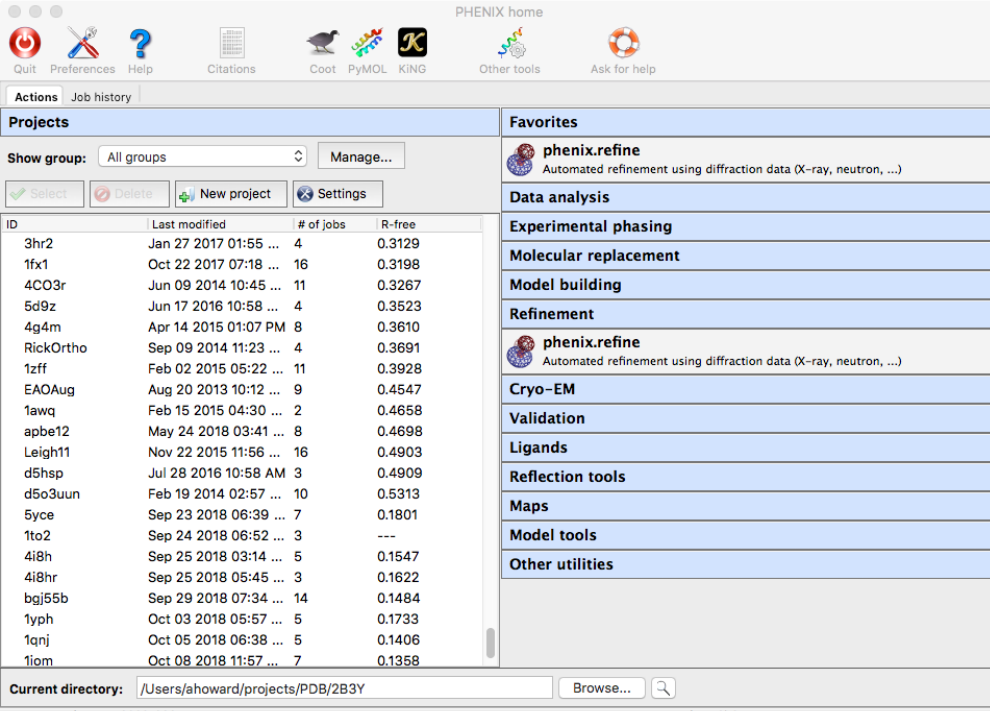# Specifics for assignment due 22 October 2018

You have already obtained access to phenix and Coot. The point of this assignment is to actually get experience using these toolkits (remember that "Coot" stands for "crystallographic object-oriented toolkit") to manipulate a crystal structure.

Begin by loading the three files related to bovine trypsin structures that were given in the assignment itself, namely the coordinate file for structure 2XTT, the sequence for structure 4I8H, and the structure factors for 4I8H. Put all three of those files in a folder (or "directory" if you prefer traditional Unix nomenclature). I encourage you to *look* those three files: they're all text files, so you can read them with any text editor or word processor. The exercise, then, is to rotate and translate the coordinates for structure 2XTT into the data for structure 4I8H, and then refine that structure and manually fit the coordinates so they more precisely align with the electron density while retaining their chemical integrity. While you're looking at things, I suggest you go to the Protein Data Bank (rcsb.org) and look at both of those structures' descriptions on the PDB site. Spend a few minutes looking at the structures' descriptions.
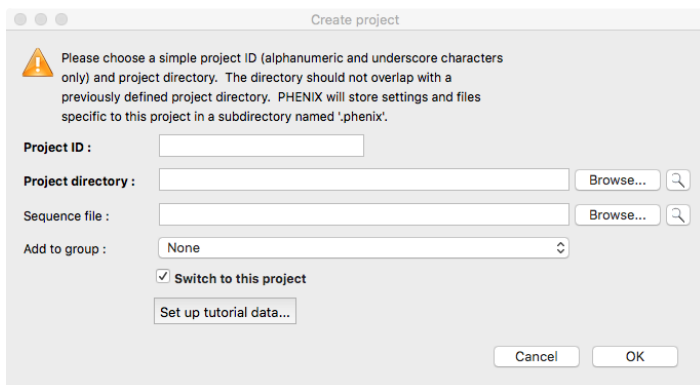
Having loaded the files into your directory, you can now invoke phenix. You should see a top-level menu that looks like this, but without the list of already-completed projects on the left side:

Click the button that says "New project" near the upper left-hand corner. You should see a menu that looks like this:

Type in a reasonable project ID, e.g. 4i8hs, and click on "Project directory". Use that browser to find the folder where you put your data: in my example, that would be /Users/myname/trypsin. Click on the folder name in the browser menu and click "Open". You'll return to the "Create project" menu, where you can specify a sequence file by clicking the browser to the right of "Sequence file:". Find the sequence file (4i8h.fasta.txt), click it, and click "Open". Leave the rest of the "Create project" menu alone and click OK. That will return you to the PHENIX home menu.

Click on "Molecular replacement" on the right to show the following:

Click on "Phaser-MR (simple one-component interface)", to bring up a menu that looks like this:
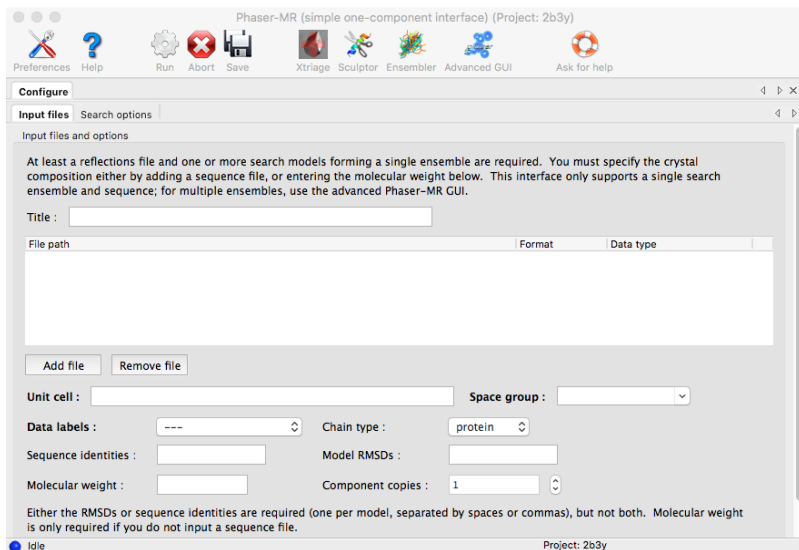
Enter a title, e.g. "phaser for 4I8H using 2xtt's coordinates", and then click on "Add file". You'll see a file browser containing your three input files. Select all three of them and click "Open": you can do this one at a time or all at once. One way or the other, when you're returned to the Phaser-MR menu, you should be able to see all three files, and their "Data type" entries should match what you expect them to be:

2xtt.pdb should be listed as "Search model", 4i8h-sf.cif should be listed as "X-ray data", and 4i8h.fasta.txt should be listed as "Sequence".

