JCTC Journal of Chemical Theory and Computation

A Revised Density Function for Molecular Surface Calculation in Continuum Solvent Models

Xiang Ye, Jun Wang, and Ray Luo*

Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900

Abstract: A revised density function is developed to define the molecular surface for the numerical Poisson−Boltzmann methods to achieve a better convergence and a higher numerical stability. The new density function does not use any predefined functional form but is numerically optimized to reproduce the reaction field energies computed with the solvent excluded surface definition. An exhaustive search in the parameter space is utilized in the optimization, using a wide-range of training molecules including proteins, nucleic acids, and peptides in both folded and unfolded conformations. A cubic-spline function is introduced to guarantee good numerical behavior of the new density function. Our test results show that the average relative energy errors computed with the revised density function are uniformly lower than 1% for both training and test molecules with different sizes and conformations. Our transferability analysis shows that the performance of the new method is mostly size and conformation independent. A detailed analysis further shows that the numerical forces computed with the revised density function converge better with respect to the grid spacing and are numerically more stable in tested peptides.

Introduction

Solvation is one of the essential determinants of the structure and function of proteins and nucleic acids.[1−14] To model solvation interactions in classical molecular simulations, explicit solvent models that explicitly represent every atom in a solvent molecule are natural choices. However, limitations of explicit solvent models have also been recognized. Apparently, the computational cost is a concern when explicit solvent models are used. Often overlooked are the pre-equilibration and sampling needs. Indeed the exponentially large phase space of explicit solvent degrees of freedom makes the convergence of simulations very challenging. Analyses of these limitations prompted pioneers in molecular simulations to propose implicit representations to model solvation, especially in biomolecular applications.[1−14]

Implicit solvent models replace explicit solvent interactions with an equivalent energetic term based on a mean field approximation. Accuracy and transferability often requires decomposing the mean field potential into electrostatic and nonelectrostatic components and modeling the two components separately. Such an approach can reduce the computation cost but has been found to provide a certain degree of accuracy in the treatment of solvation interactions.[1−14] With over 20 years of developments, implicit solvent models, especially those based on the Poisson−Boltzmann (PB) theory, have been widely accepted in studies of solvation interactions. In the PB-based implicit solvent models, the electrostatic interaction or, more fundamentally, the electrostatic potential is assumed to obey the classical PB equation:

$$\nabla \cdot [\varepsilon \nabla \phi] = -\rho - \lambda \sum q_i n_i^0 \exp[-\beta q_i \phi] \tag{1}$$

where ε is the dielectric constant, ϕ is the electrostatic potential, ρ is the solute charge density, λ has a value of 0 wherever mobile ions cannot penetrate and a value of 1 where they can, n_i^0 is the number density of mobile ions of type i in the bulk solution, q_i is the charge of the mobile ions of type i, and $\beta = 1/kT$. Here k is the Boltzmann constant, and T is the temperature. When the Boltzmann factor is close to zero, eq 1 can be linearized as

* Corresponding author. Telephone: (949) 824-9528. E-mail: ray.luo@uci.edu.

$$\nabla \cdot [\varepsilon \nabla \phi] = -\rho + \lambda \sum \beta q_i^2 \phi n_i^0 \qquad (2)$$

Adaptation of the PB solvent models to molecular simulations requires a numerical solution of the three-dimensional (3-D) partial differential equation. However, the numerical procedure has been a bottleneck, largely limiting the application to calculations with static structures only. The difficulty lies in the numerical procedure to apply such solvent models, which involves discretization of the partial differential equation into a system of linear or nonlinear equations that tends to be rather large; it is not uncommon to have millions of unknowns in biochemical applications. In addition, the setup of the system before the numerical solution and postprocessing to obtain energies and forces are both nontrivial. Three major discretization methods are widely used in biomolecular applications. The most commonly used approach is the finite-difference method.[15−32] In this method, the physical properties of the solution, such as atomic charges and dielectric constants, are mapped onto rectangular grid points, and a discrete approximation to the governing partial differential equation is produced. The second approach is the finite-element method,[33−38] which approximates the potential by a superposition of a set of basis functions. A linear or nonlinear system for the coefficients produced by the weak formulation has to be solved. The third approach is the boundary-element method.[39−52] In the boundary-element method, the Poisson or PB equation is solved for either the induced surface charge[39−41,43,45,46,49,52] or the normal component of the electric displacement[42,44,47,48,50,51] on the dielectric boundary between the solute and the solvent.

Due to the computational expense for solving the PB equation numerically, considerable efforts have been invested in approximating the solution of the PB equation via methods, such as the semianalytical generalized Born (GB) model,[53−63] the induced multipole model,[64] the dielectric screening model,[65,66] and others. The pairwise GB model, in particular, has been widely accepted as an efficient estimation of the solution of the PB equation, as recently reviewed.[5,6,9−11]

A crucial component of all implicit solvent models within the PB framework is the dielectric model, i.e., the dielectric constant distribution of a given solution system. Typically, a solution system is divided into the low-dielectric interior and the high-dielectric exterior by a molecular surface. That is to say that the molecular surface is used as the dielectric interface between the two piece-wise dielectric constants. In numerical PB calculations, such as the finite-difference methods, discretization of the molecular surface is required. One possible approach is to build the molecular surface analytically and then to map it onto a grid.[67−69] However, analytical procedures can be quite time-consuming and do not necessarily offer any advantages for finite-difference calculations because the surface must, in any case, be mapped onto a grid lattice. In this study, we focus on representations of molecular surface for numerical solutions of the PB equation.

In numerical solutions of the PB equation, the solvent excluded surface (SES) is the most used surface definition.[24,26]

Indeed, recent comparative analyses of both PB-based and TIP3P solvent models show that the SES definition is reasonable in calculation of reaction field energies and electrostatic potentials of mean force of hydrogen-bonded and salt-bridged dimers with respect to the TIP3P explicit solvent.[70−72] However, a previous test of the SES definition in the finite-difference solution indicates that it is numerically unstable for molecular dynamics.[73] Similar numerical difficulty was also observed in the pairwise GB method when the SES definition was used.[74] A major limitation of the SES definition is the re-entry volume; it is found that in simulations of proteins at room temperature, large re-entry volumes generated by nonbonded atoms come and go as often as every femtosecond when the nearby atoms undergo vibrational motion.[73] Thus, extremely large surface derivatives with respect to atomic coordinates may occur in the SES definition. In addition, surface cusp may also exist given certain combinations of atom and probe radii and arrangements of atoms.

The van der Waals (VDW) surface, or the hard sphere surface, represents the low-dielectric molecular interior as a union of atomic VDW spheres. With the VDW definition, surface derivatives with respect to the atomic coordinates are much better behaved, different from the SES definition. However, surface cusp at the joint between any two spheres may still cause instability in numerical solutions and in force interpretations, just like the situation in the SES definition. In addition, there exist many nonphysical high (solvent) dielectric pockets inside the solute interior when the VDW definition is used, as discussed in ref 75. These small buried "solvent pockets" result in a molecular interior too hydrophilic, which would cause proteins to unfold. In addition, the complex dielectric interface due to the buried solvent pockets also results in an unsmooth field distribution, leading to unstable dynamics simulations. Considering these limitations, the modified VDW definition was proposed. The basic idea of the modified VDW definition is to use the solvent accessible surface (SAS) definition for fully buried atoms and the VDW definition for fully exposed atoms.[73] This is realized with a set of conformation-dependent modified VDW radii, whose calculation requires the solvent accessible surface area of all atoms to determine their solvent accessibility.[73] Apparently the definition of modified VDW radii has to be smooth to be any use for dynamics simulations. The standard VDW surface can then be generated with the modified VDW radii. The harmonic dielectric smoothing is also applied to smooth the dielectric transition between solvent and solute.[76] Apparently, the dielectric distribution within and around buried atoms is very smooth, i.e., it is all part of the solute low dielectric. However, the dielectric distribution around exposed atoms can still show spatial fluctuation, as in the original VDW surface. Thus, an additional step in the modified VDW definition is used to smooth the spatial fluctuation around exposed atoms,[73] though it is difficult to implement and hard to be optimized to reproduce the SES.

