Getting started with ChimeraX

The objective of this little exercise is to get you acquainted with ChimeraX.

Authors & contributors of this work:
Françoise Ochsenbein, B3S/AMIG, CEA/I2BC
Raphaël Guérois, B3S/AMIG, CEA/I2BC
Thibault Tubiana, B3S/IMAPP, CEA/I2BC
Chloé Quignot, PF/BIOI2, CEA/I2BC

Summary

Step 1 – Before you start

ChimeraX is the new version of USCFChimera and is free (for academics). It also has interesting features enabling interactive analysis of AlphaFold's outputs.

>Install ChimeraX:

=> Downloadable from their website for any OS (Windows, Linux, Mac):
https://www.cgl.ucsf.edu/chimerax/download.html

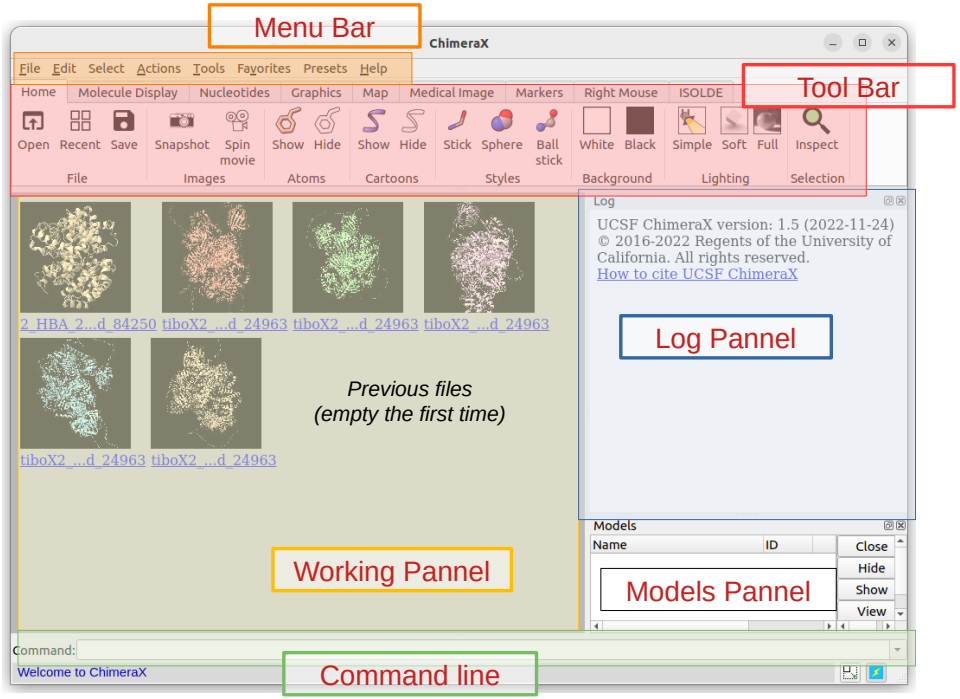
NB: CEA and I2BC members should have ChimeraX accessible within the app store

>Recommendations:

- Using an **external mouse** can make manipulating your structure(s) within ChimeraX easier
- ChimeraX **help pages** are very complete:
  https://www.rbvi.ucsf.edu/chimerax/docs/user/index.html

Step 2 – Open ChimeraX

The ChimeraX window is composed of the **Menu Bar** (*url*) & **Tool Bar** (*url*) across the top, the **command line** (*url*) area at the bottom and several detachable and moveable pannels in between, of which:



Log Pannel:
Any action is recorded here and clickable to open its respective help page

Models Pannel:
All objects (e.g. molecules) are listed here and are given an ID *(url)*

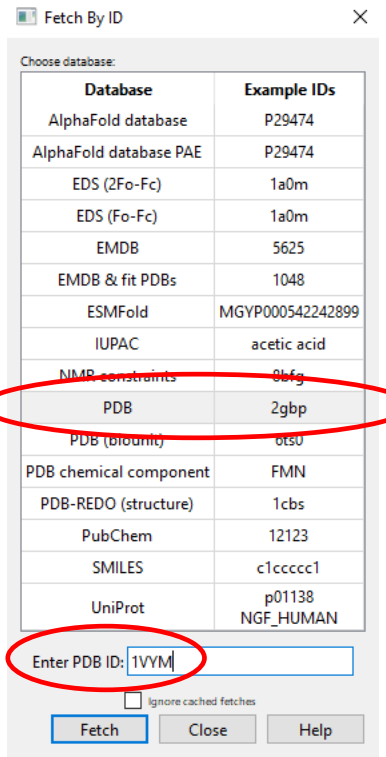
Working Pannel:
Here's where you'll visualise your objects *(url)*

Step 3 – Upload a structure

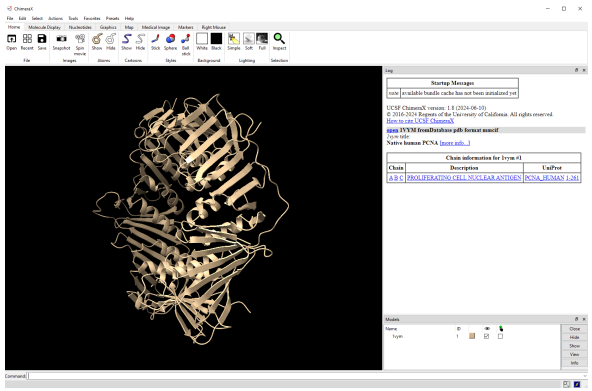
You can open a file directly via the `File > Open` menu or fetch from the Protein Data Bank using the ID of an entity:

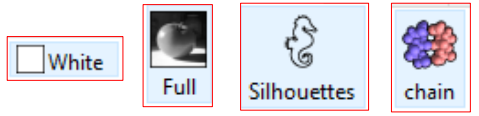
1. `File > Fetch By ID`

PDB ID : 1VYM

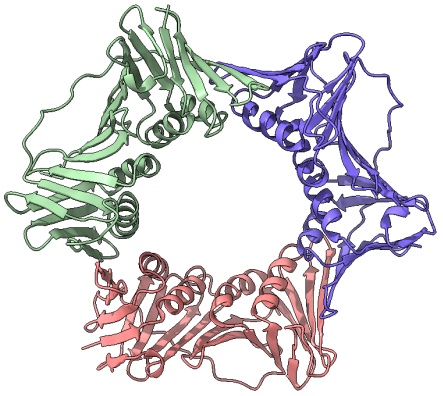


Automatically fetches the PDB coordinates from the reference PDB database https://www.rcsb.org/



2. Find the tools : 

to switch it to the representation below



Question 1: How many different proteins ?