You should be able to see the unit cell for 4i8h in the line below the file menus. You can leave the defaults as they are for all but one of the items on the Phaser-MR menu. The one thing you'll need to specify is the "Sequence identities", which you can set to 90, meaning 90% accurate. Then click on the other tab within that menu, which is labeled "Search options". You should a menu that looks like this, apart from the directory names:
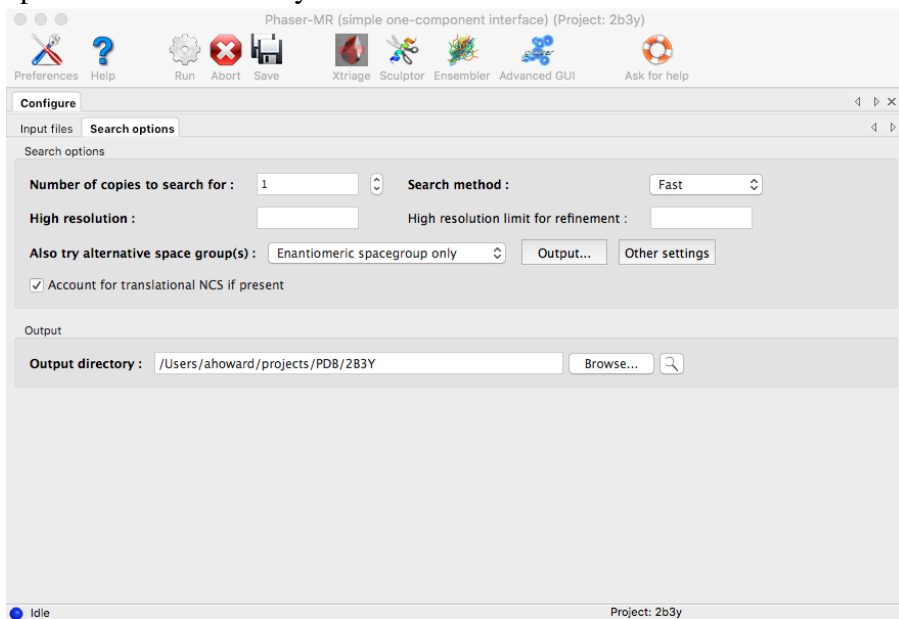
You could leave everything on this menu alone, but to save a bit of time I'd recommend pulling down the pulldown labeled as "Also try alternative space group(s):" and click "No alternatives". Then close that tab. The result should look something like this:
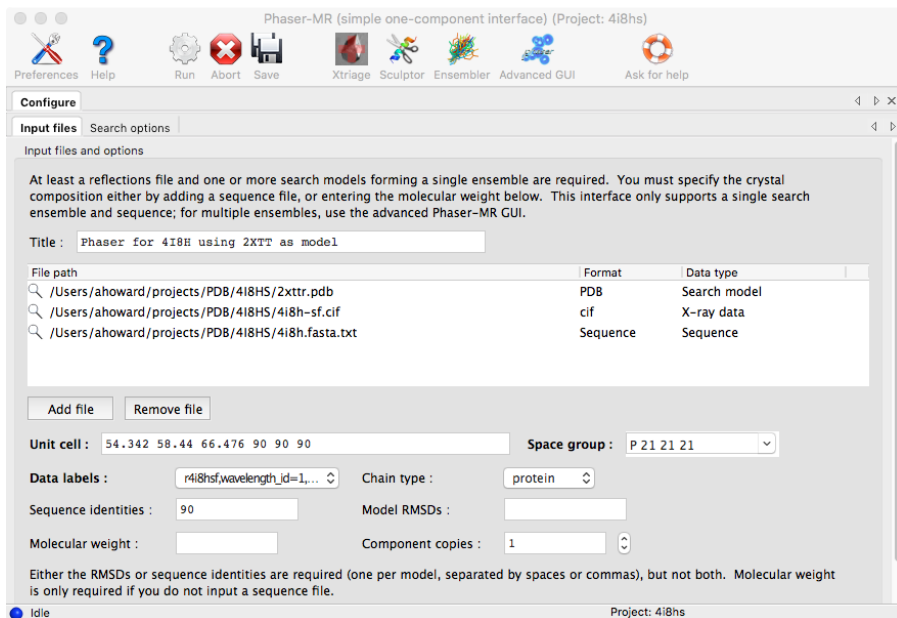
Go up to the options at the top of phenix's "Phaser-MR" menu and click "Run". It will give you two choices: "Run now" and "Edit parameters and run". Click "Run now" and let it start. You should see some output in the Phaser-MR box, and after about fiteen minutes (on my laptop, anyway) you should get an output that looks something like this:
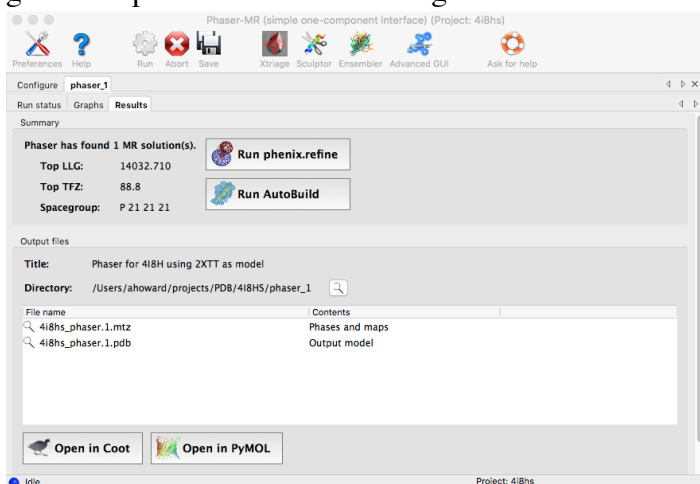
Do look at the numbers near the upper left-hand corner. The log-likelihood gain (LLG) should be above 5000, and the TFZ should be above 40. If it is, you're ready to move on.

