UNIVERSITY OF CALIFORNIA, SAN DIEGO

Characterizing Biomolecular Recognition and Solvation with End-Point Free Energy Calculations and Implicit Solvent Models

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy
in

Chemistry

by

Jessica M.J. Swanson

Committee in charge:

Professor J. Andrew McCammon, Chair
Professor Elizabeth Komives, Co-chair
Professor David A. Case
Professor Jose Onuchic
Professor Wei Wang
Professor Peter Wolynes

2006

The dissertation of Jessica M.J. Swanson is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Co-chair

_____
Chair

University of California, San Diego

2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The work that follows is a manifestation of the great teachers and role-models that have helped direct my curiosity and shape my passions/desire to contribute to a better future. First and foremost I thank my advisor, Dr. J. Andrew McCammon, whose guidance was invaluable and whose academic ability and grace I will forever hope to emulate. I thank my Committee members, Drs. Komives, Onuchic, Wang, and Wolynes for their time, efforts and advice. I owe a special thanks to Dr. David A. Case from the Scripps Research Institute whose voluntary committee service and research advice was greatly appreciated. I thank the UCSD Chemistry and Biochemistry faculty whose talent and emphasis on a collaborative science make this graduate program both a challenge and pleasure. I thank the many great teachers that inspired me through my formative years, especially Dr. Katja Lindenberg whose teaching skills are unsurpassed. Finally, I thank Dr. Alexei Stuchebrukhov who introduced me to the field of computational biophysics and continues to be an esteemed mentor and colleague.

Although computational chemistry is often done alone, my most insightful and productive moments came in the presence of collaboration. I am grateful to have worked with many distinguished professors including Drs. Benoit Roux, Charlie Brooks, and Nathan Baker, and many talented colleagues, especially Drs. Richard Henchman, Stewart Adcock, Joachim Dzubiella, John Mongan, Donald Hamelberg, Chia-en Chang, and Tushar Jain. I am very thankful to the Molecular Biophysics Training Grant and the Center for Biological Physics for providing financial support a rich academic community to engage with and learn from. I am also grateful to the entire McCammon group for many entertaining if not productive discussions.

Since work without play ceases to be fun, I thank my wonderful friends, especially Amaya, Stacia, and Eileen, for adding balance and joy to my life. I thank my family for being that essential net I call home; my Mother for her amazing capacity to love and her never-faltering support; and finally, my father who blessed me with his insatiable curiosity, amazes me with his academic insight, and who *never* ceases to believe in me.

Chapter 2 is a reprint in full of material that appeared in. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy Jessica M.J. Swanson, Richard H. Henchman, and J. Andrew McCammon. Biophysical Journal, 86 67-74, January 2004. I was the primary researcher and author of this work.

Chapter 3 is a reprint in full of material that appeared in The Entropic Cost of Protein-Protein Association: A Case Study on Acetylcholinesterase Binding to Fasciculin-2, David D.L. Minh, J.M. Bui, C.E. Chang, T. Jain, J.M.J. Swanson, J.A. McCammon, Biophys. J. 89, L25-L27 (2005). I was a secondary researcher and author of this work.

Chapter 4 is a reprint in full of material that appeared in Optimized radii for Poisson-Boltzmann calculations with the AMBER force field, Swanson, J.M.J., S.A. Adcock, and J.A. McCammon, J. Chem. Theory Comput. (3); 484-493 (2005). I was the primary researcher and author of this work.

Chapter 5 is a reprint in full of material that appeared in Limitations of atom-centered dielectric functions in implicit solvent models. Jessica M.J. Swanson, John Mongan and J. Andrew McCammon. Journal of Physical Chemistry B, 109(31) 14769-14772, August 2005. I was the primary researcher and author of this work.

Chapter 6 is a reprint in full of material that will appear in Optimizing the Poisson dielectric boundary with explicit solvent forces: Lessons learned with atom-centered dielectric functions. , Jessica M.J. Swanson, J. Wagoner, N. A. Baker, and J.A. McCammon, J Chem Theory and Comput (submitted). I was the primary researcher and author of this work.

Chapter 7 is a reprint in full of material that appeared in Coupling Hydrophobicity, Dispersion, and Electrostatics in Continuum Solvent Models. Dzubiella J., J.M.J. Swanson, J.A. McCammon, Phys. Rev. Lett.96, 087802 (2006). I was a secondary researcher and author of this work.

Chapter 8 is a reprint in full of material that appeared in Coupling nonpolar and polar solvation free energies in implicit solvent models. Dzubiella J., J.M.J. Swanson, J.A. McCammon, J. Chem. Phys.12, 084905 (2006). I was a secondary researcher and author of this work.

VITA

April 11, 1977	Born Austin, Texas
1996-97	Undergraduate Technician in Analytical Chemistry, Los Alamos National Laboratory
1998-99	Undergraduate Researcher in Computational Chemistry - University of California, Davis
2000	BS Biochemistry, University of California, Davis
2002-2003	Funded Trainee - Molecular Biophysics Training Grant (MBTG), University of California, San Diego
2002-2004	Teaching Assistant, University of California, San Diego
2003-2005	Student Committee Member – MBTG, University of California, San Diego
2003-2006	Fellow Center for Biological Physics, University of California, San Diego
2005	Co-Lecturer, University of California, San Diego
2006	Ph.D. in Chemistry, University of California, San Diego

PUBLICATIONS

Optimizing Poisson dielectric boundaries: Lessons learned with atom-centered dielectric functions Swanson J.M.J., J. Wagoner, N.A. Baker, J. A. McCammon, J. Chem. Theory and Comput. (submitted)

Coupling nonpolar and polar solvation free energies in implicit solvent models Dzubiella J., J.M.J. Swanson, J.A. McCammon, J. Chem. Phys. 124, 084905 (2006).

Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models Dzubiella J., J.M.J. Swanson, J.A. McCammon, Phys. Rev. Lett. 96, 087802 (2006).

The Entropic Cost of Protein-Protein Association: A Case Study on Acetylcholinesterase Binding to Fasciculin-2 Minh, D.D.L., J.M. Bui, C.E. Chang, T. Jain, J.M.J. Swanson, J.A. McCammon, Biophys. J. in press (2005).

Limitations of Atom-Centered Dielectric Functions in Implicit Solvent Models Swanson, J.M.J., J. Mongan, and J.A. McCammon, J. Phys. Chem. B, 109, 31, 14769-14772 (2005).

Optimized radii for Poisson-Boltzmann calculations with the AMBER force field Swanson, J.M.J., S.A. Adcock, and J.A. McCammon, J. Chem. Theory Comput. (3); 484-493 (2005).

Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy Swanson, J.M.J., R.H. Henchman, J.A. McCammon. Biophys. J., 86, 67-74 (2004).

Computer Simulation of Water in Cytochrome C Oxidase. Xuehe Zheng, Dmitri Medvedev, Jessica Swanson, Alexei Stuchebrukhov, BBA, 1557: 99-107 (2003).

FIELDS OF STUDY

Major Field: Physical Chemistry

Studies in computational chemistry and modeling
Professor Alexei A. Stuchebrukhov, University of California, Davis

Studies in computational biophysics and dynamics
Professor J. Andrew McCammon, University of California, San Diego

Studies in electrostatics and implicit solvation
Professor Nathan A. Baker, Washington University School of Medicine

ABSTRACT OF THE DISSERTATION

Characterizing Biomolecular Recognition and Solvation with End-Point Free Energy Calculations and Implicit Solvent Models

by

Jessica M. J. Swanson

Doctor of Philosophy in Chemistry

University of California, San Diego, 2006

Professor J. Andrew McCammon, Chair

Professor Elizabeth Komives, Co-chair

Biomolecular recognition, the nexus of individual molecules and function, and solvation, the medium in which it occurs, are two of the most important, well-studied, and yet enigmatic phenomena in biology. This thesis explores the combination of theory and computational algorithms to study these two phenomena. End-point free energy calculations are used to predict and characterize biomolecular recognition. A theoretical framework is presented which properly accounts for standard state conditions and highlights several areas of potential improvement. A novel method is presented for calculating the association free energy, which arises from one molecule losing its translational and rotational freedom upon binding. The entropic penalties of protein-ligand and protein-protein associations are compared and the physical trends that affect this penalty are presented. Compatibility between the implicit and explicit solvent energies is shown to be especially important in end-point free energy analysis. Energetic compatibility, in addition to implicit solvent accuracy, can be improved by fitting implicit solvent parameters to explicit solvent simulations. Several variations of this approach are presented, each of which optimizes the solute radii which define the dielectric boundary between the low dielectric solute and the high dielectric solvent in Poisson-based implicit solvent models. The radii are first optimized to reproduce explicit solvent charging free energies from explicit solvent simulations.

Then, atomic forces are shown to increase the optimization efficiency and improve the resulting implicit solvent parameters. Molecular surfaces are compared to atom-centered dielectric functions, and the latter are shown to create unphysical high dielectric regions in spaces between atoms which are too small for water molecules to penetrate. Optimized radii are presented for molecular surfaces and spline-smoothed surfaces created for stable and efficient force calculations. Finally, the coupling between polar and nonpolar solvation energies is discussed. This coupling has never been accounted for in an implicit solvent framework, but has been demonstrated in solutes ranging from small molecules to nanosolutes, including several biological systems. A new implicit solvent formalism is presented which accounts for coupling by expressing the system free energy as a functional of the solvent volume exclusion function.

Chapter One

Introduction

Biomolecular recognition, the specific association of two biomolecules, is of fundamental importance in the study of biological systems; it is the phenomenon that initiates and perpetuates most cellular processes from replication, transcription, and translation, to cell trafficking and neuronal signaling. Measuring the probability of molecular association is an essential tool in experimental biology. It is most commonly accomplished with binding assays which report concentrations of bound and unbound molecules. Similarly, predicting the probability of association is an important tool in theoretical biology. Theory and computational simulations can provide access to systems that can not be analyzed in the laboratory and insight into the microscopic nature of molecular association. Such insights not only help us characterize biological systems but help us mimic and constructively manipulate them. In fact most pharmaceuticals interfere with or use molecular recognition to alter the function of specific biomolecules. Similarly, many biomimetic sensing approaches use recognition to signal the presence of specific molecules. For these reasons, predicting the probability of molecular association, i.e. binding affinities, is essential in rational drug design; developing theoretical techniques for doing so has been an active area of medically-motivated research.

Noncovalent association is traditionally thought to be driven by inter-solute Coulombic attraction, the burial of hydrophobic surfaces and concomitant release of bound water molecules. The process can, however, be intriguingly complex. Examples of interesting binding interactions include negative electrostatic potential in anion binding sites, increased entropic freedom for proteins and water molecules upon binding, and a diverse array of specific solute-solvent

interactions (e.g. bridging water molecules in binding interfaces).[1] Such complexity and diversity not only deserves but demands continued theoretical and experimental attention. The first half of this thesis explores the prediction and characterization of noncovalent biomolecular association with end-point free energy and entropy calculations.

Since this phenomenon is largely influenced by the surrounding aqueous environment, properly accounting for the influence of water is an additional focus. In fact accounting for solvation is essential in almost all biophysical queries, so the need for accurate and efficient solvation models pervades computational biophysics. The most rigorous models represent solvent molecules atomistically while more efficient models represent water implicitly by replacing individual molecules with a linearly polarizable continuum. Improving the accuracy and fundamental theory of such implicit solvent models is the focus of the second half of this thesis.

1.1 Predicting Biomolecular Recognition with Free Energy and Entropy Calculations

The probability that any physical process will occur is logarithmically related the change in free energy associated with that process. Experimentalists generally measure the probability of molecular association directly with the concentrations of bound and unbound molecules, i.e. equilibrium constants. Theoreticians, on the other hand, predict association by calculating the change in free energy upon binding, i.e. a binding free energy. The statistical mechanics that connects binding free energies (or equilibrium constants) to the microscopic interactions between the solutes (hereafter referred to as protein and ligand) and solvent molecules has been well known for some time.[2, 3] The computational application of this theory, however, remains a constantly evolving challenge.[1, 4, 5] Several formalisms have been developed, each with differing degrees of efficiency and accuracy. The most rigorous methods use particle based

simulations to measure the change in free energy of either alchemical or positional transformations. The alchemical transformations make use of a thermodynamic cycle in which the bound molecule (ligand A) is transformed into a new molecule (ligand B) in the bound complex and in the liquid environment.[6] Free energy perturbation (FEP) or thermodynamic integration (TI) are used to calculate the free energies for these two simple processes, the difference of which is the binding free energy of ligand B relative to ligand A. Because these simulations require adequate sampling of the transformation intermediates, they are most amenable to simple transformations between similar ligands. They are ideal for the final stages of drug development when a limited number of similar drug candidates need to be accurately compared and each can be afforded several days of simulation time.[7]

In order to calculate an absolute binding affinity, the ligand must be transformed completely out of existence (i.e. interaction) in the bound complex and aqueous environment. This process, often referred to as the 'double decoupling' method, has historically been limited to small ligands due to its computational expense.[8] It was recently applied to a set of eight larger ligands on the Folding at Home distributed computing architecture.[9] An alternative approach uses a positional transformation to pull the ligand out of the binding pocket. Integrating the free energy along the pulling trajectory (i.e. the potential of mean force; PMF) yields the absolute binding affinity. This approach has also seen limited application because of the high computational demand and challenge of sampling along the pulling reaction coordinate. However, a recent advance introduced by Woo and Roux[10] circumvents sampling limitations by restricting the ligand's position, orientation, and conformation to that found in the complex. These so called 'staged free energy calculations' have been applied to a range of ligands and show great promise for attaining accurate absolute binding affinities.[11] They will likely be challenging when the ligand is significantly buried such that the protein either has to be forced open and closed or passed through to reach the unbound form.

Despite continued development and progress, the particle-based approaches remain too computationally expensive for most large and/or complex systems and for high-throughput applications, such as the early stages of rational drug design where libraries of thousands of drug candidates have to be screened for potential leads. For such applications, replacing portions of the system, such as the solvent, with continuum models decreases computational demands significantly. Continuum solvation models represent water as a linearly polarizable medium. The most commonly used methods solve either the Poisson, the Poisson-Boltzmann[12, 13], or the generalized Born[14, 15] equations for electrostatic solvation effects and use a term proportional to the molecule's solvent accessible surface area (SASA)[16, 17] for the nonelectrostatic effects. These methods will be the focus of chapters 4-7. Combining continuum models with perturbation approaches decreases simulation times as recently demonstrated by Lee et al. who used continuum PMF simulations to calculate the absolute binding affinity of same system that will be explored in chapter 2.[18]

Another technique that reduces computational demand is to limit the sampling to only the two end points of the binding reaction, i.e. the bound and unbound states. Taking the difference between the total free energies of the bound and unbound systems yields the absolute binding free energy. The most popular end-point free energy method was co-developed by the late Peter Kollmann and David. A. Case in the late nineties.[19] It quickly became known by its acronym, MM/PBSA, so named because it averages molecular mechanics (MM) enthalpies and Poisson-Boltzmann surface area (PBSA) solvation free energies of conformations generated with an explicit solvent simulation for the total system free energy. Although it is not included in the acronym, the entropic contribution to the binding free energy is usually approximated with methods such as normal mode analysis or quasi-harmonic analysis. Continued development of MM/PBSA-like approaches have been highly sought after by the rational drug design community for several reasons. First, they offer insight in to the microscopic phenomena and energetic trends that govern association. These insights can be used to improve physically motivated

docking and scoring functions, which typically include energy approximations and empirical entropy corrections. Second, they can be used to re-score drug candidates which have been selected by less accurate docking algorithms. The "relaxed complex method" uses the technique to incorporate protein flexibility into the drug design protocol.[20] In the relaxed complex method an ensemble of protein conformations is generated with standard or accelerated molecular dynamics, a select number of ligands are then docked to these conformations, and finally the docked structures are rescored with MM/PBSA for a more accurate ranking of potential drugs.

When this thesis work began, MM/PBSA had received over five years of widespread development and application. It was still, however, lacking the theoretical foundation necessary to account for standard state conditions and to attain rigorous results. This foundation is presented in chapter 2 with a derivation that connects end-point free energy approaches, such as MM/PBSA, to statistical thermodynamics. This work highlights several shortcomings of previous MM/PBSA implementations, including: the failure to account for the association free energy, which results from one molecule's loss of translational and rotational freedom from its standard state concentration; inaccurate and inconsistent conformational free energies, those which result from changes in both molecules' intramolecular motions; and incompatibility between the explicit and implicit solvent models. A simple protein-ligand system is used to present a novel method for calculating the association free energy directly from a molecular dynamics simulation, and to demonstrate the challenges of calculating the conformational free energy.

The theoretical formalism presented in chapter 2 clearly defines the association free in terms of the ligand's loss of translational and rotational freedom upon binding. The entropic contribution to the association free energy has been well understood for over 40 years,[21] but calculating the bound molecule's residual freedom has been a continually evolving challenge. Chapter 2 provides a brief history of efforts to meet this challenge and compares them to our presented method, which measures the bound ligand's residual freedom directly from the

molecular dynamics simulation. A protein-small molecule system is used to emphasize the importance of using the principal modes of translational and rotational motion and accounting for correlations between these modes. In chapter 3, we extend this analysis from protein-small molecule association to the first application to date of protein-protein association. Chapter 3 additionally explores improvements on single-well and coupling approximations of the ligand's translational and rotational freedom with histogram analysis.

Decomposing a binding free energy into specific contributions such as conformational and association entropies is difficult to do rigorously. The necessary assumptions that are discussed in chapters 2 and 3 introduce small errors into energy components. Nevertheless, calculating energetic components yields insights into the physical driving forces behind association. For example, comparing association entropies for water-protein, small molecule-protein, and protein-protein binding reveals several interesting trends in the entropic penalty of association. Small molecules with weak protein-ligand interactions have the most translational and rotational freedom in the bound form. In fact, water molecules have been shown to actually gain entropy upon binding.[22] Similarly, small ligands, such as butanone in chapter 2, have significant translational and rotational freedom and low association entropic penalties. In contrast, large ligands, such as the protein-protein complex studied in chapter 3, have much less translational and rotational freedom and larger entropic penalties. These findings point to a simple trend: larger ligands, stronger intermolecular interaction energies, and larger binding interfaces all increase the entropic penalty of association.

1.2 Improving Poisson-based Implicit Solvent Models

One of the main limitations of accuracy in end-point free energy methods is a lack of compatibility between the implicit and explicit energy models. This compatibility is a necessary requirement of any method that energetically couples implicit and explicit solvent models

including continuum dynamics, pKa calculations, hybrid implicit-explicit solvation schemes, etc. It is particularly important in end-point free energy calculations since the solute-solute interaction energy and the solutes' solvation energies are large and opposing numbers which largely cancel for a comparatively small total free energy. Thus, relatively small errors in either MM electrostatic or PB solvation energies result in large free energy errors. In order to minimize these errors the implicit solvent model can be empirically fit to explicit solvent quantities such as solvation free energies and forces.

The dynamic and thermodynamic properties of a solvated solute are strongly influenced by microscopic solute-solvent interactions and the resulting microscopic structure and organization of the water that surrounds the solute. Therefore predicting solvation free energies and forces accurately for a variety of solutes requires a formalism that accounts for solute-solvent interactions (electrostatic and nonelectrostatic) with sufficient detail. Microscopic models (e.g. explicit solvent simulations) are successful in this regard because they treat these interactions explicitly. Implicit solvent models offer a less physically accurate but more computationally efficient *macroscopic* approach, replacing individual molecular interactions with an implicit representation of water. It has long been recognized that the lack of physical accuracy must be compensated by fitting empirical parameters to known quantities in an atom and/or system specific manner,[23] e.g. fitting Born radii to solvation free energies.[24, 25]

The relationship between explicit and implicit solvent models is explored in chapters 4 through 8. Chapter 4 through 6 focus on the polar portion of implicit solvent models, specifically those which solve the Poisson or Poisson-Boltzmann equation for the electrostatic potential of a system as a function of the solute's charge distribution and a spatially varying dielectric coefficient. The resulting solvation energies and forces are particularly sensitive to the nature and location of the boundary which separates the low dielectric solute from the high dielectric solvent.

In chapter 4 the solute radii used to define this boundary are optimized with explicit solvent charging free energies. Two dielectric boundary definitions are explored. The first, and most commonly used, definition is an abrupt transition from low to high dielectric values at the molecular surface, as defined by Lee and Richards[26] or Connolly[27]. This surface is defined by the contact surface between a solvent probe and the solute atoms as the probe is rolled around the solute. The second dielectric boundary, designed for stable and efficient continuum force calculations, defines a smooth dielectric transition at the solute surface with overlapping *atom-centered* polynomial or Gaussian functions.[28, 29] Changing the nature of the dielectric transition also changes its optimal location. Therefore solute radii are optimized for either the molecular or the spline smoothed surface definitions in chapter 4.

In the molecular surface definition, the spaces in between solute radii (herein called *interstitial spaces*) which are two small for the solvent probe to penetrate are included in the low dielectric solute. In atom-centered definitions interstitial spaces are included in the high dielectric solvent region. Although interstitial high dielectrics are not physically accurate, their role in continuum calculations has been not fully appreciated in recent literature. Chapter 5 explores the magnitude and impact of interstitial high dielectric regions formed by the spline-smoothed surfaces. The magnitude is shown to be quite large for a variety of protein systems and their presence is shown to introduce errors in solvation free energies and the free energy surfaces of nonbonded interactions.

In chapter 6, a novel approach of implicit solvent optimization is presented in which explicit solvent atomic forces are used instead of molecular solvation free energies. This proves to be much more efficient and robust as it provides atomic information on the performance of an implicit solvent model. It can be applied to any system with explicit solvent parameters; thus it can be used to determine the optimal continuum parameters when experimental solvation energies are unavailable and explicit solvent charging free energies are computationally prohibitive. This chapter continues to probe the limitations of atom-centered dielectric functions and shows that

errors are more significant in large, globular solutes. Despite these limitations, spline smoothed surfaces show surprising accuracy for small, compact systems.

The concept of polar and nonpolar coupling is introduced in chapter 6. As previously mentioned canonical implicit solvent models divide the solvation free energy into polar and nonpolar contributions. This division is rigorously based on a thermodynamic cycle in which the solute charges are turned off in vacuum, the neutral solute cavity is solvated, and the solute is recharged in the solvent environment. The polar contribution is treated with electrostatic calculations as previously described. The nonpolar contribution, due to hydrophobic and dispersion interactions, is commonly, though inadequately[30, 31], treated with a solvent accessible surface area (SASA) model.[16, 17, 32] Both polar and nonpolar models depend critically on the location of the solvent-solute interface. Similarly the location of the solvent-solute interface can vary according to local polar and nonpolar solvent-solute potentials. This suggests that polar and nonpolar solvation contributions are coupled and should be treated as coupled in implicit solvation formalisms. In fact the strong coupling that exists between polar and nonpolar solvation contributions has been emphasized in many recent studies on everything from small amphiphiles to protein folding.[33, 34] In chapters 7 and 8 a novel continuum formalism is presented which accounts for polar and nonpolar coupling by expressing the system free energy as a functional of the solute volume exclusion function. Unlike existing implicit solvent approaches, the solvent accessible surface is an output of this theory. Accounting for coupling is expected to be crucial for a complete characterization of biomolecular solvation. The theory presented in chapters 7 and 8 is only limited by crude curvature and dielectric descriptions. Future efforts to improve these descriptions based on physical rational and further empirical corrections may lead to the first implicit solvent framework that can describe the solvation of diverse solutes without system-dependent-fit-parameters and that captures the sensitive balance of polar and nonpolar solvation contributions.

In summary, this dissertation explores the theory and methodology of end-point free energy and association entropy calculations for the prediction of noncovalent biomolecular association, and implicit solvation models for the representation of biomolecular solvation.

Chapter Two

Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy

ABSTRACT

The prediction of absolute ligand-receptor binding affinities is essential in a wide range of biophysical queries, from the study of protein-protein interactions to structure based drug design. End-point free energy methods, such as the Molecular Mechanics Poisson Boltzmann Surface Area (MM/PBSA) model, have received much attention and widespread application in recent literature. These methods benefit from computational efficiency as only the initial and final states of the system are evaluated, yet there remains a need for strengthening their theoretical foundation. In this chapter a clear connection between statistical thermodynamics and end-point free energy models is presented. The importance of the association free energy, arising from one molecule's loss of translational and rotational freedom from the standard state concentration, is addressed. A novel method for calculating this quantity directly from a molecular dynamics simulation is described. The challenges of accounting for changes in the protein conformation and its fluctuations from separate simulations are discussed. A simple first order approximation of the configuration integral is presented to lay the ground work for future efforts. This model has been applied to FKBP12, a small immunophilin that has been widely studied in the drug industry for its potential immunosuppressive and neuroregenerative effects.

2.1 Introduction

The theoretical prediction of binding affinities is one of the most important problems in computational biochemistry. It complements experimental analysis and adds molecular insight to the macroscopic properties measured therein. It serves as a cornerstone in disease research and rational drug design where accurate scoring functions remain a challenge. It is no wonder, then, that computational models aimed at the prediction of binding affinities have been highly sought after for over half a century and are the subject of frequent reviews [1, 2, 4, 5].

The theory underlying binding affinities has been well described by many, yet the complexity and accuracy of its application has varied. The most rigorous methods involve alchemical or structural transformations such as Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) [6, 35]. The accuracy of these methods relies on equilibrium sampling of the entire transformation path, from an initial to a final state. The computational demand of adequate sampling makes relative binding affinities between similar ligands the most amenable targets of FEP and TI. Relative binding affinities between diverse ligands and absolute binding affinities pose more of a challenge.

End-point free energy models, wherein only the initial and final states of the system are evaluated, present a desirable alternative to perturbation simulations. They are less computationally expensive making them suitable for a greater variety of systems and problems. They are typically based on partitioning the free energy into a sum of enthalpic and entropic contributions [36-38]. Frameworks that use implicit solvent approximations reduce computational demands even further. Although all such models are founded in statistical mechanics, there is a need for strengthening the theoretical framework of many to account for standard state dependence and entropic considerations. Other implicit solvent, end-point models have thorough theoretical descriptions [39-41], yet their remains a need for further analyses

regarding which contributions to include, how to measure them, and which approximations are appropriate to make.

This chapter focuses on providing a clear theoretical foundation for end-point free energy models. Two issues that have been inconsistently applied in previous analyses are highlighted; the association free energy, which results from one molecule's loss of translational and rotational freedom from the standard state, and the conformational free energy due to changes in both molecules' intramolecular motions. An implicit solvent approximation is used to evaluate the initial and final equilibrium ensembles generated during explicit solvent MD simulations. The association free energy is thoroughly discussed and measured from the simulation. Determining the conformational free energy represents the most challenging aspect of this work and of all such methods as it is tied to the evaluation of the internal configuration integral of the bound and free systems. A first order approximation assumes that the changes in conformational freedom are minimal and that the energy landscape can be characterized from a sufficiently long MD simulation. This simplification serves as a necessary stepping stone for more advanced evaluations of the configuration integral.

