Journal of Computer-Aided Molecular Design, **15:** 411–428, 2001.
KLUWER/ESCOM
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases

Todd J. A. Ewing[a,b], Shingo Makino[a,c], A. Geoffrey Skillman[a,d] and Irwin D. Kuntz[a,*]
[a]*Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA, 94143-0446, USA;* [b]*Camitro Corporation, 4040 Campbell Ave, Menlo Park, CA 94025, USA;* [c]*Central Research Laboratory, Ajinomoto Co. Inc., 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki-shi 210, Japan;* [d]*Open Eye Scientific Software, 3350 Winische Way, Santa Fe, NM 87501, USA*

Received 13 March 2000; accepted 12 January 2001

Key words: database searching, flexible docking, incremental construction, molecular docking, virtual screening

Abstract

In this paper we describe the search strategies developed for docking flexible molecules to macomolecular sites that are incorporated into the widely distributed DOCK software, version 4.0. The search strategies include incremental construction and random conformation search and utilize the existing Coulombic and Lennard-Jones grid-based scoring function. The incremental construction strategy is tested with a panel of 15 crystallographic testcases, created from 12 unique complexes whose ligands vary in size and flexibility. For all testcases, at least one docked position is generated within 2 Å of the crystallographic position. For 7 of 15 testcases, the top scoring position is also within 2 Å of the crystallographic position. The algorithm is fast enough to successfully dock a few testcases within seconds and most within 100 s. The incremental construction and the random search strategy are evaluated as database docking techniques with a database of 51 molecules docked to two of the crystallographic testcases. Incremental construction outperforms random search and is fast enough to reliably rank the database of compounds within 15 s per molecule on an SGI R10000 cpu.

Introduction

The computational study of molecular recognition is a important component of structure-based drug design. The molecular docking problem is generally cast as a problem of finding the low-energy binding modes of a small molecule, or ligand, within the active site of a macromolecule, or receptor, whose structure is known. Solving the docking problem computationally requires an accurate representation of the molecular energetics as well as an efficient algorithm to search the potential binding modes.

Although computational algorithms have been described for simultaneously searching the conformation space of the receptor and one, or several ligands, such as simulated annealing [1], evolutionary algorithms [2–7], and other heuristic searches [8, 9], they are largely devoted to studies of individual ligands. Much active research is still required for the development of algorithms sufficiently fast to process large databases of potential ligands. Initial attempts to perform this task made the problem tractable by fixing the conformation of both the ligand and receptor [10, 11]. Recent development of database search algorithms has sought to include limited treatment of molecular flexibility for the receptor [12] and the ligand [13, 14, 19]. There has emerged a consensus on the efficient treatment of ligand flexibility using the anchor and grow – or incremental construction – technique for reconstructing crystal complexes [15–18] and in addition for processing small molecule databases [14, 19, 21].

The original DOCK algorithm addressed rigid body docking using a geometric matching algorithm to superimpose the ligand onto a negative image of the binding pocket [10, 22]. Important features for database processing [11] including force-field based

*To whom correspondence should be addressed.

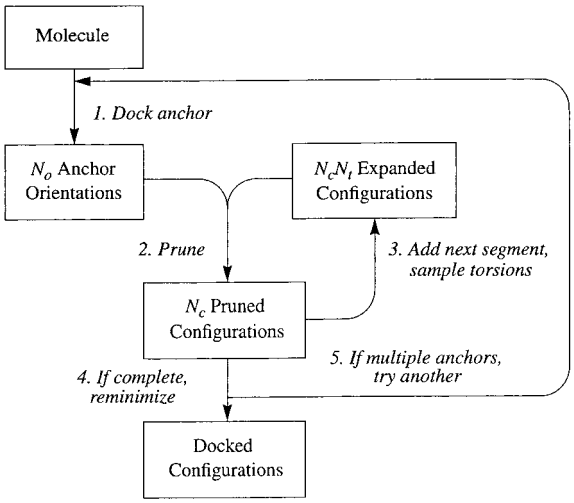


Figure 1. Flowchart of the Anchor and Grow docking algorithm.

scoring [23] and on-the-fly optimization [24] were introduced in earlier versions. The current release, version 4.0, incorporates an improved matching algorithm for rigid body docking [25] and an algorithm for flexible ligand docking. It is the purpose of this paper to evaluate the DOCK 4.0 algorithm's treatment of ligand flexibility – incremental construction and random search.

Incorporation of a new search algorithm into a widely distributed scientific tool, such as DOCK, places a unique constraint on the development process. In addition to the issue of performance, we must also address the issue of usability. The algorithm must be intuitive and easy to control by the investigator. This requires that the algorithm be constructed with the following design goals: we maximize the control that the user has over the docking process using adjustable, execution-time parameters; we simultaneously minimize the number of necessary parameters; and we maximize the intuitive nature of the parameters so that adjustments made to them have predictable consequences.

In our evaluation of the algorithm we will consider the speed and accuracy of the search technique. Like previous authors, we will test how accurately it can reconstruct crystal complexes. In addition, we will examine the speed of the algorithm by monitoring convergence with respect to increasing sampling. Finally, we will examine the performance of the algorithm when processing a test database of compounds.

Methods

Anchor and grow (incremental construction)

The anchor-and-grow docking procedure is illustrated in Figure 1. In overview, a rigid portion of the ligand, the anchor, is docked using a geometrical matching procedure (step 1). The resulting anchor positions are used to initialize a pruned conformation search (steps 2 and 3). The conformation search is performed breadth-first on each anchor position simultaneously, with the most promising partially-built conformations retained during each stage of the search. When complete, each conformation is locally optimized (step 4) to relieve any strain incorporated during the construction process. If additional portions of the ligand are suitable as anchors, the process can be repeated (step 5).

Anchor selection

Prior to the first docking step, a preprocessing step is performed to preorganize the ligand atoms into rigid segments and to select the molecular anchor. The preprocessing step is illustrated in Figure 2. All preprocessing is performed at run-time and requires no modification of the ligand database.

The detection of rotatable bonds (Figure 2A) is performed in two phases. First, bonds contained in molecular rings are isolated. Ring flexibility is not considered directly in the algorithm presented here. For molecules in which ring flexibility plays an important role, different ring conformations can be generated by other modeling programs and treated as independent instances in the molecular database. In a second step, non-ring bonds are compared to a library of flexible bond definitions. Both the flexible bond definitions and the allowed torsion positions can be easily modified by the user. Each rotatable bond is assigned its allowed torsion positions which are also read in from an editable text file. Bonds can be excluded from rotation by flagging them in the molecule input file.

Based on the location of rotatable bonds, the molecule is then divided into rigid, overlapping segments (Figure 2B). The anchor segment is selected either by the user or automatically from the set of rigid segments. Automatic selection is based on segment size. If multiple anchors are desired, then all overlapping segments that meet a size cutoff are selected. Otherwise, the largest overlapping segment is selected. If key functional group interactions must be satisfied by the compound in order to bind, then the user may

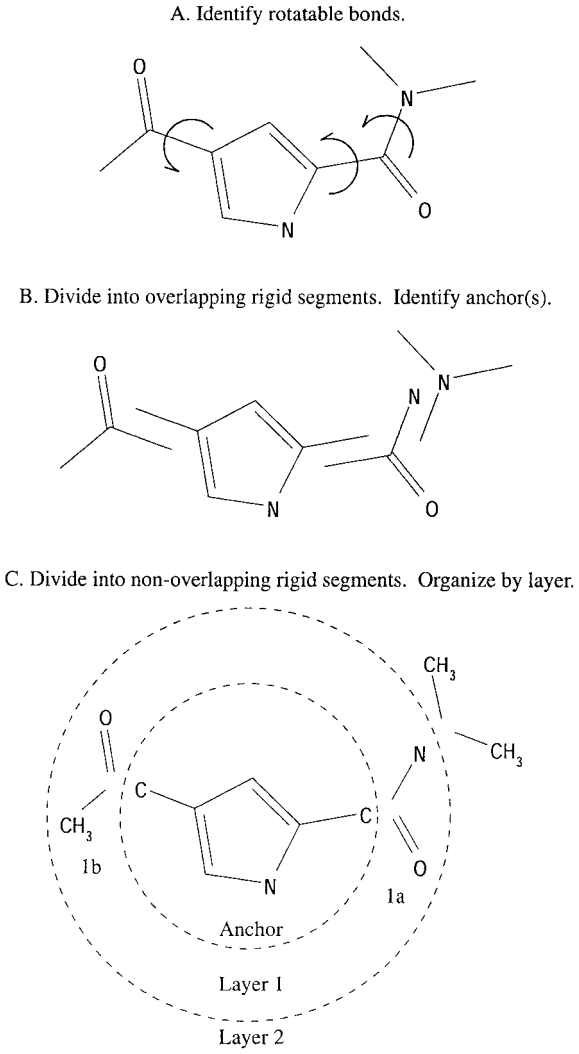


Figure 2. Atom Pre-organization and anchor selection.

override the automatic selection rules by defining the anchor in the molecule input file.