Question 2: How would you call such a protein assembly ?

a. Homotrimer
b. Heterotrimer
c. Monomer
d. Polymer

Step 4 – Mouse commands and selection

→ *Play around with the following commands to get familiar with them & check they work for you*

- **Move your structures around** in the Working pane:

  - Left click + hold **on the background or protein** to **rotate** in 2 different ways
  - Central click + hold **on the protein(s)** to **translate** (or Left click + Shift)
  - Scroll to **zoom** in & out
  - Right click : Depends on what's selected in the « Right Mouse » tab in the Toolbar (default : translate)

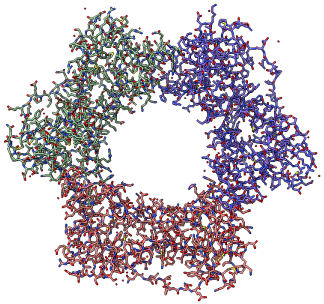
- **Select & deselect regions of your structure** in the Working pane ***(Selections are highlighted in green)***:

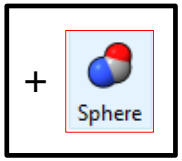
  - Ctrl + Left click + drag over an area of your protein to **select** this region
  - Ctrl + Left click on a specific residue/element to **select** it
  - Ctrl + Shift + Left click/drag over an area/residue to **add to or remove from the current selection** depending if the area/residue is already in the selection or not
  - Ctrl + Left click on the background to **clear the current selection**

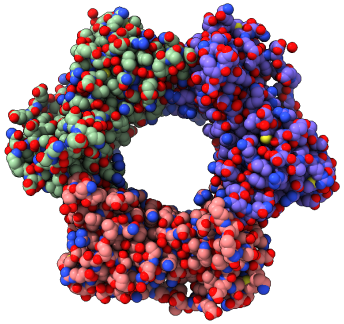
  Use Up/Down arrow keys to expand the current selection to the lower/higher resolution level (e.g. from a residue to the whole secondary structure it is contained in)

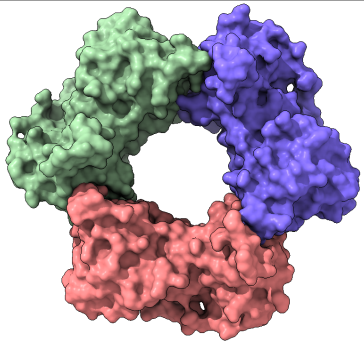
Step 5 – Changing representations

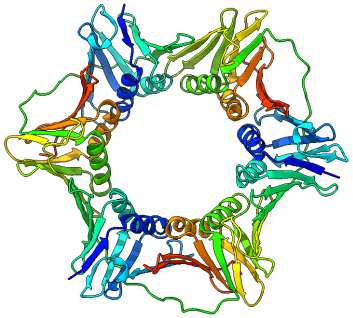
> Depending on what you want to look at/analyse, different representations will help to highlight different aspects of your molecule(s)



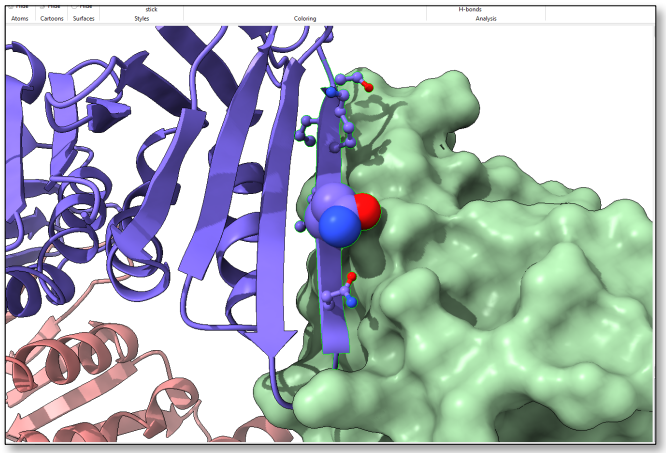








Question 3: Combining Selections and Representations actions, obtain the following mixed representation in your Working pannel :

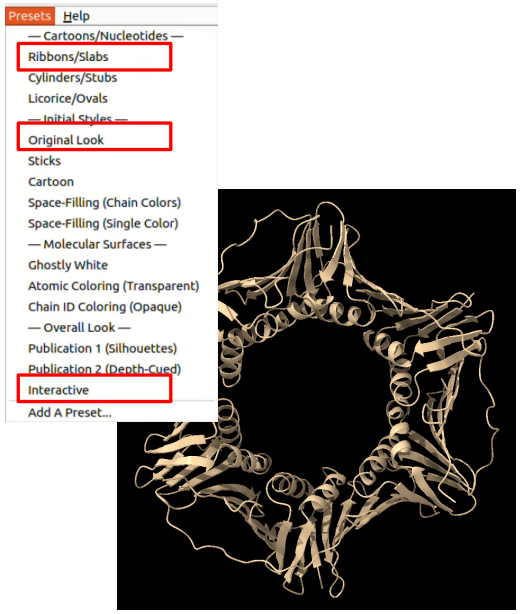


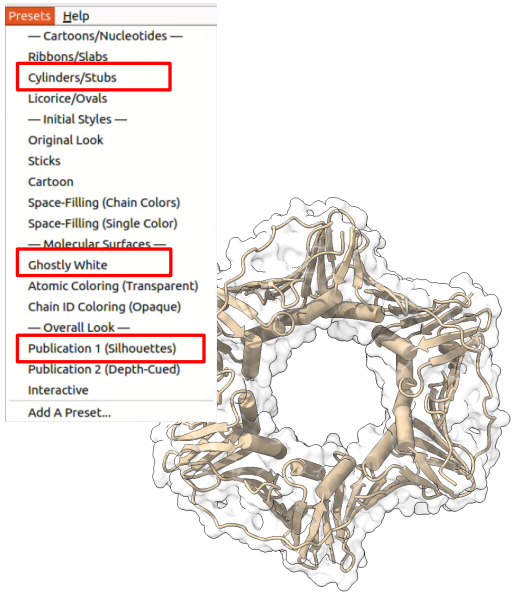
Hints :

