Creating Artificial Binding Pocket Boundaries To Improve the Efficiency of Flexible Ligand Docking

W. Michael Brown*,† and David L. Vander Jagt‡

Department of Computational Biology, Sandia National Laboratories, Albuquerque, New Mexico 87123, and Department of Biochemistry and Molecular Biology, University of New Mexico School of Medicine, Albuquerque, New Mexico 87131

Received April 30, 2004

Traditionally, algorithms for binding site characterization or identification focus on the problem of identifying atoms within a macromolecule that might be responsible for ligand binding. In this manuscript, we focus on the binding pocket problem from a different perspective as a challenge of calculating an artificial binding pocket boundary that is sufficient to isolate binding pocket volume. The approach involves the calculation of a *macromolecule encapsulating surface* (MES) that separates binding pocket volume from outside space. We show that the MES can be used to increase the efficiency of flexible docking as implemented in AutoDock 3.0. The most significant improvement in docking efficiency is seen when the entire protein is searched and results show additional support for the use of AutoDock, in and of itself, as a feasible tool for binding-site identification for cases in which a protein ligand is known.

INTRODUCTION

Biological metabolism is largely regulated via enzymatic catalysis of biochemical reactions. In this process, a substrate interacts with a macromolecule enzyme whose role is to decrease energy barriers in order to facilitate chemical reaction under physiologic conditions. Most often, this substrate ligand "binds" to a concave region (termed binding pocket here) within the enzyme's surface. The concave nature of the binding pocket within an enzyme would seem to serve several purposes—it provides for interatomic interactions sufficient to facilitate favorable enthaplic energies, it provides for enzyme−substrate specificity, it orients ligand atoms needing chemical modification with active-site atoms or other ligands, and it can lower the energetic cost of assuming ligand conformations that favor reaction (transition-state conformations). Therefore, the binding pocket structure is of intense biological interest for those who wish to understand or possibly alter biochemical mechanisms.

There are many computational tools that focus on binding pocket structure in their analyses. These include approaches to binding-site identification, ligand docking software, de novo design/lead optimization software, and software rendering abstract graphical representations of ligand-binding pocket interactions. Ligand docking tools aim to predict if a ligand will bind to a macromolecule, and if so, the binding conformation(s) assumed by the ligand. De novo design and lead-optimization programs take the binding pocket (possibly with a lead inhibitor bound) and attempt to build ligands predicted to elicit high binding affinity. Most of these algorithms can be roughly grouped into two categories based on their characterization of search space. *Geometry-based algorithms* identify the binding pocket and the ligand as a set of interaction sites (for examples see refs 1−4). These algorithms then solve the problem of matching complementary interaction sites to find the fit resulting in the highest score. *Force field-based algorithms,* on the other hand, utilize a potentially infinite search space, performing optimization of an objective function describing binding energy parametrized by ligand orientation (for examples see refs 5−8).

Typically, both types of algorithm impose some limitation on the search space in order to achieve acceptable run times. Geometry-based algorithms must choose those macromolecule atoms in the binding pocket that are essential to ligand binding. Force-field based algorithms typically focus on a limited volume of the macromolecule presumed to contain the binding pocket. The difficulty for geometry-based algorithms is picking out those surface atoms that would be important for ligand binding. Likewise in force-field based approaches, how do we decide which volume, out of the infinite space continuous with the binding pocket is necessary for ligand binding?

The problem arose for us when working with a lead optimization program which performs systematic addition of molecule fragments onto a seed molecule (for examples of this type of algorithm see refs 7 and 8). An important bound to the combinatorial explosion inherent in these types of algorithms is the binding pocket boundary, defined by the repulsive interactions of the macromolecule atoms with those of the growing ligand.[8] The problem, however, is the infinite space continuous with the binding pocket that might be utilized for ligand growth. While a cutoff based on van der Waals interaction energies has been used in the past, this approximation might not be justified; it is not uncommon to find among crystallographic structures ligand atoms that do not exhibit "tight packing" with macromolecule atoms.

A potential solution to this problem is the calculation of binding pocket boundaries in Cartesian space that can be

* Corresponding author phone: (505)284-8938; fax: (505)284-9379; e-mail: wmbrown@sandia.gov.
† Sandia National Laboratories.
‡ University of New Mexico School of Medicine.

10.1021/ci049853r CCC: $27.50
Published on Web 06/03/2004

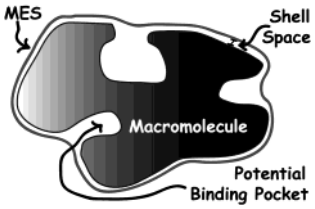

Figure 1. Concept of the macromolecule encapsulating surface (MES).

used to limit the volume that is considered for occupancy by ligand atoms. In addition to the steric boundary formed by macromolecule atoms, such a boundary would necessitate artificial components in order to separate binding pocket volume from outside space. Despite the potential advantages of such an approach, to our knowledge, no algorithm has been published to date with demonstrated efficacy for defining an isolated volume sufficient to encompass ligand atoms in known binding modes. While numerous algorithms have been developed for binding site identification,[9−17] these approaches focus on the identification of macromolecule atoms responsible for ligand binding do not make an explicit definition of the binding pocket boundary which encompasses its volume or, if they do, neglect to quantify the ability to sufficiently encapsulate a bound ligand within this volume. That is, even if such an algorithm is able to identify 100% of the macromolecule atoms in contact with a known ligand, this should not imply that the approach isolates a binding pocket volume sufficient to encompass all ligand atoms within. While binding-site volume statistics are often reported for these algorithms, these volumes correspond to the binding pocket as defined by the algorithm and therefore do not necessarily correlate with volume occupied by the ligand.

For example, we consider the binding site identification program CAST.[10] The algorithm makes explicit definition of the binding pocket volume in terms of a polytope calculated from a weighted Delaunay triangulation of the macromolecule atoms. For at least 5.9% of the test cases, the ligand binding pocket volume is smaller than the ligand volume[10] and therefore cannot encompass the ligand. For the other test cases, although the volume is sufficient in magnitude, data are not given to determine if the binding pocket *boundaries* sufficiently encompass the ligand or not. If the algorithm is used to calculate the binding pockets for cytosine methyltransferase (1HMY.pdb), a member of the test set for the software, we see that the calculated ligand binding pocket does not encompass the α-amino acid atoms of the S-adenosylmethionine ligand in its binding mode, despite the fact the magnitude of the volume (314 Å^3) is sufficient to encompass the ligand and the fact that the α-amino acid atoms exhibit interactions with residues 60, 62, and 98−100 of the protein. Although the algorithm successfully identifies a region of the protein responsible for ligand binding, the calculated binding pocket does not encompass the known ligand.

Therefore, in this manuscript we have focused on the binding pocket problem as a challenge of calculating an artificial binding pocket boundary that is sufficient to isolate the binding pocket volume that will encompass an *entire* ligand when positioned in its binding mode. The approach involves the calculation of a *macromolecule encapsulating surface* (MES). The MES (Figure 1) is defined to be the surface that separates binding pocket volume from volume outside of the macromolecule. The ability to calculate such a surface is desirable because not only does it facilitate a means for binding pocket characterization in a traditional sense but also it defines a volume that might be used to increase the efficiency of force-field based algorithms. The volume considered for the lead optimization algorithms described above could be limited without the requirement for van der Waals potential cutoffs and docking algorithms might be made more efficient by isolation of binding pocket volume from other empty volume included in the search space.

