

DIFFERENTIAL GEOMETRY BASED MULTISCALE MODELING OF SOLVATION

By

Zhan Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Applied Mathematics

2011

ABSTRACT

DIFFERENTIAL GEOMETRY BASED MULTISCALE MODELING OF SOLVATION

By

Zhan Chen

Solvation is an elementary process in nature and is of paramount importance to many sophisticated chemical, biological and biomolecular processes. The understanding of solvation is an essential prerequisite for the quantitative description and analysis of biomolecular systems. Implicit solvent models, particularly those based on the Poisson-Boltzmann (PB) equation for electrostatic analysis, are established approaches for solvation analysis. However, *ad hoc* solvent-solute interfaces are commonly used in the implicit solvent theory and have some severe limitations.

We have introduced differential geometry based solvation models which allow the solvent-solute interface to be determined by the variation of a total free energy functional. Our models extend the scaled particle theory (SPT) of nonpolar solvation models with a solvent-solute interaction potential. The nonpolar solvation model is completed with a PB theory based polar solvation model. In our Eulerian formulation, the differential geometry theory of hypersurface is utilized to define and construct smooth interfaces with good stability and differentiability, for use in characterizing the solvent-solute boundaries and in generating continuous dielectric functions across the computational domain. Some techniques from the geometric measure theory are employed to rigorously convert a Lagrangian formulation of the surface energy into an Eulerian formulation, so as to bring all energy terms on an equal footing. In our Lagrangian formulation, the differential geometry theory of surfaces is used to provide a natural description of solvent-solute interfaces. By optimizing the total free energy

functional, we derive a coupling of the generalized Poisson-Boltzmann equation (GPBE) and the generalized geometric flow equation (GGFE or also called Laplace-Beltrami equation) for the electrostatic potential and the construction of realistic solvent-solute boundaries, respectively. The coupled partial differential equations (PDEs) are solved with iterative procedures to reach a steady state, which delivers the desired solvent-solute interface and electrostatic potential for many problems of interest. These quantities are utilized to evaluate the solvation free energies, protein-protein binding affinities, etc.

The above proposed approaches have been extensively validated. Extensive numerical experiments have been designed to validate the present theoretical models, to test the computational methods, and to optimize the numerical algorithms. Solvation analysis of both small compounds and proteins are carried out to further demonstrate the accuracy, stability, efficiency and robustness of the present new models and numerical approaches. Comparison is given to both experimental and theoretical results in the literature.

Moreover, to account for the charge rearrangement during the solvation process, we also propose a differential geometry based multiscale solvation model which makes use of electron densities computed directly from a quantum mechanical approach. We construct a new total energy functional, which consists of not only polar and nonpolar solvation contributions, but also the electronic kinetic and potential energies. We show that the quantum formulation of our solvation model improves the prediction of our earlier models, and outperforms some explicit solvation analysis.

ACKNOWLEDGMENT

First of all, I would like to express my deepest and sincerest gratitude to my dissertation advisor, Dr. Guowei Wei for his invaluable supports, guidance and personal care. It is beyond my words of appreciation that he is a far-sighted person with a broad interdisciplinary perspective, and that he has a unique creative insight which often sheds light on undiscovered potentialities of a subject. More than being a polymath, he cares about his graduate students so much that I have not seen any other advisor like him. Throughout these years, he has been cultivating my passion in science and academic aspirations. He has enormous patience on me, even during the most challenging times in my Ph.D. study. His spirit of pursuing academic dreams and perfection has a life-long influence on me. It has been an enormously enjoyable experience working with him.

My gratitude also goes to my defense committee members, Dr. Changyi Wang, Dr. Leslie Kuhn, Dr. Chichia Chiu, and Dr. Moxun Tang for their expertise and precious time. I also would like to thank Dr. Zhengfang Zhou for his generous supports and unceasing cares for my wife and me throughout our Ph.D. program, and Dr. Keith Promislow for his kindness and the inspiring classroom discussions, and Dr. Nathan Baker for his supports and intense interests in our work of solvation.

Moreover, I would like to thank Dr. Shan Zhao for his continuing helps and encouragement in my research. Dr. Zhao is the one who taught me how to write FORTRAN codes after I joined this group. My thanks also go to the rest of current and former members in Dr. Wei's group: Dr. Yongcheng Zhou, Dr. Yuhui Sun, Dr. Sining Yu, Dr. Weihua, Geng, Dr. Duan Chen, Dr. Siyang Yang, Ms. Qiong Zheng, Mr. Langhua Hu, Ms. Jin Kyoung Park,

Ms. Yajie Yu, Ms. Weijuan Zhou, Mr. Wangheng Liu, Mr. Manfeng Hu, Ms. Shuailing Wang, Mr. Qi Zhao, Mr. Kelin Xia, Mr. Huibing Zhu, for their valuable discussions and helps in my research and daily life. Our group is like a big loving family. I will always treasure our friendships. I am also grateful to Ms. Barbara Miller, Graduate Secretary in the Department of Mathematics, for her generous helps during my graduate study, and to Mr. Nick Boros for proofreading a part of my dissertation.

Finally, from my very deep heart I thank Yuting, my wife and loved one, for her constant love, understanding and supports throughout my Ph.D. study. She is always patient to me and always has trusts in me. To her, words can not express my gratitude.

To my family: Baokun, Ziyang, Yan, Xiang and Yuting

TABLE OF CONTENTS

List of Tables	x
List of Figures	xii
1 Introduction	1
1.1 Introduction to solvation models	1
1.1.1 Biological background	1
1.1.2 Polar solvation models	3
1.1.3 Poisson-Boltzmann theory	6
1.1.4 Nonpolar solvation models	8
1.2 Molecular interface definitions	10
1.3 Quantum mechanical continuum models	12
1.4 Limitations of current models	15
1.5 Mathematical issues and numerical challenges	18
1.5.1 Geometry, PDE and interface	19
1.5.2 Geometric flow equation	20
1.5.3 Highly accurate and efficient solver for interface problems	21
1.5.4 Self-consistent iterative methods	24
1.6 The rest of this thesis	25
2 Eulerian formulation	27
2.1 Theory and model	29
2.1.1 Solute-solvent boundary	30
2.1.2 Total free energy functional	31
2.1.2.1 Polar free energy functional	31
2.1.2.2 Non-polar free energy functional	32
2.1.3 Governing equations	35
2.2 Methods and algorithms	38
2.2.1 Discretization schemes of the governing equations	38
2.2.1.1 The generalized Poisson-Boltzmann equation	39
2.2.2 Acceleration procedures	40
2.2.2.1 Precondition of the PB solver	40
2.2.2.2 Initial guess of the generalized PB solution	41
2.2.2.3 Convergence criteria in the generalized PB solver	42
2.2.3 Dynamical coupling of the generalized Poisson Boltzmann and geometry flow equations	42
2.2.3.1 Approach I	43

2.2.3.2	Approach II	44
2.2.4	Evaluation of the solvation free energy	45
2.3	Numerical test and validation	47
2.3.1	The behavior of the coarea formula	47
2.3.2	Accuracy of the generalized PB solver	49
2.3.3	Convergence of boundary profile and dielectric function	51
2.3.4	Consistency of iteration procedures	55
2.3.5	Efficiency of the accelerated iteration procedure	58
2.3.6	Impact of potentials in the geometric flow equation	61
2.4	Applications	63
2.4.1	Set of 17 test molecules	63
2.4.2	Solvation free energy of proteins	66
2.4.3	Twenty two proteins	70
2.5	Chapter conclusions	73
3	Quantum formulation	75
3.1	Theory and model	77
3.1.1	Charge density based polar free energy functional	77
3.1.2	Quantum mechanical energy functional	79
3.1.2.1	Kinetic energy	81
3.1.2.2	Potential energy	82
3.1.3	Total free energy functional	83
3.1.4	Governing equations	84
3.2	Numerical methods and algorithms	88
3.2.1	Solution of the generalized Poisson-Boltzmann equation	88
3.2.2	Solution of the generalized Kohn-Sham equation	92
3.2.3	Evaluation of the solvation free energy	95
3.2.4	Dynamical coupling of involved PDE equations	98
3.3	Numerical test and validation	101
3.3.1	Validation of the cancellation of self-interaction energy	102
3.3.2	Validation of data translation and unit conversion	104
3.3.3	Accuracy of solvation free energies computed by the present model	106
3.4	Applications	107
3.4.1	Solvation free energies of 24 small molecules	108
3.4.2	Solvation free energies of 16 molecules	110
3.4.3	Solvation free energies of 3 larger molecules	115
3.5	Chapter conclusions	119
4	Lagrangian formulation	121
4.1	Theory and model	123
4.1.1	Solvation free energy functionals	123
4.1.1.1	Polar solvation functional	124
4.1.1.2	Total free energy functional of solvation	125
4.1.2	Surface variation	125

4.1.3	Governing equations	132
4.2	Methods and algorithms	134
4.2.1	Interconversion between the Lagrangian and Eulerian representations	134
4.2.1.1	Embedding the Lagrangian dynamics into the Eulerian representation	135
4.2.1.2	Transform from the Lagrangian representation to the Eulerian representation	137
4.2.1.3	Transform from the Eulerian representation to the Lagrangian representation	138
4.2.1.4	Numerical surface integral and volume integral in the Eulerian representation	140
4.2.2	Dynamic coupling of the Poisson-Boltzmann and geometric flow equations	142
4.3	Validation	143
4.3.1	Validation of the interface extraction scheme	144
4.3.2	Effect of interaction potentials	146
4.3.2.1	Surfaces of a diatom system	147
4.3.2.2	Surfaces of a four-atom system	148
4.3.2.3	Surfaces and electrostatic potentials of a protein	151
4.3.3	Isosurface function value and minimal surfaces	152
4.3.4	Convergence of surface area, volume and energy	154
4.4	Application	156
4.4.1	Free energy calculations	157
4.4.1.1	Solvation energies of 17 compounds	157
4.4.1.2	A set of 23 proteins	158
4.4.2	Salt effect on protein-protein binding energies	162
4.5	Chapter conclusions	167
5	Thesis achievements and future work	170
5.1	Contributions	170
5.2	Future work	174
A	Solution of the generalized Laplace-Beltrami equation and the ADI scheme	178
B	PB equation in different forms	185
C	Differential geometry theory preliminary	188
	Bibliography	192

LIST OF TABLES

2.1	Areas computed from the coarea formula for bounded open sets	48
2.2	Errors and convergence orders for the generalized PB solver ($\epsilon_1 = 1$)	51
2.3	Comparison between two iteration approaches	56
2.4	Effect of relaxation factor α on final results	57
2.5	Effect of the number of intermittency in Approach II	58
2.6	CPU time analysis from original schemes	59
2.7	Speedup from adjustment of initial guess and preconditioner in PB solver	59
2.8	Influence of convergence criteria on electrostatic solvation free energy and computational time for the diatomic system	60
2.9	Comparison of CPU time (s) in the iteration procedures with and without accelerations	61
2.10	Effects of potentials on the solvent-solute boundary	63
2.11	Comparison of free energies (kcal/mol) for 17 compounds	64
2.12	Comparison of electrostatic solvation free energies (kcal/mol) obtained from the MFCC-CPCM, the present model (OSM) and MIBPB.	68
2.13	Electrostatic solvation free energies for 22 proteins	71
3.1	Comparison of total electrostatic energy (kcal/mol) and electrostatic solvation energy (kcal/mol) obtained with the partial charge approach.	104
3.2	Comparison of total electrostatic energy (kcal/mol) and electrostatic solvation energy (kcal/mol) obtained with the direct use of charge density.	105

3.3	Comparison of solvation energy components between present results and those of Wang <i>et al</i> [238] for three small molecules.	106
3.4	Solvation free energy (kcal/mol) and its decomposition.	107
3.5	Comparison of solvation free energies (kcal/mol) obtained from the present model and experimental data for 24 small molecules.	108
3.6	Solvation free energy (kcal/mol) decomposition for a set of 21 molecules.	112
3.7	Comparison of free energies (kcal/mol) for 16 compounds.	114
3.8	Free energies (kcal/mol) for 16 compounds using structures from Pubchem data.	117
3.9	Solvation free energies (kcal/mol) of 3 large molecules and corresponding CPU time (hour).	118
4.1	Comparison of surface areas (\AA^2), volumes (\AA^3) and energies (kcal/mol) for two small systems.	145
4.2	Electrostatic solvation free energies (kcal/mol), surface areas (\AA^2) and volumes (\AA^3) of protein 451c with different solvent-solute interactions.	151
4.3	Surface areas (\AA^2) for different surface definitions	153
4.4	RMS error with different nonpolar potentials.	158
4.5	Predicted and experimental total solvation free energies for 17 small compounds.	160
4.6	Comparison of electrostatic solvation free energies of 23 proteins.	162
4.7	Comparison of binding affinities of two proteins complexes from current simulations and those from Bertonati et al's paper.	166

LIST OF FIGURES

1.1	A solvation free energy cycle adapted from Levy et al. [130]. The total solvation energy (1) is decomposed into several steps: “charging” the solute in solvent (6) and vacuum (2), including attractive dispersive solute-solvent interactions in solvent (5) and vacuum (3), and cavity formation associated with repulsive solute-solvent interactions (4). The energy associated with Step (7) is generally termed a “nonpolar solvation energy” while the difference in energies associated with Steps (1) and (7) is generally considered as “polar solvation energy”. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.	4
2.1	The cross line of S and $(1 - S)$ of a diatomic system described in Section 2.3.3	29
2.2	The cross line profile of $\epsilon(S)$ of a diatomic system described in Section 2.3.3. Here, we have set $\epsilon_s = 80$ and $\epsilon_m = 1$.	36
2.3	The evolutionary profiles of the S function at cross section $(x, y, 0.05)$ in a diatomic system plotted from four intermediate states.	53
2.4	The time evolution histories of the electrostatic solvation free energy, $F(\text{Volume})$ and $J(\text{Area})$, where $F(\text{Volume}) = \text{volume}/5-180$ and $J(\text{Area}) = (\text{surface area})/5-200$.	55
2.5	Correlation between experimental data and the present optimized surface model (OSM)(also results from Nicholls’) in electrostatic solvation free energies of 17 compounds.	67
2.6	Correlation between MFCC-CPCM [152] and the present optimized surface model (OSM) in electrostatic solvation free energies of 8 proteins.	69
2.7	Correlation between MIBPB-III and the present model (OSM) in electrostatic solvation free energies of 22 proteins	70
2.8	Differences between electrostatic solvation free energies obtained from the MIBPB and the present model with original radii (Radii0) or enlarged radii (Radii1).	72

2.9	Surface potential display of one protein (PDBID: 1frd) at different isosurfaces	72
3.1	Flowchart of the numerical solution of the coupled PDEs.	99
3.2	Correlation between experimental data and the present optimized surface model with quantum correction (OSMQ) in solvation free energies of 24 small molecules.	111
3.3	Illustration of surface electrostatic potentials of four small compounds at their corresponding isosurfaces $S = 0.50$. (a) Glycerol triacetate; (b) 1,1-diethoxyethane; (c) Bis-2-chloroethyl ether; (d) Dimethoxymethane.	111
3.4	Correlation between experimental data [160] and the present optimized surface model with quantum mechanics (OSMQ) in solvation free energies of 16 compounds.	116
3.5	Illustration of surface electrostatic potentials of three compounds at their corresponding isosurfaces $S = 0.50$. (a) Phorbol; (b) Phorbol12,13-dibutyrate; (c) Staurosporine.	118
4.1	Illustration of surface morphology of a diatom system with radii 1.87\AA and coordinates $(x, y, z) = (-2.2, 0, 0)$, and $(2.2, 0, 0)$ under different solute-solvent interactions. Top Left: The MMS (no potential); Top Right: The surface obtained under a repulsive potential ($V_i^{\text{rep,LJ}}$); Bottom Left: The surface obtained under a full L-J potential (V^{LJ}); Bottom Right: The surface obtained under a full L-J potential and an electrostatic potential. It can be seen that the repulsive potential produces a “fat” surface, while an attractive potential or an electrostatic potential leads to a “slim” surface.	147
4.2	Illustration of surface morphology of a four-atom system with radii 1.87\AA and coordinates $(x, y, z) = (-3.40, 0, 0)$, $(3.40, 0, 0)$, $(0, -2.94, 0)$ and $(0, 2.94, 0)$ under different solute-solvent interactions. Top Right: The surface obtained under a repulsive potential ($V_i^{\text{rep,LJ}}$); Bottom Left: The surface obtained under the full L-J potential (V^{LJ}); Bottom Right: The surface obtained under the full L-J potential and an electrostatic potential. Again, it can be seen that the repulsive potential produces a “fat” surface; while an attractive potential or an electrostatic potential leads to a “slim” surface.	149

4.3	The projection of electrostatic surface potentials of protein 451c onto different surfaces obtained under various solvent-solute interactions. Top Left: Attractive surface; Top Right: The MMS; Bottom Left: Repulsive surface; Bottom Right: Polar surface. It is noted that the repulsive surface is a “fat” surface comparing to the MMS; while an attractive surface or a polar surface is a “slim” surface.	150
4.4	Difference of surface areas (\AA^2) between MMS and various resulting surfaces generated under different constant pressure effects.	155
4.5	Decreasing of surface area ($\times 10^2 \text{\AA}^2$), $J(\text{volume})=(\text{volume}-2)$ ($\times 10^2 \text{\AA}^3$) and total solvation free energy (kcal/mol) in diethyl propanedioate as the number of iterations increases.	155
4.6	Correlation of solvation free energy between previous optimized smooth surface (OSS) model and the present optimized molecular surface (OMS) model in the set of 17 compounds listed in Table 4.5.	159
4.7	Correlation of electrostatic solvation free energy between the present optimized molecular surface (OMS) model, and previous models, such as the optimized smooth surface (OSS), the MIBPB-III and the MMS for 23 proteins listed in Table 4.6.	161
4.8	Difference of electrostatic solvation free energies between the OMS model and previous OSS and MMS models for 23 proteins listed in Table 4.6.	163
4.9	Protein-protein complexes. Left: Protein complex 1emv; Right: Protein complex 1beb.	164
4.10	The salt dependence of the binding affinities of two protein complexes. Left: Protein 1emv; Right: Protein 1beb. Here OMS data are computed by our optimized molecular surface (OMS) model. NLPB data are taken from Bertoni et al’s paper [26].	165

Chapter 1

Introduction

1.1 Introduction to solvation models

1.1.1 Biological background

Almost all important biological processes in nature, including signal transduction, DNA recognition, transcription, post-translational modification, translation, protein folding and protein ligand binding, naturally occur in water, which comprises 65-90% of cellular mass. The understanding of solvation is an elementary prerequisite for the quantitative description and analysis of the above-mentioned processes. Solvation involves the energetics of interactions between solute molecules and solvent molecules or ions in the aqueous environment. Solute-solvent interactions are typically described by solvation energies (or closely related quantities): the free energy of transferring the solute from a vacuum to the solvent environment of interest (e.g., water at a certain ionic strength), as shown in more detail in Figure 1.1. Solvation free energy is a physical quantity that can be measured experimentally.

Although millions of organic compounds are known now, only several thousands of com-

pounds have experimental data being reported for the solvation free energy. It is mainly due to experimental difficulties associated with the precise measurement, particularly for those compounds with low solubility and/or low volatility [182, 174]. Because of low solubility and/or low volatility, accurate and time-consuming measurement is required with highly sensitive instruments. Unfortunately, many important organic compounds belong to this category. Moreover, attentions need to be paid on the chemical stability of solute under investigation. Therefore, the experimental study of solvation free energy still remains expensive, laborious and is sometimes inaccurate.

Computational approaches provide an alternative method to obtain the solvation free energy. Solvation free energies can be calculated by a variety of computational methods, ranging from very time-consuming quantum mechanical approaches [111, 183, 148, 118] to simple phenomenological modifications of Coulomb's law. Solvation models can be roughly divided into two main classes [186, 239, 203, 200]: explicit solvent models that describe the solvent in molecular or atomic detail [179], and implicit solvent models that generally replace the explicit solvent with a dielectric continuum [8, 10, 65, 109, 186, 117]. Explicit solvent models provide the detailed information on molecular constitutions, and generally require extensive sampling to extract meaningful thermodynamic, statistical or kinetic properties of interest. Whereas, implicit solvent models focus on the biomolecules of interest, and take a mean field approximation for solvent properties. Because of their fewer degrees of freedom, implicit solvent methods have become popular for many applications in molecular simulation [7, 82, 9, 69].

To help the calculation of solvation energy, one can conceptually break up the solvation process as follows: #1 in this figure can be decomposed into two basic processes: a

“nonpolar” process of inserting the uncharged solute into solvent (#7) and a “polar” process of charging the solute in vacuum (#2) and solvent (#6). The free energy change in #7 is called the nonpolar solvation energy. The difference of energies associated with #6 and #2 is called the “charging” or polar solvation energy and represents the solvent’s effect on the solute charging process. The polar portion of solvation originates from electrostatic interactions, which are ubiquitous for any system of charged or polar molecules, such as biomolecules (proteins, nucleic acids, lipid bilayers, sugars, etc.) in their aqueous environment [240, 65, 69, 105, 200, 239, 203, 86, 204, 7, 82, 9]. The nonpolar portion describes the remaining contributions, including the surface tension, mechanical work, and attractive solvent-solute dispersion interactions.

1.1.2 Polar solvation models

Electrostatic interactions are ubiquitous in nature. For biomolecular systems in aqueous environment, the analysis of molecular solvation and electrostatics is of great importance to research in chemistry, biophysics, medicine and nano-technology. Implicit solvent models are widely used in such an analysis which can be classified into two general types: quantitative analysis and qualitative study. One of the primary quantitative application in computational biology and chemistry has been the calculation of thermodynamic properties. Implicit solvent methods “pre-equilibrate” the solvent and mobile ions, thus effectively pre-compute the solvent contribution for a system [186]. Such pre-equilibration is particularly evident in MM/PBSA models [246, 216, 173, 220, 149], which combine implicit solvent approaches with molecular mechanical models to evaluate binding free energies from an ensemble of biomolecular structures [16, 2, 133, 162, 145, 223, 161, 251, 162, 95, 150, 132, 131]. These methods

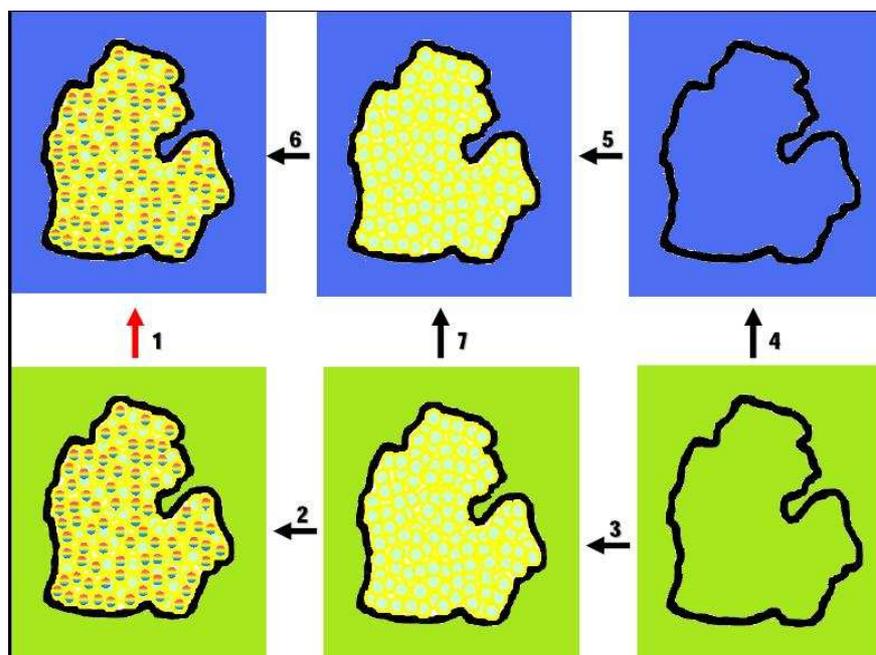


Figure 1.1: A solvation free energy cycle adapted from Levy et al. [130]. The total solvation energy (1) is decomposed into several steps: “charging” the solute in solvent (6) and vacuum (2), including attractive dispersive solute-solvent interactions in solvent (5) and vacuum (3), and cavity formation associated with repulsive solute-solvent interactions (4). The energy associated with Step (7) is generally termed a “nonpolar solvation energy” while the difference in energies associated with Steps (1) and (7) is generally considered as “polar solvation energy”. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

have been employed to interpret experimental titration curves, analyze residue contributions in protein-protein and protein-ligand binding energetics, examine structural/functional consequences of RNA nucleotide protonation, etc. Another quantitative application of implicit solvent models is the evaluation of biomolecular kinetics where implicit solvent models are generally taken to compute solvation forces for molecular Langevin dynamics [219, 180, 181, 142, 141], Brownian dynamics [147, 89, 76, 195], or continuum diffusion [49, 50, 255, 210, 211] simulations. A major qualitative study of implicit solvent methods is the visualization and qualitative analysis of electrostatic potentials on and around biomolecular surfaces [241, 177, 11, 7]. Visualization has become a standard procedure in the analysis of biomolecular structures, including ligand-receptor binding, drug design, macromolecular assembly, protein-nucleic acid complexes, protein-protein interactions, enzymatic mechanism study, etc.

The polar solvation energy is generally associated with a difference in charging free energies in vacuum and solvent (see Figure 1.1 (#2) and (#6)). Polar solvation process and electrostatic effect are described by a variety of implicit solvent models [240, 186, 239, 200, 65, 92, 120, 41, 193, 233, 198, 10, 9, 221]; however, the most widely-used ones are Poisson-Boltzmann (PB) models [105, 69, 7, 125, 86, 200, 65, 117], generalized Born (GB) methods [68, 15, 229, 167, 92, 263, 120, 226, 155, 41, 102] and polarizable continuum models (PCM) [52, 227, 113, 218, 35, 14, 59]. Polarizable continuum models are proposed to model the solvent either as polarizable dielectrics or as conductor-like media, and treat the solute compound by the quantum mechanical means [52, 227, 113, 218, 35, 14, 59]. These approaches have often been used in reactive kinetics where quantum mechanical descriptions are desired. Generalized Born methods are relatively fast, but are not as accurate as the PB

methods [9, 166, 68, 166, 63, 230, 226]. They are often employed in high-throughput applications such as molecular dynamics [15, 229, 203, 82, 120, 68, 167, 63, 41]. PB methods can be formally derived from more detailed theories [22, 159, 107] and provide a more accurate, although somewhat slower, approach for evaluating polar solvation properties [63, 166, 15]. Moreover, unlike most generalized Born methods, PB models offer a global description for the electrostatic properties, therefore making them uniquely suited to visualization and other studies [138, 29, 66, 234, 89, 76, 64, 196, 211] where the electrostatic information is required for both inside and outside a biomolecule.

1.1.3 Poisson-Boltzmann theory

Mathematically, the PB equation [105, 69, 7, 125, 86, 199, 65] is a nonlinear elliptic partial differential equation (PDE) which is solved for the electrostatic potential. It is a continuum model at equilibrium state, which dictates the solvent with a piecewise dielectric constant and ionic charge density by the Boltzmann distribution. The PB equation can be derived by the Gauss law and the Boltzmann distribution law [108]. Additionally, in the physical point of view, the free energy of the system must be minimized at the equilibrium state. Therefore, a total electrostatic free energy functional may be developed based on the PB theory, then the PB equation can also be obtained by the variational principle [199]. The standard formula of the PB equation is the following:

$$-\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi) - \sum_{i=1}^{N_c} Q_i n_i^0 e^{-\phi Q_i/k_B T} = \sum_j^{N_m} q_j \delta(\mathbf{r} - \mathbf{r}_j) \quad (1.1)$$

where ϵ is piecewise constant and depends on interface, being ϵ_s in solvent and ϵ_m in solute; ϕ is electrostatic potential. Here q_j is the partial charge on an atom located at \mathbf{x}_j , Q_i is the

charge of ion species i , N_c is the number of ion species, k_B is the Boltzmann constant, T is the temperature, N_m is the total number of solute atoms, and n_i^0 is the bulk concentration of the i th ionic species. Note that the PB equation can appear in different forms according to purposes as well as unit representation (detailed description can be found in Appendix B)

The PB theory is approximate and, as a result, has several well-known limitations which can affect its accuracy [105, 159, 107, 69, 239, 198, 81, 56, 221, 222, 191, 45]. These limitations have been reviewed in the literature and will only be briefly summarized here. First, most continuum models assume linear and local solvent response [239, 198, 81, 22]. However, nonlinear solvent response (usually through dielectric saturation or electrostriction), can be important in regions of strong electric field [239, 198, 81]. Biologically-relevant examples of nonlinear solvent response have been found near highly charged ions, biomolecules, and other interfaces. Nonlocal solvent response generally involves the finite non-zero size of water and its unique hydrogen bonding with solute and other solvent molecules. Such nonlocal response can be important in describing the orientation of water at biomolecular interfaces [38], differing solvation of cations and anions, and the solvation of asymmetric charge distributions. The second major limitation is the mean-field treatment of ions in PB theory [107, 159, 105]. Mean field models assume that each ion experiences only the average influence of the other ions in solution. Such averaging precludes detailed ion-ion interactions involving steric repulsion of ions (or their solvation shells) and Coulombic interaction of ions, including repulsion and attractive pairing. The mean field assumption thereby eliminates correlations and fluctuations which can have important energetic and structural consequences for solutions of divalent and multivalent ions surrounding highly-charged molecules such as nucleic acids [56, 221, 222, 191, 45]. As suggested by the limitations above, PB models also neglect de-

tailed ion-solvent interactions which eliminate differences between ion species in solution and thereby prevent effects analysis of specific ion species – which can be important in biophysical modeling. However, despite these limitations, PB methods are still very important for biomolecular structural analysis, modeling, and simulation. Furthermore, these limitations are currently being addressed through new implicit solvent models [5, 56, 159, 221, 175] and hybrid treatments [232, 13, 128, 165, 156] which extend the applicability of the PB theory while preserving some of its computational efficiency through pre-averaging solvent and ion response.

1.1.4 Nonpolar solvation models

Poisson-Boltzmann methods provide *polar* solvation energies and therefore must be complemented by *nonpolar* solvation models to provide a complete view of solvent-solute interactions. As illustrated in Figure 1.1, nonpolar solvation is generally associated with the insertion of the uncharged solute into solvent. There are many nonpolar solvation models available. The most commonly used one is solvent-accessible surface area (SASA) models. They states that nonpolar solvent-solute interactions are proportional to the area of the solvent-solute interface. It is worth to note that the proportional constant varies dramatically in the literature because the energies of other processes are also assumed to be proportional to SASA [77]. Roughly speaking, SASA models are based on the scaled particle theory (SPT) [213, 178] which actually includes the energy of surface tension effect and the mechanical work of immersing a particle into the solvent. Moreover, studies indicates that nonpolar distribution should depend on the solvent-accessible volume and surface, with a crossover to SASA when the size of solute is large [235]. Recent work by Levy, Gallic-

chio, and others [90, 92, 130, 91, 235] has demonstrated the importance of nonpolar solvent models which include treatment of attractive solute-solvent dispersion terms (#5 in Figure 1.1) as well as models of solvent-solvent repulsive interactions (#4 in Figure 1.1), which are described by both area and volume contributions [235]. Based on these considerations, in the present work, we use the following model for nonpolar solvation free energies [235]

$$G_{\text{np}} = \gamma \cdot \text{Area} + p \cdot \text{Vol} + \int_{\Omega_s} \rho_s U_{ss} d\vec{r}, \quad (1.2)$$

where γ is the surface tension, "Area" is the solvent-excluded area of the solute, p is the hydrodynamic pressure, "Vol" is the solvent-excluded volume of the solute, ρ_s is the solvent density, Ω_s denotes the solvent accessible region, and U_{ss} is the solvent-solute van der Waals (vdW) interaction potential. The first two terms in Eq. (1.2) are those from the SPT model [213, 178]. The first term is the surface energy. It measures the disruption of intermolecular and/or intramolecular bonds that occurs when a surface is created. The second term is the mechanical work of creating the vacuum of a biomolecular size in the solvent. The third term represents the attractive dispersion effects near the solvent-solute interface which has been shown by Wagoner and Baker [235] to play a crucial role in accurate nonpolar solvation analysis. In general, U_{ss} can be obtained by the sum of the interaction of individual atoms in Ω_m with the solvent continuum in Ω_s under the assumption that the nonpolar solute-solvent potential is pairwise: $U_{ss} = \sum_i V_i^{\text{vdW}}$. This model of nonpolar solvation has been demonstrated to give good agreement with explicit solvent solvation forces on proteins [235] and RNA hairpins [71]. Work by Levy and co-workers has demonstrated the good performance of a similar model [90, 92, 130, 91].

In the present work, we further allow the solvent density ρ_s to be a function of position in

general. In particular, we split the solvent density ρ_S into the sum of atomic or ionic density distribution functions $\rho_S = \sum_i \rho_{S,i}$. The distribution of an individual solvent component can be computed by integral equations or other approaches, such as Monte Carlo and generalized Langevin equation [231, 88, 23]. This design of solvent density allows us to recover the nonlinear and nonlocal effects of the solvent-solute interactions.

1.2 Molecular interface definitions

The separation of discrete and continuum domains in implicit solvent models requires an interface to indicate the separation of solute atoms from the surrounding solvent. Naturally, such an interface can be regarded as the surface or the profile of a molecule. The definition of molecular profiles, or molecular graphics traces back to Corey and Pauling in 1950s [58], who tried to depict the profiles of amino acids, peptides and proteins from X-ray crystallography. In quantum chemistry, molecular graphics are often associated with the shapes of polynomial functions that provide approximation to electron wavefunctions. In fact, since the electron wavefunction changes its distribution under different environments, molecular profiles change accordingly. Commonly used interface definitions in implicit solvent models include the van der Waals surface, the solvent accessible surface [126], and the molecular surface (MS) [185, 57]. In certain sense, these interface definitions determine the performance of implicit solvent models because all of the physical properties of interest, including electrostatic free energies, biomolecular surface areas, molecular cavitation volumes, solvation free energies, and pK_a values are very sensitive to these interface definitions [70, 72, 163, 217].

The use of PB model encounters some challenges in molecular dynamics regarding stability and accuracy. For example, the widely used molecular surface based PB model results

in forces which are unstable over time, lack analytical expression, are sensitive to grid discretization, and converge poorly [101]. Moreover, a discontinuous dielectric definition leads to numerical instability regardless of the location of boundary [217]. Additionally, more physically realistic surface definitions are desired because of the argument that the macroscopic physical properties must vary continuously. To overcome these difficulties, overlapping atom-centered Gaussian or polynomial functions have been proposed to define the solute surface, with smooth transitions between low and high dielectric values [101, 112]. Although continuous dielectric functions give rise to an improvement of stability and computational efficiency, some recent work demonstrates that most of them are physically incorrect, which will be discussed more in Section 1.4.

The recent development of a new class of molecular interfaces that incorporate the fundamental laws of physics starts with the construction of partial differential equation (PDE) based molecular surface by Wei et al. in 2005 [245]. This approach distinguishes itself from many other PDE based surface smoothing methods [249, 256] by utilizing only atomic information, i.e., atomic coordinates and radii, instead of an existing surface. The atomic information is embedded in the Eulerian formulation and a family of hypersurfaces are evolved in time under the PDE operator, which is designed to control the curvature and surface tension. The generalized molecular surface is subsequently extracted from the final hypersurface by a level-set type of approach [245]. This PDE based surface construction procedure generates well defined molecular surfaces for both small molecules and large proteins [245]. To our knowledge, geometric PDE based approach was the first of its kind for molecular surface construction. A further progress in the development of a “physical interface” was the introduction of the minimal molecular surface (MMS) that minimizes a surface free energy

functional by the variational principle and leads to the mean curvature flow in 2006 [18, 19]. To our knowledge, MMSs were the first set of biomolecular surfaces that had ever been constructed by means of variational principles. The construction of the MMS was driven by the desire to understand the true physical boundary of a biomolecule in solvent. As a physical concept, the solvent-solute interface should be in general determined by the minimization of the free energy of a macromolecule in the aquatic environment. The MMS is constructed by using essentially the same procedure as that developed in the first PDE based surface generation method [245]. Another desirable property of the MMS is that it is free of geometric singularities. The MMS model was applied to the calculation of electrostatic solvation free energies of 26 proteins [20].

1.3 Quantum mechanical continuum models

In most implicit solvent models, the solute is described as a collection of fixed atomic point charges, which describes molecular polarizations at the atomistic level of resolution. However, it is well-known that charge rearrangement plays an important role in the solvation process of proteins in various cases [99]. Similar arguments were used to incorporate the quantum mechanics (QM) description in the classical implicit solvent theory [227, 228, 61]. The resulting QM version of continuum models, called quantum mechanical continuum models [52, 227, 113, 218, 35, 14, 59], offer the possibility of carrying out accurate quantum calculations in solution and near interfaces. Quantum mechanical continuum models provide a framework to describe the QM effect in solvent analysis [224]. However, this description is often compromised by the use of a pre-determined solvent-solute interface model. To integrate a continuum model with a QM description, reaction field potential, i.e., the electric

field induced by the polarized solvent, has been introduced as a unifying concept. It is obtained from the electrostatic computation in the framework of continuum models. It also exists in the Hamiltonian of the solute in the quantum calculation [224, 237, 42]. Therefore, the quantum formulation of the continuum model involves two problems: (1) the classical electrostatic problem of determining the solvent reaction field potential with the quantum mechanically calculated charge density; (2) the quantum mechanical problem of calculating the electron charge density with fixed nucleus charges in the presence of the reaction field potential. These two problems need to be resolved simultaneously. To carry out these computations, a intuitive self-consistent iterative procedure can be constructed to resolve the quantum problem for electron distribution and the classic electrostatic problem for the reaction field potential [224, 99, 227, 237, 42].

After computing the QM charge density, there are still two ways to implement the solvation analysis. The first approach is to use the continuous QM charge density directly in the Poisson-Boltzmann equation. The second approach is to fit the QM charge density into the atomic point charges, and then use the point charges as the source term in the Poisson-Boltzmann equation. Various schemes have been proposed to compute atomic partial charges with certain efficiency and convenience. The simplest way for atomic partial charge assignments is the Mulliken analysis method [157]. In this approach, the charge is distributed according to the orbital occupation. Many other schemes have also been proposed, including the natural bond orbital analysis, the distributed multipole analysis (DMA), the wavefunction mapping ‘Class IV’ model, the electrostatic potential expansion and analysis, etc [184, 214, 202]. Presently, the most widely used method for estimating atomic partial charges is the least-squared electrostatic potential (ESP) fitting approach.

It was first proposed by Momany and has subsequently been implemented in different ways with different choices of grid points where the electrostatic potentials are calculated [154]. Examples of such potential-based methods are CHELP, CHELPG, and the Merz-Kollman scheme [60, 205, 28, 53]. Hu and Yang have recently developed a new object function to improve the quality and especially the numerical stability of the ESP fitting [110]. ESP fitting methods are not only widely used in the simulation with molecular mechanical (MM) force fields, but also extended to the QM/MM simulations as well as the molecular polarization calculation [110]. However, there are some well-known deficiencies in the atomic partial charge approaches [201, 110]. First, atomic partial charge is not observable, i.e., it can not be definitely determined by experimental data or directly obtained from quantum calculations. Therefore, it is a term lacking a rigorous and consistent definition [110]. Additionally, the approximation of quantum mechanical electron-electron interactions by simple Coulombic interactions between atomic partial charges leads to inaccurate calculations. Moreover, results from different methods or definitions may show different numerical dependences upon the QM level of the theory, basis sets used, and the choices of the number and location of grid points. Finally, there is a concern about the transferability of the atomic partial charges in different molecules. To avoid these problems, the direct use of the quantum charge density in the continuum dielectric theory was proposed [227, 237].

The quantum mechanical problem of determining the electron charge density was solved originally limited to the Hartree-Fock level, which is a traditional way to obtain complicated many-electron wavefunctions. Density functional theory (DFT) was proposed in 1960s to provide the ground state of a many-electron system in terms of a single electron wavefunction [106, 122]. In DFT, the Kohn-Sham equation is the Schrödinger equation of a fictitious

system. It has been more and more popular for quantum calculations in solid state physics since the 1970s due to its low computational cost when compared with traditional approaches. Moreover, the results of DFT calculations have been considered accurate enough especially from 1990s when approximations used in the theory were greatly refined to better model the exchange and correlation interactions. [127, 21, 33]. DFT is now a leading method for electronic structure calculations in chemistry, physics and nano-technology. Therefore, the incorporation of DFT to continuum solvation methods becomes a routine approach in methods with the QM description of solute and the continuum description of solvent [227].

1.4 Limitations of current models

Current two-scale implicit solvation models have a severe limitation that undermines their performance in practical applications. While traditional surface definitions have found much success in biomolecular modeling and computation [212, 139, 62, 123, 25, 73, 114, 136], they are simply *ad hoc* divisions of the solute and solvent regions of the problem domain. In reality, the solvation is a physical process and its equilibrium state should be determined by fundamental laws of physics. Moreover, these surface definitions confront many challenges. First of all, as mentioned earlier, from the fundamental physical point of view, macroscopic properties should vary continuously. Any description of the permittivity changing instantaneously from one point to another is incorrect. Secondly, they admit non-smooth interfaces, i.e., cusps and self-intersecting surfaces, that lead to well-known instability in molecular simulations due to extreme sensitivity to atomic positions, radii, etc [188]. Thirdly, each pre-determined surface definition has its own limitations. For example, the van der Waals surface (VDW) is not smooth in space. And there are a lot of un-physical solvent pock-

ets inside the solute which cause the fluctuation of the electrostatic field [141]. the widely used molecular surface (MS) accompanying with the discontinuous dielectric function is not smooth in time for the molecular dynamic due to its definition, while it is much smoother in space than VDW and embodies the ratio between contact surface and reentry surface which maybe an important information of surface roughness. The solvent accessible surface (SAS) is not suitable in terms of solvation energies, which are the strongest validation of the PB theory. These difficulties associated with traditional discontinuous dielectric functions often drive the use of alternative “smoothed” solvent-solute interface definitions [112, 101] by applying overlapping atom-centered Gaussian or polynomial functions. However, smoothed interface definitions increase computational cost [70, 72]. Moreover, interatomic crevices and buried pockets of high dielectric, which are too small for a solvent molecule to occupy, are introduced. Furthermore, they often overestimate the electrostatic solvation free energy [217]. Finally, the wide range of surface definitions has often led to confusion and misuse of parameter (radii) sets developed for implicit solvent calculations with specific surface definitions.

It is celebrated that the construction of partial differential equation (PDE) based molecular surface by Wei et al. in 2005 [245] generates well defined molecular surfaces for both small molecules and large proteins [245]. Moreover, the construction of minimal molecular surface (MMS) is determined by the minimization of the free energy of a macromolecule in the aquatic environment, and it is free of geometric singularities [20]. However, the MMS, which incorporates only the minimization of the free energy associated with the surface tension, offers only an approximation to the true physical boundary of a biomolecule in solvent. Therefore, to account for other important effects that determine the solvent-solute inter-

face, new potential driven geometric flows (PDGFs) needs to be proposed that allow the incorporation of many other potential effects in surface formation and evolution.

Another criticism of implicit solvent models is the lack of uniqueness in polar and nonpolar decomposition of the solvation process [148] and the neglect of the polar-nonpolar coupling as well as solvent-solute interactions [4, 32, 38, 55, 74, 85, 87]. Dzubiella et al [74, 75] considered this problem by adding a solvent-solute coupling (interaction) term to the total free energy functional discussed by Sharp and Honig [200] and Gilson et al [97]. A feature of this new model is that surface tension energy and mechanical work of immersing a molecule into the solvent were also included in the total free energy functional. However, their initial work does not provide a protocol for the construction of molecular interfaces and a systematical analysis of solvation energy for macromolecules. Recently, Cheng et al. [48] have extracted solvent-solute interfaces from the free energy functional of Dzubiella et al [74, 75] in a setting very similar to our earlier Eulerian geometric PDE approaches of biomolecular surfaces and solvation analysis [245, 18, 19, 20].

In the earlier PDE based molecular surface models [245, 20, 74] the solute is described as a collection of fixed atomic point charges, which, together with the charge described in the continuum approximation of the surrounding medium, give rise to the total charge source for the Poisson-Boltzmann equation. The atomic partial charges describe molecular polarizations at the atomistic level of resolution. This approach is able to evaluate many physical and chemical properties. However, it can not cover the whole range of properties of interest. In particular, the charge rearrangement in the solute molecule during the transfer from the gas phase to solution has not been taken into account in the calculation of solvation free energies. Consequently, the highly accurate analysis of solvent-solute surfaces is discounted

by the estimation of charge rearrangement during the solution process. Additionally, those earlier solvation models depend on parameters from the existing molecular mechanical force field parameters [147, 12, 112, 116], which are typically parameterized for certain class of (macro-) molecular systems and may not be appropriate for other class of molecules. Therefore, it is desired to develop a quantum mechanical (QM) description of the solute molecule in PDE based molecular surface models, which gives accurate, self-consistent and force field independent charge arrangement treatment during the solavtion process.

1.5 Mathematical issues and numerical challenges

Based on previous discussions, it is clear that there are various mathematical issues involved in the modeling of solvation. In this work, the solvent-solute boundary can be modeled in the framework of the differential geometry theory of surfaces and manifolds, which is employed to result in new coupled geometric and potential flows for the generation of a physical solvent-solute boundary and the optimization of solvation energy. Technically, the smoothness of the resulting solute-solvent boundary is ensured by coupled geometric and potential flows of parabolic type. A variational framework is established to couple different parts of the solvation contributions. Governing equations are derived by variational principles. Other than the modeling strategy, implementations of models encounter many numerical challenges, which have attracted great mathematical interests for the past several decades. For instance, the multidomain and multiscale treatment of both systems results in discontinuous coefficient based interface problems and a singular source in the partial differential equation. Highly accurate and efficient numerical schemes are desired to handle these singularities in the application of biological systems. Additionally, numerical convergence and efficiency of the

self-consistent iteration need to be explored for the derived coupled governing equations. The involved mathematical issues and associated numerical challenges in model implementations are outlined as follows:

1.5.1 Geometry, PDE and interface

In this thesis, we consider the solvent-solute boundary as a two-dimensional (2D) differential manifold embedded into a 3D Euclidean space, or a hypersurface, in a Riemannian manifold. The differential geometry theory of surfaces and manifolds is employed. The above-mentioned minimal molecular surface, which incorporates only the minimization of the free energy associated with the surface tension, offers only an approximation to the true physical boundary of a biomolecule in solvent. To account for other important effects that determine the solvent-solute interface, we have recently proposed a framework of potential driven geometric flows (PDGFs) that allow the incorporation of many other potential effects in surface formation and evolution [17]. The PDGFs are inherently multiscale in nature, and enable the incorporation of microscopic interactions, such as van der Waals potentials, into the macroscopic curvature evolution.

From a mathematical point of view, the molecular surfaces are constructed from the geometric partial differential equation (GPDE) in this thesis. In general, GPDE is a PDE which controls the motion of curves or surfaces and is merely formulated by the geometric measure theory. It is related to geometric analysis, manifold theory, topology, PDE, calculus of variation, and geometric measure theory. GPDE can be applied to the motion of interfaces problems in the physical or chemistry field, e.g., dissolution, combustion, erosion in the biology field, biomembrane vesicle problem, and the construction of protein surface. GPDE is

also applied to computational geometry, computer graphics, image processing edge detection, noise removal, image restoration. Moreover, it can be used to obtain some optimal properties such as surface area minimization, total energy minimization, etc.

The surfaces generated by GPDE possess some attractive features such as smoothness and a clear geometric sense. GPDE can be constructed by an energy functional based variational approach. In this method, an energy functional with a physical target is formed. Then, the first order variation of the energy functional gives rise to geometric PDEs. In the framework of variational approach, many famous geometric flow equations are derived such as the mean curvature flow, the Willmore flow, etc.

Because it is difficult to gain explicit solutions of the GPDEs, numerical solutions are necessary. Numerical solutions of Geometric PDEs can be obtained by the generalized finite difference method, the finite element methods and the level set like methods.

