New Expi293 suite of products for structural biology, inducible expression, and protein labeling

Learn more

Protein Science (1998), 7:938–950. Cambridge University Press. Printed in the USA.

Flexible ligand docking using conformational ensembles

DAVID M. LORBER AND BRIAN K. SHOICHET

Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, Chicago, Illinois 60611-3008

(RECEIVED September 30, 1997; ACCEPTED December 17, 1997)

Abstract

Molecular docking algorithms suggest possible structures for molecular complexes. They are used to model biological function and to discover potential ligands. A present challenge for docking algorithms is the treatment of molecular flexibility. Here, the rigid body program, DOCK, is modified to allow it to rapidly fit multiple conformations of ligands. Conformations of a given molecule are pre-calculated in the same frame of reference, so that each conformer shares a common rigid fragment with all other conformations. The ligand conformers are then docked together, as an ensemble, into a receptor binding site. This takes advantage of the redundancy present in differing conformers of the same molecule. The algorithm was tested using three organic ligand protein systems and two protein-protein systems. Both the bound and unbound conformations of the receptors were used. The ligand ensemble method found conformations that resembled those determined in X-ray crystal structures (RMS values typically less than 1.5 Å). To test the method's usefulness for inhibitor discovery, multi-compound and multi-conformer databases were screened for compounds known to bind to dihydrofolate reductase and compounds known to bind to thymidylate synthase. In both cases, known inhibitors and substrates were identified in conformations resembling those observed experimentally. The ligand ensemble method was 100-fold faster than docking a single conformation at a time and was able to screen a database of over 34 million conformations from 117,000 molecules in one to four CPU days on a workstation.

Keywords: conformation; flexibility; molecular docking; structure-based drug design

Molecular docking algorithms fit molecules together in complementary fashions. The technique has attracted increasing attention as a way to predict the geometries of biomolecular complexes (for reviews see Kuntz et al., 1994; Malby et al., 1994; Janin, 1995) and to discover novel ligands as leads for drug design (Bartlett et al., 1989; Kuntz, 1992).

Despite important successes (Bartlett et al., 1989; DesJarlais et al., 1990; Bodian et al., 1993; Olson & Goodsell, 1993; Ring et al., 1993; Rutenber et al., 1993; Shoichet et al., 1993; Strynadka et al., 1996a), molecular docking faces several methodological problems. These include predicting the relative binding affinities of different possible complexes, identifying binding sites on receptors, and allowing for molecular flexibility in the docking event. Together, these and related challenges make up "the docking problem" (Connolly, 1985). Here, we will consider one aspect of the docking problem, that of allowing for ligand flexibility in molecular docking.

Docking algorithms began by treating molecules as rigid bodies (Rose, 1978; Wodak & Janin, 1978; Kuntz et al., 1982; Kuhl et al., 1984; Connolly, 1985; Jiang & Kim, 1991). This approach is successful if the conformational change on complex formation is small. In such situations, docking programs have predicted the structures of molecular complexes (Janin, 1995; Shoichet & Kuntz, 1996; Strynadka et al., 1996a) and discovered novel ligands (Bartlett et al., 1989; DesJarlais et al., 1990; Bodian et al., 1993; Ring et al., 1993; Rutenber et al., 1993; Shoichet et al., 1993). However, as conformational change becomes more significant, the accuracy of rigid body docking programs diminishes. For example, specific hydrogen bonds may be missed due to residue flexibility (Shoichet et al., 1993) and larger scale features are missed due to hinge-bending motions (Rutenber et al., 1993; Strynadka et al., 1996a). Likewise, many novel ligands discovered by docking programs have been fairly rigid; many flexible ligands may be missed because they are docked as rigid bodies in the wrong conformation.

Investigators have long recognized the importance of ligand flexibility in the docking process. The difficulty has been allowing for flexibility without greatly increasing calculation times. The number of possible ligand conformations rises in proportion to the power of the number of the bonds rotated. Hence, for an organic molecule with ten rotatable bonds, the number of possible conformations would be 59,049, if only three minima were considered per bond, allowing for six minima leads to 3.48×10^9 conformations. Regardless of increases in computer speed, advances in algorithms are needed to address a problem that grows exponentially.

Reprint requests to: Brian K. Shoichet, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, IL 60611-3008; e-mail: b-shoichet@nwu.edu.

One approach has been to use Monte Carlo dynamics simulation and simulated annealing to sample ligand flexibility (Goodsell & Olson, 1990; Caflisch et al., 1992; Stoddard & Koshland, 1992; Goodsell et al., 1996; Wasserman & Hodge, 1996). In this approach, a molecular mechanics force-field orients the ligand in a binding site and adjusts its conformation. The binding site is either held static (Goodsell & Olson, 1990) or allowed some flexibility (Wasserman & Hodge, 1996). An advantage of this method is that it can be applied to a variety of molecular docking problems. Its energies are comparable to those of other force-field methods, and the intramolecular energies may be integrated with the intermolecular energies. A disadvantage is that the method can be slow to identify global minima and can spend significant time exploring local minima.

To overcome the exponential problem inherent in most flexible docking methods, fragment or incremental construction algorithms divide ligands into modular pieces. A particular fragment is docked, and the full ligand is built up incrementally; alternatively, separately docked fragments are joined together. These methods have been used to design new ligands (Miranker & Karplus, 1991, 1995; Moon & Howe, 1991; Bohm, 1992; Lewis et al., 1992; Ho & Marshall, 1993) and to sample ligand conformational space (DesJarlais et al., 1986; Leach & Kuntz, 1992; Sandak et al., 1995; Rarey et al., 1996; Welch et al., 1996). Fragment methods eliminate many possible combinations early in the construction of the molecule, pruning the search tree to their advantage. Like the Monte Carlo methods, they can select conformations according to different site geometries. Recent applications of fragment methods (Leach & Kuntz, 1992; Rarey et al., 1996; Welch et al., 1996) are relatively efficient and accurate. A drawback is that the initial docking of the fragment or fragments necessarily reduces the amount of information for the fitting part of the algorithm, since the fragment contains less information than does the full ligand. Also, ligands can be split up in many different ways, and the exact identity of the lead fragment or fragments can affect the accuracy of the docking result (Rarey et al., 1996). The approach is nevertheless robust enough to search diverse sets of ligands automatically (Welch et al., 1996).

Genetic algorithms (Oshiro et al., 1995; Verkhivker et al., 1996; Jones et al., 1997) approach the problem by representing possible ligand conformations and orientations in a modular fashion. Using a series of operations resembling genetic crosses and mutations, followed by selection against a scoring function, increasingly favorable populations of possible complexes are propagated. The quality of the result depends upon the starting genes, the number of mutations and crossover operations, as well as the accuracy and radius of convergence of the scoring function. These methods have accurately predicted molecular complexes. Presently, genetic algorithms are relatively slow for flexible ligand docking.

Flexible ligand docking is an area of active investigation. The methods that we have sketched above, and those that are related to them, have had some important successes and are rapidly developing. Still, a general solution to the flexible docking problem has not been proposed. The current algorithms remain prone to sampling problems and retain fairly long calculation times, despite their considerable improvement on brute force methods. There is room for further methods development in this field.

Here, we describe a new algorithm for examining the binding of flexible ligands to proteins. Although this method is not a general solution to the problem of flexible ligand docking, it is relatively efficient and accurate, and it can search large databases of flexible molecules rapidly.

Methods

Overview of the approach

The method begins by considering two characteristics of flexible ligand docking. First, generating an orientation of a ligand in a binding site may be separated from calculating a conformation of that ligand in that particular orientation. Second, multiple conformations of a given ligand usually have some portion in common (e.g., internally rigid atoms such as ring systems), and therefore, contain redundancies. Since many different orientations of a ligand are sampled in a binding site, this redundancy can lead to the internally rigid atoms being sampled repeatedly.

To reduce the redundancy of docking multiple conformations, the ligand may be broken down into a rigid fragment and a conformationally flexible fragment. If the ensemble of ligand conformations to be considered is generated before docking, so that each conformation places the rigid fragment identically, then the rigid fragment may be represented once for the set of conformations. By extension, when the ensemble of conformations is docked into a receptor binding site, it is only necessary to calculate the placement of the rigid fragment once, irrespective of the number of conformations of the given molecule. Further, orienting the rigid fragment in the site defines the placement of each conformation of the flexible portion of the molecule. For an ensemble of n ligand conformations, calculating an orientation in the site is performed once, and the evaluation of the fits, which will differ from conformation to conformation, is performed n times (Fig. 1).

