Parametrization of a Generalized Born/Solvent-Accessible Surface Area Model and Applications to the Simulation of Protein Dynamics

Jiang Zhu, Yunyu Shi, and Haiyan Liu*

Key Laboratory of Structural Biology, University of Science and Technology of China (USTC), Chinese Academy of Science (CAS), Hefei, Anhui, 230026, China, and School of Life Science, USTC, Hefei, Anhui, 230026, China

Received: January 10, 2002; In Final Form: March 15, 2002

A generalized Born/surface area (GB/SA) water model (Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014) in combination with the GROMOS96 force field is parametrized with the specific goal of application to protein simulations. The parameters are determined by a combined procedure of fitting to the free energies of solvation of amino acid-mimicking small molecules, rationally considering the effects of the solvent model on hydrogen-bond interactions, and most critically, carrying out long-time trial molecular dynamics simulations on two structurally distinct proteins, the B1 domain of streptococcal protein G and the bovine pancreatic trypsin inhibitor. The refined set of parameters is tested by carrying out a simulation on a third protein, the chymotrysin inhibitor 2. For all three proteins, we report separate 3 ns simulations using, respectively, the re-parametrized GB/SA model, the explicit solvent model, and another implicit solvent model which is mainly based on the solvent-accessible surface area. The simulations with the GB/SA model show excellent agreement with crystal structures and explicit solvent simulations. The root-mean-square error of the final GB/SA model in the free energies of solvation for 16 amino acid analogues is 1.12 kcal/mol. We believe that the procedure of considering structurally distinct proteins in parametrization and running final tests on unrelated proteins is essential for avoiding biasing the model toward stabilizing certain types of protein structures. The results indicate that the model can be widely used in future studies of protein conformations and dynamics.

I. Introduction

Solvent environment has profound influences on the structures, dynamics, and functions of biological molecules. The proper representation of this environment is one of the main challenges in the molecular modeling of biological systems. Explicit solvent models have been extensively used in molecular dynamics (MD) and Monte Carlo (MC) simulations of biomolecules.[1−3] However, high computational expenses associated with the explicit treatment of thousands of solvent degrees of freedom strongly inhibited its applications to a wider range of problems, such as protein folding or large-scale conformational transitions of proteins.[4]

Continuum models of solvation avoid the explicit treatment of solvent molecules. Instead, the solvent is treated as a continuous medium, the applicability of its macroscopic dielectric properties empirically extended to microscopic, atomic-level simulations.[5−15] In continuum models, the free energy of solvation is usually considered as the sum of three terms:

$$\Delta G_{sol} = \Delta G_{cav} + \Delta G_{vdw} + \Delta G_{pol} \quad (1)$$

where ΔG_{cav} is a solvent−solvent cavity term, representing the free energy of creating a solute-sized cavity in the solvent continuum; ΔG_{vdw} accounts for free energy contributions from van der Waals interactions between the solute and solvent; ΔG_{pol} represents free energy contributions from electrostatic interactions between the solute and solvent. The first two terms in eq 1 are often combined into a single term proportional to the solvent-accessible surface area (SASA) of the solute,[5,6,16−18]

$$\Delta G_{cav} + \Delta G_{vdw} = \sum_{i=1}^{N} \sigma_i A_i \quad (2)$$

where A_i is the SASA of atom i and σ_i is an empirical atomic solvation parameter for atom i.

The ΔG_{pol} term results from the polarization of the solvent continuum by the solute charge distribution. A most commonly used approach to obtain ΔG_{pol} is by numerically solving the Poisson−Boltzmann (PB) equation.[19−31] However, the computational costs associated with numerical PB equation solvers usually do not allow for them to be directly used within MD or MC simulations that extensively sample the conformational space of the solute. Indirect combinations of explicit solvent simulations with numerical PB solvers are possible. Indeed, exciting progress has been made in recent years on the calculation of free energies by applying continuum solvent models to solute conformations sampled by separate explicit solvent simulations.[32−40]

An alternative approach to the calculation of ΔG_{pol} is the generalized Born (GB) model proposed by Still and coworkers.[11] In this model, the dielectric component of the free energy of solvation is approximated by the following sum of

* Author to whom correspondence should be addressed at School of Life Science, University of Science and Technology of China, Hefei, Anhui, 230026, China. Telephone/Fax: (+86) 551-360-7451. E-mail: hyliu@ustc.edu.cn.

10.1021/jp020058v CCC: $22.00
Published on Web 04/17/2002

self- and pairwise terms:

$$\Delta G_{pol} = -166.0\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{(r_{ij}^2 + \alpha_{ij}^2 e^{-D_{ij}})^{0.5}} \quad (3)$$

where $\alpha_{ij} = (\alpha_i\alpha_j)^{0.5}$ and $D_{ij} = r_{ij}^2/(2\alpha_{ij})^2$, r_{ij} is the distance between atoms i and j with atomic charges q_i and q_j, respectively, and ϵ is the dielectric constant of the solvent. α_i is the so-called effective Born radius of atom i. The terms with $i = j$ in eq 3 represent the effective Born free energy of solvation of charge i. When $i \neq j$ and r_{ij} is sufficiently large, the denominator in eq 3 approaches r_{ij}, as it should if the solvent was a continuum filling the entire space. For every charged atom i, the effective Born radius α_i (in Å) is related to the effective Born free energy of solvation, $\Delta G'_{pol,i}$ (in kcal/mol), of a reference system through the Born formula

$$\alpha_i = -\frac{166.0}{\Delta G'_{pol,i}}\left(1 - \frac{1}{\epsilon}\right) \quad (4)$$

where the reference system is, by definition, the same as the original solute/continuum solvent system, except that in the reference system atom i bears unit charge and all other atoms are neutral.

Exact calculation of α_i for every charged atom in a macromolecule through numerically solving the PB equation is time-consuming and of little practical use. The potential of GB models in the simulation of macromolecules has come from recent progress in approximation methods to compute α_i with sufficient accuracy and efficiency. An approximate analytical formula proposed by Still et al.[41] forms the basis of several recent parametrizations and applications of the GB model for protein systems.[42−50] In Still's work, the parameters contained in the formula have been optimized to minimize the differences between the effective Born radii calculated by the finite difference Poisson−Boltzmann (FDPB) method and by the formula. A database that consists of numerous organic molecules and biopolymers has been employed in the parametrization.[41] This model has been applied to study the folding of the C-terminal β-hairpin of the GB1 domain.[42] It has also been shown that the model gives accurate free energies of solvation for small organic molecules when combined with the OPLS force field[11,41,43] or the MMFF force field.[44,45] Brooks and coworkers adapted the parameters to the CHARMM united-atom and all-atom force fields.[46−50] They showed that the model can reproduce free energies of solvation computed by FDPB method and can be used in molecular dynamics simulation of proteins.

