Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening

TODD J. A. EWING, IRWIN D. KUNTZ

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143-0446

ABSTRACT: The DOCK program explores possible orientations of a molecule within a macromolecular active site by superimposing atoms onto precomputed site points. Here we compare a number of different search methods, including an exhaustive matching algorithm based on a single docking graph. We evaluate the performance of each method by screening a small database of molecules to a variety of macromolecular targets. By varying the amount of sampling, we can monitor the time convergence of scores and rankings. We not only show that the site point−directed search is tenfold faster than a random search, but that the single graph matching algorithm boosts the speed of database screening up to 60-fold. The new algorithm, in fact, outperforms the bipartite graph matching algorithm currently used in DOCK. The results indicate that a critical issue for rapid database screening is the extent to which a search method biases run time toward the highest-ranking molecules. The single docking graph matching algorithm will be incorporated into DOCK version 4.0.   © 1997 by John Wiley & Sons, Inc.   *J Comput Chem* **18**: 1175−1189, 1997

Introduction

With the advent of high resolution X-ray crystallography and NMR, structural chemists and biologists can study biomacromolec-

ular interactions in atomic detail. This information, combined with computational and visualization tools, has helped spawn the field of structure-based ligand design. A common step in the design cycle is the process of molecular docking, in which possible binding geometries of a molecule with a macromolecule are studied.

The docking process can be divided into two parts: a search algorithm and a scoring algorithm. The search algorithm should sample the degrees of freedom of the ligand:macromolecule system suffi-

ciently to include the true binding mode(s). The scoring algorithm should represent the thermodynamics of interaction sufficiently to distinguish the true binding mode(s) from all others explored.

Because of the computationally expensive nature of the search problem, many different solutions have been proposed. Docking with molecular dynamics and Monte Carlo algorithms has been explored,[1,2] including simulated annealing[3-5] and MCSS[6] methods. Other docking protocols consider molecular flexibility, including rotamer search,[7-9] distance geometry,[10] and genetic algorithm[11-13] methods. To make the search tractable for processing a large set of molecules, the molecular components are often treated as rigid objects. With this approximation, researchers have used systematic searching,[14] pattern recognition,[15-17] graph theoretical,[18-21] and other superposition[22] techniques to dock molecules.

The UCSF DOCK program belongs to the group of methods employing the rigid body assumption and uses graph theoretical techniques. Because of its speed, the program is often used to screen a large database of molecules, selecting potential ligands of a receptor target.[23] In this article, the term ''ligand'' is used loosely; it refers to any small molecule whose binding is under study. The term ''receptor'' refers to the macromolecule whose binding pocket is being explored.

The core of the DOCK search algorithm is the superimposition of ligand atoms onto predefined site points[24,25] that map out the negative image of the binding site. A matching process is used to determine which ligand atoms and site points are to be superimposed.[26] Multiple orientations are generated this way, with each receiving a score assessing the intermolecular interactions. This score is based on the intermolecular terms of a molecular mechanics force field.[27] Recently, an optimization procedure has been added that adjusts each orientation to improve the intermolecular interactions.[28,29]

In this work, we critically evaluate several matching algorithms for the docking process, including an exhaustive matching algorithm. The exhaustive algorithm was presented by Bron and Kerbosch as a method to detect cliques in an undirected graph.[30] It was later incorporated into a docking algorithm by Crippen and coworkers.[10,19] It has many attractive features, so we chose to evaluate it in the context of rigid molecular docking, score optimization, and database screening. As an exhaustive search, it avoids some of the artifacts encountered by the current matching

method. For example, with a nonexhaustive algorithm, adjusting parameters to increase the total amount of sampling can reduce the amount of sampling of certain binding modes.[28] Increased sampling, with the new algorithm, will always retain binding orientations found with less sampling, leading to a proper superset of binding modes. An exhaustive algorithm also does not require the additional parameters controlling the heuristics of the nonexhaustive search.[26]

Although the matching algorithms formally treat the ligand and receptor as rigid objects, they can readily be incorporated into a flexible docking scheme.[10,21,31-33] In future work, we will investigate how best to divide up a flexible docking problem into smaller rigid parts.

To evaluate the performance of the new matching algorithm, we propose a new assessment protocol based on screening a small database of molecules. Because we specifically intend to minimize any artifacts due to the quality of the scoring function in this work, we will not use experimental measurements as the standard, but instead, the global minimum of our current scoring function. We will also evaluate random, and partially random, search algorithms as controls with which to put the current DOCK performance in perspective. These control algorithms let us investigate fundamental issues of orientational sampling, such as the effect of using site points to guide the search.

Methods

BIPARTITE DOCKING GRAPH

Since the first release of DOCK, the search process has been driven by a matching procedure in which subsets of ligand atoms and receptor site points are identified that have equivalent internal distances.[18] Matching is formulated as a graph theoretical problem in which the ligand atoms and receptor site points are separate sets of nodes in a bipartite graph.[26] A match is defined as a set of compatible edges which connect a subset of ligand nodes with an equal number of receptor nodes. For the edges to be compatible, the distances among ligand nodes must map to equivalent distances among receptor nodes. An example of match formation is depicted in Figure 1. As this figure illustrates, to extend a match, all possible edges (including bad edges) must be considered; distance comparisons are used to identify and discard bad edges. Matches are extended until there is a

sufficient number of nodes in the match to define a unique orientation of the ligand.

Since version 2.0, DOCK avoided considering some bad edges with a pruning method involving distance binning.[26] Nodes were preorganized in distance bins, such that, for each seed node, sets of nodes in discrete distance intervals from the seed were identified. These bins guide match extension from a seed edge (connecting seed nodes), ensuring that candidate edges are compatible with the seed edge. They do not, however, ensure compatibility with other nonseed edges already included in the match. For example, the binning algorithm would avoid considering the bad edge in step II of Figure 1, but not the bad edge in step III. The storage requirements for this algorithm grow as $N_n N_b N_{n/b} \leq N_n^3$, where N_n is the number of nodes (ligand atoms or receptor site points). N_b is the number of distance bins and grows with the longest distance and the inverse of the bin width. $N_{n/b}$ is the number of nodes in each distance bin which grows with N_n and the bin width.

SINGLE DOCKING GRAPH

Kuhl et al. proposed merging the bipartite docking graphs into a single docking graph,[19] which is then amenable to clique detection techniques developed by Bron and Kerbosch.[30] In a single docking graph, each node represents a pairing of an atom with a site point. Each edge identifies adjacent nodes, or two nodes for which both atom components and site point components are separated by equivalent distances. The docking graph is represented by an adjacency matrix in which each nonzero element identifies adjacent nodes. An adjacency matrix for the example depicted in Figure 1 is presented in Figure 2. The chief advantage with this representation is that all necessary distance comparisons are made during the construction of the adjacency matrix. Consequently, during matching, the adjacency matrix is used as a rapid filter to ensure that no bad edges are ever considered. This type of matching is presented in Figure 3. Although the single docking graph is one step removed from the intuitive appeal of the bipartite docking graph, it enables a more efficient solution to the docking problem.

