PROTEINS: Structure, Function, and Genetics 13:206–222 (1992)

A Multiple-Start Monte Carlo Docking Method

Trevor N. Hart and Randy J. Read
Department of Medical Microbiology and Infectious Diseases, University of Alberta, Edmonton, Alberta, Canada, T6G 2H7

ABSTRACT **We present a method to search for possible binding modes of molecular fragments at a specific site of a potential drug target of known structure. Our method is based on a Monte Carlo (MC) algorithm applied to the translational and rotational degrees of freedom of the probe fragment. Starting from a randomly generated initial configuration, favorable binding modes are generated using a two-step process. An MC run is first performed in which the energy in the Metropolis algorithm is substituted by a score function that measures the average distance of the probe to the target surface. This has the effect of making buried probes move toward the target surface and also allows enhanced sampling of deep pockets. In a second MC run, a pairwise atom potential function is used, and the temperature parameter is slowly lowered during the run (Simulated Annealing). We repeat this procedure starting from a large number of different randomly generated initial configurations in order to find all energetically favorable docking modes in a specified region around the target. We test this method using two inhibitor-receptor systems: *Streptomyces griseus* proteinase B in complex with the third domain of the ovomucoid inhibitor from turkey, and dihydrofolate reductase from *E. coli* in complex with methotrexate. The method could consistently reproduce the complex found in the crystal structure searching from random initial positions in cubes ranging from 25 Å to 50 Å about the binding site. In the case of SGPB, we were also successful in docking to the native structure. In addition, we were successful in docking small probes in a search that included the entire protein surface.**

Key words: molecular docking, Monte Carlo, simulated annealing, rational drug-design, dihydrofolate reductase, proteinase inhibitors

INTRODUCTION

One of the striking achievements of structural biochemistry is the potential to directly visualize and understand the interaction of therapeutic agents with specific macromolecules to which they bind. The primary source of this information is the structures of drug-receptor complexes that have been solved to high resolution by X-ray crystallography. In addition to giving a detailed picture of binding, this information makes it possible to predict modifications to the ligand to improve binding affinity. This could complement the more established rational drug design methods, based on the structure of the drug and analogues, such as QSAR (Quantitative Structure-Activity Relationships). However, a crystal structure of a *potential* drug target holds, in principle, the information necessary to predict the structure of a ligand—with high binding affinity—without reference to the binding mode of a known ligand. Since an existing binding mode is not relied upon, the predictions would essentially be generating ligands "from scratch." The value of such an ab initio drug design is clear, since one of the major stumbling blocks in drug research is finding a lead compound from which a suitable drug may be derived.

At present, drug design based on crystal structures relies on chemical intuition and the use of a high-speed graphics computer [1–3]. This technique is useful in finding small changes to the ligand that will improve binding affinity or specificity, particularly if a crystal structure for the drug-receptor complex is available. It is extremely difficult to design new chemical structures based on visual inspection of the chemical properties (hydrogen bonding sites, hydrophobic regions, steric shape, etc.) of the proposed drug target site; however, see Beddell.[4] The number of ways the numerous types of chemical structures could potentially fit in the site is staggering. In order to maximize the chances of success, we need to consider as many chemical structures and potential binding modes as possible. This problem is best approached using computational methods.

In general, then, the basic problem is to determine the most thermodynamically favorable mode(s) of binding between a target macromolecule of known structure and a given probe whose structure is ei-

Received May 9, 1991; revision accepted July 22, 1991.
Address reprint requests to Dr. Trevor Hart, Department of Medical Microbiology and Infectious Diseases, 1-41 Medical Sciences Building, University of Alberta, Edmonton, Canada T6G 2H7.

ther fixed or to be determined (in the case of flexible probes). In the literature, this is known as the Docking Problem. In the case of a rigid (or approximately rigid) probe, this is essentially a 6-dimensional search over translational and rotational degrees of freedom of the probe (relative to the target), in some selected region around the potential binding site; in general, when N internal degrees of freedom (for probe and/or target) are included, we have a $(6 + N)$-dimensional search. Since solvated thermodynamics calculations are extremely expensive in terms of computer time, the optimal binding modes are sought in terms of a more easily calculated scoring function, typically involving steric complementarity and distance geometry, such as Kuntz,[5,6] Bacon and Moult,[7] and Yue,[8] or energetics using pair-potentials, such as Goodsell and Olson.[9] To perform a docking, the target site must be identified along with a region around it within which the search will take place. An initial position and orientation for the probe is selected, either randomly generated or user selected, and a procedure is then invoked to optimize the score function of the probe. The best scores might then undergo further (more computationally expensive) screening to eliminate physically unrealistic situations.

In this work, we introduce a new docking method based on *Monte Carlo* (MC) techniques. The MC method explores the thermodynamically accessible states of a system by generating small random changes in the system configuration that are either *accepted* or *rejected* according to a rule, the Metropolis algorithm. Acceptance is a function of the change in energy δE and a parameter T, which may be physically interpreted as a temperature. Whereas the algorithm favors decreases in energy, increases are possible: for high values of T increases are almost as likely; for small values of T increases are very infrequent. A common technique used in MC involves changing the parameter T during the run: T starts at a high value, so that small energy barriers can be jumped and the configuration can move between local minima; then T is gradually lowered as the run progresses and the configuration settles down into a local minimum. This method is known as *Simulated Annealing*. Since the method can circumvent energy barriers, it is particularly useful when there are many local minima from which a global minimum is sought. Here we use the names "Monte Carlo" and "Simulated Annealing" almost synonymously.

We apply the MC technique to the docking of a rigid molecular fragment with a rigid macromolecule (usually a protein) of known three-dimensional structure. For each run, we apply two separate procedures: first, using a "steric conflict" scoring function, which effectively eliminates bad steric contacts, and then performing a more standard MC procedure with a pairwise empirical potential function. The data for any given docking are generated using a large number of independent runs, with each run starting from a randomly generated initial position and orientation. The initial positions, as well as the subsequent random orientations generated during the run, are constrained to fall inside a cube specified by the user (for our tests we chose cube sizes from 25 Å to 60 Å, along a cube side). We refer to the technique of using a large number of independent runs as *Multiple-Start Monte Carlo* (MSMC).

Since the ultimate goal of docking is to design drugs, it is important to consider the various strategies that might be used to approach the larger problem of predicting effective inhibitors from a target structure. One possibility is to perform a search of a large structure database, such as the Cambridge Database,[10] and test the ability for each probe to bind to the target structure. This strategy has been pursued in the work of Kuntz.[11] A disadvantage of this approach is that the possible ligands are restricted to occur within a well-defined set. An alternative is to use a fragment assembly strategy for the design of new ligands, in which members of a library of molecular fragments are docked independently to a potential drug target and then assembled to form new molecules. In essence, this approach has been initiated by Goodford,[12] with fragments that are essentially functional groups represented as spherical probes. One advantage of this approach is that the search for ligands is not restricted to a fixed molecular database, although care must be taken to screen out possibilities that are chemically unreasonable.

Ultimately, we plan to adopt this latter strategy toward drug design, in which molecules are to be assembled from independently docked, rigid fragments taken from a library, as opposed to the spherical probe approach of Goodford. In docking a fragment, we must consider that it need not be in the global energy minimum in order for it, along with other fragments, to build a good inhibitor. Indeed, to take only the global energy minimum would greatly restrict the number of possible molecules one could build. Therefore, in docking fragments for this purpose, it is necessary to find all of the *good* binding modes, not just the *best* one. By using a large number of independent runs, our MSMC approach is aimed specifically at compiling a list of the energetically favorable modes of binding for a fragment.

We have tested our method on two inhibitor-receptor systems: *Streptomyces griseus* proteinase B in complex with the third domain of the ovomucoid inhibitor from the turkey,[13] and dihydrofolate reductase from *E. coli* in complex with methotrexate.[14] Our objective was to dock the inhibitor, as a rigid fragment, to the binding site. In the first case, we used probes of various sizes, from small dipeptide analogues up to the whole inhibitor. For each probe,

the crystal structure was close to an energy minimum found by the program. In the second case, we used probes for two distinct fragments of methotrexate as well as the whole inhibitor. We found that the crystal structure was well represented among the low energy local minima and that the correspondence between the two fragments of the inhibitor was sufficiently good that we could, in principle, design methotrexate from these fragments.

Although our results are promising, we wish to emphasize that the development of docking methods, generally, is at an early stage. The objective of docking—to search molecular configuration space for favorable binding modes—is computationally very intensive and requires the use of simplified models for the molecular interactions. Once search methods are proven, they may then be applied with more realistic molecular interactions, and results with greater predictive power may be obtained. Hence important but computationally expensive effects such as solvent screening, hydrophobic interactions, and molecular flexibility are not included in the energy potentials in our tests. We are attempting to evaluate the effectiveness of our method for searching a molecular surface for the favorable binding modes—our results show that it is an effective and efficient method. Since these results are obtained without extending the limits of technology (a complete run generally taking less than 1 cpu-day using one processor on a Silicon Graphics Iris 4D/120), more sophisticated energy calculations might realistically be performed on the fastest computer hardware.