Molecular Surface Calculation

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1159**

The density approaches have recently been developed and can be used for numerical PB solutions. Either a Gaussian-like function or a smoothed step function has been explored in previous developments.[77,78] In these approaches, a distance-dependent density/volume exclusion function is used to define each atomic volume or dielectric constant directly. This is in contrast to the hard-sphere definition of atomic volume, as in the VDW or the SES definition. Note that the use of a smooth function allows the "boundary region" or the "solute/solvent transition region" extends both inward and outward, i.e., the abrupt dielectric transition in the classical two-dielectric model has been replaced with a smooth dielectric transition region of finite width. Indeed, extension inward is necessary in the new definition to reproduce the results of the classical two-dielectric energetics for small training molecules.[77,78] In doing so, the trailing tail outside the cavity radii can be used to smooth out the small cracks and crevices formed by neighboring atoms. Apparently, any surface cusps are removed by the use of smooth density functions. However, their agreements with the classical two-dielectric model for typical macromolecules are not very good, as will be shown below even if the agreements can be excellent for the small training molecules.[77,78]

In this development, we have revisited the density function strategy by combining it with the modified VDW definition to balance numerical stability and model quality for numerical PB applications. Specifically, we have optimized the new density function to reproduce, as much as possible, the reaction field energies based on the SES definition due to its reasonable agreement with explicit solvent models. A large and diversified training set of biomolecules is used during the optimization. A separate test set of biomolecules is also used to validate its performance. Its benefit in improving the convergence and the numerical stability of electrostatic force calculations is also discussed.

Methods

Finite-Difference Method. In this work, we focus on numerical solution of the linearized PB equation. However, the proposed new numerical surface procedure can certainly be applied to the full nonlinear PB equation, which has been implemented in the Amber/PBSA program.[25,30−32] A widely used numerical method is the finite-difference method. It requires mapping the problem domain onto a lattice of grid points. The grid points are connected with grid edges. Solute atomic charge distribution is mapped onto grid points, and the dielectric constant distribution is mapped onto the centers of grid edges. The linearized PB equation can then be converted into a linear system with the finite-volume scheme. Under this discretization scheme, the partial differential equation can be written as follows at each grid point:

$$
\begin{aligned}
&-h^2\varepsilon\left(i-\tfrac{1}{2},j,k\right)\ \ [\phi(i-1,j,k)-\phi(i,j,k)]\\
&-h^2\varepsilon\left(i+\tfrac{1}{2},j,k\right)\ \ [\phi(i+1,j,k)-\phi(i,j,k)]\\
&-h^2\varepsilon\left(i,j-\tfrac{1}{2},k\right)\ \ [\phi(i,j-1,k)-\phi(i,j,k)]\\
&-h^2\varepsilon\left(i,j+\tfrac{1}{2},k\right)\ \ [\phi(i,j+1,k)-\phi(i,j,k)]\\
&-h^2\varepsilon\left(i,j,k-\tfrac{1}{2}\right)\ \ [\phi(i,j,k-1)-\phi(i,j,k)]\\
&-h^2\varepsilon\left(i,j,k+\tfrac{1}{2}\right)\ \ [\phi(i,j,k+1)-\phi(i,j,k)]\\
&+\kappa^2\qquad\qquad \phi(i,j,k)\qquad\qquad =h^{-3}q(i,j,k)
\end{aligned}
\tag{3}
$$

where h is the spacing in each dimension, i, j, and k are the grid indexes along the x, y, and z axes, respectively, and $\varepsilon(i \mp 1/2, j, k)$ is the dielectric constant between grids (i, j, k) and $(i \mp 1, j, k)$. Both $\varepsilon(i, j \mp 1/2, k)$ and $\varepsilon(i, j, k \mp 1/2)$ are defined similarly. In the Boltzmann term, κ^2 absorbs all the related coefficients, and $q(i, j, k)$ is the total charge within the cubic volume centered at (i, j, k). We can use several methods to solve the linear system, such as Gauss−Seidel, Jacobi, successive over relaxation, conjugate gradient, and so on.[79,80]

Dielectric Distribution Model. A key issue in the solution of eq 3 is how to map the dielectric constant distribution on all grid edges. In biomolecular calculations, the dielectric constant distribution often adopts a piece-wise constant model, where the dielectric within the molecular surface is assigned to that of the solute and the dielectric outside the molecular surface is assigned to that of the solvent. Within this model, the dielectric constant on a grid edge, apparently, should be assigned to the dielectric constant in this region where the two neighbor grid points belong. However, when the two neighbor grid points belong to different dielectric regions, i.e., when the grid edge is a boundary grid edge, its dielectric constant is nontrivial to assign because the dielectric constant is discontinuous across the interface. One simple treatment is the use of harmonic average (HA) of the two dielectric constants at the center of grid edges across the solute/solvent boundary.[76] For example, if $(i − 1, j, k)$ and (i, j, k) belong to solute and solvent regions, respectively, then there must be an intersection point on the grid edge between $(i − 1, j, k)$ and (i, j, k). Denote a as the distance from the intersection point to the grid point $(i − 1, j, k)$ and b as the distance from the same intersection point to the grid point (i, j, k). In HA, $\varepsilon(i − 1/2, j, k)$ is defined as

$$
\varepsilon\left(i-\tfrac{1}{2},j,k\right)=\cfrac{h}{\cfrac{a}{\varepsilon(i-1,j,k)}+\cfrac{b}{\varepsilon(i,j,k)}}
\tag{4}
$$

This strategy has been shown to improve the convergence of reaction field energies with respect to the grid spacing and to reduce the grid dependence of the solvation energetics.[76] The smoothed dielectric constant transition across the solute/solvent interface also makes it possible to compute dielectric boundary force via a variational approach proposed by Gilson et al.[81] Apparently, only the intersection points between the molecular surface and the boundary edges are needed to utilize eq 4 to assign dielectric constants at boundary grid edges.

A Revised Density Function Strategy. In this study, we explored representing the molecular surface *indirectly* or *implicitly* with a molecular density function. That is to say that the two-dimensional (2-D) molecular surface is represented as an equi-density surface of a 3-D density function. The strategy appears to complicate the numerical problem by increasing the dimensionality of the procedure. However, the primary aim is to reduce the numerical instability arising from using hard spheres in classical molecular surface representation. In addition, the density function strategy naturally fits in the level set method that can be utilized to compute, numerically, many properties of the molecular surface in a rather straightforward manner for the finite-difference methods, as will be described.

The central idea of our revised density function is the same as previous attempts to apply density functions to molecular surface or volume presentations,[77,78] i.e., to use a function that smoothly maps out the interior of the SES of a molecule. However, we require here that it also reproduces the solvation energetics computed with the hard-sphere-based SES definition as much as possible. Specifically, atoms are described by atomic density functions, and these are combined using a composite molecular density function that can be used to calculate discretized molecular surface, i.e., the intersection points between the molecular surface and the boundary grid edges. As discussed above, these points are used to define the dielectric distribution of the solution system by the harmonic average method.[76]

Specifically, given the nth atom centered at $\mathbf{r}_n$, its density ρ_n is defined as

$$\rho_n = \rho_n(x) \tag{5}$$

with $x = (d - r_c)/2r_p$. Here, $d = |\mathbf{r} - \mathbf{r}_n|$ is the distance to the atomic center ($\mathbf{r}_n$), r_c is the VDW radius of the atom, and r_p is the solvent probe radius. The domain of the independent variable, x, is then set to be $[-r_c/2r_p, 1]$, where $x = -r_c/2r_p$ corresponds to the atomic center and $x = 1$ corresponds to one probe diameter away from the atomic VDW surface. Note also that $x = 0$ corresponds to the atomic VDW surface, when x is so defined. Apparently the domain of the dependent variable, the density value (ρ_n), needs to satisfy certain constraints to be physical or reasonable and to be smooth for numerical PB methods. Here, the density function is required to satisfy:

$$\rho_n(x) > 1, \quad x < 0$$
$$\rho_n(x) = \begin{cases} 1, & x = 0 \\ 0, & x = 1 \end{cases} \tag{6}$$

The constraints are meant to specify that: (i) the density function is always positive; and (ii) the density value within any atomic VDW surface is guaranteed to be ≥ 1 within a molecule, as will become clear below. A main point of this study is not to use any prescribed function form but to optimize a numerical function, i.e., a table lookup function, which satisfies the above constraints, and to achieve the best possible agreement with a given benchmark for a specific training set. The details of the benchmark, the training set, and the quality measure will be discussed below. Of course, to guarantee smoothness and good numerical behaviors the

cubic-spline interpolation is used to interpolate the function value within the allowed range of x.