The single docking graph representation has also been implemented in the FLOG docking program.[21] This program heavily prunes the matching search tree using a minimum-residual search heuristic. Although it examines all possible nodes at each branch point, it only pursues the node with
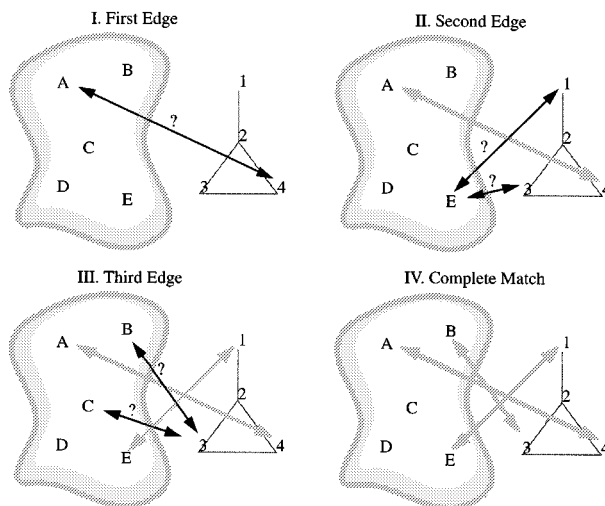

FIGURE 1. Bipartite graph matching algorithm. The receptor site points (A–E) and ligand atoms (1–4) are separate sets of nodes in a bipartite graph.[26] I. A first (seed) edge is considered. Of the 20 seed edges to be tried (5 points ∗ 4 atoms), we first consider A4 for this example, in three dimensions, such a match would superimpose atom 4 onto site point A. This match would fix three of six orientational degrees of freedom. II. Second edges are considered. Of the 12 edges to be tried (4 points left ∗ 3 atoms left), we consider E3. Because AE > 43, we discard E3 as a second edge. Then we consider E1. Because AE = 41, we retain E1 as a second edge. In three dimensions, this match would superimpose atoms 4 and 1 onto points A and E, respectively. This match would fix two more orientational degrees of freedom. III. Third edges are considered. Of the six edges to be tried (3 points left ∗ 2 atoms left), we consider C3. Though AC = 43, EC < 13 so we must discard C3 as a third edge. Then we consider B3. Because AB = 43 and EB = 13, we retain B3 as a third edge. This match fixes the last of six orientational degrees of freedom. IV. The match is large enough to define a unique orientation which superimposes atoms 4, 1, and 3 onto site points A, E, and B, respectively.

the smallest difference in ligand and receptor distances with respect to the most recently added node in the match. Backtracking is only allowed at the seed level, where all possible nodes are pursued to initiate matching. As a result, the total number of matches can never exceed the number of nodes.

Here we propose implementing the single docking graph representation combined with a variation of the exhaustive clique detection method discussed by Bron and Kerbosch.[30] A clique is defined as a set of fully adjacent nodes (i.e., a completely connected subgraph) which cannot be

FIGURE 2. Adjacency matrix for single docking graph algorithm. This matrix identifies all adjacent nodes for the example given in Figure 1. Each node is defined as a site point–ligand atom pair, for instance, A4. For two nodes to be adjacent, the intra-atom distance must be equal to the intra–site point distance. For example, matrix element (A4, E1) is turned on because $\overline{AE} = \overline{41}$. The matrix is symmetric.

further enlarged without adding a nonadjacent node. Much attention is given to the intractable nature of the maximum clique problem. It is classified as NP-complete because the solution time grows faster than any polynomial expression of the problem size.[19] For the application of molecular docking, however, we are not trying to find the single, largest clique. The process of matching is in fact a process of finding completely connected subgraphs within an undirected graph: a less restrictive, and therefore more tractable problem than finding cliques and maximum cliques. Although Bron and Kerbosch actually present two methods and recommend a bounding technique for clique detection, we find their original, brute-force method sufficient for finding fully connected subgraphs in a manner efficient for molecular docking.

MATCHING PARAMETERS

In our molecular docking implementation, we use two parameters to determine node adjacency: a distance tolerance and a distance minimum. The distance tolerance parameter addresses experimental uncertainty in the ligand and target structures.
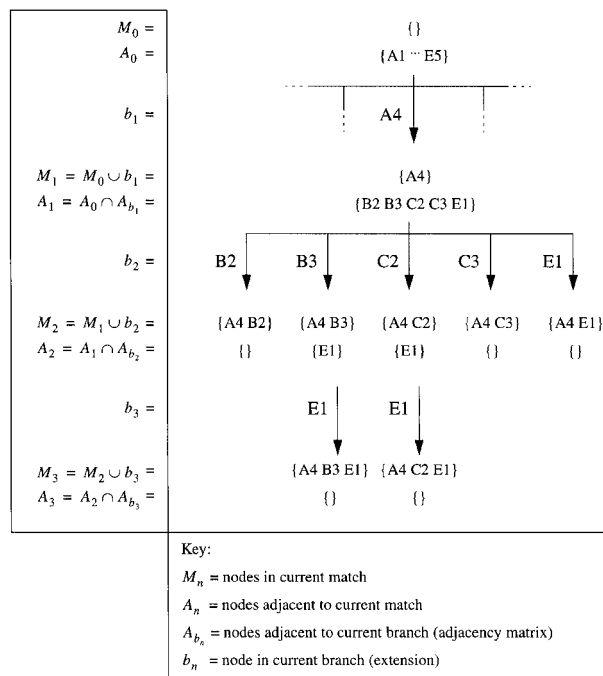

FIGURE 3. Single docking graph matching algorithm. This figure depicts the entire search path-starting from the same seed used in Figure 1. Matching begins with the match set, M_0, empty and the adjacency set, A_0, maximally filled. A match is extended by finding the union of the previous match, M_{n-1}, and a branch node, b_n, which is selected from A_{n-1}. The new adjacency set, A_n, is the intersection of A_{n-1} with the set of all nodes adjacent to b_n, which is A_{b_n} and is taken from the adjacency matrix presented in Figure 2. Match extension continues until A_n is empty. Matches with three or more nodes define a unique ligand orientation (see text and Fig. 4). The match set, {A4 B3 E1}, corresponds to the solution presented in Figure 1.

The distance minimum parameter enables the search to focus on the longer, more relevant internal distances. If adjacency information is stored in a matrix, then, for most docking situations, the matrix can be very large. The matrix size grows as $(N_{lig} N_{rec})^2$, where N_{lig} is the number of nonhydrogen ligand atoms and N_{rec} is the number of receptor site points. Because these matrices are sparse (ca. 1% elements typically occupied), we store only the nonzero elements of each row of the matrix as an integer list of nodes. The probability, p_{on}, of an element being nonzero is a function of the distance tolerance and distance minimum. The memory requirement for the adjacency lists is $p_{on}(N_{lig} N_{rec})^2$. We presort each adjacency list so that the process of finding the common elements of two lists (the $A_n = A_{n-1} \cap A_{b_n}$ steps in Fig. 3)

can be performed on a once-through basis at a speed comparable to the use of the complete matrix.