MATERIALS AND METHODS

The MC method was originally devised to calculate thermodynamic properties of many-body systems. The theory of statistical thermodynamics states that any disordered system in equilibrium can be represented as a probability distribution in the set of accessible states; the MC method is a computer algorithm to generate a finite representation of this distribution. This algorithm is designed on the principle of *importance sampling*:[15] states are randomly generated with frequency that is directly related to their statistical significance in the thermodynamic system. In thermodynamic averaging, no statistical weight (i.e., Boltzmann distribution) for states is required, since this is reflected directly in the frequency with which the states are generated in the simulation. In MC, this means that the lowest energy states (most thermodynamically probable states, for sufficiently low temperature) are by far the ones most frequently generated. The Metropolis algorithm represents the Boltzmann distribution using a very simple scheme: given a current state, a new state is randomly generated (usually close, in some sense, to the current one) and accepted as the next state of the system if either: (1) it is of lower energy than the current state, or (2) a (uniformly distributed) random number $r \epsilon [0,1]$ satisfies $r < \exp(-\delta E/kT)$; otherwise it is rejected and the current state is kept as the new one. Case (2) means that increases in energy are accepted with frequency $\exp(-\delta E/kT)$. It is simple to show that in the limit of a large number of states, the frequency with which a state occurs is proportional to the Boltzmann probability of that state.[16]

Our interest in the MC method is in finding energy minima in a complex system. The technique of *Simulated Annealing* (SA), a variation of the MC method, has been applied to numerous minimization problems,[9,17–19] some of which have been intractable by other methods. In SA, the temperature parameter T is varied during the course of the run (k is the Boltzmann constant, expressed in appropriate units), according to a schedule that may depend on the configuration history. Experience shows that, generally, the most efficient method of relaxation is to start at a high temperature, where large energy barriers can easily be jumped, and to decrease the temperature gradually during the course of the run, down to the value where the calculation is to be performed. SA works best for problems involving many local minima, in which standard minimization methods tend to become trapped. Because of the complexity of molecular surfaces, the method is well suited to the docking problem.

Our implementation of SA is the program BOXSEARCH, written in the 'C' programming language. The following is a basic sketch of the program. First, coordinates for the target and probe are read in from files in protein data bank (PDB) format,[20] and the atoms are organized into neutral or integral charge groups according to a residue library (partial charges and van der Waals parameters are assigned at this time). Since a cutoff is used for the energy calculation (typically 8 Å), the groups from the target are collected into lists found within distinct (8 Å) cubes, in order to reduce the time of the energy calculation. A distance-to-surface grid (see below) is then read in to provide the data for the "floating" algorithm and to define the search cube within which the probes will be constrained. Following this, a steric contact grid is read in; this serves to screen out atom positions that have bad steric contacts with the target, eliminating the need to proceed with computationally expensive energy calculations. Finally, the annealing schedule and other control data are read in.

The main loop is then invoked. This starts by generating a completely random position and orientation for the probe within the search cube, followed by a floating run to first minimize the bad contacts to within a given tolerance of the floating score function. In essence, probes that interpenetrate with the target are allowed to move (or "float") to the surface in this step. If the score falls below the tol-

erance within a given number of steps, it is then put to a simulated annealing run; otherwise it is rejected and another initial random state is generated. The floating algorithm is very inexpensive to run, and very rarely (typically 1 in 1,000 for the parameters we used) fails to produce an acceptable state. The energy-driven MC involves a fixed number of Metropolis steps at a fixed number of temperatures, after which the state's final energy is recorded. If the energy is below a specified value, it is subjected to a further 200 runs at $kT = 10^{-4}$ kcal/mole, after which the atomic coordinates are printed into an output file in PDB format.

The floating algorithm uses a precalculated grid, which we call the *floating grid,* to efficiently determine if any given point is inside the target protein, and if so, determine its distance to the nearest point on the target's surface. This grid is calculated by the program BOXGRID, which reads in the target coordinates and user-defined parameters (such as grid stepsize, box position, and dimensions) and outputs a binary file of values that represent the distance from that point to the nearest point that doesn't overlap with the target within a given distance, 3 Å in our runs. For points that are outside the target protein surface (i.e., at least 3 Å from any target atom) the grid value is zero. Essentially, the grid represents the distance to the target surface. The algorithm is straightforward and the implementation is relatively inexpensive to run, particularly since the grid is calculated only once for a given region being searched.

A floating run is an MC run, in which an "average overlap" score function replaces the usual energy function. The score function is evaluated by finding the closest grid point for each atom in the probe, summing over all probe atoms, and dividing by the total number of atoms in the probe. Atoms that fall outside the search cube (and therefore outside the grid) are given a maximum possible score value (the diagonal distance of the cube), which effectively restricts the probe to remain within the cube. We took our score value tolerance to be between 0.2 to 0.7, which means that the floating algorithm will stop when the probe has a score of less this value. For instance, we would have a score of 0.2 if 20% of the atoms were 1 Å below the target surface, still a very close contact. However, these contacts generally did not pose a problem for the energy-driven MC, since we found that typically less than 10% of the final states had positive energies. In contrast, tests run without the floating procedure had great difficulty with states that overlapped with the target, with more than 50% of the final states having positive energy. As a side benefit, starting states that are buried in the target near deep binding pockets may tend to float into the pocket. This could enhance the sampling of pockets that, without conformational changes, might even be inaccessible from the surface. In essence, docking is allowed to occur from inside the target.

In our implementation of MSMC, we have sought a tradeoff between using multiple starts and extensive MC runs. Since our goal is to randomly "map out" configuration space, enough multiple starts need to be set up to give statistical weight to the favorable modes that are found. For this reason, we chose to depart from the more standard simulated annealing schedules in which the system is allowed to equilibrate at each temperature step before proceeding to the next one. Instead, we ran a fixed number of MC trials at each temperature, regardless of the number that were accepted or rejected. This allowed us to control the amount of computer time for the program to finish while changing the total number of starts.

In order to deal with the large amount of data produced by a run and to sort out duplicate dockings, we perform a simple cluster analysis on the data produced by a run. This is done by the program CLUSTER, which takes the output from a BOXSEARCH run and sorts the dockings by energy. It then starts with the lowest energy state and counts all others in the list that are within 2 Å RMS, removing them from the list. These dockings are regarded as being equivalent, and only the lowest energy representative is kept. This procedure is then repeated with the remaining list until all states have been placed in a cluster. The coordinates of the representatives are then printed into a file, along with their energies, placement in the list and the number of equivalent states found. When ranking a state, we will refer to its place in the list (from the lowest energy) relative to the total number in the list.

The most costly part of the BOXSEARCH program is undoubtedly the energy calculation in the MC runs. However, the program speed is significantly improved by using a steric contact grid, which scores on a 0.5 Å grid the inverse square distance to the nearest heavy atom in the target. The grid dimensions were taken to be identical to the floating grid. The contact grid is generated with the program CONTGRID. Whenever an energy calculation is required, the grid point nearest to each heavy atom in the probe is found. If the steric grid value S of any of these fall above $0.3/A^2$ (and hence is within approximately 1.8 Å from a heavy atom in the target), the energy is given a value $E = 10^{10}(1 + S)$ kcal/mole, and the normal energy calculation is not performed. For consistency, any normal energy calculations that exceed the value $E_{max} = 10^{10}$ are cut off at that value. For atoms outside the grid boundaries, a maximum score value is assigned. Effectively, the sides of the search box are impenetrable and the probe is constrained to stay within the box.

We chose to use simple pairwise potential functions using extended atoms, with standard van der

Waals and electrostatic terms (i.e., no explicit hydrogen-bond potentials). The van der Waals parameters are taken from the work of Hagler.[21–24] All terms were subject to an 8 Å cutoff, and no explicit solvent effects were accounted for. Our basic philosophy was to use the simplest energy function available and to dock to the favorable energy minima that it gives. Of course, our method could readily be adapted to more complex energy functions, or possibly explicit solvent effects. Our peptide charges were based on a library obtained from John Moult, which again is based on the work of Hagler. Since the lack of solvent effects meant that the interactions involving net charge groups tended to be greatly overestimated, we chose to rescale all net charge groups to have neutral charge, following the method used for the GROMOS vacuum potentials.[25] The alternative to this is to introduce a dielectric constant, but this has the unacceptable effect of changing hydrogen-bond distances, since the hydrogen-bond geometry is determined by both charge interactions and van der Waals forces.

Another reason to avoid net charges is that they can artificially enhance or suppress docking results, since net charge opposites will attract each other over long distances and cause oversampling of states involving charge-charge interactions. In the case of MTX docking to DHFR, where ionic interactions play an important role in binding, the number of successful dockings greatly increases when net charges are used. In contrast, most electrostatic interactions in the OMTKY3-SGPB system are H-bonds, and the presence of net charges around the active site can tend to pull away states that might otherwise dock successfully.

For nonpeptide groups, we chose charges based on a careful comparison with the peptide charges. Since a significant problem of choosing charges is one of maintaining consistency with the chosen van der Waals parameters, rather than finding the exact electron-density, we chose not to directly use quantum mechanical methods. Another reason for this approach is that we end up rescaling net charge groups (as above), so our charges can only roughly correspond to the actual electron-density distributions.