The most straightforward method for calculation of the MES would be to define the surface as the convex hull of the macromolecule. In this manner, any concave regions within a macromolecule would be considered as binding pocket space, and it would be very unlikely that any ligand-binding mode would include atoms positioned outside of this surface. The problem with this approach is that the volume isolated is highly dependent on the shape of the macromolecule—more so than the topology of the binding pocket. A simple qualitative description of the problem can be seen in that the volume isolated for a roughly 'V' shaped protein is much larger than for that of a roughly spherical protein—regardless of the binding pocket shape. Therefore, we have investigated two distinct algorithms for calculation of the MES and make a brief comparison between the two. While we do not present the method for binding pocket characterization based on the MES herein, we seek solution to the simple problem of a set a parameters that minimize binding pocket volume while maintaining a general solution to an artificial boundary sufficient to encapsulate known ligands for the majority of test cases.

In the second part of the manuscript, we show how the MES can be used to improve the efficiency of flexible ligand docking as implemented by the program AutoDock[18] by isolating binding pocket volume as defined by the MES in an energetic fashion. In addition to the case where the binding pocket is known and docking is performed at this site, we also consider the case where the binding pocket for a ligand is not known. In this application, termed blind docking,[19] the entire macromolecule is considered for ligand interactions. For each of these cases, we perform testing on 14 ligand−protein complexes for which the binding mode is known. We show that by enforcing the MES, the efficiency of the docking algorithm can be improved.

CALCULATION OF THE MES

The SR Algorithm. The point of the MES calculation is to define a surface that encapsulates the macromolecule and separates binding pocket volumes from those outside the macromolecule. The observation that concave regions within a protein characterize binding pockets suggests a surface with some restraint such that the surface cannot curve "inside" into binding pocket space. We can then consider the macromolecule to have a minimum surface covering the "outside" of the molecule and adjust this minimum surface such that the restraint is satisfied. In this respect, the minimum surface describes the minimum volume that can be encompassed by the MES.

Using known definitions, this could be accomplished by considering the solvent-accessible surface[20] to represent the

minimum surface. We could then create the MES by expanding this surface such that a curvature restraint (based on a differential geometry definition[21]) is satisfied. We have used the fact that we do not want the surface to curve under and into binding pockets to create a simpler algorithm that is applicable in a discrete setting. In this approach, the surface is considered to be a set of points where each point is restricted to lie on a given ray $[\theta, \phi]$ in spherical coordinates defined by an origin at the macromolecule center of mass. Redefining the surface restraint and minimum surface under this restriction as given below, we calculate the MES by starting with an initial surface that satisfies the restraint, and compressing this surface toward the minimum surface under the assertion that the surface restraint is satisfied.

The initial surface is a sphere, centered at the macromolecule center of mass (O), and tessellated at a resolution (R) with the set of points I $[\rho = ||\boldsymbol{a}||, \theta, \phi]$ where $||\boldsymbol{a}||$ is set such that the surface encompasses the entire van der Waals surface. The sphere is tessellated at evenly distributed intervals for θ and ϕ such that the resolution is satisfied in an approximate manner:

$$(1)\phi_{\text{int}} = \pi/\left(\text{int}\left(\left(\frac{\pi}{2\cdot\sin^{-1}[R/(2\cdot||a||)]}\right)\right) + 1\right) \quad (1)$$

$$\theta_{\text{int}}(\phi) = 2\pi/\left(\text{int}\left(\left(\frac{2\pi}{2\cdot\sin^{-1}[R/(2\cdot||a||\cdot\sin\phi)]}\right)\right) + 1\right) \quad (2)$$

During tessellation, neighbors to a surface point are stored in the surface point data structure so that the surface restraint can be enforced in a discrete manner. Each surface point is assigned two top neighbors, two side neighbors, and two bottom neighbors.

Using the van der Waals hard sphere approximation[22] for steric considerations, we define the minimum surface under the spherical coordinate restriction as the set of surface points M $[\rho, \theta, \phi]$ indexed for each surface point i and parametrized by the shell space (S). The shell space is the smallest distance allowed between the minimum surface and the van der Waals surface of the macromolecule (Figure 1). The minimum surface can then be calculated as the set of points $M_i[\rho_i, \theta_i, \phi_i]$ where ρ_i is the set minimum of P indexed for each macromolecule atom j. That is, for each surface point i, its position relative to each macromolecule atom j is considered to find the minimum value for ρ. If we let $\boldsymbol{b_j}$ represent the vector between I_i (the initial surface point) and the center of macromolecule atom j, $\boldsymbol{a}$ represents the vector between the I_i and O, and vdW(j) represents the van der Waals radius of atom j (Figure 2), then P can be calculated as

$$\rho_j \in P,\ \rho_j = ||a|| - \left(\frac{b_j^T a}{||a||} - (S_j'^2 - ||c||^2)^{1/2}\right) \quad (3)$$

$$S_j' = \text{vdW}(j) + S \quad (4)$$

$$c_j = b_j - proj_a b_j \quad (5)$$

The minimum surface provides a terminating condition so that the MES cannot overlap with the macromolecule. No change in the resulting MES will be observed regardless of whether it is calculated as the solvent accessible surface (with an appropriate probe radius) or by the equations presented above.

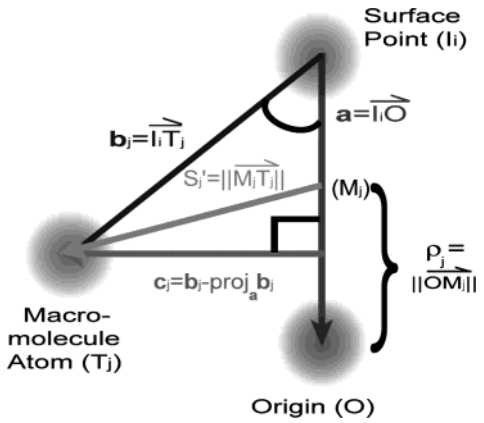

Figure 2. Vector diagram illustrating the calculation of the minimum surface. Within the minimum surface, a surface point i is placed at the minimum value for ρ found during calculation using each macromolecule atom j.

The surface restraint controls the compression of the surface toward the origin such that the MES does not enter binding pockets. In spherical coordinates, we define the surface restraint as a limit on $\Delta\rho$ based on the separation between surface points which is expressed in terms of $\Delta\theta$ and $\Delta\phi$. Because this separation is dependent on ρ at constant θ and ϕ, the limit for $\Delta\rho$ is different in the direction of increasing and decreasing ρ. Therefore, the general form of the restraint can be expressed in terms of $f(\rho)$ which provides the maximum decrease in ρ as a function of θ and ϕ and $g(\rho)$ which describes the maximum increase and is consistent with $f(\rho)$.

$$f(\rho,\phi,\theta) \le d\rho \le g(\rho,\phi,\theta) \quad (6)$$

$$f(\rho) = -g[\rho - f(\rho)] \quad (7)$$

Because the surface restraint should be uniform around a sphere at a given ρ, $\Delta\theta$, and $\Delta\phi$ are expressed in terms of ω which is the angle between the rays originating at the origin and passing through $[\theta_2,\phi_2]$ and $[\theta_1,\phi_1]$:

$$\omega(d\theta,d\phi) \equiv 2\cdot\sin^{-1}\left(\frac{2 - \cos(d\theta) - \cos(d\phi)}{2}\right)^{1/2} \quad (8)$$

