Continuum Solvation Models in the Linear Interaction Energy Method

Jens Carlsson, Martin Andér, Martin Nervall, and Johan Åqvist*

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden

Received: November 29, 2005; In Final Form: April 19, 2006

The linear interaction energy (LIE) method in combination with two different continuum solvent models has been applied to calculate protein−ligand binding free energies for a set of inhibitors against the malarial aspartic protease plasmepsin II. Ligand−water interaction energies are calculated from both Poisson−Boltzmann (PB) and Generalized Born (GB) continuum models using snapshots from explicit solvent simulations of the ligand and protein−ligand complex. These are compared to explicit solvent calculations, and we find close agreement between the explicit water and PB solvation models. The GB model overestimates the change in solvation energy, and this is caused by consistent underestimation of the effective Born radii in the protein−ligand complex. The explicit solvent LIE calculations and LIE-PB, with our standard parametrization, reproduce absolute experimental binding free energies with an average unsigned error of 0.5 and 0.7 kcal/mol, respectively. The LIE-GB method, however, requires a constant offset to approach the same level of accuracy.

1. Introduction

Computer-based estimation of protein−ligand binding affinities is of great interest for pharmaceutical applications and can contribute to efficient development of new drugs. The free energy of binding can be estimated from Monte Carlo (MC) or molecular dynamics (MD) simulations by using free-energy perturbation (FEP) and thermodynamic integration (TI) techniques.[1] Although accurate in theory, FEP/TI methods are computationally demanding and suffer from severe convergence problems that are mostly connected to the annihilation and creation of atoms in simulations. To use this approach in computer-aided drug design, where large libraries of compounds are screened, shorter computation times and faster convergence is necessary. For these reasons, it is useful to find approximate expressions for the free energy; efforts to date range from simple scoring functions to more advanced approximations of FEP and TI based on MD or MC simulations.[2] The linear interaction energy (LIE) method[3] is a semiempirical model that has become widely used to predict protein−ligand binding affinities. In LIE, the free ligand in water and the solvated protein−ligand complex are simulated and from these two calculations the ligand−surrounding electrostatic and van der Waals energies are collected. The binding free energy is then evaluated from

$$\Delta G_{\mathrm{bind}}^{\mathrm{LIE}} = \beta(\langle U_{\mathrm{l-s}}^{\mathrm{el}}\rangle_{\mathrm{bound}} - \langle U_{\mathrm{l-s}}^{\mathrm{el}}\rangle_{\mathrm{free}}) + \alpha(\langle U_{\mathrm{l-s}}^{\mathrm{vdW}}\rangle_{\mathrm{bound}} - \langle U_{\mathrm{l-s}}^{\mathrm{vdW}}\rangle_{\mathrm{free}}) + \gamma \quad (1)$$

where $\langle\ \rangle$ represents MD or MC averages of intermolecular electrostatic (el) and van der Waals (vdW) energies for the bound and free ligand with its surroundings (l−s). The first term in eq 1 is derived using the linear response approximation, which predicts that β is equal to 0.5 (see ref 4 for details). Although this approximation holds very well for calculating solvation energies of ionic compounds, Åqvist and Hansson found that there are deviations from linear response for uncharged solutes.[4] This can be taken into account by changing the value of β between 0.33 and 0.5 depending of the chemical nature of the ligand, and including these corrections improved the LIE estimates of protein−ligand affinities significantly.[5] The second term in eq 1 represents the nonpolar contributions to the free energy of binding, and they are assumed to have a linear relationship with the surrounding van der Waals energies. This approximation is based on the observation that solvation energies of nonpolar compounds are linearly correlated with the surrounding van der Waals energies.[3,6] The scaling factor, α, is empirical, and a value of 0.18 has worked well for a large number of systems.[1,5,7] The parameter γ is a constant term that can be added to get the correct absolute binding free energies. It depends on the nature of the binding site, that is, it is the same for a set of ligands that bind to the same receptor, and has recently been shown to correlate with the hydrophobicity of the protein active site.[7]

To obtain converged free-energy calculations using MD or MC simulations, it is necessary for one to sample not only the energetically available conformations of the solute but also the solvent degrees of freedom. A continuum solvent model can efficiently replace the water by treating it as a continuous medium having a dielectric constant of 80. By using this approach, one can obtain the average solvation properties of water without averaging over the interactions of thousands of explicit water molecules, which in simulations is connected to large fluctuations in solute−solvent and solvent−solvent energies. For these reasons, continuum models are now used widely in different applications (see refs 8 and 9 for recent reviews), but there are cases in which a continuum treatment of water can cause problems. For example, crystal structures of protein−ligand complexes often reveal that conserved water molecules bridge interactions between ligand and protein and it is not clear if these interactions are reproduced accurately by a continuum model.[10] Unfortunately, there are rather few comparisons of explicit and continuum solvent binding free-energy calculations,[11] and it is therefore of considerable interest to further evaluate the usefulness of the continuum approach in this context.

* Corresponding author. E-mail: aqvist@xray.bmc.uu.se. Phone: +46 16 471 4109. Fax: +46 18 536971.

10.1021/jp056929t CCC: $33.50
Published on Web 05/25/2006

The Poisson−Boltzmann (PB) continuum model, for zero salt concentration, reduces to solving the Poisson equation

$$\nabla\cdot\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) + 4\pi\rho(\mathbf{r}) = 0 \quad (2)$$

where $\epsilon(\mathbf{r})$ is the dielectric constant, $\phi(\mathbf{r})$ is the electrostatic potential, and $\rho(\mathbf{r})$ is the charge distribution of the system. By numerically solving the PB equation[12] for a solute in vacuum ($\epsilon = 1$) and the dielectric medium ($\epsilon = 80$), one can obtain the electrostatic solvation energy from

$$\Delta G_{\mathrm{el}}^{\mathrm{PB}} = \frac{1}{2}\sum_i q_i\,(\phi^{80}(\mathbf{r}_i) - \phi^{1}(\mathbf{r}_i)) \quad (3)$$

where the sum is over all (partial) charges of the solute. In recent years, the approximate Generalized Born (GB) continuum model[13] has also become popular. The GB equation has the form

$$\Delta G_{\mathrm{el}}^{\mathrm{GB}} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_i\sum_j \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i\alpha_j\,\mathrm{e}^{-r_{ij}^2/(4\alpha_i\alpha_j)}}} \quad (4)$$

