Segger Tutorial – SegFit Dialog

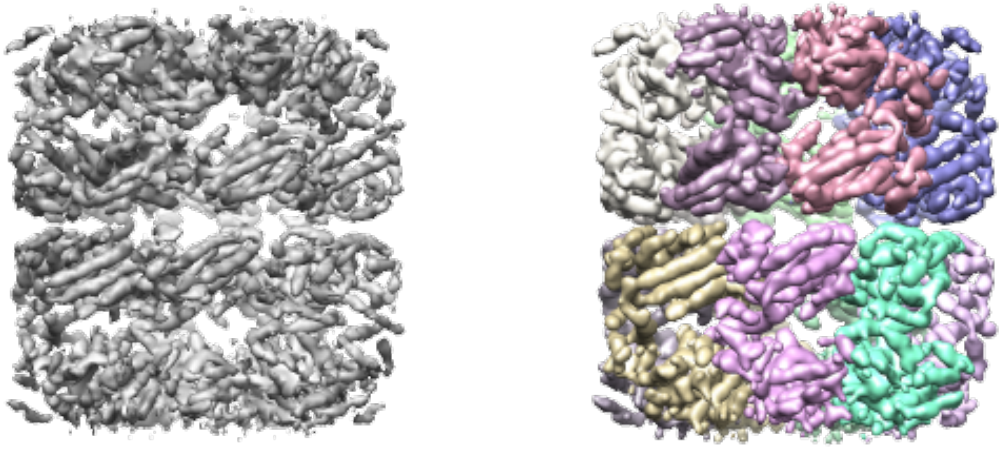
Last updated: Jan. 31, 2021 (Segger v2.5.4, Chimera Version 1.13)
By: Greg Pintilie (gregdp@stanford.edu)

1. Segmenting the map

Segger can help rigidly dock atomic structures (or other density maps) into a map by aligning the model with segmented regions. The process starts with segmenting the map. If you haven't already, start with the Segger Dialog tutorial.

As a simple example we will use here the density map of GroEL at 4.2Å resolution (EMDB:5001). This map can be download from this link, or via File -> Fetch by ID -> EMDB: 5001.

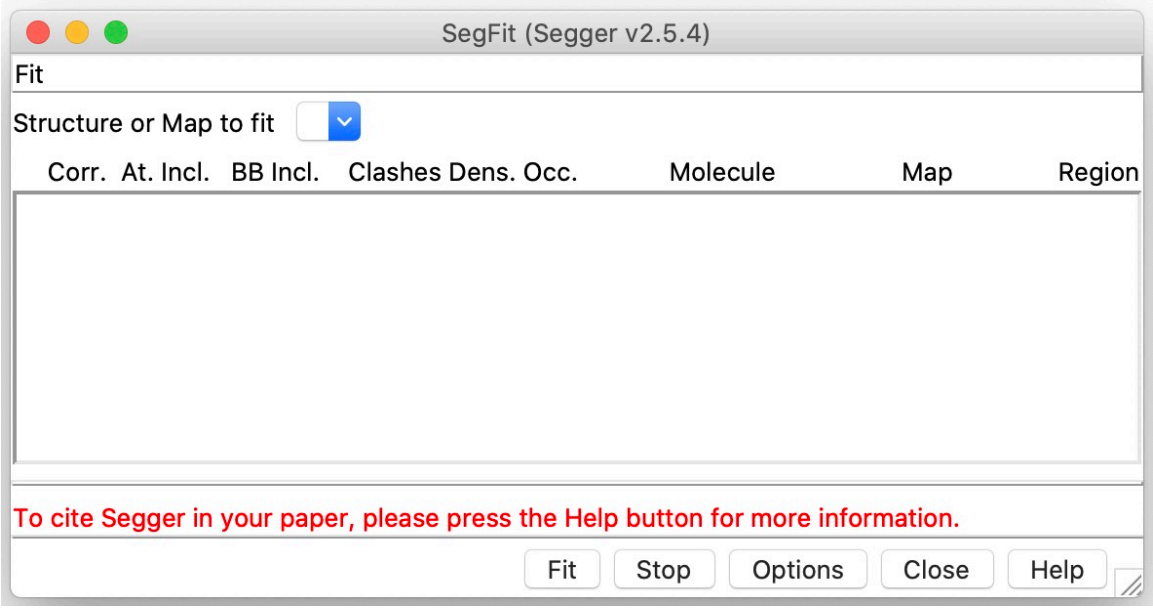
It is segmented here using the Segger dialog at a threshold of 0.9, with 2 steps of size 7. The result is 14 regions, with each region corresponding to a single protein. The density map is shown below on the left, and the 14 segmented regions are on the right:



Note: For this example we used large smoothing steps intentionally. When taking smaller steps, less accurate segmentation regions were obtained. This is likely due to noise in the density map. Applying more smoothing in the first step is helping to suppress this noise.

