Evaluation of Poisson Solvation Models Using a Hybrid Explicit/Implicit Solvent Method

Michael S. Lee*,†,‡ and Mark A. Olson†

Department of Cell Biology and Biochemistry, U.S. Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Frederick, Maryland 21702, and Computational and Information Sciences Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland 21005

Received: August 11, 2004; In Final Form: January 12, 2005

Implicit solvent methods have become popular tools in the field of protein dynamics simulations, yet evaluation of their validity has been primarily limited to comparisons with experimental and theoretical data for small molecules. In this paper, we use a recently developed hybrid explicit/implicit solvent methodology to evaluate the accuracy of several Poisson-based implicit solvent models. Specifically, we focus on the calculation of electrostatic solvation free energies of various fixed conformations for two proteins. We show that, among various dielectric boundary definitions, the Lee−Richards molecular surface has the best agreement with hybrid solvent results. Furthermore, certain modifications of the molecular surface Poisson protocol provide varied results. For instance, simple modifications of atomic radii on charged residues generally improve absolute errors but do not significantly reduce relative errors among conformations. On the other hand, using a water-probe radius of 1.0 Å, as opposed to the standard value of 1.4 Å, to generate the molecular surface, moderately improves both absolute and relative results.

I. Introduction

Implicit solvent methods[1−7] have emerged as important tools for biomolecular simulations.[8,9] They have been applied to ab initio protein structure prediction,[10−12] protein pK_a determination,[3,4,13] mutational free-energy computations,[14−16] binding free-energy calculations,[17,18] and general simulation studies.[19] Implicit solvent simulations are advantageous because they are often significantly faster than explicit solvent calculations.[20,21] Furthermore, complications regarding convergence of the solvent degrees of freedom are virtually eliminated.[4]

Nonetheless, a key question regarding implicit solvent models is, what is their accuracy relative to the more expensive explicit solvent simulations?[22,23] One concern is that implicit solvent models have fundamental limitations because local hydrogen-bonding interactions between surface protein atoms and water molecules are lacking.[24] Also, when implicit models are used in a molecular dynamics (MD) simulation, distortions from the original X-ray structure are evident.[19,25] Moreover, the fine details in protein conformational landscapes are smoothed out.[26,27] It is unclear if implicit solvent models are sufficiently accurate for certain difficult applications such as calculations of ligand-binding free energy[18] and protein pK_a's.[28]

In this work, we focus on one particular class of implicit solvent models known as Poisson models.[1] Poisson solvent models are considered the gold standard for simpler implicit solvent approaches such as generalized Born (GB) models.[29] Alternative constructs such as the Langevin dipole[4] and screened Coulomb potential[5] methods may have comparable accuracies; however, they are not considered here. The basis behind Poisson methods for aqueous systems is that the solvent is represented as a high-dielectric continuum, typically modeled as $\epsilon = 80$, and the solute is characterized as a low-dielectric material, for example, $\epsilon = 1$. The Poisson method involves an iterative numerical solution of the Poisson equation to obtain an electrostatic potential map and subsequently an electrostatic solvation free energy. The Poisson algorithm is useful for estimating the solvation energy[1] and electrostatic characteristics[30] of a fixed solute conformation. In addition, fast analytical formulations have facilitated the use of Poisson theory in simulation techniques such as MD and Monte Carlo.[31,32]

The absence of local interactions in the Poisson model can be compensated somewhat through the modification of atomic van der Waals (vdW) radii.[33,34] Another critical issue is determining which dielectric boundary definition will lead to the most accurate results. Most researchers utilize the molecular surface (MS) definition first proposed by Lee and Richards.[35] This definition has physically realistic assumptions about the possible locations of solvent molecules. The MS definition is similar to the vdW volume definition yet removes interstitial gaps between atoms where water molecules are unlikely to sample. However, the discontinuous nature of the MS dielectric boundary is problematic for theoretical and practical reasons.[31] Alternatively, molecular volumes, built up as a superposition of atomic functions, provide smooth solute−solvent boundaries; however, they often lack an optimal treatment of the interstitial space between atoms.[36] A recently devised superposition definition, which has been applied to GB theory,[36] attempts to mimic the salient features of the MS while retaining an analytical formulation.

Given various molecular boundary definitions, it is difficult to improve upon implicit solvent theory without sufficient experimental or theoretical benchmarks to gauge accuracy. Experimental benchmarks might include protein residue pK_a's,[37] ligand binding,[18] and mutational free energies.[15] However, a comparison to the experiment is often obfuscated by three factors: the inability to generate a convergent statistical ensemble of solute conformations,[38] the difficulty in estimating entropic contributions, and the fundamental accuracy limitations of the force field. On the other hand, there has been little attempt in the literature thus far to compare implicit solvation energies of fixed protein conformations and potential energy landscapes

* Author to whom correspondence should be addressed. E-mail: michael.lee@amedd.army.mil.
† U.S. Army Medical Research Institute of Infectious Diseases.
‡ U.S. Army Research Laboratory.

10.1021/jp046377z This article not subject to U.S. Copyright. Published 2005 by the American Chemical Society
Published on Web 02/19/2005

against explicit solvent approaches.[24,39−41] Zhang et al. compare implicit and explicit approaches for calculating the binding free energy of a peptide to a rigid protein[39] and find reasonable agreement between the two approaches. Perhaps the scarcity of studies such as this one is due to the fact that it is believed that thermodynamic quantities such as the solvation energy are slowly convergent with respect to simulation time. Furthermore, the number of water molecules necessary to achieve convergence with respect to the system size is large without an appropriate reaction field to compensate for the bulk limit. In addition, charging free-energy calculations, which are the precursor to obtaining the solvation energy, are technically difficult with the commonly used periodic boundary techniques because artificial real/virtual solute interactions must somehow be removed.[42] Alternatively, finite water-cluster methods that use spherical boundaries[43,44] are more suitable to calculate the solvation energies of fixed protein structures with a reasonable degree of precision. The spherical constraint on the simulation volume, however, can become very costly and wasteful for extended protein structures.

We introduce a hybrid explicit/implicit solvent approach,[45] which consists of a solute encapsulated by a modestly sized layer of water molecules (∼10 Å). A GB model is used to treat the reaction field of the irregularly shaped simulation volume. In our current implementation, the simulation volume is held fixed, which is adequate for the charging free-energy calculation of a fixed solute. Furthermore, our approach employs a multigrid technique that greatly reduces the computational cost of the expensive pairwise electrostatic and GB terms. Because of the multigrid algorithm and the need for a relatively small number of explicit water molecules, the hybrid solvent method provides an efficient means for calculating the solvation free energies of protein conformations. In this work, we use our hybrid solvent method to compute solvation free energies for two sets of protein conformations and utilize these values as a benchmark for various Poisson-based implicit solvent protocols. The accuracy of the hybrid approach compared to those of fully explicit solvent methods has yet to be determined. A preliminary study on the accuracy of the hybrid method versus that of the established particle mesh Ewald procedure has been reported elsewhere[76]. Furthermore, in this work, we make comparisons with another established hybrid solvent approach for the calculation of charging free energies in solution. Nonetheless, the hybrid approach will require extensive further testing to ensure that it is a generally applicable and reliable method for biomolecular simulations.

II. Methods

Hybrid Solvent Method. The benchmark explicit solvent calculations in this work were performed using a recently developed hybrid explicit/implicit solvent method. We briefly describe the method here and defer detailed discussion of the multigrid enhancements to another work.[45] The basic idea of our method, depicted schematically in Figure 1, is to restrain several layers of water molecules to a simulation volume that uniformly encapsulates the solute. The infinite expanse outside the simulation volume is approximated as a dielectric continuum using GB theory. The GB model allows us to accurately calculate the implicit solvent contribution for an irregularly shaped simulation volume. Most prior hybrid solvation schemes are limited to simulation volumes with simple geometric shapes, such as spheres and rectangular prisms.[43,46,47]

Specifically, as can be seen in Figure 1, the simulation volume consists of interlocking spheres, resembling the shape of the initial conformation of the solute.[48,49] Enlarged spheres of radius

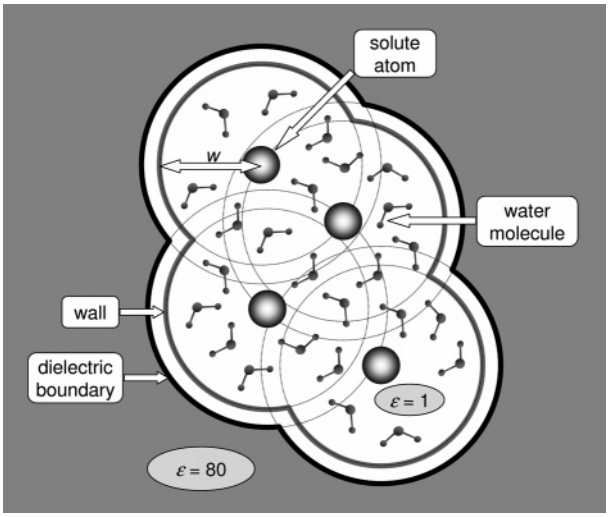

Figure 1. Schematic of the hybrid explicit/implicit solvent approach. The simulation volume consists of a collection of solute atoms and a layer of water molecules. The reaction field associated with the bulk dielectric (gray outer region) boundary is treated by a GB approximation. The light circles are used in defining the wall and dielectric boundaries, but they otherwise play no physical role in a simulation. The exact extents of the wall and dielectric boundary as a function of the layer width, *w*, are specified in the text.

w are constructed around each heavy atom and joined together to encapsulate the explicit solvent molecules. To ensure a smooth transition between spheres of different atoms, we developed a volume function similar to the one proposed by Im et al.[50] Our volume function, $V(\vec{r})$, is defined such that $V = 1$ signifies the internal region and $V = 0$ signifies the external region. The volume function resulting from a superposition of spherical functions around each atom, i, positioned at $\vec{x}_i$ has the following mathematical form:

$$V(\vec{r}) = 1 - \prod_i \left[1 - v\left(\frac{|\vec{r} - \vec{r}_i| - w_i^0}{w_i^1 - w_i^0}\right)\right] \qquad (1)$$

where

$$v(s) = \begin{cases} 1 & s \le 0 \\ 1 - 6s^5 + 15s^4 - 10s^3 & 0 < s < 1 \\ 0 & s \ge 1 \end{cases} \qquad (2)$$

The width parameters, w_i^0 and w_i^1, define the radial end points of the spherical function on atom i. The polynomial in eq 2 has continuity up to second derivatives at its piecewise end points.

