Journal of Computational Physics 464 (2022) 111340

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

A regularization approach for solving the super-Gaussian Poisson-Boltzmann model with heterogeneous dielectric functions

Siwen Wang [a], Yuanzhen Shao [a], Emil Alexov [b], Shan Zhao [a,*]

[a] Department of Mathematics, University of Alabama, Tuscaloosa, AL 35487, USA
[b] Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 18 August 2021
Received in revised form 25 May 2022
Accepted 25 May 2022
Available online 31 May 2022

Keywords:
Poisson-Boltzmann equation
Gaussian dielectric model
Singular charge source
Regularization
Finite-difference method
Electrostatic free energy

ABSTRACT

A regularization method is introduced for the first time in the literature for treating charge singularities in the heterogeneous dielectric Poisson-Boltzmann (PB) model. In the existing Gaussian and super-Gaussian electrostatic analysis, the singular sources in terms of Dirac delta functions are approximated by trilinear interpolation, which introduces a large error in potential solutions, and has to rely on an error cancellation for delivering acceptable estimates of the electrostatic free energy. To overcome this difficulty, a dual decomposition of potential and dielectric function is carried out in the proposed regularization, so that the charge singularities can be analytically captured by the Coulomb potential, while the reaction field potential satisfies a regularized PB equation with a new source term. A rigorous analysis has been conducted to show that a super-Gaussian density, instead of a Gaussian one, is required to guarantee a small neighborhood around each atom center with a nearly homogeneous dielectric medium, so that the source term can be well defined in the sense of distribution. Moreover, the well-posedness of the regularized formulation has been proved, and the regularity of the weak solution has been clarified. In the numerical implementation, special considerations are given in calculating the source term, by using analytical differentiations, instead of numerical approximations, whenever possible. The accuracy, convergence, efficiency, and robustness of the proposed regularization are numerically verified via benchmark studies. It is found that the regularization is more accurate than the trilinear method, as well as produces a faster convergence in energy estimation. Moreover, the grid artifact or artificial grid energy is completely eliminated in the present finite difference PB model.

1. Introduction

In physiological studies, most biomolecular processes occur in water, which takes more than half of the cell mass. When these biological processes being analyzed, an unavoidable prerequisite for quantitative analysis is understanding the solvation between the macromolecule and its surrounding aqueous environment. Among all forces and energies involved in the solvation analysis, the electrostatic force is a long-range force dominating all other forces at long distances. Thus, the elec-

* Corresponding author.
E-mail address: szhao@ua.edu (S. Zhao).

https://doi.org/10.1016/j.jcp.2022.111340
0021-9991/

trostatic analysis is indispensable for solvated biological processes, by studying interactions between charged solutes such as proteins, DNAs and RNAs, and mobile ions contained in the solvent.

As an implicit solvent approach, the Poisson-Boltzmann (PB) model [5,25,38] treats the solute and solvent as dielectric continuum, and assumes that the mobile ions follow the Boltzmann distribution in water. Such a mean field approach leads to a nonlinear elliptic partial differential equation (PDE), whose source term involves partial charges carried by the macromolecule located at atom centers. As a standard method for electrostatic analysis, the PB model has been widely used for studying energies and dynamics of various biomolecular processes [5]. Recently, the PB model has been applied to study the conformational changes of the SARS-CoV-2 spike protein in [28].

In the classical PB model [5,25,38], a sharp interface is assumed at the solute-solvent boundary, which splits the domain into two subdomains, i.e., the solute region with a low dielectric constant and the water region with a high dielectric constant. Across the dielectric interface which is usually modeled by a molecular surface, the electrostatic potential and its flux are continuous, while the normal derivative of the potential is discontinuous. To handle the dielectric jump, advance numerical algorithms have to be designed for solving sharp interface PB equation [3,11,15,19,20,32]. Moreover, the commonly used molecular surface models are known to admit geometric singularities, such as cusps and self-intersecting surfaces [6].

There exist many improved PB models in the literature [1,6,7,10,13,16,22,30,31,36,37,44,48], that feature a smooth solute-solvent boundary, i.e., the dielectric function varies smoothly from the macromolecule to the solvent over a narrow band. From modeling point of view, a smooth solute-solvent boundary is more physical than a sharp interface by taking into account the definition of dielectric coefficient and the atomistic nature of the biomolecular system [22]. Basically, the dielectric coefficient of water molecules or dipoles in the solute-solvent boundary is determined by the polarizability of the dipole in responding to the electrostatic field. Due to local interactions with the macromolecule, such as Van der Waals forces, water molecules close to the macromolecule tend to have a lower polarizability. Moreover, such a polarizability will increase gradually, when one moves away from the macromolecule, instead of undergoing a sharp jump. This is the main justification for developing the so-called diffuse interface PB models [1,6,7,13,16,37,44,48], in which the solute and solvent are still assumed to be two homogeneous dielectric media, away from the smooth solute-solvent boundary.

Besides the diffuse interface models, various heterogeneous dielectric PB models [10,22,26,30,31,35,36,39,41] have been developed to represent the dielectric distribution of the solute as a space dependent function. This paper will pay particular attention to Gaussian [30,31] and super-Gaussian dielectric distributions [22], which were introduced to mimic the effect of random conformational changes of the macromolecule on the solvation free energy. This is realized by defining Gaussian or super-Gaussian density functions to describe atom specific heterogeneity. Compared with the classical PB model, the Gaussian PB model produces more accurate solvation free energies for small molecules [30,31], and better pKa predictions for proteins, RNAs and DNAs [42,43]. Moreover, an ensemble averaged electrostatic free energy could be captured by using a single protein structure in Gaussian and super-Gaussian models [10,36], which is much more efficient than the usual ensemble calculation involving thousands of steps of PB computations, together with molecular dynamics or Monte Carlo simulations.

Sharing some similarities with the Gaussian PB model [30,31], the super-Gaussian PB model [22] defines a diffuse interface function to explicitly characterize the molecular subdomain, solute-solvent boundary, and water subdomain. While the dielectric function of the Gaussian model is C^0 but not C^1 continuous in the vacuum state, it is at least C^2 continuous for both water and vacuum states in the super-Gaussian model. Moreover, by treating the maximum dielectric value of the macromolecule as an adjustable parameter, the super-Gaussian formulation is able to model the dielectric property of interior cavities and ion-channels [22]. As to be shown in this work, in order to generalize the regularization formulation to heterogeneous dielectric PB models, a super-Gaussian density, instead of Gaussian one, is required.

The main focus of this study is on treating point charges carried by the macromolecule in the super-Gaussian PB model. As in the other PB models, these partial charges are expressed as Dirac delta distributions, and their summation forms a singular source term for the PB equation [5,25,38]. In traditional numerical studies, the singular charges are directly discretized in grid based computations. For example, a trilinear method is commonly used in finite difference discretization, in which the singular charges are distributed to the neighboring grid nodes with finite values [33]. In fact, all existing simulations of Gaussian and super-Gaussian PB models [22,30,31] are realized via the trilinear method. In a finite element variational form, a point charge can be evaluated through the trial function, by means of the definition of the delta function [17]. Nevertheless, when singular charges are directly discretized, one actually attempts to approximate an unbounded potential solution by finite numerical values, which is doomed to be inaccurate. Therefore, the singular charge source is known to be a significant challenge for both mathematical analysis and numerical computation of the PB equation [12,24,46].

For sharp interface PB model, various regularization methods [8,9,12,14,19,20,23,27,46,49] have been developed, which provide a perfect treatment to the singular charge source of the PB equation. In these regularization methods, the potential is split into two or three components. The key of the perfect treatment is to analytically capture the singularities of the potential by using one component, which satisfies a Poisson's equation with the same singular source. Mathematically, this singular component or Coulomb potential can be expressed in terms of Green's functions, which blow up at atom centers, but can be analytically computed at other locations. After removing the singular component, other potential components become bounded so that their numerical discretization and mathematical analysis become manageable. Nevertheless, an accuracy reduction issue has been noticed in Ref. [23], i.e., while these regularization formulations are consistent with the same PB equation, some methods could be significantly less accurate than the others. To understand this discrepancy, four

popular regularizations were compared by using the same discretization in [29]. Through tracking the source of error and conducting numerical analysis correspondingly, an accuracy recovery technique has been introduced in [29] so that all four methods yield the same high precision.

However, the feasibility for generalizing the regularization method to diffuse interface PB models [1,6,7,13,16,48] is unclear until our recent works [44,45]. This is due to the fact that the fundamental solution to a Poisson's equation with the PB singular source, but involving a space dependent dielectric function, is not analytically attainable. To alleviate this issue, a semi-analytical approach has been introduced in Ref. [47] to approximate the fundamental solution, which unfortunately is applicable only to simple geometric shapes. Recently, the first regularization formulation for the diffuse interface PB model was reported in [44,45], in which the breakthrough is realized via a dual decomposition for both potential and dielectric functions, i.e., besides the usual decomposition of the potential into Coulomb and reaction-field components, the dielectric function is also split into a constant base and a space varying part. By requiring the Coulomb component satisfying a Poisson's equation with the constant dielectric base, the above mentioned difficulty is simply bypassed and the singularities can be captured again by the usual Green's functions. By combining other components into one equation, the reaction-field potential satisfies a PB equation with a new smooth source. This new regularization for the diffuse interface PB model has been benchmarked by comparing with the trilinear method [33] and semi-analytical method [47], and further validated by calculating energies for real proteins [44,45]. More recently, a rigorous analysis has been conducted in [37] to show that the regularization formulation developed in [44,45] is well-posed. Moreover, when the underlying diffuse interface approaches to the sharp interface, the regularization of the diffuse interface PB model converges to that of the sharp interface PB model, in terms of both the reaction-field potential and electrostatic free energy.

For the first time in the literature, the regularization formulation will be introduced in the super-Gaussian PB model to treat singular charges in this work. The dual decomposition strategy introduced in the diffuse interface PB model [44,45] will again play a central role for the new development. However, there are two major difficulties that need to be addressed in the present study, due to a heterogeneous dielectric function inside the protein region. First, in the diffuse interface PB model, the shifted dielectric function is identically zero around atom centers so that the source term can be well defined in a regular sense [44,45]. However, in the super-Gaussian PB model, the shifted dielectric function is just nearly vanishing, not identically equals to zero, in a small neighborhood around each atom center. Thus, a rigorous analysis has to be carried out to examine the regularities of the source term and reaction-field potential at charge centers. Moreover, the requirement on the dielectric density function will be investigated so that the source term can be defined in the distributional sense. The well-posedness of the proposed regularization will also be proved. Second, the numerical computation of the new source term deserves further attention, because it involves a space dependent dielectric function. Analytical differentiations, instead of numerical approximations, will be adopted whenever possible, so that the proposed regularization could be accurately implemented.

In the present implementation of the super-Gaussian PB model [22], we will replace the minimal molecular surface (MMS) [6,7] by the Gaussian convolution surface (GCS) introduced in [44] to define a surface function for characterizing the smooth solute-solvent boundary. Both MMS and GCS produce smooth diffuse interfaces, and can robustly handle various protein geometries. Nevertheless, based on the fast Fourier transform (FFT) algorithm, the GCS is much more efficient than the MMS in surface generation. Moreover, the control of the smoothing solute-solvent boundary, including its width and decay rate of surface function, becomes easier in case of the GCS.

The rest of the paper is structured with the following sections. Section 2 first revisits the super-Gaussian PB model. Then, a new regularization approach is proposed, along with a detailed analysis of the new source term. The well-posedness of the regularized PB equation for the reaction-field potential is analyzed, and numerical implementation details are offered. In Section 3, the proposed regularization method is first validated by being compared with the trilinear method for solving a manufactured example with the analytical potential solution. Biological application to real molecules and proteins will be considered. In Section 4, a brief discussion is offered. Finally, in Appendix B, the construction of the analytical solution for the manufactured example is presented.

2. Theory and algorithm

2.1. Super-Gaussian Poisson-Boltzmann model

In this subsection, the super-Gaussian Poisson-Boltzmann (PB) model proposed in [22] will be introduced. Moreover, a different surface function will be employed to characterize the subdomains.

Consider a macromolecule, for example, a protein being immersed into an aqueous solvent. Define a large enough cubic domain Ω in $\mathbb{R}^3$ for this three dimensional (3D) solute-solvent system. In the super-Gaussian PB model, the domain Ω consists of three regions: an interior domain Ω_i for the solute, an exterior domain Ω_e for the solvent, and a transition layer Ω_t in between Ω_i and Ω_e as a smooth solute-solvent boundary. Define the interface between Ω_i and Ω_t as Γ_i, while that between Ω_t and Ω_e as Γ_e. An illustration of subdomains is shown in Fig. 1 (a). The subdomains can be characterized by a diffuse interface or smooth surface function $S(\mathbf{r})$ for $\mathbf{r} \in \Omega$, which equals to one and zero, respectively, in Ω_i and Ω_e. In Ω_t, as $\mathbf{r}$ travels from the interior protein to the exterior solvent, $S(\mathbf{r})$ decays from one to zero, so that $S(\mathbf{r})$ is at least a C^2 continuous function over the entire domain Ω. See Fig. 1 (b).

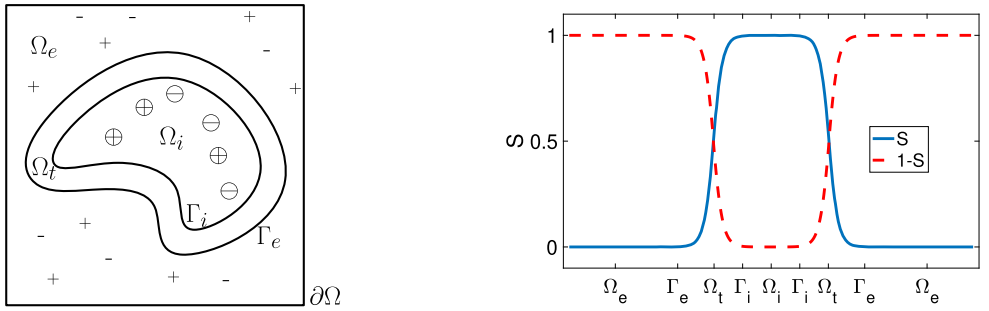

Fig. 1. (a) The subdomain setting used in the super-Gaussian PB model. (b) The subdomains are characterized by a diffuse interface function S, which is plotted along a straight line.

The minimal molecular surface (MMS) [6,7] is employed in the original super-Gaussian PB model [22] for calculating $S(\mathbf{r})$. In the present study, the Gaussian convolution surface (GCS) introduced in [44] will be employed instead. Both MMS and GCS produce smooth diffuse interfaces. Nevertheless, based on the fast Fourier transform (FFT) algorithm, the GCS is much more efficient than the MMS in surface generation. The details about the definition and generation of the MMS and GCS are offered in Appendix A.