With the definition of atomic density functions, we can now define a molecular density function as

$$\rho_{mol}(\mathbf{r}) = 1 - \prod_n (1 - \rho_n) \tag{7}$$

in terms of a repeated product over atomic density functions.[77,78] Expanding this expression we obtain

$$\rho_{mol}(\mathbf{r}) = \sum_n \rho_n - \sum_{n>m} \rho_n\rho_m + \sum_{n>m>l} \rho_n\rho_m\rho_l + \ldots$$
$$= \rho_{sum}(\mathbf{r}) + \text{"intersection terms"} \tag{8}$$

where $\rho_{sum}(\mathbf{r}) = \sum_n \rho_n$ is the linear summation of all atomic terms, and the higher products, i.e., the "intersection terms", represent corrections for over or under counting of atomic intersections.[77,78] Use of $\rho_{mol}(\mathbf{r})$ would produce a molecular volume almost the same as the "classical" VDW volume, which is not desired in numerical PB applications as reviewed in the introduction.[77,78] The original density function (ρ_{mol}) was defined in such a way that $\rho_{mol} > 1$ corresponds the VDW volume. If only the leading term ρ_{sum} were used, then $\rho_{sum} > 1$ would correspond to an overestimation of the VDW volume. Corrections from the intersection terms in ρ_{mol} were used to successively correct the volume so that it eventually leads to a volume consistent with the VDW volume. That is to say that when only the leading term is used, the definition of $\rho_{sum} > 1$ has the potential to capture the "re-entry volume", in addition to the VDW volume when properly optimized. A similar idea was used in previous work.[77,78]

Combination with the Modified VDW Surface. Note that the molecular volume defined by a density function crucially depends upon the function form. A function defined to decay faster to zero would lead to a smaller molecular volume, i.e., more solvent-exposed interatomic crevices would exist. In contrast, a function defined to decay slower to zero would lead to a larger molecular volume, i.e., less solvent-exposed interatomic crevices would exist. Our initial analysis of the density function approach shows that its sensitivity to the function form makes it very difficult to reproduce the SES definition.

The limitation can be attributed to the different requirements between defining the solvent-exposed surface for exposed atoms and the solvent-excluded volume for buried atoms. Our analysis shows that the density function has to be defined to decay to zero faster to capture the solvent-exposed surface in the SES definition. However, the density function has to be defined to decay to zero slower to capture the solvent-excluded volume in the SES definition. Since more atoms on peptides are solvent-exposed, while more atoms in proteins are solvent-excluded, the density function tends to make the proteins too hydrophilic, if it is optimized with respect to the peptides. In contrast, the density function tends to make the peptides too hydrophobic, if it is optimized with respect to the proteins. Overall, it is too difficult to find a compromise that would work for different sized molecules.

We explored combining the density function with the modified VDW approach reviewed in the introduction to

Molecular Surface Calculation

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1161**

overcome its limitation. Specifically, the density function is revised as follows:

  A. Compute SAS with unmodified VDW radii (r_c).

  B. Compute modified VDW radii (r_c^m) based on atomic SAS (see below).

  C. Calculate the composite density function, $\rho_{sum}(\mathbf{r}) = \sum_n \rho_n$, with the modified VDW radii, i.e., with r_c replaced by r_c^m in eq 5.

The modified VDW radius of atom i is defined to be smoothly dependent upon atomic SAS as follows:

$$r_{c,i}^m = \begin{cases} r_{c,i} + \frac{1}{2}\eta r_p \left[1 + \cos\left(\pi \frac{A_i^s}{A^c}\right) \right], & A_i^s < A^c \\ r_{c,i}, & A_i^s \geq A^c \end{cases} \quad (9)$$

where $r_{c,i}^m$ is the modified radius of atom i, $r_{c,i}$ is the unmodified VDW radius of atom i, $\eta \in [0,1]$ controls how much increment to be added to the unmodified VDW radius, r_p is the solvent probe radius, A_i^s is the relative solvent accessibility of atom i, and $A^c \in [0,1]$ is the cutoff relative solvent accessibility. Equation 9 shows that only VDW radii of buried or somewhat exposed atoms are incremented, while VDW radii of highly exposed atoms are not incremented.[73]

Thus, when the density function is defined with the modified VDW radii, we can achieve the goal of raising the density function value higher within the molecular interior by effectively pushing the atomic VDW surface outward, while still preserving the low-density values for solvent-exposed atoms that are needed to define the solvent-exposed surface. As will be shown below, a more consistent performance can be realized among different sized molecules when the modified VDW radii are used in the density function approach. Of course, parameters η and A^c should be optimized, along with the density function, to best reproduce the reaction field energies with the SES definition.

Use of Numerical Function to Represent the Atomic Density Function. Up to this point, we have yet to define the atomic density function, except its overall properties. As discussed, a main point of this study is not to use any prescribed function form but to optimize a numerical function for $\rho_n(x)$ to achieve the best possible agreement between $\rho_{sum}(\mathbf{r})$ and the given benchmark for a specific training set. Specifically, five intervals in the range of [0, 1] were used to optimize the numerical function. These are then interpolated over the domain of interest with the cubic-spline interpolation to guarantee a continuous and smooth $\rho_n(x)$.[82] It is found that use of more intervals does not improve the quality of the agreement with the SES definition. Of course, the use of fewer intervals reduces the quality of the agreement.

Apparently, it is perfectly reasonable to include intervals <0 in the optimization, but these intervals are already within the atomic VDW volumes, which are guaranteed to be within the molecular surface and do not contribute to the optimization quality of the function. Inclusion of these intervals only increases the difficulty of the optimization problem. Instead, we extended the density function all the way to the atomic center with a linear function with the slope of the cubic-spline function at $x = 0$. For example, the atomic density
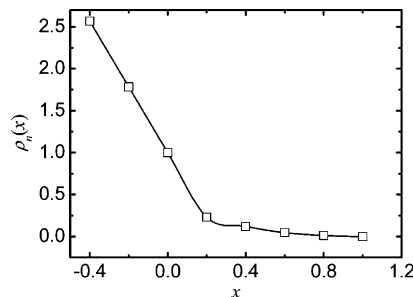

Figure 1. Optimized atomic density function.

function optimized within the range of [0, 1] is shown in Figure 1, which also plots two more intervals <0. Note also that the numerical function within the first interval [0, 0.2] is almost linear, even if the cubic-spline interpolation is used (Figure 1).

As described above, the numerical density function is optimized with a benchmark set. Here the PB reaction field energies in the SES definition were used in this study. Our choice of the SES definition as a benchmark was based on recent comparative analyses of both PB and TIP3P solvent models. These studies show that the SES definition is a reasonable surface definition in the calculation of reaction field energies and electrostatic potentials of mean force of hydrogen-bonded and salt-bridged dimers, with respect to the TIP3P explicit solvent.[70−72] Finally, to ensure transferability and universality of the density function, we used a large set of diversified biomolecules, both folded and unfolded conformations, in the optimization of the density function.

In summary, the following parameters were optimized: (i) the four density function values at $x = 0.20$, 0.40, 0.60, and 0.80, respectively, to determine $\rho_n(x)$; and (ii) η and A^c to determine the optimal modified VDW radii (eq 9). The optimization was conducted with respect to the benchmark reaction field energies computed with the SES definition. The average error (i.e., unsigned relative deviation) between the two sets of reaction field energies was used as the optimization measure. Note that the function values at $x = 0$ and 1 have been fixed as 1.0 and 0.0, respectively, according to eq 6. In this study, we used a three-step systematic scan of the parameter space to optimize the parameters. At step one, an initial scan in the resolution of 0.2 was used for all six parameters. At step two, a refinement scan in the resolution of 0.05 was used in the reduced search space (± 0.2), centered on the best parameter set from step one. Finally, at step three, a second refinement scan in the resolution of 0.01 was used in the reduced search space of (± 0.05), centered on the best parameter set from step two. The optimized parameters are shown in the Results and Discussion Section.

