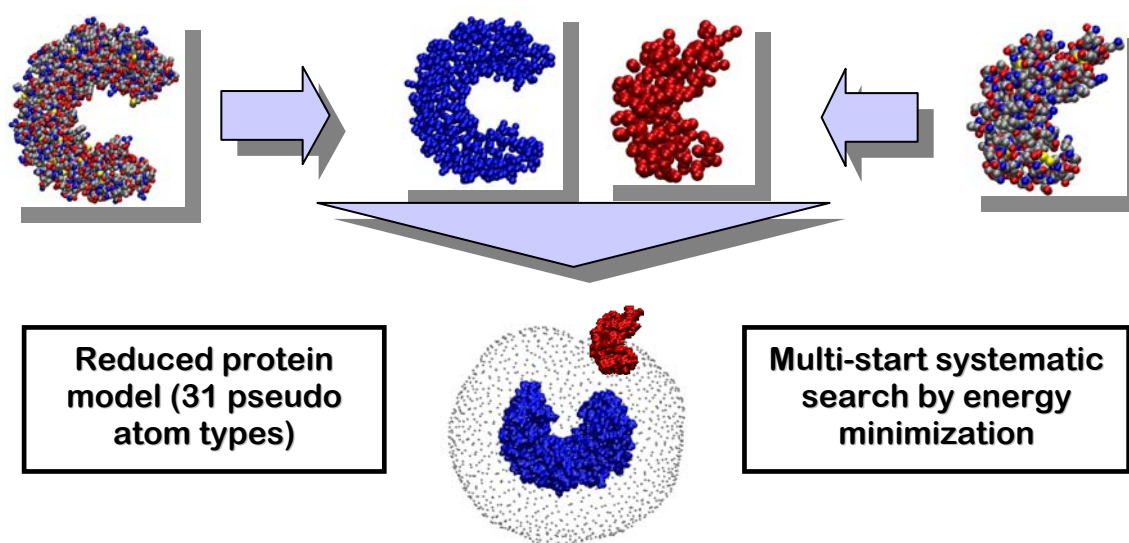


EMBO Practical Course
Docking Predictions of Protein-Protein Interaction
Barcelona, Oct. 14-17, 2008

**Protein-protein docking with *Attract* using a
reduced protein model**

Martin Zacharias, Jacobs University Bremen, School of Engineering and Science,
Campus Ring 6, D-28759 Bremen, e-mail: m.zacharias@jacobs-university.de



Content:

1.	Protein-protein docking by energy minimization	2
2.	The reduced protein model in <i>Attract</i>	2
3.	The <i>Attract</i> program	4
4.	Auxiliary programs to setup docking searches and to analyse results	7
5.	Example 1: Rigid docking of an enzyme-inhibitor complex: Comparison of docking bound and unbound structures	8
6.	Example 2: Inclusion of side chain rotamer copies during docking	9
7.	Normal mode calculation of proteins using an elastic network mode	10
8.	Example 3: Docking of Taxi-Inhibitor to Xylanase: accounting for global changes during docking	11
9.	Example 4: Docking of a protein fragment to HIV-capsid protein	12
10.	Example 5: Docking of a helical peptide to a Helix-binding domain of SRP (signal recognition particle)	13

1. Protein-protein docking by energy minimization

The protein-protein docking program ATTRACT (1-6) employs energy minimization in rotational and translational degrees of freedom (+ potential conformational variables) of one protein partner (ligand) with respect to the second protein (receptor). Flexibility of the partner structures can be taken into account by representing flexible surface side chains (and also loops) as multiple conformational copies. In case of side chains the conformational copies represent possible rotameric states of the surface side chains. The best fitting conformational copies are automatically selected during docking simultaneously with the minimization in translational and rotational degrees of freedom of the protein partners. Global flexibility of the binding partners can be included approximately by energy minimization in pre-calculated normal mode directions of the binding partners (3,6).

For systematic docking studies one of the proteins (usually the smaller protein, called the ligand protein) is used as probe and placed at various positions on the surface of the second fixed (receptor) protein. To select regularly spaced starting points a probe radius that is slightly larger than the maximum distance of any receptor atom from the ligand center is used. At each starting position on the receptor protein various initial ligand protein orientations are generated. The docking from each start position consists of a series of energy minimizations of the ligand protein with respect to the receptor protein. During the first minimization a harmonic restrain between the center of the fixed protein and the closest C α -pseudo atom of the ligand protein can be applied. This first minimization serves to generate a close contact between the two proteins. For the subsequent energy minimizations the ligand protein is typically free to move to the closest energy minimum. Approximately 10000-15000 complexes (in case of medium-sized protein partners with < 150 residues) can be energy-minimized to low residual gradients in about 1 h on a 2 GHz Linux PC.

