Multiscale generalized Born modeling of ligand binding energies for virtual database screening

Hao-Yang Liu[†], **Sam Z. Grinter**[†], and **Xiaoqin Zou**[*]

Department of Physics & Astronomy, Department of Biochemistry, Dalton Cardiovascular Research Center, and Informatics Institute University of Missouri, Columbia, MO 65211

Abstract

Generalized Born (GB) models are widely used to study the electrostatic energetics of solute molecules including proteins. Previous work demonstrates that GB models may produce satisfactory solvation energies if accurate effective Born radii are computed for all atoms. Our previous study showed that a GB model which reproduces the solvation energy may not necessarily be suitable for ligand binding calculations. In this work, we studied binding energetics using the exact GB model, in which Born radii are computed from the Poisson-Boltzmann (PB) equation. Our results showed that accurate Born radii lead to very good agreement between GB and PB in electrostatic calculations for ligand binding. However, recently developed GB models with high Born radii accuracy, when used in large database screening, may suffer from time constraints which make accurate, large-scale Born radii calculations impractical. We therefore present a multiscale GB approach in which atoms are divided into two groups. For atoms in the first group, those few atoms which are most likely to be critical to binding electrostatics, the Born radii are computed accurately at the sacrifice of speed. We propose two alternative approaches for atoms in the second group. The Born radii of these atoms may simply be computed by a fast GB method. Alternatively, the Born radii of these atoms may be computed accurately in the free state, then a variational form of a fast GB method may be used to compute the change in Born radii experienced by these atoms during binding. This strategy provides an accuracy advantage while still being fast enough for use in the virtual screening of large databases.

Keywords

generalized Born model; ligand-protein binding energies; molecular docking; solvation

1 Introduction

The process of structure-based drug design would benefit greatly from the ability to compute protein-ligand binding affinity accurately, but today's computing power places sharp restrictions on what can be done [1–3]. This compromise between accuracy and efficiency is particularly evident in large database screening, in which the time available to consider any one compound is brief. One substantial bottleneck is accounting for the electrostatic effects of water. Water molecules can be modeled explicitly, but this is very computationally intensive due to the many degrees of freedom available to the molecules enveloping the protein and ligand. A less time-consuming approach is to treat water as a continuum dielectric medium, referred to as implicit solvent models. There are two widely-used types of implicit solvent models [5]. In the first type, the Poisson-Boltzmann (PB) equation is solved numerically by

[*]Corresponding author: Xiaoqin Zou, 134 Research Park, Columbia, MO 65211, zoux@missouri.edu, 573-882-6045 (tel.), 573-884-4232 (fax).
[†]H.-Y. Liu and S.Z. Grinter contributed equally to this work.

finite-difference, finite-element, or boundary-element methods [6–9]. The computational cost of this physics-based approach still limits its practicality for large database screening. The second type is the generalized Born (GB) model [10]. It is essentially an approximation of the PB method and can typically be performed faster (see [11,12] for review). In developing GB models, electrostatic properties calculated with the PB method are often used to calibrate the empirical parameters introduced in GB models and also to test GB model accuracy. It was found that with very accurate Born radii, GB models can closely reproduce the PB solvation energies of macromolecules [13].

In recent years, a variety of models have been developed to calculate the effective Born radii. The early models assumed the Coulomb field approximation [10,14–20], in which the electric displacement ($\mathbf{D}_i$) induced by an atomic point charge is crudely estimated by a Coulomb field ($\mathbf{D}_i \sim \frac{q_i \mathbf{r}}{r^3}$). This is not true in general for heterogeneous dielectric media, such as a protein surrounded by water [11]. Additional approximations have been introduced to improve the computational efficiency with minor sacrifices in accuracy. One example is the method proposed by Hawkins, Cramer, and Truhlar (which we denote the HCT GB model) [14]. All these models yield some level of accuracy in Born radii calculations. Recently, empirical corrections to the Coulomb field approximation were proposed [21–23], and very recently, an integral expression referred to as GB-R6 was developed for Born radii calculations [24–26]. Analysis showed that GB-R6 methods give very accurate electrostatic solvation free energies [27]. In summary, there are numerous models for Born radii calculations, with a wide range of accuracy and efficiency. Given that certain atoms are likely to be more influential than others in determining the electrostatic solvation energy, we propose a multiscale method in which the Born radii of certain atoms are calculated using accurate GB models, while other atoms are accounted for using more rapid approximations, in order to produce a better compromise between accuracy and efficiency. In this work, to treat certain atoms accurately we used the exact GB model, in which Born radii are derived from the PB equation (PGB model), despite that it is unreasonably computationally demanding in practice. We use the PGB model to represent the most accurate GB models.

The paper is organized as follows. The method section will first briefly review the GB and PB methods and explain how the PB equation is used to derive exact Born radii, then will introduce a variational form of the HCT GB model for substantially more efficient Born radii calculations. Several criteria will be presented for validation of GB models using PB results as a reference. The results section will describe our multiscale GB approach and validate various multiscale models using the proposed criteria. Finally, the paper will conclude with a brief discussion of the proposed models.

