



MRC Laboratory  
of Molecular  
Biology

# EMDA tutorial

Rangana Warshamanage  
Software Fayre

XXV IUCr congress  
15.08.2021

# EMDA

- I/O and format conversion
- Likelihood-based methods (overlay, magnification, bestmap etc...)
- Validation metrics (RCC, FCC, FSC, OCC)
- General purpose tools

# Obtaining EMDA

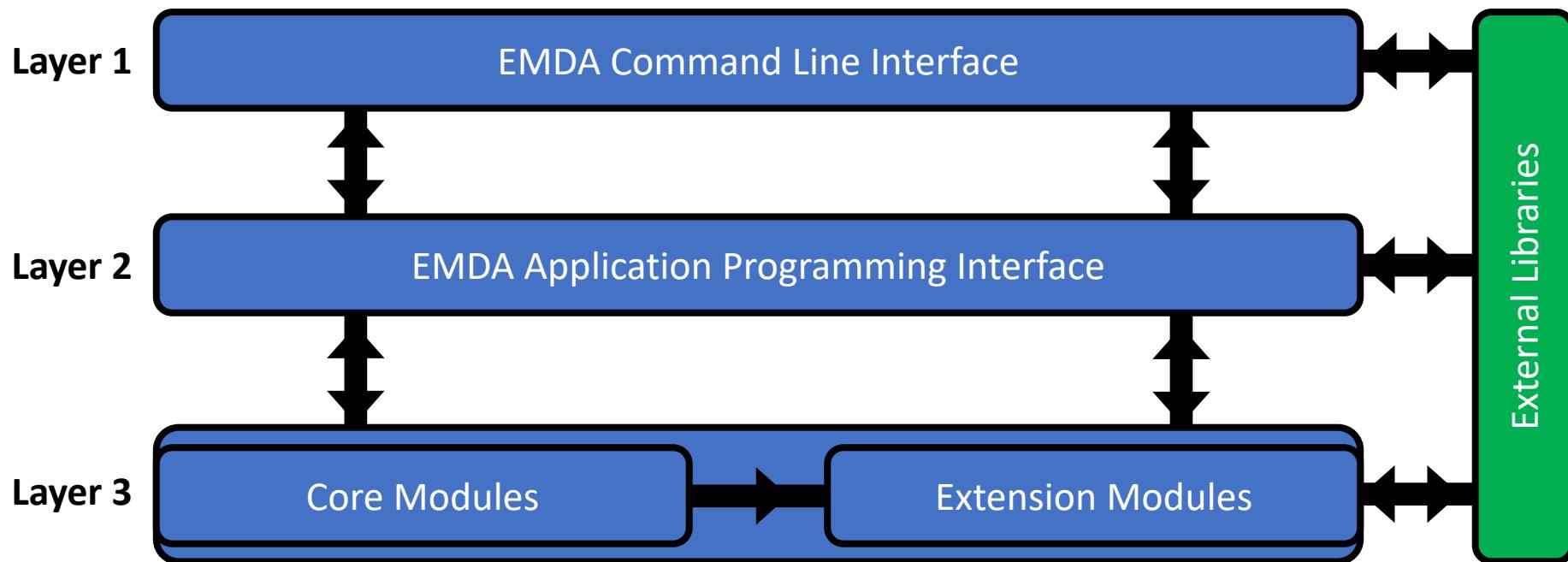
- Open source (MPL-2.0 license)
- Included in Python Package Index (PyPi)  
`pip install emda`
- Part of CCP-EM suit

# Dependencies

- Standard python packages – Numpy, Scipy, Pandas, Matplotlib
- Other python packages – mrcfile, gemmi, servalcat, proshade
- CCP4(REFMAC)

# EMDA

## Electron Microscopy Data Analytical toolkit



# Command-Line Interface

## Example:

```
>>> emda mapmask --map map1.mrc --knl 5 --prb 0.99 --itr 3 --res 15
```

# Application Programming Interface

# Part 1: Local correlation calculation

# Example 1: Detecting model errors

The Goal is to use local correlation in real space to detect errors in the model.  
In this example we use EMD-5623 entry.

Step 1: Download maps and model

Note that you won't be able to find half data for this entry in EMDB, but you will find them @  
<https://www.emdataresource.org/EMD-5623>

Make sure you have EMD-5623-half-1.map.gz, EMD-5623-half-2.map.gz, emd\_5623.map.gz and 3j9i.cif  
in your working directory. Now unzip all maps.

3j9i.cif

EMD-5623-half-1.map

EMD-5623-half-2.map

emd\_5623.map

## Step 2: refine the model

Since we're calculating local correlation between the map and the model, we need to refine the model w.r.t the map.

Note that we use the fullmap (average of half maps) as the map in the refinement because we want to get the correct distribution of atomic B values. If we use the primary map as the refinement target, we cannot be sure that we get the correct distribution because the sharpening affects the B values. On the other hand, since the fullmap is blurred, the whole distribution is shifted, but the shape of the distribution is not affected.

We can use servalcat to do the refinement.

```
$ servalcat refine_spa --model ../3j9i.cif --resolution 3.3 --halfmaps  
..../EMD-5623-half-1.map ..../EMD-5623-half-2.map --ncycle 10
```

Make sure you have refined.pdb in your working directory. You may look at the refined.pdb in COOT.

Also, you may open primary map, halfmaps and refined model in COOT or Chimera, just to make sure that all maps and model are properly aligned.

Note that EMDA does not optimized map-model fit during local correlation calculation. It assumes that all maps and models are properly aligned beforehand.

You may use other ways to refine the model w.r.t the fullmap.

### Step 3: Calculating local correlation

Now, we're ready to calculate local correlation in EMDA.

To do that, copy (or make a symbolic link) the refined.pdb into the directory where you have halfmaps.

```
$ emda rcc --h1 EMD-5623-half-1.map --h2 EMD-5623-half-2.map --mdl refined.pdb  
--res 3.3 -nrm --knl 2
```

This command does several things.

1. First, it computes real space local correlation map using normalized and weighted half maps (normalization and weighting are done in Fourier space).
2. Also, it outputs the normalized and weighted fullmap that will be used in the subsequent map-model local correlation calculation.
3. Computes the map from the model (refined.pdb) using REFMAC to the given resolution.
4. Apply weights and perform normalization on the calculated map
5. Finally, computes real space local correlation map between the normalized and weighted fullmap and the calculated map.

Output of this operation

**Normalized and weighted maps:**

- bin\_normalized\_fullmap.mrc, bin\_normalized\_halfmap1.mrc, bin\_normalized\_halfmap1.mrc, bin\_normalized\_modelmap.mrc

**Local correlation maps:**

- rcc\_halfmap\_smax2.mrc, rcc\_fullmap\_smax2.mrc, rcc\_fullmap\_star\_smax2.mrc, rcc\_mapmodel\_smax2.mrc

## Step 4: Analyzing correlation maps

A convenient way to look at a correlation map is to use it as a map to colour another map.

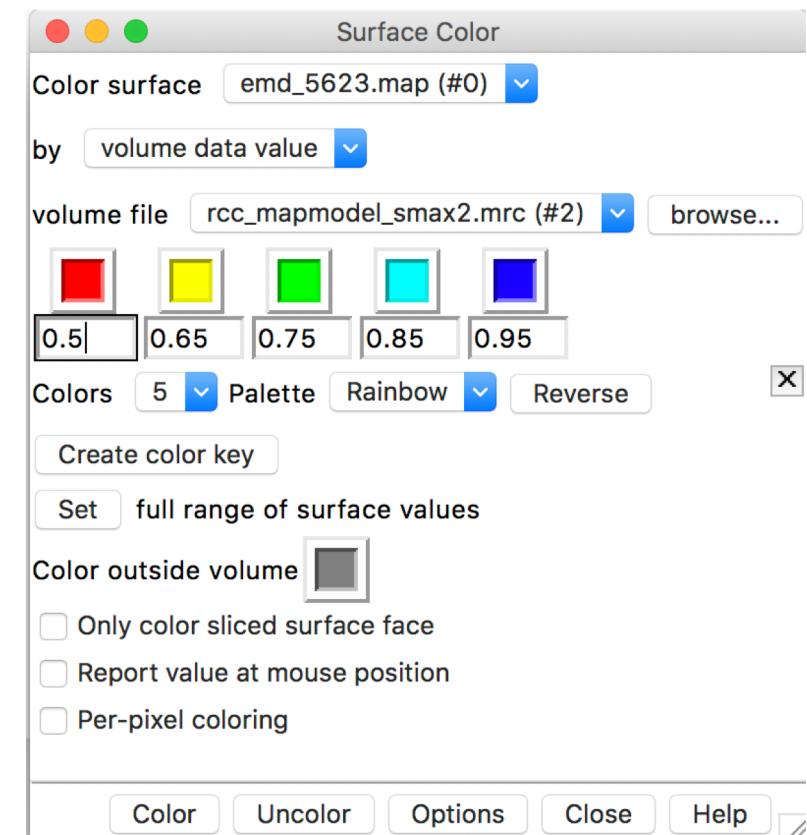
In this case, we use rcc\_fullmap\_star\_smax2.mrc and rcc\_mapmodel\_smax2.mrc to colour the primary map.

rcc\_fullmap\_star\_smax2.mrc is the map with  $\sqrt{CC_{full}}$  correlation

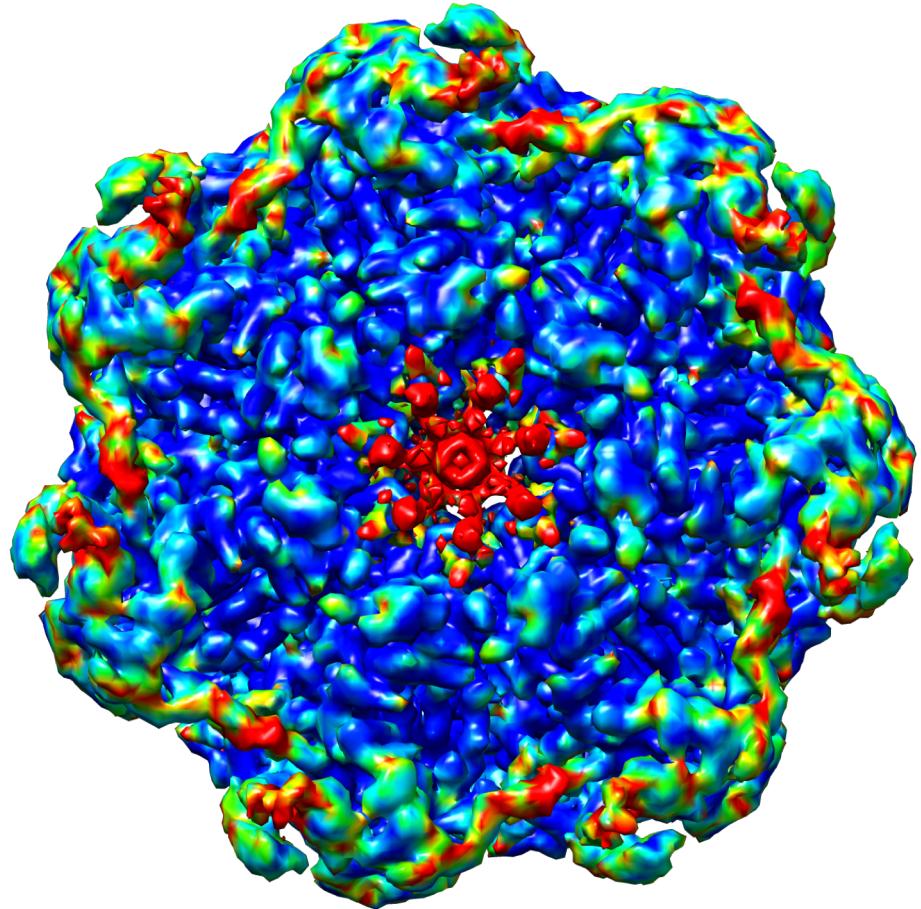
rcc\_mapmodel\_smax2.mrc is the local correlation map between fullmap and the model-based map (using normalized and weighted densities)

