Journal of Computer-Aided Molecular Design **17:** 755–763, 2003.

Comparative study of several algorithms for flexible ligand docking

Badry D. Bursulaya, Maxim Totrov, Ruben Abagyan & Charles L. Brooks III*
Department of Molecular Biology (TPC6), The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 1 August 2003; accepted in revised form 12 November 2003

Key words: AutoDock, DOCK, FlexX, Gold, ICM

Abstract

We have performed a comparative assessment of several programs for flexible molecular docking: DOCK 4.0, FlexX 1.8, AutoDock 3.0, GOLD 1.2 and ICM 2.8. This was accomplished using two different studies: docking experiments on a data set of 37 protein–ligand complexes and screening a library containing 10,037 entries against 11 different proteins. The docking accuracy of the methods was judged based on the corresponding rank-one solutions. We have found that the fraction of molecules docked with acceptable accuracy is 0.47, 0.31, 0.35, 0.52 and 0.93 for, respectively, AutoDock, DOCK, FlexX, GOLD and ICM. Thus ICM provided the highest accuracy in ligand docking against these receptors. The results from the other programs are found to be less accurate and of approximately the same quality. A speed comparison demonstrated that FlexX was the fastest and AutoDock was the slowest among the tested docking programs. The database screening was performed using DOCK, FlexX and ICM. ICM was able to identify the original ligands within the top 1% of the total library in 17 cases. The corresponding number for DOCK and FlexX was 7 and 8, respectively. We have estimated that in virtual database screening, 50% of the potentially active compounds will be found among $\approx$ 1.5% of the top scoring solutions found with ICM and among $\approx$ 9% of the top scoring solutions produced by DOCK and FlexX.

Introduction

Molecular docking has become a useful tool in drug discovery efforts. The screening of large databases for possible lead compounds is becoming a routine procedure. Early approaches to the docking problem, such as the original DOCK algorithm [1], considered both a rigid ligand and receptor and used shape complementarity to identify the native-like orientation of the ligand. While advantageous in terms of computational cost, rigid docking has very limited applicability since the large majority of small ligands are flexible [2]. To circumvent the constraints of rigid-body docking, several low-energy conformations of the ligand can be pre-generated, then docked, and the best solution(s) can be chosen according to a scoring/energy function [3, 4]. However, this approach quickly becomes inefficient for ligands with multiple flexible bonds. To avoid the combinatorial 'explosion' associated with ligands containing many rotatable bonds, several incremental construction algorithms, such as Hammerhead [5], DOCK4.0 [1, 6] and FlexX [7] have been implemented. At the same time, improvement in computing power and advances in the energy calculation techniques such as the introduction of a grid-based receptor field representation [8, 9] and the use of internal coordinates [10–15], make the simulations of continuously flexible ligands computationally feasible.

A number of docking programs (AutoDock [16], MCDOCK [17], ICM-dock [18, 19], QXP [20], GOLD [21, 22] and CHARMM [15, 23]) utilize this observation and implement Metropolis Monte-Carlo or genetic algorithms to search for the global minimum of the energy function in the continuous conformational space of the ligand. It is clear that the systematic investigation of existing docking approaches would be helpful in selecting those algorithms and

*To whom correspondence should be addressed. E-mail: brooks@scripps.edu; Fax: +1-858/784-8688.

energy/scoring functions that are optimal. The basic principles behind such an assessment have been described by Vieth et al. [15, 24]. These studies separate the role of the scoring/energy function from that of the docking algorithm and describe a general framework for the necessary conditions that must be satisfied for success in docking. Additionally, several studies have been reported recently comparing the relative performance of the various docking and screening techniques [25–29]. However, so far these studies have involved a very limited number of receptors (between 1 and 4) and, in some cases only a small number of ligands. Given the large diversity in such properties of the ligands as flexibility, size, hydrophobicity and polarity, tests on a significantly larger data set are necessary to evaluate objectively the advantages of various docking approaches [15, 24].

