FiPD

Introduction

FiPD (**Fi**lter for **P**roteins **D**ocking) reads the output of a docking program, Zdock, and selects, from the thousands of solutions listed there, the ones which are compatible with (depending on the chosen options):

- docking of a membrane protein with another membrane protein, ie with the dimerization (and possibly polimerization) of membrane proteins. To achieve this a cutoff is imposed on the inclination that the "ligand" protein may have with respect to the original orientation and the shift along the main axis. The inclination criterium is based on the assumption that an excessive inclination would mean that the proteins is not in a suitable orientation to cross the membrane. The shift criterium is based on the hypothesis that it is derirable for the ligand protein not to "stick" out of the membrane much more than in the original position.

- a set of distances specified by the user. The distances between selected pairs of atoms in the target and the ligand are calculated and solutions accepted when all the distances are below specified tresholds.

FiPD can also pre-process the input pdb file in order to ensure the correct orientation of the starting structure with respect to the membrane (see below).

FiPD is released as statically-linked binaries for GNU-linux on i386 and X86-64 platforms. You can get it at:

`http://www.csbl.unimore.it/software/software.htm`

for problems or bugs you can email mseeber@unimore.it.

If you use it, please cite:

<div align="center">

Casciari D., Seeber M. and Fanelli F.;
Quaternary structure predictions of transmembrane proteins starting from the monomer: a docking-based approach;
BMC Bioinformatics (2006), 7:340

</div>

Requirements

In order to effectively enforce the criteria for membrane compatibility, two conditions must be met:

- the ligand in its starting position must be correctly superimposed on the target protein

- the ligand in its position must be aligned with the Z-axis, meaning the virtual membrane is in the XY plane

No particular preprocessing is required in case the distance selection aproach is used.

Method

Membrane Compatibility

The inclination is defined as the angle between the original Z-axis and the one resulting from the transformation listed in the docking output file.This is thus directly compared with the chosen cutoff.
In the zdock output only the second rotation (θ) modifies the inclination. This is thus directly compared with the chosen cutoff.
In the escher output both the first and the second rotation affect it. The angle θ to be compared with the cutoff is then calculated as:

$$\theta = arcos(cos(angle_x) * cos(angle_y))$$

On the other hand, the zoffset is the shift that is applied to the structure along the Z-axis. This value is listed in a single entry in both files, so the comparison in straightforward.

Membrane Alignment

The previous criterium is meaningful only if the starting, reference structure is properly aligned with the membrane, *ie* with the XY plane. Membrane topology predictions can be used to identify the interface residues. FiPD uses this information to align the protein. A multilinear regression is run over the coordinates of the interface atoms and the protein is then rotated to make these planes parallel to the XY plane.

Distance Constraints

The distance between two atoms is defined as the square root of the sum of the squares of the difference between the coordinates of the atoms themselves:

$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Clustering

Filtered solutions can still number in the hundreds, so, to help in spotting the best ones and avoid redundancy, FiPD can cluster the accepted poses. The clustering can be done with 2 different algorithms, *Hierarchical* and *Quality Threshold* (see details at `http://en.wikipedia.org/wiki/Data_clustering`). Both give more or less the same results, with performance being better in one way or the other depending on the system and the choosen cutoff - hierarchical is faster for sparse systems and/or low cutoff values (*ie* small clusters), qt is best for dense systems and/or high cutoffs (*ie* large clusters).

Usage

FiPD needs the output file from the docking program and the pdb of the starting structures of the ligand and target proteins. FiPD writes two log files, accepted.log and rejected.log, in which it lists the solutions and the criteria it used to accept/reject them. It also writes the pdbs resulting form the accepted trasformations.

Pdb Format - Warning

Attention must be paid to the fact that ZDock requires "irregular" pdbs, in which the last columns are modified in content and format with respect to the pdb standard (a pre-processing program in the ZDock package takes care of this modification). These pdbs should be avoided and "regular" ones used for FiPD filtering, namely the ones which the *zdock*-modified pdb are derived from. *Note: this is not strictly necessary when using the latest version of FiPD*