The electrostatic interaction of this solute-solvent system is governed by the nonlinear Poisson-Boltzmann (PB) equation in the dimensionless form [22]

$$-\nabla\cdot(\epsilon(\mathbf{r})\nabla u(\mathbf{r}))+(1-S(\mathbf{r}))\kappa^2\sinh(u(\mathbf{r}))=\rho(\mathbf{r}),\quad \text{in } \Omega, \tag{1}$$

where $u(\mathbf{r})$ is the electrostatic potential and $\epsilon(\mathbf{r})$ is the dielectric function. Assume the protein contains N_m atoms, with partial charges q_j in terms of the fundamental charge e_c located at the atom centers $\mathbf{r}_j$ for $j=1,2,\ldots,N_m$. The singular source of Eq. (1), due to these partial charges, is given as

$$\rho(\mathbf{r})=4\pi\frac{{e_c}^2}{k_BT}\sum_{j=1}^{N_m}q_j\delta(\mathbf{r}-\mathbf{r}_j),\quad \text{in } \Omega, \tag{2}$$

where k_B is the Boltzmann constant and T is the temperature. In the super-Gaussian PB model, we require that $\mathbf{r}_j\in\Omega_i$ for all j. Consequently, $\rho(\mathbf{r})$ is only defined within Ω_i and $S(\mathbf{r})$ has actually been dropped in the source term, i.e., $S(\mathbf{r})\rho(\mathbf{r})=\rho(\mathbf{r})$ in Eq. (1). With $S(\mathbf{r})$, the coefficient of the PB nonlinear term $(1-S)\kappa^2$ is at least C^2 continuous, where the modified Debye-Hückel parameter κ takes a constant value $\kappa^2=\left(\frac{2N_Ae_c^2}{100k_BT}\right)I=8.486902807\ \text{Å}^{-2}I$. Here N_A is the Avogadro's Number and I is the molar ionic strength. Assume the domain Ω is large enough, so that the dielectric function is constant ϵ_{out} on the outer boundary $\partial\Omega$. Then, a Dirichlet boundary condition can be assumed

$$u(\mathbf{r})=u_b(\mathbf{r}):=\frac{{e_c}^2}{k_BT}\sum_{j=1}^{N_m}\frac{q_j}{\epsilon_{out}|\mathbf{r}-\mathbf{r}_j|}e^{-|\mathbf{r}-\mathbf{r}_j|\sqrt{\frac{\kappa^2}{\epsilon_{out}}}},\quad \text{on } \partial\Omega. \tag{3}$$

In the super-Gaussian PB model [22], the dielectric function $\epsilon(\mathbf{r})$ is defined as a heterogeneous function. In particular, each atom of the protein is regarded as a "soft sphere" with a density function. For the j^{th} atom, the density at the position $\mathbf{r}$ is given as [22]

$$g_j(\mathbf{r})=\exp\Big[-\Big(\frac{|\mathbf{r}-\mathbf{r}_j|^2}{\sigma^2R_j^2}\Big)^m\Big], \tag{4}$$

where $\mathbf{r}_j$ is the center of the j^{th} atom, R_j is the Van der Waals (VDW) radius of the j^{th} atom and σ is the relative variance. The integer order m is typically required to be $m\geq 2$ in the super-Gaussian model. When the order $m=1$, Eq. (4) is actually the density function used in the Gaussian PB model [30,31]. When the order m goes to the infinity, one attains a "hard sphere" density, i.e., $g_j=1$ inside the VDW ball and $g_j=0$ otherwise.

With the density for each atom being calculated, a total density function for all atoms is defined as

$$g(\mathbf{r})=1-\prod_{j=1}^{N_m}[1-g_j(\mathbf{r})]. \tag{5}$$

Note that Eq. (5) can describe the density for overlapped region covered by multiple atoms. For instance, a cross term $g_i g_j$ accounts for the density of the overlap region due to the i^{th} and j^{th} atoms. Moreover, in order to explicitly model the dielectric properties of protein cavities, the maximal dielectric value of the macromolecule is controlled by a parameter ϵ_{gap} in the super-Gaussian model [22], and the dielectric distribution within the protein region is defined as

$$\epsilon_{in}(\mathbf{r}) = \epsilon_m g(\mathbf{r}) + \epsilon_{gap}[1 - g(\mathbf{r})], \tag{6}$$

where the constants ϵ_m and ϵ_{gap} are the dielectric values at the atom centers and in a gap region, respectively. By combining (5) with (6), we arrive at a concise form

$$\epsilon_{in}(\mathbf{r}) = \epsilon_m + (\epsilon_{gap} - \epsilon_m) \prod_{j=1}^{N_m} [1 - g_j(\mathbf{r})]. \tag{7}$$

Since $\epsilon_{gap} > \epsilon_m$ and $g_j(\mathbf{r})$ has a range of $[0, 1]$, this means that ϵ_m and ϵ_{gap} are, respectively, the minimal and maximal dielectric values of the protein.

Making use of the surface function $S(\mathbf{r})$ to characterize the subdomains, the dielectric function in the entire domain Ω is defined as [22]

$$\epsilon(\mathbf{r}) = S(\mathbf{r})\epsilon_{in}(\mathbf{r}) + [1 - S(\mathbf{r})]\epsilon_{out}. \tag{8}$$

Inside the protein region Ω_i, we have $S(\mathbf{r}) = 1$ so that the inhomogeneity of the super Gaussian dielectric distribution is retained. In the exterior region Ω_e, we have $S(\mathbf{r}) = 0$. Then $\epsilon(\mathbf{r})$ becomes a constant ϵ_{out}, which will be taken as 1 and 80, respectively, for the vacuum state and water state. As $S(\mathbf{r})$ decays smoothly from one to zero over Ω_t, $\epsilon(\mathbf{r})$ will change from $\epsilon_{in}(\mathbf{r})$ to ϵ_{out}. Over the entire domain Ω, $\epsilon(\mathbf{r})$ is at least C^2 continuous.

2.2. A novel regularization for the super Gaussian PB model

A new regularization method is proposed in this work to analytically handle the singular source of the super-Gaussian PB model. For this purpose, a dual decomposition in potential and dielectric functions will be conducted as in the regularization formulation for the diffuse interface PB equation [44,45]

$$\begin{cases} u(\mathbf{r}) = u_C(\mathbf{r}) + u_{RF}(\mathbf{r}), & \text{for } \mathbf{r} \in \Omega, \\ \epsilon(\mathbf{r}) = \epsilon_m + \hat{\epsilon}(\mathbf{r}), & \text{for } \mathbf{r} \in \Omega. \end{cases} \tag{9}$$

In Eq. (9), the potential u is split into a Coulomb component u_C and a reaction field component u_{RF}, while the dielectric function is decomposed to be a constant base value plus space variant part $\hat{\epsilon}(\mathbf{r})$. Consider the outside medium being water with $\epsilon_{out} = 80$. Then $\epsilon(\mathbf{r})$ takes minimal values $\epsilon(\mathbf{r}) = \epsilon_m$ only at atom centers, so that $\hat{\epsilon}(\mathbf{r}) \geq 0$ for $\mathbf{r} \in \Omega$. Assume that $m \geq 2$ and an appropriate σ value such as $\sigma = 1$ is chosen in the super Gaussian density. This essentially guarantees that in a small neighborhood surrounding each atom center $\mathbf{r}_j$, $\epsilon(\mathbf{r})$ is almost ϵ_m and $\hat{\epsilon}$ is nearly vanishing, which is a crucial property for the proposed regularization. In the exterior region Ω_e, $\hat{\epsilon} = \epsilon_{out} - \epsilon_m$.

To capture the singularity in the potential function, the Coulomb potential u_C is assumed to satisfy a homogeneous Poisson's equation with the singular source ρ given by (2) [44,45]

$$\begin{cases} -\epsilon_m \Delta u_C(\mathbf{r}) = \rho(\mathbf{r}) & \text{in } R^3, \\ u_C(\mathbf{r}) = 0 & \text{as } |\mathbf{r}| \to \infty. \end{cases} \tag{10}$$

It is obvious that the singular component u_C is actually the Green's function $G(\mathbf{r})$

$$u_C(\mathbf{r}) = G(\mathbf{r}) := \frac{e_c^2}{k_B T} \sum_{j=1}^{N_m} \frac{q_j}{\epsilon_m |\mathbf{r} - \mathbf{r}_j|}. \tag{11}$$

After removing the singular component, the reaction field potential $u_{RF}(\mathbf{r})$ becomes bounded everywhere.

The PB equation (1) can be expanded according to the dual decomposition

$$-\nabla \cdot (\hat{\epsilon} \nabla u_C) - \nabla \cdot (\hat{\epsilon} \nabla u_{RF}) - \epsilon_m \Delta u_C - \epsilon_m \Delta u_{RF} + (1 - S)\kappa^2 \sinh(u_C + u_{RF}) = \rho, \quad \text{in } \Omega. \tag{12}$$

By using (10) and (12), the singular source term ρ can be canceled

$$-\nabla \cdot (\hat{\epsilon} \nabla u_C) - \nabla \cdot (\hat{\epsilon} \nabla u_{RF}) - \epsilon_m \Delta u_{RF} + (1 - S)\kappa^2 \sinh(u_C + u_{RF}) = 0, \quad \text{in } \Omega. \tag{13}$$

By re-arranging the terms, we arrive at a regularized PB equation for the reaction field component

$$-\nabla\cdot(\epsilon\nabla u_{RF})+(1-S)\kappa^2\sinh(u_{RF}+G)=\nabla\cdot(\hat{\epsilon}\nabla G),\quad \text{in } \Omega, \tag{14}$$

where we have replaced $u_C(\mathbf{r})$ by the known function $G(\mathbf{r})$, and combined two terms for $u_{RF}(\mathbf{r})$ into one on the left-hand side. The new source on the right-hand side involves the dielectric function $\hat{\epsilon}$ and the gradient of Green's function, which is known analytically

$$\nabla G(\mathbf{r})=-\frac{e_c^2}{k_BT}\sum_{j=1}^{N_m}\frac{q_j(\mathbf{r}-\mathbf{r}_j)}{\epsilon_m|\mathbf{r}-\mathbf{r}_j|^3}. \tag{15}$$

In order to analyze the new source term of Eq. (14), we first discuss the regularity of $\hat{\epsilon}\nabla G$. Throughout the domain Ω, $\hat{\epsilon}$ is a C^2 continuous function, while ∇G is smooth everywhere except at all atom centers, where it is singular. Since all atom centers are located inside Ω_i, we will examine the regularity of $\hat{\epsilon}\nabla G$ for Ω_i only. Within Ω_i, with $S=1$, we have

$$\hat{\epsilon}(\mathbf{r})=(\epsilon_{gap}-\epsilon_m)\prod_{j=1}^{N_m}[1-g_j(\mathbf{r})],\quad \text{in } \Omega_i. \tag{16}$$

We note that $\hat{\epsilon}$ is nearly vanishing in a small neighborhood around each atom center, which is different from the case in the diffuse interface PB model [44,45], in which $\hat{\epsilon}$ identically equals to zero. Thus, a rigorous mathematical analysis becomes indispensable for examining the regularity.

Proposition 2.1. *When the order $m=1$ in the super Gaussian model, as $\mathbf{r}$ approaches to the i^{th} atom center $\mathbf{r}_i$, the limit of $\hat{\epsilon}\nabla G$ does not exist.*

Proof. At the i^{th} atom center $\mathbf{r}_i$, we have $\hat{\epsilon}(\mathbf{r}_i)=0$, while $\nabla G(\mathbf{r})$ goes to infinity as $\mathbf{r}$ approaches to $\mathbf{r}_i$. In particular, out of N_m terms in the summation of $\nabla G(\mathbf{r})$, only the i^{th} term is singular at $\mathbf{r}_i$, while all other terms are bounded at $\mathbf{r}_i$. Thus, the products of $\hat{\epsilon}$ with the other N_m-1 terms are all vanishing at $\mathbf{r}_i$. The singularity could only occur for the i^{th} term of $\nabla G(\mathbf{r})$, which is denoted as

$$\nabla G_i(\mathbf{r})=-\frac{e_c^2}{k_BT}\frac{q_i(\mathbf{r}-\mathbf{r}_i)}{\epsilon_m|\mathbf{r}-\mathbf{r}_i|^3}:=C_i\frac{\mathbf{r}-\mathbf{r}_i}{|\mathbf{r}-\mathbf{r}_i|^3}, \tag{17}$$

where C_i is a constant. Similarly, we will single the contribution of i^{th} term out in $\hat{\epsilon}(\mathbf{r})$ by defining $\hat{\epsilon}_i(\mathbf{r})=(\epsilon_{gap}-\epsilon_m)\prod_{j=1,j\neq i}^{N_m}[1-g_j(\mathbf{r})]$, so that

$$\hat{\epsilon}(\mathbf{r})=\hat{\epsilon}_i(\mathbf{r})(1-g_i(\mathbf{r}))=\hat{\epsilon}_i(\mathbf{r})\left(1-\exp\Big(-\frac{|\mathbf{r}-\mathbf{r}_i|^{2m}}{\sigma^{2m}R_i^{2m}}\Big)\right),\quad \text{in } \Omega_i.$$

Note that $\hat{\epsilon}_i(\mathbf{r}_i)$ is nonzero but finite. For the ease of our analysis, a coordinate translation is assumed to be conducted so that $\mathbf{r}_i$ becomes the origin of the new coordinate system, i.e., $\mathbf{r}_i=\mathbf{0}$. Denote $\mathbf{r}=(x,y,z)^T$. For $m=1$, the x component of the interested limit is given as

$$\lim_{\mathbf{r}\to\mathbf{0}}C_i\hat{\epsilon}_i(\mathbf{r})\left(1-\exp\Big(-\frac{x^2+y^2+z^2}{\sigma^2R_i^2}\Big)\right)\frac{x}{(x^2+y^2+z^2)^{3/2}}. \tag{18}$$

Consider a limiting path with $y=nx$ and $z=kx$ for some constants n and k with $x\to 0$. The above limit becomes

$$\begin{aligned}
&C_i\hat{\epsilon}_i(\mathbf{0})\lim_{x\to0}\frac{x\left[1-\exp\Big(-\frac{x^2(1+n^2+k^2)}{\sigma^2R_i^2}\Big)\right]}{x^3(1+n^2+k^2)^{3/2}}\\
=&C_i\hat{\epsilon}_i(\mathbf{0})\lim_{x\to0}\frac{\left[1-\exp\Big(-x^2\frac{1+n^2+k^2}{\sigma^2R_i^2}\Big)\right]}{x^2(1+n^2+k^2)^{3/2}}\\
=&C_i\hat{\epsilon}_i(\mathbf{0})\lim_{x\to0}\frac{-\exp\Big(-x^2\frac{1+n^2+k^2}{\sigma^2R_i^2}\Big)\Big(-2x\frac{1+n^2+k^2}{\sigma^2R_i^2}\Big)}{2x(1+n^2+k^2)^{3/2}}\\
=&C_i\hat{\epsilon}_i(\mathbf{0})\lim_{x\to0}\frac{\exp\Big(-x^2\frac{1+n^2+k^2}{\sigma^2R_i^2}\Big)}{\sigma^2R_i^2(1+n^2+k^2)^{1/2}}\\
=&\frac{C_i\hat{\epsilon}_i(\mathbf{0})}{\sigma^2R_i^2(1+n^2+k^2)^{1/2}},
\end{aligned}$$

where the L'Hopital's Rule has been applied. Since the final results are path dependent, the limit considered in Eq. (18) does not exist. Thus, for $m = 1$, as $\mathbf{r}$ approaches to the i^{th} atom center $\mathbf{r}_i$, the limit of $\hat{\epsilon}\nabla G$ does not exist. □

We next investigate the regularity of $\hat{\epsilon}\nabla G$ and $\nabla\epsilon \cdot \nabla G$. Throughout the domain Ω, $\hat{\epsilon}$ and $\nabla\epsilon$ are smooth functions, while ∇G is smooth everywhere except at all atom centers, where it is singular. Thus, we can conclude that $\hat{\epsilon}\nabla G$ and $\nabla\epsilon \cdot \nabla G$ are smooth everywhere except at atom centers. We will focus on the atom centers in the following Theorem.