RESULTS

To test our method, we chose crystallographic structures of complexes between a drug or inhibitor and a protein. Our basic objective was to pull apart the complex and to see if BOXSEARCH could find the observed complex, using a fairly large search area around the binding site. Thus we assumed the general location of the binding site is known, but nothing else. In addition to docking the whole ligand, we also attempted to dock various fragments to test the feasibility of building up the ligand from the fragments that comprise it, without knowing a priori how these fragments fit together. Obviously, an advantage to using complex structures as test cases is that there is a known solution to look for in the results. Unfortunately, a method that successfully reassembles a complex will not necessarily work when docking a new ligand. The components of a complex refined with stereochemical restraints are certain to be almost perfectly complementary. In contrast, even interactions characterized as lock-and-key typically involve atomic shifts on the order of 0.5 Å, which relieve bad contacts and increase complementarity. Therefore, it is necessary to perform tests using the native conformation of a target protein, when the structure is available.

Some explanation regarding the choice of fragments may be necessary. Although our method is designed to be used with rigid fragments, most of our test fragments have some degrees of freedom. In general, we can work around the problem of flexible fragments by selecting several different likely conformations and docking each one independently. Essentially, this approach treats different conformations of the same molecule as different fragments. For our tests, however, we have chosen to dock the fragments only in the conformation found in the actual inhibitor. There are two reasons for this choice. First, there is obviously considerably less computer time involved in docking a single conformation, as opposed to docking maybe 10 or more conformations. Second, and more important, we have no way of assessing whether or not a fragment docked in some other conformation is "right" or "wrong," since we will generally have no experimental evidence to go by. We can only compare the docking of fragments with experiments after we have assembled fragments into a complete molecule, which would be done either "by hand" or using some computationally automated scheme. That, however, is beyond the scope of this work.

Docking of OMTKY3 to SGPB

Streptomyces griseus proteinase B (SGPB) is a member of the chymotrypsin family of serine proteinases. We examined the crystal structure of this protein in complex with the third domain of the ovomucoid inhibitor from turkey (OMTKY3), a protein inhibitor of SGPB, solved in the laboratory of M.N.G. James.[13] As with other protein inhibitors of serine proteinases, the reactive site loop of OMTKY3 is highly complementary to the active site of SGPB. It is believed to bind like a normal peptide substrate, but the reactive site peptide bond between residues 18I and 19I (the letter "I" indicates a residue of the inhibitor) is cleaved very slowly.[26] In addition, there are a number of other sites on the peptide backbone of the reactive site loop that contribute to the high affinity of the inhibitor. By independently docking small peptide fragments, we wanted to see if we could reconstruct the backbone

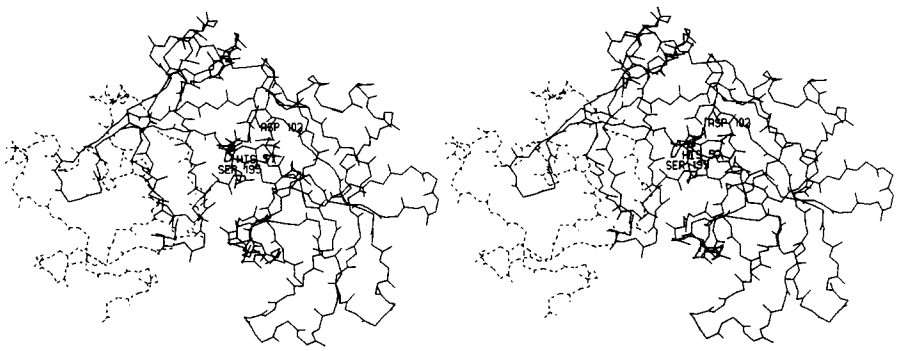

Fig. 1. The crystal structure of OMTKY3 (dashed lines) bound to SGPB. Only backbone atoms are shown, along with the residues of the catalytic triad Ser195, His57, Asp102.

conformation of the reactive site loop. The binding of OMTKY3 to SGPB is shown in Figure 1.

Another reason to choose this structure is that the native SGPB (without bound inhibitor) also has been solved,[27,28] so we could test the effect on the docking method of not applying the small conformational changes inherent in using a complexed structure. The conformational changes in the active site are relatively small, typically less than 0.5 Å.[13] However, since van der Waals contacts are extremely sensitive to atom movements, even small changes can have significant effects on the ability to dock fragments to the binding site. Because the atom shifts are so small, this system makes an ideal test case to determine the sensitivity to conformational changes in a "best case" situation.

We defined the search region as a 40 Å cube centered on a region outside the active site of SGPB, selected so that the whole inhibitor could fit within the cube. Because the whole inhibitor has a diameter of about 27 Å, its center can move within a box of approximately 13 Å on an edge. For smaller fragments we used a 50 Å box, since we wanted to cover as wide an area of the surface as possible. A 0.5 Å grid was used for the floating runs. We used a fixed set of temperature steps for the simulated annealing runs, starting at a temperature of kT = 10 kcal/mole and going down to $kT = 10^{-4}$ kcal/mole (see Table I). We tested a number of fragments taken from the OMTKY3 structure, including a 3-residue fragment from the reactive site loop (FRAG1) and 4 different dipeptide fragments in different conformations (FRAG2-5) (see Table II).

For our tests to be successful, we require that the binding mode from the crystal structure appear, within some tolerance, in the list of energetically favorable configurations that BOXSEARCH generates. In all cases, we counted a docking to be successful if it was within 2 Å RMS from the crystal structure for the inhibitor. Generally, we consider a docking to within 1 Å RMS to be good, and one between 1 Å and 2 Å RMS to be marginal. However, dockings within 2 Å RMS will usually converge to the same docking on energy minimization; hence, by selecting the 2 Å cutoff, we are implying that subsequent energy minimization would be performed to give better results. To avoid duplication this would be carried out after the CLUSTER program had been run on the docking results. At present, we have not developed the software to perform this step.

Generally, we would expect the crystal structure to be the lowest energy configuration, but this may not be the case because of deficiencies in the potential functions. In addition, the subfragments from the crystal structure need not be in global energy minima because they are free to find new interactions that are not possible when they are part of the whole ligand. But in any case, we do expect the crystal structures of the fragments to be in or near local minima of reasonably low energy. In all tests, we were successful in docking OMTKY3 and its fragments to both the native and complexed SGPB structures. As expected, results are somewhat better in the complexed case.

The docking of the whole OMTKY3 inhibitor was performed in a 40 Å box using 20,000 independent runs. For docking to complexed SGPB, the lowest energy docking (see Fig. 2a) was 0.41 Å RMS from the crystal structure with an energy of −68.3 kcal/mole. Data for SGPB complex runs are given in Table III. The docking to the native structure was, as expected, not as good, although it was still the lowest energy in the run. The lowest energy docking (see Fig. 2b) was 1.74 Å RMS from the crystal structure complex with an energy of −51.3 kcal/mole. In all native dockings, the position of the docked fragments is compared to the position of OMTKY3 relative to SGPB after the native and complexed structures have been overlapped. The data for the SGPB native runs are given in Table IV.

It is interesting to compare the dockings with the true minima for the potential functions. In a sense, this will be a fairer test of the method, since we can expect dockings to be only as good as the potential

TABLE I. The Annealing Schedule

Step No.	kT (kcal/mole)	No. of runs	Step No.	kT (kcal/mole)	No. of runs
1	10	5	6	1.0	5
2	8.0	5	7	0.5	10
3	6.0	5	8	0.25	10
4	4.0	5	9	0.1	50
5	2.0	5	10	10^{-4}	50

TABLE II. OMTKY3 Fragments Used in Docking to SGPB*

Fragment	Members (residues)	ψ (degrees)	ϕ (degrees)
OMTKY3	7I–56I (whole inhibitor)	NA	NA
FRAG1	17I–19I (complete residues)	NA	NA
FRAG2	16I–17I (dipeptide analogue)	146.6	−69.0
FRAG3	18I–19I (dipeptide analogue)	45.6	−83.0
FRAG4	19I–20I (dipeptide analogue)	153.0	−98.7
FRAG5	20I–21I (dipeptide analogue)	114.8	−148.8

*The ψ and ϕ angles defining the fragment conformation are given. The letter "I" denotes a residue from OMTKY3.

functions will allow. To test this, we performed a separate minimization of the inhibitor from the crystallographically observed docking and found that it moved 0.28 Å RMS with an energy of −70.9 kcal/mole. The best docking was 0.19 Å RMS from the energy minimum. The minimum was found by running BOXSEARCH in a "minimization" mode (with a very low temperature) until it performed 500 MC steps without changing energy. In this mode, BOXSEARCH performs only one run (instead of multiple runs) and starts the probe at the coordinates as they are read from the probe file. For native SGPB, the energy minimum was 0.72 Å RMS from the crystal complex with an energy of −53.6 kcal/mole. The best docking was 1.29 Å RMS from this minimum.

We performed dockings on a number of fragments of the OMTKY3 inhibitor. The first fragment, denoted as FRAG1, was defined to be the three reactive site residues 17I-19I of OMTKY3, which display high complementarity to the active site. The lowest energy docking (see Fig. 3a) to the complexed SGPB differed from the crystal structure for the fragment by 1.10 Å RMS, with an energy of −28.7 kcal/mole. Hence the global energy minimum for the FRAG1 again corresponded to the crystal structure. In the case of the native SGPB dockings, the lowest energy docking (with an energy of −32.3 kcal/mole) did not correspond to the crystal structure: the lowest energy docking within 2 Å RMS occurred at 1.76 Å, and had an energy of −19.3 kcal/mole (see Fig. 3b).

