

MULTI DOMAIN ARRANGEMENT

Version 1.0 created by Björn Hellenkamp

Requirements

The program runs either with Matlab R2016b (or higher) or with Matlab runtime R2016b.

The runtime version is freely available (means that you don't need any Matlab license):

<https://de.mathworks.com/products/compiler/mcr/>

PyMol Version 1.5 or higher should be installed to visualize the output structures (pml-files).

Example data

Arranging two domains of Hsp90

In this first example the following two domains of Hsp90 (PDB structures) will be arranged:

- 2cg9_MNA.pdb (Prearranged M-Domain and N-domain of Chain A)
- 2cg9_MNB.pdb (Prearranged M-Domain and N-domain of Chain B)

Therefore, the following parameter file and position file were prepared:

- OpenHsp90_MM.par.txt (see below for table description)
- OpenHsp90_MM.pos.txt (see below for table description)

First, a distance file (*.dis.txt) is generated by fitting the pre-corrected FRET efficiency histograms with a probability distribution analysis. The parameter file (*.par.txt) indicates required measurement parameters for the fit. The position file (*.pos.txt) assigns a position number to the aminoacid number of the protein, to the domain number and to the monomer number in case of multimeric proteins (the x,y,z-coordinates of the positions are not used here, but for the following arrangement).

Then, different arrangement procedures can be executed (**see Fig.1**) based on the following position file, which contains the average positions of the dye accessible volumes, and the generated distance file, which contains the network of average distances and distance fluctuations:

- OpenHsp90_MM.pos.txt (see below for table description)
- OpenHsp90_MM.dis.txt (see below for table description)

Arranging three domains of Hsp90

In this second example the following three domains of Hsp90 (PDB structures) will be arranged:

- 2cg9_MNA.pdb (Prearranged M-Domain and N-domain of Chain A)
- 2cg9_MNB.pdb (Prearranged M-Domain and N-domain of Chain B)
- 2cg9_CC.pdb (C-Domains Chain A and Chain B)

Therefore, the following parameter file and position file were prepared:

- OpenHsp90_MMCC.par.txt (see below for table description)
- OpenHsp90_MMCC.pos.txt (see below for table description)

First, a distance file (*.dis.txt) is generated by fitting the pre-corrected FRET efficiency histograms with a probability distribution analysis. The parameter file (*.par.txt) indicates required measurement parameters for the fit. The position file (*.pos.txt) assigns a position number to the aminoacid number of the protein, to the domain number and to the monomer number in case of multimeric proteins (the x,y,z-coordinates of the positions are not used here, but for the following arrangement).

Then, different arrangement procedures can be executed (**see Fig.1**) based on the following position file, which contains the average positions of the dye accessible volumes, and the generated distance file, which contains the network of average distances and distance fluctuations:

- OpenHsp90_MMCC.pos.txt (see below for table description)
- OpenHsp90_MMCC.dis.txt (see below for table description)

1. Select files required for 2.-5.

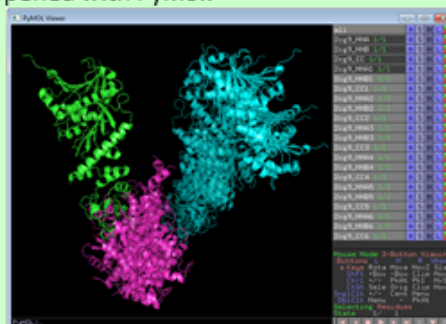
- A. Subfolder with the eff. histograms
- B. Parameter file with measurement parameters for each histogram
- C. Position file with label coordinates
- D. Distance file with average distances, standard deviations and errors

2. Fit efficiency histograms (requires A, B, C)

Function fits efficiency histograms specified in the parameter file assuming # states (Gaussian distance distributions) with s.d. between min. and max. Then, it generates a distance file with the obtained mean values and s.d. of the Gaussian distance distributions.