Theorem 2.2. *When the order $m \geq 2$ in the super Gaussian model, as $\mathbf{r}$ approaches to the i^{th} atom center $\mathbf{r}_i$, we have*

$$\lim_{\mathbf{r}\to\mathbf{r}_i} \hat{\epsilon}\nabla G = \mathbf{0} \quad \textit{and} \quad \lim_{\mathbf{r}\to\mathbf{r}_i} \nabla\epsilon \cdot \nabla G = 0. \tag{19}$$

Moreover, $\nabla\epsilon \cdot \nabla G \in C^{1-}(\overline{\Omega})$, i.e., $\nabla\epsilon \cdot \nabla G$ is Lipschitz continuous on Ω up to the boundary.

Proof. We first prove the second part of (19). Consider $\mathbf{r} \in \Omega_i$ with $S(\mathbf{r}) = 1$, we have

$$\nabla\epsilon(\mathbf{r}) = (\epsilon_m - \epsilon_{gap}) \sum_{i=1}^{N_m} \nabla g_i(\mathbf{r}) \prod_{k=1,k\neq i}^{N_m} [1 - g_k(\mathbf{r})], \quad \text{in } \Omega_i. \tag{20}$$

Using Eq. (17), we have $\nabla G = \sum_{j=1}^{N_m} \nabla G_j$. The inner product can be written down explicitly

$$\begin{aligned}
\nabla\epsilon \cdot \nabla G &= (\epsilon_m - \epsilon_{gap}) \sum_{i=1}^{N_m} \prod_{k=1,k\neq i}^{N_m} [1 - g_k] \sum_{j=1}^{N_m} [\nabla g_i \cdot \nabla G_j] \\
&= (\epsilon_m - \epsilon_{gap}) \sum_{i=1}^{N_m} \prod_{k=1,k\neq i}^{N_m} [1 - g_k] \left\{ \nabla g_i \cdot \nabla G_i + \sum_{j=1,j\neq i}^{N_m} [\nabla g_i \cdot \nabla G_j] \right\}.
\end{aligned} \tag{21}$$

The gradient of the i^{th} super Gaussian density function $g_i(\mathbf{r})$ can be expressed as

$$\nabla g_i(\mathbf{r}) = -\frac{2m}{\sigma^{2m} R_i^{2m}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^{2m}}{\sigma^{2m} R_i^{2m}} \right) \left(|\mathbf{r} - \mathbf{r}_i|^2 \right)^{m-1} (\mathbf{r} - \mathbf{r}_i). \tag{22}$$

Obviously, we have $\nabla g_i(\mathbf{r}_i) = \mathbf{0}$. Thus, the last summation in (21) is vanishing at $\mathbf{r} = \mathbf{r}_i$

$$\sum_{j=1,j\neq i}^{N_m} [\nabla g_i(\mathbf{r}_i) \cdot \nabla G_j(\mathbf{r}_i)] = 0,$$

because ∇G_j is finite at $\mathbf{r}_i$ for $j \neq i$. For calculating the desired limit, we just need to consider

$$\begin{aligned}
&\nabla g_i(\mathbf{r}) \cdot \nabla G_i(\mathbf{r}) \\
&= -\frac{2m}{\sigma^{2m} R_i^{2m}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^{2m}}{\sigma^{2m} R_i^{2m}} \right) \left(|\mathbf{r} - \mathbf{r}_i|^2 \right)^{m-1} (\mathbf{r} - \mathbf{r}_i) \cdot C_i \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \\
&= -\frac{2mC_i}{\sigma^{2m} R_i^{2m}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^{2m}}{\sigma^{2m} R_i^{2m}} \right) |\mathbf{r} - \mathbf{r}_i|^{2m-3}.
\end{aligned} \tag{23}$$

For $m \geq 2$, we have

$$\lim_{\mathbf{r}\to\mathbf{r}_i} \nabla g_i(\mathbf{r}) \cdot \nabla G_i(\mathbf{r}) = 0.$$

Thus, the second part of the limiting equation (19) holds.

In a sufficiently small neighborhood, $B(\mathbf{r}_i, R)$, of $\mathbf{r}_i$, by the mean value theorem and (22)

$$\begin{aligned}
|1 - g_i(\mathbf{r})| = |g_i(\mathbf{r}_i) - g_i(\mathbf{r})| &= \left| \int_0^1 (\mathbf{r} - \mathbf{r}_i) \cdot \nabla g_i(t(\mathbf{r} - \mathbf{r}_i) + \mathbf{r}_i)\, dt \right| \\
&= \frac{2m}{\sigma^{2m} R_i^{2m}} \left| \int_0^1 t^{2m-1} \exp\left(-\frac{|t(\mathbf{r} - \mathbf{r}_i) + \mathbf{r}_i|^{2m}}{\sigma^{2m} R_i^{2m}} \right) |\mathbf{r} - \mathbf{r}_i|^{2m}\, dt \right| \leq C_R |\mathbf{r} - \mathbf{r}_i|^{2m}.
\end{aligned}$$

Then inside $B(\mathbf{r}_i, R)$, it holds that

$$\left|\hat{\epsilon}(\mathbf{r}) \cdot \nabla G_i(\mathbf{r})\right| \leq \left(\epsilon_{gap} - \epsilon_m\right) C_R C_i |\mathbf{r} - \mathbf{r}_i|^{2m-2}.$$

This proves the first part of (19).

To prove the additional regularity $\nabla\epsilon \cdot \nabla G \in C^{1-}(\overline{\Omega})$, it suffices to check the regularity of $\nabla\epsilon \cdot \nabla G$ in a neighborhood of $\mathbf{r}_i$ for $i = 1, \cdots, N_m$. For each i, there exists some $R_i > 0$ such that

$$\nabla g_i \cdot \nabla G_i + \sum_{j=1, j\neq i}^{N_m} [\nabla g_i \cdot \nabla G_j] \in C^{1-}(\overline{B(\mathbf{r}_i, R_i)})$$

in view of (17), (22) and (23). Combining with the expression (21), this shows that $\nabla\epsilon \cdot \nabla G \in C^{1-}(\overline{B(\mathbf{r}_i, R_i)})$ for all $i = 1, \cdots, N_m$. Because $\nabla\epsilon \cdot \nabla G \in C^{\infty}(\overline{\Omega} \setminus \{\mathbf{r}_1, \cdots, \mathbf{r}_{N_m}\})$, we immediately deduce that $\nabla\epsilon \cdot \nabla G \in C^{1-}(\overline{\Omega})$. □

We will restrict ourselves to the case $m = 2$ from now on. Since, at all atom centers, $\hat{\epsilon}\nabla G$ is vanishing, we infer that $\hat{\epsilon}\nabla G \in C(\overline{\Omega})$ and its gradient is defined in the distributional sense. Further,

$$\nabla \cdot (\hat{\epsilon}\nabla G) = \nabla\hat{\epsilon} \cdot \nabla G + \hat{\epsilon}\Delta G, \quad \text{in } \Omega \setminus \{\mathbf{r}_1, \cdots, \mathbf{r}_{N_m}\}.$$

Note that $\epsilon_m \Delta G = -\rho$ is a distribution in Ω. Given any $f \in C_c^{\infty}(\Omega)$,

$$\int_{\Omega} \hat{\epsilon}(\mathbf{r})\Delta G(\mathbf{r}) f(\mathbf{r})\, d\mathbf{r} = -4\pi \frac{e_c^2}{k_B T \epsilon_m} \sum_{j=1}^{N_m} q_j \hat{\epsilon}(\mathbf{r}_j) f(\mathbf{r}_j) = 0,$$

because $\hat{\epsilon}$ vanishes at charge centers, which implies $\hat{\epsilon}\Delta G = 0$ in the sense of distribution. Thus, this term can be dropped from the new source

$$\nabla \cdot (\hat{\epsilon}\nabla G) = \nabla\hat{\epsilon} \cdot \nabla G = \nabla\epsilon \cdot \nabla G, \quad \text{in } \Omega, \tag{24}$$

where we have applied $\nabla\epsilon = \nabla\hat{\epsilon}$, because ϵ and $\hat{\epsilon}$ differ by a constant ϵ_m.

Based on the above analysis, we propose to solve the reaction field potential by the following regularized PB equation

$$\begin{cases} -\nabla \cdot (\epsilon\nabla u_{RF}) + (1 - S)\kappa^2 \sinh(u_{RF} + G) = \nabla\epsilon \cdot \nabla G & \text{in } \Omega, \\ u_{RF} = u_b - G & \text{on } \partial\Omega. \end{cases} \tag{25}$$

Since $(1 - S) = 0$ inside Ω_i, the nonlinear term in (25) is also vanishing in Ω_i. This guarantees that the reaction field potential u_{RF} will be bounded and smooth in Ω. Once u_{RF} is computed from (25), the potential solution of the PB equation (1) is simply given by $u = u_{RF} + G$. It is noted that in the regularized PB equation (25), one does not need to know $\hat{\epsilon}$. In other words, the dielectric function decomposition $\epsilon = \epsilon_m + \hat{\epsilon}$ is used only in the derivation, and all computations can be carried out based on $\epsilon(\mathbf{r})$ only. The details on the numerical implementation of the smooth source $\nabla\epsilon \cdot \nabla G$ will be offered in a later subsection.

2.3. Well-posedness of the super-Gaussian Poisson-Boltzmann model

Theorem 2.3. *The regularized PB equation (25) has a unique weak solution $u_{RF} \in H^1(\Omega)$. Moreover, the solution $u_{RF} \in C^2(\Omega)) \cap C(\overline{\Omega})$.*

Proof. (i) We first establish the existence and uniqueness part. To this end, a further two-component decomposition $u_{RF} = u_l + u_n$ will be introduced in such a way that

$$\begin{cases} -\epsilon_m \Delta u_l = \nabla\epsilon \cdot \nabla G & \text{in} \quad \Omega, \\ u_l = u_b - u_C & \text{on} \quad \partial\Omega \end{cases} \tag{26}$$

and

$$\begin{cases} -\nabla \cdot (\epsilon\nabla u_n) + (1 - S)\kappa^2 \sinh(u_n + u_l + G) = \nabla \cdot (\hat{\epsilon}\nabla u_l) & \text{in} \quad \Omega, \\ u_n = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{27}$$

One can extend $(u_b - u_C)|_{\partial\Omega}$ to be a function $h \in C^{\infty}(\overline{\Omega})$ by possibly multiplying $u_b - u_C$ by a smooth cut-off function to remove the singularities. Then $\widetilde{u}_l := u_l - h$ solves

$$\begin{cases} -\epsilon_m \Delta \widetilde{u}_l = \nabla\epsilon \cdot \nabla G - \epsilon_m \Delta h & \text{in} \quad \Omega, \\ \widetilde{u}_l = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{28}$$

We can infer from [18, Corollary 1] that (28) has a unique solution $\widetilde{u}_l \in H^2(\Omega) \cap H_0^1(\Omega)$ and in turn $u_l \in H^2(\Omega)$. We thus have $\nabla \cdot (\hat{\epsilon}\nabla u_l) \in L_2(\Omega)$.

To study (27), we define the energy functional

$$I[v] = \int_\Omega \left[\frac{\epsilon}{2} |\nabla v|^2 + (1-S)\kappa^2 \cosh(v + u_l + G) - v\nabla \cdot (\hat{\epsilon}\nabla u_l) \right] d\mathbf{r}, \quad v \in H_0^1(\Omega).$$

By the Hölder, Poincaré and Young's inequalities,

$$\int_\Omega \left[v\nabla \cdot (\hat{\epsilon}\nabla u_l) \right] d\mathbf{r} \le L(\delta) \|\nabla \cdot (\hat{\epsilon}\nabla u_l)\|_{L_2(\Omega)}^2 + \delta \|\nabla v\|_{L_2(\Omega)}^2,$$

for any $\delta > 0$ and a continuous and decreasing function $L : \mathbb{R}_+ \to \mathbb{R}_+$. Choosing $\delta > 0$ sufficiently small, we can immediately verify that

$$I[v] \ge C_1 \|\nabla v\|_2^2 - C_2,$$

for some constants $C_1 > 0$ and C_2. This implies that $I[\cdot]$ is coercive. It is a simple task to check the strict convexity of the functional. Hence the direct method of Calculus of Variation yields the existence and uniqueness of a global minimizer of $I[\cdot]$. Here v solves (27) iff it is a critical point of $I[\cdot]$. The uniqueness of a critical point of $I[\cdot]$ is an immediate consequence of the strict convexity of $I[\cdot]$. We thus infer that u_n is the uniqueness weak solution to (27), and thus $u_{RF} := u_l + u_n$ uniquely and weakly solves (25) in $H^1(\Omega)$.

(ii) Next, we will show u_{RF} is essentially bounded. From the Sobolev embedding $H^2(\Omega) \hookrightarrow C^{1/2}(\overline{\Omega})$, cf. [2, Theorem 4.12], we conclude that $\|u_l\|_\infty \le C_3$ for some constant C_3. To show the essential boundedness of u_n, we further decompose it into $u_n = u_{n,1} + u_{n,2}$ in such a way that

$$\begin{cases} -\nabla \cdot (\epsilon \nabla u_{n,1}) = \nabla \cdot (\hat{\epsilon}\nabla u_l) & \text{in} \quad \Omega, \\ u_n = 0 & \text{on} \quad \partial\Omega \end{cases} \tag{29}$$

and

$$\begin{cases} -\nabla \cdot (\epsilon \nabla u_{n,2}) + (1-S)\kappa^2 \sinh(u_{n,2} + u_{n,1} + u_l + G) = 0 & \text{in} \quad \Omega, \\ u_n = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{30}$$

Since $\hat{\epsilon}\nabla u_l \in H^1(\Omega) \hookrightarrow L_6(\Omega)$, it follows from [21, Theorem 8.16] that $\|u_{n,1}\|_\infty \le C_4$ for some constant $C_4 > 0$. We already know that

$$|G(\mathbf{r})| \le C_5, \quad \mathbf{r} \in \Omega \setminus \Omega_i$$

for some $C_5 > 0$. Define

$$W = \{\mathbf{r} \in \Omega : u_{n,2} \le C_3 + C_4 + C_5\}.$$

Multiplying both sides of (30) by $(u_{n,2} - C_3 - C_4 - C_5)_+$ and integrating over Ω yield

$$\int_W \left[\epsilon |\nabla u_{n,2}|^2 + (1-S)\kappa^2 \sinh(u_{n,2} + u_{n,1} + u_l + G)(u_{n,2} - C_3 - C_4 - C_5)_+ \right] d\mathbf{r} = 0.$$

Because both terms in the integrand are non-negative, this implies that

$$u_{n,2} \le C_3 + C_4 + C_5 \quad \text{a.e. in } \Omega \setminus \Omega_i.$$

Similarly, we can show that

$$u_{n,2} \ge -C_3 - C_4 - C_5 \quad \text{a.e. in } \Omega \setminus \Omega_i.$$

Applying [21, Theorem 8.16] once more to (30) yields that $\|u_{n,2}\|_\infty \le C_6$ for some C_6.

(iii) From Steps (i) and (ii), one can readily infer that $(1-S)\sinh(u_{RF} + G) \in H^1(\Omega)$. In view of the C^2-continuity of ϵ, applying [21, Theorem 8.10] to (25), we conclude that $u_{RF} \in H^3(\Omega')$ for any subdomain Ω' such that $\overline{\Omega}' \subset \Omega$. The Sobolev embedding theorem, cf. [2, Theorem 4.12], implies that $u_{RF} \in C^1(\Omega)$. Since both $(1-S)\sinh(u_{RF} + G)$ and $\nabla\epsilon \cdot \nabla G$ are Lipschitz, we infer from [21, Theorem 6.13] that $u_{RF} \in C^2(\Omega)) \cap C(\overline{\Omega})$. Note that a cube satisfies an exterior sphere condition. □

2.4. Electrostatic free energy

In biomolecular simulations, one common usage of the PB model is to compute the electrostatic free energy, which measures the difference in the polar solvation energy of the macromolecule between the water phase and vacuum phase [25,33]. This amounts to solving the PB equation for water state and a Poisson equation for vacuum state, subject to the same singular charge sources.