where q_i is the atomic charge of atom i and r_{ij} is the distance between atoms i and j. The effective Born radius, α_i, depends on the surrounding atoms and can be viewed as the average distance from the atom to the dielectric boundary. The effective Born radii can be defined more precisely in terms of PB solvation energies as

$$\alpha_i = -\frac{166}{\Delta G_{\mathrm{el},i}^{\mathrm{PB}}}\left(1 - \frac{1}{\epsilon}\right) \quad (5)$$

where $\Delta G_{\mathrm{el},i}^{\mathrm{PB}}$ is the polarization free energy (kcal/mol) with a unit charge on atom i and all other charges of the solute set to zero. The main advantage of the method is, however, that the radii can be estimated from a simpler analytical formula, which in most cases is parametrized to reproduce radii from eq 5. Remarkably, the GB model can often reproduce electrostatic solvation energies within a few percent of the corresponding PB calculation, but with considerably shorter computation times. For macromolecules, a general weakness of all GB models is the difficulty to calculate accurate radii for buried atoms.[14,15] It has also been shown that although the total free energy of solvation often is a good approximation of the corresponding PB calculation, partitioning these into self ($i = j$ in eq 4) and shielding ($i \neq j$ in eq 4) term energies may yield incorrect results.[16,17] Although this is not a problem for applications where the entire solute's solvation energy is sought, it is of major importance for calculation of interaction energies. For example, in the LIE method, only the terms of eq 4 in which the ligand is involved are used in the calculation of ligand−water interaction energies and if the self and interaction terms of the protein are not balanced correctly, then the GB model cannot be used for these types of calculations.

Continuum electrostatic ligand−water interaction energies are here incorporated in the LIE method by using an equivalent form of eq 1

$$\Delta G_{\mathrm{bind}}^{\mathrm{LIE}} = \beta(\langle U_{\mathrm{l-prot}}^{\mathrm{el}}\rangle_{\mathrm{bound}} + \langle U_{\mathrm{l-wat}}^{\mathrm{el}}\rangle_{\mathrm{bound}} - \langle U_{\mathrm{l-wat}}^{\mathrm{el}}\rangle_{\mathrm{free}}) + \alpha\Delta\langle U_{\mathrm{l-s}}^{\mathrm{vdW}}\rangle + \gamma \quad (6)$$

where the ligand−surrounding electrostatic energies have been partitioned into ligand−protein (l−prot) and ligand−water (l−wat) components. The van der Waals term (vdW) represents the same difference as in eq 1. By extracting snapshots of the free ligand and the protein−ligand complex from explicit water calculations, one can estimate the ligand−water interaction energies ($U_{\mathrm{l-wat}}^{\mathrm{el}}$) using continuum models. For a ligand free in solution, the electrostatic ligand−water interaction energy is twice the calculated continuum solvation energy of transfer of the ligand from vacuum to a homogeneous medium of dielectric constant 80, that is, $\langle U_{\mathrm{l-wat}}^{\mathrm{el}}\rangle_{\mathrm{free}} = 2\Delta G_{\mathrm{el}}^{\mathrm{PB/GB}}$. This relationship is based on the fact that the continuum solvent models used in this work, similar to the LIE method, rely on the linear response approximation. For the bound state, the ligand−water interaction energy is twice the difference between one calculation where all charges and radii of the protein−ligand complex are present and a second calculation, where the only difference from the first is that all ligand charges are turned off. In the GB model, the bound-state ligand−water interaction energy can be calculated by including only the terms to which the ligand contributes in eq 4.

Using MD simulations to sample solute degrees of freedom and estimate the solvent contribution a posteriori from a continuum model has become an attractive approach in calculations of protein−ligand binding affinities.[18,19,24,42] One example is the MM-PBSA method proposed by Kollman and coworkers.[18,19] In this method, the configurational free energy of a state is calculated from

$$G = \langle E_{\mathrm{MM}}\rangle + \langle E_{\mathrm{PBSA}}\rangle - T\langle S_{\mathrm{MM}}\rangle \quad (7)$$

where $\langle E_{\mathrm{MM}}\rangle$ is the average molecular mechanics energy (both intra- and intermolecular energies), $\langle E_{\mathrm{PBSA}}\rangle$ is the solvation energy obtained from PB and weighted-surface-area (SA) calculations, and $T\langle S_{\mathrm{MM}}\rangle$ is the solute entropy estimated from quasiharmonic (QHA) or normal mode (NMA) analysis.[18] The binding free energy is then calculated from

$$\Delta G_{\mathrm{bind}}^{\mathrm{MM-PBSA}} = G_{\mathrm{complex}} - G_{\mathrm{receptor}} - G_{\mathrm{ligand}} \quad (8)$$

For ligand binding calculations all energies in MM-PBSA are usually estimated from one explicit solvent MD simulation of the complex.[20] All terms for the complex in eq 7 are then obtained using snapshots from this simulation. The two remaining energy contributions in eq 8 (G_{receptor} and G_{ligand}) are calculated by simply removing the ligand or receptor from the trajectory. This approach is taken because if separate simulations of the receptor and ligand were to be carried out, then the difference in eq 8 would be dominated by fluctuations in intramolecular protein−protein energies, leading to very serious convergence problems. In the "single-trajectory" approach, all energetic contributions from the MM force field except receptor−ligand interaction energies cancel. This may, at first glance, seem similar to the LIE approach, but in contrast, the MM-PBSA method assumes that the conformation of the ligand and receptor is the same in both the bound and unbound state, which is very questionable. In the LIE method, however, the receptor conformational response to binding is implicitly taken into account by different weight factors. For example, in the classical test case of ion hydration the electrostatic linear response factor of $^1/_2$ accurately encompasses all of the conformational response of the solvent[4] and the same goes, for example, for ion binding to crown ethers.[21] Also, the entropic term in MM-PBSA has been shown to be difficult to estimate. NMA is a drastic simplification and is often based just on a single or a few structures of the solutes.[20] QHA might seem to be a better alternative because it is based on the atomic covariance matrix obtained from the simulation rather than a single minimized

structure. However, using a single simulation of the complex and retaining only the partial covariance matrix corresponding to either the free ligand or receptor would be inconsistent. In addition, the accuracy of QHA for flexible molecules has been questioned recently.[22] Surprisingly, the MM-PBSA approach has given some impressive results for ligand binding[20,23] even though the "single trajectory" approach must be considered a rather crude approximation.