The other fragments, FRAG2 to FRAG5, are a collection of dipeptide fragment analogues that are in several selected conformations. The ϕ-ψ angles for the respective fragments are chosen to agree with those found in residue sequences 16I–17I, 18I–19I, 19I–20I, 20I–21I of OMTKY3. The dipeptide analogue chosen was N-isopropyl-isobutylamide, which is essentially an Ala-Ala dipeptide in which the terminal amino and carboxyl groups have been replaced by methyl groups. The idea behind using the methyl termini is that we want to test for favorable interactions with the peptide group alone, without having to consider the electrostatic properties of the atoms at the termini. We could then assemble ligands by substituting atoms from other, possibly nonpeptide, probes at the termini. The structure of FRAG2 is shown in Figure 4: the ϕ-ψ angles for all the fragments are listed in Table II.

We used a 50 Å box for the runs with FRAG2 to FRAG5 and achieved successful dockings for each fragment. Hence, for example, the fragment FRAG2 found low energy dockings within 2 Å RMS of the atoms from 16I–17I in OMTKY3. In addition, some fragments docked in the correct mode corresponding to some of the other fragments in the list. For example, there were dockings in the FRAG2 run (corresponding to 16I–17I) that docked within 1.2 Å RMS from 18I–19I. The dockings did not generally represent the global energy minimum, since the minimum energy for each fragment in the correct docking mode varied somewhat. For example, the lowest energy docking within 2 Å RMS for FRAG2 had an energy of −9.8 kcal/mole, whereas the lowest energy for the entire run was −18.4 kcal/mole. In contrast, the correct docking of lowest energy for FRAG3 had an energy of −16.8 kcal/mole, which is quite close to the lowest energy for the run, −19.5 kcal/mole. The dockings to complexed SGPB for FRAG2 to FRAG5 are shown in Figure 5a–d.

It is important to emphasize that BOXSEARCH only prints the coordinates of states that fall below a specified energy. We generally selected this cutoff energy according to the fragment we were trying to dock. Some fragments docked correctly at or near the global energy minimum, and we could use a low cutoff energy, whereas others docked at higher energies and required the use of a higher cutoff energy. Consequently, the number of states in the CLUSTER listing varied considerably between these extremes. In general, the placement of the ranking is more important than the total list size. For example, referring to Table III, we note that FRAG3 ranked 2nd out of 9 clusters, but its placement in the list would still be 2 had the cutoff energy been raised and the size of the list increased.

For the fragments FRAG2 to FRAG5, we also performed dockings over the whole surface of complexed SGPB by defining a 60 Å box centered near the geometrical center of the protein. The data for these runs are given in Table V. Some of the fragments docked successfully, whereas others had dif-

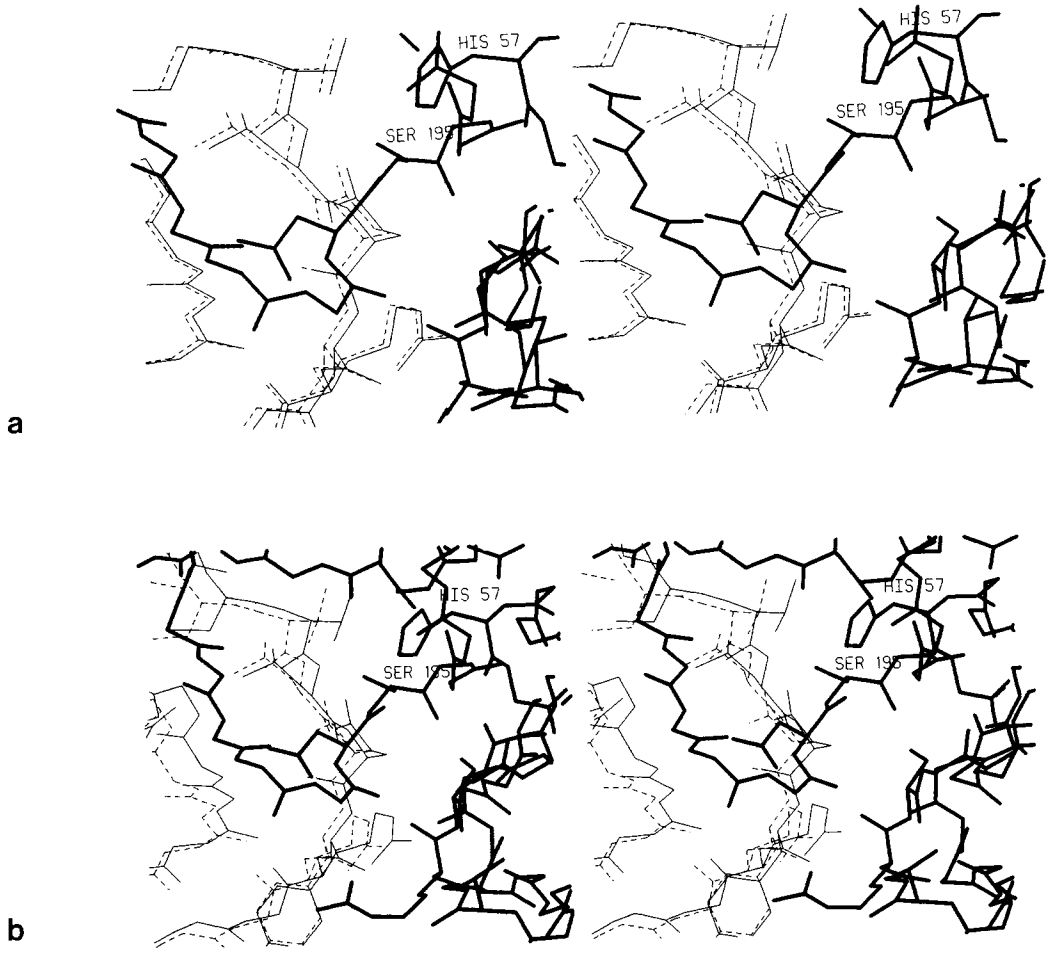

Fig. 2. Docking of the whole OMTKY3 inhibitor (dashed lines) overlapping the crystal structure for OMTKY3 (thin lines) with **(a)** complexed, and **(b)** native SGPB (thick lines). Only sidechains that are in contact between the inhibitor and the proteinase are shown.

TABLE III. Results From Docking the OMTKY3 Fragments to SGPB From Complex*

Fragment	No. runs	Box size (Å)	No. correct dockings	RMS (Å)	E_{docked} (kcal/mole)	E_{min} (kcal/mole)	Ranking vs. total clusters
OMTKY3	2×10^4	40	6	0.41	−68.3	−68.3	1/26
FRAG1	2×10^4	40	3	1.10	−28.7	−28.7	1/143
FRAG2	2×10^4	50	3	0.96	−9.8	−18.4	97/257
FRAG3	2×10^4	50	6	0.88	−16.8	−19.5	2/9
FRAG4	2×10^4	50	12	1.14	−13.1	−19.1	14/253
FRAG5	2×10^4	50	5	1.10	−8.4	−17.1	361/738

*Dockings are considered correct if they fall within 2.0 Å RMS of the complexed structure. The ranking gives the placing of the lowest energy correct docking in the complete docking list, after the cluster analysis.

ficulty finding the minima (in particular FRAG5, which had no successful dockings), in spite of a 4-fold increase in the number of independent runs. One explanation for this is in the way that floating runs are performed. Since most of the volume is inside the protein, the floating runs will be used extensively to get good starting states. However, since the floating tends to move states to the nearest point on the surface, concave regions of the surface will get better sampling than the convex regions. Our results tend to support this hypothesis, since we were more successful in docking FRAG3 and FRAG4, which are deep in the binding pocket, than we were for FRAG2 and FRAG5, which are more towards the edge of the binding site. In contrast, one expects lower energy states to have higher sampling, by a factor of $\exp(\Delta E/kT)$, when MC runs are sufficiently long to allow the system to attain equi-

TABLE IV. Results From Docking the OMTKY3 Fragments to Native SGPB*

Fragment	No. runs	Box size (Å)	No. correct dockings	RMS (Å)	E_{docked} (kcal/mole)	E_{min} (kcal/mole)	Ranking vs. total clusters
OMTKY3	2×10^4	40	4	1.74	−51.3	−51.3	1/129
FRAG1	2×10^4	40	2	1.79	−19.3	−32.3	51/617
FRAG2	2×10^4	50	2	1.00	−10.0	−18.1	197/1058
FRAG3	2×10^4	50	4	1.22	−16.0	−18.2	7/395
FRAG4	2×10^4	50	4	1.29	−11.5	−18.4	58/405
FRAG5	2×10^4	50	5	1.21	−8.6	−18.5	488/1014

*Dockings are considered correct if they fall within 2.0 Å RMS of the complexed structure. The ranking gives the placing of the lowest energy correct docking in the complete docking list, after the cluster analysis.