To illustrate this method, a small, fairly rigid protein-ligand system, FK506 binding protein (FKBP12) and the ligand 4-hydroxy-2-butanone, was chosen. FKBP12 is an immunophilin that, when bound by the immunosuppressant drug FK506, blocks early T-cell activation via calcineurin inhibition. Smaller ligands that mimic FK506 as potential immunosuppressive drugs have been highly sought after. In an attempt to characterize its binding properties, the crystal structure of FKBP12 bound by several small molecules including 4-hydroxy-2-butanone (BUT) was determined [42]. With only six heavy atoms and four rotatable bonds, BUT was one of the smallest ligands to bind FKBP12 with a measured binding affinity, K_i, of 500 μM. Despite the current method's exclusion of the changes in conformational free energy, which is expected to be positive, the calculated change in free energy was only 10 kJ/mol lower than that measured in experiment. The small magnitude of this

discrepancy is consistent with the low binding affinity of the ligand, which is unlikely to substantially perturb the protein's conformation or fluctuations.

First, the theoretical framework will be described. Some of the foundation presented in the literature [2] will be reviewed for a complete description. The simulation methods and numerical results will then be presented. Evaluation of the association free energy will be compared to previously published methods and deviations from experimental results will be discussed. Finally the method and its potential to contribute to future efforts will be summarized.

2.2 Theory

We are interested in calculating the standard change in free energy upon non-covalent molecular association. Consider the following reaction:

$$A + B \Leftrightarrow AB \tag{2.1}$$

where A represents the protein, B the ligand, and AB the protein-ligand complex. Each molecule can be described by a sum of translational, rotational and internal modes of freedom. Upon binding, the ligand's external translational and rotational motions become internal motions of the complex. According to classical statistical mechanics, after the kinetic contributions of each species have cancelled [2], the standard change in free energy can be expressed as a ratio of configuration integrals:

$$\Delta G^{\circ}_{AB} = -RT \ln\left(\frac{C^{\circ}}{8\pi^2}\right)\left(\frac{Z_{N,AB}Z_{N,O}}{Z_{N,A}Z_{N,B}}\right) + P^{\circ}\left\langle \Delta V_{AB} \right\rangle \tag{2.2}$$

where R is the gas constant, T is the absolute temperature, C° is the standard state concentration (typically 1 M or 1 molecule/1660 Å^3), N is the number of solvent molecules, and $P^{\circ}\left\langle \Delta V_{AB} \right\rangle$ is the pressure-volume work associated with changing the system size from the replacement of two free molecules by one bound species. The last term is generally considered to be negligibly small

in water at 1 atm. It is important to note that all mass dependent terms have cancelled in equation 2.2. This is a direct result of the equal kinetic contribution to the partition function of the bound and the free species. The configuration integral of the protein, *A*, in solution is:

$$Z_{N,A} = \int \mathrm{e}^{-\beta U(r_A, r_S)} dr_A dr_S \qquad 2.3$$

where $U(r_A, r_S)$ is the potential energy as a function of all solute coordinates, r_A, and solvent coordinates, r_S, and β is the reciprocal of the product of the Boltzmann constant and temperature. A similar equation gives $Z_{N,B}$ for the ligand. The configuration integral of the solvent alone is:

$$Z_{N,O} = \int \mathrm{e}^{-\beta U(r_S)} dr_S \qquad 2.4$$

The ratio of configuration integrals in equation 2.2 can be simplified with an implicit solvent approximation:

$$\frac{Z_{N,A}}{Z_{N,O}} = Z_A = \int \mathrm{e}^{-\beta[U(r_A)+W(r_A)]}\, dr_A \qquad 2.5$$

where,

$$W(r_A) \equiv -RT \ln\left(\frac{\int \mathrm{e}^{-\beta \Delta U(r_A, r_S)}\, \mathrm{e}^{-\beta U(r_S)}\, dr_S}{\int \mathrm{e}^{-\beta U(r_S)}\, dr_S} \right) \qquad 2.6$$

represents the solvation free energy of species *A*, and the quantity $\Delta U(r_A, r_S)$ is $U(r_A, r_S) - U(r_A) - U(r_S)$. Analogous equations hold for the complex and ligand. The complex, however, contains six degrees of freedom that represent the residual translational and rotational motions of the bound ligand. In order to account for these modes of motion, it is helpful to introduce a set of coordinates, $\delta_B \equiv (x_1, x_2, x_3, \xi_1, \xi_2, \xi_3,)$, which define the bound ligand's position and orientation with respect to the protein. The complete complex configuration integral is:

$$Z_{AB} = \int \mathrm{e}^{-\beta[U(r_A, r_{B'}, \delta_B) + W(r_A, r_{B'}, \delta_B)]} dr_A dr_{B'} d\delta_B \qquad 2.7$$

where $r_{B'}$ represents the bound ligand's remaining internal coordinates and δ_B spans conformations where *A* and *B* form a complex. As will be seen below, the displacements of δ_B in the dynamics of the complex are very small. It is therefore reasonable to assume that the higher order coupling terms in the potential energy function due to the effect of the ligand's translational/rotational motions on either species' internal vibrational motions are very small. Thus, the potential and solvation energies in equation 2.7 are separable:

$$U(r_A, r_{B'}, \delta_B) + W(r_A, r_{B'}, \delta_B) \cong U_1(\delta_B) + W_1(\delta_B) + U_2(r_A, r_{B'}) + W_2(r_A, r_{B'}) \qquad 2.8$$

If one defines a potential of mean force [43] for a particular ligand position and orientation, δ_B :

$$\begin{aligned} \omega(\delta_B) &\equiv -RT \ln\left(\int \mathrm{e}^{-\beta(U(r_A, r_{B'}, \delta_B) + W(r_A, r_{B'}, \delta_B))} dr_A dr_{B'} \right) \\ &= U_1(\delta_B) + W_1(\delta_B) - RT \ln \int \mathrm{e}^{-\beta[U_2(r_A, r_{B'}) + W_2(r_A, r_{B'})} dr_A dr_{B'} \end{aligned} \qquad 2.9$$

Equation 2.9 shows that the ligand's potential and solvation energies are equal to within a constant of the potential of mean force.

A similar assumption about the correlation between translational and rotational motions, permits further decomposition of $U(\delta_B) \cong U(x_1, x_2, x_3) + U(\xi_1, \xi_2, \xi_3)$ and $W(\delta_B) \cong W(x_1, x_2, x_3) + W(\xi_1, \xi_2, \xi_3)$. These separate contributions can be directly measured from a MD simulation as described in the methods section. Substituting equations 2.5 and 2.9 into 2.2 we have:

$$\Delta G^{\circ}_{AB} = -RT \ln\left[\frac{C^{\circ} z^{trans}_{B'} z^{rot}_{B'}}{8\pi^2} \frac{Z_{AB'}}{Z_A Z_B} \right] \qquad 2.10$$

where $z^{trans}_{B'} = \int \mathrm{e}^{-\beta[U(x_1, x_2, x_3) + W(x_1, x_2, x_3)]} dx_1, dx_2, dx_3$, $z^{rot}_{B'} = \int \mathrm{e}^{-\beta[U(\xi_1, \xi_2, \xi_3) + W(\xi_1, \xi_2, \xi_3)]} d\xi_1, d\xi_2, d\xi_3$ and $Z_{AB'} = \int \mathrm{e}^{-\beta U(r_A r_{B'})} dr_A dr_{B'}$. Equation 2.10 holds the most challenging aspect of this work, the evaluation of many-dimensional configuration integrals. As a first order approximation, one can

assume that the energetic landscape of each species has an energy and a volume that can be determined from a sufficiently long MD simulation:

$$Z_A = \int \mathrm{e}^{-\beta[U(r_A)+W(r_A)]} dr_A \approx z_A^{int} \mathrm{e}^{-\beta\langle E_A \rangle} \qquad 2.11$$

$\langle E_A \rangle \equiv \langle U(r_A) + W(r_A) \rangle$ represents the average molecular mechanics plus solvation energy over the simulation and z_A^{int} is the internal configuration integral. Equivalent equations hold for the ligand, Z_B, and the complex $Z_{AB'}$. If one assumes that the volumes of configuration space occupied by the ligand and protein change negligibly upon association, that is, $z_A^{\mathrm{int}} z_B^{\mathrm{int}} = z_{AB'}^{\mathrm{int}}$, then all internal configuration integrals cancel in the ratio, leaving:

$$\Delta G_{AB}^{\circ} = -RT \ln\left(\frac{C^{\circ} z_{B'}^{trans} z_{B'}^{rot}}{8\pi^2} \right) + \left(\langle E_{AB'} \rangle - \langle E_A \rangle - \langle E_B \rangle \right) \qquad 2.12$$

Alternatively, the volume of configuration space occupied by each species can be approximated and the changes in conformational entropy can be included, as described in the discussion section.

2.3 Methods

2.3.1 Molecular Dynamics Protocol

The coordinates of the ligand, the protein and the complex were taken from the 1.85 Å resolution complex crystal structure in the Brookhaven Protein Data Bank (PDB code 1D7J) [42]. The free ligand was optimized with Gaussian 98 [44] at the Hartree-Fock level with the 6-31G* basis set. It was assigned RESP charges as implemented in the Antechamber module from AMBER 7.0 [45]. The complex was prepared in three steps. First, the program GRID [46] was used to add 142 buried and first shell water molecules to 126 crystal waters already present. WHAT IF [47] was then used to place hydrogens and to assign favorable protonation states of

histidine residues, as well as the favorable orientations of glutamine and asparagine sidechains. Finally, the system was placed in a 80.2 Å x 78.9 Å x 78.9 Å TIP3 water box with the Leap module from AMBER 7.0 [48]. One of the bulk water molecules was replaced with a chloride ion to neutralize the system.

Simulations of the complex, protein, and ligand were run under constant N,P,T conditions with AMBER 7.0. Periodic boundary conditions, particle-mesh Ewald treatment of the electrostatics, and SHAKE-enabled 2 fs time-steps were employed. The protein and ligand heavy atoms were restrained during a 500 step minimization. Restraints were maintained through a 40 ps gradual warming from 0 to 300 K under constant volume and temperature conditions (N,V,T). Ten picoseconds of constant pressure and temperature (N,P,T) allowed the system to reach the proper density. A minor modification of the Sander module, allowed a linear release of the heavy atom restraints over 30 ps. Unrestrained N,P,T completed the equilibration phase, and three nanoseconds of production phase was collected.

2.3.2 Energetic Analysis

The binding affinity was approximated from both a single simulation, in which the protein and ligand structures were taken from the complex simulation, and from separate simulations. Snapshots taken every 2 ps from the 3 ns of production phase simulation were evaluated for a total of 1500 structures. The molecular mechanics energy, U_{MM}, was evaluated in a single MD step in the Sander module using an infinite cutoff for non-bonded interactions. The solvation free energy can be decomposed into electrostatic and non-electrostatic components, $W_{PBSA} = W_{PB}^{eel} + W_{SA}^{np}$. The electrostatic contribution to the solvation free energy, W_{PB}^{eel}, was calculated with the Adaptive Poisson Boltzmann Solver (APBS)[49]. The interior of the protein was given a dielectric constant of one, in agreement with simulation conditions. The reference system had a solvent dielectric of one and 0 M salt concentration. The solvated system

had a solvent dielectric of 78.4 and 100 mM salt concentration. The electrostatic energy of the reference system was subtracted from that of the solvated system to yield the solvation energy. Harmonic smoothing was used to define the protein boundary. Finally, the nonpolar contribution to the solvation free energy, W_{SA}^{np}, was approximated with the commonly used solvent accessible surface area (SASA) model, $\Delta W_{SA}^{np} = \gamma(\mathrm{SASA}) + \beta$, where γ = 0.00542 kcal/mol Å^2 and β = 0.92 kcal/mol [50]. The SASA was estimated with a 1.4 Å solvent-probe radius as implemented in Sander.

2.3.3 Ligand Translational Freedom

The bound ligand's translational configuration integral, $z_{B'}^{trans}$, can be conceptually linked to the volume of space that its center of mass occupies through the simulation. As previously mentioned, $\delta_{B'}$ in equation 2.7 spans conformations where A and B form a complex. Thus, this analysis is only valid for simulations where the ligand remains bound to the protein. The effective volume was measured with the quasi-harmonic model, which relies on the assumption that the translational motion can be described by a multivariate Gaussian probability distribution. Superimposition of every snapshot according to protein C-alpha atoms defined a static protein reference system and an average ligand structure. Centered at the origin, the ligand's center of mass covariance matrix was then evaluated, accounting for the possible coupling of motions along different axes. The resulting eigenvalues, λ_i, describe the variance Δx_i^2 along each principal axis by $\lambda_i = \Delta x_i^2$. The equipartition theorem allows one to relate the variance to the force constant of the classical harmonic oscillator as the average potential energy for one dimension is $\langle U(x) + W(x) \rangle = (1/2)\kappa \langle \Delta x^2 \rangle \cong (1/2) k_B T$, such that $\kappa \cong k_B T / \langle \Delta x^2 \rangle$. Thus, $z_{B'}^{trans}$ can be calculated as:

$$z_{B'}^{trans} = \int e^{\left(-\kappa_1 \Delta x_1^2 / 2k_B T\right)} dx_1 \int e^{\left(-\kappa_2 \Delta x_2^2 / 2k_B T\right)} dx_2 \int e^{\left(-\kappa_3 \Delta x_3^2 / 2k_B T\right)} dx_3$$
$$= (2\pi)^{3/2} \left(\left\langle \Delta x_1^2 \right\rangle \left\langle \Delta x_2^2 \right\rangle \left\langle \Delta x_3^2 \right\rangle \right)^{1/2} \quad 2.13$$

2.3.4 Ligand Rotational Freedom

The ligand's rotational freedom, $z_{B'}^{rot}$, was accounted for in a similar manner. Quaternions, an elegant alternative to Euler angles, were used to represent the ligand's rotational motion. Quaternions are hypercomplex numbers that can be represented as a linear combination of a scalar (a_1) and a vector ($\vec{n} \equiv [a_2, a_3, a_4]$) :

$$q = a_1 + a_2 i + a_3 j + a_4 k = (a_1, \vec{n}) \quad 2.14$$

The rotation of point p through angle ϕ about a normalized axis $\vec{n} \equiv [\vec{n}_x, \vec{n}_y, \vec{n}_z]$ can be computed with the quaternion q and its complex conjugate q*.

$$q = \cos\left(\frac{\phi}{2}\right) + \sin\left(\frac{\phi}{2}\right)\vec{n}_x i + \sin\left(\frac{\phi}{2}\right)\vec{n}_y j + \sin\left(\frac{\phi}{2}\right)\vec{n}_z k$$
$$p' = qpq^* \quad 2.15$$

The rotational transformation of any point about three axes is the product of three quaternion. Thus, the transformation of each ligand snapshot, within the protein reference binding pocket, was described by the product of three quaternion, each defining the rotation about one axis. When the angles of rotation are small the cross terms of this product will be negligibly small.

$$q = q_1 q_2 q_3 \cong 1 + \sin\left(\frac{\phi_1}{2}\right) i + \sin\left(\frac{\phi_2}{2}\right) j + \sin\left(\frac{\phi_3}{2}\right) k \quad 2.16$$

This small angle approximation reduces the product of three quaternion to a single quaternion which is sinusoid ally related to three angles of rotation. The covariance matrix was evaluated to account for coupling between axes. The resulting eigenvalues were related to a spring force constant assuming a Gaussian distribution and $z_{B'}^{rot}$ was evaluated according to equation 2.13,

replacing $(\Delta x_1, \Delta x_2, \Delta x_3)$ with $(\Delta\phi_1, \Delta\phi_2, \Delta\phi_3)$. Although the present analysis assumes that the bound ligand's translational and rotational motions are dominated by a single minimum energy well, it is easily extendible to multiple minima.

2.4 Results and Discussion

2.4.1 Energetic Convergence

Each simulation reached a satisfactory equilibrium after 100 ps as indicated by the total system energy. The protein's energetic contributions as a function of time in the simulation of the complex are shown in Figure 2.1. Similar plots were obtained for the complex and ligand from the complex simulation as well as the protein and ligand from the separate simulations. The variation in the solvation energy and the molecular mechanics energy (Figure 2.1a) are anti-correlated, yielding a fairly stable total energy (Figure 2.1b). This is further supported by the average energetic contributions and standard deviations of the complex simulation evaluations, shown in Table 2.1. The standard deviations of the molecular mechanics and solvation energies are consistently four to five percent while that of the total energy is less than one percent.

Table 2.1: Average Complex, Protein, and Ligand Energies (kJ/mol)*

Simulation	$\langle U_{MM} \rangle$*	$\langle W_{PB} \rangle^{\dagger}$	$\langle W_{SA} \rangle^{\ddagger}$	$\langle G_{MM/PBSA} \rangle^{\S}$
Complex	−5759 (281)	−5123 (247)	134.0 (2.9)	−10,748 (130)
Protein	−5605 (273)	−5134 (241)	135.1 (2.8)	−10,604 (130)
Ligand	−50 (11)	−46 (3)	8.5 (0.1)	−88 (10)
Com-Pro-Lig	−103 (26)	57 (15)	−9.6 (0.5)	−56 (15)

* average of 1500 snapshots from 100 – 3100 ps, standard deviations in parentheses.

[#] molecular mechanical energy,

[§] electrostatic solvation energy, [¶] non-polar solvation energy,

[‖] MM/PBSA energy: $\langle G_{MM/PBSA} \rangle = \langle U_{MM} \rangle + \langle W_{PB} \rangle + \langle W_{SA} \rangle$

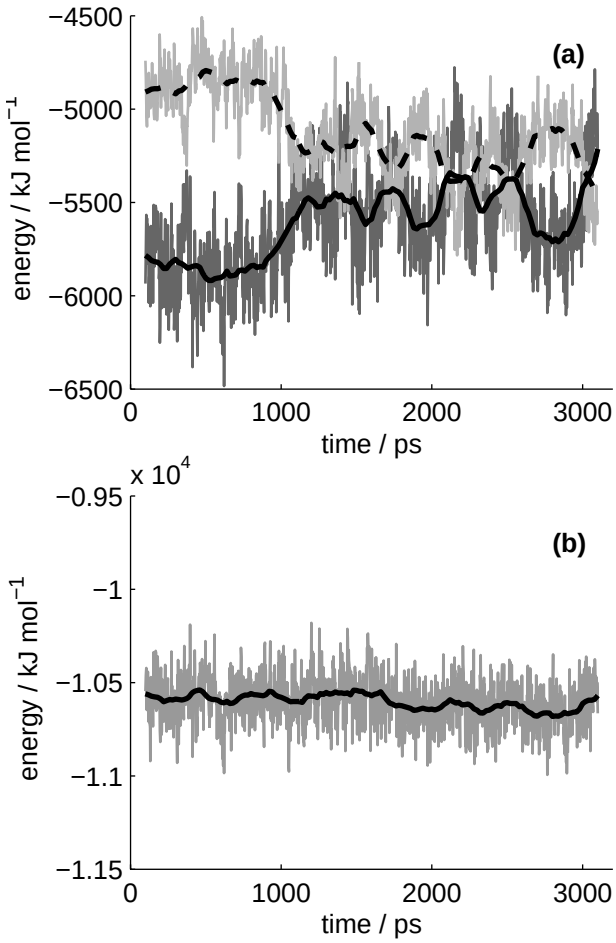

Figure 2.1: Running Protein Energies. (a) The protein's solvation energy, light grey, and molecular mechanics energy, dark grey, across 3 ns of simulation. The darker solid and dashed lines represent a 100 ps running averages. (b) The protein's total energy , $E_{TOT} = U_{MM} + W_{PBSA}$, and running average.

2.4.2 Relaxation Energies and Protein Flexibility

The ideal mimicry of an *in vitro* binding event would be to run three separate simulations and calculate the energetic components of each. This would include the effects of the conformational changes upon binding, e.g. protein flexibility. The relaxation energy would be captured in the molecular mechanics and solvation energy, and the conformational free energy would be captured by a complete evaluation of the configuration integral. This evaluation relies on sufficient sampling of configuration space, which remains a major challenge on the timescale of MD simulations.

In an approximate single simulation evaluation, the protein and ligand structures are taken from the complex simulation. This, in theory, assumes that the structures and conformational freedom of the protein and ligand change negligibly upon binding. In practice, taking all structures from a single simulation cancels the noise that would result from sampling inconsistencies and the error inherent in force-field and implicit solvation energies. Although the analysis based on simulations of separate species (results not shown) generated similar trends to the single simulation analysis, it was clearly dependent on simulation length and dominated by noise. A striking representation of this phenomenon is shown in Figure 2.2, where the differences in energetic contributions are given as a function of time for both the single and the separate simulations. It should be noted that the corresponding structures from the free and bound simulations can not be equated for any given timeframe. Thus plot A is a non-physical measurement. Given the commutative nature of averages, however, the total energies, shown as the smoothed dark line, are the quantitative results of the molecular mechanic and solvation free energies. The same axis scales are used to emphasize the noise of the separate simulations compared to the single simulation.

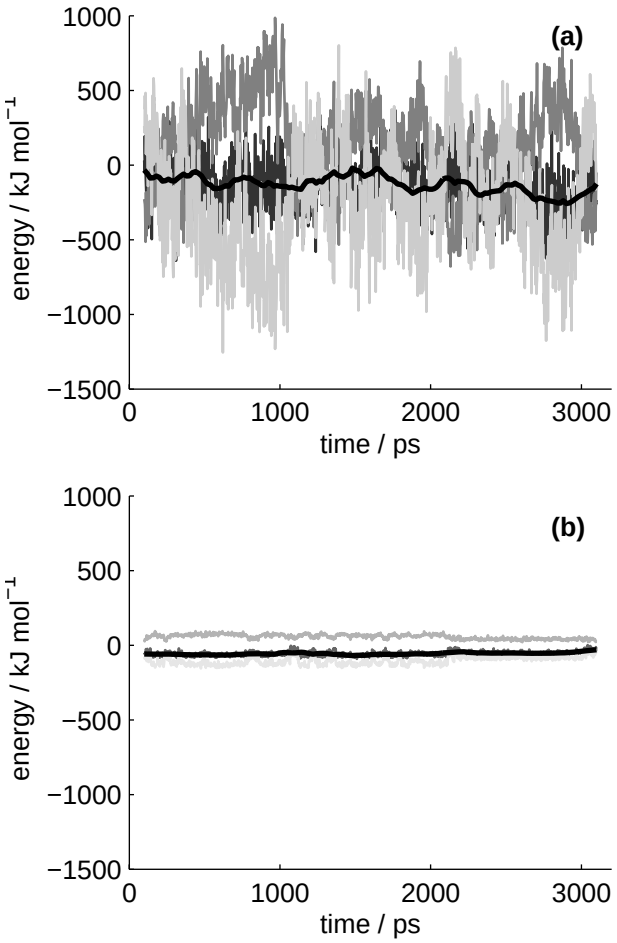

Figure 2.2: Running MM Binding Energy. The change in free energy ($E_{AB} - E_{A} - E_{B}$) of each snapshot for (a) the separate and (b) the single simulation evaluations. The lightest shade is the molecular mechanics energy, the middle is the solvation energy, the darkest is the total, and the smooth black line is a 100 frame running average of the total energy.

While it was clear that the protein sampling was insufficient, the small ligand sampling was extensive. It was possible, therefore, to capture the ligand's relaxation energy, ΔE_{LR}, which is the difference between the total energy of the ligand from the complex simulation and that from the free simulation. The final calculated binding free energy and its components, including the ligand relaxation energy of 1.7 kJ/mol, are shown in Table 2.2.

Table 2.2: Contributions to the free energy (kJ/mol)

$\Delta G_{MM/PBSA}$*	$\Delta G_{association}$†	ΔE_{LR}‡	ΔG_{CALC}§	ΔG_{EXP}¶
−56	23.3	1.7	−31	−18.9

* MM/PBSA energy, # association entropy, § ligand relaxation energy,
¶ total calculated free energy, ‖ experimental free energy

2.4.3 Association Free Energy

At 1 M standard concentration, a free molecule has 1660 Å^3 (C° = 1/1660 Å^3 in equation 2.12) of translational freedom and $8\pi^2$ of rotational freedom. Upon association, one solute molecule loses translational and rotational freedom while released solvent molecules gain translational and rotational of freedom. As previously described, the solvent's enthalpic and entropic contributions are accounted for in the implicit approximation of the solvation free energy. The solute's contribution, which is presently described as the association free energy, was directly measured from the simulation (see Methods). To provide some context for this evaluation, a brief, and therefore incomplete, historical account of comparable theoretical studies on the association free energy is helpful.

The free energy change, and particularly the entropic cost, due to one molecule's loss of translational and rotational freedom has been well recognized for over 40 years [51]. These degrees of freedom do not disappear but are transformed into internal motions within the complex. The range of these motions determines the magnitude of the entropic cost. More tightly bound ligands will have a higher entropic cost than loosely bound ligands. Quantifying the ligand's residual translational and rotational motions, however, is not an easy task. Many authors have estimated them with cubic box translational and isotropic rotational approximations, such that $T\Delta S_{trans} = RT \ln(\Delta x^3/1660\mathring{A}^3)$ and $T\Delta S_{rot} = RT \ln(\Delta\theta^3/8\pi^2)$.

Finkelstein and Janin [52] assumed that the atomic motions in crystals were representative of any bound ligand's motion. Using Debye-Waller temperature factors, they

estimated a standard deviation of 0.25 Å along three principal axes, resulting in a translational entropic cost of -15 kcal/mol. Since the magnitudes of rotational oscillations in crystals were unknown at the time, they assumed a similar angular displacement from $\delta\theta = 2\delta x / d$, where d is the distance to the ligand interface. This resulted in a rotational entropic cost of -7.2 kcal/mol and a total association entropy of -22.2 kcal/mol.

Tidor and Karplus [53] took a different approach. Using normal mode analysis (NMA) to study insulin dimerization, they found the internal vibrational modes of the complex increased, contributing -7.2 kcal/mol to the binding free energy. Although the six introduced modes of motion are included in this estimate, it is impossible to separate them to account for the range of the bound ligand's motion or the exact association entropy. Assuming no change in internal vibrational modes and estimating the free energy change due to complete loss of rotational and translational motion from gas phase ($T\Delta S$ = -27.3 kcal/mol), they reported an association entropy around -20 kcal/mol.

Hermans and Wang [8] presented the first complete evaluation of an absolute binding free energy with FEP. In this study they evaluated the effective volume of the bound ligand in two independent ways. First, they applied translational restraints to the ligand in the standard state gas phase. Releasing the restraints in the protein environment and taking the difference in free energies for the two processes, they measured the association entropy (-7 kcal/mol). Second, they estimated the ligand's positional and orientational root mean square displacement (RMSD) directly from the simulation. It should be emphasized that these two methods of obtaining the effective volume, using RMSD's versus the energetically measured volume, are very different. The point, in this case, is a methodological one as the two are similarly small. The calculated RMSD volume, 0.184 $Å^3$, and the energetically measured volume, 0.4 $Å^3$, result in -5.0 kcal/mol and -5.4 kcal/mol entropic contributions respectively.