The advantage of an exhaustive algorithm is that, when sampling is increased, the search is guaranteed to include the search space explored at the lower sampling level. This property helps to avoid sampling artifacts encountered with the bipartite matching algorithm.[28] Despite its exhaustive nature, it does not undergo a combinatorial explosion for larger systems because of user control over the sampling parameters. Typical sampling parameters for a docking scenario having fewer than 30 ligand nonhydrogen atoms and fewer than 50 target site points are: four nodes minimum for a match, and 0.5-Å distance tolerance and 2.0-Å distance minimum for node adjacency. Because the search never explores invalid branches, search time grows as a function of the number of distance-constrained solutions rather than the number of possible unconstrained solutions. Therefore, docking larger molecules into larger sites can be made nearly as rapid if a smaller distance tolerance or larger distance minimum is chosen. Because memory reserved for the adjacency lists is dynamically allocated, the memory burden is adjusted as well.

MINIMUM MATCH SIZE

Some confusion exists in the literature over how many atoms and site points must be in a match to define a unique orientation. The orientation is generated by a ligand transformation which has a translation and a rotation component. The translation vector has three degrees of freedom. The rotation matrix has four degrees of freedom. Three of them are represented by the Euler angles. The fourth is represented by the sign of the determinant. A rotation with a positive determinant retains the handedness of the object it transforms, while one with a negative determinant will reverse the handedness of the object. If one knows in advance whether to reverse the handedness of the object, then only the three Euler angles need to be determined. For example, when docking a chiral ligand in which only one stereoisomer is relevant (e.g., protein or peptide ligands), only the positive-determinant rotation matrix would be of interest. When docking a ligand available as a racemate, then both transformations would be of interest. The FLOG program,[21] for instance, routinely samples both mirror images a ligand, even when the ligand is achiral. When the sign of the determinant is known in advance, the six degrees of freedom of
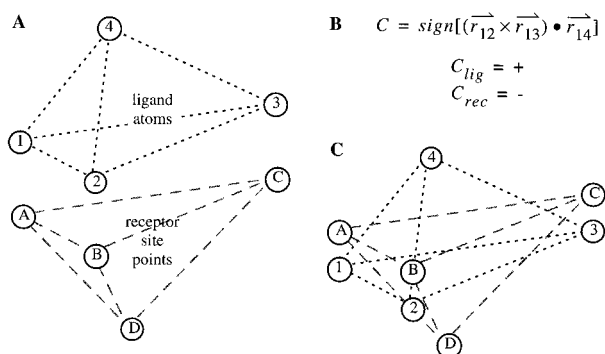

FIGURE 4. Relative chirality of match points. (**A**) Distance matching might identify a set of atoms {1 2 3 4} to match with a set of site points {A B C D}, such that A is with 1, B is with 2, and so on. Although the internal distances within the atoms are equivalent to those within the site points, the handedness is opposite. (**B**) We define the relative chirality, C, according to the sign of the triple product. For any given four-point match, the probability that the relative chiralities are the same is 50%. (**C**) If the chirally opposed sets are superimposed without inverting the chirality of the ligand set, then the resulting least squares fit is poor.

the rotation and translation are uniquely determined by a match set containing three nonlinear atoms and site points.

Processing a larger match set causes only one transformation to be allowed. As illustrated in Figure 4, when a one-to-one mapping has been made between two sets of four nonplanar points, then each set can be assigned a relative chirality. This is true even if the points come from an achiral molecule or from a set of site points where chirality is ambiguous. If the relative chiralities are the same, then the ligand can be oriented normally. If the relative chiralities are opposite, then either the ligand is inverted when oriented, or if that is not desired, then the match is discarded. In fact, all larger matches that are supersets of the discarded match are also discarded, because they too yield inconsistent matches. If these steps are not taken, then the resulting orientations will poorly superimpose the ligand atoms and receptor site points in the match set, even though all the distance tolerances are met (Fig. 4C).

ADDITIONAL SEARCH CONSTRAINTS

The systematic design of the matching algorithm makes it well suited to incorporate specialized search constraints. Some examples, although not assessed in this study, are mentioned because

they have been shown to be useful elsewhere. To avoid oversampling particular binding modes, orientational degeneracy checking has been studied.[17,29] In the new matching algorithm, a degenerate orientation is detected as a degenerate match whose nodes are a subset of nodes in a larger match. In other words, only subgraphs that are true cliques need be processed. As another example, chemical information can be used to guide the matching process using labeled atoms and site points. Only nodes composed of a chemically compatible atom and site point are used to seed or extend a match. Much like the repellent node implementation of Kuhl et al.,[19] matches adjacent to chemically incompatible nodes are discarded. In addition, sampling can be focused on particular regions of the active site by defining critical site point clusters. This technique is similar to the approach used in targeted DOCK[34] and FLOG,[21] except that clusters can be of arbitrary size and number. The matching process automatically restricts itself to make sure all matches include members from each cluster. The new matching

algorithm lends itself so well to these constraints that when activated, they contribute a negligible computational overhead, and can lead to considerable speed improvements for database searches.[21,34]

CONTROL METHODS

We will test a total of five methods (Table I) to isolate specific aspects of the search process. These methods range in complexity from a completely random search to the bipartite and single graph procedures described previously. We begin with the uniform random transformation (URT) method, which explores a predefined rectangular volume enclosing the active site. It is the most simple and ''hypothesis-free'' of the methods tested here. URT will indicate the minimum level of performance that we expect from any docking algorithm. The uniform random matching (URM) method explores the irregularly shaped volume described by the collection of site points. It will show the performance gains, if any, of using a ''negative image''

TABLE I.
Search Methods.

Abbrev.	Method	Description	Hypothesis Tested
URT	Uniform random transformation	• Construct rectangular volume enclosing site • Randomly move molecule center of mass within volume • Randomly rotate • Each molecule in database sampled uniformly	• Random method used as reference • Rectangular enclosure is sufficient Site point description is unnecessary
URM	Uniform random matching	• Match random subsets of atoms with random subsets of site points • Superimpose match atoms onto match site points with a least squares fit • Each molecule in database sampled uniformly	• An irregular volume to describe the site is more efficient
BRM	Biased random matching	• Match randomly (like URM) • SGM controls amount of sampling for each molecule in database	• An irregular volume is more efficient • Spending more time on molecules which match better is more efficient
SGM	Single graph matching	• Using single graph, exhaustively match subsets of atoms with subsets site points with equivalent internal distances (DOCK 4.0)	• Fitting some atoms precisely onto of some site points is more efficient
BGM	Bipartite graph matching	• Using bipartite graph, nonexhausitively match using binning algorithm (DOCK 3.5)	• Fitting some atoms precisely onto some site points is more efficient • Binning algorithm more efficient

approach to map out the binding site. The biased random matching (BRM) method is identical to URM, except it uses the new matching algorithm to determine the number of random matches to try for each molecule. Once the number of matches has been determined, BRM uses completely random selections of nodes to form the actual matches used to generate molecule orientations. Because BRM is a hybrid approach, it is meant to help isolate the source of any differences between URM and the new matching algorithm. The single graph matching (SGM) algorithm uses the new matching algorithm to determine both the number of matches and the actual orientations to try for each molecule. SGM will reveal the performance gains, if any, of using site points to not only map out the most interesting binding site volume, but to also direct the positioning of individual ligand atoms. The bipartite graph matching (BGM) method is the existing DOCK 3.5 matching algorithm. It will reveal the advantage, if any, of using a nonexhaustive search method with a longest distance first heuristic. BGM is described last because, within the spectrum of different search methods, its algorithm is the most elaborate.

