**Fitting PolIII heterotrimer subunit C82 in a cryo-EM map representing elongating PolIII (EMD-3178)**

Homology model of C82 subunit was built based on the crystal structure of human RPC62 (PDB ID: 2XUB) : **C82\_Model\_on2xub\_A.pdb**.

To fit the homology model in the EM map, the model has to be placed at the location of the subunit and further refined to accomodate any conformational changes pertaining to elongating PolII complex.

Analysis of local resolution of this part of the volume suggests that the local resolution in this region ranges from 4-6Å. The map is globally sharpened using an optimal B-factor. However, global sharpening may lead to over-sharpening in areas of lower local resolution (or inadequate sharpening in locally higher resolution regions). As this subunit density has relatively lower local resolution, one may benefit from blurring or low-pass filtering the map. The secondary structure features become more evident in a low-pass filtered map (**emd\_3178\_lp5.mrc**) than in the globally sharpened map (emd\_3178.map). We will use the low-pass filtered map for fitting the atomic model of C82.

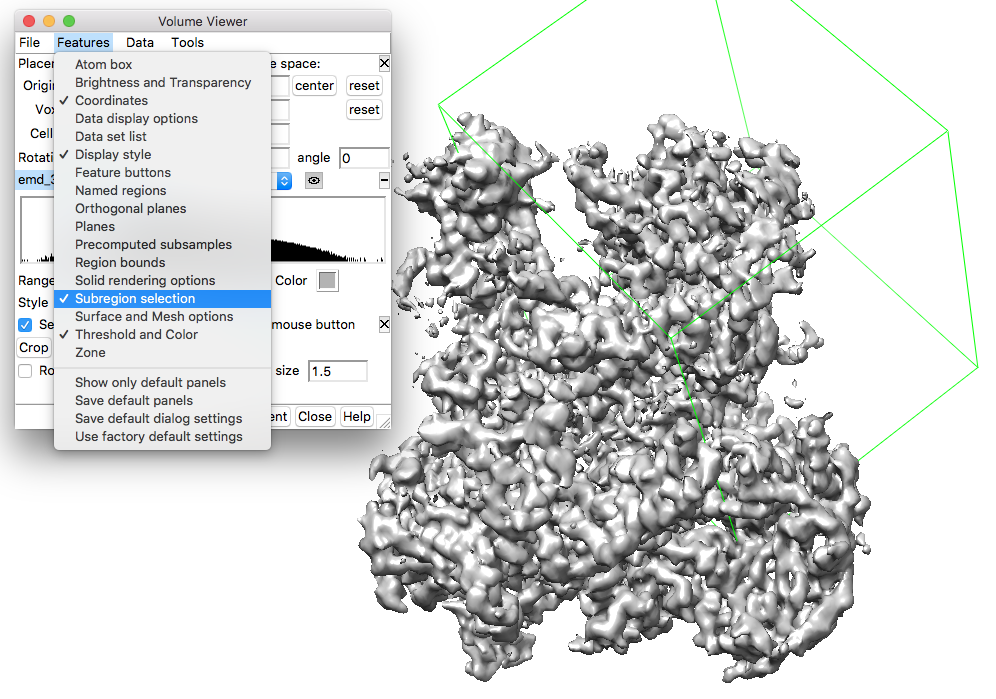
1. **Rigid-body docking of the homology model (Chimera)**

To reduce the search space (and improve accuracy of the search), cut out a section of the map volume corresponding to the heterotrimer domain.

1.1. Crop a sub-volume corresponding to the heterotrimer domain

This can be done with tools in the *Volume Viewer*

Volume Viewer : *Features/Subregion selection*



Enable mouse middle click to adjust the crop box. Centre the crop box at the heterotrimer domain and adjust the sides (middle click on the sides and drag) to cover the required region.

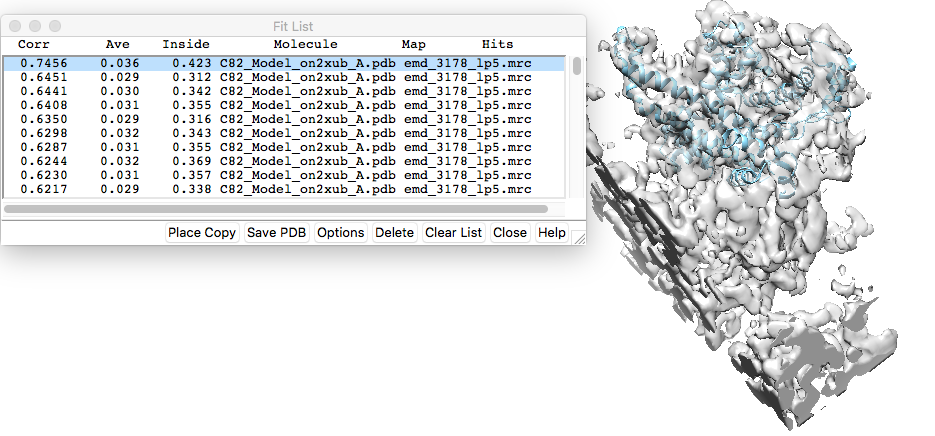
Click *Crop* on the *Subregion selection* window to get the cropped map. The cropped map can be saved (**emd\_3178\_lp5\_cropped.mrc**) using *File/Save map as* option in Volume viewer.