Lazaridis et al. [40] evaluated the ranges of deviation in the ligand's center of mass and orientation, described with Euler angles, from a dynamics simulation. They weighted these

ranges according to their probability distributions. It is not clear whether they evaluated these deviations along the principal axes or along an arbitrary reference frame. Our results indicated that similar range assumptions resulted in significantly larger translational and rotational motions that were sensitive to simulation length. This could explain the smaller translational and rotational entropic contributions measured in this study.

Luo and Sharp [39] used quasi-harmonic analysis of short simulations to account for the ligand's translational, rotational, as well as internal vibrational motions. They assumed that the rotational motion was isotropic and divided by a factor of $3^{3/2}$ to yield; $T\Delta S_{rot} = RT \ln(\sigma^3/(6\pi)^{1/2})$. They measured association entropies between -1.5 kcal/mol and -7.5 kcal/mol for four different ligands.

As described in the methods section, this chapter presents a similar evaluation of the association free energy using the quasi-harmonic model. The covariance matrix accounts for coupled motions in different dimensions and defines the principal components, capturing a more accurate variation than an arbitrary reference frame. Quaternions were found to be a desirable alternative description of angular motions, eliminating the cumbersome conversion to Euler angles. They smoothly converted into a covariance matrix and produced three different eigenvalues. This finding discourages the assumption that rotational motion is isotropic. As summarized in Table 2.3, the ligand experienced 1.72 $Å^3$ of translational motion and 6.57 radians of rotational motion. This correlates to a free energy change of 17.1 kJ/mol and 6.2 kJ/mol, respectively. Thus, the total association free energy was 23.3 kJ/mol. If one assumes that the translational and orientational motions of the ligand within the complex can in fact be described as classical harmonic oscillator displacements, then this total free energy of association can be separated into enthalpic and entropic components. The six configurational degrees of freedom would contribute an equipartition enthalpy of 3RT ≈ 7.5 kJ/mol. The remainder, about 15.8 kJ/mol, then represents the entropic cost of limiting the ranges of translational and rotational motion.

Table 2.3: Contributions to the association free energy.

z_{trans}*	ΔG_{trans}*	z_{rot}‡	ΔG_{rot}†	$\Delta G_{association}$†§
1.72	17.1	6.57	6.2	23.3

*All energies reported in kJ/mol,
Translational configuration integral (Å3),
§ Rotational configuration integral, ¶ association free energy.

2.4.4 Conformational Free Energy

Detailed evaluations of the configuration integrals in equation 2.10 would inherently capture the exact changes in conformational free energy upon binding. This remains, to date, computationally infeasible. Changes in intramolecular conformational free energy have traditionally been approximated with quasi-harmonic analysis, normal mode analysis, or side-chain rotational analysis. Yet the validity and accuracy of these methods remain questionable. In the current study, quasi-harmonic analysis was extensively explored (data not shown). Although the results followed the expected trends, making the calculated free energy of binding less favorable, they were clearly sensitive to simulation length. Similar to the separate simulation analysis, this lack of convergence indicates inadequate sampling. This is likely compounded by a large noise to signal ratio due to the weak binding nature of this ligand. A system with stronger interactions may prove more amenable to analysis. Given the challenges of a weak binding system and the excluded protein relaxation energy and configurational free energy, both of which are expected to be slightly positive, it is encouraging to find the calculated binding free energy only 10 kJ/mol lower than that measured in experiment.

2.5 Conclusions

Although the theory of binding affinity calculations has been discussed by many previous authors, it remains an ongoing topic of research. The implementation of end-point free energy models has improved with increasing computational resources and thoughtful design. A

connection between theory and implementation was the focus of this paper. This chapter has discussed the statistical mechanical basis for the change in free energy upon binding and its link to obtaining this quantity from a molecular dynamics trajectory. The importance of the standard state dependence has been emphasized. A novel method for evaluating a bound ligand's residual translational and rotational motion from an MD simulation has been presented and these quantities have been used to calculate the association free energy. Finally, an attempt has been made to establish the proper groundwork for end point free energy calculations such that future efforts can focus on the inclusion of protein relaxation energies and changes in conformational free energy.

This chapter is a reprint in full of material that appeared in. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy Jessica M.J. Swanson, Richard H. Henchman, and J. Andrew McCammon. Biophysical Journal, **86** 67-74, January 2004. I was the primary researcher and author of this work.

Chapter Three

The Entropic Cost of Protein-Protein Association: A Case Study on Acetylcholinesterase Binding to Fasciculin-2

ABSTRACT

Protein-protein association is accompanied by a large reduction in translational and rotational (external) entropy. In this chapter a 15 ns MD simulation of acetylcholinesterase (AChE) in complex with fasciculin 2 (Fas2) is used to estimate the loss in external entropy using quasiharmonic analysis and histogram-based approximations of the probability distribution function. The external entropy loss of AChE-Fas2 binding, about 30 cal/mol K, is found to be significantly larger than most previously characterized protein-ligand systems. However, it is less than the entropy loss estimated by Finkelstein and Janin (1989), which was based on atomic motions in crystals.

3.1 Introduction

Entropic changes upon noncovalent binding are difficult to estimate because extensive phase space sampling is required [54]. For example, to calculate the external (translational and rotational) entropy, the range of residual motion in the complex must be known. Finkelstein and Janin [55] were the first to estimate residual motion; using the average fluctuation of crystal atoms they estimated an entropy loss of 50 cal/mol K. Many studies since have calculated the external entropy of protein-ligand systems [39, 56, 57], but none to date have targeted protein-

protein association. In this chapter, the external entropy loss is estimated using MD sampling and integrating over different forms of the probability distribution function (PDF).

3.2 Methods

This analysis is applied to a 15 ns MD trajectory of mouse acetylcholinesterase (AChE) in complex with fasciculin-2 (Fas2) [58, 59]. AChE, which catalyzes the hydrolysis of the neurotransmitter acetylcholine in cholinergic synapses, is inhibited by the snake toxin Fas2, a high affinity binder with a slow dissociation rate. The initial external entropy for AChE-Fas2 is given by the Sackur-Tetrode equation and a rotational entropy term.[3] For the final state, the entropy of restricted modes is estimated using the Gibbs entropy for translational and rotational modes. Assuming no change in momenta and moments of inertia upon binding, the entropic change is,

$$\Delta S_{ext} = R \ln\left(\frac{C^\circ}{8\pi^2}\right) - R\int p_{ext}(q) \ln p_{ext}(q) dq \quad , \qquad 3.1$$

where C° is standard concentration (1 molecule/1660 Å^3) and $p_{ext}(q)$ is the external mode PDF.

Molecular dynamics simulations of the AChE-Fas2 complex have been previously described [58, 59]. The proteins are assumed to be bound over the course of the simulation. After 1.1 ns of equilibration, snapshots from every 10 ps of a 14.9 ns trajectory were superimposed according to the C-α atoms of an alignment molecule, either AChE or Fas2, allowing a reference structure to be defined by their average. Translational coordinates were defined by the center of mass of the other (hereafter referred to as the test) molecule. Quaternions were used for least squares alignment of the test molecule with respect to the reference structure, and were converted to Euler angles through the small angle approximation

[57]. Principal components analysis was performed by diagonalizing the zero-mean covariance matrices for the rotational and translational motions. The calculations were performed using either two 3x3 or a single 6x6 covariance matrix. The eigenvalues λ_i are equivalent to the variances σ_i^2 in each principal axis.

In quasiharmonic (QH) analysis [60], the molecule is assumed to have a Gaussian distribution, leading to a Gibbs entropy of $-R\ln\sqrt{2\pi e\sigma_i^2}$ for each dimension. The PDF was also estimated by direct histogram analysis of the simulation. After projecting the coordinates into principal component space, histograms were constructed with 70 bins in each dimension, allowing $p_{ext}(q)$ to be calculated by normalizing the histogram and numerically integrating the configurational integral using the composite Simpson's rule. In Gaussian fit methods, a single Gaussian (sGF) or the sum of two Gaussians (dGF) was fitted to the histogram using non-linear least-squares regression. The configuration integrals were numerically evaluated by adaptive Lobatto quadrature over three standard deviations from the extrema. To validate the integration procedure, it was tested on the sGF and found to match the analytical result.

3.3 Results

In the AChE-Fas2 simulation, most of the external coordinate histograms are singly peaked, except for the doubly peaked most important components (Figure 3.1). Thus, the dGF PDF more closely follows the simulation histogram than other PDFs. The histogram shape is due to sampling of several local minima over time. The similarity between dGF and direct histogram entropies is advantageous in the limit of extensive ensemble sampling. However, in this limit, the value of the harmonic oscillator assumption is debatable. A multiple Gaussian fit may be most applicable in cases where multiple nearby energy wells are sparsely sampled.

When there are multiple energy wells, the accuracy of QH analysis is questionable [61]. If the wells are distant from one another, assuming a normal distribution will overestimate the entropy. To illustrate this principle, the entropy of a double-well distribution was compared to the QH value (Figure 3.2). As separation is increased, the double well entropy stabilizes while the QH entropy continues to increase.

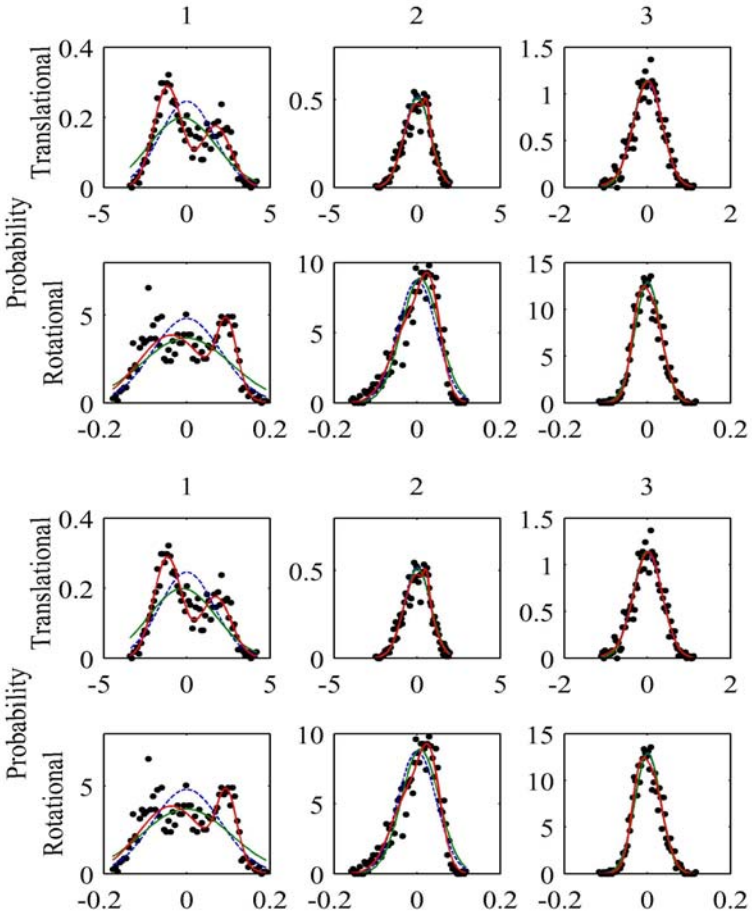

Figure 3.1: Probability Distributions of Translational and Rotational Degrees of Freedom with AChE (top) and Fas2 (bottom) as alignment molecules. Black dots signify the histogram-derived probabilities; the dashed blue line is the QH distribution; the green line is sGF; the red line is dGF. Translation units are Å and rotational units are radians.

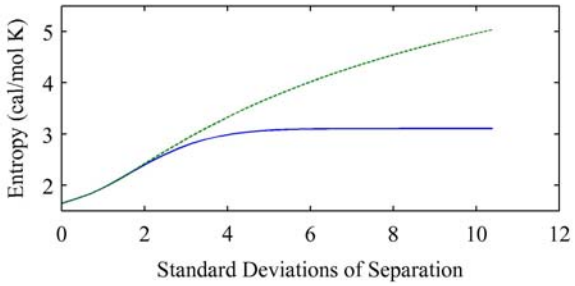

Figure 3.2: QH (green) and Gibbs (blue) entropies as a function of well separation. The double-well distribution was generated using the sum of two equally weighted Gaussians and varying the distance between peaks.

Calculated entropies are qualitatively similar for every technique (Table 3.1). Entropy losses are slightly lower, by less than 1 cal/mol K, when a single Gaussian is used, in agreement Figure 3.2 calculations. If Fas2 is used as the alignment molecule, the rotational entropy of the complex is estimated to be about 2 cal/mol K higher. This may be due to difficulty of accurately aligning to a smaller reference molecule, leading to artefactual phase space excursions. One way to remove this artefact is to treat the six external modes all together.

Table 3.1: External Entropy Changes (cal/mol K) with AChE as alignment molecules. Decoupled entropies denote values obtained from a 6x6 covariance matrix. Error values for AChE meassurements, enclosed in parentheses, are the standard deviation of 500 entropies calculations from 500 randomly selected snapshots.

Method	Translation	Rotation	External	Decoupled
Hist	−18.7(0.02)	−10.9(0.02)	−29.5(0.03)	−31.7(0.19)
QH	−18.2(0.01)	−10.6(0.10)	−28.8(0.15)	−31.1(0.16)
sGF	−17.7(0.03)	−10.4(0.02)	−28.1(0.03)	−31.2(0.26)
dGF	−18.3(0.48)	−11.0(0.11)	−29.3(0.50)	−31.9(0.55)

The degree of coupling between translational and rotational modes is given by $S_{coupling} = S_{combined} - (S_{trans} + S_{rot})$. Entropic analysis was performed for a 6x6 covariance matrix encompassing all external modes. Because generating this matrix combines angular and

Cartesian coordinates, the eigenvectors are extremely sensitive the units used in the equation. Therefore, dimensionless units of length $x^* = x/(C°)^{1/3}$ and angle $l^* = l/(8\pi)^{1/3}$ were used to generate a the matrix and histograms (Fig 5). In the QH method, when either AChE or Fas2 were used as the alignment molecule, the entropy change estimate is -31.1 cal/mol K. The similarity of external entropy values evinces the successful decoupling of translational and rotational entropy terms. The coupling terms for AChE and Fas2 alignment molecules are 2.4 and 5.0 cal/mol K, respectively. Retrospectively, it is clear that considering the translational and rotational degrees of freedom separately led to artefactual motions in rotational phase space.

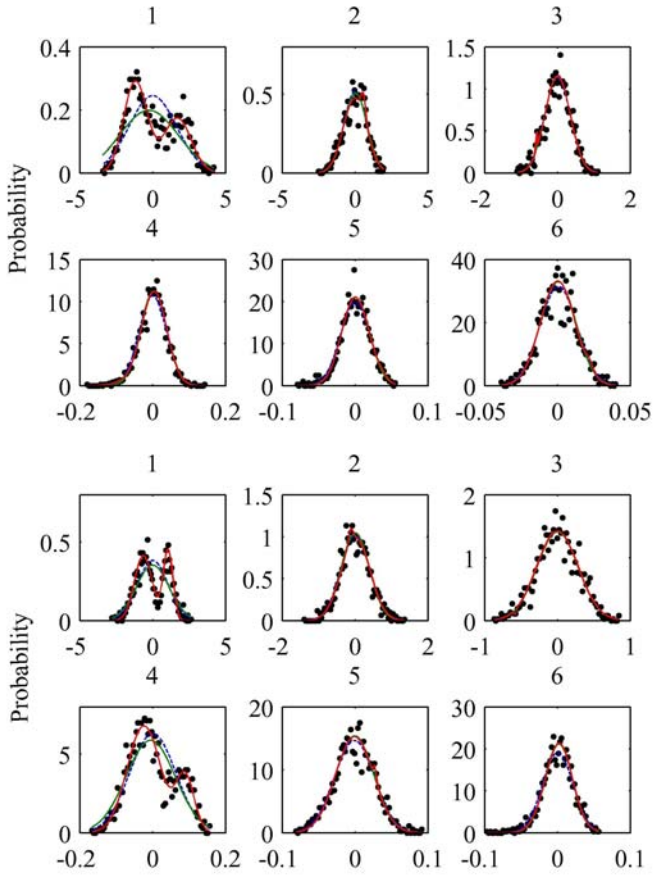

Figure 3.3: Probability Distributions of External Degrees of Freedom, Fas2 as alignment molecule. Black dots signify the histogram-derived probabilities; the dashed blue line is the QH distribution; the green line is sGF; the red line is dGF.

3.4 Conclusions

Molecular dynamics simulations provide evidence for enhanced conformational fluctuations of AChE-Fas2 in complex in comparison to apo form [59]. These fluctuations may lead to increased internal entropy to compensate for the external entropy loss upon binding. The external entropy loss upon protein-protein complexation is of greater magnitude than for protein-water binding (0 to 7 cal/mol K) [62] and protein-small molecule systems [39, 56, 57]. For example, in chapter 2 the association of FKBP and 4-hydroxy-2-butanone leads to

translational and rotational entropy changes of 10.6 and 2.0 cal/mol K, respectively. With protein-protein complexation, the larger interaction surface leads to a more substantial reduction of the external entropy, particularly for the rotational degrees of freedom. However, the entropy loss estimated is less than expected from analysis of atomic motions in crystals [55], even for the tight binding AChE-Fas2 complex. This is a reasonable result reflecting the enhanced fluctuations of proteins in solution.

This chapter is a reprint in full of material that appeared in *The Entropic Cost of Protein-Protein Association: A Case Study on Acetylcholinesterase Binding to Fasciculin-2*, David D.L. Minh, J.M. Bui, C.E. Chang, T. Jain, J.M.J. Swanson, J.A. McCammon, Biophys. J. 89, L25-L27 (2005). I was a secondary researcher and author of this work.

Chapter Four

Implicit Solvation Free Energies: Optimized Radii for Poisson-Boltzmann Calculations with the AMBER Force Field

ABSTRACT

Implicit solvent models are a standard tool for assessing the electrostatics of biomolecular systems. The accuracy of quantitative predictions, such as pKa values, transfer free energies, binding energies, and solvation forces, is strongly dependent on one's choice of continuum parameters: the solute charges, dielectric coefficient, and radii, which define the dielectric boundary. To ensure quantitative accuracy, these parameters can be benchmarked against explicit solvent simulations. In this chaptert two sets of optimized radii are presented to define either abrupt or cubic-spline smoothed dielectric boundaries in Poisson-Boltzmann calculations of protein systems with AMBER (parm99) charges. Spline smoothing stabilizes the electrostatic potential at the molecular surface, allowing for continuum force calculations. Most implementations, however, require significantly different radii than the abrupt boundary surfaces. The optimal continuum radii are initially approximated from the solvent radial charge distribution surrounding each atom type. A genetic algorithm is then used to fine-tune the starting values to reproduce charging free energies measured from explicit solvent simulations. The optimized radii are tested on four protein-like polypeptides. The results show increased accuracy of molecular solvation energies and atomic forces relative to commonly used continuum parameter sets. These radii are suitable for Poisson-Boltzmann calculations with the

AMBER force field and offer energetic congruence to any model that combines molecular mechanics and Poisson-Boltzmann solvation energies.

4.1 Introduction

Continuum solvation models are a useful link between the microscopic and macroscopic realms of theoretical biochemistry, and are widely used to evaluate the electrostatics of biomolecular systems. One of the most common models solves the Poisson Equation (PE) [63] or the Poisson-Boltzmann Equation (PBE)[64, 65] equation as a function of the solute's charge density, a spatially dependent dielectric coefficient, and, in the case of the PBE, ion concentration. The resulting electrostatic potential can be used, for example, to predict electrostatic complementarity of molecular surfaces, to estimate electrostatic binding and transfer free energies, to approximate pK_a values, and to supply electrostatic forces for use in molecular dynamics.

The accuracy of continuum models is dependent on the parameters that are used to define the solute charges, the solvent and solute dielectric coefficients, and the atomic radii which define the dielectric boundary. It has been shown that the average solvent charge density in the continuum model is a sharply peaked function in the region of dielectric discontinuity.[23] The solvent charge density gives rise to the solvent reaction field. Thus, quantitative results are especially sensitive to the location of the solvent-solute dielectric boundary, i.e., the molecular surface. To appreciate where the dielectric boundary should be, it is helpful to relate its macroscopic and microscopic descriptions. Early studies comparing atomistic simulations of aqueous solutions of monatomic ions to the Born model were helpful in elucidating this connection.[25, 66, 67] These studies revealed significant differences in the solvent structure surrounding anions and cations with the same ionic radius but different

solvation energies. The Born model correctly predicts disparate anionic and cationic solvation energies if, and only if, the radii used to define the continuum dielectric boundary correspond with the first peak in the microscopic solvent density. Thus, the optimal Born radii are not the ionic radii, a function of the ion alone, but those that trace out solvent excluded cavities, which are a function of both the ion and the surrounding solvent structure.

Optimal biomolecular radii could, presumably, be measured in a similar manner. Biomolecules are, however, more complicated because the solvent structure surrounding each atom is influenced by neighboring atoms and the solute's conformation. Continuum radii for biomolecules must, therefore, balance accuracy for a given conformation with robustness across multiple conformations. Though the solvent structure can still be used to approximate the location of the dielectric boundary, the radii that define that boundary must ultimately be benchmarked against quantitative observables such as solvation energies or forces. Several continuum parameter sets, consisting of solute charges and radii that complement a specified protein dielectric constant and molecular surface definition, have been optimized to reproduce either experimental solvation energies[68, 69] or explicit solvent simulations[70, 71]. These two methods will be considered in turn.

Referencing continuum models against experimentally determined solvation free energies is an appealing approach because there are many small molecule data to draw upon and because computational models should ultimately be benchmarked by experiment. There are, however, two disadvantages to this approach. First, it necessitates estimating the non-polar contribution to the solvation free energy since electrostatic and non-electrostatic contributions can not be clearly distinguished in experiment. Solvation energies, rigorously defined as the reversible work involved in transferring a molecule from gas phase to bulk solvent, are typically separated into electrostatic and non-electrostatic contributions via a thermodynamic cycle. This cycle involves cavity formation, the introduction of solute-solvent vdW interactions, and the

introduction of solute-solvent electrostatic interactions; $\Delta G_{solv} = \Delta G_{cavity} + \Delta G_{vdW} + \Delta G_{elec}$. PB calculations only account for the electrostatic contribution, ΔG_{elec} . The remaining nonpolar contribution is commonly approximated by the Solvent Accessible Surface Area (SASA) model, $\Delta G_{np} = \Delta G_{cavity} + \Delta G_{vdw} \cong \gamma\, \Delta\mathrm{SASA}$, despite its inadequacy as discussed by Gallicchio *et al.*.[30] The second disadvantage is that one is limited to the small, neutral molecules for which experimental results are available. Accordingly, the transferability of these parameters to biomolecular systems such as proteins, where backbone hydrogen bonding and charged residues likely affect solvent-solute interactions, is unknown.

Benchmarking continuum parameters against explicit solvent simulations avoids the aforementioned challenges; simulations can be applied to charged molecules that mimic biomolecular systems, and thermodynamic cycles can be used to separate solvation energies into electrostatic and non-electrostatic contributions. In the first leg of the thermodynamic cycle, the 'growth' phase, the nonpolar contributions are accounted for by growing a neutral solute cavity into bulk solvent and introducing solute-solvent vdW interactions. In the next leg, the 'charging' phase, the electrostatics contributions are measured by introducing solvent-solute charge-charge interactions. The main disadvantage to using explicit solvent simulations is the error inherent in modern force fields, limited sampling, and simulations techniques. It is encouraging that attempts to minimize these errors have become increasingly successful.[72, 73] Simulations additionally offer energetic congruence to methods that combine continuum solvation and molecular mechanics energies such as end-point free energy calculations [19, 57], pKa calculations[74, 75] continuum dynamics [76], and constant pH molecular dynamics.[77]

It is also possible to compare the forces acting on specific atoms from explicit solvent simulations and continuum models.[78] Continuum force evaluations require a smooth dielectric boundary because abrupt boundaries often result in numerical instability in the

electrostatic potential. Techniques such as Gaussian[28] and cubic spline[29] based volume functions have been introduced to smooth the dielectric boundary, making continuum force calculations and continuum dynamics possible. Spline smoothed surfaces are now standard options in several Poisson Boltzmann solvers.[49, 79] To maintain quantitative accuracy, however, these implementations require significant rescaling of the radii used to define the dielectric boundary. If the radii are not rescaled, solvation energies and forces tend to be overestimated by 10 to 40% (results not shown). Nina *et al.* used explicit solvent simulations to optimize of a set of radii for the CHARMM22 force field for abrupt boundary definitions[70] and later rescaled these radii for spline smoothed boundary definitions.[70, 71]. Since the location of the dielectric boundary is highly dependent on the solute charge distribution, the Nina *et al.* radii are not transferable to other solute charge definitions. Thus, a similar effort is needed for other force fields as has been suggested by a number of authors.[76, 78, 80]

In this chapter two sets of optimized radii are presented for the AMBER (parm99) force field for either abrupt or cubic spline smoothed dielectric boundary definitions. The radii are initially approximated from the solvent charge distributions measured during explicit solvent simulations and then optimized with a genetic algorithm (GA) to reproduce explicit solvent charging energies. Several commonly used continuum parameter sets are tested on the model compounds and four protein-like polypeptide chains. The latter were included to demonstrate the transferability of these parameters to protein systems. Both sets of optimized radii improve the accuracy of continuum solvation energies and the smooth boundary radii improve the correlation between explicit and implicit forces. These radii are suitable for PB quantitative measurements with AMBER partial charges and are recommended for methods that combine AMBER molecular mechanics and PB solvation energies.

The next section describes the model systems and the methodology used in the explicit solvent simulations, the continuum calculations and the genetic algorithm optimizations.

Section 3 defines the radius groups, discusses the explicit solvent charging energies, compares continuum solvation energies from different parameter sets, verifies radii transferability to proteins with protein test cases, and presents the quantitative affects of these radii on atomic forces. Section 4 summarizes the work and possible future directions.

4.2 Methods

4.2.1 Model Systems

The explicit solvent simulations, continuum calculations, and genetic algorithm optimizations were divided into two stages. First, the protein backbone atoms were optimized using 14 polyalanine peptides of varying lengths in common secondary structure conformations. Each conformation (see Table 4.1) was modeled from fragments of either lysozyme (pdb code 1ati) or crambin (pdb code 1ejg). The fragments were mutated to polyalanine and terminated with neutral blocking groups with the MMTSB Tool Set.[81] Second, the side chain radii were optimized using 20 nonzwiterionic N-acetyl-X-N'-methylamide dipeptides where X represents one of the twenty standard amino acids. Two conformations of each side chain dipeptide were used. The first conformation, chosen for the sake of comparison with previous optimizations[70], used extended backbone phi and psi angles (180º, 180º) and the most frequent side chain dihedral angles from a Dunbrack backbone-independent rotamer library. The second conformation used a much more common backbone conformation (-60º, -40º) and the most frequent rotamers from a Dunbrack backbone-dependent rotamer library.[82] All model conformations are provided in the Supporting Information.

Table 4.1: The 14 polyalanine peptides used to optimize the backbone radii.

