JCTC Journal of Chemical Theory and Computation

Optimizing the Poisson Dielectric Boundary with Explicit Solvent Forces and Energies: Lessons Learned with Atom-Centered Dielectric Functions

Jessica M. J. Swanson,* Jason A. Wagoner, Nathan A. Baker, and J. A. McCammon

Howard Hughes Medical Institute, Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received July 1, 2006 ; Revised Manuscript Received September 25, 2006

Abstract: Accurate implicit solvent models require parameters that have been optimized in a system- or atom-specific manner on the basis of experimental data or more rigorous explicit solvent simulations. Models based on the Poisson or Poisson−Boltzmann equation are particularly sensitive to the nature and location of the boundary which separates the low dielectric solute from the high dielectric solvent. Here, we present a novel method for optimizing the solute radii, which define the dielectric boundary, on the basis of forces and energies from explicit solvent simulations. We use this method to optimize radii for protein systems defined by AMBER *ff*99 partial charges and a spline-smoothed solute surface. The spline-smoothed surface is an atom-centered dielectric function that enables stable and efficient force calculations. We explore the relative performance of radii optimized with forces alone and those optimized with forces and energies. We show that our radii reproduce the explicit solvent forces and energies more accurately than four other parameter sets commonly used in conjunction with the AMBER force field, each of which has been appropriately scaled for spline-smoothed surfaces. Finally, we demonstrate that spline-smoothed surfaces show surprising accuracy for small, compact systems but may have limitations for highly solvated protein systems. The optimization method presented here is efficient and applicable to any system with explicit solvent parameters. It can be used to determine the optimal continuum parameters when experimental solvation energies are unavailable and the computational costs of explicit solvent charging free energies are prohibitive.

I. Introduction

Properly accounting for solvation effects is a long-standing and constantly evolving challenge in computational biophysics. Both the dynamic and thermodynamic properties of a solvated molecule are strongly influenced by the microscopic structure and organization of the water that surrounds it. Predicting solvation free energies and forces accurately across different chemical architectures requires a formalism that collectively accounts for electrostatic, nonelectrostatic, and specific (e.g., hydrogen bonding) solute−solvent interactions. Microscopic formalisms treat these interactions explicitly with an atomistic representation of water. Macroscopic formalisms offer a less physically accurate but more computationally efficient approach: they replace individual molecular interactions with an implicit representation of water, most often as a linearly polarizable continuum.[1] The efficiency of implicit solvent models makes them ideal for large systems and computationally expensive problems such as extensive conformational sampling or high-throughput analyses. The lack of physical detail in implicit solvent models can be compensated by parametrization, for example, fitting Born radii to solvation free energies,[2,3] which generally results in more accurate quantitative predictions for specific

* Corresponding author phone: 858-822-2771; fax: 858-534-4974; e-mail: jswanson@ucsd.edu.

10.1021/ct600216k CCC: $37.00
Published on Web 11/15/2006

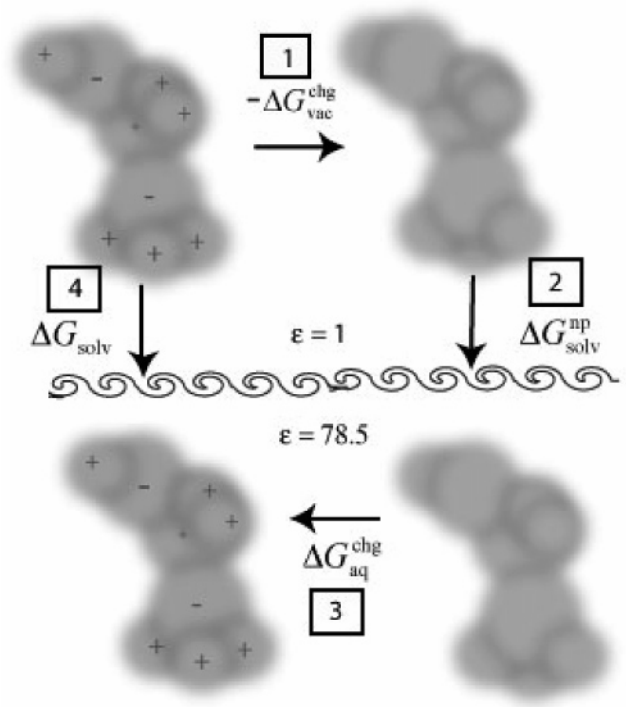

Figure 1. Solvation thermodynamic cycle in which solvation occurs in three path-dependent phases and according to which the total solvation free energy, ΔG_{solv}, is decomposed into polar, $\Delta G^{p}_{solv} = \Delta G^{chg}_{aq} - \Delta G^{chg}_{vac}$, and nonpolar, ΔG^{np}_{solv}, contributions.

solutes but questionable transferability between solutes and different solute conformations.

Most implicit solvent models divide the solvation free energy into polar and nonpolar contributions, $\Delta G_{solv} = \Delta G^{p}_{solv} + \Delta G^{np}_{solv}$. This division is rigorously based on a thermodynamic cycle (Figure 1) in which the solute charges are turned off in a vacuum (phase 1: $-\Delta G^{chg}_{vac}$), the neutral solute is solvated (phase 2: ΔG^{np}_{solv}), and the solute is recharged in the aqueous environment (phase 3: ΔG^{chg}_{aq}). The nonpolar contribution, because of hydrophobic and dispersion interactions, is commonly, though inadequately,[4,5] treated with a solvent-accessible surface area (SASA) model.[6−8] Improvements on this model include independently accounting for dispersion contributions,[4] atom-dependent SASA coefficients,[9] and more recently including an additional term which is proportional to the solvent-accessible volume.[5]

The polar contribution is the difference between charging the system in a vacuum and doing so in the solvent environment, $\Delta G^{p}_{solv} = \Delta G^{chg}_{aq} - \Delta G^{chg}_{vac}$. It can be approximated in a number of ways,[1] but we will focus on methods which solve the Poisson equation,[10−12]

$$-\nabla\cdot[\epsilon(r)\,\nabla\,\phi(r)] = 4\pi\,\rho(r)$$

for the electrostatic potential of the system, $\phi(r)$, given a position-dependent dielectric coefficient, $\epsilon(r)$, and the solute charge distribution, $\rho(r)$. Solving the electrostatic potential of a solute in a vacuum, ϕ_{vac}, and the aqueous environment, ϕ_{aq}, yields the polar solvation free energy

$$\Delta G^{p}_{solv} = -\frac{1}{2}\int \rho(r)[\phi_{aq}(r) - \phi_{vac}(r)]\,dr$$

The parameters which go into both the polar and nonpolar approximations are generally optimized with solvation energies from experiments[13] or explicit solvent simulations.[5,14] Although energies calculated with simulations are only as accurate as the underlying explicit solvent force fields, they can be quantitatively separated into polar and nonpolar contributions. The polar parameters in Poisson calculations include the solute charge distribution and a spatially varying dielectric coefficient. Energies and forces obtained from the Poisson equation are particularly sensitive to the nature and location of the boundary which separates the low dielectric solute ($\epsilon = 1-20$) from the high dielectric solvent ($\epsilon \approx 80$).

Traditionally, the dielectric boundary has been an abrupt transition from low to high dielectric values at the molecular surface, as defined by Lee and Richards[15] or Connolly[16]. Much work has gone into optimizing solute radii for molecular surfaces such that implicit solvent calculations reproduce accurate solvation energies.[13,17,18] Although molecular surfaces work well for most purposes involving static biomolecular structures, they are problematic for continuum solvent dynamics simulations because they are computationally costly and result in unstable forces which vary rapidly with changes in molecular conformation.[19] These problems can be avoided by defining a smooth dielectric transition at the solute surface with overlapping atom-centered polynomial or Gaussian functions.[19,20] The spline-smoothed surface introduced by Im et al.,[20] for example, is particularly well-suited for continuum solvent dynamics simulations because it results in numerically stable and efficient force calculations. Changing the nature of the dielectric transition, however, also changes its optimal location as defined by the solute radii. Not only do spline-smoothed surfaces have a gradual dielectric transition, they also define the low dielectric (internal) volume with a topology that is fundamentally different from the molecular surface. In particular, the spline-smoothed surface introduces interstitial spaces between atoms with significantly higher dielectric values than the remainder of the low dielectric solute volume. For these reasons, radii appropriate for an abrupt transition at the solute surface are not appropriate for spline-smoothed surfaces;[21] in fact, they result in solvation energies and forces which are overestimated by 10−40%.[22,23]

Here, we present a novel and efficient method for optimizing the Poisson dielectric boundary on the basis of explicit solvent simulations. Following the work of Wagoner and Baker,[5] we demonstrate how mean atomic solvation forces can be used alone or in addition to molecular solvation energies to optimize the solute radii for a given partial charge set and surface definition. This approach is closely related to force-matching techniques used in multiscale models[24,25] and provides atomically detailed information about the performance of implicit solvent models. It is significantly more efficient than previous optimization schemes which have relied on either experimental solvation free energies[13,17,26−28] or explicit solvent charging free energies.[18,21,23,29] Its simplicity and efficiency make it ideal for large, unusual,

or highly charged solutes for which solvation energies are unavailable and charging free energies are computationally expensive.

