Student level introduction to autoBUSTER tutorial: Refinement and model building of a molecular replacement from a homology model of SULT1A3 Question sheet.

See: https://www.globalphasing.com/buster/wiki/index.cgi?AutoBusterExampleIntro

Your name: ***Oliver Smart Model Answers***

Part A: calculating a map for the initial MR model.

Question A1: what do the following options do in autobuster refine?:

-p	**specify pdb file to be refined**
-m	**specify the mtz file for refinement**
-M MapOnly	**just calculate a map without refinement**
-d	**specify output directory**

Question A2: what are the R and Rfree values reported by autoBUSTER refine?
R value: **0.4317 (or 43%)**
Rfree value: **0.4207 (or42%)**
Note from these values you cannot tell whether the MR has been successful or not.

For questions A3, A4, A5. You should be using coot to look at the autobuster map of the initial MR model around residue TRP A 155:

Question A3:.Residue 155 is a TRP in sult1a3 but is a PHE in estrogen sulfotransferase. Are these amino acids similar? How do they differ.
tryptophan and phenylalanine side chains are both hydrophobic rings. tryptophan is larger than PHE and has an NH group that can form hydrogen bonds. The substitution F155W is regarded as conservative in the sequence alignment (marked by a :).
Question A4: Is residue A155 complete in the MR model?
No it should have a double ring but it is incomplete looking like leucine. The reason is that BALBES cuts back a residue to whatever atoms are in common between the template and target.
Question A5: What do the autobuster maps show about A155?
The difference map has a strong positive feature showing there should be extra atoms. The 2Fo-Fc map is flat and about the right size for a tryptophan ring.

For question A6 to A9 look at residue A 272 GLN that is adjacent to TRP A155. This residue is a GLN in sult1a3 but is a LEU in estrogen sulfotransferase.

Question A6:.Is the sidechain of A272 complete? Is it in density?
No the GLN lacks the =O and NH2 groups. It is not in density.
Question A7: How are LEU and GLN sidechains similar and how do they differ.
They are similar size with the same number of atoms. But leucine is hydrophobic and glutamine has amide group that can form hydrogen bonds. The substitution L272Q is regarded as not conservative in the sequence alignment (marked by a .).
Question A8: looking at the autoBUSTER maps around the backbone of A272, is there indication that the model needs correction?
Yes the CA atom is not centered in the 2Fo-Fc and there is a big blob of positive difference density showing the backbone should be moved.
Question A9: What is the backbone secondary structure around residue A272?
alpha-helical, the carbonyl oxygen of A272 forms a hydrogen bond to the main chain N of residue 276. i, i+4 hydrogen bonds

For question A10 to A14 you should be using the "Validate" "Unmodelled Blobs" feature to look for unmodelled blobs in the difference map.

Question A10:.look for arginine and lysine residues that are close to the blobs identified. What residues are close to the blobs found?

There is one disconnected blob near residues A48 LYS, A130 ARG and A257 ARG. The other blob has a similar shape and involves the same residues in the B chain.

Question A11: Are arginine and lysine amino acid side chains normally charged at physiological pH?.

Yes they are both positively charged.

Question A12: Looking at the unmodelled blob in the A chain are the lysine and arginine residues close to the protein surface..

No they are both buried.

Question A13. Display just the difference map then use the + and – keys .to change the contour level to 4.0 sigma. What do you see? At what sigma level do the blobs disappear?

There are two blobs one close to A48 LYS that disappears at 5.8 sigma the other close to A130 ARG that disappears at 6.8 sigma.

Question A14. What kind of atom would be expected to produce large difference map peaks? Re-read background – what do you think the difference density is for?

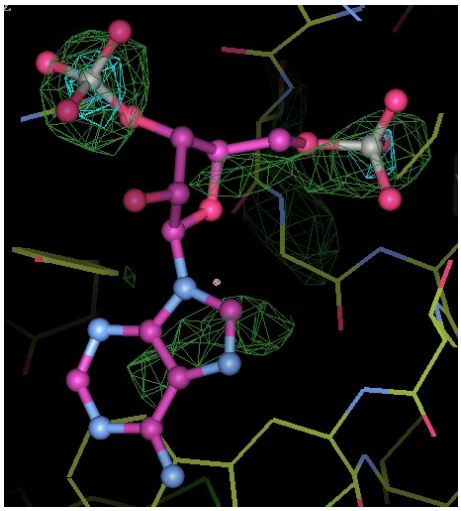
This difference density is likely to be due to an atom that scatters X-rays more than carbon, oxygen or nitrogen. In this case the enzyme uses PAPS as a cosubstrate and PAP was soaked in the experiment. The difference density will be for the PAP with the difference peaks being the phosphorous atoms of the phosphate groups. The lysine and arginine residues have positive charges that complement the negatively charged phosphates of PAP.

