

# **RosettaDock examples and cases**

## **CONTENTS**

### **A) SCORE**

Cases 1-3

### **B) PREPACK**

Cases 4-5

### **C) GLOBAL RUN**

Cases 6-8

### **D) LOCAL RUN**

Cases 9-10

### **E) CONSTRAINTS**

Cases 11-15

### **F) SIMPLE DOCKING MODES**

Cases 16-17

### **G) FULL PRODUCTION RUN**

Case 18

## **REFERENCES**

## A) SCORE

### COMMON I/O FILES

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting or native structure of the protein-protein complex  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

Rosetta\_print.log : standard Rosetta file that prints information for the run

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides<sup>1</sup>

#### 1)

Score a low resolution structure

**Rosetta.exe ZZ 1brs \_ -dock -score -s 1brs.pdb >**

**Rosetta\_print.log**

#### OUTPUT

ZZ1brs.sc : score file with energy terms for the low resolution structure

#### 2)

Score a high resolution structure in Full-Atom(FA) mode

**Rosetta.exe ZZ 1brs \_ -dock -score -dockFA -s 1brs.pdb >**

**Rosetta\_print.log**

#### OUTPUT

ZZ1brs.fasc : score file with energy terms for the high resolution structure

#### 3)

Score a high resolution structure without repacking on the interface

**Rosetta.exe ZZ 1brs \_ -dock -score -dockFA -dock\_score\_norepack -s 1brs.pdb >**

**Rosetta\_print.log**

#### OUTPUT

ZZ1brs.sc : score file with energy terms for the low resolution structure

In the low resolution representation, each residue is represented by the four backbone atoms (N, Ca, C and O) and one pseudo-atom, the “centroid” to represent the side-chain. The score is decomposed into residue environment and pair terms(d\_env, d\_pair) and contact, vdw(bump), alignment scores and other constraints(d\_sc, d\_fab).<sup>2</sup> In the high resolution score the proteins are represented in atomistic (full atom) configuration. The forcefield is based on CHARMM and other empirical parameterizations.

The score is decomposed in several full atom energy terms:

van der Waals repulsion and attraction(fa\_atr, fa\_rep), solvation(fa\_sol) sidechain

H-bonds (hb\_sc) rotamer probability (Dunbrack<sup>3</sup>) score (fa\_dun) electrostatic residue pair potentials

(fa\_pair) and coulombic term with linear dependent dielectric constant (d\_elec) short range (sr) and long range (lr) backbone H-bonds (hb\_srbb, hb\_lrbb)<sup>2</sup>

High resolution energy terms according to RosettaDock <sup>2,4-6</sup>	names in the score file
van der Waals repulsion	fa_atr
van der Waals repulsion	fa_rep
rotamer probability (Dunbrack)	fa_dun
electrostatic residue pair potentials	fa_pair
coulombic dielectric dependent	d_elec
short range H-bonds	hb_srbb
long range H-bonds	hb_lrbb
implicit Gaussian solvation term	fa_sol
sidechain H-bonds	hb_sc

## B) PREPACK

To save computation time, the side-chains of the protein monomer are added and optimized before docking (PREPACK) and to prevent errors in docking due to irregularities (e.g. crystal contacts) in the native structure. Rosetta uses Rapid Side-Chain Optimization for pre-packing the proteins. Using a simulated annealing protocol, side-chain rotamers can be completely re-optimized. The side-chain rotamers are rapidly optimized by cycling through each side-chain position in random order and replacing the current rotamer with the lowest energy rotamer available at that position.<sup>7</sup> During the pre-packing of the monomer components, disulfides are allowed to form, and extra rotamer conformations are allowed for cysteine residues.<sup>2</sup>

### COMMON I/O FILES

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting or native structure of the protein-protein complex  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

before.pdb : starting structure  
away.pdb : slide partners away and not repack  
1brs.ppk.pdb : **prepacked complex to use for runs**  
ZZ1brs.fasc : score file with full energy terms for the above structures  
1brs\_prepack.out : standard log/output Rosetta file  
Rosetta\_print.log : standard Rosetta file that prints information for the run

#### 4)

Prepack full (separate the components, repack separately, put them back, and write PDB)

