Time-Efficient Docking of Similar Flexible Ligands

Matthias Rarey, Bernd Kramer, Christoph Bernd, and Thomas Lengauer

German National Research Center for Information Technology (GMD)
Institute for Algorithms and Scientific Computing (SCAI)
53754 Sankt Augustin, Germany
{Matthias.Rarey | Bernd.Kramer | Christoph.Bernd | Thomas.Lengauer}@gmd.de

We present a new method for predicting binding modes and energies for a set of similar ligands if the structure of the receptor is known. The method is an extension of a flexible docking algorithm, called FlexX, which is currently under development at GMD [1]. FlexX follows an *incremental construction* strategy, which consists of three phases.

The first phase, the *base selection*, selects a connected part of the ligand, called the *base fragment*. This phase is the only one, which is not yet automated. In the *base placement* phase, a set of energetically favorable placements of the base fragment into the active site is computed. The algorithm in this phase has been adapted from applications in the area of pattern recognition and is described in [2]. In the third phase, the *complex construction*, we use a simple, iterative greedy strategy in order to add the remaining part of the ligand. This strategy was first used in the de-novo (peptide) design tool GROW [3]. The ligand is divided into fragments $f_0, \ldots, f_n$, such that f_0 is the base fragment and each set of fragments $f_0, \ldots, f_i$ is a connected part of the ligand. In the i-th iteration, the fragment f_i is joined to all placements found so far among all possible conformations of a finite energetically favorable set. The model of conformational flexibility of the ligand is equal to that of the conformational search program MIMUMBA, developed by Klebe [4]. For the new placements obtained, the binding energy is estimated by an empirical energy function adapted with minor changes from Böhm [5] and the 500 highest ranking placements are taken into the next iteration. Various algorithmic techniques are used to make this approach time-efficient. An extended description of the complex construction algorithm can be found in [1].

In our current version, a typical drug molecule with a flexible ring system and about 10 rotatable bonds can be docked in 2–3 minutes on a workstation. Energy estimations and RMS deviations for a set of known receptor-ligand complexes taken from the PDB are shown in table 1. The complete run time for all examples sums up to about 19 minutes on a SUN SPARC-station 20.

For docking a set of similar ligands a preprocessing step detects similar parts of the ligands such that the molecules can be arranged in a tree structure: Each leaf of the tree represents a ligand, an interior tree node represents the common part of all ligands rooted in it. This structure enables the docking algorithm to reuse partial results achieved in the docking of previous ligands. With this preprocessing step, a significant speedup of a factor 2–3 can be achieved for a set of similar ligands. Thus, the amortized run time is below one minute per ligand on a workstation. The approach has been tested with a set of 108 thrombin inhibitors with known binding constants.

Although the correlation between the predicted and experimental binding energies is relatively weak, the program predicts the right tendency of the binding strength and can therefore be used for filtering out bad candidates from the ligand set (see fig. 1). Note that we use a generally applicable energy function which is not calibrated to a special receptor by learning data. In addition, the program produces a set of possible binding modes for each ligand just as in the single-ligand-docking mode, which can be of help in understanding the binding process and in designing new potential inhibitors.

Keywords: molecular docking, flexible docking, receptor ligand interaction, molecular flexibility, conformational analysis, drug design, protease inhibitors, free energy of binding

References

[1] M. Rarey, B. Kramer, and T. Lengauer. In *Proc. of the 3rd International Conference on Intelligent Systems in Molecular Biology*, pages 300–308. AAAI Press, Menlo Park, California, 1995.

[2] M. Rarey, S. Wefing, and T. Lengauer. Arbeitspapier 897, GMD, 1995, accepted for *J. Comput.-Aided Mol. Design.*

[3] J.B. Moon and W.J. Howe. *Protein Struct. Funct. Genet.*, 11:314–328, 1991.

[4] G. Klebe and T. Mietzner. *J. Comput.-Aided Mol. Design*, 8:583–606, 1994.

[5] H.-J. Böhm. *J. Comput.-Aided Mol. Design*, 8:243–256, 1994.