Zhou et al. have proposed the SGB-LIE method,[24] in which they used the surface-GB model[25] for generating trajectories/configurations and for evaluation of ligand−water interaction energies. In these simulations, the number of atoms is reduced drastically by replacing explicit solvent with an efficient GB model and computation times decrease significantly, but it is not clear whether GB continuum models are accurate enough to generate realistic trajectories of biomolecules.[26−28] The binding free energy for the protein−ligand complexes were estimated as a sum of three weighted terms: the change in ligand−surrounding van der Waals and electrostatic energies, and a cavity term. In their evaluation of ligand−water interaction energies, a different and somewhat questionable approach was taken. For the bound state, only half of the shielding term in the GB equation ($i \neq j$ in eq 4) was used in the evaluation of ligand−water interaction energies for terms comprising one ligand and one protein atom, while theoretically the full shielding term should be included. This can be understood by considering the free energy of charging the ligand in the protein binding site, which is what the electrostatic LIE term represents. The estimated free energy in LIE is the difference between the two endpoints of this transformation, which is obtained by removing the energies involving only the protein in eq 4. Still, using three free scaling parameters, the SGB-LIE method has given good agreement with experimental binding data.[24]

Here, PB and GB solvent models are used to calculate ligand−water interaction energies for a set of eight inhibitors of the malarial aspartic protease plasmepsin II (Plm II). The continuum calculations are compared to recently published explicit water calculations on the same ligands.[29] We calculate binding free energies using the three different solvation models (explicit water, PB, and GB), compare these to experimental data, and discuss the possible advantages of using continuum models in the linear interaction energy approach.

2. Methods

The explicit water simulations of the eight Plm II inhibitors have been described elsewhere, and the compounds are shown in Table 1.[29] In the work of Ersmark et al.,[29] the inhibitors were superimposed onto the scaffold of a similar ligand with a known structure. Two possible binding modes of the inhibitors were identified, and explicit water MD simulations were carried out for both orientations (binding mode 1 and mode 2). The MD simulations were carried out with the program Q[30] using the GROMOS87[31] force field. LIE calculations predicted that all ligands had the same mode of binding (mode 2),[29] and the present work is based on new simulations of these models. For all of the ligands, we obtain free energies close to those published earlier.

The continuum ligand−water interaction energies were calculated using snapshots from explicit water MD simulations of the protein−inhibitor complex and the free inhibitor.[29] Between 400 and 900 snapshots from 1 to 2 ns trajectories were used for each ligand both in the bound and free states. In accordance with the explicit water calculations, only residues within 20 Å of the ligand center were included in the calculations. All PB calculations were carried out with the DelPhi[32,33] program. The grid spacing was set to 0.5 Å, and the solute filled 80% of the total grid box. The salt concentration was set to zero, and the dielectric constants of the solutes and

TABLE 1: Plm II Inhibitors Used in the Calculations and Experimental Inhibitory Activities (nM)[29]

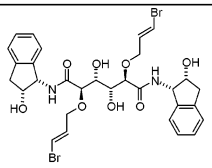
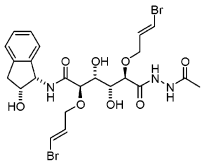
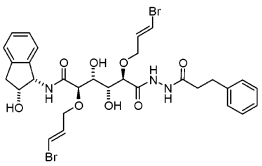
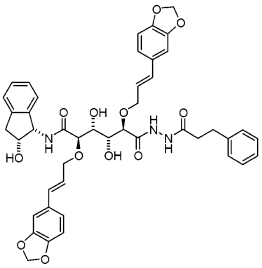
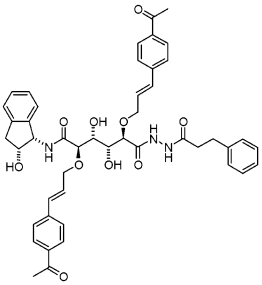
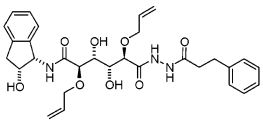
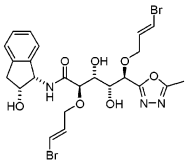
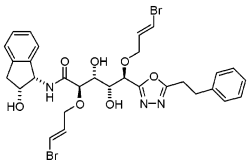
Compound	Structure	Enzyme K_i (nM)
2		47
5		359
9		7
14		11
15		145
16		161
19		5000
23		2900

TABLE 2: Ligand−surrounding (l−s) van der Waals (vdW) and Electrostatic (el) Energies (kcal/mol) from Explicit Solvent Simulations

ligand	$\langle U^{el}_{l-s}\rangle_{free}$	$\langle U^{vdW}_{l-s}\rangle_{free}$	$\langle U^{el}_{l-s}\rangle_{bound}$	$\langle U^{vdW}_{l-s}\rangle_{bound}$
2	-74.7 ± 0.3	-49.6 ± 0.4	-74.0 ± 0.2	-102.6 ± 1.3
5	-72.8 ± 0.7	-46.4 ± 0.3	-74.2 ± 1.0	-90.4 ± 0.3
9	-70.5 ± 0.4	-56.9 ± 0.3	-73.4 ± 0.9	-108.2 ± 0.4
14	-75.1 ± 0.9	-71.1 ± 0.2	-75.8 ± 0.7	-128.9 ± 0.3
15	-85.0 ± 0.5	-69.2 ± 1.4	-83.4 ± 0.9	-127.5 ± 0.4
16	-69.9 ± 1.3	-47.4 ± 0.5	-72.3 ± 0.4	-94.8 ± 0.4
19	-73.8 ± 0.1	-47.6 ± 0.2	-74.4 ± 0.7	-87.5 ± 0.1
23	-73.5 ± 0.4	-55.0 ± 0.5	-73.4 ± 0.2	-103.4 ± 0.0