**Rosetta.exe ZZ 1brs \_ -dock -dock\_min -prepack\_full -ex1 -ex2 -s 1brs.pdb >**

**Rosetta\_print.log**

#### OUTPUT

repacked\_away.pdb : slide partners away and repack  
1brs.reppk.pdb : re-prepacked complex

#### 5)

Prepack\_full, with off-rotamer sampling of side chain

**Rosetta.exe ZZ 1brs \_ -dock -dock\_min -prepack\_rtmin -ex1aro -ex2 -s 1brs.pdb >**

**Rosetta\_print.log**

#### OUTPUT

minimized\_away.pdb : slide partners away and repack  
1brs.remin.pdb : re-prepacked complex

**ex1aro** option includes extra aromatic residues plus **ex1** conformations for chi1 angles

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides<sup>1</sup>

## C) GLOBAL RUN

In GLOBAL docking search (or blind prediction) the orientation of the two protein molecules is unknown and it is necessary to sample the complete conformational space starting with an adequate number of different random orientations.

### 6)

How to DOCK two proteins GLOBALLY (such as the PREPACKED structure of the Barnase-Barstar Complex<sup>1</sup> employing a low-resolution, rigid-body, Monte Carlo search followed by simultaneous all-atom optimization of backbone displacement and side-chain conformations using the PREPACKED structure (Only residues with side-chain centroid positions within 8Å° of a side-chain centroid of the other protein partner are included in the packing routine)

```
Rosetta.exe ZZ 1brs -dock -dock_mcm -ex1 -ex2aro_only -fake_native -s 1brs.pdb >  
Rosetta_print.log
```

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting structure of the protein-protein complex (Barnase-Barstar)  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

Rosetta\_print.log : standard Rosetta file that prints information for the run  
ZZ1brs\_0001.pdb: final complex structure after docking  
ZZ1brs.fasc : score file with full energy terms for the above structures

#### BENCHMARK

Around 10 minutes with VisualStudio.NET (8.0) C++ (cl) for an Intel Xeon 5150@2.66GHz and Windows XP 64bit OS

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides<sup>1</sup>



7)

How to DOCK two proteins GLOBALLY (after we have PREPACKED the structure) employing a low-resolution, rigid-body, Monte Carlo search followed by simultaneous all-atom optimization of backbone displacement and side-chain conformations and improved modeling of side chain conformations using Rotamer Trial Minimization protocol.

```
Rosetta.exe ZZ 1brs _ -dock -dock_mcm dock_rtmin -ex1 -ex2aro_only -fake_native -s 1brs.pdb > Rosetta_print.log
```

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting structure of the protein-protein complex (Barnase-Barstar)  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

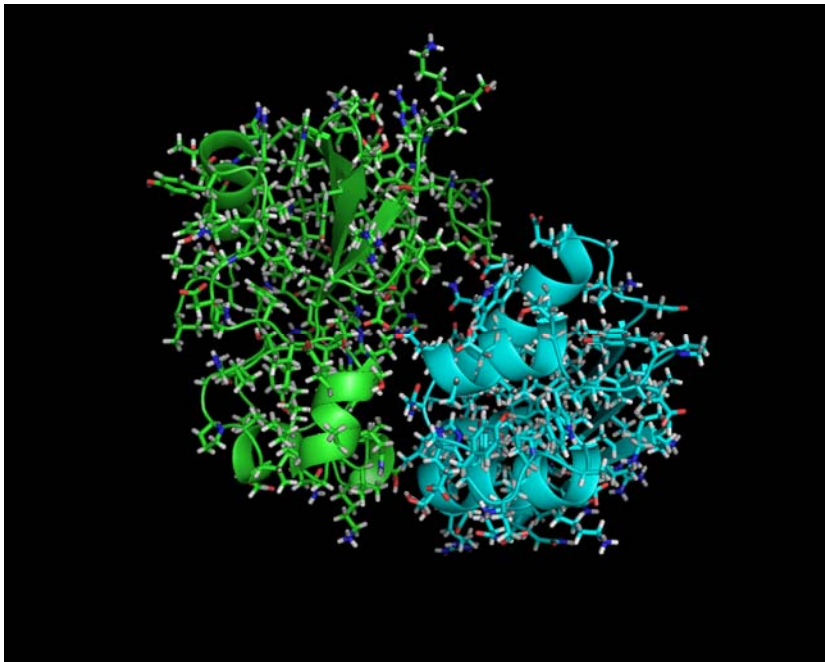
#### OUTPUT

Rosetta\_print.log : standard Rosetta file that prints information for the run  
ZZ1brs\_0001.pdb: final complex structure after docking  
ZZ1brs.fasc : score file with full energy terms for the above structures

