Theoretical and practical course in protein biochemistry, biophysics and structural biology: Computer lab handbook

Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos

Faculdade de Medicina de Ribeirão Preto

20 to 31 March 2023

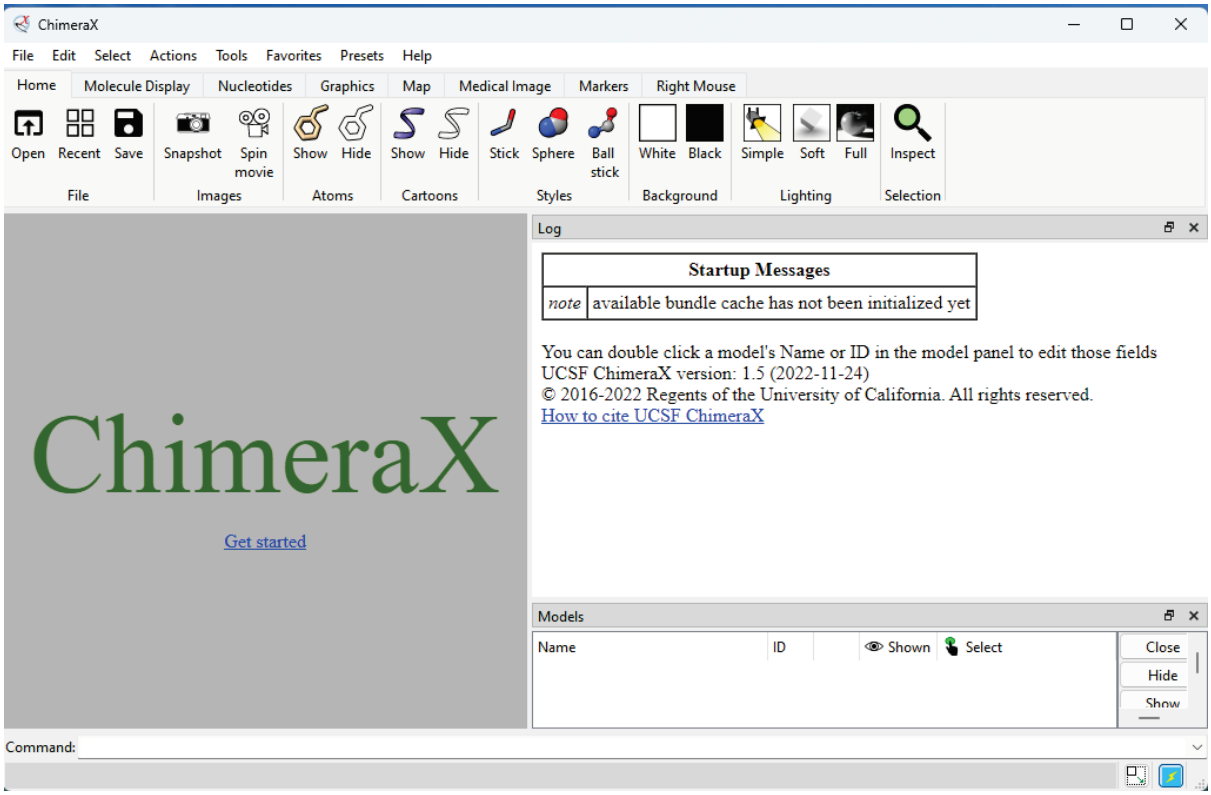
Computer lab 1 – Visualising proteins

Overview

Today we will have a look at the structures of the proteins we've been working with this week. We will inspect the structures using the programme ChimeraX[1], which you can download from https://www.cgl.ucsf.edu/chimerax/download.html.

Visualising EGFP

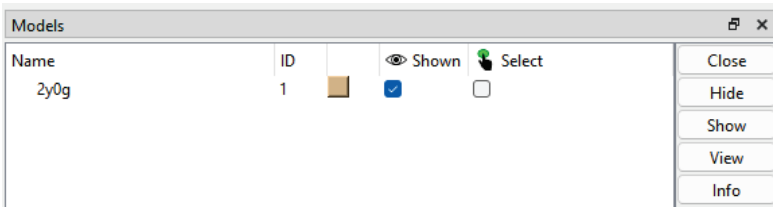
Open ChimeraX[2]



Enter the following command in the command bar to download the structure of EGFP from the Protein Data Bank:

```
open 2y0g
```

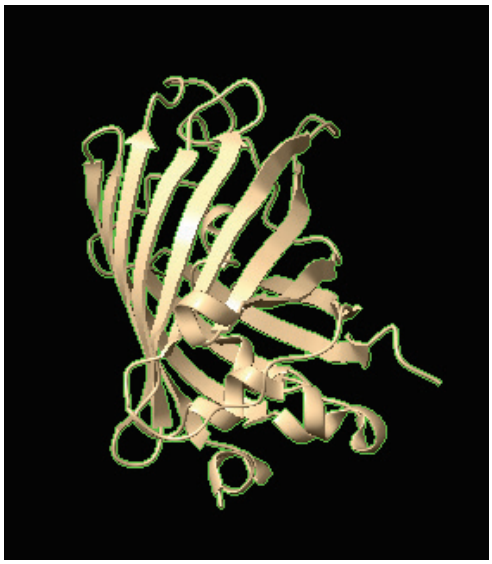
This has opened the protein with identifier #1, as you can see in the **Models** panel. You're currently looking at the protein in 'cartoon' view, showing the protein backbone:



[1] 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.

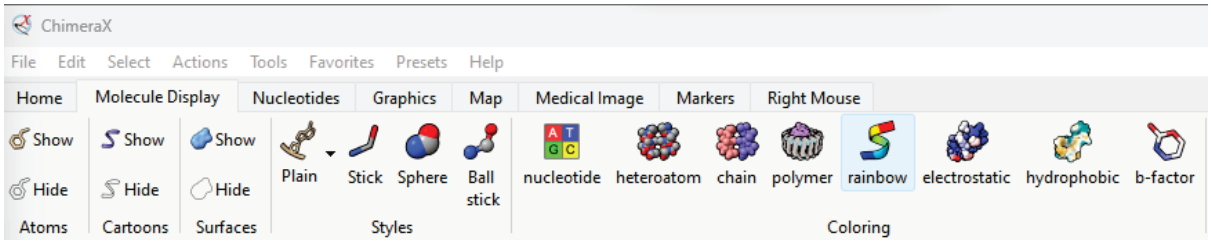
[2] This tutorial assumes you are running ChimeraX version 1.5

You can use the models panel to hide or view different structures or maps, and to select models. For example, try clicking on 'select'. You see that the whole model now has a green outline because it is selected.



Also, the command equivalent of what you did is shown in the **Log** panel:

Try colouring the protein backbone as a rainbow (blue at N terminus to red at C terminus) so that you can see the direction of the peptide chain. You can do this using the toolbar:

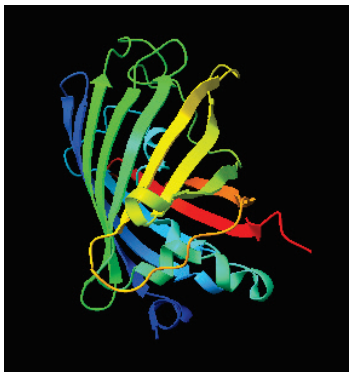


You can also see the command that was used in the log:

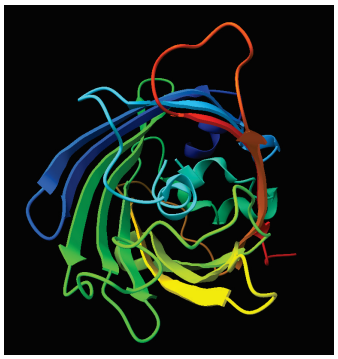
```
rainbow sel
```

And unselect everything (to get rid of the green outline)

```
select clear
```



You can use the left mouse button to rotate the view of the molecule, the middle mouse button to translate the view, and the scroll wheel to zoom in/out. Note that you're not moving the molecule, you're moving the 'camera' through which you're looking at the molecule (its position generally remains unchanged…more on that later). Try finding a view where you're looking down the middle of the beta barrel:



We might want to return to this viewing orientation later, so let's save the view:

```
view name barrel
```

Try reorienting the molecule and then returning to this view:

```
view barrel
```

And to get a list of the views you're generated, type:

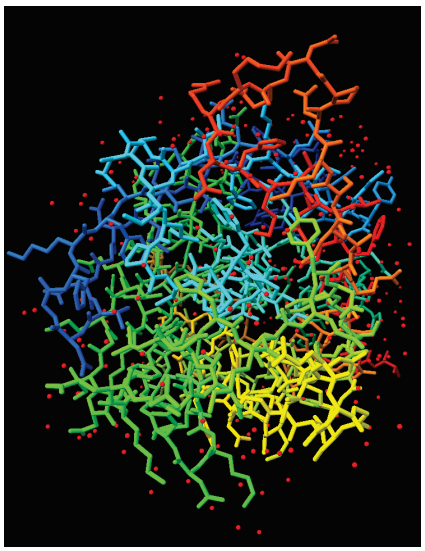
```
view list
```

You can then click on the entries you see in the **Log** panel to return to those views:

view list
Named views: barrel

Hopefully you're looking down the barrel again! Let's now have a look at all the atoms in the molecule:

```
hide ribbons
show atoms
```



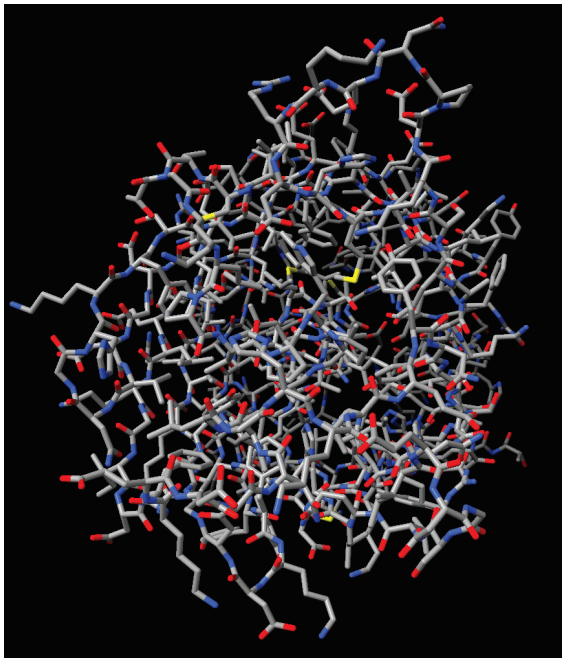
That's a lot more complicated! All the red dots are the water molecules that were ordered in the molecule when its structure was solved. Let's remove them for the moment:

```
delete solvent
```

And let's colour the atoms by (hetero)atom type:

```
color byelement
```

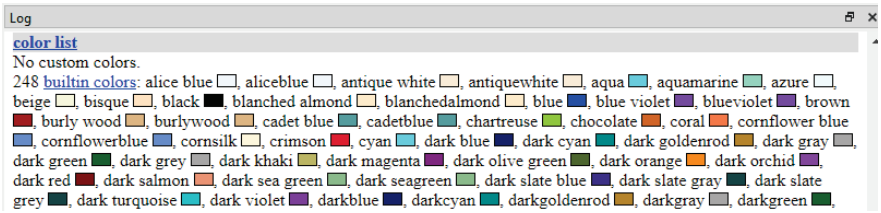
Now we see all the atoms coloured by atom type (blue = nitrogen, red = oxygen, yellow = sulfur, carbon=grey).



Hrm, that is a bit drab, let's try and find a nicer colour. You can ask ChimeraX for a list of pre-defined colours:

```
color list
```

Look at the **Log** panel to see the list:


Log

color list

No custom colors.

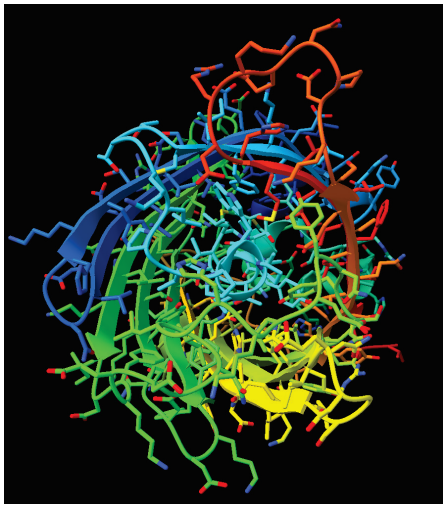
248 builtin colors: alice blue, aliceblue, antique white, antiquewhite, aqua, aquamarine, azure, beige, bisque, black, blanched almond, blanchedalmond, blue, blue violet, blueviolet, brown, burly wood, burlywood, cadet blue, cadetblue, chartreuse, chocolate, coral, cornflower blue, cornflowerblue, cornsilk, crimson, cyan, dark blue, dark cyan, dark goldenrod, dark gray, dark green, dark grey, dark khaki, dark magenta, dark olive green, dark orange, dark orchid, dark red, dark salmon, dark sea green, dark seagreen, dark slate blue, dark slate gray, dark slate grey, dark turquoise, dark violet, darkblue, darkcyan, darkgoldenrod, darkgray, darkgreen,

That's a lot to choose from. Select a colour for your carbon atoms (I'm going for a light green):

```
color C lime
```

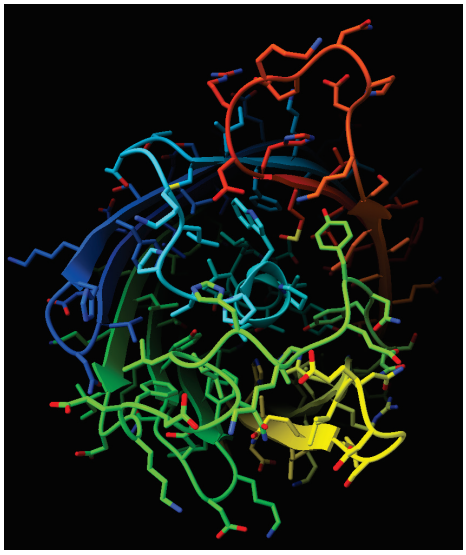
Have a look at the molecule – it's a bit hard to navigate when there is so much going on, so let's show the cartoon again but keep the side chains of the amino acids and show the rainbow view again, but this time colour all the atoms by atom type *except for the carbon atoms*:

```
cartoon
rainbow
color byhetero
```



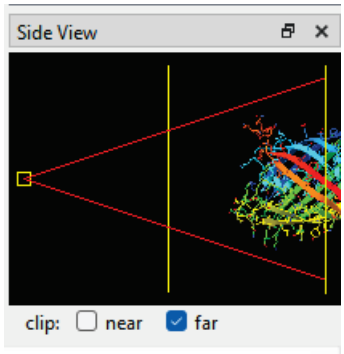
Great. Now we're getting a nicer picture, but let's try and reduce the clutter by 'clipping' the view so that we reduce the depth of field of the image. First off, we can try and set the far clipping plane so residues near the back of the molecule fade to darkness

```
clip far 5
```

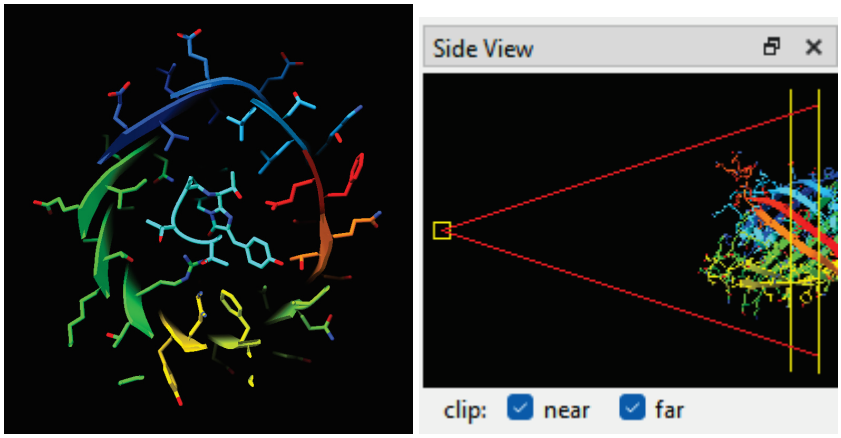


It can become difficult to understand what's happening with clipping planes, but thankfully ChimeraX includes a handy way of visualising the clipping planes that are being applied to your molecule via the Side View panel:

```
tool show "Side View"
```



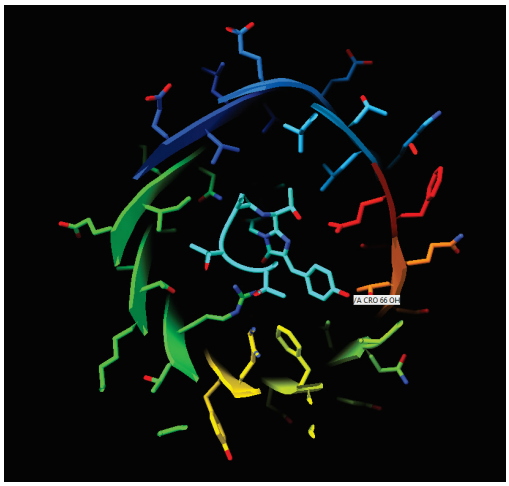
You can use this tool to turn on the near clipping plane (tick the box) and adjust its position by clicking on the yellow horizontal line and dragging. Try moving the clipping plane so that you can clearly see the tyrosine-like residue in the middle of the beta barrel:



Let's save this view tool:

```
view name barrel-closeup
```

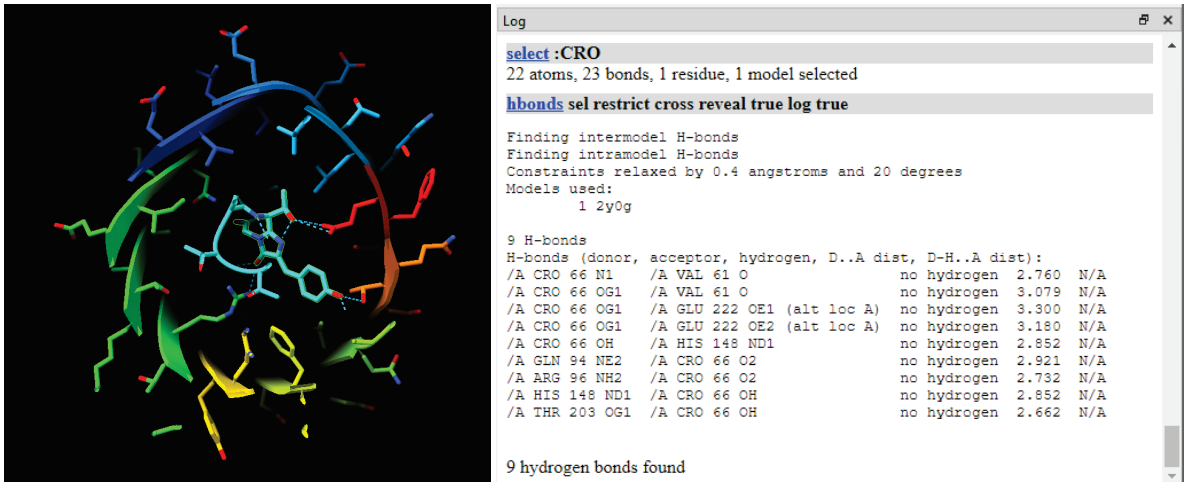
If you click on any atom in this residue you can see that it isn't actually a tyrosine residue, it's a Gly-Tyr-Gly chromophore (named 'CRO' in the structure) and this residue is key to the fluorescence of EGFP!