TEST SYSTEM

We assess the performance of the search methods in the following way. We dock a set of 100 molecules, chosen randomly from the set of uncharged, medium-sized, and generally rigid molecules in the available chemicals database (ACD).[35] In our study, a medium-sized molecule is one with 15 to 35 nonhydrogen atoms. A generally rigid molecule is one with no single bonds except those attaching hydrogen atoms, attaching terminal nonhydrogens (i.e., methyl or hydroxyl groups), or participating in ring structures. Molecules meeting these three criteria compose 40% of the ACD. For each molecule in the test set, a single CONCORD[36]-generated conformation is used.

We see several advantages to using such a data set of molecules to test the search methods. First, the docking conditions represent a close approximation to the typical application of DOCK to database screening. Not only can we study the convergence of score for each molecule, we can study the convergence of relative scores, or rankings, of the set of molecules. Second, the docking conditions allow us to explore a multitude of diverse molecular shapes so that our results are less subject to potential artifacts of a particular ligand: receptor system. Although some may argue that

studying a set of known, potent ligands would be more relevant, we counter that the databases DOCK searches often do not contain potent binders, and that DOCK frequently finds micromolar inhibitors to serve as lead compounds.[23] By choosing a random subset of molecules, we, in fact, will arrive at a set that best represents the typical array of molecules tested. By biasing the subset to include medium-sized, generally rigid molecules, we also focus on that portion of the database which is best treated by rigid molecular docking.

For each search method, we perform multiple docking runs, and vary the amount of sampling from zero to a value at which the docking results converge. The key sampling parameters for each method are listed in Table II. Scoring and opti-

TABLE II.
Sampling Parameters.

URT	
Total orientations	0–900,000[a]
URM	
Total orientations	0–100,000[a]
Nodes min/max	3/3[b]
BRM	
Distance tolerance	0–0.9 Å[a]
Distance minimum	2.0 Å[c]
Nodes min/max	3/3[b]
SGM	
Distance tolerance	0–0.6 Å[a]
Distance minimum	2.0 Å[c]
Nodes min/max	4/10[d]
BGM	
Distance tolerance	1.5 Å[e]
Ligand bin width	0.1–0.9 Å[a,f]
Receptor bin width	0.1–0.9 Å[a,f]
Ligand bin overlap	0.1–0.5 Å[a,f]
Receptor bin overlap	0.1–0.5 Å[a,f]
Nodes min/max	4/4[e]

[a] For some test systems, the upper limit was not reached if docking results converged early.
[b] Set to three, because, as the size of a random match increases, the least-squares superposition procedure increasingly biases the orientation toward the centroid of the site points.
[c] Set large enough to exclude atoms sharing a covalent bond.
[d] Minimum of four chosen so that chirality could be used to filter matches. Maximum of ten is somewhat inconsistent with value chosen for BGM, but we presume any effects of this would be small.
[e] Chosen as historical default.
[f] Minimum value not zero because of a numerical instability of the algorithm.

mization parameters are listed in Table III. The range of values in these tables correspond to timings of less than 0.1 second/molecule to more than 100 seconds/molecule on a modern workstation (see subsequent text).

For each docking run, several properties are computed that compare the results from any specific run to the best results from all runs combined (assumed to contain the global minimum). These properties are summarized in Table IV. When these values are plotted versus time, the convergence of each property can be monitored. We assume that a better search algorithm will lead to more rapid convergence.

When considering the average behavior of each property, we compute both the usual mean and also a rank-weighted mean. The rank-weighted mean is more sensitive to the behavior of the top scoring molecules, which are of most interest in database screening runs. Although many kinds of weighting functions could be chosen for this pur-

TABLE III.
Scoring and Optimization Parameters.

Type	Force field
Bump maximum	3[a]
Dielectric	4r
Grid spacing	0.3 Å
Interpolation	Trilinear
Convergence criteria	0.1 kcal/mol[b]
Maximum iterations	500[c]

[a] Nonzero maximum allows some orientations with limited Van der Waals clashes with the receptor to be recovered by the minimizer.
[b] A relatively tight convergence criteria was selected to reduce noise in the score evaluation, so that differences between methods were more directly attributable to differences in sampling. The rank correlation would be especially vulnerable to such noise.
[c] A large iteration limit was also selected to reduce noise in the score evaluation by preventing the minimizer from terminating prematurely.

TABLE IV.
Comparison Methods.

Method	Definition	Equation	Rank weighted equation	Range
Average relative score	• For each molecule in the docking run, normalize its score by the best score it ever received in the site • Then compute the average over the molecules in the set	$$\dfrac{\sum\limits_{i=1}^{N} S_i/S_i^{min}}{N}$$	$$\dfrac{\sum\limits_{i=1}^{N} \dfrac{S_i/S_i^{min}}{i}}{\sum\limits_{i=1}^{N} \dfrac{1}{i}}$$	[0, 1]; unitless
Rank Correlation	• Assign a rank, y_i, to each molecule based on its best score in the docking run • Then correlate y_i with the rank of each molecule based on the best score it ever received in the site, x_i	$$\dfrac{\sum\limits_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{N} (x_i - \bar{x})^2}$$	$$\dfrac{\sum\limits_{i=1}^{N} \dfrac{1}{i}(x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^{N} \dfrac{1}{i}(x_i - \bar{x})^2}$$	[−1, 1]; unitless
Average RMS error	• For each molecule in the docking run, compute the RMS error of its predicted orientation compared to that which received the best score for that site • Then, compute the average over the molecules in the set	$$\dfrac{\sum\limits_{i=1}^{N} R_i}{N}$$	$$\dfrac{\sum\limits_{i=1}^{N} \dfrac{R_i}{i}}{\sum\limits_{i=1}^{N} \dfrac{1}{i}}$$	[0, ∞]; in angstroms

TABLE V.
Receptor Structures from PDB[37,38] Used for Test Systems.