#### BENCHMARK

Around 12 minutes with VisualStudio.NET (8.0) C++ (cl) for an Intel Xeon 5150@2.66GHz and Windows XP 64bit OS

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides.<sup>1</sup>



8)

How to dock two partners such as the cohesin–dockerin complex<sup>8</sup> randomizing the orientation one of them and given bound coordinates for one of the partners (**GLOBAL search**)?<sup>9</sup>

```
Rosetta.exe ZZ 1ohz _-dock -dock_mcm -ex1 -ex2aro_only -norepack2 -randomize1 -fake_native -s 1ohz.pdb > Rosetta_print.log
```

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1ohz.pdb : starting structure of the protein-protein complex<sup>8</sup>

#### OUTPUT

ZZ1ohz\_0001.pdb: final complex structure after docking  
ZZ1ohz.fasc : score file with full energy terms for the above structures  
Rosetta\_print.log : standard Rosetta file that prints information for the run

#### REFERENCE

From the work and runs of Daily et al.<sup>9</sup> for CAPRI round 3-5 (<http://capri.ebi.ac.uk>), (<http://graylab.jhu.edu/>)

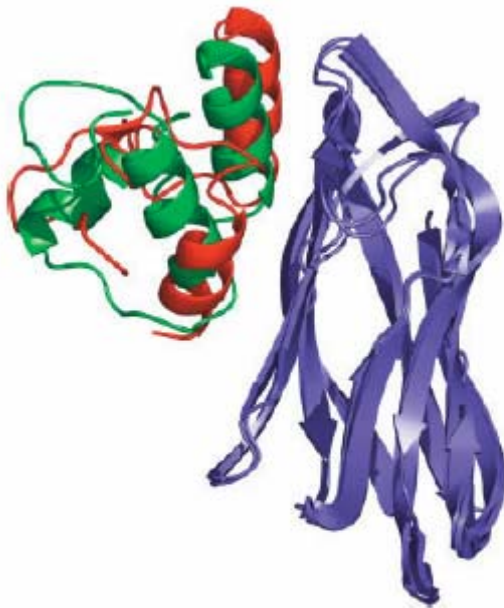


Fig. 2. Homology docking prediction in CAPRI. T11: Cohesin–dockerin. Blue: cohesin; red: predicted dockerin homology model; green: native dockerin.<sup>18</sup> 42% contacts, 6.11 Å Lrms, 1.17 Å Irms.

## D) LOCAL RUN

In addition to the **GLOBAL** docking searches, we perform perturbation studies(**LOCAL SEARCHES**) to explore **the nature of the docking energy funnel**. These studies are performed on both bound complexes with native side-chain conformations removed and on unbound monomer components superimposed on the native, bound complex structures. Random starting positions are created by translating one Protein partner by Gaussian random distances of 3 Å standard deviation along the line of protein centers and 8 Å standard deviation in the two perpendicular directions, and by spinning the mobile partner by a Gaussian random angle of 8° standard deviation around the axis of centers and by tilting a Gaussian random angle of 8° standard deviation off the axis in a random direction. In this way, a set of 1000 random starting positions create a diffuse cloud that covers a reasonable area (~20 Å radius rmsd) with moderate density around the native ligand position.<sup>2</sup>

9)

How to DOCK two partners perturbing the initial structure of the complex and improved modeling of side chain conformations using Rotamer Trial Minimization (**LOCAL search**)?<sup>10</sup>