3. Average Arrangements for screening possible local minima (requires C, D)

Function spatially arranges domains structures that fulfill the network of average distances with a $\chi^2 < 1$. The resulting 3D structures are saved and can be directly opened with PyMol:



MDA

1. Select files

A. Histogram folder: histograms Open

B. Parameter file: OpenHsp90_MM.par.txt Open

C. Position file: OpenHsp90_MM.pos.txt Open

D. Distance file: OpenHsp90_MM.dist.txt Open

Refresh list

2. Fit efficiency histograms (multiple PDA) - requires A,B,C

Fit

states: 1 Min. s.d.: 0 Max. s.d.: 30

Distance output file: OpenHsp90_MM.dist.txt

3. Arrange average structures for a max. χ^2 - requires C,D

Average Arr. Max. χ^2 : 1 Open structures Open results

4. Arrange average structures from distance subsets - requires C,D

Error Arr. # reduced distances: 2 # subsets: 10 Open structures Open results

5. Arrange dynamic structures from Gaussian distances - requires C,D

Dynamic Arr. Open structures Open results

Dynamic Domain Arrangement - Version 1.0 - created by Björn Hellenkamp
Generating structures...
Successfully generated structures: minima.pml
Successfully generated structures: minima.results.txt

Message box

4. Error Arrangements for screening the distance error space (requires C, D)

Function generates # distance subsets each with # reduced distances. For each subset, the optimal domain arrangement is generated. The resulting 3D structures are saved.

5. Dynamic Arrangements for visualizing the structural fluctuations (requires C, D)

Function generates and saves an ensemble of 3D structures based on the network of Gaussian distance distributions.

Fig. 1: Follow the steps 1-5 for analyzing the structural ensemble and its fluctuations of a multi-domain protein. Complete sample data for Hsp90's open state is provided.

Table specification

B. Parameter table

Each row specifies the measurement of one FRET pair.

Position1:	Number of the first position of the FRET pair (the number belongs to the row number of the position table)
Position2:	Number of second first position of the FRET pair (the number belongs to the row number of the position table)
HistogramName:	Path to pre-corrected FRET efficiency histogram
FoersterRadius:	Förster radius of the FRET pair
Int.Threshold:	Intensity threshold that was used for the burst search
Gamma:	Gamma factor
DirectExc:	Acceptor direct excitation factor
Leakage:	Donor leakage factor (cross talk)
EffMin:	Minimum efficiency value used for the distance convolution
EffMax:	Maximum efficiency value used for the distance convolution
CombinedAniso:	Combined anisotropy used for error estimation
DeviationClosedState:	Deviation between the distance for the closed state and the respective x-ray structure used for error estimation

C. Position table

Each row specifies one labeling position.

DomainNr:	Number of the domain which the position belongs to
MonomerChain:	Number of the monomer chain which the position belongs to (for non multimeric proteins set all to 1)
PositionNr:	Number of the position (typically the amino acid number)
Mean.X:	x-coordinate of the mean position of the dye accessible volume (calculated with the FPS software: http://www.mpc.hhu.de/software/fps.html)
Mean.Y:	y-coordinate of the mean position of the dye accessible volume

Mean.Z: z-coordinate of the mean position of the dye accessible volume

PDBName: Name of the PDB file that belongs to the domain number

D. Distance table

Each row specifies one distance, its fluctuations (at a timescale of 1 ms) and errors.

Position1: Donor position numbered as in the position table

Position2: Acceptor position numbered as in the position table

Distance: Distance in Angstrom

DeviationRight: Distance deviation to the right

DeviationLeft: Distance deviation to the left (normally the same as the right deviation, as symmetric Gaussians are assumed)

ErrorRight: Squared sum of anisotropy error, transformed efficiency error and a relative error from the closed state

ErrorLeft: Squared sum of anisotropy error, transformed efficiency error and a relative error from the closed state