**Model Building and Refinement Practical**

**Introduction**

In this practical we will continue working with the CD44 experimental phases we determined in the MAD/SAD phasing practical.  We will begin where the previous practical finished, by inspecting a CD44 model which has been automatically built by the program Buccaneer.

1. **Getting Started with CCP4 GUI2**
   1. Launch ccp4i2 by double clicking on the icon on the desktop
   2. You will be presented with a “Welcome” screen.  Click on the link to “Start a new crystallography project”
   3. In the “Name of project/folder” field, enter cd44.  Click the “Select Directory” button and browse to /home/crystal/cd44
   4. A “Project Viewer” window will open for project cd44
2. **Inspect Output from Automatic Model Building for Errors and Make Corrections**
   1. We will use the ccp4i2 project database to organize our data – normally this would already be populated with data from data processing and structure solution, but since we are starting part way through the process for the purposes of the practical we will begin by importing data into ccp4i2.
      1. From the “Task Menu” select the “Import merged data, sequences, alignments or coordinates” and launch the “Import” task. Browse to /home/crystal/cd44 and select the file cd44\_bucaneer.mtz. Run the task.
      2. From the “Task Menu” select the “Import merged data, sequences, alignments or coordinates” and launch the “Import a coordinate set” task. Browse to /home/crystal/cd44 and select the file cd44\_bucaneer.pdb. Run the task.
      3. In the section marked “Output Data”, right-click on the atomic model icon and select “View > View in COOT”. If you are prompted about nomenclature errors, just click Yes.
      4. From “File > Auto Open MTZ...” open the file cd44\_buccaneer.mtz
      5. Use the scroll wheel of the mouse to change the contour level of the electron density map. Scroll to a value near 1.0 rmsd (this value is displayed in the top right corner of the graphics window).
      6. Click on the “Map” button in the top right-hand corner of the graphics window and select the “FWT PHWT” map for use in refinement.
      7. 1.1.5 Click on the “R/RC” button in the top right-hand corner of the graphics window to open the Refinement and Regularization control panel. Under “Weight Matrix”, set the Refinement Weight to 20.
   2. Automatic model building will only rarely produce a model which is both complete and correct. It is helpful to compare the known sequence of CD44 with the model generated by ARP/wARP.
      1. Select “Validate > Alignment vs. PIR...” and choose cd44\_buccaneer.pdb as the model.  Then choose to link chain A and the file cd44.seq (you may need to browse to /home/crystal/cd44). A “Residue Mismatches” panel showing residues present in the sequence file but different or absent from the current model will be generated.
      2. From the “Residue Mismatches” panel, select “Mutate A 2 UNK to  Ala”
         1. Inspect the map at this point.  Since Buccaneer has built an Ala residue for any unknown residues not docked with the protein sequence during model building (and marked them UNK) there is no need to add any atoms at this point. We can use the “Simple Mutate” tool (the https://lh4.googleusercontent.com/En9ew-yg6MdPoeSPWc-M8Crk9juqnoh5fi9ae2cFYwISrhXfNT-0-6EYntHPFl7n6sIspc509bQdSRwU_WLkoTh_zMhYNx0fpkdY2qikjRt4MaSUbPqITRTWzAx8bhY5icon on the toolbar) to tell coot that this residue is in fact an Ala.
      3. From the “Residue Mismatches” panel, select “Mutate A 21 UNK to  Asn”
         1. Inspect the map at this point. It should be quite clear where the side chain of Asn 21 should be placed. Use the “Mutate & AutoFit” tool (https://lh4.googleusercontent.com/z9A0lTFhVm1V2V1aqFW-jb9sZ3R-HSWUongX7vGJG0rHxIyfsi1nEzTqlUVvHFmsbNfsnHXlcAWQtNI8wclGP-cuEndeAQ8VyrRIQtW8G0rIQIBEiGfkUaR9X7M9TtTi) to mutate UNK 21 to Asn.
      4. The small loop from A 22 to A 25 has proved more difficult to build automatically (it looks like auto-tracing has followed a side-chain rather than the main-chain at one point). Fortunately, the map in this region look quite good so we will try to complete this region of the model using the loop fitting tools in coot.
         1. Use the “Delete Item...” tool ( https://lh3.googleusercontent.com/uIJXEWh7BoZHHHJcWTTs-i1SebINox2BT0mbNwG_VpSMGfzHJM0s705eUw1AyeyLd7mnojKMO5sGyHs2ePFmVYEmeqBXVO1mGQz9hs3gWMgf2R0WNHDT06SLupXWjcA) to remove the range of residues from A 22 to A 25. You can do this either by deleting one residue at a time by deleting “Residue/Monomer” or remove them all at once using “Delete Zone” and clicking on the first and last residues to be deleted.