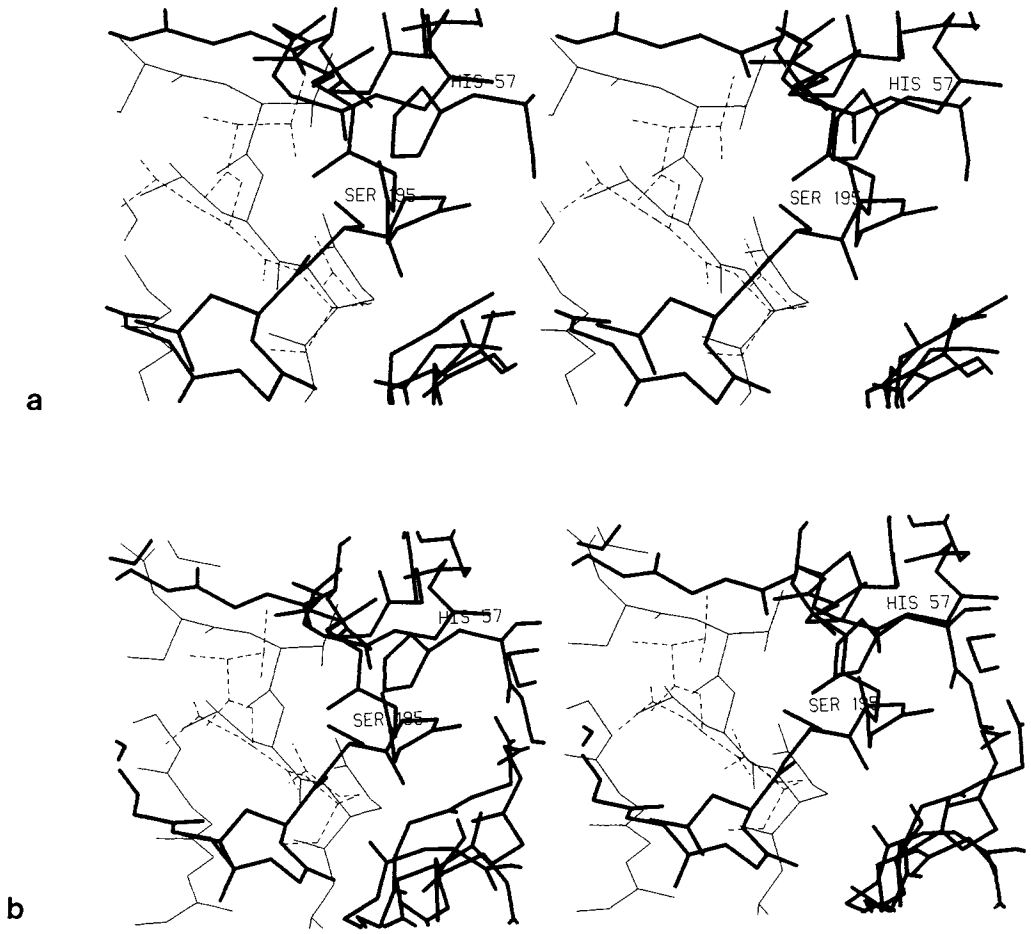

Fig. 3. Docking of FRAG1 (dashed lines) overlapping the crystal structure for OMTKY3 (thin lines) with **(a)** complexed, and **(b)** native SGPB (thick lines).

librium. However, our runs are extremely short and the resulting set of states is generally far from equilibrium.

Although there is some deviation from the crystal structure for the dipeptide analogues, we feel that the correspondence is sufficiently good that we could reconstruct the backbone shape of the reactive site loop from the independently docked fragments. This basically comes down to a fundamental question: are the minima of the independently docked fragments close to their corresponding positions as part of the whole inhibitor? In Figure 6, we show the lowest energy correct dockings for FRAG3-5 with the backbone of the reactive site loop. Here, we would join the C^{α} atoms from neighboring residues to form the inhibitor backbone (note that not all atoms from the fragments are shown). This figure illustrates that this approach is feasible provided we allow a

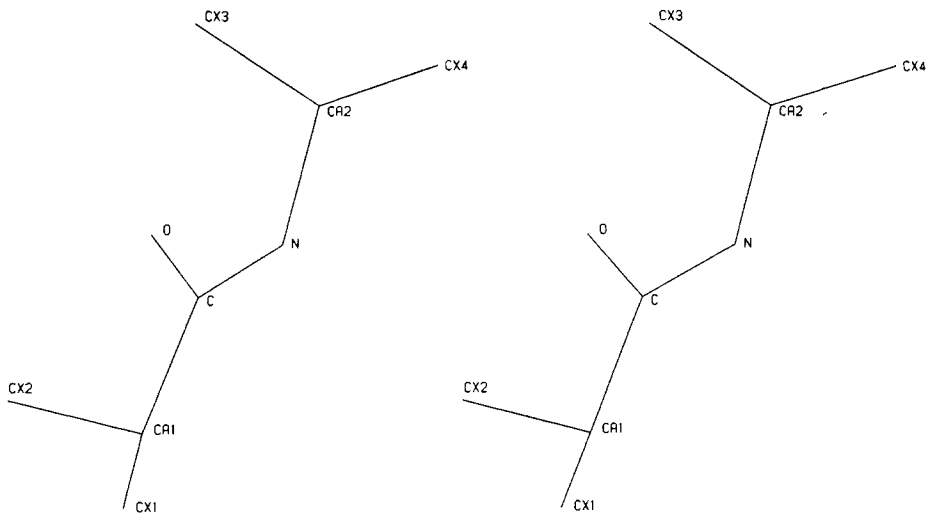

Fig. 4. Stereo view of the dipeptide fragment analogue FRAG2. The other dipeptide analogues are of the same structure but in different conformations.

sufficiently large tolerance in the position of the atoms that are to be joined.

Docking of MTX to DHFR

The other system tested was the structure of *E. coli* dihydrofolate reductase (DHFR) in complex with the inhibitor methotrexate (MTX), solved in the laboratory of J. Kraut.[14] Methotrexate is a potent inhibitor of DHFR and is a widely used anticancer drug. In our tests, we wanted to determine how each subfragment contributed to the energy of binding in order to see if we could "build up" methotrexate from these independently docked subfragments.

The DHFR-MTX system offered difficulties not present in the SGPB-OMTKY3 system, primarily due to the presence of charged groups on MTX. At physiological pH, it is generally thought that the N1 nitrogen is protonated and that the carboxylate groups are not protonated, giving a net charge of −1.0e. Of course, this is further complicated by the fact that the positive charge is at one end, and the two negatives are at the other. The α-carboxylate of the glutamic acid portion of MTX forms a strong ionic interaction with Arg57. Therefore, our simplified energy model—using neutral-charge groups—suffered at the presence of these obviously important ionic interactions. Nonetheless, we were determined to be consistent in our treatment of charge interactions, and therefore we chose not to modify our treatment of net charge groups (i.e., we neutralized all charge groups).

We used three probes: the entire methotrexate molecule (MTX), a fragment containing only the pteridine ring and its amino side groups (MF1), and a fragment containing both the aminobenzoyl and glutamyl groups (MF2). The chemical structures of MTX, MF1 and MF2 are given in Figures 7a–c. We defined the search region to be a 25 Å cube centered approximately in the middle of the binding site. The floating and temperature parameters (Table I) were the same as for the SGPB-OMTKY3 system: the floating tolerance was set to 0.5 for MTX and MF2 and to 0.7 for MF1. These changes were necessary since the binding pocket more completely surrounds the inhibitor, and there is less leeway for movement.

The dockings for the whole MTX inhibitor found 5 repeats within 2 Å RMS of the crystal structure. The lowest energy docking of these was within 0.65 Å RMS of the crystal structure for MTX, with an energy of −31.9 kcal/mole. This docking, shown in Figure 8, was among the lowest energies for the run, ranking 6th in the CLUSTER list. The data for all DHFR runs are shown in Table VI.

The lowest energy for the MTX run was −43.3 kcal/mole and was positioned along the NADPH binding grove. Since MTX does not actually bind at these positions, we must blame the energy potentials for assigning lower energies than should be the case. On closer examination (see Fig. 9) we see that, whereas there are favorable charge and van der Waals contacts, the hydrophobic benzyl moiety is in close proximity to a number of polar groups, specifically the carbonyl and amide groups of Met16, and the hydroxyl group of Thr123. The free energy cost of these contacts would be considerable, since a number of bound solvent molecules would be forced out of low energy binding sites. This demonstrates a serious problem in neglecting hydrophobic and solvent effects.

The results for MF1 were also very good: the lowest energy correct docking was 0.45 Å RMS from the

crystal structure with an energy of −21.3 kcal/mole. The global minimum for the run was −24.2 kcal/mole. For MF2, the correct docking was considerably lower in the list: it ranked 284 out of 650, 1.0 Å RMS from the crystal structure and with an energy of −15.6 kcal/mole. The global minimum for the run was −29.7 kcal/mole. The correct dockings for MF1 and MF2 are shown in Figure 10. The relatively poor ranking of MF2 again shows the problem of neglecting solvent effects, since a major source of binding free energy is due to the placement of the benzyl moiety in a hydrophobic pocket formed by Leu28, Phe31 and Ile50 (see Fig. 8).

