# A Molecular Mechanics / Grid Method for Evaluation of Ligand–Receptor Interactions

**BROCK A. LUTY, ZELDA R. WASSERMAN,***
**PIETER F. W. STOUTEN, and C. NICHOLAS HODGE**
*Computer-Aided Drug Design Group, DuPont Merck Pharmaceutical Company, Experimental Station, P.O. Box 80353, Wilmington, Delaware 19880-0353*

**MARTIN ZACHARIAS**
*Health and Science Center, University of Colorado, Denver, Colorado 80262*

**J. ANDREW McCAMMON**
*Department of Chemistry, University of Houston, Houston, Texas 77204-5641*

*Received 18 March 1994; accepted 8 August 1994*

## ABSTRACT

We present a computational method for prediction of the conformation of a ligand when bound to a macromolecular receptor. The method is intended for use in systems in which the approximate location of the binding site is known and no large-scale rearrangements of the receptor are expected upon formation of the complex. The ligand is initially placed in the vicinity of the binding site and the atomic motions of the ligand and binding site are explicitly simulated, with solvent represented by an implicit solvation model and using a grid representation for the bulk of the receptor protein. These two approximations make the method computationally efficient and yet maintain accuracy close to that of an all-atom calculation. For the benzamidine/trypsin system, we ran 100 independent simulations, in many of which the ligand settled into the low-energy conformation observed in the crystal structure of the complex. The energy of these conformations was lower than and well-separated from that of others sampled. Extensions of this method are also discussed. 

*Author to whom all correspondence should be addressed.

*Journal of Computational Chemistry, Vol. 16, No. 4, 454–464 (1995)*

CCC 0192-8651 / 95 / 040454-10

## Introduction

Specific binding of a ligand to a macromolecular receptor is accomplished through steric fit and complementary chemical interactions between the two species. Although the conformation of the ligand may change significantly upon binding, the macromolecular receptor typically undergoes only local rearrangements. When approximate structural information (such as the conformation of the unliganded form of an enzyme) is known, computational methods can provide information about the accessible conformations of the complex and the strengths of interactions within the complex.[1] This quantitative information, combined with chemical knowledge, can help guide the medicinal chemist in a rational design process to produce a ligand with high affinity and specificity for a given receptor.

Empirical energy functions which treat each atom of the ligand, receptor, and surrounding solvent explicitly are potentially the most accurate. If transferable atomic parameters are available, then the accessible conformations and energetics of the system can be estimated with a minimum number of assumptions. However, the computational demands of estimating the energy of a molecular system grow as the square of the number of atoms. Therefore, calculating the energy for more than a relatively small number of conformations of a macromolecular complex can quickly become computationally prohibitive.

A common simplification, which we employ in this study, is to ignore the details of the surrounding solvent molecules and to consider only the average effects of the solvent on the explicit atoms. This implicit solvent assumption eliminates a significant number of calculations. For example, the small 58-residue protein bovine pancreatic trypsin inhibitor (BPTI) consists of about 600 atoms. To solvate the protein properly requires on the order of 2360 water molecules,[2] thus adding more than 7000 atomic sites and nominally increasing computational demands over 150-fold. However, when using an implicit solvent assumption, care must be taken to choose approximations which account for the enthalpic and entropic effects (and dynamic, if necessary) of the solvent as accurately as possible.

Another common simplification to make the macromolecular energy calculations more tractable is to employ a distance cutoff.[3] In this approximation, it is assumed that an atom interacts only with the atoms in a spherical region surrounding that atom. All the interactions with atoms outside this region are assumed negligible and are ignored or are accounted for in a simple manner. The computational demands of the energy calculation using a cutoff scale as a constant (the average number of neighboring atoms in a sphere of the cutoff radius) times the number of atoms in the system. However, recent studies[4,5] have shown that a relatively large cutoff distance (e.g., over 14 Å) must be used to account for long-range electrostatic effects properly. Because the number of neighboring atoms scales as the third power of the cutoff distance, the scaling constant can become large and thereby dramatically reduce the benefits of using a cutoff.

A promising method for calculating the interactions of a ligand with a macromolecule which does not require a distance cutoff is based on the assumption that the macromolecule can be held completely rigid.[6–8] Because the macromolecule does not move, a three-dimensional potential grid surrounding the receptor can be precalculated. Because the grid only needs to be calculated a single time, the extra cost of including all the interactions without a cutoff distance is within computational limits. The interaction energy of various positions of the ligand in the binding site of the receptor is estimated by simply reading values off the grid at the coordinates of the ligand. This method is computationally efficient, reducing the number of calculations to be proportional to the number of atoms in the ligands plus the setup cost of precalculating the grid. Unfortunately, due to the steep nature of the steric interaction energy, conformations with small overlaps of the ligand and receptor will not be allowed. With small rearrangements of the receptor, these conformations may be energetically favorable. Various approximations (such as softening up the repulsion interaction[7,9] or allowing a certain number of bad contacts) have been proposed, but unfortunately these approximations jeopardize the accuracy of a quantitative answer. In addition, the rapidly varying nature of the steric repulsion is difficult to represent accurately on a grid (this will be discussed in more detail below).