It looks like the terminal hydroxyl group of this chromophore makes a hydrogen bond with a threonine residue in the beta barrel – let's draw all the hydrogen bonds made by the chromophore and measure their distances

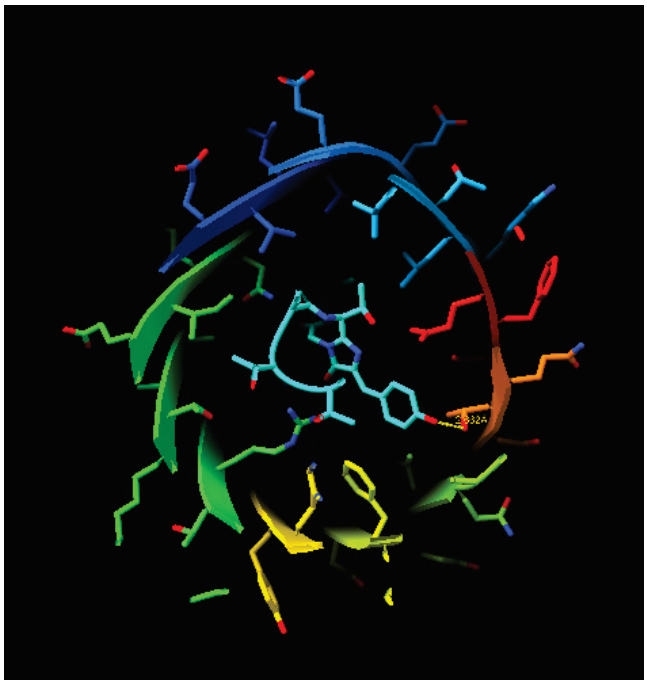
```
select :CRO
hbonds sel restrict cross reveal true log true
```

That command is a bit more complex, it first selects all residues named 'CRO' (the chromophore) and then shows all hydrogen bonds made by this selection, restricted to ones that are between the selection and other molecules, it displays (reveals) them on the image and it logs their distances.



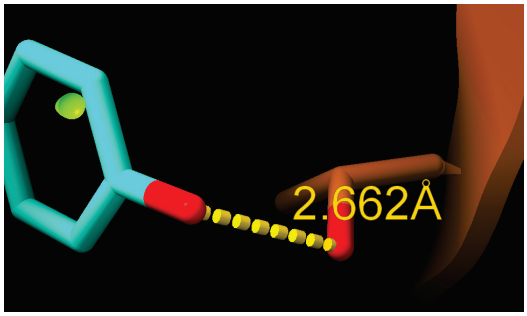
You can see that the hydrogen bond between the terminal hydroxyl (OH) of CRO and the hydroxyl group of threonine 203 is 2.66 Å. Let's show just that bond by deselecting the CRO residue (residue number 66), hiding all the hydrogen bonds, then drawing just that bond manually and showing the distance between the two atoms on the image:

```
~sel
~hbonds
distance :66@OH :203@OG1
```

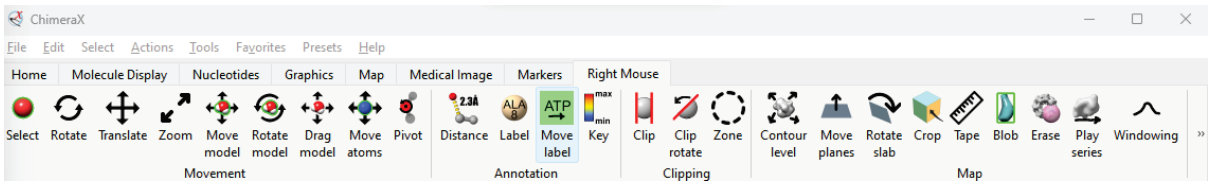


It's a bit hard to see, so let's zoom in on that:

```
view :66@OH :203@OG1
```

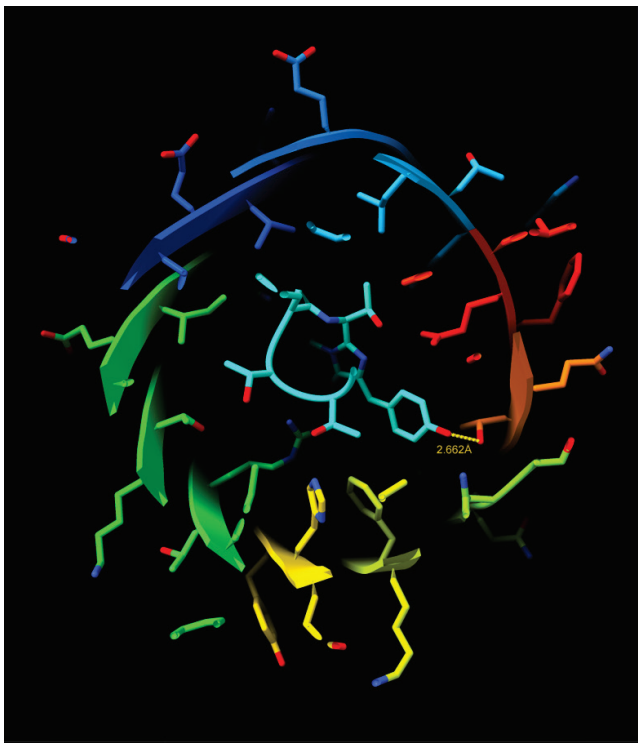


That's a bit too close! Try zooming out and changing the clipping planes until you have a nice view of the chromophore and the hydrogen bond. You can also move the label to a better position using the 'move label' mouse mode – this lets you click and drag on the label with the left mouse button to move it:

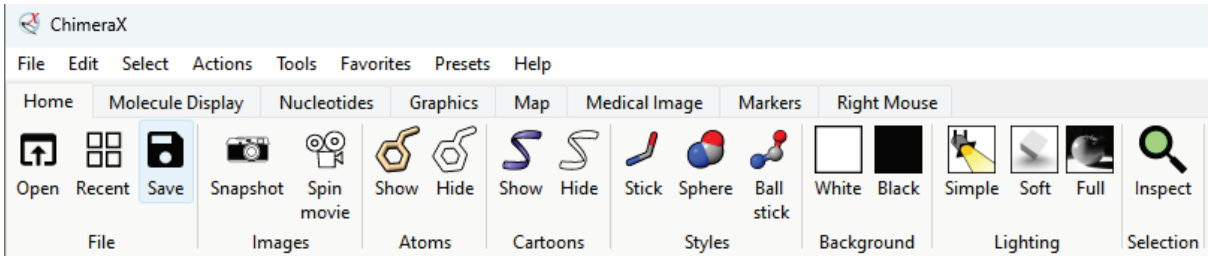


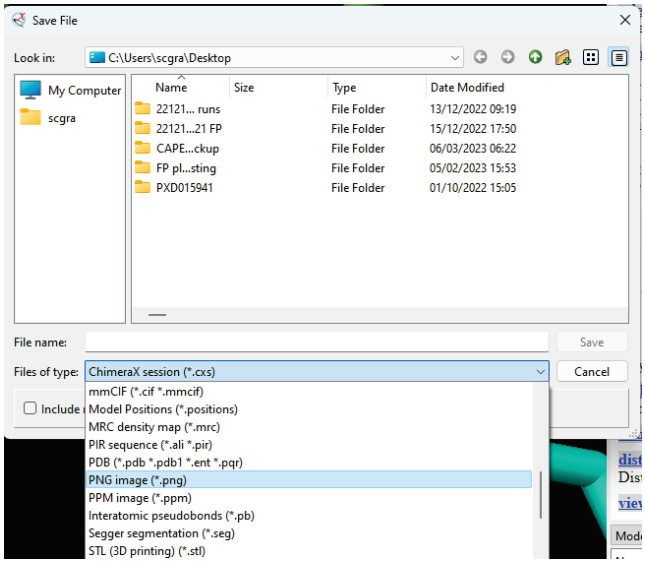
Once you have a nice view, you can save it and you can save an image of the H-bond to your desktop:

```
save ~/Desktop/CRO_H-bond.png
```



Hopefully you'll see the file appear on your desktop. If you have trouble, you can always use the save command from the Home tab of the Menu ribbon and the set the file type to a png (or whatever other image type you wish):



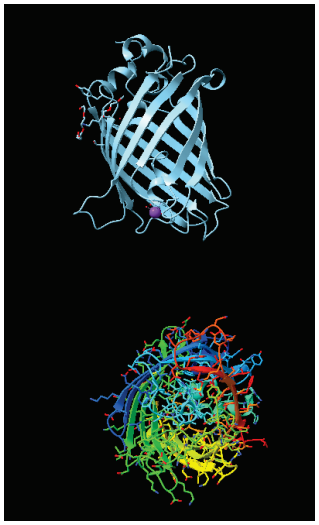


Don't forget to also save the view in case you want to return to the same orientation later:

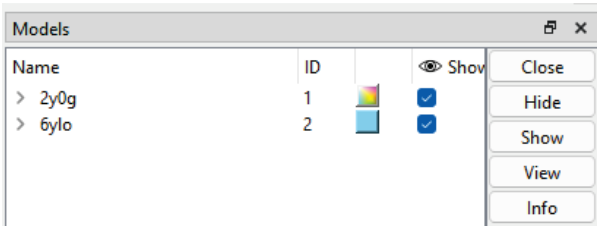
```
view name CRO_H-bond
```

Now that we've made a simple image, let's try comparing EGFP to mTurquoise2. Delete the distance, change the mouse mode to right-button selects residues, download the mTurquoise2 structure and then reset the view:

```
~distance
mouse right select
open 6YLO
view
```

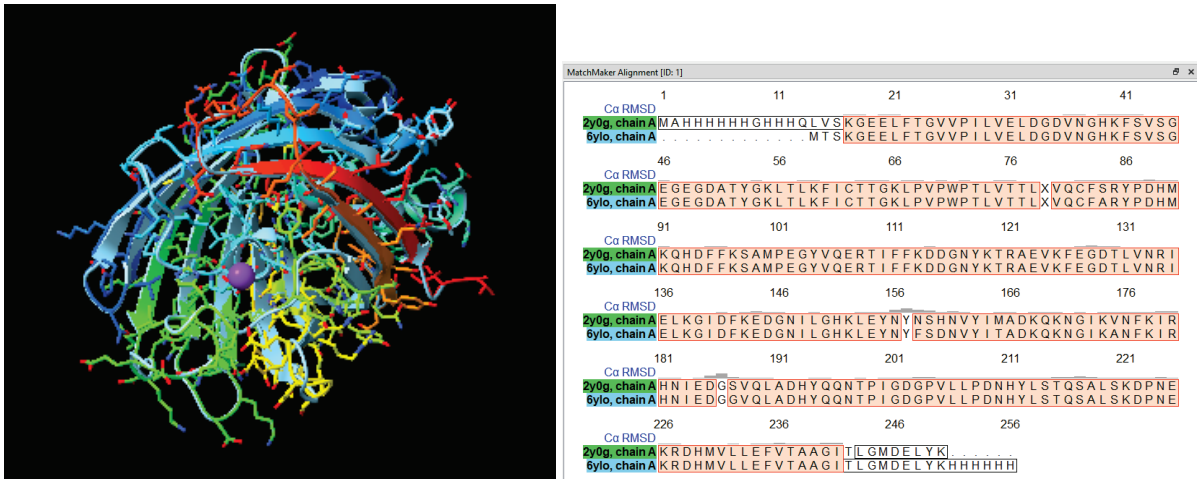


You see that we now have two models in the **Models** panel. They are named models 1 and 2, so you can change the view (etc) of each one individually:

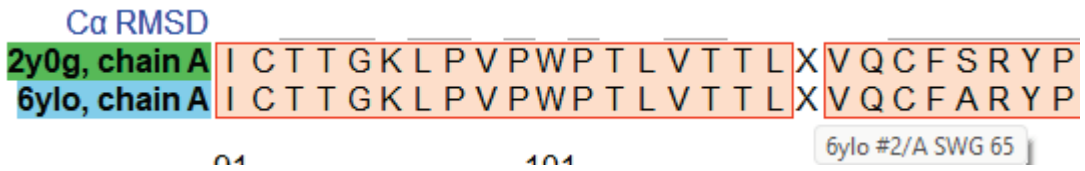


We want to show the mTurquoise2 superposed onto EGFP – this is an instance where we'll need to move the molecule, rather than move the view. This can be done with the MatchMaker tool, that can also generate a sequence alignment at the same time:

```
mm #2 to #1 showAlignment true
```



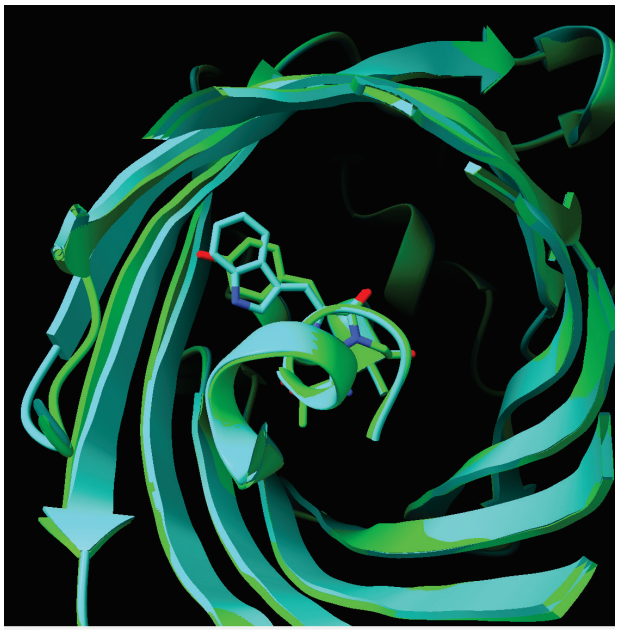
This seems to have done a good job – you can see that the EGFP and mTurquoise2 were both expressed as His-tagged proteins, but the tags were at opposite ends of the protein. Note also that the CRO residue of EGFP is called 'X' in the alignment because it's a non standard amino acid. If you hover your mouse over the equivalent residue of mTurquoise2 in the alignment you can see that this protein has a SWG residue (Ser-Trp-Gly) at this position:



It's a bit hard to see everything with all the side chains, so let's get rid of them and let's show the two chromophores. Also, let's make sure that EGFP is green and mTurquoise is blue, let's use the `name` command to make a named selection "chromophores" that includes both residues (CRO *or* SWG, where the `|` symbol represents or), and let's change the view to zoom in on the chromophores:

```
hide atoms
name chromophores :CRO | :SWG
show chromophores atoms
color #1@C* lime
color #2@C* turquoise
view chromophores
```

Your orientation will probably be end-on to the chromophore at this point:



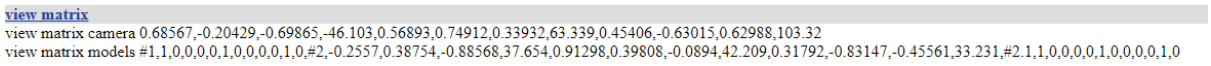
Let's instead look at the side of the barrel. You can set the view to a specific orientation (one I prepared earlier!) using the following command, which sets an explicit view matrix:

```
view matrix camera 0.68567,-0.20429,-0.69865,-33.748,0.56893,0.74912,0.33932,57.338,0.45406,-0.63015,0.62988,92.181
```

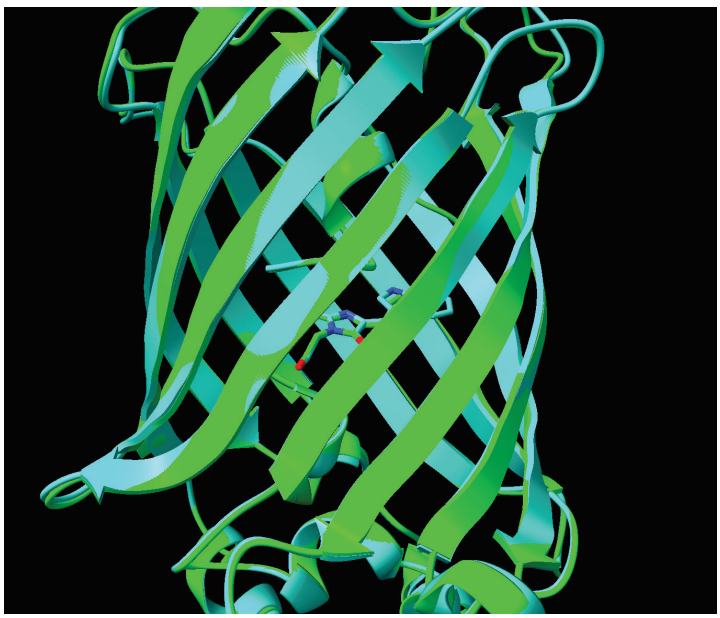
(Note: All one line, and there can't be any spaces between the numbers…don't ask me why!)

You can get the current camera matrix, which defines the view, and model matrices, which define any movements of the models, using the command

```
view matrix
```

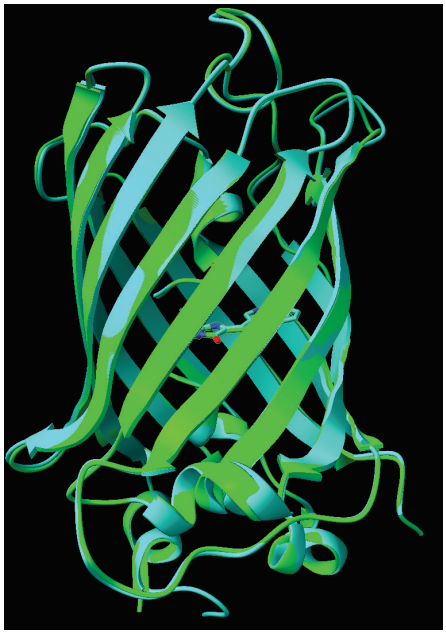


Note in the log panel how the matrix for model #2 (mTurquoise2) isn't the identity matrix – that's because we moved the model when we superposed it onto EGFP.