Besides the approach of Still et al., Hawkins et al. have proposed another approximate pairwise method to compute the effective Born radius analytically.[51,52] This model has been adapted to the AMBER force field[53,54] and applied to study a number of nucleic acids and proteins.[55−58]

In the current work, we investigate some important issues associated with the parametrization of the GB/SA model with the specific goal of applying the model in protein simulations. Starting from the original model of Still et al., a set of parameters suitable for molecular dynamics simulations of proteins with the GROMOS96 force field[59] are developed. The parameters are determined by a combined procedure of (i) fitting to the free energies of solvation of amino acid-mimicking small molecules; (ii) rationally considering the effects of the solvent model on hydrogen-bond interactions; and (iii) most critically, carrying out trial molecular dynamics simulations on two structurally distinct proteins, namely, the B1 domain of streptococcal protein G (the GB1 domain) and the bovine pancreatic trypsin inhibitor (BPTI), respectively. The refined set of parameters is tested by carrying out a simulation on a third protein, the chymotrypsin inhibitor 2 (CI2). We choose such a procedure to avoid biasing the model toward stabilizing certain types of protein structures. For all three proteins, we carried out separate simulations using, respectively, the re-parametrized GB/SA model, the explicit water model, and an implicit solvent model based on the solvent-accessible surface areas and the distance dependent dielectric constant. We compare key structural and dynamics properties of the proteins in these simulations.

II. Materials and Methods

II.A. Theory. Following Still et al., we use the following analytical function to compute the effective Born free energy of solvation of the reference system:[41]

$$\Delta G'_{pol,i} = \left(1 - \frac{1}{\epsilon}\right)\left(\frac{-166.0}{R_{vdw-i} + \phi + P_1} + \sum^{stretch}\frac{P_2 V_j}{r_{ij}^4} + \sum^{bend}\frac{P_3 V_j}{r_{ij}^4} + \sum^{non\text{-}bonded}\frac{P_4 V_j \,CCF}{r_{ij}^4}\right) \quad (5)$$

where r_{ij} is the distance between atoms i and j, R_{vdw-i} is the van der Waals radius of atom i, φ the dielectric offset, and V_j is the volume of atom j. In the above formula lengths are in Å and energies are in kcal/mol. In the reference system, each atom j excludes a small volume of solvent continuum from the environment of atom i. The first term in eq 5 corresponds to the usual Born solvation contribution and P_1 is a single-atom scaling factor. The r_{ij}^{-4} dependency in the remaining terms comes from the reasoning that the contribution of a small volume of solvent dielectric continuum to the free energy of solvation is proportional to V/r_{ij}^4, where V is the volume.[60] The parameters P_2-P_4 and the function CCF empirically account for the overlaps between solvent volumes excluded by different atoms, P_2 being the stretch scaling factor, P_3 the bend scaling factor, P_4 the nonbonded scaling factor, and CCF a close contact function for nonbonded interactions containing an additional soft cutoff parameter P_5.[41] In Still's original model,[41] the close contact function CCF is employed to reduce the effective volumes of overlapping atoms when a nonbonded atom pair come within a certain portion of the sum of their van der Waals radii:

$$CCF = 1.0, \; if\left(\frac{r_{ij}}{R_{vdw-i} + R_{vdw-j}}\right)^2 > \frac{1}{P_5}$$

otherwise

$$CCF = \left\{0.5\left[1.0 - \cos\left\{\left(\frac{r_{ij}}{R_{vdw-i} + R_{vdw-j}}\right)^2 p_5\pi\right\}\right]\right\}^2 \quad (6)$$

It is worth noting that the first three terms in eq 5 vary little among different solute conformations. Therefore, they can be treated as constants and calculated at the beginning of a dynamics simulation.

II.B. Data Sets for Parametrization. Throughout our parametrization process, the free energies of solvation of 23 small molecules[61−65] (Table 1) have been used as criteria to screen the initial parameter sets. Only those parameter sets that lead to reasonably small errors in these free energies were further

TABLE 1: Free Energies of Solvation of Amino Acid-Mimicking Small Molecules[a]

molecules	residue	ΔG_{exp}	ΔG_{calc}	error
	Neutral Amino Acids			
methane	Ala	1.93	0.67	−1.26
propane	Val	1.98	1.01	−0.97
isobutene	Leu	2.28	1.13	−1.15
butane	Ile	2.15	1.14	−1.01
toluene	Phe	−0.89	−0.52	0.37
4-cresol	Tyr	−6.13	−5.21	0.92
3-methylindole	Trp	−5.91	−4.19	1.72
methanol	Ser	−5.11	−5.21	−0.10
ethanol	Thr	−5.01	−4.59	0.42
methanethiol	Cys	−1.24	0.02	1.26
methylethyl sulfide	Met	−1.49	0.91	2.40
acetic acid	Asp	−6.70	−5.30	1.40
propionic acid	Glu	−6.47	−4.96	1.51
acetamide	Asn	−9.71	−8.42	1.29
propionamide	Gln	−9.41	−8.13	1.28
N-butylamine	Lys	−4.29	−3.09	1.20
N-propylguanidine	Arg	−10.91	−12.12	−1.21
4-methylimidazole	His	−10.25	−9.30	0.95
trans-*N*-methylacetamide	bb	−10.08	−4.73	5.35
				1.12
	Ionized Amino Acids			
acetate ion	Asp (−)	−80.65	−84.29	−3.64
propionate ion	Glu (−)	−79.16	−82.97	−3.81
N-butylammonium	Lys (+)	−69.24	−83.61	−14.37
N-*p*-guanidinium	Arg (+)	−66.07	−63.60	−2.47

[a] Free energies and errors are in kcal/mol. The free energies presented here are calculated using the GB/SA model with final set of parameters.

investigated by trial protein simulations. The 23 molecules, 19 neutral and 4 ionized, include analogous compounds of amino acid backbone and side chains. For consistency, the geometries of these molecules have been optimized with the GROMOS96 force field.[59]

Trial simulations of two proteins (the GB1 domain and BPTI) have been used to guide the parametrization process. These two proteins are rather distinct in their primary, secondary, and tertiary structures. The GB1 domain is a 56-mer with more than 66% secondary structure contents and no disulfide bond. It forms a tight, well-packed structure in solution, with a four-stranded β-sheet and a helix. BPTI is a 64-mer with 45% secondary structure contents and three disulfide bonds. Its secondary structures are short helices. The final set of parameters has been applied to simulate CI2, a 65-mer peptide with medium contents of secondary structures (46%) and with no disulfide bond.

II.C. Explicit Solvent Simulations. Explicit solvent simulations of the above three proteins have been carried out for comparisons with implicit solvent models. The GROMOS96 condensed-phase force field[59] and the SPC water model[66] have been used. The protonation states of these proteins have been modeled for neutral pH, counterions added to neutralize the system and truncated octahedron periodic boundary conditions chosen with a solvent shell of 0.8 nm. Nonbonded interactions have been treated using a twin-range cutoff method with generalized reaction field corrections,[67] short-range interactions within 0.8 nm evaluated every time step, medium-range interactions between 0.8 and 1.4 nm updated every 10 time steps, and corrections corresponding to electrostatic interactions beyond 1.4 nm approximated by the reaction field generated by a dielectric continuum with a dielectric constant of 54. A time step of 2 fs was used in the simulations, covalent bond lengths constrained by SHAKE[68] with a tolerance of 10^{-4}. The temperature and pressure were kept constant by weak coupling[69] to external heat and isotropic pressure baths, the relaxation times being 0.1 ps and 0.5 ps, respectively. The following protocol has been used to generate equilibrated protein−solvent systems as starting points of the sampling simulations. After energy minimization, a series of simulations have been performed at 50 K (20 ps), 100 K (20 ps), 150 K (20 ps), 200 K (20 ps), 250 K (20 ps), and 300 K (50 ps), respectively. The positions of all non-hydrogen protein atoms were restrained in the simulations, with successively decreasing restraining force constants. The sampling simulations are 3 ns MD simulations at 300 K with no restraint. Coordinates and energies have been recorded every 0.5 ps during the sampling.

