NCMI Workshop on Single Particle Reconstruction, Validation and Analysis

March 14-17, 2011

Segger Tutorial

Segmentation of Density Maps and Rigid Body Docking of Structures

Greg Pintilie

Baylor College of Medicine

Introduction

Segger is a tool for segmenting 3D density maps, and for quick rigid-body docking of structures into density maps using segmented regions.

This tutorial will show how to use Segger using several density maps from the EMDB and corresponding PDB structures.

Technical details, quantitative and objective evaluation of the methods used in Segger are beyond the scope of this tutorial. For such information, please refer to the following paper:

> G.D. Pintilie, J. Zhang, T.D. Goddard, W. Chiu, and D.C. Gossard, "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions," *Journal of Structural Biology*, vol. 170, Jun. 2010, pp. 427-438.
>
> Online: http://www.ncbi.nlm.nih.gov/pubmed/20338243
> PDF: http://ncmi.bcm.edu/ncmi/software/segger/segger_jsb.pdf

Other online resources for Segger include the following web sites:

1. http://ncmi.bcm.edu/ncmi/software/segger/
2. http://www.cgl.ucsf.edu/chimera/segger-howto-jun2010/segger.html
3. http://www.cgl.ucsf.edu/chimera/1.5.2/docs/ContributedSoftware/segger/segment.html

Segger is open software, written in Python. The code can be obtained and contributed to at the SourceForge site: http://sourceforge.net/projects/segger/

<u>1. GroEL at 11.5Å resolution</u>

1) In Chimera, open the map (File -> Open …)

2) In the Open File dialog, navigate to the folder where you downloaded the maps, and select emd_1080.map.

- A surface of the map will appear in the main window (Figure 1.1)
- The Volume Viewer dialog will appear as well (Figure 1.2)

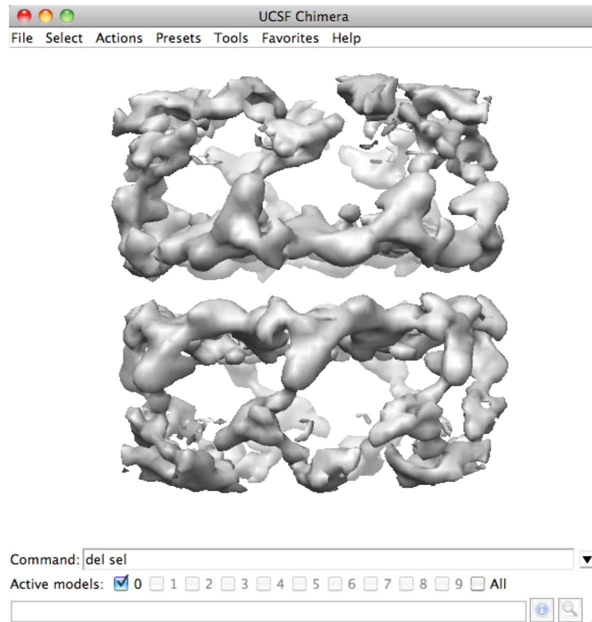

Figure 1.1

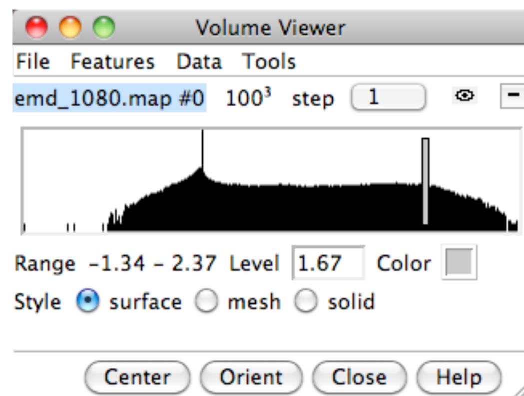

Figure 1.2

- In the Volume Viewer dialog (Figure 1.2), a histogram of the density values in the map is shown. The vertical bar represents a "level" or density value – this value is used to extract the surface shown in the main window (Figure 1.1). You can change the level by click+dragging the bar, or by entering a new value in the Level field (just under the histogram).
- When segmenting the map, the same density level shown in the Volume Viewer dialog will be used: only grid points with density values above this level will be included in the segmentation.
- Before segmenting, you should first choose a level in the Volume Viewer dialog. Typically the level is chosen so that the surface captures as much of the proteins to be segmented as possible, without including too much background noise.
- For maps in the EMDB, there is usually a "suggested contour level". For this map it's 0.852 (http://emsearch.rutgers.edu/atlas/1080_visualization.html).

3) Drag the bar in the Volume Viewer dialog until the Level value is close to 0.852, or enter the number 0.852 in the Level field and press enter.

4) Open the Segment Map dialog by selecting from the top menu bar in the main window Tools -> Volume Data -> Segment Map (Figure 1.3).

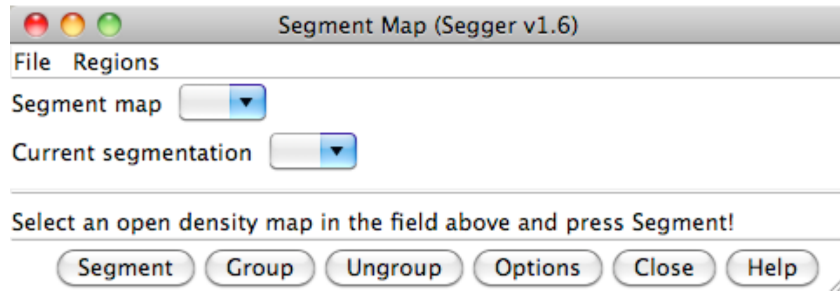

Figure 1.3. The Segment Map dialog

5) Select the map to be segmented in the "Segment map" field. This field is a drop-down menu which will show a list of open density maps. After you select emd_1080.map from this list, you will see the name appear in the field (Figure 1.4).