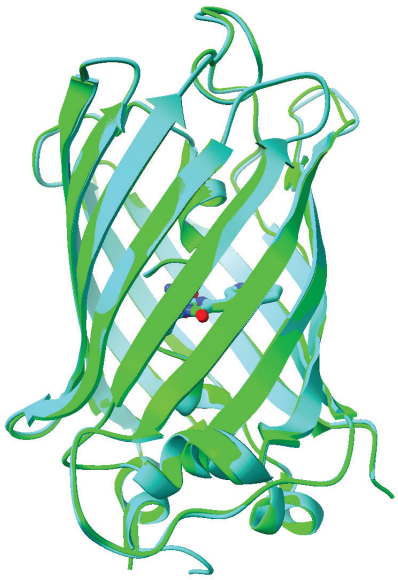
That's a bit close, so zoom out a bit and set the clipping planes to give more depth of field, and save the fiew in case we want to return to it:

```
zoom 0.6
clip far 15
clip near -15
view name side-on
```

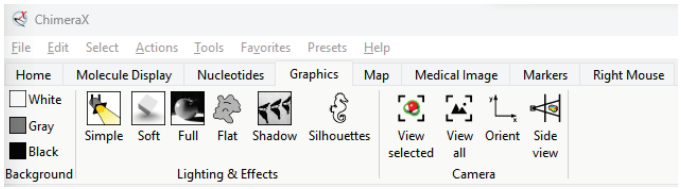


Let's make a figure where the bonds of the chromophores are a bit fatter and the background is white:

```
size chromophores stickRadius 0.5
set bgColor white
```



We can change the lighting modes using commands on the toolbar – experiment and find a mode you like:



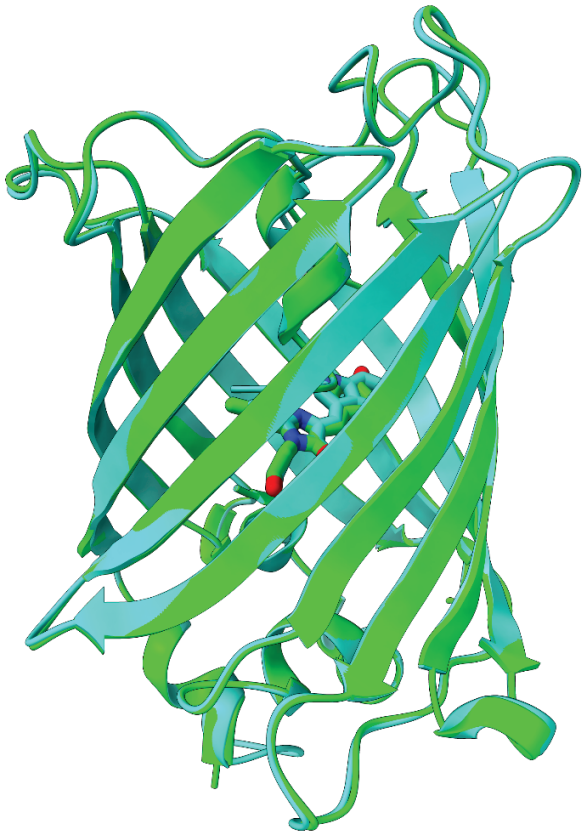
(Note that you can follow what commands are being performed in the **Log** panel). I like the following:

```
lighting flat
lighting full
lighting shadows false
```

I'd like to output a very high resolution image that I could use for a journal figure (4 inches high at 600 dpi [dots per inch]) that has been antialiased (supersampled) to remove and 'jagged' lines:

```
save ~/Desktop/EGFP_mTurquoise2_overlay.png supersample 3 height 2400
```

This might take a minute…Once it's done, if you look on your desktop you'll see the file – note how it's a much larger file than the previous image you saved.

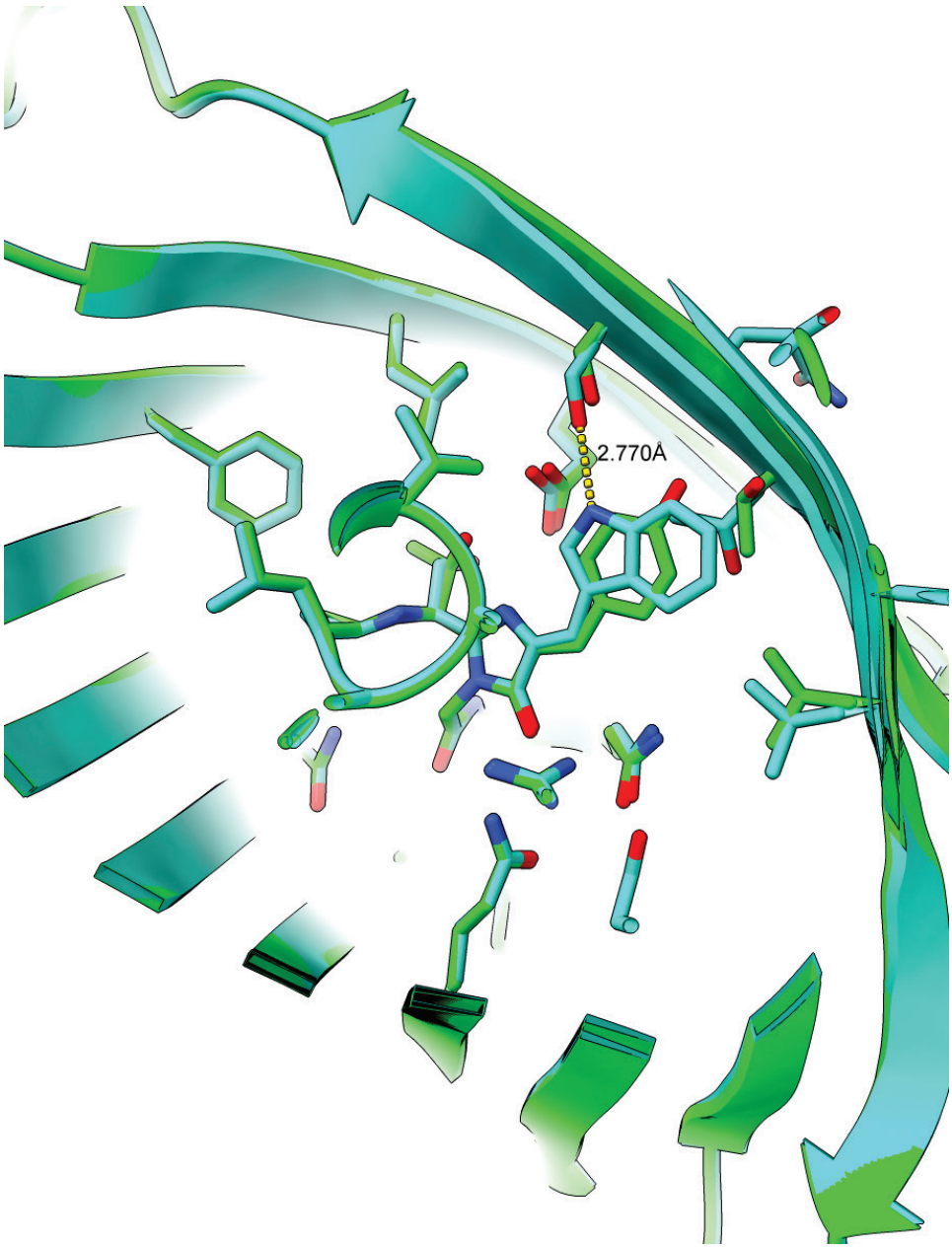


Lastly, let's look at the environment of the chromophores. Reset the bond width to the default (0.2 Å wide):

```
view CRO_H-bond
size chromophores stickRadius 0.2
sel zone chromophores 5 protein residues true
show sel atoms
sel clear
```

This command shows all residues within 5 Å of the CRO chromophore, in either model, then shows the atoms and clears the selection. You can see that the mTurquoise2 chromophore doesn't interact with the threonine residue we identified, but it does form a hydrogen bond to the hydroxyl residue of serine 205. Let's draw that H-bond, with black text for the distance label, then save a second image:

```
distance #2:65@NE1 #2:205@OG
label color black
save ~/Desktop/EGFP_mTurquoise2_chromophore.png supersample 3 height 1200
```



Lastly, save the session so you can come back to it later on if you want to:

```
save ~/Desktop/EGFP_mTurquoise2.cxs
```

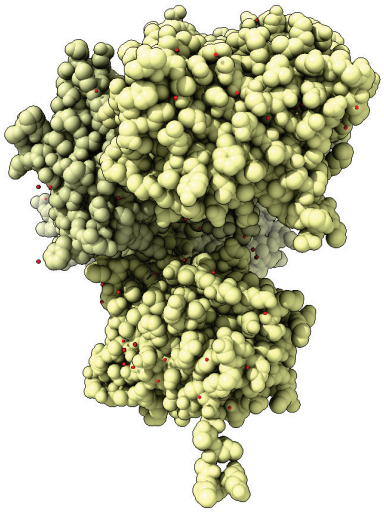
And close the ChimeraX session, ready for the next exercise:

```
close
```

Inspecting the EGFP:nanobody interaction

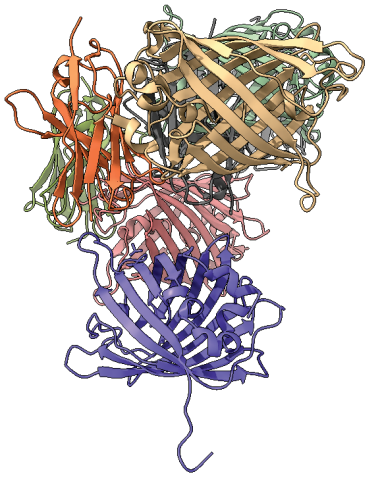
Next, we'll inspect the interaction between EGFP and the anti-GFP nanobody that we purified.

```
open 3ogo
```



It's a bit hard to see what's going on here, so let's go to ribbon view (one of ChimeraX's built-in view types) and colour the chains different colours

```
preset ribbons
color bychain
```



There seem to be more than two chains here! This is because we're looking at the crystallographic asymmetric unit, which can have more than one molecule (or complex). A quick trip to the PDB website for this structure (https://www.rcsb.org/structure/3OGO) shows us that chain A,B,C and D are EGFP, whereas chains E,F,G and H are the nanobody:

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details
Green fluorescent protein	A, B, C, D	247	Aequorea victoria	**Mutation(s)**: 3 **Gene Names:** GFP

Entity ID: 2

Molecule	Chains	Sequence Length	Organism	Details
GFP-nanobody	E, F, G, H	123	Camelus dromedarius	**Mutation(s)**: 0

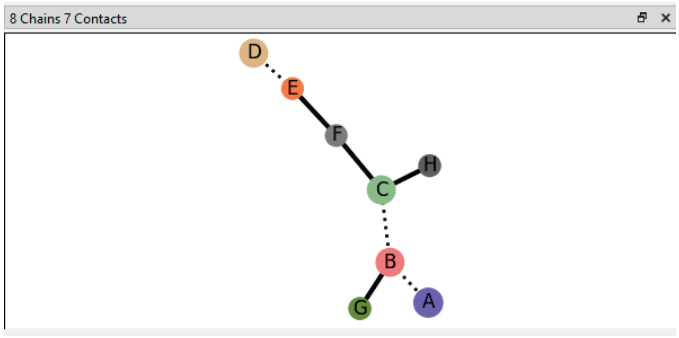
This information is also in the Log panel when you open the structure:

3ogo title:
Structure of the GFP:GFP-nanobody complex at 2.8 A resolution in spacegroup P21212 [more info...]

Chain information for 3ogo #1		
Chain	**Description**	**UniProt**
A B C D	Green fluorescent protein	GFP_AEQVI
E F G H	GFP-nanobody	

Let's look at the interactions between chains, to see which EGFP and nanobody pairs go with each other:

```
interfaces #1 & protein
```

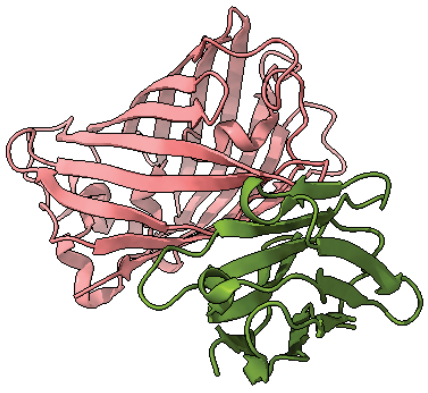


Log

interfaces #1 & protein
7 buried areas: C F 687, B G 683, E F 639, C H 348, C B 336, D E 330, A B 308

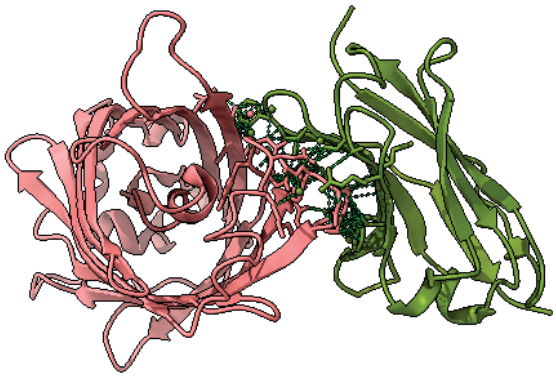
The log tells us the buried surface area between each molecule, and the interfaces with the most buried area are likely to be the biologically meaningful ones. In this case, CF and BG are likely to be biologically relevant interactions, whereas the others are likely not meaningful. Let's focus on BG for the moment:

```
name EGFP /B
name nano /G
hide ribbons
hide atoms
show EGFP | nano ribbons
```



Let's inspect the contacts made between the nanobody and EGFP:

```
contacts nano restrict cross reveal true log true sele true
```



Log

contacts nano restrict cross reveal true log true select true

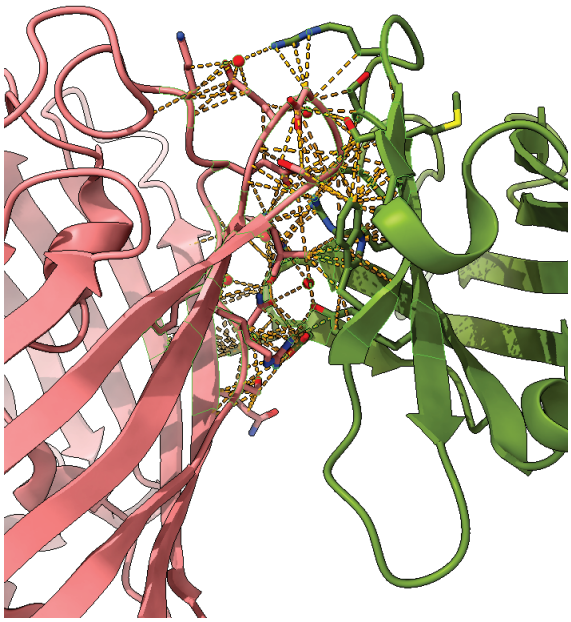
```
Allowed overlap: -0.4
H-bond overlap reduction: 0.4
Ignore contacts between atoms separated by 4 bonds or less
Detect intra-residue contacts: False
Detect intra-molecule contacts: True

102 contacts
     atom1           atom2       overlap  distance
/G HOH 137 O    /B GLU 172 CG     0.566    2.774
/G SER 34 OG    /B GLU 142 CD     0.524    2.816
/G ARG 36 NH2   /B GLU 142 CD     0.359    3.161
/G TRP 48 CD1   /B VAL 176 CG1    0.357    3.283
/G TRP 48 CZ2   /B SER 175 CA     0.341    3.299
/G SER 53 CB    /B GLU 172 O      0.338    2.962
/G ARG 36 NH2   /B ASN 170 CB     0.318    3.202
/G ASN 100 ND2  /B ASN 144 CB     0.252    3.268
/G ARG 36 CD    /B VAL 176 CG1    0.229    3.531
/G ARG 36 NH1   /B GLY 174 C      0.188    3.062
/G HOH 137 O    /B GLU 172 CD     0.187    3.153
/G ASN 100 OD1  /B TYR 145 O      0.178    2.662
/G SER 34 CB    /B GLU 142 OE2    0.169    3.131
/G GLY 102 O    /B ALA 206 CB     0.131    3.169
/G TRP 48 CH2   /B SER 175 CA     0.107    3.533
/G ARG 36 NH1   /B GLY 174 CA     0.103    3.417
/G TRP 48 CG    /B VAL 176 CG1    0.101    3.389
```

Wow – that's a lot of contacts. The command selected all the atoms at the interface,so let's zoom in on them and colour the non-carbon atoms by element, plus color the dashed lines for the contacts orange:

```
view sel
color byhetero
contacts sel color orange
```

You might need to rotate the view using the left mouse button to find a clear view:



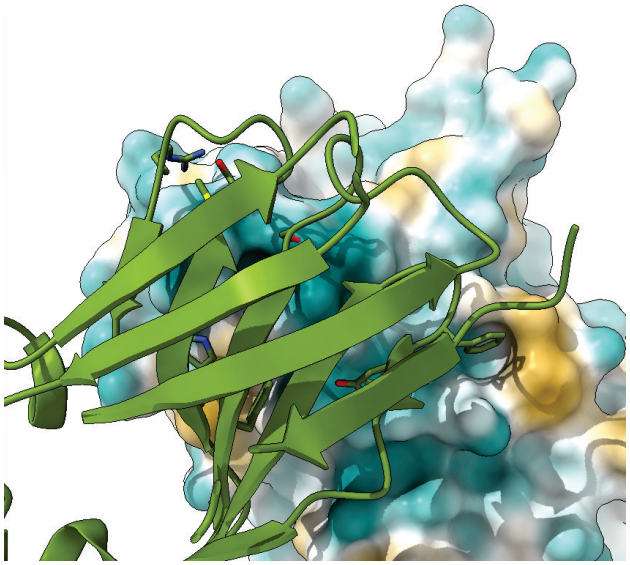
We can look whether there are any clashes between EGFP and the nanobody:

```
clashes nano restrict cross reveal true continuous true
```

The **Log** panel shows that there aren't any (it would say if there were). To get a better idea of the nature of the interaction interface, we can remove the contacts and show the EGFP surface coloured by hydrophobicity (molecular lipophilicity potential):

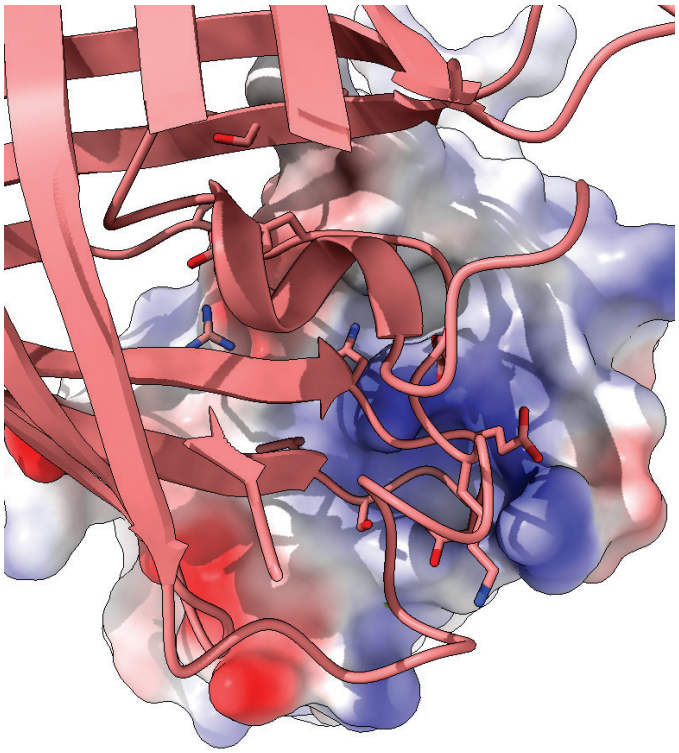
```
~contacts
sel clear
hide solvent
mlp EGFP
```

Again, use the left mouse button to rotate the view so you can inspect the footprint of the nanobody bound to the EGFP