         2. Check the fit of residue Asn 21, paying special attention to the position of its carbonyl oxygen.
         3. From the “Calculate” menu, select “Fit Loop... > Fit Loop by Rama Search…”. Make sure that molecule “cd44\_buccaneer.pdb” and chain “A” are selected. Enter residue numbers 22 and 25 as the beginning and end of the region to be built. The sequence for the loop to be built is “GRYS”. Click the “Fit Loop” button and watch as coot build your loop for you.
         4. The model now looks like it fits the map a great deal better, but there is still some room for improvement so we will carry interactively refine the region we have just built. Use the “Real Space Refine Zone” (https://lh5.googleusercontent.com/yjuPhz7wQIKPvlVe9tdn3z_OTI5FzJCQ6y9MlvS2QY41OgCIuR1_iTJiXRsBfweAvrPkamF9ueZMF3PSKNp2qJHxz7Ok_HhAAyzk6fdCxRwOqSUBeJ5sPF7utXxn0_5q) tool and click on residues immediately before and after the loop we have just built - residues Asn 21 and Ile 26 would be good ones to select.
         5. At this point a putative refined model will be displayed with carbon atoms shown in white and a panel of ‘traffic light’ indicators will appear indicating the quality of model geometry in the refined region. Are you happy with the position the putative refined residues have adopted? Do the traffic light indicators all show green, signifying good model geometry?  If so, accept the refinement and continue.  If not, try to improve the model, perhaps with help from a demonstrator.
      5. From the “Residue Mismatches” panel, select “Mutate A 93 UNK to THR”
         1. Inspect the map at this point.  To me, it looks like residue A 93 has been somewhat misplaced but residues A 94-96 are placed well in the map. We can trust their placement in this density sufficiently to trust that there are no residues incorrectly missing or added and we can therefore extrapolate the sequence back from A 97 Asp.
         2. Use the “Delete” tool to remove residue A 93.
         3. Coot allows us to mutate a range of residues at once, and can attempt to fit the newly placed sidechains in density automatically. From the “Calculate” menu, select “Mutate Residue Range”.
         4. Make sure that molecule “cd44\_buccaneer.pdb” and chain “A” are selected. Enter 94 and 96 as the beginning and end of the range and the sequence “SQY”. Tick the box to Autofit mutated residues. Click the “Mutate” button. Once again, this could use a little improvement so use the “Real Space Refine Zone” tool on the range you have just mutated.  It is usually a good idea to extend the refinement one or two residues beyond the region you have just built, so in this case residues 94 and 97 would be good beginning and ending points.
         5. We are left with a model where residues A 92 and A 93 have not been built.  Looking at the map in this region it is very difficult to see where the main chain should be traced, so we are better off leaving the residues absent. With luck, future refinement will improve the map sufficiently to allow these residues to be built.
      6. From the “Residue Mismatches” panel, select “Insert A 152”
         1. Inspect the map at this point.  There does not seem to be any density to support adding any more residues to the C-terminus of the current model.
      7. Use the “Go To Atom” tool (https://lh5.googleusercontent.com/mbXKu8_iAxBsZQU9-evvHZWx2o3OPY-4XDaCgDdsdyCJYo7HUpTZ6SNLyqbngfD0gXKblmcduie9tlYnr_EVytj-0KXZtx9772IgAw71IyiPtlDR8Ul_0Frmanvkovo) to return to residue 2 of chain A.
         1. You can now progress along the polypeptide chain by pressing “space” to move to the next residue (N to C) and “shift+space” to return to the previous residue (C to N).
         2. Work your way along the polypeptide backbone inspecting the fit between the model and the electron density map. You may attempt to correct any errors in the model using the “Auto Fit Rotamer” (https://lh3.googleusercontent.com/u8D7I7vkmvCJh8GKGQO2b9Ql2q8MMOzg17HpA8zRGNXuIz3a3CPXXpLFtBvWuH1CMZINpQciMI0AEiPO9LTIzb26AWuOhkJrIHJrMiDtctsiYZVaReYMp-BfhVE3rmWE)and “Realspace Refine Zone” tools.
         3. When you reach residue Asp 5 you will notice that the model fits the map very poorly. Use the Auto-Rotamer tool to correct the orientation of Asp 5.
