

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



**UNIVERSITY OF
CAMBRIDGE**

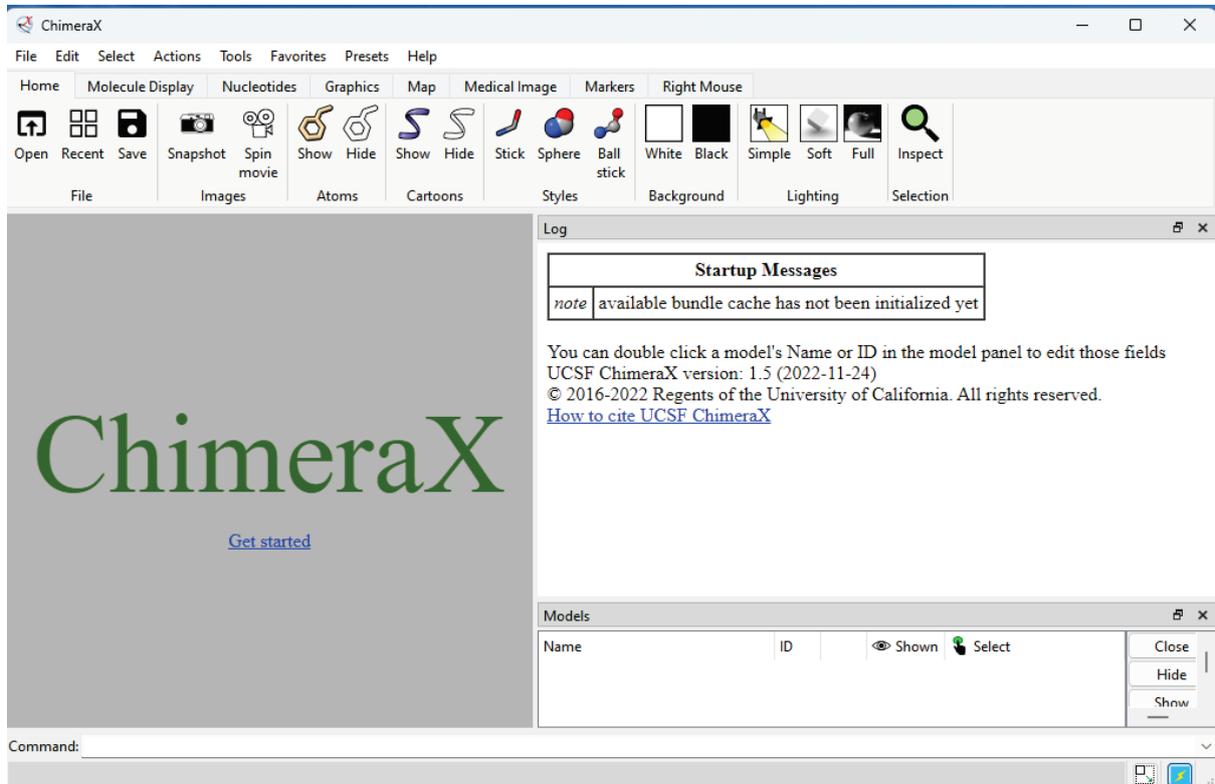
Computer lab 1 – Visualising proteins

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 ChimeraX¹, which you can download from <https://www.cgl.ucsf.edu/chimerax/download.html>.

Visualising EGFP

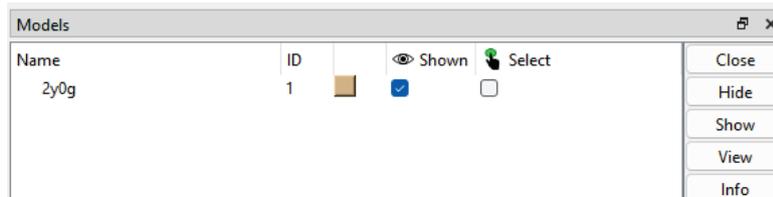
Open ChimeraX²



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

```
open 2y0g
```

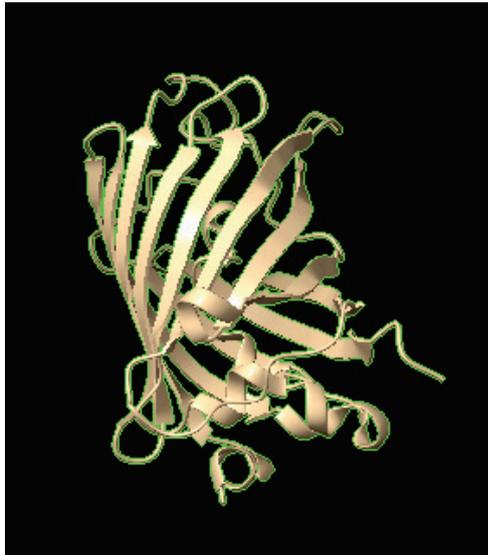
This has opened the protein with identifier #1, as you can see in the **Models** panel. You're currently looking at the protein in 'cartoon' view, showing the protein backbone:



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

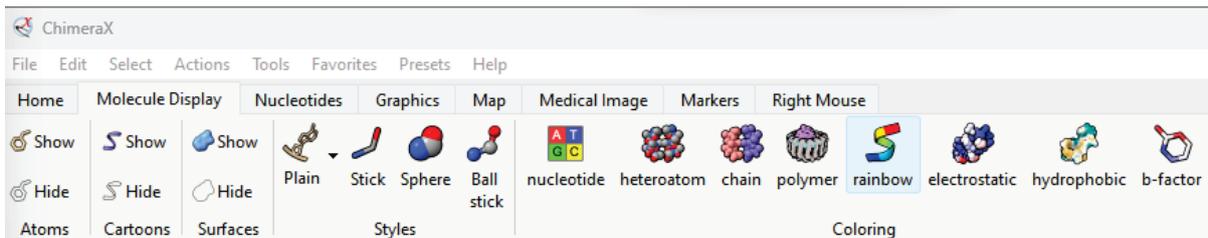
You can use the models panel to hide or view different structures or maps, and to select models. For example, try clicking on 'select'. You see that the whole model now has a green outline because it is selected.



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

```
Log
select add #1
2067 atoms, 1846 bonds, 490 residues, 1 model selected
```

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:

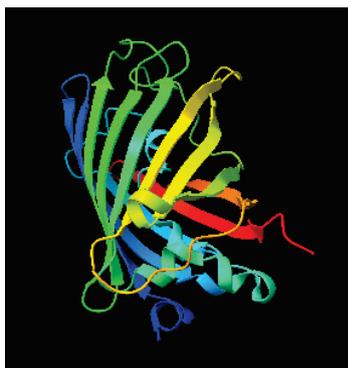


You can also see the command that was used in the log:

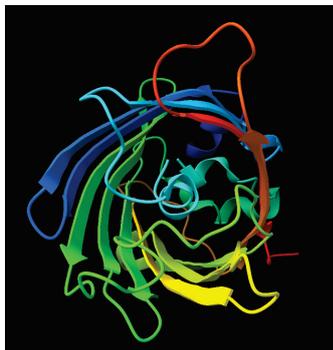
```
rainbow sel
```

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
```

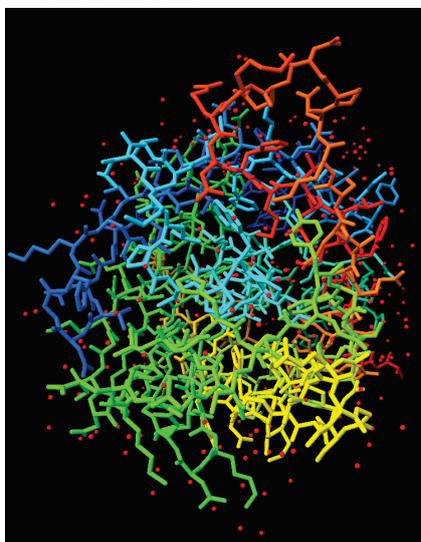
You can then click on the entries you see in the **Log** panel to return to those views:

[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
```



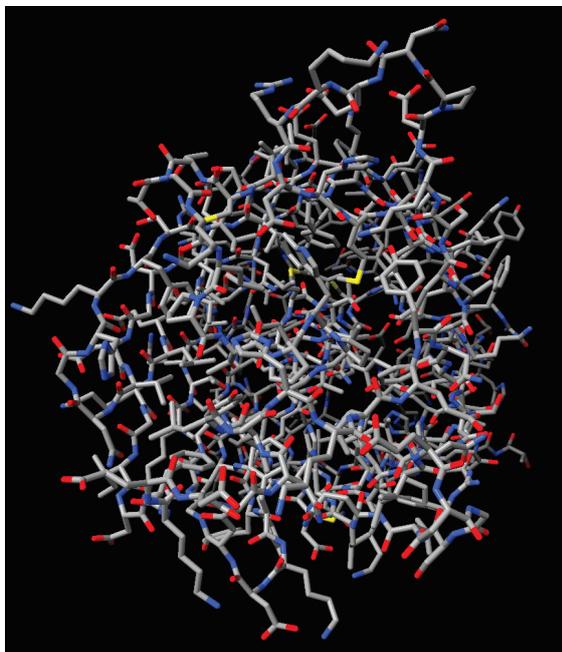
That's a lot more complicated! All the red dots are the water molecules that were ordered in the molecule when its structure was solved. Let's remove them for the moment:

```
delete solvent
```

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

```
color byelement
```

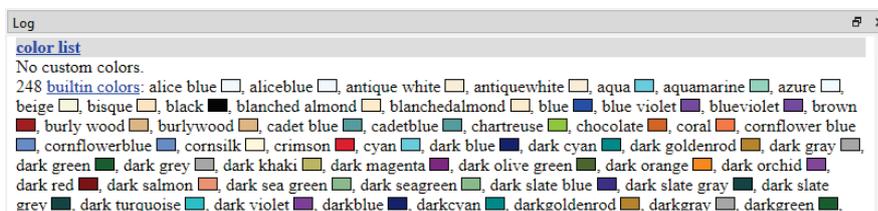
Now we see all the atoms coloured by atom type (blue = nitrogen, red = oxygen, yellow = sulfur, carbon=grey).



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:

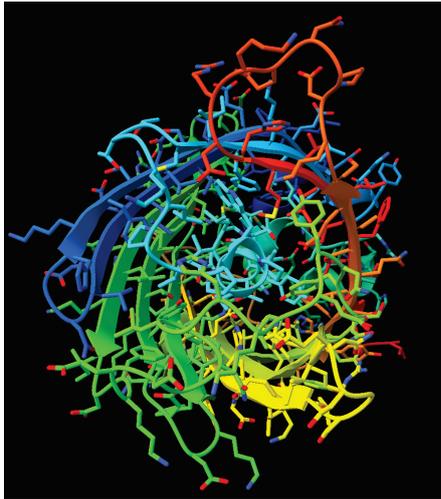


That's a lot to choose from. Select a colour for your carbon atoms (I'm going for a light green):

```
color C lime
```

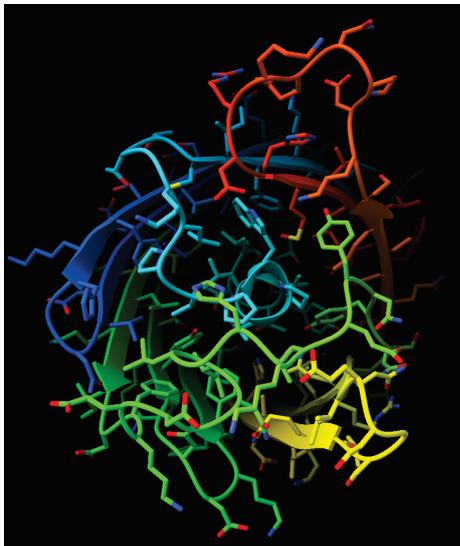
Have a look at the molecule – it's a bit hard to navigate when there is so much going on, so let's show the cartoon again but keep the side chains of the amino acids and show the rainbow view again, but this time colour all the atoms by atom type *except for the carbon atoms*:

```
cartoon
rainbow
color byhetero
```



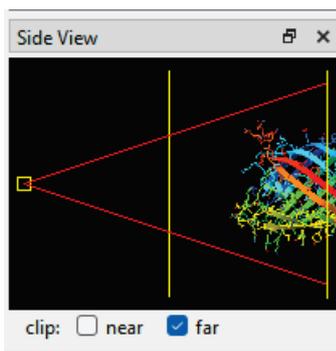
Great. Now we're getting a nicer picture, but let's try and reduce the clutter by 'clipping' the view so that we reduce the depth of field of the image. First off, we can try and set the far clipping plane so residues near the back of the molecule fade to darkness

```
clip far 5
```

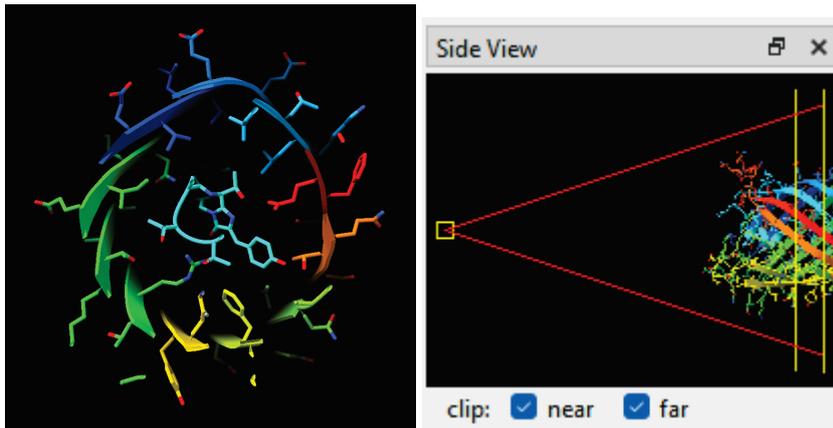


It can become difficult to understand what's happening with clipping planes, but thankfully ChimeraX includes a handy way of visualising the clipping planes that are being applied to your molecule via the Side View panel:

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



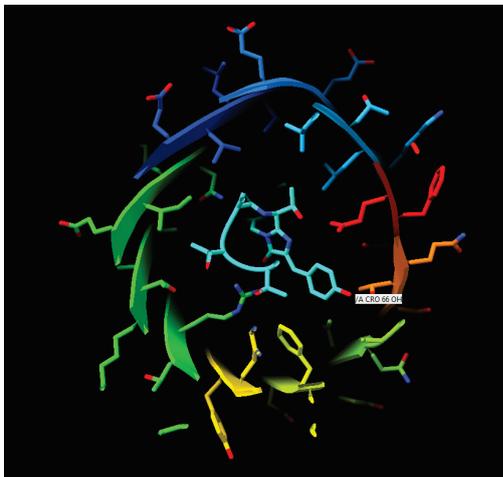
You can use this tool to turn on the near clipping plane (tick the box) and adjust its position by clicking on the yellow horizontal line and dragging. Try moving the clipping plane so that you can clearly see the tyrosine-like residue in the middle of the beta barrel:



Let's save this view tool:

```
view name barrel-closeup
```

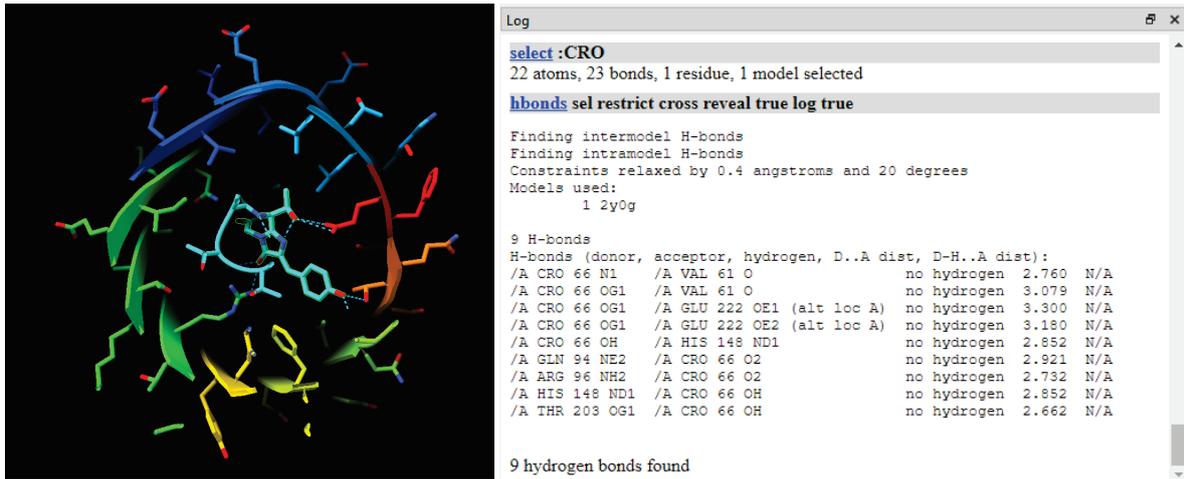
If you click on any atom in this residue you can see that it isn't actually a tyrosine residue, it's a Gly-Tyr-Gly chromophore (named 'CRO' in the structure) and this residue is key to the fluorescence of EGFP!



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.

