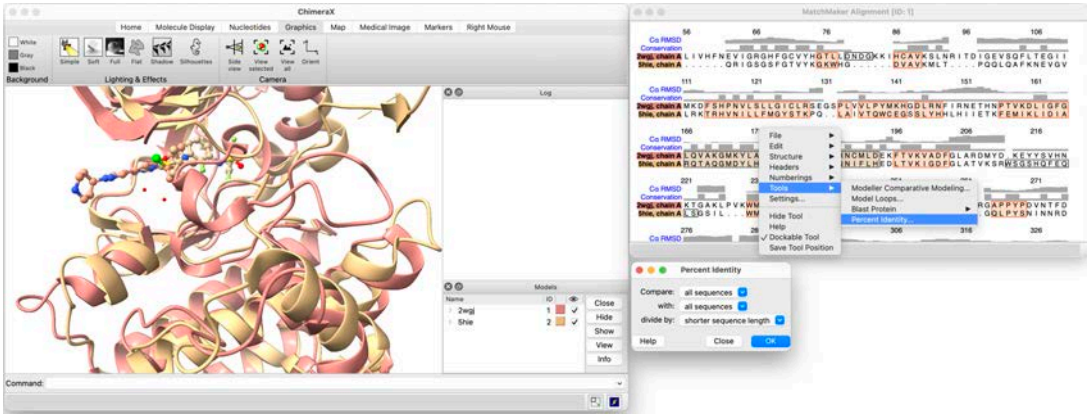
Make a right-click over the “MatchMaker Alignment” panel. In the new menu, select “**Headers>Conservation**”. This will add a new histogram above the sequence alignment, showing the sequence conservation per residue. We see that the sequence conservation is generally higher in secondary structure elements.



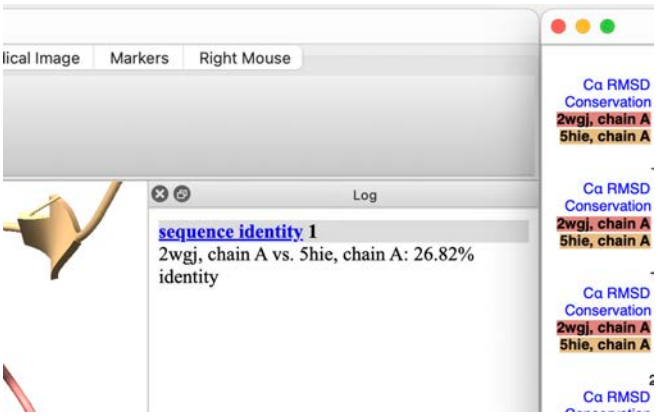
Perform the following to calculate the percentage of sequence identity between the two molecules.

First display the “Log Panel” by selecting “**Tools>Log**”.

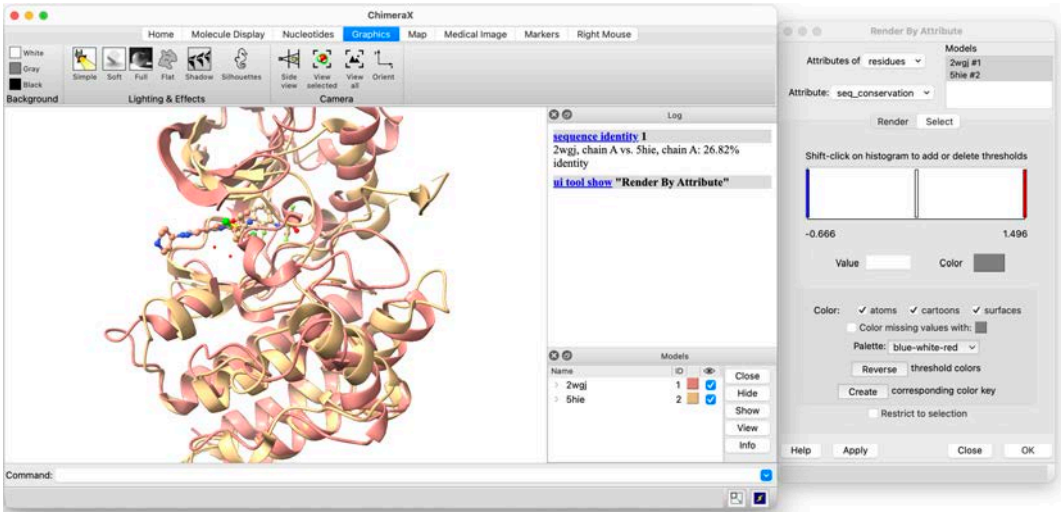
Then, make a right click over the “MatchMaker Alignment” panel and select “Tools>Percent Identity…”



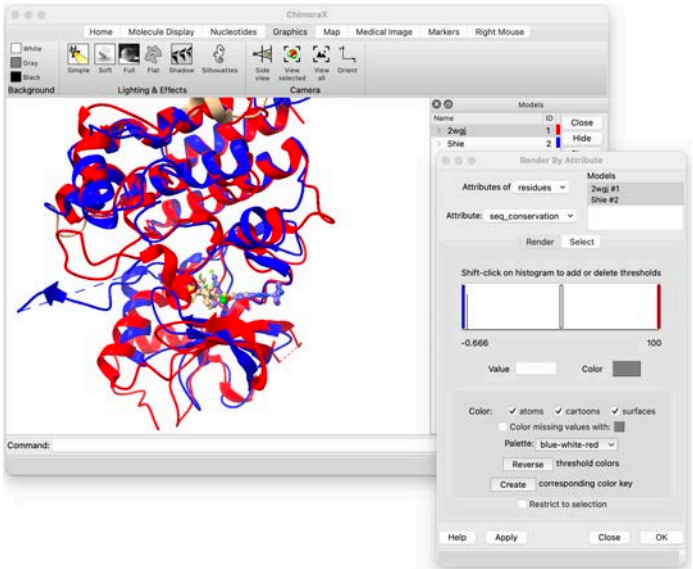
Finally, click “OK” in the new “Percent Identity” panel. The result, 26.82%, will appear in the “Log” panel.



It is possible to color the structures according to the percentage of conservation in the sequence alignment. For this, open the “Render By Attribute” panel using “**Depiction>Render By Attribute**”. In this new panel, select “residues” for “Attributes of” and “seq_conservation” for “Attribute”. The “Render” tab indicates that conserved residues will be colored in red, and others in blue. We can keep this coloring.



Finally click “Apply” using the default values for the different options.

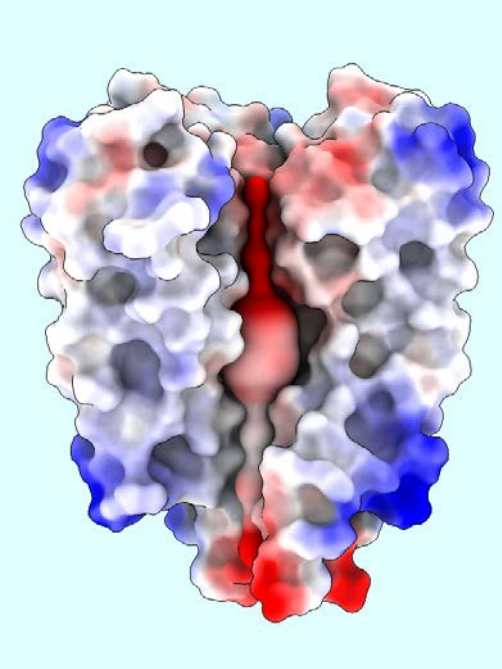


Going further with UCSF ChimeraX

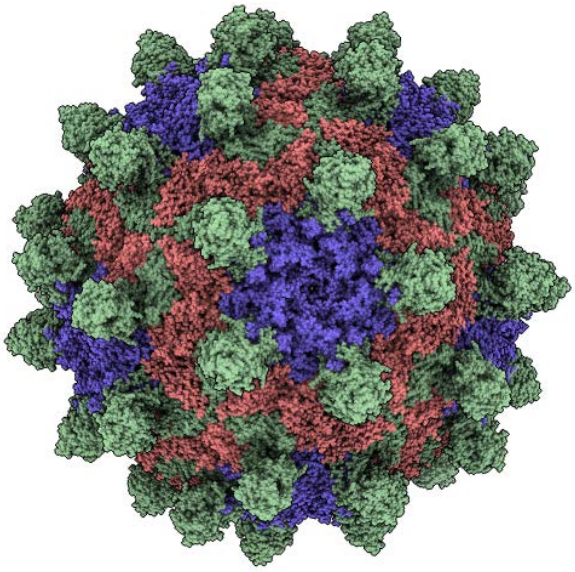
These exercises give only a very limited overview of what ChimeraX is capable of. You can find a detailed documentation, as well as some tutorials, at the following address: https://www.rbvi.ucsf.edu/chimerax/docs.

Here are some examples of images produced using ChimeraX that were taken from the official Website.

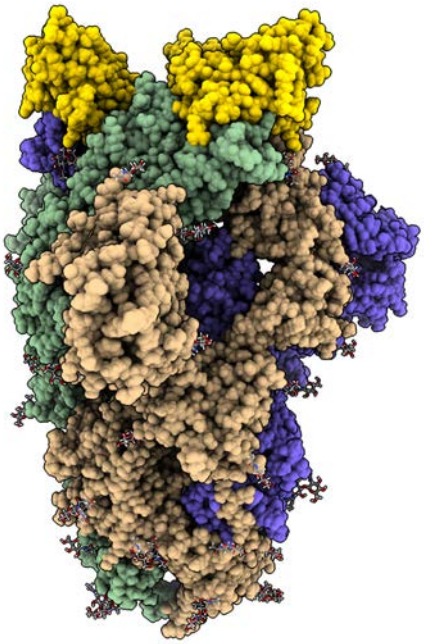
MOLE Channel Coloring by Property



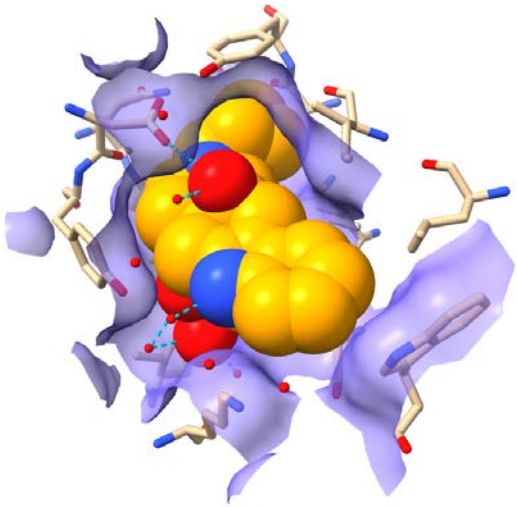
Capsid of Hepatitis D Virus



SARS-CoV-2 spike structure



Delta opioid receptor binding site

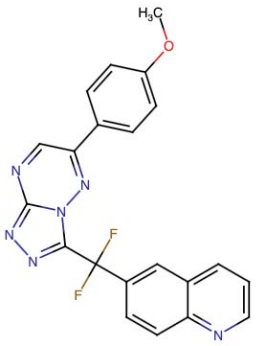


Practice Session 2:
Ligand-protein docking with SwissDock

Antoine Daina, Ute Röhrig, Vincent Zoete – 2024
Contact : vincent.zoete@sib.swiss or antoine.daina@sib.swiss

Exercise 5. Docking of a discovery compound into the protein c-Met with SwissDock

The purpose of this exercise is to use the tool **SwissDock** to dock the new molecule described below into a c-MET tridimensional structure to predict a probable binding mode. Then intermolecular interactions will be visualized within the ChimeraX molecular graphics environment.