2. The reduced protein model in Attract

The Attract docking minimization employs a reduced protein model which is intermediated between a residue-based representation and full atomic resolution. Each residue is represented by up to 4 pseudo atoms (2 for the backbone and up to 2 for each side chain) allowing to approximately account for the dual character of some amino acid side chains (e.g. hydrophobic and hydrophilic parts of a side chain). Small amino acid side chains (Ala, Asp, Asn, Ser, Thr, Val, Pro) are represented by one pseudo atom (geometric mean of side chain heavy atoms) whereas larger and more flexible side chains are represented by two pseudo atoms (for details, see reference 1).

The repulsive and attractive LJ-parameters describe approximately the size and physico-chemical character of the side chain chemical groups. Systematic tests of the model on “bound” protein partners indicates that rigid-body-minimization of the

experimental complex structures yields energy-minimized complex structures with an Rmsd (root mean square deviation) of the ligand protein from the experimental position of $\sim 1\text{-}2 \text{ \AA}$ [1-2] which is comparable to energy minimization using atomistic models.

Note, in contrast to a previous force field version the backbone is represented by 2 pseudo atoms located at the position of the N-atom and O-atom of the backbone carbonyl group. With this representation a more realistic description of the polar character of the protein backbone is possible. The parameters have been improved by optimizing the ranking of near-native solution with respect to non-native decoy complexes (unpublished result). The reduced representation is otherwise identical to the previous version. The energy function consists of pair-wise soft Lennard-Jones type functions and an electrostatic interaction term with a distance dependent dielectric constant ($\epsilon(r)=15r$) for the interaction of charged residues. As illustrated in Figure 1 the scoring function differs from a standard Lennard-Jones-type function in that it contains a saddle point instead of an energy minimum for certain types of pseudo atom pairs (those that are repulsive).

1. M. Zacharias (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* 12, 1271-1282.
2. M. Zacharias (2005) ATTRACT: Protein-Protein Docking in CAPRI Using a Reduced Protein Model. *Proteins* 60, 252-256.
3. A. May, and M. Zacharias (2005) Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochem. Biophys. Acta*, 1754, 225-231.
4. K. Bastard, Prevost, C., and M. Zacharias (2006) Accounting for loop flexibility during protein-protein docking. *Proteins*, 62, 956-969.
5. A. May and M. Zacharias (2008) Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins*, 69, 774-780.
6. A. May and M. Zacharias (2008) Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins*, 70, 794-809.

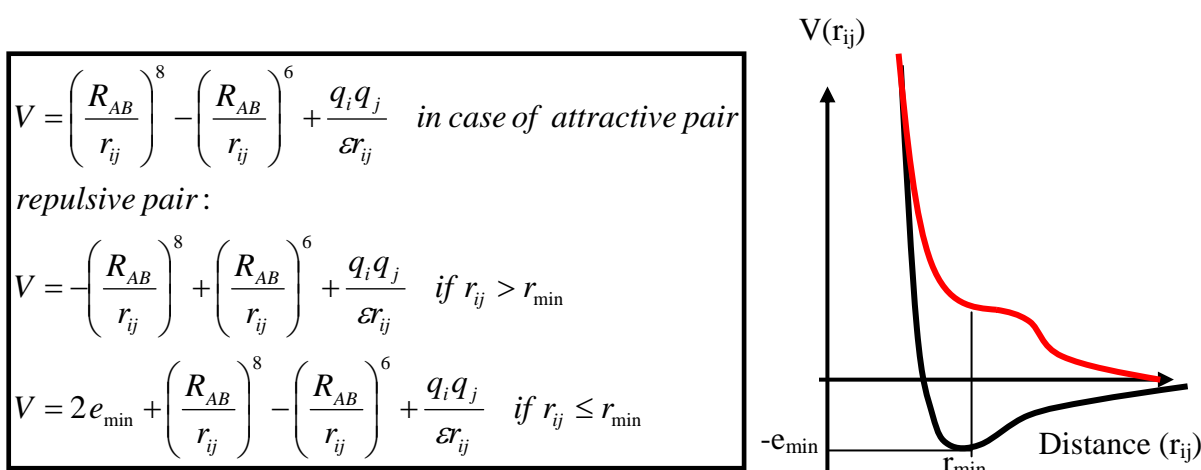


Figure 1: Docking scoring function in Attract. In case of an attractive pair (black line) an r^{-8}/r^{-6} -Lennard-Jones-type potential is used. For a repulsive pair (red curve) the energy minimum is replaced by a saddle point.

3. The Attract program

The Attract program exists as a version written in Fortran and a version written in Python/C++ (ptools C++ library, can be used a Python module: to be published). Both versions use the same input files and input formats. During the course we will work with a Fortran version of the program.

Two separate executables can be generated. "attractsingl" is for docking minimization of single complexes. The executable is located in \$path/bins. "attractmulti" can be used to perform a systematic docking search (located in \$path/bin). The variable \$path indicates the location of the Attract main directory.