Code	Structure	Resolution	R factor	Site description
121D	DNA dodecamer with Netrospin[39]	2.2 Å	0.198	Site is broad, presenting two continuous binding sites in the major and minor grooves of the DNA dodecamer; highly polar
1ULB	Purine nucleoside phosphorylase with guanine[40]	2.75 Å	0.204	Site has two pockets; one is broad and centrally located, another where actual ligand binds is peripheral and solvent excluded
3DFR	Dihydrofolate reductase with NADPH and methotrexate[41]	1.7 Å	0.152	Site has a deep, centrally located binding pocket; mixed polar and nonpolar regions
4FAB	Fab fragment with fluorescein[42]	2.7 Å	0.215	Site is shallow with three pockets formed by the six hypervariable loops; generally nonpolar
9HVP	HIV-1 protease with A-74704[43]	2.8 Å	0.182	Site is a long, narrow tube, which completely penetrates protein; mixed polar and nonpolar regions

pose, we chose to use the reciprocal of the rank for convenience.

To make sure that our conclusions are generalizable, we analyzed the methods using five different receptor sites listed in Table V. These sites were chosen from the list of complexes of high resolution, well-refined structure with a ligand having a well-defined binding position. They were also chosen based on very different-shaped binding sites. The chief features of each site are shown in Table V.

Results and Discussion

DOCKING CONDITIONS

The test cases were prepared for docking in the standard way. Site points were constructed using the sphere generation accessory program of DOCK with default parameters.[18] We selected the cluster of site points which occupied the binding site of the actual ligand in the crystal complex. Within this cluster, we merged the positions of tightly grouped site points using a 2-Å cutoff. The final number of site points used for each receptor ranged from 30 to 60.

RESOURCE USAGE

All docking calculations were performed on Silicon Graphics Indigo2 workstations, equipped with 200 MHz R4400 processors and 128 MB of RAM, so timings are consistent among the different methods. Several weeks of computer time were required to complete all runs. All methods required approximately 13 MB of RAM to store the scoring grids. The URT and URM methods required negligible additional memory for matching and orienting. BRM and SGM required up to 0.1 MB of RAM for matching arrays. BGM required 1 MB of RAM for matching arrays.

TEST SYSTEMS

Selected results for the 3DFR test system are presented in Figures 5, 6, and 7 to illustrate the type of data we collected. As shown in Figure 5, the weighted average score generally converges asymptotically to an optimum as sampling increases. The scores from all matching methods converge to within 90% of the optimum in about 10 seconds/molecule, whereas the URT method requires about 100 seconds/molecule. The weighted rank correlation in Figure 6 also shows convergent behavior, but with some interesting differences. It goes through much wider fluctuations, indicating that small changes in score have large effects on the rankings of the top scoring molecules. It appears to discriminate among the different methods, selecting BRM and SGM as superior, BGM and URM as next best, and URT as worst again. In particular, BRM and SGM both show a rapid initial rise, indicating that, with very
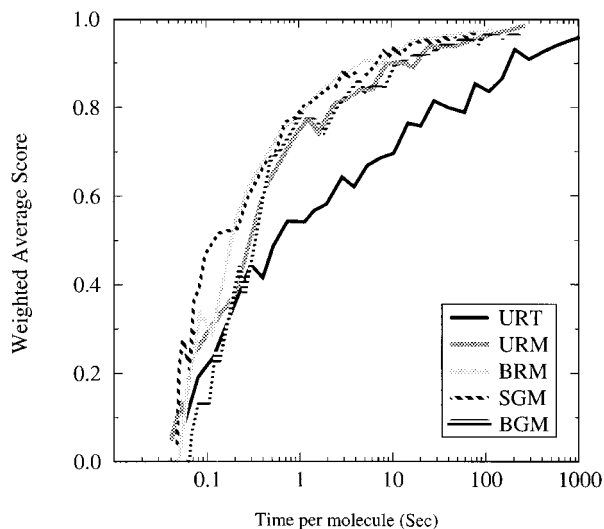
FIGURE 5. Weighted average score for molecules docked to 3DFR using SPHGEN site points. Each curve represents a search algorithm in Table I. Each data point is a weighted average of the score for all molecules in a particular run using the equation in Table IV.

little sampling, these methods come closest to predicting the rankings of the top scoring molecules. The convergence of weighted RMSD in Figure 7 indicates how long it takes the different methods to predict reproducibly the same binding mode of the top scoring molecules. Although the two top performing methods converge in predicted score
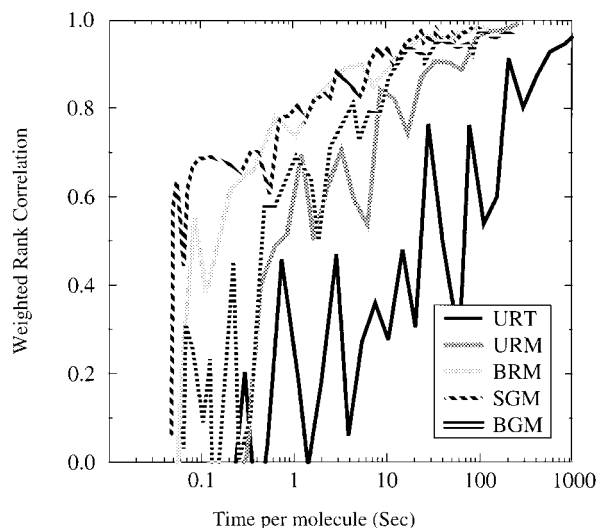

FIGURE 6. Weighted rank correlation for molecules docked to 3DFR using SPHGEN site points. Each data point is an average of the rank correlation for all molecules in a particular run using the equations in Table IV.
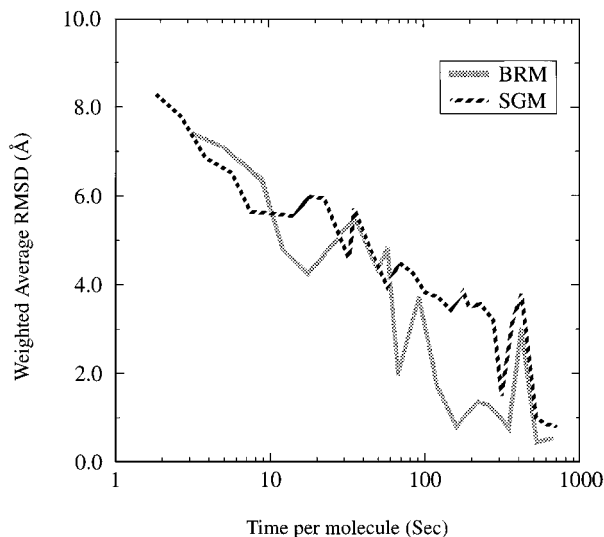

FIGURE 7. Weighted average RMSD for molecules docked to 3DFR using SPHGEN site points. Only the two best search algorithms from Table I are presented. Each data point is an average of the RMSD for all molecules in a particular run using the equations in Table IV.

and ranking in about 10 seconds, they require about 500 seconds before they consistently predict the same binding mode. This result indicates that, for these molecules in this site, several good scoring orientations must exist that are close in score but distant in space. We found similar results for the other sites as well.