Using this definition, we define the surface restraint function $f(\rho)$ as

$$\Delta\rho(\Delta\theta,\Delta\phi) \ge \frac{2\cdot\rho\sin\left(\frac{\omega}{2}\right)\sin\alpha}{\cos\left(\alpha - \frac{\omega}{2}\right)\omega} \quad (9)$$

which is derived from Figure 2. The restraint describes the maximum drop (*compression*) of a surface point $S_2[\rho_2,\theta_2,\phi_2]$ relative to a neighboring surface point $S_1[\rho_1,\theta_1,\phi_1]$ normal to the chord with endpoints $[\rho_1,\theta_2,\phi_2]$ and $[\rho_1,\theta_1,\phi_1]$. It is designed to create a restraint that is uniform at different values for ρ, despite the changing impact of $\Delta\theta$ and $\Delta\phi$ on surface point spacing. The angle α in eq 9 is an adjustable *compressibility angle* that can be used to control the degree

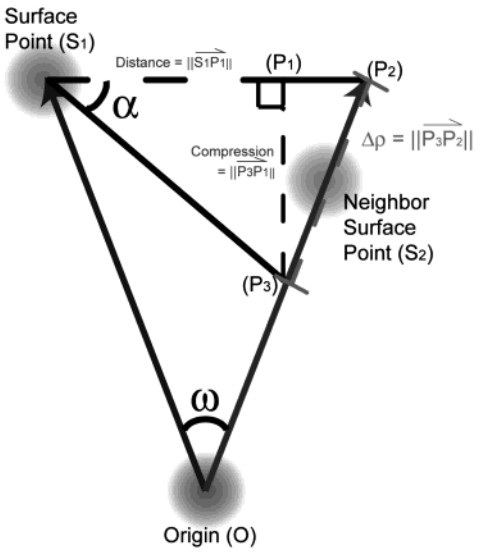

Figure 3. Illustration of the surface restraint [$\Delta\rho(S_1,S_2,\omega)$] for the MES based on two neighboring surface points. The surface is adjusted based on the value given for α.

to which the surface can drop into concave regions as described in Figure 3:

$$\alpha = \tan^{-1}\left(\frac{compression}{dist}\right) \qquad (10)$$

By adjusting α, the MES can be adjusted to gradually change the volume for a given binding pocket (increasing α will decrease binding pocket volume).

Because the algorithm starts with a surface that satisfies the surface restraint, the MES can be calculated iteratively by decreasing ρ for each surface point such that the surface restraint is always satisfied. Therefore, the function $g(\rho)$ does not need to be evaluated. The terminating condition for the algorithm occurs when no surface point can move, either because it has reached its point on the minimum surface or because the surface restraint (which is calculated based on all 6 neighbors) prevents further movement. There is no requirement that the macromolecule be homeomorphic to a sphere in order for the algorithm to work, over, the restriction does exist that the surface must be describable by a function in spherical coordinates.

The growth rate for the algorithm $O(n \cdot p)$ is limited by the minimum surface calculation where n represents the number of surface points and p represents the number of macromolecule atoms. This complexity can be reduced to $O(n)$ by hashing based on the spherical coordinate angles of atoms in the macromolecule, however, in its current form, the runtimes allow for MES calculation to be performed on very large protein databases within hours on modern PCs.

The algorithm has been implemented in C++ and supports output of the MES either as a triangulated surface (where two triangles are generated for each vertex—one with its top neighbors and one with a top and side neighbor) or as a PDB formatted file where each vertex is represented as an encapsulating atom. An example of the MES calculated for lactate dehydrogenase (1LDG.pdb) using the SR algorithm described above is shown in Figure 4.

The GB Algorithm. The second algorithm implemented for MES calculation is grid based (GB) and considers binding pockets as the space between atom pairs in a concept

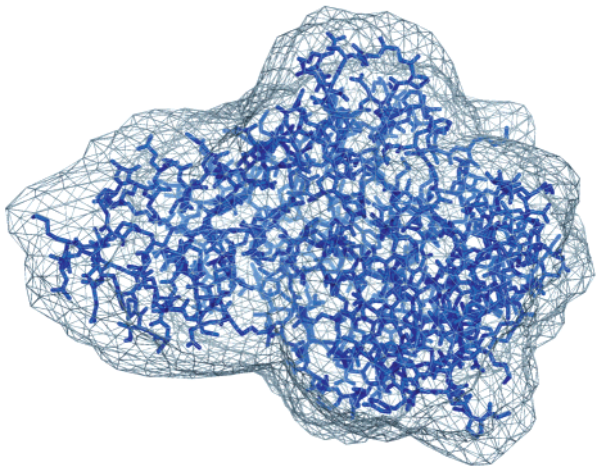

Figure 4. The MES calculated for lactate dehydrogenase (1LDG) using the SR algorithm with α = 0.7.

analogous to that of the programs Pocket[15] and LigSite.[16] Both Pocket and LigSite operate on a regular grid lattice and define binding pocket space as grid points lying between macromolecule atoms. In Pocket, a grid point is determined to be between two macromolecule atoms if there exists atoms positioned on both sides of the grid point in either of the xy, xz, or yz planes. Therefore, the software suffers from a dependence of macromolecule orientation with respect to the Cartesian axes, and improperly oriented macromolecules can result in pockets that are improperly characterized or not found at all.[16] LigSite reduces this dependence by also scanning along the four cubic diagonals. Here, we have removed this dependence by considering a grid point to be part of a binding pocket if it lies between any two atoms, regardless of the atoms' positions relative to planes defined in Cartesian space. According to this definition, *any* concave space within the macromolecule will be considered to be binding pocket volume and therefore a binding pocket *diameter* factor is introduced to limit the maximum distance between atoms that are considered to compose binding pocket space.

The algorithm is performed on a regularly spaced grid lattice (with the grid spacing represented as *resolution*) encompassing the macromolecule or the area of interest. Each grid point is marked as empty, creating the initial surface, which is null. The minimum surface is calculated by *filling* all grid points that are overlapped by a macromolecule atom's van der Waals radius. The next step in the algorithm involves creating cylindrical segments with endpoints at all atom pairs. Atom pairs cannot be restricted to only solvent accessible atoms, due to empty space within the macromolecule that is not solvent accessible. The length of such a segment is set to the distance between the two atom centers minus the van der Waals radius of each atom. Any segment whose length is larger than the *diameter* factor is rejected. The radius of each segment is set to the maximum of the radii of the atoms that form the segment endpoints. Creation of the MES then occurs by *filling* any grid points that are overlapped by the cylindrical segments. The remaining *empty* grid points represent the outside of the molecule, and the MES which separates empty grid points from those that are filled can be calculated by triangulation at these points. Incorporation of the shell space (Figure 1) can be accomplished by adding the shell space parameter to the radii of all cylindrical

segments. The result is an alternative MES algorithm where binding pocket volume is controlled by a diameter factor instead of a compression factor. Using a diameter factor of infinity, the volume encompassed by the MES can be no greater than that encompassed by the convex hull, and a decrease in the diameter factor is correlated with a decrease in the encompassed volume.

