Pairwise GB/SA Scoring Function for Structure-based Drug Design

Hao-Yang Liu,[†] Irwin D. Kuntz,[‡] and Xiaoqin Zou*[,†]

Dalton Cardiovascular Research Center and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, and Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143

Received: November 14, 2003

An accurate and efficient energy model to compute free energies of ligand binding to proteins is essential to understanding protein−ligand interactions, particularly to structure-based drug design. A bottleneck to derive such an energy model is how to account for the solvent effect. A grid-based free energy model was proposed (Zou et al. *J. Am. Chem. Soc.* **1999**, *121*, 8033−43),[1] based on the generalized-Born (GB/SA) model of solvation for small molecules (Still et al. *J. Am. Chem. Soc.* **1990**, *112*, 6127−9).[2] Although the grid-based free energy model shows much improved results than the current DOCK force field energy model (Meng et al. *J. Comput. Chem.* **1992**, *13*, 505−24)[3] by taking into account the solvent effect, the computational speed for the free energy model restricts the model from a direct drug screen of a huge compound database. Rather, the model is used for a second-step screening after a pre-screening by the force field method. Here, we present a fast algorithm, the pairwise free energy model, for ligand binding affinity calculations. Specifically, a pairwise descreening approximation (Hawkins et al. *Chem. Phys. Lett.* **1995**, *246*, 122−9)[4] is used in calculations of the electrostatic energy contribution. A procedure is also developed to account for the low-dielectric region that might form between the ligand and the receptor during docking processes. The method has been tested on database screening for dihydrofolate reductase, trypsin, and a fatty acid-binding protein. We obtain similar results compared with the grid-based free energy model but with much less computation efforts. Our pairwise algorithm takes approximately 0.5 s per orientation (with minimization) on a Silicon Graphics Octane R12000 workstation, rapid enough to be used for direct screening of a database or combinatorial library.

1. Introduction

Quantitation of molecular forces is essential for understanding the basis of biological structures and functions and designing therapeutic interventions. A reliable and efficient model to compute free energies of ligand binding to proteins is crucial as thousands of protein atoms (and thus large-scale degrees of freedom) are involved. The model is especially important for structure-based drug design, as hundreds of thousands of molecules need to be screened, among which only a few molecules would have high binding affinities to the target protein and would become potent therapeutic inhibitors after optimization. A long-standing bottleneck in developing energy functions for ligand binding is how to account for the solvent effect.[5−7]

Upon ligand binding, some or all of the water molecules that surround the ligand and the receptor binding site are stripped away (i.e., desolvation). As a polar group favors aqueous environment, there will be an energy penalty when the group is desolvated. This energy penalty will be high especially for desolvation of charged groups.[8] Ignorance of this desolvation effect will result in misprediction of highly charged molecules to undergo strong binding reactions.[9,10] Because of the complicated geometries of the protein and the ligand, it is challenging to calculate their desolvation energies and electrostatic interaction energy upon ligand binding.

The most straightforward method for desolvation correction is to treat solvent molecules explicitly via molecular dynamics or Monte Carlo simulations of binding (for recent reviews, see refs 11 and 12). However, these approaches take an enormous amount of computational time and thus are impractical for screening large numbers of molecules. A simplified description of the solvent effect is to treat solvents as a high-dielectric continuum (see refs 13−15 for reviews). A widely used approach is to solve the Poisson−Boltzmann (PB) equations by finite-difference,[16−18] finite-element,[19,20] or boundary-element[21] numerical methods. A further simplification of the PB approach that has received considerable recent attention is the generalized Born (GB) method proposed by Still and coworkers[2] (see ref 22 and 23 for reviews). The GB model extends the formalism of the Born model for ionic solvation[8] to treat solutes of arbitrary shape. It can be thought of as a relatively simple analytical approximation which for real molecular geometries mimic the physics of the partial differential Poisson equation. As a result, the GB approach is relatively fast, yet, produces agreeable results for solvation electrostatics, pK_a shifts, and octanol/water partition coefficients (log *P*) for small molecules and proteins.[2,24−29] Recently, the GB model has been used to calculate the electrostatic contribution to ligand−receptor binding energies via grid-based (i.e., volume-integration-based)[1,30,31] and surface-area-based[32] approaches. By also accounting for the hydrophobic effect in terms of the solvent-accessible surface area (SA) and for the van der Waals interactions, our grid-based GB/SA scoring function has been applied to calculate free energies of ligand binding,[1] leading to

* To whom correspondence should be addressed. E-mail: zoux@missouri.edu. Phone: 573-882-6045. Fax: 573-884-4232.
[†] University of Missouri.
[‡] University of California, San Francisco.

10.1021/jp0312518 CCC: $27.50
Published on Web 04/06/2004

much improved scoring compared with the simple distant-dependent dielectric used in earlier versions of the DOCK program.[3,33,34]

The time-consuming step in the grid-based or volume-integration-based GB approach is to calculate atomic Born solvation radii on the fly, which limits its efficient application to macromolecular systems. To solve this problem, several atom-pair-based approximations have been proposed for Born solvation radii calculations.[4,35−41] These approximations have been applied to MD simulations for proteins and nucleic acid helices.[40−46]

In this paper, we present a fast algorithm for ligand-protein binding affinity calculations by using a pairwise Born radius approximation.[4,35,46] With similar accuracy compared to our grid-based scoring model,[1] the new model is nearly 2 orders of magnitude faster and may be used for direct database screening for lead compounds. The outline of this paper is as follows. First, we briefly review the GB/SA model and our previous application to ligand binding. Next we will describe how to apply the pairwise Born radius approximation to our scoring function, including calculating the Born radii on the fly and accounting for possible void formation between the protein and misaligned ligands. Then, the model is tested on database screening for dihydrofolate reductase (dhfr), trypsin, a fatty acid-binding protein (fabp), together with crystal complexes with their inhibitors. Finally, conclusions and discussions are presented.

2. Method

2.1. Overview of the Generalized Born/Surface Area (GB/SA) Model and Its Application to Ligand−Receptor Binding. *2.1.1. The Born Model.* Before we review the generalized Born model, we will first describe the Born formula.[8] Using the classical electrostatic theory, it is easy to derive that the electrostatic component of the solvation free energy (i.e., the polarization energy) of a simple ion of radius a and charge q is given by

$$\Delta G_{\text{Born}} = G^{\text{solvent}} - G^{\text{vacuum}} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon_w}\right) \tag{1}$$

where ϵ_w is the dielectric constant of water (78.3). The negative value of ΔG_{Born} indicates that ions favor an aqueous environment.