We test this ligand-ensemble docking method against five well-studied enzyme-ligand complexes (Table 1). In three of these, the ligand is an organic molecule varying in complexity from 5 to 13 rotatable bonds (Fig. 2). In the remaining systems, the ligand is a protein for which we consider a small subset of possible flexible groups. We also test the applicability of this method to novel inhibitor discovery. In inhibitor discovery applications, multi-compound databases are docked into a binding site. The databases typically have 100,000 to 200,000 molecules, and the goal is to find novel candidate molecules that complement the binding site. We dock multi-compound multi-conformer databases against two of the enzymes to test the usefulness of the ensemble method for the discovery of novel lead compounds. Here, we will usually consider the enzymes both in the conformation that they adopt when the ligand is bound (bound conformation) and in the conformation they adopt when their structure is determined in the absence of the ligand (unbound conformation). The exception is the protein-protein docking cases, where we only consider the unbound conformations of the docking molecules. Although this ensemble method has several drawbacks, to which we shall return, we find that it can efficiently dock flexible molecules, typically to within 1.5 Å RMS of the experimental structure. The ensemble method can find known, flexible ligands from among diverse candidates in database searches.

Modifications to DOCK

The rigid body docking program DOCK fits molecules into receptors in two steps: calculating an orientation for the molecule in a binding site, and then evaluating the fit of the molecule in that site. To calculate an orientation, DOCK determines the internal distances between sites pre-calculated in the receptor binding region, and matches these to the internal distances between ligand atoms (Kuntz et al.,

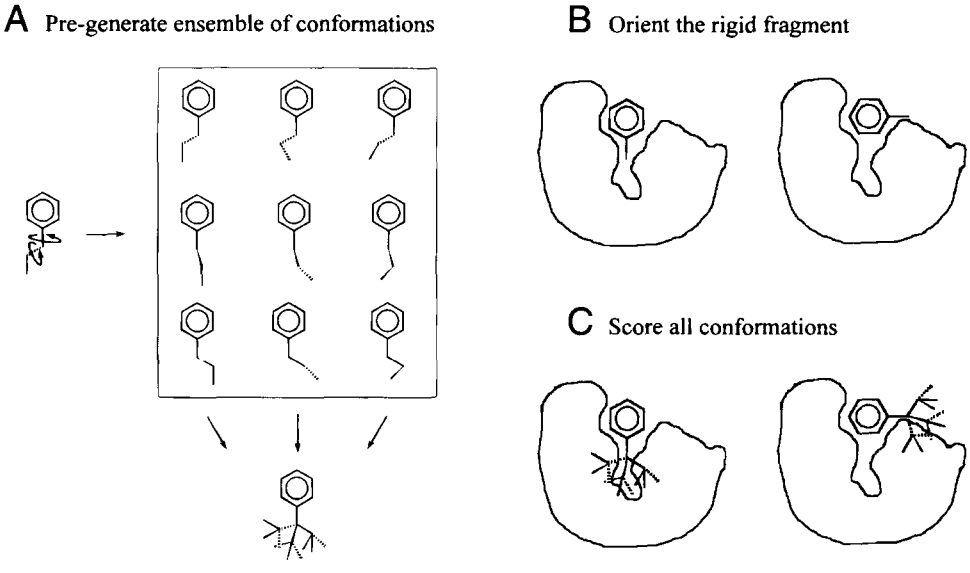

Fig. 1. Overview of the ligand ensemble method. **A:** The largest group of internally rigid atoms is fixed in position and the conformational space of the rest of the molecule is systematically sampled at 60° or 120° increments. **B:** The rigid fragment common to all conformations of the molecule is oriented in the binding site. **C:** All flexible fragments of the molecule are scored in the orientation of the rigid fragment.

1982). The receptor sites may be thought of as pseudo-atom positions; they are regions that are pre-organized to complement ligand atoms. We commonly use the sphere method of Kuntz (Kuntz et al., 1982) to identify these pseudo-atom positions on the receptor, but other criteria may be used (Shoichet, 1996). Upon finding a set of at least four matching receptor-ligand points, DOCK calculates a translation-rotation matrix that is used to generate an orientation for the ligand in the site. The orientation of the molecule is then scored for fit to the receptor. In the calculations presented here, the score measures steric and electrostatic complementarity.

Figure 1 illustrates the flow of the new algorithm. DOCK was modified to allow calculation of ligand orientations using only the internally rigid atoms of the ligand (Fig. 1B); previously, all ligand atoms were used in receptor-ligand point matching. All ligand atoms are still used to score an orientation. The translation-rotation matrix for orienting the rigid fragment of the molecule is applied to the flexible atoms to generate an orientation for scoring. The same matrix is applied to all conformations in a given orientation (Fig. 1C). All atoms of the ligand must be in the same frame of reference for this to be successful. The absolute frame of reference is unimportant and can change from molecule to molecule (Fig. 1A).

Systematic generation of ligand conformations

Multiple conformations of ligands were generated in the same frame of reference as the largest group of internally rigid atoms using a script written in the SYBYL Programming Language within the program SYBYL (Tripos Inc., St. Louis, MO). The rigid fragment was identified as the largest ring system in each molecule including all atoms attached by one bond to the ring. Conforma-

Table 1. *Protein crystal structures. Protein structures used in docking calculations*

Enzyme/Ligand	Res. (Å)	PDB id	Reference
Dihydrofolate reductase (alone)	2.4	6dfr	Bystroff et al., 1990
DHFR/MTX Complex	1.7	3dfr	Bolin et al., 1982
Thymidylate synthase (alone)	2.1	3tms	Perry et al., 1990
TS/dUMP Complex	2.0	1syn	Stout & Stroud, 1996
Lactate dehydrogenase (alone)	2.0	6ldh	Abad-Zapatero et al., 1987
LDH/NADH Complex	2.2	9ldb	Dunn et al., 1991
Trypsin (alone)	1.6	2ptn	Walter et al., 1982
BPTI (alone)	1.5	4pti	Marquart et al., 1983
Trypsin/BPTI Complex	1.9	2ptc	Marquart et al., 1983
TEM-1 (alone)	1.8	—	Strynadka et al., 1996a
BLIP (alone)	2.1	—	Strynadka et al., 1994
TEM-1/BLIP Complex	1.7	—	Strynadka et al., 1996b

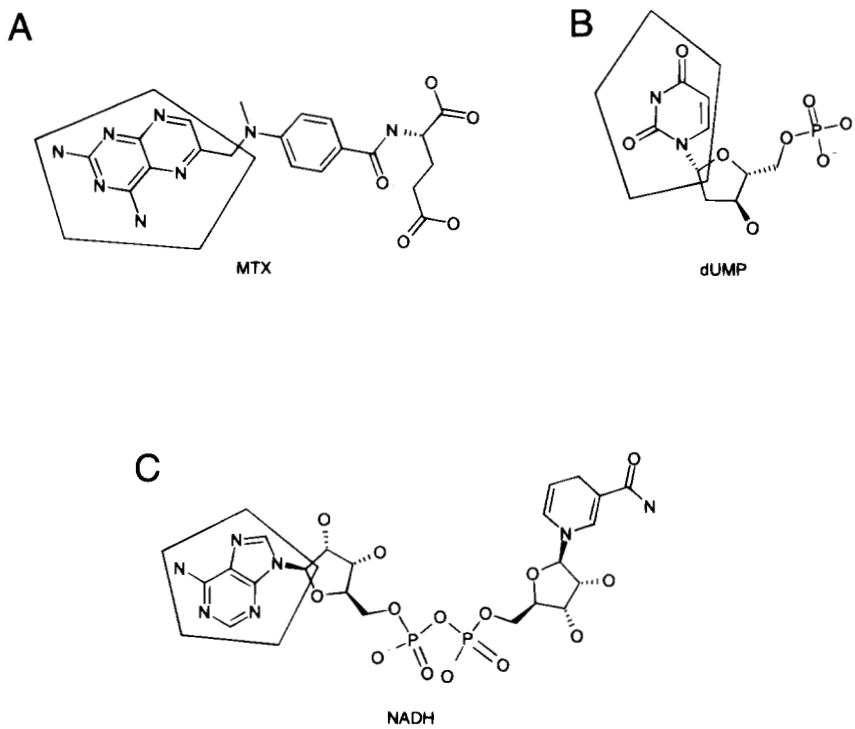

Fig. 2. Organic ligands used in docking tests. The outline regions of the three small-molecule ligands represent the rigid fragments used for docking. **A:** Methotrexate (MTX). **B:** 2′-deoxyuridine 5′-monophosphate (dUMP). **C:** Nicotinamide adenine dinucleotide (NADH).

tions were generated by rotating all single, non-terminal, acyclic bonds in specified increments. All rings were considered to be conformationally rigid. Bonds with three-fold symmetry (e.g., phosphate, phosphonate, sulfate, sulfonate, methyl, and charged amino) were not rotated. For molecules with fewer than five rotatable bonds, rotation increments were 60°. For larger molecules, the bond closest to the rigid fragment was rotated in 60° increments and all other bonds were rotated in 120° increments. In molecules containing more than nine rotatable bonds, amino, carboxyl, and hydroxyl groups were not rotated. If fewer than 13 rotatable bonds remained, conformers were generated, and up to 500 of the lowest energy conformers were written to a ligand-ensemble in preparation for docking (Fig. 1A). Molecules containing more than 13 rotatable bonds were excluded for these tests. All organic ligands had Gasteiger-Marsili charges as assigned by SYBYL (Meng et al., 1992). Selected residues were rotated in the uncomplexed forms of the two protein ligands (Asp49 and Phe142 in β-lactamase inhibitory protein, BLIP, and Lys15, Arg17, and Arg39 in bovine pancreatic trypsin inhibitor, BPTI) in 120° increments. The rest of the protein was used as the rigid fragment. For BPTI, 417 conformations were calculated and for BLIP, 162 conformations were calculated.

