Flexible fitting of atomic models in cryo-EM maps using Flex-EM and assessment of local density fit
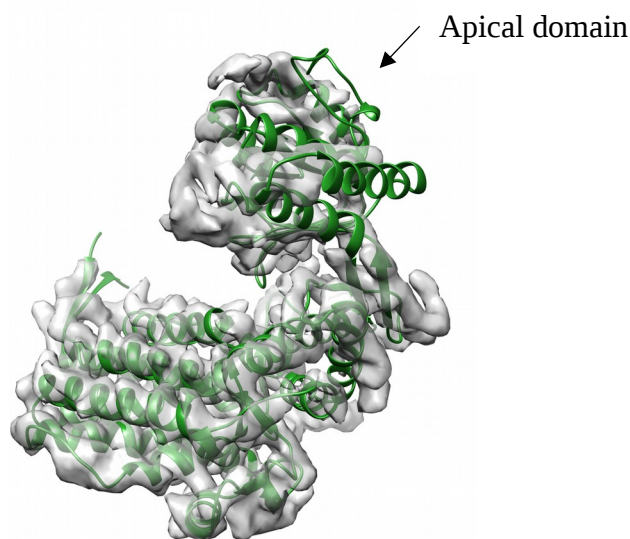
EMBO course, Birkbeck College, 10th Sept 2019

In this tutorial, we will use Flex-EM[1,2] to fit a homology model of GroEL (modelled based on a ADP-bound structure, PDB ID: 4KI8) in a map of GroEL (4.1 Å, FSC 0.143) in the unbound form (EMDB ID: 6422). The homology model was built using MODELLER[3] that uses an alignment with template structure to build a homology model by optimizing spatial restraints. Flex-EM uses MODELLER for the molecular dynamics runs, with the density score added to the calculations.

Input files:

1. We will use a single subunit density of the map which is segmented using Segger[4] tool in Chimera[5] (emd_6422_16_regions.mrc).
2. The starting homology model (4ki8_fit.pdb) was rigidly fitted in the segmented density using Fit-in-Map tool in Chimera.

Start Chimera and open emd_6422_16_regions.mrc and 4ki8_fit.pdb. It can be seen that the apical domain of GroEL is partly out of the density and need to be fitted into the density.

Apical domain

We will use Flex-EM with rigid-body restraints in a hierarchical way starting with larger rigid bodies (sub-domains) in the initial run, followed by another Flex-EM run with relatively smaller rigid bodies (secondary structures). The initial step simulates large body movements whereas the secondary structure fits are optimized in the second stage.

We can use rigid body restraints on larger segments or sub-domains in the Flex-EM run to fit the apical domain into the density.

Rigid body restraints have to be listed in a text file and this file has to be added as input for Flex-EM. Each line in the rigid body restraint file has the set of segments which are part of this rigid body, where each segment is defined by the start and end residue of the segment.

For example:

10:A 20:A 50:A 70:A
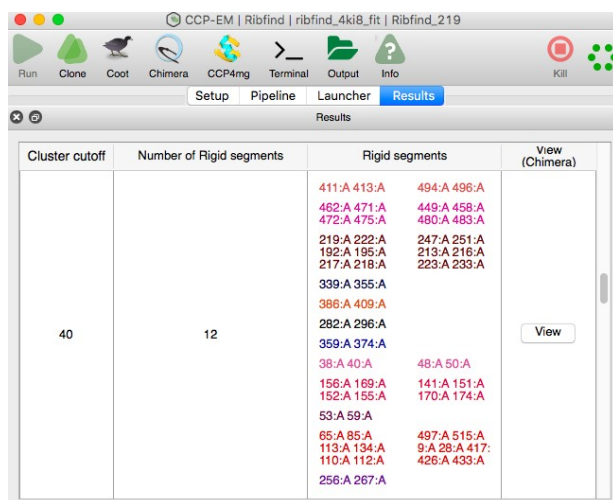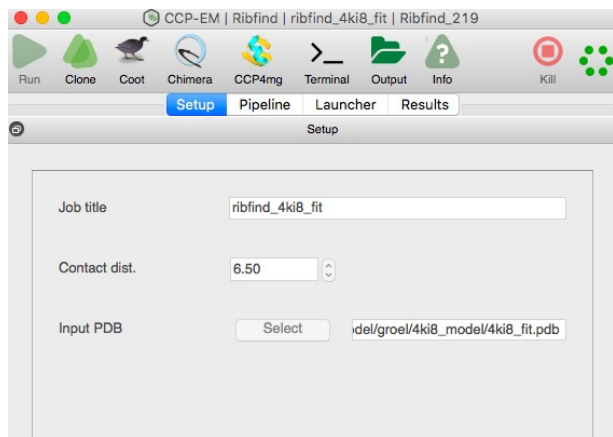
100:B 130:B 100:A 120:A

adds residues 10 to 20 and 50 to 70 of chain A to one single rigid body, and 100 to 130 of chain B and 100 to 120 of chain A to another rigid body.