In this article, we try to balance necessary details and computational demands by defining a binding site region of the receptor, to be treated explicitly, combined with a grid representation of the bulk of the receptor. With this method, we are able to mimic the local detailed response of the receptor to the binding of a flexible ligand while

accounting for all interactions without resorting to a distance cutoff. The method we suggest is a combination of the stochastic boundary molecular mechanics methods proposed by Berkowitz and McCammon[10] and the particle–particle, particle–mesh concepts proposed by Hockney and Eastwood.[11] Although both methods have been used extensively, they have not been used in what we consider an effective combination to study the process of a ligand binding to a macromolecular receptor.

## Energy Function

We assume that the energy of a molecular system can be described as a sum of strain energies

$$E_{\text{strain}} = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \quad (1)$$

and nonbonded interaction terms:

$$E_{\text{nonbond}} = \sum \sum_{\text{all pairs}} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon r_{ij}} - S_i f_j \exp\left(\frac{-r_{ij}^2}{2\sigma^2}\right) + \sum \sum_{\text{h bonds}} \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \quad (2)$$

### STRAIN ENERGY

The three terms in eq. (1) represent the difference in energy between a geometry in which the bond lengths, bond angles, and dihedral angles have ideal values and the actual geometry.[12] The computational cost of these terms grows linearly with the number of atoms and is a small percentage (typically less than 5%) of the total computational cost for evaluating the energy in a macromolecular system.

### VAN DER WAALS ENERGY

The first two terms in eq. (2) represent the van der Waals interaction as a sum of a steep repulsion and attractive dispersion term. These interactions take place between all nonbonded atoms in the system and, although they are relatively short ranged, can account for a significant amount of energy.

### ELECTROSTATIC INTERACTION ENERGY

The third term in eq. (2) represents the electrostatic interaction between the explicit atoms. In this work, the effects of the implicit solvent on the electrostatic interactions are approximated using an effective dielectric function.

For the protein rhodanese, Gilson and Honig[13] used the finite difference solution of the Poisson equation to show that a simple dielectric function was unable to reproduce adequately the details of the electrostatic interactions in the system. For a charge on the surface of the protein, however, they found that the average electrostatic interaction at a given distance was approximately given by Coulomb's law with a dielectric constant proportional to the distance.

Repeating the calculations for the protein trypsin in complex with the inhibitor benzamidine, we have found similar behavior. Although the potentials at particular atoms in trypsin show significant deviations, the average potential at a given distance generated by a charge in benzamidine is represented reasonably well by a simple linear distance-dependent dielectric. For 20 positions of benzamidine randomly located within 8 Å of the binding site of trypsin, we found that the electrostatic interaction energy estimated assuming $\epsilon = r$, with $r$ measured in Å, correlated well ($R^2 = 0.96$) with the interaction energies calculated using solutions of the finite difference Poisson equation (range of energies = 0 to −38 kcal/mol). Like Gilson and Honig, we found that the interaction energies calculated with the distance-dependent dielectric were approximately 1.4 times as large as the values calculated using the finite difference solution of the Poisson equation. Given the uncertainty and computational demands of the finite difference method and the preliminary nature of this study, we have chosen to use the linear distance-dependent dielectric in this work.

### DESOLVATION ENERGY

The fourth term in eq. (2) is a simple representation of the free energy of interaction of an atom with the implicit solvent. It was derived by considering the transfer of an atom from an environment completely surrounded by solvent to one in which it has explicit atomic neighbors.[2] Upon transfer,

the neighbors are assumed to displace a volume of solvent surrounding the atom proportional to the volume of the neighbor ($f_j$) times a Gaussian envelope function. The proportionality constant, termed the solvation parameter ($S_i$), is determined for each of six atomic types [C(aliphatic), C(aromatic), S, N/O, $O^-$, and $N^+$] from experimental free energies. These experimentally determined values include electrostatic and dispersion desolvation of the atoms as well as the entropic gain for the release of solvent. For the details of the occupancy desolvation method, see Stouten et al.[2]

We define the electrostatic desolvation as the total desolvation penalty calculated using the aforementioned solvation parameters minus the value calculated assuming that the system was composed of atoms with the solvation properties of aliphatic carbon. We can then compare the electrostatic desolvation penalty calculated with this method to the values calculated using the finite difference Poisson equation (see Gilson and Honig[14]).

For 20 randomly generated positions of benzamidine near the trypsin binding pocket, we find that the electrostatic desolvation penalty of the receptor calculated using the occupancy method correlates well ($R^2 = 0.87$) with the corresponding values calculated by solving the finite difference Poisson equation (range of receptor desolvation energies = 1.5 to 18.6 kcal/mol). We find a slightly better correlation ($R^2 = 0.95$) for the desolvation of benzamidine (range of ligand desolvation energies = 1.0 to 19.0 kcal/mol). The energies calculated using the simple occupancy method were a factor of 1.2 to 1.3 larger than the finite difference results. Again, there is significant scatter in the results, but given the uncertainty and computational demands of the finite difference method, we have chosen to use the much simpler occupancy desolvation model in this work.

