Journal of Computer-Aided Molecular Design, **15:** 157–171, 2001.
KLUWER/ESCOM
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Ligand-receptor docking with the Mining Minima optimizer

Laurent David**, Ray Luo & Michael K. Gilson*
Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850, USA

Received 8 February 2000; Accepted 3 October 2000

Key words: flexible ligand docking, Mining Minima, optimization, receptor

Summary

The optimizer developed for the Mining Minima algorithm, which uses ideas from Genetic Algorithms, the Global Underestimator Method, and Poling, has been adapted for use in ligand-receptor docking. The present study describes the resulting methodology and evaluates its accuracy and speed for 27 test systems. The performance of the new docking algorithm appears to be competitive with that of previously published methods. The energy model, an empirical force field with a distance-dependent dielectric treatment of solvation, is adequate for a number of test cases, although incorrect low-energy conformations begin to compete with the correct conformation for larger sampling volumes and for highly solvent-exposed binding sites that impose little steric constraint on the ligand.

Introduction

The Mining Minima method [1] computes molecular free energies by rapidly identifying the most stable conformations of a molecule or a complex and integrating the Boltzmann factor in the corresponding energy wells. The method requires an efficient method for locating the low-energy conformations of a molecular system, and we have developed a novel optimization algorithm to accomplish this task. This optimization algorithm draws on ideas from the Global Underestimator method [2] and Genetic Algorithms [3–5], and also has connections with the Poling method [6] and Tabu Search [7–9]. Because of its efficiency, the optimizer developed for the Mining Minima algorithm has the potential to be useful in other applications as well. The present report describes the use of this optimizer in protein-ligand docking. The intent is to identify the most stable bound conformation of a receptor-ligand complex as an aid to molecular design and as a first step in ranking the stabilities of protein-ligand complexes.

*To whom correspondence should be addressed; E-mail: gilson@umbi.umd.edu
**Current address: Département d'Informatique et de Recherche Opérationnelle Université de Montréal, Québec, H3C 3J7, Canada.

The Mining Minima optimizer has been adapted as follows for this purpose. First, because no free energy calculation is done, the time-consuming integration of the Boltzmann factor within each energy well is removed. Second, an exclusion zone of uniform dimensions is placed around each energy minimum as it is discovered, in order to avoid rediscovering it in future docking iterations. Third, ligand-receptor interactions are now evaluated via precomputed potential grids, and an upper limit is placed on the van der Waals repulsions between ligand and receptor in order to speed the search for low energy conformations. Finally, heuristic parameters within the optimization algorithm are adjusted to optimize performance. The present paper compares the performance of the resulting algorithm with published results for the docking programs PRO_LEADS [8], AutoDock [10], FlexX [11], MCDOCK [12], the method of Wang et al. [13] and GOLD [14].

The paper is organized as follows. The Methods section describes the optimization algorithm and its adaptation for use in ligand-docking, along with the test systems and the energy model. The Results section first describes preliminary tests and adjustments based upon ligand-protein systems used previously to assess the PRO_LEADS algorithms [8]. Then the adjusted method is analysed and compared with other docking

methods by reusing prior test cases, including systems for which the GOLD algorithm failed to reproduce the experimentally determined bound conformations. Finally, the method is tested for sampling regions of increasing size.

Methods

In the present study, the energy is optimized with respect to the position and orientation of the ligand, along with a user-selected set of rotatable bonds (dihedral angles) in the ligand. The receptor is currently treated as rigid, but mobilizing selected rotatable bonds in the receptor will be conceptually straightforward. The following subsections describe the optimization algorithm, the energy model, and the tests used to evaluate the method.

Docking Algorithm

A typical docking run here consists of 5–100 individual optimizations, each of which yields one docked configuration. Gradient minimization techniques are not used. Thus, neither forces nor second derivatives of the energy are calculated. Instead, the optimization is guided by the trends in and history of a series of energy evaluations. The procedure is as follows.

Gradually focusing a random search
The central method of the optimizer involves generating a large number of candidate conformations at random within a search region, and keeping track of which conformation has the lowest energy at any given iteration. The search region, which is defined in terms of a search range for each dihedral angle and for overall translation and rotation of the ligand, is always centered around the current lowest energy conformation found to date. When a new conformation is found to be more stable than the current lowest-energy conformation, the search region is re-centered on the new lowest-energy conformation. The search region is gradually narrowed during the search until it has essentially converged on a single conformation. Note that each successive conformation is generated randomly and without reference to prior conformations, so the search is not blocked by energy barriers. This is in contrast to methods like Metropolis Monte Carlo that generate the new conformation by modifying the current conformation.

The total number of conformations generated during one optimization series is $N_{\text{hunt}} \approx 10\,000$. The factor f by which the search range is narrowed relative to the initial range is computed as $f = e^{AN_{\text{done}}/N_{\text{hunt}}}$, where N_{done} is the number of samples already taken and A is an adjustable parameter empirically set to -3. The starting range for each dihedral angle is $[0, 2\pi]$, and overall rotation of the ligand is initially sampled over all possible orientations, as defined by 3 Euler angles.

This 'hunt' phase of the optimization is repeated $N_{\text{search}} \approx 4$ times, and the lowest energy conformation found over all N_{search} attempts is advanced to a 'fine-tuning' phase. In the fine-tuning phase, the selected conformation is refined by applying the same narrowing-in procedure, except that the initial search range is restricted to the vicinity of the conformation of interest. More specifically, the initial search range is set to the excluded zone (see next two paragraphs) surrounding the energy minimum. Only 1 final conformation will ever be reported from a single excluded zone. The number of samples taken in the fine tuning phase, N_{fine}, is typically 10 000. Thus, the total number of energy evaluations for each docked configuration equals $N_{\text{search}} \times N_{\text{hunt}} + N_{\text{fine}}$.

This narrowing-in technique is based loosely on the global underestimator method, which iterates the following steps. A collection of local energy minima is generated; the coordinates and energies of these minima are used to construct a concave-up parabola of energy versus conformation that lies at or below each of the local minima; and a new set of local minima is then generated in the vicinity of the global minimum of the parabolic function, with the idea that the global energy minimum of the actual energy function probably lies near the minimum of the parabolic 'underestimator' function. We conjectured that the minimum of the parabolic function tends to lie near the lowest energy configuration found to date and that one could therefore obtain similar results by simply narrowing the search around the lowest-energy configuration found to date, without constructing the parabolic underestimator.

Promoting the discovery of new energy minima
It is wasteful for a search algorithm to spend time rediscovering previously discovered energy minima. The present algorithm avoids this problem placing an explicit excluded zone around each docked conformation that has been generated to date. During the 'hunt' phase, a conformation that lies in any of the excluded zones is immediately discarded without incrementing N_{done} (see above). An excluded zone is centered on an

energy minimum and is defined by a range for each ligand dihedral angle, along with 6 ranges for the overall translations and rotations of the ligand. In order for a test conformation to be discarded, every dihedral angle and all 6 translational/rotational coordinates must lie within their respective excluded ranges. (For example, a single dihedral angle in an excluded range does not lead to exclusion of a trial conformation.) The sizes of the exclusion ranges are parameters of the algorithm, as described below. Previously described optimization methods that also guide the search away from previously discovered low-energy conformations include Poling [6] and Tabu Search [7-9].

Recombining partial solutions

Optimization algorithms sometimes generate partially correct solutions on the way to finding the global energy optimum. For example, an alpha-helix-forming peptide may be helical at the N-terminus but frayed at the C-terminus, or vice versa. Recombining such partial solutions can speed the discovery of the global optimum. Thus, Genetic Algorithms [3-5] recombine partial solutions via crossovers to accelerate discovery of the global optimum.

In the present algorithm, a docking run yields a growing list of energy minima. These are stored and used for recombination in the hunt phase of subsequent energy minimizations. For example, 5 configurations are available for recombination during the 6th minimization. During the hunt phases of the 6th minimization, each time a new random configuration is generated, one of the saved minima is selected for possible recombination. There is a probability $p = 1 - e^{ABN_{\text{done}}/N_{\text{hunt}}}$ that a given coordinate (positional, orientational or conformational) in the new, randomly generated configuration will be replaced with the corresponding coordinate from the old energy minimum. The adjustable parameter B is empirically set to 0.1. Note that recombination affects the new configuration but, unlike in a Genetic Algorithm, it does not alter the stored configuration.

Energy model

The energy of a proposed protein-ligand conformation depends upon the position and orientation of the ligand, and upon the values of the ligand dihedral angles selected as rotatable ($\boldsymbol{\phi}$). The position and orientation of the ligand are defined, respectively, via the 3 Cartesian coordinates $\mathbf{r}$ of the ligand atom closest to its center of coordinates at the start of the calculation, and 3 Euler angles that rotate the ligand around this atom.

The internal energy of the ligand as a function of its dihedral angles, $E^L(\boldsymbol{\phi})$, is computed with the CHARMM 22 force-field [15], using the distance-dependent dielectric model $\epsilon_{ij} = 4r_{ij}$. The CHARMM parameters were generated with the program Quanta98 (see below) [16]. In the present study, the receptor conformation is kept fixed so its energy does not vary and may be neglected. The ligand-receptor interaction energy $E^{LR}(\boldsymbol{\phi})$ is computed with three precomputed grids of 'potential fields' generated by the receptor protein [17, 18], as follows.

An electrostatic grid is computed with $\epsilon_{ij} = 4r_{ij}$ and the CHARMM22 charges of the protein. A repulsive Lennard-Jones grid accounts for steric interactions between ligand and receptor: each receptor atom i contributes A_i/r^{12} to the steric potential ϕ^{steric} at each grid point, where r is the distance between the atom and the grid point and $A_i \equiv \sqrt{\epsilon_i}\sigma_i^6$, with ϵ_i and σ_i equal to the usual Lennard-Jones parameters of atom i. An attractive Lennard-Jones grid accounts for dispersion interactions between ligand and receptor: each receptor atom contributes $-B_i/r^6$ to the dispersion potential ϕ^{disp} at each grid point, where $B_i \equiv \sqrt{\epsilon_i}\sigma_i^3$. During docking, the interaction of the ligand with the receptor is evaluated as:

$$E^{LR}(\mathbf{r}, \boldsymbol{\zeta}, \boldsymbol{\phi}) = \sum_{j}^{N_{\text{atoms}}} \min(E_{\max}, [q_j \phi_j^{\text{elec}} + A_j \phi_j^{\text{steric}} + B_j \phi_j^{\text{disp}}]) \quad (1)$$

where j indexes the atoms of the ligand and the potentials ϕ_j at ligand atom j are obtained from the appropriate grids by trilinear interpolation. Note that the ligand-receptor interaction energy is capped at $E_{\max} \in [2, \infty]$ kcal/mol in the present study in order to eliminate 'noise' in the ligand-receptor interaction energy and thus allow the steric energy to indicate reliably the degree of ligand-receptor overlap.

In the present study, the grid has a 0.2 Å spacing and dimensions of $30 \times 30 \times 30$ Å, except as otherwise noted. For the present tests, where the position of the ligand is known experimentally, the grids are centered on the ligand atom closest to the center of coordinates of the ligand in its crystal configuration, in accordance with the PRO_LEADS test protocol. An additional energy term $E^{\text{Off}}(\mathbf{r}, \boldsymbol{\zeta}, \boldsymbol{\phi})$ adds an energy penalty of 15×10^6 kcal/mol for any atom that lies off the grid. However, the grids are normally large enough that this case does not arise.

In summary, the total energy of a trial conformation of the ligand is

$$E(\mathbf{r}, \boldsymbol{\zeta}, \boldsymbol{\phi}) = E^{L}(\boldsymbol{\phi}) + E^{LR}(\boldsymbol{\phi}, \mathbf{r}, \boldsymbol{\zeta}) + E^{\text{Off}}(\mathbf{r}, \boldsymbol{\zeta}, \boldsymbol{\phi}) \quad (2)$$

where each term is defined above.

Details of the calculations

Three sets of test cases were examined: one drawn from a prior study of the docking program PRO_LEADS; another drawn from prior docking studies with AutoDOCK [10], FlexX [11], MCDOCK [12], and the multistep method of Wang et al. [13]; and the third drawn from a prior docking study with GOLD [14]. Table 1 provides more information on the receptor-ligand complexes examined here.

The program Quanta [16] was used to add all hydrogen atoms to the ligands and polar hydrogens only to the proteins, with the exception of Cytochrome P450CAM (**2cpp**), for which all hydrogens were added with CHARMM [45] in order to facilitate the incorporation of heme parameters. All Asp, Glu, Lys and Arg residues were treated as ionized, as were the peptide chain termini. Except as otherwise noted, histidines were assumed neutral with the proton on Nδ1. Ligand carboxylic acid groups were treated as ionized, except for methotrexate in **3dfr** and **4dfr** [23], in order to match the test conditions of Westhead et al. [8]. Methotrexate site B was studied in the case of **4dfr** and the buried oxamate site in the case of **1ldm** [26]. For the first two test sets, the hydrogen coordinates of the complexes were optimized with steepest-descent energy minimization (500 steps or until RMS force < 1 kcal/mol) with only hydrogen atoms free, while no energy minimizations were performed for the third set.

For the PRO_LEADS test cases, all water molecules were included as part of the receptor in order to match the test conditions used in the previous study [8]. Water molecules were also retained for **4dfr**, in order to match the test conditions used for Autodock [10]. For all other test cases, however, the water molecules were removed.

For the GOLD comparisons, some systems had undefined atoms. These were treated as follows. For Adipocyte Lipid-Binding Protein **1lic**, the chemistry of the oxidized cysteine is indefinite [43] as is the charge state of the ligand's sulfonic acid. Here, both are treated as ionized sulfonates. (We do not know how these groups were treated in the GOLD study.) The propanoic acid, which is far from the binding site, was omitted for simplicity. In the crystal structure of the Major Urinary Protein **1mup**, the side-chains of Asn 39, Arg 64, Lys 80, Glu 150 and Asp 157 are truncated after the Cβ; these residues were treated as alanines. Similarly, Glu 65 includes no side-chain coordinates and is here treated as a glycine. None of these residues is expected to interact significantly with the ligand, based upon the structure. The ligand from **1mup**, 2-(sec-butyl) thiazoline, was simulated in two different protonation states, the imine nitrogen being protonated or neutral. Similarly, Acetylcholinesterase, **1acl**, has multiple undefined side chains that extend into the solvent; again, all of these are replaced by alanines. The ligand 5-Fluoro 2′-Deoxyuridine 5′-Monophosphate from **1tdb** is treated as dianionic as suggested by Hardy et al. [46], and Glu 111 of Thymidylate Synthase, which is not fully defined, is replaced by an Alanine. The ligand SCH38057 from **1hri** can have two different protonation states, with a protonated or a neutral imidazole. The electron density map for **1hri** suggests a chloride ion adjacent to the imidazole of SCH38057, and this was taken to suggest that the imidazole is cationic. However, we ran calculations with the imidazole both neutral and charged. Based upon the low pH at which the crystal structure was solved, we treated the ligand's histidine in **1mcr** as ionized.

All ligand dihedral angles involving two heavy atoms and not involved in rings or controlling only a methyl group are treated as degrees of freedom, except that the two hydroxyl dihedrals of XK263 in its complex with HIV-1 protease [24] were held fixed in order to allow comparison with the calculations of Westhead et al. [8].

Exclusion ranges (see Promoting the Discovery of New Energy Minima, above) were adjusted through trial studies of the thrombin/argatroban (**1etr**) system. Runs of 20 dockings were carried out with $E_{\max} = 2$ kcal/mol, $N_{\text{hunt}} = 20\,000$, $N_{\text{search}} = 5$. The percentages of configurations with RMSD ≤ 1.5Å were 12, 15, and 8, for exclusion ranges of (5°, 2°, 0.25 Å), (15°, 7.5°, 0.5 Å), and (20°, 15°, 1 Å), respectively, where each triplet represents the excluded ranges for dihedral angles, overall rotation, and translation, respectively. Based upon these results, exclusion ranges of (15°, 7.5°, 0.5 Å) are used here, except as otherwise noted.

Table 1. Ligand-receptor systems

Description	PDB	N_{rot}	Res	Chrg	N_{atoms}
Influenza virus Neuraminidase/DANA	**1nsd**[a]	4	2.2	−1	36
Thrombin/NAPAP	**1ets**[b]	6	2.3	1	69
Dihydrofolate reductase/Methotrexate	**3dfr**[c]	7	1.7	0	55
Thrombin/Argatroban	**1etr**[d]	7	2.3	1	72
HIV-1 protease/XK263	**1hvr**[e]	8	1.8	0	84
Beta-Trypsin/Benzamidine	**3ptb**[f]	0	1.7	1	18
Lactate dehydrogenase/Oxamate 2	**1ldm**[g]	0	2.1	−1	8
Cytochrome P450CAM/Camphor	**2cpp**[h]	0	1.6	0	46
Immunoglobulin F_{AB} fragment/Phosphocholine	**2mcp**[i]	4	3.1	−1	24
Streptavidin/Biotin	**1stp**[j]	5	2.6	−1	31
Trypsinogen/Pancreatic Trypsin Inhibitor/Ile-Val	**3tpi**[k]	6	1.9	0	38
Carboxypeptidase A_α/Glycyl-L-Tyrosine	**3cpa**[l]	7	2.0	−1	30
Dihydrofolate Reductase/Methotrexate	**4dfr**[m]	7	1.7	0	33
H-RAS P21/Guanosine-5′-[B,G-Methylene] Triphos.	**121p**[n]	10	1.5	−2	46
Hemagglutinin/Sialic Acid	**4hmg**[o]	10	3.0	−1	39
Thermolys./N-(1-Carboxy-3-Phenylpropyl)-*L-*Leu-*L-Trp	**1tmn**[p]	13	1.9	−2	66
HIV-1 protease/VAC	**4phv**[q]	15	2.1	0	88
HIV-1 protease/KNI-272	**1hpx**[r]	15	2.0	0	87
Major Urinary Protein/Thiazoline	**1mup**[g1]	2	2.4	−1 (0)	23 (22)
Immunoglobulin/N-Acetyl-L-His-D-Pro-OH	**1mcr**[g2]	5	2.7	0	39
Thymidylate Synthase/5-F 2′-Deoxyuridine 5′-Monophosphate	**1tdb**[g3]	5	2.65	−2	31
Transthyretin/Thyroxine	**1eta**[g4]	6	1.7	0	34
Capsid Protein VP1 of Human Rhinovirus 14/SCH 38057	**1hri**[g5]	9	3.0	1 (0)	43 (42)
Acetylcholinesterase/Decamethonium	**1acl**[g6]	11	2.8	2	56
Intestinal Fatty Acid Binding Protein/Oleate	**1icn**[g7]	16	1.7	−1	55
Adipocyte Lipid-Binding Protein/Hexadecanesulfonic Acid	**1lic**[g8]	16	1.6	0	54
HIV-1 protease/Hydroxyethylene Isostere	**1aaq**[g9]	18	2.5	1	89

First five rows: comparisons with PRO_LEADS. Subsequent rows: comparisons with other docking methods. PDB: Protein Data Bank [19, 20] accession code; Res: resolution (Å) of crystal structure; Chrg: net charge assigned to ligand; N_{atoms}: number of atoms in ligand; N_{rot}: number of rotatable bonds in the ligand. a: [21]. b: [22]. c: [23]. d: [22]. e: [24]. f: [25]. g: [26]. h: [27]. i: [28]. j: [29]. k: [25]. l: [30]. m: [23]. n: [31]. o: [32]. p: [33]. q: [34]. r: [35]. g1: [**?**]. g2: [37]. g3: [38]. g4: [39]. g5: [40]. g6: [41]. g7: [42]. g8: [43]. g9: [44].

Results and discussion

Five receptor-ligand systems used previously to assess optimizers within the PRO_LEADS method [8] are studied first. These systems, listed in the first five rows of Table 1, provide a basis for adjusting heuristic parameters in the present docking algorithm and for comparing with the prior results. Then the results for a larger set of systems (lines 6–18 of Table 1) are compared with published results from FlexX [11], MCDOCK [12], Autodock [10], and the multistep approach of Wang et al. [13]. The last nine test cases are ones for which the GOLD algorithm failed for a variety of different reasons [47]. Finally, searches over larger sampling volumes are considered.

Comparison with PRO_LEADS methods

The docking program PRO_LEADS [8] was previously used to compare optimization algorithms for finding low-energy ligand conformations similar to experiment. Given a maximum of 200 000 energy evaluations for each docking, the most successful algorithms were found to be Simulated Annealing, Tabu Search, and a Genetic Algorithm. These carefully described studies provide a useful standard against which the present method can be evaluated.

Table 2. Success rates (%) (see text) for PRO_LEADS methods and for the present algorithm. SA: simulated annealing; TS: Tabu Search; GA: Genetic Algorithm; SR(n): success rate of new algorithm for the first n dockings of a 100 docking run, based upon 11 independent runs

System	PRO_LEADS			Present study		
	SA	TS	GA	SR(5)	SR(10)	SR(100)
1nsd	40	88	57	100	100	100
1ets	3	8	11	9	27.3	100
3dfr	90	93	76	90.9	100	100
1etr	30	39	13	36.4	72.8	100
1hvr	65	58	59	81.8	100	100

Conditions were matched as closely as possible to those used previously. Thus, crystallographic water molecules were retained as part of the receptor; the ligand atom closest to the center of coordinates of the ligand was allowed to sample positions in a $4 \times 4 \times 4$ Å cube around its crystallographic location; and success rates were reported as a function of the number of energy evalutions. Note, however, that the prior study used a different energy model, the Piecewise Linear Potential of Gehlhaar et al. [48] combined with a simplified internal energy function for the ligand. Here, the energy function is CHARMM 22 with nonbonded energy cutoffs ($E_{\max}$) of 2, 5 or ∞ kcal/mol for the interaction of each ligand atom with the receptor.

Eleven independent runs of 100 sequential dockings were performed with $N_{\text{hunt}} = 10\,000$, $N_{\text{search}} = 4$ and $N_{\text{fine}} = 2000$, for a total of 42 000 energy evaluations per docked configuration. The first 5 dockings of each run therefore require 210 000 energy evaluations and can be compared directly with the results of Westhead et al., because the latter used 200 000 energy evaluations per run. On an SGI R10000 processor, 5 dockings required 125–400 s of CPU time, depending upon the size of the ligand, while the PRO_LEADS methods required on the order of 500–700 s on an HP 735 computer.

Table 2 summarizes the results of the calculations with $E_{\max} = 2$ kcal/mol and compares them with the PRO_LEADS results. The figure of merit is the 'success rate' (SR) [8]. For the PRO_LEADS results, this is the percentage of 500 independent dockings yielding a configuration with RMSD ≤ 1.5 Å from the crystal structure. For the present method, the success rate is the percentage of the 11 runs for which the lowest-energy configuration found in the first 5, 10, or 100 dockings, thus, the best guess to date, has an RMSD ≤ 1.5 Å. Because the number of energy evaluations for the first 5 dockings is close to the value of 200 000 used in the PRO_LEADS tests, the success rate for the first 5 dockings, SR(5) can be compared directly with the PRO_LEADS results. These results suggest the new algorithm is at least as effective as those implemented in PRO_LEADS. Continuing the docking series to 10 and then 100 configurations yields continuing improvement in the results, indicating that the search algorithm is not trapped in an incorrect configuration.

Figure 1 further illustrates the results by graphing the RMSD of the lowest energy configuration to date, averaged over all 11 runs, versus the number of docking attempts. Results for $E_{\max}$ values of ∞, 5 and 2 kcal/mol illustrate the importance of the energy cutoff in the docking process, and show that low-energy, accurate docked configurations are generated within 15 attempts for every case except **1ets** (thrombin/NAPAP) which requires about 40 attempts. Setting $E_{\max} = 5$ produces small improvements in the RMSD of low-energy conformations in certain cases, especially in longer runs, but the best results for short runs are obtained with $E_{\max} = 2$.

Comparisons with AutoDock, FlexX, MCDOCK and a hybrid docking algorithm

This section compares the present method with published results for AutoDock [10], FlexX [11], a hybrid algorithm [13] and MCDOCK [12]. For ligands with 10 or more rotatable bonds, $N_{\text{hunt}} = N_{\text{fine}}$ was set to 20 000, N_{search} to 4, and $E_{\max}$ to 2 kcal/mol. For the less flexible ligands, we used $N_{\text{hunt}} = N_{\text{fine}} = 10\,000$, $N_{\text{search}} = 4$ and $E_{\max} = 3$ kcal/mol, based upon an impression that smaller ligands are best docked with a slightly larger energy cutoff. The remaining parameters are as described above. Five independent runs of 100 dockings were carried out for each ligand-receptor system.

Accuracy

Table 3 presents the energy and RMSD of the single most stable configuration found in the first 5, 20, and 100 dockings of the five runs, and Table 4 presents the success rates (see above) for the same runs. Table 3 includes published results of the docking algorithms for the same systems. It is worth noting that the AutoDock results are for the same receptor-ligand systems that were used to calibrate the AutoDock energy model

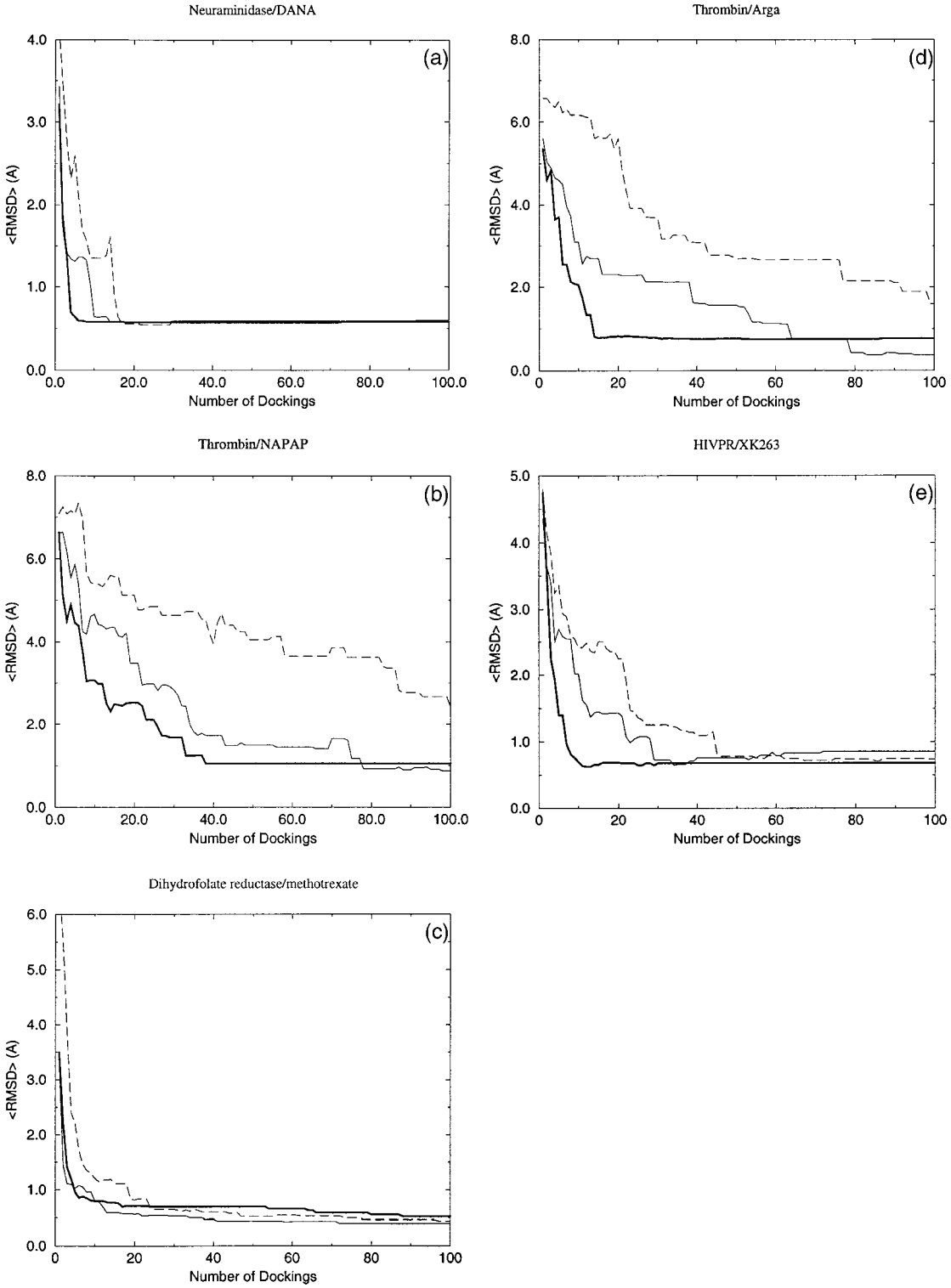

Figure 1. (a) Average over 11 independent docking runs of the RMSD of the lowest energy structure relative to the crystal structure, for neuraminidase/DANA (**1nsd**). Dashed line: No energy cutoff. Thin line: E_{max} 5 kcal/mol. Heavy line: E_{max} 2 kcal/mol. (b) Same as (a) for thrombin/NAPAP (**1ets**). (c) Same as (a) for DHFR/methotrexate (**3dfr**). (d) Same as (a) for thrombin/argatroban (**1etr**). (e) Same as (a) for HIV-1 protease/XK263 (**1hvr**).

Table 3. Comparison of docking algorithms

System	N_{rot}	5 × 5		5 × 20		5 × 100		AD	FLX	HYB	MCD
		E	R	E	R	E	R				
3ptb	0	−35.30	0.4	−35.30	0.4	−35.30	0.4	0.21	–	–	–
1ldm1	0	−31.80	0.6	−31.80	0.6	−31.80	0.6	–	0.6	–	0.4
2cpp	0	−35.60	0.4	−35.60	0.4	−35.60	0.4	0.81	–	–	–
2mcp	4	−30.20	1.4	−30.40	1.5	−30.60	1.8	0.85	–	–	–
1stp	5	−50.70	0.4	−50.70	0.4	−50.90	0.4	0.69	0.81	1.0	0.41
3tpi	6	−68.00	0.3	−68.00	0.3	−68.00	0.3	–	0.58	1.39	0.25
3cpa	7	−70.20	1.2	−71.80	1.2	−71.80	1.2	–	3.08	0.62	0.83
4dfr	7	−63.50	0.8	−64.50	0.6	−64.60	0.7	0.95	–	–	–
1hvr	8	−92.30	0.9	−92.60	0.9	−92.60	0.9	0.76	–	–	–
121p	10	−158.20	1.3	−169.20	1.1	−175.90	0.9	–	2.00	1.02	0.99
4hmg	10	−34.90	1.0	−35.10	1.0	−35.10	1.0	1.11	–	–	–
1tmn	13	−55.40	2.7	−68.60	2.2	−72.20	0.7	–	0.87	1.93	1.84
4phv	15	−79.30	0.8	−91.30	1.7	−98.70	0.3	–	1.04	1.97	0.55
1hpx	15	−74.40	1.7	−91.80	2.6	−92.70	2.9	–	–	–	–

E: energy (kcal/mol) of most stable configuration found so far in the first 5, 20, or 100 dockings of 5 independent docking runs. R: RMSD of this lowest-energy configuration. Results for other models are RMSD in Å. AD: Autodock [10]. FLX: FlexX [11]. HYB: hybrid algorithm [13]. MCD: MCDOCK [12].

Table 4. Success rates (%) (see text) with the present algorithm, as function of the number of dockings in a run

System	SR(5)	SR(20)	SR(100)
3ptb	100	100	100
1ldm	60	100	100
2cpp	100	100	100
2mcp	40	80	40
1stp	60	100	100
3tpi	100	100	100
3cpa	100	100	100
4dfr	60	100	100
1hvr	80	100	100
121p	40	40	100
4hmg	80	80	100
1tmn	0	0	80
4phv	20	0	80
1hpx	0	0	0

and thus may not represent full-fledged tests of that method.

With the present method, 5 runs of 5 dockings yield a low-energy configuration with an RMSD of 1.5 Å or less for 12 of the 14 systems. Indeed, the results for the first 5 dockings are similar in accuracy to those of other published methods. Extending the present runs yields new configurations of lower energy, especially for ligands with 10 or more rotatable dihedral angles. Some of the new configurations are closer to the crystal structures, as in the case of **1tmn**, but others are less accurate, notably in the case of **1hpx**. Note that these new, low-energy configurations are important, because energy is typically used to judge the relative affinity of various ligands for a receptor.

For 2 systems, the RMSD of the most stable energy minimum discovered is greater than 1.5 Å, and the success rates in Table 4 are correspondingly low. Closer examination shows that these dockings are still quite reasonable, as now discussed.

2mcp The most stable docked conformation for this phosphorylcholine-antibody F_{ab} complex has an RMSD of 1.8 Å. Nonetheless, the algorithm positions the two ionized moieties of the ligand correctly, and it is only the intervening chain that is seriously misplaced (Figure 2).

1hpx The most stable configuration found for KNI-272 in HIV-1 protease has an RMSD of 2.9 Å. However, as illustrated in Figure 2, this large RMSD results from deviations in the C-terminal naphthyl moiety of the ligand which, in the crystal structure of the complex, projects into the solvent and has large crystallographic B-factors typical of surface side-chains [35]. Furthermore a conformational study performed on KNI-272 in solution shown a quite high variation of this C-terminal part [50]. Thus, the docked con-

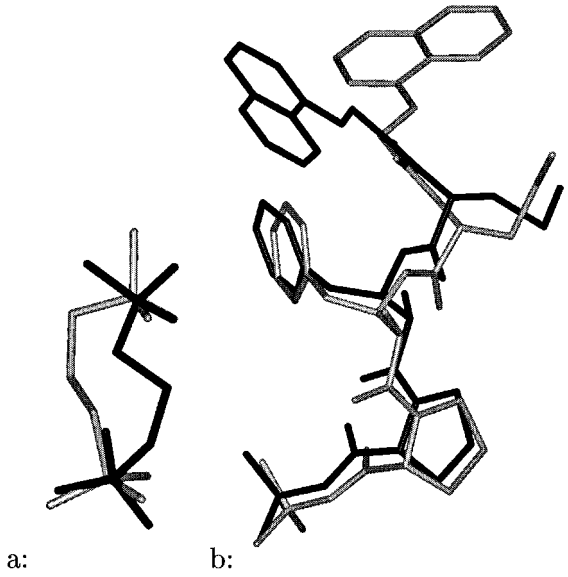

Figure 2. Superpositions of the crystal structures (in black) and the global minima found by the present algorithm (in grey) for **2mcp** (a) and **1hpx** (b). Hydrogen atoms are not included.

figuration agrees well with the well-resolved portions of the crystal structure, and the part of the ligand that deviates from the crystal structure is known to be flexible.

The performance of the algorithm for **2cpp** is also of considerable interest. For **2cpp**, the docking algorithm yields a 100% success rate (RMSD 0.4 Å) in 5 dockings, but this result is obtained only when the hydroxyl hydrogen of Tyr 96 is positioned to form a hydrogen-bond to the carbonyl oxygen of the camphor ligand. Thus, a test run with the hydrogen rotated away from the camphor yielded an RMSD of 2.4 Å. Examination of the crystal structure strongly suggests that the hydrogen bonding pattern used in the successful docking runs is correct, because it allows Thr 101 to donate a hydrogen-bond to Tyr 96 even while Tyr 96 donates to the camphor. If Tyr 96 donated to Thr 101, the hydroxyl hydrogen of Thr 101 would have no acceptor. This analysis of **2cpp** illustrates the sensitivity of the scoring (energy) function to the conformation of the receptor. Clearly, it will be important in the future to incorporate some receptor degrees of freedom in the optimization procedure; i.e., to treat part of the receptor as flexible during docking.

Speed

The available data suggest that the speed of the present method is in the same general range as that of a number of previously described docking algorithms. The present algorithm requires $N_{\text{hunt}}N_{\text{search}} + N_{\text{fine}}$ energy evaluations to generate one docked conformation, and at least 5 such conformations should be generated in order to have a high probability of yielding a good match to experiment, under the present conditions. If, to be more conservative, we allow for 5 independent runs of 5 dockings each ($5 \times 5 = 25$ dockings), then the runs in Table 3 correspond to 1.25×10^6 energy evaluations for molecules with fewer than 10 rotatable bonds, and 2.5×10^6 for the more flexible molecules (see above). Each individual docking requires 0.1 CPU-min (oxamate/LDH) to 5 CPU-min (1 tmn) on an SGI R10000 processor, with a median time for the ligands studied here of about 1.2 min. Thus, the 5×5 results were typically completed in about 30 min and a single run of 100 dockings takes about 120 min. For comparison, the AutoDock calculations referenced in our Table 3 used 10 docking runs of 1.5×10^6 energy evaluations each, where each run required 4.5–41.3 min on an SGI R4400 computer. MCDOCK required about 0.2–16 min to generate a docked configuration [12]. It is difficult to assess the timings of the hybrid method of Wang et al. because the published description does not seem to state the number of docked configurations needed to obtain converged results. However, FlexX appears to runs faster than the present algorithm. When applied to 200 protein-ligand complexes [51], FlexX required 0.1–492 s (mean 93) on a SUN Ultra-30 workstation. (The RMSD of the highest-ranked solutions relative to the crystal structure averaged 4.0 Å.) Note, though, that the ligands in the FlexX study, averaging 23 atoms, are smaller than those in the present study, averaging 48 atoms.

Tests for GOLD's challenging cases

Another important docking algorithm, GOLD, speeds its search for stable ligand conformations by generating possible hydrogen bonding patterns between the ligand and the receptor. GOLD has shown good performance in extensive testing [47], and the test cases already discussed above include 7 that were used to test GOLD. Thus, GOLD yielded 'good' results for **4dfr** and **4phv**; 'close' results for **1ldm**, **3cpa**, and **1tmn**; 'errors' for **1etr**; and 'wrong' results for **2mcp**. The present algorithm was deemed 'successful' (RMSD within 1.5 Å) for all of these cases except for **2mcp**, where the RMSD was 1.8 Å. Based upon the analysis above, our **2mcp** results would probably have been classified either 'close' or 'errors'.

Table 5. Performance of docking algorithm as a function of the size (Å) of the cubic search box and of the exclusion ranges (Excl. Ranges) for dihedral rotations, overall rotation, and translation, respectively. Other symbols defined above

Box length	N_{hunt}	Excl. ranges	5 × 5		5 × 20		5 × 100	
			E	R	E	R	E	R
5	10000	15.0–7.5–0.5	−92.6	0.9	−92.7	0.9	−92.7	0.9
10	10000	15.0–7.5–0.5	−92.1	0.5	−92.5	0.9	−92.6	0.9
15	10000	15.0–7.5–0.5	−90.7	0.8	−92.1	0.9	−92.6	0.9
15	50000	15.0–7.5–0.5	−71.4	1.5	−92.1	0.9	−92.7	0.9
15	10000	15.0–7.5–1.0	−89.9	0.9	−92.1	0.9	−92.7	0.9
15	50000	15.0–7.5–1.0	−90.1	0.9	−92.7	0.9	−92.7	0.9

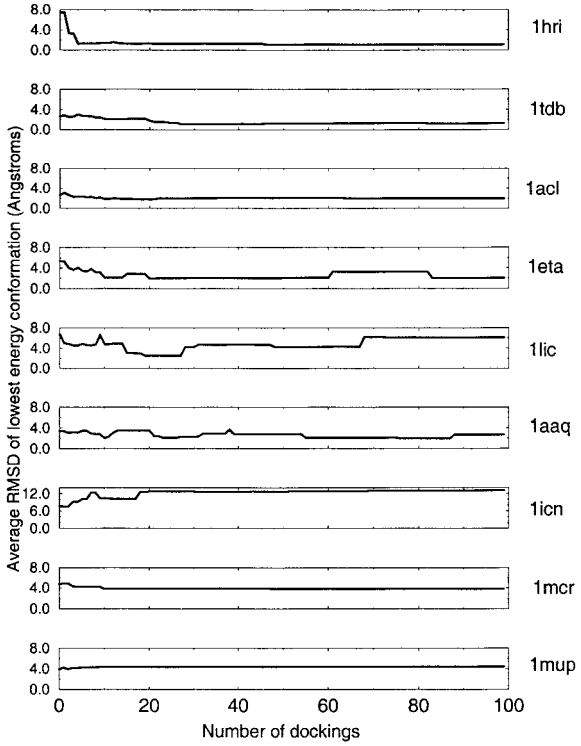

Figure 3. Average RMSD, over 5 independent docking runs, of the lowest energy conformation found to date for the GOLD test cases.

The present algorithm was evaluated further via tests for 9 additional test cases drawn from the 28 systems (lines 19–28 of Table 1 of the GOLD paper) that the GOLD algorithm failed to dock correctly ('error' or 'wrong' [47]) for various reasons. Three failures (**1hri**, **1acl**, **1mup**) were attributed to a lack of hydrogen-bonding groups, required for the GOLD algorithm; two failures (**1icn**,**1lic**) were attributed to excessive flexibility of the ligand; two (**1mcr**, **1eta**) were attributed to problems with the force-field; and two (**1aaq**, **1tdb**) could not be explained. Our results for these challenging tests are assessed not only with the 1.5 Å RMSD criterion used above, but also by subjective review in accordance with the prior study [47]. (The significance of this distinction is clear from the fact that Table 3 in the GOLD study shows that 16 of the 71 docking tests considered successful ('good' or 'close') deviated from experiment by more than 1.5 Å RMSD.) Here, as for the previous computations, 5 runs of 100 dockings were performed for each test case. All computations used $N_{hunt} = N_{fine} = 20\,000$, $N_{search} = 4$, $4 \times 4 \times 4$ Å^3 translational ranges, and $E_{max} = 2$ kcal/mol.

Figure 3 summarizes the results with graphs of the 5-run mean of the RMSD for the lowest energy conformation found to date in each run (as in Figure 1). Figure 4 shows the lowest energy structure found for each of these complexes over all docking runs, along with the experimental structures. Good results were obtained in 6 cases (**1hri**, **1tdb**, **1acl**, **1eta**, **1lic**, **1aaq**), and poor results were obtained for the remaining 3 cases (**1icn**, **1mcr**, **1mup**), as now discussed.

1hri Docking the ligand (SCH 38057) in its protonated form into Capsid Protein VP1 of Human Rhinovirus 14 yielded a global energy minimum with an RMSD of 1.2 Å (Figures 3 and 4). The less probable [40] neutral form yielded an RMSD of 10 Å. Comparison of the structures did not reveal any specific interactions between the ligand and the receptor that appeared to be responsible for the change in conformation when the charge was changed. As noted previously [47], the failure of GOLD in this case resulted from GOLD's restriction to conformations in which hydrogen bonds are formed between the ligand and receptor, because

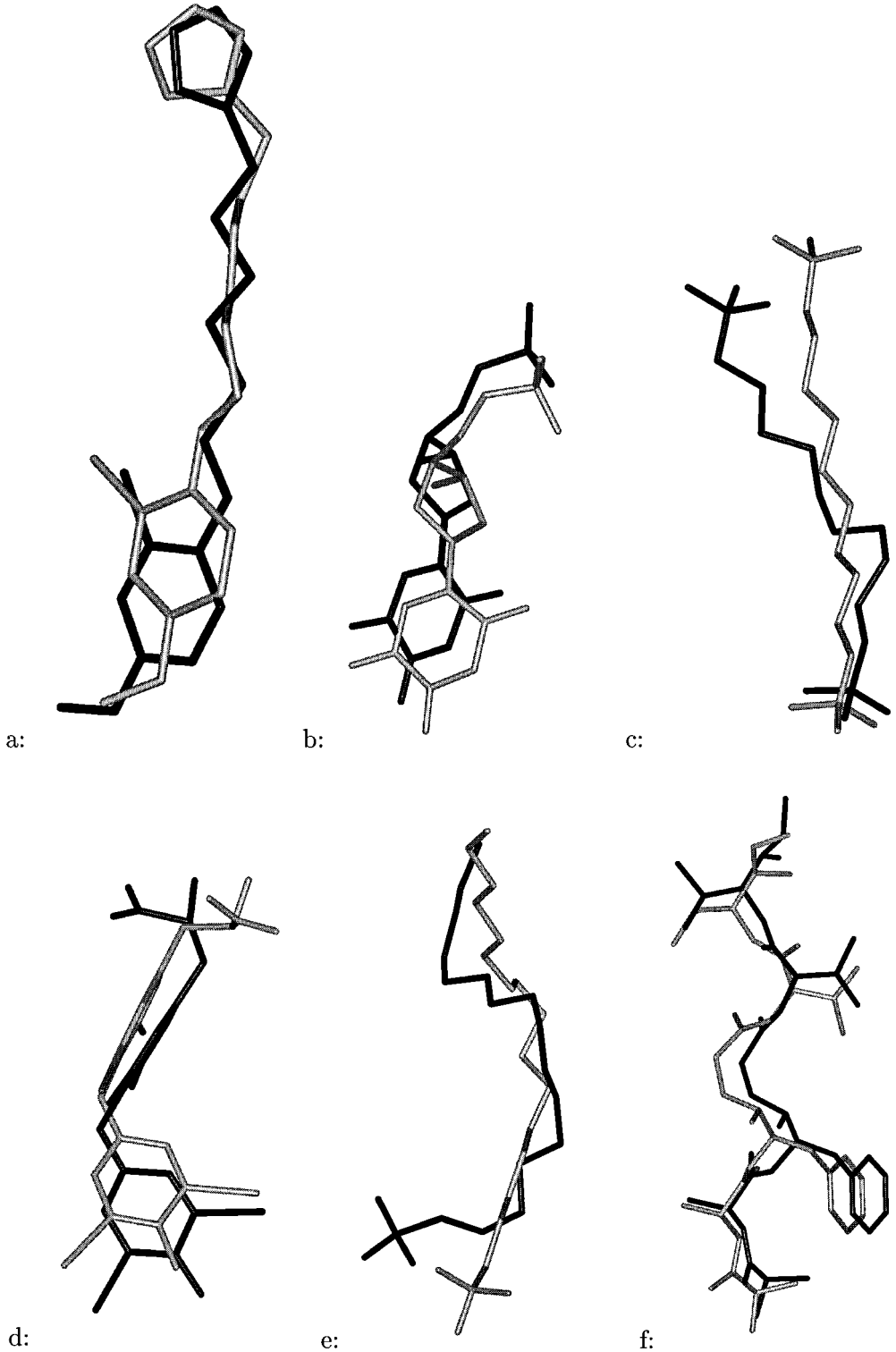

Figure 4. Superpositions of the crystal structures (in black) and the global energy minima (in grey) found by the present algorithm for the third set of test cases. Hydrogen atoms are not included in the drawings. a: **1hri**. b: **1tdb**. c: **1acl**. d: **1eta**. e: **1lic**. f: **1aaq**. g: **1icn**. h: **1mcr**. i: **1mup**.

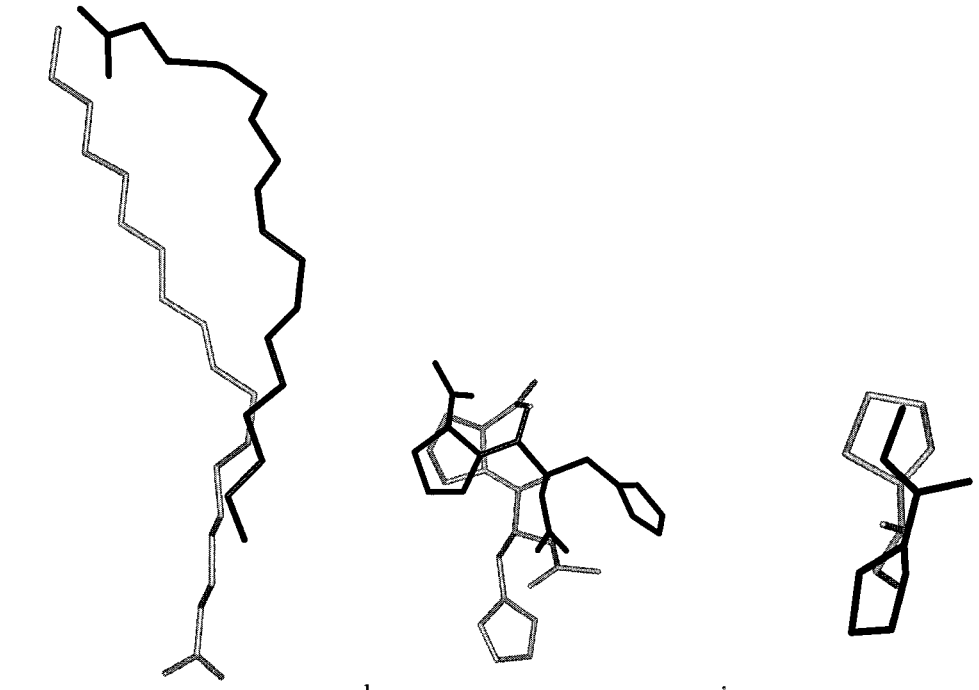

Figure 4. Continued

no such hydrogen bonds are observed in the crystal structure.

1tdb The ligand, 5-Fluoro 2′-Deoxyuridine 5′-Monophosphate, docked into Thymidylate Synthase with an RMSD of 1.3 Å. It is not clear why GOLD did not dock this ligand accurately.

1acl Acetylcholinesterase has a long, cylindrical active site pocket, with one end deep in the protein and the other opening to the surface. Two Glu residues lie at the base of the binding site. The ligand, decamethonium, consists of a long aliphatic chain with a tetramethylammonium group at each end. The combination of steric constraint and electrostatics guides the present docking algorithm to a global energy minimum with an RMSD of 2.0 Å. Most of the deviation from the crystal conformation is localized in the tetramethylammonium group near the entry of the binding site. GOLD could not be used for this ligand because it lacks hydrogen bond donors and acceptors [47].

1eta Thyroxine is successfully docked into the transthyretin receptor with an RMSD of 2.1 Å relative to experiment [39]. Most of the deviation is localized in the carboxylic moiety, which is solvent accessible. GOLD's difficulties in this case were attributed to an inadequate treatment of a short I...O interaction.

1lic The docking of hexadecanesulfonic acid into the Adipocyte Lipid-Binding Protein [43] gives a global energy minimum at 2.4 Å RMSD from the crystal structure. The main deviations in this structure are localized at the sulfonic acid head, while the aliphatic chain is well positioned. The quality of this result is surprising, given the size of the cavity and our neglect of the 10 crystallographic water molecules in the binding site. The relatively high mean RMSD in Figure 3 results from the presence of a somewhat less stable energy-minimum in which the ligand is rotated about 180° relative to the correct conformation, resulting in an RMSD of 13 Å. Use of a larger translational search region for **1lic** actually improves the docking results (see below) because it allows the ligand to find an even lower-energy structure with RMSD 2.2 Å. GOLD's difficulty with **1lic** probably results in large part from the paucity of hydrogen bonding groups on the ligand [47].

1aaq The algorithm docks this inhibitor into the active site of HIV protease in two orientations related by a 180° rotation whose axis matches the dimer symmetry axis of the protease. Because of the symmetry of the

protease, the energies of the two orientations differ by only 1 kcal/mole, and we regard both orientations of the inhibitor as legitimate, and allow for this rotation in computing the RMSD values in Figure 3. The global minimum, illustrated in Figure 4, has an RMSD of 1.1 Å relative to experiment [44].

1icn Fitting oleate into R106Q-Intestinal Fatty Acid Binding Protein poses a challenge similar to that of **1lic**; i.e., a large, binding cavity containing multiple crystallographic water molecules, and a long, flexible aliphatic chain. In our lowest energy conformation, the orientation of the ligand is reversed relative to the crystal conformation, with an RMSD of 13 Å, and the docked conformation with lowest RMSD (2.4 Å) has an energy 4.7 kcal/mol higher than the global minimum. These results are not changed significantly when the crystallographic waters are included as part of the receptor. The reversed orientation of this ligand in the global minimum appears to result in part from a favorable interaction (3.5 Å) between the carboxylate group of the ligand and the ammonium group of Lys 25 at the protein surface. Surprisingly, the crystallographic conformation of the ligand provides the carboxylate group with neither hydrogen bonds nor salt-bridges at its location within the binding site of this R106Q mutant. Thus, the bound conformation of this ligand is difficult to rationalize, independent of any docking calculations. (In contrast, the crystal structure of the wild-type receptor with the similar ligand myristate shows a tight salt-bridge between the carboxylate of the ligand and Arg 106 deep within the binding site.) Indeed, the electron density of the end of the ligand within the binding site is smeared and was interpreted as showing multiple configurations of the carboxylate group[42]. It is thus conceivable that the observed electron density is consistent with the calculated docked orientation of oleate, which would provide a strong stabilizing interaction for the carboxylate moiety with Lys 25.

1mcr The present docking algorithm yields a global energy minimum 4.0 Å away from the experimental conformation for this complex of N-Acetyl-L-His-D-Pro-OH with an immunoglobulin complex, whether or not the ligand His is treated as ionized. As noted by Jones et al. [47], this system may be difficult to model without careful treatment of the NH-π-interaction observed in the crystal structure.

1mup Docking of thiazoline into the Major Urinary Protein generates a global minimum with an RMSD of 4.0 Å, whether the imine is treated as neutral or protonated, and whether or not the crystallographic water molecules are included as part of the receptor. GOLD failed for this case because no hydrogen bonds are formed between ligand and receptor in the crystal structure, but it is not clear why the present docking algorithm fails here.

Table 6. Success rates for docking algorithm as a function of the size (Å) of the cubic search box and of the exclusion ranges (see caption of prior tables for symbols)

Box length	N_{hunt}	Excl. ranges	SR(5)	SR(20)	SR(100)
5	10000	15.0–7.5–0.5	80	100	100
10	10000	15.0–7.5–0.5	80	100	100
15	10000	15.0–7.5–0.5	20	40	100
15	50000	15.0–7.5–0.5	0	60	100
15	10000	15.0–7.5–1.0	20	60	100
15	50000	15.0–7.5–1.0	20	60	100

Larger sampling volumes

The calculations described above used a translational volume of 64 Å^3. The consequences of increasing this translational search volume were investigated for several of the test systems.

An initial exploration was carried out for the XK263-HIV-1 protease (**1hvr**) system. In these tests, the nonbonded grids were extended to $44 \times 44 \times 44$ Å^3 in order to guarantee that sampling would not be artificially constrained. Each line in Table 5 summarizes the results of five runs of 20 dockings, while Table 6 gives the corresponding success rates. As the search volume increases from 125 to 3375 Å^3, the success rate drops, but 5 runs of 10 dockings each still yields a low-RMSD, low-energy configuration. Surprisingly, increasing N_{hunt} from 10 000 to 50 000 does not improve the results significantly, by the measures used here. Closer inspection reveals that configurations with RMSD of 1.54 Å and 1.51 Å appear early in two of these longer runs, but these do not contribute to the computed success rate, because the criterion for success is 1.5 Å. Increasing the translational exclusion range (see Methods) from 0.5 to 1.0 Å does not markedly alter the results.

Further tests were carried out for 6 of the challenging GOLD test systems discussed above: **1hri**, **1tdb**, **1acl**, **1eta**, **1aaq** and **1lic**. The sampling box was increased to $10 \times 10 \times 10$ Å^3, and the nonbonded grids were extended to $40 \times 40 \times 40$ Å^3 to accomodate the larger sampling ranges. Here $N_{\text{hunt}} = N_{\text{fine}} = 20\,000$ and $E_{\text{max}} = 2$ kcal/mol. The results are summarized

Table 7. Success rates (see Table 2) for 5 independent docking runs with translational sampling volumes of $10 \times 10 \times 10$ Å

System	SR(5)	SR(10)	SR(100)	RMSD (Å)
1hri	60	80	100	1.5
1tdb	40	60	80	1.3
1acl	60	100	100	2.0
1eta	0	20	0	5.6
1lic	20	80	100	2.2
1aaq	20	20	80	1.7

in Table 7, where the success rates (SR) are now based upon subjective review of the ligand conformations, and the final column lists the RMSD of the lowest-energy conformation found. The docking procedure continues to perform quite well except for **1eta**, for which the global energy minimum now deviates significantly from the crystal conformation. As noted above, the results for **1lic** actually improve relative to the smaller search volume because the global energy minimum positions the central atom of the ligand outside the smaller sampling region.

Conclusions

The docking algorithm described here relies upon a novel optimization technique that borrows ideas from the Global Underestimator method [2], Genetic Algorithms [3, 4, 5], and Poling [6] or Tabu Search [7]. Comparisons with published data suggest that the new method is competitive, in terms of both speed and accuracy, with other energy-based methods. In most tests, a configuration with RMSD less than 1.5 Å was found within 25 dockings, though extending the runs yielded incorrect configurations of significantly lower energy in a few cases. There was no correlation between the number of rotatable bonds and the failure rate. Indeed, one of the most rigid ligands, thiazoline (**1mup**) was not docked well, while reasonably good results were obtained for some of the more flexible ligands, such as hexadecanesulfonic acid (**1lic**) and VAC (**4phv**), so long as adequate sampling was used. Calculations with extended search regions also perform quite well. Finally, test calculations for **2cpp** highlight the importance of correctly positioning hydroxyl hydrogens of the receptor prior to docking the ligand into the receptor site. Given that the positions of hydrogens – and other atom-types – depend to some degree upon the positioning of the ligand, this observation highlights the importance of optimizing the conformation of the receptor site as part of the docking procedure. In summary, the results presented in this paper motivate continued development of the docking algorithm, with the goals of increasing computational speed and incorporating receptor flexibility.

Acknowledgements

This work was supported by the National Institutes of Health (GM54053 to MKG) and the National Institute of Standards and Technology.

References

1. Head, M.S., Given, J.A. and Gilson, M.K., J. Phys. Chem., 101 (1997) 1609.
2. Phillips, A.T., Rosen, J.B. and Walke, V.H., Dimacs Series in Discrete Math. Theoret. Comput. Sci., 23 (1995) 181.
3. Holland, J.H. Adaptation in Natural and Artificial Systems. MIT Press, Cambridge, MA, 1975.
4. DeJong, An analysis of the behavior of a class of genetic adaptive systems, Phd Thesis, Michigan, 1975.
5. Goldberg, D.E. Genetic Algorithms in Search, Optimisation and Machine Learning. Addison Wesley, Reading, MA, 1989.
6. Smellie, A., Teig, S.L. and Towbin, P., J. Comput. Chem., 16 (1995) 171.
7. Glover, F. and Laguna, M. Tabu Search, Kluwer Academic Publishers, Dordrecht, 1997.
8. Westhead, D.R., Clark, D.E. and Murray, C.W., J. Comput. Aid. Mol. Des., 11 (1997) 209.
9. Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R. and Eldridge, M.D., Proteins, 33 (1998) 367.
10. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comp. Chem., 19 (1998) 1639.
11. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
12. Liu, M. and Wang, S., J. Comput. Aid. Mol. Des., 13 (1999) 435.
13. Wang, J.W., Kollman, P.A. and Kuntz, I.D., Proteins, 36 (1999) 1.
14. Jones, G., Willett, P. and Glen, R.C., J. Mol. Biol., 245 (1995) 43.
15. MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C.; Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R. Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M., J. Phys. Chem. B102 (1998) 3586.
16. The QUANTA Molecular Modeling Program. Molecular Simulation, Incorporated.
17. Pattabiraman, N., Levitt, M., Ferring T.E. and Langridge, R., J. Comput. Chem., 6 (1985) 432.

18. Given, J.A. and Gilson, M.K., Proteins, 33 (1998) 475.
19. Bernstein, F.C., Koetzle, T.F., Williams, T.F., Meyer, G.J.B., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi,T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
20. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucl. Acids Res., 28 (2000) 235.
21. Burmeister, W.P., Henrissat, B., Bosso, C., Cusack, S. and Ruigrok, R.W.H., Structure, 1 (1993) 19.
22. Brandstetter, H., Turk, D., Hoeffken, H.W., Grosse, D., Stuerzebecher, J., Martin, P.D., Edwards, B.F.P. and Bode, W., J. Mol. Biol., 226 (1992) 1085.
23. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J. , J. Biol. Chem. 257 (1999) 13650.
24. Lam, P., Jadhav, P., Eyermann, C., Hodge, N., Ru, Y., Bacheler, L., Meek, J., Otto, M., Rayner, M., Wong, Y., Chang, C.-H., Weber, P., Jackson, D., Sharpe, T. and Erickson-Viitanen, S., Science, 263 (1994) 380.
25. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Huber, R., Acta Crystallogr., B39 (1983) 480.
26. Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G., J. Mol. Biol., 198 (1987) 445.
27. Poulos, T.L., Finzel, B.C. and Howard, A.J., J. Mol. Biol., 87 (1987) 687.
28. Padlan, E.A., Cohen, G.H. and Davies, D.R., Ann. Inst. Pasteur Immunol., 136 (1985) 271.
29. Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J. and Salemme, F.R., Science, 243 (1989) 85.
30. Christianson, D.W. and Lipscomb, W.N., Proc. Natl. Acad. Sci. USA, 83 (1986) 7568.
31. Krengel, U., Schlichting, I., Scheidig, A., Frech, M., John, J. Lautwein, A., Wittinghofer, F., Kabsch, W. and Pai, E.F., Nato Asi Ser., A220 (1991) 183.
32. Weis, W.I., Bruenger, A.T., Skehel, J.J. and Wiley, D.C., J. Mol. Biol., 212 (1990) 737.
33. Monzingo, A.F. and Matthews, B.W., Biochemistry, 23 (1984) 5724.
34. Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.
35. Baldwin, E.T., Bhat, T.N., Gulnik, S., Liu, B., Topol, I.A., Kiso, Y., Mimoto, T., Mitsuya, H. and Erickson, J.W., Structure, 3 (1995) 581.
36. Bocskei, Z., Groom, C.R., Flower, D.R., Wright, C.E., Phillips, S.E.V., Cavaggioni, A., Findlay, J.B.C. and North, A.C.T., Nature, 360 (1992) 186.
37. Edmundson, A.B., Harris, D.L., Fan, Z.-C., Guddat, L.W., Schley, B.T., Hanson, B.L., Tribbick, G. and Geysen, H.M., Proteins Struct. Funct. Genet., 16 (1993) 246.
38. Perry, K.M., Carreras, C.W., Chang, L.C., Santi, D.V. and Stroud, R.M., Biochemistry, 32 (1993) 7116.
39. Hamilton, J.A., Steinrauf, L.K., Braden, B.C., Liepnieks, J.J., and Benson, M.D., J. Biol. Chem., 268 (1993) 2416.
40. Oren, D.A., Jacobo-Molina, A., Williams, R.L., Kamer, G., Rubenstein, D.A., Li, Y., Rozhon, E., Cox, S., Buontempo, P., O'Connell, J., Schwartz, J., Miller, G., Bauer, B., Versace, R., Pinto, P., Ganguly, A., Girijavallabhan, V. and Arnold, E., J. Mol. Biol., 230 (1993) 857.
41. Harel, M., Schalk, I., Ehret-Sabattier, L., Bouet,F., Goeldner, M., Hirth, C., Axelsen, P., Silman, I. and Sussman, J., Proc. Natl. Acad. Sci. USA, 90 (1993) 9031.
42. Eads, J.C., Sacchettini, J.C., Komminga, A. and Gordon, J.I., Cell, 78 (1994) 325.
43. Lalonde, J.M., Bernlohr, D.A. and Banaszak, L.J., Biochemistry, 33 (1994) 4885.
44. Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek, Jr. T.A., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Minnich, M., Petteway, S.R., and Metcalf, B.W., Biochemistry, 31 (1992) 6646.
45. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States,D.J., Swaminathan S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
46. Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V. and Stroud, R.M., Science, 235 (1987) 448.
47. Jones, G., Willett, P. and Glen, R.C., Leach, A.R., and Taylor, R., J. Mol. Biol., 207 (1997) 727.
48. Gehlhaar, D.G., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J. and Freer, S.T., Chem. Biol., 2 (1995) 317.
49. Jano, I., Compt. Rend. Acad. Sci. Paris, 261 (1965) 103.
50. David, L., Luo, R. and Gilson, M.K., J. Phys. Chem. A 103 (1999) 1031.
51. Kramer, B., Rarey, M. and Lengauer, T., Proteins Struct. Funct. Genet., 37 (1999) 228.