Generalized Born Model Based on a Surface Integral Formulation

Avijit Ghosh, Chaya Sendrovic Rapp, and Richard A. Friesner*

Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027

Received: June 8, 1998; In Final Form: October 16, 1998

We have derived a surface-area-based version of the generalized Born model (S-GB) as a well-defined approximation to the boundary element formulation of the Poisson−Boltzmann (PB) equation. The relationship of the surface area methodology to the volume-integration-based approach of Still and co-workers is elucidated. On the basis of insights obtained from these results, we then develop empirical correction schemes which yield significant improvements in accuracy, as compared to the uncorrected GB model, in reproducing accurate solutions of the Poisson−Boltzmann equation. A large suite of energetic comparisons of GB, corrected S-GB, and PB for multiple conformations of peptides and proteins is presented.

1. Introduction

Accurately modeling solvation in molecular systems is crucial to quantitatively understanding the chemical and physical properties which underly many biomolecular processes. Continuum solvent models offer the possibility of locating well-defined minima on the energy surface with tractable computational costs. With the advent of modern computational tools, rigorous continuum methods using the Poisson−Boltzmann (PB) equation have enjoyed a resurgence in the literature,[1−3] in part due to the ability to reproduce the energetics of solute−solvent interactions as compared with both experimental data and theoretical explicit solvent approaches, at a fraction of the computational cost of the latter.

Nonetheless, the computational expense involved in solving the PB equation, calculated by means of current finite difference,[2] finite element,[4,5] or boundary element methods,[6,7] is significantly larger than the cost of calculating the energy of the solute with a molecular mechanics force field. Hence, it is inefficient to use the PB equation for extensive sampling of the conformational phase space in molecular mechanics calculations.

We have shown recently that simpler models such as a distance-dependent dielectric or a surface-area-based approach do not yield quantitative energetics in comparison with PB.[8] However, one approach that appears quite promising is the generalized Born (GB) model of Still and co-workers.[9,10] We have demonstrated that the GB model performs reasonably well in rank-ordering solvated peptide conformations as compared to the accurate solution of the PB equations.[8] However, the model did display some systematic errors, which appeared to increase as the net charge on the system became larger. Furthermore, tests were not carried out for larger systems, e.g., proteins in ref 8.

Our preliminary data on proteins, which will be discussed below, indicate that while qualitative relative energetics of the GB model are still quite good, the energy offsets increase considerably with molecular size, and there are occasionally serious errors in rank-ordering important low-energy structures. These results have motivated us to develop a new version of the GB model, based upon a surface integral rather than a volume integral formulation. From this work, we derive insight into the formal basis of the GB model (not particularly apparent from the original papers of the Still group).[9] Furthermore, we develop a series of empirical corrections that significantly reduce the deviation from PB results, in terms of both absolute energies and relative conformational rankings.

The paper is organized as follows. In section 2, we develop the underlying theory, showing the connection between the GB model, the PB equation, and the surface and volume formulations of the former. Section 3 presents our empirical correction schemes and describes the protocols by which parameters are optimized therein. Section 4 displays a database of results, ranging from small molecules to a number of conformations of the protein ribonuclease A, generated via simulated annealing in an attempt to predict the geometry of a loop of length 12 (this will be the subject of another publication).[11] For all structures, we compare the predictions of the accurate solution of the PB equation, the GB model as implemented in the Still group's Macromodel/Batchmin suite of molecular modeling programs, and our new surface GB (S-GB) model. Without the correction terms, the S-GB results are in quantitative agreement with the Macromodel GB results, as is mandated by the mathematical equivalence that we show in section 2. When the correction terms are added, significant improvements in the deviations of the absolute energies and mean square deviations of relative energies (as compared to the PB results) are obtained.

The present paper is focused upon demonstrating the improved accuracy that we have obtained with our empirical correction formalism. As the method requires surface rather than volume integration, we expect that the S-GB model will ultimately become more efficient computationally than the corresponding volume-based GB model as the size of the system increases. If one does comparisons with the accurate numerical integration methods in the Macromodel, increases in speed are immediately apparent. However, the analytical approximation schemes in Macromodel, and (in effect) multiple time-scale approach in which the Born α values are evaluated infrequently, are quite fast, albeit at the cost of increased deviations from the PB results. We are in the process of developing our own fast simulation schemes, and these will be the subject of a subsequent paper.

10.1021/jp982533o CCC: $15.00
Published on Web 12/10/1998

2. Theory

2.1. Derivation of an Approximation for the Energy of a Single-Point Charge in a Dielectric Cavity from the Poisson−Boltzmann Equation. We begin with the Poisson−Boltzmann equation,

$$\nabla\cdot\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{1}$$

$\epsilon(\mathbf{r})$ is the dielectric constant, $\rho(\mathbf{r})$ is the charge density, and $\phi(\mathbf{r})$ is the electrostatic potential. The spatial dependence of $\epsilon(\mathbf{r})$ is specified by the solvent-accessible surface of the molecule: for $\mathbf{r}$ inside the molecular cavity, $\epsilon(\mathbf{r}) = \epsilon_{\text{internal}}$ (typically assigned a value between 1 and 4), while for $\mathbf{r}$ outside, $\epsilon = \epsilon_{\text{solvent}}$. As has been discussed by many authors (e.g., ref 12 and references therein), optimization of the atomic radii which define the cavity, as well as other correction terms, is required to achieve quantitative accuracy compared to experiment. In this paper, however, we will be concerned only with the accuracy of the solution of eq 1 given a specification of the dielectric boundary.

For point charges in a dielectric cavity, one can reformulate eq 1 into a surface integral formulation involving the induced polarization charge. We shall not repeat this work here, instead referring the reader to refs 6, 7, and 13. The constitutive equation of this formulation, designated in the literature as the boundary element method, is given by

$$\phi(\mathbf{r}) = \sum_k \frac{q_k}{|\mathbf{r} - \mathbf{r}_k|} + \int_S \frac{\sigma(\mathbf{R})\, \mathrm{d}^2\mathbf{R}}{|\mathbf{r} - \mathbf{R}|} \tag{2}$$

where q_k is the charge on atom k, $\mathbf{r}_k$ is its coordinate, and $\sigma(\mathbf{R})$ is the induced polarization charge density on the dielectric boundary at point $\mathbf{R}$, where $\mathbf{R}$ represents the vector of integration over the surface of the molecule.

