Parameter optimization in differential geometry based solvation models

Bao Wang[1] and G. W. Wei[1,2,3,a)]
[1]*Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, USA*
[2]*Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA*

(Received 12 August 2015; accepted 22 September 2015; published online 7 October 2015)

Differential geometry (DG) based solvation models are a new class of variational implicit solvent approaches that are able to avoid unphysical solvent-solute boundary definitions and associated geometric singularities, and dynamically couple polar and non-polar interactions in a self-consistent framework. Our earlier study indicates that DG based non-polar solvation model outperforms other methods in non-polar solvation energy predictions. However, the DG based full solvation model has not shown its superiority in solvation analysis, due to its difficulty in parametrization, which must ensure the stability of the solution of strongly coupled nonlinear Laplace-Beltrami and Poisson-Boltzmann equations. In this work, we introduce new parameter learning algorithms based on perturbation and convex optimization theories to stabilize the numerical solution and thus achieve an optimal parametrization of the DG based solvation models. An interesting feature of the present DG based solvation model is that it provides accurate solvation free energy predictions for both polar and non-polar molecules in a unified formulation. Extensive numerical experiment demonstrates that the present DG based solvation model delivers some of the most accurate predictions of the solvation free energies for a large number of molecules. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4932342]

I. INTRODUCTION

Biological processes, such as signaling, gene regulation, transcription, and translation, govern the cell growth, cellular differentiation, fermentation, fertilization, germination, etc., in living organisms. Chemical processes, such as oxidation, reduction, hydrolysis, nitrification, polymerization, and so forth, underpin biological processes. Physical processes, particularly solvation, are involved in all the aforementioned chemical and biological processes. Therefore, a prerequisite for the understanding of chemical and biological processes is to study the solvation process. As a physical process, solvation does not involve the formation and/or breaking of any covalent bond but is associated with solvent and solute electrostatic, dipolar, induced dipolar, and van der Waals interactions.

Experimentally, solvation can be analyzed by the measurement of solvation free energies. Theoretically, solvation can be investigated by quantum mechanics, molecular mechanics, integral equation, implicit solvent models, and simple phenomenological modifications of Coulomb law. The implicit solvent models are known to balance the computational complexity and the accuracy in the solvation free energy prediction, and thus, offer an efficient approach.

The general idea of implicit solvent models is to treat the solvent as a dielectric continuum and describe the solute in atomistic detail.[23,41,43,59,62] The total solvation free energy is decomposed into non-polar and polar parts. There is a wide variety of ways to carry out this decomposition. For example, non-polar energy contributions can be modeled in two stages: the work of displacing solvent when adding a rigid solute to the solvent and the dispersive non-polar interactions between the solute atoms and surrounding solvent. The polar part is due to the electrostatic interactions and can be approximated by generalized Born (GB),[2,15,24,31,38,44,51,56,66,68,87] polarizable continuum (PC),[67] and Poisson-Boltzmann (PB) models.[1,23,29,46,62,84,86] Among them, GB models are heuristic approaches to polar solvation energy analysis. PC models resort to quantum mechanical calculations of induced solute charges. PB methods can be formally derived from Maxwell equations and statistical mechanics for electrolyte solutions[7,40,52] and therefore offer the promise of handling large biomolecules with sufficient accuracy and robustness.[2,22,55]

Conceptually, the separation between continuum solvent and the discrete (atomistic) solute introduces an interface definition. This definition may take the form of analytic functions[36–38] or nonsmooth boundaries dividing the solute-solvent domains. The van der Waals surface, solvent accessible surface,[47] and molecular surface (MS)[58] are devised for this purpose and have found their success in biophysical calculations.[8,20,27,42,45,48,49,63] It has been noticed that the performance of implicit solvent models is very sensitive to the interface definition.[25,26,54,64] This comes as no surprise because many of these popular interface definitions are *ad hoc* divisions of the solute and solvent domains based on rigid molecular geometry and neglecting solute-solvent energetic interactions. Additionally, geometric singularities[18,60] associated with these surface definitions incur enormous computational instability[77,78,86] and lead to conceptual difficulty in interpreting the sharp interface.[13]

a)Author to whom correspondence should be addressed. Electronic mail: wei@math.msu.edu

The differential geometry (DG) theory of surfaces[75] and associated geometric partial differential equations (PDEs) provide a natural description of the solvent-solute interface. In 2005, Wei and his collaborators introduced curvature-controlled PDEs for generating molecular surfaces in solvation analysis.[73] The first variational solvent-solute interface, namely, the minimal molecular surface (MMS), was constructed in 2006 by Wei and coworkers based on the DG theory of surfaces.[4–6] MMSs are constructed by solving the mean curvature flow, or the Laplace-Beltrami (LB) flow, and have been applied to the calculation of electrostatic potentials and solvation free energies.[6,16] This approach was generalized to potential-driven geometric flows, which admit physical interactions, for the surface generation of biomolecules in solution.[3] While our approaches were employed and/or modified by many others[17,76,80,81] for molecular surface and solvation analysis, our geometric PDE[73] and variational surface models[3,4,6] are, to our knowledge, the first of their kind for solvent-solute interface and solvation modeling.

Since the surface area minimization is equivalent to the minimization of surface free energies, due to a constant surface tension, this approach can be easily incorporated into the variational formulation of the PB theory[33,61] to result in DG-based full solvation models,[10,70] following a similar approach by Dzubiella *et al.*[28,83] Our DG-based solvation models have been implemented in the Eulerian formulation, where the solvent-solute interface is embedded in the three-dimensional (3D) Euclidean space and behaves like a smooth characteristic function.[10] The resulting interface and associated dielectric function vary smoothly from their values in the solute domain to those in the solvent domain and are computationally robust. An alternative implementation is the Lagrangian formulation[11] in which the solvent-solute boundary is extracted as a sharp surface at a given isovalue and subsequently used in the solvation analysis, including non-polar and polar modeling.

One major advantage of our DG based solvation model is that it enables the synergistic coupling between the solute and solvent domains via the variation procedure. As a result, our DG based solvation model is able to significantly reduce the number of free parameters that users must "fit" or adjust in applications to real-world systems.[65] It has been demonstrated that physical parameters, i.e., pressure and surface tension obtained from experimental data, can be directly employed in our DG-based solvation models for accurate solvation energy prediction.[21] Another advantage of our DG based solvation model is that it avoids the use of *ad hoc* surface definitions and its interfaces, particularly ones generated from the Eulerian formulation,[10] are free of troublesome geometric singularities that commonly occur in conventional solvent-accessible and solvent-excluded surfaces.[19,60] As a result, our DG based solvation model bypasses the sophisticated interface techniques required for solving the PB equation.[32,77,78] In particular, the smooth solvent-solute interface obtained from the Eulerian formulation[10] can be directly interpreted as the physical solvent-solute boundary profile. Additionally, the resulting smooth dielectric boundary can also have a straightforward physical interpretation. The other advantage of our DG based solvation model is that it is natural and easy to incorporate the density functional theory (DFT) in its variational formulation. Consequently, it is able to reevaluate and reassign the solute charge induced by solvent polarization effect during the solvation process.[13] The resulting total energy optimization process recreates or resembles the solvent-solute interactions, i.e., polarization, dispersion, and polar and non-polar coupling in a realistic solvation process. Recently, DG based solvation model has been extended to DG based multiscale models for non-equilibrium processes in biomolecular systems.[12,14,70,72,74] These models recover the DG based solvation model at the equilibrium.[74]

Recently, we have demonstrated[16] that the DG based non-polar solvation model is able to outperform many other methods[30,57,69] in solvation energy predictions for a large number non-polar molecules. The root mean square error (RMSE) of our predictions was below 0.4 kcal/mol, which clearly indicates the potential power of the DG based solvation formulation. However, the DG based full solvation model has not shown a similar superiority in accuracy, although it works very well.[10,11] Having so many aforementioned advantages, our DG based solvation models ought to outperform other methods with a similar level of approximations. One obstacle that hinders the performance of our DG based *full* solvation model is the numerical instability in solving two strongly coupled and highly nonlinear PDEs, namely, the generalized Laplace-Beltrami (GLB) equation and the generalized PB (GPB) equation. To avoid such instability, a strong parameter constraint was applied to the non-polar part in our earlier work,[10,11] which results in the reduction of our model accuracy.

The objective of the present work is to explore a better parameter optimization of our DG based solvation models. A pair of conditions is prescribed to ensure the physical solution of the GLB equation, which leads to the well-posedness of the GPB equation. Such a well-posedness in turn renders the stability of solving the GLB equation. The stable solution of the coupled GLB and GPB equation enables us to optimize the model parameters and produce the highly accurate prediction of solvation free energies. Some of the best results are obtained in the solvation free energy prediction of more than a hundred molecules of both polar and non-polar types.

The rest of this paper is organized as follows. To establish the notation and facilitate further development, we present a brief review of our DG based solvation models in Section II. By using the variational principle, we derive the coupled GLB and GPB equations. Necessary boundary conditions and initial values are prescribed to make this coupled system well-posed. Section III is devoted to parameter learning algorithms. We develop a protocol to stabilize the iterative solution process of coupled nonlinear PDEs. We introduce perturbation and convex optimization methods to ensure stability of the numerical solution of the GLB equation in coupling with the GPB equation. The newly achieved stability in solving the coupled PDEs leads to an appropriate optimization of solvation free energies with respect to our model parameters. In Section IV, we show that for more than a hundred of compounds of various types, including both polar and non-polar molecules, the present DG solvation model offers some of the most accurate solvation free energy prediction with the overall RMSE of 0.5 kcal/mol. This paper ends with a conclusion.