Starting structures for conformer generation were obtained from the Brookhaven Protein Databank (PDB) (Bernstein et al., 1977), and the three-dimensional Available Chemicals Directory (ACD) from MDL, Inc. (Guner et al., 1991). The ACD is a collection of approximately 150,000 commercially available molecules, for which three-dimensional coordinates have been calculated using the CONCORD program (Rusinko et al., 1987). Only one conformation for each molecule is represented in the database. The ligand ensembles derived from the receptor-bound conformation of the ligands (PDB ensembles) included the crystallographic conformations as known "correct" structures. The ligand ensembles derived from the ACD (ACD ensembles) were generated without reference to the crystallographic structure (Fig. 2).

Additionally, three multi-compound databases were generated: a "two-rings" database that contains 5,671 molecules with two fused six-membered rings, including pteridines such as methotrexate (MTX); a database of 281 molecules each containing a ring-sugar-phosphate motif, including nucleotides such as 2′-deoxyuridine 5′-monophosphate (dUMP), and a database of about 117,000 molecules from the ACD. The database of two-ring structures included forty-seven 2,4-diaminopteridines that were modeled in both their charged and uncharged pteridine ring forms (N1 nitrogen protonated or unprotonated, respectively). Both single conformer and multi-conformer (up to 500 conformations per molecule) versions of these two databases were generated. The DOCK database format was modified so that for each molecule the rigid fragment was represented once, followed by multiple conformations of the flexible fragment.

Preparation of test systems

Three organic-ligand receptor systems were used in both the bound and unbound conformations of the receptors. The three systems, dihydrofolate reductase/methotrexate (DHFR/MTX), thymidylate synthase/2′-deoxyuridine-5′-monophosphate (TS/dUMP), and lactate dehydrogenase/nicotinamide adenine dinucleotide (LDH/NADH) were obtained from the PDB (Table 1). For ease of comparison, the three unbound forms of the receptors were super-

imposed on the respective bound receptor conformations by minimizing RMS deviations (RMSDs) in conserved α-carbon coordinates using MidasPlus (Ferrin et al., 1988). The binding site on each receptor was defined as the residues within 10 Å of the complexed ligand and a solvent-accessible molecular surface was generated using the MS program (Connolly, 1983). A steric map was calculated with DISTMAP (Shoichet et al., 1992) using limits of 2.4 Å for polar atoms and 2.6 Å or 2.8 Å for non-polar atoms and an electrostatic potential map was calculated with DelPhi (Gilson & Honig, 1987) for the binding site. Steric limits of 2.0 Å for polar atoms and 2.2 Å for non-polar atoms were used in the complexed form of lactate dehydrogenase and 2.0 Å for all atoms in the uncomplexed form. Additionally, two protein-ligand intersections were allowed in the uncomplexed form of this receptor to accommodate receptor conformational change on ligand binding. Receptor spheres were generated using SPHGEN (Kuntz et al., 1982). The clusters of spheres were modified with a re-clustering algorithm (Shoichet et al., 1992) and by hand. The same surfaces, maps, and spheres were used for all experiments with a given form of a receptor. Solvation corrections were calculated for all organic molecule ligands with HYDREN (Rashin, 1990; Shoichet & Kuntz, 1993), and were subtracted from the ligand-protein interaction energies. For MTX, dUMP, and NADH, the energy for desolvating the molecules was calculated to be 102.5, 107.4, and 159.9 kcal/mol, respectively.

The two protein ligand systems, trypsin/BPTI and β-lactamase (TEM-1)/BLIP, were prepared similarly to the organic ligand systems. The ensembles of uncomplexed ligands were docked into the uncomplexed form of the receptors. The ligand scores were not corrected for solvation and two and seven ligand-receptor contacts were allowed for BPTI and BLIP, respectively. Focusing (Shoichet et al., 1992) was used in both protein-protein systems. The steric contact limits for trypsin and TEM-1 were 2.0 Å and 2.2 Å, respectively, for polar atoms and 2.4 Å, 2.6 Å, respectively, for non-polar atoms.

Docking

We first evaluated the ability of the new algorithm to dock a ligand into an active site using only its rigid fragment to determine if the fragment contained enough information. The three organic ligands, in their receptor-bound conformations, were each docked to the complexed form of their receptors using all ligand atoms for orientation generation. The values for the electrostatic interaction energy and the RMSDs from these calculations were treated as controls. The fragmented forms of the ligands (i.e., rigid fragment and single flexible fragment) were then docked to the same receptors. These calculations were repeated using the unbound forms of the receptors. In DOCK, several user-defined parameters affect the number of orientations that are calculated. For each calculation, these orientation parameters (typically bin sizes (Shoichet et al., 1992)) were optimized to give orientations with favorable scores for the various ligands. Because of the differing sizes of the full versus the fragmented ligands, orientation parameters often differed between calculations that used the full ligand for orientation calculation and those that only used the rigid fragments. For both the full ligand and rigid fragment calculations, the bin sizes typically were in the range of 0.2 Å to 0.6 Å. The same scoring parameters were used, whether the full molecule or only the rigid fragment was used, for orientation calculations.

The ability of the new algorithm to identify the most complementary conformation of a ligand from a ligand-ensemble was then evaluated by docking the three ligand-ensembles into their respective receptors. Each receptor was considered in its bound and unbound forms. These calculations used the same docking parameters, as in the single fragmented conformer calculations. To determine if the conformation generation method could find appropriate conformations without starting from the crystallographic structure of the ligand, as would be needed for novel ligand discovery, the ACD ensembles were docked using the same parameters. Two multi-compound, multi-conformer databases were docked to test the ability of the algorithm to identify known ligands from a diverse set of possibilities. Again, both the bound and unbound forms of the receptors were used.

Results

In each case, the performance of the ensemble method was compared to the performance of the rigid body method in terms of accurate ligand placement, interaction energy, and the speed of the calculation. We considered the performance of the ensemble method in a series of docking calculations of increasing complexity: (1) Calculating orientations using all ligand atoms and then repeating using only the rigid fragment (all ligand atoms are used for scoring in both cases) to evaluate the significance of the information loss that occurs in the ensemble method. (2) Docking conformational ensembles of the ligands that included the crystallographic conformation to evaluate the time required to dock the ensemble. (3) Docking conformational ensembles that were unbiased by, and did not include, the crystallographic conformation to evaluate the ability of this method to generate a structure similar to the experimental structure. (4) Docking multi-compound, multi-conformer databases to evaluate the ability of DOCK to identify known ligands from an ensemble database.

1. Significance of information loss

The new algorithm depends upon the ability to split ligand molecules into rigid and flexible fragments and use only the rigid fragments to calculate orientations (all ligand atoms are used to evaluate ligand-receptor fit). This involves a loss of information compared to the use of the entire molecule for calculating an orientation. To determine how detrimental this information loss would be, we docked single conformations of the ligands, first using all ligand atoms in orientation generation and then using only the rigid fragments. The fits generated using all atoms had low RMSDs from the crystallographic configurations, as did the fits generated using only the rigid fragments (Table 2). The energy scores for the all-atom orientations were typically slightly better than those for ligands oriented based only on their rigid fragments. Using only the rigid fragments to orient the ligand decreased the number of orientations identified, and consequently, the calculation time by approximately three-fold (Table 2). The calculations were repeated for the unbound forms of the receptors. Here, the fits generated typically had slightly higher RMSDs from the experimental complex structures, and the scores were worse (Table 2).

2. Calculation times for ensemble docking

We then docked ligand conformational ensembles generated from, and including, the complexed crystallographic conformation of the

Table 2. *Fragment vs. all-atom docking of the ligands*[a]

Enzyme/ligand	Fragment docked	RMS (Å)	Time (s)	Number of orientations	Score[b] (kcal/mol)
		Complexed form of receptor			
DHFR/MTX	All atoms	0.11	14.1	12,460	−20.7
DHFR/MTX	Internally rigid	0.40	10.6	1,242	−26.8
TS/dUMP	All atoms	0.41	11.5	31,281	−78.4
TS/dUMP	Internally rigid	0.41	2.8	6,314	−77.8
LDH/NADH	All atoms	0.81	41.6	54,428	47.8
LDH/NADH	Internally rigid	0.54	13.8	11,015	57.4
		Uncomplexed form of receptor			
DHFR/MTX	All atoms	0.82	36.9	87,111	3.6
DHFR/MTX	Internally rigid	0.88	7.0	16,335	14.1
TS/dUMP	All atoms	0.67	16.7	50,949	−13.8
TS/dUMP	Internally rigid	0.73	1.2	1,554	3.4
LDH/NADH	All atoms	0.72	35.9	52,006	104.9
LDH/NADH	Internally rigid	0.61	7.9	4,100	112.2

[a]The effect of using only a subset of ligand atoms to dock the crystallographic conformations of the ligands. The information loss is expressed as a reduced number of orientations. The effect of the information loss is presented as RMSD and score.