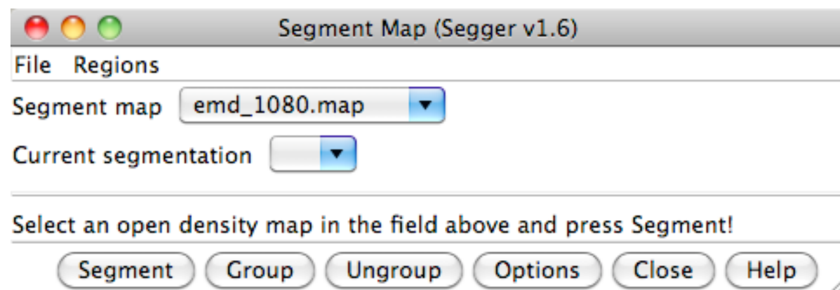

Figure 1.4. Segment Map dialog with emd_1080.map selected for segmentation.

6) Press the **Segment** button at the bottom of the Segment Map dialog.

- After a few seconds, you will see the results in the main window (Figure 1.5).
- The Segment Map dialog will also update (Figure 1.6).
    - The "Current segmentation" field will now have the text emd_1080.seg
    - Next to that field, the number of segmented regions is reported (28 regions)
    - Results are described in the text just above the buttons at the bottom: "428 watershed regions, smoothed 3 voxels to get 28 regions".

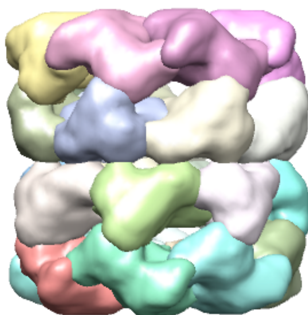

Figure 1.5

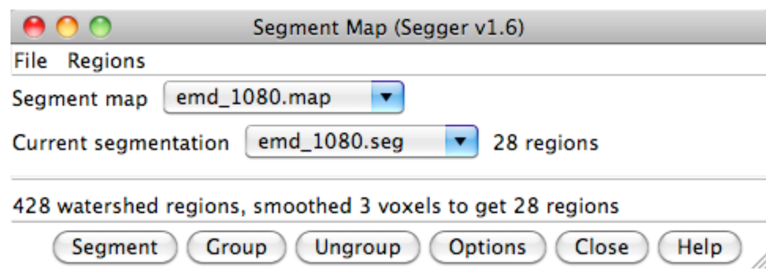

Figure 1.6

- The segmentation process starts with the watershed method, which produces regions corresponding to local density maxima in the density map. Each region extends to the voxels with the lowest density value between adjacent local maxima.
- The watershed process generally produces too many regions (428 in this example). Hence the map is smoothed, and regions are grouped based on smoother maps. In this case, after smoothing we end up with 28 regions. The more smoothing that is done, the fewer regions that will be obtained.

7) In this example, we are looking for 14 proteins. Thus we want to take the smoothing process a bit further. To do this, press the **Group** button at the bottom of the Segment Map dialog. When no regions are selected, this action smooths the map one more step and groups the current regions based on this smoothed map. The result will be 14 regions (Figures 7, 8).

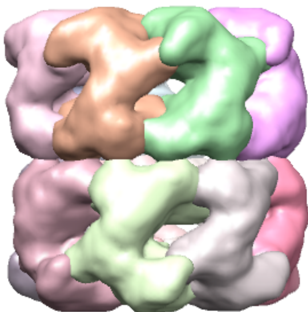

Figure 1.7

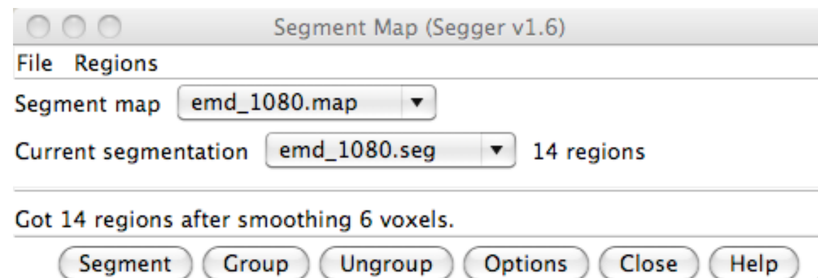

Figure 1.8

8) When pressing the **Segment** button, several parameters are used by default. To see these parameters, press the **Options** button at the bottom of the Segment Map dialog. You should see the dialog expand as in Figure 1.9.

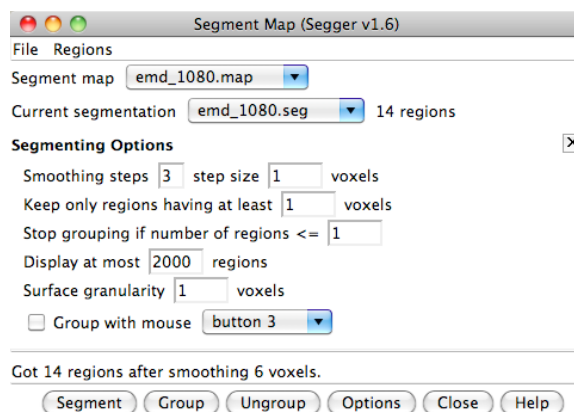

Figure 1.9. Segment Map dialog with options shown.

- Smoothing steps and step size
  - The number of smoothing steps is how many times the map is smoothed for the purpose of grouping regions. The smoothing is done using a Gaussian filter.
  - The step size is the amount of smoothing at each step. It translates into the filter width (i.e. twice the standard deviation of the Gaussian kernel).
- Keep only regions having at least ___ voxels.
  - The value entered in this field is used to remove regions in the initial segmentation that are smaller than the given size. This is useful in maps with lost of noise.
- Stop grouping if number of regions <= ___
  - The number entered here acts as a lower bound on how many regions will be produced. In this example, we might enter 14 since we know there are 14 proteins that we want to segment in the map.
- Display at most ____ regions
  - Chimera will slow considerably when more than several thousand region surfaces are shown. Hence when the segmentation contains more than this many regions, some of them are not shown. They are still there and will appear as part of larger regions when grouping further.