II. THE DG BASED SOLVATION MODEL

The free energy functional for our DG based full solvation model can be expressed as[10,11,71]

$$G[S,\Phi] = \int \left\{ \gamma|\nabla S| + pS + (1-S)U + S\left[-\frac{\epsilon_m}{2}|\nabla\Phi|^2 + \Phi\ \rho_m\right] + (1-S)\left[-\frac{\epsilon_s}{2}|\nabla\Phi|^2 - k_B T \sum_\alpha \rho_{\alpha 0}\left(e^{-\frac{q_\alpha \Phi}{k_B T}} - 1\right)\right]\right\} d\mathbf{r}, \quad \mathbf{r} \in \mathbb{R}^3, \tag{1}$$

where γ is the surface tension, p is the relative pressure difference between solvent and solute, and U denotes the solvent-solute non-electrostatic interactions represented by the semi-discrete and semi-continuum Lennard-Jones potentials in the present work. Here, $0 \le S \le 1$ is a hypersurface or simply surface function that characterizes the solute domain and embeds the 2D surface in $\mathbb{R}^3$, whereas $1 - S$ characterizes the solvent domain.[10] One may consider S as the position-dependent volume fraction of the solute. Additionally, Φ is the electrostatic potential and ϵ_s and ϵ_m are the dielectric constants of the solvent and solute, respectively. Here, k_B is the Boltzmann constant, T is the temperature, $\rho_{\alpha 0}$ denotes the reference bulk concentration of the αth solvent species, and q_α denotes the charge valence of the αth solvent species, which is zero for an uncharged solvent component. We use ρ_m to represent the charge density of the solute. The charge density is often modeled by a point charge approximation

$$\rho_m = \sum_j^{N_m} Q_j \delta(\mathbf{r} - \mathbf{r}_j),$$

where Q_j denoting the partial charge of the jth atom in the solute. Alternatively, the charge density computed from the DFT, which changes during the iteration or energy optimization, can be directly employed as well.[13]

Note that although the surface tension γ and the relative pressure p are mostly positive for most solvent-solute systems, physically, they can be negative for some solvent-solute systems as well.

In Eq. (1), the first three terms consist of the so called non-polar solvation free energy functional while the last two terms form the polar one. After the variation with respect to S, we obtain an elliptic equation for the surface function S,

$$\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|}\right) + V = 0, \tag{2}$$

where the potential driven term is given by

$$V = -p + U + \frac{\epsilon_m}{2}|\nabla\Phi|^2 - \Phi\ \rho_m - \frac{\epsilon_s}{2}|\nabla\Phi|^2 - k_B T \sum_\alpha \rho_{\alpha 0}\left(e^{-\frac{q_\alpha \Phi}{k_B T}} - 1\right).$$

It is a standard procedure to seek the solution of Eq. (2) by converting it into a parabolic equation.[3] As such, we construct the following GLB equation:[10,11]

$$\frac{\partial S}{\partial t} = |\nabla S|\left[\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|}\right) + V\right]. \tag{3}$$

As in the non-polar case, solving generalized Laplace-Beltrami equation (3) generates the solvent-solute interface through the surface function S.

Additionally, variation with respect to Φ gives rise to the GPB equation,

$$-\nabla \cdot (\epsilon(S)\nabla\Phi) = S\rho_m + (1-S)\sum_\alpha q_\alpha \rho_{\alpha 0} e^{-\frac{q_\alpha \Phi}{k_B T}}, \tag{4}$$

where $\epsilon(S) = (1-S)\epsilon_s + S\epsilon_m$ is the generalized permittivity function. As shown in our earlier work,[10,71] $\epsilon(S)$ is a smooth dielectric function gradually varying from ϵ_m to ϵ_s. Thus, the solution procedure of the GPB equation avoids many numerical difficulties of solving elliptic equations with discontinuous coefficients[78,79,82,85,86] in the standard PB equation.

GLB (3) and GBP (4) equations form a highly nonlinear system, in which the GLB equation is solved for the interface profile S of the solute and solvent. The interface profile determines the dielectric function $\epsilon(S)$ in the GPB equation. The GPB equation is solved for the electrostatics potential Φ that behaves as an external potential in the GLB equation. The strongly coupled system should be solved in self-consistent iterations.

For GLB equation (3), the computational domain is $\Omega/\Omega_m^{\text{vdW}}$, where Ω_m^{vdW} is the solute van der Waals domain given by $\Omega_m^{\text{vdW}} = \bigcup_i B(r_i^{\text{vdW}})$. Here, $B(r_i^{\text{vdW}})$ is the ith ball in the solute centered at $\mathbf{r}_i$ with van der Waals radius r_i^{vdW}. We apply the following Dirichlet boundary condition to $S(\mathbf{r},t)$:

$$S(\mathbf{r},t) = \begin{cases} 0, & \forall \mathbf{r} \in \partial\Omega \\ 1, & \forall \mathbf{r} \in \partial\Omega_m^{\text{vdW}} \end{cases}. \tag{5}$$

The initial value of $S(\mathbf{r},t)$ is given by

$$S(\mathbf{r},0) = \begin{cases} 1, & \forall \mathbf{r} \in \partial\Omega_m^{\text{ext}}, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $\partial\Omega_m^{\text{ext}}$ is the boundary of the extended solute domain constructed by $\Omega_m^{\text{ext}} = \bigcup_i B(r_i^{\text{vdW}} + r^{\text{probe}})$. Here, $B(r_i^{\text{vdW}} + r^{\text{probe}})$ has an extended radius of $r_i^{\text{vdW}} + r^{\text{probe}}$ with r^{probe} being the probe radius, which is set to 1.4 Å in the present work.

For GPB equation (4), the computational domain is Ω. We set the Dirichlet boundary condition via the Debye-Hückel expression,

$$\Phi(\mathbf{r}) = \sum_{i=1}^{N_m} \frac{Q_i}{\epsilon_s |\mathbf{r} - \mathbf{r}_i|} e^{-\bar{\kappa}|\mathbf{r}-\mathbf{r}_i|}, \quad \forall \mathbf{r} \in \partial\Omega, \tag{7}$$

where $\bar{\kappa}$ is the modified Debye-Hückel screening function,[11] which is zero if there is no salt molecule in the solvent. Note that no interface condition[77] is needed as S and $\epsilon(S)$ are smooth functions in general for $t > 0$. Consequently, resulting GBP (4) equation is easy to solve.

To compare with experimental solvation data, one needs to compute the total solvation free energy, which, in our DG based solvation model, is obtained as

$$\Delta G = \Delta G^{\mathrm{P}} + G^{\mathrm{NP}}, \tag{8}$$

where ΔG^{P} is the electrostatic solvation free energy,

$$\Delta G^{\mathrm{P}} = \frac{1}{2} \sum_{i=1}^{N_m} Q_i \left[\Phi(\mathbf{r}_i) - \Phi_h(\mathbf{r}_i)\right], \tag{9}$$

where Φ_h is the solution of the above GPB model in a homogeneous system, obtained by setting a constant permittivity function $\epsilon(\mathbf{r}) = \epsilon_m$ in the whole domain Ω. The non-polar energy G^{NP} is computed by

$$G^{\mathrm{NP}} = \int \left[\gamma |\nabla S| + pS + (1 - S)U\right] d\mathbf{r}. \tag{10}$$

The DG based solvation model is formulated as a coupled GLB and GPB equation system, in which the GLB equation provides the solvent solute boundary for solving the GPB, while the GPB equation produces the external potential in the GLB equation for the evolution of the surface function S. The solution procedure for this coupled system has been discussed in our earlier work.[10,11] Essentially, for the GLB equation, an alternating direction implicit (ADI) scheme was utilized for the time integration, in conjugation with the second order finite difference method for the spatial discretization. The GPB equation was discretized by a standard second order finite difference scheme and the resulting algebraic equation system was solved by using a standard Krylov subspace method based solver.[10,11]

III. PARAMETRIZATION METHODS AND ALGORITHMS

To solve the above coupled equation system, a set of parameters that appeared in the GLB equation, namely, surface tension γ, hydrodynamic pressure difference p, and the product of solvent density and well depth parameter of the jth atom $\tilde{\varepsilon}_{j\alpha} \doteq \rho_\alpha \varepsilon_j$, should be predetermined. Unfortunately, this coupled system is unstable at the certain choices of parameters. Specifically, for certain V, one may have $S > 1$ or $S < 0$, which leads to unphysical $\epsilon(S)$ and unphysical solution of GPB equation (4) and thus gives rise to a divergent S. This instability can seriously reduce the model accuracy.[10,11]

For a concise description of our algorithm, we assume that there is only one solvent component (water) and denote the parameter set as

$$\mathbf{P} = \{\gamma, p, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_{N_T}\}, \tag{11}$$

where N_T is the number of types of atoms in the solute molecule.

As mentioned in the previous part, the parameter set $\mathbf{P}$ used in solving the coupled PDEs should meet two requirements, namely, the stability of solving the coupled PDEs and the optimal prediction of the solvation free energy (or fitting the experimental solvation free energy in the best approach). Based on these two criteria, we introduce a two-stage numerical procedure to optimize the parameter set and solve the coupled PDEs:

- Explore the stability conditions of the coupled PDEs by introducing an auxiliary system via a small perturbation.
- Optimize the parameter set by an iteratively scheme satisfying the stability constraint.

A. Stability conditions

In this part, we investigate the stability conditions for the numerical solution to coupled PDEs (3) and (4). The basic idea is to utilize a small perturbation method. It is known that omitting the external potential in the GLB equation yields the LB equation,

$$\frac{\partial S}{\partial t} = |\nabla S| \nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|}\right). \tag{12}$$

This equation is of diffusion type and is well posed with the Dirichlet type of boundary conditions provided $\gamma > 0$. Numerically, it is easy to solve Eq. (12) to yield the profile of the solvent solute boundary.

After solving LB equation (12), we use the generated smooth profile of the solvent solute boundary to determine the permittivity function in the GPB equation. For simplicity, we consider a pure water solvent,

$$-\nabla \cdot (\epsilon(S) \nabla \Phi) = S \rho_m. \tag{13}$$

Without the external potential, the system of Eqs. (12) and (13) can be solved stably by first solving the LB equation and then the GPB equation.

Motivated by the above observation, if the external potential is dominated by the mean curvature term, the stability of coupled GPB and GLB equations can be preserved. Based on numerical experiments, the Lennard-Jones interaction between the solvent and solute is usually small since this term is constrained by the non-polar free energy in our model. In our method, we enforce the following constraint conditions to make the coupled system well-posed in the numerical sense:

$$\gamma > \gamma_0 > 0 \tag{14}$$

and

$$|p| \leq \beta \gamma, \tag{15}$$

where γ_0 and β are some appropriate positive constants.