For questions A15 and A16 you should be have superposed 1aqu.pdb onto the A chain of the MR model.

Question A15:.What molecule is found in 1aqu.pdb into the unmodelled blob next to A130 ARG? Are the phosphate atoms in this ligand bound where you expect (see question A14)?

In 1aqu PAP is bound into this site. The residue number is 301 and the 3 letter code is 3AP. The phosphates P are in the high difference density peaks.

Question A16:.Produce as pretty a picture of the PAP molecule and difference density and attach it here



The difference density from the initial MR map at the expected PAP binding site. The difference density is contoured at 3 sigma (green) and 5 sigma (cyan) together with expected binding position for the PAP molecule from homology with EST.

Part B: Using autoBUSTER to do a quick refinement from the SULT1A3 MR model

For questions B1 to B4 you should be looking at partB.log for initial warning messages

Question B1:.How many very bad bond lengths, very bad bond angles and very bad ideal distance constants are there in the initial structure? Are there any chiral atoms that are inverted?

There are 80 bad bond length, 90 bad bond angles and 67 bad contacts. In addition there are 8 chiral inverted atoms (all are C beta in ILE).

Question B2:. What are the initial Rfact, Rfree, rms bond length and rms bond angle? (Hint look for the Ncyc=0 line in the table).

Initial Rfact=0.433 Rfree=0.422 rms bond=0.029Angstroms rms angle 2.1 degrees

Question B3:. Scroll to the WARNINGS at the start of BIG cycle 2. How many warning are there now? What residue is mentioned in the warnings.

At the start of big cycle 2 there are only 4 bad bond lengths,. 1 bad bond angle and 16 ideal contacts. There are no chiral inversions. Residue A89 GLU and B89 GLU are mentioned

Question B4:. Now scroll to the end of table reporting cycle 2 How have the Rfact and Rfree changed. What about the rms bond and rms bond angle

The Rfact and Rfree have both dropped to 0.404 rms bond to 0.010 Angstroms and rms angle to 1.0 degrees.

Conclusion refinement allows the model to adjust the models atoms positions and B factors to better fit the X-ray data and have good geometry.

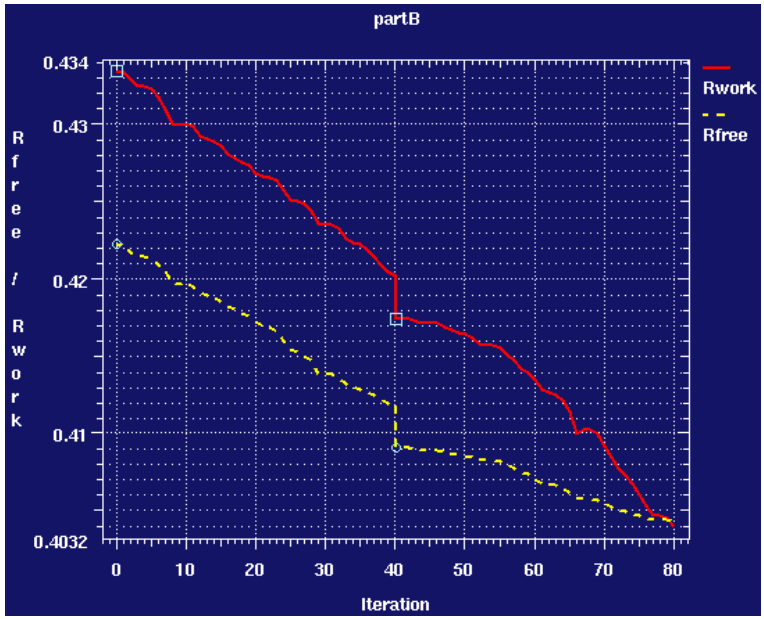
For question B5 and B6 you should be looking at looking at a graph of Rwork and Rfree produced by the graph_autobuster_R command. Rwork is the Rfactor measuring the agreement between calculated and observed structure factors that are used in the refinement. Rfree is the Rfactor for the validation set that are not used in the refinement. In this case Rfree starts below Rwork by chance.

Question B5:How Rwork and Rfree vary during this refinement?