The algorithm, while simple, suffers from a time complexity of $O(n \cdot p^2)$ for n grid points and p macromolecule atoms. This poses a serious problem, as for a 5000-atom protein there are 25 000 000 segments. If the grid has 10 000 lattice points, then $2.5 \cdot 10^{11}$ segment intersection evaluations are required. Therefore, the following steps are taken to decrease the run time:

First, a sweeping plane[23] is used for minimum surface calculation, such that only those macromolecule atoms lying within a range in the x dimension capable of intersecting a set of boundary atoms lying at a specified x coordinate will be evaluated. Second, in cases where the grid does not encompass the entire macromolecule (i.e. active-site centered grids), all segments that do not pass through the grid are rejected. Third, a sweeping plane approach is also used for segment intersection checks such that only those segments passing through the x coordinate of the macromolecule atom center are evaluated. Finally, and most significant to run time, is the inactivation of any segments where the grid point at the segment center has been removed in minimum surface calculation. This approximation is justified because it suggests that two smaller segments approximately overlap the inactivated segment.

Testing the MES. It is difficult to decide when MES calculation is successful due to the fact that the binding pocket is a subjective concept. Certainly, however, we can determine failures in MES generation when the surface intersects with ligands in known binding modes. Therefore, we tested each algorithm on a set of structures of enzyme−ligand complexes by removing the ligand, calculating the MES for the macromolecule, replacing the ligand, and testing for intersection with the MES. Ideally, a set of parameters for which the MES provides a suitable boundary for the majority of known binding pockets can be found.

We tested both algorithms on a set of 80 enzyme−ligand complexes with unique enzyme classifications. The PDB codes for the 80 structures were 1A16, 1EJN, 1FKF, 1HVR, 1STP, 2CPP, 2MCP, 2YPI, 3PTB, 4DFR, 4FUA, 4HMG, 4STD, 5CNA, 8GSS, 1A0L, 1A42, 1ADD, 1ADS, 1AQL, 1B3N, 1BIB, 1BLL, 1BYB, 1CSH, 1CZI, 1DWB, 1ELA, 1ENU, 1FMP, 1FPP, 1FUT, 1GAI, 1GKY, 1GOS, 1GOX, 1GPB, 1GTX, 1HMY, 1HYT, 1LDG, 1MBI, 1MNS, 1OYB, 1PBE, 1PDA, 1PHP, 1QPN, 1RBP, 1ROB, 1RPA, 1RTF, 1SMR, 1SNC, 1STO, 1TDE, 1TLP, 1ULB, 2ACK, 2CND, 2CTC, 2CUT, 2DHC, 2DKB, 2ER9, 2HCK, 2NPX, 2SIM, 2XIS, 2YHX, 3CLA, 3PGM, 3SGA, 4PAD, 4TMK, 5ENL, 5P2P, 8EST, 1B15, and 9NSE. Aside from the 1FPP test case (where ligand binding involved multiple subunits), one subunit was isolated from each test case for MES calculation. The assignment of van der Waals radii were made according to those reported by Li and Nussinov.[24] Each algorithm was tested using a shell space of 0 and 2 Å using a resolution of 0.5 Å. The SR algorithm was tested at α values ranging from 0 to $\pi/3$, and the GB algorithm was tested using *diameters* ranging from 0 to 15 Å. Intersection between the MES and ligand atoms was tested by checking each triangle comprising the MES for intersections with each sphere defined by the van der Waals radius of each ligand atom.

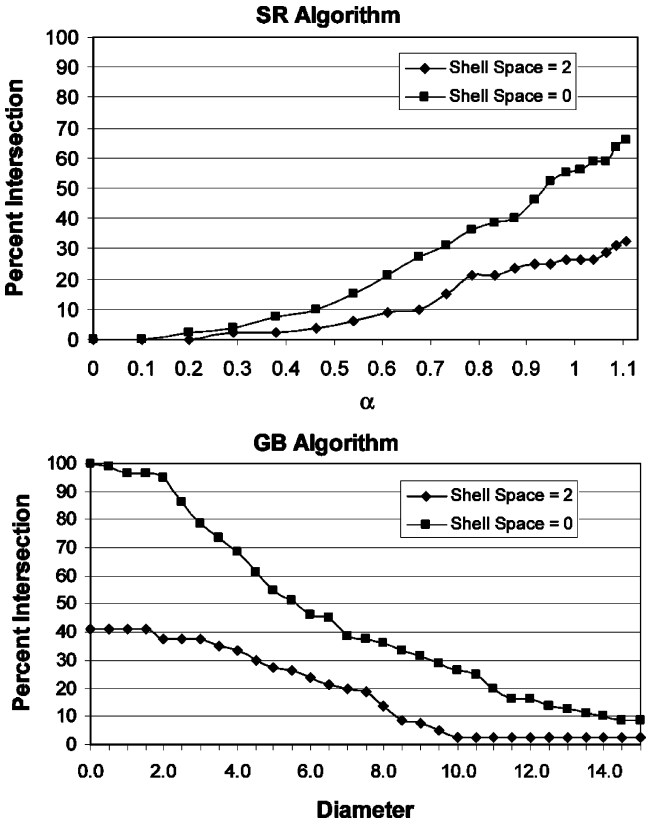

Figure 5. Results from MES testing: (top) percentage of ligands intersecting the MES as calculated with the SR algorithm at varying values for α ($n = 80$); (bottom) percentage of ligands intersecting the MES as calculated with the GB algorithm at varying diameter restraints ($n = 80$).

The results from the testing of the SR algorithm and the GB algorithm are shown in Figure 5. Using $\alpha = 0.7$ and a shell space of 2 Å, the SR algorithm performed well for 90% of the cases. In these cases, the MES provides a boundary for each known binding pocket that is sufficient to encapsulate the ligand. For values of α below 0.7, the percent intersection gradually decreases to zero; however, the surface becomes poorly defined around the macromolecule. That is, as α approaches zero, the surface becomes more "sphere-like", and while this MES is sufficient to provide a boundary to the known binding pocket in question, it does not provide boundaries to other potential binding pockets within the macromolecule. Therefore, if the binding pocket is not known a priori, the MES is unusable at values for α below 0.7 for most macromolecules.

The cases for which the SR algorithm failed involved shallow binding pockets with large openings. In these cases, it is difficult to define a surface restraint that will not "dip in" to the binding pocket while maintaining the ability to wrap around the rest of the macromolecule. The GB algorithm does not suffer from these problems and generates a MES that is much more general for the definition of binding pocket boundaries. At a shell space of 2 Å and a diameter of 10, the calculated MES is suitable for 97.5% of the test cases. Therefore, we have chosen the GB algorithm with these parameters as the default for MES calculation.

The two test cases that failed under these conditions (1FKF and 3SGA) were proteases complexed with peptide-based inhibitors that protruded out and above their surrounding atoms. For the 3SGA, test case, a suitable MES can be calculated by adjusting the diameter to 18 Å. The 1FKF test case, however, presents a ligand-binding mode with ligand atoms that are not contained within any concave space in the macromolecule. Therefore, no suitable MES can be calculated based on the algorithms presented in this manuscript.

Run times for both algorithms were recorded on a 1.9 GHz Pentium 4 Dell Precision WorkStation 340 running Red Hat Linux 8.0. The run times for the SR algorithm were recorded using $\alpha = \pi/4$ and shell space of 2 Å and for the GB algorithm using a diameter of 10 Å and a shell space of 2 Å. The average run time for the SR algorithm (using all 80 test cases) was 1.5 s with all but 1 test case requiring less than 4 s. The run times for the GB algorithm were similar with an average of 1.6 s and a maximum of 3.88 s for all test cases.