TABLE 3: Binding Free Energies (kcal/mol) Using the LIE, LIE-PB, and LIE-GB Models[a]

ligand	ΔG^{expt}_{bind}	ΔG^{LIE}_{bind}	ΔG^{LIE-PB}_{bind}	ΔG^{LIE-GB}_{bind}
2	-10.1	-9.3 ± 0.5	-10.2 ± 0.3	-9.7 ± 0.4
5	-8.9	-8.4 ± 0.6	-6.7 ± 0.3	-6.9 ± 0.4
9	-11.2	-10.2 ± 0.6	-10.7 ± 0.5	-10.0 ± 0.5
14	-10.9	-10.6 ± 0.6	-10.7 ± 0.4	-12.0 ± 0.3
15	-9.4	-10.0 ± 0.8	-9.8 ± 0.5	-9.1 ± 0.3
16	-9.3	-9.3 ± 0.7	-8.6 ± 0.4	-9.3 ± 0.5
19	-7.3	-7.4 ± 0.3	-7.0 ± 0.2	-8.2 ± 0.1
23	-7.6	-8.7 ± 0.3	-8.6 ± 0.4	-11.0 ± 0.2
<|error|>		0.5	0.7	1.2

[a] For LIE and LIE-PB, the parameters $\alpha = 0.18$, $\beta = 0.33$, and $\gamma = 0$ were used. For the LIE-GB model, $\alpha = 0.18$, $\beta = 0.33$, and $\gamma = -3.6$ kcal/mol were used. Experimental binding free energies were calculated from $\Delta G^{exp}_{bind} = RT \ln K_i$.

water were set to 1 and 80, respectively. For each calculation, 500 iterations were performed to ensure that convergence was obtained. The GB calculations were based on the model presented by Qui et al.,[15] hereafter referred to as GB^{Still}, and performed with the program Q.[30] For the PB and GB calculations, all of the radii, except for polar hydrogens, were set to the GROMOS87 van der Waals radii (0.5σ) used in interactions with water oxygen atoms. For polar hydrogens, a radius of 1.0 Å was used.

3. Results and Discussion

The ligand−surrounding electrostatic and van der Waals energies from the explicit water simulations are presented in Table 2. Using the standard LIE parametrization[5] (for these compounds: $\beta = 0.33$, $\alpha = 0.18$, and $\gamma = 0.0$), the explicit solvent calculations reproduce experimental binding free-energy values with an average unsigned error of 0.5 kcal/mol (cf. Table 3 and Figure 1), which is impressive considering the size and chemical diversity of the inhibitors. The structural analysis of the binding of these inhibitors to Plm II based on the MD simulations also gave directions for further optimization of possible drugs against malaria.[29] It can be seen from Table 2 that the nonpolar contribution is responsible for the main part of the overall binding free energy in all cases. However, the small but significant electrostatic contributions are critical for establishing the correct ranking among the compounds, as has been observed earlier.[29] For example, the relative binding free energy between **9** and **15** is determined entirely by the 4.5 kcal/mol difference in the electrostatic contribution ($\Delta\Delta\langle U^{el}_{l-s}\rangle$) and this effect can be traced to a deficient solvation of the acetylphenyl side chains in **15** when bound to the enzyme. It is also worth noting that if the electrostatic coefficient in eq 1 is set to zero, while optimizing α, or α and γ, then significantly worse models are obtained with mean unsigned errors of about 0.8 kcal/mol. These models are only marginally better than the null hypothesis of setting all binding free energies equal to the average ($\beta = 0$, $\alpha = 0$, and $\gamma = -9.34$), and the main difference with respect to the latter is that ligand size (related to the hydrophobic contribution) is captured through the van der Waals energies. A more detailed analysis that relates the binding free energy for each inhibitor to interactions with the protein can be found in ref 29 and will not be discussed here.

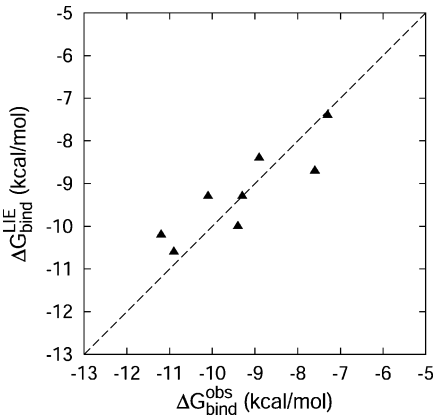


Figure 1. Calculated and experimental binding free energies for the eight Plm II inhibitors using the standard LIE model with explicit solvent (kcal/mol).

To compare explicit and continuum solvent representations, we calculated the ligand−water interaction energies for the bound and free state of the ligand with the explicit, PB, and GB^{Still} models. It is most relevant, however, to compare the change in ligand−water interaction upon binding, $\Delta\langle U^{el}_{l-wat}\rangle = \langle U^{el}_{l-wat}\rangle_{bound} - \langle U^{el}_{l-wat}\rangle_{free}$, for the different solvation models. Some of the systematic differences between explicit and continuum solvent representations will cancel when using this approach, such as the size of the simulation sphere (20 Å vs infinite radius). In addition, the difference in electrostatic solvation energy is likely to be less sensitive to choice of atomic radii and grid resolution. It should also be noted that our objective here is not the parametrization of atomic radii to reproduce explicit solvent calculations, which would require extensive FEP calculations to be carried out.

3.1. Possion−Boltzmann Model. The calculated PB ligand−water interaction energies are shown in Table 4 and are consistently somewhat more positive than the corresponding explicit solvent results. The average unsigned deviation is 3.8 and 4.4 kcal/mol for the free and bound state, respectively. Because the deviations are of similar magnitude for both the bound and free state, the change in PB electrostatic ligand−water interaction energies are in good agreement with the explicit solvent calculations, which are shown in Table 5 and Figure 2. The average unsigned deviation for these energy differences is 1.7 kcal/mol, which corresponds to an average free-energy deviation of 0.6 kcal/mol when multiplied by the β coefficient. The convergence of the PB calculations is illustrated for inhibitor **19** in Figure 3, where the ligand−water interaction energies for 200 snapshots using PB and explicit solvent models are shown. The variance of the PB energies is three times smaller than that for the explicit water calculations. Note, however, that the fluctuations of the PB interaction energies are still fairly large, thus implying that conformational averaging/sampling is indeed essential for flexible molecules. The correlation between the explicit and continuum energies is also evident from Figure 3. The average convergence error for all ligands in the free state, calculated as the difference between the first and second half of the simulation, is about half as large with PB compared to the explicit solvent calculations. In the bound state, solute−solute and solute−solvent hydrogen bonds are formed and