9) Enter different numbers of smoothing steps and press **Segment**.

- The results for 0, 1, 2, 3, 6 are shown in Figure 1.10, from left to right respectively.
- This illustrates the overall segmentation process – the regions after 0 steps are the watershed regions. As more steps are taken, regions become larger and fewer.
- The result on the right, with 14 regions, can also be obtained by entering a large number of steps (e.g. 20), and 14 in the "Stop grouping if number of regions <=" field.

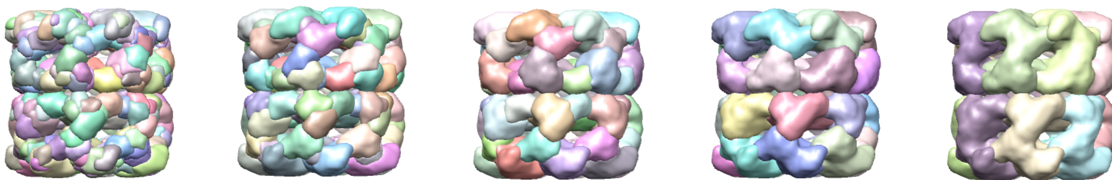

Figure 1.10. Resulting regions for number of steps 0,1,2,3,6, from left to right.

- Regions are shown as surfaces that enclose all the voxels within each region. The regions are stored in a data structure that appears in the Model Panel as a single model (emd_1080.seg). To open the model panel, choose from the menu bar in the main window Tools -> General Controls -> Model Panel. This model can be shown/hidden just like any other model.

10) Select a region in the segmentation by moving the mouse pointer over it, and pressing Ctrl + Left Button. The region will become highlighted with a green outline. (Figure 1.11).
11) From the menu bar in the Segment Map dialog, choose Regions -> Show only selected.

- This will hide all other regions except the selected one. (Figure 1.12)
- You can hide the density map either in the Model Panel dialog (click the checkbox next to emd_1080.map under the S column), or in the Volume Viewer dialog (click the eye-like icon to the left of the minus sign). Then only the region will be visible (Figure 1.13).

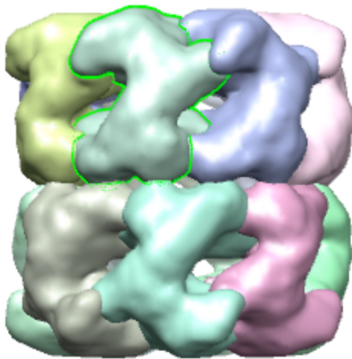
Figure 1.11

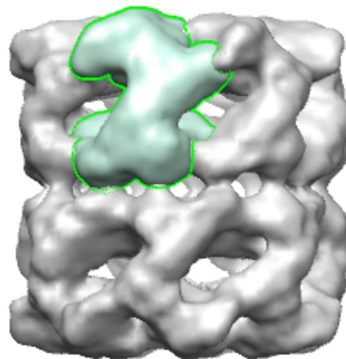
Figure 1.12

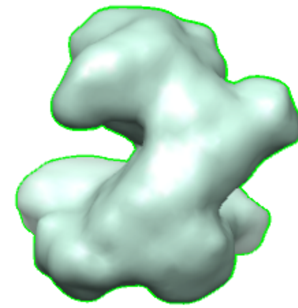
Figure 1.13

12) From the Segment Map dialog top menu, select Regions -> Mask map with selected.

- This extracts from the original density map a new map where all density values outside this region are set to 0.
- In the Model Panel and Volume Viewer dialogs, a new density map will appear, having the name emd_1080_masked.
- This map can be saved as a new map. In Volume Viewer, choose File -> Save map as. Alternatively, you can also directly create this masked map and save the file at the same time – with the region selected, choose from the Segment Map dialog top menu File -> Save selected regions to .mrc file.

We can also fit a structure directly to the region:

13) First load, the structure 1XCK_chain.pdb from the segger_data directory.
14) Select a single chain from the structure, e.g. by selecting a single residue (Ctrl + Left Button), and the pressing the Up Arrow on the keyboard.
15) Open the Fit to Segments dialog from the top menu in the main window by selecting Tools -> Volume Data -> Fit to Segments. (Figure 1.14)
16) Click on the Structure to fit field, which is a drop down menu that will show a list of open structures, and select 1XCK_chain.pdb

- Note that if you hadn't opened the 1XCK_chain.pdb file yet, it would be shown in the list under the In <>/segger_data : entry, wher <> is the location of the segger_data folder. When you select an entry here, the file is automatically opened. This is a shortcut for opening structures in the same directory as the map that was segmented.

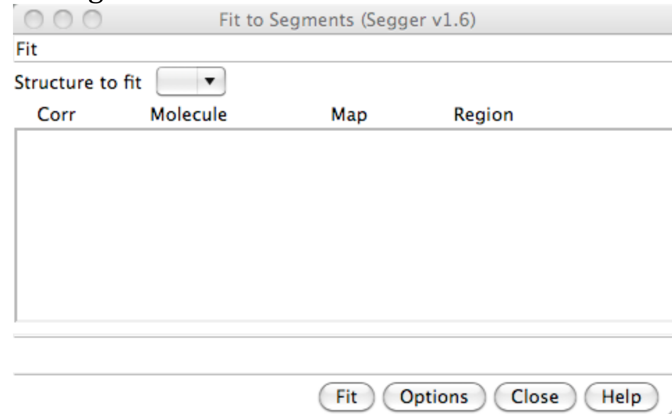

Figure 1.14. The Fit to Segments dialog.

17) With the segmented region still selected, and the structure 1XCK_chain.pdb selected in the Structure to fit field, press the **Fit** button in the Fit to Segments dialog.