Polypeptide[a]	Residues	Description[b]	H-bonding
l-beta1	17-20	type II	Yes
l-beta2	36-39	type I'	Yes
l-beta3	39-42	type I	Yes
l-beta4	59-62	type IV	No
l-beta5	60-63	type I	No
l-beta6	85-88	type VIII	No
l-helix1	5-14	type H	Yes
l-helix2	109-114	type H	Yes
l-hairpin1	42-53	class 3:5 IG	Yes
l-hairpin2	51-59	class 4:4	Yes
c-beta1	17-20	type I	Yes
c-beta2	42-45	type IV	No
c-helix1	7-20	type H	Yes
c-helix2	23-30	type H	Yes

[a] Peptides taken from lysozyme (1aki) and crambin (1ejg). [b] Secondary structure descriptions provided by PDBsum.[83]

4.2.2 Explicit Solvent Simulations

The AMBER parm99 force field converted to CHARMM format was used in all simulations. Hydrogen atoms were first energy minimized in vacuum with 50 steps of steepest descent followed by 1000 steps of the Adopted Basis Newton Raphson (ABNR) method. All solute atoms were then fixed for the duration of the simulation. Each model compound was solvated in a sphere of explicit TIP3P water molecules that extended 6.5 Å beyond the dipeptides and 10.0 Å beyond the polyalanine peptides. This resulted in 3-4 hydration shells around every solute atom. Running simulations with larger and smaller solvent shells verified that the chosen dimensions were sufficient for energetic convergence. The spherical solvent boundary potential (SSBP) model including Kirkwood's mulitpolar expansion reaction field was used to approximate the influence of bulk water beyond the explicit water sphere.[84] This model alleviates many of the difficulties that result from perturbing charged systems with periodic boundary conditions and has been shown to give reliable results for proteins, nucleic acids, and, most recently, small molecules.[23, 73, 85] The solvent was first energy minimized

with 50 steps of steepest descent followed by 1000 steps of the ABNR method and then equilibrated for 100 ps.

All simulations employed Langevin dynamics at constant temperature (300 K) using SHAKE enabled 2 fs time steps, infinite cutoffs for nonbonded interactions, and a friction constant corresponding to a relaxation time of 5 ps applied to water oxygen atoms. The pre-equilibrated fully charged systems were simulated for 200 ps to obtain the solvent charge distribution surrounding each solute atom. The solvent charge distributions were used to verify that the continuum radii were properly grouped and to estimate their starting values.

The charging free energies were measured with free energy perturbation (FEP) simulations run in the PERT module of CHARMM.[79] Each simulation consisted of 10 windows in which the solute's charge was scaled by a thermodynamic coupling parameter λ varying by ± 0.1 from 0 to 1 and then from 1 to 0 according to, $q(\lambda) = \lambda q_{final}$. The weighted histogram analysis method (WHAM) was used to combine the results of the individual windows to calculate the total charging free energy. WHAM is a self-consistent iterative procedure that optimizes the distribution of data from separate simulations and thus decreases the amount of sampling required for convergence.[86] Two tests were used to ensure that the simulations were converged. First, the standard error, calculated as half the difference between the forward and reverse WHAM post processed results and shown in Table 4.4, was required to be less than 3% of the total energy. Second, simulations of twice the length were required to be within 2% of the original simulations for the largest and most charged model compounds (results not shown). Different window lengths were required for convergence by the backbone and side chain model compounds; the polyalanine peptides were equilibrated for 10 ps followed by 40 ps of collection, while the dipeptides showed convergence in 5 ps of equilibration followed by 20 ps of collection.

4.2.3 Continuum Calculations

All continuum calculations were performed with the Adaptive Poisson Boltzmann Solver (APBS)[49] using zero bulk ionic strength, a temperature of 300 K, a solvent dielectric of 78.4, and a solute dielectric of 1. The PB grid was centered on each solute and extended at least 20 Å beyond its dimensions. A grid resolution of 0.25 Å was proven sufficient for energetic convergence by comparing calculations with 0.15 Å grid resolution, which resulted in a correlation coefficient of 0.9995 and an average absolute error (AAE) of 0.97 kcal/mol. Solute charges were distributed onto grid points using a cubic B-spline discretization. Electrostatic solvation free energies were calculated from the energetic difference in the solvated ($\varepsilon_{bulk} = 78.4$) and gas phase systems ($\varepsilon_{bulk} = 1$). The molecular surface was defined by the interface of a 1.4 Å solvent probe and the solute radii.

Some of the choices made in the continuum calculations protocol warrant explanation. First, the molecular surface traced by rolling a solvent probe, r = 1.4 Å, around the solute atoms was used to define the location of the abrupt dielectric boundary. Previous PB radii optimizations[70] have used the van der Waals (vdW) surface made up of overlapping solute atoms which excludes the interstitial spaces that lie between solute crevices from the molecular volume. Although the vdW surface definition works well for small molecules, it often results in buried high dielectric pockets in larger molecules such as proteins.[87] These pockets can change quantitative results significantly, making the radii less robust across similar conformations. The vdW surface is particularly problematic in continuum dynamics as high dielectric pockets can appear and disappear as frequently as every time step, resulting in numerical instability.[76] Second, for the spline smoothed surfaces, cubic B-spline charge discretization was used instead of trilinear interpolation to avoid large orientational

artifacts.[28] Finally, a solute dielectric of 1 was chosen for consistency with the non-polarizable force field and fixed solute conformation.

4.2.4 GA Optimization

The GA was used to optimize the radii to their final values. The GA is an efficient stochastic optimization method that has been widely applied to minimization problems because it is ideally suited for multiple-dimensional global search problems where the search space contains multiple local minima and the search variables may or may not be correlated. The GA begins with the generation of an initial population of a given number of solutions. The fitness of each solution is evaluated and a new population is generated via selection, crossover and mutation. This process is repeated until a desired fitness is reached or the maximum number of generations exceeded.

In the initial evolutions, populations of 50 solutions were run for 100 generations. In subsequent evolutions, populations of 100 solutions run for 50 generations. Solutions in the initial population of a given evolution were generated from a uniform distribution +/- 0.10 Å around the starting radii. In subsequent generations, solutions were generated via a process of selection followed by crossover or mutation. Selections were performed with the Stochastic Universal Sampling algorithm,[88] which is designed to give zero bias in the selection and minimal spread. It selects solutions with a probability proportional to their fitness:

$$p_i = \frac{\mathrm{F}_i}{\sum_{j=1}^{N} \mathrm{F}_j} \qquad 4.1$$

where p_i is the probability that solution i will be selected and F_i is its fitness. The fitness function was normalized to ensure that fitness scores remained between 0 (poor) and 1 (perfect):

$$F_i = \frac{1}{(1 - AAE)_i} \quad 4.2$$

where AAE_i is the average absolute error of solution *i*. A uniform crossover process was applied to 90% of the population. Each crossover event consisted of randomly distributing the radii from two previous-generation solutions to two new-generation solutions. Mutations were applied to 20% of the population and consisted of perturbing a solution's radii using a Gaussian distribution with a standard deviation of ± 0.05 Å. The evolution was terminated if the total fitness and the best solution converged to the same value for 6 consecutive generations. In the absence of convergence a new evolution was started with the best radii from the previous evolution.

4.3. Results and Discussion

4.3.1 Radii Grouping and Starting Values

In optimization procedures, the number of parameters that can be meaningfully optimized is generally limited by the number of reference values. Although there is likely an ideal radius for every atom in every conformation, it is desirable to find a set of radii that are robust across multiple conformations. It is generally accepted that atoms in similar chemical environments have comparable surrounding solvent structures and thus similar optimal continuum radii. The factors directly influencing solvent structure are the atom's charge, vdW parameters, and structural neighbors. It may seem appealing to use AMBER atom types to define a set of continuum radius groups. This does not work, however, since AMBER atom types are distinguished by vdW parameters and often have significantly different charges in different residues. Instead, atoms were initially grouped according to similar chemical environments. The groups were then tested for similar surrounding solvent structure.

The solvent charge distribution surrounding each atom type was measured from 100 ps of Langevin dynamics of solvent plus rigid, fully-charged solute molecules. The first peak in the solvent charge distribution corresponds to the closest explicit water molecules and can be used to approximate the optimal continuum radius. Figure 4.1a shows almost identical solvent charge distributions for the carboxyl oxygen in both conformations of the aspartate and glutamate residues. As expected, the first peak is positive due to electropositive water hydrogens crowding around the negatively charged carboxyl oxygen. Conversely, the first peak in the solvent charge distribution surrounding positively charged atoms, such as the amine nitrogen in arginine and lysine (results not shown), are negative due to electronegative water oxygen atoms. Figure 4.1b shows the solvent charge distribution for alpha-carbons in alanine, asparagine, lysine, and glycine. It demonstrates a slightly smaller solvent excluded volume for the alpha-carbon in glycine, likely due to the absence of a side chain. Thus, glycine's alpha carbon was put in a separate radius group from the rest of the alpha carbons. Similar plots were used to classify all of the radius groups which are shown in Table 4,2 along with their starting values.

There are two important distinctions between the optimizations presented in this chapter and those previously presented by Nina et al,.[70] which also used the solvent charge distribution to approximate starting radii. First, non-zero hydrogen radii were used. This was deemed important because FEP charging free energies deviated 2-13% depending upon hydrogen placement (results not shown). GB models have also shown sensitivity to hydrogen radii.[89] Although the GB and PB models are fundamentally different, both are hopeful methods for DNA dynamics and both will likely need to fine tuned their hydrogen radii. Second, although the starting radii work well with vdW surface definitions, as used by Nina *et al,* they clearly underestimated solvation effects with molecular surface definitions. Decreasing

the starting values by 10% reduced the number of evolutions necessary to converge the radii to their optimal values.

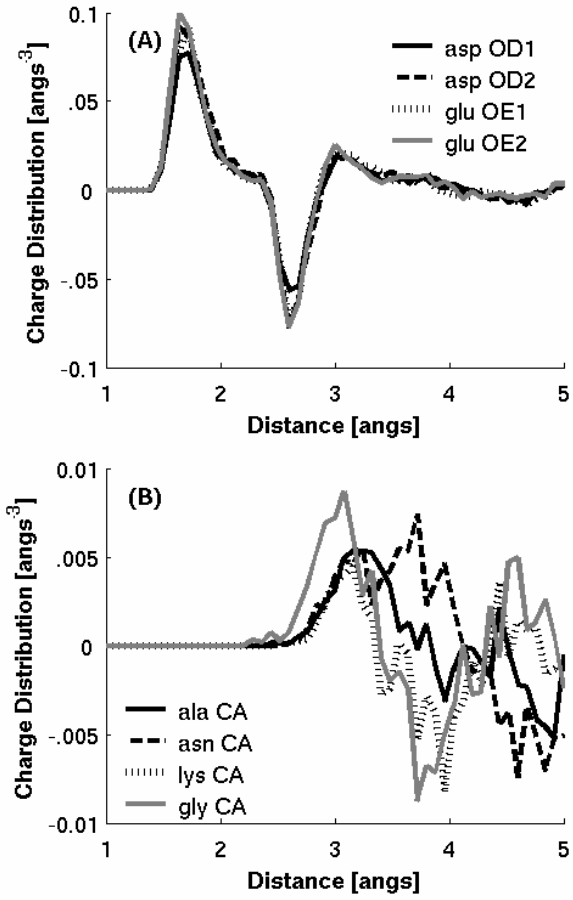

Figure 4.1: Radial solvent charge distributions show that similar chemical environments result in similar solvent structures around (A) the electronegative carboxyl oxygen in asp and glu as well as (B) the alpha carbons of ala, asn, and lys. The alpha carbon of gly reveals a unique solvent structure.

Table 4.2: Radius groups with GA starting values and the final optimized values.

Atom Name[a]	Residues	Start Value	Final Value[b]	Final Value[c]
	Backbone			
C	all	2.30	1.903	2.170
O	all	1.58	1.454	1.742
N	all	2.50	1.835	2.262
CA	all except G	2.80	1.591	2.339
CA	G	2.60	1.654	2.133
CAY/CAT	ACE,NME	2.51	2.595	2.375
	Side Chains			
CB	D,E,C,H,M,F,S,T,W,Y	2.55	2.087	2.370
CB	A,R,N,Q,I,L,K,V	2.75	1.829	2.063
CG*	R,Q,I,L,K,M,T,V	2.49	2.039	2.329
CG	H,F,W,Y	2.10	1.651	2.290
CG/CD	N,Q,D	2.19	1.995	2.257
CG	E	2.45	1.942	2.432
CB/CG/CD	P	2.70	2.008	2.157
CD	R,K	2.81	2.034	2.303
CD*	I,L	2.45	1.897	2.103
CD*/CE*/CZ	H,F,W,Y	2.05	1.837	2.122
CE	M	2.40	1.902	2.157
CZ/CE	R,K	2.66	2.020	2.414
OD*/OE*	N,Q,D,E	1.55	1.516	1.727
OG*	S,T	1.65	1.562	1.832
OH	Y	1.72	1.738	2.022
NE,NH*,NZ	R,K	2.48	1.523	1.861
ND2/NE2	N,Q	2.12	2.222	2.453
ND1,NE2	H	1.90	1.436	1.782
NE1	W	2.11	1.898	2.147
SG/SD	C,M	2.00	1.978	2.169
	Hydrogens[d]			
type H	bb HN	1.20	1.600	1.967
type H	bound to N	1.20	1.119	1.379
type HO/HS	bound to O/S	1.00	1.201	1.406
type H1/HP	polar	1.31	1.914	2.033
type HC/HA	nonpolar	1.30	0.840	1.321

[a] Radius groups are distinguished by AMBER atom names for all heavy atoms and by atom type for hydrogen atoms. [b] Final radii for abrupt dielectric surfaces and [c] spline smoothed surfaces [d] Hydrogens specified by atom type with type 'H' divided into two groups; amide backbone 'HN' and all other N-bound hydrogens.

4.3.2 FEP Simulations

The approximate starting radii were fine tuned to reproduce explicit solvent FEP charging free energies. These energies, shown in tables 4.3 and 4.4, were used as the target values in the GA optimizations. All simulations were well converged as previously described, and as indicated by the reported standard errors. Values for the polyalanine peptides ranged from -31 to -81 kcal/mol depending on their length and conformation, while those of the side chain dipeptides ranged from -12 to -83 kcal/mol. The neutral nonpolar residues had the lowest magnitudes, -12 to -16 kcal/mol; the polar residues were slightly higher, -16 to -28 kcal/mol; and the charged residues were the highest, -61 to 83 kcal/mol. Comparing the charging free energies between dipeptide conformations demonstrates a moderate range of conformational sensitivity. Lysine demonstrated the largest range with a charging free energy of -61 kcal/mol for one conformation and -77 kcal/mol for the other.

Table 4.3: FEP charging and continuum solvation energies for the 14 polyalanine peptides using AMBER and optimized radii.

Polypep	WHAM		AMBER	Opt[a]	Opt[b]
l-beta1	-34.17	(0.06)	-31.87	-35.44	-33.53
l-beta2	-32.69	(0.05)	-25.67	-30.68	-31.85
l-beta3	-30.89	(0.48)	-27.88	-31.40	-30.79
l-beta4	-31.42	(0.31)	-28.89	-32.26	-32.88
l-beta5	-33.69	(0.03)	-32.18	-34.83	-32.87
l-beta6	-31.09	(0.06)	-27.27	-31.53	-32.34
l-helix1	-65.29	(0.45)	-53.78	-64.15	-65.31
l-helix2	-49.18	(0.42)	-43.36	-48.18	-51.49
l-hairpin1	-66.50	(1.33)	-55.70	-66.14	-66.63
l-hairpin2	-81.45	(1.06)	-69.17	-81.60	-81.43
c-beta1	-34.72	(0.24)	-33.64	-34.54	-33.90
c-beta2	-40.90	(0.24)	-36.47	-40.88	-41.06
c-helix1	-68.91	(0.44)	-54.99	-66.39	-67.15
c-helix2	-51.91	(0.04)	-44.01	-51.05	-51.77

All energies in kcal/mol. Standard errors are reported as half the difference between the forward and reverse WHAM energies. a Abrupt and b spline smoothed dielectric boundary definitions.

Table 4.4: Explicit solvent charging free energies (WHAM) and continuum solvation energies for the 20 dipeptide amino acid mimics using AMBER and optimized radii.

Res	WHAM		AMBER	Opt[a]	Opt[b]	Res	WHAM		AMBER	Opt[a]	Opt[b]
Nonpolar Groups											
Gly	-13.91	(0.10)	-14.35	-14.43	-13.93	**Thr**	-16.25	(0.18)	-17.06	-16.69	-16.52
Gly2	-18.45	(0.15)	-18.51	-18.36	-18.11	**Thr2**	-19.58	(0.30)	-19.68	-19.11	-19.24
Ala	-14.01	(0.18)	-13.77	-14.26	-13.82	**Cys**	-18.18	(0.36)	-18.69	-18.17	-18.11
Ala2	-17.86	(0.02)	-16.97	-16.92	-16.90	**Cys2**	-17.97	(0.07)	-18.39	-17.96	-17.95
Val	-12.48	(0.17)	-11.22	-12.58	-12.45	**Tyr**	-19.78	(0.13)	-20.49	-20.42	-19.77
Val2	-18.79	(0.45)	-16.99	-17.36	-17.73	**Tyr2**	-25.30	(0.23)	-25.32	-25.08	-24.26
Leu	-12.58	(0.02)	-11.72	-12.68	-12.40	**Asn**	-20.88	(0.10)	-23.09	-20.89	-20.86
Leu2	-18.45	(0.06)	-16.77	-17.97	-17.94	**Asn2**	-23.38	(0.47)	-24.96	-23.36	-23.12
Ile	-11.99	(0.06)	-10.87	-12.01	-12.00	**Gln**	-18.40	(0.47)	-20.16	-18.43	-18.68
Ile2	-19.29	(0.06)	-17.25	-17.52	-17.85	**Gln2**	-27.63	(0.24)	-29.29	-26.40	-26.35
Pro	-14.75	(0.01)	-13.29	-14.76	-14.74	**His**	-23.63	(0.07)	-24.24	-24.77	-24.26
Pro2	-16.84	(0.16)	-14.66	-16.75	-16.98	**His2**	-22.12	(0.16)	-21.13	-22.52	-21.56
Phe	-16.42	(0.46)	-14.68	-16.45	-16.43	**Charged Polar Groups**					
Phe2	-21.95	(0.01)	-20.16	-21.66	-20.97	**Arg**	-60.99	(0.44)	-69.19	-63.89	-63.05
Trp	-19.76	(0.04)	-19.50	-19.73	-19.73	**Arg2**	-77.28	(0.70)	-81.88	-76.10	-74.50
Trp2	-22.23	(0.45)	-23.25	-23.14	-21.93	**Lys**	-65.99	(0.06)	-71.66	-67.44	-68.71
Met	-12.98	(0.02)	-12.76	-13.12	-13.09	**Lys2**	-82.65	(1.14)	-85.93	-81.69	-80.33
Met2	-19.53	(0.23)	-18.40	-18.85	-19.04	**Asp**	-77.86	(0.47)	-72.05	-77.90	-78.02
Neutral Polar Groups						**Asp2**	-79.38	(0.18)	-71.06	-77.35	-77.33
Ser	-17.34	(0.03)	-18.28	-17.33	-17.29	**Glu**	-79.00	(0.51)	-72.09	-79.09	-78.99
Ser2	-19.60	(0.18)	-21.03	-19.68	-19.82	**Glu2**	-76.31	(0.24)	-69.47	-76.30	-76.30

All energies in kcal/mol. Conformation 1 and 2 have phi/psi angles of (180º,180º) and (-40º,-60º) respectively. Standard errors are reported as half the difference between the forward and reverse WHAM energies. [a] Abrupt and [b] spline smoothed dielectric boundary definitions.

4.3.3 GA Optimizations

Results for the GA optimizations were typical for a highly dimensional rough energy landscapes. Specifically, different radius sets with similar fitness values were often encountered. After 6 evolutions the optimal abrupt and smooth boundary radii, as presented in Table 4.2, had AAEs of 0.5433 and 0.6355 respectively.

4.3.4 Continuum Calculations Testing Parameter Sets

In order to gauge the importance of these optimized radii they were compared to several commonly used continuum parameter sets: AMBER (parm99) charges combined with parm99 vdW radii, Bondi radii, and the previously optimized Nina *et al.* radii in addition to PARSE charges combined with PARSE radii. Figure 4.2 shows the resulting continuum solvation

energies compared to explicit solvent charging energies. The relative performance between parameter sets is more clearly illustrated in Figure 4.3a where the continuum deviation for each model compound is represented as a single point.

Each parameter set was tested for the appropriate solute dielectric and surface definitions for the comparisons made in figures 4.2 and 4.3a. Failure to use the appropriate solute dielectric and surface definitions results in overestimating or underestimating solvation effects. For example, the PARSE parameters, originally optimized with a solute dielectric of 2, consistently overestimate solvation effects when used with a solute dielectric of 1. Likewise the Nina *et al.* radii, which were optimized with a vdW surface, underestimate solvation effects with a molecular surface definition. Although changing either the solute dielectric or the surface representation generally increases or decreases solvation energies it has little effect on the relative solvation energies between different solutes. This is demonstrated in Figure. 4.3b with the Bondi radii where using a vdW surface and solute dielectric of 1 overestimates solvation effects and a molecular surface with a solute dielectric of 2 underestimates them.

The AMBER parm99 vdW radii were tested because they are frequently chosen for use with parm99 charges despite the fact that they have never been optimized for continuum models. Optimal performance, found with a molecular surface definition and a solute dielectric of 1, yielded an AAE of 3.18 kcal/mol. The Bondi radii [90] were tested because they are the most common choice for the intrinsic radii used in generalized Born models.[14] Using a solute dielectric of 1 and a molecular surface representation they performed the best out of the un-optimized parameters, with an AAE of 2.27 kcal/mol.

The Nina *et al.* radii were tested to query the transferability of radii between the AMBER and CHARMM force fields. They were not expected to work perfectly with the AMBER charges because continuum charge and radius definitions are strongly interdependent. It has been postulated, however, that the two force fields should have similar solute-solvent

interactions, similar charging free energies, and thus similar continuum radii.[91] To test this, FEP simulations were run with the CHARMM force field. They yielded substantially different charging free energies and solvent radial distributions. In continuum calculations using a solute dielectric of 1 and a vdW surface, the Nina *et al.* radii combined with AMBER charges resulted in an AAE of 2.67 kcal/mol. This is an improvement over parm99 vdW radii but worse than the Bondi radii. It is far worse than their performance with the CHARMM charges when compared to FEP simulations run with the CHARMM force field (results not shown), which had an AAE of 0.69 kcal/mol. Differences in the FEP charging free energies, the solvent radial distribution functions, and the optimal continuum radii for AMBER and CHARMM are likely due to differences in partial charges.

Finally, the PARSE parameters were tested as they have likely been the most respected and frequently used continuum charge and radius definitions since their development in 1994.[68] Their performance relative to the explicit solvent model was worse than expected. Regardless of the surface and solute dielectric definitions, they consistently overestimated solvation effects on the model systems. Optimal performance was found with a molecular surface definition and a solute dielectric of 2 resulting in an AAE of 5.22 kcal/mol.

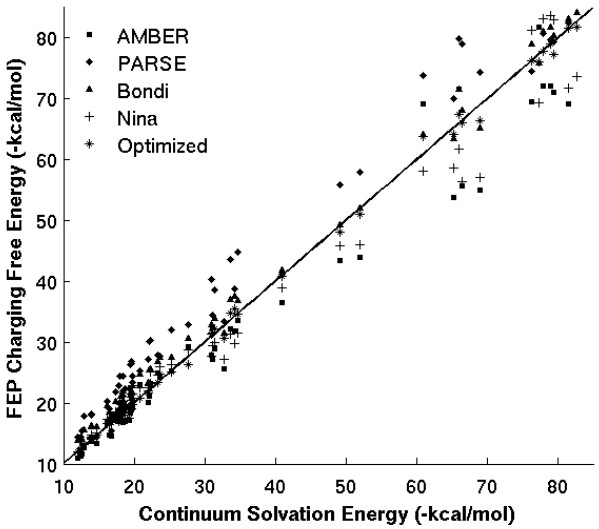

Figure 4.2: Explicit solvent charging free energies versus continuum solvation free energies calculated with the AMBER charges and different radii.

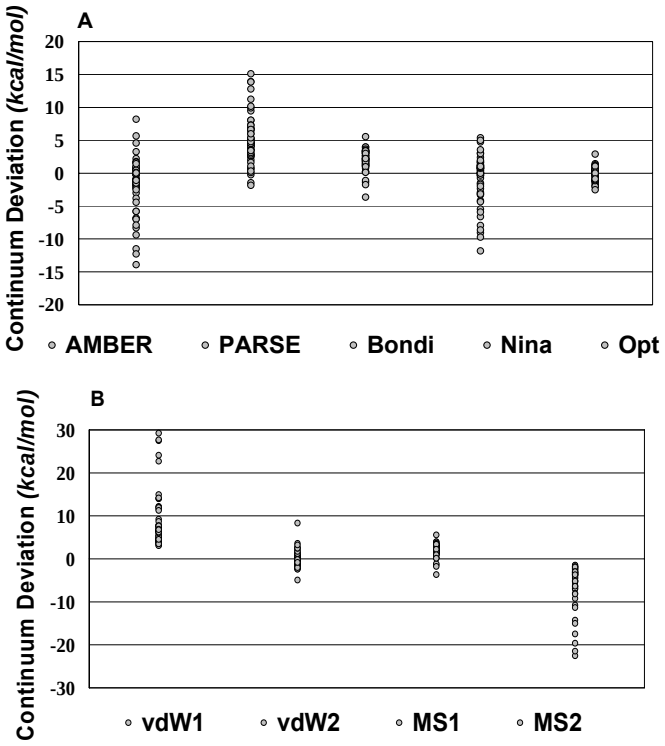

Figure 4.3: Continuum solvation energy deviations from explicit solvent reference values. Each circle represents one model compound. (A) Five different parameter sets are compared with their optimal surface definitions and solute dielectrics: PARSE charges with PARSE radii, a molecular surface, and solute dielectric of 2; AMBER charges with AMBER radii a molecular surface and solute dielectric of 1; AMBER charges with Bondi radii a molecular surface and solute dielectric of 1; AMBER charges with Nina radii a vdW surface and solute dielectric of 1; and AMBER charges with the newly optimized radii a molecular surface and solute dielectric of 1. (B) The effect of different surface and solute dielectric definitions on solvation energies calculated with the Bondi radii.

4.3.5 Protein Test Cases

The optimized radii clearly perform better on the model compounds for which they were optimized, but this is a biased test. As an unbiased test and to evaluate their transferability to proteins, FEP simulations were run on four short polypeptide chains: Trpcage (1l2y),[92] the C-terminal fragment (41-56) of protein G (2gb1),[93] the C-peptide of ribonuclease A (8rat)[94], and the first helix taken from lysozyme (residues 5-14) with all of the native side chains present. The proteins were prepared and simulated with the same procedures used on the model compounds. Simulations were held to the same requirements for energetic convergence.