Both programs require a receptor and a ligand structure in reduced representation, an input file (called attract.inp) and a parameter file (parmu.par). The receptor and ligand structures files are in a pdb type format and have the following format:

The reduced model protein structures can be easily generated from atomic resolution standard pdb-files using the \$path/bin/reduce program:

\$path/bin/reduce file.pdb

This generates a reduced model pdb-file termed:

filer.pdb (note the "r" inserted between "file" and ".pdb" to mark the file as containing a reduced representation of the protein). The format is an extended pdb-format with additional columns for pseudo-atom type, charge, conformational copy flag and re-weighting of interactions, respectively (example given below):

ATOM	1	N	LEU	1	27.879	23.918	-15.453	30	0.000	0	1.00
ATOM	2	CA	LEU	1	28.021	22.462	-15.140	32	0.000	0	1.00
ATOM	3	C	LEU	1	29.269	22.227	-14.272	32	0.000	0	1.00
ATOM	4	O	LEU	1	30.225	22.993	-14.322	31	0.000	0	1.00
ATOM	5	CSE	LEU	1	27.140	21.439	-16.720	15	0.000	0	1.00
ATOM	6	N	VAL	2	32.859	20.518	-13.086	30	0.000	0	1.00
ATOM	7	CA	VAL	2	34.086	20.379	-13.831	32	0.000	0	1.00
ATOM	8	C	VAL	2	34.777	19.035	-13.555	32	0.000	0	1.00
ATOM	9	O	VAL	2	34.993	18.666	-12.397	31	0.000	0	1.00
ATOM	10	CSE	VAL	2	34.960	21.533	-13.792	29	0.000	0	1.00

The "attract.inp" file has the following format:

```
5      0      0
73.98360 -53.41780 -13.62360 0.00050
30 2 1 1 0 0 0 0 1 2500.00
30 2 1 1 0 0 0 0 1 1500.00
40 2 1 1 0 0 0 0 1 100.00
60 2 1 1 0 0 0 0 0 50.00
60 2 1 1 0 0 0 0 0 50.00
```

The first row in the input indicates the number of successive minimizations (5 in the case above), the two 0s on the first line indicate that no soft modes for receptor or ligand are used

Second row: restraining coordinates for pushing the ligand on the surface of the protein (usually the center coordinates of the receptor protein), the fourth term is the force constant for the restraining potential (should not be larger than 0.001 RT/Å²)

The next 5 lines indicate the minimization conditions for each of the five docking minimizations (the number of lines must equal the number of minimizations chosen in the first line)

1. term: number of EM steps
2. term: minimization method (1: steepest descend (only used for testing), 2: variable metric)
3. include rotational forces (if = 1)
4. include translational forces (if = 1)
5. include soft modes for receptor (if =1)
6. include soft modes for ligand (if =1)
7. number of ligand soft modes
8. number of receptor soft modes
9. add a restraining contribution (using parameters from the second input line), (if =1)
10. cutoff radius (squared, means 100.0 corresponds to a cutoff=10.0 Angstrom)

The selectivity of the current energy function is optimized for a short cutoff ($rcut^2=50.0$). A series of minimizations (with decreasing cutoff) is necessary because the pairlist to calculate the interactions is only calculated at the beginning of each minimizations (the variable metric minimizer converges faster if one calculates the pairlist only once)

3.1 Performing a single docking run:

```
$path/bins/attractsingle receptorr.pdb ligandr.pdb parmu.par
```

This command minimizes the placement of ligandr.pdb with respect to receptorr.pdb according to the attract.inp file (and parameter file parmu.par). For each minimization the program outputs a coordinate file: lig00N.pdb (N: EM number) after the final minimization an additional final set of coordinates is generated: o.pdb (final ligand structure); rezo.pdb (final receptor structure).

On the screen the following is printed:

final energy, non-electrostatic energy, electrostatic energy (distance dependent dielectric 15r), receptor energy, convergence (norm of residual force), rmsd of the final structure from standard.pdb, fraction of native contacts, fraction of native receptor atoms, fraction of native ligand atoms, 3 rotational coordinates, 3 translational coordinates of ligand

3.2 Performing a systematic docking search:

The input file (attract.inp) is the same for a systematic search. However, a few more files are necessary for a systematic search. With the program \$path/bin/translate one can generate approximately evenly distributed starting points on the surface of the receptor structure. The starting points are located at a distance that is slightly larger than the radius of the ligand.

```
$path/bin/translate receptorr.pdb ligandr.pdb > translate.dat
```

One can check the starting points using rasmol for visualization.