In summary, the original problem is transformed into optimizing parameters in the following system to attain the best solvation free energy fitting with experimental results:

$$\begin{cases} \dfrac{\partial S}{\partial t} = |\nabla S| \left[\nabla \cdot \left(\gamma \dfrac{\nabla S}{|\nabla S|} \right) - p + U + \dfrac{1}{2}\epsilon_m |\nabla \Phi|^2 - \dfrac{1}{2}\epsilon_s |\nabla \Phi|^2 \right], \\ -\nabla \cdot (\epsilon(S)\nabla \Phi) = S\rho_m, \\ \gamma > \gamma_0 > 0, \\ |p| \leq \beta\gamma. \end{cases} \tag{16}$$

Note that the potential $\rho_m \Phi$ is omitted in GLB equation (16), because we have already enforced the Dirichlet boundary condition in the GLB equation, while ρ_m is inside the van der Waals surface.

Remark 1. Based on large amount of numerical tests, it is found that there is no need to enforce the constraint conditions on the parameters that appear in the Lennard-Jones term. When this term is used to fit the solvation energy with experimental results, the parameters can be bounded in a small neighborhood of 0 automatically during the fitting procedure. These parameters essentially do not affect the numerical stability.

B. Self-consistent approach for solving the coupled PDEs

In this part, we propose a self-consistent approach to solve the coupled GLB and GPB equations for a given set of parameters. Basically, the coupled system is solved iteratively until both the electrostatic solvation free energy ΔG^{P} given in Eq. (9) and the surface function S are both converged. Here, the surface function is said to be converged provided that the surface area and enclosed volume are both converged.

We present an algorithm for solving the following coupled systems:

$$-\nabla \cdot (\epsilon(S)\nabla \Phi) = S\rho_m \tag{17}$$

and

$$\frac{\partial S}{\partial t} = |\nabla S| \left[\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|} \right) + V_e \right], \tag{18}$$

where V_e is the external potential which is defined as

- **Auxiliary system:** $V_e = \frac{1}{2}(\epsilon_m - \epsilon_s)|\nabla \Phi|^2$.
- **Full system:** $V_e = -p + U + \frac{1}{2}(\epsilon_m - \epsilon_s)|\nabla \Phi|^2$.

Dirichlet boundary conditions are employed for both GPB (17) and GLB (18) equations with auxiliary and full external potentials, giving rise to a well-posed coupled system. The smooth profile of the solvent-solute boundary enables the direct use of the second order central finite difference scheme to achieve the second order convergence in discretizing the GPB equation. The biconjugate gradient scheme is used to solve the resulting algebraic equation system. The GLB equation of both the auxiliary and full systems can be solved by the central finite difference discretization of the spatial domain and the forward Euler time integrator for the time domain discretization.

Remark 2. For the sake of simplicity, in the current work, we employed the central finite difference scheme for spatial domain discretization in both GPB and GLB equations, and forward Euler integrator for the time domain discretization of GLB equation. For stability consideration, in the discretization of the GLB equation, the discretization step size of temporal and spatial domain satisfies the Courant-Friedrichs-Lewy condition. To accelerate the numerical integration, a multigrid solver can be employed for GBP equation, and an alternating direction implicit scheme,[10] which is unconditionally stable, can be utilized for the temporal integration. However, detailed discussion of these accelerated schemes is beyond the scope of the present work.

A pseudocode is given in Algorithm 1 to offer a general framework for solving the coupled GLB and GPB equations in a self-consistent manner. The outer iteration controls the convergence of the GPB equation through measuring the change of electrostatic solvation free energy in two adjacent iterations, while the inner iteration controls the convergence of the GLB equation based on the variation of surface areas and enclosed volumes through the surface function S. The

ALGORITHM 1. Self-consistent algorithm for the coupled GPB and GLB system.

procedure GPB-GLB-Solver
  Initialize: $\Delta G_1^{\mathrm{P}} = 0$, $\Delta G_2^{\mathrm{P}} = 100$, $\mathrm{Area}_1 = 0$, $\mathrm{Area}_2 = 100$, $\mathrm{Vol}_1 = 0$, $\mathrm{Vol}_2 = 100$
  do while ($|\Delta G_1^{\mathrm{P}} - \Delta G_2^{\mathrm{P}}| > \epsilon_1$)
    $\Delta G_1^{\mathrm{P}} \leftarrow \Delta G_2^{\mathrm{P}}$
    do while ($|\mathrm{Area}_1 - \mathrm{Area}_2| > \epsilon_2$ .and. $|\mathrm{Vol}_1 - \mathrm{Vol}_2| > \epsilon_3$)
      $\mathrm{Area}_1 \leftarrow \mathrm{Area}_2$, $\mathrm{Vol}_1 \leftarrow \mathrm{Vol}_2$.
      Update the surface function S by solving the GLB equation (18).
      $\mathrm{Area}_2 = \int_\Omega |\nabla S| d\mathbf{r}$, $\mathrm{Vol}_2 = \int_\Omega S d\mathbf{r}$.
    enddo
    Solve the GPB equation (17) in both vacuum and solvent with the previous updated surface profile.
    Update the polar solvation free energy ΔG_2^{P} according to Eq. (9).
  enddo

variables ΔG_1^P, ΔG_2^P, Area_1, Area_2, Vol_1, and Vol_2 denote the electrostatic solvation free energy, surface area, and volume enclosed by the surface of two immediate iterations, respectively. The parameters ϵ_1, ϵ_2, and ϵ_3 are the threshold constants and all set to 0.01 in the current implementation.

Remark 3. In solving the GLB equation, during each updating, to ensure the stability, instead of the fully update, we update it partially, i.e., the updated solution is the weighted sum of the new solution of the current GLB solution S_{new} and the old solution of the GLB equation in the previous step S_{old},

$$S = a_1 S_{\text{new}} + (1 - a_1) S_{\text{old}}, \tag{19}$$

where a_1 is a constant and set to 0.5 in the present work.

C. Convex optimization for parameter learning

In this part, we present the parameter optimization scheme. In our approach, parameters start from an initial guess and then are updated sequentially until reaching the convergence. Here, the convergence is measured by the root mean square (RMS) error between the fitted and experimental solvation free energies for a given set of molecules.

Consider the parameter optimization for a given group of molecules, denoted as $\{T_1, T_2, \ldots, T_n\}$. As discussed above, the parameter set is **P**. To optimize the parameter set **P**, we start from GPB equation (17) and the auxiliary system of GLB equation (18) with $\gamma = 0.05$. After solving the initial coupled system by using Algorithm 1, we obtain the following quantities for each molecule in the training set:

$$\left\{ \Delta G_j^P, \text{Area}_j, \text{Vol}_j, \left(\sum_{i=1}^{N_m} \delta_i^1 \int_{\Omega_s} \left[\left(\frac{\sigma_s + \sigma_1}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{12} - 2 \left(\frac{\sigma_s + \sigma_1}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{6} \right] d\mathbf{r} \right)_j, \tag{20}$$

$$\cdots, \tag{21}$$

$$\left(\sum_{i=1}^{N_m} \delta_i^{N_T} \int_{\Omega_s} \left[\left(\frac{\sigma_s + \sigma_{N_T}}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{12} - 2 \left(\frac{\sigma_s + \sigma_{N_T}}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{6} \right] d\mathbf{r} \right)_j \right\}, \tag{22}$$

where $j = 1, 2, \ldots, n$. Here, N_m and N_T denote the number of atoms and types of atoms in a specific molecule. The last few terms involve semi-discrete and semi-continuum Lennard-Jones potentials.[10] Additionally,

$$\delta_i^j = \begin{cases} 1, & \text{if atom } i \text{ belongs to type } j, \\ 0, & \text{otherwise,} \end{cases}$$

where $i = 1, 2, \ldots, N_m$; $j = 1, 2, \ldots N_T$; $\sigma_i, i = 1, 2, \ldots, N_T$ is the atomic radius of the ith type of atoms. Therefore, atoms of the same type have a common atomic radius and fitting parameter $\tilde{\varepsilon}$.

The predicted solvation free energy for molecule j can be represented as

$$\Delta G_j = \Delta G_j^P + \gamma \text{Area}_j + p \text{Vol}_j + \tilde{\varepsilon}_1 \left(\sum_{i=1}^{N_m} \delta_i^1 \int_{\Omega_s} \left[\left(\frac{\sigma_s + \sigma_1}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{12} - 2 \left(\frac{\sigma_s + \sigma_1}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{6} \right] d\mathbf{r} \right)_j \tag{23}$$

$$+ \cdots + \tilde{\varepsilon}_{N_T} \left(\sum_{i=1}^{N_m} \delta_i^{N_T} \int_{\Omega_s} \left[\left(\frac{\sigma_s + \sigma_{N_T}}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{12} - 2 \left(\frac{\sigma_s + \sigma_{N_T}}{\|\mathbf{r} - \mathbf{r}_i\|} \right)^{6} \right] d\mathbf{r} \right)_j. \tag{24}$$

We denote the predicted solvation free energy for the given set of molecules as $\Delta \mathbf{G}(\mathbf{P}) \doteq \{\Delta G_1, \Delta G_2, \ldots, \Delta G_n\}$, which is a function of the parameter set **P**, and denote the corresponding experimental solvation free energy as $\Delta \mathbf{G}^{\text{Exp}} \doteq \{\Delta G^{\text{Exp}1}, \Delta G^{\text{Exp}2}, \ldots, \Delta G^{\text{Exp}n}\}$.

Then, the parameter optimization problem in the coupled PDEs given by Eq. (16) can be transformed into the following regularized and constrained optimization problem:

$$\min_{\mathbf{P}} \left(\|\Delta \mathbf{G}(\mathbf{P}) - \Delta \mathbf{G}^{\text{Exp}}\|_2 + \lambda \|\mathbf{P}\|_2 \right), \tag{25}$$

subject to

$$\gamma \geq \gamma_0 \tag{26}$$

and

$$|p| \leq \beta \gamma, \tag{27}$$

where $\| * \|_2$ is the L_2 norm of the quantity $*$ and λ is the regularization parameter chosen to be 10 in the present work to ensure the dominance of the first term and avoid over-fitting. Here, γ_0 and β are set, respectively, to 0.05 and 0.1 in the present implementation, which guarantees the stability of the coupled system according to a large amount of numerical tests.

It is obvious that objective function (25) in the optimization is a convex function, meanwhile the solution domain restricted by constraints (26) and (27) forms a convex domain. Therefore, the optimization problem given by Eqs. (25)-(27) is a convex optimization problem, which was studied by Grant and Boyd.[34,35]

After solving the above convex optimization problem, parameter set **P** is updated and used again in solving the coupled GLB and GPB system, i.e., Eqs. (18) and (17). Repeating the above procedure, a new group of predicted

ALGORITHM 2. Parameters learning for a given group of molecules.