II.D. GB/SA Implicit Solvent Simulations. Stochastic dynamics[70] (SD) simulations have been used for the implicit solvent simulations. The GB/SA model has been used in conjunction with the GROMOS96 condensed-phase force field.[59] Unless specified below, the setups have been the same as the explicit solvent simulations. The nonbonded interactions were calculated without cutoff. Considering that the SASA calculation is time-consuming but the contributions of the SASA terms to the free energies are relatively small, the atomic SASAs and their derivatives have been updated every 10 time steps. The starting structures of the implicit solvent simulations were the equilibrated structures from the explicit simulations followed by 20 ps implicit solvent simulations using the corresponding solvent model, during which the positions of non-hydrogen atoms have been restrained. The sampling simulations are 3 ns SD simulations with no restraint.

II.E. The Implicit Solvent Model Based on the Solvent-Accessible Surface Area and Distance-Dependent Dielectric Constant (RD/SA). We have previously parametrized another implicit solvent model for using with the GROMOS96 vacuum force field. It is based on a simple, analytical method proposed by Hasel et al.[71] to compute approximate solvent-accessible surface areas and an earlier parametrization within the GROMOS87 force field by Fraternali et al.[72] We found that a distance-dependent dielectric constant ($\epsilon = r/\text{Å}$) approximating the long-range electrostatic screening effects significantly improves the performance of the model. We termed this model RD/SA. The parametrization of this model (results not shown) have been based on six proteins of different sizes, and its capability to discriminate native and miss-folded protein structures (results not shown) has been proven to be comparable to a solvent-exclusion model proposed by Lazaridis.[73,74] The RD/SA model has been applied in conjunction with the GROMOS96 vacuum force field[59] (zero total charges assigned to ionized side chains) to simulate the three proteins employed in this work. The setups of these simulations have been the same as the GB/SA simulations.

III. Results and Discussions

III.A. Atomic van der Waals Radii in the GB/SA Model. To be consistent with the GROMOS96 force field, we chose to use the same van der Waals radii as in the GROMOS96 force field for most of the atom types. Exceptions are some of the oxygen, nitrogen, and polar hydrogen atom types (see Table 2). In the calculation of the effective Born radii each atom must be associated with a single van der Waals radius. However, for some of these atom types, multiple values of van der Waals radius may be derived from the GROMOS96 force field in which an atom can be associated with different van der Waals parameters depending on the other atom involved in the van der Waals interaction. After trial and error (results not shown), we decided to use the same van der Waals radius of 1.477 Å for all oxygen atom types. This is the van der Waals radius of hydroxyl oxygen derived from the force field. For all nitrogen atom types, we decided to use the same van der Waals radius

TABLE 2: van der Waals Radii and Other Parameters in the GB/SA Model[a]

type code	atom type	description	R_{vdw}	σ_i
1	O	carbonyl oxygen (C═O)	1.477	0.005
2	OM	carboxyl oxygen (C═O)	1.477	0.005
3	OA	hydroxyl, sugar, or ester oxygen	1.477	0.005
4	OW	water oxygen		
5	N	peptide nitrogen (NH)	1.540	0.005
6	NT	terminal nitrogen (NH_2)	1.540	0.005
7	NL	terminal nitrogen (NH_3)	1.540	0.005
8	NR	aromatic nitrogen	1.540	0.005
9	NZ	arginine NH (NH_2)	1.540	0.005
10	NE	arginine NH (NH)	1.540	0.005
11	C	bare carbon	1.681	0.005
12	CH_1	aliphatic or sugar CH-group	1.900	0.005
13	CH_2	aliphatic or sugar CH_2-group	1.960	0.005
14	CH_3	aliphatic CH_3-group	1.937	0.005
15	CH_4	methane	1.855	0.005
16	CR1	aromatic CH-group	1.871	0.005
17	HC	hydrogen bound to carbon	1.187	0.005
18	H	hydrogen not bound to carbon	1.187	0.000
19	DUM	dummy atom		
20	S	sulfur	1.654	−0.005

P_1-P_5 parameters in the current GB/SA model	
P_1	0.0073
P_2	0.921
P_3	6.211
P_4	15.236
P_5	1.130

[a] The van der Waals radii are in Å, the σ_i is in kcal/mol Å^2. The parameters presented here are used in the re-parametrized GB/SA model, in which the backbone hydrogen atoms are "merged" into the bonded nitrogen atoms.

of 1.54 Å. There is no van der Waals interaction associated with hydrogen atoms of type code 18 (polar hydrogen atoms) in the GROMOS96 force field. For this hydrogen type we decided to use the same van der Waals radius as that of hydrogen atoms with type code 17.

III.B. The Solvent-Accessible Surface Area Term. For the SASA term, the coefficient σ ranges from 0.005 kcal/mol Å^2 to 0.010 kcal/mol Å^2 in various previous implementations of the GB/SA model. The contribution of this term to the free energy of solvation is relatively small. We decided to use $\sigma =$ 0.005 kcal/mol Å^2 for all atom types except for the sulfur atom type ($\sigma = -0.005$ kcal/mol Å^2) and hydrogen atoms of type code 18 ($\sigma = 0$).

III.C. Effective Born Radii of Backbone Hydrogen Atoms. For a macromolecule force field to be successful to any extent, it must model hydrogen bonds properly. Special care has been taken in the development of modern protein force fields so that hydrogen bonds, especially backbone−backbone hydrogen bonds, are represented properly. In GROMOS this comes from balanced interatomic electrostatic and van der Waals interactions. It is important that this balance is still maintained in combination with the GB/SA model. We compared two options in the GB/SA parametrization. In the first option (the "ordinary" backbone hydrogen option), the effective Born radius of any backbone hydrogen atom is calculated using the hydrogen atom's own van der Waals radius and environment. In the second option (the "merged" backbone hydrogen option), the effective Born radius of any backbone hydrogen atom is taken to be the same as that of the corresponding backbone nitrogen atom. The "merged" backbone hydrogen option proved to outperform the "ordinary" backbone hydrogen option in most of our trial simulations.

III.D. The Determination of P_5. In the formulations of eqs 5 and 6, P_5 is a critical parameter that determines how the GB model will influence the interactions between nonbonded atoms in close contact with each other. For small molecules, their free energies of solvation are rather insensitive to changes in P_5 because they lack such interactions. For macromolecules, however, the choice of P_5 would be crucial because such interactions contribute substantially to their conformational stability and dynamics. It has been previously noted that the GB solvation energies may contain systematic errors for large systems, and FDPB calculations on macromolecules have been utilized to refit the entire parameter set.[50]

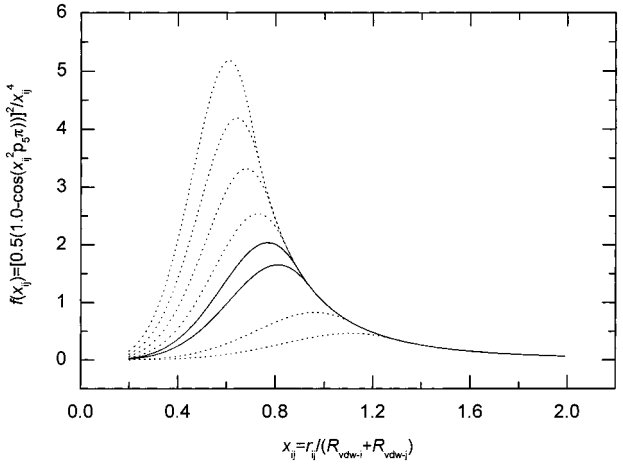

Figure 1. The function $f(x_{ij})$ ($f(x_{ij}) = [0.5(1.0 - \cos(x_{ij}^2 \cdot P_5 \cdot p))]^2/x_{ij}^4$, $x_{ij} = r_{ij}/(r_{vdw-i} + r_{vdw-j})$) with various values of P_5 factor. For dotted lines, the P_5 factor is set to 0.6, 0.8, 1.4, 1.6, 1.8, and 2.0 from below to above. For the solid lines, the P_5 factor is set to 1.13 (final parameter) and 1.254 (Still's original parameter) from below to above.