- First a density map is simulated for the structure, which will briefly appear and then be hidden. This map will appear in the Model Panel and Volume Viewer dialogs as 1XCK_chain_r8.1_sp2.7.mrc. The numbers represent the resolution (8.1Å), deduced from the map, and the grid spacing, 2.7Å, also taken from the density map.
- The structure and simulated density map are then aligned to the region. After a few seconds, the regions will become transparent, and the structure will have been fit (Figure 1.15).
- A new entry will appear in the list in the Fit to Segments dialog (Figure 1.16). The line represents the fit and includes the structure, map, region it was fit to, and density cross-correlation of the fit (0.9176). You can select this line anytime while the structure and map are still open to recreate the fit (e.g. after moving the structure to another region).

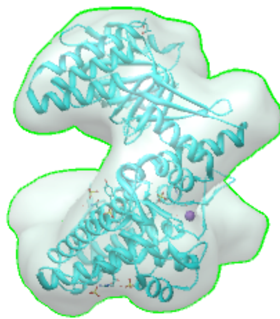

Figure 1.15

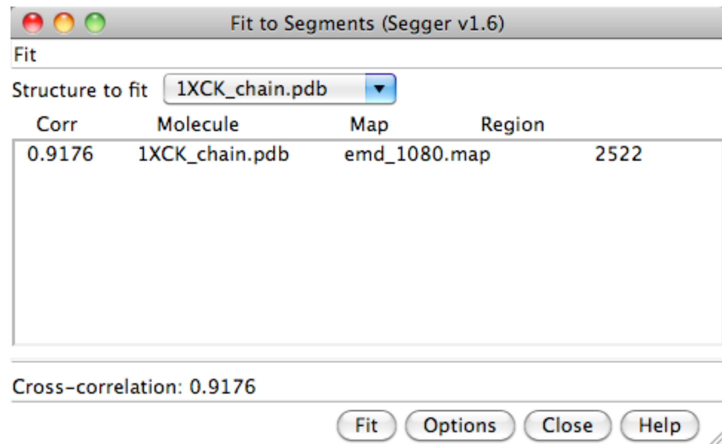

Figure 1.16

- The Fit to Segments dialog also has further options that can be revealed by pressing the **Options** button at the bottom of the dialog. Figure 1.17.

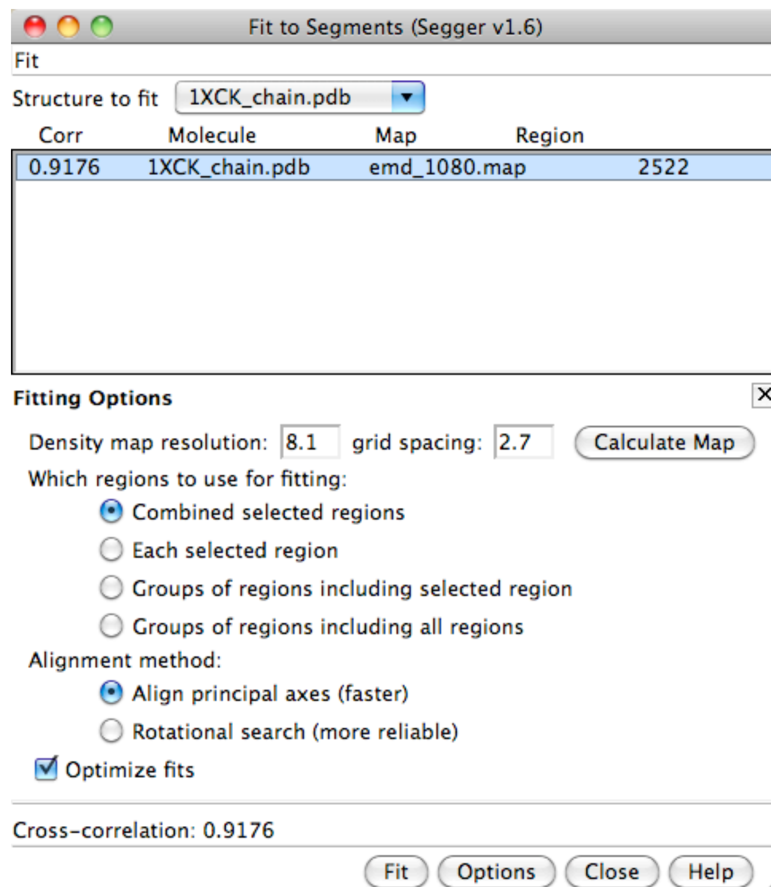

Figure 1.17. The Fit to Segments dialog with Options shown.

18) In the Segment Map dialog, select Regions -> Show all to show all the regions again.
19) Add another region to the current selection by moving the mouse pointer over it and pressing Ctrl + Shift + Left Button. You should now have two regions selected (Figure 1.18).
20) In the Fit to Segments dialog, select "Each selected region" under "Which regions to use for fitting", and press the **Fit** button.

- The same structure will be fit to both selected regions in turn, and two new lines will appear in the central box in the dialog.
- The structure will be shown placed in the last fit that was performed, and all regions that were used for fitting will be made transparent. (Figure 1.19).
- By selecting each line, the structure will be moved to the corresponding position.

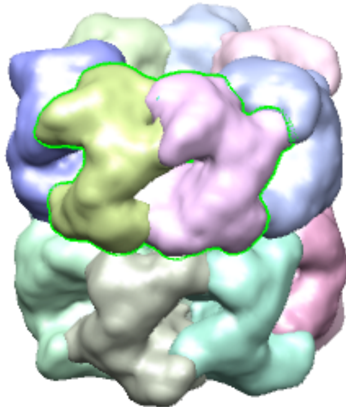
Figure 1.18

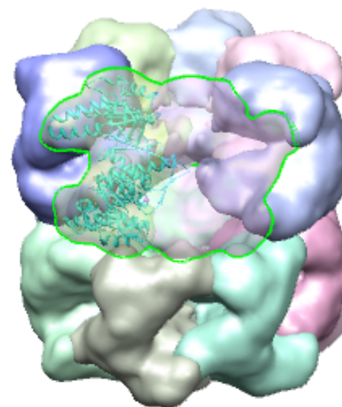
Figure 1.19

21) Select in the Fit to Segments dialog, from the top menu bar, Fit -> Place molecule copies.

