INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
*Int. J. Numer. Meth. Biomed. Engng.* 2011; **27**:1964–1981
Published online 21 June 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.1450

# Pseudo-time-coupled nonlinear models for biomolecular surface representation and solvation analysis

Shan Zhao*,†

*Department of Mathematics, University of Alabama, Tuscaloosa, AL 35487, U.S.A.*

SUMMARY

This paper presents an improved mathematical model for biomolecular surface representation and solvation analysis. Based on the Eulerian formulation, the polar and nonpolar contributions to the solvation process can be equally accounted for via a unified free energy functional. Using the variational analysis, two nonlinear partial differential equations, that is, one Poisson–Boltzmann (PB) type equation for electrostatic potential and one geometric flow type equation defining the solute–solvent interface, can be derived. To achieve a more efficient and stable coupling, we propose a time-dependent PB equation by introducing a pseudo-time. The system of nonlinear partial differential equations can then be coupled via the standard time integration. Furthermore, the numerical treatment of nonlinear terms becomes easier in the present model. Based on a hypersurface function, the biomolecular surface is represented as a smooth interface. However, for the nonlinear PB equation, such a smooth interface may cause unphysical blowup because of the existence of nonvanishing values within the solute domain. A filtering process is proposed to circumvent this problem. Example solvation analysis of various compounds and proteins was carried out to validate the proposed model. The contributions of electrostatic interactions to the protein–protein binding affinity are studied for selected protein complexes. Copyright © 2011 John Wiley & Sons, Ltd.

Received 11 January 2011; Revised 18 March 2011; Accepted 11 April 2011

KEY WORDS: biological surface formation and evolution; potential driving geometric flows; generalized nonlinear Poisson–Boltzmann equation; multiscale models; solvation free energy

## 1. INTRODUCTION

Solute–solvent interactions are of central importance in the atomic scale description of complex biological processes because under physiological conditions, such processes naturally occur in aqueous environment. Solute–solvent interactions are typically described by solvation energies, which have both polar and nonpolar contributions. The polar portion of solvation accounts for the electrostatic interaction, which has a long range of influence on charged molecules, such as water, aqueous ions, and amino or nucleic acids. The nonpolar portion consists of the contributions due to the surface tension, mechanical work, and attractive solvent–solute dispersion interactions.

Two types of solvation models have been developed in the literature [1]: explicit and implicit solvent models. In explicit solvent models, the solvent is represented in molecular or atomic details, whereas in implicit solve models, it is treated as a continuous dielectric medium. Being computationally more expensive, explicit solvent models offer the greatest detail and potential for accuracy in solvation analysis [2]. By sacrificing the molecular details of the solvent, the implicit solvent models focus only on the biomolecules of interest [3]. Thus, the implicit solvent approaches usually can handle much larger biological systems so that such approaches play an important role in many applications of molecular simulation [4–6].

*Correspondence to: Shan Zhao, Department of Mathematics, University of Alabama, Tuscaloosa, AL 35487, U.S.A.
†E-mail: szhao@bama.ua.edu

For implicit solvent models, two different approaches are widely used to describe the polar solvation process, that is, the generalized Born (GB) methods and Poisson–Boltzmann (PB) methods. Being computationally faster, the GB methods are commonly used in high-throughput molecular dynamics applications [7]. They are usually parameterized and accessed by the PB theory. Derived based on rigorous biophysical theories, the PB theory formulates the electrostatic potential via a nonlinear elliptic partial differential equation (PDE). This nonlinear PDE can be linearized for cases where the interactions between the mobile ions and the solute electrostatic potential are weak. The PB methods, in which the PB equation is solved by means of numerical approximations [1, 4, 5, 8], usually offer a more accurate analysis of solvation in comparison with the GB methods, even though the PB methods are often slower. Moreover, the PB methods are well suited for problems in which global information about electrostatic properties are required for the purpose of visualization and surface analysis because the PB methods provide a global solution for electrostatic potential and surrounding field.

In all implicit solvent models, an interface location to separate the discrete solute region and the continuous solvent domain is needed in molecular simulations. For the PB and GB models, such an interface can be regarded as the solute–solvent boundary for defining the dielectric constant and ion accessibility coefficients. For nonpolar models, such an interface can be regarded as the surface or the profile of a molecule that, in turn, determines the area and volume of the molecule. Commonly used solvent–solute interface models include the van der Waals surface, the solvent accessible surface [9], and the molecular surface [10, 11]. Obviously, the interface definition affects the performance of implicit solvent models because all of the physical properties of interest in molecular simulations are very sensitive to the interface definition.

Even though the conventional surface models have found a great success in biomolecular modeling, a number of difficulties exist in the definition of the solute–solvent interface from the mathematical point of view. First, all traditional molecular surface models are *ad hoc* partitions to separate solute atoms from the surrounding solvent. In fact, the choice of solute–solvent interface in the PB theory remains to be a controversial issue in the literature [12, 13]. There lacks a physical principle to justify one surface definition over others. Second, the choice of solute–solvent interface becomes even more challenging in the PB simulations because of the use of filtered or smoothed versions [14, 15] of the conventional surface models in the numerical solutions of the PB equation. This is because the traditional surface definitions admit geometric singularities, such as cusps and self-intersecting surfaces [16, 17], that could lead to instability in molecular simulations. The smoothed surfaces, which are usually attained via some image processing techniques, can improve the regularities of the molecular surfaces. However, additional computational artifacts are also introduced [18]. Finally, different surface definitions and different smoothing techniques may require different parameters or radii in implicit solvent simulations. A wide range of such parameter sets may lead to confusion and misuse.

One way to overcome the aforementioned difficulties is to treat the definition of the solvent–solute interface as a part of a free energy minimization process based on the fundamental laws of physics. Mathematically, the free energy optimization can be accomplished with the Euler–Lagrange variation, which typically leads to a PDE as the molecular surface model. Converging to the same global extrema, the PDE approach is easier to implement numerically than the direct minimization procedure. Moreover, biomolecular surfaces generated by the PDE models are typically free of geometric singularities. The first PDE based molecular surface was constructed by Wei *et al.* [19] in 2005. This approach distinguishes itself from many other PDE-based surface smoothing methods [20, 21] by utilizing only atomic information, that is, atomic coordinates and radii, instead of an existing surface. The atomic information is embedded in the Eulerian formulation, and the molecular surface is generated via a gradient-driven diffusion process and a subsequent level-set extraction [19].

Driven by the desire to understand the true physical boundary of a biomolecule in a solvent, we have recently proposed a mean curvature flow PDE model of bimolecular surfaces, that is, the minimal molecular surface (MMS) [22, 23], that minimizes the surface free energy of the macromolecule in the aquatic environment. Based on the differential geometry theory, the intrinsic curvature force is used to drive the biosurface evolution and formation. For simple models of solute–solvent interactions, interface free energy minimization is often equivalent to surface area minimization and gives



DOI: 10.1002/cnm

rise to the MMS. To account for more complicated solute–solvent interactions, we have further proposed potential driven geometric flows (PDGFs) [24] that embed many other potential effects into surface formation and evolution. The PDGFs are inherently multiscale in nature and enable the incorporation of microscopic interactions, such as van der Waals potentials, into the macroscopic curvature evolution. Physical properties, such as free energy optimization (area decreasing) and incompressibility (volume preserving), were realized in the PDGF variational formulation.

Recently, a family of differential geometry-based multiscale models have been developed by Wei [25] for various chemical and biological systems, including protein complex, molecular motors, fuel cells, ion channels, DNA packing, nanofluidic systems, and virus evolution. These novel models describe not only the structure and solute–solvent interactions but also the dynamics and transport of various chemical and biomolecular systems. In a general framework, four types of PDEs are derived for different parts of complex systems, including fluid dynamics, molecular dynamics, electrostatic interactions, and surface dynamics. For excessively large systems, a generalization by replacing the expensive molecular dynamics with a macroscopic elastic description is also presented [25]. The general usage of various PDE-governing equations and the necessary coupling of them for a series of chemical and biological systems ranging from simple to complex are discussed [25].

More recently, the particular formulation and computational realization of the differential geometry-based model by Wei [25] for analyzing the equilibrium properties of solvation have been presented by Chen *et al.* in both the Eulerian and Lagrangian formulations [26, 27]. The minimization of the total free energy functional, which encompasses not only the curvature effect, the mechanical work and the electrostatic potential, but also the solute–solvent interactions, is conducted. The variation of the total free energy functional leads to two coupled nonlinear PDEs—one PB type equation for electrostatic potential and one geometric flow type equation defining the solute–solvent interface.

Mathematically, the Lagrangian formulation makes the direct use of differential geometry theory of manifolds and enables the sharp solute–solvent interface being directly evolved according to governing PDEs [27]. The Lagrangian formulation can be rigorously converted into the Eulerian formulation by means of geometric measure theory [26]. In the Eulerian formulation, the biomolecular surface is embedded in a hypersurface function or a level set function, and such a function is evolved under prescribed PDEs. It is interesting to note that besides the regular sharp interface of solute–solvent, the Eulerian formulation allows a smooth interface, that is, an overlapping transition region as the solute–solvent boundary [26]. Mathematically, the coarea theorem is used to represent the surface area density in the free energy functional of the Eulerian formulation. Computationally, the smooth interface possesses good stability and differentiability and generates a continuous dielectric profile so that a simpler numerical discretization can be employed.