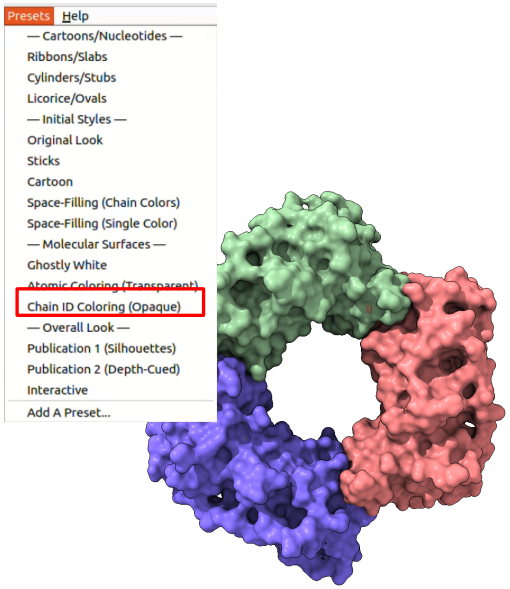
- Colour by chain
- Chains A & B cartoon, Chain C as surface
- Show atoms in Beta sheet of Chain A closest to Chain C
- Show residue 179 as spheres

Step 5 – Changing representations

You can also play around and combine Presets in the top Menu bar… or create your own preset(s)





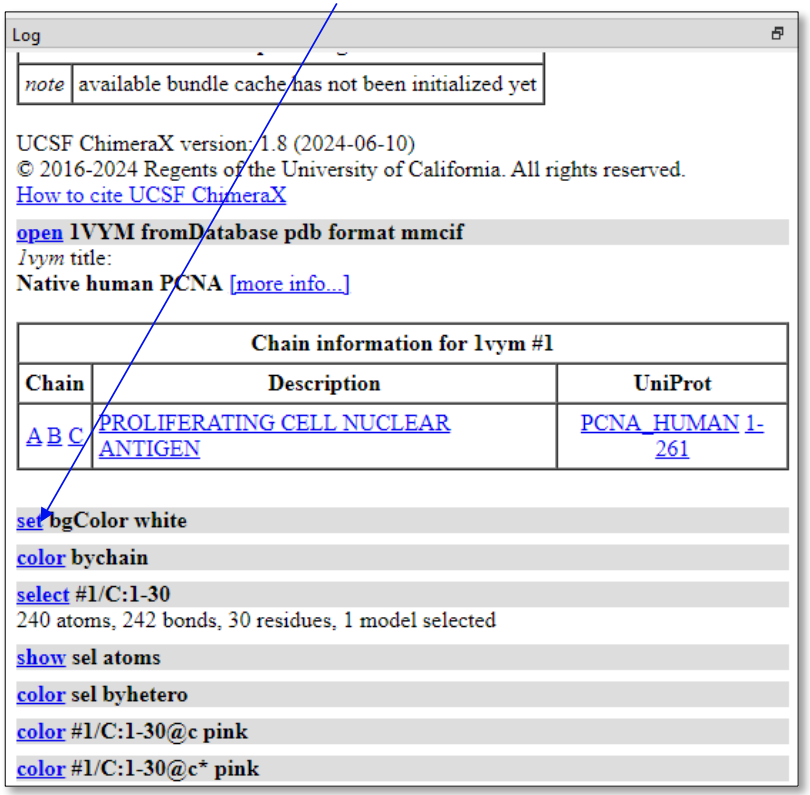


Step 6 – Command line usage and log

You don't have to use the command line in ChimeraX for basic functions but it offers more options and is quite easy to master, especially thanks to the logs

> All the commands are reported in the Log pane
those coming from clicking or typing, making it easier to transition to the command line

Link to the ChimeraX html help pages



Command syntax :

> command + option not needing a value
color bychain
command name option name

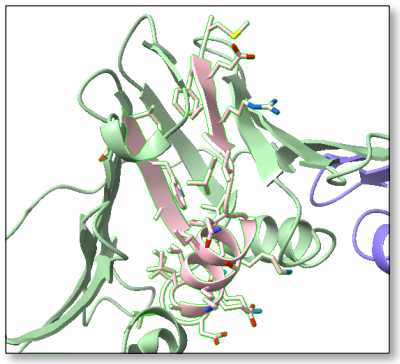
> command + option needing a value
set bgColor white
command name option value option name

To know the options available => click hyperlink to open help page

Step 6 – Command line usage and log

Question 4: Follow the steps below using the **select**, **show**, **color** and **view** commands in the command line area:

1) Reinitialise the view as in **Step 3**
2) Command: select #1/C:1-30
3) Command: show sel atoms
4) Command: color sel byhetero
5) Command: color #1/C:1-30@c pink
6) Command: color #1/C:1-30@c* pink
7) Command: view sel

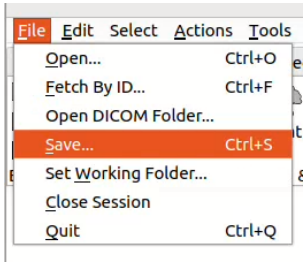


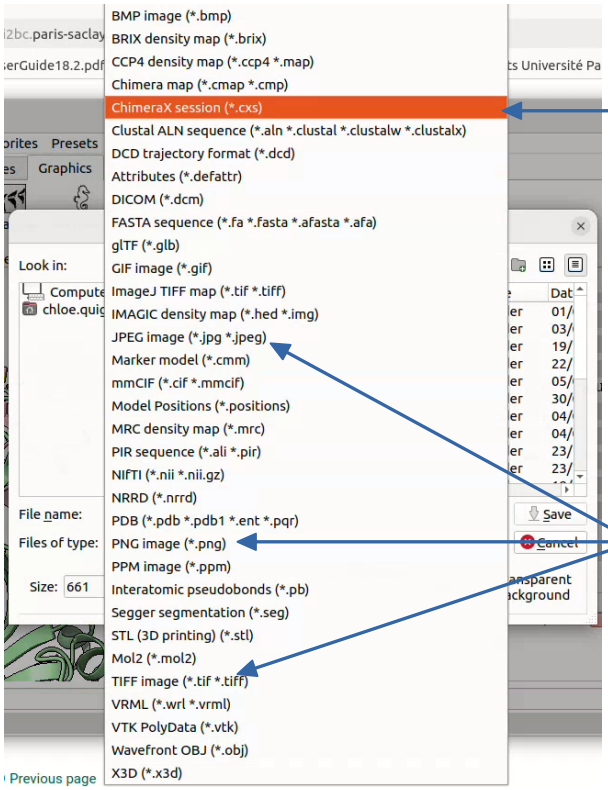
Selection syntax for elements in your structure :

Levels	Models	Chains	Residues	Atoms
Symbol	#	/	:	@
Usage examples	**#1** = model 1 **#1,5** = model 1 & 5	**/A** = all chains A **#1/A,B** = chains A & B of model 1 only	**:1-25** = all residues with numbers 1 to 25 **:tyr** = all Tyrosines **#1/A:1-30&:tyr** = all Tyrosine residues with numbers between 1 and 25 and who belong to chain A of model 1	**@c** = only C of the peptide backbone **#1@c*** = wildcard * is used to allow for all carbons of model 1

Step 7 – How to save your work?

