



# CCP-EM Refmac Tutorial

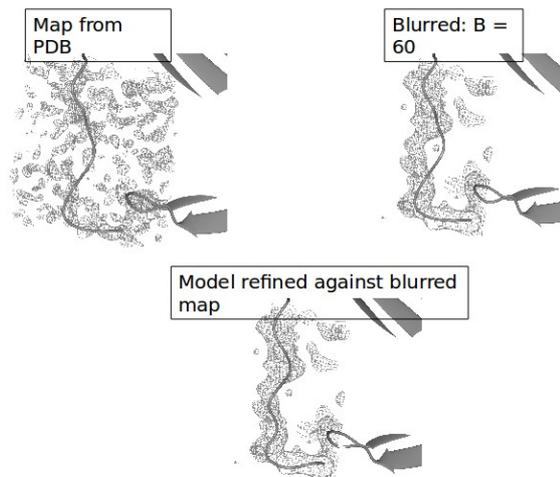
*Act without doing; work without effort. Think of the small as large and the few as many. Confront the difficult while it is still easy; accomplish the great task by a series of small acts.*

- Laozi

## **Tutorial 1 – Map inspection with Refmac-SB (sharpen/blur)**

- 1) An (in)famous case! Use EMDB (<https://www.ebi.ac.uk/pdbe/emdb/>) to download EMD-2984 map and the associated PDB file. 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 Refmac-SB.
- 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 'Refmac-SB' 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<sup>2</sup>) 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 ( $\langle |F| \rangle$ ) vs resolution ( $1 / \text{Å}$ ) 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 411+321.



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 small model with synthetic data

We will use a small model and synthetic data to allow quick running of Refmac. The data was kindly provided by Maya Topf (Birkbeck).

Starting model: 1ake\_start.pdb

Map: 1ake\_4-5A.mrc

### 1) Map inspection

As with the first tutorial check the map to see if blurring/sharpening is appropriate in this case. Is it necessary?

## **2) What's my starting model like (Coot / Molprobity)?**

Using coot load the starting model (File -> Open Coordinates). Then check Ramachandran plot (Validate -> Ramachandran plot). Are there any outliers?

The Molprobity server (Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.) will generate a more detailed assessment of the geometric quality of the starting model. Using a browser:

<http://molprobity.biochem.duke.edu/>

'Chose file' -> 1ake\_start.pdb.

Click 'Continue' then 'Analyze geometry without all-atom contacts' and then 'Run programs to perform these analyses' using the default options. Oh dear... does this look good, would you be happy to deposit this structure?!

## **3) Preparing for refinement**

The aim of refinement is to simultaneously improve the model's fit the experimental data and the expected geometric restraints. Therefore Refmac should improve the model's geometry. Before running Refmac it's good practise to visually inspect the model and the map.

Use coot to load the model (File -> Open Coordinates) and load map (File → Open Map). By default coot only displays a small window of density centred around the selected atom. You can use Edit -> Map Parameters -> Map Radius to increase this.

Does the map and model look aligned? If the model and map are not aligned set the the 'Find in map' option in the Refmac GUI to True.

## **3) 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 (4.5 Å). Decide if you want to find the model in map and ('Find in map') and also whether to apply map sharpening/blurring ('Refinement options -> Map sharpen'). To start the refinement process hit 'Run'. This will take a few minutes (N.B. this is synthetic data, with larger maps and models the process will be slower...).

#### b) Results – overall refinement metrics

Once the process is finished ('hex-em' logo will turn solid green). 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? Are these values good and should you expect such values for real data (i.e. not synthetic maps)?

#### 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?

### **4) Validation**

#### a) Visual inspection

Use coot to load the refined model (refined.pdb). To compared with the starting model you can also load the find in map model (in\_map.pdb). This is the starting model which has been fitting into the map using rigid body fitting with Molrep. Visually can you see what has changed? Use the Ramachandran plot (Validate -> Ramachandran plot) has this improved for the refined model?

#### b) Molprobit geometric validation

Use Molprobit to assess the geometric quality of the refined model. Has it improved compared with the starting model? Is it good enough to deposit?

Why is the model not as good as one would hope... the synthetic map is generated from 1ake.pdb. Try running this through Molprobit – from the main page insert '1ake' into the 'PDB/NDB code' box. What are the Molprobit metrics for the deposited PDB?

The geometry errors in the deposited PDB therefore also present in the synthetic map. In this case this is an old entry (from 1992) and refinement programs have improved since then. A useful resource is the PDB\_REDO server ([http://www.cmbi.ru.nl/pdb\\_redo/](http://www.cmbi.ru.nl/pdb_redo/)). Here the deposited models and data are re-refined using modern refinement and validation software. Now lets try some real data.

### ***Tutorial 3 (Advanced) – Refinement and validation with real data***

There are more advanced tutorials available here:

[http://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/em\\_fitting/em\\_fitting.html](http://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/em_fitting/em_fitting.html)

N.B. these are large datasets run using scripts. They will be incorporated in CCP-EM soon...