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 10 Å resolution native density map. The map was generated from adenylate kinase crystal structure (PDB code: 1AKE:A) using 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 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.

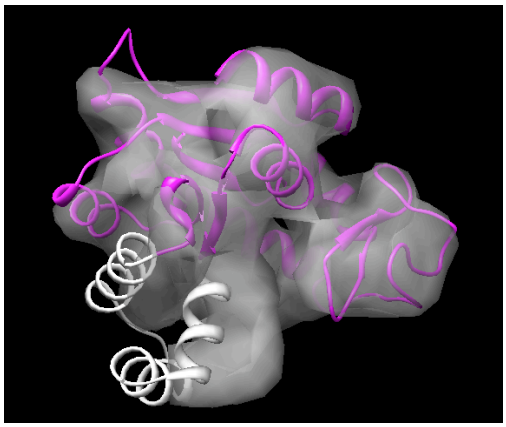
Start Chimera and open **1akeA_10A.mrc** and **mdl1.pdb**.

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.



1. Flexible fitting using secondary structure elements as rigid bodies

To improve the fit will run Flex-EM using MODELLER(version 9 and above). In the current application of Flex-EM we will employ only 2 iterations of simulated annealing molecular dynamics (MD) refinement. 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 (note that the origin is in Å):

```
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 (for the purpose of this tutorial we are going to do only 2 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

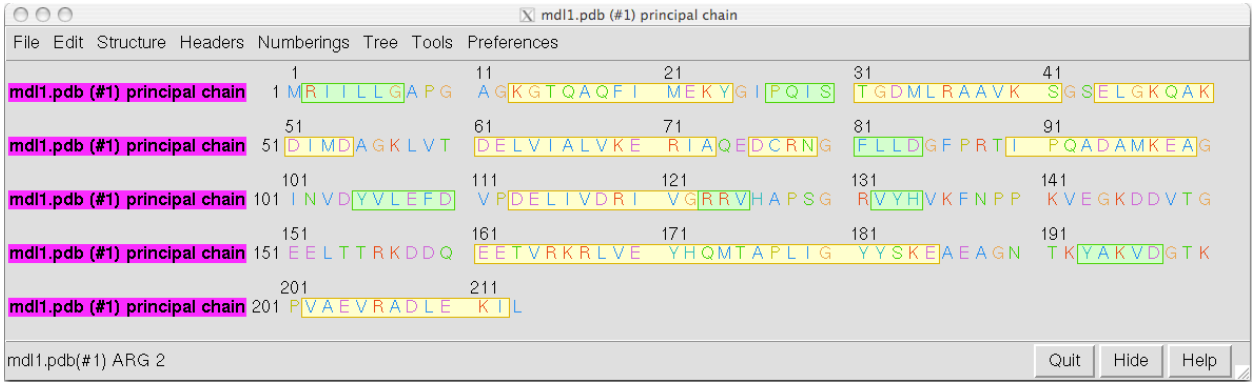
```
nedit rigid.txt
```

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

```
mod9.14 flex-em.py > flex-em.log &
```

Note that there are different versions of MODELLER so you might need to run a different version from mod9.14 depending on the version installed on your machine

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. You can also save the CC values into a text file

and open it in Excel in order to see the convergence of the score.

Open **1_md/final1_mdcg.pdb** and change it into ribbon representation. 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 from 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.

2. Hierarchical flexible fitting with Flex-EM/RIBFIND

It is possible to reduce 'overfitting' by first running Flex-EM with coarser rigid bodies and then "release" by running it with rigid bodies defined as secondary-structure elements (as above). To get the larger rigid bodies you can use the RIBFIND server: http://ribfind.ismb.lon.ac.uk.

You can start the refinement first using the rigid bodies downloaded from RIBFIND server (with the set that contains the maximal number of rigid bodies, in this case 2). Click on move the *cutoff value* to 11 and then "*Download rigid body file*".

Call the file **rigid_ribfind.txt** not to confuse it with **rigid.txt**. The rigid bodies should look like this:

```
#RIBFIND_clusters: 6.5 14.0 2 8 10 2.8
13 24 202 212 2 6 28 30 81 84 105 110 193 197 161 185 7 12 25 27 186
192 198 201
31 41 44 54 61 73 90 95 42 43 55 60
#individual_rigid_bodies 2
113 119
123 126 132 134 153 154
```

You can select the rigid bodies in chimera in order to visualise them on **mdl1.pdb** by typing the following in the Chimera *command line*:

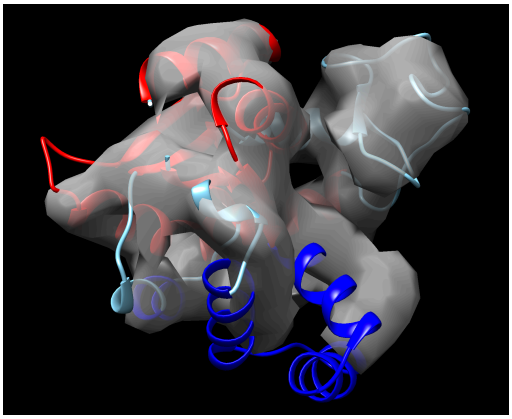
```
color red #1:13-24,202-212,2-6,28-30,81-84,105-110,193-197,161-185,7-
12,25-27,186-192,198-201
```

Press *Enter*. Then type:

```
color blue #1:31-41,44-54,61-73,90-95,42-43,55-60
```

and press *Enter* again.

The two large rigid bodies should be coloured red and blue respectively.



Copy **flex-em.py** to **flex-em2.py** by typing the command:

```
cp flex-em.py flex-em2.py
```

In the new input file (**flex-em2.py**) edit the following lines:

init_dir = 2

rigid_filename = 'rigid_ribfind.txt'

Re-run Flex-EM by typing

```
mod9.14 flex-em2.py > flex-em2.log &
```

Once this run is finished open the refined structure (2_md/**md2_2.pdb**) in Chimera. Calculate the Cα RMSD of your refined model from the native structure. Has the model improved? Report the RMSD value.

To further improve the result use this model as an initial model for further refinement using the rigid bodies as in part 1 of the tutorial (**rigid.txt**). First copy **flex-em.py** to **flex-em3.py** and **md2_2.pdb** from **2_md/** to the location where flex-em3.py is present. Then edit **flex-em3.py:**

input_pdb_file= 'md2_2.pdb'

init_dir = 3

After two iterations of Flex-EM refinement note that the overall Cα RMSD between the final model (3_md/**final3_mdcg.pdb)** and the native structure is similar to the initial refined structure (without using RIBFIND rigid bodies). However, if you calculate the Cα RMSD of the individual helices, you will find that the RMSD for helix 90-95 has reduced significantly.

Links:

RIBFIND software and documentation: http://ribfind.ismb.lon.ac.uk

Flex-EM software and documentation: http://topf-group.ismb.lon.ac.uk/flex-em/

MODELLER software, documentation and tutorials can be found on: http://www.salilab.org/modeller/

References:

Pandurangan AP, Topf M. RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps. *Bioinformatics* 28:2391-3, 2012.

Pandurangan AP, Topf M. Finding rigid bodies in protein structures: Application to flexible fitting into cryoEM maps. *J Struct Biol.* 177:520-531, 2012.

Pandurangan AP, Shakeel S, Butcher SJ, Topf M Combined approaches to flexible fitting and assessment in virus capsids undergoing conformational change. *J Struct Biol* 185:427–439, 2014.

Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A. Protein structure fitting and refinement guided by CryoEM density. *Structure* 16:295-307, 2008.

N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.