GBr^6: A Parameterization-Free, Accurate, Analytical Generalized Born Method

Harianto Tjong and Huan-Xiang Zhou*

Department of Physics and Institute of Molecular Biophysics and School of Computational Science, Florida State University, Tallahassee, Florida 32306

Received: September 25, 2006; In Final Form: January 6, 2007

The Poisson−Boltzmann (PB) equation is widely used for modeling electrostatic effects and solvation for macromolecules. The generalized Born (GB) model has been developed to mimic PB results at substantial lower computational cost. Here, we report an analytical GB method that reproduces PB results with high accuracy. The analytical approach builds on previous work of Gallicchio and Levy (*J. Comput. Chem.* **2004**, *25*, 479), and incorporates an improvement, proposed by Grycuk (*J. Chem. Phys.* **2003**, *119*, 4817), of the Coulomb-field approximation used in most GB methods. Tested against PB results, our GB method has an average unsigned relative error of only 0.6% for a representative set of 55 proteins and of 0.4% and 0.3%, respectively, for folded and unfolded conformations of cytochrome *b*562 sampled in molecular dynamics simulations. The dependencies of the electrostatic solvation free energy on solute and solvent dielectric constants and on salt concentration are fully accounted for in our method.

Introduction

The treatment of solvation effect is a fundamental problem in biomolecular modeling. One may choose to include explicit water molecules in molecular dynamics (MD) or Monte Carlo (MC) simulations of solute molecules. Such simulations can provide a realistic description of energetic effects.[1−3] However, due to the large number of degrees of freedom, explicit solvent simulations require considerable computer resources to obtain meaningful results, especially due to the extensive equilibration and averaging required for the solvation energy of the solute. On the other hand, implicit solvation methods, which typically approximate the solvent as a continuum, are much more efficient computationally and may provide a more intuitive understanding. In general, the solvation free energy in implicit solvent models is separated into a nonpolar term, ΔG_{np}, and an electrostatic term, ΔG_{elec}. The latter accounts both for the direct interactions of the solute charges with the solvent and for the screening of Coulomb interactions among the solute charges. In this paper, we present an accurate, analytical method for calculating ΔG_{elec}.

Our method is based on modeling the solvent as a uniform high-dielectric that responds to the partial charges of a low-dielectric solute. The standard approach to such solvation models is numerical solution of the Poisson−Boltzmann (PB) equation,[4−14]

$$\nabla\cdot\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) - 4\pi\sum_i ez_ic_{i0}\exp[-ez_i\phi(\mathbf{r})/k_BT] \quad (1)$$

where ϵ is the dielectric constant as a function of the position vector $\mathbf{r}$, ρ is the solute charge density, and ϕ is the electrostatic potential, all of which are functions of the position vector $\mathbf{r}$. The second term on the right-hand side arises from the Boltzmann distribution of mobile ions in the solvent (k_BT: thermal energy); ions of species i have valency z_i and bulk concentration c_{i0}. Upon linearization, the PB equation becomes

$$\nabla\cdot\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \epsilon_s\kappa^2\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad (2)$$

where $\kappa^2 = 8\pi e^2I/\epsilon_sk_BT$, with I, the ionic strength; and ϵ_s, the solvent dielectric constant. Various numerical methods for solving the PB equation have been developed. UHBD,[7] DelPhi,[4−6,10] and APBS,[9] for instance, are based on finite difference methods and have been generally accepted as benchmarks for the calculation of the electrostatic solvation free energy.[15,16] Highly accurate results can be achieved by focusing on fine grids, with consequent increase in computational cost in terms of both CPU time and memory. The CPU time required per energy evaluation often far exceeds that of explicit solvent simulations.

One way to circumvent the demand of the PB model on computing time is to introduce simpler models. Such models are especially desired if one wants to do MD or MC simulations in implicit solvent, where fast energy and force calculations are a prerequisite. A promising alternative implicit solvent model that has now gained wide attention is the generalized-Born (GB) formalism,[17−29] inspired by the Born formula[30] for the solvation energy of ions:

$$\Delta G_{elec} = -(1/\epsilon_i - 1/\epsilon_s)Q^2/2R \quad (3)$$

This result can be derived from the PB equation for a spherical solute (with radius R and dielectric constant ϵ_i) having all the charge (Q) located at the center. Because of their simplicity, GB methods in general are faster than both PB and explicit water simulation methods. In the past few years, GB methods have been improved to have good agreement with PB results[15,16] as well as with MD simulations results,[31,32] especially in the absence of salts.

The method that we report here implements the GB model, with several well-known weaknesses in previous methods

* Corresponding author. Phone, (850) 645-1336. Fax, (850)644-7244. E-mail: zhou@sb.fsu.edu.

10.1021/jp066284c CCC: $37.00
Published on Web 02/20/2007

resolved. First, most existing GB methods are based on the so-called Coulomb-field approximation, which may lead to as much as 100% errors in the solvation energy.[33−35] To address this serious problem, Grycuk[35] proposed an alternative approximation that was shown to have higher accuracy for spherical solutes. We implemented Grycuk's formulation and show that it significantly improves accuracy for proteins.