1.5.2 Geometric flow equation

Much of the recent development in implicit solvent models is due to the use of geometric flows [245, 18, 20, 17, 243], particularly mean curvature flows, which have been of considerable interest in applied mathematics for decades [172, 187, 197, 84, 100, 153, 170, 190, 192, 197, 208, 247]. Earlier research work and part of present research are focused on image processing [172, 187], computer vision, materials design [197] and surface smoothing [249, 256]. Computational techniques using the level set theory were devised by Osher and Sethian [172, 187, 197] and have been further developed and applied by many others [37, 54, 206]. An alternative approach for image analysis is to minimize a functional in the framework of the Mumford-Shah variational functional [158], and/or the Euler-Lagrange formulation of

variation [30, 36, 134, 171, 187, 189]. Wei introduced some of the first family of high-order geometric flow equations for image analysis [242]. In fact, the nonlinear production term in these high-order operators provides a framework to accommodate the PDGF in our later formation for macromolecular surfaces. Their high-order geometric flow equations have led to many interesting applications [242, 244, 215, 144, 96, 39]. Mathematical analysis of these high order equations in Sobolev space was carried out by Bertozzi and Greer [27, 103, 104], who proved the existence and uniqueness of the solution to a case with H^1 initial data and a regularized operator. A similar analysis was performed by Xu and Zhou [250]. Wei and Jia also introduced a coupled geometric flow equation system for image edge detection [244]. Such an algorithm works extremely well with texture images. Recently, Wei and his collaborators have proposed an evolution operator based single-step method for image denoising and enhancement [215]. Most recently, a family of differential geometry based multiscale models has been developed by Wei for chemical and biomolecular systems, including fuel cells, ion channels, DNA packing, nanofluidic systems, and virus evolution [243]. These models describe not only the structure, but also the dynamics and transport of the above mentioned chemical and biomolecular systems.

1.5.3 Highly accurate and efficient solver for interface problems

In general, electrostatic energy is much larger than the non-electrostatic part so that the accuracy of electrostatic potential calculation based on the Poisson-Boltzmann (PB) equation plays a critical role in controlling the accuracy of the total solvation free energy. Therefore, numerical methods that are able to deliver highly accurate solution of the PB equation is desirable. In the present Lagrangian model, there exists a sharp solvent-solute interface and

it leads to a discontinuous dielectric constant definition in the PB equation. When the discontinuous dielectric profile is applied across the interface, the standard numerical methods, including the centered finite differences scheme, lose their accuracy and convergent order. This problem is aggravated by complex geometric shapes, possible geometric singularity, and singular charges of biomolecules. In the worst-case scenario, the standard numerical methods do not converge at all for complex irregular solvent-solute interfaces [252, 93].

The solution of elliptic equations with discontinuous coefficients and singular sources is a challenging problem in computational mathematics. In order to achieve high-order numerical accuracy, it is indispensable to develop mathematical interface techniques. Peskin pioneered the immersed boundary method (IBM) [124, 176] to address this class of problems. Recently, many other elegant methods have been proposed, including the ghost fluid method [79, 137], the upwinding embedded boundary method [34], finite-volume-based methods [164], and integral equation methods [151]. A major advance in the field was due to LeVeque and Li [129], who proposed a remarkable second order sharp interface scheme, the immersed interface method (IIM) [129, 135]. Chen and Strain discussed a piecewise-polynomial discretization and Krylov-accelerated multigrid for elliptic interface problems [44]. However, these interface techniques have not been implemented for the Poisson-Boltzmann equation in the context of realistic biomolecules.

Wei and his coworkers have recently proposed a highly accurate algorithm, the matched interface and boundary (MIB) method [254, 253, 258, 261, 260] for solving elliptic equations. Many essential ideas of the current MIB method were introduced in earlier interface schemes for solving Maxwell's equations [258]. The MIB is of arbitrarily high-order accuracy in principle, and sixth-order accurate MIB schemes have been constructed [253, 261]. Wei's

group has developed three generations of MIB based PB solvers, MIBPB-I [259], MIBPB-II [252], and MIBPB-III [93]. The MIBPB-I is the first PB solver that explicitly enforces the flux continuity conditions at the dielectric interface in a biomolecular context; however, it cannot maintain its designed order of accuracy in the presence of molecular surface singularities, such as cusps and self-intersecting surfaces commonly occurred in biomolecular systems [188]. This problem was first addressed in the MIBPB-II by utilizing an advanced MIB technique developed by Yu et al. [253]; however, the MIBPB-II still loses its accuracy when the mesh size is as large as half of the smallest van der Waals radius, because of the interference of the interface and singular charges. To split the singular charge part of the solution [262, 43, 31], a Dirichlet to Neumann mapping approach [51] was designed in the MIBPB-III, which is by far the most accurate and stable PB solver. To our knowledge, the MIBPB method is the only existing method that is able to offer second order accuracy in solving the Poisson-Boltzmann equation with discontinuous coefficients, singular sources and arbitrarily complex interfaces. The MIBPB is a few orders of magnitude more accurate at a given mesh size and about three times faster at a given accuracy than some traditional PB solvers [93].

The most important idea in all interface techniques is to take care of interface conditions, which may vary from systems to systems. Complex interface conditions are needed for the Helmholtz equation [257] and Maxwell's equations [258]. For the Poisson-Boltzmann equation, interface conditions are the following

$$\begin{aligned}
[\phi]_{\Gamma} &= \phi^+(\mathbf{r}) - \phi^-(\mathbf{r}) = 0 \\
[\epsilon\phi]_{\Gamma} &= (\epsilon_S \nabla \phi^+) \cdot \mathbf{N} - (\epsilon_m \nabla \phi^-) \cdot \mathbf{N} = 0.
\end{aligned} \tag{1.3}$$

where ϕ^+ and ϕ^- are the electrostatic potential inside and outside the solvent-solute surface, respectively. Different methods may have different strategies in dealing with these conditions. The MIB method has a unique set of procedures in implementing Eq. (1.3). The interested reader is referred to earlier work [254, 253, 258, 261, 260].

In this work, we make use of MIBPB-III scheme for PDE based interface problems. We take dielectric constants $\epsilon_m = 1$ and $\epsilon_s = 80$ in our calculations. We use the Dirichlet far-field boundary condition and the electrostatic potential values at the boundary are practically obtained by the sum of potential contributions from individual atomic charges with an exponential decay factor [93]. The MIBPB-III is used to handle discontinuous dielectric constants, complex geometry and charge singularity. Note that although the geometry is complex, there is no geometric singularities, such as cusps and intersecting surfaces, in the biomolecular surfaces generated by our approaches. The extraction of surface information is carried out by the marching cubes algorithm embedded in our codes.

1.5.4 Self-consistent iterative methods

It will be seen, through the optimization of the solvation energy, that the resulted generalized Poisson-Boltzmann (PB) equation and the generalized potential driven geometric flow equation are fully coupled. The optimized electrostatic potential is obtained by solving the Poisson-Boltzmann equation in which solvent-solute interface is used to determine the dielectric constant and the domain decomposition. The interface is generated by the solution of the potential driven geometric flow equation, which in turn depends on the electrostatic potential. In other words, the solution of the potential driven geometric flow equation requires the knowledge of electrostatic potential, while the solution of the generalized PB equation

requires the input of the interface definition.

Therefore, the coupled generalized geometric flow equation and the generalized PB equation need to be solved simultaneously, in the present differential geometry based solvation model. The existence and the uniqueness of their solution, under the biomolecular context, can be an interesting mathematical issue. Numerically, both the convergence and efficiency of the solutions of the coupled system will be given quite a bit of attentions.

In practice, this coupled nonlinear system can be solved by an iterative procedure, until a self-consistency is reached. Iterative methods are often the only choice for nonlinear equations. In computational mathematics, an iterative method is a mathematical procedure that leads to approximate solutions for a class of problems. The most common iterative method is Newton's method. Other well-known examples include the Gummel's method, the steepest descent method, and the conjugate iterative method. An iterative method is called convergent if the corresponding solution sequence converges for given initial approximations. A mathematically rigorous convergence analysis of an iterative method need to be performed; however, heuristic-based iterative methods are also commonly used.

1.6 The rest of this thesis

The rest of this thesis is organized as follows. In Chapter 2, we present the Eulerian formulation of our differential geometry based solvation models. The Eulerian analysis of biomolecular surfaces utilizes the well-known coarea theorem from the geometric measure theory. The resulting operator from surface area minimization can be identified as a generalized Laplace-Beltrami operator in a 3D dimension. Chapter 4 is devoted to the Lagrangian formulation. Lagrangian analysis of biomolecular surfaces makes the direct use of differential

geometry theory of surfaces and manifolds. The surface minimization leads to the Laplace-Beltrami operator defined in 2D domain, or the mean curvature operator. In Chapter 4, the connection of two representations is analyzed in the present work. The structure of governing equations, and the accuracy and efficiency of two formulations are compared. The objective of Chapter 3 is to incorporate a quantum mechanical description of charge density into our earlier differential geometry based solvation model, which is described in Chapter 2. This thesis is concluded by discussing the main achievements and future work.

Chapter 2

Eulerian formulation

This chapter presents a differential geometry based model for the analysis and computation of the equilibrium property of solvation. Differential geometry theory of surfaces is utilized to define and construct smooth interfaces with good stability and differentiability for use in characterizing the solvent-solute boundaries and in generating continuous dielectric functions across the computational domain. A total free energy functional is constructed to couple polar and nonpolar contributions to the solvation process. Geometric measure theory is employed to rigorously convert a Lagrangian formulation of the surface energy into an Eulerian formulation so as to bring all energy terms on an equal footing. By minimizing the total free energy functional, we derive coupled generalized Poisson-Boltzmann equation (GPBE) and generalized geometric flow equation (GGFE) for the electrostatic potential and the construction of realistic solvent-solute boundaries, respectively. By solving the coupled GPBE and GGFE, we obtain the electrostatic potential, the solvent-solute boundary profile, and the smooth dielectric function, and thereby improve the accuracy and stability of implicit solvation calculations. We also design efficient second order numerical schemes for the solution of

the GPBE and GGFE. Matrix resulted from the discretization of the GPBE is accelerated with appropriate preconditioners. An alternative direct implicit (ADI) scheme is designed to improve the stability of solving the GGFE. Two iterative approaches are designed to solve the coupled system of nonlinear partial differential equations. Extensive numerical experiments are designed to validate the present theoretical model, test computational methods, and optimize numerical algorithms. Example solvation analysis of both small compounds and proteins are carried out to further demonstrate the accuracy, stability, efficiency and robustness of the present new model and numerical approaches. Comparison is given to both experimental and theoretical results in the literature.

This chapter is organized as follows. Section 2.1 is devoted to the theoretical foundation of the present differential geometry based solvation model. A variational framework is established to couple different parts of the solvation contributions. Governing equations are derived by variational principles. The solution of the governing equations leads to physical solvent-solute boundaries and accurate solvation free energies. Numerical methods and algorithms are discussed in Section 2.2. Schemes of the second order numerical accuracy are designed for the construction and evolution of solute characteristic function. Appropriate preconditioners are used for solving the generalized Poisson-Boltzmann equations. The coupled equations are solved by two iterative schemes. Section 2.3 presents validation and analysis of the proposed numerical approaches. The accuracy and convergence of various computational schemes, including the surface area formulation based on the geometric measure theory, are carefully tested to ensure their computational reliability and efficiency. The applications of the proposed theories, methods and algorithms are considered to two sets of compounds: small molecules and proteins in Section 2.4. Finally, this chapter ends with

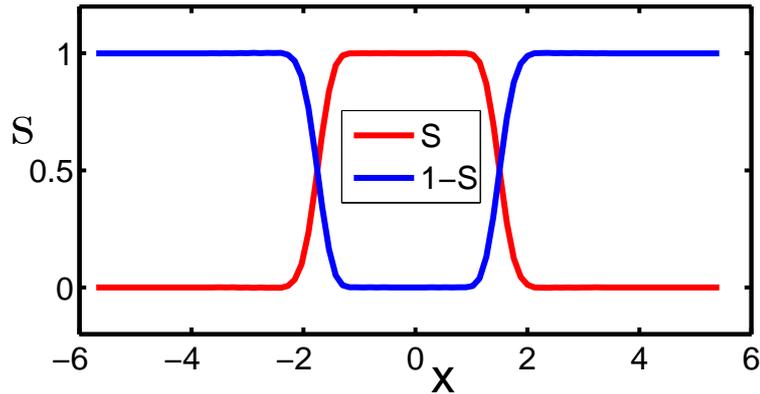


Figure 2.1: The cross line of S and $(1 - S)$ of a diatomic system described in Section 2.3.3

concluding remarks.

2.1 Theory and model

In this section, a differential geometry based model of solvation is briefly described for macromolecules and their aquatic environment that are near equilibrium. More details about the differential geometry based multiscale formalism, particularly dynamics and transport aspects, can be found elsewhere [243]. For a system near equilibrium, the density of charged particles in the solvent can be approximated by the Boltzmann distribution, which considerably reduces the number of degrees of the freedom of the solvation system. Alternatively, the Nernst-Planck equations or the full set of the Navier-Stokes equations might be utilized to describe systems that are far from the equilibrium [243].

2.1.1 Solute-solvent boundary

Let us consider a multi-domain setting of a macromolecule and solvent system. The macromolecule is described in discrete atomic detail, while the aqueous solvent is treated as a continuum. Therefore, the domain $\Omega \in \mathbb{R}^3$ is essentially divided into two (types of) regions, i.e., aqueous solvent domain Ω_s and macromolecular domain Ω_m . Therefore, one has $\Omega = \Omega_s \cup \Omega_m$. However, because electron wavefunctions of the solvent and the solute overlap at the atomic scale, Ω_s and Ω_m should overlap with each other at the boundary of molecules and solvent, i.e., $\Omega_b = \Omega_s \cap \Omega_m \neq \emptyset$, where Ω_b is the region of solvent-solute boundary. Therefore, we propose a characteristic function $S : \mathbb{R}^3 \rightarrow \mathbb{R}$ to characterize this overlapping solvent-solute boundary. As such, $S(x)$ is a description function or a characteristic function of the solute domain, i.e., it is one ($S = 1$) inside the biomolecule and zero ($S = 0$) in the aquatic solvent. However, S takes a value between zero and one at the solvent-solute boundary region. Therefore, $(1 - S)$ is a description function or a characteristic function for the solvent domain. The profiles of S and $(1 - S)$ are depicted in Figure 2.1 for a simple system. It is seen that there is a transition region at the solvent-solute boundary where the solvent and the solute regions overlap. Obviously, in our model, the evaluation of all the solvent-solute properties depends on S . Physically, S and thus the profile of solvent-solute boundary, must be determined by the energy minimization principle. Therefore, our task is to identify the energy functional that to be optimized. This task is accomplished via the differential geometry of surfaces and manifolds in the present work.

2.1.2 Total free energy functional

The solvation process of macromolecules involves a number of interactions. As discussed in the Introduction, typically, the free energy of solvation models consists of polar and nonpolar contributions, as well as polar and nonpolar interactions.

2.1.2.1 Polar free energy functional

The polar part is standardly represented by electrostatic interactions, which are of special importance because of their long range and influence on polar or charged molecules including water, aqueous ions, and amino or nucleic acids. They are also some of the most important aspects that determine the physical and chemical properties of biomolecules, such as protein folding, protein-DNA binding, gene expression and regulation, etc. The widely used free energy functional of the electrostatic system was given by Sharp and Honig [199] and Gilson et al [97]. However, their formulation is based on a sharp interface that separates the solvent and solute domains. In our formulation, we incorporate the function S into the polar solvation free energy functional

$$G_p = \int_{\Omega} \left\{ S \left[\rho_m \phi - \frac{1}{2} \epsilon_m |\nabla \phi|^2 \right] + (1 - S) \left[-\frac{1}{2} \epsilon_s |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 (e^{-\phi Q_i / k_B T} - 1) \right] \right\} d\mathbf{r} \quad (2.1)$$

where ϕ is the electrostatic potential whose domain is the whole computational domain Ω , and ϵ_s and ϵ_m are the dielectric constants of the solvent and solute, respectively. Here $\rho_m = \sum_j q_j \delta(\mathbf{r} - \mathbf{x}_j)$ is the density of molecular charges, with q_j being the partial charge on an atom located at \mathbf{x}_j , Q_i is the charge of ion species i , N_c is the number of ion species,

k_B is the Boltzmann constant, T is the temperature, and n_i^0 is the bulk concentration of the i th ionic species. The term associated with S is the electrostatic free energy of the solute and that with $(1 - S)$ is the electrostatic free energy of the solvent.

The above electrostatic free energy functional is inherently multidomain in nature and the domain is divided into the solute subdomain and the solvent subdomain as indicated by S and $1 - S$, respectively. These subdomains do not have to be mutually exclusive. A discrete description of the solute and a continuum description of the solvent are also employed in Eq. (2.1) in the framework of the implicit solvent treatment, in which the charge density of mobile ions follows the Boltzmann distribution. Moreover, it will be demonstrated that the present electrostatic free energy functional is able to reproduce the classical Poisson-Boltzmann equation when a sharp solvent-solute interface is used and S becomes a Heaviside function. Finally, we note that the terms that are quadratic in the potential gradient in Eq. (2.1) have negative signs. Therefore, the free energy will be optimized instead of being minimized. In this work, we have adopted the earlier sign convention in the field [199, 97].

2.1.2.2 Non-polar free energy functional

For the nonpolar contribution, we consider the nonpolar solvation free energy functional discussed in the Introduction.

$$G_{\text{np}} = \gamma \cdot \text{Area} + p \cdot \text{Vol} + \rho_0 \int_{\Omega_s} U^{\text{att}} d\mathbf{r} \quad (2.2)$$

To obtain a functional relation for S , it is necessary to rewrite nonpolar free energy

formulation in terms of $S(\mathbf{r})$. The enclosed volume of biomolecule can be given by

$$\text{Vol} = \int_{\Omega_m} d\mathbf{r} = \int_{\Omega} S(\mathbf{r})d\mathbf{r}. \quad (2.3)$$

Similarly the attractive dispersion term can be rewritten in the form

$$\rho_0 \int_{\Omega_s} U^{\text{att}} d\mathbf{r} = \rho_0 \int_{\Omega} (1 - S(\mathbf{r}))U^{\text{att}} d\mathbf{r}, \quad (2.4)$$

where we assume that the solvent bulk density ρ_0 is a constant in space.

Typically, one expresses the area of a unique surface as a surface integration over the biomolecular boundary in the Lagrangian formulation. However, this approach does not work directly in our formulation because no sharp solvent-solute boundary is assumed. In fact, the concept of the surface area cannot be defined in the same manner as in the sharp surface case. For a smooth boundary, there are infinitely many surfaces and the surface area can be defined as a weighted mean of this family of surfaces. Additionally, for practical purpose, we need an appropriate Eulerian formulation for the surface area so that we can put all energy contributions on an equal footing. Therefore, we need to convert the surface integral into a volume one. To this end, we make use of the coarea formula in the geometric measure theory [78]. For a scalar field B in \mathbb{R}^3 , with C^1 continuity condition, integrating a function f over its isolevel c in a region Ω can be done directly by a volume integral over Ω through the expression

$$\int_{\mathbb{R}} \int_{B^{-1} \cap \Omega} f d\sigma dc = \int_{\Omega} \|\nabla B\| f(\mathbf{r}) d\mathbf{r}, \quad (2.5)$$

where c denotes an isovalue of B , and B^{-1} represents the c -isosurface, i.e., the set of points $\{\mathbf{r}_c\}$ such that $B(\mathbf{r}_c) = c$. Here, the coarea formula prescribes a relationship between the sum of area integrals and a global volume integral. In our case, we introduce the concept of mean surface area of the family of isosurfaces which are subsets of point satisfying $\{S(\mathbf{r}) = y\}$, where $0 \leq y \leq 1$. Therefore the mean surface area can be given by a volume integral as

$$\begin{aligned} \text{Area} &= \int_0^1 \int_{S^{-1}(c) \cap \Omega} d\sigma dc \\ &= \int_{\Omega} \|\nabla S(\mathbf{r})\| d\mathbf{r}. \end{aligned} \quad (2.6)$$

Note that $\nabla S \neq 0$ only in the region of the solvent-solute boundary. Numerical test of this formulation will be presented in Section 2.3.1.

Finally, the electrostatic free energy functional is complemented by the nonpolar free energy functional to give the total free energy functional of solvation for biomolecules at equilibrium

$$\begin{aligned} G_{\text{total}} &= \int_{\Omega} \gamma \|\nabla S(\mathbf{r})\| + pS(\mathbf{r}) + \rho_0(1 - S(\mathbf{r}))U^{\text{att}} \\ &+ S(\mathbf{r}) \left[\rho_m \phi - \frac{1}{2} \epsilon_m |\nabla \phi|^2 \right] \\ &+ (1 - S(\mathbf{r})) \left[-\frac{1}{2} \epsilon_s |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \right] d\mathbf{r}. \end{aligned} \quad (2.7)$$

Note that the polar and nonpolar parts are coupled via the characteristic function S , which is determined by the total energy optimization instead of the surface free energy optimization as done in our earlier work [20]. The above total free energy expression provides a basis for

the evaluation of the solvation free energy and a starting point for the derivation of governing equations for the solvation analysis. A similar coupling of polar and nonpolar interactions was described previously by Dzubiella and co-workers [74, 75]; however, the implementation of non-polar interactions and the representation of continuum and discrete domains differ significantly from the present work.

2.1.3 Governing equations

The solvation free energy functional is a functional in terms of characteristic function S and potential ϕ . The integral is taken over the whole space. From the physical point of view, there should exist an optimal function $S(\mathbf{r})$ and an optimal potential ϕ at the equilibrium state which optimize the total energy. Since S and ϕ can vary independently in our formulation, to optimize G_{total} , it is necessary that

$$\frac{\delta G_{\text{total}}}{\delta \phi} \Rightarrow S\rho_m + \nabla \cdot [(1-S)\epsilon_s + S\epsilon_m]\nabla\phi + (1-S) \sum_{i=1}^{N_c} n_i^0 Q_i e^{-\phi Q_i/k_B T} = 0 \quad (2.8)$$

and

$$\begin{aligned} \frac{\delta G_{\text{total}}}{\delta S} \Rightarrow & -\nabla \cdot \left(\gamma \frac{\nabla S}{\|\nabla S\|} \right) + p - \rho_0 U^{\text{att}} + \rho_m \phi - \frac{1}{2} \epsilon_m |\nabla \phi|^2 \\ & + \frac{1}{2} \epsilon_s |\nabla \phi|^2 + k_B T \sum_{i=1}^{N_c} n_i^0 (e^{-\phi Q_i/k_B T} - 1) = 0, \end{aligned} \quad (2.9)$$

where $\nabla \cdot \left(\gamma \frac{\nabla S}{\|\nabla S\|} \right)$ is a generalized Laplace-Beltrami operator, which is a generalization of the usual Laplacian operator to a smooth manifold [17, 243]. In general, γ can be a function of the position $\gamma = \gamma(\mathbf{r})$ to reflect surface hydrophobicity at different locations. However,

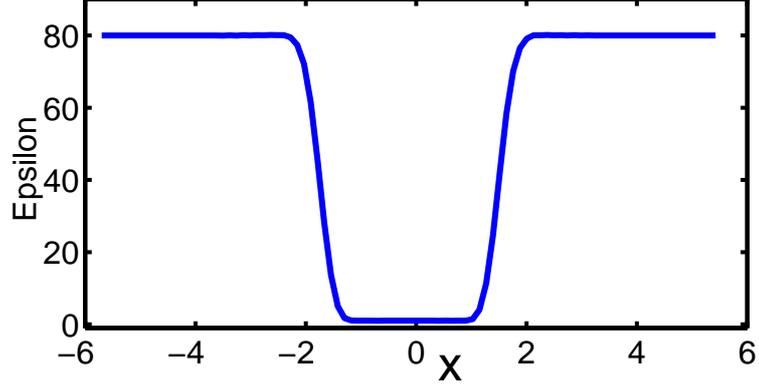


Figure 2.2: The cross line profile of $\epsilon(S)$ of a diatomic system described in Section 2.3.3. Here, we have set $\epsilon_s = 80$ and $\epsilon_m = 1$.

it is treated as a constant in our present computation. From Eq. (2.8) we result in the generalized Poisson-Boltzmann equation (GPBE)

$$-\nabla \cdot (\epsilon(S)\nabla\phi) = S\rho_m + (1 - S) \sum_{i=1}^{N_c} n_i^0 Q_i e^{-\phi Q_i/k_B T}, \quad (2.10)$$

where the dielectric function is given by

$$\epsilon(S) = (1 - S)\epsilon_s + S\epsilon_m. \quad (2.11)$$

This expression provides a smooth dielectric profile. Figure 2.2 shows the cross line of the dielectric function $\epsilon(S)$ of a diatomic system. It is seen that there is a smooth transition region for the dielectric constant to change from ϵ_s to ϵ_m . The solution procedure of Eq. (2.10) can differ very much from that of the standard PB equation, due to the smooth dielectric function. Particularly, many mathematical difficulties of solving elliptic equations with discontinuous coefficients [258, 261, 260, 254, 253] can be avoided in the present generalized Poisson-Boltzmann equation.

For a weak electrostatic potential, i.e., $\phi \ll 1$, one can linearize the generalized PB equation

$$-\nabla \cdot (\epsilon(S)\nabla\phi) + (1 - S)\bar{\kappa}^2\phi = S\rho_m, \quad (2.12)$$

where $\bar{\kappa}$ is a modified Debye-Hückel screening function describing the ion strength [108].

Furthermore, the solution of Eq. (2.9) leads to a “physical solvent-solute boundary” (S). As discussed in earlier work [20, 17, 243], the solution of this elliptic partial differential equation can be attained via a parabolic partial differential equation

$$\frac{\partial S}{\partial t} = \sqrt{\|\nabla S\|} \left[\nabla \cdot \left(\gamma \frac{\nabla S}{\|\nabla S\|} \right) + V \right], \quad (2.13)$$

where the generalized “potential” V is defined as

$$V = -p + \rho_0 U^{\text{att}} - \rho_m \phi + \frac{1}{2} \epsilon_m |\nabla \phi|^2 - \frac{1}{2} \epsilon_s |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right). \quad (2.14)$$

Note that Eq. (2.13) has the same differential operator as the mean curvature flow equation [20], except for the extra external source term. Therefore, it is a special case of the potential driven geometric flow equation proposed in our earlier work [17]. In Eq. (2.13), as $t \rightarrow \infty$, the initial profile of S evolves into a steady state solution, which solves the original Eq. (2.9).

It is interesting to see that the sharp solvent-solute interface and the standard PB equation, as well as related interface conditions, can be recovered from Eq. (2.10). For a sharp interface, S becomes a Heaviside function, having value 1 for the solute subdomain and 0 for the solvent subdomain. As such, the smooth transition region in the dielectric function disappears and the dielectric function becomes discontinuous. Then, Eq. (2.10) reduces to

the classical form of the Poisson-Boltzmann equation [108]

$$\begin{aligned} -\epsilon_m \nabla^2 \phi_m &= \rho_m & \forall \mathbf{r} \in \Omega_m \\ -\epsilon_s \nabla^2 \phi_s &= \sum_j Q_j c_j e^{(-\phi_s Q_j / k_B T)} & \forall \mathbf{r} \in \Omega_s \end{aligned}$$

with appropriate interface conditions

$$\phi_s = \phi_m, \quad \text{and} \quad \epsilon_m \nabla \phi_m \cdot \mathbf{n} = \epsilon_s \nabla \phi_s \cdot \mathbf{n} \quad \forall \mathbf{r} \in \Gamma, \quad (2.15)$$

where ϕ_m and ϕ_s represent the potential in the solute domain Ω_m and solvent domain Ω_s , respectively, Γ denotes the sharp interface, and \mathbf{n} is the normal vector of the solvent-solute sharp interface.

Note that the generalized Poisson-Boltzmann Eq. (2.10) and the potential driven geometric flow equation (2.13) are strongly coupled. Therefore, these two equations have to be solved by appropriate iterative procedures. This aspect will be discussed in Section 2.2.3.

2.2 Methods and algorithms

This section presents a variety of computational methods and algorithms for the solution of the generalized Poisson-Boltzmann equation and the generalized geometric flow equation.

2.2.1 Discretization schemes of the governing equations

We design second order finite difference schemes for governing equation derived from the free energy optimization.

2.2.1.1 The generalized Poisson-Boltzmann equation

For the solution of the generalized PB equation, the finite difference scheme is utilized in this study. The continuous dielectric definition allows us to obtain accurate solution by using the standard second order center difference scheme. Let the pixel (i, j, k) represent the position (x_i, y_j, z_k) . The discretized form of Eq. (2.12) is

$$\begin{aligned}
& \epsilon(x_i + \frac{1}{2}h, y_j, z_k)[\phi(i + 1, j, k) - \phi(i, j, k)] \\
+ & \epsilon(x_i - \frac{1}{2}h, y_j, z_k)[\phi(i - 1, j, k) - \phi(i, j, k)] \\
+ & \epsilon(x_i, y_j + \frac{1}{2}h, z_k)[\phi(i, j + 1, k) - \phi(i, j, k)] \\
+ & \epsilon(x_i, y_j - \frac{1}{2}h, z_k)[\phi(i, j - 1, k) - \phi(i, j, k)] \\
+ & \epsilon(x_i, y_j, z_k + \frac{1}{2}h)[\phi(i, j, k + 1) - \phi(i, j, k)] \\
+ & \epsilon(x_i, y_j, z_k - \frac{1}{2}h)[\phi(i, j, k - 1) - \phi(i, j, k)] \\
= & (1 - S(i, j, k))\bar{\kappa}^2\phi(i, j, k)h^2 - S(i, j, k)q(i, j, k)/h
\end{aligned} \tag{2.16}$$

where h is the grid spacing, and $q(i, j, k)$ is the fractional charge at grid point (x_i, y_j, z_k) , which is resulted from the interpolation of the charge density ρ_m . The second order interpolation (i.e., the trilinear mapping) is used to distribute charges. Thus, the discretized PB equation can be cast into the standard linear algebraic equation system of the form $AX = B$, where X is the solution of the interest, A is the discretization matrix and B is the source term associated with the continuum and discrete charges. The boundary condition is built up by the far field condition and practically obtained by the sum of potential contributions from individual atomic charges with a decay factor from the continuum charge strength κ [93]. We have explored the use of the biconjugate gradient method as the linear solver.

Matrix acceleration is discussed in a later section. The initial guess of the solution is set to 0 and the convergence tolerance is set as 10^{-6} . It is shown in the test section that the PB solver is able to deliver the designed second-order accuracy. The solution of geometry flow equation (2.13) is described in detail in Appendix A.

2.2.2 Acceleration procedures

The computational efficiency of the solution process is an important issue and can be a bottleneck for further applications of the present model. Particularly, when this model is applied to molecular dynamic simulation, the generalized PB equation and geometry flow equation are to be solved up to millions of times. Therefore, any nontrivial improvement in computational efficiency will make the present model more feasible to many practical applications in chemical and biological systems.

2.2.2.1 Precondition of the PB solver

The linear algebraic system of the discretized PB equation can be solved by two major approaches: direct methods and iterative methods. Large memory requirement prohibits direct methods to be used in the matrix resulted from large chemical and biological systems. Widely used iterative methods, including Gauss-Seidel and successive over-relaxation (SOR), work well for the generalized PB equation, but typically converge slowly. Conjugate Gradient method is quite efficient for symmetric and positively definite matrices. However, the sparse matrix A resulted from Eq. (2.16) is seven-fold banded but non-symmetric because the dielectric distribution function is not a constant and varies in the transition region. The biconjugate gradient (BiCG) method can be a good choice for

non-symmetric systems and has been adopted in a variety of our MIB schemes [258, 261, 260, 254, 253], but attentions are still to be paid in regard to the convergence issue. We have studied the application of pre-conditioners in two linear solver libraries, the SLATEC (http://people.sc.fsu.edu/~burkardt/f_src/slatec/slatec.html) and the PETSc (<http://www.mcs.anl.gov/petsc/petsc-as/>) to the solution of the PB equation [40]. It turned out that combination of stabilized biconjugate gradient method (BiCG) and the blocked Jacobi preconditioner (BJAC) from the PETSc and the combination of the orthomin method (OM) and the incomplete LU factorization preconditioner (ILU) from the SLATEC performed better compared to other tested solvers, preconditioners and their combinations [40]. In this study, we focus on the combination of the ILU and the OM from the SLATEC, which is easy to incorporate into the present iteration procedure and provides a stand-alone package, while the PETSc needs to be pre-installed before being used. In Section 2.3, we further demonstrate the improvement by the combination of pre-conditioners and the iterative linear solvers.

2.2.2.2 Initial guess of the generalized PB solution

A good initial guess is always desired for the speedup of the iterative PB solver. Normally, the initial guess can be simply set to 0 because it is complicated and computationally expensive to find good ones. However, in our iteration procedure, it is found that the electrostatic potential distribution does not change dramatically from the prior calculation due to the small adjustment in dielectric from the prior step. Therefore, it is beneficial to take the prior potential as a good guess for solving the linear system. It turns out that the generalized PB solver converges faster when the electrostatic potential from the previous iteration is used as an input.

2.2.2.3 Convergence criteria in the generalized PB solver

The convergence criterion directly influences the accuracy and CPU cost of the solution of the generalized PB equation. The smaller convergence criterion, the more accurate the solution of linear system is. However, the smaller convergence criterion requires more iterations and longer CPU time in the iterative solution process. Therefore, it is worthwhile to find a criterion which is a good compromise between the accuracy and the efficiency. Typically, a value of 10^{-6} is used in many chemical and biological applications. Later on we will numerically investigate the effect of convergence criterion on the electrostatic solvation energy, mean surface area and mean volume which are used to compute the total solvation free energy. With the 10^{-8} as a standard, we will examine the efficiency and the accuracy for several relaxed convergence criteria, such as 10^{-6} , 10^{-4} , 10^{-3} , 10^{-2} and 10^{-1} .

2.2.3 Dynamical coupling of the generalized Poisson Boltzmann and geometry flow equations

As described in Section 2.1, the present differential geometry based solvation model prescribes a procedure to set up the total free energy functional of the solvation. By the variational principle, we obtain generalized coupled PB equation (2.10) and potential driven geometric flow equation (2.10). The solution of these coupled nonlinear equations provides a “physical” dielectric profile $\epsilon(S)$ and the electrostatic potential ϕ and thereby enables the calculation of the solvation free energy. The solution of the potential driven geometric flow equation (2.10) requires the knowledge of ϕ , while the solution of the generalized PB equation (2.10) requires the input of S and $\epsilon(S)$. Therefore, in principle, the geometric flow equation needs to be solved simultaneously with the generalized PB equation until a self-consistency is reached.

In this study, we explore two self-consistent iteration procedures.

2.2.3.1 Approach I

The iteration process can be carried out by breaking up the process into an iterative sequence of two steps as follows: Starting with an initial guess of characteristic function S , one finds out the temporary electrostatic potential ϕ by solving the generalized PB equation with a given initial S . Once the electrostatic potential is obtained, the electrostatic energy can be calculated. The second step is to solve the time-dependent generalized geometric flow equation for S with a prior calculated potential ϕ . In this step, the time integration can usually reach a quasi-steady state. With the updated quasi-steady S , one can come back to the first step for the next cycle until the solvation free energy converges to within a pre-determined criteria. However, in practice, simply re-inserting S into the PB solver may diverge. Because the quasi-steady S may vary dramatically during the iteration. Note that all changes in S are concentrated around the solvent-solute boundary, as the final solution of the potential driven geometric flow equation reflects the balance between the intrinsic curvature energy and the external potential terms. A large change in S near the solvent-solute boundary in turn leads to much adjustment in the electrostatic potential which differs much in the solute and solvent regions. To avoid this problem, we adopt a simple relaxation algorithm: the characteristic function S used for the PB solver is a linear combination of the previous one S_{old} and the newly generated one S_{new}

$$S = \alpha S_{\text{new}} + (1 - \alpha) S_{\text{old}}, \quad 0 < \alpha < 1. \quad (2.17)$$

It turns out that the convergence of the generalized PB equation is guaranteed if α is small enough. In this work, α can be taken in the range from 0.1 to 0.7. The choice of α is explored later. Note that this approach may fail sometimes when the generalized geometric flow equation blows up due to a large variation in the temporary electrostatic potential. We therefore utilize a similar procedure for the electrostatic potential used in the evolution of the generalized geometric flow equation

$$\phi = \alpha' \phi_{\text{new}} + (1 - \alpha') \phi_{\text{old}}, \quad 0 < \alpha' < 1, \quad (2.18)$$

where ϕ_{old} and ϕ_{new} are previous and newly resolved electrostatic potentials, respectively. This treatment can avoid the blow-up of the generalized geometric flow solution.

2.2.3.2 Approach II

To reduce the dramatic changes in S and ϕ , we can consider a straightforward way in which solving generalized PB equation follows each time-step integration of the generalized geometry flow equation. However, this treatment makes the whole iterative procedure computationally over expensive as many more PB solution processes are required. Additionally, it may not be necessary since the change in the S from one time step to another one is so small that the change in the corresponding potential distribution should be very small. Indeed, it is practical to update electrostatic potential after a number of time steps (i.e., 10 to 100 steps) in the generalized geometry flow equation integration rather than every time step. We call the number of time integration between two ϕ updatings the number of intermittency N_{step} . This approach effectively speeds up the whole process. Additionally, the relaxation algorithm given in Eqs. (2.17) and (2.18) can also be used here to guaran-

tee the convergence. Moreover, in this approach, one better starts the iterative process by solving the S from Eq. (2.13) without the electrostatic potential term. So that the later iteration procedure can focus on the impact of electrostatic potential to the generation of the solvent-solute boundary. This treatment reduces the total iteration number and save the computational time significantly.

In fact, there is a relationship between Approach I and Approach II. When the number of the time integrations becomes larger and larger, Approach II returns to Approach I. In Section 2.3, we systematically study the difference between these two approaches. This can be done by comparing the impacts of different approaches on the resulting total solvation free energy, surface area and volume of the solute molecule. It is found that these two approaches lead to the same results. This, to some degree, indicates the reliability and validity of the proposed iteration procedures.

2.2.4 Evaluation of the solvation free energy

Once the electrostatic potential ϕ and the characteristic function S are obtained, the solvation free energy is given by

$$\Delta G = G - G_0 \tag{2.19}$$

where G_0 is the energy calculated from the homogeneous environment with $\epsilon_s = \epsilon_m = 1$ without the nonpolar energy part. Therefore, we have

$$\Delta G = G_p + G_{np} - G_0. \tag{2.20}$$

The expressions of G_p and G_{np} are taken from Eq. (2.7). Here $G_p - G_0$ can be considered as the electrostatic solvation free energy. In all calculations presented here except for salt effect calculation, mobile ions will be set to zero corresponding to a solution without salt.

Therefore we have the following approximation

$$G_p = \int_{\Omega} S(r)\rho_m\phi d\mathbf{r} - \frac{1}{2} \int_{\Omega} \epsilon(S(r))|\nabla\phi|^2 d\mathbf{r} \approx \frac{1}{2} \int_{\Omega} S(r)\rho_m\phi d\mathbf{r}. \quad (2.21)$$

Discretizing the integral yields

$$G_p = \frac{1}{2} \sum_{i=1}^{N_m} Q(\mathbf{r}_i)\phi(\mathbf{r}_i), \quad (2.22)$$

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

$$\Delta G_p = G_p - G_0 = \frac{1}{2} \sum_{i=1}^{N_m} Q(\mathbf{r}_i)(\phi(\mathbf{r}_i) - \phi_0(\mathbf{r}_i)), \quad (2.23)$$

where ϕ and ϕ_0 are electrostatic potentials in the presence of the solvent and the vacuum, respectively. Here ϕ is computed from the generalized Poisson equation (2.10) using the continuous dielectric distribution

$$-\nabla \cdot (\epsilon(S)\nabla\phi(\mathbf{r})) = S\rho_m \quad (2.24)$$

where $\epsilon(S)$ and ρ_m are the same as the ones in Eq. (2.10). The homogeneous solution ϕ_0 is computed with $\epsilon(S) = \epsilon_p$ in the whole domain. The nonpolar part, G_{np} , is computed exactly by Eq. (2.2).

2.3 Numerical test and validation

This section provides systematic validations for the computational algorithms and schemes proposed in the last two sections. We first examine the behavior of the coarea formula, then continue testing through various equation solvers, and finally check the impact of the potential term in our generalized geometric flow equation.

2.3.1 The behavior of the coarea formula

As described earlier, the coarea formula plays an important role in describing the mean surface area of an infinite family of smooth solvent-solute boundaries by a volume integral. This Eulerian formulation puts the free energy of the surface area and other free energies in an equal footing. Usually, the isovalue of the surface area in the coarea formula can be any positive real number. But for the present derivation, it is limited to be between 0 and 1 because S is defined as a characteristic function of the solute. Here, we numerically explore the behavior of the coarea formula in a bounded open set. To this end, we design some test cases as follows: Let B be a smooth function with a specific expression according to the geometry in the coarea formula, we set

$$\text{mean surface area of } \{\mathbf{x} | 0 < B(\mathbf{x}) < 1\} = \int_0^1 \int_{B^{-1} \cap \Omega} d\sigma dc = \int_{\Omega} \|\nabla B\| d\mathbf{x} \quad (2.25)$$

where Ω is a bounded open set. Therefore, the mean surface area has the same value as the volume of open set $\Omega \cap \{\mathbf{x} | 0 < B(\mathbf{x}) < 1\}$. Computationally, integrating over the norm of the gradient of B gives rise to the corresponding mean surface area. The volume integral of

a density function f is just simply approximated by

$$\int_{\Omega} f(x, y, z) d\mathbf{r} \approx \sum_{(i,j,k) \in J} f(x_i, y_j, z_k) h^3 \quad (2.26)$$

where the summation is over J , the set of points inside Ω , and (x_i, y_j, z_k) is the coordinates of grid points (i, j, k) . Table 2.1 lists the numerical results and exact values of the surface areas for the following cases

(a) A unit sphere: $\Omega = \{(x, y, z) | x^2 + y^2 + z^2 \leq 1\}$ and $B = \sqrt{x^2 + y^2 + z^2}$

(b) A cylinder: $\Omega = \{(x, y, z) | x^2 + y^2 < 1, -4 \leq z \leq 4\}$ and $B = \sqrt{x^2 + y^2}$

(c) A ellipsoid: $\Omega = \{(x, y, z) | (x/a)^2 + (y/b)^2 + (z/c)^2 \leq 1\}$ and

$B = \sqrt{(x/a)^2 + (y/b)^2 + (z/c)^2}$, where $a = 20/7, b = 25/14$, and $c = 25/14$.

It is evident that the numerical result converges to the exact value. The errors from the

Table 2.1: Areas computed from the coarea formula for bounded open sets

case	grid spacing					exact value
	0.5	0.25	0.1	0.05	0.025	
Sphere	4.00	4.00	4.15	4.17	4.18	4.189
Cylinder	22.50	23.25	24.49	24.84	25.01	25.133
Ellipsoid	37.75	37.97	38.10	38.17	38.16	38.163

cylinder are slightly larger than those from the sphere and ellipsoid because the cylinder has non-smooth edges. However, the errors are small for all cases. Therefore, we conclude that the mean value of the areas of the family of smooth solvent-solute surfaces indeed converges to the area of the corresponding sharp surface. Thus, the present definition of the mean surface area of an infinite family of smooth surfaces is an important generalization of the classic concept of the area of a sharp surface.

2.3.2 Accuracy of the generalized PB solver

In this section, we investigate the accuracy of the proposed numerical solvers. The generalized geometric flow equation (A.8) has the same differential operator as the mean curvature flow [20] except for the extra source terms. Previously, we have numerically proved that the explicit Euler algorithm delivers the reliability and convergence of the solution of geometric flow equations, and the finite central different scheme is of second order convergence in space [17]. Here, we focus on the test of the accuracy of the generalized PB solver with the proposed dielectric function $\epsilon(S)$. Although the discretization form of the second-order finite difference PB expression has been used for other continuous dielectric definitions [112], the accuracy of this approach has not been examined. Moreover, it is worthwhile to validate the generalized PB solver due to its different settings of dielectric function, i.e., the $\epsilon(S)$ profiles generated by the geometric flow equation. For this purpose, we construct a benchmark test of a simple one-ball system. We examine the convergence order and the accuracy of the finite difference scheme in solving the generalized PB equation. In particular, we consider a modified Poisson equation with a designed dielectric definition given by

$$\epsilon(\mathbf{r}) = \epsilon_1 Su(\mathbf{r}) + \epsilon_2 \cdot (1 - Su(\mathbf{r})) \quad (2.27)$$

where $\mathbf{r} = (x, y, z)$, ϵ_1 and ϵ_2 are two constants to be determined, and

$$Su(\mathbf{r}) = \begin{cases} 1 & \text{if } \sqrt{x^2 + y^2 + z^2} < a; \\ -2 \left(\frac{b - \sqrt{x^2 + y^2 + z^2}}{b - a} \right)^3 + 3 \left(\frac{b - \sqrt{x^2 + y^2 + z^2}}{b - a} \right)^2 & \text{if } a \leq \sqrt{x^2 + y^2 + z^2} \leq b; \\ 0 & \text{if } b < \sqrt{x^2 + y^2 + z^2}; \end{cases} \quad (2.28)$$

where $a < b$. Note that through the definition of ϵ and $Su(\mathbf{r})$, this designed case has the same features of the dielectric definition as that in our model. The exact solution is designed to be

$$\phi_0(\mathbf{r}) = \cos(x) \cos(y) \cos(z). \quad (2.29)$$

Then the modified equation becomes

$$\nabla \cdot (\epsilon \nabla \phi) = \epsilon_x \nabla_x \phi_0 + \epsilon_y \nabla_y \phi_0 + \epsilon_z \nabla_z \phi_0 + \epsilon \nabla^2 \phi_0 \quad (2.30)$$

where

$$\nabla^2 \phi_0(x, y, z) = -3 \cos(x) \cos(y) \cos(z)$$

$$\nabla_x \phi_0 = -\sin(x) \cos(y) \cos(z)$$

$$\nabla_y \phi_0 = -\sin(y) \cos(x) \cos(z)$$

and

$$\nabla_z \phi_0 = -\sin(z) \cos(x) \cos(y).$$

For ϵ , if $a \leq \sqrt{x^2 + y^2 + z^2} \leq b$, we have

$$\epsilon_i(\mathbf{r}) = 6(\epsilon_2 - \epsilon_1) \left(\frac{b - \sqrt{x^2 + y^2 + z^2}}{b - a} \right) \left(\frac{\sqrt{x^2 + y^2 + z^2} - a}{b - a} \right) \left(\frac{i}{(b - a)\sqrt{x^2 + y^2 + z^2}} \right) \quad (2.31)$$

where $i = x, y, z$. Otherwise, $\epsilon_x = \epsilon_y = \epsilon_z = 0$. Table 2.2 lists the computed errors under different mesh sizes with $a = 1$ and $b = 3$. The standard absolute norm error measurement L_∞ is employed. Here ϵ_1 is fixed to be 1 and ϵ_2 is taken to be 80 or 10. The second order convergence in space is observed for the scheme. Furthermore, it is found that a large ϵ_2 may slightly deteriorate the convergence.

Table 2.2: Errors and convergence orders for the generalized PB solver ($\epsilon_1 = 1$)

spacing	$\epsilon_2 = 80$		$\epsilon_2 = 10$	
	L_∞	order	L_∞	order
1	0.22		0.13	
0.5	8.13*10-2	1.65	3.18*10-2	2.02
0.25	2.06*10-2	1.99	7.97*10-3	2.00
0.125	5.44*10-3	1.94	1.98*10-3	2.01

2.3.3 Convergence of boundary profile and dielectric function

In the present model, the characteristic function S defines the solvent-solute boundary. Consequently, it can significantly affect the solution of the generalized Poisson-Boltzmann equation, the surface area and volume, and thus, plays a key role in the solvation free energy calculation. To illustrate the evolution and the convergence of the generalized geometric

flow equation and corresponding S function, we solve Eq. (2.13) without the electrostatic potential term in this test. However, the electrostatic solvation free energy at a given time can still be calculated. The expression of attractive interaction U_i^{att} needs to be given explicitly in order to solve the geometric flow equation. Here we consider the 6-12 Lennard-Jones (L-J) pair potential to model U_i^{att} (see the description in Appendix A). All the calculations in this work are carried out by using the WCA decomposition.

To illustrate the S profile and evolution, we consider a diatomic system with the van der Waals radius 2.2\AA and coordinates $(x, y, z) = (-3, 0, 0)$ and $(1.4, 0, 0)$. The spacing and time stepping are chosen as $h = 0.25\text{\AA}$ and $\tau = h^2/4.5$, respectively. The solvent probe radius is set to 2\AA , which is used for the initial value setting and constraint construction. In fact a much small solvent probe radius can ensure the correct surface topology [20]. The computational domain is set to $[-8.70, 7.05] \times [-5.7, 5.55] \times [-5.7, 5.55]$. Thus, the size of computational system is $64 \times 46 \times 46$. The L-J parameters are set as follows: σ_i is taken from atomic radius and σ_s is chosen to be 0.65\AA . We set well depth $\epsilon_i = 0.039$ kcal/mol and bulk density coefficient $\rho_0/\gamma = 2$, where, $\gamma = 1/15$ kcal/(mol \AA^2). To compute the electrostatic solvation free energy during the evolution of solvent-solute boundary, 1 unit charge is set on the center of each atom. We choose the dielectric constants $\epsilon_m = 1$ and $\epsilon_s = 80$, respectively. We set pressure coefficient $p/\gamma = 0.2$. A different γ is used for real systems.

The evolution process of diatomic solvent-solute boundary is depicted through a group of cross section profiles of the S function in Figure 2.3 where the values of S from a set of points of $(x, y, 0.05)$ are described. The cross sections start with a relatively fat-shaped interface which reflects the solvent accessible density. Here, $S = 1$ inside the molecular domain and $S = 0$ in the solvent domain. Then the solvent-solute boundary is driven inward by

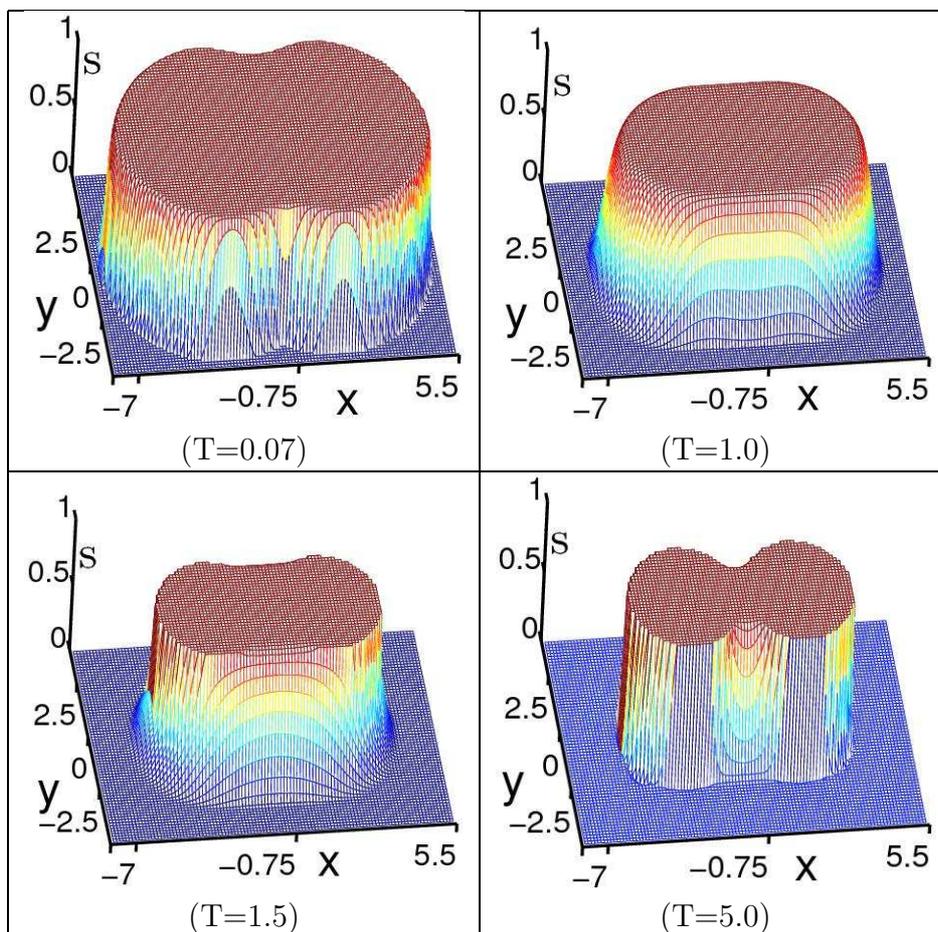


Figure 2.3: The evolutionary profiles of the S function at cross section $(x, y, 0.05)$ in a diatomic system plotted from four intermediate states.

intrinsic geometric curvature effect and external potential. At the same time, there appears a transition region between the solvent and the solute. Finally a convergence is reached with a balance among intrinsic geometric curvature effect, different potentials and enforced constraints. To have a clear idea about the distribution feature of the S function, we draw a cross line from the cross section graph at $T = 5$ along $x = -0.75$ which are shown in Figure 2.1, where the functions of S and $1 - S$ are described together. It can be seen that the S function in the transition region is rather smooth. Once the S function is determined, the dielectric function $\epsilon(S)$ is calculated according to Eq. (2.11). Here the dielectric function $\epsilon(S)$ corresponding to the S function in Fig (2.1) is also exhibited in Figure 2.2. It has a pattern similar to $1 - S$ but with different values. It is important to note that the dielectric function $\epsilon(S)$ is also very smooth at the solute-solvent boundary. That is why the classical finite difference scheme can be applied to solve the generalized PB equation without reducing the accuracy of the solution.

Once the solution of the generalized PB equation is computed, the electrostatic solvation free energy can be calculated. Therefore, the time history of the free energy along with the evolution of solvent-solute boundary can be recorded. To illustrate the convergence pattern of the solvation free energy, we compute the electrostatic solvation free energies in intermediate states during the time evolution. The results are shown in Figure 2.4. In order to show evolution histories of the surface area, volume and solvation free energy together in one plot, we plot two linear functions $F(\text{Volume})$ and $J(\text{Area})$ which have the same time dependence as the volume and the surface area, respectively. Here T denotes the time span and $N = \frac{T}{\tau}$ represents the number of computational steps in the generalized geometric flow solver. It is found that the solvation free energy decreases with respect to the time evolution, which

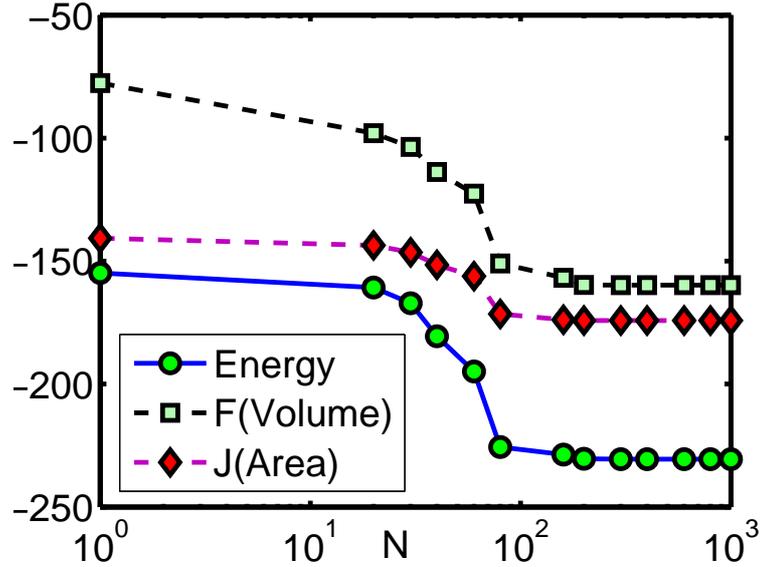


Figure 2.4: The time evolution histories of the electrostatic solvation free energy, $F(\text{Volume})$ and $J(\text{Area})$, where $F(\text{Volume}) = \text{volume}/5 - 180$ and $J(\text{Area}) = (\text{surface area})/5 - 200$.

is consistent with our theoretical formulation. It is observed that the solution of our model converges to a steady state in terms of volume (\AA^3), area (\AA^2) and electrostatic solvation free energy (kcal/mol). Moreover, to obtain the results at the steady state, $N = 200$ or $T = 3.5$ is large enough to be taken as the stopping time in our geometry flow solver for this system. Normally, it takes a longer evolution time for a large system to set down to the steady state. The total integration time could be considerably shortened had a small probe radius been used [20].

2.3.4 Consistency of iteration procedures

If the electrostatic potential effect is taken into account during the solvent-solute boundary evolution, the iteration procedure has to be used to update the electrostatic potential repeatedly. As described earlier, there are two possible iterative approaches which can be explored to solve the coupling system, in which the simple relaxation algorithm guarantees

the convergence of the algorithm. The question is whether these two approaches lead to the same outcome.

Table 2.3: Comparison between two iteration approaches

	2 atoms		Gly	
	Approach I	Approach II	Approach I	Approach II
Energy	-231.18	-231.18	-12.44	-12.44
Surface area	128.67	128.67	271.91	272.02
Volume	100.72	100.83	287.85	287.93

To study the consistency between these two approaches, the above mentioned diatomic system is employed as well as a small molecule called glycerol triacetate (Gly) from a set of 17 test compounds whose detailed information is given in Section 2.4. The self-consistent iteration is performed until the electrostatic solvation free energy converges to within 0.01 kcal/mol. The electrostatic solvation free energy(kcal/mol), surface area (\AA^2) and volume(\AA^3) resulted from these two different methods are compared. The results are shown in Table 2.3. Here we take $\alpha = 0.5$ for both methods. The electrostatic potential ϕ is updated in every 15 steps of the generalized geometric flow integrations in Approach II. It is evident that the results from these two approaches are almost the same. Therefore, they can be alternatives for each other at least in small molecular systems. But for large protein systems, as we mentioned, it is better to use the second approach to avoid the possible blowup in the generalized geometry flow caused by unpredictable large changes in the temporary electrostatic potential. Thus, in the following tests and applications, the second method is applied except specified.

In Approach II, the relaxation factor α and the number of intermittency N_{step} need to be determined. We are interested in knowing whether the relaxation factor α plays a role in the final result. Similarly, it is important to know whether the N_{step} makes a difference in

the converged result. We address these issues by examining the effects of α and N_{step} on the electrostatic solvation free energy, surface area and enclosed volume. The above mentioned diatomic system is used here again.

Table 2.4: Effect of relaxation factor α on final results

α	0.1	0.2	0.5	0.7	0.8
Energy	-231.26	-231.18	-231.18	-231.18	divergence
Surface area	100.73	100.94	100.83	100.71	
Volume	128.65	128.62	128.67	128.71	

It is known that a stable α value is between 0 and 1 but can not be very close to 1. We consider a number of α values in the diatomic system while keeping other settings fixed. Table 2.4 shows the electrostatic solvation free energy, surface area and volume for $\alpha = 0.1, 0.2, 0.5$ and 0.7. It is found that the procedure diverges when $\alpha \geq 0.8$. However, convergence is achieved as long as the relaxation factor α is small enough. Once the convergence is achieved there is no much difference in the final outcome. We therefore take $\alpha = 0.5$ in the following tests and applications.

To study the effect of the number of intermittency, we take $N_{\text{step}} = 5, 10, 15, 40$ and 100, while fixing $\alpha = 0.5$ and other settings. The results are listed in Table 2.5. It is seen that all values obtained from different number of intermittency are very close to each other. However, a numerically too large or too small N_{step} is not preferable. If N_{step} is too large, Approach II goes back to the first one. If N_{step} is too small, the iterative process may stop too early because the perturbation in each iteration is so weak that the convergence criteria is satisfied unexpectedly sometimes. In addition, small step number makes the whole process computationally expensive.

Table 2.5: Effect of the number of intermittency in Approach II

N_{step}	100	40	15	10	5
Energy	-231.18	-231.19	-231.18	-231.17	-231.11
Surface area	100.83	100.83	100.83	100.89	101.07
Volume	128.63	128.63	128.67	128.48	128.63

2.3.5 Efficiency of the accelerated iteration procedure

We study the efficiency of the accelerated self-consistent iteration in this section. At the beginning we analyze the CPU time usage based on an original combination of methods: Biconjugate Gradient (BiCG) method for the generalized PB solver and a widely used explicit scheme for the generalized geometric flow (GGF) solver. In addition, as commonly used in the linear system of the PB equation, we take 10^{-6} as the initial convergence criteria and set the first guess of the electrostatic potential in each generalized PB run to be 0. The above mentioned diatomic and Gly systems will be utilized through this efficiency test. Table 2.6 lists the breakup of time spending in the different parts of the self-consistent iteration procedure. It is seen that for these two systems the major computation cost lies in the routines of the generalized PB solver and the generalized geometric flow solver (more than 90%). Therefore, the total time will be dramatically reduced when efficient accelerations are achieved in the generalized PB solver and generalized geometric flow solver. Note that all of the computations are performed on a SGI Altix 350 workstation with a 1.4 GHz Itanium processor and 4 GB memory. Additionally, we explore the improvement made to the generalized PB solver, the generalized geometric flow solver, and consequently to the total time cost. First, we combine an appropriate preconditioner with the iterative solver of the linear system. Additionally, we make use of the prior electrostatic potential as a first guess for the next PB run. Moreover, we obtain the approximations through the relaxation of the

Table 2.6: CPU time analysis from original schemes

	2 atoms		Gly	
	Time(s)	%	Time(s)	%
Total	23.95		58.4	
GF	4.95	21	11.31	19
PB	18.25	76	45.24	77
Other	0.75	3	2.03	4

convergence criteria of the linear solver. Finally, we employ the ADI scheme to integrate the generalized geometric flow equation.

First of all, we do the following improvement: Take the prior potential solution as the first guess of each run of the linear solver, then replace BiCG scheme with a combination of the preconditioner and the iterative solver (ILU/OM), while keeping other settings unchanged.

Table 2.7: Speedup from adjustment of initial guess and preconditioner in PB solver

h	size	BiCG	BiCG ¹	ILU/OM	Speedup
1	$17 \times 12 \times 12$	0.0557(252)	0.0322(152)	0.0371(50)	1.50
0.5	$32 \times 23 \times 23$	0.775(419)	0.467(248)	0.420(82)	1.85
0.25	$64 \times 46 \times 46$	17.676(841)	10.410(490)	6.947 (166)	2.54
0.125	$127 \times 92 \times 92$	525.74(2771)	263.11(1371)	130.76 (410)	4.02

Table 2.7 gives the total computational costs of the generalized PB solver in the diatomic system as well as the total iteration numbers which are inside the parenthesis. The third column lists the time spending for original schemes, the fourth one makes use of prior potential as a first guess and the fifth one records the time spending from the usage of the preconditioner and the new first guess setting. It is seen that the gain of speedup is related to the size of system: The larger size is the system, the more acceleration is achieved. For a $127 \times 92 \times 92$ system, combination of the above two implementations can obtain a speedup up to a factor of 4, while a single adjustment does not give much impressive improvement.

It is also found that although the total iteration number reduces dramatically by adding the preconditioner, the total computational cost is reduced with a much smaller factor. The reason is that the PB solver with a preconditioner takes more time in each step.

Next, we study the impact of the convergence criteria to the electrostatic solvation free energy of the diatomic system. Table 2.8 summarizes the calculated electrostatic solvation free energies and total time cost of the PB solver under different convergence criteria. It indicates that 10^{-4} is good enough to deliver accurate results. In fact, 10^{-2} is still fine but 10^{-1} is clearly unacceptable. In this study, we take 10^{-4} as the convergence criteria of the linear system in the following calculations except specified. Because it is able to save much time compared to 10^{-6} while at the same time maintains the accuracy to a satisfied level. In practical application, one might use 10^{-2} . A further reduction in computational time is possible if one sets the probe radius to $r_p = 0.25r_{\text{vdW}}$, where r_{vdW} is the van der Waals radius [20].

Table 2.8: Influence of convergence criteria on electrostatic solvation free energy and computational time for the diatomic system

Criteria	10^{-8}	10^{-6}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
Energy (kcal/mol)	-231.17	-231.17	-231.19	-231.28	-231.07	-239.80
PB Time (s)	10.44	6.95	4.17	3.09	2.05	0.90

Finally, we implement the ADI scheme in the generalized geometric flow solver. Thus we can use a much larger time increment than that used in an explicit scheme without the stability concern. For example, if grid spacing is $h = 0.25$, the time step size can be taken as large as 0.2 for the ADI scheme to be a good balance between accuracy and efficiency, while it has to be less than 0.02 in the explicit scheme. The acceleration of the generalized geometric flow solver can be found in Table 2.9, which is obtained by applying all of the

speedup strategies we have discussed to the diatomic and Gly systems. This table shows all the time spending for major routines in the iterative process before and after the acceleration. It indicates that speedup in the PB solver can reach a factor of 4 or 5. However, the speedup in the total time is not as impressive as in the PB solver. It is about a factor of 3 in the Gly system and about a factor of 2 in the diatomic system. The reason is that the acceleration in the generalized geometric flow equation through the ADI can not have the same speedup factor as that of the PB solver. The electrostatic solvation free energies are also given in the table for a comparison before and after the speedup. Little difference in energies is observed due to varying schemes and the approximation.

Table 2.9: Comparison of CPU time (s) in the iteration procedures with and without accelerations

	2 atoms			Gly		
	without	with	Speedup	without	with	Speedup
Total	23.95	8.87	2.70	58.40	16.28	3.59
GF	4.95	3.67	1.35	11.13	5.57	2.00
PB	18.25	4.45	4.10	45.24	8.71	5.19
Other	0.75	0.75		2.03	2.00	
Energy	-231.17	-231.18		-12.44	-12.44	

2.3.6 Impact of potentials in the geometric flow equation

The potential source terms in the generalized geometric flow equation include pressure, long-ranged attractive dispersion interaction and electrostatic potential. The solution (S) of the generalized geometric flow reflects a balance between the intrinsic geometric curvature effect and several external potentials at the equilibrium. In this section, we illustrate the impact of involved potentials to the characteristic function S , and consequently to the solvation free energy. Although in our model there is no explicit surface definition to be demonstrated, the

impacts of these potentials can be reflected by volume, area and the electrostatic solvation free energy. In particular, if the flow is driven inward by a potential, the volume should become smaller, and an outward driving makes the volume larger. These are true at least during the early stage of the solvent-solute boundary evolution. In fact, they are not true for the system near the equilibrium. The present study is carried out through two proteins (PDB ID 1ajj and 1fca) from protein data bank (PDB). Their detailed coordinates and parameters are given in the application section of 22 proteins. Without any potential term, the mean curvature flow equation is driven purely by intrinsic geometric curvature effect, which leads to the minimal molecular surface (MMS) [20]. With the MMS as a reference, each time we use one additional potential term in Eq. (2.13) to produce a new characteristic function S which will be used in the PB solver to calculate the electrostatic potential. In other words, starting with the MMS, we attain the different characteristic functions S with either pressure, attractive non-polar potential, or electrostatic potential separately. Only when electrostatic potential term is taken into account, is it needed to run the self-consistent iteration process for the solution of the coupled system. Table 2.10 gives the calculated volume and electrostatic solvation free energy under each potential term. We also calculate the solvation free energy when all the potentials are turned on. It is seen that all the potentials involved here drive the flow inward so that there are more solvent components between or around two spherical solutes when just an individual potential is turned on. This is consistent with the experimental observations [236] and our previous studies [17].

Table 2.10: Effects of potentials on the solvent-solute boundary

	1ajj		1fca	
	Volume	Energy	Volume	Energy
MMS	6601.9	-975.6	9345.7	-1082.2
Pressure	6195.1	-1032.1	8866.6	-1123.2
Attractive	5533.8	-1139.3	8107.2	-1192.8
Electrostatic	6585.8	-1061.4	9329.3	-1112.4
Total potential	5381.6	-1165.6	7886.8	-1211.9

2.4 Applications

We consider two types of problems in this section. First, we apply our new approach to a set of 17 small molecules. Then, some protein examples are studied. The Dirichlet boundary condition is used for both the generalized Poisson-Boltzmann equation and the generalized geometric flow equation as in our previous calculations [259, 252, 93, 20]

2.4.1 Set of 17 test molecules

We apply our optimized surface model (OSM) of solvation to compute the solvation free energies of a set of 17 small compounds. This test set was studied by Nicholls et al [160] using a number of approaches, including quantum mechanics, PB theory etc. An important aspect about this test set is that experimental data are available. Therefore, solvation free energies predicted from our new model can be compared with both experimental values and other numerical results. Moreover, these compounds are considered as a challenging test set for computational methods because the existence of polyfunctional or interacting polar groups, which lead to strong solvent-solute interactions.

In our calculation, we set the dielectric constants $\epsilon_m = 1$ and $\epsilon_s = 80$. We use γ as fitting parameter, and its initial value is set to $\gamma = 1/15$ kcal/(molÅ²) to compute other

Table 2.11: Comparison of free energies (kcal/mol) for 17 compounds

Compound	G_{np}	ΔG_p	ΔG	Exptl	Error
glycerol triacetate	2.27	-12.44	-10.16	-8.84	-1.32
benzyl bromide	1.40	-4.89	-3.49	-2.38	-1.11
benzyl chloride	1.35	-5.02	-3.68	-1.93	-1.75
m-bis(trifluoromethyl)benzene	2.22	-3.22	-1.00	1.07	-2.07
N,N-dimethyl-p-methoxybenzamide	1.96	-9.20	-7.24	-11.01	3.77
N,N-4-trimethylbenzamide	1.86	-7.67	-5.81	-9.76	3.95
bis-2-chloroethyl ether	1.45	-4.22	-2.77	-4.23	1.46
1,1-diacetoxyethane	1.65	-8.24	-6.59	-4.97	-1.62
1,1-diethoxyethane	1.52	-4.40	-2.88	-3.28	0.40
1,4-dioxane	1.01	-5.65	-4.64	-5.05	0.41
diethyl propanedioate	1.82	-7.85	-6.03	-6.00	-0.03
dimethoxymethane	1.03	-4.52	-3.50	-2.93	-0.57
ethylene glycol diacetate	1.59	-8.43	-6.84	-6.34	0.50
1,2-diethoxyethane	1.55	-4.31	-2.76	-3.54	0.78
diethyl sulfide	1.22	-2.39	-1.17	-1.43	0.26
phenyl formate	1.37	-7.84	-6.48	-4.08	-2.40
imidazole	0.82	-11.27	-10.45	-9.81	-0.64

γ -dependent parameters, see Eq. (2.13). We choose $\rho_0/\gamma = 2$ by comparing the bulk density 0.033\AA^{-3} and the possible γ value. For micro-molecular systems, pressure p can be very small and sometimes is neglected in the calculation [48]. But here we still take it into account and set p/γ to 0.2. Note that in the numerical simulation, all ratio parameters here are treated as dimensionless. For L-J parameters, σ_s is chosen to be 0.65\AA as a good fitting solvent radius and σ_i is the solute atomic radii [235]. Note that due to the continuum representation of solvent in our model, the 6-12 Lennard Jones potential formula (A.2) differs from the standard version — the distance used in our formula is no longer the distance between the centers of solute atoms and the centers of solvent atoms but the distance between a specific position in the solvent area and the centers of solute atoms. This should make the setting of well depth ϵ_i different from the ones taken from AMBER

or OPLS force fields. However, the performance of the L-J potential should be similar, i.e., the value of the L-J potential in the solvent caused by a solute atom only depends on the distance from the center of the atom. It implies that the value of L-J potential caused by a solute atom should be a constant on the van der Waals surface of the atom. In other words, $\epsilon_i \left[\left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^{12} - 2 \left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^6 \right] = D_i$ if \mathbf{r} is on the vdW surface of the atom. Here the constant D_i should have different values for various types of atoms. For simplicity we use a uniform constant D to determine the value of ϵ_i given σ_s and σ_i . In the present calculation, we pick 1.0 for D and the WCA expression is chosen as the attractive van der Waals potential. We choose grid spacing $h = 0.25\text{\AA}$ and time stepping $\tau = h^2/4.5$. Here, γ (kcal/(mol \AA^2)) serves as a fitting parameter, and its final value is 0.0065 kcal/(mol \AA^2).

Structure and charge information of the 17 compounds are adopted from those of Nicholls et al [160] and can be obtained from the supporting information of their paper. In particular, charges are taken from the OpenEye-AM1-BCC v1 parameters [115]. Atomic coordinates and radii are based on their new parametrization called ZAP-9 in which certain types of radii are adjusted by them from Bondi radii to improve the agreement with experimental free energy. With these structures and charges parameters, the root mean square error (RMS) obtained in their paper is 1.71 ± 0.05 kcal/mol via the explicit solvent model. And the smallest RMS error of their single - conformer Poisson-Boltzmann approach is 1.87 ± 0.03 kcal/mol [160]. Such a large RMS error indicates the challenge of this test set. Usually, different surface definitions in implicit solvent models should have their own optimal radii set. In particular, a continuous dielectric definition based model is supposed to have radii set with larger values than those of a discontinuous dielectric model. Otherwise, the calculated free energy does not give a good fitting to experimental data [217]. This also occurs in the

present model. Therefore, we multiply the radii from ZAP-9 by a common factor 1.1. It turns out this treatment leads to a good agreement with experimental data in terms of electrostatic solvation free energies and total solvation free energies. The results are summarized in Table 2.11, which gives a comparison between calculated and experimental values of solvation free energies of 17 compounds. RMS error of the present model is 1.76 kcal/mol which is similar to that of Nicholls et. al, i.e., 1.87 kcal/mol. This RMS error is competitive to that of the explicit solvent approach (1.71 ± 0.05 kcal/mol) under the same charge and structure parameters set [160]. This may be credited to the more satisfactory nonpolar terms and the enforcement of the potential driven geometric flow. Here, as expected, major errors are from the calculation of benzamides which are between 3.5 and 4.0 kcal/mol, see Figure 2.5. Without these benzamide compounds, the RMS error drops from 1.76 kcal/mol to 1.24 kcal/mol. This problem with benzamides is likely due to radius adjustment for the carbonyl oxygens and tertiary nitrogens in ZAP 9 under the OpenEye-AM1-BCC v1 charges [160]. In other words, these large errors from benzamides can not be avoided if both OpenEye-AM1-BCC v1 charge and corresponding optimized ZAP 9 radii are used in PB approaches. Based on these considerations, one possible approach for improvement is to create a new charge set more appropriate for the PB approach with the same ZAP radii. It may be realized by introducing quantum mechanical corrections to our model to take care of charge density. However, this aspect will be investigated in next chapter 3.

2.4.2 Solvation free energy of proteins

Validation by using a set of 17 molecules has shown that proposed differential geometry based solvation model works well for the energy prediction of small compounds. Since small

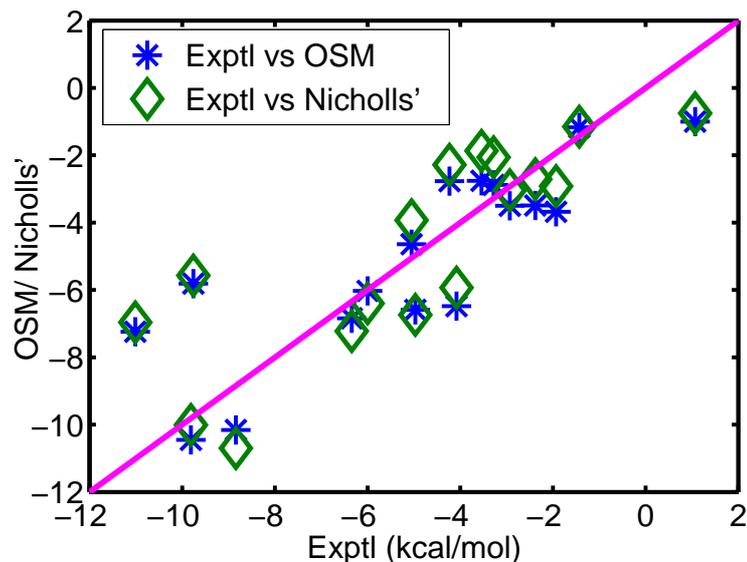


Figure 2.5: Correlation between experimental data and the present optimized surface model (OSM)(also results from Nicholls') in electrostatic solvation free energies of 17 compounds.

molecules are accessible to more accurate computational means, such as quantum mechanical calculations, one of the main purposes of developing the present optimized surface model (OSM) is to attack relatively large macromolecules. To this end, we consider a test set of proteins employed by Mei et al [152]. For this set, the total number of residues ranges from 21 to 275. The initial structures of all proteins are taken from the protein data bank (PDB). The hydrogen atoms, which are typically missing from the X-ray data, are added to the structures to obtain full all-atom models with optimized hydrogen bondings. Partial charges at atomic sites and atomic van der Waals radii in angstroms are assigned from the CHARMM27 force field [146]. All of these operations, i.e, the transformation from PDB files to PQR files, can be easily done with a software PDB2PQR. Parameters of the present calculation are set in the same way as those for 17 compounds except for $N_{\text{step}} = 2$. Similar to the treatment of the 17-compound set, the radii from the CHARMM force field need to be multiplied by a common factor of 1.1. Our results are summarized in Table 2.12.

For a comparison, The results of Mei et al are listed in Table 2.12 as well. Their results obtained from the molecular fractionation with conjugate caps and conductor-like polarizable continuum model (MFCC-CPCM). This is an approximate quantum approach that divides the macromolecule into fragments, such that the quantum calculations at HF/6-31 G level and B3LYP/6-31 G level can be applied. The solvation effect is estimated vis the polarizable continuum method with the classic molecular surface [152]. It is seen from the table that there are relatively large deviations, up to 28%, between results obtained by the present OSM and those of the MFCC-CPCM. These derivations might due to the different methodologies, computational environments and structures. In fact, the results from two different quantum basis sets have up to 10% deviation for protein Amyloid. Another deviation between results of two quantum basis sets is about 5% for the protein BPTI.

Table 2.12: Comparison of electrostatic solvation free energies (kcal/mol) obtained from the MFCC-CPCM, the present model (OSM) and MIBPB.

Protein	PDBID	No.of residues	ΔG_p (kcal/mol)		
			MIBPB-III	MFCC-CPCM [152]	OSM
RP71955	1RPB	21	-184.68	-267.60	-192.23
Amyloid	1AMC	28	-861.65	-886.01(-798.72)	-852.68
Crambin	1CBN	46	-303.80	-361.52	-304.84
BPTI	1BPI	58	-1301.9	-1332.71(-1263.52)	-1281.19
Calbindin	1CDN	75	-2188.96	-2259.62	-2195.42
Ubiquitin	1UBQ	76	-1170.61	-997.02	-1148.81
Lysozyme	2BLX	129	-1913.40	-1887.71	-1898.07
Subtilisin	1SBT	275	-1896.5	-2062.2	-2001.4

As the largest deviation between the results from the proposed OSM and that of the MFCC-CPCM is quite large, we consider another independent approach, the MIBPB [259, 252, 93], to evaluate the present method. A specific MIBPB code, the MIBPB-III which has the treatment of geometric and charge singularities [93], is employed in our calculations.

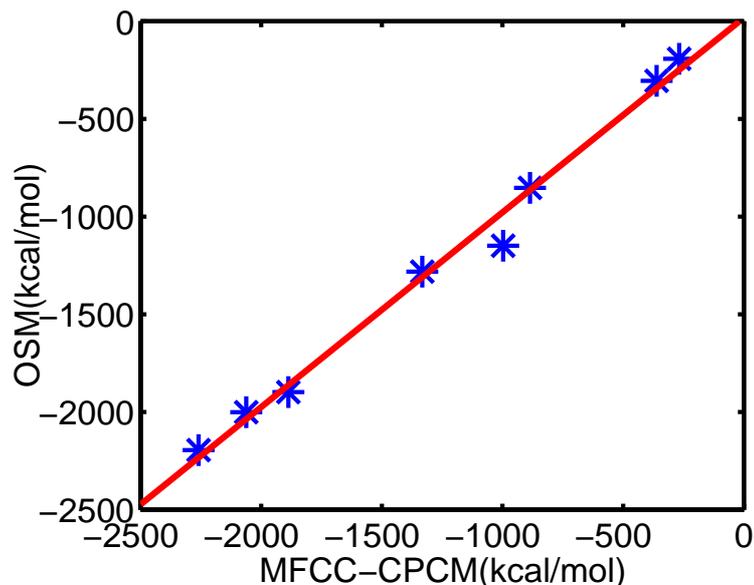


Figure 2.6: Correlation between MFCC-CPCM [152] and the present optimized surface model (OSM) in electrostatic solvation free energies of 8 proteins.

MIBPB-III has been intensively calibrated in the past and is the only known second accurate method for solving the Poisson-Boltzmann equation with both molecular surfaces and partial charges represented by the Dirac delta functions. To deliver such an accuracy, the MIBPB-III has built in the MIB scheme [261, 254] and Dirichlet to Neumann mapping [93]. Similar to the present approach, the structural data of MIBPB-III is also prepared with the PDB2PQR software. As such, we can eliminate the difference due to the different treatment of initial data. However, the MIPPB utilizes the classic PB equation and the molecular surface, while the present method has a generalized PB equation, and an optimized smooth surface. It is seen from Table 2.12 that solvation energy results from the present OSM and from the MIBPB have an excellent agreement on most proteins except for Subtilisin. For this protein, the difference of energies from two methods is about 5%.

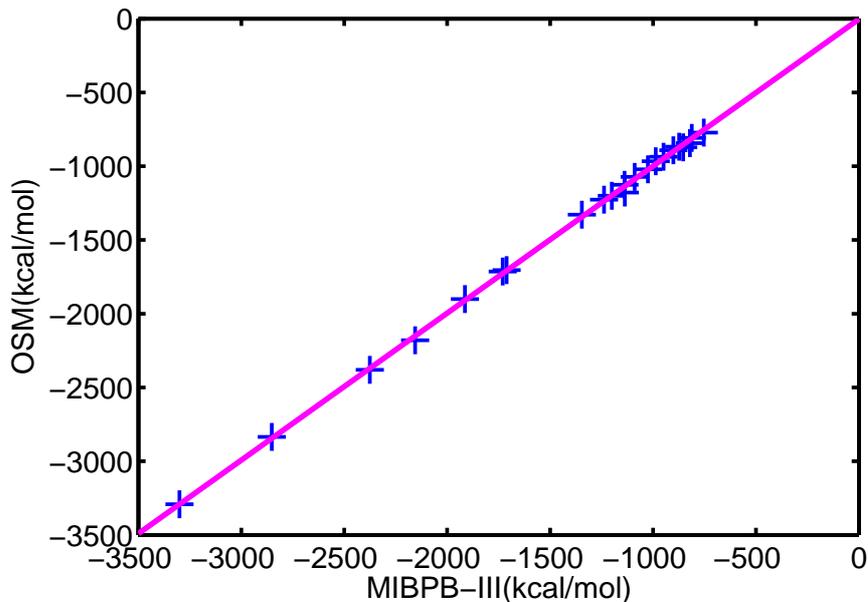


Figure 2.7: Correlation between MIBPB-III and the present model (OSM) in electrostatic solvation free energies of 22 proteins

2.4.3 Twenty two proteins

Encouraged by the good consistency of the proposed method and the MIBPB-III, we further compare these approaches by a larger set of protein molecules — twenty two proteins that have been frequently used in previous studies [83, 252, 93, 17]. The implementation of these two methods is the same as that described in the last section.

Table 2.13 shows the results from the present continuous dielectric model, denoted as “Radii1” in the table, and those by MIBPB-III. It turns out that electrostatic solvation energies obtained via our minimization process are very close to those based on the MIBPB-III. This can also be seen through Figure 2.7 which shows that the results between them are quite linearly correlated. The correlation coefficient is 0.999.

It is still interesting to understand how important it is to use a slightly enlarged radius in smooth surface models [217]. To this end, we carry out the present calculations by using

Table 2.13: Electrostatic solvation free energies for 22 proteins

PDB-ID	No. of atoms	ΔG_p (kcal/mol)		
		MIBPB-III	Radii1	Radii0
1ajj	519	-1137.2	-1178.5	-1362.6
1bbl	576	-986.8	-965.94	-1158.7
1bor	832	-853.7	-871.4	-1066.5
1fca	729	-1200.1	-1200.6	-1340.9
1frd	1478	-2852.2	-2844.8	-3173.4
1fxd	824	-3299.8	-3291.9	-3496.9
1hpt	858	-811.6	-808.2	-1039.1
1mbg	903	-1346.1	-1328.2	-1535.4
1neq	1187	-1730.1	-1713.9	-2049.3
1ptq	795	-873.1	-866.2	-1064.5
1r69	997	-1089.5	-1072.7	-1294.0
1sh1	702	-753.3	-771.8	-973.8
1svr	1435	-1711.2	-1704.6	-2073.7
1uxc	809	-1138.7	-1125.7	-1350.9
1vii	596	-901.5	-892.0	-1052.1
2erl	573	-948.8	-935.8	-1067.3
2pde	667	-820.9	-843.0	-1049.3
451c	1216	-1024.6	-1020.6	-1291.8
1a2s	1272	-1913.5	-1900.3	-2155.0
1a7m	2809	-2155.5	-2179.8	-2666.1
1a63	2065	-2373.5	-2380.5	-2912.0
1vjw	828	-1237.9	-1226.6	-1411.4

the original CHARMM22 van der Waals radii, denoted as “Radii0”. This result is also listed in Table 2.13. It is seen that results from the original CHARMM22 van der Waals radii can have over 20% deviations from those of “Radii1”. This helps to come to a conclusion that for continuous dielectric models, it is necessary to enlarge atomic radii obtained from widely used force fields. Otherwise, the results will be inconsistent with those of other analysis. This is in agreement with the observation in the literature [217]. The necessity of using larger radii is also shown clearly in Fig 2.8 by the differences of electrostatic solvation free energies obtained from the MIBPB-III and the present calculations with original radii (Radii0) or

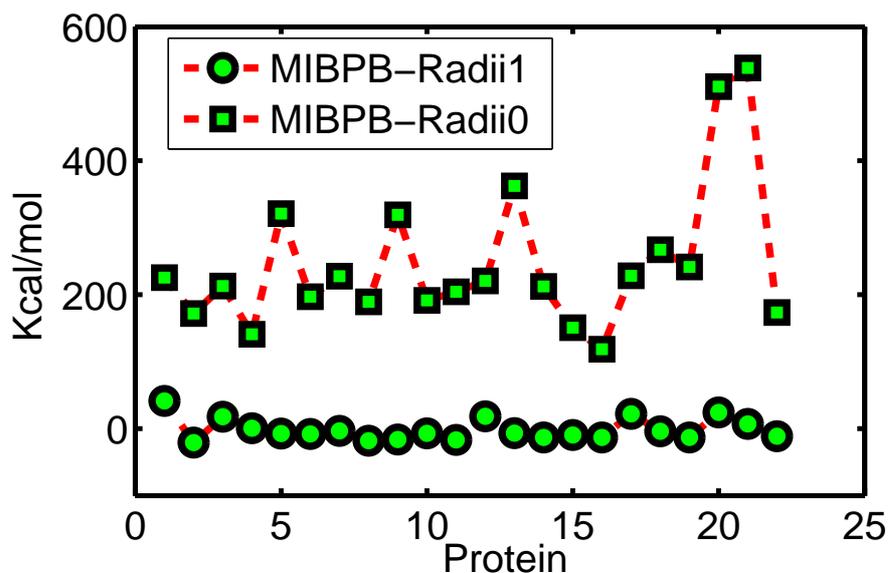


Figure 2.8: Differences between electrostatic solvation free energies obtained from the MIBPB and the present model with original radii (Radii0) or enlarged radii (Radii1).

enlarged radii (Radii1).

Additionally, it is useful to demonstrate that the electrostatic potential function computed in the present OSM can be illustrated at arbitrary isosurface of the characteristic function S . This is done by first computing a sharp surface at a given S value, then projecting the ϕ value on the isosurface of a given S value. Figure 2.9 shows three plots of the electrostatic potentials at $S = 0.25, 0.5$ and 0.75 . A comparison of these potentials indicates

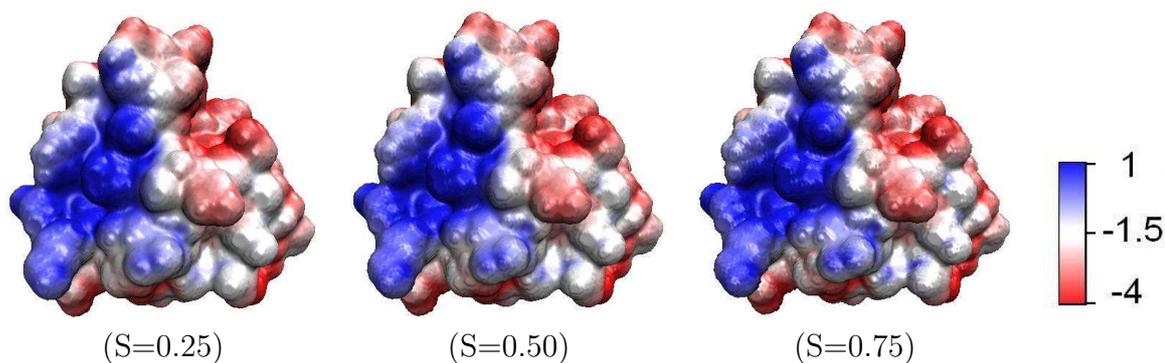


Figure 2.9: Surface potential display of one protein (PDBID: 1frd) at different isosurfaces

the fast/slow electrostatic potential changing regions in the solvent-solute boundary. These regions are also interactive regions in the protein-protein or protein-ligand interactions.

Finally, it remains an important issue to further improve the computational efficiency, although systematical efforts have been made in this work to reduce CPU cost. Since the coupled generalized PB and geometry flow equations are needed to evolve self-consistently to reach the steady state, it takes more CPU time for the present method to calculate the total free energy than some existing approaches that compute the polar and nonpolar energies separately.

2.5 Chapter conclusions

This chapter presents a novel differential geometry based solvation model. A crucial concept in the present model is the characteristic function or the description function of solute molecules which is one inside the solute domain and zero inside the solvent. Near the solvent-solute boundary, the characteristic function gradually changes from one to zero over a region of transition. The exact position and width of the transition region are determined by a variational framework, which is formulated based on the total solvation free energy. As a key ingredient of the present framework, the total energy encompasses coupled polar and nonpolar contributions. The polar solvation free energy functional is described by the electrostatic theory at equilibrium, while the nonpolar solvation free energy functional consists of surface energy, mechanical work and attractive solvent-solute interactions. Both the polar and nonpolar solvation free energies are coupled through the characteristic function S . In the present work, geometric measure theory is utilized to convert the Lagrangian formulation of the surface into appropriate Eulerian formulation. By variation of the total

solvation free energy functional with respect to the characteristic function and electrostatic potential, a generalized geometric flow equation for the electrostatic potential and a generalized Poisson-Boltzmann equation for the characteristic function are obtained. Unlike the standard Poisson-Boltzmann equation, the generalized Poisson-Boltzmann admits a smooth dielectric profile governed by the generalized geometric flow equation, which provides a physical description of the true solvent-solute dielectric boundary, according to the variational principle. The generalized geometric flow equation balances the intrinsic geometric curvature effect and external potential due to mechanical work, solvent-solute interactions, and the electrostatic potential. The solution of the generalized geometric flow equation and the generalized Poisson-Boltzmann equation leads to quantities for the direct evaluation of the solvation free energy.

Chapter 3

Quantum formulation

The objective of this chapter is to incorporate a quantum mechanical description of charge density into our earlier differential geometry based solvation model, which is described in Chapter 2. To this end, we hope to develop a more accurate and self-consistent multiscale approach for the solvation analysis of both small and large molecules. The advantages of the present quantum formulation of the differential geometry based multiscale solvation models are follows. First, compared with our earlier solvation models, the present model is able to provide more accurate descriptions of charge arrangement during the solvation process and leads to more accurate prediction of solvation free energies. Additionally, the present multiscale model reduces the dependence of our earlier solvation models on the existing molecular mechanical force field parameters, which are typically parameterized for certain class of (macro-) molecular systems and may not be appropriate for other class of molecules. Therefore, the present model can be applied to a wider class of molecules. Moreover, compared with other existing QM based solvation models [227, 237, 42], the present model avoids the use of unphysical solvent-solute interfaces. The solvent-solute

boundary in the present model is described by the differential geometry theory of surfaces. Finally, a systematical framework is established to incorporate polar energy, nonpolar energy and quantum energy into a total energy functional. The optimization of the total energy functional leads to coupled governing equations for a set of important state functions, such as electrostatic potential, electronic density, and solvent-solute boundary profile. This set of state functions gives rise to theoretical predictions of solvation free energy, electrostatic profile and solvent-solute interface of the solvent-solute complex.

This chapter is organized as follows. Section 3.1 is devoted to the theoretical formulation of our differential geometry based quantum model of solvation. We provide a detailed description of various solvation free energy functionals. Three governing equations, i.e., the generalized Poisson-Boltzmann equation, the potential driven geometric flow (i.e., generalized Laplace-Beltrami) equation, and the Kohn-Sham equation are derived from the total energy functional via the Euler-Lagrange variation. Numerical methods and algorithms are presented in Section 3.2. This section offers detailed schemes for the solution of the above-mentioned three governing equations. The dynamical coupling of these three equations is achieved by an efficient iterative procedure. A formula for the solvation free energy estimation is also derived from the multiscale total energy functional. The present multiscale model is validated by numerical tests using a number of molecules in Section 3.3. To establish a valid approach, we have examined consistency of the electron density with the Poisson equation. The unit conversion between conventions used in our Poisson solver and that in a DFT software is discussed. The results from the present multiscale model is compared with those in our previous methods and those in the literature. Applications to three sets of molecules are given in Section 3.4. Some of these sets are computationally challenging. We

demonstrate that the present model performs well in the prediction of solvation free energies. This chapter ends with a conclusion.

3.1 Theory and model

In this section, we first prescribe the polar and nonpolar free energy functionals based on our differential geometry theory of the solvent-solute interface introduced in section 2.1.1 of Chapter 2. We then give an expression for the quantum mechanical energy of electrons. In the present work, the quantum mechanical energy of electrons is also treated as a part of the multiscale total energy for the solvation system. Governing equations for the solvation process are derived by the Euler-Lagrange variational principle.

3.1.1 Charge density based polar free energy functional

The solvation process involves both intermolecular and intramolecular interactions. Solvation analysis has been following certain convention, which may not be precisely consistent with that in other fields. Typically, solvation interactions are classified into polar type and non-polar type. The polar type of interactions is often modeled by the Poisson-Boltzmann (PB) equation with appropriate point charges at atomic central positions. In the original electromagnetic theory, the charge source of the electric potential is to be “free charges”. However, in biophysics, due to the atomistic nature of the description, the point charges are obtained by the fitting of the electron density distribution of either a charged molecule or a charge-neutral molecule into its atomic centers. Such point charge information is often stored in the database of popular software packages, such as CHARMM [146]. Therefore, the polar interactions include both charge and polarization effects inside the molecule. Note

that the effect of the rearrangement of electron charges during the solvation process needs to be computed twice, once before and once after the solvation. Polar interactions are also called electrostatic interactions. However, not all electrostatic interactions are described by the PB equation. Strictly, the electrostatic potential solved from the PB equation represents Coulombic type of interactions between charges. However, many other intermolecular interactions, such as London dispersion interactions, Debye (induced dipole) interaction, ion-dipole interactions and dipole-dipole interactions are also electrostatics in origin, and are not represented by the PB equation [121, 1, 5, 6, 194].

Sharp and Honig [199] and Gilson et al. [97] have given a formulation for the electrostatic free energy functional. However, their formulation is based on a given static sharp solvent-solute interface. In the present work, we follow our earlier definition of differential geometry based electrostatic free energy functional in Chapter 2

$$G_p = \int_{\Omega} \left\{ S \left[\rho_m \phi - \frac{1}{2} \epsilon_m |\nabla \phi|^2 \right] + (1 - S) \left[-\frac{1}{2} \epsilon_s |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \right] \right\} d\mathbf{r} \quad (3.1)$$

where Q_i is the charge of i th ionic species, N_c is the total number of ionic species, k_B is the Boltzmann constant, T is the temperature, and n_i^0 is the bulk concentration of the i th ionic species. Here, ϵ_s and ϵ_m are the permittivity, or dielectric constants of the solvent and solute domains, respectively. The ϵ is unity in vacuum, but assumes different values in different environments. In solvation analysis, ϵ is usually set to 1 or 2 in the solute domain and to 80 in the solvent domain. In Eq. (3.1), ρ_{total} is the total charge density of the

molecule and is given by

$$\begin{aligned}\rho_{\text{total}} &= qn(\mathbf{r}) - qn_n(\mathbf{r}) \\ &= qn(\mathbf{r}) - q \sum_I Z_I \delta(\mathbf{r} - \mathbf{R}_I),\end{aligned}\tag{3.2}$$

where q is the unit charge of an electron, $n(\mathbf{r})$ is the electron density, $n_n(\mathbf{r})$ is the nucleus density, and Z_I and \mathbf{R}_I are the atomic number and the position vector of nucleus I , respectively.

In Eq. (3.1), the term associated with S is the electrostatic free energy of the solute and that associated with $(1 - S)$ is the electrostatic free energy of the solvent. In our model, the surface function S will be determined by the total energy optimization. Additionally, non-polar solvent-solute interactions are modeled in the same framework as that in Section 1.1.4.

3.1.2 Quantum mechanical energy functional

In the present multiscale model, we need to evaluate the total charge density $\rho_{\text{total}}(\mathbf{r})$ by quantum mechanical principles or *ab initio* approaches. However, the *ab initio* calculation of the electronic structure of a macromolecule is intractable at present due to the large number of degrees of freedom. A vast variety of theories and algorithms have been developed in the literature to reduce the dimensionality of this many-body problem. One of the simplest *ab initio* approaches is the Hartree - Fock (HF) method, which replaces instantaneous Coulombic electron-electron repulsion interactions with a mean-field average. A variational procedure is used to minimize the energy. An alternative of the HF method is the density functional theory (DFT), which is originated from the Thomas-Fermi model. DFT represents the electronic

structure (principally the ground state) of a many-body system as a functional of a single electron density. As usual in many-body electronic structure calculations, the nuclei of the molecule of interest are treated by the Born-Oppenheimer approximation (i.e., as fixed) in DFT to generate a static external potential in which the electrons are moving. The self-consistent iterations are utilized to minimize the total energy of the system. Recently, DFT has become one of the most popular and versatile methods available in computational physics, computational chemistry and computational biology. In the present work, we incorporate the DFT description of the electronic structure of the solute molecule into our differential geometry based solvation model.

Despite the improvement in computer hardware and software for the quantum mechanical calculation, computational costs are still a major concern for the QM simulation of large molecules of interest. Therefore, so-called order- N algorithms [98, 169], in which the computer time and memory scale linearly with the simulated system size, become increasingly important. Though the plane wave basis set has advantages over local basis sets in terms of avoiding basis-set superposition error as well as convergence concerns, it is difficult to be used in the implementation of the $O(N)$ method in DFT. As such, a localized basis set is normally taken to develop fully self-consistent $O(N)$ DFT algorithms. Along this line, a software package named SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) was developed [209, 3]. It is based on a flexible linear combination of atomic orbitals (LCAO) basis set and essentially perfect $O(N)$ scaling. Therefore, it allows very fast simulations using minimal basis sets and very accurate calculations with complete multiplezeta and polarized bases [169, 168]. Moreover, the pseudopotential is used in SIESTA to avoid the calculation of core electrons and to achieve the expansion of a smooth (pseudo-

) charge density on a uniform spatial grid domain, which further accelerates the speed of quantum calculations.

3.1.2.1 Kinetic energy

Combining DFT with our differential geometry based solvation formulation, we define the kinetic energy functional as

$$G_{\text{kin}}[n] = \sum_j \int S(\mathbf{r}) \frac{\hbar^2}{2m} |\nabla \psi_j(\mathbf{r})|^2 d\mathbf{r} \quad (3.3)$$

where $m(\mathbf{r})$ is the position-dependent electron mass, $\hbar = \frac{h}{2\pi}$ with h being the Planck constant, and $\psi_j(\mathbf{r})$ are the Kohn-Sham orbitals. Here, the total electron density n is given by

$$n(\mathbf{r}) = \sum_i |\psi_i|^2, \quad (3.4)$$

where the summation is over all the Kohn-Sham orbitals. Note that orbitals $\{\psi_j\}$ are subject to the orthonormality constraint

$$\int S \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) d\mathbf{r} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \quad (3.5)$$

Obviously, Eq. (3.5) is an approximation which is valid as long as the boundary represented by the characteristic function S is sufficiently far away from atomic centers of the solute molecule. This is true in our model.

3.1.2.2 Potential energy

Without external potentials, the electrostatic potential energy of nuclei and electrons can be represented by the Coulombic interactions among the electrons and nuclei. There are three groups of electrostatic interactions: interactions between nuclei, interactions between electrons and nuclei, and interactions between electrons. Because of the Born-Oppenheimer approximation, interactions between nuclei do not directly have an impact on the structure of electrons in DFT.

According to the Coulombic law, the repulsive interaction between electrons can be expressed as the Hartree term

$$U_{ee}[n] = \frac{1}{2} \int \frac{q^2 n(\mathbf{r}) n(\mathbf{r}')}{\epsilon(\mathbf{r}) |\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (3.6)$$

where q is again the unit charge of an electron, $\epsilon(\mathbf{r})$ is the position dependent electric permittivity, and \mathbf{r} and \mathbf{r}' are positions of two interacting electrons. Equation (3.6) gives rise to a nonlinear function in terms of electron density n . Therefore, the problem of solving the electronic structure has to be resolved by self-consistent iterations.

Additionally, the attractive interactions between electrons and nuclei are given by

$$U_{en}[n] = - \sum_I \frac{q^2 n(\mathbf{r}) Z_I}{\epsilon(\mathbf{r}) |\mathbf{r} - \mathbf{R}_I|}. \quad (3.7)$$

Finally, we write the total potential energy functional as

$$G_{\text{potential}} = \int_{\Omega} S(\mathbf{r}) (U_{ee}[n] + U_{ne}[n] + E_{\text{XC}}[n]) d\mathbf{r}, \quad (3.8)$$

where the last term, E_{XC} is the exchange-correlation potential, which includes all the many-particle interactions in the solute molecule. In general, the exact form of the exchange-correlation potential is not known. There are good approximations in the practical applications, such as the local-density approximation, the local spin-density approximation, and generalized gradient approximations. A detailed elaboration of the exchange-correlation potential is beyond the scope of the present work.

3.1.3 Total free energy functional

Intuitively, it may appear that the total free energy functional is the summation of the polar, non-polar, kinetic and potential energy. However, such an approach will lead to some double counting because of the coupling among different energy terms. For example, the electrostatic energy depends on the charge density, which, in turn, depends on the kinetic and potential energies of electrons. Additionally, the electrostatic potential serves as an unknown in the polar energy functional, meanwhile it serves as an input in the potential energy of electrons. To see this connection, we need to solve the Poisson equation in vacuum ($\epsilon = 1$)

$$-\nabla^2 \phi_v(\mathbf{r}) = \rho_{\text{total}}^v(\mathbf{r}), \quad (3.9)$$

where ϕ_v is the electrostatic potential in vacuum and $\rho_{\text{total}}^v = n_v + n_n$ with $n_v(\mathbf{r})$ being the electron density in vacuum. The solution of Eq. (3.9) is

$$\phi_v(\mathbf{r}) = \int \frac{qn_v(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \sum_I \frac{qZ_I}{|\mathbf{r} - \mathbf{R}_I|}. \quad (3.10)$$

Note that Eq. (3.10) is the exact total Coulombic potential of electron-electron interactions and electron-nucleus interactions. As such, we do not need to include $U_{ee}[n]$ and $U_{en}[n]$ terms in the total free energy functional.

Finally, we propose a multiscale total free energy functional for biomolecules at equilibrium

$$\begin{aligned}
G_{\text{total}}[S, \phi, n] = & \int_{\Omega} \{ \gamma |\nabla S(\mathbf{r})| + pS(\mathbf{r}) + (1 - S(\mathbf{r}))\rho_0 U_{ss} \\
& + (1 - S(\mathbf{r})) \left[-\frac{1}{2}\epsilon_s |\nabla\phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-Q_i\phi/k_B T} - 1 \right) \right] \\
& + S(\mathbf{r}) \left[\rho_m \phi - \frac{1}{2}\epsilon_m |\nabla\phi|^2 \right] + S(\mathbf{r}) \left[\sum_j \frac{\hbar^2}{2m} |\nabla\psi_j|^2 + E_{\text{XC}}[n] \right] \} d\mathbf{r}
\end{aligned} \tag{3.11}$$

where the first row is the non-polar energy functional, the second row is the electrostatic energy functional and the last row is the electronic energy functional. As discussed above, the term $\rho_{\text{total}}\phi$ also contributes to the Coulombic potentials of electron-electron and electron-nucleus interactions. This total free energy functional provides a starting point for the derivation of governing equations and a basis for the evaluation of solvation free energies.

3.1.4 Governing equations

The total free energy functional (3.11) is a function of characteristic function S , electrostatic potential ϕ and electron density n . The governing equations of these quantities can be obtained by the first variation of the total free energy functional (3.11). From the mathematical point of view, there should exist optimal functions $S(\mathbf{r})$, $\phi(\mathbf{r})$ and a set of orbitals $\{\psi_j\}$ at the equilibrium state in which the solvation free energy is optimized. The variational procedure

for $S(\mathbf{r})$, $\phi(\mathbf{r})$ and $\{\psi_j\}$ is described below.

First, by the variation of Eq. (3.11) with respect to the electrostatic potential ϕ , we have

$$\begin{aligned} \frac{\delta G_{\text{total}}}{\delta \phi} = 0 \Rightarrow \\ S\rho_m + \nabla \cdot [(1-S)\epsilon_s + S\epsilon_m]\nabla\phi + (1-S) \sum_{i=1}^{N_c} n_i^0 Q_i e^{-Q_i\phi/k_B T} = 0 \end{aligned} \quad (3.12)$$

The Euler-Lagrange equation is used in the above variation. Equation (3.12) is the generalized Poisson-Boltzmann (GPB) equation [243, 46]

$$-\nabla \cdot (\epsilon(S)\nabla\phi) = S\rho_{\text{total}} + (1-S) \sum_{i=1}^{N_c} n_i^0 Q_i e^{-Q_i\phi/k_B T}, \quad (3.13)$$

where the dielectric function is given by

$$\epsilon(S) = (1-S)\epsilon_s + S\epsilon_m. \quad (3.14)$$

This is a smooth function. It is clear that the GPB equation utilizes a smooth dielectric profile. There is a smooth transition region for the dielectric function to change from ϵ_s to ϵ_m . Therefore, the solution procedure of the present GPB equation (3.13) avoids many numerical difficulties of solving elliptic equations with discontinuous coefficients [258, 261, 260, 254, 253] in the classical PB equation. Furthermore, in a solvent without salt, the GPB equation is simplified to the generalized Poisson equation

$$-\nabla \cdot (\epsilon(S)\nabla\phi) = S\rho_{\text{total}}. \quad (3.15)$$

Both Eqs. (3.13) and (3.15) are similar to our earlier results in Chapter 2. However, in the present multiscale model, the charge source ρ_{total} is to be determined by solving the Kohn-Sham equations, rather than by the fixed charges $\rho_{\text{fix}} = \sum_j q_j \delta(\mathbf{r} - \mathbf{r}_j)$.

Additionally, by the variation of Eq. (3.11) with respect to the surface function S , we have

$$\begin{aligned} \frac{\delta G_{\text{total}}}{\delta S} = 0 \Rightarrow \\ -\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|} \right) + p - \rho_0 U_{ss} - \frac{1}{2} \epsilon_m |\nabla \phi|^2 + \frac{1}{2} \epsilon_s |\nabla \phi|^2 \\ + k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-Q_i \phi / k_B T} - 1 \right) + \rho_m \phi + \sum_j \frac{\hbar^2}{2m} |\nabla \psi_j|^2 + E_{\text{XC}}[n] = 0 \end{aligned} \quad (3.16)$$

In Eq (3.16), $\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|} \right)$ is a generalized Laplace-Beltrami operator, which is a generalization of the usual Laplacian operator to a smooth manifold of macromolecular surface [17, 243]. In general, γ can be a function of the position $\gamma = \gamma(\mathbf{r})$ to account for the surface hydrophobicity at different locations of the macromolecule. For simplicity, it is treated as a constant in our present computation. By solving Eq. (3.16), we obtain a “physical solvent-solute boundary” (S). As discussed in earlier work [20, 17, 243], the solution of this elliptic partial differential equation (PDE) can be attained via a parabolic PDE

$$\frac{\partial S}{\partial t} = |\nabla S| \left[\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|} \right) + V \right], \quad (3.17)$$

where the generalized “potential” V is defined as

$$\begin{aligned}
V = & -p + \rho_0 U_{ss} + \frac{1}{2} \epsilon_m |\nabla \phi|^2 - \frac{1}{2} \epsilon_s |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-Q_i \phi / k_B T} - 1 \right) \\
& - \rho_{\text{total}} \phi - \sum_j \frac{\hbar^2}{2m} |\nabla \psi_j|^2 - E_{\text{XC}}[n]
\end{aligned} \tag{3.18}$$

where the electronic potentials in last row do not contribute much to V at equilibrium. This is due to the fact that they are essentially confined inside the solute molecular domain. Note that Eq. (3.17) has the same structure as the potential driven geometric flow equation defined in Chapter 2. As $t \rightarrow \infty$, the initial profile of S evolves into a steady state solution, which solves the original Eq. (3.16) with an optimal surface function S .

Finally, to derive the equation for the electronic wavefunctions, we need to incorporate the constraint as shown in Eq. (3.5) into the total energy functional. This can be easily done with a family of Lagrange multipliers $\sum_i E_i \left(\delta_{ij} - \int S \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}) d\mathbf{r} \right)$. Therefore, by the variation of Eq. (3.11) with respect to the wavefunction $\psi_j^*(\mathbf{r})$ and subject to the constraint, we have

$$\begin{aligned}
& \frac{\delta \left[G_{\text{total}} + \sum_i E_i \left(\delta_{ij} - \int S \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}) d\mathbf{r} \right) \right]}{\delta \psi_j^*} = 0 \Rightarrow \\
& \left(-\frac{\hbar^2}{2m} \nabla^2 + U_{\text{eff}} \right) \psi_j = E_j \psi_j,
\end{aligned} \tag{3.19}$$

where the Lagrange multiplier constants E_i have been interpreted as energy expectation values. Equation (3.19) is the Kohn-Sham equation in which the effective Kohn-Sham potential is defined as

$$U_{\text{eff}}(\mathbf{r}) = q\phi + V_{\text{XC}}[n], \tag{3.20}$$

where $V_{\text{XC}}[n] = \frac{dE_{\text{XC}}[n]}{dn}$ with $q\phi$ being the potential contribution from Coulombic interactions. It is to be calculated by the GPB equation (3.12) with a given total charge density. Apparently, Eq. (3.19) does not directly depend on the surface function S , so that existing DFT packages can be used in our computations with a minor modification as described in Section 3.2.2.

It is seen that the generalized Poisson-Boltzmann equation (3.13), the generalized Laplace-Beltrami equation (3.17) and the Kohn-Sham equation (3.19) are strongly coupled to each other. Therefore, these three equations have to be solved by appropriate iterative procedures. This aspect is discussed in the next section.

3.2 Numerical methods and algorithms

3.2.1 Solution of the generalized Poisson-Boltzmann equation

The solution of generalized Laplace-Beltrami equations has been studied in appendix A, including the details of some discretization schemes.

In solvation analysis, the generalized PB equation (3.15) is subject to the Dirichlet boundary condition [93]

$$\phi(\mathbf{r}) = \sum_j^{N_a} \frac{q_j}{\epsilon_s |\mathbf{r} - \mathbf{r}_j|}, \quad \forall \mathbf{r} \in \partial\Omega, \quad (3.21)$$

where q_j is the total fixed charge of the j th solute atom. One option is to use the point charges from a force field model such as CHARMM. However, in the present work, we

consider the following Dirichlet boundary condition

$$\phi(\mathbf{r}) = \int \frac{\rho_{\text{total}}(\mathbf{r}')}{\epsilon_S |\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad \forall \mathbf{r} \in \partial\Omega, \quad (3.22)$$

where the boundary condition is nonlinear — it depends on the electron density n and thus needs to be implemented iteratively.

The standard second order center difference scheme is applied in this study to solve Eq. (3.15). An accurate solution can be obtained due to the continuous dielectric definition $\epsilon(S)$. Let the pixel (i, j, k) represent the position (x_i, y_j, z_k) . The discretized form of Eq. (3.15) is

$$\begin{aligned} & \left[\epsilon \left(x_i + \frac{1}{2}h_x, y_j, z_k \right) [\phi(i+1, j, k) - \phi(i, j, k)] \right. \\ & + \epsilon \left(x_i - \frac{1}{2}h_x, y_j, z_k \right) [\phi(i-1, j, k) - \phi(i, j, k)] \left. \right] \frac{1}{h_x^2} \\ & + \left[\epsilon \left(x_i, y_j + \frac{1}{2}h_y, z_k \right) [\phi(i, j+1, k) - \phi(i, j, k)] \right. \\ & + \epsilon \left(x_i, y_j - \frac{1}{2}h_y, z_k \right) [\phi(i, j-1, k) - \phi(i, j, k)] \left. \right] \frac{1}{h_y^2} \\ & + \left[\epsilon \left(x_i, y_j, z_k + \frac{1}{2}h_z \right) [\phi(i, j, k+1) - \phi(i, j, k)] \right. \\ & + \epsilon \left(x_i, y_j, z_k - \frac{1}{2}h_z \right) [\phi(i, j, k-1) - \phi(i, j, k)] \left. \right] \frac{1}{h_z^2} \\ & = -S(i, j, k)\rho_m(i, j, k) \end{aligned} \quad (3.23)$$

where h_x, h_y and h_z are the grid spacings at x, y and z directions, respectively. Here, $\rho_{\text{total}}(i, j, k)$ is the charge density at grid point (x_i, y_j, z_k) , which is calculated from the electronic charge density $n(\mathbf{r})$ and nucleus density n_n . The implementation of ρ_{total} will

be discussed in the next paragraph. As such, the discretized PB equation can be converted into the standard linear algebraic equation system of the form $AX = B$, where X is the solution of interest, A is the discretization matrix and B is the source term associated with the charge density. It has been shown previously that the PB solver is able to deliver the designed second-order accuracy in Chapter 2.

On the right hand side of Eq (3.23), the charge density at each grid point should be given. As an efficient approach, atomic charges have been widely used to approximate the charge density of electrons and nuclei, especially for large molecules of general interest. Therefore, the point charge approach has gained much popularity in PB solvers as well as PB applications [147, 12, 112, 116]. Nevertheless, charge assignment at atomic centers is a nontrivial issue. The deficiencies of the atomic point charge approach have been discussed in the Introduction. The direct implementation of the quantum mechanical charge density can avoid errors caused by the atomic point charge approximation. Moreover, this approach frees us from the electrostatic potential fitting process. To carry it out in the finite different scheme, the total charge density $\rho_{\text{total}}(i, j, k)$, which consists of the electron density $n(\mathbf{r})$ and nucleus density n_n , needs to be prescribed at each grid point of the computational domain. In particular, the nucleus density $n_n(\mathbf{r})$, which is considered as stationary and located at the center of atoms, can be distributed to the nearest eight neighboring grid points by the second order interpolation (i.e., the trilinear mapping). The distributed nucleus core point charges are converted into the nucleus charge density by dividing point charges with the volume of the unit grid. Finally, the total charge density at each grid point is obtained by the summation of nucleus density and the electronic charge density which is directly taken from the quantum mechanical calculation.

However, a new issue arises from the above treatment of nuclei. Since each nucleus core charge is split into its eight neighboring grid points, it is easy to find out that short range interactions are biased and self-interaction energies are artificially introduced. This is due to the interactions of grid charges within one single atom. It exists even in the absence of solvent and any other charges. Apparently, this is a pure artifact due to the finite difference approach and must be eliminated. Within the partial charge approach the artifact can be canceled out mainly by calculating the PB equation twice, one in vacuum and the other in the solvent. It turns out that this strategy also works well here. Numerical tests regarding this cancellation of self-interaction energies are demonstrated later. It is important to point out that numerically if one implements the quantum mechanical calculation with a non-frozen core method, the remaining error from the self-interaction cancellation is still too large to be neglected. In other words, the above cancellation strategy may fail when one applies a non-frozen core approach. Therefore, frozen core approaches, such as pseudopotential methods, must be applied in our quantum calculations here. Because frozen core approaches dramatically reduce the number of charges in each nucleus and thus implicitly decrease implementation errors.

The biconjugate gradient method is a good choice in solving the PB equation. However, as we have demonstrated in Chapter 2, the combination of stabilized biconjugate gradient method (BiCG) and the blocked Jacobi preconditioner (BJAC) from PETSc (<http://www.mcs.anl.gov/petsc/petsc-as/>), as well as the combination of the orthomin method (OM) and the incomplete LU factorization preconditioner (ILU) from SLATEC, speeds up the process of the PB solution. In this study, we apply the combination of ILU and OM from SLATEC. In our iteration procedure, the prior electrostatic potential is taken as a good initial guess

for the followed linear system solving procedure. It turns out that this treatment makes the generalized PB solver converge much faster than simply setting the initial guess to be 0 [46]. Additionally, the convergence tolerance is set to be 10^{-4} as a good compromise between the accuracy and efficiency.

3.2.2 Solution of the generalized Kohn-Sham equation

The generalized Kohn-Sham equation (3.19) admits all-electron and all-nucleus potentials. The direct solution of Eq. (3.19) is very expensive for macromolecules. Therefore, further simplifications are necessary. In particular, because classical DFT methods have been developed in the past few decades, the solution of Eq. (3.19) needs to make use of existing DFT methods.

Note that the Coulombic potential functionals shown in Eqs. (3.6) and (3.7) involve spatially varying dielectric constants, which reflect the solvation process. The related spatially varying electrostatic potential is built in the generalized Poisson-Boltzmann equation (3.13), whose solution gives rise to the electrostatic potential energy $q\phi$ used in the generalized Kohn-Sham equation (3.19). In contrast, the standard Kohn-Sham equation is for a molecular system in vacuum and its Coulombic potentials are of the form of $q\phi_V$ where ϕ_V is given by Eq. (3.10) with the total charge density in vacuum described in the next section. The effective potential in the generalized Kohn-Sham equation (3.19) can be written as

$$U_{\text{eff}}[n] = q\phi + V_{\text{XC}}[n] = q\phi_{\text{RF}} + U_{\text{eff}}^0(\mathbf{r}), \quad (3.24)$$

where

$$\phi_{\text{RF}} = \phi - \phi_0 \quad (3.25)$$

is called the reaction field potential. Here ϕ_0 is the solution of the Poisson equation in the homogeneous medium with the charge source $\rho_{\text{total}}(\mathbf{r})$

$$-\nabla \cdot \epsilon_0 \nabla \phi_0(\mathbf{r}) = \rho_{\text{total}}(\mathbf{r}), \quad (3.26)$$

where $\rho_{\text{total}}(\mathbf{r})$ is obtained from the generalized Kohn-Sham equation (3.19). In Eq. (3.24), $U_{\text{eff}}^0(\mathbf{r})$ is the Kohn-Sham potential

$$U_{\text{eff}}^0(\mathbf{r}) = q\phi_0 + V_{\text{XC}}[n]. \quad (3.27)$$

In the present work, $U_{\text{eff}}^0(\mathbf{r})$ is represented by the traditional Kohn-Sham potential. Consequently, a vast variety of computational techniques developed for the traditional Kohn-Sham DFT can be utilized in the present work. What we need to do in solving the generalized Kohn-Sham equation (3.19) is to add a reaction field potential $q\phi_{\text{RF}}$ to an existing Kohn-Sham DFT solver.

The most important issues in the solution of the Kohn-Sham equation are the selection of the exchange-correlation potential and the use of the pseudopotential. The pseudopotential approach eliminates the complicated effects of core electrons and allows the expansion of a smooth (pseudo-) charge density on a uniform spatial grid. In this approach, the chemically active valence electrons are dealt with explicitly, while the core electrons are ‘frozen’ and considered together with the nuclei as fixed non-polarizable ion cores. With the pseudopotential approximation, the formalism of the total energy functional needs to be modified,

which leads to the following expression of a Kohn-Sham effective potential [209]

$$U_{\text{eff}}^0(\mathbf{r}) = \sum_I V_I^{\text{local}}(\mathbf{r}) + \sum_I V_I^{\text{nonlocal}} + V^H(\mathbf{r}) + V_{\text{XC}}(\mathbf{r}) \quad (3.28)$$

where $V^H(\mathbf{r})$ and $V_{\text{XC}}(\mathbf{r})$ are total Hartree and exchange-correction potentials, respectively. Here, V_I^{local} and V_I^{nonlocal} are the local part and the nonlocal part of the pseudopotential of atom I . For elaborated discussions of the above potentials, we refer the reader to an excellent review [209].

In the present work, SIESTA (Spanish initiative for the electronic structure of thousands of atoms), a quantum mechanical package of high efficiency, is utilized for solving our generalized Kohn-Sham equation (3.19). SIESTA possesses the ability to perform the density functional theory (DFT) simulations of more than a thousand atoms. The details of the package has been extensively described [209]. It develops a self-consistent density functional method using the standard norm-conserving pseudopotential and a flexible numerical linear combination of atomic orbital (LCAO) basis sets with essential perfect $O(N)$ scaling, in which the computer CPU time and memory scale linearly with the simulated system size. The exchange and correlation are treated within the Kohn-Sham DFT. Both the local density approximation and local spin density approximation (LDA/LSDA), as well as the generalized gradient approximation (GGA) are allowed. Moreover, SIESTA permits very fast simulations by using minimal basis sets as very accurate calculations with complete multiple-zeta and polarizable bases. Therefore, it can provide a general scheme to perform quantum calculations with requirements ranging from being very fast to being very accurate. For all of the simulations in the present work, the default double- ζ plus single polarization (DZP) bases are used. The MeshCutoff is set as 125 Rydberg and the LDA is applied. The

SolutionMethod is set to be ‘diagon’.

3.2.3 Evaluation of the solvation free energy

The solvation free energy is the energy required or released from the transfer of a unit of solute molecules from vacuum to a solvent. By definition, it is calculated by the difference of the total energies in solution and in vacuum

$$\Delta G_{\text{total}} = G_{\text{total}}[S, \phi, n] - G_{\text{vacuum}}[\phi_v, n_v] \quad (3.29)$$

where ϕ_v is the electrostatic potential in vacuum and n_v is the solute electronic density in vacuum, which is defined in terms of the electronic wavefunctions of the solute in vacuum

$$\psi_j^v(\mathbf{r}) \quad n_v(\mathbf{r}) = \sum_j |\psi_j^v(\mathbf{r})|^2. \quad (3.30)$$

In Eq. (3.29), $G_{\text{total}}[S, \phi, n]$ is given in Eq. (3.11) and $G_{\text{vacuum}}[\phi_v, n_v]$ denotes the total energy functional in vacuum

$$\begin{aligned} G_{\text{vacuum}}[\phi_v, n_v] = & \int \left[\rho_{\text{total}}^v \phi_v - \frac{1}{2} \epsilon |\nabla \phi_v|^2 \right. \\ & \left. + \sum_j \frac{\hbar^2}{2m} |\nabla \psi_j^v|^2 + E_{\text{XC}}[n_v] \right] d\mathbf{r} \end{aligned} \quad (3.31)$$

where $\rho_{\text{total}}^v = qn_v - qn_n$ is the total charge density in vacuum. For simplicity, we have omitted the ionic density $k_B T \sum_{i=1}^{N_c} n_i^0 (e^{-Q_i \phi_v / k_B T} - 1)$ in Eq. (3.29). Note that the variation of $G_{\text{vacuum}}[\phi_v, n_v]$ gives rise to the standard Poisson equation (3.9) and the

standard Kohn-Sham equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{eff}}^v\right)\psi_j^v = E_j^v\psi_j^v, \quad (3.32)$$

where E_j^v and ψ_j^v are appropriate eigenvalue and eigenfunction of Hamiltonian $H^v = -\frac{\hbar^2}{2m}\nabla^2 + U_{\text{eff}}^v$.

However, there is a technical difficulty in the direct evaluation of $G_{\text{total}}[S, \phi, n]$. Namely, the integration of the quantum mechanical terms in Eq. (3.11) requires the S function profile, which is not involved in most Kohn-Sham DFT software packages. Therefore, in the present work, we evaluate the solvation free energy by the following approximation

$$\Delta G_{\text{total}} = G_{\text{np}} + \Delta G_{\text{p}} + \Delta G_{\text{QM}} \quad (3.33)$$

where G_{np} , ΔG_{p} and ΔG_{QM} are the non-polar, polar and quantum mechanical contributions, respectively. The non-polar solvation free energy does not exist in vacuum, and its form in solution is given by

$$G_{\text{np}}[S] = \int [\gamma|\nabla S(\mathbf{r})| + pS(\mathbf{r}) + \rho_0(1 - S(\mathbf{r}))U_{ss}] d\mathbf{r}. \quad (3.34)$$

By using the Gauss' divergent theorem and integration by parts, it is easy to show that the polar solvation energy is given by

$$G_{\text{p}}[S, \phi, n] = \frac{1}{2} \int_{\Omega_s} \rho_{\text{total}} \phi d\mathbf{r}. \quad (3.35)$$

Similarly, the polar solvation energy in vacuum is $\frac{1}{2} \int \rho_{\text{total}}^v \phi_v d\mathbf{r}$. Therefore, one may com-

pute the change of the polar solvation energy by $\frac{1}{2} \left(\int_{\Omega_s} \rho_{\text{total}} \phi d\mathbf{r} - \int \rho_{\text{total}}^v \phi_v d\mathbf{r} \right)$. However, this expression leads to a situation that the quantum mechanical contribution ΔG_{QM} cannot be evaluated in SIESTA because of the lack of required potential terms. Additionally, such an expression is inconsistent with the conventional electrostatic solvation free energy of the form

$$\Delta G_{\text{p}} = \frac{1}{2} \int_{\Omega_s} \rho_{\text{total}} [\phi - \phi_0] d\mathbf{r} = \frac{1}{2} \int_{\Omega_s} \rho_{\text{total}} \phi_{\text{RF}} d\mathbf{r}. \quad (3.36)$$

Therefore, in the present solvation analysis ΔG_{p} is calculated by Eq. (3.36), which leads to two remaining electrostatic potential terms $\frac{1}{2} \left(\int_{\Omega_s} \rho_{\text{total}} \phi_0 d\mathbf{r} - \int \rho_{\text{total}}^v \phi_v d\mathbf{r} \right)$. These terms are combined with the rest of the quantum energy functionals to compute the change of the quantum mechanical energy as

$$\Delta G_{\text{QM}} = \sum_j \left[\langle \psi_j | H^0 | \psi_j \rangle - \langle \psi_j^v | H^v | \psi_j^v \rangle \right]. \quad (3.37)$$

where $H^0 = -\frac{\hbar^2}{2m} \nabla^2 + U_{\text{eff}}^0$. Note that wavefunctions $\{\psi_j\}$ are computed with the full Hamiltonian in the solution. The main advantage of the quantum mechanical energy change given in Eq. (3.37) is that it can be easily computed by using existing DFT software packages as discussed in the last section.

The current formula of the solvation free energy is systematically derived from the framework of the differential geometry based solvation model. It consists of the non-polar energy G_{np} , the electrostatic solvation free energy ΔG_{p} , and the change of the solute self-energy ΔG_{QM} due to the redistribution of electrons in the solvation process. It is of interest to see that the formulation of the present solvation analysis is consistent with that in the litera-

ture [224, 237, 238], which is basically computed by using a good chemical intuition. The reliability and accuracy of the current model are further validated by a comparison of the present prediction with experimental data, as well as with that in the literature in Section 3.3.

3.2.4 Dynamical coupling of involved PDE equations

As described in Section 3.1.4, the total charge density in the solution is obtained by solving the Kohn-Sham equation in the presence of the reaction field potential $\phi_{\text{RF}} = \phi - \phi_0$, which is computed by solving the PB equation and the Poisson equation, i.e., Eqs. (3.13) and (3.26). On the other hand, the solution of the PB equation requires the quantum mechanically calculated charge density, the surface profile S , and the dielectric profile $\epsilon(S)$ which are generated by solving the generalized Laplace-Beltrami equation (LBE). Moreover, the potential in the generalized Laplace-Beltrami equation contains the terms associated with the electrostatic potential from the PB equation and the charge density from the Kohn-Sham equations. In principle, the Laplace-Beltrami equation, the generalized PB and Kohn-Sham equations need to be solved simultaneously until the convergence is reached, i.e, the solvation energy of two runs agrees with each other within a prescribed tolerance. This can be achieved via a self-consistent iteration procedure.

In practice, we adopt an inner-outer iterative strategy to implement the self-consistent procedure. The inner iterations concern the solution of the coupled generalized PB equation and the Laplace-Beltrami equation. These iterations have been carried out in Chapter 2, except for the different representation of the continuous charge density. In the present work, the inner iterations are combined with the solution of the Kohn-Sham equation during the

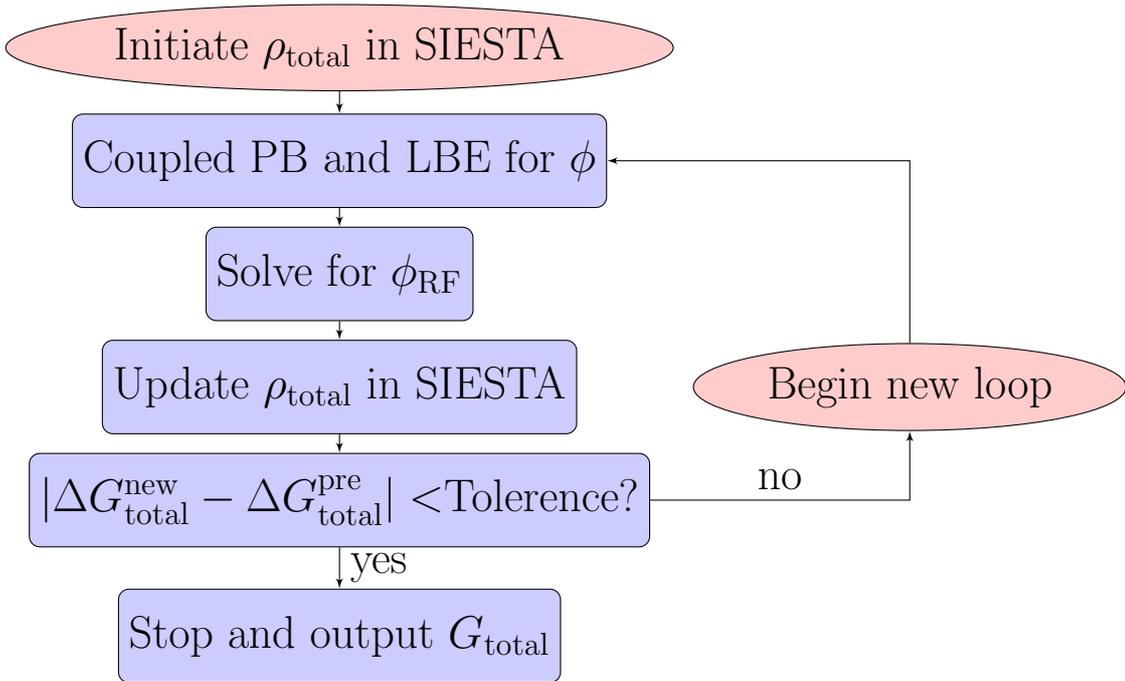


Figure 3.1: Flowchart of the numerical solution of the coupled PDEs.

outer iterations. More details are described in the following and can be seen from a work flow chart in Figure 3.1 as well.

Step 0 (Generation of the solute quantum energy in vacuum and initialization of the charge density): We carry out a quantum mechanical calculation in vacuum with SIESTA to obtain an initial total charge density. The solute quantum energy in vacuum is recorded for computing ΔG_{QM} in the output step.

Step 1 (Inner iteration of the coupled PB and Laplace-Beltrami equations): Given a total charge density distribution, as described in Chapter 2, a temporary electrostatic potential ϕ can be generated by solving the generalized PB equation with a temporary S . It is followed by the evolution of time-dependent generalized Laplace-Beltrami equation for a number of time iterations. With the updated intermediate S , one can update the electrostatic potential via the PB equation. This cycle repeats until the electrostatic solvation energy converges

within a pre-determined criterion. Note that the relaxation algorithm should be used to guarantee the convergence.

Step 2 (Generation of a reaction field potential): Solve the PB equation in a homogeneous medium with the same total charge density as that in the previous step. Then the reaction field potential ϕ_{RF} is obtained by the difference between the electrostatic potential from the previous step and the current Poisson calculation in the homogeneous medium.

Step 3 (Solution of the Kohn-Sham equation): Run the SIESTA program again to obtain a new total charge density by incorporating the computed reaction field potential into the Kohn-Sham Hamiltonian.

Step 4 (Calculation of the solvation energy and convergence check): Calculate the total solvation free energy with the resulting ϕ , S and $\{\psi_j\}$. Go back to Step 1 until it converges. When convergence is reached, one ends the iteration and outputs the solvation free energy.

To output the charge density in SIESTA, one needs to set `SaveRho` to be true in the input `fdf` file, while `SpinPolarized` is false according to the fact that all tested molecules in this work are neutral. Eventually, SIESTA generates an `XV` file to store lattice vectors and atomic positions, together with an `RHO` file to record the values of the charge quantity on the grid points. However, the standard input for our PB solver is a Gaussian CUBE format file which contains the origin, grid spacing, atomic coordinates and charge densities. Therefore, a CUBE format file must be created based on the information from the `XV` file and `RHO` file to transfer the charge density data from SIESTA to the PB solver. It can be carried out by a subroutine `grid2cube.x`, which is included in the SIESTA package. During the translation, the coordinates of the origin are shifted to make the molecule roughly appear at the center of the computational domain.

After the reaction field potentials are obtained by the PB solver, they are regarded as an external solvent potential effect and must be incorporated into the Kohn-Sham quantum calculation in SIESTA. As such, reaction field potentials have to be passed into SIESTA during the self-consistent iteration process. It can be done by adding them into the variable of total potential named V_{scf} in the subroutine file `dhsf.F` under “Siesta-3.0-b/Src”. Therefore, a data translation procedure is required to pass the reaction field potential from the PB solver into SIESTA. Furthermore, attention needs to be paid to the unit conversion. In particular, the unit of distance used in SIESTA is Bohr, while it is Angstrom in the PB solver. The units of potential are $e_C/\text{Angstrom}$ and $\text{Rydberg}/e_C$ in the present PB solver and in SIESTA, respectively. Here e_C denotes the fundamental charge used as the unit of a point charge in both the PB solver and SIESTA.

3.3 Numerical test and validation

This section provides systematical validations for the proposed model and computational algorithms. The performance of SIESTA has been tested and described in the literature [209]. The generalized Laplace-Beltrami equation (3.17) has the same differential operator as our earlier mean curvature flow [20] except for the extra source terms. Previously, we have numerically proved that the explicit Euler algorithm delivers the reliability and convergence of the solution of the Laplace-Beltrami equation, and the finite central difference scheme is of second-order convergence in space [17]. For the PB solver, given the partial charges, it has also been numerically proved to be of second-order convergence in Chapter 2. However, here the source term in the PB equation is no longer represented by the partial charges adopted from existing molecular mechanical force fields, such as AMBER or CHARMM. In

this work, the source term is represented by the charge density obtained directly from the quantum calculation. There are new concerns from this different charge strategy. First of all, regarding the distribution of nucleus charges, which are much larger than partial charges, it is crucial to know whether the self-interaction energy artifact within a single atom leads to non-neglected bias even after the treatment of the energy cancellation. Secondly, it is important to check whether the implementation is correct in terms of the data translation and unit conversion between different solvers during the self-consistent iteration procedure. Based on these considerations, we first examine the cancellation of self-interaction artificial energy, then continue to check the data translation and unit conversion between the PB solver and SIESTA. Finally, we demonstrate the overall accuracy of our model in the calculation of solvation free energies, as well as the solvent effect on the solute electronic structure, by a comparison with experimental data.

3.3.1 Validation of the cancellation of self-interaction energy

As described earlier, the use of the finite difference scheme in the solution of the Poisson-Boltzmann equation results in the artifact of self-interaction energy which needs to be removed. Although it is common to assign the partial charge at the center of each atom to mimic the effect of electrostatic interactions, this approach also has a similar issue of artificially introduced self-interaction energy using the finite different scheme. It is known that the cancellation of self-interaction energy works fine with the partial charge formalism by computing the PB equation twice. However, the direct use of the quantum mechanical charge density in the PB equation behaves differently. Because the magnitude of distributed nucleus charges is much larger than that of partial charges, the direct use of the quantum

mechanic charge density leads to much larger self-interaction energy in the finite difference scheme. Therefore, the accuracy of the PB solver for the reaction field potential and the solvation calculation should heavily depend on the cancellation procedure. To our knowledge, with the direct use of the quantum charge density in the PB equation, no numerical test has been done for the impact of self-interaction, neither has the performance of the artifact energy cancellation been examined. To validate our approach, we test 16 small molecules whose partial charges can be obtained from the literature. The details of structure data and parameter setting are described in Section 3.4. In particular, a uniform grid size 0.25\AA is applied to the computation with the partial charge approach. However, in quantum calculation the grid dimension is the same both in the PB solver and in SIESTA. It is automatically generated in SIESTA by setting the MeshCutoff equal to 125 Rydberg. Therefore, the grid size in the present simulation is no longer uniform and varies with different molecules. Take the water molecule as example, when cutoff energy is 125 Rydberg, the grid size is $h_x \times h_y \times h_z = 0.1190\text{\AA} \times 0.1192\text{\AA} \times 0.1182\text{\AA}$, which is fine enough for the solvation calculation. Table 3.1 and 3.2 list the total electrostatic energies both in vacuum and in solution for these 16 molecules, together with the electrostatic solvation free energies which are the difference between the total electrostatic energies in vacuum and in solution. It is found that, as expected, the self-interaction energies with the quantum charge density are much larger than those with the partial charge treatment. The former is hundreds of times larger than the latter. However, through the cancellation, the electrostatic solvation free energies are very close to each other. It can be concluded that most of the self-interaction energy artifact can be removed in the present finite difference scheme with either the fixed partial charge source or the charge density source. Therefore, the direct use of the quan-

tum charge density in the PB solver with the finite difference scheme is validated. Note that although the total electrostatic energies vary under different mesh cutoff energies, the resulting electrostatic solvation free energies show convergence.

Table 3.1: Comparison of total electrostatic energy (kcal/mol) and electrostatic solvation energy (kcal/mol) obtained with the partial charge approach.

Compound	Vacuum	Solution	ΔG_p
glycerol triacetate	2443.64	2456.10	-12.46
benzyl chloride	210.88	215.93	-5.04
m-bis(trifluoromethyl)benzene	1472.87	1476.12	-3.25
N,N-dimethyl-p-methoxybenzamide	1068.47	1077.68	-9.21
N,N-4-trimethylbenzamide	866.22	873.91	-7.69
bis-2-chloroethyl ether	315.5	319.77	-4.23
1,1-diacetoxyethane	1754.93	1763.19	-8.25
1,1-diethoxyethane	613.59	618.04	-4.45
1,4-dioxane	316.81	322.47	-5.65
diethyl propanedioate	1726.22	1734.06	-7.84
dimethoxymethane	517.65	522.18	-4.55
ethylene glycol diacetate	1768.99	1777.45	-8.46
1,2-diethoxyethane	484.14	488.46	-4.32
diethyl sulfide	133.21	135.59	-2.39
phenyl formate	876.50	884.3	-7.85
imidazole	944.38	955.65	-11.27

3.3.2 Validation of data translation and unit conversion

In this section, we demonstrate the reliability of data translation and unit conversion between the PB solver and SIESTA by comparing the results from the current calculations with ones from the literature. As stated before, the solution of the PB solver depends on the input of the quantum charge density from SIESTA. The polarization of electron cloud, in turn, requires the input of the reaction field potential obtained from the solution of the PB solver. Therefore, the reliability of charge density data passed into the PB solver can be tested by

Table 3.2: Comparison of total electrostatic energy (kcal/mol) and electrostatic solvation energy (kcal/mol) obtained with the direct use of charge density.

Compound	Vacuum	Solution	ΔG_p
glycerol triacetate	704426.41	704439.14	-12.73
benzyl chloride	341037.68	341043.97	-6.29
m-bis(trifluoromethyl)benzene	669772.11	669774.74	-2.63
N,N-dimethyl-p-methoxybenzamide	533551.77	533563.01	-11.24
N,N-4-trimethylbenzamide	396325.28	396335.07	-9.79
bis-2-chloroethyl ether	478232.71	478239.05	-6.34
1,1-diacetoxyethane	471728.52	471736.99	-8.47
1,1-diethoxyethane	332415.49	332421.71	-6.22
1,4-dioxane	286515.14	286521.15	-6.00
diethyl propanedioate	434796.12	434805.20	-9.08
dimethoxymethane	258225.34	258230.45	-5.11
ethylene glycol diacetate	434943.17	434952.17	-9.00
1,2-diethoxyethane	343486.83	343493.67	-6.85
diethyl sulfide	223784.37	223788.70	-4.32
phenyl formate	328961.95	328968.86	-6.91
imidazole	199535.85	199549.02	-13.16

the solution of the PB equation. Meanwhile, the validity of reaction field potential values is illustrated via the change of the solute self-energy due to the polarization of electrons. For the purpose of a comparison, results by Wang et al [238] are used because of similar energy decomposition in the solvation analysis. In their work, they studied the polarization of electron cloud during the solvation process as well. A quantum mechanical calculation based on density functional theory and the Poisson-Boltzmann equation was implemented. Finally, by chemical analysis the solvation free energy was also decomposed into an electrostatic solvation free term, a change term in the solute self-energy and a non-polar term. The software UHBD together with a pre-determined solvent excluded surface was used for the solution of the PB equation in their work [238]. Table 3.3 displays the comparison in solvation free energy components for three small molecules which are water, NH_3 and methanol. It is

evident that the results of solvation components from two different methods are comparable to each other. This consistency proves the appropriate data translation process used in different forms of computation domains, as well as the correct unit conversion between the PB solver and SIESTA.

Table 3.3: Comparison of solvation energy components between present results and those of Wang *et al* [238] for three small molecules.

Compound	ΔG_p (kcal/mol)		ΔG_{QM} (kcal/mol)		G_{np} (kcal/mol)	
	OSMQ	Wang	OSMQ	Wang	OSMQ	Wang
Water	-7.55	-7.36	0.86	1.01	0.39	0.59
Methanol	-6.18	-5.53	0.63	0.77	0.57	0.77
NH ₃	-5.12	-6.77	0.87	0.92	0.46	0.63

3.3.3 Accuracy of solvation free energies computed by the present model

Besides the data translation and unit conversion, the overall accuracy of the present model still needs to be further verified by comparing with experimental data. In particular, it is of crucial importance to check the accuracy of the total solvation energy as well as the solvent effect on the solute electronic structure. The contribution of the solvation free energy from the polarization of electron cloud can be decomposed into two parts. As shown in Table 3.4, prior to the polarization, the solvent interacts with the solute based on its vacuum electronic distribution which gives rise to the electrostatic solvation energy ΔG_p^v using the vacuum charge density. When the polarization takes place, the electron cloud is redistributed to reach a more favorable interaction with the solvent. This generates a gain for the solvation free energy $\Delta\Delta G_p = \Delta G_p - \Delta G_p^v$. However, the redistribution of the electron cloud leads to the change of the interactions between electrons and nuclei and those between electrons. It

causes an unfavorable decrease in the solvation energy (ΔG_{QM}). Therefore, the total energy contribution of the polarization is calculated by the sum of ΔG_{QM} and $\Delta\Delta G_{\text{p}}$. Table 3.4 summarizes the numerical results of the total solvation free energy and the decomposition for the above mentioned three small molecules. Here, $\Delta G_{\text{total}} = \Delta G_{\text{p}} + G_{\text{np}} + \Delta G_{\text{QM}}$. As demonstrated in the table, the total solvation energies compare well with experimental data [237]. Moreover, it is expected from the classical linear response theory that the loss from the distortion of electron cloud is equal to about half of the gain from the solute-solvent interaction energy [224]. It is evident that our results are quantitatively in accord with the theory. The slightly higher ratio (about 0.6) may be caused by the change in the exchange-correction energy term within the DFT calculations.

Table 3.4: Solvation free energy (kcal/mol) and its decomposition.

Compound	ΔG_{p}^v	ΔG_{p}	$\Delta\Delta G_{\text{p}}$	ΔG_{QM}	$ \frac{\Delta G_{\text{QM}}}{\Delta\Delta G_{\text{p}}} $	ΔG_{total}	Exptl [237]
Water	-6.25	-7.55	-1.30	0.86	0.66	-6.31	-6.30
Methanol	-5.18	-6.18	-1.00	0.63	0.63	-4.98	-5.11
NH3	-3.86	-5.12	-1.26	0.87	0.69	-3.79	-4.29

3.4 Applications

We consider three types of applications in this section. First, we apply our new multiscale model to a set of 24 small molecules. Then, a challenging set of 16 molecules is studied. Finally, three larger molecules are taken for efficiency and robustness test. The Dirichlet boundary condition is used for both the generalized Poisson-Boltzmann equation and the generalized Laplace-Beltrami equation as in our previous calculations [259, 252, 93, 20].

Table 3.5: Comparison of solvation free energies (kcal/mol) obtained from the present model and experimental data for 24 small molecules.

Compound	N_a	G_{np}	ΔG_p	ΔG_{QM}	ΔG_{total}	Exptl [237]	Error
Ethanol	9	0.74	-6.33	0.61	-4.98	-5.01	0.03
Propionamide	12	1.02	-12.19	2.14	-9.03	-9.41	0.38
H ₂ O	3	0.39	-7.55	0.86	-6.31	-6.30	-0.01
Phenol	13	1.21	-7.88	1.01	-5.66	-6.60	0.94
Methanethiol	6	0.7	-4.29	0.50	-3.08	-1.24	-1.84
Propionic acid	11	1.02	-7.73	0.87	-5.84	-6.47	0.63
Acetamide	9	0.82	-14.23	2.65	-10.76	-9.71	-1.05
Acetonitrile	10	0.87	-8.23	1.36	-5.99	-3.9	-2.09
Ethanethiol	9	0.87	-4.53	0.47	-3.19	-1.3	-1.89
Aniline	14	1.23	-9.34	1.15	-6.97	-5.49	-1.48
Methanol	6	0.57	-6.18	0.63	-4.98	-5.11	0.13
Acetic acid	8	0.83	-8.67	1.03	-6.81	-6.7	-0.11
1-methylcytosine	16	1.32	-23.9	5.96	-16.62	-18.4	1.78
Pyridine	11	0.98	-7.61	1.49	-5.14	-4.7	-0.44
9-methyladenine	18	1.45	-18.5	3.17	-13.88	-13.6	-0.28
1-methyluracil	15	1.28	-15.35	2.78	-11.29	-14.0	2.72
NH ₃	4	0.46	-5.12	0.87	-3.79	-4.29	0.50
4-cresol	16	1.35	-7.94	1.00	-5.58	-6.13	0.55
4-Methylimidazole	12	1.04	-15.88	3.84	-11.00	-10.25	-0.76
Methylethyl sulfide	12	1.03	-4.83	0.55	-3.25	-1.49	-1.76
n-Butylamine	16	1.08	-7.19	0.81	-5.29	-4.29	-1.00
3-Methylindole	19	1.42	-10.40	1.66	-7.32	-5.91	-1.41
Methylamine	7	0.54	-7.70	1.07	-6.09	-4.5	-1.59
Benzene	12	1.20	-4.69	0.51	-2.98	-0.9	-2.08
RMS error							1.31
Average error							1.06

3.4.1 Solvation free energies of 24 small molecules

Encouraged by the successful reproduction of the solvation free energies for the above three small molecules, we apply the present differential geometry based multiscale model and algorithms to the solvation analysis of an extended set of 24 small neutral organic molecules. In this application, all geometric structures are taken from the Pubchem database

(<http://pubchem.ncbi.nlm.nih.gov>). The required pseudopotential input files for SIESTA in the psf format are conveniently produced through a pseudopotential generator web (www.tddft.org/programs/octopus). Atomic radii for the LB equation are adopted from a new parametrization of ZAP-9 used by Nicholls [160] and in Chapter 2. Specifically, the radii of hydrogen, carbon, oxygen, nitrogen, chlorine, fluorine and sulfur are set to be 1.1Å, 1.87Å, 1.76Å, 1.40Å, 1.82Å, 2.4Å and 2.15Å, respectively. Note that different surface definitions in implicit solvent models should have their own optimal radii set [235]. In particular, it is found numerically that a continuous dielectric definition based model is supposed to have radii of slightly larger values than those of a sharp interface based model [46]. Therefore, we multiply the radii from ZAP-9 by a factor. In practice, a factor of 1.1 is used for all atomic radii in a molecule of more than 14 atoms. However, if a molecule has less than 15 atoms, the factor is given by a formula $1.02 + (\text{Max}(0, N_a - 5)) * 0.008$, where N_a represents the total number of atoms. Numerically it turns out that the solvation energy predictions are sensitive to the radii factor for small molecules with less than 15 atoms. In this case, the fewer atoms, the smaller factor.

In our calculation, since the polarization is treated explicitly with the quantum mechanical calculation, we set the dielectric constant in the solute region as $\epsilon_m = 1$, while $\epsilon_s = 80$ for the solvent region. Other parameters are set in a similar way as in Chapter 2: We choose $\rho_0/\gamma = 2$ and take into account the pressure by setting $p/\gamma = 0.2$. Note that in the numerical simulation, all ratio parameters here are treated as dimensionless. For L-J parameters, σ_s is chosen to be 0.65Å and σ_i is the solute atomic radii [235]. Due to the continuum representation of solvent in our model, the 6-12 Lennard Jones potential formula (A.2) differs from the standard version—the distance used in our formula is no longer the distance between

the centers of solute atoms and the centers of solvent atoms but the distance between a specific position in the solvent region and the centers of solute atoms. Therefore, the setting of well depth ε_i differs from the ones taken from AMBER or OPLS force fields. As we did in Chapter 2, it is determined by an equality, that is, $\varepsilon_i \left[\left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^{12} - 2 \left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^6 \right] = D_i$ if \mathbf{r} is on the vdW surface of the atom. Here the constant D_i should have different values for various types of atoms. For simplicity we use a uniform constant $D = 1.0$ to determine the value of ε_j . The grid dimension is the same both in the PB solver and in SIESTA, and it depends on the MeshCutoff energy value in SIESTA, which is 125 Rydberg in current simulations. The time stepping of $\tau = h_x^2/4.5$ is used, where h_x is the grid spacing at the x direction. Finally, $\gamma = 0.0065 \text{ kcal}/(\text{mol}\text{\AA}^2)$ obtained from the previous work in Chapter 2 is applied to compute the total non-polar solvation energy.

Table 3.5 summarizes the numerical results of the solvation free energies of 24 molecules. The root mean square error (RMS) of 1.31kcal/mol is obtained, which indicates a very good agreement between the present prediction and experimental data [237]. The agreement can also be seen from Figure 3.2. Moreover, to examine consistency, the ratio of $\Delta\Delta G_p$ and ΔG_{QM} is computed and listed in Table 3.6. It is evident that the numerical ratio is always about 0.6. Therefore, the reliability and consistency of the present model have been illustrated.

3.4.2 Solvation free energies of 16 molecules

Application of the set of 24 small molecules has shown that the proposed solvation model works well for the solvation free energy calculation of small molecules. One of the motivations in developing the present optimized surface model with the quantum charge density is to

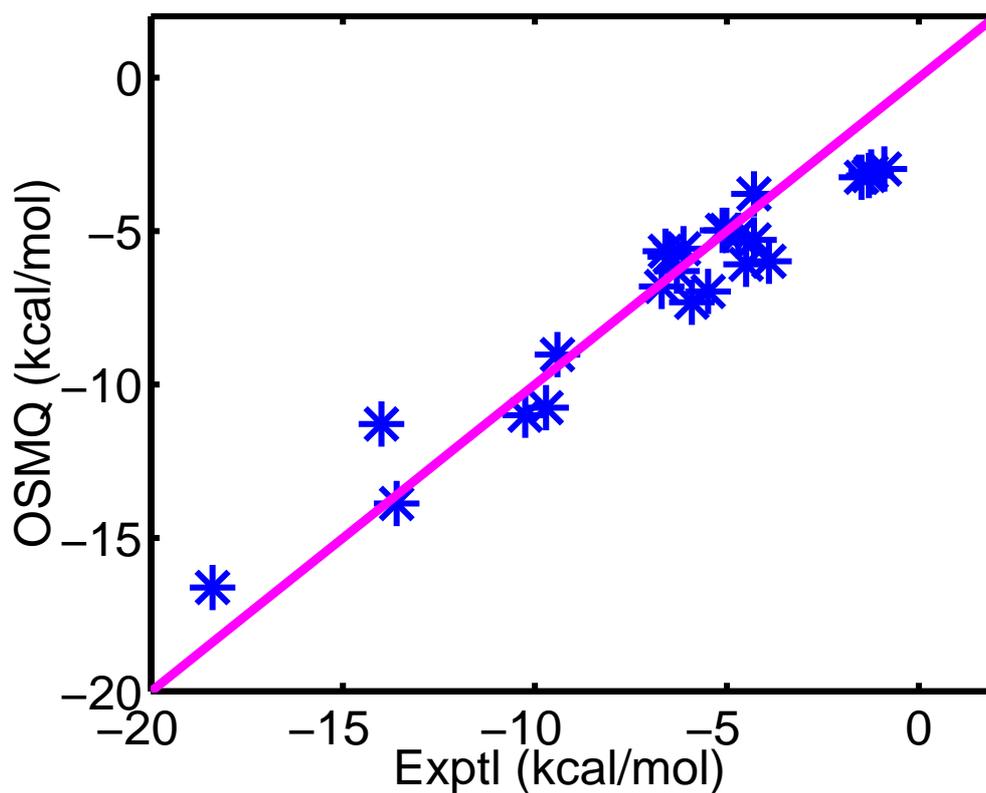


Figure 3.2: Correlation between experimental data and the present optimized surface model with quantum correction (OSMQ) in solvation free energies of 24 small molecules.

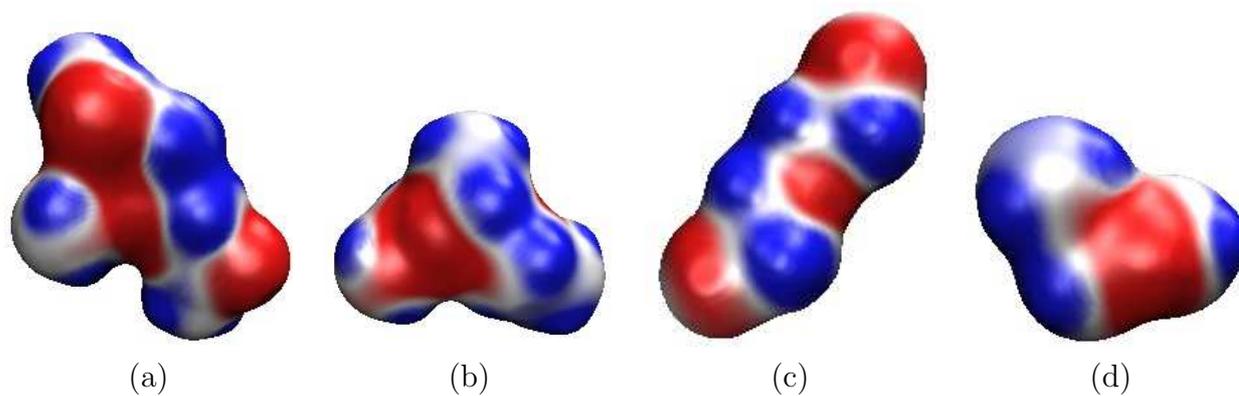


Figure 3.3: Illustration of surface electrostatic potentials of four small compounds at their corresponding isosurfaces $S = 0.50$. (a) Glycerol triacetate; (b) 1,1-diethoxyethane; (c) Bis-2-chloroethyl ether; (d) Dimethoxymethane.

Table 3.6: Solvation free energy (kcal/mol) decomposition for a set of 21 molecules.

Compound	ΔG_p^v	ΔG_p	$\Delta\Delta G_p$	ΔG_{QM}	$ \frac{\Delta G_{QM}}{\Delta\Delta G_p} $
Ethanol	-5.42	-6.33	-0.91	0.61	0.67
Propionamide	-8.84	-12.19	-3.35	2.14	0.64
Phenol	-6.30	-7.88	-1.58	1.01	0.64
Methanethiol	-3.44	-4.29	-0.85	0.50	0.59
Propionic acid	-6.31	-7.73	-1.42	0.87	0.62
Acetamide	-10.13	-14.23	-4.10	2.65	0.65
Acetonitrile	-6.04	-8.23	-2.19	1.36	0.62
Ethanethiol	-3.77	-4.53	-0.76	0.47	0.62
Aniline	-7.52	-9.34	-1.82	1.15	0.63
Acetic acid	-7.01	-8.67	-1.66	1.03	0.62
1-methylcytosine	-14.68	-23.90	-9.22	5.96	0.65
Pyridine	-5.3	-7.61	-2.31	1.49	0.64
9-methyladenine	-13.42	18.50	-5.08	3.17	0.62
1-methyluracil	-10.88	-15.35	-4.47	2.78	0.62
4-cresol	-6.40	-7.94	-1.54	1.00	0.65
4-Methylimidazole	-10.03	-15.88	-5.85	3.84	0.66
Methylethyl sulfide	-3.94	-4.83	-0.89	0.55	0.62
n-Butylamine	-5.96	-7.19	-1.23	0.81	0.66
3-Methylindole	-7.80	-10.40	-2.60	1.66	0.64
Methylamine	-6.10	-7.70	-1.60	1.07	0.67
Benzene	-3.72	-4.69	-0.97	0.52	0.53

deal with a relatively challenging set of compounds, which was studied by Nicholls *et al.* [160] and in Chapter 2 where the PB theory and fixed partial charges were used. This set is challenging to compute because of the existence of polyfunctional or interacting polar groups, which lead to strong solvent-solute interactions. The challenge has been illustrated numerically in the previous work [160, 46]. In particular, with the OpenEye-AM1-BCC v1 charge and corresponding optimized ZAP 9 radii, the root mean square error (RMS) obtained by Nicholls *et al.* is 1.71 ± 0.05 kcal/mol via an explicit solvent model. The smallest RMS error of their single - conformer Poisson-Boltzmann approach is 1.87 ± 0.03 kcal/mol [160]. By using our previous optimized surface model (OSM) with OpenEye-AM1-BCC v1 charges,

a better performance in the solvation calculation could be attained. However, RMS error was still 1.76 kcal/mol. Large errors from benzamides can not be avoided if both OpenEye-AM1-BCC v1 charge and ZAP 9 radii are used in the PB approaches. Errors from the calculation of benzamides were still between 3.5 and 4.0 kcal/mol.

Therefore, hoping possible improvement, we introduce quantum mechanical corrections to take care of the charge density. As we did before, structure data of this set of 16 molecules is taken from the paper of Nicholls *et al.* [160]. In particular, atomic coordinates are taken from their supporting information, which have been already optimized by using Gaussian03 package in vacuum with B3LYP/6-31G**. The atomic radii are still based on their new parametrization ZAP-9 and multiplied by a common factor 1.1. All other parameters needed in current model are set in the same way as described for the above set of 24 molecules. Note that in the previous papers [160, 46], the set contains 17 molecules. Here we remove a compound (benzyl bromide) because it has atomic species Br for which we can not obtain a proper pseudopotential file from the mentioned pseudopotential generator website. Since errors from the calculation of benzyl bromide previously was about 1 kcal/mol which is much lower than RMS, exclusion of benzyl bromide should make the RMS increase in previous results.

The results are summarized in Table 3.7 which lists the values of solvation free energies for different components and gives a comparison of total solvation free energies between calculated and experimental values [160]. It shows that the RMS error of 16 molecules from the present model is 1.50 kcal/mol, which is much better than that from the explicit method of 1.71 kcal/mol [160]. It is also better than our earlier OSM result of 1.76 kcal/mol in

Table 3.7: Comparison of free energies (kcal/mol) for 16 compounds.

Compound	G_{np}	ΔG_p	ΔG_{QM}	ΔG_{total}	Exptl
glycerol triacetate	2.24	-12.73	1.35	-9.13	-8.84
benzyl chloride	1.35	-6.29	0.82	-4.11	-1.93
m-bis(trifluoromethyl)benzene	2.24	-2.63	0.36	-0.03	1.07
N,N-dimethyl-p-methoxybenzamide	1.97	-11.24	1.85	-7.42	-11.01
N,N-4-trimethylbenzamide	1.86	-9.79	1.52	-6.41	-9.76
bis-2-chloroethyl ether	1.45	-6.34	0.55	-4.34	-4.23
1,1-diacetoxyethane	1.65	-8.47	0.92	-5.90	-4.97
1,1-diethoxyethane	1.50	-6.22	0.52	-4.20	-3.28
1,4-dioxane	1.00	-6.00	0.66	-4.35	-5.05
diethyl propanedioate	1.81	-9.08	0.89	-6.38	-6.00
dimethoxymethane	1.03	-5.11	0.48	-3.59	-2.93
ethylene glycol diacetate	1.59	-9.00	1.03	-6.38	-6.34
1,2-diethoxyethane	1.54	-6.85	0.61	-4.70	-3.54
diethyl sulfide	1.22	-4.32	0.41	-2.69	-1.43
phenyl formate	1.37	-6.91	0.89	-4.65	-4.08
imidazole	0.95	-14.10	3.29	-9.86	-9.81
RMS error				1.50	
Average error				1.08	

Chapter 2. Therefore, a conclusion can be reached that nontrivial improvement is made using the charge density directly computed from quantum mechanical calculations. Figure 3.3 depicts the surface electrostatic potentials of four compounds at their corresponding isosurfaces $S = 0.50$. These surface potential profiles correlate with the surface electron density distribution and chemical properties of the molecules.

Unfortunately, as shown in Table 3.7, errors from two amide compounds are still quite large. This problem leads us to further explore the source of their errors. Note that in this work, structural parameters are pre-determined and have not been optimized during the quantum calculation of the electronic density profile. Therefore, the aforementioned large errors must be due to the structural parameters. To prove this assumption, we carry out the

present solvation analysis of the 16 compounds based on the structural data provided from the Pubchem database (<http://pubchem.ncbi.nlm.nih.gov/>). Our new results are listed in Table 3.8. Indeed, errors from two amide compounds are significantly reduced. However, the RMS error of the set (i.e., 1.50 kcal/mol) is exactly the same as that computed by using ZAP-9 structural parameters [160], because larger errors from other compounds.

Note that our approach belongs to the so called “blind test” [160] in which the same set of atomic parameters is used for all compounds. Certainly, our errors can be further reduced if atomic parameters are chosen based on molecular information. As such, one is allowed to use different atomic parameters based the chemical constitution and function groups of a molecule. For example, carbons within the same molecule can have different atomic radii. However, such approaches can no longer be called a “blind test” as discussed by Nicholls *et al.* [160]

3.4.3 Solvation free energies of 3 larger molecules

The overall accuracy of the proposed model has been examined by the above two sets of small molecules. Reliability, robustness and consistency are numerically illustrated as well. As far as efficiency is concerned, the proposed model is expected to be slower than traditional continuum models in the solvation analysis due to the involvement of QM calculations. However, the computational cost of the present model can be much less than the traditional QM calculation and existing quantum mechanical continuum solvation models. This is due to the following three reasons. First, implicit description of the solvent is adopted to dramatically reduce the number of degrees of freedom. Second, the time-consuming quantum calculation

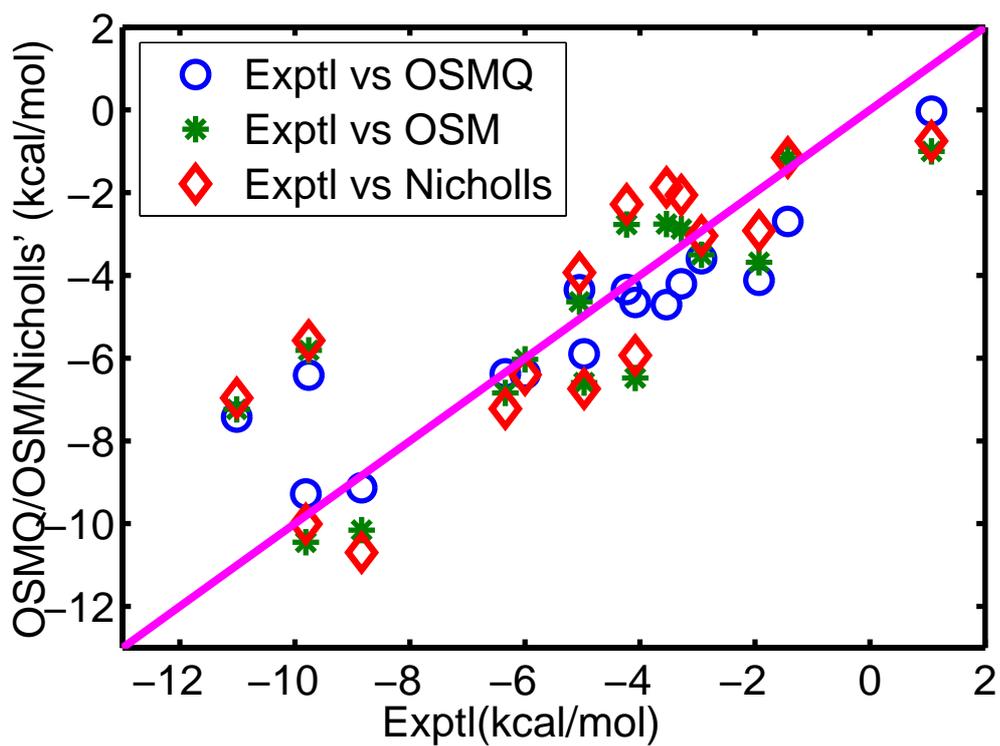


Figure 3.4: Correlation between experimental data [160] and the present optimized surface model with quantum mechanics (OSMQ) in solvation free energies of 16 compounds.

Table 3.8: Free energies (kcal/mol) for 16 compounds using structures from Pubchem data.

Compound	G_{np}	ΔG_p	ΔG_{QM}	ΔG_{total}	Error
glycerol triacetate	2.33	-13.01	1.44	-9.23	-0.39
benzyl chloride	1.34	-6.10	0.81	-3.94	-2.01
m-bis(trifluoromethyl)benzene	2.22	-3.29	0.41	-0.66	-1.73
N,N-dimethyl-p-methoxybenzamide	1.94	-12.00	2.04	-8.02	2.99
N,N-4-trimethylbenzamide	1.85	-10.59	1.77	-6.97	2.79
bis-2-chloroethyl ether	1.44	-5.85	0.58	-3.82	0.41
1,1-diacetoxyethane	1.66	-8.51	0.90	-5.94	-0.97
1,1-diethoxyethane	1.53	-6.07	0.58	-3.96	-0.68
1,4-dioxane	1.05	-5.13	0.49	-3.59	1.46
diethyl propanedioate	1.83	-11.78	1.62	-8.33	-2.33
dimethoxymethane	1.06	-4.61	0.34	-3.21	-0.28
ethylene glycol diacetate	1.68	-9.63	1.4	-6.91	-0.57
1,2-diethoxyethane	1.72	-5.68	0.44	-3.52	-0.02
diethyl sulfide	1.22	-4.57	0.51	-2.84	-1.41
phenyl formate	1.35	-6.61	0.90	-4.36	-0.28
imidazole	0.82	-13.16	2.97	-9.28	-0.53
RMS error					1.50
Average error					1.18

has been accelerated by the pseudopotential and minimum basis set in the framework of a linear scaling density functional theory in SIESTA. Furthermore, EPS charge fitting process, which depends on the definition of partial charges and the choices of sampling points, is avoided in this model by the direct use of the charge density. Therefore, it is believed that with powerful computer facilities, the current model can be a good choice to handle complex systems such as large drug molecules, amino acids as well as moderately large proteins. To examine the performance of this model for larger molecules, three molecules are chosen, including phorbol (54 atoms), Staurosporine (66 atoms, a potent protein kinase C inhibitor which enhances cAMP-mediated responses in human neuroblastoma cells), and phorbol12,13-dibutyrate (71 atoms, an effective activator of calcium-activated, phospholipid-

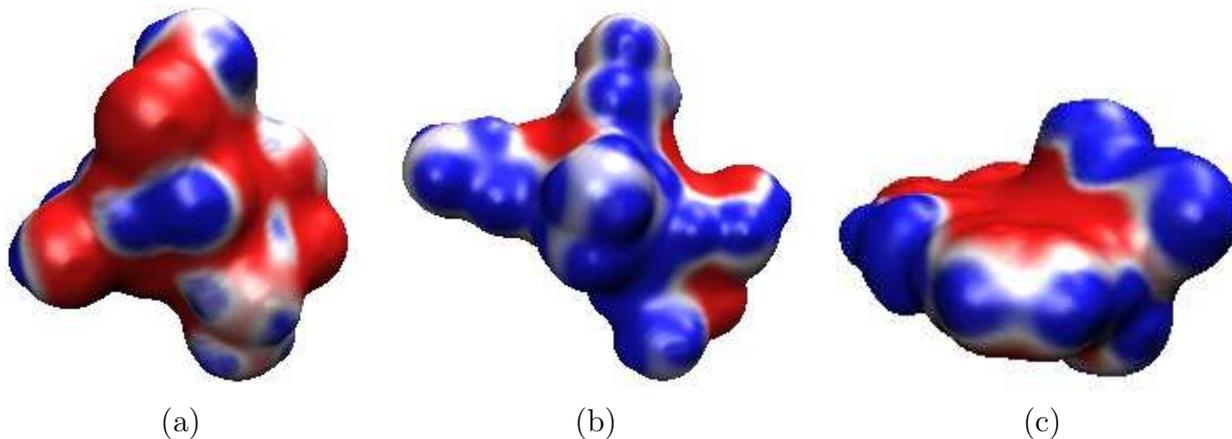


Figure 3.5: Illustration of surface electrostatic potentials of three compounds at their corresponding isosurfaces $S = 0.50$. (a) Phorbol; (b) Phorbol12,13-dibutyrate; (c) Staurosporine.

dependent protein kinase C). Solvation free energies of these molecules are computed and the computational cost is recorded with two quad-core Xeon 2.3 GHz processors. Our results are listed in Table 3.9. It is also shown that the ratios of $\Delta\Delta G_p$ and ΔG_{QM} are still about 0.6, which is consistent with those in smaller molecule calculations. In Figure 3.5, we illustrate the surface electrostatic potentials for the three compounds. The isosurface $S = 0.50$ is chosen for the plot. As shown in Chapter 2, different isosurfaces may exhibit different electrostatic characteristic.

Table 3.9: Solvation free energies (kcal/mol) of 3 large molecules and corresponding CPU time (hour).

Compound	$\Delta\Delta G_p$	ΔG_{QM}	$ \frac{\Delta G_{QM}}{\Delta\Delta G_p} $	G_{np}	ΔG_{total}	Time
phorbol	-3.02	1.94	0.64	3.24	-15.53	0.44
phorbol12,13-dibutyrate	-4.94	2.91	0.59	4.60	-17.43	0.97
Staurosporine	-8.39	4.78	0.57	3.79	-18.53	0.82

3.5 Chapter conclusions

The polar solvation energy was computed by the Poisson-Boltzmann equation with partial charges adopted from molecular mechanical force fields. Generally, partial charges in force fields are parameterized for certain class of molecules and may not be accurate for other molecules. Additionally, the fixed charge pattern does not describe the charge arrangement during the solvation process. This drawback limits the accuracy and utility of our earlier solvation models. The present work addresses such a limitation of our differential geometry based solvation models by the incorporation of quantum density into our earlier models.

We believe that solvation is subject to the fundamental law of physics. As such, all the important state functions, including the quantum density of the solute, must be determined by a multiscale total free energy functional, which has the contribution from the quantum mechanical energy of electrons at the finest scale. Therefore, we construct a new multiscale total free energy functional which includes the electron kinetic energies and potential energies. Apart from the earlier two governing equations, i.e., the generalized Poisson-Boltzmann equation for the electrostatic potential and the generalized Laplace-Beltrami equation for the solvent-solute boundary, one additional governing equation, the Kohn-Sham equation for electronic structures, is also derived from the total energy functional by the Euler-Lagrange variation. The solution of these three governing equations give rise to the desirable minimal free energy of solvation.

Numerical methods and algorithms are discussed for the solution of three coupled partial different equations (PDEs). The Poisson-Boltzmann and the Laplace-Beltrami equations (i.e., a generalized Laplace-Beltrami equation) are solved in a similar manner as that in Chapter 2. The Kohn-Sham equation is solved twice, once for the solute in vacuum and

the other in solution. By means of the reaction field potential, we can relate our Kohn-Sham Hamiltonian in both situations to the standard one, so that existing computational software packages can be utilized. The present work has developed a protocol to make use of SIESTA (Spanish initiative for the electronic structure of thousands of atoms), an efficient linear scaling DFT package, for the solution of the electronic density. Appropriate iteration procedures are developed to dynamically couple three governing equations and ensure the convergence of the solution.

The present multiscale model is validated by the solvation analysis of realistic molecules whose experimental solvation free energies are available. We have particularly analyzed the stability and consistency of the present model when atomic partial charges in our previous Poisson solver are replaced by the continuous density distribution.

Chapter 4

Lagrangian formulation

This chapter presents a Lagrangian formulation of our differential geometry based solvation models. Therefore, we consider the solvent-solute boundary as a 2-dimensional (2D) differentiable manifold embedded in a 3D Euclidean space or a hypersurface in a Riemannian manifold. Then the subsequent free energy optimization can be carried out on the 2D manifold. For example, the area of a solvent-solute interface is modeled as a surface integration over the biomolecular manifold. The Lagrangian representation of biomolecular surfaces has a few utilities/advantages. First, it provides an essential basis for biomolecular visualization, surface electrostatic potential map and visual perception of biomolecules. Additionally, it is consistent with the conventional setting of implicit solvent theories and thus, many existing theoretical algorithms and computational software packages can be directly employed. Finally, the Lagrangian representation does not need to resort to artificially enlarged van der Waals radii as often required by the Eulerian representation in solvation analysis.

The main goal of the present chapter is to explore the Lagrangian formulation of the differential geometry based solvation model, and analyze the similarity and difference of two

differential geometry based formulations. The mathematical analysis of biomolecular surfaces in the Lagrangian formulation is quite different from that of the Eulerian formulation. The Lagrangian analysis of biomolecular surfaces makes the direct use of differential geometry theory of surfaces and manifolds. The surface minimization leads to the Laplace-Beltrami operator, or the mean curvature operator. Whereas the Eulerian analysis of biomolecular surfaces utilizes the coarea theorem of the geometric measure theory. The resulting operator from surface area minimization can also be identified as a generalized Laplace Beltrami operator in a higher dimension. The connection of two representations is analyzed in this chapter. The structure of governing equations, and the accuracy and efficiency of two formulations are compared.

The rest of this chapter is organized as follows. In Section 4.1, we present the Lagrangian formulation of differential geometry based solvation models. The total free energy functional of solvation is constructed, and followed by detailed analysis of on-manifold variations. Such variations produce coupled potential driven geometric flow and Poisson-Boltzmann equations. Section 4.2 is devoted to the computational methods and algorithms. We discuss different realizations of biomolecular surfaces, including direct on-manifold evolution and Eulerian embedding approaches. Algorithms for Eulerian and Lagrangian inter converting are also discussed. A solution strategy for the coupled equations is designed and analyzed. The proposed PDE methods are of second order in convergence. Method for solvation free energy calculations is also given. The proposed methods and algorithms are validated in Section 4.3. A large number of numerical examples are designed to test the numerical accuracy, convergence order and the computational efficiency of numerical methods and algorithms. The proposed differential geometry based solvation model is applied to two classes of prob-

lems in Section 4.4. First, we examine the solvation free energy calculation of a set of small molecules. We then extend our solvation calculations to a set of proteins. Results are compared with experimental data, those obtained by using our earlier Eulerian formulation and those obtained by the classic molecular surface definition. Finally, we consider two cases of salt-regulated protein-protein interactions. The protein binding affinities are computed by using the proposed new models. The resulting binding affinities compare well with experimental data in the literature.

4.1 Theory and model

This section presents the differential geometry based solvation model. We first discuss a few free energy functionals of solvation. These functionals may appear exactly the same as the solvation energy expressions in the classic theory of solvation. However, a fundamental difference is that, in the present approach, the solvent-solute interface will be obtained by the variational principle, rather than by a prefixed surface, such as the van der Waals surface or the molecular surface. The governing equations for the solvation system, including the Poisson-Boltzmann equation and the generalized Laplace Beltrami equation, will be derived from the first variation.

4.1.1 Solvation free energy functionals

This subsection presents a few solvation models and establishes the notations for the further development of differential geometry based solvation analysis. A polar solvation model is described before a nonpolar solvation model is given. The total solvation free energy functional is designed as the combination of the polar and nonpolar components.

4.1.1.1 Polar solvation functional

The free energy functional of the electrostatic system was given exactly by Sharp and Honig [199], and Gilson et al [97]. A sharp solvent-solute interface is assumed in their free energy expression

$$G_{\text{p}} = \int_{\Omega} \left(\lambda_m \rho_m \phi - \frac{1}{2} \epsilon |\nabla \phi|^2 - k_B T \lambda_s \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \right) d\mathbf{r} \quad (4.1)$$

where the involved notations are defined as those in Eulerian formulation, while the permittivity $\epsilon(\mathbf{r})$ (also called dielectric coefficient) and the ionic function $\lambda_m(\mathbf{r})$ and $\lambda_s(\mathbf{r})$ are defined as [7, 143, 93]

$$\epsilon(\mathbf{r}) = \epsilon_m \lambda_m + \epsilon_s \lambda_s \quad (4.2)$$

and

$$\lambda_m(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in \Omega_m \\ 0 & \mathbf{r} \in \Omega_s \end{cases}, \quad \lambda_s(\mathbf{r}) = \begin{cases} 0 & \mathbf{r} \in \Omega_m \\ 1 & \mathbf{r} \in \Omega_s \end{cases} \quad (4.3)$$

where, the computational domain is divided into two subdomains, $\Omega = \Omega_m \cup \Omega_s$, with Ω_m and Ω_s denoting the solute and solvent accessible regions, respectively. The domains Ω_m and Ω_s are separated by an interface Γ . Here, $\epsilon_m = \epsilon_0 \epsilon_m$ and $\epsilon_s = \epsilon_0 \epsilon_s$ are the permittivities of the macromolecule and the solvent, respectively, where ϵ_0 is the permittivity of vacuum and ϵ_α ($\alpha = m, s$) are relative permittivities. We treat ϵ_α as constants. Note that in the classic PB theory, these functions explicitly depend only on the radii of the solute [7, 93] — they do not depend on Γ . However, in the present theory, these functions not only implicitly depend on the radii of the solute, but also explicitly depend on the interface Γ , which in

turn, depends on the total energy functional.

4.1.1.2 Total free energy functional of solvation

The electrostatic free energy functional is complemented by nonpolar free energy functional described in Section 1.1.4 to give the total free energy functional of solvation for biomolecules at equilibrium

$$G_{\text{total}} = \gamma \cdot \text{Area} + p \cdot \text{Vol} + \int_{\Omega_s} \rho_s U_{ss} d\mathbf{r} + \int_{\Omega_m} \left(\rho_m \phi - \frac{\epsilon_m}{2} |\nabla \phi|^2 \right) d\mathbf{r} - \int_{\Omega_s} \left(\frac{\epsilon_s}{2} |\nabla \phi|^2 + k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \right) d\mathbf{r}, \quad (4.4)$$

where we have used the coefficient definitions provided in Eqs. 4.2 and 4.3 to simplify the integrals. This total free energy functional might be subject to a variety of corrections and modifications.

4.1.2 Surface variation

The total free energy functional (4.4) is an important component of the present differential geometry based solvation model. However, it does not provide a protocol for practical solvation analysis. This subsection describes the variation principle which leads to desirable governing equations for the surface formation and evolution, and for the evaluation of the electrostatic potential.

In our previous minimal molecular surface (MMS) model [18, 20], the surface variation was accomplished via the Euler-Lagrange equation in the Eulerian representation. In the present Lagrangian representation, we perform the optimization of G_{total} on manifold Ξ

with respect to the solvent-solute interface Γ . In the spirit of differential geometry, the interface can be represented as a closed surface in the 3D Euclidean space and denoted as $\Gamma(u_1, u_2)$, which depends on the two real parameters u_1 and u_2 . The solute region, $\Omega_m(\Gamma)$ and the solvent region, $\Omega_s(\Gamma)$ can be regarded as functions of $\Gamma(u_1, u_2)$. We use $\frac{\delta(\cdot)}{\delta\Gamma}$ to denote the first variation of (\cdot) with respect to surface definition Γ ,

$$\begin{aligned} \frac{\delta G_{\text{total}}}{\delta\Gamma} = & \frac{\delta}{\delta\Gamma} \left[\gamma \cdot \text{Area} + p \cdot \text{Vol} + \int_{\Omega_s} \rho_s U_{ss} d\mathbf{r} + \int_{\Omega_m} \rho_m \phi d\mathbf{r} - \frac{1}{2} \int_{\Omega_s} \epsilon_s |\nabla\phi|^2 d\mathbf{r} \right. \\ & \left. - \frac{1}{2} \int_{\Omega_m} \epsilon_m |\nabla\phi|^2 d\mathbf{r} - \int_{\Omega_s} k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) d\mathbf{r} \right]. \end{aligned} \quad (4.5)$$

We set $\frac{\delta G_{\text{total}}}{\delta\Gamma} = 0$ to construct the governing equation that describes the optimized solvent-solute interface. To carry out this variation, we express the surface area and volume as the following integrals

$$\text{Area} = \int_{\Xi} d\sigma \quad \text{and} \quad \text{Vol} = \int_{\Omega_m} d\mathbf{r}, \quad (4.6)$$

where $d\sigma$ represents the infinitesimally small surface element on the solute-solvent interface. The complete first variation formula of G can be obtained by adding up the variation of each term from Eq. (4.5). To this end, we consider a surface element $\mathbf{f}(u_1, u_2)$ and its infinitesimal displacement in the normal direction

$$\mathbf{f}^{(\varepsilon)}(u_1, u_2) := \mathbf{f}(u_1, u_2) + \varepsilon \varphi(u_1, u_2) \cdot \mathbf{N}(u_1, u_2) \quad (4.7)$$

where \mathbf{N} is the outward unit normal direction and φ is an arbitrary C^2 function. In other words, we consider a one-parameter family $\mathbf{f}^{(\varepsilon)}$ of surface elements and the unperturbed

surface element is a particular case, $\mathbf{f} = \mathbf{f}^{(0)}$. The tangent vectors of $\mathbf{f}^{(\varepsilon)}$ are given by

$$\frac{\partial \mathbf{f}^{(\varepsilon)}}{\partial u_i} = \frac{\partial \mathbf{f}}{\partial u_i} + \varepsilon \frac{\partial \varphi}{\partial u_i} \mathbf{N} + \varepsilon \varphi \frac{\partial \mathbf{N}}{\partial u_i}. \quad (4.8)$$

To analyze the impact of the perturbation, we examine the first fundamental form of differentiable manifolds

$$\begin{aligned} g_{ij}^{(\varepsilon)} &= \left\langle \frac{\partial \mathbf{f}^{(\varepsilon)}}{\partial u_i}, \frac{\partial \mathbf{f}^{(\varepsilon)}}{\partial u_j} \right\rangle \\ &= g_{ij} + 2\varepsilon \varphi \left\langle \frac{\partial \mathbf{f}}{\partial u_i}, \frac{\partial \mathbf{N}}{\partial u_j} \right\rangle + \varepsilon^2 \left(\varphi^2 \left\langle \frac{\partial \mathbf{N}}{\partial u_i}, \frac{\partial \mathbf{N}}{\partial u_j} \right\rangle + \frac{\partial \varphi}{\partial u_i} \frac{\partial \varphi}{\partial u_j} \right) \\ &= g_{ij} - 2\varepsilon \varphi h_{ij} + O(\varepsilon^2). \end{aligned} \quad (4.9)$$

We therefore have

$$\left. \frac{\partial g_{ij}^{(\varepsilon)}}{\partial \varepsilon} \right|_{\varepsilon=0} = -2\varphi h_{ij}. \quad (4.10)$$

By virtue of surface elements $f^{(\varepsilon)}$ and the first fundamental form, we can carry out the

surface variation of the area as follows

$$\begin{aligned}
\frac{\delta A}{\delta \Gamma} &= \frac{\delta}{\delta \Gamma} \left(\int_{\Xi} d\sigma \right) \\
&= \frac{\delta}{\delta \Gamma} \left(\int_U \sqrt{g} du_1 du_2 \right) \\
&= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\int_U \sqrt{g^{(\varepsilon)}} du_1 du_2 \right) \\
&= \left(\int_U \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \sqrt{g^{(\varepsilon)}} du_1 du_2 \right) \\
&= \int_U \frac{1}{2\sqrt{g}} \left(g_{22} \frac{\partial g_{11}^{(\varepsilon)}}{\partial \varepsilon} \Big|_{\varepsilon=0} + g_{11} \frac{\partial g_{22}^{(\varepsilon)}}{\partial \varepsilon} \Big|_{\varepsilon=0} - 2g_{12} \frac{\partial g_{12}^{(\varepsilon)}}{\partial \varepsilon} \Big|_{\varepsilon=0} \right) du_1 du_2 \\
&= - \int_U \varphi \frac{1}{g} (h_{11}g_{22} + h_{22}g_{11} - 2h_{12}g_{12}) \sqrt{g} du_1 du_2 \\
&= - \int_U \varphi 2H \sqrt{g} du_1 du_2 \\
&= - \int_{\Xi} \varphi 2H d\sigma,
\end{aligned} \tag{4.11}$$

where g is the Gram determinant, $g = \text{Det}(g_{ij}) = g_{11}g_{22} - g_{12}^2$ and H the mean curvature $H = \frac{1}{2g}(h_{11}g_{22} + h_{22}g_{11} - 2h_{12}g_{12})$, where g_{ij} and h_{ij} are defined in Appendix C.

The first variation of the volume enclosed by the manifold Ξ is derived as follows. We first express the volume enclosed by the surface element $\mathbf{f}^{(\varepsilon)}$ as a Taylor expansion in terms of ε

$$\text{Vol}(\mathbf{f}^{(\varepsilon)}) = \text{Vol}(\mathbf{f}) + \varepsilon \frac{\delta \text{Vol}}{\delta \Gamma} + O(\varepsilon^2). \tag{4.12}$$

As shown in the above calculation, we can pursue the volume variation with respect to Γ by means of the variation with respect to ε

$$\frac{\delta \text{Vol}}{\delta \Gamma} = \frac{\partial \text{Vol}(\mathbf{f}^{(\varepsilon)})}{\partial \varepsilon} \Big|_{\varepsilon=0} = \frac{\partial (\text{Vol}(\mathbf{f}^{(\varepsilon)}) - \text{Vol}(\mathbf{f}))}{\partial \varepsilon} \Big|_{\varepsilon=0}. \tag{4.13}$$

It follows [24]

$$\begin{aligned}
\frac{\delta \text{Vol}}{\delta \Gamma} &= \left. \frac{\partial(\text{Vol}(f(\varepsilon)) - \text{Vol}(f))}{\partial \varepsilon} \right|_{\varepsilon=0} \\
&= \left. \frac{\partial \int_0^\varepsilon \int_U \sqrt{\text{Det} J} du_1 du_2 d\tau}{\partial \varepsilon} \right|_{\varepsilon=0} \\
&= \int_U \sqrt{\text{Det} J} du_1 du_2 \Big|_{\tau=\varepsilon=0} \\
&= \int_U \varphi \sqrt{g} du_1 du_2 \\
&= \int_{\Xi} \varphi d\sigma,
\end{aligned} \tag{4.14}$$

where matrix J is defined as

$$J = \begin{pmatrix} \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_1}, \frac{\partial \mathbf{f}(\tau)}{\partial u_1} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_1}, \frac{\partial \mathbf{f}(\tau)}{\partial u_2} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_1}, \frac{\partial \mathbf{f}(\tau)}{\partial \tau} \right\rangle \\ \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_2}, \frac{\partial \mathbf{f}(\tau)}{\partial u_1} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_2}, \frac{\partial \mathbf{f}(\tau)}{\partial u_2} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_2}, \frac{\partial \mathbf{f}(\tau)}{\partial \tau} \right\rangle \\ \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial \tau}, \frac{\partial \mathbf{f}(\tau)}{\partial u_1} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial \tau}, \frac{\partial \mathbf{f}(\tau)}{\partial u_2} \right\rangle & \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial \tau}, \frac{\partial \mathbf{f}(\tau)}{\partial \tau} \right\rangle \end{pmatrix}, \tag{4.15}$$

and similarly, $\mathbf{f}(\tau) = \mathbf{f} + \tau\varphi \cdot \mathbf{N}$, $\tau \in (0, \varepsilon)$. We have carried out the calculation of inner products

$$\left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_i}, \frac{\partial \mathbf{f}(\tau)}{\partial \tau} \right\rangle \Big|_{\tau=0} = \left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_i}, \varphi \mathbf{N} \right\rangle \Big|_{\tau=0} = 0, \tag{4.16}$$

where $\left\langle \frac{\partial \mathbf{f}(\tau)}{\partial u_i}, \varphi \mathbf{N} \right\rangle$ vanishes since $\frac{\partial \mathbf{f}(\tau)}{\partial u_i}$ and \mathbf{N} are orthogonal to each other. Moreover, we have

$$\left\langle \frac{\partial \mathbf{f}(\tau)}{\partial \tau}, \frac{\partial \mathbf{f}(\tau)}{\partial \tau} \right\rangle = \varphi^2, \tag{4.17}$$

where we use the fact that $\frac{\partial \mathbf{f}(\tau)}{\partial \tau} = \varphi \mathbf{N}$. Therefore, we can compute the determinant of the

J matrix as

$$\text{Det}J|_{\tau=0} = \text{Det} \begin{pmatrix} \left\langle \frac{\partial \mathbf{f}}{\partial u_1}, \frac{\partial \mathbf{f}}{\partial u_1} \right\rangle & \left\langle \frac{\partial \mathbf{f}}{\partial u_1}, \frac{\partial \mathbf{f}}{\partial u_2} \right\rangle & 0 \\ \left\langle \frac{\partial \mathbf{f}}{\partial u_2}, \frac{\partial \mathbf{f}}{\partial u_1} \right\rangle & \left\langle \frac{\partial \mathbf{f}}{\partial u_2}, \frac{\partial \mathbf{f}}{\partial u_2} \right\rangle & 0 \\ 0 & 0 & \varphi^2 \end{pmatrix} = \varphi^2 g. \quad (4.18)$$

Moreover, the above derivation process can be extended to the first variation of a general volume integral $\int_{\Omega_m} F(\mathbf{r}) d\mathbf{r}$.

$$\begin{aligned} \frac{\delta \left(\int_{\Omega_m} F d\mathbf{r} \right)}{\delta \Gamma} &= \left. \frac{\partial \int_0^\varepsilon \int_U F(u_1, u_2, \tau) \sqrt{\text{Det}J} du_1 du_2 d\tau}{\partial \varepsilon} \right|_{\varepsilon=0} \\ &= \int_{\Xi} F(u_1, u_2) \varphi d\sigma \end{aligned} \quad (4.19)$$

where F represents a general integrable function of u_1 and u_2 , and is defined in the whole domain. Similar to the volume variation, we end up with a surface integral.

Furthermore, because $\int_{\Omega} F d\mathbf{r}$ is independent of the surface variation, we have

$$\frac{\delta \left(\int_{\Omega} F d\mathbf{r} \right)}{\delta \Gamma} = 0. \quad (4.20)$$

Then, it follows that

$$\frac{\delta \left(\int_{\Omega_s} F d\mathbf{r} \right)}{\delta \Gamma} = \frac{\delta \left(\int_{\Omega} F d\mathbf{r} - \int_{\Omega_m} F d\mathbf{r} \right)}{\delta \Gamma} = - \int_{\Xi} F \varphi d\sigma. \quad (4.21)$$

Therefore, the first variation of all other volume integration terms in Eq. (4.5) can be

attained by replacing F with appropriate integrands in Eq. (4.19) and Eq. (4.21)

$$\frac{\delta \left(\int_{\Omega_m} \rho_m \phi d\mathbf{r} \right)}{\delta \Gamma} = \int_{\Xi} \rho_m \phi d\sigma, \quad (4.22)$$

$$\frac{\delta \left(\int_{\Omega_s} \epsilon_s |\nabla \phi|^2 d\mathbf{r} \right)}{\delta \Gamma} = - \int_{\Xi} \epsilon_s |\nabla \phi|^2 \varphi d\sigma, \quad (4.23)$$

$$\frac{\delta \left(\int_{\Omega_m} \epsilon_m |\nabla \phi|^2 d\mathbf{r} \right)}{\delta \Gamma} = \int_{\Xi} \epsilon_m |\nabla \phi|^2 \varphi d\sigma, \quad (4.24)$$

$$\frac{\delta \left(\int_{\Omega_s} \rho_s U_{ss} d\mathbf{r} \right)}{\delta \Gamma} = - \int_{\Xi} \rho_s U_{ss} \varphi d\sigma \quad (4.25)$$

and

$$\begin{aligned} & \frac{\delta \left(\int_{\Omega_s} k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) d\mathbf{r} \right)}{\delta \Gamma} \\ &= - \int_{\Xi} k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \varphi d\sigma, \end{aligned} \quad (4.26)$$

where $\nabla \phi$ on the surface represents the first derivative of potential ϕ .

Substituting Eqs. (4.11), (4.14), (4.22) (4.23), (4.24), (4.25) and (4.26) into Eq. (4.5) yields

$$\begin{aligned} \frac{\delta G_{\text{total}}}{\delta \Gamma} &= \int_{\Xi} \left[-2\gamma H + p - \rho_s U_{ss} + \rho_m \phi + \frac{1}{2} \epsilon_s |\nabla \phi|^2 - \frac{1}{2} \epsilon_m |\nabla \phi|^2 \right. \\ &\quad \left. + k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \right] \varphi d\sigma = 0. \end{aligned} \quad (4.27)$$

Since φ is an arbitrary function, the following condition must be satisfied for each point on the optimized interface

$$\begin{aligned}
W_n &\equiv -2\gamma H + p - \rho_s U_{ss} + \rho_m \phi + \frac{1}{2}\epsilon_s |\nabla\phi|^2 - \frac{1}{2}\epsilon_m |\nabla\phi|^2 \\
&\quad + k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right) \\
&= 0
\end{aligned} \tag{4.28}$$

It is noted that this condition was also obtained in our earlier work, but was derived in the Eulerian representation using different mathematical techniques [243, 46].

4.1.3 Governing equations

The directional derivative of solvation free energy functional G in the direction of a normal variation φ can be expressed as

$$D_\varphi G_{\text{total}}(\mathbf{f}) = \left. \frac{\partial G_{\text{total}}(\mathbf{f}^{(\varepsilon)})}{\partial \varepsilon} \right|_{\varepsilon=0} = \int_{\Xi} W_n \varphi d\sigma. \tag{4.29}$$

If we choose $\varphi(u_1, u_2) = -W_n$, then

$$D_\varphi G_{\text{total}}(\mathbf{f}) = - \int_{\Xi} W_n^2 d\sigma \leq 0$$

. This means that the total free energy decreases along the normal direction when $\varphi(u_1, u_2) = -W_n$ until it reaches a local minimal. Therefore the evolution $\mathbf{f}^{(\varepsilon)} = \mathbf{f} - \varepsilon W_n \mathbf{N}$ leads to a steady state and associated solvent-solute interface with strictly smaller energy. This analysis motivates the following potential driven geometric flow equation for the optimization of

the solute-solvent interface

$$\frac{\partial \mathbf{X}}{\partial t} = -W_n \mathbf{N}, \quad (4.30)$$

where $\mathbf{X} \in \Xi \subset \mathbb{R}^3$ is a position vector on the evolving manifold Ξ . Equation (4.30) is a Lagrange formulation of generalized geometric flows and its structure has been discussed in our earlier work [17]. This approach is computationally efficient but may have difficulties in handling topological changes during the biomolecule surface evolution. Numerical schemes for the solution of geometric flow equations is similar to that in Appendix A.

The electrostatic potential ϕ is governed by the Poisson-Boltzmann equation for traditional continuum biomolecular electrostatics applications. In the present approach, the Poisson-Boltzmann equation can be easily derived from the variation of the full free energy functional in Eq. (4.4) with respect to the electrostatic potential ϕ for a fixed interface Γ via the Euler Lagrange equation [199, 97, 243]

$$\frac{\delta G_{\text{total}}}{\delta \phi} = 0 \Rightarrow -\nabla \cdot (\epsilon(\mathbf{r}) \nabla \phi) - \lambda_s(\mathbf{r}) \sum_{i=1}^{N_c} Q_i n_i^0 e^{-\phi Q_i / k_B T} = \lambda_m \rho_m. \quad (4.31)$$

This Poisson-Boltzmann equation admits a sharp solvent-solute interface Γ and is quite different from the generalized Poisson-Boltzmann equation derived in Chapter 2, which has an optimized smooth surface (OSS).

Note that the Poisson-Boltzmann equation (4.31) and the potential driven geometric flow equation (4.30) are fully coupled. The existence and the uniqueness of their solution under the biomolecular context can be an interesting mathematical issue. In practice, these two equations have to be solved in a self-consistent manner. This aspect is discussed in 4.2.2. The solution of the Poisson-Boltzmann equation is subject to the far field boundary condi-

tion, which in practice can be approximated by the Dirichlet type of boundary conditions commonly used in the continuum electrostatic modeling (see Eq. (B.5)).

4.2 Methods and algorithms

This section discusses the solution strategies for a pair of coupled equations, the potential driven geometric flow and the Poisson-Boltzmann equation.

4.2.1 Interconversion between the Lagrangian and Eulerian representations

The generalized geometric flow equation (4.30) is in the Lagrangian representation which is well suited for boundary element or finite element type of methods. Although this Lagrangian formulation of geometric flow models is relatively easy to implement in many applications, such as surface smoothing, it can lead to computational difficulties in the case where there are topological changes, such as surface breaking or merging. These difficulties originate from the fact of singularity development on the manifold which is supposed to be smooth and differentiable. These topological changes commonly occur in biomolecular surface constructions and molecular dynamics applications. One way to overcome these obstacles is to use the Eulerian formulation [245, 18, 19, 20, 17]. The essential idea of resolving the difficulty of a “singular manifold” or non-smooth surface is to embed the problem in a higher dimensional space such that the embedded function is smooth and differentiable. Another way to avoid the difficulty of singular manifolds is to use a hybrid Eulerian-Lagrangian approach for the biomolecular surface generation. In such a hybrid approach, the surface evolution is carried out mostly in the Lagrangian representation, but is temporally switched to the

Eulerian representation whenever there is a singularity development on the manifold. In the rest of this subsection, we discuss computational tools for the interconversion between the Lagrangian and Eulerian representations.

4.2.1.1 Embedding the Lagrangian dynamics into the Eulerian representation

To embed a Lagrangian operator into its Eulerian representation, we introduce arbitrary hypersurface function $S(\mathbf{r})$ with $\mathbf{r} \in \mathbb{R}^3$. The earlier function $\chi(u_1, u_2)$ can be obtained by solving $S = 0$. For example, if $S = X_0x^2 + Y_0y^3 + Z_0z + D_0$, where, X_0, Y_0 and Z_0 are constants, and $z_0 \neq 0$, then one can set

$$\chi(t) = z = -x_0x^2 - y_0y^3 - d_0, \quad x_0 = \frac{X_0}{Z_0}, y_0 = \frac{Y_0}{Z_0}, d_0 = \frac{D_0}{Z_0}. \quad (4.32)$$

Via this example, it is easy to verify that the unit norm vector defined in Eq. (C.5) can also be expressed in term of S

$$\mathbf{N} = \frac{(-\chi_1, -\chi_2, 1)}{\sqrt{g}} = \frac{\nabla S}{\|\nabla S\|}. \quad (4.33)$$

Then the desired surface can be represented as a set of points with a constant value of function S

$$\Xi = \{\mathbf{r} | S(\mathbf{r}) = L\} \quad (4.34)$$

where L is an isosurface value. By the chain rule

$$\begin{aligned} \frac{\partial S}{\partial t} &= \frac{\partial S}{\partial \mathbf{X}} \cdot \frac{\partial \mathbf{X}}{\partial t} \\ &= \nabla S \cdot \frac{\partial \mathbf{X}}{\partial t} \\ &= -W_n \nabla S \cdot \mathbf{N}, \end{aligned} \quad (4.35)$$

where \mathbf{X} is a 3D position vector confined to the manifold Ξ . Due to Eq. (4.33), one has

$$\begin{aligned}\frac{\partial S}{\partial t} &= -\|\nabla S\|W_n \\ &= \|\nabla S\| \left[2\gamma H - p + \rho_s U_{ss} - \rho_m \phi - \frac{1}{2}\epsilon_s |\nabla \phi|^2 \right. \\ &\quad \left. + \frac{1}{2}\epsilon_m |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i/k_B T} - 1 \right) \right]\end{aligned}\quad (4.36)$$

where all terms should be expressed in terms of level surface function S . In particular, the surface mean curvature H must be rewritten in terms of S . The explicit form of mean curvature (C.7) implies the equality $2H = \nabla \cdot \mathbf{N}$ which gives

$$H = \frac{1}{2} \left[\frac{\partial}{\partial u_1} \left(\frac{\chi_1}{\sqrt{g}} \right) + \frac{\partial}{\partial u_2} \left(\frac{\chi_2}{\sqrt{g}} \right) \right] = \frac{1}{2} \nabla \cdot \left(\frac{\nabla S}{\|\nabla S\|} \right). \quad (4.37)$$

One can easily verify this relation by Eq. (4.32). Equation (4.37) connects the Lagrangian representation of the Laplace-Beltrami operator with its Eulerian representation. Eventually, the potential driven geometric flow equation in the Eulerian form is obtained for the optimized solvent-solute interface

$$\begin{aligned}\frac{\partial S}{\partial t} &= \|\nabla S\| \left[\nabla \cdot \left(\frac{\gamma \nabla S}{\|\nabla S\|} \right) - p + \rho_s U_{ss} - \rho_m \phi - \frac{1}{2}\epsilon_s |\nabla \phi|^2 \right. \\ &\quad \left. + \frac{1}{2}\epsilon_m |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i/k_B T} - 1 \right) \right] \\ &= \|\nabla S\| \left[\gamma \nabla \cdot \left(\frac{\nabla S}{\|\nabla S\|} \right) + V \right]\end{aligned}\quad (4.38)$$

where

$$V = -p + \rho_s U_{ss} - \rho_m \phi - \frac{1}{2} \epsilon_s |\nabla \phi|^2 + \frac{1}{2} \epsilon_m |\nabla \phi|^2 - k_B T \sum_{i=1}^{N_c} n_i^0 \left(e^{-\phi Q_i / k_B T} - 1 \right). \quad (4.39)$$

Equation (4.38) is the same as the generalized geometric flow equation derived in Chapter 2 using the characteristic function and geometric measure theory . This consistency lays the foundation for switches forwards and backwards between the Eulerian and Lagrangian representations and the development of hybrid methods for biomolecular surfaces. In fact, Eq. (4.38) has the same structure as that of our potential and curvature driven geometric flows and V is essentially the generalized potential defined in our earlier work [17]. Eq. (4.38) is subject to the similar boundary and initial conditions as those of geometric PDEs described in our earlier work [18, 20, 17].

4.2.1.2 Transform from the Lagrangian representation to the Eulerian representation

Usually the Lagrangian representation of surfaces is expressed in the form of triangulations [188]. To convert the Lagrangian representation of surfaces into the Eulerian representation, specifically, a Cartesian grid, we need to register the set of intersecting points between the surface and the Cartesian mesh. For the purpose of computing the electrostatic potential from the Poisson-Boltzmann equation which admits discontinuous coefficients across the solvent-solute interface, we have to calculate the norms at all of the intersecting points as well. A numerical algorithm for this Lagrangian-Eulerian transformation of molecular surfaces was developed by Zhou et al. [259], although the original paper did not provide implementation details.

We first set up a plane equation for each triangle on the molecular surface. For each plane equation, we compute all the intersection points between plane and the Cartesian mesh. For this set of intersecting points, we further record the subset that are located within the triangle. For this subset of intersecting points, we finally compute their norms as well by the second order finite difference scheme [259]. This algorithm has been extensively tested in our previous work [259, 252, 93, 20].

4.2.1.3 Transform from the Eulerian representation to the Lagrangian representation

Once the hypersurface function $S(x, y, z, T)$, where T is the stopping time, is obtained from the potential driven geometric flow equation (4.38) the solvent-solute interface can be extracted easily as an isosurface, $S(x, y, z, T) = L$, where isosurface value $L = (1 - \delta)S_0$. Here S_0 is set as the initial amplitude in geometric flow equation and $1 > \delta > 0$. In our earlier work [18, 20], we have chosen $S_0 = 1000$ and a small δ . Recently, in our differential geometry based multiscale models [243], we have designated S as characteristic function of the solute and chosen $0 \leq S \leq 1$. There is no need to specify L in such a formulation. In the present work, we choose $L = S_{\max}/2$, where S_{\max} is the maximum of S . This choice is computationally stable and delivers correct MMSs, when the potential term is absent. It is to point out that S here is quite different from the S used in our Eulerian formulation [46]. In the present work, S only serves as a hypersurface function for evolving and extracting the solvent-solute interface and can take any real value. Numerically, isosurface extraction can be done by existing software such as MATLAB. However, for further electrostatic analysis, we need a Cartesian representation of the interface locations and the associated norm values. Therefore, we construct a stand-alone algorithm to extract interface information. To this

end, the marching cubes method is adopted [140]. For a given grid partition, the marching cubes algorithm simply deals with a local meshing problem by processing each cell or cube independently. Each vertex of a cube can be either greater or less than the threshold value L , giving 256 different scenarios. In considering the symmetry and complementarity, there are only 15 canonical configurations in each cell needed to be addressed for the local meshing [140, 67] A look-up table is a quite efficient local triangulation or Cartesian algorithm. The marching cubes method can be modified in many ways to improve the accuracy, efficiency, robustness, and topological correctness. Auxiliary binary tree structures are typically employed in the range-space approaches, such as kd-tree method and interval tree method, to speed up the marching cubes method. For a structured grid dataset, geometric searching methods exploiting spatial coherence can be simpler and more efficient than the range space approaches.

To implement this scheme, first all points inside or outside the surface must be identified according to S value in the Cartesian grid domain. The surface must intersect those cube edges where one vertex is outside and the other is inside the surface. Therefore the surface intersection points and their normal directions can be approximated by linear interpolation. For instant, to compute an intersection point \mathbf{r}_o between an inside grid point \mathbf{r}_1 with value S_1 and an outside grid point \mathbf{r}_2 with value S_2 , the distance d between \mathbf{r}_o and \mathbf{r}_1 is calculated by

$$L = S_1 * (1 - d) + S_2 * d \quad (4.40)$$

$$d = \frac{S_1 - L}{S_1 - S_2}$$

where L is the isosurface function value. Obviously, with known positions \mathbf{r}_1 and \mathbf{r}_2 , distance

d determines the position of the intersection point \mathbf{r}_o . To calculate the norm vector at \mathbf{r}_o , we need to compute the normal vectors at \mathbf{r}_1 and \mathbf{r}_2 . In general, the normal direction of a grid point (x_i, y_j, z_k) can be estimated by

$$\begin{aligned} n_x(x_i, y_j, z_k) &= \frac{S_{i+1,j,k} - S_{i-1,j,k}}{2\Delta x} \\ n_y(x_i, y_j, z_k) &= \frac{S_{i,j+1,k} - S_{i,j-1,k}}{2\Delta y} \\ n_z(x_i, y_j, z_k) &= \frac{S_{i,j,k+1} - S_{i,j,k-1}}{2\Delta z} \\ \mathbf{n} &= (n_x, n_y, n_z) \end{aligned} \tag{4.41}$$

where n_x, n_y and n_z are the x, y and z components of the norm, respectively. Thus, the norm at the intersecting point \mathbf{r}_o , denoted as \mathbf{n}_o , can be interpolated through \mathbf{n}_1 and \mathbf{n}_2 , the norms of \mathbf{r}_1 and \mathbf{r}_2 , respectively

$$\mathbf{n}_o = (1 - d)\mathbf{n}_1 + d\mathbf{n}_2. \tag{4.42}$$

Clearly, the choice of $L = S_{\max}/2$ offers the best computational accuracy and stability. The unit norm \mathbf{N}_o at the intersecting point can be easily computed as $\mathbf{N}_o = \frac{\mathbf{n}_o}{\|\mathbf{n}_o\|}$. This algorithm is used in our calculation of unit norms at the intersecting points. Obviously, higher-order algorithms can be easily constructed when they are needed.

4.2.1.4 Numerical surface integral and volume integral in the Eulerian representation

Very often, we need to accurately carry out surface integration and volume integration in the Eulerian representation. The surface integral of a density function f can be approximated

by [207]

$$\int_{\Xi} f(x, y, z) d\sigma = \int_{\Omega} f(x, y, z) \delta(d(x, y, z)) d\mathbf{r} \approx \sum_{i,j,k} f(x_i, y_j, z_k) \tilde{\delta}_{i,j,k} h^3 \quad (4.43)$$

where (x_i, y_j, z_k) is the coordinate of grid point (i, j, k) , $d(x, y, z)$ is the distance of a point (x, y, z) defined in Ω from the interface Γ , h is the uniform grid size, and $f(x, y, z)$ is the surface density function defined on Γ . The delta function $\tilde{\delta}_{i,j,k}$ is given by

$$\tilde{\delta}_{i,j,k} = \tilde{\delta}_{i,j,k}^{(+x)} + \tilde{\delta}_{i,j,k}^{(-x)} + \tilde{\delta}_{i,j,k}^{(+y)} + \tilde{\delta}_{i,j,k}^{(-y)} + \tilde{\delta}_{i,j,k}^{(+z)} + \tilde{\delta}_{i,j,k}^{(-z)} \quad (4.44)$$

where $\tilde{\delta}_{i,j,k}^{(\pm\alpha)}$, $(\alpha = x, y, z)$ are discrete delta functions [207]. To carry out integration exactly on the interface, we use the following discrete surface integration formula [94]

$$\int_{\Xi} f(x, y, z) d\sigma \approx \sum_{(i,j,k) \in I} \left(f(x_o, y_j, z_k) \frac{|n_x|}{h} + f(x_i, y_o, z_k) \frac{|n_y|}{h} + f(x_i, y_j, z_o) \frac{|n_z|}{h} \right) h^3 \quad (4.45)$$

where (x_o, y_j, z_k) is the intersecting point of the interface and the x meshline that passes through (i, j, k) , and n_x is the x component of the normal vector at (x_o, y_j, z_k) . Similar relations exist between (x_i, y_o, z_k) and n_y , and (x_i, y_j, z_o) and n_z . Since Eq. (4.45) has already taken into account the contribution from irregular grid points outside the interface, the summation is restricted to I , the set of irregular grid points inside or on the interface [94]. The derivation of Eq. (4.45) is lengthy and is omitted here but it can be seen elsewhere [94]. The surface area can be calculated by setting $f = 1$ in Eq. (4.45). The error of the surface integration depends on the grid resolution and was observed to be approximately second-order convergence [94].

The volume integral of the density function f can be simply approximated by [94]

$$\int_{\Omega_m} f(x, y, z) d\mathbf{r} \approx \sum_{(i,j,k) \in J} f(x_i, y_j, z_k) h^3 \quad (4.46)$$

where the summation is over J , the set of points inside the interface.

4.2.2 Dynamic coupling of the Poisson-Boltzmann and geometric flow equations

The solution of potential driven geometric flow equation, or the generalized Laplace-Beltrami equation, is discussed in Appendix A, including many discretization schemes. In general, electrostatic energy is much larger than the non-electrostatic part so that the accuracy of electrostatic potential calculation based on the Poisson-Boltzmann (PB) equation plays a critical role in controlling the accuracy of the total solvation free energy. Therefore, numerical methods that are able to deliver highly accurate solution of the PB equation is desirable. In this work we apply matched interface and boundary method (MIB) to impose interface conditions for high accuracy. This challenge is addressed in Section 1.5.3.

As described earlier, optimized electrostatic potential ϕ is obtained by solving the Poisson-Boltzmann equation (4.31) in which solvent-solute interface Γ is used to determine ϵ and λ values. The interface Γ is generated by the solution of the potential driven geometric flow equation (4.38) which in turn depends on the electrostatic potential. Therefore, the geometric flow equation and the PB equation need to be solved simultaneously in the present differential geometry based solvation model. In practice, this coupled nonlinear system can be solved by an iterative procedure: First solving the PB equation with a fixed interface Γ for

ϕ ; Then obtaining the interface Γ from solving the potential driven geometry flow equation with a fixed potential ϕ . A more detailed algorithm is

1. Start with an initial solvent-solute interface, such as a solvent accessible surface.
2. Solve the Poisson-Boltzmann equation (4.31) for the potential with the initial solvent-solute interface.
3. Obtain new solvent-solute interface by solving the potential driven geometric flow equation (4.38) with the updated potential.
4. Calculate the solvation free energy with the resulting ϕ and Γ .
5. Go back to Step 2 until it converges.

The initial solvent-solute interface can be set by the solvent accessible surface with a probe radius of 1.4 Å. Another way to define the initial geometry is to use the interface obtained from solving the potential driven geometric flow equation (4.38) without the potential term V . Both approaches lead to the same result. In this study, we take the latter. The iteration will be stopped if the values of total solvation free energy converge to within a designated small criteria value which is 0.01kcal/mol for small molecules and 0.1kcal/mol for proteins in this chapter. Formulism of solvation free energy evaluation is similar to that in Chapter 2.

4.3 Validation

This section is devoted to the validation of the proposed differential geometry based solvation model and a number of computational algorithms used in the present work. The

overall accuracy of solvation free energy calculation depends on the reliability and accuracy of the solution of the geometric flow equation and the PB solver, surface and volume integrations, and the interface extraction process. The explicit Euler algorithm guarantees the reliability and convergence of the solution of the geometric flow equation. The finite central different scheme is of second order convergence in space and first-order in time although it is computationally expensive [17]. The MIBPB-III has been verified to be of second-order in convergence even with molecular surface singularities of proteins [93]. Therefore, it will be second order accurate for the present application. In fact, the biomolecular surfaces generated with geometric flows are free of geometric singularities, which is computationally easier.

We first examine the impact of the interaction potentials to the surface morphology, and surface electrostatic potentials. A few small molecular systems and 23 protein molecules are used in this examination. We then check the behavior of the surface area under different potential interactions. In particular, we verify whether the proposed minimal molecular surface (MMS) [20] provides the extreme surface area for various molecules and proteins. Finally, we investigate the convergence of the proposed iterative procedure for solving the coupled Poisson-Boltzmann and geometric flow equations.

4.3.1 Validation of the interface extraction scheme

The numerical algorithm for surface integrations has second-order convergence [207, 94]. However, the MIBPB-III here has been modified for our purpose to admit the present optimized molecular surface (OMS) as the solvent-solute interface definition. This implies that the reliability of the present MIBPB-III solver depends on the interface extracted by the

marching cube algorithm. Moreover, the performance of our surface integration and volume integration algorithms is also determined by the resulting interface instead of some pre-determined interface such as the molecular surface [185]. Therefore, it is worthwhile to validate the interface extraction procedure and algorithm in terms of surface area (\AA^2), surface enclosed volume (\AA^3) and electrostatic solvation free energy (kcal/mol). In general, there is no analytical result for electrostatic energy except for the one-atom system due to Kirkwood [119]. For the one-atom system without interaction potential, the resulting solvent-solute interface from the geometry flow evolution is a sphere with the same radius as the initial one, for which the PBE admits an analytical solution. The surface area and volume can be calculated analytically. Therefore, we consider a dielectric sphere of radii 2\AA with a unit charge at the center. We set $\frac{\rho}{\gamma} = 0.5$, $S_0 = 1000$ and $L = 500$. Table 4.1 lists the numerical results under different grid resolutions h , which are compared with the exact solution. The convergence in space is observed and a satisfactory result is attained.

Table 4.1: Comparison of surface areas (\AA^2), volumes (\AA^3) and energies (kcal/mol) for two small systems.

h	One Atom			Diatom		
	Area	Volume	Energy	Area	Volume	Energy
0.5	48.86	34.00	-84.92	95.24	71.00	-238.03
0.25	49.04	33.56	-82.92	99.28	72.73	-233.66
0.125	50.09	33.52	-82.08	100.5	73.20	-232.37
Referenced Value	50.265	33.510	-81.98	100.34	71.18	-233.67

Another test is done with a diatom system. It has been illustrated previously that molecular surface of a diatom system can be reproduced by our PDE based approach at an appropriate constant potential value [17]. In particular, when the van der Waals (VDW) radii of two atoms are 2\AA , the generated surface with $\frac{\rho}{\gamma} = -0.222$ will be almost identical visually

to the molecular surface with probe radius $r_m = 1.4\text{\AA}$. In this setting, the corresponding solvation free energy, surface area and volume calculated by our numerical procedure are compared with those based on molecular surface. To calculate the electrostatic solvation energy, a unit charge is set at the centers of two atoms $(-2.3,0,0)$ and $(2.3,0,0)$. The numerical results are summarized in Table 4.1. For a comparison, the reference molecular surface area and volume of this diatom system are obtained by using the MSMS program [188] with probe radius 1.4\AA and density 100. The electrostatic solvation energy $(-233.67\text{ kcal/mol})$ based on molecular surface is calculated by the original MIBPB-III [93]. A good agreement is also observed from this test.

4.3.2 Effect of interaction potentials

In this section, we illustrate the impact of potentials to the generation of solvent-solute interface, consequently to the solvation analysis. Since the effect of the pressure term has already been shown in our previous study, we focus our attention on the study of potential effects which include short-ranged repulsive interaction, long-ranged attractive dispersion interaction and electrostatic potential effect. Here we consider 12-6 Lennard-Jones pair potential to model V_i^{LJ} . For the purpose of demonstration, all the surface profiles here are constructed by using 6-12 decomposition and based on the geometric flow Eq. (4.36) in absence of the pressure term and the ionic effect

$$\frac{\partial S}{\partial t} = \sqrt{1 + \|\nabla S\|^2} \left[\nabla \cdot \left(\frac{\nabla S}{\sqrt{1 + \|\nabla S\|^2}} \right) + \frac{\rho_S}{\gamma} V^{\text{LJ}} - \left(\frac{1}{2} \frac{\epsilon_S}{\gamma} |\nabla \phi|^2 - \frac{1}{2} \frac{\epsilon_m}{\gamma} |\nabla \phi|^2 \right) \right]. \quad (4.47)$$

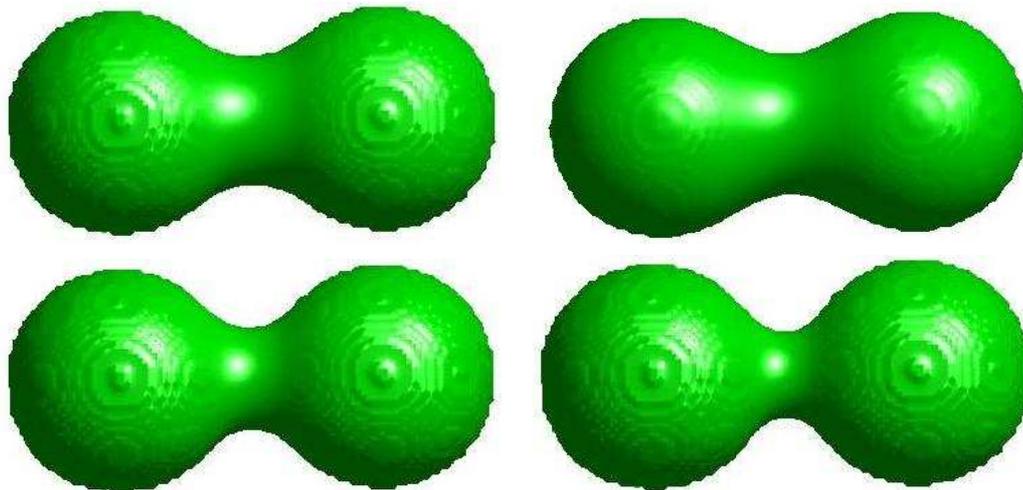


Figure 4.1: Illustration of surface morphology of a diatom system with radii 1.87\AA and coordinates $(x, y, z) = (-2.2, 0, 0)$, and $(2.2, 0, 0)$ under different solute-solvent interactions. Top Left: The MMS (no potential); Top Right: The surface obtained under a repulsive potential ($V_i^{\text{rep,LJ}}$); Bottom Left: The surface obtained under a full L-J potential (V^{LJ}); Bottom Right: The surface obtained under a full L-J potential and an electrostatic potential. It can be seen that the repulsive potential produces a “fat” surface, while an attractive potential or an electrostatic potential leads to a “slim” surface.

Without any potential term, this geometric flow equation leads to the minimal molecular surface (MMS) [20]. The effects of those three potentials are demonstrated by a diatom system, a four-atom system and finally a protein molecule which is also used to illustrate the potential impacts on surface electrostatic potential analysis. In the present computation, we have treated the solvent density ρ_s as homogeneous.

4.3.2.1 Surfaces of a diatom system

We first consider a diatom system with van der Waals radius 1.87\AA and coordinates $(x, y, z) = (-2.2, 0, 0)$ and $(2.2, 0, 0)$. Mesh size $h = 0.04\text{\AA}$ is used. The L-J parameters are set as follows: σ_i is taken from atomic radius and σ_s is chosen to be 0.65\AA . Well depth $\bar{\epsilon}_i = 0.035$ kcal/mol and bulk density coefficient $\frac{\rho_s}{\gamma} = 2$. To account for electrostatic

potential effect, a unit charge is set on the center of each atom and we choose $\frac{\epsilon_s}{\gamma} = 80 * \epsilon_s$ and $\frac{\epsilon_m}{\gamma} = 80 * \epsilon_m$. We use $\epsilon_m = 1$ and $\epsilon_s = 80$ for dielectric constants in solute and solvent, respectively. Figure 4.1 illustrates the different potential effects on the surface morphology for the diatom system. We systematically change the potential effects to generate different surfaces. We begin with no potential effect, which leads to the minimal molecular surface (Top Left of Figure 4.1), then add the repulsive part of the L-J potential (Top Right of Figure 4.1), then add an attractive interaction (Bottom Left of Figure 4.1) and finally add the electrostatic potential effect (Bottom Right of Figure 4.1). It can be seen that the repulsive potential produces a “fat” surface, while an attractive potential or electrostatic potential leads to a “slim” surface. In other words, with a purely repulsive interaction turning on, there is less bulk area between or around two spherical solutes, while more bulk area with attractive potential or electrostatic potential turning on. This result is consistent with experimental observations [236]

4.3.2.2 Surfaces of a four-atom system

The effects of potentials on the surface generation are further demonstrated by a four-atom system in Figure 4.2 with van der Waals radius 1.87 \AA and coordinates $(x, y, z) = (-3.40, 0, 0), (3.40, 0, 0), (0, -2.94, 0)$ and $(0, 2.94, 0)$. The needed parameters in Eq. (4.47) are set as the same as the above diatom system except for setting $1/2$ charge at the center of each atom. We also systematically change the potential effects by beginning with no potential which leads to the MMS in Figure 4.2(Top Left), then add the repulsive part of L-J potential (Top Right of Figure 4.2), then add the attractive part in Figure 4.2(Bottom Left) and finally add the electrostatic potential effect in Figure 4.2(Bottom Right). The impact of potentials is similar to that in the diatom system. It is found that the size of hole

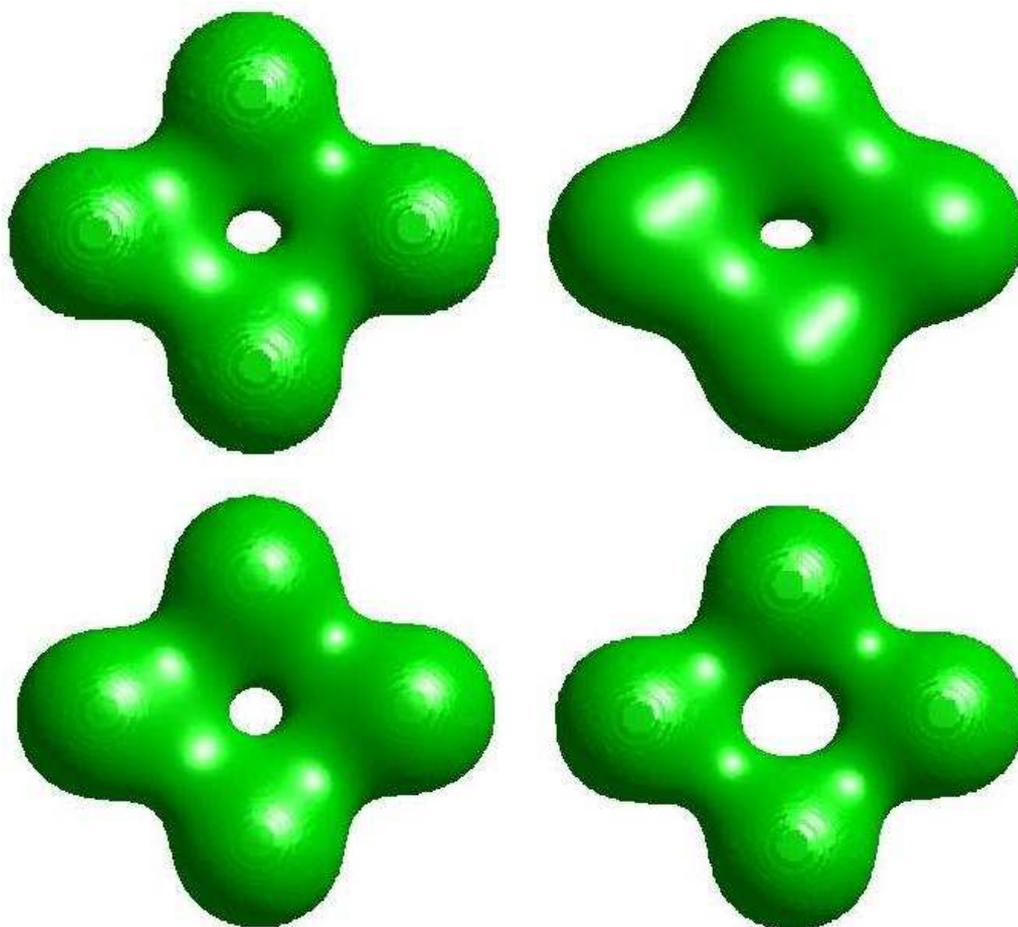


Figure 4.2: Illustration of surface morphology of a four-atom system with radii 1.87\AA and coordinates $(x, y, z) = (-3.40, 0, 0), (3.40, 0, 0), (0, -2.94, 0)$ and $(0, 2.94, 0)$ under different solute-solvent interactions. Top Right: The surface obtained under a repulsive potential ($V_i^{\text{rep,LJ}}$); Bottom Left: The surface obtained under the full L-J potential (V^{LJ}); Bottom Right: The surface obtained under the full L-J potential and an electrostatic potential. Again, it can be seen that the repulsive potential produces a “fat” surface; while an attractive potential or an electrostatic potential leads to a “slim” surface.

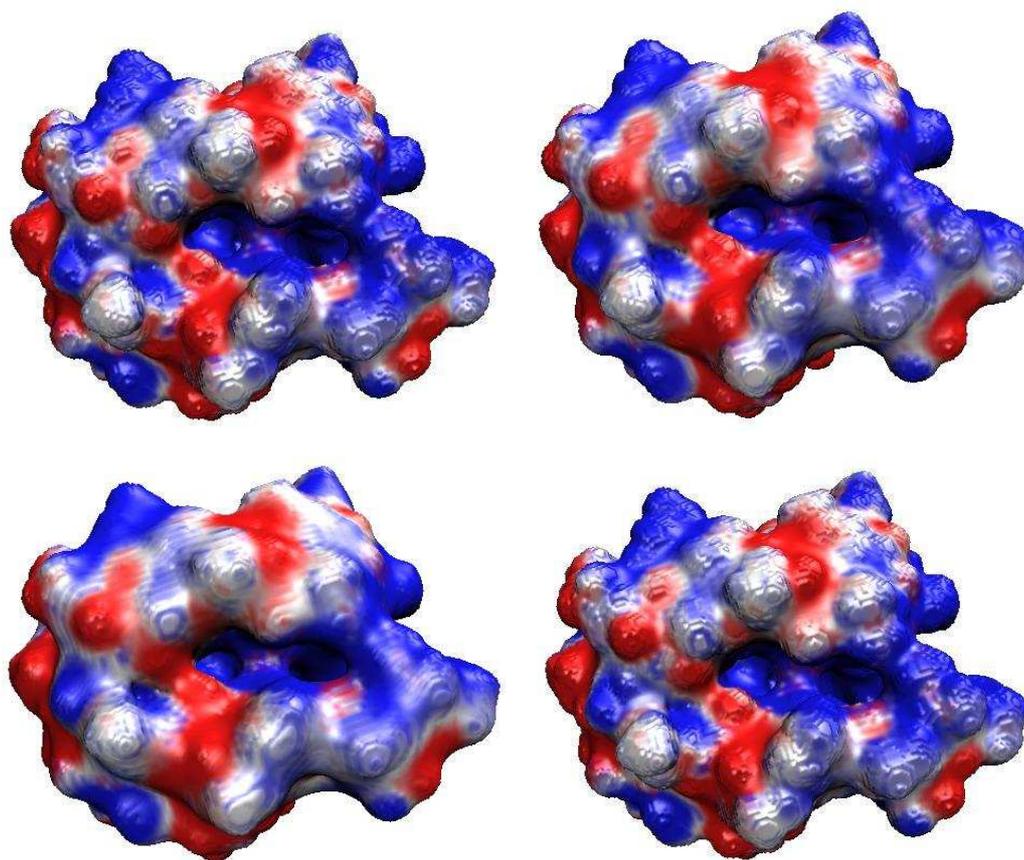


Figure 4.3: The projection of electrostatic surface potentials of protein 451c onto different surfaces obtained under various solvent-solute interactions. Top Left: Attractive surface; Top Right: The MMS; Bottom Left: Repulsive surface; Bottom Right: Polar surface. It is noted that the repulsive surface is a “fat” surface comparing to the MMS; while an attractive surface or a polar surface is a “slim” surface.

in the four atoms changes dramatically when varying non-bonding potentials. This would imply that the size of pocket or cavity inside a protein can be dramatically changed under different electrostatic potentials and/or solute-solute interaction potentials. Therefore, it may result in a significant difference in physical properties of biological systems, which can dramatically influence the selectivity and gating mechanism of ion channels.

Table 4.2: Electrostatic solvation free energies (kcal/mol), surface areas (\AA^2) and volumes (\AA^3) of protein 451c with different solvent-solute interactions.

Surface	Energy	Area	Volume
MMS	-724.3	3695.0	12881.9
Repulsive	-635.2	3805.6	13458.3
Attractive	-897.9	3904.9	11635.6
Polar	-838.1	3702.9	12595.7

4.3.2.3 Surfaces and electrostatic potentials of a protein

Having illustrated the effects of various potentials to surface generations for simple artificial systems, we now consider their impacts to the surface morphology and the solvation analysis of proteins. For this purpose, we choose a protein called heme-binding protein, Fe(II) cytochrome C551 from the organism *Pseudomonas aeruginosa* (PDB ID: 451c). For the structure, extra water molecules that are attached to proteins are excluded and hydrogen atoms are added to obtain a full all-atom model. Partial charges at atomic sites and atomic van der Waals radii in angstroms are taken from the CHARMM22 force field [146]. To show the potential effects, each time we keep one and only one potential term in Eq. (4.47) to produce a new surface which is used in our PB solver to calculate the electrostatic potential. Starting with the MMS, the surface is called a repulsive surface when only a repulsive potential term is added, an attractive surface when only an attractive interaction is added and a polar surface if only the electrostatic potential effect is allowed. The needed parameters for potential expressions are set in the same way as in the 17 compounds which is described in Section 4.4.1.1. Surface electrostatic potentials are plotted on the corresponding surfaces in Figure 4.3. Meanwhile, electrostatic solvation free energies, surface areas and volumes are calculated and listed in Table 4.2. Potential effects similar to the surface generations of

the diatom and four-atom systems are observed. Moreover, it is interesting to note that the minimal molecular surface (MMS) has the smallest surface area among them so that it indeed minimizes the surface free energy. This is consistent with the mathematical proof that the mean curvature flow equation leads to the minimal surface area. Yet, the MMS does not possess the minimal volume. Instead, the attractive solvent-solute interaction leads to the minimal volume. The repulsive solvent-solute interaction gives rise to the largest volume. These results might appear to be counterintuitive. However, one has to keep in mind that proteins are partially charged molecules. The electrostatic free energy plays a dominant role in the solvent-solute interactions. There is an obvious correlation between the solute volume and the electrostatic free energy: The larger the solute volume is, the smaller the electrostatic free energy is in absolute value. Therefore, the repulsive potential interaction leads to the smallest electrostatic solvation free energy in absolute value, which is an indication of the hydrophobic nature of the repulsive potential interaction. As expected, the attractive solvent-solute interaction leads to the largest electrostatic solvation free energy, reflecting the hydrophilic nature of the attractive solvent-solute interaction. It is believed that the results in Table 4.2 are very helpful to the understanding of the sophistication of solvation.

4.3.3 Isosurface function value and minimal surfaces

The minimal molecular surface (MMS) proposed in our earlier work [20] was based entirely on the differential geometry theory of surfaces. Although the minimal surface theory is mathematically rigorous, the resulting surface might not be exactly the one with the minimal surface area, when the evolution of the Laplace-Beltrami operator is carried out in the Eulerian representation. This problem is due to the surface extraction process. There are

Table 4.3: Surface areas (\AA^2) for different surface definitions

PDB-ID	No.of atoms	Area	
		OMS	MMS
1ajj	519	2201.4	2046.7
1bbl	576	2657.6	2434.1
1bor	832	2946.9	2683.9
1bpi	898	3274.9	3017.4
1cbn	648	2401.4	2212.7
1fca	729	2728.7	2474.1
1frd	1478	4467.2	3994.2
1fxd	824	3037.3	2762.5
1hpt	858	3368.3	3013.8
1mbg	903	3163.2	2831.3
1neq	1187	4829.0	4295.5
1ptq	795	2959.4	2685.8
1r69	997	3124.8	2806.3
1sh1	702	2808.4	2515.4
1svr	1435	4796.4	4247.9
1uxc	809	2916.1	2630.6
1vii	596	2571.2	2269.3
2erl	573	2380.4	2162.9
2pde	667	2782.0	2527.9
451c	1216	4184.7	3688.5
1a2s	1272	4507.5	3968.7
1a63	2065	7184.8	6369.7
1a7m	2809	7939.4	6918.9

infinitely many ways to select an isosurface value. Our tests indicate that the MMS can be obtained if we choose $L = \frac{S_0}{2}$. Therefore, we set the isosurface function value to 500 in the present solvation free energy calculations rather than a value very close to 1000 which was used in our earlier paper [20]. The results in Table 4.2 were obtained in this manner. In this subsection, we further illustrate that the MMS indeed gives rise to the smallest surface area for a set of 23 proteins. Moreover, we also study the impact of the pressure to the surface area for a couple of given proteins.

We consider a set of 23 proteins in the present study. The detail of preparation and treatment of protein data is described in Section 4.4.1.2. Two types of surfaces are generated in the present work. The first type is the MMS constructed by the mean curvature flow [20]. The second type of surfaces is called optimized molecular surface (OMS) generated by using the present differential geometry based solvation model. Results are listed in Table 4.3. As expected, the surface areas from the MMS model are always smaller than those from the OMS model. Essentially, the optimization of total free energy differs much from the minimization of the surface free energy.

The possession of the minimal surface area in the MMS can be further demonstrated as follows: we consider situations where only the constant pressure (p) is added into the mean curvature flow equation to cause a perturbation of the MMS. For our purpose, two arbitrarily chosen protein systems (PDB-IDs: 1ajj and 451c) from the set of 23 proteins are explored. In our simulation, we set $p = -0.4, -0.3, -0.2, -0.1, 0, 0.2, 0.3, 0.4$ and 0.5 . The minimal molecular surface is obtained when $p = 0$. Figure 4.4 illustrates the difference of surface areas (\AA^2) between various resulting surfaces generated under different p values and the MMS for these two protein systems. It is clearly seen that a small perturbation around MMS leads to a larger surface area comparing to that of the MMS. In other words, the MMS indeed has the minimal surface area.

4.3.4 Convergence of surface area, volume and energy

In this section, we also illustrate numerically the convergence and decreasing pattern of total solvation free energy during the time integration, which has been shown theoretically in the process of the model derivation. To this end, a small compound named diethyl propanedioate

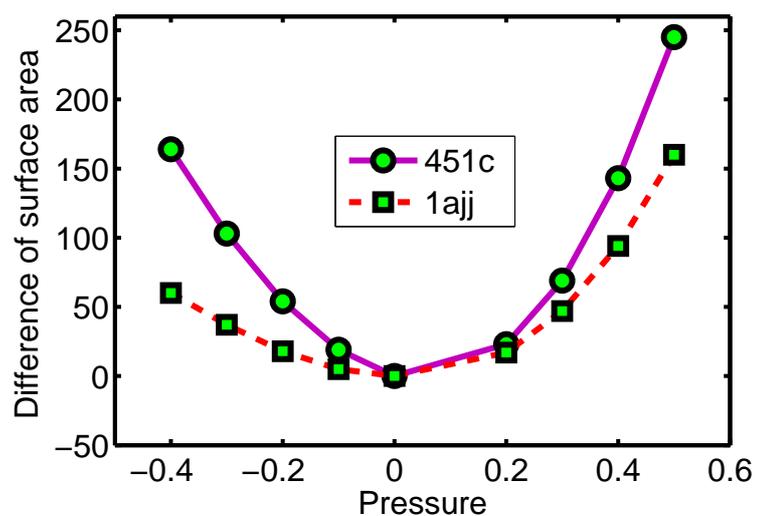


Figure 4.4: Difference of surface areas (\AA^2) between MMS and various resulting surfaces generated under different constant pressure effects.

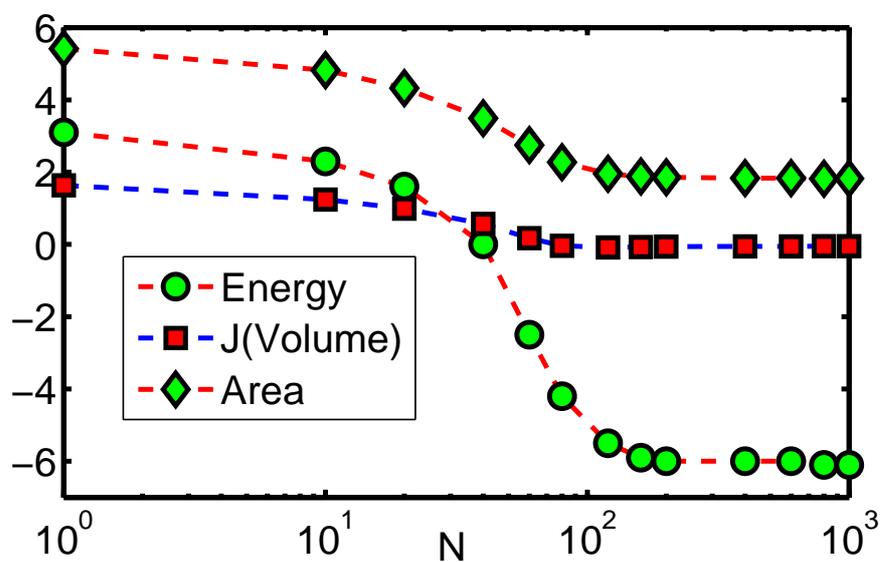


Figure 4.5: Decreasing of surface area ($\times 10^2 \text{\AA}^2$), $J(\text{volume}) = (\text{volume} - 2) (\times 10^2 \text{\AA}^3)$ and total solvation free energy (kcal/mol) in diethyl propanedioate as the number of iterations increases.

has been chosen from a set of 17 compounds described in Section 4.4.1.1. The time history of the total solvation free energies along with the evolution of solvent-solute boundary is recorded. To illustrate the convergence pattern of the solvation free energy, we compute the total solvation free energies in the intermediate states during the time evolution. The results are shown in Figure 4.5. Here T denotes the time span and $N = \frac{T}{\tau}$ represents the number of computational steps in the generalized geometric flow solver. In order to put surface area, enclosed volume and total solvation energy together in one evolution picture, we illustrate $J(\text{volume})$, which is a linear function of volume and shares the same pattern with volume, rather than volume. It is found that the total solvation free energy decreases with respect to the time evolution, which is consistent with our theoretical formulation. It is observed that the solution of our model converges to a steady state in terms of volume (\AA^3), area (\AA^2) and total solvation free energy (kcal/mol).

4.4 Application

In this section, we consider the application of the proposed differential geometry model to the calculations of solvation free energies and salt effects on the protein-protein binding affinity. Previously, we have developed an optimized smooth surface (OSS) model in Chapter 2[46] via the Eulerian formulation of the differential geometry based solvation model. It has been demonstrated that OSS model successfully reproduces not only the solvation free energy of small molecules but also the electrostatic solvation free energies of proteins. Although the present optimized molecular surface (OMS) model is derived by using the same framework of free energy functional optimization, the solvent-solute interfaces are entirely different in two models. It is important to verify whether their results are consistent with each other.

For a comparison, we choose the same set of 17 compounds used in Chapter 2. Thus the results from the OSS model are taken directly from the earlier work. In addition, we also choose a subset of 23 proteins from 30 proteins studied in Chapter 2. The protein-protein binding affinity is investigated by using two protein systems.

4.4.1 Free energy calculations

4.4.1.1 Solvation energies of 17 compounds

This test set of 17 small compounds was studied by Nicholls et al. [160] using a number of approaches, including quantum mechanical methods, the PB theory etc. It is considered as a challenging test set for computational methods because of the existence of polyfunctional or interacting polar groups, which leads to strong solvent-solute interactions. The nonpolar solvent-solute interaction potential in the present model provides a potentially efficient means to deal with strong solvent-solute interactions.

In our calculations, we set the initial amplitude $S_0 = 1000$ and isosurface function value $L = \frac{S_0}{2}$. Other parameters are set in the same way as that in our previous work. Again, here γ (kcal/(molÅ²)) serves as a fitting parameter and will have different values for different expressions of the nonpolar potential. Details are listed in Table 4.4. Typically, only attractive solvent-solute interactions contribute to the dispersion effects in the third term of Eq. (1.2). Here, we have three choices for the dispersion effect: $V^{\text{att,WCA}}$, $V^{\text{att},6/12}$ and V^{LJ} . It turns out that the use of full L-J potential expression can offer the smallest root mean square (RMS) error for the set of 17 compounds. Therefore, it will be chosen from now on for the further study in this chapter except specified.

Structure and charge information of compounds are adopted from those in Nicholls's

Table 4.4: RMS error with different nonpolar potentials.

Potential	$V^{\text{att,WCA}}$	$V^{\text{att,6/12}}$	V^{LJ}
γ	0.0077	0.0094	0.0074
RMS (kcal/mol)	1.77	1.83	1.75

paper [160] too, which can be obtained from the supplementary information of the paper. The results of the present full L-J potential model are summarized in Table 4.5. It gives a comparison between the predicted and experimental values of solvation free energies of 17 compounds. The RMS error of the present model is 1.75 kcal/mol. This RMS error is competitive with the explicit solvent approach (1.71 ± 0.05 kcal/mol) under the same charge and structure parameters set [160]. Moreover, it is interesting to note that the RMS is almost the same as the one obtained from our earlier optimized smooth surface (OSS) model using similar γ value (i.e., 0.0065 kcal/(mol \AA^2) vs 0.0074 kcal/(mol \AA^2)). This consistency can also be seen through Figure 4.6 which shows that the results from the OSS and the present OMS are linearly correlated. The correlation coefficient is 0.999. It may reveal at least two facts. First, in the framework of free energy optimization, the calculated results using the Lagrangian representation and the Eulerian representation should be similar to each other. Additionally, a satisfactory nonpolar term and the enforcement of the potential driven geometric flow really play a critical role in the analysis of solvation free energies.

4.4.1.2 A set of 23 proteins

The set of 17 compounds has already shown the present approach’s ability to predict the total solvation free energy of small compounds. Tests on proteins are needed to demonstrate the capacity on the large system of interest. Encouraged by the success in the application to

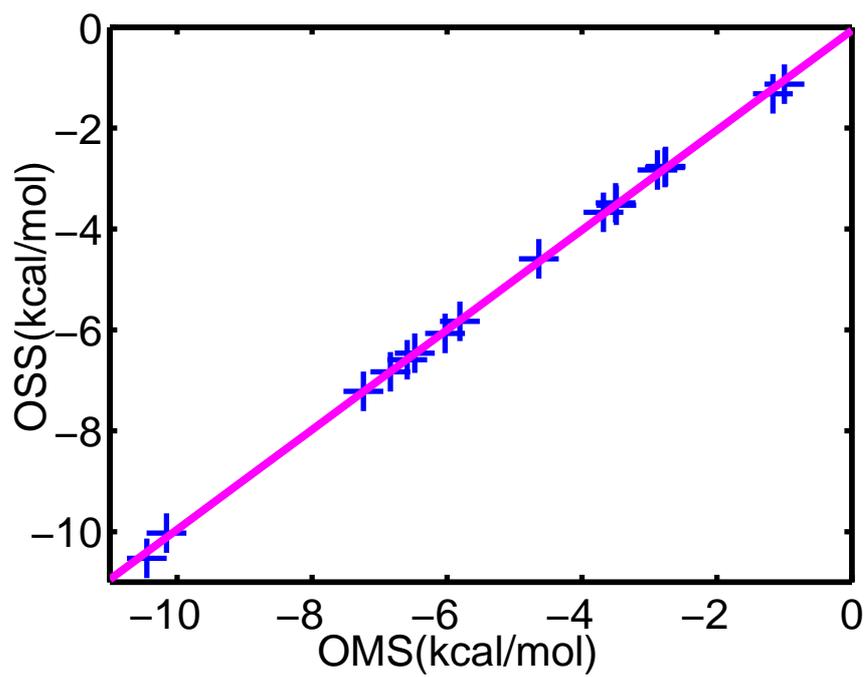


Figure 4.6: Correlation of solvation free energy between previous optimized smooth surface (OSS) model and the present optimized molecular surface (OMS) model in the set of 17 compounds listed in Table 4.5.

Table 4.5: Predicted and experimental total solvation free energies for 17 small compounds.

Compound	G_{np}	ΔG_p	ΔG	Exptl	Error
	(kcal/mol)				
glycerol triacetate	2.33	-12.36	-10.03	-8.84	-1.19
benzyl bromide	1.39	-4.87	-3.47	-2.38	-1.09
benzyl chloride	1.36	-5.06	-3.70	-1.93	-1.77
m-bis(trifluoromethyl)benzene	2.22	-3.30	-1.07	1.07	-2.14
N,N-dimethyl-p-methoxybenzamide	1.99	-9.22	-7.22	-11.01	3.79
N,N-4-trimethylbenzamide	1.91	-7.84	-5.93	-9.76	3.83
bis-2-chloroethyl ether	1.44	-4.16	-2.71	-4.23	1.52
1,1-diacetoxyethane	1.67	-8.21	-6.53	-4.97	-1.56
1,1-diethoxyethane	1.55	-4.63	-3.08	-3.28	0.20
1,4-dioxane	1.01	-5.64	-4.62	-5.05	0.43
diethyl propanedioate	1.87	-7.75	-5.88	-6.00	0.12
dimethoxymethane	1.02	-4.64	-3.62	-2.93	-0.69
ethylene glycol diacetate	1.62	-8.40	-6.78	-6.34	0.44
1,2-diethoxyethane	1.57	-4.40	-2.83	-3.54	0.71
diethyl sulfide	1.22	-2.40	-1.17	-1.43	0.26
phenyl formate	1.37	-7.82	-6.45	-4.08	-2.37
imidazole	0.80	-11.56	-10.76	-9.81	-0.95

small compounds, we further consider a set of realistic proteins and compare the results with those from previous optimized smooth surface (OSS) model and MIBPB-III [93] with pre-determined molecular surfaces (MSs), which is defined as the inner surface smoothly traced by a probe sphere as it rolls over the atomic sphere [185, 57]. Twenty three proteins, a test set used in previous studies [17, 46], are chosen for the present calculations. All structures and partial charges are obtained in the same way as the 451c system which is described before. Table 4.6 shows the results of the present model, and those of the OSS and the MIBPB-III. Results from the OSS and the MIBPB-III have been proved to be very close to each other and they are competitive to those from quantum mechanic approaches [46]. Like in the set of 17 compounds, results from the OSS and the OMS also show quite consistency. The correlation coefficient between them are 0.999. This can also be observed in the Figure

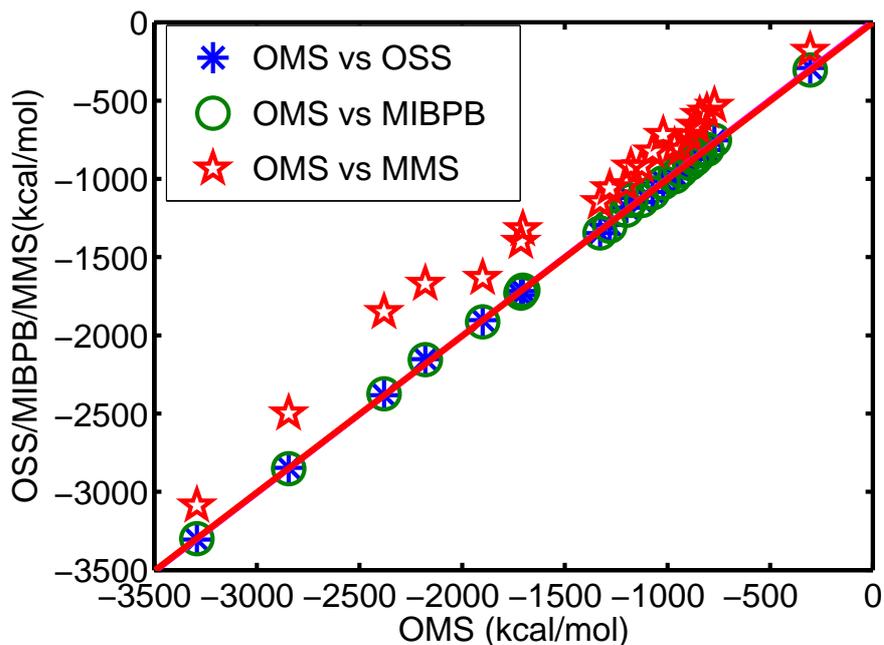


Figure 4.7: Correlation of electrostatic solvation free energy between the present optimized molecular surface (OMS) model, and previous models, such as the optimized smooth surface (OSS), the MIBPB-III and the MMS for 23 proteins listed in Table 4.6.

4.7. Therefore, this observation further convinces us that in the framework of differential geometry based free energy optimization, the OSS and the OMS can be alternative to each other in the aspect of solvation analysis. Since both of them share similar energy functional expressions and take into account the key feature of total energy optimization at equilibrium. If we remove the external potential effects in the surface evolution which are derived from the energy optimization, the present OMS model returns to our previous minimal molecular surface (MMS) model and the calculated results of solvation energies deviate dramatically. Table Figure 4.8 demonstrates the difference of electrostatic solvation free energy between the OSS and MMS models, as well as the difference between the OSS and OMS models. This once again indicates the importance of polar-nonpolar coupling and solute-solvent interaction in implicit solvent modeling and solvation analysis.

Table 4.6: Comparison of electrostatic solvation free energies of 23 proteins.

PDB-ID	No. of atoms	ΔG_p (kcal/mol)			
		MIBPB-III	OSS	OMS	MMS
1ajj	519	-1137.2	-1178.5	-1122.3	-921.0
1bbl	576	-986.8	-965.9	-972.0	-792.3
1bor	832	-853.7	-853.7	-836.3	-665.9
1bpi	898	-1301.9	-1281.2	-1295.1	-1060.0
1cbn	648	-303.8	-304.8	-291.0	-181.0
1fca	729	-1200.1	-1200.6	-1184.1	-1040.0
1frd	1478	-2852.2	-2844.8	-2846.7	-2499.5
1fxd	824	-3299.8	-3291.9	-3306.1	-3087.1
1hpt	858	-811.6	-808.2	-815.6	-570.0
1mbg	903	-1346.1	-1328.2	-1346.9	-1148.7
1neq	1187	-1730.1	-1713.9	-1742.9	-1401.6
1ptq	795	-873.1	-866.2	-872.9	-660.2
1r69	997	-1089.5	-1072.7	-1082.7	-824.4
1sh1	702	-753.3	-771.8	-753.9	-532.1
1svr	1435	-1711.2	-1704.6	-1716.7	-1321.3
1uxc	809	-1138.7	-1125.7	-1147.9	-919.3
1vii	596	-901.5	-892.0	-907.0	-724.2
2erl	573	-948.8	-935.8	-944.4	-812.2
2pde	667	-820.9	-843.0	-812.3	-591.3
451c	1216	-1024.6	-1020.6	-1016.8	-718.2
1a2s	1272	-1913.5	-1900.3	-1902.8	-1633.0
1a63	2065	-2373.5	-2380.5	-2382.6	-1851.0
1a7m	2809	-2155.5	-2179.8	-2152.6	-1699.9

4.4.2 Salt effect on protein-protein binding energies

Finally, we consider the application of our differential geometry based solvation model to the calculations of salt effect on the protein-protein binding. This is the first time that our new model is applied to the study of the salt effect. The ion concentration plays an important role in the stability and even reactivity of biomolecules. This application can be further extended to the binding affinity analysis of ligands, peptide, proteins, nucleic acids, and membrane proteins. To this end, the potential terms caused by mobile ions need to be

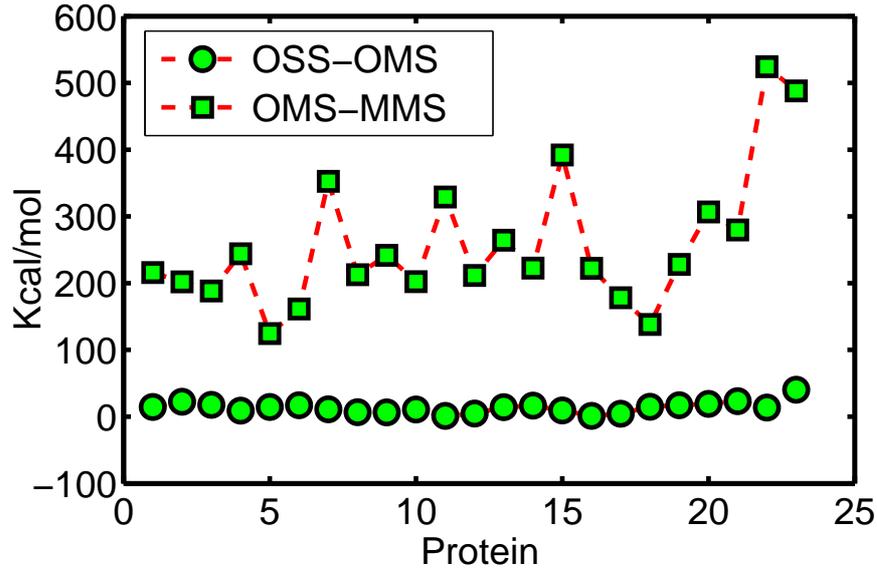


Figure 4.8: Difference of electrostatic solvation free energies between the OMS model and previous OSS and MMS models for 23 proteins listed in Table 4.6.

restored in our calculation.

The full Poisson-Boltzmann (4.31) is coupled to the geometric flow Eq. (4.38) to obtain the solvation free energy for proteins in the salted solvent. The solution procedure for the nonlinear PB equation was described in our work [40]. The coupling of the nonlinear PB equation and potential driven geometric flow equation is discussed in Section 4.2.2. For low salt concentration and weak electrostatic potential, the linearized Poisson-Boltzmann equation discussed in Section 4.1.3 can be applied.

For the binding free energy, only the electrostatic component and particularly, its salt dependence are studied. The total binding free energy which includes many other terms that do not depend on the salt concentration, does not need to be calculated. Then the electrostatic component of the binding energy (ΔG_p) is found as the difference of the electrostatic

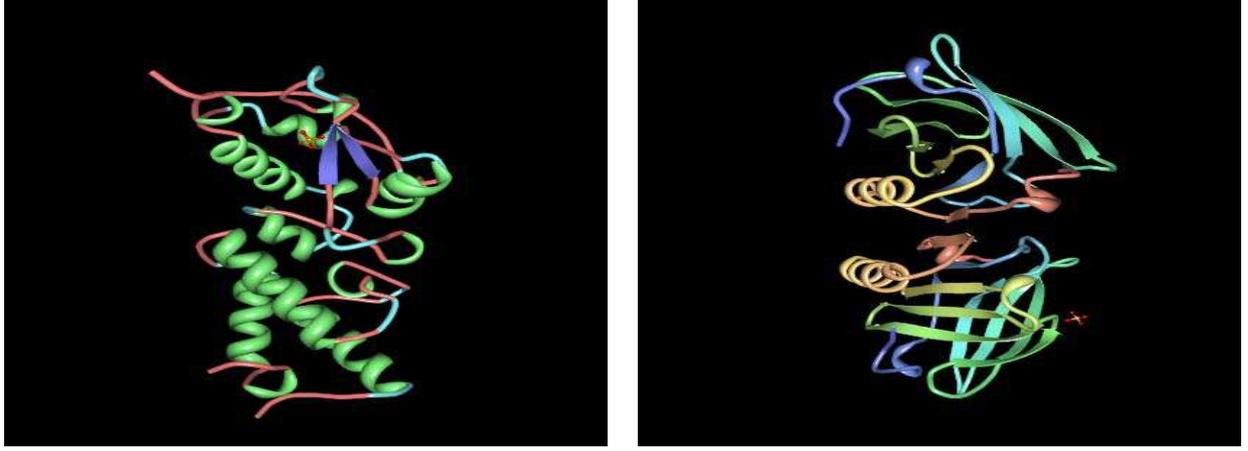


Figure 4.9: Protein-protein complexes. Left: Protein complex 1emv; Right: Protein complex 1beb.

free energies of the complex and those of the free molecules

$$\Delta G_p(I) = G_p^{AB}(I) - G_p^A(I) - G_p^B(I), \quad (4.48)$$

where $G_p^{AB}(I)$, $G_p^A(I)$ and $G_p^B(I)$ are the electrostatic free energies of the complex AB, and the monomers A and B, respectively, at a given ionic strength, I . The salt dependence of the binding free energy $\Delta\Delta G_p(I)$ is thus the difference in the electrostatic components of the binding energies, Eq. (4.48), attained at some salt concentration I and at zero salt concentration

$$\begin{aligned} \Delta\Delta G_p(I) &= \Delta G_p(I) - \Delta G_p(I=0) & (4.49) \\ &= \{G_p^{AB}(I) - G_p^{AB}(I=0)\} \\ &\quad - \{G_p^A(I) - G_p^A(I=0)\} \\ &\quad - \{G_p^B(I) - G_p^B(I=0)\}, \end{aligned}$$

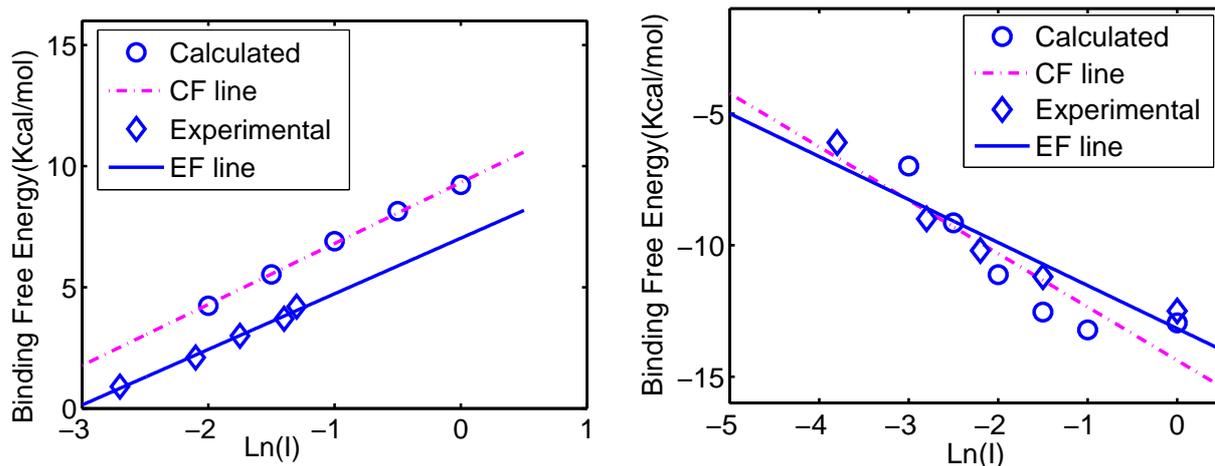


Figure 4.10: The salt dependence of the binding affinities of two protein complexes. Left: Protein lemV; Right: Protein 1beb. Here OMS data are computed by our optimized molecular surface (OMS) model. NLPB data are taken from Bertonati et al's paper [26].

where each energy term at different ionic strengths can be calculated via Eq. (2.23). In general, the nonlinear PB equation should be used for the evaluation of salt effects on the protein-protein binding. However, as shown by Bertonati et al. [26], both the linearized PB (LPB) and the nonlinear PB (NLPB) can be applied to calculate salt effects when the ionic strength of the salt is weak and the net charges of the binding complex and individual molecules are relatively small. The results obtained with the LPB were very close to those obtained with the NLPB. This encourages us to use LPB in this section to reduce the computational complexity.

To test the utility of our new model in the calculation of salt effects on the protein-protein binding, a hetero-dimeric and a homo-dimeric complex are selected for our study. These cases were considered by Bertonati et al [26]. In the experiment, NaCl is used for the salt with a pH value of 3. As shown in Figure 4.9, each protein complex encompasses two separated pieces. The structures and charges of them are attained in the same way as earlier 23 proteins, so are the needed parameters in the potential driven geometric flow equation. The salt

Table 4.7: Comparison of binding affinities of two proteins complexes from current simulations and those from Bertonati et al’s paper.

PDB code	Complex charge	Surface Area (\AA^2)	Charge of the free monomers	Bertonati [26]		OMS
				LPB	NLPB	
1emv	-3	1465	B=+5; A=-8	1.29	1.31	2.40
1beb	+26	1167	A=B=+13	-2.48	-1.53	-2.02

dependence of the binding free energy from NLPB simulation by Bertonati et al. as well as our OMS model is shown in Figure 4.10, where the binding free energy $\Delta\Delta G_p(I)$ is plotted as a function of the logarithm of the salt strength I . Additionally, binding affinities are summarized in Table 4.7, in which the first four columns describe the properties of proteins and the last two columns are the binding affinities extracted from the slope of the lines in Figure 4.10. Note that the calculation is performed by assuming that all Arg, Asp, Glu and Lys residues are ionized in both free and bound states. It is seen that our model clearly reproduces the experimental observation, i.e., for the hetero-dimeric complex, the binding free energies increase with the increasing ionic strength; while for the homo-dimeric complex, the affinity is negative. Moreover, as shown in the table the quantities of the binding affinity obtained from simulations with the present OMS model are in good agreement with those obtained by LPB and NLPB methods in Bertonati et al’s paper in which NaCl is used for salt with a pH value of 3. Note that in the case of Lactoglobulin dimer, the results obtained with all acidic groups neutral are shown.

Note that proteins considered in the present work have fixed conformations before and after their interaction. Ions are treated as structureless. Therefore, the salt effect studied in the present work does not include the effect of ions on biomolecule conformation and complex

formation due to different ion species [80, 225]. It takes more sophisticated models to study the impact of the different ion species to molecular reaction rates and conformational change. Such an aspect is beyond the scope of the present work.

4.5 Chapter conclusions

The objective of the present work is to explore an alternative formulation, the Lagrangian formulation of differential geometry based multiscale solvation models. The Lagrangian representation of biomolecular surfaces is suitable for the visualization, surface electrostatic potential map and visual perception of biomolecules. It can be directly employed in the implicit solvent models and existing software packages. Finally, the Lagrangian representation has an advantage that it avoids artificially enlarging van der Waals radii as often required by smooth surface models [235].

In the present approach, the discrete and continuum domains are separated by a sharp solvent-solute interface, which naturally constitutes a smooth and differentiable manifold enclosing the biomolecule of interest. The time evolution of the manifold is governed by the potential driven geometric flow, a mathematical framework introduced in our previous work [242, 243, 17]. The specific potential driven geometric flow equation used in the present work is derived via the first variation of the total free energy functional of solvation in the Lagrangian representation. Such a derivation differs much from our earlier derivation using the Eulerian representation and geometric measure theory [243].

Although there are some similarities in expressions of coupled PB equation and geometric flow equation between our previous optimized surface (OSS) model in Chapter 2 and the present optimized molecular surface (OMS) model, there are important differences to be

spelled out. First of all, the solute-solvent interface definitions in two models are fundamentally different. In the OSS model, the solute and solvent region is described by a continuous characteristic function denoted by $0 \leq S \leq 1$. In contrast, in the present OMS model, solute and solvent regions are strictly separated by a 2D differentiable manifold. The function S in Eq. (4.38) only serves as a hypersurface function for the geometric surface evolution. This difference has a dramatic computational implication. The generalized Poisson-Boltzmann equation with an OSS is much easier to solve than the OMS is. However, a formal comparison of this computational aspect is beyond the scope of the present work. Moreover, in the potential driven geometric flow equation (4.38), $\nabla\phi^+ \neq \nabla\phi^-$ because of the discontinuity of $\nabla\phi$ inside and outside the solute-solvent interface. However, $\nabla\phi^+ = \nabla\phi^-$ in the overlap region of the OSS model due to continuous dielectric definition. Further, dielectric constant $\epsilon(\mathbf{x})$ in the PB equation is defined in a totally different way: $\epsilon(S)$ is a function of S in the OSS model, and there exists a smooth transition region from the low dielectric to the high dielectric. In contrast, $\epsilon(\mathbf{x})$ is piecewise constant in the present model. In other words, here $\epsilon = \epsilon_s$ in solvent and $\epsilon = \epsilon_m$ in solute, respectively. Further, a generalized Poisson-Boltzmann equation was derived in the OSS model. Whereas, we formally end up with the classical Poisson-Boltzmann equation in the present theory, although it is coupled to the potential driven geometric flow equation. Yet, the present OMS brings up a number of mathematical issues, including the singularity formation on the manifold, and Eulerian embedding of Lagrangian dynamics. Finally, there are many computational problems associated with the Lagrangian formulation of our differential geometry based solvation model too. For instant, the current discontinuous definition of ϵ leads to dramatic accuracy reduction in the standard numerical schemes for the elliptic equations with discontinuous coefficients

and singular sources [254, 253, 258, 261, 260]. To overcome this difficulty, we have incorporated the highly accurate MIB scheme into our PB solver [259, 252, 93]. In addition, many other computational issues, such as hybrid Lagrangian and Eulerian dynamics, level set methods, isosurface extraction, surface integration, and Dirichlet to Neumann mapping [93], are relevant in the present Lagrangian representation.

Chapter 5

Thesis achievements and future work

5.1 Contributions

As described in Chapter 1, solvation is an elementary process in nature that has a great impact on many sophisticated physical, chemical, and biological processes. Therefore, the importance of the understanding of solvation cannot be overemphasized. The major contribution of this thesis is in the construction of a series of novel differential geometry based multiscale solvation models for chemical and biomolecular systems. We have extended our earlier variational formulation of the surface free energy to the analysis of total solvation free energy via the differential geometric theory of surfaces. As a key ingredient of the present framework, the total energy functional encompasses coupled polar and non-polar contributions with a self-consistent interface definition. The true physical boundary of a biomolecule in a solvent, as a physical concept, is determined by the optimization of the total free energy for the solvation equilibrium. As such, a natural description of the solvent-solute interface is provided by the differential geometric theory of surfaces and implemented by the generalized

Laplace -Beltrami equation, rather than simply ad hoc divisions of the solute and solvent regions. Moreover, rigorous mathematical derivations have been demonstrated to obtain the coupled PDE system in the spirit of the variational principle. Additionally, efficient and robust computational algorithms have been designed for the 3D simulation. Finally, solvation analysis of both small compounds and proteins are carried out to further display the accuracy, stability, efficiency and robustness of the proposed new models and the associated numerical approaches. Comparison is made with both experimental and theoretical results in the literature.

Biologically, this thesis provides a self-consistent treatment of the dielectric boundary in all energy terms in the implicit solvent model based on the Poisson-Boltzmann theory.the non-polar free energy terms. It simultaneously optimizes the total solvation free energy. This is considered as a novel and important advance in the continuum treatment of molecular free energies. Moreover, we propose a new theory that allows one to predict the solvation free energy of a biomolecule in a computationally inexpensive way. The relative simplicity of the theory could make this a useful daily tool for researcher to describe general trends in solvation behavior of different biomolecular systems. Furthermore, our method potentially allows one to investigate the ionic effects on the thermodynamics of biomolecules.

In our Eulerian formulation, the surface, separating the low dielectric interior from the high dielectric solvent, is for the first time treated through a 3-D function S . S function takes a value of 1 or 0 in the solute and solvent regions, respectively, and smoothly varies between these values in the interface region. Then the van der Waals interaction, mechanical work and the surface free energy terms are introduced as functions of S . The total energy can then be optimized by simultaneous solution of the generalized Poisson-Boltzmann (GPBE) equa-

tion and the generalized geometric flow equation (GGFE) through the first order variation with respect to the potential and S function. As such, realistic solvent-solute boundaries are constructed. By solving the coupled GPBE and the GGFE, we obtain the electrostatic potential, the solvent-solute boundary profile, and the smooth dielectric function, and thereby improve the accuracy and stability of implicit solvation calculations. We also design efficient second order numerical schemes for the solution of the GPBE and GGFE. The matrix, which results from the discretization of the GPBE, is accelerated with appropriate preconditioners. An alternative direct implicit (ADI) scheme is designed to improve the stability of solving the GGFE. Two iterative approaches are constructed to solve the coupled system of nonlinear partial differential equations. Extensive numerical experiments are designed to validate the present theoretical model, to test computational methods, and to optimize numerical algorithms.

Generally, partial charges from the existing force fields are parameterized for certain class of molecules and may not be accurate for other molecules. Additionally, the fixed charge pattern does not describe the charge rearrangement during the solvation process. This drawback limits the accuracy and utility of our earlier solvation models. The quantum formulation work in this thesis addresses such a limitation by the incorporation of quantum electron density in our earlier models. To this end, we initially construct a new multiscale total free energy functional, which includes the electron kinetic energies and potential energies. By means of the reaction field potential, we can relate the full Kohn-Sham Hamiltonian to the standard Hamiltonian, so that existing computational software packages can be utilized. We have developed a protocol to make use of the SIESTA (Spanish initiative for the electronic structure of thousands of atoms), an efficient linear scaling Density Functional Theory pack-

age, to obtain the solution of the electron density. Appropriate iteration procedures are developed to dynamically couple three governing equations and ensure the convergence of the solution.

In the Lagrangian formulation of our differential geometry based solvation model, the solvent-solute interface is modeled as a 2D manifold embedded into a 3D space. It is suitable for visualization, surface electrostatic potential mapping and visual perception of biomolecules. The Lagrangian formulation is can be directly employed in the implicit solvent model based existing software packages. Moreover, the Lagrangian representation has an advantage in that it avoids artificially enlarging van der Waals radii as is often required by smooth surface models. In this thesis, we also analyze the connections, similarities and differences between the Eulerian and Lagrangian formulations of the solvation models. Such analysis is important to the understanding of our differential geometry based solvation models. Finally, besides the solvation free energies, our Lagrangian formulation model is utilized to evaluate the protein-protein binding affinities.

Most of the materials of this thesis are adopted from the following publications:

- **Zhan Chen** and Guo-wei Wei, “Differential geometry based solvation model III: Quantum formulation.”, *submitted to Journal of chemical physics*
- **Zhan Chen**, Nathan A. Baker and Guo-Wei Wei, “Differential geometry based solvation model II: Lagrangian formulation”, *Journal of Mathematical Biology*, in press, 2011.
- **Zhan Chen**, Nathan A. Baker and Guo-Wei Wei, “Differential geometry based solvation model I: Eulerian formulation”, *Journal of Computational Physics*, 229, vol. 22, pp.8231-8258, 2010

The following publications are also closely related to the present thesis:

- Duan Chen, **Zhan Chen** and G.W. Wei, "Quantum dynamics in continuum for proton channel transport II: Variational interface". *in press, 2011*
- Guo-Wei Wei and **Zhan Chen** , "Multiscale models for nano-bio systems", *Proceedings of CMBE: "2nd international conference on computational and mathematical biomedical engineering*, 19-22,2011
- Duan Chen, **Zhan Chen**, Changjun Chen, Weihua Geng and Guo-Wei Wei, "MIBPB: A software package for electrostatic analysis", *Journal of Computational Chemistry*, 32 vol.4,pp.756-770, 2011 , 2010
- P. W. Bates, **Zhan Chen**, Y.H. Sun, G.W. Wei and Shan Zhao, "Geometric and potential driving formation and evolution of biomolecular surfaces", *Journal of Mathematical Biology*, 59, vol.2, pp.193-231, 2009

5.2 Future work

In our current research, implicit solvent models assume local and linear solvent responses. Further, we concentrate on the solvation calculations in the differential geometry based models. As far as our future work is concerned, models, numerical simulations and mathematical proofs can be improved and extended as follows:

1. A simple homogeneous solvent density is employed in the present work. In future work we will consider the implementation of the solvent variation. In particular, such a variation can be computed by the integral equation approaches of solutions. A combination of the integral equation theories with the present differential geometry based models will lead to

better predictions. This is because the non-uniform density fluctuations of solvent around the solute-solvent interface can be taken into account.

2. Another important extension of the present work is the implementation of the implicit solvation model based molecular dynamics (MD). Currently, the Poisson-Boltzmann (PB) based molecular dynamics algorithms have not been commonly used in the practical simulation of macromolecules. Major hurdles to this development include limits in accuracy, stability, speed and reliability. The multiscale models proposed in this thesis have given rise to a new promise for the development of the PB based molecular dynamics. The accuracy and stability problems in the previous PB based MD methods will not appear in our new model. This is because they are free of interface singularity. Moreover, in all differential geometry based models, the force expressions differ much from those in the classic Poisson-Boltzmann based MD algorithms.

3. The resulting nonlinear PDE systems in our models pose challenges to mathematicians. Numerically, we have shown the existence and local uniqueness of the solutions, which lead to the optimization of the solvation free energy. Moreover, the convergence of the iterative algorithms has been displayed numerically. However, rigorous mathematical proofs have not been studied yet. In future work, upon the promise of numerical performance, we can study these properties using rigorous mathematical tools.

4. The understanding of solvation is an elementary prerequisite for the quantitative description and analysis of a variety of biological, physical and chemical processes. At an equilibrium state, many models are expected to return to the solvation model. Therefore, our differential geometry based solvation models can provide a classic framework for testing the validity of other proposed implicit solvent based biological models, such as ion channel trans-

port models, proton transport models, protein-protein binding simulations, etc. Moreover, it can be incorporated into other related models for accurate simulations in the future.

APPENDICES

Appendix A

Solution of the generalized Laplace-Beltrami equation and the ADI scheme

The solution of the generalized Laplace-Beltrami equations (also called the geometric flow equation) Eq.(2.13) or Eq.(3.17) is studied here. First of all, to solve the geometry flow equation, the expression of solvent-solute interaction potential U_{SS} must be prescribed explicitly. Although U_{SS} includes many unspecified interactions, we consider the following form

$$U_{SS}(\mathbf{r}) = \sum_i U_i(\mathbf{r}), \quad (\text{A.1})$$

where $U_i(\mathbf{r})$ is the potential due to the i th atom in the solute molecule. One possible choice of $U_i(\mathbf{r})$ is the following Lennard-Jones (L-J) 6-12 pair potential

$$U_i^{\text{LJ}}(\mathbf{r}) = \varepsilon_i \left[\left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^{12} - 2 \left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^6 \right] \quad (\text{A.2})$$

where ε_i is the well-depth parameter, and σ_i and σ_s are solute atomic and solvent radii, respectively. Here \mathbf{r} is the point of interest and \mathbf{r}_i is a position vector of an atom in the solute molecule. The L-J potential can be divided into attractive term U^{att} and repulsive term U^{rep} in different ways. It can be a “6-12” decomposition as follows:

$$\begin{aligned} U_i^{\text{att},6/12}(\mathbf{r}) &= -2\varepsilon_i \left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^6 \\ U_i^{\text{rep},6/12}(\mathbf{r}) &= \varepsilon_i \left(\frac{\sigma_i + \sigma_s}{|\mathbf{r} - \mathbf{r}_i|} \right)^{12}. \end{aligned} \quad (\text{A.3})$$

Alternatively, it can also be a “WCA” decomposition based on the original WCA model [130]

$$U_i^{\text{att},\text{WCA}}(\mathbf{r}) = \begin{cases} -\varepsilon_i(\mathbf{r}) & 0 < |\mathbf{r} - \mathbf{r}_i| < \sigma_i + \sigma_s \\ U_i^{\text{LJ}}(\mathbf{r}) & |\mathbf{r} - \mathbf{r}_i| \geq \sigma_i + \sigma_s, \end{cases} \quad (\text{A.4})$$

$$U_i^{\text{rep},\text{WCA}}(\mathbf{r}) = \begin{cases} U_i^{\text{LJ}}(\mathbf{r}) + \varepsilon_i(\mathbf{r}) & 0 < |\mathbf{r} - \mathbf{r}_i| < \sigma_i + \sigma_s \\ 0 & |\mathbf{r} - \mathbf{r}_i| \geq \sigma_i + \sigma_s. \end{cases} \quad (\text{A.5})$$

As indicated in Chapter 2, the WCA attractive potential provides good results for solvation. Therefore, all the calculations in the present study are carried out by using the WCA decomposition. Note that due to the fast decay of the potential, only those solute atoms which

locate near the solvent-solute boundary are needed to include in the evaluation of U_{SS} .

Additionally, a necessary step in solving our generalized Laplace-Beltrami equations is to determine all the physical parameters involved. Because of the choice of the polar and nonpolar separation and the continuum representation of solvent in our model, not all parameters from the literature are suitable. In particular, surface tension γ serves as a fitting parameter in our model due to the ambiguity of its specific value in atomic-scale models [130, 160, 235]. Therefore, we rewrite the generalized potential driven geometric flow equation as

$$\begin{aligned}\frac{\partial S}{\partial t} &= |\nabla S| \left[\nabla \cdot \left(\gamma \frac{\nabla S}{|\nabla S|} \right) + V \right] \\ &= |\nabla S| \gamma \left[\nabla \cdot \left(\frac{\nabla S}{|\nabla S|} \right) + V_\gamma \right]\end{aligned}\tag{A.6}$$

where $V_\gamma = \frac{V}{\gamma}$. Therefore, in addition to the Lennard Jones parameters $\bar{\epsilon}_i$, σ_s and σ_i , other parameters including p/γ , ρ_s/γ , ϵ_s/γ , and ϵ_m/γ need to be pre-determined in the solution of the generalized potential driven geometry flow equation.

The discretization scheme used here for the solution of the generalized geometry flow equation (A.6) is similar to what we designed previously [20, 17]. It can be rewritten in the form

$$\begin{aligned}\frac{\partial S}{\partial t'} &= \frac{(S_x^2 + S_y^2)S_{zz} + (S_x^2 + S_z^2)S_{yy} + (S_y^2 + S_z^2)S_{xx}}{S_x^2 + S_y^2 + S_z^2} \\ &\quad - \frac{2S_x S_y S_{xy} + 2S_x S_z S_{xz} + 2S_z S_y S_{yz}}{S_x^2 + S_y^2 + S_z^2} + \sqrt{S_x^2 + S_y^2 + S_z^2} V_\gamma,\end{aligned}\tag{A.7}$$

where $t' = t\gamma$. To obtain the discretized form, we introduce the following notations. We consider a discrete time $t_n := n\tau$ where n is a non-negative integer and τ is the time stepping

size. We denote S_{ijk}^n to be the discretized form of $S(x_i, y_j, z_k, t_n)$. An explicit scheme of the generalized geometry flow equation is given by

$$S_{ijk}^{n+1} - S_{ijk}^n := [v_x \delta_x^2 + v_y \delta_y^2 + v_z \delta_z^2] S_{ijk}^n + \tau f_{ijk}^n, \quad (\text{A.8})$$

where

$$f_{ijk}^n = \left\{ -2 \frac{S_x S_y S_{xy} + S_x S_z S_{xz} + S_z S_y S_{yz}}{S_x^2 + S_y^2 + S_z^2} + \sqrt{S_x^2 + S_y^2 + S_z^2} V \right\}_{ijk}^n$$

$$v_x = \tau \left\{ \frac{(S_y^2 + S_z^2)}{S_x^2 + S_y^2 + S_z^2} \right\}_{ijk}^n,$$

$$v_y = \tau \left\{ \frac{(S_x^2 + S_z^2)}{S_x^2 + S_y^2 + S_z^2} \right\}_{ijk}^n,$$

$$v_z = \tau \left\{ \frac{(S_x^2 + S_y^2)}{S_x^2 + S_y^2 + S_z^2} \right\}_{ijk}^n,$$

where

$$\delta_x^2 S_{ijk}^n = (S_{(i-1)jk}^n - 2S_{ijk}^n + S_{(i+1)jk}^n)/h^2$$

$$\delta_y^2 S_{ijk}^n = (S_{i(j-1)k}^n - 2S_{ijk}^n + S_{i(j+1)k}^n)/h^2$$

$$\delta_z^2 S_{ijk}^n = (S_{ij(k-1)}^n - 2S_{ijk}^n + S_{ij(k+1)}^n)/h^2$$

$$\{S_x\}_{ijk}^n = (S_{(i+1)jk}^n - S_{(i-1)jk}^n)/2h$$

$$\{S_y\}_{ijk}^n = (S_{i(j+1)k}^n - S_{i(j-1)k}^n)/2h$$

$$\{S_z\}_{ijk}^n = (S_{ij(k+1)}^n - S_{ij(k-1)}^n)/2h$$

$$\{S_{xy}\}_{ijk}^n = (S_{(i+1)(j+1)k}^n + S_{(i-1)(j-1)k}^n - S_{(i+1)(j-1)k}^n - S_{(i-1)(j+1)k}^n)/4h^2$$

$$\{S_{xz}\}_{ijk}^n = (S_{(i+1)j(k+1)}^n + S_{(i-1)j(k-1)}^n - S_{(i+1)j(k-1)}^n - S_{(i-1)j(k+1)}^n)/4h^2$$

and

$$\{S_{yz}\}_{ijk}^n = (S_{i(j+1)(k+1)}^n + S_{i(j-1)(k-1)}^n - S_{i(j+1)(k-1)}^n - S_{i(j-1)(k+1)}^n)/4h^2.$$

Equation (A.6) is solved with the Dirichlet boundary condition $S(\mathbf{r}, t) = 0, \quad \forall \mathbf{r} \in \partial\Omega$.

For the initial value of S , we consider

$$S(x, y, z, 0) = \begin{cases} 1, & (x, y, z) \in D_{sa} \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.9})$$

where we define the domain enclosed by the solvent accessible surface to be $D_{sa} = \bigcup_{i=1}^{N_a} \{\mathbf{r} : |\mathbf{r} - \mathbf{R}_i| < r_i + r_p\}$, with r_i and r_p being atomic van der Waals radius and the probe radius, respectively. Here, \mathbf{R}_i is the atomic center position vector of the i th solute atom and N_a denotes the total number of atoms for a given macromolecule. To protect the van der Waals surface and make the computation more efficient, we only update the values of $S(x, y, z, t)$ at the points in between the van der Waals surface and the solvent accessible surface; i.e., $(x, y, z) \in D_{sa}/D_{vdW}$, where D_{vdW} is the domain enclosed by van der Waals surface $D_{vdW} = \bigcup_{i=1}^{N_a} \{\mathbf{r} : |\mathbf{r} - \mathbf{R}_i| < r_i\}$. Numerically, to avoid possible zeros in the denominator of Eq. (A.6) we add a very small number, such as 10^{-7} , into the denominator expression, which does not affect the result at all.

For simplicity, the widely used explicit Euler scheme can be applied to the solution of the generalized Laplace-Beltrami equation for the time integration. The Euler scheme can

be combined with the second order central difference scheme for the spatial discretization [20]. Nevertheless, this algorithm is not very efficient because a very small time stepping size is required to guarantee the stability of the time integration. Therefore, an alternative direction implicit (ADI) scheme is desirable. The ADI scheme is second order in both spatial and time discretizations. It builds in a fast $O(N)$ Thomas algorithm to solve the tridiagonal linear system and thus is very efficient. The ADI algorithm is unconditionally stable and allows a much larger time stepping size than does the explicit Euler scheme. An splitting algorithm based ADI scheme reported in our earlier work [17] was the fastest scheme among the tested ones under typical accuracy requirement for the mean curvature flow. Considering the similarity of the current differential operator and the mean curvature flow, we adopt the ADI scheme to speed up our generalized geometric flow solver. To this end, we modify Eq. (A.8) as

$$\left(1 - \frac{v_x}{2}\delta_x^2 - \frac{v_y}{2}\delta_y^2 - \frac{v_z}{2}\delta_z^2\right) S_{ijk}^{n+1} = \left(1 + \frac{v_x}{2}\delta_x^2 + \frac{v_y}{2}\delta_y^2 + \frac{v_z}{2}\delta_z^2\right) S_{ijk}^n + \tau f(S_{ijk}^n) \quad (\text{A.10})$$

It follows that

$$\begin{aligned} & \left(1 - \frac{A_x}{2}\right) \left(1 - \frac{A_y}{2}\right) \left(1 - \frac{A_z}{2}\right) S_{ijk}^{n+1} \\ &= \left[\left(1 + \frac{A_x}{2}\right) \left(1 + \frac{A_y}{2}\right) \left(1 + \frac{A_z}{2}\right) - \frac{A_x A_y A_z}{4} \right] S_{ijk}^n + \tau f(S_{ijk}^n) \end{aligned} \quad (\text{A.11})$$

where

$$A_x = v_x \delta_x^2, A_y = v_y \delta_y^2, A_z = v_z \delta_z^2 \quad (\text{A.12})$$

Here $v_x, v_y, v_z, \delta_x^2, \delta_y^2$ and δ_z^2 are defined in Section 2.2.1. The following multi-step imple-

mentation can be carried out.

Step 1:

$$\left(1 - \frac{A_x}{2}\right) S_{ijk}^{n+\frac{1}{3}} = \left(1 + \frac{A_x}{2} + A_y + A_z\right) S_{ijk}^n + \tau f(S_{ijk}^n) \quad (\text{A.13})$$

Step 2:

$$\left(1 - \frac{A_y}{2}\right) S_{ijk}^{n+\frac{2}{3}} = S_{ijk}^{n+\frac{1}{3}} - \frac{A_y}{2} S_{ijk}^n \quad (\text{A.14})$$

Step 3:

$$\left(1 - \frac{A_z}{2}\right) S_{ijk}^{n+1} = S_{ijk}^{n+\frac{2}{3}} - \frac{A_z}{2} S_{ijk}^n. \quad (\text{A.15})$$

Appendix B

PB equation in different forms

Assuming that there are only two mobile ion species and all ions are univalent, we can treat them as positive and negative ions with charge $+e_c$ and $-e_c$, where e_c is the electron charge.

Then the nonlinear Poisson-Boltzmann (NLPB) equation (1.1) becomes [93, 108]

$$-\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi) + \bar{\kappa}^2(\mathbf{r}) \left(\frac{k_B T}{e_c} \right) \sinh \left(\frac{e_c \phi}{k_B T} \right) = \lambda_m \rho_m, \quad (\text{B.1})$$

where $\bar{\kappa}$ is the modified Debye-Hückel screening function describing ion strength and is determined by

$$\bar{\kappa}^2 = \left(\frac{2\lambda_s N_a e_c^2}{1000 k_B T} I_s \right), \quad (\text{B.2})$$

where N_a the Avogadro's number, and I_s the ion strength in the unit of mole. Numerically, when $T = 298\text{K}$, the value of $\bar{\kappa}^2$ can be obtained via $\bar{\kappa}^2 = 0.675365 \text{ \AA}^{-2} I_s$. Note that Debye-Hückel parameter $\bar{\kappa}$ can also be expressed as [97]

$$\bar{\kappa}^2 = \frac{\lambda_s}{k_B T} \sum_{i=1}^{N_c} n_i^0 Q_i^2. \quad (\text{B.3})$$

Equations (B.2) and (B.3) are equivalent to each other via the following formula [108]

$$I_s = \frac{1}{2e_c^2} \sum_{i=1}^{N_c} n_i^0 Q_i^2 = \frac{1000M}{N_a}, \quad (\text{B.4})$$

where M is the bulk concentration of ions in the unit of mole per cubic centimeter $\left(\frac{\text{mol}}{\text{cm}^3}\right)$ for both positive and negative ionic charges. Equation (B.1) is subject to the far-field boundary condition $\phi(\infty) = 0$. However, the Dirichlet boundary condition is used in practical computations

$$\phi(\mathbf{r}) = \sum_i \phi_i = \sum_{i=1}^{Nm} \frac{q_i}{\epsilon_s |\mathbf{r} - \mathbf{r}_i|} e^{-\bar{\kappa} |\mathbf{r} - \mathbf{r}_i| / \epsilon_s} \quad \forall \mathbf{r} \in \partial\Omega, \quad (\text{B.5})$$

where ϕ_i is the exact solution of a single ion in a homogeneous media. The linear superposition in Eq. (B.5) is very accurate if the macromolecule domain Ω_m is sufficiently away from the boundary $\partial\Omega$.

Let define a dimensionless potential u through $u = e_c \phi / k_B T$, one yields another formulation of the nonlinear PB equation in terms of u [40]

$$-\nabla \cdot (\epsilon(\mathbf{r}) \nabla u) + \bar{\kappa}^2(\mathbf{r}) \sinh(u) = \frac{e_c}{k_B T} \lambda_m \rho_m. \quad (\text{B.6})$$

If the potential is very weak, i.e., $u \ll 1$, one can numerically solve the following linearized PB (LPB) equation

$$-\nabla \cdot (\epsilon(\mathbf{r}) \nabla u) + \bar{\kappa}^2(\mathbf{r}) u(\mathbf{r}) = \frac{e_c}{k_B T} \lambda_m \rho_m. \quad (\text{B.7})$$

Note that in the Poisson-Boltzmann theory, there are two unit conventions in the literature that differs by a factor of 4π . Specifically, the convention used by Sharp and Honig [200], and in some of our earlier work [93, 40] has a factor of 4π in the Poisson-Boltzmann

equation. Whereas, the convention used by Gilson et al. [97] and in our recent work [243] as well as the present derivation, the 4π factor does not appear. Therefore, care is needed in the comparison of the electrostatic potentials computed by these two conventions.

Appendix C

Differential geometry theory preliminary

The solvent-solute boundary can be considered as a 2-dimensional (2D) differentiable manifold embedded in a 3D Euclidean space or a hypersurface in a Riemannian manifold. For example, the subsequent free energy optimization can be carried out on the 2D manifold. Consider a C^2 immersion $\mathbf{f} : U \rightarrow \mathbb{R}^{n+1}$, where $U \subset \mathbb{R}^n$ is an open set and \bar{U} is compact[248]. Here $\mathbf{f}(\mathbf{u}) = (f_1(\mathbf{u}), f_2(\mathbf{u}), \dots, f_{n+1}(\mathbf{u}))$ is a hypersurface element (or a position vector), and $\mathbf{u} = (u_1, u_2, \dots, u_n) \in U$. Tangent vectors (or directional vectors) of \mathbf{f} are $X_i = \frac{\partial \mathbf{f}}{\partial u_i}, i = 1, 2, \dots, n$. The Jacobi matrix of the mapping \mathbf{f} is given by $D\mathbf{f} = (X_1, X_2, \dots, X_n)$. The first fundamental form is a symmetric, positive definite metric tensor of \mathbf{f} , given by $I(X_i, X_j) := (g_{ij}) = (D\mathbf{f})^T \cdot (D\mathbf{f})$. Its matrix elements can also be expressed as $g_{ij} = \langle X_i, X_j \rangle$, where \langle, \rangle is the Euclidean inner product in \mathbb{R}^n , $i, j = 1, 2, \dots, n$.

Let $\mathbf{N}(\mathbf{u})$ be the unit normal vector given by the Gauss map $\mathbf{N} : U \rightarrow R^{n+1}$,

$$\mathbf{N}(u_1, u_2, \dots, u_n) := X_1 \times X_2 \cdots \times X_n / \|X_1 \times X_2 \cdots \times X_n\| \in \perp_{\mathbf{u}\mathbf{f}}, \quad (\text{C.1})$$

where “ \times ” denotes the cross product. Here $\perp_{\mathbf{u}\mathbf{f}}$ is the normal space of \mathbf{f} at point $\mathbf{X} = \mathbf{f}(\mathbf{u})$, where the position vector \mathbf{X} differs much from tangent vectors X_i . The normal vector \mathbf{N} is perpendicular to the tangent hyperplane $T_{\mathbf{u}\mathbf{f}}$ at \mathbf{X} . Note that $T_{\mathbf{u}\mathbf{f}} \oplus \perp_{\mathbf{u}\mathbf{f}} = T_{\mathbf{f}(\mathbf{u})}\mathbb{R}^n$, the tangent space at \mathbf{X} . By means of the normal vector \mathbf{N} and tangent vector X_i , the second fundamental form is given by

$$II(X_i, X_j) = (h_{ij})_{i,j=1,2,\dots,n} = \left(\left\langle -\frac{\partial \mathbf{N}}{\partial u_i}, X_j \right\rangle \right)_{ij}. \quad (\text{C.2})$$

The mean curvature can be calculated from $H = \frac{1}{n} h_{ij} g^{ji}$, where we use the Einstein summation convention, and $(g^{ij}) = (g_{ij})^{-1}$.

For $n = 2$, which fits into our purpose, let us choose $\mathbf{f}(\mathbf{u}) = (\mathbf{u}_1, \mathbf{u}_2, \chi)$, where $\chi(u_1, u_2)$ is a function of interest. We have the first fundamental form:

$$(g_{ij}) = \begin{pmatrix} 1 + \chi_1^2 & \chi_1 \chi_2 \\ \chi_1 \chi_2 & 1 + \chi_2^2 \end{pmatrix}, \quad (\text{C.3})$$

where $\chi_i = \frac{\partial \chi}{\partial u_i}$, $i = 1, 2$. The inverse matrix of (g_{ij}) is given by

$$(g^{ij}) = \frac{1}{g} \begin{pmatrix} 1 + \chi_2^2 & -\chi_1 \chi_2 \\ -\chi_1 \chi_2 & 1 + \chi_1^2 \end{pmatrix}, \quad (\text{C.4})$$

where $g = \text{Det}(g_{ij}) = 1 + \chi_1^2 + \chi_2^2$ is the Gram determinant. The normal vector can be computed from Eq. (C.1)

$$\mathbf{N} = \frac{(-\chi_1, -\chi_2, 1)}{\sqrt{g}}, \quad (\text{C.5})$$

The second fundamental form is given by $(h_{ij}) = \left(\frac{1}{\sqrt{g}} \chi_{u_i u_j} \right)$, i.e., the Hessian matrix of χ .

The explicit form for the mean curvature operator can be written as

$$H = \frac{1}{2g} (h_{11}g_{22} + h_{22}g_{11} - 2h_{12}g_{12}) \quad (\text{C.6})$$

$$= \frac{1}{2} \left[\frac{\partial}{\partial u_1} \left(\frac{\chi_1}{\sqrt{g}} \right) + \frac{\partial}{\partial u_2} \left(\frac{\chi_2}{\sqrt{g}} \right) \right]. \quad (\text{C.7})$$

In Section 4.1.3, we show that the mean curvature operator can be expressed in a (3D) formulation.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] A. Abrashkin, D. Andelman, and H. Orland. Dipolar poisson-boltzmann equation: Ions and dipoles close to charge interfaces. *PHYSICAL REVIEW LETTERS*, 99:077801, 2007.
- [2] J. Antosiewicz, J. A. McCammon, and M. K. Gilson. The determinants of pK_a s in proteins. *Biochemistry*, 35(24):7819–7833, 1996.
- [3] E. Artacho. The siesta method; developments and applicability. *J. Phys.: Condens. Matter*, 20:064208, 2008.
- [4] H. S. Ashbaugh. Convergence of molecular and macroscopic continuum descriptions of ion hydration. *Journal of Physical Chemistry B*, 104(31):7235–7238, 2000.
- [5] C. Azuara, E. Lindahl, P. Koehl, H. Orland, and M. Delarue. PDB_Hydro: incorporating dipolar solvents with variable density in the Poisson-Boltzmann treatment of macromolecule electrostatics. *Nucleic Acids Research*, 34:W38–W42, 2006.
- [6] C. Azuara, H. Orland, M. Bon, P. Koehl, and M. Delarue. Incorporating dipolar solvents with variable density in poisson-boltzmann electrostatics. *Biophysical Journal*, 95:5587C5605, 2008.
- [7] N. A. Baker. Poisson-Boltzmann methods for biomolecular electrostatics. *Methods in Enzymology*, 383:94–118, 2004.
- [8] N. A. Baker. Biomolecular applications of Poisson-Boltzmann methods. In K. B. Lipkowitz, R. Larter, and T. R. Cundari, editors, *Reviews in Computational Chemistry*, volume 21. John Wiley and Sons, Hoboken, NJ, 2005.

- [9] N. A. Baker. Improving implicit solvent simulations: a Poisson-centric view. *Current Opinion in Structural Biology*, 15(2):137–43, 2005.
- [10] N. A. Baker, D. Bashford, and D. A. Case. Implicit solvent electrostatics in biomolecular simulation. In B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schutte, and R. Skeel, editors, *New Algorithms for Macromolecular Simulation*. Springer, 2006.
- [11] N. A. Baker and J. A. McCammon. Electrostatic interactions. In P. Bourne and H. Weissig, editors, *Structural Bioinformatics*, pages 427–440. John Wiley & Sons, Inc., New York, 2003.
- [12] N. A. Baker, 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 of the United States of America*, 98(18):10037–10041, 2001.
- [13] N. K. Banavali, W. Im, and B. Roux. Electrostatic free energy calculations using the generalized solvent boundary potential method. *Journal of Chemical Physics*, 117(15):7381–8, 2002.
- [14] V. Barone, M. Cossi, and J. Tomasi. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *Journal of Chemical Physics*, 107:3210–3221, 1997.
- [15] D. Bashford and D. A. Case. Generalized Born models of macromolecular solvation effects. *Annual Review of Physical Chemistry*, 51:129–152, 2000.
- [16] D. Bashford and M. Karplus. pK_a 's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 29(44):10219–25, 1990.
- [17] P. W. Bates, Z. Chen, Y. H. Sun, G. W. Wei, and S. Zhao. Geometric and potential driving formation and evolution of biomolecular surfaces. *J. Math. Biol.*, 59:193–231, 2009.
- [18] P. W. Bates, G. W. Wei, and S. Zhao. The minimal molecular surface. *arXiv:q-bio/0610038v1*, [q-bio.BM], 2006.
- [19] P. W. Bates, G. W. Wei, and S. Zhao. The minimal molecular surface. *Midwest Quantitative Biology Conference*, Mission Point Resort, Mackinac Island, MI:September 29 – October 1, 2006.

- [20] P. W. Bates, G. W. Wei, and S. Zhao. Minimal molecular surfaces and their applications. *Journal of Computational Chemistry*, 29(3):380–91, 2008.
- [21] A. D. Becke. Density functional thermochemistry iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.
- [22] D. Beglov and B. Roux. Solvation of complex molecules in a polar liquid: an integral equation theory. *Journal of Chemical Physics*, 104(21):8678–8689, 1996.
- [23] D. Beglov and B. Roux. An integral equation to describe the solvation of polar molecules in liquid water. *Journal of Physical Chemistry B*, 101(39):7821–6, 1997.
- [24] M. Berger and B. Gostiaux. *Differential Geometry: Manifolds, Curves, and Surfaces*. Springer-Verlag, 1988.
- [25] C. A. S. Bergstrom, M. Strafford, L. Lazorova, A. Avdeef, K. Luthman, and P. Artursson. Absorption classification of oral drugs based on molecular surface properties. *Journal of Medicinal Chemistry*, 46(4):558–570, 2003.
- [26] C. Bertonati, B. Honig, and E. Alexov. Poisson-Boltzmann calculations of nonspecific salt effects on protein-protein binding free energy. *Biophysical Journal*, 92:1891–1899, 2007.
- [27] A. L. Bertozzi and J. B. Greer. Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes. *Communications on Pure and Applied Mathematics*, 57(6):764–790, 2004.
- [28] B. H. Besler, J. Merz, K. M., and P. A. Kollman. Atomic charges derived from semiempirical methods. *J. Comput. Chem.*, 11:431–439, 1990.
- [29] N. Blomberg, R. R. Gabdouliline, M. Nilges, and R. C. Wade. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins*, 37(3):379–387, 1999.
- [30] P. Blomgren and T. Chan. Color TV: total variation methods for restoration of vector-valued images. *Image Processing, IEEE Transactions on*, 7(3):304–309, 1998.
- [31] A. H. Boschitsch and M. O. Fenley. Hybrid boundary element and finite difference method for solving the nonlinear Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 25(7):935–955, 2004.

- [32] M. Bostrom, F. W. Tavares, D. Bratko, and B. W. Ninham. Specific ion effects in solutions of globular proteins: Comparison between analytical models and simulation. *Journal of Physical Chemistry B*, 109(51):24489–94, 2005.
- [33] K. Burke, J. Werschnik, and E. K. U. Gross. Time-dependent density functional theory: Past, present, and future. *J. Chem. Phys.*, 123:062206, 2005.
- [34] W. Cai and S. Z. Deng. An upwinding embedded boundary method for Maxwell’s equations in media with material interfaces: 2d case. *J. Comput. Phys.*, 190:159–183, 2003.
- [35] E. Cancès, B. Mennucci, and J. Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *Journal of Chemical Physics*, 107:3032 – 3041, 1997.
- [36] V. Carstensen, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997.
- [37] T. Cecil. A numerical method for computing minimal surfaces in arbitrary dimension. *Journal of Computational Physics*, 206(2):650–660, 2005.
- [38] D. S. Cerutti, N. A. Baker, and J. A. McCammon. Solvent reaction field potential inside an uncharged globular protein: A bridge between implicit and explicit solvent models? *The Journal of Chemical Physics*, 127(15):155101, 2007.
- [39] Q. Chang, X. Tai, and L. Xing. A compound algorithm of denoising using second-order and fourth-order partial differential equations. *NUMERICAL MATHEMATICS-THEORY METHODS AND APPLICATIONS*, 2:353–376, 2010.
- [40] D. Chen, Z. Chen, C. Chen, W. H. Geng, and G. W. Wei. MIBPB: A software package for electrostatic analysis. *J. Comput. Chem.*, in press, 2010.
- [41] D. Chen, G. W. Wei, X. Cong, and G. Wang. Computational methods for optical molecular imaging. *Communications in Numerical Methods in Engineering*, 25:1137–1161, 2009.
- [42] J. Chen, L. Noodleman, D. Case, and D. Bashford. Incorporating solvation effects into density functional electronic structure calculations. *J. Phys. Chem.*, 98:11059–11068, 1994.

- [43] L. Chen, M. J. Holst, and J. Xu. The finite element approximation of the nonlinear poisson–boltzmann equation. *SIAM Journal on Numerical Analysis*, 45(6):2298–2320, 2007.
- [44] T. Chen and J. Strain. Piecewise-polynomial discretization and Krylov-accelerated multigrid for elliptic interface problems. *J. Comput. Phys.*, 16:7503–7542, 2008.
- [45] Y. G. Chen and J. D. Weeks. Local molecular field theory for effective attractions between like charged objects in systems with strong Coulomb interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20):7560–5, 2006.
- [46] Z. Chen, N. A. Baker, and G. W. Wei. Differential geometry based solvation models I: Eulerian formulation. *J. Comput. Phys.*, 229:8231–8258, 2010.
- [47] Z. Chen, N. A. Baker, and G. W. Wei. Differential geometry based solvation models II: Lagrangian formulation. *J. Math. Biol.*, accepted, 2010.
- [48] L. T. Cheng, J. Dzubiella, A. J. McCammon, and B. Li. Application of the level-set method to the implicit solvation of nonpolar molecules. *Journal of Chemical Physics*, 127(8), 2007.
- [49] Y. Cheng, J. K. Suen, Z. Radi, S. D. Bond, M. J. Holst, and J. A. McCammon. Continuum simulations of acetylcholine diffusion with reaction-determined boundaries in neuromuscular junction models. *Biophysical Chemistry*, 127(3):129–39, 2007.
- [50] Y. Cheng, J. K. Suen, D. Zhang, S. D. Bond, Y. Zhang, Y. Song, N. A. Baker, C. L. Bajaj, M. J. Holst, and J. A. McCammon. Finite element analysis of the time-dependent Smoluchowski equation for acetylcholinesterase reaction rate calculations. *Biophysical Journal*, 92(10):3397–406, 2007.
- [51] I. L. Chern, J.-G. Liu, and W.-C. Weng. Accurate evaluation of electrostatics for macromolecules in solution. *Methods and Applications of Analysis*, 10(2):309–28, 2003.
- [52] M. Chiba, D. G. Fedorov, and K. Kitaura. Polarizable continuum model with the fragment molecular orbital-based time-dependent density functional theory. *Journal of Computational Chemistry*, 29:2667–2676, 2008.
- [53] L. E. Chirlian and M. M. Francl. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.*, 8:894–905, 1987.

- [54] D. L. Chopp. Computing minimal surfaces via level set curvature flow. *Journal of Computational Physics*, 106(1):77–91, 1993.
- [55] I. Chorny, K. A. Dill, and M. P. Jacobson. Surfaces affect ion pairing. *Journal of Physical Chemistry B*, 109(50):24056–60, 2005.
- [56] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach. Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophysical Journal*, 93(9):3202–3209, 2007.
- [57] M. L. Connolly. Analytical molecular surface calculation. *Journal of Applied Crystallography*, 16(5):548–558, 1983.
- [58] R. B. Corey and L. Pauling. Molecular models of amino acids, peptides and proteins. *Rev. Sci. Instr.*, 24:621–627, 1953.
- [59] M. Cossi, V. Barone, R. Cammi, and J. Tomasi. Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chemical Physics Letters*, 255:327–335, 1996.
- [60] S. R. Cox and D. E. Williams. Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.*, 2:304–323, 1981.
- [61] C. J. Cramer and D. G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99:2161–2200, 1999.
- [62] P. B. Crowley and A. Golovin. Cation- π interactions in protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics*, 59(2):231–239, 2005.
- [63] L. David, R. Luo, and M. K. Gilson. Comparison of generalized Born and Poisson models: Energetics and dynamics of HIV protease. *Journal of Computational Chemistry*, 21(4):295–309, 2000.
- [64] M. E. Davis, J. D. Madura, J. Sines, B. A. Luty, S. A. Allison, and J. A. McCammon. Diffusion-controlled enzymatic reactions. *Methods in Enzymology*, 202:473–497, 1991.
- [65] M. E. Davis and J. A. McCammon. Electrostatics in biomolecular structure and dynamics. *Chemical Reviews*, 94:509–21, 1990.

- [66] F. De Rienzo, R. R. Gabdoulline, M. C. Menziani, P. G. De Benedetti, and R. C. Wade. Electrostatic analysis and Brownian dynamics simulation of the association of plastocyanin and cytochrome F. *Biophysical Journal*, 81(6):3090–3104, 2001.
- [67] C. Dietrich, C. E. Scheidegger, J. Schreiner, J. L. D. Comba, L. P. Nedel, and C. T. Silva. Edge transformations for improving mesh quality of marching cubes. *IEEE Transactions on Visualization and Computer Graphics*, 15(1):150–159, 2009.
- [68] B. N. Dominy and C. L. Brooks, III. Development of a generalized Born model parameterization for proteins and nucleic acids. *Journal of Physical Chemistry B*, 103(18):3765–3773, 1999.
- [69] F. Dong, B. Olsen, and N. A. Baker. Computational methods for biomolecular electrostatics. *Methods in Cell Biology*, 84:843–70, 2008.
- [70] F. Dong, M. Vijaykumar, and H. X. Zhou. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophysical Journal*, 85(1):49–60, 2003.
- [71] F. Dong, J. A. Wagoner, and N. A. Baker. Assessing the performance of implicit solvation models at a nucleic acid surface. *Physical Chemistry Chemical Physics*, 10:4889–902, 2008.
- [72] F. Dong and H. X. Zhou. Electrostatic contribution to the binding stability of protein-protein complexes. *Proteins*, 65(1):87–102, 2006.
- [73] A. I. Dragan, C. M. Read, E. N. Makeyeva, E. I. Milgotina, M. E. Churchill, C. Crane-Robinson, and P. L. Privalov. DNA binding and bending by HMG boxes: Energetic determinants of specificity. *Journal of Molecular Biology*, 343(2):371–393, 2004.
- [74] J. Dzubiella, J. M. J. Swanson, and J. A. McCammon. Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. *Physical Review Letters*, 96:087802, 2006.
- [75] J. Dzubiella, J. M. J. Swanson, and J. A. McCammon. Coupling nonpolar and polar solvation free energies in implicit solvent models. *Journal of Chemical Physics*, 124:084905, 2006.
- [76] A. H. Elcock, R. R. Gabdoulline, R. C. Wade, and J. A. McCammon. Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. *Journal of Molecular Biology*, 291(1):149–162, 1999.

- [77] A. H. Elcock, D. Sept, and J. A. McCammon. Computer simulation of protein-protein interactions. *J. Phys.Chem. B*, 105:1504–1518, 2001.
- [78] H. Federer. Curvature Measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.
- [79] R. P. Fedkiw, T. Aslam, B. Merriman, and S. Osher. A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method). *J. Comput. Phys.*, 152:457–492, 1999.
- [80] M. V. Fedorov, J. M. Goodman, and S. Schumm. To switch or not to switch: The effects of potassium and sodium ions on α -poly-l-glutamate conformations in aqueous solutions. *J. Am. Chem. Soc.*, 131:10854–0856, 2009.
- [81] M. V. Fedorov and A. A. Kornyshev. Unravelling the solvent response to neutral and charged solutes. *Molecular Physics*, 105(1):1–16, 2007.
- [82] M. Feig and C. L. Brooks III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol.*, 14:217 – 224, 2004.
- [83] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and I. Brooks, C. L. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry*, 25(2):265–284, 2004.
- [84] X. Feng and A. Prohl. Analysis of a fully discrete finite element method for the phase field model and approximation of its sharp interface limits. *Mathematics of Computation*, 73:541–567, 2004.
- [85] M. Fixman. The Poisson-Boltzmann equation and its application to polyelectrolytes. *Journal of Chemical Physics*, 70(11):4995–5005, 1979.
- [86] F. Fogolari, A. Brigo, and H. Molinari. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, 15(6):377–92, 2002.
- [87] J. Forsman. A simple correlation-corrected Poisson-Boltzmann theory. *Journal of Physical Chemistry B*, 108(26):9236–45, 2004.

- [88] P. H. Fries and G. N. Patey. The solution of the hypernetted-chain approximation for fluids of nonspherical particles. a general method with application to dipolar hard spheres. *Journal of Chemical Physics*, 82:429 – 440, 1985.
- [89] R. R. Gabdouliline and R. C. Wade. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods-a Companion to Methods in Enzymology*, 14(3):329–341, 1998.
- [90] E. Gallicchio, M. M. Kubo, and R. M. Levy. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *Journal of Physical Chemistry B*, 104(26):6271–85, 2000.
- [91] E. Gallicchio and R. M. Levy. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *Journal of Computational Chemistry*, 25(4):479–499, 2004.
- [92] E. Gallicchio, L. Y. Zhang, and R. M. Levy. The SGB/NP hydration free energy model based on the surface generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. *Journal of Computational Chemistry*, 23(5):517–29, 2002.
- [93] W. Geng, S. Yu, and G. W. Wei. Treatment of charge singularities in implicit solvent models. *Journal of Chemical Physics*, 127:114106, 2007.
- [94] W. H. Geng and G. W. Wei. Multiscale molecular dynamics via the matched interface and boundary (mib) method. *J. Comput. Phys.*, in press, 2010.
- [95] R. E. Georgescu, E. G. Alexov, and M. R. Gunner. Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins. *Biophysical Journal*, 83(4):1731–1748, 2002.
- [96] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Image sharpening by flows based on triple well potentials. *Journal of Mathematical Imaging and Vision*, 20:121–131, 2004.
- [97] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *Journal of Physical Chemistry*, 97(14):3591–3600, 1993.
- [98] S. Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71:1085–1123, 1999.

- [99] V. Gogonea and K. M. Merz. Fully quantum mechanical description of proteins in solution. combining linear scaling quantum mechanical methodologies with the poisson-boltzmann equation. *J. Phys. Chem. A*, 103:5171–5188, 1999.
- [100] J. Gomes and O. D. Faugeras. Using the vector distance functions to evolve manifolds of arbitrary codimension. *Lecture Notes in Computer Science*, 2106:1–13, 2001.
- [101] J. A. Grant, B. T. Pickup, and A. Nicholls. A smooth permittivity function for Poisson-Boltzmann solvation methods. *Journal of Computational Chemistry*, 22(6):608–640, 2001.
- [102] J. A. Grant, B. T. Pickup, M. T. Sykes, C. A. Kitchen, and A. Nicholls. The Gaussian Generalized Born model: application to small molecules. *Physical Chemistry Chemical Physics*, 9:4913–22, 2007.
- [103] J. B. Greer and A. L. Bertozzi. H-1 solutions of a class of fourth order nonlinear equations for image processing. *Discrete and Continuous Dynamical Systems*, 10:349–366, 2004.
- [104] J. B. Greer and A. L. Bertozzi. Traveling wave solutions of fourth order pdes for image processing. *SIAM Journal on Mathematics Analysis*, 36:38–68, 2004.
- [105] P. Grochowski and J. Trylska. Continuum molecular electrostatics, salt effects and counterion binding. a review of the Poisson-Boltzmann theory and its modifications. *Biopolymers*, 89(2):93–113, 2007.
- [106] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–871, 1964.
- [107] C. Holm, P. Kekicheff, and R. Podgornik. *Electrostatic effects in soft matter and biophysics; NATO Science Series*. Kluwer Academic Publishers, Boston, 2001.
- [108] M. J. Holst. *Multilevel Methods for the Poisson-Boltzmann Equation*. University of Illinois at Urbana-Champaign, Numerical Computing Group, Urbana-Champaign, 1993.
- [109] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–9, 1995.
- [110] H. Hu, Z. Y. Lu, and W. T. Yang. Fitting molecular electrostatic potentials from quantum mechanical calculations. *Journal of Chemical Theory and Computation*, 3:1004 – 1013, 2007.

- [111] B. Husowitz and V. Talanquer. Solvent density inhomogeneities and solvation free energies in supercritical diatomic fluids: A density functional approach. *The Journal of Chemical Physics*, 126(5):054508, 2007.
- [112] W. Im, D. Beglov, and B. Roux. Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications*, 111(1-3):59–75, 1998.
- [113] R. Improta, V. Barone, G. Scalmani, and M. J. Frisch. A state-specific polarizable continuum model time dependent density functional theory method for excited state calculations in solution. *Journal of Chemical Physics*, 125(054103), 2006.
- [114] R. M. Jackson and M. J. Sternberg. A continuum model for protein-protein interactions: Application to the docking problem. *Journal of Molecular Biology*, 250(2):258–275, 1995.
- [115] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method. *Journal of Computational Chemistry*, 21(2):132–146, 2000.
- [116] B. Jayaram, D. Sprous, and D. L. Beveridge. Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field. *Journal of Physical Chemistry B*, 102(47):9571–9576, 1998.
- [117] R. Jinnouchi and A. B. Anderson. Electronic structure calculations of liquid-solid interfaces: Combination of density functional theory and modified Poisson-Boltzmann theory. *PHYSICAL REVIEW B*, 77(245417), 2008.
- [118] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel. Progress in ab initio qm/mm free-energy simulations of electrostatic energies in proteins: Accelerated qm/mm studies of pk(a), redox reactions and solvation free energies. *Journal of Phys. Chem. B*, 113:1253–1272, 2009.
- [119] J. G. Kirkwood. Theory of solution of molecules containing widely separated charges with special application to zwitterions. *J. Comput. Phys.*, 7:351 – 361, 1934.
- [120] P. Koehl. Electrostatics calculations: latest methodological advances. *Current Opinion in Structural Biology*, 16(2):142–51, 2006.

- [121] P. Koehl, H. Orland, and M. Delarue. Beyond the poisson-boltzmann model: Modeling biomolecule-water and water-water interactions. *PHYSICAL REVIEW LETTERS*, 102:087801, 2009.
- [122] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133– A1138, 1965.
- [123] L. A. Kuhn, M. A. Siani, M. E. Pique, C. L. Fisher, E. D. Getzoff, and J. A. Tainer. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *Journal of Molecular Biology*, 228(1):13–22, 1992.
- [124] M. C. Lai and C. S. Peskin. An immersed boundary method with formal second-order accuracy and reduced numerical viscosity. *J. Comput. Phys.*, 160:705–719, 2000.
- [125] G. Lamm. The Poisson-Boltzmann equation. In K. B. Lipkowitz, R. Larter, and T. R. Cundari, editors, *Reviews in Computational Chemistry*, pages 147–366. John Wiley and Sons, Inc., Hoboken, N.J., 2003.
- [126] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55(3):379–400, 1971.
- [127] C. T. Lee, W. T. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. *Physical Review B*, 37:785 – 789, 1988.
- [128] M. S. Lee, J. Salsbury, F. R., and M. A. Olson. An efficient hybrid explicit/implicit solvent method for biomolecular simulations. *Journal of Computational Chemistry*, 25(16):1967–78, 2004.
- [129] R. J. LeVeque and Z. L. Li. The immersed interface method for elliptic equations with discontinuous coefficients and singular sources. *SIAM J. Numer. Anal.*, 31:1019–1044, 1994.
- [130] R. M. Levy, L. Y. Zhang, E. Gallicchio, and A. K. Felts. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *Journal of the American Chemical Society*, 125(31):9523–9530, 2003.
- [131] H. Li, A. D. Robertson, and J. H. Jensen. The determinants of carboxyl pKa values in turkey ovomucoid third domain. *Proteins*, 55(3):689–704, 2004.

- [132] H. Li, A. D. Robertson, and J. H. Jensen. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4):704–21, 2005.
- [133] J. Li, C. L. Fisher, J. L. Chen, D. Bashford, and L. Noodleman. Calculation of redox potentials and pKa values of hydrated transition metal cations by a combined density functional and continuum dielectric theory. *Inorganic Chemistry*, 35(16):4694–702, 1996.
- [134] Y. Li and F. Santosa. A computational algorithm for minimizing total variation in image restoration. *IEEE Transactions on Image Processing*, 5(6):987–95, 1996.
- [135] Z. L. Li and K. Ito. Maximum principle preserving schemes for interface problems with discontinuous coefficients. *SIAM J. Sci. Comput.*, 23:339–361, 2001.
- [136] V. J. Licata and N. M. Allewell. Functionally linked hydration changes in escherichia coli aspartate transcarbamylase and its catalytic subunit. *Biochemistry*, 36(33):10161–10167, 1997.
- [137] X. D. Liu, R. P. Fedkiw, and M. Kang. A boundary condition capturing method for Poisson’s equation on irregular domains. *J. Comput. Phys.*, 160:151–178, 2000.
- [138] D. R. Livesay, P. Jambeck, A. Rojnuckarin, and S. Subramaniam. Conservation of electrostatic properties within enzyme families and superfamilies. *Biochemistry*, 42(12):3464–3473, 2003.
- [139] J. R. Livingstone, R. S. Spolar, and M. T. Record Jr. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry*, 30(17):4237–44, 1991.
- [140] W. E. Lorensen and H. E. Cline. Marching cubes: a high resolution 3d surface reconstruction algorithm. *Computer Graphics*, 21:163–169, 1987.
- [141] Q. Lu and R. Luo. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *Journal of Chemical Physics*, 119(21):11035–11047, 2003.
- [142] R. Luo, L. David, and M. K. Gilson. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *Journal of Computational Chemistry*, 23(13):1244–53, 2002.

- [143] B. A. Luty, M. E. Davis, and J. A. McCammon. Solving the finite-difference non-linear Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 13:1114–1118, 1992.
- [144] M. Lysaker, A. Lundervold, and X. C. Tai. Noise removal using fourth-order partial differential equation with application to medical magnetic resonance images in space and time. *IEEE Trans. Imag. Proc.*, 12:1579–1590, 2003.
- [145] C. M. MacDermaid and G. A. Kaminski. Electrostatic polarization is crucial for reproducing pKa shifts of carboxylic residues in turkey ovomucoid third domain. *Journal of Physical Chemistry B*, 111(30):9036–44, 2007.
- [146] J. MacKerell, A. D., D. Bashford, M. Bellot, J. Dunbrack, R. L., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. Reiher, W. E., B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.
- [147] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. Electrostatics and diffusion of molecules in solution - simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications*, 91(1-3):57–95, 1995.
- [148] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. Perspective on foundations of solvation modeling: The electrostatic contribution to the free energy of solvation. *Journal of Chemical Theory and Computation*, 4(6):877–887, 2008.
- [149] I. Massova and P. A. Kollman. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society*, 121(36):8133–43, 1999.
- [150] W. M. Matousek, B. Ciani, C. A. Fitch, B. E. Garcia-Moreno, R. A. Kammerer, and A. T. Alexandrescu. Electrostatic contributions to the stability of the GCN4 leucine zipper structure. *Journal of Molecular Biology*, 374(1):206–19, 2007.
- [151] A. Mayo. The fast solution of Poisson’s and the biharmonic equations on irregular regions. *SIAM J. Numer. Anal.*, 21:285–299, 1984.

- [152] Y. Mei, C. G. Ji, and J. Z. H. Zhang. A new quantum method for electrostatic solvation energy of protein. *J. Chem. Phys.*, 125(094906), 2006.
- [153] K. Mikula and D. Sevcovic. A direct method for solving an anisotropic mean curvature flow of plane curves with an external force. *Mathematical Methods in the Applied Sciences*, 27(13):1545–1565, 2004.
- [154] F. A. Momany. Determination of partial atomic charges from ab initio molecular electrostatic potentials. application to formamide, methanol, and formic acid. *J. Phys. Chem.*, 82:592–601, 1978.
- [155] J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, and A. Onufriev. Generalized Born model with a simple, robust molecular volume correction. *Journal of Chemical Theory and Computation*, 3(1):159–69, 2007.
- [156] Y. Mu, Y. Yang, and W. Xu. Hybrid hamiltonian replica exchange molecular dynamics simulation method employing the Poisson–Boltzmann model. *Journal of Chemical Physics*, 127(8), 2007.
- [157] R. S. Mulliken. Electronic population analysis on lcaomo [linear combination of atomic orbital-molecular orbital] molecular wave functions. *I. J. Chem. Phys.*, 23:1833–1840, 1955.
- [158] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [159] R. R. Netz and H. Orland. Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions. *European Physical Journal E*, 1(2-3):203–14, 2000.
- [160] A. Nicholls, D. L. Mobley, P. J. Guthrie, J. D. Chodera, and V. S. Pande. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *Journal of Medicinal Chemistry*, 51(4):769–79, 2008.
- [161] J. E. Nielsen, K. V. Andersen, B. Honig, R. W. W. Hooft, G. Klebe, G. Vriend, and R. C. Wade. Improving macromolecular electrostatics calculations. *Protein Engineering*, 12(8):657–662, 1999.
- [162] J. E. Nielsen and G. Vriend. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Proteins*, 43(4):403–412, 2001.

- [163] M. Nina, W. Im, and B. Roux. Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophysical Chemistry*, 78(1-2):89–96, 1999.
- [164] M. Oevermann and R. Klein. A cartesian grid finite volume method for elliptic equations with variable coefficients and embedded interfaces. *Journal of Computational Physics*, 219:749–769, 2006.
- [165] A. Okur, L. Wickstrom, M. Layten, R. Geney, K. Song, V. Hornak, and C. Simmerling. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *Journal of Chemical Theory and Computation*, 2(2):420–433, 2006.
- [166] A. Onufriev, D. Bashford, and D. A. Case. Modification of the generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B*, 104(15):3712–3720, 2000.
- [167] A. Onufriev, D. A. Case, and D. Bashford. Effective Born radii in the generalized Born approximation: the importance of being perfect. *Journal of Computational Chemistry*, 23(14):1297–304, 2002.
- [168] P. Ordejon. Linear scaling ab initio calculations in nanoscale materials with siesta. *Phys. Stat. Sol.*, 217:335, 2000.
- [169] P. Ordejon, E. Artacho, and J. M. Soler. Self-consistent order-n density-functional calculations for very large systems. *Phys. Rev. B*, 53:10441–10444, 1996.
- [170] S. Osher and R. P. Fedkiw. Level set methods: An overview and some recent results. *Journal of Computational Physics*, 169(2):463–502, 2001.
- [171] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4):919–940, 1990.
- [172] S. Osher and J. E. Sethian. Fronts Propagating with Curvature-dependent Speed: Algorithms based on the Hamilton-Jacobi Formulation. *Journal of Computational Physics*, 79:12–49, 1988.
- [173] C. S. Page and P. A. Bates. Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? *Journal of Computational Chemistry*, 27(16):1990–2007, 2006.

- [174] D. S. Palmer, V. P. Sergiievskiy, F. Jensen, and M. V. Fedorov. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *J. Chem. Phys.*, 133(044104), 2010.
- [175] R. Penfold, S. Nordholm, B. Jansson, and C. E. Woodward. A simple analysis of ion-ion correlation in polyelectrolyte solutions. *Journal of Chemical Physics*, 92(3):1915–1922, 1990.
- [176] C. S. Peskin. Numerical analysis of blood flow in the heart. *Journal of Computational Physics*, 25(3):220–52, 1977.
- [177] D. Petrey and B. Honig. GRASP2: Visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods in Enzymology*, 374:492–509, 2003.
- [178] R. A. Pierotti. A scaled particle theory of aqueous and nonaqueous solutions. *Chemical Reviews*, 76(6):717–726, 1976.
- [179] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Advances in Protein Chemistry*, 66:27–85, 2003.
- [180] N. V. Prabhu, M. Panda, Q. Y. Yang, and K. A. Sharp. Explicit ion, implicit water solvation for molecular dynamics of nucleic acids and highly charged molecules. *J. Comput. Chem.*, 29:1113–1130, 2008.
- [181] N. V. Prabhu, P. Zhu, and K. A. Sharp. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *Journal of Computational Chemistry*, 25(16):2049–2064, 2004.
- [182] E. L. Ratkova, G. N. Chuev, V. P. Sergiievskiy, and M. V. Fedorov. An accurate prediction of hydration free energies by combination of molecular integral equations theory with structural descriptors. *J. Phys. Chem. B*, 114(37):12068–2079, 2010.
- [183] M. R. Reddy, U. C. Singh, and M. D. Erion. Ab initio quantum mechanics-based free energy perturbation method for calculating relative solvation free energies. *Journal of Computational Chemistry*, 28(2):491–4, 2007.
- [184] A. E. Reed, L. A. Curtiss, and F. Weinhold. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.*, 88:899–926, 1988.

- [185] F. M. Richards. Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Bioengineering*, 6(1):151–176, 1977.
- [186] B. Roux and T. Simonson. Implicit solvent models. *Biophysical Chemistry*, 78(1-2):1–20, 1999.
- [187] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. In *Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics : computational issues in nonlinear science*, pages 259–268, Amsterdam, The Netherlands, The Netherlands, 1992. Elsevier North-Holland, Inc.
- [188] M. F. Sanner, A. J. Olson, and J. C. Spohner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38:305–320, 1996.
- [189] G. Sapiro and D. L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *Image Processing, IEEE Transactions on*, 5(11):1582–1586, 1996.
- [190] A. Sarti, R. Malladi, and J. A. Sethian. Subjective surfaces: A geometric model for boundary completion. *International Journal of Computer Vision*, 46(3):201–221, 2002.
- [191] A. Savelyev and G. A. Papoian. Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *Journal of the American Chemical Society*, 129(19):6060–1, 2007.
- [192] C. Sbert and A. F. Solé. 3d curves reconstruction based on deformable models. *Journal of Mathematical Imaging and Vision*, 18(3):211–223, 2003.
- [193] M. Schaefer and M. Karplus. A comprehensive analytical treatment of continuum electrostatics. *Journal of Physical Chemistry*, 100(5):1578–1599, 1996.
- [194] M. J. Schnieders, N. A. Baker, P. Ren, and J. W. Ponder. Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. *Journal of Chemical Physics*, 126:124114, 2007.
- [195] D. Sept, A. H. Elcock, and J. A. McCammon. Computer simulations of actin polymerization can explain the barbed-pointed end asymmetry. *Journal of Molecular Biology*, 294(5):1181–1189, 1999.

- [196] D. Sept and J. A. McCammon. Thermodynamics and kinetics of actin filament nucleation. *Biophysical Journal*, 81(2):667–674, 2001.
- [197] J. A. Sethian. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *Journal of Computational Physics*, 169(2):503–555, 2001.
- [198] Y. Y. Sham, I. Muegge, and A. Warshel. The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. *Biophysical Journal*, 74(4):1744–1753, 1998.
- [199] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *Journal of Physical Chemistry*, 94:7684–7692, 1990.
- [200] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules - theory and applications. *Annual Review of Biophysics and Biophysical Chemistry*, 19:301–332, 1990.
- [201] E. Sigfridsson and U. Ryde. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.*, 19(4):377–395, 1998.
- [202] A. C. Simmonett, A. T. B. Gilbert, and P. M. W. Gill. An optimal point-charge model for molecular electrostatic potentials. *Mol. Phys.*, 103:2789–2793, 2005.
- [203] T. Simonson. Macromolecular electrostatics: continuum models and their growing pains. *Current Opinion in Structural Biology*, 11(2):243–252, 2001.
- [204] T. Simonson. Electrostatics and dynamics of proteins. *Reports on Progress in Physics*, 66(5):737–87, 2003.
- [205] U. C. Singh and P. A. Kollman. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.*, 5:129–145, 1984.
- [206] P. Smereka. Semi-implicit level set methods for curvature and surface diffusion motion. *Journal of Scientific Computing*, 19(1):439–456, 2003.
- [207] P. Smereka. The numerical approximation of a delta function with application to level set methods. *Journal of Computational Physics*, 211(1):77–90, 2006.
- [208] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *Image Processing, IEEE Transactions on*, 7(3):310–318, 1998.

- [209] J. M. Soler. The siesta method for ab initio order-n materials simulation. *J. Phys.: Condens. Matter*, 14:2745–2779, 2002.
- [210] Y. Song, Y. Zhang, C. L. Bajaj, and N. A. Baker. Continuum diffusion reaction rate calculations of wild-type and mutant mouse acetylcholinesterase: Adaptive finite element analysis. *Biophysical Journal*, 87(3):1558–66, 2004.
- [211] Y. Song, Y. Zhang, T. Shen, C. L. Bajaj, J. A. McCammon, and N. A. Baker. Finite element solution of the steady-state Smoluchowski equation for rate constant calculations. *Biophysical Journal*, 86(4):2017–2029, 2004.
- [212] R. S. Spolar, J. H. Ha, and M. T. Record Jr. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 86(21):8382–8385, 1989.
- [213] F. H. Stillinger. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Solution Chem.*, 2:141 – 158, 1973.
- [214] A. J. Stone and M. Alderton. Distributed multipole analysis methods and applications. *Mol. Phys.*, 56:1047–1064, 1985.
- [215] Y. H. Sun, P. R. Wu, G. W. Wei, and G. Wang. Evolution operator based single-step method for image processing. *Int. J. Biomed. Imaging*, 83847:1–27, 2006.
- [216] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysical Journal*, 86(1):67–74, 2004.
- [217] J. M. J. Swanson, J. Mongan, and J. A. McCammon. Limitations of atom-centered dielectric functions in implicit solvent models. *Journal of Physical Chemistry B*, 109(31):14769–72, 2005.
- [218] Y. Takano and K. N. Houk. Benchmarking the conductor-like polarizable continuum model (cpcm) for aqueous solvation free energies of neutral and ionic organic molecules. *Journal of Chemical Theory and Computation*, 1(1):70–77, 2005.
- [219] C. Tan, L. Yang, and R. Luo. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *Journal of Physical Chemistry B*, 110(37):18680–18687, 2006.

- [220] J. J. Tan, W. Z. Chen, and C. X. Wang. Investigating interactions between HIV-1 gp41 and inhibitors by molecular dynamics simulation and MM-PBSA/GBSA calculations. *Journal of Molecular Structure: THEOCHEM*, 766(2-3):77–82, 2006.
- [221] Z. J. Tan and S. J. Chen. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *Journal of Chemical Physics*, 122:044903, 2005.
- [222] M. Tanaka and A. Y. Grosberg. Giant charge inversion of a macroion due to multivalent counterions and monovalent coions: Molecular dynamics study. *Journal of Chemical Physics*, 115(1):567–574, 2001.
- [223] C. L. Tang, E. Alexov, A. M. Pyle, and B. Honig. Calculation of pKas in RNA: On the structural origins and functional roles of protonated nucleotides. *Journal of Molecular Biology*, 366(5):1475–96, 2007.
- [224] D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard, and B. Honig. Accurate first principles calculation of molecular charge distribution and solvation energies from ab initio quantum mechanics and continuum dielectric theory. *J. Am. Chem. Soc.*, 116:11875 – 11882, 1994.
- [225] I. Terekhova, A. O. Romanova, R. S. Kumeev, and M. V. Fedorov. Selective Na^+/K^+ effects on the formation of α -cyclodextrin complexes with aromatic carboxylic acids: Competition for the guest. *J. Phys. Chem. B*, 114(37):12607–2613, 2010.
- [226] H. Tjong and H. X. Zhou. GBr6NL: A generalized Born method for accurately reproducing solvation energy of the nonlinear Poisson-Boltzmann equation. *Journal of Chemical Physics*, 126:195102, 2007.
- [227] J. Tomasi, B. Mennucci, and R. Cammi. Quantum mechanical continuum solvation models. *Chem. Rev.*, 105:2999–3093, 2005.
- [228] J. Tomasi and M. Persico. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.*, 94:2027–2094, 1994.
- [229] V. Tsui and D. A. Case. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *Journal of the American Chemical Society*, 122(11):2489–2498, 2000.
- [230] V. Tsui and D. A. Case. Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *Journal of Physical Chemistry B*, 105(45):11314–11325, 2001.

- [231] D. M. Tully-Smith and H. Reiss. Further development of scaled particle theory of rigid sphere fluids. *Journal of Chemical Physics*, 53(10):4015–25, 1970.
- [232] A. Vitalis, N. A. Baker, and J. A. McCammon. ISIM: A program for grand canonical Monte Carlo simulations of the ionic environment of biomolecules. *Molecular Simulation*, 30(1):45–61, 2004.
- [233] A. Vitalis and R. V. Pappu. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry*, 30(5):673–99, 2009.
- [234] R. C. Wade, R. R. Gabdouliline, and F. De Rienzo. Protein interaction property similarity analysis. *International Journal of Quantum Chemistry*, 83(3-4):122–127, 2001.
- [235] J. A. Wagoner and N. A. Baker. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proceedings of the National Academy of Sciences of the United States of America*, 103(22):8331–6, 2006.
- [236] A. Wallquist and B. J. Berne. Computer-simulation of hydrophobic hydration forces stacked plates at short-range. *Journal of Physical Chemistry*, 99:2893–2899, 1995.
- [237] M. L. Wang and C. F. Wong. Calculation of solvation free energy from quantum mechanical charge density and continuum dielectric theory. *J. Phys. Chem. A*, 110:4873–4879, 2006.
- [238] M. L. Wang, C. F. Wong, J. H. Liu, and P. X. Zhang. Efficient quantum mechanical calculation of solvation free energies based on density functional theory, numerical atomic orbitals and poissonboltzmann equation. *Chemical Physics Letters*, 442:464–467, 2007.
- [239] A. Warshel and A. Papazyan. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Current Opinion in Structural Biology*, 8(2):211–7, 1998.
- [240] A. Warshel, P. K. Sharma, M. Kato, and W. W. Parson. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1764(11):1647–76, 2006.
- [241] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *Journal of Molecular Biology*, 157(4):671–9, 1982.

- [242] G. W. Wei. Generalized Perona-Malik equation for image restoration. *IEEE Signal Processing Letters*, 6(7):165–7, 1999.
- [243] G. W. Wei. Differential geometry based multiscale models. *Bulletin of Mathematical Biology*, 72:1562 – 1622, 2010.
- [244] G. W. Wei and Y. Q. Jia. Synchronization-based image edge detection. *Europhysics Letters*, 59(6):814, 2002.
- [245] G. W. Wei, Y. H. Sun, Y. C. Zhou, and M. Feig. Molecular multiresolution surfaces. *arXiv:math-ph/0511001v1*, pages 1 – 11, 2005.
- [246] P. Weinzinger, S. Hannongbua, and P. Wolschann. Molecular mechanics PBSA ligand binding energy and interaction of efavirenz derivatives with HIV-1 reverse transcriptase. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 20(2):129–134, 2005.
- [247] T. J. Willmore. *Riemannian Geometry*. Oxford University Press, USA, 1997.
- [248] K. Wolfgang. *Differential Geometry: Curves-Surface-Manifolds*. American Mathematical Society, 2002.
- [249] G. Xu, Q. Pan, and C. L. Bajaj. Discrete surface modeling using partial differential equations. *Computer Aided Geometric Design*, 23(2):125–145, 2006.
- [250] M. Xu and S. L. Zhou. Existence and uniqueness of weak solutions for a fourth-order nonlinear parabolic equation. *JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS*, 325:636–654, 2007.
- [251] A. S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. On the calculation of pK(a)s in proteins. *Proteins-Structure Function and Genetics*, 15(3):252–265, 1993.
- [252] S. Yu, W. Geng, and G. W. Wei. Treatment of geometric singularities in implicit solvent models. *Journal of Chemical Physics*, 126:244108, 2007.
- [253] S. Yu and G. W. Wei. Three-dimensional matched interface and boundary (MIB) method for treating geometric singularities. *Journal of Computational Physics*, 227:602–632, 2007.
- [254] S. Yu, Y. Zhou, and G. W. Wei. Matched interface and boundary (MIB) method for elliptic problems with sharp-edged interfaces. *Journal of Computational Physics*, 224(2):729–756, 2007.

- [255] D. Zhang, J. Suen, Y. Zhang, Z. Radic, P. Taylor, M. Holst, C. Bajaj, N. A. Baker, and J. A. McCammon. Tetrameric mouse acetylcholinesterase: Continuum diffusion rate calculations by solving the steady-state Smoluchowski equation using finite element methods. *Biophysical Journal*, 88(3):1659–65, 2005.
- [256] Y. Zhang, G. Xu, and C. Bajaj. Quality meshing of implicit solvation models of biomolecular structures. *Computer Aided Geometric Design*, 23(6):510–30, 2006.
- [257] S. Zhao. High order matched interface and boundary methods for the helmholtz equation in media with arbitrarily curved interfaces. *J. Comput. Phys.*, 229:3155–3170, 2010.
- [258] S. Zhao and G. W. Wei. High-order FDTD methods via derivative matching for Maxwell’s equations with material interfaces. *Journal of Computational Physics*, 200(1):60–103, 2004.
- [259] Y. C. Zhou, M. Feig, and G. W. Wei. Highly accurate biomolecular electrostatics in continuum dielectric environments. *Journal of Computational Chemistry*, 29:87–97, 2008.
- [260] Y. C. Zhou and G. W. Wei. On the fictitious-domain and interpolation formulations of the matched interface and boundary (MIB) method. *Journal of Computational Physics*, 219(1):228–246, 2006.
- [261] Y. C. Zhou, S. Zhao, M. Feig, and G. W. Wei. High order matched interface and boundary method for elliptic equations with discontinuous coefficients and singular sources. *Journal of Computational Physics*, 213(1):1–30, 2006.
- [262] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn, and M. Levitt. Finite-difference solution of the Poisson-Boltzmann equation: complete elimination of self-energy. *Journal of Computational Chemistry*, 17:1344–1351, 1996.
- [263] J. Zhu, E. Alexov, and B. Honig. Comparative study of generalized Born models: Born radii and peptide folding. *Journal of Physical Chemistry B*, 109(7):3008–22, 2005.