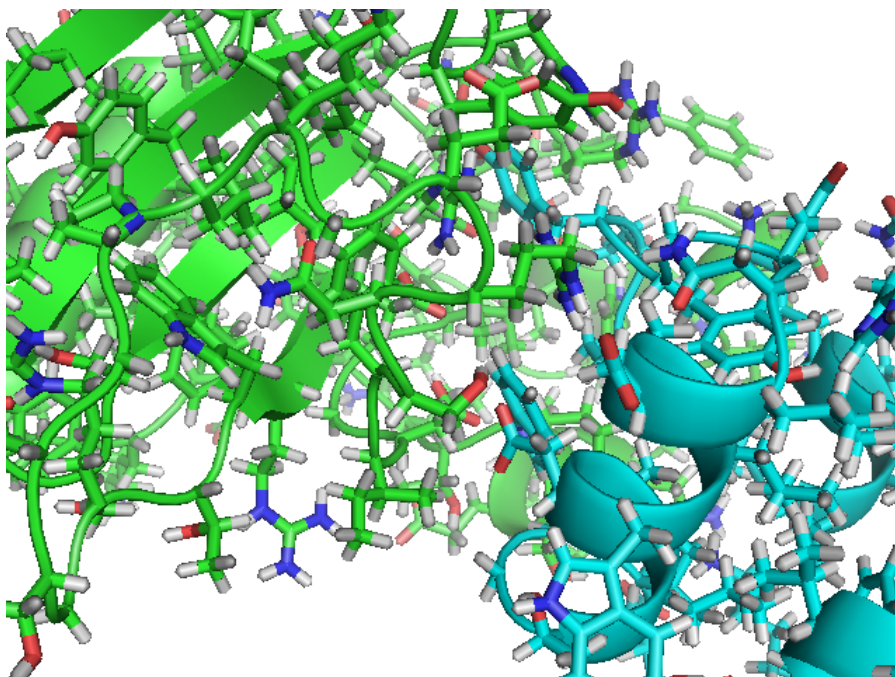
```
Rosetta.exe ZZ 1brs _ -dock -dock_mcm -dock_rtmin -dock_pert 5 10 10 -ex1 -ex2aro_only -s 1brs.pdb > Rosetta_print.log
```

### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting structure of the protein-protein complex (Barnase-Barstar)<sup>1</sup>  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

### OUTPUT

Rosetta\_print.log : standard Rosetta file that prints information for the run  
ZZ1brs\_0001.pdb: final complex structure after docking  
ZZ1brs.fasc : score file with full energy terms for the above structures





10)

How to DOCK two partners (antigen-antibody) perturbing the initial structure of the complex (**LOCAL search**) and using the disulfide logfile and antibody information for the one partner?<sup>11</sup>

```
Rosetta.exe ZZ 1ahw _ -dock -dock_mcm -ex1 -ex2aro_only -fake_native -fab1  
-dock_pert 3 8 8 -fix_disulf 1ahw.fixdisulf -use_disulf_logfile 1ahw.disulflog  
-norepack_disulf -s 1ahw.pdb > Rosetta_print.log
```

#### INPUT

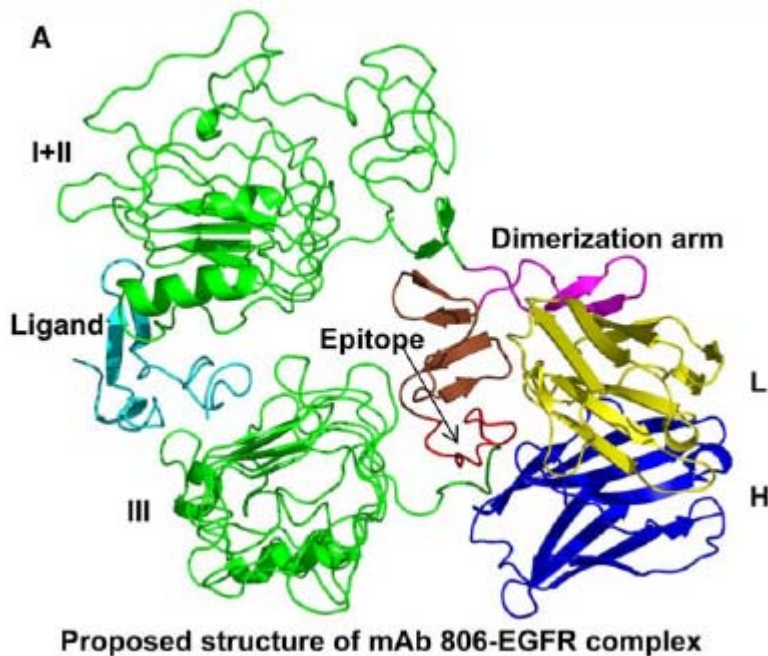
paths.txt : file to setup paths for the input, output structures and databases  
1ahw.pdb : starting structure of the protein-protein complex (immunoglobulin/tissue factor)  
1ahw.fixdisulf : disulfide information (use the script makefixdisulf.py )  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