We use this approach to optimize a set of Poisson radii for protein systems defined by spline-smoothed surfaces and AMBER *ff*99 partial charges.[30] We compare radii that have been optimized with forces alone to those optimized with forces and energies, showing that the latter are superior when both forces and energies are important. We test our optimized radii against four other commonly used Poisson parameter sets,[13,28,30,31] each of which has been optimally scaled for spline-smoothed surfaces. Finally, we explore the effects of atom-centered surface definitions on radii optimizations and discuss their limitations in large, globular solutes. Combining the polar parameters developed here with the recently developed and complementary nonpolar parameters[5] should greatly increase the accuracy of the Poisson-based implicit solvent framework for protein systems.

II. Theory

A straightforward statistical mechanical formalism for implicit solvent models[1,32] begins with the potential energy of a solute in an aqueous medium decomposed into

$$U_{\mathrm{TOT}}(X,Y) = U(X) + U(Y) + U(X,Y) \tag{1}$$

where X and Y are the solute and solvent degrees of freedom, respectively, $U(X)$ is the intramolecular solute potential, $U(Y)$ is the solvent−solvent potential, and $U(X,Y)$ is the solute−solvent potential. The system's free energy is then

$$-\beta G = \ln\left(\int \mathrm{d}X \int \mathrm{d}Y\, \mathrm{e}^{-\beta[U(X)+U(Y)+U(X,Y)]}\right) \tag{2}$$

where $\beta = (k_{\mathrm{B}}\mathrm{T})^{-1}$ is the inverse thermal energy. Equation 2 can be used to derive a potential of mean force, $W(X)$, in the usual way:[1,33]

$$\mathrm{e}^{-\beta W(X)} = \frac{\int \mathrm{e}^{-\beta[U(X)+U(Y)+U(X,Y)]}\,\mathrm{d}Y}{\int \mathrm{e}^{-\beta[U(Y)]}\,\mathrm{d}Y} \tag{3}$$

Note that the integral of $\mathrm{e}^{-\beta W(X)}$ over the solute degrees of freedom gives $\mathrm{e}^{-\beta\Delta G_{\mathrm{solv}}}$ where ΔG_{solv} is the solute's solvation free energy. Choosing the unsolvated solute as the reference state, the solvation free energy of a particular solute conformation X is $\Delta W(X) = W(X) - U(X)$. This solvation free energy can be decomposed into polar and nonpolar terms according to a path-dependent process: phases 1−3 of the thermodynamic cycle shown in Figure 1. The coupling between polar and nonpolar solvation interactions, which has been demonstrated in several studies (see Dzubiella et al.[34,35] for discussion), is not explicitly defined in this thermodynamic cycle but implicitly included in the charging phase, $\Delta G_{\mathrm{aq}}^{\mathrm{chg}}$. Decomposition into polar and nonpolar solvation contributions begins by expressing the solute−solvent potential as a sum of polar and nonpolar contributions,[1] $U(X,Y) = U^{\mathrm{p}}(X,Y) + U^{\mathrm{np}}(X,Y)$, such that

$$\Delta W(X) = \Delta W^{\mathrm{np}}(X) + \Delta W^{\mathrm{p}}(X) \tag{4}$$

and

$$\mathrm{e}^{-\beta\Delta W^{\mathrm{np}}(X)} = \frac{\int \mathrm{e}^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)]}\,\mathrm{d}Y}{\int \mathrm{e}^{-\beta[U(Y)]}\,\mathrm{d}Y} \tag{5}$$

and

$$\mathrm{e}^{-\beta\Delta W^{\mathrm{p}}(X)} = \frac{\int \mathrm{e}^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)+U^{\mathrm{p}}(X,Y)]}\,\mathrm{d}Y}{\int \mathrm{e}^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)]}\,\mathrm{d}Y} \tag{6}$$

These contributions can be calculated in two successive stages of a free energy perturbation or thermodynamic integration[36] according to

$$\Delta W^{\mathrm{np}}(X) = \int_0^1 \mathrm{d}\lambda_1 \left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \lambda_1} \right\rangle_{(\lambda_2=0)} \tag{7}$$

and

$$\Delta W^{\mathrm{p}}(X) = \int_0^1 \mathrm{d}\lambda_2 \left\langle \frac{\partial U^{\mathrm{p}}(X,Y)}{\partial \lambda_2} \right\rangle_{(\lambda_1=1)} \tag{8}$$

where λ_1 and λ_2 are coupling parameters that scale the nonpolar and polar solvent−solute interactions from 0 (off) to 1 (on). The polar contribution is therefore a "charging free energy" but is often referred to as the polar solvation free energy. The mean forces can be obtained by differentiation of $W(X)$ with respect to the solute coordinates[1], X

$$\langle F^{\mathrm{np}}(X)\rangle = -\frac{\partial W^{\mathrm{np}}(X)}{\partial X} = -\left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{np}} \tag{9}$$

and

$$\langle F^{\mathrm{p}}(X)\rangle = -\frac{\partial W^{\mathrm{p}}(X)}{\partial X} = -\left\langle \frac{\partial U^{\mathrm{p}}(X,Y)}{\partial \mathrm{X}} + \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{p}} + \left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{np}} \tag{10}$$

where $\langle\cdots\rangle_{\mathrm{np}}$ and $\langle\cdots\rangle_{\mathrm{p}}$ denote the ensemble averages with nonpolar only and full (nonpolar and polar) solute−solvent interactions, respectively.

In this work, we will focus on polar contributions, $\Delta W^{\mathrm{p}}(X) \equiv \Delta G_{\mathrm{solv}}^{\mathrm{p}}$ and $\langle F^{\mathrm{p}}(X)\rangle \equiv \bar{F}^{\mathrm{p}}$. A similar treatment of nonpolar contributions, presented by Wagoner and Baker,[22] was utilized to optimize nonpolar implicit solvent parameters with explicit solvent forces.[5] The goal of the present work is to optimize polar parameters, specifically the solute radii used to define spline-smoothed dielectric boundaries in protein systems, by fitting Poisson solvation free energies and forces to explicit solvent charging free energies and mean atomic forces.

Our approach assumes that polar parameters are independent of the solute's conformation and polar−nonpolar coupling. Though this is clearly an approximation,[34,35] it may be a reasonable approximation for biomolecular systems because they are predominantly composed of similar subunits (e.g., amino and nucleic acids) with significant solute−solvent dispersion interactions. Previous efforts have shown that optimizing complementary polar and nonpolar parameters in a system- or atom-dependent manner can result in

Table 1. Force and Force−Energy-Optimized Radii

atom name[a]	residues	opt. radii[b]	opt. radii[c]
	Backbone		
C	all	2.338	2.317
O	all	1.766	1.732
N	all	2.331	2.399
CA	all except G	2.425	2.428
CA	G	2.122	2.353
CAY/CAT	ACE,NME	2.022	2.148
	Side Chains		
CB	D,E,C,H,M,F,S,T,W,Y	2.128	2.231
CB	A,R,N,Q,I,L,K,V	2.209	2.576
CG*	R,Q,I,L,K,M,T,V	2.414	2.414
CG	H,F,W,Y	2.147	2.136
CG/CD	N,Q,D	2.197	2.353
CG	E	2.195	2.341
CB/CG/CD	P	2.506	2.440
CD	R,K	2.316	2.442
CD*	I,L	2.640	2.559
CD*/CE*/CZ	H,F,W,Y	2.282	2.217
CE	M	1.940	1.972
CZ/CE	R,K	2.398	2.461
OD*/OE*	N,Q,D,E	1.729	1.756
OG*	S,T	1.956	2.101
OH	Y	1.871	2.108
NE,NH*,NZ	R,K	2.323	2.328
ND2/NE2	N,Q	2.122	2.397
ND1,NE2	H	1.927	2.040
NE1	W	1.880	1.960
SG/SD	C,M	2.406	2.337
	Hydrogens[d]		
type H	bb HN	1.244	1.388
type H	bound to N	1.228	1.302
type HO/HS	bound to O/S	0.999	0.917
type H1/HP	polar	1.905	2.019
type HC/HA	nonpolar	1.809	1.680

[a] Radius groups are distinguished by AMBER atom names for all heavy atoms and by atom type for hydrogen atoms. [b] Final radii [Å] from force optimizations. [c] Final radii [Å] from force and energy optimizations. [d] Hydrogens specified by atom type with type "H" divided into two groups: amide backbone "HN" and all other N-bound hydrogens.

fairly accurate, though sometimes system-dependent, solvation frameworks.[13,26,27] The SM*x* models, for example, can predict solvation free energies for neutral molecules with errors less than 1 kcal/mol and have shown recent success in treating charged solutes and ions with the addition of an explicit water molecule.[26,27] Thus, combining the presented polar parameters with complimentary nonpolar parameters[5] should result in a complete Poisson-based implicit solvent framework for protein systems that is accurate to the exclusion of polar−nonpolar coupling, conformational dependence, and the limitations of mean-field theories.