### HYDROGEN BOND ENERGY

The final two terms in eq. (2) represent the energy of hydrogen bonds in the system. These terms, which are only summed over atoms which can participate in hydrogen bonds, are present in the AMBER force field which was used in this work.[12] In other force fields, the energetics of a hydrogen bond are accounted for by a careful balancing of the van der Waals and electrostatic components.

## Grid Representation

As mentioned in the Introduction, the nonbonded terms which cause the computational effort of the interaction energy calculation to grow as the total number of atoms squared can be significantly reduced if a region of the system can be assumed to be rigid. The potential due to the rigid atoms can be calculated at any point $p$ in space using eq. (2)[6-8] For the first three terms, we construct three grids:

$$\phi^A(p) = \sum_{j \in \text{rigid atoms}} \frac{(A_{jj})^{1/2}}{r_{jp}^{12}} \quad (3)$$

$$\phi^B(p) = \sum_{j \in \text{rigid atoms}} -\frac{(B_{ij})^{1/2}}{r_{jp}^{6}} \quad (4)$$

$$\phi^{ES}(p) = \sum_{j \in \text{rigid atoms}} \frac{q_j}{\epsilon r_{jp}} = \sum_{j \in \text{rigid atoms}} \frac{q_j}{r_{jp}^2} \quad (5)$$

The fourth term (desolvation) requires two grids:

$$\phi^{DES,EXPL}(p) = \sum_{j \in \text{rigid atoms}} f_j \exp\left(\frac{-r_{jp}^2}{2\sigma^2}\right) \quad (6)$$

$$\phi^{DES,BULK}(p) = \sum_{j \in \text{rigid atoms}} S_j \exp\left(\frac{-r_{jp}^2}{2\sigma^2}\right) \quad (7)$$

where if a mobile atom $i$, with solvation parameter $S_i$ and fragmental volume $f_i$, is located at grid point $p$, then the desolvation of the explicit atom $i$ by the rigid atoms is equal to $S_i\phi^{DES,EXPL}(p)$ and the desolvation of the rigid bulk atoms by atom $i$ is equal to $f_i\phi^{DES,BULK}(p)$.

Grids corresponding to the hydrogen bonding potential were constructed in the same manner as the preceding equations. But, as implemented in the method described in the next section, they were found to have a negligible contribution and were not used.

To calculate the potential at any point in space, we use a simple trilinear interpolation scheme from the closest eight grid points.[15] The total interaction energy between the mobile and fixed atoms

is then:

$$E = \sum_{i \in \text{mobile atoms}} \left[ \begin{array}{c} (A_{ii})^{1/2} \phi^{A}(x_i) + (B_{ii})^{1/2} \phi^{B}(x_i) \\ + q_i \phi^{\text{ES}}(x_i) \\ + S_i \phi^{\text{DES,EXPL}}(x_i) \\ + f_i \phi^{\text{DES,BULK}}(x_i) \end{array} \right] \quad (8)$$

where $x_i$ are the coordinates of mobile atom $i$.

A three-dimensional force grid can be calculated by taking the negative of the gradient of the potential grid at each grid point. The force on each atom can then be estimated by using trilinear interpolation of the force grid. Notice that in the application described later in this article, and in applications for which the method is intended, the mass of the rigid macromolecular receptor will be relatively large. To simplify calculations, we assume that all forces and torques on the macromolecular receptor are negligible.

To examine the error in replacing explicit functions by a grid, we constructed grids of (20 Å)$^3$ ($40^3$ grid points with spacing 0.5 Å) representing a single atom at the origin. The atom was assigned parameters typical of the AMBER force field ($r^*$ = 1.5 Å, $\epsilon$ = 0.15 kcal/mol, $q$ = 0.25 electrons) and solvation parameters typical of a charged atom ($S_i$ = −28.3, $f_i$ = 12.0).

An explicit atom having the same parameters was then placed at a random position between 2.3 and 3.5 Å from the origin, and the energy of the system was calculated. The results for 250 trial positions are presented in Figure 1.

The Coulombic interaction (filled circles) follows the analytical results (solid line) very well up to the closest distance examined (error less than 0.05 kcal/mol). The dispersion interaction (filled squares) deviates from the analytical result (solid line) by less than 0.05 kcal/mol for distances larger than 2.5 Å but begins to show a slight deviation (0.2 kcal/mol) at the closest distance examined (2.3 Å). In contrast, the repulsive interaction (filled triangles) begins to show significant deviations (greater than 1 kcal/mol) at separation distances less than 2.5 Å. The solvation term (filled diamonds) is relatively small for the system and is represented well at all separations.

Also shown in Figure 1 is the total interaction energy of the two atoms assuming that they are of opposite charge (open circles). For this system, the minimum in the potential energy is at approximately 2.5 Å separation. As mentioned earlier, at this distance the grid representation of the repulsive term includes significant error. This, in turn, causes the total interaction energy to deviate from the analytical result.