Through the File menu in the Menu Bar, then choose your format :





ChimeraX session

Image of your current view

Or through the Command line :

To save your session :

```
Command : save  mypath/mysession.cxs
```

To capture high quality images :

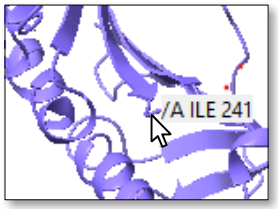
```
Command : save  mypath/mypicture.png width 3000 supersample 5
```

Step 8 – Sequence analysis

There are different ways of linking « sequence » information (i.e. chain name, residue type, residue number) and protein structure :

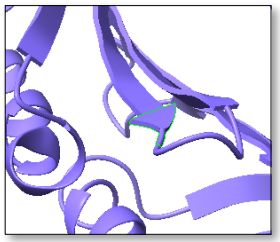
- **Method 1: Mouse**

Hover the mouse over structural elements for >1 sec



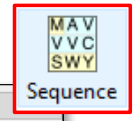
- **Method 2: Command line**

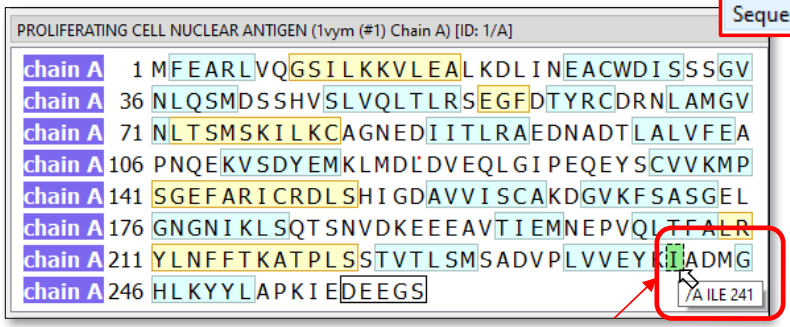
```
Command: select #1/A:241
```



- **Method 3: Sequence Pannel**

Tools > Sequence > Show sequence viewer *or*

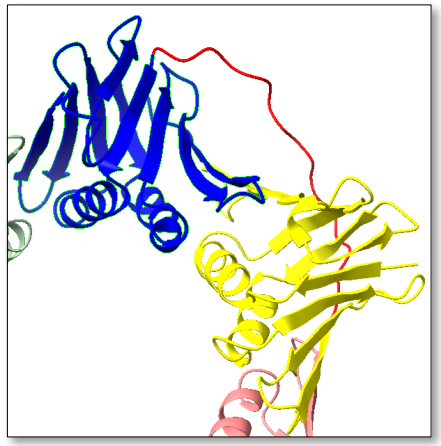




Residues can be selected with mouse Left click + drag

Question 5: In method 3, the sequence pane contains yellow, blue and white boxes. What do they correspond to ? What about residues with no box ?

- ?
- ?
- ?



Question 6: From the sequence pannel, identify the longest loop. What is the index of its 1st and last residue ?

Color the longest loop red

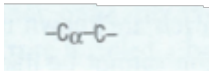
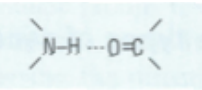
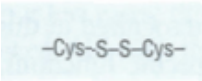
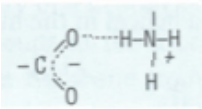
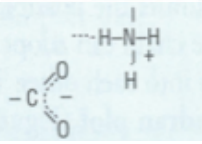
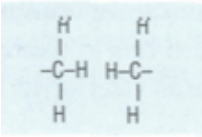
Select all the residues before the longest loop and color them yellow.

Select all the residues after the longest loop and color them blue.

Question 7: Compare the yellow Nter and the blue Cter domains. See anything striking ?

Step 9 – Hydrogen bond networks

> Many chemical interactions stabilise proteins (polypeptides)...

Interaction	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)	
Covalent bond	-	1.5Å	356kJ/mole (610kJ/mole for a C=C bond)	
Hydrogen bond	Donor (here N), and acceptor (here O) atoms <3.5Å	3.0Å	2-6kJ/mole in water 12.5-21kJ/mole if either donor or acceptor is charged	
Disulfide bond	-	2.2Å	167kJ/mole	
Salt bridge	Donor (here N), and acceptor (here O) atoms <3.5Å	2.8Å	12.5-17kJ/mole may be as high as 30kJ/mole for fully or partially buried salt bridges, less if the salt bridge is external	
Long-range electrostatic interaction	Depends on dielectric constant of medium. Screened by water. 1/r dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water	
Van der Waals interaction	Short range. Falls off rapidly beyond 4Å separation. 1/r^6 dependence	3.5Å	4kJ/mole (4-17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)	

*...of which **hydrogen bonds :** can be formed between different types of atoms in proteins :*

- Between mainchain atoms (peptidic backbone atoms) ➔ bases of helices and strands (=secondary structure)
- Between sidechain and mainchain atoms ➔ frequent in turns
- Between sidechain atoms ➔ frequent at the surface of proteins
- With water molecules ➔ some of them can be observed in Xray structures