2 Methods

2.1 Brief review of the generalized Born model

GB models approximate the electrostatic component of the solvation energy (referred to as the polarization energy and denoted G_{pol}) as follows [10]:

$$G_{\mathrm{pol}} = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\sum_i^N \frac{q_i^2}{\alpha_i} - \frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\sum_{i\neq j}^N \frac{q_i q_j}{f_{GB}(r_{ij})}$$

(1)

where the solvation energy is defined as the energy cost to transfer a solute molecule from a low dielectric medium to a high dielectric medium. For Eq. 1, the low dielectric medium is a homogeneous environment and has the same dielectric constant as that of the solute molecule, denoted ε_p. The high dielectric medium is water, taken to have dielectric constant $\varepsilon_w = 80$. q_i

and q_j are the charges of atoms i and j in the molecule, and r_{ij} is the distance between them. N is the total number of solute atoms.

The first term on the right-hand side of Eq. 1 is the self-energy term in the polarization energy; its accuracy depends on the accuracy of the effective Born radii. The effective Born radius of atom i, α_i, is a variable in GB models representing how deeply the atom is buried in the molecule. The second term on the right hand side of Eq. 1 is the cross-energy term.

The $f_{GB}(r_{ij})$ function is introduced to capture the distance-dependence of the damping effect of water on the electrostatic interaction between atoms i and j. The most commonly-used form of $f_{GB}(r_{ij})$ was proposed by Still and colleagues [10]:

$$f_{GB}(r_{ij})= \sqrt{r_{ij}^2+\alpha_i\alpha_j\exp[-r_{ij}^2/(4\alpha_i\alpha_j)]}$$

(2)

When two atoms are far apart, $f_{GB}(r_{ij})$ approaches r_{ij} and the electrostatic interaction between the two becomes Coulombic. When $i=j$ (and therefore $r_{ij}=0$), $f_{GB}(r_{ij})$ becomes α_i, the effective Born radius of the atom, and the interaction energy becomes self-energy. When the distance is in between, the $f_{GB}(r_{ij})$ form is purely empirical.

2.2 Accurate Born radii derived from the Poisson-Boltzmann Method

The PB method is an implicit solvation approach in which water is treated as a continuum dielectric medium and the PB equation is solved. If the ionic effect can be ignored, the PB equation is reduced to the following Poisson equation:

$$\nabla \cdot \varepsilon(r)\nabla\varphi(r)= - 4\pi\rho(r)$$

(3)

where φ is the electric potential in a continuum of varying dielectric constant $\varepsilon(r)$, and $\rho(r)$ is the charge density.

For certain geometries, the Poisson equation can be solved analytically, but for the more general case of an arbitrarily-shaped solute surrounded by water, it can only be solved numerically. Once the electrostatic potential is calculated, the total electrostatic energy of the system is given by

$$G_{\text{elec}}=\frac{1}{2} \int_V \rho(r)\varphi(r)dV$$

(4)

For a set of point charges, the integration becomes a summation

$$G_{\text{elec}}=\frac{1}{2}\sum_i q_i\varphi(r_i)$$

(5)

Then, the electrostatic energy of solvation ($G_{\text{pol}}(PB)$) for transferring the solute from a homogeneous environment ($\varepsilon_o = \varepsilon_p$) to an aqueous environment ($\varepsilon_o = \varepsilon_w$) is given by the difference in G_{elec}:

$$G_{\text{pol}}(\text{PB}) \equiv \Delta G_{\text{solv,es}} = \frac{1}{2} \sum_i q_i (\varphi^{\varepsilon_o = \varepsilon_w}(r_i) - \varphi^{\varepsilon_o = \varepsilon_p}(r_i))$$

(6)

To derive the exact Born radius of atom i in a solute molecule, consider a molecule that has the same geometry as the solute but with no point charges except a +1 charge at atom i [13]. The electrostatic solvation energy of the model molecule ($G_{\text{pol,i}}$) is then related to the Born radius (α_i) as follows [28,29]:

$$G_{\text{pol,i}}(\text{PB}) = -\frac{1}{2} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \frac{1}{\alpha_i}$$

(7)

Therefore, $\alpha_i(\text{PB})$ can be calculated by

$$\alpha_i(\text{PB}) = -\frac{1}{2} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) \frac{1}{G_{\text{pol,i}}(\text{PB})}$$

(8)

Here, the notation $\alpha_i(PB)$ is used in order to distinguish it from the α_i calculated from approximate approaches such as GB. $\alpha_i(PB)$ is accurate and therefore also referred to as the perfect Born radius [13].

In the present work, DelPhi program [6] was used for the PB calculations. The VDW radii and charge assignments were the same as in the GB calculations. The dielectric constants for solute and environment were set to 4 and 80, respectively. The grid spacing was set to 0.33 Å. The grid number was set so that the system occupies 90% of the grid volume. The default values were used for all other control parameters in the program.

2.3 The HCT GB method for fast Born radii calculations

With the Coulomb field approximation, the Born radius of atom i (α_i) can be calculated as

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

(9)

where a_i is the radius of atom i and the integral is over the high dielectric media outside atom i. The integral is often replaced with the pairwise summation in fast GB methods, simplifying Eq. 9 to

$$\frac{1}{\alpha_i} \simeq \frac{1}{a_i} - \frac{1}{4\pi} \sum_j \int_j \frac{1}{r^4} dV = \frac{1}{a_i} - \sum_j \mathcal{H}(a_i, a_j, r_{ij})$$