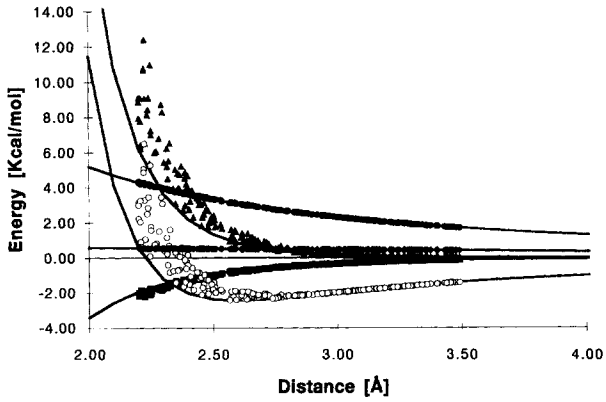


**FIGURE 1.** Interaction energy (kcal / mol) of an explicit atom with an implicit atom represented by a grid with a 0.5 Å grid spacing. Both atoms were assigned nonbonded parameters typical of the AMBER force field $r^*$ = 1.5 Å, $\epsilon$ = 0.15 kcal / mol, $q$ = 0.25 electrons and solvation parameters typical of charged atoms $S_i$ = −28.3 cals / (mol Å$^3$), $f_i$ = 12.0 Å$^3$. The components of the interaction energy are displayed individually: Coulombic interaction (filled circles), dispersion interaction (filled squares), repulsion interaction (filled triangles), and the solvation interaction (filled diamonds). The open circles are the total interaction energy assuming that the two atoms are of opposite charge.

This error can be reduced by decreasing the grid spacing. Presently, we are using $40^3$ grids with a grid spacing of 0.5 Å to cover a (20 Å)$^3$ region of space. With each grid point stored as a four-byte real number, the five grids occupy about 1.3 Mbytes of memory. Decreasing the grid spacing by a factor of 2 would decrease the error in the representation of the repulsion interaction and still allow the grids to fit in the memory of a typical workstation. But in the future we intend to employ multiple conformations of the rigid receptor, which, in turn, would require multiple sets of grids and further increase demands on memory. Furthermore, in the method proposed next, an explicit atom will never significantly overlap with an implicit atom. For these reasons, we select a 0.5-Å grid spacing for the work described here.

## Method

The system we have chosen to test our approach is the binding of the small charged inhibitor ben-

zamidine to the 223-residue protein trypsin. This system was chosen because of the extensive literature on binding of a diverse set of ligands to the serine proteases.[16,17] Structural studies led us to expect only limited rearrangements of the atoms in the binding site upon formation of the complex.[18] This will minimize any error associated with holding the bulk of the receptor completely rigid in this initial study.

The crystal structure of the benzamidine/trypsin complex, 3ptb.pdb, available from the Brookhaven Protein data bank,[19] was used as a starting point for the calculations. Polar hydrogens were added to the complex using the molecular modeling program AMBER, with titrating residues assigned their normal protonation states at neutral pH. The active site Histidine (57) was assumed to be singly protonated at the delta nitrogen. The complex was minimized using the AMBER force field to remove any steric overlap; deviation of the nonhydrogen atoms from the crystal structure was less than 0.4 Å.

The binding site of trypsin, to be represented explicitly, was chosen as residues 57, 99, 189–195, 213–220, and 228. This defines a 154-atom region around the $S_1$ binding pocket. The remaining 1885 bulk atoms were assumed to be rigid and were represented implicitly on the grids. This is illustrated in Figure 2a, in which the binding site and ligand atoms are displayed as spheres and only a backbone trace of the rest of the protein is shown. The partitioning of the receptor into binding site and bulk is arbitrary, but the binding site should include all residues which could interact directly (i.e., approach van der Waals overlap) with the ligand and residues which could undergo significant conformational adjustments upon complex formation (such as those of flexible loops or flaps).

The binding site region was further subdivided into two regions: a fully mobile region and a buffer region. The buffer atoms were defined as any binding site atom which had a separation from a bulk atom of less than 0.5 Å plus the sum of their van der Waals radii. These 99 buffer atoms are restrained to the relaxed crystallographic coordinates using a harmonic potential function with a force constant of 1 kcal/(mol Å$^2$). This potential is intended to mimic the steric and electrostatic forces of the bulk protein on the buffer region atoms. The restrained atoms are represented in Figure 2a and more schematically in Figure 2b as dark spheres.