1: **procedure** PARAMETERS-LEARNING
2: **Initialize**: $\mathrm{Err}_1 = 0$, $\mathrm{Err}_2 = 100$
3: Solve the coupled GPB and GLB system, where GLB utilizes the auxiliary equation (18).
4: Solve the constrained optimization problem Eqs. (25)-(27) to obtain the initial parameter set $\mathbf{P}_0$.
5: Update Err_1 to be the RMS error between experimental and predicted results in the above step.
6: **do while** $(|\mathrm{Err}_1 - \mathrm{Err}_2| > \epsilon_4)$
7: $\mathrm{Err}_2 \leftarrow \mathrm{Err}_1$.
8: Solve the coupled GPB and GLB system, where GLB system with parameters set $\mathbf{P}_0$.
9: Solve the constrained optimization problem Eqs. (25)-(27) to get the updated parameters set $\mathbf{P}$.
10: Update Err_1 to be RMS error between experimental and predicted results in the previous optimization step.
11: Update $\mathbf{P}_0 \leftarrow \mathbf{P}$.
12: **enddo**

solvation free energies together with a new group of parameters is obtained. This procedure is repeated until the RMS error between the predicted and experimental solvation free energies in two sequential iterations is within a given threshold.

D. Algorithm for parameter optimization and solution of the coupled PDEs

Based on the preparation made in Subsections III B and III C, namely, the self-consistent approach for solving the coupled GLB and GPB system and the parameter optimization, we provide the combined algorithm for the parameter optimization and solving the coupled system for a given set of molecules.

Algorithm 2 offers a parameter learning pseudocode for a given group of molecules. This algorithm is formulated by combining outer and inner self-consistent iterations. The outer iteration controls the convergence of the optimized parameters via two controlling parameters, Err_1 and Err_2, denoting the RMS error between predicted and experimental solvation free energies in two sequential iterations. The inner iteration implements the solution to the GLB and GPB equations by Algorithm 1.

The threshold parameter ϵ_4 is set to 0.01 in the present work.

IV. NUMERICAL RESULTS

In this section, we present the numerical study of the DG based solvation model using the proposed parameter optimization algorithms. We first explore the optimal solvent radius used in the van der Waals interactions. Due to the

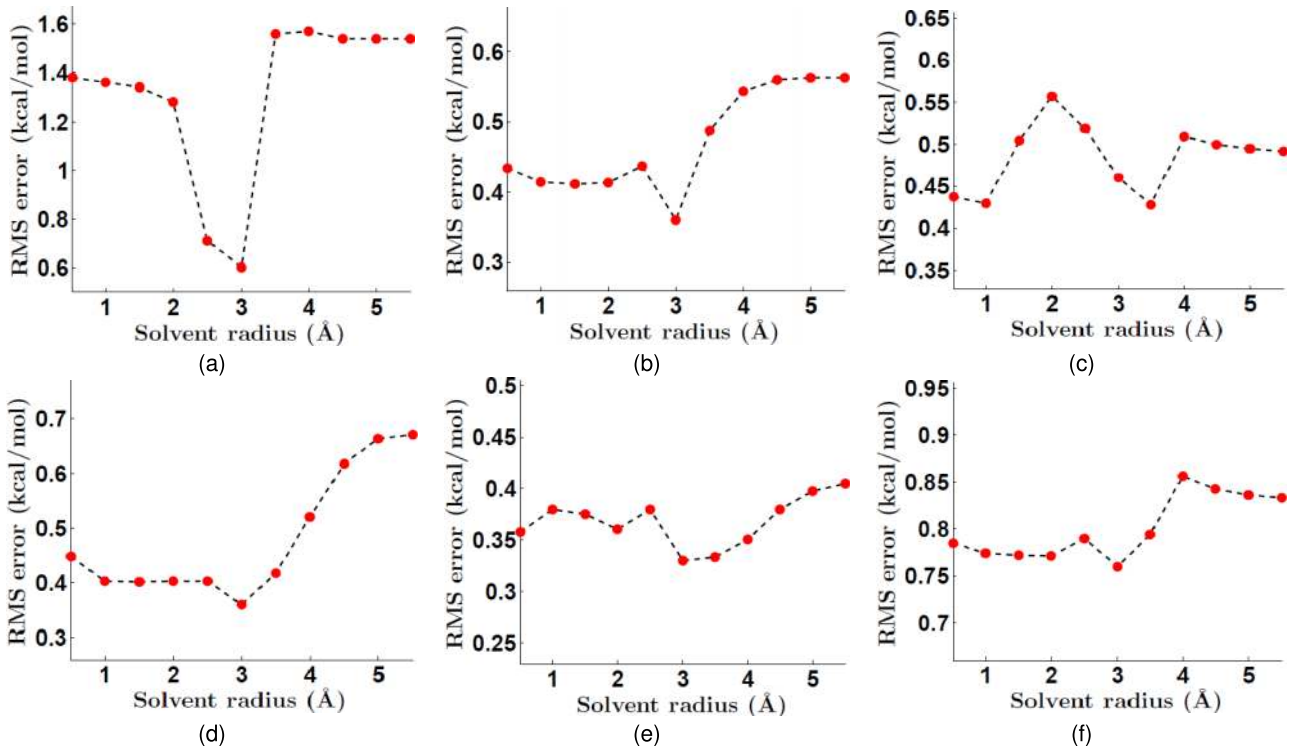

FIG. 1. The relations between the solvent radii and the RMS errors. (a) SAMPL0 test set; (b) alkane set; (c) alkene set; (d) ether set; (e) alcohol; (f) phenol set. Notably, there is a common local minimum at the solvent radii 3.0 Å for all test sets except for the alkene set.

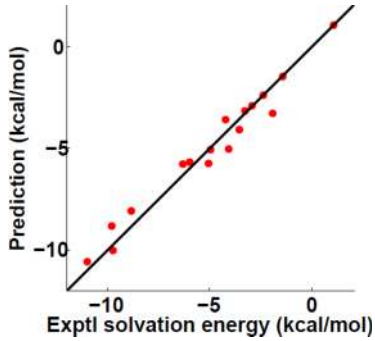

FIG. 2. The predicted and experimental solvation free energy for the 17 molecules in the SAMPL0 test set.

high nonlinearity, the solvent radius cannot be automatically optimized and its optimal value is obtained via searching the parameter domain. We show that for a group of molecules, there is a local minimum in the RMS error when the solvent radius is varied. The corresponding optimal solvent radius is adopted for other molecules. Additionally, we consider a large number of molecules with known experimental solvation free energies to test the proposed parameter optimization algorithms. These molecules are of both polar and non-polar types and are divided into six groups: the SAMPL0 test set,[53] the alkane, alkene, ether, alcohol, and phenol types.[50] It is found that our DG based solvation model works really well for these molecules. Finally, to demonstrate the predictive power of the present DG based solvation model, we perform a fivefold cross validation[39] for alkane, alkene, ether, alcohol, and phenol types of molecules. It is found that training and validation errors are of the same level, which confirms the ability of our model for the solvation free energy prediction.

The SAMPL0 molecule structural conformations are adopted from the literature with ZAP 9 radii and the OpenEye-AM1-BCC v1 charges.[53] For other molecules, structural conformations are obtained from FreeSolv.[50] General AMBER force field (GAFF) is utilized for the charge assignment.[9] The van der Waals radii as well as the atomic radii of hydrogen, carbon, and oxygen atoms are set to 1.2, 1.7, and 1.5 Å, respectively. The grid spacing is set to 0.25 Å in all of our calculations (discretization and integration). The computational domain is set to the bounding box of the solute molecule with an extra buffer length of 6.0 Å.

A. Solvent radius

In the present semi-discrete and semi-continuum Lennard-Jones potential, $\tilde{\epsilon}_k \int_{\Omega_s} \left[\left(\frac{\sigma_s+\sigma_i}{\|\mathbf{r}-\mathbf{r}_i\|} \right)^{12} - 2 \left(\frac{\sigma_s+\sigma_i}{\|\mathbf{r}-\mathbf{r}_i\|} \right)^{6} \right] d\mathbf{r}$, the positions $\mathbf{r}_i$, $(i = 1, 2, \ldots, N_m)$ are the coordinates of solute atoms, while $\mathbf{r}$ is not the position of a regular solvent atom or molecule. Since the solvent is treated as a continuum, $\mathbf{r}$ varies, in principle, continuously over the whole solvent domain. The distance $\|\mathbf{r} - \mathbf{r}_i\|$ is scaled by the sum of solvent radius σ_s and solute radii σ_i. Because of the explicit representation of solute atoms, solute atomic radii σ_i are set to their van der Waals radii, the radii that define the van der Waals surface, which is used for setting up the boundary condition for the GLB equation. However, the continuum treatment of the solvent prevents us to simply associate σ_s with the radius of the solvent molecule. Unlike the fully discrete Lennard-Jones potential in explicit solvent models, the semi-discrete and semi-continuum Lennard-Jones potential in our DG based solvation model describes the "interaction" of a solute atom with an arbitrary position in the solvent domain. In numerical approximation, the arbitrary position is represented by a grid mesh. Therefore, one cannot simply take the solvent radius in the present model as the radius of individual (discrete) solvent molecules. Additionally, it is noted that the solvent radius in the present work and solvent probe radius in the Poisson-Boltzmann theory are two different concepts. In the present work, solvent radius σ_s is considered as an optimization parameter. Note that due to the nonlinear nature, this optimization cannot be carried out together or mixed with the parameter optimization discussed in Sec. III.