The resulting charging free energies are shown in Table 4.5 in addition to continuum deviations with the AMBER, PARSE, Bondi, and newly optimized radii. The surface definitions and solute dielectrics found optimal for the model compounds, as described in the previous section, were used for consistency. The optimized radii perform the best on abrupt boundary surfaces with an AAE of 5.96 kcal/mol. All abrupt boundary radius sets overestimated solvation energies by 20-40% when used with the spline smoothed surfaces (results not shown). The smooth boundary radii, presented in the 9th column, are a significant improvement, but still overestimate solvation effects slightly. The source of this systematic overestimation is unknown. It can be compensated for by using a higher solute dielectric, which, as previously mentioned, shifts all solvation energies down but has little effect relative continuum deviations. As shown in the final column, the AAE can be decreased to 4.42 kcal/mol by using a solute dielectric of 1.15 instead of 1.0.

Table 4.5: FEP charging energies and continuum deviations for 4 protein-like polypeptides.

Protein	WHAM-40ps	WHAM-60ps	AMBER	PARSE	Bondi	Nina	Opt[a]	Opt[b]	Opt[c]
trpcage	-309.85 (2.00)	-308.63 (1.89)	-20.47	-27.41	1.10	-2.97	0.51	19.57	-3.32
protein G	-314.37 (0.51)	-314.60 (0.25)	-17.78	14.58	21.55	70.55	15.74	19.72	4.86
ribonucleaseA	-167.77 (1.55)	-165.25 (1.21)	-6.21	38.76	11.69	45.41	3.82	17.37	3.83
lysozyme	257.88 (3.89)	-259.79 (2.53)	4.99	-3.09	1.93	31.16	3.77	5.99	-5.69
AAE			**12.36**	**20.96**	**9.07**	**37.52**	**5.96**	**15.66**	**4.43**

All energies in kcal/mol. FEP standard errors are reported as half the difference between the forward and reverse WHAM energies. [a] Abrupt and [b] spline smoothed dielectric boundary definitions with solute dielectric of 1. [c] spline smoothed boundary with solute dielectric of 1.15.

4.3.6 Spline Smoothed Radii and Atomic Forces

If abrupt boundary radii are used with a spline smoothed dielectric boundary, then solvation forces, similar to solvation energies, tend to be overestimated by 10 to 40% (results not shown). The width of the spline smoothed dielectric transition region, often referred to as

the *spline window*, has been shown to effect solvation energies.[71] Thus, the optimal spline smoothed radii will be dependent on the spline window width. Previous optimizations have shown that reasonable quantitative accuracy can be maintained by rescaling abrupt-boundary radii by a single factor that is dependent on the spline window width.[71] This, however, assumes that all radii are equally affected by the spline smoothing. This assumption was tested by optimizing a set of radii for a single spline window of 0.3 Å with the GA. These radii were compared to abrupt boundary radii scaled by a single scaling factor. A simplex optimization showed that the optimal single scaling factor was $R_{spline} = (R_{abrupt} + 0.3) * 1.0107$. The independently optimized radii, shown in Table 4.2, are significantly different from those obtained from the single scaling factor. The AAEs for the independently optimized and single scaled radii were 0.66 kcal/mol and 2.11 kcal/mol respectively.

Recently, Wagoner *et al.* suggested comparing continuum and explicit solvent forces instead of molecular solvation energies to test the continuum model.[78] To obtain the explicit solvent electrostatic forces they averaged the total minus non-polar solvent-solute interactions from a simulation of a 131-residue protein in a fixed conformation. They used the AMBER parm99 vdW radii and demonstrated reasonable correlation between continuum and explicit solvent forces with a correlation coefficient r = 0.88. As expected, however, the continuum forces were systematically overestimated by a factor of 2.2, as indicated by the slope from linear regression analysis. To see what effect the optimized radii have on atomic forces, the same procedure was carried out with Trpcage and the C-terminus of G-Protein. Explicit solvent forces were reasonably converged after 200ps of simulation. This was tested by comparing forces from a 500 ps simulation which resulted in a correlation coefficient of 0.9987 and an AAE of 0.10 kcal/mol. Figure 4.4a shows the improved correlation for atomic forces calculated with the optimized radii, r = 0.94, versus the AMBER radii, r = 0.84. The AMBER radii still

systematically overestimate the continuum forces by a factor of 1.99 while the optimized radii slightly underestimate the forces by a factor of 0.80. Figure 4.4b shows the optimized atomic forces according to atom type and demonstrates that forces on oxygen atoms are most often underestimated. Comparing specific atomic forces from continuum and explicit solvent models may prove useful in future optimization efforts.

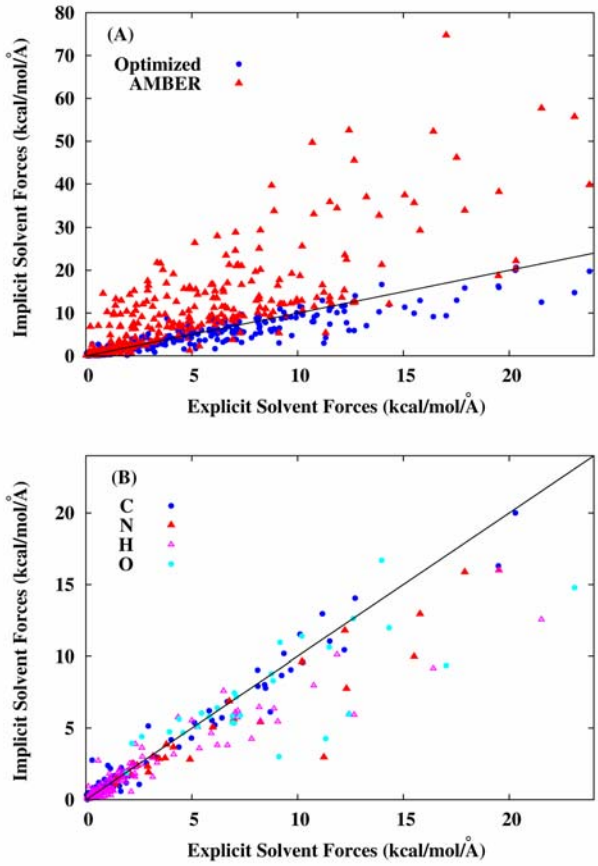

Figure 4.4: Comparison of explicit solvent and continuum forces. Lines indicate y = x NOT best fit. A) Continuum forces calculated with AMBER radii in red: correlation coefficient 0.84; slope 1.99; intercept 0.03 kcal mol-1 Å-1. Continuum forces calculated with optimized radii in blue: correlation coefficient 0.94; slope 0.80; intercept 0.29 kcal mol-1 Å-1. B) Continuum forces calculated with optimized radii only distinguishing different elements by color.

4.4. Conclusions

Computational biochemistry is a field with multiple levels of models and theories whose computational requirements scale proportionally with accuracy. The success of the lower level, less accurate, more efficient models relies upon their connection to the higher level, more accurate, less efficient models while the latter are complimented by the former to characterize complex biological systems. As methodology advances, opportunities for improved synergy between levels of theory and models are created. As microscopic, explicit solvent simulations become more accurate and the numerical methods used in continuum models improve, there is an increasing opportunity and need to benchmark continuum models on microscopic simulations.

The main results of this paper are two sets of optimized continuum radii, presented in Table 4.2, for PB calculations with the AMBER force field using either abrupt or spline smoothed dielectric boundary definitions. Thirty-one radius groups were defined based on similar chemical environments. Both sets of optimized radii improve quantitative agreement with microscopic simulations when compared to AMBER parm99 vdW, Bondi, and Nina *et al.* radii as well as the PARSE parameters. The Bondi radii combined with a protein dielectric of 1 and a molecular surface definition were the best alternative choice for AMBER charges. The PARSE parameters, shown to work optimally with a solute dielectric of 2 and a molecular surface definition, were less accurate. The abrupt boundary radii work quite well for protein systems. The smooth boundary radii offer a significant improvement over abrupt boundary radii used on spline smoothed surfaces, but still overestimated the solvation energies of four protein-like polypeptides. This overestimation can be compensated for by using a solute dielectric of 1.15 instead of 1.0. The smooth boundary radii greatly improve agreement between continuum and explicit solvent atomic forces.

Future efforts will focus on identifying the source of deviations of the smoothed boundary radii on protein systems as well as small molecule parameterization and coupling these parameters with nonpolar solvation effects. If one hopes to use these radii on systems involving small molecule ligands, it will be important to characterize the appropriate radii for certain small molecules as well. It will also be of great use to test the complementarity of these electrostatic parameters with various nonpolar solvation models by comparing total solvation energies to either explicit solvent simulations or experimental solvation energies. It is expected that the nonpolar and polar contributions to continuum solvation models will be coupled and that the greatest degree of accuracy will require them to be treated as such.

This chapter is a reprint in full of material that appeared in *Optimized radii for Poisson-Boltzmann calculations with the AMBER force field*, Swanson, J.M.J., S.A. Adcock, and J.A. McCammon, J. Chem. Theory Comput. (3); 484-493 (2005). I was the primary researcher and author of this work.

Chapter Five

Limitations of Atom-Centered Dielectric Functions in Implicit Solvent Models.

ABSTRACT

Many recent advances in Poisson-Boltzmann and Generalized Born implicit solvent models have used atom-centered polynomial or Gaussian functions to define the boundary separating low and high dielectric regions. In contrast to the Lee and Richards molecular surface, atom-centered surfaces result in inter-atomic crevices and buried pockets of high dielectric which are too small for a solvent molecule to occupy. We show that these interstitial high dielectric regions are of significant magnitude in globular proteins, that they artificially increase solvation energies, and that they distort the free energy surface of non-bonded interactions. These results suggest that implicit solvent dielectric functions must exclude interstitial high dielectric regions in order to yield physically meaningful results.

5.1 Introduction

Continuum solvent models have become an increasingly useful tool in the characterization of biomolecular systems. The most popular such methods employ either the Poisson-Boltzmann (PB) or Generalized Born (GB) models, treating the solute as a set of point charges in a low dielectric cavity and the surrounding solvent as a uniform high dielectric medium. The PB model is generally considered to be more accurate and is often used to

benchmark GB models. GB has found more extensive application in dynamical simulations, however, because it is computationally efficient and more amenable to force calculations.

One of the main challenges in the use of PB for dynamics has been the determination of numerically stable and accurate forces. Most PB calculations have used a dielectric boundary based on the molecular surface (MS) as defined by Lee and Richards,[26] which results in forces that are unstable over time, lack analytical definition, converge poorly, and are sensitive to grid discretization.[28] Furthermore, an abrupt dielectric transition results in numerical instability regardless of the location of the boundary. Recent advances in PB methods have avoided these difficulties by using overlapping atom-centered Gaussian or polynomial functions to define the solute surface, resulting in analytically defined, differentiable dielectric functions with smooth transitions between low and high dielectric values.[28, 29] These dielectric definitions increase force stability and computational efficiency. However, unless modifications such as those presented by Lu and Luo[76] or Lee *et al.*[95] are employed, they result in inter-atomic crevices and buried pockets of high dielectric which are too small for a solvent molecule to occupy.[18, 76, 95] It was originally postulated that the consequences of these regions, henceforth called *interstitial high dielectrics*, would be minimal and that either a MS or an atom-centered surface definition should be physically and theoretically equivalent,[28, 96] but more recent work has suggested that atom-centered surfaces are physically flawed.[18, 95] Nevertheless, implicit solvent models based on unmodified atom-centered dielectric functions are becoming increasingly popular in the biophysical community.[28, 29, 78, 96-98]

Here we report results showing that atom-centered surfaces create interstitial high dielectric regions of significant magnitude in globular proteins, increase solvation energies, and distort the free energy surface of non-bonded interactions. Although similar results are expected for most atom-centered smoothed dielectric boundaries, we focus on the spline

surfaces (SS) introduced by Im *et al.* [29] and implemented in the Adaptive Poisson-Boltzmann Solver (APBS),[49] the PBEQ module in CHARMM,[79] and the GBSW model.[97]

5.2 Methods

The protein surface and energy calculations were performed with APBS 0.3.2 using a grid resolution of 0.2 Å. To facilitate comparison between the different dielectric boundaries, we have chosen to use the Nina *et al.* optimized radii to define the van der Waals surface (vdWS), MS, and SS.[70, 71] For the results with different surface definitions to be comparable, the radii must be rescaled for the MS and particularly for the SS; we have performed the recommended rescaling. To check that results were not specific to Nina *et al.* radii, they were verified with AMBER optimized radii,[99] parm22 radii,[79] and Bondi radii.[90] The latter two were augmented by the spline window width (w = 0.3 Å) for the SS: a generic but very reasonable scaling. APBS versions 0.3.2 and earlier have a flaw in the MS algorithm that, in our testing, overestimates MS volumes by 2-5% and underestimates solvation energies by 1-3%. Results shown here were calculated with a modified algorithm that corrects this problem. The energy calculations were performed with zero bulk ionic strength, a temperature of 300 K, a solvent dielectric of 80, a solute dielectric of 1 and charges from the CHARMM22 all hydrogen force field in accordance with the radii.

Implicit solvent potentials of mean force (PMFs) for hydrogen bond formation were calculated by combining solvation, coulomb and vdW energies. PB solvation energies were calculated with APBS and the same parameters as used for protein solvation energies except for a finer grid resolution of 0.1 Å. GBMV and GBSW solvation energies, coulombic energies and vdW energies were calculated with CHARMM 31a1. AMBER GB solvation energies were calculated using the igb=1 model in AMBER 8. Because AMBER GB models are not

compatible with atoms having zero radius, the hydrogen radii were increased to 0.8 Å. This made the outer surfaces of the hydrogen atoms approximately coincident with the surfaces of the atoms to which they were bound. The explicit solvent PMFs were calculated by WHAM from results of umbrella sampling in TIP3P solvent. Umbrella sampling was carried out using the PMEMD module of sander, modified to apply harmonic restraints to only the y and z coordinates of the peptides. Due to the use of different force fields for implicit and explicit solvent measurements, no quantitative comparison should be made. However, explicit solvent potentials calculated with the CHARMM force field have the same general shape.[100]

5.3 Results and Discussion

To probe the magnitude of interstitial high dielectric regions in globular proteins we compared the solute volumes generated by vdWS, MS, and SS. Figure 5.1 shows the MS and SS dielectric values on a plane intersecting a structure of Intestinal Fatty Acid Binding Protein (IFABP) taken from a molecular dynamics (MD) simulation. Although MD conformations might be expected to contain more interstitial high dielectrics than NMR or crystal structures, surprisingly similar plots were obtained for all the systems in Table 5.1. The solute volumes, reported in Table 5.1, show a consistent trend across all 6 structures: substantial interstitial high dielectrics with the SS definition. Within the MS volume the SS renders 65 – 262 $Å^3$ of interstitial high dielectric space with a value of 80 and an additional 502 - 1121 $Å^3$ with values over 20. The total interstitial high dielectric space ranges from 12% for the crystal structures to 15% for the NMR and MD structures. The quantitative effects of these regions on electrostatic solvation energies are shown in Table 5.2; SS energies are overestimated by 11 – 21%, only slightly less than the overestimation by vdWS.

The volume and dielectric value of the interstitial spaces created by the SS can be decreased by using a larger spline smoothing window for SS or longer Gaussian tails for Gaussian surfaces, but interstitial high dielectrics can not be eliminated altogether.[28] Unfortunately, this overestimates the size of solvent exposed atoms and creates unphysical bulges around overlapping and adjacent atoms.[95] These expanded dielectric boundaries yield solvation energies and forces that are severely underestimated. For example when a spline window of 1.0 Å is applied to the systems in Table 5.1, interstitial high dielectrics with $\varepsilon > 20$ are essentially eliminated, but the volume of lowered dielectric outside the MS is increased dramatically and the solvation energies are underestimated by 35% to 48%. This over or under-estimation of solvation energies by SS has also been reported by Lee *et al.*[6] who used hybrid explicit/implicit solvation energies to test various continuum surface definitions.

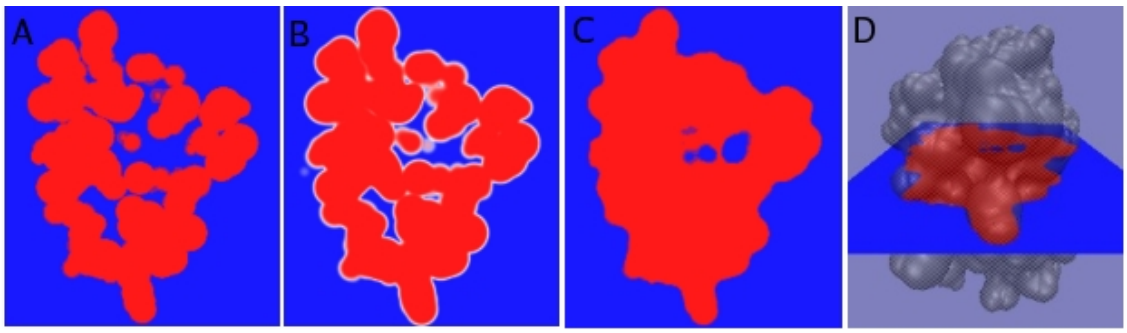

Figure 5.1: Dielectric maps of IFABP. Dielectric values on a plane intersecting IFABP for the vdWS (A), SS (B) and MS (C). Red regions have $\varepsilon = 1$, blue have $\varepsilon = 80$ and white regions have intermediate dielectric values. The location of the intersecting plane is shown in (D).

Table 5.1:. Solute volumes for different surface definitions. Solute volumes [Å^3] for FKBP12 (1fkg), lysozyme (1aki), protease (1kzk) and IFABP (1ael). (b) The SS is larger because almost 1/4 of its volume has dielectric (ε) values less than 80 but greater than 1. (c) Interstitial high dielectric volumes are measured by the volume within the MS that has high dielectric values in the SS. The total percentage of interstitial high dielectrics relative to the MS volume, shown in parenthesis, range from 11% to 14%, increasing for the MD and NMR structures.

protein	vdWS	MS	SS	interstitial high dielectric					
	$\varepsilon = 1$	$\varepsilon = 1$	$\varepsilon < 80$	$1 < \varepsilon < 20$		$20 < \varepsilon < 80$		$\varepsilon = 80$	
1FKG-xtal	14863.3	16087.5	18508.6	1345.8	(8.4%)	502.5	(3.1%)	65.3	(0.4%)
1AKI-xtal	17495.8	18954.9	21473.0	1519.4	(8.0%)	608.1	(3.2%)	100.5	(0.5%)
1KZK-xtal	27528.0	29937.1	33885.2	2512.8	(8.4%)	1006.9	(3.4%)	139.7	(0.5%)
1KZK-md	27617.3	30245.4	34081.5	2576.2	(8.5%)	1120.9	(3.7%)	202.3	(0.7%)
1AEL-nmr	17589.4	19361.9	21869.6	1532.7	(7.9%)	797.0	(4.1%)	196.8	(1.0%)
1AEL-md	19318.5	21585.4	24579.4	1930.0	(8.9%)	1031.1	(4.8%)	262.5	(1.2%)

Table 5.2: Electrostatic solvation energies (kcal/mol) for different surface definitions. The SS yields energies much larger than the MS and similar to the vdWS. Percentages of vdWS and SS overestimation relative to the MS energies are given in parenthesis.

protein	MS	vdWS	SS
1fkg-xtal	−1475.8	−1680.4 (13.9%)	−1638.3 (11.0%)
1aki-xtal	−1724.3	−2034.7 (18.0%)	−1976.7 (14.6%)
1kzk-xtal	−2196.0	−2585.6 (17.7%)	−2499.9 (13.8%)
1kzk-md	−2162.5	−2549.1 (17.9%)	−2475.9 (14.5%)
1ael-nmr	−2247.4	−2838.2 (26.3%)	−2727.0 (21.3%)
1ael-md	−1979.3	−2368.9 (19.7%)	−2301.9 (16.3%)

While the ability to calculate atomic forces in a PB model is an important advance, such forces can have useful application only if the potential they are derived from accurately represents the physics of the system. In particular, solvation models employed in dynamical simulations must be capable of accurately calculating high energy as well as low energy conformations. Therefore, dynamics may constitute a more demanding test of a solvation model than calculating solvation energies of static structures, which tend to be dominated by low energy configurations of atoms.

Hydrogen bonds are of particular interest in simulations of biomolecules; since solvation effects make a large contribution to these interactions, the PMF for the separation of a hydrogen bond can be used as a test of the quality of a solvation model. The PMF of hydrogen bonding between the delta hydrogen and the epsilon nitrogen of two delta protonated histidines calculated with a variety of solvation methods is shown in Figure 5.2. Both PB and GB results based on a MS dielectric boundary faithfully represent the important features of the explicit solvent PMF: a narrow minimum and a significant barrier to separation of the hydrogen bond. The energetic barrier in the MS PMFs comes about because the electrostatic energy rises rapidly as soon as the hydrogen bond participants are separated, but the solvation energy does not substantially increase in magnitude (become more negative) until the bond is sufficiently

separated that the solvent probe will fit between the participants. The SS based implicit solvent models have good performance near the minimum and at long distances, but fail to capture the appropriate energetic barrier. At the separation where the MS PMF energy peaks, the SS has large interstitial high dielectrics, which result in a more negative solvation energy. This produces an artifactual minimum -- or in less extreme cases, a shoulder – near a location where the PMF should have a maximum. Discrepancies between MS and SS PMFs are most dramatic for interactions between sterically bulky groups, where the magnitude of the interstitial high dielectrics is largest, but are observed to a greater or lesser degree across a variety of hydrogen bond and salt bridge systems as shown in figures 5.3 through 5.6. AMBER (igb=1) GB results, based on the model of Hawkins, Cramer and Truhlar[101], illustrate that the lack of a barrier to separation is a general feature of implicit solvent models that allow interstitial high dielectrics. In comparison to the SS results, the AMBER GB PMF is somewhat broader near the minimum, but is smoother and avoids the second minimum seen in the SS PMF. For simplicity, apolar solvation contributions have been ignored in the implicit solvent PMFs presented here. A traditional surface area apolar term changes the depth of the minimum, but has no appreciable effect on the discrepancies between MS and SS PMFs.

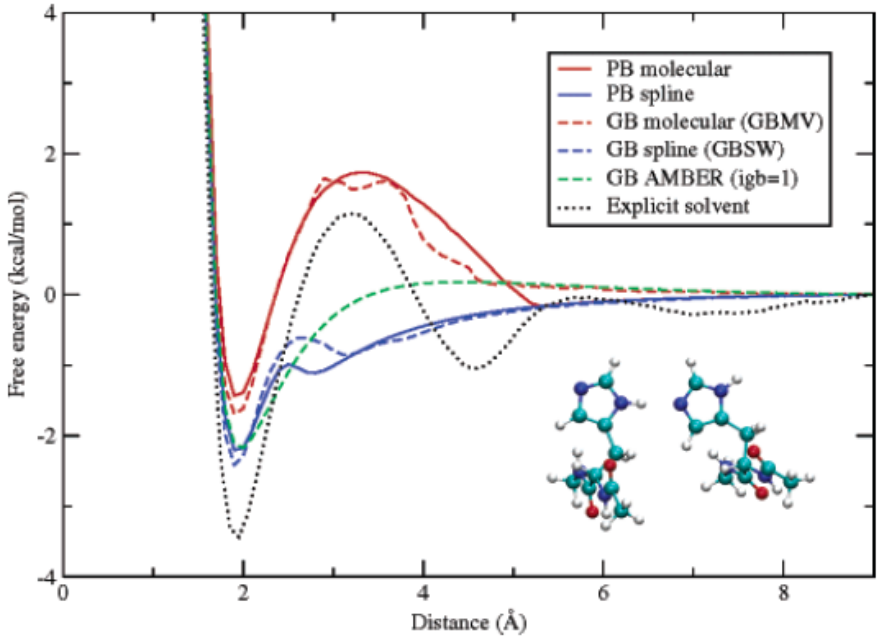

Figure 5.2: PMF for the illustrated histidine-histidine hydrogen bond. Distances are measured between the hydrogen and nitrogen atoms participating in the bond. Spline-based dielectric boundaries fail to capture the free energy barrier to hydrogen bond separation because they allow interstitial high dielectrics near the hydrogen bond as it is separated.

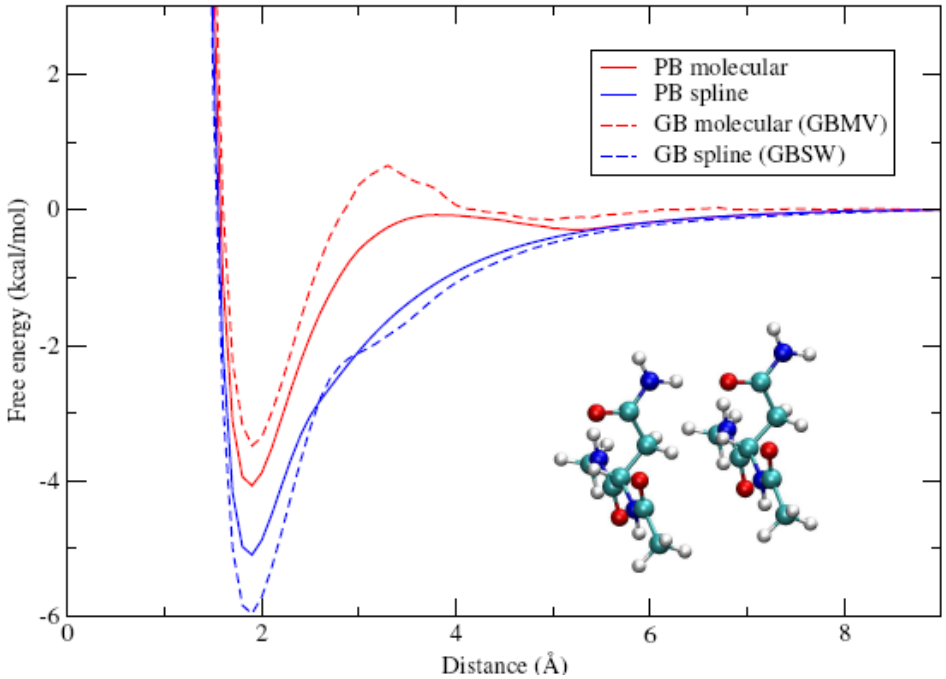

Figure 5.3: PMF for the illustrated asparagine-asparagine hydrogen bond. Distances are measured between the hydrogen and oxygen atoms participating in the bond.