We found the weighted and unweighted forms of the average score and rank correlation to be the most relevant in assessing database screening performance. This gives us four measurements of five methods over five sites. Instead of presenting 100 different curves, we have condensed each curve into a single value—the convergence time—which represents the time at which the 90% threshold value is passed (and not recrossed). In Table VI, we present the convergence times along with the speed improvement factor of each method compared to the URT method.

SCORE CONVERGENCE

With respect to the unweighted average score in Table VI(A), all matching methods show roughly equivalent convergence. and outperform URT by a factor of 10. Therefore, on average, site points provide a much more succinct description of the active site than the smallest enclosing box, especially when searching a site that is large or difficult

TABLE VI.
Convergence Properties of Search Methods.[a]

Site	Search Methods				
	URT	URM	BRM	SGM	BGM
(A) Unweighted Average Score					
121D	200 (1x)	10 (20x)	20 (10x)	10 (20x)	8 (30x)
1ULB	200 (1x)	10 (20x)	20 (8x)	20 (10x)	8 (20x)
3DFR	100 (1x)	9 (10x)	10 (10x)	10 (10x)	20 (6x)
4FAB	30 (1x)	7 (4x)	8 (4x)	7 (5x)	7 (4x)
9HVP	200 (1x)	7 (30x)	5 (40x)	5 (40x)	6 (30x)
Mean	100 (1x)	9 (10x)	10 (10x)	9 (10x)	9 (10x)
(B) Rank-weighted average score					
121D	400 (1x)	10 (40x)	6 (80x)	3 (100x)	8 (50x)
1ULB	1000 (1x)	60 (20x)	30 (40x)	20 (70x)	30 (40x)
3DFR	200 (1x)	20 (9x)	10 (20x)	8 (30x)	20 (10x)
4FAB	90 (1x)	7 (10x)	4 (20x)	4 (20x)	7 (10x)
9HVP	200 (1x)	7 (30x)	2 (100x)	4 (40x)	5 (40x)
Mean	300 (1x)	20 (20x)	7 (40x)	6 (50x)	10 (30x)
(C) Unweighted rank correlation					
121D	400 (1x)	20 (20x)	10 (30x)	7 (60x)	10 (40x)
1ULB	1000 (1x)	60 (20x)	40 (20x)	30 (30x)	80 (10x)
3DFR	400 (1x)	50 (9x)	20 (20x)	10 (30x)	40 (10x)
4FAB	200 (1x)	30 (7x)	10 (20x)	20 (7x)	50 (3x)
9HVP	700 (1x)	20 (40x)	8 (90x)	10 (60x)	20 (50x)
Mean	500 (1x)	30 (10x)	20 (30x)	20 (30x)	30 (20x)
(D) Rank-weighted rank correlation					
121D	400 (1x)	20 (30x)	10 (40x)	7 (60x)	20 (30x)
1ULB	1000 (1x)	60 (20x)	20 (60x)	20 (70x)	80 (20x)
3DFR	600 (1x)	100 (6x)	10 (40x)	10 (40x)	20 (30x)
4FAB	400 (1x)	10 (30x)	3 (100x)	4 (90x)	10 (30x)
9HVP	300 (1x)	7 (40x)	2 (100x)	4 (60x)	3 (80x)
Mean	500 (1x)	20 (20x)	7 (70x)	8 (60x)	20 (30x)

[a] For each sampling method and receptor site, we find the amount of sampling, in seconds, beyond which all values are within 90% of the maximum. With the time of URT as the reference, the relative speed factor of each method is reported in parentheses. The geometric mean over all receptor site values is reported at bottom. Because of the large uncertainty in these values, all values are rounded to one significant digit.

to define *a priori*. Because the site points were generated based on general considerations of shape, this result should be generalizable to other ''negative image'' techniques, like the shape-based critical point methods of Nussinov and Wolfsen,[44] and the energetic probe methods of Goodford.[45] The rank-weighted average score in Table VI(B) shows more discrimination among the matching methods. While the two uniform random methods (URT and URM) had more difficulty converging with the top-scoring molecules, BRM and SGM actually converged more quickly. This implies that trying a uniform number of orientations for the

molecules in the database is inefficient with respect to processing the top-scoring molecules. BGM performed better than URM, but not as well as either BRM or SGM.

RANK CONVERGENCE

The convergence of the rank correlation further confirms the differences between methods. The unweighted rank correlations in Table VI(C) again show a general 10- to 30-fold advantage of using site points to dock the molecules. Interestingly, the time required to get the rank correct is two- to

fivefold greater than the time to get the average score [Table VI(A)] correct. The weighted rank correlations in Table VI(D) also further discriminate among the methods. URM and BGM still show the 20- to 30-fold advantage over URT. BRM and SGM again outperform URT by 60- to 70-fold. Of all measures, this last one is arguably the most relevant to database screening, because absolute scores are generally not used as strict cutoffs, but instead the rankings are used to select some subgroup of molecules whose number is amenable to further processing. Most often, it is the top-scoring subgroup of most interest, so using a rank-weighted correlation should focus our attention on how the methods treat this particular subgroup of molecules. Therefore, it appears that, of the methods investigated here, BRM and SGM are the best suited for database screening.

BGM VERSUS SGM

The single graph matching method clearly outperforms the existing bipartite graph matching method by up to twofold in speed. This result may appear counterintuitive, because SGM is exhaustive, whereas BGM uses heuristics to speed the search. However, the precomputing of the adjacency matrix and the rapid processing of the adjacency lists show that reformulating the problem into an efficient form can be just as effective as using heuristics. SGM has the additional advantage of requiring fewer fundamental matching parameters than BGM (Table II).

BRM VERSUS SGM

Why does a simplistic random matching algorithm, BRM, perform so competitively with the distance matching algorithm, SGM, and even outperform BGM? Does this suggest that distance matching is an unnecessarily complicated solution to docking? We seek to resolve this question by breaking the problem into three parts.

1. What is the disadvantage of sampling each molecule uniformly?
2. Why does distance matching sample molecules nonuniformly?
3. Is an orientation from random matching just as good as one from distance matching?

First, spending the same time on each molecule may result in spending too little time on the better scoring molecules. One feature of the force field scoring function is that it tends to favor larger molecules.[21] For the set of molecules used in this study, we have plotted the best score for each molecule against its size in Figure 8. Although there is some trend, the correlation coefficient is not large. A stronger trend exists in Figure 9, relating molecule size to bump filtering. Large molecules have a greater propensity to bump into receptor atoms when oriented in the site. Because we use a bump filter in DOCK to discard poor orientations before the more computationally expensive scoring and optimization steps, we are more likely to discard an orientation of a large molecule than that of a small molecule. Forcing a uniform number of matches per molecule would then result in a size-biased attrition through the bump filter and, overall, spending less time on the potentially better scoring, larger molecules.