A more refined model could use the experimentally determined crystallographic thermal factors to determine the harmonic force constant for each atom in the buffer region.[20] A force constant of 1 kcal/(mol Å$^2$) corresponds to a $B$ factor of about 47 Å$^2$. This is a relatively large value compared to the experimental values, which average approximately 13 Å$^2$ for the buffer atoms. Because the structure of the receptor was based on the complex with the benzamidine ligand, we chose this rela-

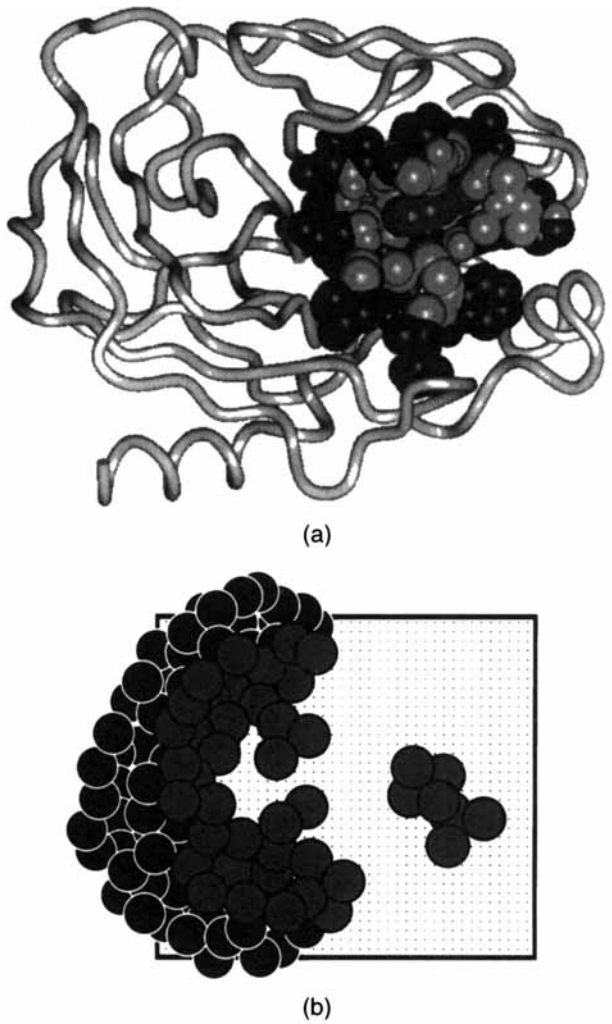


**FIGURE 2.** (a) Molecular graphic representation of the three-dimensional structure of the system. The binding site and ligand atoms which were represented explicitly in the calculation are displayed as spheres. Only the backbone trace of the remaining bulk receptor is displayed. (b) Schematic representation of the trypsin / benzamidine system. The dots represent the mapping of the potential from the bulk of the receptor onto a three-dimensional grid surrounding the active site. In both (a) and (b), the restrained binding site atoms are displayed as dark spheres and the mobile atoms are displayed as lighter gray spheres. In (a) the atoms of benzamidine (although completely mobile) have been slightly darkened to offset them from the rest of the mobile atoms.

tively weak constraint to try to remove this bias and to be able to use this same model for other structurally diverse inhibitors.

The remaining 55 atoms of the binding site and 13 atoms of the ligand are fully mobile and interact with each other and the buffer region directly through the potential function eq. 2. These atoms are represented as the lighter gray spheres in Figures 2a and 2b. These mobile atoms interact with the remaining bulk of the receptor through the potential grids (illustrated by the grid points in Fig. 2b). Note that these atoms are separated from the bulk atoms by the buffer region, thus avoiding the error described earlier, which results from overlapping with the implicit atoms.

It is this separation of fully mobile atoms from the bulk atoms which makes the hydrogen bonding interactions between the regions insignificant. The hydrogen bonding interaction is relatively short ranged, more so than the van der Waals interactions. Also, in contrast to the van der Waals interactions, the hydrogen bonding interaction potential only operates between atoms which are capable of donating or receiving a hydrogen bond. In our studies, the energetic contributions of the hydrogen bonding potential between atoms in the bulk and mobile atoms was insignificant (less than 0.001 kcal/mol) and was ignored.

This scheme lends itself to implementation in existing molecular mechanics software. The molecular mechanics program only needs information about the binding site and ligand atoms. Within the force/energy subroutine of the program, a call is made to an external subroutine, to which is passed the coordinates of all the binding site and ligand atoms. The subroutine uses the preconstructed potential grids to calculate the force and energy of the system due to the implicit bulk atoms and returns these values to the main program. In this study, we also used the subroutine to calculate the desolvation forces and energies which are not included in AMBER.

AMBER allows the user to hold a section of the system rigid and skip the calculation of forces between these bulk atoms. We estimate that the number of nonbonded calculations using the "belly" option is proportional to $N_{\text{bulk}} N_{\text{dynamic}} + N_{\text{dynamic}}(N_{\text{dynamic}} - 1)/2$, where $N_{\text{dynamic}}$ is the number of binding site atoms plus the number of ligand atoms. We estimate that the dominant calculation in mixed grid/explicit atom method will be proportional to $N_{\text{dynamic}}(N_{\text{dynamic}} - 1)/2$. For trypsin we had 1885 bulk atoms and 167 binding site and ligand atoms. This gives a theoretical speedup of over 23-fold relative to the belly calculation without a cutoff. As mentioned in the Introduction, one can also use a distance cutoff in the energy evaluation to reduce computational costs. In the belly calculation, this implies that each dynamic atom only interacts with other dynamic atoms and bulk atoms within the cutoff radius.