ZZ1ahw\_0001.pdb: final complex structure after docking  
ZZ1ahw.fasc : score file with full energy terms for the above structures  
1ahw.disulflog : logfile for disulfide bridges  
Rosetta\_print.log : standard Rosetta file that prints information for the run

#### REFERENCE

From the work and runs of Sivasubramanian et al.<sup>11</sup> ( <http://graylab.jhu.edu/> )





## E) CONSTRAINTS

How to DOCK two proteins such as Barnase & Barstar<sup>1</sup> in a blind prediction with constraints (site(reward) & distance)?

### COMMON I/O FILES

#### INPUT

paths.txt  
1brs.pdb : starting unbound structure  
1brs.cst : site constraint for the residues  
1brs.dst : distance constraints for a pair of residues

#### OUTPUT

ZZ1brs.fasc : score file with full energy terms for the above structures  
Rosetta\_print.log : standard Rosetta file that prints information for the run  
ZZ1brs\_0001.pdb : final complex structure after docking

#### 11)

Rosetta.exe ZZ 1brs \_ -dock -dock\_mcm -randomize1 -randomize2 -fake\_native -ex1 -s 1brf.pdb > 1-Rosetta\_print.log

#### INPUT

**1brs.cst** : site constraint for the residues 1 & 2 of chain A , partner 1

#### 12)

Rosetta.exe ZZ 1brs \_ -dock -dock\_mcm -randomize1 -randomize2  
-fake\_native -ex1 -s 1brf.pdb > Rosetta\_print.log

#### INPUT

**1brf.cst** : site constraint for the residue 64 of chain D , partner 2

#### 13)

Rosetta.exe ZZ 1brs \_ -dock -dock\_mcm -randomize1 -randomize2  
-fake\_native -ex1 -s 1brf.pdb > Rosetta\_print.log

#### INPUT

**1brs.cst** : site constraint for the last residue 86 of chain D , partner 2

#### 14)

Rosetta.exe ZZ 1brs \_ -dock -dock\_min -fake\_native -s 1brf.pdb > Rosetta\_print.log

#### INPUT

**1brs.cst** : site constraint for the last residue 1,3 of chain D , partner 2

#### 15)

Rosetta.exe ZZ 1brs \_ -dock -dock\_mcm -randomize1 -randomize2  
-fake\_native -ex1 -s 1brs.pdb Rosetta\_print.log

#### INPUT

**1brs.cst** : site constraint for the last residue 1 of chain D , partner 2  
**1brs.dst** : 0.8nm distance constraint for the residue 1 of chain A, 29 chain D

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides<sup>1</sup>

## F) SIMPLE DOCKING MODES

### 16

How to DOCK two proteins such as Barnase & Barstar in a blind prediction with centroid search + repacking of interface side-chains?

**Rosetta.exe ZZ 1brs \_ -dock -dockFA -fake\_native -s 1brs.pdb > Rosetta\_print.log**

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting or native structure of the protein-protein complex (Barnase & Barstar)<sup>1</sup>  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

ZZ1brs.fasc : score file with full energy terms for the above structures  
Rosetta\_print.log : standard log/output Rosetta file  
ZZ1brs\_0001.pdb: final complex structure after docking

### 17

How to DOCK two proteins such as Barnase & Barstar in a blind prediction with centroid search + repacking of interface side-chains + + one cycle of rigid-body minimization?

**Rosetta.exe ZZ 1brs \_ -dock -dock\_min -fake\_native -s 1brs.pdb > Rosetta\_print.log**

#### INPUT

paths.txt : file to setup paths for the input, output structures and databases  
1brs.pdb : starting or native structure of the protein-protein complex (Barnase & Barstar)  
Rosetta.exe : the executable file that is compiled from Rosetta++ source

#### OUTPUT

ZZ1brs.fasc : score file with full energy terms for the above structures  
Rosetta\_print.log : standard log/output Rosetta file  
ZZ1brs\_0001.pdb: final complex structure after docking

**1brs.pdb** is the PDB file for Barnase-Barstar Complex without disulfides<sup>1</sup>

## G) FULL PRODUCTION RUN

In order to achieve an optimum LOCAL(see section D) and GLOBAL(see section C) search we have to produce a large number of final structures and to analyze/evaluate the convergence of the run with hierarchical statistical analysis of the clusters of the decoys.