III. Methods

Force-based optimizations involve three steps, which are described in more detail below. First, explicit solvent forces are collected for each solute in the training set in one or more fixed conformations. Next, the continuum calculations are set up such that energies and forces are well converged, that is, independent of grid resolution, boundary conditions, and so forth. Finally, the Poisson atomic radii are optimized to reproduce the explicit solvent forces. Optimizations using both energies and forces additionally require calculating explicit solvent charging free energies with thermodynamic integration or free energy perturbation simulations as previously reported.[23]

Training Sets. To test the limitations of our methodology, we began by optimizing *each* atomic radius in three nonzwiterionic *N*-acetyl-X-*N*′-methylamide dipeptides, where X represents alanine with ϕ/ψ angles of 180°/180° (ala1), alanine with ϕ/ψ angles of −60°/−40° (ala2), and serine with ϕ/ψ angles of 180°/180° (ser). As previously described,[23] using simple PARSE, Bondi, or AMBER-vdW atom types limits the accuracy of the resulting parameters because of significant differences in the solvation structures, energies, and forces for atoms within these atom types. Therefore, we used 31 "atom types" or radius groups (Table 1) which were previously identified as having similar chemical signatures and surrounding solvation structures according to explicit solvent charge distribution functions.[23]

The training and test sets are summarized in Table 2. We used two sets of model systems in our force-based optimizations. The first set was two conformations of intestinal fatty acid binding protein (IFABP) representing highly populated conformational clusters from an a priori fully equilibrated explicit solvent simulation.[22] The second set was two short polypeptide chains: Trpcage (PDB ID 1l2y)[37] and the C-terminal fragment (residues 41−56) of protein G (PDB ID 2gb1),[38] hereafter called G-peptide. In the energy- and force-based optimizations, we additionally used the explicit solvent charging free energies of 20 *N*-acetyl-X-*N*′-methylamide dipeptides, where X represents each of the amino acids (ϕ/ψ angles of 180°/180°) and seven polyalanine peptides with common secondary structures (three β turns, two helices, and two β hairpins). Each of the polyalanine peptides was modeled from fragments of lysozyme (PDB ID 1ati) or crambin (PDB ID 1ejg), which were mutated to polyalanine and capped with neutral blocking groups. The specific peptide conformations and explicit solvent charging free energies were previously reported.[23]

Testing Sets. To determine if the optimized radii are transferable to molecules outside of the training set, we used the forces and solvation energies of the other test systems (e.g., the Trpcage and G-peptide forces for the IFABP-force-trained radii) as well as the solvation energies of the following molecules: Trpcage, G-peptide, two additional proteinlike polypeptides, five new polyalanine peptides, and 20 amino acid dipeptides in a new side chain and backbone conformation (ϕ/ψ angles of −60°/−40°).[23] We also used the forces of IFABP in six new conformations. The relative root-mean-square deviation (RMSD) values for all eight IFABP structures ranged from 1.2 to 2.2 Å.[22] All of the aforementioned systems were used to compare the optimized radii to four other continuum parameter sets: AMBER *ff*99 partial charges combined with *ff*99 van der Waals (vdW) radii,[30] Bondi radii,[31] and a set of radii recently published by Luo et al.,[28] as well as the PARSE parameters (charges

Table 2. Solutes Used to Train and Test Radii on the Basis of Explicit Solvent Forces or Energies

training sets solute[s]	quantity used	trained radii	additional test sets solute[s]	quantity used
IFABP conformations 1 & 2	forces	f-opt 1	IFABP conformations 3−8	forces
Trpcage & G-peptide	forces	f-opt 2 and fe-opt	20 dipeptides conformation 2	energies
20 dipeptides conformation 1	energies	fe-opt	polyalanine peptides 8−12	energies
polyalanine peptides 1−7	energies	fe-opt	4 proteinlike peptides	energies

and radii).[13] Each of the comparison parameter sets was scaled for use with spline-smoothed surfaces as described below.

Explicit Solvent Forces. As shown in eq 10, the polar solvation forces are calculated by subtracting the nonpolar solvation forces from the total forces ($\bar{F}^{p} = \bar{F}^{np+p} - \bar{F}^{np}$). The nonpolar forces are averaged from an ensemble generated with only nonpolar (vdW) solute−solvent interactions, while the total forces are averaged over an ensemble generated with both polar (electrostatic) and nonpolar solvent−solute interactions. Thus, the calculation of polar forces requires two simulations: one with all solute−solvent interactions and one with only nonpolar solute−solvent interactions (i.e., solute charges set to zero).

The IFABP simulations were performed, as previously reported,[5] with the AMBER *ff*99[30] force field using AMBER 7 software.[39] Eight conformations were obtained by clustering a 2.5 ns simulation which started from a nuclear magnetic resonance structure (PDB ID 1ael)[40] and was run in TIP3P water in the presence of ∼160 mM NaCl at 300 K and 1 atm of pressure. The clusters were calculated with the single linkage algorithm available in GROMACS[41] using a 1.0 Å cutoff over all main-chain atoms. The eight resulting clusters accounted for 81% of the trajectory structures. Each conformation was then constrained via belly dynamics[39] and simulated for 1.05 ns with SHAKE-enabled 2 fs time steps under isobaric−isothermal conditions. The forces for conformations 3−6 were averaged from 250 snapshots taken every 4 ps from the last nanosecond of simulation. The simulations of conformations 1 and 2 were extended an additional 4 ns, and forces were averaged from 1000 snapshots taken every 4 ps from the last 4 ns of simulation. Comparing these forces with those averaged from the last nanosecond of an independent 1.05 ns simulation resulted in an average squared residual of 0.085 56 kcal/mol Å, indicating that the forces were well-converged.

Trpcage, G-peptide, and the three dipeptides (ala1, ala2, ser) were simulated with the AMBER *ff*99 converted to CHARMM format and the CHARMM software (version 31a1).[42] The solutes were protonated and solvated in a sphere of TIP3P water molecules that extended 10.0 Å beyond the solute, resulting in three to four hydration shells. The spherical solvent boundary potential including Kirkwood's multipolar expansion reaction field was used to approximate the influence of bulk water beyond the explicit water sphere.[43] This model was previously used in the free energy perturbation simulations to calculate charging free energies[23] because it alleviates difficulties that result from perturbing charged systems with periodic boundary conditions. It was chosen here for consistency with the molecular charging free energies. It has been shown to give reliable results for proteins, nucleic acids, and small molecules.[1,18,44] All solute atoms were restrained to their original coordinates. The solvent was first energy-minimized with 50 steps of steepest descent followed by 1000 steps of the ABNR method and then equilibrated for 200 ps. Langevin dynamics were employed at a constant temperature (300 K) using SHAKE-enabled 2 fs time steps, infinite cutoffs for nonbonded interactions, and a friction constant corresponding to a relaxation time of 5 ps applied to water oxygen atoms.

The peptide forces were averaged from snapshots taken every 0.2 ps over 500 ps of pre-equilibrated simulation for a total of 2500 conformations. Comparing forces from the first 200 ps with those from the full 500 ps resulted in a root-mean-square deviation (RMSD_$\bar{F}^{p}$) of 0.187 kcal/mol/Å and an average relative error of 0.0884 kcal/mol/Å, demonstrating reasonable convergence after only 200 ps. Comparing duplicate 500 ps simulations of the dipeptides dropped RMSD_$\bar{F}^{p}$ to 0.05−0.09 kcal/mol/Å, which was considered sufficient convergence. Because the IFABP and peptide systems were used in independent optimizations, the effects of the different simulation protocols for these two systems are expected to be minor.