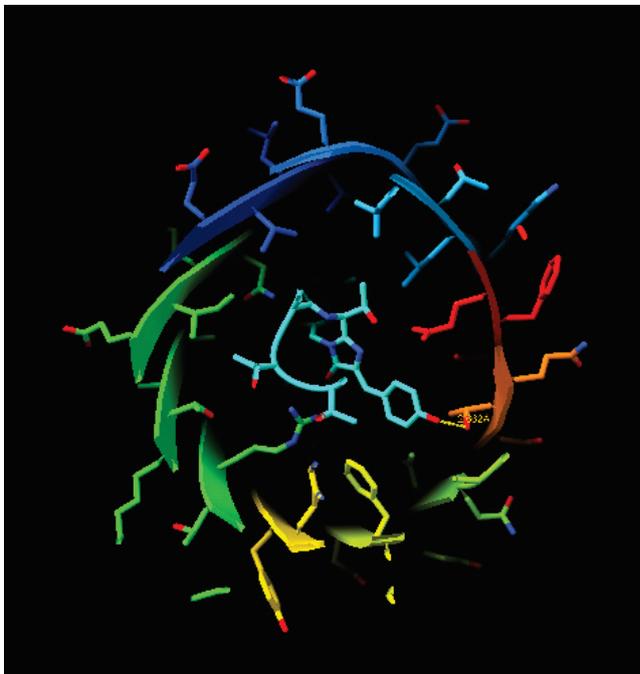


You can see that the hydrogen bond between the terminal hydroxyl (OH) of CRO and the hydroxyl group of threonine 203 is 2.66 Å. Let's show just that bond by deselecting the CRO residue (residue number 66), hiding all the hydrogen bonds, then drawing just that bond manually and showing the distance between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

```

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

```

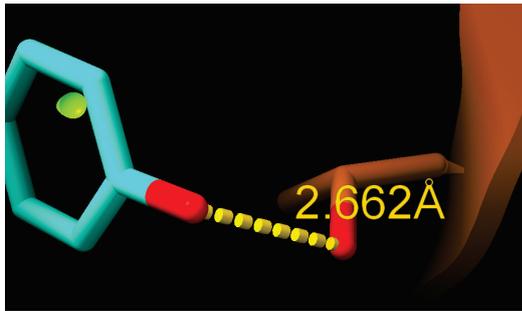


It's a bit hard to see, so let's zoom in on that:

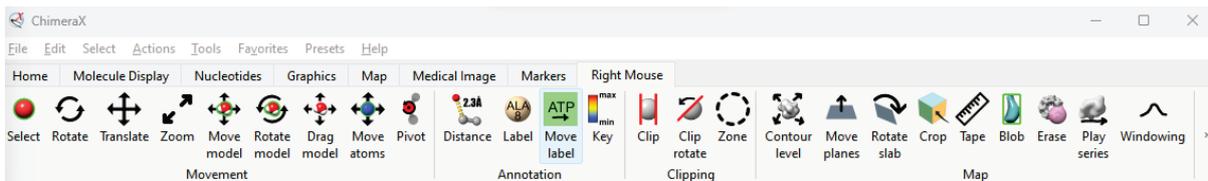
```

view :66@OH :203@OG1

```

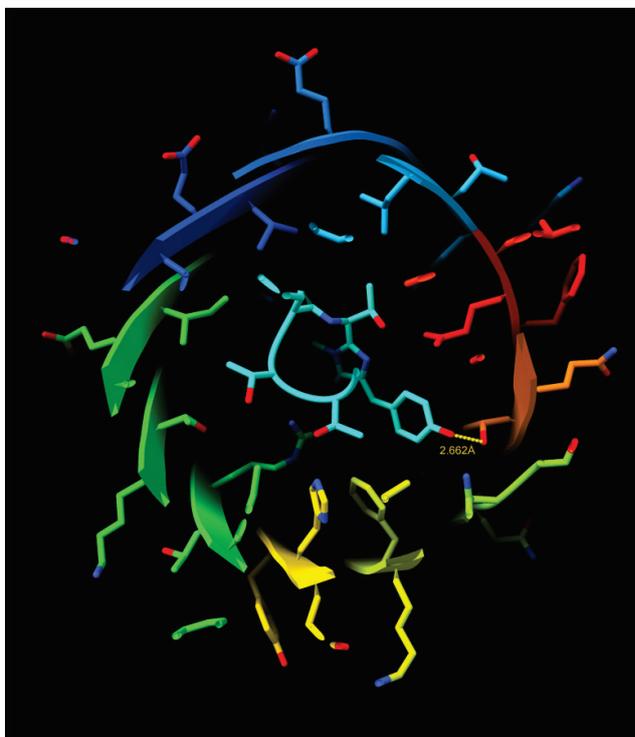


That’s a bit too close! Try zooming out and changing the clipping planes until you have a nice view of the chromophore and the hydrogen bond. You can also move the label to a better position using the ‘move label’ mouse mode – this lets you click and drag on the label with the left mouse button to move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

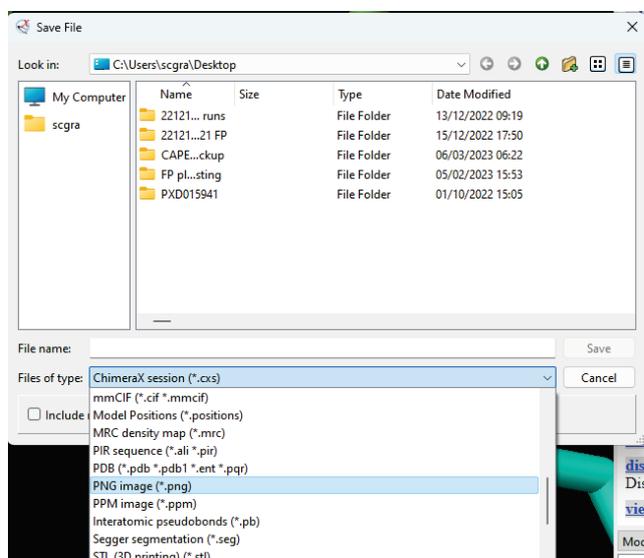
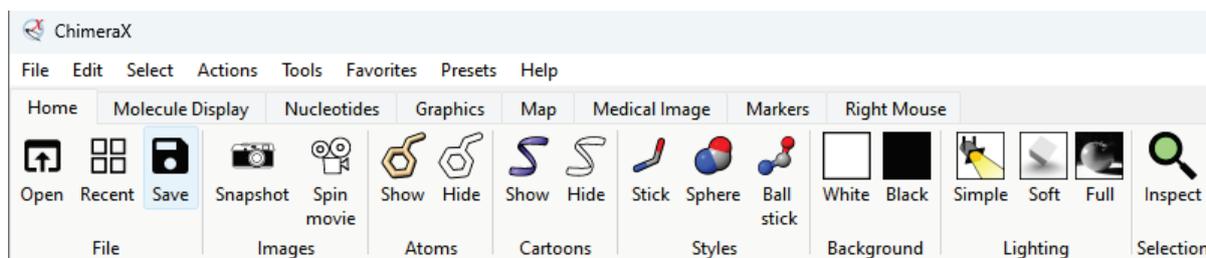


Once you have a nice view, you can save it and you can save an image of the H-bond to your desktop:

```
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):

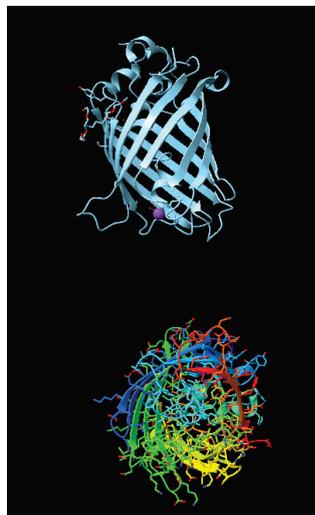


Don't forget to also save the view in case you want to return to the same orientation later:

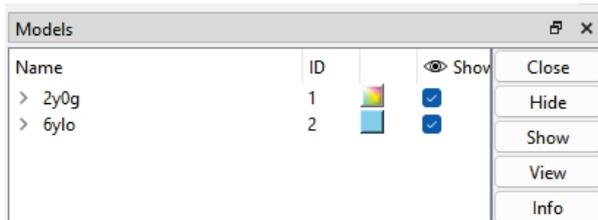
```
view name CRO_H-bond
```

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
```

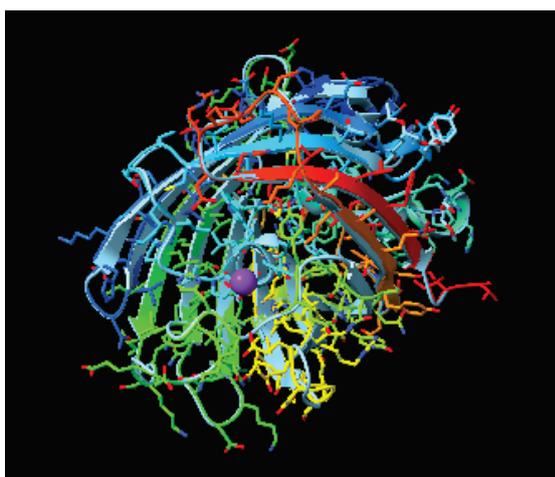


You see that we now have two models in the **Models** panel. They are named models 1 and 2, so you can change the view (etc) of each one individually:



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
```



Ca RMSD		1	11	21	31	41
2y0g, chain A	MAHHHHHHGHHHQLVSKG	EELFTGVVPI	LVELDGDVNGHKFSVSG			
6ylo, chain AMTSKG	EELFTGVVPI	LVELDGDVNGHKFSVSG			
Ca RMSD		46	56	66	76	86
2y0g, chain A	E	EGGDATYGKLLKFI	CTTGKLPVPWPTLVTTL	XVQCFSRYPDHM		
6ylo, chain A	E	EGGDATYGKLLKFI	CTTGKLPVPWPTLVTTL	XVQCFARYPDHM		
Ca RMSD		91	101	111	121	131
2y0g, chain A	K	QHDFFKSAMPEGYVQERT	I	FFKDDGNYKTRAEVKFE	GD	TLVNR I
6ylo, chain A	K	QHDFFKSAMPEGYVQERT	I	FFKDDGNYKTRAEVKFE	GD	TLVNR I
Ca RMSD		136	146	156	166	176
2y0g, chain A	E	LKGIDFKEDGNILGHKLE	YNYNSHN	VYIMADKQKNGIKVNF	KIR	
6ylo, chain A	E	LKGIDFKEDGNILGHKLE	YNYNSHN	VYIMADKQKNGIKVNF	KIR	
Ca RMSD		181	191	201	211	221
2y0g, chain A	H	NI	EDG	SVQLADHYQQNTPI	GDGPVLLPDNHYLSTQ	SALSKDPNE
6ylo, chain A	H	NI	EDG	SVQLADHYQQNTPI	GDGPVLLPDNHYLSTQ	SALSKDPNE
Ca RMSD		226	236	246	256	
2y0g, chain A	K	RDH	MVLL	EFVTAAGIT	LG	MDELYK
6ylo, chain A	K	RDH	MVLL	EFVTAAGIT	LG	MDELYKHHHHHH

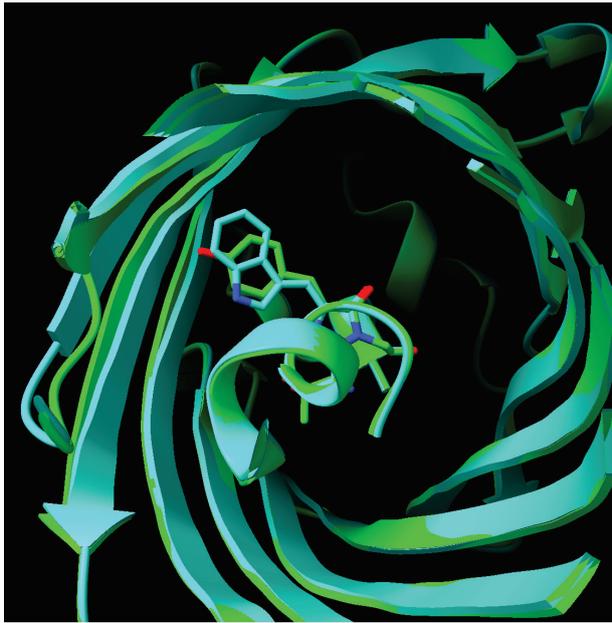
This seems to have done a good job – you can see that the EGFP and mTurquoise2 were both expressed as His-tagged proteins, but the tags were at opposite ends of the protein. Note also that the CRO residue of EGFP is called ‘X’ in the alignment because it’s a non standard amino acid. If you hover your 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:

Ca RMSD		01	101	6ylo #2/A SWG 65
2y0g, chain A	I	CTTGKLPVPWPTLVTTL	X	VQCFSRYP
6ylo, chain A	I	CTTGKLPVPWPTLVTTL	X	VQCFARYP

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!)

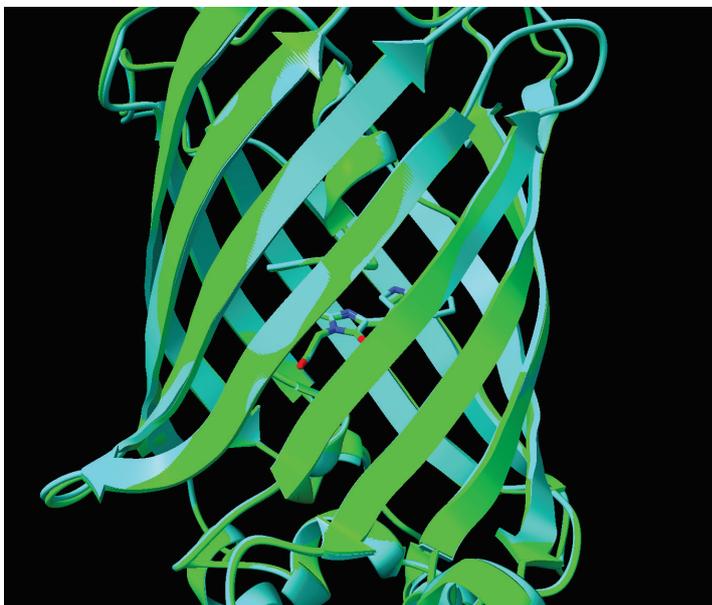
You can get the current camera matrix, which defines the view, and model matrices, which define any movements of the models, using the command

```
view matrix
```

view matrix

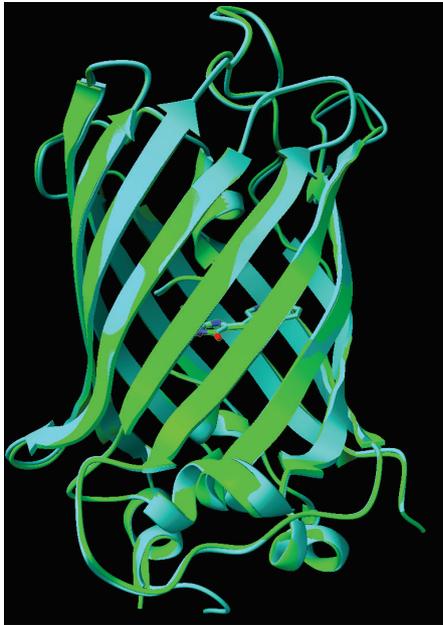
```
view matrix camera 0.68567,-0.20429,-0.69865,-46.103,0.56893,0.74912,0.33932,63.339,0.45406,-0.63015,0.62988,103.32
view matrix models #1,1,0,0,0,1,0,0,0,1,0,#2,-0.2557,0.38754,-0.88568,37.654,0.91298,0.39808,-0.0894,42.209,0.31792,-0.83147,-0.45561,33.231,#2,1,1,0,0,0,1,0,0,0,1,0
```

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.



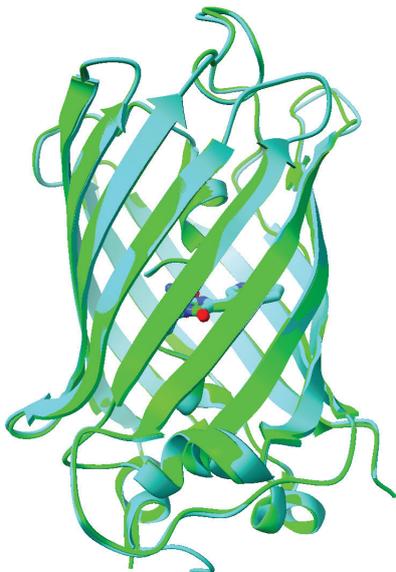
That's a bit close, so zoom out a bit and set the clipping planes to give more depth of field, and save the view in case we want to return to it:

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

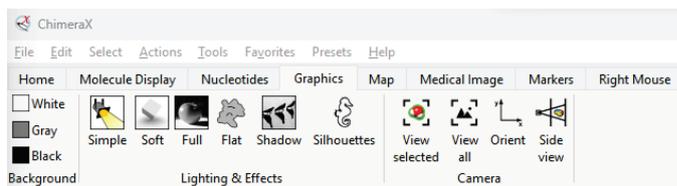


Let's make a figure where the bonds of the chromophores are a bit fatter and the background is white:

```
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:



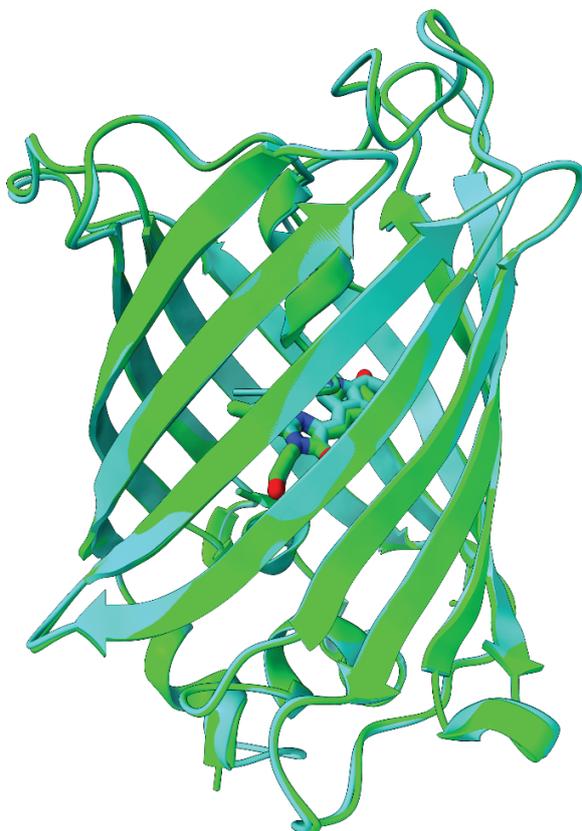
(Note that you can follow what commands are being performed in the **Log** panel). I like the following:

```
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.

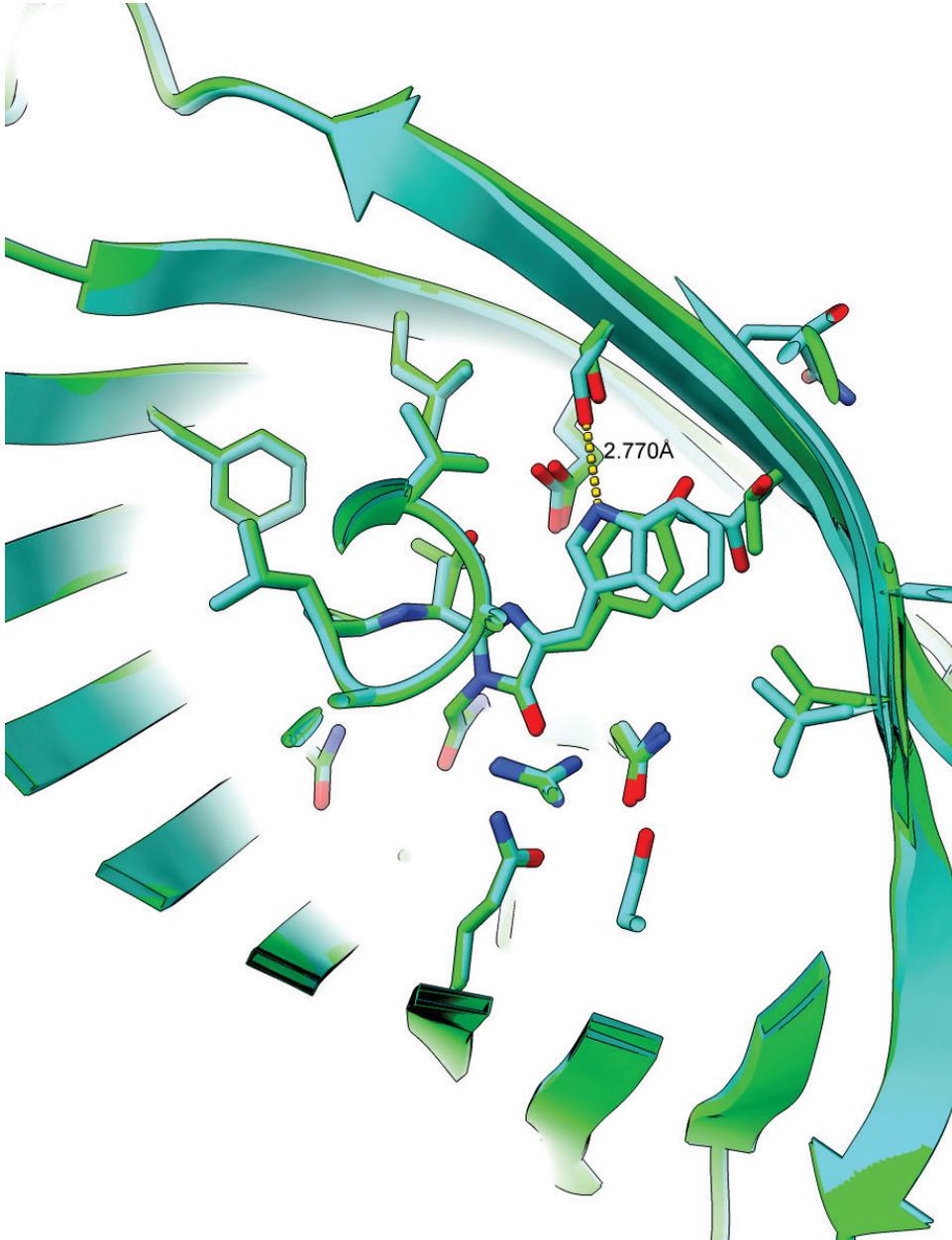


Lastly, let's look at the environment of the chromophores. Reset the bond width to the default (0.2 Å wide):

```
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
```

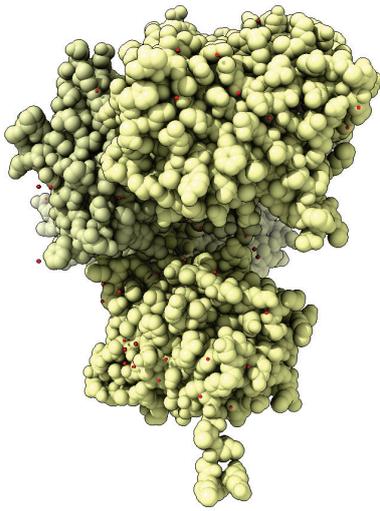
And close the ChimeraX session, ready for the next exercise:

```
close
```

Inspecting the EGFP:nanobody interaction

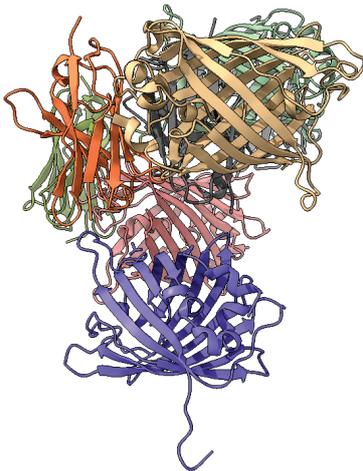
Next, we'll inspect the interaction between EGFP and the anti-GFP nanobody that we purified.

```
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

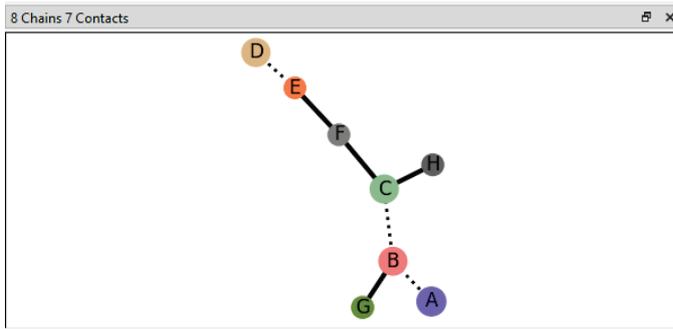
This information is also in the Log panel when you open the structure:

3ogo title:
 Structure of the GFP:GFP-nanobody complex at 2.8 Å 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:

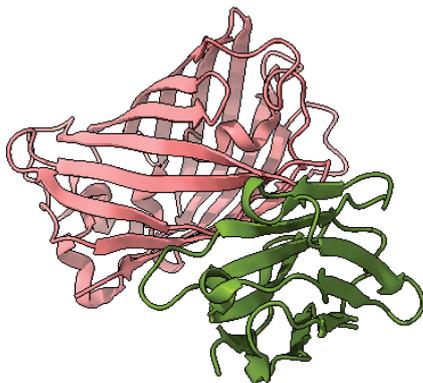
```
interfaces #1 & protein
```



```
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
```

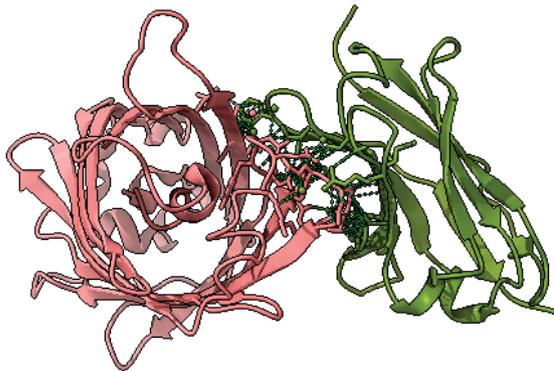
The log tells us the buried surface area between each molecule, and the interfaces with the most buried area are likely to be the biologically meaningful ones. In this case, CF and BG are likely to be biologically relevant interactions, whereas the others are likely not meaningful. Let's focus on BG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

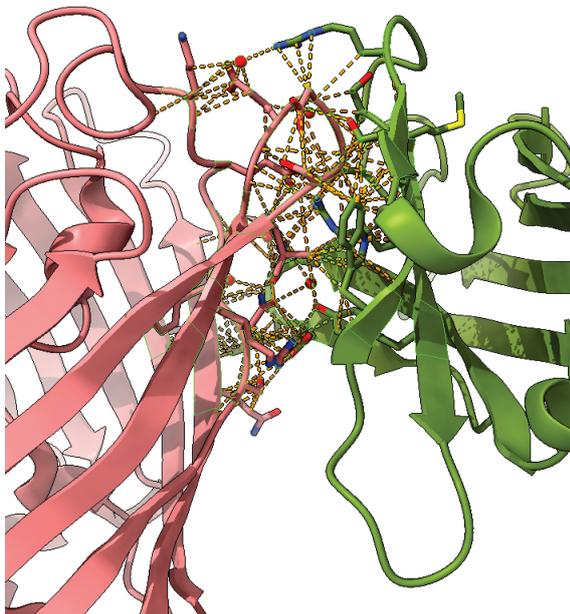
Allowed overlap: -0.4
H-bond overlap reduction: 0.4
Ignore contacts between atoms separated by 4 bonds or less
Detect intra-residue contacts: False
Detect intra-molecule contacts: True

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:



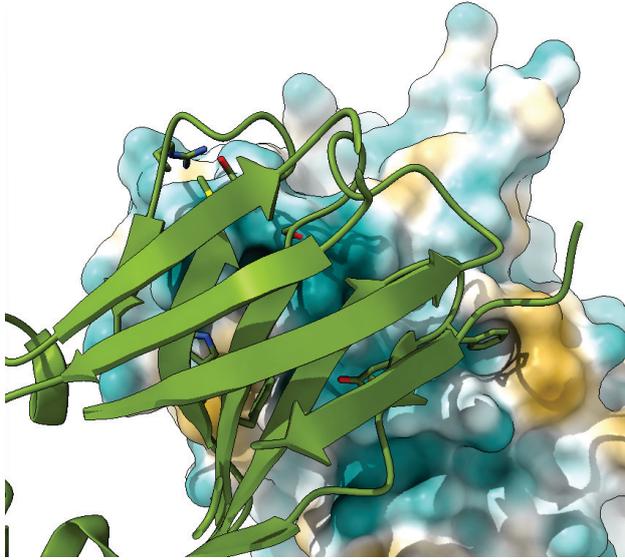
We can look whether there are any clashes between EGFP and the nanobody:

```
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):

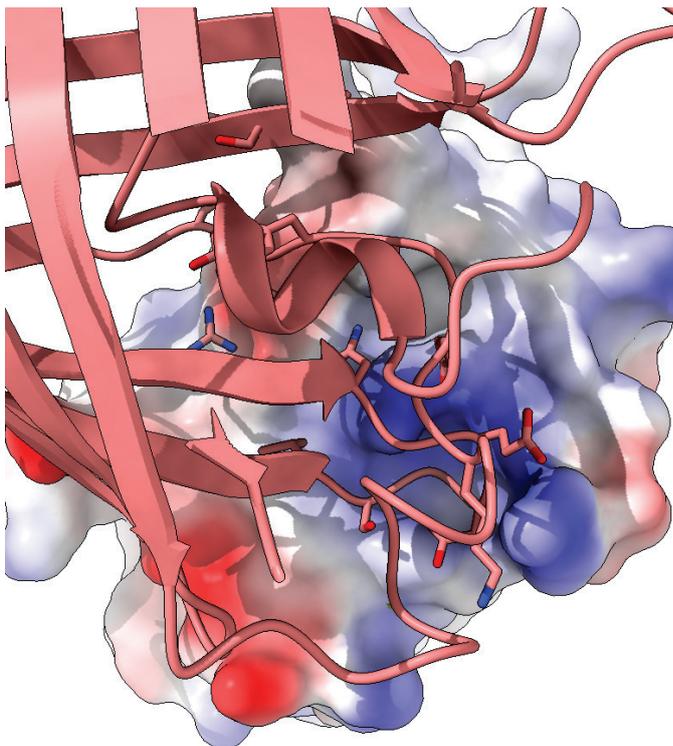
```
~contacts
sel clear
hide solvent
mlp EGFP
```

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



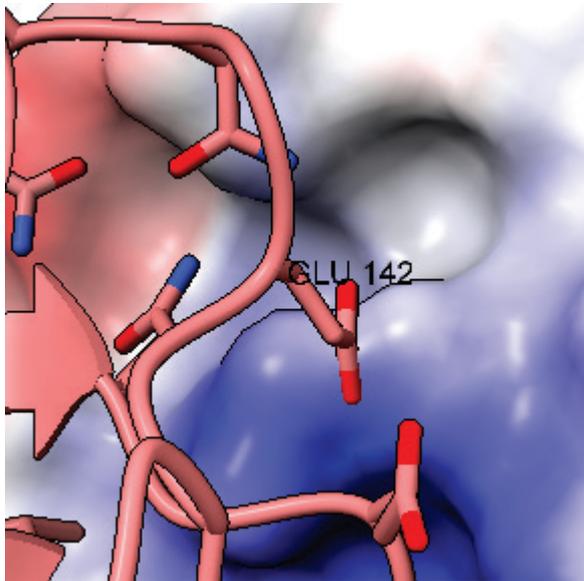
Blue is more hydrophilic and gold is more lipophilic. Have a careful look at the surface, you can see both polar and hydrophobic interactions. Let's now look at the electrostatics of the nanobody (ChimeraX currently struggles to calculate electrostatics for the EGFP because of the CRO residue – it confuses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

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



Blue is positively charged and red negative¹. 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
```

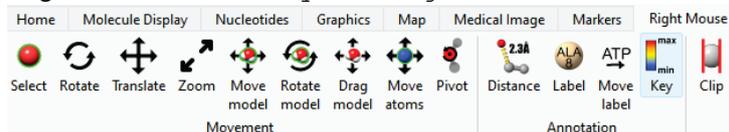


But the real interest is why some fluorescent proteins were bound by the nanobody and others weren't, so lets load the structure of mCherry and align it to EGFP:

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

Ca RMSD	1	11	21	31	41
3ogo, chain B	MAHHHHHH			SSGSVSKG	EELFTG...VV
6ylm, chain A	HHHHHHG	MASMTGG	QQMGRDLY	DDDDKDP	ATMVSKG
Ca RMSD	51	61	71	81	91
3ogo, chain B	PILVELDGD	VNGHKFS	VSGEGEG	DATYGKLT	LKFICT.TGKLPVPWPTL
6ylm, chain A	RFKVHMEG	SVNGHEF	EIEGEGE	GRPYEGT	QTAKLKVT.KGGPLPFAWDILS
Ca RMSD	101	111	121	131	141
3ogo, chain B	VTTLVQC	FSRYPDH	MKQH	DFFKSAM	PEGYVQERTIFFKDDGNYKTR
6ylm, chain A	PQFXSKA	YVKHPAD	P.....	DYLLKLS	FPEGFKWERVMNFEDGVVTVT
Ca RMSD	151	161	171	181	191
3ogo, chain B	AEVKFEG	DTLVNRI	ELKGI	DFKEDG	NILGHKLEYNYN
6ylm, chain A	QDSSLQ	DGEFIYK	VKLRGT	NFSPDGP	V.MKKTMGWEASSERMYPE..DG
Ca RMSD	201	211	221	231	241
3ogo, chain B	GIKVNF	KIRHNI	EDGSV	QLADHY	QQNTPI.GDGPVLLP
6ylm, chain A	ALKGEI	KQRLK	LKDG	GGHYDA	EVKTTYKA..KKPVQLPGAYNVN
Ca RMSD	251	261	271		
3ogo, chain B	DPNEKR	DHMLV	LEFVTA	AGITL	GMDELYK
6ylm, chain A	H.NEDY	TIV	VEQY	ERA	EGRHSTGGMDELYKA

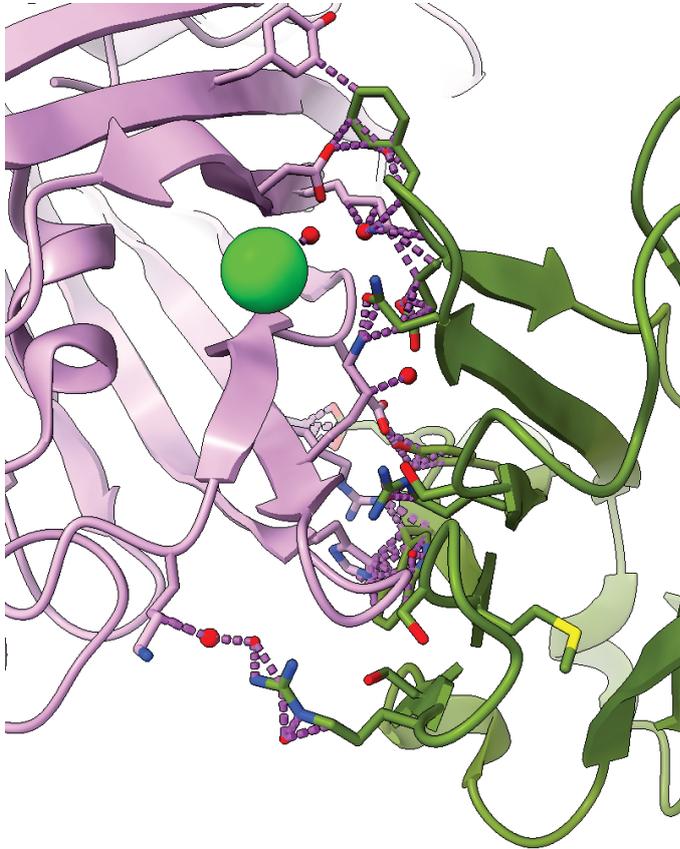
¹ You can add a legend (colour key) for the electrostatics by adding the text key true to the coulomb command, or by clicking Annotation > Key in the Right Mouse tab of the Ribbon:



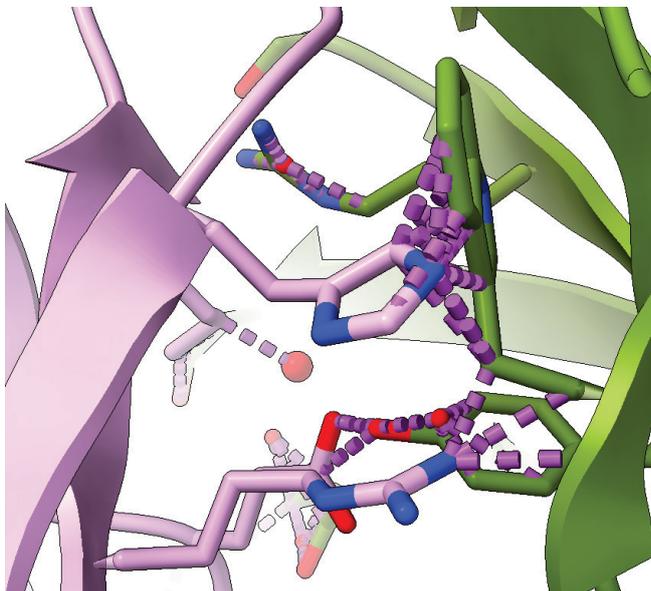
These proteins aren't as closely conserved as EGFP and mTurquoise2. We can have a quick look at why we think mCherry might not bind the nanobody. Are there clashes between the proteins?

```

~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

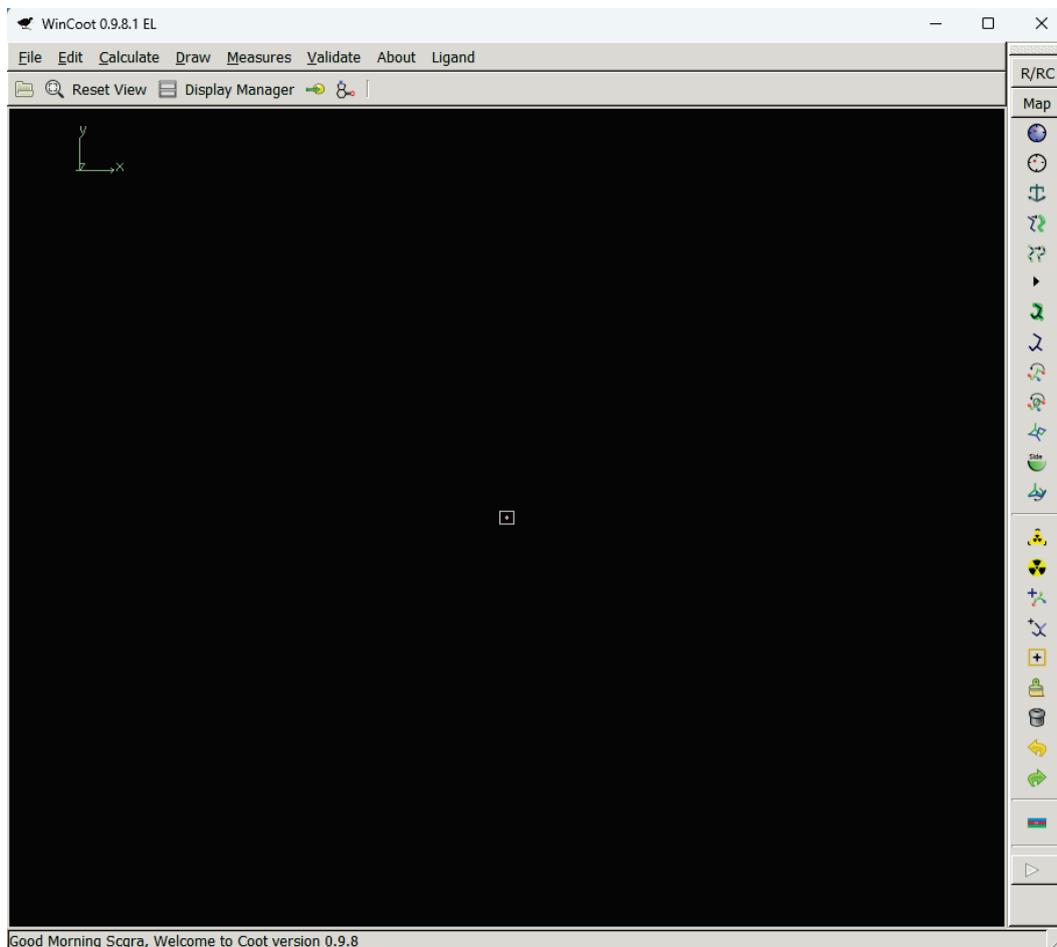
Overview

Today we will re-look at the structures of EGFP in complex with the anti-GFP nanobody, but using COOT¹ 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

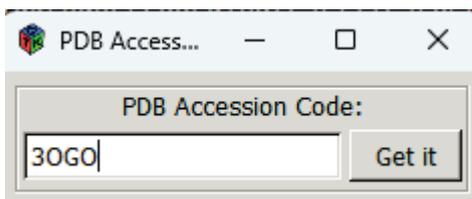
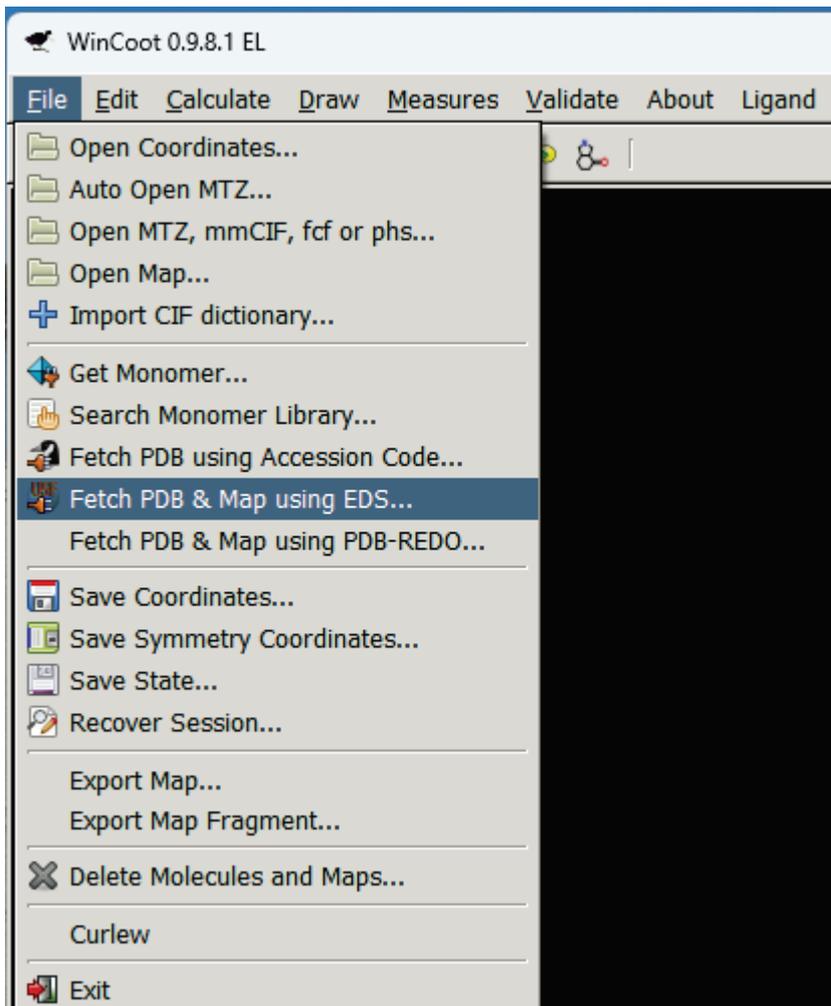
Open COOT²



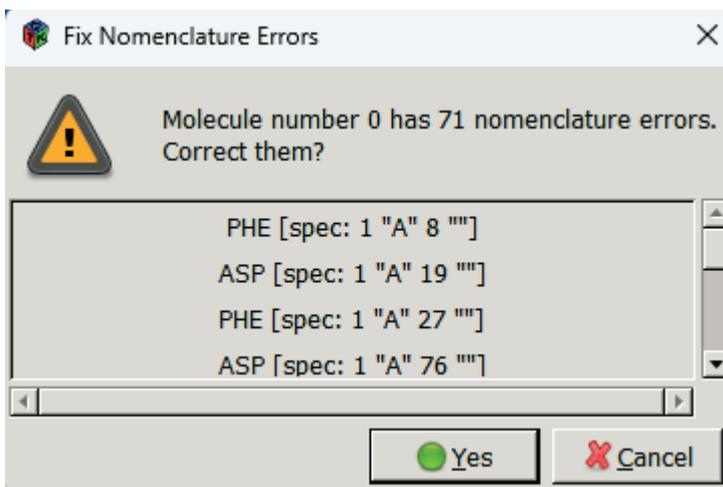
Use the build-in data retrieval tools to download the structure of the complex, plus associated experimental data (electron density maps), from the Electron Density Server hosted by the EBI (<https://www.ebi.ac.uk/pdbe/>) using the **File > Fetch PDB & Map using EDS...** tool:

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

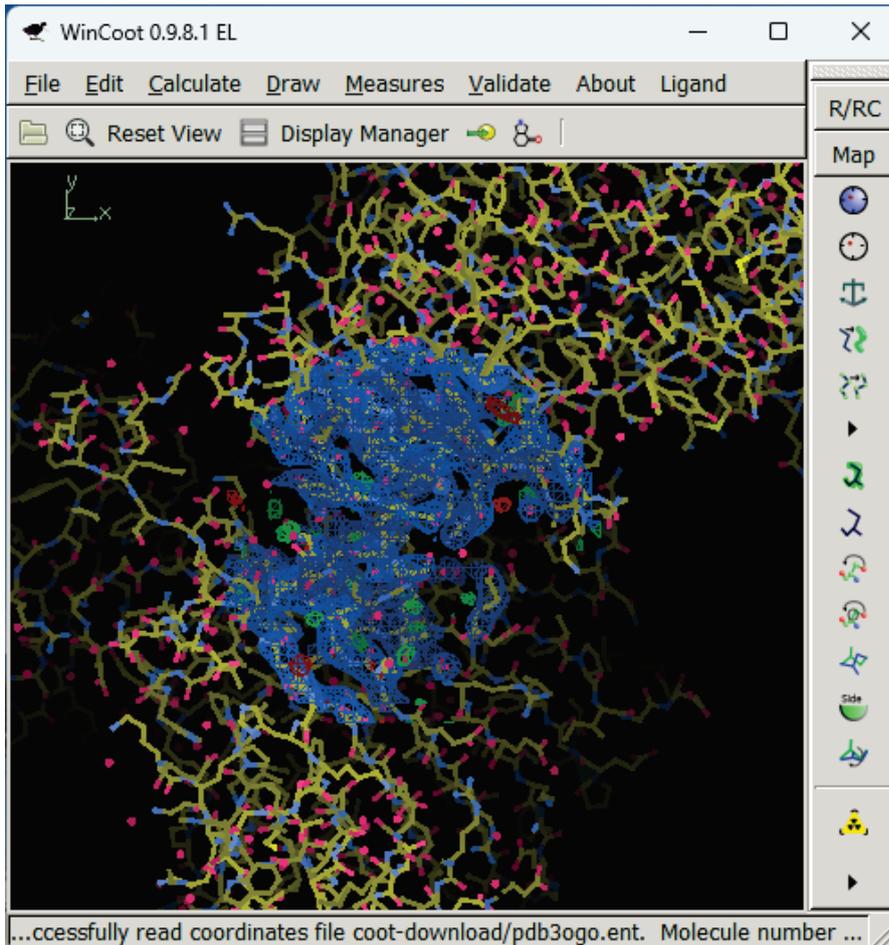
² This tutorial assumes you are running WinCoot version 0.9.8



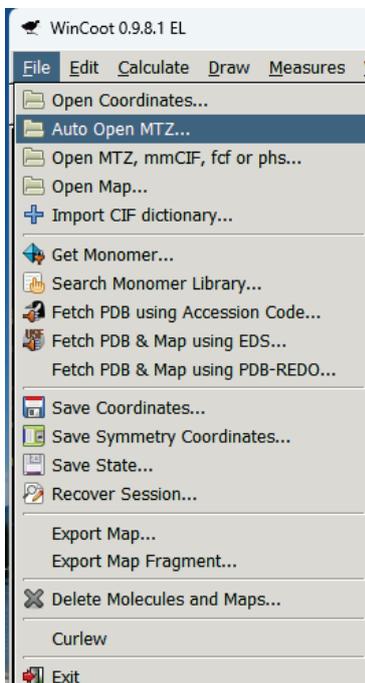
COOT will warn you about some 'nomenclature errors', because the names of some atoms don't match official guidelines – this is fine, you can hit Yes (this isn't anything to worry about).



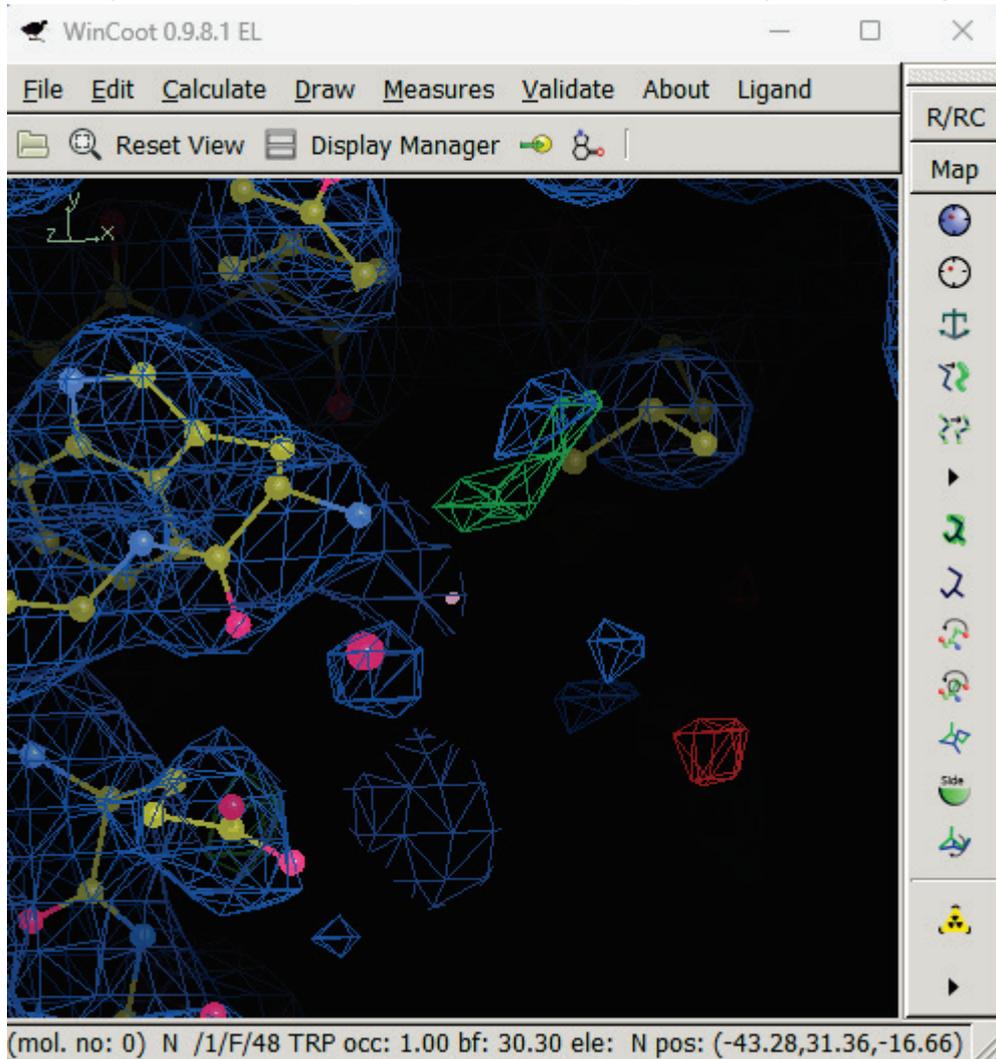
This should show you both a protein model and an electron density map (blue chicken wire).



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

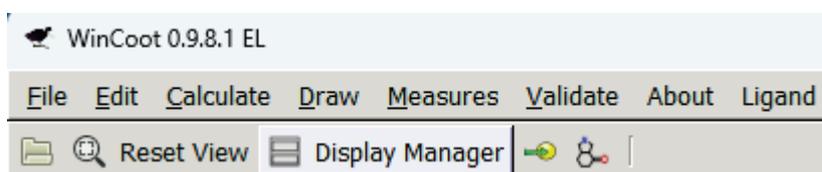


Once you've successfully opened your molecule and electron density map, try zooming in and out (right mouse button), rotating (left mouse button), translating (middle mouse button) and changing the map contour level (mouse scroll wheel). What you're looking at is as follows:

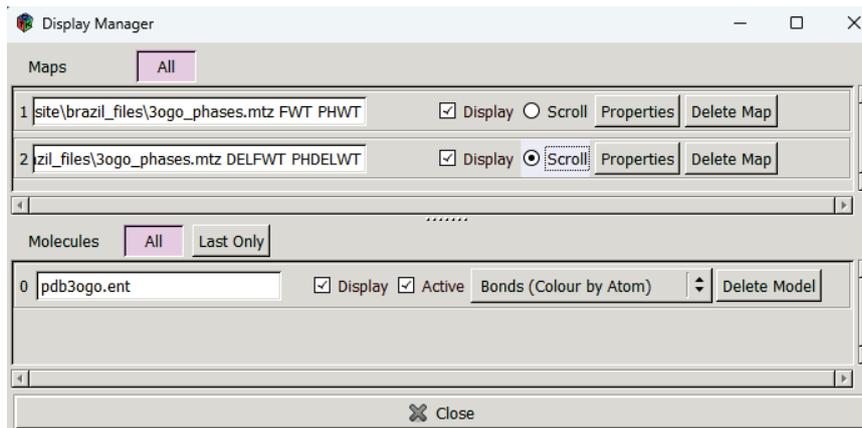


- Balls and sticks = Protein (model) and water atoms
- Blue chicken wire = $2F_o - F_c$ electron density, showing where electrons were observed in the experiment
- Green chicken wire = Positive $F_o - F_c$ density, shows where the model predicts fewer electrons than were observed in the experiment
-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

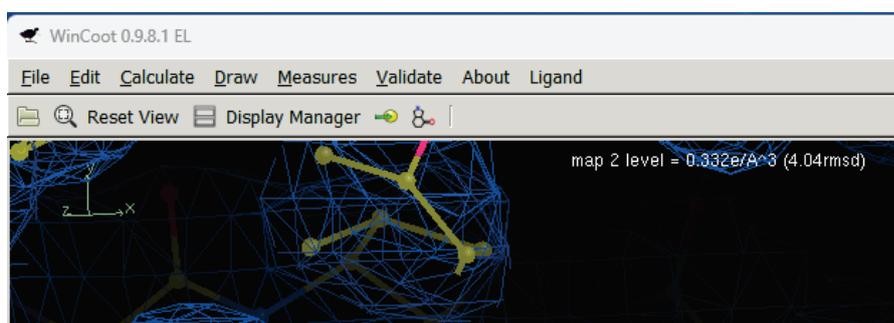
Note that the $F_o - F_c$ maps can be quite noisy, especially for a final refined structure (where all the big positive and negative peaks have been fixed and we're only looking at noise). You can change the contour level (signal:noise) level of the $F_o - F_c$ map by first opening the **Display Manager**:



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

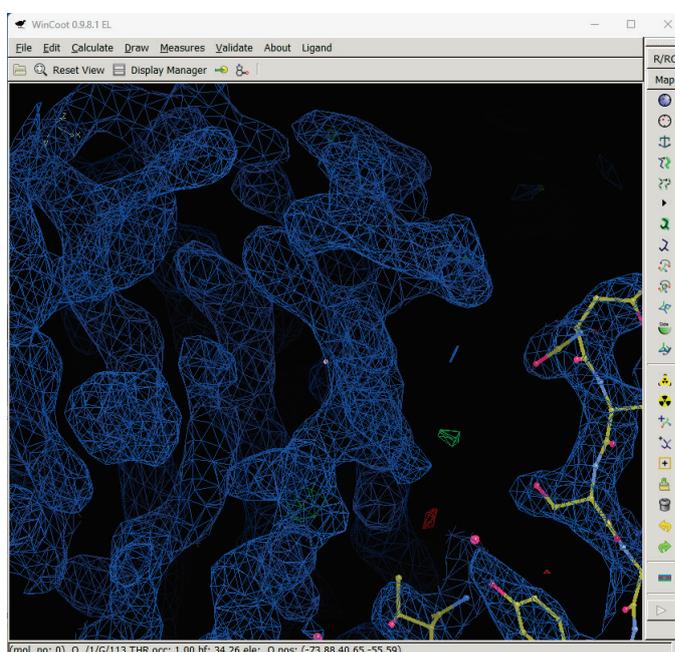


Then click back onto the main window and try scrolling your middle mouse button. The map level will be shown in the top of the main window – you want to aim for a contour of ~ 4 RMSD.

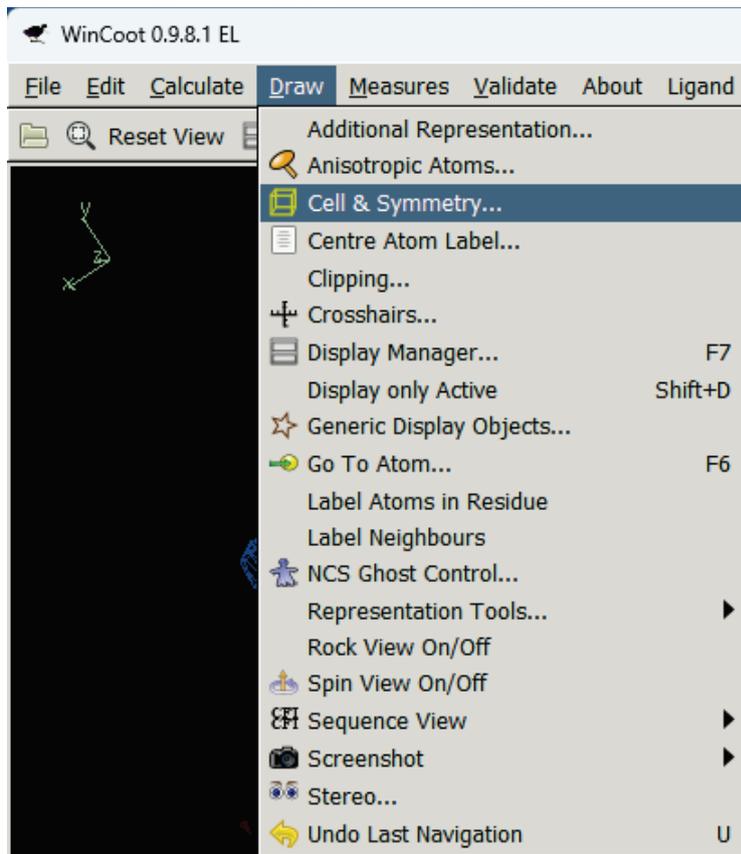


Now use the Display manager to select the 2Fo-Fc map (1, FWT) and set the contour to something like 1.5 RMSD.

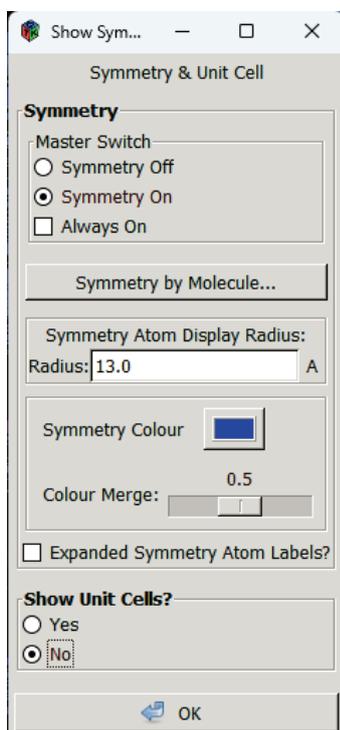
If you move around the structure (middle mouse) you can see that, for the most part, the model fits nicely into the middle of the electron density. However, if you scroll far enough you'll see areas of space where there is electron density (blue map) but no model.



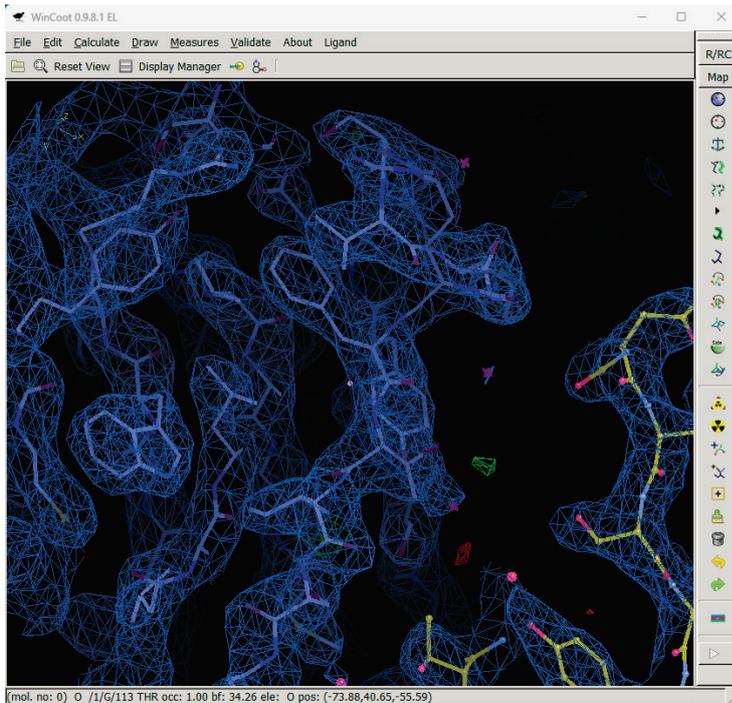
This is because the crystal is 'infinite' due to symmetry – COOT is automatically generating the density for areas of space by combining the information from the crystallographic asymmetric unit with the crystal symmetry. This is useful! We can use the same principle to show symmetry-related atoms from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



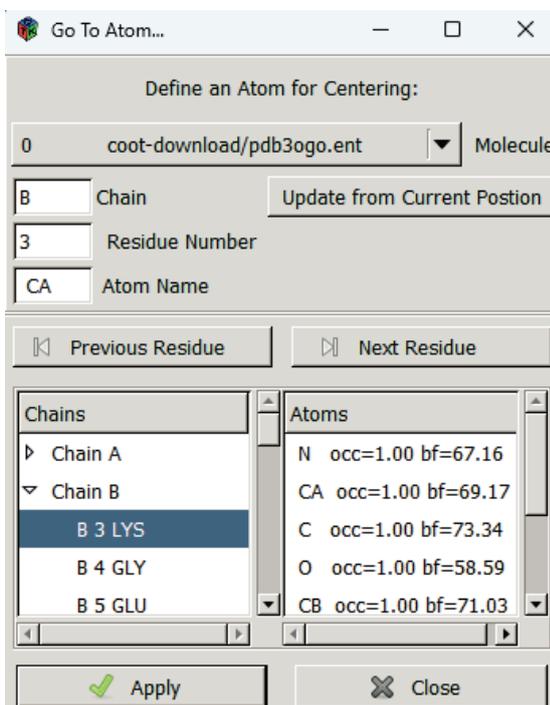
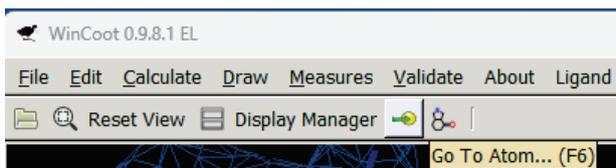
Turn the Symmetry Master Switch to On and hit OK:



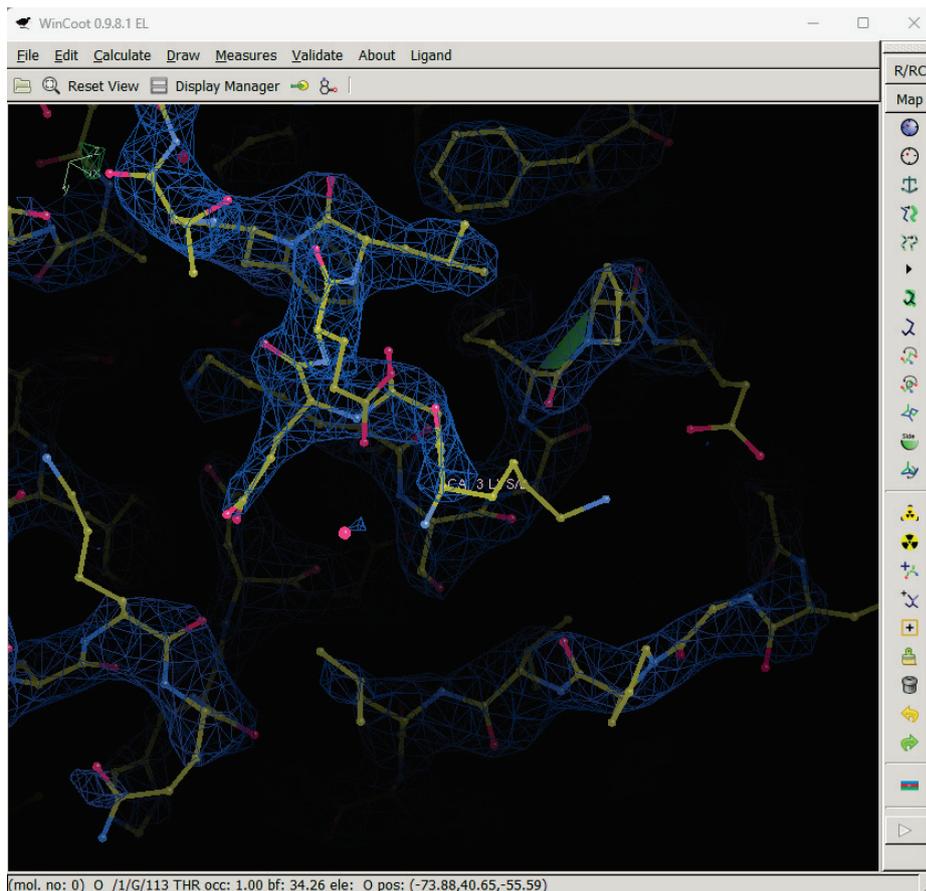
If you zoom out you can now see atoms in the symmetry-related density



We're now ready to inspect the molecule, but where to start – the structure is huge! Let's start by navigating 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.



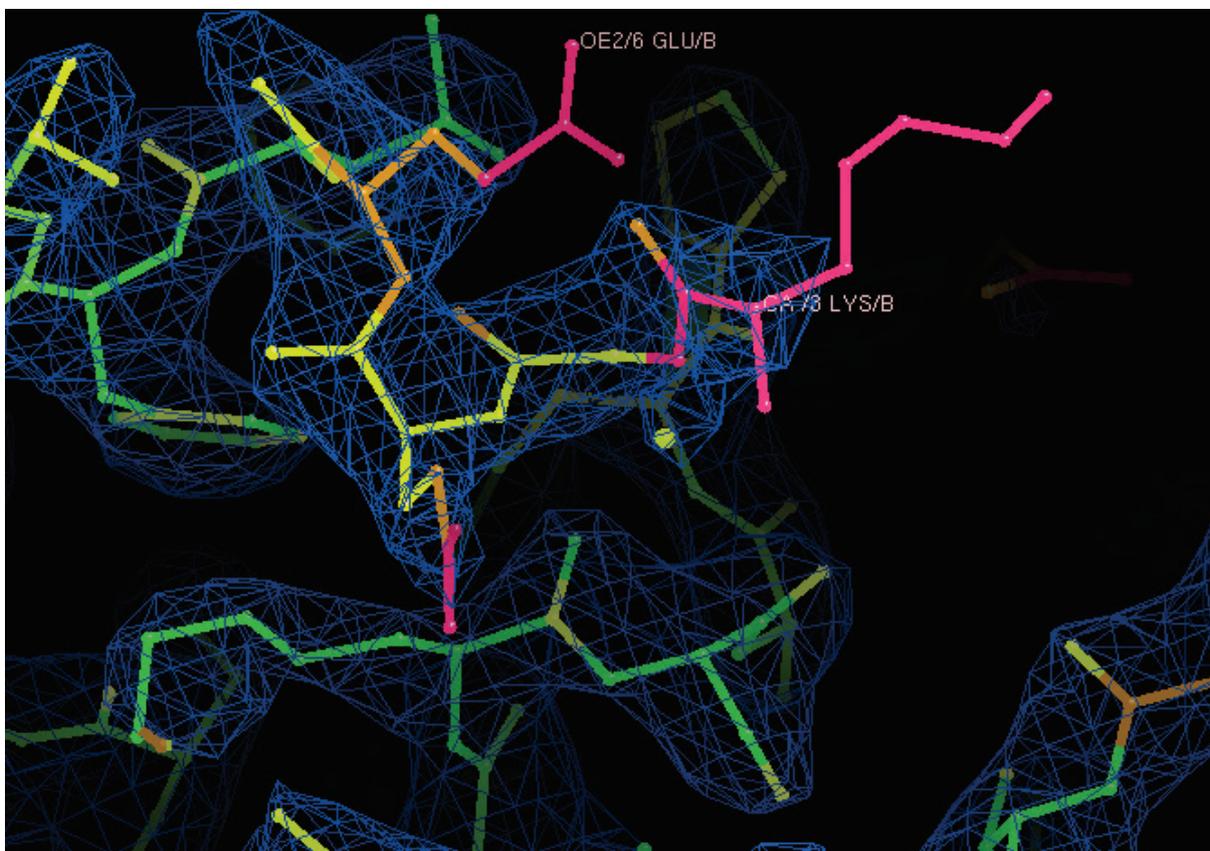
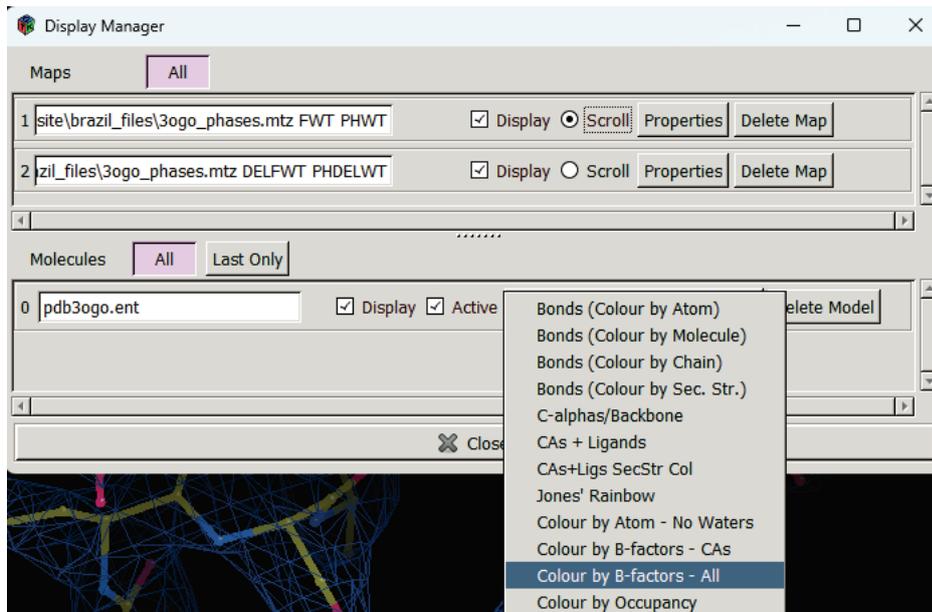
Why? This is because the side chain of this lysine is mobile. The maps in crystallographic experiments show the time and space average positions of electron density in the crystal – if atoms are highly mobile then these will average to (near) zero and you have a lack of electron density. This is presumably what happened for residue 2 of this EGFP molecule, which isn't modelled in the structure¹. If you go to residue 6 (Go To Atom, Glu 6) you can also note that glutamic acid has poor density too:



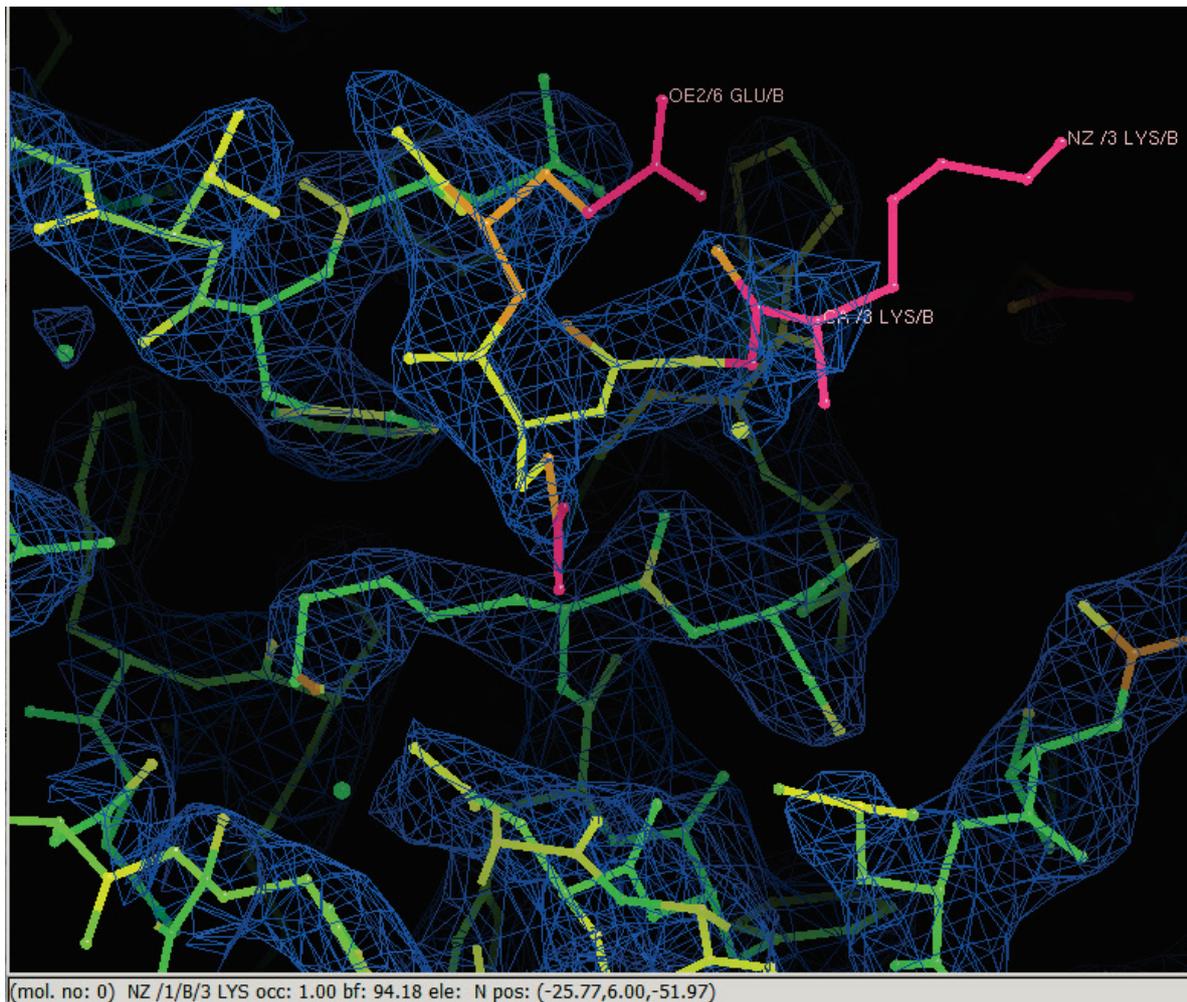
If you scroll the 2Fo-Fc density up and down you can see some weak density appear for this side chain, whereas the lysine really is very poorly ordered (almost no density).

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:



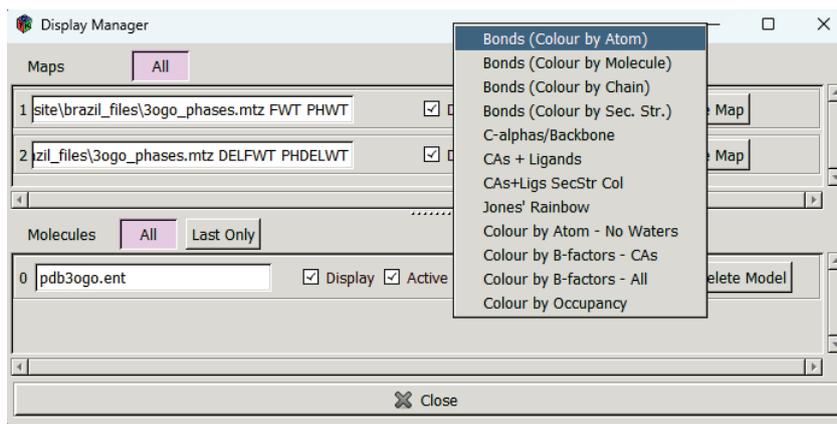
You see how the residues with poor electron density also have high B factors (red), whereas the residues in strong density have lower B factors (green). You can also hold `Shift` and left mouse button click on any residue to label that residue and show its B factor in the bottom panel of the COOT window:



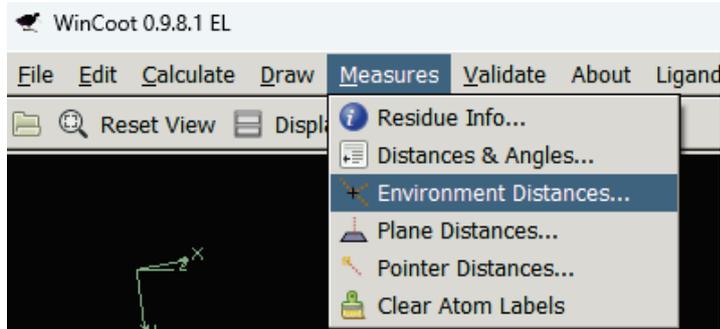
The final nitrogen atom of this lysine (part of the side chain amino group) has a very high temperature factor, so it is highly mobile.

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!**

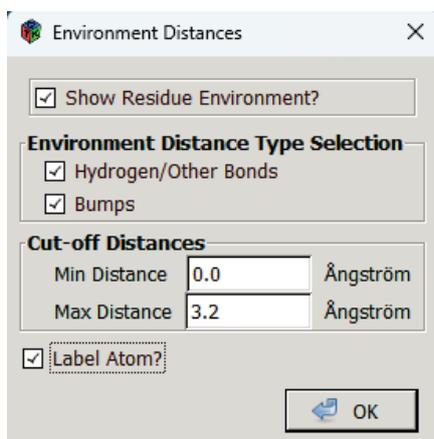
Let's go back to per-atom (element type) colouring:



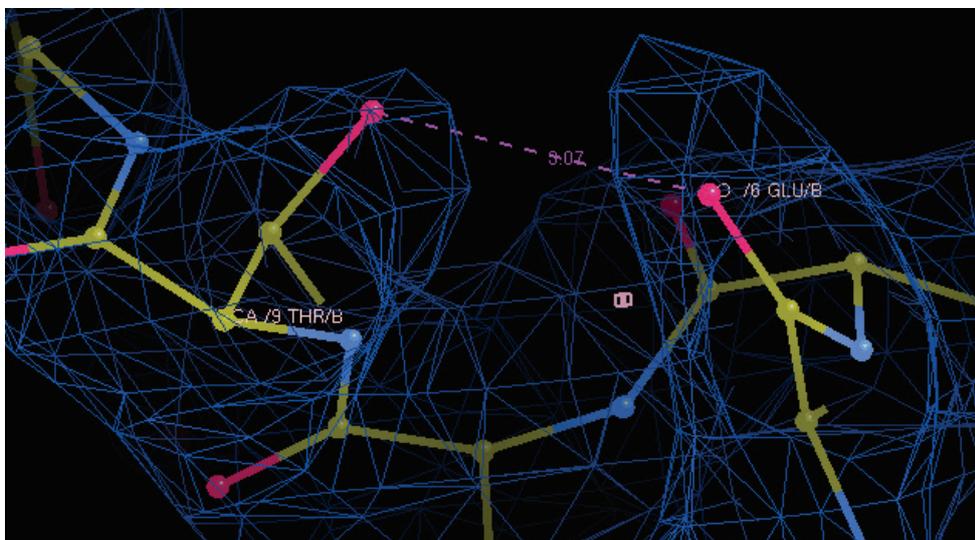
Step through the structure residue by residue. You can move to the next residue using the `Space` bar on your keyboard and go back using `Shift + Space`. When you get to Thr 9 you can see that the side chain hydroxyl group looks like it's making a hydrogen bond to the carbonyl oxygen of Glu 6. You can 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:



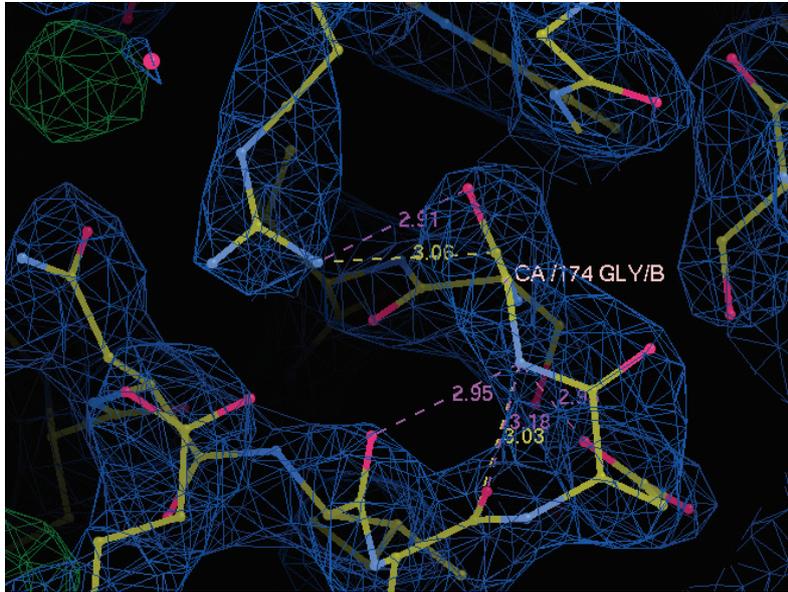
You can see now that the Thr side chain is indeed making a 3.07 Å hydrogen bond¹:



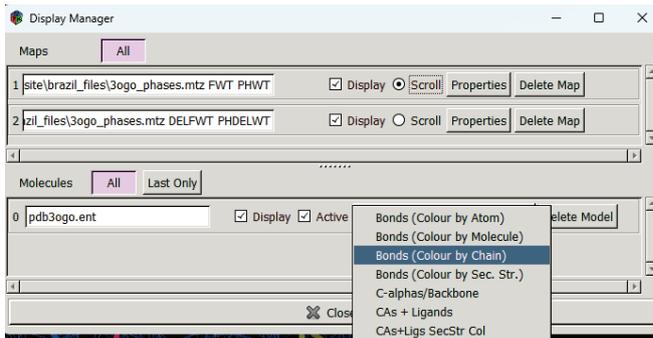
¹ If the text is a bit hard to read you can change the font size using **Edit > Preferences**, then click on **Others** and go to the **Fonts** tab where you can change the size

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 carbonyl 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 Ctrl + 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?

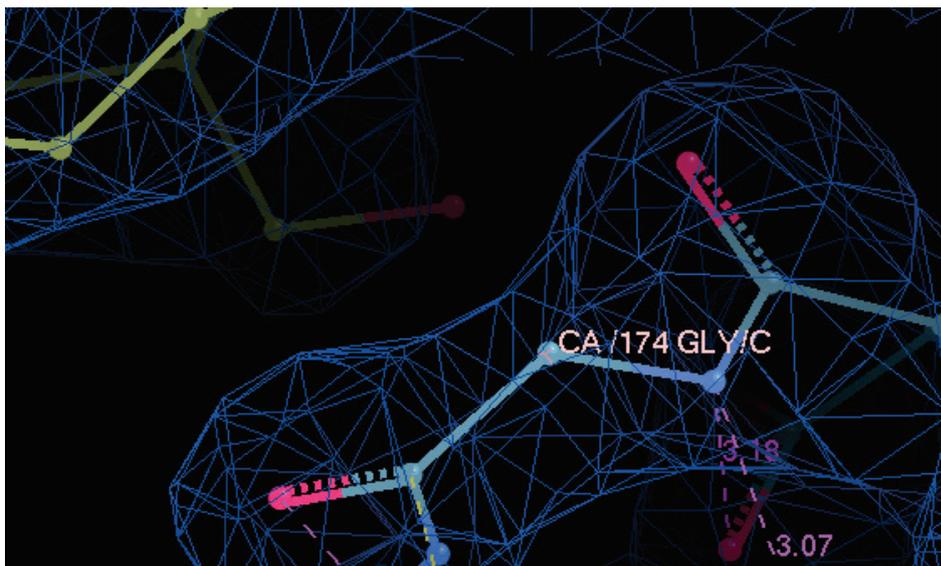


Turn on Bonds (Colour by Chain) in the Display Manager:

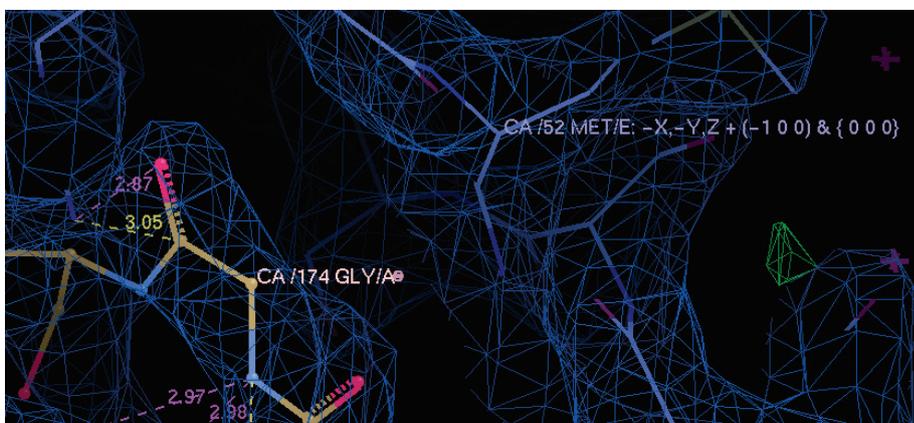


You can actually see (just – the colouring is a bit subtle) that the arginine is in a different chain because it has a different carbon atom colour. This is the nanobody that the EGFP is bound to, i.e. the interaction interface that we were looking at yesterday. It looks like the interface is very well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

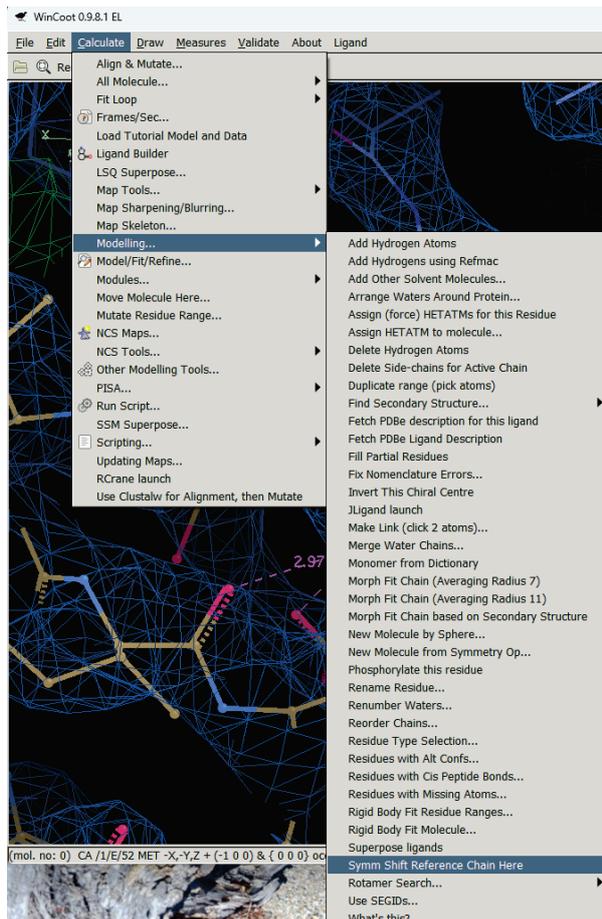
Chain B is just one of the four copies of EGFP in the asymmetric unit that was crystallised in this experiment (the other copies are chains A, C and D). COOT is clever about this and lets you hop between the same residue *on different chains* using the `o` key (as in ‘o’ for orange, between ‘i’ and ‘p’ 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:

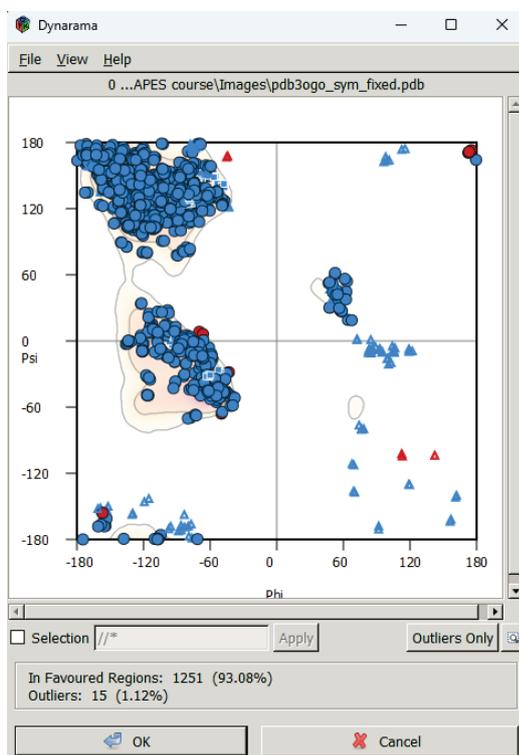


In COOT you can actually change which symmetry copy of the chain is in the PDB file, because it’s more convenient to have the ‘real’ complexes close to each other. Try doing this for chains E and H by middle mouse clicking on a residue in the symmetry-related nanobody then shifting the reference chain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



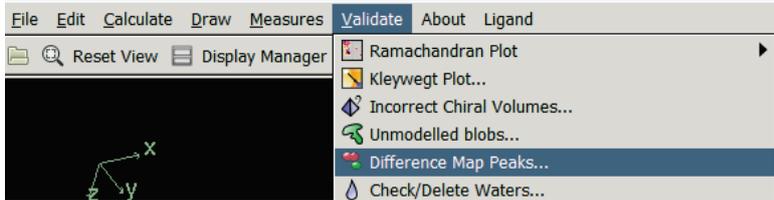
You can save this file for later use using **File > Save Coordinates**.

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:

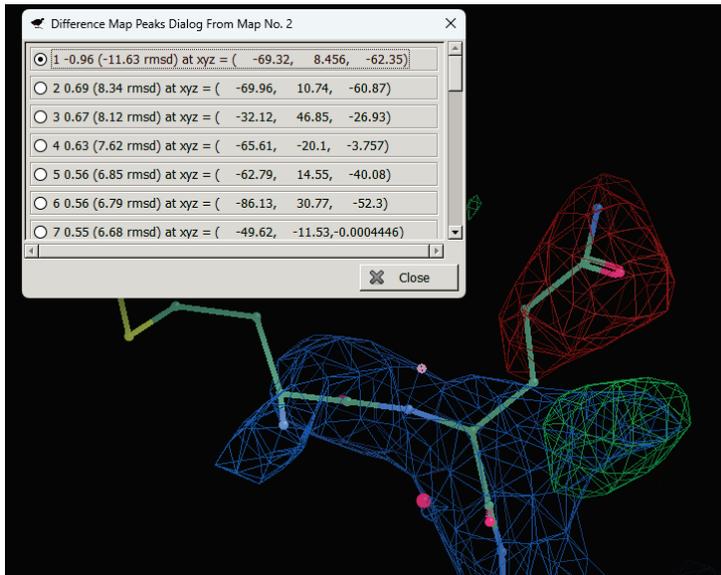


The fact that there are outliers doesn't 100% mean these residues are in the wrong position, but they warrant further inspection. Try clicking on one of the red shapes (outliers). COOT will take you to that residue where you can have a look at the density and judge for yourself – is the density strong?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

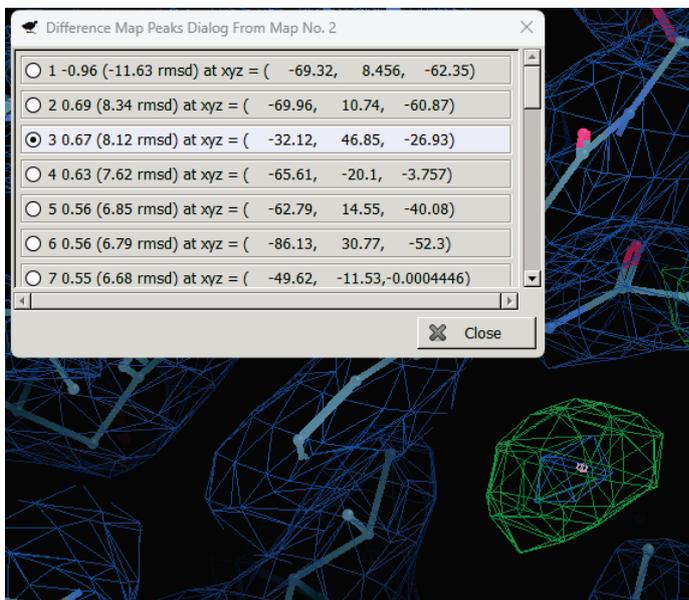
Next, try looking at the difference map peaks:



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 structure 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>).

Structure Summary | 3D View | Annotations | Experiment | Sequence | Genome | Versions

3OGO
Structure of the GFP:GFP-nanobody complex at 2.8 Å resolution in spacegroup P21212
PDB DOI: 10.2210/pdb3OGO/pdb Entry: 3OGO supersedes: 3MIQ
Classification: **FLUORESCENT PROTEIN/IMMUNE SYSTEM**
Organism(s): Aequorea victoria, Camelus dromedarius
Expression System: Escherichia coli
Mutation(s): Yes

Deposited: 2010-08-17 Released: 2010-08-25
Deposition Author(s): Kubala, M.H., Kovtun, O., Alexandrov, K., Collins, B.M.

Experimental Data Snapshot
Method: X-RAY DIFFRACTION
Resolution: 2.80 Å
R-Value Free: 0.253
R-Value Work: 0.202
R-Value Observed: 0.204

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.250
Clashscore		12
Ramachandran outliers		0.1%
Sidechain outliers		7.1%
RSRZ outliers		1.3%

This is version 1.2 of the entry. See complete [history](#).

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:

The screenshot shows a web-based molecular structure viewer. The main window displays a protein structure (3OGO) in a ribbon representation. The structure is colored in shades of orange and brown. The top of the window shows the sequence of the protein, with residues 1-100 and 101-200 visible. The right-hand side of the window contains a control panel with the following sections:

- Structure:** 3OGO | Structure of the GFP:GFP-n...
 - Type: Assembly
 - Asm Id: 2: Author Defined As...
 - Dynamic Bonds: X Off
 - Nothing Focused
- Measurements:** (empty)
- Structure Motif Search:** (empty)
- Components:** 3OGO
 - Preset: + Add
 - Polymer: Cartoon
 - Non-standard: Ball & Stick
 - Water: Ball & Stick
 - Unit Cell: P 21 21 2
- Density:** 3OGO
 - 2Fo-Fc σ : 1.5
 - Fo-Fc(+ve) σ : 3
 - Fo-Fc(-ve) σ : -3
 - Entry: 3ogo
 - View: Around Focus
- Quality Assessment:** (empty)

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:

This is a close-up screenshot of the 'Structure' control panel. It shows the following details:

- Structure:** 3OGO | Structure of the GFP:GFP-n...
- Type:** Assembly
- Asm Id:** 2: Author Defined As...
- Assembly Selection:** A dropdown menu is open, showing four options:
 - 1: Author Defined Assembly
 - 2: Author Defined Assembly** (highlighted)
 - 3: Author Defined Assembly
 - 4: Author Defined Assembly
- Dynamic Bonds:** X Off
- Current Selection:** LYS 175 [auth 166] | B

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).

Structure Summary | **3D View** | Annotations | Experiment | Sequence | Genome | Versions

3OGO Display Files | Download Files

Structure of the GFP:GFP-nanobody complex at 2.8 Å resolution in spacegroup P21212 Help

Sequence of 3OGO | Struct... | Chain | 1: Green fluor... | B

MAHHHHHSSGVSKGEEELFTGVVPIFVLDGVDVNGHNFVSVSGEGEGDATTYGLKILKFICTTGLKLVVFWFTLVITL CROVQCFSTRYPDRHQDHFYKSAPEGVVQERTIFFFK
 112 115 118 121 124 127 130 133 136 139 142 145 148 151 154 157 160 163 166 169 172 175 178 181 184 187 190 193 196 199 202 205 208 211 214
 DDGNYKTRAEVKFEGDILVNRILKGIIDFKEDGNILGHKLEVINSHNVYINADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPFVLLPDNHYLSTQSALSKDPNE