Close the Phaser window and look at your main Phenix menu. Partway down you'll see a category called "Refinement". Click that and click on "phenix.refine". You'll see a menu that includes an Input files menu. Populate that file menu with three files: your sequence file, which should already be there; your structure-amplitude file, which is the file 4i8h-sf.cif, and the output coordinate file from your phaser run, which you can find by clicking on the folder icon for the phaser run (probably phaser_1), and finding the output PDB file, which will have a name ending in something like 4i8hs_phaser_1.pdb. Leave the other items in their default form, except the Wavelength, which for this experiment is 0.5904. After you've put those names and that value in, your Input data tab should look like this:
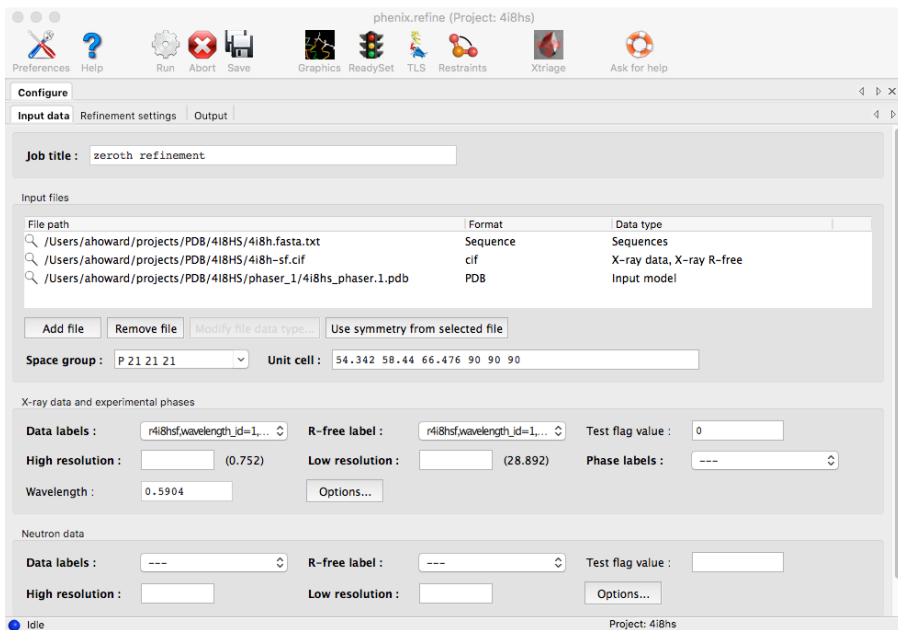
Now click on the second tab, labeled "Refinement settings". Turn on "XYZ (reciprocal-space)" and "XYZ (real-space)", "Occupancies", and "Individual B-factors" if they're not already turned on. Then click "Select Atoms", bringing up a new menu. You should then click on the "Refinement Strategy" tab. At the bottom of that menu is a question that says "Common configurations:" Pull that pulldown down and click on "All atoms except water anisotropic":