In general we face 2 problems of docking:

- a) BOUND docking where we have the initial bound structure from the coordinates of the crystal that is provided from X-ray or NMR methods. In this case we need to explore the energy landscape around the bound (native) complex, to refine and optimize the low resolution structure using LOCAL search usually.
- b) UNBOUND docking when we use initial unbound structures of the protein partners either from experimental methods (PDB) or a modeled structure (Rosetta Abinitio<sup>12</sup>). In this case we employ a GLOBAL search usually, except if we know some biological information for the interface of the 2 partners (see section E).

### 18)

For a real GLOBAL run we need the **prepacked** structure (1br2.pdb=1brs.ppk.pdb, see section B) and a **calibration** run to estimate the maximum negative score (-100) in order to filter all the decoys that have score larger than -100.

```
Rosetta.exe AB 1br2 _-dock -dock_mcm -randomize1 -randomize2  
-fake_native -scorefilter -100 -ex1 -s 1br2.pdb -nstruct 100000 > Rosetta_print.log
```

#### INPUT

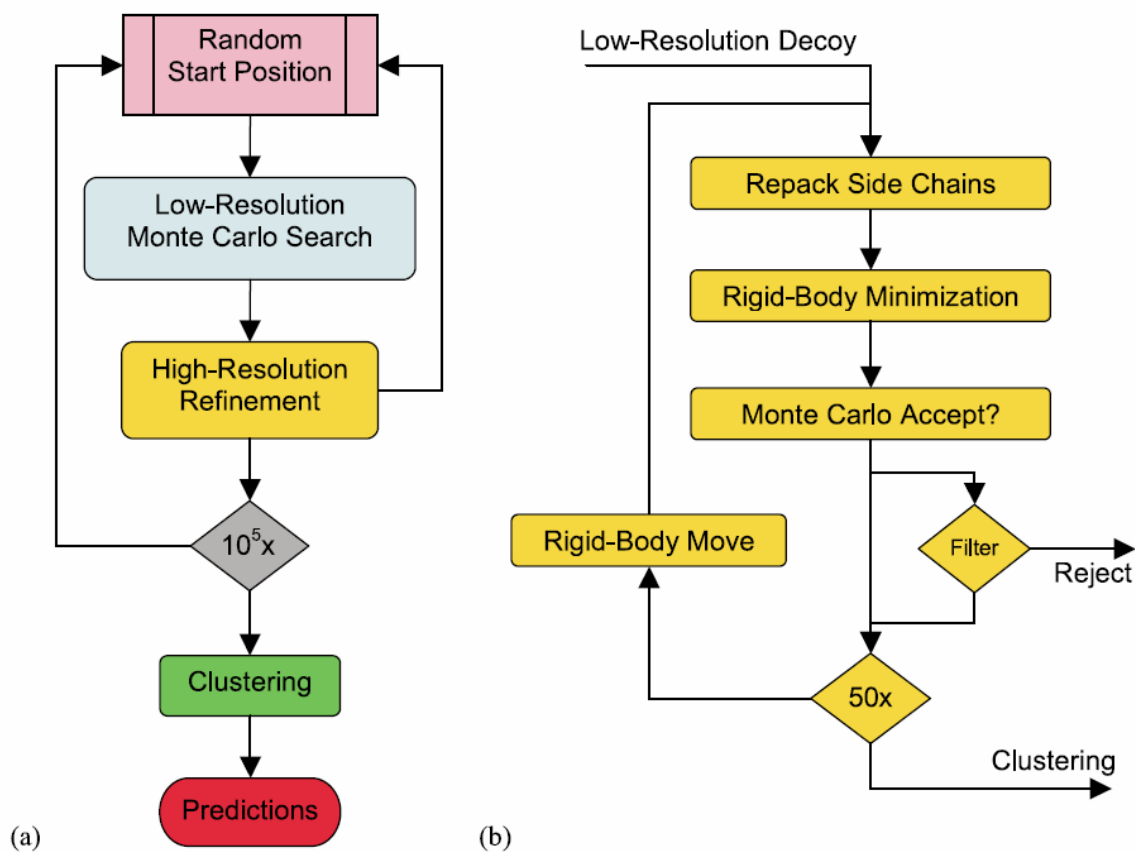
paths.txt  
1br2.pdb : starting unbound structure