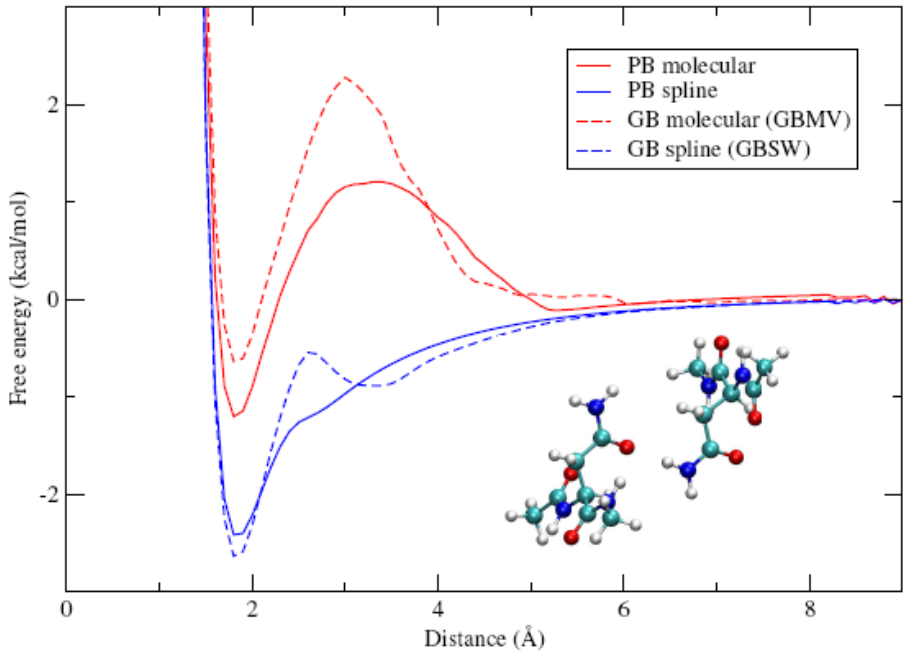

Figure 5.4: PMF for the illustrated asparagine-asparagine hydrogen bond, which is similar to that seen in Figure 5.3, but with one of the bonding partners rotated 180 degrees about the axis of the hydrogen bond. Distances are measured between the hydrogen and oxygen atoms participating in the bond.

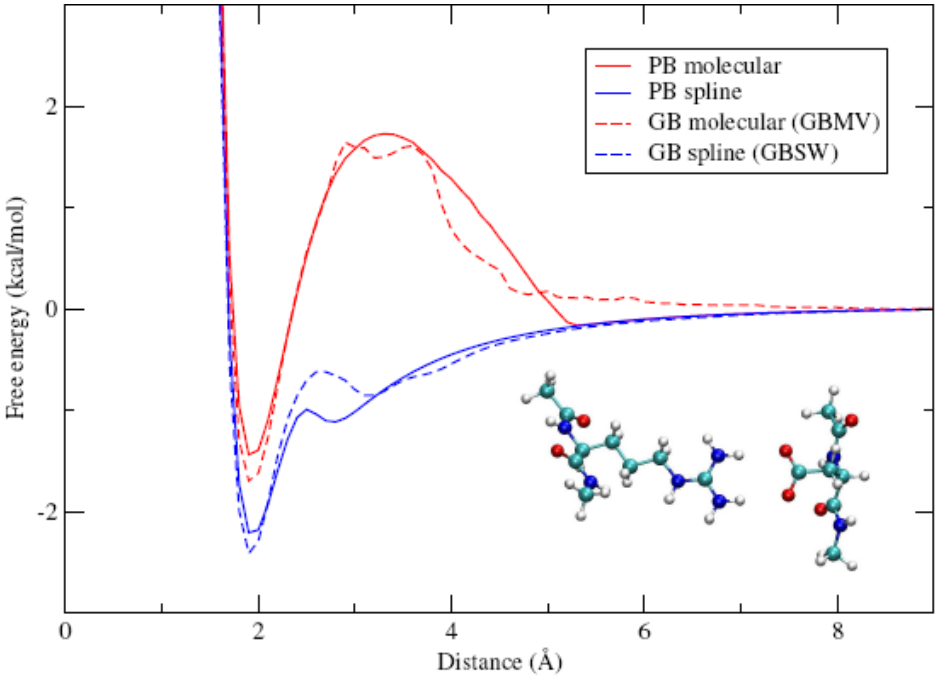

Figure 5.5: PMF for the illustrated arginine-aspartate salt bridge. Distances are measured between the hydrogen and oxygen atoms participating in the bond.

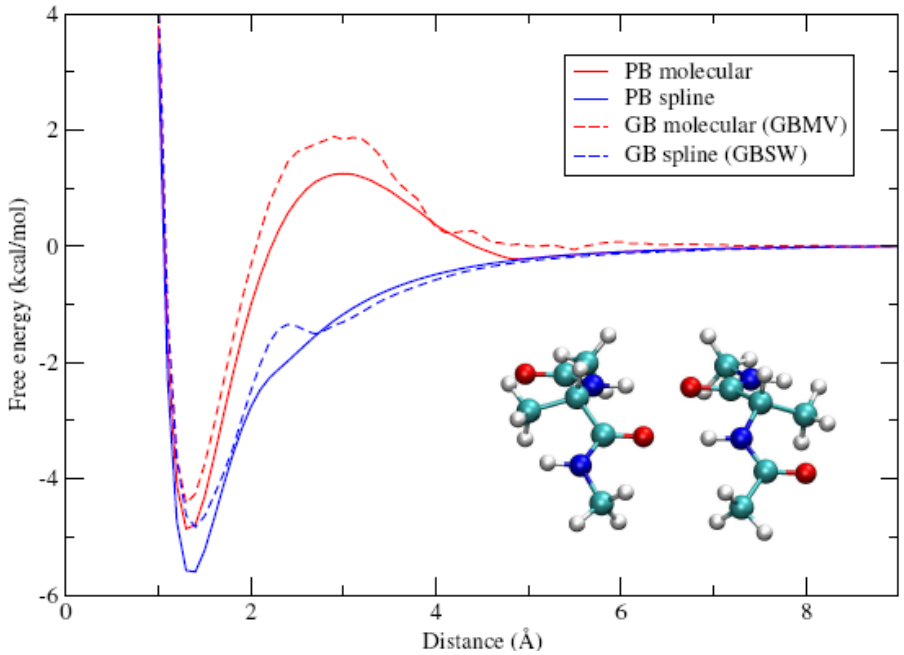

Figure 5.6: PMF for the hydrogen bonding between alanine backbones atoms. Distances are measured between the hydrogen and oxygen atoms participating in the bond.

5.4 Conclusions

The introduction of atom-centered dielectric functions has been a significant advance for PB force calculations. They can be analytically defined and easily smoothed allowing for numerical stability and increased efficiency. However, this paper demonstrates that atom-centered surfaces produce large volumes of interstitial high dielectrics in globular proteins which artificially overestimate solvation energies and distort the free energy profile of non-bonded interactions such as hydrogen bonds and salt bridges. Dynamical simulations conducted using these dielectric boundaries will sample incorrect conformational ensembles. These findings suggest that although the optimal surface definition should be smooth and differentiable it should also exclude interstitial high dielectrics as the MS does. Dielectric

boundaries that address this issue, such as those proposed by Luo *et al*.[76], and Lee *et al*.[95] will be critical for further improvement of PB and GB models.

This chapter is a reprint in full of material that appeared in Limitations of atom-centered dielectric functions in implicit solvent models. Jessica M.J. Swanson, John Mongan and J. Andrew McCammon. Journal of Physical Chemistry B, 109(31) 14769-14772, August 2005. I was the primary researcher and author of this work.

Chapter Six

Optimizing the Poisson Dielectric Boundary with Explicit Solvent Forces: Lessons Learned with Atom-Centered Dielectric Functions

ABSTRACT

Implicit solvent models rely on system- and/or atom-specific parameters which have been optimized to accurately reproduce solvation energies obtained from experiments or more detailed explicit solvent simulations. Models based on the Poisson or Poisson-Boltzmann equation are particularly sensitive to the nature and location of the boundary which separates the low dielectric solute from the high dielectric solvent. Here we present a novel method for optimizing the solute radii, which define the dielectric boundary, based on explicit solvent forces and charging free energies. We use this method to optimize radii for protein systems defined by AMBER *ff*99 partial charges and a spline-smoothed solute surface. The spline-smoothed surface is an atom-centered dielectric function which was introduced by Im and co-workers for stable and efficient force calculations. We explore the relative performance of radii optimized with forces alone and those optimized with forces and energies. Our radii reproduce the explicit solvent forces and energies more accurately than four other parameter sets commonly used in conjunction with the AMBER force field, each of which has been appropriately scaled for use with spline-smoothed surfaces. Finally, we show that spline-smoothed surfaces show surprising accuracy for small, compact systems, but may have limitations for highly-solvated protein systems. The presented optimization method is efficient

and applicable to any system with explicit solvent parameters. It can be used to determine the optimal continuum parameters when experimental solvation energies are unavailable and the computational costs of explicit solvent charging free energies are prohibitive.

6.1 Introduction

Properly accounting for solvation effects is a long-standing and constantly evolving challenge in computational biophysics. Both the dynamic and thermodynamic properties of a solvated molecule are strongly influenced by the microscopic structure and organization of the water that surrounds it. Predicting solvation free energies and forces accurately across different chemical architectures requires a formalism that collectively accounts for electrostatic, nonelectrostatic, and specific (e.g. hydrogen bonding) solute-solvent interactions. Microscopic formalisms treat these interactions explicitly with an atomistic representation of water. Macroscopic formalisms offer a less physically accurate but more computationally efficient approach: they replace individual molecular interactions with an implicit representation of water, most often as a linearly polarizable continuum.[23] The efficiency of implicit solvent models makes them ideal for large systems and computationally expensive problems such as extensive conformational sampling or high throughput analyses. The lack of physical accuracy can be compensated by parameterization, e.g. fitting Born radii to solvation free energies,[24, 25] which generally results in more accurate quantitative predictions for specific solutes but questionable transferability between solutes and solute conformations.

Most implicit solvent models divide the solvation free energy into polar and nonpolar contributions, $\Delta G_{\text{solv}} = \Delta G^{\text{p}}_{\text{solv}} + \Delta G^{\text{np}}_{\text{solv}}$. This division is rigorously based on a thermodynamic

cycle (Figure 6.1A) in which the solute charges are turned off in vacuum ($-\Delta G_{\text{vac}}^{\text{chg}}$), the neutral solute cavity is solvated ($\Delta G_{\text{solv}}^{\text{np}}$), and the solute is recharged in the solvent environment ($\Delta G_{\text{aq}}^{\text{chg}}$). The nonpolar contribution, due to hydrophobic and dispersion interactions, is commonly, though inadequately[30, 31], treated with a solvent accessible surface area (SASA) model.[16, 17, 32] Improvements on this model include independently accounting for dispersion contributions[30] and, more recently, including a term proportional to the solvent volume in addition that proportional to surface area.[31]

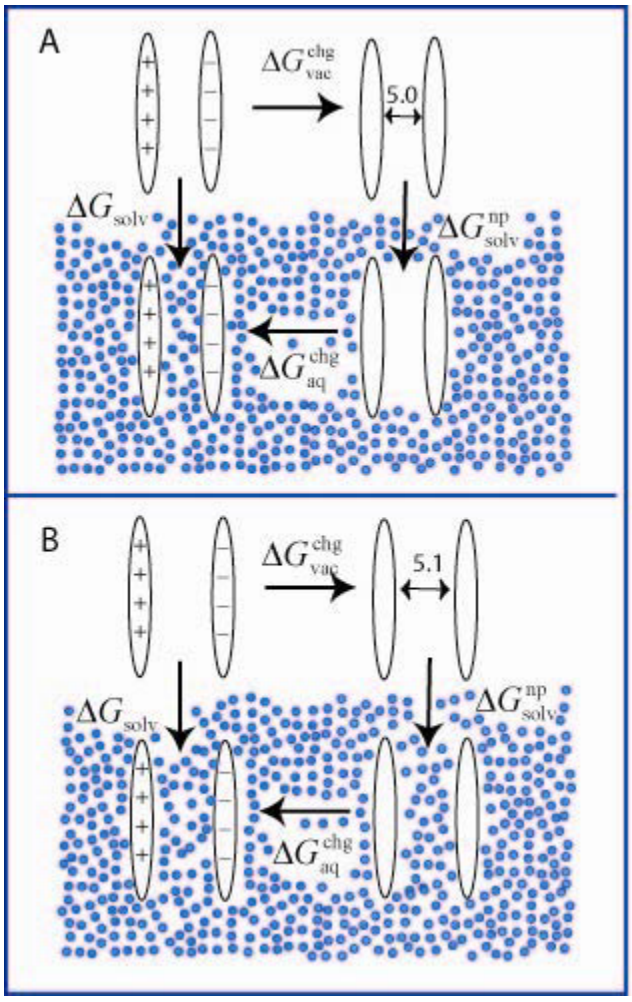

Figure 6.1: Solvation thermondynamic cycle in which solvation occurs in three path-dependent phases and according to which the total solvation free energy, ΔG_{solv} , is decomposed into polar, $\Delta G_{solv} = \Delta G_{solv}^{p} + \Delta G_{solv}^{np}$, and nonpolar, ΔG_{solv}^{np} , contributions. (A) Depicts to parallel plates separated 5.0 Å. The hydrophobic forces for the neutral plates are greater than the attractive dispersion forces such that the region between the neutral plates is void of water. (B) The same plates separated 5.1 Å with dispersion forces that are stronger than the hydrophobic forces such that the region between the neutral plates is fully solvated.

The polar contribution is the difference between charging the system in vacuum and the solvent environment $\Delta G_{solv}^{p} = \Delta G_{aq}^{chg} - \Delta G_{vac}^{chg}$. It can be approximated in a number of ways,[23] but we will focus on methods which solve the Poisson equation,[13, 102, 103]

$$\nabla \cdot [\varepsilon(r) \nabla \phi(r)] = -4\pi\rho(r),$$

for the electrostatic potential of the system, $\phi(r)$, given a position-dependent dielectric coefficient, $\varepsilon(r)$, and the solute charge distribution, $\rho(r)$. Solving the electrostatic potential of a solute in vacuum, ϕ_{vac}, and in the solvent environment, ϕ_{aq}, yields the polar solvation free energy,

$$\Delta G^{\text{p}}_{\text{solv}} = -1/2 \int \rho(r) \left[\phi_{\text{aq}}(r) - \phi_{\text{vac}}(r) \right] dr \, .$$

The parameters which go into both the polar and nonpolar approximations are generally optimized against solvation energies from experiment[68] or explicit solvent simulations[31, 104]. The polar parameters in Poisson calculations include the solute charge distribution and a spatially-varying dielectric coefficient. Energies and forces obtained from the Poisson equation are particularly sensitive to the nature and location of the boundary which separates the low dielectric solute ($\varepsilon = 1 - 20$) from the high dielectric solvent ($\varepsilon \cong 80$).

Traditionally, the dielectric boundary has been an abrupt transition from low to high dielectric values at the molecular surface, as defined by Lee and Richards[26] or Connolly[27]. Much work has gone in to optimizing solute radii for molecular surfaces such that implicit solvent calculations reproduce accurate solvation energies.[68, 69, 85] Although molecular surfaces work well for most purposes involving static biomolecular structures, they are problematic for continuum solvent dynamics simulations because they are computationally costly and result in unstable forces which vary rapidly with changes in molecular conformation.[28] These problems can be avoided by defining a smooth dielectric transition at the solute surface with overlapping atom-centered polynomial or Gaussian functions.[28, 29] One such surface, the spline-smoothed surface introduced by Im et al.[29], shows great potential for continuum dynamics due to its smoothness which improves force calculation efficiency and numerical stability. Changing the nature of the dielectric transition, however, also changes its optimal location as defined by the solute radii. The spline-smoothed surface not only has a

gradual dielectric transition, it is also fundamentally different from the molecular surface; it defines interstitial spaces between atoms as high dielectric solvent as opposed to low dielectric solute. For these reasons, radii appropriate for an abrupt transition at either the van der Waals or molecular surface are not appropriate for spline-smoothed surfaces[71]; in fact, they result in solvation energies and forces which are overestimated by 10-40%.[78, 99]

Here we present optimized Poisson radii for protein systems defined by spline-smoothed surfaces and AMBER *ff*99 partial charges.[105] We compare our optimized radii to four other commonly-used Poisson parameter sets,[68, 90, 105, 106] each of which has been properly scaled for spline-smoothed surfaces. We explore the effects of atom-centered surface definitions on radii optimizations and discuss their limitations in protein systems. More importantly, we present an efficient method for optimizing the Poisson dielectric boundary based on explicit solvent simulations. Following the work of Wagoner and Baker[31, 107], we demonstrate how mean atomic solvation forces can be used in addition to molecular solvation energies to optimize the solute radii for a given partial charge set and surface definition. This approach is closely related to force-matching techniques used in multiscale models[108, 109] and provides atomically-detailed information about the performance of implicit solvent models. It is significantly more efficient than previous optimization schemes which have relied on either experimental solvation free energies[68, 69, 106, 110, 111] or explicit solvent charging free energies.[70, 71, 85, 99] Finally, this parameterization approach can be applied to large, unusual, or highly charged solutes for which solvation energies are unavailable and charging free energies are computationally expensive.

Wagoner and Baker used nonpolar solvation forces from explicit solvent simulations to successfully optimize nonpolar implicit solvent parameters which are useful in the prediction of solvation forces and energies.[112] In the present work, we compare Poisson radii that have been optimized with forces alone to those optimized with forces and energies, showing that the

latter are superior when both forces and energies are important. Combining these polar parameters with the recently developed and complimentary nonpolar parameters[112] should greatly increase the accuracy of the Poisson-based implicit solvent framework for protein systems.

6.2 Theory

As described by Roux and Simonson,[23] the statistical mechanics formalism for implicit solvent models begins with the potential energy of a solute in an aqueous medium decomposed into

$$U_{TOT}(X,Y) = U(X) + U(Y) + U(X,Y), \tag{6.1}$$

where X and Y are the solute and solvent degrees of freedom, respectively, $U(X)$ is the intramolecular solute potential, $U(Y)$ is the solvent-solvent potential, and $\mathrm{U(X,Y)}$ is the solute-solvent potential. The system's free energy is

$$-\beta G(X,Y) = \ln\left(\int dX \int dY e^{-\beta[U(X)+U(Y)+U(X,Y)]}\right). \tag{6.2}$$

where $\beta=(k_B T)^{-1}$ is the inverse thermal energy. Equation. 6.2 can be used to derive a potential of mean force, $W(X)$, in the usual way[3, 23]:

$$e^{-\beta W(X)} = \frac{\int e^{-\beta[U(X)+U(Y)+U(X,Y)]} dY}{\int e^{-\beta[U(Y)]} dY}, \tag{6.3}$$

Note that the integral of $e^{-\beta W(X)}$ over the solute degrees of freedom gives $e^{-\beta \Delta G_{solv}}$ where ΔG_{solv} is the solvation free energy. Choosing the unsolvated solute as the reference state, the solvation free energy of a particular solute conformation X is $\Delta W(X) = W(X) - U(X)$. To

decompose the solvation free energy into polar and nonpolar terms, one must first express the solute-solvent potential as a sum of polar and nonpolar contributions[23], $U(X,Y) = U^{\mathrm{p}}(X,Y) + U^{\mathrm{np}}(X,Y)$, such that

$$\Delta W(X) = \Delta W^{\mathrm{np}}(X) + \Delta W^{\mathrm{p}}(X), \tag{6.4}$$

$$e^{-\beta\Delta W^{\mathrm{np}}(X)} = \frac{\int e^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)]}dY}{\int e^{-\beta[U(Y)]}dY}, \text{ and} \tag{6.5}$$

$$e^{-\beta\Delta W^{\mathrm{p}}(X)} = \frac{\int e^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)+U^{\mathrm{p}}(X,Y)]}dY}{\int e^{-\beta[U(Y)+U^{\mathrm{np}}(X,Y)]}dY}. \tag{6.6}$$

These contributions can be calculated in two successive stages of a free energy perturbation or thermodynamic integration[6] according to

$$\Delta W^{\mathrm{np}}(X) = \int_0^1 d\lambda_1 \left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \lambda_1} \right\rangle_{(\lambda_2=0)} \text{ and} \tag{6.7}$$

$$\Delta W^{\mathrm{p}}(X) = \int_0^1 d\lambda_2 \left\langle \frac{\partial U^{\mathrm{p}}(X,Y)}{\partial \lambda_2} \right\rangle_{(\lambda_1=1)}, \tag{6.8}$$

where λ_1 and λ_2 are coupling parameters that scale the nonpolar and polar solvent-solute interactions from 0 (off) to 1 (on). The polar contribution is therefore a 'charging free energy', but is often referred to as the polar solvation free energy. The mean forces can be obtained by differentiation of $W(X)$ with respect to the solute coordinates[23], X,

$$\left\langle F^{\mathrm{np}}(X)\right\rangle = -\frac{\partial W^{\mathrm{np}}(X)}{\partial X} = -\left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{np}} \quad \text{and} \qquad 6.9$$

$$\left\langle F^{\mathrm{p}}(X)\right\rangle = -\frac{\partial W^{\mathrm{p}}(X)}{\partial X} = -\left\langle \frac{\partial U^{\mathrm{p}}(X,Y)}{\partial \mathrm{X}} + \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{p}} + \left\langle \frac{\partial U^{\mathrm{np}}(X,Y)}{\partial \mathrm{X}} \right\rangle_{\mathrm{np}}, \qquad 6.10$$

where $\langle \ldots \rangle_{\mathrm{np}}$ and $\langle \ldots \rangle_{\mathrm{p}}$ denote the ensemble averages with nonpolar only and full (nonpolar and polar) solute-solvent interactions, respectively. In this work we will focus on polar contributions, $\Delta W^{\mathrm{p}}(X) \equiv \Delta G^{\mathrm{p}}_{\mathrm{solv}}$ and $\left\langle F^{\mathrm{p}}(X)\right\rangle \equiv \bar{F}^{\mathrm{p}}$. A similar treatment of nonpolar contributions has been presented by Wagoner et al.[107]

It is important to emphasize that the decomposition expressed in equations. 6.4-6.6 describes the process of solvation as the second and third phases of the path dependent thermodynamic cycle shown in Figure 6.1. The coupling that exists between polar and nonpolar (i.e., hydrophobic and dispersion) interactions (see Dzubiella et al.[113, 114] for discussion) is not explicitly defined in this thermodynamic cycle, but is implicitly included in the charging phase, $\Delta G^{\mathrm{chg}}_{\mathrm{aq}}$. The magnitude of this coupling is extremely sensitive to the molecular makeup and conformation of the system. Consider, as an extreme example, a very small conformational change in a system composed of two parallel plates separated by fixed distances of 5.0 Å and 5.1 Å as depicted in Figure 6.1. The dispersion forces during phase 2, $\Delta G^{\mathrm{np}}_{\mathrm{solv}}$, will be much greater for system (B) than system (A). Similarly, the change in the dispersion forces during phase 3, $\Delta G^{\mathrm{chg}}_{\mathrm{aq}}$, will be significantly less for system (B) than system (A). The polar and nonpolar parameters that would reproduce these phenomena would be very different for the two systems, despite their similarity. This illustrates that it is not possible to define a set of independent polar and nonpolar parameters that will be accurate across widely disparate

molecular systems or for all system conformations. Fortunately, biomolecular systems are predominantly composed of similar subunits (e.g. amino and nucleic acids) with significant solute-solvent dispersion interactions; thus, conformational changes will rarely result in such extreme changes in the solvation structure and density surrounding biomolecules.

It has been shown that optimizing complimentary polar and nonpolar parameters in a system- and/or atom-dependent manner can result in fairly accurate, though sometimes system-dependent, solvation frameworks.[68, 110, 111] The SMx models, for example, can predict solvation free energies for neutral molecules with errors less than 1 kcal/mol and have shown recent success in treating charged solutes and ions with the addition of an explicit water molecule.[110, 111] The goal of the present work is to fit the polar solvation energies and forces calculated with the Poisson equation to those calculated with explicit solvent simulations and thereby optimize a set of radii to define the Poisson spline-smoothed dielectric boundary for protein systems. Combining these polar parameters with the nonpolar parameters recently developed by Wagoner and Baker[112] should greatly increase the accuracy of the Poisson-based implicit solvent framework for protein systems.

6.3 Methods

The force-based optimizations involve three steps, described in more detail below. First, explicit solvent forces are collected for each solute in the training set in one or more fixed conformations. Next, the continuum calculations are set up such that energies and forces are well converged, i.e. independent of grid resolution, boundary conditions, etc. Finally, the Poisson atomic radii are optimized to reproduce the explicit solvent forces. Optimizations using both energies and forces additionally require calculating explicit solvent charging free energies with thermodynamic integration or free energy perturbation simulations as previously reported.[99]

6.3.1 Training Sets

To test the limitations of our methodology, we began by optimizing *each* atomic radius in three non-zwiterionic N-acetyl-X-N'-methylamide dipeptides, where X represents alanine with phi/psi angles 180°/180° (ala1), alanine with phi/psi angles -60°/-40° (ala2), and serine with phi/psi angles 180°/180° (ser). As previously described,[99] using simple PARSE- or Bondi- or AMBER vdW-like atom types limits the accuracy of our parameter set due to significant differences in the solvation structures, energies, and forces for atoms within these type groups. Therefore, we used 31 'atom types' or radius groups (Table 6.2) which were previously identified as having similar chemical signatures and surrounding solvation structures according to explicit solvent charge distribution functions.[99] The importance each group was further tested based on final optimization fitness values and unique optimal solutions.

The training and test sets are summarized in Table 6.1. Two model systems were used in our force-based optimizations. The first was 2 conformations of intestinal fatty acid binding protein (IFABP) representing highly populated conformational clusters from a prior fully-equilibrated explicit solvent simulation.[78] The second was two short polypeptide chains: Trpcage (1l2y)[92] and the C-terminal fragment (residues 41-56) of protein G (2gb1)[93], hereafter called G-peptide. The energy and force-based optimizations additionally used the explicit solvent charging free energies of 20 N-acetyl-X-N'-methylamide dipeptides where X represents each of the amino acids (phi/psi angles 180°/180°) and 7 polyalanine peptides with common secondary structures (3 beta turns, 2 helices, and 2 beta hairpins). Each of the polyalanine peptides was modeled from fragments of lysozyme (pdb code 1ati) or crambin (pdb code 1ejg) which were mutated to polyalanine and capped with neutral blocking groups. The specific peptide conformations and explicit solvent charging free energies were previously reported.[99]

Table 6.1: Solutes used to training and test sets

Training Sets			Additional Test Sets	
Solute[s]	**Quantity**	**Trained Radii**	**Solute[s]**	**Quantity**
IFABP conformations 1 & 2	forces	force optimized 1	IFABP conformations 3-8	forces
Trpcage & G-peptide	forces	force optimized 2	20 dipeptides conformation 2	energies
Trpcage & G-peptide	forces	force-energy optimized	poly-alanine peptides 8-12	energies
20 dipeptides conformation 1	energies	force-energy optimized	4 protein-like peptides	energies
poly-alanine peptides 1-7	energies	force-energy optimized		

6.3.2 Testing Sets

To see if the optimized radii are transferable to molecules outside of the training set, we used the forces and solvation energies of the opposing test systems as well as the solvation energies of the following molecules: Trpcage, G-peptide, 2 additional protein-like polypeptides, 5 new polyalanine peptides, and 20 amino acid dipeptides in a new side chain and backbone conformation (phi/psi angles -60°/-40°).[99] We also used the forces of IFABP in 6 new conformations. The relative RMSD values for all 8 IFABP structures ranged from 1.2 – 2.2 Å.[78] All of the aforementioned systems were used to compare the optimized radii to 4 other continuum parameter sets: AMBER *ff*99 partial charges combined with *ff*99 van der Waals (vdW) radii,[105] Bondi radii,[90] and a set of radii recently published by Luo et al.,[106] as well as the PARSE parameters (charges and radii).[68] Each of the comparison parameter sets was scaled for use with spline-smoothed surfaces as described below.