 222 225 228 231 234 237 240 243 246 249 252 255 258 261 264 267 270 273 276 279 282 285 288 291 294 297 300 303 306 309 312 315 318 321 324 327 330 333 336 339 342 345 348 351 354 357 360 363 366 369 372 375 378 381 384 387 390 393 396 399 402 405 408 411 414 417 420 423 426 429 432 435 438 441 444 447 450 453 456 459 462 465 468 471 474 477 480 483 486 489 492 495 498 501 504 507 510 513 516 519 522 525 528 531 534 537 540 543 546 549 552 555 558 561 564 567 570 573 576 579 582 585 588 591 594 597 600 603 606 609 612 615 618 621 624 627 630 633 636 639 642 645 648 651 654 657 660 663 666 669 672 675 678 681 684 687 690 693 696 699 702 705 708 711 714 717 720 723 726 729 732 735 738 741 744 747 750 753 756 759 762 765 768 771 774 777 780 783 786 789 792 795 798 801 804 807 810 813 816 819 822 825 828 831 834 837 840 843 846 849 852 855 858 861 864 867 870 873 876 879 882 885 888 891 894 897 900 903 906 909 912 915 918 921 924 927 930 933 936 939 942 945 948 951 954 957 960 963 966 969 972 975 978 981 984 987 990 993 996 999 1002 1005 1008 1011 1014 1017 1020 1023 1026 1029 1032 1035 1038 1041 1044 1047 1050 1053 1056 1059 1062 1065 1068 1071 1074 1077 1080 1083 1086 1089 1092 1095 1098 1101 1104 1107 1110 1113 1116 1119 1122 1125 1128 1131 1134 1137 1140 1143 1146 1149 1152 1155 1158 1161 1164 1167 1170 1173 1176 1179 1182 1185 1188 1191 1194 1197 1200 1203 1206 1209 1212 1215 1218 1221 1224 1227 1230 1233 1236 1239 1242 1245 1248 1251 1254 1257 1260 1263 1266 1269 1272 1275 1278 1281 1284 1287 1290 1293 1296 1299 1302 1305 1308 1311 1314 1317 1320 1323 1326 1329 1332 1335 1338 1341 1344 1347 1350 1353 1356 1359 1362 1365 1368 1371 1374 1377 1380 1383 1386 1389 1392 1395 1398 1401 1404 1407 1410 1413 1416 1419 1422 1425 1428 1431 1434 1437 1440 1443 1446 1449 1452 1455 1458 1461 1464 1467 1470 1473 1476 1479 1482 1485 1488 1491 1494 1497 1500 1503 1506 1509 1512 1515 1518 1521 1524 1527 1530 1533 1536 1539 1542 1545 1548 1551 1554 1557 1560 1563 1566 1569 1572 1575 1578 1581 1584 1587 1590 1593 1596 1599 1602 1605 1608 1611 1614 1617 1620 1623 1626 1629 1632 1635 1638 1641 1644 1647 1650 1653 1656 1659 1662 1665 1668 1671 1674 1677 1680 1683 1686 1689 1692 1695 1698 1701 1704 1707 1710 1713 1716 1719 1722 1725 1728 1731 1734 1737 1740 1743 1746 1749 1752 1755 1758 1761 1764 1767 1770 1773 1776 1779 1782 1785 1788 1791 1794 1797 1800 1803 1806 1809 1812 1815 1818 1821 1824 1827 1830 1833 1836 1839 1842 1845 1848 1851 1854 1857 1860 1863 1866 1869 1872 1875 1878 1881 1884 1887 1890 1893 1896 1899 1902 1905 1908 1911 1914 1917 1920 1923 1926 1929 1932 1935 1938 1941 1944 1947 1950 1953 1956 1959 1962 1965 1968 1971 1974 1977 1980 1983 1986 1989 1992 1995 1998 2001 2004 2007 2010 2013 2016 2019 2022 2025 2028 2031 2034 2037 2040 2043 2046 2049 2052 2055 2058 2061 2064 2067 2070 2073 2076 2079 2082 2085 2088 2091 2094 2097 2100 2103 2106 2109 2112 2115 2118 2121 2124 2127 2130 2133 2136 2139 2142 2145 2148 2151 2154 2157 2160 2163 2166 2169 2172 2175 2178 2181 2184 2187 2190 2193 2196 2199 2202 2205 2208 2211 2214 2217 2220 2223 2226 2229 2232 2235 2238 2241 2244 2247 2250 2253 2256 2259 2262 2265 2268 2271 2274 2277 2280 2283 2286 2289 2292 2295 2298 2301 2304 2307 2310 2313 2316 2319 2322 2325 2328 2331 2334 2337 2340 2343 2346 2349 2352 2355 2358 2361 2364 2367 2370 2373 2376 2379 2382 2385 2388 2391 2394 2397 2400 2403 2406 2409 2412 2415 2418 2421 2424 2427 2430 2433 2436 2439 2442 2445 2448 2451 2454 2457 2460 2463 2466 2469 2472 2475 2478 2481 2484 2487 2490 2493 2496 2499 2502 2505 2508 2511 2514 2517 2520 2523 2526 2529 2532 2535 2538 2541 2544 2547 2550 2553 2556 2559 2562 2565 2568 2571 2574 2577 2580 2583 2586 2589 2592 2595 2598 2601 2604 2607 2610 2613 2616 2619 2622 2625 2628 2631 2634 2637 2640 2643 2646 2649 2652 2655 2658 2661 2664 2667 2670 2673 2676 2679 2682 2685 2688 2691 2694 2697 2700 2703 2706 2709 2712 2715 2718 2721 2724 2727 2730 2733 2736 2739 2742 2745 2748 2751 2754 2757 2760 2763 2766 2769 2772 2775 2778 2781 2784 2787 2790 2793 2796 2799 2802 2805 2808 2811 2814 2817 2820 2823 2826 2829 2832 2835 2838 2841 2844 2847 2850 2853 2856 2859 2862 2865 2868 2871 2874 2877 2880 2883 2886 2889 2892 2895 2898 2901 2904 2907 2910 2913 2916 2919 2922 2925 2928 2931 2934 2937 2940 2943 2946 2949 2952 2955 2958 2961 2964 2967 2970 2973 2976 2979 2982 2985 2988 2991 2994 2997 3000 3003 3006 3009 3012 3015 3018 3021 3024 3027 3030 3033 3036 3039 3042 3045 3048 3051 3054 3057 3060 3063 3066 3069 3072 3075 3078 3081 3084 3087 3090 3093 3096 3099 3102 3105 3108 3111 3114 3117 3120 3123 3126 3129 3132 3135 3138 3141 3144 3147 3150 3153 3156 3159 3162 3165 3168 3171 3174 3177 3180 3183 3186 3189 3192 3195 3198 3201 3204 3207 3210 3213 3216 3219 3222 3225 3228 3231 3234 3237 3240 3243 3246 3249 3252 3255 3258 3261 3264 3267 3270 3273 3276 3279 3282 3285 3288 3291 3294 3297 3300 3303 3306 3309 3312 3315 3318 3321 3324 3327 3330 3333 3336 3339 3342 3345 3348 3351 3354 3357 3360 3363 3366 3369 3372 3375 3378 3381 3384 3387 3390 3393 3396 3399 3402 3405 3408 3411 3414 3417 3420 3423 3426 3429 3432 3435 3438 3441 3444 3447 3450 3453 3456 3459 3462 3465 3468 3471 3474 3477 3480 3483 3486 3489 3492 3495 3498 3501 3504 3507 3510 3513 3516 3519 3522 3525 3528 3531 3534 3537 3540 3543 3546 3549 3552 3555 3558 3561 3564 3567 3570 3573 3576 3579 3582 3585 3588 3591 3594 3597 3600 3603 3606 3609 3612 3615 3618 3621 3624 3627 3630 3633 3636 3639 3642 3645 3648 3651 3654 3657 3660 3663 3666 3669 3672 3675 3678 3681 3684 3687 3690 3693 3696 3699 3702 3705 3708 3711 3714 3717 3720 3723 3726 3729 3732 3735 3738 3741 3744 3747 3750 3753 3756 3759 3762 3765 3768 3771 3774 3777 3780 3783 3786 3789 3792 3795 3798 3801 3804 3807 3810 3813 3816 3819 3822 3825 3828 3831 3834 3837 3840 3843 3846 3849 3852 3855 3858 3861 3864 3867 3870 3873 3876 3879 3882 3885 3888 3891 3894 3897 3900 3903 3906 3909 3912 3915 3918 3921 3924 3927 3930 3933 3936 3939 3942 3945 3948 3951 3954 3957 3960 3963 3966 3969 3972 3975 3978 3981 3984 3987 3990 3993 3996 3999 4002 4005 4008 4011 4014 4017 4020 4023 4026 4029 4032 4035 4038 4041 4044 4047 4050 4053 4056 4059 4062 4065 4068 4071 4074 4077 4080 4083 4086 4089 4092 4095 4098 4101 4104 4107 4110 4113 4116 4119 4122 4125 4128 4131 4134 4137 4140 4143 4146 4149 4152 4155 4158 4161 4164 4167 4170 4173 4176 4179 4182 4185 4188 4191 4194 4197 4200 4203 4206 4209 4212 4215 4218 4221 4224 4227 4230 4233 4236 4239 4242 4245 4248 4251 4254 4257 4260 4263 4266 4269 4272 4275 