The induced polarization charge $\sigma(\mathbf{r})$ on the surface is related to the normal component of the electric field, $\mathbf{E}(\mathbf{r})$ at the surface, approaching from inside the cavity, by the relation

$$\sigma(\mathbf{r}) = \frac{1}{4\pi}(1 - 1/\epsilon)\mathbf{E}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) \tag{3}$$

The normal component of the electric field, $E(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})$, at the surface can be decomposed into two components, a Coulombic contribution from the internal charges q_k,

$$\mathbf{E}_{(\text{Coul.})}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = \sum_k q_k \frac{(\mathbf{r} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|^3} \tag{4}$$

and an integral contribution from the induced polarization charge distribution on the surface,

$$\mathbf{E}_{(\text{ind.})}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 2\pi\sigma(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) + \int_S \frac{\sigma(\mathbf{R})(\mathbf{r} - \mathbf{R})\cdot\mathbf{n}(\mathbf{r})}{|\mathbf{r} - \mathbf{R}|^3}\, \mathrm{d}^2\mathbf{R} \tag{5}$$

The electrostatic free energy U can be found directly from the induced polarization charge by

$$U = -\frac{1}{2}\int_S \sigma(\mathbf{R})\phi_0(\mathbf{R})\, \mathrm{d}^2\mathbf{R} \tag{6}$$

where the factor of $^1/_2$ arises from the work needed to produce the polarization of the dielectric. Here, $\phi_0(\mathbf{R})$ is the bare Coulombic potential from the internal charges,

$$\phi_0(\mathbf{R}) = \sum_k \frac{q_k}{|\mathbf{R} - \mathbf{r}_k|} \tag{7}$$

and, as before, q_k and $\mathbf{r}_k$ are the charge and coordinate of atom k. In boundary element codes,[6,7] the induced surface charge density is found iteratively using only the contribution from the internal charges (eq 4) as the initial guess for the normal component of the electric field.

An important special case of the PB equation that can be solved exactly is that of a Born sphere, a single-point charge q_k in the center of a spherical dielectric cavity of with radius r. For simplicity, consider a model in which the dielectric inside the sphere is 1 and that of the solvent is ϵ. In this case, the induced charge density at the surface is

$$\sigma(\mathbf{r}) = \sigma = \frac{1}{4\pi}(1 - 1/\epsilon)\frac{q_k}{r^2} \tag{8}$$

which, using eq 6 and eq 7, immediately yields

$$U_{\text{sphere}} = -\frac{1}{8\pi}(1 - 1/\epsilon)\int_S \frac{{q_k}^2}{|\mathbf{R} - \mathbf{r}_k|^3}\, \mathrm{d}^2\mathbf{R} \tag{9}$$

Carrying out the integration in eq 9 yields

$$U_{\text{sphere}} = -\frac{1}{2}(1 - 1/\epsilon)\frac{{q_k}^2}{r} \tag{10}$$

An interesting feature of the solution of the Born model is that the exact surface charge density, eq 3, is equal to that specified only by the normal component of the electric field arising from the charge distribution inside the molecule (eq 4). The exact equivalence is a consequence of the symmetry of the Born model, which leads to eq 5 reducing to zero.

While perfect spherical symmetry will not be manifested in a realistic molecular problem, one might hypothesize that for systems in which the surface is locally convex, eq 4 would provide a reasonable approximation to the electric field and, hence, the surface charge density. Furthermore, empirical corrections representing the effects of the second term, eq 5, might be available in a systematic form.

Specifically, by using only the Coulombic contribution to the electric field (eq 4) as a basis for the induced polarization charge (eq 3) and integrating using eq 6 and eq 7, we arrive at our approximation for the solvation free energy of a single-point charge q_k at $\mathbf{r}_k$ in a cavity of arbitrary shape, U_{se}, given by

$$U_{\text{se}} = -\frac{1}{8\pi}(1 - 1/\epsilon)\int_S \frac{{q_k}^2}{|\mathbf{R} - \mathbf{r}_k|^4}(\mathbf{R} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{r})\, \mathrm{d}^2\mathbf{R} \tag{11}$$

One can easily verify that eq 11 reduces to eq 9 for the case of a spherical cavity. This is the formulation that we will use in the S-GB model described below.

2.2. GB Model: Surface and Volume Formulations. It is well-known that the total electrostatic contribution to the solvation free energy for point charges in the molecular cavity can be decomposed pairwise. That is, the electrostatic energy U can be rewritten as

$$U = \sum_k \hat{U}_{\text{se}}(q_k, \mathbf{r}_k) + \sum_{i\neq j} \hat{U}_{\text{pr}}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j) \tag{12}$$

$\hat{U}_{\text{se}}$ corresponds to the solution of the Poisson−Boltzmann

equation for a single charge q_k at position $\mathbf{r}_k$ in the molecular dielectric cavity, defined by $\epsilon(\mathbf{r})$, while $\hat{U}_{pr}$ corresponds to the screened Poisson−Boltzmann interaction energy between two charges. In the following discussion, $\acute{U}_{se}$ and $\acute{U}_{pr}$ represent the single-charge energy and the pair interaction energy for Still's formulation of the GB model. In the GB model, $\hat{U}_{se}$ is approximated by

$$\acute{U}_{se} = \frac{3}{8\pi}(1 - 1/\epsilon)\int_V \frac{q_k^2}{|\mathbf{R} - \mathbf{r}_k|^4}\, d^3\mathbf{R} \tag{13}$$

Still's formulation of the GB model, eq 13, can be shown to be equivalent to eq 11 above by means of Green's theorem:

$$\int_V \nabla\cdot\mathbf{A}\, d^3\mathbf{r} = \int_S \mathbf{A}\cdot\mathbf{n}(\mathbf{r})\, d^2\mathbf{r} \tag{14}$$

Choosing

$$\mathbf{A}(\mathbf{r}) = \frac{\mathbf{R} - \mathbf{r}_k}{|\mathbf{R} - \mathbf{r}_k|^4} \tag{15}$$

we obtain the result

$$\acute{U}_{se} = \frac{3q_k^2}{8\pi}(1 - 1/\epsilon)\int_V \frac{\mathbf{R} - \mathbf{r}_k}{|\mathbf{R} - \mathbf{r}_k|^5}\, d^3\mathbf{R} \tag{16}$$

$$\acute{U}_{se} = -\frac{q_k^2}{8\pi}(1 - 1/\epsilon)\int_V \nabla\cdot\frac{\mathbf{R} - \mathbf{r}_k}{|\mathbf{R} - \mathbf{r}_k|^4}\, d^3\mathbf{R} \tag{17}$$

$$\acute{U}_{se} = -\frac{q_k^2}{8\pi}(1 - 1/\epsilon)\int_S \frac{(\mathbf{R} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})}{|\mathbf{R} - \mathbf{r}_k|^4}\, d^2\mathbf{R} \tag{18}$$

which is equivalent to the formula U_{se} in eq 11.