#### OUTPUT

ZZ1br2.fasc : score file with full energy terms for the above structures  
Rosetta\_print.log : standard log/output Rosetta file  
ZZ1br2.fasc : score file with full energy terms for the above structures and also the root mean square deviation from the starting structure (rms)  
ZZ1br2\_0001.pdb : first successful decoy  
ZZ1br2\_0002.pdb : second successful decoy  
ZZ1br2\_0003.pdb : third successful decoy  
...  
ZZ1br2\_100000.pdb : last successful decoy

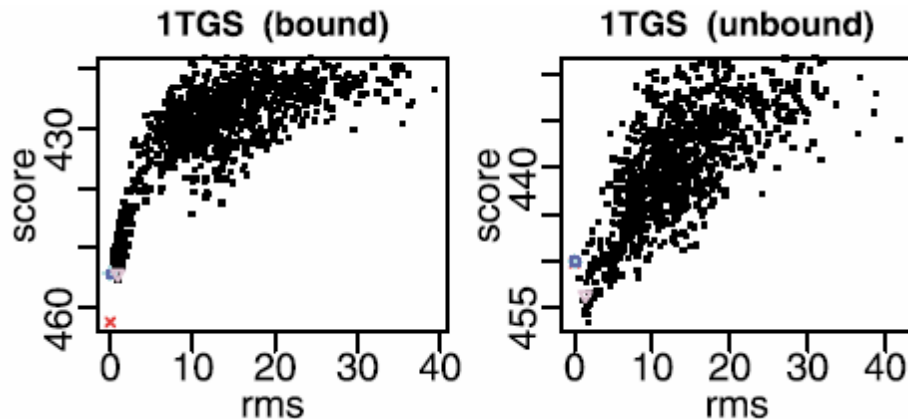


“The voyage over the free energy surface during one refinement cycle. The steps are: (1) a random perturbation (rigid-body translation and rotation) moves the structure on the potential surface; (2) a packing step optimizes the side-chain positions, thus changing the energy surface; (3) an explicit minimization finds the nearest local minimum accessible via a rigid-body translation and rotation. Start and finish positions are compared by the Metropolis criterion, and the cycle is repeated 50 times.”<sup>2</sup>

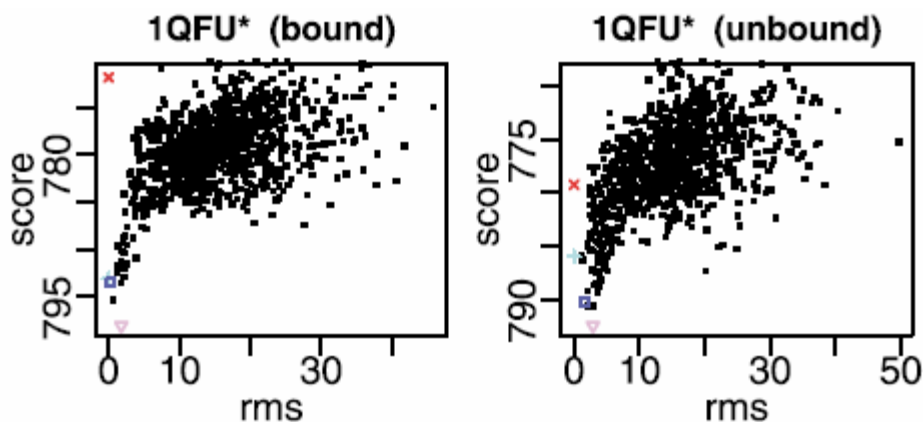


Flowchart of the docking algorithm, the generation of decoys and the clustering, to predict and compare with experiments.<sup>2</sup> “The search procedure is repeated to create approximately 100000 decoys per target.