In our work, the P_1-P_4 factors are fixed at the same values as in Still's model. This choice together with our choices of atomic van der Waals radii and σ lead to reasonable free energies of solvation for 23 amino acid-mimicking small molecules (Table 1). P_5, however, remains poorly determined.

To understand how different choices of P_5 may influence the interatomic nonbonded interactions, we rewrite the last term in eq 5 as $f(x_{ij})$ (the distance-independent pre-factor and P_4 left out):

$$f(x_{ij}) = \frac{[0.5(1.0 - \cos x_{ij}^2 P_5 \pi)]^2}{x_{ij}^4}$$

where

$$x_{ij} = \frac{r_{ij}}{R_{vdw-i} + R_{vdw-j}} \quad (7)$$

This function with different choices of P_5 is plotted in Figure 1. The function corresponds to the effect of the presence of atom j on $\Delta G'_{pol,i}$ if there were only two atoms. It first increases when the distance between atoms i and j decreases, then reaches a maximum, and decreases afterward. The effective Born radius of atom i varies in the same manner. This is consistent with the intuitive definition of the effective Born radius, which is the distance from the atomic center to the molecular dielectric boundary. When the atomic dielectric boundary of atoms i and j are in contact, the molecular dielectric boundary is the farthest from the atomic center of i and, as a result, the effective Born radius reaches its maximum value. Sensible choice of P_5 should allow the CCF function to start switching off the V/r^4 contribution at the position where the atomic dielectric boundary overlaps. Still et al. proposed an empirical offset of −0.009 Å

from the atomic radii amended by the P_1 factor to define the atomic dielectric boundary.[11,41]

Considering the effects of the model on balanced hydrogen-bond interactions, we believe that the importance of intramolecular hydrogen bonds must also be taken into account in the determination of P_5. The influence of different choices of P_5 on the stability of hydrogen bonds in a molecular dynamics simulation can be rationalized as follows. The CCF term should mimic the solvent-induced barrier in hydrogen-bonding interactions, the CCF term plus the corresponding solute/solute interactions giving the potential of mean force for hydrogen-bond formation in water. If P_5 is too large, some equilibrium hydrogen-acceptor distances would fall on the right side of the maximum of $f(x_{ij})$, the GB term would not give the correct solvent-induced hydrogen-bond barrier on the potential of mean force profile. Instead, the barriers for existing hydrogen bonds to break would be decreased. On the other hand, a too small P_5 would result in too small solvation effects in the hydrogen-bonding interactions. A reasonable assumption about P_5 is that the maximum equilibrium hydrogen-bonding distance measured by the distance between the hydrogen and acceptor atom coincides with the distance at which the function $f(x_{ij})$ reaches its maximum. That is,

$$P_5 = \left(\frac{R_{\text{vdw-receptor}} + R_{\text{vdw-H}}}{r_{\text{receptor-H}}}\right)^2 \tag{8}$$

On the basis of observed backbone−backbone hydrogen-bonding distances in proteins, we use a value of 2.5 Å as the maximum hydrogen-bonding distance and computed that a value of 1.13 should be used for P_5 with our van der Waals parameters. This rational choice of P_5 has proven to produce the best results among our trial simulations. The suggestion of using eq 8 to determine P_5 remains to be justified in the parametrization of GB models for other force fields.

III.E. Free Energies of Solvation of Small Molecules. The free energies of solvation for the 23 small molecules are listed in Table 1. The root-mean-square (RMS) error given by the refined set of GB/SA parameters is 1.12 kcal/mol for 16 neutral molecules not including those mimicking the peptide backbone and the side chains of cysteine and methionine, respectively. In a recent study on the free energy of solvation for neutral analogues of amino acid side chains using the GROMOS force field, anomalous behavior has also been observed for sulfur-containing molecules: the calculation gave positive free energies of hydration whereas the experimental results are negative.[75] This may result from the facts that the force field is not polarizable, and sulfur atoms bear relatively small point charges (−0.064 *e* in protonated cysteine and zero in methionine). To partially compensate for this deficiency we assigned a negative σ_i of −0.005 *kcal/molÅ²* to sulfur atoms to model their weak hydrophilicity. This treatment, however, has little influence on the simulated properties of proteins. The relatively poor agreement between the calculated and experimental free energies for the peptide backbone-mimicking compound cannot be improved without changing the charge distribution specified by the force field or considering polarizable models. Although merging the backbone polar hydrogen atoms with their bonded nitrogen atoms in calculating their effective Born radius slightly worsens the calculated free energy of solvation for the small molecule, the protein simulations strongly benefit from this treatment (see below). For the solvation of the ionized species, the errors given by the refined GB/SA parameter set are reasonable for molecules mimicking aspartic acid, glutamic acid, and arginine, respectively, but much larger (−14.37 kcal/mol) for the molecule mimicking lysine. In general the results for the solvation free energies of small molecules indicate that the GB/SA model with the refined set of parameters can be used with GROMOS96 force field in protein simulations.

III.F. Structural Deviations in Protein Simulations. Figure 2 shows the root-mean-square (RMS) positional deviations from the crystal structures as functions of simulation time for backbone atoms involved in secondary structures. For the GB1 domain and BPTI the RMS deviations in several trial GB/SA simulations using different parameter sets are also shown.

Figure 2 demonstrates that for both the GB1 domain and BPTI, the GB/SA simulations using the final parameter set are more stable than the other sets, giving consistently smaller RMS deviations. The original GB/SA model cannot be used directly with the GROMOS96 force field, resulting in too large RMS deviations (reaches 0.35 nm in 3 ns) and partial unfolding of the GB1 domain. This probably reflects the imbalance in hydrogen-bond interactions in the direct combination of the two models. The GB1 domain, which is richer in secondary structure, seems to be more prone to such imbalance than BPTI and its RMS deviations significantly larger, although the native structure of BPTI is also unstable under the model. Using the "merged" backbone hydrogen atom option without fixing P_5 significantly enhances the structural stability of the GB1 domain, presumably because this treatment gives more balanced consideration of solvent screening on different atoms involved in the same hydrogen bond than the "ordinary" backbone hydrogen atom option. The effects of the "merged" backbone hydrogen atom treatment are obvious in the simulation of the GB1 domain, in which the RMS deviation converges to around 0.2 nm. However, the RMS deviations for the secondary-structure-lacking BPTI still increased gradually through out the first 2 ns of the simulation. The RMS deviations indicate that the final parameter set with both the "merged" backbone hydrogen option and a modified P_5 probably gives the best-balanced hydrogen-bonding interactions.

It is interesting to compare the GB/SA simulations with the RD/SA simulations using vacuum force field. Although the RMS deviations in the RD/SA simulations converge more rapidly, they generally correspond to much larger structural distortions (Figure 2).