The key quantities for approximating the interaction terms $\hat{U}_{pr}$ in the GB model are the "Born α" parameters. These are the effective radii of a hypothetical Born sphere having the same electrostatic energy of a single energy term. That is, given $U_{se}(q_k, \mathbf{r}_k)$ for a particular atom k, one can use eq 10 to solve for the radius of a Born sphere having the same contribution.

Still's original formulation for the electrostatic free energy uses the following term for a pair of atoms i and j and the distance between them r_{ij}

$$\acute{U}_{pr} = -\frac{1}{2}(1 - 1/\epsilon)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}} \tag{19}$$

where

$$\alpha_{ij} = \sqrt{\alpha_i \alpha_j} \tag{20}$$

and

$$D = \frac{r_{ij}^2}{(2\alpha_{ij})^2} \tag{21}$$

One can easily verify that using eq 19, the total electrostatic energy (eq 12) reduces to the Born solution (eq 10), in the case of superimposed charges. Furthermore, as r_{ij} becomes large, eq 19 approaches a Coulombic pair term, as expected.

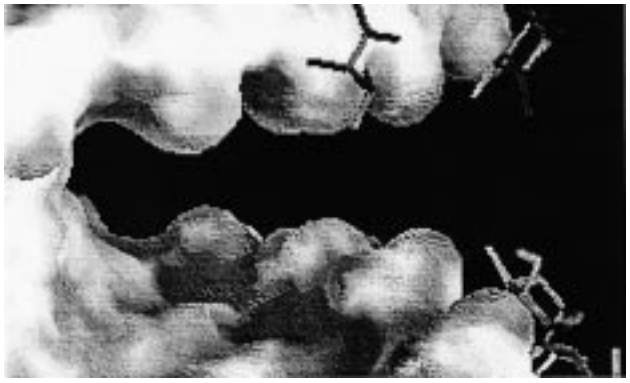

Figure 1. Molecular surface of a short DNA fragment being bound by chromomyosin. Generated using GRASP: Anthony Nicholls, Kim Sharp, and Barry Honig. Proteins, Structure, Function, and Genetics; 1991, Vol. 11, No. 4, pp 281ff.

3. Empirical Corrections to the S-GB Model

3.1. Overview. We have developed our correction schemes by examining systematic deviations between the primitive S-GB results (which, as shown above, are mathematically exactly equivalent to the volume integral formulation of Still and coworkers) and the PB results for a large database of molecules. We have discovered that these deviations can be in large part corrected via relatively simple empirical modifications of the formalism. We also attempt to justify our modifications via physical arguments, although these arguments are not quantitatively derived from the PB equation. Corrections are presented both for the calculation of U_{se} and for U_{pr}.

Two types of corrections are described below. We refer to the first type as short-range corrections, because these are determined by the local environment of the charge or pair of charges. The idea here is that the distance from the surface and number of nearest neighbors can be used to systematize much of the difference between our approximate induced charge and the real induced charge. Consider, for example, a carbonyl group in which the oxygen is maximally exposed to solvent. It is reasonable that the single-charge energy (or, equivalently, the Born α) associated with this oxygen should deviate from the S-GB initial guess in a similar fashion for all such carbonyls, without regard to the rest of the molecule. In actual fact, these regularities turn out to be more reliable than one would have expected a priori, a fortunate aspect of our investigations.

The second type of correction we refer to as long range. While the short-range corrections provide remarkably accurate results for small molecules or even for larger systems which are basically convex in the shape of their surface, problems arise in situations like the one shown in Figure 1 where significant concavities in the surface lead to a strong interaction between surface charges in relatively distant regions, in effect screening the source charge from a portion of the induced surface charge. This deviation is not surprising, as the GB model is exact for a particle at the center of a sphere, where the entire surface is strictly convex. Indeed, a criterion for all of our corrections is that for the Born model, the correction terms are rigorously zero. We detect long-range contributions by examining the dot product of the vector from the charge to the surface element with the surface normal; only terms where this dot product is in the neighborhood of −1 make significant contributions to the long-range corrections. The specific implementation of each correction term is described in detail below.

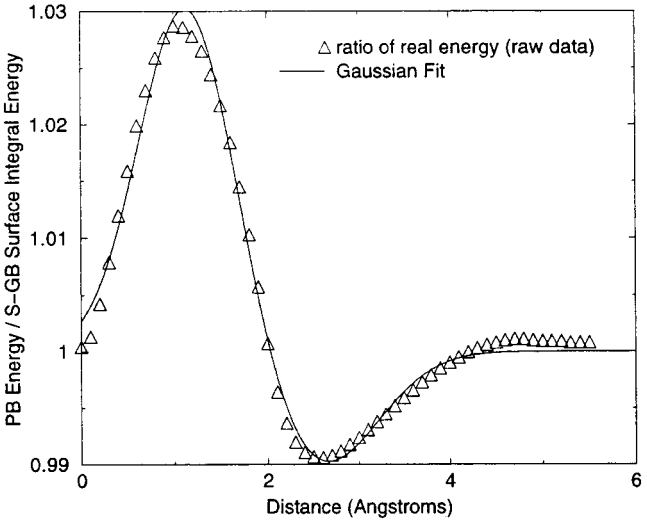

Figure 2. Short-range correction for a single atom and a neighboring dielectric cavity: $R_{srp}(r_1, r_2, \beta)$.