Second, the GB model does not have additional dependence on solute and solvent dielectric constants beyond the prefactor $(1/\epsilon_i - 1/\epsilon_s)$ present in the Born formula. Exact results[36] for spherical solutes with off-center charges show that the solvation energy, in general, has far more complex dependences on solute and solvent dielectric constants. Recently, we found a simple formula that accurately predicts the PB solvation energies for all combinations of ϵ_i and ϵ_s, from the PB calculation on a single set of ϵ_i and ϵ_s values.[37] Our GB method incorporates this formula, which allows the solvation energy to be accurately obtained from a single GB calculation, for a wide range of solute and solvent dielectric constants.

Third, though salt effects have been considered in some studies,[23,38,39] this issue does not appear to have been systematically addressed. Our GB method fully and accurately accounts for salt effects.

Theory and Implementation

The "r^6" Approximation. The formulation of the GB model is well-known.[17,34] To motivate the introduction of Grycuk's improved approximation[35] in our GB method, we briefly outline the formulation.

In the GB model, the electrostatic contribution to the free energy of solvation is given by

$$\Delta G^0_{GB} = -(1/\epsilon_i - 1/\epsilon_s)\sum_{i,j} q_i q_j / 2 f_{ij} \tag{4}$$

where f_{ij} is a function of the distance, r_{ij}, between the charges q_i and q_j on atoms i and j. The form originally proposed and most widely used is[17]

$$f_{ij} = [r_{ij}^2 + B_i B_j \exp(-r_{ij}^2/4B_iB_j)]^{1/2} \tag{5}$$

By design, eq 5 gives $f_{ii} = B_i$. The self-energy of atom i is given by the Born formula, with B_i playing the role of the solute radius. B_i is, hence, referred to as the Born radius. The Born radii have to be optimized such that ΔG^0_{GB} agrees with the PB counterpart, ΔG_{PB}, as much as possible.

The conventional approach to determining the Born radii is based on the Coulomb-field approximation, which assumes that the electric displacement is Coulombic in form,[34] leading to

$$\frac{1}{B_i} = \frac{1}{4\pi}\int_{solvent} \frac{d^3\mathbf{r}}{(\mathbf{r} - \mathbf{r}_i)^4} \tag{6}$$

In the above expression, $\mathbf{r}_i$ is the location of the ith atom, and the integration covers the (infinite) solvent dielectric. Grycuk's formulation[35] replaces eq 6 with

$$\frac{1}{B_i^3} = \frac{3}{4\pi}\int_{solvent} \frac{d^3\mathbf{r}}{(\mathbf{r} - \mathbf{r}_i)^6} \tag{7}$$

Equation 7 will be referred to as the "r^6" approximation, which also explains why our method is called GBr^6.

Analytical Implementation. We adapted the analytical GB method of Gallicchio and Levy[27] to implement GBr^6. The solvent is the whole space without the solute, which is taken to be the union of the atomic van der Waals (vdW) spheres of the solute molecule. Equation 7 can then be transformed into

$$\frac{1}{B_i^3} = \frac{1}{a_i^3} - \frac{3}{4\pi}\int_{\substack{solute \\ r'>a_i}} \frac{d^3\mathbf{r}'}{r'^6} \tag{8}$$

where a_i is the vdW radius of atom i and $\mathbf{r}' = \mathbf{r} - \mathbf{r}_i$. If no other solute atoms were present, one would have $B_i = a_i$. The Born radius of an atom thus reflects the degree of its burial inside the solute molecule. Less burial leads to a smaller Born radius and a greater contribution to the solvation energy. Physically, more exposed atoms induce stronger polarization fields in the solvent.

To a zeroth order approximation, the volume integral of eq 8 can be broken into contributions of individual atoms. The contribution of atom j ($\neq i$) is the integral over the region of its vdW sphere, which lies outside atom i,

$$Z_{ji} = \frac{3}{4\pi}\int_{\substack{r'>a_i \\ r''<a_j}} \frac{d^3\mathbf{r}'}{r'^6} \tag{9}$$

where $\mathbf{r}'' = \mathbf{r} - \mathbf{r}_j$. The correction to the zeroth order approximation will be discussed shortly. First, we give the results for Z_{ij} in four possible situations:

1. Atoms j and i do not intersect; i.e., interatomic distance $r_{ij} > a_i + a_j$.

$$Z_{ji} = \frac{a_j^3}{(r_{ij}^2 - a_j^2)^3} \tag{10}$$

2. Atoms j and i intersect, but neither is completely inside the other.

$$Z_{ji} = \frac{1}{16 r_{ij}}\left[-6\left(\frac{1}{a_i^2} - \frac{1}{(r_{ij}+a_j)^2}\right) + 8 r_{ij}\left(\frac{1}{a_i^3} - \frac{1}{(r_{ij}+a_j)^3}\right) - 3(r_{ij}^2 - a_j^2)\left(\frac{1}{a_i^4} - \frac{1}{(r_{ij}+a_j)^4}\right)\right] \tag{11}$$

3. Atom i is completely inside atom j.

$$Z_{ji} = \frac{1}{a_i^3} - \frac{a_j^3}{(a_j^2 - r_{ij}^2)^3} \tag{12}$$

4. Atom j is completely inside atom i. In this case, atom j does not contribute, and $Z_{ji} = 0$.

Corresponding results for the Coulomb-field approximation (eq 6) were given by Gallicchio and Levy.[27] We note that the latter results also involve logarithmic functions in addition to polynomials. As we will see in the next section, the logarithmic functions lead to ~10% of additional CPU time.

