

# Blind docking of drug-sized compounds to proteins with up to a thousand residues

Csaba Hetényi<sup>a,\*</sup>, David van der Spoel<sup>b</sup>

<sup>a</sup> Department of Biochemistry, Eötvös Loránd University, 11C Pázmány P. sétány, 1117 Budapest, Hungary

<sup>b</sup> Molecular Biophysics Group, Department of Cell and Molecular Biology, Uppsala University, Box 596, 75124 Uppsala, Sweden

Received 14 November 2005; revised 20 December 2005; accepted 24 January 2006

Available online 31 January 2006

Edited by Robert B. Russell

**Abstract** Blind docking was introduced for the detection of possible binding sites and modes of peptide ligands by scanning the entire surface of protein targets. In the present study, the method is tested on a group of drug-sized compounds and proteins with up to a thousand amino acid residues. Both proteins from complex structures and ligand-free proteins were used as targets. Robustness, limitations and future perspectives of the method are discussed. It is concluded that blind docking can be used for unbiased mapping of the binding patterns of drug candidates.

© 2006 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

**Keywords:** Active site; Binding pocket; Rational design; Pharmacophore; AutoDock; Molecular interaction

## 1. Introduction

In silico molecular docking is one of the most powerful techniques of structure-based drug design [1]. Most applications of docking tools focus on the (supposed) primary binding region. However, there are cases in which the information on the binding region is missing. The AutoDock [2]-based blind docking (BD) approach [3] was introduced previously to search the entire surface of proteins for binding sites while simultaneously optimizing the conformations of the peptides. The results of BD were regarded as “very encouraging” in a recent review [4]. BD [5–7] and the recommended search parameters [8–12] have been used for solving various problems such as design of inhibitors [5], comparison of microtubule-stabilizing agents [7] and exploring substrate binding modes [8]. Because of the apparent success of the approach [4–12], we decided to perform further systematic tests on a set of 43 ligand–protein complexes which was previously used in a comprehensive study on the selectivity of binding of aromatic compounds [13]. In the previous study [13] searching was restricted to the surrounding of the primary binding site

(for a list of the complexes, refer to [Supplementary material, Table A](#)), here we use BD on the entire protein surfaces. The set contains drug-sized aromatic ligands with relatively few free rotations. All of these have some (positive or negative) biological effects and some (e.g., cancer drugs tamoxifen in 3ert and methotrexate in 4dfr) of them are actual medicines (see [Fig. 1](#)).

## 2. Methods

In the present study, the original parameters [3] of BD ([Supplementary material, Table C](#)) were used in combination with an evaluation scheme based on binding free energy ( $\Delta G$ ) and root mean square deviation (RMSD) calculated between the crystallographic and the docked ligand conformations (RMSD, [Supplementary material, Scheme A](#)). By definition, the entire protein surfaces were subjected to the BD search. For every 5th complexes of [Table A](#) (starting with the 1st row) and for the system with the largest protein (1b70) coordinates of the ligand-free proteins were obtained from the protein databank (PDB). In two cases, where the unbound proteins were not available (complexes 1a0q and 1gaf) the next systems (1a53 and 1guh, respectively) were involved in the study. The selected ten ligand-free proteins ([Supplementary material, Table B](#)) were superimposed on the corresponding protein–ligand complex structures and used for BD as described previously.

## 3. Results and discussion

### 3.1. Test of BD search on large protein targets

The results of the BD calculations on the 43 proteins (ligand-bound conformations, marked with the corresponding PDB codes, [Table A](#)) are summarized in [Table 1](#) (for a detailed list of results, refer to [Supplementary material, Table D](#)). For 34 of 43 systems the BD search identified the crystallographic binding site and mode of ligands as the energy minimum of the whole BD job, i.e., all 100 docking trials (runs). In terms of averages and standard deviations ([Table 1](#)) the corresponding ranks contain energetically uniform members, with a significant population in most of the cases. In six of the remaining nine cases (1dy4, 1e7a, 1eqg, 1livb, 1ngp, 3pcn) the native ligand position was ranked in the best 2nd–7th ranks and in three cases (1hz4, 1ju4, 1pth) an additional 1–3 accumulative BD jobs were necessary to locate the native binding mode (for details of accumulative BD refer to [Supplementary material, Scheme A](#)). In two out of the nine cases (1dy4, 1ngp) the native binding mode was also placed in the 2nd rank in the restricted docking study [13] indicating

\*Corresponding author. Fax: +36 1 3812172.