Many mathematical and computational issues remain unexplored in the aforementioned differential geometry-based solvation models [25–27]. For example, the coupling of electrostatic potential and biomolecular surface evolution deserves a further study. Mathematically, the PB type equations are nonlinear elliptic equations, whereas the geometric flow equations are nonlinear parabolic equations. An iterative procedure was used to couple two PDEs [26, 27]. To ensure the stability, the hypersurface function and electrostatic potential computed for the next stage have to be balanced by the ones in the previous stage according to a pair of controlling parameters [26, 27]. Also, the solution of the boundary value problems for electrostatic potential was not conducted at every time step in integrating the hypersurface function but at certain number of time steps, which amounts to another coupling parameter. A more efficient coupling with less controlling parameters is obviously desired. Moreover, the solvation analysis and computation were realized via linearized PB equations in the existing simulations [26, 27]. The solution of nonlinear PB equations has not been considered.

The objective of the present work is to introduce a pseudo-time-coupled nonlinear PDE model for the theoretical modeling of biomolecular surface and solvation process. The proposed model will be employed for the solvation analysis and computation in the Eulerian formulation with smooth solute–solvent interfaces and can be similarly extended to the Lagrangian formulation with sharp interfaces. By treating the nonlinear PB equation as the steady state solution of a time-dependent process, the overall coupling can be accomplished via standard time integration techniques and controlled simply by time increments. The nonlinear PB equation thus can be treated in the same



DOI: 10.1002/cnm

manner as the linearized PB equation, with a moderate growth in CPU time. Moreover, a subtle problem is found when solving nonlinear PB equations with the smooth interface of the solute–solvent. Unphysical blowing-up solutions could be encountered at the smooth transition region of the solute–solvent boundary. A filtering process is considered so that the unphysical blowup can be prevented in numerical simulation.

The rest of paper is organized as follows. In Section 2, we first establish the theoretical framework of the Eulerian formation of differential geometry-based solvation models. Two governing nonlinear PDEs are derived by variational principles. Through the introduction of time coupling, the details of the proposed nonlinear model will be presented. Numerical validations and applications to various benchmark biological systems are considered in Section 3. Finally, this paper ends with a conclusion.

## 2. MATHEMATICAL MODELS AND NUMERICAL ALGORITHMS

### 2.1. Solute–solvent boundary

Consider a solute macromolecule being immersed in an aqueous solvent environment. As in the implicit solvent models, the macromolecule is described in discrete atomic detail, whereas the aqueous solvent is treated as a continuum. Thus, the entire domain $\Omega \subset \mathbb{R}^3$ of this solute–solvent system naturally consists of two regions, that is, macromolecular domain $\Omega_m$ and aqueous solvent domain $\Omega_s$. Here, $\Omega = \Omega_s \cup \Omega_m$. Besides a sharp interface of solute–solvent boundary, the Eulerian formulation also allows a smooth interface, that is, an overlapping transition region $\Omega_b$ as the solute–solvent boundary. Here, $\Omega_b = \Omega_s \cap \Omega_m \neq \emptyset$. This multi-domain setting can be represented by a characteristic function of the solute domain $S : \mathbb{R}^3 \to \mathbb{R}$. It is defined to be one ($S = 1$) inside the biomolecule and zero ($S = 0$) in the aquatic solvent. A continuous transition with values from one to zero is assumed by $S$ at the solute–solvent boundary region $\Omega_b$. Obviously, $(1 - S)$ is the characteristic function for the solvent domain. Sample profiles of $S$ and $(1 - S)$ along one straight line are depicted in Figure 1(a). It can be seen that the characteristic functions $S$ and $(1 - S)$ are zero in the opposite regions but take nonzero values at the solute–solvent boundary where the solvent and solute regions overlap. In the present study, the physical energy minimization principle will be employed to determine the characteristic functions and the profile of the solute–solvent boundary, which in turn, decide various physical and chemical properties of biomolecules.

### 2.2. Total free energy functional

A combination of polar and nonpolar components is considered for the total free energy functional. The polar component, that is, the polar solvation energy, is due to electrostatic interaction—a process that is of long range in nature and ubiquitous for any system of charged or polar molecules. For a sharp interface of solute–solvent, a widely used free energy functional of the electrostatic systems is due to Sharp and Honig [28] and Gilson *et al.* [29]. For a smooth interface of solute–solvent,

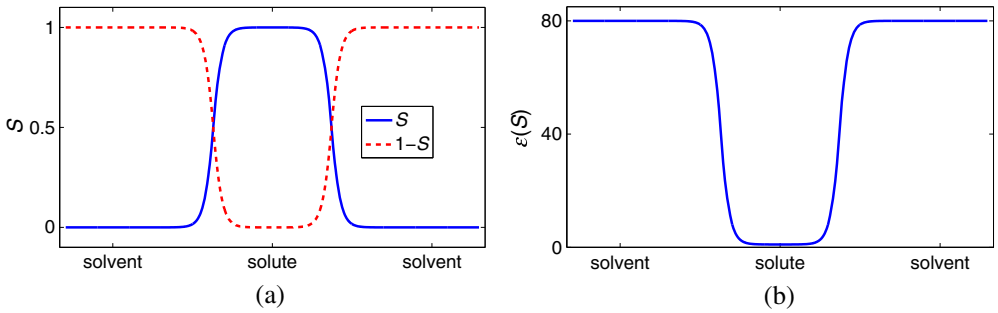


Figure 1. (a) The characteristic functions $S$ and $(1 - S)$ of the solute and solvent regions along one straight line. (b) The dielectric function $\epsilon(S)$ along one straight line with $\epsilon_s = 80$ and $\epsilon_m = 1$.



DOI: 10.1002/cnm

a polar solvation functional with the characteristic function $S$ being appropriately incorporated is suggested by Wei [25] and Chen *et al.* [26]

$$G_{\mathrm{p}} = \int_{\Omega} \left\{ S\left[\rho_{\mathrm{m}}\phi - \frac{\epsilon_{\mathrm{m}}}{2}|\nabla\phi|^2\right] + (1-S)\left[-\frac{\epsilon_{\mathrm{s}}}{2}|\nabla\phi|^2 - \mathrm{k_B}T\sum_{j=1}^{N_c} c_j(\mathrm{e}^{-q_j\phi/\mathrm{k_B}T}-1)\right]\right\} \mathrm{d}\mathbf{r}, \quad (1)$$

where $\phi$ is the electrostatic potential defining on the entire domain $\Omega$, $\mathrm{k_B}$ is the Boltzmann constant, $T$ is the temperature, $c_j$ is the bulk concentration of the $j$th ionic species, $q_j$ is the charge of the $j$th ionic species, $N_c$ is the number of ionic species, and $\rho_{\mathrm{m}} = \sum_j Q_j\delta(\mathbf{r}-\mathbf{x}_j)$ is the canonical density of molecular free charges, with $Q_j$ being the partial charge on an atom located at $\mathbf{x}_j$. Here, $\epsilon_{\mathrm{m}}$ and $\epsilon_{\mathrm{s}}$ are the electric permittivities of the macromolecule and the solvent, respectively.

The first term in the integrand of Equation (1) is the electrostatic free energy of the solute, whereas the second term is the electrostatic free energy of the solvent, in which the charge density of mobile ions is assumed to follow the Boltzmann distribution. The quadratic terms of the potential gradient take negative signs according to the convention of the field [28, 29]. Consequently, the polar free energy will be optimized instead of being minimized.

The nonpolar solvation free energy functional proposed by Wagoner and Baker [30] is considered for the nonpolar contribution,

$$G_{\mathrm{np}} = \gamma(\mathrm{Area}) + p(\mathrm{Vol}) + \rho_0\int_{\Omega_{\mathrm{s}}} U^{\mathrm{att}}\mathrm{d}\mathbf{r}, \quad (2)$$

where $\gamma$ is the surface tension, Area is the surface area of the macromolecule, $p$ is the hydrodynamic pressure, Vol is the volume occupied by the molecule of interest, $\rho_0$ is the solvent bulk density, and $U^{\mathrm{att}}$ is the attractive portion of the van der Waals potential at point $\mathbf{r}$. The first term is the surface energy, which measures the disruption of intermolecular and/or intramolecular bonds that occurs when a surface is created. The second term is due to the mechanical work of creating the vacuum of a biomolecular size in the solvent. The third term represents the attractive dispersion effects near the solute–solvent interface.

The nonpolar free energy functional (2) was originally constructed for a sharp interface of solute–solvent [30]. For the present smooth interface of solute–solvent, the concepts of the surface area and volume cannot be defined in the same manner as in the sharp surface case. As suggested by Wei [25], the differential geometry measure theory should be employed to convert these concepts into the Eulerian formulation so that all energy contributions can be treated in an equal footing. Essentially, there are infinitely many surfaces for a smooth boundary case, a geometry property can be defined as a weighted mean of this family of surfaces. Consequently, the nonpolar functional $G_{\mathrm{np}}$ will be in terms of the characteristic functions $S(\mathbf{r})$ and $(1-S(\mathbf{r}))$. In particular, the volume of biomolecule will be rewritten as

$$\mathrm{Vol} = \int_{\Omega_{\mathrm{m}}} \mathrm{d}\mathbf{r} = \int_{\Omega} S(\mathbf{r})\mathrm{d}\mathbf{r}. \quad (3)$$

Similarly the van der Waals dispersion term is given by

$$\rho_0\int_{\Omega_{\mathrm{s}}} U^{\mathrm{att}}\mathrm{d}\mathbf{r} = \rho_0\int_{\Omega}(1-S(\mathbf{r}))U^{\mathrm{att}}\mathrm{d}\mathbf{r}. \quad (4)$$

By using the coarea formula [25], the mean surface area can be given by a volume integral as

$$\mathrm{Area} = \int_{\Omega} \|\nabla S(\mathbf{r})\|\mathrm{d}\mathbf{r}. \quad (5)$$

We note that $\nabla S(\mathbf{r}) \neq 0$ only in the transition region of the solute–solvent boundary.