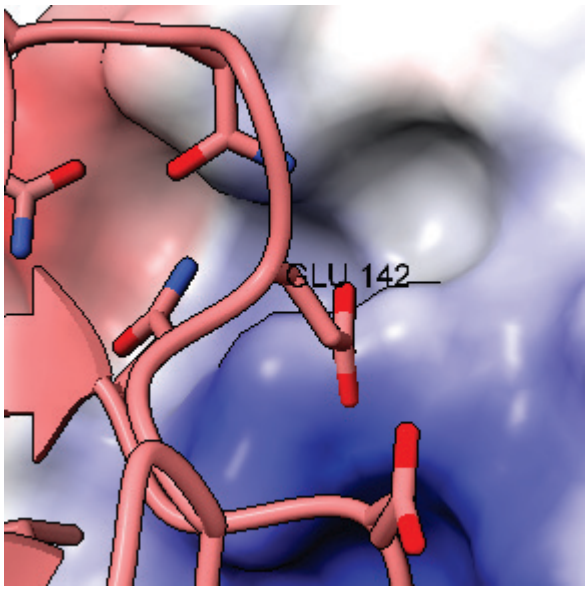
Blue is more hydrophilic and gold is more lipophilic. Have a careful look at the surface, you can see both polar and hydrophobic interactions. Let's now look at the electrostatics of the nanobody (ChimeraX currently struggles to calculate electrostatics for the EGFP because of the CRO residue – it confuses the electrostatic potential calculation program):

```
hide EGFP surface
show nano surface
coulomb nano
```



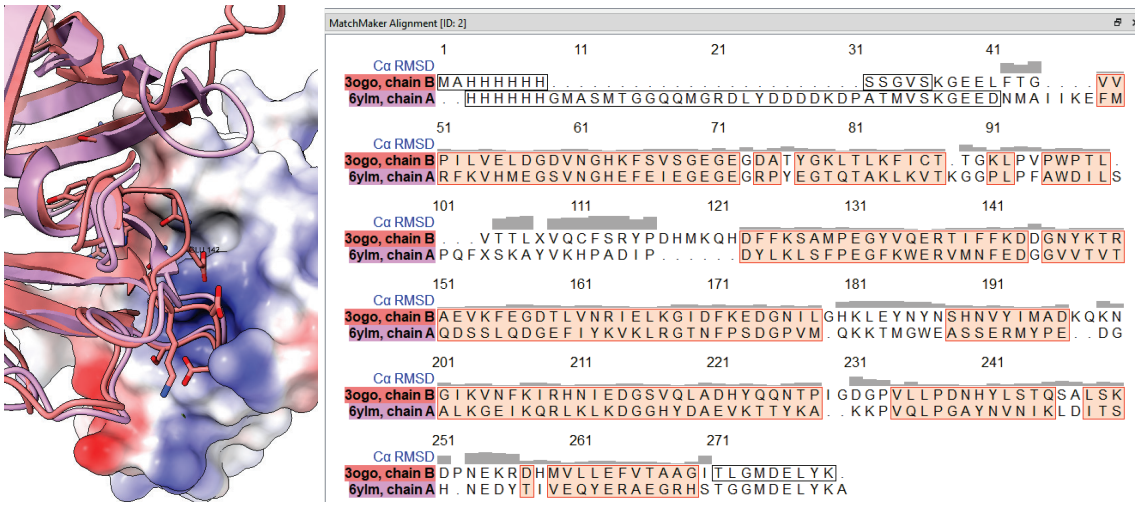
Blue is positively charged and red negative[1]. See how the glutamate 142 of the EGFP is close to a positively charged patch on the nanobody, for example:

```
~clip
view matrix camera -0.91909,0.37322,0.12637,-66.091,-0.36603,-0.68991,-0.62454,0.25204,-0.14591,-0.62026,0.77071,-21.203
label EGFP & :142
lighting shadows false
```

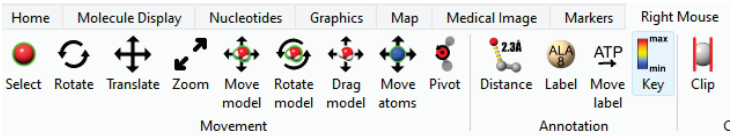


But the real interest is why some fluorescent proteins were bound by the nanobody and others weren't, so lets load the structure of mCherry and align it to EGFP:

```
open 6ylm
name mCherry #2/A
mm mCherry to EGFP showAlignment true
```

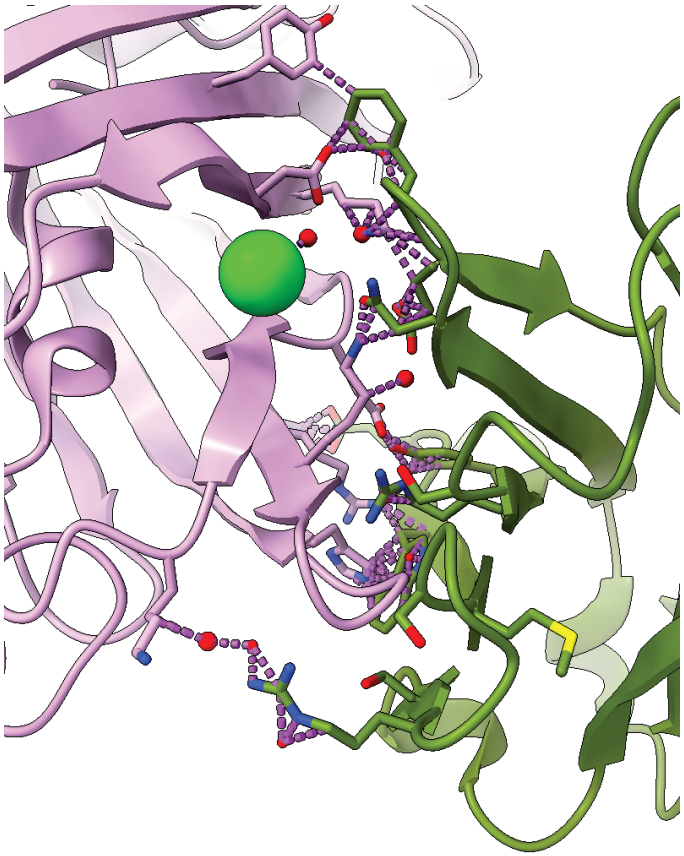


[1] You can add a legend (colour key) for the electrostatics by adding the text `key true` to the coulomb command, or by clicking `Annotation > Key` in the `Right Mouse` tab of the **Ribbon**:

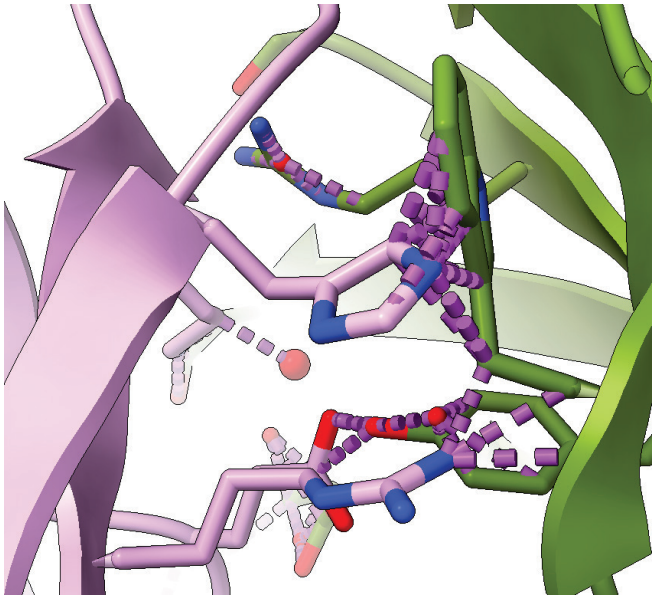


These proteins aren't as closely conserved as EGFP and mTurquoise2. We can have a quick look at why we think mCherry might not bind the nanobody. Are there clashes between the proteins?

```
~surface
~label
hide EGFP atoms
hide EGFP cartoon
clashes nano restrict cross reveal true
```



There are lots of potential clashing residues, including extensive clashes between nanobody Trp48 and His172 of mCherry.



This explains our pull-down results, so we can save our session then remove the mCherry protein from the session

```
save ~/Desktop/EGFP_nano_mCherry.cxs
delete mCherry
```

Why don't you now delete mCherry and have a look at mTurquoise to see whether you can identify the extent of clashes between this protein and the nanobody. You'll need to superpose the mTurquoise onto EGFP using matchmaker (mm). Also, see if you can work out how to superpose the other EGFP:nanobody complex (chains C and F) onto chains B and G that we were inspecting. Are they in similar conformations? (Hint: you might need to clone the #1 molecule)

Computer lab 2 – Visualising proteins

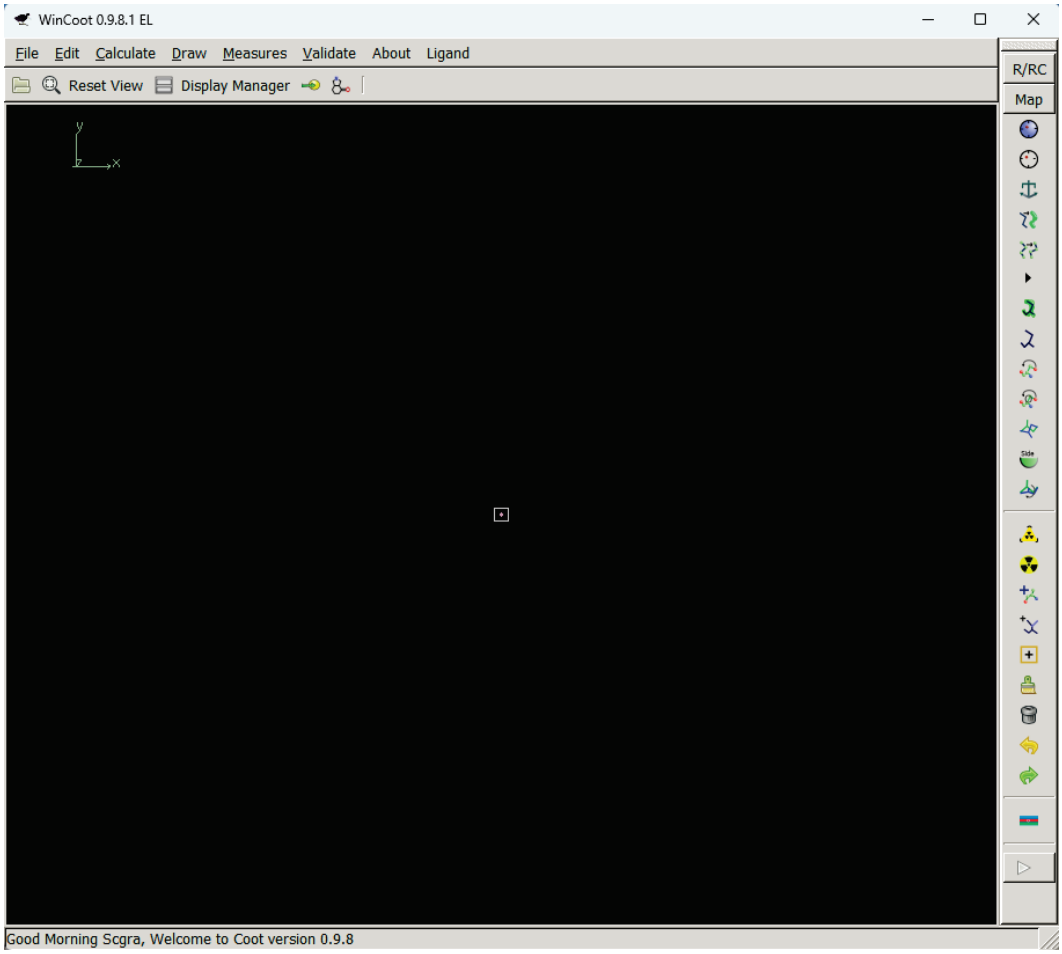
Overview

Today we will re-look at the structures of EGFP in complex with the anti-GFP nanobody, but using COOT[1] to inspect the electron density maps, which is the experimental data into which the model was built. We will also use the validation tools built into COOT and available at the PDB to inspect the properties of the model and validate whether it is high quality. You can download COOT from the following websites, depending on your operating system:

- Windows: http://bernhardcl.github.io/coot/wincoot-download.html
- Linux: https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/binaries/release/ (go to bottom of page for latest releases)
- MacOS: Follow instructions at https://scottlab.ucsc.edu/xtal/wiki/index.php/Stand-Alone_Coot

Inspecting the EGFP:anti-GFP nanobody complex electron density

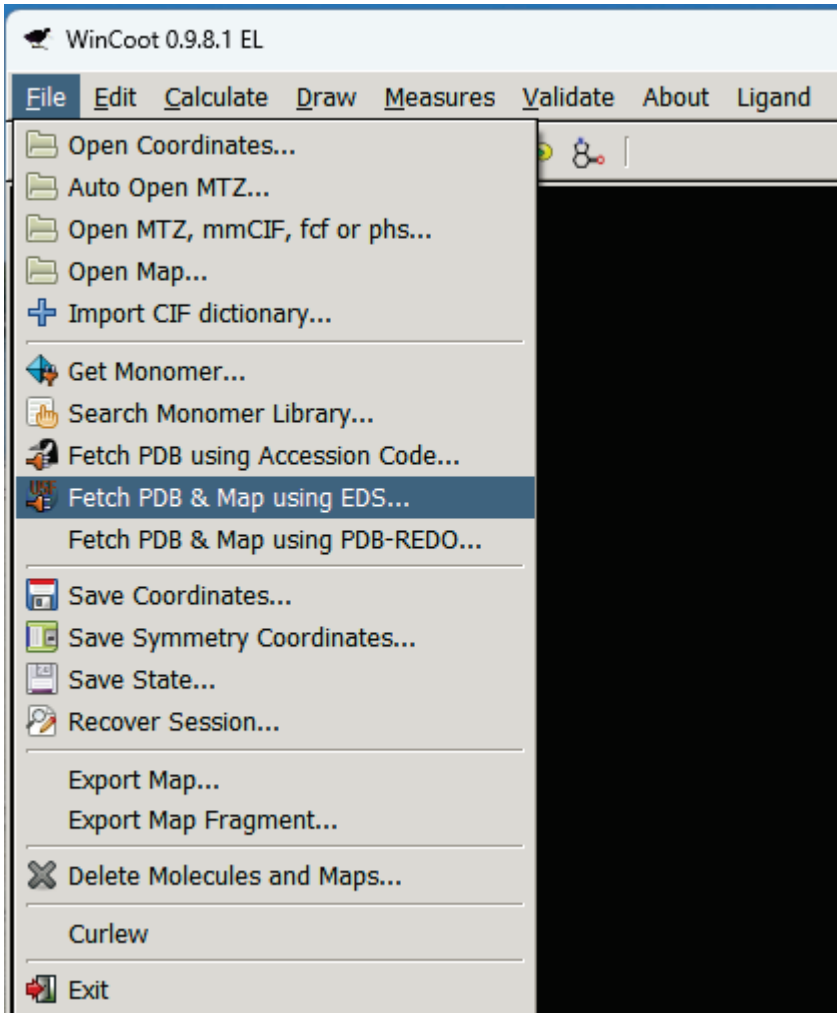
Open COOT[2]

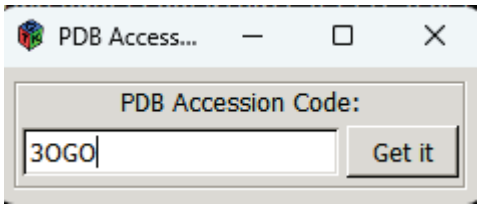


Use the build-in data retrieval tools to download the structure of the complex, plus associated experimental data (electron density maps), from the Electron Density Server hosted by the EBI (https://www.ebi.ac.uk/pdbe/) using the **File > Fetch PDB & Map using EDS…** tool:

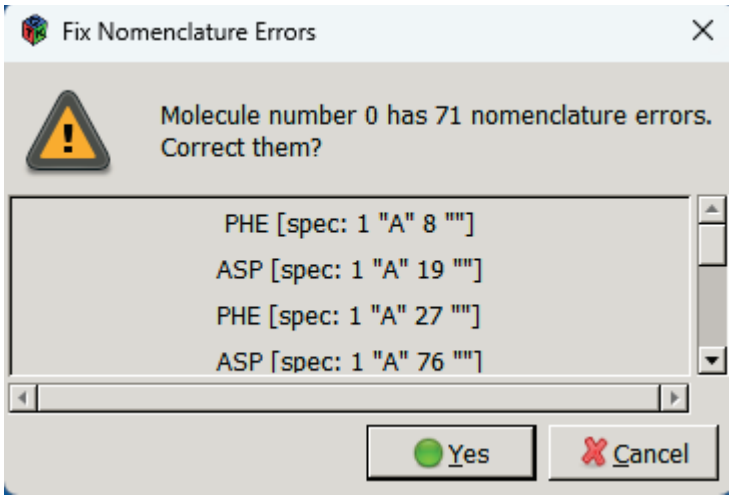
[1] P Emsley 1, B Lohkamp, W G Scott, K Cowtan (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486-501

[2] This tutorial assumes you are running WinCoot version 0.9.8

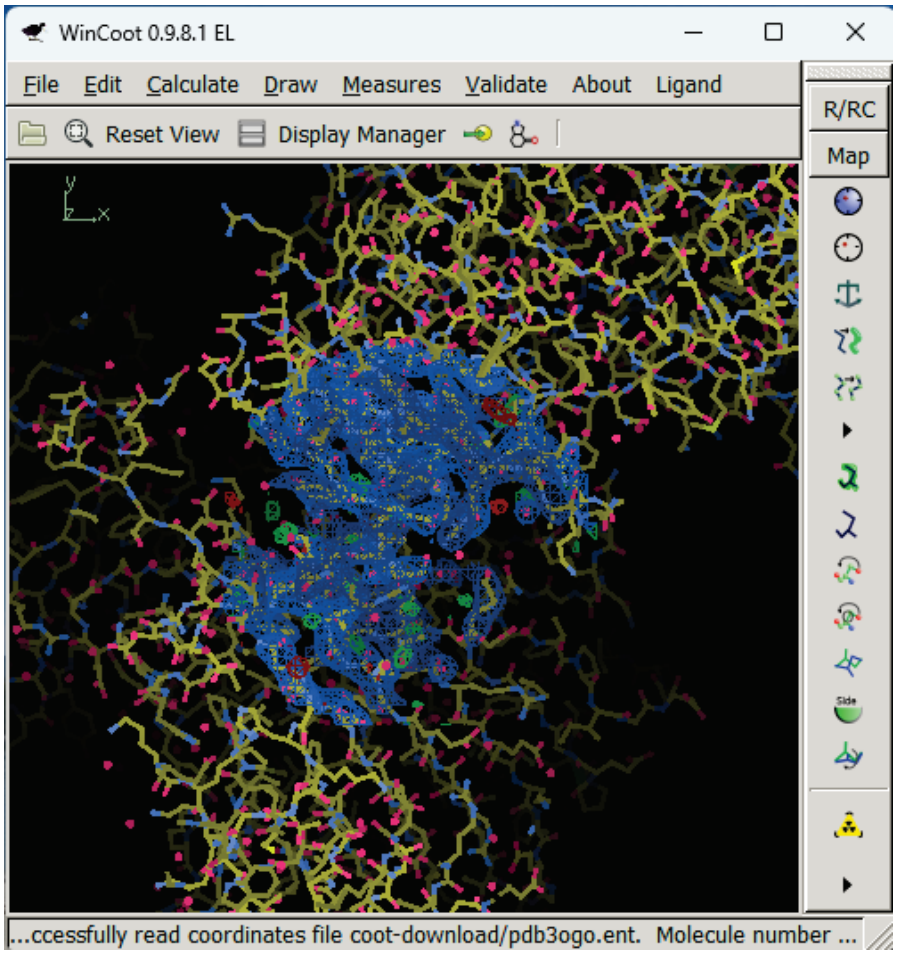




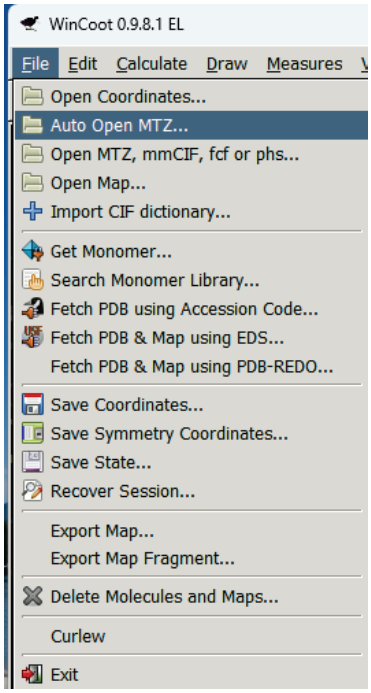
COOT will warn you about some 'nomenclature errors', because the names of some atoms don't match official guidelines – this is fine, you can hit Yes (this isn't anything to worry about).