A brief help is obtained by typing at the command line the option -h.

```
****************************************************************
*  FiPD: a GPLed program to filter from a (z)docking output  *
*     suitable solutions for protein-protein interaction     *
****************************************************************
Command line arguments:
 -z  input file   (zdock  output analysis)
 -p  pdb list     (pdb list - needs target)
 -f  output file  (fipd output files basename)
 -0  -            (avoid solutions pdb output)
 -t  target       (overrides name in input file)
 -l  ligand       (overrides name in input file)
 -o  offset       (max z-offset,    default 5.0 A )
 -a  angle        (max angle with z, default 0.5rad)
 -d  distfile     (list of distances to monitor)
 -c  cutoff       (activates clustering module)
 -b  cutoff begin (activates multiple clustering module)
 -s  cutoff step  (activates multiple clustering module)
 -e  cutoff end   (activates multiple clustering module)
 -r  -            (forces use of qt algorithm in clustering)
 -x  out.xyz      (outputs xyz of CAs for external clustering)
 -i  file.pdb     (input pdb for membrane alignment)
 -m  file.pdb     (output pdb for membrane alignment)
 -1  sele1        (selection for membrane alignment or clustering)
 -2  sele2        (second (optional) selection for membrane alignment)

This is FiPD, v. 0.1m - http://www.csbl.unimore.it/software/software.htm
```

Options:

- -z : starts zdock output analysis module. This filters zdock solutions based on membrane topology or distances criteria

- -p : starts filtering on a number of pdb listed in the given file (one pdb per line)

- -f : when filtering zdock solutions, fipd generates the pdb of the accepted solutions: this options allows to specify the basename of these pdb files - a progressive number is added.

- -t : the name of the target (receptor) pdb file. This is necessary with the -p option, and overrides what is written in the zdock output when used with the -z option

4

- -l : this overrides what is written in the zdock output when used with the -z option and is ignored with the -p option.

- -o : the offset threshold (see paper 1 for details)

- -a : the angle threshold (see paper 1 for details)

- -d : filtering the solutions on the basis of a set of distances (see paper 2 for details), this option gives the name of a file with the list of distances to monitor and their thresholds - see below for selection sintax

- -c : activates clustering of filtered solutions with the given cutoff; default algorithm in hierarchical

- -b : simultaneous multiple clusterings can be run with a range of cutoffs; this option gives the lower value

- -s : in multiple clustering, this gives the step for the cutoff range

- -e : in multiple clusteringm this gives the upper value

- -r : in clustering, this forces the use of the Quality Threshold algorithm

- -x : this asks fipd to output a concatenated xyz file for further analysis by external programs

- -i : starts preprocessing module that align the given pdb to the XY plane according to membrane topology predictions (these must be run separately and fed to fipd)

- -m : in the alignment module, the name of the output (modified) pdb file

- -1 : the selection string for the atoms predicted to be at the first water-membrane interface (necessary for -i option) - see below for selection syntax

- -2 : the selection string for the atoms predicted to be at the second water-membrane interface (one is sufficient, but 2 gives better results)

Options -z, -p and -i are mutually exclusive

Selection Syntax

The selection routine employs a string structured as follows:
/segname/resnumber/atomtype

Wild cards such as * (any number of any character), ? (any single character), [abc] (any single character among a, b and c) and [!abc] (any single character except a, b and c) are supported.

/*/*/CA	⇒ all alpha carbons
/MOL1/*/*	⇒ all atoms in segment MOL1
/MOL?/**	⇒ all atoms in MOL1, MOL2, MOLA etc.
/*/*/C[AB]	⇒ all alpha and beta carbons

Moreover, it supports ksh-style pattern matching, such as:

?(pattern-list)	⇒ The pattern matches if zero or one occurrences of any of the patterns in the pattern-list allow matching the input string.
(pattern-list)	⇒ The pattern matches if zero or more occurrences of any of the patterns in the pattern-list allow matching the input string.
+(pattern-list)	⇒ The pattern matches if one or more occurrences of any of the patterns in the pattern-list allow matching the input string.
@(pattern-list)	⇒ The pattern matches if exactly one occurrence of any of the patterns in the pattern-list allows matching the input string.
!(pattern-list)	⇒ The pattern matches if the input string cannot be matched with any of the patterns in the pattern-list.

Where pattern-list is a — separated list of patterns. A dash can be used to indicate a range of values in the residue number

/*/*/@(CA|C|N)	⇒ backbone atoms
/*/*/!(CA|C|N|O|H|OT1|OT2)	⇒ non-backbone atoms
/MOL1/@(1|3|5)/*	⇒ residues 1, 3 and 5 of segment MOL1
/MOL1/@(1-5)/*	⇒ residues 1 to 5 of segment MOL1
/MOL1/@(1-5—10)/*	⇒ redisued 1 to 5 and 10 of segment MOL1

Selection for distance constraints are structured as follows:
/segname/resnumber/atomtype : /segname/resnumber/atomtype : threshold
Each string (/.../.../...) must select one and only one atom; the distance constraints are listed in a file, one per line.
Selections for membrane alignment are given on the command line, protected by "'" and should select at least three atoms per selection.
fipd.x -i file_in.pdb -m file_out.pdb -1 "/.../.../..."