To colour the primary map by correlation maps,

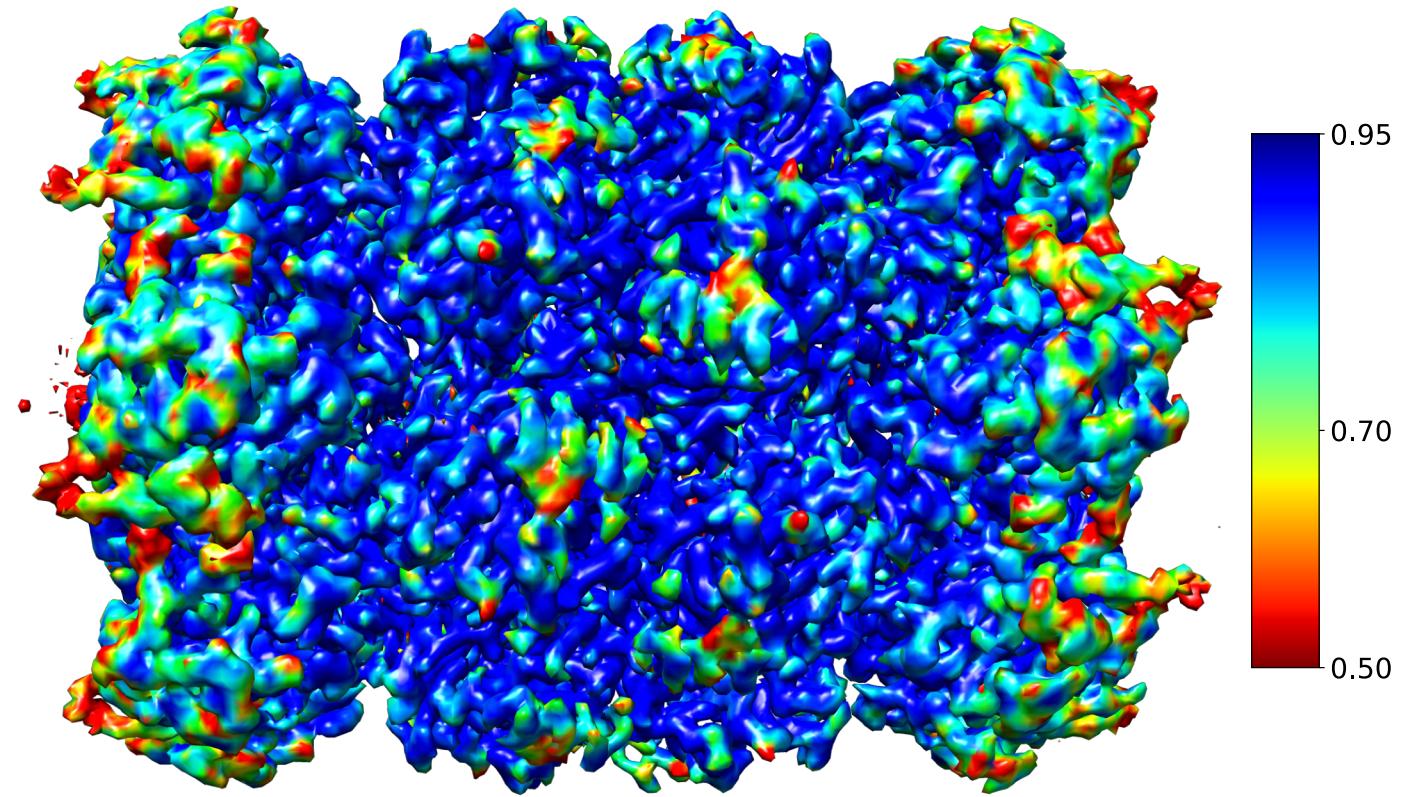
1. open primary map (emd\_5623.map),  
rcc\_fullmap\_star\_smax2.mrc and  
rcc\_mapmodel\_smax2.mrc in Chimera (you may use ChimeraX, as well)
2. In Volume Viewer – Tools > Surface Color  
Color surface – emd\_5623.map  
By – volume data value  
Volume file – rcc\_mapmodel\_smax2.mrc
3. Then go to Options and click on Set button.  
That will automatically choose values for colour contours.  
But we would like to set those values by hand.  
So, give following values (0.5, 0.65, 0.75, 0.85, 0.95),  
and press Color.



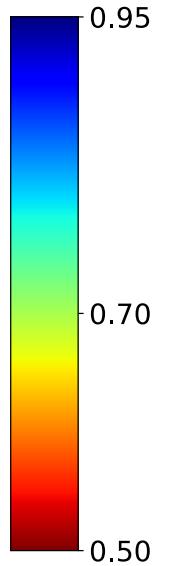
EMD-5623 primary map coloured by map-model local correlation



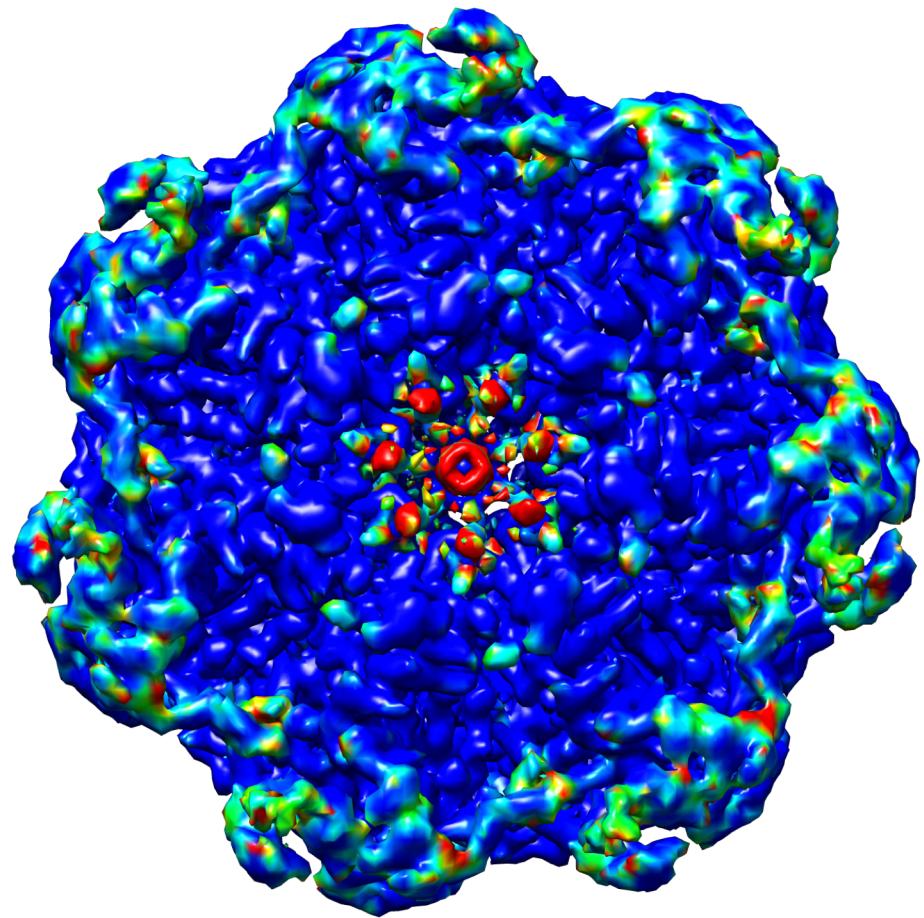
Top view



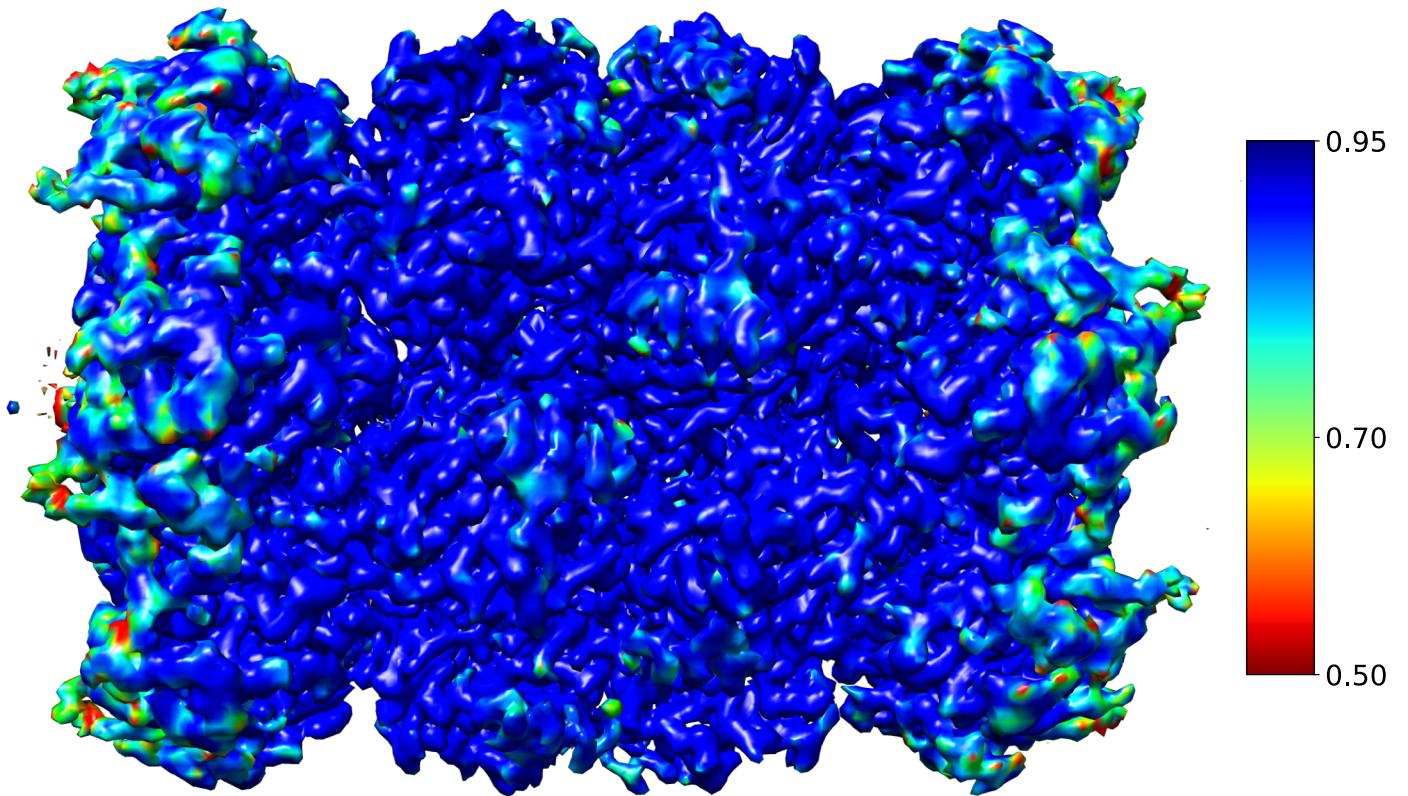
Side view



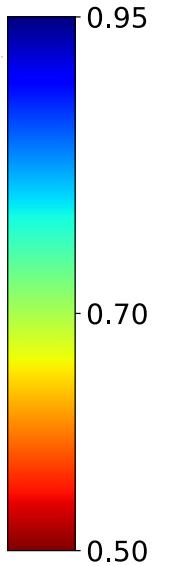
EMD-5623 primary map coloured by  $\sqrt{CC_{full}}$  local correlation