- This duplicates the structure in the position for that particular fit. To save the structure to a file in this position, select Fit -> Save chosen fit molecules.

22) Select all the fits in the Fit to Segments dialog with Ctrl + Left Button on the first entry and Ctrl + Shift + Left Button on the last entry, and choose Fit -> Delete Fits from list.

23) Close all models by choosing from the menu bar in the main window File -> Close Session.

2. GroEL at 4.2Å resolution

1) Open the map emd_5001.map and set the level in the Volume Viewer dialog to 0.6.

2) In the Segment Map dialog, choose emd_5001.map in the “Segment map” field, and set “Smoothing steps” to 4, “step size” to 4 voxels, and press the **Segment** button (Figure 2.1).

- This will produce 14 regions, as shown in Figure 2.2

3) Enter 40 smoothing steps, step size of 1, set “stop if number of regions <= “ to 14, and press the **Segment** button.

- This will also produce 14 regions (stopping after 14 steps), however the regions are a bit different – compare Figure 2.3 and 2.4.
- This result is a bit more undesirable, i.e. part of a region looks disconnected. This happens in maps that are a bit more noisy. Since in step 2 we used a higher step size, the larger degree of smoothing helps to reduce the effects of noise for the first grouping step.

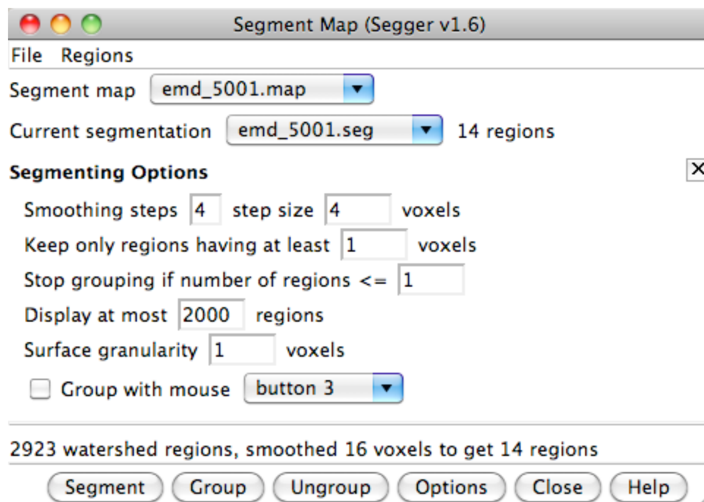

Figure 2.1

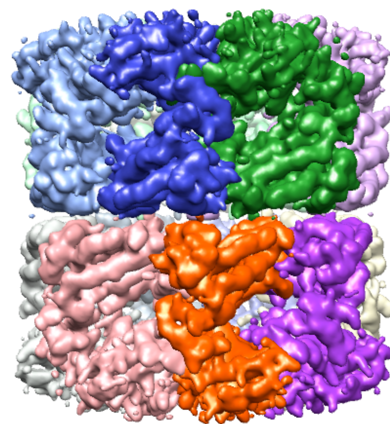
Figure 2.2

4) Enter 4 smoothing steps, step size of 4, set “Keep regions having at least ___ voxels” to 20, and press the **Segment** button.

- This acts to remove the smaller regions (in the initial watershed segmentation), which reduces some of the smaller ‘islands’ visible in the previous segmentations. A region from this segmentation is shown in Figure 2.5.

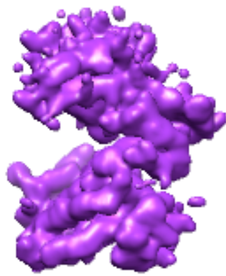

Figure 2.3. Resulting region after step 2.

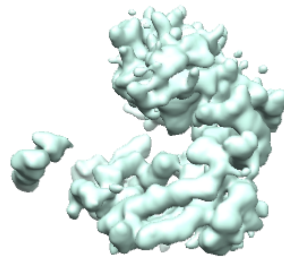

Figure 2.4. Resulting region after step 3.

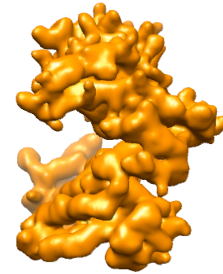

Figure 2.5. Resulting region after step 4.

5) Select one of the segmented regions, and in the Segment Map dialog select Region -> Show only selected.

6) In the Fit to Segments dialog, select 1XCK_chain.pdb in the "Structure to fit" field, set "Which regions to use for fitting" to "Combined selected regions", and press the **Fit** button.

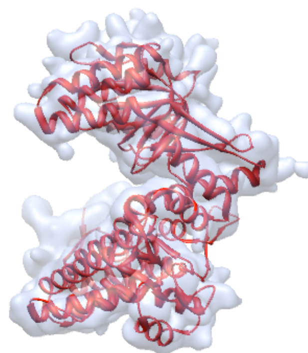

Figure 2.6

7) Close all models by choosing from the menu bar in the main window File -> Close Session.

3. Mm-cpn at 4.3Å resolution

1) Open the map emd_5137.map and set the level in the Volume Viewer dialog to 0.3.