Implicit Solvent Forces. All implicit solvent energies and forces were calculated with the Adaptive Poisson−Boltzmann Solver (APBS; http://apbs.sf.net/) version 0.4.0.[45] APBS employs the analytical method of Im et al.[20] to evaluate polar solvation forces. In particular, the dielectric function, $\epsilon(r)$, is defined by overlapping atom-centered cubic polynomials

$$\epsilon(r) = 1 + (\epsilon_s - 1)\prod_i H_i(|r - r_i|)$$

where ϵ_s is the dielectric coefficient for the solute, H_i is the

$$H_i(r) = \begin{cases} 0 & r < R_i - w \\ \dfrac{-1}{4w^3}(r - R_i + w) + \dfrac{3}{4w^2}(r - R_i + w)^2 & R_i - w < r < R_i + w \\ 1 & r \geq R_i + w \end{cases}$$

polynomial for each atomic sphere, r is the distance from a point in the system and atom i, R_i is that atom's radius, and w is the half-width of the transition region.[20] The resulting total solvation force is composed of three terms: a reaction field force due to interaction of the solute charge with the total electric field, a dielectric boundary force due to spatial variation in the dielectric function, and an ionic boundary force for systems with ions. The last term was zero in this study because we used zero bulk ionic strength. The abruptness of the transition between the solute and solvent

values in these spline-smoothed surfaces is controlled by the spline window half-width, w, a user-defined parameter in APBS. We optimized our radii with the recommended value, $w = 0.3$ Å. We additionally scaled these radii for spline windows 0.2 Å $\leq w \leq$ 1.0 Å with both force- and force−energy-based genetic algorithm optimizations (as described below). The Poisson equation was solved at a temperature of 300 K with the multiple Debye−Hückel sphere boundary condition and cubic B-spline charge discretization. We used the experimental solvent dielectric of 78.4, which differs from the reported dielectric constants[46,47] of 91−104 for TIP3P water[48] but is consistent with typical values used with implicit solvent models. Calculations run with the PARSE parameters used a solute dielectric of 2.0 in agreement with their original development.[13] All other calculations used a solute dielectric of 1.0, which is appropriate for calculations that explicitly sample solute conformations and desirable for consistency with nonpolarizable force fields. It is important to ensure that the chosen continuum parameters, particularly the grid parameters, provide sufficient spatial resolution to ensure the convergence of energies and forces. We compared the energies and forces calculated with successively finer grids (down to 0.10 Å) and selected production calculation grid parameters that resulted in relative errors below 0.5%. Our final grids had resolutions of either 0.2 or 0.25 Å, depending on the solute, and overall dimensions which extended beyond the solute by 35% of its length in each dimension.

Genetic Algorithm Optimizations. We used a genetic algorithm to optimize the radii from various starting values (AMBER *ff*99 vdW, Bondi, and our previously published smooth boundary radii[23]) to their final values. Populations of 50 solutions were run for 10 to 100 generations. The initial populations were generated from a uniform distribution that varied from 0.1 to 0.3 Å around the starting values, depending on the optimization. The fitness of each solution was evaluated, and the next population was generated via selection, crossover, and mutation. Selections were based on the stochastic universal sampling algorithm,[49] which chooses solutions on the basis of a probability proportional to their fitness

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{11}$$

where p_i is the probability that solution i will be selected and f_i is its fitness defined by

$$f_i = (1 + \mathrm{RMSD\_}\bar{F}_i^{\mathrm{p}})^{-1} \tag{12}$$

where RMSD_$\bar{F}^{\mathrm{p}}$ is the standard deviation of all 3N force components in an N-atom system for the radii solution set, i. The fitness scores range from 0 (poor) to 1 (perfect). A uniform crossover process in which two solutions from the previous generation were randomly distributed to two new solutions was applied to either 80% or 90% of the population, while mutations selected from a Gaussian distribution with standard deviations which ranged from ±0.025 to ±0.1 Å were applied to anywhere from 5% to 30% of the population. The various different crossover and mutation values were tested to ensure that the final solutions were not an artifact of limiting genetic algorithm run parameters. This process was repeated until a desired fitness was reached or the maximum number of generations was exceeded.

As we will demonstrate, accurately fit solvation forces do not necessarily ensure accurate solvation energies. Thus, we ran a second set of optimizations that incorporated the solvation energies of 20 amino acid dipeptides and seven polyalanine peptides as well as the Trpcage and G-peptide solvation forces into the fitness function according to

$$f_i = [1 + 1/2(\mathrm{RMSD\_}\bar{F}_i^{\mathrm{p}} + \mathrm{AAE\_}\Delta G_{p\_i}^{\mathrm{solv}})]^{-1} \tag{13}$$

where RMSD_$\bar{F}^{\mathrm{p}}$ is defined as before and AAE_$\Delta G_{p_i}^{\mathrm{solv}}$ is the average absolute error for the 27 solutes' polar solvation energies for solution set i. Because the optimal RMSD_$\bar{F}^{\mathrm{p}}$ and AAE_$\Delta G_{p_i}^{\mathrm{solv}}$ terms are not known a priori, it is not possible to define a fitness function that would weight the forces and energies equally. Equation 13 was chosen to give a small bias to energies over forces based on previously computed AAE_$\Delta G_{p_i}^{\mathrm{solv}}$ values (∼0.9−4.0 kcal/mol) which were slightly larger than the RMSD_$\bar{F}^{\mathrm{p}}$values (∼0.7−1.5 kcal/mol/Å).

Scaling Radii for Spline-Smoothed Surfaces. Because the spline-smoothed surface brings high dielectric values closer to each atom's center than the molecular surface does, radii that work well with molecular surfaces overestimate solvation energies and forces by 10−40% when used with spline-smoothed surfaces.[22,23] A first-order correction to this overestimation is to scale each radius by a single factor, x, according to

$$R_{\mathrm{new}} = R_{\mathrm{orig}} + x \tag{14}$$

where R_{new} is the new radius and R_{orig} the original. The four parameter sets chosen to benchmark our optimized radii are all intended for use with molecular surfaces. Therefore, to make a fair comparison, we had to determine the best scaling factor, x_{best}, for each set of radii. We chose to do this using force-only and force- and energy-based optimization schemes analogous to those used in our radii optimizations described above. We used the same approach to scale *down* our spline-smoothed radii from Table 1 for molecular surfaces in order to compare the two surface topologies (molecular and spline-smoothed) directly. Finally, we scaled our radii (Table 1), which use a spline window half-width of $w = 0.3$ Å, for different spline window half-widths ranging from $w = 0.2$ to 1.0 Å. We selected this scaling function (eq 14) over that used in previous optimizations, $R_{\mathrm{new}} = x(R_{\mathrm{orig}} + w)$,[21] because it resulted in slightly higher final fitness values and converged in one to three fewer generations.

IV. Results

Dipeptide Test Case. It is important know the optimal degree of accuracy one can expect from the presented methodology, that is, the instrinsic limitations of force- and force−energy-based parametrizations. To address this question, we generated "perfect (Poisson) radii"[50] by optimizing *every* atom's radius in three simple dipeptides, alanine in

Table 3. Individually Optimized Radii

	force-optimized ala1 radii		force−energy-optimized ala1 radii	
	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]
ala1	−1.753 (0.290)	0.147 (0.009)	0.000 (0.005)	0.244 (0.025)
ala2	−1.594 (0.336)	0.358 (0.013)	0.152 (0.073)	0.432 (0.026)
ser	−3.314 (0.247)	0.344 (0.020)	−1.633 (0.164)	0.333 (0.046)
	force-optimized ala2 radii		force−energy-optimized ala2 radii	
	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]
ala1	−1.865 (0.359)	0.302 (0.009)	−0.054 (0.103)	0.287 (0.020)
ala2	−1.906 (0.265)	0.160 (0.003)	−0.001 (0.005)	0.322 (0.046)
ser	−3.810 (0.391)	0.412 (0.011)	−1.820 (0.168)	0.342 (0.027)
	force-optimized ser radii		force−energy-optimized ser radii	
	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]	$\Delta\Delta G_{solv}$[a]	RMSD_F[b]
ala1	−0.603 (0.210)	0.291 (0.036)	1.476 (0.171)	0.422 (0.027)
ala2	−0.450 (0.209)	0.390 (0.009)	1.603 (0.148)	0.616 (0.024)
ser	−2.420 (0.233)	0.200 (0.009)	0.000 (0.002)	0.418 (0.025)
	force-optimized radii		force−energy-optimized radii	
	$\Delta\Delta G_{solv}$[c]	RMSD_F[d]	$\Delta\Delta G_{solv}$[c]	RMSD_F[d]
ala1	0.376	0.262	0.217	0.265
ala2	0.569	0.444	0.449	0.476
ser	−0.831	0.560	−0.842	0.598

[a] The *average* solvation energy deviation [kcal/mol]. [b] *Average* force RMSD [kcal/mol/Å] for three dipeptides using individually force and force−energy-optimized radii. Statistics collected from 30 independent genetic algorithm optimizations. Standard deviations are given in parentheses. [c] The solvation energy deviation [kcal/mol]. [d] Force RMSD [kcal/mol/Å] for the three dipeptides using the force and force−energy-optimized radii presented in Table 1.

two different conformations (ala1 and ala2) and serine (ser). We used both the force- and the force−energy-based approaches (eqs 11 and 12, respectively). To ensure that the results were independent of the genetic algorithm run and the parameters, the optimizations were repeated 30 times varying the initial population distribution from ±0.1 to 0.4 Å, the mutation distribution from ±0.025 to 0.1 Å, and the mutation ratio from 0.05 to 0.4 (i.e., mutations were applied to 5−40% of the solutions in the next generation). The resulting force and polar solvation energy statistics (shown in Table 3) demonstrate errors that are larger than the inherent errors in the explicit and implicit calculations. The explicit solvation energy and force errors were $\Delta\Delta G_p^{solv} \approx 0.02-0.17$ kcal/mol and RMSD_$\bar{F}^p \approx 0.05-0.09$ kcal/mol Å, respectively, while the continua were $\Delta\Delta G_p^{solv} \approx 0.03$ kcal/mol and RMSD_$\bar{F}^p \approx 0.05$ kcal/mol Å.