The last step in preprocessing is organizing the molecule atoms of the molecule into non-overlapping segments arranged concentrically around the anchor (Figure 2C). Each rotatable bond is assigned to a non-anchor segment. The atom bordering each rotatable bond is transferred to the inner segment, since it is unaffected by rotation of the bond. The net result is that each segment contains only the atoms directly affected by rotation about its associated rotatable bond. This process is repeated if multiple anchor segments (see below) are used.

Anchor docking

The orientation search of the anchor segment provides a rapid, preliminary search of binding modes (step 1 in Figure 1). In DOCK 4.0, the orientation search can be performed using a variety of protocols. For most applications, the best protocol is geometric matching, in which a subset of ligand atoms are superposed onto a subset of receptor site points providing each subset contains equivalent internal distances [10, 22, 25]. Additional random search protocols are available [26].

In DOCK 4.0, the matching-based orientation search can be run manually or automatically. When matching is manual, the user specifies geometric parameters like the distance tolerance, and the program builds all matches which fit the parameters. The matches are used to build orientations. Valid orientations are those which survive the bump filter (N_o in Figure 1). They are scored and minimized. With automated matching, the user specifies the number of valid orientations (N_o) directly, and the program will perform nested cycles of matching, until N_o non-bumping orientations have been formed.

The choice of whether to use automated matching depends on the docking problem. Automated matching is recommended for docking single ligands because it provides easy and intuitive control of orientation sampling. Manual matching is recommended for docking a database of ligands, because it biases the sampling towards molecules that contain more internal distance similarity to the site points. Manual matching is required when using other matching features like chemical matching [27] and critical site point matching [28].

Conformation search

As an overview, the conformation search (steps 3 and 4 in Figure 1) starts from the anchor positions generated in the anchor docking step. The anchor positions are pruned using an algorithm tuned to retain the best scoring and the most different positions. This pruning is repeated during each stage of the conformation search (see below). Each cycle of the conformation search involves adding a molecule segment to the current set of partial binding configurations and sampling the appropriate torsion positions of the intervening rotatable bond. This leads to an expanded population of binding configurations which is subsequently pruned based on score and positional diversity.

Conformation expansion

Segments are added to the current set of partial binding configurations in an orderly fashion, starting with the innermost layer and proceeding outward. Multiple copies are made, depending on the number of torsion positions searched for the intervening bond. Given N_c binding configurations in the current ensemble and N_t torsion positions for the current bond, then $N_c N_t$ new configurations will be built in a cycle. The default settings specify that two torsions are used for bonds between two sp^2 hybridized atoms, three torsions between two sp^3 atoms, and six torsions between an sp^2 and an sp^3 atom.

Conformation optimization

The partial binding configurations can be locally optimized at each step of construction in order to minimize the sum of ligand intramolecular energy and ligand-receptor intermolecular energy. If the newly added bond is flagged in the definition file as optimizable then the torsion angle for that bond is included in a short process of energy minimization. The default definition file specifies that bonds with double bond character are sampled at 0 and 180 deg, but are not allowed to move during energy minimization. Additional bonds may be included from inner segment layers to help prevent partial configurations from getting trapped in dead-end constructions. The rotation and translation of the anchor (and consequently the entire configuration) may also be adjusted to relieve strain. These minimization options are under user control. Although optimization is an expensive process, it significantly reduces the amount of sampling required, and actually improves the overall accuracy and speed. In general, optimization of the torsion angles is recommended, along with the anchor position and torsions from two inner layers.

Configuration pruning

After expansion and minimization, the set of partial binding configurations are pruned back down to an approximately constant size, N_c. Pruning is necessary to avoid the exponential growth of a systematic conformation search. Binding positions are pruned according to score and position so that both the best scoring and most different binding positions are retained from each cycle.

Pruning is performed top-down with a single pass algorithm. All binding positions are ranked according to score. The top-ranked position is set aside as the reference position and all others are considered candidates for removal. The weighted root-mean-squared distance (wRMSD) between the reference position, x^r, and each candidate, x^c, is computed according to Equation 1.

$$\text{wRMSD} = \left(\frac{\sum_{i=1}^{N_{\text{atoms}}} L_i[(x_i^c - x_i^r)^2 + (y_i^c - y_i^r)^2 + (z_i^c - z_i^r)^2]}{\sum_{i=1}^{N_{\text{atoms}}} L_i} \right)^{1/2} \quad (1)$$

where L_i is the layer to which atom i belongs. The RMSD is weighted in this fashion to make it more sensitive to the position of atoms in the outer layers of segments. The outer segments are more important because they have a greater influence over the position of subsequently added segments.

Each candidate is then evaluated for removal based on its rank and wRMSD using the inequality shown in Equation 2. If the factor is greater than the value of the configurations per cycle input parameter, the candidate is removed. This factor favors low ranking configurations as well as spatially different configurations. It is important to maintain a heterogeneous population of configurations to reduce the chance of getting trapped in a local minima during the conformation search. The next best scoring configuration which survives the first pass of removal is then set aside and used as a reference configuration for the second round of pruning, and so on. The time dependence of this algorithm is less than quadratic with respect to the number of configurations in the set. However, the time cost of this step is small and generally insignificant compared to the time cost of computing and optimizing the score of each configuration.

$$\left(\frac{\text{Rank of candidate}}{\text{wRMSD}} \right) > \text{configurations per cycle.} \quad (2)$$

This pruning method attempts to balance the twin goals of recovering the best scoring and the most different binding configurations without introducing additional user parameters. In practise, the number of binding positions that are retained is up to a factor of two greater than *configurations per cycle*, and depends on the size of the ligand and binding site.

Final optimization

After all cycles of segment addition are complete, the pruned set of complete binding configurations is locally optimized once again (step 4 in Figure 1) to relieve strain developed during the construction process. The rotation, translation, and all appropriate torsions are optimized using the energy minimization protocol discussed below.

Time demand

The scoring and optimization steps of the calculation tend to consume the majority of the cpu cycles, making the time demand of the algorithm sensitive to the number of times that score optimization is applied during construction. For the typical run, in which score optimization is applied to the positions of the initial anchor fragment as well as the positions and torsions of the partially-built configurations, the time demand of the algorithm can be approximated by Equation 3.

$$\text{Time} = \text{const.} \cdot \mathrm{N_0} + \text{const.} \cdot \mathrm{N_c} \cdot \mathrm{N_b} \cdot \mathrm{N_t} \qquad (3)$$

With score optimization, the cost of the anchor docking step tends to depend linearly on the number of positions searched, N_o. Likewise, the cost of the conformation search tends to depend linearly on the average number of pruned configurations, N_c, the number of rotatable bonds, N_b, and the average number of torsion positions per rotatable bond, N_t.

