Tutorial on flexible fitting with Flex-EM

Local rigid fitting methods do not fully utilize the information in the cryoEM map. Often, the conformation of the atomic structure may differ from the conformation represented by the map. Also, if a homology model is used for the fitting, the model may suffer from loop distortions, movement of secondary structure elements, or other errors introduced in the comparative modelling procedure itself (e.g. incorrect alignment). These problems can be overcome by using a flexible fitting method, whereby the protein structure is optimized simultaneously using the fit into the EM map and the regular scoring function incorporating stereochemistry and nonbonded interactions.

For this tutorial, a homology model of *e-coli* adenylate kinase (generated by Modeller using a homologous structure, PDB code: 1DVR:B) is refined within its density map. The map was generated from adenylate kinase crystal structure (PDB code: 1AKE:A) with Chimera *molmap* command (using *sigmaFactor*=0.225). To improve the fit between the model and the density map we will run the Flex-EM program, using the Modeller.

Download the files needed for the tutorial from:

http://topf-group.ismb.lon.ac.uk/flex-em/data.tar.gz

For this part of the tutorial you should be familiar with the basic features of the Chimera or any other molecular visualization software that allows you to display density maps. We will use Chimera for demonstration.

Structure Chimera and open **1akeA_10A.mrc** and **mdl1.pdb**.

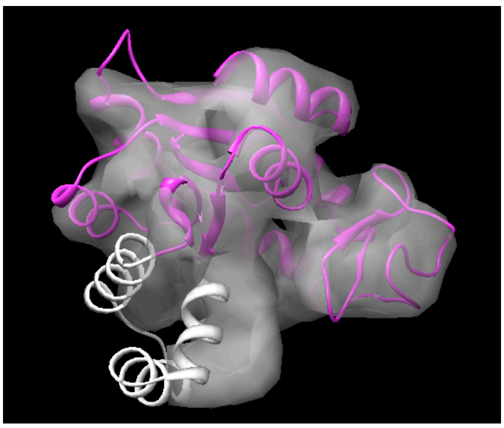
You can see in the Volume Viewer dialog that the size of 1ake_10A.map is 22^3 voxels where each voxel size is 3 Å/pixel. The resolution of this map is 10 Å. The origin index should be 2.1647 -4.4603 1.8033.

mdl1.pdb is the homology model of *e-coli* adenylate kinase, already fitted in the map. To get a clearer view of the fit change the display to ribbons and make the map transparent.

From visual inspection you can see that the fit can be improved. For example, colour residues 31-74 by typing the command

```
color white #1:31-74
```

in the entry field at the bottom of the Chimera Graphics Window and press Enter. You can see that the residues in white lie outside the density.



To improve the fit will run Flex-EM using Modeller(9v7). In the current application of Flex-EM we will employ 2 iterations of simulated annealing molecular dynamics (MD) optimisation. We first have to edit the control file **flex-em.py.** Open the file by typing:

```
nedit flex-em.py
```

Edit the parameters below *INPUT PARAMETERS*:

1. Define the mode of optimization:

    optimization = 'MD'.

2. Set the input parameters of the atomic structure that you want to fit:

    input_pdb_file='mdl1.pdb'

3. Edit the EM map parameters:

```
em_map_file = '1akeA_10A.mrc'       # name of EM density map (mrc)
format='MRC'                        # map format: MRC or XPLOR
apix=3                              # voxel size: A/pixel
box_size=22                         # size of the density map (cubic)
resolution=10.0                     # resolution
x= -6.494; y=13.381; z=-5.410       # origin of the map (in Å)
```

4. Specify the directory in which you want the results to be found:

    init_dir = 1

This will produce the results in a directory called **1_md**.

5. Specify the number of simulated annealing iterations:

    num_of_iter = 2

and close flex-em.py.

In order to reduce the conformational space that has to be searched during this procedure, groups of atoms are defined and moved as rigid bodies (*e.g.*, domains, sub-domains, secondary structure elements). This is done by editing the file **rigid.txt**. Open the file by typing