DISCUSSION

In developing a docking method, we have aimed to solve the combinatorial problem of considering the many ways a molecule could conceivably interact with the target. We wish to understand which methods and strategies might be most effective in developing ab initio drug design techniques. We chose a Monte Carlo technique because of its proven effectiveness in a wide range of minimization problems and because its random nature allows for many different avenues in configuration space to be explored in a complete fashion. We chose to use energetics as our main criterion of fit since this is ultimately the most physically realistic criterion that can be computed in a reasonable amount of computer time (as opposed to the very costly free energy techniques).

Recently, Goodsell and Olson[9] have developed a method of docking in which simulated annealing incorporates the grid energy evaluation method of Goodford[12] and allows for internal degrees of freedom of the probe. They had good success docking ligands from a single arbitrary position 3 to 5 Å from the binding site. Whereas our method also uses MC as a minimization method and energy evaluation to drive the minimization, our implementation and goals are quite different from theirs. In essence, our method attempts to recover *all* favorable states of a rigid probe in a given region, duplicating the results of a systematic search. Because of this, we require that a great number of initial states be produced, which limits the amount of computation for each run. Also, we use the floating procedure, a new method, which is the key to being able to get reasonable results using such short MC runs. The floating procedure essentially allows initial states that are buried inside the target to move to better docking positions, so we can start with states that are uniformly distributed within a given volume, rather than starting them only from the outside of the target. The method of Goodsell and Olson runs much faster than ours, mainly because of their incorporation of grids to do the energy calculation. In principle, our method could adopt this innovation and would be considerably more efficient. As implemented, both methods suffer from the problems of not including solvent effects, or conformational changes in the target, in an appropriate way.

It is interesting to compare our method to a systematic search that is not directed by energetic or chemical constraints. Such a search would simply move the probe through every possible position and orientation, by steps that are separated by a given RMS tolerance, and within a specified search region. To estimate, in general, the number of states involved, let L be the dimensions of the box, R be the probe radius (for simplicity, assume the probe to be spherical), and let d_{tol} be the RMS step size between the states. The number of states that are generated by considering only the translations is approximately:

$$N_{\text{trans}} \approx \left[\frac{L - 2R}{d_{\text{tol}}}\right]^3,$$

where the term $L - 2R$ arises since the probe is constrained to remain completely inside the box. To estimate the number of states generated by rotations, we first observe that the RMS difference between two states that differ by a rotation of angle α about any axis is:

$$d_{\text{RMS}}(\alpha) = \left[\frac{4}{5}(1 - \cos(\alpha))\right]^{1/2} R.$$

This result is easily found by integrating over the sphere using spherical coordinates and assuming the rotation is about the z-axis. We invert this expression to get the angle α_{tol} of rotation giving the RMS difference d_{tol}:

$$\alpha_{\text{tol}} = \cos^{-1}\left[1 - \frac{5}{4}\left(\frac{d_{\text{tol}}}{R}\right)^2\right].$$

This gives the angular stepsize for the rotational part of the systematic search. Using an "efficient" rotation sampling scheme,[29] the total number of distinct rotations is (with α_{tol} in radians)

$$N_{\text{rot}} \approx \frac{1}{\pi}\left[\frac{2\pi}{\alpha_{\text{tol}}}\right]^2\left[\frac{\pi}{\alpha_{\text{tol}}}\right],$$

and the total number of states (and therefore, energy evaluations) is

$$N = N_{\text{rot}} N_{\text{trans}}.$$

Fig. 5. Lowest energy correct dockings for FRAG2-FRAG5 (dashed lines) along with the corresponding residues from OMTKY3 (thin lines) and complexed SGPB (thick lines). **(a)** FRAG2 docking compared with 16I-17I. The primary electrostatic interaction is a hydrogen bond formed with the amide nitrogen of Gly216. **(b)** FRAG3 docking compared with 18I–19I. An important electrostatic interaction is a hydrogen bond formed with the amide nitrogen of Asp194. **(c)** FRAG4 docking compared with 19I–20I. The primary electrostatic interaction is a hydrogen bond formed with the carbonyl oxygen of Arg41. **(d)** FRAG5 docking compared with 20I–21I. The primary electrostatic interaction is a hydrogen bond formed with the amide nitrogen of Arg41.

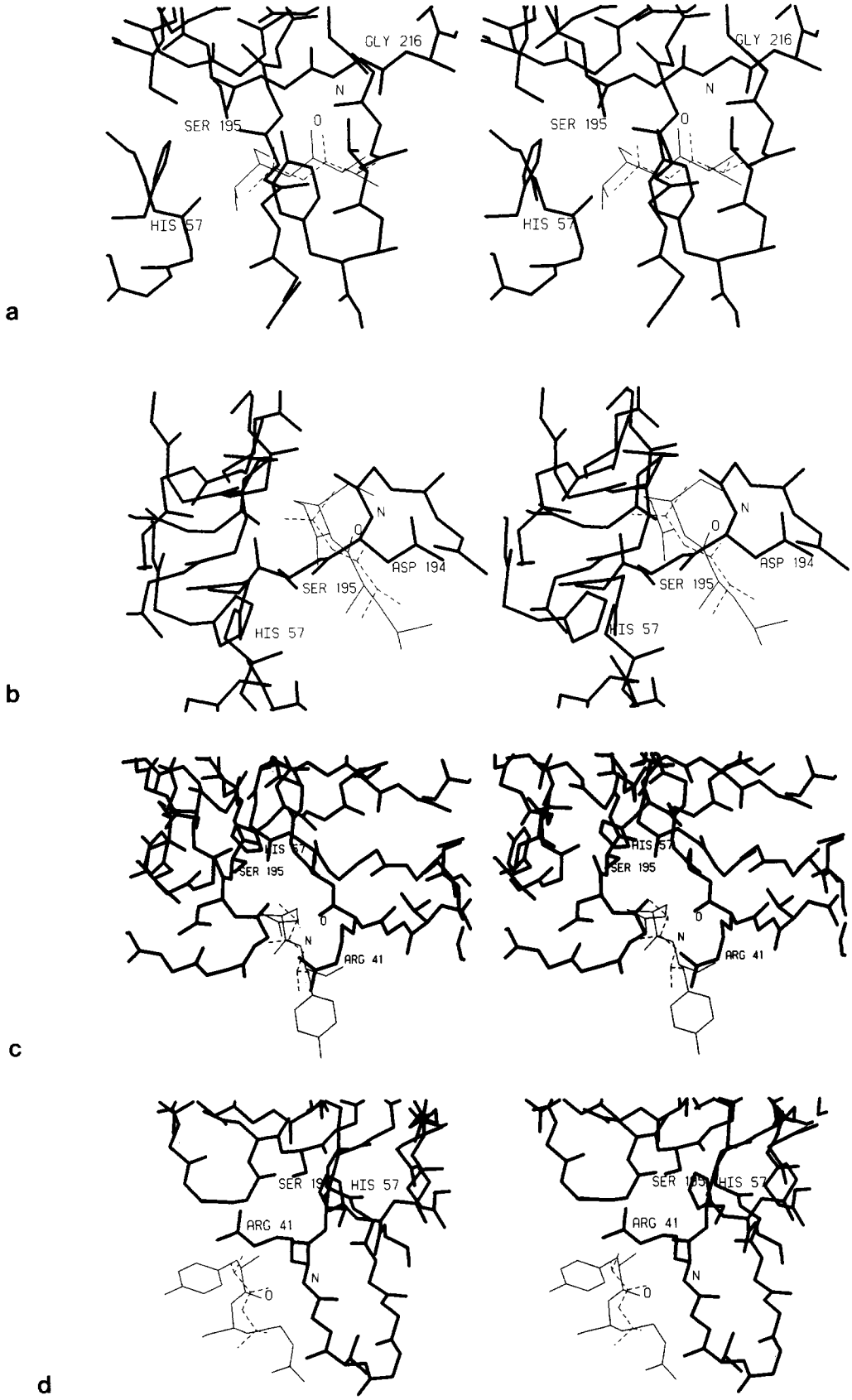

Fig. 5. Legend appears on page 216.

TABLE V. Results From Docking the OMTKY3 Fragments to SGPB From Complex for Dockings Over the Whole Protein Surface*

Fragment	No. runs	Box size (Å)	No. correct dockings	RMS (Å)	E_{docked} (kcal/mole)	E_{min} (kcal/mole)	Ranking vs. total clusters
FRAG2	8×10^4	60	2	0.79	−10.2	−18.3	363/1051
FRAG3	8×10^4	60	17	1.32	−18.0	−20.7	3/995
FRAG4	8×10^4	60	7	1.21	−12.3	−19.6	80/1097
FRAG5	8×10^4	60	0	NA	NA	−18.7	NA

*Dockings are considered correct if they fall within 2.0 Å RMS of the complexed structure. The ranking gives the placing of the lowest energy correct docking in the complete docking list, after the cluster analysis.