We can group the docking algorithms based on the approach to the conformational search of the flexible ligand into the following two major classes: algorithms that try to fit the ligand into the binding pocket of the protein by matching (geometrically, chemically, energetically etc.); and algorithms that find an optimal ligand conformation by solving a global energy optimization problem. In this study we attempt to evaluate the performance of several programs employing algorithms of both classes. The first group of algorithms is represented by the programs DOCK and FlexX. The second group is represented by the programs AutoDock, GOLD and ICM. All of these programs are publicly available and were obtained by us under either academic (DOCK 4.0, AutoDock 3.0, and ICM 2.8) or demo (FlexX 1.8, GOLD 1.2) licenses. In order to compare the performance of these programs in terms of accuracy and database ranking efficiency, we undertook two kinds of studies, respectively: cross-docking experiments involving protein-ligand complexes with different ligands but the same protein, and database screening against the same receptor. We have chosen to perform the cross-docking experiments in order to avoid an artificial bias inherent in docking of ligands into receptors derived from the crystal structure of their own complexes. The dataset for the docking experiments included 11 groups of complexes, each group containing between two and eight members. The library for the screening experiments contained ligands from the docking dataset and an additional 10,000 molecules randomly selected from a database of commercially available compounds (ACD reagent library) distributed by MDL.

The outline of the paper is as follows. First we briefly review the algorithms employed by the programs. Next we describe input data preparation and docking protocols. Finally we present results from the docking and screening experiments.

Algorithms

Here we briefly outline the docking algorithms employed by AutoDock, DOCK, FlexX, GOLD and ICM. The reader is referred to the original papers for a detailed account [6, 7, 14, 16, 22].

AutoDock and GOLD

AutoDock explores the conformational space of the ligand using the Lamarkian genetic algorithm (LGA), which is a hybrid of a genetic algorithm (GA) with an adaptive local search (LS) method [16]. In this approach, the ligand's state is represented as a chromosome, which is composed of a string of real-valued genes describing the ligand location (three coordinates), orientation (four quaternions) and conformation (one value for each torsion). The simulation is started by creating a random population of individuals. It is followed by a specified number of generation cycles, each consisting of the following steps: mapping and fitness evaluation, selection, crossover, mutation and elitist selection. Each generation cycle is followed by a local search. The solutions are scored using an energy-based scoring function, which includes terms accounting for short-ranged van der Waals and electrostatic interactions, loss of entropy upon ligand binding, hydrogen bonding and solvation.

Similarly, GOLD employs a genetic algorithm. The ligand's state is encoded by a chromosome, representing its conformation and hydrogen bonding. The conformation of the ligand is represented by a binary string, in which every byte encodes for one torsional angle. Each torsion is allowed to vary between $-180°$ and $180°$ in step-sizes of 1.4 Å. Two integer strings encode mappings suggesting possible hydrogen bonds between the protein and the ligand. The first of these strings encodes a mapping of acceptors in the ligand to the donor hydrogens in the protein. The second string encodes a mapping of donor hydrogens in the ligand to the acceptor atoms in the protein. On decoding a chromosome, GOLD utilizes least-squares fitting to form as many of these hydrogen bonds as possible. In the evolutionary development of the ligand conformations the program employs an island model, in

which several subpopulations of chromosomes are created at the beginning instead of one large population. The genetic operations include migration of individual chromosomes between the subpopulations, crossover and mutation. In order to preserve diversity within the population GOLD employs a niching technique, namely, when adding a new individual to the population, the number of individuals in the population that inhabit the same niche as the new chromosome is determined. If there are more than a specified number of individuals in the niche, then the new individual replaces the worst member of the niche rather than the worst member of the total population. Two individuals share the same niche if the RMSD between their donor and acceptor coordinates is less than 1.0 Å. The fitness of a new individual is assessed using a scoring function, which includes energy terms accounting for hydrogen bonding, short-ranged van der Waals interaction between the ligand and protein, and the ligand internal energy. The latter is a sum of ligand steric and torsional energies.

DOCK and FlexX

Both DOCK [6] and FlexX [7, 30] employ an incremental reconstruction algorithm. In this algorithm rigid anchor (DOCK) or base (FlexX) fragments are identified first. At the next step, the selected fragment is placed into the active site of the receptor using a sphere-matching procedure (DOCK) or a hashing technique (FlexX). The complete ligand is constructed by adding the remaining components step by step. At each step of reconstruction a specified number of optimal partial solutions are selected for the next extension step. In DOCK the solutions are scored using energy, contact or chemical scoring functions. The energy scoring function, which was used in this study, includes van der Waals and electrostatic components. In FlexX the scoring is done using a modified Böhm scoring function, which includes the following terms: entropic, which accounts for loss of entropy upon ligand binding; hydrogen bonding; ionic, accounting for electrostatic interactions; aromatic, which accounts for interactions between aromatic groups; and lipophilic, which accounts for hydrophobic interactions. All terms, except the entropic term, are scaled by a corresponding heuristic distance and an angle dependent penalizing function.

ICM

ICM performs flexible ligand docking via global optimization of the energy function [12–14]. The energy terms include the internal energy of the ligand based on the ECEPP/3 force field, as well as van der Waals, hydrogen-bonding, electrostatic and hydrophobic ligand/receptor interaction terms pre-calculated on the grid for computational efficiency [19]. A Monte-Carlo Minimization (MCM) procedure in the internal coordinate space [14] is employed to find the global minimum of the energy function. Each step of the algorithm consists of a random conformational change of two types, torsional or positional, followed by the local minimization. Torsional move involves complete randomization of a single arbitrarily chosen torsion angle. A positional move involves a pseudo-Brownian random translation and rotation of the ligand as a whole [14]. ICM uses an analytical gradient minimizer, which finds the local minima of the energy function more rapidly than the simplex minimizer or the stochastic search alone. To improve convergence, multiple MC runs from several starting positions are performed. The VLS (virtual library screening) scoring function used in ICM consists of the internal force-field energy of the ligand and the ligand/receptor interaction energy. The latter includes van der Waals terms, a hydrophobicity term based on the solvent accessible surface buried upon binding, an electrostatic solvation term calculated using a boundary-element solution of the Poisson equation, hydrogen-bond interaction terms and an entropic term proportional to the number of flexible torsions in the ligand [19].

Methods

Input data preparation and algorithms comparison methodology

(1) The proteins which we have chosen for docking and database screening experiments satisfy the following criteria: they have at least two entries with different ligands in the protein databank (PDB); they do not form covalent bonds with their respective ligands; the majority of their ligands have a relatively large number of rotatable bonds (see Table 1).

(2) The ligand input files were prepared according to the following procedure. First, we extracted ligand coordinates from the corresponding PDB file and assigned chemical bonds, partial charges and added hydrogen atoms employing the bond increment rules

Table 1. Results of the docking experiments.

Complex	N^a_{rot}	AutoDock[b]	DOCK[b]	FlexX[b]	ICM[b]	GOLD[b]
			Trypsin			
3ptb	3	0.80	0.59	1.11	0.49	1.09
1tng	2	0.62	0.86	1.08	0.71	1.89
1tnj	3	1.21	1.56	1.73	2.17	1.90
1tnk	4	1.69	1.87	1.70	2.53	3.08
1tni	5	2.61	5.26	2.73	3.40	4.93
1tnl	1	0.41	2.08	3.74	1.93	1.61
1tpp	7	1.80	3.25	1.95	1.71	2.33
1pph	11	5.14	3.91	3.27	1.44	4.23
			Cytochrome P-450$_{cam}$			
1phf	1	2.09	2.39	4.68	1.23	4.42
1phg	5	3.52	5.57	4.87	0.46	4.20
2cpp	3	3.40	2.48	0.44	2.53	3.49
			Neuraminidase			
1nsc	12	1.40	4.86	6.00	1.80	1.02
1nsd	11	1.20	4.51	1.56	1.04	0.96
1nnb	11	0.92	4.51	0.92	1.09	0.84
			Carbocypeptidase A			
1cbx	5	1.33	3.13	1.32	0.82	1.87
3cpa	8	2.26	6.48	1.51	0.77	1.87
6cpa	16	8.30	8.30	9.83	1.60	4.96
			L-Arabinose binding protein			
1abe	4	0.16	1.87	0.55	0.36	0.18
1abf	5	0.48	3.25	0.76	0.61	0.50
5abp	6	0.48	3.89	4.68	0.88	0.59
			ε-Thrombin			
1etr	15	4.61	6.66	7.26	0.87	5.99
1ets	13	5.06	3.93	2.11	6.22	2.39
1ett	11	8.12	1.33	6.24	0.99	1.30
			Thermolysin			
3tmn	10	4.51	7.09	5.30	1.36	3.96
5tln	14	5.34	1.39	6.33	1.42	1.60
6tmn	20	8.72	7.78	4.51	2.60	8.54
			Penicillopepsin			
1apt	30	1.89	8.06	5.95	0.88	8.82
1apu	29	9.10	7.58	8.43	2.02	10.7
			Intestinal FABP			
1icm	13	1.80	3.99	2.94	1.11	2.30
1icn	17	3.99	3.88	2.95	1.35	2.05
2ifb	15	3.09	1.43	8.94	1.04	2.61
			Ribonuclease T_1			
1gsp	4	2.67	1.16	3.71	0.54	0.70
1rhl	7	0.96	0.71	1.15	3.53	1.08
1rls	7	0.98	1.75	4.33	0.79	1.16
			Carbonic anhydrase II			
1cil	6	5.81	2.78	3.52	2.00	6.04
1okl	5	8.54	5.65	4.22	3.03	3.55
1cnx	13	10.9	7.35	6.83	2.09	6.32

[a] Number of rotatable bonds in the ligand.
[b] RMSD values in Å.

from MMFF94 [31] as implemented in ICM [12–14]. All carboxylic acid and phosphoric acid groups were ionized and all amino-groups were protonated. Next, all torsion bonds were randomized and local minimization was performed for the ligands in the gas phase using the MMFF94 force field in ICM. After that each set of ligand coordinates was modified in such a way that the ligand center of geometry was superimposed with that of the reference ligand. Finally, the ligand coordinates were written into MOL2 and PDB format files.

(3) The receptor input files were prepared according to the following procedure. First, we removed from the corresponding PDB file all water molecules, ligand atoms and those ions that did not belong to the active site of the receptor. Next hydrogen atoms were added and partial charges were assigned. This was followed by a local minimization to relieve potential bad contacts. The minimization was performed in the presence of restraints to maintain the protein conformation very near that observed in the experimental model. Finally, the receptor coordinates were written into MOL2 and PDB format files.

(4) The docking experiments were performed on the same computer and the CPU time required for docking was recorded. The docking protocols are described in the next section. We emphasize here that all methods use the same receptor coordinates and start with exactly the same initial location, conformation and orientation of the ligands. The length of the docking experiments was controlled by the default or recommended parameter settings. We observed that doubling the length of the docking experiments (AutoDock and ICM) did not improve significantly the accuracy of the solutions.

(5) For each docking method only the best scoring solution per complex was saved. Different algorithms were compared based on the root-mean-square deviation (RMSD) of heavy atoms of the best predicted structures from the corresponding crystal structures. If the ligand has local topological symmetry at single bonds, whose torsion angle can be changed by a rotation of less than 360° without changes in the global conformation of the ligand, the RMSD of alternative orientations was calculated and the smallest one was kept for the purpose of comparison of different algorithms. The coordinates of the ligand structure, used for RMSD calculations, were obtained by superimposing its crystallographic protein coordinates with the receptor coordinates used for the docking.

(6) In order to quantify the ligand docking quality and to compare performance of the different methods, we introduced a docking accuracy (DA) function following the idea of Vieth et al. [24], which makes use of RMSD values and measures how accurately the ligands – members of a particular group – are docked by a given method:

$$DA = f_{rmsd\leq 2} + 0.5(f_{rmsd\leq 3} - f_{rmsd\leq 2}), \quad (1)$$

where $f_{rmsd\leq a}$ indicates the fraction of ligands docked into a given receptor with RMSD less or equal to a Å. The docking accuracy of the method for a particular receptor is zero if $f_{rmsd\leq 3}$ is zero.

Docking protocols

In all algorithms studied here, the receptor is treated as a rigid body and a grid potential is used to evaluate the scoring functions. This simplification allows one to perform docking more efficiently, which is especially crucial in database screening.

AutoDock. AutoDock requires the receptor and ligand coordinates in MOL2 format. Nonpolar hydrogen atoms were removed from the receptor file and their partial charges were added to the corresponding carbon atoms. The program *Mol2topdbqs* was used to transform the receptor MOL2 file into the PDBQS format file containing the receptor atom coordinates, partial charges and solvation parameters. The program *AutoTors* was used to transform the ligand MOL2 file into a PDBQ file, merge nonpolar hydrogen atoms and define torsions. The grid calculations were set up with the utility *Mkgpf3* and maps were calculated with the program *AutoGrid*. The grid maps were centered on the ligand's binding site and were of dimension $61 \times 61 \times 61$ points. The grid spacing was 0.375 Å yielding a receptor model that included atoms within 22.9 Å of the reference binding site center. The default parameter settings generated by the program *Mkdpf3* were used for docking. For each complex 10 dockings were performed. The initial population was set to 50 individuals; maximum number of energy evaluations was 2.5×10^5; maximum number of generations was 27,000. The other parameters provided by the default setting were the same as in Ref. [16].

GOLD. GOLD requires the receptor and ligand coordinates in any of the following formats: PDB, MOL, SDF or MOL2 format. The active site origin was specified by the center of geometry of the reference ligand. The active site radius was 10 Å. The following default parameters were used: number of islands was 5, population size was 100, number of genetic operations was 100,000 and niche size was 2.

DOCK. DOCK requires the following receptor files: a MOL2 format file containing coordinates of all atoms; a PDB file containing heavy atoms coordinates only; a PDB file containing heavy atoms excluding those of the active site. The active site atoms included those receptor atoms that were within 6.5 Å from the reference ligand atoms. The ligand coordinates were provided in MOL2 format. The site points for the ligand docking were identified using the *SPHGEN* program. The number of docking points did not exceed 50. The energy score was employed for the orientational and conformational search. Grid maps were calculated using the program *Grid*, with grid spacing of 0.5 Å. An energy cutoff distance of 10 Å was employed. Electrostatic interactions were calculated with distance dependent dielectric constant. The dielectric factor was set to 4. Proteins were represented by a united atom model. Flexible bonds and anchors were automatically identified by DOCK. The conformational search was done using the torsion driver. The clash overlap was set to 0.5. The top 25 conformations were retained during each cycle of the search. Multiple anchors were allowed, with the minimum number of heavy atoms in the anchor set to 10. An orientational search was performed with automated matching. The maximum number of orientations was 500 for docking experiments and 100 for the database search. A local energy minimization of orientations and conformations of the ligand and anchor was performed. The ligand reminimization was turned on. The default minimization parameters were employed.

FlexX. FlexX requires a MOL2 format file for the ligand and a PDB format file for the receptor. The default settings as provided with the FlexX 1.8 package were used for flexible docking and database screening. The conformational flexibility of the ligand is modeled by a discrete set of preferred torsional angles for acyclic single bonds. The rings were considered rigid, since the program *CORINA* for treating multiple conformations of the rings was not included in the distribution. The active site and the interaction surface of the receptor were defined by using a reference ligand and a 6.5 Å cutoff distance. Base fragments were selected automatically. The maximum number of base fragments was 4. The base fragment was

placed into the active site using two algorithms. The first one superimposes triples of interaction centers of a base fragment with triples of compatible interactions in the active site. The second algorithm, called matching, is used when the base fragment had fewer than three interaction centers. The sampling was done with 400 solutions per partial solution at each iteration of incremental construction.

ICM. Grid maps were calculated with a grid spacing of 0.5 Å. Docking was performed with a default script provided by ICM, version 2.8.169. During the docking, either one of the torsional angles of the ligand was randomly changed or a pseudo-Brownian move was performed. Each random change was followed by 100 steps of local conjugate-gradient minimization. The new conformation was accepted or rejected according to a Metropolis rule using a temperature of 600 K. The length (number of Monte Carlo steps) of the docking run as well as the length of local minimization was determined automatically by an adaptive algorithm, depending on the size and number of flexible torsions in the ligand.

Results

All docking experiments and database screenings were performed on an SGI R10000 equipped with a single 195 MHz IP28 processor and 128 MB main memory.

Docking experiments

The following receptors were used in docking experiments: trypsin (PDB entry 3ptb), cytochrome P-450$_{cam}$ (PDB entry 1phf), neuraminidase (PDB entry 1nsc), carboxypeptidase A (PDB entry 1cbx), L-arabinose binding protein (PDB entry 1abe), ϵ-thrombin (PDB entry 1etr), thermolysin (PDB entry 3tmn), penicillopepsin (PDB entry 1apt), intestinal fatty-acid binding protein (PDB entry 1icm), ribonuclease T_1 (PDB entry 1gsp), and carbonic anhydrase II (PDB entry 1cil). Most receptors have three ligand members with the exception of trypsin, which has 8 members, and penicillopepsin, which has 2 members.

The complexes for which cross-docking experiments were performed and the RMSD values are given in Table 1. From the results in this table, we see that $\approx$ 46%, 30%, 35%, 46% and 76% of ligand molecules are docked correctly within 2 Å RMSD of their known binding modes by, respectively, AutoDock, DOCK, FlexX, GOLD and ICM. The docking accuracy of each program, which measures a fraction of molecules docked with acceptable accuracy, is summarized in Table 2. From this table it follows that an average docking accuracy of AutoDock, DOCK, FlexX, GOLD and ICM is, respectively, 0.47, 0.31, 0.35, 0.52 and 0.93. Thus ICM is the most accurate in predicting the correct protein–ligand conformation. We note that it gets a score of 1 for five receptors, which means that all members of those receptors are docked within an RMSD of less than 2 Å from the corresponding crystal structures. Other methods are less accurate in their predictions. In particular, we observe that there are several receptors for which some programs fail to produce any acceptable solution. Those are ε-thrombin, thermolysin and carbonic anhydrase II in case of AutoDock; neuraminidase, carboxypeptidase and penicillopepsin in case of DOCK; thermolysin, penicillopepsin and carbonic anhydrase II in case of FlexX; and cytochrome P450, penicillopepsin and carbonic anhydrase II in case of GOLD. On average the docking accuracy of GOLD, AutoDock and FlexX is approximately the same and that of DOCK is slightly less.

Table 2. Docking accuracies of algorithms.

Receptor	AutoDock	DOCK	FlexX	ICM	GOLD
Trypsin	0.81	0.56	0.69	0.75	0.56
Cytochrome P450$_{cam}$	0.17	0.33	0.33	0.83	0.00
Neuraminidase	1.00	0.00	0.67	1.00	1.00
Carboxypeptidase	0.50	0.00	0.67	1.00	0.67
L-Arabinose	1.00	0.33	0.67	1.00	1.00
ϵ-Thrombin	0.00	0.33	0.17	0.67	0.50
Thermolysin	0.00	0.33	0.00	0.83	0.33
Penicillopepsin	0.50	0.00	0.00	1.00	0.00
Intestinal FABP	0.33	0.33	0.33	1.00	0.67
Ribonuclease T_1	0.83	1.00	0.33	0.67	1.00
Carbonic anhydrase II	0.00	0.17	0.00	0.67	0.00
Average	0.47	0.31	0.35	0.93	0.52

The average docking CPU times of the methods are given in Table 3. The average docking time increases in the following order: FlexX, DOCK, ICM, Gold and AutoDock. The low docking speed of GOLD and AutoDock suggests that at the present time it is not suitable for database screening on a single processor computer. We note, however, that compared to the overall cost in the drug development process, the com-

Table 3. Average docking times (s) of algorithms.

Receptor	AutoDock	DOCK	FlexX	ICM	GOLD
Trypsin	391	51	26	65	165
Cytochrome P450$_{cam}$	291	29	82	40	273
Neuraminidase	620	98	72	99	269
Carboxypeptidase	624	88	92	147	437
L-Arabinose	353	37	31	39	288
ε–Thrombin	1174	421	83	336	676
Thermolysin	789	170	65	238	500
Penicillopepsin	1122	412	77	645	840
Intestinal FABP	560	138	29	234	489
Ribonuclease T_1	668	69	56	71	352
Carbonic anhydrase II	519	55	88	92	388
Average	646	143	64	182	356

putational cost is of minor importance. Moreover, such costs are anticipated to continue to decrease owing to rapid developments in the computer industry. Thus, when comparing different docking algorithms more emphasis should be attached to their accuracy rather than the computational cost. With this in mind we conclude, based on the results of docking experiments, that the Internal Coordinates Method (ICM) has the best docking performance among the algorithms we examined.

Before proceeding to the library screening results, we note that in all docking experiments we used receptor and ligand MOL2 files generated by ICM. However, according to the AutoDock, DOCK and FlexX manuals, it is recommended to use SYBYL to generate the ligand and receptor (AutoDock and DOCK) input files. The only difference between input files generated by SYBYL and ICM is in partial atomic charges. ICM uses the bond charge increment method from MMFF94 [31] to assign partial atomic charges. In order to be certain that our results are not influenced by the partial charges provided by MMFF94 bond-charge increments, we repeated the docking experiments for AutoDock, DOCK and FlexX using charges generated by SYBYL: Kollman charges for receptor atoms and Gasteiger charges for ligand atoms. Our findings indicate that the accuracy of AutoDock, DOCK and FlexX with SYBYL charges was the same as with MMFF94 charges.

Database screening

The same proteins that were used in the docking experiments have been chosen to perform screening of a ligand library. As was mentioned in the Introduction, this library contains ligands that were used for the docking experiments and an additional 10,000 molecules selected randomly from the ACD reagent library. While alternative libraries could have been employed, the characteristics of the ligands in this were similar to those of the ligands from the protein receptors: the mean number of rotatable bonds in the ACD set was 8, which is close to the value of 9 found for the ligands in Table 1. The purpose of the screening experiments is to find out how well the programs distinguish the original ligands of the complexes among a large database of molecules [2, 15, 24]. This has tremendous practical implications, since a good docking algorithm may allow one to cut significantly the cost of the drug discovery process by reducing the fraction of compounds in a ligand library that needs to be analyzed experimentally. The results of screening performed by DOCK, FlexX and ICM are summarized in Table 4. First we note that out of 37 original ligands ICM places 17 ligands within the top 1% of scanned solutions, while the corresponding number for DOCK and FlexX is 7 and 8, respectively. The easiest receptor for screening experiments is L-Arabinose binding protein, since all algorithms assign very high scores to its original ligands.

In order to better quantify the results from the database screening test, we use the cumulative probability, $P(\rho)$, for the original ligands to be found in the top ρ percent of scanned solutions:

$$P(\rho) = \frac{N_{\leq\rho}}{N_{lig}}, \tag{2}$$

where $N_{\leq\rho}$ is a number of original ligands that can be found within the top ρ percent of scanned solutions and N_{lig} is a total number of original ligands in the library. In our case N_{lig} equals 37. The cumulative probability, $P(\rho)$, for DOCK, FlexX and ICM is given in Table 5. The probability of finding an original ligand within the smallest percentage of the rank ordered binding affinities is highest in the case of ICM. To be specific in comparing library screening results, we use a percent of the scanned database that contains approximately half of the original ligands. This criterion can be employed in virtual screening in order to select drug leads. The actual activity of the selected molecules can be later verified in screening experiments. It is evident from the data compiled in Table 5 that $\approx$ 50% of ori-

Table 4. Results of the virtual database screening.

Complex	DOCK[a]	FlexX[a]	ICM[a]
Trypsin			
3ptb	20.3	14.0	0.1
1tng	28.7	30.8	0.3
1tnj	45.1	47.9	1.1
1tnk	58.1	37.1	14.6
1tni	43.6	38.5	0.6
1tnl	82.9	39.1	51.6
1tpp	4.00	0.3	0.4
1pph	37.0	0.6	0.3
Cytochrome P-450$_{cam}$			
1phf	26.9	13.2	3.7
1phg	60.6	1.4	12.7
2cpp	7.3	12.6	11.7
Neuraminidase			
1nsc	19.0	7.2	5.01
1nsd	9.4	3.0	0.44
1nnb	6.7	1.4	0.43
Carboxypeptidase A			
1cbx	36.8	2.7	0.2
3cpa	14.2	1.4	0.3
6cpa	8.5	2.7	3.4
L-Arabinose binding protein			
1abe	0.5	0.2	0.01
1abf	1.2	0.2	0.03
5abp	0.2	0.4	0.02
ε-Thrombin			
1etr	5.6	0.5	1.6
1ets	0.4	0.3	6.7
1ett	0.9	9.0	0.06
Thermolysin			
3tmn	34.3	3.5	2.8
5tln	66.0	0.6	28.5
6tmn	15.0	8.5	77.6
Penicillopepsin			
1apt	0.3	41	1.1
1apu	3.8	80	12.8
Intestinal FABP			
1icm	10.1	68.8	2.6
1icn	3.4	74.4	0.5
2ifb	7.1	99.3	0.5
Ribonuclease T_1			
1gsp	4.4	10.7	0.6
1rhl	5.6	10.2	0.4
1rls	68.9	7.4	12.3
Carbonic anhydrase II			
1cil	35.0	3.7	71.7
1okl	18.5	19.3	30.9
1cnx	0.3	14.5	90.6

[a] Percent of molecules in library which score equally or higher than a given ligand.

Table 5. Cumulative probability.

ρ[a]	DOCK	FlexX	ICM
0.1	0.000	0.000	0.135
0.2	0.027	0.054	0.162
0.4	0.135	0.135	0.351
0.6	0.162	0.216	0.459
1.0	0.189	0.216	0.459
2.0	0.216	0.297	0.540
5.0	0.324	0.432	0.676
8.0	0.459	0.486	0.676
10	0.540	0.568	0.703
20	0.676	0.730	0.838
50	0.867	0.892	0.892
100	1.000	1.000	1.000

[a] Percent of total database.

ginal ligand molecules are found among $\approx$ 1.5% of the top ICM scoring solutions and among $\approx$ 9% of top scoring solutions produced by DOCK and FlexX. These results suggest that virtual screening can reduce the cost of drug discovery by limiting the number of compounds suitable for screening experiments. In addition, a list of potential drug candidates produced by ICM is $\approx$ 6 times smaller than that of DOCK and FlexX.

As a sidenote, we found that the database screening results with DOCK are improved somewhat if the chemical scoring function is employed instead of energy score. This contrasts the results of docking experiments, where no influence of the choice of scoring function was found.