Both Rwork and Rfree drop during refinement but the drop in Rwork is larger. At the end of the refinement Rwork and Rfree are about the same. Further refinement can be expected to produce Rwork lower than Rfree as can be normally expected. The sharp eyed will note that at iteration 40 there is a sudden drop in both R's this is at the end of big cycle 1 and start of big cycle 2. The reason is that the solvent mask is recalculated at the start of each big cycle but then becomes inaccurate as the model moves. The accurate values at the start of each big cycle are marked by squares/circles.

Question B6:Do you think the refinement has converged.

Clearly not both Rs are still dropping so lets run a longer refinement.



Part C: Using autoBUSTER to do a long refinement from the SULT1A3 MR model

For questions C1 you should be looking at the graph of Rwork and Rfree in the run

Question C1: What does the graph show? Has Rwork dropped below Rfree? What is the approximate %Rfree and % gap between Rwork and Rfree.

The graph shows that the Rwork and Rfree have plateaued so further progress will need revision of the model (started in D). Rwork has dropped below Rfree as can be expected. Final Rfree is just below 32% around 10% below the starting value. The gap is about 4.5% which is reasonable.

For questions C2 & C3 look at partC.log

Question C2: Are there any geometry warnings at the start of the last big cycle?

No there are no WARNING messages after cycle 3.

Note that geometry WARNING are only for really bad geometry problems just because they are no warnings does not mean the geometry is OK (see below).

Question C3: What is the final rms bond and angle deviation in run partC?

Final rms bond 0.010Angs, rms angle 1.1 degrees.

Note that by default autoBUSTER refine adjusts the weight on the X-ray function to achieve an rms bond of 0.010 Angs.

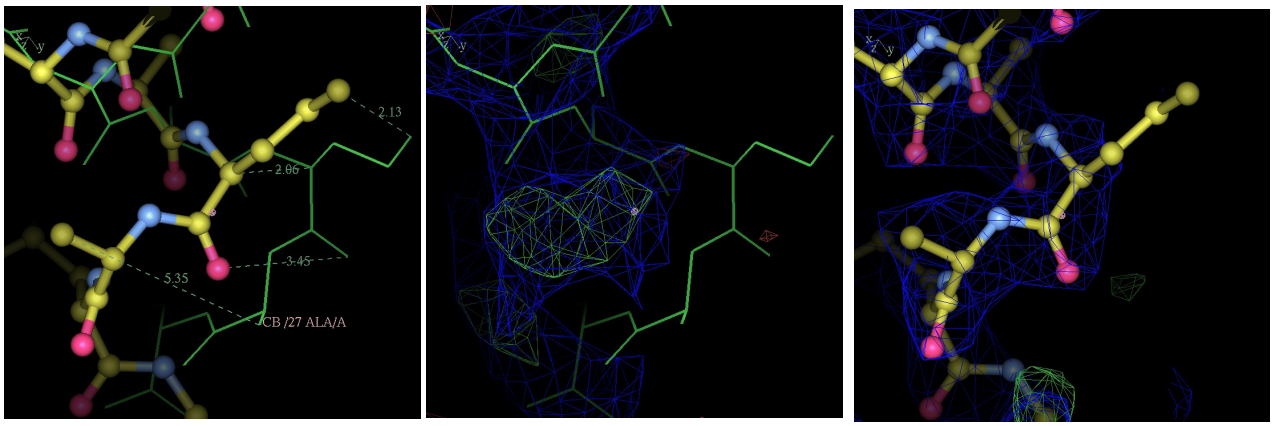
For questions C4 & C5 you should be using coot to look at the initial and final structures of the partC refinement for an area with big movements.

Question C4: What is the biggest atomic displacement you can find? (Hint look around residue A 27).

Atom CB from residue ALA A27 is displaced 5.4 Angstoms. Atom CG from residue PRO A50 is displaced 5.0 Angstoms.

Question C5: Now load the partA/refine.mtz and partC/refine.mtz to get the maps and look at the area with the big displacement. What has the refinement done? Does the 2Fo-Fc map look the same at the start and finish. If you wish illustrate your answer with snapshot pictures.

The protein has swung into 2Fo-Fc density. The 2Fo-Fc density itself is less noisy. Final difference density shows where sidechains have to completed or rebuilt.