Random search

An alternative search technique was also developed, in which multiple random conformations are docked independently. This technique is much like the the 'flexibase' approach [13], in which the molecule database is seeded with multiple conformations of each molecule, and each conformation is docked independently as a rigid molecule. DOCK can reproduce this technique as a default if the flexible ligand option is requested, but Anchor-First docking is not requested. Multiple random conformations of each molecule are constructed and docked sequentially. Sampling is controlled by specifying the number of conformations per molecule and the number of orientations per conformation. The conformation search is actually controlled in a novel way. In order to allow a more thorough search of more flexible molecules, the number of conformations per rotatable bond is actually selected. This way, a molecule with 4 rotatable bonds will be represented with 4N conformations and a molecule with 8 rotatable bonds with 8N conformations, and so on. The complexity of the random search then grows linearly with the number of rotatable bonds, just like the anchor and grow technique. The technique can also be performed without conformation search, so that each molecule is docked in its static conformation, but with full torsion and position relaxation using simplex optimization.

Scoring

With the anchor and grow strategy, the scoring function is used to guide intermediate stages of the search. An increased burden is placed on the accuracy of the scoring function, since entire modes of binding could be missed because of mistakes in the calculated interactions of a portion of the complex. In this work, the existing molecular mechanical scoring function is retained, but the importance of continued development of more accurate scoring functions cannot be emphasized enough. The anchor and grow strategy constrains the type of scoring function to be atom pairwise decomposable since the interactions of a partially-built molecule are evaluated.

Intermolecular terms

The intermolecular interaction of the small molecule with the macromolecule is described with the non-bonded terms of the AMBER molecular mechanic potential [29] precomputed on a three-dimensional grid [23]. In the current version of DOCK, the user can select the following options: whether a united atom or all-atom model is used, the exponents of the attractive and repulsive Lennard-Jones terms, and a function representing the dielectric behavior [29]. During the construction of flexible molecules, the interactions of the partially built portions are approximated by leaving the VDW terms and partial charges unperturbed. Since functional groups are often split into different segments, this approximation may cause short-term inaccuracies, especially when dipole interactions like hydrogen bonding are important. Some further study in how to treat these interactions is warranted.

Intramolecular terms

Because of the conformational freedom given to flexible molecules during construction, some treatment of intramolecular energy is necessary. The chief concern is to prevent internal clashes. To this end, the non-bonded Lennard–Jones and Coulombic terms are computed for interactions between atoms in different

rigid segments. Atoms within a rigid segment are excluded, since their contribution is a constant. Since only torsional degrees of freedom are explored, no bond length or angle terms are included. As a first level of approximation, bonds with a low barrier to rotation are allowed to freely rotate without the evaluation of an explicit torsion potential. All other bonds are prevented from free rotation. Speed is enhanced by maintaining a record of the interactions of each segment with all internal segments to avoid recalculation when no relative movement has occurred.

Optimization

The thoroughness of sampling is enhanced with on-the-fly score optimization [24]. The optimization technique used in DOCK is the simplex method [30, 31]. During anchor docking, rigid body minimization is applied to each anchor orientation. Six simplex variables are used for rotation and translation. The quaternion is used to represent the rotation angles because of enhanced numerical stability [32]. Additional simplex variables are used to represent rotatable bonds. When torsions are included from multiple layers, movement of the inner torsions causes greater perturbation than movement of outer torsions. To compensate for this distortion, the step size of an inner torsion is scaled by the number of segment layers it is from the current layer.

An important part of minimization is the criteria for termination. The progress of minimization is monitored by checking the difference in score of the best scoring and worst scoring simplex vertex. When the variation is within some tolerance, minimization terminates. Simplex minimization can be vulnerable to premature convergence. As a consequence, additional cycles of minimization are launched if the current minimization has moved a specified distance in simplex space. The vector distance of the simplex variables is used. The vector is normalized with respect to the step sizes of each simplex variable.

An important consequence of using a simplex minimizer is that the results of each run are dependent on how the random number generator was seeded. In this work, multiple runs are performed in which different random number seeds are used. Results can be averaged over the set of runs. Standard deviations can be computed to assess stability and errors.

Evaluation techniques

Rebuilding X-ray complexes

The flexible docking algorithm is tested in several ways. A set of high and medium resolution X-ray crystal complexes is used to assess the accuracy and speed of the algorithm. The complexes included in the study are presented in Table 1. Since the binding position and conformation of the ligand molecule is known, we can test how well the docking algorithm reconstructs the binding mode. This is a simpler problem than ab initio docking, since the receptor conformation is not searched. These testcases have been used to verify other flexible docking methods [14, 17]. The portions of the ligands selected as the anchor segment are consistent with the selections used in the validation of FlexX. The ligands in the set span a range of sizes and flexibility, possessing from 2 to 13 rotatable bonds.

All testcases are prepared using a uniform protocol. The ligand and macromolecules are processed in SYBYL to add hydrogens and load partial charges (SYBYL Molecular Modeling System, Version 6.3, Tripos Associates, Inc., St. Louis, MO). As described in Meng et al. [23], AMBER united atom charges are used for the macromolecule and Gasteiger-Marsili charges are used for the ligand. Crystallographic waters are removed, except for a few testcases where key active site waters are included. Table 2 lists the testcases for which special considerations of active site modeling are made. Histidine residues are modeled as fully protonated (HIP), except active site histidines listed in Table 2 for which the protonation states can be inferred from hydrogen bonding patterns. If metal ions or cofactors are present, they are included in the receptor model as fixed elements of the binding site with standard Van der Waals parameters and Gasteiger-Marsili partial charges incorporating all formal charges if present. The importance of including the waters, ion and cofactors in the active site is assessed for several testcases by repeating the docking with and without the additional elements.

A molecular surface of the binding site is prepared using MS from the MIDAS package [44], by including only protein residues and modeled waters, ions and cofactors within 8 Å of the ligand position. The DOCK utility, SPHGEN, is used to compute site points [10]. The first cluster which overlapped the ligand is used. SPHGEN can be used as a *de novo* tool to identify binding sites based on shape alone, but the quality of such a prediction is not assessed here. A scoring grid is then constructed to enclose all the site points plus an

Table 1. X-ray crystal complexes used to evaluate docking algorithm.

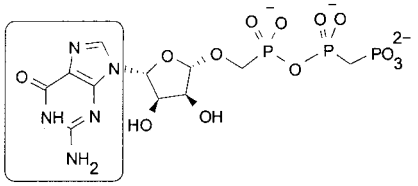
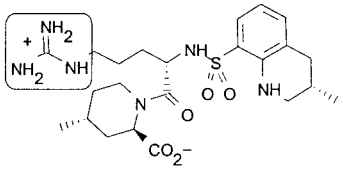
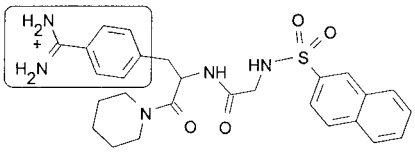
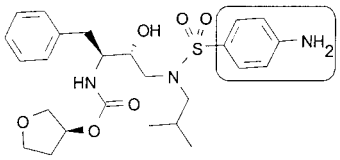
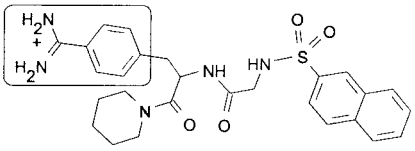
Name	Receptor	Ligand (Anchor segment highlighted by box)	Resolution, Å	Refinement
121p[33]	H-Ras P21	 Guanosine-5'-[B,G-Methylene] Triphosphate	1.54	0.195
1dwc[34]	Thrombin	 Argatroban	3.0	0.150
1dwd[34]	Thrombin	 NAPAP	3.0	0.156
1hpv[35]	HIV-1 Protease	 VX-478	1.9	0.192
1ppc[36,37]	Trypsin	 NAPAP	1.8	0.181

Table 1. Continued.