The wall potential and the dielectric volume are both defined in terms of the simulation volume. The wall potential is a repulsive boundary that confines the solute and the explicit water molecules to the simulation volume and is defined as a function of the simulation volume:

$$E_{\text{wall}}(\vec{r}) = \frac{1}{[E_{\text{wall}}^{\max}]^{-1} + V(\vec{r})} - \frac{1}{[E_{\text{wall}}^{\max}]^{-1} + 1} \qquad (3)$$

such that $V = 1$ maps to $E_{\text{wall}}(\vec{r}) = 0$ and $V = 0$ maps to $E_{\text{wall}}(\vec{r}) \approx E_{\text{wall}}^{\max} - 1$, where $E_{\text{wall}}^{\max} = 30$ kcal/mol in this work. Because we use a finite value for $E_{\text{wall}}^{\max}$, the wall potential is actually penetrable. This turns out to be useful at the beginning of a simulation when some water molecules are initially placed outside the simulation boundary. As a MD simulation proceeds, these excess water molecules can proceed beyond the simulation

volume without incurring unphysically large energies and forces. After a short period of dynamics, these excess water molecules can be deleted from the simulation if so desired. For this study, the width parameters for the wall potential were derived empirically: $w_i^0(\text{wall}) = w - 0.6$ Å and $w_i^1(\text{wall}) = w + 0.4$ Å. The wall parameters and dielectric values described below were derived by trial and error to obtain a reasonable bulk water radial density profile in a sphere.[45] Roughly speaking, a wall potential is analogous to the vdW repulsion term of the bulk solvent. It follows that one should also include an attractive component.[46] However, we did not incorporate an attractive wall potential because our initial tests indicated that such a potential would have only a relatively small impact.[45]

Furthermore, we neglected to incorporate a surface angular polarization term that would enhance isotropy of the water molecules near the edge of the simulation volume.[43,46,51] Our own studies (not shown) and evidence from the work of others[44,52] suggest that, regardless of the attempts to force angular isotropy on the water molecules near the boundary, surface dipolar artifacts still occur. The naïve approach for calculating the charging free energy (CFE) of a charged solute in a cluster of water molecules produces values that are ~10−15 kcal/mol different from the CFEs obtained by the charge-summation (or P-summation) approach applied to properly corrected particle mesh Ewald calculations.[44] Nonetheless, the alternative molecule-based summation (M-summation) procedure applied to periodic boundary calculations leads to CFE values closer to the ones obtained by hybrid cluster methods (results not shown). Controversy in the literature still exists over which procedure, P summation or M summation, produces the correct values.[53,54] Comparisons with experimental data seem to be inconclusive so far because, for cations Na^+ and K^+, P summation is in good agreement with experimental data.[52] However, for anions such as Cl^-, M summation appears to be in better agreement. Darden et al.[44] proposed a procedure that brings spherical cluster methods into good agreement with P-summation approaches for atomic ions. However, extension of their approach to nonspherical systems is beyond the scope of this work. Darden and co-workers[44] also suggest that, because differences between standard and P-summation CFEs are relatively constant for a specific cluster size, a simple constant term may be sufficient. In this work, we did not modify calculated CFEs for the sake of simplicity. However, it is understood that absolute CFEs may be shifted by a product of the net charge of the solute with 10−15 kcal/(mol·atomic charge).

The implicit solvent model in the hybrid approach utilizes a dielectric volume that has the same form as that of the simulation volume with empirically derived width parameters of $w_i^0(\text{dielectric}) = w + 1.8$ Å and $w_i^1(\text{dielectric}) = w + 2.2$ Å.[45] The boundary of the dielectric volume is approximately 2 Å farther from the spherical centers than the wall potential. This distance is roughly in line with the vdW radius of TIP3P oxygen (~1.8 Å). The GB energy for a hybrid simulation is defined by a modified form of the Still equation, which has properties similar to those of the original formulation[2]

$$E_{\text{solv}} = -k\left(1 - \frac{1}{\epsilon}\right)\sum_{ij}\frac{q_i q_j}{r_{ij} + \frac{1}{2}(\alpha_i + \alpha_j)\exp[-2r_{ij}/(\alpha_i + \alpha_j)]} \quad (4)$$

where $k \sim 166.0$ (kcal/mol)·Å/(atomic charge)^2, ϵ is the dielectric boundary of the solvent (in this work, $\epsilon = 80$), i and j are indices over all of the atoms, q_i is the partial charge of atom i, r_{ij} is the distance between atoms i and j, and α_i is the Born radius of atom i. Equation 4 has two beneficial properties that the original Still formula lacks. First, experiments in our lab have determined that eq 4 provides results closer to those of analytical Poisson theory for the special case of random points inside a sphere[46] (results not shown). Second, unlike the original Still formula, eq 4 requires no extra square root evaluations after the electrostatic term has been computed. In this work, we assumed a fixed simulation and dielectric volume for the entire MD run. This simplification allowed us to precalculate Born radii on a grid of points labeled m using the accurate empirical approximation[36,55]

$$\alpha_m^{-1} = \left(1 - \frac{1}{\sqrt{2}}\right)\left(R^{-1} - \frac{1}{4\pi}\int\int_R^\infty \frac{V_{\text{dielectric}}(\vec{r})}{|\vec{r} - \vec{x}_m|^4}\,dr\,d\Omega\right) + \left(\frac{1}{4R^4} - \frac{1}{4\pi}\int\int_R^\infty \frac{V_{\text{dielectric}}(\vec{r})}{|\vec{r} - \vec{x}_m|^7}\,dr\,d\Omega\right)^{1/4} \quad (5)$$

where $V_{\text{dielectric}}(\vec{r})$ is the dielectric volume defined in eq 1 and R is an asymptotic atomic radius set to 1.5 Å, which is approximately the average Born radius of the oxygen and hydrogen of an isolated TIP3 water molecule. Equation 5 provides Born radii that are very highly correlated with analytically determined self-energies of random points inside a sphere using the Kirkwood solution to the Poisson model[46] (results not shown). However, the Born radii are linearly transformed, $\alpha_m' = 1.0649\alpha_m - 0.316$, to obtain a better fit to Poisson solvation self-energy results. Note that when a water molecule pierces through the dielectric volume, it does not obtain an infinite solvation energy as it does in other Poisson-based spherical cluster models.[46] Instead, it gains a finite solvation energy related to our specified asymptotic Born radius, R. Thus, in our method, a water molecule can escape the simulation volume without incurring unphysical energies or forces.[46] The precalculated Born radii grid has a cubic cell size of 2 Å extending over a rectangular region encompassing the entire simulation volume. With this grid, the Born radius of each atom is obtained rapidly by means of cubic interpolation from nearest-neighbor grid points.[45,56] The numerical integration procedure to obtain Born radii for each grid point m from eq 5 is described elsewhere.[36,55]

The most computationally intensive components of this method are the nonbonded energy terms: Lennard-Jones, electrostatic, and GB. A commonly employed technique in molecular mechanics to reduce the computational effort is a distance cutoff. This is usually appropriate for the Lennard-Jones term, which diminishes quickly as the inverse sixth power of distance. The pairwise electrostatic and GB terms, nonetheless, have much slower inverse distance decay. This problem is most evident for ionized functional groups whereby large cutoffs are required to obtain converged results. To deal with this problem, a variety of methods have been devised such as the local reaction field,[57] fast multipoles,[58] and multigrid methods.[56] Because the multigrid method is more flexible than the multipole-based methods, we extended the multigrid method of Skeel et al.[56] to incorporate the GB potential.[45] The general framework of the multigrid method for pairwise interactions is to divide the Coulomb potential into local and soft components.

$$K(r) = K_{\text{soft}}(r) + K_{\text{local}}(r) = \frac{1}{r} \quad (6)$$

The local component, which is defined by a distance cutoff criterion, is performed exactly. The soft component is performed on a hierarchy of charge and potential grids, where the appropriate resolution of the grid becomes coarser as the distance increases. The highest-resolution charge grid is built by cubic

interpolation of the atomic charges. The next-highest-resolution grid is obtained by cubic interpolation of the highest-resolution charge grid, and so on.

Our extension of the method of Skeel et al.[56] involves several components.[45] First, we use a different soft Coulomb interaction kernel that becomes completely flat at half the short-range cutoff, a.

$$K^0_{\text{soft}}(r) = \begin{cases} \frac{18}{3a^5}\left(r - \frac{a}{2}\right)^4 - \frac{16}{3a^4}\left(r - \frac{a}{2}\right)^3 + \frac{31}{24a}, & \frac{a}{2} < r < a \\ \frac{31}{24a}, & r < \frac{a}{2} \end{cases} \tag{7}$$

This flat region makes it easier to remove the artifacts associated with the multigrid treatment of excluded 1−2, 1−3, and 1−4 interactions. Equation 7 is also continuous up to second derivatives.[45] The GB potential is split into local and soft components merely by replacing r_{ij} in eq 4 with the inverses of the local and soft kernels, respectively. The Born radii for the multigrid evaluation of the soft component are derived by cubic interpolation of the Born radii grid. In this work, we used a short-range cutoff, a, equal to 8 Å, unless otherwise specified. This cutoff value provides accuracy comparable to that of a standard distance-based truncation cutoff of 26 Å.[45] Further details of our multigrid implementation can be found elsewhere.[45]

Simulation Protocol. All of the MD simulations in this work were performed with the CHARMM program[59] using the PARAM22 empirical force-field potential.[60] The equations of motion were integrated with a Langevin dynamics algorithm at constant temperature (298 K) using a friction constant of 5 ps^{-1} on the TIP3 oxygen atoms and an integration time step of 2 fs. Covalent bonds between the heavy atoms and hydrogens were constrained by the SHAKE algorithm.[61] The Lennard-Jones potential was truncated at the same radius as that of the local electrostatics term, using a 1-Å-wide potential switching function.

The general procedure for adding explicit water molecules to a solute consists of two steps: (1) determining the number of water molecules needed to fill the simulation volume and (2) carving out these water molecules from a large block. The number of desired water molecules is calculated by first determining the water simulation volume (WSV), which is equal to the difference between the expanded solute volume, where each heavy atom radius is set to the specified width, w, and the standard solute vdW volume. Molecular volumes in this study were determined using the *coor volume* command in CHARMM. The desired number of water molecules equals the WSV multiplied by the standard bulk water density, 0.0334 $Å^{-3}$. A large cubic box of water molecules is overlaid onto the solute, and water molecules that are less than 3.1 Å from a heavy atom in the solute are deleted. This is followed by an iterative procedure in which water molecules beyond a certain cutoff distance from all of the solute's heavy atoms are deleted. The procedure repeats the deletion with decreasing cutoffs until the number of water molecules remaining is less than or equal to the desired number of water molecules. Often, in actual system setups, after the deletion procedure, some water molecules will be initially positioned outside the wall potential. Because of the finite nature of the wall potential, these water molecules will drift away during a simulation. Thus, a large spherical boundary potential, using the *MMFP* module in CHARMM, is placed around the system to prevent the unbounded drift of water molecules that initially escape from the smaller finite boundary of eq 3. These escaped water molecules do not significantly impact the free-energy calculation because their small Born radii (<1.5 Å) lead to greatly attenuated electrostatic interactions with the solute. In theory, some of the excess water molecules could return into the simulation volume because there is no barrier. In practice, with a large spherical boundary potential, this effect appears to be negligibly small for the simulation time scales used in this work.