[b]Corrected for the electrostatic component of solvation using the method of Rashin (1990). The solvation corrections were 102.5 kcal/mol for MTX, 107.4 kcal/mol for dUMP, and 159.9 kcal/mol for NADH.

ligands. This addresses the ability of the algorithm to treat multiple conformations efficiently, and to recognize the "correct" conformation from the ensemble. Docking each ligand ensemble to its receptor (bound form) produced fits with RMSD values and interaction energies identical to those obtained from docking the crystallographic conformation using only the rigid atoms (Tables 2, 3). From among the 500 conformations present, the conformation from the ensemble that best fit the receptor was the crystallographic conformation. Docking MTX and NADH to the unbound form of DHFR and LDH, respectively, also resulted in the crystallographic conformation of the ligand receiving the best score. The crystallographic conformation of dUMP had a favorable interaction energy score with the unbound conformation of TS, but another similar (RMSD 0.159 Å) conformation had a better interaction energy. The time per conformation was decreased approximately 25-fold compared to the single conformer docking calculations (Tables 2, 3; Fig. 3).

3. Comparison of computer generated conformations to experimental structures

The next question was could the method find conformations resembling the experimental structure when beginning with the computationally generated ACD conformation of a ligand. This ACD conformation was dissimilar to the receptor-bound conformation. This involved identifying at least one conformation in the starting multi-conformer list that had a low RMSD from the experimental conformation and then being able to recognize this conformation once docked. Among the 500 low-energy conformations calculated for MTX, dUMP, and NADH, the lowest RMSDs from the experimental structures (PDB) were 1.19 Å for MTX, 0.50 Å for dUMP, and 1.67 Å for NADH. Generating 1,000 conformations of NADH produced a lowest RMS of 1.40 Å. The lowest RMSDs between the rotated residues of the protein ligands and the complexed crystal structures were 2.28 Å for BLIP and 1.31 Å for BPTI.

When only one conformation, that found in the ACD, of the three organic molecules was docked into the bound conformations of the receptors, no fits were found. The unbound conformation of TS could accommodate the ACD conformation of dUMP, but only with a relatively high RMSD and a poor energy score (RMS 3.67 Å, interaction energy −8.3 kcal/mol; Table 4). The unbound conformations of DHFR and LDH did not accommodate the conformations of MTX and NADH found in the ACD when docked as rigid bodies. Conversely, docking the conformational ensembles that began from the calculated ACD structures produced complexes that resembled those determined by X-ray crystallography (Table 3, Figs. 4–6). The RMSDs were slightly higher when dock-

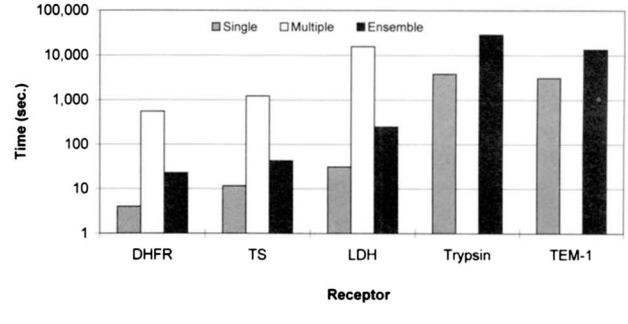

Fig. 3. Multiple vs. ensemble ligand docking. The graph illustrates the time requirements (in seconds, on a log scale) for the various docking methods with the five test systems. The first data set (hashed) represents the time required to dock a single rigid conformation of the ligand. The second data set (white) is the time required to dock 500 conformations of the ligands as individual molecules. This series of tests was completed only on the organic ligands due to the time requirements expected for the protein ligands. The last set of data (black) shows the time required to dock ensembles of conformations (500 conformations for the organic molecules and 417 and 162 for BPTI and BLIP, respectively).

Table 3. *ACD- vs. PDB-derived ligand structures for ensemble docking*[a]

Enzyme/ligand	Source	Confs.	RMS (Å)	Time (s)	Number of orientations	Score[b] (kcal/mol)
Complexed form of receptor						
DHFR/MTX	PDB	500	0.40	28.2	1,242	−26.8
DHFR/MTX	ACD	500	0.98	19.3	985	−13.5
TS/dUMP	PDB	500	0.41	42.8	6,314	−77.8
TS/dUMP	ACD	500	0.80	33.7	4,778	−89.3
LDH/NADH	PDB	500	0.54	125.6	11,015	57.4
LDH/NADH	ACD	1,000	1.88	393.4	11,565	84.1
LDH/NADH	ACD	500	12.86	151.3	11,565	117.1
Uncomplexed form of receptor						
DHFR/MTX	PDB	500	0.88	294.7	16,335	14.1
DHFR/MTX	ACD	500	1.31	105.4	17,027	−3.2
TS/dUMP	PDB	500	0.89	20.0	1,554	−17.1
TS/dUMP	ACD	500	0.70	13.5	1,422	−18.0
LDH/NADH	PDB	500	0.53	84.8	4,100	112.2
LDH/NADH	ACD	1,000	2.75	159.2	3,770	114.9
LDH/NADH	ACD	500	2.75	81.3	3,770	114.9
Protein-protein docking (h)						
Uncomplexed BPTI to		1	7.73	1.1	1,272,117	−17.1
uncomplexed trypsin		417	1.14	8.0	1,176,909	−29.5
Uncomplexed BLIP to		1	3.20	0.9	243,731	−5.6
TEM-1		162	2.90	3.7	533,521	−9

[a]The ability to generate a useful ensemble of conformers beginning with a computed structure that differs from the experimental or complexed structure. ACD ensembles begin with the CONCORD-derived conformation of the ligand found in the database. PDB ensembles begin with, and include, the experimentally determined conformation of the ligand. The difference in the number of orientations between single and multiple conformers in protein docking is a result of differences in the focusing used in the docking runs (Shoichet et al., 1992).

[b]Corrected for electrostatic component of solvation.

Table 4. *Database searching*[a]

Enzyme	Number of Confs.	Number of Cmpds.	Time (h)	Average number of orientations per ligand	RMS (Å)	Known ligand results Score[b] (kcal/mol)	Known ligand results Rank
Single conformation database							
Complexed DHFR	5,761	5,761	0.58	668	—	—	—
Uncomplexed DHFR	5,761	5,761	1.40	2,043	8.32	91.9[c]	4,834/5,761
Complexed TS	281	281	0.31	10,761	—	—	—
Uncomplexed TS	281	281	0.51	18,470	3.67	−8.3[d]	8/281
Multi conformation database							
Complexed DHFR	867,822	5,656	0.94	554	1.20	−12.5[c]	38/5,656
Uncomplexed DHFR	867,822	5,656	2.96	2,030	1.34	−7.4[c]	66/5,656
Complexed TS	88,487	263	0.27	5,187	0.77	−89.2[d]	1/263
Uncomplexed TS	88,487	263	0.18	1,035	2.71	−31.5[d]	2/263
Full multi conformation database							
Complexed DHFR	33,715,748	117,240	23.5	233	1.2	−12.5[c]	324/117,240
Complexed TS	33,715,748	117,240	80.9	4,416	0.77	−89.2[d]	80/117,240

[a]The ability of the new algorithm to rapidly screen a large number of compounds was evaluated. The rank of a known flexible ligand and the RMSD to the experimentally determined structure gives an indication of the quality of results. The multi conformation database searches consist of a subset of the single search molecules containing fewer than 12 rotatable bonds. Among the best scoring 60 compounds from the complexed DHFR search (5,656 compounds) were 27 pterins. These including 4-(N-(2,4-diamino-6-pteridinylmethyl)-N-methylamino benzoic acid, aminopterin, 2,4-diamino-6,7-diphenylpteridine, triamterene, methotrexate, 6-(o-tolyl)-2,4,7-triaminopteridine, L-biopterin, 6-formylpterin, and xanthopterin/2-amino-4,6-pteridinediol (Blaney et al., 1984).

[b]Corrected for electrostatic component of solvation.

[c]Score of MTX.

[d]Score of dUMP.