Implicit Molecular Surface Representation by the Level-Set Method. Once $\rho_{sum}(\mathbf{r})$ is defined on grid points, we need to know where the equi-density surface, $\rho_{sum}(\mathbf{r}) = 1$, intersects all grid edges to use the harmonic average method to set up the dielectric constants of all boundary grid edges, see eq 4. In addition, we also need the surface normal direction in the force calculation, as will be shown next.

We explored computing these numerical surface properties with the level set method.[82,83] In the level set method, a scalar

function, $\varphi(\mathbf{r})$, i.e., the level set function, is used to represent the surface implicitly. For example, to describe a 2-D spherical surface with the radius of 1, we can define a level set function as $\varphi(\mathbf{r}) = |\mathbf{r}|^2 - 1$ in the 3-D space. Thus, the spherical surface consists of all the points satisfying the condition of $\varphi(\mathbf{r}) = 0$. In addition, for any point $\mathbf{r}$, we can easily determine it is within the surface, if $\varphi(\mathbf{r}) < 0$ or outside of the surface, if $\varphi(\mathbf{r}) > 0$. It should be pointed out that for a specific surface ($\varphi(\mathbf{r}) = 0$), there may be many different level set functions because the requirement of $\varphi(\mathbf{r}) = 0$ cannot uniquely define the 3-D function. To utilize the level set method to define the molecular surface, we need to define a level set function so that $\varphi(\mathbf{r}) = 0$, i.e., the zero level set corresponds to the molecular surface. In addition, positive and negative function values denote the solvent and the solute sides of the surface, respectively. Thus, our revised density function can be used to define a level set function as

$$\varphi(\mathbf{r}) = 1 - \rho_{sum}(\mathbf{r}) \qquad (10)$$

With these preparations, we proceed to compute the interaction point of a boundary grid edge and a molecular surface, as follows. Without loss of generality, suppose that this is an x-edge flanked by two grid points x_1 and x_2. The level set function values are φ_1 and φ_2, respectively. Apparently, we have $\varphi_1 \times \varphi_2 < 0$, since the sign of the level set function, defined by eq 10, changes when crossing the molecular surface and the intersection point is between x_1 and x_2. Next, we choose a third grid point, x_3, from the two grid points that flank the x-edge in the x-direction, i.e., it is right next to x_1 or x_2 and has the same y and z coordinates. Given the three grid points with their x coordinates as x_1, x_2, x_3 and with corresponding level set functions as φ_1, φ_2, φ_3, respectively, a quadratic function $\varphi = a_2 x^2 + a_1 x + a_0$ can be determined to pass through three points (x_1, φ_1), (x_2, φ_2), and (x_3, φ_3). Thus, the intersection point is simply the root of the quadratic equation $a_2 x^2 + a_1 x + a_0 = 0$, within $[x_1, x_2]$. It has been shown that the error in the calculated intersection point scales as $O(h^2)$.[82,83]

With the level set method, calculation of other surface properties is also straightforward. In this study, the surface normal direction is needed at every intersection point and can be computed as[82,83]

$$\mathbf{n} = \frac{\nabla\varphi}{|\nabla\varphi|} = \frac{(\varphi_x, \varphi_y, \varphi_z)}{(\varphi_x^2 + \varphi_y^2 + \varphi_z^2)^{1/2}} \qquad (11)$$

where φ_x is the derivative of φ, with respect to x. Other symbols are defined similarly. The simplicity of the level set method makes it well suited to the finite-difference method, where these derivatives can be interpolated with accuracy of $O(h^2)$.[82,83]

Computation Details

Electrostatic Energy and Force Calculation. The electrostatic reaction field energy was computed via the dielectric polarization charges, which were calculated using the Gauss law and the grid potential obtained from the finite-difference solution of the PB equation.[26] In the finite-difference method,

the dielectric polarization charges are located on the boundary grid points, i.e., the grid points surrounded by nonuniform dielectric grid edges. To improve the convergence of reaction field energy, the polarization charges were first projected onto the molecular surface, according to the procedure described by Rocchia et al.,[26] before they were used to compute the reaction field energy as a pairwise summation of Coulombic interactions between atomic and polarization charges.

It is well-known that the electrostatic force density can be derived through the divergence of the Maxwell stress tensor ($\mathbf{P}$) as[84,85]

$$\begin{aligned} \mathbf{f} = \nabla \cdot \boldsymbol{P} &= \frac{\partial}{\partial x}(i \cdot \boldsymbol{P}) + \frac{\partial}{\partial y}(j \cdot \boldsymbol{P}) + \frac{\partial}{\partial z}(k \cdot \boldsymbol{P}) \\ &= \rho^f \mathbf{E} - \frac{1}{8\pi}E^2\nabla\varepsilon - \Delta\Pi\nabla\lambda \end{aligned}$$

$$(12)$$

where ρ^f is the fixed charge density, $\mathbf{E}$ is the electric field, ε is the dielectric constant, $\Delta\Pi$ is the excess osmotic pressure,[86] and λ is the Stern layer defined so that it is 1 in regions accessible to the mobile ions and 0 elsewhere. This is consistent with the formulation of Gilson et al.[81]

Equation 12 shows that there are three components in the total electrostatic forces: (i) the Coulombic and reaction field forces acting on the atomic charges, $\rho^f\mathbf{E}$; (ii) the dielectric boundary forces, or pressure acting on the dielectric boundary, $-(1/8\pi)E^2\nabla\varepsilon$; and (iii) the ionic boundary forces, or pressure on the ionic boundary. Since the Coulombic forces can be computed analytically by pairwise summation of Coulombic interactions among atomic charges, only the rest of the force components were computed numerically.

Similar to the treatment of reaction field energy, dielectric polarization charges can be used to improve the convergence of reaction field forces, with respect to the grid spacing. Specifically, the reaction field forces were calculated by the pairwise summation of the Coulombic interactions between polarizarion and atomic charges.[26] The computation of dielectric boundary forces requires the derivative of the dielectric constant. Thus, only a smoothly varied dielectric model, such as eq 4, can be used in eq 12,[85] where the dielectric constants on grid edges across the dielectric boundary (molecular surface) were assigned as weighted harmonic averages of solvent and solute dielectric constants. The finite-difference procedure to implement $-(1/8\pi)E^2\nabla\varepsilon$, given the weighted harmonic averages of dielectric boundary grid edges, has been described in detail by Gilson et al.[81] Note that the dielectric boundary force element is always along the direction of the gradient of dielectric constant, which is the normal direction of the molecular surface. Thus, computation of the dielectric boundary force element requires the numerical calculation of surface normal vectors, which was computed by eq 11.

Due to the fact that the SES surface is not differentiable, the dielectric boundary force elements are distributed to nearby atoms in an ad hoc manner, as follows. For the contact portion of the SES, the surface force elements are distributed to the closest atom sphere. For the re-entry portion of the SES, the dielectric boundary force elements are distributed to the two nearest atom spheres proportional to the inverse

Molecular Surface Calculation

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1163**

Table 1. Training (1−3) and Test Sets (4−8) Used in This Study[a]

structure set	no. structures	no. residues	rmsd (Å)
1: PDB1	290	16−517	0.00
2: PGB	314	56	0.01−13.80
3: HPN1	300	16	0.02−11.63
4: PDB2	289	19−497	0.00
5: ENH	261	54	0.02−10.50
6: SHG	284	57	0.16−13.22
7: helix	501	19	0.02−8 0.95
8: HPN2	451	16	0.02−11.28

[a] Root-mean-square deviation from the experimental native structure is defined as rmsd.