In practice, we measured the speedup by taking 10,000 molecular dynamics steps of the benzamidine/trypsin system using the belly option of AMBER without a cutoff and with 12, 9, and 7 Å residue-based cutoffs. The results are summarized in Table I, where the central processing unit (CPU) times (in seconds) are listed in the third row and the speedup relative to the belly calculation without a cutoff is given in the fourth row. Also shown in rows 1 and 2 of the table are the van der Waals and electrostatic interaction energies of the benzamidine, with the bulk atoms of trypsin in the relaxed crystal structure (the initial structure prior to the molecular dynamics runs). These values are given in kcal/mol, and the error in the approximate methods relative to the method without a cutoff is given in parentheses.

**TABLE I.**
**Van der Waals and Electrostatic Interaction Energies (kcal / mol) Calculated for Benzamidine in the Active Site of Trypsin.**

| | Without a cutoff | 12 Å residue-based cutoff | 9 Å residue-based cutoff | 7 Å residue-based cutoff | Combined MM / grid |
|---|---|---|---|---|---|
| vdW energy | −3.87 | −3.76 (0.12) | −3.49 (0.38) | −2.86 (1.02) | −3.81 (0.06) |
| ES energy | 10.09 | 9.26 (−0.83) | 8.95 (−1.14) | 4.22 (−5.86) | 10.10 (0.01) |
| CPU time | 6160 | 2143 | 1182 | 705 | 320 |
| Speedup | 1 | 2.9 | 5.2 | 8.7 | 19.3 |

The error in the approximate methods relative to the calculation without cutoffs is shown in parentheses. CPU time in seconds for 10,000 molecular dynamics steps and speedup of the approximate methods relative to the calculation without cutoffs. All calculations were performed on an SGI Challenge.

As can be seen in Table I, the 12-Å cutoff gives energies with less than 1 kcal/mol of error with almost a threefold increase in computational speed over the no-cutoff calculation. The errors with a 9-Å cutoff are still reasonable (on the order of 1 kcal/mol) with a fivefold enhancement. A cutoff of 7 Å, although giving an almost ninefold improvement in speed, involves significant errors. In comparison, the combined grid/molecular mechanics method, which we propose, has small errors and gives an almost 20-fold improvement in speed over the belly option without a cutoff.

## Automated Docking

As an application of the method, we investigated the automated docking of benzamidine into the binding site of trypsin. A starting conformation of benzamidine near the binding site was generated by a random rotation and translation (up to a maximum translation of 8 Å) of the ligand from the crystallographic position. The average root mean square deviation (rmsd) of the benzamidine starting coordinates from the corresponding crystallographic coordinates was 14.5 Å (minimum = 10.2 Å, maximum = 18.6 Å). The constant temperature molecular dynamics portion of the AMBER program was then used to search the conformational space available to the system. Visualization of the dynamics simulation demonstrated that the benzamidine was sampling a large region of space surrounding the binding site. If the benzamidine drifted past the edge of the grid, an artificial harmonic force was applied to reflect the ligand back onto the grid. The trajectory was propagated at 300 K for 20 ps using a 2-fs timestep. The energy and coordinates of the trajectory were saved to disk every 50 steps (0.1 ps). The CPU time on a single processor of an SGI Challenge was approximately 15 minutes for a 20-ps trajectory. This calculation was repeated with 100 different starting positions.

Because each starting configuration is independent of the others, the entire calculation was easily run in parallel (at the script level) on six processors of the eight-processor SGI Challenge, with a turnaround time of about 4 hours.

In Figure 3, the potential energy of the system averaged over the last 2 ps of each simulation is plotted against the rmsd of the benzamidine from the crystallographic positions averaged over the same period. The potential energy is relative to the

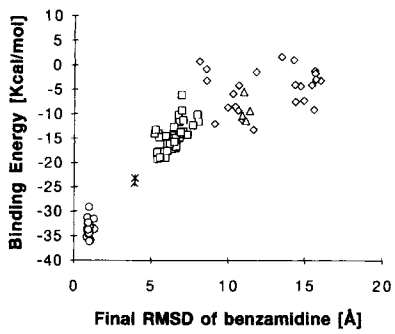


**FIGURE 3.** The binding energy of benzamidine to trypsin (kcal / mol) versus the RMSD (Å) of the benzamidine atoms from the corresponding crystallographic coordinates averaged over 2 ps. Some of the results of the 100 independent trajectories are grouped into clusters. (See text for details.)

sum of the average potential energy of the benzamidine and uncomplexed inhibitor calculated in separate 10-ps simulations. Typically, the rmsd of the mobile portion of the binding site atoms from the corresponding crystallographically determined positions was about 1.0 Å. The rmsd of the buffer atoms was typically less than 0.5 Å.

The final coordinates of the atoms of the benzamidine were used to cluster the terminal conformations into groups. An arbitrary criterion—that a member of a group have an rmsd less than 0.5 Å from another member of that group—resulted in one cluster of 44, one cluster of 23, one cluster of 4, 7 clusters of 2, and 15 final positions of benzamidine which were no closer than 0.5 Å to any other final positions.