DOI: 10.1002/cnm

The total free energy functional of solvation for biomolecules at equilibrium is given by the sum of the electrostatic components and nonpolar components

$$G_{\text{total}} = \int_{\Omega} \Big\{ \gamma \|\nabla S(\mathbf{r})\| + pS(\mathbf{r}) + \rho_0(1-S(\mathbf{r}))U^{\text{att}} + S(\mathbf{r}) \Big[\rho_{\text{m}}\phi - \frac{\epsilon_{\text{m}}}{2}|\nabla\phi|^2\Big] + (1-S(\mathbf{r}))\Big[-\frac{\epsilon_{\text{s}}}{2}|\nabla\phi|^2 - \text{k}_{\text{B}}T\sum_{j=1}^{N_c} c_j\,(\text{e}^{-q_j\phi/\text{k}_{\text{B}}T}-1)\Big]\Big\}\,\text{d}\mathbf{r}. \tag{6}$$

The proposed pseudo-time-coupled nonlinear model will be derived based on this total free energy functional.

### 2.3. Variational analysis

The total free energy functional (6) can be regarded as a functional of two independent functions: the characteristic function $S$ and electrostatic potential $\phi$. Thus, the total free energy can be minimized via an optimal function $S(\mathbf{r})$ and an optimal potential $\phi$ at the equilibrium state. Mathematically, such an optimal state can be achieved via the variational analysis

$$\frac{\delta G_{\text{total}}}{\delta S} = 0 \quad \text{and} \quad \frac{\delta G_{\text{total}}}{\delta \phi} = 0. \tag{7}$$

This leads the following two nonlinear PDEs [25, 26]

$$S\rho_{\text{m}} + \nabla\cdot([(1-S)\epsilon_{\text{s}} + S\epsilon_{\text{m}}]\nabla\phi) + (1-S)\sum_{j=1}^{N_c} c_j q_j \text{e}^{-q_j\phi/\text{k}_{\text{B}}T} = 0 \tag{8}$$

$$-\nabla\cdot\left(\gamma\frac{\nabla S}{\|\nabla S\|}\right) + p - \rho_0 U^{\text{att}} + \rho_{\text{m}}\phi - \frac{\epsilon_{\text{m}}}{2}|\nabla\phi|^2 + \frac{\epsilon_{\text{s}}}{2}|\nabla\phi|^2 + \text{k}_{\text{B}}T\sum_{j=1}^{N_c} c_j\left(\text{e}^{-q_j\phi/\text{k}_{\text{B}}T}-1\right) = 0 \tag{9}$$

where $\nabla\cdot\left(\gamma\frac{\nabla S}{\|\nabla S\|}\right)$ is a generalized Laplace–Beltrami operator, which is a generalization of the usual Laplacian operator to a smooth manifold. The parameter $\gamma$ can be a function of the position $\gamma = \gamma(\mathbf{r})$ to reflect surface hydrophobicity at different locations. For simplicity, it will be treated as a constant in the present study.

By defining an overall dielectric function to be

$$\epsilon(S) = (1-S)\epsilon_{\text{s}} + S\epsilon_{\text{m}}, \tag{10}$$

Equation (8) is actually a generalized PB equation [25, 26] of the form

$$-\nabla\cdot(\epsilon(S)\nabla\phi) = S\rho_{\text{m}} + (1-S)\sum_{j=1}^{N_c} c_j q_j \text{e}^{-q_j\phi/\text{k}_B T}. \tag{11}$$

One dielectric profile of the dielectric function $\epsilon(S)$ is shown in Figure 1(b). A smooth transition of dielectric values from $\epsilon_{\text{s}}$ to $\epsilon_{\text{m}}$ is clearly seen. From the numerical point of view, such a smooth dielectric profile prevents many computational difficulties in solving elliptic equations with discontinuous coefficients [31, 32]. For a given characteristic function $S(\mathbf{r})$, the electrostatic potential $\phi$ can be solved from Equation (11) via some standard spatial discretization methods, such as the central difference.

As in our previous MMS model [22, 23] and PDGF model [24], the solution of the nonlinear elliptic PDE will be attained via the following nonlinear parabolic PDE [25, 26]:

$$\frac{\partial S}{\partial t} = \sqrt{\|\nabla S\|}\Big[\nabla\cdot\Big(\gamma\frac{\nabla S}{\|\nabla S\|}\Big) + V\Big], \tag{12}$$



DOI: 10.1002/cnm

where $V$ can be regarded as a generalized potential with the form

$$V = -p + \rho_0 U^{\text{att}} - \rho_{\text{m}}\phi + \frac{\epsilon_{\text{m}}}{2}|\nabla\phi|^2 - \frac{\epsilon_{\text{s}}}{2}|\nabla\phi|^2 - \text{k}_{\text{B}}T\sum_{j=1}^{N_c} c_j\,(\text{e}^{-q_j\phi/\text{k}_{\text{B}}T} - 1) \tag{13}$$

Obviously, Equation (12) is a special case of our previous PDGF equation [24]. Integrating with respect to the pseudo-time $t$, the initial profile of $S$ evolutes in a steady state solution, which satisfies the original Equation (9). This final characteristic function $S(\mathbf{r})$ determines the physical solute–solvent boundary.

The formulation of sharp interface of solute–solvent can be obtained via the present variational formulation under certain limit. In particular, by taking $S$ to be a Heaviside function with value one for the solute subdomain and zero for the solvent subdomain, the smooth transition region disappears. Consequently, the dielectric function becomes discontinuous. Then, Equation (11) reduces to the classical PB equation with appropriate regularity conditions at the sharp interface [25, 27].

### *2.4. Pseudo-time-coupled governing equations*

The generalized PB equation (11) and the PDGF equation (12) constructed in [25, 26] are strongly coupled. The Lagrangian formulation of the two counterpart PDEs have been similarly generated in [27]. We explore a more efficient coupling of electrostatic potential $\phi$ and biomolecular surface evolution of $S(\mathbf{r})$ in the present work. In the original works [26, 27], an iterative procedure, in which two PDEs are solved alternatively, was employed to couple two processes. To avoid instability problems, a relaxation algorithm was used [26, 27]. For example, at one step with a characteristic function $S_{\text{old}}$, one solves Equation (12) to attain a new function $S_{\text{new}}$. The characteristic function $S$ to be used in the next PB solver is a weighted average

$$S = \alpha S_{\text{old}} + (1-\alpha)S_{\text{new}}, \tag{14}$$

which introduces one coupling parameter $\alpha$. The potential $\phi$ is treated in the same manner with another coupling parameter $\alpha'$, although normally $\alpha' = \alpha$. Moreover, the solution of the generalized PB equation (11) requires the solution of a large linear algebra system, which implies a longer CPU time. Thus, another coupling parameter $N_{\text{step}}$ was introduced in [26, 27] to enhance the efficiency. Basically, one alternatively solves the PB equation (11) once and the geometric flow equation (12) $N_{\text{step}}$ times. For large biomolecular systems, $N_{\text{step}}$ could be as small as $N_{\text{step}} = 2$ [26, 27]. Finally, the existing computations [26, 27] are mainly based on the linearized PB equations, the numerical coupling of nonlinear PB equations has not been explored.

In the present study, we propose to solve the nonlinear PB equation (11) through a time-dependent process. This actually brings the solution of PB equation (11) into the same footing as the solution of geometric flow equation (12). In particular, we consider the following nonlinear parabolic PDE:

$$\frac{\partial\phi}{\partial t} = \nabla\cdot(\epsilon(S)\nabla\phi) + S\rho_{\text{m}} + (1-S)\sum_{j=1}^{N_c} c_j q_j \text{e}^{-q_j\phi/\text{k}_{\text{B}}T}. \tag{15}$$

Here, the time $t$ is treated as the same variable $t$ in Equation (12). The coupling of two time-dependent processes (15) and (12) can be achieved as in the commonly encountered unsteady systems. Neither special coupling procedure nor relaxation algorithm is required. Two processes will be coupled via standard time integration techniques with a time increment $\Delta t$. The standard results in numerical analysis can be used to ensure the stability of the coupling. However, due to the different intrinsic time scale, different $\Delta t$ should be used for two processes.

Assume that all mobile ions are univalent in the solvent. We can treat them as positive and negative ions with charge $+e_{\text{c}}$ and $-e_{\text{c}}$, where $e_{\text{c}}$ is the charge of an electron. We further assume that away from the solute, the bulk concentrations of both positive and negative ions are the same.