         4. When you reach residue Arg 11 you will notice that the model fits the map poorly. Try using the “Auto Fit Rotamer” and “Real Space Refine Zone” to fix this error. You will notice that the resultant model is still a rather poor fit with the electron density map.
         5. Fortunately it is possible to intervene manually in cases such as this where the refinement has become trapped in a false minimum.  Use the mouse to drag the refined model (the one with the carbon atoms displayed in white) into the electron density.  You should find that the Arg sidechain will snap neatly into the map once you have dragged it in the right direction.  As long as you are happy with the geometry of this refined model, accept the refinement.  NB. *It is also possible to drag individual atoms by dragging with Ctrl+left mouse button.  In some cases this can be very helpful.*
         6. His 17 has been built without a side chain, but there is clear electron density present, so you can make use of the “Mutate & AutoFit” tool to place the sidechain.
         7. Continue to work your way around chain A, fixing errors where you find them. You may find at least one error that the flip peptide (https://lh6.googleusercontent.com/9FRpDmf-R2XhnHwZUjuEKtRWmYHSrHewh2EJkIL_e92OvagVgayowSYTZJ_dxL1xg_c-6tFakeIocrB1_pNSRJNkANWGrWwV0w8kTYOXhWYKjYhkPYn4jLX_fApBacRa) tool will help you fix.
         8. When you have reached the end of chain A (or when the demonstrators tell you that you have used enough time on this part of the practical) continue to section 2.2.8.
      8. Buccaneer did not do quite so well building the second copy of cd44 and has split it into two separate chains (B and C). Chain B contains most of the model, consisting of residues 22-151.  Although they may contain some small differences, at this early stage of refinement it is reasonable to assume that the two chains are at least similar to each other, so we can use coot to copy our edited and improved chain A to provide a good approximation of the second molecule.
         1. First we want to remove chain C, since it will be in the way.
         2. Open the “Display Manager” and change the display of your model from “Bonds (Colour by Atom)” to “C-alphas/Backbone”.
         3. Zoom out by dragging upwards with the right mouse button until you can see all of chain C (it consists of two beta-strands joined by a hairpin). Shit-left clicking on a residue will identify it, so you can make sure that you have correctly identified chain C.
         4. Open the “Delete Item...” tool, select “Delete Zone” and click on both ends of chain C.
         5. Now you can copy chain A onto chain B. From the “Extensions” menu, select “NCS > Copy NCS Chain...”
         6. Make sure that molecule “cd44\_buccaneer.pdb” and chain “A” are selected. Click “OK”.
   3. Select “File > Save to ccp4i2”.  If you accept the default values here, you will save your model as “cd44\_buccaneer-coot-0.pdb”
      1. Select “File > Exit” to quit coot.
3. **Refinement in Refmac5**We will now use the program refmac5 from the ccp4 suite to refine our corrected model against our reflection data.  We will make use of the new NCS tools in the latest version of refmac5.
   1. From the Task Menu, select the “Refinement” section and launch the “Refinement – REFMAC5” task
      1. In the section marked “Use data from job” check that the job marked “Manual model building – COOT” is selected.
      2. In the “Reflections” field, select the reflections from cd44\_buccaneer
      3. In the “Free R set” field, select the Free R set from imported merged data
      4. Leave fields “Phases”, TLS coefficients” and “Reference model” may be left as “...is not used” although all can be useful in some circumstances. We will in fact use TLS refinement, but we will allow REFMAC5 to determine TLS groups automatically, which it will do by assigning one group per polypeptide chain.
      5. Select the “Options” tab.
         1. In the parameters section, tick the box to “Use TLS parameters”. The two new fields that appear can be left at their default values.
         2. In the restraints section, tick the box to “Use non-crystallographic symmetry (NCS) restraints. The two new fields that appear can be left at their default values.
   2. Run the job.
   3. Open the Results tab for job you have just run. *This will probably happen automatically.*
      1. At the top of this report is a table showing the statistics from the refinement job. The final stats are presented in a table and the change in these values is plotted by refinement cycle.
      2. We would expect both R and Rfree to have fallen during a successful refinement. In addition, R and Rfree should not diverge from each other too greatly - this would be an indicator of over refinement.  A difference of approximately 0.05 is a good rule of thumb, although this may vary with resolution.