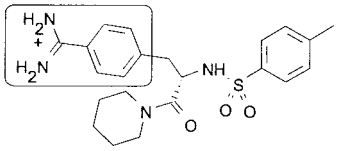
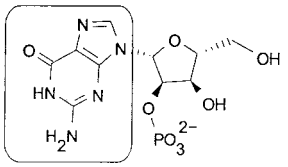
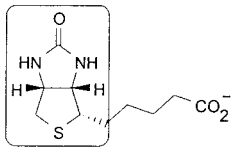
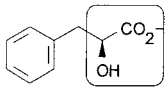
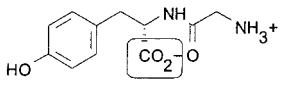
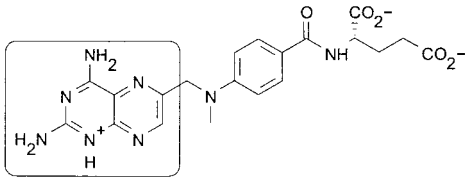
Name	Receptor	Ligand (Anchor segment highlighted by box)	Resolution, Å	Refinement
1pph[36,37]	Trypsin	 3-TAPAP	1.9	0.167
1rnt[38]	Ribonuclease T1 Isozyme	 2'-Guanylic Acid	1.9	0.191
1stp[39]	Streptavidin	 Biotin	2.6	0.22
2ctc[41]	Carboxypeptidase A	 L-Phenyl Lactate	1.4	0.161
3cpa[42]	Carboxypeptidase A	 Glycyl-L-Tyrosine	2.0	not refined
3dfr[43]	Dihydrofolate Reductase	 Methotrexate	1.7	0.152

Table 2. Specific modeling considerations of each test case

Testcase	Protonation state of active site histidines[a]	Metal Ions	Waters/cofactors
121p		Mg^{2+}	
1dwd	His72 = HID		
1hpv			WAT201 interacts with ILE 50A and ILE 50B
1ppc and 1pph	His57 = HID		
1rnt	His40 = HIP His92 = HIE		
1tmn	His142 = HID His146 = HID His231 = HIP	Zn^{2+}	
2ctc	His69 = HIE His196 = HIE	Zn^{2+}	WAT29 interacts with zinc
3cpa	His69 = HIE His196 = HIE	Zn^{2+}	
3dfr			NADPH with Gasteiger-Marsili charges

[a]HID = delta nitrogen protonated; HIE = epsilon nitrogen protonated; HIP = both imidazole nitrogens protonated leading to a full net charge on residue.

extra 8 Å margin along each axis. These preparatory calculations took less than an hour for each testcase.

Docking calculations are then carried out for each testcase. Default values are selected for all input parameters except the main sampling parameters, which are varied to explore the influence of sampling. Since only a single ligand is docked at a time, automated matching is used. Consequently, the chief input parameter controlling the anchor orientation search is the number of anchor positions to explore. The main parameter controlling the conformation search is the number of configurations saved at each cycle of pruning. These parameters are varied over range of two orders of magnitude using the six settings listed in Table 3. For each level of sampling, five runs are performed using different random number seeds.

Table 3. Sampling conditions used to rebuild X-ray complex testcases.

Runs[a]	Anchor positions	Configurations per cycle
1a–1e	10	1
2a–2e	30	3
3a–3e	100	10
4a–4e	300	30
5a–5e	600	60
6a–6e	1000	100

[a]Five duplicate runs were performed at each sampling level, varying only the number used to seed the random number generater.

Database screening

An important application of DOCK is the screening of a molecule database. For a flexible docking algorithm to be appropriate to screen a large database, the time spent processing each molecule must be on the order of tens of seconds or less. Of course, the amount of sampling is under direct user control, so sampling can be made arbitrarily low in order to achieve sufficient speed. The critical issue is how much sampling is required to get an accurate ranking of the database. The quantitative measurement of screening accuracy and its convergence is discussed below.

In this work, we use streptavidin (1stp) and dihydrofolate reductase (3dfr) as the test sites to which

a set of 49 randomly selected molecules from the Current Medicinal Chemicals (CMC) molecular database (MDL Information Systems, San Leandro, CA) are screened. The test database is seeded with biotin and methotrexate so that at least one tight binding molecule is included.

In order to put the performance of our flexible docking algorithm in context, we will use two reasonable alternative algorithms as controls. The first control algorithm is a random search technique, in which multiple random conformations are docked independently. DOCK can accomodate this technique if the flexible ligands option is selected, but anchor-first docking is not selected. The second control algorithm is rigid docking, using the single molecule conformation present in the database. The conformations are generated by CONCORD [45]. DOCK can accomodate this technique if the flexible ligands option is not selected. Running DOCK in this fashion is much like using DOCK3.0 or earlier, except that the new matching algorithm is used [25].

To calculate the screening accuracy at a given level of sampling, the molecule rankings must be compared with an ideal or best possible ranking. Since we want to isolate the performance of the sampling algorithm from any errors in scoring, we cannot base the target rankings on actual binding data. Instead, the target ranking will be based on the best possible score for each molecule at a very high level of sampling. In practice, the best score for each molecule is determined by taking the best from all runs. In order to compare the scores among different molecules, only the score of the ligand:receptor interaction, or intermolecular term, will be used. This is based on the assumption that the intramolecular terms for the docked configuration after minimization are equivalent to the same terms in the uncomplexed ensemble of configurations. In addition, a 0.5 kcal/mol penalty is applied per heavy atom to offset the tendency of the current scoring function to favor larger molecules. Implementation of explicit desolvation terms into the scoring function should also address this problem. Given the rankings, the screening accuracy is computed according to the weighted rank correlation (WRC) expressed in Equation 4. This term is based on the Spearman Rank Coefficient, but weighted to be more sensitive to differences among high ranking members.

$$\mathrm{WRC} = \frac{\sum_{i=1}^{N} \frac{1}{x_i}(x_i - \overline{x})\,(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} \frac{1}{x_i}(x_i - \overline{x})^2 \sum_{i=1}^{N} \frac{1}{y_i}(y_i - \overline{y})^2}} = \frac{\sum_{i=1}^{N} \frac{1}{x_i}(x_i - \overline{x})\,(y_i - \overline{y})}{\sum_{i=1}^{N} \frac{1}{x_i}(x_i - \overline{x})^2} \tag{4}$$

In this equation, x_i is the rank of molecule i based on its best score over all runs, and y_i is the rank of molecule i based on its score from a particular run. The $1/x_i$ term biases the contribution of the top-ranked members. This bias is consistent with the notion that database screening is used primarily to identify the top scoring molecules, and any inaccuracies of molecules lower on the list are of diminishing importance. WRC varies from negative one for perfect anti-correlation to positive one for perfect correlation.

Results

Rebuilding X-ray complexes

Docking accuracy

The accuracy of the docking algorithm was evaluated with 12 crystallographic testcases. The results of these dock runs are presented in Table 4. The target interaction score for each testcase was determined by performing energy minimization directly on the crystallographic complex, allowing the ligand position and conformation to be adjusted according to the simplex method. Five independent minimization runs were performed for each complex. The best minimized scores are presented in Table 4A, along with a RMS distance from the original crystallographic position. The values of energy score vary from −33 kcal/mol (3cpa + zinc) to −-132 kcal/mol (121 p). The large, negative score of 121p results from the presence of four negative formal charges on the triphosphate ligand. The interaction energy of biotin with streptavidin (1stp), −34 kcal/mol, is less favorable by comparison to the other complexes, which is inconsistent with measured affinity actually being greater. The interaction energy is inconsistent with the actual binding free energy because many complexation and entropy terms are ignored, including desolvation terms. The

deficiencies of using the intermolecular molecular mechanics potential are well known [13, 23, 47, 48]. The purpose of this work is not to predict binding constants, but to show that even with a minimal scoring function, the docking search algorithm succeeds in finding reasonable binding modes. The simplex optimizer found locally minimum positions between 0.3 and 1.1 Ångstroms from the original crystallographic position.