To validate the hybrid layer approach as a means of obtaining the CFEs in solution, we also performed spherical simulations with either the hybrid method or the spherical solvent boundary potential (SSBP) method[46] of Beglov and Roux. The validation CFE studies were run with the same parameters as those of the hybrid layer CFE studies save for the spherical simulation setups, which are specified as follows. First, the center of mass of the solute is translated to the origin. For the hybrid sphere method, a fixed dummy atom, used to define the center of the simulation volume, is added and positioned at the origin. The simulation sphere radius, R_{sphere}, is calculated as a sum of the specified layer width, w, and the maximal distance between a solute atom and the origin. This simple protocol could be improved because it does not provide the optimally smallest simulation sphere that can encompass the solute and still provide, at minimum, a width, w, of water molecules.[62] After R_{sphere} is obtained, the total simulation volume is computed as $(4/3)\pi R_{\text{sphere}}^3$, while the WSV and target number of water molecules are obtained analogously to the layer setup (see above). At this point, the SSBP and hybrid sphere setups diverge. The hybrid method assumes a fixed specified radius, R_{sphere}. Once again, several water molecules may initially escape the finite boundary potential. To achieve a fair comparison to the SSBP method, we remove these escaped water molecules using a simple corrective protocol. The procedure consists of running 0.4 ps of dynamics followed by the deletion of all of the water molecules outside the boundary wall, that is, those with an oxygen distance to the origin of greater than w_i^1(wall). This method eliminates most, if not all, of the water molecules that initially had drifted out of the simulation volume. The SSBP method, on the other hand, requires no such corrective procedure because it dynamically defines its radius as the maximal distance between any water molecule and the origin.

The SSBP method is run with default parameters,[46] except that the number of Kirkwood multipoles is set to 20 versus the default value of 15. This modification improves the accuracy of the dielectric continuum component. The SSBP method is augmented with the extended electrostatics method[63] so that a small, short-range electrostatic cutoff can be employed.

Calculation of the Electrostatic CFE in Solution. We are concerned with calculating the CFE in solution, ΔG_{CFE}, of an arbitrary solute in a fixed conformation. This quantity is simply the sum of the electrostatic solvation free energy, $\Delta G^{\text{elec}}_{\text{solv}}$, and the vacuum solute electrostatic energy, $E^{\text{elec}}_{\text{solute}}$,

$$\Delta G_{\text{CFE}} = \Delta G^{\text{elec}}_{\text{solv}} + E^{\text{elec}}_{\text{solute}} \tag{8}$$

The CFE in solution is a more demanding benchmark than solvation energy because the electrostatic solvation energy has a high negative correlation to the solute electrostatic energy.[64] Besides, the CFE in solution is often the relevant quantity for thermodynamic studies.

To obtain ΔG_{CFE}, we calculate the free energy necessary to transform the solute from uncharged (state $\lambda = 0$) to fully charged (state $\lambda = 1$), using the thermodynamic integration (TI) formula[65]

$$\Delta G_{\mathrm{CFE}} = \int_0^1 \langle \partial E/\partial\lambda \rangle \, \mathrm{d}\lambda \tag{9}$$

In our approach, n simulation windows are run, corresponding to values $\lambda = 1/2n$, $3/2n$,..., $(2n - 1)/2n$. In each window, the force term, $\partial E/\partial\lambda$, is statistically averaged. Then, a straight line, $y = a_0 + a_1\lambda$, is fit through the force terms to arrive at an electrostatic CFE, $\Delta G_{\mathrm{CFE}} = a_0 + a_1/2$. We justify the use of a linear fit in the Results section.

Implicit Solvent Protocol. We compared the solvation energies obtained from the hybrid method to a commonly used implicit solvation scheme, the Poisson model. Calculation of the solvation energy of a molecule using the Poisson model can be performed in a variety of software packages, which all obtain similar results given the same dielectric boundary definition.[29] We used the PBEQ module in CHARMM[59] for this study. The protocol for Poisson solvation energy involves the iterative solution of the Poisson equation on a cubic grid:

$$\nabla[\epsilon(\vec{r}) \, \nabla\psi(\vec{r})] = -4\pi\rho(\vec{r}) \tag{10}$$

where $\psi(\vec{r})$ is the electrostatic potential, $\rho(\vec{r})$ is the charge density, and $\epsilon(\vec{r})$ is the dielectric boundary, defined such that $\epsilon = 1$ signifies the solute and $\epsilon = 80$ signifies the bulk solvent. A common dielectric boundary is the one of Lee and Richards.[35] The MS is defined by first superimposing spheres for each atom with radius $R_i + r_{\mathrm{probe}}$ and then rolling a water probe of radius r_{probe} across this boundary to carve out regions where the water probe can re-enter. The resultant MS is the vdW surface of the solute plus regions where a water probe cannot access. The charge density, $\rho(\vec{r})$, is computed on a grid by the trilinear interpolation of the partial atomic charges. Details on the numerical solution of the differential equation (eq 10) are available elsewhere.[66]

An alternative to the MS dielectric representation involves a transformed superposition of spherically symmetric atomic functions.[31,50,55] The benefit of such a description is that analytical forces can be obtained, which enables Poisson-based geometry optimization and MD simulations. Furthermore, a smooth description of the boundary can lead to improved computational speed and reduced grid artifacts.[31] Despite these advantages, it has been questioned whether such a definition is more or less physical than the generally adopted MS definition.[29] The main problem with the superposition volume approach is finding the right-sized atomic tail that forms physically reasonable volumes when atoms are combined at different distances. For example, if the tails of the atomic function are too small, the dielectric volume will have gaps between two nearby atoms. On the other hand, tails that are too large lead to unphysical bulges when two atoms are touching or covalently bound.[36] We recently proposed an analytical dielectric boundary approach that ameliorates these issues somewhat and attempts to mimic the MS.[36] Nevertheless, in this study, we evaluate the simple smoothing window (SW) dielectric volume definition, $\epsilon_{\mathrm{SW}}(\vec{r})$, of Im et al.,[50] as implemented in the PBEQ module:

$$\epsilon_{\mathrm{SW}}(\vec{r}) = 1 + (\epsilon_{\mathrm{solv}} - 1)\prod_i H_i(|\vec{r} - \vec{r}_i|) \tag{11}$$

where

$$H_i(t) = \begin{cases} 0 & t \le R_i - w_{\mathrm{SW}} \\ -\dfrac{1}{4{w_{\mathrm{SW}}}^3}(t - R_i + w_{\mathrm{SW}})^3 + \dfrac{3}{4{w_{\mathrm{SW}}}^2}(t - R_i + w_{\mathrm{SW}})^2 & R_i - w_{\mathrm{SW}} < t < R_i + w_{\mathrm{SW}} \\ 1 & t \ge R_i + w_{\mathrm{SW}} \end{cases} \tag{12}$$

TABLE 1: CFEs (kcal/mol) of Small Solutes Using Various Layer Sizes of Explicit Solvent and an 8-Å Short-Range Cutoff[a]

layer width (Å)	Na^+	Cl^-	H_2O	ethanol
6	−104.4	−82.6	−8.7	−12.5
8	−104.5	−82.6	−8.4	−12.3
10	−105.3	−83.0	−8.7	−12.3
12	−105.6	−83.9	−8.8	−12.3
24	−105.0	−84.0	−8.4	−12.1

[a] A TI was performed over 10 equally spaced 100-ps windows. The convergence errors are estimated to be ∼0.5 kcal/mol on the basis of a few repeated simulations that were started with different random seeds.

The smoothing parameter, w_{SW}, corresponds to the user-specified tail size of the atomic function, and ϵ_{solv} is set to 80. An efficient GB model mimicking the SW boundary Poisson scheme was recently proposed.[21]

Protein Conformation Test Sets. Two sets of protein conformations were used in this work. The first set consists of 120 conformations of the 36-residue chicken villin headpiece (PDB code: 1V11). These conformations were derived from a protocol in which low-resolution lattice model simulations[67] were reconstructed to an all-atom model.[68] Three subsets of 40 structures were generated from (1) short lattice simulations starting from extended chains, (2) longer lattice-folding simulations, and (3) simulations starting from the native conformation. The second set contains 60 conformations of the 62-residue immunoglobulin binding domain of protein L (PDB code: 2PTL). This set is a randomly chosen subset of the 215 protein-L structures used in previous works.[29,36,55] The original 215 protein-L structures were obtained in an unpublished work by extracting cluster centers from high-temperature unfolding simulations performed in explicit solvent using the TIP3P and PARAM19 force fields.[69] For both protein sets, nonpolar hydrogen atoms were added using the HBUILD facility in CHARMM[59] and the N and C termini were unblocked.