APPLICATION TO DOCKING

Genetic algorithms (GAs) have been shown to be efficacious for flexible ligand docking in a number of software applications,[18,25−29] typically via optimization of an objective function describing the binding affinity of a ligand. Here, we ask whether the efficiency of the GA can be improved by "helping" the algorithm to differentiate ligand conformations that lie outside the binding pocket. Presumably, by forcing high fitness energies onto conformations that lie outside the binding pocket, the populations will quickly converge toward conformations whose atoms are inside the binding pocket. This should limit the fitness evaluation of rotation and conformation degrees of freedom to relevant areas of the macromolecule. This is tested here via modification of AutoDock 3.0 to incorporate a MES boundary during search.

AutoDock 3.0 is a program for flexible ligand docking that can be parametrized to run dockings based on the Metropolis algorithm, genetic algorithm (GA), or Lamarckian genetic algorithm (LGA).[18] The LGA adds a local search operator parametrized by frequency in order to improve performance at local minima. The program uses precalculated energetic grids (calculated by AutoGrid 3.0) for interatomic energy evaluations. The dimensions of the grid determine the range of translations for the search space. For a given docking "job", AutoDock is parametrized to perform a certain number of runs. At completion, the best conformations from each run are clustered into bins based on similarity measured by RMSD and binding energy.

Prior to publication, AutoDock was validated via reproduction of structural data from 7 ligand−enzyme complexes previously determined by X-ray crystallography (1HVR, 1STP, 2CPP, 2MCP, 3PTB, 4HMG, 4DFR). We have used these 7 test cases to evaluate the impact of the MES boundary on docking efficiency along with an additional 7 test cases (1A16, 1EJN, 1PHP, 2YPI, 4FUA, 4STD, 5CNA). Each case was tested with and without the MES using the GA and the LGA. For the LGA, two types of tests were considered. In the first, testing is based on an energetic grid 22.5 $Å^3$ in volume located at the crystallographic ligand structure's center of mass. In the second, we consider the potential use of AutoDock for ligand docking when the binding pocket is not known a priori. In these cases, the grid is dimensioned according to the bounding box for the *entire* enzyme.

We have chosen to enforce the MES boundary within AutoDock in an energetic fashion via the creation of artificial repulsive energies outside the binding pocket. Because these energies can be precalculated into the potential grids, this approach requires no additional computational time for boundary enforcement during the docking. The GB algorithm for calculation of the MES occurs on a grid, and therefore the most straightforward method to enforce the boundary might be to set each grid point lying outside the MES with a constant high-energy value. However, a constant energy outside of the MES provides no gradient information for local search algorithms, and the decision on the value for the constant energy is not straightforward. Therefore, we have considered an approach where grid points determined to be outside the MES according to the GB algorithm are considered to be *encapsulating atoms.* It is hoped that this approach will provide an energetic surface outside the MES that is similar in nature to the steric boundary created by macromolecule atoms.

The encapsulating atom approach is taken into account during MES calculation by considering each grid point to be a hard sphere with a specified van der Waals radius. Then, any encapsulating atom intersecting a macromolecule atom or cylindrical segment is removed, leaving a set of encapsulating atoms lying outside of the macromolecule binding pocket. The *repulsive* van der Waals energies from the encapsulating atoms are incorporated into AutoDock 3.0 by addition of a repulsive Lennard-Jones 12−6 boundary term (ΔG_{MES}) into the free energy equation:

$$\Delta G_{MES} = \Delta G_{vdW} \sum \left(\frac{A_{ic}}{r_{ic}^{12}} - \frac{B_{ic}}{r_{ic}^{6}} \right)$$

$$\Delta G_{MES} = 0, \text{ if } \Delta G_{MES} > 0 \qquad (11)$$

The term is identical to the van der Waals term within the AutoDock force field with the exception that the energy is truncated at 0 to prevent any negative potentials between ligand atoms and encapsulating atoms. In the equation, *A* and *B* represent atom type parameters as specified in ref 18, *r* represents the distance from the ligand atom center to the encapsulating atom center, and ΔG_{vdW} represents an empirically determined coefficient for all van der Waals interactions.[18] The summation occurs over all encapsulating atoms indexed at *c* for each ligand atom indexed at *i*.

The encapsulating atom representation of the MES is easily incorporated into AutoGrid 3.0. The encapsulating atoms are appended onto the macromolecule PDB file, and a special atom type is given to the boundary atom. The force field need only be modified to recognize this atom type and reject its term in the summation if the energy is negative. Here, we have used a van der Waals radius of 2 Å and Lennard-Jones parameters (*A* and *B* in eq 11) equal to those used for carbon. The MES for each test case was generated, *without the ligand in place*, using the GB algorithm with default parameters (*diameter* = 15, *shell space* = *2*). The algorithm was run using a resolution of 0.5 Å; however, for any two encapsulating atoms within 2.5 Å of each other, one was

Table 1. Reference Ligand Binding Mode Data

PDB code	RMSD (Å)	docked energy (kcal/mol)	free bond torsions
1HVR	0.711	−21.37	10
1STP	0.599	−9.65	5
2CPP	0.844	−7.42	0
2MCP	0.956	−5.94	4
3PTB	0.539	−8.48	0
4DFR	1.051	−11.76	11
4HMG	0.767	−7.58	10
1A16	1.286	−8.61	4
1EJN	0.570	−11.78	8
1PHP	0.830	−13.31	9
2YPI	0.995	−6.05	3
4FUA	0.764	−6.04	5
4STD	0.988	−6.62	4
5CNA	0.969	−7.41	6

removed so that no two encapsulating atoms were closer than 3 Å. The MES was generated only within the grid whose dimensions matched those of the grids used here for the AutoGrid calculations. An example of the impact of the MES boundary on the potential surface for a carbon ligand atom can be seen in Figure 6 where the isosurface at 2.5 kcal·mol^{-1} is plotted for dihydrofolate reductase.

GA efficiency was compared between the MES and control cases by analysis of averages of the final number of correct answers, the final best energies, the energies of correct individuals, the final number of conformational cluster bins, and the population of the top 10 bins. The stochastic nature of the GA and LGA makes comparison difficult due to statistical sampling error. Therefore, we have attempted to reduce this error with the use of random number generator seeds. In this manner, we can compare the effects of the MES when the same initial population is used in the MES case and in the control. The docking methods, including enzyme and ligand preparation, were identical to those reported in ref 18 aside from a difference in the energy evaluation limit ($2.5 \cdot 10^5$ used here) and number of runs (40 used here). For each job type consisting of 40 runs, 40 jobs were executed to get a measure of reproducibility.

An additional problem with GA analysis, at least in this case, is that the determination of a "correct" answer is somewhat ambiguous. This stems from the fact that the search space is infinite, that structures determined by crystallography have significant uncertainty, and that the force-field used here for docking relies on certain simplifications in order to make run-times acceptable. For this analysis, a *reference structure* ligand-binding mode was determined for each test case and assumed to represent the global minimum according to the force field. The reference structure was generated using the best conformation found from a docking job consisting of 200 runs and $1.5 \cdot 10^6$ energy evaluations. The dockings used to find the reference structure were performed without incorporation of a MES boundary. The RMSDs of each reference structure from the crystallographic conformations along with the docking energy and number of rotatable bonds considered by AutoDock are listed in Table 1. For the analyses presented below, a conformation was considered to be "correct" if the conformation was within 1 Å RMSD from the reference compound and had an energy no more than 1 kcal·mol^{-1} greater than the reference energy.