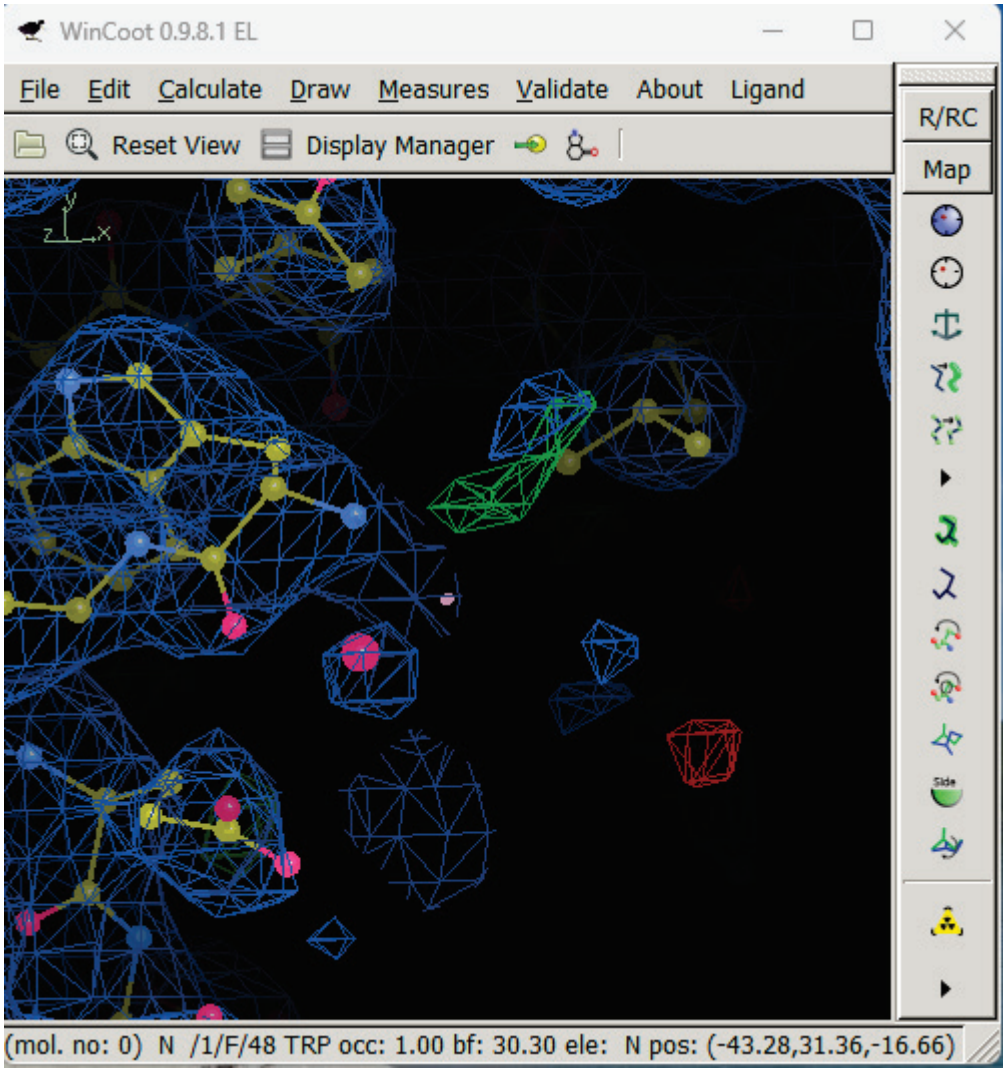
This should show you both a protein model and an electron density map (blue chicken wire).



If it doesn't (it was a bit buggy for me the other day) you can download the electron density map from the course website (http://www.atomicvirology.path.cam.ac.uk/brazil_files/3ogo_phases.mtz) and open it using the **File > Auto Open MTZ…** tool in COOT

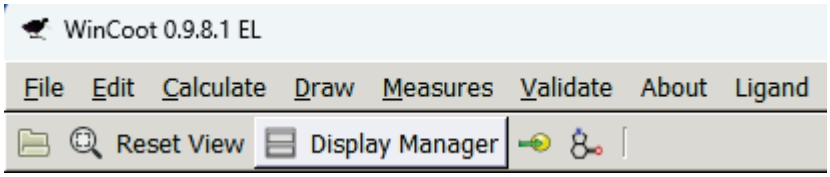


Once you've successfully opened your molecule and electron density map, try zooming in and out (right mouse button), rotating (left mouse button), translating (middle mouse button) and changing the map contour level (mouse scroll wheel). What you're looking at is as follows:

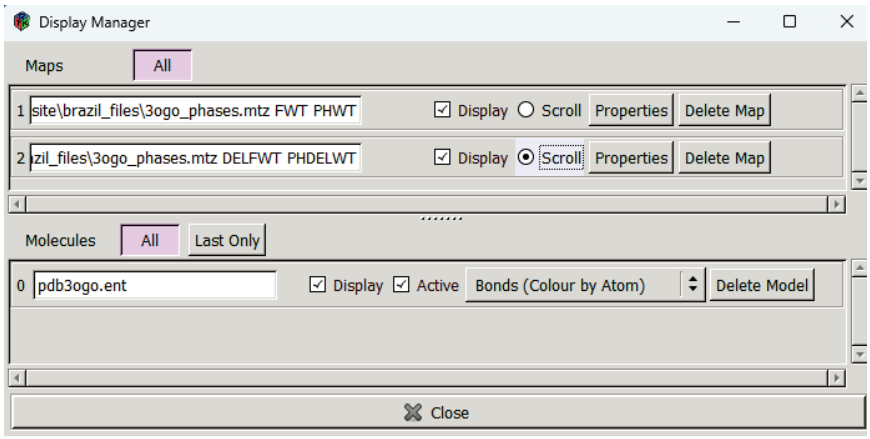


- Balls and sticks = Protein (model) and water atoms
- Blue chicken wire = 2Fo-Fc electron density, showing where electrons were observed in the experiment
- Green chicken wire = Positive Fo-Fc density, shows where the model predicts fewer electrons than were observed in the experiment
- Red chicken wire = Negative Fo-Fc density, shows where the model predicts more electrons than were observed in the experiment

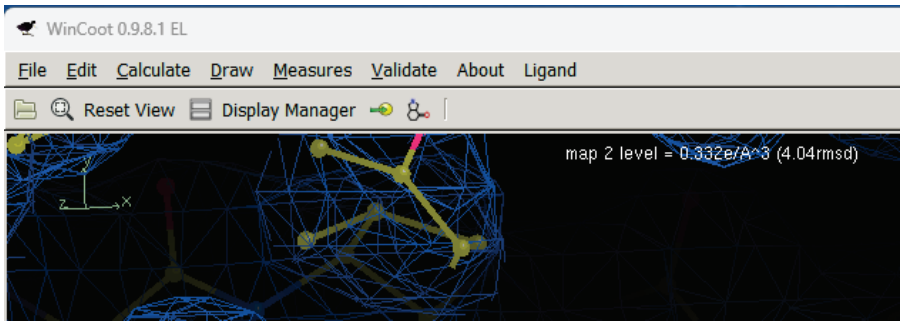
Note that the Fo-Fc maps can be quite noisy, especially for a final refined structure (where all the big positive and negative peaks have been fixed and we're only looking at noise). You can change the contour level (signal:noise) level of the Fo-Fc map by first opening the **Display Manager**:



Toggle Scroll to be selected for the 2, DELFWT map (which is the Fo-Fc map):

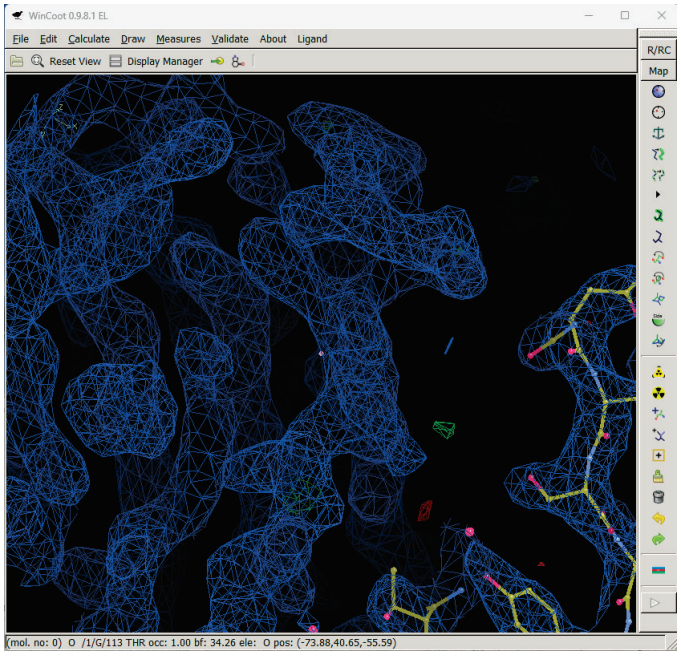


Then click back onto the main window and try scrolling your middle mouse button. The map level will be shown in the top of the main window – you want to aim for a contour of ~4 RMSD.

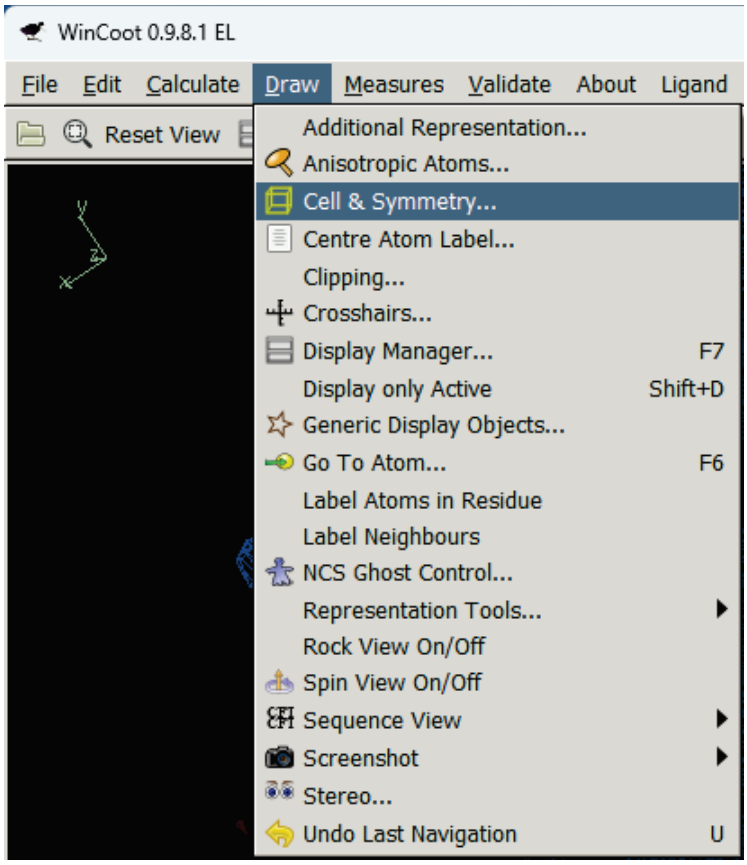


Now use the Display manager to select the 2Fo-Fc map (1, FWT) and set the contour to something like 1.5 RMSD.

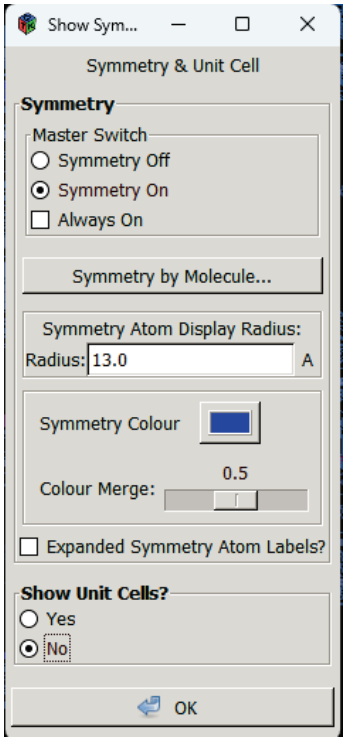
If you move around the structure (middle mouse) you can see that, for the most part, the model fits nicely into the middle of the electron density. However, if you scroll far enough you'll see areas of space where there is electron density (blue map) but no model.



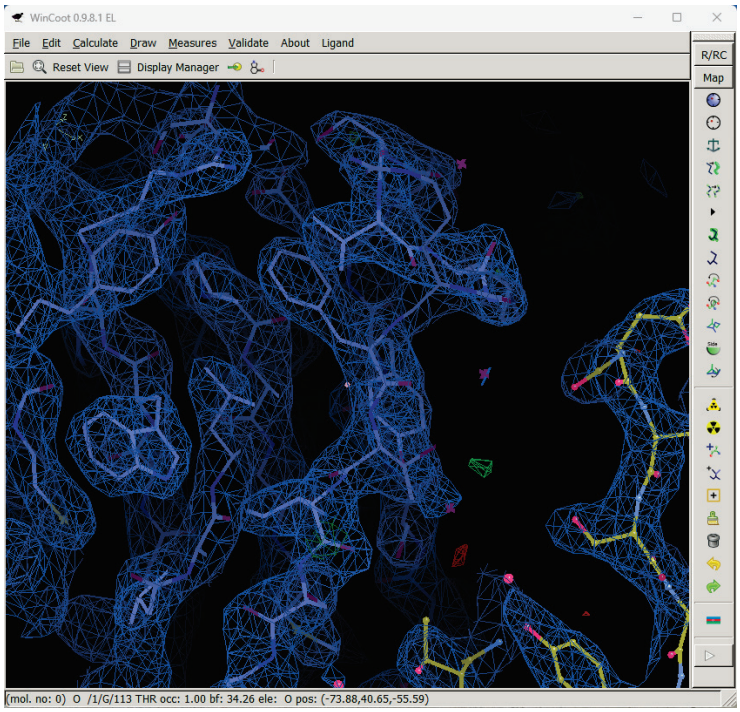
This is because the crystal is 'infinite' due to symmetry – COOT is automatically generating the density for areas of space by combinging the information from the crystallographic asymmetric unit with the crystal symmetry. This is useful! We can use the same principle to show symmetry-related atoms from our model using the **Draw > Cell & Symmetry…** tool:



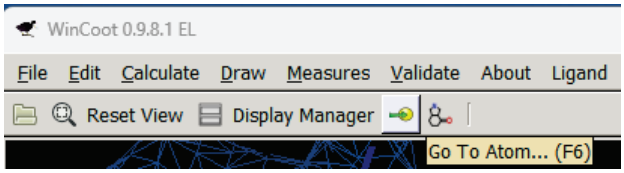
Turn the Symmetry Master Switch to On and hit OK:

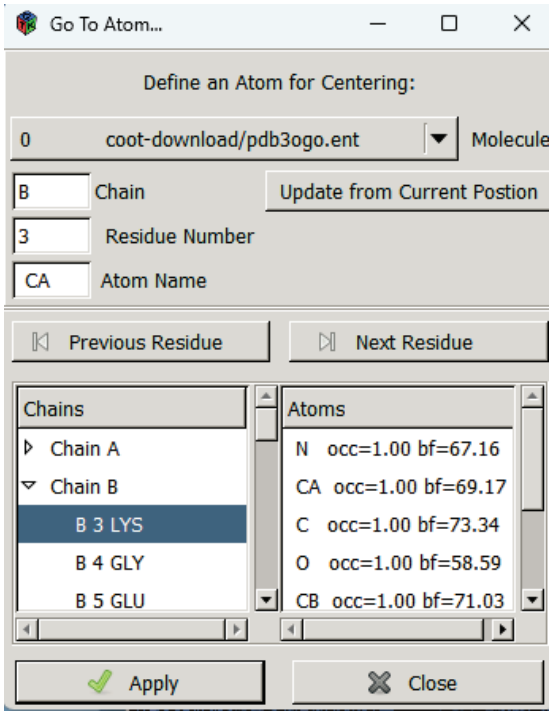


If you zoom out you can now see atoms in the symmetry-related density

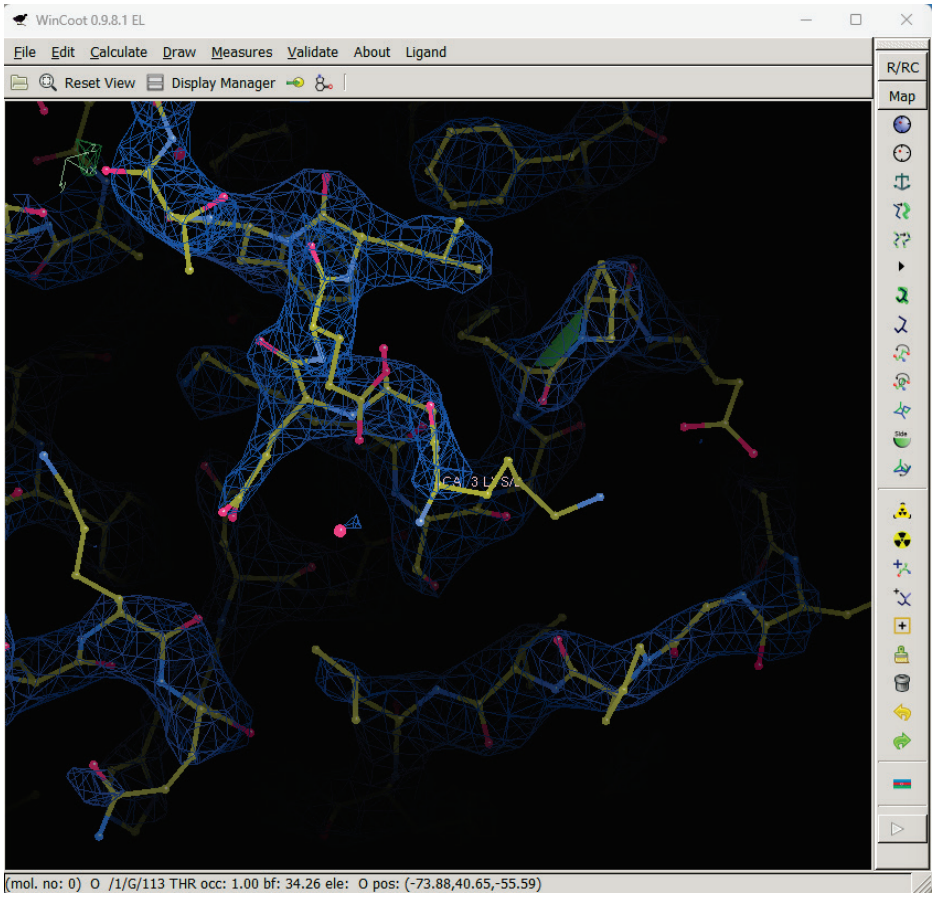


We're now ready to inspect the molecule, but where to start – the structure is huge! Let's start by natigating to the first residue of Chain B, the EGFP molecule we were looking at yesterday. Do this using the Go To Atom… window and then double-click on the first residue of chain B, which is a lysine residue

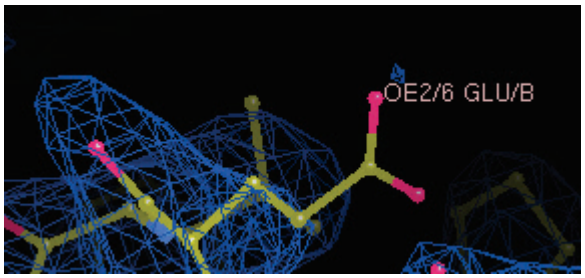




Go back to the main window. Lysine 3 is in the middle of the view, but you can see that the electron density for the side chain isn't as strong as for the main chain.