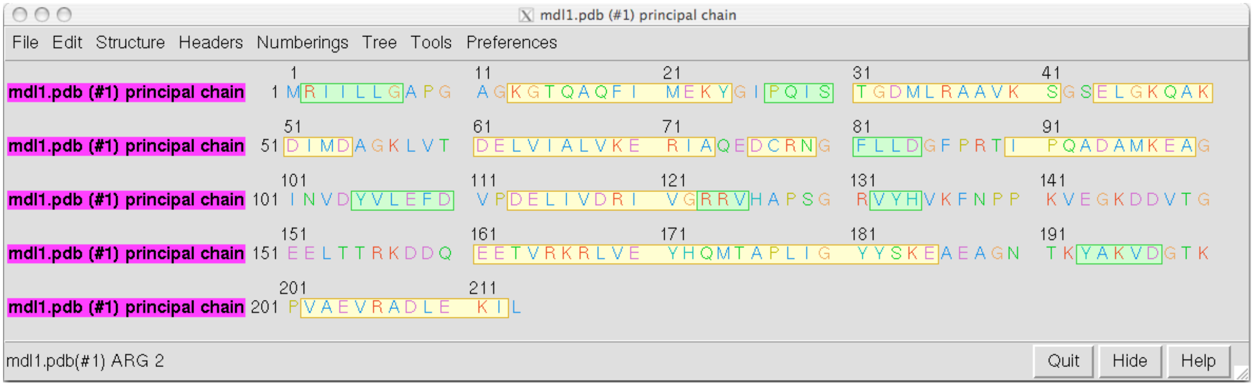
```
nedit rigid.txt
```

The file uses the following format:

- Comment lines begin with '#' (*e.g.*, describing the rigid body: '#domain', '#helix', '#beta').

- Other lines: each line describes one rigid body by specifying the initial and final residue of each of the segments in that rigid body (*e.g.*, '2 6 28 30' means that residues 2-6 and 28-30 will be included in the same rigid body).

In this tutorial we use the secondary structure elements of mdl1.pdb as rigid bodies. This level of description is typically suitable for a map of 10 Å resolution. To obtain those elements you could use Chimera. Open the Model Panel dialog and select mdl1.pdb by clicking on the *ID* column next to it (#1). Click the *sequence* bar on the side menu. A new panel corresponding to the sequence of mdl1.pdb will appear. The secondary structure elements are indicated on top of the sequence – beta strands are coloured green and alpha helices are coloured yellow (based on the PDB file or on DSSP (Kabsch & Sander, Biopolymers 1983). Placing the left mouse on top of an amino-acid letter will show its corresponding number at the bottom left of the sequence panel.



You probably noticed that we already edited the file using the secondary structure elements of mdl1.pdb as rigid bodies. Check that the numbers are correct.

Run the program by typing

```
mod9v7 flex-em.py > flex-em.log &
```

or:

```
modpy.sh python flex-em.py > flex-em.log &
```

(modpy.sh is in the Modeller /bin directory).

Apart from the output file **flex-em.log**, the program generates a number of files in the **1_md** directory. After each iteration of simulated annealing a pdb file with the latest refined coordinates is generated (*e.g.*, **md1_1.pdb** is generated after one cycle of simulated annealing). You can open these files in Chimera to see the progression of the optimisation. On completion, the program generates a file (**final1_mdcg.pdb**) containing the final structure that has been refined by flexible fitting into the map.

To look at the change in CC during the optimization you can type the following command:

```
grep "Mod-EM" flex-em.log | awk '{printf "%7.4f\n", $7}' | more
```

(You can press *ctrl C* to stop this command). When the optimisation is finished you could look directly at the final CC value by typing:

```
grep "Mod-EM" flex-em.log | awk '{printf "%7.4f\n", $7}' | tail -1
```

Report the initial and final values of the CC.

Open the final structure in Chimera **(final_mdcg.pdb)** and change the display to ribbons**.** By visual inspection of the initial model (mdl1.pdb, in magenta) and the final model (final1_mdcg.pdb, in cyan), it is clear that several secondary structure elements have moved towards the density.

The level of refinement can be understood via comparison to the native structure (which in this case is already known). We will compare both models (initial and final) with the native structure using the Cα Root Mean Square Deviation (Cα RMSD) measure.

Open the native structure **1akeA.pdb** in Chimera and change it into ribbon representation**.** To calculate the Cα RMSD between the initial model and the native structure type the command:

```
rmsd #1:1-213@ca #3:1-213@ca
```

in the Chimera command line.

You can see that the Cα RMSD (shown in the bottom left of Chimera Graphics Window) is ~4.5 Å. Now check the RMSD to the native structure of your refined model by typing:

```
rmsd #2:1-213@ca #3:1-213@ca
```

Using the density information, the Cα RMSD of the model from the native structure has been reduced from ~4.5 Å to ~2.3 Å.

Useful Links:

Chimera: http://www.cgl.ucsf.edu/chimera/

Modeller: http://www.salilab.org/modeller/

Flex-EM:

http://salilab.org/Flex-EM/

http://topf-group.ismb.lon.ac.uk/flex-em/cluster_fits.html