A summary of the best scoring dock positions is presented in Table 4B. The RMSD of the top scoring position is an important measure of accuracy. The relative positions of the top scorers with respect to the crystallographic positions vary from 0.9 Å (3dfr) to 6.8 Å (1dwd) away. In 7 of the 15 testcases, the best scoring docked positions are less than 2 Å away. Of the 8 testcases docking farther than 2 Å, 6 had a score more favorable than the corresponding minimization-only score, which indicates that innaccuracies in the scoring function are causing the deviations. With the significant role of related potential functions in the refinement of macromolecular structures, it might be worthwhile exploring inconsistencies in scoring paradigms. Of the remaining 2 testcases (1hpv and 1hpv + wat), the best scoring positions are greater than 2 Ångstroms away, with scores less favorable than the minimization-only score, indicating that for these testcases, the search method is insufficient within the settings tried here. The largest deviations in position occurred with the trypsin and thrombin inhibitors (1dwc, 1dwd, 1ppc and 1pph) in which the guanidinyl and benzamidinyl groups were docked correctly, but the remainder of the ligands adopt non-crystallographic positions.

Including crystallographic waters, ions and cofactors had some effect on the top scoring positions but was not a critical determinant. The inclusion of the crystallographic water in the 1hpv structure had only a small effect on the best scoring position (1 kcal/mol and 0.1 Å). The presence of the zinc ion in the 3cpa structure caused the docking algorithm to select a position 4.5 Å from the crystallographic position, compared to 1.9 Å without the zinc ion, and prefering it by 5 kcal/mol. The apparent cause of the movement is that in the crystallographic mode, the carboxylate of the dipeptide ligand interacts with an arginine side chain, however, when the zinc ion is present, the scoring function favors placing the carboxylate to interact with the zinc ion. The presence of the NADH cofactor in the 3dfr structure caused a 7 kcal/mol change in the best score, but the position was the same distance from the crystallographic position.

Each dock run produces a number of binding positions beyond the top scoring position which approach even closer to the crystallographic position. The RMSD of the closest approaching position is another important measure of docking accuracy. These positions are listed in Table 4C. The RMSD data reveal that in all systems the ligand is positioned no closer than 0.5 Å from the crystallographic position. In 13 of 15 cases, a position is found within 1.5 Å. In all cases, a position is found within 2.0 Å. These closest positions are often ranked near to the top scoring position. In 13 of 15 cases, the nearest position is ranked in the top one-third of the positions generated for the given run.

Docking speed

The results of the evaluation of docking speed are presented in Table 5 and Table 6. All timings are based on a 150 Mhz R10000 Octane processor from Silicon Graphics. We have found this cpu to be up to three-fold faster at docking calculations than the older 200 Mhz R4400 processor in the Indigo2 from Silicon Graphics. The fastest successful dock run for each testcase is listed in Table 5, where success is defined as at least one docked position within 2 Ångstroms of the X-ray position. The fastest timings ranged from 1 second for 2ctc up to 600 s for 1dwc. The influence of including cofactors in the active sites had interesting effects on docking, speeding up 1hpv and 3dfr, but slowing 3cpa. A stricter standard of success is to assess sampling convergence, or when all the runs for a given amount of sampling attain the 2 Å cutoff. These data are presented in Table 6. The testcase for which sampling converged the fastest was again 2ctc which required 2 s. The most difficult testcases did not converge within the standard range of sampling conditions (100 configurations per cycle). Testcases 1dwc and 1hpv require greater than 1000 s for converged docking. The inclusion of cofactors made convergence more difficult for 1hpv and 3cpa, and had no effect on 3dfr.

Case study

The docking results from one of the moderately difficult testcases, 1ppc, highlight the issues typically encountered with docking analysis. The Trypsin:NAPAP complex contains a flexible ligand with 9 rotatable bonds. Figure 3A shows the intermediate progress dur-

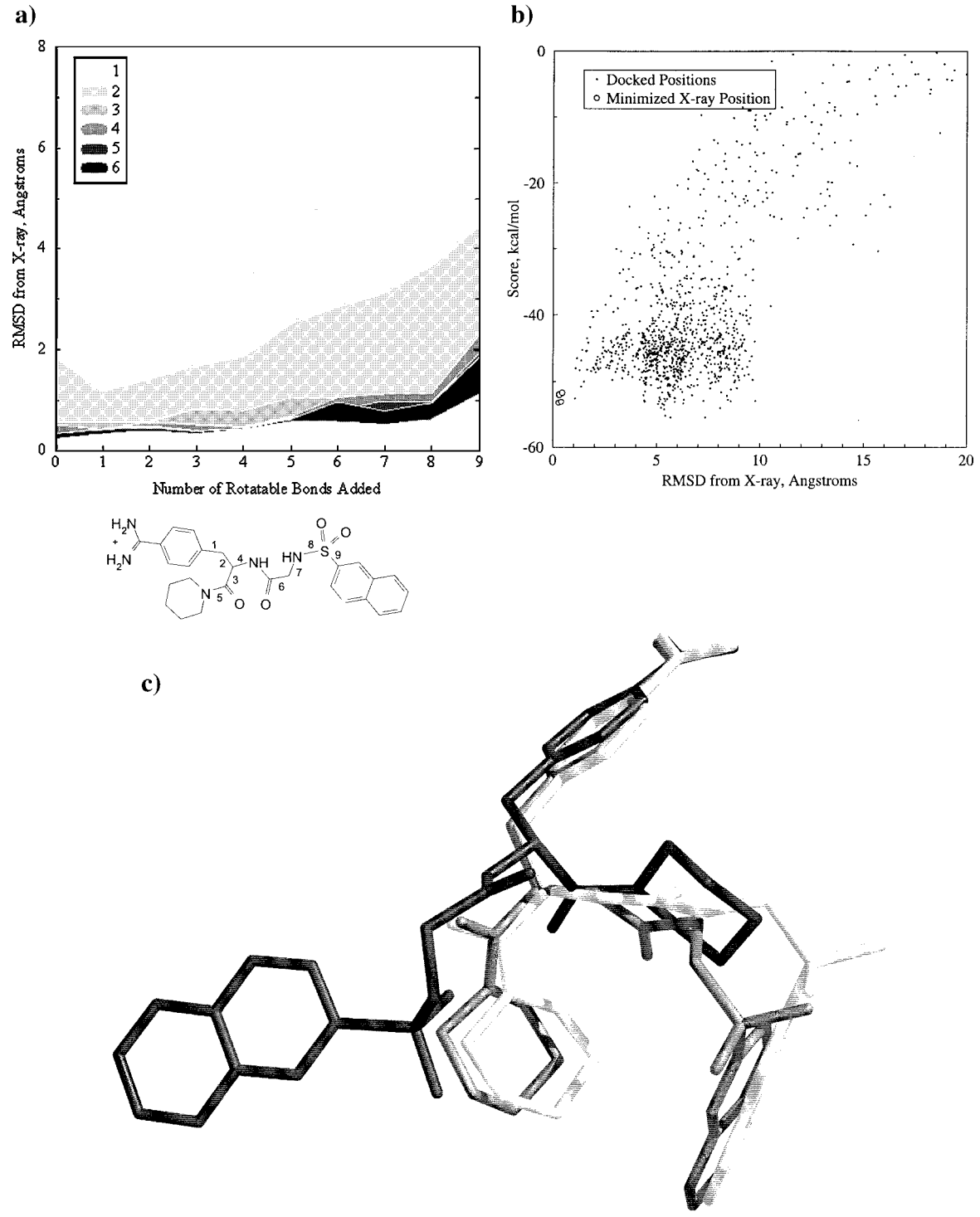


Figure 3. Analysis of trypsin:NAPAP testcase (1ppc). A. Construction Progress. During each cycle of the conformation search, the population of pruned conformations is surveyed and the member closest to the X-ray position is selected. The RMS distance of the closest member from the X-ray position is plotted. The left-most data points (zero torsions) correspond to the anchor positions. The right-most data points (9 torsions) correspond to completely-built, docked conformations. The progress of five independent, high sampling (1000 anchor orientations, 100 conformations per cycle) runs is plotted. B. Scatter Plot of Docked versus Minimized Positions. The score of each docked position is plotted against its RMS Distance from the X-ray position. For comparison, points describing the minimization of the X-ray position are shown in open circles. C. Overlay of Docked Positions. The best-scoring docked position (black) and the nearest docked position (gray) are superimposed on the X-ray position (white). Figure was generated with UCSF MidasPlus [44].