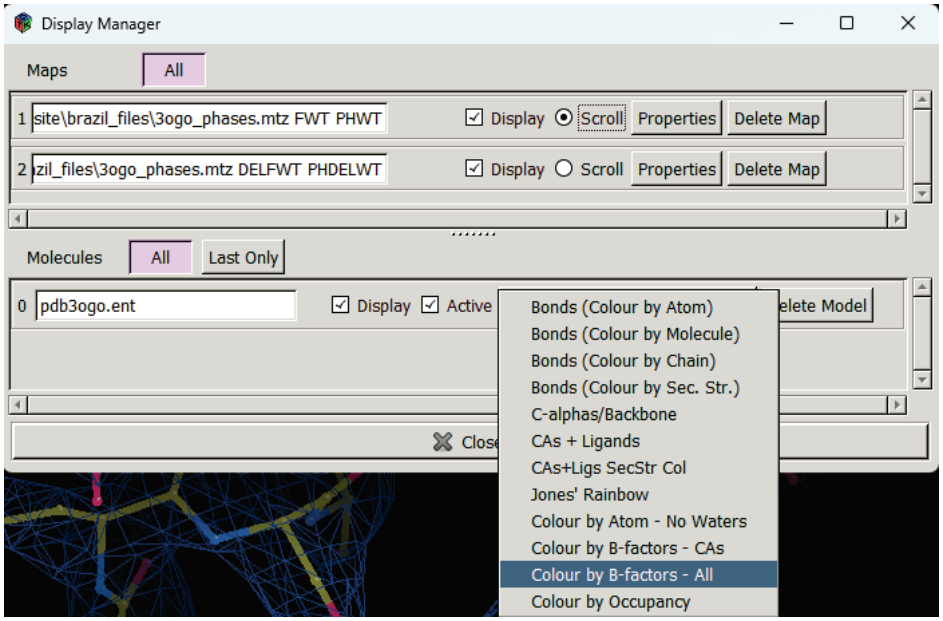
Why? This is because the side chain of this lysine is mobile. The maps in crystallographic experiments show the time and space average positions of electron density in the crystal – if atoms are highly mobile then these will average to (near) zero and you have a lack of electron density. This is presumably what happened for reisude 2 of this EGFP molecule, which isn't modelled in the structure[1]. If you go to residue 6 (Go To Atom, Glu 6) you can also note that glutamic acid has poor density too:

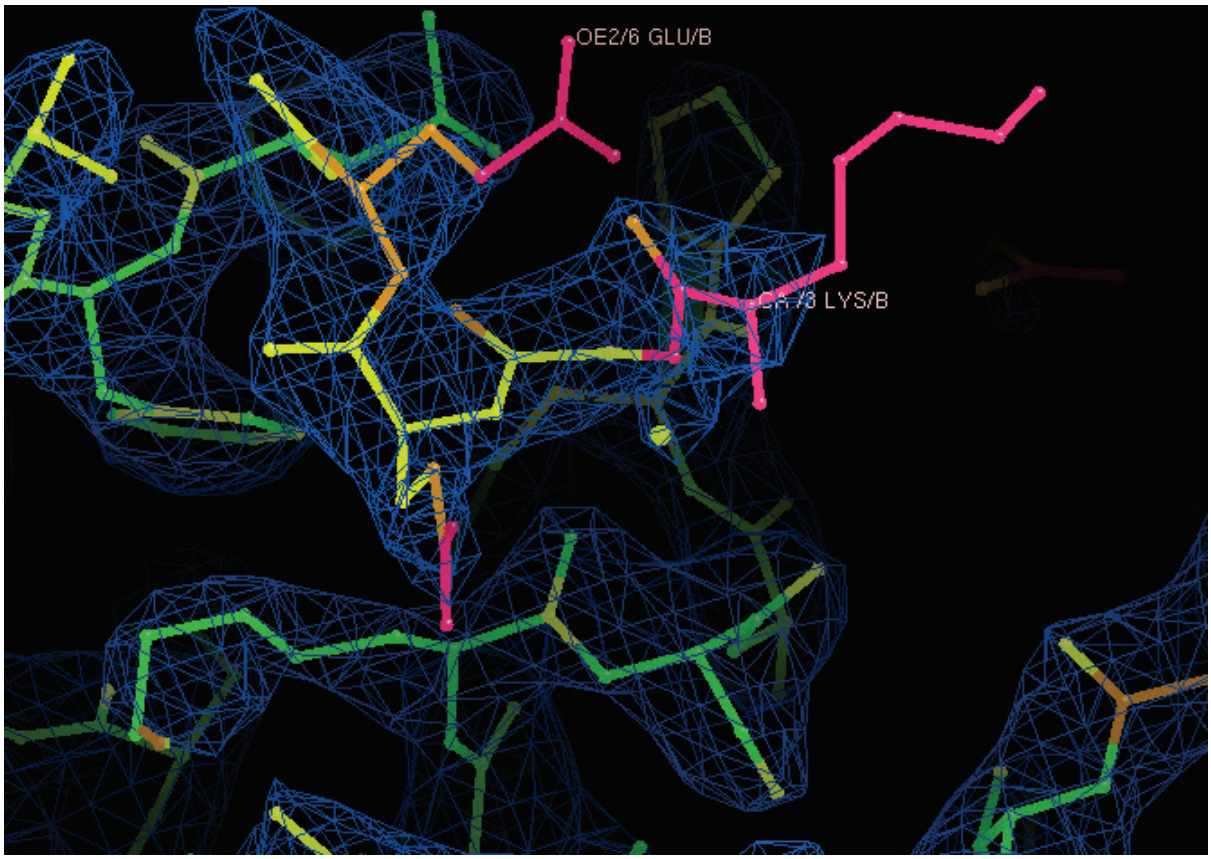


If you scroll the 2Fo-Fc density up and down you can see some weak density appear for this side chain, whereas the lysine really is very poorly ordered (almost no density).

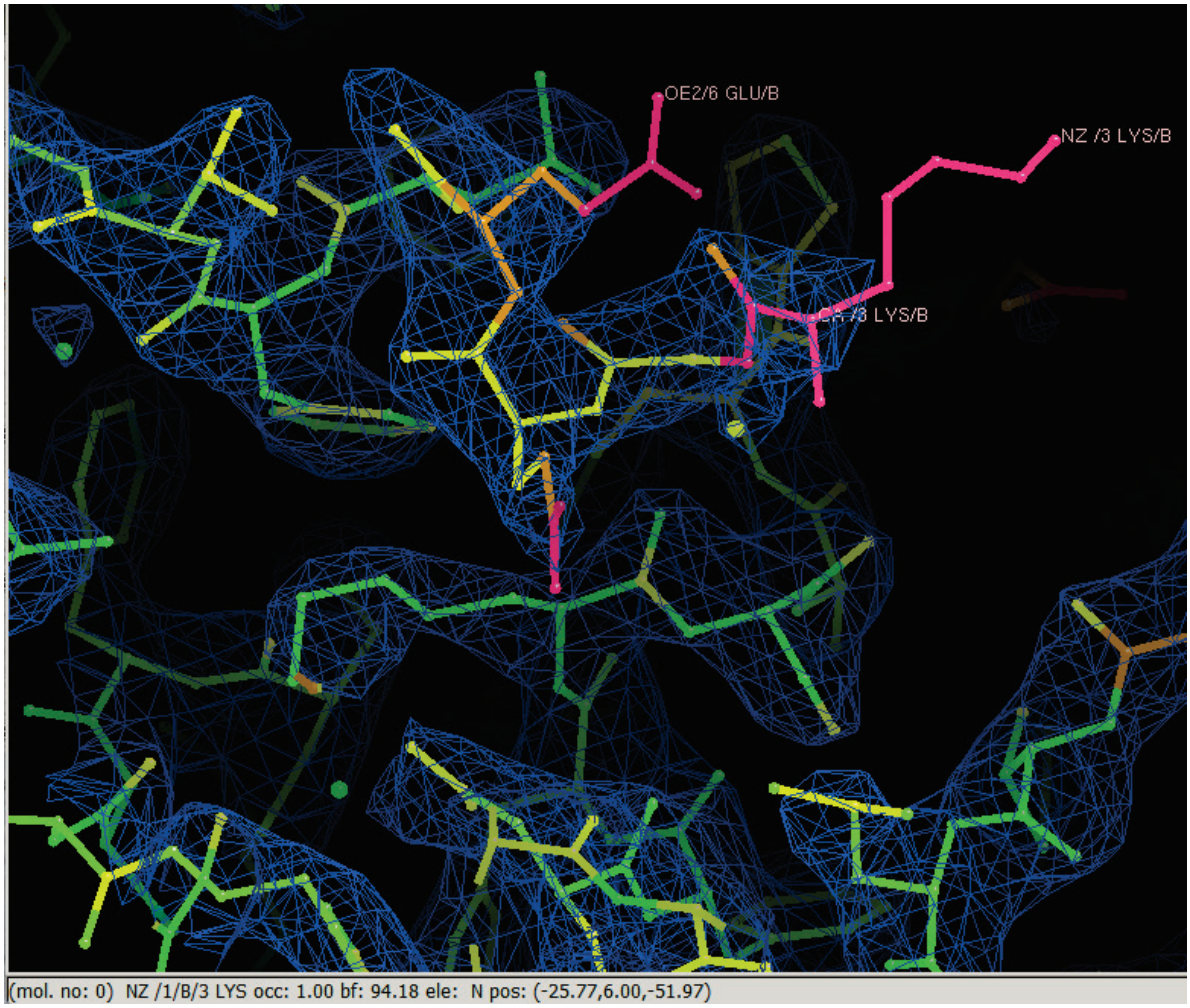
[1] Crystallographers are generally conservative types, so we tend not to model residues if we can't see them in electron density (even though we know from the sequence that *chemically* that the residue is most likely present). Opinions vary on what to do with side chains that are poorly ordered, but the consensus is to put the side chain somewhere chemically plausible if the electron density is weak (as not having the side chain makes no chemical sense).

There is actually information in the model about how mobile each atom is, this is the 'B factor' (aka atomic displacement parameter, ADP, or temperature factor) of each atom. We can colour the model by B factor using the Display Manager:





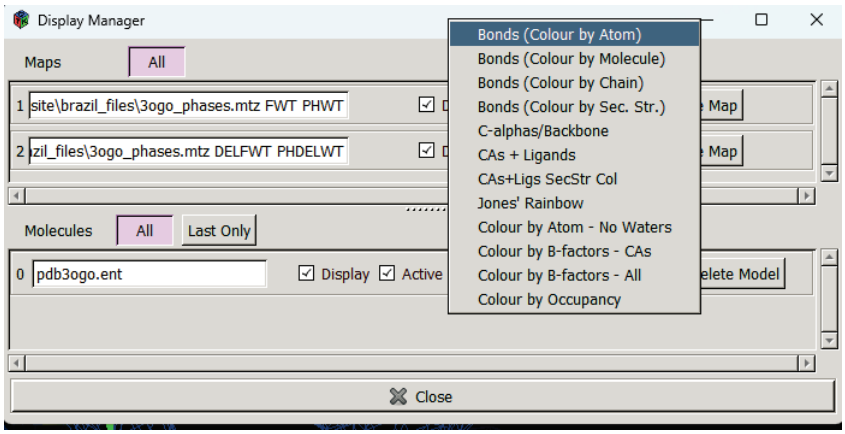
You see how the residues with poor electron density also have high B factors (red), whereas the residues in strong density have lower B factors (green). You can also hold `Shift` and left mouse button click on any residue to label that residue and show its B factor in the bottom panel of the COOT window:



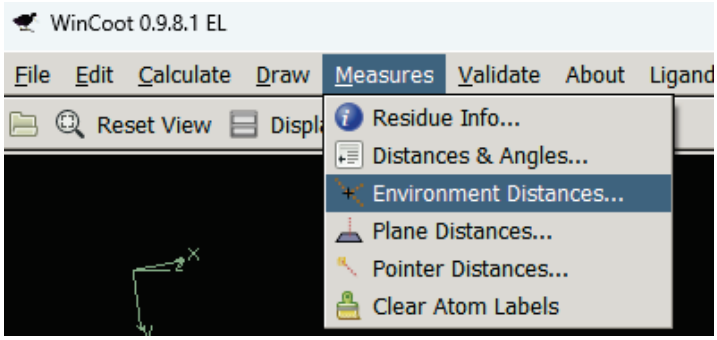
The final nitrogen atom of this lysine (part of the side chain amino group) has a very high temperature factor, so it highly mobile.

This is an important point – If you were to remove the map and just look at the model without thinking about the temperature factors you might think that the lysine side chain was well ordered and definitely where it sits in the model. The reality is that it's super-flexible and samples lots of different conformations. **It is critical that you at least look at B factors, and when possible look at the electron density maps, before designing any experiments based on someone else's structural data!**

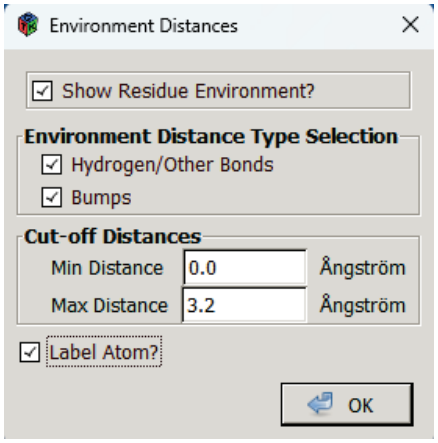
Let's go back to per-atom (element type) colouring:



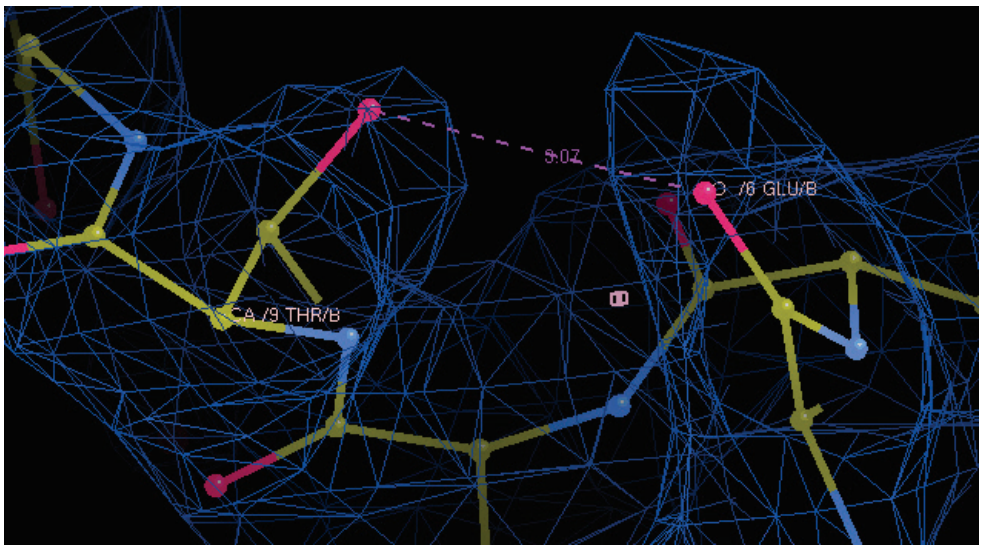
Step through the structure residue by residue. You can move to the next residue using the `Space` bar on your keyboard and go back using `Shift + Space`. When you get to Thr 9 you can see that the side chain hydroxyl group looks like it's making a hydrogen bond to the carbonyl oxygen of Glu 6. You can show all the hydrogen bonds and close contacts of the residue you're currently centred on using the **Measures > Environment Distances** tool:



I like to show the residue environment and label the atom:



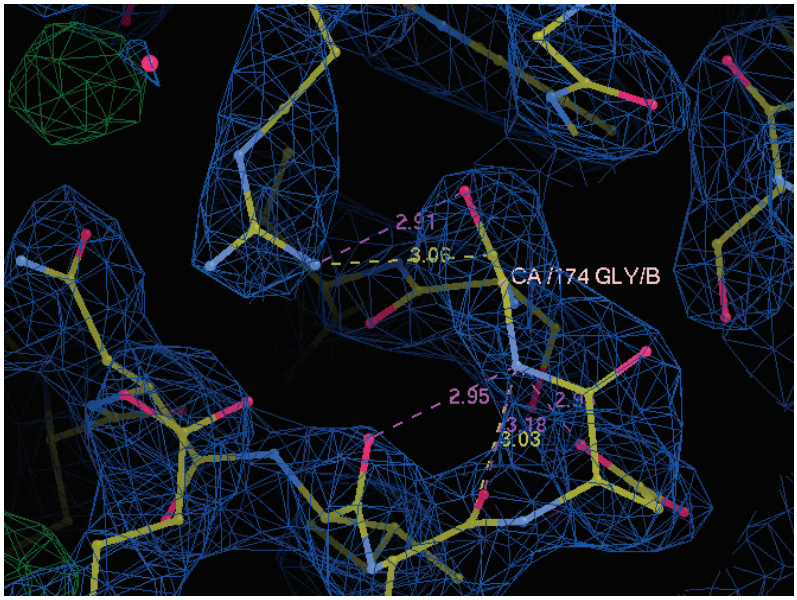
You can see now that the Thr side chain is indeed making a 3.07 Å hydrogen bond[1]:



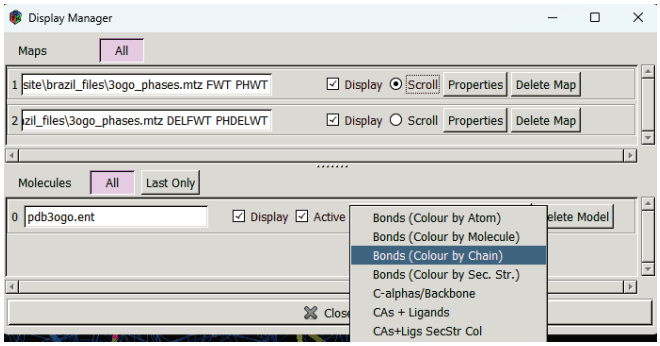
[1] If the text is a bit hard to read you can change the font size using **Edit > Preferences**, then click on **Others** and go to the **Fonts** tab where you can change the size

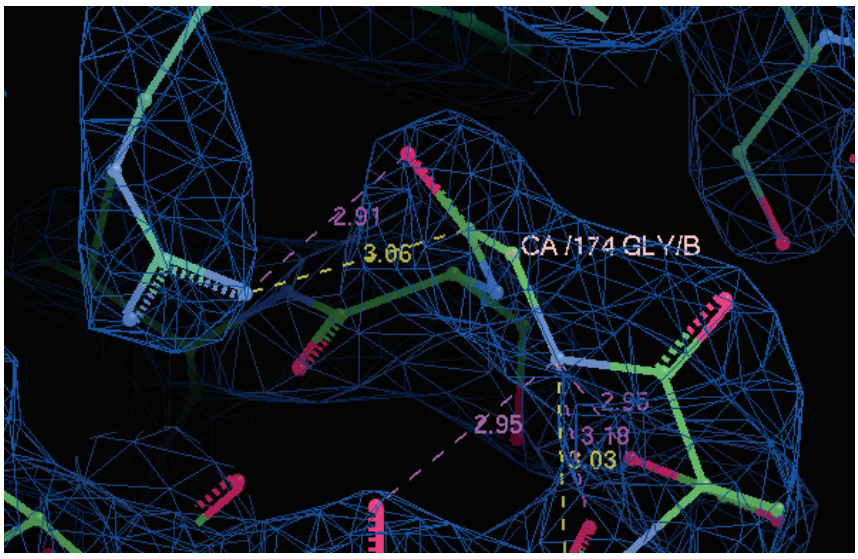
Walk through the structure a bit more and you'll find that residue 13 is at the start of a beta-strand. You can see the hydrogen bonds between the carbony oxygen and amide nitrogen atoms of adjacent strands that hold this anti-parallel beta barrel together.