SMILES: `COC1=CC=C(C=C1)C1=NN2C(=NN=C2C(F)(F)C2=CC3=CC=CN=C3C=C2)N=C1`

Access the web interface of **SwissDock** by typing the following URL https://www.swissdock.ch in a web browser (preferably *Google Chrome* or *Mozilla Firefox*).

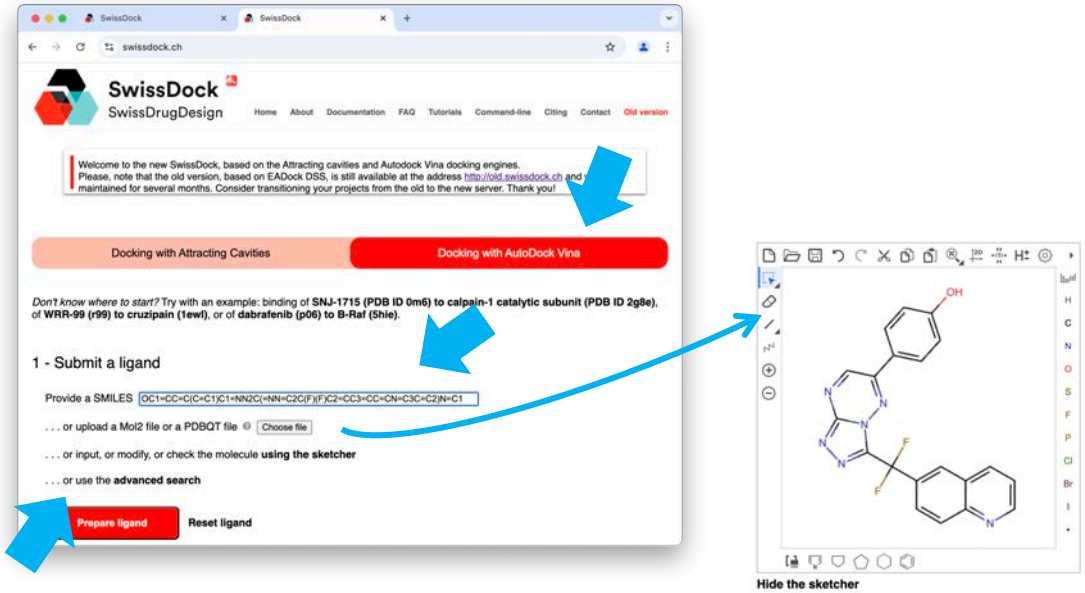
1 – Select the AutoDock Vina engine and submit the ligand

First click on the "**Docking with AutoDock Vina**" right-hand tab, which becomes red.

Then input the molecule described above as ligand. For this, you can either:

- copy/paste the SMILES into the "**Provide a SMILES**" text box, *or*
- draw the chemical structure in the sketcher (click on "**using the sketcher**").

Note: the content of the text box and the sketcher are synchronized



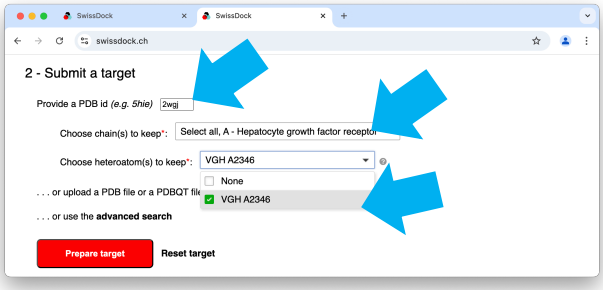
Finally, click on the "**Prepare ligand**" button (which turns red when active).
It takes a few seconds for the molecule to be prepared.

2 – Submit the target

You can provide the *PDB id* of the X-ray structure of c-MET crystallized with Crizotinib, by **typing "2wgj"** and <enter>.

- Choose to **keep** only **the chain A** – Hepatocyte growth factor receptor
- Choose to **keep** only **VGH A2346**
  This 'residue' corresponds to Crizotinib, useful to place the searching box. It will be removed later (next section).

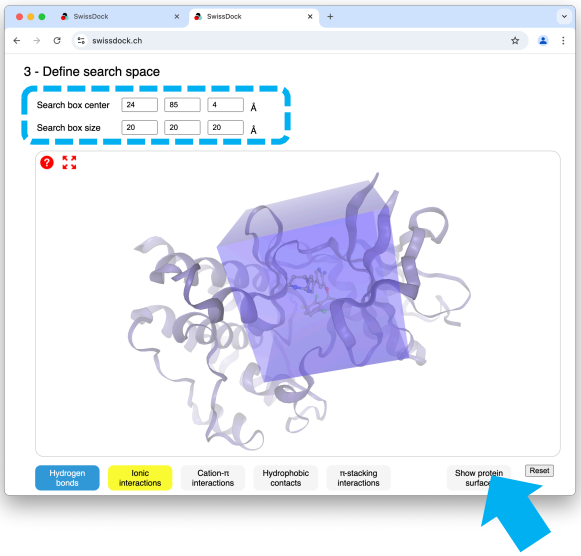
Note: In case you don't know the PDB id or want to get more info about it, or would like to choose a different target structure, you can click on "advanced search".



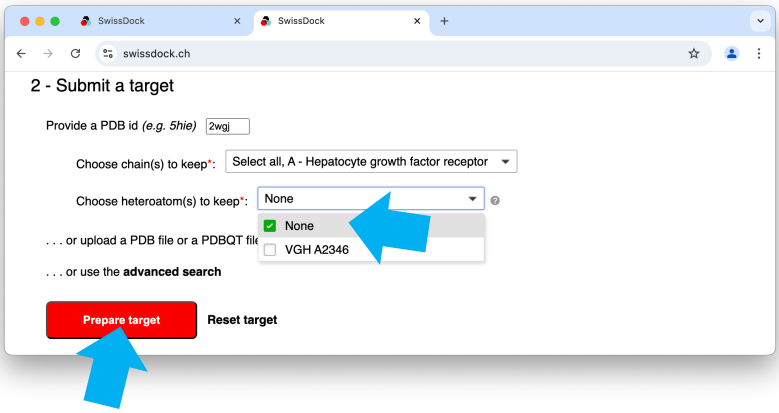
3 – Define search space

The poses will be searched on a defined region on the protein surface. This approach is called *focused docking*, as opposed to *blind docking* (for a search all over the protein surface).
Here, one makes the hypothesis that your molecule has the same binding pocket as typical c-Met inhibitors like Crizotinib.

With the help of the **Search box center** and **size**, place a cube of (20Å)^3 so that it encompasses the entire cavity that accomodates the co-crystallized Crizotinib. You can switch on and off the protein surface to help you.

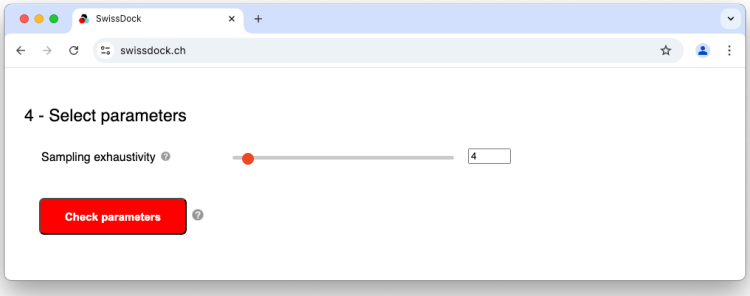


You can now remove the Crizotinib from the protein structure to prepare. Go back to the “Submit target” section and in **heteroatroms to keep,** select “**None**” in the drop-down menu



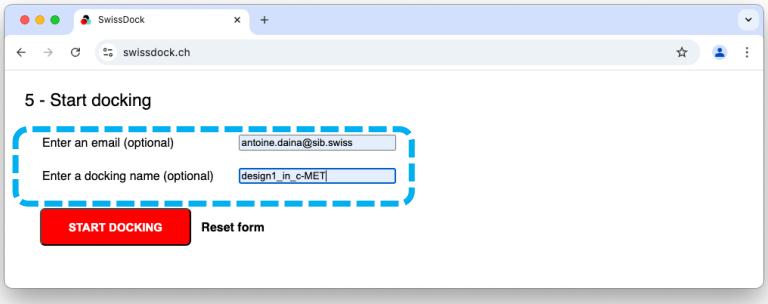
Finally, click on the “**Prepare target**” button (which turns red when active).
It takes several seconds for the protein to be set up.

4 – Select parameters



Set exhaustivity to 4 (default)
Click on the “**Check parameters**” button (which turns red when active).
It can take some time to validate the parameterization.
Once this step completed, you will receive an estimated duration of your docking.

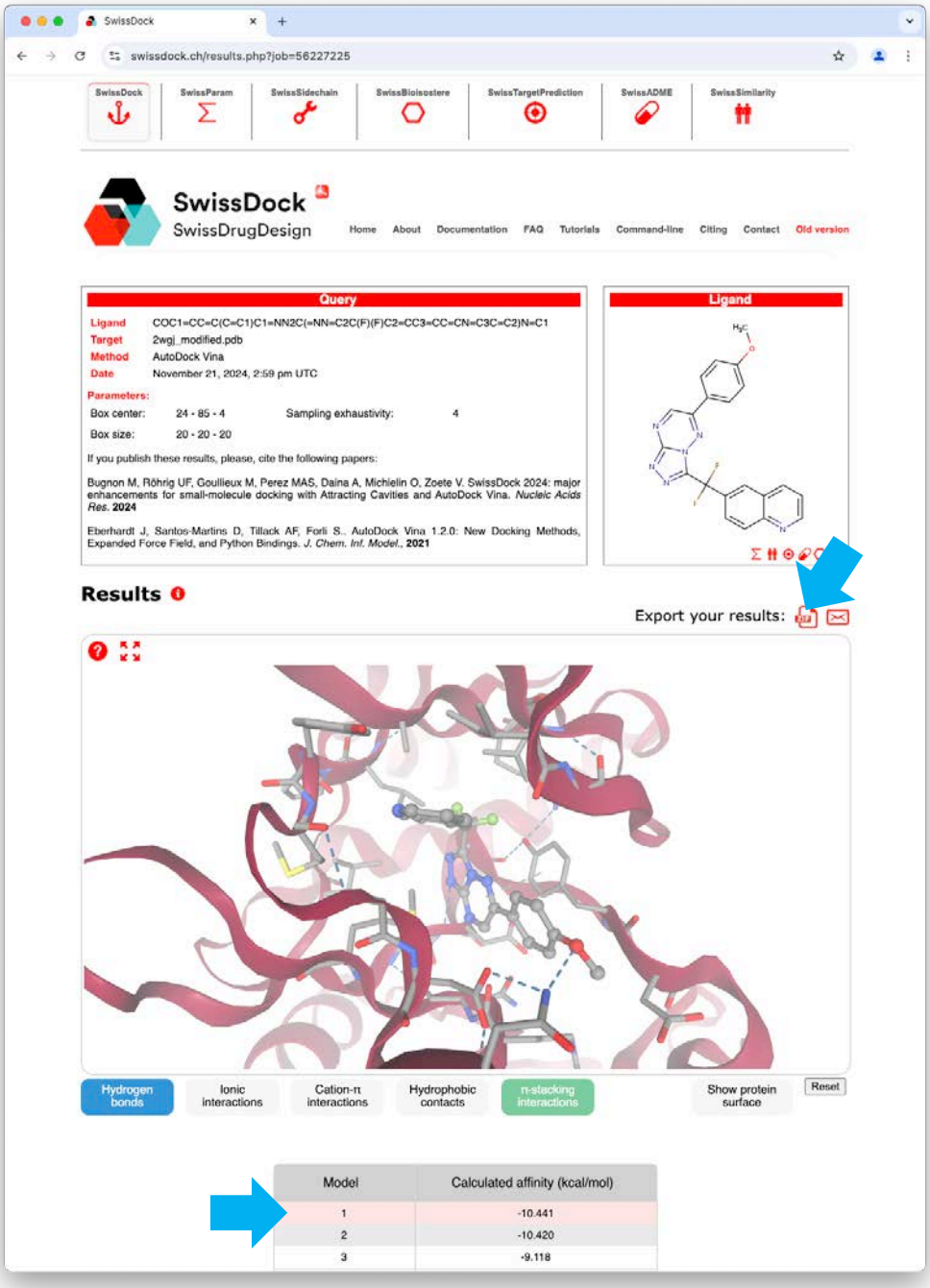
5 – Start docking



Enter an **email address** and a **job name**.
Click on the “**START DOCKING**” button (which turns to red when active).
It should take about 30 secondes to one minute, once your run has started (there is a queueing system on SwissDock server).