4278 4281 4284 4287 4290 4293 4296 4299 4302 4305 4308 4311 4314 4317 4320 4323 4326 4329 4332 4335 4338 4341 4344 4347 4350 4353 4356 4359 4362 4365 4368 4371 4374 4377 4380 4383 4386 4389 4392 4395 4398 4401 4404 4407 4410 4413 4416 4419 4422 4425 4428 4431 4434 4437 4440 4443 4446 4449 4452 4455 4458 4461 4464 4467 4470 4473 4476 4479 4482 4485 4488 4491 4494 4497 4500 4503 4506 4509 4512 4515 4518 4521 4524 4527 4530 4533 4536 4539 4542 4545 4548 4551 4554 4557 4560 4563 4566 4569 4572 4575 4578 4581 4584 4587 4590 4593 4596 4599 4602 4605 4608 4611 4614 4617 4620 4623 4626 4629 4632 4635 4638 4641 4644 4647 4650 4653 4656 4659 4662 4665 4668 4671 4674 4677 4680 4683 4686 4689 4692 4695 4698 4701 4704 4707 4710 4713 4716 4719 4722 4725 4728 4731 4734 4737 4740 4743 4746 4749 4752 4755 4758 4761 4764 4767 4770 4773 4776 4779 4782 4785 4788 4791 4794 4797 4800 4803 4806 4809 4812 4815 4818 4821 4824 4827 4830 4833 4836 4839 4842 4845 4848 4851 4854 4857 4860 4863 4866 4869 4872 4875 4878 4881 4884 4887 4890 4893 4896 4899 4902 4905 4908 4911 4914 4917 4920 4923 4926 4929 4932 4935 4938 4941 4944 4947 4950 4953 4956 4959 4962 4965 4968 4971 4974 4977 4980 4983 4986 4989 4992 4995 4998 5001 5004 5007 5010 5013 5016 5019 5022 5025 5028 5031 5034 5037 5040 5043 5046 5049 5052 5055 5058 5061 5064 5067 5070 5073 5076 5079 5082 5085 5088 5091 5094 5097 5100 5103 5106 5109 5112 5115 5118 5121 5124 5127 5130 5133 5136 5139 5142 5145 5148 5151 5154 5157 5160 5163 5166 5169 5172 5175 5178 5181 5184 5187 5190 5193 5196 5199 5202 5205 5208 5211 5214 5217 5220 5223 5226 5229 5232 5235 5238 5241 5244 5247 5250 5253 5256 5259 5262 5265 5268 5271 5274 5277 5280 5283 5286 5289 5292 5295 5298 5301 5304 5307 5310 5313 5316 5319 5322 5325 5328 5331 5334 5337 5340 5343 5346 5349 5352 5355 5358 5361 5364 5367 5370 5373 5376 5379 5382 5385 5388 5391 5394 5397 5400 5403 5406 5409 5412 5415 5418 5421 5424 5427 5430 5433 5436 5439 5442 5445 5448 5451 5454 5457 5460 5463 5466 5469 5472 5475 5478 5481 5484 5487 5490 5493 5496 5499 5502 5505 5508 5511 5514 5517 5520 5523 5526 5529 5532 5535 5538 5541 5544 5547 5550 5553 5556 5559 5562 5565 5568 5571 5574 5577 5580 5583 5586 5589 5592 5595 5598 5601 5604 5607 5610 5613 5616 5619 5622 5625 5628 5631 5634 5637 5640 5643 5646 5649 5652 5655 5658 5661 5664 5667 5670 5673 5676 5679 5682 5685 5688 5691 5694 5697 5700 5703 5706 5709 5712 5715 5718 5721 5724 5727 5730 5733 5736 5739 5742 5745 5748 5751 5754 5757 5760 5763 5766 5769 5772 5775 5778 5781 5784 5787 5790 5793 5796 5799 5802 5805 5808 5811 5814 5817 5820 5823 5826 5829 5832 5835 5838 5841 5844 5847 5850 5853 5856 5859 5862 5865 5868 5871 5874 5877 5880 5883 5886 5889 5892 5895 5898 5901 5904 5907 5910 5913 5916 5919 5922 5925 5928 5931 5934 5937 5940 5943 5946 5949 5952 5955 5958 5961 5964 5967 5970 5973 5976 5979 5982 5985 5988 5991 5994 5997 6000 6003 6006 6009 6012 6015 6018 6021 6024 6027 6030 6033 6036 6039 6042 6045 6048 6051 6054 6057 6060 6063 6066 6069 6072 6075 6078 6081 6084 6087 6090 6093 6096 6099 6102 6105 6108 6111 6114 6117 6120 6123 6126 6129 6132 6135 6138 6141 6144 6147 6150 6153 6156 6159 6162 6165 6168 6171 6174 6177 6180 6183 6186 6189 6192 6195 6198 6201 6204 6207 6210 6213 6216 6219 6222 6225 6228 6231 6234 6237 6240 6243 6246 6249 6252 6255 6258 6261 6264 6267 6270 6273 6276 6279 6282 6285 6288 6291 6294 6297 6300 6303 6306 6309 6312 6315 6318 6321 6324 6327 6330 6333 6336 6339 6342 6345 6348 6351 6354 6357 6360 6363 6366 6369 6372 6375 6378 6381 6384 6387 6390 6393 6396 6399 6402 6405 6408 6411 6414 6417 6420 6423 6426 6429 6432 6435 6438 6441 6444 6447 6450 6453 6456 6459 6462 6465 6468 6471 6474 6477 6480 6483 6486 6489 6492 6495 6498 6501 6504 6507 6510 6513 6516 6519 6522 6525 6528 6531 6534 6537 6540 6543 6546 6549 6552 6555 6558 6561 6564 6567 6570 6573 6576 6579 6582 6585 6588 6591 6594 6597 6600 6603 6606 6609 6612 6615 6618 6621 6624 6627 6630 6633 6636 6639 6642 6645 6648 6651 6654 6657 6660 6663 6666 6669 6672 6675 6678 6681 6684 6687 6690 6693 6696 6699 6702 6705 6708 6711 6714 6717 6720 6723 6726 6729 6732 6735 6738 6741 6744 6747 6750 6753 6756 6759 6762 6765 6768 6771 6774 6777 6780 6783 6786 6789 6792 6795 6798 6801 6804 6807 6810 6813 6816 6819 6822 6825 6828 6831 6834 6837 6840 6843 6846 6849 6852 6855 6858 6861 6864 6867 6870 6873 6876 6879 6882 6885 6888 6891 6894 6897 6900 6903 6906 6909 6912 6915 6918 6921 6924 6927 6930 6933 6936 6939 6942 6945 6948 6951 6954 6957 6960 6963 6966 6969 6972 6975 6978 6981 6984 6987 6990 6993 6996 6999 7002 7005 7008 7011 7014 7017 7020 7023 7026 7029 7032 7035 7038 7041 7044 7047 7050 7053 7056 7059 7062 7065 7068 7071 7074 7077 7080 7083 7086 7089 7092 7095 7098 7101 7104 7107 7110 7113 7116 7119 7122 7125 7128 7131 7134 7137 7140 7143 7146 7149 7152 7155 7158 7161 7164 7167 7170 7173 7176 7179 7182 7185 7188 7191 7194 7197 7200 7203 7206 7209 7212 7215 7218 7221 7224 7227 7230 7233 7236 7239 7242 7245 7248 7251 7254 7257 7260 7263 7266 7269 7272 7275 7278 7281 7284 7287 7290 7293 7296 7299 7302 7305 7308 7311 7314 7317 7320 7323 7326 7329 7332 7335 7338 7341 7344 7347 7350 7353 7356 7359 7362 7365 7368 7371 7374 7377 7380 7383 7386 7389 7392 7395 7398 7401 7404 7407 7410 7413 7416

Computer lab 3 – AlphaFold2 and ColabFold

Overview

Today we will have a go at running AlphaFold2 and ColabFold via Google Colab. This is a free¹ 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:

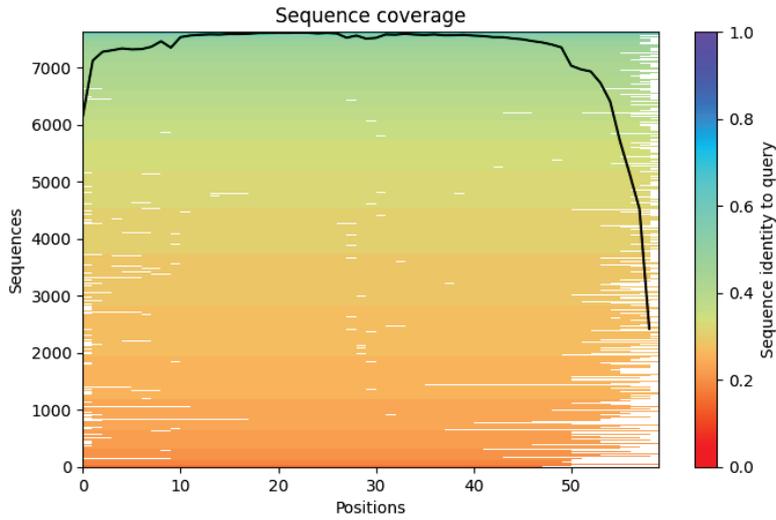
<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

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*):

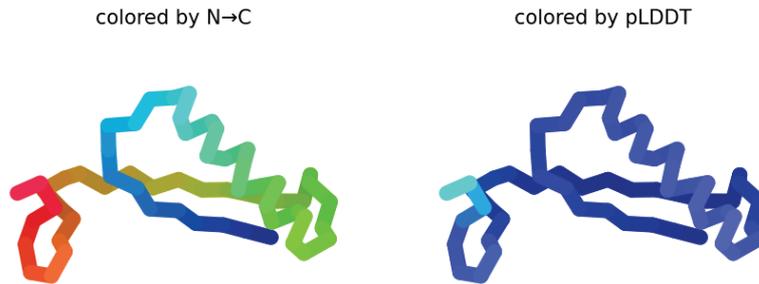
To start the prediction, just click on **Runtime > Run all**:

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:

▶ Display 3D structure

rank_num:

color:

show_sidechains:

show_mainchains:

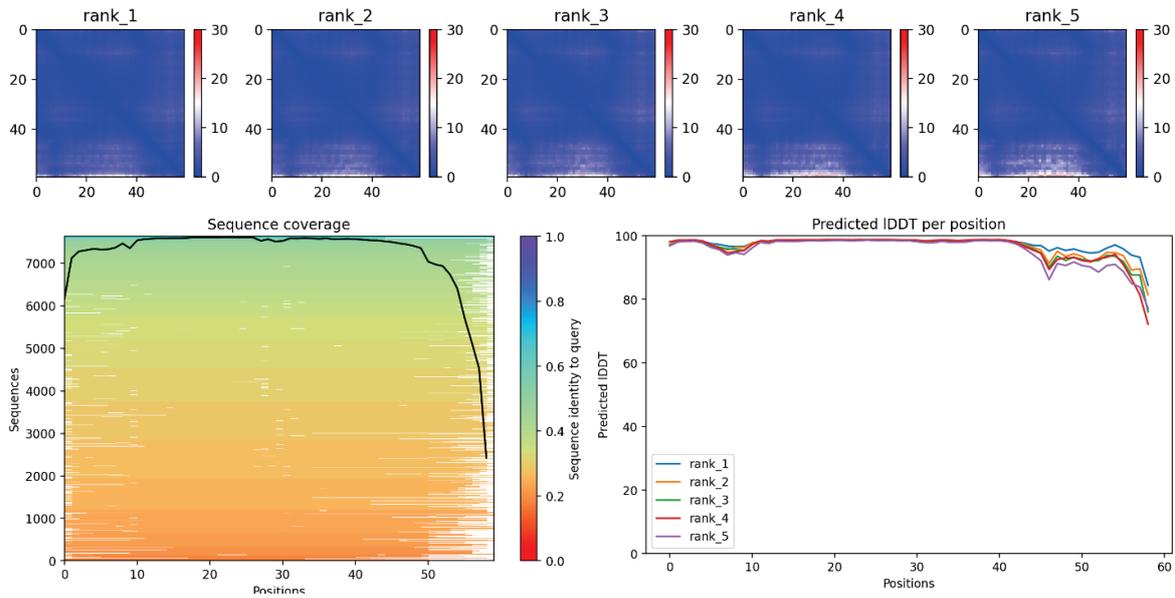
[Show code](#)



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

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

Plots for test_a5e17



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):