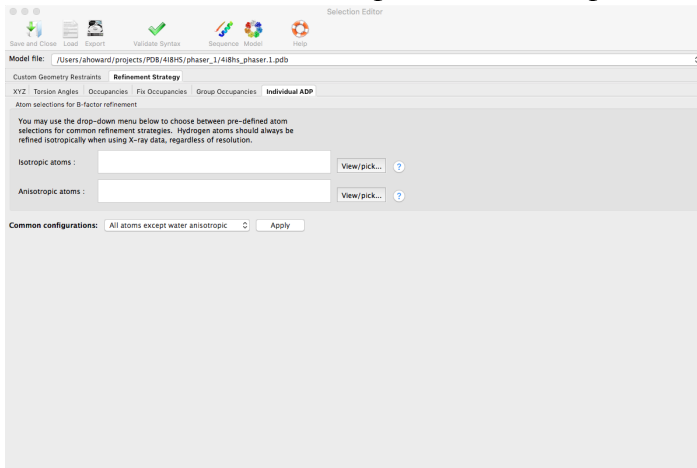
Click "Apply", and the screen will look like
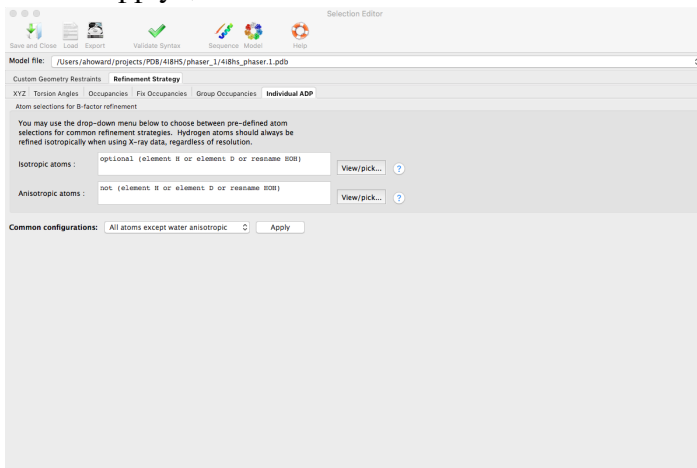
Then click "Save and Close" in the upper lefthand corner, and you'll be returned to the Options menu. You'll make a few changes on that Options menu, and then you'll be ready to run phenix.refine. Specifically, under "Targets and weighting", shift the "Target function" to ML, turn on "Optimize X-ray/stereochemistry weight" and "Optimize X-ray/ADP weight" if they're not already on. Click "Refinement target weights". You'll see a menu with some choices. Take the defaults, except change the "Stereochemistry weight scale (wc)" to 1.2. Click OK on that menu so it can disappear. In the "Other options" section, turn on "Automatically add hydrogens to model", "Automatically correct N/Q/H errors", and "Update waters" if they're not already on. Click the "Number of processors" according to the number of separate CPUs in your computer; for many of you that number will be 4. The final result should look something like
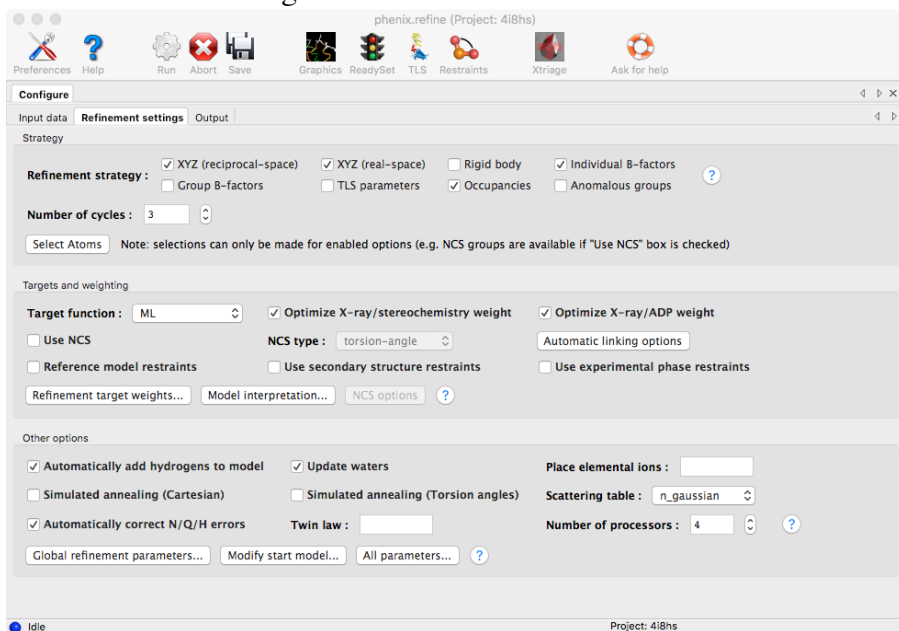
If that's roughly what you see, click "Run" and "Run now" as you did for the Phaser run.