(10)

where the summation is over each atom j that is not atom i. $\mathcal{H}$ is the analytical function derived by Hawkins, Cramer, and Truhlar [14], expressed as

$$\mathcal{H}(a_i, a_j, r_{ij}) = \frac{1}{2}\left[\left(\frac{1}{L_{ij}} - \frac{1}{U_{ij}}\right) + \left(\frac{a_j^2}{4r_{ij}} - \frac{r_{ij}}{4}\right)\left(\frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2}\right) + \frac{1}{2r_{ij}}\ln\frac{L_{ij}}{U_{ij}}\right]$$

(11)

where

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

(12)

and

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

(13)

To compensate for the problem that the contribution of atom j is over-counted if it intersects a third atom, the HCT GB model [14] introduced a scaling factor, S_j, to a_j in the $\mathcal{H}$ function:

$$\frac{1}{\alpha_i} \simeq \frac{1}{a_i} - \sum_j \mathcal{H}(a_i, S_j a_j, r_{ij})$$

(14)

The atom type-dependent scaling factors $\{S_j\}$ were parameterized to fit with PB calculations.

Eq. 14 is atom-based and additive. The contribution of each atom to α_i can be calculated individually. When the ligand and the receptor form a complex, Eq. 14 can be generalized and the Born radius of atom i after binding is given by

$$\frac{1}{\alpha_i} = \frac{1}{a_i} - \sum_j \mathcal{H}(a_i, S_j a_j, r_{ij}) - \sum_k \mathcal{H}(a_i, S_k a_k, r_{ik})$$

(15)

where the last term is the contribution from each atom k of the binding partner (e.g., k refers to ligand atoms if i is a protein atom and vice versa).

2.4 A variational form of the HCT GB method is useful in accurate Born radii calculations

Although computationally efficient, the HCT GB method does not provide satisfactory accuracy for Born radii calculations [29,30]. We therefore propose a variational form of the HCT method, denoted as the ΔHCT method, which provides an acceptable approximation of the change in inverse Born radii experienced by atoms during binding, as long as the change is small. In our ΔHCT method, the effect of the binding partner as a pure dielectric medium appears as the change in $1/\alpha$, which is

$$\Delta\frac{1}{\alpha_i} = \frac{1}{\alpha_i}(Bound) - \frac{1}{\alpha_i}(Free) = -\sum_k \mathcal{H}(a_i, S_k a_k, r_{ik})$$

(16)

Consequently, the effective Born radius of an atom in the bound state can be obtained by subtracting the contribution due to the binding process from its value in the free state:

$$\frac{1}{\alpha_i}(Bound) = \frac{1}{\alpha_i}(Free) - \sum_k \mathcal{H}(a_i, S_k a_k, r_{ik})$$

(17)

For situations in which the Born radii of atoms in the free state are already known, this method is more efficient. Unlike the original HCT method, the ΔHCT method only sums the contributions from atoms of binding partners. Additionally, if accurate free-state Born radii are available, the ΔHCT method can be much more accurate than the HCT method (see the Results section).

2.5 Validating GB models

To evaluate our GB calculations, we considered their ability to accurately reproduce several electrostatic quantities, using PB as a reference. The first of these was the electrostatic solvation energy of a single solute molecule, namely, the energy cost for moving the protein ($G_{pol, R}$) or the ligand ($G_{pol, L}$) from a homogeneous environment ($\varepsilon_o = \varepsilon_p$) to an aqueous environment ($\varepsilon_o = \varepsilon_w$). Previous studies have shown that $G_{pol, L}$ calculated with most GB models agrees well with the corresponding PB value (*e.g.*, [29]). Therefore, we focused on $G_{pol, R}$ in the present work.

Though the majority of work in developing Born radii models uses accurate solvation energy as the sole criterion for validation, the sufficiency of this criterion for GB models has been questioned [17,31,32]. Our previous systematic study showed that a model that successfully reproduces the solvation energy is not necessarily suitable for ligand binding calculations [29]. We therefore used additional quantities for evaluation. The second quantity used for evaluation was the electrostatic component of the binding energy, G_{POL}. G_{POL} was calculated from the following formula [29,33]:

$$G_{POL} = G_{pol,LR} - G_{pol,L} - G_{pol,R} + G_{Coulomb}$$

(18)

where $G_{pol, LR}$ is the electrostatic solvation energy of the ligand-receptor complex (LR). $G_{Coulomb}$ is the Coulomb interaction between the ligand and the receptor in the homogeneous environment ($\varepsilon_o = \varepsilon_p$). Detailed explanations are given in [29].

Lastly, we considered the ability of our GB calculations to reproduce the components of G_{POL}. G_{POL} can be decomposed by rewriting Eq. 18 as [19,29]

$$G_{POL} = G_{R\ desolv} + G_{L\ desolv} + G_{screened\ es}$$

(19)

where $G_{R\ desolv}$ and $G_{L\ desolv}$ are the partial desolvation energies of the receptor and the ligand, respectively. $G_{screened\ es}$ is the screened electrostatic interaction energy between the receptor and the ligand. Our previous study [29] suggested that $G_{R\ desolv}$ is the most challenging variable to match between GB and PB. Once $G_{R\ desolv}$ and G_{POL} of the GB calculations agreed with the corresponding PB results, the rest of the components in G_{POL} generally agreed well. We therefore focused only on $G_{R\ desolv}$ among these three components.