2) In the Segment Map dialog, choose emd_5137.map in the “Segment map” field, and, as in Figure 3.1,

- set “Smoothing steps” to 6
- set “step size” to 2 voxels
- set “Keep only regions having at least 30 voxels”
- press the **Segment** button

- This produces 16 regions, however the apical section is segmented all together, forming a disc-like extension at the top (Figure 3.2)

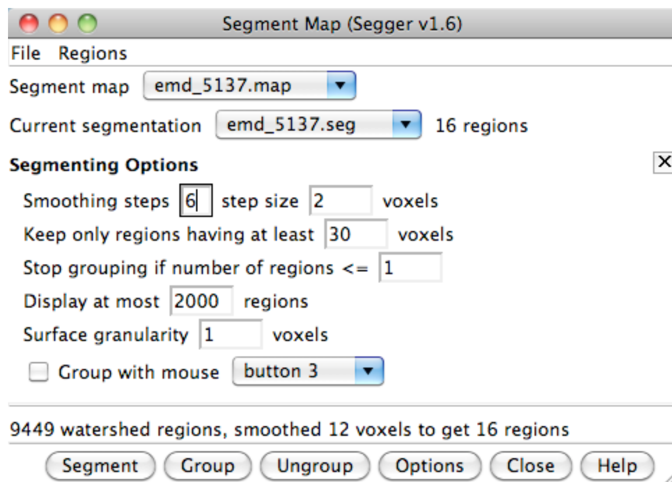

Figure 3.1

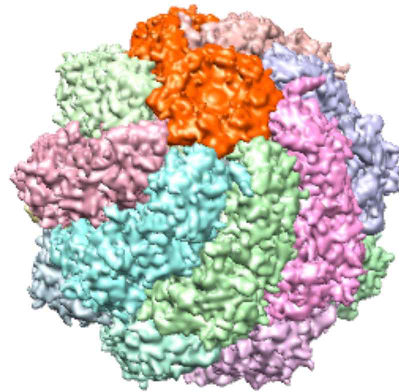

Figure 3.2

3) Select the region with the disc-like section at the top by moving the mouse pointer over it and pressing Ctrl + Left Button. Then press the **Ungroup** button in the Segment Map dialog.

- This action takes the selected region and “ungroups” the regions that were grouped together to create it. The result is two regions. The top region still has the form of a disc (Figure 3.3)
- After “ungrouping”, the regions that were ungrouped remain selected. If for some reason the ungrouping did not produce a desirable result, it can easily be undone by pressing the **Group** button with the same regions selected. When more than one region is selected, this button does not do the default “smooth and group” process, as described in section 1 of this tutorial – instead it takes the selected regions and groups them together.

4) Select the disc-like region by itself by moving the mouse pointer over it and pressing Ctrl + Left Button (alternatively, you can remove the other non-disc like region from the selection by moving the mouse pointer over it and pressing Shift +

Ctrl + Left Button, to leave just the disc-like region selected), and then press the **Ungroup** button again.

- This now produces 8 regions (Figure 3.3), which are the apical segments of the 8 proteins that are arranged in a circle in this protein complex.

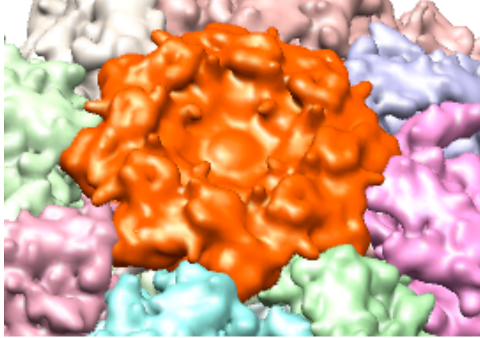

Figure 3.2

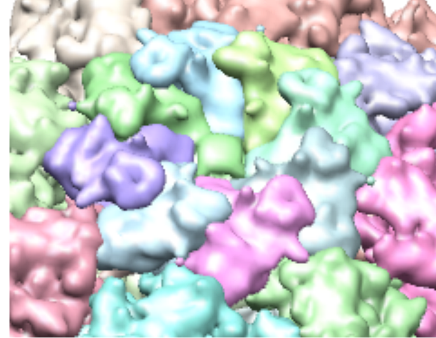

Figure 3.3

5) Select the two regions that correspond the same protein, as shown in Figure 3.4, and press the **Group** button.

- After the two regions are grouped, the resulting region remains selected. In the Segment Map dialog, select Region -> Show only selected, to see this region by itself (Figure 3.5).

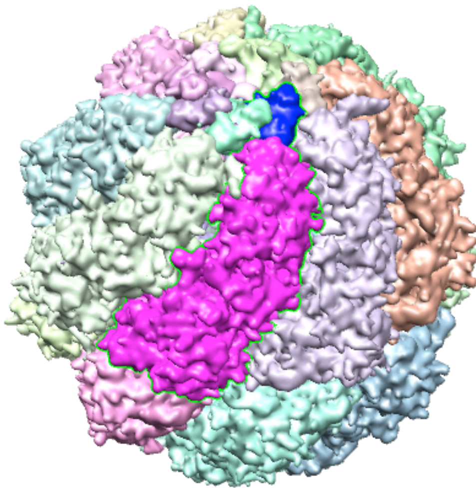

Figure 3.4

Figure 3.5

- In this section, we interactively modified the segmentation. Segger creates segmentations quickly and easily, however the results are not perfect, i.e. it won't segment each protein from each other exactly as it should be. This interactive process allows you to modify the segmentation results.
- The edited segmentation can be saved to a file so that it can be reloaded later. To do this, select File -> Save segmentation, or File -> save segmentation as… in the Segment Map dialog.

6) In the Fit to Segments dialog, select the structure 3LOS_chain.pdb in the "Structure to fit" field, and press the **Fit** button.

- The results should be as shown in Figure 3.6

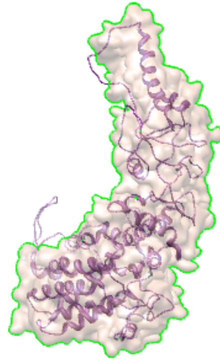

Figure 3.6

4. Closed RyR1 channel at 9.6Å resolution

1) Open the map emd_1275.map and set the level in the Volume Viewer dialog to 1.0.