III. Results

Calibration of the Layer Width. In Table 1, we evaluate the accuracy of the hybrid method at different solvation widths for calculating the electrostatic CFE of four small molecules using a short-range cutoff of 8 Å. It appears that, with a 10−12-Å layer, the free energies converge roughly to an asymptotic limit. This width range corresponds to approximately three shells of water surrounding a solute. This is reasonably consistent with experimental studies, which indicate that the first two layers of water molecules surrounding a protein surface have properties deviating from those of the bulk solvent.[70] Furthermore, as a verification of our method, our converged results are consistent with the results of Beglov and Roux.[46] They used the same force field with a different water-cluster approach and obtained charging free energies of −105.1 kcal/mol for sodium and −8.8 kcal/mol for water using 100 explicit water molecules (∼9-Å layer). With an alternative water/ion force field,[71,72] we also found reasonable agreement (Na^+, −110 kcal/mol; Cl^-, −79 kcal/mol) with the external potential solvation energies reported by Darden et al. (Na^+, −108 kcal/mol; Cl^-, −82 kcal/mol).[44]

Small Model Compounds and the S1 Set. Next, we evaluated the solvation free energies of 20 amino acid model

TABLE 2: Solvation Free Energies (kcal/mol) for the Hybrid and Poisson Methods[a]

name	SSBP	hybrid	Poisson	Poisson/S1
		Nonpolar		
Gly	−13.3	−12.5	−12.6	−12.6
Ala	−11.9	−11.4	−11.8	−11.8
Val	−10.4	−10.3	−9.9	−9.9
Leu	−11.4	−11.9	−11.6	−11.6
Ile	−10.8	−10.2	−10.0	−10.0
Pro	−16.5	−15.9	−14.2	−14.2
Phe	−14.6	−14.7	−13.3	−13.3
Trp	−17.9	−17.9	−16.9	−16.9
Met	−11.2	−10.8	−11.2	−11.2
		Neutral Polar		
Ser	−19.7	−19.1	−19.9	−19.9
Thr	−18.1	−18.4	−18.5	−18.5
Cys	−15.5	−15.1	−15.3	−15.3
Tyr	−19.0	−18.8	−19.4	−19.4
Asn	−20.8	−19.8	−20.5	−20.5
Gln	−16.7	−15.1	−15.4	−15.4
His-δ	−25.7	−24.7	−25.0	−25.0
		Charged Polar		
Asp^-	−91.9	−90.2	−79.5	−91.0
Glu^-	−87.8	−88.3	−75.6	−86.8
Lys^+	−72.9	−76.9	−82.9	−76.8
Arg^+	−66.0	−70.2	−78.6	−69.7

[a] A TI for the hybrid method (12-Å layer) was performed over 10 equally spaced 30-ps windows. The SSBP explicit solvent results are obtained from Nina et al.[73] S1 indicates PARAM22 atomic radii with a small number of modifications (see text).

compounds[34] using a 12-Å water layer and an infinite cutoff (no multigrid). In Table 2, solvation energies of the capped amino acids [$CH_3C(=O)-X-NHCH_3$] are compared between hybrid and fully implicit solvent schemes. The structures of these model compounds were the same as those specified by Roux and co-workers[34] so that a direct comparison with their work could be made. The hybrid simulations involved TI over 10 30-ps windows (6-ps equilibration/24-ps production). The estimated error in the hybrid simulations for not using an infinite-sized layer and infinite simulation time is ∼1 kcal/mol. One can see that, for the uncharged species, the implicit solvent scheme has an excellent correspondence to the explicit approach, with a root-mean-square deviation (RMSD) error of 0.7 kcal/mol. Among uncharged residues, only Pro, Phe, and Trp have errors greater than 1 kcal/mol.

However, for the charged groups Asp^-, Glu^-, Lys^+, and Arg^+, there are significant deviations between the Poisson and hybrid methods. Corrections to the hybrid method associated with transforming to the P-summation-like solvation energies would correspond to even larger deviations. Generally speaking, compared to the PARAM22/TIP3 model, the Poisson method overestimates the solvation energies of positively charged residues and underestimates the solvation energies of negatively charged residues. One method for alleviating this situation is to modify some of the atomic radii of the side-chain atoms to bring the Poisson results into better agreement with the explicit solvent results.[34] We found, for instance, that the following scheme, which we term S1, improves the Poisson results for the charged groups: reducing the carboxylate oxygen radii of Asp, Glu, and the C terminus by 0.3 Å, augmenting the ammonium nitrogen radii of Lys and the N terminus by 0.19 Å, and increasing the side-chain nitrogen radii of Arg by 0.2 Å. Of course, this particular set of radii modifications is nonunique because there are multiple atoms on each side chain that can be modified. Furthermore, given the estimated error of ∼1 kcal/mol for the hybrid method, an exact fitting to the hybrid results is probably not constructive. Also, it should be noted that the S1 protocol is not intended as a general protocol for use by other researchers. Instead, it is used to illustrate how radii modifications, which are designed to reduce large Poisson solvation errors in small molecules, perform in calculating protein solvation energies.

TABLE 3: Comparison of Various Poisson Solvation Schemes vs Hybrid Scheme for Electrostatic CFEs (kcal/mol) of 120 Conformations of the Villin Headpiece[a]

method	cell size	probe radius	RMSD	shifted RMSD	*R*	slope
MS	0.5	1.4	17.7	6.9	0.969	0.96
MS	0.5	1.0	8.0	6.6	0.974	0.87
MS	0.25	1.4	37.6	6.0	0.982	1.07
MS	0.25	1.2	28.2	4.9	0.985	1.02
MS	0.25	1.1	23.3	4.6	0.987	1.00
MS	0.25	1.0	18.2	4.4	0.987	0.97
MS	0.25	0.9	13.1	4.4	0.987	0.95
vdW	0.25	0.0	57.6	11.0	0.925	0.73
vdW w/NBR	0.25	0.0	10.4	10.1	0.948	1.08
SW-1	0.25		14.2	9.8	0.961	1.14
SW-2	0.25		11.4	7.5	0.966	0.85
MS w/S1	0.25	1.4	28.3	5.9	0.985	1.10
MS w/S1	0.25	1.0	8.5	4.1	0.989	1.00
MS w/NBR	0.25	1.4	73.5	16.3	0.958	1.41
null			1317.0	27.5	0	

[a] The hybrid method employed a 10-Å water layer and an 8-Å short-range cutoff. MS indicates the Poisson method with a Lee−Richards MS dielectric boundary, and vdW signifies the vdW dielectric boundary. Cell sizes and probe radii are in Å. S1 indicates PARAM22 atomic radii with a small number of modifications (see text). SW-1 refers to the SW Poisson method[50] with $w_{SW} = 0.4$ Å using modified atomic radii, as prescribed elsewhere.[73] SW-2 indicates the SW Poisson method[50] with $w_{SW} = 0.5$ Å and all of the atomic radii augmented by 0.35 Å. NBR is the atomic radii set of Nina et al.[34] The term "null" corresponds to a zero CFE. RMSD indicates a root-mean-square deviation. Shifted RMSD is the RMSD calculated after the two sets being compared are each subtracted by their mean. *R* designates the correlation coefficient. The slope is obtained from a linear least-squares fit.

Our model compound study repeats part of the works of Nina et al.,[34,73] except for several key differences. First, we used between 361 (glycine) and 470 (arginine) water molecules in our layer-based simulation volume. In comparison, Nina et al. chose a simulation sphere of 150 water molecules for all of the model compounds. Also, they employed a smaller total simulation time of 200 ps in their free-energy approach. Furthermore, they selected the vdW[34] and SW boundary[73] dielectric definitions for their Poisson-based implicit solvent calculations. Using the standard PARAM22 atomic radii[60] and vdW dielectric definition, Nina et al.[34] obtained a RMSD of 1.4 kcal/mol versus that of the explicit solvent for the 16 nonpolar and neutral polar groups. Most likely, the reason they obtained worse correspondence than our results for this particular test set is because the vdW and SW volume definitions are inferior to the MS definition for Poisson calculations.

Charging Free Energies of Protein Conformations. Although comparisons between hybrid and implicit solvation energies for small model compounds, such as capped amino acids, have been previously presented in the literature,[34,73] little work has been done in comparing the solvation energies of proteins.[24,39] In Tables 3 and 4, the CFEs for several protein conformations are compared using various metrics: RMSD, shifted RMSD, correlation coefficient *R*, and a best-fit slope. Scatter-plot comparisons are also presented in Figures 2 and 3. As mentioned above, we look at the CFEs (see eq 6) as opposed to electrostatic solvation energies because the former is a stricter

TABLE 4: Comparison of Various Poisson Solvation Schemes vs the Hybrid Scheme for Electrostatic CFEs (kcal/mol) of 60 Conformations of Protein L[a]

method	cell size	probe radius	RMSD	shifted RMSD	*R*	slope
MS	0.5	1.4	89.0	14.5	0.742	0.893
MS	0.5	1.0	17.2	10.1	0.837	0.840
MS	0.25	1.4	131.2	12.8	0.756	0.799
MS	0.25	1.2	96.0	9.7	0.845	0.797
MS	0.25	1.1	78.6	9.0	0.862	0.775
MS	0.25	1.0	62.0	9.3	0.854	0.743
MS	0.25	0.9	46.8	9.8	0.834	0.717
vdW	0.25	0.0	77.2	16.5	0.483	0.381
vdW w/NBR	0.25	0.0	18.5	17.9	0.461	0.431
SW-1	0.25		37.7	16.6	0.531	0.491
SW-2	0.25		15.7	13.3	0.675	0.551
MS w/S1	0.25	1.4	77.1	11.5	0.795	0.810
MS w/S1	0.25	1.0	10.7	9.0	0.861	0.723
MS w/NBR	0.25	1.4	191.0	18.0	0.570	0.662
null			1785.8	17.7	0	

[a] The hybrid method employed a 10-Å water layer and a 12-Å short-range cutoff. MS indicates the Poisson method with a Lee−Richards MS dielectric boundary, and vdW signifies the vdW dielectric boundary. Cell sizes and probe radii are in Å. S1 is PARAM22 with a small set of modified atomic vdW radii (see text). SW-1 refers to the SW Poisson method [50] with $w_{SW} = 0.4$ Å using modified radii, as prescribed elsewhere.[73] SW-2 indicates the SW Poisson method[50] with $w_{SW} = 0.5$ Å and all of the PARAM22 atomic radii augmented by 0.35 Å. NBR is the atomic radii set of Nina et al.[34] The term "null" corresponds to a zero CFE. RMSD indicates a root-mean-square deviation. Shifted RMSD is the RMSD calculated after the two sets being compared are each subtracted by their mean. *R* designates the correlation coefficient, and the slope is obtained from a linear least-squares fit.

measure.[64] The CFEs for the hybrid method (10-Å layer) were obtained by TI simulations over five 60-ps windows (10-ps equilibration/50-ps production). First, it can be seen that there is a noticeable improvement in relative accuracy progressing from a cell size of 0.5−0.25 Å. It should be pointed out, however, that the 0.25-Å calculations are approximately 8 times slower than the 0.5-Å runs. Furthermore, for a Poisson calculation at 0.25-Å resolution, the error associated with the rotational variance of single conformation is roughly 1 kcal/mol for this protein (results not shown). Next, it appears that, as the water-probe radius is decreased from 1.4 to 1.0 Å, correspondence with the hybrid solvent method improves. The benefit of using a 1.0-Å probe radius versus one of 1.4 Å is also illustrated in Figures 2 and 3. At a 0.9-Å probe radius, the results start to degrade slightly. Taken at an extreme, results with a zero-radius probe, that is, the vdW volume definition, are significantly poorer. In fact, for the protein-L set, vdW Poisson calculation has no better discriminatory power than a guess of zero for the electrostatic CFE, that is, the null hypothesis. Finally, we consider that the errors of the best Poisson methods for protein L are nearly twice as large as those for villin. Errors in terms of percentage of the total CFE (0.3% for villin and 0.5% for protein L) make the comparison somewhat closer because the CFEs for protein L are larger. However, the discrepancy may be due to the chemical composition of the proteins or the disparate ways that the conformations were generated.