In summary, we evaluated the accuracy of our GB calculations by comparing their values for $G_{pol, R}$, G_{POL} and $G_{R\ desolv}$ with PB-derived values. Details of how to calculate these three

variables using both methods are described in [29]. The differences between our GB models' predicted values and PB values were averaged over the testing set using the root mean square deviation (*rmsd*):

$$rmsd = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$$

(20)

where y_i and x_i are the values of $G_{pol,R}$, G_{POL} or $G_{R\ desolv}$ calculated with the GB and PB approaches for the *i*-th protein-ligand complex, respectively. *N* is the total number of complexes in the testing set, which was 32.

2.6 Preparation of the test set

The test set consisted of the 32 protein-ligand complexes from our previous work [29], listed in Table 1. The coordinates of the complexes were obtained from the Protein Data Bank [34]. Water molecules were removed. The AMBER all-atom force field parameters were used in the present study [35]. The protein atoms were assigned with AMBER all-atom charges and the ligand atoms were assigned with Gasteiger charges, both using SYBYL software (Tripos, Inc.). The complexes being selected cover a wide variety of protein families and ligand types. Moreover, the electrostatic component of the binding energies (G_{POL}) calculated with the PB method spans a wide range, from −10 kcal/mol to +35 kcal/mol.

3 Results

3.1 Perfect GB method offers accurate binding electrostatics

As shown previously by Onufriev *et al.*, the perfect GB model (see Methods) gives solvation energies of macromolecules [13] which agree very well with values computed directly from the PB equation. It was unknown whether accurate Born radii would also allow the GB model to closely reproduce PB-derived ligand binding energies, which is a more stringent case. To address this, we computed perfect Born radii (see Eq. 8) for all the atoms in our 32 test complexes and calculated the electrostatic binding energies (G_{POL}), the receptor partial desolvation energies ($G_{R\ desolv}$), and the receptor full solvation energies ($G_{pol,R}$). The 32 complexes used are listed in Table 1.

Figure 1 compares G_{POL}, $G_{R\ desolv}$, and $G_{pol,R}$ calculated with this perfect-radii GB model versus the values calculated with the PB method. G_{POL} *rmsd* was 2.84 kcal/mol (panel a). $G_{R\ desolv}$ *rmsd* was 1.57 kcal/mol (panel b). The $G_{pol;R}$ values also agreed very well with the PB method (*rmsd* = 16.07 kcal/mol, panel c), which is consistent with the findings of Onufriev *et al.* [13]. We therefore conclude that, if the Born radii are calculated very accurately, the GB approach is adequate for studying the electrostatics of ligand binding.

3.2 The multiscale GB method improves efficiency

It is currently impractical to compute highly accurate Born radii for all of the atoms when doing large virtual database screening. Therefore, we sought to improve the tradeoff between accuracy and efficiency in the GB approximation by selecting only those atoms near the ligand for accurate radii calculation. Those atoms near the ligand we refer to as primary atoms; the distant atoms we call secondary atoms. In order to demonstrate this multiscale strategy, we computed the radii of primary atoms in both free and bound states perfectly (PGB method, Eq. 8) to avoid being restricted to a specific high-accuracy GB model and thus losing generality. In practice, one would not use the PGB method in database screening as it is more computationally expensive than the PB method and offers no accuracy advantage over it. For

secondary atoms, we used the very rapid HCT method [14]. For this PGB/HCT model (see Table 2), a cutoff radius of 5 Å provides satisfactory results with a G_{POL} *rmsd* value of 3.09 kcal/mol and $G_{R\ desolv}$ *rmsd* 4.02 kcal/mol. For this cutoff, the radii of only a small fraction of the atoms were derived from PB, 6.8% on average. For the sake of comparison, we also tested the GB model using HCT Born radii for all atoms. $G_{R\ desolv}$ *rmsd* was 5.32 kcal/mol and $G_{R\ desolv}$ *rmsd* was 7.72 kcal/mol (Table 2, last row). Accurately computing the Born radii of only a small percent of the atoms allows the GB model to produce acceptable binding energies and $G_{R\ desolv}$ values.

It was unknown whether or not the small Born radii changes experienced by secondary atoms affect the results enough to warrant computing the Born radii of bound state secondary atoms after having already computed the Born radii of free state secondary atoms. To investigate this, we used perfect Born radii for primary atoms as above, but we used the free state perfect Born radii in both the free and bound-state secondary atoms (denoted PGB/staticPGB in Table 2), essentially disregarding any change in the Born radii of secondary atoms. While not as accurate as the perfect GB model, this model was close with a G_{POL} *rmsd* of 3.11 kcal/mol versus PB-derived values, and a $G_{R\ desolv}$ *rmsd* of 1.97 kcal/mol versus PB-derived values. Accordingly, we modified the above PGB/HCT model so that the bound state secondary atoms reused their free state Born radii, instead of updating them. This PGB/staticHCT model (see Table 2) gave results which are nearly identical to PGB/HCT results in accuracy. The upper panels of Figure 2 plots the deviation between the results given by this model and PB-derived values in terms of *rmsd* for various choices of cutoff radius. Panel a gives the G_{POL} *rmsd* values and panel b gives the $G_{R\ desolv}$ *rmsd* values. It is notable that a short cutoff of 3 Å yielded a G_{POL} *rmsd* of 9.77 kcal/mol and a $G_{R\ desolv}$ *rmsd* of 7.57 kcal/mol, which were much worse than the *rmsd* values yielded from applying the crude HCT GB model to every atom (last row of Table 2). Despite that the Born radii were calculated more accurately for primary atoms than secondary atoms, the poor accuracy in binding energetic calculations suggests the importance of the cutoff distance in primary atom selection. An unpublished plot of PGB/HCT *rmsd* values for various cutoff distances is nearly identical to Figure 2. Again, the 5 Å cutoff offers an effective compromise with a $G_{R\ desolv}$ *rmsd* of 2.88 kcal/mol and $G_{R\ desolv}$ *rmsd* was 3.98 kcal/mol, insignificantly different from the PGB/HCT *rmsd* values.