3.2. Single-Charge Energies. 3.2.1. Short-Range Correction. The single-charge correction is based on the fact that a real molecular dielectric cavity for a molecule is a superposition of many atomic cavities. Let $\epsilon_k(\mathbf{r})$ represent the dielectric cavity for atom k; then, the molecular dielectric cavity for N atoms can be represented as

$$\epsilon_{mol}(\mathbf{r}) = \min(\epsilon_1(\mathbf{r}), \epsilon_2(\mathbf{r}), ..., \epsilon_k(\mathbf{r}), \epsilon_{k+1}(\mathbf{r}), ..., \epsilon_N(\mathbf{r})) \quad (22)$$

The deviations in U_{se} for atom k from the Born approximation (eq 10) are therefore due to the superposition of other dielectric cavities. The simplest example containing the effects of cavity overlap is that arising from a pair of atomic cavities:

$$\epsilon_{mol}(\mathbf{r}) = \min(\epsilon_1(\mathbf{r}), \epsilon_2(\mathbf{r})) \quad (23)$$

Consider, for example, a charged sphere of radius 1.0 Å (r_1) and an uncharged sphere (representing only a dielectric cavity) also of radius 1.0 Å (r_2). Letting R_{srp} be the ratio of the true Poisson−Boltzmann energy $\acute{U}_{se}$ using a probe radius of 1.4 Å to our model approximation U_{se} (eq 11), R_{srp} can be mapped out as a function of distance (in Å) between the two spheres as shown in Figure 2.

Also shown in Figure 2 is a fit to these data using an expansion in two Gaussians:

$$R_{srp}(r_1, r_2, \beta) = \beta_1 e^{-\beta_2(d-\beta_3)^2} + \beta_4 e^{-\beta_5(d-\beta_6)^2} + 1 \quad (24)$$

The parameter d is designed to measure the degree to which the second uncharged dielectric cavity protrudes from the initial sphere. Therefore, d is defined to be zero if the second sphere is contained inside the first and to be the maximum distance of protrusion otherwise. The β_i are fitting parameters used to fit the model Gaussians to the actual ratio $\acute{U}_{se}/U_{se}$ for this particular diatomic. Hence, the correct U_{se} for this particular diatomic can be represented as

$$U'_{se,pair} = R_{srp}(r_1, r_2)U_{se} \quad (25)$$

When the neighboring cavity is close to the charged atom ($<(r_1 + r_2)/2$), the surface is roughly an elongated ellipsoid. In this case, eq 11 underestimates slightly the true contribution to the energy, given the additional convex surface area. As the distance between the charged cavity and the uncharged cavity increases, the center of the dielectric cavity crosses the edge of the charged atom (around 2 Å in Figure 2). The molecular surface develops an invagination, resulting in a slight overestimation of the true electrostatic energy in the same way as cases involving the long-range screening correction in the next section. Finally, as the uncharged cavity moves even further away from the source charge, it can no longer contribute significantly to the true electrostatic energy, restoring the ratio R_{srp} back to unity as expected.

TABLE 1: Representative Values of β for the Short-Range Correction $R_{srp,i}$ (Eq 28)

atom radius, Å							
charged	uncharged	β_1	β_2	β_3	β_4	β_5	β_6
1.0	1.0	0.031 1	1.82	1.15	−0.010 4	1.16	2.54
1.9	1.9	0.034 7	0.485	2.25	0	0	0
2.1	2.1	0.035 2	0.381	2.51	0	0	0
1.1	1.9	0.008 38	1.29	1.39	−0.002 52	1.63	2.79
1.9	1.1	0.104	0.999	1.57	0	0	0
1.1	2.1	0.006 63	1.23	1.41	−0.001 97	1.42	2.8
2.1	1.1	0.123	0.937	1.72	0	0	0
1.9	2.1	0.027 8	0.471	2.29	0	0	0
2.1	1.9	0.0437	0.398	2.28	0	0	0

TABLE 2: Scaling Parameter $S(N_i)$ (Eq 26) for the Short-Range Correction

van der Waal's neighbors (N)	$S(N)$
1	1.4
2	1.5
3	1.9
4	1.8

The parameters β_i are calculated by comparing the PB energy for a diatomic system of all possible radii (with 0.01-Å grid spacing) with the surface integral approximation U_{se} (eq 11) and carrying out a least-squares fit. For atoms with more than a single neighbor, this function is calculated over all neighbors and scaled empirically by a single parameter S, depending only on the number of van der Waals neighbors N_i that particular atom i has, where a van der Waal neighbor is defined as a neighboring atom that is closer than the average of the van der Waal radius of the original atom and itself.

$$R_{sr,i} = S(N_i)^{-1}(\sum_k (R_{srp,i}(r_i, r_k) - 1)) + 1 \quad (26)$$

$S(N)$ was found by examining a large database of random hexapeptides and calculating both U_{se}, the energy calculated via our surface integral formulation, and $\acute{U}_{se}$, the true Poisson−Boltzmann energy for charges located at all possible atomic positions in each member of the set. The single-charge energy using only the short-range correction with the scaling parameter $S(N_i)$ is calculated as

$$U'_{se,i} = U_{se,i}R_{sr,i} \quad (27)$$

$$U'_{se,i} = U_{se,i}(S(N_i)^{-1}(\sum_k (R_{srp,i}(r_i, r_k) - 1)) + 1) \quad (28)$$

Representative values of β and the values for $S(N_i)$ can be found in Tables 1 and 2.

3.2.2. Long-Range Correction. The long-range term corrects for the interactions between the induced charge on nonadjacent portions of the surface. This becomes significant when there are invaginations as shown, e.g., in Figure 1.

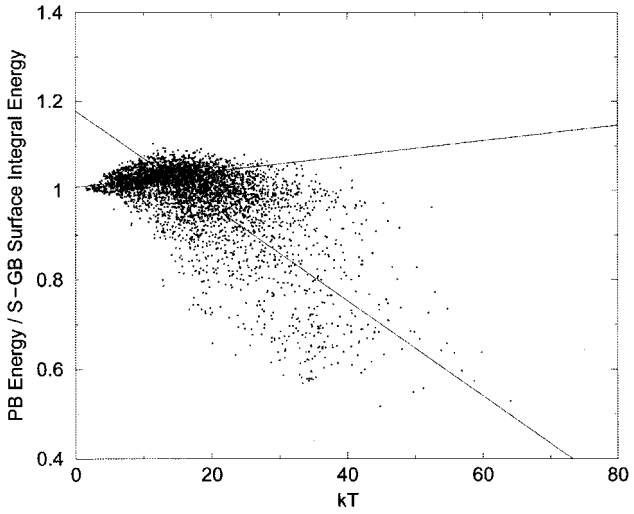

Figure 3. $\hat{U}_{se,k}/U'_{se,k}$ with respect to O_k (axis in kT)**.**

TABLE 3: Long-Range Correction Parameters for R_{lr} (Eq 31)