We utilize a brute force approach for the solvent radius selection or optimization. Six sets of test examples are utilized to explore appropriate solvent radius. The SAMPL0 test

TABLE I. The solvation free energy prediction for the SAMPL0 set. Energy is in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	$\Delta G^{\mathrm{Exp53}}$	Error
Glycerol triacetate	−10.60	2.53	−8.07	−8.84	−0.77
Benzyl bromide	−4.31	1.93	−2.38	−2.38	0.00
Benzyl chloride	−4.45	1.18	−3.27	−1.93	1.34
m-bis (trifluoromethyl) benzene	−2.62	3.70	1.08	1.07	−0.01
N,N-dimethyl-p-methoxybenzamide	−8.35	−2.22	−10.57	−11.01	−0.45
N,N-4-trimethylbenzamide	−6.93	−3.09	−10.03	−9.76	0.27
bis-2-chloroethyl ether	−3.73	−0.14	−3.59	−4.23	−0.64
1,1-diacetoxyethane	−7.07	2.00	−5.07	−4.97	0.10
1,1-diethoxyethane	−3.58	0.43	−3.15	−3.28	−0.13
1,4-dioxane	−5.36	−0.38	−5.74	−5.05	0.69
Diethyl propanedioate	−7.07	1.40	−5.67	−6.00	−0.33
Dimethoxymethane	−4.09	1.19	−2.90	−2.93	−0.03
Ethylene glycol diacetate	−7.66	1.90	−5.76	−6.34	−0.58
1,2-diethoxyethane	−3.64	0.45	−4.09	−3.54	0.55
Diethyl sulfide	−2.21	0.76	−1.47	−1.43	0.04
Phenyl formate	−7.10	2.08	−5.02	−4.08	0.94
Imidazole	−11.54	2.71	−8.83	−9.81	−0.98
RMS					0.60

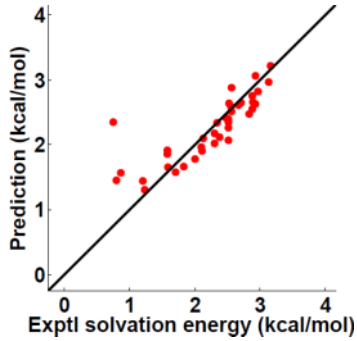

FIG. 3. The predicted and experimental solvation free energies for 38 alkane molecules.

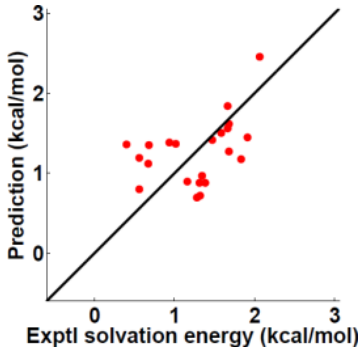

FIG. 4. The predicted and experimental solvation free energies for 22 alkene molecules.

TABLE II. The solvation free energy prediction for the alkane set. All energies are in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	ΔG^{Exp}[50]	Error
Octane	−0.13	2.89	2.76	2.88	0.12
Ethane	−0.04	1.70	1.66	1.83	0.17
Propane	−0.05	1.83	1.78	2.00	0.22
Cyclopropane	−0.08	2.43	2.35	0.75	−1.60
Isobutane	−0.07	2.09	2.02	2.30	0.28
2,2-dimethylbutane	−0.07	2.34	2.27	2.51	0.24
Isopentane	−0.07	2.19	2.12	2.38	0.26
2,3-dimethylbutane	−0.07	2.41	2.34	2.34	0.00
3-methylpentane	−0.08	2.43	2.35	2.51	0.16
Methylcyclopentane	−0.10	1.76	1.66	1.59	−0.07
n-butane	−0.07	2.03	1.96	2.10	0.14
Isohexane	−0.09	2.49	2.40	2.51	0.11
2,4-dimethylpentane	−0.09	2.57	2.48	2.83	0.35
Methylcyclohexane	−0.10	1.68	1.58	1.70	0.12
n-pentane	−0.08	2.25	2.17	2.30	0.13
Hexane	−0.09	2.51	2.42	2.48	0.06
Cyclohexane	−0.10	1.40	1.30	1.23	−0.07
Nonane	−0.14	3.11	2.97	3.13	0.16
Heptane	−0.11	2.73	2.62	2.67	0.05
Cyclopentane	−0.10	1.54	1.44	1.20	−0.24
Cycloheptane	−0.11	1.56	1.45	0.80	−0.65
Cyclooctane	−0.12	1.69	1.57	0.86	−0.71
Neopentane	−0.06	2.13	2.07	2.51	0.44
2,2,4-trimethylpentane	−0.08	2.74	2.66	2.89	0.23
3,3-dimethylpentane	−0.07	2.58	2.51	2.56	0.05
2,3-dimethylpentane	−0.08	2.72	2.64	2.52	−0.12
2,3,4-trimethylpentane	−0.08	2.96	2.88	2.56	−0.32
1,2-dimethylcyclohexane	−0.10	2.02	1.92	1.58	−0.34
3-methylhexane	−0.09	2.74	2.65	2.71	0.06
3-methylheptane	−0.11	2.94	2.83	2.97	0.14
1,4-dimethylcyclohexane	−0.11	2.02	1.91	2.11	0.20
2,2-dimethylpentane	−0.08	2.64	2.56	2.88	0.32
2-methylhexane	−0.10	2.73	2.63	2.93	0.30
Decane	−0.16	3.37	3.21	3.16	−0.06
Propylcyclopentane	−0.12	2.21	2.09	2.13	0.03
cis-1,2-dimethylcyclohexane	−0.09	1.95	1.86	1.58	−0.28
2,2,5-trimethylhexane	−0.09	3.15	3.06	2.93	−0.13
Pentylcyclopentane	−0.15	2.73	2.58	2.55	−0.04
RMS					0.36

TABLE III. The solvation free energy prediction for the alkene set. All energies are in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	ΔG^{Exp}[50]	Error
Ethylene	−0.27	0.96	0.69	1.28	0.59
Isoprene	−0.62	1.97	1.35	0.68	−0.67
but-1-ene	−0.29	1.17	0.88	1.38	0.50
Butadiene	−0.56	1.75	1.19	0.56	−0.63
pent-1-ene	−0.30	1.57	1.27	1.68	0.41
prop-1-ene	−0.32	1.03	0.71	1.32	0.61
2-methylprop-1-ene	−0.37	1.26	0.89	1.16	0.27
Cyclopentene	−0.37	1.17	0.79	0.56	−0.23
2-methylbut-2-ene	−0.40	1.28	0.87	1.31	0.44
2,3-dimethylbuta-1,3-diene	−0.65	2.01	1.36	0.40	−0.95
3-methylbut-1-ene	−0.27	1.45	1.18	1.83	0.65
1-methylcyclohexene	−0.38	1.50	1.11	0.67	−0.45
penta-1,4-diene	−0.53	1.91	1.38	0.93	−0.45
hex-1-ene	−0.30	1.81	1.50	1.58	0.08
hexa-1,5-diene	−0.51	1.88	1.37	1.01	−0.36
hept-1-ene	−0.33	2.17	1.84	1.66	−0.18
hept-2-ene	−0.34	1.96	1.62	1.68	0.06
4-methyl-1-pentene	−0.26	1.71	1.45	1.91	0.46
2-methylpent-1-ene	−0.33	1.75	1.42	1.47	0.05
non-1-ene	−0.36	2.81	2.45	2.06	−0.39
trans-2-heptene	−0.34	1.90	1.56	1.66	0.10
trans-2-pentene	−0.30	1.26	0.96	1.34	0.38
RMS					0.46

set[53] is a benchmark having 17 molecules. Additionally, we consider 38 alkane, 22 alkene, 17 ether, 25 alcohol, and 18 phenol molecules. The solvent radius is varied from 0.5 Å to 5.5 Å away from van der Waals surface. Due to the fast decay property of the Lennard-Jones interactions, the above setting enables the full inclusion of the Lennard-Jones interactions in our model. Figure 1 depicts the RMS errors of six test sets at different solvent radii calculated from the present DG based solvation model. In Figure 1(a), the result clearly demonstrates that with the increase of the solvent radius, the RMS error decreases dramatically initially. The minimum appears at 3.0 Å. The further increase of the solvent radius leads to a rapid jump in the RMS error before it stabilizes around 1.54 kcal/mol. It is noted that 3.0 Å is much larger than the commonly used solvent probe radius of 1.4 Å in Poisson-Boltzmann theory based implicit solvent models. For other five test sets, although the behavior of the RMS errors differs in each case, essentially all the RMS errors have a local minimum at the solvent radius of 3 Å. Therefore, in all the following computations, the solvent radius is set to 3.0 Å.

B. Optimization results

In this section, we illustrate the performance of our parameter optimization algorithms. First, we provide the regression results of the SAMPL0 test set.[53] Figure 2 shows the predicted and experimental solvation free energies based on the present model and optimization method. It is obvious that predicted solvation free energies are highly consistent with the experimental ones. The RMS error is 0.60 kcal/mol.

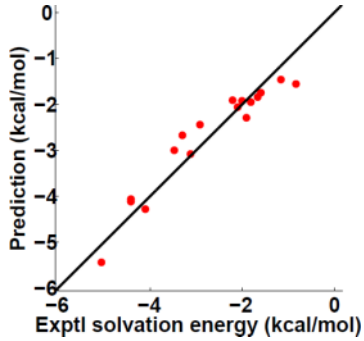

FIG. 5. The predicted and experimental solvation free energy for the 17 ether molecules.

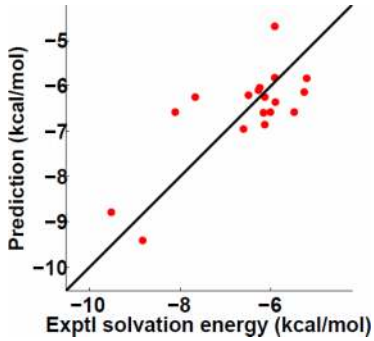

FIG. 7. The predicted and experimental solvation free energy for the 18 phenol molecules.

Table I shows the breakup of polar, non-polar, and total predicted solvation free energies. The experimental values and errors are also provided.[53]

Compared to our earlier prediction[10] in which the same model is employed but the parameters were not optimized in the present manner, the RMS error decreases dramatically from previous 1.76 kcal/mol to 0.60 kcal/mol for the same test set. Note that the present RMS error (0.60 kcal/mol) is also significantly smaller than that of the explicit solvent approach (1.71 ± 0.05 kcal/mol) and that obtained by the PB based prediction (1.87 kcal/mol) under the same structure, charge, and radius setting.[53] The present results confirm the efficiency of the proposed new parameter optimization algorithms and demonstrate the accuracy and power of our DG based solvation models.

Additionally, we investigate the solvation free energies prediction of two families of non-polar molecules, alkane and alkene, which were studied previously by using our DG based non-polar solvation model.[16] In the following, we demonstrate that the present DG based full solvation model can provide the same level of accuracy in the solvation free energy prediction for alkane and alkene molecules.