In summary, accurately computing the Born radii of only a small proportion of the atoms in our test complexes gave adequate accuracy. The rapid HCT method was adequate for secondary atoms, and the Born radii change of secondary atoms could be ignored without introducing substantial inaccuracies. Consequently, when speed is desired, we suggest the following strategy: a high accuracy GB model (such as GBR6 [25,27] or GBMV [21,22]) is used for primary atoms in both the free and bound states, and a rapid method such as HCT is used for secondary atoms in the free state (ignoring the Born radii change of bound state secondary atoms).

3.3 The ΔHCT method is useful in the multiscale GB method

Some applications may require higher accuracy than was achieved using HCT for secondary atoms as above, so we also tested the multiscale GB method using a more accurate approach for secondary atoms. As before, the Born radii of primary atoms were computed using perfect GB, in order to generally represent high accuracy GB models. However this time, the Born radii of secondary atoms in the free state were computed using perfect GB, while those of bound state secondary atoms were computed using ΔHCT. Our ΔHCT is a modification of the original HCT formalism [14] which computes the change in effective Born radius in order to overlay these changes on the accurate Born radii computed for atoms in the free state (see Section 2.4). In screening a large database of ligands against one protein, the accurate Born radii computed for secondary atoms in the free state would be a one-time calculation to be

reused in each docking, making this approach computationally practical. See PGB/PGB/ΔHCT in Table 2 for the accuracy of this method with a cutoff of 5 Å.

The lower panels of Figure 2 plot the inaccuracy of the PGB/PGB/ΔHCT approach for various choices of cutoff distance in terms of *rmsd* versus PB-derived values. Panel c gives the G_{POL} *rmsd* values and the panel d gives the $G_{R\ desolv}$ *rmsd* values. In this case, a cutoff distance of 4.5 Å from the ligand for primary atoms offers a nice tradeoff between number of primary atoms and accuracy. For this cutoff, 5.7% of the atoms in a complex are primary, on average. The G_{POL} *rmsd* is 2.81 kcal/mol versus PB-derived values, comparable to the PGB/staticHCT model, while the $G_{R\ desolv}$ *rmsd* is 1.96 kcal/mol versus PB-derived values, a great improvement over the PGB/staticHCT model. As mentioned previously, $G_{R\ desolv}$ tends to be the most difficult component to match between GB and PB models, and accurate G_{POL} and $G_{R\ desolv}$ values tend to generalize to the accuracy of GB model as a whole [29]. One might expect, given that PGB/HCT and PGB/staticHCT were nearly identical in accuracy, that the change in the Born radius of secondary atoms can simply be disregarded. However, for every choice of cutoff distance tested, using ΔHCT to estimate this change offered at least a slight G_{POL} and $G_{R\ desolv}$ accuracy advantage over disregarding the change. Moreover, since the ΔHCT method is much faster than highly accurate GB models, PGB/PGB/ΔHCT is advantageous over PGB/staticPGB.

4 Discussion

GB models face a compromise between accuracy and efficiency. An important part of this compromise is in their method of approximating the effective Born radius of an atom. GB models using perfect Born radii give values for the electrostatic binding energy and its components which show minimal difference from those given by the Poisson-Boltzmann method, suggesting the performance of accurate GB models such as GBR6 [25,27] and GBMV [21,22] would be close to the performance of the PB method on ligand binding. Calculating perfect Born radii is highly computationally demanding (much slower than the PB method itself) and we therefore do not recommend it be used for primary atoms in practice. Instead, we suggest GBR6 or GBMV for primary atoms, since they have been shown to have very good accuracy.

Although we used HCT to calculate the Born radii of secondary atoms in this work, we expect any fast GB method to be suitable. For example, the GB method proposed by Onufriev *et al.* [36] is also a good choice since it is considerably more accurate than HCT while almost as fast. We also propose an alternative approach for dealing with secondary atoms. In this alternative, a high-accuracy GB model (such as GBR6 or GBMV) is used to compute the effective Born radii of secondary atoms, while in each bound state for the same protein, these radii are reused with an additional adjustment by ΔHCT (see Methods) to estimate the change in their effective Born radii during binding. In screening a large database of ligands against one protein, this approach is computationally practical since the free state Born radii may be computed only once and reused. Additionally, this second approach was found to be accurate, an advantage which is especially manifest in evaluating the difficult $G_{R\ desolv}$ component.

It would be ideal to also compare the results of the perfect-radii GB model or the multiscale GB approaches with experimental measurements. Unfortunately the measured binding affinities lump many complex interactions such as electrostatic interactions, van der Waals interactions, hydrophobic interactions and conformational entropic effect. Therefore, true values of G_{POL}, $G_{R\ desolv}$, and $G_{pol,R}$ are unavailable. However, we may use the error of the calculated binding energies with MM/PBSA relative to experimental affinities as a crude gauge. MM/PBSA combines PB with additional calculations of nonpolar contributions. The rigorous MM/PBSA calculations with careful protonation assignments done by Wittayanarakul

et al. on HIV-1 pro-tease inhibitors [37] showed an rmsd value of around 6 kcal/mol. Other MM/PBSA studies [38] had an rmsd of about 10 kcal/mol [37]. Thus, the rmsd error of the perfect-radii GB model (less than 3 kcal/mol) and errors of the multiscale GB approaches (about 3 kcal/mol) are acceptable.