6 – SwissDock output page and download of the results

The output page appears in a new tab upon docking completion.



The docking output can be analyzed in the web page thanks to the interactive 3D window. Click on a docking pose (“Model”) to display it, and visualize it inside the binding site. You can display different types of intermolecular interactions, as well as the protein surface by clicking on the buttons below the 3D window.

For deeper analyses, you are going to download the results to be then opened in ChimeraX (next section).
Click on the Export ZIP icon, an archive file **<job_name>.zip** will be saved locally on your hard-drive (e.g. in the Downloads folder).
Navigate to this location on the hard-drive of your computer. Locate the **<job_name>.zip** archive, **uncompress it** to create a sub-folder containing SwissDock input and output files. Some of these files will be opened in ChimeraX.

7 – Analyze docking results in ChimeraX

Open ChimeraX but clicking on the icon.

*Don't forget to save **ChimeraX sessions (.cxs)** from time to time for easy recovering in case of mistake!*

> **>File >Save… Go on a prefered location on your hard-drive,** give a name and select Files of type: ChimeraX session (*.cxs)
> Click the **"Save"** button**.**
> *Details on p.13*

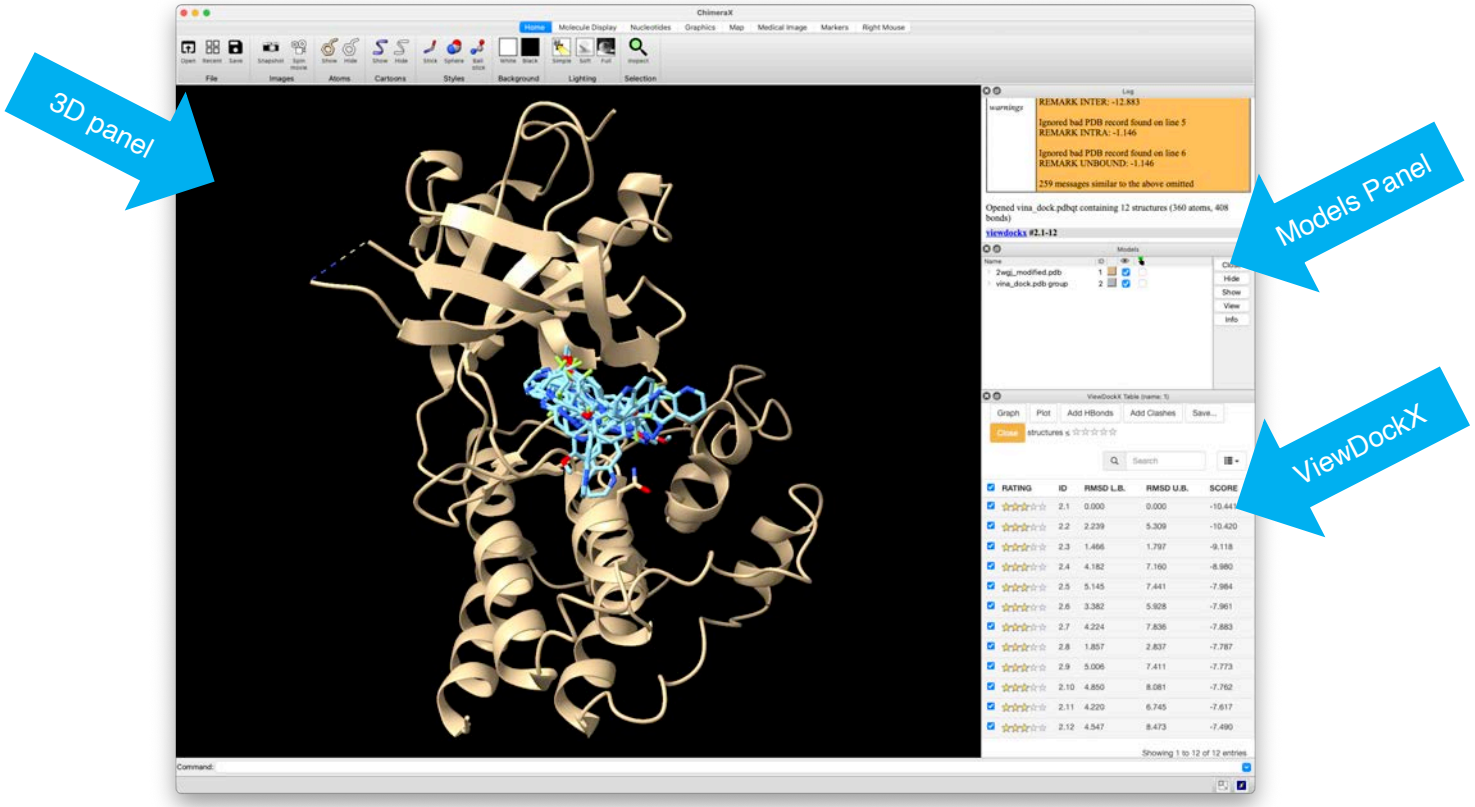
Open the c-MET structure used as docking target which is described in a PDB file located in the sub-folder of your job created at the location defined in the previous section.

> **>File >Open…**
> Navigate to (e.g. Downloads/design1_in_c-MET), select **2wgj_modified.pdb** and click **"Open".**

Open the docking result which is in a PDBQT file at the same location on your hard drive.

> **>File >Open…**
> Navigate to (e.g. Downloads/design1_in_c-MET), select **vina_dock.pdbqt** and click **"Open".**

In the main *3D panel*, are displayed the c-Met protein as cartoon (Model #1, as shown in the *Models panel*) and all docking poses as sticks (Model group #2), which are detailed in the *ViewDockX panel*.



Calculate and display the solvent-accessible surface of the protein

Select the protein, for instance:

Select >Residues >Standard Amino Acids

The structure is highlighted by thin green lines showing that it is selected.

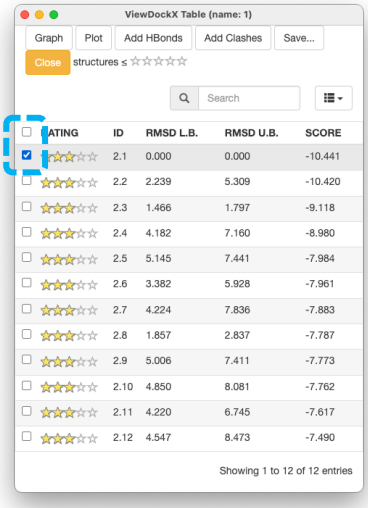
Actions >Surface >Show

Clear selection by **ctrl click** in an empty space of the *3D panel*.

- *How many docking solutions are proposed?*
- *Are they well accomadated in the envisaged cavity?*
- *What is the best score? In what range of K_i corresponds this score?*

Display the best scored pose only.

Untick the general check box in ViewDockX (on the top) and tick the one in the first line (model #2.1).



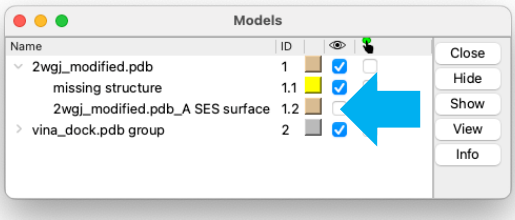
You can go through each docking solution by pressing the ↓ **key**.

It is also possible to display multiple poses by ticking multiple checkboxes**.**

Go back to display the first docking solution only (as in the picture above).

Hide the surface of the protein through the *Models panel*.

Develop the model #1 by **clicking the ">" on the left and untick the corresponding "eye" checkbox #1.2**



Calculate and display hydrogen-bonds

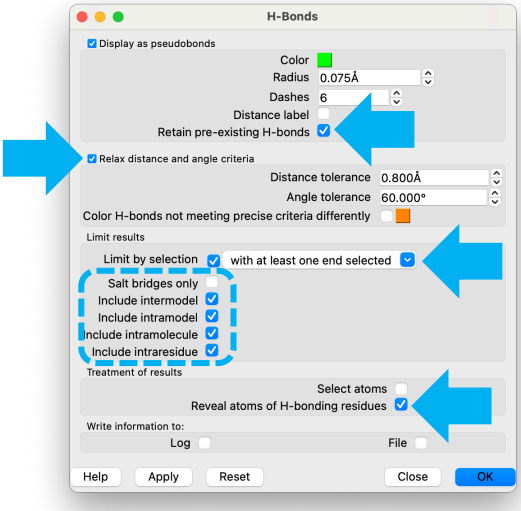
Evaluate the intermolecular **hydrogen-bonds** between this pose and the nearby amino acids that could explain (at least partly) the recognition of the your molecule by c-MET.

Select the first docking solution by:

- typing `select #2.1` in the command line, or
- holding the **ctrl key** and clicking on one of its atom in the 3D panel, then **press** ↑

The molecule is highlighted by thin green lines showing that it is selected.

>Tools >Structure Analysis >H-Bonds



Check that the H-Bonds settings are as shown n the image on the left to:

- Retain already shown H-bonds
- Relax criteria that define H-bonds
- Only includes H-bonds with seleted atoms
- Include intra- and intermolecular H-bonds
- Display residues involved if hidden.

Press the **"Apply"** button.
The computed H-Bonds are dispalyed in dashed lines. Hover the mouse on the line to show the details.
Note: AutoDock Vina considers hydrogen atoms as implicit and diplays only polar hydrogens of the ligand.

- *How many intermolecular H-bonds are proposed for this first docking solution?*
- *Note the residues involved, if the polar atom belongs to the backbone or to the side chain of the amino acid, and if it is a H-bond donor or an acceptor.*

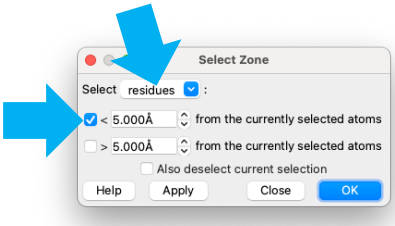
Repeat the calculation, display and analysis of hydrogen-bonds for one (or two) docking solution(s) that can be considered as realistic, for instance those with significant best scores (most negatives).