In addition to the starting points, starting orientations need to be generated. The number of steps in the Euler angles (theta, phi, rot) are encoded in the file rotation.dat:

```
7      6
0.0    1
30.0   4
60.0   8
90.0  12
120.0   8
150.0   4
180.0   1
```

With the following command one can run a systematic search:

```
$path/bins/attractmulti receptorr.pdb ligandr.pdb parmu.par
```

This command performs a systematic multi-start docking minimization by placing the ligand initially at the positions stored in translate.dat and generating different orientations according to file rotation.dat. These initial placements are then energy minimized and if the minimization converged and gives a score below a preset threshold the final structure is written to standard output. To the file "out.dat" for each final complex structure the following is written:

Number, final energy, non-electrostatic energy, electrostatic energy (distance dependent dielectric 15r), receptor energy, convergence (FNORM), rmsd of the final structure from standard.pdb, fraction of native contacts, fraction of native receptor atoms, fraction of native ligand atoms, 3 rotational coordinates, 3 translational coordinates of ligand and 5 displacements in normal modes. The header of an "out.dat" file is shown below:

```
1 -12.345 -12.34 0.0 0.0 0.4E-07 26.81 0.007 0.041 0.055 -1.863 2.011 1.876
109.80 22.51 -2.22 0.0 0.0 0.0 0.0 0.0
2 -11.027 -8.76 -2.3 0.0 0.3E-10 22.66 0.000 0.055 0.000 1.877 -1.425 3.896
103.45 14.34 3.80 0.0 0.0 0.0 0.0 0.0
3 -12.345 -12.34 0.0 0.0 0.6E-06 26.81 0.007 0.041 0.055 1.862 5.153 5.018
109.80 22.51 -2.22 0.0 0.0 0.0 0.0 0.0
4 -14.536 -12.10 -2.4 0.0 0.1E-06 28.41 0.000 0.000 0.000 1.931 4.389 0.956
124.35 17.75 -8.81 0.0 0.0 0.0 0.0 0.0
5 -17.537 -12.46 -5.1 0.0 0.4E-06 22.63 0.000 0.082 0.000 1.410 6.205 5.451
112.15 19.91 11.32 0.0 0.0 0.0 0.0 0.0
6 -12.104 -12.10 0.0 0.0 0.1E-05 30.93 0.000 0.014 0.000 4.942 4.853 2.031
105.97 16.16 -14.59 0.0 0.0 0.0 0.0 0.0
```

3.3 Accounting for side chain conformational changes by multiple rotamer copies of selected side chains

The program reduce accepts several different side chain conformations for one residue if marked in the following manner (in the atomic resolution structure file). The rotameric states can be generated using the program *rotam* (see auxiliary programs).

```
ATOM 99 SG CYS 13 -2.797 -11.546 -16.236
ATOM 100 N LYS 14 -6.029 -11.602 -19.390
ATOM 101 CA LYS 14 -7.022 -11.674 -20.468
ATOM 102 C LYS 14 -6.883 -12.730 -21.490
ATOM 103 O LYS 14 -7.669 -12.639 -22.510
ATOM 104 CB LYS A 14A -7.069 -10.265 -21.092
ATOM 105 CG LYS A 14A -7.489 -9.253 -19.992
ATOM 106 CD LYS A 14A -7.516 -7.892 -20.667
ATOM 107 CE LYS A 14A -7.909 -6.784 -19.729
ATOM 108 NZ LYS A 14A -7.936 -5.452 -20.389
TER
ATOM 109 CB LYS B 14B -7.069 -10.265 -21.092
ATOM 110 CG LYS B 14B -8.107 -10.272 -22.246
ATOM 111 CD LYS B 14B -8.105 -8.864 -22.816
ATOM 112 CE LYS B 14B -9.069 -8.696 -23.959
ATOM 113 NZ LYS B 14B -9.068 -7.318 -24.517
TER
ATOM 114 N ALA 15 -5.999 -13.689 -21.329
```

After application of the "\$path/bin/reduce" command his example results in 2 rotamer copies (for Lys) also for the reduced representation of the protein structure (marked by conformer 1 and 2, respectively, in 11th column):

ATOM	67	CSE	CYS	13	-4.067	-11.882	-16.859	7	0.000	0	1.00
ATOM	68	N	LYS	14	-6.029	-11.602	-19.390	30	0.000	0	1.00
ATOM	70	CA	LYS	14	-7.022	-11.674	-20.468	32	0.000	0	1.00
ATOM	72	C	LYS	14	-6.883	-12.730	-21.490	32	0.000	0	1.00
ATOM	74	O	LYS	14	-7.669	-12.639	-22.510	31	0.000	0	1.00
ATOM	77	CB	LYS	14	-7.489	-9.253	-19.992	16	0.000	1	1.00
ATOM	80	CE	LYS	14	-7.909	-6.784	-19.729	17	1.000	1	1.00
ATOM	83	CB	LYS	14	-8.107	-10.272	-22.246	16	0.000	2	1.00
ATOM	86	CE	LYS	14	-9.069	-8.696	-23.959	17	1.000	2	1.00
ATOM	88	N	ALA	15	-5.999	-13.689	-21.329	30	0.000	0	1.00

The Attract program can automatically switch between the two copies during docking minimization in order to select the best fitting rotamer (example 2).