2.1.2. The Generalized Born (GB) Model. The GB model[2] is widely used to calculate solvation free energies and pK_a shifts of organic molecules.[4,24−27,35−39] The model introduces the Coulomb field approximation for the electric displacement due to the charge of atom i by ignoring the reaction field.[22,25] In other words, atoms other than i in the molecule are served only as a solute dielectric medium with no charges for the calculation of the polarization energy of atom i. By adapting the Born formalism for this "self-energy" term, Still et al. proposed

$$G_{\text{pol},i} = -\frac{q_i^2}{2\alpha_i}\left(1 - \frac{1}{\epsilon_w}\right) \tag{2}$$

where the effective Born radius of atom i, α_i, is defined as

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{4\pi}\int_{in, r>a_i} \frac{1}{r^4}\,dV \tag{3}$$

where the integration is over the interior region of the molecule excluding the volume of atom i and r is the distance to the center of i. That is, the Born radius α_i is the radius of a hypothetical sphere which has the same polarization energy as the self-energy of atom i of the molecule. This variable quantitatively measures how much atom i is screened from water by other atoms in the molecule. Equation 3 shows that the Born radii are independent of atomic charges and dielectric constants but molecular geometries. The equation also shows that a Born radius (α_i) is larger than its corresponding atomic radius (a_i).

For the electrostatic component of the pair solvation energy, Still and co-workers introduced the following approximation:

$$G_{\text{pol,ij}} = -\frac{q_i q_j}{f_{\text{GB}(r_{ij})}}\left(1 - \frac{1}{\epsilon_w}\right) \tag{4}$$

where q_i and q_j are the charges of atom i and j, $f_{\text{GB}(r_{ij})} = \{r_{ij}^2 + \alpha_i\alpha_j \exp[-r_{ij}^2/(4\alpha_i\alpha_j)]\}^{1/2}$, and α_i and α_j are the corresponding effective Born radii. The f_{GB} function has the following important feature: It approaches r_{ij} as the two atoms are far apart, simplifying $G_{\text{pol},ij}$ to the Coulomb interaction energy; f_{GB} approaches the Born radius as $r_{ij} \rightarrow 0$, simplifying $G_{\text{pol},ij}$ to the self-energy given in eq 2.

The total polarization energy of a molecule with an arbitrary shape is the sum of the self-and pair energy terms

$$G_{\text{pol}} = \sum_i^N G_{\text{pol},i} + \sum_i^{N-1}\sum_{j>i}^N G_{\text{pol},ij} \tag{5}$$

$$= -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^N \frac{q_i^2}{\alpha_i} - \left(1 - \frac{1}{\epsilon_w}\right)\sum_i^{N-1}\sum_{j>i}^N \frac{q_i q_j}{f_{\text{GB}(r_{ij})}} \tag{6}$$

$$= -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^N\sum_j^N \frac{q_i q_j}{f_{\text{GB}(r_{ij})}} \tag{7}$$

where N is the total number of atoms in a molecule.

2.1.3. The Grid-Based Free Energy Model. By applying the GB model to ligand binding, we have derived the following formula for calculations of the electrostatic component of the ligand−receptor binding energy, G_{POL}:[1]

$$G_{\text{POL}} = G_{\text{screened es}} + G_{\text{L desolve}} + G_{\text{R desolve}} \tag{8}$$

where the three terms on the right-hand side are screened electrostatic energy, partial desolvation energy of the ligand, and the partial desolvation energy of the receptor, respectively. These terms are calculated by

$$G_{\text{screened es}} = \sum_i^L\sum_j^R \frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\epsilon_w}\right)\sum_i^L\sum_j^R \frac{q_i q_j}{f_{\text{GB}}^{\text{LR}}(r_{ij})} \tag{9}$$

$$G_{\text{L desolve}} = \frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^L\sum_j^L \left[\frac{q_i q_j}{f_{\text{GB}}^{\text{L}}(r_{ij})} - \frac{q_i q_j}{f_{\text{GB}}^{\text{LR}}(r_{ij})}\right] \tag{10}$$

$$G_{\text{R desolve}} = \frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^R\sum_j^R \left[\frac{q_i q_j}{f_{\text{GB}}^{\text{R}}(r_{ij})} - \frac{q_i q_j}{f_{\text{GB}}^{\text{LR}}(r_{ij})}\right] \tag{11}$$

where L, R, and LR represent the ligand alone, the receptor alone, and the ligand−receptor complex, respectively. f_{GB}^{L}, f_{GB}^{R} and $f_{\text{GB}}^{\text{LR}}$ are calculated in the corresponding environments.

An alternative way to decompose G_{POL} is as follows:

$$G_{POL} = G_{pol,LR} - G_{pol,L} - G_{pol,R} + G_{Coulomb} \quad (12)$$

Here, $G_{pol,X}$ represents the polarization energy of the X molecule, where X stands for the complex (LR), the receptor (R), or the ligand (L). $G_{Coulomb}$ is the Coulombic energy between L and R. Each term can be written explicitly as

$$G_{pol,L} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^L \sum_j^L \frac{q_i q_j}{f_{GB}^L(r_{ij})} \quad (13)$$

$$G_{pol,R} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^R \sum_j^R \frac{q_i q_j}{f_{GB}^R(r_{ij})} \quad (14)$$

$$G_{pol,LR} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_w}\right)\sum_i^{L+RL+R} \sum_j \frac{q_i q_j}{f_{GB}^{LR}(r_{ij})} \quad (15)$$

$$G_{Coulomb} = \sum_i^L \sum_j^R \frac{q_i q_j}{r_{ij}} \quad (16)$$

Simple algebraic calculations show that eqs 7−11 are consistent with eqs 12−16. The physical meaning of eqs 12−16 is further explained in section 3.2.

For the nonelectrostatic energy terms, first, we have considered two different types of VDW interactions: the ligand−protein VDW interactions and the VDW interactions between water molecules and the solute (ligand or protein). The VDW interactions between the ligand and the receptor are characterized by the Lennard-Jones 6−12 potential. Neglecting the solute−solvent VDW interactions would over-estimate the VDW contributions because water molecules usually form good contact with the protein before ligand binding, and the ligand simply replaces the water molecules for VDW contact. The solute−solvent VDW interactions can be rather accurately characterized with explicit solvent treatment but not with the continuum solvent treatment like in this paper. For simplification, we have assumed that the van der Waals interactions between the solute and the continuum solvent (water) is proportional to the solvent-accessible surface area (SA).[1] Second, we have also adapted the simple approximation that the hydrophobic energy is proportional to the nonpolar SA.[4,35,36,47] Our strategy is equivalent to the approaches used by Hawkins et al. and Qiu et al., in which polar atoms and nonpolar atoms are associated with different surface tentions (i.e., σ) for nonelectrostatic interactions.[4,35,36] The total energy of ligand binding, $G_{binding}$, is then given by the GB/SA scoring function[1]

$$G_{binding} = G_{LR}^{solvent} - G_L^{solvent} - G_R^{solvent} = G_{POL} + \beta VDW + \sigma_1 \Delta(SA_{hp}) - \sigma_2 \Delta(SA) \quad (17)$$

where VDW denotes the Lennard-Jones 6−12 potential, $\Delta(SA_{hp})$ and $\Delta(SA)$ denote the change in the hydrophobic and total solvent-accessible surface area upon ligand binding. Each energy term is scaled by a coefficient parameter, β, σ_1, and σ_2, respectively.