DOI: 10.1002/cnm

Then, by applying the Boltzmann distribution law to both ions, we have the following nonlinear time-dependent PB equation

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (\epsilon(S)\nabla\phi) + S\rho_{\mathrm{m}} - (1-S)\bar{\kappa}^2\Big(\frac{\mathrm{k_B}T}{e_{\mathrm{c}}}\Big)\sinh\Big(\frac{e_{\mathrm{c}}\phi}{\mathrm{k_B}T}\Big). \tag{16}$$

Here, $\bar{\kappa}$ is the modified Debye–Huckel parameter and is defined as

$$\bar{\kappa}^2 = \left(\frac{2\,\mathrm{N_A}e_{\mathrm{c}}^2}{1000\,\mathrm{k_B}T}\right) I_{\mathrm{s}}, \tag{17}$$

where $\mathrm{N_A}$ is the Avogadro's number and $I_{\mathrm{s}}$ is the ion strength in the unit of mole. Numerically, when $T = 298$ K, we have $\bar{\kappa}^2 = 67.5365(\mathrm{nm})^{-2} I_{\mathrm{s}}$. Under the same assumptions, the Boltzmann distribution term in the Equation (13) should be similarly modified. This gives rise to a new generalized potential $V$

$$V = -p + \rho_0 U^{\mathrm{att}} - \rho_{\mathrm{m}}\phi + \frac{\epsilon_{\mathrm{m}}}{2}|\nabla\phi|^2 - \frac{\epsilon_{\mathrm{s}}}{2}|\nabla\phi|^2 - \bar{\kappa}^2\left(\frac{\mathrm{k_B}T}{e_{\mathrm{c}}}\right)^2\left[\cosh\left(\frac{e_{\mathrm{c}}\phi}{\mathrm{k_B}T}\right) - 1\right]. \tag{18}$$

to drive the geometric flow equation (12).

If the electrostatic potential is very weak, that is, $e_{\mathrm{c}}\phi \ll \mathrm{k_B}T$, the nonlinear PB equation can be approximated by the following time-dependent linearized PB equation

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (\epsilon(S)\nabla\phi) + S\rho_{\mathrm{m}} - (1-S)\bar{\kappa}^2\phi. \tag{19}$$

Meanwhile, to avoid a vanishing contribution, we take the first two terms in the Taylor expansion of $\cosh(\cdot)$ so that the generalized potential $V$ becomes

$$V = -p + \rho_0 U^{\mathrm{att}} - \rho_{\mathrm{m}}\phi + \frac{\epsilon_{\mathrm{m}}}{2}|\nabla\phi|^2 - \frac{\epsilon_{\mathrm{s}}}{2}|\nabla\phi|^2 - \frac{1}{2}\bar{\kappa}^2\phi^2. \tag{20}$$

In the generalized potential $V$, the attractive interaction with respect to the $i$th partial charge at location $\mathbf{r}_i$, that is, $U_i^{\mathrm{att}}$, is modeled based on the 12-6 Lennard-Jones potential [26, 27]

$$U_i^{\mathrm{LJ}}(\mathbf{r}) = \beta_i\left[\Big(\frac{\sigma_i + \sigma_s}{\|\mathbf{r} - \mathbf{r}_i\|}\Big)^{12} - 2\Big(\frac{\sigma_i + \sigma_s}{\|\mathbf{r} - \mathbf{r}_i\|}\Big)^{6}\right], \tag{21}$$

where $\beta_i$ is the well-depth parameter, and $\sigma_i$ and $\sigma_s$ are solute atomic and solvent radii, respectively. The attractive portion $U^{\mathrm{att}}$ and the repulsive portion $U^{\mathrm{rep}}$ can then be obtained via a decomposition of the Lennard-Jones potential. Two types of decompositions are commonly used. The first one is the so-called '6-12' decomposition

$$U_i^{\mathrm{att,6/12}}(\mathbf{r}) = -2\beta_i\Big(\frac{\sigma_i + \sigma_s}{\|\mathbf{r} - \mathbf{r}_i\|}\Big)^{6}, \quad U_i^{\mathrm{rep,6/12}}(\mathbf{r}) = \beta_i\Big(\frac{\sigma_i + \sigma_s}{\|\mathbf{r} - \mathbf{r}_i\|}\Big)^{12}. \tag{22}$$

Based on the Weeks–Chandler–Anderson (WCA) theory [33], the WCA decomposition is the other one

$$U_i^{\mathrm{att,WCA}}(\mathbf{r}) = \begin{cases} -\beta_i(\mathbf{r}), & 0 < \|\mathbf{r} - \mathbf{r}_i\| < \sigma_i + \sigma_s \\ U_i^{\mathrm{LJ}}(\mathbf{r}), & \|\mathbf{r} - \mathbf{r}_i\| \geqslant \sigma_i + \sigma_s \end{cases} \tag{23}$$

$$U_i^{\mathrm{rep,WCA}}(\mathbf{r}) = \begin{cases} U_i^{\mathrm{LJ}}(\mathbf{r}) + \beta_i(\mathbf{r}), & 0 < \|\mathbf{r} - \mathbf{r}_i\| < \sigma_i + \sigma_s \\ 0, & \|\mathbf{r} - \mathbf{r}_i\| \geqslant \sigma_i + \sigma_s \end{cases} \tag{24}$$

Following Chen *et al.* [26], the WCA decomposition is employed in the present study.



DOI: 10.1002/cnm

### 2.5. Numerical discretization

We consider the spatial and temporal discretization of the proposed time-dependent PDEs. Standard numerical algorithms will be employed to solve the time-dependent nonlinear PB equation (16) and PDGF equation (12) with the potential $V$ given by Equation (18). The solution of the linearized model can be conducted in the same manner.

We first prescribe boundary conditions. The characteristic function $S$ becomes zero away from the biomolecule. Thus, a Dirichlet zero condition is naturally imposed for $S(\mathbf{r})$ on $\partial\Omega$—the boundary of the computational domain $\Omega$. For electrostatic potential $\phi$, the far-field condition holds at the infinity, that is, $\phi(\infty) = 0$. For a finite domain $\Omega$, the approximate analytical condition can be employed

$$\phi(\mathbf{r}) = \sum_{i=1}^{N_{\mathrm{m}}} \frac{Q_i}{\epsilon_{\mathrm{s}}|\mathbf{r}-\mathbf{r}_i|} \mathrm{e}^{-\bar{\kappa}|\mathbf{r}-\mathbf{r}_i|/\sqrt{\epsilon_{\mathrm{s}}}}. \tag{25}$$

This condition (25) is actually a linear superposition of Coulomb's law for a series of $N_{\mathrm{m}}$ partial charges $Q_i$ at positions $\mathbf{r}_i$ and is very accurate when $\partial\Omega$ is far away from the macromolecule subdomain.

The standard central difference scheme and explicit Euler's scheme are employed for spatial and temporal discretization of time PB equation (16). Consider a uniform mesh partition of the computational domain $\Omega$. Without the loss of generality, we assume the grid spacing $h$ in all $x$, $y$, and $z$ directions to be the same. A uniform grid is used for the time advancing too with the time increment being $\Delta t$. By denoting $\phi^n_{i,j,k} = \phi(x_i, y_j, z_k, t_n)$ and $S^n_{i,j,k} = S(x_i, y_j, z_k, t_n)$, Equation (16) is discretized to be

$$\begin{aligned}\phi^{n+1}_{i,j,k} = \phi^n_{i,j,k} &- \Delta t(1 - S^n_{i,j,k})\bar{\kappa}^2\Big(\frac{\mathrm{k_B}T}{e_{\mathrm{c}}}\Big)\sinh\Big(\frac{e_{\mathrm{c}}\phi^n_{i,j,k}}{\mathrm{k_B}T}\Big) + \frac{\Delta t}{h}S^n_{i,j,k}Q(x_i, y_j, z_k) \\ &+ \frac{\Delta t}{h^2}\Big(\epsilon(x_{i+\frac{1}{2}}, y_j, z_k)(\phi^n_{i+1,j,k} - \phi^n_{i,j,k}) + \epsilon(x_{i-\frac{1}{2}}, y_j, z_k)(\phi^n_{i-1,j,k} - \phi^n_{i,j,k}) \\ &+ \epsilon(x_i, y_{j+\frac{1}{2}}, z_k)(\phi^n_{i,j+1,k} - \phi^n_{i,j,k}) + \epsilon(x_i, y_{j-\frac{1}{2}}, z_k)(\phi^n_{i,j-1,k} - \phi^n_{i,j,k}) \\ &+ \epsilon(x_i, y_j, z_{k+\frac{1}{2}})(\phi^n_{i,j,k+1} - \phi^n_{i,j,k}) + \epsilon(x_i, y_j, z_{k-\frac{1}{2}})(\phi^n_{i,j,k-1} - \phi^n_{i,j,k})\Big)\end{aligned} \tag{26}$$

where $Q(x_i, y_j, z_k)$ is the fractional charge at grid point $(x_i, y_j, z_k)$, which is obtained by using the trilinear interpolation to distribute all charges in the charge density $\rho_{\mathrm{m}}$. It is noted that the nonlinear $\sinh(\cdot)$ term imposes no additional difficulty numerically because it is based on the known $\phi$ values at time $t_n$.