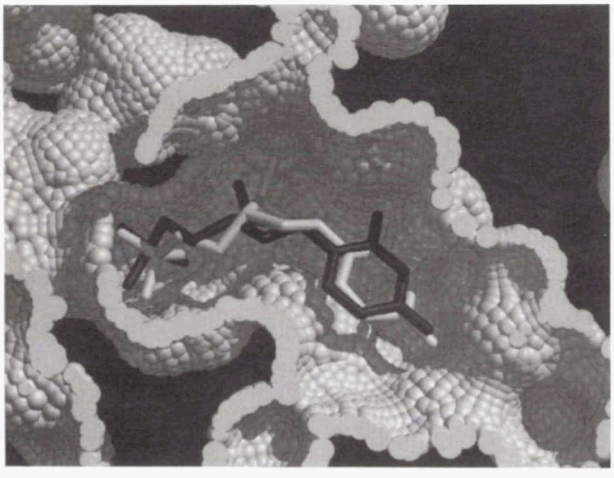

Fig. 4. The experimental (white) and the best fit from ensemble docking (black) orientations of dUMP in the binding site of TS, represented by its molecular surface (Connolly, 1983). This figure was generated with MidasPlus (Huang, 1989) as were all other binding site figures.

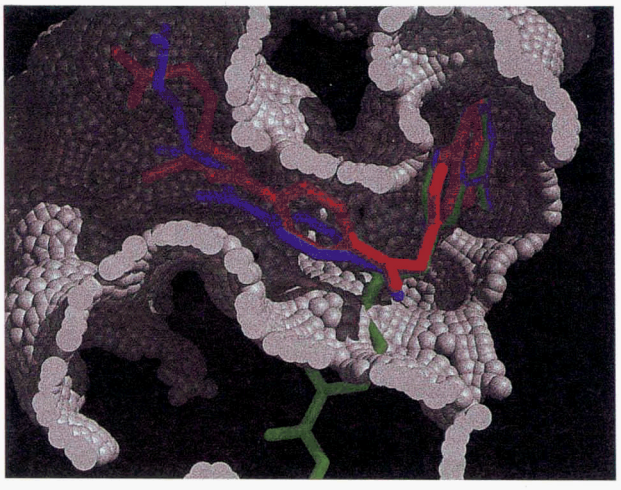

Fig. 6. The experimental (blue) and the best fit from ensemble docking (red) orientations of MTX in the binding site of DHFR (molecular surface). The conformation of MTX found in the ACD (green) has been manually superimposed on the experimental structure (alignment with the pteridine rings) to illustrate that the conformation found in the ACD cannot fit into the binding site.

ing the ACD-derived ensembles for MTX and dUMP compared to the PDB-derived ensembles. The electrostatic interaction energies were similar for the best scoring compounds from each ensemble. It was necessary to double the number of conformations in the ACD-derived ensemble of NADH to 1,000 to identify conformations that had favorable docking geometries. Docking this ACD-derived ensemble into both the bound and unbound conformations of LDH produced higher RMSDs (1.88 and 2.75 Å, respectively), and weaker interaction energies than did the crystallographic ensemble. For all calculations the time per conformation was decreased approximately 25-fold compared to the single conformer docking calculations (Tables 2, 3).

Docking ensembles of protein conformations led to lower RMSDs (α-carbons) and better electrostatic interaction scores than docking the single uncomplexed form of the ligands. The time difference between the single and ensemble docking is more complex to determine because focusing was used. Focusing allows for some favorable ligand orientations to be examined in greater detail. Therefore, it is possible that increasing the number of conformations will increase the number of orientations. Treating the calculation times at face value, the time per conformation was decreased approximately 50-fold (Table 3).

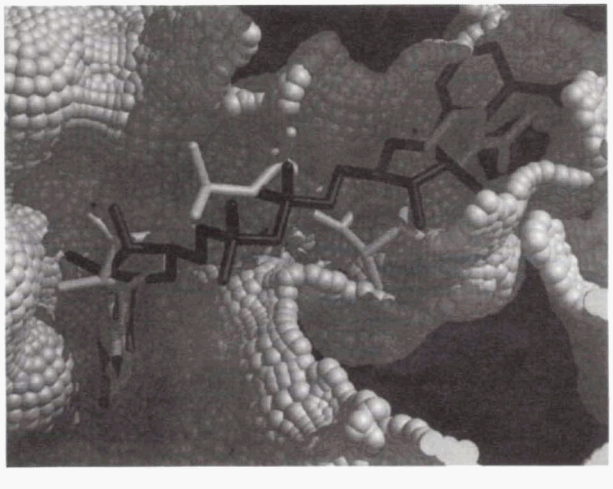

Fig. 5. The experimental (white) and the best fit from ensemble docking (black) orientations of NADH in the binding site of LDH (molecular surface).

4. Database screening

Finally, we tested the ability of the method to treat ligand flexibility in the docking of multi-compound databases into DHFR and TS. Docking the databases using only one conformation per molecule led to poor fits for known flexible ligands of the two enzymes. The orientations found were dissimilar to those seen in the experimental structures and had poor energy scores. Conversely, docking the databases with multiple conformations per molecule led to favorable fits for known flexible ligands. The orientations resembled the crystallographic structures and had favorable energy scores. This is illustrated by considering the ability to identify dUMP as a ligand for TS and MTX as a ligand for DHFR in the database screens (Table 4). When docking the single conformation form of each database into the respective receptors, the two known ligands, MTX and dUMP, ranked in the bottom half of compounds or did not fit the receptor. When docking multiple conformations (average of 152 per compound) of the "two-rings" database into the complexed form of DHFR, MTX ranked 38th of the 5,656 molecules searched. Its energy scores and orientations resembled those for the single ligand docking calculations (Tables 2, 3). Similar fits and scores were found for several analogs of MTX also known to inhibit DHFR (Table 4). Likewise, docking multiple conformations (average 337 per compound) from the ring-sugar-phosphate database of molecules dUMP was ranked first of the 263 molecules searched for TS. Its energy scores and orientations resembled those for the single ligand docking calculations

(Tables 2, 3). Similar fits and scores were found for several analogs of dUMP also known to bind TS. These analogs included the halide uridine monophosphates, deoxycytosine monophosphate, and thymidine monophosphate. Full database screens examined 34 million conformations from about 117,000 compounds. Known ligands ranked well and the run times varied between one and four days (Table 4).

Discussion

Perhaps the area where the ligand conformational ensemble method had the most dramatic impact was in the multi-compound, multi-conformation database screens. We, therefore, begin our discussion by considering the ability of the ligand conformational ensemble method to treat large databases of flexible molecules. We then return to consider aspects of the method that may interest the specialist.

Database screening

Docking a database of flexible ligands substantially altered the docking results. The small rigid molecules that predominated the rigid database screen (Fig. 7) were replaced in the flexible screen by larger, more flexible ligands that better complimented the binding site (Table 4, Fig. 8). In the database searches against DHFR and TS, docking ligand conformational ensembles was approximately 100-fold faster than docking each conformer of each ligand individually would have been (Table 4). When the "two-rings" database with only single conformations was docked into DHFR, MTX ranked poorly when docked to the ligand-bound form and did not fit into the unbound form of the enzyme (Fig. 6). When the "two-rings" database with multiple conformations was docked into DHFR, MTX ranked 38th of the 5,656 compounds searched. Analogs of MTX that are known to inhibit DHFR also ranked well and adopted reasonable conformations in the site (Table 4; Fig. 8). This was true for both the bound and unbound conformations of DHFR. The rank and score of dUMP docking into TS also improved by adding multiple conformations. The structures calculated from the searches resembled the experimental structures with RMSDs of less than 1.5 Å (Table 4). The exception was the orientation of dUMP found for the unbound conformation of TS. Here, key residues such as Arg23, Arg178′, and Arg179′, which bind to the phosphate of the nucleotide in the dUMP/TS complex, have swung out of the binding site. This reduces the energy of interaction of the "crystallographic" configuration and allows for other configurations to be seen.