The time-consuming step for binding energy calculations with the GB/SA scoring function is to calculate the Born radii, which takes more than 90% of the total CPU time. Specifically, we have used a grid-based summation to compute the integral in eq 3 for the Born radius α_i[1]

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{dV}{4\pi}\sum_k^{in, r_{ik} > a_i} \frac{1}{r_{ik}^4} \quad (18)$$

where k represents a cell in the grid space that is occupied by the solute molecule but not atom i, dV is the unit volume of the cell, and r_{ik} is its distance to the center of atom i. The summation calculations are time-consuming. Moreover, any displacement of the surrounding atoms, such as when binding occurs, can cause changes in the atomic Born radii. Consequently, the Born radii need to be calculated for every ligand and receptor atoms before and after binding, which further prolongs the computing time. A new algorithm for Born radii calculations is needed to improve the efficiency of our scoring function.

2.2. Pairwise Formula for Born Radii Calculations. Recently several rapid, analytical approximations are proposed for Born radii calculations[4,35−41] (see ref 22 for review). The basic idea is to approximate the integral in eq 3 by a sum of contributions from each atom. If no two atoms overlap in a solute molecule, eq 3 will reduce to

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{4\pi}\sum_{j, j\neq i}^N \int_{\text{sphere j}} \frac{1}{r_{ij}^4} dV \quad (19)$$

For real solute molecules, Hawkins et al. proposed the following pairwise (or atom-based) approximation to account for the atomic overlaps:[4]

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \sum_{j, j\neq i}^N \int_{a_i}^{\infty} \frac{dr}{r^2} H_{ij}(r_{ij}, a_j)\, dr \quad (20)$$

where a_j represents the radius of atom j, r_{ij} represents the distance between atoms i and j, and H_{ij} is the fraction of the area of a sphere of radius r centered at atom i that is shielded by a sphere centered at atom j with a scaled radius $S_j a_j$.

The scaling factor S_j is unity if there is no overlap between any atomic spheres. In this case, eq 20 is exact and reduces to eq 19. For real molecules, however, if the S_j scaling factors were not used, the volume integral in eq 20 would be overestimated, causing systematic errors in Born radii calculations. Generally, $S_j < 1$, and thus, the effective volume for each atom j is reduced.

H_{ij} can be calculated analytically,[4] and eq 20 becomes[4,35]

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{2}\sum_j \mathscr{I}(a_i, S_j a_j, r_{ij}) \quad (21)$$

$$= \frac{1}{a_i} - \frac{1}{2}\sum_j \left[\left(\frac{1}{L_{ij}} - \frac{1}{U_{ij}}\right) + \left(\frac{S_j^2 a_j^2}{4 r_{ij}} - \frac{r_{ij}}{4}\right)\left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2}\right) + \frac{1}{2 r_{ij}} \ln\frac{L_{ij}}{U_{ij}}\right] \quad (22)$$

where

$$L_{ij} = \begin{cases} 1 & \text{if } a_i \geq r_{ij} + S_j a_j \\ \max(a_i, r_{ij} - S_j a_j) & \text{if } a_i < r_{ij} + S_j a_j \end{cases} \quad (23)$$

TABLE 1: Pairwise Scaling Parameters

atom	C	H	N	O	P	S
S_x	0.72	0.85	0.79	0.85	0.86	0.96

and

$$U_{ij} = \begin{cases} 1 & \text{if } a_i \geq r_{ij} + S_j a_j \\ r_{ij} + S_j a_j & \text{if } a_i < r_{ij} + S_j a_j \end{cases} \quad (24)$$

Notice that the summation in eq 22 is over the total number of atoms in consideration (usually <2000, see section 2.4). In contrast, the summation in eq 18 is over the total number of grid points (usually > 200 000). The atom-based or pairwise formula, eq 22, therefore greatly reduces the computational time on calculations of the Born radii.

For the receptor (R) atoms, we adopt the scaling parameters developed in the Tinker molecular modeling package[46] (shown in Table 1), where S_X^R depends only on its atom type X.

For the ligand (L) atoms, because the ligand atoms are more densely packed (i.e., more overlapped) than the receptor atoms, the S_X scaling factors should be reduced for the ligand atoms. In this paper, we set $S_X^L = 0.9S_X^R$ for all atom types.

2.3. Low-Dielectric Correction on Void Formation between the Protein and Misaligned Ligands. When performing virtual database screening, some ligand orientations being generated block only the mouth of the binding pocket and form a void in between. In this void region, the dielectric constant would be much less than that of the bulk water, because the buried water molecules therein, if any, have decreased entropy. The effect of the reduced effective dielectric constant enhances the electric interactions between the ligand and the receptor and thus offsets the entropic lost. To account for this effect, one needs to characterize the irregular boundary of the low-dielectric region, which is very time-consuming. The use of solvent-accessible surface rather than molecular surface[31] is a good strategy but cannot fully solve this problem, because the size of the low-dielectric region can be as large as a few water molecules.

To save computational time, we define the core atoms in a known inhibitor as "buried atoms" to roughly define the low-dielectric region.[1] For instance, methotrexate (MTX) is a well-known inhibitor of dihydrofolate reductase (dhfr); the crystal structure of the MTX-dhfr complex has been solved (PDB ID: 4dfr).[48] We define the bound MTX atoms excluding the dicarboxylate functionality up to the amide as the low-electric buried atoms for database screening against dhfr. Similarly, benzamidine is a small-size inhibitor of trypsin with known bound crystal structure (PDB ID: 3ptb); we set all atoms in benzamidine as buried atoms for trypsin. For the fatty acid-binding protein (fabp, PDB ID: 2ifb), we define the buried atoms as the carboxylate group and attached 7 methyl groups of the known inhibitor, palmitic acid (PLM). We define the radii of the buried atoms as follows: 0.5 Å for a buried atom that replaces a hydrogen atom of the inhibitor and 1.0 Å otherwise.

Obviously, our defined buried atoms overlap with some ligand atoms. To account for this overlap, we define the scaling parameter of a buried atom k as

$$S_{buried} = \begin{cases} 1 & \text{if } N_{deep} = 0 \\ 0.8 & \text{if } N_{deep} = 1 \\ 0 & \text{if } N_{deep} > 1 \text{ or } r_{ik} = 0 \end{cases} \quad (25)$$

where N_{deep} characterizes how much the buried atom k overlaps with other ligand atoms, and r_{ik} is its distance to the ligand atom

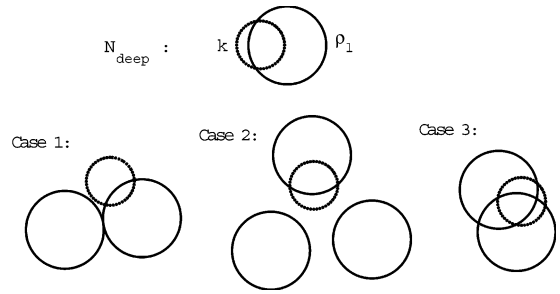

Figure 1. Illustration of atomic deep overlap for low-electric corrections. Solid circles (larger circles) represent ligand atoms, and dotted circles (smaller circles) represent low-electric buried atoms. Upper panel: An example of severe (or deep) overlap, that is, the center of the buried atom k falls into the scaled sphere of atom l ($r_{lk} < S_l a_l$). Lower panels: (1) No ligand atoms have severe overlap with the buried atom, $N_{deep} = 0$ for this buried atom; (2) Only one ligand atom has severe overlap with the buried atom, $N_{deep} = 1$; (3) Two ligand atoms have severe overlap with the buried atom, $N_{deep} = 2$.

i. Specifically, N_{deep} is defined as the number of the ligand atoms with which the buried atom k has severe overlap. A severe overlap means that the center of k falls into the scaled sphere of atom i, i.e., $r_{ik} < S_i a_i$. Figure 1 illustrates examples of $N_{deep} = 0$, 1 and > 1, respectively.