Let's skip forward now to residue 174 (using the Go To Atom window, or press Crtl + G and then type B174 and press Enter). The carbonyl oxygen of this residue is making a hydrogen bond to the side chain of an arginine, but is this from the same molecule?



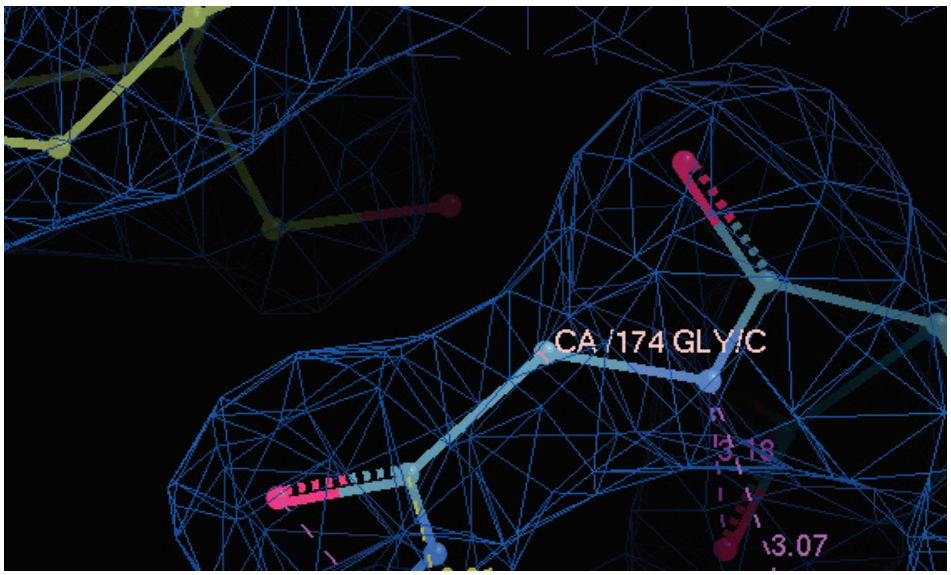
Turn on Bonds (Colour by Chain) in the Display Manager:



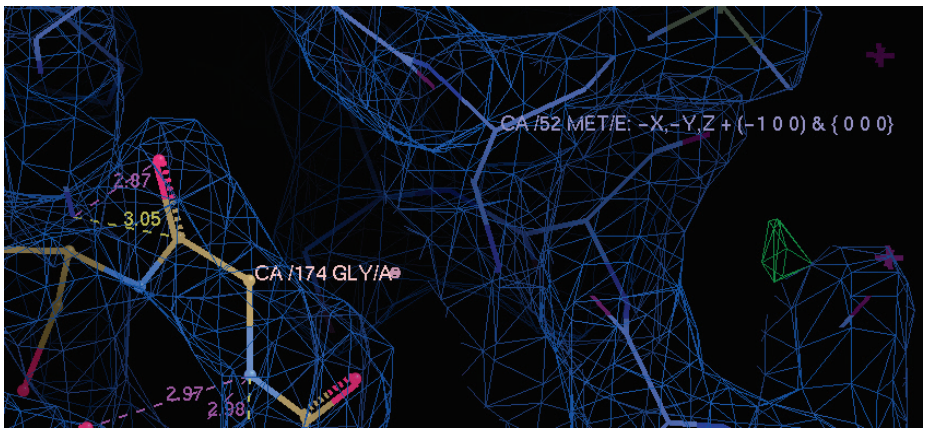


You can actually see (just – the colouring is a bit subtle) that the arginine is in a different chain because it has a different carbon atom colour. This is the nanobody that the EGFP is bound to, i.e. the interaction interface that we were looking at yesterday. It looks like the interface is very well ordered because the electron density is strong and clear. Have a look at some of the interactions that hold the nanobody and EGFP together.

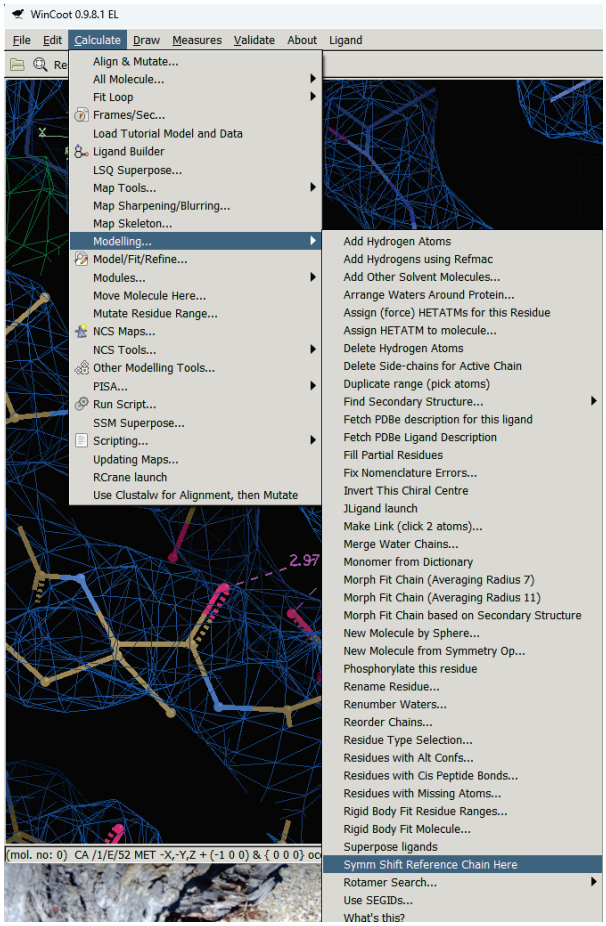
Chain B is just one of the four copies of EGFP in the asymmetric unit that was crystallised in this experiment (the other copies are chains A, C and D). COOT is clever about this and lets you hop between the same residue *on different chains* using the o key (as in 'o' for orange, between 'i' and 'p' on the keyboard!). Try it! You'll see that the conformation of the EGFP residue is similar in each chain, as is the interaction with the nanobody. This gives us some confidence that the interaction we're looking at is biologically meaningful and not an artefact of crystallisation. You can click the middle mouse button to centre on a residue in the other chains, or you can use the p key to centre on the residue that's closest to the centre of the screen.



You'll also see that for chains A and D the nearest nanobody chain is actually a symmetry-related one. For example, EGFP chain A interacts with the nanobody chain E:

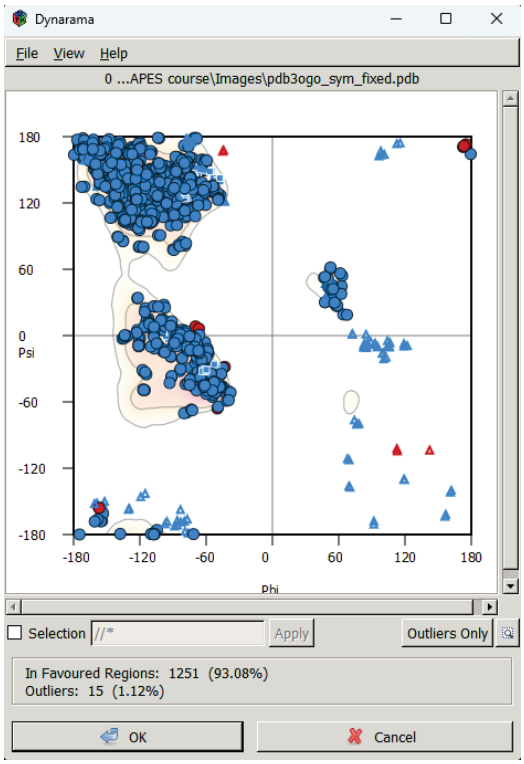


In COOT you can actually change which symmetry copy of the chain is in the PDB file, because it's more convenient to have the 'real' complexes close to each other. Try doing this for chains E and H by middle mouse clicking on a residue in the symmetry-related nanobody then shifting the reference chain to that position in space:



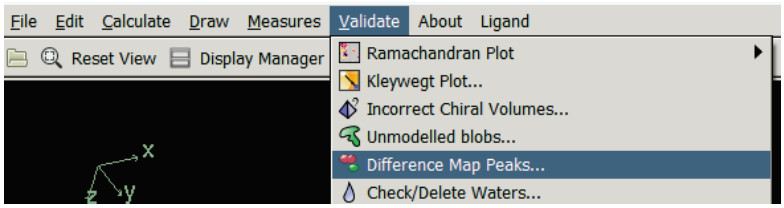
You can save this file for later use using **File > Save Coordinates**.

Let's now have a look at some of the structure validation tools in COOT. If you select **Validate > Ramachandran Plot** for the model you see that there are a number of outliers:

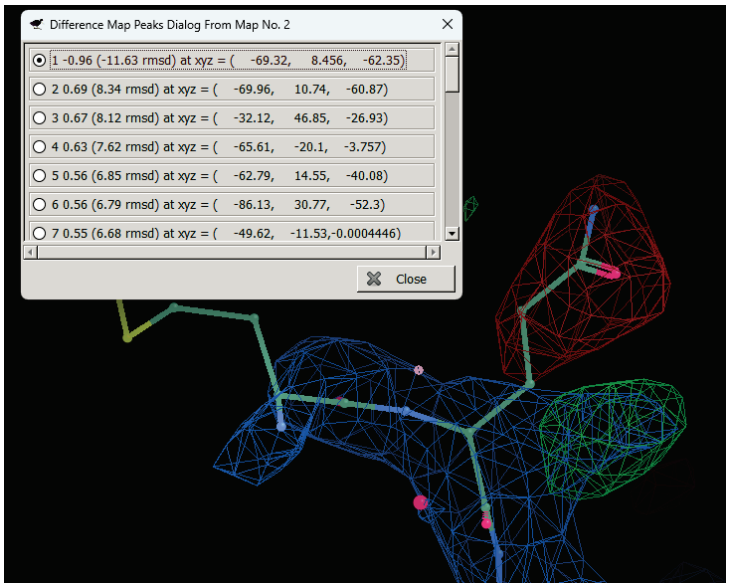


The fact that there are outliers doesn't 100% mean these residues are in the wrong position, but they warrant further inspection. Try clicking on one of the red shapes (outliers). COOT will take you to that residue where you can have a look at the density and judge for yourself – is the density strong? Is the model in the centre of the density? Does the model look like it's making sensible hydrogen bonds and non-bonded interactions?

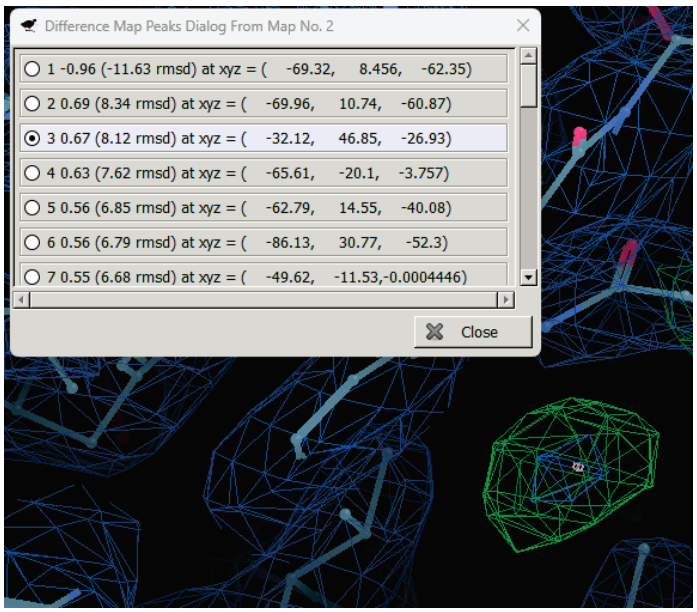
Next, try looking at the difference map peaks:



If you click on the first or second one you can see a pretty clear example of a side chain in the wrong position:



The third one down the list is clearly a water molecule that wasn't modelled



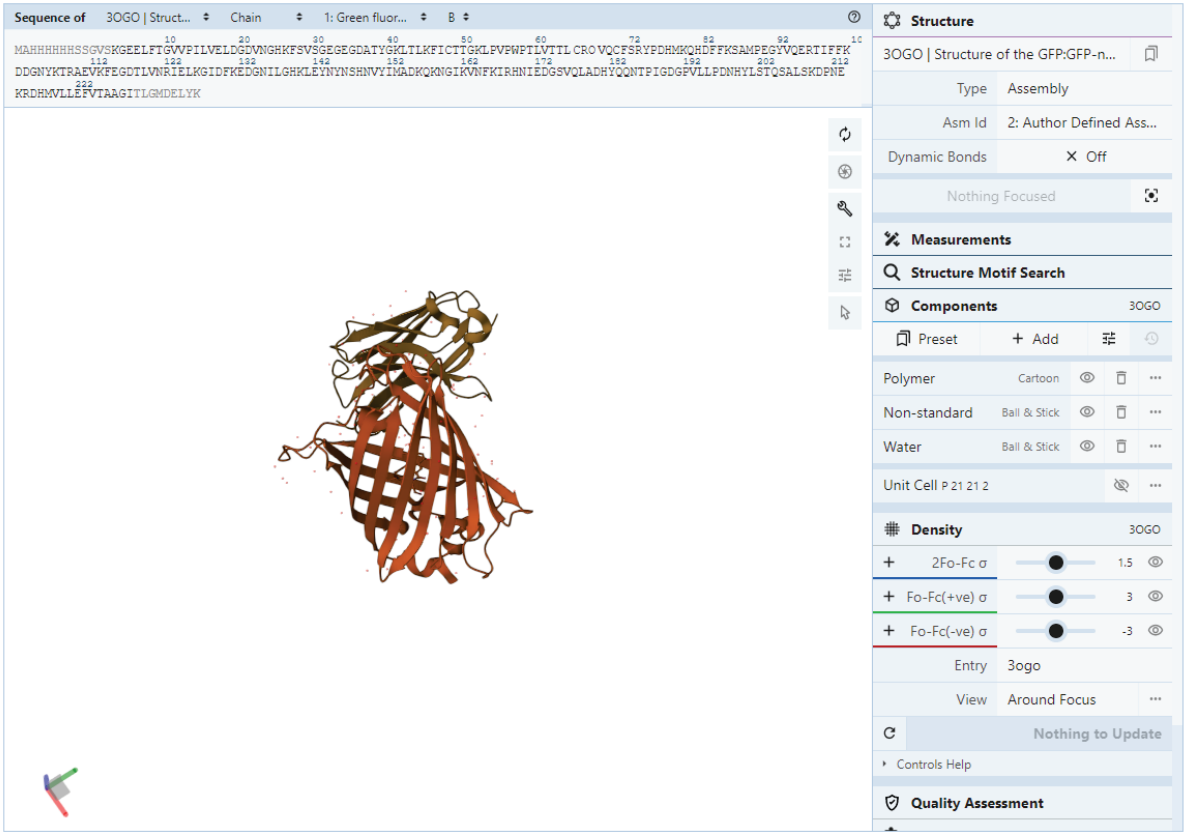
The point here isn't to make fun of the authors of this structure, but just to show again how it helps to look at the results of the experiment (the electron density) yourself before designing big experiments based on a published structure. But how do you get an overview of how good/bad a structure is? One easy way is to visit the page for that struture on the PDB website.

PDB website – summary reports, biological units and MOL*

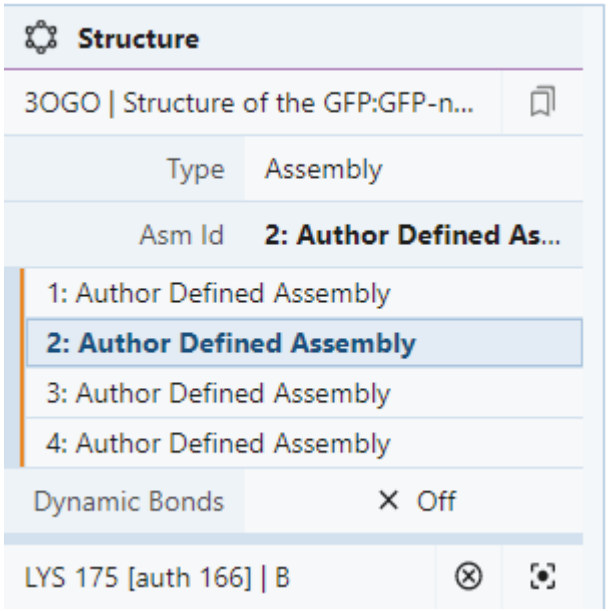
The PDB summary report accompanies every structure in the PDB – let's look at the one for our EGFP-nanobody complex (https://www.rcsb.org/structure/3OGO).

The summary graphic shows at a glance you how this structure compares to all the other structures in the PDB (black filled boxes) and to structures at the same resolution (black hollow boxes). You can see that the structure is a bit poorer than average for sidechain outliers and clashscores across the whole PDB, but is not so bad when you factor in the resolution of the structure (2.8 Å). If you click on the full report you can get a lot more detailed information but it's perhaps a bit too complex for the casual observer.