We first review how the electrostatic free energy is calculated in the original super-Gaussian PB model. For simplicity, consider a linearized PB (LPB) model, which is commonly used when the electrostatic potential is weak. In this case, $\sinh(u)$ can be approximated by u, so that the electrostatic interactions can be modeled by a LPB equation

$$-\nabla\cdot(\epsilon(\mathbf{r})\nabla u(\mathbf{r}))+(1-S(\mathbf{r}))\kappa^2 u(\mathbf{r})=\rho(\mathbf{r}),\quad \text{in } \Omega, \tag{31}$$

subject to the same source term (2) and boundary condition (3). Here $\epsilon(\mathbf{r})$ is produced by the super-Gaussian dielectric model with $m\geq 2$ and $\epsilon_{out}=80$. By solving (31) numerically, one attains $u(\mathbf{r})$ for the water state. For the vacuum state, one solves a Poisson equation [22]

$$-\nabla\cdot(\epsilon_v(\mathbf{r})\nabla v(\mathbf{r}))=\rho(\mathbf{r}),\quad \text{in } \Omega, \tag{32}$$

which is derived by taking $\kappa=0$ in the LPB equation (31) since no ions are presented in the vacuum. On $\partial\Omega$, we have a boundary condition $v(\mathbf{r})=v_b(\mathbf{r})$, where $v_b(\mathbf{r})$ is obtained by setting $\kappa=0$ in (3). The dielectric function $\epsilon_v(\mathbf{r})$ in (32) is defined by Eq. (8) with $\epsilon_{out}=1$. By solving (32) numerically, one attains $v(\mathbf{r})$ for the vacuum state. With the potential solutions of two states, the electrostatic free energy is then calculated as

$$E=\frac{1}{2}k_BT\int_\Omega\sum_{j=1}^{N_m}q_j\delta(\mathbf{r}-\mathbf{r}_j)(u(\mathbf{r})-v(\mathbf{r}))d\mathbf{r}=\frac{1}{2}k_BT\sum_{j=1}^{N_m}q_j(u(\mathbf{r}_j)-v(\mathbf{r}_j)). \tag{33}$$

We next discuss how to compute electrostatic free energy in the proposed regularization method. In the water state, the potential decomposition is conducted for the solution of the LPB equation (31): $u=u_{RF}+u_C=u_{RF}+G$. Then the reaction field potential u_{RF} satisfies a regularized LPB equation

$$\begin{cases}-\nabla\cdot(\epsilon(\mathbf{r})\nabla u_{RF}(\mathbf{r}))+(1-S(\mathbf{r}))\kappa^2u_{RF}(\mathbf{r})=\nabla\epsilon(\mathbf{r})\cdot\nabla G(\mathbf{r})-(1-S(\mathbf{r}))\kappa^2G(\mathbf{r}), & \text{in } \Omega,\\ u_{RF}(\mathbf{r})=u_b(\mathbf{r})-G(\mathbf{r}) & \text{on } \partial\Omega.\end{cases} \tag{34}$$

In the solute domain Ω_i, $(1-S)\kappa^2G=0$, while in other subdomains, G is well-defined. Thus, the second source term of (34), i.e., $(1-S)\kappa^2G$, is smooth over the entire domain Ω. By solving (34) numerically, one attains $u_{RF}(\mathbf{r})$ for the water state. A similar potential decomposition is carried out in the vacuum state for the Poisson equation (32): $v=v_{RF}+v_C=v_{RF}+G$. Here the singular component v_C is essentially u_C and is also analytically represented by the Green's function G. The reaction field potential satisfies a regularized Poisson equation

$$\begin{cases}-\nabla\cdot(\epsilon_v(\mathbf{r})\nabla v_{RF}(\mathbf{r}))=\nabla\epsilon_v(\mathbf{r})\cdot\nabla G(\mathbf{r}), & \text{in } \Omega,\\ v_{RF}(\mathbf{r})=v_b(\mathbf{r})-G(\mathbf{r}) & \text{on } \partial\Omega.\end{cases} \tag{35}$$

The source $\nabla\epsilon_v\cdot\nabla G$ is Lipschitz. In particular, we have $\nabla\epsilon_v\cdot\nabla G=\nabla\epsilon\cdot\nabla G$ in Ω_i and $\nabla\epsilon_v\cdot\nabla G=0$ in Ω_e. In Ω_t, $\nabla\epsilon_v\cdot\nabla G$ shall be computed independently. By solving (35) numerically, one attains $v_{RF}(\mathbf{r})$ for the vacuum state. The electrostatic free energy is still defined by (33). However, since $v_C=u_C$, the singular components are simply canceled. Thus, the free energy can be computed by using reaction field components only

$$E=\frac{1}{2}k_BT\int_\Omega\sum_{j=1}^{N_m}q_j\delta(\mathbf{r}-\mathbf{r}_j)(u_{RF}(\mathbf{r})-v_{RF}(\mathbf{r}))d\mathbf{r}=\frac{1}{2}k_BT\sum_{j=1}^{N_m}q_j(u_{RF}(\mathbf{r}_j)-v_{RF}(\mathbf{r}_j)). \tag{36}$$

In the present study, the electrostatic free energy will be reported in the unit of kcal/mol. Since the numerically calculated potentials are available only on grid nodes, a linear interpolation has to be conducted to evaluate these potentials at charge centers $\mathbf{r}_j$ in (33) and (36).

2.5. Numerical discretization and implementation

In this subsection, we offer details on the numerical discretization of the regularized LPB equation (34) and numerical implementation of source terms. For a given protein structure, we first select a large enough computational domain Ω and construct a uniform mesh partition with N_x, N_y, N_z being the number of the grid points in x, y, and z directions, respectively. Without the loss of generality, we assume the grid spacing h in all x, y, and z directions to be the same, i.e., $h=\Delta x=\Delta y=\Delta z$, with the unit Å. For a function f defined at a node (x_i,y_j,z_k), we denote $f_{i,j,k}=f(x_i,y_j,z_k)$.

The central finite difference method is employed to directly discretize the divergence form of the regularized LPB equation (34). For example, at the node (x_i, y_j, z_k), the x derivative is approximated as

$$\left[\frac{\partial}{\partial x}(\epsilon\frac{\partial u_{RF}}{\partial x})\right]_i = \frac{1}{h^2}\left[\epsilon_{i+\frac{1}{2}}\left((u_{RF})_{i+1}-(u_{RF})_i\right)-\epsilon_{i-\frac{1}{2}}\left((u_{RF})_i-(u_{RF})_{i-1}\right)\right]+O(h^2), \tag{37}$$

where we have dropped indexes for y_j and z_k for simplicity.

In the present study, the diffuse interface function $S(\mathbf{r})$ in (34) is generated by the Gaussian convolution surface (GCS) [44], which is accelerated by the fast Fourier transform (FFT) algorithm. The dielectric function $\epsilon(\mathbf{r})$ is constructed according to (8), which depends on $S(\mathbf{r})$ as well as several super Gaussian parameters. Similar to [44], the band-width of Ω_t will be taken as 3 Å in the GCS, which allows a smooth variation of $\epsilon(\mathbf{r})$ in Ω_t. Since the GCS surface function $S_{i,j,k}$ is generated numerically, $\epsilon_{i,j,k}$ is only known at nodes (x_i, y_j, z_k). Thus, the half-grid values needed in central difference approximation, such as in (37), are obtained via an average

$$\epsilon_{i+\frac{1}{2},j,k} = \frac{\epsilon_{i,j,k}+\epsilon_{i+1,j,k}}{2}+O(h^2). \tag{38}$$

The finite difference discretization of the left-hand side of Eq. (34) gives rise to a sparse linear system with dimension N_3-by-N_3, where $N_3 = N_x \times N_y \times N_z$. A biconjugate gradient iterative solver is employed to solve this linear system for $u_{RF}(x_i, y_j, z_k)$. The order of accuracy of the entire central difference discretization is two.

The implementation of the source terms of Eq. (34) requires further attention. Since $G_{i,j,k}$ and $\nabla G_{i,j,k}$ are known analytically, the second source term $(1-S)\kappa^2 G$ can be calculated directly at (x_i, y_j, z_k). The first source term $\nabla\epsilon\cdot\nabla G$ depends on several parts. Whenever possible, analytical differentiation instead of finite difference approximation is used for calculating $\nabla\epsilon$. Using Eq. (8), one has

$$\nabla\epsilon(\mathbf{r}) = S(\mathbf{r})\nabla\epsilon_{in}(\mathbf{r})+\nabla S(\mathbf{r})(\epsilon_{in}(\mathbf{r})-\epsilon_{out}), \quad \text{in } \Omega. \tag{39}$$

For $(x_i, y_j, z_k)\in\Omega_e$, we have $S=0$ and $\nabla S=\mathbf{0}$, so that $\nabla\epsilon=\mathbf{0}$ in the solvent subdomain Ω_e. In the solute subdomain Ω_i with $S=1$ and $\nabla S=\mathbf{0}$, $\nabla\epsilon=\nabla\epsilon_{in}$ can be computed analytically. Using the definition of ϵ_{in}, one attains

$$\nabla\epsilon_{in}(\mathbf{r}) = (\epsilon_m-\epsilon_{gap})\sum_{n=1}^{N_m}\nabla g_n(\mathbf{r})\prod_{l=1,l\neq n}^{N_m}[1-g_l(\mathbf{r})] = (\epsilon_m-\epsilon_{gap})\sum_{n=1}^{N_m}\left(\frac{1-g(\mathbf{r})}{1-g_n(\mathbf{r})}\right)\nabla g_n(\mathbf{r}).$$

For $m\geq 2$, we have the analytical form

$$\nabla\epsilon_{in}(\mathbf{r}) = (\epsilon_{gap}-\epsilon_m)\sum_{n=1}^{N_m}\frac{2m}{\sigma^{2m}R_n^{2m}}\left(\frac{1-g(\mathbf{r})}{1-g_n(\mathbf{r})}\right)g_n(\mathbf{r})\left(|\mathbf{r}-\mathbf{r}_n|^2\right)^{m-1}(\mathbf{r}-\mathbf{r}_n), \tag{40}$$

for $\mathbf{r}\in\Omega$, except at all atom centers $\mathbf{r}_n$, $n=1,2,\ldots,N_m$. When $(x_i, y_j, z_k)\in\Omega_i$, $\nabla\epsilon_{i,j,k}$ can be calculated analytically using Eq. (40). We note that if the grid node (x_i, y_j, z_k) happens to be the n^{th} atom center, i.e., $(x_i, y_j, z_k)=\mathbf{r}_n$, the denominator $1-g_n(\mathbf{r}_n)=0$ in (40). Nevertheless, this is not an issue in our computation. According to Theorem 2.2, the first source term is vanishing at all atom centers. Thus, in this case, one should not apply (40), but just simply takes $\nabla\epsilon\cdot\nabla G=0$. When $(x_i, y_j, z_k)\in\Omega_t$, ϵ_{in} and $\nabla\epsilon_{in}$ are still calculated analytically. By approximating the gradient of S using the central difference, e.g.,

$$\left.\frac{\partial S}{\partial x}\right|_{i,j,k} = \frac{S_{i+1,j,k}-S_{i-1,j,k}}{2h}+O(h^2),$$

$\nabla\epsilon_{i,j,k}$ can be computed according to Eq. (39). The smooth source term $\nabla\epsilon_v\cdot\nabla G$ for the regularized Poisson equation (35) in vacuum state can be calculated similarly.

For a comparison, the traditional trilinear method [33] will also be employed to treat singular charges contained in ρ of the un-regularized LPB equation (31) and Poisson equation (32). For each partial charge q_n located at $\mathbf{r}_n$, one will find a cubic cell containing this charge. One will then distribute the charge q_n into eight corner nodes of the cell, which gives rise to eight numerical source values $Q_{i,j,k}$ for each partial charge. The Laplacian term of LPB equation (31) and Poisson equation (32) will be discretized by the same second order central differences (37). Denote the resulting potential solutions as $u(x_i, y_j, z_k)$ and $v(x_i, y_j, z_k)$, respectively, for water and vacuum states. Because the Dirac delta functions are approximated by finite quantities, the trilinear method produces a significant error at charge centers. However, in calculating the electrostatic free energy by using Eq. (33), only the difference between $u(x_i, y_j, z_k)$ and $v(x_i, y_j, z_k)$ is needed. By using the same trilinear interpolation in computing both $u(x_i, y_j, z_k)$ and $v(x_i, y_j, z_k)$, this difference will significantly cancel the error associated with the delta function approximation. We will illustrate this point in the next section.

3. Numerical experiments

In this section, we will examine the proposed regularization method by providing simulation results, and compare it with the trilinear method. A manufactured example with analytical potential solutions will be studied first. Then, we report results of electrostatic free energy calculation for one atom system and small molecules. After a convergence analysis by using one protein, a set of real proteins will finally be employed to further compare the regularization and trilinear methods. In the super-Gaussian PB model, the following parameter values are chosen: $m = 2$, $\sigma = 1$, $\epsilon_m = 1$, and $I = 0.15$ m. For the water state and vacuum state, ϵ_{out} is chosen as 80 and 1, respectively. In all computations, the length is reported with an unit Å, and the electrostatic free energy has the unit of kcal/mol. In all figures and tables, the regularization and trilinear methods will be denoted as REG and TRI, respectively.

3.1. A manufactured example with analytical solution

For the sharp interface PB model, a new numerical algorithm is usually benchmarked by considering a Kirkwood sphere with analytical potential or energy values [4,19,20]. However, no analytical example is known for the Gaussian dielectric PB models [22,30,31], due to a complicated dielectric profile. In order to validate the proposed regularization method, we will consider a one-atom system and construct an analytical potential.

Consider a one atom system with a unit charge $q_0 = 1$ located at $\mathbf{r}_0 = (0, 0, 0)$. The interfaces Γ_i and Γ_e are spheres with radii being $r_i = 2$ and $r_e = 5$, respectively. A tanh-like surface function is defined, which smoothly varies in the radial direction

$$S(\mathbf{r}) = \frac{1}{2} - \frac{1}{2}\tanh(k(\frac{|\mathbf{r}| - r_i}{r_e - r_i} - \frac{1}{2})), \tag{41}$$

where a large enough $k = 14$ is employed so that $S \approx 1$ in Ω_i and $S \approx 0$ in Ω_e. In particular, at the charge center $\mathbf{r}_0$, we have that $S(\mathbf{r}_0)$ numerically equals to 1, i.e., $|S(\mathbf{r}_0) - 1| = 6.44e - 15$. With $m = 2$ and $\sigma = 1$, the super-Gaussian density function is $g(\mathbf{r}) = \exp(-|\mathbf{r}|^4/r_i^4)$. Then, ϵ_{in} can be calculated as

$$\epsilon_{in}(\mathbf{r}) = \epsilon_m + (\epsilon_{gap} - \epsilon_m)\left(1 - \exp\left(-\frac{|\mathbf{r}|^4}{r_i^4}\right)\right), \tag{42}$$

and the super-Gaussian dielectric function is given as

$$\epsilon(\mathbf{r}) = S(\mathbf{r})\epsilon_{in}(\mathbf{r}) + (1 - S(\mathbf{r}))\epsilon_{out}, \tag{43}$$

where $\epsilon_m = 1$, $\epsilon_{out} = 80$ and $\epsilon_{gap} = 8$.