It would be interesting to compare the computational efficiency with some of the existing accurate GB methods. The GBMV method in particular stands out as providing high accuracy at a moderate computational cost. The work of Feig *et al.* [5] found that single calculations of electrostatic solvation energies with GBMV require only about five times as much time as the HCT method. Notice that the computational time for GB methods is primarily spent on Born radii calculations and energy summations. The energy summation step is identical for different GB methods, which takes about half of the computational time for Born radii derivation with the HCT method according to our estimate. Therefore, the time for Born radii calculations with GBMV is about seven folds of the time for the HCT method. Based on these estimations, using the multiscale GB approach with GBMV for primary atoms and HCT for secondary atoms is about four to six times faster on Born radii calculations than using GBMV for all protein atoms without sacrificing much accuracy. The efficiency improvement would be much more prominent if GBR6 and other accurate GB models are used. For example, the rigorous numerical volume integration GB method [17,19] is about 50 times slower than the HCT method, and requires careful treatment of the low-dielectric crevices between protein atoms [36]. The efficiency improvement of the multiscale method over high accuracy GB methods is useful for large database screening. Future work in prioritizing the calculation of accurate Born radii in the multiscale GB approach may produce additional efficiency improvements.

Acknowledgments

References

1. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Curr Opin Chem Biol 2002;6:439–46. [PubMed: 12133718]

2. Brooijmans N, Kuntz ID. Annu Rev Biophys Biomol Struct 2003;32:335–73. [PubMed: 12574069]

3. Gilson MK, Zhou H. Annu Rev Biophys Biomol Struct 2007;36:21–42. [PubMed: 17201676]

4. Adcock SA, McCammon JA. Chem Rev 2006;106:1589–615. [PubMed: 16683746]

5. Feig M, Onufriev A, Lee MS, Im M, Case DA, Brooks CL III. J Comput Chem 2004;25:265–84. [PubMed: 14648625]

6. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B. J Comput Chem 2002;23:128–137. [PubMed: 11913378]

7. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Proc Natl Acad Sci USA 2001;98:10037–41. [PubMed: 11517324]

8. Grant JA, Pickup BT, Nicholls A. J Comput Chem 2001;22:608–40.

9. Totrov M, Abagyan R. Biopolyms Peptide Sci 2001;60:124–133.

10. Still WC, Tempczyk A, Hawley RC, Hendrickson T. J Am Chem Soc 1990;112:6127–9.

11. Bashford D, Case DA. Annu Rev Phys Chem 2000;51:129–52. [PubMed: 11031278]

12. Chen J, Brooks CL III, Khandogin J. Curr Opin Struct Biol 2008;18:140–8. [PubMed: 18304802]

13. Onufriev A, Case DA, Bashford D. J Comput Chem 2002;23:1297–304. [PubMed: 12214312]

14. Hawkins GD, Cramer CJ, Truhlar DG. Chem Phys Lett 1995;246:122–9.

15. Hawkins GD, Cramer CJ, Truhlar DG. J Phys Chem 1996;100:19824–39.

16. Qiu D, Shenkin PS, Hollinger FP, Still WC. J Phys Chem A 1997;101:3005–14.

17. Scarsi M, Apostolakis J, Caflisch A. J Phys Chem A 1997;101:8098–8106.

18. Ghosh A, Rapp CS, Friesner RA. J Phys Chem B 1998;102:10983–90.

19. Zou X, Sun Y, Kuntz ID. J Am Chem Soc 1999;121:8033–43.

20. Dominy BN, Brooks CL III. J Phys Chem B 1999;103:3765–73.

21. Lee MS, Feig M, Salsbury FR Jr, Brooks CL III. J Comput Chem 2003;24:1348–56. [PubMed: 12827676]

22. Lee MS, Salsbury FR Jr, Brooks CL III. J Chem Phys 2002;116:10606–14.

23. Im W, Lee MS, Brooks CL III. J Comput Chem 2003;24:1691–702. [PubMed: 12964188]

24. Grycuk T. J Chem Phys 2003;119:4817–26.

25. Tjong H, Zhou H. J Phys Chem B 2007;111:3055–61. [PubMed: 17309289]

26. Tjong H, Zhou H. J Chem Phys 2007;126:195102. [PubMed: 17523838]

27. Mongan J, Svrcek-Seiler WA, Onufriev A. J Chem Phys 2007;127:185101–10. [PubMed: 18020664]

28. Rankin KN, Sulea T, Purisima EO. J Comput Chem 2003;24:954–62. [PubMed: 12720316]

29. Liu H, Zou X. J Phys Chem B 2006;110:9304–13. [PubMed: 16671749]

30. Zhu J, Alexov E, Honig B. J Phys Chem B 2005;109:3008–22. [PubMed: 16851315]

31. Jayaram B, Liu Y, Beveridge DL. J Chem Phys 1998;109:1465–71.

32. Scarsi M, Caflisch A. J Comput Chem 1999;20:1533–6.

33. Liu HY, Kuntz ID, Zou X. J Phys Chem B 2004;108:5453–62.

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Research 2000;28:235–42. [PubMed: 10592235]

35. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz K, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc 1995;117:5179–97.