1.2. Fit the C82 model in the cropped map

Open the homology model **C82\_Model\_on2xub\_A.pdb**. To search the model in cropped volume, run the following command on Chimera command line (*Favourites/Command Line*)

**fitmap #1 #0 search 200 resolution 5.0**

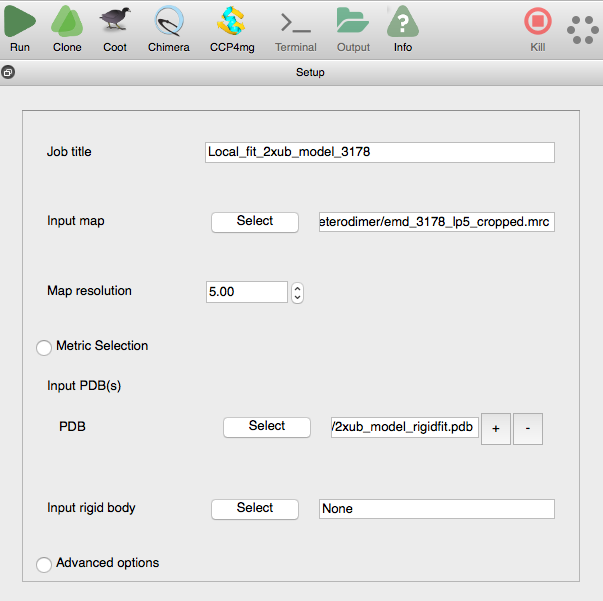
The model will be placed at 200 random starting points in the map and locally searched to maximize agreement with the density. Once the search is complete, a *Fit List* window appears with a list of model locations ranked by fit to density.

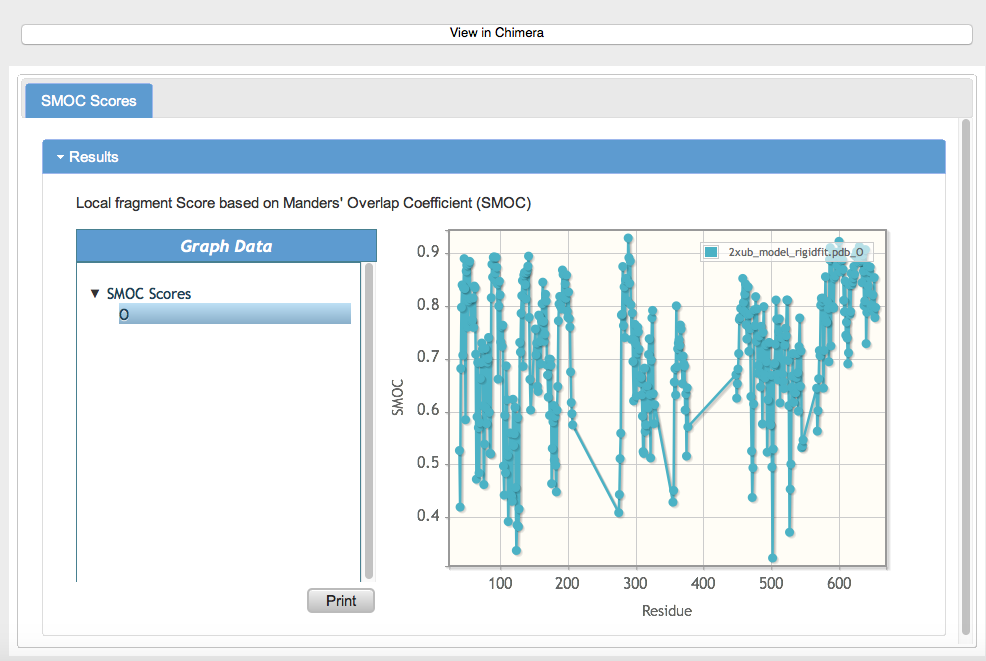


The fitted model can be saved (**2xub\_model\_rigidfit.pdb**) relative to the map from *File/Save PDB* (with *Save relative to model* enabled and the volume selected as the model relative to which the coordinates are saved).