RESULTS

For the genetic algorithm, substantially improved results were obtained when MES boundaries were enforced. As expected, results from the 2CPP test case were invariable due to the fact that the MES surface did not pass through the energetic grids. For three of the test cases (1PHP, 1STP, and 4DFR), there was over an 80% increase in the number of correct individuals, and there were significant increases in the others (Figure 7). In the extremes of improvement, there was, based on averages of 40 jobs (1600 runs), a 92% increase in correct individuals (1STP), a 1.62 kcal·mol^{-1} decrease in best energies (1STP), a 21% decrease in the number of conformational bins (2MCP), and a 26.6% increase in population of the top 10 bins (2MCP) [each value taken as the best from the 14 test cases]. The energies for correct individuals were relatively invariant between the control and MES cases.

For the LGA with default grid sizes, much less dramatic results were seen. However, significant improvements were seen in two of the test cases—a 23.8% increase in correct individuals for the 4DFR test case and a 91.2% percent increase for the 1PHP test case (Figure 8). In the 1PHP control case, 18 conformations out of the 40, on average, were within an RMSD of 1 Å but failed to reach the energy threshold of the reference structure (which was also calculated without the MES boundary). The deficiency in energy in these cases was almost entirely due to incorrect positioning of oxygen atoms on the terminal diphosphate of the adenosine diphosphate ligand. While it is tempting to believe that the artificial boundary forced correct positioning of these atoms in the MES cases, the terminal diphosphate is distal from the surface and is positioned deep within the binding pocket. Nonetheless, the average number of correct individuals increases from 20.48 to 39.15 when the MES boundary is used. This is accompanied by an average 1.28 kcal·mol^{-1} decrease in best energies, a 0.91 kcal·mol^{-1} decrease in energy of correct individuals, a 46.7% decrease in conformational bins, and a 5.86% increase in population of the top 10 bins.

For the LGA applied to blind docking (where the entire protein was searched for ligand binding), the most dramatic results were seen (Figure 9). For five of the test cases, the number of correct individuals was doubled when the MES boundary was used. This improvement is significant, as for three of the test cases (4DFR, 1EJN, and 1HVR), the average number of correct individuals did not reach 1 in the control. The tests failed for only two of the test cases (2MCP and 4HMG). In these cases, the respective ligands found lower binding energies in different binding pockets within the proteins. Therefore, no comparison was made between the control and MES cases for these protein complexes. For the other test cases, in the extremes of improvement there was, based on averages of 40 jobs, a 1190% increase in correct individuals (1HVR − 0.28 vs 3.55 correct in the MES case), a 1.8 kcal·mol^{-1} decrease in best energies (1HVR), a 0.18 kcal·mol^{-1} decrease in the average energy of correct individuals (1PHP), a 45% decrease in the average number of conformational bins (2CPP), and a 56% increase in the population of the top 10 bins.

There is no simple trend between the increase in docking efficiency and ligand degrees of freedom or volume outside

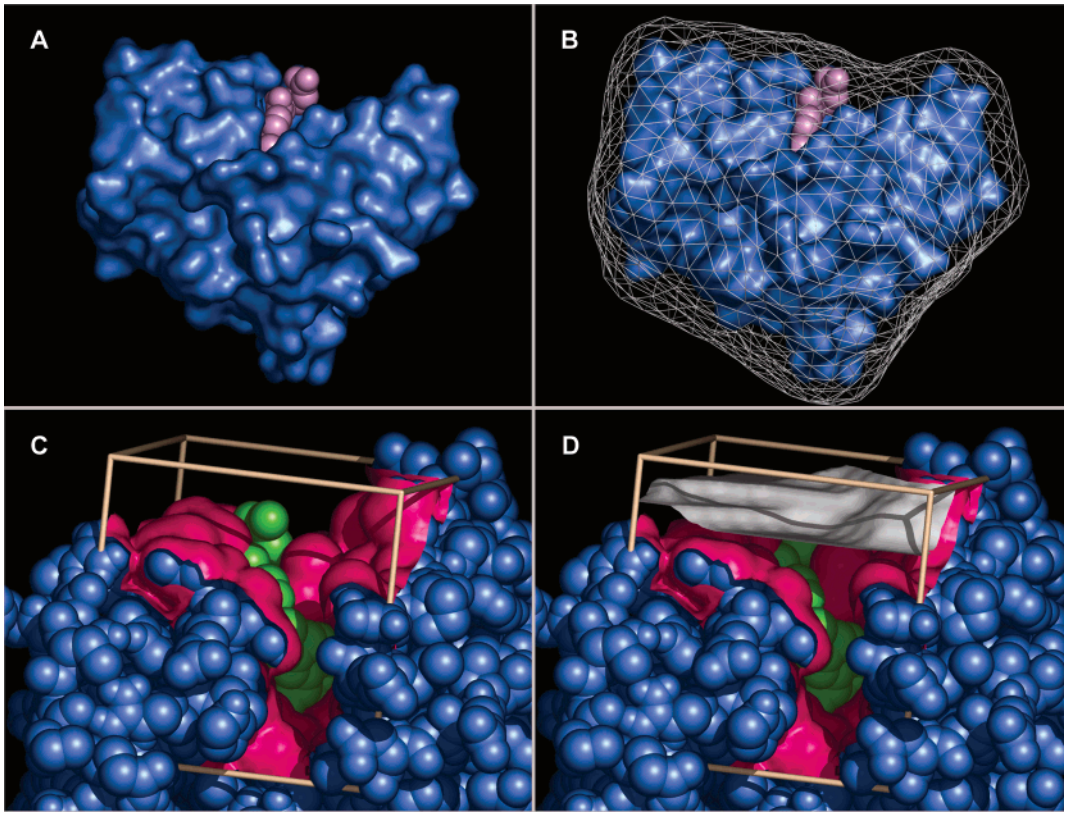

Figure 6. Dihydrofolate reductase (4DFR) and the calculated MES rendered using PyMol:[31] (a) the solvent accessible surface for the protein (blue) with the ligand CPK model shown in purple; (b) the MES calculated for the macromolecule using the GB algorithm with a diameter of 10 Å (white); (c) the isosurface (pink) calculated at 2.5 kcal·mol^{-1} for a carbon probe atom using AutoGrid 3.0 [The protein CPK model is shown in blue with ligand atoms in green. The boundaries for the docking grid are shown in beige. Any carbon atom whose center is placed "below" the isosurface experiences a potential greater than 2.5 kcal·mol^{-1}.]; and (d) the isosurface calculated using MES boundaries. The difference is plotted in white for clarity. Any carbon atom whose center is placed "above" the white isosurface experiences a potential greater than 2.5 kcal·mol^{-1}. Isosurfaces were calculated using the marching cubes algorithm.[32]

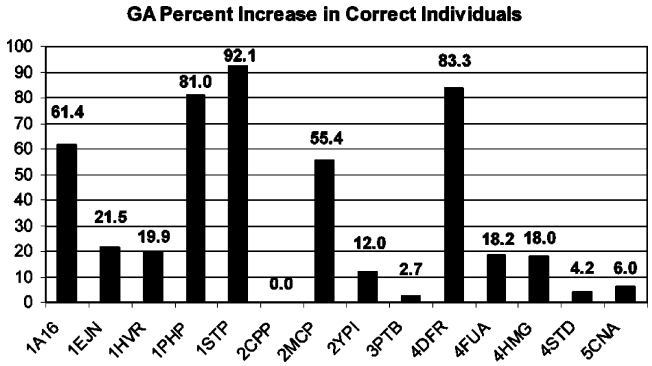

Figure 7. Percent increase in the average number of correct individuals for the GA when the MES boundary is enforced ($n = 40$). The control and the MES test cases are run with the same initial starting populations for 5000 generations.