Second, nonuniform sampling arises in distance matching because the number of matches is related to the number of internal distances that the ligand has in common with the site points. Larger molecules have more internal distances, and so will tend to have more in common. Therefore, larger molecules tend to generate more matches than smaller molecules. However, a molecule that is not larger, but rather similar in shape to the binding site, will also have more internal distances
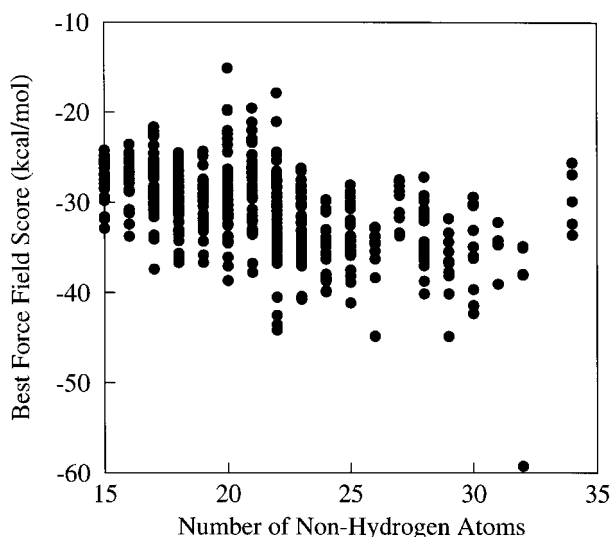

FIGURE 8. How molecule size affects force field score. The best force field score for each molecule is plotted versus the number of nonhydrogen atoms in the molecule. The plots for each receptor are pooled into this single plot to show the overall trend. Fitting a line to these data yields an R^2 value of 0.24.
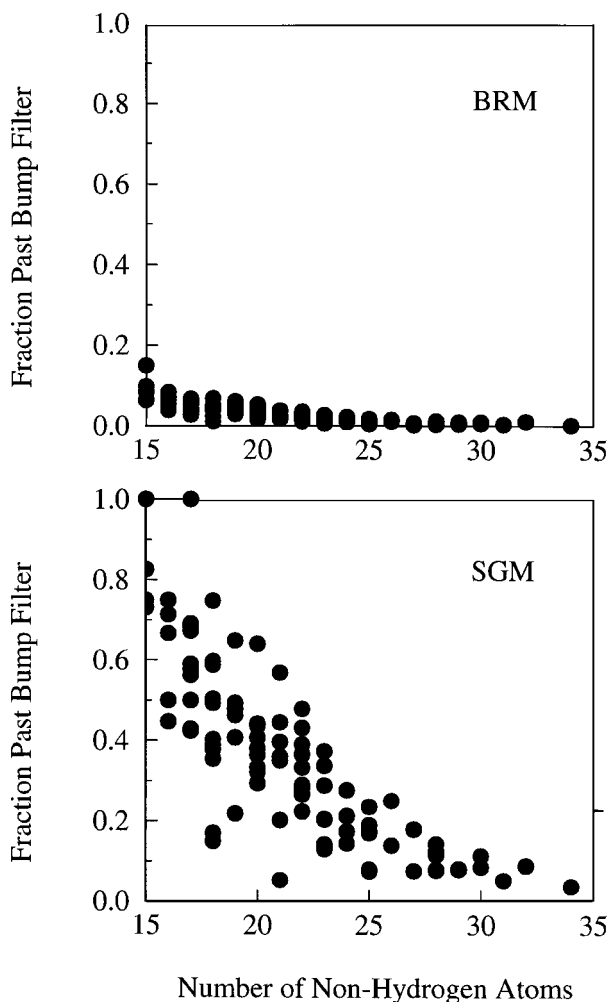
FIGURE 9. How molecule size affects bump filtering. In a given docking run, we compute the fraction of orientations that pass the bump filter for each molecule, and then plot this value against the size of the molecule. (**Top**) BRM method. Average filter rate is 0.031. (**Bottom**) SGM method. Average filter rate is 0.36.

in common. Therefore, distance matching will spend more time on molecules that are complementary in shape to the receptor. Thus, the BRM method benefits from using distance matching to determine the number of orientations to try simply because it will bias its efforts toward the larger and/or more complementary molecules.

Third, an orientation from random matching is, on average, not as good as one from distance matching. We can check the quality of these orientations by examining how they survive the bump filter as depicted in Figure 9. The orientations from distance matching are at least ten times more likely to make it past this filter, so they are indeed

superior. To understand why this superiority does not translate directly into faster docking, we must consider the computational bottleneck of the current implementation of DOCK. If the early docking steps—matching, orienting, and bump filtering—were to consume a dominant portion of the total cpu time, then SGM would be up to ten times faster than BRM. On the other hand, if the later docking steps—scoring and optimization—were to consume a dominant portion of the total cpu time, then BRM would be equivalent in speed to SGM. With the current implementation of the minimizer, the scoring and optimization steps indeed consume an overwhelming portion of cpu time ($> 90\%$), so the ability of distance matching to form high-quality orientations is for now unrewarded. BRM is kept competitive because, in concert with heavy pruning by the bump filter, it forms orientations that are suitable starting points for optimization.

BRM VERSUS BGM

The fact that BRM outperforms BGM by up to a factor of two points out the critical importance of how processing time is allocated among the different molecules in the database. For maximum efficiency, a search algorithm must bias its efforts toward the most highly ranked molecules. BGM indeed has such a bias built in, but the heuristics have the effect of reducing the magnitude of bias inherent to the unrestrained matching found in SGM and BRM.

FUTURE DIRECTIONS

It appears sensible to incorporate the single graph matching technique into version 4.0 of DOCK, because its results are of high quality and its potential for speed improvement is high. The speed gains might arise from additional orientation filtering or by fundamental improvements in the optimization technique. The new matching algorithm will provide a solid algorithmic foundation on which to base further development of molecular docking, including the addition of sophisticated search constraints like chemical labels and critical clusters, as well as the explicit treatment of ligand flexibility.

Until a faster scoring and optimization method is implemented, it may be useful to preprocess the set of orientations generated by matching. For instance, an implementation of degeneracy checking has been tried in which similar orientations are

removed at the matching stage.[29] Because orienting is also relatively facile, it implies that degeneracy checking could instead be performed as an RMS deviation calculation between real orientations. Rarey et al. present a rapid RMSD evaluation technique based directly on the rotation matrices[17] that would be applicable. It might also be possible identify a unique property of orientations that lead to the best scores upon minimization (e.g., degree of surface overlap with receptor). We intend to further study the nature of orientations generated by BRM that pass the bump filter, to see why they manage to score so well upon minimization. Based upon this knowledge, a filter could be constructed that enriches the set of orientations generated by regular matching prior to minimization.

An additional avenue for improvement would be to try alternative methods to generate site points. The current shape-based site points may not be entirely consistent with a force-field-based scoring function.[21] Including force-field considerations during site point construction, might make the site points more relevant as points on which to position ligand atoms, thereby increasing the quality of docked orientations.