TABLE 4: Electrostatic Ligand−Water (l−wat) Interaction Energies (kcal/mol) for the Plm II Inhibitors from Explicit, PB, and GB^{Still} Water Models

	$\langle U^{el}_{l-wat}\rangle_{free}$			$\langle U^{el}_{l-wat}\rangle_{bound}$		
ligand	explicit	PB	GB^{Still}	explicit	PB	GB^{Still}
2	-74.7 ± 0.3	-67.7 ± 0.1	-70.5 ± 0.0	-12.0 ± 0.4	-7.6 ± 0.2	2.0 ± 0.0
5	-72.8 ± 0.7	-71.9 ± 0.2	-74.0 ± 0.4	-14.2 ± 0.5	-8.2 ± 0.2	-0.1 ± 0.2
9	-70.5 ± 0.4	-66.4 ± 0.4	-68.3 ± 0.4	-9.8 ± 0.1	-7.3 ± 0.2	3.9 ± 0.1
14	-75.1 ± 0.9	-73.1 ± 0.1	-75.0 ± 0.3	-11.2 ± 0.5	-9.6 ± 0.2	-4.5 ± 0.9
15	-85.0 ± 0.5	-82.2 ± 0.2	-85.2 ± 0.0	-14.5 ± 0.3	-11.1 ± 0.1	-1.2 ± 0.6
16	-69.9 ± 1.3	-67.2 ± 0.4	-69.3 ± 0.7	-15.0 ± 0.3	-10.2 ± 0.7	-3.6 ± 0.5
19	-73.8 ± 0.1	-68.8 ± 0.3	-67.3 ± 0.1	-19.7 ± 0.4	-13.7 ± 0.1	-4.8 ± 0.2
23	-73.5 ± 0.4	-67.3 ± 0.1	-63.9 ± 0.2	-16.4 ± 0.9	-10.2 ± 0.0	-3.0 ± 0.8
<|error|>		3.8	3.1		4.4	12.7

TABLE 5: Change in Electrostatic Ligand−Water (l−wat) Interaction Energies (kcal/mol), $\Delta\langle U^{el}_{l-wat}\rangle = \langle U^{el}_{l-wat}\rangle_{bound} - \langle U^{el}_{l-wat}\rangle_{free}$, for the Plm II Inhibitors from Explicit, PB, and GB^{Still} Water Models

	$\Delta\langle U^{el}_{l-wat}\rangle$		
ligand	explicit	PB	GB^{Still}
2	62.7 ± 0.7	60.2 ± 0.3	72.5 ± 0.1
5	58.6 ± 1.2	63.7 ± 0.4	74.0 ± 0.6
9	60.7 ± 0.5	59.1 ± 0.6	72.2 ± 0.5
14	63.8 ± 1.4	63.5 ± 0.2	70.5 ± 1.2
15	70.5 ± 0.7	71.1 ± 0.3	84.0 ± 0.6
16	54.9 ± 1.5	57.0 ± 1.0	65.7 ± 1.2
19	54.1 ± 0.5	55.2 ± 0.4	62.5 ± 0.3
23	57.0 ± 1.3	57.1 ± 0.2	60.8 ± 1.0
<|error|>		1.7	10.0

broken during the simulations and the fluctuations of the separate ligand−water and ligand−protein components are not entirely relevant because there is strong compensation between the terms. Nevertheless, the convergence errors here are also about half as large for the bound-state ligand−water interaction energies when using the PB model. To test the dependence of our results on the chosen radii set, we also calculated ligand−water interaction energies using the PARSE[34] parameter set (results not shown). The relative differences in ligand−water interaction energies obtained using the PARSE radii are in excellent agreement with the force-field radii result (Figure 4), but a constant value of −5.6 kcal/mol has to be added to each PARSE energy to reproduce the absolute explicit solvent contribution to the binding free energies. Hence, the force-field radii seem to be a more appropriate choice of radii for these calculations and the result might also reflect that PARSE[34] radii were derived using a different set of partial charges and a solute dielectric constant equal to two.

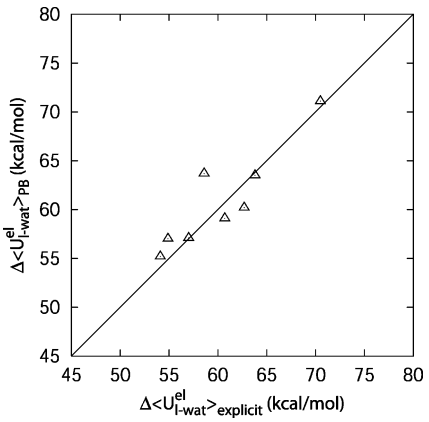


Figure 2. Change in electrostatic ligand−water interaction energies (kcal/mol) between the free and bound states for the PB and explicit solvent models.

To calculate binding free energies with the PB solvent representation (LIE-PB), we insert the ligand−water interaction energies into eq 6. Using the standard LIE parametrization ($\beta = 0.33$, $\alpha = 0.18$, and $\gamma = 0.0$), we find that the average unsigned error compared to experiment is 0.7 kcal/mol and the calculated values are shown in Table 3 and Figure 5. For all inhibitors, except ligand **5**, the experimental binding affinity is reproduced with similar accuracy as in the explicit water calculations. The binding free energy for ligand **5** is overestimated with 2.2 kcal/mol, which is considerably worse than that for the explicit solvent calculations. Surprisingly, the binding affinity of ligand **9**, which is similar to **5**, is predicted quite well. This result seems to be partially connected to very specific ligand−water interactions in the protein binding site, which might not be described accurately by a continuum model. For example, there are three structurally conserved water molecules during the MD simulation when ligand **5** is bound to Plm II, while in the complex with ligand **9** only one of these is present (Figure 6).