In the present study, the same spatial–temporal discretization is employed for the PDGF equation (12) too. The parameters involved in Equation (12) are chosen according to the literature [26, 27]. In particular, the surface tension $\gamma$ will be chosen as a fitting parameter taking a constant value throughout the domain. Thus, Equation (12) becomes

$$\frac{\partial S}{\partial t} = \gamma\sqrt{\|\nabla S\|}\left[\nabla\cdot\left(\frac{\nabla S}{\|\nabla S\|}\right) + \frac{V}{\gamma}\right], \tag{27}$$

Moreover, by using the Cartesian components, Equation (12) can be rewritten in a full detailed form

$$\begin{aligned}\frac{\partial S}{\partial t} =& \gamma\frac{(S_x^2 + S_y^2)S_{zz} + (S_x^2 + S_z^2)S_{yy} + (S_y^2 + S_z^2)S_{xx}}{S_x^2 + S_y^2 + S_z^2} \\ &- \gamma\frac{2S_xS_yS_{xy} + 2S_xS_zS_{xz} + 2S_zS_yS_{yz}}{S_x^2 + S_y^2 + S_z^2} + \sqrt{S_x^2 + S_y^2 + S_z^2}V,\end{aligned} \tag{28}$$

where the subscripts of $S$ denote partial derivatives, for example, $S_x = \partial S/\partial x$. For simplicity, we will not present the discretization details of Equation (28), which are almost the same as our previous results given in [22–24]. Symbolically, we have the Euler's scheme of the form

$$S^{n+1} = S^n + \Delta t F(S^n, V^n), \tag{29}$$



DOI: 10.1002/cnm

where $S^n$ and $V^n$ are the collection of function values at time $t_n$ and $F$ is the finite difference discrete operator approximating the right hand side of Equation (28).

The coupling of PB equation and geometric flow equation is thus simply achieved via simultaneous time advancing in the present study. It is found that the whole system is a stiff system mathematically in the sense that one time process requires a much smaller $\Delta t$ than the other. Thus, in order to save the computational work, two different $\Delta t$ values are employed. In particular, the $\Delta t$ value for the PB equation is about 140 times smaller than that for the geometric flow equation.

### 2.6. Computational procedure

The computation of the proposed time-dependent nonlinear solvation models is carried out in three steps. At the first step, the initial profile of the characteristic function $S(\mathbf{r})$ is generated. Consider a macromolecule with total $N_a$ number of atoms. Denote the center and radius of the $i$th atom to be $\mathbf{r}_i = (x_i, y_i, z_i)$ and $r_i$, respectively, for $i = 1, 2, \ldots, N_a$. We then define the domain enclosed by the solvent accessible surface to be $D = \bigcup_{i=1}^{N_a} \{\mathbf{r} : |\mathbf{r} - \mathbf{r}_i| < r_i + r_p\}$, where $r_p$ is the probe radius. At $t = 0$, we first define $S$ to be

$$S(x, y, z, 0) = \begin{cases} 1, & (x, y, z) \in D \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

Then, the hypersurface function $S$ is evolved according to the geometric flow equation (28) or its numerical realization (29) in the absence of the driving potential $V$ to a stopping time $T_S$. Typically, we choose $T_S = 3.5$ to generate the initial profile of $S$. Also, in this first step and subsequent time integration, we only update the values of $S(x, y, z, t)$ at the points in between the van der Waals surface and solvent accessible surface, i.e., $(x, y, z) \in \bigcup_{i=1}^{N_a} \{\mathbf{r} : r_i < |\mathbf{r} - \mathbf{r}_i| < (r_i + r_p)\}$. This is for the purpose of protecting the van der Waals surface and making the computation more efficient. Numerically, to avoid a vanishing value in the denominator of Equation (28), we add a very small number, say $10^{-7}$, into the denominator, which does not affect the accuracy of our computations.

At the second step, we generate the initial profile of the electrostatic function $\phi(\mathbf{r})$. Based on the hypersurface function $S(\mathbf{r})$ constructed in the first step, we solve the following linearized PB equation once

$$-\nabla \cdot (\epsilon(S)\nabla\phi) + (1 - S)\bar{\kappa}^2\phi = S\rho_m. \tag{31}$$

Based on a standard finite difference discretization, the standard biconjugate gradient method with the incomplete LU factorization preconditioner is used to solve the resulting linear algebraic system.

At the final step, the simultaneous time coupling of two processes are considered. To achieve a steady state coupling, a long enough coupling time $T_c = 10$ is employed. An early convergence criterion is also allowed; that is, the computation stops if the correction in the free energy of solvation is less than $10^{-6}$. For the PDGF equation (12), a large $\Delta t_S = h^2/4.5$ is enough to ensure the stability. Nevertheless, a very small $\Delta t_\phi = h^2/630$ is used for the time-dependent nonlinear PB equation (16) for a stable result. In other words, by advancing $S$ for one time step, we will update $\phi$ for 140 time steps.

### 2.7. Filtering process

Finally, we propose a filtering process to stabilize the time integration of the proposed nonlinear equations. It is found in the present study that the nonlinear terms of the electrostatic potential $\phi$, that is, $\sinh\left(\frac{e_c\phi}{k_B T}\right)$ in Equation (16) and $\cosh\left(\frac{e_c\phi}{k_B T}\right)$ in Equation (18), could become very large so that the time stepping becomes unstable. Let us investigate this subtle issue by considering the nonlinear PB equation. The counterpart of the nonlinear term $(1 - S)\bar{\kappa}^2\left(\frac{k_B T}{e_c}\right)\sinh\left(\frac{e_c\phi}{k_B T}\right)$ appears in the classical nonlinear PB equation too. Nevertheless, such a term is assumed to be separated from the molecule or solute domain by an ion-exclusion layer [34]. Moreover, a sharp interface between the ion-exclusion layer and the solvent domain is employed in the classical model. Thus, the nonlinear term is vanishing within the solute domain in the classical nonlinear PB equation. On the other hand, for the present smooth interface of solute–solvent, the switch off is controlled via



DOI: 10.1002/cnm

a characteristic function $(1-S)$, which, unfortunately, is not completely zero at the outer region of the macromolecule subdomain. Thus, there could exist certain places which are very close to partial charges, but in which $(1-S)$ values are not zero. Due to the Coulomb's law, the electrostatic potential $\phi$ is quite large in such regions. This is exponentially amplified by the sinh(·) function then. Then, unphysical blowing-up solutions could be generated in the smooth biomolecular surface model when considering nonlinear PB equations. Such a problem becomes less evident for the linearized PB equation.

In Figure 2, the nonzero values of the nonlinear term $g(S,\phi) := (1-S)\bar{\kappa}^2\left(\frac{k_B T}{e_c}\right)\sinh\left(\frac{e_c\phi}{k_B T}\right)$ are plotted throughout the domain $\Omega$ at one time instant. It can be seen that majority values are of small magnitude, except for a very small portion of values that are as high as $10^8$ in magnitude. We propose to employ a post-processing filtering to rectify this theoretical pitfall. A 'soft' filtering with a threshold value $\delta$ can be employed

$$g(S,\phi) = \delta, \quad \text{if } g(S,\phi) > \delta. \tag{32}$$

Alternatively, a 'hard' filtering may be utilized

$$g(S,\phi) = 0, \quad \text{if } g(S,\phi) > \delta. \tag{33}$$

In the present study, the soft thresholding with $\delta = 1$ is found to be able to overcome this difficulty in all tested biological systems. The same filtering is applied to the cosh(·) term in Equation (18).

## 3. NUMERICAL VALIDATION AND APPLICATION

In this section, we carry out numerical experiments to test the performance of the proposed nonlinear solvation models. The application of the proposed models to real biological systems will be considered too. In all studies, a cubic computational domain is chosen. The size of the computational domain is large enough so that the error introduced by the approximate boundary condition (25) is negligible. A uniform mesh with mesh size $h$ along all three dimensions is employed.

### *3.1. Model parameters and solvation free energy*

Following References [26] and [27], the Lennard-Jones parameters are chosen as follows: $\sigma_i$ is taken from atomic radius and $\sigma_s$ is chosen to be 0.065 nm as a good fitting solvent radius. Given $\sigma_s$ and $\sigma_i$, the value of $\beta_i$ can be solved from the equation $\beta_i\left[\left(\frac{\sigma_i+\sigma_s}{\|\mathbf{r}-\mathbf{r}_i\|}\right)^{12} - 2\left(\frac{\sigma_i+\sigma_s}{\|\mathbf{r}-\mathbf{r}_i\|}\right)^{6}\right] = D_i$, which

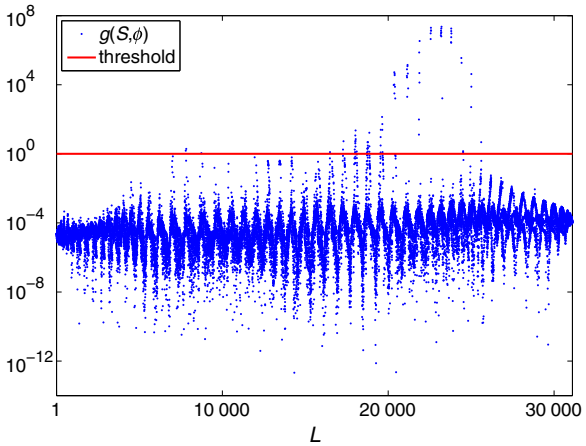


Figure 2. Nonzero values of $g(S,\phi) = (1-S)\bar{\kappa}^2\left(\frac{k_B T}{e_c}\right)\sinh\left(\frac{e_c\phi}{k_B T}\right)$ at time $T_c = 1$. Here, the electrostatic potential $\phi$ is generated based on the protein 1ajj using $h = 0.05$ nm. All grid nodes $(x_i, y_j, z_k)$ in the domain $\Omega$ have been re-arranged into one array with index $L = 1, 2, \ldots, N_g$, where $N_g$ is the total number of nonzero $g(S,\phi)$ values.