We can also use RIBFIND[6] to generate rigid body files, which can then be used as an input file to Flex-EM. RIBFIND clusters secondary structures that are closely in contact along with intermittent loops into rigid bodies. Cluster cut-offs are used to consider the percent of residues in a secondary structure that are expected to be in contact (with those in another secondary structure) to form rigid bodies. E.g. cluster cut-off of 100% considers each secondary structure as a separate rigid body while a cut-off of 50% groups together secondary structure where half of residues are in contact.

We will be using CCP-EM GUI interface to run RIBFIND and Flex-EM.

A. Running RIBFIND

1. Launch the CCP-EM GUI
2. Use the starting model (4ki8_fit.pdb) as input for the RIBFIND task in CCPEM interface.
3. Click 'Run' to start RIBFIND and the results will appear in the 'Results' tab. A list of rigid bodies identified at different cluster cutoffs will be listed along with a 'view' button to see the rigid bodies colored in Chimera and the parts of the protein which do not form any rigid bodies are colored in white. A cluster cutoff of 40% groups together secondary structures into compact subdomains (lower cutoffs start to group domains together).

We will use the rigid body file corresponding to the cluster cut-off of 40 for the Flex-EM run (4ki8_fit_denclust_40.txt). The rigid body files can be accessed in the launcher tab and the job directory with different rigid body files can be accessed on the terminal by clicking the terminal icon on the interface.
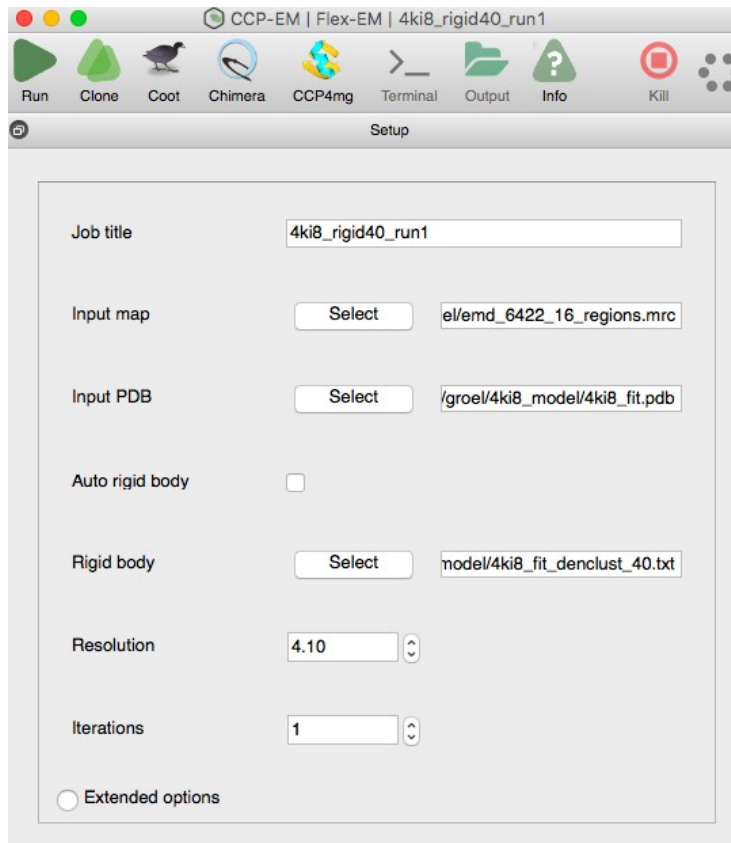
B. Running Flex-EM

1. Select task Flex-EM from CCP-EM GUI interface
2. For the Flex-EM run, use the starting model (4ki8_fit.pdb), segmented map (emd_6422_16_regions.mrc) and the rigid body file (from RIBFIND, corresponding to 40% cluster cutoff) as input. Enter map resolution (4.1Å).