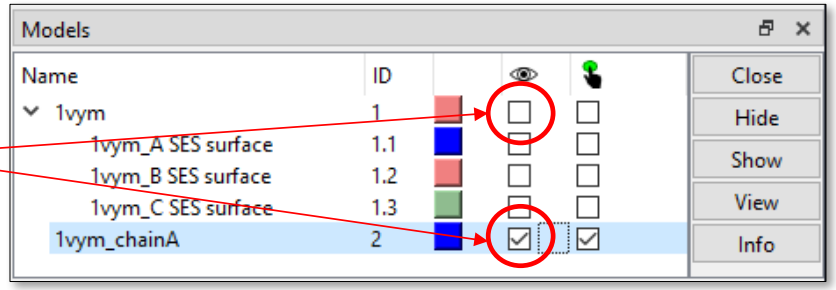
Step 9 – Hydrogen bond networks

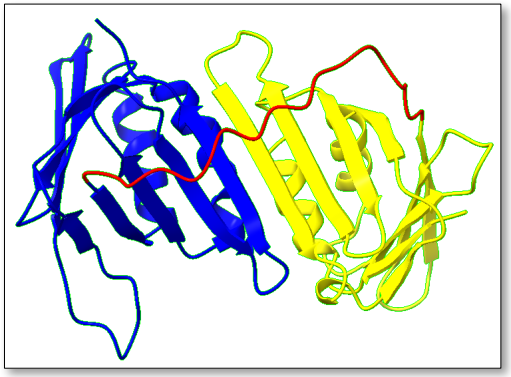
*1. Let's **focus on a single chain** (chain A) in our protein (1VYM aka. object/model #1 in the models pannel)*

The « `combine` » command can be used to duplicate model #1 into a second model (ID : #2, name : 1vym_chainA).
The « `delete` » command is used to remove chains /B and /C from model #2

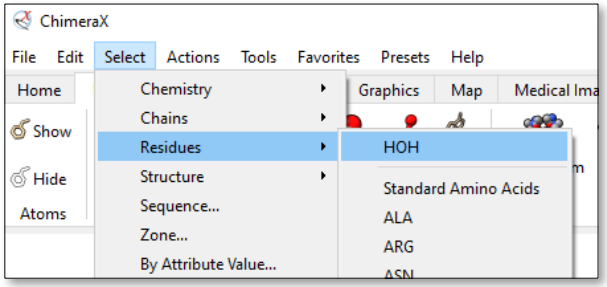
```
Command: combine #1 name 1vym_chainA
Command: delete #2/B,C
```

Toggle on and off the visualisation boxes to view only the new model #2 containing only chain /A



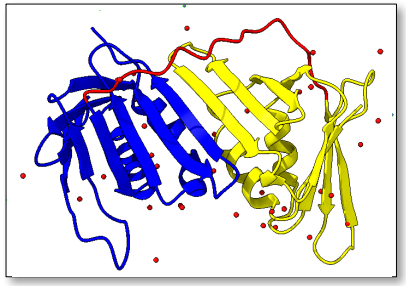


*2. **Display the water molecules** using the following :*



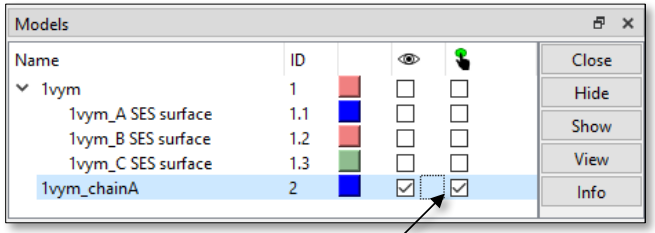
+

=



Step 9 – Hydrogen bond networks

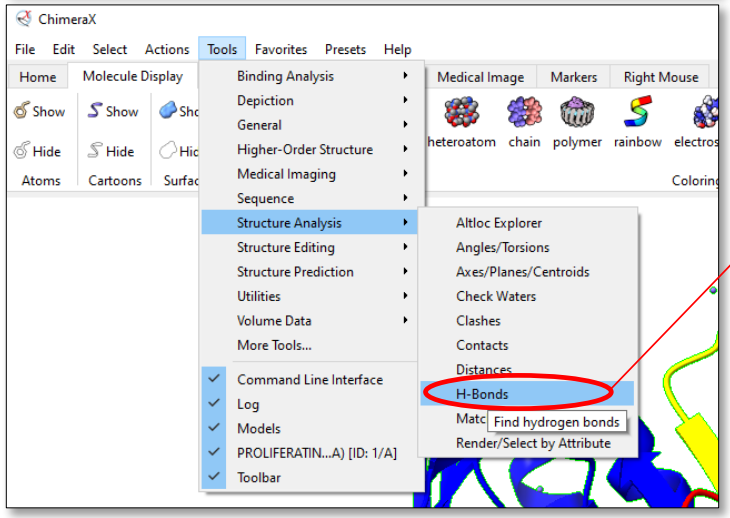
*3. **Detect hydrogen bonds** with the following instructions :*