Z_{ji} overcounts the contribution of atom j if its region outside atom i intersects a third atom. To overcome this problem, Gallicchio and Levy[27] cleverly introduced a scaling coefficient, s_{ji}. This was calculated from the self-volume, that is, the portion of the vdW sphere that belongs exclusively to atom j as a fraction of its vdW volume. The volume of overlapping spheres according to the Poincaré inclusion−exclusion principle is[40]

$$V = \sum_j V_j - \sum_{k>j} V_{jk} + \sum_{l>k>j} V_{jkl} - \ldots \tag{13}$$

where V_j is the vdW volume of atom j; V_{jk} is the intersection volume of atoms j and k; V_{jkl} is the intersection volume of atoms j, k, and l; and so forth. The self-volume V_j' is found as

$$V_j' = V_j - \frac{1}{2}\sum_k V_{jk} + \frac{1}{3}\sum_{l>k} V_{jkl} - ... \quad (14)$$

Note that atom i is not excluded in calculating V_j'. In calculating s_{ji}, the reduction of the self-volume of atom j by atom i is added back, leading to

$$s_{ji} = \frac{V_j' + \frac{1}{2}V_{ji}}{V_j} \quad (15)$$

This formula gives the correct result, $s_{ji} = 1$, when no other atoms intersect atom j. Following Grant and Pickup,[41] Gallicchio and Levy approximated intersection volumes by Gaussian integrals. This approximation works well for intersections among heavy atoms. Hydrogens are deeply buried into the attached heavy atoms, and as such, the approximation by Gaussian integrals incurs significant errors. Gallicchio and Levy designed a simple remedy by totally neglecting contributions of hydrogens to intersection volumes. The final result for the Born radius is given by

$$\frac{1}{B_i} = \left(\frac{1}{a_i^3} - \sum_{j\neq i} s_{ji}Z_{ji}\right)^{1/3} \quad (16)$$

For the sake of comparison, we also coded the version of Gallicchio and Levy's GB method, which they called AGBNP. Our coded version will be referred to as GBr^4/AGBNP.

Dependence on Solute and Solvent Dielectric Constants. To model the dependence of solvation free energy on dielectric constants beyond the prefactor $(1/\epsilon_i - 1/\epsilon_s)$ in eq 4, we incorporate in GBr^6 a simple formula found previously for PB results:[37]

$$\frac{\Delta G_{PB}(\epsilon_i, \epsilon_s)}{1/\epsilon_i - 1/\epsilon_s} = \frac{\Delta G_{PB}(\epsilon_{ir}, \epsilon_{sr})}{1/\epsilon_{ir} - 1/\epsilon_{sr}} f(\epsilon_i/\epsilon_s) \quad (17)$$

The factor $f(\epsilon_i/\epsilon_s)$ scales the solvation energy, $\Delta G_{PB}(\epsilon_{ir}, \epsilon_{sr})$, calculated at a reference set of solute and solvent dielectric constants (ϵ_{ir} and ϵ_{sr}, respectively) into the corresponding quantity at any desired combination of dielectric constants. It has the form

$$f(\epsilon_i/\epsilon_s) = \frac{A + 2B\epsilon_i/\epsilon_s}{1 + 2\epsilon_i/\epsilon_s} \quad (18)$$

With $\epsilon_{ir} = 2$ and ϵ_{sr}, $= 78.5$, the coefficients A and B were given by

$$A = -1.63 \times 10^{-3}|Q|^{0.65} + 2.18 \times 10^{-6}N_{atom} + 1.016 \quad (19)$$

$$B = 3.31 \times 10^{-2}|Q|^{0.65} - 4.77 \times 10^{-5}N_{atom} + 0.683 \quad (20)$$

where Q and N_{atom} are the net charge and number of atoms, respectively, of the solute.

This scaling factor was introduced into GBr^6, leading to

$$\Delta G_{GB} = \Delta G_{GB}^0 f(\epsilon_i/\epsilon_s) \quad (21)$$

as the final result for the solvation free energy for our GB method.

Salt Effects. Following the work of Srinivasan et al.,[23] we accounted for salt effects by modifying eq 4 into

$$\Delta G_{GB}^0 = -\sum_{i,j}[1/\epsilon_i - \exp(-\alpha\kappa f_{ij})/\epsilon_s]q_iq_j/2f_{ij} \quad (22)$$

The scaling parameter α was introduced to accommodate the ion exclusion radius in the PB model. Its value was set to 0.73 by Srinivasan et al. Our method departed from that work in two respects. First, we applied the dielectric scaling of eq 21. Second, we treated α as a floating parameter, instead of a constant, that was optimized for individual ionic strengths.