E-mail address: [csabahete@yahoo.com](mailto:csabahete@yahoo.com) (C. Hetényi).

**Abbreviations:** BD, blind docking; PDB, protein databank; RMSD, root mean square deviation



Fig. 1. The match of the crystallographic (red) and the minimum energy blind docked (yellow) conformations of methotrexate, the largest ligand molecule investigated (system 4dfr). The size of the ligand molecules did not affect the results of BD in the present study.

that the reason of these results is not the insufficient BD search. The average  $\text{RMSD}_{\text{min}}$  (RMSD corresponding to the energy minimum of the rank) of the 43 systems ( $1.0 \pm 0.7 \text{ \AA}$ ) is similar to the value calculated from the results of restricted docking [13] for the same set ( $1.2 \pm 0.7 \text{ \AA}$ ). This comparison shows, that in the case of drug-sized compounds, both the AutoDock scoring function and the Lamarckian genetic algorithm with the pseudo-Solis and Wets local search method can be applied to the large BD search space, i.e., the whole target surface solely by tuning the search parameters (Supplementary material, Table C). In the original BD study [3] the largest protein was 316 AA. In the present study, proteins with up to 1040 residues were involved in the calculations and 16 of 43 systems have more than 316 AAs. In seven of these 16 cases, including the largest protein investigated (Fig. 2) the native binding conformation was in the 1st rank, i.e., as the energy minimum of 100 trials. In the other nine cases the binding mode was correctly reproduced in terms of RMSD, but placed in higher ranks due to higher binding energy (for explanation, refer to Section 3.4).

### 3.2. Protein flexibility: robustness and limitations

BD to the 10 ligand-free protein structures (marked with U in Table 1 and Table D) provides additional information on the sensitivity of BD on protein flexibility. Such information may be useful for the situations, where only the unbound protein is available for the calculations, as expected for most real applications. In eight of the selected 10 cases the ranking of docked conformations with the best RMSD-s were identical or lower (better) compared to the results obtained for the corresponding proteins from complexes (previous section, Table 1) which demonstrates the robustness of BD. In two cases (1b70<sup>U</sup> and livb<sup>U</sup>) the best-RMSD-solution moved to higher ranks (rank serial numbers increased with 2 and 1, respectively). At 1b70<sup>U</sup>, a turn of  $180^\circ$  (relative to the 1b70 complex) of the amide group of a central glutamine residue spoiled the favorable H-bonding pattern with the ligand at the binding site (Fig. 2). This resulted in higher  $\Delta G$ -s and higher ranking if compared with 1b70 (Table 1). However, the corresponding RMSD has not increased dramatically, due to the remaining (e.g., hydrophobic) interactions at the site. It should be remarked, that in these systems only moderate changes can be observed between the bound and li-

gand-free protein structures (see  $C_\alpha$ -RMSD-s in Table B, Supplementary material). For these systems with moderate flexibility in the active site BD proved to be robust, but obviously BD alone may prove insufficient for systems with a higher degree of induced fit upon ligand binding. To overcome this problem, methods which handle structural flexibility [14] could be used in post-docking mode with the (prerequisite) binding positions and conformations of ligands found by BD as input.

### 3.3. Ligand flexibility

Neither the number of flexible torsions in the ligands (tabulated in Table A, Supplementary material), nor the size of the ligands affects the accuracy of the results of BD for the investigated systems (Fig. 1). The computational cost (efficiency) of the BD runs does depend on ligand flexibility. For systems with the smallest (1mpj) and largest (4dfr) ligand molecules BD runs took 5 and 22 min (Opteron 2 GHz), respectively.