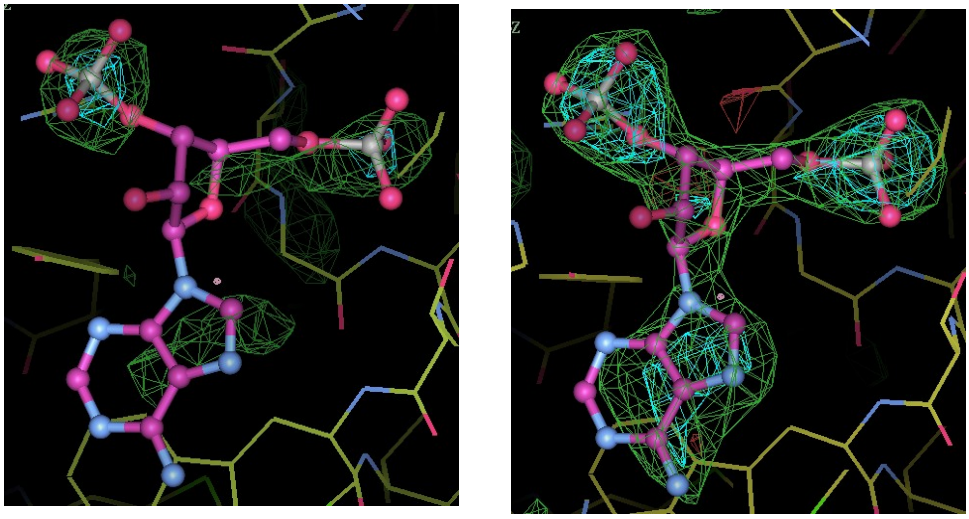


Refinement from the initial MR model shows large movements around residue A27. Left panel shows the initial model in green lines and the autoBUSTER refined model in ball lines. Center panel shows the initial model and its 2Fo-Fc map contoured at 1.5 sigma blue and difference map contoured at 3 sigma in green/red. Right panel shows the final model and maps. The green blob at the bottom of the model is due to residue A30 PRO being incomplete.

For questions C6 you should be using coot to look at the difference density at the expected PAP binding site after refinement.

Question C6: What has happened to the difference density at the expected PAP binding site compared to that in the initial MR map.

The density for the as yet unmodelled PAP molecule has become very much clearer:



Difference density for the unmodelled PAP compared to its expected position from 1aqu.pdb. On the left is the difference density from the initial MR solution. On the right the density after refinement is show. In both cases difference density contoured at 3 sigma is in green and 5 sigma in cyan. It should be noted that the PAP binding site is well conserved so that the density improvement is due to overall model improvement.

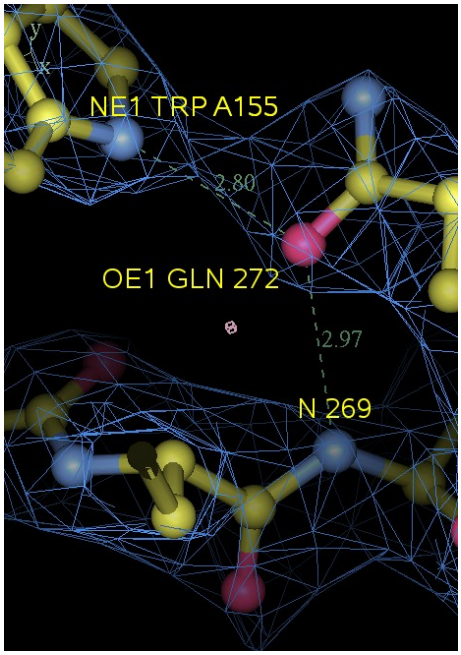
Part D: Manually building three residues using coot

For questions D1 you should be running coot looking at the model and map from partC looking at residue A 30

Question D1: What amino acid is at residue number 30 in the A chain? Is the amino acid complete? Does the density indicate how the model should corrected?
Residue A 30 is a proline (PRO). It is incomplete – it should have a ring to the main chain nitrogen. There is difference density showing where the missing (CE) atom should be located.
Now follow the instructions on the wiki page to correct this residue.
Question D2: What happens when click on an atom and drag it in coot? What happens when you ctrl-click and drag an atom? What do the traffic lights indicate?
When you drag the molecule around in coot then coot rerefines it when you release. Ctrl-click and drag moves a single atom around rather than distorting the whole molecule.The traffic lights indicate whether bonds, angles, planes, chirals etc can be satisfied.