Figures 3 and 4 depict the predicted and experimental solvation free energies for 38 alkane and 22 alkene molecules, respectively. Tables II and III list the polar, non-polar, total, and experimental solvation free energies for both families of solute molecules, respectively. Except for one alkane molecule, namely, cycloprotane, whose predicted error is 1.60 kcal/mol, the errors for all other molecules are within 1 kcal/mol. The RMS errors of these two families are 0.36 and 0.46 kcal/mol, respectively. This level of accuracy is similar to our earlier results obtained by using our DG based non-polar solvation model,[16] which does not involve the electrostatic (polar) model and is computationally easier to optimize.

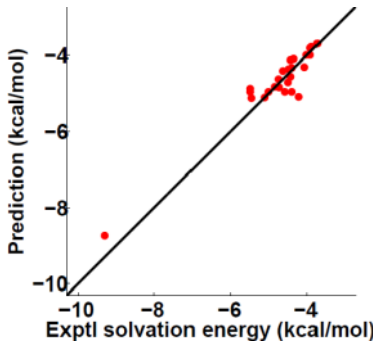

FIG. 6. The predicted and experimental solvation free energy for the 25 alcohol molecules.

It is interesting to note that for both alkane and alkene molecules, the polar solvation free energy contribution is very small and the non-polar part dominates the solvation free energy contribution, which explains why the DG based non-polar solvation model works extremely well for the solvation free energy prediction of alkane and alkene molecules.[16] Further, note that for almost all the alkane molecules, the polar solvation free energies ΔG^{P} are of magnitude 0.01 kcal/mol, while alkene molecules have slightly larger magnitude polar free energies, which further verifies that alkene molecules have a stronger polarity than alkane molecules in general.

Finally, we analyze three classes of polar solute molecules, namely, ether, alcohol, and phenol molecules. Figures 5–7 illustrate the predicted and experimental solvation free energies for 17 ether, 25 alcohol, and 18 phenol molecules, respectively. Tables IV–VI list the polar, non-polar, total, and experimental solvation free energies for the corresponding

TABLE IV. The solvation free energy prediction for the ether set. All energies are in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	$\Delta G^{\mathrm{Exp50}}$	Error
Ethoxyethane	−4.08	2.33	−1.75	−1.59	0.16
2-methyltetrahydrofuran	−4.10	1.43	−2.67	−3.30	−0.63
Tetrahydrofuran	−4.36	1.36	−3.00	−3.47	−0.47
1-propoxypropane	−3.75	2.29	−1.46	−1.16	0.30
Methoxymethane	−4.55	2.26	−2.29	−1.91	0.36
Tetrahydropyran	−4.17	1.09	−3.07	−3.12	−0.05
1-butoxybutane	−3.88	2.33	−1.55	−0.83	0.72
Trimethoxymethane	−7.57	3.51	−4.06	−4.42	−0.36
Methoxyethane	−4.35	2.29	−2.06	−2.10	−0.04
1-methoxypropane	−4.08	2.24	−1.84	−1.66	0.18
2-methoxypropane	−4.12	2.20	−1.92	−2.01	−0.09
1-ethoxypropane	−4.26	2.32	−1.94	−1.81	0.13
1,3-dioxolane	−6.09	1.81	−4.28	−4.10	0.18
2,5-dimethyltetrahydrofuran	−3.86	1.42	−2.44	−2.92	−0.48
1,1,1-trimethoxyethane	−7.58	3.46	−4.12	−4.42	−0.30
2-methoxy-2-methyl-propane	−3.88	1.97	−1.91	−2.21	−0.30
1,4-dioxane	−7.09	1.66	−5.44	−5.06	0.38
RMS					0.36

TABLE V. The solvation free energy prediction for the alcohol set. All energies are in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	$\Delta G^{\mathrm{Exp50}}$	Error
Ethylene glycol	−6.98	−1.76	−8.73	−9.30	−0.57
butan-1-ol	−3.33	−1.51	−4.84	−4.72	0.12
Ethanol	−3.49	−1.47	−4.96	−5.00	−0.04
Methanol	−3.69	−1.41	−5.10	−5.10	0.00
propan-1-ol	−3.34	−1.48	−4.82	−4.85	−0.03
propan-2-ol	−3.26	−1.36	−4.62	−4.74	−0.12
pentan-1-ol	−3.36	−1.61	−4.97	−4.57	0.40
2-methylpropan-2-ol	−3.10	−1.27	−4.37	−4.47	−0.10
2-methylbutan-2-ol	−2.95	−1.17	−4.12	−4.43	−0.31
2-methylpropan-1-ol	−3.20	−1.50	−4.70	−4.50	0.20
butan-2-ol	−3.09	−1.32	−4.40	−4.62	−0.22
Cyclopentanol	−3.20	−1.68	−4.88	−5.49	−0.61
4-methylpentan-2-ol	−2.65	−1.05	−3.69	−3.73	−0.04
Cyclohexanol	−3.21	−1.92	−5.13	−5.46	−0.33
hexan-1-ol	−3.43	−1.53	−4.96	−4.40	0.56
heptan-1-ol	−3.48	−1.62	−5.09	−4.21	0.88
2-methylbutan-1-ol	−3.27	−1.29	−4.56	−4.42	0.14
cycloheptanol	−3.07	−1.89	−4.96	−5.48	−0.52
2-methylpentan-3-ol	−2.86	−0.93	−3.78	−3.88	−0.10
pentan-3-ol	−3.01	−1.08	−4.10	−4.35	−0.25
4-heptanol	−2.90	−1.10	−3.99	−4.01	−0.02
2-methylpentan-2-ol	−2.93	−1.08	−4.00	−3.92	0.08
2,3-dimethyl-2-butanol	−2.89	−0.93	−3.82	−3.91	−0.09
hexan-3-ol	−3.04	−1.27	−4.31	−4.06	0.25
pentan-2-ol	−3.10	−1.23	−4.33	−4.39	−0.06
RMS					0.33

families of solute molecules. The RMS errors of these three families are 0.36, 0.33, and 0.76 kcal/mol, respectively.

From the results listed in Tables IV–VI, we note that for ether molecules, all the non-polar energies are positive which neutralizes some polar contributions to the total solvation free energies. For the alcohol molecules, the non-polar energies are all negative, which enhance the contributions of the polar contributions to the total solvation free energies. The attractive van der Waals interactions between alcohol molecules and water solvent must be very strong. Physically, there are strong solvent-solute hydrogen bonds that make alcohol molecules easily solvated. These solvent-solute interactions are described by the strong attractive van der Waals interactions in the present model. As for the phenol molecules, there is a mixed pattern for the non-polar contributions.

TABLE VI. The solvation free energy prediction for the phenol set. All energies are in the unit of kcal/mol.

Name	ΔG^{P}	G^{NP}	ΔG	$\Delta G^{\mathrm{Exp50}}$	Error
3-hydroxybenzaldehyde	−9.17	0.39	−8.78	−9.52	−0.74
4-hydroxybenzaldehyde	−9.60	0.19	−9.41	−8.83	0.58
o-cresol	−5.32	−1.04	−6.36	−5.90	0.46
m-cresol	−5.71	−0.86	−6.57	−5.49	1.08
Phenol	−5.81	−0.14	−6.95	−6.61	0.34
p-cresol	−5.80	−1.05	−6.85	−6.13	0.72
naphthalen-1-ol	−5.50	−0.75	−6.25	−7.67	−1.42
3,4-dimethylphenol	−5.72	−0.49	−6.21	−6.50	−0.29
2,5-dimethylphenol	−5.34	−0.48	−5.82	−5.91	−0.09
4-tert-butylphenol	−5.55	0.86	−4.69	−5.91	−1.22
2,4-dimethylphenol	−5.55	−1.03	−6.58	−6.01	0.57
3,5-dimethylphenol	−5.69	−0.41	−6.10	−6.27	−0.17
naphthalen-2-ol	−5.85	−0.72	−6.57	−8.11	−1.54
2,3-dimethylphenol	−5.47	−1.13	−6.60	−6.16	0.44
2,6-dimethylphenol	−5.07	−1.07	−6.14	−5.26	0.88
3-ethylphenol	−5.67	−0.37	−6.04	−6.25	−0.21
4-propylphenol	−5.79	−0.05	−5.84	−5.21	0.63
4-ethylphenol	−5.76	−0.48	−6.24	−6.13	0.11
RMS					0.76

TABLE VII. The partition of the molecules into sub-groups.

Molecule	Group 1	Group 2	Group 3	Group 4	Group 5
Alkane	8	8	8	7	7
Alkene	5	5	5	4	4
Ether	4	4	3	3	3
Alcohol	5	5	5	5	5
Phenol	4	4	4	3	3

The above study of a large variety of molecules indicates that our DG based solvation model together with the proposed parameter optimization algorithms can provide very accurate predictions of solvation free energies for both polar and non-polar solute molecules.

C. Fivefold cross validation

Having verified that our DG based solvation model with the optimized parameters provides very good regression results, we perform a fivefold cross validation to further illustrate the predictive power of the present method for independent data sets. Specifically, the parameters learned from a group of molecules can be employed for the blind prediction of other molecules.

To perform the fivefold cross validation, each type of molecules is subdivided into five sub-groups as uniformly as possible, and Table VII lists the number of molecules in each sub-group for each type of molecules. In our parameters optimization, we leave out one sub-group of molecules and use the rest of molecules to establish our DG based solvation model. The optimized parameters are then employed for the blind prediction of solvation free energies of the left out sub-group of molecules.

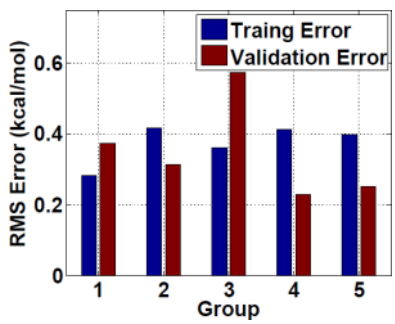

FIG. 8. The bar plot of the training and validation errors of alkanes.