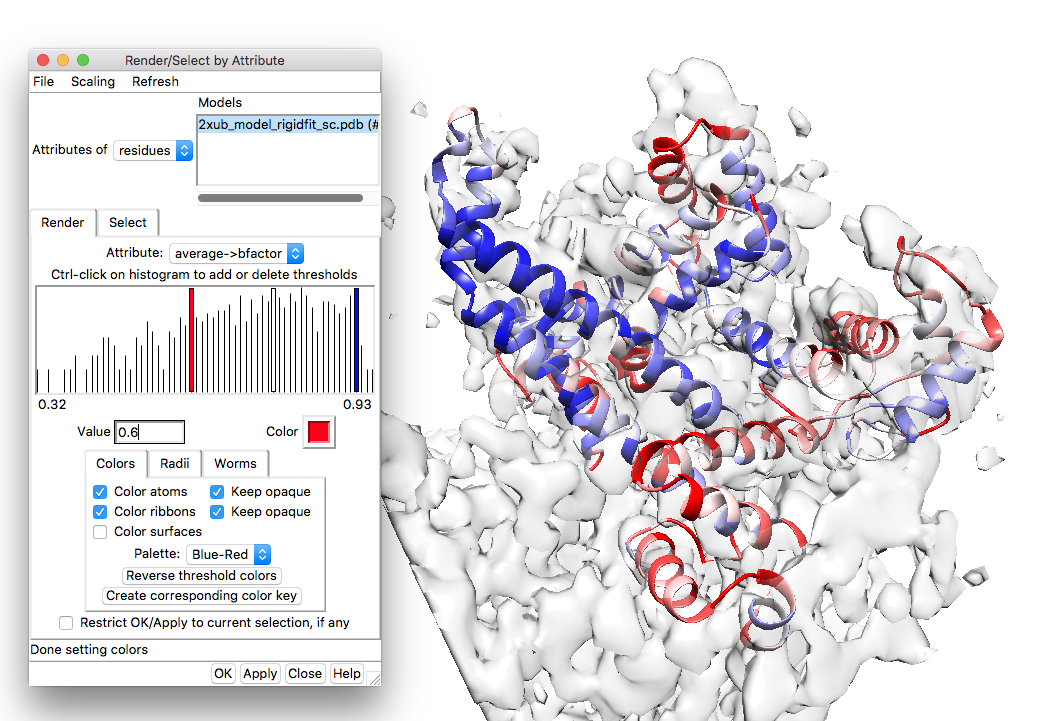
1. **Examining model fit in density**

The local fit of the atomic model in density can be examined using the TEMPy:LocScore tool in the CCP-EM interface.



The plot of local score per residue appears on the *Results* tab. The plot shows that the range of scores vary from 0.32 to 0.93. Hence there are local regions of the atomic model that are not in good agreement with the density. 

This can also be visualised in Chimera by clicking the *View in Chimera* button. This opens the model and map in Chimera with the model backbone colored by SMOC scores. The range of scores used for coloring can be adjusted further by looking at the score distribution using Render/Select by Attribute window (Tools->Render by Attribute->Structure Analysis, choose residues/average bfactor). Here we use 0.6 (red), 0.75 (white) and 0.9 (blue) for the color range.



1. **Flexible fitting of the homology model**

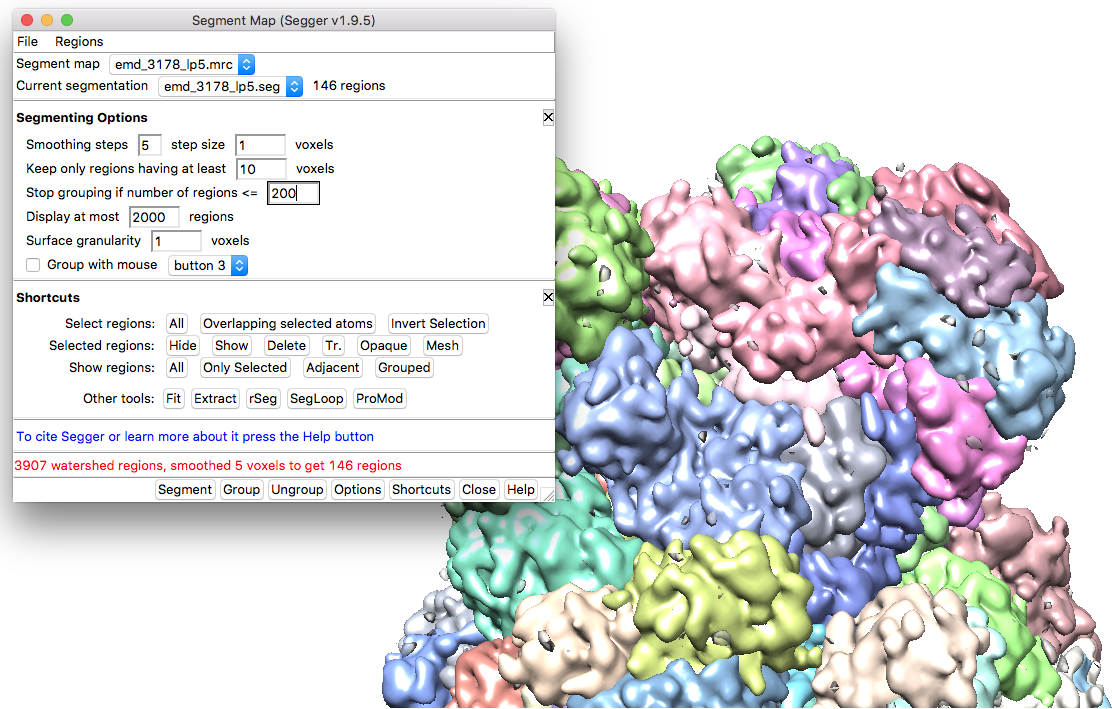
The rigidly fitted model has regions that are not fitted well in the density and hence require further flexible fitting and refinement.

For flexible fitting of the atomic model in volume density, it is often necessary to extract volume segment corresponding to the molecule of interest. Using a segment instead of the full map, makes the fitting process computationally efficient and minimizes chances of fitting into the density of neighboring subunits. We will extract a segment from the map including density *regions* traced by the rigidly fitted model.

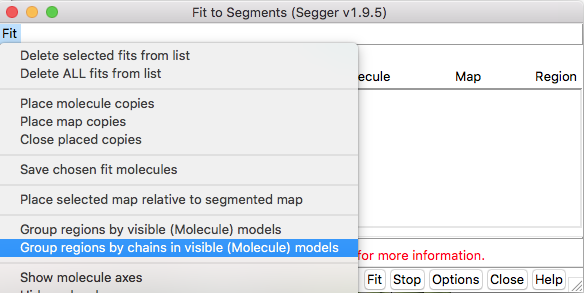
**3.1. Map segmentation with Segger**

Open **emd\_3178\_lp5.mrc** and **2xub\_model\_rigidfit.pdb** in a new chimera window. Set contour *Level* to 0.006 in the *Volume Viewer* to cover most of the non-background density. Set volume opacity to 0.5 (*Color* button).