Conclusions

We compared six different docking programs in their ability to flexibly dock 37 molecules into 11 different receptors. We found that $\approx$ 46%, 30%, 35%, 46% and 76% of the molecules are docked correctly within 2 Å RMSD of their known binding modes by, respectively, AutoDock, DOCK, FlexX, GOLD and ICM.

The screening of a database containing 10,037 molecules against 11 different proteins was also performed. The screening results indicate that DOCK and FlexX find 7 (8) original molecules within the top 1% of docking solutions, while ICM finds 17 such molecules. From these findings we estimated that in order to find 50% of the potentially active compounds, $\approx$ 1.5% of the top ICM scoring solutions and $\approx$ 9% of the top scoring solutions produced by

DOCK and FlexX need to be selected for experimental verification.

Our results suggest that all docking programs studied here do a reasonable job in docking and database screening experiments and should aid significantly the drug discovery process. However, ICM consistently outperformed other programs and its use for virtual library screening seems to be most advantageous.

Acknowledgements

The authors would like to thank Dr A. Olson for the AutoDock 3.0 academic license; Dr I.D. Kuntz for the DOCK 4.0 academic license; the German National Research Center for Information Technology for the FlexX 1.8 evaluation license; Molsoft for the ICM 2.8 academic license; and the Cambridge Crystallographic Data Centre for the GOLD 1.2 evaluation license. We are also grateful to the National Institutes of Health (GM37554) for financial support.

References

1. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
2. Roche, O., Kiyama, R. and Brooks, C.L., III, J. Med. Chem., 44 (2001) 3592.
3. Kearsley, S.K., Underwood, D.J., Sheridan, R.P., and Miller, M.D., J. Comput.-Aided Mol. Des., 8 (1994) 565.
4. Leach, A.R., J. Mol. Biol., 235 (1994) 345.
5. Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
6. Ewing, T.J.A. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1176.
7. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
8. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
9. Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., J. Comput.-Aided Mol. Des., 8 (1994) 153.
10. Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A., J. Phys. Chem., 79 (1975) 2361.
11. Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
12. Mazur, A.K. and Abagyan, R.A., J. Biomol. Struct. Dyn., 6 (1989) 815.
13. Abagyan, R. and Argos, P., J. Mol. Biol., 225 (1992) 519.
14. Abagyan, R., Totrov, M. and Kuznetsov, D., J. Comput. Chem., 15 (1994) 488.
15. Vieth, M., Hirst, J.D., Dominy, B.N., Daigler, H. and Brooks, C.L., III, J. Comput. Chem., 19 (1998) 1623.
16. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comput. Chem., 19 (1998) 1639.
17. Liu, M. and Wang, S., J. Comput.-Aided Mol. Des., 13 (1999) 435.
18. Totrov, M. and Abagyan, R., Proteins, Suppl 1 (1997) 215.
19. Totrov, M. and Abagyan, R. In: Protein–ligand Docking as an Energy Optimization Problem. The Thermodynamics of Receptor–Inhibitor Interactions, ed. R.B. Raffa. Wiley and Sons, New York, 2000.
20. McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Des., 11 (1997) 333.
21. Jones, G., Willett, P. and Glen, R.C., J. Mol. Biol., 245 (1995) 43.
22. Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
23. Wu, G., Robertson, D.H., Brooks, C.L., III, and Vieth, M., J. Comput. Chem., 24 (2003) 1549.
24. Vieth, M., Hirst, J.D., Kolinski, A. and Brooks, C.L., III, J. Comput. Chem., 19 (1998) 1612.
25. Ha, S., Andreani, R., Robbins, A. and Muegge, I., J. Comput.-Aided Mol. Des., 14 (2000) 435.
26. Charifson, P.S., Corkery, J.J., Murcko, M.A. and Walters, W.P., J. Med. Chem., 42 (1999) 5100.
27. Diller, D.J. and Verlinde, L.M.J., J. Comput. Chem., 20 (1999) 1740.
28. Knegtel, R.M., Bayada, D.M., Engh, R.A., von der Saal, W., van Geerestein, V.J. and Grootenhuis, P.D., J. Comput.-Aided Mol. Des., 13 (1999) 167.
29. Bissantz, C., Folkers, G. and Rognan, D., J. Med. Chem., 43 (2000) 4759.
30. Kramer, B., Rarey, M. and Lengauer, T., Proteins, 37 (1999) 228.
31. Halgren, T.A., J. Comput. Chem., 17 (1995) 490–641.