2) In the Segment Map dialog, choose emd_1275.map in the "Segment map" field, and, as in Figure 4.1,

    - set "Smoothing steps" to 3
    - set "step size" to 2 voxels
    - set "Keep only regions having at least 0 voxels"
    - press the **Segment** button
- This produces 48 regions

3) Rotate the view as shown in Figure 4.2

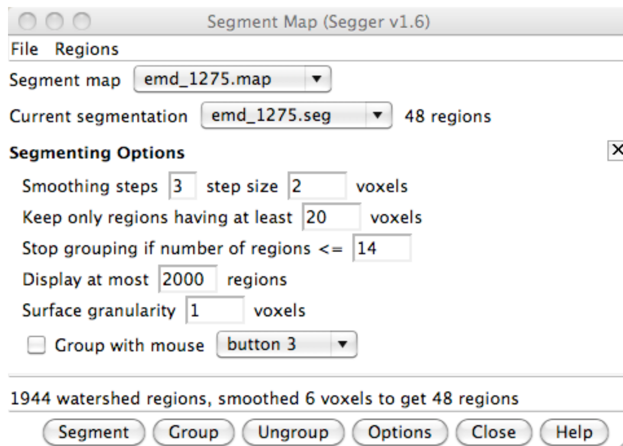

Figure 4.1

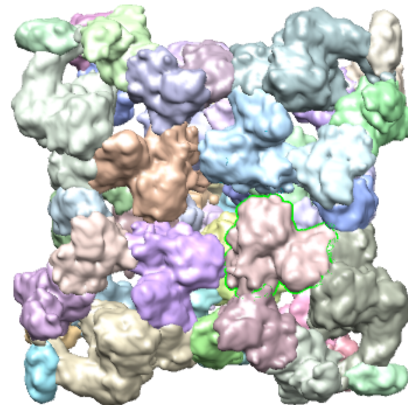

Figure 4.2

4) Select the region shown in Figure 4.3, which is close to the center, and in the Segment Map dialog, select Region -> Show only selected. You should then see only this region, as shown in Figure 4.4 (you will have to hide the density map as well to see only the region).

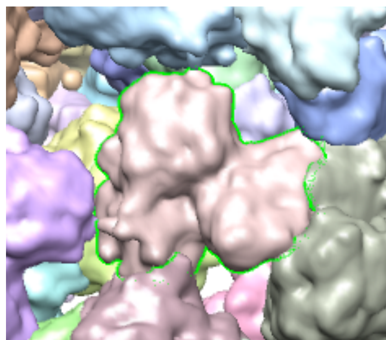

Figure 4.3

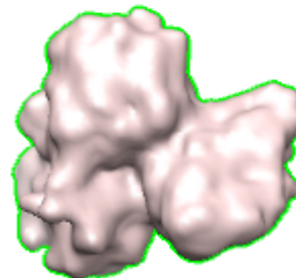

Figure 4.4

5) In the Fit to Segments dialog, in the Structure to Fit field, select 2XOA.pdb

6) Set “Which regions to use for fitting” to “Combined selected regions”, “Alignment method to “Align principal axes”, and press the **Fit** button.

- The resulting fit is shown in Figure 4.5.

7) Set “Alignment method” to “Rotational search” and press the **Fit** button.

- The resulting fit is shown in figure 4.6.
- Note the fits obtained in step 6 and here are different – the one obtained here has a higher density cross-correlation.
- The method used in step 6 is faster, however the method used here is more thorough, and hence a better fit is found. Method 6 works well for quick fits, but if the fit doesn’t look right, this more thorough method should be tried.
- This method is not meant to replace more thorough ‘exhaustive search’ methods as done by programs such as Foldhunter, Situs, and ADP_EM. It is only meant to facilitate quick interactive visualization of possible fits in various areas of the map, and for quick comparisons of segmented regions to structures.

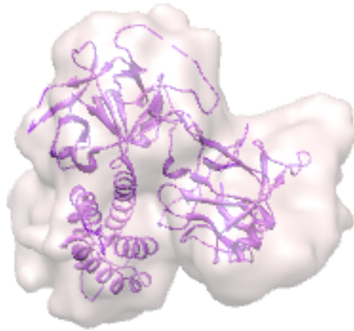

Figure 4.5

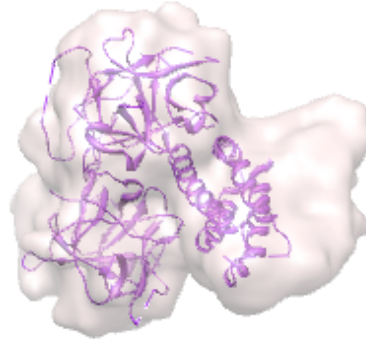

Figure 4.6

5. Ribosome at 6.4 Å resolution

1) Open the map emd_5030.map and set the level in the Volume Viewer dialog to 1.4.

2) In the Segment Map dialog, choose emd_5030.map in the "Segment map" field, and, as in Figure 5.1,

- set "Smoothing steps" to 6
- set "step size" to 4 voxels
- set "Keep only regions having at least 0 voxels"
- press the **Segment** button

- This produces 2 regions, representing the large and small subunits. We will now "ungroup" the small subunit to look for the elongation-factor protein.