Top view

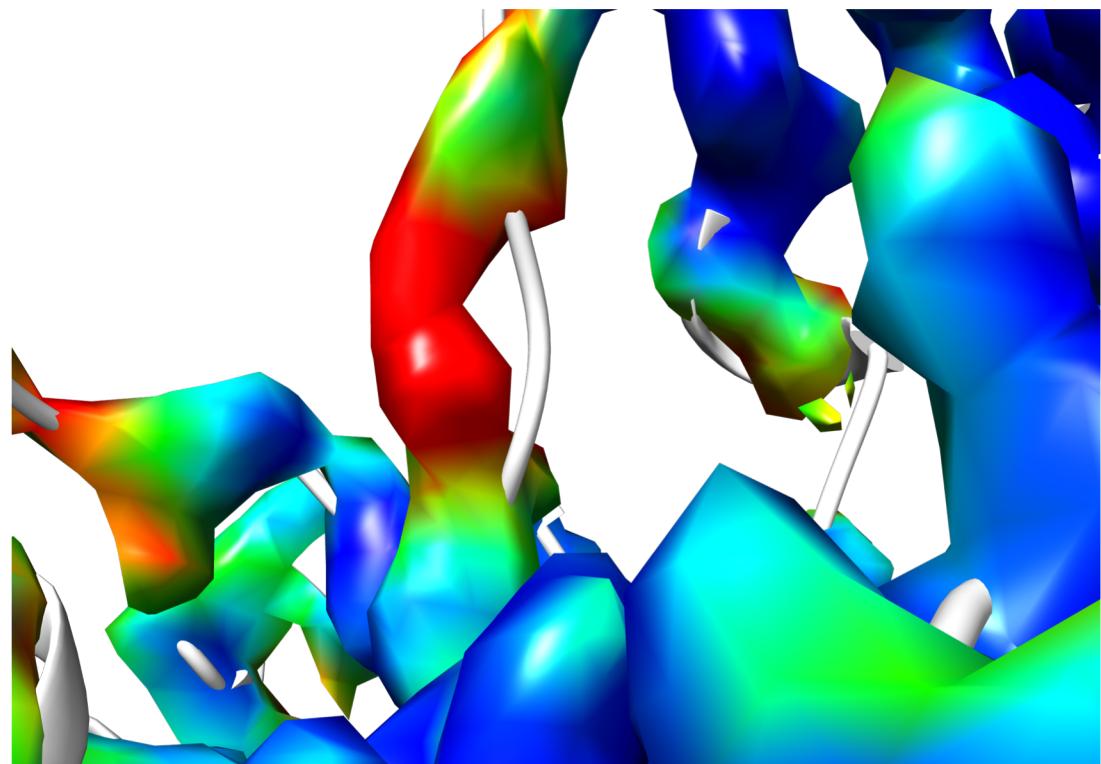


Side view

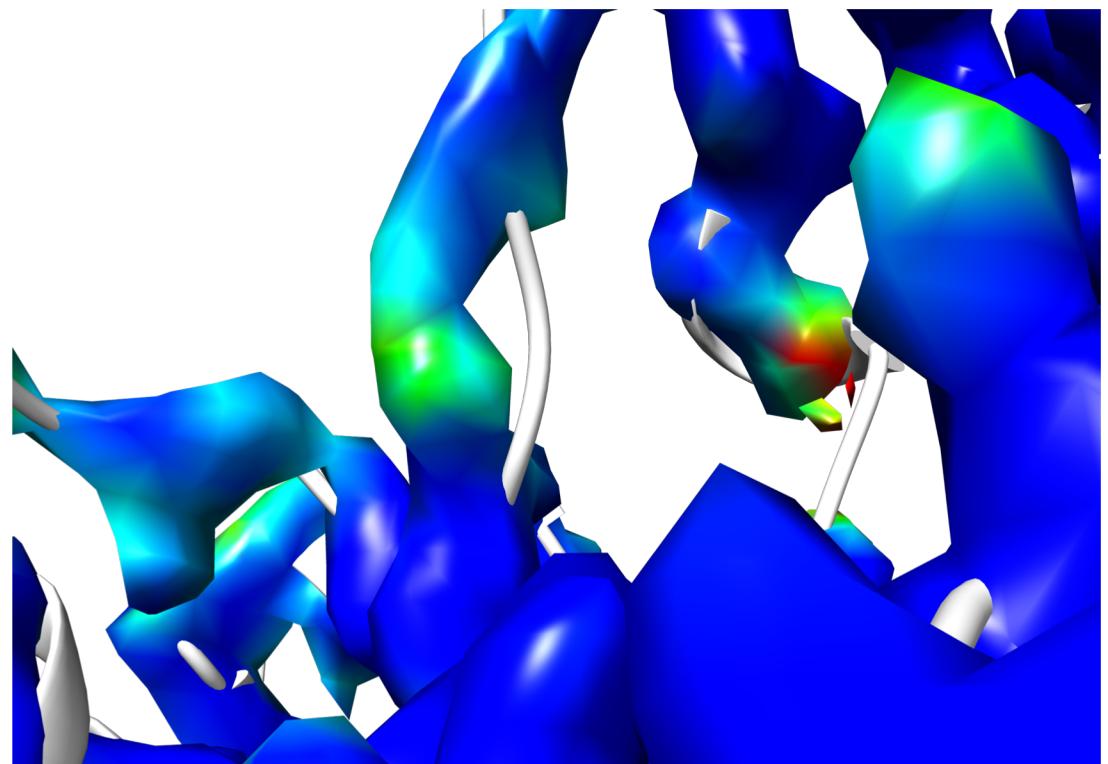


## Some regions of interest (compared with model)

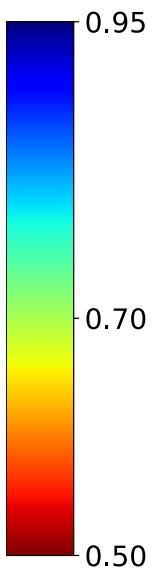
U/52-54



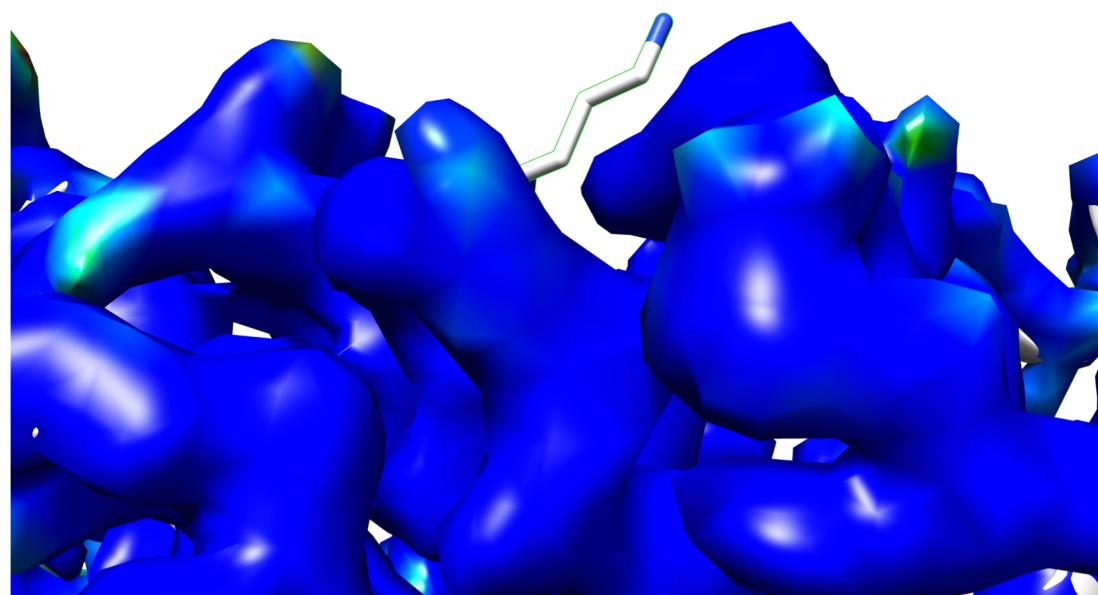
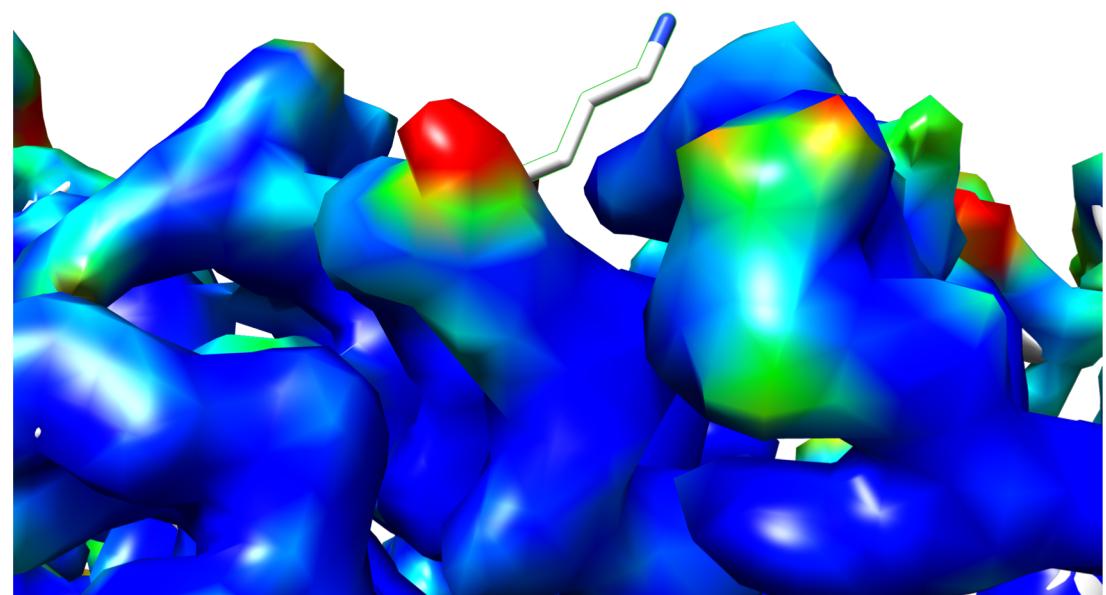
Map-model CC



$$\sqrt{CC_{full}}$$



Chain N/LYS



Map-model CC

$$\sqrt{CC_{full}}$$

# Example 2: Detecting unmodeled density

In this example, we use real space local correlation to detect unmodeled density in EMD-11203 map.

Step 1: Download data

```
$ emda fetch --emd 11203 --all
Successfully created the directory /Users/ranganaw/MRC/REFMAC/tutorial_data/emda_rcc/EMD-
11203/
claimed resol 2.6
Mask Id: emd_11203_msk_1.map
Half1id: emd_11203_half_map_2.map.gz
Half2id: emd_11203_half_map_1.map.gz
Fetched 11203
```

Step 2:

Refine the model 6zge that you just downloaded against the fullmap using servalcat (as you did in example 1)

Step 3: Calculate real space local correlation maps using EMD-11203 halfmaps and the refined.pdb

```
$ emda rcc --h1 emd_11203_half1.map --h2 emd_11203_half2.map --mdl refined.pdb --res 2.6 --knl 2 --nrm
```

Note: This calculation may take some time. Be patient ☺

Output of this operation

**Normalized and weighted maps:**

- bin\_normalized\_fullmap.mrc, bin\_normalized\_halfmap1.mrc, bin\_normalized\_halfmap1.mrc, bin\_normalized\_modelmap.mrc

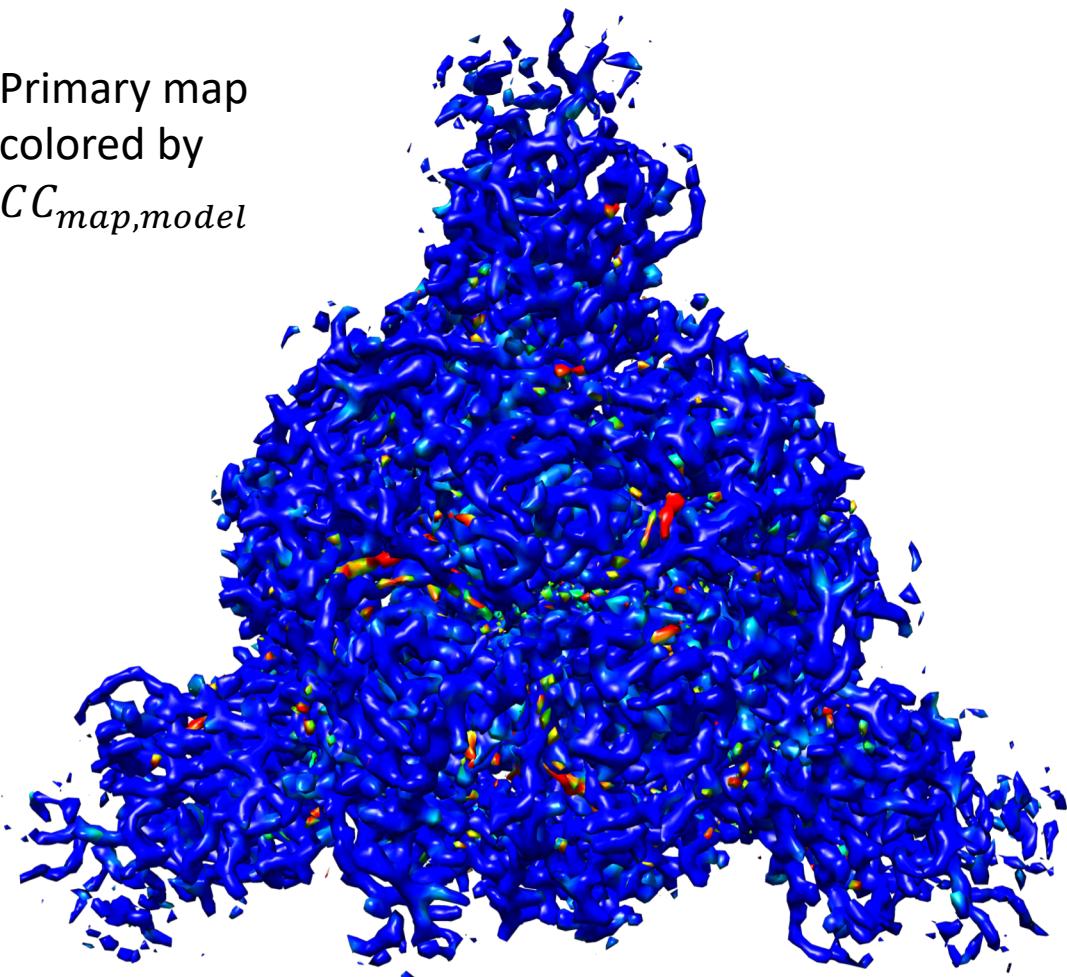
**Local correlation maps:**

- rcc\_halfmap\_smax2.mrc, rcc\_fullmap\_smax2.mrc, rcc\_fullmap\_star\_smax2.mrc, rcc\_mapmodel\_smax2.mrc