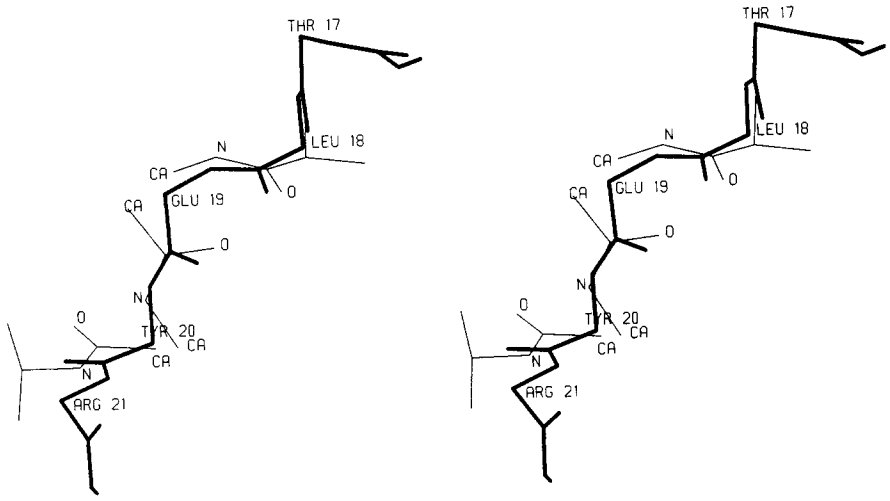

Fig. 6. Overlapping of the lowest energy correct dockings for FRAG3, FRAG4, and FRAG5 (thin lines) with the backbone of OMTKY3 (thick lines). Not all atoms in the fragments are shown. This illustrates that we could, in principle, build the backbone conformation of the inhibitor from the independently docked fragments. See text for explanation.

To consider an example, take the docking of the OMTKY3 to SGPB. Here we have L = 40Å, R = 13Å, and we will take d_{tol} = 1Å. Then the above formulas give the total number of energy evaluations to generate the systematic search as

$$N \approx 6 \times 10^7.$$

In our docking runs on SGPB-OMTKY3, we have a total 2×10^4 runs, each involving 150 energy evaluations, giving a total number of energy evaluations as

$$N = 3 \times 10^6.$$

In this case, our method is better than a systematic search by an order of magnitude. We want to emphasize, however, that we are comparing our method to a completely unconstrained systematic search: sophisticated use of stereo-chemical constraints (for example, the method of Bacon and Moult[7]) would reduce the number of states considerably. In contrast, a systematic search would have to be coupled with energy minimization in order not to miss energy minima.

In testing our method, we have aimed at answering a number of questions. First, we wanted to determine the overall effectiveness of MC in searching for minima, and particularly its effectiveness at finding the correct answer, as represented by the crystal structure of the complex. Second, we wanted to examine the differences between trying to dock to complexed and native structures in order to examine the effect of conformational changes in the target. Third, we wanted to try to determine which basic strategies might be most effective in designing compounds ab initio from the target structure alone.

In all the tests we performed, BOXSEARCH was able to duplicate the complexed crystal structure, although it sometimes was not the lowest energy found. In docking OMTKY3 to SGPB, the crystal structure was represented by the lowest energy docking; in docking MTX to DHFR, the crystal structure was represented by a low energy docking, although there were several dockings of lower energy. The dockings of MF2 to DHFR indicate that the energy ranking of MTX may be explained as an

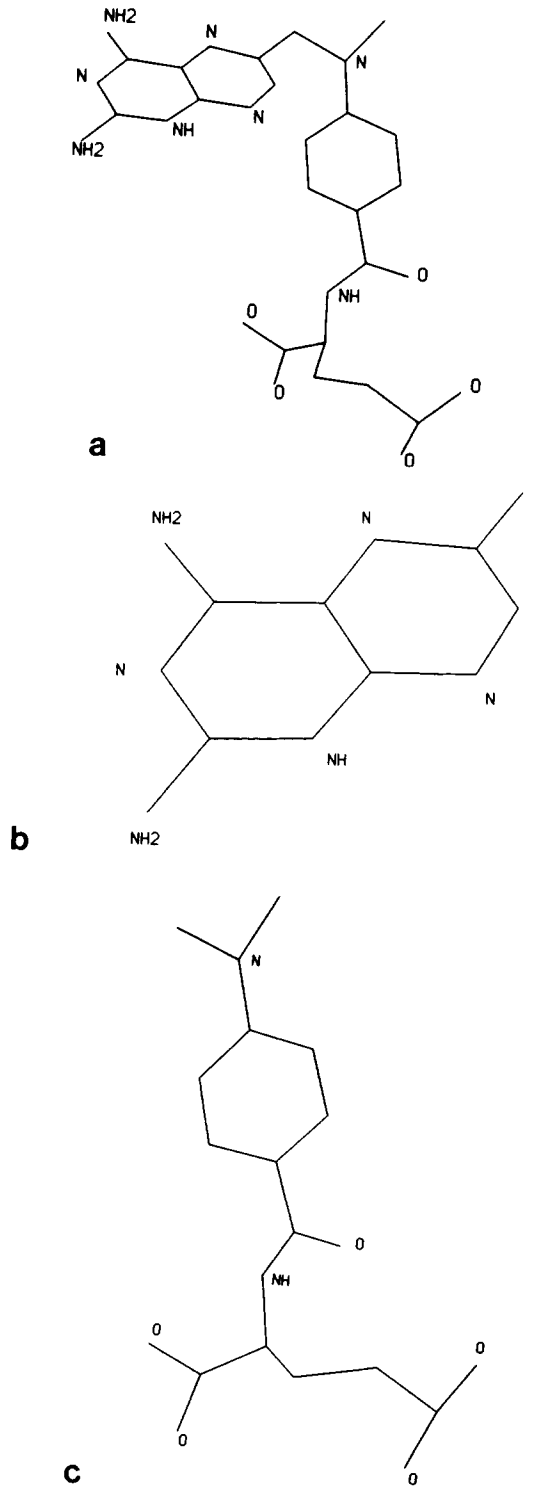

Fig. 7. Chemical structures for **(a)** MTX, **(b)** MF1, and **(c)** MF2.

inadequacy in the potential energy functions. Of course, better energy evaluation is the means of solving this problem, and fortunately there is much research being done on the problem of finding better potential parameters and functional forms. We chose to use the most straightforward energy function in order to focus testing on the method's ability to search the configuration space, rather than building the most accurate representation of the molecular interactions.

The dockings to native SGPB raise the issue of the effect of conformational changes on docking. Even though the conformational changes in the protein are small, the effect on the quality of the dockings is significant. In particular, we want to be able to predict the binding energy accurately, since this will determine where dockings appear in the list and affect the magnitude of the combinatorial problem of matching independently docked fragments to build ligands. There are basically two situations to consider: small changes that involve slight shifting of atoms (typically less than 1 Å) and large conformational changes in main or side chains of up to 3 Å or more. In the case of SGPB, all active site conformational changes are of the former type. For DHFR, there are probably both types present, particularly since flexible surface side-chains are involved in binding the inhibitor.

There are a number of approaches that can be taken to this problem. One possibility is to handle small conformational changes by using "soft" van der Waals potential functions. This could involve taking a form that increases less rapidly as $r \rightarrow 0$ than the $1/r^{12}$ or $1/r^{9}$ form. However, this would be an ad hoc approach, since the form is not derived on the basis of a physical principle. An alternative might be to calculate the approximate changes in target atom positions given a fixed probe position, and then calculate the energy at these shifted positions, including a compensating harmonic energy term that would tend to minimize the shifts. The computational cost would be approximately double that of simple pairwise potentials. We are currently developing an implementation of this method.

Of course, these methods could only deal with small conformational shifts. An alternative that might work for both small and medium conformational changes in the target is to perform an energy minimization prior to each energy evaluation, again with the probe in a fixed position. To avoid large deviations from X-ray coordinates, the minimization could be restricted to residues within a fixed distance of the probe. This procedure would allow bad contacts to be removed with small shifts but would also allow large changes in side chain conformation. The computational cost of such a scheme would certainly test the limits of existing technology.

The methods we have developed and presented here are a first step toward the ab initio design of lead compounds from the known three-dimensional structure of the drug target. At the present state of development, the primary value of this tool is to indicate the directions for future work and to suggest

TABLE VI. Results for Docking Fragments to DHFR*

Fragment	No. runs	Box size (Å)	No. correct dockings	RMS (Å)	E_{docked} (kcal/mole)	E_{min} (kcal/mole)	Ranking vs. total clusters
MTX	2×10^4	25	5	0.65	−31.9	−41.3	6/59
MF1	2×10^4	25	12	0.45	−21.3	−24.2	18/283
MF2	2×10^4	25	4	1.00	−15.6	−29.7	284/650

*Dockings are considered correct if they fall within 2.0 Å RMS of the complexed structure. The ranking gives the placing of the lowest energy correct docking in the complete docking list, after the cluster analysis.