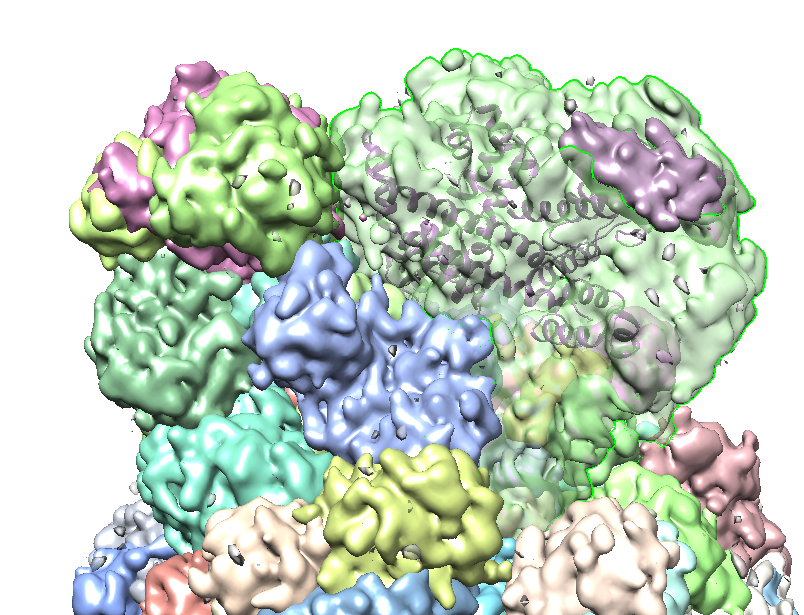
The Segger tool in Chimera (*Tools/Volume Data/ Segment map*) uses a watershed algorithm to segment volume starting from local maximas and iteratively filters the volume to groups neighboring maximas into larger segments. The grouping terminates when the number of segments falls below a cutoff (provided by user). Use the parameters in the figure below to run Segger. Here we stop grouping the segments if the number of regions <= 200 (and we get 146 segment regions). One may use 100 instead of 200 to get a coarser segmentation (larger segments), if the conformational changes required to fit the model is larger.



On the *Segment Map* window, click on *Shortcuts* and under *Shortcuts* click *Other tools: Fit.* A *Fit to Segments* window appears. Use the *Group regions by chains in visible models* option to group the segments that cover the starting model.



This will group the segments based on the rigidly fitted model and provide a single segment that covers the model.



Ctrl click on the grouped segment to select it and in the *Segment Map* window, use *Regions/Make transparent* to make the segment transparent. *File/Save selected regions to .mrc file* to save the segment map (**emd\_3178\_lp5\_regions\_bymodel.mrc**).

**3.2. Flexible fitting with Flex-EM**

Next, we will use Flex-EM1,2 to fit the homology model of C82 in the map of elongating PolIII (EMDB ID: 3178). The homology model was built using MODELLER3 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 the volume around C82 segmented using Segger4 tool in Chimera5 (**emd\_3178\_lp5\_regions\_bymodel.mrc**).

2. The starting homology model (**2xub\_model\_rigidfit.pdb**) was rigidly fitted in the segmented density using Chimera (*fitmap*).

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.

Rigid body restraints are 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 RIBFIND6 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.

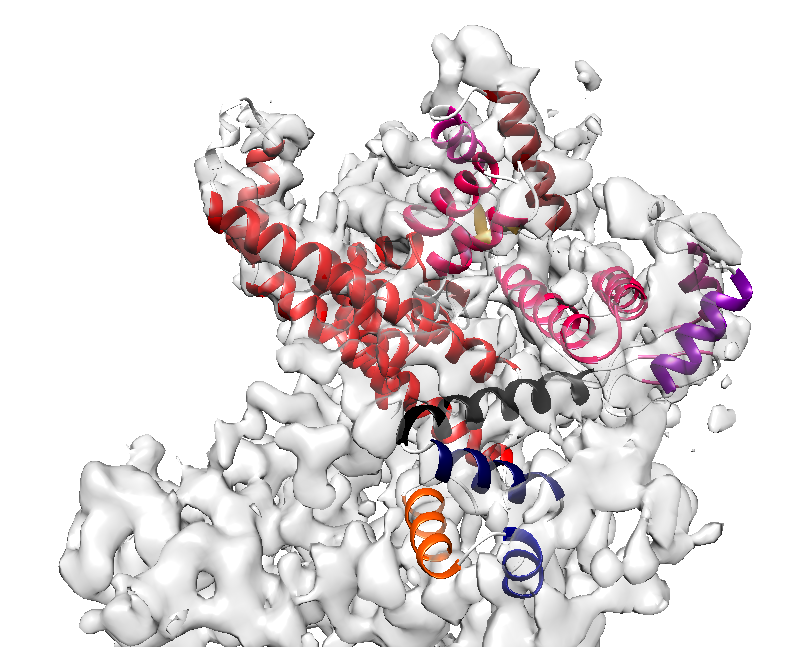
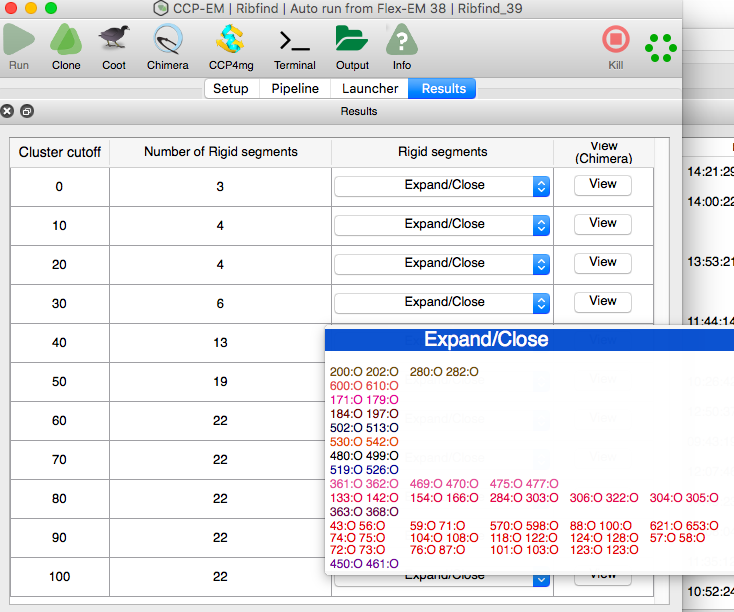
**We will be using CCP-EM GUI interface to run RIBFIND and Flex-EM.**