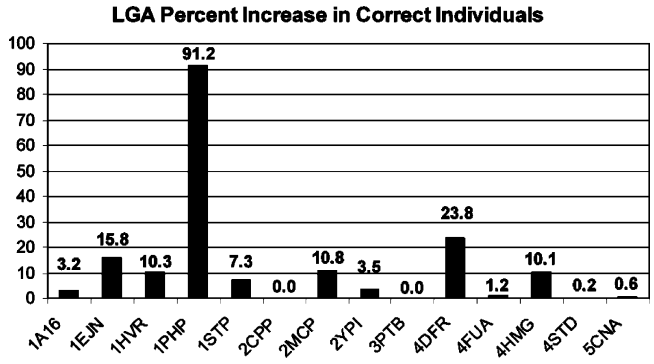

Figure 8. Percent increase in the average number of correct individuals for the LGA when the MES boundary is enforced ($n = 40$). The control and the MES test cases are run with the same initial starting populations for ~2.5·10^5 energy evaluations.

of the macromolecule (estimated by summing the volume of empty grid points). However, the test cases where the largest impact was seen generally involved a relatively large number of ligand degrees of freedom (1HVR, 4DFR, and 1PHP). The influence on docking efficiency is highly sensitive to the parameters used for MES calculation (data not shown), and therefore algorithms capable of calculating input parameters based on protein topology would be beneficial. It is also noteworthy that no optimization of genetic algorithm parameters for use with MES boundaries was performed. It is expected that adjustment of parameters for MES incorporation will offer further improvement in the results; however, these experiments have not been performed.

We also tested the idea of population seeding using the assertion that the center atom of every individual in an initial population was positioned within the binding pocket. The results were indistinguishable from the controls in most cases, however, and therefore the data pertaining to population seeding was left out.

DISCUSSION

We have presented in this manuscript two distinct algorithms designed for computation of artificial binding pocket boundaries which can be used in software where a limit to the volume considered for ligand atom occupancy is desired.

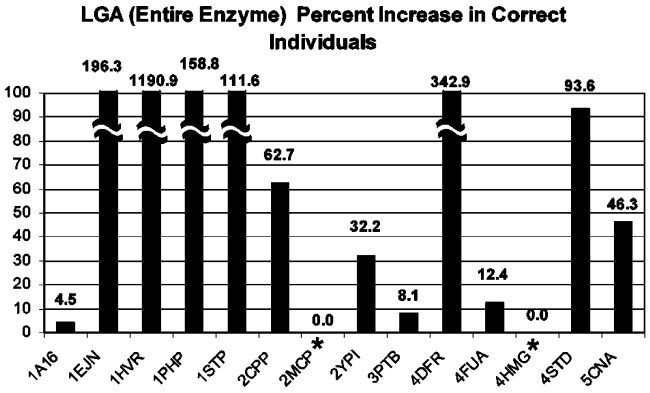

Figure 9. Percent increase in the average number of correct individuals for the LGA when the entire protein is considered for ligand interactions using the MES boundary ($n = 40$). The grid dimensions are set to the bounding box for the protein, and the MES is calculated without the ligand in place for all test cases. *For the 2MCP and 4HMG test cases, the ligands assume binding modes with lower energies in different binding pockets than those identified by crystallography and therefore no comparisons were made.

Parameters for the algorithms were determined based on the sole criterion of a calculated MES that effectively encapsulates the ligand for the majority of test cases. Based on this criterion, the GB algorithm provides a general method for boundary definition that does not require excessive volumes in order to work for a large percentage of test cases. While we have shown that the MES calculated by this method is sufficient to increase the efficiency of flexible ligand docking, we have not considered many of the other important issues in binding pocket characterization (ligand volume vs binding pocket volume, identification of accessible macromolecule atoms for ligand interaction, etc.). Because such analyses require description of a method for calculation of a binding pocket surface, we have left them for a subsequent paper.

We will, however, describe briefly the concept for calculating the binding pocket surface from the MES and potential advantages of such an approach for binding pocket characterization. The MES encloses a volume accessible to ligand atoms, and therefore the binding pocket surface (both van der Waals and artificial) can be defined in an analogous manner to the solvent accessible surface[20] as the surface created by rolling a probe sphere around the binding pocket and the MES. Using this definition, the binding pocket surface can be calculated with only slight modification to existing surface algorithms. Then, the MES can facilitate binding pocket characterization and, possibly, binding-site prediction. We have implemented the approach in the software *Binding Pocket Surveyor*. The software which provides the ability for MES calculation and binding pocket characterization will be made freely available following a publication of methods.

Most algorithms previously described for binding pocket characterization work by evaluating a descriptor on either a set of grid points[9,14−17] or on a set of probes[11−13] (i.e. spheres placed tangential to protein atoms or molecular fragments placed around the macromolecule). The descriptors used by these algorithms have included probe interaction energies,[12] accessibility parameters (surface accessibilities,[9] burial counts of nearby protein atoms,[13] or volumes that become inaccessible due to atom fattening[14]), an angular condition that identifies spheres within concave regions,[11] or the presence of surrounding protein atoms collinear with the grid point.[15−17] Most of these algorithms, like the SR algorithm presented here, admittedly suffer from failures in cases where the binding pocket has a large opening or mouth.[10] Accessibility parameters, angular limitations, and probe interaction energies seek regions of the macromolecule where "tight packing" with ligand atoms can occur. The problem, as mentioned earlier, is that all ligand atoms need not exhibit tight packing with the macromolecule. For algorithms that use atom fattening or the discrete flow approach,[10] binding pockets cannot be identified if among the infinite number of possible cross sections of a pocket, one is not larger than the mouth opening of the pocket. For the discrete flow implementation in CAST,[10] this corresponds to a failure in 22% of the test cases.

Therefore, the concept used by the binding pocket characterization programs Pocket and LigSite might provide the most general method for binding pocket characterization as these algorithms do not suffer from these problems. However, the binding pocket characterization in these algorithms is dependent on the orientation of the macromolecule with respect to the Cartesian axes. Because the GB algorithm does not suffer from this dependence on orientation, it holds promise as a method for binding pocket characterization. We have shown, herein, that the GB algorithm can successfully encapsulate ligand atoms for 97.5% of the test cases using default parameters and therefore should not miss the identification of any accessible atoms needed for ligand interactions. This still leaves to be addressed, however, the issues of identifying a *minimum* number of accessible atoms and a *minimum* volume needed for ligand binding.

An additional asset of the algorithms presented in this paper is the ability to gradually adjust the volume considered to be the binding pocket according to experimental need. The sensitivity of adjustment in the GB algorithm is dependent on the distances between atoms in the binding pocket and macromolecule and thus to the topology of the protein. Therefore, the SR algorithm might still find use in cases where a fine control over binding pocket space is desired. For the default parameters presented in this paper, adjustment was required to encapsulate the ligand volume for one test case. In the other test case, however, no suitable MES could be defined as ligand atoms occupied space outside of the convex hull of the macromolecule (ligand atoms were positioned outside of any concave space). In both cases, the macromolecules were proteases and the ligands were peptide-based. Because proteases exhibit catalytic activity through protein−protein interactions, substrate-derived ligands might be expected to contain atoms which are not directly involved in binding (as is the case for the natural substrates) and cannot be encapsulated by binding pocket volume. For these cases, the MES should not be used.