The cluster of 23 is represented as open circles in Figure 3. This group shows an rmsd, on average, of approximately 1 Å from the crystallographic position and is lower in energy than any of the other clusters by at least 5 kcal/mol and more typically 10 kcal/mol. Based on the similarity of the final positions of benzamidine to the crystallographically observed coordinates, we consider this to be the correct docking of benzamidine into trypsin.

An interesting cluster of 2 is represented by asterisks in Figure 3. These two structures had rmsd's from the crystallographic positions of about 4 Å. Upon visual inspection, they were found to place only one of the amidino nitrogens in proximity of the ASP 189 at the bottom of the pocket (the benzamidine has both amidino nitrogens at the bottom of the pocket interacting with the aspartic acid). The second amidino nitrogen was pointed in

a direction out of the pocket into solvent. These structures were approximately 8.9 kcal/mol higher in energy than the average of 23 correct structures. The largest portion of the energy was due to a 7.6 kcal/mol loss in electrostatic energy.

The largest cluster of 44 is represented by the open boxes in Figure 3. These structures showed an average rmsd of 6.4 Å from the crystal structure. Upon visual examination, the benzamidine was found to be rotated by approximately 180 degrees from the crystallographic position (about the axis normal to the plane of the benzene ring). These backward benzamidines are able to retain most of the van der Waals interactions of the correct orientation with approximately a 2 kcal/mol increase in energy. However, they pay a large (almost 13 kcal/mol on average) penalty for lack of favorable electrostatic interactions.

The cluster of 4 is shown in Figure 3 as open triangles. Visualization revealed that in this conformation the benzamidine was at the surface of the binding pocket, with the amidine groups participating in electrostatic interactions with GLN 192. The rest of the members of all the clusters are not distinguished and are represented as open diamonds. On average, these structures had about 12 kcal/mol less favorable electrostatic interactions and about 11 kcal/mol less favorable van der Waals interactions than the correct orientations.

No relationship among the 23 successful starting positions was visually discernible. The rmsd of these 23 starting positions from the crystallographically observed binding coordinates were on average 14.0 Å, with a minimum of 11.3 Å and a maximum of 16.8 Å. These values can be compared to the rmsd of the other 77 starting positions from the observed binding coordinates, which averaged 14.7 Å with a minimum of 10.2 Å and a maximum of 18.6 Å.

It is important to note that on average, all the noncorrect final orientations paid about a 2 kcal/mol penalty in constraint energy for distorting the active site. It is difficult to assess the importance of this penalty. The plasticity of the binding site is unknown and probably varies from atom to atom. Replacing the assumed 1 kcal/(mol Å$^2$) harmonic force constant with values determined from experimentally determined thermal factors of the crystallographic complex removes the subjectivity of the value but may restrain the site too closely to the conformation with that particular inhibitor. The sensitivity of the results to this parameter and methods of deriving force constants from structural analysis of proteins are being investigated.

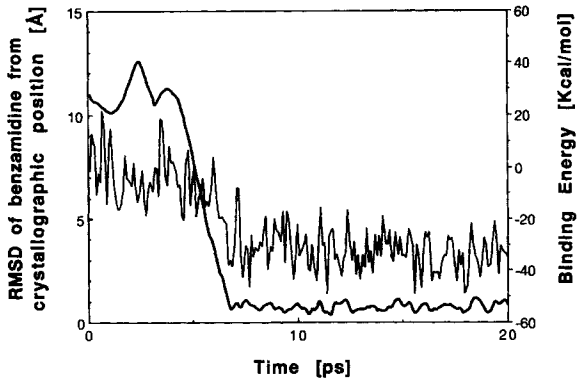


**FIGURE 4.** The time course of a trajectory which resulted in the benzamidine binding in the correct manner to trypsin. The thick line is the rmsd (Å) of the coordinates of the substrate from the corresponding coordinates of the crystal structure as a function of time (picoseconds). The thin line is the potential energy of the system (kcal/mol) relative to the unbound substrate/receptor as a function of time.

The binding energy is, in fact, the binding free energy with respect to the solvent but is lacking terms accounting for the entropy change of the explicit atoms. Using a value of 15 kcal/mol for the loss of translational and rotational entropy of the ligand upon binding[21] and ignoring any change in the entropy of internal rotations of the ligand or the receptor, we find a binding free energy of approximately −20 kcal/mol. This value is quite a bit larger than the experimentally measured value of −6 kcal/mol[17] but is not unexpected given the simplicity of the model. It is hoped that when relative binding energies of ligands with the same number of degrees of freedom are estimated, the error in this term will cancel. However, if ligands with different numbers of degrees of freedom are compared, or even ligands with the same number of degrees of freedoms which bind in different manners, then a less arbitrary method of estimating the loss in entropy upon complexation will be necessary.

In Figure 4, we present the time course of one of the trajectories which resulted in the binding of benzamidine in the correct orientation. The thick line is the rmsd of the benzamidine from the crystallographic position, and the thin line is the effective potential energy (again shown relative to the time average potential energy of the uncomplexed benzamidine and trypsin). It can be seen in Figure 4 that once the benzamidine reaches a cer-

tain position, it rapidly drops into the binding site with a corresponding drop in energy.