It is worth noting that the performance of the refined GB/SA parameter set in the simulation of CI2 is as good as in the simulations of the GB1 domain and BPTI, while only the latter two have been considered by extensive trial simulations (not limited to those reported in Figure 2). We also note that in comparison with GB/SA simulations of proteins reported in the literature, our simulations are significantly longer. Simulations of such length may be necessary: in a considerable portion of our trial simulations, significant structural distortions are observed only after more than a few hundred picoseconds to nanosecond of simulation time.

III.G. Structural Fluctuations in Protein Simulations. We examined the C_α positional RMS fluctuations of the three proteins in the explicit solvent simulations, the RD/SA simulations and the GB/SA simulations (Figure 3). As references, the "experimental" fluctuations have also been estimated using the crystallographic isotropic B-factors.

For all three proteins, the positional fluctuations produced by the GB/SA simulations are at least in as close agreement with the crystallographic B-factors as the explicit solvent simulations. For the GB1 domain, the absolute fluctuations estimated using the B-factors are much smaller than all the simulations. The RMS fluctuations produced by the final GB/SA model are similar to that generated by the explicit solvent simulation but smaller than that by the RD/SA model. For BPTI, the RMS fluctuations generated by the final GB/SA model are

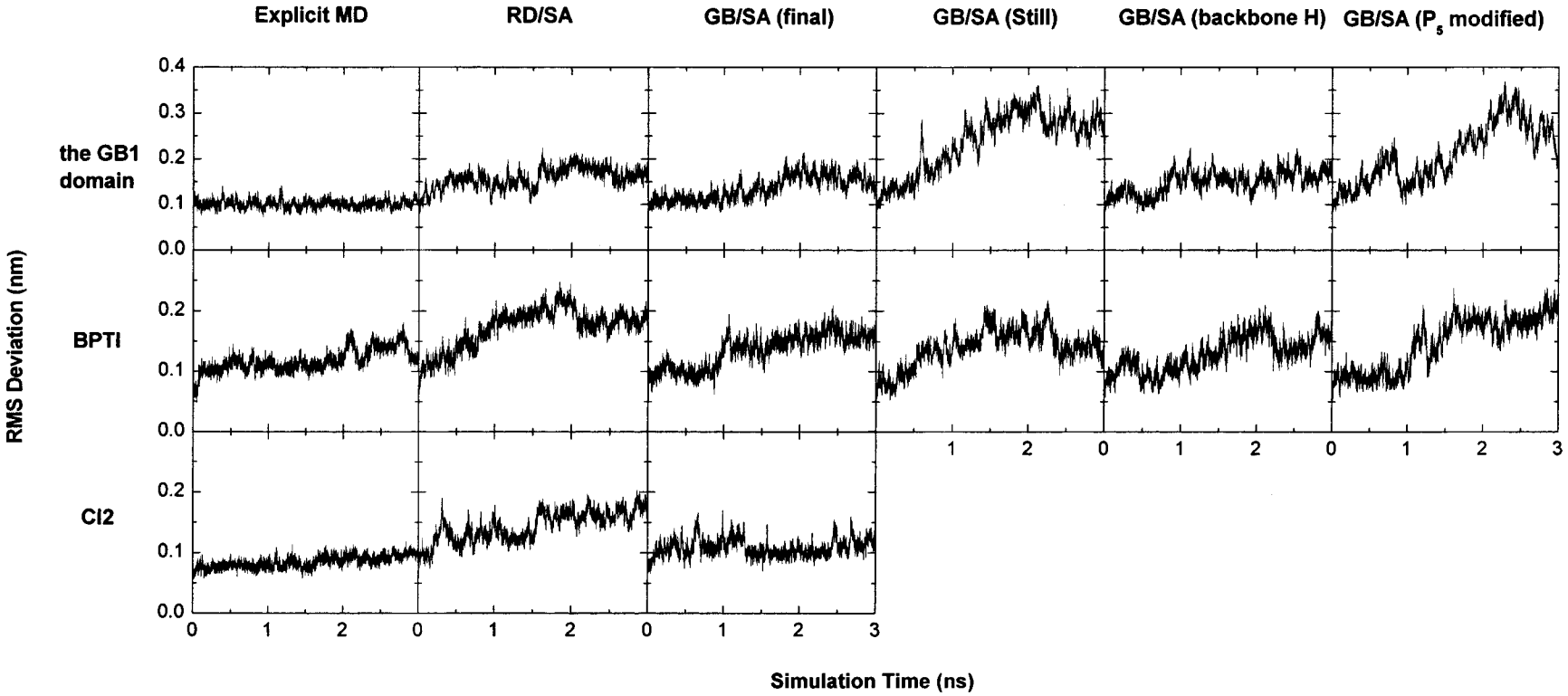

Figure 2. The root-mean-square (rms) deviations from crystal structures as functions of simulation time. From left to right: explicit water model, RD/SA model, GB/SA models using the final parameters, Still's original parameters, Still's original parameters with backbone hydrogen atoms merged with nitrogen atoms (see text), and P_5 set to 1.13 without merging the hydrogen atoms. For CI2 there is no trial GB/SA simulation data. Only C_α atoms of residues contained in secondary structures are considered.