Number of iterations depend on the extent to which the model has to move to fit into the map. 1 iteration is sufficient in this case (also in the interest of time) as we will follow this with another run with each secondary structure as rigid bodies.

The 'atom disp' (in extended option) refers to maximum atom displacement in every MD step (in Å), can be reduced if smaller rigid bodies (e.g. individual secondary structures) are used.
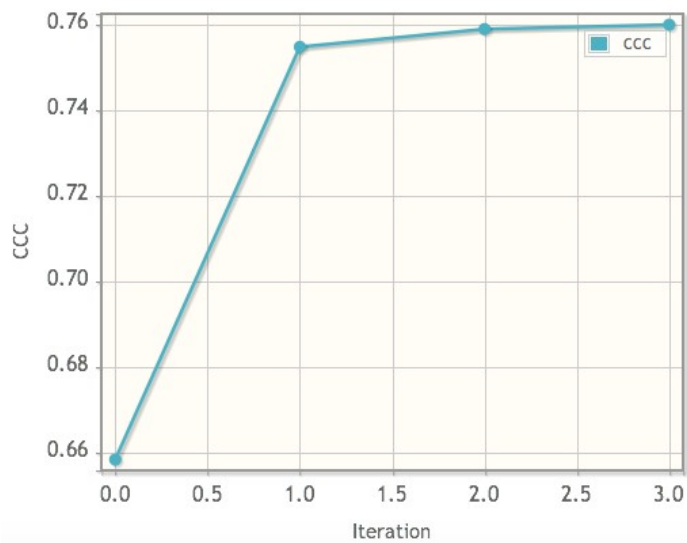
Input the rigid body file (4ki8_fit_denclust_40.txt)

4. 'Run' the Flex-EM job to start the flexible fitting process. Each iteration will take about ~30 mins in this case.

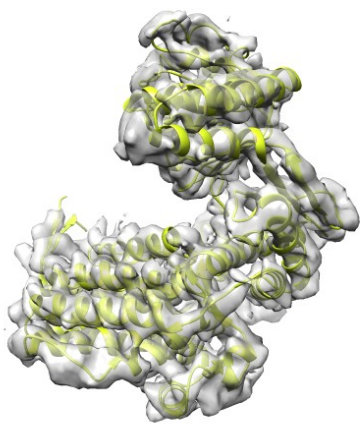Whilst the job is running the green hexagon in the top right will rotate.

You can see the progression through the various stages in the pipeline on the left. Green is finished, blue running, grey yet to start (red failed).

Clicking on the stage will display log file for that stage. Double clicking on the log file window will launch your prefered text editor with the log file.