36. Onufriev A, Bashford D, Case DA. J Phys Chem B 2000;104:3712–20.

37. Wittayanarakul K, Hannongbua S, Feig M. J Comput Chem 2008;29:673–85. [PubMed: 17849388]

38. Wang W, Kollman PA. Proc Natl Acad Sci USA 2001;98:14937–42. [PubMed: 11752442]
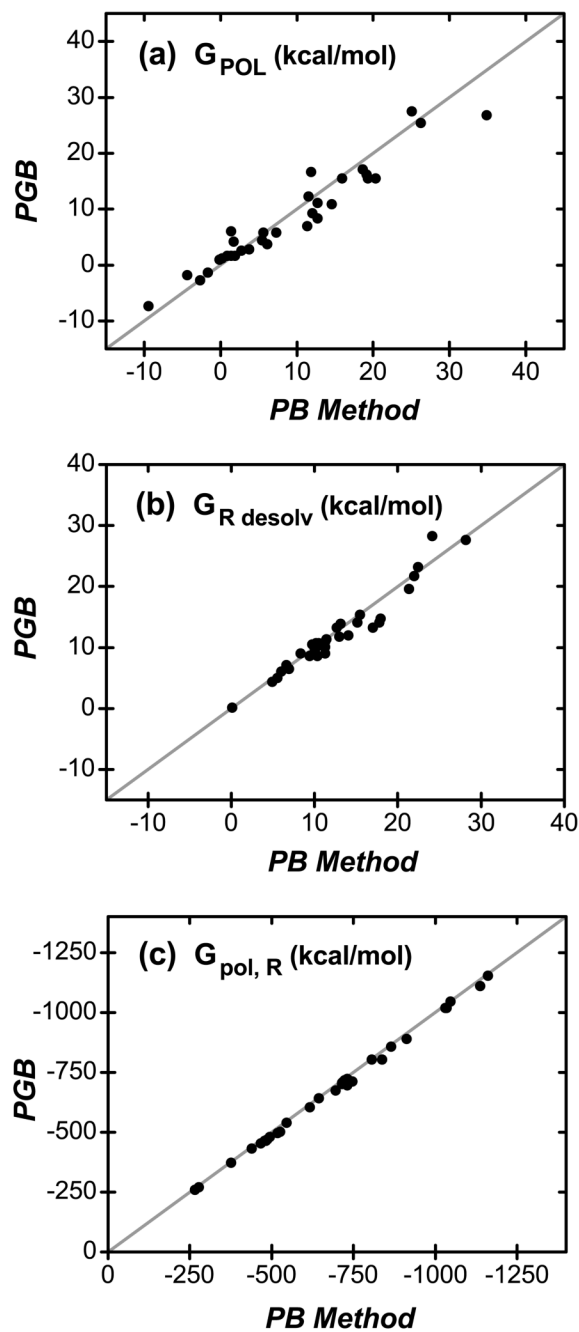
Figure 1.
Electrostatic quantities computed via the perfect GB method versus values derived directly from PB. Panel a compares the electrostatic binding energies given by these two methods; panel b, the receptor partial desolvation energy; panel c, the protein electrostatic solvation energy.
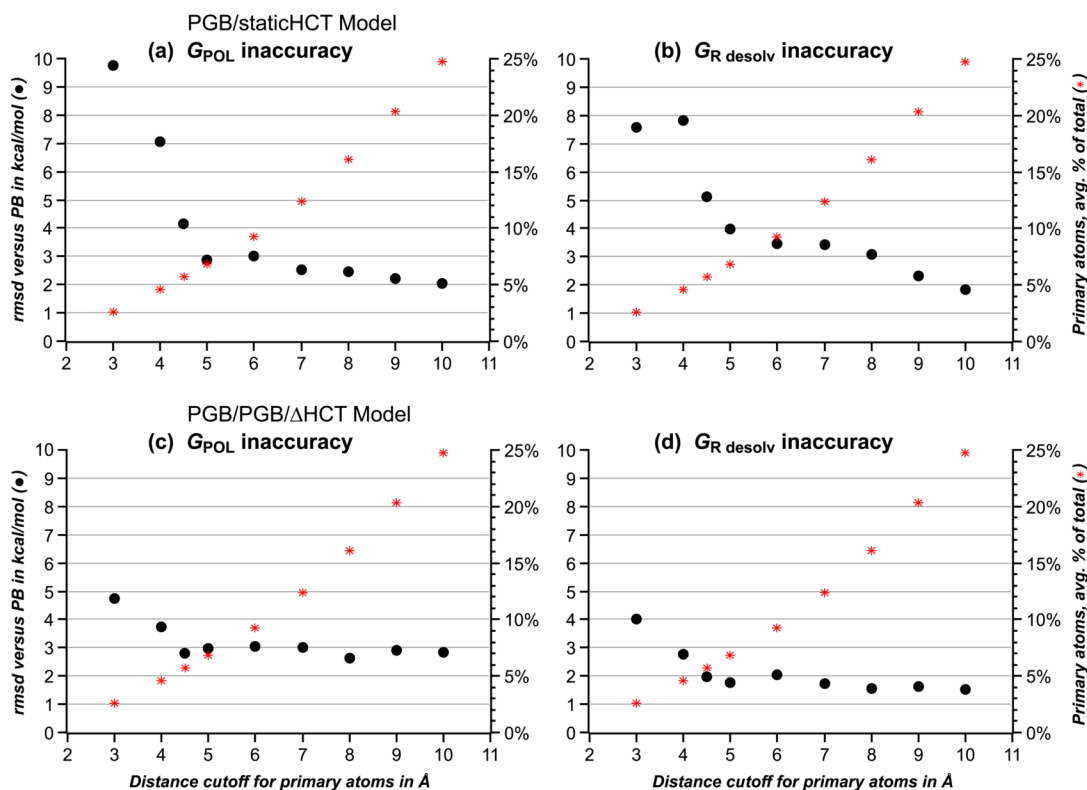
Figure 2.
Accuracy of suggested models versus cutoff distance for primary atoms. Black dots give the accuracy of the PGB/staticHCT model in reproducing PB-derived G_{POL} values (panel a) and $G_{R\ desolv}$ values (panel b) in terms of *rmsd*, also accuracy of the PGB/PGB/ΔHCT model in reproducing PB-derived G_{POL} values (panel c) and $G_{R\ desolv}$ values (panel d). The distance cutoff is the maximum distance an atom may be from the ligand before it is defined as a secondary atom rather than a primary atom. Red stars give the average percentage of complex atoms considered primary for each distance cutoff.