Table 4. Accuracy of rebuilding X-ray complexes.

Testcase	A. Minimize only		B. Dock – best score			C. Dock – best position		
	Score[a]	RMSD[b]	Run[c]	Score	RMSD	Run	Score	RMSD
121p	−132	0.4	6e	−130	1.9	6c	−108	1.2
1dwc	−51	0.4	5e	−57	5.3 d	6b	−47	1.9
1dwd	−49	0.8	6a	−55	6.8 d	4a	−48	1.1
1hpv	−59	0.6	6e	−57	4.2 e	6c	−37	1.4
1hpv + wat	−59	0.3	6c	−58	4.1 e	3c	−50	1.6
1ppc	−53	0.3	4c	−56	5.3 [d]	3a	−48	1.0
1pph	−47	0.3	5e	−51	2.9 [d]	4b	−50	0.8
1rnt	−49	0.8	6c	−54	1.4	5d	−50	1.0
1stp	−34	0.8	6a	−38	1.0	5a	−34	0.6
1tmn	−66	0.3	6a	−67	2.8 d	6c	−55	0.5
2ctc	−39	0.7	6d	−41	1.3	6d	−39	0.6
3cpa	−42	0.8	6d	−45	1.9	5c	−41	0.8
3cpa + zinc	−33	1.1	6d	−50	4.5 d	6a	−42	1.3
3dfr	−70	0.6	6b	−74	0.9	6b	−72	0.7
3dfr + nad	−73	0.4	6c	−81	0.9	4d	−79	0.7

[a]Units of kcal/mol.
[b]RMS distance from X-ray position in Ångstroms.
[c]Identifer of dock run. Sampling condition listed in Table 3.
[d]Inconsistent docking because of innaccurate scoring function.
[e]Inconsistent docking because of insufficient search.

Table 5. Sampling conditions for fastest run to find a position within 2 Å RMSD from X-ray structure

Testcase	Run	Number of configurations per cycle	CPU Time (R10000 seconds)
121p	3e	10	70
1dwc	5a	60	600
1dwd	3e	10	80
1hpv	4c	30	300
1hpv + wat	3c	10	100
1ppc	2a	3	20
1pph	3d	10	50
1rnt	2c	3	10
1stp	2a	3	5
1tmn	3c	10	90
2ctc	1c	1	1
3cpa	1b	1	3
3cpa + zinc	3a	10	30
3dfr	2b	3	20
3dfr + nad	1a	1	10

Table 6. Sampling conditions for fastest runs find a position within 2 Å RMSD from X-ray position

Testcase	Set of Runs	Number of Configurations per Cycle	Average CPU Time (R10000 Seconds)
121p	5	60	300
1dwc	6[a]	100[a]	1000[a]
1dwd	6	100	500
1hpv	6[b]	100[b]	1000[b]
1hpv + wat	6[c]	100[c]	1000[c]
1ppc	5	60	300
1pph	6	100	300
1rnt	4	30	90
1stp	2	3	5
1tmn	5	60	600
2ctc	2	3	2
3cpa	2	3	9
3cpa + zinc	4	30	70
3dfr	4	30	200
3dfr + nad	4	30	200

[a]1 out of 5 runs meet 2 Ångstrom cutoff.
[b]4 out of 5 runs meet 2 Ångstrom cutoff.
[c]2 out of 5 runs meet 2 Ångstrom cutoff.

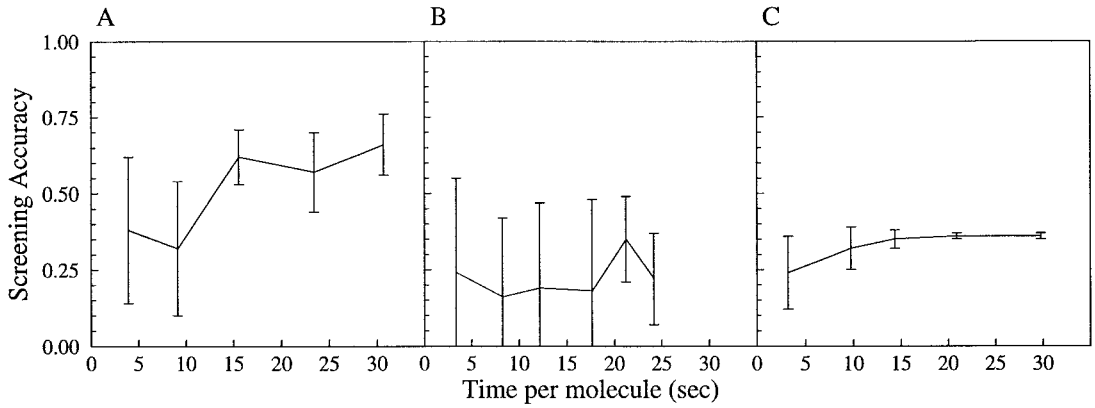


Figure 4. Performance of screening against 1stp. A. Incremental construction. B. Dock multiple ligand conformations independently. C. Dock CONCORD conformation. The rank accuracy (Equation 4) is computed as the weighted correlation of the single-run rankings compared to the overall rankings. The weightings place greater emphasis on the relative positions of the top-scoring compounds.

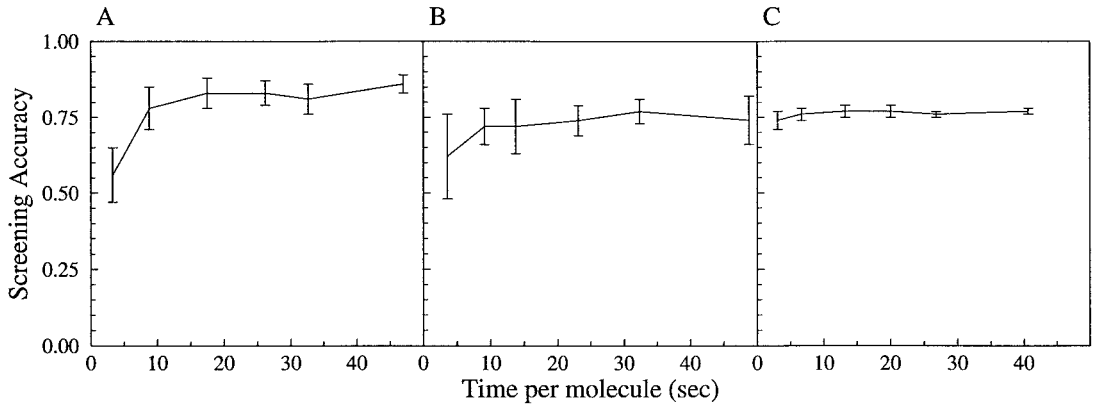


Figure 5. Performance of screening against 3dfr. A. Incremental construction. B. Dock multiple ligand conformations independently. C. Dock CONCORD conformation.

ing the docking calculation over a range of sampling conditions. Each shaded area corresponds to the upper and lower RMS distance of the nearest position from the X-ray position over the set of 5 runs for each level of sampling. At the two lowest levels of sampling, the docking calculation fails because the anchor benzamidine segment is placed incorrectly. At higher levels of sampling, the anchor segment and all outer segments are placed correctly up through the last segment, which shows a step-wise increase in the total RMS error from 1 to 2 Å. The scatter plot of docked positions at high sampling, Figure 3B, shows a narrow cluster of dockings 1 Ångstrom from the X-ray position with the same interaction scores as the minimized X-ray scores. There is a broad, highly populated cluster of positions from 3 Å to 10 Å from the X-ray position with interaction scores more favorable than the X-ray position. This phenomenon occurs on occasion, indicating an error in the scoring function which causes the docking calculation to favor an alternative binding mode, while simultaneously showing the robustness of the search algorithm which maintains sufficient breadth to be able to reproduce the less favored, but more accurate crystallographic position. The key docked positions are depicted in Figure 3C, showing that the 1 Å position is nearly exact except for the placement of the sulfonamide and the flipping of the naphthyl system, which are placed last in the construction process. The top scoring position positions the benzamidine segment correctly, but flips the naphthyl sulfonamide portion with the piperidinyl group, although managing to overlap much of the same volume of the crystallographic position.