### 3.4. Competition for the binding sites between the ligand and solvent molecules. Multiple binding sites

It should be remarked, that docking calculations generally use ‘dry’ protein molecules for the search, i.e., all ions, water molecules etc., are removed from the coordinate files before docking. In six out of nine cases where the native binding mode did not belong to the 1st rank, inspection of the original PDB files showed, that the low-energy binding sites of the first ranks found for the ligand during BD are occupied by water molecules (or other solvent) in the PDB structure. This can be due to the energetically favorable protein–solvent interactions at those sites, but it is also possible that the crystallographic complexes do not include all binding sites/modes of the ligands. In the systems (1ev3, 1mpj, 1qiz, 1tym) where insulin oligomers were used as targets in this study, multiple crystallographic binding sites at the protein interfaces were reproduced, showing the applicability of BD for multiple binding site search. Although some methods have been proposed for the modeling of ligand–solvent competition ‘on-line’, i.e., during docking simulations [15], or ‘off-line’ with mixed maps for the restricted search space [16], there is no trivial solution for BD yet. However, there is no alternative to using a dry target if multiple sites are searched for since water molecules covering the putative sites may hinder entrance of the ligand molecules.

### 3.5. Recommendations for BD of drugs

(1) In 3 cases (1hz4, 1ju4, 1pth) additional, accumulative BD jobs were necessary to find the native ligand conformation. In these cases the previously found representative ligand conformations (one per rank) were merged with the protein structure and these molecular complexes were used as docking targets in the next job. This procedure can be useful in BD calculations aimed at mapping all possible binding sites and can be automated by setting a limit criterion in terms of, e.g., binding free energy (Supplementary material, Scheme A). (2) In general,  $0.55 \text{ \AA}$  grid spacing (Supplementary material, Table B) was adequate for the BD search of the drug-sized compounds in the present study to obtain acceptable RMSD-s. However, in one case (1ju4) a re-docking was performed for the located binding site with  $0.375 \text{ \AA}$  grid spacing and the fit was refined from  $4.136$  to  $0.629 \text{ \AA}$  (Supplementary material, Table D).

Table 1  
Results of the blind docking calculations (abridged)

PDB	Job #	Rank #	$\Delta G_{\min}$	RMSD <sub>min</sub>	Population	$\Delta G_{\text{avg}}$	$\Delta G_{\text{sdev}}$
1a0q	1	1	-9.16	2.212	16	-8.96	0.21
1a53	1	1	-10.03	0.646	50	-9.61	0.29
1a53 <sup>U</sup>	1	1	-10.55	1.223	54	-9.91	0.45
1a8u	1	1	-6.48	0.439	100	-6.48	0.00
1alw	1	1	-6.41	2.658 <sup>a</sup>	86	-6.23	0.10
1az8	1	1	-11.99	0.544	83	-11.48	0.23
1az8 <sup>U</sup>	1	1	-11.26	0.987	82	-10.59	0.33
1b70	1	1	-8.72	0.891	42	-8.64	0.05
1b70 <sup>U</sup>	1	3	-7.15	1.023	22	-7.06	0.06
1bjz	1	1	-12.88	0.567	100	-12.80	0.04
1c83	1	1	-11.30	0.541	100	-11.13	0.07
1c84	1	1	-10.45	0.862	93	-10.11	0.24
1c85	1	1	-9.96	0.741	100	-9.86	0.05
1c85 <sup>U</sup>	1	1	-8.95	1.557	70	-8.76	0.20
1ca7	1	1	-7.89	0.854	91	-7.84	0.04
1d1q	1	1	-10.85	0.545	99	-10.75	0.08
1dy4	1	2	-8.79	0.777	13	-8.37	0.36
1e7a	1	2	-6.07	1.023	73	-6.03	0.03
1ecv	1	1	-11.24	0.674	100	-10.85	0.30
1ecv <sup>U</sup>	1	1	-8.64	1.166	25	-8.07	0.48
1eqg	1	4	-7.64	0.727	56	-7.59	0.02
1ev3	1	1	-4.95	1.075	10	-4.95	0.01
1f5k	1	1	-7.45	0.432	57	-7.45	0.00
1fiw	1	1	-9.00	0.832	100	-8.99	0.01
1gaf	1	1	-10.00	0.409	62	-9.46	0.39
1guh	1	1	-11.50	0.792	17	-10.58	0.66
1guh <sup>U</sup>	1	1	-11.01	1.180	22	-10.06	0.90
1hd2	1	1	-5.32	0.739	100	-5.31	0.01
1hdu	1	1	-8.60	0.525	68	-8.42	0.15
1hz4	2	3	-5.42	0.490	1	-5.42	—
1ivb	1	3	-6.46	0.200	26	-6.31	0.10
1ivb <sup>U</sup>	1	4	-5.47	2.717	28	-5.35	0.11
1ju4	3 <sup>b</sup>	1	-5.09	0.629	7	-5.09	0.00
1kel	1	1	-12.25	1.932	55	-11.32	0.65
1mpj	1	1	-3.88	0.465	54	-3.87	0.01
1ngp	1	2	-7.35	0.691	36	-7.23	0.11
1pth	4	8	-3.95	2.450	3	-3.95	0.01
1pth <sup>U</sup>	3	4	-4.60	2.660	5	-4.59	0.01
1qiz	1	1	-4.63	2.481	25	-4.61	0.01
1rfn	1	1	-8.61	0.573	100	-8.60	0.00
1sri	1	1	-9.06	1.006	47	-8.63	0.27
1tnj	1	1	-7.47	1.964	84	-7.27	0.07
1tym	1	1	-5.89	1.830	71	-5.79	0.06
1tym <sup>U</sup>	1	1	-5.06	1.919	12	-5.01	0.06
2ay5	1	1	-9.50	2.085	22	-9.18	0.15
3cpa	1	1	-8.74	0.757	44	-8.23	0.22
3ert	1	1	-9.84	1.646	58	-9.38	0.23
3pax	1	1	-6.14	1.208	100	-6.02	0.05
3pcn	1	7	-5.11	2.568	12	-5.00	0.06
3pcn <sup>U</sup>	1	2	-5.24	2.658	3	-4.94	0.27
43ca	1	1	-5.17	0.419	100	-5.16	0.01
4dfr	1	1	-13.35	1.086	19	-12.54	0.93
4ts1	1	1	-6.94	0.504	76	-6.68	0.13