Table 1

Complexes studied in the present work

The testing set used in the present work for validating our GB calculations. The binding energies, G_{POL}, as computed using the PB method, are listed for each complex, as well as the resolution of each structure. See the text for detail.

Complex	G_{POL} by PB in kcal/mol	Resolution in Å	Protein/Ligand
3ptb-benz	−1.58	1.7	trypsin/benzamidine
4dfr-mtx	−4.46	1.7	dihydrofolate reductase/methotrexate
2ifb-PLM	1.70	2.0	fatty acid binding protein/C15COOH
6abp-ara	12.70	1.67	L-arabinose binding protein M108L/L-arabinose
1rnt-2gp	15.88	1.9	ribonuclease T1/2-GMP
1fkf-fk5	3.80	1.7	FK506 binding protein/FK506
1drf-fol	0.80	2.0	dihydrofolate reductase/folate
1hsl-his	−0.18	1.89	histidine binding protein/Histidine
1aaq-psi	18.60	2.5	HIV-1 protease/hydroxyethylene isostere
1ppc-nap	1.93	1.8	trypsin/NAPAP
1tng-amc	−2.64	1.8	trypsin/aminomethylcyclohexane
1pgp-6pg	17.85	2.5	6-PGDH/6-phosphogluconic acid
1pph-tap	1.27	1.9	trypsin/3-TAPAP
7abp-fca	14.47	1.67	l-arabinose binding protein M108L/D-fucose
2gbp-glc	12.04	2.9	galactose binding protein/galactose
1nsd-dan	20.34	1.80	neuraminidase/DANA
1rbp-rtl	5.33	2.00	retinol-binding protein/retinol
1ajv-nmb	12.65	2.0	HIV-1 protease/AHA006
1a9t-two	34.96	2.0	PNP/9-deazainosine
1dwd-mid	7.36	3.0	thrombin/NAPAP
1ets-pap	11.61	2.3	thrombin/NAPAP
1abe-ara	11.35	1.7	L-arabinose binding protein/L-arabinose
1mdq-mal	19.08	1.9	maltose binding protein A301GS/maltose
1l83-bnz	0.12	1.7	lysozyme/benzene
4ts1-tyr	5.66	2.5	tyrosyl-transfer RNA synthetase/tyrosine
2mcp-PC	−0.12	3.1	immunoglobulin/phosphocholine
1ulb-gun	2.75	2.75	PNP/guanine
3fx2-fmn	26.24	1.9	flavodoxin/FMN
8atc-pal	1.36	2.5	aspartate carbamoyltransferase/PALA
1nnb-dan	19.32	2.80	neuraminidase/DANA
2pk4-aca	−9.47	2.25	plasminogen kringle 4/aminocaproic acid
7tim-pgh	6.12	1.9	trisephosphate isomerase/phosphoglycolohydroxamate

Table 2

Summary of selected models

Accuracy of selected GB calculations in reproducing PB derived values for G_{POL} and $G_{R\,desolv}$. The last column lists, for each calculation, the methods used to determine Born radii for primary and secondary atoms. If the method used to determine Born radii for secondary atoms in the free state differs from the method used to compute their Born radii in the bound state, both cases are listed.

Model (cutoff distance)	G_{POL} *rmsd* in kcal/mol	$G_{R\,desolv}$ *rmsd* in kcal/mol	Born radii method
PGB	2.84	1.57	PB for all atoms
PGB/PGB/ΔHCT (5 Å)	2.96	1.75	PB for primary atoms, PB for secondary atoms in free state, ΔHCT for bound state secondary atoms
PGB/staticPGB (5 Å)	3.11	1.97	PB for primary atoms, PB for secondary atoms in free state, secondary atom Born radii change ignored
PGB/staticHCT (5 Å)	2.88	3.98	PB for primary atoms, HCT for secondary atoms in free state, secondary atom Born radii change ignored
PGB/HCT (5 Å)	3.09	4.02	PB for primary atoms, HCT for secondary atoms
HCT	5.32	7.72	HCT for all atoms