Your phenix.refine window will obtain some new tabs. Initially you'll see the "Run status" tab, but you can click back and forth between that tab and the "Log output" tab during the refinement run. Phaser will spawn a Coot window so you can see what your structure looks like while you're refining it. With an ultra-high-resolution data set like this it will take about an hour to complete the refinement, so find something productive to do while it's running. At a step partway through, the "Run status" tab will look like
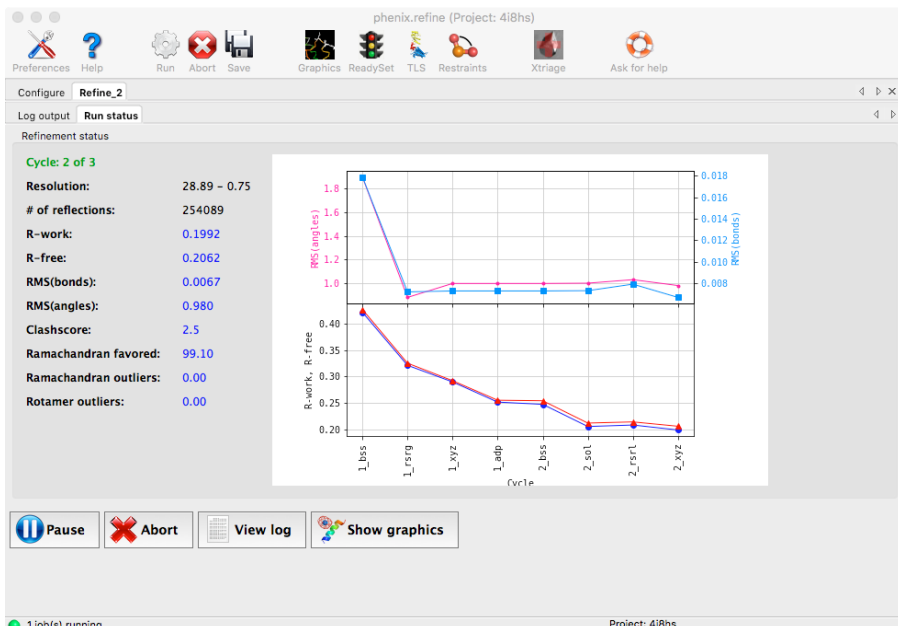
The final version of the "Run status" tab will look like
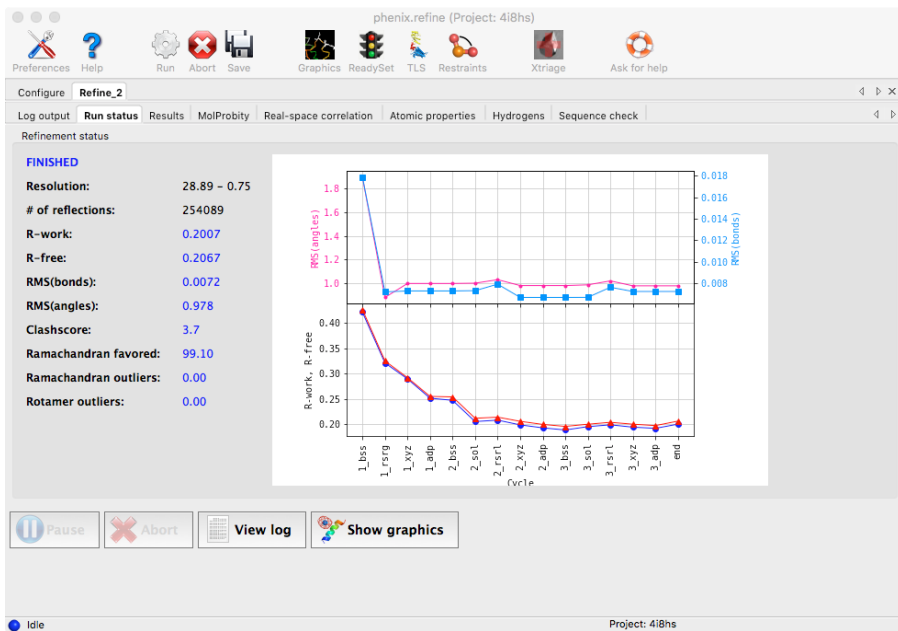
There will be other tabs that appear as your run of phenix.refine ends, including a MolProbity analysis, which will look like
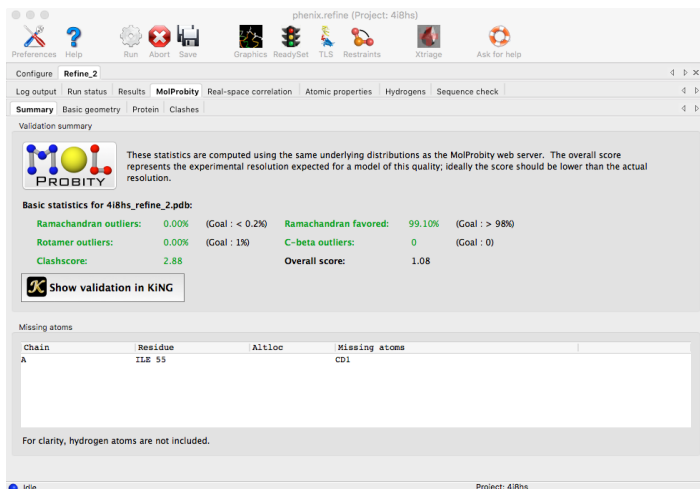
You're finished running phenix.refine and can move on to examining your structure with Coot.