Test Systems and PB Solutions. Extensive tests of GBr^6 against the PB benchmark were carried on a set of 55 proteins. These proteins, listed in Table 1, were collected from the Protein Data Bank (http://www.rcsb.org/pdb) using the following criteria: sequence identity <10%, resolution better than 1.0 Å, and number of residues <250. The net charges and total number of atoms are also listed in Table 1. For PDB structures without hydrogen atoms, hydrogen atoms were added with the LEAP module in the AMBER package,[42] then the energy was minimized in vacuum with heavy atoms fixed.

To explore the potential use of GBr^6 in MD simulations, we also collected conformations from explicit solvent MD simulations of cytochrome *b*562 (PDB entry 1yyj).[43] One thousand folded conformations were collected from a 10.5-ns trajectory at 300 K, and 500 unfolded conformations were collected from a 19-ns trajectory at 500 K. The modified parm99 of AMBER[44,45] were used for the MD simulations. There was no particular reason for selecting this protein, except that we had the trajectories available.

PB results were obtained by using the UHBD program.[7] Calculations were carried out with the solute boundary specified either as the vdW surface or as the molecular surface (MS). The latter was done by adding the "nmap 1.4, nsph 500" option in the UHBD input file. GBr^6 was designed for the vdW surface, and a series of studies have shown that PB calculations using the vdW surface gives much better agreement with experimental data than using the MS.[46−50] The difference between the two surfaces lies in the numerous small crevices that cannot be accessed by a 1.4-Å spherical probe. These crevices are assigned to be part of the solvent dielectric in the vdW specification but part of the solute dielectric in the MS specification. Given that protein structures are dynamic and water can penetrate into the protein interior,[51] the vdW specification would also appear more physical. However, use of the MS is still quite popular in PB calculations, and a number of GB methods were tailored to mimic PB results obtained using the MS. Hence, we obtained PB results using the MS for the sake of benchmarking such GB methods. All UHBD calculations used coarse grid with a 1.5-Å spacing, followed by a fine grid with a 0.5-Å spacing. The dimensions of the coarse and fine grids were 160 × 160 × 160 and 200 × 200 × 200 for the 55 PDB structures, 100 × 100 × 100 and 140 × 140 × 140 for folded cytochrome *b*562 conformations, and 200 × 200 × 200 and 240 × 240 × 240 for unfolded cytochrome *b*562 conformations. In all PB and GB calculations, the temperature was 300 K.

Protein atoms were assigned charges of the AMBER force field.[52] For most calculations, the atoms were assigned the Bondi

TABLE 1: List of 55 Proteins and Their Properties

PDB	N_{atom}	Q	ΔG_{PB}[a]	ΔG^0_{GB}[a]	PDB	N_{atom}	Q	ΔG_{PB}[a]	ΔG^0_{GB}[a]	PDB	N_{atom}	Q	ΔG_{PB}[a]	ΔG^0_{GB}[a]
1a6m	2435	2	−1327.5	−1339.4	1k4i	3253	−6	−1898.0	−1909.4	1unq	1966	−3	−1667.8	−1666.8
1aho	967	−2	−634.8	−628.4	1kth	894	0	−710.8	−708.1	1vb0	921	3	−541.7	−543.4
1byi	3383	−4	−1746.7	−1748.9	1l9l	1230	11	−1514.5	−1526.9	1vbw	1058	8	−885.0	−896.7
1c75	987	−4	−685.7	−685.0	1m1q	1265	−7	−1166.1	−1159.4	1w0n	1756	−5	−1178.5	−1174.6
1c7k	1929	−5	−1190.0	−1191.8	1nls	3564	−7	−2274.3	−2290.1	1wy3	560	1	−376.6	−379.2
1cex	2867	1	−1404.8	−1415.2	1nwz	1912	−6	−1334.7	−1330.9	1x6z	1741	0	−1052.8	−1056.0
1eb6	2572	−15	−2493.9	−2491.0	1od3	1900	−3	−987.7	−984.9	1x8q	2815	−1	−1729.9	−1733.7
1ejg	678	0	−279.5	−278.5	1ok0	1076	−5	−756.1	−750.5	1xmk	1268	1	−777.0	−792.1
1etl	145	0	−140.2	−140.1	1p9g	529	4	−364.5	−366.0	1yk4	770	−8	−920.3	−911.7
1exr	2240	−25	−4535.9	−4529.4	1pq7	3065	4	−1254.5	−1249.1	1zzk	1252	1	−778.6	−787.8
1f94	982	1	−588.7	−588.0	1r6j	1230	0	−653.9	−656.1	2a6z	3432	−3	−1776.0	−1790.5
1f9y	2535	−5	−1424.5	−1420.1	1ssx	2750	8	−1282.0	−1292.7	2bf9	560	−2	−447.4	−441.7
1g4i	1842	−1	−1171.8	−1174.9	1tg0	1029	−12	−1567.5	−1554.5	2chh	1624	−3	−1037.4	−1039.9
1g66	2794	−2	−1429.4	−1426.7	1tqg	1660	−7	−1423.1	−1425.3	2cws	3400	−3	−1562.7	−1579.0
1gqv	2143	7	−1251.4	−1251.3	1tt8	2676	1	−1271.5	−1269.6	2erl	573	−6	−573.1	−562.7
1hje	179	1	−134.8	−135.6	1u2h	1526	4	−997.1	−1004.9	2fdn	731	−8	−834.7	−827.0
1iqz	1171	−17	−2262.2	−2246.9	1ucs	997	0	−499.3	−501.6	2fwh	1830	−6	−1101.9	−1102.5
1iua	1207	−1	−629.6	−633.3	1ufy	1926	−3	−1120.9	−1125.3	3lzt	1960	8	−1264.7	−1271.8
1j0p	1597	8	−1376.8	−1379.0										