For questions D3 you should be running coot and have rebuilt residues 30, 155 and be considering the placement of residue 272

Question D3 How should the sidechain of GLN 272 be placed and what interactions can it make?
It should be placed so the the amide carbonyl oxygen (C=O) forms hydrogen bonds to the NH in 155 trp and main chain NH from residue 269. Good hydrogen bond is suggested by the geometry of the interactions with distance of 2.8Å and 3.0Å (before refinement) and the nitrogens pointing so that the NH bond would point directly towards the carbonyl oxygen. If placed the other way around the NH groups would unfavorably point towards one another.



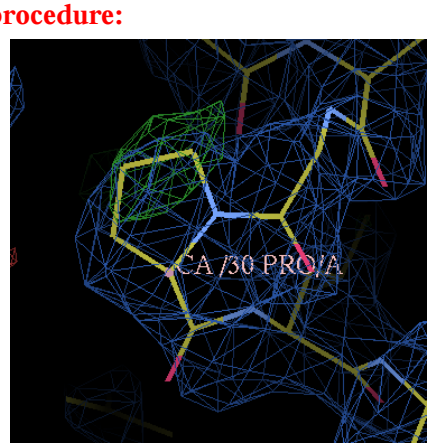
Question D4 Is there any difference around the changes you made?
No the 2Fo-Fc density is very nice. No difference density in the area.

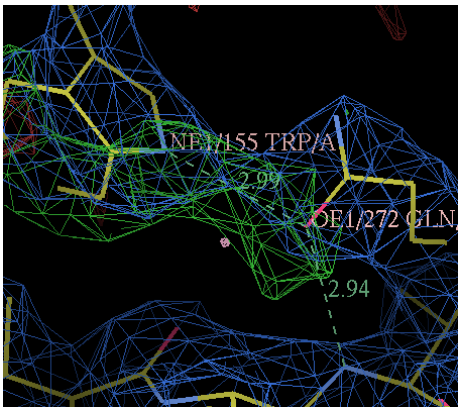
Part E: Using coot automated tools to rebuild and refit all the side chains

For question E1 you should have used coot to fill partial residues in the partC model/map

Question E1: Has the automated sidechain fixer made the same changes to residue 30, 155 and 272 that you did.

Yes residues 30, 155 and 272 have been placed very sensibly by the automatic procedure:





For question E2 you should have refined the structure with the coot fixed up side chains.

Question E2: What is the Rfree at the end of partE refinement? How does this compare to the Rfree at the end of partC?.

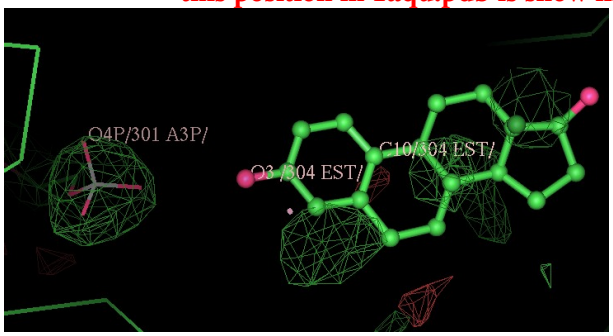
Rfree 0.289 compared to 0.315 at the end of partC.

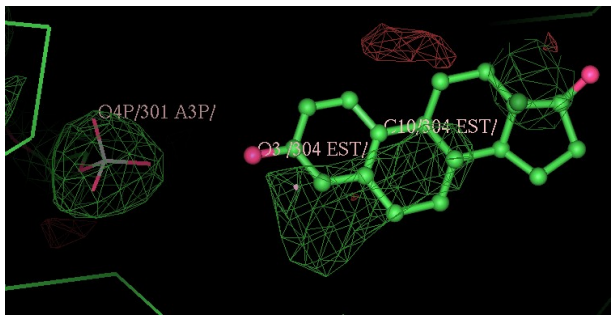
Only for advanced users with enough time.

Question E3: Has improving the side chains of the model improved the density at the expected PAP binding site? What about at the expected dopamine binding site? Hint dopamine could be expected to bind somewhere here the phenol like end of estradiol in 1aqu.pdb For question E2 you should have refined the structure with the coot fixed up side chains.

The density in the PAP site is so clear in partC it is not improved by partE. The difference peak for the stronger phosphate now disappears at 13 sigma rather than 10.8 sigma! So there is a stronger signal.

However for the "dopamine site" the difference density at 3 sigma does improve – there is a hint that something the size of dopamine (phenol with a tail) may be bound. Difference density at 3 sigma from partC (before side chain rebuild) on left and that from partE on right. The estradiol bound at this position in 1aqu.pdb is show in green ball/stick:





There is not enough density to place the dopamine (yet) but if we further improve the model then we may get enough!

Part F: **

To be written