Database screening

The convergence properties of screening a database of molecules using different docking techniques are studied. The first testcase studied is streptavidin (1stp). The 51 molecules in the dataset (including biotin) were docked and ranked with increasing levels of sampling. The results are shown in Figure 4. At low levels of sampling (5 to 10 s per molecule), all three methods are equivalent in accuracy, although the single conformation method produces more consistent results. At higher sampling (15 to 30 s per molecule), the incremental construction method shows clear superiority (50 to 75% accuracy) compared to the independent conformation method (0 to 50% accuracy) and the single conformation method (30 to 35% accuracy).

The convergence properties of database screening were also evaluated for the dihydrofolate reductase complex (3dfr) and are shown in Figure 5. The 51 molecule dataset included the dhfr inhibitor methotrexate. At low levels of sampling (3 to 10 s per molecule) the single conformation method (72 to 77% accuracy) is marginally better than the incremental construction method (45 to 85% accuracy) and the multiple independent conformations method (45 to 75% accuracy). But at higher sampling (15 to 50 s per molecule), the incremental construction method (75 to 90% accuracy) outperforms the other two which fail to improve.

Since the docking algorithms appeared to have more difficulty ranking the test dataset of molecules for binding to streptavidin compared to dihydrofolate reductase, the testcases were studied further. The distribution of scores for the dataset are shown in Figure 6. The interaction score of biotin with streptavidin is among the best in the set. However, the interaction score of methotrexate with dhfr is significantly different from the other molecules in the set, causing the apparent accuracy in screening to be inflated as a result. In fact, even when methotrexate is docked rigidly using the non-crystallographic CONCORD conformation, its score is within the top 5 scores of the set of molecules docked flexibly.

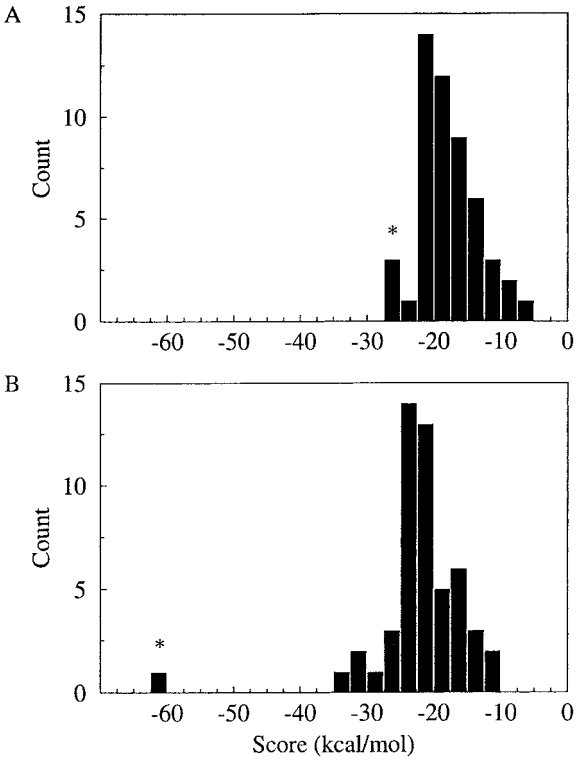


Figure 6. Distribution of database scores for 1stp and 3dfr. Asterisk marks where natural ligand is located in distribution.

Discussion

Incremental construction is a best-first strategy which traverses a small portion of the possible search space. Given that the crystallographic position of a molecule is at a global minimum of the binding site energy landscape, the incremental strategy works best when the partially-built fragments of the molecule also occupy energy minima (either local or global). An exhaustive search of binding positions must search the 6 translational and rotational degrees of freedom, plus all conformational degrees of freedom. The incremental construction strategy must search the same 6 degrees of rigid body freedom for the anchor segment, but uses a linearly-constrained conformation search. For a typical molecule with 8 rotatable bonds and 3 low-energy torsion positions per bond, the exhaustive conformational search requires 3^8, or 6561, evaluations. Using a generous list length of 50, the incremental construction strategy would require $3 \times 8 \times 25$, or 600, evaluations, which is a speed-up by a factor of 10. For molecules in which the partially built fragments lie near energy minima, the list length can be reduced further. For more intractable problems, the incremental list can be lengthened. Alternatively, a systematic conformation search can be coupled with rigid docking for each conformation. Both techniques are available with DOCK 4.0.

Considering the simplicity of the scoring function that is currently implemented, the docking method is

Table 7. Comparison of the accuracy of different docking methods[a]

Testcase	DOCK	GOLD[5,6,46]	FlexX[17]	Wang, et. al.[20]	Makino et. al.[19]
121p	1.2/1.9		1.1/2.0	1.0/1.0	
1dwc	1.9/5.3		1.2/2.7	1.2/1.2	
1dwd	1.1/6.8	1.7/1.7	0.6/2.1	1.4/2.0	
1hpv	1.4/4.2				
1hpv + wat	1.6/4.1				
1ppc	1.0/5.3				
1pph	0.8/2.9				n.a./1.9
1rnt	1.0/1.4		1.0/1.5		n.a./0.9
1stp	0.6/1.0	0.6/0.7	0.8/0.8	1.0/1.0	n.a./1.0
1tmn	0.5/2.8	1.5/1.7	0.9/0.9	1.7/1.9	
2ctc	0.6/1.3	0.2/0.3	0.6/0.6		
3cpa	0.8/1.9	0.9/1.5	1.1/3.1	0.6/0.6	
3cpa + zinc	1.3/4.5				
3dfr	0.7/0.9	0.8/1.4[b]	0.9/1.3[b]	1.4/1.4[b]	n.a./1.2[b]
3dfr + nad	0.7/0.9				

[a]Paired values for each method are reported, representing the RMS distances of the nearest docked position and the best-scoring docked position from the X-ray crystallographic position in Ångstroms. The testcases listed here for the alternative methods represent a subset of the actual testcases reported.
[b]Actual target 4dfr.

remarkably accurate at reconstructing the testcases. The current scoring function neglects desolvation effects, which can dominate the prediction of the interaction energy. Consequently, the magnitude of the docking scores deviate from known binding energies by an order of magnitude. However, the placement of the ligand is less sensitive because of approximately equivalent desolvation terms for the same ligand binding to different areas of the same binding site. It is an important validation of the search algorithm that it can perform well despite this deficiency. Further refinement of the scoring function will likely result in increasing docking accuracy.

The accuracy of the dock results can be compared directly with other methods for which the same testcases have been studied. Table 7 summarizes the RMS distances of the docked positions for the method presented here and for four published methods, including a genetic algorithm method (GOLD[5,6]) and three incremental construction methods. The nearest docked position generated by our method is equivalent to the results of the other methods. The best scoring position is occasionally less accurate than what is reported by the other methods (1dwc, 1dwd and 1tmn). One source for this inconsistency may be that some of the alternative methods use empirical scoring functions which may be better tuned for – perhaps even trained on – those particular testcases. A similar result would occur if the other methods were not allowed to explore far enough away from the X-ray position to uncover the alternative binding positions seen here for 1dwc and 1dwd.

For a set of 15 complexes, the method is able to generate a docked position within 2 Å of the crystallographic position for all the cases. Because of the small set size, it is difficult to estimate the probability that the method will be successful for an arbitrary test case. Other workers have attempted to characterize such statistics by considering much larger test sets. The GOLD method[5,6,46] reports a docked position within 2.0 Å for 80% of 99 testcases, while FlexX[17] reports 69% of 200 testcases.

