

# Manual for FLEA

## FFLD Leader Clustering

### Version 1.0

Fabian Dey and Amedeo Caffisch

To improve this documentation, please send comments and feedback to:

Amedeo Caffisch

E-mail: [caffisch@bioc.unizh.ch](mailto:caffisch@bioc.unizh.ch)

FAX: (+41 44) 635 68 62

## 1 Function

Clusters poses generated by FFLD using the leader algorithm. The poses are read from a multi-PDB file generated by FFLD, the corresponding energies are extracted from a single column table with the corresponding energy values. The total number of poses present is evaluated by the number of energy entries in the user supplied list. The number of atoms per pose is extracted from the template file. These two values are then used to read the PDB file containing the poses. The poses are then sorted according to energy and clustered with the leader algorithm, where the similarity between two poses is measured with :

$$sim(AB) = \frac{S_{AB}}{\max(S_{AA}, S_{BB})} \quad \text{with} \quad (1)$$

$$sim(AB) = [0, 1] \quad \text{and} \quad (2)$$

$$S_{XY} = \sum_i^{n_x} \sum_j^{n_y} w_{ij} e^{-\gamma r_{ij}} \quad (3)$$

where  $n_x, n_y$  are the number of atoms of  $X$  and  $Y$ ,  $r_{ij}$  is the interatomic distance in Å,  $\gamma$  is a user-defined coefficient and  $w_{ij}$  is a unit element matrix. The similarity is not

measured for atoms with element type “1” (see 4).

The clustered poses are written to disk in either the PDB- or the MOL2-format with the pose energy stored in the header of the corresponding file.

## 2 Outcome

There are three different possibilities of how FLEA can finished clustering:

- All poses are clustered.
- FLEA stops the clustering procedure if the maximal number of clusters has been generated.
- FLEA stops the clustering procedure if a pose is encountered in the energy sorted list which has a higher energy than the user-defined cutoff.

## 3 Input parameters

- **-n / --numb**  
Defines the maximal number of sorted poses that should be clustered.
- **-t / --template**  
Template MOL2-file which is used to write out MOL2-files and to build the element matrix. The elements are derived with the help of the parameter file (see 4).
- **-e / --energyfile**  
Single column list of all the energy values of the poses, which is needed to sort the poses according to energy.
- **-c / --cutoff**  
Clustering cutoff above which two poses are considered “similar”.
- **-r / --requested**  
Defines the maximal number of clusters to be produced.
- **--expfactor**  
The exponential factor  $\gamma$  needed for the computation of the similarity (see equation 3).
- **--outputname**  
Prefix of the files written to disk by FLEA :  
[prefix]\_clus#.mol2 / \_clus#.pdb
- **--output**  
Output format (“pdb” or “mol2”).
- **--energycutoff**  
Poses with a energy higher than the value specified are ignored in the clustering step.

## 4 Parameter file

The first line denotes the number of lines in which atom types are stored. The following lines denote the atom type and the corresponding element as an integer number. Atom types with the element set to “1” are not considered for the similarity evaluation, which provides the user with the possibility to specify how the similarity is measured between two poses (e.g. omitting non-polar hydrogens).

110	
B	5
C	6
C3	6
C4	6
C5R	6
C5RE	6
.	
.	
.	