PDB, protein databank code; U, unbound (ligand-free) protein; Job #, number of the accumulative jobs; Rank #, serial number of the Rank;  $\Delta G_{\min}$ , the minimum of AutoDock free energy of binding (kcal/mol) values of the members of Rank; Population, population of the Rank (the maximum value is 100 corresponding to a docking job, i.e., 100 docking runs); RMSD<sub>min</sub>, root mean square deviation (Å) of the conformation conjugated to  $\Delta G_{\min}$ . Averages ( $\Delta G_{\text{avg}}$ ) and standard deviations ( $\Delta G_{\text{sdev}}$ ) are calculated for the rank.

<sup>a</sup>The crystallographic ligand used for comparison has erroneous structure.

<sup>b</sup>In case of 1ju4, Job 3 was a re-docking with 0.375 Å grid spacing focused on the previously located (Job 2:Rank 2) binding site.

Such re-dockings are of limited computational cost (10 docking runs usually suffice) and can be recommended for all BD studies. (3) In general, post-docking refinement with, e.g., normal mode methods [14] accounting for protein flexibility at the docked complexes may be advantageous to increase precision of ranking.

### 3.6. Future applications of BD

In combination with experimental techniques such as site-directed mutagenesis, BD can be a useful tool for mapping of binding modes of drug candidates on protein targets and even the selection of new protein targets (protein screening [13]) for existing drugs.