To construct an analytical potential for this one-atom system, we consider a Poisson's equation with $\kappa = 0$

$$-\nabla \cdot (\epsilon \nabla u) = 4\pi \frac{e_c^2}{k_B T} q_0 \delta(\mathbf{r} - \mathbf{r}_0) + f(\mathbf{r}), \tag{44}$$

where $f(\mathbf{r})$ is an additional source. As in the usual PB models, the potential u can be decomposed into reaction-field and coulomb components $u = u_{RF} + G$, where $G = \frac{e_c^2}{k_B T}\frac{q_0}{\epsilon_m |\mathbf{r} - \mathbf{r}_0|}$. It is noted that, since $S(\mathbf{r}_0)$ is slightly different from 1, $\epsilon(\mathbf{r}_0) \neq \epsilon_m$. Nevertheless, the difference between $\epsilon(\mathbf{r}_0)$ and ϵ_m is around $1.0e - 14$. For simplicity, we assume them to be the same numerically, and still use ϵ_m in the Green's function G. In this study, we assume that u_{RF} takes the form

$$u_{RF}(\mathbf{r}) = -\frac{e_c^2}{k_B T}\exp\left(-\frac{|\mathbf{r}|^2}{r_e^2}\right). \tag{45}$$

By substituting the analytical expressions for u_{RF} and G into the Poisson equation (44), one can derive the expression for the additional source $f(\mathbf{r})$. See Appendix B for more details.

Assume a computational domain $\Omega = [-10, 10]^3$, which is partitioned by a uniform mesh with the spacing $h = 20/(N - 1)$ in all directions. Here N will be chosen as an even integer so that the singular charge point is not located on the grid nodes. We will solve this one-atom system by the trilinear method, as well as two regularization implementations. In all cases, the Dirichlet boundary condition is obtained by the exact solution.

For the trilinear method, one directly discretizes the original Poisson equation (44) for a numerical solution $u(x_i, y_j, z_k)$. For the regularization approach, we will solve the regularized Poisson equation for the reaction field component

$$-\nabla \cdot (\epsilon \nabla u_{RF}) = \nabla\epsilon \cdot \nabla G + f. \tag{46}$$

In the regularization or REG scheme, $\nabla\epsilon$ is implemented as discussed above. In particularly, $\nabla\epsilon_{in}$ is calculated analytically by (40), while ∇S is approximated by the central difference. For the present example, we have implemented another regularization scheme, i.e., REG-2, in which $\nabla\epsilon$ is simply approximated by the central difference, e.g.,

Table 1
Numerical convergence of the trilinear and two regularization methods for the manufactured example.

Norm	N	TRI		REG		REG-2	
		Error	Order	Error	Order	Error	Order
L_2	50	7.25e-1		3.86e-1		8.14e-1	
	100	3.07e-1	1.23	1.12e-1	1.77	2.08e-1	1.94
	200	1.86e-1	0.72	3.03e-2	1.87	5.31e-2	1.96
	400	1.28e-1	0.53	7.76e-3	1.96	1.34e-2	1.99
L_∞	50	6.26e+1		7.01e-0		5.61e+1	
	100	7.43e+1	-0.24	1.47e-0	2.22	1.74e+1	1.66
	200	1.33e+2	-0.83	3.90e-1	1.90	4.77e-0	1.86
	400	2.63e+2	-0.98	9.88e-2	1.97	1.24e-0	1.93

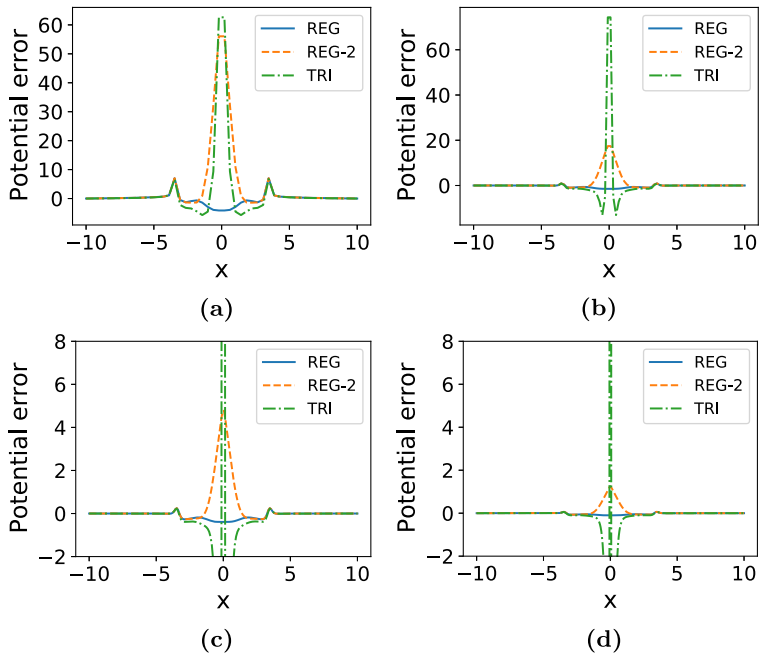

Fig. 2. Plots of numerical errors along a x line with $y = h/2$ and $z = h/2$ for the manufactured example. (a) $N = 50$; (b) $N = 100$; (c) $N = 200$; (d) $N = 400$. Large enough windows are used in first two subplots, while a small window is used in last two subplots so that errors of two regularization schemes could be visualized.

$$\left.\frac{\partial \epsilon}{\partial x}\right|_{i,j,k} = \frac{\epsilon_{i+\frac{1}{2},j,k} - \epsilon_{i-\frac{1}{2},j,k}}{h} + O(h^2).$$

By taking $N = 50$, 100, 200, and 400, the L_2 and L_∞ errors of three numerical schemes are reported in Table 1. It can be seen that both regularization methods achieve a second order convergence in L_2 and L_∞ norms, while the REG scheme is more accurate than the REG-2. For the TRI scheme, the L_2 error converges, but the L_∞ error becomes larger and larger as N increases. In Fig. 2, errors of three numerical schemes are depicted along a x grid line with $y = z = h/2$. It can be seen that the largest trilinear error is always located at the charge center, while outside Ω_i, the trilinear error is quite similar to those of two regularization schemes. This suggests that the failure of the trilinear approximation is solely caused by the singular charge. Due to the use of finite difference to approximate the source term $\nabla\epsilon$, the REG-2 scheme initially produces a large error at the charge center, which is just slightly smaller than that of TRI for $N = 50$. Nevertheless, as N becomes larger, the REG-2 error at the charge point decays rapidly toward zero.

By using an analytical source implementation, the REG error is always negligible at the charge center. It can be observed in Fig. 2 that the maximum error of the REG scheme is located around $x = \pm 3.5$, where the surface function S has the largest gradient value. This means the numerical error of the REG scheme is solely due to finite difference approximations, because the charge singularity is well taken care of in the proposed regularization and implementation. The present study also demonstrates the advantage of analytical source implementation over finite different approximation. For this reason, we will focus only on the REG scheme in the rest of studies.

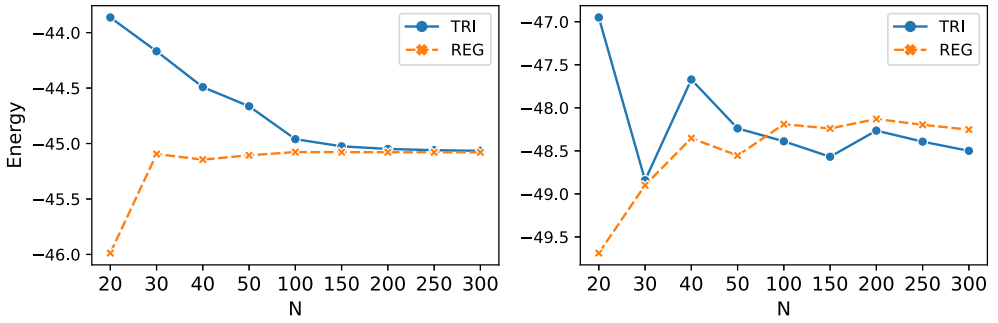

Fig. 3. Electrostatic free energies of one atom system calculated by the regularization method and trilinear method. The left subfigure is based on the tanh-like surface and the right one is by using the Gaussian convolution surface (GCS).

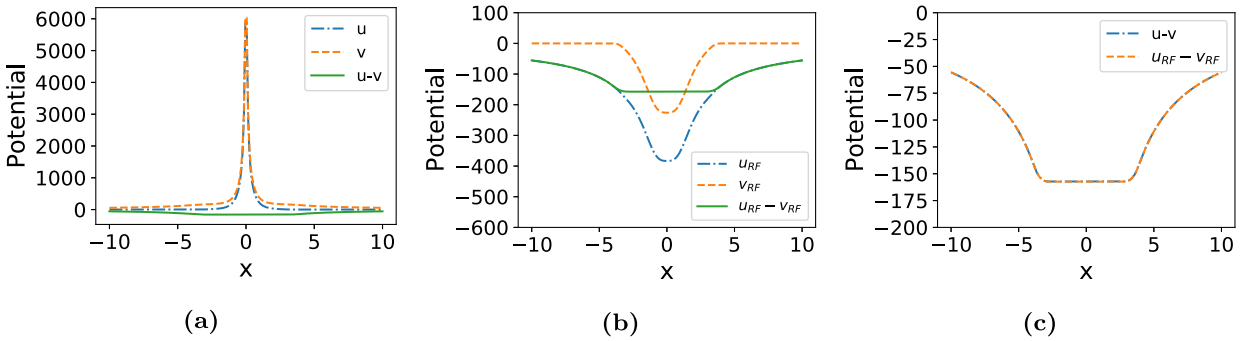

Fig. 4. Line plots of potentials along a x line with $y = h/2$ and $z = h/2$ for the one-atom system with the tanh-like surface. (a) PB potential u, vacuum potential v, and their difference $u - v$, calculated by the trilinear method; (b) PB potential u_{RF}, vacuum potential v_{RF}, and their difference $u_{RF} - v_{RF}$, calculated by the regularization method; (c) Comparison of trilinear and regularization potential differences.

3.2. Electrostatic free energy of one atom

We next calculate the electrostatic free energy of the one atom system studied above. The model parameters are still $q_0 = 1$, $\mathbf{r}_0 = (0, 0, 0)$, $r_i = 2$, and $r_e = 5$. Two diffuse interface settings will be examined. First, the same tanh-like surface will be used, with S calculated by Eq. (41). Second, the Gaussian convolution surface (GCS) developed in [44] will be employed to compute another S function. Based on the S function, one can then define $\epsilon(\mathbf{r})$ and $\epsilon_v(\mathbf{r})$, respectively, for water and vacuum states. In this example and the following simulations, we will take $\epsilon_{gap} = 4$.

Using the procedures described in Section 2, we will calculate the electrostatic free energy by regularization and trilinear methods. In particular, one solves the LPB equation (31) and Poisson equation (32) by the trilinear method, respectively, for potentials u and v. Then Eq. (33) is adopted to compute the electrostatic free energy E, which essentially depends on the difference of u and v. For the regularization method, one solves the regularized LPB equation (34) and regularized Poisson equation (35), respectively, for u_{RF} and v_{RF}. The difference between them is used in Eq. (36) for calculating E. We note that no exact energy value is known for the present super-Gaussian PB model of the one atom system.

By testing a set of even integers for N, the energies calculated by both methods are shown in the left chart of Fig. 3 for the tanh-like surface. It can be seen that the energies of trilinear and regularization methods converge to the same limit. For the regularization method, the energy curve becomes almost flat after $N > 50$, and the energy at $N = 300$ is -45.0792, which will be treated as the reference energy. The trilinear energy also converges toward the reference energy, but at a slower rate. In particular, the trilinear energy at $N = 300$ is -45.0660, whose error in comparing with the reference energy is about 0.0132. For a comparison, the energy error of the regularization method at $N = 100$ is already as small as 0.0020. The means that the regularization is much more accurate than the trilinear method in calculating electrostatic free energy.

It is noted that even though the trilinear method involves a large and divergent approximation error at charge center, it may produce convergent energy. To see this point, we plot the potentials of both methods along a x grid line with $y = z = h/2$, where $N = 200$. It is observed in Fig. 4 that for both PB potential u and vacuum potential v of the trilinear method, a huge error happens again at the charge center, due to the poor approximation of the singular charge. Since these two charge errors are almost the same for u and v, they are simply canceled in $u - v$ for calculating the electrostatic free energy. In other words, the artificial grid energy of the trilinear method could be suppressed by an error cancellation. Thus, it can be seen from Fig. 4 (a) that $u - v$ becomes pretty flat at the charge center, and the huge error at the atom center disappears. For the regularization method, both u_{RF} and v_{RF} are free of charge errors at the atom center, and

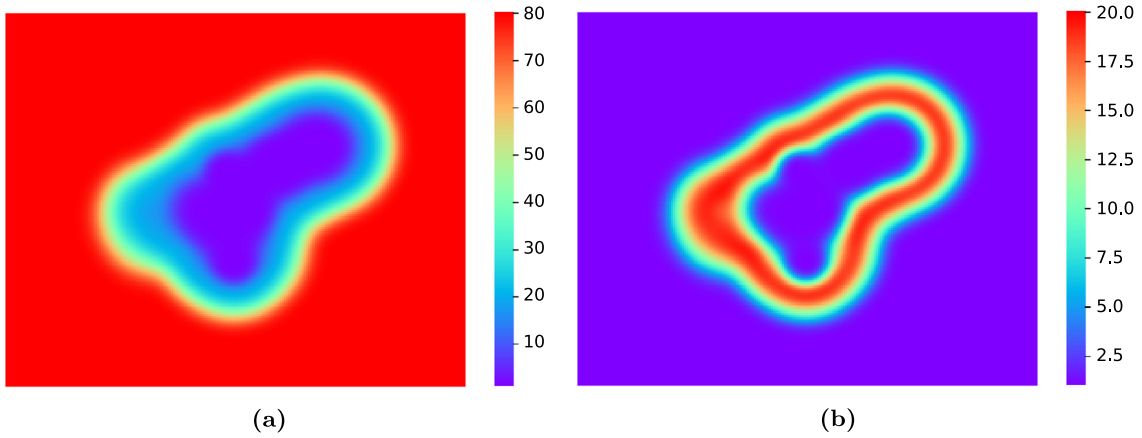

Fig. 5. Heat-map plots of the dielectric functions cross a plane with $z =$ const for the GLY. (a) $\epsilon(\mathbf{r})$ for the water state; (b) $\epsilon_v(\mathbf{r})$ for the vacuum state. For illustration purpose, ϵ_{gap} is taken as 20 in the plots. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

their difference $u_{RF} - v_{RF}$ is also quite flat at the atom center. In Fig. 4 (c), the potential differences of the trilinear and regularization methods are plotted together. It can be seen that they agree with each other very well. The minor difference between them cannot be visualized in this figure.

The tanh-like surface is defined only for the one atom system. To efficiently generate diffuse interfaces for general protein systems, a GCS has been introduced in [44]. To test the GCS in the super-Gaussian PB model, we first apply it to the one atom system. The estimated energies of regularization and trilinear methods are depicted in the right chart of Fig. 3. Now, with $S(\mathbf{r})$ being constructed numerically, the regularization method will be implemented as mentioned in Section 2. Since the GCS is computed discretely, itself has a convergence issue. Moreover, it has been shown in [44] that the GCS convergence is oscillatory, not monotonically. This affects the energy convergence. It can be seen from Fig. 3 Right that the calculated energies of regularization and trilinear methods are all fluctuated, and the regularization energy is larger than the trilinear one. It also can be observed that the oscillation amplitude of the regularization is less than that of the trilinear method, indicating that the regularization could be more robust in energy computation.