Each final decoy is rescored, reducing the weight on the repulsive van der Waals energy term and including the surface area-based solvation term. The 200 best-scoring decoys are then clustered on the basis of pairwise root-mean-squared distance (rmsd) using a hierarchical clustering algorithm. Structures within a 2.5 Å clustering threshold are designated as a set, and the lowest-scoring decoy within the set represents that position. The clusters with the most members are selected as the final predictions, ranked according to the cluster sizes. The cluster size, or the degeneracy of the docked position, may be related to the entropy of the bound complex”<sup>2</sup>



Perturbation studies on enzyme/inhibitor complexes. Plots show rmsd versus score for 1000 decoys created from random starting positions near the native structure. Bound indicates that backbone coordinates were taken from the bound complex; Unbound indicates that backbone coordinates were taken from one or both unbound monomer components.<sup>2</sup>



Perturbation studies on antibody/antigen complexes.<sup>2</sup>

RosettaDock allows full side chain flexibility during the docking and it was the first docking protocol to explicitly model all side chain conformations at the interface. The resulting predictions of RosettaDock were tested successfully<sup>5,9</sup> in CAPRI (<http://capri.ebi.ac.uk>), a community that assesses different docking protocols. High-resolution modeling of protein complexes is now feasible, provided that the backbone of the protein partners does not change significantly upon binding.

Current development concentrates on backbone flexibility and conformational changes.

RosettaDock has been extended to allow not only Protein-Protein docking, but also the docking of ligands to proteins,<sup>13</sup> and proteins to crystal surfaces.<sup>14</sup>

## REFERENCES

1. Buckle, A. M.; Schreiber, G.; Fersht, A. R., Protein-Protein Recognition - Crystal Structural-Analysis of a Barnase Barstar Complex at 2.0-Angstrom Resolution. *Biochemistry* **1994**, 33, (30), 8878-8889.
2. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology* **2003**, 331, (1), 281-299.
3. Dunbrack, R. L., Jr.; Cohen, F. E., Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **1997**, 6, (8), 1661-81.
4. Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D., Protein structure prediction using rosetta. *Numerical Computer Methods, Pt D* **2004**, 383, 66-+.
5. Schueler-Furman, O.; Wang, C.; Baker, D., Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins* **2005**, 60, (2), 187-94.
6. Gray, J. J.; Moughon, S. E.; Kortemme, T.; Schueler-Furman, O.; Misura, K. M.; Morozov, A. V.; Baker, D., Protein-protein docking predictions for the CAPRI experiment. *Proteins* **2003**, 52, (1), 118-22.
7. Kuhlman, B.; Baker, D., Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, 97, (19), 10383-10388.
8. Carvalho, A. L.; Dias, F. M. V.; Prates, J. A. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romao, M. J.; Fontes, C. M. G. A., Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, 100, (24), 13809-13814.
9. Daily, M. D.; Masica, D.; Sivasubramanian, A.; Somarouthu, S.; Gray, J. J., CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock. *Proteins-Structure Function and Bioinformatics* **2005**, 60, (2), 181-186.
10. Wang, C.; Schueler-Furman, O.; Baker, D., Improved side-chain modeling for protein-protein docking. *Protein Sci* **2005**, 14, (5), 1328-39.
11. Sivasubramanian, A.; Chao, G.; Pressler, H. M.; Wittup, K. D.; Gray, J. J., Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure* **2006**, 14, (3), 401-14.
12. Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C. E.; Baker, D., Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins* **2001**, Suppl 5, 119-26.
13. Meiler, J.; Baker, D., ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins-Structure Function and Bioinformatics* **2006**, 65, (3), 538-548.
14. Makrodimitris, K.; Gray, J. In *Modeling statherin structure binding to hydroxyapatite [001] crystal surface* <http://oasys2.confex.com/acs/232nm/techprogram/P986054.HTM>, 232nd ACS National Meeting, (Molecular Mechanics and Simulation, Computers in Chemistry Division), San Francisco, CA, 2006; ACS, Ed. American Chemical Society:2006.

These cases for RosettaDock prepared and benchmarked in different platforms in 2007 by:

Kosta Makrodimitris, Ph.D.

Johns Hopkins University, Chemical & Biomolecular Engineering

[kosta@jhu.edu](mailto:kosta@jhu.edu)