As discussed in the Introduction, the MM-PBSA method[18] assumes that the conformation of the ligand is the same in both protein and water, which may be a severe approximation for large flexible ligands**.** To test this approach, we extracted conformations of ligand **2** from the simulation of the complex and calculated the ligand−water interaction energy, $\langle U^{el}_{l-wat}\rangle_{free}$, using PB for the isolated ligand. The average ligand−water interaction energy is then -76.6 ± 0.3 kcal/mol, which differs by 8.9 kcal/mol from the calculation based on the explicit water simulation of the free ligand and would thus make the free energy of binding 3 kcal/mol more positive for ligand **2**. This example shows that, although PB is less conformationally sensitive than explicit solvation, the assumption of identical ligand conformations in the bound and free states used in MM-PBSA is clearly questionable.

Previously, continuum solvent models have also been used in various methods similar to the LIE approach.[35,36] These studies were concerned mainly with improving the speed of the LIE method by using a continuum water representation in combination with different sampling techniques. Good results were obtained for large sets of ligands using fast methods such as minimization or MD simulations in vacuo employing a distance-dependent dielectric for electrostatic interactions.[35,36] Using sampling in vacuo is a fast method for exploring conformational space, but it might yield unrealistic trajectories.[27] The use of a single structure will be sensitive to choice of minimization scheme, and from our experience conformational sampling (in explicit solvent) generally improves calculated binding free energies even for empirical scoring functions.[37]

3.2. Generalized Born Model. The calculated GB^{Still} ligand−water interaction energies for the free state are very similar to the explicit solvent results and have an average unsigned

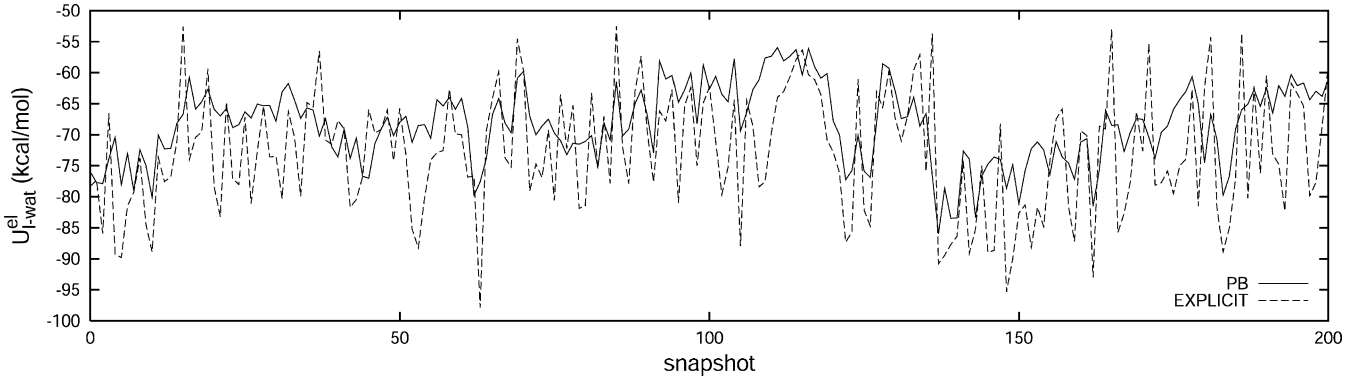


Figure 3. Ligand−water interaction energy (kcal/mol) along an MD trajectory for ligand **19** free in water using explicit (dashed line) and PB (solid line) water representations.

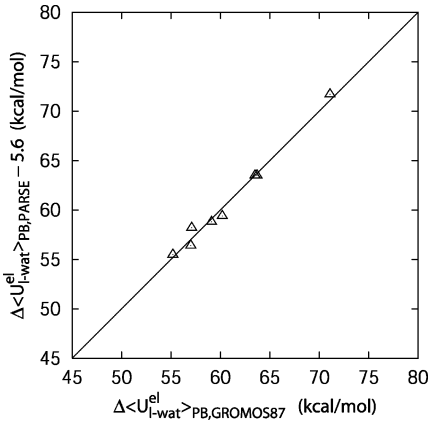


Figure 4. Change in electrostatic ligand−water interaction energies (kcal/mol) upon binding for the PB model using GROMOS87 van der Waals radii (0.5σ) and PARSE radii. The PARSE energies have been shifted with −5.6 kcal/mol.

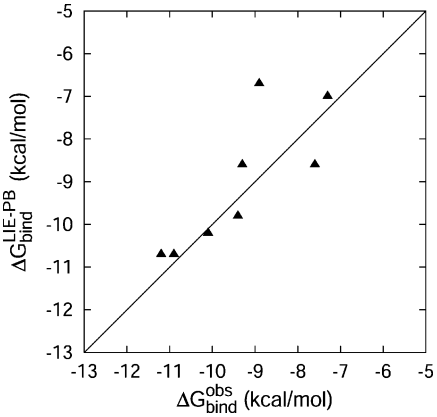


Figure 5. Calculated LIE-PB and experimental binding free energies (kcal/mol) for the eight PlmII inhibitors.

deviation of 3.1 kcal/mol, but the bound-state interaction energies, however, are overestimated by about 13 kcal/mol for each ligand (Table 4). Consequently, the change in ligand−water interaction energies calculated with the GB^{Still} model are not in agreement with either the explicit or PB solvation model (Table 5). Still, it can be seen from Figure 7 that the essential features of the change in solvation upon binding are captured by the GB^{Still} model and that a systematic shift of all of the calculated differences by −10.0 kcal/mol would improve the result significantly.

A correction for the GB^{Still} model offset can, of course, also be incorporated into the constant γ in eq 1 ($\gamma = -3.6$ kcal/mol) and binding free energies were calculated in this way, utilizing the original LIE parametrization ($\beta = 0.33$ and $\alpha = 0.18$). The experimental binding free energies are then reproduced with an average unsigned error of 1.2 kcal/mol (Table 3 and Figure 8). Good agreement with experimental data for six out of eight inhibitors is obtained. The binding free energy of inhibitor **5** is, in accordance with the PB calculations, again overestimated while inhibitor **23** is predicted to bind too strongly to PlmII. A possible explanation for the latter observation is found by comparing the difference between the ligand−surrounding electrostatic energies in the free state. The GB^{Still} energies are 2−3 kcal/mol more negative than the corresponding PB energies for all inhibitors except **19** and **23**, for which the ligand−surrounding energies instead increase. The overestimated binding affinity of inhibitor **23** appears therefore to be related to the oxadiazole ring, which is present in both **19** and **23**. The effective Born radii for the polar atoms of the group might be overestimated because of the proximity of the atoms in the ring, leading to an increased solvation energy and underestimated binding free energy.