parameter	value
γ_1	0.385
γ_2	1.008
γ_3	−0.236
γ_4	1.178

The correction is essentially empirical in nature and is a simple functional of the portion of the single energy integral (eq 11), O_k, that comes from the invaginated portion of the surface. This contribution is estimated via the formula

$$O_k = \int_S -\frac{1}{|\mathbf{R}-\mathbf{r}_k|^4} e^{-0.5([(\mathbf{R}-\mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})]/|\mathbf{R}-\mathbf{r}_k|)} (\mathbf{R}-\mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})\, d^2\mathbf{R} \quad (29)$$

where $\mathbf{r}_k$ is the coordinate of atom k, $\mathbf{n}(\mathbf{R})$ the surface normal, and $\mathbf{R}$ the vector of integration. The integrand is simply that previously used in eq 11 multiplied by an exponential of the dot product $(\mathbf{R}-\mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})$, which is taken to be a measure of the invagination of the surface with respect to the atomic coordinate.

Figure 3 represents a plot of $\hat{U}_{se,k}/U'_{se,k}$ against O_k. The major portion of the correction is clearly due to charge screening at large values of O_k. Additionally, at small values, there is a small correction for a slight underestimation of the true PB energy. The two linear functions in Figure 3 represent the functional correction $R_{lr,k}(O_k)$ and can be written out as follows:

$$R_{lr,k} = \gamma_1 O_k + \gamma_2 \qquad O_k < 18 \text{ kT} \quad (30)$$

$$R_{lr,k} = \gamma_3 O_k + \gamma_4 \qquad O_k > 18 \text{ kT} \quad (31)$$

The specific values for γ are given in Table 3. With both the long- and short-range corrections, the approximate U_{se} can be written as

$$U''_{se,k} = R_{lr,k}(O_k) U'_{se,k} \quad (32)$$

$$U''_{se,k} = U_{se,i}(S(N_i)^{-1}(\sum_k (R_{srp,i}(r_i, r_k) - 1)) + 1) R_{lr,k}(O_k) \quad (33)$$

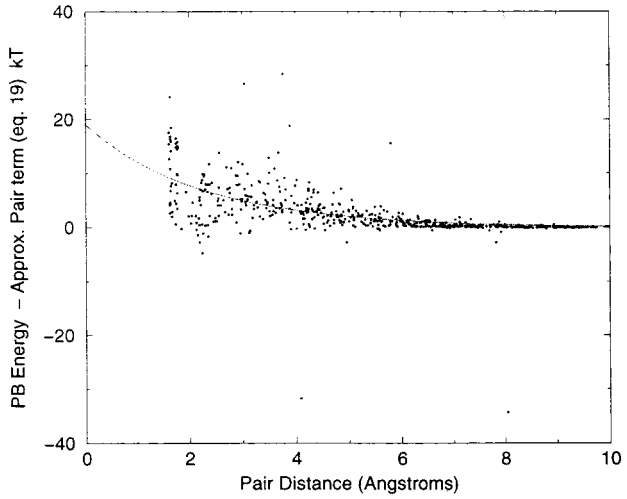

Figure 4. Difference between PB pair term and eq 19 (α_i calculated from $\hat{U}_{se}$) (in kT) as a function of pair distance.

TABLE 4: Parameters for Pair Function Correction R_{pr} (Eq 34)

no. of van der Waal's neighbors			
atom 1	atom 2	δ_1	δ_2
0	0	18.966	0.447 41
0	1	7.5011	0.325 88
0	2	9.5131	0.359 55
0	3	8.8575	0.335 03
0	4	10.183	0.387 92
1	1	15.379	0.456 11
1	2	4.2971	0.293 09
1	3	5.3653	0.331 93
1	4	5.7968	0.525 24
2	2	2.411	0.278 92
2	3	3.6653	0.235 52
2	4	3.6077	0.356 07
3	3	5.1236	0.580 91
3	4	3.6172	0.315 63
4	4	5.5207	0.467 65

3.3. Pair Function Energies. As shown in section 2.2, $\acute{U}_{pr}$ (eq 19) represents Still's approximate pair interaction term for the total electrostatic solvation energy (eq 12). Figure 4 plots $\hat{U}_{pr} - \acute{U}_{pr}$ or the difference between the exact Poisson−Boltzmann interaction term and the approximate one using the exact the Poisson−Boltzmann single energies $\hat{U}_{se}$. Figure 4 shows that eq 19 consistently underestimates the real interaction energy of the system, in effect overestimating the local neighborhood surface charge. The overestimation can be partly attributed to the lack of dielectric and, hence, molecular surface between the charges that is there for the two hypothetical spherical cavities. As a pair term inherently involves a pair of atoms, a correction of the functional form

$$R_{pr} = \delta_1 e^{-\delta_2 r_{ij}} \quad (34)$$

is used, where δ_1 and δ_2 are the fitting parameters. The corrected electrostatic pair term incorporates Still's pair term, $\acute{U}_{pr}$ (eq 19), as follows:

$$U_{pr} = \acute{U}_{pr} + R_{pr} \quad (35)$$

This is done for pairs of atoms grouped by the number of van der Waals neighbors, where a van der Waals neighbor is defined

TABLE 5: GBSA/S-GB Peptide Calculations

chain length	no. of conformers	no. of atoms	GBSA			S-GB		
			abs error	rel error	rel square error	abs error	rel error	rel square error
4	534	58	10.229	2.4339	9.1371	2.027	2.0207	6.2728
6	430	93	3.5423	3.0095	14.399	4.7132	1.9982	6.3078
8	307	139	19.291	2.5101	10.127	1.6839	2.2415	8.0249
14	95	245	32.016	4.4931	34.478	5.3434	2.8237	13.68
18	70	280	23.703	4.2121	38.811	2.2127	3.3032	22.869
22	6	362	45.414	3.9866	22.278	2.9309	1.1543	1.8909
26	33	436	52.495	3.9753	28.839	2.6493	3.5993	20.786
30	15	620	65.929	4.9359	37.954	10.801	2.452	8.9455
35	38	608	55.889	3.0498	15.202	4.3012	3.8048	22.31
43	16	487	89.477	5.5338	46.736	14.296	5.3112	40.987

TABLE 6: Ribonuclease A (Generated by Minimization of Substructure followed by Minimization of Entire Protein)