Table 2. Average Error, Maximum Error, And Error Spread (Average/Max/Spread) for Each of the Training and Test Sets[a]

structure set	revised density	density	MVDW
1: PDB1	0.47/1.82/3.11%	1.26/6.58/9.40%	2.01/6.66/8.07%
2: PGB	0.76/2.38/3.05%	0.57/3.50/4.77%	4.39/6.75/5.37%
3: HPN1	0.76/3.26/5.90%	1.84/3.72/5.36%	4.37/9.32/10.47%
4: PDB2	0.45/2.40/3.86%	1.27/8.44/10.85%	1.95/5.95/6.83%
5: ENH	0.58/1.72/2.50%	0.56/2.23/2.90%	3.18/5.27/4.39%
6: SHG	0.89/2.47/3.01%	0.71/2.98/4.32%	4.72/7.42/6.11%
7: helix	0.54/1.87/3.66%	1.29/2.82/5.61%	3.88/7.67/7.46/%
8: HPN2	0.58/3.38/5.53%	1.24/3.12/5.95%	3.89/8.52/9.22%

[a] Revised density: density function definition with modified VDW radii. Density: density function definition with unmodified VDW radii. MVDW: VDW surface definition with modified VDW radii.

of the distances of the two atom spheres. This procedure is also used to distribute the dielectric boundary force elements for the surface defined by the revised density function.

Finally, the ionic boundary forces are $\sim O(10^{-2})$ smaller than those of the reaction field forces and the dielectric boundary forces in water,[81] so that we only focus on the reaction field forces and dielectric boundary forces in the performance analysis of the revised density function below. In addition their performance is apparently more closely related to how the Stern layer is defined, which is beyond the scope of this development.

Training and Test Sets. To cover the highly heterogeneous molecular surface topologies, eight different structure sets were used in this study to calibrate and to test the revised density function (see Table 1). Sets 1 and 4 contain a large set of Protein Data Bank (PDB) structures, ranging from small peptides to very large biomolecules (with more than 550 residues) and covering a wide variety of native protein folds.[87] Sets 2, 5, and 6 are unfolded protein conformations of three small global proteins from high-temperature unfolding simulations: 1PGB (alpha/beta), 1ENH (all-alpha), and 1SHG (all-beta). Sets 3, 7, and 8 are from unfolding simulations of three peptides. Set 3 is a hairpin from 1PGB, set 7 is a helix from 1PGB, and set 8 is a hairpin from 1SHG. These sets are used to test how well reaction field energies are reproduced for different native and nonnative conformations of the same protein. Finally, sets 1−3 were used as training sets, and sets 4−8 were used as test sets. All molecular structures were processed with Leap in Amber9[88] and held static in all calculations. As described above, the optimization of the density function was conducted with the reaction field energies of chosen training molecules. In the surface potential analysis, the native structures of 1PGB, 1ENH, and 1SHG and the p53 DNA binding domain with and without DNA were used.

Other Details. In all calculations, the ion concentration was set to 0. The dielectric constant of the solvent was set to 80, i.e., for water, while the solute was set to 1. The solvent probe was set to be 0.6 Å. Our use of an unusually small probe radius was based on our previous comparative analyses of the numerical PB and TIP3P solvent models. In these analyses, the reaction field energies of small molecules were found to be not very sensitive to the different probe sizes, but the electrostatic potentials of mean force of the hydrogen-bonded or the salt-bridged dimers were quite sensitive to the probe radius used, and a solvent probe radius of 0.6 Å

was found best to reproduce the TIP3P solvent among the tested values.[72] Subsequent analysis of the ion pairs on peptides and proteins also indicates that the probe radius of 0.6 Å can best reproduce the TIP3P solvent [Tan and Luo, manuscript in preparation]. The finite-difference grid spacing was set to be 0.5 Å, if not mentioned otherwise. A two-level electrostatic focusing was used to speed up the assignment of electrostatic boundary condition. The coarse grid spacing was 2.0 Å. The dimension of the coarse grid was set to be twice as large as the dimension of the solute to secure the quality of the nonperiodic boundary condition in the PB calculations. The finite-difference convergence criterion was set to be 10^{-3}. All other parameters are set to be default as in the PBSA program of Amber 9.[25,73,88]

Results and Discussion

Optimization of the Revised Density Function. As described in the method, the revised density function was optimized by a systematic search of the six parameters with respect to a benchmark set of reaction field energies computed with the SES definition. The optimized parameters, with the lowest average unsigned error with respect to the benchmark set, are as follows: $\rho_n(x) = 0.21, 0.15, 0.05,$ and 0.01 at $x = 0.2, 0.4, 0.6,$ and 0.8, respectively; $\eta = 0.57$; and $A^c = 0.08$. Once the discrete density function values are given, the cubic-spline function is utilized to interpolate the density function within the range of [0, 1] for x. Note that $x = 0$ corresponds to the atomic VDW surface, and the atomic density function is defined for $x < 0$ as a linear function with the slope at $x = 0$. Figure 1 plots the cubic-spline interpolated function with the optimized discrete density function values. To fully understand the performance gain in combining the density function with the modified VDW surface, the pure density function approach, i.e., with density function defined with the unmodified VDW radii, was also optimized with respect to the same benchmark set. The optimized parameters are as follows: $\rho_n(x) = 0.26, 0.21,$ 0.05, and 0.01 at $x = 0.2, 0.4, 0.6,$ and 0.8, respectively.

It can be seen in Table 2, the revised density function with the modified VDW radii, denoted as "revised density", can achieve average errors less than 1% (PDB1: 0.47; PGB: 0.76; and HPN1: 0.76%) for all three training sets, a respectable agreement. The corresponding values of the density function with the unmodified VDW radii, denoted as "density",
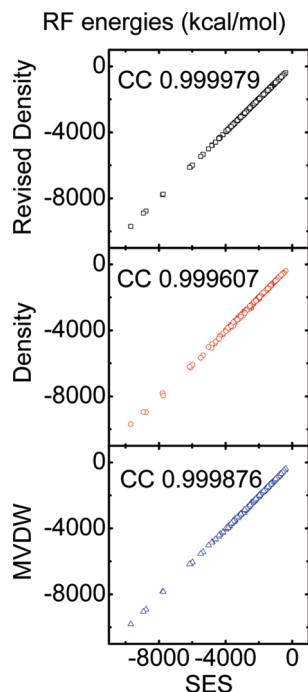
Figure 2. Top: Correlation between reaction field (RF) energies computed with the SES definition and those computed with the revised density definition (correlation: 0.999979, slope: 0.999505, offset: 5.30848, and rms relative deviation: 0.57%). Middle: Correlation between RF energies computed with the SES definition and those computed with the density definition (correlation: 0.999607, slope: 1.01631, offset: 11.1488, and rms relative deviation: 1.79%). Bottom: Correlation between RF energies computed with the SES definition and those with the MVDW definition (correlation: 0.999876, slope: 1.01112, offset: −15.306, and rms relative deviation: 2.17%).

(PDB1: 1.26; PGB: 0.57; and HPN1: 1.84%) and modified VDW definition, denoted as "MVDW", are relatively larger (PDB1: 2.01; PGB: 4.39; and HPN1: 4.37%). The average errors of the revised density definition are also lower than those of the density and the MVDW definitions in the test sets. Overall the average error of every training or test set is lower than 1% (0.45−0.89%). For practical purposes, it is also important, however, to minimize maximum errors and error spreads. Here, the error spread is defined as the difference between the maximum positive and negative errors. It can be seen from Table 2 that the unsigned maximum errors by the revised density definition are smaller than those of the density and the MVDW definitions (revised density: 1.72−3.38%; density: 2.23−8.44%; and MVDW: 5.27−9.32%) and that the error spreads of the revised density definition are narrower (revised density: 2.50−5.90%; density: 2.90−10.85%; and MVDW: 4.39−10.47%). Overall, the revised density definition can achieve a more consistent performance with lower average errors, maximum errors, and error spreads than those of the density and MVDW definitions in both the training and test sets.