DOI: 10.1002/cnm

holds if $\mathbf{r}$ is on the van der Waals surface of the atom. Here, the constant $D_i$ should take different values for various types of atoms. For simplicity, we use a uniform value $D_i = 1$. The dielectric constants are taken as $\epsilon_m = 1$ and $\epsilon_s = 80$. The default value of the ionic strength in the present paper is set to be $I_s = 0.1$ The surface tension $\gamma$ is treated as a fitting parameter, with its initial value being $\gamma = 1/15$ to scale other parameters. We set the bulk density coefficient to be $\rho_s/\gamma = 2$ and choose pressure coefficient as $p/\gamma = 0.2$. The final values of $\gamma$ are different for various real systems [26, 27]. Finally, we note that the atomic radii are enlarged according to [26]. In particular, the radii from the CHARMM force field need to be multiplied by a common factor of value 1.1.

To test the proposed solvation model, we will benchmark our results by evaluating the solvation free energy. The total free energy functional of solvation does not directly provide the total solvation free energy. Actually, one needs to calculate the difference of the macromolecular system in the solvent and in the vacuum because the nonpolar energy describes the change in free energy. With the steady state solution $S(\mathbf{r})$ and $\phi(\mathbf{r})$, the solvation free energy can be computed as

$$\Delta G = G - G_0, \tag{34}$$

where $G_0$ is the polar free energy calculated from the homogeneous (vacuum) environment with $\epsilon_s = \epsilon_m = 1$ and is independent of nonpolar energy. In a more detailed form, we have

$$\Delta G = G_{np} + (G_p - G_0). \tag{35}$$

Here, $G_p - G_0$ can be regarded as the electrostatic solvation free energy. The expressions of $G_p$ and $G_{np}$ can be derived from Equation (6). In the present study, the polar part is evaluated as

$$G_p = \frac{1}{2}\int_\Omega S(\mathbf{r})\rho_m\phi(\mathbf{r})d\mathbf{r} = \frac{1}{2}\sum_{i=1}^{N_m} Q(\mathbf{r}_i)\phi(\mathbf{r}_i), \tag{36}$$

where $Q(\mathbf{r}_i)$ is the $i$th partial charge at location $\mathbf{r}_i$ in the biomolecule and $N_m$ is the total number of partial charges. Similarly, the electrostatic solvation free energy can be calculated as

$$\Delta G_p = G_p - G_0 = \frac{1}{2}\sum_{i=1}^{N_m} Q(\mathbf{r}_i)(\phi(\mathbf{r}_i) - \phi_0(\mathbf{r}_i)), \tag{37}$$

where $\phi$ and $\phi_0$ are electrostatic potentials in the presence of the solvent and the vacuum, respectively. The nonpolar part, $G_{np}$, is computed exactly according to Equation (2).

### *3.2. Numerical convergence*

We first numerically test the convergence of the proposed time coupling of two nonlinear PDEs. A protein system (PDB ID: 1ajj) is chosen for this purpose. To focus mainly on the coupling stage, we set a very small initiation time $T_S = 0.3$ in the first step. The initial value of $\phi(\mathbf{r})$ is then generated as usual in the second step. Now, in the third step, we consider a total of $N$ time steps for advancing the hypersurface function $S(\mathbf{r})$, i.e., $N = T_c/\Delta t_S$. Here, the above mentioned time increment values $\Delta t_S = h^2/4.5$ and $\Delta t_\phi = h^2/630$ are used. And we choose $h = 0.05$ nm and $N = 1000$. The time histories of the scaled solvation free energy, area, and volume of this protein are depicted in the Figure 3 (a) as the nonlinear cases. Because these three values are of difference scales, they are properly re-scaled in the Figure 3 (a) in order to attain a better view. In particular, they are scaled as the follows: scaled volume = volume/200 − 20 ($10^{-3}$nm$^3$), scaled area = area/100 − 10 ($10^{-2}$nm$^2$), and scaled energy = energy/50 + 25 (kcal/mol). It can be observed that the steady state solution is numerically achieved when $N$ is large. Moreover, it is obvious that the total solvation free energy decreases with respect to the time evolution, as predicted by the variational theory.

It is also interesting to compare the time convergence of the proposed linear and nonlinear PB models. In particular, we compare the results of time-dependent nonlinear PB equation (16) and time-dependent linearized PB equation (19). With the same numerical setting as in the previous case, we updated the linearized PB equation also until $N = 1000$. The time historical values of the linearized PB model are also shown in the Figure 3(a). The visual differences between scaled values



DOI: 10.1002/cnm

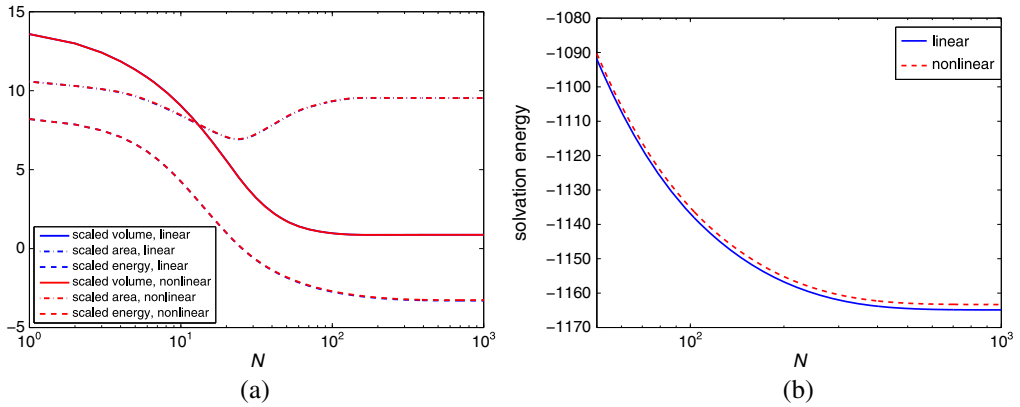


Figure 3. The time evolution histories of volume, area, and electrostatic solvation free energy for the protein 1ajj: (a) the scaled quantities for entire time history; (b) the un-scaled free energy in the later time.

of nonlinear and linear PB models are found to be very small. To see the difference, the original energy values of both nonlinear and linear PB models are also shown in Figure 3(b). It is seen that the difference gradually grows with respect to time, even though it remains to be small eventually. To gain a quantitative understanding, the starting and final un-scaled values of the solvation free energy, area, and volume obtained by both models are listed in Table I. It can be observed that starting from the almost identical initial values, the final values of nonlinear and linearized PB models are of very minor difference. This is because the electrostatic potential is rather weak for the present system so that the linearized model is a very good approximation to the nonlinear model. Also, the present numerical study demonstrates that the proposed time coupling procedure works equally well in both nonlinear and linearized cases. There is usually a moderate increment in terms of CPU time in comparing the nonlinear model with the linearized model. However, without the proposed time coupling, the nonlinear models could demand far more CPU time.

### *3.3. Solvation energies of small compounds*

We next consider some real biological applications of the proposed nonlinear models. The solvation free energies of a set of 17 small compounds are studied first. This test set has been studied by using many different methods in the literature [26, 27, 35]. This is an excellent benchmark set because experimental measurements are available. On the other hand, this test set is considered to be computationally challenging because of the existence of polyfunctional or interacting polar groups, which leads to strong solute–solvent interactions.

The structure and charge information of the 17 compounds are adopted from those of Nicholls *et al.* [35]. The OpenEye-AM1-BCC v1 parameters are used for the charges, whereas atomic coordinates and radii are based on a new parameterization introduced by Nicholls *et al.*, that is, the so-called ZAP-9 form. Basically, in ZAP-9 parameterization, certain types of radii are adjusted from Bonds radii so that a better agreement with experimental data is achieved. As mentioned earlier, the atomic radii are enlarged by a factor of 1.1 in the proposed model, and we choose $I_s = 0.1$. For the fitting parameter $\gamma$, an initial value of $\gamma = 1/15$ is employed as before, whereas its final value

Table I. Electrostatic solvation free energies (kcal/mol), areas ($10^{-2}$nm$^2$) and volumes ($10^{-3}$nm$^3$), calculated by linear and nonlinear PB models for the protein 1ajj.

| | Starting values | | | Final values | | |
|---|---|---|---|---|---|---|
| Model | Area | Volume | $\Delta G$ | Area | Volume | $\Delta G$ |
| Linear | 2557.22 | 7716.66 | −590.06 | 2452.99 | 5174.19 | −1164.89 |
| Nonlinear | 2557.22 | 7716.65 | −590.07 | 2453.40 | 5173.97 | −1163.33 |



DOI: 10.1002/cnm

is $\gamma = 0.0094$ in the present test. For the present small biological systems, we employ a dense mesh with $h = 0.025$ nm so that our computation could be more accurate. The numerical results are summarized in Table II. In comparison with the experiment data, the errors of the proposed solvation model are also reported. The root mean square (RMS) error of our time-coupled nonlinear model is found to be 1.83 kcal/mol, which is slightly larger than those of the differential geometry models [26,27]. Nevertheless, such RMS error is still considered to be good in comparison with the original results of Nicholls *et al.* [35]. In particular, by using the explicit solvent model, their RMS error is still as large as $1.71 \pm 0.05$ kcal/mol, whereas the smallest RMS error of their single-conformer PB approach is $1.87 \pm 0.03$ kcal/mol. Thus, our present result is competitive to the existing results in the literature.