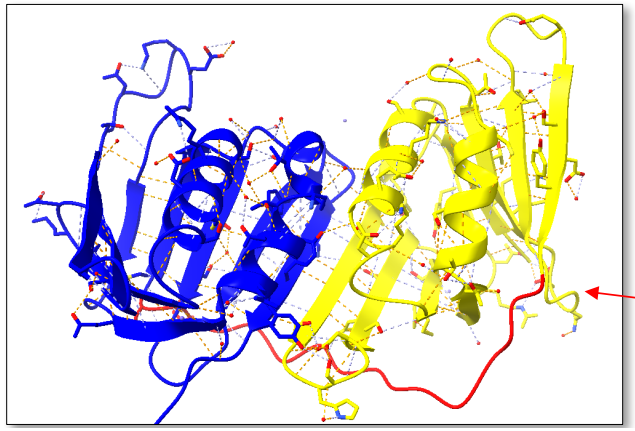


Select only Model #2

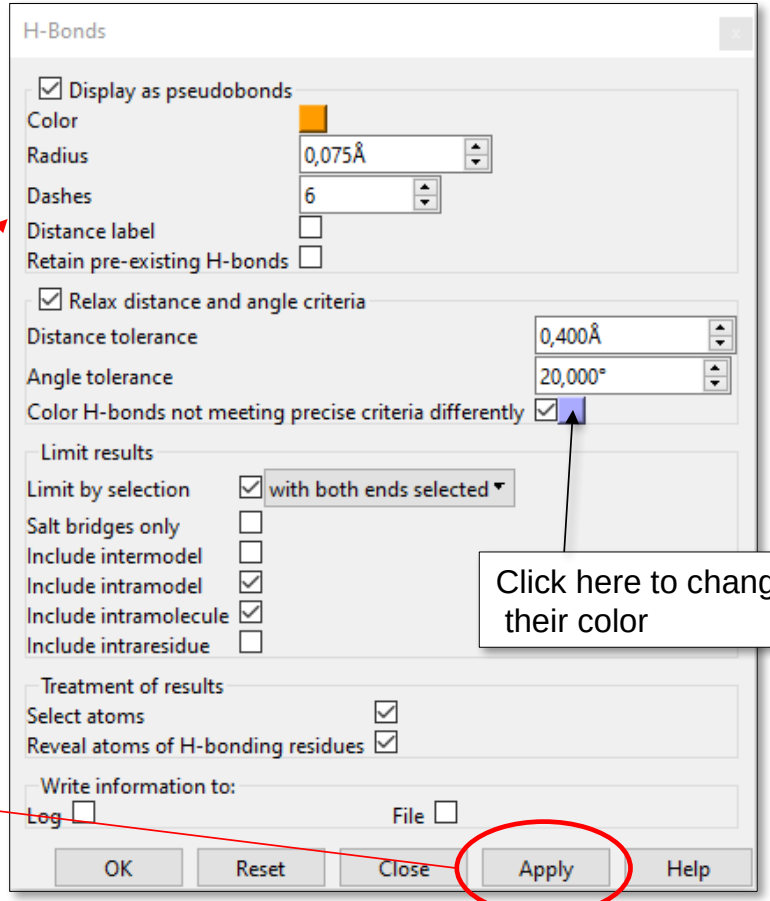
can be also done with command line

Open the Menu
Tools -> Structure Analysis -> H-Bonds :





The H-Bonds pannel can be moved around or inserted on top of the Log pane (access through tabs)



Click here to change their color

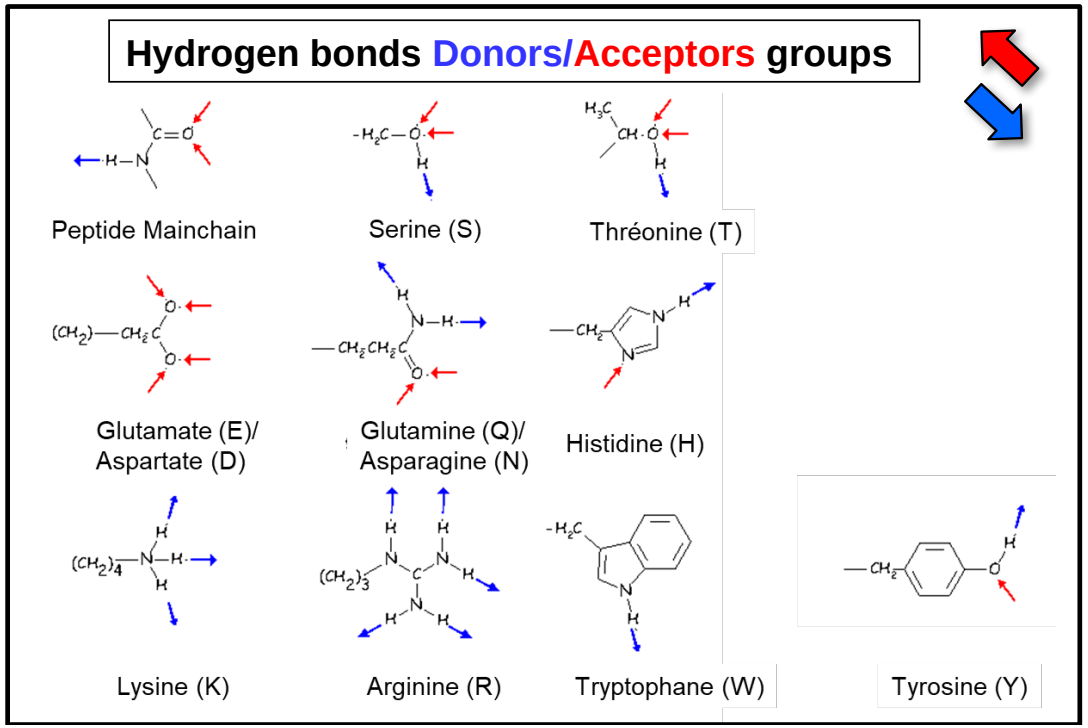
Color N and O atoms with + heteroatom

Step 9 – Hydrogen bond networks

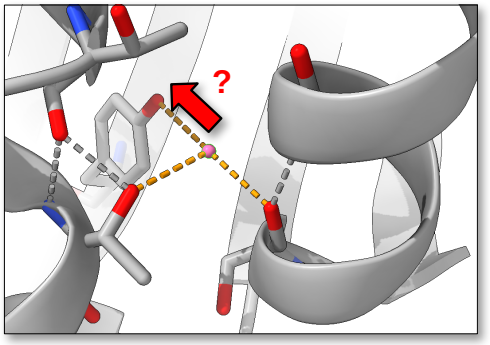
Question 8: Inspect the hydrogen-bond network. The Log provides you with some information.
Among the 167 strict hydrogen bonds found, how many are buried in the core of the proteins? Is this expected?

Question 9: Identify the single water molecule buried in the core. How is it connected to the protein sidechains?
What are the index of the amino-acids it is hybrogen-bonded to?

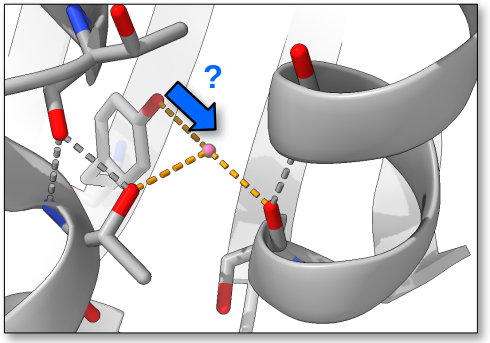
Question 10: With X-ray, we generally can't see hydrogen electronic densities except with very high resolutions.
Using the molecular rules provided below, is the Tyrosine (A) Acceptor or (B) Donor of the pink water molecule ?



(A) **Tyr Acceptor ?**



(B) **Tyr Donor ?**



Step 10 – Let’s dive into the protein core

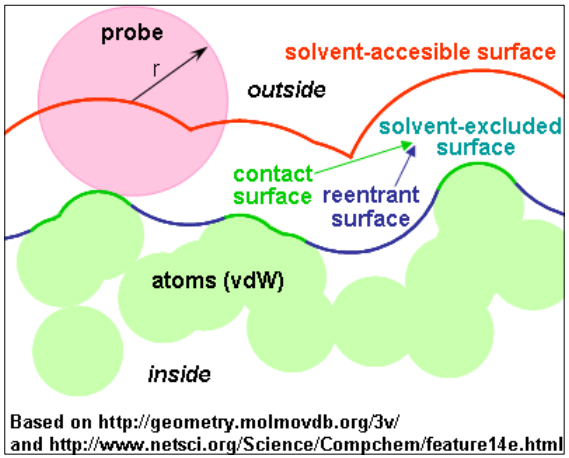
First, we need to select all the amino acids buried in the protein.