Although systematic differences in absolute GB and PB solvation energies are observed quite often,[38] these may cancel when differences are taken. Therefore, it is interesting to try to find the reason for the difference between PB and GB^{Still} estimates of ligand−water interaction energies. Although PB, GB^{Still}, and explicit solvent interaction energies for the free state of the ligands are quite similar, there are large differences for the bound-state interaction energies. This observation suggests that the GB^{Still} model has difficulties estimating ligand−water interaction energies in the protein−ligand complex. It should be pointed out that the GB^{Still} model was parametrized to work with a different force field than the one used here, which could have some effect on the accuracy of the model. However, this is probably not the main reason because GROMOS87[31] and OPLS[39] radii are quite similar. Otherwise, the errors can arise from two known problems of current GB models: either the effective Born radii are not estimated correctly[17] or the form[16] of eq 4 might be inappropriate. To investigate the former of these possibilities, we calculated the exact effective Born radii, according to eq 5, for a snapshot of ligand **9** in the bound and unbound states. The GB^{Still} effective Born radii for the free ligand (Figure 9) are in good agreement with the "exact" radii. For the enzyme−ligand complex, however, the radii are underestimated consistently (Figure 10). Our results are in agreement with the known difficulty of estimating Born radii of buried atoms in proteins and are also similar to a recent comparison of different GB models, where the model applied here was found to underestimate the effective Born radii when used with the GROMOS96 force field.[14] In this context, it should also be mentioned that there are numerous methods that can be

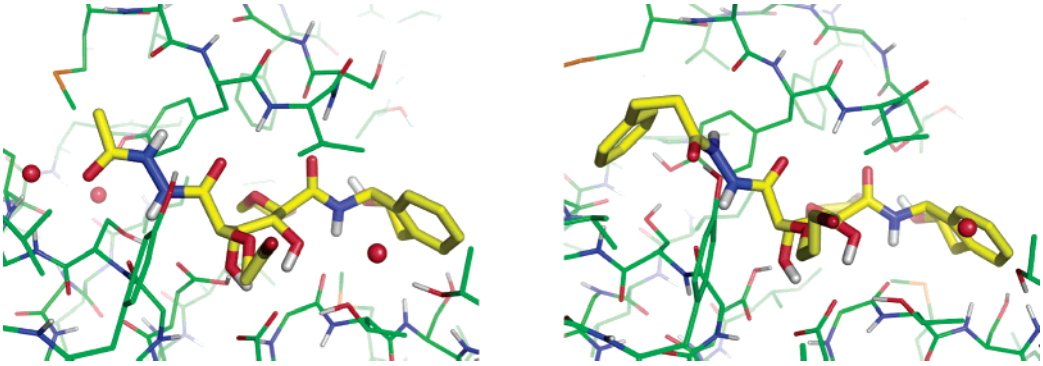


Figure 6. Conserved water molecules during MD simulations (depicted as red spheres) with ligand **5** (left) and **9** (right) bound to PlmII.

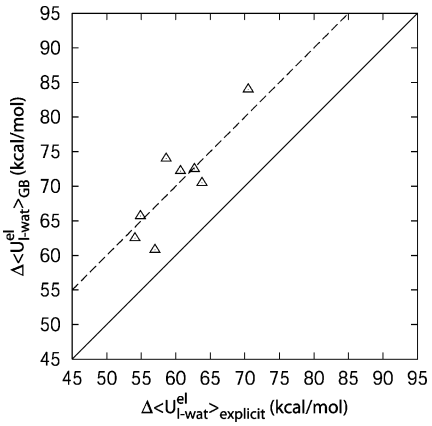


Figure 7. Change in electrostatic ligand−water interaction energies (kcal/mol) upon binding for the GB^{Still} and explicit solvent models. The dashed line shows the ideal correlation with a constant offset of −10.0 kcal/mol added to each calculated GB^{Still} value.

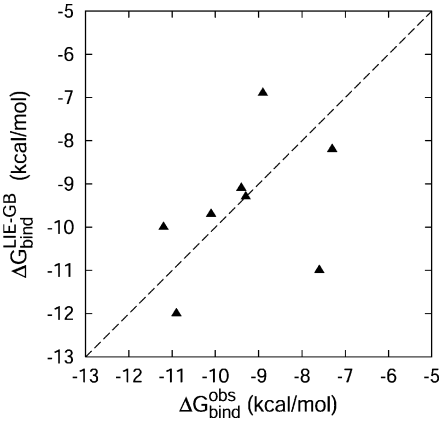


Figure 8. Calculated LIE-GB and experimental binding free energies (kcal/mol) for the eight PlmII inhibitors with a constant $\gamma = -3.6$ kcal/mol used in the LIE equation.

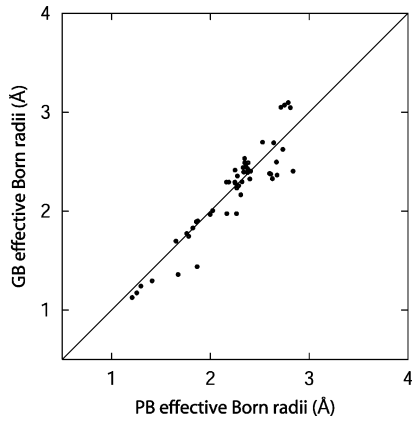


Figure 9. PB (exact) and GB^{Still} effective Born radii (Å) for one snapshot of ligand **9** free in water.