Accounting for the effect of the low-electric region, eq 22 for Born radii calculations remains the same before ligand binding, but changes to the following equation after ligand binding:

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{2}\sum_j \mathscr{H}(a_i, S_j a_j, r_{ij}) - \frac{\frac{1}{\epsilon_c} - \frac{1}{\epsilon_w}}{1 - \frac{1}{\epsilon_w}} \sum_k^{\text{buried}} \mathscr{H}(a_i, S_k a_k, r_{ik}) \quad (26)$$

where the third term on the right represents the contribution from the buried atoms, and ϵ_c is their dielectric constant. In this paper, we set $\epsilon_c = 1$, namely, the dielectric constant of the buried region is that of the vacuum. Equation 26 then reduces to

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \frac{1}{2}\sum_j \mathscr{H}(a_i, S_j a_j, r_{ij}) - \sum_k^{\text{buried}} \mathscr{H}(a_i, S_k a_k, r_{ik}) \quad (27)$$

2.4 Implementation of the Pairwise GB/SA Scoring Function in DOCK. Equation 17 gives the GB/SA scoring function, in which the effective Born radii can be calculated using the pairwise formulas, eqs 22 (before ligand binding) and 27 (after ligand binding). The pairwise GB/SA scoring function has been implemented in the DOCK software package[49] as an alternative scoring function, as briefly described below.

(1) To save computational time, only atoms within a user-specified box are used for binding energy calculations (eqs 8−17). The box is defined to enclose the ligand with an extra margin of the cutoff distance.

(2) The electrostatic energy calculations (eqs 9−11) require evaluation of the Born radii. The Born radius of every receptor atom within the user-defined box before ligand binding is precalculated using eq 22, in which the aforementioned cutoff is used for the summation. Note that, in order to calculate the Born radii of the receptor atoms near the box boundaries, their

TABLE 2: Effects of Cutoff on Electrostatic Energy Terms (in kcal/mol)

cutoff(Å)	$G_{screenedes}$	$G_{Ldesolve}$	$G_{Rdesolve}$	G_{POL}
(A) Bound Crystal Structure of dhfr-(Protonated) MTX (PDB ID: 4dfr)				
8	−68.7	37.5	16.0	−15.2
10	−69.3	38.1	15.8	−15.4
12	−68.1	38.1	15.9	−14.1
(B) Bound Crystal Structure of Trypsin−Benzamidine (PDB ID: 3ptb)				
8	−29.8	11.4	9.6	−8.8
10	−29.2	11.7	8.8	−8.7
12	−29.9	12.0	8.4	−9.5
(C) Bound Crystal Structure of fabp−PLM (PDB ID: 2ifb)				
8	−30.1	23.6	9.2	2.6
10	−30.1	24.1	9.4	3.4
12	−30.9	24.3	9.5	2.8

neighboring atoms within the cutoff distance but outside the box are taken into account. The results are saved for future ligand binding calculations. The Born radii of the ligand atoms and the ligand effect on the Born radii of the receptor atoms within the box are calculated on the fly. Because the summations involved in these calculations are over the number of atoms (2 orders of magnitude smaller than the number of grid points), the computational speed is significantly improved as compared with the grid-based approach.[1]

(3) By definition of the f_{GB} function (see section 2.1.2), for receptor atoms remote from the ligand, $(1/r_{ij} - 1/(f_{GB}^{LR}(r_{ij})))$, $(1/(f_{GB}^{L}(r_{ij})) - 1/(f_{GB}^{LR}(r_{ij})))$, and $(1/(f_{GB}^{R}(r_{ij})) - 1/(f_{GB}^{LR}(r_{ij})))$ approach 0. Namely, these remote receptor atoms contribute little to G_{POL}. As a test, we took the crystal structures of three protein−ligand complexes from the Protein Data Bank (PDB)[48] and calculated the electrostatic energies of these complexes at different cutoffs (8, 10, and 12 Å). Each cutoff was associated with a grid box of different size. Specifically, the grid boxes were generated using the SHOWBOX program of DOCK, with margins of 8 Å (for cutoff = 8 Å), 10 Å (for cutoff = 10 Å), and 12 Å (for cutoff = 12 Å), respectively. To allow a direct comparison, we did not minimize the ligand structures. The results are listed in Table 2, indicating that a cutoff of 8 Å appeared to be sufficient for electrostatic energy calculations.

(4) We use the minimization approach utilized in the current DOCK software[49,3] to optimize the ligand position on the six translational and rotational degrees of freedom during GB/SA energy scoring. Strictly speaking, the atomic Born radii vary as the ligand position varies. As a typical optimization step involves 50∼200 orientations with small RMSD (root-mean-square deviation) variations, we choose not to update the Born radii, the partial desolvation energy of the ligand, ($G_{L\ desolve}$) and the partial desolvation energy of the receptor ($G_{R\ desolve}$) during optimization except for the final ligand position in order to save computational time. Our tests show that this is a good approximation.

For the rest of the paper, we name this atom-based scoring function as the pairwise GB/SA function.

2.5. Optimization for the Parameters, β, σ_1, and σ_2. The goal of the scoring functions for structure-based drug design is to correctly predict the lead compounds. Thus, we use the following two criteria to determine the appropriate parameter regime for our GB/SA scoring function.

First, our predicted free energies of binding should have good correlations with experimental findings. We optimize the following error function representing the correlations

$$\mathrm{Err}_1 = \frac{1}{N}\sum_{i=1}^{N}[G_{\mathrm{binding},i}^{\mathrm{pred}} - (aG_{\mathrm{binding},i}^{\mathrm{exp}} + b)]^2 \qquad (28)$$

where $G_{\mathrm{binding},i}^{\mathrm{pred}}$ and $G_{\mathrm{binding},i}^{\mathrm{exp}}$ represent the calculated and measured binding energies of each inhibitor, N represents the total number of known inhibitors in the training set, and a and b are the parameters obtained from least-squares fitting.

Second, the scoring function should be able to distinguish good inhibitors from the rest compounds in any given chemical database. We minimize the error function Err_2, the number of false positive molecules (i.e., noninhibitive molecules that have more negative scores than the known inhibitors). In summary, the appropriate parameter regime should optimize all these error functions simultaneously.