[a] Calculated, in kcal/mol, for $\epsilon_i = 2$, $\epsilon_s = 78.5$, and $I = 0$, with the solute boundary defined as the vdW surface.

radii: C, 1.7 Å; N, 1.55 Å; O, 1.5 Å; S, 1.8 Å; and H, 1.2 Å.[53] We also specifically tested GBr^6 for different atomic radii, with the radius of each type of atoms decreasing or increasing by up to 0.3 Å.

Accuracy of GB methods in the absence of salts was measured by unsigned relative error in the solvation free energy, ΔG_{GB}, in comparison to the corresponding PB result, ΔG_{PB}. This was calculated as $<|(\Delta G_{GB} - \Delta G_{PB})/\Delta G_{PB}|>$, averaging over a set of proteins or a set of conformations. In addition, systematic differences from PB were checked by the signed relative error, $<(\Delta G_{GB} - \Delta G_{PB})/\Delta G_{PB}>$; worst cases were identified by the maximum unsigned relative error within a set of proteins or conformations.

These relative errors do not provide a sensitive measure of the quality of the salt effects calculated by a GB method, since the salt-induced change in solvation free energy is overwhelmed by the solvation free energy at zero salt. Therefore, we decided to report the root-mean-square-deviation (rmsd) of the GB salt effects from the PB counterparts for a set of proteins; here, salt effect specifically refers to the salt-induced change in solvation free energy. As a reference, we also report the average PB salt effect.

Results and Discussion

Performance of GBr^6 over a Wide Range of Dielectric Constants. As shown in Table 1, the 55 proteins from the PDB cover a wide range of total charge, from −25 to +11. Among them, 15 have positive net charges, 30 have negative net charges, and 7 are neutral. They also traverse a wide range of sizes (from ∼150 to ∼3600 total atoms) and a variety of structures and, hence, seem to serve well as a test set.

The ΔG^0_{GB} results obtained from GBr^6 and the corresponding ΔG_{PB} results obtained by UHBD for $\epsilon_i = 2$ and $\epsilon_s = 78.5$ are listed in Table 1. Very good agreement can be seen. Overall, the unsigned relative error is only 0.5%. There is no systematic difference from the PB benchmark, as the signed relative error is only 0.1%. The maximum error is 1.9% (for PDB 1xmk).

In most applications of the PB equation to proteins, the solute dielectric constant is set between 1 and 4. When ϵ_i is changed from 2 to 1, 3, and 4, the differences between ΔG^0_{GB} and ΔG_{PB} are still small, with unsigned relative errors at 0.9, 1.1, and 1.8%, respectively. However, the differences are systematic, with ΔG^0_{GB} consistently underestimating the magnitude of ΔG_{PB} at the lower value of ϵ_i while overestimation occurring at the higher values of ϵ_i.

TABLE 2: Performance of GBr^6 on the Set of 55 Proteins

(ϵ_i, ϵ_s)	unsigned relative error, %	signed relative error, %	maximum error, %
(1, 78.5)	0.6	−0.1	1.8
(2, 78.5)	0.5	0.1	1.9
(3, 78.5)	0.6	0.2	2.0
(4, 78.5)	0.6	0.2	2.0
(1, 10.3)	0.8	0.3	2.4
(2, 10.3)	1.1	0.1	3.7
(3, 10.3)	1.5	0.03	5.0
(4, 10.3)	1.8	0.2	6.2

The introduction of the dielectric constant dependent scaling factor, via eq 21, resolves the systematic errors. The resulting solvation free energy, ΔG_{GB}, has unsigned relative errors reduced to 0.5% or 0.6% for all four values of ϵ_i (see Table 2).

Most GB methods have been concerned with water as the solvent. It is desirable to extend the application of the GB model to other solvents, e.g., for calculating the free energy of transfer between two different solvents. The dielectric-constant-dependent scaling factor was found to be able to predict the PB solvation energy for a solvent with a much lower dielectric constant.[37] We therefore tested the accuracy of eq 21 in reproducing ΔG_{PB} for *n*-octanol, with $\epsilon_s = 10.3$. Even for the much lower solvent dielectric constant, the unsigned relative errors of GBr^6 are still very small, ranging from 0.8 to 1.8% for $\epsilon_i = 1$, 2, 3, and 4. There does not appear to be any systematic errors (Table 2).