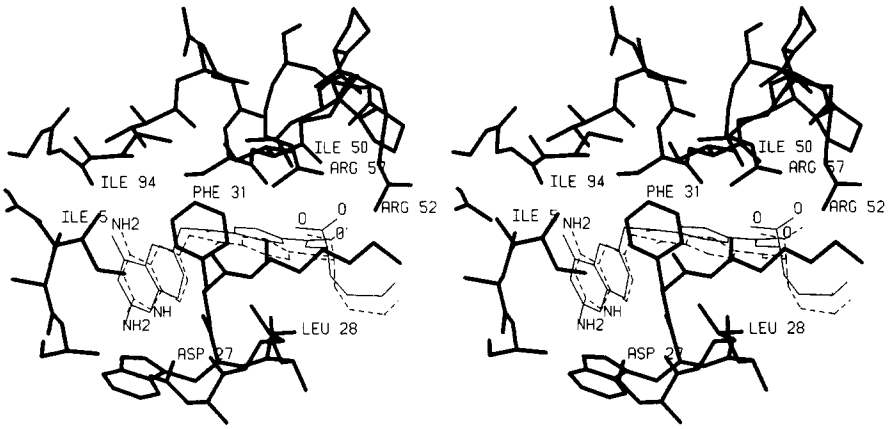

Fig. 8. Lowest energy correct docking for MTX (dashed lines) with the crystal structure for MTX (thin lines) and DHFR (thick lines). MTX forms electrostatic interactions with the backbone of Ile5 and Ile94, and with the sidechains of Asp27, Arg52, and Arg57. The residues Leu28, Phe31, and Ile50 form a hydrophobic pocket.

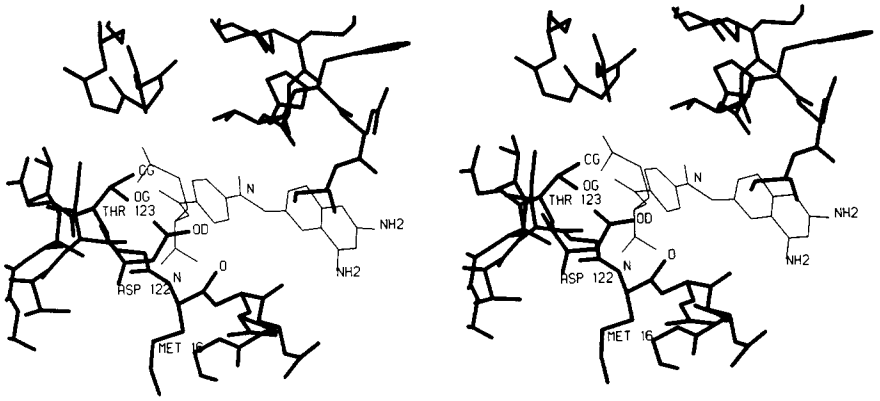

Fig. 9. Global energy minimum docking found for MTX (thin lines) with DHFR (thick lines). The docking occurs in the NADPH binding pocket. See text for discussion.

which types of techniques might be likely to work in making viable predictions. Generally we find that the Monte Carlo method, in our implementation, works well in predicting correct binding modes for probes to complexed structures and is able to predict the correct binding to native structures when the conformational changes are small. By making improvements to the energy calculations by incorporating solvent and entropic effects, and by accounting for conformational changes in the target, we hope to be able to make reliable predictions for the binding of small fragments to a native protein of known structure.

ACKNOWLEDGMENTS

This research was supported by the Alberta Heritage Foundation for Medical Research (AHFMR) and the Medical Research Council of Canada. We thank the AHFMR for support in the form of a fellowship (T.H.) and a scholarship (R.J.R.). We thank

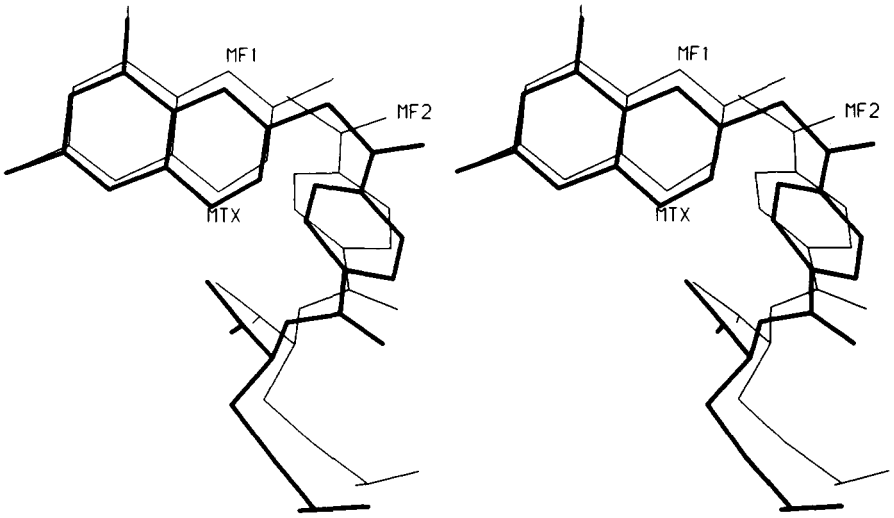

Fig. 10. The lowest energy dockings for MF1 and MF2 (thin lines) within 2 Å of the crystal structure, shown here with the crystal position for MTX (thick line). This figure demonstrates that we could, in principle, design MTX from the two independently docked fragments.

John Moult for helpful discussions and for supplying us with his potential function library, David Bacon, Colin Broughton, and Marie Fraser for helpful discussions, and Natalie Strynadka for help with the figures.

REFERENCES

1. Goodford, P.J. Drug design by the method of receptor fit. J. Med. Chem. 27:557–564, 1984.
2. Hol, W.G.J. Protein crystallography and computer graphics—toward rational drug design. Angew. Chem. Int. Ed. Engl. 25:767–778, 1986.
3. Abraham, D.J. X-ray crystallography and drug design. In: "Computer-Aided Drug Design: Methods and Applications," (Perun, T.J., Propst, C.L., eds.) New York: Marcel Dekker, New York, 1989: 93–132.
4. Beddell, C.R., Goodford, P.J., Kneen, G., White, R.D., Wilkinson, S., Wootton, R. Substituted benzaldehydes designed to increase the oxygen affinity of human haemoglobin and inhibit the sickling of sickle erythrocytes. Br. J. Pharmac. 82:397–407, 1984.
5. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161:269–288, 1982.
6. DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. J. Med Chem. 29:2149–2153, 1986.
7. Bacon, D., Moult, J. in preparation.
8. Yue, S.Y. Distance-constrained molecular docking by simulated annealing. Protein Engineering 4:177–184, 1990.
9. Goodsell, D.S., Olson, A.J. Automated docking of substrates to proteins by simulated annealing. Proteins: Structure, Function, and Genetics 8:195–202, 1990.
10. Allen, F.H., Kennard, O., Taylor, R. Systematic analysis of structural data as a research technique in organic chemistry. Acc. Chem. Res. 16:146–153, 1983.
11. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D., Venkataraghavan, R. 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, 1988.
12. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med Chem. 28:849–857, 1985.
13. Read, R.J., Fujinaga, M., Sielecki, A.R., James, M.N.G. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. Biochemistry 22:4420–4433, 1983.
14. Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., Kraut, J. Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution I. General features and binding of methotrexate. J. Biol. Chem. 257:13650–13662, 1982.
15. Binder, K. Introduction: Theory and technical aspects of monte carlo simulations. In: "Monte Carlo Methods in Statistical Physics," (Binder, K., ed.). New York: Springer-Verlag, 1979: 1–45.
16. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
17. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. Optimization by simulated annealing. Science 220:671–680, 1983.
18. Subbiah, S., Harrison, S.C. A simulated annealing approach to the search problem of protein crystallography. Acta Cryst. A45:337–342, 1989.
19. Wilson, S.R., Cui, W. Applications of simulated annealing to peptides. Biopolymers 29:225–235, 1990.
20. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
21. Hagler, A.T., Huler, E., Lifson, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. J. Am. Chem. Soc. 96:5319–5327, 1974.
22. Lifson, S., Hagler, A.T., Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the CO···H-hydrogen bonds. J. Am. Chem. Soc. 101:5111–5121, 1979.
23. Hagler, A.T., Lifson, S., Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields. J. Am. Chem. Soc. 101:5122–5130, 1979.
24. Hagler, A.T., Dauber, P., Lifson, S. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 3. The CO···HO hydrogen bond and the analysis of the energetics and packing of carboxylic acids. J. Am. Chem. Soc. 101:5131–5141, 1979.
25. Aqvist, J., van Gunsteren, W.F., Leijonmarck, M., Tapia, O. A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein: Secondary structure mo-

tion in a 150 picosecond trajectory. J. Mol. Biol., 183:461–477, 1985.
26. Read, R.J., James, M.N.G. Introduction to the protein inhibitors: X-ray crystallography. In Barrett, A.J., Salvesen, G. (eds): "Proteinase Inhibitors." New York: Elsevier, 1986: 301–336.
27. Delbaere, L.T.J., Hutcheon, W.L.B., James, M.N.G., Thiessen, W.E. Tertiary structure differences between microbial serine proteases and pancreatic serine enzymes. Nature 257:758–763, 1975.
28. Sawyer, L., Sielecki, A.R., James, M.N.G. unpublished coordinates.
29. Lattman, E.E. Optimal sampling of the rotation function. In Rossmann, M.G. (ed.): "The Molecular Replacement Method: A Collection of Papers on the Use of Non-Crystallographic Symmetry." New York: Gordon and Breach, 1972: 179–185.