      3. It is also important to check that the model resulting from refinement conforms to expected protein geometry.  The summary table at the end of the refmac5 log file lists final values for Bond Length and Bond Angle showing the rmsd from library values.  The average rms for these values in the restraint library is listed earlier in the log file - it is 0.022 for bond lengths and 1.943 for bond angles.  The values for your refined model should be lower than these library averages, ideally substantially lower.  If this is not the case, you will need to re-run the refmac5 job.  If the geometry is acceptable you can continue to section 4.
   4. During a refinement job, refmac5 attempts to optimise the model against two separate targets - the experimentally measured structure factor amplitudes and prior knowledge of protein geometry.  The weighting given to each of these targets during refinement is of critical importance to achieving a successful refinement, and the correct weighting can be very sensitive to data resolution.  By default, refmac5 will attempt to automatically determine this weight but it will often require manual intervention for optimisation.
      1. Under the Refinement Stats table in the Results tab you were inspecting in 3.2 you will see reported the weight applied to the X-ray term during the refinement job you have just run.
      2. In the Job List, right-click on the refinement job you have just run and select “Clone” from the menu.
      3. In the “Options” tab, change the “Weight restraints” option from automatic to manual. Enter a value lower than that reported in section 3.3.1 to tighten the restraints on geometry. I would suggest a possible value of 0.05, but the only way to arrive at a suitable value for a given dataset and model is to test possible values and inspect the output.  
         **NB.** This value will be very sensitive to both the resolution and quality of your reflection data. A lower value will lead to more tightly restrained geometry whilst a higher value will weight more heavily towards the experimental X-ray terms.
      4. Once again, check the Results tab resulting from your refinement job.  Do you think the statistics reported by refmac5 now indicate a more acceptable model?
4. **Inspect output from refinement and check quality of the resulting model using validation tools. When rebuilding and refining a protein model it is very easy to make small mistakes, particularly at low resolution. It is therefore very important to cross-check your protein model against the large body of prior knowledge regarding protein geometry.  This is the process of validation.**
   1. From the bottom of this new Results tab, click on “Manual model building – COOT”. This job is automatically populated with the output from your refinement job, so you can simply Run the job.
   2. Set up the maps and restraints as described in section 2.1.
   3. Exactly what validation and editing are needed at this point will depend both on what editing you carried out in section 2.2 and exactly how your refinement job was run in section 3.2 or 3.3.  Here are some suggestions - please note that you are unlikely to have time to fully rebuild and validate this model during the practical session, so try to fix no more than 5 problems using each validation tool in order to get a feel for the tools.
      1. Open “Validate > Difference Map Peaks” The correct map and model should be selected by default.  The default sigma level (5.0) is also sensible, so click “Find Peaks”.  A list of peaks will then be generated - work your way down them, correcting problems as you find them. Don’t worry about adding solvent molecules at this point - we’ll cover that in section 3.3
      2. Open “Validate > Ramachandran plot” and select the current model.  An interactive Ramachandran plot will be displayed, with any outliers shown in red.  Click on any outliers you find - are there any problems with the model that you can fix?  A hint - the flip peptide (https://lh6.googleusercontent.com/N0QOGAlrzIdxoIXBlknYclKCBSYw0PYbFhrUJDJDJnQPtylpm5MNnulesil3TLeJr54u_jIm-ohO3mlR7_5r8zovY2FT2UDNbjJoNH3WjsQE9Ai4RfS6hryV1mo9nmw) tool may be useful.
      3. Open “Validate > Geometry Analysis” and select the current model.  A histogram plotting geometry for each chain in the model will be displayed, with small green bars indicating good geometry and large orange/red bars suggesting problems that need to be fixed.  Are there any problems in need of attention?
   4. Whilst examining the structure, you will have noticed that there are numerous peaks in the electron density map that would be well explained by ordered water molecules.  Coot contains tools to help you add waters to the model.
      1. Open “Calculate > Other Modelling Tools > Find Waters”
      2. The default values in the find waters dialog are acceptable for our purposes, so click “find Waters” to proceed.
      3. The new waters have been added to chain D, and it is important to inspect them and make sure that they have all been added in appropriate places. Use the “Go To Atom” tool and select residue 1 in chain D.
      4. Does this water molecule look correct? If so, hit the ‘space’ bar to move onto the next water.  If not, either move or delete the water molecule as required.
   5. Upon correcting as many errors with the model as possible, you would once again save your coordinate model and use it as input for a further round of refinement, repeating successive rounds of rebuilding and refinement until no further errors need to be corrected.  If you have sufficient time, carry out another round of refinement as described in sections 2.3 to 3.4. Have the R and Rfree values improved relative to those observed in 3.4.3 as a result of your editing?