Supporting Information to the manuscript

'PARADOCKS – A framework for molecular docking with population-based metaheuristics'

René Meier,[*,†,‡,⊥] Martin Pippel,[†,⊥] Frank Brandt,[¶] Wolfgang Sippl,[†] and Carsten Baldauf[*,§,‖,¶]

Department of Pharmaceutical Chemistry, Martin-Luther Universität Halle-Wittenberg, Wolfgang-Langenbeckstr. 4, 06120 Halle/Saale, Germany, Research Center Pharmaceutical Engineering GmbH, Inffeldgasse 21a/II, 8010 Graz, Austria, Biotechnologisches Zentrum der TU Dresden, Tatzberg 47/49, 01307 Dresden, Germany, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, 200031 Shanghai, P. R. China, and Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

E-mail: rene@paradocks.org; caba@paradocks.org

[*]To whom correspondence should be addressed
[†]MLU Halle-Wittenberg
[‡]RCPE Graz
[¶]BIOTEC TU Dresden
[§]CAS-MPG PICB, Shanghai
[‖]HITS, Heidelberg
[⊥]Contributed equally to this work

Ligand Conformation

The method to generate ligand conformations produces at least one ligand conformation with an *rmsd* of less than 2 Å for 89 % of all complexes from the PDBbind *core set*.[1] The complete *rmsd* distribution of the docking results of the PDBbind *core set* is shown in fig. Figure 1.

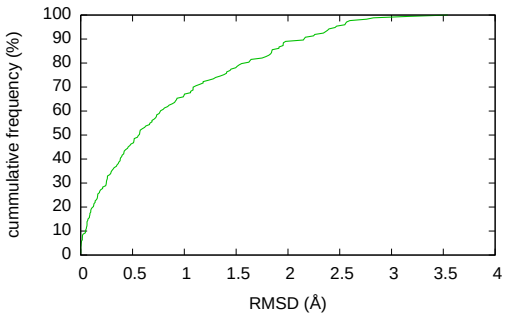

Figure 1: Distribution of lowest *rmsd* value of all docking solutions and the respective crystal structure calculated by a least square fitting algorithm.

Timings and parallel efficiency

The computing time is of innate importance for the application of molecular docking techniques especially when performing virtual screenings with large libraries of compounds. Parallel computing offers a chance for speed up and is especially interesting as there is a clear trend towards multi-core CPUs and workstations.

As an example we used a docking simulation with 50 consecutive runs of the non-nucleoside reverse transcriptase inhibitor TNK-651 to HIV-1 reverse transcriptase (PDB: 1JLA).[2] As test platform a HP server with 2.53 GHz Intel Xeon CPUs was used. Employing one processor, the computing time of GOLD for the given problem is slightly less than half of the computing time PARDOCKS needs. As PARADOCKS is capable of parallel processing, one way to overcome the speed limitation is the use of multiple cores. As apparent from Table 1, PARADOCKS with 4 CPU cores reaches the speed of GOLD with 1 CPU core. With numbers of processors above 8, the scaling basically collapses. Reason for that is the high overhead of communication compared to the effectice computing time for fitness function evaluation.

Table 1: Timings and speed-ups for PARADOCKS/p-Score and GOLD.

Docking approach	# CPUs	Time (s)	Speed-up
	1	3555	
	2	2392	1.5
PARADOCKS	3	1857	1.9
	4	1660	2.1
	6	1392	2.6
	8	1228	2.9
GOLD	1	1591	

p-Score Parameter

Table 2: d_{ij} parameter in p-Score

element	geometry	d_{ij} in Å
C	tetrahedral	2,1
C	trigonal planar aromatic	2,0
C	trigonal planar	1,9
C	trigonal planar	1,9
C	linear	1,8
N	tetrahedral	1,8
N	trigonal planar	1,75
N	linear	1,75
O	tetrahedral	1,65
O	trigonal planar	1,55
O	in H_2O	1,75
S	tetrahedral	2,1
S	trigonal planar	2,0
P	any	2,0
F	any	1,5
Cl	any	1,75
Br	any	1,9
I	any	2,05
Si	any	2,0
metall ion	any	1,25

The following parameter were used in Equation (5), Equation (6) and Equation (7):

- $k_1 = 0.9\mathring{A}$
- $k_2 = 170^\circ$
- $k_3 = 170^\circ$.

References

(1) Wang, R.; Fang, X.; Lu, Y.; Wang, S. *J. Med. Chem.* **2004**, *47*, 2977–2980.

(2) Ren, J.; Nichols, C.; Bird, L.; Chamberlain, P.; Weaver, K.; Short, S.; Stuart, D. I.; Stammers, D. K. *J. Mol. Biol.* **2001**, *312*, 795–805.

(3) Sippl, W.; Contreras, J. M.; Parrot, I.; Rival, Y. M.; Wermuth, C. G. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 395–410.

(4) Contreras, J. M.; Parrot, I.; Sippl, W.; Rival, Y. M.; Wermuth, C. G. *J. Med. Chem.* **2001**, *44*, 2707–2718.

(5) Thomson Scientific, Philadelphia, PA, USA, *World Drug Index*, 2001.