TABLE 3: Deviation in Effective Born Radii Calculations between the Grid-Based Approach and the Pairwise Approach

	mean value of $1/\alpha$ (Å^{-1})	RMSD (Å^{-1})	relative error
receptor (trypsin)	0.32	0.026	8.1%
receptor (dhfr)	0.34	0.025	7.4%
ligand	0.51	0.019	3.7%

3. Results

3.1. Deviation of the Calculated Effective Born Radii from the Grid-Based Calculations. To test the pairwise formulas for effective Born radii (α), we have carried out the following calculations using both the atom-based (pairwise) and grid-based approach: We have calculated the Born radii of every receptor atom of dhfr and trypsin, respectively. We have also calculated the Born radii of every ligand atom of 100 randomly selected ligand molecules from the Available Chemicals Directory (ACD, distributed by Molecular Design Ltd., San Leandro, CA). The results are shown in Table 3. Using the grid-based calculation as a reference, we define the average value of $1/\alpha$, its RMSD, and relative error as

$$\overline{1/\alpha} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i^{\mathrm{grid}}} \qquad (29)$$

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\alpha_i^{\mathrm{atom}}} - \frac{1}{\alpha_i^{\mathrm{grid}}}\right)^2} \qquad (30)$$

$$\text{relative error} = \mathrm{RMSD}/\overline{1/\alpha} \qquad (31)$$

where N is the number of atoms in a molecule.

As shown in Table 3, the relative errors are less than 10%, indicating that the pairwise calculations well reproduce the effective Born radii results from the grid-based calculations.

3.2. Comparison of the Electrostatic Energy Components with the PB Approach. To further test the pairwise GB formulas, we have compared the electrostatic energy components obtained with the pairwise GB approach (eq 8) with the finite-difference PB (Poisson−Boltzmann) results. Specifically, we have chosen the following inhibitor-protein complexes with known binding energies and crystal structures:[48] protonated

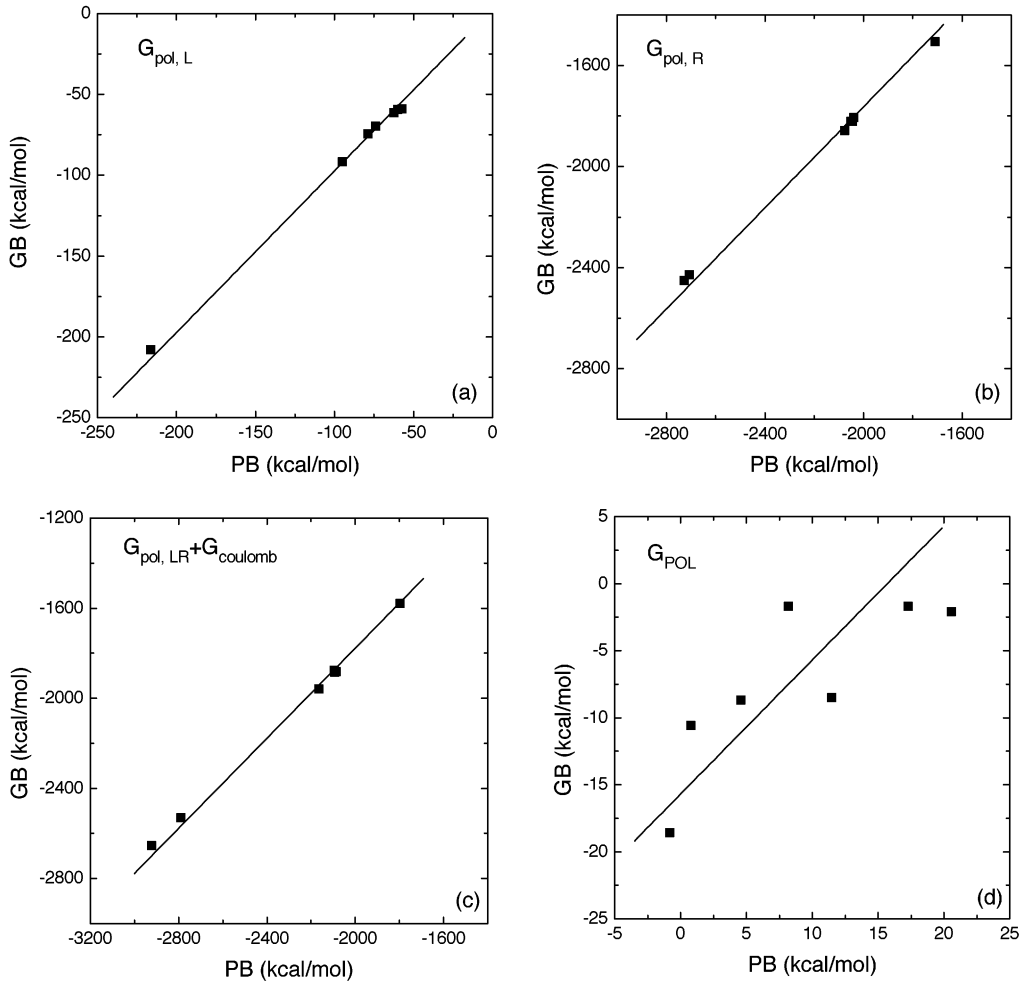

Figure 2. Comparison of the electrostatic component of the ligand−receptor binding energy, G_{POL} (see eq 12) with the GB and PB approaches. Comparisons are also made on individual terms of G_{POL} in eq 12. Every fitted line has a slope of 1. The GB data and PB data have good correlations except with some shifts.

methotrexate (MTX)-dhfr (PDB ID: 4dfr), protonated trimethoprim (TMP)-dhfr (0dfr), benzamidine-trypsin (3ptb), APPA-trypsin (1tpp), TAPAP-trypsin(1pph), NAPAP-trypsin(1ppc), and palmitic acid (PLM)-fabp (2ifb).

For the pairwise GB calculations, we used 0.4 Å for grid spacing and 8 Å for distance cutoff. With the PB approach, we used the DelPhi software,[50] with the same vdw radii and charge assignments as in the GB calculations. The dielectric constant of the ligand and protein molecules was set to 1 for consistency with the GB approach and with the AMBER protein parameter setup.[33,34] The grid spacing in PB was set to 0.5 Å, and "percentage fill of grid" was set to 90%. Other control parameters were set to default values of the software. To avoid the unphysical "grid energies",[50] we used the three-step thermodynamic process described in ref 51 (see Figure 2 therein) to calculate G_{POL} with DelPhi. Specifically, the process starts at the state before binding; namely, the ligand (L) and the receptor (R) are far apart in aqueous solution. The first step is to desolvate L and R separately and to calculate the respective polarization energies. The second step is to allow L and R to form a complex (LR) in a vacuum; the corresponding energy difference can be calculated by Coulomb's law (noticed that the dielectric constant is 1 everywhere). The final step is to solvate the LR complex and to calculate the polarization energy of the complex. The electrostatic component of the binding energy, G_{POL}, can then be calculated with eq 12 (see section 2.1.3).