2. SegFit Dialog

The interface for fitting structures/maps to segments can be opened from the Volume menu in Chimera, under Volume Data -> SegFit, or via Shortcuts -> Other tools: SegFit on the Segger dialog. This interface is shown below.



Next we'll obtain a structure of a single protein to fit to each segmented region shown above.

3. Obtaining the structure of a single protein

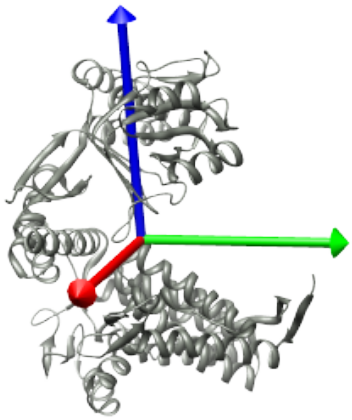
Open the structure of GroEL either:

- Download PDB:1xck, then open in Chimera form File menu.
    - Note you can also conveniently open files that are in the same folder as the density map from the SegFit dialog, from the drop-down 'Structrue or Map to Fit', which lists all maps and models in the same folder as the density map.
- Or via File -> Fetch by ID… -> PDB: 1XCK:

Next we want to isolate the structure of a single protein from this structure of the entire complex:

- One way of doing this is to select a single chain of this structure, e.g. chain A, by *Ctrl+Click* on any one atom or part of a ribbon, and then pressing the *Up* key on the keyboard. Once a single protein is selected, it can be saved to an individual file using File / Save PDB... (make sure *Save selected atoms only* is checked in the dialog that appears, and enter a name, e.g. *1xck_A.pdb*. Then use the drop-down menu to the right of *Structure to fit* once again to select and open this structure by itself.
- Alternatively, you can select all chains in the *1xck.pdb* structure other than A, and execute the command *del sel* in the command line under the Chimera main window (you may have to enable this from the Settings panel if it's not already enabled).

Once the structure of a single protein is selected in the drop-down menu to the right of *Structure to fit*, its principal axes can be shown from the **Fit** menu at the top of the **SegFit** dialog, by selecting **Show molecule axes**. The structure and its principal axes are shown in the image below.

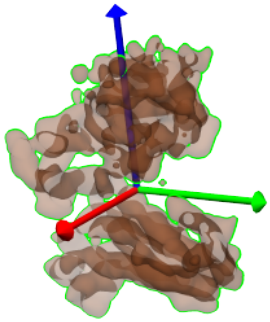


4. Fitting the structure by aligning it to a region

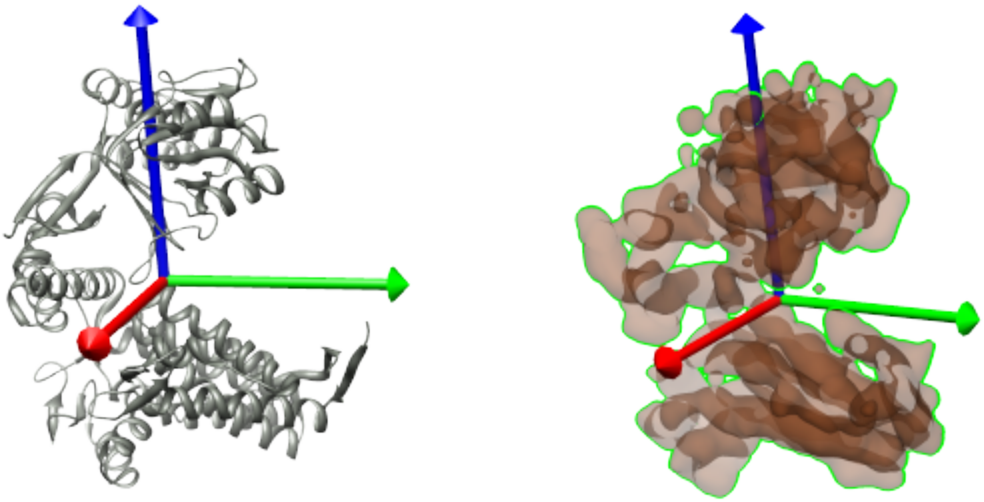
We can now align this structure to a segmented region in the density map. One way to do this is using the principal-axes transform. To illustrate this process:

- make sure the segmentation of GroEL is visible in the main Chimera window,
- select a single region corresponding to any one of the proteins
- choose **Show only selected** from the **Regions** menu in the **Segger** dialog (this will hide all other regions but the one selected)
- choose **Make transparent** from the **Regions** menu in the **Segger** dialog (this will let us see through the region's surface)
- choose **Show axes for selected** from the **Regions** menu in the **Segger** dialog

The image below shows what you should see upon doing all that:



Here are the images of the structure and the region side by side, with their principal axes shown:

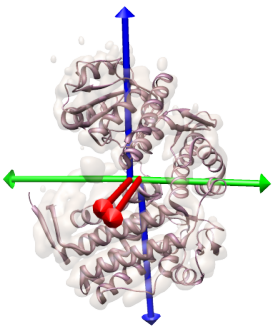


These two images show how the principal axes of the structure and of the region are roughly the same. We can use this to quickly align the structure to the region, by matching centers and principal axes. This is what we mean by *principal axes transform*. Once the alignment is done, the *Fit in Map* method implemented in Chimera is run to refine the alignment, an to produce the correct fit of the structure in the map.

To perform the alignment:

- make sure the structure is selected in the field to the right of *Structure to fit* in the **SegFit** dialog
- press the **Options** button in the same dialog
- enter **5** to the right of *Density map resolution:*
- press the **Fit** button at the bottom of the same dialog.

Shortly, the structure will have been moved to fit right into the region, as shown below:

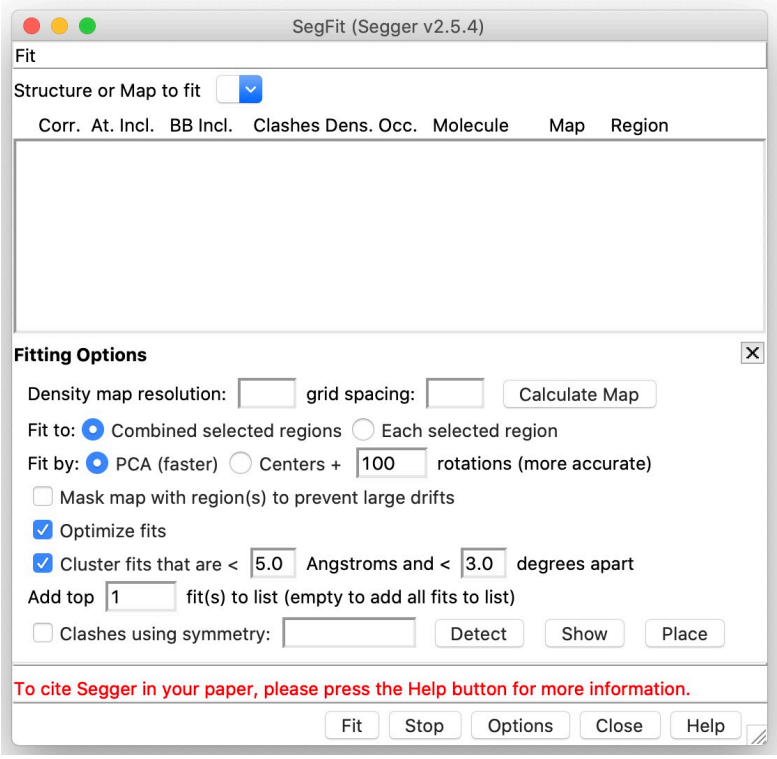


Notes:

- Showing the principal axes of the structure or the region is not required to complete the alignment process. They are computed automatically as needed. They are shown here for illustrative purposes.
- The principal axes in the image above are not pointing the same way. In fact, the principal axes are the eigenvectors of a covarience matrix. An eigenvector gives the direction of each axis, but the*signs* of these directions are ambiguous. When performing the alignment, the signs are flipped to generate 4 possible transforms. The alignment that gives the highest cross correlation is kept. Only non-reflecting transforms are considered, i.e. transforms in which either none or two of the three axes are flipped. Any other transform where odd number of axes are flipped result in reflections about one or more axes.
- The aligment process here only achieves a *rigid fit*. This assumes the structure of the molecule being fit should be the same in both the cryo-EM and crystallographic states for a good fit to be obtained. The latter may not always be true, for example some proteins may have different conformations under different conditions. In such cases, a flexible-fitting method, such as Direx orMDFF should be used.

