High Resolution Atomic Model Refinement in CCP-EM

Refining a flexibly fitted AF2 model of C82 subunit in the map of elongating Pol III (EMD-3178)

In this practical we will consider steps essential for high resolution model building starting from an existing experimental or predicted structure. We will solve the following problems using tasks from the CCP-EM software suite:

- *Inspecting* model quality with **Coot** and **Model Validation**
- *Automated refinement* of the atomic model with **Refmac-Servalcat**
- *Manual refinement* of the atomic model with **Coot**

We will focus on high resolution refinement of the C82 subunit of Pol III (from Practical 1 tutorial) against density maps from single particle cryoEM. We will use the following data:

- **emd_3178_5fj8.mrc**
  - *Full map, 3.9 Å resolution*
- **emd_3178_run1_half1_class001_unfil.mrc**
  - *Unsharpen and unfiltered halfmap 1 from 3D reconstruction*
- **emd_3178_run1_half2_class001_unfil.mrc**
  - *Unsharpen and unfiltered halfmap 2 from 3D reconstruction*
- **flexem_mdcg.pdb**
  - *FlexEM fitted model from AF2 prediction*

Part 1) *Inspecting* model quality with **Coot**

From the CCP-EM task bar launch Coot (or from the terminal type coot) and load the model (**flexem_mdcg.pdb**) and map(**emd_3178_5fj8.mrc**) and set the view.

  - `File -> Open Coordinates...`
  - `File -> Open Map...`

- To customise Coot for cryoEM:
  - `Calculate -> Modules -> Cryo-EM`
    - This will load an additional Cryo-EM menu
- Move the view to the centre of the density and sets the contour level
  - `Cryo-EM -> Go to Map Molecule Middle`

  If you don't see the map, but you do see it listed in the Display Manager, then it could be that the contour level is too high – try decreasing the map contour level (scroll down with the mouse, or press "-" a few times).

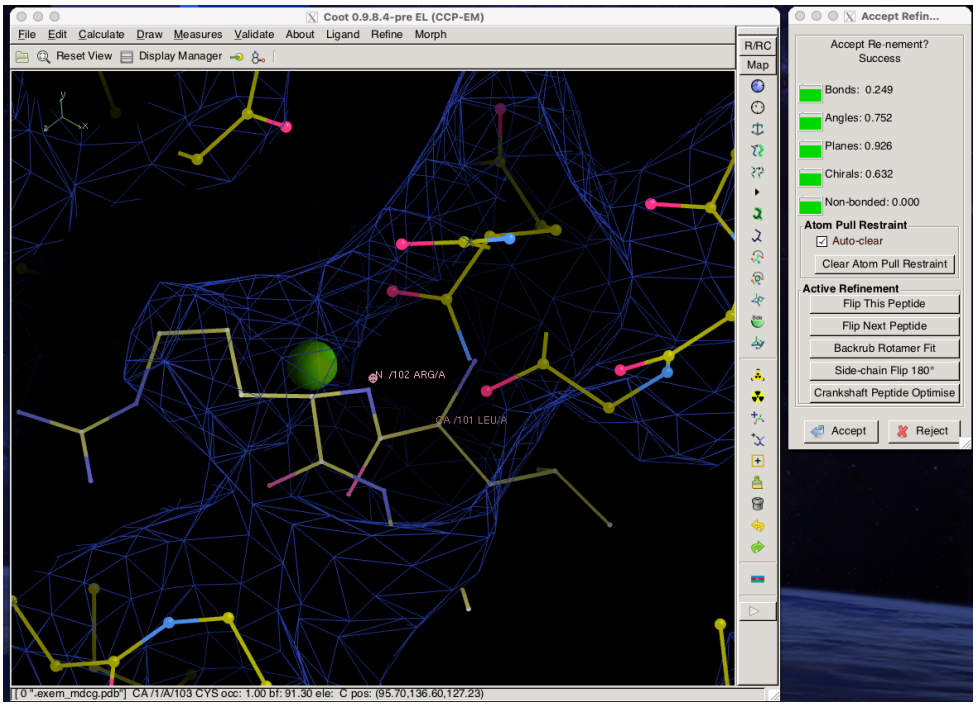
- It can be useful to display the box that encapsulates the map:
    - `Draw -> Cell & Symmetry -> Show Unit Cells? -> Yes`
- Then zoom in/out until you see the whole box (right click and drag, or use the “m” and “n” keys).
- Increase the map radius in order to display the whole map (e.g. to 70 Å in this case):
    - `Edit -> Map Parameters -> Map Radius EM.`
- If using a MacBook without a mouse:
    - `X11 -> Preferences -> Emulate 3 Button mouse`

  This will simulate the pressing of the middle and right buttons when you use it in conjunction with Option and Command keys.