## Conclusions and Extensions

For a system in which the approximate location of the binding site is known and only local adjustments occur upon complexation, we believe that the modest computational demands of the method we propose will allow for efficient searching of the conformational space of the binding complex. Although there is considerable variation in size among binding sites, there is little proportionality between size of the protein and that of binding site. Because the calculation grows as $N^2_{\text{dynamic}}$, the simulation will remain tractable for large systems. Calculation of the grids, which scale as $N^2_{\text{bulk}}$, will become lengthy for large proteins, although the size of the grids (proportional to the dimensions of the binding site) will remain small enough to fit easily into the memory of modern workstations. The grids need be calculated only once because they do not depend on the ligand studied or on its position.

Our results indicate that the method works well for a relatively inflexible molecule such as benzamidine. When the ligand has multiple rotatable bonds the problem becomes much more complex, but that is a separate issue that goes beyond the scope of this article. What has been addressed here is the accuracy of the molecular mechanics/grid model. With confidence in the formalism, one can go on to consider the binding of flexible ligands.

A second more important step will be the ability to make quantitative predictions regarding the binding affinity of an uncharged ligand and the relative binding affinities of a set of ligands. This work is currently underway and examines a number of serine protease inhibitors, with careful attention being paid to weakness in the energy function (e.g., the implicit solvent terms) and in the method (e.g., estimation of entropy loss of the explicit atoms and dependence of results on the structural restraints). These results will be compared to experimentally measured values and results from more detailed computational models using explicit solvents.

As mentioned earlier, the assumption of holding the bulk receptor completely rigid can be at least partially relaxed. Although we do not expect a small rearrangement of a sidechain far from the binding site to have a large influence, large global changes of the macromolecular receptor will be important. An obvious example is the enzyme lysozyme, which undergoes a large hinge-bending motion upon complexation with its substrate.[1] In this situation it may be possible to construct multiple sets of grids which represent different conformations of the bulk. After a certain interval in the searching process, another conformation of the bulk receptor could be tried and the energy evaluated. The new conformation could be accepted or rejected based on a Metropolis criterion. A number of bulk coformations should be constructed which are significantly different and cover conformational space. If the reaction coordinate is known, such as in lysozyme, this could be used; if not, then normal mode analysis could be used to help find the modes responsible for particular large conformational changes.

In conclusion, we feel that the present approach, which has only moderate computational demands, is a promising one for the study of complex formation and evaluation of potential ligands. It occupies the middle ground between the methods which can find binding sites over larger regions of space by assuming a rigid receptor and a less complex energy function, and the computer alchemy methods which can give accurate answers and intricate details, but at a relatively high computational cost.

## Acknowledgments

We thank Dr. Michael K. Gilson for his comments on the manuscript and stimulating discussions on the energetics of ligand binding. B. A. L. is grateful to DuPont Merck for providing a postdoctoral fellowship during which this work was completed. J. A. M. is supported in part by grants from the National Institute of Health and the Robert A. Welch foundation.

## References

1. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.
2. P. F. W. Stouten, C. Frömmel, H. Nakamura, and C. Sander, *Mol. Simulation*, **10**, 97 (1993).
3. R. J. Loncharich and B. R. Brooks, *Proteins*, **6**, 32 (1989).
4. H. Schreiber and O. Steinhauser, *Biochemistry*, **31**, 5856 (1992).

5. H. E. Alper, D. Bassolino, and T. R. Stouch, *J. Chem. Phys.*, **98,** 9798 (1993).
6. P. J. Goodford, *J. Med. Chem.*, **28, 849 (1985).**
7. D. S. Goodsell and A. J. Olson, *Proteins*, **8,** 195 (1990).
8. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comp. Chem.*, **13,** 505 (1992).
9. F. Jiang and S.-H. Kim, *J. Mol. Biol.*, **219,** 79 (1991).
10. M. Berkowitz and J. A. McCammon, *Chem. Phys. Lett.*, **90,** 215 (1982).
11. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, IOP Publishing Ltd., Bristol, England, 1988.
12. P. K. Weiner and P. A. Kollman, *J. Comp. Chem.*, **2,** 287 (1981).
13. M. K. Gilson and B. H. Honig, *Proteins*, **3,** 32 (1988).
14. M. K. Gilson and B. Honig, *Proteins*, **4,** 7 (1988).
15. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1986.
16. F. Markwardt, H. Landmann, and P. Walsmann, *Eur. J. Biochem.*, **6,** 502 (1968).
17. M. Mares-Guia, D. L. Nelson, and E. Rogana, *J. Am. Chem. Soc.*, **99,** 2331 (1977).
18. M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, *Acta Crystallog.*, **B39,** 480 (1983).
19. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112,** 535 (1977).
20. C. L. Brooks III, A. Brünger, and M. Karplus, *Biopolymers*, **24,** 843 (1985).
21. A. Fersht, *Enzyme Structure and Mechanisms*, W. H. Freeman, San Francisco, CA, 1977.