5. Once the Flex-EM run is finished, the cross-correlation scores for models from each iteration is plotted in 'Results' tab (results of 3 iterations are shown here just to compare the models over three iterations):
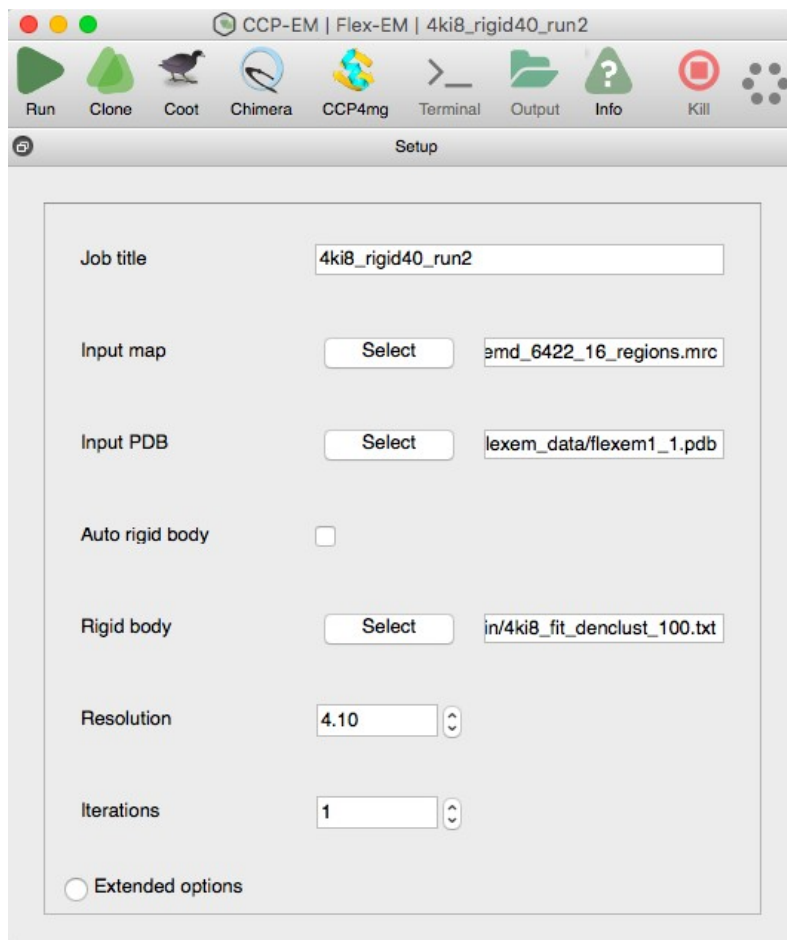
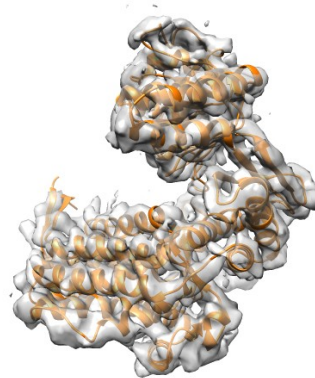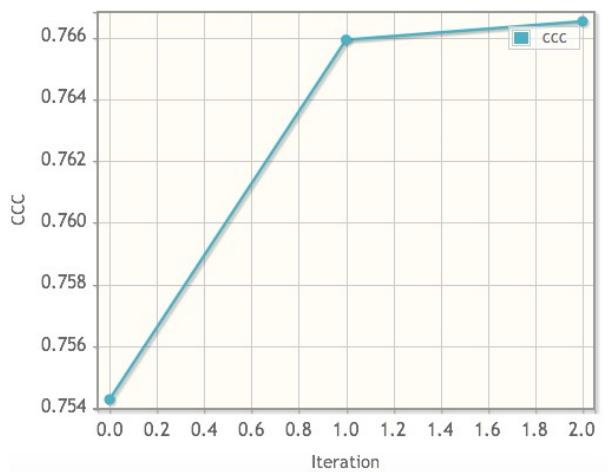The fitted models from each iteration can be accessed from the launcher or can be accessed in terminal.

To further fit the model in the density, the secondary structures can be treated as rigid bodies in the next Flex-EM run (rigid body file: 4ki8_fit_denclust_100.txt). We can use the output from the first iteration of the above run (flexem1_1.pdb in the flexem_data folder) as input for the next Flex-EM run.
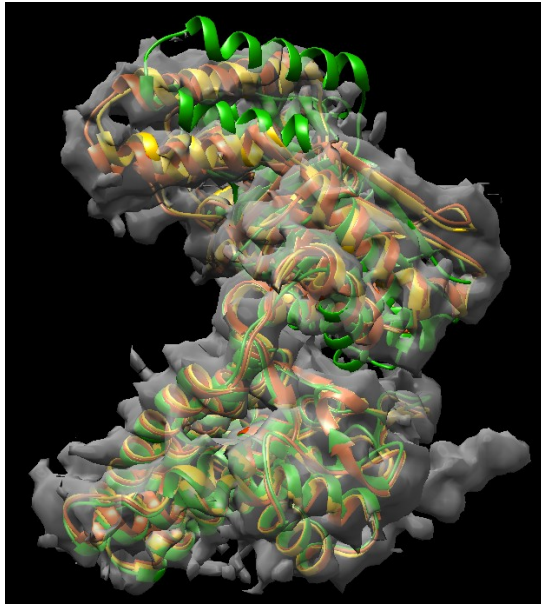
6. Launching the second Flex-EM run where secondary structures can be treated as rigid bodies. (can use the clone option in the GUI)

The output from the second Flex-EM run shows further improvement in the cross-correlation of model with density.
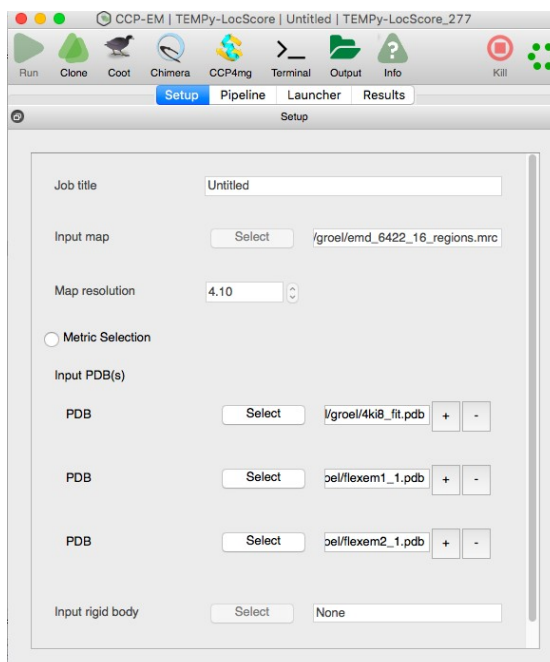