Some researchers have suggested that, if many atomic radii are scaled appropriately to obtain accurate small-molecule solvation energies, one would obtain a more accurate solvent description for proteins.[33,34] For example, to reduce the discrepancies between vdW Poisson and explicit solvent solvation energies for the 20 model tripeptides of the amino acids, Nina et al.[34] opted to make multiple modifications of the atomic vdW radii specified in the PARAM22 force field. It was assumed that such a scheme would translate to improved solvation energies for proteins. Nonetheless, as one can see in Tables 3 and 4, the elaborate radii modifications of Nina et al., dubbed here as NBR,[34] only reduce the absolute errors, not the relative errors, associated with the vdW volume Poisson method. It has also been suggested that the SW boundary Poisson method can be combined with the NBR radii.[73] This procedure, which we attempted with $w_{SW} = 0.4$ Å and called SW-1, is still not as accurate as the MS Poisson method with no radii modifications. We also found that the NBR radii should probably not be used in conjunction with the MS Poisson method. On the other hand, our radii modification scheme, S1, which only involves four atom types, only slightly affects the shifted RMSD and correlation coefficients. However, the S1 protocol improves absolute solvation energies significantly. This may prove useful

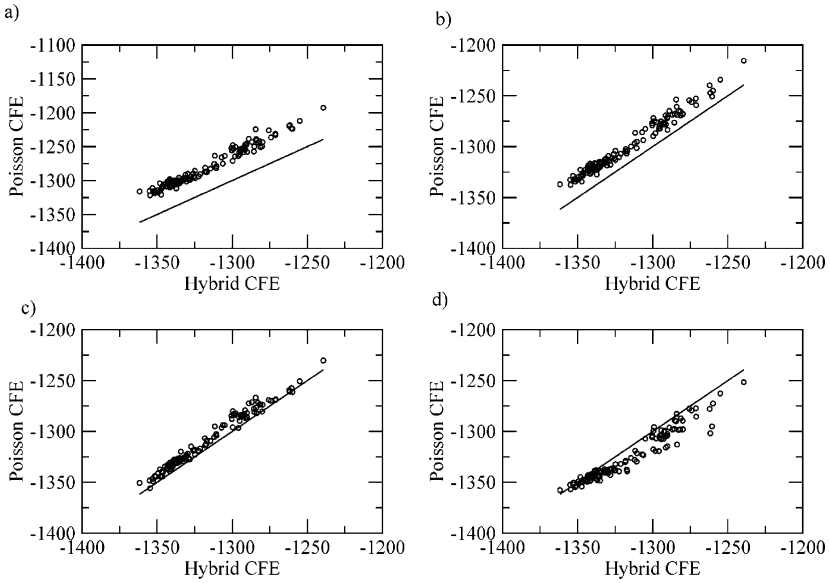

Figure 2. Comparisons of various Poisson models against the hybrid method for the CFEs of 120 conformations of the villin headpiece: (a) MS Poisson ($r_{probe} = 1.4$ Å) using PARAM22 radii; (b) MS Poisson ($r_{probe} = 1.0$ Å) using PARAM22 radii; (c) MS Poisson ($r_{probe} = 1.0$ Å) using PARAM22/S1 radii; (d) PBSW ($w_{SW} = 0.5$ Å) using PARAM22 atomic radii augmented by 0.35 Å. All units are kcal/mol. The straight lines indicate the ideal correspondence, $y = x$.

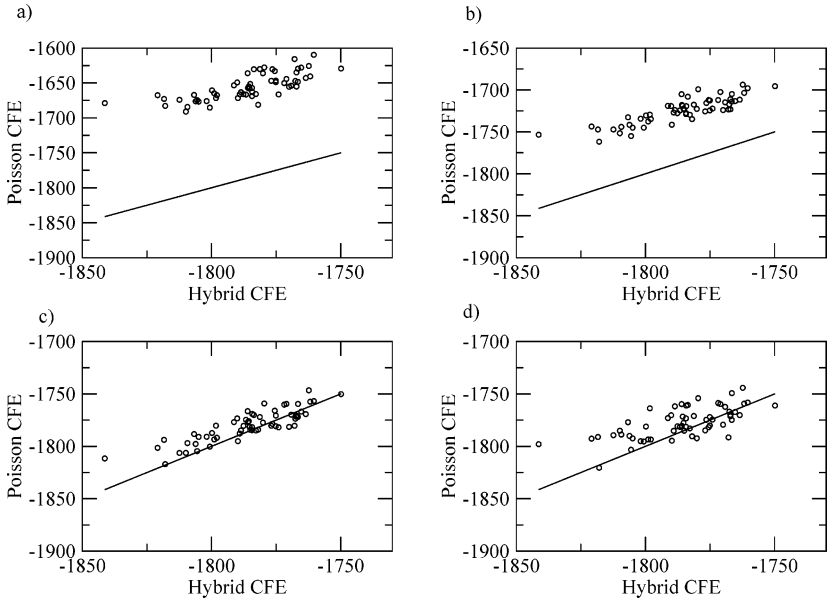

Figure 3. Comparisons of various Poisson models against the hybrid method for the CFEs of 60 conformations of protein L: (a) MS Poisson (r_{probe} = 1.4 Å) using PARAM22 radii; (b) MS Poisson (r_{probe} = 1.0 Å) using PARAM22 radii; (c) MS Poisson (r_{probe} = 1.0 Å) using PARAM22/S1 radii; (d) PBSW (w_{SW} = 0.5 Å) using PARAM22 atomic radii augmented by 0.35 Å. All units are kcal/mol. The straight lines indicate the ideal correspondence, $y = x$.

in the context of protein−ligand modeling, where each molecule desolvates the other upon binding.[74]

We also attempted to improve upon the SW dielectric boundary definition of Im et al.[50] Through trial and error, we determined that a simple scheme, designated as SW-2, which involves augmenting all of the atomic radii by 0.35 Å and using a switching function width of 0.5 Å, provides results on par with those of the MS Poisson method, using a probe radius of 1.4 Å for the protein-L data set. This particular scheme is somewhat transferable to the villin set.

Single-Residue Analysis of Two Villin Structures. Insight can be gained by looking at the CFE of each individual residue embedded in the fully charged protein even though, by construction, these values do not add up to the total CFE of the protein. A total of 72 individual hybrid charging simulations were performed corresponding to each residue for two villin structures. One villin structure was the first member of the native-like set, and the other structure was the first member of the extended set. Each simulation incorporated a 10-Å layer of water around the entire protein. Table 5 summarizes the results of these simulations. The errors in residue CFEs of a single native villin structure are presented in Figure 4. The trend here is similar to that seen in the model compound data in Table 2. For instance, in Figure 4, only the charged groups Asp, Glu, Lys, and Arg and the N and C termini are in error significantly greater than 1 kcal/mol. On the other hand, the Poisson implicit solvent model does a remarkable job in treating uncharged residues even in a protein environment. Interestingly, using a 1.0-Å probe radius versus a 1.4-Å probe radius improves the relative errors between the extended and native structures but not the absolute values. Perhaps this indicates that the smaller probe radius leads mainly to an improved cancellation of errors.

To understand the effect of the S1 radii set, we focused on the CFEs for each ionized residue on the two villin structures in Figures 5 and 6. As one can see in these figures and in Table 5, the S1 modifications significantly improve the absolute residue CFEs. Nonetheless, when the differences between the errors for the two structures in Figure 7 and Table 5 are examined, it is seen that the benefits of radii modification are somewhat muted. There is a clear cancellation of errors for the standard radii calculations. However, the S1 radii actually slightly increase the individual relative errors between these two structures for the calculations using a water-probe radius of 1.0 Å. Combined with the total CFE results, it would seem that the benefit of using S1 radii for relative energies is evident but not substantial.

TABLE 5: Statistics for the Errors in the CFEs of the MS Poisson (0.25-Å Cell Size) Method for Each of the 36 Residues of Two Villin Structures[a]

probe radius/ radii set	RMSD all	RMSD neutral	RMSD ionized
	Native		
1.4 Å	5.9	0.7	10.1
1.0 Å	5.9	0.8	10.2
1.4 Å/S1	1.9	0.8	3.1
1.0 Å/S1	1.7	0.8	2.7
	Extended		
1.4 Å	6.1	0.9	10.4
1.0 Å	6.1	0.7	10.4
1.4 Å/S1	2.0	0.9	3.3
1.0 Å/S1	2.0	0.8	3.2
	Native−Extended		
1.4 Å	1.8	0.8	2.8
1.0 Å	1.2	0.8	1.7
1.4 Å/S1	1.6	0.9	2.4
1.0 Å/S1	1.5	0.9	2.3

[a] The benchmark is the hybrid results with an 8-Å cutoff, a 300-ps total simulation time, and a 10-Å water layer. S1 indicates that the S1 atomic radii were used.

Validation of the Hybrid Layer Model. There are multiple critical issues to consider regarding the validity of the hybrid layer solvent model as a benchmark for Poisson calculations. In Table 6, we consider the effects of separately extending the short-range cutoff, layer size, and simulation time for 12 of the villin headpiece structures (the first 4 from each subset). As a control, we also compare the results against simulation runs that use the same parameters as the standard protocol except for a different random seed for the initial velocities. The reference

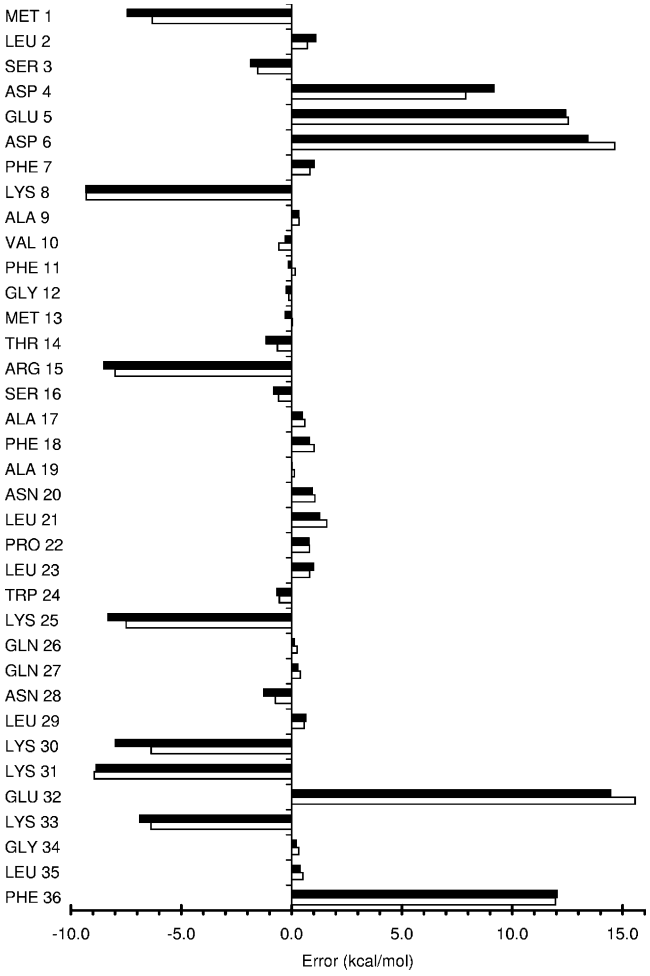

Figure 4. Deviations in the CFE of each residue in a native conformation of the villin headpiece between the MS Poisson (PARAM22 radii) and hybrid methods. Filled rectangles indicate a probe radius of 1.0 Å; open rectangles indicate a probe radius of 1.4 Å.