3.3. Impact of rotation and shifting

We then investigate the sensitivity of grid rotation and shifting. Physically, when the concerned biomolecule is shifted or rotated to a new position, the electrostatic free energy will not be altered. However, because the finite difference method is a grid-based numerical algorithm for solving the PB equation, the numerical energy is grid dependent. Therefore, for a new finite difference PB model, it is important to test the grid sensitivity with respect to rotation and shifting [31], in which a so-called artificial grid energy is usually considered to measure the numerical artifact.

In the present study, we consider a small compound, i.e., glycerol triacetate or GLY. The charges, atomic coordinates, and radii are defined based on a parameterization presented in [34]. By taking $h = 0.1$, an initial domain $[-7, 14] \times [-10, 8] \times [-7, 12]$ is chosen, and is partitioned into a uniform mesh. The dielectric functions $\epsilon(\mathbf{r})$ and $\epsilon_v(\mathbf{r})$ of the GLY in the water and vacuum states are displayed, respectively, in Fig. 5 (a) and (b). It can be seen that inside the Van der Waals region, two dielectric functions are the same. Outside the smooth solute-solvent boundary, the dielectric function goes to 80 and 1, respectively, for the water and vacuum states. Moreover, no geometrical symmetry can be observed in Fig. 5. Hence, this is a good example to test the sensitivity of grid rotation and shifting.

In the shifting test, we will shift the GLY in z direction with an increment 0.01. In other words, for all atomic coordinates, we will keep x and y values unchanged and add z values by 0.01. For each shifted structure, the computational domain will be automatically generated. After shifting ten steps, the relative position of the shifted structure with respect to the uniform mesh is essentially the same as the initial one, because $h = 0.1$. Then, the numerical energy should become the same as the original one. In the rotation test, we will fix z coordinate values, and rotate (x, y) values within the xy-plane with respect to $(x, y) = (0, 0)$. By using an angular increment of 30°, the (x, y) coordinates will be rotated back to the original ones after 12 steps. For each rotated structure, a new computational domain will be automatically generated for calculating the energy.

We first consider a synchronized coordinate change that will be applied to both PB potential and vacuum potential. For the trilinear method, this means that the biomolecular structure will be changed simultaneously in calculating potential u for the LPB equation (31) and v for the Poisson equation (32). Similarly, for the regularization method, the biomolecular structure will be changed at the same time for both u_{RF} and v_{RF}. The sensitivity results of the trilinear and regularization methods are depicted in Fig. 6 (a) and (c), respectively, for rotation and shifting. It can be seen that the regularization energy is barely affected by the rotation or shifting – only a very minor fluctuation is displayed. For the trilinear method, we can see that initial energy value is slightly different from that of the regularization. The fluctuation is more severe than

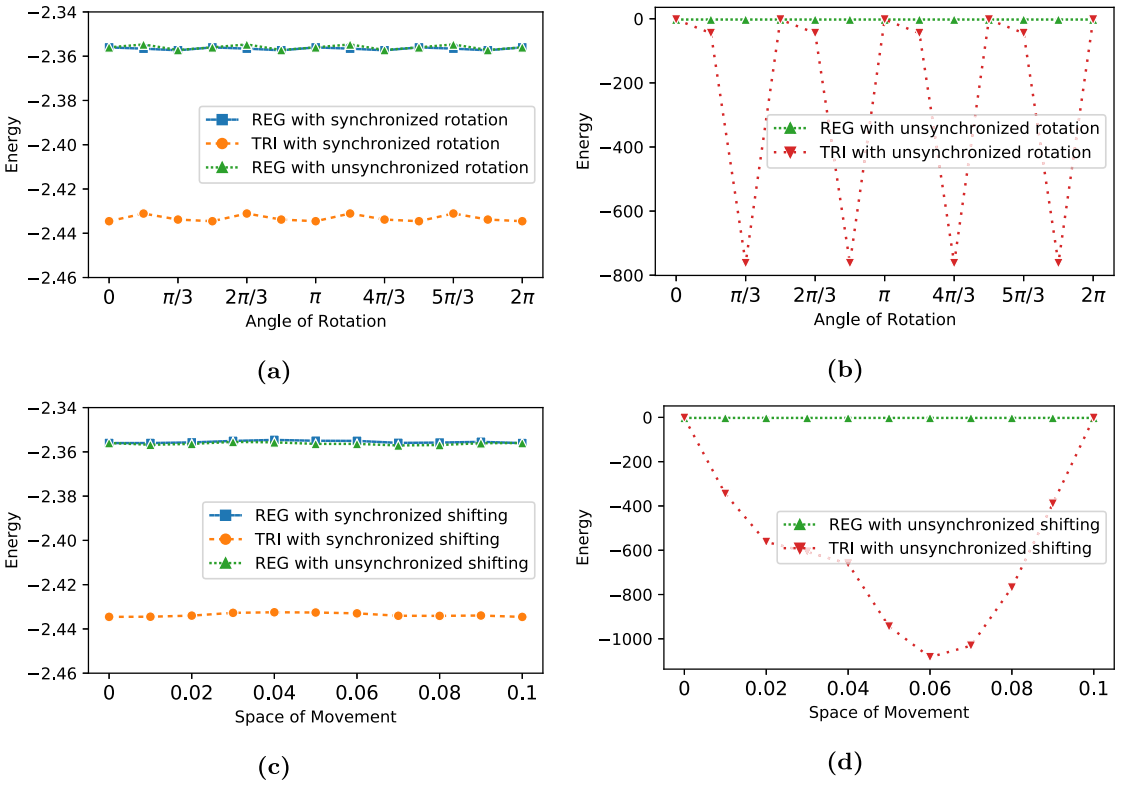

Fig. 6. Grid sensitivity tests of rotation and shifting for the GLY. (a) Synchronized rotation; (b) Unsynchronized rotation; (c) Synchronized shifting; (d) Unsynchronized shifting.

the regularization, but it is still acceptable. Due to the synchronized coordinate change, the structure and the corresponding numerical grid are the same in calculating u and v, so that the error cancellation can still be taken advantage of for the trilinear method. This is why the trilinear method is still acceptable.

We next consider an unsynchronized coordinate change. For the trilinear method, instead of Eq. (33), the electrostatic free energy will be calculated as

$$E = \frac{1}{2}k_B T \sum_{j=1}^{N_m} q_j u(\mathbf{r}_j') - \frac{1}{2}k_B T \sum_{j=1}^{N_m} q_j v(\mathbf{r}_j), \tag{47}$$

where the atomic coordinates $\mathbf{r}_j$ will be fixed in calculating the potential v in the vacuum, while $\mathbf{r}_j'$ will be rotated or shifted in calculating the potential u in the solvent. Similarly, in the regularization method, the electrostatic free energy will be computed as

$$E = \frac{1}{2}k_B T \sum_{j=1}^{N_m} q_j u_{RF}(\mathbf{r}_j') - \frac{1}{2}k_B T \sum_{j=1}^{N_m} q_j v_{RF}(\mathbf{r}_j). \tag{48}$$

The numerical results of unsynchronized rotation and shifting are illustrated, respectively, in Fig. 6 (b) and (d). Significant errors can be seen now for the trilinear method. For the shifting, the largest deviation happens at 0.06 with the energy being around -1100. For the rotation, the energy oscillates dramatically with maximal deviations around -800. At rotation angles of degree 90, 180, 270, the energy is close to that of degree 0. It is impressive to see that the regularization energy is barely affected by such unsynchronized rotation or shifting. The energy curves look like straight lines in Fig. 6 (b) and (d). To see these curves more clearly, they are also plotted in Fig. 6 (a) and (c). Comparing with the previous synchronized results, one cannot tell much difference for the unsynchronized results. This indicates that the proposed regularization method is very robust with respect to grid rotation and shifting. In other words, the grid artifact or artificial grid energy is completely eliminated in the present finite difference PB model.

Table 2
Electrostatic free energy (kcal/mol) of the protein 1XMK and the corresponding CPU time in seconds for different mesh sizes. The CPU time of the regularization and trilinear methods differs in three major parts: surface generation, source generation, and algebraic solution of the linearized PB equation.

			CPU time (s)					
	Energy		Surface generation		Source generation		Solving PB Eq.	
h	REG	TRI	REG	TRI	REG	TRI	REG	TRI
1.0	-360.9206	-315.4097	9.68E−2	4.74E−2	5.50E+1	5.36E+0	1.06E+1	5.73E+0
0.5	-343.8260	-357.6610	7.78E−1	3.71E−1	3.92E+2	2.15E+1	1.15E+2	5.91E+1
0.25	-338.8072	-345.4913	7.06E+0	3.15E+0	2.97E+3	8.79E+1	1.67E+3	8.04E+2
0.125	-337.5360	-342.2154	6.96E+1	2.73E+1	2.33E+4	3.94E+2	2.62E+4	1.15E+4

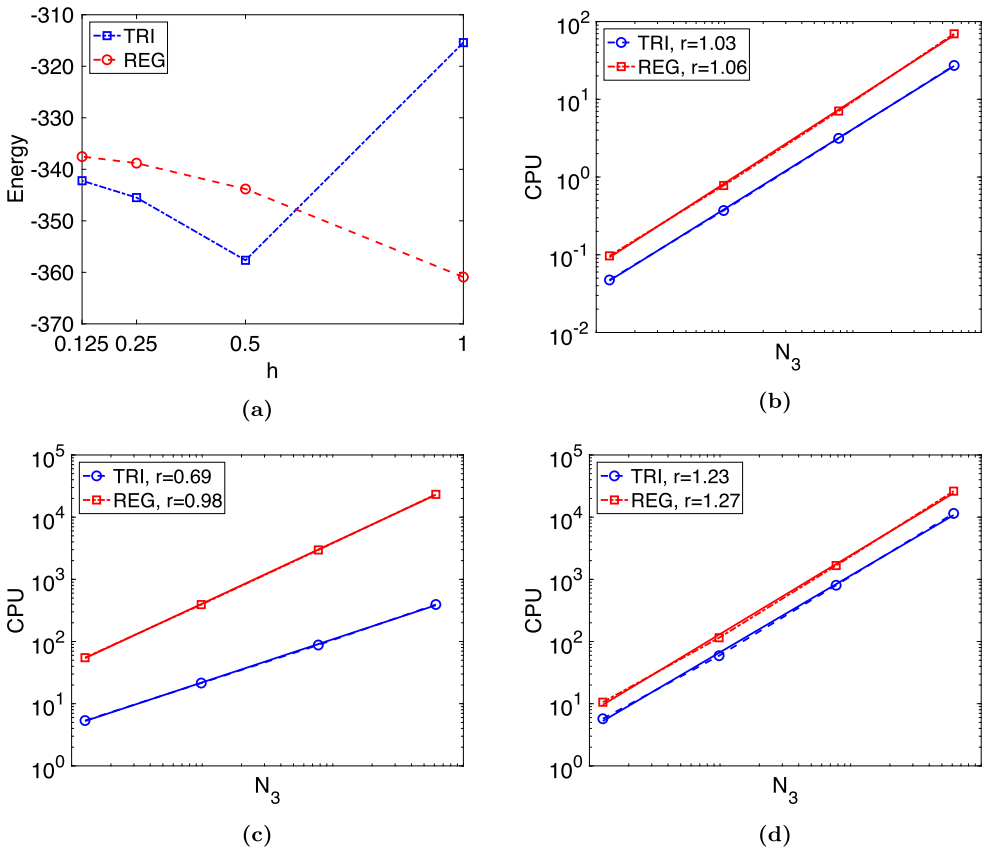

Fig. 7. Convergence and efficiency tests for the protein 1XMK, where h is spacing and N_3 is the total degree of freedom. (a) Convergence of the electrostatic free energy with respect to h; (b) CPU time against N_3 for surface generation; (c) CPU time against N_3 for source generation; (d) CPU time against N_3 for solving the PB equation. In all CPU plots, the solid lines represent the least-squares fitted linear trends. The corresponding complexity ratio r is also reported.

3.4. Convergence and efficiency

We next study the convergence and efficiency by considering a protein with protein databank (PDB) ID: 1XMK. By using a fixed domain $\Omega = [6, 56] \times [3, 56] \times [9, 54]$, different mesh sizes are examined with $h = 1$, 0.5, 0.25, and 0.125. The calculated electrostatic free energy and corresponding CPU time are reported in Table 2.

We first examine the energy convergence. It can be seen in Fig. 7 (a) that the regularization method converges monotonically to certain limit, while the energy curve of the trilinear method is not monotonic. As h becomes smaller, the gap between energies predicted by the two methods is shrinking. It looks like that both methods will eventually converge to the same limit, as $h \to 0$. Moreover, the convergence speed of each method could be assessed by considering the energy difference after a mesh refinement. For instance, it can be seen from Table 2 that the energy difference between $h = 0.125$ and $h = 0.25$ is about 1.33 and 3.28, respectively, for the regularization and trilinear method. In fact, the energy difference of the regularization is always less than that of the trilinear method. This suggests that the regularization method converges faster than the trilinear method.

We next investigate the CPU time. Based on the same super-Gaussian PB model, the regularization and trilinear methods share a lot of common components in implementation. Thus, in biomolecular simulations, the execution time of two methods are different mainly in three parts, i.e., surface generation, source generation, and algebraic solution of the PB equation. We will examine these three parts one by one.

First, in the surface generation, the CPU time for setting up the GCS function S is measured in the trilinear method. For the regularization, extra time is needed for calculating the gradient of S. Thus, it is seen in Table 2 that the CPU time of the trilinear method is less than half of that of the regularization for this part. In Fig. 7 (b), the CPU time in surface generation is plotted against the total degree of freedom $N_3 = N_x \times N_y \times N_z$. Moreover, a least square linear fitting is conducted in log-log scale to determine the order of complexity, i.e., CPU=$O(N_3^r)$. With $r = 1.03$ and $r = 1.06$, respectively, for trilinear and regularization methods, one can conclude that the CPU time in surface generation scales linearly with respect to N_3. This is consistent with the finding in Ref. [44].

Second, the regularization is much more expensive than the trilinear method in source generation. For the trilinear method, the main execution time is spent on searching for the grid cubes that contain partial charges. The cost of trilinear interpolation depends on the total number of charges, which does not change when one refines the mesh. Thus, the source generation for the trilinear method is cheap and does not grow too much with respect to N_3. This can be observed in Fig. 7 (c) that the trilinear CPU scales as $O(N_3^{0.69})$ for source generation. For the regularization method, the computation of two source terms in Eq. (34) involves a lot of finite difference approximations and analytical differentiations, and has to be carried on all nodes throughout the domain. Thus, such computations are expensive. Fortunately, as can be seen in Fig. 7 (c) that the complexity of this part scales linearly with respect to N_3. Consequently, for the finest mesh with $h = 0.125$, the source generation is not the most time consuming part of the regularization method.

Third, the same matrix and iterative algebraic solver (biconjugate gradient algorithm) are used in the regularization and trilinear methods. Since the source term of the regularization becomes more complicated, the algebraic solution demands more CPU time than that of the trilinear method. It can be observed in Fig. 7 (d) that the complexity order of this part is the largest, about $O(N_3^{1.25})$. Thus, iterative solution of the PB equation shall be the most expensive part for large systems.

Combining three CPU parts, one can conclude that the regularization is more expensive than the trilinear method. In particular, the CPU time of the regularization is about 4 to 6 times larger than that of the trilinear method for a fixed h. However, it can be observed from Table 2 that the regularization provides a more accurate energy prediction. In fact, by using $h = 0.5$, the energy calculated by the regularization is more accurate than that of the trilinear method with $h = 0.25$, while the total CPU of the regularization at $h = 0.5$ is smaller than that of the trilinear method at $h = 0.25$. Therefore, by allowing to use a coarse mesh, the regularization is actually more cost-efficient than the trilinear method.