We next consider the correlation between the reaction field energies computed with the revised density definition and those with the SES definition for test set PDB2. As shown in Figure 2, all data fall on the diagonal line with almost no scatter. Indeed, the correlation coefficient is 0.999979, the
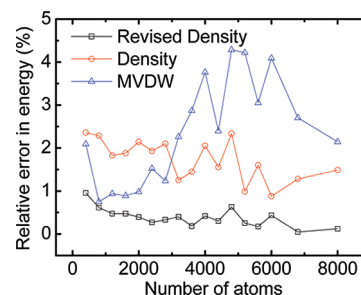

Figure 3. Size dependence of relative errors in reaction field energies for the revised density (square), the density (circle), and the MVDW (triangle) definitions. To improve clarity, the numbers of atoms are binned with a width of 400, and the error is averaged for each bin before plotting.

slope is 0.999505, the offset is 5.30848, and the root-mean-square (rms) relative deviation is 0.57%. In contrast, the correlation between the density definition and the SES definition is 0.999607, the slope is 1.01631, the offset is 11.1488, and the rms relative deviation is 1.79%. The correlation between the MVDW and the SES definitions is 0.999876, the slope is 1.01112, the offset is −15.306, and the rms relative deviation is 2.17%.

Transferability Considerations. While the overall relative error in reaction field energy should be as small as possible, it is also important that the error is both system and structure independent. We investigated two aspects in this regard: errors with different system sizes and different conformations (loosely packed/unfolded or folded conformations). To study the transferability of the new method, we calculated the errors versus number of atoms for the PDB2 set. Figure 3 shows that the error range of the revised density definition does not change much as the number of atoms increases. This confirms that the performance of the revised density and the density definitions are mostly size independent, though the revised density definition clearly agrees better with the SES definition. On the other hand, the size-dependence effect of the MVDW definition is obvious; the error increases as the number of atoms increases.

Besides the size dependence, we also studied the conformation dependence of the revised density definition. We unfolded two proteins (SHG/ENH) and two peptides (helix/HPN2) with high-temperature (500 K) molecular dynamics simulations and analyzed the correlation between energy errors and conformations for about 250−500 collected snapshots for each molecule. Figure 4 shows that the conformation dependence of the revised density definition is lower, i.e., the energy errors are uniform over the different conformations, shown as different rmsd's. It is worth pointing out that in the SHG and ENH test sets, the reaction field energies of the revised density definition are similar to those of the density definition. In the helix and HPN2 test sets, however, the results of the revised density definition are much better than both the density and MVDW definitions. This demonstrates the more consistent performance of the revised density definition over different sized molecules.

Convergence and Stability of Atomic Electrostatic Forces. Given the overall improved agreement in the computed reaction field energies between the revised density

Molecular Surface Calculation

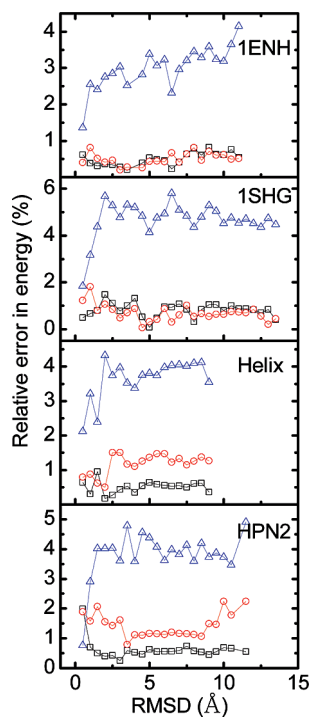*J. Chem. Theory Comput., Vol. 6, No. 4, 2010* **1165**

Figure 4. Conformation dependence of relative unsigned errors in reaction field energies for the revised density (square), the density (circle), and the MVDW (triangle) definitions. To improve clarity, the rmsd's are binned with a width of 0.5 Å, and the error is averaged for each bin before plotting.

Figure 5. Left: Correlation between reaction field (RF) forces computed with the revised density definition and those computed with the SES definition. Right: Correlation between dielectric boundary (DB) forces computed with the revised density definition and those computed with the SES definition. All forces are computed with a grid spacing of 1/8 Å.

and the SES definitions, we studied the convergence and the numerical stability of atomic electrostatic forces, which are important for stable dynamics simulations. The convergence quality of atomic electrostatic forces is measured by the correlation and the rmsd of atomic forces at a typical coarse grid spacing (1/2 Å) with respect to those at a fine grid spacing (1/8 Å). The numerical stability of atomic forces is measured by the average standard deviations of individual atomic forces computed with different finite-difference grid origins. Here 64 different finite-difference grid origins were used to analyze the numerical uncertainty of different methods. Two small systems, the native helix (from test set helix) and the native hairpin (from test set HPN2) structures, were selected in this analysis because they can be processed with the finest tested grid spacing, 1/8 Å, on our local computer servers.

Consistency of the Two Methods. While the two surface definitions are clearly different, we have tried to optimize the revised density definition so that the reaction field energies computed with the two surface definitions agree as much as possible. Thus, we expect the correlation of atomic forces computed with the two surface definitions to be reasonably well, at least at the finest grid spacing tested, 1/8 Å. The correlations between the two sets of forces are shown for the two tested molecules in Figure 5. The reaction field (RF) forces are shown on the left, with the correlation coefficients of 0.98730 for the helix and 0.98912 for the hairpin, indicating that the RF forces by SES are reasonably well reproduced by the revised density function. Similarly, the right panel of Figure 5 plots the correlations of dielectric
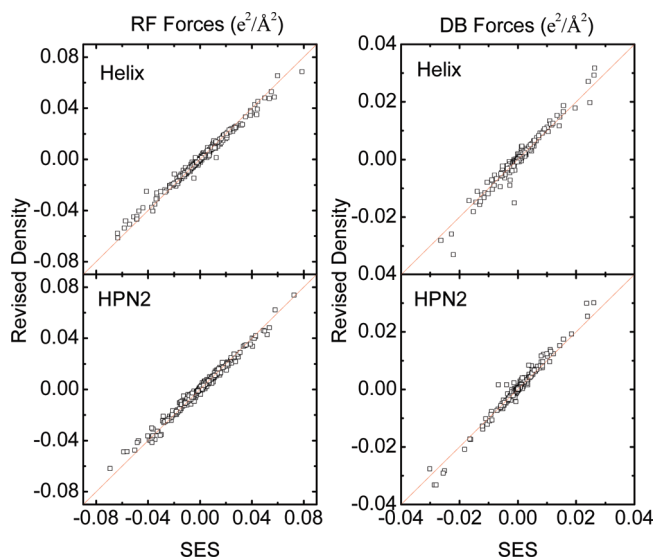
boundary (DB) forces between the two methods for the two tested molecules. The correlation coefficients are 0.95137 for the helix and 0.95848 for the hairpin. Due to the physical nature of the dielectric boundary forces, which act on the molecular surface, they are more sensitive to the exact location of molecular surface. Therefore, the correlation of DB forces is, in principle, lower than that of the RF forces between the two surface definitions.

Convergence of Forces. We next analyzed the convergence of forces at typical coarse grid spacing (1/2 Å) for biomolecular simulations for both the revised density and the SES definitions. The convergence measures show that the revised density definition converges faster than the SES definition in the numerical calculation of RF forces (Table 3) and DB forces (Table 4). The correlations of RF forces for the revised density function are 0.99411 for the helix and 0.98917 for the hairpin, respectively, while the corresponding correlations for the SES definition are 0.96625 and 0.91612, respectively. The rms deviations of RF forces computed with the revised density function are a factor about two smaller than the corresponding rms deviations with the SES definition. The convergence comparison for the DB forces is similar. The correlations of DB forces computed with the revised density function are 0.93229 for the helix and 0.95588 for the hairpin, respectively, while the corresponding correlations computed with the SES definition are 0.90205 and 0.84782, respectively. The rms deviations of DB forces computed with the revised density function are a factor of about two smaller than the corresponding rms deviations with the SES definition.