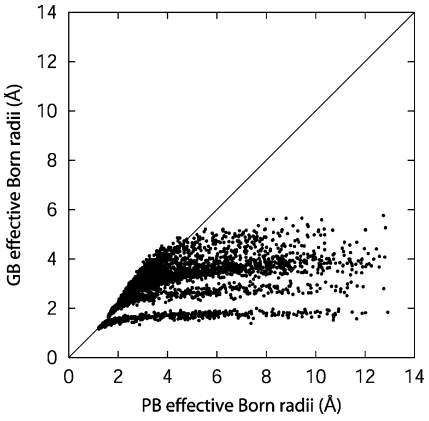


Figure 10. PB (exact) and GB^{Still} effective Born radii (Å) for one snapshot of ligand **9** in complex with Plm II. Radii for both ligand and protein atoms are shown.

used to estimate effective Born radii and many of these have been shown to reproduce PB solvation energies and radii considerably better than the GB^{Still} model.[14,38]

The calculated exact radii can be used in eq 4 to test if the results would improve.[17] As an example, this was carried out for one snapshot of bound and free ligand **9**, and the results are presented in Table 6. For the free ligand, the GB result is similar to PB, but using the exact radii (GB$^{exact\ radii}$) yields better agreement for self and shielding terms compared to the radii of Qui et al.[15] For the bound state, there is a significant difference between PB and GB electrostatic energies. The self term is considerably too negative because the calculated radii in the complex are too small, but the error is reduced by an error of the same magnitude in the shielding term. In total, the ligand−water interaction energy is overestimated by 8.6 kcal/mol using GB^{Still} radii, but if exact radii are used then the self and shielding terms are improved significantly and GB^{Still} comes quite close to the PB result. From our results, it is evident that eq 4 yields considerably better agreement with the PB model given that the correct effective Born radii are used. However, for calculations of binding free energies, where sometimes very small differences in energy are sought, it is still not completely clear if the GB model can achieve the required accuracy. For example, it should be noted that the total electrostatic ligand−surrounding energy for the free ligand deviates by 5 kcal/mol from the PB result using the exact radii.

TABLE 6: Self ($i = j$ in Eq 4) and Shielding ($i \neq j$ in Eq 4) Components of the Ligand−Water Interaction Energies (kcal/mol) for One Snapshot of Ligand 9 in the Bound and Free State[a]

ligand	component	PB	GB^{Still}	$GB^{exact\ radii}$
free	self	−560.1	−583.3 (23.2)	−560.1 (0.0)
	shielding	491.9	512.8 (20.9)	497.2 (5.3)
	total	−68.3	−70.5 (2.3)	−62.9 (5.3)
bound	self	−240.7	−410.0 (169.3)	−240.7 (0.0)
	shielding	234.8	412.6 (177.9)	233.6 (1.2)
	total	−6.0	2.6 (8.6)	−7.2 (1.2)

[a] The unsigned errors compared to the PB model are given in parentheses after each value.

4. Conclusions

In this study, we have compared explicit and continuum solvent representations in the LIE method. For a set of eight Plm II inhibitors, the change in ligand−water electrostatic energy from the PB continuum models agrees well with explicit solvent simulations. The difference in interaction energies calculated with the GB^{Still} model showed a systematic shift from the PB calculations, and a significantly better agreement was obtained if a constant offset is added to each calculated value. For the most approximate description of solvent, namely the (empirical) GB^{Still} model, we found that the effective Born radii in the protein−ligand complex were underestimated consistently and this was also shown to be the main reason for the discrepancy between GB and PB descriptions of inhibitor binding. The present work, despite a relatively small dataset, indicates that good agreement for the change in solvation energy upon binding can be obtained by simply using the van der Waals radii of the applied force field. We want to stress, however, that reliable (absolute) continuum solvation energies can, in principle, only be obtained by careful parametrization of the radii based on explicit water calculations. This has been performed for some force fields,[40,41] but not the one used here.

For binding free energies, the original parametrization of the LIE method can be used with both explicit and PB solvent models and the calculated binding free energies are in good agreement with experimental values, which is somewhat remarkable considering the approximative nature of continuum models. The comparison of force-field-derived radii with those of the PARSE data set[34] for the PB method, however, shows that the radii parametrization is important, particularly if accurate absolute energies are sought. For the GB^{Still} solvation model, the relative binding free energies are also in fair agreement with experiment, but because there was a systematic shift in the ligand−water interaction energies, even with a seemingly appropriate radii set, a different value of the constant γ was then required to reproduce absolute binding free energies. The results presented herein illustrate the usefulness of the LIE method in combination with short sampling times and continuum solvent models, in agreement with the conclusions of Zhou et al.[24] To further assess the usefulness of continuum models for estimating binding free energies, evaluation of other datasets and more accurate GB models should be carried out. This approach could then be applied to larger sets of inhibitors and contribute to fast and efficient ligand design.

Acknowledgment. Support from the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF/Rapid) is gratefully acknowledged.

References and Notes

(1) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. *Adv. Protein Chem.* **2003**, *66*, 123.
(2) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2645.
(3) Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385.
(4) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512.
(5) Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27.
(6) Carlson, H. A.; Jorgensen, W. L. *J. Phys. Chem.* **1995**, *99*, 10667.
(7) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. *J. Comput. Chem.* **2004**, *25*, 1242.
(8) Baker, N. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137.
(9) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217.
(10) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers* **2003**, *68*, 16.
(11) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591.
(12) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(13) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(14) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008.
(15) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(16) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.
(17) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.
(18) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
(19) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.
(20) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786.
(21) Marelius, J.; Hansson, T.; Åqvist, J. *Int. J. Quantum Chem.* **1998**, *69*, 77.
(22) Chang, C. E.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, *1*, 1017.
(23) Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. *J. Med. Chem.* **2002**, *45*, 1412.
(24) Zhou, R. H.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388.
(25) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
(26) Krol, M. *J. Comput. Chem.* **2003**, *24*, 531.
(27) David, L.; Luo, R.; Gilson, M. K. *J. Comput. Chem.* **2000**, *21*, 295.
(28) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(29) Ersmark, K.; Nervall, M.; Hamelink, E.; Janka, L. K.; Clemente, J. C.; Dunn, B. M.; Blackman, M. J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. *J. Med. Chem.* **2005**, *48*, 6090.
(30) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Modell.* **1998**, *16*, 213.
(31) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos B. V.: Nijenborgh, Groningen, The Netherlands, 1987.
(32) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.
(33) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.
(34) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(35) Zoete, V.; Michielin, O.; Karplus, M. *J. Comput. Aided Mol. Des.* **2003**, *17*, 861.
(36) Huang, D.; Caflisch, A. *J. Med. Chem.* **2004**, *47*, 5791.
(37) Marelius, J.; Ljungberg, K. B.; Åqvist, J. *Eur. J. Pharm. Sci.* **2001**, *14*, 87.
(38) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265.
(39) Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(40) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(41) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2005**, *1*, 484.
(42) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins* **2000**, *39*, 393.