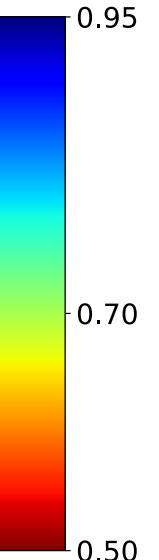
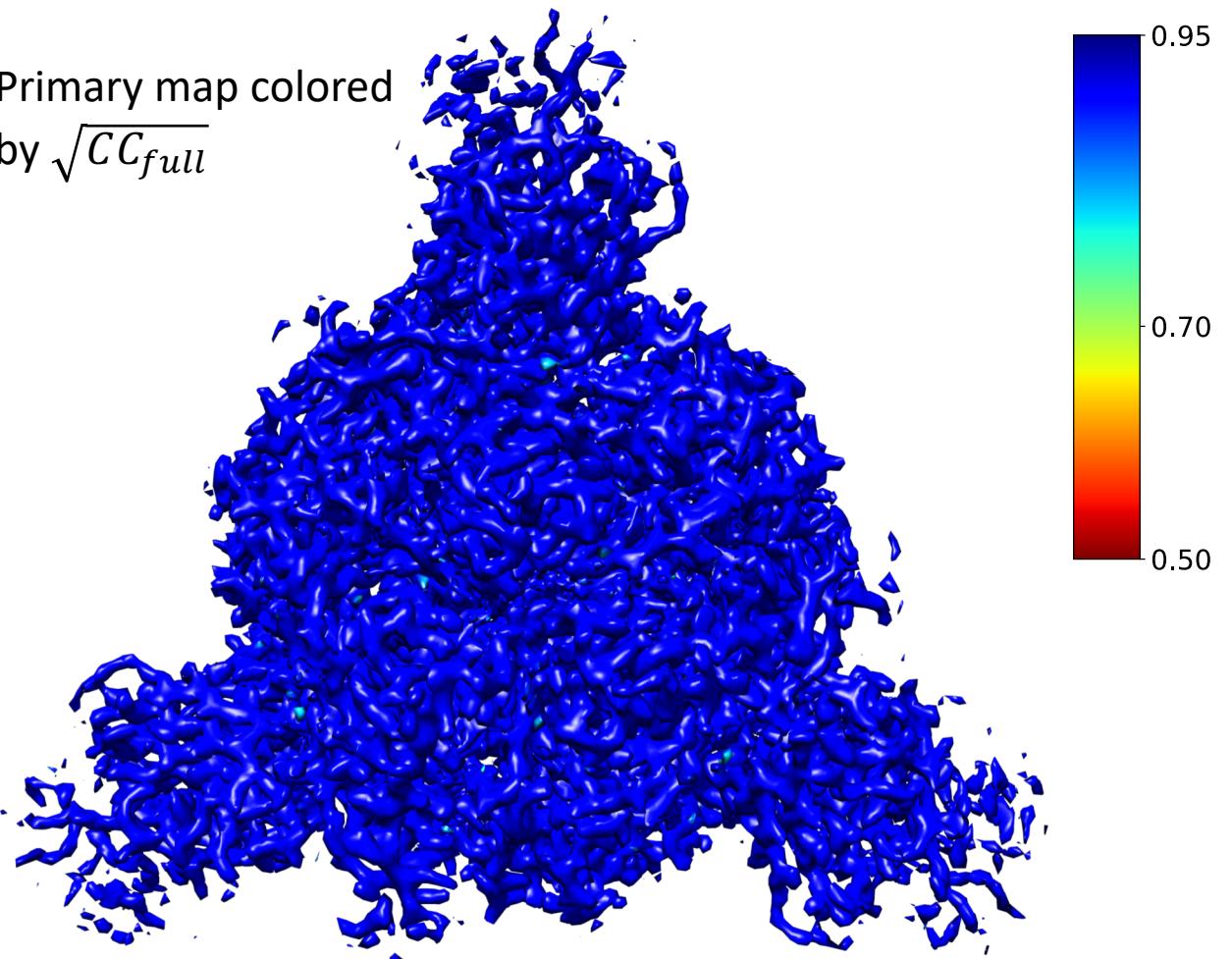
## Step 4: Analyzing correlation maps.

As before, let's colour the primary map by `rcc_fullmap_star_smax2.mrc` and `rcc_mapmodel_smax2.mrc`. Use the same procedure that explained in example 1.

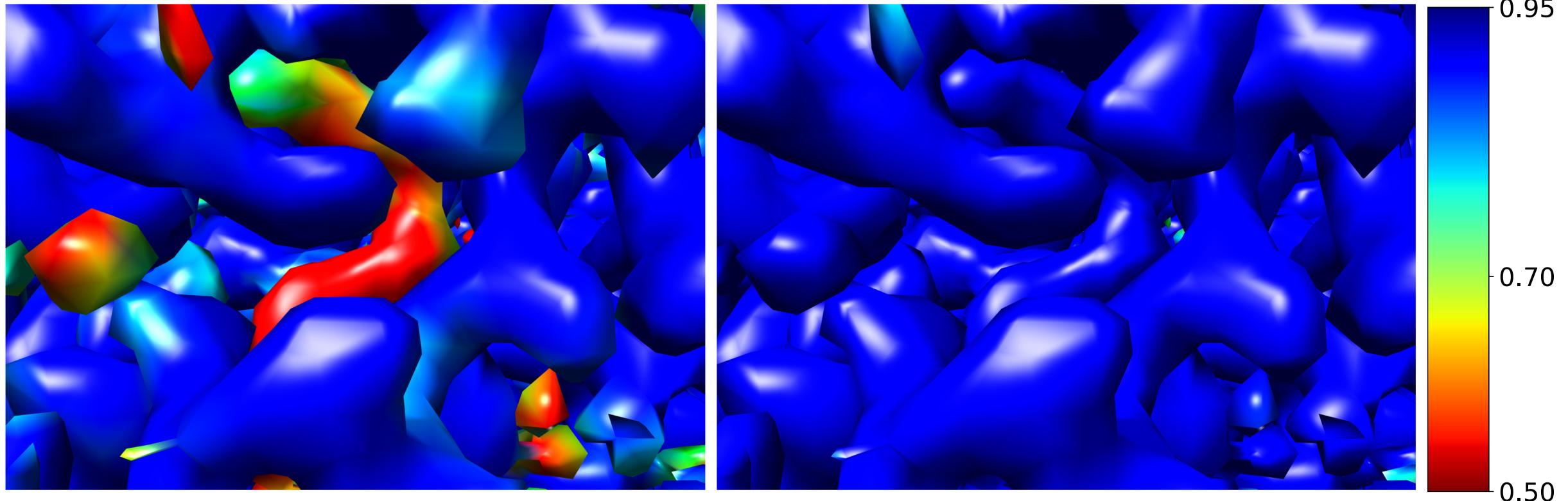
Primary map  
colored by  
 $CC_{map,model}$



Primary map colored  
by  $\sqrt{CC_{full}}$



Magnified view of an unmodeled density



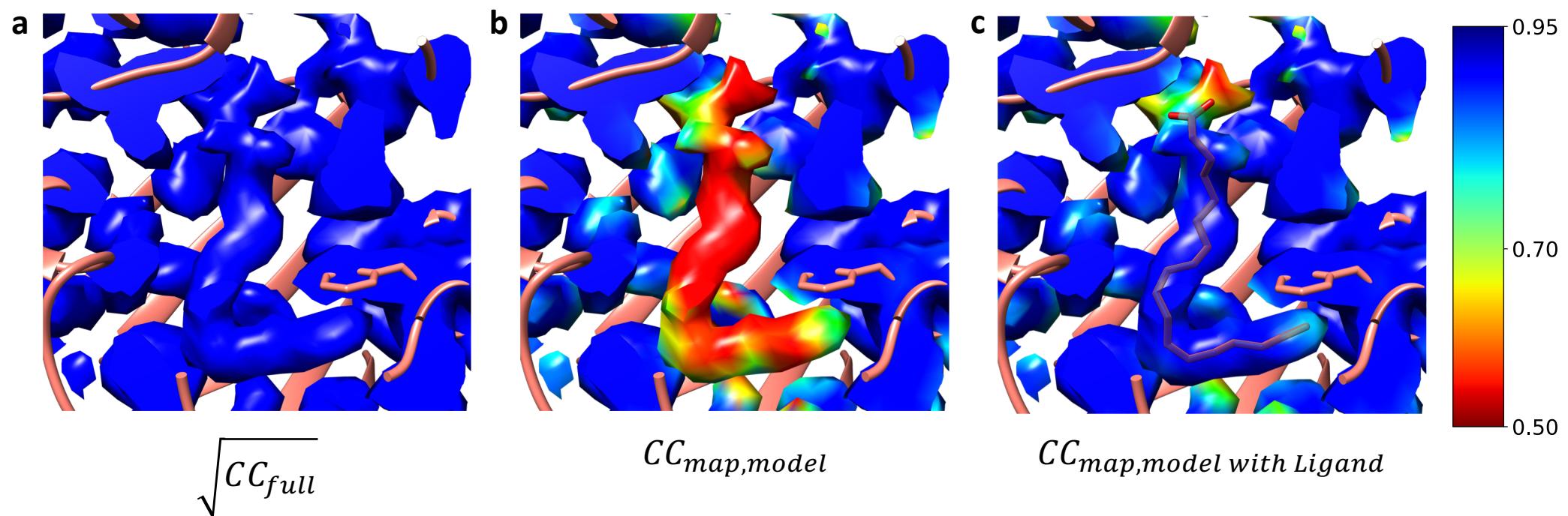
Primary map colored by  $CC_{map,model}$

Primary map colored by  $\sqrt{CC_{full}}$

No correspondence in the model for this density

Step 5: Use homology model 6z5b to account for this density as Linoleic acid.

1. use Chimera to fit the homology model to EMD-11203 map
2. refine coordinates and B values using servalcat (REFMAC)
3. Recalculate map-model local correlation using newly refined model.



**a, b** - unmodelled density in the primary map colored by  $\sqrt{CC_{full}}$  and  $CC_{map,model}$ , respectively.  
**c** – same density coloured by the  $CC_{map,model}$  after the ligand has been fitted  
(using homology model 6z5b).

## Step 6: Atomic correlation values

Atomic correlation values can be obtained from correlation maps by interpolation.

To help you with this, I have put a script at

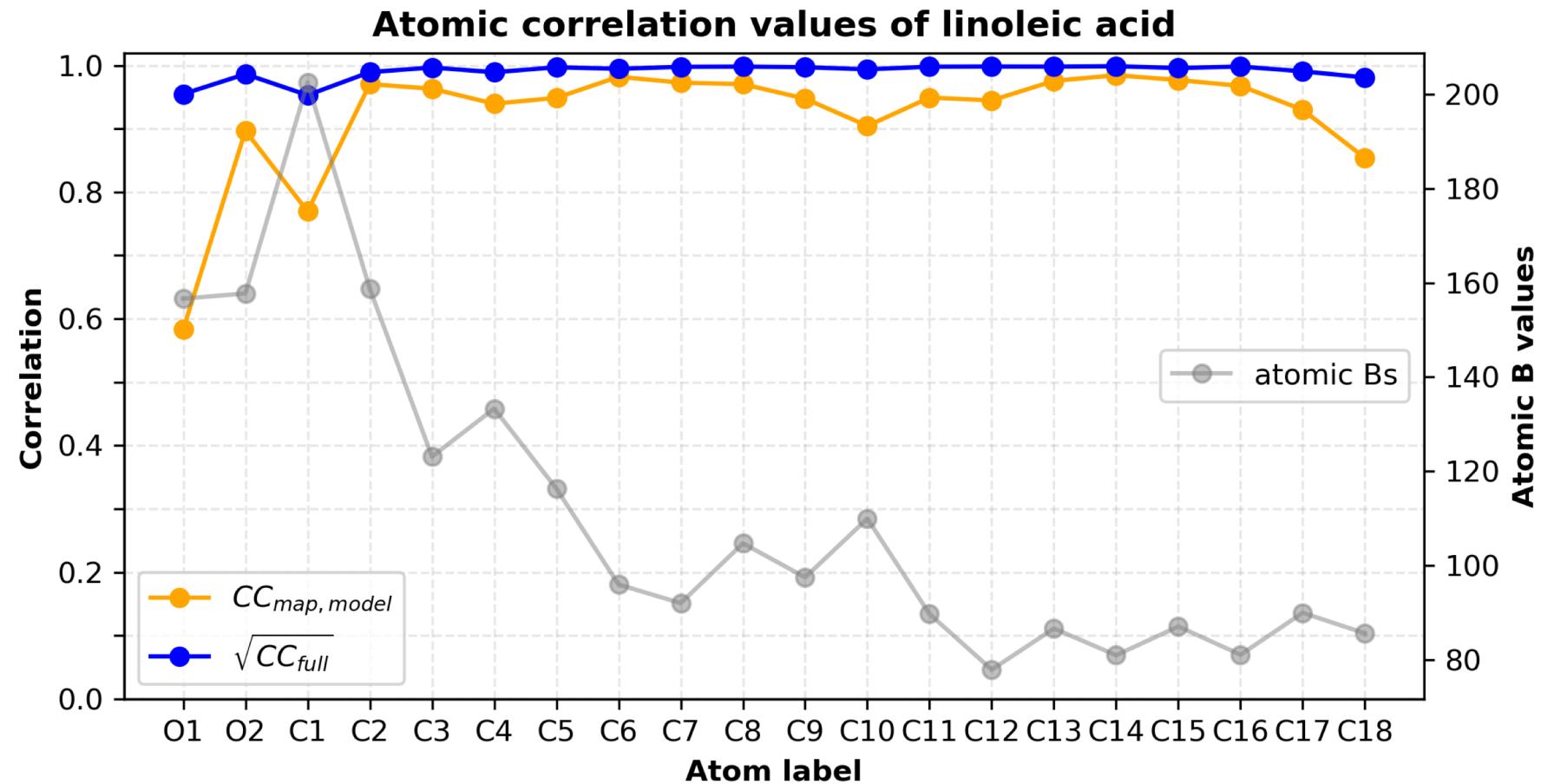
[https://gitlab.com/ccpem/emda/-/blob/master/tests/interpolate\\_at\\_atompos.py](https://gitlab.com/ccpem/emda/-/blob/master/tests/interpolate_at_atompos.py)