There are five key points to take away from this example. First, the genetic algorithm solution space is highly frustrated. The majority of optimizations generated solutions of similar fitness but unique parameters, indicating that there are multiple solutions even for simple molecules with independently fit radii. Second, the explicit solvent forces cannot be perfectly reproduced even with independently fit radii. Whether this discrepancy is caused by the limited dielectric definition (i.e., the union of spline-smoothed spheres) or limitations in the Poisson equation itself (i.e., the assumption of linear and local solvent response) is not clear. It is our expectation that all of these limitations contribute to the differences between explicit and implicit solvation forces but that the limited dielectric definition is the major source of error. This analysis suggests that the limit of accuracy for implicit solvent forces calculated with this surface is RMSD_$\bar{F}^p \gtrsim 0.15$ kcal/mol Å. Third, each solute has a unique set of optimal continuum radii. When the force-optimized radii for the serine peptide are used on alanine or vice versa, the average RMSD_$\bar{F}^p$ increases by 0.14 to 0.21 kcal/mol Å. Similarly, when the force−energy-optimized radii are switched, the average $\Delta\Delta G_p^{solv}$ decreases from ∼0.0 to between −1.820 and −0.054 kcal/mol. Because serine and alanine share many of the same atom types, these increased errors indicate that neighboring atoms *change* the optimal continuum radius for a specific atom type. Fourth, the solute's conformation also changes the optimal continuum radii. When the radii are switched for ala1 and ala2, we observe a similar, though less dramatic, increase in the average RMSD_$\bar{F}^p$ by 0.15 to 0.16 kcal/mol Å and a change in the average $\Delta\Delta G_p^{solv}$ to between −0.054 and 0.152 kcal/mol. Although the goal of optimizing radii with multiple solutes and conformations is to find a solution set which is transferable to similar solutes and to different solute conformations, it is not likely that such a solution set will be more accurate than radii that are individually optimized for these three very similar peptides. Thus, one can expect the limit of accuracy of force-optimized radii used on varying

Table 4. Summary of Force RMSDs and Solvation Energy Average Absolute Errors (AAEs) for Optimized Radii and Comparison Parameter Sets

		RMSD_F [kcal/mol/Å]		AAE_ΔG_{solv} [kcal/mol]			
parameters[a]	x_{best}	Trpcage & G-peptide[b]	IFABP[c]	di/poly-ala peptides[d]	di/poly-ala peptides[e]	protein	peptides[f]
Amber-f	0.378	0.870	1.020	4.953	4.719	20.296	7.60%
Amber-fe	0.281	1.055	1.222	1.986	1.873	14.997	6.02%
Parse-f	0.236	1.801	2.438	4.766	6.346	26.608	10.99%
Parse-fe	0.310	1.822	2.178	2.944	3.342	11.213	4.00%
Parse0-f	0.236	1.812	2.456	4.186	4.346	27.778	11.47%
Parse0-fe	0.269	1.816	2.327	3.929	3.342	14.725	6.23%
Bondi-f	0.311	1.216	1.460	3.314	2.356	28.122	11.29%
Bondi-fe	0.373	1.240	1.443	1.088	0.971	7.014	2.99%
Luo-f	0.403	0.923	1.014	5.565	5.173	15.108	5.67%
Luo-fe	0.283	1.150	1.360	1.902	1.949	26.472	10.74%
IFABP Opt-f	na	0.750	0.906	4.136	3.799	11.379	4.35%
optimized-f	na	0.656	0.971	1.035	0.832	7.4545	3.00%
optimized-fe	na	0.743	1.029	0.453	0.536	5.7164	2.04%

[a] Optimized radii and comparison parameter sets scaled by scaling factor, x_{best} [Å], that was fit with forces (-f) or forces and energies (-fe) according to $R_{new} = R_{orig} + x$. [b] Trpcage and G-peptide forces used in force and force−energy training sets. [c] IFABP forces used in IFABP force optimization training set. [d] Dipeptides and polyalanine peptides used in force−energy training set. [e] Dipeptide and polyalanine peptides *not* used in training set. [f] Four proteinlike peptides *not* used in training sets.

conformations and across unique amino acid architectures to be RMSD_$\bar{F}^p \approx 0.29$ kcal/mol Å and that of force−energy-optimized radii to be $\Delta\Delta G_p^{solv} \approx 0.05$ kcal/mol. Our optimized radii show errors just above these limits for both forces and solvation energies (Table 3).

Finally, the best-fit forces do not necessarily result in accurate solvation energies. In fact, the solvation energies are overestimated by the individually force-optimized radii by between 0.45 and 3.81 kcal/mol! This apparent contradiction to the theoretical framework presented in section II is caused by a fundamental difference in the implicit and explicit solvent potentials. The relationship presented in eq 10 *is* exact *within* a given solvent model, and it can be used to fit the implicit potential of mean force to the explicit potential of mean force, as we are attempting to do here. However, fitting the directional derivative (i.e., force) of the potential of mean force from one solvent model to the other does not guarantee that the change in free energy will be equally well-fit. In fact, the approximate, at best, agreement between implicit and explicit solvent potentials of mean force for salt bridges and hydrogen bonds[51,52] suggests that more accurate forces would necessitate less accurate energies. The results shown in Tables 3 and 4 support this hypothesis.

Protein Radii Optimizations. Similar to the dipeptides, the solution space for the force- and force−energy-based optimizations was also highly frustrated with many similarly well-fit solutions. The various different starting radii (AMBER *ff*99 vdW, Bondi, and our previously optimized radii) had no distinguishable effect on the final solutions. However, the choice of starting radii did affect the speed of the optimizations, with the latter two sets being the most efficient. Increasing the number of target values generally decreases the frustration in the solution space. Thus, using 3N atomic solvation forces as opposed to one solvation energy for an N-atom solute makes the optimization substantially more efficient and decreases the likelihood that the final solution is locally trapped far from the global minimum. The force-based optimizations reached complete convergence; that is, the same solution set was found for five consecutive generations, in a single evolution of 20−80 generations. The force−energy-based optimizations never reached complete convergence but converged to a minimum fitness deviation (ΔRMSD_$\bar{F}^p \leq 0.002$) in one to two evolutions of ∼65 generations. Both the force- and force−energy-based optimizations were considerably faster than our previous optimizations, which used solvation energies alone and took six or more evolutions to reach a minimum fitness convergence.

Ideally, both continuum forces and energies would be faithfully predicted by a single parameter set. However, the discrepancy in implicit and explicit potentials demonstrated by the dipeptides suggests that the accuracy of one or the other must be sacrificed. This was indeed the case for the protein radii as shown in Table 4. The force-optimized radii performed 5.6−11.7% better on forces, while the force−energy-optimized radii perform 35.6−56.2% better on polar solvation energies. Depending on the particular problem at hand, accuracy in individual forces or overall energies may be more important; thus, both sets are presented in Table 1. However, we recommend the force−energy-optimized radii for general applications where accuracy in both forces and energies is important. The force-optimized radii presented in column 2 of Table 1 are those from Trpcage and G-peptide optimizations. As described below, radii obtained from these two systems provide much better global accuracy than those fit from IFABP simulations.

Effects of Interstitial High Dielectrics. The two model systems used in the force-based optimizations resulted in strikingly different solution sets. The IFABP radii were 1−10% larger than the Trpcage and G-peptide radii and underestimated solvation energies by 1−26% (Table 4). Close inspection of the IFABP atomic forces that were being significantly over- and underestimated revealed the source of this discrepancy: the significant outliers were atoms found

near interstitial high dielectrics (IHDs). IHDs are regions which have a higher dielectric value (i.e., are solvent) in atom-centered dielectric functions but a low dielectric value (i.e., are solute) in molecular surfaces. They are, by definition, too small for an explicit water molecule to penetrate. The dielectric constant in these regions is the same as the solvent's ($\epsilon \approx 80$) in a van der Waals surface (the envelope surrounding unsmoothed atomic spheres) but can take on a range of values ($1 < \epsilon \leq 80$) in spline-smoothed surfaces. IHDs can be identified by comparing molecular and spline-smoothed surface topologies. As described in the Methods, it was necessary to scale down our spline-smoothed radii (Table 1) in order to define a realistic molecular surface that could be compared to our optimal spline-smoothed surfaces. The scaling factors and solvation energies for molecular surfaces are presented in the Supporting Information. Using unscaled spline-smoothed radii would place the molecular surface too far from the solute, overestimating the solute volume and, hence, overestimating the volume of IHDs. The optimal scaling factor for the force-optimized radii was $x_{best} = -0.275$ Å.