We are working to extend the current docking protocol to include ligand and receptor flexibility. Ligand flexibility can be incorporated in several ways. The most straightforward approach is to dock multiple conformations of the ligand separately.[21] A potentially more efficient method is to use distance geometry to build a ligand conformation that fits a subset of receptor site points.[10] Another viable option is the ''divide and conquer'' strategy, in which a flexible molecule is broken into rigid fragments, the fragments are docked independently, and the molecule is rebuilt from adjacent fragment orientations.[31-33] Limited receptor flexibility is being investigated during the step of score evaluation by superimposing multiple receptor conformers on a single score potential grid.[46]

Conclusion

We evaluated various search algorithms for automated molecular docking that range in complexity from purely random to site point–directed, graph-theoretical matching methods. Our basis of comparison was how quickly each docking method could correctly score and rank a database of molecules. Over a broad range of active site environments, it is at least tenfold more efficient to use a collection of site points to describe the active site search volume than to use the smallest enclosing box. Using graph theoretical matching techniques boosts this relative efficiency higher. The bipartite graph matching used in the current DOCK version 3.5 improves efficiency up to 30-fold. The single graph matching proposed for DOCK version 4.0 improves efficiency up to 60-fold. Because single graph matching is not only faster, but also less complicated than the bipartite graph matching used in version 3.5, we feel it will be an important advance in docking technology.

References

1. A. Caflisch, P. Niederer, and M. Anliker, *Proteins*, **13**, 223 (1992).

2. A. Di Nola, D. Roccatano, and H. J. C. Berendsen, *Proteins*, **19**, 174 (1994).

3. D. S. Goodsell and A. J. Olson, *Proteins*, **8**, 195 (1990).

4. J. B. Moon and W. J. Howe, *Proteins*, **11**, 314 (1991).

5. R. Abagyan, M. Totrov, and D. Kuznetsov, *J. Comput. Chem.*, **15**, 488 (1994).

6. A. Miranker and M. Karplus, *Proteins*, **11**, 29 (1991).

7. A. R. Leach and I. D. Kuntz, *J. Comput. Chem.*, **13**, 730 (1992).

8. A. R. Leach, *J. Mol. Biol.*, **235**, 345 (1994).

9. M. Y. Mizutani, N. Tomioka, and A. Itai, *J. Mol. Biol.*, **243**, 310 (1994).

10. A. S. Smellie, G. M. Crippen, and W. G. Richards, *J. Chem. Inf. Comput. Sci.*, **31**, 386 (1991).

11. R. S. Judson, E. P. Jaeger, and A. M. Treasurywala, *Theochem J. Mol. Struct.*, **114**, 191 (1994).

12. G. Jones, P. Willett, and R. C. Glen, *J. Mol. Biol.*, **245**, 43 (1995).

13. C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, *J. Comput.-Aided Mol. Design*, **9**, 113 (1995).

14. Y. P. Pang and A. P. Kozikowski, *J. Comput.-Aided Mol. Design*, **8**, 683 (1994).

15. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, and I. A. Vakser, *Proc. Natl. Acad. Sci. USA*, **89**, 2195 (1992).

16. D. Fischer, S. L. Lin, H. L. Wolfson, and R. Nussinov, *J. Mol. Biol.*, **248**, 459 (1995).

17. M. Rarey, S. Wefing, and T. Lengauer, *J. Comput.-Aided Mol. Design*, **10**, 41 (1996).

18. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, *J. Mol. Biol.*, **161**, 269 (1982).

19. F. S. Kuhl, G. M. Crippen, and D. K. Friesen, *J. Comput. Chem.*, **5**, 24 (1984).

20. M. C. Lawrence and P. C. Davis, *Proteins*, **12**, 31 (1992).

21. M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, *J. Comput.-Aided Mol. Design*, **8**, 153 (1994).

22. H.-J. Bohm, *J. Comput.-Aided Mol. Design*, **8**, 623 (1994).

23. I. D. Kuntz, *Science*, **257**, 1078 (1992).

24. D. R. Ferro and J. Hermans, *Acta Crystallogr.*, **A33**, 345 (1977).

25. Private communication from C. M. Oshiro and G. Golub.

26. B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, *J. Comput. Chem.*, **13**, 380 (1992).

27. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.*, **13**, 505 (1992).

28. E. C. Meng, D. A. Gschwend, J. M. Blaney, and I. D. Kuntz, *Proteins*, **17**, 266 (1993).

29. D. A. Gschwend and I. D. Kuntz, *J. Comput.-Aided Mol. Design*, **10**, 123 (1996).

30. C. Bron and J. Kerbosch, *Commun. ACM*, **16**, 575 (1973).

31. R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, *J. Med. Chem.*, **29**, 2149 (1986).

32. M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard, *Proteins*, **19**, 199 (1994).

33. H.-J. Bohm, *J. Comput.-Aided Mol. Design*, **6**, 61 (1992).

34. R. L. DesJarlais and J. S. Dixon, *J. Comput.-Aided Mol. Design*, **8**, 231 (1994).

35. Distributed by Molecular Design Ltd., San Leandro, CA.

36. A. Rusinko et al., *J. Chem. Inf Comput. Sci.*, **29**, 251 (1989).

37. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).

38. E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng, in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*, F. H. Allen, G. Bergerhoff, and R. Seivers, Eds., Data Commission of the International Union of Crystallography, Bonn, 1987, pp. 107–132.

39. L. Tabernero, N. Verdaguer, M. Coll, I. Fita, G. A. Van der Marel, J. H. Van Boom, A. Rich, and J. Aymami, *Biochemistry*, **32**, 8403 (1993).

40. S. E. Ealick, Y. S. Babu, C. E. Bugg, M. D. Erion, W. C. Guida, J. A. Montgomery, and J. A. Secrist III, *Proc. Natl. Acad. Sci. USA*, **88**, 11540 (1991).

41. J. T. Bolin. D. J. Filman, D. A. Matthews, R. C. Hamlin, and J. Kraut, *J. Biol. Chem.*, **257**, 13650 (1982).

42. J. N. Herron, X. He, M. L. Mason, E. W. Voss Jr., and A. B. Edmundson, *Proteins*, **5**, 271 (1989).

43. J. Erickson, D. J. Neidhart, J. Van Drie, D. J. Kempf, X. C. Wang, D. W. Norbeck, J. J. Plattner, J. W. Rittenhouse, M. Turon, N. Wideburg, W. E. Kohlbrenner, R. Simmer, R. Helfrich, D. A. Paul, and M. Knigge, *Science*, **249**, 527 (1990).

44. S. L. Lin. R. Nussinov, D. Fischer, and H. J. Wolfsen, *Proteins*, **18**, 94 (1994).

45. C. A. Reynolds, R. C. Wade, and P. J. Goodford, *J. Mol. Graph.*, **7**, 103 (1989).

46. R. M. A. Knegtel, I. D. Kuntz, and C. M. Oshiro, *J. Mol. Biol.*, **266**, 424 (1997).