Download this script to your working directory (where you have correlation maps) and run as

```
$ python interpolate_at_atompos.py rcc_fullmap_star_smax2.mrc  
rcc_mapmodel_smax2.mrc refined_withLA.pdb
```

**Note that your `rcc_mapmodel_smax2.mrc` is the correlation map you generated with the model with linoleic acid.**

# Atomic correlation values interpolated



## Part 2: Magnification refinement

Goal is to check the magnification of EMD-7770 and EMD-10574 cryo-EM maps w.r.t crystallography reference model 3dyp

### Beta-galactosidase

Reference:

3dyp (X-ray model, 1.75 Å)  
(Juers et al. 2009)

Target maps:

1. EMD-7770 (EM map, 1.9 Å)  
(Bartesargi et al. 2018)
  
2. EMD-10574 (EM map, 2.2 Å)  
(Saur et al. 2020)

## Step 1: Download data

```
$ emda fetch --emd 7770 --all
```

```
Successfully created the directory /Users/ranganaw/MRC/REFMAC/tutorial_data/EMD-7770/
claimed resol 1.9
```

```
Mask Id: emd_7770_msk_1.map
```

```
Half1id: emd_7770_half_map_1.map.gz
```

```
Half2id: emd_7770_half_map_2.map.gz
```

```
Fetched 7770
```

```
$ emda fetch --emd 10574 --all
```

```
Successfully created the directory /Users/ranganaw/MRC/REFMAC/tutorial_data/EMD-10574/
claimed resol 2.2
```

```
Mask Id: None
```

```
Half1id: emd_10574_half_map_2.map.gz
```

```
Half2id: emd_10574_half_map_1.map.gz
```

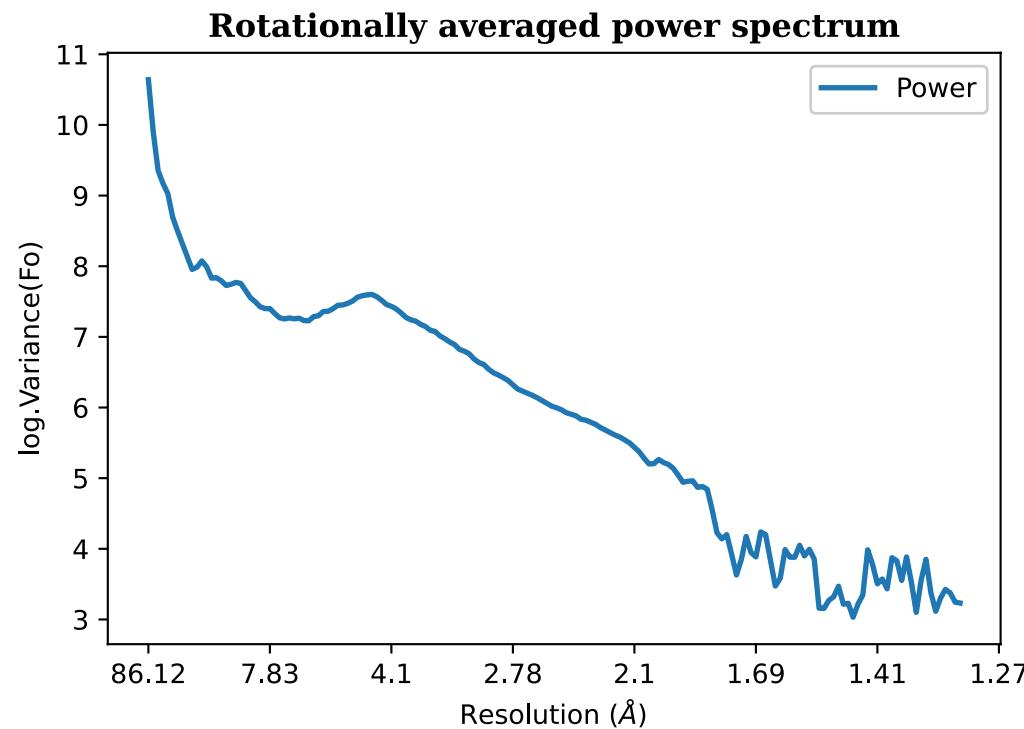
```
Fetched 10574
```

Download **3dyp** model from RCSB – <https://www.rcsb.org/structure/3DYP>

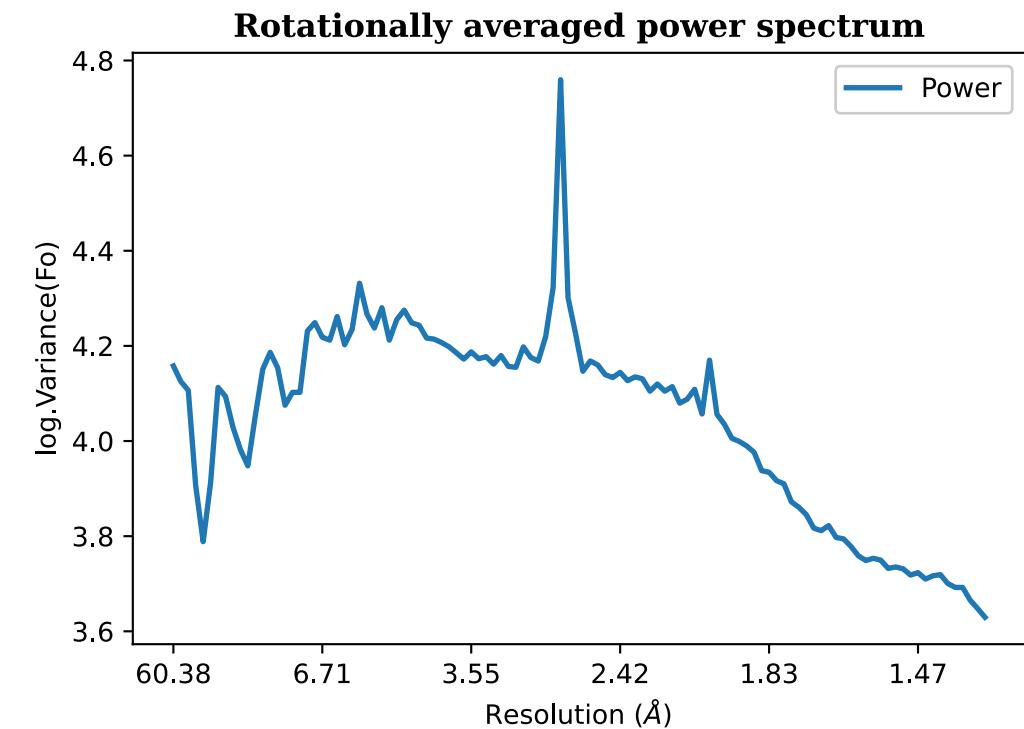
Step 2: Analyze EM maps emd\_7770.map and emd\_10574.map

Calculate their power spectra

```
$ emda power --map emd_7770.map
```



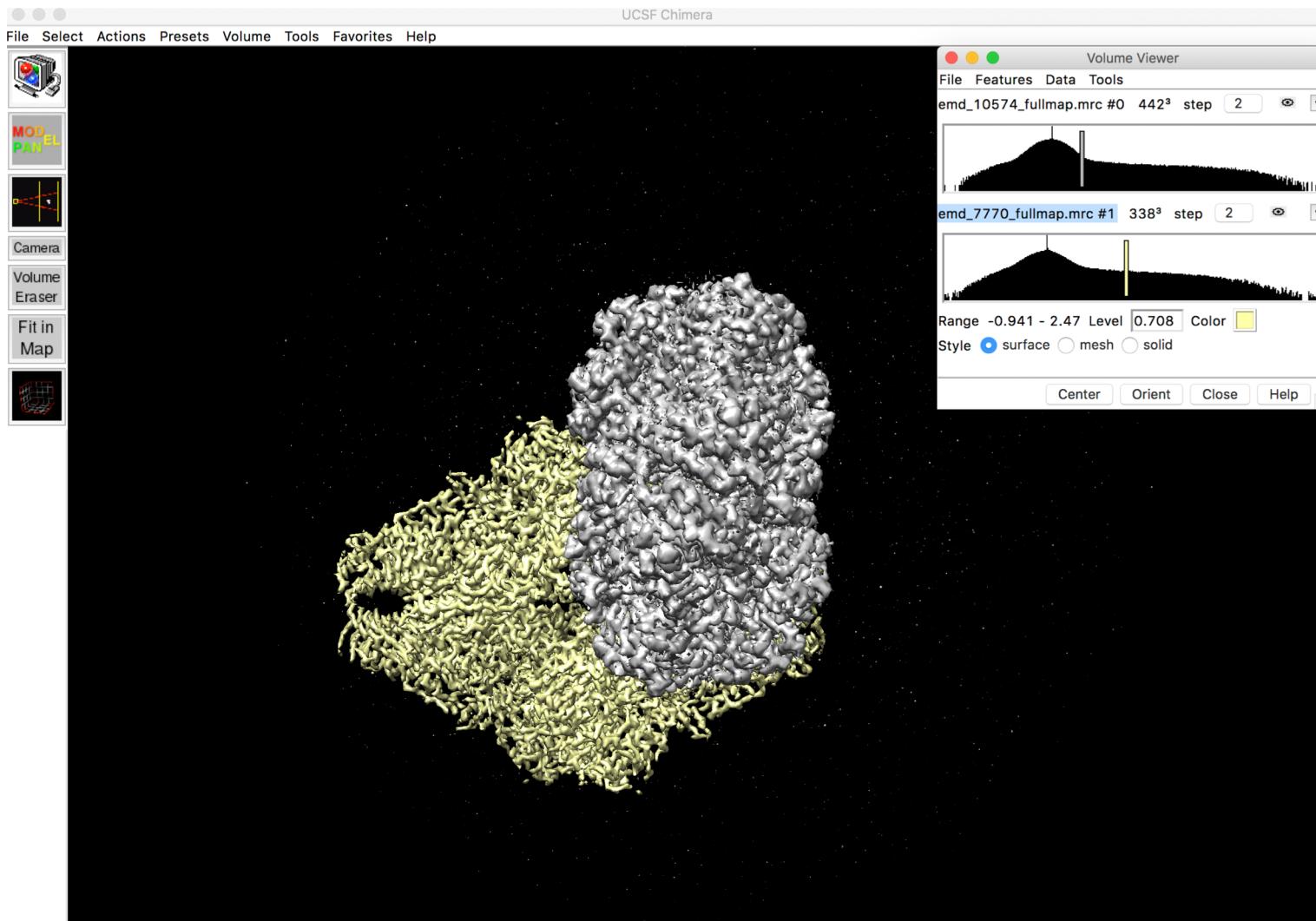
```
$ emda power --map emd_10574.map
```



