John von Neumann Institute for Computing

Analysis and Optimization of the Flex-Screen Docking Approach Using DUD Benchmarking Database

D. Kokh, B. Fischer, W. Wenzel

published in

From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **36**, ISBN 978-3-9810843-2-0, pp. 181-183, 2007.

http://www.fz-juelich.de/nic-series/volume36

Analysis and Optimization of the Flex-Screen Docking Approach Using DUD Benchmarking Database

Daria Kokh[1], **Bernhard Fischer**[2], **and Wolfgang Wenzel**[2]

[1] FB.C - Mathematik und Naturwissenschaften,
Bergische Universität Wuppertal, 42097 Wuppertal, Germany
E-mail: kokh@uni-wuppertal.de

[2] Institut für Nanotechnologie, Forschungszentrum Karlsruhe, GmbH,
Postfach 3640, 76021 Karlsruhe, Germany
E-mail: {*bernhard.fischer, wenzel*}@int.fzk.de

Screening performance of the all-atom Flex-Screen docking approach including receptor flexibility is investigated by using of a "directory of useful decoys", DUD. DUD is a bias-corrected benchmarking database based on 40 different target proteins, where each receptor is associated with a native ligand, a set of annotated ligands, and a set of decoy molecules that are unlikely to be binders. The docking performance is evaluated using two criteria: 1) geometrical fidelity of the docked poses compared to those of the experimental structures; 2) enrichment of annotated ligands among their decoys, which shows the ability of the docking calculations to distinct between true positives and nonbinder molecules with the same physical properties. Based on these results the scoring function and the receptor side-chain rearrangement procedure are optimized.

1 Method

FlexScreen[1] is an all-atom docking approach based on the stochastic tunneling method of the energy minimization and a simple atomistic scoring function that contains a sum of the Van-der-Waals, electrostatic, hydrogen-bond and salvation energies. The VdW and hydrogen-bond parameters are taken from OPLSAA[2] and AutoDock[3], respectively, the partial charges of the receptors are computed with MOE, and the atomic salvation parameters are optimized as described below. The method enables rotation up to 15 side-chain bonds of the receptor.
Scoring performance of the Flex-Screen approach is benchmarked by using of the DUD database[4] based on 40 target proteins of different classes with available ligand-bound X-ray crystal structures. For each protein the database includes: 1) crystal structures of the receptor and its native ligand; 2) a set of the annotated ligands that should in principle dock well (15-450 molecules); 3) a set of the decoys (about 36 molecules for each annotated ligand) that resemble the particular ligand in physical properties, but differ from the ligand topologically, so that they are likely to be nonbinders.

2 Results

2.1 Optimization of the Salvation Energy Parameters

Salvation energy is described as a sum of energies for the individual atoms that are assumed to be proportional to the solvent accessible surface area and atomic salvation parameter, ASP[5]. All atoms are divided into two groups: those responsible for 1) hydrophobic and

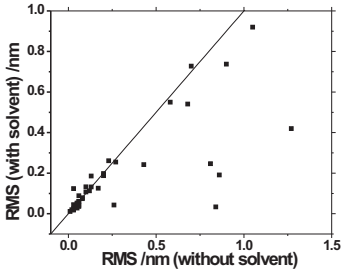

Figure 1. RMS deviation of the docked poses from the crystal ones for optimized ASPs versus the same values computed without salvation energy. 40 protein-ligand structures are included.

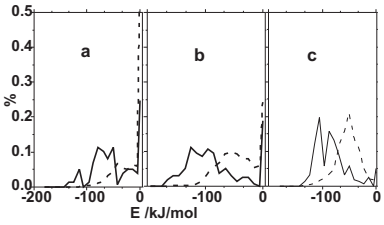

Figure 2. Distributions of annotated ligands in percentage (solid lines) and decoys (dashed lines) of the androgen receptor plotted as a function on their binding energies: (a) receptor is rigid; (b) 10 receptor side chains are flexible; (c) 15 receptor residues are shifted by 0.5 nm away from the cavity centre.

2) hydrophilic effects, so that only two ASPs have to be optimized. As an optimization criterion we use the sum of the RMS deviations of the docked conformations of the native ligands from the experimental ones.

Inclusion of the salvation energy in the scoring function in general improves docking poses for most of the ligands (see Fig.1), although some of them (10%) fail to find correct conformation regardless salvation effects and, therefore, need additional analysis.

2.2 Receptor Side Chain Rearrangement

Docking screen of the annotated ligand sets has shown that in some cases binding mode cannot be found because of high-energy clashes between protein and ligand atoms arising from the VdW term. To improve docking efficiency we enable receptor rearrangement by using of two approaches:1) rotation of up to 15 receptor side-chain bonds, 2) shift (by about 0.25-0.5 nm) of the receptor residues that are involved in clashes away from the binding pocket centre. Fig.2 demonstrates how both methods influence docking efficiency of annotated ligands and their decoys.

It is important to note, that receptor flexibility either improves or at least does not change scoring performance of the docking method. Although both approaches help to reduce a number of nondocking ligands and increase absolute value of the binding energy, the second method has been found to be usually more effective and notably less expensive.

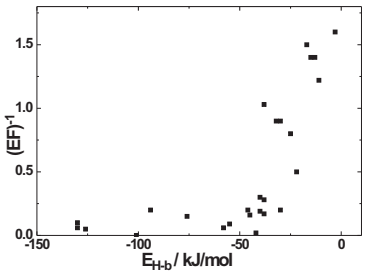

Figure 3. EF^{-1} for 28 receptors plotted as a function of the average energy of hydrogen bonds formed by docked ligands.

2.3 Enrichment of Annotated Ligands Among their Own Decoys

Efficiency of the docking method selectivity is estimated by computing of the inverse enrichment factor, EF^{-1}, for each receptor, described as a relation of the top-scoring decoys to the top-scoring annotated ligands (expressed as a percentage of the total number of the decoys and ligands, respectively). "top-scoring" molecule means that the absolute value of its binding energy is larger than at least 80% of that for the corresponding native ligand. A small value of EF^{-1} shows that annotated ligands dock notably better than their decoys for the particular receptor, whereas $EF^{-1} > 1$ indicates that the docking approach cannot distinguish between molecules with similar physical but different chemical properties.
Docking results of the preliminary calculations without salvation effects, summarized in Fig.3, indicate that: 1) the screening performance of the method is quite effective for about 70% of the targets, and 2) the docking efficiency depends strongly on the hydrogen-bond energy, e.g. essentially all molecules can find appropriate binding mode by unspecific interaction. Since most of the receptors with large value of EF^{-1} have open binding pockets, we expect that inclusion of the salvation energy in the scoring function will improve docking performance in these cases. These calculations are in progress.

References

1. H. Merlitz and W. Wenzel, *High Throughput in-silico Screening against Flexible Protein Receptors* , Lecture Notes in Computer Science **465**, 3045, 2004.
2. W. L. Jorgensen and N. A. McDonald, *Development of an all-atomic force field for heterocycles properties of liquid pyridine and diazenes* , J. Mol. Struct. **424**, 145, 1997.
3. G. M. Morris, D. S. Goodsell, R. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding energy function* , J. Comptut. Chem. **19**, 1639, 1998.
4. N. Huang, B. K. Shoichet, and J. J. Irwin *Benchmark sets for molecular docking* , J. Med. Chem. **49**, 6789-6801, 2006.
   `http://blaster.docking.org/dud/`
5. D. Eisenberg and A. D. McLachlan, *Comparison of of stochastic optimization methods for receptor-ligand docking* , Nature **319**, 199-203, 1986.