Visualise the model and see how it fits in the density. To assess the model geometry go to:

- `Validate -> Ramachandran Plot`
- What distribution would you expect and does this agree? What is the reason for this?
    - `Validate -> Geometry Analysis`
- Can you find areas with poor local geometry? Click on the outliers e.g. Arg73, Arg102 etc.
    - This will centre on the residue that needs to be improved.

Part 2) *Manual* model refinement with **Coot**



We’ll use Coot’s interactive refinement tools to fix some of these errors.

- Manual refinement with Real Space Refine Zone

    - Click the third button on Coot's button menu (Blue circle under R/RC and Map)
    - Now select the "Zone" to refine. To do this click on the starting atom (~1 residue before the residue that needs to be fixed) and finished atom (~1 residue after).
    - This area is now active and these atoms can be manually moved by clicking and dragging. Try and do this until the geometry scores are acceptable
        - Also try Backrub Rotamer Fit and/or Crankshaft Peptide Optimise
        - When you are happy press Accept (or Reject to cancel any changes)
    - Re-run geometry analysis and see if your residue has improved
        - `Validate -> Geometry Analysis`

How many residues need to be improved… too many. It's more efficient to do automated refinement.

Part 3) *Automated* model refinement with **Refmac-Servalcat**

Refmac was orgininally written to refine structures from X-ray crystallography and is part of the CCP4 suite. Servalcat is written to configure Refmac for use with cryoEM data. It ensures maps can be used (instead of .mtz structure factors), maps are sharpened and weighted, SPA symmetry is used where needed, electron scatter factors are used and halfmap validation can be performed.

`(a) Run refinement`

- In the CCP-EM interface, select the Refmac Servalcat task.
- To scale and calculate difference maps Refmac Servalcat now uses unfiltered and unsharpened half maps. These should be taken from the final 3D refinement process.
- Masked refinement (previously called local refinement) creates a sub-volume around the atomic model and refines this area only. This speeds up refinement dramatically if you have a partial model and reports statistics for this area only. We will use this here.
- Enter the following parameters:

    - `Title: Initial Refmac Job`
    - `Input model: flexem_mdcg.pdb`
    - `Resolution (3.9 Å)`
    - `Half map 1: emd_3178_run1_half1_class001_unfil.mrc`
    - `Half map 2: emd_3178_run1_half2_class001_unfil.mrc`
    - `Masked refinement: select`

- Now run the job (this might take a few minutes, depending on computing power). Whilst waiting you can continue to try to manually improve the original structure in Coot.

`(b) Inspect the refinement statistics`

- Once the job has finished, click on the "Results" tab. Look at the refinement statistics table, as well as the graphs.
- Were 20 cycles sufficient, i.e. has the refinement converged?

- Is there evidence that refinement has improved the model, or made it worse? Consider statistics representing fit to the data (FSC average), as well as geometric quality (Rms bond/angle/chiral).
- Aside from the coordinate positions, what other atomic parameters are being refined? Below are the first lines of the `flexem_mdcg.pdb` structure, what does each ATOM column refer to?

```
EXPDTA    THEORETICAL MODEL, MODELLER 10.3 2022/11/02 17:08:46
REMARK   6 MODELLER OBJECTIVE FUNCTION:      2253.0767
REMARK   6 MODELLER BEST TEMPLATE % SEQ ID: 100.000
ATOM      1  N   VAL A  26     132.008 128.372 150.888  1.00 44.72           N
ATOM      2  CA  VAL A  26     132.159 127.107 150.134  1.00 44.72           C
ATOM      3  CB  VAL A  26     130.822 126.468 149.910  1.00 44.72           C
ATOM      4  CG1 VAL A  26     130.257 126.014 151.266  1.00 44.72           C
ATOM      5  CG2 VAL A  26     129.926 127.472 149.163  1.00 44.72           C
ATOM      6  C   VAL A  26     132.772 127.371 148.802  1.00 44.72           C
ATOM      7  O   VAL A  26     133.425 128.392 148.595  1.00 44.72           O
```

- Why are the statistics so good even though only a single domain is present?

(c) Visual inspection

Click on the “Coot” button at the top-left of the interface. Coot will open with two maps and both the model before and after refinement loaded and displayed.

Click Display manager and look at the two maps:

  - diffmap.mtz FWT PHWT corresponds to the sharpened and weighted full map combined from the input half maps.
  - diffmap.mtz DELFWT PHDELWT corresponds to the sharpened and weighted difference map between full map and model.
    - Hide the difference map for now

- Inspect the models. Can you see any evidence of changes in the model, or improvements to local model quality?
- Zoom out so that you can see the whole model(s). Open the Display Manager. Hide the map, and change the representation of both models to “CAs + Ligands”. Now repeatedly toggle the display of one of the models on and off. What differences can you perceive between the models?
- Change the representation of both the models to “Colour by B-factors - All”. Set the B-factor range
  - `Edit -> Settings -> B factor bonds scale factor…`
    - `Set B factor scale to 0.1 for both structures`
- Can you see differences between the two structures? Where did the B-factors from for the FlexEM structure? Look at the density map in regions with high and low B-factors, do you see differences in the quality of the data?