We have demonstrated the use of the MES to improve the efficiency of flexible ligand docking in the genetic algorithms applied in AutoDock. The stochastic nature of the GA and the expensive CPU cost for fitness evaluations makes GA analysis difficult. Exact mathematical models of the GA are limited to simple, impractical applications,[30] and therefore it is difficult to answer such questions as "How

can GA efficiency be improved?". In the spirit of traditional GA theory, however, it might be expected that the population would quickly converge toward translations within or near the binding pocket, as these individuals have binding affinities improved by many orders of magnitude. Indeed, observation of best energies as the GA progresses reveals a rapid initial drop in energies as steric overlap is alleviated. Based on these considerations, MES boundary enforcement would seem to have little impact on docking efficiency.

The key observation, however, is that ligand rotation and conformation degrees of freedom are not independent from translation variables. The optimum ligand conformation at one translation is almost certainly different from that at another translation. However, the fitness value makes no attempt to distinguish between translation and conformational fitness. It therefore seems reasonable that during convergence of translation variables toward the binding pocket, there is an associated convergence toward values for the rotation and conformational degrees of freedom that are not relevant to conformations where ligand atoms lie entirely within the binding pocket.

It is therefore presumed that enforcing MES boundaries during ligand docking helps to prevent the loss of relevant conformational variables during early convergence by distinguishing in an energetic fashion between those conformations that exist inside or outside the binding pocket. The LGA incorporates local search, which may help to reintroduce relevant conformation values that were lost during translation convergence. This would explain the less dramatic improvement seen for MES incorporation during LGA search. However, this also suggests that the impact of local search is lessened in the bound cases and that docking efficiency in these cases can be improved by lowering the frequency of expensive local searches. This seems to be the circumstance for the test cases 1HVR and 1STP (in data not reported); however, no full optimization involving all test cases was performed.

Incorporation of MES boundaries for flexible docking had the most significant impact when the entire enzyme was searched for ligand binding modes. This application allows for docking when a binding pocket is not known a priori or an unknown site with the potential for allosteric regulation is being considered. The validity of these blind docking experiments within AutoDock was first demonstrated by Hetenyi and van der Spoel using 8 protein/peptide ligand complexes.[19] The results presented here provide additional support for the feasibility of AutoDock as a binding-site identification tool using 14 protein/organic ligand complexes. For 12 out of the14 test cases, a binding mode within a 1 Å RMSD from the crystallographic structure was found. The fact that the blind docking results without the use of the MES (presented here and previously) are characterized by a very low percentage of successful runs indicates the importance of the MES for docking as a means of binding site identification.

ACKNOWLEDGMENT

Funding for this work was provided by the U.S. Army/DOD Breast Cancer Program, grant numbers DAMD17-00-0368 and DAMD17-00-0372.

REFERENCES AND NOTES

(1) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. Dock 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.
(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(3) Bohm, H. J. Site-Directed Structure Generation by Fragment-Joining. *Perspect. Drug Discovery Des.* **1995**, *3*, 21−33.
(4) Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. Pro-Ligand: An Approach to De Novo Molecular Design. 1. Application to the Design of Organic Molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 13−32.
(5) Trosset, J.; Scheraga, H. ProDock: Software Package for Protein Modeling and Docking. *J. Comput. Chem.* **1999**, *20*, 412−427.
(6) Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: A Computer Program for Identification of Drug Interaction Sites in Macromolecules and Drug Leads from Chemical Databases. *J. Comput. Chem.* **2001**, *22*, 1750−1771.
(7) Luo, Z.; Wang, R.; Lai, L. Rasse: A New Method for Structure-Based Drug Design. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1187−1194.
(8) Budin, N.; Majeux, N.; Tenette-Souaille, C.; Caflisch, A. Structure-Based Ligand Design by a Build-up Approach and Genetic Algorithm Search in Conformational Space. *J. Comput. Chem.* **2001**, *22*, 1956−1970.
(9) Stahl, M.; Bur, D.; Schneider, G. Mapping of Proteinase Active Sites by Projection of Surface-Derived Correlation Vectors. *J. Comput. Chem.* **1999**, *20*, 336−347.
(10) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. *Protein Sci.* **1998**, *7*, 1884−1897.
(11) Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(12) Ruppert, J.; Welch, W.; Jain, A. N. Automatic Identification and Representation of Protein Binding Sites for Molecular Docking. *Protein Sci.* **1997**, *6*, 524−533.
(13) Brady, G. P., Jr.; Stouten, P. F. Fast Prediction and Visualization of Protein Binding Pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383−401.
(14) Kleywegt, G.; Jones, T. Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **1994**, *50*, 178−185.
(15) Levitt, D.; Banaszak, L. Pocket − a Computer-Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino-Acids. *J. Mol. Graphics Modell.* **1992**, *10*, 229−234.
(16) Hendlich, M.; Rippmann, F.; Barnickel, G. LigSite: Automatic and Efficient Detection of Potential Small Molecule-Binding Sites in Proteins. *J. Mol. Graphics Modell.* **1997**, *15*, 359-&.
(17) Exner, T.; Keil, M.; Moeckel, G.; Brickmann, J. Identification of Substrate Channels and Protein Cavities. *J. Mol. Model.* **1998**, *4*, 340−343.
(18) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(19) Hetenyi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2002**, *11*, 1729−1737.
(20) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(21) Coxeter, H. *Introduction to Geometry*, 2nd ed.; Coxeter, H., Ed.; Wiley: New York, 1969; Chapter 19, pp 414−434.
(22) Richards, F. M. The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density. *J. Mol. Biol.* **1974**, *82*, 1−14.
(23) De Berg, M.; Van Kreveld, M.; Overmars, M.; Otfried, S. In *Computational Geometry*, 2nd ed.; De Berg, M., Ed.; Springer: New York, 2000; Chapter 2, pp 20−29.
(24) Li, A. J.; Nussinov, R. A Set of van der Waals and Coulombic Radii of Protein Atoms for Molecular and Solvent-Accessible Surface Calculation, Packing Evaluation, and Docking. *Proteins Struct. Funct. Genet.* **1998**, *32*, 111−127.
(25) Hou, T.; Wang, J.; Chen, L.; Xu, X. Automated Docking of Peptides and Proteins by Using a Genetic Algorithm Combined with a Tabu Search. *Protein Eng.* **1999**, *12*, 639−647.
(26) Oshiro, C.; Kuntz, I.; Dixon, J. Flexible Ligand Docking Using a Genetic Algorithm. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 113−130.
(27) Jones, G.; Willett, P.; Glen, R.; Leach, A.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(28) Taylor, J.; Burnett, R. Darwin: A Program for Docking Flexible Molecules. *Proteins Struct. Funct. Genet.* **2000**, *41*, 173−191.
(29) Yang, J.; Kao, C. Flexible Ligand Docking Using a Robust Evolutionary Algorithm. *J. Comput. Chem.* **2000**, *21*, 988−998.
(30) Nix, A. E.; Vose, M. D. Modeling Genetic Algorithms with Markov Chains. *Ann. Math. Artif. Intell.* **1991**, *5*, 79−88.
(31) *PyMol Software*; DeLano Scientific: San Carlos, CA, 2002.
(32) Lorensen, W. E.; Cline, H. E. Marching Cubes: A High Resolution 3D Surface Reconstruction Algorithm. *Comput. Graphics* **1987**, *21*, 163−169.

CI049853R