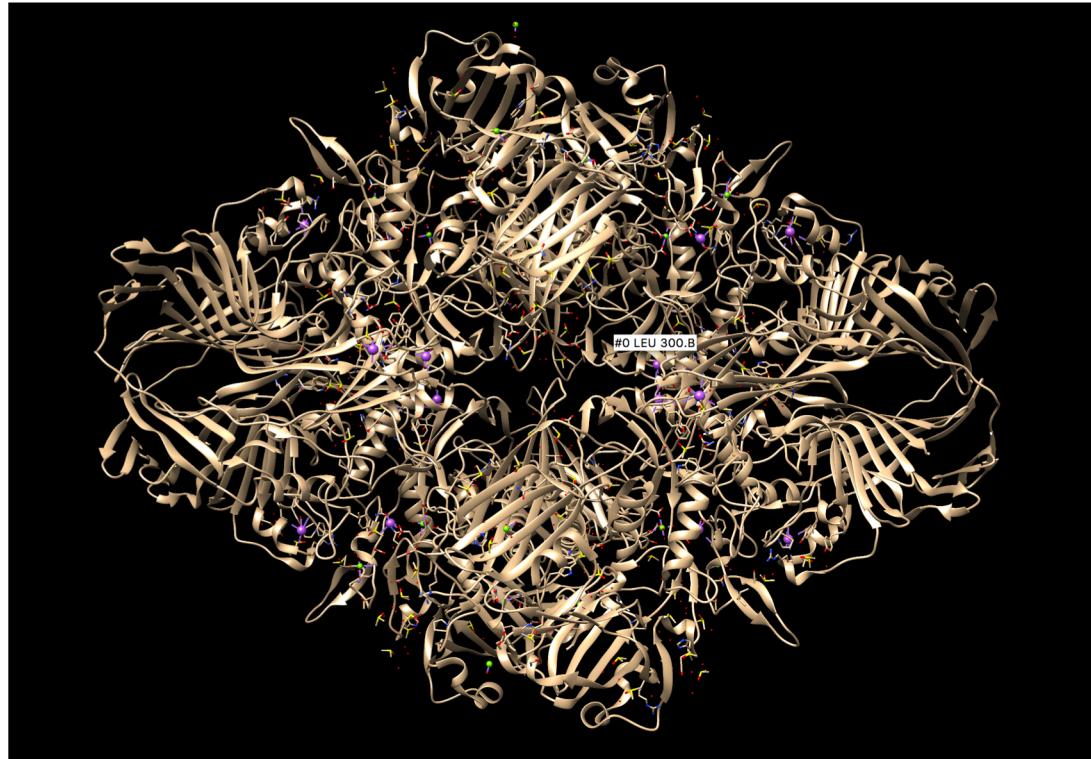
Since EMD-10574 map is filtered, let's use fullmaps for both entries.

To get the combined map (average of both half maps)

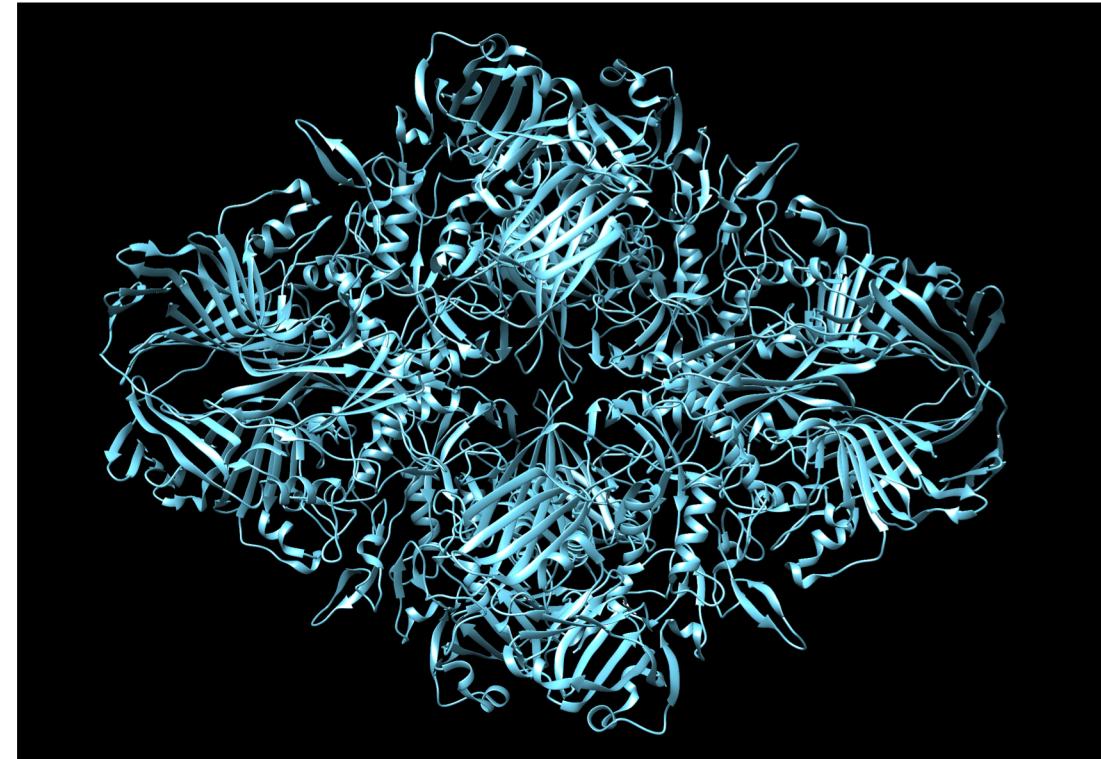
```
$ emda half2full --h1 emd_7770_half1.map --h2 emd_7770_half2.map --out emd_7770_fullmap.mrc  
$ emda half2full --h1 emd_10574_half1.map --h2 emd_10574_half2.map --out emd_10574_fullmap.mrc
```



We need just the polymer from our crystallography model

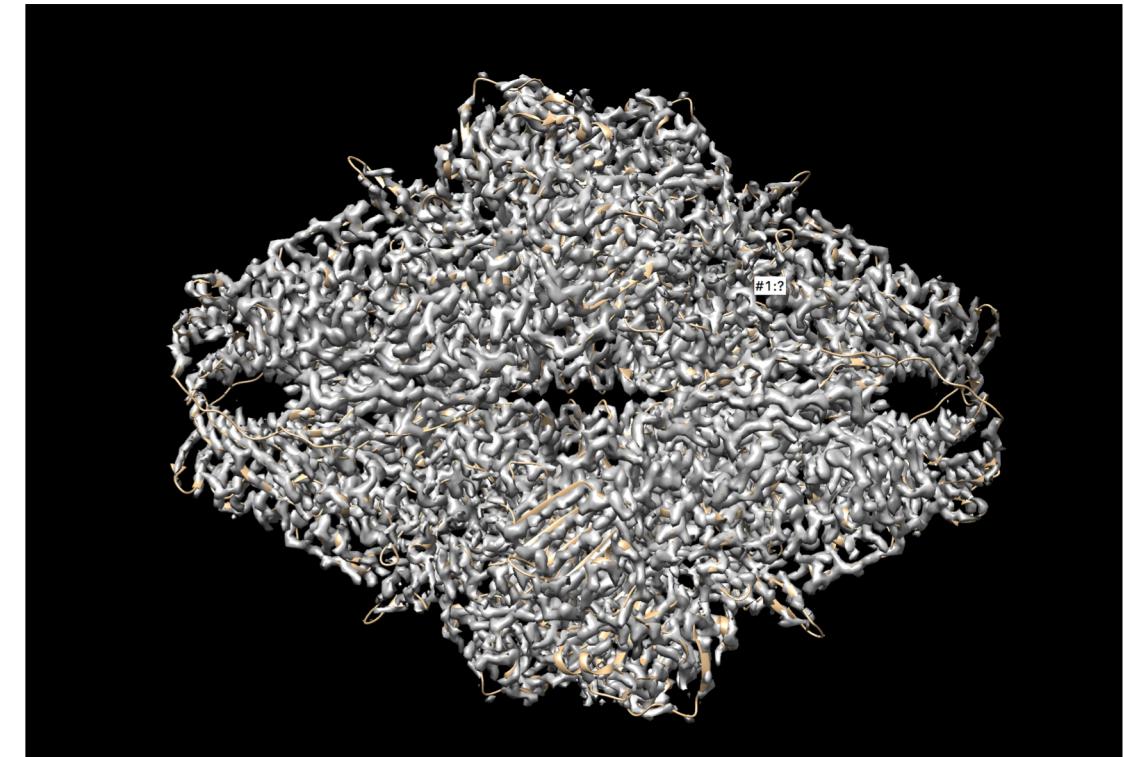
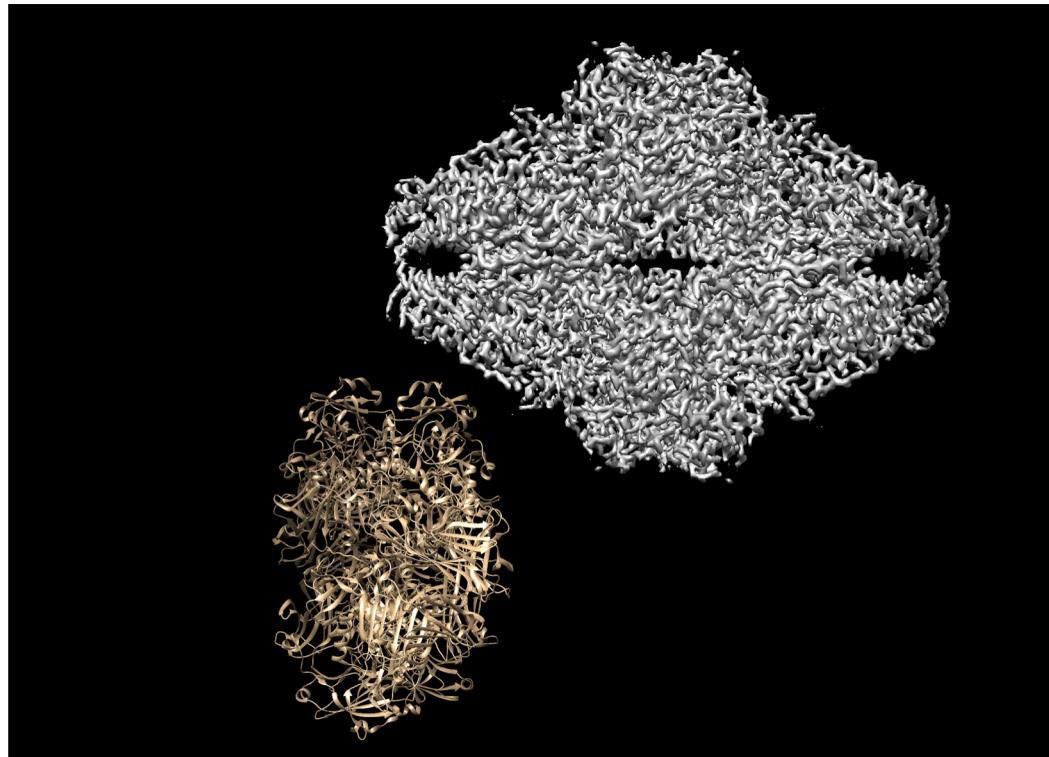


3dyp



3dyp-polymer

### Step 3: Fit crystallography model to emd\_7770\_fullmap



Save the fitted model w.r.t EM map as reference\_model.pdb

Step 3: Calculate the map from the reference model up to 1.9 Å.

Use the same sampling and cell as of emd\_7770\_fullmap.mrc

To get information about emd\_7770\_fullmap.mrc

```
$ emda info --map EMD-7770/emd_7770_fullmap.mrc
EMD-7770/emd_7770_fullmap.mrc (338, 338, 338) [215.30601501 215.30601501 215.30601501]
Unit cell: [215.30601501 215.30601501 215.30601501 90.          90.
           90.          ]
Sampling: (338, 338, 338)
Pixel size: 0.637
Origin: [0, 0, 0]
```

Now, let's calculate the model-based map

```
$ emda model2map --mdl reference_model.pdb --res 1.9 --dim 338 338 338 --cel 215.30601501
215.30601501 215.30601501 90 90 90
```

model2map uses REFMAC to generate map from the model.

In a moment modelmap\_refmac.mrc will be available in the working directory.

Rename it as

```
$ mv modelmap_refmac.mrc reference_map.mrc
```

Your reference map is now ready to use in magnification refinement.

### Step 3: Resampling target maps on the reference

Since the reference\_map.mrc was generated on the emd\_7770\_fullmap.mrc both have the same sampling and pixel sizes.