This three-step thermodynamic process is able to get rid of the unphysical "grid energies" arising from the simple energy subtraction between the states before and after binding.[50]

In Figure 2, we plotted the electrostatic energy results with the pairwise GB approach and the PB approach. The GB data show reasonable correlation with the PB data, with a systematic shift. The fitted line in Figure 2d has a slope of 1 and a shift of 16.3 kcal/mol. We have also plotted the comparisons on individual terms in eq 12. Similar shifts have been reported previously on PB and GB comparisons.[41,44,52]

The average CPU time for calculating the electrostatic energy of ligand binding for a single ligand orientation is about 30s with DelPhi, using the above three-step thermodynamic process. The average CPU time with the GB approach is about 0.5s (including minimization of the ligand orientations). All calculations were carried out on a Silicon Graphics Octane R12000 workstation.

3.3. Ranking of Known Inhibitors. Three systems are tested in this paper: dihydrofolate reductase (dhfr), trypsin, and a fatty acid-binding protein (fabp). Around 10 000 molecules randomly selected from the ACD database are used as the compound database. We chose a random database so as to provide one of the error functions (Err_2) for our scoring function. Most of the molecules in this database are unlikely inhibitors and would have more positive binding energies than the known inhibitors. For each compound, a single CONCORD[53]-generated conformation was used, and the best orientation was selected by using

TABLE 4: Energy Scores and Rankings of Known Inhibitors with the Force Field Scoring Function and the Pairwise GB/SA Scoring Function

	ΔG_{exp} (kcal/mol)	force field score	rank	G_{POL} (kcal/mol)	VDW (kcal/mol)	Δ (SA) (Å ^{2})	Δ (SA_{hp}) (Å ^{2})	$G_{binding}$[a] (kcal/mol)	rank
MTX	−11.7[a]	−70.0	7	−18.63	−25.78	−922	−504	−32.89	1
TMP	−12.1[b]	−28.9	2402	−8.50	−20.90	−720	−497	−22.67	2
benzamidine	−6.4[c]	−28.9	2988	−10.55	−19.77	−400	−236	−22.31	6
APPA	−7.9[d]	−33.0	711	−14.07	−18.82	−477	−241	−24.23	3
TAPAP	−8.0[e]	−43.2	36	−1.51	−34.21	−740	−479	−22.91	5
NAPAP	−8.4[e]	−50.5	3	−1.86	−38.48	−850	−541	−25.72	1
PLM	−7.4[f]	−31.2	587	−1.65	−23.80	−958	−780	−21.10	1

[a] (β, σ_1, σ_2)= (0.6, 0.025, 0.015) References for experimental binding affinity data: (a) Poe et al. *J. Biol. Chem.* **1979**, *254*, 8143. (b) Kuyper et al. *J. Med. Chem.* **1985**, *28*, 303. (c) Mares-Guia et al. *J. Biol. Chem.* **1965**, *240*, 1579. (d) Sturzebecher et al. *Acta Biol. Med. Ger.* **1976**, *35*, 1665. (e) Turk et al. *FEBS Lett.* **1991**, *287*, 133. (f) Lowe et al. *J. Biol. Chem.* **1987**, *262*, 5931.

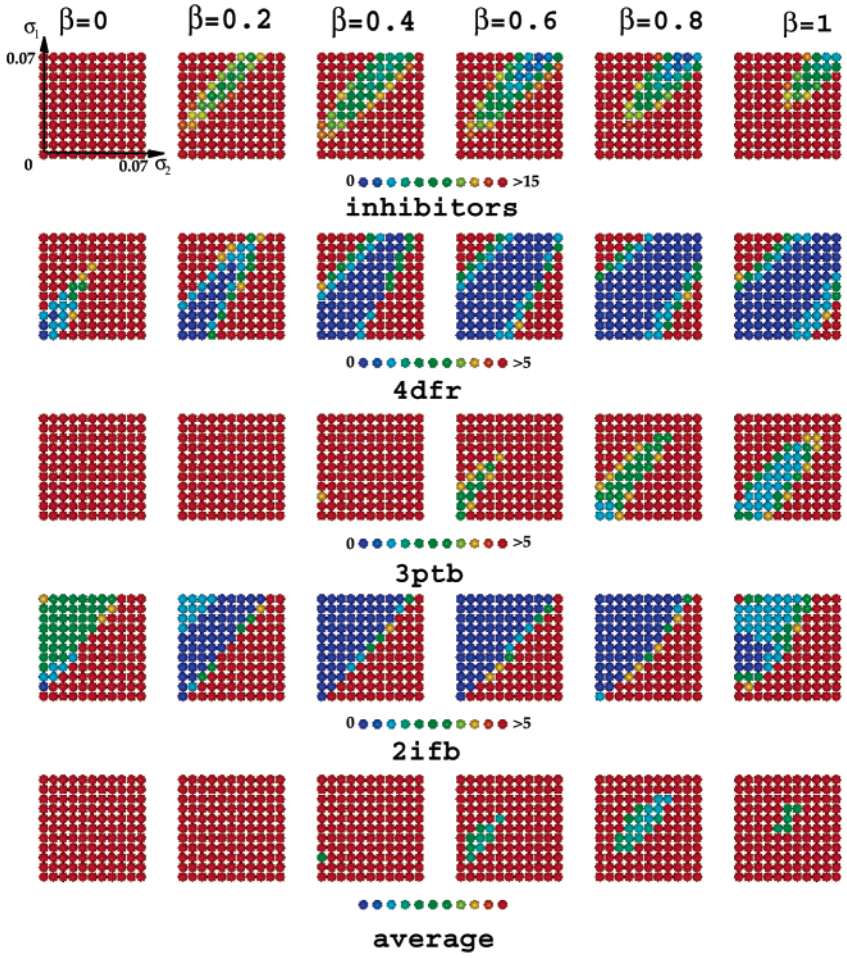

Figure 3. Physical parameter search. For each panel, the vertical and horizontal axes represent σ_1 and σ_2 in the unit of kcal/mol/Å^2. Different columns represent different β (as marked). The color represents the magnitude of the error functions, with low errors in blue and high errors in red. The first row shows the error function (Err_1) for the training set of known inhibitors. The second to fourth rows show the error functions (Err_2) for database screening against dhfr, trypsin and fabp, respectively. The last row shows the optimized error function, which is defined as the average value of the normalized error functions in the first four rows.

DOCK 4.0[49] for further pairwise GB/SA scoring. We ranked the compound database by applying the pairwise GB/SA scoring function. We also scored the known protein inhibitors tested in section 3.2 for comparison. We used 0.4 Å for grid spacing and 8 Å for distance cutoff. The pairwise algorithm takes about 0.5 s per orientation (including minimization with regard to the six translational and rotational degrees of freedom and an update of the Born radii and energies for the final ligand position) on a Silicon Graphics Octane R12000 workstation, around 50 times faster than the grid-based energy function.[1] The results are listed in Table 4.