Let’s work back on model #1 and delete solvent molecules

```
Command: delete solvent & #1
```

Use the command « measure », combined with the parameter « sasa » for solvent-accessible surface area

```
Command: measure sasa #1
```



➔ The solvent-accessible surface of every residue is stored as an « Attribute » property

List in the Log pane all the sasa values with the command « info »:

```
Command: info residues #1 attribute area
```

info residues #1 attribute area
residue id #1/A:1 area 30.18797279072163 index 0
residue id #1/A:2 area 1.0646300099787567 index 1
residue id #1/A:3 area 47.8403790261102 index 2
residue id #1/A:4 area 0.3419030365784721 index 3
residue id #1/A:5 area 84.01697544772601 index 4
residue id #1/A:6 area 3.4891860678588102 index 5
residue id #1/A:7 area 56.43023147636802 index 6
residue id #1/A:8 area 96.44069189106486 index 7
residue id #1/A:9 area 0.03742603620693785 index 8
residue id #1/A:10 area 22.052144377305652 index 9

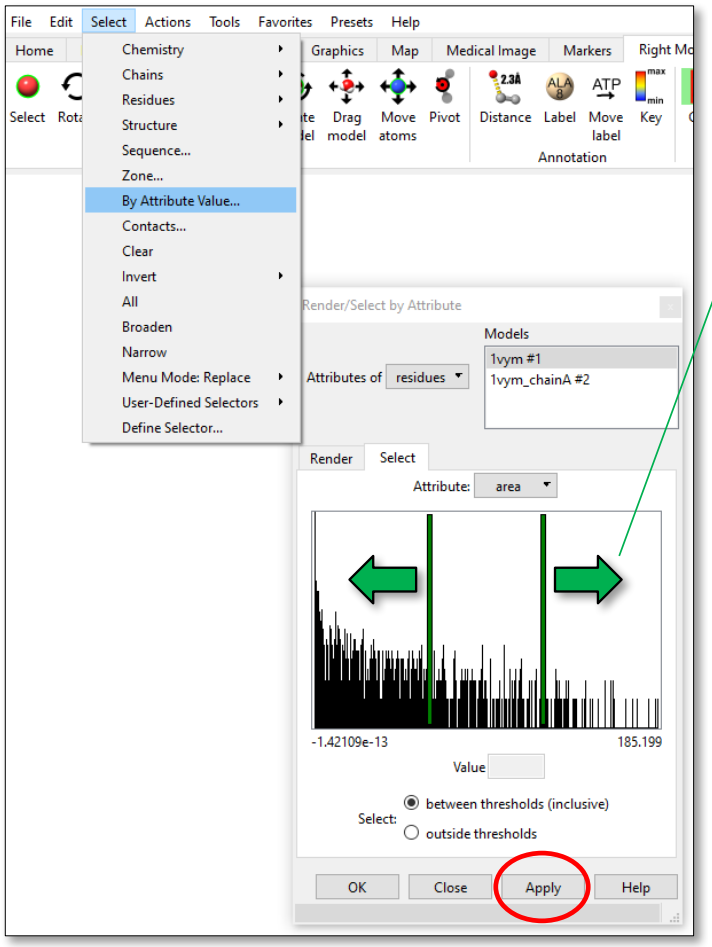
Buried

Solvent exposed

Step 10 – Let’s dive into the protein core

Now, we can use the menu `Select -> By Residue Attribute ->`

Open the window to define the thresholds for selection based on the Attribute « area ».

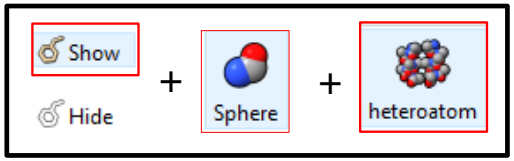


Select the value of solvent accessibility up to ~15 Å^2:

In the Log you can see the command line to use and refine the threshold you wish:

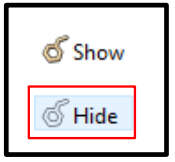
```
Command: select #1::area>=-1 & ::area<=15.0
```

*The attribute is accessed with the **::** symbol. Here we use the « **area** » attribute*




```
Command: select backbone & #1
```

+




```
Command: select #1
```

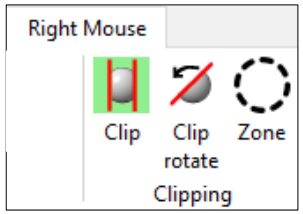
+

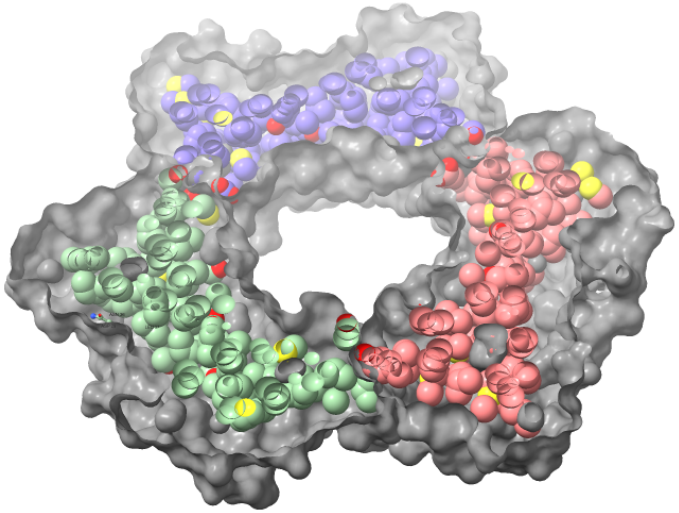

```
Command: color #1 gray surface
```

Step 10 – Let’s dive into the protein core

You can clip the 3D structures in order to analyse their interior.