Figure 2 shows a series of slices through the dielectric map of IFABP. The magnitude of IHDs, demonstrated by comparing the molecular surface (panel A) and the spline-smoothed surface (panel B) through a single plane of intersection, is consistent through the entire structure. Both surfaces are rendered in panels C−E to show three examples of atoms with significantly overestimated continuum forces: tyr70 OH, ser109 OG, and glu85 N are next to IHDs with volumes of 41.06, 20.44, and 13.47 Å^3, respectively. This overestimation occurs regardless of the parameter set used to define each atom's radius; it is 3.6, 4.5, and 2.9 kcal/mol/Å and 2.8, 1.9, and 3.7 kcal/mol/Å for the Trpcage/G-peptide and IFABP force-optimized radii, respectively. Moreover, analyses of the explicit solvent simulations demonstrate that these IHDs have explicit solvent occupancies of 0.431, 0.33, and 0.0 in regions with dielectric values above 70 ($70 \leq \epsilon \leq 78.4$) and 0.0 in all regions less than 70 ($10 < \epsilon < 70$), indicating that they *are* mostly buried and *should* have low dielectric values.

Figure 3 shows that the percentage of IFABP atoms with large force deviations increases as the volume of IHDs within 2.0 Å of the atom's radial boundary increases. Forces from all eight IFABP conformations were qualified as "large" if they were greater than a given cutoff value; this value was allowed to vary to test sensitivity (see Figure 3). Although the plotted data were generated with the force-optimized radii from Table 1, other radii yielded very similar results. This trend lends support to the notion that the force deviations are related to the presence of IHDs. Figure 4 shows that IFABP has a larger distribution of force deviations than Trpcage and G-peptide, when defined by either the force-optimized or scaled Bondi radii. The same trend was found for all of the tested parameter sets, as indicated by consistently larger RMSD_$\bar{F}^p$ values for IFABP than Trpcage and G-peptide (Table 4). This suggests that IHDs introduce errors in the entire protein, not just in proximal atoms.

IFABP is a highly solvated protein with a significant number of solvent-exposed residues in "buried" fatty acid binding sites. It is more susceptible to the formation of IHDs than Trpcage or G-peptide, which have compact structures.

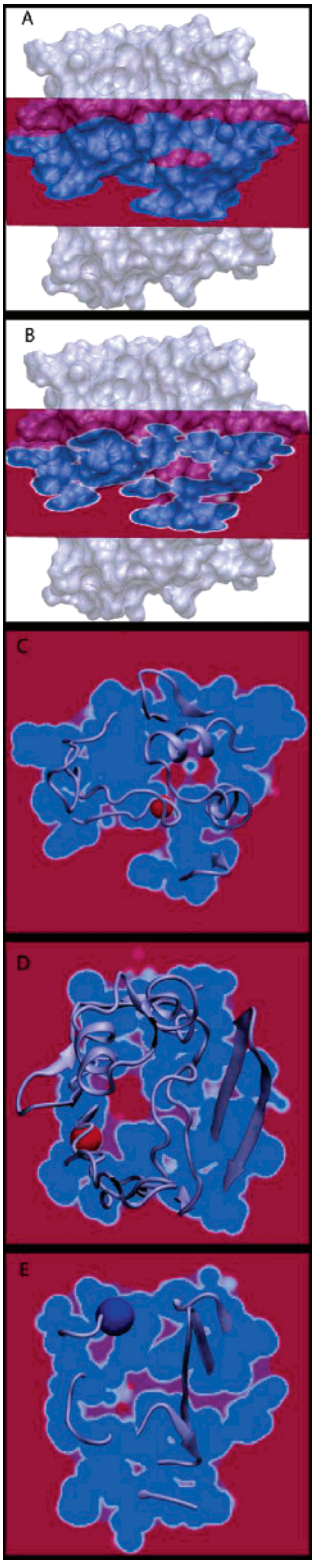

Figure 2. Dielectric maps of IFABP. Planes intersecting IFABP high dielectric ($\epsilon = 80$) regions in red, low dielectric ($\epsilon = 1$) in blue, and values between in white. Comparing the molecular surface (A) and the spline smoothed surface (B) shows the magnitude of IHDs created by the latter. A top-down view of both surfaces shows that IHDs (grey regions) are proximal to (C) tyr 70 OH, (D) ser109 OG, and (E) glu 85 N.

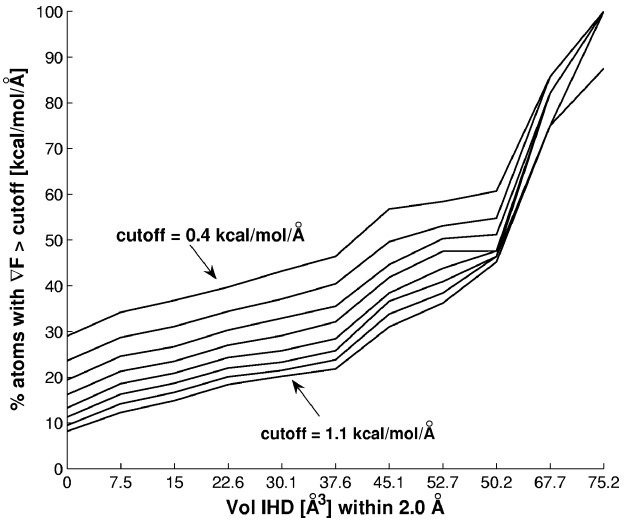

Figure 3. Percentage of IFABP atoms with large force deviations as a function of the volume of IHD ($\epsilon > 10$) within 2.0 Å of the atom's radial boundary. Deviation cutoff values ranged from 0.4 to 1.1 kcal/mol Å. All eight IFABP conformations were used with the force-optimized radii (scaled down for the molecular surface used to identify the IHDs).

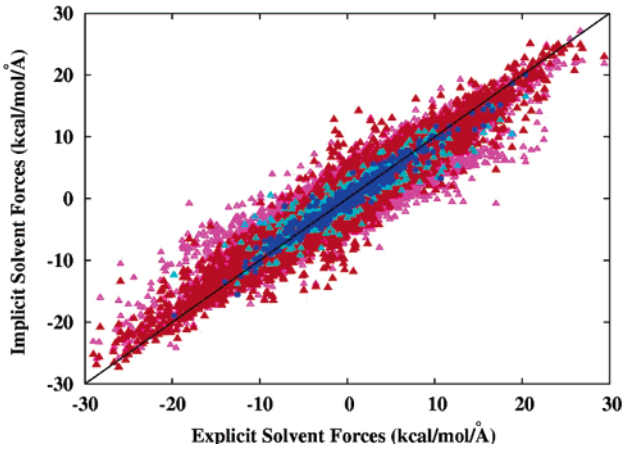

Figure 4. Comparison of explicit solvent and implicit solvent forces for IFABP and Trpcage/G-peptide showing a larger distribution of deviations for IFABP. Bondi radii are in pink (IFABP) and turquoise (Trpcage/G-peptide), while force optimized radii are in red (IFABP) and dark blue (Trpcage/G-peptide).

These findings point to a simple explanation for the different force-optimized solution sets: the IFABP-optimized radii were systematically increased to compensate for errors caused by large IHDs. The Trpcage and G-peptide-optimized radii were less affected; consequently, they perform better on polar solvation energies and forces for every tested molecule other than IFABP. They will be more widely applicable to compact systems but will not compensate for errors resulting from excessively large IHDs in globular solutes.

In retrospect, we believe that the formation of IHDs was also the source of differences between our previously presented "sharp-boundary" and spline-smoothed radii.[23] The sharp-boundary radii, fit to molecular surfaces, could not be scaled to perform as well on spline-smoothed surfaces as the spline-smoothed radii. Moreover, the spline-smoothed radii overestimated solvation energies for larger peptide systems because they were fit to small peptides. It is now clear that this overestimation was most likely caused by the spline-smoothed surfaces affecting the dielectric definition and resulting solvation forces of large solutes much more than those of small peptides.

It has been suggested that the width of the spline-smoothing window, *w*, can be increased to counteract the deleterious effects of IHDs.[19] Although it is not possible to eliminate them completely, their magnitude and dielectric value *can* be significantly decreased with larger smoothing windows. Unfortunately, this moves the dielectric boundary too far from surface atoms and creates unphysical bulges around overlapping and adjacent atoms, both of which result in underestimated solvation energies and forces. If it were possible to distinguish surface and buried atoms, then using larger radii or larger smoothing windows on buried atoms could significantly reduce the IHDs and related errors. For most globular systems, however, a clean distinction is impossible because many atoms are both exposed and buried. Thus, more involved modifications of the spline-smoothed surfaces, such as those proposed by Lu and Luo[53] and Lee et al.,[54] seem necessary for consistently accurate forces and energies. Despite these limitations, the surprisingly accurate solvation energies and forces in Table 4 suggest that spline-smoothed surfaces work quite well for compact systems.