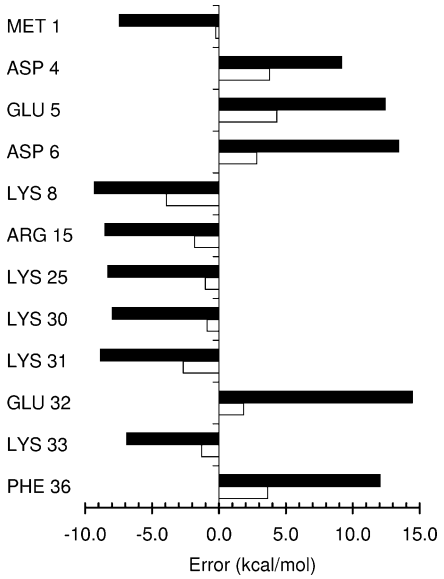

Figure 5. Deviations in the CFEs of each ionized residue in a *native* conformation of the villin headpiece between the MS Poisson (r_{probe} = 1.0 Å) and hybrid methods. Filled rectangles indicate standard PARAM22 atomic radii; open rectangles correspond to the PARAM22/S1 atomic radii.

calculations are the same as those used above in our implicit solvent benchmarking. First, one should note that extension of

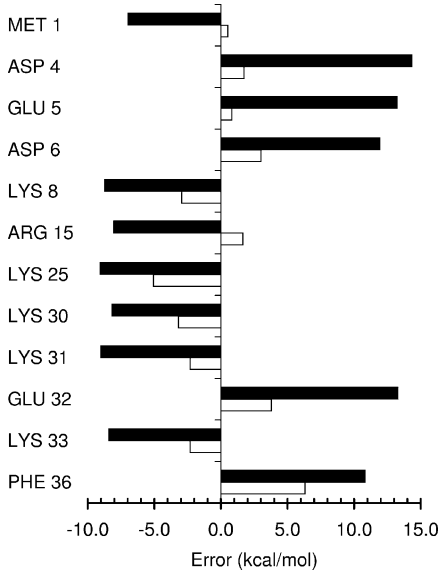

Figure 6. Deviations in the CFEs of each ionized residue in an *extended* conformation of the villin headpiece between the MS Poisson (r_{probe} = 1.0 Å) and hybrid methods. Filled rectangles indicate standard PARAM22 atomic radii; open rectangles correspond to the PARAM22/S1 atomic radii.

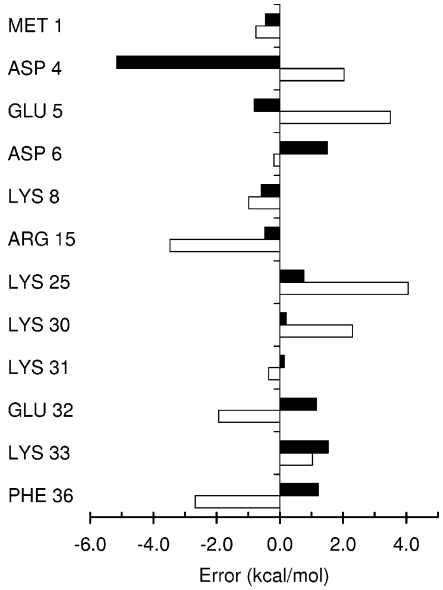

Figure 7. Relative deviations in the CFEs for each ionized residue between an extended and native conformation (native minus extended) of the villin headpiece between the MS Poisson (r_{probe} = 1.0 Å) and hybrid methods. Filled rectangles indicate standard PARAM22 atomic radii; open rectangles correspond to the PARAM22/S1 atomic radii.

the simulation layer to 12 Å seems to introduce a smaller deviation than the control comparison. This is a promising indicator that we have achieved convergence with respect to the system size. Next, it appears that increasing the simulation time to 600 ps per window does modify the results somewhat. However, in an absolute sense, the results from the doubled-simulation-time protocol deviate only 0.5 kcal/mol more than those of the control. This outcome suggests that there is some amount of convergence to be gained by increasing the simulation time. Finally, the most noteworthy deviation comes from extending the short-range cutoff. There is clearly an absolute error associated with our small cutoff of 8 Å. Nevertheless, the fact that the relative error is within the deviations of the control indicates that an 8-Å cutoff is probably sufficient for assessing

TABLE 6: Comparison of CFEs (kcal/mol) for the Hybrid Method Using Different Parameters[a]

water layer (Å)	short-range cutoff (Å)	total simulation time (ps)	RMSD	shifted RMSD
10[b]	8	300	2.3	2.1
10	8	600	2.8	2.0
10	12	300	3.7	1.9
12	8	300	1.7	1.4

[a] The test set contains the first four structures from each subset of the villin headpiece structures ($N = 12$). The reference calculations employed the following parameters: 10-Å water layer, short-range cutoff = 8 Å, and total simulation time = 300 ps. [b] The same parameters as those of the reference calculations were used except for the seed used in random number generation for initial velocity assignments.

the relative conformational energies of villin. Overall, combining the factors of a lack of an infinite system, time, and cutoff leads to estimates of ~4 kcal/mol RMSD error and ~2 kcal/mol shifted RMSD error for the villin set regarding these particular issues. For the protein-L set, we only evaluated the effect of the short-range cutoff size because this seemed to be the most significant aspect in the villin set. We determined that a 12-Å cutoff is necessary to achieve reasonably converged results with respect to the cutoff. Compared to using a 16-Å cutoff for 10 random conformations, the RMSDs are ~5 kcal/mol for both absolute and relative energies. Thus, we estimate the RMSD error of our protein-L hybrid calculations, associated with short-range cutoff, layer size, and simulation time, to be ~5 kcal/mol.

One obvious question is, how does the hybrid solvent model fare against other hybrid or explicit solvent approaches? It is beyond the scope of this work to compare our hybrid approach with purely explicit solvent approaches that employ periodic boundary conditions[75] because of a few unsolved technical issues. An Ewald-type treatment, such as particle-mesh Ewald,[76] is difficult to perform because it is not clear how to properly remove interactions between the real solute and its periodic images.[42] Alternatively, a periodic boundary conditions (PBC) simulation with large cutoffs, but not large enough to allow for real−virtual solute interactions, requires a reaction field correction to the cutoff. The problem with this type of reaction-field correction is that the neglected long-range solute−solute interactions are implicitly treated as though there is a dielectric boundary of water associated with the distant atoms. Furthermore, PBC reaction-field corrections may require reparametrization of the force field to restore correct bulk water properties.[77]

With these technical difficulties for PBC simulations, we are left with comparisons to alternative established hybrid approaches such as the SSBP[46] or surface-constrained all-atom solvent model.[43] As mentioned in the Methods section, we have chosen to compare our results to those of the SSBP method. Table 7 presents comparisons of the SSBP method to our hybrid approach using a small set of villin conformations ($N = 8$; first 4 native-like, first 4 lattice). Conformations of villin in the extended structure set produce simulation spheres that are too large for our current simulation protocols and, thus, have been neglected in this comparison. As one can see, the deviations between the SSBP model and the hybrid sphere model are quite substantial (~22 kcal/mol) despite the fact that both calculations use nearly the same simulation volume. To understand the discrepancies between the SSBP and hybrid sphere results, one must be aware of the differences between the two methods. The most significant difference between the two approaches lies in the dissimilar methods used in partitioning the short- and long-range electrostatic and continuum interactions. In the hybrid method, the GB and electrostatic pairwise short-range interactions are truncated at the same cutoff. In contrast, the SSBP method truncates the electrostatic interaction but does not pare the continuum component, regardless of the specified cutoff. The accuracy of the SSBP continuum component is only tuned by the number of chosen Kirkwood multipoles. We predicted that a balance of the cutoffs in the GB and electrostatic terms would make the hybrid method more quickly convergent to the cutoff length through a cancellation of errors. Figure 8 illustrates how the results for the SSBP and hybrid methods change as a function of the cutoff length. As one can see, both the SSBP and hybrid methods converge to the no-cutoff limit as the cutoff increases. The hybrid results, however, converge with more noise. In fact, by chance and error cancellation, the 12-Å cutoff hybrid result is very close to the infinite cutoff value. In addition, one can see that the MS Poisson value using a 1.0-Å water probe is quite close to the infinite cutoff value. Although this result is not statistically significant, it does reinforce our premise that there can be considerable error cancellation with the Poisson method, even without any atomic radii modification.

Also surprisingly, the SSBP and hybrid results for this single conformation converge to virtually the same result at infinite cutoffs. One would expect there to be some discrepancies in the CFE results at the infinite cutoff limit because the SSBP and hybrid methods have notable algorithmic differences. For example, as mentioned above, the SSBP method uses a truncated Kirkwood multipole expansion to treat the continuum and our hybrid approach, instead, uses GB theory. Each continuum method has its drawbacks. On the one hand, GB theory is fundamentally an approximation to Poisson theory. On the other hand, in the Kirkwood method, the dielectric boundary definition is discontinuously sharp and water molecules must be confined inside the simulation volume. If a water molecule were to pierce the boundary in the SSBP method, an infinite solvation energy would result.[46] Also, the accuracy of the Kirkwood method is dependent upon the number of multipoles specified by the user. Another difference between the SSBP and hybrid methods is how the angular polarization of water molecules near the boundary is treated. The SSBP method attempts to impose angular isotropy but falls short. As a result, computations of the CFE of sodium ions are close to those values expected for external solvation energies.[46] The hybrid method, in its present form, does not have an angular isotropy correction. A final difference between the two methods is the type of ensemble each method tries to produce. The SSBP method is, by design, a constant-pressure algorithm, whereas the hybrid method is, in its current form, a constant-volume method. Finally, a note on computational times is as follows: SSBP calculations on a single conformation using an 8-Å short-range cutoff with extended electrostatics require between 65 and 180 h of computational time. This compares to ~16 h for the hybrid layer using an 8-Å cutoff and 24−63 h for the hybrid sphere calculations, also using an 8-Å cutoff.