```

Package and download results
- If you are having issues downloading the result archive, try disabling your adblocker and run this cell again. If that fails click on the little folder icon to the left, navigate to file: jobname.result.zip, right-click and select Run cell (Ctrl+Enter) cell executed since last change (shot)
  executed at 15:29 (0 minutes ago)
  executed in 0.267 s
  
```

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
  
```

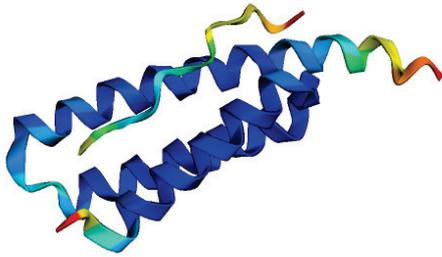
Put the two sequences into the **Input protein sequence(s)** cell separated by a colon (:)

```

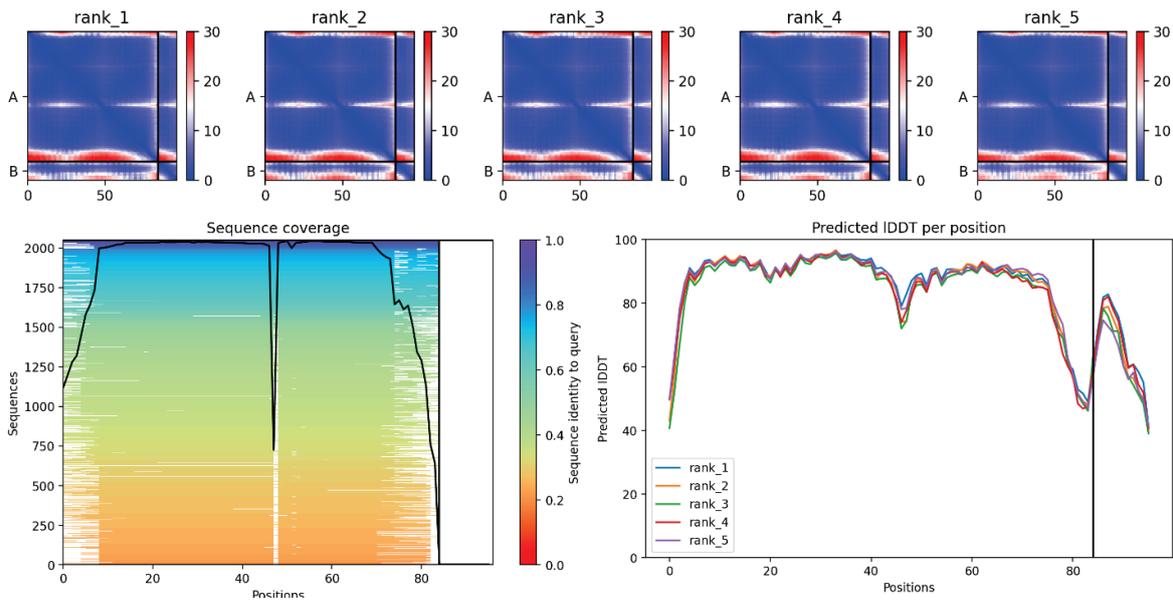
Input protein sequence(s), then hit Runtime -> Run all
query_sequence: "MTTLTLQKAIDLVTKATEEDKAKNYEEALRLYQHAVEYFLHAIKYEAHSDKAKESIRGKCMQYLDRAEKLKDYLRSKDKQSKKP:VHLPDVPVEPPP"
  
```

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

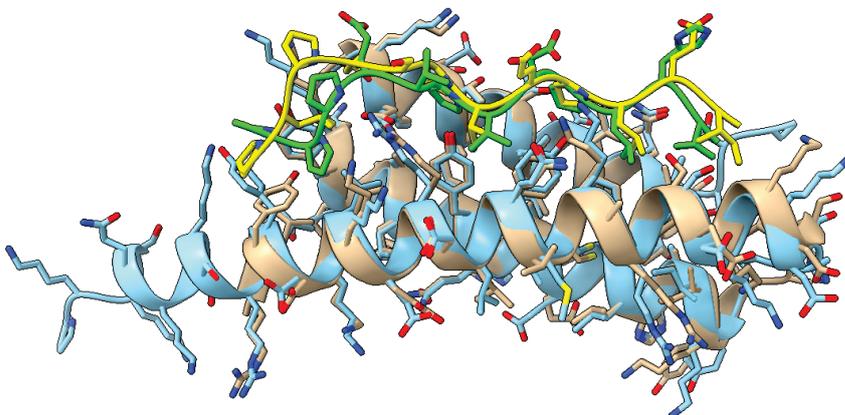
The model isn't super confident for the CHMP6 peptide, but overall the model is reasonably confident in the relative orientation of the peptide versus VPS4:



Plots for test_ea7fa



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...