If your Coot window from within phenix.refine is up and running, you can use it; if not, you'll need to invoke Coot specifically and load up the PDB file (Refine_2/4i8hs_refine_2.pdb, or a similar name) from *File → Open coordinates*, and the mtz file (Refine_2/4i8hs_refine_2.mtz or similar) from *File → Auto Open MTZ*. You should see both the model and the map superimposed on parts of the model. I won't try to walk you through every detail of manual fitting, but I'll point out a few things. Before getting into the mechanics of fitting, start your Coot run by invoking symmetry: *Draw → Cell&Symmetry*: say "Yes" to the *Master Switch* at the top, and click Apply and OK. Also, get the program to show you close contacts by clicking *Measures → Environment Distances*: click on *Show Residue Environment* and click OK. Finally, click *Calculate → Model/Fit/Refine* and, on the resulting menu, click *Select Map*, and click OK. Once you've done that, you may proceed.

1. Go to *Validate → Difference Map Peaks*, accept the defaults, and click *Find Peaks*. You'll get a list of about 240 peaks, sorted by the absolute value of the difference between the observed density and the calculated density. Probably the topmost peak will be near the carbonyl oxygen of Asn 54, the carbonyl oxygen of Val 57, the OE2 of Glu 62, the OE1 of Glu 52, and two water molecules. This is actually the location of a calcium ion in this structure. So on the *Model/Fit /Refine* menu, click *Place Atom At Pointer*, pull down the pulldown to replace *New Molecule* with the name of your current PDB file, i.e. *25 ...HS/Refine_2/4i8hs_refine_2.pdb* or something similar. Then click Ca (calcium) and click OK. That should place the calcium ion right where it needs to be.