By minimizing the error functions we previously defined for database screening and for the training set of known inhibitors, we tried to pinpoint the appropriate parameter regime for β, σ_1, and σ_2. The results are shown in Figure 3. The error functions are plotted in colors, with low errors in blue and high errors in red. The top row shows the error function (Err_1) for the training set. The next three rows show the error functions (Err_2) for database screening against dhfr, trypsin, and fabp, respectively. The appropriate parameter regime is obtained by optimizing these four error functions, which should be the light blue and green regions in the last row. The narrow yet finite size of the light blue and green regions indicates that there is no unique "best" parameter set. Instead, a range of "reasonable" choices are allowed, implying that the parameter choices are undetermined.

TABLE 5: Results of Orientational Test[a]

orientations	force field score	force field rank	$G_{POL} + VDW$	rank by $G_{POL} + VDW$	RMSD	VDW
A	−31.4	1	−20.55	5	0.36	−23.3
B	−31.2	2	−22.89	4	0.35	−23.2
C	−26.5	3	−13.03	7	0.56	−24.2
D	−26.0	4	−23.21	3	0.36	−23.8
E[f b]	−20.7	5	8.48	23	10.20	−23.5
F[f]	−20.5	6	6.12	14	10.19	−23.8
G[f]	−20.5	7	6.23	16	10.22	−23.7
H[f]	−20.3	8	6.16	15	10.20	−24.4
I[f]	−19.7	9	6.00	13	10.20	−25.4
J[f]	−19.5	10	6.66	19	10.22	−23.6
K[b]	−18.6	13	−24.27	1	0.32	−23.8
L	−17.4	16	−23.28	2	0.60	−16.3

[a] Superscript "f" indicates a flipped orientation. [b] Shown in Figure.

We found that the parameter set (β, σ_1, σ_2) = (0.6, 0.025, 0.015) chosen from the blue and green regions in the bottom panels not only yields a good correlation with the experimental binding affinity data for known inhibitors but also gives top rankings to these inhibitors out of the ACD compound database. Using this parameter set, least-squares fitting of Err_1 gives $a = 1.02$ and $b = -15.5$ (see eq 28). The fact that the slope a is close to 1 indicates that the predicted binding affinities differ from the corresponding measured binding affinities by a constant b within the error tolerance.

Importantly, the pairwise scoring function together with the parameter set (0.6, 0.025, 0.015) also produces good rank ordering for these enzyme inhibitors. For our random compound database of 10 000 molecules, TMP and MTX are ranked first and second in affinity for dhfr; NAPAP, APPA, TAPAP, and benzamidine are ranked nos. 1, 3, 5, and 6 for trypsin; and PLM is ranked top 1 for fabp. As a comparison, only NAPAP is ranked well for trypsin (no. 3) and MTX is ranked well for dhfr (no. 7) by the force field scoring function, whereas the other inhibitors are mis-ranked.

Though σ_1 and σ_2 show strong covariance in Figure 3, we are unable to remove either parameter because the shift in the quasi-linear relationship depends on the value of β. We chose σ_1 of 0.025 kcal/mol/Å^2 not only because this value locates in the allowable blue and green region but also because it is consistent with the finding from entropic measurements.

3.4. Orientational Test. Finally, we will test the ability of the pairwise GB/SA scoring function to identify good ligand orientations out of a sampling of orientations, an important criterion for a successful scoring function. We used the fabp−PLM complex as an example. We started with a random ligand orientation by performing a random displacement (>50 Å) and random rotation on the bound PLM position in the crystal structure. We then docked the PLM into the binding site of fabp, using DOCK to generate 50 orientations (i.e., setting "maximum_orientations" to 50). Each orientation was scored by the force field scoring function and the pairwise GB/SA scoring function, respectively. Optimizations are allowed upon scoring. Selected results are shown in Table 5.

As shown in Table 5, there are both "false positives" and "false negatives" among the force field scoring predictions. For example, orientations E−J are flipped from the position of PLM in the crystal structure (marked by a superscript "f"); however, their rankings by the force field scoring function are high (ranging from 5 to 10). These "false positive" orientations are ranked poorly by the pairwise GB/SA scoring function. Notice that the flipped orientations and correct orientations have very similar van der Waals scores (see the last column). Therefore, a contact score cannot differentiate the "flips". The differences in the electrostatic component of our scoring function enable us to identify the "flips". On the other hand, orientations K and L are close to the PLM orientation in the fabp−PLM crystal structure but receive poor force field scores. These "false negative" orientations are correctly identified by the pairwise GB/SA scoring function. Figure 4 has shown the crystal structure orientation of PLM in cyan, the "false negative" orientation K in black, and the "false positive" orientation E in yellow, and the bound fabp-PLM crystal receptor structure in green. To display the orientational flip, the oxygen atoms in the important carboxylate group of PLM are colored in red.

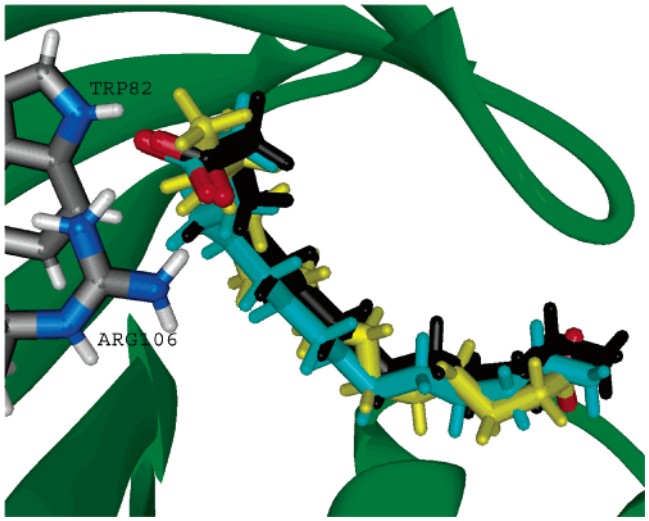

Figure 4. Results of the orientational test. The crystal structure of the fabp-PLM complex is shown, with fabp in green and PLM in cyan. The "false negative" and "false positive" orientation from force filed scoring are shown in black and yellow, respectively. The oxygen atoms of PLM are plotted in red.

4. Discussion

Based on the continuum solvent model, we have developed a computationally efficient method to account for the desolvation effect in ligand-protein interactions. When calculating energies of ligand binding, the DOCK force field scoring function, despite good success, does not account for the dependence of charge−charge interactions on the local environment. It also tends to underestimate electrostatic interactions between charges in close proximity and totally ignores the desolvation effect. The pairwise GB/SA scoring function offers physically reasonable and computationally efficient calculations of the electrostatic component of the binding energies by accounting for these effects. Furthermore, two nonelectrostatic terms are also included in

the free energy model: the change of van der Waals interactions between the solutes (L and R) and water upon ligand binding and a hydrophobic component. The last term comprises the entropic penalty associated with the reorganization of the solvent molecules around the ligand and the protein.