A. Running RIBFIND

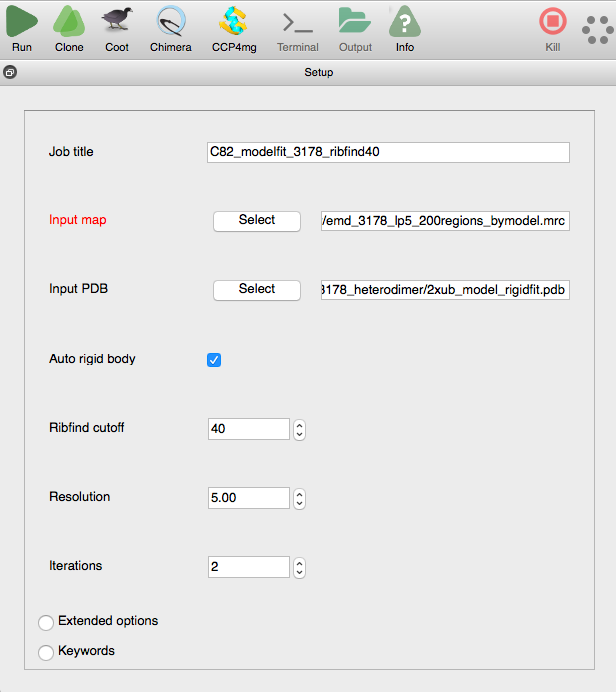
1. Launch the CCP-EM GUI

2. Use the starting model (2xub\_model\_rigidfit.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 these sub-domains together).

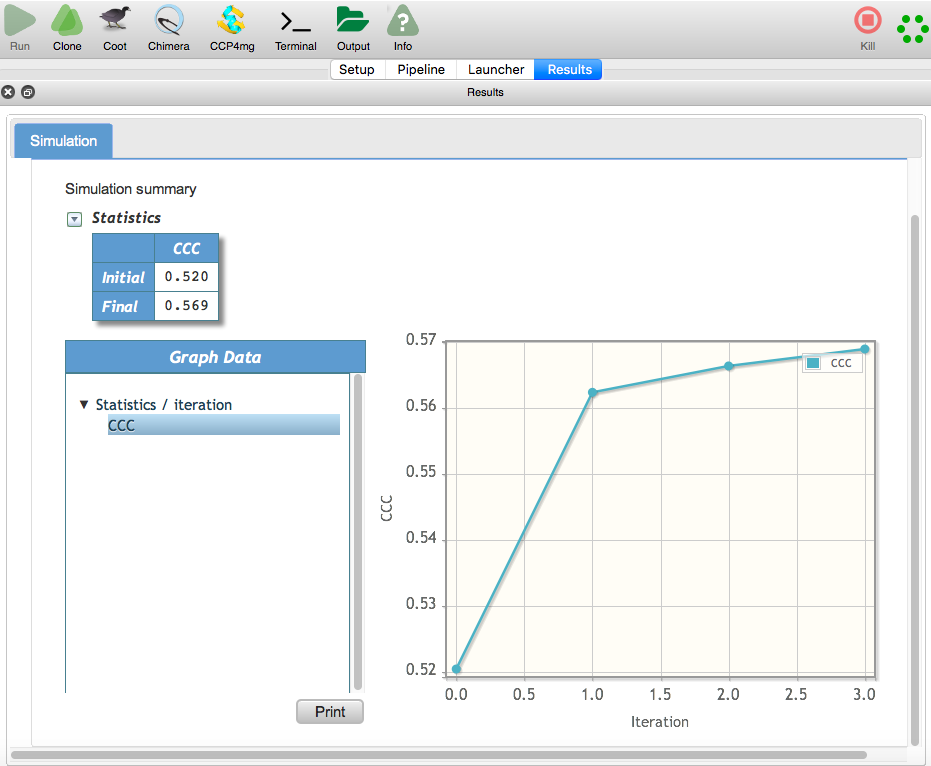


We can use the rigid body file corresponding to the cluster cut-off of 40 for the Flex-EM run OR RIBFIND can be run as part of FlexEM with 40% cutoff. Open Flex-EM task in the CCP-EM gui and fill in the input parameters.