2. The next highest peak will probably be on methionine 86, where the side-chain has two possible conformations, and since the sulfur scatters so effectively, it has a big difference peak. If you know how to add alternate conformations, do so; if you don't, skip this one.

3. The next highest peak will probably be an odd-shaped blob near the NE2 of His 40 and somewhat near the N of gly 175 and the Cα of Gln 174. This is actually a sulfate ion, so click *Place Atom At Pointer*, *SO4*, and *OK*. This will place the sulfate within the density but misorient it, so go up to the top of the Model/Fit/Refine menu, click Real Space Refine Zone, and click two atoms within your sulfate. Then click "Regularize Zone" and do the same thing. You may have to move it around a little using *Rotate / Translate Zone* to get it right.

4. The next several positive peaks will probably be water molecules. To add these, click *Place Atom At Pointer*, leave the identity as *Water*, and click OK: then move on to the next one.

5. Somewhere around peak 8 you will find a zone that looks like a benzene ring with three atoms

attached to one of the carbons. That's actually a benzamidine inhibitor that the experimenters inserted into the enzyme to keep it stable.

Click Extensions → Modeling → Add Other Solvent Molecules, and click *Add a new Residue Type...*

A dialog box will appear that says *Add new 3-letter-code/comp-id*; answer that with BEN which is the three-letter code for benzamidine. Then go back to the Solvent Ligands list, and you'll see benzamidine at the bottom of the list. Click on it, and it will magically appear in the structure. It will probably even be correctly oriented, but you may need to *Real Space Refine* it a little.

6. There are some amino acid residues within the protein whose side chains are seriously misaligned. Among them are Asn 56, Asn 61, Lys 139, Tyr 22, and Pro 74. Go to *Draw → Go to Atom* and click on the CB of each of these residues. You will see it in the center of your screen, and you can correct its orientation. In most cases you can do this with the *Rotamers...* button on the Model/Fit/Refine menu: click *Rotamers* and then step through the possible rotamers until you find one that matches the density. Click that one and then clean it up with *Real Space Refine Zone*.

7. If you look closely at the output of your MolProbity analysis, you'll see that it indicates that an atom is missing from Ile 55. Go to that residue and you'll see that, indeed, CD1 is missing. The easy way to put it in is to go to the *Model/Fit/Refine* menu and click *Mutate & Auto Fit*. A menu with the amino acids will appear; click on ILE, and the program will properly place the isoleucine.

8. Somewhere around peak 14 you will find another sulfate, in this case near the surface, not far from the amine nitrogen of gly 154. Do the same thing as you did with the first sulfate.

Once you've played around with improving the map one residue at a time like this, try a bulk approach: Click *Extensions → All Molecule → Stepped Refine*. A dialog box with a pdb filename will appear. Click OK. The program will step through the entire sequence, fitting each residue as it goes. In many places, nothing will happen; in a few, something will happen.

When you've done enough manual and automated improvements, save your work: *File → Save Coordinates*
and then respond to the dialog box when it asks for a filename.

What I will ask you to submit are:

1. The logfile from the phaser run.
2. The .txt output file from the phenix.refine run.
3. The .pdb output file from the phenix.refine run.
4. The .pdb output file from the Coot effort.

As always, I'm happy to walk you through some of these steps.