3.5. Electrostatic free energies of proteins

We finally consider the electrostatic free energy of 25 proteins, which are randomly chosen from a protein data set studied in Ref. [10]. The energies calculated by using trilinear and regularization methods based on three mesh sizes are reported in Table 3. For both methods, when h varies from 1, 0.5, to 0.25, the energies converge to the finest mesh $h = 0.25$. Also, a general agreement between the regularization and trilinear results at $h = 0.25$ can be seen.

To better illustrate the accuracies of two methods, we plot the relative energy differences in Fig. 8. First, for one method and a protein, the self-convergence is tested by taking the energy predicted by $h = 0.25$ as the reference value E_{ref}. Then the errors for $h = 1$ and 0.5 are assessed by the relative energy difference $\frac{E-E_{ref}}{E_{ref}}$. Such differences are plotted in Fig. 8 (a) and (b), respectively, for regularization and trilinear methods. While both methods are self-convergent, it is observed that the errors of the regularization are much smaller than those of the trilinear method. Moreover, the convergence pattern of the trilinear method is always oscillatory, i.e., $\frac{E-E_{ref}}{E_{ref}}$ is positive for $h = 1$ and negative for $h = 0.5$. For the regularization, the convergence could be either oscillatory or monotonic. In Fig. 8 (c), we take the regularization energy at $h = 0.25$ as the reference E_{ref} to examine the accuracy of the trilinear method. It can be seen that the error of the trilinear method decreases as h is reduced from 1 to 0.25. This suggests that the trilinear energy converges to the regularization energy. Therefore, in all cases, the regularization method is more accurate than the trilinear method in calculating electrostatic free energies of proteins.

A comparison between the present super-Gaussian PB model and the sharp interface PB model [5,25,38] is also considered. For this purpose, the MIBPB web server (https://weilab.math.msu.edu/MIBPB/) is employed. The MIBPB method is a second order convergent PB solver that rigorously treats charge singularities, as well as geometrical singularities involved in the molecular surface [11,19]. Because the MIBPB energy converges very fast with respect to h, we calculate the MIBPB energies by just using $h = 0.5$ Å. Except for one protein (PDB ID: 1CBN) that the MIBPB web server reports an error, electrostatic free energies of other 24 proteins are obtained. A comparison between the MIBPB energy and those calculated by the trilinear and regularization methods at $h = 0.25$ Å is shown in Fig. 9. It can be seen that the super Gaussian PB energy is usually weaker than the MIBPB energy, i.e., the former has a smaller magnitude. By taking the MIBPB energy as the independent variable x, a least-squares linear fitting is conducted for the regularization energy y, which gives a linear function $y = 0.4960x + 253.7996$. The energy difference here is mainly due to the underlying PB models. In the super Gaussian PB model, the ionic strength is scaled by $(1 - S)$. Thus, the ionic strength is increasing in Ω_t and is in full strength only for

Table 3
Electrostatic free energy (kcal/mol) of 25 proteins calculated by the trilinear and regularization methods. For each case, three mesh sizes are tested.

PDB -ID	Atom number	Electrostatic free energy					
		Trilinear			Regularization		
		h=1	h=0.5	h=0.25	h=1	h=0.5	h=0.25
1AHO	962	-211.2636	-248.1750	-238.2240	-228.5431	-237.2506	-233.1279
1C75	985	-651.9008	-690.7268	-680.2420	-675.1506	-675.4569	-673.4696
1J0P	1597	-940.1217	-1014.6565	-993.8498	-979.8352	-989.0494	-981.9095
1TG0	1029	-1793.6217	-1835.2729	-1824.0573	-1830.3678	-1811.3032	-1814.1010
1X8Q	2815	-542.4612	-630.3998	-606.4268	-583.1636	-602.5760	-593.1467
1CBN	639	-73.3233	-89.5271	-84.2385	-86.0809	-84.8137	-82.0041
1G6X	888	-609.7857	-645.4802	-635.0196	-605.9305	-629.6001	-628.4701
1IUA	1207	-210.3572	-248.1770	-237.2259	-238.8805	-236.8073	-231.6764
1L9L	1226	-1674.2821	-1723.6808	-1709.9624	-1700.3133	-1695.7472	-1699.1604
1M1Q	1265	-821.4304	-877.8622	-863.3227	-840.7822	-857.3641	-853.9345
1NWZ	1912	-725.5706	-791.8178	-774.0107	-793.2792	-769.0200	-763.5936
1OK0	1076	-470.4574	-506.2608	-496.1842	-490.7971	-493.6079	-490.7643
1TQG	1660	-779.2361	-823.3382	-809.0508	-826.5917	-805.5121	-800.8380
1VB0	913	-251.5802	-286.9582	-275.2815	-282.2747	-275.0322	-269.7315
1VBW	1056	-828.6500	-865.4070	-854.4223	-832.8698	-848.7617	-847.4552
1W0N	1756	-574.7337	-634.0710	-619.6113	-593.3313	-615.8816	-611.0642
1X6X	1732	-397.5570	-457.1552	-439.1164	-440.1687	-438.2086	-430.1511
1XMK	1268	-315.4097	-357.6610	-345.4913	-360.9206	-343.8260	-338.8072
1ZUU	868	-478.8526	-515.3313	-505.1730	-498.0964	-501.0076	-498.9569
1ZZK	1252	-415.9231	-462.4152	-449.9412	-434.1725	-445.3290	-442.6274
2FDN	731	-846.6849	-871.4588	-864.1814	-856.0617	-859.8413	-858.8722
2FMA	924	-333.8799	-371.9273	-362.3285	-361.9970	-360.0563	-356.6664
2FWH	1830	-661.0665	-714.0925	-698.4507	-678.1039	-695.4554	-689.5961
2H5C	2755	-714.1789	-776.7490	-757.7038	-739.3847	-755.2416	-747.9770
2IDQ	1596	-505.4323	-554.5314	-541.3744	-545.1966	-537.5704	-533.4612

outside Ω_t or 3 Å away from the protein. For the sharp interface PB model, the full ionic strength is applied immediately outside the protein, so that the solute-solvent electrostatic interactions are stronger. This is why the MIBPB energy becomes stronger. Nevertheless, certain correlations can still be seen in Fig. 9. In fact, the Pearson correlation coefficient between the MIBPB and regularization energies is as high as 0.791.

4. Conclusion

In this paper, a regularization method is introduced for the first time in the literature for treating charge singularities in the heterogeneous dielectric Poisson-Boltzmann (PB) model. The heterogeneous dielectric PB models, including Gaussian PB model [10,30,31] and super-Gaussian PB model [22,36], have achieved a great success in various biological applications involving the electrostatic analysis. However, the singular charges are simply handled by the trilinear method in the existing studies, which essentially approximate the delta function by a finite quantity. This results in a large grid artifact and a slow convergence in energy calculations. In the proposed regularization, the singularities are analytically captured by Green's functions so that the grid artifact is completely eliminated and a fast energy convergence is observed.

The proposed regularization for the super-Gaussian PB model is generalized from that for the diffuse interface PB model [37,44,45]. The key in the regularization formulation is the dual decomposition first presented in [45], i.e., besides the decomposition of the potential into Coulomb and reaction-field components, the dielectric function is also split into a constant based and a space varying part. Nevertheless, in order to develop a valid regularization formulation for the heterogeneous dielectric PB models, an extensive study has been carried out in this paper to deal with theoretical and numerical difficulties. Mathematically, a rigorous analysis has to be conducted to examine the new source term of the regularized PB equation, because the arguments for the diffuse interface PB model [37,44,45] cannot be directly extended to the present study. In particular, the shifted dielectric function is almost vanishing in a small neighborhood around each center, while it is identically zero in the diffuse interface case [37,44,45]. Consequently, it is shown that a super-Gaussian density, instead of a Gaussian one, is required for the validity of the regularization formulation. Moreover, the proposed regularization is proved to be well-posed and its solution is C^2 continuous throughout the domain. Numerically, the standard finite difference algorithm is simply employed to solve the regularized PB equation. In order to secure a more accurate implementation, analytical differentiations, instead of numerical approximations, are adopted whenever possible in calculating the new source term involving a space dependent dielectric function.

In numerical experiments, various examples are considered to validate the proposed regularization and to compare with the trilinear method. The trilinear method heavily relies on an error cancellation to produce acceptable results in energy calculations, i.e., the huge errors due to approximating singular charges by finite quantities are the same in both solvent and vacuum states, so that the electrostatic free energy calculated as the difference of two states can be free of such errors. It is shown in the present study that when error cancellation is unavailable, such as in case of an unsynchronized

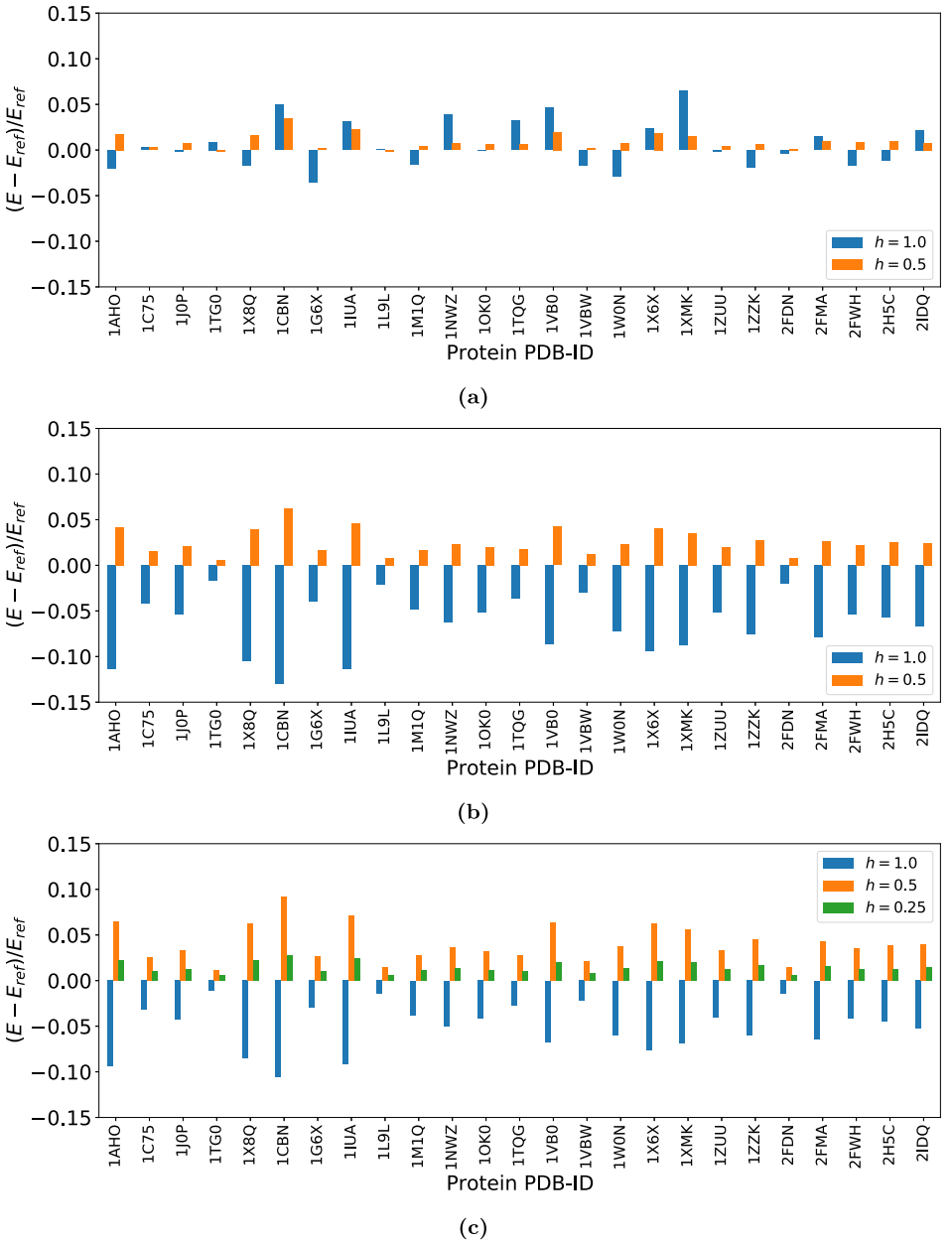

Fig. 8. (a) Self-convergence of the regularization energy. Here E_{ref} is the regularization energy at $h = 0.25$, and E is the regularization energy at $h = 1$ or 0.5; (b) Self-convergence of the trilinear energy. Here E_{ref} is the trilinear energy at $h = 0.25$, and E is the trilinear energy at $h = 1$ or 0.5; (c) Convergence of the trilinear energy toward the regularization energy. Here E_{ref} is the regularization energy at $h = 0.25$, and E is the trilinear energy at $h = 1$, 0.5, or 0.25.

rotation of shifting, an incredibly large error will be produced, which completely ruins the biomolecular simulation. On the contrary, the unsynchronized rotation or shifting does not affect the regularization at all, and the grid artifact or artificial grid energy is completely eliminated in the present finite difference PB model. When the error cancellation is allowed in the trilinear method, the regularization is still more accurate by producing a fast energy convergence. Nevertheless, the regularization is about 4 - 6 times more expensive than the trilinear method, based on the same degree of freedom. Fortunately, the regularization method could yield a better energy prediction by using a courser mesh. From cost-efficiency point of view, the regularization could still be faster than the trilinear method, in order to achieve a certain precision in energy calculation.

In our future studies, the development of more accurate and efficient numerical algorithms for solving the regularized PB equation of the super-Gaussian model will be investigated. This regularization will also be implemented in the super-Gaussian module of the DelPhi package [36], and its application to biological problems will be explored.

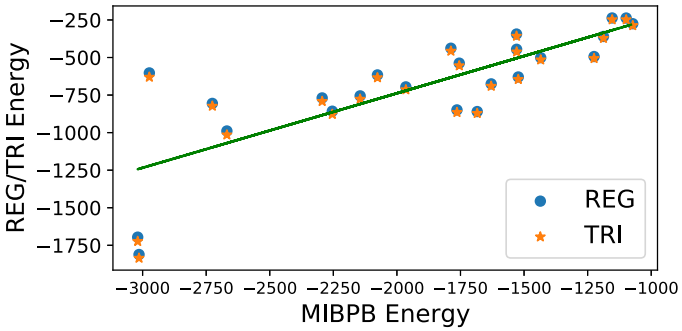

Fig. 9. Comparison of the electrostatic free energies of 24 proteins obtained from the present super-Gaussian PB model and the MIBPB sharp-interface PB model. For the present model, both trilinear and regularization energies are shown.

Funding

The research of Zhao was supported in part by the National Science Foundation (NSF) grants DMS-1812930 and DMS-2110914. The research of Alexov was supported in part by the National Institutes of Health (NIH) grant R01GM093937 and NSF grant DMS-1812597.

CRediT authorship contribution statement

Siwen Wang: Software, Validation, Visualization, Writing – original draft. **Yuanzhen Shao:** Formal analysis, Methodology, Writing – original draft. **Emil Alexov:** Conceptualization, Funding acquisition, Writing – review & editing. **Shan Zhao:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Smooth solute-solvent boundary

Consider a protein with a total N_m atoms, with the center and radius of each atom being $\mathbf{r}_n$ and r_n, respectively. By treating each atom as a hard sphere, the Van der Waals (VdW) surface is defined as the smallest envelope enclosing the union of all spheres. By augmenting each atomic radius by 3 Å, another VdW surface can be defined, which will be referred as to an extended solute-accessible surface (ESAS). The diffuse interface function $S(\mathbf{r})$ should satisfy a subdomain requirement that $S(\mathbf{r}) = 1$ inside the VdW surface and $S(\mathbf{r}) = 0$ outside the ESAS surface. In the original super-Gaussian PB model [22], $S(\mathbf{r})$ is generated by the minimal molecular surface (MMS) [6,7], while in the present study, a more efficient Gaussian convolution surface (GCS) [44] is used to calculate $S(\mathbf{r})$.