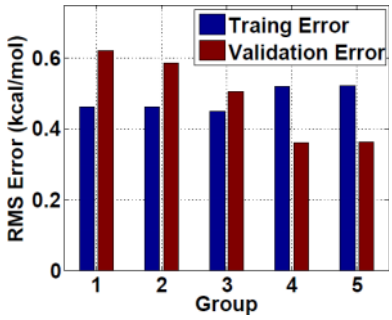

FIG. 9. The bar plot of the training and validation errors of alkenes.

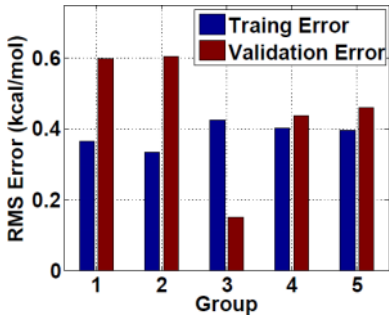

FIG. 10. The bar plot of the training and validation errors of the ethers.

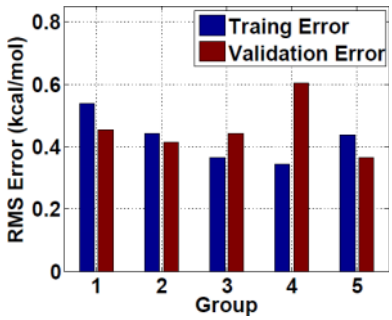

FIG. 11. The bar plot of the training and validation errors of alcohols.

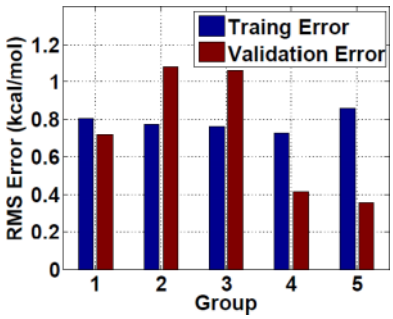

FIG. 12. The bar plot of the training and validation errors of phenols.

Figures 8–12 demonstrate the cross validation results of the alkane, alkene, ether, alcohol, and phenol molecules, respectively. It is seen that training and validation errors are similar to each other, which verifies the ability of our model in the blind prediction of solvation free energies.

In the real prediction of the solvation free energy for a given molecule of unknown category, we can first assign it to a given group, and then employ the DG based solvation model with the optimal parameters learned for this specific group for a blind prediction.

V. CONCLUSION

DG based solvation models have had a considerable success in solvation analysis.[10,11,13,70] Particularly, our DG based non-polar solvation model was shown to offer some of the most accurate solvation energy predictions of various non-polar molecules.[16] However, our DG based full solvation model is subject to numerical instability in solving the GLB equation, due to its coupling with the GPB equation. To stabilize the coupled GLB and GPB equations, a strong constraint on the van der Waals interaction was applied in our earlier work,[10,11,13] which hinders the parameter optimization of our DG based solvation model. In the present work, we resolve this problem by introducing new parameter optimization algorithms, namely, perturbation method and convex optimization, for the DG based solvation model. New stability conditions are explicitly imposed to the parameter selection, which guarantees the stability and robustness of solving the GLB equation and leads to constrained optimization of our DG based solvation model. The new optimization algorithms are intensively validated by using a large number of test molecules, including the SAMPL0 test set,[53] alkane, alkene, ether, alcohol, and phenol types of solutes. Regression results based on our new algorithms are consistent extremely well with experimental data. Additionally, a fivefold cross validation technique is employed to explore the ability of our DG based solvation models for the blind prediction of the solvation free energies for a variety of solute molecules. It is found that the same level of errors is found in the training and validation sets, which confirms our model's predictive power in solvation free energy analysis. The present DG based full solvation model provides a unified framework for analyzing both polar and nonpolar molecules. In our future work, we will develop machine learning approaches for the robust clustering of solute molecules of interest into appropriate categories so as to better predict their solvation free energies.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant Nos. IIS-1302285 and DMS-1160352, NIH Grant No. R01GM-090208, and MSU Center for Mathematical Molecular Biosciences Initiative. G.W.W. thanks Nathan Baker for valuable comments.

[1]Baker, N. A., "Improving implicit solvent simulations: A Poisson-centric view," Curr. Opin. Struct. Biol. **15**(2), 137–143 (2005).
[2]Bashford, D. and Case, D. A., "Generalized Born models of macromolecular solvation effects," Annu. Rev. Phys. Chem. **51**, 129–152 (2000).
[3]Bates, P. W., Chen, Z., Sun, Y. H., Wei, G. W., and Zhao, S., "Geometric and potential driving formation and evolution of biomolecular surfaces," J. Math. Biol. **59**, 193–231 (2009).
[4]Bates, P. W., Wei, G. W., and Zhao, S., "The minimal molecular surface," e-print arXiv:q-bio/0610038v1 [q-bio.BM] (2006).
[5]Bates, P. W., Wei, G. W., and Zhao, S., "The minimal molecular surface," in Midwest Quantitative Biology Conference, Mission Point Resort, Mackinac Island, MI, 29 September–1 October 2006.
[6]Bates, P. W., Wei, G. W., and Zhao, S., "Minimal molecular surfaces and their applications," J. Comput. Chem. **29**(3), 380–391 (2008).

[7]Beglov, D. and Roux, B., "Solvation of complex molecules in a polar liquid: An integral equation theory," J. Chem. Phys. **104**(21), 8678–8689 (1996).
[8]Bergstrom, C. A. S., Strafford, M., Lazorova, L., Avdeef, A., Luthman, K., and Artursson, P., "Absorption classification of oral drugs based on molecular surface properties," J. Med. Chem. **46**(4), 558–570 (2003).
[9]Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., III, T. E. C., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R., Wu, X., York, D. M., and Kollman, P. A., *Amber 2015* (University of California, San Francisco, 2015).
[10]Chen, Z., Baker, N. A., and Wei, G. W., "Differential geometry based solvation models I: Eulerian formulation," J. Comput. Phys. **229**, 8231–8258 (2010).
[11]Chen, Z., Baker, N. A., and Wei, G. W., "Differential geometry based solvation models II: Lagrangian formulation," J. Math. Biol. **63**, 1139–1200 (2011).
[12]Chen, D., Chen, Z., and Wei, G. W., "Quantum dynamics in continuum for proton transport II: Variational solvent-solute interface," Int. J. Numer. Methods Biomed. Eng. **28**, 25–51 (2012).
[13]Chen, Z. and Wei, G. W., "Differential geometry based solvation models III: Quantum formulation," J. Chem. Phys. **135**, 194108 (2011).
[14]Chen, D. and Wei, G. W., "Quantum dynamics in continuum for proton transport—Generalized correlation," J. Chem. Phys. **136**, 134109 (2012).
[15]Chen, D., Wei, G. W., Cong, X., and Wang, G., "Computational methods for optical molecular imaging," Commun. Numer. Methods Eng. **25**, 1137–1161 (2009).
[16]Chen, Z., Zhao, S., Chun, J., Thomas, D. G., Baker, N. A., Bates, P. B., and Wei, G. W., "Variational approach for non-polar solvation analysis," J. Chem. Phys. **137**, 084101 (2012).
[17]Cheng, L. T., Dzubiella, J., McCammon, A. J., and Li, B., "Application of the level-set method to the implicit solvation of non-polar molecules," J. Chem. Phys. **127**(8), 084503 (2007).
[18]Connolly, M. L., "Analytical molecular surface calculation," J. Appl. Crystallogr. **16**(5), 548–558 (1983).
[19]Connolly, M. L., "Depth buffer algorithms for molecular modeling," J. Mol. Graphics **3**, 19–24 (1985).
[20]Crowley, P. B. and Golovin, A., "Cation-pi interactions in protein–protein interfaces," Proteins: Struct., Funct., Bioinf. **59**(2), 231–239 (2005).
[21]Daily, M., Chun, J., Heredia-Langner, A., Wei, G. W., and Baker, N. A., "Origin of parameter degeneracy and molecular shape relationships in geometric-flow calculations of solvation free energies," J. Chem. Phys. **139**, 204108 (2013).
[22]David, L., Luo, R., and Gilson, M. K., "Comparison of generalized Born and Poisson models: Energetics and dynamics of HIV protease," J. Comput. Chem. **21**(4), 295–309 (2000).
[23]Davis, M. E. and McCammon, J. A., "Electrostatics in biomolecular structure and dynamics," Chem. Rev. **94**, 509–521 (1990).
[24]Dominy, B. N. and Brooks, III, C. L., "Development of a generalized Born model parameterization for proteins and nucleic acids," J. Phys. Chem. B **103**(18), 3765–3773 (1999).
[25]Dong, F., Vijaykumar, M., and Zhou, H. X., "Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar," Biophys. J. **85**(1), 49–60 (2003).
[26]Dong, F. and Zhou, H. X., "Electrostatic contribution to the binding stability of protein–protein complexes," Proteins **65**(1), 87–102 (2006).
[27]Dragan, A. I., Read, C. M., Makeyeva, E. N., Milgotina, E. I., Churchill, M. E., Crane-Robinson, C., and Privalov, P. L., "DNA binding and bending by HMG boxes: Energetic determinants of specificity," J. Mol. Biol. **343**(2), 371–393 (2004).
[28]Dzubiella, J., Swanson, J. M. J., and McCammon, J. A., "Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models," Phys. Rev. Lett. **96**, 087802 (2006).
[29]Fogolari, F., Brigo, A., and Molinari, H., "The Poisson–Boltzmann equation for biomolecular electrostatics: A tool for structural biology," J. Mol. Recognit. **15**(6), 377–392 (2002).
[30]Gallicchio, E., Kubo, M. M., and Levy, R. M., "Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation," J. Phys. Chem. B **104**(26), 6271–6285 (2000).
[31]Gallicchio, E., Zhang, L. Y., and Levy, R. M., "The SGB/NP hydration free energy model based on the surface generalized Born solvent reaction field and novel non-polar hydration free energy estimators," J. Comput. Chem. **23**(5), 517–529 (2002).
[32]Geng, W., Yu, S., and Wei, G. W., "Treatment of charge singularities in implicit solvent models," J. Chem. Phys. **127**, 114106 (2007).
[33]Gilson, M. K., Davis, M. E., Luty, B. A., and McCammon, J. A., "Computation of electrostatic forces on solvated molecules using the Poisson–Boltzmann equation," J. Phys. Chem. **97**(14), 3591–3600 (1993).
[34]Grant, M. and Boyd, S., "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by Blondel, V., Boyd, S., and Kimura, H. (Springer-Verlag Limited, 2008), pp. 95–110.
[35]Grant, M. and Boyd, S., CVX: Matlab software for disciplined convex programming, version 2.1, 2014, http://cvxr.com/cvx.
[36]Grant, J. and Pickup, B., "A Gaussian description of molecular shape," J. Phys. Chem. **99**, 3503–3510 (1995).
[37]Grant, J. A., Pickup, B. T., and Nicholls, A., "A smooth permittivity function for Poisson–Boltzmann solvation methods," J. Comput. Chem. **22**(6), 608–640 (2001).
[38]Grant, J. A., Pickup, B. T., Sykes, M. T., Kitchen, C. A., and Nicholls, A., "The Gaussian generalized born model: Application to small molecules," Phys. Chem. Chem. Phys. **9**, 4913–4922 (2007).
[39]Hastie, T., Tibshirani, R., and Friedman, J., "The elements of statistical learning: Data mining, inference, and prediction," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, 2009).
[40]Holm, C., Kekicheff, P., and Podgornik, R., *Electrostatic Effects in Soft Matter and Biophysics*, Nato Science Series (Kluwer Academic Publishers, Boston, 2001).
[41]Honig, B. and Nicholls, A., "Classical electrostatics in biology and chemistry," Science **268**(5214), 1144–1149 (1995).
[42]Jackson, R. M. and Sternberg, M. J., "A continuum model for protein–protein interactions: Application to the docking problem," J. Mol. Biol. **250**(2), 258–275 (1995).
[43]Jinnouchi, R. and Anderson, A. B., "Electronic structure calculations of liquid–solid interfaces: Combination of density functional theory and modified Poisson–Boltzmann theory," Phys. Rev. B **77**, 245417 (2008).
[44]Koehl, P., "Electrostatics calculations: Latest methodological advances," Curr. Opin. Struct. Biol. **16**(2), 142–151 (2006).
[45]Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D., and Tainer, J. A., "The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures," J. Mol. Biol. **228**(1), 13–22 (1992).
[46]Lamm, G., "The Poisson–Boltzmann equation," in *Reviews in Computational Chemistry*, edited by Lipkowitz, K. B., Larter, R., and Cundari, T. R. (John Wiley and Sons, Inc., Hoboken, NJ, 2003), pp. 147–366.
[47]Lee, B. and Richards, F. M., "The interpretation of protein structures: Estimation of static accessibility," J. Mol. Biol. **55**(3), 379–400 (1971).
[48]Licata, V. J. and Allewell, N. M., "Functionally linked hydration changes in escherichia coli aspartate transcarbamylase and its catalytic subunit," Biochemistry **36**(33), 10161–10167 (1997).
[49]Livingstone, J. R., Spolar, R. S., and Record, Jr., M. T., "Contribution to the thermodynamics of protein folding from the reduction in water-accessible non-polar surface area," Biochemistry **30**(17), 4237–4244 (1991).
[50]Mobley, D. L. and Guthrie, J. P., "Freesolv: A database of experimental and calculated hydration free energies, with input files," J. Comput.-Aided Mol. Des. **28**, 711–720 (2014).
[51]Mongan, J., Simmerling, C., McCammon, J. A., Case, D. A., and Onufriev, A., "Generalized Born model with a simple, robust molecular volume correction," J. Chem. Theory Comput. **3**(1), 159–169 (2007).
[52]Netz, R. R. and Orland, H., "Beyond Poisson–Boltzmann: Fluctuation effects and correlation functions," Eur. Phys. J. A **1**(2-3), 203–214 (2000).
[53]Nicholls, A., Mobley, D. L., Guthrie, P. J., Chodera, J. D., and Pande, V. S., "Predicting small-molecule solvation free energies: An informal blind test for computational chemistry," J. Med. Chem. **51**(4), 769–779 (2008).
[54]Nina, M., Im, W., and Roux, B., "Optimized atomic radii for protein continuum electrostatics solvation forces," Biophys. Chem. **78**(1-2), 89–96 (1999).
[55]Onufriev, A., Bashford, D., and Case, D. A., "Modification of the generalized Born model suitable for macromolecules," J. Phys. Chem. B **104**(15), 3712–3720 (2000).
[56]Onufriev, A., Case, D. A., and Bashford, D., "Effective Born radii in the generalized Born approximation: The importance of being perfect," J. Comput. Chem. **23**(14), 1297–1304 (2002).
[57]Ratkova, E. L., Chuev, G. N., Sergiievskyi, V. P., and Fedorov, M. V., "An accurate prediction of hydration free energies by combination of molecular