Fitting Options

The **SegFit** dialog after the **Options** button is pressed is shown below:



A description of the options, going in order from top to bottom:

Density map resolution and grid spacing

- During the fitting process, a density map is generated for the structure. This density map is used to compute a cross-correlation score for the fit. After the structure is aligned to one or more region, the structure and its density map are further moved so as to increase the cross-correlation score (the same thing that Volume Data / Fit in Map interface does). The density map of the structure should be generated at a resolution that is approximately the same as the reported resolution for the density map into which the structure is being fit, and roughly the same grid spacing. Before pressing the **Fit** button, enter this resolution (and optionally a grid spacing) in these fields. Note that by default, the resolution and grid spacing are set to 3**g* and *g* respectively, where *g* is the grid spacing in the map selected in the Segger dialog.

Fit to:

- *Combined selected regions:* This is the default, and in this mode, the structure is aligned to the selected region. If more than one regions are selected, the structure is aligned to all the selected regions combined.
- *Each selected region:* The structure is aligned to each selected region. This means that more than one alignment will be made, and multiple fits will result. (To see all fits, make sure to increase the number in the box next to "Add top ____ fit(s) to list", or leave it empty to add all fits). Note that if no regions are selected, and this mode is chosen, then the structure is aligned to each region in the current segmentation. This is useful when, much like in this example using GroEL, all proteins in the complex have approximately the same structure.

Fit by:

- *PCA:* This alignment method uses the principal axes as described above and is quite fast, especially compared to an alternative, exhaustive search. However it may not always find the right fit. It is likely to fail for structures that have shapes for which the principal axes are completely ambiguous (e.g. sphere, rod, or cube-like shapes).
- *Centers + [N] rotations:* Use this method when you want to make sure you find the best fit. With this method, the structure is first placed so its center is the same as that of the selected region(s). Then it is rotated through N evenly spaced orientations. You can enter any value for N (it is 100 by default). 100 should be adequate for very thorough searches, but you can try a smaller number if you are in a rush as that will complete faster. A higher N may be needed at higher resolutions, as there are more local minima to get stuck in.
    - Note: This process is more like an exhaustive search, except only orientations (3 degrees of freedom) are searched through, rather than both position and orientation (6 degrees of freedom) as in exhaustive search.

Mask map with region to prevent large drifts

- When this option is checked, before fitting, the entire map is masked with the selected region(s), so that voxels except the regions in the selected region(s) become 0. This is useful if 'Optimize fits' is enabled but large drifts are noticed. Checking this option will prevent such drifts, and the structure will remain in the same area as the region.

Optimize fits

- When this option is checked, after generating each alignment to a region, the fit will be further optimized (i.e. the structure will be moved in the average gradient direction, until convergence is reached). When used, this process will typically produce better fits, but in lower resolution maps, it can result in large drifts. The entire fitting process will also be slower when this option is checked.

Cluster fits that are < [d] Angstroms and < [a] degrees apart

- During rotational search with say 100 different orientations, and if the Optimize fits button is checked, then it is very likely that some of the resulting fits after optimization are more or less the same. By the same, meaning that it would take only a small displacement or rotation to make them exactly the same. It's quite handy to cluster such fits together, especially if you will be looking at more than just the top fit (see next point). Check this option if you would like this to happen. You can also adjust the cluster criteria, i.e. how far (d) the fits should be apart and what angle (a) they must vary by before they are taken to be in different clusters.

Add top [N] fit(s) to list (empty to add all fits to list)

- Here you can choose how many of the fits tried (and clustered, optionally) should be added to the list. Once in the list, you can click through them to inspect each one, or compute statistics on them.

Clashes using symmetry

- When this option is checked, symmetric copies of the structure will be placed (if symmetry is detected in the map), and a clash score will be computed. The clash score is the fraction of atoms in the fitted structure that are within 3 Angstroms of any atom in a symmetric copy.
- You can press the **Detect** button to see if your map has any detectable symmetry (this is a Chimere built-in function), **Show** to apply the symmetry relationships to the currently selected fit, or **Place** to apply symmetry to the selected structure (with Place, the structure does not have to be fit first as with Show).

5. Aligning the structure to multiple regions

- Back to the GroEL example, make sure no regions are selected, and then make sure the field for "Add top ____ fit(s) to list" is empty (so that all fits will be added to the list).
- Also check the *Each selected region* under *Which regions to align the structure to*, and press the **Fit** button.

    - Under this mode, the structure is aligned to each selected region. But if no regions are selected, then the structure is aligned to each region.
- After pressing "Fit", you should see 16 entries in the list of fits.
- Select all the resulting fits by clicking on the first one, and then shift-click on the last, then choose "Place molecule copies" from the Fit menu at the top of the window.
- The results are shown in the image below: