A SUPER-GAUSSIAN POISSON-BOLTZMANN MODEL FOR ELECTROSTATIC SOLVATION ENERGY CALCULATION: SMOOTH DIELECTRIC DISTRIBUTIONS FOR PROTEIN CAVITIES AND IN BOTH WATER AND VACUUM STATES

by

TANIA HAZRA

SHAN ZHAO, COMMITTEE CHAIR
DAVID HALPERN
KABE MOEN
BRENDAN AMES
WEIHUA SU

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2018

ABSTRACT

Calculations of electrostatic potential and solvation energy of macromolecules are essential for understanding the mechanism of many biological processes. In the classical implicit solvent Poisson-Boltzmann (PB) model, the macromolecule and water are modeled as two-dielectric media with a sharp border. However, the dielectric property of interior cavities and ion-channels is difficult to model in a two-dielectric setting. In fact, whether there are water molecules or cavity-fluid inside a protein cavity remains to be an experimental challenge. Physically, this uncertainty affects the subsequent solvation free energy calculation. In order to compensate this uncertainty, a novel super-Gaussian dielectric PB model is introduced in this work, which devices an inhomogeneous dielectric distribution to represent the compactness of atoms and characterize empty cavities via a gap dielectric value. Moreover, the minimal molecular surface level set function is adopted so that the dielectric profile remains to be smooth when the protein is transfer from water phase to vacuum. A nice feature of this new model is that as the order of super-Gaussian function approaches the infinity, the dielectric distribution reduces a piecewise constant of the two-dielectric model. Mathematically, a simple effective dielectric constant analysis is introduced in this work to benchmark the dielectric model and select optimal parameter values. Computationally, a pseudo-time alternative direction implicit (ADI) algorithm is utilized for solving the super-Gaussian PB equation, which is found to be unconditionally stable in a smooth dielectric setting. Solvation free energy calculation of a Kirkwood sphere and various proteins is carried out to validate the super-Gaussian model and ADI algorithm. One macromolecule with both cavity-fluids and empty cavities is employed to demonstrate how the cavity uncertainty in protein structure can be bypassed through dielectric modeling in biomolecular electrostatic analysis.

Keywords: Poisson-Boltzmann equation, Gaussian dielectric model, super-Gaussian dielectric model, minimal molecular surface (MMS), pseudo-time approach, alternating direction implicit (ADI), electrostatic free energy.

DEDICATION

To my family, mentors and friends, and to all those who supported me to fulfill my dream.

LIST OF ABBREVIATIONS AND SYMBOLS

$\bar{\kappa}$	Modified Debye Huckel parameter
$\Delta\hat{\epsilon}$	Difference between EDCs in water and vacuum states
Δ	Laplacian
δ	Dirac delta function
ϵ	Dielectric constant distribution of the entire solute-solvent system
ϵ_2	Two dielectric model
ϵ_G	Dielectric constant distribution generated from Gaussian function
ϵ_m	Dielectric constant of the atomic region
ϵ_s	Dielectric constant of the solvent region
ϵ_{gap}	Dielectric constant of the cavity fluid
ϵ_{in}	Dielectric distribution of inside molecular part consists of ϵ_m and ϵ_{gap}
ϵ_{max}	Maximal dielectric value in the macromolecule-cavity
ϵ_{out}	Dielectric constant of outside molecular part: solvent or vacuum
ϵ_{sG}	Dielectric distribution generated from super-Gaussian function
Γ	Molecular interface
$\hat{\epsilon}$	Effective dielectric constant
κ	Debye Huckel parameter

$\mathbb{R}$	Real Numbers
Ω	Entire solute-solvent domain
Ω_b	Boundary region near the molecular interface
Ω_m	Solute (molecular) region
Ω_s	Solvent Region
ρ_m	Source term in the non-linear Poisson Boltzmann equation
σ	Relative variance
$\tilde{\epsilon_G}$	Gaussian model in vacuum phase (surface-cut model)
$\vec{r_j}$	Position vector of the center of j^{th} atom
$\vec{r}$	Position vector at (x_i, y_j, z_k)
a	Right end point of x, y, z arrays in one atom model
D	Distance between two atoms in four atoms model
d	Influence domain extension function
e_c	Fundamental charge
g_0	Total density of a macromolecule produced by Gaussian function
g_0^s	Total density of a macromolecule produced by super-Gaussian function
g_i	Density function of i^{th} atom produced by Gaussian function
g_i^s	Density function of i^{th} atom produced by super-Gaussian function
I_s	Ionic strength
K_B	Boltzmann Constant

m	Power raised to the super-Gaussian function
N_m	Total number of atoms present in a macromolecule
q_j	Partial charge of j^{th} atom
R_0	Radius of the atom in one atom model
R_i	Radius of i^{th} atom
S	Hyper-surface function
T	Room temperature
T_e	Stopping time in numerical computation
u	Electrostatic potential for the solvent (water) phase
u_0	Electrostatic potential for the vacuum phase
ADI	Alternating Direction Implicit
CNS	Coupled Nonlinear Solvation
EDC	Effective Dielectric Analysis
IIM	Immersed Interface Method
MD	Molecular Dynamics
MIB	Matched Interface and Boundary
MMS	Minimal Molecular Surface
MSMS	Michel Sanner Molecular Surface
PB	Poisson-Boltzmann
PDE	Partial Differential Equation

RPB	Regularized Poisson Boltzmann
SAS	Solvent Accessible Surface
SES	Solvent Excluded Surface
VDW	Van der Wall

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Shan Zhao for his continuous support that directed me to this dissertation. Back in 2016, it was challenging for me to prove the unconditional stability of the numerical method we were working on, but thanks to Dr. Zhao for his encouragement for introducing a new mathematical model! I was privileged to work with him in an NSF (DMS-1318898) and Simons Foundation (award 524151) supported Project.

I would also like to thank Dr. David Halpern, Dr. Kabe Moen, Dr. Brendan Ames and Dr. Weihua Su for their generous support to accomplish my dissertation work. I am very grateful to them for the valuable discussions we had related to my research work.

I appreciate Dr. David Cruz-Uribe for mentoring me during the period of my job-search and hiring interviews. His guidance helped me to secure my first job, after which, I was able to invest sufficient time in preparing my dissertation.

I am lucky to have Sheik Ahmed Ullah, Siwen Wang, Zhihan Wei, Hongsong Feng, Arum Lee, Dr. Duc Nguyen, Dr. Chuan Li, Dr. Lunji Song, Leighton Wilson and Benjamin Jones as my research group members. Their questions and feedbacks always refined my learning. I thank all the professors at the UA Mathematics Department for their invaluable help. I also thank Ms. Farley Michele, Ms. Marcia Black, and Ms. Natalie Lau, for providing me with all kinds of administrative support. I feel grateful to Mr. Nathan Jackson, Ms. Camille Steiner, Ms. Jil Chambless and Dr. Yuanyuan Song for their huge support in building my teaching career here at the University of Alabama. It is really hard to express how helpful the entire mathematics department was for the last 5 years of my graduate study. Thanks to all my colleagues at the Department for being helpful and creating such a friendly environment.

I am very grateful to my entire family for nurturing my ambition. I would not be here if my father did not heal my algebra anxiety. Many thanks to my high school tutors, Panchugopal Kalowar and Dr. Surajit Chattopadhyay, for helping me build my math foundation and also for encouraging me to pursue doctoral study.

My special thanks to my husband Soumyadip Acharyya, who motivated me in life, teaching and research. Whenever I needed, I was fortunate to have Soumyadip to share my thoughts.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Calculations of electrostatic potential and solvation energy of macromolecules are essential for understanding the mechanism of biological processes. However, these calculations cannot be done analytically for irregularly shaped objects, and so computational methods must be applied. There are two major approaches for solvation energy analysis, i.e., explicit models and implicit models [1]. Explicit models treat water as individual molecules; on the contrary, implicit models average the effect of solvent phase as continuum media [1–3]. Compared to explicit models, implicit models are more efficient; therefore they can handle much larger systems [3, 4], however, it comes with the price of losing some atomic information and having the ambiguity of how to describe the dielectric properties of the system, the solute, and the water phases.

As a partial differential equation (PDE) model for electrostatics of biomolecules, the Poisson-Boltzmann (PB) equation is a widely used implicit solvent method [3]. Traditionally, a two-dielectric approach is employed in the PB model to describe the dielectric properties: a biomolecule is assigned a low dielectric constant while the surrounding water phase is considered as a high dielectric constant medium. A dielectric interface is assumed at the macromolecule-water boundary, which is usually modeled as a molecular surface. The most commonly used definitions of the macromolecule-water boundary are the Van der Waals (VDW) surface [5], the solvent accessible surface (SAS) [6], and the solvent excluded surface (SES) [7, 8]. However, these "hard sphere" molecular surface models are known to admit geometric singularities, such as cusps and self-intersecting surfaces [20].

To avoid geometric singularities associated with "hard sphere" definitions of the molecular surface, "soft sphere" models have been developed [9–11], where each atom is outlined by a Gaussian density distribution function. While dealing with multiple atoms, the summation of these Gaussian soft clouds forms an electron density map which generates Gaussian molecular surfaces at appropriate isosurfaces or level sets to approximate the rigid VDW surface, SAS, or SES. The density maps based on volumes can also be generated by other smoothly decaying functions [12] or by maximizing the Gaussian functions and then post-processing using a low-pass filtering [13]. The models based on Gaussian surfaces are particularly useful for fast and robust molecular surface mesh generations [12, 14, 15].

In most studies of Gaussian surfaces, the PB equation is still solved in a two-dielectric setting by generating an iso-surface as the dielectric boundary. Across such a sharp interface, the PB solution loses its regularity. In order to avoid accuracy reduction in numerical discretization near the interface, sophisticated interface algorithms have to be adopted for handling the dielectric jump in solving the PB equation. With rigorous interface treatments, the matched interface and boundary (MIB) method [17, 19] and the immersed interface method (IIM) [16] can improve the accuracy significantly, but they develop complexity in the algorithm to a certain extent which reduces the computational efficiency.

Instead of using a sharp molecular surface definition, smooth or smeared molecular surfaces have also been introduced in the literature [18, 20–22, 53], in which a smooth transition is assumed in between solute and solvent domains. For instance, by using the Euler-Lagrange variation of the free energy minimization, Bates *et al.* [20, 53] introduced a variational PDE model for molecular surface generation. Neglecting other solute-solvent interactions, this model is simplified to be the surface area minimization, and gives rise to the minimal molecular surface (MMS). Cheng *et al.* [18] have employed the level set approach to minimize a free energy functional for coupling the

polar-nonpolar interaction at the solvent-solute interface, and the corresponding PDE model involves contributions from electrostatic effects, pressure, Gauss and mean curvatures, and others. A phase-field variational approach has been developed in [21] to represent the solute-solvent interface via a double-well potential in the free energy functional. The convergence of the phase field free energy functionals and forces to their sharp interface limits have been rigorously proved [22]. By using these smooth molecular surfaces, simple numerical methods can be employed for solving the PB equation, and complicated interface treatments are unnecessary.

Besides the above mentioned PB models with essentially two homogeneous media, heterogeneous dielectric models have also been introduced in the literature [1, 25–29, 31], in which the dielectric function ϵ is not uniform and varies within the structure of the molecule. Physically, such an inhomogeneity, expressed as the different polarizability and flexibility, is well-documented for the amino acids [23, 24]. Mathematically, heterogeneous dielectric distributions provides an alternative means to mimic the effect of conformation changes of the macromolecule on the solvation energy, because dielectric distributions reflect the structure-energy relations via screening of the electrostatic interactions within the solute and between the solute and solvent [55, 56].

This dissertation will pay particular attention to an inhomogeneous model, i.e., the Gaussian dielectric PB model [1, 31], which was developed with an aim to provide a "correct" description of the dielectric property of the macromolecule, i.e., beginning with macromolecule interior and moving toward the macromolecular surface and further into the water phase, the ability of the corresponding medium to respond to local electrostatic field constantly increases [30]. This dielectric model has been found to outperform the traditional two dielectric model in many biological applications, including a better agreement with experimentally measured solvation energy of small molecules [1, 31] and a better prediction of the pKa's of ionizable groups against thousands of experimentally measured pKa's in various proteins [37, 38]. The Gaussian

dielectric model has also demonstrated the feasibility of approximating the ensemble average polar solvation energy by calculating a single macromolecular structure and without resorting to expensive molecular dynamics or Monte Carlo simulations [57].

Even though it has achieved a great success in biology, there is still a large room for the Gaussian dielectric model [1, 31] to be further improved, particularly from the mathematical point of view. In this dissertation, we will focus on three major aspects for our modeling efforts. First, there is a theoretical gap between the Gaussian dielectric model and the classical two-dielectric model, in the sense that the Gaussian dielectric function will not converge to the piecewise dielectric constants under any limit. It is desired to establish such a convergence, at least for individual atoms, so that the theory and the algorithm developed for one model can be potentially extended to the other. Second, the maximal dielectric value ϵ_{max} of the macromolecule is not determined by the solute itself in the existing Gaussian model, instead it is inflated by the external water dielectric value (usually taken as $\epsilon = 80$). This causes some concerns physically, because in free energy calculations, the same protein and the same inhomogeneous ϵ should be calculated for the outside phase being the vacuum. Third, in the existing Gaussian dielectric model [1, 31], the inhomogeneous dielectric profile of the macromolecule is generated based on the water state first. Then a surface cut with an empirical iso-value is conducted to preserve the same inhomogeneous profile for the vacuum state. Consequently, the ϵ function becomes discontinuous, because outside the surface cut, $\epsilon = 1$ is simply used for the vacuum. A modified surface cut technique has been reported recently [57], which results in a C^0 but not C^1 continuous dielectric function. Mathematically, the Gaussian model still cannot be claimed as a smooth dielectric model.

In this work, several simple but mathematically critical improvements will be introduced to address the aforementioned three concerns. First, we propose a super-Gaussian density function as a "soft sphere" model for each atom, which includes

the Gaussian function as a special case of the order $m = 1$. On the other hand, it approaches the two-dielectric model in the limit of order m going to infinity. In practice, $m = 3$ or 4 achieves a good trade-off in our modeling and simulations. Second, a maximal dielectric value ϵ_{max} of the macromolecule is introduced in our new dielectric model, which is independent of the outside phase (water or vacuum). Physically, ϵ_{max} can be justified by considering the compactness of atoms, so that it is related to the cavities and channels which are frequently encountered in biomolecules. The dielectric variation due to cavities can hardly be modeled by two-dielectric models, while for inhomogeneous models, the determination of dielectric values for cavities is still in its infancy [41], because such cavity regions could be empty or filled with crystallographic water. The proposed super-Gaussian model will provide a manipulable tool for biologists to study this issue. Finally, the minimal molecular surface (MMS) [20, 44] will be employed to represent solute and solvent regions in the super-Gaussian model. The main purpose of such representation is not defining a molecular surface. Instead, the MMS allows us to represent both water and vacuum states in the one equation, by simply changing the exterior dielectric value to be 80 or 1. The interior dielectric profile for proteins remains unchanged in this process. With these improvements, the new Gaussian model becomes a smooth dielectric model with the dielectric functions being C^∞ continuous in both water and vacuum states.

Besides the modeling studies, this work will also investigate efficient numerical algorithms for solving the smooth dielectric PB equation. Another important parameter in the nonlinear PB equation is the Debye-Huckel parameter or Debye length κ, which is vanishing in the solute region. Instead of treating κ as a piecewise constant, it will be defined in the same manner as the dielectric function ϵ, by using the MMS characteristic function. The consequent numerical issue is that κ could be nonzero in certain places which are traditionally regarded inside the solute region, but are now in the transition layer between solute and solvent media [43]. Therefore, the

hyperbolic nonlinear term of the PB equation could take huge values at such places, so that the conventional numerical methods become unstable [43]. To suppress the nonlinear instability, a pseudo-time continuation approach with analytical integration of the nonlinear term has been proposed in the literature [43, 45, 54]. Based on finite difference spacial discretization, efficient alternating direction implicit (ADI) schemes have been developed for pseudo-time integration [43, 45, 54]. However, such ADI schemes could not achieve unconditional stability in treating two-dielectric PB equations. For the present super-Gaussian dielectric model and well filtered MMS representation, the pseudo-time ADI schemes will be unconditionally stable for solving the nonlinear PB equation.

The rest of the dissertation is organized as follows. Chapter 2 introduces the super-Gaussian dielectric PB model with a few parameters. An effective dielectric constant (EDC) analysis is proposed for determining the best fitting parameters, and the role of hypersurface function generated from MMS is discussed. In chapter 3, the super-Gaussian PB equation is discretized by using a pseudo-time alternating direction implicit (ADI) algorithm. Model validation and convergence, accuracy, and stability of the ADI algorithm are experimented by calculating solvation free energy for a single atom system in chapter 4. The proposed model and algorithm are further verified in chapter 5, by considering various proteins. Particular attention will be paid on the studies of a real protein with both cavity-fluids and empty cavities. This dissertation ends with some conclusions.

CHAPTER 2

MATHEMATICAL MODELING

In this chapter, we will first briefly describe the existing models, including the two-dielectric Poisson-Boltzmann (PB) model and Gaussian dielectric PB model. Then, a super-Gaussian dielectric PB model will be introduced. An effective dielectric approach will be employed to systematically study the influence of the adjustable parameters of the new model in various settings.

2.1 Two-dielectric Poisson-Boltzmann model

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 classical two-dielectric PB model, the domain Ω is divided by a molecule surface Γ into two parts, namely the inner solute domain Ω_m and the outer solvent domain Ω_s such that $\Omega = \Omega_m \cup \Omega_s$ and $\Omega_m \cap \Omega_s = \Gamma$. Denote the boundary of Ω as $\partial\Omega$. For $\vec{r} \in \mathbb{R}^3$, the electrostatic potential u of this system is governed by the nonlinear Poisson-Boltzmann equation and its most commonly used dimensionless form [50, 54] is given as

$$-\nabla \cdot (\epsilon(\vec{r})\nabla u(\vec{r})) + \kappa^2 \sinh(u(\vec{r})) = \rho_m(\vec{r}), \tag{2.1}$$

where the singular source term is

$$\rho_m(\vec{r}) = 4\pi \frac{{e_c}^2}{k_B T} \sum_{j=1}^{N_m} q_j \delta(\vec{r} - \vec{r_j}). \tag{2.2}$$

On the outer boundary $\partial\Omega$, a Dirichlet boundary condition can be assumed

$$u(\vec{r}) = \frac{{e_c}^2}{k_B T} \sum_{j=1}^{N_m} \frac{q_j}{\epsilon_s |\vec{r} - \vec{r_j}|} e^{(-|\vec{r}-\vec{r_j}|\sqrt{\frac{\bar{\kappa}^2}{\epsilon_s}})}. \tag{2.3}$$

In the two-dielectric PB model, the dielectric function $\epsilon(\vec{r})$ is assumed to be a piecewise constant

$$\epsilon(\vec{r}) = \begin{cases} \epsilon_m, & \vec{r} \in \Omega_m \\ \epsilon_s, & \vec{r} \in \Omega_s. \end{cases} \tag{2.4}$$

In the present study, we will take $\epsilon_m = 1$ for the protein and $\epsilon_s = 80$ for the water. Similarly, the modified Debye-Hückel parameter κ is a piecewise constant. It vanishes in Ω_m, i.e., $\kappa = 0$, while in Ω_s $\kappa = \bar{\kappa}$, where $\bar{\kappa}^2 = 8.486902807\text{Å}^{-2} I_s$ and I_s is the ionic strength of the solvent. Here, k_B is the Boltzmann constant with $k_B T = 0.5921830$ kcal/mol at $T = 298K$, and e_c is the fundamental charge and q_j is the partial charge for the j^{th} atom in the solute, centered at $\vec{r_j}$. Moreover, e_c and q_j have the same units, and $e_c^2 = 332.06364$ kcal/mol. The total number of atoms present in the solute macromolecule is denoted by N_m.

The energy released when the solute macromolecule is dissolved in solvent is know as the free energy of solvation. The polar component of solvation free energy can be calculated in the PB model by computing the difference between total electrostatic free energy of the macromolecule in the solvent and in a vacuum. In particular, for the two-dielectric PB model, the solvation energy is defined as

$$\Delta G = G_s - G_0 = \frac{1}{2} \int_\Omega \rho_m (u(\vec{r}) - u_0(\vec{r})) d\vec{r} \tag{2.5}$$

where $u(\vec{r})$ is the solution of the PB equation (2.1), while $u_0(\vec{r})$ is the electrostatic potential of the macromolecule in the vacuum. The vacuum state is obtained by taking $\epsilon(\vec{r}) = 1$ throughout and setting the ionic strength $I_s = 0$. Consequently, $\kappa = 0$ in the PB equation (2.1) and $\bar{\kappa} = 0$ in the boundary condition (2.3). Thus, $u_0(\vec{r})$ is in fact the

solution of a Poisson equation

$$-\Delta u_0 = \rho_m, \tag{2.6}$$

with the same singular source (2.2).

2.2 Gaussian dielectric PB model

In order to overcome some inherent difficulties associated with the two-dielectric PB model, a Gaussian dielectric PB model has been proposed in [1, 31] to provide a "correct" description of the dielectric property of the macromolecule. Physically, at the atomistic level of detail, any system in molecular biophysics is made up of macromolecules immersed in water, and can be considered as a multitude of atoms: atoms of water molecules and amino acids (nucleic acids). It thus makes sense to study a smooth dielectric PB model, in which one avoids to define a solute-solvent boundary or molecular surface. Instead, an appropriate definition of the dielectric function $\epsilon(\vec{r})$ is assumed in the entire domain Ω. Moreover, it is known that beginning with the macromolecule interior and moving toward the macromolecular surface and further into the water phase, the ability of the corresponding medium to respond to the local electrostatic field constantly increases [30]. Hence, one should expect that $\epsilon(\vec{r})$ in the water state increases smoothly from the solute region to the solvent region. Finally, allowing $\epsilon(\vec{r})$ to be inhomogeneous gives us flexibility in modeling different polarizability of the amino acids [23, 24], and mimicking the effect of conformation changes of the macromolecule on the solvation energy [55–57].

A "soft sphere" approach by introducing a density function for each atom seems to be a natural model to fulfill all of the above considerations. This motivated the development of the Gaussian dielectric PB model [1, 31] in the water state. Suppose the macromolecule contains a total of N_m atoms and the density [11, 51, 52] at the position

$\vec{r}$ for the i^{th} atom is given by

$$g_i(\vec{r}) = \exp\left[\frac{-|\vec{r} - \vec{r_i}|^2}{\sigma^2 R_i^2}\right] \tag{2.7}$$

where $\vec{r_i}$ is the center of the i^{th} atom, R_i is the Van der Waals radius of the i^{th} atom and σ is the relative variance. Once the density for each atom is generated, the total density function for the atoms and overlapped area covered by multiple atoms is given by

$$g_0(\vec{r}) = 1 - \prod_{i=1}^{N_m}[1 - g_i(\vec{r})] \tag{2.8}$$

where the cross term such as $g_i g_j$ accounts for the density of the overlap region due to the i^{th} and j^{th} atoms. Also the total density function ensures that the overlap region has more density than the density generated by a single atom. The range of the function g_0 is $[0, 1]$. Finally, the dielectric distribution is derived as a weighted convex combination

$$\epsilon_G(\vec{r}) = g_0(\vec{r})\epsilon_m + (1 - g_0(\vec{r}))\epsilon_s \tag{2.9}$$

where ϵ_m and ϵ_s are the dielectric constants in the molecule and water respectively. Similar to the two-dielectric model, we will take $\epsilon_m = 1$ and $\epsilon_s = 80$ in the present study.

By simply replacing $\epsilon(\vec{r})$ in the PB equation (2.1) by $\epsilon_G(\vec{r})$, the Gaussian PB model has achieved a great success in various biophysical applications [1, 31]. However, there are several theoretical concerns which call for further improvements. First, the Gaussian dielectric function $\epsilon_G(\vec{r})$ will never converge to the piecewise constant defined in (2.4) under any circumstance. Moreover, the maximal value of $\epsilon_G(\vec{r})$ of a macromolecule or inside the solute region is actually determined by $\epsilon_s = 80$. If a different parameter value other than 80 is used for the water phase, the same protein will have different dielectric values! Finally, in electrostatic free energy calculations, a

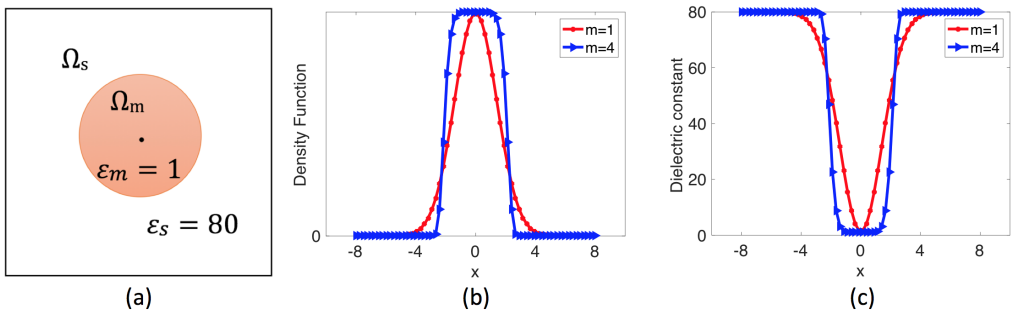


Figure 2.1: (a) A single atom is immersed into water, (b) density function of one atom for $m = 1$ (Gaussian), $m = 4$ (super-Gaussian of order 4) and (c) dielectric distribution of the single atom system calculated by Eq. (2.9) for Gaussian and super-Gaussian (order 4) densities.

surface cut of $\epsilon_G(\vec{r})$ at an iso-value 20 is conducted to introduce a sharp boundary Γ [1, 31]. Inside Γ, the dielectric function of the vacuum state is the same as that in the water state, i.e., $\tilde{\epsilon}_G(\vec{r}) = \epsilon_G(\vec{r})$, while outside Γ, $\tilde{\epsilon}_G(\vec{r}) = 1$. One then solves the Poisson equation

$$-\nabla \cdot (\tilde{\epsilon}_G \nabla u_0(\vec{r})) = \rho_m, \tag{2.10}$$

for the electrostatic potential u_0 in vacuum, and then computes the solvation free energy by (2.5). Note that $\tilde{\epsilon}_G$ is discontinuous in (2.10) so that various difficulties associated with the two-dielectric PB equation may not be avoided. Recently, a further modification to $\tilde{\epsilon}_G$ has been introduced in [57], which results in a C^0 but not C^1 continuous function. Rigorously speaking, the Gaussian dielectric PB model still cannot be claimed as a smooth dielectric model, because of $\tilde{\epsilon}_G$ in the vacuum state.

2.3 Super-Gaussian dielectric PB model

To address the first concern of the Gaussian model, we propose to define the density of the i^{th} atom as a super-Gaussian function

$$g_i^s(\vec{r}) = \exp\Big[-\Big(\frac{|\vec{r} - \vec{r_i}|^2}{\sigma^2 R_i^2}\Big)^m\Big]. \tag{2.11}$$

Note that with the order $m = 1$, $g_i^s(\vec{r})$ becomes the original Gaussian density function $g_i(\vec{r})$. To illustrate the idea, we first consider the dielectric distribution defined by (2.8) and (2.9) and simply replace $g_i(\vec{r})$ by $g_i^s(\vec{r})$. A virtual comparison of the corresponding Gaussian and super-Gaussian distributions for a single atom system is depicted in Fig. 2.1. It can be seen that the super-Gaussian function or higher order Gaussian has a flat-top density and a rapid while smooth transition at the solute-solvent border area. As m goes to infinity, $g_i^s(\vec{r})$ approaches a step function that equals one inside the Van der Waals (VDW) sphere with the center $\vec{r_i}$ and radius R_i and equals zero outside the sphere. Consequently, the dielectric distribution shown in Fig. 2.1 (c) will converge to the piecewise constant of the two-dielectric model, i.e., Eq. (2.4). Therefore, the super-Gaussian density includes both the Gaussian density and piecewise constant (of the two dielectric model) as special cases. In practice, we will consider the order m in the range of $\{1, 2, \ldots, 8\}$, which maintains enough smoothness when the function is sampled on a discrete grid. In Fig. 2.2, we depict the super-Gaussian dielectric distributions for a one-atom system by using different order m and relative variance σ. The optimal selection of these parameter values will be discussed later.

To address the second concern of the Gaussian model, we introduce a parameter ϵ_{gap} to represent the maximum dielectric value of the macromolecule. In particular, we similarly define the total density function as

$$g_0^s(\vec{r}) = 1 - \prod_{i=1}^{N_m} [1 - g_i^s(\vec{r})]. \tag{2.12}$$

A new dielectric distribution is proposed within a protein region

$$\epsilon_{in}(\vec{r}) = \epsilon_m g_0^s(\vec{r}) + \epsilon_{gap}[1 - g_0^s(\vec{r})], \tag{2.13}$$

where the constants ϵ_m and ϵ_{gap} are defined as the reference dielectric values at the atom centers and in a gap region, respectively, with $\epsilon_{gap} > \epsilon_m$. By substituting (2.12)

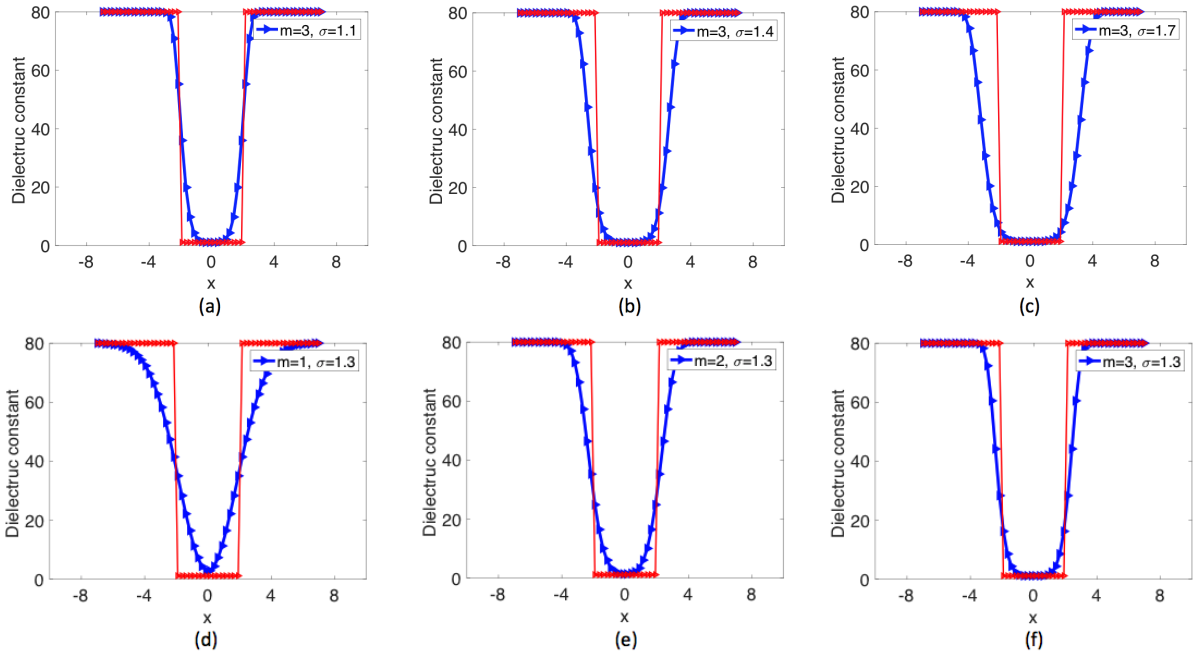


Figure 2.2: The dielectric distributions generated by the super-Gaussian functions for a one-atom system. In all figures, the red line represents the piecewise constant of the two dielectric model.

into (2.13), we have an equivalent form of ϵ_{in}

$$\epsilon_{in}(\vec{r}) = \epsilon_m + (\epsilon_{gap} - \epsilon_m) \prod_{i=1}^{N_m} [1 - g_i^s(\vec{r})]. \tag{2.14}$$

It is then clear that ϵ_m and ϵ_{gap} are, respectively, the minimal and maximal dielectric values of the protein, independent of the outside medium.

The physical idea underlying (2.13) or (2.14) is that the permittivity at a loosely packed region of a protein shall be higher than that in a densely packed region, because the former region has a higher polarization or allows a larger conformational change. In a densely packed region, the charged atoms and amino acid chains are harder to shift from their average equilibrium positions when an electric field is placed, so that the polarization or density of induced electric dipole moments is weaker. Moreover, cavities have to be taken into account in an inhomogeneous dielectric model. Crystallographic

waters may be trapped inside some large cavities. The polarization of water molecules inside cavities is smaller than the bulky water molecules in a solvent due to their restricted degree of freedom, but it is still much higher than that of protein. This suggests that $\epsilon_m < \epsilon_{gap} \leq \epsilon_s$. An appropriate value for ϵ_{gap} depends on the real protein system and will be determined through analytical and numerical means in this work. Also, we will take $\epsilon_m = 1$ and $\epsilon_s = 80$ as in the other models.

To address the third concern of the Gaussian model, we propose to provide a certain description of the solute and solvent domain on top of the dielectric distribution. Without such a description, the original Gaussian PB model has to conduct a surface cut operation for the vacuum state, which compromises the regularity of the system. We note that traditional molecular surfaces, including the VDW surface [5], the solvent accessible surface (SAS) [6], and the solvent excluded surface (SES) [7, 8], could not fulfill our goal here, because the smoothness still cannot be maintained across a sharp solute-solvent interface. Instead, we propose to employ the minimal molecular surface (MMS) [20, 53], which is defined as the unique surface that is of the smallest area and encloses all VdW balls. Physically, the MMS model is attained through the surface free energy minimization. Mathematically, the Euler-Lagrange variation of the free energy leads to a mean curvature flow partial differential equation (PDE), which can be solved by a fast algorithm developed in [44]. The numerical solution provides not only the MMS, but also a level set function or hypersurface function $S(\vec{r})$ defining the solute and solvent regions in a smooth manner, see Fig. 2.3 (a) for an illustration.

The hypersurface function $S(\vec{r})$ of the MMS model [20, 44, 53] was originally used for representing the protein region, with $S = 1$ inside all VDW balls and $S = 0$ outside the SAS (based on a probe radius 1.4 Å). A smooth transition from one to zero is obtained through numerical PDE solution. In the proposed super-Gaussian PB model, we will make use of $(1 - S)$ to present the exterior region so that both water and

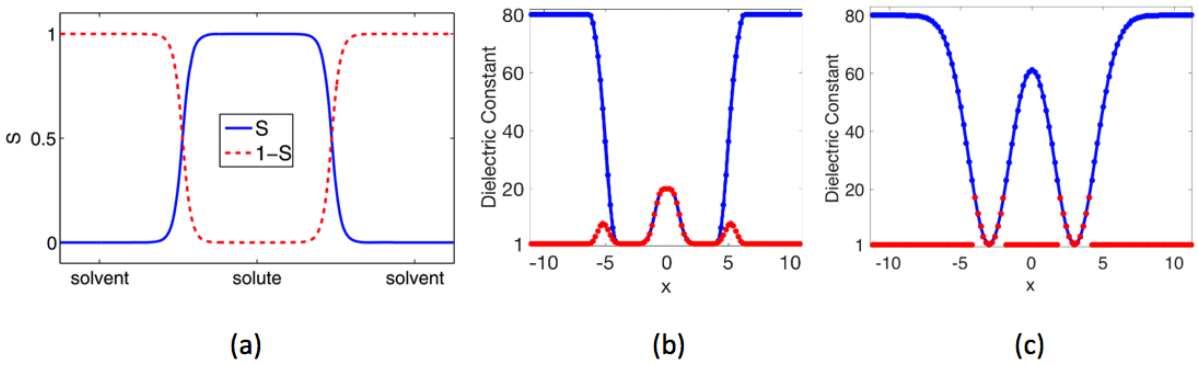


Figure 2.3: (a) The hypersurface functions S and $(1-S)$ of the solute-solvent region along a straight line. (b) The blue and red curves depict the dielectric function $\epsilon_{sG}(\vec{r})$ of the super-Gaussian model in the water and vacuum phases respectively. (c) The blue and red curves depict the dielectric functions $\epsilon_G(\vec{r})$ and $\tilde{\epsilon}_G(\vec{r})$ of the Gaussian model in the water and vacuum phases respectively.

vacuum phases could be modeled in one equation

$$\epsilon_{sG}(\vec{r}) = S(\vec{r})\epsilon_{in}(\vec{r}) + [1 - S(\vec{r})]\epsilon_{out}, \tag{2.15}$$

where the constant ϵ_{out} determines the dielectric value far away from the protein. Note that $S = 1$ inside the VDW region so that inhomogeneity of the super-Gaussian dielectric distribution is retained. By setting ϵ_{out} to be 1 or 80, one simply switches from vacuum phase to water phase.

In the proposed super-Gaussian dielectric model, the PB equation is modeled as

$$-\nabla \cdot (\epsilon_{sG}\nabla u) + (1 - S)\bar{\kappa}^2 \sinh(u) = S\rho_m, \tag{2.16}$$

where we have similarly inserted the hypersurface function $S(\vec{r})$ for both the source and nonlinear terms. Note that $\bar{\kappa}$ is a constant, not a piecewise constant in our notation. The switch off of the nonlinear term relies on $(1 - S)$, which has some impact numerically [43]. Similarly, the electrostatic potential u_0 in the vacuum is calculated by

neglecting the nonlinear term

$$-\nabla \cdot (\epsilon_{sG} \nabla u_0) = S\rho_m, \tag{2.17}$$

because $\bar{\kappa} = 0$. Of course, in this Poisson equation, we shall take $\epsilon_{out} = 1$ for defining $\epsilon_{sG}(\vec{r})$ in (2.15). One can then compute the solvation free energy by (2.5).

In the super-Gaussian model, the dielectric function $\epsilon_{sG}(\vec{r})$ remains C^∞ continuous in both water and vacuum states. This is illustrated by considering a two-atoms system in Fig. 2.3 (b), in which $m = 3$, $\sigma = 1.3$, and $\epsilon_{gap} = 20$. It can be observed that inside the solute region, $\epsilon_{sG}(\vec{r})$ is identical for both water and vacuum phases. Near the solute-solvent boundary, the dielectric value produces a smooth bump, because $\epsilon_{gap} = 20$ allows a large ϵ away from the center of the atom. Further away from atoms, the hypersurface function $S(\vec{r})$ plays a dominant role so that ϵ decays to $\epsilon_{out} = 1$ smoothly. For a comparison, the dielectric functions $\epsilon_G(\vec{r})$ and $\tilde{\epsilon}_G(\vec{r})$ of the Gaussian model, for water and vacuum phase respectively, are depicted in Fig. 2.3 (c). In the water phase, the maximal value of $\epsilon_G(\vec{r})$ is determined by $\epsilon_s = 80$, so that it is higher than that of $\epsilon_{sG}(\vec{r})$. In the vacuum phase, by conducting a surface cut at 20, $\tilde{\epsilon}_G(\vec{r}) = \epsilon_G(\vec{r})$ inside two atoms. Nevertheless, $\tilde{\epsilon}_G(\vec{r})$ is discontinuous at atom boundaries.

2.4 Effective dielectric constant analysis

In the proposed super-Gaussian dielectric model, there are three adjustable parameters, i.e., the order m, the relative variance σ which determines the window width of Gaussian and super-Gaussian distributions, and ϵ_{gap} which controls the maximal dielectric value of the solute. In this section, we will explore the impact of these parameters on the final heterogeneous dielectric function $\epsilon(\vec{r})$ and find certain means for selecting suitable values of these parameters in real applications. We are also interested in a comparison among three dielectric models, i.e., the classical

two-dielectric function (2.4), the Gaussian dielectric distribution (2.9), and the super-Gaussian one (2.15), which will be referred to as Model I, II, and III, respectively, in this section.

In principal, the solvation free energy calculation is an ideal means for validating dielectric PB models and calibrating parameters. For example, the solvation energies produced by the Gaussian dielectric model have been compared with experimental results for some organic small molecules [1]. Nevertheless, measurement of solvation energies for macromolecules is still an experimental challenge. Thus, to access the Gaussian dielectric model for proteins, explicit solvent molecular dynamics (MD) simulations have been conducted to generate referencing solvation energies to compare with the PB results [1]. We will not repeat such studies in this dissertation based on the following considerations. First, the electrostatic potential calculated from solving the PB equation contributes only to the polar part of the total solvation free energy [42], while the nonpolar part of such energy depends on the volume of the macromolecule, the area of the molecular surface, and the attractive VDW dispersion effects near the solute-solvent interface. Thus, the impact of the dielectric function on the total solvation free energy is really hard to be singled out. Consequently, the use of experimentally available solvation free energies to optimize the parameters of the new dielectric model is not an easy job. Second, the three dielectric models considered in this work involve different underlying molecular surfaces, and hence have different areas and volumes, so that a comparison in terms of solvation free energy seems to be inappropriate. Third, the electrostatic solvation free energy which is just the polar part of total solvation energy could be a better choice. Indeed, in Ref. [1], the Gaussian dielectric model is primarily benchmarked by studying electrostatic solvation energies for many proteins, which are generated by MD simulations. However, we note that MD simulations are very time consuming.

In this paper, we propose an effective dielectric constant (EDC) analysis as a

simple means to access different dielectric models. Consider some simple systems with a few atoms immersed in the water. We first generate the dielectric function $\epsilon(\vec{r})$ by a model over a certain domain Ω. We then define the effective dielectric constant as

$$\hat{\epsilon} = \frac{\int_\Omega \epsilon(\vec{r}) d\vec{r}}{\int_\Omega d\vec{r}}, \tag{2.18}$$

which measures, in an average sense, the resistance encountered when forming an electric field in this solute-solvent system. The EDC can be calculated either analytically or numerically, and enables us to investigate the role of each parameter in the super-Gaussian distribution.

To select suitable parameter values, we will benchmark the EDC of the super-Gaussian model against that of the two-dielectric model, and report the relative difference between them in our studies. Note that this does not mean that we treat the two-dielectric PB model as the "correct" model to compare with. In fact, the original purpose of Gaussian type models is to improve the two-dielectric PB model. However, in practice, the two-dielectric function is still the most widely used setting for the PB equation. It thus makes sense that a new dielectric model should not deviate from the two-dielectric model too much. With the EDC analysis, we can ensure the super-Gaussian model agrees with the two-dielectric model in a mean field sense. This could potentially persuade more biologists to use the new model, because more freedom is available now for modeling purpose. However, we also note that with similar EDC values, the electrostatic solvation energies produced by the two-dielectric and the super-Gaussian models could still be significantly different.

In the super-Gaussian model, the minimal molecular surface (MMS) is calculated by using the fast algorithm developed in [44]. Through the EDC analysis, we will choose the relative variance for the Gaussian dielectric model around the value 1, namely, $\sigma \in \{0.8, 0.9, ..., 1.3\}$. When we upgrade the density function from Gaussian to

super-Gaussian, the corresponding relative variance will be changed and depends on the choice of $m \in \{1, 2, ...8\}$. Finally, as we consider the super-Gaussian dielectric model (ϵ_{sG}) for the inhomogeneous macromolecule interior, we need to decide the preference of $\epsilon_{gap} \in \{2, 4, ..., 8, 10, 20, 40, 80\}$ for different solute-solvent system. This selection depends on the cavity inside the solute. In the two-dielectric model, the solvent excluded surface (SES) is chosen as the molecular surface defining the solute-solvent boundary, and will be calculated by using the MSMS package [58]. We refer to [20] for a detailed comparison between MMS and MSMS.

2.4.1 Effective dielectric constant analysis with one atom

We first conduct the effective dielectric constant (EDC) analysis for a single atom solute-solvent system in the water phase. Consider a sphere with radius $R_0 = 2$ and center at the original. A large enough domain $\Omega = [-a, a]^3$ is chosen with $a = 8$. See Fig. 2.1(a) for an illustration. By taking $\epsilon_s = 80$ and $\epsilon_m = 1$, three dielectric models are studied in this dissertation. By comparing the EDCs of three models, we can find the optimal values of parameter σ and m for the one atom system.

Model I: In the two-dielectric model, $\epsilon_2(\vec{r})$ is defined as a piecewise constant as in Eq. (2.4). The EDC can be calculated analytically in this case

$$\hat{\epsilon}_2 = \frac{\int_\Omega \epsilon_2 d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\epsilon_s(2a)^3 - (\epsilon_s - \epsilon_m)(\frac{4}{3}\pi R_0^3)}{(2a)^3} = 79.3537. \tag{2.19}$$

Model II: In the Gaussian dielectric model, $\epsilon_G(\vec{r})$ is calculated by (2.9). The EDC $\hat{\epsilon_G}$ for ϵ_G is calculated through numerical integration:

$$\hat{\epsilon}_G = \frac{\int_\Omega \epsilon_G d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\int_\Omega [\epsilon_m g_0 + \epsilon_s(1 - g_0)] d\vec{r}}{(2a)^3}, \tag{2.20}$$

where g_0 is given by the equation (2.8), and $\hat{\epsilon}_G$ only depends on the relative variance σ. By taking $\sigma = \{0.8, 0.9, 1.0, 1.1, 1.2, 1.3\}$, the EDC results are reported in Fig. 2.4. It

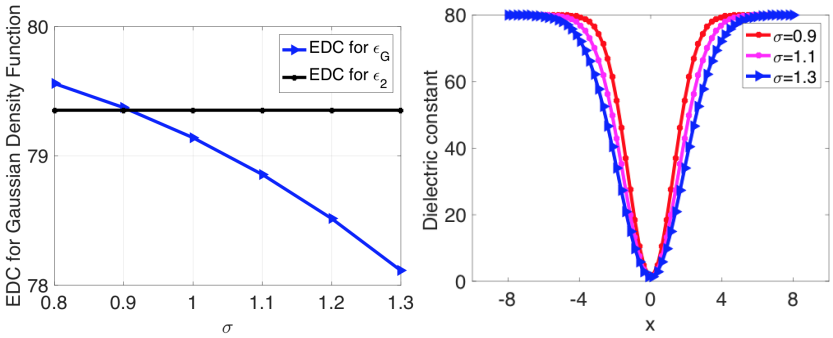


Figure 2.4: Comparison of effective dielectric constant of Gaussian dielectric model with different σ.

can be seen from Fig. 2.4(a) that out of the six discrete numbers being considered, $\sigma = 0.9$ obviously provides the best fit to $\hat{\epsilon}_2$. This is in excellent agreement with the existing study, in which the optimal value obtained through molecular dynamics simulations is $\sigma = 0.93$ [57]. The slice plot of $\epsilon_G(\vec{r})$ is given in Fig. 2.4(b) for several σ values. Physically, the relative variance controls the upper half window-width of the function ϵ_G. As σ increases, the window becomes wider at the upper half section of ϵ_G which belongs to the solvent region, while has less impact to the bottom half section. Due to this broadening effect of σ, the EDC decreases as σ increases, as can be seen in Fig. 2.4(a).

Model III: For a comparison, we will also consider the super-Gaussian function for the one-atom system. Nevertheless, we note that with only one atom, there is no cavity or gap region in the solute. Consequently, ϵ_{gap} is physically undefined in this system. For this reason, we will not study the actual super-Gaussian dielectric model. Instead, in the Gaussian dielectric model (2.9), we simply replace $g_0(\vec{r})$ by the super-Gaussian density function $g_0^s(\vec{r})$ defined by Eq. (2.12). Let us denote the corresponding dielectric model as ϵ_G^s. This enables us to investigate the roles of the order m and relative variance σ in the one-atom system. Numerical integration is

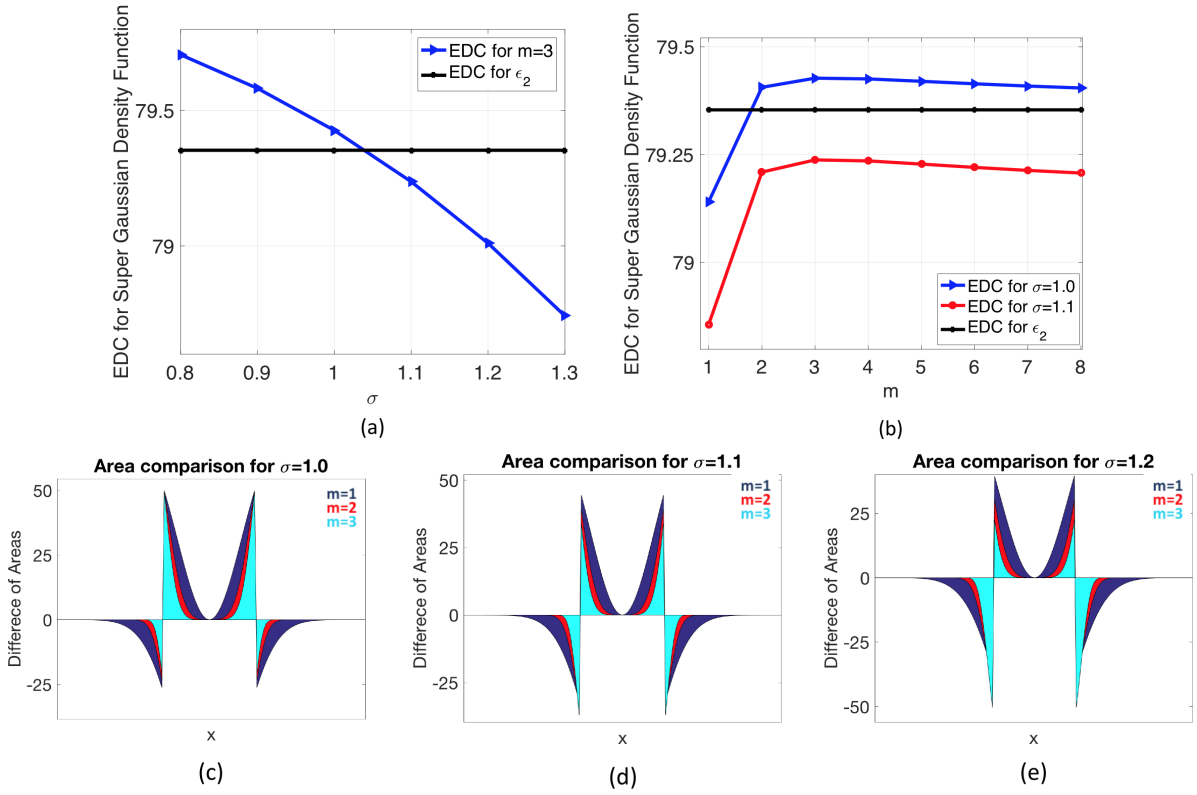


Figure 2.5: (a) The EDC of the super-Gaussian density function with power $m = 3$, (b) the EDC for $\sigma = 1.0$ and $\sigma = 1.1$, (c)-(e) Comparison of area differences between the compensated dielectric curves $\epsilon_G^s(\vec{r}) - \epsilon_2(\vec{r})$ for $\sigma = 1.0, 1.1, 1.2$ and $m = 1, 2, 3$.

carried out to calculate the EDC similarly

$$\hat{\epsilon}_G^s = \frac{\int_\Omega \epsilon_G^s d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\int_\Omega [\epsilon_m g_0^s + \epsilon_s(1 - g_0^s)] d\vec{r}}{(2a)^3}. \tag{2.21}$$

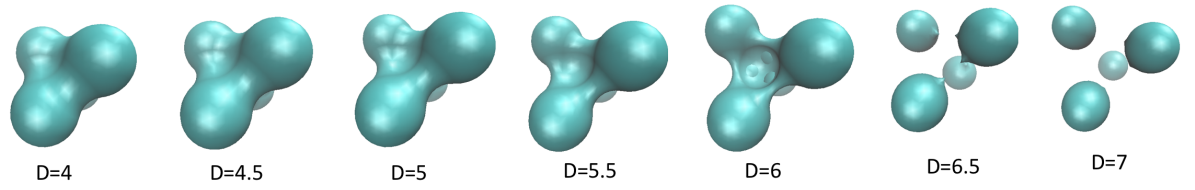
We first vary σ without changing m. Similarly to the previous case, it is found that $\hat{\epsilon}_G^s$ decreases as σ increases for a fixed m, see Fig. 2.5(a) for the case $m = 3$. Moreover, for a larger m, the optimal σ value becomes larger. For example, for $m = 3$, the optimal σ value is larger than 1 now. Next, by fixing σ, the effect of changing m is shown in Fig. 2.5(b). It can be seen that the EDC increases quickly when m changes from 1 to 2, achieves a maximum around $m = 3$ or $m = 4$, and then declines slowly. Asymptotically, for $\sigma = 1$, the EDC of the super-Gaussian density should approach

that of the two-dielectric model, as $m \to \infty$, i.e., $\lim_{m\to\infty} \epsilon_G^s = \epsilon_2$. This confirms that the super-Gaussian dielectric function approaches the two-dielectric function when m goes to infinity. For $\sigma = 1.1$, the EDC curve is simply a shift of that of $\sigma = 1.0$ downwardly with the optimal orders being $m = 3$ or $m = 4$. For the other σ values, we have seen the same pattern that the EDC values for $m > 4$ are quite close to those of $m = 3$ or $m = 4$. Thus, in our numerical computations, we usually choose $m = 3$ or $m = 4$ with an optimized σ.

Since the EDC values change significantly for $1 \leq m \leq 3$, it is interesting to further compare the difference among them from a different perspective. In Fig. 2.5 (c) - (e), we plot the compensated dielectric curves, i.e., $\epsilon_G^s(\vec{r}) - \epsilon_2(\vec{r})$, over the cross section plane $y = 0$. As can been from these figures, the compensated curves are positive inside the atom, because $\epsilon_G^s \geq 1$ and $\epsilon_2 = 1$. Right outside the atom boundary, ϵ_2 becomes 80, so that the compensated curves immediately drops to negative numbers. As the radius keeps increasing, ϵ_G^s approaches 80 so that the compensated curves vanish at both ends. We note that due to the symmetry of this system, the net area obtained by integrating each compensated curve in such a two-dimensional (2D) setup essentially captures the volume difference between the EDC values for the super-Gaussian and two-dielectric models. For each σ, when m becomes larger, the areas for both positive and negative regions shrink significantly. This is essentially why the EDC lines change dramatically for $1 \leq m \leq 3$ in Fig. 2.5 (b). Comparing with the different σ values, it seems that $\sigma = 1.1$ produces more balanced net areas.

2.4.2 Effective dielectric constant analysis with four atoms

We next study a four-atom system immersed in water so that a cavity region can be formed. This enables us to explore the role of ϵ_{gap} in the super-Gaussian model for the water phase. To this end, consider a regular tetrahedron with all sides having the same length D. Four atoms are defined by using the vertices of the tetrahedron as centers and with a radius 2. By fixing the center of this tetrahedron as the origin of the



(a) MSMS generated surfaces for D=4, 4.5, ...7 and probe radius =1.5

(b) MMS generated surfaces for D=4, 4.5, ...7, dx=0.25, iso-value=0.9 and probe radius =1.5

Figure 2.6: (a) The solvent excluded surface (SES) generated by the MSMS package for D=4, 4.5, ..., 7. (b) The minimal molecular surface (MMS) for D=4, 4.5, ..., 7.

coordinate, we will vary D from 4 to 7. The illustrations of four atoms with $D = 4$ and $D = 7$ are shown in Fig. 2.7 (a) and Fig. 2.8 (a), respectively. A large enough domain $\Omega = [-a, a]^3$ is chosen with $a = 11$. By taking $\epsilon_s = 80$ and $\epsilon_m = 1$, the effective dielectric constant (EDC) in Eq. (2.18) is computed via numerical integration in all cases.

For the Gaussian dielectric model (2.9), $\epsilon_G(\vec{r})$ in the water phase is mainly determined by the positions and radii of the four atoms. For the two-dielectric model (2.4) and super-Gaussian model (2.15), the dielectric function is greatly influenced by the underlying molecular surfaces. In particular, in the two-dielectric model, $\epsilon_2(\vec{r}) = 1$ inside the solvent excluded surface (SES) and $\epsilon_2(\vec{r}) = 80$ outside. The SES is generated by the MSMS package [58] in the present study, see Fig. 2.6 for MSMS with different D values. It is seen that the solute domain initially becomes larger as D increases. However, as D keeps increasing, the reentry region in between the four atoms becomes smaller and smaller. Self-intersecting singularities are developed for $D = 6$ and $D = 6.5$. When $D = 7$, the system becomes four isolated balls, because with a probe radius 1.5, the probe sphere can freely pass the gaps between atoms. For the

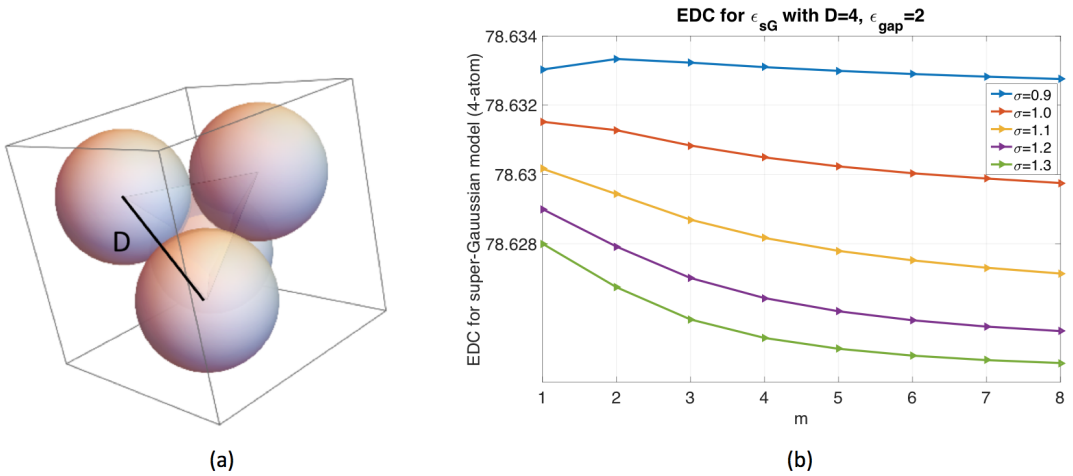


Figure 2.7: (a) Four-atom system in water solvent, where each atom-center is placed at the vertex of a regular tetrahedron of side $D = 4$Å. (b) Effective Dielectric Constant $\hat{\epsilon}_{sG}$ for different m and σ with $\epsilon_{gap} = 2$.

super-Gaussian model, $\epsilon_{sG}(\vec{r})$ is calculated based on the hypersurface function $S(\vec{r})$ of the minimal molecular surface (MMS) [44]. For a comparison, the MMS iso-surfaces with $S = 0.9$ at different D values are also shown in Fig. 2.6. A similar pattern as in the MSMS can be seen, i.e., the solute domain increases initially and then shrinks as D increases. Nevertheless, the MMS gives isolated atoms at an earlier D value, and never runs into geometrical singularities [20].

We first study the super-Gaussian model with fixed D and ϵ_{gap} values. Since the hypersurface function $S(\vec{r})$ plays an additional role in calculating dielectric distributions, the optimal m and σ results could be different from those of the one-atom system, because our previous study did not involve $S(\vec{r})$. We consider two extreme cases, $D = 4$ and $D = 7$, for studying m and σ.

With $D = 4$ and atomic radius 2, four balls are touching each other, leaving little space in between them. We thus fix $\epsilon_{gap} = 2$ in this scenario, and calculate the effective dielectric constant (EDC) $\hat{\epsilon}_{sG}$ for $\sigma \in \{0.9, 1.0, ..., 1.3\}$ and $m = 1, 2, ..., 8$, see Fig. 2.7 (b). As we have discussed for the one atom case before, with the increment of σ, the upper window part of super-Gaussian function becomes wider and it reduces the EDC.

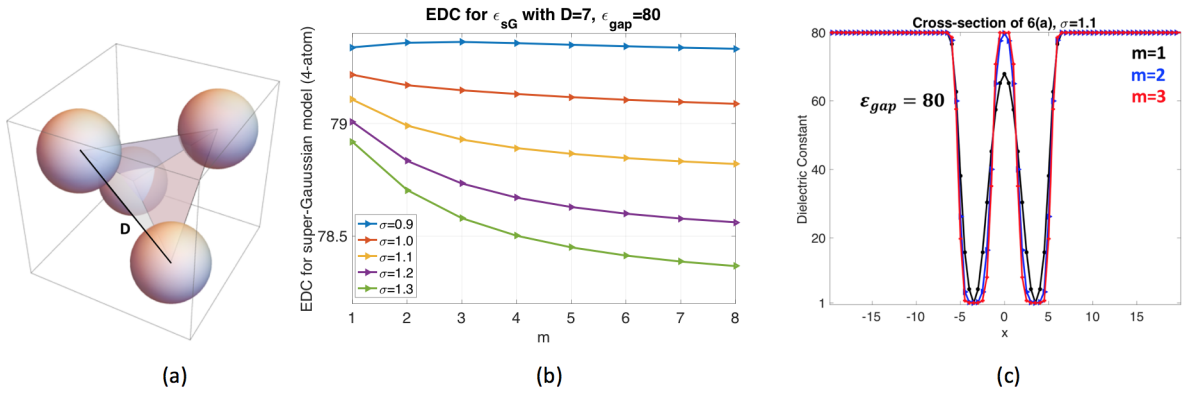


Figure 2.8: (a) Four-atom system in water solvent, where each atom-center is placed at the vertex of a regular tetrahedron of side $D = 7$Å. (b) Effective Dielectric Constant $\hat{\epsilon}_{sG}$ for different m and σ with $\epsilon_{gap} = 80$ (c) Cross-section of ϵ_{sG} for the model 8(a) which contains only two atoms.

We observe the same behavior in the 4-atom system too. Also, higher m values broaden the lower window-width too and it decreases the EDC $\hat{\epsilon}_{sG}$ of the super-Gaussian dielectric model as well. As we record the EDC, we observe very small variations in $\hat{\epsilon}_{sG}$ for different σ and m values, i.e., $78.626 < \hat{\epsilon}_{sG} < 78.634$. This insensitiveness indicates that the dielectric distribution is essentially dominated by the MMS hypersurface function $S(\vec{r})$ and $\epsilon_{gap} = 2$ for the current system with no cavities. In particular, the choice of $\epsilon_{gap} = 2$ does not let the dielectric distribution ϵ_{sG} bump up inside the small room in between four atoms, see Fig. 2.7 (a). From the parameter selection point of view, we will still suggest to use $m = 3$ or $m = 4$, while any choice of σ does not make much difference for $D = 4$.

When $D = 7$, a probe with radius 1.5 can freely access the interior of the four atoms. Both MSMS and MMS give isolated spheres in Fig. 2.6. Physically, the internal region should be treated as solvent. Thus we take $\epsilon_{gap} = 80$ in the super-Gaussian model, and calculate the EDC $\hat{\epsilon}_{sG}$ for $\sigma \in \{0.9, 1.0, ..., 1.3\}$ and $m = 1, 2, ..., 8$. As can be shown in Fig. 2.8 (b), $\hat{\epsilon}_{sG}$ is also decaying when m or σ is large. But the range of EDC values is quite large now, i.e., from 78.3666 to 79.3623, due to $\epsilon_{gap} = 80$. For a comparison, we consider the two-dielectric model, and the EDC value for the present

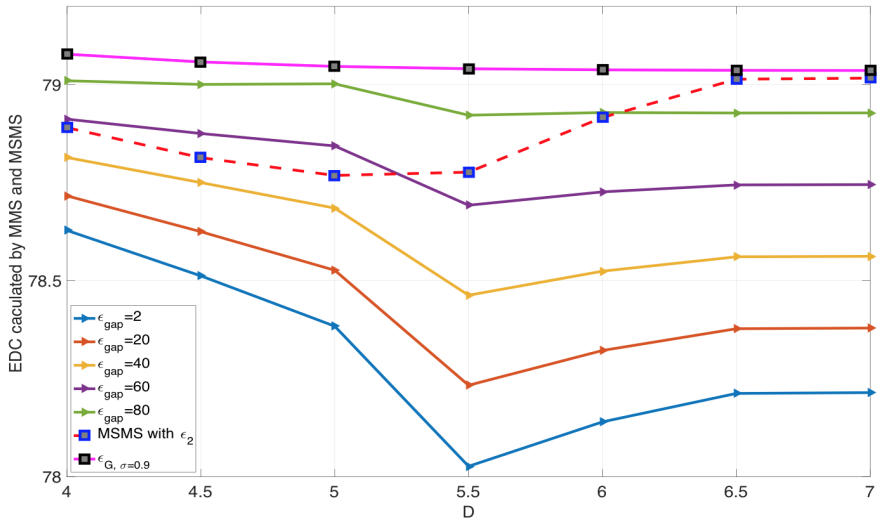


Figure 2.9: Comparison of EDC curves for $\hat{\epsilon}_{sG}$ (super-Gaussian model based on MMS), $\hat{\epsilon}_2$ (two-dielectric model based on MSMS) and $\hat{\epsilon}_G$ (Gaussian model).

setting is calculated as $\hat{\epsilon}_2 = 79.0171$. If we choose $\sigma = 1.0$, it can be seen that $\hat{\epsilon}_{sG}$ could approach $\hat{\epsilon}_2$ when $m \to \infty$. Again, this justifies our theory that the two-dielectric model is a limiting case of the proposed super-Gaussian model as m goes to infinity. For practical computations, a finite m shall be used. For the parameter combinations shown in Fig. 2.8 (b), $\hat{\epsilon}_{sG}$ produces a good approximation to $\hat{\epsilon}_2$ when $(\sigma, m) = (1.2, 1)$, $(1.1, 2)$ or $(1.1, 3)$. Nevertheless, for $m = 1$, the dielectric function actually does not reach 80 in the interior region, see Fig. 2.8 (c). Instead, associated with $\sigma = 1.1$, $m = 2$ or 3 would be a better choice.

Next, we study the super-Gaussian model with varying D and ϵ_{gap} values. As shown in the previous studies, with the presence of the hypersurface function $S(\vec{r})$, the changes of m and σ do not alter the EDC $\hat{\epsilon}_{sG}$ too much, especially for compactly packed regions. Hence, we will simply fix $\sigma = 1.1$ and $m = 3$ in the following, which are optimal values for $D = 7$. By considering $\epsilon_{gap} = \{2, 20, 40, 60, 80\}$, the EDC curves of $\hat{\epsilon}_{sG}$ with respect to D are depicted in Fig. 2.9. For a comparison, the EDC results of the two-dielectric and Gaussian models are also shown in Fig. 2.9. Here the Gaussian results are generated with the optimal $\sigma = 0.9$.

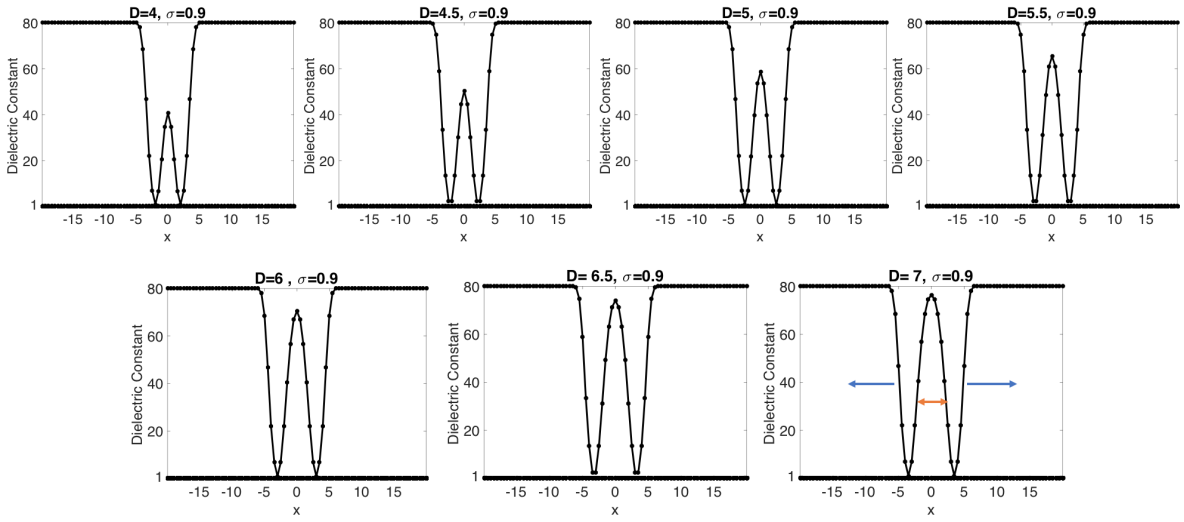


Figure 2.10: Gaussian dielectric model ϵ_G with $\sigma = 0.9$ for 4-atom cross section which consists of two atoms only. D varies from 4 to 7.

Model I: In the two dielectric model, we have $\epsilon_2 = 1$ within the four atoms and inside the MSMS surface in between the atoms, and $\epsilon_2 = 80$ otherwise. The EDC $\hat{\epsilon}_2$ is actually determined by the total volume of the solute domain. Therefore, the change of $\hat{\epsilon}_2$ in Fig. 2.9 can be related to the volume change in Fig. 2.6(a). In particular, as D increases from 4 to 5, $\hat{\epsilon}_2$ becomes smaller initially and achieves a minimal around $D = 5$. This is because the volume of the solute domain becomes larger in this period. Note that the volume increment is simply because the dimension of the system is larger, while the torus surface actually becomes thinner and thinner. Thus, the volume becomes smaller later, despite of the further increment of the dimension D. Consequently, $\hat{\epsilon}_2$ bounces up, and reaches a constant level for $D = 6.5$ and $D = 7$, for which the volumes are almost the same.

Model II: In the Gaussian model, the dielectric function ϵ_G defined in (2.9) only depends on the position and radii of the atoms, and there is no molecular surface behind it. Thus, as one can see in Fig. 2.9, when D increases, the EDC $\hat{\epsilon}_G$ is monotonically and slowly decreasing. To gain an in-depth understanding, we plot ϵ_G along a line passing two atom centers, see Fig.2.10. With fixed radii, the Gaussian

distributions for two atoms are unchanged as D increases. Hence, the increment of D only affects the dielectric value in between two atoms, which is higher and higher. This is why the EDC $\hat{\epsilon}_G$ behaves monotonically. For a very large D value, the Gaussian distribution is very close to the one for the one-atom system, for which $\sigma = 0.9$ is known to be the optimal value. Consequently, for $D = 6.5$ and $D = 7$, $\hat{\epsilon}_G$ is quite close to $\hat{\epsilon}_2$.

Model III: The EDC $\hat{\epsilon}_{sG}$ of the super-Gaussian model displays a similar pattern as $\hat{\epsilon}_2$ of the two-dielectric model for most ϵ_{gap} values except the limiting case $\epsilon_{gap} = 80$. However, the pattern of $\hat{\epsilon}_{sG}$ is not solely determined by the volume inside the MMS isosurface, because ϵ_{sG} is a function of space - changing in between the minimal value $\epsilon_m = 1$ and the maximal value ϵ_{gap} inside the solute domain. As Fig. 2.6(b) shows, the MMS generated isosurfaces are connected for $D = 4, 4.5$ and 5. Then from $D = 5.5$, the surfaces are disconnected and the four atoms are just isolated balls. Due to this topological change in the 4-atom system, there is a significant change in $\hat{\epsilon}_{sG}$ from $D = 5$ to $D = 5.5$. Before 5.5, as D increases from 4 to 5, the volume of solute domain enclosed by the MMS isosurface increases, while the connecting surfaces along the edges of the tetrahedron shrink inward. This volume increment induces the decrement of $\hat{\epsilon}_{sG}$. It is interesting to note that $\hat{\epsilon}_{sG}$ keeps decreasing from $D = 5$ to $D = 5.5$. This does not necessarily mean the isolated balls at $D = 5.5$ have larger volumes than the connected MMS region at $D = 5$. In fact, the volume at $D = 5.5$ is still large, because the four balls are much fatter than those for a bigger D value. Moreover, with a fat enough ball, ϵ_{sG} has the potential to reach its maximum, i.e. ϵ_{gap}. The combining effect of volume and ϵ_{sG} distribution determines the minimum of $\hat{\epsilon}_{sG}$ in Fig. 2.9 for most ϵ_{gap} values. As D becomes even bigger, the radii of MMS balls decreases so that $\hat{\epsilon}_{sG}$ becomes larger. Also, for all ϵ_{gap} values, the EDC $\hat{\epsilon}_{sG}$ is almost the same for both $D = 6.5$ and $D = 7$. For the limiting case $\epsilon_{gap} = 80$, it turns out that particular MMS shape does not affect $\hat{\epsilon}_{sG}$, because $\epsilon_{gap} = \epsilon_s = 80$. Basically, $\hat{\epsilon}_{sG}$ just takes two values, one for a connected region and another for isolated balls.

In comparison of the EDC results of three models, we found that the Gaussian model is significantly different from the other two, while two-dielectric and super-Gaussian models share the similar physics: the volume of solvent accessible region is determined by the size of the cavity in a convex manner, so that the dependence of the EDC on the cavity size is concave. Moreover, besides the MMS hypersurface function $S(\vec{r})$, the ϵ_{sG} is also affected by the adjustable parameters m, σ, and ϵ_{gap}. If one changes m or σ, the EDC lines of $\hat{\epsilon}_{sG}$ in Fig. 2.9 will be shifted up or down, and the concave feature shall be the same. If one wishes to match $\hat{\epsilon}_{sG}$ with $\hat{\epsilon}_2$, Fig. 2.9 suggests that a larger ϵ_{gap} should be employed for a larger D. In other words, the optimal ϵ_{gap} depends on the size of the cavity.

2.4.3 Effective dielectric constant analysis in both water and vacuum phases

In our last EDC analysis, we consider both water and vacuum states in the solvent region. As we know, the electrostatic solvation energy is calculated as the energy difference of the macromolecule in between the water and vacuum. Physically, the homogeneous or inhomogeneous dielectric distribution of the protein should remain unchanged in both states so that the energy difference makes sense. Consequently, the difference of $\epsilon(\vec{r})$ should not depend on a particular dielectric model for the solute, but relates to the solvent domain and property.

For the following experiments, we will calculate the EDC difference between water and vacuum phases. For this purpose, we need to explicitly specify the dependence of the solvent dielectric constant for the three models. In both two-dielectric and Gaussian models, we thus have $\epsilon_2(\vec{r}, \epsilon_s)$ and $\epsilon_G(\vec{r}, \epsilon_s)$, respectively, while the super-Gaussian model takes the form $\epsilon_{sG}(\vec{r}, \epsilon_{out})$. The EDC difference is defined as

$$\Delta\hat{\epsilon} = \frac{\int_\Omega \Big[\epsilon(\vec{r}, 80) - \epsilon(\vec{r}, 1)\Big] d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\int_\Omega \epsilon(\vec{r}, 80) d\vec{r}}{\int_\Omega d\vec{r}} - \frac{\int_\Omega \epsilon(\vec{r}, 1) d\vec{r}}{\int_\Omega d\vec{r}}, \tag{2.22}$$

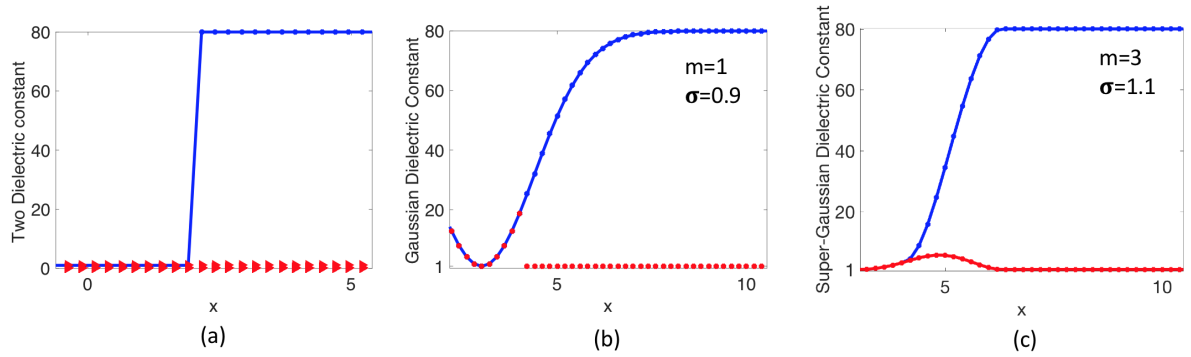


Figure 2.11: Dielectric models (a) ϵ_2: discontinuous in water phase and continuous in vacuum phase, (b) ϵ_G: continuous in water phase and discontinuous in vacuum phase ("surface-cut"), (c) ϵ_{sG}: continuous in both water and vacuum states

where $\epsilon \in \{\epsilon_2, \epsilon_G, \epsilon_{sG}\}$. We note that because one solves different PDEs, i.e., the PB equation in the water phase and the Poisson equation in the vacuum phase, the EDC difference may not have directly influence the electrostatic solvation energy. Nevertheless, $\Delta\hat{\epsilon}$ is still a useful quantity for investigating different dielectric models.

In Fig. 2.11, illustrations of the three models in both states are depicted. For the two-dielectric model, $\epsilon_2(\vec{r}, 80)$ is discontinuous, while $\epsilon_2(\vec{r}, 1)$ is continuous because $\epsilon_m = 1$ in the present study. If $\epsilon_m > 1$, $\epsilon_2(\vec{r}, 1)$ is discontinuous too in the vacuum. For the Gaussian dielectric model, $\epsilon_G(\vec{r}, 80)$ is continuous in the water phase, but it is discontinuous in the vacuum phase, due to a surface-cut. In the proposed super-Gaussian model, both $\epsilon_{sG}(\vec{r}, 80)$ and $\epsilon_{sG}(\vec{r}, 1)$ are continuous, respectively, in the water and vacuum states. Another thing that can be observed in Fig. 2.11 is that inhomogeneous solute dielectric models will impact solvent region nearby. In particular, for the Gaussian model, the dielectric values near the protein are influenced by the parameter σ in the water phase. In the super-Gaussian model, such values are affected by both m and σ for both water and vacuum phases.

We will consider the same four atom system of the previous study. For simplicity, we test only one case with $D = 7$ for computing $\Delta\hat{\epsilon}$.

Model I: In the two-dielectric model, the molecular surface is generated by the

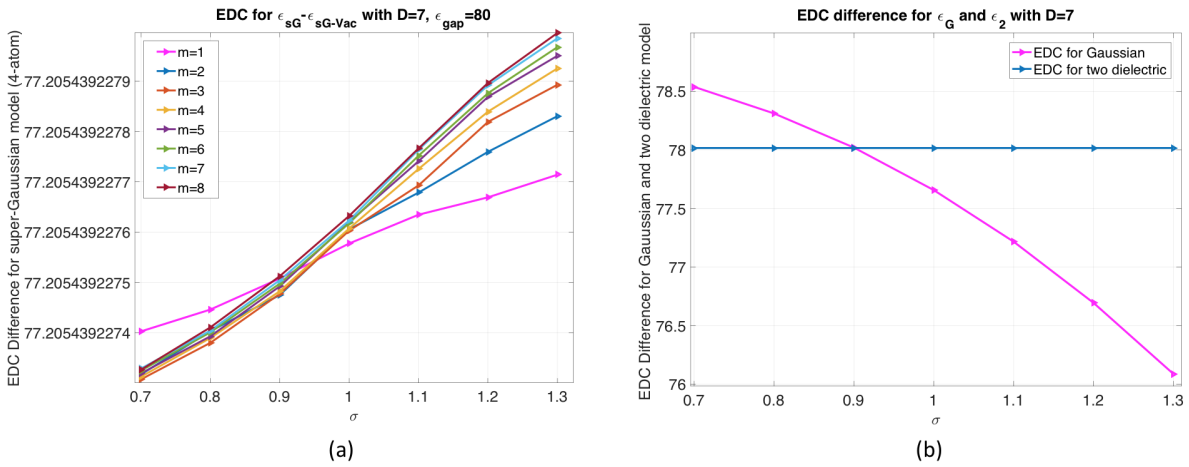


Figure 2.12: (a) EDC difference $\Delta\hat{\epsilon}_{sG}$ for different $m \in \{1, 2, ..., 8\}$ and $\sigma \in \{0.7, 0.8, ..., 1.3\}$, (b) $\Delta\hat{\epsilon}_G$ for different σ and $\Delta\hat{\epsilon}_2$

MSMS package. Since $\epsilon_2(\vec{r}, 1) = 1$ throughout the domain Ω, we have simply

$$\Delta\hat{\epsilon}_2 = \frac{\int_\Omega \Big[\epsilon_2(\vec{r}, 80) - 1\Big] d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\int_\Omega \epsilon_2(\vec{r}, 80) d\vec{r}}{\int_\Omega d\vec{r}} - 1 = \hat{\epsilon}_2 - 1. \tag{2.23}$$

For $D = 7$, we have $\Delta\hat{\epsilon}_2 = 78.0171$ numerically, which is exactly one unit less than the EDC $\hat{\epsilon}_2$ studied in the previous study.

Model II: For the Gaussian model, we consider several σ values, i.e., $\sigma = 0.7, ..., 1.3$. According to (2.22), the EDC difference could be calculated by considering the water and vacuum phases separately. In the water phase, based on $\epsilon_G(\vec{r})$ given in (2.9), the EDC value is within $(77.13635, 79.54655)$. In the vacuum phase, when $\epsilon_G(\vec{r})$ exceeds 20, a surface-cut is conducted to set dielectric constant as zero [1], see Fig. 2.11(b). The EDC for the vacuum case is within $(1.00738, 1.04756)$. By taking the difference, $\Delta\hat{\epsilon}_G$ is within the range of $(76.08879, 78.53918)$. Moreover, $\Delta\hat{\epsilon}_G$ depends on σ significantly, see Fig.2.12(b). In the same figure, $\Delta\hat{\epsilon}_2 = 78.0171$ is shown as a constant line. From the parameter selection point of view, this figure shows again $\sigma = 0.9$ is an optimal value for the Gaussian model. This is because with $D = 7$, four atoms are completely separated, so that the present result is consistent with a

single atom study. However, from a different perspective, the dependence of $\Delta\hat{\epsilon}_G$ on σ indicates that the Gaussian model negatively impacts on the dielectric value in the solvent region. The inhomogeneous model here is designed for the protein and should be confined within the solute. Unfortunately, this is not the case for the Gaussian model.

Model III: In the super-Gaussian model, we also fix $\epsilon_{gap} = 80$ for $D = 7$. Different parameter values are tested for $m \in \{1, 2, ..., 8\}$ and $\sigma \in \{0.7, 0.8, ..., 1.3\}$. By taking $\epsilon_{out} = 80$ in the water and $\epsilon_{out} = 1$ in the vacuum, we note that in Eq. (2.15), $S(\vec{r})\epsilon_{in}(\vec{r})$ is simply canceled out when computing EDC difference:

$$\Delta\hat{\epsilon}_{sG} = \frac{\int_\Omega \Big[\epsilon_{sG}(\vec{r}, 80) - \epsilon_{sG}(\vec{r}, 1)\Big] d\vec{r}}{\int_\Omega d\vec{r}} = \frac{\int_\Omega 79\Big[1 - S(\vec{r})\Big] d\vec{r}}{\int_\Omega d\vec{r}}. \tag{2.24}$$

This is confirmed numerically. In Fig. 2.12(a), $\Delta\hat{\epsilon}_{sG}$ is plotted against σ for different m values. The vertical values change from 77.2054392274750 to 77.2054392279961, for which the difference takes place at the tenth decimal place. Thus, the EDC difference is solely dominated by the MMS hypersurface function $S(\vec{r})$. With $S(\vec{r})$, the impact of the super-Gaussian model is confined within the solute, as shown by the present EDC analysis.

Both the two-dielectric and the super-Gaussian models yield a constant $\Delta\hat{\epsilon}$. But this does not mean that the change of parameter values will have no impact on electrostatic solvation energy in the super-Gaussian model. For example, a different choice of (m, σ) pair will produce a different "bump" in the vacuum case in Fig. 2.11(c). Such a bump near solute-solvent boundary is driven by a combined mechanism: away from atom centers ϵ_{in} becomes larger, while away from the protein, it will damp out ϵ to one. Because ϵ_{in} depends on (m, σ), the height and width of the bump depend on the pair. Moreover, since different PDEs will be solved in the water and vacuum phases, the electrostatic solvation energy will rely on (m, σ) in practice.

2.4.4 Discussions

In this chapter, we have carried out an effective dielectric constant (EDC) analysis for three cases, which helps us to understand the role of each parameter, including m, σ, and ϵ_{gap}, in the super-Gaussian dielectric function $\epsilon_{sG}(\vec{r})$. For the EDC difference studied in the third case, it is independent of these parameters, and just relies on the MMS hypersurface function $S(\vec{r})$. For the first case without involving $S(\vec{r})$, the impact of m and σ on the EDC $\hat{\epsilon}_{sG}$ has been identified. A comprehensive EDC analysis has been conducted for the second order, which tells us more about parameters. Basically, with a fixed ϵ_{gap}, optimal m and σ can be established. Moreover, due to the influence of $S(\vec{r})$, the super-Gaussian model behaves robustly with respect to m and σ, in the sense that the EDC will not change too much for different m and σ values. Furthermore, our analysis indicates that ϵ_{gap}, which determines the maximum of $\epsilon_{sG}(\vec{r})$ inside the solute, should be larger when the size or volume of the cavity increases. For proteins without cavities, we usually recommend a small value, such as $\epsilon_{gap} = 2$, which does not deviate too much from $\epsilon_m = 1$ of the two-dielectric model. On the other hand, the selection of ϵ_{gap} for proteins with cavities is not an easy task in practical computations. In our opinion, physical considerations have to be taken into account, so that the super-Gaussian PB model can capture as many atomic details as possible in the continuum electrostatics modeling.

For proteins containing cavities and channels, one critical issue on selecting ϵ_{gap} is whether a cavity is empty or filled with pure water or crystallographic water. Trapped water molecules tend to interact with the protein via either hydrogen bonds or Van der Walls forces. As a consequence of these interactions, the water molecules are considered to lose their flexibility. Thus, the cavity water (or fluid) could have a smaller dielectric constant in comparison to the bulk water, but it is still larger than that of the amino acids. Moreover, the size or volume of the cavity is also important. Depending on the cavity size, confined water molecules exhibit a different ability to reorient in response

to the local electrostatic field which affects their rotational polarizability. This further alters the dielectric value. Furthermore, the situation becomes more complicated in the ion channel modeling, in which the Poisson equation is used to calculate the force encountered by permanent ions. In this scenario, besides dielectric values for the protein and bulky water, one also needs to specify ϵ for water in the pore even in classical models. Physically, the dielectric value in the ion-channel shall be higher than that in regular cavities, owing to the mobility of ions. Therefore, the optimal ϵ_{gap} has to be determined based on a particular macromolecule, and varies for different systems. Ideally, in-depth physical investigation or biological simulation shall be carried out for selecting a proper dielectric value for cavities and pores. For instance, Brownian dynamics simulations have been conducted in [41] to decide ϵ values for protein channels to be used in solving the PB equation.

CHAPTER 3

NUMERICAL ALGORITHM

In this chapter, we discuss how to discretize the PB equation (2.16) and Poisson equation (2.17) in the proposed super-Gaussian model. In solving the two-dielectric PB equation (2.1), special interface treatments [16, 17, 19] are required for high order spatial discretizations, in order to handle the non-smoothness of the solution across the dielectric interface. Such a difficulty is simply bypassed in the smooth PB equation (2.16), because the solution now is C^∞ continuous throughout the domain. However, the nonlinearity term $(1-S)\bar{\kappa}^2 \sinh(u)$ in (2.16) introduces additional challenges numerically. In particular, near the solute-solution boundaries, the MMS characteristic function S is changing from one for solute to zero for solvent (see Fig. 2.3(a) for an illustration). Thus $(1-S)$ is not completely zero at some places for which the two-dielectric model will treat them as solute domain. If such a place happens to be close to an atom center, the magnitude of the potential u is not small then. Consequently, $\sinh(u)$ will be exponentially large. Even though $(1-S)$ is very close to zero, $\sinh(u)$ could still be dominant in many cases. This yields the so-called nonlinear instability, which has been observed in other smooth PB models before [42, 43]. In present study, we will employ the analytical treatment introduced in [43, 54] to overcome the nonlinear instability within a pseudo-time framework.

Consider a uniform mesh in both space and time. Without the loss of generality, we assume the grid spacing h in all x, y, and z directions to be the same. Denote the time increment as Δt. For a function u at a grid point (x_i, y_j, z_k) and time instant t_n, we denote $u^n_{i,j,k} = u(x_i, y_j, z_k, t_n)$.

3.1 Pseudo-time solution of the Poisson-Boltzmann equation

In the pseudo-time approach, a pseudo-time derivative will be added to the PB equation [42]. Consequently, (2.16) becomes a time dependent PB equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (\epsilon_{sG} \nabla u) - (1 - S)\bar{\kappa}^2 \sinh(u) + S\rho_m, \quad \text{in } \Omega, \tag{3.1}$$

with the same boundary condition (2.3). By using a trivial initial value $u = 0$, one numerically integrates (3.1) for a sufficiently long time period to reach a stable state. The solution to the original nonlinear PB equation (2.16) is essentially recovered by the steady state solution of the pseudo-time dependent process (3.1).

A first order time splitting scheme [43, 54] will be employed for solving (3.1). The time stepping of (3.1) over the time interval $[t_n, t_{n+1}]$ can be carried out in two stages

$$\frac{\partial w}{\partial t} = -(1 - S)\bar{\kappa}^2 \sinh(w), \quad \text{with } w^n = u^n \tag{3.2}$$

$$\frac{\partial v}{\partial t} = \nabla \cdot (\epsilon_{sG} \nabla v) + S\rho_m, \quad \text{with } v^n = w^{n+1} \tag{3.3}$$

We then set $u^{n+1} = v^{n+1}$. The numerical solution u^{n+1} differs from the direct solution of (3.1) by an error on the order one, i.e., $O(\Delta t)$. A second order time splitting has also been developed in [43, 54], by dividing the process into three stages.

The nonlinear sub-system (3.2) is integrated analytically. For the region inside VDW balls with $S(\vec{r}) = 1$, we do not need to solve this equation. We will just simply set $w^{n+1} = w^n$. When $S(\vec{r}) < 1$, the nonlinear term is calculated as [43]

$$w^{n+1} = \ln\Big[\frac{\cosh(\frac{1}{2}(1 - S)\bar{\kappa}^2 \Delta t) + \exp(-w^n)\sinh(\frac{1}{2}(1 - S)\bar{\kappa}^2 \Delta t)}{\exp(-w^n)\cosh(\frac{1}{2}(1 - S)\bar{\kappa}^2 \Delta t) + \sinh(\frac{1}{2}(1 - S)\bar{\kappa}^2 \Delta t)}\Big]. \tag{3.4}$$

In the MMS generation, we have carefully filtered the results from the fast algorithm [44], so that the hypersurface function is strictly between 0 and 1, i.e., $S \in [0, 1]$. Together with (3.4), this guarantees that the present PB algorithm is free of

nonlinear instability.

3.2 Alternating direction implicit (ADI) scheme

A Douglas-Rachford type alternating direction implicit (ADI) scheme will be applied to solve the linear parabolic equation (3.3). To this end, an implicit Euler spatial-temporal discretization of (3.3) is formulated first

$$v_{i,j,k}^{n+1} = v_{i,j,k}^{n} + \Delta t(\delta_x^2 + \delta_y^2 + \delta_z^2)v_{i,j,k}^{n+1} + \Delta t S_{i,j,k} Q_{i,j,k} \tag{3.5}$$

where $Q_{i,j,k}$ is the fractional charge at grid point (x_i, y_j, z_k), which is obtained by using the trilinear interpolation to distribute all charges in the charge density ρ_m. Here δ_x^2, δ_y^2 and δ_z^2 are the central difference operators along x,y and z directions respectively,

$$\begin{aligned}
\delta_x^2 v_{i,j,k}^n &= \frac{1}{h^2}\Big(\epsilon(x_{i+1/2}, y_j, z_k)(v_{i+1,j,k}^n - v_{i,j,k}^n) + \epsilon(x_{i-1/2}, y_j, z_k)(v_{i-1,j,k}^n - v_{i,j,k}^n)\Big) \\
\delta_y^2 v_{i,j,k}^n &= \frac{1}{h^2}\Big(\epsilon(x_i, y_{j+1/2}, z_k)(v_{i,j+1,k}^n - v_{i,j,k}^n) + \epsilon(x_i, y_{j-1/2}, z_k)(v_{i,j-1,k}^n - v_{i,j,k}^n)\Big) \\
\delta_z^2 v_{i,j,k}^n &= \frac{1}{h^2}\Big(\epsilon(x_i, y_j, z_{k+1/2})(v_{i,j,k+1}^n - v_{i,j,k}^n) + \epsilon(x_i, y_j, z_{k-1/2})(v_{i,j,k-1}^n - v_{i,j,k}^n)\Big)
\end{aligned}$$

where we have dropped the subscript sG in the ϵ function for simplicity. In these finite difference discretizations, the dielectric function is needed on half grid nodes, such as $\epsilon(x_{i+1/2}, y_j, z_k)$. Because the MMS hyperfunction is obtained numerically, we only know S function on grid nodes, i.e., $S_{i,j,k} = S(x_i, y_j, z_k)$. In the present, we will first generate ϵ_{sG} on (x_i, y_j, z_k) grid nodes. Then a linear interpolation at (x_i, y_j, z_k) and (x_{i+1}, y_j, z_k) is conducted for determining $\epsilon(x_{i+1/2}, y_j, z_k)$.

In the ADI scheme, instead of solving a three-dimensional (3D) linear system,

(3.5) is solved in x, y, and z directions alternatively

$$
\begin{aligned}
(1-\Delta t\delta_x^2)v_{i,j,k}^* &= [1+\Delta t(\delta_y^2+\delta_z^2)]v_{i,j,k}^n + \Delta t S_{i,j,k}Q_{i,j,k} \\
(1-\Delta t\delta_y^2)v_{i,j,k}^{**} &= v_{i,j,k}^* - \Delta t\delta_y^2 v_{i,j,k}^n \\
(1-\Delta t\delta_z^2)v_{i,j,k}^{n+1} &= v_{i,j,k}^{**} - \Delta t\delta_z^2 v_{i,j,k}^n
\end{aligned} \tag{3.6}
$$

By eliminating the intermediate solutions v^* and v^{**}, one can show that the Douglas-Rachford ADI Scheme (3.6) is a higher order perturbation of the implicit Euler scheme (3.5) [43]. The overall temporal order is one, because both the time splitting and ADI schemes are first order accurate in time. For smooth solutions, the finite difference discretization has order two in space. Moreover, since only one-dimensional (1D) linear systems shall be solved in each stage of the ADI scheme (3.6) and such 1D systems are tridiagonal, the algebraic computation is very efficient based on the Thomas algorithm. The complexity of each time step is on the order of $O(N)$, where N is the degree of freedom in all of x, y, and z directions.

We note that the same ADI scheme has been previously applied to solve the PB equation in a two-dielectric setting with a sharp interface [54] and in a coupled system [43], in which the MMS hyperfunction $S(x, y, z)$ is evolved in time as well. However, such an ADI scheme is conditionally stable for real proteins, even though the scheme is fully implicit. As to be illustrated in our numerical studies, with a C^∞ dielectric setting, the ADI scheme now becomes unconditionally stable in protein studies.

3.3 Poisson equation in the vacuum phase

For the vacuum phase, ϵ_{sG} is calculated by (2.15) with $\epsilon_{out} = 1$. With inhomogeneous dielectric values inside the protein, the Poisson equation (2.17) cannot be solved by the fast Poisson solver as in the two-dielectric PB model. Instead of solving the Poisson equation (2.17) as a boundary value problem, we will solve it via a

pseudo-time approach too. This is motivated by the fact that there is usually a systematic error cancellation, when one applies the same algorithm for solving the PB equation in water phase and the Poisson equation in vacuum phase [65]. Thus, we rewrite the Poisson equation (2.17) in the vacuum phase into a time dependent one

$$\frac{\partial u_0}{\partial t} = \nabla \cdot (\epsilon_{sG} \nabla u_0) + S\rho_m, \quad \text{in } \Omega. \tag{3.7}$$

Then, the ADI discretization of (3.7) is exactly the same as that for Eq. (3.3).

3.4 Electrostatic free energy

After solving the time dependent PB and Poisson equations until the steady state, we denote the convergent solution, respectively, to be $u(x_i, y_j, z_k)$ and $u_0(x_i, y_j, z_k)$, where (x_i, y_j, z_k) is a grid node. To calculate the electrostatic free energy defined in (2.5), we first note that this definition is valid in super-Gaussian model too, i.e.,

$$\Delta G = \frac{1}{2} \int_\Omega \rho_m (u(\vec{r}) - u_0(\vec{r})) d\vec{r} = \frac{1}{2} \int_\Omega S\rho_m (u(\vec{r}) - u_0(\vec{r})) d\vec{r}. \tag{3.8}$$

This is because the charge density ρ_m is nonzero only inside the VDW atoms, for which S always equals to one. In the present study, the electrostatic free energy is calculated based on grid node values

$$\Delta G = \frac{1}{2} \sum_i \sum_j \sum_k Q_{i,j,k} (u(x_i, y_j, z_k) - u_0(x_i, y_j, z_k)), \tag{3.9}$$

where the summation is conducted for all (i, j, k) nodes for which $Q_{i,j,k}$ is nonzero, i.e., surrounding the singular charges in ρ_m. Moreover, electrostatic potentials u and u_0 are usually rescaled by a constant 0.592183 corresponding to room temperature (298K) so that they are in units of kcal/mol/e_c.

CHAPTER 4

NUMERICAL VALIDATION

In this chapter, we will solve the NPB equation on a sphere, for which an analytical solution of electrostatic free energy is available in a two-dielectric setting. This enables us to validate the proposed super-Gaussian dielectric model and select model parameters. Numerically, we will also verify the convergence and stability of the pseudo-time ADI method.

4.1 Benchmark problem

Consider a single charge q at the center of a sphere with radius r_0. Here we take $q = 1e_c$ and $r_0 = 2\text{Å}$, and assume the center being the origin of our coordinate system. An analytical solution of electrostatic free energy ΔG is admissible if we assume a two-dielectric setting: $\epsilon = \epsilon_m$ inside the sphere and $\epsilon = \epsilon_s$ outside. By taking $\epsilon_m = 1$ and $\epsilon_s = 80$, we have

$$\begin{aligned}\Delta G &= -\frac{q^2}{2}\Big(\frac{1}{\epsilon_m} - \frac{1}{\epsilon_s}\Big)\frac{1}{r_0} \quad e_c^2/\text{Å} \\ &= -\frac{q^2}{2}\Big(\frac{1}{\epsilon_m} - \frac{1}{\epsilon_s}\Big)\frac{1}{r_0} \times 332.06364 \quad kcal/mol \\ &= -81.9782 \quad kcal/mol \end{aligned} \tag{4.1}$$

4.2 Modal validation and parameters

In the super-Gaussian model, we take $\epsilon_{out} = 80$ for the water phase. Then the ADI method is employed for solving the pseudo-time NPB equation (3.1). The computational domain is taken as $\Omega = [-8, 8]^3$. On the boundary $\partial\Omega$, the Dirichlet boundary condition (2.3) is assumed in a single charge setup. By using a initial

condition $u = 0$, Eq. (3.1) will be numerically integrated until the steady state. Similarly, in the vacuum state with $\epsilon_{out} = 1$, the pseudo-time Poisson equation (3.7) will be solved by the ADI method with the corresponding boundary condition. Then, the electrostatic solvation energy can be computed by (3.9). Numerically, the same spacing is used in all three directions $h = \Delta x = \Delta y = \Delta z$. We will take $h = 0.5$ as in most PB computations.

In the previous chapter, we have discussed about the choice of m, σ and ϵ_{gap} through the Effective Dielectric Constant (EDC) analysis. For the present problem, we will examine these parameters again by considering free energy. We note that with different dielectric setting, our super-Gaussian results will not converge to the analytical value, which is based on a two-dielectric setting. However, it makes sense to adjust parameters so that the new dielectric model could produce energy values that are close to the two-dielectric model. This is particular convenient if people want to use it to replace an existing two-dielectric PB solver in a large package. For this reason, we will simply take $\epsilon_{gap} = 2$, which gives the least difference in comparing with $\epsilon_m = 1$ within the sphere.

By considering $m \in \{1, 2, ..., 8\}$ and $\sigma \in \{0.9, 1.0, ..., 1.3\}$ for the super-Gaussian function ϵ_{sG}, the steady state energies are shown in Fig. 4.1. The exact value -81.9782 kcal/mol is also shown for a reference. A few pairs of (m, σ) are found to produce good approximations to the two dielectric model, i.e., $(1.2, 5), (1.2, 6), (1.2, 7), (1.2, 8)$ and $(1.3, 3)$. Among them, we will mainly focus on $m = 3$ and $\sigma = 1.3$ in the following energy calculations, to avoid using a large m.

4.3 Numerical convergence and stability

By fixing $m = 3$, $\sigma = 1.3$, and $\epsilon_{gap} = 2$, we investigate the performance of the pseudo-time ADI algorithm. By taking $\Delta T_e = 0.01$, we first examine the steady state convergence. The time history given in Fig. 4.2 (a) shows that ΔG is decreasing monotonically before the steady state is reached. The stopping criterion issue of the

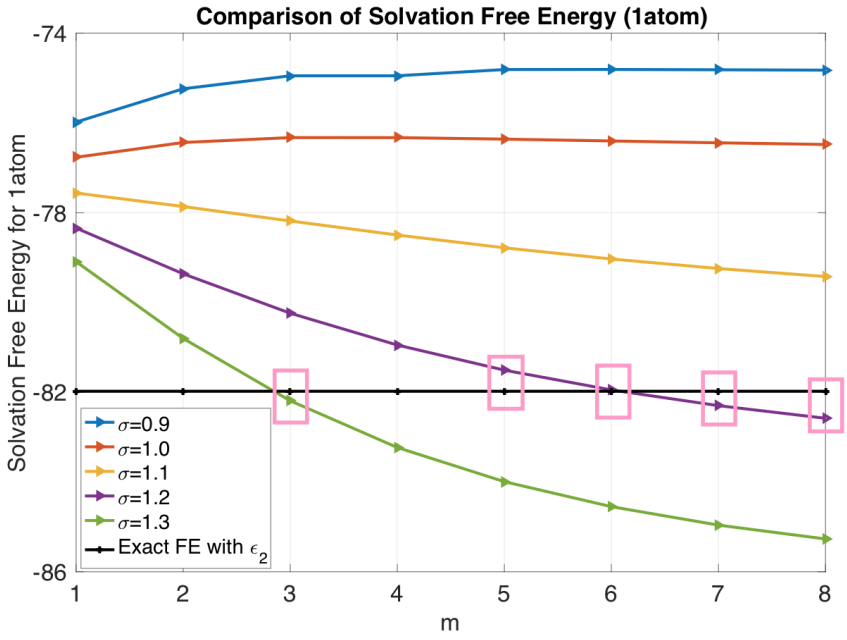


Figure 4.1: Solvation free energy for $\sigma \in \{0.9, 1.0, ..., 1.3\}$ and $m \in \{1, 2, ..., 8\}$ with $h = 0.5$.

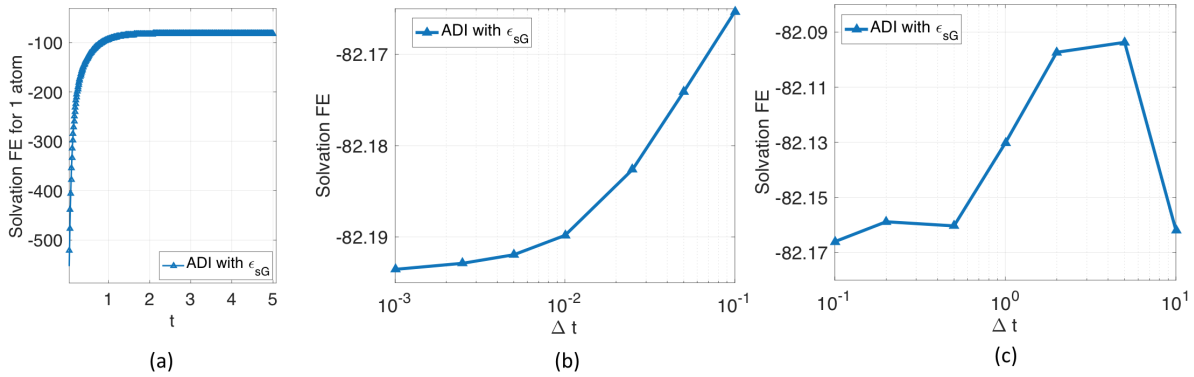


Figure 4.2: Pseudo-time ADI algorithm for one atom system. (a) Steady state convergence; (b) Temporal accuracy; (c) Stability.

pseudo-time ADI algorithm has been discussed in [43] for two-dielectric PB equation. Basically, the computation will stop if either $t \geq T_e$ or the energy difference in between two time steps is less than a tolerance. We note that in the existing pseudo-time PB studies based on two-dielectric media [43, 45, 54], the steady state is usually reached fairly quick, around $T_e = 4$. For the present inhomogeneous dielectric medium, one has

to choose $T_e \in [50, 100]$ to guarantee convergence. We will take $T_e = 100$ in the following studies, unless specified otherwise.

We next examine the temporal accuracy of the ADI algorithm. With $T_e = 100$, free energies are generated by using different Δt, see Fig. 4.2 (b). Obviously, as Δt becomes smaller and smaller, the free energy approaches a certain limiting value. The vertical range is actually quite small. In practice, $\Delta T_e = 0.01$ is enough to produce a reliable energy estimate.

We finally examine the stability of the pseudo-time ADI algorithm. We note that in a two-dielectric setting, this ADI algorithm does not achieve the unconditional stability, even though it is fully implicit [54]. In particular, to fulfill the stability requirement, one has to choose $\Delta t = h^2/20$ in protein studies [54]. Because Δt is small, the resulting algorithm could be inefficient, when T_e is large. With the C^∞ continuous ϵ function in both water and vacuum states, the pseudo-time ADI algorithm is unconditionally stable in the super-Gaussian model. We demonstrate this by taking some large Δt values and conduct each computation with $10,000$ time steps. As can be seen in Fig. 4.2 (c), the energy value with a large Δt could be slightly different. Nevertheless, the ADI algorithm remains stable for any large Δt. By considering both convergence and efficiency, we will choose $T_e = 100$ and $\Delta T_e = 0.01$ in the following studies.

CHAPTER 5

BIOLOGICAL APPLICATION

In this chapter, we further explore the performance of the super-Gaussian PB model and ADI algorithm by studying free energies of protein systems. We first discuss how a real protein is implemented in the super-Gaussian model. Then, we test different parameter values for a particular protein. With a reasonable choice of domain and parameters, we study solvation free energies for a set of proteins. We finally consider a protein with cavities to demonstrate how cavities can be represented via inhomogeneous dielectric distributions. In all studies, a large enough computational domain Ω is assumed and a uniform mesh with $h = \Delta x = \Delta y = \Delta z = 0.5$ is adopted.

5.1 Protein structure preparation and simulation setup

We have collected a set of proteins from the RCSB protein data bank (pdb). In this collection, the proteins consist of at least 500 atoms. Usually, we download the pdb format which is a standard representation for macromolecular structure data obtained from X-ray diffraction and NMR studies. This format preserves the details of water molecules, ions, nucleic acids, ligands etc. With the aid of the PDB2PQR program from the APBS package, we extract three important data for each atom involved in the protein, i.e., centers $\vec{r}_i = (x_i, y_i, z_i)$, radius R_i, and partial charge q_i, for $i = 1, 2, \ldots N_m$. These data are stored in two files, one with extension .xyzr which contains numerical values for $\vec{r}_i$ and R_i in four columns. Another file with extension .xyzq contains numerical values for $\vec{r}_i$ and q_i.

The density function of the super-Gaussian model defined in (2.11) and (2.12) depends on the centers and radii of all atoms. It is time-consuming if one computes the

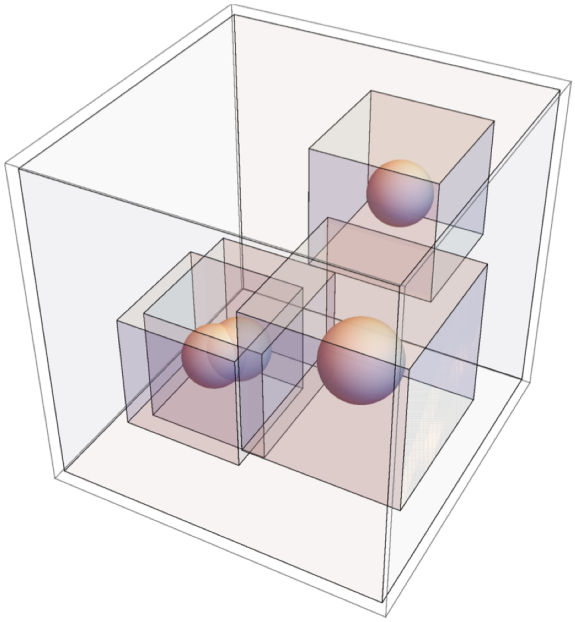


Figure 5.1: Influence domain

density of every atom by using the entire domain Ω. In fact, the density of the i^{th} atom $g_i^s(\vec{r})$ decays quickly away from its center $\vec{r}_i$, so that one does not need to calculate this function in the far field. By carefully examining the numerical truncation so that it will not affect the subsequent computations, we have introduced an influence domain for each atom, which is defined as a cubic box with dimension $[-D, D]^3$ and centered at $\vec{r}_i$. See Fig. 5.1 for an illustration. In particular, in our considerations, we consider maximum relative variance σ as 1.3. The influence domain dimension possibly depends on the radius R_i and the order of the super-Gaussian function m. An empirical function is found to be satisfactory in our computations: $D = 2R_i\left(1 + m^{\frac{-m}{2}}\right)$, which takes its maximum $D = 4R_i$ at $m = 1$. As a monotonically decreasing function, D will be very close to its asymptotic value $\lim_{m\to\infty} 2R_i\left(1 + m^{\frac{-m}{2}}\right) = 2R_i$ when m is large. In other words, for $m = 1$ (in Gaussian density function), the dimension of the influence cube is four times of the atomic radius (R_i) and as $m \to \infty$, the cube's dimension shrinks down to double of the radius.

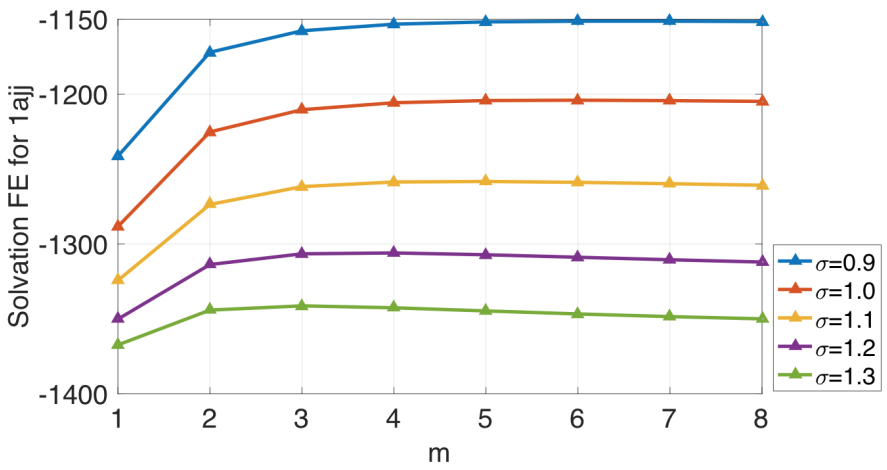


Figure 5.2: Solvation free energy for protein 1ajj, $m \in \{1, 2, ..., 8\}$ and $\sigma \in \{0.9, 1.0, ..., 1.3\}$

5.2 Solvation free energies of proteins

For studying our super-Gaussian model on proteins, we first experiment the ADI algorithm with ϵ_{sG} on a sample protein, say 1ajj (pdb id) for different m and σ. Since 1ajj does not contain any cavity inside the molecular surface, we set $\epsilon_{gap} = 2$. The performance of the pseudo-time ADI with ϵ_{sG} is recorded in Fig. (5.2). Here we considered $\Delta t = 0.01$ and $T = 100$. The solvation free energy for 1ajj at different (m, σ) values ranges from $[-1367.4, -1151.55]$. For a fixed σ, the increment of m from 1 to 3 gives rise to a higher energy, while the energy declines slowly as m is even larger. Numerically, the energy difference between $m = 3$ and $m > 3$ is not significant, which justifies our usual choice of $m = 3$. Nevertheless, the choice of σ does have a strong impact on energies, as shown in Fig. 5.2. Without comparing with results from other computational models, we will continue to use $\sigma = 1.3$ for simplicity.

We next investigate the pseudo-time ADI algorithm by fixing $m = 3$, $\sigma = 1.3$ and $\epsilon_{gap} = 2$. We first consider the steady state convergence by using $\Delta t = 0.01$. The time-lapse data is displayed in Fig. 5.3(a). Here the stopping criteria of the numerical computation are the same as those described in section 4.3. The solvation free energy

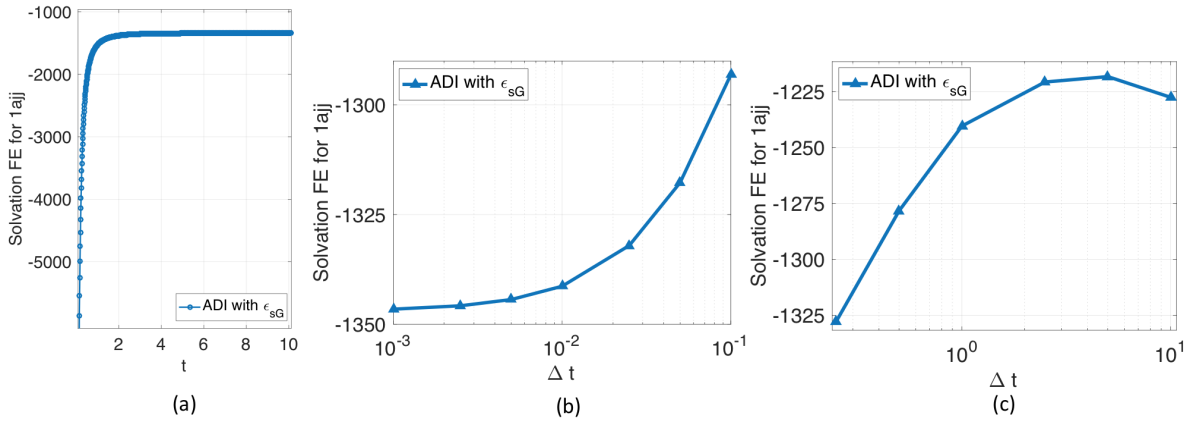


Figure 5.3: Pseudo-time ADI algorithm for the protein 1ajj. (a) Steady state convergence; (b) Temporal accuracy; (c) Stability.

for the protein 1ajj reaches the steady state around $T = 8$. Next, for the temporal accuracy in protein 1ajj case, we consider $T = 100$ and different time steps in Fig. 5.3(b). As Δt decreases the solvation free energy approaches -1341.34 kcal/mol. At last, we experiment the stability of the pseudo-time dependent ADI scheme with ϵ_{sG} for the protein 1ajj. For this purpose, we take $\Delta t \in \{0.25, 0.5, 1, 2.5, 5, 10\}$ and $T = 10^4 \Delta t$ to validate stability in Fig. 5.3(c). The result shows that the super-Gaussian ADI scheme is unconditionally stable for the protein 1ajj case.

We finally study a set of 23 proteins with the size (number of atoms) ranging from 519 to 2809. These proteins do not contain any cavity either. Therefore, we fix $\epsilon_{gap} = 2$. Regarding the (m, σ) pair, we keep $(3, 1.3)$ in ϵ_{sG}. The pseudo-time dependent ADI experiment is conducted with $\Delta t = 0.01$ and $T = 100$. The free energies calculated by the super-Gaussian model are listed in in Table 5.1. For reference, we also show two literature results, i.e., the pseudo-time coupled nonlinear solvation (CNS) model [43] with $\Delta t = \frac{h^2}{18}$ and h=0.5Å, and the two-component regularized PB (RPB) model [61] with $h = 0.25$Å. In the CNS model [43], the solvation free energy including both polar and apolar parts is reported, which in the RPB model [61], electrostatic free energy of the two-dielectric PB equation is reported. Thus, these

Table 5.1: Solvation Free Energy for Proteins in kcal/mol

No. of atoms	PDB ID	Ref. [43]	Ref. [61]	Present
519	1ajj	-1260.6	-1139.48	-1341.32
573	2erl	-919.8	-952.36	-1013.59
576	1bbl	-977.2	-988.40	-1186.34
596	1vii	-893.6	-902.31	-1031.52
648	1cbn	-255.5	-303.33	-398.27
667	2pde	-881.6	-820.97	–992.62
702	1sh1	-819.2	-753.99	-962.02
729	1fca	-1221.8	-1204.44	-1337.86
795	1ptq	-869.6	-873.32	-1057.50
809	1uxc	-1151.7	-1139.25	-1363.33
824	1fxd	-3347.0	-3321.39	-3073.39
832	1bor	-928.8	-853.47	-1120.57
858	1hpt	-790.4	-811.56	-1019.58
898	1bpi	-1283.4	-1304.37	-1450.90
903	1mbg	-1328.7	-1353.31	-1501.26
997	1r69	-1048.2	-1088.62	-1225.83
1187	1neq	-1710.3	-1731.71	-1991.43
1216	451c	-978.5	-1025.66	-1219.22
1272	1a2s	-1842.5	-1921.20	-1951.26
1435	1svr	-1750.6	-1711.11	-2039.08
1478	1frd	-2881.3	-2862.50	-2867.16
2065	1a63	-2423.9	-2374.41	-2881.10
2809	1a7m	-2141.3	-2160.34	-2527.79

energy results are necessarily close to the present ones. For example, for larger protein size, if the number of atoms exceed 2000 then the absolute energy difference between the super-Gaussian and RPB exceeds 350kcal/mol. Nevertheless, as can be observed from Fig. 5.4, the energies of three models are quite consistent with each other.

Along with the solvation free energy calculation, we record the CPU time from each protein simulation against the protein size [Fig.5.5]. Here also, we recorded the CPU time consumed for $\Delta t = 0.01$, $T = 100$. We note that the electrostatic potential $u(\vec{r})$ from the NPB equation (3.1) and $u_0(\vec{r})$ from the Poisson equation (3.7) both are solved by ADI method with super-Gaussian dielectric distribution.

Figure 5.4: Comparing the super-Gaussian results with coupled nonlinear solvation (CNS) and regularized PB (RPB) models.

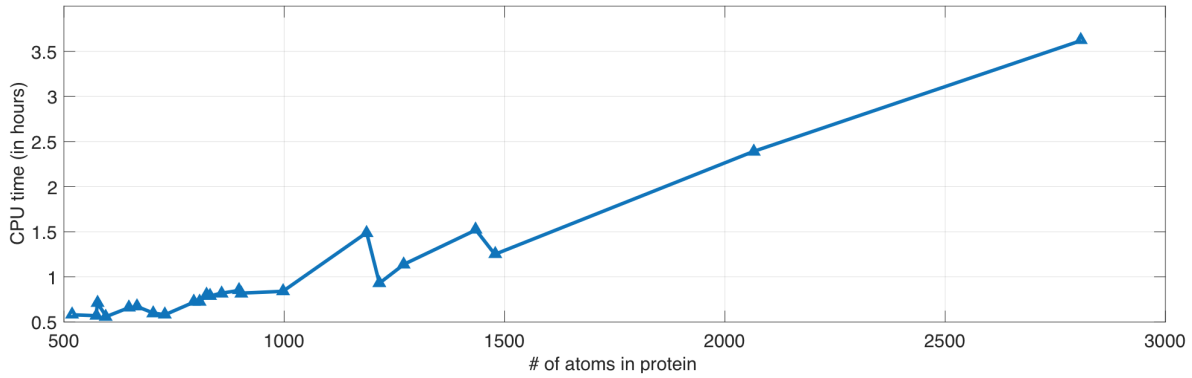


Figure 5.5: CPU time

5.3 Protein with cavities

In this section, we investigate a protein with interior cavities and discuss how ϵ_{gap} should be adjusted to compensate the cavity impact on the electrostatic free energy. It is known in the literature that the cavities in protein could be filled with water molecules. Experimentally, it is very challenging to identify the water molecules inside the protein cavities with the crystallographic analysis. Computationally, these cavity

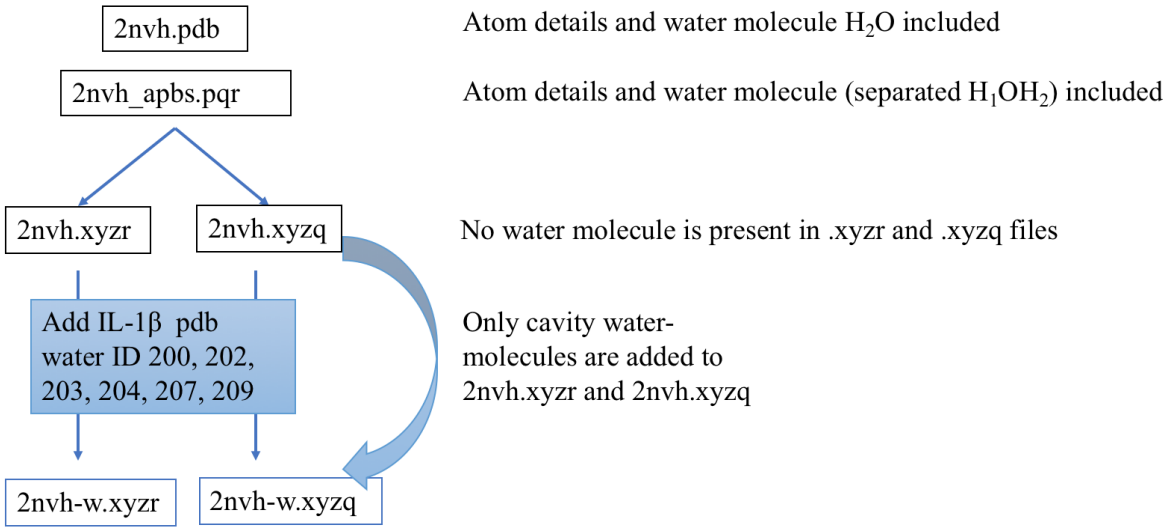


Figure 5.6: Inserting 6 water molecules in the protein IL-1β (pdb ID 2nvh)

water molecules play very important roles in solvation analysis. It is thus of great interest to numerically test the impact of cavity water molecules on electrostatic free energy in the present super-Gaussian PB model.

In our numerical experiment, we focus on a protein IL-1β (pdb ID 2nvh), whose cavity structure has been well studied in the literature. It has been confirmed by using the electron density experiments that water molecules are present in several cavities of IL-1β [63]. In particular, there are a few cavities with volumes in the range of 16Å^3 to 45Å^3 containing a total of 6 water molecules [63]. Moreover, there is a cavity with volume 39Å^3, for which electron density could not determine if water molecules exist in this cavity or not.

To study cavities with and without water molecules, we will process the protein IL-1β as illustrated in Fig. 5.6. We first note that in the protein preparation procedure discussed above, all water molecules will be removed in the final files, i.e., in .xyzr and .xyzq files, while all water molecules are included in the .pqr file produced by the PDB2PQR web server `http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/`. Furthermore, the atom IDs of six cavity water molecules are reported in [64]. This enables us to

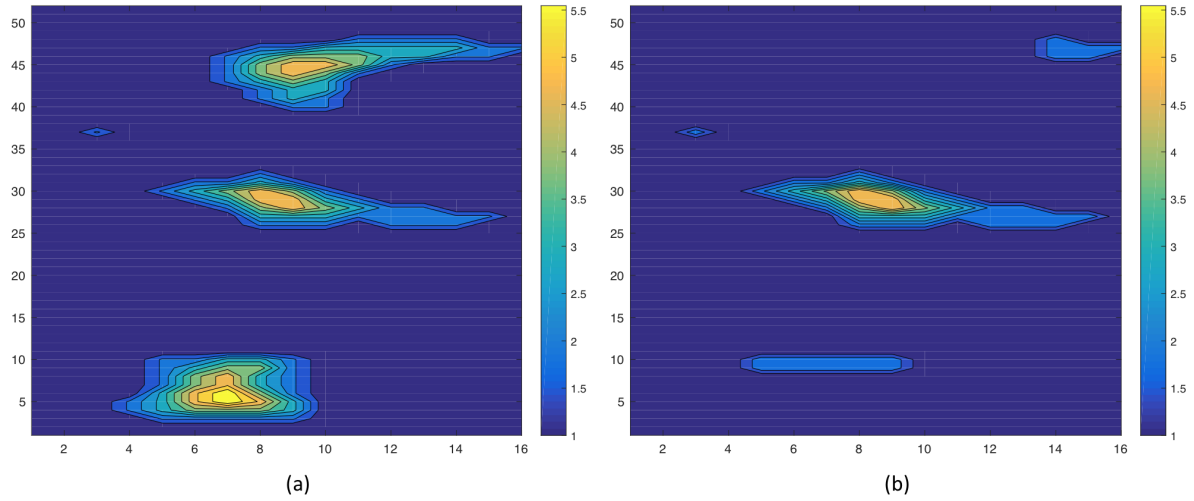


Figure 5.7: The super-Gaussian dielectric distribution with $\epsilon_{gap} = 6$. (a) 2nvh (b)2nvh-w.

identify these six molecules in the .pqr file and insert the corresponding hydrogen and oxygen atoms (located in the cavities) into 2nvh.xyzr and 2nvh.xyzq. These modified files will be called 2nvh-w.xyzr and 2nvh-w.xyzq. Computationally, we have generated two sets of workable files: one without water in cavities (2nvh) and another with 6 water molecules in some cavities (2nvh-w).

After adding water molecules, we note that one cavity with volume 39Å^3 is still empty. To see this, we compare the super-Gaussian dielectric function ϵ_{sG} of 2nvh and 2nvh-w in Fig. 5.7. Here we take $\epsilon_{gap} = 6$. By choosing a zoomed x-y cross section, we are able to capture three cavities of 2nvh in one contour plot (left figure). After adding water molecules, two of three cavities are filled so that ϵ values are reduced in these two locations (right figure). The cavity in the center remains unchanged in both 2nvh and 2nvh-w cases, which is the only one visible for 2nvh-w.

We then study the energy difference between two structures 2nvh and 2nvh-w based on the super-Gaussian PB model. A methodical mutation analysis [62] indicates that inserting one water molecule into cavities generally produces 1-2 kcal/mol energy gain. This helps us to quantitatively examine our inhomogeneous dielectric model with cavity modeling. By using the same parameter pair $(m, \sigma) = (3, 1.3)$, we first take

Table 5.2: Energy gain in 2nvh-w in kcal/mol

ϵ_{gap}	FE for 2nvh	FE for 2nvh-w	Energy gain in 2nvh-w
2	-2531.74	-2527.84	3.90
3	-2392.29	-2389.64	2.65
4	-2276.94	-2275.28	1.66
5	-2178.40	-2177.54	0.85
6	-2092.38	-2092.20	0.18
7	-2016.12	-2016.50	-0.39

$\epsilon_{gap} = 2$. The energy gain of 2nvh-w over 2nvh is around 4 kcal/mol for inserting six water molecules, which agrees with the theoretical estimate very well.

We have further studied the energy difference for $\epsilon_{gap} = 2, 3, \ldots, 7$. Table 5.2 shows the energy gains of 2nvh and 2nvh-w in kcal/mol with $\Delta t = 0.1$ and end time $T = 50$. The idea behind this study is that we can compensate the absence of water molecules in cavities by raising the dielectric value ϵ_{gap} of the cavity fluid in the computational modeling. Consequently, one can represent the water molecules without physically adding them by just using a larger ϵ_{gap} value in the dielectric model. Indeed, as can be shown in Table 5.2, the energy gain becomes smaller and smaller as ϵ_{gap} is increased. At round $\epsilon_{gap} = 6$, the difference between the solvation free energies of 2nvh and 2nvh-w is almost zero, i.e., around 0.1 kcal/mol. Our recommendation is that for proteins with cavities but one does not know if there are water molecules inside or not (such as the one shown in Fig. 5.7), one can model water molecules computationally by setting $\epsilon_{gap} = 6$ or higher. We also believe that the magic number $\epsilon_{gap} = 6$ for this example relates to the cavity size or volume. This parameter set up works well if the volume of the cavities approximately $\leq 40\text{Å}^3$. If we have large volume cavities, we may need to increase the value of ϵ_{gap}.

CHAPTER 6

CONCLUSION

We started with a brief introduction to the PDE model for electrostatics of bio-molecules which uses the PB equation. The consideration of the two dielectric model in the literature, different types of existing molecular surfaces, solute-solvent interactions and electrostatic interactions were described in chapter 1. Also, we discussed the choice of minimal molecular surface (MMS) to generate the molecular interface. Mainly we pay attention to improve the non-homogeneous Gaussian dielectric model, and we introduce the new super-Gaussian dielectric model. Chapter 2 explained how the super-Gaussian dielectric model converging of to the discrete two dielectric model with higher order, the preservation of maximum dielectric value inside the protein cavities and maintaining smoothness in water and vacuum both the states. The super-Gaussian dielectric model depends on the hypersurface function (generated by MMS) across the molecular boundary and some parameters such as relative variance, the order of super-Gaussian density function and the maximum dielectric constant of the cavity fluid inside the solute. Effective dielectric constant analysis determines the choice of those parameters according to a solute-solvent system.

In equation NPB solving, the analytical integration of the nonlinear term in a first order time splitting framework avoids the difficulties in the direct treatment of the strong nonlinearity. Then an implicit Euler based ADI scheme is used to solve the linear heat equation. The super-Gaussian dielectric distribution is employed in central finite difference space discretization. This ADI scheme with the super-Gaussian dielectric model becomes unconditionally stable. We have verified our new model with a single atom system and 23 proteins with no cavity inside. Finally, we validate the

ADI method with super-Gaussian dielectric profile successfully on a real protein with several water molecules at the interior cavities.

Chapter 5 opens new avenues to explore the super-Gaussian model on more proteins with internal cavities. We already have some proteins containing the interior cavity size (64-108)Å^3. With these experiments, we hope to provide a better range of cavity fluid dielectric constant. Also, it is reasonable to establish a relation between the volume of the interior cavities and maximal dielectric constant for the cavity water molecules.

REFERENCES

[1] L. Li, C. Li, Z. Zhang and E. Alexov, On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi, *J. Chem. Theory Comput.*, **9**, 2126-2136, (2013).

[2] Che, J., J. Dzubiella, B. Li, and J. A. McCammon, Electrostatic free energy and its variations in implicit solvent models, *The Journal of Physical Chemistry B*, **112**, 3058-3069 (2008).

[3] Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences*, **98**, 10037-10041, (2001).

[4] Li, C., L. Li, J. Zhang, and E. Alexov. 2012. Highly efficient and exact method for parallelization of grid-based algorithms and its implementation in DelPhi. *Journal of Computational Chemistry*, **33**: 1960-1966, (2012).

[5] Pang, X., and H. X. Zhou. Poisson-Boltzmann Calculations: van der Waals or Molecular Surface? *Commun Comput Phys.*, **13**, 1–12, (2013).

[6] B. Lee and FM Richards, Interpretation of protein structure: estimation of static accessibility, *J. Molecular Bio.*, **55**, 379-400, (1973).

[7] FM Richards, Areas, volumes, packing and protein structure, *Annual Review of Biophysics and Bioengineering*, **6**, 151-176, (1977).

[8] ML Connolly, Analytical molecular surface calculation, *Journal of Applied Crystallography*, **16**, 548-558 (1983).

[9] Blinn JF. A generalization of algegraic surface drawing. *ACM Transactions on Graphics*, **1**, 235–256, (1982).

[10] Duncan BS, Olson AJ. Shape analysis of molecular surfaces. Biopolymers, **33**, 231–238, (1993).

[11] Grant, J. A., and B. Pickup. A Gaussian description of molecular shape. *Journal of Physical Chemistry*, **99**, 3503-3510, (1995).

[12] Chen M, Lu B. TMSmesh: a robust method for molecular surface mesh generation using a trace technique. *Journal of Chemical Theory and Computation*, **7**, 203–212, (2011)

[13] Giard J, Macq B. Molecular surface mesh generation by filtering electron density map, *International Journal of Biomedical Imaging*, **2010**, 923780 (9 pages), (2010)

[14] Zhang Y, Xu G, Bajaj C. Quality meshing of implicit solvation models of biomolecular structures, *Computer Aided Geometric Design*, **23**, 510–530, (2006)

[15] Yu Z, Holst MJ, Cheng Y, McCammon JA. Feature-preserving adaptive mesh generation for molecular shape modeling and simulation, *Journal of Molecular Graphics and Modelling*, **26**, 1370–1380, (2008)

[16] Z.H. Qiao, Z.L. Li, T. Tang, A finite difference scheme for solving the nonlinear Poisson-Boltzmann equation modeling charged spheres. *J. Comput. Math.* **24**, 252–264, (2006)

[17] Chen, D. A., Z. Chen, C. J. Chen, W. H. Geng, and G. W. Wei. Software News and Update MIBPB: A Software Package for Electrostatic Analysis. *Journal of computational chemistry*, **32**, 756-770 (2011).

[18] Cheng L-T, Dzubiella J, McCammon JA, Li B, Application of the level-set method to the solvation of nonpolar molecules. *J Chem Phys*, **127**, 084503, (2007).

[19] Zhou, L., and S. A. Siegelbaum. Effects of surface water on protein dynamics studied by a novel coarse-grained normal mode approach. *Biophys J*, **94**, 3461-3474, (2008).

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

[21] Zhao, Y., Kwan, Y.-Y., Che, J., Li, B., McCammon, J. A., Phase-field approach to implicit solvation of biomolecules with Coulomb-field approximation. The Journal of Chemical Physics, 139(2), 024111. http://doi.org/10.1063/1.4812839, (2013).

[22] S. Dai, B. Li, J. Liu; Convergence of phase-field free energy and boundary force for molecular solvation, Archive for Rational Mechanics and Analysis,**227**, 105-147,(2018).

[23] M. Hammel. Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur. Biophys. J.*, **41**, 789-799, (2012).

[24] M. Kokkinidis, N.M. Glykos, and V.E. Fadouloqlou. Protein flexibility and enzymatic catalysis. *Adv. Protein Chem. Struct. Biol.*, **87**, 181-218, (2012).

[25] Nymeyer, H., and H. X. Zhou. A method to determine dielectric constants in nonhomogeneous systems: application to biological membranes. *Biophysical journal*, **94**, 1185-1193, (2008).

[26] Song, X. An inhomogeneous model of protein dielectric properties: Intrinsic polarizabilities of amino acids. *The Journal of chemical physics*, **116**, 9359, (2002).

[27] Voges, D., and A. Karshikoff. A model of a local dielectric constant in proteins. *The Journal of chemical physics*, **108**, 2219, (1998).

[28] Alexov, E. G., and M. R. Gunner. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophysical journal*, **72**, 2075-2093, (1997).

[29] Alexov, E. G., and M. R. Gunner. Calculated protein and proton motions coupled to electron transfer: electron transfer from QA- to QB in bacterial photosynthetic reaction centers. *Biochemistry*, **38**, 8253-8270, (1999).

[30] Simonson, T., and D. Perahia. Internal and interfacial dielectric properties of cytochrome c from molecular dynamics in aqueous solution. *Proceedings of the National Academy of Sciences*, **92**, 1082-1086, (1995).

[31] L. Li, C. Li, and E. Alexov, On the modeling of polar component of solvation energy using smooth Gaussian-based dielectric function, *J. Theory Comput. Chem.*, **13**, 1440002, (2014).

[32] Alexov, E., Miksovska, J., Baciou, L., Schiffer, M., Hanson, D. K., Sebban, P., and Gunner, M. R., Modeling the effects of mutations on the free energy of the first electron transfer from QA-to QB in photosynthetic reaction centers, *Biochemistry*, **39:20**, 5940–5952, (2000).

[33] Alexov, E. Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: comparing experimental with computed pK(a)s. *Proteins-Structure Function and Bioinformatics*, **50**, 94-103, (2003).

[34] Alexov, E., E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. E. Nielsen, D. Farrell, T. Carstensen, M. H. Olsson, J. K. Shen, J. Warwicker, S. Williams, and J. M. Word. Progress in the prediction of pKa values in proteins. *Proteins*, **79**, 3260-3275, (2011).

[35] Li, C., M. Petukh, L. Li, and E. Alexov. Continuous development of schemes for parallel computing of the electrostatics in biological systems: implementation in DelPhi. *Journal of Computational Chemistry*, **34**, 1949-1960, (2013).

[36] Li, C., L. Li, M. Petukh, and E. Alexov. Progress in developing Poisson-Boltzmann equation solvers. Molecular Based Mathematical Biology 1. *Molecular Based Mathematical Biology*, **1**, 42-62, (2013).

[37] Lin Wang, Lin Li, and Emil Alexov. "pKa predictions for proteins, RNAs and DNAs with the Gaussian dielectric function using DelPhiPKa." *Proteins*, doi: 10.1002/prot.24935 (2015)

[38] Lin Wang, Min Zhang, and Emil Alexov. "DelPhiPKa Web Server: Predicting pKa of proteins, RNAs and DNAs." *Bioinformatics*, doi: 10.1093/bioinformatics/btv607, (2015)

[39] Lin Li, Chuan Li, Emil Alexov; "On the Modeling of Polar Component of Solvation Energy using Smooth Gaussian-based Dielectric Function"; *Journal of Theoretical and Computational Chemistry*, **13**, 1440002, (2014).

[40] Jia Z, Li L, Chakravorty A, Alexov E. "Treating ion distribution with Gaussian-based smooth dielectric function in DelPhi." *J Comput Chem.*, **38**, 1974-1979, (2017)

[41] Ng, Jin, Vora, Taira, Krishnamurthy, Vikram, Chung, Shin-Ho, Estimating the dielectric constant of the channel protein and pore. *European biophysics journal*, ,**37**, 213-222, (2008).

[42] Zhao, S., Pseudo-time-coupled nonlinear models for biomolecular surface representation and solvation analysis, *Int. J. Numer. Method. Bio. Engrg.*, **27**, 1964-1981, (2011).

[43] Zhao, S., Operator splitting ADI schemes for pseudo-time coupled nonlinear solvation simulations, *J. Comput. Phys.*, **257**, 1000-1021, (2014).

[44] Tian, W., Zhao, S., A fast ADI algorithm for geometric flow equations in biomolecular surface generation, *Int. J. Numer. Meth. Biomed. Engrg.*, **30**, 490-516, (2014).

[45] L. Wilson and S. Zhao, Unconditionally stable time splitting methods for the electrostatic analysis of solvated biomolecules, International Journal of Numerical Analysis and Modeling, **13**, 852-878, (2016).

[46] A. Sayyed-Ahmad, K. T., and Ortoleva, P. Efficient solution technique for solving the poisson-boltzmann equation. *Journal of Computational Chemistry*, **25**, 1068–1074, (2004)

[47] A.I. Shestakov, J. M., and Noy, A. Solution of the nonlinear poisson-boltzmann equation using pseudo-transient continuation and the finite element method. *J. of Colloid Interface Sci.*, **247**, 62–79, (2002)

[48] J. Douglas, Jr. and D. Peaceman, Numerical solution of two-dimensional heat flow problems, *Amer. Inst. Chem. Engrg. J.*, **1**, 505-512, (1955)

[49] J. Douglas, Jr., On the numerical integration of $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial t}$ by implicit methods, *J. Soc. Indust. Appl. Math.*, **3**, 42-65, (1955).

[50] B.Z. Lu, Y.C. Zhou, M.J. Holst, and J.A. McCammon, Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications, *Commun. Comput. Phys.*, **3**, 973-1009, (2008).

[51] Grant, J. A., B. T. Pickup, and A. Nicholls. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J Comput Chem*, **22**, 608-640, (2001).

[52] Im, W., D. Beglov, and B. Roux. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications*, **111**, 59-75, (1998).

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

[54] Geng, W. H., Zhao, S., Fully implicit ADI schemes for solving the nonlinear Poisson-Boltzmann equation, *Molecular Based Mathematical Biology*, **1**, 109-123, (2013).

[55] Warshel, A.; Russell, S. T. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys*, **17**, 283-422, (1984).

[56] Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochim Biophys Acta*, **1764**, 1647-76, (2006).

[57] A. Chakravorty, Z. Jia, L. Li, S. Zhao, and E. Alexov, Reproducing the Ensemble Average Polar Solvation Energy of a Protein from a Single Structure: Gaussian-Based Smooth Dielectric Function for Macromolecular Modeling, *Journal of Chemical Theory and Computation*, accepted, (2018).

[58] M. Sanner, A. Olson, and J. Spehner, Reduced surface: an efficient way to compute molecular surfaces, *Biopolymers*, **38**, 305-320, (1996).

[59] Surface Calculation Failures and Workarounds, UCSF Computer Graphics Laboratory / March 2014, Retrieved from https://www.rbvi.ucsf.edu/chimera/docs/UsersGuide/surfprobs.html

[60] Comparison Study of Polar and Nonpolar Contributions to Solvation Free Energy, *Journal of Chemical Information and Modeling*, **57**, 2539-2553, (2017).

[61] W. Geng and S. Zhao, A two-component Matched Interface and Boundary (MIB) regularization for charge singularity in implicit solvation, *Journal of Computational Physics*, **351**, 25-39, (2017).

[62] Takano,K.,Y.Yamagata, and K.Yutani, Buried water molecules contribute to the conformational stability of a protein. *Protein Eng.*, **16**, 5-9, (2003).

[63] Determination of solvent content in cavities in IL-1β using experimentally phased electron density, Michael L. Quillin, Paul T. Wingfield, Brian W. Matthews, *Proceedings of the National Academy of Sciences*, **103**, 19749-19753, (2006).

[64] David J. Huggins,Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations, *Biophysical Journal*, **108**, 928–936, (2015).

[65] W. Deng, J. Xu, and S. Zhao, On developing stable finite element methods for pseudo-time simulation of biomolecular electrostatics, *Journal of Computational and Applied Mathematics*, **330**, 456-474, (2018).