Numerical Stability of Forces. Finally, we analyzed the numerical stability of numerical electrostatic forces for both the revised density and the SES definitions. Our stability measure shows that the revised density definition is more stable than that of the SES definition in the numerical calculation of RF and DB forces (Tables 3 and 4); reductions

Table 3. Convergence and Numerical Stability of Reaction Field forces ($e^2/\text{Å}^2$) for the Tested Helix and Hairpin at the Coarse Grid Spacing of 1/2 Å[a]
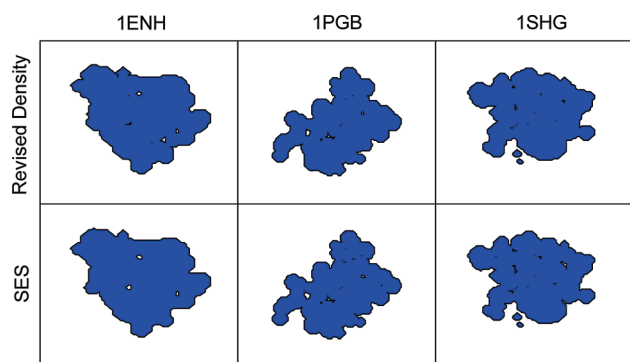
		revised density			SES		
	1/h	CC	rmsd ($\times 10^{-4}$)	σ ($\times 10^{-4}$)	CC	rmsd ($\times 10^{-4}$)	σ ($\times 10^{-4}$)
helix	2	0.99411	12.7	3.40	0.96625	22.5	5.04
	8	NA	NA	0.192	NA	NA	0.274
hairpin	2	0.98917	18.5	5.40	0.91612	39.5	7.59
	8	NA	NA	0.238	NA	NA	0.326

[a] SES: solvent excluded surface definition. CC: correlation coefficient between the forces at 1/2 and 1/8 Å. rmsd: rms deviation between the forces at 1/2 and 1/8 Å. σ: average standard deviation of individual atomic forces.

Table 4. Convergence and Numerical Stability of Dielectric Boundary Forces ($e^2/\text{Å}^2$) for the Tested Helix and Hairpin at the Coarse Grid Spacing of 1/2 Å[a]

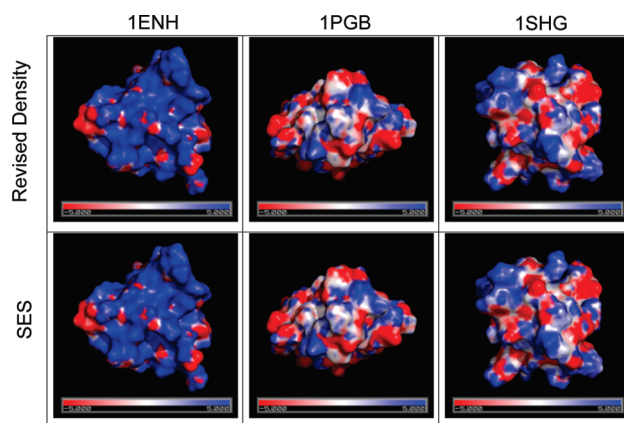|		revised density			SES		
	1/h	CC	rmsd ($\times 10^{-4}$)	σ ($\times 10^{-4}$)	CC	rmsd ($\times 10^{-4}$)	σ ($\times 10^{-4}$)
helix	2	0.93229	16.2	7.08	0.90205	22.4	15.9
	8	NA	NA	1.58	NA	NA	2.27
hairpin	2	0.95588	12.8	12.5	0.84782	33.0	24.9
	8	NA	NA	2.64	NA	NA	2.77

[a] See Table 3 for more information.

Figure 6. Molecular surface determined by the revised density and SES definitions for folded conformations of 1ENH, 1PGB, and 1SHG, respectively.

Figure 7. Surface electrostatic potential (in kT/mol-e) for protein 1ENH, 1PGB, and 1SHG, respectively (from left to right). Upper panel: computed with the revised density definition. Lower panel: computed with the SES definition. Potential is visualized in PyMol using a continuous color scale. Blue: positive values; white: zero; and red: negative.

in the standard deviations by a factor of about two were observed when the revised density definition was used for both types of forces.

Reproduction of Molecular Surface. Note that we have solely relied on reaction field energies in the optimization of the revised density definition. This is reasonable because the measure based on reaction field energies is quite sensitive to the exact locations of molecular surface. Nevertheless, we have also studied the quality of the revised density definition in reproducing the molecular surfaces with the SES definition. Figure 6 shows the molecular surfaces for folded 1ENH, 1PGB, and 1SHG. It can be seen that both definitions are quite consistent for these folded proteins. Figure S-1 in the Supporting Information shows the molecular surfaces for unfolded 1ENH, 1PGB, and 1SHG with the two definitions, respectively. Here slightly more discrepancy can be observed in the computed molecular surface. Overall, the molecular surface of both folded and unfolded conformations in the SES definition can be reasonably reproduced by the revised density function.

Reproduction of Surface Electrostatic Potential. We further analyzed the surface electrostatic potential of several selected proteins with the revised density definition. Figure

7 shows the surface potentials for 1PGB, 1ENH, and 1SHG calculated with the revised density function and the SES definition, respectively. Clearly, the two surface potential maps agree very well for the three selected proteins. For the larger p53 tetramer with and without DNA, the agreement is also excellent, as shown in Figure S-2 in the Supporting Information.

Limitation of the Density Function Approaches. A limitation of the revised density definition concerns the use of a "unconventionally" small solvent probe, 0.6 Å, which was optimized to reproduce the electrostatic potential of mean forces of dimers and surface salt bridges on macromolecules in the TIP3P explicit solvent. It is likely that the use of a small probe may result in a more hydrophilic interior in loosely packed macromolecules. Thus, we also tried to use the traditional solvent probe of 1.4 Å in our training of the revised density function. However, our analysis shows that the revised density definition can perform no better than the MVDW definition, even in the training sets (data not shown).

Thus, the limitation of the method to applications with a smaller solvent probe radii is clearly an issue that needs further considerations, to make it a more general numerical procedure for molecular surface calculations. Nevertheless, since our ultimate motivation in the study was to reproduce as close as possible the explicit solvent energetics, and it is unclear how quantitative loosely packed the molecular interior would be without detailed explicit solvent simulations, we delay further improvement to a future study. In addition, it is worth highlighting the intrinsic difference between all density function definitions and the SES definition to represent the molecular surface. Indeed, the multibody re-entry region in the SES definition is very difficult, if not possible, to reproduce with any additive density function approach.

Concluding Remarks

A revised density function is developed to define the molecular surface for dielectric assignment in the numerical PB methods. The density function is optimized through an exhaustive search using a benchmark data set of reaction field energies, computed with the solvent excluded surface for a wide-range of proteins and peptides in both folded and unfolded conformations. A uniformly low average unsigned error of less than 1% in reaction field energies can be achieved for both the training and test sets. The maximum errors and error spreads are also smaller than the modified van der Waals and the original density function definitions, upon which the new density function was developed. We further studied the transferability of the revised density function. Our data show that the average unsigned errors change little as the molecular size increases, confirming the size independence of the method. We also studied the conformation dependence of the revised density function with unfolded conformations of two proteins and two peptides. It turns out that there is little conformation dependence in the revised density function. Next, we studied the convergence and the stability of numerical electrostatic forces computed with the revised density function. The analysis shows that the revised density function can improve both the convergence and numerical stability by a factor of two over the solvent excluded surface. The analysis of computed molecular surfaces shows that the revised density function can well reproduce the solvent excluded surfaces for the three tested proteins in both folded and unfolded conformations. Finally, we also analyzed the performance of the revised density function in reproducing the surface electrostatic potential. The analysis shows that the agreements are very good for different sized proteins, whether it is neutral or highly charged.

Nevertheless, it is worth pointing out the limitation of the current development documented here. First, it is important to highlight that the numerical instability of the dielectric boundary forces is still higher than that of the reaction field forces. This is, in part, due to the use of grid-independent surface polarization charges in the computation of the reaction field energy and forces, which improves the convergence and stability in the computed energies and forces. In contrast, the dielectric boundary forces were computed directly with the finite-difference grid potential without further treatment. This is, in part, responsible for their larger error and uncertainty. We are actively working to improve the convergence and the stability of dielectric boundary forces by utilizing the grid-independent surface polarization charges, which would further enhance their numerical performance when combined with the revised density function. In addition, partition of dielectric boundary force elements on the solvent/surface interface to nearby atoms is clearly the next natural step before the method can be applied to routine molecular dynamics simulations. We are actively working on both issues in this group.

