Refinement against cryo-EM data

Tom Burnley, Oleg Kovalevskiy, Garib Murshudov

Instruct Icknield 2018

Tutorial 1 – Map inspection with MRCtoMTZ (sharpen/blur)

1) An (in)famous case! Use EMDB (https://www.ebi.ac.uk/pdbe/emdb/) to download EMBD-2984 map and the associated PDB file (PDB code 5a1a). Note the authors claimed resolution. Use chimera or coot to exam the map – does the map look visually over-sharpened? We will quantify this using MRCtoMTZ.

2) Launch the main CCP-EM gui via 'ccpem' at the command line. If don't have an active project please use 'Add project' button to create one.

3) Run 'MRCtoMTZ' and input the map and model from the EMDB. Enter the claimed resolution and set the desired B-factors to blur or sharp (e.g. try blur: 20, 40, 60 and sharp: 20 $Å^2$) and hit run. Whilst it's running – what is this doing; e.g. what happens to the high resolution components when a blurring B-factor is applied?

4) The plot shows the mean structure factor amplitude (<|F|>) vs resolution (1 / Å) plot appears can you see what happened to the high resolution components when the map was a) sharpened and b) blurred?

5) Can you see anything unusual for the plot of the unsharpened map? Should the map be blurred or sharpened and if so what are the implications for the overall resolution?

6) To visually inspect effect of map sharpening hit the coot button. This will display the model and multiple maps using each of the selected blurred/sharpen coefficients. See if you can locate the loop region shown in the talk / below. Hint: using the Display Manager, in the Molecules section select 'Colour by B-factor' and look for areas with high B-factor (coloured red). Cheat: Try chain A around residue 732.

7) Using Coot's 'Rigid Space Refine Zone' (blue sphere on toolbar try and fix the loop area identified above. You can click once on a residue at the end of erroneous loop region followed by clicking again on residues at the opposite end.

8) What are the limitations of overall blurring / sharpening? Can you suggest a good work-around now and in the future?

9) What do you consider more important overall resolution or map/model quality?!

Tutorial 2 – Refinement of a haemoglobin structure

Starting model: 5me2.pdb
Map: emd_3488.map
Resolution: 3.2 Å

N.B. We are using the original deposition of the Hb model: 5ME2. This was sub-optimally fitted and provides a good example to work with. However it has now been superseded by 5NI1.

1) Map and model inspection

Using coot load the starting model (File -> Open Coordintates) and the map.
As with the first tutorial check the map to see if blurring/sharpening is appropriate in this case. Is it necessary? Examine, is it a good fit of the model into the map?

Then check Ramachandran plot (Validate -> Ramachandran plot) and geometrical quality (Validate -> Geometry analysis) of the deposited model.

2) Refinement

a) Running Refmac5

Launch the CCP-EM GUI ('ccpem' in terminal) and select the 'Refmac5' task. Set the starting model, map and the resolution (3.2 Å). Model is already fitted into the map, so you don't need to find it – Find in map should be false. All other options should be defaults – let us see how default parameters work. To start the refinement press 'Run' button.

b) Results – overall refinement metrics

First lets examine the refinement statistics. Have the overall FSC and R-factors improved? Which of these two metrics is the most important for EM refinement and why? How RMS angles and bonds changed and what it means?

c) Results – statistics / cycle

The number of cycles to use can depend on a number factors including the quality of the experimental data quality and the starting model as well as the patience of the user! The default number of cycles is 20. Using the metrics in the statistics / refinement cycle tab would you judge that the refinement has converged? What would happen if you used 3 cycles or 300 cycles?

3) Validation

a) Visual inspection

Use coot to load the starting and refined models (refined.pdb). Visually can you see what has changed?

Examine fitting of haems (residue 201) in chains B and C. Is it better after refinement? Examine residues 92 and 118-119 of the chain C and their fit into the map.

Use the Ramachandran plot (Validate -> Ramachandran plot) and geometrical quality (Validate -> Geometry analysis) to check how they changed after refinement.

Colour the model by B-factors (in Display Manager select ‘Colour by B-factors - All’). Check the model for very low B-factors (ice blue, residues 36-37 of chain B, for instance). Do you expect B-factors that low in the cryoEM structure?

4) Re-refinement with modified parameters

During the last inspection we have spotted that geometrical quality of the model became a little bit worse, also presence of very low B-factors suggest that original map was over-sharpened. In response to that, we can now modify refinement parameters

a) Basic parameters

Fill all the basic parameters (PDB file, map, resolution, etc) as in the previous run, starting from the original model and structure.

b) Advanced parameters

Refmac cycles – shall we change it? During the log-file inspection we saw that 20 cycles were enough to reach the plateau, also it did not stay at the plateau for too long, so 20 cycles are good for this case (but you may want to change it for real-life case).

Weight. This is key parameter to control your geometrical quality. Go to the REFMAC5 log file from the previous run (Pipeline tab in the CCPEM job window, select ‘Refmac refine’ in the Task pipeline). Scroll it from the bottom until you see ‘CGMAT cycle number = 19’, which is one of the last refinement cycles. Few lines below that you will spot something like ‘Weight matrix 3.8000850E-03’, which means Refmac automatically set weight to 3.8*10^-3 = 0.0038. As we saw from our inspection, geometrical quality of the model became slightly worse after refinement, so we want to tight geometry up. For doing that, we need to switch

off ‘Weight Auto’ option and specify a value lower than 0.0038 into ‘Weight’ field. Let us put 0.0005 for a big difference.

Map sharpen. We observed very low B-factors in the model after refinement, which means over-sharpening (sounds like typical problem!). Correspondingly, we can try to blur the map and refine against blurred map. Blurring is the opposite action for sharpening, so the sharpening with negative parameter will result in blurring! Correspondingly, we want to put some negative value into ‘Map sharpen’ box for blurring, let us try -40.

Now you can start refinement (Run button) with these options.

5) Validation of the new refinement run

a) Overall refinement metrics

First lets examine the refinement statistics. Have the overall FSC and R-factors improved? What about RMS angles and bonds? Are they different comparing to the first refinement run? What does this difference mean?

b) Visual inspection

Use coot to load the starting and refined models (refined.pdb). Also you can load the model after first refinement. Visually can you see what has changed?

As before, examine fitting of haems (residue 201) in chains B and C. Check residues 92 and 118-119 of the chain C and their fit into the map.

Use the Ramachandran plot (Validate -> Ramachandran plot) and geometrical quality (Validate -> Geometry analysis) to check how they changed after refinement. Is it better now? How would you compare fit into the map after first and second refinements? Which one would you prefer?

Colour both models after first and second refinement by B-factors (in Display Manager select ‘Colour by B-factors - All’). Check the parts that had very low B-factors after first refinement (residues 36-37 of chain B, for instance). Are they better now?