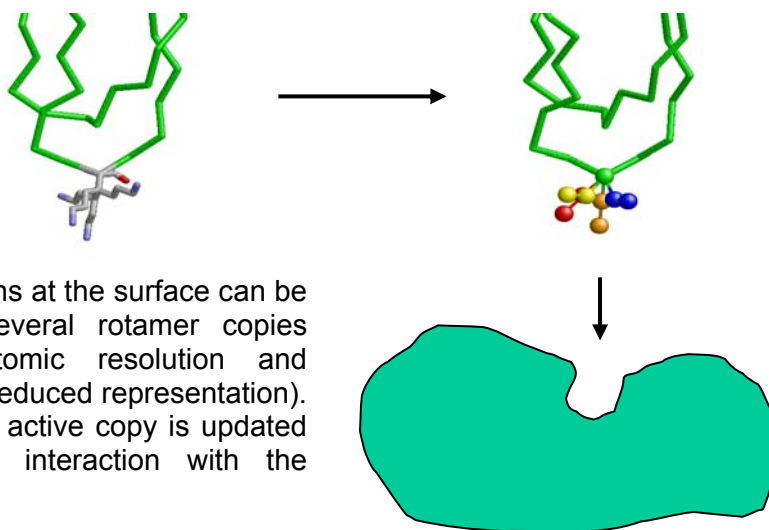


Figure 2: Side chains at the surface can be represented by several rotamer copies (generated at atomic resolution and translated into the reduced representation). During docking the active copy is updated depending on its interaction with the protein partner.

4. Auxiliary programs to setup docking searches and to analyse results

```
$path/bin/modesca structureca.pdb
```

This program calculates elastic network modes according to Hinsen (Proteins, 1998) for the CA-backbone only. It generates a file "eigen.out" which contains all eigenvectors and eigenvalues.

```
$path/bin/viewe 0.005 structureca.pdb eigen.out
```

This program generates a series of deformed structures (using structureca.pdb as reference) in the 10 softest normal modes (taken from eigen.out).

```
$path/bin/compare structure1ca.pdb structure2ca.pdb eigen.out Nmodes
```

This program calculates a best possible deformation of structure1 in the Nmodes (softest) modes to best approximate structure2. The deformed structure is stored in final.pdb.

```
$path/bin/collect ligand.pdb output.dat receptor.pdb > minima.pdb
```

This program reads an Attract output.dat file and reduces the output such that every docking minimum occurs just once (stored in output.red). This file also contains the information how often a minimum was found (last column). In addition, it stores a file

output.tra (in pdb-format) that contains the Cartesian center coordinates of all docking minima. The file minima.pdb contains all docking minima (just one).

```
$path/bin/rmsca structure1.pdb structure2.pdb
```

This program superimposes two structures structure1.pdb/structure2.pdb with respect to all CA-atoms and stores it in out.pdb. In addition, a file rmsd.dat is generated that contains the Rmsd of each CA-atom after best superposition.

```
$path/bin/rotam structure.pdb rcut
```

The program generates several side chain rotamers for residues indicated in the file "list.dat" (in the working directory) according to possible side chain dihedral angles listed in file "rota.dat" (in the working directory). It produces an output file "reg.pdb" and "rota.pdb". The "reg.pdb" file contains the original structure but with the residue numbering starting with 1. The file "rota.pdb" contains the coordinates including several rotameric states for those residues indicated in "list.dat". Rotamers for which more than 2 side chain atoms come closer than rcut (usually set to 4.0 Angstrom) to any other protein atom are eliminated. Note, that the program internally assumes the residue numbering starts with 1. Therefore it is wise to run it twice (use reg.pdb as target) and also use the reg.pdb file to identify the residues (and there numbers) for which side chain copies should be generated.

5. Example 1: Rigid docking of an enzyme-inhibitor complex: Comparison of docking bound and unbound structures

This example employs a trypsin-inhibitor system (pdb-entry: 2ptc). Four docking searches with a limited number of start placements are performed with either both partners in bound form, one partner in unbound form and both partners in unbound form. In this case the conformational differences between bound and unbound forms are relatively small. Systematic searches are performed with a small set of start points stored in "translate_red.dat".

1. generation of reduced protein models of each structure:

```
$path/bin/reduce 2PTC_b_l.pdb  
$path/bin/reduce 2PTC_b_r.pdb  
$path/bin/reduce 2PTC_u_l.pdb  
$path/bin/reduce 2PTC_u_r.pdb
```

2. use the ligand structure (extracted from the complex) as reference structure (standard.pdb):

```
cp 2PTC_b_lr.pdb standard.pdb
```

It might be a good idea to perform single docking runs first to check how far docked structures deviate from experiment if one starts from the placements found in the experimental complex structure.