6.3.3 Explicit Solvent Forces

As shown in equation. 6.10, the polar solvation forces are calculated by subtracting the nonpolar solvation forces from the total forces, ($\overline{F}^{\,p} = \overline{F}^{\,np+p} - \overline{F}^{\,np}$). The nonpolar forces are averaged from an ensemble generated with only nonpolar (vdW) solute-solvent interactions, while the total forces are averaged from an ensemble generated with both polar (electrostatic) and nonpolar solvent-solute interactions Thus, polar forces require two simulations, one with

all solute-solvent interactions and one with only nonpolar interactions (i.e., solute charges set to zero).

The IFABP simulations were performed, as previously reported,[31] with the AMBER *ff*99[105] force field using AMBER 7 software.[115] Eight conformations were obtained by clustering a 2.5 ns simulation which started from an NMR structure (pdb code 1ael)[116] and was run in TIP3P water in the presence of ~160 mM NaCl at 300K and 1 atm pressure. Each conformation was then constrained via belly dynamics and simulated for 1.05 ns with SHAKE enabled 2 fs time steps under isobaric-isothermal conditions. The forces for conformations 3-6 were averaged from 250 snapshots taken every 4 ps from the last 1 ns of simulation. The simulations of conformations 1 and 2 were extended an additional 4 ns and forces were averaged from 1000 snapshots taken every 4 ps from the last 4 ns of simulation. The average squared residual for forces from 1 and 5 ns was 0.08556 kcal/mol Å.

Trpcage, G-peptide, and the three dipeptides (ala1, ala2, ser) were simulated with the AMBER *ff*99 converted to CHARMM format and the CHARMM software (version 31a1).[79] The solutes were protonated and solvated in a sphere of TIP3P water molecules that extended 10.0Å beyond the solute resulting in 3-4 hydration shells. The spherical solvent boundary potential including Kirkwood's multipolar expansion reaction field was used to approximate the influence of bulk water beyond the explicit water sphere.[84] This model was previously used in the free energy perturbation simulations to calculate charging free energies[99] because it alleviates difficulties that result from perturbing charged systems with periodic boundary conditions. It was chosen here for consistency with the molecular charging free energies. It has been shown to give reliable results for proteins, nucleic acids, and small molecules.[23, 73, 85] All solute atoms were restrained to their original coordinates. The solvent was first energy minimized with 50 steps of steepest descent followed by 1000 steps of the ABNR method and then equilibrated for 200 ps. Langevin dynamics were employed at constant temperature (300

K) using SHAKE enabled 2 fs time steps, infinite cutoffs for nonbonded interactions, and a friction constant corresponding to a relaxation time of 5 ps applied to water oxygen atoms.

The peptide forces were averaged from snapshots taken every 0.2 ps over 500 ps of pre-equilibrated simulation for a total of 2500 conformations. Comparing forces from the first 200 ps with those from the full 500 ps resulted in a root mean square deviation ($\mathrm{RMSD\_}\overline{F}^{\mathrm{p}}$) of 0.187 kcal/mol Å and an average relative error of 0.0884 kcal/mol Å, demonstrating reasonable convergence after only 200 ps. Comparing duplicate 500 ps simulations of the dipeptides dropped $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}}$ to 0.05-0.09 kcal/mol Å, which was considered sufficient convergence. Since the IFABP and peptide systems were being used in independent optimizations, any small differences that may have resulted from different simulation protocols were not a major concern.

6.3.4 Implicit Solvent Forces

All implicit solvent energies and forces were calculated with the Adaptive Poisson-Boltzmann Solver (APBS; http://apbs.sf.net/) version 0.4.0.[49] APBS employs the analytical force evaluation method of Im. et al.[29] In order to maintain numerical stability, the transition between high and low dielectric regions must be gradual; thus, the solute surface proposed by Im et al. and used in APBS is defined by atom-centered cubic polynomials. The resulting total solvation force is composed of three terms: a reaction field force due to interaction of the solute charge with the total electric field, a dielectric boundary force due to spatial variation in the dielectric function, and an ionic boundary force for systems with ions. The last term was zero in this study since we used zero bulk ionic strength. The abruptness of the transition between the solute and solvent values in these spline-smoothed surfaces is controlled by the spline window width, *w,* a user defined parameter in APBS. We optimized our radii with the

recommended value, w = 0.3 Å. We additionally scaled these radii for spline windows 0.2 Å ≤ w ≤ 1.0 Å with both force and force-energy based genetic algorithm optimizations (as described below). We solved the Poisson equation at a temperature of 300 K with the multiple Debye-Hückel sphere boundary condition, and cubic B-spline charge discretization. We used the experimental solvent dielectric of 78.4. Calculations run with the PARSE parameters used a solute dielectric of 2.0 in agreement with their original development.[68] All other calculations used a solute dielectric of 1.0, which is appropriate for calculations which explicitly sample solute conformations and desirable for consistency with nonpolarizable force fields. It is important to ensure that the chosen continuum parameters, particularly the grid parameters, are appropriate for energetic and force convergence. We did so by comparing the energies and forces calculated with successively-finer grids (down to 0.10 Å), requiring the relative errors to be less than 0.5%. Our final grids had resolutions of either 0.2 Å or 0.25 Å depending on the solute, and dimensions which extended beyond the solute by 35% of its length in each dimension.

6.3.5 Genetic Algorithm Optimizations

We used a genetic algorithm to optimize the radii from various starting values (AMBER *ff*99 vdW, Bondi, and our previously published smooth boundary radii[99]) to their final values. Populations of 50 solutions were run for 10 to 100 generations. The initial populations were generated from a uniform distribution which varied from 0.1 Å to 0.3 Å around the starting values, depending on the optimization. The fitness of each solution was evaluated and the next population was generated via selection, crossover and mutation. Selections were based on the Stochastic Universal Sampling algorithm,[88] which selects solutions based on a probability proportional to their fitness,

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}, \tag{6.11}$$

where p_i is the probability that solution i will be selected and f_i is its fitness defined by,

$$f_i = \left(1 + \mathrm{RMSD\_}\overline{F}_i^{\,p}\right)^{-1}, \tag{6.12}$$

where $\mathrm{RMSD\_}\overline{F}^{\,p}$ is the standard deviation of all 3N force components in an N-atom system for the radii solution set, i. The fitness scores range from 0 (poor) to 1 (perfect). A uniform crossover process in which two solutions from the previous generation were randomly distributed to two new solutions was applied to either 80% or 90% of the population, while mutations selected from a Gaussian distribution with standard deviations which ranged from ±0.025 Å to ±0.1 Å were applied to anywhere from 5% to 30%. The various different crossover and mutation values were tested to ensure that the final solutions were not an artifact of limiting genetic algorithm run parameters. This process was repeated until a desired fitness is reached or the maximum number of generations exceeded.

As we will demonstrate, accurately fit solvation forces do not necessarily ensure accurate solvation energies. Thus, we ran a second set of optimizations that incorporated the solvation energies of 20 amino acid dipeptides and 7 polyalanine peptides as well as the Trpcage and G-peptide solvation forces into the fitness function according to,

$$f_i = \left(1 + 1/2(\mathrm{RMSD\_}\overline{F}_i^{\,p} + \mathrm{AAE\_}\Delta G_{p\_i}^{solv})\right)^{-1}, \tag{6.13}$$

where $\mathrm{RMSD\_}\overline{F}^{\,p}$ is the defined as before and $\mathrm{AAE\_}\Delta G_{p\_i}^{solv}$ is the average absolute error for the 27 solutes' polar solvation energies for solution set, i.

6.3.6 Scaling Radii for Spline-Smoothed Surfaces

Radii which work well with molecular surfaces will overestimate solvation energies and forces by 10-40% when used with spline-smoothed surfaces.[78, 99] Therefore, a direct comparison between the four comparison parameter sets and our optimized radii was not possible. Instead, the optimal scaling factor, x, for each parameter set was determined according to $R_{new} = R_{f/fe} + x$, where R_{new} is the new radius and $R_{f/fe}$ is the force or force-energy-optimized radius from Table 6.2. The scaling function was chosen over a function used in previous optimizations, $R_{new} = x(R_{f/fe} + w)$, because it resulted in slightly higher final fitness values and converged in one to three fewer generations. Both were used, each with the respective optimized radii. We also scaled our radii for different spline window widths according to, $R_w = R_{0.3} + x$, where R_w is the new radius for w = 0.2 to 1.0 Å and $R_{0.3}$ is the w = 0.3 Å force or force-energy-optimized radius from Table 6.2.

6.4 Results

6.4.1 Dipeptide Test Case

It is important know what degree of accuracy one can expect from the presented methodology, i.e., the limitations of force and force-energy based optimizations. To address this question, we generated "perfect (Poisson) radii"[117] by optimizing every atom's radius in three simple dipeptides, alanine in two different conformations (ala1, ala2) and serine (ser). We used both the force and the force-energy based approaches (equations. 6.11 and 6.12, respectively). To ensure that the results were independent of the genetic algorithm run and parameters, the optimizations were repeated 30 times varying the initial population distribution from ± 0.1 to 0.4 Å, the mutation distribution from +/- 0.025 to 0.1 Å, and the mutation ratio from 0.05 to 0.4 (i.e., mutations were applied to 5% to 40% of the solutions in the next generation). The resulting force and polar solvation energy statistics (shown in Table 6.1) demonstrate errors that are larger than the inherent errors in the explicit and implicit

calculations. The explicit solvation energy and force errors were $\Delta\Delta G_{\mathrm{p}}^{\mathrm{solv}} \approx$ 0.02-0.17 kcal/mol and $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}} \approx$ 0.05-0.09 kcal/mol Å, respectively, while the continuum were $\Delta\Delta G_{\mathrm{p}}^{\mathrm{solv}} \approx$ 0.03 kcal/mol and $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}} \approx$ 0.05 kcal/mol Å.

Table 6.1: Individually-optimized radii solvation energy and force results for ala1, ala2, and ser.

	force optimized ala1 radii				force-energy optimized ala1 radii			
	$\Delta\Delta G_{solv}$ [a]		**RMSD_F** [b]		**$\Delta\Delta G_{solv}$** [a]		**RMSD_F** [b]	
ala1	-1.753	(0.290)	0.147	(0.009)	0.000	(0.005)	0.244	(0.025)
ala2	-1.594	(0.336)	0.358	(0.013)	0.152	(0.073)	0.432	(0.026)
ser	-3.314	(0.247)	0.344	(0.020)	-1.633	(0.164)	0.333	(0.046)
	force optimized ala2 radii				force-energy optimized ala2 radii			
	$\Delta\Delta G_{solv}$ [a]		**RMSD_F** [b]		**$\Delta\Delta G_{solv}$** [a]		**RMSD_F** [b]	
ala1	-1.865	(0.359)	0.302	(0.009)	-0.054	(0.103)	0.287	(0.020)
ala2	-1.906	(0.265)	0.160	(0.003)	-0.001	(0.005)	0.322	(0.046)
ser	-3.810	(0.391)	0.412	(0.011)	-1.820	(0.168)	0.342	(0.027)
	force optimized ser radii				force-energy optimized ser radii			
	$\Delta\Delta G_{solv}$ [a]		**RMSD_F** [b]		**$\Delta\Delta G_{solv}$** [a]		**RMSD_F** [b]	
ala1	-0.603	(0.210)	0.291	(0.036)	1.476	(0.171)	0.422	(0.027)
ala2	-0.450	(0.209)	0.390	(0.009)	1.603	(0.148)	0.616	(0.024)
ser	-2.420	(0.233)	0.200	(0.009)	0.000	(0.002)	0.418	(0.025)
	force optimized radii				force-energy optimized radii			
	$\Delta\Delta G_{solv}$ [c]		**RMSD_F** [d]		**$\Delta\Delta G_{solv}$** [c]		**RMSD_F** [d]	
ala1	0.376		0.262		0.217		0.265	
ala2	0.569		0.444		0.449		0.476	
ser	-0.831		0.560		-0.842		0.598	

a) The *average* solvation energy deviation [kcal/mol] and (b) *average* force RMSD [kcal/mol Å] for three dipeptides using individually force and force-energy-optimized radii. Statistics collected from 30 independent genetic algorithm optimizations. Standard deviations given in parentheses. c) The solvation energy deviation [kcal/mol] and (d) force RMSD [kcal/mol Å] for the three dipeptides using the force and force-energy-optimized radii presented in Table 6.2.

There are five key points to take away from this example. First, the genetic algorithm solution phase space is highly frustrated. The majority of optimizations generated similar but unique solutions, indicating that there are multiple solutions with similar fitness even for simple

molecules with independently fit radii. Second, the explicit solvent forces can not be perfectly reproduced. Whether this discrepancy is caused by the limited dielectric definition (i.e., the union of spline-smoothed spheres) or limitations in Poisson equation itself (i.e., the assumption of linear and local solvent response) is not clear. It is our expectation that all of these limitations factor in, but that the limited dielectric definition is a major source of error. This puts the limit of accuracy for implicit solvent forces calculated with this surface at $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}}$ $\gtrsim$ 0.15 kcal/mol Å. Third, each solute molecule is unique. When the force-optimized radii for serine are used on alanine or vice-versa, the average $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}}$ increases from 0.15–0.20 to 0.29-0.41 kcal/mol Å. Similarly, when the force-energy-optimized radii are switched, the average $\Delta\Delta G_{\mathrm{p}}^{\mathrm{solv}}$ increases from between −0.001 and 0.000 kcal/mol to between −1.820 and −0.054 kcal/mol. This increase indicates that neighboring atoms change the optimal continuum radius for any given atom type. Third, the solute's conformation is important. When the radii are switched for ala1 and ala2, we observe a similar, though less dramatic, increase in the average $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}}$ to between 0.30 and 0.36 kcal/mol Å and the average $\Delta\Delta G_{\mathrm{p}}^{\mathrm{solv}}$ to between −0.054 and 0.152 kcal/mol. Although the goal of optimizing radii with multiple solutes and conformations is to find a solution set which is transferable, there is no guarantee that they will be more accurate than those individually optimized for such similar solutes. Thus, one can expect the limit of accuracy of force-optimized radii used on varying conformations, across unique amino acid architectures to be $\mathrm{RMSD\_}\overline{F}^{\mathrm{p}} \approx 0.29$ kcal/mol Å, and that of force-energy-optimized radii to be $\Delta\Delta G_{\mathrm{p}}^{\mathrm{solv}} \approx 0.05$ kcal/mol. Our optimized radii show errors just above these limits for both forces and solvation energies (Table 6.1).

As observed by Wagoner and Baker for alkane solvation energies[107], better-fit forces do not necessarily translate into better fit solvation energies. In fact, the solvation energies are

overestimated by the individually force-optimized radii by between 0.45 to 3.81 kcal/mol, which is 0.073 to 3.433 kcal/mol larger than solvation energy deviations of the force-optimized radii presented in Table 6.2. This apparent contradiction to the theoretical framework presented in section II is caused by fundamental differences in the implicit and explicit solvent potentials. The relationship presented in Eqn. 10 *is* exact *within* a given solvent model, and it can be used to fit the implicit potential of mean force to the explicit potential of mean force, as we are attempting to do here. However, fitting the directional derivative (i.e. slope) of the potential of mean force from one solvent model to the other does not guarantee that the change in free energy will be equally well fit. In fact, the approximate, at best, agreement between implicit and explicit solvent potentials of mean force for salt bridges and hydrogen bonds[100, 118] suggests that more accurate forces (slopes) would necessitate less accurate energies ($\Delta G_{\mathrm{p}}^{\mathrm{solv}}$). The results shown in tables 6.1 and 6.3 confirm this expectation.

Table 6.2: Force and force-energy-optimized radii.

atom name[a]	residues	opt. radii[b]	opt. radii.[c]
	Backbone		
C	all	2.338	2.307
O	all	1.766	1.760
N	all	2.331	2.348
CA	all except G	2.425	2.365
CA	G	2.122	2.225
CAY/CAT	ACE,NME	2.022	2.095
	Side Chains		
CB	D,E,C,H,M,F,S,T,W,Y	2.128	2.050
CB	A,R,N,Q,I,L,K,V	2.209	2.344
CG*	R,Q,I,L,K,M,T,V	2.414	2.492
CG	H,F,W,Y	2.147	2.126
CG/CD	N,Q,D	2.197	2.232
CG	E	2.195	2.275
CB/CG/CD	P	2.506	2.311
CD	R,K	2.316	2.392
CD*	I,L	2.640	2.659
CD*/CE*/CZ	H,F,W,Y	2.282	2.319
CE	M	1.940	1.976
CZ/CE	R,K	2.398	2.360
OD*/OE*	N,Q,D,E	1.729	1.729
OG*	S,T	1.956	1.979
OH	Y	1.871	1.961
NE,NH*,NZ	R,K	2.323	2.355
ND2/NE2	N,Q	2.122	2.227
ND1,NE2	H	1.927	1.986
NE1	W	1.880	1.915
SG/SD	C,M	2.406	2.420
	Hydrogens[d]		
type H	bb HN	1.244	1.283
type H	bound to N	1.228	1.225
type HO/HS	bound to O/S	0.999	0.991
type H1/HP	polar	1.905	1.886
type HC/HA	nonpolar	1.809	1.774

a) Radius groups are distinguished by AMBER atom names for all heavy atoms and by atom type for hydrogen atoms. b) Final radii [Å] from force-only optimizations. c) Final radii [Å] from force-energy optimization. d) Hydrogens specified by atom type with type 'H' divided into two groups; amide backbone 'HN' and all other N-bound hydrogens.

6.4.2 Protein Radii Optimizations

Similar to the dipeptides, the solution phase space for the force and force-energy based optimizations was highly frustrated with many similarly well-fit solutions. The various different starting radii (AMBER *ff*99 vdW, Bondi, and our previously optimized radii) had no

distinguishable effect on the final solutions, but did affect the speed of the optimizations, the latter two being the most efficient. Increasing the number of target values generally decreases the frustration in the solution phase-space. Thus using 3N atomic solvation forces as opposed to one solvation energy for an N-atom solute makes the optimization substantially more efficient and decreases the likelihood that the final solution is locally trapped far from the global minimum. The force based optimizations reached complete convergence, i.e. the exact same solution set was found for 5 consecutive generations, in a single evolution of 20 – 80 generations. The force-energy based optimizations never reached complete convergence, but converged to a minimum fitness deviation ($\Delta\text{RMSD\_}\overline{F}^{\text{p}} \leq 0.002$) in 1-2 evolutions of ~ 65 generations. Both methods are considerably faster than our previous optimizations which used solvation energies alone and took 6 or more evolutions to reach a minimum fitness convergence.

Ideally both continuum forces and energies would be faithfully predicted by a single parameter set. However, the discrepancy in implicit and explicit potentials demonstrated by the dipeptides suggests that the accuracy of one or the other must be sacrificed. This was indeed the case for the protein radii as shown in Table 6.3. The force-optimized radii performed 1.8% better on forces than the force-energy-optimized radii. The force-energy-optimized radii perform 14.1-17.2% better on polar solvation energies. Depending on the problem at hand, forces or energies may be more important; thus, both sets are presented in Table 6.2. The force-energy-optimized radii are recommended when both forces and energies are important. The force-optimized radii (column 2) are those from Trpcage and G-peptide optimizations since they perform much better than those from IFABP for reasons which we will now explore.

6.4.3 Effects of Interstitial High Dielectrics

The two model systems used in the force based optimizations resulted in strikingly different solution sets. The IFABP radii were 1-10% larger than the Trpcage and G-peptide radii, and underestimated solvation energies by 1-26% (Table 6.3). Close inspection of the IFABP atomic forces that were being significantly over and under-estimated revealed the source of this discrepancy. The significant outliers were atoms found proximal to interstitial high dielectrics (IHDs). IHDs are regions which have a higher dielectric value (i.e., are solvent) in atom-centered dielectric functions, but a low dielectric value (i.e. are solute) in molecular surfaces. They are too small for an explicit water molecule to penetrate. The dielectric constant in these regions is the same as the solvent's ($\varepsilon \approx 80$) in a vdW surface, but can take on a range of values ($1 < \varepsilon \leq 80$) in spline-smoothed surfaces. Figure 6.2 shows a series of slices through the dielectric map of IFABP. The magnitude of IHDs demonstrated by comparing the molecular surface (panel A) and the spline-smoothed surface (panel B) through a single plane of intersection, is consistent through the entire structure. Both surfaces are rendered in panels C through E to show just 3 examples of atoms proximal to IHDs with significantly overestimated continuum forces (tyr70 OH, glu63 OE1, and glu85 N). This overestimation occurs regardless of the parameter set used to define each atom's radius; it is 3.6, 5.2, and 2.9 kcal/mol Å, and 2.8, 3.9, and 3.7 kcal/mol Å for the Trpcage/G-peptide and IFABP force-optimized radii, respectively. Figure 6.3 shows the proportion of forces with large deviations increases with the volume of IHDs within 1 Å of the atom's radial boundary. The deviation cut-off values were chosen just below the $\mathrm{RMSD}\_\overline{F}^{\mathrm{p}}$ for IFABP conformations 1 and 2 using the force-optimized radii (from Table 6.2). This trend lends support to the notion that the force deviations are related to the formation of IHDs. Figure 6.4 shows that IFABP has a larger force deviations than Trpcage and G-peptide, when defined by either the force-optimized or scaled Bondi radii.

This illustrates that IHDs introduce errors in the entire protein, not just in proximal atoms. Finally, the statistics in Table 6.3 show that IFABP has larger larger $\mathrm{RMSD}\_\overline{F}^{\mathrm{p}}$ values for each of the tested parameter sets.

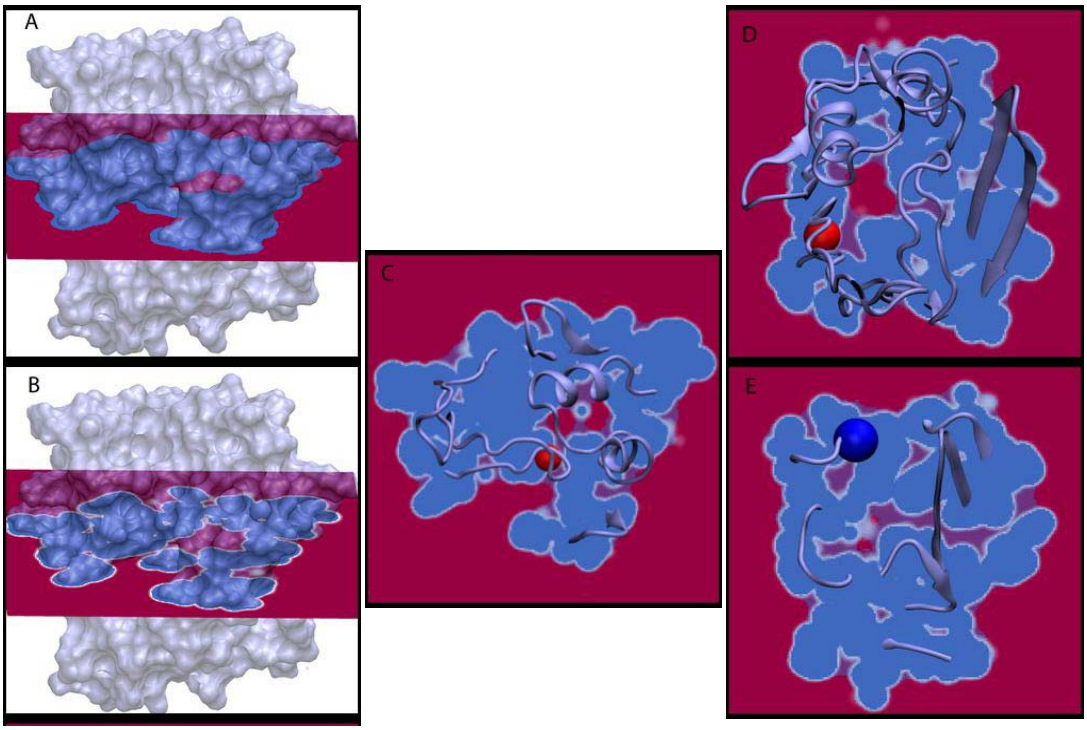

Figure 6.2: Dielectric maps of IFABP. Dielectric values of planes intersecting IFABP in which red represents high dielectric (ε=80), blue represents low dielectric (ε=1), and white is in between. Comparing the molecular surface (A) and the spline smoothed surface (B) shows the magnitude of IHD's created by the latter. A top-down view of both surfaces shows that IHD's (purple regions) are proximal to (C) tyr 70 OH, (D) glu63 OE1, and (E) glu 85 N.

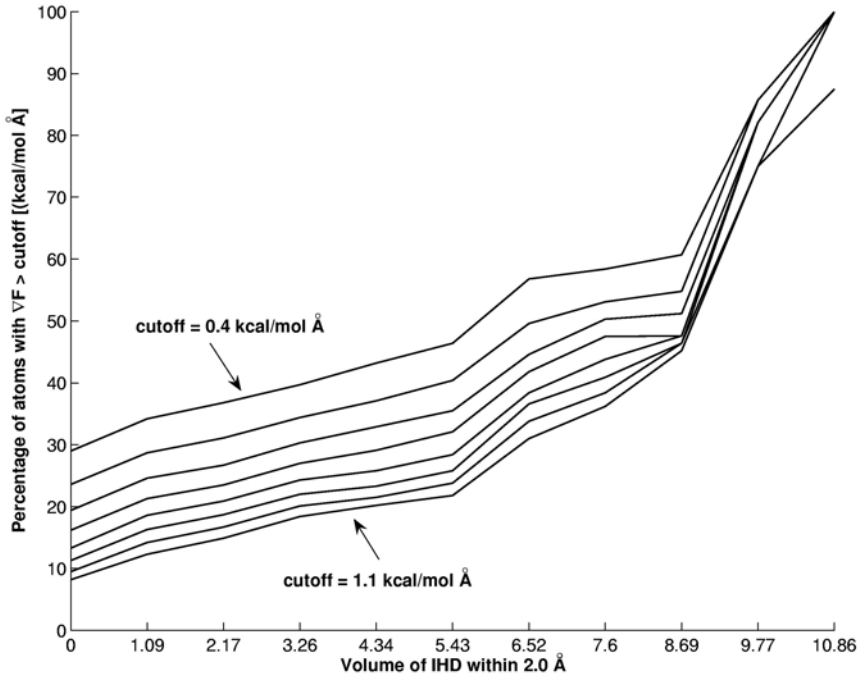

Figure 6.3: Percentage of IFABP atoms with large force deviations as a function of IHD ($\varepsilon \geq 60$) density within 1.0 Å of the atom's radial boundary. Deviation cutoff values

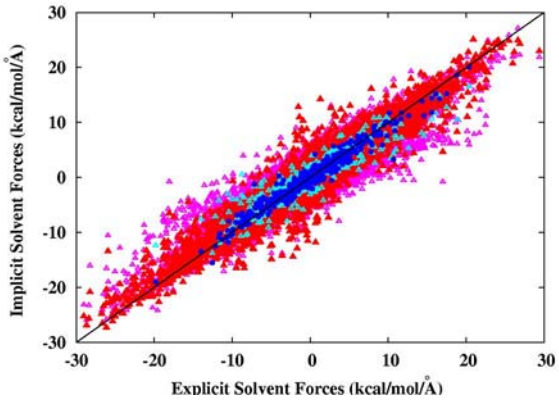

Figure 6.4: Comparison of expicit solvent and implicit solvent force components for IFABP and Trpcage/G-peptide forces showing increased distribution of the former. Bondi radii are in pink (IFAPB) and turquoise (Trpcage/G-peptide) while force optimized radii are in red (IFAPB) and dark blue(Trpcage/G-peptide).