PDB entry	# Solut.	best ΔG ΔG	RMSD	RMSD $\leq$ 1.1 ΔG	RMSD	rank	ΔG_{exp}
1DWD	319	-39.6	1.33	-37.2	0.90	12	-48.6
1LDM	15	-37.6	0.66	-37.6	0.66	1	-30.8
1RNT	81	-37.6	2.06	-28.3	1.39	32	-28.5
1STP	159	-30.2	1.09	-30.2	1.09	1	-76.4
1TMN	237	-39.5	2.00	-38.0	1.09	7	-41.6
1ULB	76	-18.0	0.66	-18.0	0.66	1	-30.2
2CTC	37	-30.8	2.14	-29.1	0.65	6	-22.2
2PHH	42	-22.9	0.66	-22.9	0.66	1	-26.7
3CPA	218	-37.8	3.08	-29.5	1.72	94	-22.1
3PTB	20	-24.5	0.56	-24.5	0.56	1	-27.2
3TPI	221	-22.5	1.32	-20.6	1.05	21	-24.5
4DFR	84	-63.6	0.90	-63.6	0.90	1	-55.3
4PHV	13	-38.4	1.08	-38.4	1.08	1	-52.2
4TLN	105	-31.6	5.44	-17.5	1.61	95	-21.2
4TS1	15	-20.8	0.54	-20.8	0.54	1	-32.0
5TIM	3	-12.9	1.60	-10.1	0.85	2	-13.1
6RSA	153	-33.7	8.44	-31.0	0.86	11	-28.5

Table 1: Protein-ligand complexes used as test cases for the flexible docking algorithm.
Columns: number of solutions, ΔG, and RMS deviations from the crystal structure of the energetically highest ranking solution, the highest ranking solution with an RMSD $\leq$ 1.1Å (or solution with best RMSD if no solution with RMSD $\leq$ 1.1Å found), and the experimentally determined free binding energy (ΔG is given in kJ/mol, RMSD in Å).
Comments on the examples: 4DFR: water molecule 403 and 405 included; 4PHV: water molecule 1 included; 6RSA: ring systems of the ligand held fixed in given conformation because of the vanadium atom in the ring system.

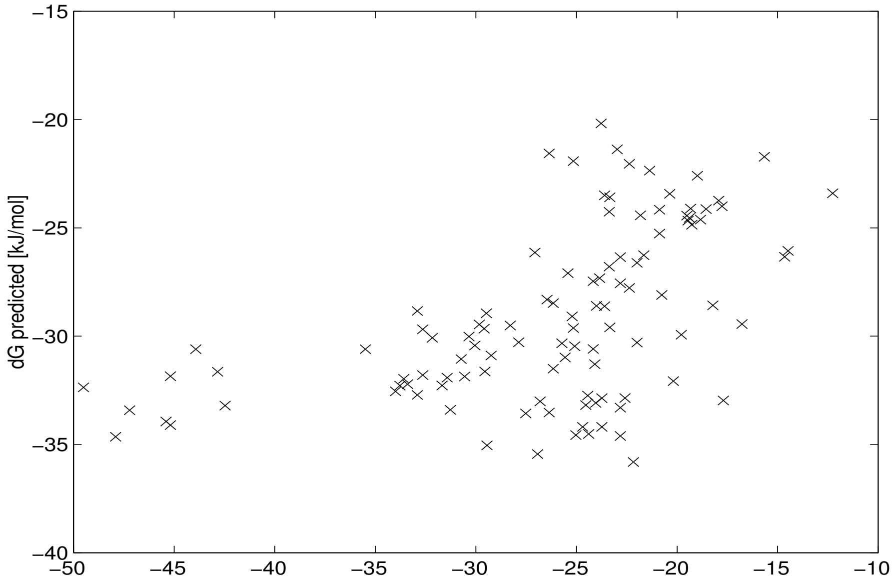

Figure 1: Comparison of predicted ΔG with experimental ΔG for a set of 108 potential thrombin inhibitors.