Solvation free energies for 1000 folded and 500 unfolded conformations of cytochrome *b*562 calculated by GBr^6 are also in good agreement with the corresponding PB results. With $\epsilon_i = 1$ and $\epsilon_s = 78.5$, the unsigned errors of ΔG_{GB} relative to ΔG_{PB} are only 0.4% and 0.3%, respectively, for the folded and unfolded conformations. The unfolded conformations expectedly have many large variations in size, with the radius of gyration ranging from 12.3 to 21.2 Å (compared with a range of 15.0−17.6 Å for the folded conformations). The expanded conformations necessitated the use of a significantly larger grid in the UHBD calculations (fine grids with dimensions of 240 × 240 × 240 for unfolded conformations versus 140 × 140 × 140

TABLE 3: Performance of Different GB Methods on the Set of 55 Proteins

method	solute surface	PB target	total CPU time (s)[a]	relative errors (%)[b] $\epsilon_i = 1$	relative errors (%)[b] $\epsilon_i = 4$
GBr^6	vdW	vdW	37.0	0.6; −0.1; 1.8	0.6; 0.2; 2.0
GBr^4/AGBNP	vdW	vdW	41.9	25.8;−25.8; 39.2	23.8;−23.8; 37.2
GBr^4/AGBNP	vdW	MS	41.9	2.8; 1.8; 18.3	4.3; 3.9; 22.3
GB^{OBC1}	vdW	MS	44.3	9.8; 9.8; 24.5	12.1; 12.1; 28.7
GB^{OBC2}	vdW	MS	44.7	4.8; 2.2; 17.0	5.9; 4.4; 21.0
ALPB1	vdW	MS	44.6[c]	9.1; 9.1; 23.7	9.4; 9.4; 25.4
ALPB2	vdW	MS	44.3[c]	4.6; 1.6; 16.3	4.8; 1.9; 17.9
GMBV old	MS	MS	68.2	1.3; 1.1; 7.2	3.2; 3.2; 10.8
GBMV new	MS	MS	67.9	2.1; 2.1; 8.4	4.2; 4.2; 12.1

[a] All calculations were performed on a 3.06 GHz Intel Xeon dual processor Linux computer. [b] Unsigned, signed, and maximum relative errors are listed for each method at either of the two solute dielectric constants. [c] Not included is the CPU time, totaling 4 s, required for precalculation with the *elsize* program.

for folded conformations). The increase in grid dimension resulted in a 10-fold increase in CPU time. In contrast, the CPU time of GBr^6 was virtually the same for the folded and unfolded conformations.

Evaluation against Other GB Methods. To assess the performance of GBr^6 against previous developments, we also carried out calculations using several well-known GB methods on the same test set of 55 proteins with the same atomic charge and radius assignments. They are GBr^4/AGBNP (our coded version of the method of Gallicchio and Levy[27]), GB^{OBC1} and GB^{OBC2} of Onufriev et al.,[28] GBMV2 of Brooks and co-workers,[15,25] and ALPB of Sigalov et al.[39]

Calculations with GB^{OBC1} and GB^{OBC2} were done by running AMBER 8[42] with flag *igb* set to 2 and 5, respectively. The only difference of these two methods lies in the parametrization of the Coulomb-field Born radius (eq 6),[28]

$$\frac{1}{B_i} = \frac{1}{a_i - 0.09} - \frac{1}{a_i}\tanh(\alpha\Psi - \beta\Psi^2 + \gamma\Psi^3) \quad (23)$$

with $\alpha = 0.8$, $\beta = 0$, and $\gamma = 2.91$ for GB^{OBC1} and $\alpha = 1.0$, $\beta = 0.8$, and $\gamma = 4.85$ for GB^{OBC2}. ALPB is a method that postprocesses the results of GB^{OBC} (similar in spirit to eq 21) and is implemented in AMBER 9. Calculations with this method were performed just as with GB^{OBC}, except with an additional flag *alpb* set to 1 and an additional parameter *arad* set to a value precalculated with the *elsize* program.

Calculations with GBMV2 (angular grid with $N_\phi = 5$) were performed by running CHARMM c31b1.[54] This GB method also entails parametrization of the Born radius, in the form[25]

$$B_i = \frac{S}{C_0 A_4 + C_1 A_7} + D \quad (24)$$

where A_4 is the Coulomb-field result (eq 6) and A_7 is an analogous quantity with the exponent of the integrand changed from 4 to 7. There are, again, two parameter sets, old with $C_0 = 1 - 1/2^{1/2}$, $C_1 = 1$, $S = 0.9085$, and $D = -0.102$,[25] and new with $C_0 = 0.2966$, $C_1 = 1.0369$, $S = 0.9114$, and $D = -0.0637$.[15]

In GBr^6 and AGBNP, the solute boundary is defined by the vdW surface. This is also the case for GB^{OBC} and ALPB, but these two methods were parametrized to mimic PB with the solute boundary defined by the MS. In GBMV, the solute boundary is specifically defined as the MS. To make a fair comparison between the methods, errors of the methods are measured against their respective target PB results, calculated with the solute boundary set to either the vdW surface or MS. GBr^4/AGBNP is compared with both sets of PB results, because we found that its errors from PB with MS are actually less than from PB with a vdW surface. For the assessment of the methods, the solvent dielectric constant was 78.5, and the solute dielectric constant was either 1 or 4.