- Compare the Ramachandran plots corresponding to the two models and the geometry analysis
    - `Validate -> Ramachandran Plot`
    - `Validate -> Geometry Analysis`
- Has the structure improved?

Part 4) Cross-validation and optimising refinement weight with **Refmac-Servalcat**

Now let's see if we can improve the model by adjusting refinement parameters. We want to improve the fit-to-data (as judged by the FSC), without overly negatively affecting the geometry (agreement with prior knowledge) or overfitting to the data (judged by difference between fit to working and free half maps).

Firstly re-rerun the refinement as before but also select Cross validation. To clone the previous Refmac5 refinement job in CCP-EM (double click on the job, and then select "Clone" from the top left of the window) and select cross validation

- `Title: Initial Refmac Job`
- `Input model: flexem_mdcg.pdb`
- `Resolution (3.9 Å)`
- `Half map 1: emd_3178_run1_half1_class001_unfil.mrc`
- `Half map 2: emd_3178_run1_half2_class001_unfil.mrc`
- **`Cross validation: select`**
- `Masked refinement: select`

Run this job and whilst it's running try running some alternative weights in parallel. Use the clone button and then deselect auto weight and set auto weight scale to:

- **`Refinement Options -> Auto weight: deselect`**
- **`Refinement Options -> Auto weight scale: 0.5`**

You can run a third job in parallel with the auto weight scale set to:

- **`Refinement Options -> Auto weight scale: 0.1`**

Once all three have finished, compare the refinement statistics from this job and the original one.

- Inspect the work and free FSC in the validation tab. How is this affected by the weight?
- Which of the weighting schemes would you use?
- Do you need to run an additional weight scheme?

Select the optimum weight and increase the number of cycles such that the refinement converges e.g. 100-200 cycles.

Part 5) *Validation with* **Model Validation** task

Using the output of your best Refmac-Servalcat run we will perform detailed model validation to assess the quality of the model and see if further areas need to be addressed. The Model:Validation task runs a series of global and local validation scores and clusters the results together to highlight areas which require attention.

To run click **Validation:model** in the CCP-EM task menu and set the following:

- PDB: refined.pdb (from the best Refmac job)
- Input map: emd_3178_5fj8.mrc
- Map resolution: 3.9
- Method Selection -> Local density fit (TEMPy): Select
- Contour level: 0.06*

** For deposited maps the author recommended level can be found on the EMDB entry page under the experiment tab.*

Once the job has completed click through the results pages for the local and global score. In particular look at: Global: Molprobity scores and Local: Per-residue SMOC scores. Do you think this model could be improved?

Click on the “Fix refinded_0” button and this will load Coot with the map, model and a list of residues to investigate. See if you fix some of the issues in the model.

Once you have manually refined some of the problematic regions save the structure

- File -> Save Coordinates

And re-run Refmac-Servalcat and Validation:model. Model refinement typically requires several iterations of these steps until refinement converges and no further improvements can be made.

When using structures from the PDB/EMDB it is always worth re-refining them yourself. No structure in the PDB is perfect and some can contain serious errors therefore it is worth checking. Also the software for refinement and validation improves over time (we hope!) therefore older structures may also benefit from revisiting with these improved tools.

Part 6) Correcting larger regions in **Coot**

The starting structure has some large regions of density that need to be rebuilt. The alpha helix spanning Glu570 to Asn580 is in the wrong conformation. The density in this region is weak so to help we can add additional restraints to the real space fitting. To configure this

- Toolbar -> R/RC (Refine/Regularise control
  - Use Torsion Restraints [select]

- `Ramachandran Restraints [select if necessary]`
- `Mainchain Restraints -> Alpha Helix [select]`
- `Weight matrix -> Estimate`

The N-terminus is in the wrong conformation compared with the deposited structure and requires changes from Val26 to Asp45. This is difficult to do when the neighbouring domains are missing. Load the original structure (pdb5fj8.ent) for reference and see if you can recreate the correct conformation.

Refmac Servalcat references:

Yamashita, K., Palmer, C. M., Burnley, T., Murshudov, G. N. Cryo-EM single particle structure refinement and map calculation using Servalcat. *Acta Cryst* D77, 1282-129, 2021.

Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-EM. Nicholls, R.A., Tykac M., Kovalevskiy, O., & Murshudov, G.N. *Acta Cryst* D74, 492-505, 2018.

CCP-EM reference:

Burnley, T., Palmer, C.M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Cryst* D73, 469-47, 2017.

Contact:

Do please report any issues or bugs.... it's much appreciated and helps us make the software better:

ccpem@stfc.ac.uk