3) Rotate the view as shown in Figure 5.2

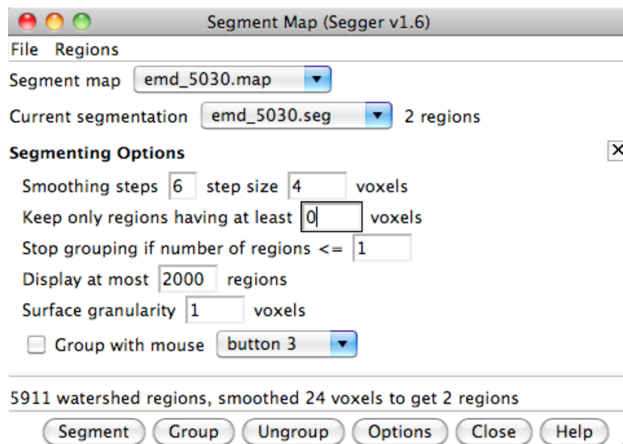

Figure 5.1

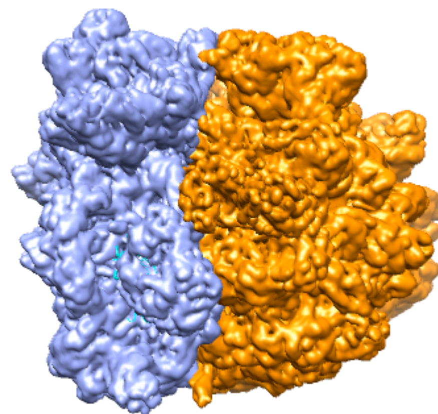

Figure 5.2

4) Select the small subunit (which is blue and on the left side in the image above), and in the Segment Map dialog select Regions -> Show only selected. You should see the small subunit alone, as in Figure 5.3.

5) In the Segment Map dialog, press the **Ungroup** button. This will split up the small subunit into two regions, as in Figure 5.4.

6) Select the bottom region, and press the **Ungroup** button again. You should see further smaller regions as in Figure 5.5.

7) Select the region shown in Figure 5.6 and in the Segment Map dialog select Regions -> Show only selected. You should see the region by itself as shown in Figure 5.6

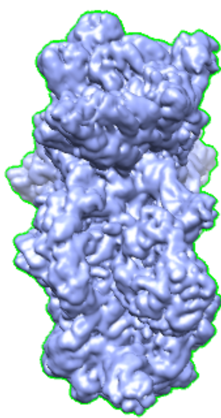

Figure 5.3

Figure 5.4

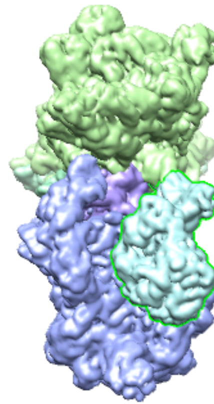

Figure 5.5

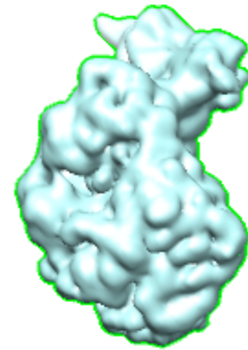

Figure 5.6

8) With the region in Figure 5.6 selected, press the **Ungroup** button. You should see regions as in Figure 5.7.

9) Select one of the regions as shown in Figure 5.8

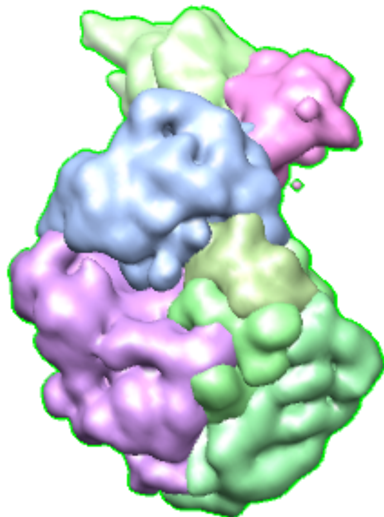

Figure 5.7

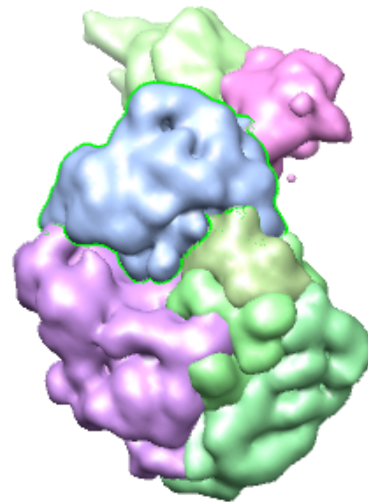

Figure 5.8

10) In the Fit to Segments dialog, select ribosome_EF.pdb in the "Structure to fit" field.

11) For "Which regions to use for fitting" select "Groups of regions including selected region", for "Alignment method" select "Align principal axes", and press the **Fit** button.

- This method find suitable groups of regions around the selected region and fits the structure to each of these groups. The final fit should be as shown in Figure 5.9
- When compared to the segmented region in Figure 5.6, shown again in Figure 5.10, the structure matches quite well. However the region includes some parts of the RNA, which makes fitting directly to this region not work.
- This example illustrates how Segger can be used to do a coarse initial segmentation, and then ungroup regions to look for smaller segments within the initial larger regions. It also illustrates how to automatically look for sub-groups of regions that match a particular structure.

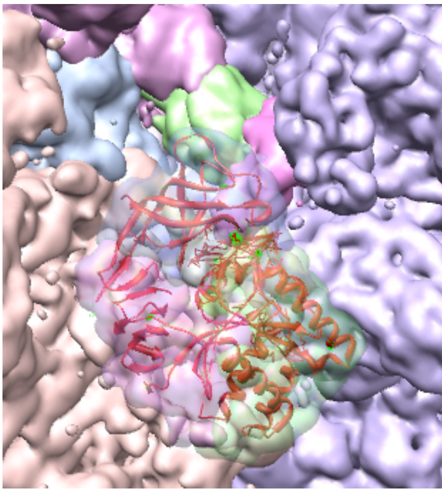
Figure 5.9

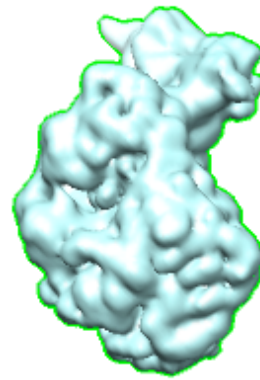
Figure 5.10