Delphi		GBSA		S-GB	
conformer no.	energy, kT	conformer no.	energy, kT	conformer no.	energy, kT
44	−2802.89	48	−2495.43	44	−2767.22
36	−2797.47	44	−2467.41	48	−2763.54
21	−2773.59	36	−2437.18	36	−2750.17
11	−2767.27	11	−2435.29	21	−2734.61
12	−2758.92	29	−2429.16	29	−2724.94
5	−2756.76	12	−2426.33	11	−2724.81
34	−2754.84	14	−2424.71	12	−2722.21
48	−2749.9	21	−2416.34	33	−2716.22
33	−2749.01	18	−2415.74	14	−2713.06
37	−2747.12	30	−2410.32	37	−2711.05
29	−2746.78	1	−2409.5	34	−2709.25
6	−2737.39	33	−2408.5	5	−2705.65
10	−2735.8	34	−2408.47	27	−2704.67
17	−2734.54	27	−2407.26	1	−2703.39
30	−2733.91	5	−2403.61	10	−2700.52
1	−2733.3	37	−2401.23	17	−2698.41
19	−2733.3	40	−2398.33	30	−2697.12
14	−2731.43	17	−2393.22	18	−2691.45
27	−2723.01	19	−2392.4	6	−2691.38
26	−2719.98	6	−2386.06	19	−2689.53
45	−2716.32	10	−2385.85	45	−2679.47
18	−2716.16	45	−2380.51	26	−2678.26
42	−2712	26	−2374.28	42	−2674.61
13	−2696.17	7	−2373.35	7	−2670.21
7	−2692.96	42	−2367.35	40	−2669.77
25	−2685.88	25	−2364.53	25	−2663.15
40	−2682.89	13	−2350.7	13	−2661.26
	av error (GBSA)			330.61 kT	
	sq error (GBSA)			10984 kT	
	av error (S-GB)			33.368 kT	
	sq error (S-GB)			1220.2 kT	

TABLE 7: Ribonuclease A (Generated by Simulated Annealing of Substructure followed by Minimization of Entire Protein)

Delphi		GBSA		S-GB	
conformer no.	energy, kT	conformer no.	energy, kT	conformer no.	energy, kT
34	−2741.89	22	−2464.29	34	−2770.66
25	−2727.75	34	−2463.41	22	−2761.17
22	−2712.41	16	−2404.44	16	−2720.18
13	−2705.15	24	−2396.69	25	−2716.16
16	−2698.95	13	−2395.46	33	−2691.15
12	−2691.84	33	−2379.77	24	−2689.96
9	−2681.08	25	−2377.5	13	−2687.16
24	−2680.49	19	−2361.37	12	−2681.93
33	−2677.39	9	−2348.43	19	−2673.26
30	−2657.87	12	−2347.92	17	−2652.96
19	−2653.42	17	−2345.7	30	−2651.59
17	−2652.14	30	−2339.6	39	−2644.42
5	−2651.07	21	−2335.74	38	−2644.23
35	−2644.08	39	−2329.36	9	−2640.76
38	−2640.92	5	−2319.23	42	−2635.62
21	−2636.68	35	−2313.56	35	−2634.03
39	−2622.79	42	−2312.72	21	−2631.16
42	−2622	28	−2308.17	5	−2628.1
28	−2618.01	38	−2307.38	28	−2624.73
18	−2596.03	36	−2306.03	36	−2615.3
31	−2596.01	18	−2274.69	31	−2591.93
36	−2591.54	1	−2254.47	18	−2587.51
45	−2590.04	29	−2253.2	45	−2584.2
37	−2569.12	45	−2250.62	37	−2558.44
11	−2568.12	31	−2249.75	29	−2554.48
1	−2547.98	11	−2240.56	1	−2549.43
29	−2535.94	37	−2238.29	11	−2542.88
40	−2529.21	40	−2216.76	40	−2528.96
4	−2496.85	23	−2205.95	26	−2513.91
3	−2479.59	3	−2203.35	4	−2510.19
26	−2479.41	4	−2194.29	3	−2497.23
6	−2474.2	26	−2191.04	23	−2493.25
8	−2456.69	27	−2177.25	27	−2479.79
27	−2452.01	6	−2154.4	6	−2463.16
23	−2436.63	14	−2145.47	8	−2451.73
14	−2431.93	8	−2133.54	41	−2448.53
10	−2431.26	7	−2128.91	7	−2445.61
7	−2431.21	41	−2116.38	14	−2442.50
41	−2412.98	10	−2095.12	10	−2422.03
	av error (GBSA)			306.2 kT	
	sq error (GBSA)			94438 kT	
	av error (S-GB)			16.578 kT	
	sq error (S-GB)			440.4 kT	

as the neighboring atom which is closer than the average of the van der Waals radii of the original atom and the neighbor. The best fit values for δ are given in Table 4. This form does not converge to the single-charge energy as r_{ij} approaches 0. This is because the correction term involves pairs of atoms that are at the smallest, 1 Å apart. However, its performance represents a clear improvement for the relevant portion of the distance space where the correction term is being used to fix the interaction energy.

3.4. Numerical Implementation. To obtain the results in this paper, we implemented a simple numerical quadrature on the molecular surface based upon grid points associated with each atomic sphere that are assigned equal weights. This is by no means the optimally efficient approach to the problem. A subsequent paper will present a more sophisticated approach, incorporating multipole expansions and providing algorithms for the evaluation of the analytical gradient of the solvation free energy. Here, however, the only relevent issue is convergence of the integral, which we have checked by systematic increases in the density of mesh points. All results, repeated below, represent converged energies, using 400 grid points per surface atom.