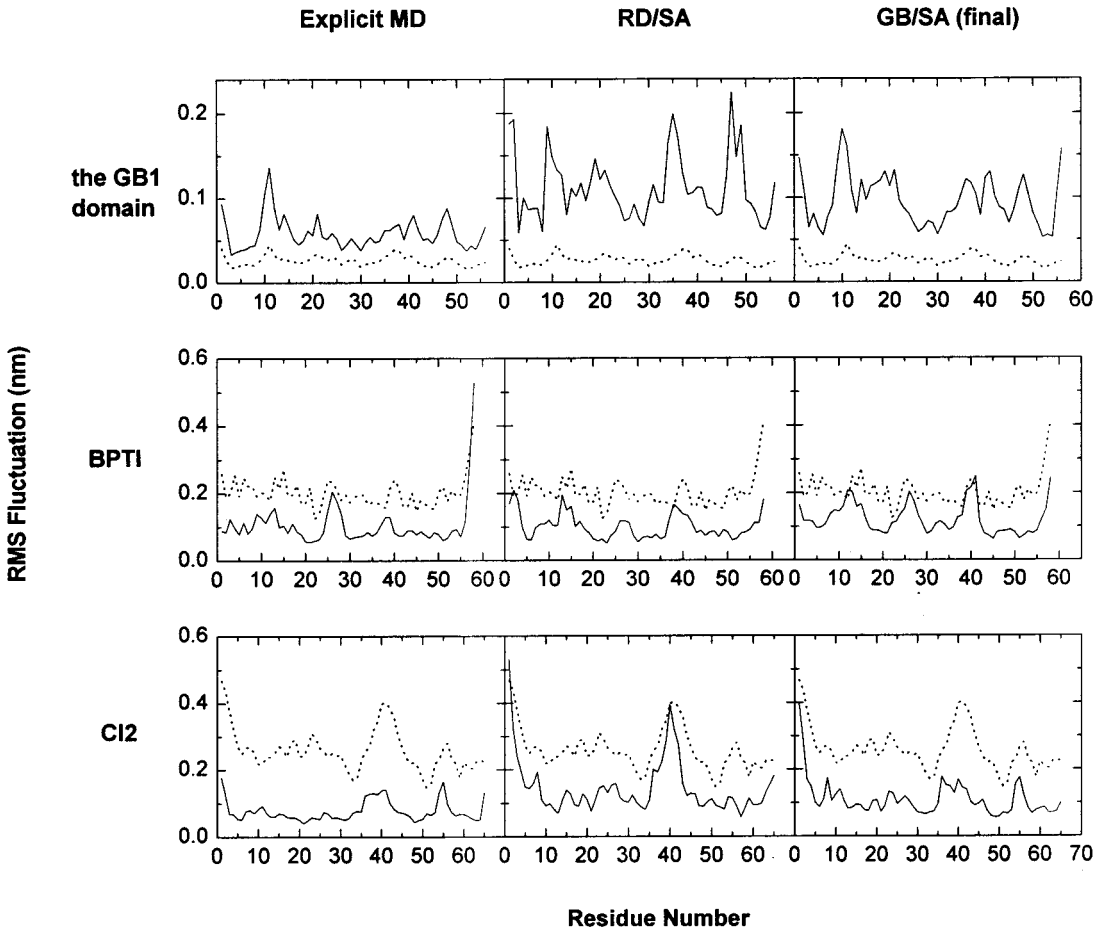

Figure 3. Root-mean-square fluctuations of C_α atom positions from simulations using different models. The simulations are 3 ns in length. From left to right: explicit water model, RD/SA model, and final GB/SA model. The positional fluctuations estimated from the isotropic B-factors are shown as references (dotted lines).

in the closest agreement with the B-factors, especially for peptide regions around the N and C terminals and residues 13, 26, and 40. For CI2, the GB/SA simulation also leads to the closest agreement with the B-factors. For all three proteins, the positional fluctuations produced by the RD/SA model seem to be the least reliable. For example, while the B-factors and both the GB/SA and the explicit water models indicate larger fluctuations around residue 55 of CI2, the RD/SA simulation indicates the contrary.

III.H. Structural Compactness and Surface Areas in Protein Simulations. Good indicators of inappropriate treatment of solvation in a protein simulation include parameters measuring the overall compactness and exposed surface areas of the protein. Figures 4, 5, and 6 show the radius of gyration (R_g), hydrophobic SASA (S_{pho}), and hydrophilic SASA (S_{phi}) of the three proteins as functions of the simulation time.

The explicit solvent simulation results are, in general, consistent with the crystal structures. There are, however, some small but obvious discrepancies. The R_gs from the simulations are somewhat smaller than that calculated from the crystal structures. Especially for CI2, the explicit solvent simulation resulted in structures more compact than the crystal structure. Also for CI2, its exposed hydrophobic surface area tends to be larger in the explicit solvent simulation than in the crystal structure. And for the GB1 domain, its hydrophilic surface area in the explicit solvent simulation tends to be smaller than in the crystal structure (Figure 6).

In general, the reparametrized GB/SA model performs as well as the explicit solvent model in reproducing the radii of gyration and the total hydrophobic/hydrophilic areas of the three proteins. For the GB1 domain and CI2, the R_g values produced by the GB/SA simulations are reasonable and in better agreement with the crystal structures than both the explicit solvent simulations and the RD/SA simulations. Although the R_g of BPTI generated by the GB/SA simulation are slightly smaller (by about 0.05 nm) than the crystal structure, it is far better than that produced by the RD/SA implicit solvent simulation and reaches equilibrium after approximate 1 ns. Both the crystal structure values of S_{pho} and S_{phi} for the GB1 domain and BPTI are reproduced by the 3 ns GB/SA simulations. For CI2, its S_{pho} and S_{phi} in the GB/SA simulation show small deviations from the crystal structure values. These deviations, however, have the same magnitudes as the deviations produced by the explicit water simulation.

The RD/SA simulations led to the largest deviations from the crystal structures. Lacking electrostatic repulsion between oppositely charged side chains and crudely representing the screening effects of the solvent, the RD/SA model tends to generate structures too compact. The R_g values produced by the RD/SA simulations are much smaller than the crystal structures and the other simulations. For example, the R_g of BPTI decreased from 1.11 nm to 0.975 nm in 3 ns RD/SA simulation. The S_{pho} values in the RD/SA simulations are consistently larger than the explicit solvent simulations for all three proteins. They are also larger than the corresponding crystal structure values for the GB1 domain and BPTI. The RD/SA model also caused S_{phi} values to decrease significantly from their crystal structure values in the simulations. Physically, the surface-area-propor-

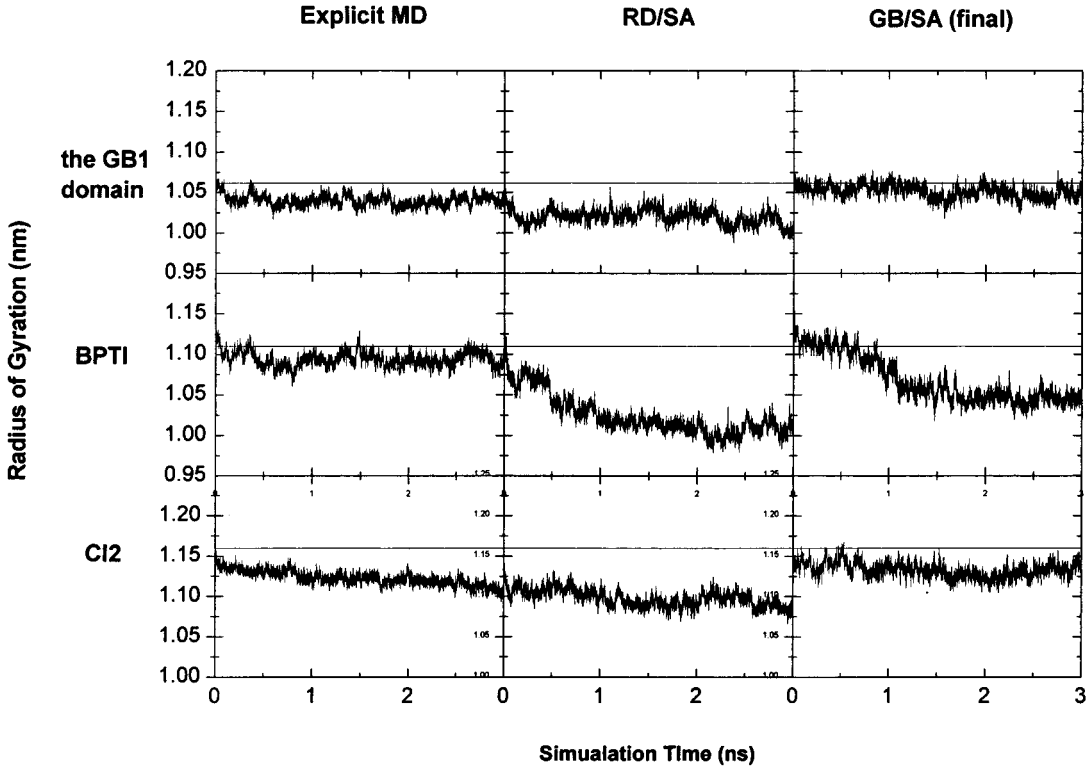

Figure 4. The radius of gyration (R_g) as functions of simulation time. From left to right: simulations using explicit water model, RD/SA model, and final GB/SA model. R_g of the crystallographic structures are shown as references (straight lines).