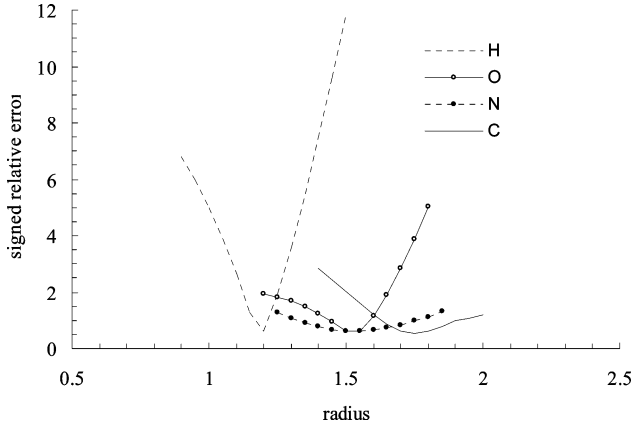

Figure 1. The variation of unsigned relative errors of GBr^6 with atomic radii.

As shown in Table 3, the unsigned error of GBr^6 is smaller than that of the best previous method, GBMV2 old, by a factor of 2 for $\epsilon_i = 1$ and by a factor of 5 for $\epsilon_i = 4$. Previous methods, including GBMV2 old, all systematically overestimate the magnitude of the solvation energy. In contrast, GBr^6 does not have significant systematic errors. The postprocessing method ALPB does reduce the systematic errors of GB^{OBC} somewhat, but the unsigned relative errors are still over 4.5%.

GBr^6 also appears to be the fastest method (Table 3). GBr^4/AGBNP took ~10% extra CPU time, mainly because of additional logarithmic functions in this method. The GB^{OBC} and ALPB methods are slightly slower than GBr^4/AGBNP. The most accurate previous method, GBMV2, took ~50% more time than all other methods. As a reference, the total CPU time for the UHBD calculations on the 55 proteins was 1000-fold greater than that for GBr^6.

Variation of Atomic Radii. Different sets of atomic radii have been used in PB and GB calculations. To assess the impact of atomic radii on the accuracy of GBr^6, we varied the radii of C, O, N, and H from the Bondi values by up to 0.3 Å. This range covers most of the radius values used in PB and GB calculations.

Figure 1 shows the variation of the unsigned relative errors of GBr^6 with the atomic radii. The minimum error occurs at or close to the Bondi values. The increase in error toward extreme values of radii is small for N, moderate for C and O, and steep for H. The steep increase is likely related to the crude treatment of hydrogens in determining the scaling coefficient s_{ji}.

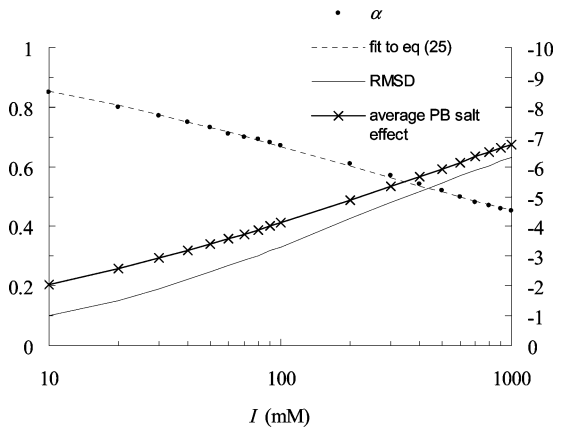

Figure 2. Comparison of GBr^6 and PB salt effects. The scale on the left ordinate is for α, the fit to eq 25, and the rmsd of GBr^6 from PB; the scale on the right ordinate is for the average PB salt effect. rmsd and average salt effect are in kcal/mol. PB and GBr^6 calculations were done on the set of 55 proteins, with $\epsilon_i = 4$ and $\epsilon_s = 78.5$. The ion exclusion radius was 2 Å in the PB calculations.

Salt Effects. Incorporating salt effects into the GB model is very desirable, since salts are present under physiological conditions. In GBr^6, salt effects are calculated according to eqs 21 and 22. For each ionic strength, the value of the parameter α was selected to minimize the rmsd of the GB salt effects from the corresponding PB results on the set of 55 proteins. Figure 2 shows that the optimal α values have a systematic dependence on the ionic strength. This dependence is well fitted to the function

$$\alpha = \frac{1 + 0.0169I^{1/2}}{1 + 0.075I^{1/2}} \qquad (25)$$

For the 55 proteins, the average PB salt effect changes from −2.0 kcal/mol at $I = 10$ mM to −6.8 kcal/mol at $I = 1000$ mM. Over the same range of ionic strength, the rmsd of GBr^6 from PB changes from 0.1 to 0.6 kcal/mol (Figure 2). The accuracy of the salt effects calculated by GBr^6 thus seems very encouraging.