Comparison to Other Parameter Sets. To gauge the accuracy of our optimized radii, we compared them to four other continuum parameter sets commonly used in conjunction with the AMBER force field. These included the AMBER *ff*99 charges combined with *ff*99 vdW radii,[30] Bondi radii,[31] and a set of radii recently published by Luo et al,[28] as well as PARSE charges combined with two variants of PARSE radii.[13] As previously mentioned, these radii were intended for use with molecular surfaces and so had to be scaled for use with the spline-smoothed surfaces. The scaling factors were determined with both force-only and force−energy-based genetic algorithm optimizations as described above for the radii presented in Table 1. With only one parameter to fit, these optimizations converged in one to five generations. Replica optimizations found the same scaling factors, indicating that the results were independent of the run parameters.

Two variants of the PARSE parameters were tested. They differed only in the value of the nonpolar hydrogen radius. The first set used 1 Å, as intended in the original publication, while the second used 0 Å, the standard value in PARSE parameter implementation. The first set ($R_{\mathrm{Hnp}} = 1$ Å) performed marginally better on both forces and energies, validating Sitkoff's intention for the PARSE radii to be the same as the standard Pauling vdW radii (with the exception of hydrogens which were decreased from 1.2 to 1.0 Å).[13]

The solvation energy deviations for the training set dipeptides and polyalanine peptides with different radii sets are shown in Figure 5. The deviations are not consistent across the various scaled radii, but they are consistently larger than the optimized radii deviations. A similar range of deviations was calculated for the test set peptides (see the

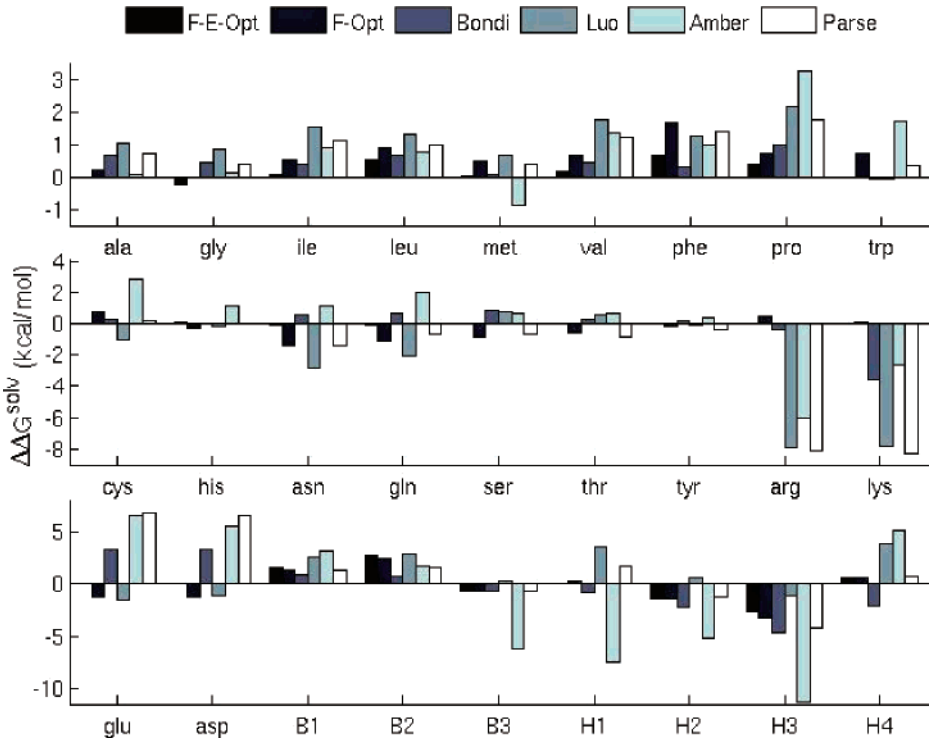

Figure 5. Solvation energy deviations for training set dipepetides (ala, gly, ... asp) and polyalanine peptides (B ≡ β-turn, H ≡ α-helix or β-hairpin) using the (F-E) force and energy optimized radii, (F) force optimized radii, Bondi, Luo, AMBER ff99, and PARSE parameters.

Supporting Information, Figure 7), but the comparison parameter sets (Bondi, Luo, AMBERff99, and PARSE) were inconsistently accurate for the two dipeptide conformations. A summary of the force and energy results is presented in Table 4. The scaling factors for the force and force−energy optimizations were significantly different for each of the parameter sets. Nevertheless, they followed several trends demonstrated by our optimized radii: the force-optimized radii performed better on forces and worse on energies than the force- and energy-optimized radii, the spread in deviations for IFABP were larger than those for Trpcage and G-peptide, and the solvation energy deviations for the amino acid dipeptides and polyalanine peptides in the test set were comparable or less than those in the training set. The only unexpected result was a larger solvation energy deviation for the proteinlike peptides with the Luo force and energy-optimized radii than the Luo force-optimized radii. Out of the comparison parameter sets, the AMBER and Luo radii performed best on forces while the Bondi radii performed best on solvation energies. As expected, our force and force−energy-optimized radii (presented in Table 1) performed better on forces and energies than all of the comparison sets.

Scaled Radii for Different Spline Window Half-Widths and Molecular Surfaces. The spline window half-width used in our optimizations ($w = 0.3$ Å) will not be optimal for every problem or be chosen by every user. It has been shown, however, that increasing or decreasing $w = 0.3$ Å radii by a window half-width-dependent scaling factor recovers accurate solvation energies.[21] As described in the Methods, we determined the best scaling factors, x_{best}, according to eq 14 for a range of spline window half-widths ($w = 0.2$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 Å) with force- and force-and-energy-based optimizations. The optimizations converged in one to five generations, and replica optimizations found the same scaling factors, indicating that the results were independent of the run parameters. The final results are presented in Table 5 and plotted in Figure 6. The scaling values for the force-optimized and the force-and-energy-optimized radii are well fit by quadratic polynomials, $x = -0.0518w^2 + 0.5043w - 0.1475$ and $x = -0.2078w^2 + 0.667w - 0.1823$, respectively. These equations can be used to approximate x_{best} (eq 14) for any spline window half-width but will likely yield unreliable results for widths less than 0.1 Å or larger than 1.1 Å.

Table 5. Scaling Factors [x_{best}] for Different Spline Window Half-Widths [w] for (a) Force and (b) Force−Energy-Optimized Radii

	scaling factor x_{best} [Å]	
spline half-width w [Å]	force[a]	force−energy[b]
0.2	−0.051	−0.058
0.3	0.000	0.000
0.4	0.051	0.051
0.5	0.093	0.099
0.6	0.129	0.143
0.7	0.178	0.182
0.8	0.225	0.218
0.9	0.267	0.250
1.0	0.303	0.277

If only solvation energies are desired, molecular surfaces can be used to avoid artifacts introduced by IHDs in spline-smoothed surfaces. Although our optimized radii are not ideal, they can be scaled *down* for molecular surfaces. This

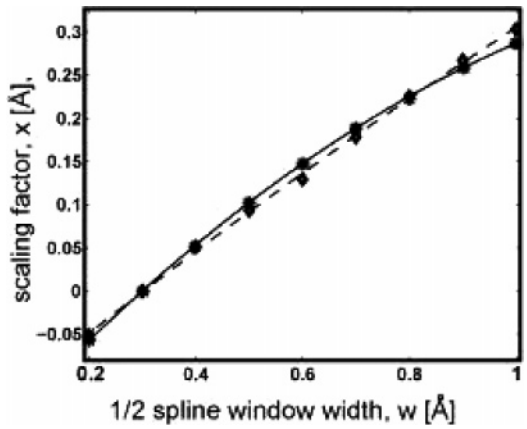

Figure 6. Scaling factors for different spline window half-widths. The force results (diamonds) are best fit by $x = -0.0518w^2 + 0.5043w - 0.1475$ (dashed). The force−energy results (stars) are best fit by $x = -0.2078w^2 + 0.667w - 0.1823$ (solid).

is analogous to scaling the comparison parameter set's molecular surface radii *up* (i.e., made larger) for spline-smoothed surfaces. As described in the Methods, we used eq 14 to scale our force-only and our force-and-energy-optimized radii (Table 1) for both abrupt and harmonically smoothed molecular surfaces. The scaling values and resulting solvation energies (see Supporting Information) indicate that these radii work well for molecular surfaces but not as well as they work for spline-smoothed surfaces.

V. Conclusions

We have presented two sets of radii optimized to define spline-smoothed dielectric boundaries in Poisson electrostatics calculations on proteins defined with AMBER *ff*99 partial charges. These optimizations were based on explicit solvent simulations and, thus, offer forces and energies consistent with explicit solvent mean polar forces and charging free energies. More importantly, we have presented a method of optimizing the Poisson dielectric boundary on the basis of atomic forces, which is substantially more efficient and widely applicable than previous optimization approaches. It can be used, for example, on highly charged, large, or unusual systems for which experimental solvation energies are unavailable and explicit solvent charging free energies are computationally challenging or prohibitive.