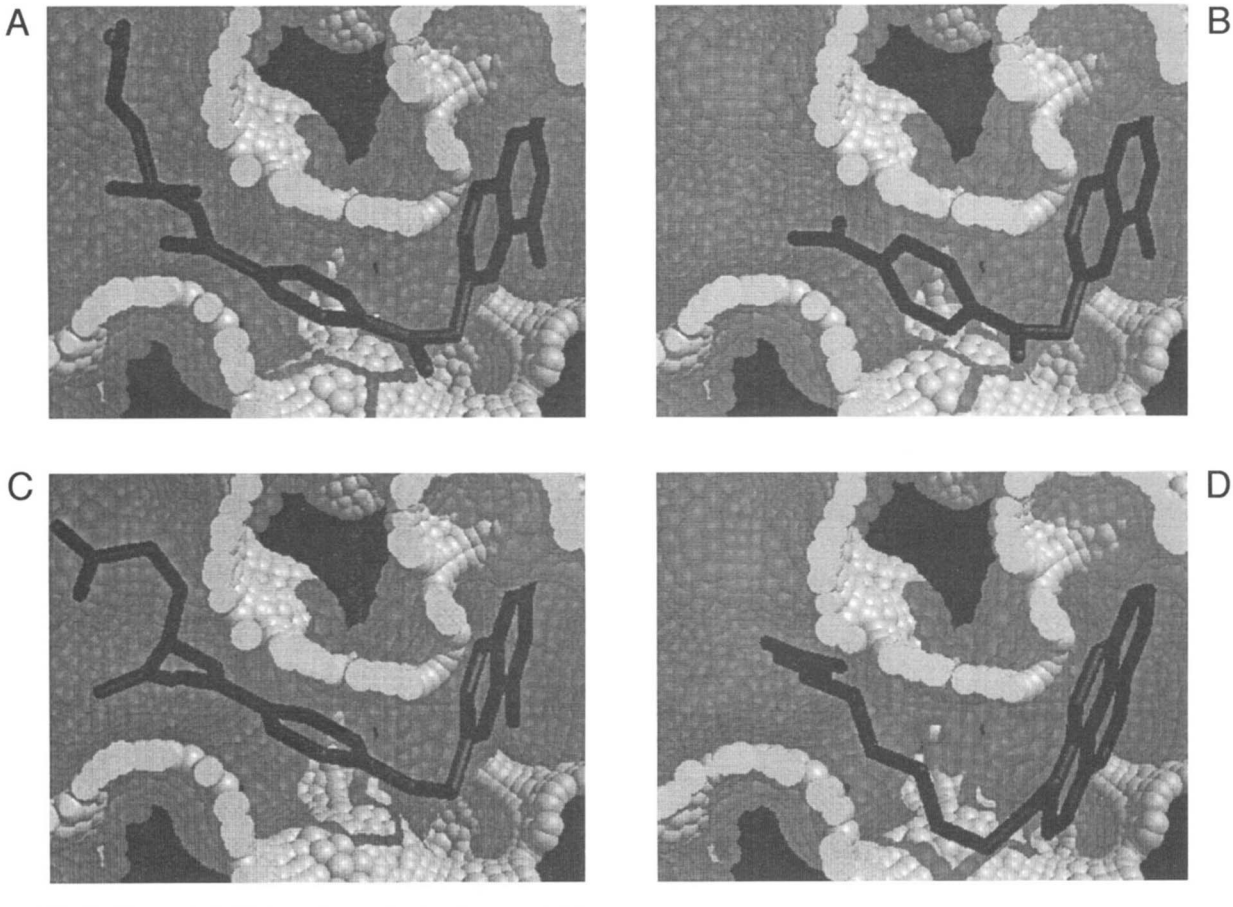

Fig. 7. Characteristic high scoring molecules from the DOCK search of the 5,656 compound, 867,822 conformation "two-rings" database docked into DHFR. **A:** Methotrexate (ranked 38th). **B:** 4-(n-(2,4-diamino-6-pteridinylmethyl)-n-methylamino (ranked 2nd). **C:** Aminopterin (ranked 4th). **D:** 10-(1-pyrene)-10-ketodecanoic-acid (ranked 1st). Flexible molecules predominated in this search.

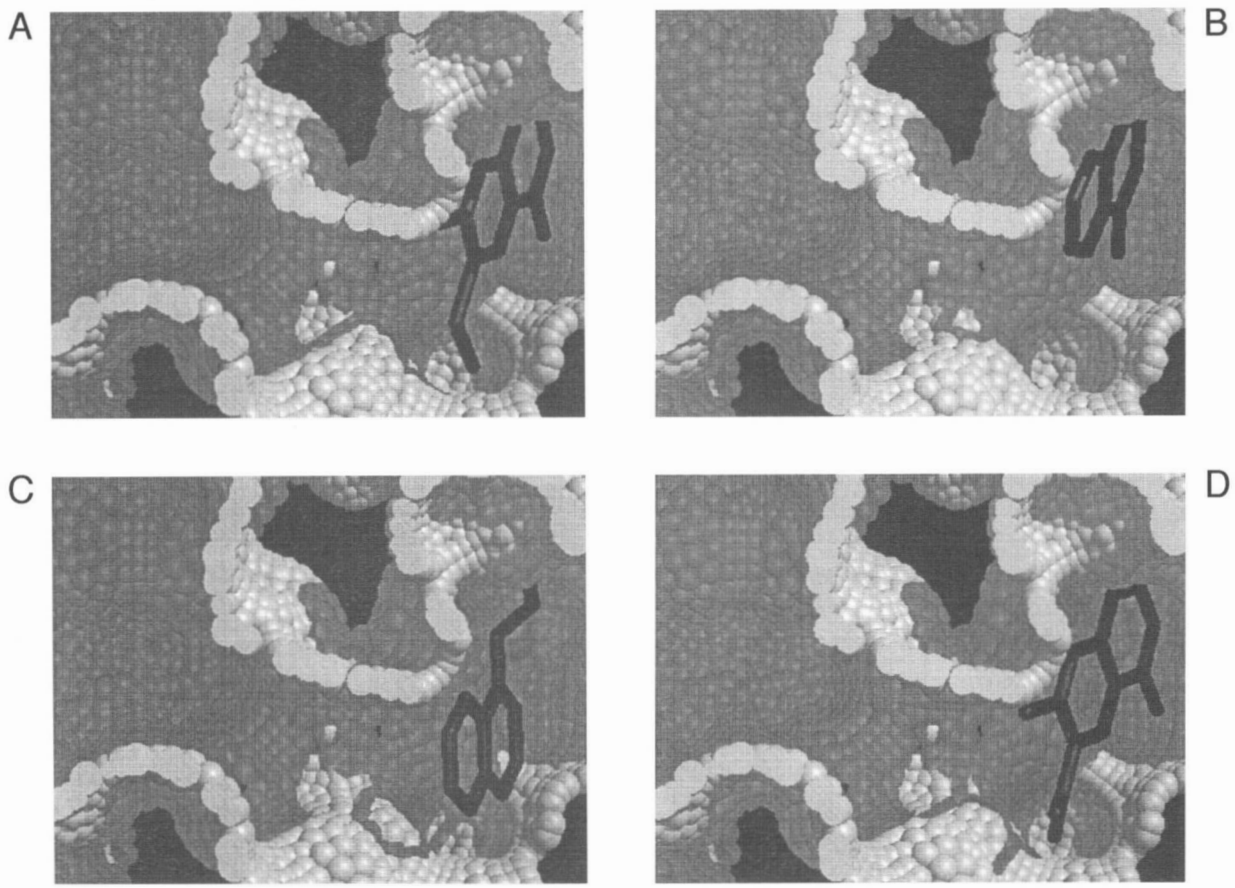

Fig. 8. Characteristic high scoring molecules from the DOCK search of the 5,761 compound, single conformation "two-rings" database docked into DHFR. **A:** 6-methylthio-2,4,7-triaminopteridine (ranked 1st). **B:** 2,4-diaminopteridine ranked (2nd). **C:** n-1-naphthylethylenediamine dihydrochloride (ranked 3rd). **D:** 2,4-diamino-7-methyl-6-pteridinyl methyl ketone (4th). Small rigid molecules predominated in this search.

The increased efficiency of the ensemble method allowed us to screen large databases of molecules rapidly. The method was able to screen over 117,000 compounds, made up of close to 34 million conformations, for complementarity to DHFR in one CPU day on an Indigo2 workstation. A similar sized database was screened against TS in four CPU days. In both cases, known flexible ligands were found in conformations that closely resembled those determined experimentally. These times are comparable to docking calculations on single conformation databases. In effect, this method makes docking of large databases of flexible molecules practical using relatively modest computational resources.

Docking time

We now turn to consider methodological details that determine the efficiency and the accuracy of the method. The ligand ensemble algorithm reduces redundancy by only evaluating the fit of a common rigid fragment once for a set of ligand conformations in the same orientation. In the absence of extensive energy minimization (Ewing & Kuntz, 1997), the orienting step is the rate limiting step of DOCK. Previously, the time (t) required to dock a series of conformations (n) of a given molecule was proportional to the number of conformations being docked (oriented and scored) (Equation 1). Each conformation was oriented (o) and then each atom (a) was scored (s). In the ligand-ensemble method, fewer atoms are used for orienting the ligands (leading to o'), and a single orientation is calculated for all n conformations. The number of atoms being scored in each orientation is smaller by the number of atoms in the rigid fragment (leading to a'); all n conformations are still scored. The time required to dock a series of conformations now scales with the number of conformations being scored instead of the number being oriented and scored (Equation 2).

$$t = n(o + as) \tag{1}$$

$$t = o' + na's. \tag{2}$$

Information loss and computational resources

Although docking ligands as ensembles of conformations is much more efficient than the "brute force" approach of one conformation at a time (Miller et al., 1994), there are several potential drawbacks. Information is lost when only the rigid fragments are used for generating an orientation, as opposed to using all atoms. The ensemble method will fail for ligands that lack internally rigid atoms. Relying on a fixed set of conformations for the ligands

prevents the method from tailoring conformations to particular sites, as can be done with genetic algorithms (Oshiro et al., 1995; Verkhivker et al., 1996; Jones et al., 1997) and fragment-based build-up methods (Leach & Kuntz, 1992; Rarey et al., 1996; Welch et al., 1996). We also impose a computer memory and disk space burden by having to calculate the conformations in advance.