Visualize other intermolecular interactions.

Display only and select what you consider the best docking pose (with the **select** command or by the crtl clicking method).
Display the amino acids close to the selected pose:

Select > Zone…

In the Select Zone window, select **residues** and tick **"< 5.000Å** from the currently selected atoms".
Click on "**OK**"



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

- *Give two examples of intermolecular hydrophobic interactions.*
- *Can you spot intermolecular aromatic interactions?*
- *Can you identify salt-bridges?*
- *Propose one targeted mutation that could validate the predicted binding mode of your compound in c-MET.*

Make the best possible JPEG image that illustrates how your design molecule is predicted to bind to c-MET, with the solvent accessible surface of the protein (possibly transparent or cropped). Emphasize, select and label (once selected **>Actions >Label >Residue >Name and Number**) the amino acids involved in important intermolecular interactions, like hydrogen-bonds or aromatic stacking.

>File >Save As… and select **JPEG** as Files of type.
*The "**Snapshot**" button in the top tool bar is also a good and easy way of doing images, if you don't need to control all parameters.*

Exercise 6. Structure-based design in c-Met

Perform structure-based optimization yourself!

Are you able to design and dock one or two analogues of your compound chemically modified to optimize the binding to c-MET?

The easiest way could be to copy/paste the SMILES of the original molecule (p. 52) into the submission page of SwissDock and apply the chemical modifications in the molecular sketcher.

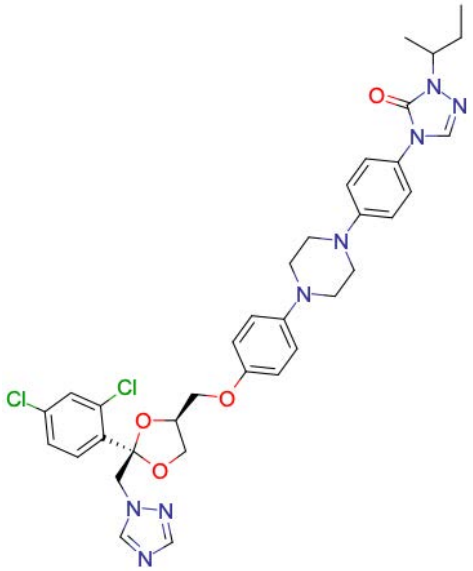
Perform the docking with the same engine (AutoDock Vina) into the same protein structure and with the same parameters so that you can import the results in ChimeraX to directly compared the docking solutions and scores of the different designed compounds.

Practice Sessions 3-6
Ligand-based Drug Design

Antoine Daina, Vincent Zoete – 2024
Contact : vincent.zoete@sib.swiss or antoine.daina@sib.swiss

Practice Session 3. Ligand-based virtual screening with SwissSimilarity

Before the actual workshop, we will perform a the simple preliminary virtual screening of a commercially available antifungal drug, **Itraconazole** (Sporanox®), which inhibits fungal 14α-demethylase, important for cell wall synthesis. We will search for similar molecules inside the library of all drugs approved by the FDA. This will illustrate the concept of ligand-based virtual screening, and show how to calculate enrichment. Here are the 2D structure and the SMILES of Itraconazole:

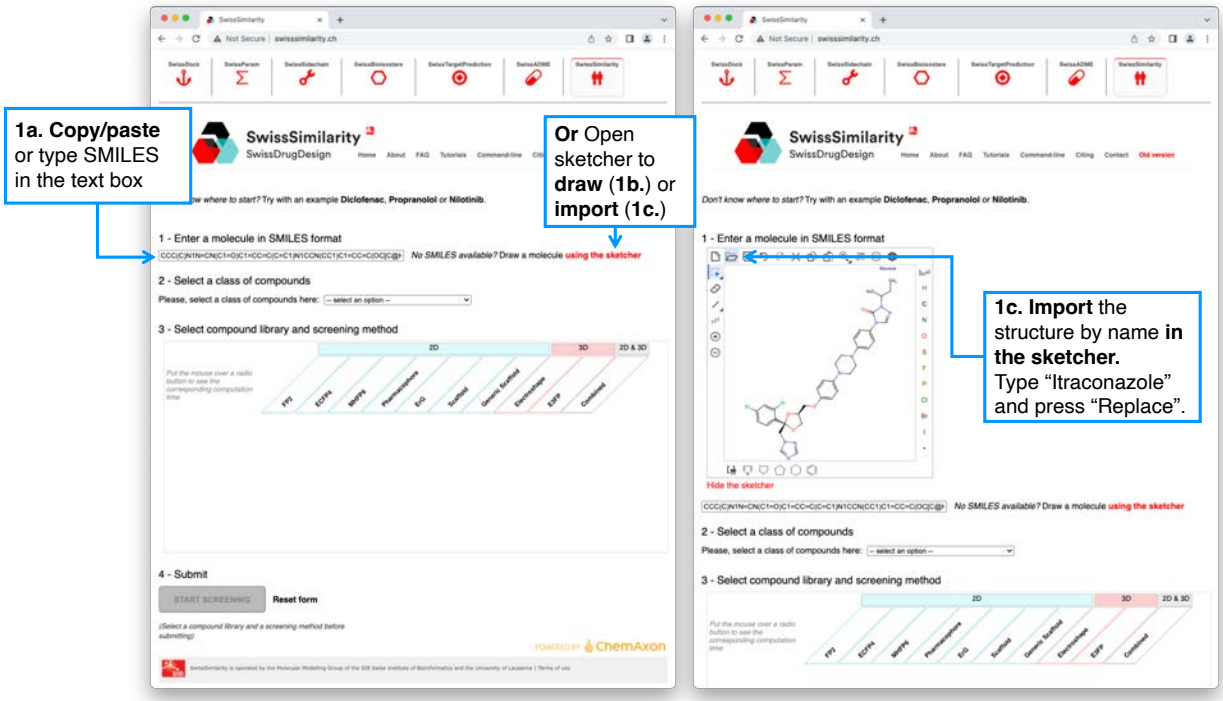


SMILES: `CCC(C)N1N=CN(C1=O)C1=CC=C(C=C1)N1CCN(CC1)C1=CC=C(OC[C@H]2CO[C@@](CN3C=NC=N3)(O2)C2=CC=C(Cl)C=C2Cl)C=C1`

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

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

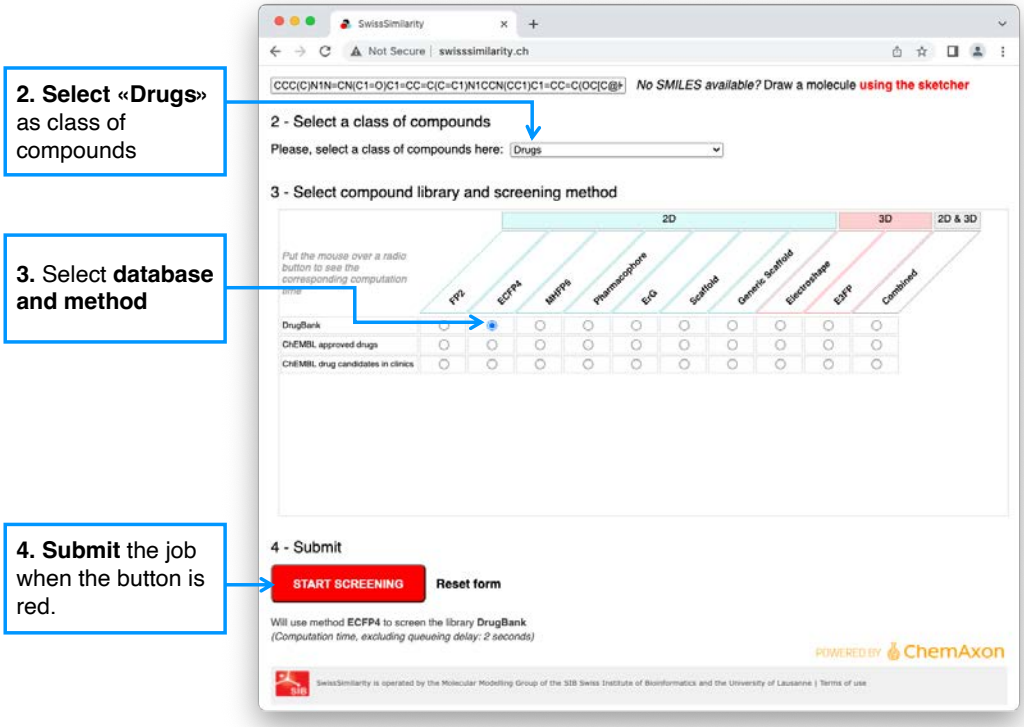
   Please note that the SMILES in the text field and the 2D structure are synchronized.



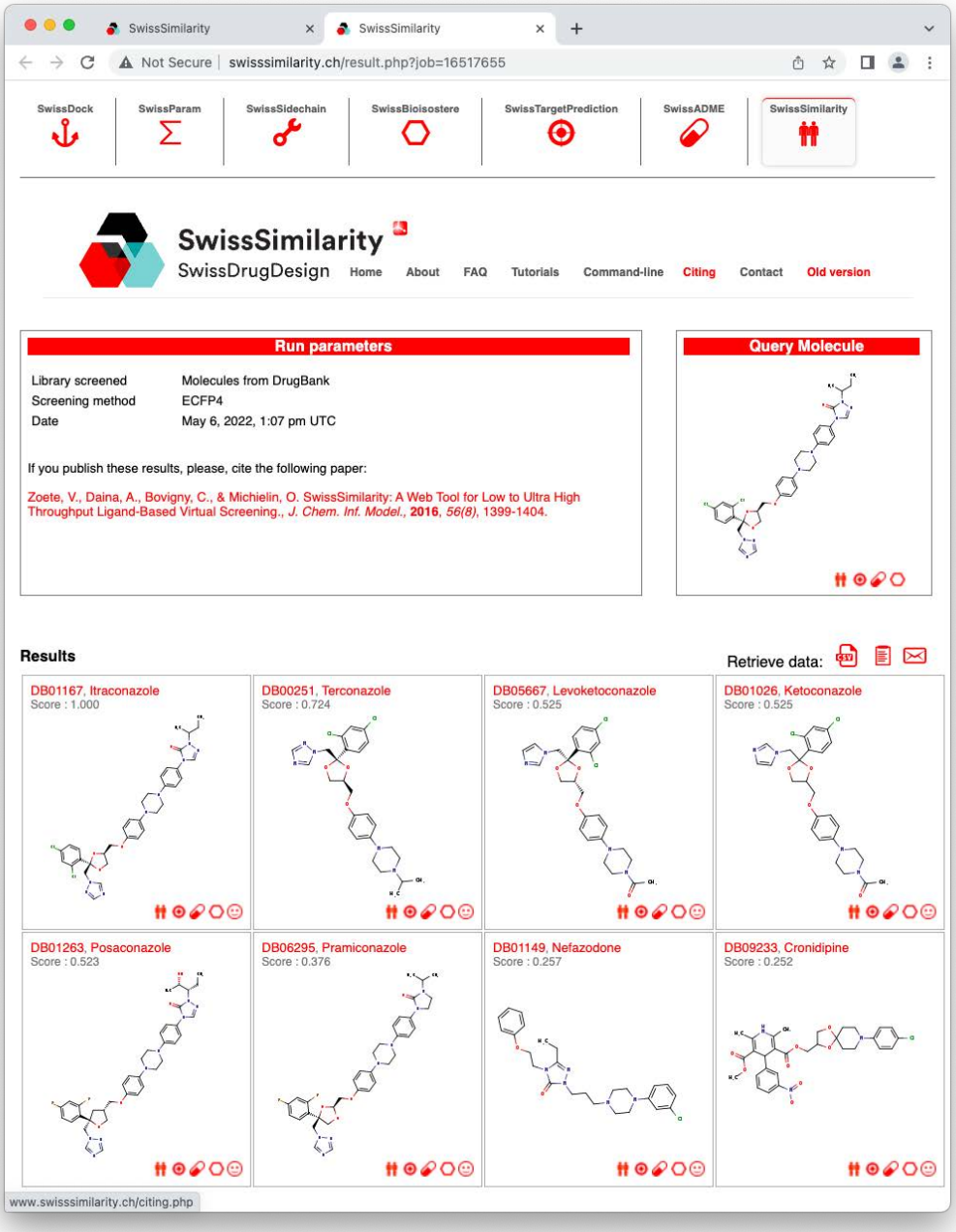
2. Select “Drugs” as the class of compounds in the drop-down menu.