The pairwise formula for Born radii calculations increases the computational speed by around 50-folds comparing to the previous grid-based algorithm. Since it takes only 0.5s to score one orientation (including minimization of the orientation) on a Silicon Graphics Octane R12000 workstation, the pairwise GB/SA scoring function can be used for direct screening of a database or combinatorial library. It can also be used for post-DOCK screening. For example, for a database that contains 100 000 molecules, if we first use DOCK to find the best orientation for each molecule for further scoring, it will take less than a day to use the pairwise scoring function to screen these 100 000 molecules.

Our generalized pairwise GB/SA model for ligand binding differs from the pairwise GB/SA scheme for single molecule solvation in several ways. First, compared with the electrostatic component of the solvation energy for a single molecule which has only one term, the electrostatic contribution to binding free energies is composed of three terms; the screened ligand−receptor electrostatic energy, the partial desolvation energy of the ligand, and the partial desolvation energy of the receptor. Second, during virtual database screening, it is not uncommon to find ligand orientations that occlude the mouth region of the binding pocket, forming a void in between. The dielectric constant in this void region would be low because the buried water molecules therein, if any, have limited mobility due to hydrogen bonding. This unique low-dielectric effect in ligand binding has been accounted for in this paper. Third, the van der Waals interactions involved in ligand binding are more complex than in single molecule solvation. Here, we have used the Lennard-Jones 6−12 potential to characterize the van der Waals interactions between the ligand and the receptor and assumed the ligand-solvent and receptor-solvent van der Waals interactions to be linearly proportional to their solvent-accessible surface areas.

There are several ways to further develop our pairwise scoring function. First, physically reliable charge assignments give correct molecular polarity and therefore may contribute to more accurate calculations of the electrostatic component of binding energies.[1,9,54] For instance, one may use the atomic partial charges derived from quantum mechanical approaches (e.g., RESP[55]) or semiempirical quantum mechanical approaches (e.g., AMSOL[56]). Second, the GB/SA model including the scaling parameters, S_js (see eq 20), need to be re-parametrized for ligand binding studies. The re-parametrization has been done for small ligands with the AMBER force field[39] and for proteins and nucleic acids with the CHARMM force field.[41] Yet, the re-parametrization has not been worked out for ligand−protein complex systems. The Poisson−Boltzmann (PB) model can be used as a reference, using one of the PB solvers such as MEAD,[57] DelPhi,[50] ZAP,[18] UHBD,[58] and APBS.[59] Third, it may be a necessity to include an explicit hydrogen bonding term accounting for hydrogen bond formation between the ligand and the protein. Classical electrostatic treatment may underestimate energy contributions from hydrogen bonds formed between polar groups due to the ignorance of charge redistribution. Upon hydrogen bonding formation, charges are found to be redistributed among the acceptor group and the donor group.[60−63] Thus, accounting for such a charge redistribution effect may improve our model. Fourth, the current scoring function has left out conformational entropy terms. Last, more test systems need to be examined with the final scoring function. Exploration of these issues is underway.

To summarize, we have developed a rapid and reasonably accurate model to include the solvation effect for ligand binding affinity calculations. A procedure is also developed to account for the low-dielectric region that might form between the ligand and the receptor during docking processes. It is possible to use our algorithm to screen a large database directly.

Acknowledgment. The authors thank Dr. Robert Rizzo for his valuable comments on the manuscript. Support from Tripos, Inc. (St. Louis, MO) and MDL Information Systems, Inc. (San Leandro, CA) is gratefully acknowledged. Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081). This work is supported by the Research Board Award of the University of Missouri, AHA grant 0265293Z (Heartland Affiliate), NIH grant DK61529 (X.Z.) and NIH grant GM31497 (I.D.K.).

References and Notes

(1) Zou, X.; Sun Y.; Kuntz I. D. *J. Am. Chem. Soc.* **1999**, *121*, 8033−43.
(2) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−9.
(3) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505−24.
(4) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−9.
(5) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847−63.
(6) Leach, A. R. *Molecular Modeling: Principles and Applications*; Longman: Singapore, 1996
(7) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71−90.
(8) Born, M. *Z. Phys.* **1920**, *1*, 45−8.
(9) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins* **1999**, *34*, 4−16.
(10) Kangas, E.; Tidor, B. *J. Chem. Phys.* **2000**, *112*, 9120−31.
(11) Lamb, M. L.; Jorgensen, W. L. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−57.
(12) Wang, W.; Donini, O., Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−43.
(13) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−200.
(14) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.
(15) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243−52.
(16) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327−35.
(17) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−45.
(18) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608−40.
(19) You, T. J.; Harvey, S. C. *J. Comput. Chem.* **1993**, *14*, 484−501.
(20) Cortis, C. M.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1570−90.
(21) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 815−20.
(22) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−52.
(23) Tobias, D. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 253−61.
(24) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−99.
(25) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **1997**, *101*, 8098−106.
(26) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **1998**, *102*, 3637−41.
(27) Luo, R.; Head, M. S.; Moult, J.; Gilson, M. K. *J. Am. Chem. Soc.* **1998**, *120*, 6138−46.
(28) Best, S. A.; Merz, K. M.; Reynolds, C. H. *J. Phys. Chem. B* **1999**, *103*, 714−26.
(29) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712−20.
(30) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. *Proteins* **1999**, *37*, 88−105.
(31) Majeux, N.; Scarsi, M.; Caflisch, A. *Proteins* **2001**, *42*, 256−68.
(32) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591−607.
(33) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Prefeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(34) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(35) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−39.
(36) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−14.
(37) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−90.
(38) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571−6.
(39) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465−71.
(40) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629−33.
(41) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, *103*, 3765−73.
(42) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9.
(43) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426−34.
(44) Tsui, V.; Case, D. A. *J. Phys. Chem. B* **2001**, *105*, 11314−25.
(45) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891−913.
(46) Ponder, J. W. http://dasher.wustl.edu/tinker
(47) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−76.
(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−42.
(49) Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175−89.
(50) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128−137.
(51) Gilson, M. K.; Honig, B. *Proteins* **1988**, *4*, 7−18.
(52) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Fiesner, R. A. *J. Phys. Chem. B* **1997**, *101*, 1190−7.
(53) Rusinko, A., III,; Sheridan, R. P.; Nilakantan, R.; Nilakantan, K. S.; Bauman, N.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251−5.
(54) Wei, B. Q. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W., Shoichet, B. K. *J. Mol. Biol.* **2002**, *322*, 339−55.
(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.
(56) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 6.7.2; University of Minnesota: Minneapolis, MN, 2001.
(57) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473−86.
(58) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *91*, 57−95.
(59) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−41.
(60) Scheiner, S. *Hydrogen Bonding*; Oxford University Press: New York, 1997.
(61) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997.
(62) van der Vaart, A.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1999**, *121*, 9182−90.
(63) Hobza, P.; Havlas, Z. *Chem. Rev.* **2000**, *100*, 4253−64.