Physically, the MMS is defined as the unique surface that is of the smallest surface area, while encloses all VdW spheres. Mathematically, the MMS can be derived from the Euler-Lagrange variation of the surface free energy, which leads to a mean curvature flow equation subject to the subdomain requirement [6,7]

$$\frac{\partial S}{\partial t}(\mathbf{r}, t) = \nabla \cdot \left(\frac{\nabla S(\mathbf{r}, t)}{\|\nabla S(\mathbf{r}, t)\|} \right) + V(\mathbf{r}), \tag{A.1}$$

where $V(\mathbf{r})$ represents the total external potential, and the diffuse interface function $S(\mathbf{r})$ is obtained as the steady state solution of Eq. (A.1). An unconditionally stable alternating direction implicit (ADI) algorithm has been developed in [40] for Eq. (A.1), whose computational complexity in each time step scales as $O(N_3)$ for N_3 being the spatial degree of freedoms. A sharp MMS can be obtained as an iso-surface of $S(\mathbf{r})$ [6,7].

In the GCS model [44], a standard solute-accessible surface (SAS) is first considered, which is a VdW surface after augmenting each atomic radius by 1.5 Å. A Heaviside function is defined with $H(\mathbf{r}) = 1$ inside the SAS and $H(\mathbf{r}) = 0$ outside the SAS. One then convolutes $H(\mathbf{r})$ with a Gaussian kernel function

$$K(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2} \right). \tag{A.2}$$

In real computations, the Heaviside function and Gaussian kernel are defined discretely on a uniform mesh. The discrete convolution in three-dimensions is realized through a fast Fourier transform (FFT) algorithm with the total complexity being $O(N_3)$. After convolution, the subdomain requirement is enforced in the post-processing, and the resulted diffuse interface function is $S(\mathbf{r})$ [44].

Appendix B. The manufactured example

In the section 3, a manufactured example is considered for a one-atom system. The derivation details of this example are offered here. Following the proposed regularization, after the dual decomposition, the reaction-field potential satisfies a regularized Poisson equation

$$-\nabla\cdot(\epsilon\nabla u_{RF}) = \nabla\epsilon\cdot\nabla G + f, \tag{B.1}$$

where the Green's function G is known and the reaction-field potential u_{RF} is assumed to be given

$$u_{RF}(\mathbf{r}) = -C\exp\left(-\frac{|\mathbf{r}|^2}{r_e^2}\right), \quad G(\mathbf{r}) = C\frac{q_0}{\epsilon_m|\mathbf{r}|}, \tag{B.2}$$

with the constant $C = \frac{e_c^2}{k_B T}$. The unknown extra source term f can be determined by plugging (B.2) into (B.1).

The present problem can be recasted into a spherical coordinate (r,θ,φ). Due to the rotational symmetry of the domain, all involved functions do not change in θ and φ directions, so that they can be regarded as functions in the radical direction $r = |\mathbf{r}|$ only. Consequently, the regularized Poisson equation (B.1) can be reduced to an ordinary differential equation (ODE) with respect to r. For instance, the left-hand side can be expanded as

$$-\nabla\cdot(\epsilon(\mathbf{r})\nabla u_{RF}(\mathbf{r})) = -\frac{1}{r^2}\frac{d}{dr}\left(r^2\epsilon\frac{du_{RF}}{dr}\right) = -\frac{2}{r}\epsilon\frac{du_{RF}}{dr} - \frac{d\epsilon}{dr}\frac{du_{RF}}{dr} - \epsilon\frac{d^2u_{RF}}{dr^2}.$$

Then $f(r)$ can be expressed by

$$f(r) = -\frac{2}{r}\epsilon\frac{du_{RF}}{dr} - \frac{d\epsilon}{dr}\frac{du_{RF}}{dr} - \epsilon\frac{d^2u_{RF}}{dr^2} - \frac{d\epsilon}{dr}\frac{dG}{dr}. \tag{B.3}$$

Based on Eq. (B.2), the derivatives of G and u_{RF} can be given as

$$\frac{du_{RF}}{dr} = \frac{2Cr}{r_e^2}e^{-\frac{r^2}{r_e^2}}, \quad \frac{d^2u_{RF}}{dr^2} = \frac{2C}{r_e^2}e^{-\frac{r^2}{r_e^2}} - \frac{4Cr^2}{r_e^4}e^{-\frac{r^2}{r_e^2}}, \quad \frac{dG}{dr} = -\frac{Cq_0}{\epsilon_m}r^{-2}.$$

The derivative of ϵ is calculated as

$$\frac{d\epsilon}{dr} = \frac{dS}{dr}\epsilon_{in}(r) + S\frac{d\epsilon_{in}(r)}{dr} - \epsilon_{out}\frac{dS}{dr} \tag{B.4}$$

where $\epsilon_{in}(r)$ is given by Eq. (42) with $r = |\mathbf{r}|$. Its derivative can be calculated as

$$\frac{d\epsilon_{in}}{dr} = (\epsilon_{gap} - \epsilon_m)\frac{4r^3}{r_i^4}e^{-\frac{r^4}{r_i^4}}.$$

The derivative of S is

$$\frac{dS}{dr} = -\frac{k}{2(r_e - r_i)}\operatorname{sech}^2(k(\frac{r - r_i}{r_e - r_i} - \frac{1}{2})).$$

With these derivatives, one can compute $\frac{d\epsilon}{dr}$ by Eq. (B.4). Finally, the source $f(r)$ can be calculated by Eq. (B.3) at every grid node in Ω.

References

[1] A. Abrashkin, D. Andelman, H. Orland, Dipolar Poisson-Boltzmann equation: ions and dipoles close to charge interfaces, Phys. Rev. Lett. 99 (7) (2007) 077801.

[2] R.A. Adams, J.J.F. Fournier, Sobolev Spaces, second edition, Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003.

[3] S. Ahmed Ullah, S. Zhao, Pseudo-transient ghost fluid methods for the Poisson-Boltzmann equation with a two-component regularization, Appl. Math. Comput. 380 (2020) 125267.

[4] S. Amihere, W. Geng, S. Zhao, Benchmarking electrostatic free energy of the nonlinear Poisson-Boltzmann model for the Kirkwood sphere, Commun. Inf. Syst. 22 (2022).

[5] N.A. Baker, Improving implicit solvent simulations: a Poisson-centric view, Curr. Opin. Struct. Biol. 15 (2) (2005) 137–143.

[6] P. Bates, Z. Chen, Y. Sun, G.-W. Wei, S. Zhao, Geometric and potential driving formation and evolution of biomolecular surfaces, J. Math. Biol. 59 (2) (2009) 193–231.

[7] P.W. Bates, G.-W. Wei, S. Zhao, Minimal molecular surfaces and their applications, J. Comput. Chem. 29 (3) (2008) 380–391.

[8] P. Benner, V. Khoromskaia, B. Khoromskij, C. Kweyu, M. Stein, Regularization of Poisson–Boltzmann type equations with singular source terms using the range-separated tensor format, SIAM J. Sci. Comput. 43 (1) (2021) A415–A445.

[9] Q. Cai, J. Wang, H.-K. Zhao, R. Luo, On removal of charge singularity in Poisson–Boltzmann equation, J. Chem. Phys. 130 (14) (2009) 04B608.

[10] A. Chakravorty, Z. Jia, L. Li, S. Zhao, E. Alexov, Reproducing the ensemble average polar solvation energy of a protein from a single structure: Gaussian-based smooth dielectric function for macromolecular modeling, J. Chem. Theory Comput. 14 (2) (2018) 1020–1032.
[11] D. Chen, Z. Chen, C. Chen, W. Geng, G.W. Wei, MIBPB: a software package for electrostatic analysis, J. Comput. Chem. 32 (2011) 657–670.
[12] L. Chen, M.J. Holst, J. Xu, The finite element approximation of the nonlinear Poisson–Boltzmann equation, SIAM J. Numer. Anal. 45 (6) (2007) 2298–2320.
[13] L.-T. Cheng, J. Dzubiella, J.A. McCammon, B. Li, Application of the level-set method to the implicit solvation of nonpolar molecules, J. Chem. Phys. 127 (8) (2007) 084503.
[14] I.-L. Chern, J.-G. Liu, W.-C. Wang, et al., Accurate evaluation of electrostatics for macromolecules in solution, Methods Appl. Anal. 10 (2) (2003) 309–328.
[15] R. Chowdhury, R. Egan, D. Bochkov, F. Gibou, Efficient calculation of fully resolved electrostatics around large biomolecules, J. Comput. Phys. 448 (2022) 110718.
[16] S. Dai, B. Li, J. Lu, Convergence of phase-field free energy and boundary force for molecular solvation, Arch. Ration. Mech. Anal. 227 (1) (2018) 105–147.
[17] W. Deng, J. Xu, S. Zhao, On developing stable finite element methods for pseudo-time simulation of biomolecular electrostatics, J. Comput. Appl. Math. 330 (2018) 456–474.
[18] S.J. Fromm, Potential space estimates for Green potentials in convex domains, Proc. Am. Math. Soc. 119 (1) (1993) 225–233.
[19] W. Geng, S. Yu, G. Wei, Treatment of charge singularities in implicit solvent models, J. Chem. Phys. 127 (11) (2007) 114106.
[20] W. Geng, S. Zhao, A two-component matched interface and boundary (mib) regularization for charge singularity in implicit solvation, J. Comput. Phys. 351 (2017) 25–39.
[21] D. Gilbarg, N.S. Trudinger, Elliptic Partial Differential Equations of Second Order, second edition, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 224, Springer-Verlag, Berlin, 1983.
[22] T. Hazra, S.A. Ullah, S. Wang, E. Alexov, S. Zhao, A super-Gaussian Poisson–Boltzmann model for electrostatic free energy calculation: smooth dielectric distribution for protein cavities and in both water and vacuum states, J. Math. Biol. 79 (2) (2019) 631–672.
[23] M. Holst, J.A. McCammon, Z. Yu, Y.C. Zhou, Y. Zhu, Adaptive finite element modeling techniques for the Poisson-Boltzmann equation, Commun. Comput. Phys. 11 (1) (2012) 179–214.
[24] M.J. Holst, The Poisson-Boltzmann Equation: Analysis and Multilevel Numerical Solution, PhD thesis, UIUC, 1994.
[25] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, Science 268 (5214) (1995) 1144–1149.
[26] L. Hu, G.-W. Wei, Nonlinear Poisson equation for heterogeneous media, Biophys. J. 103 (4) (2012) 758–766.
[27] B.N. Khoromskij, Range-separated tensor decomposition of the discretized Dirac delta and elliptic operator inverse, J. Comput. Phys. 401 (2020) 108998.
[28] A. Kucherova, S. Strango, S. Sukenik, M. Theillard, Computational modeling of protein conformational changes - application to the opening sars-cov-2 spike, J. Comput. Phys. 444 (2021) 110591.
[29] A. Lee, W. Geng, S. Zhao, Regularization methods for the Poisson-Boltzmann equation: comparison and accuracy recovery, J. Comput. Phys. 426 (2021) 109958.
[30] L. Li, C. Li, E. Alexov, On the modeling of polar component of solvation energy using smooth Gaussian-based dielectric function, J. Theor. Comput. Chem. 13 (03) (2014) 1440002.
[31] L. Li, C. Li, Z. Zhang, E. Alexov, On the dielectric “constant” of proteins: smooth dielectric function for macromolecular modeling and its implementation in delphi, J. Chem. Theory Comput. 9 (4) (2013) 2126–2136.
[32] M. Mirzadeh, M. Theillard, A. Helgadottir, D. Boy, F. Gibou, An adaptive, finite difference solver for the nonlinear Poisson-Boltzmann equation with applications to biomolecular computations, Commun. Comput. Phys. 13 (1) (2013) 150–173.
[33] A. Nicholls, B. Honig, A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation, J. Comput. Chem. 12 (4) (1991) 435–445.
[34] A. Nicholls, D.L. Mobley, J.P. Guthrie, J.D. Chodera, C.I. Bayly, M.D. Cooper, V.S. Pande, Predicting small-molecule solvation free energies: an informal blind test for computational chemistry, J. Med. Chem. 51 (4) (2008) 769–779. PMID: 18215013.
[35] H. Nymeyer, H.-X. Zhou, A method to determine dielectric constants in nonhomogeneous systems: application to biological membranes, Biophys. J. 94 (4) (2008) 1185–1193.
[36] S.K. Panday, M.H. Shashikala, A. Chakravorty, S. Zhao, E. Alexov, Reproducing ensemble averaged electrostatics with super-Gaussian-based smooth dielectric function: application to electrostatic component of binding energy of protein complexes, Commun. Inf. Syst. 19 (4) (2019).
[37] Y. Shao, M. McGowan, S. Wang, E. Alexov, S. Zhao, Convergence of a diffuse interface Poisson-Boltzmann (pb) model to the sharp interface pb model: a unified regularization formulation, Appl. Math. Comput. (2022), submitted for publication.
[38] K.A. Sharp, B. Honig, Electrostatic interactions in macromolecules - theory and applications, Annu. Rev. Biophys. Biophys. Chem. 19 (1990) 301–332.
[39] X. Song, An inhomogeneous model of protein dielectric properties: intrinsic polarizabilities of amino acids, J. Chem. Phys. 116 (21) (2002) 9359–9363.
[40] W. Tian, S. Zhao, A fast alternating direction implicit algorithm for geometric flow equations in biomolecular surface generation, Int. J. Numer. Methods Biomed. Eng. 30 (4) (2014) 490–516.
[41] D. Voges, A. Karshikoff, A model of a local dielectric constant in proteins, J. Chem. Phys. 108 (5) (1998) 2219–2227.
[42] L. Wang, L. Li, E. Alexov, pka predictions for proteins, rnas, and dnas with the Gaussian dielectric function using delphi pka, Proteins, Struct. Funct. Bioinform. 83 (12) (2015) 2186–2197.
[43] L. Wang, M. Zhang, E. Alexov, DelPhiPKa web server: predicting pKa of proteins, RNAs and DNAs, Bioinformatics 32 (4) (10 2015) 614–615.
[44] S. Wang, E. Alexov, S. Zhao, On regularization of charge singularities in solving the Poisson-Boltzmann equation with a smooth solute-solvent boundary, Math. Biosci. Eng. 18 (2) (2021) 1370–1405.
[45] S. Wang, A. Lee, E. Alexov, S. Zhao, A regularization approach for solving Poisson’s equation with singular charge sources and diffuse interfaces, Appl. Math. Lett. 102 (2020) 106144.
[46] D. Xie, New solution decomposition and minimization schemes for Poisson-Boltzmann equation in calculation of biomolecular electrostatics, J. Comput. Phys. 275 (2014) 294–309.
[47] C. Xue, S. Deng, Unified construction of Green’s functions for Poisson’s equation in inhomogeneous media with diffuse interfaces, J. Comput. Appl. Math. 326 (2017) 296–319.
[48] Y. Zhao, Y.-Y. Kwan, J. Che, B. Li, J.A. McCammon, Phase-field approach to implicit solvation of biomolecules with Coulomb-field approximation, J. Chem. Phys. 139 (2) (2013) 024111.
[49] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn, M. Levitt, Finite-difference solution of the Poisson–Boltzmann equation: complete elimination of self-energy, J. Comput. Chem. 17 (11) (1996) 1344–1351.