Open the fitted model from the first iteration (md1_1.pdb/final1_mdcg.pdb) in Chimera from the launcher along with the map to view the model fit in density. This model has been copied to the flexem_data folder as flexem2_1.pdb. Open the starting model (4ki8_fit.pdb) as well to compare.

C. Assessment of improvement in local density fit

To compare the fitted models and evaluate the improvement in local density fit, we will use TEMPy SMOC (segment-based Manders' overlap coefficient) score[2,7]. A local overlap coefficient is calculated over voxels covered by each residue.

1. Select TEMPy Local scores task in the CCP-EM GUI interface
2. Input the map. Input pdb files to be scored : initial model 4ki8_fit.pdb, flexem1_1.pdb and flexem2_1.pdb
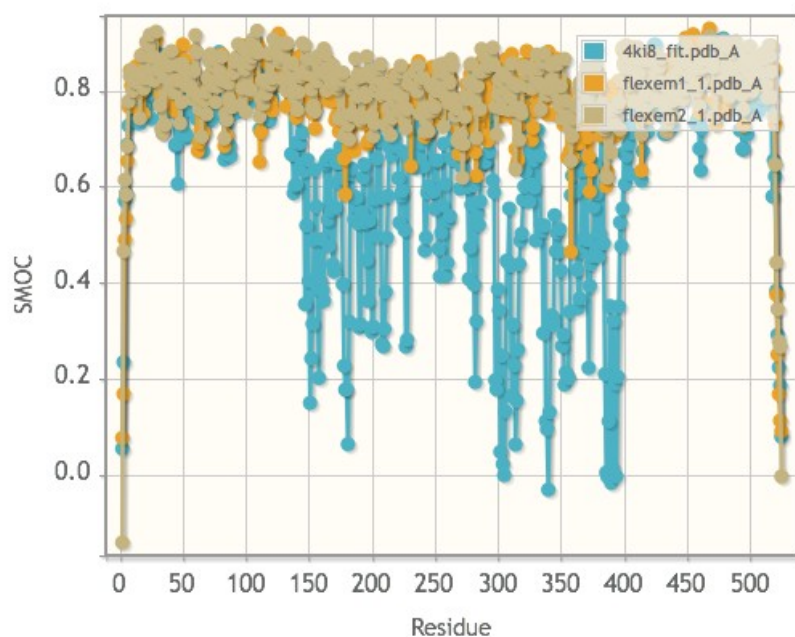

Click   Run and wait for the Results to be displayed.

We can see that the overall local fit to density improves along the chain. When you click 'View in Chimera' the models open colored based on the SMOC scores. The range of values

7

for coloring can be adjusted in the 2D Labels window and a label can be placed by mouse click and drag. OR close the 2D Labels and use (Tools->Render by Attribute->Structure Analysis, choose residues/average bfactor)

Any low scoring areas can be further fixed in Coot[8] and or followed by a round of Refmac[9] refinement to further improve the fit. Note that the residue at each chain terminus of the chain is partly outside the segmented map (scores higher when using the full map as input).

References:

1. Topf, M. *et al.* Protein Structure Fitting and Refinement Guided by Cryo-EM Density. *Structure* **16**, 295–307 (2008).

2. Joseph, A. P. *et al.* Refinement of atomic models in high resolution EM reconstructions using Flex-EM and local assessment. *Methods San Diego Calif* **100**, 42–49 (2016).

3. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

4. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).

5. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

6. Pandurangan, A. P. & Topf, M. Finding rigid bodies in protein structures: Application to flexible fitting into cryoEM maps. *J. Struct. Biol.* **177**, 520–531 (2012).

7. Joseph, A. P., Lagerstedt, I., Patwardhan, A., Topf, M. & Winn, M. Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Struct. Biol.* **199**, 12–26 (2017).

8. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot. Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

9. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).