3. In the dynamic table of databases and screening methods, click on the radio button to **select “DrugBank”** to search for similar compounds to Itraconazole using “**ECFP4**” fingerprints.

4. Submit the computation by clicking the **“START SCREENING” 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 a few seconds, the DrugBank compounds most similar to Itraconazole, from a chemical point of view, are displayed in a new tab.



- Looking at the output page, let's answer these questions:
  - *How many DrugBank drugs have been found?*
  - *Are these structures chemically similar to the one of Itraconazole?*

- Keep this page open for the next exercise.

Exercise 7 (preliminary). Enrichment of antifungal drugs molecules in FDA-approved drugs.

Calculate the **enrichment factor** (EF) obtained by screening the DrugBank database for Itraconazole, using ECFP4 fingerprints.

There are **2726 FDA-approved drugs** in the DrugBank database, among which **28 are antimycotic azole derivatives** targeting fungal 14α-demethylase**.** All their names **end with** suffixes **-conazole or -imazole**.

The rate of antimycotic azoles in the whole DrugBank database is $\mathbf{r_{db} = 28 / 2726 = 0.01.}$

*Calculate the enrichment factor (**EF**) at **top 8** (i.e. at 0.3%)*

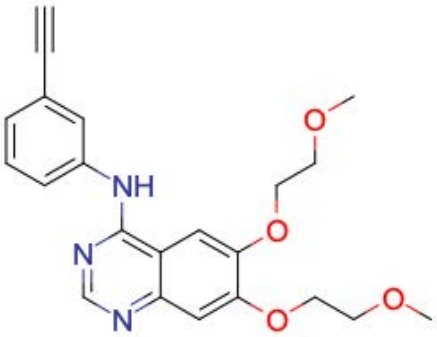
- *How many of these* antimycotic azoles *can be found in the top 8 of your screening* **(n_{screen})** ?
- *The rate at top 8:* $\mathbf{r_{screen} = n_{screen} / 8}$

$$EF = \frac{r_{screen}}{r_{db}}$$

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

Let's start with the workshop!

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 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 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.
  (Please note, a detailed video tutorial is available at:
  http://www.swisssimilarity.ch/tutorials.php)

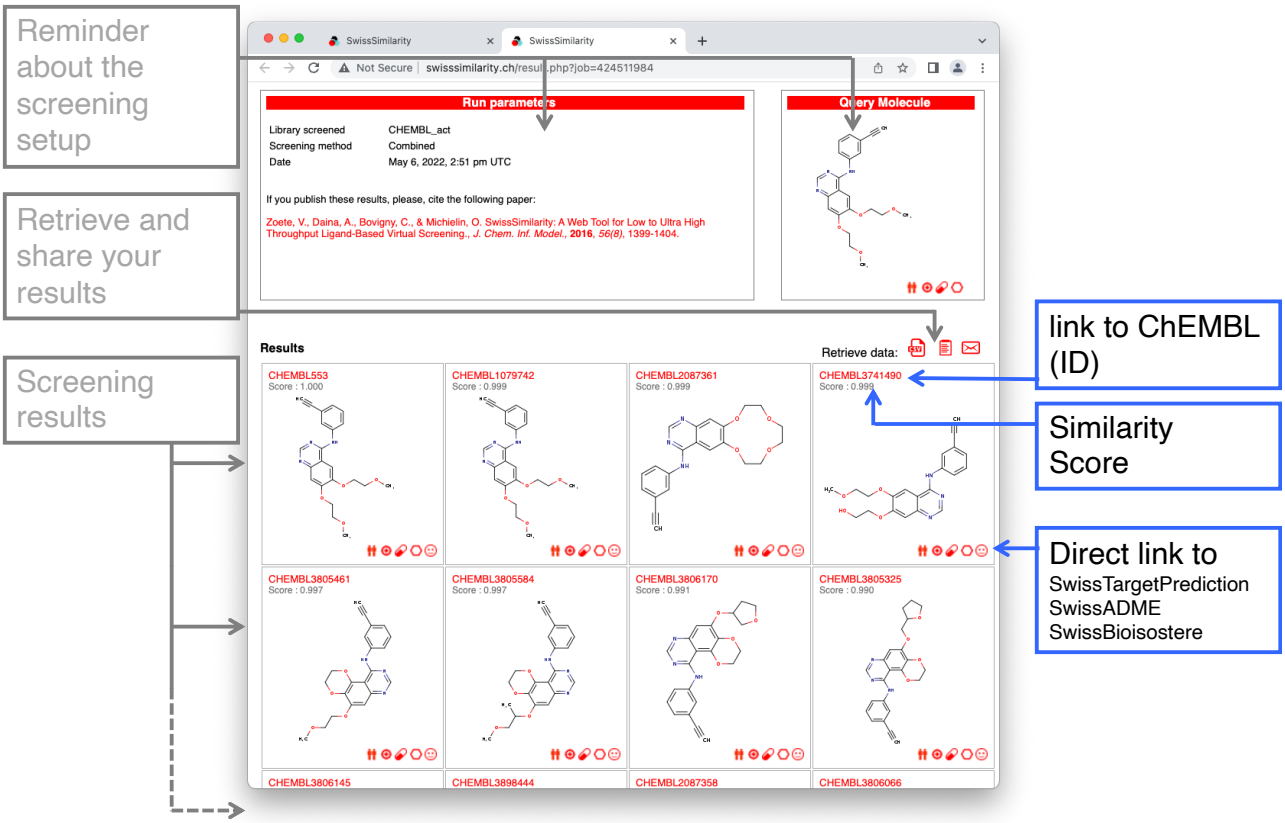
- Use one of the following options to **input the molecule** to be screened against (i.e. Erlotinib):
  a. copy/paste or type the SMILES in the text box, or
  b. draw the structure in the sketcher, or
  c. import the structure (by name) in the sketcher.

- Select "**Bioactive**" as the class of compounds in the drop-down menu.

- Click on the radio button in the table to **select both the screening method and the library** to screen. Here it is proposed to evaluate the library "**ChEMBL (actives only)**" for similarity with Erlotinib using a "**Combined**" method. The latter makes the process screen the library with both 2D-fingerprint (FP2) and 3D-electroshape (ElectroShape) to return a consensus value, corresponding to a combined score obtained 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 a reference molecule and the selection of a 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 in a new tab of your web browser. In the meantime, you can follow the progression of the calculations thanks to the blue bar.



📣 Given those 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.

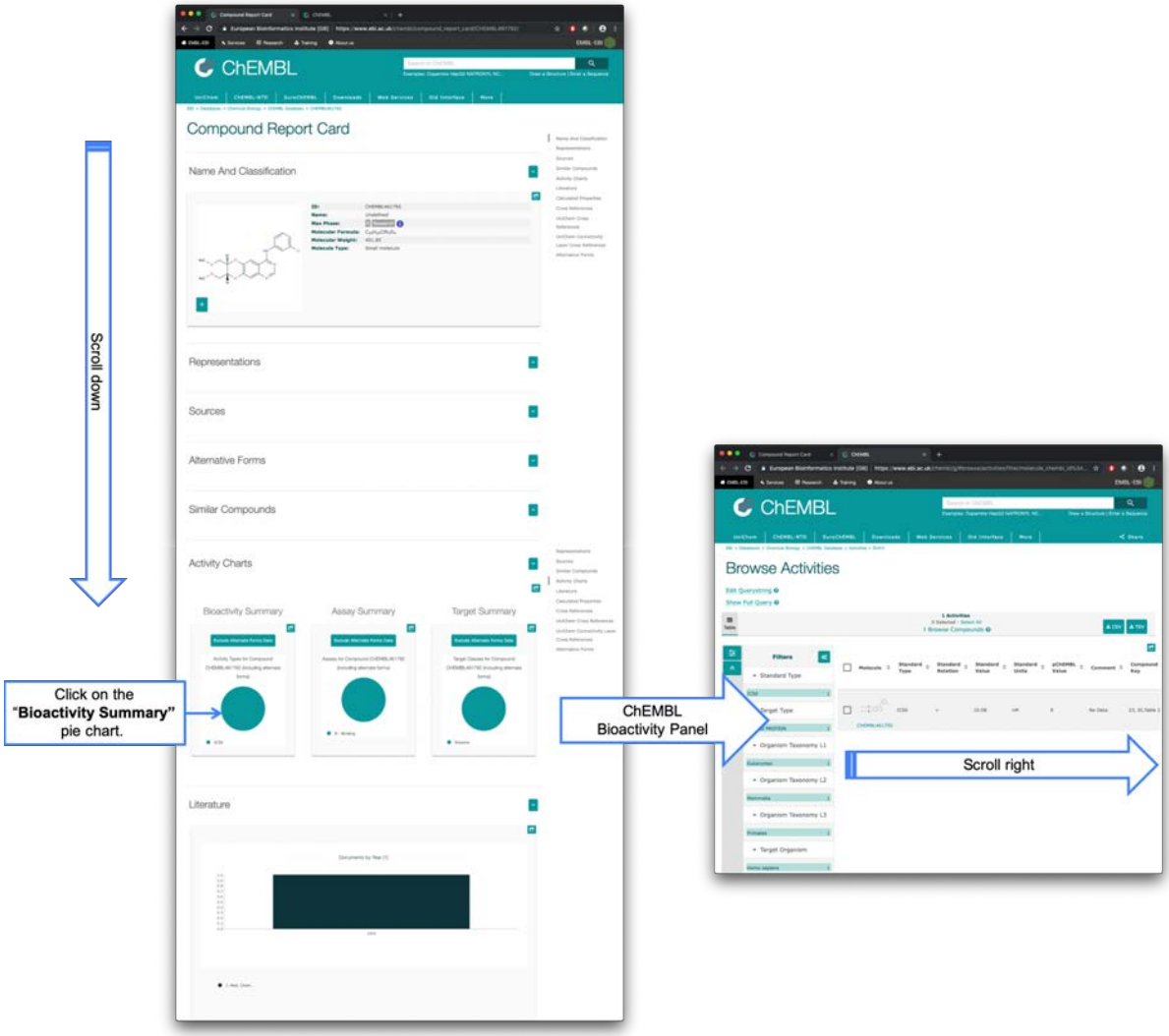
You can copy, bookmark or email the URL of your SwissSimilarity result page.