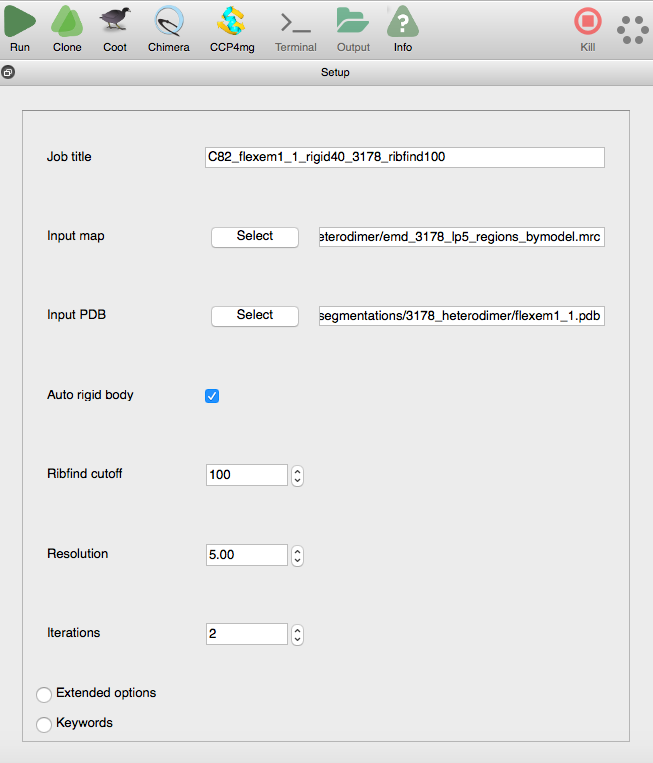


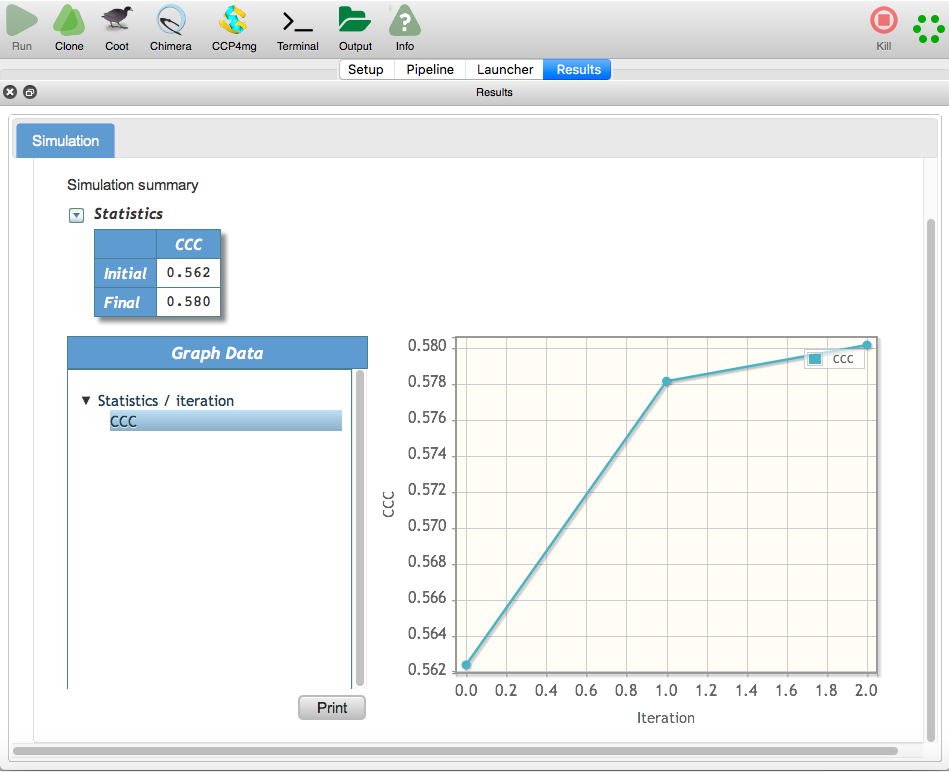
‘Run’ the Flex-EM job for 2 iterations to start the flexible fitting process. Each iteration will take about ~20 mins in this case.

Once the Flex-EM run is finished, the cross-correlation scores for models from each iteration is plotted in ‘Results’ tab. The cross-correlation with density appears to settle between 0.56 and 0.57, after the first iteration. The results of 3 iterations are shown here, just to demonstrate.



To further fit the model in the density, the secondary structures can be treated as rigid bodies in the next Flex-EM run (cluster cutoff 100). We can use the output from the first iteration of the above run (flexem1\_1.pdb in the Flex-EM data folder) as input for the next Flex-EM run.