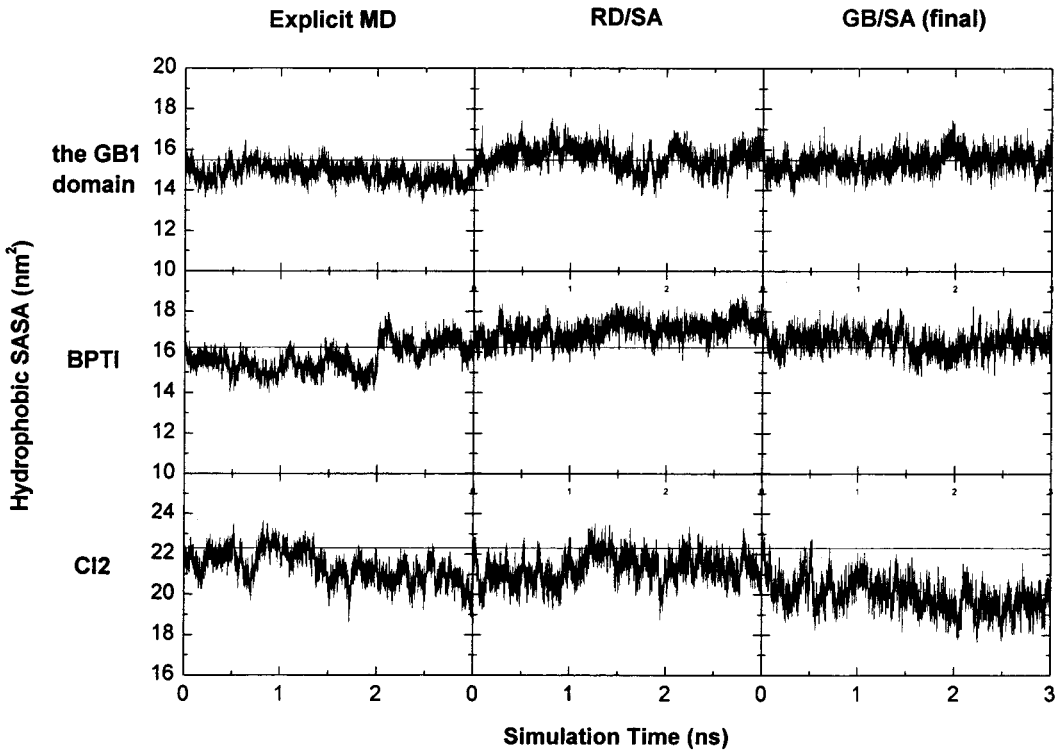

Figure 5. The total hydrophobic solvent-accessible surface areas (S_{pho}) as functions of simulation time. From left to right: the simulations using explicit water model, RD/SA model, and final GB/SA model. S_{pho} of the crystallographic structures are shown as references (straight lines).

tional hydrophilic term in the RD/SA model is a poor representation of the hydrophilic effects. It seems that within such an approximation it is difficult to obtain satisfactory results on different overall structural parameters simultaneously.

III.I. Comparisons between the GB Results and the Numerical PB Results. The GB model is an approximation to the continuum dielectric theory, which is usually quantified via PB calculations. Although in this particular work we have not optimized the GB model specifically to reproduce PB results, there exist strong correlations between the PB results and our GB model results for the solvation free energies of proteins. We compared the free energies of solvation calculated by the GB model and by the fast adaptive multi-grid boundary element (FAMBE) method.[32,33] The calculations have been performed

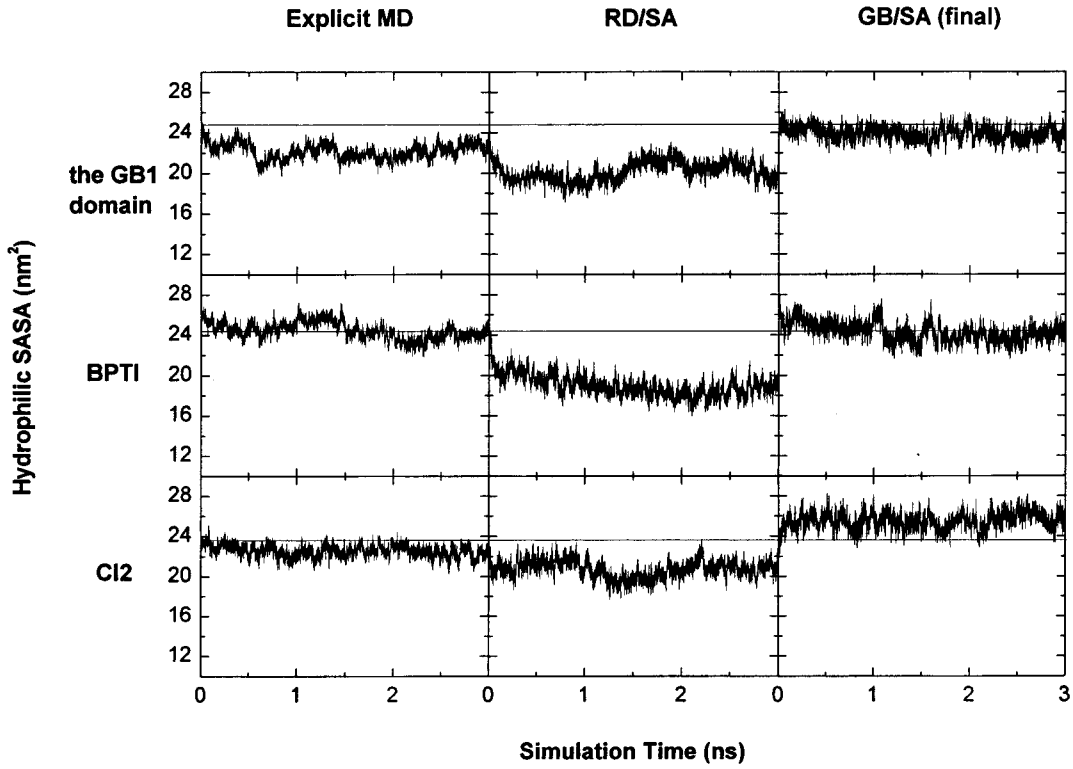

Figure 6. The total hydrophilic solvent-accessible surface areas (S_{phi}) as functions of simulation time. From left to right: the simulations using explicit water model, RD/SA model, and final GB/SA model. S_{phi} of the crystallographic structures are shown as references (straight lines).

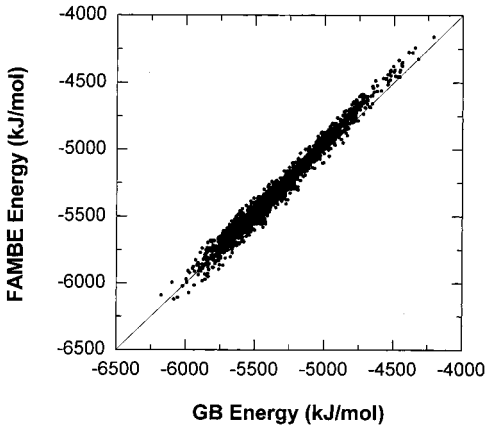

Figure 7. Correlation between the free energies of solvation calculated by the GB model and a numerical PB solver (FAMBE) for the GB1 domain, BPTI and CI2. Calculations are performed on 1800 conformers (600 for each of the three proteins) evenly extracted from the GB/SA simulation. The straight line denotes the case in which the GB free energy is equal to the FAMBE free energy.

on 1800 conformations (600 for each of the three proteins studied in this work) evenly extracted from the GB/SA simulations. The van der Waals radii and atomic charges used in the FAMBE calculation are the same as in the GB/SA simulation. The results are shown in Figure 7 and the correlation between the results of the two models is 0.99.