integral equations theory with structural descriptors," J. Phys. Chem. B **114**(37), 12068–12079 (2010).
[58]Richards, F. M., "Areas, volumes, packing, and protein structure," Annu. Rev. Biophys. Bioeng. **6**(1), 151–176 (1977).
[59]Roux, B. and Simonson, T., "Implicit solvent models," Biophys. Chem. **78**(1-2), 1–20 (1999).
[60]Sanner, M. F., Olson, A. J., and Spehner, J. C., "Reduced surface: An efficient way to compute molecular surfaces," Biopolymers **38**, 305–320 (1996).
[61]Sharp, K. A. and Honig, B., "Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation," J. Phys. Chem. **94**, 7684–7692 (1990).
[62]Sharp, K. A. and Honig, B., "Electrostatic interactions in macromolecules - theory and applications," Annu. Rev. Biophys. Biophys. Chem. **19**, 301–332 (1990).
[63]Spolar, R. S., Ha, J. H., and Record, Jr., M. T., "Hydrophobic effect in protein folding and other noncovalent processes involving proteins," Proc. Natl. Acad. Sci. U. S. A. **86**(21), 8382–8385 (1989).
[64]Swanson, J. M. J., Mongan, J., and McCammon, J. A., "Limitations of atom-centered dielectric functions in implicit solvent models," J. Phys. Chem. B **109**(31), 14769–14772 (2005).
[65]Thomas, D., Chun, J., Chen, Z., Wei, G. W., and Baker, N. A., "Parameterization of a geometric flow implicit solvation model," J. Comput. Chem. **24**, 687–695 (2013).
[66]Tjong, H. and Zhou, H. X., "GBr6NL: A generalized Born method for accurately reproducing solvation energy of the nonlinear Poisson–Boltzmann equation," J. Chem. Phys. **126**, 195102 (2007).
[67]Tomasi, J., Mennucci, B., and Cammi, R., "Quantum mechanical continuum solvation models," Chem. Rev. **105**, 2999–3093 (2005).
[68]Tsui, V. and Case, D. A., "Molecular dynamics simulations of nucleic acids with a generalized Born solvation model," J. Am. Chem. Soc. **122**(11), 2489–2498 (2000).
[69]Wagoner, J. A. and Baker, N. A., "Assessing implicit models for non-polar mean solvation forces: The importance of dispersion and volume terms," Proc. Natl. Acad. Sci. U. S. A. **103**(22), 8331–8336 (2006).
[70]Wei, G. W., "Generalized Perona-Malik equation for image restoration," IEEE Signal Process. Lett. **6**, 165–167 (1999).
[71]Wei, G. W., "Differential geometry based multiscale models," Bull. Math. Biol. **72**, 1562–1622 (2010).
[72]Wei, G.-W., "Multiscale, multiphysics and multidomain models I: Basic theory," J. Theor. Comput. Chem. **12**(8), 1341006 (2013).
[73]Wei, G. W., Sun, Y. H., Zhou, Y. C., and Feig, M., "Molecular multiresolution surfaces," e-print arXiv:math-ph/0511001v1, 1–11 (2005).
[74]Wei, G.-W., Zheng, Q., Chen, Z., and Xia, K., "Variational multiscale models for charge transport," SIAM Rev. **54**(4), 699–754 (2012).
[75]Willmore, T. J., *Riemannian Geometry* (Oxford University Press, USA, 1997).
[76]Yu, Z. Y. and Bajaj, C., "Computational approaches for automatic structural analysis of large biomolecular complexes," IEEE/ACM Trans. Comput. Biol. Bioinf. **5**, 568–582 (2008).
[77]Yu, S. N., Geng, W. H., and Wei, G. W., "Treatment of geometric singularities in implicit solvent models," J. Chem. Phys. **126**, 244108 (2007).
[78]Yu, S. N. and Wei, G. W., "Three-dimensional matched interface and boundary (MIB) method for treating geometric singularities," J. Comput. Phys. **227**, 602–632 (2007).
[79]Yu, S. N., Zhou, Y. C., and Wei, G. W., "Matched interface and boundary (MIB) method for elliptic problems with sharp-edged interfaces," J. Comput. Phys. **224**(2), 729–756 (2007).
[80]Zhao, S., "Pseudo-time-coupled nonlinear models for biomolecular surface representation and solvation analysis," Int. J. Numer. Methods Biomed. Eng. **27**, 1964–1981 (2011).
[81]Zhao, S., "Operator splitting adi schemes for pseudo-time coupled nonlinear solvation simulations," J. Comput. Phys. **257**, 1000–1021 (2014).
[82]Zhao, S. and Wei, G. W., "High-order FDTD methods via derivative matching for Maxwell's equations with material interfaces," J. Comput. Phys. **200**(1), 60–103 (2004).
[83]Zhou, S. G., Cheng, L. T., Sun, H., Che, J. W., Dzubiella, J., Li, B., and McCammon, J. A., "Ls-vism: A software package for analysis of biomolecular solvation," J. Comput. Chem. **36**, 1047–1059 (2015).
[84]Zhou, Y. C., Feig, M., and Wei, G. W., "Highly accurate biomolecular electrostatics in continuum dielectric environments," J. Comput. Chem. **29**, 87–97 (2008).
[85]Zhou, Y. C. and Wei, G. W., "On the fictitious-domain and interpolation formulations of the matched interface and boundary (MIB) method," J. Comput. Phys. **219**(1), 228–246 (2006).
[86]Zhou, Y. C., Zhao, S., Feig, M., and Wei, G. W., "High order matched interface and boundary method for elliptic equations with discontinuous coefficients and singular sources," J. Comput. Phys. **213**(1), 1–30 (2006).
[87]Zhu, J., Alexov, E., and Honig, B., "Comparative study of generalized Born models: Born radii and peptide folding," J. Phys. Chem. B **109**(7), 3008–3022 (2005).