You can play with tools located in the « right mouse » menu:





Fraction of space occupied by atoms in different media

Water	0.36
Cyclohexane	0.44
Ethanol	0.47
Compacted Spheres	0.71 (NaCl cristal)

PS : You might need to run an extra command in order to see the interior of the volume like in this picture.
-> you can try : `surface cap false`

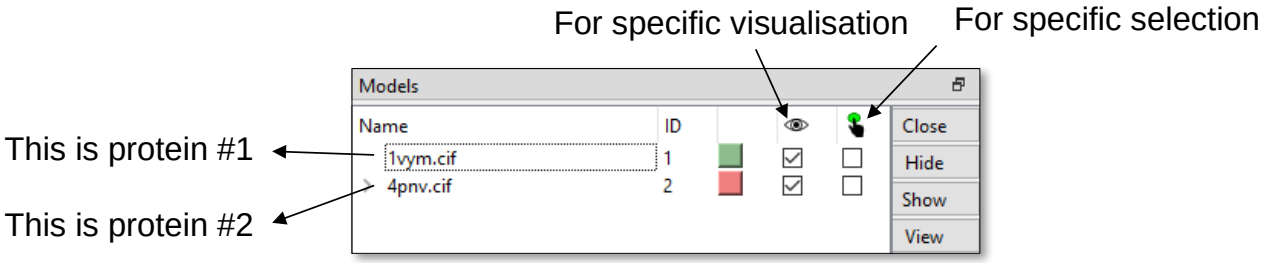
Question 11: Visualise the interior and appreciate the level of compacity within the core. From the table above, guess what is the average fraction of space occupied by atoms in protein cores ?

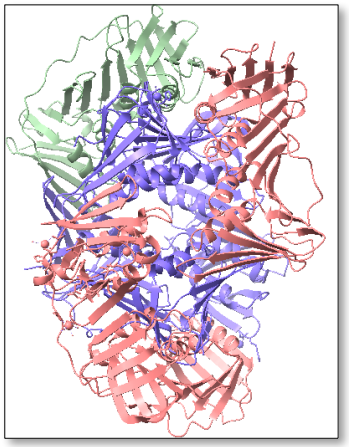
Question 12: How many polar atoms can you recognise in the core in one subunit ? Does this seem favoured ?

Step 11 – Evolution of protein structures

The PCNA ring architecture exists in all eukaryotes but also in prokaryotes. Let's compare the structures between the Human and the *E. coli* homologs.

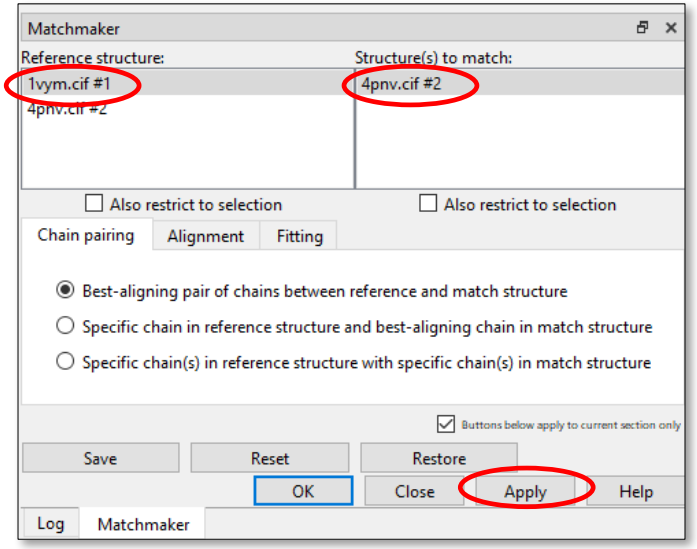
1. Load the structure of the E. coli homologue **4PNV**





Messy ! We need to superimpose both structures !

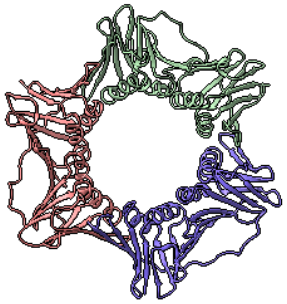
Open the Menu
Tools -> Structure Analysis -> MatchMaker :



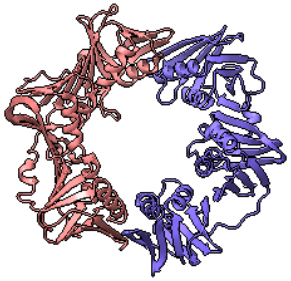
Step 11 – Evolution of protein structures

It might help to get a side by side view of both proteins : Command: **`tile`**

To undo this « tile » organisation : Command: **`~tile`**



Eukaryotic



Prokaryotic

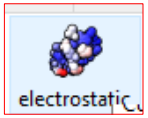
Question 13: What is the major difference you can notice between the human and the *E. coli* systems?
Hint : Show both in cartoon representation only & colour them both by chain

Question 14: What evolutionary process could have given rise to such differences?
Could you imagine what the composition of the ancestral complex would've been?

Step 11 – Evolution of protein structures

Electrostatic forces are crucial for specific molecular interactions and molecular assembly.
Let's compute the electrostatics potential at the surface of the two protein systems.

A quick and dirty mapping can be obtained with the menu button :

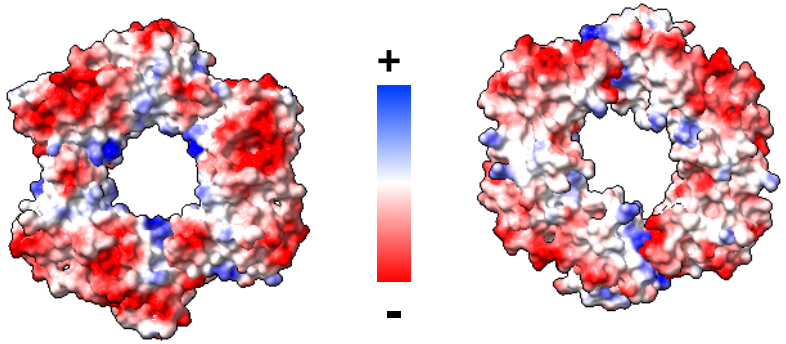


Thresholds can be tuned with the palette keyword :

```
Command: coulombic #1#2 palette -5,red:0,white:5,blue key true
```

Thresholds values can be be changed

To show & change colour scale



Eukaryotic Prokaryotic

Question 15: What common property has been maintained in evolution at the center of the ring ? What are the amino acid types responsible for this property ? What could be the other macromolecule likely to interact with such a ring ?

Open the 6GIS PDB structure on top of the human PCNA structure. Superimpose with the structure 1VYM and analyse the contacts with the partners.

Question 16: What can you say about these contacts ?