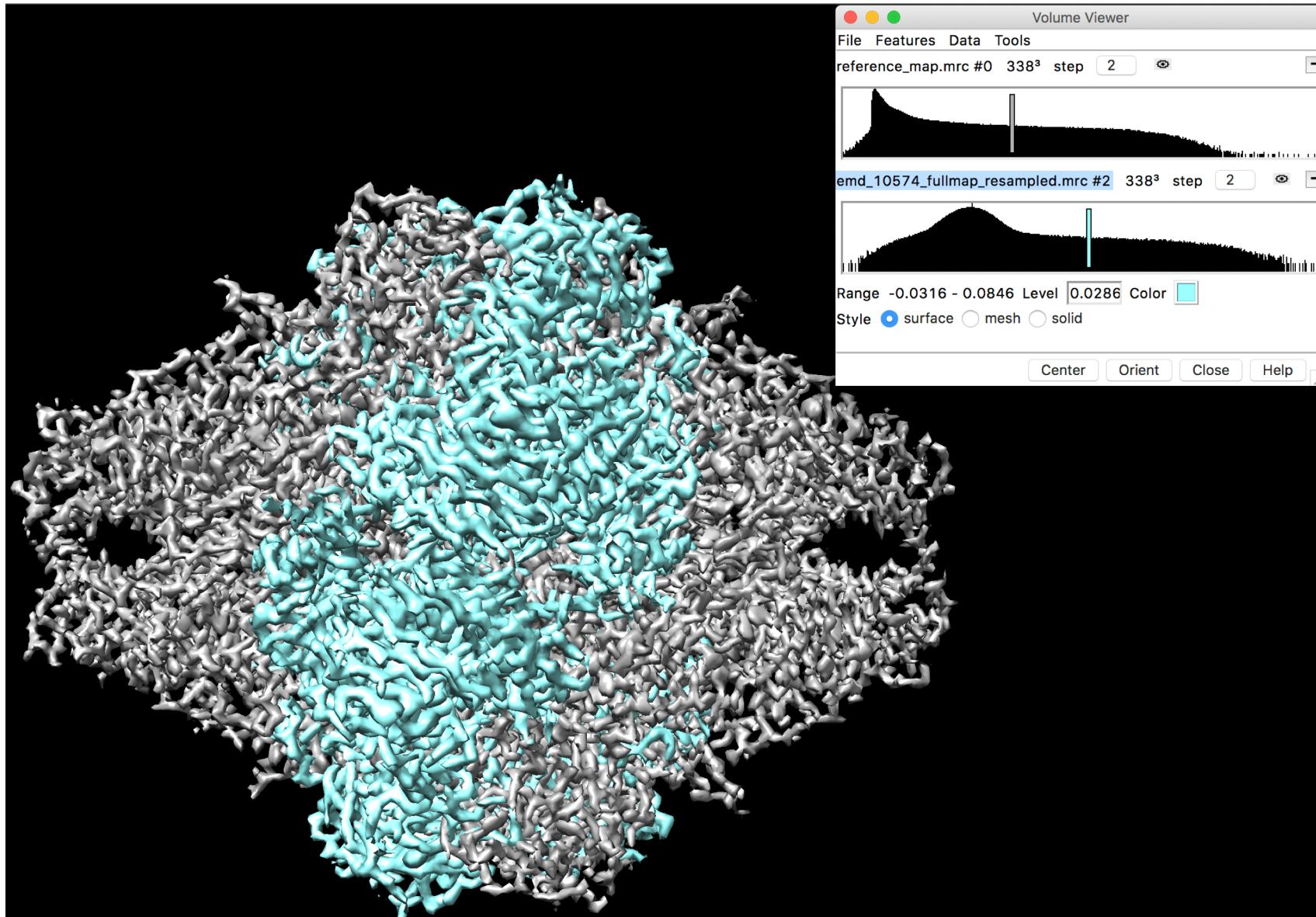
But, emd\_10574\_fullmap.mrc has different sampling and thus different pixel sizes.  
We need to make sure all maps have the same sampling and pixel size.

Now, let's resample emd\_10574\_fullmap.mrc on reference\_map.mrc

```
$ emda resamplemap2map --map1 reference_map.mrc --map2 EMD-10574/emd_10574_fullmap.mrc --out  
emd_10574_fullmap_resampled.mrc  
reference_map.mrc (338, 338, 338) [215.30601501 215.30601501 215.30601501]  
EMD-10574/emd_10574_fullmap.mrc (442, 442, 442) [300.55999756 300.55999756 300.55999756]  
Current pixel size: [0.68, 0.68, 0.68]  
Target pixel size: [0.637, 0.637, 0.637]  
arr.shape: (442, 442, 442)  
Resizing in Fourier space and transforming back  
[-30 -30 -30]  
upsampling...  
cropping image...
```

Find emd\_10574\_fullmap\_resampled.mrc in the working directory

Let's have a look at emd\_7770\_fullmap.mrc and emd\_10574\_fullmap\_resampled.mrc in Chimera



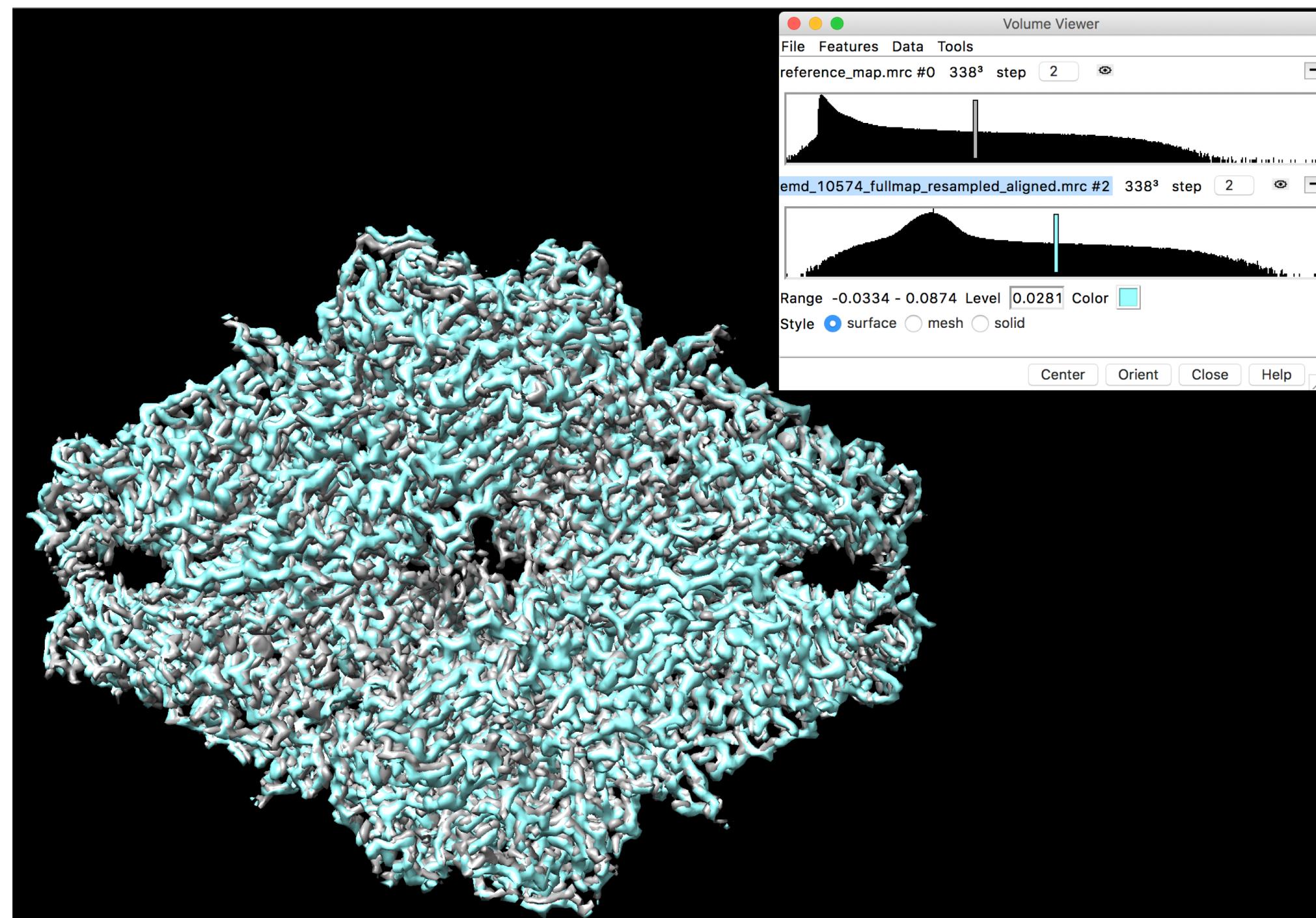
Note that these two maps are almost perpendicular to each other.

#### Step 4: aligning emd\_10574\_fullmap\_resampled.mrc onto reference\_map.mrc

The idea is to get their orientations approximately correct. During magnification refinement, their orientations are optimized.

There are many ways to align those two maps.

1. Use Chimera – keep the reference map static and bring the other map on to it.
2. Use Molrep to align them.
3. Since the principle axes of these two maps are almost perpendicular to each other (and distinct), we can use PCA analysis on their variance-covariance matrices and match the principle components. I have gone in this path and the result is shown below. (PCA analysis was done outside of EMDA)



As you see, they  
are pretty well  
aligned.

You started to see  
the magnification  
difference.

Now, let's refine  
magnifications.

## Step 5: Magnification refinement

Refining magnification of emd\_7770\_fullmap.mrc against the reference\_map.mrc

```
$ emda magref --map EMD-7770/emd_7770_fullmap.mrc --ref reference_map.mrc
```

<truncated>

Optimising overlay.....Done

Optimising map magnification...

ifit	cycle#	func	
val.	magnification		

1	0	2377414.3486	1.0000
---	---	--------------	--------

1	1	2375491.7329	1.0027
---	---	--------------	--------

1	2	2375491.7310	1.0027
---	---	--------------	--------

magnification error (%): 0.268

Magnification corrected map

emda\_magcorrectedmap\_1.mrc was written.

## Step 5: Magnification refinement

Refining magnification of emd\_10574\_fullmap\_resampled\_align.mrc against the reference\_map.mrc

```
$ emda magref --map emd_10574_fullmap_resampled_aligned.mrc --ref reference_map.mrc
```

<truncated>

Optimising overlay.....Done

Optimising map magnification...

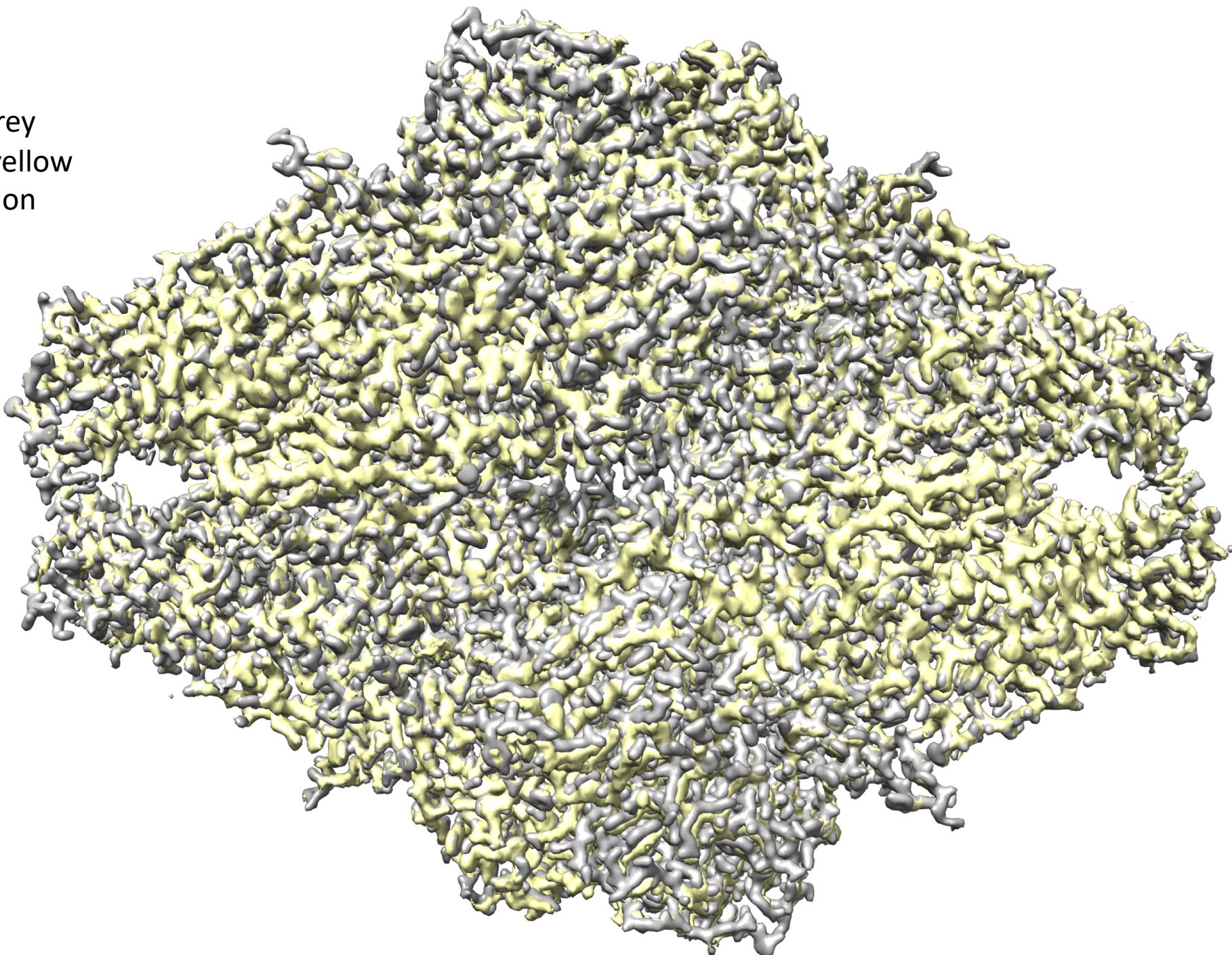
ifit	cycle#	func	
val.	magnification		
1	0	1613534.2744	1.0000
1	1	1549638.7018	1.0171
1	2	1549638.6560	1.0171
1	3	1549638.6560	1.0171
magnification error (%): 1.708			