It is interesting to note that the SSBP results are closer to those of the MS Poisson model than the hybrid data, as seen in both Figure 8 and Table 7. We believe that this agreement is somewhat fortuitous. In Figure 8, the error associated with using an 8-Å cutoff for the SSBP method is roughly 10 kcal/mol, placing the SSBP result right in the middle of the two MS Poisson values. Also, as seen in Table 7, the shifted RMSD for the SSBP method is in slightly better agreement with that of the Poisson model compared to other hybrid protocols. Another notable fact is that the shifted RMSD values between the Poisson model and both the hybrid layer and sphere models are roughly

TABLE 7: Comparison of Various Methods for Calculating CFEs[a]

method	SSBP[b]	hybrid sphere[c]	hybrid layer[d] (12 Å)	hybrid layer[d] (8 Å)	MS[e] (1.4 Å)	MS[e] (1.0 Å)
SSBP[b]						
sphere[c] (8 Å)	21.6 (3.6)					
layer[d] (12 Å)	24.6 (3.6)	4.0 (2.6)				
layer[d] (8 Å)	28.3 (4.6)	7.4 (3.1)	4.2 (2.1)			
MS[e] (1.4 Å)	13.1 (4.7)	33.9 (5.1)	36.9 (5.0)	40.5 (5.6)		
MS[e] (1.0 Å)	9.3 (4.3)	14.2 (5.4)	16.9 (5.1)	20.5 (5.6)	20.5 (1.8)	
MS[e] (1.0 Å) w/ S1	18.4 (6.3)	6.2 (4.6)	7.8 (3.2)	11.3 (3.6)	30.1 (6.2)	11.3 (6.8)

[a] The test set contains the first four structures of the native and lattice subsets ($N = 8$). In each cell, the first value corresponds to the RMSD and the second value corresponds to shifted RMSD. [b] The SSBP method of Beglov et al. with a minimum 10-Å solute−boundary distance using an 8-Å short-range cutoff (see the Methods section for details). [c] The hybrid method using a spherical simulation volume with a minimum 10-Å solute−boundary distance and an 8-Å short-range cutoff. [d] The hybrid method using a 10-Å layer simulation volume, and the short-range cutoff is specified in parentheses. [e] MS indicates the Poisson method with a MS dielectric boundary and using the water probe specified in parentheses.

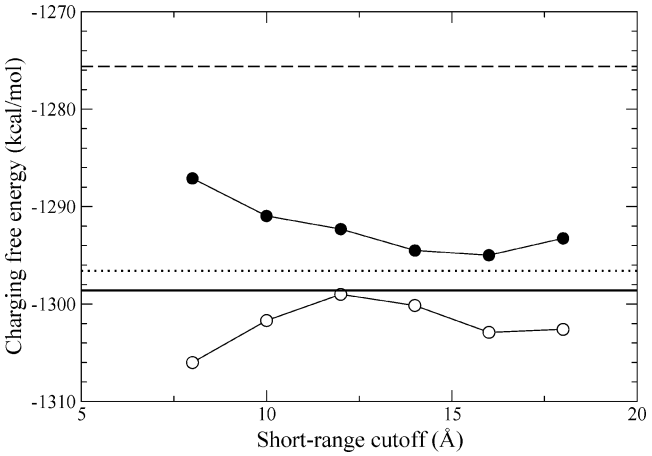

Figure 8. Convergence of CFE as a function of the short-range cutoff for a single villin conformation for hybrid sphere and SSBP simulations. Dark circles correspond to SSBP simulations, and open circles indicate hybrid sphere simulations. The horizontal solid line corresponds to a no-cutoff limit of both the SSBP and hybrid sphere simulations. The horizontal dotted line indicates the MS Poisson method with a 1.0-Å water probe, and the dashed line signifies the MS Poisson method with 1.4-Å water probe.

the same. Nonetheless, it should be remembered that the Poisson model is an implicit solvent model and, therefore, not a suitable benchmark to determine which hybrid method is better.

In Table 7, the differences in the results between a hybrid layer and hybrid sphere model are also considered. The discrepancy between the sphere and layer models using the same cutoff is ∼7 kcal/mol. The deviation is a bit smaller when the layer model has a larger cutoff of 12 Å. The shifted RMSD between the layer and sphere models is ∼2−3 kcal/mol. Overall, we deduce that curvature effects associated with water molecules near the boundary in our layer calculations introduce noticeable absolute errors and minor relative errors. However, given the computational feasibility of a layer calculation versus that of a sphere calculation, especially for extended conformations, the curvature errors are tolerable for this study and, perhaps, amenable to an ad hoc systematic correction.

Another question is, how do the calculated CFEs deviate depending on the scheme one uses to treat isolated protein cavities where water molecules might get trapped? A rigorous solution to this problem would involve calculating the chemical potential of a water molecule inside each identified cavity and retaining the molecule if the chemical potential were favorable. Nonetheless, we analyzed this issue by evaluating the statistics on the subset structures where no isolated cavities that could fit a 1.4-Å probe sphere could be identified. We used the program VOLBL[78] to determine that 8 villin structures have one void and 41 protein-L structures have between one and seven voids. For the villin set, when the subset with voids is removed, the comparisons between the MS Poisson results, using the S1 radii and a 1.0-Å water-probe radius, and the hybrid results are virtually identical to the full set results. For the protein-L set, when only the nonvoid subset is considered, the MS Poisson results, using the S1 radii and a 1.0-Å water-probe radius, had a RMSD of 12.3 kcal/mol and a shifted RMSD of 8.2 kcal/mol, which is in line with the full set results.

A further issue regarding the validity of our hybrid solvent results is our treatment of non-neutral solutes. The villin structures have a +2 charge, and the protein-L structures have a −2 charge. As we mentioned above, there is some ambiguity in the literature regarding the proper summation procedure for calculating ionic CFEs. Different methods can lead to as much as a 20 kcal/mol discrepancy per atomic charge. In our method and in other cluster approaches, a definite surface dipole artifact exists even for a bulk water droplet that leads to about a −10 to −15 kcal/(mol·atomic charge) potential at the center of the spherical cluster (results not shown). This artifact is characterized by the tendency of hydrogen atoms in water molecules to point outward near the boundary. This can be explained by the fact that water molecules near the boundary are trying to sacrifice the least number of intermolecular hydrogen bonds and, therefore, choose to point one of their hydrogens outward. The effect of the surface artifact on the CFEs never goes to zero because the surface dipole density scales to the extent of the square of the cluster. If we correct the CFEs of atomic ions for the surface artifact using the approach outlined by Darden et al.,[44] we find good agreement with their results (results not shown). However, the discrepancy with uncorrected Poisson results increases by another ∼14 kcal/mol for both sodium and chloride ions. Thus, to bring the Poisson results into agreement with hybrid results, one would need to modify the radii even further from their initial force-field values.

Nonetheless, we do not believe surface dipole artifacts are going to be critical for practical applications. The fact is that charged solutes such as charged proteins are, in reality, surrounded by counterions. If one were to model-in counterions in a cluster approach, the effect of the surface artifact on the CFEs would be greatly reduced because the net system would be neutral and interactions with the surface dipoles would decay asymptotically to zero with increasing system size. Furthermore, we assert that an optimal implicit solvent theory should account for counterions if the solute has a net charge. Thus, we predict that the ambiguity associated with the different possible explicit solvent results used for parametrization may eventually disappear.

A few other issues are worth mentioning. As pointed out above, because both of our test proteins had a net charge, it is plausible that counterions may be needed to improve physical realism.[79] This should be a topic of further investigation. Another issue is that we used a linear fit to convert the force data to free energies. To validate this choice, we performed fifth-order polynomial fits for some of the data sets. We found that the small-molecule results were nearly unchanged and the protein conformation results had an absolute RMSD of less than 2 kcal/mol and a shifted RMSD of ~0.5 kcal/mol. Finally, the specific force field we used in this study is but one of many possible force fields. It is well-known that deviations in potentials of mean force do exist among different solvent models and force fields.[27] Thus, the conclusions we have drawn in this study could potentially change if another force field were employed. This matter should also be a topic of future study.

IV. Discussion

Although the results in this work are only a small subset of the possible comparisons between implicit and hybrid solvent schemes, some of the key issues regarding implicit solvent techniques are considered. First, the model compound and residue charging results indicate that the standard MS-based Poisson solvation technique is very good at reproducing the solvation energies of neutral chemical groups. This result is impressive given such a simple dielectric continuum model. Perhaps charge neutrality implies that a cancellation of errors takes place between the positive and negative components of the solute. Furthermore, the smaller partial atomic charges of a neutral group may produce small enough local disturbances of water molecules such that the continuum dielectric treatment is an adequate approximation.

On the other hand, the Poisson method fares much worse for charged groups in model compounds and proteins. This result is well-known.[34,80−82] For small charged model compounds, it is trivial to modify a few atomic radii to improve the Poisson energies as we have done here and others have elsewhere.[34,82] In certain cases, improving absolute solvation energies may lead to better descriptions of protein−ligand complexes and mutational free energies. For example, the desolvation energy associated with mutating an uncharged residue to a charged residue is likely better described by the Poisson method with our S1 set of modified radii. Nonetheless, the protein-charging free-energy results seem to indicate that simple radii modification is not a universal panacea. In fact, we reiterate that the S1 radii were introduced in this work only as an example of how radii modifications for the most glaring deficiencies of standard Poisson methods affect absolute and relative errors. For one thing, a single choice of an atomic radius may not be generally suitable for disparate protein environments such as the surface and interior. Also, the S1 radii modification scheme does not appear to improve the relative energies of conformations and, hence, may not be useful in the realm of ab initio protein structure prediction and refinement.[10,83] Generally speaking, the possibility exists that solvation around ionized groups cannot simply be treated as a two-dielectric continuum problem. Explicit solvent simulations suggest that ion solvation incurs no less than four types of deviation from simple continuum behavior:[80−82] asymmetrical water structure, electrostriction, dielectric saturation, and a nonzero electrostatic potential for a solute in its uncharged state. Asymmetrical water structure refers to the fact that the water molecules penetrate the solute boundary differently depending upon whether the oxygen or hydrogens are pointing inward. For example, anions, such as chloride and carboxylate ions, cause the innermost shell of water molecules to point their hydrogens inward. Because the hydrogens can penetrate slightly deeper into the solute than the oxygen, one might conclude that the effective solute radius should be decreased in order for implicit solvent simulations to match explicit solvent ones. The reverse is true for cations. Electrostriction refers to the situation where water molecules in the first few shells are drawn closer to the ion than they would be if the solute were uncharged. Dielectric saturation implies that the effective dielectric values of the region corresponding to the first few water shells are lower than the bulk values because the ion enforces orientational restraints on the water molecules.[82] Finally, a nonzero electrostatic potential for the uncharged solute occurs when water molecules have a slight orientational preference even at the initial uncharged state of the work function for charging an ion in a solvent.[81] The combined effect of these four phenomena is unlikely to be reproduced by only a simple modification of atomic radii.[81] More sophisticated ideas should be considered, such as the two-tiered model of Marcus,[80] which casts the electrostatic solvation energy as a sum of an electrostricted inner hydration shell and a bulk continuum. Another possible route is a distance-dependent dielectric definition such as the one of Bucher and Porter.[84] To account for water asymmetry and the fact that dissimilar protein environments should be treated differently, one might consider allowing the dielectric boundary to vary as a function of the local electric field.[81]