III.J. Computational Cost. Without using cutoff, pairlist, or other time-saving techniques, the computing time of the GB/SA model is half of that of explicit solvent simulation (with cutoff) for small proteins such as BPTI or CI2. For larger proteins, such as the dihydrofolate reductase (DHFR), the computational cost of the GB/SA model can be as much as two times of that of explicit solvent simulation. However, our initial results (not shown) indicated that by introducing cutoff, pairlist, and other time-saving techniques, the computational cost of GB/SA can be significantly reduced, especially for larger proteins.

IV. Conclusions

In this work we have re-parametrized the GB/SA model of Still and co-workers for the study of protein structures and dynamics using the GROMOS96 force field. We discussed the importance of maintaining the proper balance in nonbonded interactions, especially hydrogen-bond interactions in the development of implicit solvent models, and showed that rational choices of parameters may be derived from such considerations. Extensive trial simulations have been carried out on two structurally distinct proteins (the GB1 domain and BPTI) during parametrization and a simulation on a third protein, CI2, has been reported using the refined parameter set as a test.

For all three proteins, the GB/SA implicit solvent simulations with the refined set of parameters produced results comparable to explicit solvent simulations. Various key structural and dynamics parameters have been considered in the comparison. The GB/SA model significantly outperforms another implicit solvent model that uses solvent-accessible surface area proportional solvation terms. Although the three proteins used in our parametrization contain only around 60 residues, we found that the GB/SA model can also give satisfying results for larger proteins. For example, in a 1 ns GB/SA simulation of the dihydrofolate reductase (DHFR, 159 residues), the average RMS deviation of the last 500 ps with respect to the crystal structure is 0.149 nm versus 0.124 nm in the explicit solvent simulation. The RMS deviation of the average conformation of the last 500 ps with respect to the crystal structure is 0.131 nm versus 0.108 nm in the explicit solvent simulation. The RMS deviations have been calculated using the backbone atoms of residues contained in secondary structures.

Although our main focus has been on the development of a model which can, under certain circumstances, replace explicit

solvent in long time dynamics simulations of proteins, we have shown that the experimental free energies of solvation of amino acid-mimicking small molecules can be reproduced reasonably well. For small proteins the numerically calculated PB free energies of solvation can also be reproduced with reasonable accuracy. There are still some limitations in the current model. Without modifying the solute force field, the free energies of solvation calculated for certain small molecules show large deviations from the experimental values. Protein-specific errors in radius of gyration and exposed surface areas, although small and comparable in magnitude to explicit solvent models, still exist in long time simulations. Our results highlight the importance of using structurally distinct proteins, long trial simulations and multiple criteria in the parametrization of a solvent model. We expect the model parametrized here to have wide applications in future theoretical studies of proteins.

Acknowledgment. H. Y. Liu acknowledges financial support from the Chinese National Natural Science Foundation (Grant No. 30025013). Y. Y. Shi acknowledges support from the National Basic Research Projects (Grant No. G1999075605) and the Chinese National Natural Science Foundation (Grant No. 39990600). We gratefully thank Professor WF van Gunsteren of the Department of Chemistry, ETH-Zurich, for providing the GROMOS96 package.

References and Notes

(1) Jorgensen, W. L.; Ravimohan, C. J. *J. Chem. Phys.* **1985**, *83*, 3050.
(2) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1970**, *101*, 1913.
(3) DeBolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316.
(4) Duan, Y.; Wang, L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897.
(5) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
(6) Ooi, T.; Oobatake, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.
(7) Kang, Y. K.; Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4105, 4109, 4118.
(8) Warshel, A.; Russel, S. T. *Rev. Biophys.* **1984**, *17*, 283.
(9) Gilson, M.; Honig, B. *Proteins* **1988**, *4*, 7.
(10) Honig, B.; Sharp, K.; Yang, A. S. *J. Phys. Chem.* **1993**, *97*, 1101.
(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(12) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.
(13) Tomasi, J.; Bonaccorsi, R.; Cammi, R.; Olivares del Valle, F. J. *J. Mol. Struct. (THEOCHEM)* **1991**, *234*, 401.
(14) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(15) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(16) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
(17) Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. *J. Phys. Chem.* **1975**, *72*, 2239.
(18) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.
(19) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.
(20) Luty, B. A.; Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1992**, *13*, 768.
(21) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(22) Gilson, M.; Rashin, A.; Fine, R.; Honig, B. *J. Mol. Biol.* **1985**, *88*, 503.
(23) Klapper, I.; Hangstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47.
(24) Gilson, M.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327.
(25) Sharp, K.; Jean-Charles, A.; Honig, B. *J. Phys. Chem.* **1992**, *96*, 3822.
(26) Sitkoff, D.; Ben-Tal, N.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 2744.
(27) Rashin, A. A. *J. Phys. Chem.* **1989**, *93*, 4664.
(28) Juffer, A. H.; Botta, E. F. F.; van Keulen, B. A. M.; van der Plog, A.; Berendsen, H. J. C. *J. Comput. Phys.* **1991**, *97*, 144.
(29) You, T. J.; Harvey, S. C. *J. Comput. Chem.* **1993**, *14*, 484.
(30) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 815.
(31) Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1988**, *9*, 171.
(32) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. *Proteins* **1998**, *32*, 399.
(33) Vorobjev, Y. N.; Hermans, J. *Biophys. Chem.* **1999**, *78*, 195.
(34) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
(35) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330.
(36) Lee, M. R.; Duan, Y.; Kollman, P. A. *Proteins* **2000**, *39*, 309.
(37) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567.
(38) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265.
(39) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671.
(40) Tsui, V.; Case, D. A. *J. Phys. Chem. B* **2001**, *105*, 11314.
(41) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(42) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151.
(43) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(44) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490, 520, 553, 587.
(45) Cheng, A. L.; Best, S. A.; Merz, K. M., Jr.; Reynolds, C. H. *J. Mol. Graph. Model.* **2000**, *18*, 273.
(46) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Gao, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(47) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
(48) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902.
(49) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(50) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem. B* **1999**, *103*, 3765.
(51) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
(52) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(53) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Fergusen, D. M.; Siebel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.1*; University of San Francisco: San Francisco, CA, 1995.
(54) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571.
(55) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *109*, 1465.
(56) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629.
(57) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.
(58) Cornell, W.; Abseher, R.; Nilges, M.; Case, D. A. *J. Mol. Graph. Model.* **2001**, *19*, 136.
(59) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; Groningen Molecular Simulation (GROMOS) System; University of Groningen, The Netherlands; ETH Zurich: Switzerland, 1996.
(60) Gilson, M. K.; Honig, B. *J. Comput. Aid. Mol. Des.* **1991**, *5*, 5.
(61) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(62) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.
(63) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(64) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(65) York, D. M.; Lee, T. S.; Yang, W. T. *Chem. Phys. Lett.* **1996**, *263*, 297.
(66) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, 1981; p 331.
(67) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451.
(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(70) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. *Mol. Phys.* **1981**, *44*, 69.
(71) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103.
(72) Fraternali, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *256*, 939.
(73) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133.
(74) Lazaridis, T.; Karplus, M. *J. Mol. Biol.* **1998**, *288*, 477.
(75) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548.