Finally, our analysis shows that the density approximation does not work well for "conventional" solvent probes of 1.4 Å. Fortunately, our prior comparative analysis indicates that the numerical PB methods with a probe radius of 0.6 Å reproduces the TIP3P solvent best on the tested conformations and systems. It is likely that the use of a small probe may result in a more hydrophilic interior in loosely packed macromolecules. This apparent dilemma between the molecular interior being too hydrophilic and the best agreement with TIP3P solvent points to the limitation of the hard-sphere-based strategies to define the dielectric model for numerical PB methods. To overcome this difficulty, we may revise the molecular surface calculation into a two-step procedure. First, we shall use a large solvent probe to calculate solvent accessibility of each atom. Of course, this procedure should be optimized to be consistent with solvent accessibility simulated in explicit solvent models. For the solvent inaccessible atoms, we will augment their VDW radii with an optimized amount that would effectively fill the molecular interior, resulting in a hydrophobic interior. Second, we will still use the density function optimized with a solvent probe of 0.6 Å to compute the re-entry region among the solvent-accessible atoms. The proposed strategy is still within the hard-sphere strategies but may resolve the dilemma discussed to a certain degree.

Supporting Information Available: The larger p53 tetramer with and without DNA. The molecular surfaces for unfolded 1ENH, 1PGB, and 1SHG with the two definitions, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.

(2) Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 216.

(3) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.

(4) Sharp, K. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 234.

(5) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.

(6) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(7) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1.

(8) Baker, N. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137.

Ye et al.

(9) Chen, J. H.; Im, W. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728.

(10) Feig, M.; Chocholousova, J.; Tanizaki, S. *Theor. Chem. Acc.* **2006**, *116*, 194.

(11) Im, W.; Chen, J. H.; Brooks, C. L. In *Peptide Solvation and H-Bonds*; Elsevier Academic Press Inc: San Diego, CA, 2006; Vol. 72, p 173.

(12) Koehl, P. *Curr. Opin. Struct. Biol.* **2006**, *16*, 142.

(13) Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A. *Communications in Computational Physics* **2008**, *3*, 973.

(14) Wang, J.; Tan, C. H.; Tan, Y. H.; Lu, Q.; Luo, R. *Communications in Computational Physics* **2008**, *3*, 1010.

(15) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1989**, *10*, 386.

(16) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 47.

(17) Luty, B. A.; Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1992**, *13*, 1114.

(18) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.

(19) Forsten, K. E.; Kozack, R. E.; Lauffenburger, D. A.; Subramaniam, S. *J. Phys. Chem.* **1994**, *98*, 5580.

(20) Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, *14*, 105.

(21) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59.

(22) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.

(23) Bashford, D. *Lecture Notes in Computer Science* **1997**, *1343*, 233.

(24) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327.

(25) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244.

(26) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.

(27) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *91*, 57.

(28) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 281.

(29) Wang, J.; Cai, Q.; Li, Z. L.; Zhao, H. K.; Luo, R. *Chem. Phys. Lett.* **2009**, *468*, 112.

(30) Cai, Q.; Wang, J.; Zhao, H. K.; Luo, R. *J. Chem. Phys.* **2009**, *130*, 145101.

(31) Cai, Q.; Hsieh, M. J.; Wang, J.; Luo, R. *J. Chem. Theory Comput.* **2010**, *6*, 203.

(32) Wang, J.; Luo, R. *J. Comput. Chem.* **2010**, DOI: 10.1002/jcc.21456.

(33) Baker, N.; Holst, M.; Wang, F. *J. Comput. Chem.* **2000**, *21*, 1343.

(34) Cortis, C. M.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1591.

(35) Holst, M.; Baker, N.; Wang, F. *J. Comput. Chem.* **2000**, *21*, 1319.

(36) Chen, L.; Holst, M. J.; Xu, J. C. *SIAM J. Numerical Anal.* **2007**, *45*, 2298.

(37) Shestakov, A. I.; Milovich, J. L.; Noy, A. *J. Colloid Interface Sci.* **2002**, *247*, 62.

(38) Xie, D.; Zhou, S. *BIT Numerical Mathematics* **2007**, *47*, 853.

(39) Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. *J. Chem. Phys.* **1987**, *87*, 1107.

(40) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

(41) Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1988**, *9*, 171.

(42) Juffer, A. H.; Botta, E. F. F.; Vankeulen, B. A. M.; Vanderploeg, A.; Berendsen, H. J. C. *J. Comput. Phys.* **1991**, *97*, 144.

(43) Rashin, A. A. *J. Phys. Chem.* **1990**, *94*, 1725.

(44) Yoon, B. J.; Lenhoff, A. M. *J. Comput. Chem.* **1990**, *11*, 1080.

(45) Bharadwaj, R.; Windemuth, A.; Sridharan, S.; Honig, B.; Nicholls, A. *J. Comput. Chem.* **1995**, *16*, 898.

(46) Purisima, E. O.; Nilar, S. H. *J. Comput. Chem.* **1995**, *16*, 681.

(47) Zhou, H. X. *Biophys. J.* **1993**, *65*, 955.

(48) Liang, J.; Subramaniam, S. *Biophys. J.* **1997**, *73*, 1830.

(49) Vorobjev, Y. N.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 569.

(50) Boschitsch, A. H.; Fenley, M. O.; Zhou, H. X. *J. Phys. Chem. B* **2002**, *106*, 2741.

(51) Lu, B. Z.; Cheng, X. L.; Huang, J. F.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19314.

(52) Totrov, M.; Abagyan, R. *Biopolymers* **2001**, *60*, 124.

(53) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(54) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.

(55) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.

(56) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.

(57) Im, W.; Feig, M.; Brooks, C. L. *Biophys. J.* **2003**, *85*, 2900.

(58) Feig, M.; Im, W.; Brooks, C. L. *J. Chem. Phys.* **2004**, *120*, 903.

(59) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383.

(60) Tanizaki, S.; Feig, M. *J. Chem. Phys.* **2005**, 122.

(61) Sigalov, G.; Scheffel, P.; Onufriev, A. *J. Chem. Phys.* **2005**, 122.

(62) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156.

(63) Grant, J. A.; Pickup, B. T.; Sykes, M. J.; Kitchen, C. A.; Nicholls, A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913.

(64) Davis, M. E. *J. Chem. Phys.* **1994**, *100*, 5149.

(65) Luo, R.; Moult, J.; Gilson, M. K. *J. Phys. Chem. B* **1997**, *101*, 11226.

(66) Li, X. F.; Hassan, S. A.; Mehler, E. L. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 464.

(67) Eisenhaber, F.; Argos, P. *J. Comput. Chem.* **1993**, *14*, 1272.

(68) You, T.; Bashford, D. *J. Comput. Chem.* **1995**, *16*, 743.

(69) Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1990**, *11*, 603.

(70) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. *J. Phys. Chem. B* **2005**, *109*, 14769.

(71) Tan, C. H.; Yang, L. J.; Luo, R. *J. Phys. Chem. B* **2006**, *110*, 18680.

(72) Wang, J.; Tan, C.; Chanco, E.; Luo, R. *Phys. Chem. Chem. Phys.* **2010**, *12*, In press.

(73) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035.

(74) Chocholousova, J.; Feig, M. *J. Comput. Chem.* **2006**, *27*, 719.

(75) Alexov, E. *Proteins: Struct., Funct., Bioinf.* **2003**, *50*, 94.

(76) Davis, M. E.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 909.

(77) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608.

(78) Grant, J. A.; Pickup, B. T. *J. Phys. Chem.* **1995**, *99*, 3503.

(79) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing*; Cambridge University Press: New York, NY, 1992.

(80) Stoer, J.; Bulirsch, R. *Introduction to Numerical Analysis*, 3rd ed.; Springer-Verlag: New York, NY,2002.

(81) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.

(82) Osher, S.; Fedkiw, R. P. *Level set methods and dynamic implicit surfaces*; Springer: New York, NY, 2003.

(83) Sethian, J. A. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*; Cambridge University Press: New York, NY, 1999.

(84) Landau, L. D.; Lifshitz, E. M.; Pitaevskii, L. P. *Electrodynamics of Continuous Media*, 2nd ed.; Butterworth-Heinemann: Oxford, 1993.

(85) Ye, X.; Wang, J.; Luo, R., submitted.

(86) Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684.

(87) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265.

(88) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.