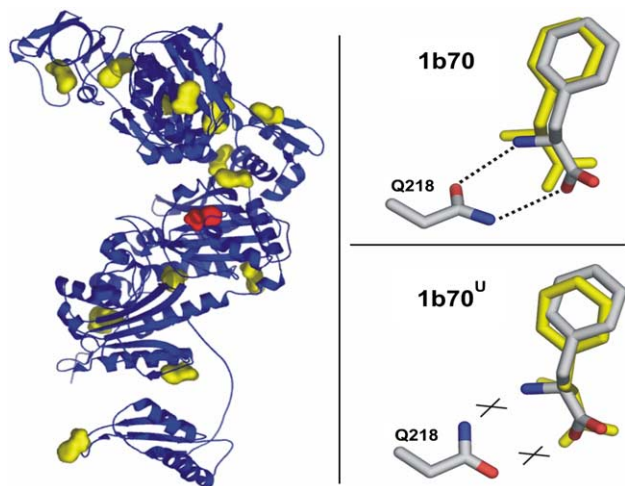


Fig. 2. The result of blind docking for phenylalanyl-tRNA synthase (blue cartoon, 1b70), a protein with more than a thousand amino acids. Representative ligand conformations of each rank and the crystallographic one are depicted as yellow and red surfaces, respectively (on the left). Due to the large protein surface, numerous putative sites can be found among which the crystallographic site was identified in the 1st or 3rd best ranks using the bound or ligand-free proteins as targets, respectively. The blind docked ligand conformations (yellow sticks) have good match with the crystallographic ligand conformation (sticks colored by atom type) if using either the bound (1b70, top on the right) or the ligand-free (1b70<sup>U</sup>, bottom on the right) protein structures. In case of 1b70<sup>U</sup> the amide group of the key H-bonding glutamine (Q218) residue is turned with ca. 180° hindering formation of the H-bonds (dotted lines) which exist in the complex form (1b70) and cause a higher  $\Delta G$  value when docking to 1b70<sup>U</sup>. Figures were prepared using PyMol [17].

*Acknowledgment:* This work was financed by the Eötvös Fellowship of the Hungarian State.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.febslet.2006.01.074](https://doi.org/10.1016/j.febslet.2006.01.074).

#### References

[1] Brooijmans, N. and Kuntz, I.D. (2003) Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 32, 335–373.

- [2] Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662.
- [3] Hetényi, C. and van der Spoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* 11, 1729–1737.
- [4] Campbell, S.J., Gold, N.D., Jackson, R.M. and Westhead, D.R. (2003) Ligand binding: functional site location, similarity and docking. *Curr. Opin. Struct. Biol.* 13, 389–395.
- [5] Jining, L., Makagiansar, I., Yusuf-Makagiansar, H., Chow, V.T.K., Siahaan, T.J. and Jois, S.D.S. (2004) Design, structure and biological activity of  $\beta$ -turn peptides of CD2 protein for inhibition of T-cell adhesion. *Eur. J. Biochem.* 271, 2873–2886.
- [6] Brown, W.M. and Vander Jagt, D.L. (2004) Creating artificial binding pocket boundaries to improve the efficiency of flexible ligand docking. *J. Chem. Inf. Comput. Sci.* 44, 1412–1422.
- [7] Pineda, O., Farràs, J., Maccari, L., Manetti, F., Botta, M. and Vilarrasa, J. (2004) Computational comparison of microtubule-stabilising agents laulimalide and peloruside with taxol and colchicine. *Bioorg. Med. Chem. Lett.* 14, 4825–4829.
- [8] Bjelic, S. and Åqvist, J. (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry* 43, 14521–14528.
- [9] Österberg, F. and Åqvist, J. (2005) Exploring blocker binding to a homology model of the open hERG K<sup>+</sup> channel using docking and molecular dynamics methods. *FEBS Lett.* 579, 2939–2944.
- [10] Li, L., Geng, X., Yonkunas, M., Su, A., Densmore, E., Tang, P. and Drain, P. (2005) Ligand-dependent linkage of the ATP site to inhibition gate closure in the K<sub>ATP</sub> channel. *J. Gen. Physiol.* 126, 285–299.
- [11] Alonso, H., Gillies, M.B., Cummins, P.L., Bliznyuk, A.A. and Gready, J.E. (2005) Multiple ligand-binding modes in bacterial R67 dihydrofolate reductase. *J. Comput.-Aided Mol. Des.* 19, 165–187.
- [12] Yonkunas, M.J., Xu, Y. and Tang, P. (2005) Anesthetic interaction with ketosteroid isomerase: insights from molecular dynamics simulations. *Biophys. J.* 89, 2350–2356.
- [13] Hetényi, C., Maran, U. and Karelson, M. (2003) A comprehensive docking study on the selectivity of binding of aromatic compounds to proteins. *J. Chem. Inf. Comput. Sci.* 43, 1576–1583.
- [14] Lindahl, E. and Delarue, M. (2005) Refinement of docked protein–ligand and protein–DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Res.* 33, 4496–4506.
- [15] Rarey, M., Kramer, B. and Lengauer, T. (1999) The particle concept: placing discrete water molecules during protein–ligand docking predictions. *Proteins* 34, 17–28.
- [16] Österberg, F., Morris, G.M., Sanner, M.F., Olson, A.J. and Goodsell, D.S. (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 46, 34–40.
- [17] DeLano, W.L. (2002) PyMol Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.