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

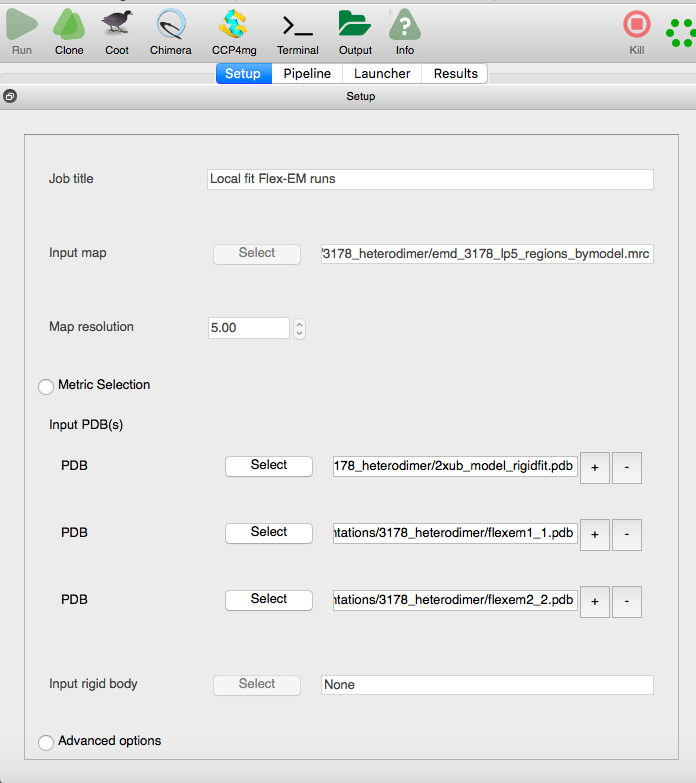
Open the fitted model from the second iteration (md1\_2.pdb or final1\_mdcg.pdb) in Chimera from the launcher along with the map to view the model fit in density. This model has been copied to the Flex-EM data folder as flexem2\_2.pdb. Open the starting model (2xub\_model\_rigidfit.pdb) as well to compare.

1. **Assessment of improvement in local density fit**

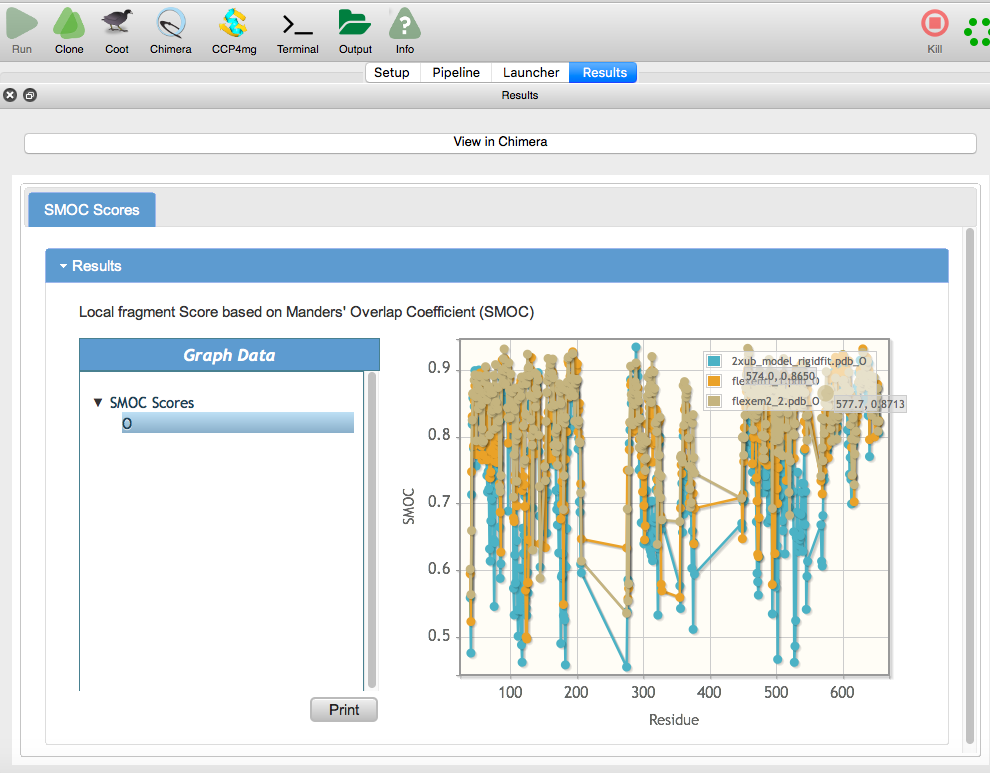
To compare the fitted models and evaluate the improvement in local density fit, we will use TEMPy SMOC (segment-based Manders’ overlap coefficient) score2,7. A local overlap coefficient is calculated over voxels covered by each residue.

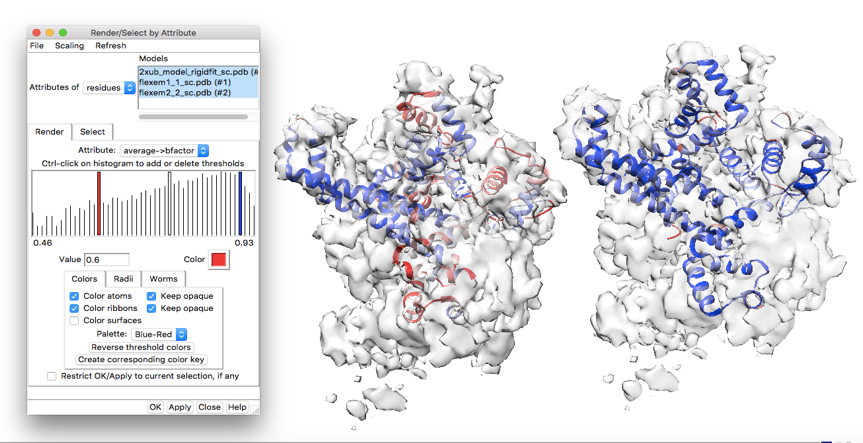
1. Select TEMPy Local scores task in the CCP-EM GUI interface

2. Input the map. Input pdb files to be scored : initial model 2xub\_model\_rigidfit.pdb, flexem1\_1.pdb and flexem2\_2.pdb



Click Run and wait for the Results to be displayed.