We have shown that optimizations based on forces alone generate radii that reproduce forces accurately but solvation energies less accurately. Optimizations based on both forces and energies, on the other hand, decrease the accuracy of forces minimally and increase solvation energy accuracy. Thus, our force−energy-optimized radii are more appropriate than our force-optimized radii when both energies and forces are important. As anticipated, our optimized radii perform better than four other parameter sets commonly used in conjunction with the AMBER force field, each of which was scaled for use with spline-smoothed dielectric boundaries. Of these comparison sets, the AMBER and Luo radii perform best on forces while the Bondi radii perform best on energies.

Finally, we have shown that atom-centered dielectric functions, such as the spline-smoothed surface used in our optimizations, form high dielectric regions in interstitial spaces, thereby limiting the accuracy of solvation forces and energies. These errors seem less pronounced in compact systems, where the IHDs fill the crevices between atoms, than in globular systems, where buried IHDs can be copious. Thus, our optimized radii, and spline-smoothed surfaces in general, are expected to work well for compact systems but will produce errors in highly solvated globular proteins. Efforts to correct atom-centered, smooth surfaces will be crucial for the continued development of Poisson-based implicit solvent models, especially in the area of continuum dynamics.

It is important, however, to appreciate the complexity of the problem at hand. The specific nature of the optimal dielectric boundary most rigorously depends on the locally defined solute−solvent potential,[34,35] a complex function of solute geometry (i.e., surface curvature) and electrostatic and nonelectrostatic solvent−solute interactions. Thus, theory and our analysis agree that an optimal continuum radius cannot be defined by atom type alone. Consequently, we anticipate that corrected smooth surfaces will reduce inaccuracies and inconsistencies but that it will take a more physically accurate implicit solvent framework which accounts for polar and nonpolar coupling to predict accurate solvation forces and energies for disparate solutes and solute conformations without system- or conformation-specific parameters.

Acknowledgment. J.M.J.S thanks John Mongan for invaluable discussion and the NSF-sponsored Center for Theoretical Biological Physics for financial support (Grant Nos. PHY-0216576 and PHY-0225630). N.A.B. and J.A.W. were supported by NIH Grant GM069702 and an Alfred P. Sloan Research Fellowship (N.A.B.). Work in the McCammon group is additionally supported by NIH, NSF, HHMI, NBCR, and Accelrys Inc.

Supporting Information Available: Additional figures of test set continuum solvation energy deviations and molecular surface results. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1−20.

(2) Rashin, A. A.; Honig, B. Reevaluation of the Born Model of Ion Hydration. *J. Phys. Chem.* **1985**, *89*, 5588−5593.

(3) Roux, B.; Yu, H. A.; Karplus, M. Molecular-Basis for the Born Model of Ion Solvation. *J. Phys. Chem.* **1990**, *94*, 4683−4688.

(4) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy=Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J. Phys. Chem.* **2000**, *104*, 6271−6285.

(5) Wagoner, J. A.; Baker, N. A. Assessing Implicit Models for Nonpolar Mean Solvation Forces: The Importance of Dispersion and Volume Terms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331−8336.

(6) Tanford, C.; Kirkwood, J. G. Theory of Protein Titration Curces. I. General Equations for Impenetratable Spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.

(7) Chothia, C. Hydrophobic Bonding and Accessible Surface Area in Proteins. *Nature (London)* **1974**, *248*, 338−339.

(8) Spolar, R. S.; Ha, J. H.; Record, M. T., Jr. Hydrophobic Effect in Protein Folding and Other Noncovalent Processes Involving Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8382−8385.

(9) Liu, T.; Ye, L.; Chen, H.; Li, J.; Wu, Z.; Zhou, R. A Combined Steepest Descent and Genetic Algorithm (SD/GA) Approach for the Optimization of Solvation Parameters. *Mol. Simul.* **2006**, *32*, 427−435.

(10) Davis, M. E.; McCammon, J. A. Electrostatics in Biomolecular Structure and Dynamics. *Chem. Rev.* **1990**, *90*, 509−521.

(11) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.

(12) Baker, N. A. Improving Implicit Solvent Simulations: A Poisson-Centric View. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137−143.

(13) Sitkoff, D.; Sharp, K.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solbent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(14) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute−Solvent Interaction Energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523−9530.

(15) Lee, B.; Richards, F. M. Interpretation of Protein Structures − Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379.

(16) Connolly, M. L. Computation of Molecular Volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118−1124.

(17) Bordner, A. J.; Cavasotto, C. N.; Anagyan, R. A. Accurate Transferable Model for Water, *n*-Octanol, and *n*-Hexadecane Solcation Free Energies. *J. Phys. Chem.* **2002**, *106*, 11009−11015.

(18) Nina, M.; Simonson, T. Molecular Dynamics of the tRNA-(Ala) Acceptor Stem: Comparison between Continuum Reaction Field and Particle-Mesh Ewald Electrostatic Treatments. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.

(19) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson−Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608−640.

(20) Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: Computation of Electrostatic Forces from Numerical Solutions to the Poisson−Boltzmann Equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(21) Nina, M.; Im, W.; Roux, B. Optimized Atomic Radii for Protein Continuum Electrostatics Solvation Forces. *Biophys. Chem.* **1999**, *78*, 89−96.

(22) Wagoner, J.; Baker, N. A. Solvation Forces on Biomolecular Structures: A Comparison of Explicit Solvent and Poisson−Boltzmann Models. *J. Comput. Chem.* **2004**, *25*, 1623−1629.

(23) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. Optimized Radii for Poisson−Boltzmann Calculations with the AMBER Force Field. *J. Chem. Theory Comput.* **2005**, *1*, 484−493.

(24) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. Effective Force Fields for Condensed Phase Systems from ab Initio Molecular Dynamics Simulation: A New Method for Force-Matching. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(25) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(26) Cramer, C. J.; Truhlar, D. G. General Parameterized Scf Model for Free-Energies of Solvation in Aqueous-Solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(27) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *J. Chem. Theory Comput.* **2005**, *1*, 1133−1152.

(28) Lwin, T. Z.; Zhou, R. H.; Luo, R. Is Poisson−Boltzmann Theory Insufficient for Protein Folding Simulations? *J. Chem. Phys.* **2006**, *124*.

(29) Nina, M.; Beglov, D.; Roux, B. Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(30) Wang, J. M.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(31) Bondi, A. *J. Chem. Phys.* **1964**, *64*, 441.

(32) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987.

(33) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: London, 1976.

(34) Dzubiella, J.; Swanson, J. M. J.; McCammon, J. A. Coupling Nonpolar and Polar Solvation Free Energies in Implicit Solvent Models. *J. Chem. Phys.* **2006**, *124*.

(35) Dzubiella, J.; Swanson, J. M. J.; McCammon, J. A. Coupling Hydrophobicity, Dispersion, and Electrostatics in Continuum Solvent Models. *Phys. Rev. Lett.* **2006**, *96*.

(36) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(37) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(38) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein-G. *Science* **1991**, *253*, 657−661.

(39) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(40) Hodsdon, M. E.; Cistola, D. P. Ligand Binding Alters the Backbone Mobility of Intestinal Fatty Acid-Binding Protein as Monitored by N-15 NMR Relaxation and H-1 Exchange. *Biochemistry* **1997**, *36*, 2278−2290.

(41) Lindahl, E.; Hess, B.; vand der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm − A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(43) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System − Solvent Boundary Potential for Computer-Simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(44) Deng, Y. Q.; Roux, B. Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.

(45) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(46) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004**, *121*, 10096−10103.

(47) Hochtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. Rationalization of the Dielectric Properties of Common Three-Site Water Models in Terms of Their Force Field Parameters. *J. Chem. Phys.* **1998**, *109*, 927−4937.

(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(49) Baker, J. E. Reducing Bias and Inefficiency in the Selection Algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic Algorithms and their Application*; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, 1987; pp 14−21.

(50) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, 1297−1304.

(51) Masunov, A.; Lazaridis, T. Potentials of Mean Force between Ionizable Amino Acid Side Chains in Water. *J. Am. Chem. Soc.* **2003**, *125*, 1722−1730.

(52) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. Limitations of Atom-Centered Dielectric Functions in Implicit Solvent Models. *J. Phys. Chem. B* **2005**, *109*, 14769−14772.

(53) Lu, Q.; Luo, R. A Poisson−Boltzmann Dynamics Method with Nonperiodic Boundary Condition. *J. Chem. Phys.* **2003**, *119*, 11035−11047.

(54) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations (vol 24, pg 1348, 2003). *J. Comput. Chem.* **2003**, *24*, 1348−1356.

CT600216K