4. Results

In section 3, we developed an approximation for the total electrostatic energy and demonstrated that parameters could be adjusted to reproduce both single energies and pair energies

TABLE 8: Ribonuclease A (Generated by Simulated Annealing of Entire Protein followed by Minimization)

Delphi		GBSA		S-GB	
conformer no.	energy, kT	conformer no.	energy, kT	conformer no.	energy, kT
9	−2777.03	9	−2428.69	12	−2732.27
32	−2772.07	1	−2416.6	9	−2729.23
12	−2771.91	12	−2416.19	32	−2718.87
8	−2755.3	8	−2415.28	20	−2718.75
20	−2753.98	32	−2412.7	1	−2717.73
29	−2750.51	21	−2411.68	8	−2717.19
14	−2745.24	14	−2406.31	14	−2709.7
1	−2744.91	30	−2405.32	2	−2709.32
2	−2742.46	20	−2401.61	29	−2709.24
11	−2738.66	2	−2401.44	21	−2701.75
21	−2733.65	29	−2400.02	11	−2701.01
27	−2732.73	11	−2391.95	30	−2699.22
30	−2729.48	27	−2386.6	27	−2695.15
17	−2722.91	10	−2379.69	17	−2690.65
10	−2721.25	13	−2375.12	10	−2681.01
13	−2718.69	17	−2371.25	13	−2680.15
6	−2644.47	6	−2321.9	6	−2624.97
	av error (GBSA)				341.94 kT
	sq error (GBSA)				11705 kT
	av error (S-GB)				36.414 kT
	sq error (S-GB)				1380.7 kT

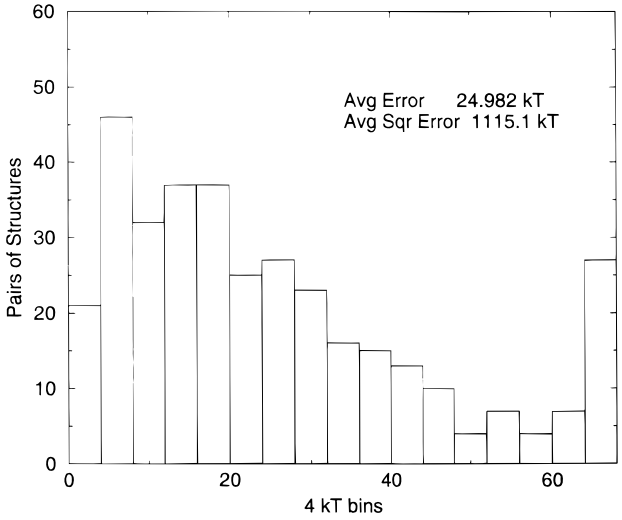

Figure 5. Relative GBSA errors for Table 6.

with greater accuracy than the original GB model. However, this does not necessarily guarantee better results for the total energy of an actual large molecule, as this is a sum of a large number of terms and one approach could have much better cancellation of error than another when all of the pairs are added up (i.e., how random are the distribution of errors?). Consequently, it is crucial to directly compare the two formulations with accurate Poisson−Boltzmann results and examine statistical measures of the deviations to see whether the absolute and relative energies of our proposed new function are indeed more accurate.

Table 5 summarizes the results of a series of calculations on various peptides that were obtained from Monte Carlo conformer searches. S-GB, in general, does better than the standard GB model in the conformational rankings for each series, as illustrated by the square relative errors for each method. Moreover, an examination of the absolute errors shows that the linear offset has, remarkably, been virtually eliminated. The absolute and relative errors are defined as deviations from the Poisson−Boltzmann error or $(\sum_i |PB_i - E_i|)/n$, where E_i is the approximate

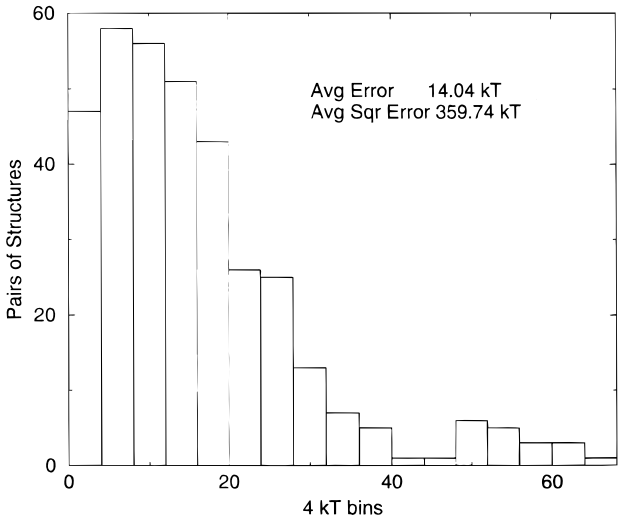

Figure 6. Relative S-GB errors for Table 6.

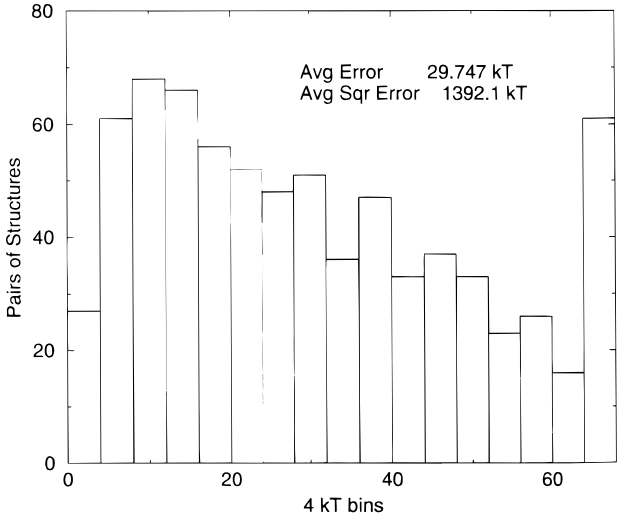

Figure 7. Relative GBSA errors for Table 7.

energy and n is the number of conformers, and the relative errors are defined as $(\sum_{i>j} |(PB_i - E_i) - (PB_j - E_j)|)/n$.

Finally, the electrostatic energies of a series of protein conformers of ribonuclease A were calculated and rank-ordered in Tables 6−8. The three sets of protein conformers arise from GB simulations in which minimization of the protein and the loop substructure were carried out in different ways, as is described in detail elsewhere.[11] The first set represents a set of conformers generated by first minimizing only the substructure followed by minimization of the entire protein. The second set of conformers represents a series of minima found by first using simulated annealing on the substructure followed by minimization of the entire protein. The third series of conformers was generated by using simulated annealing on the entire protein followed by minimization. Again, an order of magnitude decrease in the offset in the S-GB model in comparision to GBSA is apparent. In addition to calculating the mean-square relative errors, Figures 5−10 combine these relative errors in groups of 4 kT, showing again the extent to which the short- and long-range corrections can improve the standard GB model in its ability to do conformational ranking.