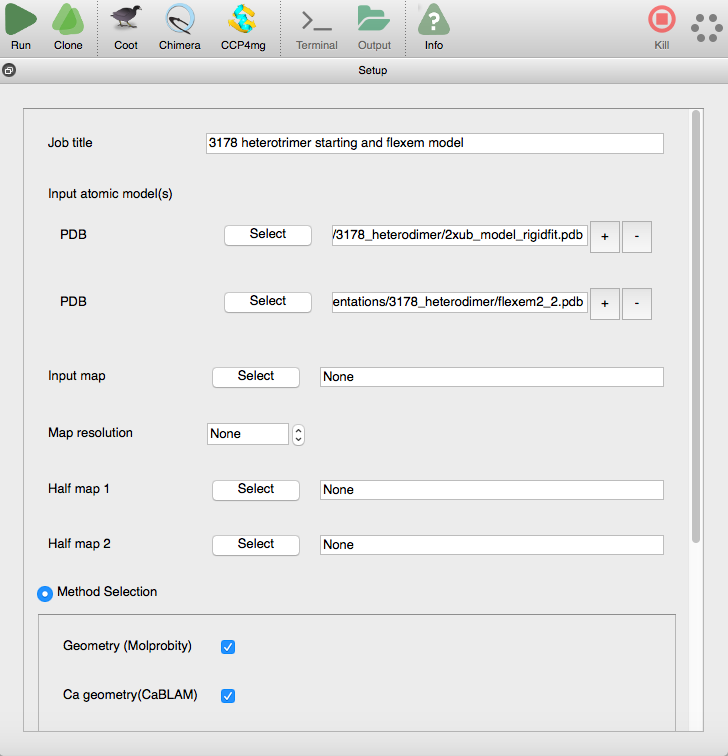


We can see that the overall local fit to density improves along the chain. When you click ‘View in Chimera’ the models open colored based on the SMOC scores. The range of values for coloring can be adjusted in the Structure Analysis window (Tools->Render by Attribute->Structure Analysis, choose residues/average bfactor)

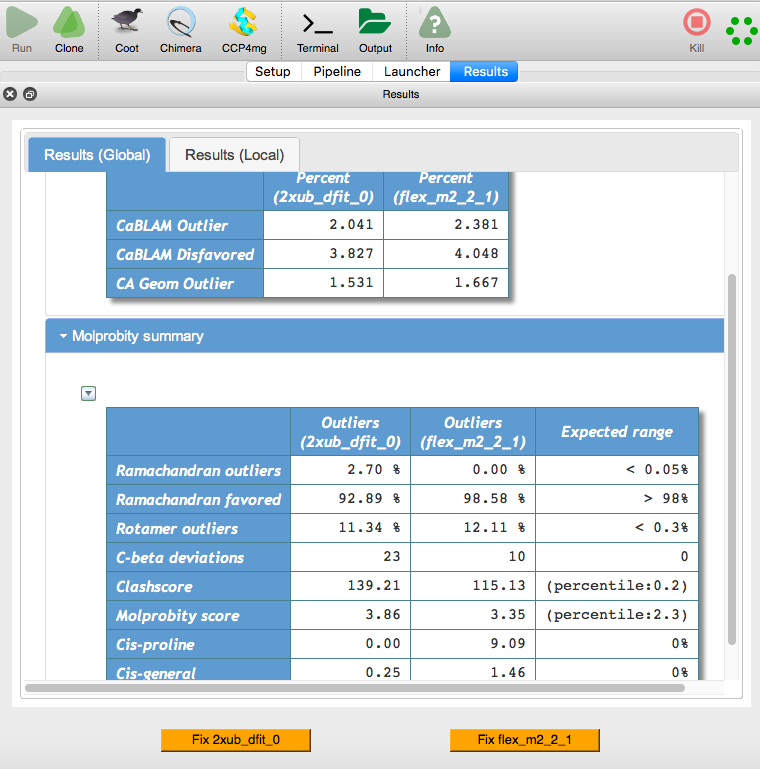
There are a few segments in the chain that are low scoring and require further inspection. Any low scoring areas can be further fixed interactively in Coot8, followed by a round of Refmac9 refinement to further improve the fit. Note that the residue at the chain terminus of the chain and chain breaks have relatively low scores. The geometry

1. **Validating model geometry**

The *Validation:Model* task in CCP-EM interface enables use of model quality evaluators like Molprobity (model geometry) and CaBLAM (backbone quality).



On the *Results (global)* tab, the summary of outlier statistics from Molprobity and CaBLAM appear.



Clearly, the model quality can be improved further by fixing clashes, Rotamer and Ramachandran outliers. Any severe issues can be fixed in Coot and the model can be refined further using Refmac wherein the quality usually improves as well.

**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).