The speed of reconstructing the crystallographic complexes is competitive with existing methods. The fastest runs for each testcase were as short as a few seconds, with most being less than 100 s with the R10000 SGI cpu (about 300 R4000 seconds). These timings are very competitive with previously reported algorithms which for similar testcases required up to 400 R3000 cpu seconds [14], 180 R4000 cpu seconds [17], 300 R4000 cpu seconds [19], and 2000 s [6]. The robustness of the search algorithm is confirmed by monitoring sampling convergence. With approximately 5-fold more sampling than the fastest run, all

runs consistently succeed at meeting the 2 Å cutoff. We encourage other researchers in this field to consider assessing the stability of their algorithms in a similar way.

The analysis of three of the testcases demonstrates the interplay of orientation and conformation sampling in determining the success of the dock run. For smaller ligands like biotin in 1stp, we observe a simple threshold of orientation sampling that needs to be exceeded before the docking is successful. For larger, more flexible ligands like NAPAP in 1ppc or the protease inhibitor in 1hpv, large increases in conformational sampling lead to incremental improvements in docking. Despite problems with the scoring function, the search algorithm can often recover binding modes near the crystallographic position. Further refinement of the scoring function should further improve sampling by helping guide the search during intermediate stages of construction.

The efficiency of the search algorithm is most critical when processing a large number of molecules. The implications of our work with processing a test database are that if very little time is allowed for calculation ($<$ 10 s per molecule), both single conformer and flexible docking are competitive. In fact, both methods should be used in parallel, and the results combined. DOCK version 4.0 supports all three modes of docking studied here. If more cpu time is available ($>$ 15 s per molecule), then the incremental construction method is superior, and should either be performed alone, or with a larger allocation of computing resources.

Conclusion

The search algorithm presented here represents a considerable advance in technology, by extending the dock algorithm to the treatment of flexible molecules. The method is suprisingly accurate considering the simplicity of the scoring function currently implemented. In addition, the software gives the user access to straightforward sampling parameters, making it well-suited for wide distribution within the modeling community. Because the algorithm emphasizes speed at only a modest cost to accuracy, it can be used to screen a large database of molecules. Future work in this area should focus on continued development of more accurate scoring functions. The search algorithm should also be extended to consider a limited search of receptor site flexibility within the time constraints of a database processing.

The algorithm is incorporated into the DOCK suite of molecular modeling programs available from UCSF. Please see the website for licensing information (http://www.cmpharm.ucsf.edu/kuntz/dock.html). All experiments reported here were performed with version 4.0.1 of DOCK.

Acknowledgements

The authors would like to gratefully acknowledge Yaxiong Sun, Diana Roe, Daniel Gschwend, and Andrew Good for their thoughtful discussions and the Computer Graphics Laboratory at UCSF for the use of their facilities and the MIDAS visualization software. Tripos Associates provided the SYBYL modeling software and MDL provided the ISIS/Base software and the CMC Database, for which we are grateful. Financial support was provided by NIH Grant GM-31497. T.J.A.E was sponsored by an NSF Graduate Research Fellowship and a Glaxo Research Fellowship.

References

1. Morris, G. M., Goodsell, D. S., Huey, R., & Olson A. J., J. Comput. Aid. Mol. Des., 10 (1996) 293.
2. Judson, R.S., Jaeger, E.P. and Treasurywala, A.M., Theochem-J. Mol. Struct., 114 (1994) 191.
3. Oshiro, C.M., Kuntz, I.D. and Dixon, J.S., J. Comput. Aid. Mol. Des., 9 (1995) 113.
4. Clark, K.P. and Ajay, J. Comput. Chem., 16 (1995) 1210. Jones, G., Willett, P. and Glen, R.C., J. Mol. Biol., 254 (1995) 43.
5. Jones, G., Willett, P., Glen, R. C., Leach, A. R. L. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
6. Gehlhaar, D.K., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J., and Freer, S.T., Chem. Biol., 2 (1995) 317.
7. Leach, A. R., J. Mol. Biol., 235 (1994) 345.
8. Westhead, D.R., Clark, D.E., and Murray, C.W., J. Comput. Aid. Mol. Des., 11 (1997) 209.
9. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
10. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
11. Knegtel, R.M.A., Kuntz, I.D. and Oshiro, C.M., J. Mol. Biol., 266 (1997) 424.
12. Miller, M. D., Kearsley, S. K., Underwood, D. J. and Sheridan, R. P., J. Comput.- Aided Mol. Des., 8 (1994) 153.
13. Welch, W., Ruppert, J. and Jain, A. N., Chem. Biol. 3 (1996) 449.
14. Leach, A.R. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 730.

15. Mizutani, M. Y., Tomioka, N. and Itai, A., J. Mol. Biol., 243 (1994) 310.
16. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
17. Rarey, M., Kramer, B. and Lengauer, T., Bioinformatics, 15 (1999) 243.
18. Makino, S., & Kuntz, I.D., J. Comput. Chem., 18 (1997), 1812.
19. Wang, J. and Kollman, P. A., in publication.
20. Rarey, M., Kramer, B. and Lengauer, T., J. Comput. Aid. Mol. Des., 11 (1997) 369.
21. Shoichet, B.K. and Kuntz, I.D., Protein Eng., 6 (1993) 723.
22. Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem. 13 (1992) 505.
23. Gschwend, D. A. and Kuntz, I. D., J. Comput. Aid. Mol. Des., 10 (1996) 123.
24. Ewing, T.J.A. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1175.
25. Ewing, T.J.A., ed., Dock 4.0 User Manual, University of California, San Francisco, 1998, http://www.cmpharm.ucsf.edu/kuntz/dock.html.
26. Shoichet, B.K. and Kuntz, I.D., Protein Eng., 6 (1993) 723.
27. DesJarlais, R.L. and Dixon, J.S., J. Comput. Aid. Mol. Des., 8 (1994) 231.
28. Pearlman, D. A., Case, D.A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P., Comput. Phys. Commun., 91 (1995), 1.
29. Nelder, J. A. and Mead, R., Comput. J., 77 (1965) 308.
30. Press, W. H. and Vetterling, W. T., Numerical Recipes: the Art of Scientific Computing, University Press, New York, NY, 1989.
31. Allen, M. P. and Tildesley, D. J., Computer Simulation of Liquids, Clarendon Press, New York, NY, 1987.
32. Krengel, U. (1991). PhD. Thesis. Heidelberg University.
33. Banner, D. W. and Hadvary, P., J. Biol. Chem., 266 (1991) 20085.
34. Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D. and Navia, M. A., J. Am. Chem. Soc., 117 (1995) 1181.
35. Bode, W., Turk, D. and Stuerzebecher, J., Eur. J. Biochem., 193 (1990) 175.
36. Turk, D., Stuerzebecher, J. and Bode, W., FEBS Lett., 287 (1991) 133.
37. Arni, R., Heinemann, U., Maslowska, M., Tokuoka, R. and Saenger, W., Acta Crystallographer, B43 (1987) 549.
38. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., and Salemme, F.R., Science, 243 (1989) 85.
39. Monzingo, A. F. and Matthews, B. W., Biochemistry, 23 (1984) 5724.
40. Teplyakov, A., Wilson, K. S., Orioli, P. and Mangani, S., Acta Crystallographica, D49 (1993) 534.
41. Christianson, D. W. and Lipscomb, W. N., Proc. Nat. Acad. Sci. USA, 83 (1986) 7568.
42. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
43. Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R., J. Mol. Graphics, 6 (1988) 13.
44. Rusinko, A., Sheridan, R. P., Nilakantan, R., Haraki, K. S., Bauman, N. and Venkataraghavan, R., J. Chem. Inf. Comput. Sci., 29 (1989) 251.
45. World Wide Web posting of GOLD evaluation results at http://panizzi.shef.ac.uk/cisrg/ gareth/gold/gold.html
46. Bohm, H.J., J. Comput. Aid. Mol. Des., 8 (1994) 243.
47. Jain, A.N., J. Comput. Aid. Mol. Des., 10 (1996) 427.