### *3.4. Solvation energies of proteins*

We next validate our model by considering some larger biological systems. A set of 23 proteins that has been studied in previous works [26, 27, 36] is employed for this purpose. To save CPU time, a larger spacing $h = 0.05$ nm is employed. The protein structures are prepared as in the previous studies [26, 27, 36]. In particular, extra water molecules that are attached to proteins are excluded in all structures, and hydrogen atoms are added to obtain full all-atom models. Partial charges at atomic sites and atomic van der Waals radii in angstroms are taken from the CHARMM22 force field. In the present computation, all model parameters are chosen as the same as those in the previous subsection, except the final value of $\gamma = 1/15$ in the present case. In other words, for large protein systems, the surface tension $\gamma$ allows to be a uniform value. The solvation free energies of these 23 proteins are listed in Table III. For a comparison, the corresponding results by using a highly accurate sharp interface PB solver, that is, MIBPB-III [36], and the original Eulerian formulation linearized PB model (EFLPB), are also given in Table III. An excellent agreement between the present results and those of the literature is clearly seen.

For the present protein systems, it is interesting to illustrate the visualization capability of the proposed implicit solvent model because the electrostatic potential $\phi(\mathbf{r})$ is well defined on the entire domain. Because the biomolecular surfaces are represented via a smooth variating characteristic function $S(\mathbf{r})$, we can generate multiple isosurfaces $S(\mathbf{r}) = C$ by using different iso-values of $C$. Here, we take the protein 1uxc as an example. The isosurfaces with $S = 0.25$, $S = 0.5$, and $S = 0.75$ are shown in Figure 4. The calculated potential $\phi$ is then projected onto these isosurfaces. Essentially, these surface potential plots could reveal the fast/slow electrostatic potential changing region on the solute–solvent boundary. Such visual information may be used to detect charged active sites, which are helpful to the further studies on the protein–protein or protein–ligand interactions.

Table II. Electrostatic solvation free energies (kcal/mol) for 17 compounds.

| Compound | $G_{\text{np}}$ | $\Delta G_{\text{p}}$ | $\Delta G$ | Experimental | Error |
|---|---|---|---|---|---|
| Glycerol triacetate | 2.25 | −12.50 | −10.24 | −8.84 | −1.40 |
| Benzyl bromide | 1.39 | −4.91 | −3.52 | −2.38 | −1.14 |
| Benzyl chloride | 1.33 | −5.05 | −3.72 | −1.93 | −1.79 |
| *m*-Bis(trifluoromethyl)benzene | 1.37 | −3.26 | −1.89 | 1.07 | −2.96 |
| *N*,*N*-Dimethyl-*p*-methoxybenzamide | 1.93 | −9.22 | −7.29 | −11.01 | 3.72 |
| *N*,*N*-4-Trimethylbenzamide | 1.82 | −7.70 | −5.88 | −9.76 | 3.88 |
| Bis-2-chloroethyl ether | 1.51 | −4.23 | −2.72 | −4.23 | 1.51 |
| 1,1-Diacetoxyethane | 1.67 | −8.26 | −6.59 | −4.97 | −1.62 |
| 1,1-Diethoxyethane | 1.55 | −4.46 | −2.90 | −3.28 | 0.38 |
| 1,4-Dioxane | 1.03 | −5.65 | −4.62 | −5.05 | 0.43 |
| Diethyl propanedioate | 1.83 | −7.87 | −6.04 | −6.00 | −0.04 |
| Dimethoxymethane | 1.08 | −4.52 | −3.44 | −2.93 | −0.51 |
| Ethylene glycol diacetate | 1.59 | −8.47 | −6.88 | −6.34 | −0.54 |
| 1,2-Diethoxyethane | 1.59 | −4.32 | −2.73 | −3.54 | 0.81 |
| Diethyl sulfide | 1.20 | −2.39 | −1.18 | −1.43 | 0.25 |
| Phenyl formate | 1.34 | −7.86 | −6.51 | −4.08 | −2.43 |
| Imidazole | 0.86 | −11.28 | −10.42 | −9.81 | −0.61 |



DOI: 10.1002/cnm

Table III. Electrostatic solvation free energies (kcal/mol) for 23 proteins.

| PDB ID | No. of Atoms | MIBPB-III [36] | EFLPB [26] | Present |
|---|---|---|---|---|
| 1ajj | 519 | −1137.2 | −1178.5 | −1161.4 |
| 1bbl | 576 | −986.8 | −965.9 | −967.0 |
| 1bor | 832 | −853.7 | −871.4 | −862.5 |
| 1bpi | 898 | −1301.9 | −1281.2 | −1279.3 |
| 1cbn | 648 | −303.8 | −304.8 | −251.0 |
| 1fca | 729 | −1200.1 | −1200.6 | −1185.9 |
| 1frd | 1478 | −2852.2 | −2844.8 | −2833.2 |
| 1fxd | 824 | −3299.8 | −3291.9 | −3317.9 |
| 1hpt | 858 | −811.6 | −808.2 | −781.1 |
| 1mbg | 903 | −1346.1 | −1328.2 | −1313.2 |
| 1neq | 1187 | −1730.1 | −1713.9 | −1699.2 |
| 1ptq | 795 | −873.1 | −866.2 | −853.0 |
| 1r69 | 997 | −1089.5 | −1072.7 | −1040.5 |
| 1sh1 | 702 | −753.3 | −771.8 | −791.5 |
| 1svr | 1435 | −1711.2 | −1704.6 | −1717.8 |
| 1uxc | 809 | −1138.7 | −1125.7 | −1142.5 |
| 1vii | 596 | −901.5 | −892.0 | −887.7 |
| 2erl | 573 | −948.8 | −935.8 | −913.9 |
| 2pde | 667 | −820.9 | −843.0 | −841.4 |
| 451c | 1216 | −1024.6 | −1020.6 | −964.1 |
| 1a2s | 1272 | −1913.5 | −1900.3 | −1825.6 |
| 1a63 | 2065 | −2373.5 | −2380.5 | −2383.1 |
| 1a7m | 2809 | −2155.5 | −2179.8 | −2074.4 |

EFLPB, Eulerian formulation linearized Poisson–Boltzmann model; PDB, Protein Date Bank.

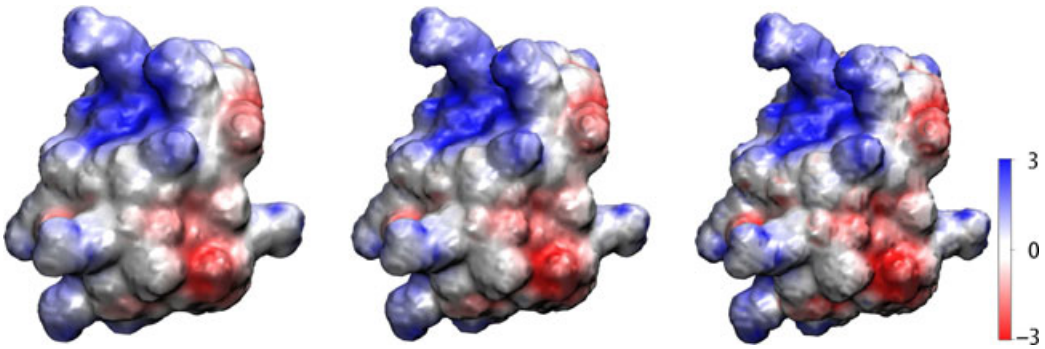


Figure 4. Plots of the surface potential of the protein 1uxc based on different isosurfaces. (a) $S = 0.25$; (b) $S = 0.5$; (c) $S = 0.75$.

### *3.5. Salt effect on protein–protein binding energies*

We finally investigate the performance of the proposed time-coupling nonlinear PB model for the evaluation of the salt effect in the protein–protein binding. The nonlinear PB equation is often used to describe the salt effects on the binding of ligands, peptides, and proteins to nucleic acids, membranes, and proteins. Physically, the binding affinity can be quantitatively represented based on the binding-free energies, which reflect the non-specific salt dependence of the formation of macromolecular complexes. The binding affinity is then calculated as the slope ratio of the salt-dependent binding energy at certain salt strength $I$ against the natural logarithm of $I$.

Mathematically, the electrostatic component of the binding energy ($\Delta G_{\mathrm{p}}$) measures the difference in the electrostatic free energy between the complex and its monomers

$$\Delta G_{\mathrm{p}}(I) = G_{\mathrm{p}}^{\mathrm{AB}}(I) - G_{\mathrm{p}}^{\mathrm{A}}(I) - G_{\mathrm{p}}^{\mathrm{B}}(I), \tag{38}$$



DOI: 10.1002/cnm

where $G_p^{AB}(I)$, $G_p^{A}(I)$ and $G_p^{B}(I)$ are the electrostatic free energies of the complex AB, and the monomers A and B, respectively, at a given ionic strength $I$. The electrostatic binding-free energy can be further split into salt independent and dependent parts. The variation of the salt dependent part of the binding-free energy $\Delta\Delta G_p(I)$ can thus be calculated as the difference in $\Delta G_p$ based on some salt strength $I$ and at the zero salt concentration, because the salt independent parts are simply canceled

$$\begin{aligned}\Delta\Delta G_p(I) &= \Delta G_p(I) - \Delta G_p(I=0)\\ &= [G_p^{AB}(I) - G_p^{AB}(I=0)] - [G_p^{A}(I) - G_p^{A}(I=0)] - [G_p^{B}(I) - G_p^{B}(I=0)]. \end{aligned} \quad (39)$$

In the present study, we evaluate the binding-free energy in various ionic concentrations for two protein complexes with different biological features (see Table IV), that is, 1beb and 1emv. The model parameters used for the earlier 23 proteins are used again in the present test. The calculated binding-free energies are depicted in Figure 5 together with the experimental results. The slope ratio or the binding affinity is calculated and reported in Table IV. For a comparison, the results attained by the Lagrangian formulation linearized PB (LFLPB) model [27] are also given in Table IV. Qualitatively, the proposed model agrees with the experimental observations; that is, for the hetero-diemric complex, the binding-free energy increases when the ionic strength is larger, whereas for the homo-diemric complex it decreases instead. Quantitatively, the binding affinities estimated by the present model are not as good as those of the LFLPB model. This is probably because the calculation of the binding affinity requires a physical cutoff to obtain two monomers A and B. Such a cutoff is consistent with the sharp interface definition of the LFLPB model but is not naturally defined for the present smooth interface with transition. A better Eulerian formulation to accommodate such a situation deserves further study.