In our test cases, most of these potential disadvantages were not realized. Although there is some loss of information on docking rigid fragments, it did not significantly diminish the quality of our results (Table 2). In two of three small-ligand cases, restricting ourselves to the 500 lowest energy ligand conformations still allowed us to find those that resembled the bound, crystallographic conformation of the ligand, RMSD less than 1.4 Å (Table 3; Figs. 4–6). Only in the case of NADH was a "correct" conformation not found in the best 500 structures generated by the SYBYL routine (Table 3). NADH is a highly flexible organic molecule with 13 rotatable bonds. The conformation closest to the experimental LDH-bound structure differed by 1.7 Å from the crystallographic conformation before docking. DOCK was unable to find a favorable fit from among this ensemble that resembled the experimental structure of NADH bound to LDH. When we expanded the number of conformations to 1,000 we were able to generate conformations closer to the experimental structure (RMS 1.4 Å) that fit better into the docking site (RMS 1.8 Å from the experimental conformation/orientation of NADH in the LDH site).

Even with 1,000 conformations, a reasonable solution was found only when we allowed for one close contact (ligand-receptor intersection) with the bound form of LDH and two close contacts with the unbound form of LDH. The fit of NADH in the LDH site is unusually close; the protein almost completely encloses the ligand. This makes this test case a difficult one for the ensemble method. Small errors in the placement of the rigid fragment can cause the flexible fragments, especially atoms at the distal end, to intersect the protein. This problem is significantly reduced in less constrained binding sites. Nevertheless, NADH probably establishes an upper bound to the degree of ligand flexibility that the conformer generation method can reasonably undertake.

Although the pre-calculation of conformations led to program arrays with large dimensions, our calculations were performed on SGI Indigo2 workstations with 64 Mb of memory; such a computer is common in the field (we were able to replicate some docking runs on a desktop PC with a Pentium Pro 200MHz CPU and 32Mb RAM). The disk space required for a 117,000 compound database with an average of 297 conformations per molecule was 16.6 GB when uncompressed and 2.6 GB when compressed (at no time does the entire database need to be uncompressed). This is a large, but not unreasonable, amount of disk space given equipment common at the workstation level. Perhaps the most onerous feature is the time required to pre-calculate conformations for the ligands. An average of 40 s is required to generate an ensemble of conformations for a typical organic molecule. For a database of 117,000 such molecules this amounts to 50 CPU days of calculation. While this time is significant, it is a one-time cost; once generated the same database can be used for different docking applications.

Protein-protein docking

By focusing on a few key residues, the method was applied to protein-protein docking. Allowing for three mobile side chains in BPTI and two in BLIP allowed us to improve on our previous docking simulations (Shoichet & Kuntz, 1991) and predictions (Shoichet & Kuntz, 1996; Strynadka et al., 1996a) with these molecules without significant increases in computational overhead. In the case of BPTI docking into trypsin, we previously had to truncate the side chains of Lys15, Arg17, and Arg39 to get reasonable fits. We also had to allow for contacts as close as 2 Å between the ligand and the receptor (Shoichet & Kuntz, 1991). By allowing for other conformations of these residues, we were able to include all ligand atoms in the docking calculation and use tighter limits on the close contacts allowed. Including conformational ensembles at these residues led to more reasonable fits in shorter calculation times than found previously. The ensembles also reduced the number of false-fits by increasing the complementarity stringency. The same can be said of docking BLIP into (β-lactamase) by allowing for conformational flexibility around Asp49 and Phe142, we found better fits in shorter times with increased stringency.

Bound vs. unbound forms of receptors

The ability to find reasonable ligand orientations in the unbound form of the receptor suggests that this method can be applied to situations where the receptor conformation may change slightly upon complex formation. It must be admitted that we chose systems where the conformational changes were relatively small, certainly compared to enzymes that have large domain movements. Still, the conformational changes were not completely trivial. Key residues in TS change conformation on nucleotide binding, leading to a binding site that better complements the phosphate moiety than in the unbound form of the receptor. This explains both the lower score of dUMP in the unbound versus the bound TS searches and its higher RMSD from the crystallographic configuration. In the unbound form of DHFR the Nζ of Lys32 is rotated into the binding site. Upon superposition of the bound and unbound structures, this nitrogen atom is approximately 1.2 Å away from the orientation of MTX in the bound enzyme. His28, which forms a hydrogen bond with a terminal carboxylate of MTX, is in a significantly different conformation in the unbound conformation of DHFR. These are potential difficulties that the docking program overcomes through orientation- and conformation-based accommodations on the part of the ligand.

General features and future directions

Several other caveats should be mentioned. We use different force fields to evaluate the internal energies of the ligands and their complementarity to the receptor. Indeed, we make no effort to integrate the two energies except to insist that the conformations that we choose are among the lowest energy n conformations, where n might typically be between 500 and 2,000. As with our earlier docking calculations (Kuntz et al., 1982; DesJarlais et al., 1988; Meng et al., 1992; Shoichet & Kuntz, 1993), the energies we report here leave out many terms thought to be important in determining ligand binding affinities. These include receptor desolvation, receptor conformational change, and changes in translational, rotational and vibrational partition functions, among others. Although the scoring function properly excludes many molecules based on steric fit or on desolvation energies, and although it properly highlights known ligands, the absolute energies that we determine are unreliable. At best, they are useful for ranking the relative energies of a set of ligands for potential binding to a common receptor site.

Notwithstanding the limitations that we have discussed, modeling ligand flexibility using conformational ensembles presents several important opportunities. The method is rapid enough to search a database of 117,000 molecules with close to 34 million conformations in one to four CPU days. It is accurate enough to identify known high-affinity ligands from this search and place them in a binding site appropriately. The algorithm is not restricted to particular torsional move-sets; since the conformational ensembles are pre-calculated, there is no pre-set constraint on what conformations are included. Neither is there a limit on the sorts of molecules that can be treated—macromolecular ligands are no more difficult than organic molecules. Although the program can only treat ensembles of a given size (the maximum used here was 2,000 conformations per ensemble), more conformations can easily be included by multiplying the ensembles as needed.

The ligand ensemble method introduces several database organization principles that may be broadly helpful. For instance, many molecules in any given large database resemble one another closely, differing by one or two atoms, but being otherwise the same. It should be possible to reduce this chemical redundancy in a multicompound database in the same manner that we reduced conformational redundancy. This would have the dual advantage of speeding, again, the docking of a database of molecules against a receptor and would allow for more diversity in the results of such a calculation. Considerable redundancy remains in even our current conformational ensembles. Thus, most conformations may be grouped into similar families within which many atom positions are identical. Organizing the conformations into hierarchies of decreasing similarity would further speed the algorithm. Most docking algorithms match chemical and structural features of molecules, and a hierarchical organization of these features should be generally useful.

Acknowledgments

We are grateful to Scott Weston for his assistance and enlightening conversations, and to E. M. Silinsky and members of our group for critically reading this manuscript. This work was partly supported by a grant from the Animal Alternatives Program of Procter and Gamble, Inc (to BKS) and by NIH training grant T32-GM08382 (to DML). We thank MDL (San Leandro, CA) for the Isis Program and the Available Chemicals Directory. The authors are members of the Drug Discovery Program of Northwestern University.

References

Abad-Zapatero C, Griffith JP, Sussman JL, Rossmann MG. 1987. Refined crystal structure of dogfish M4 apo-lactate dehydrogenase. *J Mol Biol 198*:445–467.

Bartlett PA, Shea GT, Telfer ST, Waterman S. 1989. CAVEAT: A program to facilitate the structure-derived design of biologically active molecules. In: Roberts SM, ed. *Molecular recognition*. London: Royal Society of Chemistry. pp 182–196.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structure. *J Mol Biol 112*:535–542.

Blaney JM, Hansch C, Silipo C, Vittoria A. 1984. Structure-activity relationships of dihydrofolate reductase inhibitors. *Chem Rev 84*:333–407.

Bodian DL, Yamasaki RB, Buswell RL, Stearns JF, White JM, Kuntz ID. 1993. Inhibition of the fusion-inducing conformational change of the influenza hemagglutinin by bensoquinones and hydroquinones. *Biochemistry 32*:2967–2978.

Bohm HJ. 1992. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J Comput-Aided Mol Des 6*:61–78.

Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. 1982. Crystal structures of *E. coli* and *L. casei* DHFR refined to 1.7 Å resolution. 1: General features and binding of Methotrexate. *J Biol Chem 257*:13650–13662.