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**).

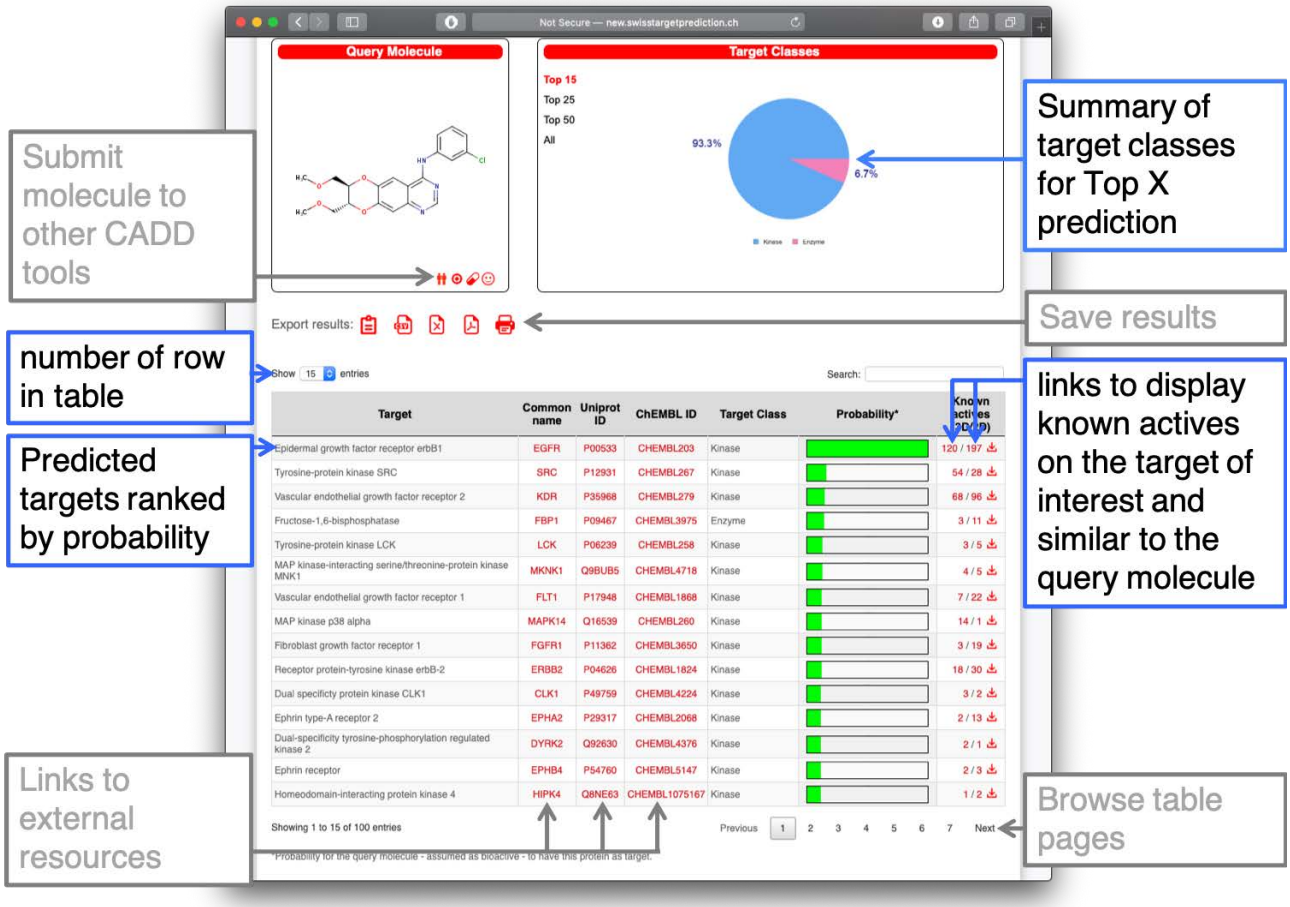
2. A first **ChEMBL** panel corresponding to entry CHEMBL461792 opens. Scroll down to the "Activity Charts" section and click on the "**Bioactivity Summary"** pie chart.
   This brings you to a second "ChEMBL Bioactivity" panel.



📣 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 protein targets**, 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 to copy/paste the SMILES in the dedicated text box, and finally click on the "Submit" button.

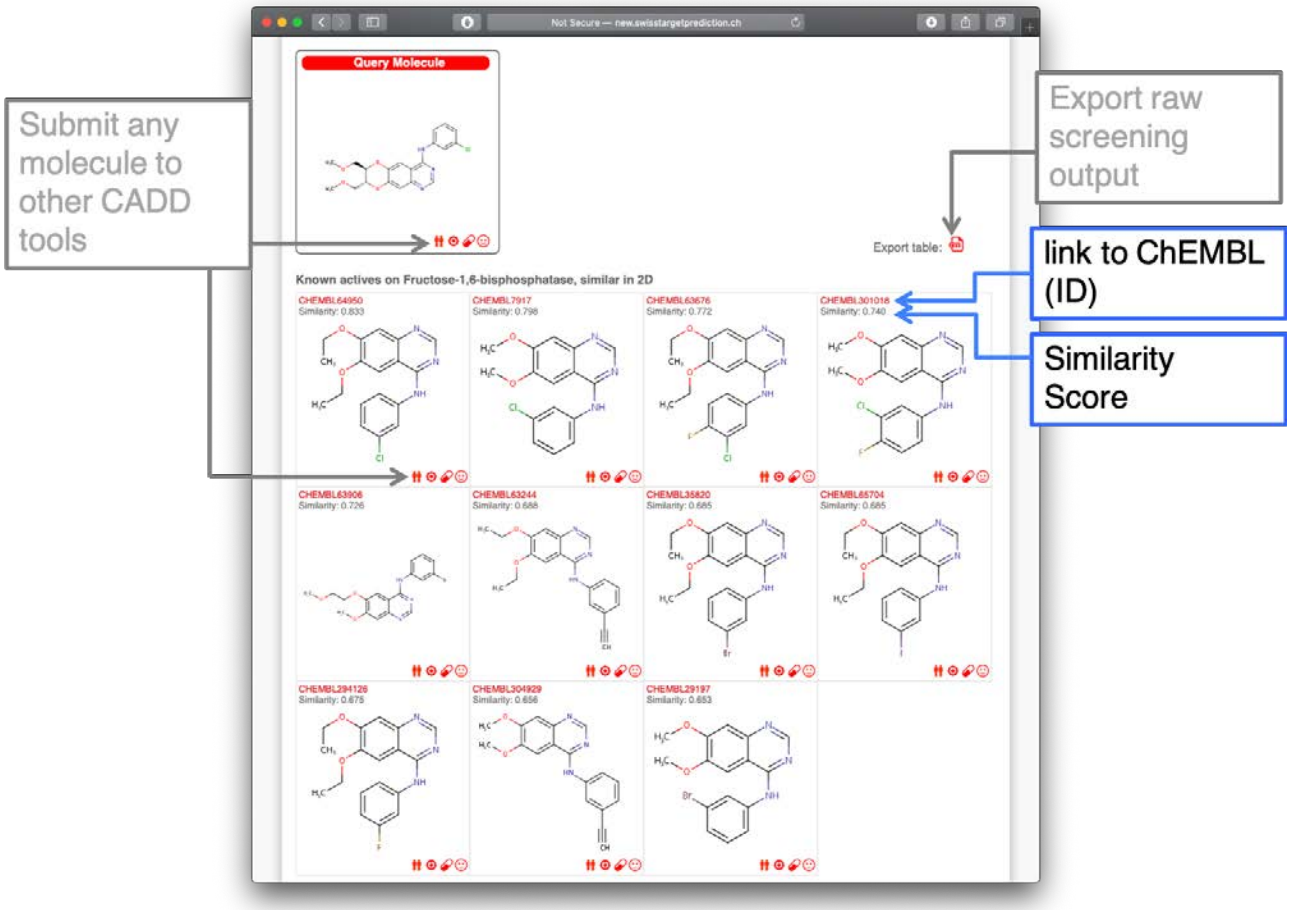
- After less than one minute, the **target prediction results** should appear in your web browser.



With those 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.



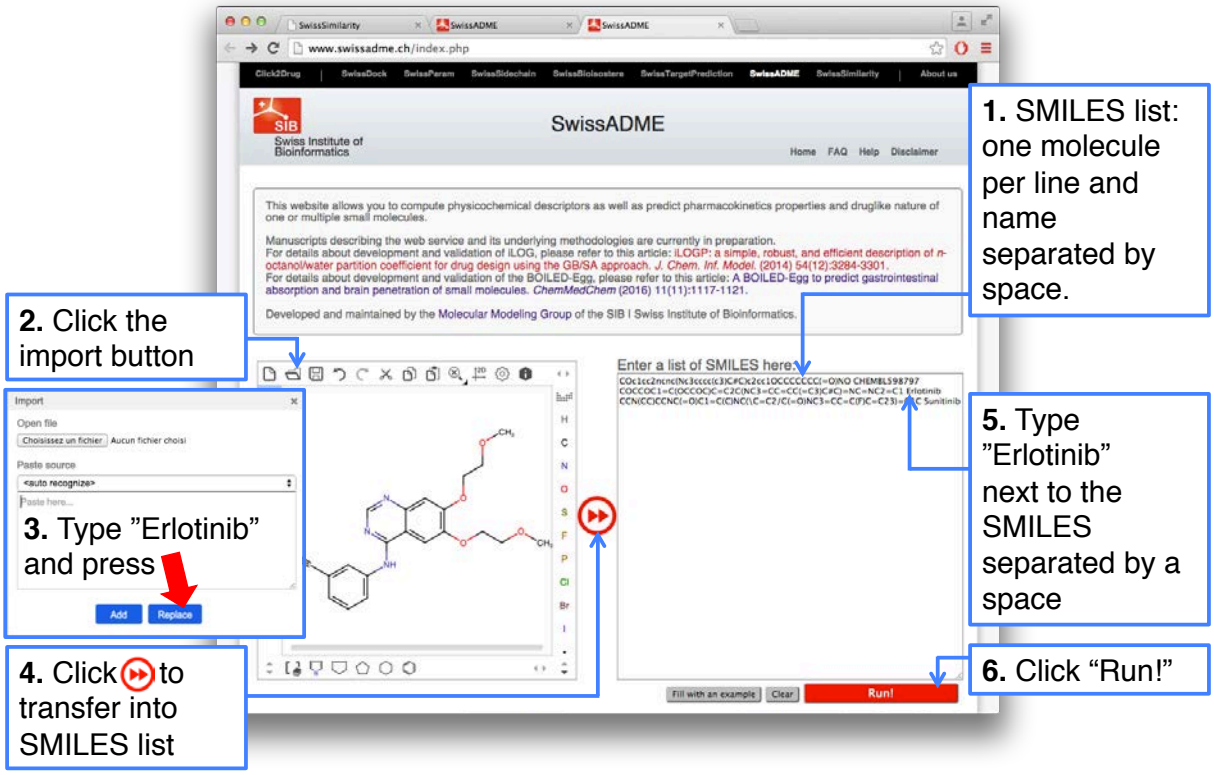
- *Are the actives most similar in 2D and in 3D to CHEMBL461792 the same compounds?*

Practice Session 5. Pharmacokinetics evaluation with SwissADME

During this session, it is proposed to estimate the druglikeness and two important ADME parameters 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: 0.961) and click on the corresponding "**pill" icon** to directly launch 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.