For comparison, the rmsd of ALPB2 from PB is 0.5 and 2.3 kcal/mol, respectively, at $I = 100$ and 1000 mM. Errors in salt effects are even greater for GBMV2 old, with rmsd at 1.8 and 4.3 kcal/mol for the two salt concentrations. The difference in performance between GBr^6 on the one hand and ALPB2 and GBMV2 old on the other cannot be attributed to the use of different solute boundaries (vdW surface versus MS). In previous studies,[48,50] it was shown that the difference in salt effects between PB vdW and PB MS is very small; among the 55 proteins here the magnitudes of the latter on average were larger by 3% at I = 100 mM and by 6% at I = 1000 mM.

Conclusions

We have reported that our new GB method, GBr^6, reproduces PB solvation free energy with high accuracy. Extensive tests show that the errors of GBr^6 average ∼0.6% and are always <2.0% for proteins with dielectric constants of 1−4 and solvated in water. This level of error is substantially lower than those of current GB methods and is achieved with less computational time.

There are three ingredients in GBr^6 that may be useful for improving other GB methods. First, the r^6 approximation seems to significantly reduce the error of the Coulomb-field approximation. As a result, the need for parametrization of Born radii (e.g., eqs 23 and 24) can be eliminated. Its implementation also allows GBr^6 to be completely analytical, a feature that carries distinct advantage for applying the method to MD or MC simulations. Second, the introduction of the dielectric scaling factor, via eq 21, allows PB results over a wide range of solute and solvent dielectric constants to be reproduced by GB. Third, the additional salt-dependent scaling parameter α extends the accuracy of GB over a wide range of salt concentrations.

Acknowledgment. This work was supported in part by Grant GM058187 from the National Institutes of Health.

Note Added After ASAP Publication. This article was released ASAP on February 20, 2007. References 39 and 54 have been changed from their original form. Citations to these references have been revised in the text to reflect these changes. The corrected version was posted on February 21, 2007.

References and Notes

(1) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
(2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(3) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.
(4) Gilson, M. K.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327.
(5) Gilson, M. K.; Honig, B. *Proteins* **1988**, *4*, 7.
(6) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.
(7) Madura, J. D.; Briggs, J. M.; Wade, R.; 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.
(8) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.
(9) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037.
(10) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.
(11) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244.
(12) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035.
(13) Prabhu, N. V.; Zhu, P.; Sharp, K. A. *J. Comput. Chem.* **2004**, *25*, 2049.
(14) Baker, N. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137.
(15) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., 3rd. *J. Comput. Chem.* **2004**, *25*, 265.
(16) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008.
(17) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrikson, R. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(18) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
(19) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(20) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
(21) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(22) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
(23) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426.
(24) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.
(25) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348.
(26) Im, W.; Lee, M. S.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1691.
(27) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479.
(28) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
(29) Yu, Z.; Jacobson, M. P.; Friesner, R. A. *J. Comput. Chem.* **2006**, *27*, 72.
(30) Born, M. *Z. Phys.* **1920**, *1*, 45.
(31) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760.
(32) Tanizaki, S.; Feig, M. *J. Chem. Phys.* **2005**, *122*, 124706.
(33) Luo, R.; Moult, J.; Gilson, M. K. *J. Phys. Chem. B* **1997**, *101*, 11226.
(34) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(35) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817.

(36) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333.
(37) Tjong, H.; Zhou, H.-X. *J. Chem. Phys.* **2006**, *125*, 206101.
(38) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem.* **2000**, *104*, 3712.
(39) Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, 124902.
(40) Petitjean, M. *J. Comput. Chem.* **1994**, *15*, 507.
(41) Grant, J. A.; Pickup, B. T. *J. Phys. Chem.* **1995**, *99*, 3503.
(42) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. AMBER 8; University of California, San Francisco, 2004.
(43) Feng, H.; Takei, J.; Lipsitz, R.; Tjandra, N.; Bai, Y. *Biochemistry* **2003**, *42*, 12461.
(44) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(45) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
(46) Vijayakumar, M.; Zhou, H.-X. *J. Phys. Chem. B* **2001**, *105*, 7334.
(47) Dong, F.; Zhou, H.-X. *Biophys. J.* **2002**, *83*, 1341.
(48) Dong, F.; Vijayakumar, M.; Zhou, H.-X. *Biophys. J.* **2003**, *85*, 49.
(49) Dong, F.; Zhou, H.-X. *Proteins* **2006**, *65*, 87.
(50) Qin, S. B.; Zhou, H. X. *Biopolymers* **2007**, submitted.
(51) Damjanovic, A.; Garcia-Moreno, B.; Lattman, E. E.; Garcia, A. E. *Proteins* **2005**, *60*, 433.
(52) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(53) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(54) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.