You can also inspect the structure directly via your web browser if you click on the 3D View tab:



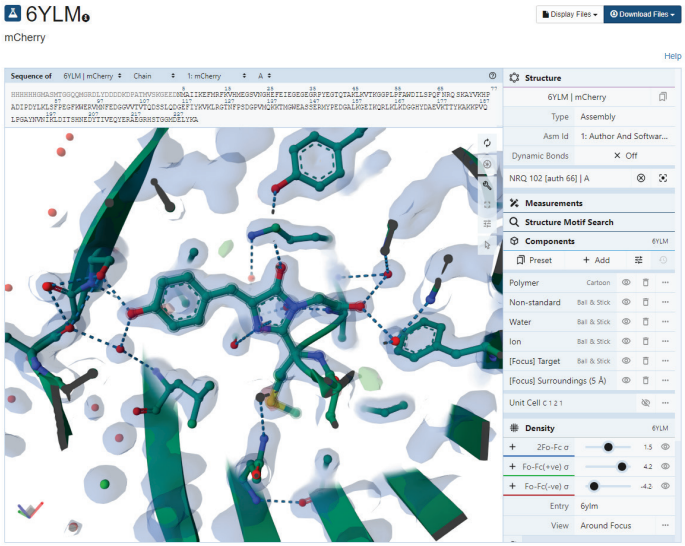
Try having a look at the structure using this browser-based viewer. For this structure you'll need to select "Author defined assembly 2" to see the interactions between chains B and G we were looking at in ChimeraX and COOT:



Left mouse button rotates, right mouse button translates, middle mouse zooms, Shift + left mouse sets the clipping planes, clicking on a residue centres on that residue and clicking into space zooms out. If you enable electron density it will only show up once you've clicked on a residue (focus area).

Now try going to the PDB pages for the other structures we looked at yesterday – are they high quality structures? Have a look at the density using the web-based viewer:

- EGFP (alone): https://www.rcsb.org/structure/2y0g
- mTurquoise2: https://www.rcsb.org/structure/6ylo
- mCherry: https://www.rcsb.org/structure/6YLM



Computer lab 3 – AlphaFold2 and ColabFold

Overview

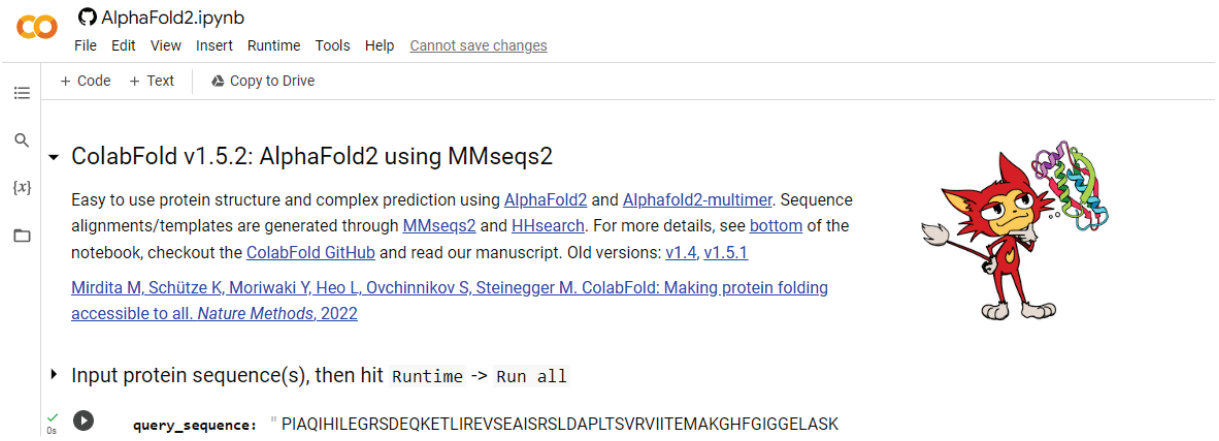
Today we will have a go at running AlphaFold2 and ColabFold via Google Colab. This is a free[1] service that lets you run programs on Google servers.

The AlphaFold2 colab notebook is available at:
https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb

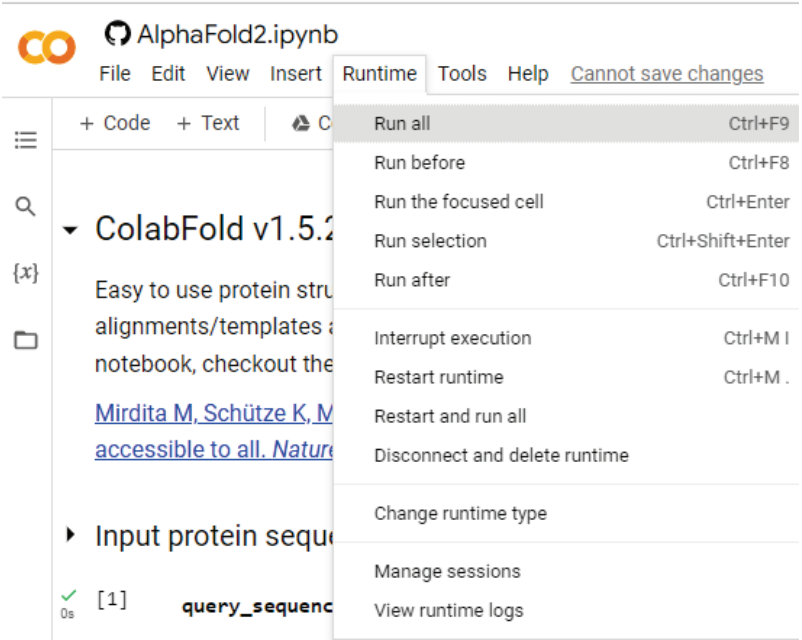
And the ColabFold colab notebook is available at:

https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb

For this tutorial we'll use ColabFold, because it's a bit quicker, but feel free to try both at a later date and compare the output. For this test we'll first predict the structure of the default query sequence (which is a 4-oxalocrotonate tautomerase enzyme from *Pseudomonas putida*):

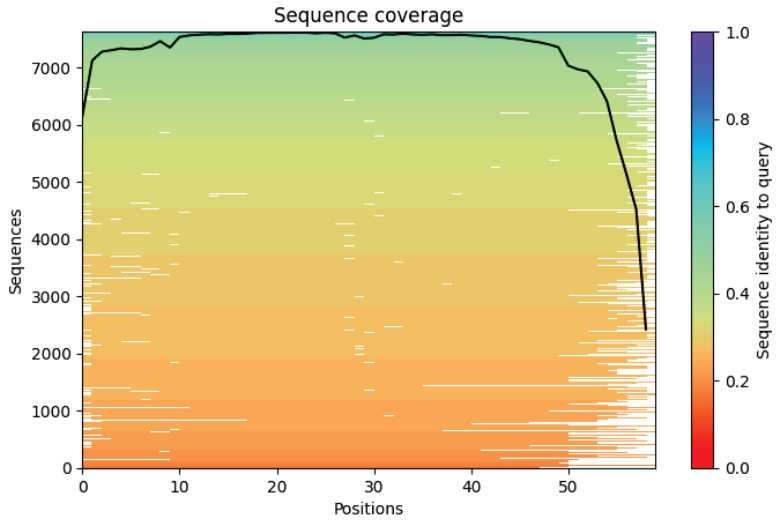


To start the prediction, just click on **Runtime > Run all**:

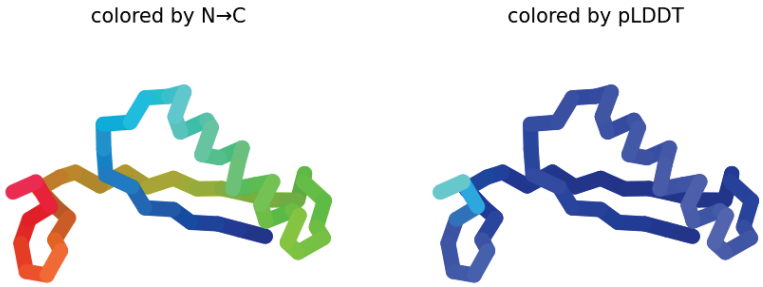


[1] Well, you can pay for ColabPro if you need to do lots of calculations, but the free version is fine for occasional use

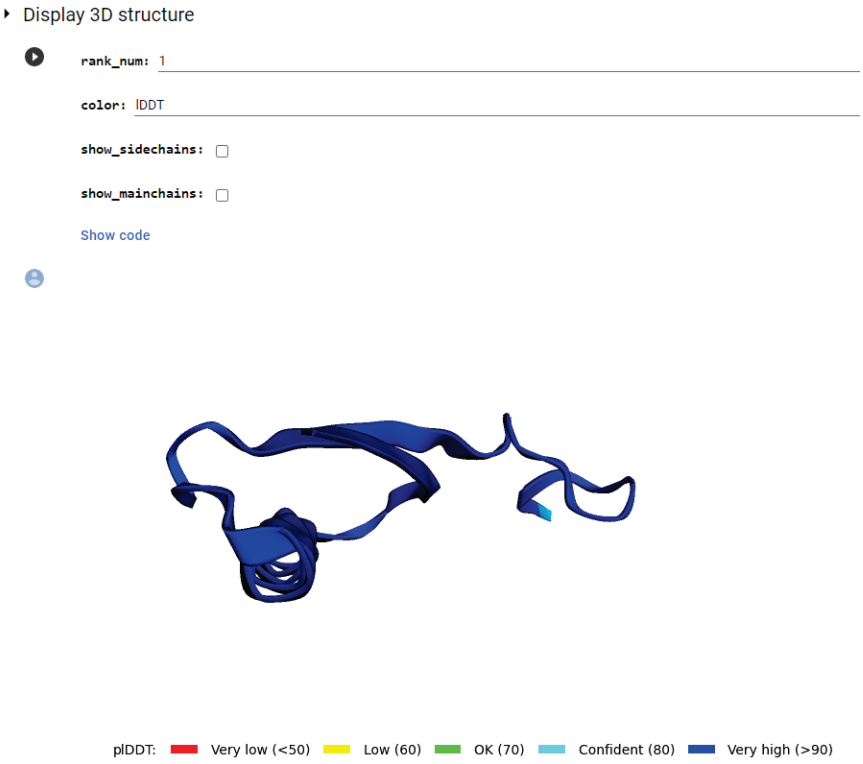
It takes about 5 min to predict the structure of this little protein, but as it goes you can see the multiple sequence alignment quality (on which the prediction depends):



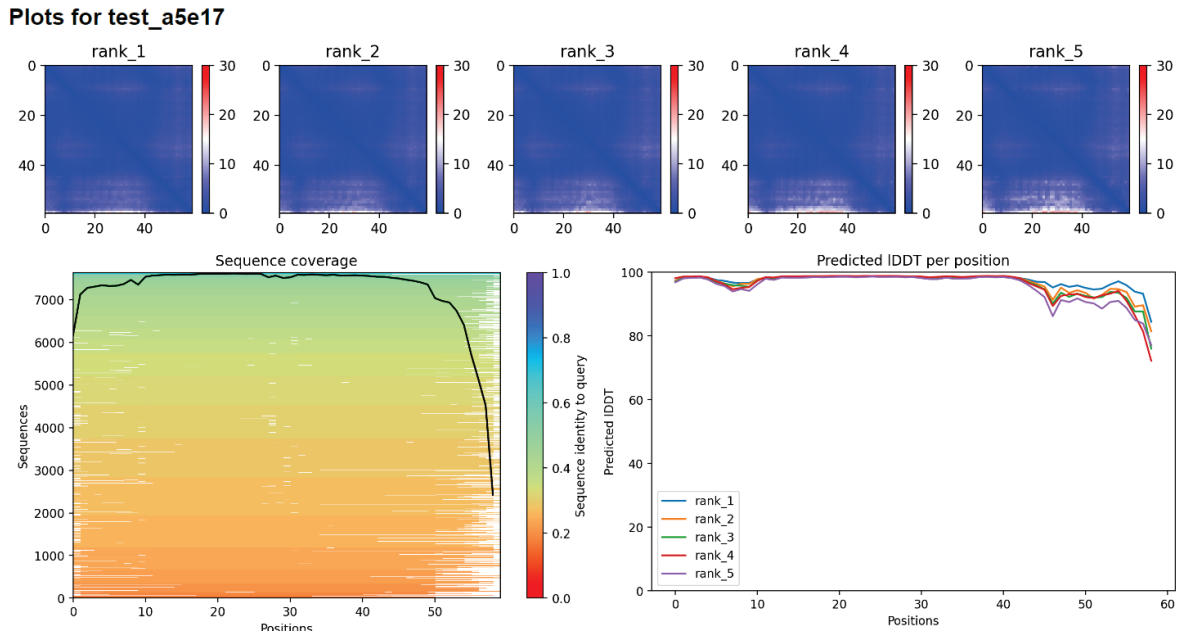
And you see the predicted structure at each iteration of the AlphaFold2 algorithm:



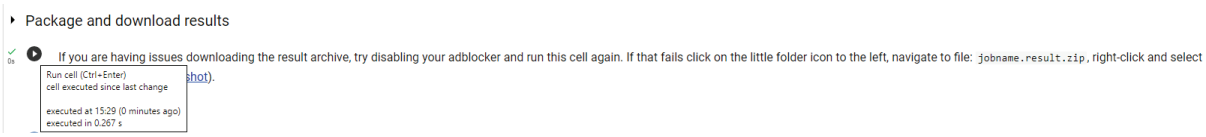
When the prediction finishes there is a simple interactive structure viewer:



And you then get the **all-important validation statistics** plotted:



This is a very high confidence prediction! The results should automatically be downloaded as a zip file, but you can re-download it by clicking on the **Run** button in the **Package and download results** code block (cell):



Simple!

Prediction of a complex

Now let's try something a bit more taxing – predicting the complex between the MIT domain of VPS4A in complex with MIM2 motif of CHMP6 (an important protein:protein interaction in membrane trafficking) from the species Ornithorhynchus anatinus (the platypus!).

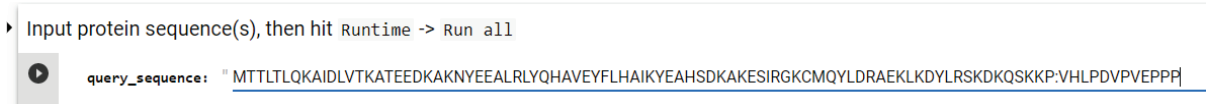
The sequences are below:

```
>XP_028930985.1:1-84 VPS4A [Ornithorhynchus anatinus]
MTTLTLQKAIDLVTKATEEDKAKNYEEALRLYQHAVEYFLHAIKYEAHSDKAKESIRGKCMQYLDRAEKLKD
YLRSKDKQSKKP
```

And:

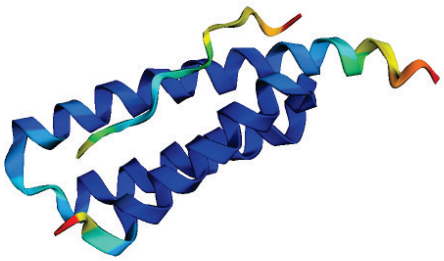
```
>XP_028912921.1:169:180 CHMP6 [Ornithorhynchus anatinus]
VHLPDVPVEPPP
```

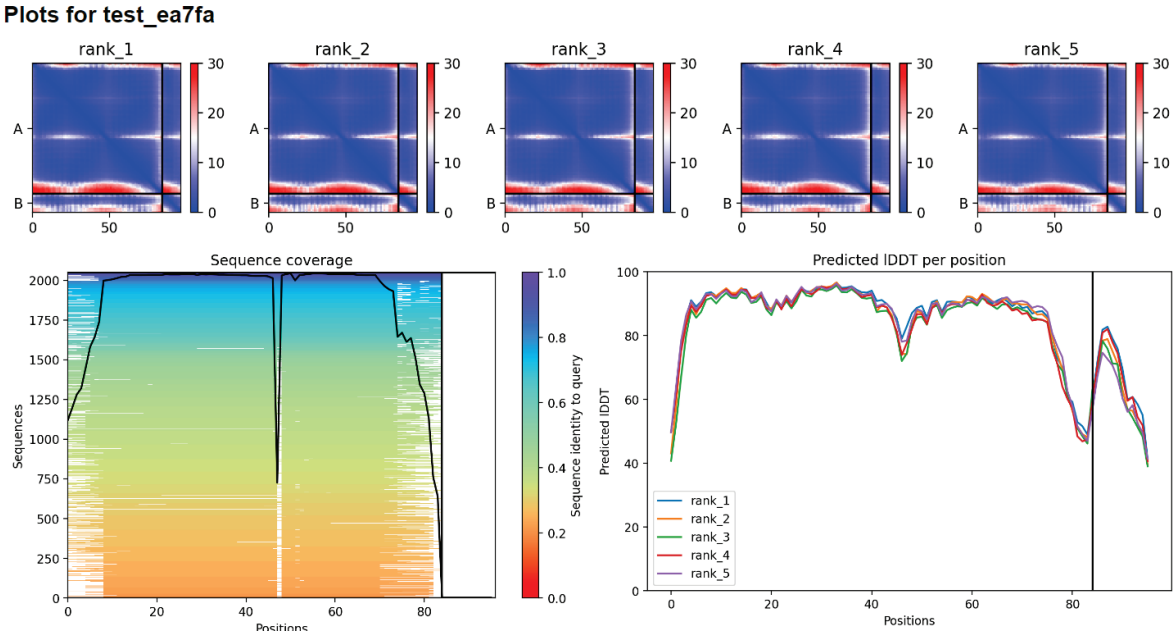
Put the two sequences into the **Input protein sequence(s)** cell separated by a colon (:)



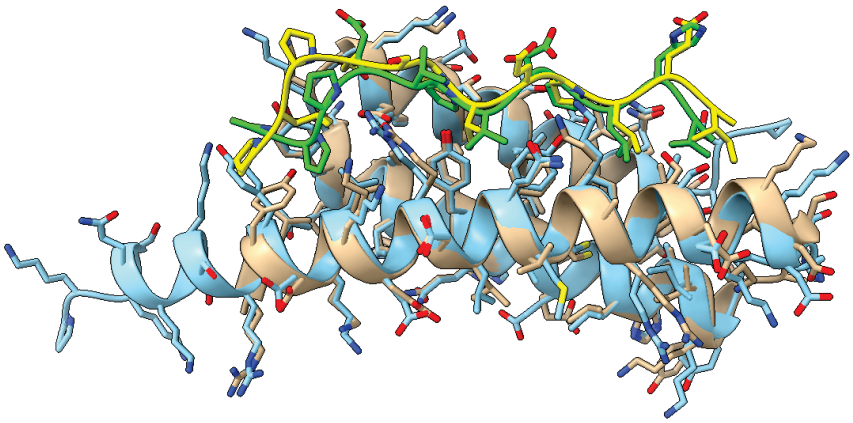
All the other defaults are sensible, so then press **Runtime > Run all** and wait a few minutes…

The model isn't super confident for the CHMP6 peptide, but overall the model is reasonably confident in the relative orientation of the peptide versus VPS4:





Based on what we know of VPS4:CHMP6 interactions this prediction is likely to be correct – if you wanted you could try downloading the complex and superposing it onto the structure of the human VPS4 MIT domain in complex CHMP6 (https://www.rcsb.org/structure/2K3W) to compare:



(Human is brown/yellow and platypus is cyan/green).

Now it's time to try a protein or complex of interest to you. Make sure you pay close attention to the statistics when analysing the results…