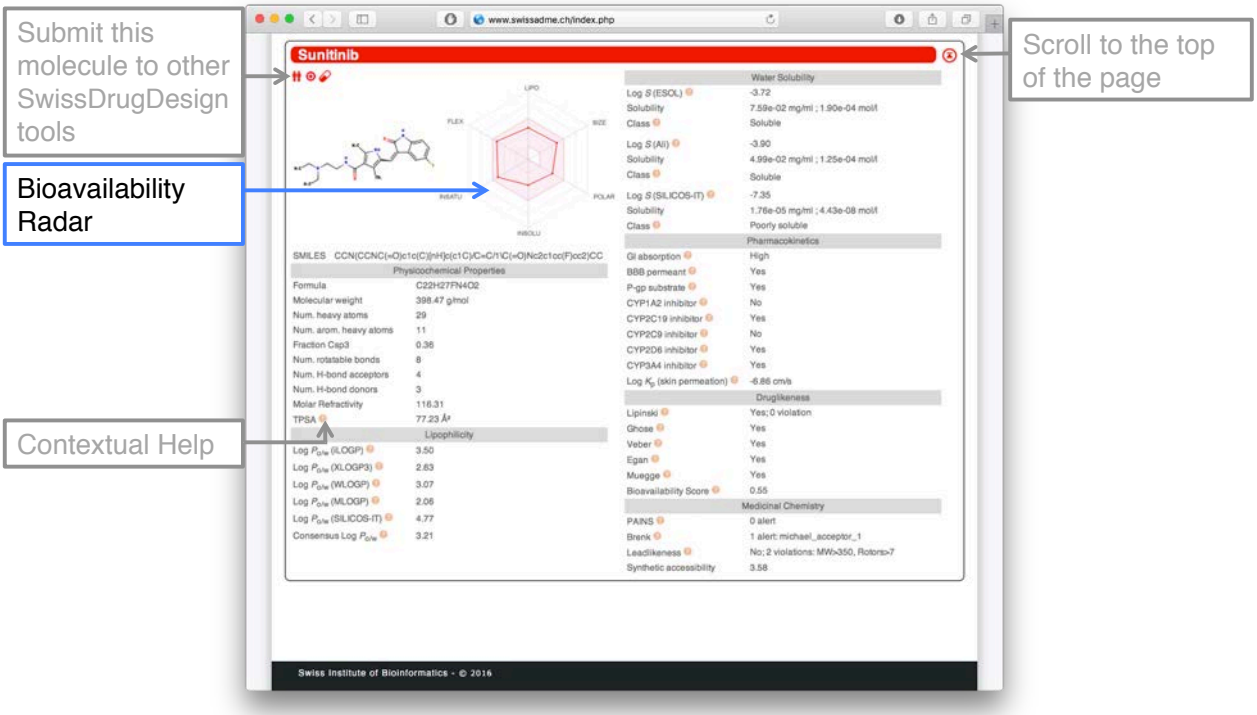


1. Scroll to the top of the SwissADME page and in the SMILES list box, type the name of the compound "CHEMBL598797" next to the SMILES, separated by a **space**. Then press **enter** to go to the next line.
2. Click on the "**Import**" button of the sketcher (second button top-left).
3. Type "Erlotinib"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

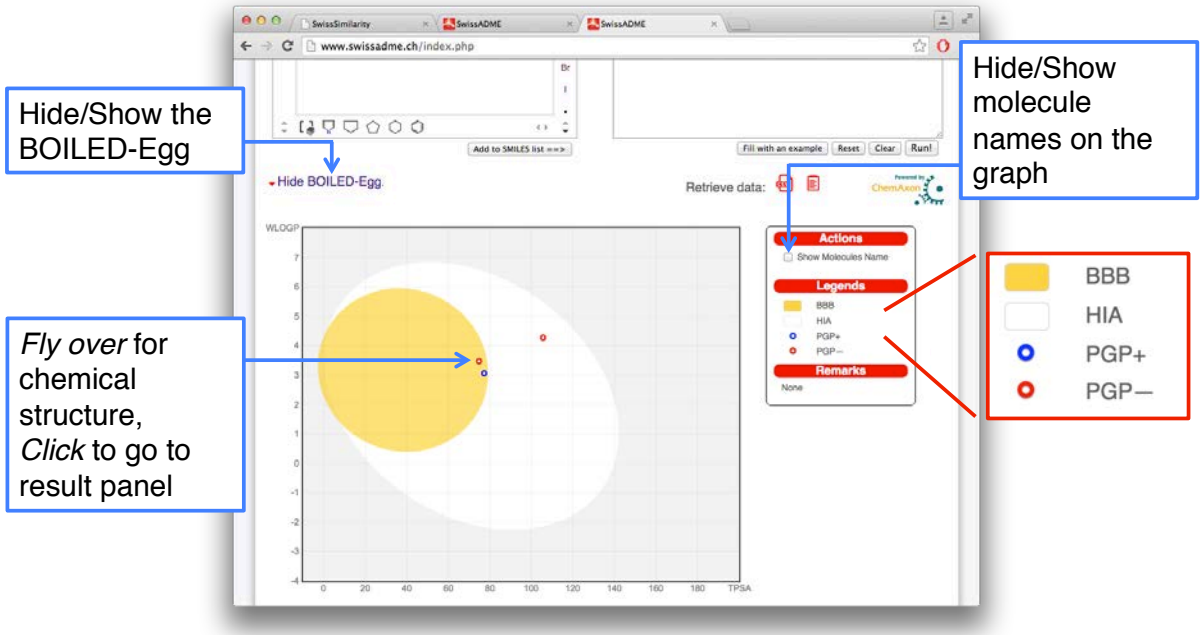
- Repeat steps 2, 3, 4 and 5 for Sunitinib (*no matter if there are blank lines in the text box*)

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 to be 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.

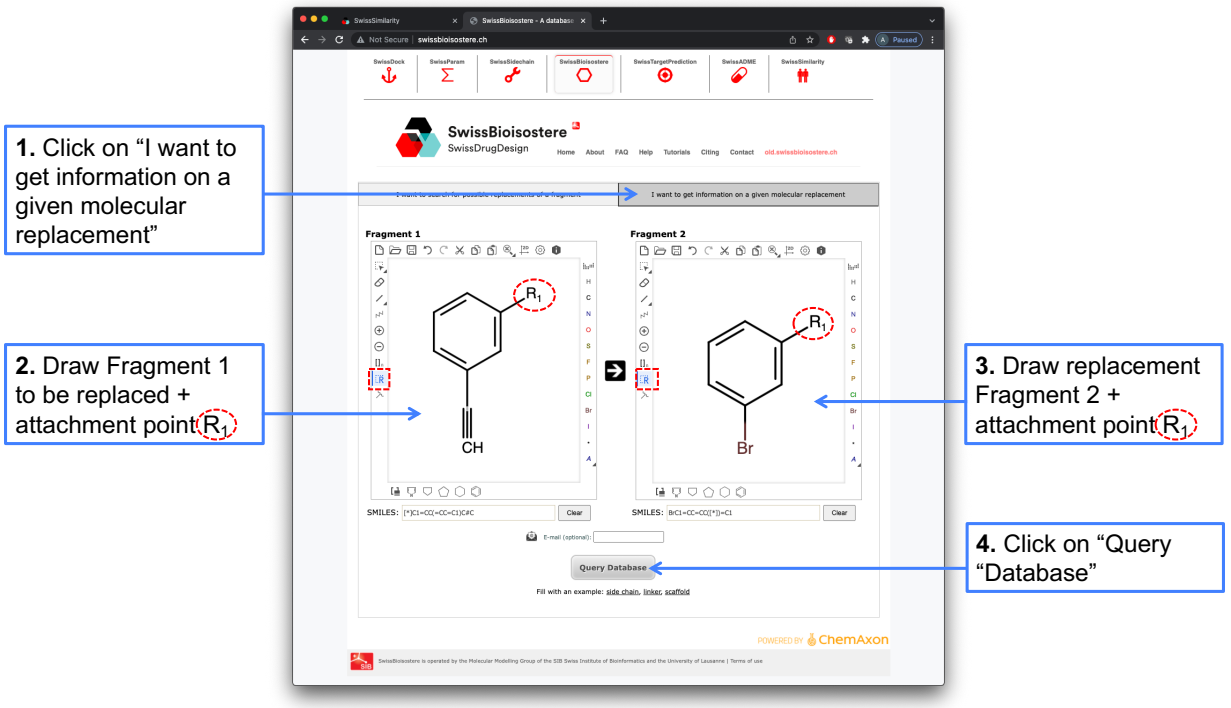
- Imagine that your endeavor consists in optimizing the properties of CHEMBL598797, which has to inhibit a kinase located in the central nervous system (CNS). Try some small chemical modifications (e.g. copy/paste SMILES in the sketcher, apply chemical modifications and transfer multiple entry lines 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

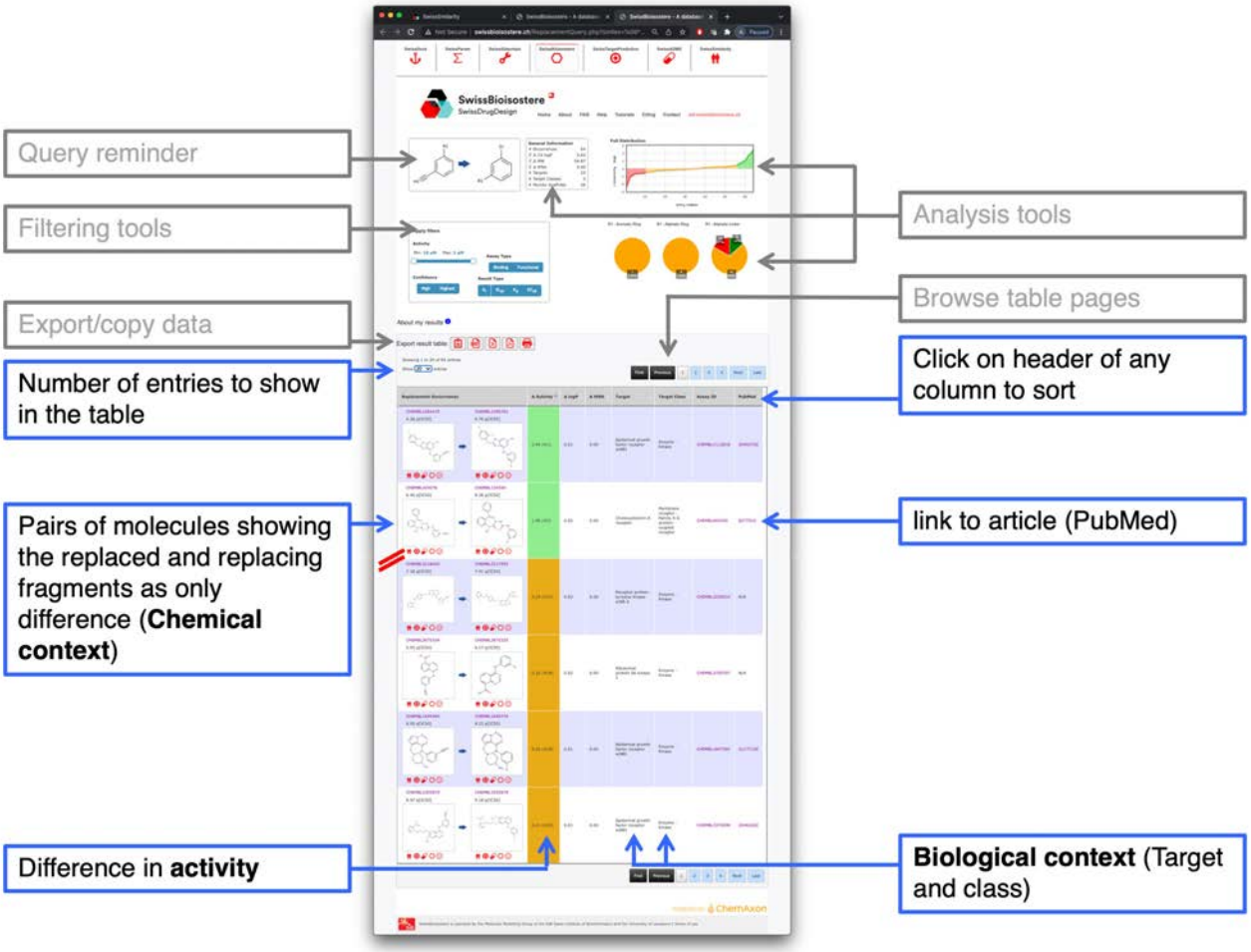
The purpose of this session is twofold. First it is proposed to analyze a specific replacement found in the virtual screening. Then we will suggest other possible bioisosteric replacements for ligand design.
Please note, detailed video tutorials are available at: http://swissbioisostere.ch/tutorials.html