3. setup start placements for docking minimization:

```
cp translate_red.dat translate.dat
```

4. perform four systematic searches:

```
$path/bin/attractmulti 2PTC_b_rr.pdb 2PTC_b_lr.pdb parmu.par >res.pdb
```



```

mv -f out.dat out_br_bl.dat

$path/bin/attractmulti 2PTC_u_rr.pdb 2PTC_b_lr.pdb parmu.par >res.pdb

mv -f out.dat out_ur_bl.dat

$path/bin/attractmulti 2PTC_b_rr.pdb 2PTC_u_lr.pdb parmu.par >res.pdb

mv -f out.dat out_br_ul.dat

$path/bin/attractmulti 2PTC_u_rr.pdb 2PTC_u_lr.pdb parmu.par >res.pdb

mv -f out.dat out_ur_ul.dat

```

5. After the systematic searches have been completed all unique minima need to be identified and sorted:

```

$path/bin/collect 2PTC_b_lr.pdb out_br_bl.dat 2PTC_b_rr.pdb 0 >
col_br_bl.pdb

$path/bin/collect 2PTC_u_lr.pdb out_ur_ul.dat 2PTC_u_rr.pdb 0 >
col_ur_ul.pdb

sort +ln out_br_bl.red > out_br_bl.srt
sort +ln out_ur_ul.red > out_ur_ul.srt

```

The generated data and docked complexes can now be analysed using gnuplot and visualized using VMD (details during the tutorial).

6. Example 2: Inclusion of side chain rotamer copies during docking

This example employs also a trypsin-inhibitor system (pdb2r9p, a human trypsin enzyme). However, in this case the conformational differences between bound and unbound forms are larger than for the first example. We will compare systematic searches (again using a reduced set of start points stored in “translate_red.dat”) employing first both partners in bound form, second the inhibitor (ligand) protein in unbound form (receptor in bound form) and third the unbound ligand protein with some rotamer copies on critical residues.

1. generation of reduced protein models of each structure:

```

$path/bin/reduce r9pl.pdb
$path/bin/reduce r9pr.pdb
$path/bin/reduce r9pu.pdb

```

2. use the ligand structure (extracted from the complex) as reference structure (standard.pdb):

```

cp r9plr.pdb standard.pdb

```

It might again be a good idea to perform single docking runs first to check how far docked structures deviate from experiment if one starts from the placements found in the experimental complex structure.

3. setup start placements for docking minimization:

```

cp translate_red.dat translate.dat

```

4. perform systematic searches (using bound or unbound inhibitor structures):

```
$path/bin/attractmulti r9prrr.pdb r9plr.pdb parmu.par >res.pdb
mv -f out.dat out_br_bl.dat
$path/bin/attractmulti r9prrr.pdb r9pur.pdb parmu.par >res.pdb
mv -f out.dat out_br_ul.dat
```

5. generate possible rotamer states for 3 critical residues:

```
$path/bin/rotam r9pu.pdb 4.0
$path/bin/rotam reg.pdb 4.0
$path/bin/reduce rota.pdb
```

4. perform a systematic search using the unbound inhibitor with side chain copies.

```
$path/bin/attractmulti r9prrr.pdb rotar.pdb parmu.par >res.pdb
mv -f out.dat out_rota.dat
```

5. analyse and compare the results (details during tutorial):

```
$path/bin/collect r9plr.pdb out_br_bl.dat r9prrr.pdb 0 > col_br_bl.pdb
$path/bin/collect r9pur.pdb out_br_ul.dat r9prrr.pdb 0 > col_br_ul.pdb
$path/bin/collect r9pur.pdb out_rota.dat r9prrr.pdb 0 > col_rota.pdb

sort +ln out_br_bl.red > out_br_bl.srt
sort +ln out_br_ul.red > out_br_ul.srt
sort +ln out_rota.red > out_rota.srt
```

7. Normal mode calculation of proteins using an elastic network model

Normal modes can be calculated using the program `$path/bin/modes`. It employs an elastic network model according to Hinsen (Proteins, 1998).

```
$path/bin/modes receptor.pdb
```

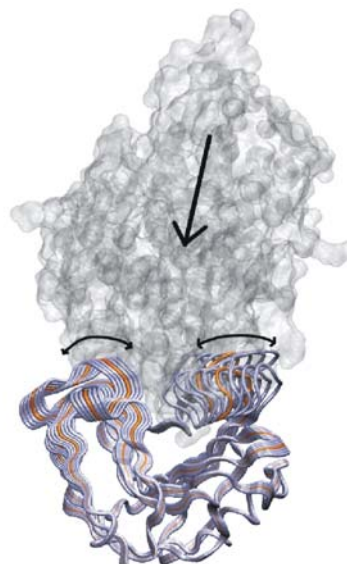
This command performs a normal mode calculation on the CA backbone of the receptor structure and adapts the modes to the full structure by moving each residue as rigid unit according to the calculated normal modes. The resulting modes are stored in the file `hm.dat`. In addition, the program generates control files (`zwei001.pdb`, `zwei002.pdb`, `zwei010.pdb`) that contain series of structures deformed in the first, second ... etc. normal modes. These structures can be visualized using Rasmol or VMD to check the deformations.