Bystroff C, Oatley SJ, Kraut J. 1990. Crystal structures of *Escherichia coli* dihydrofolate eductase. The NADP holoenzyme complex and the folate/NADP ternary complex. Substrate binding and a model for the transition state. *Biochemistry 29*:3263–3277.

Caflisch A, Niederer P, Anliker M. 1992. Monte Carlo docking of oligopeptides to proteins. *Proteins 13*:223–230.

Connolly ML. 1983. Analytical molecular surface calculations. *J Appl Crystallogr 16*:548–558.

Connolly ML. 1985. Shape complementarity at the hemoglobin $\alpha 1\beta 1$ subunit interface. *Biopolymers 25*:1229–1247.

DesJarlais RL, Seibel GL, Kuntz ID, Montellano POd, Furth PS, Alvarez JC, DeCamp DL, Babé LM, Craik CS. 1990. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc Natl Acad Sci USA 87*:6644–6648.

DesJarlais RL, Sheridan RP, Dixon JS, Kuntz ID, Venkataraghavan R. 1986. Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem 29*:2149–2153.

DesJarlais R, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID, Venkataraghavan R. 1988. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem 31*:722–729.

Dunn CR, Wilks HM, Halsall DJ, Atkinson T, Clarke AR, Muirhead H, Holbrook J. 1991. Design and synthesis of new enzymes based on the lactate dehydrogenase framework. *Phil Trans Royal Soc B 332*:177–184.

Ewing TJ, Kuntz ID. 1997. Critical evaluation of search algorithms for automated molecular docking and database screening. *J Comput Chem 18*:1175–1189.

Ferrin TE, Huang CC, Jarvis LE, Langridge R. 1988. The MIDAS display system. *J Mol Graph 6*:13–27.

Gilson MK, Honig BH. 1987. Calculation of electrostatic potentials in an enzyme active site. *Nature 330*:84–86.

Goodsell DS, Morris GM, Olson AJ. 1996. Automated docking of flexible ligands: Applications of AutoDock. *J Mol Recogn 9*:1–5.

Goodsell DS, Olson AJ. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins 8*:195–202.

Guner OF, Hughes DW, Dumont LM. 1991. An integrated approach to three dimensional information management with MACCS-3D. *J Chem Inf Comput Sci 31*:408–414.

Ho CMW, Marshall GR. 1993. Foundation: A program to retrieve all possible structures containing a user-defined minimum number of matching elements from three-dimensional databases. *J Comput-Aided Mol Des 7*:3–22.

Huang C. 1989. *Computer graphics* [Ph.D. thesis]. University of California.

Janin J. 1995. Protein-protein recognition. *Prog Biophys Mol Biol 64*:145–166.

Jiang F, Kim SH. 1991. Soft docking: Matching of molecular surface cubes. *J Mol Biol 201*:79–102.

Jones G, Willett P, Glen RC, Leach AR, Taylor R. 1997. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol 267*:727–748.

Kuhl FS, Crippen GM, Friesen DK. 1984. A combinatorial algorithm for calculating ligand binding. *J Comput Chem 5*:24–34.

Kuntz ID. 1992. Structure-based strategies for drug design and discovery. *Science 257*:1078–1082.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. 1982. A geometric approach to macromolecule-ligand interactions. *J Mol Biol 161*:269–288.

Kuntz ID, Meng EC, Shoichet BK. 1994. Structure-based molecular design. *Accts Chem Res 27*:117–123.

Leach AR, Kuntz ID. 1992. Conformational analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem 13*:730–748.

Lewis RA, Roe DC, Huang C, Ferrin TE, Langridge R, Kuntz ID. 1992. Automated site-directed drug design using molecular lattices. *J Mol Graph 10*:66–78.

Malby RL, Tulip WR, Harley VR, McKimm-Breschkin JL, Laver WG, Webster RG, Coleman PM. 1994. The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. *Structure 2*:733–746.

Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B 39*:480–490.

Meng EC, Shoichet B, Kuntz ID. 1992. Automated docking with grid-based energy evaluation. *J Comput Chem 13*:505–524.

Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. 1994. FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput-Aided Mol Des 8*:153–174.

Miranker A, Karplus M. 1991. Functionality maps of binding sites: A multicopy simultaneous search method. *Proteins 11*:29–34.

Miranker A, Karplus M. 1995. An automated method for dynamic ligand design. *Proteins 23*:472–490.

Moon JB, Howe WJ. 1991. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *Proteins Struct Funct Genet 11*:314–328.

Olson AJ, Goodsell DS. 1993. Soluble proteins: Size, shape and function. *Trends Biochem Sci 18*:65–68.

Oshiro CM, Kuntz ID, Dixon JS. 1995. Flexible ligand docking using a genetic algorithm. *J Comput-Aided Mol Des 9*:113–130.

Perry KM, Fouman EB, Finer-Moore JS, Montfort WR, Maley GF, Maley F, Stroud RM. 1990. Plastic adaptation toward mutations in proteins: Structural comparison of thymidylate synthase. *Proteins Struct Funct Genet 8*:315.

Rarey M, Kramer B, Lengauer T. 1996. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol 261*:470–489.

Rashin AA. 1990. Hydration phenomena, classical electrostatics and the boundary element method. *J Phys Chem 94*:1725–1733.

Ring CS, Sun E, McKerrow JH, Lee GK, Rosenthal PJ, Kuntz ID, Cohen FE. 1993. Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc Natl Acad Sci USA 90*:3583–3587.

Rose GD. 1978. Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature 272*:586–590.

Rusinko A, Skell JM, Balducci R, Pearlman RS. 1987. *CONCORD manual and description*. St. Louis, MO: Tripos Associates.

Rutenber E, Fauman EB, Keenan RJ, Fong S, Furth PS, Ortiz de Montellano PR, Meng E, Kuntz ID, DeCamp DL, Salto R, Rose JR, Craik CS, Stroud RM. 1993. Structure of a non-peptide inhibitor complexed with HIV-1 protease. *J Biol Chem 268*:15343–15346.

Sandak B, Wolfson HJ, Nussinov R. 1995. An automated computer vision and robotics-based technique for 3-D flexible biomolecular docking and matching. *Comput Appl Biosci 11*:87–89.

Shoichet B, Bodian DL, Kuntz ID. 1992. Molecular docking using sphere descriptors. *J Comput Chem 13*:380–397.

Shoichet B, Kuntz ID. 1991. Protein docking and complementarity. *J Mol Biol 221*:327–346.

Shoichet BK. 1996. Docking ligands to proteins. In: Sternberg MJE, ed. *Protein structure prediction*. New York: Oxford University Press. pp 263–290.

Shoichet BK, Kuntz ID. 1993. Matching chemistry and shape in molecular docking. *Protein Eng 6*:723–732.

Shoichet BK, Kuntz ID. 1996. Predicting the structures of protein complexes: A step in the right direction. *Chem Biol 3*:151–156.

Shoichet BK, Perry KM, Santi DV, Stroud RM, Kuntz ID. 1993. Structure-based discovery of inhibitors of thymidylate synthase. *Science 259*:1445–1450.

Stoddard BL, Koshland DE. 1992. Prediction of the structure of a receptor-protein complex using a binary docking method. *Nature 358*:774–776.

Stout TJ, Stroud RM. 1996. The complex of the anti-cancer therapeutic, BW1843U89, with thymidylate synthase at 2.0 Å resolution: Implications for a new mode of inhibition. *Structure 5*:67–77.

Strynadka NC, Jensen SE, Johns K, Blanchard H, Page M, Matagne A, Frere JM, James MN. 1994. Structural and kinetic characterization of a beta-lactamase-inhibitor protein. *Nature 368*:657–660.

Strynadka NCJ, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abazyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MNG. 1996a. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to tem-1 beta-lactamase. *Nature Struct Biol 3*:233–239.

Strynadka NCJ, Jensen SE, Alzari PM, James MNG. 1996b. A potent new mode of β-lactamase inhibition revealed by the 1.7 Å X-ray crystallographic structure of the TEM-1-β-lactamase complex. *Nature Struct Biol 3*:290–297.

Verkhivker GM, Rejto PA, Gehlhaar DK, Freer ST. 1996. Exploring the energy landscapes of molecular recognition by a genetic algorithm: Analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes. *Proteins 25*:342–353.

Walter J, Steigemann W, Singh TP, Bartunik H, Bode W, Huber R. 1982. On the disordered activation domain in trypsinogen. *Acta Crystallog B 38*:1462.

Wasserman ZR, Hodge CN. 1996. Fitting an inhibitor into the active site of thermolysin: A molecular dynamics study. *Proteins 24*:227–237.

Welch W, Ruppert J, Jain AN. 1996. Hammerhead: Fast, fully automated docking of flexible ligands into protein binding sites. *Chem Biol 3*:449–462.

Wodak SJ, Janin J. 1978. Docking BPTI and trypsin. *J Mol Biol 124*:323–342.