- Go back to the **SwissSimilarity result page** in your web browser, locate compound CHEMBL2087361 (rank 3) and compound CHEMBL2087355 (rank 13). *Can you point out 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. Click on the "**I want to get information on a given molecular replacement**" grey tab. A second sketcher "Fragment 2" appears on the right.
2. Draw in the *left* sketcher the fragment in CHEMBL2087361 that is replaced. Add the attachment point (R_1) with "smart R-group" in the left tool bar (numbering is automatic).
3. Draw the replacing fragment in CHEMBL2087355 in the *right* sketcher. Pay attention to add the same attachment point (R_1).
4. Start the search by clicking the button "**Query Database**" at the bottom of the page.



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

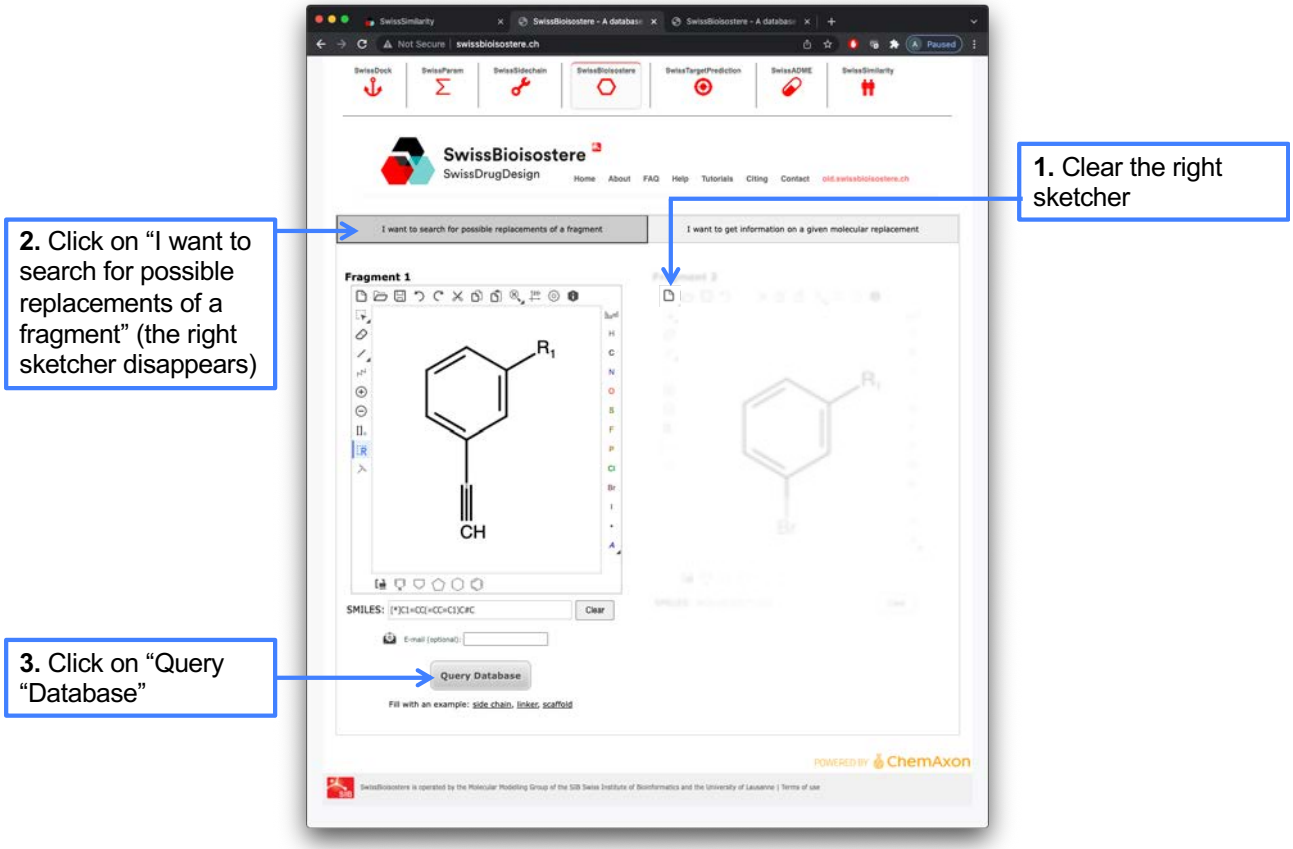


📣 Display all lines by setting "**All**" in the number of **entries to show** on the upper left corner of the result table. Let's try to answer the following questions:

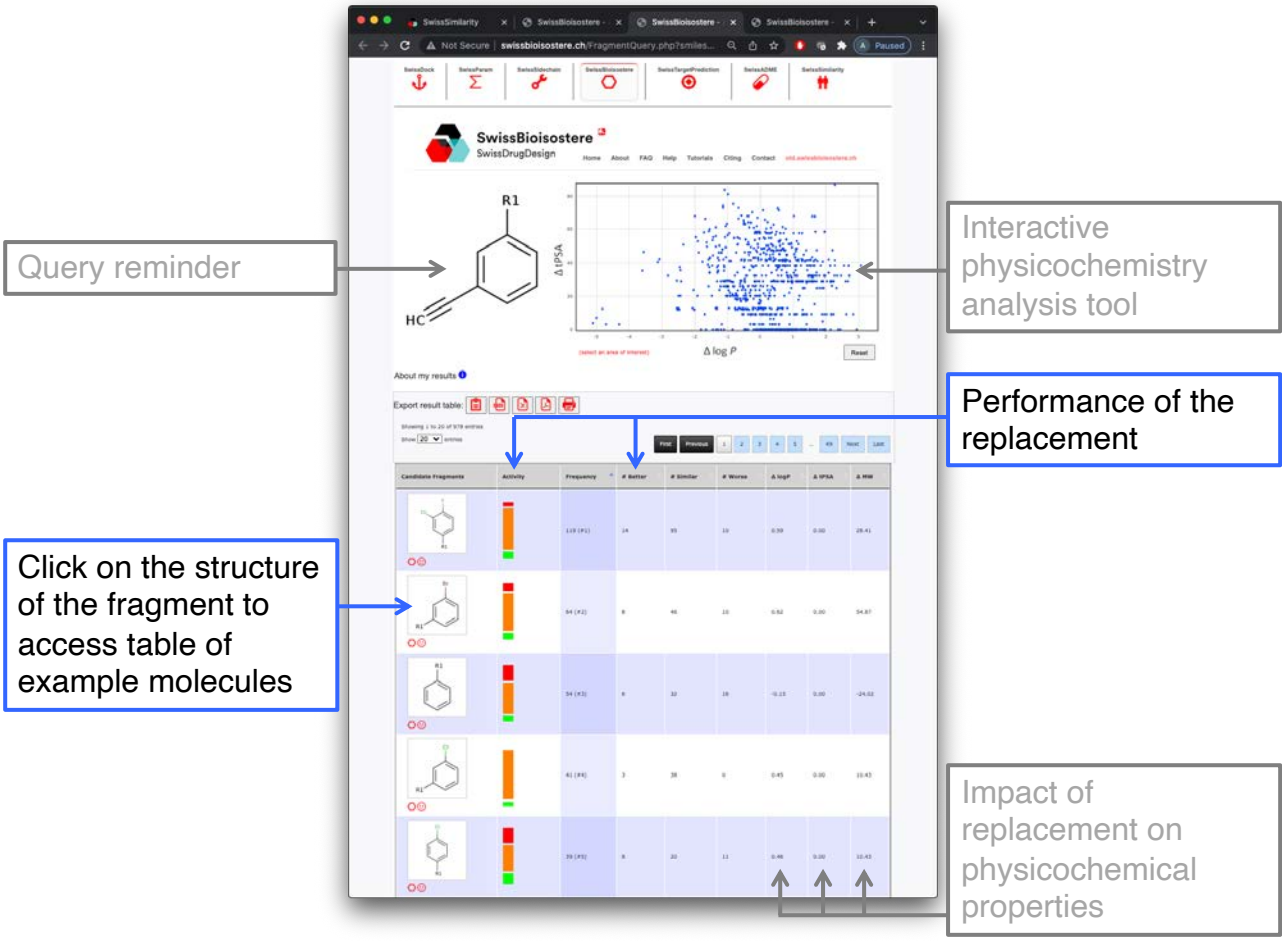
- *Overall, how many times this specific replacement was found in the literature?*
- *Generally speaking, what is the trend: increasing or decreasing or similar biological activity?*
- *In what biological context this replacement was mainly tried?*
- *How many times this replacement was found for compounds tested on our target of interest (Epidermal growth factor receptor erbB1)?*
- *What is the trend for activity on this specific protein?*
- *Can you find the entry corresponding to our case (CHEMBL2087361 to CHEMBL2087355)?*
- *Which molecule is the most potent on erbB1 among 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. **Go back to the tab** of your browser where you have made the last SwissBioisostere request.

1. Clear the *right* sketcher by clicking on its most upper-left button.
2. Click on the "**I want to search for possible replacements of a fragment**" grey tab. The right sketcher disappears.
3. Verify that the *m*-ethynylbenzene is correctly drawn in the left sketcher and 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 in a new tab.



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

- *Can you find the replacement that we studied in detail in the previous section (CHEMBL2087361 to CHEMBL2087355)?*
- *Propose two other sensible replacements for* m*-ethynylbenzene in our biological/chemical contexts and explain how you found them.*

The workshop is finished. We hope you have enjoyed it !!!