In order to include energy minimization in a set of normal modes (relaxation in normal modes) one needs to include the file `hm.dat` in the working directory (calculated by `modes`) and the `attract.inp` file looks like:

```
5      1      0
73.98360 -53.41780 -13.62360 0.00050
30 2 1 1 1 0 0 5 1 2500.00
30 2 1 1 1 0 0 5 1 1500.00
40 2 1 1 1 0 0 5 1 100.00
60 2 1 1 1 0 0 5 0 50.00
60 2 1 1 1 0 0 5 0 50.00
```

Flag 1 in the first input line indicates that the file “hm.dat” is opened and read from the working directory. The 5th entry in each EM control line indicates that normal mode relaxation is included using the first 5 (softest) modes (8th entry).

Figure 3: It is possible to include a number of collective degrees of freedom (usually normal mode vectors) during docking. Docking minimization is then performed in translational and rotational coordinates of the ligand and normal mode directions of the receptor (and/or ligand) as illustrated in the Figure. The flexible partner is shown as superposition of structures deformed in one collective mode (second partner in transparent surface representation). Note, the indicated deformations are amplified for visualization and are usually much smaller during docking.



8. Example 3: Docking of Taxi-Inhibitor to Xylanase: Accounting for global changes during docking

Basis of this example is the Xylanase/taxi-Inhibitor system (capri18 target). In this case the conformational differences between bound and unbound forms of the Xylanase (receptor) are mainly due to a global opening motion of the enzyme upon inhibitor binding. The overlap of the soft normal mode directions calculated for the unbound form of Xylanase with the conformational difference between bound and unbound forms will be compared. Systematic searches (again using a reduced set of start points stored in “translate_red.dat”) using bound or unbound Xylanase and with or without energy minimization in normal modes will be performed and compared.

1. setup (similar to previous examples). We use the bound ligand (TAXI-I: xyl.pdb) and the bound (xylr.pdb) and unbound (xyu.pdb) forms of the receptor (xylanase enzyme).

```
$path/bin/reduce xyl.pdb
$path/bin/reduce xylr.pdb
$path/bin/reduce xyu.pdb
cp xylr.pdb standard.pdb
cp attractrigid.inp attract.inp
```

2. perform systematic searches (rigid receptor, bound and unbound)

```
$path/bin/attractmulti xyrr.pdb xylr.pdb parmu.par >res.pdb
mv -f out.dat out_br_bl.dat
$path/bin/attractmulti xyur.pdb xylr.pdb parmu.par >res.pdb
mv -f out.dat out_ur_bl.dat
```

3. ENM normal mode analysis (see above paragraph on normal mode analysis)

```
grep CA xyrr.pdb > xyrca.pdb
grep CA xyur.pdb > xyuca.pdb
```

```
$path/bin/modesca xyuca.pdb
$path/bin/rmsca xyrca.pdb xyuca.pdb
mv -f rmsd.dat rmsd0.dat
```

4. analyse best possible deformation to approximate bound Xylanase structure

```
$path/bin/compare xyuca.pdb xyrca.pdb eigen.out 1
$path/bin/rmsca xyrca.pdb final.pdb
mv -f rmsd.dat rmsd1.dat
$path/bin/compare xyuca.pdb xyrca.pdb eigen.out 5
$path/bin/rmsca xyrca.pdb final.pdb
mv -f rmsd.dat rmsd5.dat
```

We will analyse the results using Rasmol and Xmgrace.

5. setup soft modes for the complete Xylanase structure

```
$path/bin/modes xyur.pdb
```

6. run a systematic search including soft mode relaxation of the receptor

```
cp attractflex.inp attract.inp
$path/bin/attractmulti xyur.pdb xylr.pdb parmu.par >res.pdb
mv -f out.dat out_hm5.dat
```

7. Analysis the results by collecting docking minima and plotting Rmsd vs. score of the docking solutions

```
$path/bin/collect xylr.pdb out_br_bl.dat xyrr.pdb 0 > col_br_bl.pdb
$path/bin/collect xylr.pdb out_ur_bl.dat xyur.pdb 0 > col_ur_bl.pdb
$path/bin/collect xylr.pdb out_hm5.dat xyur.pdb 1 > col_hm5.pdb

sort +ln out_br_bl.red > out_br_bl.srt
sort +ln out_ur_bl.red > out_ur_bl.srt
sort +ln out_hm5.red > out_hm5.srt
```

9. Example 4: Docking of a protein fragment to the HIV-capsid protein

In this example we will try to dock a protein inhibitor to a capsid domain from the HIV gag protein (a protein that is part of the virus envelope). The capsid domain structure is available in an unbound (pdb1a43) and bound form (pdb2buo). The protein fragment (termed "buol.pdb") inhibits the assembly of the HIV-Gag-protein. Docking will be performed both to the bound and unbound forms of the capsid-protein-domain ("buor.pdb" and "buou.pdb"). We will also compare rigid docking searches with searches including relaxation (minimization) in soft normal modes of the gag capsid domain.