Table IV. Comparison of binding affinities of two protein complexes.

| Complex | PDB code | Complex charge | Charge of the free monomers | Experimental data | LFLPB [27] | Present |
|---|---|---|---|---|---|---|
| E9Dnase-Im9 (10) (B–A) | 1emv | −3 | $B = +5$; $A = -8$ | 2.17 | 2.40 | 3.66 |
| Lactoglobulin dimer (57) (A–B) | 1beb | +26 | $A = B = +13$ | −1.62 | −2.02 | −3.02 |

LFLPB, Lagrangian formulation linearized Poisson–Boltzmann; PDB, Protein Data Bank.

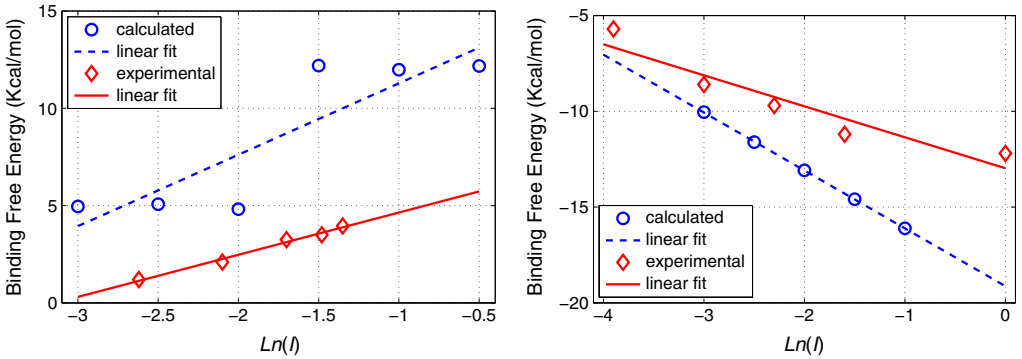


Figure 5. The salt dependence of the binding affinities of 1emv (left) and 1beb (right).



DOI: 10.1002/cnm

## 4. CONCLUSION

In summary, we have introduced an improved mathematical model for the theoretical modeling of biomolecular surface and solvation process. Built upon the previous differential geometry-based solvation models [25–27], mathematical issues on the coupling of nonlinear PDEs and the prevention of unphysical blowup have been explored. A time-dependent generalized PB equation is proposed for calculating the electrostatic potential. The coupling of the PB equation and the geometric flow equation representing the smooth solute–solvent interface can be simply achieved via standard time integration. Computationally, the treatment of the nonlinear terms poses no difficulty in the present model. Furthermore, a soft thresholding is introduced to confine these nonlinear terms. Numerical experiments are carried out for both simple molecules and complex proteins. The salt effect on protein–protein binding affinity is studied for selected protein complexes.

Because of the use of a very small time increment in solving time-dependent PB equations, the present approach is not very efficient. It has been found numerically that the proposed linearized PB model takes about five times more CPU time than the original EFLPB model [26]. On the other hand, the CPU time increment from the proposed linearized PB model to the nonlinear PB model is moderate—usually about 60–70%. However, without the proposed pseudo-time coupling, the solution of nonlinear boundary value problem when extending the EFLPB model to the nonlinear case could require far more CPU time. Issues of accelerating the pseudo-time coupling process and further application to the implicit solvent models are under our consideration.

ACKNOWLEDGEMENTS

The author thanks Zhan Chen and Guowei Wei for their useful discussions and insightful comments that improved the paper. This work was supported in part by the NSF grant DMS-1016579 and by a UA Research Grants Committee Award.

REFERENCES

1. Sharp KA, Honig B. Electrostatic interactions in macromolecules: theory and applications. *Annual Review Biophysics and Biophysical Chemistry* 1990; **19**:301–332.
2. Ponder JW, Case DA. Force fields for protein simulations. *Advances in Protein Chemistry* 2003; **66**:27–85.
3. Roux B, Simonson T. Implicit solvent models. *Biophysical Chemistry* 1999; **78**:1–20.
4. Baker NA. Improving implicit solvent simulations: a Poisson-centric view. *Current Opinion in Structural Biology* 2005; **15**:137–143.
5. Dong F, Olsen B, Baker NA. Computational methods for biomolecular electrostatics. *Methods in Cell Biology* 2008; **84**:843–870.
6. Feig M, Brooks III CL. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology* 2004; **14**:217–224.
7. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks III CL. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry* 2004; **25**:265–284.
8. Davis ME, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Computer Physics Communications* 1991; **62**:187–197.
9. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* 1973; **55**:379–400.
10. Connolly ML. Analytical molecular surface calculation. *Journal of Applied Crystallography* 1983; **16**:548–558.
11. Richards FM. Areas, volumes, packing and protein structure. *Annual Review of Biophysics and Bioengineering* 1977; **6**:151–176.
12. Dong F, Vijaykumar M, Zhou HZ. Electrostatic contribution to the binding stability of protein–protein complexes. *Proteins* 2006; **65**:87–102.
13. Dong F, Zhou HZ. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophysical Journal* 2003; **85**:49–60.
14. Grant JA, Pickup BT, Nicholls A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *Journal of Computational Chemistry* 2001; **22**:608–640.
15. Im W, Beglov D, Roux B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson–Boltzmann equation. *Computer Physics Communications* 1998; **111**:59–75.
16. Connolly ML. Molecular surface triangulation. *Journal of Applied Crystallography* 1985; **18**:499–505.



DOI: 10.1002/cnm

17. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 1996; **38**:305–320.
18. Swanson JMJ, Henchman RH, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysical Journal* 2004; **86**:67–74.
19. Wei GW, Sun YH, Zhou YC, Feig M. Molecular multiresolution surfaces, *arXiv:math–ph*, 0511001, Nov. 1 (2005).
20. Xu G, Pan Q, Bajaj C. Discrete surface modelling using partial differential equations. *Computer Aided Geometric Design* 2006; **23**:125–145.
21. Zhang Y, Xu G, Bajaj C. Quality meshing of implicit solvation models of biomolecular structures. *Computer Aided Geometric Design* 2006; **23**:510–530.
22. Bates PW, Wei GW, Zhao S. The minimal molecular surface, *arXiv:q-bio/0610038v1*, [q-bio.BM], 2006.
23. Bates PW, Wei GW, Zhao S. Minimal molecular surfaces and their applications. *Journal of Computational Chemistry* 2008; **29**:380–391.
24. Bates PW, Chen Z, Sun YH, Wei GW, Zhao S. Geometric and potential driving formation and evolution of biomolecular surfaces. *Journal of Mathematical Biology* 2009; **59**:193–231.
25. Wei GW. Differential geometry based multiscale models. *Bulletin of Mathematical Biology* 2010; **72**:1562–1622.
26. Chen Z, Baker N, Wei GW. Differential geometry based solvation model I: Eulerian formation. *Journal of Computational Physics* 2010; **229**:8231–8258.
27. Chen Z, Baker N, Wei GW. Differential geometry based solvation model II: Lagrangian formation, accepted. *Journal of Mathematical Biology* 2011. DOI: 10.1007/s00285-011-0402-z.
28. Sharp KA, Honig B. Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation. *Journal of Physical Chemistry* 1990; **94**:7684–7692.
29. Gilson MK, Davis ME, Luty BA, McCammon JA. Computation of electrostatic forces on solvated molecules using the Poisson–Boltzmann equation. *Journal of Physical Chemistry* 1993; **97**:3591–3600.
30. Wagoner J, Baker NA. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proceedings of the National Academy of Sciences of the USA* 2006; **103**:8331–8336.
31. Zhao S. High order matched interface and boundary methods for the Helmholtz equation in media with arbitrarily curved interfaces. *Journal of Computional Physics* 2010; **229**:3155–3170.
32. Zhou YC, Zhao S, Feig M, Wei GW. High order matched interface and boundary (MIB) schemes for elliptic equations with discontinuous coefficients and singular sources. *Journal of Computional Physics* 2006; **213**:1–30.
33. Weeks JD, Chandler D, Anderson HC. Role of repulsive forces in determining the equilibrium structure of simple liquids. *Journal of Chemical Physics* 1971; **54**:5237–5247.
34. Holst MJ. The Poisson–Boltzmann equation: analysis and multilevel numerical solution. *PhD thesis*, University of Illinois at Urbana-Champaign, 1993.
35. Nicholls A, Mobley DL, Guthrie PJ, Chodera JD, Pande VS. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *Journal of Medicinal Chemistry* 2008; **51**:769–779.
36. Geng W, Yu S, Wei GW. Treatment of charge singularities in implicit solvent models. *Journal of Chemical Physics* 2007; **127**:114 106.



DOI: 10.1002/cnm