5. Conclusion

We have demonstrated that corrections to the GB solvation model can yield significant improvements in the accuracy of

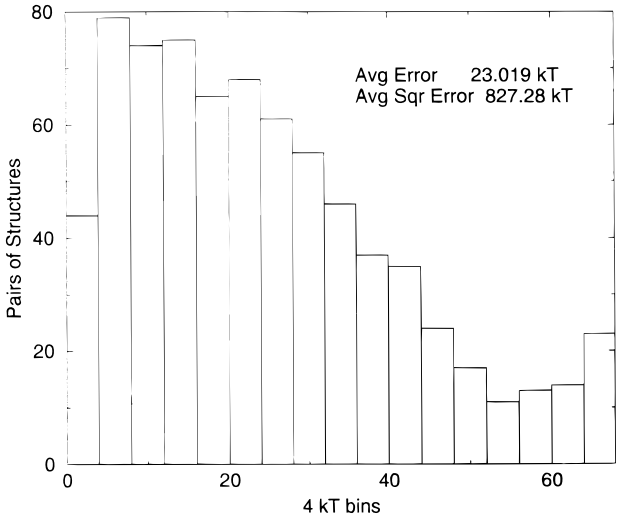

Figure 8. Relative S-GB errors for Table 7.

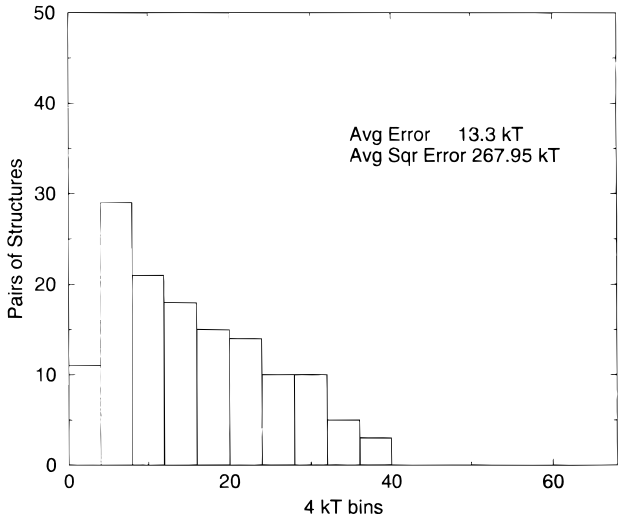

Figure 9. Relative GBSA errors for Table 8.

the approach as compared to accurate PB continuum solvation calculations. Figures 5−10 demonstrate that there is a major reduction in the fraction of large relative errors for the protein systems. The reduction in mean-square error by a factor of 2−3 is a reasonable quantitative measure of the effectiveness of the improvement. The enhanced performance of the model will be particularly useful in our planned utilization of the GB model as a fast screening method, to be followed by more accurate PB calculations on the low-lying minima; if the energy errors are restricted to a relatively small range, the number of minima that need to be examined with the more expensive PB calculations will be qualitatively diminished. The virtual elimination of the offset error could also be quite significant, for example, in comparing folded and unfolded structures (something we did not do in the present paper) or in comparing the relative energetics of protonated and unprotonated forms of a protein.

In addition to the numerical improvements discussed above, a major objective of the paper has been to reformulate the GB approach in a new analytical context. In our view, the surface formulation provides qualitatively enhanced insight into the derivation of the approximation and the reasons for its successes and failures. The correction models described here represent

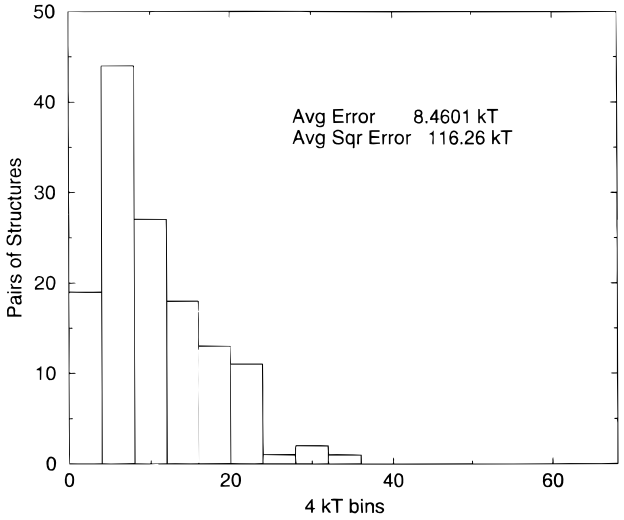

Figure 10. Relative S-GB errors for Table 8.

only an initial effort to eliminate the errors; in particular, the long-range errors still exhibit significant dispersion, and our present correction term is a crude first approximation. There are clearly additional factors, which have not been included in the model, which are generating deviations from the PB results, and further investigation should be able to uncover these and ameliorate their effects.

Finally, it is worth noting that, despite the relative complexity of the parameterization, the corrections have a negligible effect on the CPU time; the parameter development only has to be carried out once and is then stored as a preprocessing step. Development of new parameters for additional types of atomic cavities is straightforwardly performed by the prescription given here. We have also been careful to make sure that (a) the ratio of data to parameters remains very large and (b) the final test cases were not used in performing the fitting, so the test results are unbiased. Of course, additional tests are still needed to judge the performance of the model over the full range of possible molecular systems.

Acknowledgment. This paper was supported in part by a grant from the NIH (GM-52018) to R.A.F. We thank Peter Shenkin and Barry Honig for useful discussions.

References and Notes

(1) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(2) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comp. Chem.* **1987**, *9*, 327−335.
(3) Nicholls, A.; Honig, B. *J. Comp. Chem.* (1991).
(4) Cortis, C. M.; Friesner, R. A. *J. Comp. Chem.* **1997**, *13*, 1570−1590.
(5) You, T. J.; Harvey, S. C. *J. Comp. Chem.* **1993**, *14* (4), 484−501.
(6) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 815−820.
(7) Zauhar, R. J.; Morgan, R. S. *J. Comp. Chem.* **1988**, *9* (2), 171−187.
(8) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J. Phys. Chem. B* **1997**, *7*, 1190−1197.
(9) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(10) Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *16*, 3005−3014.
(11) Friesner, R. A.; Rapp, C. S. Loop geometry predictions in proteins: Capabilities and limits of current technology in a study of ribonuclease-a and interleukin-4. Submitted for publication.
(12) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100* (28).
(13) Tomasi, M. P. J. *Chem. Rev.* **1994**, *94*, 2027−2094.