EMD-7770

Reference in grey

Emd\_7770 in yellow

Before correction

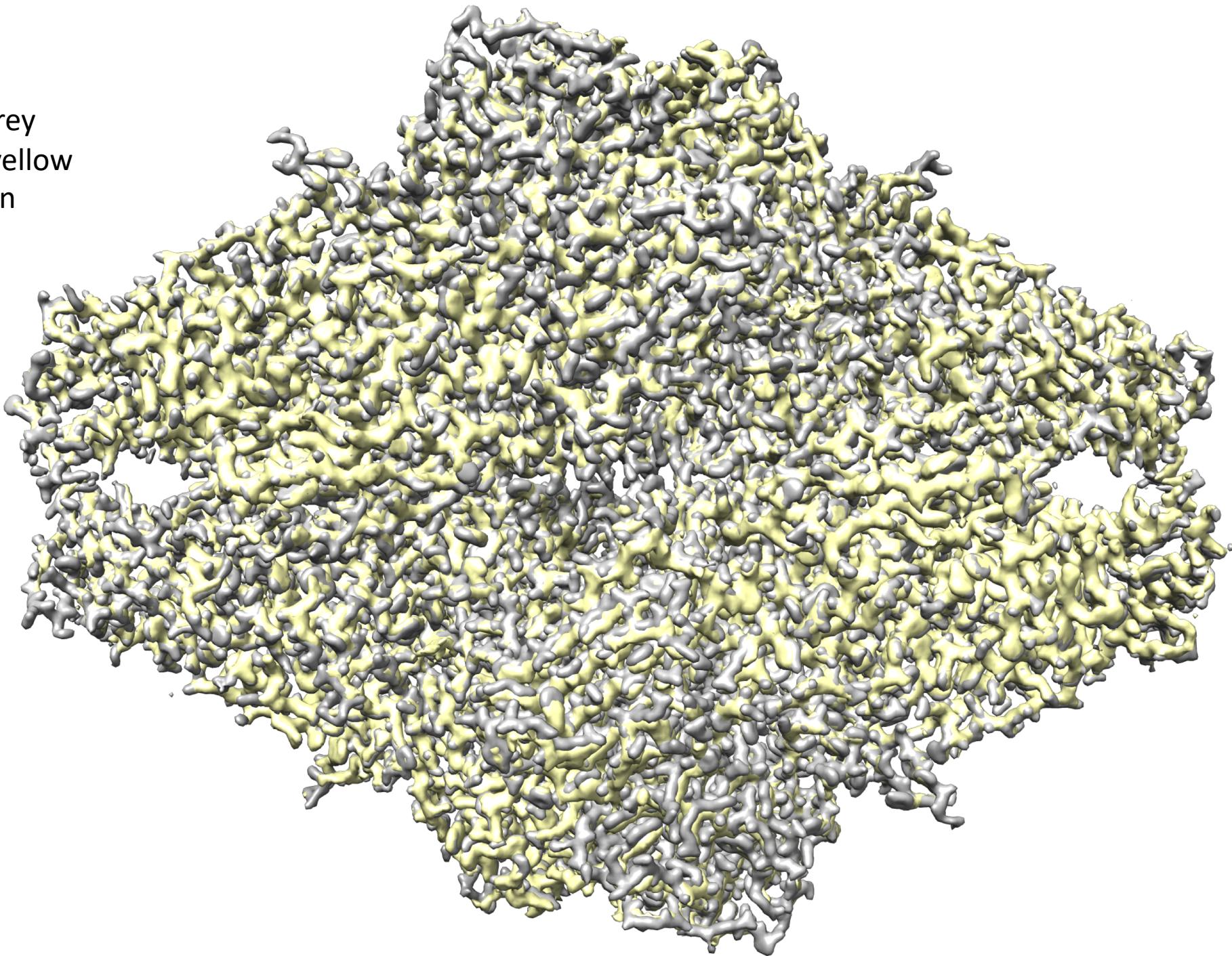


EMD-7770

Reference in grey

Emd\_7770 in yellow

After correction

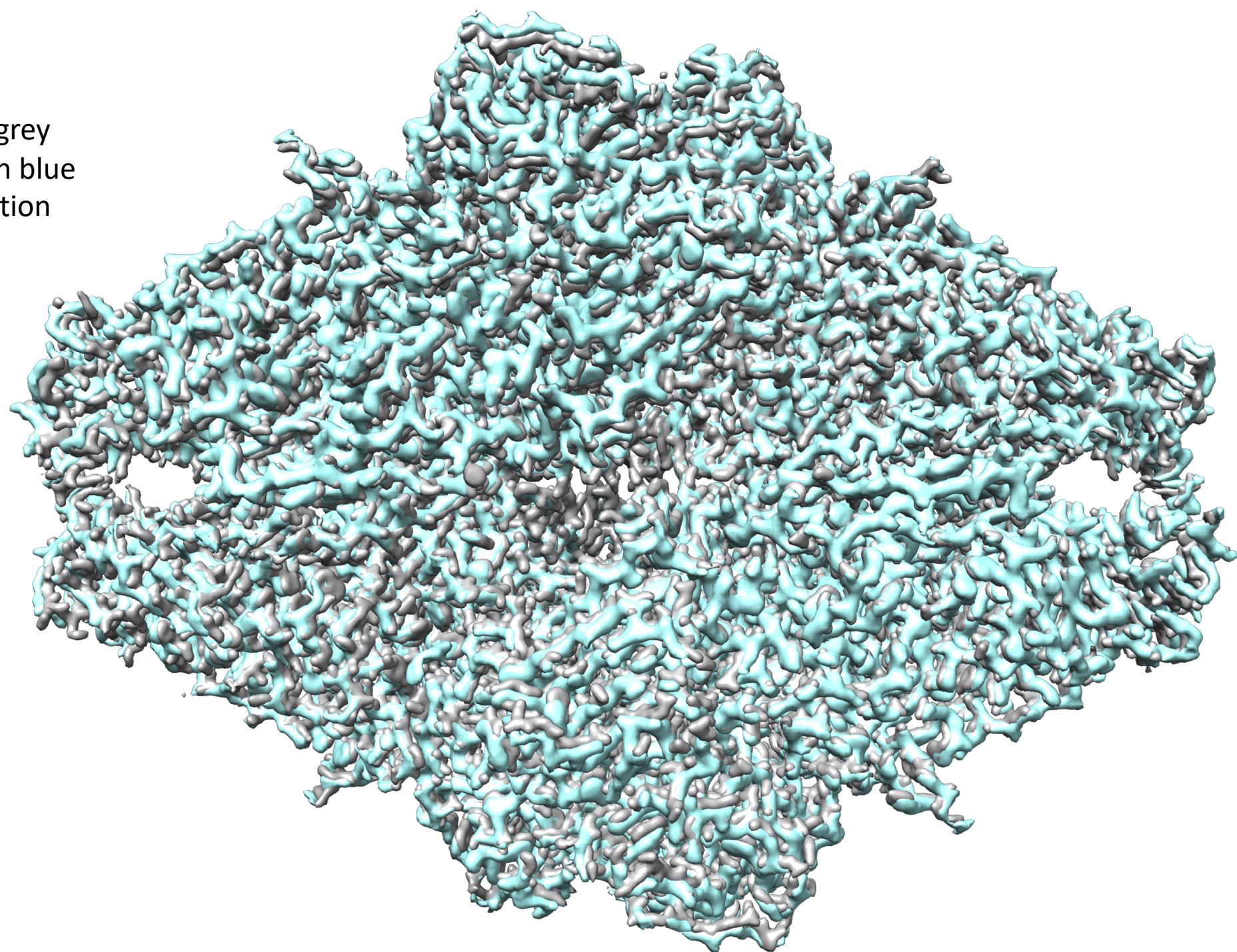


EMD-10574

Reference in grey

Emd\_10574 in blue

Before correction

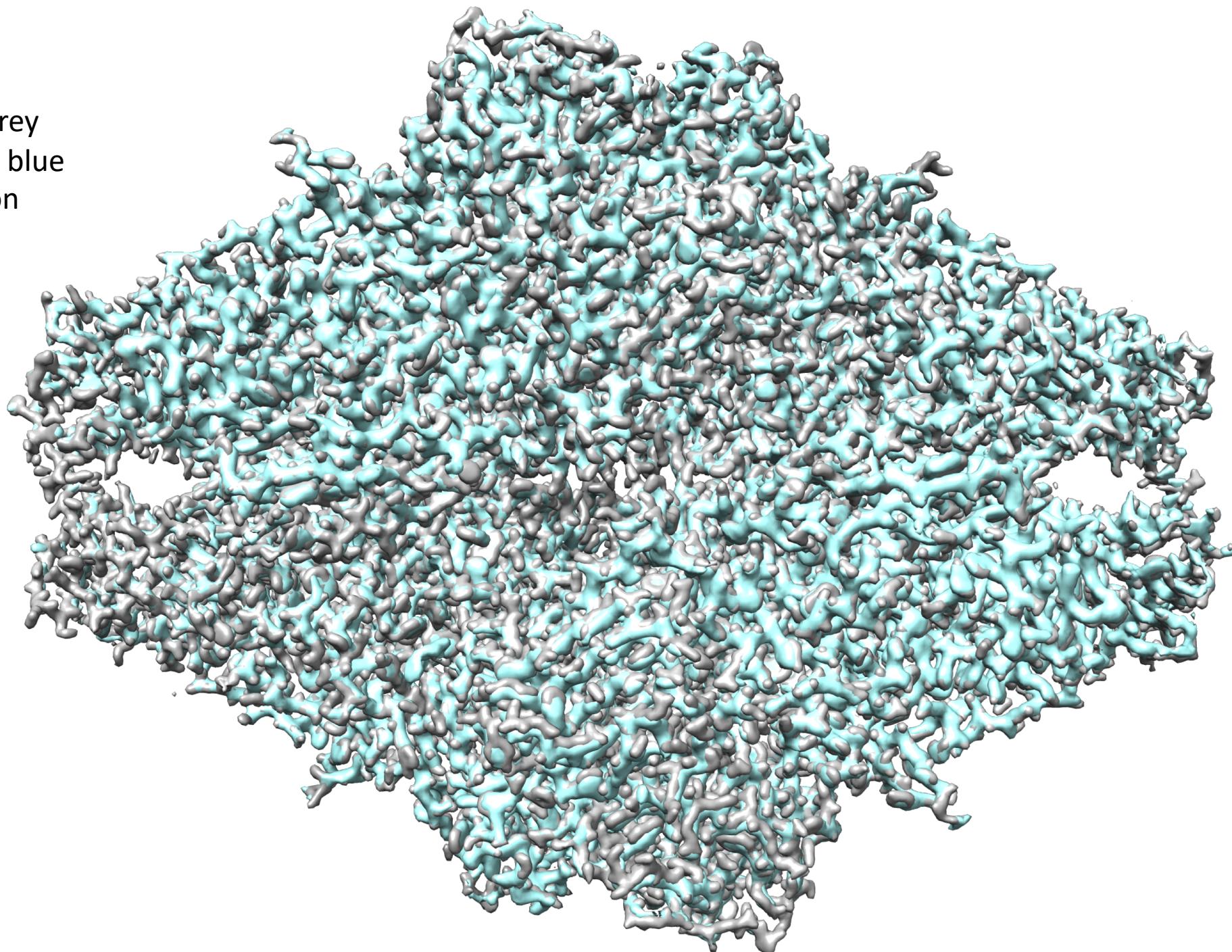


EMD-10574

Reference in grey

Emd\_10574 in blue

After correction



Let's calculate FSC between reference and those maps before and after correction