One of the most surprising findings of this study is that a slightly smaller probe radius of 1.0 Å versus the normally prescribed radius of 1.4 Å distinctly improves the correspondence between implicit and explicit solvent total CFEs for both protein sets. These results suggest that water molecules are able to access deeper into the crevices between atoms than previously thought.[85] Nonetheless, at the individual residue level, the smaller water probe only appears to have a beneficial effect on relative energies. Hence, the 1.0-Å probe radius protocol might be applicable in the context of protein structure prediction[10] and loop prediction,[86] and it may not be as useful in studies of mutational free energies[14,15,87] and pK_a's.[13] The underlying reasons for the conflicting total energy and residue energy results are unclear. However, we can at least postulate why the probe radius parameter may deviate from the standard value of 1.4 Å. One way to derive the standard 1.4-Å value is to observe that it is approximately half the experimentally determined interatomic distance between the two oxygen atoms in a water dimer (~2.946 Å/2 = 1.473 Å).[88] On the other hand, in the context of classical MD simulations, the probe radius ought to be the same as the TIP3P oxygen vdW radius (1.77 Å) near a solute atom with a zero partial charge. As the partial charge increases on a solute atom, the water-probe radius should decrease as the solute−water interaction distance decreases as a result of electrostriction. At the extreme of hydrogen-bond formation to polar atoms on the surface of the solute, corresponding to bond distances between 2.8 and 3.0 Å, water-probe radii values between 1.0 and 1.2 Å might be appropriate if one assumes that hydrogen-bond donors/acceptors in the solute have an average vdW radius of ~1.8 Å. These rough arguments may partially explain the success of using a 1.0-Å probe radius for relative energies. Furthermore, one can envision the possibility for parametrization of a spatially dependent probe radius function against explicit solvent data.

Finally, the results in this work indicate that the MS boundary definition provides better agreement with hybrid results for proteins than the alternative SW boundary definition of Im et

al.[50] We attempted to fit the SW Poisson method to the hybrid results by increasing all of the atomic radii by a fixed amount (SW-2), but we still did not achieve results quite as good as those of the MS Poisson method. Furthermore, the elaborate strategy of Nina et al.,[73] which attempts to fit the SW Poisson method to model tripeptide results, performed less adequately than a simple MS Poisson procedure in the context of total CFEs of protein conformations. Given these results, we predict that the optimal boundary definition should resemble the MS definition yet provide a smooth, rather than discontinuous, boundary.[31] Such an optimal model might be similar to the GBMV2 boundary definition,[36] which was specifically designed to be smooth, yet closely mimics the MS definition.

V. Conclusion

In this work, we compared the CFEs of various conformations of two small proteins between macroscopic implicit solvent approaches and a semi-microscopic explicit/implicit solvent method. We were able to determine that, among different protocols, the Poisson method with a MS dielectric boundary obtains the best agreement with hybrid solvent results. In addition, we observed that assigning the water-probe radius to a nonstandard value of 1.0 Å appeared to improve total energies but seemed to provide only a modest improvement to single-residue CFEs. Furthermore, we surmised that the source of errors in Poisson solvent models for both small molecules and proteins lies mainly in the treatment of ionized groups. Thus, modifying atomic vdW radii in these charged groups can improve absolute errors somewhat. Nonetheless, our simple radii modification scheme had little effect in improving the relative energies between conformations.

Acknowledgment. We thank Dr. M. Feig, Dr. L. Carlacci, and Dr. G. Hummer for helpful discussions. M.S.L. acknowledges funding support from the Army Medical Research and Materiel Command Project RIID 02-4-1R-069. Computational time for this work was provided, in part, by the U.S. Army Research Laboratory Major Shared Resource Center. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

References and Notes

(1) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7.
(2) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(3) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.
(4) Russell, S. T.; Warshel, A. *J. Mol. Biol.* **1985**, *185*, 389.
(5) Hassan, S. A.; Mehler, E. L. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 45.
(6) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133.
(7) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.
(8) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(9) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(10) Feig, M.; Brooks, C. L., III. *Proteins* **2002**, *49*, 232.
(11) Vorobjev, Y.; Hermans, J. *Protein Sci.* **2001**, *10*, 2498.
(12) Lee, M. R.; Baker, D.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 1040.
(13) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473.
(14) Langen, R.; Brayer, G. D.; Berghuis, A. M.; McLendon, G.; Sherman, F.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 589.
(15) Olson, M. A.; Reinke, L. T. *Proteins* **2000**, *38*, 115.
(16) Huo, S.; Massova, I.; Kollman, P. A. *J. Comput. Chem.* **2002**, *23*, 15.
(17) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393.
(18) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. I. *Acc. Chem. Res.* **2000**, *33*, 889.
(19) Krol, M. *J. Comput. Chem.* **2003**, *24*, 531.
(20) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765.
(21) Im, W.; Lee, M. S.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1691.
(22) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458.
(23) Wagoner, J.; Baker, N. A. *J. Comput. Chem.* **2004**, *25*, 1623.
(24) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 502.
(25) Calimet, N.; Schaefer, M.; Simonson, T. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 144.
(26) Bursulaya, B. D.; Brooks, C. L., III. *J. Phys. Chem. B* **2000**, *104*, 12378.
(27) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722.
(28) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 738.
(29) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. I. *J. Comput. Chem.* **2003**, *25*, 265.
(30) Baker, N. A.; D., S.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037.
(31) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.
(32) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244.
(33) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(34) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(35) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 151.
(36) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348.
(37) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400.
(38) Luo, R.; Gilson, M. K. *J. Am. Chem. Soc.* **2000**, *122*, 2934.
(39) Zhang, L. Y.; Gallichio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591.
(40) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(41) Felts, A. K.; Harano, Y.; Gallichio, E.; Levy, R. M. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 310.
(42) Bogusz, S.; Cheatham, T. E. I.; Brooks, B. R. *J. Chem. Phys.* **1998**, *108*, 7070.
(43) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.
(44) Darden, T.; Pearlman, D.; Pedersen, L. G. *J. Chem. Phys.* **1998**, *109*, 10921.
(45) Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967.
(46) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.
(47) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924.
(48) Beglov, D.; Roux, B. *Biopolymers* **1995**, *35*, 171.
(49) Lounnas, V.; Ludemann, S. K.; Wade, R. C. *Biophys. Chem.* **1999**, *78*, 157.
(50) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59.
(51) Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951.
(52) Herce, D. H.; Darden, T.; Sagui, C. *J. Chem. Phys.* **2003**, *119*, 7621.
(53) Vorobjev, Y. N.; Hermans, J. *J. Phys. Chem. B* **1999**, *103*, 10234.
(54) Hummer, G.; Pratt, L. R.; Garcia, A. E.; Garde, S. *J. Phys. Chem. B* **1998**, *102*, 3841.
(55) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606.
(56) Skeel, R. D.; Tezcan, I.; Hardy, D. J. *J. Comput. Chem.* **2002**, *23*, 673.
(57) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.
(58) Board, J. A.; Causey, J. W.; Leathrum, J. F.; Windemuth, A.; Schulten, K. *Chem. Phys. Lett.* **1992**, *198*, 89.
(59) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminatham, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(60) Mackerell, A. D., Jr.; Bashford, D.; Bellott, D. M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, 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, I. W. E.; 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.
(61) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(62) Lawson, C. L. *SIAM Rev.* **1965**, *7*, 415.
(63) Stote, R. H.; States, D. J.; Karplus, M. *J. Chim. Phys. Phys.-Chim. Biol.* **1991**, *88*, 2419.
(64) Scarsi, M.; Caflisch, A. *J. Comput. Chem.* **1999**, *20*, 1533.

(65) Kirkwood, J. G. In *Theory of Liquids*; Alder, B. J., Ed.; Gordon and Breach: New York, 1968.
(66) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 47.
(67) Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 475.
(68) Feig, M.; Rotkiewicz, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L. I. *Proteins* **2000**, *41*, 86.
(69) Sheinerman, F. B.; Brooks, C. L. I. *J. Mol. Biol.* **1998**, *278*, 439.
(70) Pal, S. K.; Peon, J.; Zewail, A. H. *Proc. Natl. Acad. Sci.* **2002**, *99*, 1763.
(71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, The Netherlands, 1981.
(72) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876.
(73) Nina, M.; Im, W.; Roux, B. *Biophys. Chem.* **1999**, *78*, 89.
(74) Rankin, K. N.; Sulea, T.; Purisima, E. O. *J. Comput. Chem.* **2003**, *24*, 954.
(75) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(76) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(77) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *108*, 10220.
(78) Edelsbrunner, H.; Facello, M.; Fu, P.; Liang, J. Measuring Proteins and Voids in Proteins. *Proceedings of the 28th Annual Hawaii International Conference System Sciences*; IEEE Computer Society Press: Los Alamitos, CA, 1995; Vol. 5, p 256.
(79) Ibragimova, G. T.; Wade, R. C. *Biophys. J.* **1998**, *75*, 2906.
(80) Marcus, Y. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2995.
(81) Ashbaugh, H. S. *J. Phys. Chem. B* **2000**, *104*, 7235.
(82) Babu, C. S.; Lim, C. *J. Chem. Phys.* **2001**, *114*, 889.
(83) Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 8392.
(84) Bucher, M.; Porter, T. L. *J. Phys. Chem.* **1986**, *90*, 3406.
(85) Feig, M.; Lee, M. S. Work in progress.
(86) Rapp, C. S.; Friesner, R. A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 173.
(87) Huo, S.; Massova, I.; Kollman, P. A. *J. Comput. Phys.* **2002**, *23*, 15.
(88) Halkier, A.; Koch, H.; Jorgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150.