1. Setup (similar to previous examples)

```
$path/bin/reduce buol.pdb
$path/bin/reduce buor.pdb
$path/bin/reduce buou.pdb
cp buolr.pdb standard.pdb
cp attractrigid.inp attract.inp
cp translate_red.dat translate.dat
```

2. Performing a rigid systematic search with bound or unbound receptor structure:

```
$path/bin/attractmulti buorr.pdb buolr.pdb parmu.par >res.pdb
```

```
mv -f out.dat out_br_bl.dat
$path/bin/attractmulti buour.pdb buolr.pdb parmu.par >res.pdb
mv -f out.dat out_ur_bl.dat
```

3. ENM normal mode analysis

```
grep CA buorr.pdb > buorca.pdb
grep CA buour.pdb > buouca.pdb
$path/bin/modesca buouca.pdb
$path/bin/rmsca buorca.pdb buouca.pdb
mv -f rmsd.dat rmsd0.dat
$path/bin/compare buouca.pdb buorca.pdb eigen.out 1
$path/bin/rmsca buorca.pdb final.pdb
mv -f rmsd.dat rmsd1.dat
$path/bin/compare buouca.pdb buorca.pdb eigen.out 5
$path/bin/rmsca buorca.pdb final.pdb
mv -f rmsd.dat rmsd5.dat
```

4. Systematic docking search including normal mode relaxation of the receptor

```
$path/bin/modes buour.pdb
cp attractflex.inp attract.inp
$path/bin/attractmulti buour.pdb buolr.pdb parmu.par >res.pdb
mv -f out.dat out_hm5.dat
```

5. Analysis of results

```
$path/bin/collect buolr.pdb out_br_bl.dat buorr.pdb 0 > col_br_bl.pdb
$path/bin/collect buolr.pdb out_ur_bl.dat buour.pdb 0 > col_ur_bl.pdb
$path/bin/collect buolr.pdb out_hm5.dat buour.pdb 1 > col_hm5.pdb

sort +1n out_br_bl.red > out_br_bl.srt
sort +1n out_ur_bl.red > out_ur_bl.srt
sort +1n out_hm5.red > out_hm5.srt
```

10. Example 5: Docking of a helical peptide to a Helix-binding domain of SRP (signal recognition particle)

This example is similar to the previous case, however, it potentially involves much larger global changes. Docking is performed on an alpha-helix that can bind to a helix binding domain of the signal recognition particle (SRP). The SRP is a structure that is involved in recognizing membrane bound proteins after translation and mediating its attachment to the membrane (via a hydrophobic helix). Docking involves the helix-binding domain in a bound structure (termed srpr.pdb, extracted from pdb-entry:1ry1), a homology-modelled unbound structure (srpu.pdb) and a helix-ligand protein (extracted from pdb1ry1), termed: srpl.pdb. This example will also include the possibility to relax the receptor structure in pre-calculated normal modes.

1. Setup and docking using rigid partners

```
$path/bin/reduce srpl.pdb
$path/bin/reduce srpr.pdb
$path/bin/reduce srpu.pdb
cp attractrigid.inp attract.inp
cp translate_red.dat translate.dat
$path/bin/attractmulti srpr.pdb srplr.pdb parmu.par >res.pdb
mv -f out.dat out_br_bl.dat
$path/bin/attractmulti srpur.pdb srplr.pdb parmu.par >res.pdb
```

```
mv -f out.dat out_ur_bl.dat
```

2. Analyse ENM-normal modes

```
grep CA srprrr.pdb > srprca.pdb
grep CA srpur.pdb > srpuca.pdb
$path/bin/modesca srpuca.pdb
$path/bin/rmsca srprca.pdb srpuca.pdb
mv -f rmsd.dat rmsd0.dat
$path/bin/compare srpuca.pdb srprca.pdb eigen.out 1
$path/bin/rmsca srprca.pdb final.pdb
mv -f rmsd.dat rmsd1.dat
$path/bin/compare srpuca.pdb srprca.pdb eigen.out 5
$path/bin/rmsca srprca.pdb final.pdb
mv -f rmsd.dat rmsd5.dat
```

3. Perform systematic docking including soft mode relaxation

```
$path/bin/modes srpur.pdb
cp attractflex.inp attract.inp
$path/bin/attractmulti srpur.pdb srplr.pdb parmu.par >res.pdb
mv -f out.dat out_hm5.dat
```

4. Collect docking minima and analyse results

```
$path/bin/collect srplr.pdb out_br_bl.dat srprrr.pdb 0 > col_br_bl.pdb
$path/bin/collect srplr.pdb out_ur_bl.dat srpur.pdb 0 > col_ur_bl.pdb
$path/bin/collect srplr.pdb out_hm5.dat srpur.pdb 1 > col_hm5.pdb
sort +1n out_br_bl.red > out_br_bl.srt
sort +1n out_ur_bl.red > out_ur_bl.srt
sort +1n out_hm5.red > out_hm5.srt
```

Remarks:

Commands and options to visualize docking trajectories or to plot data are not included but will be explained during the tutorial.