IFABP is a highly solvated protein with a significant number of solvent-exposed residues in the “interior” fatty acid binding sites. Therefore, one might expect it to be more susceptible to the formation of IHDs than Trpcage or G-peptide, both of which have small, compact structures. These findings point to a simple explanation for the different force-optimized solution sets; the IFABP optimized radii were systematically increased to decrease the over and under-estimation of forces proximal to IHDs, whereas the Trpcage and G-peptide optimized radii were less affected. As a result, the Trpcage/G-peptide radii perform better on polar solvation energies and forces for every tested molecule other than IFABP. They will be more widely applicable to compact systems, but will not compensate for errors resulting from excessively large IHDs in globular systems. In retrospect, we believe that the formation of IHDs was also the source of differences between our previously presented abrupt boundary and spline-smoothed radii. The former, fit to molecular surfaces, could not be scaled to perform as well on spline-smoothed surfaces as the latter. Moreover, the latter overestimated solvation energies for larger peptide systems because they were fit to small peptides, those least altered by the spline-smoothed surfaces.

It has been suggested that the width of the spline smoothing window can be increased to counteract the deleterious effects of IHDs. Although it is not possible to eliminate them completely, their magnitude and dielectric value *can* be significantly decreased with larger smoothing windows. Unfortunately, this moves the dielectric boundary too far from surface atoms and creates unphysical bulges around overlapping and adjacent atoms, both of which result in underestimated solvation energies and forces. If it were possible to distinguish surface and buried atoms, then using larger radii or larger smoothing windows on buried atoms would significantly reduce the IHDs and related errors. For most globular systems, however, a clean distinction is impossible since many atoms are both exposed and buried. Thus, more involved modifications of the spline-smoothed surfaces, such as those proposed by Lu and Luo[76]. and

Lee et al.[95], seem necessary for consistently accurate forces and energies. Despite these limitations, the surprisingly accurate solvation energies and forces in Table 6.3 suggest that spline-smoothed surfaces work quite well for reasonably compact systems.

6.4.4 Comparison to Other Parameter Sets

Several of the continuum parameter sets commonly used in conjunction with the AMBER force field were tested alongside our optimized radii to assess their relative accuracy. These included the AMBER *ff*99 charges combined with *ff*99 vdW radii,[105] Bondi radii,[90] and a set of radii recently published by Luo *et al*,.[106] as well as PARSE charges combined with two variants of PARSE radii. [68] As previously mentioned, each set of radii had to be scaled for use with the spline-smoothed surfaces. The scaling factors were determined with both force and force-energy based genetic algorithm optimizations in an analogous manner to the radii presented in Table 6.2. With only one parameter to fit, these optimizations converged in 1 to 5 generations. Replica optimizations found the same scaling factors, indicating that the results were independent of the run parameters.

Table 6.3: Summary of force RMSD's and solvation energy average absolute errors (AAE's) for optimized radii and comparison parameter sets.

		RMSD_F [kcal/mol Å]		AAE_ΔG_{solv} [kcal/mol]			
parameters[a]	scaling factor	trpcage & G-peptide*[b]	IFABP[c]	di/poly-ala peptides*[d]	di/poly-ala peptides[e]	protein peptides[f]	
Amber-f	0.378	0.870	1.020	4.953	4.719	20.296	7.60%
Amber-fe	0.281	1.055	1.222	1.986	1.873	14.997	6.02%
Parse-f	0.236	1.801	2.438	4.766	6.346	26.608	10.99%
Parse-fe	0.310	1.822	2.178	2.944	3.342	11.213	4.00%
Parse0-f	0.236	1.812	2.456	4.186	4.346	27.778	11.47%
Parse0-fe	0.269	1.816	2.327	3.929	3.342	14.725	6.23%
Bondi-f	0.311	1.216	1.460	3.314	2.356	28.122	11.29%
Bondi-fe	0.373	1.240	1.443	1.088	0.971	7.014	2.99%
Luo-f	0.403	0.923	1.014	5.565	5.173	15.108	5.67%
Luo-fe	0.283	1.150	1.360	1.902	1.949	26.472	10.74%
IFABP Opt-f	na	0.750	0.906	4.136	3.799	11.379	4.35%
Optimized-f	na	0.656	0.971	1.035	0.832	7.4545	3.00%
Optimized-fe	na	0.667	0.978	0.889	0.689	5.9838	2.34%

a) Optimized radii and comparison parameter sets scaled by scaling factor [Å], x, ($r_{new} = r_{orig} + x$) to fit forces (-f) or forces and energies (-fe). b) Trpcage and G-peptide forces used in force and force-energy training sets. c) IFABP forces used in IFABP force optimization training set. d) Di- and poly-alanine peptides used in force-energy training set. e) Di- and poly-alanine peptides *not* used in training set. f) Four protein-like peptides *not* used in training sets.

The two variants of the PARSE parameters differed only in the value of the nonpolar hydrogen radius. The first set used 1 Å, as intended in the original publication, while the second used 0 Å, the standard value in PARSE parameter implementation. The first set (R_{Hnp} = 1 Å) performed marginally better on both forces and energies, validating Sitkoff's intention for the PARSE radii to be the same as the standard Pauling vdW radii (with the exception of hydrogens which were decreased from 1.2 Å to 1.0 Å).[68]

The final results for the scaled radii (Table 6.3) show that the scaling factors for the force and force-energy optimizations were significantly different for each of the parameter sets. Nevertheless, they followed several trends demonstrated by our optimized radii: the force-optimized radii performed better on forces and worse on energies than the force-energy radii, the spread in deviations for IFABP were larger than those for Trpcage and G-peptide, and the solvation energy deviations for the amino acid dipeptides and polyalanine peptides in the test set

were of comparable or lesser value to those in the training set. The only unexpected result was the larger solvation energy deviation for the protein-like peptides from the Luo force-energy radii than the Luo force radii. Out of the comparison parameter sets, the AMBER and Luo radii performed best on forces while the Bondi radii performed best on solvation energies. As expected, our force and force-energy-optimized radii (presented in Table 6.2) performed better on forces and energies than all of the comparison sets.

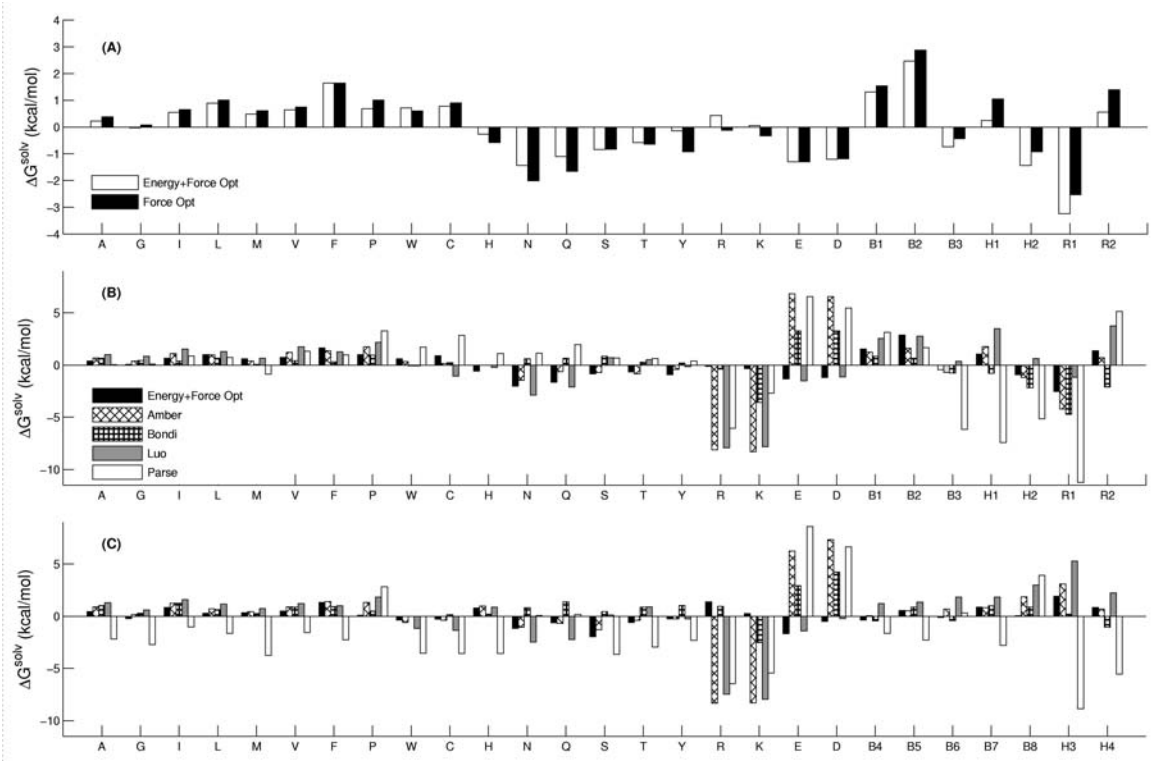

Figure 6.5: Solvation energy deviations for training set dipepetides and polyalanine peptides using (A) force and force-energy optimized radii, (B) force-energy optimized, Amber ff99, Bondi, Luo and PARSE parameters. (C) Same as (B) on test set of dipeptides and polyalanine peptides.

6.4.5 Scaled Radii for Varying Spline Window Widths

The spline window width used in our optimizations (w = 0.3 Å) will not be optimal for every problem or chosen by every user. We, therefore, scaled our radii for a range of spline window widths (w = 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 Å). Similar to the comparison parameter sets, a single scaling factor was determined with both force and force-energy optimizations.

The optimizations converged in 1 to 5 generations and replica optimizations found the same scaling factors, indicating that the results were independent of the run parameters. The final results are presented in Table 6.4 and plotted in Figure 6.6. Both the force and force-energy scaling values are well fit by quadratic polynomials, $x = -0.0509\ w^2 + 0.5027\ w - 0.1471$ and $x = -0.2036\ w^2 + 0.6741\ w - 0.1836$ respectively. These functions can be used to determine scaling values for any spline window width, but widths less than 0.1 Å or larger than 1.1 Å are not expected to yield reliable results.

Table 6.4: Scaling factors for different spline window widths for (a) force and (b) force-energy-optimized radii.

spline width	scaling factor x [Å]	
w [Å]	force[a]	force-energy[b]
0.2	-0.051	-0.057
0.3	0.000	0.000
0.4	0.051	0.054
0.5	0.093	0.103
0.6	0.129	0.147
0.7	0.178	0.188
0.8	0.225	0.226
0.9	0.267	0.258
1.0	0.303	0.287

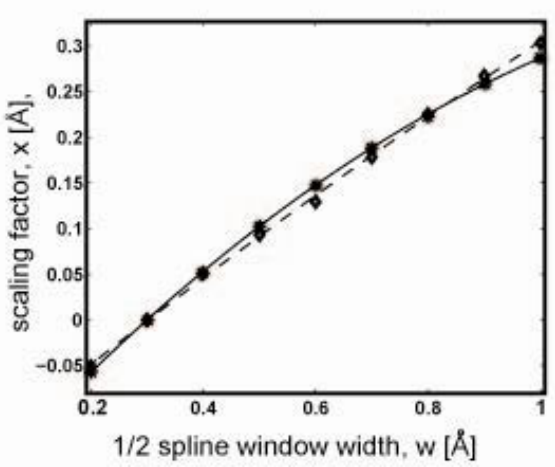

Figure 6.6: Scaling factors for different spline window widths. The force results (diamonds) are best fit by $x = -0.0509\ w^2 + 0.5027\ w - 0.1471$ (dashed). The force-energy results (stars) are best fit by $x = -0.2036\ w^2 + 0.6741\ w - 0.1836$ (solid).

6.5 Conclusions

We have presented two sets of radii optimized to define spline-smoothed dielectric boundaries in Poisson electrostatics calculations on protein systems using AMBER *ff*99 partial charges. These optimizations were based on explicit solvent simulations, and thus offer forces and energies consistent with explicit solvent polar forces and charging free energies. More importantly, we have presented a method of optimizing the Poisson dielectric boundary based on atomic forces, which is substantially more efficient and widely applicable than previous optimization approaches. It can be used, for example, on highly charged, large, or unusual systems for which experimental solvation energies are unavailable and explicit solvent charging free energies are computationally challenging or prohibitive.

We have shown that optimizations based on forces alone generate radii that reproduce forces accurately, but solvation energies less accurately. Optimizations based on both forces and energies, on the other hand, decrease the accuracy of forces minimally and increase solvation energy accuracy significantly. Thus, our force-energy-optimized radii are more appropriate than our force-optimized radii when both energies and forces are important. As anticipated, our optimized radii perform better than four other parameter sets commonly used in conjunction with the AMBER force field, each of which was scaled for use with spline-smoothed dielectric boundaries. Of these comparison sets, the AMBER and Luo radii perform best on forces while the Bondi radii perform best on energies.

Finally, we have shown that atom-centered dielectric functions, such as the spline-smoothed surface used in our optimizations, form regions of high dielectric in interstitial spaces, thereby limiting the accuracy of solvation forces and energies. These errors seem less pronounced in compact systems, where the IHDs fill the crevices between atoms, than in globular systems, where buried IHDs can be copious. Thus, our optimized radii, and spline-

smoothed surfaces in general, are expected to work well for compact systems, but will produce errors in highly-solvated globular proteins. Efforts to correct atom-centered, smooth surfaces will be crucial for the continued development of Poisson based implicit solvent models, especially in the area of continuum dynamics.

It is important, however, to appreciate the complexity of the problem at hand. The specific nature of the optimal dielectric boundary most rigorously depends on the locally defined solute-solvent potential, [113, 114] a complex and coupled function of solute geometry (i.e., surface curvature), electrostatic and nonelectrostatic solvent-solute interactions. Thus, theory and our analysis agree that an optimal continuum radius can not be defined by atom type alone. Consequently we anticipate that corrected smooth surfaces will greatly reduce inaccuracies and inconsistencies, but that it will take a more physically accurate and coupled framework to eliminate them.

This chapter is a reprint in full of material that will appear in Optimizing the Poisson dielectric boundary with explicit solvent forces: Lessons learned with atom-centered dielectric functions. , Jessica M.J. Swanson, J. Wagoner, N. A. Baker, and J.A. McCammon, J Chem Theory and Comput (submitted). I was the primary researcher and author of this work.

Chapter Seven

Coupling Hydrophobicity, Dispersion, and Electrostatics in Continuum Solvent Models

ABSTRACT

This chapter presents an implicit solvent model that couples hydrophobic, dispersion, and electrostatic solvation energies by minimizing the system Gibbs free energy with respect to the solvent volume exclusion function. The solvent accessible surface is the output of the theory. The method is illustrated with the solvation of simple solutes on different length scales and captures the sensitivity of hydration to the particular form of the solute-solvent interactions in agreement with recent computer simulations.

7.1 Introduction

Much progress has been made in the last decade in the understanding of hydrophobic solvation on different length scales [1,2]. Most of this work has been devoted to study solvation of purely repulsive, hard-sphere-like solutes, while less attention has been given to the influence and incorporation of dispersion or electrostatic contributions. Likewise, an entire field in the biophysical community has explored electrostatic solvation effects in the absence or uncoupled addition to hydrophobic considerations; see, e.g., [3] for a review. Recently, however, several computer simulations have demonstrated a strong coupling between hydrophobicity, solute-solvent dispersion attractions, and electrostatics. For example, simulations of explicit water

between platelike solutes revealed that hydrophobic attraction and dewetting phenomena are strongly sensitive to the nature of solute-solvent dispersion interactions [4,5]. Similarly, simulations of hydrophobic channels [6,7] and nanosolutes [8] have shown that electrostatic potentials strongly affect the dewetting behavior and potentials of mean force (pmf). A fully atomistic simulation of the folding of the two-domain protein BphC enzyme [9] further supported coupling by showing that the region between the two domains was completely dewetted when solvent-solute van der Waals (vdW) and electrostatic interactions were turned off, but accommodated 30% of the density of bulk water with the addition of vdW attractions, and 85%–90% with the addition of electrostatics, in accord with experimental results. Finally, Liu *et al.* recently observed a dewetting transition in the collapse of the melittin tetramer, which was strongly sensitive to the type and location of the hydrophobic residues proving that these observations apply to realistic biomolecular systems [10].

In this chapter a continuum description of solvation is proposed that explicitly couples hydrophobic, dispersion, and electrostatic contributions. The Gibbs free energy is expressed as a *functional* of the solute cavity shape, the latter given by the volume exclusion function of the solvent [11], and obtain the optimal shape by minimization. This leads to an expression similar to the Laplace-Young equation for the geometrical description of capillary surfaces [12], but in contrast to existing approaches *explicitly* includes the inhomogeneous distributions of dispersion and electrostatic contributions as well as curvature corrections. Geometry-based approaches similar to our formalism exist in related fields, such as the Helfrich description of membranes shapes [12], wetting in colloids and granular media [12], and electrowetting [13]. As opposed to other implicit solvent models, the solvent accessible surface (SAS) is an output of our theory.

We begin by verifying that this method is able to describe the solvation of small alkanes on molecular scales. We then demonstrate that our theory captures the strong sensitivity of dewetting and hydrophobic hydration to solute-solvent interactions on larger scales for a model

system of two alkane-assembled spheres. In this striking example the strong hydrophobic attraction decreases almost 2 orders of magnitude in units of the thermal energy, k_BT, and dewetting is partially or completely suppressed when realistic dispersion and electrostatic contributions are included. We expect our approach to be particularly useful in solvation studies of proteins, where the hydrophobic surfaces are highly irregular and laced with hydrophilic units [1,10] and superhydrophobic nanosolutes [15].

7.2 Theory and Methods

Let us consider an assembly of solutes with arbitrary shape and composition surrounded by a dielectric solvent in a macroscopic volume *W*. We define a subvolume *V* empty of solvent for which we can assign a volume exclusion function in space given by $v(\vec{r}) = 0$ for $\vec{r} \in V$ and $v(\vec{r}) = 1$ elsewhere. We assume that the surface bounding the volume is continuous and closed. The absolute volume *V* and interface area *S* of *V* can then be expressed as functionals of $v(\vec{r})$ via $V[v] = \int_W d^3r[1 - v(\vec{r})]$ and $S[v] = \int_W d^3r\left|\vec{\nabla}v(\vec{r})\right|$, where $\vec{\nabla} \equiv \vec{\nabla}_{\vec{r}}$ is the usual gradient operator. The density distribution of the solvent is either zero or ρ_0 defined by $\rho(\vec{r}) = \rho_0 v(\vec{r})$, where ρ_0 is the bulk density of the solvent at fixed temperature and pressure. The solutes' positions and conformations are fixed.

We suggest expressing the Gibbs free energy $G[v]$ of the system as a functional of $v(\vec{r})$ and obtaining the optimal volume and surface via minimization $\delta G[v]/\delta v(\vec{r}) = 0$. We adopt the following ansatz for the Gibbs free energy:

$$G[v] = PV[v] + \int_W d^3r\gamma(\vec{r})\left|\vec{\nabla}v(\vec{r})\right| + \int_W d^3r\rho(\vec{r})U(\vec{r}) + \frac{\varepsilon_0}{2}\int_W d^3r\left\{\vec{\nabla}\Psi(\vec{r})\right\}^2\varepsilon(\vec{r}). \qquad 7.1$$

The first term in (7.1) is the energy of creating a cavity in the solvent against the difference in bulk pressure between the liquid and vapor, $P = P_l - P_v$. The second term describes the energetic cost due to solvent rearrangement close to the cavity surface in terms of a parameter $\gamma(\vec{r})$. This interfacial energy penalty is thought to be the main driving force for hydrophobic phenomena [1]. $\gamma(\vec{r})$ is not only a solvent specific quantity but also depends on the topology of the surface in a nonlocal way [16]; i.e., it is a functional of the volume exclusion function, $\gamma(\vec{r}) = \gamma(r;[v])$. The exact form of this functional is not known. For planar macroscopic solvent-cavity interfaces $\gamma(\vec{r})$ is usually identified by the liquid-vapor surface tension, γ_{lv}, of the solvent, which we also employ here. Furthermore, we make a *local curvature approximation;* i.e., we assume that $\gamma(\vec{r})$ can be expressed solely as a function of the local mean curvature of the cavity interface, $\gamma(\vec{r}) = \gamma(H(\vec{r}))$, with $H(\vec{r}) = [\kappa_1(\vec{r}) + \kappa_2(\vec{r})]/2$, where κ_1 and κ_2 are the two principal curvatures. We then apply the first order curvature correction to γ_{lv} given by scaled-particle theory [17], the commonly used ansatz to study the solvation of hard spheres, arriving at

$$\gamma(\vec{r}) = \gamma_{lv}[1 - 2\delta H(\vec{r})], \qquad 7.2$$

where δ is a constant and positive length expected to be of the order of the solvent particle size [17]. The curvature is positive or negative for convex or concave surfaces, respectively. Note that this leads to an increased surface tension for concave surfaces. It has been shown by simulations that (2) predicts the interfacial energy of growing a spherical cavity in water rather well for radii ≥3 Å [18].

The third term in (7.1) is the total energy of the nonelectrostatic solute-solvent interaction given a density distribution $\rho_0 v(\vec{r})$. The energy $U(\vec{r}) = \sum_i U_i(\vec{r} - \vec{r}_i)$ is the sum of the short-ranged repulsive and long-ranged attractive dispersion interactions U_i between

each solute atom *i* and a solvent molecule. Classical solvation studies typically represent U_i as an isotropic Lennard-Jones (LJ) potential, $U_{LJ}(r) = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, with an energy scale ε and a length scale σ. The importance of treating dispersion interactions independently, as opposed to absorbing them into the surface tension term, has been emphasized by Gallicchio *et al.* in their study of cyclic alkanes [19].

The fourth term in (7.1) describes the total energy of the electrostatic field expressed by the local electrostatic potential $\Psi(\vec{r})$ and the position-dependent dielectric constant $\varepsilon(\vec{r})$ assuming linear response of the dielectric solvent. The electrostatic potential Ψ is evaluated by Poisson's equation, $\vec{\nabla}\cdot[\varepsilon(\vec{r})\vec{\nabla}\Psi(\vec{r})] = -\lambda(\vec{r})/\varepsilon_0$, where $\lambda(\vec{r})$ is the solute's charge density distribution. The most common approximation for $\varepsilon(\vec{r})$ is proportional to the volume exclusion function $v(\vec{r})$ [3]

$$\varepsilon(\vec{r}) = \varepsilon_v + v(\vec{r})(\varepsilon_l + \varepsilon_v); \qquad 7.3$$

where ε_v and ε_l are the dielectric constants inside and outside the volume V, respectively. Plugging in (7.2) and (7.3) in functional (7.1) and using the calculus of functional derivatives, the minimization yields

$$0 = P + 2\gamma_{lv}[H(\vec{r}) - \delta K(\vec{r})] - \rho_0 U(\vec{r}) - \frac{\varepsilon_0}{2}[\nabla\Psi(r)\varepsilon(\vec{r})]^2\left(\frac{1}{\varepsilon_l} + \frac{1}{\varepsilon_v}\right). \qquad 7.4$$

Equation 7.4 is a partial second order differential equation for the optimal solvent accessible volume and surface expressed in terms of pressure, surface curvatures, dispersion interactions, and electrostatics, all of which have dimensions of force per surface area or energy density. $K(\vec{r}) = \kappa_1(\vec{r})\kappa_2(\vec{r})$ is the Gaussian curvature and follows from the variation of the surface integral over $H(\vec{r})$ in (1). Thus, in our approach the geometry of the surface, expressed by H

and K, is directly related to the inhomogeneous dispersion and electrostatic energy contributions. Note that the SAS is presently defined with respect to the LJ centers of the solvent molecules.

In the following we illustrate solutions of (7.4) in spherical and cylindrical symmetries. For a spherical solute (4) reduces to a function of R, the radius of the solvent accessible sphere, $H = 1/R$ and $K = 1/R^2$. In cylindrical symmetry the SAS can be expressed by a one-dimensional shape function $r(z)$, where z is the coordinate on the symmetry axis and r the radial distance to it. The surface in three-dimensional space is obtained by revolving $r(z)$ around the z axis. The principal curvatures are then given by $\kappa_1 = -1/(r\sqrt{r^2+1})$ and $\kappa_2 = r''/[(r^2+1)^{3/2}]$, where the primes indicate the partial derivative with respect to z. We solve (7.4) and Poisson's equation numerically, using standard forward time relaxation schemes.

7.3 Results and Discussion

7.3.1 Molecular Length Scales

We now study the solvation of methane and ethane in water and compare our results to the simple point charge (SPC) explicit water simulations by Ashbaugh *et al.* [20], in which the alkanes are modeled by neutral LJ spheres [21].We fix the liquid-vapor surface tension for SPC water at 300 K to $\gamma_{lv} = 65\,\mathrm{mJ/m^2}$ [18]. Since we deal with water under ambient conditions, the pressure term can be neglected and the length δ remains the only free parameter. For methane we can reproduce the simulation solvation energy ΔG with a fit $\delta = 0.85$ Å. This is in good agreement with Huang *et al.* [18] who measured $\delta = 0.76 \pm 0.05$ Å for SPC water. Solving the cylindrically symmetric problem for the diatomic ethane with the same $\delta = 0.85$ Å, we obtain a

fit-parameter-free ΔG = 11.40 kJ/mol, which is only 7% larger than the simulation results. Alternatively, the best fit δ = 0.87 Å reproduces the simulation energy exactly. This is surprisingly good agreement given the crude curvature correction we apply and the fact that the large curvature of the system varies locally in space. This supports the validity of our continuum approach down to a molecular scale.

The curvature and shape functions $H(z)$, $K(z)$, and $r(z)$ are plotted in Fig. 1 together with the vdW surface and the canonical SAS obtained from rolling a probe sphere with a typically chosen radius r_p =1.4 Å over the vdW surface [14]. Away from the center of mass $|z| \geq 1$ Å the curvatures follow the expected trends $H \cong 1/R$ and $K \cong 1/R^2$ with $R \cong 3.1$ Å for the spherical surfaces. The surface resulting from our theory is smaller than the canonical SAS, and is smooth at the center of mass (z = 0) where the canonical SAS has a kink. Thus, our surface has a smaller mean curvature at z = 0 and an almost zero Gaussian curvature, which is typical for a cylinder geometry for which one of the principal curvatures is zero. These results may justify the use of smooth surfaces in coarse-grained models of closely packed hydrocarbon surfaces, a possibility we now explore with solvation on larger length scales where dewetting effects can occur.

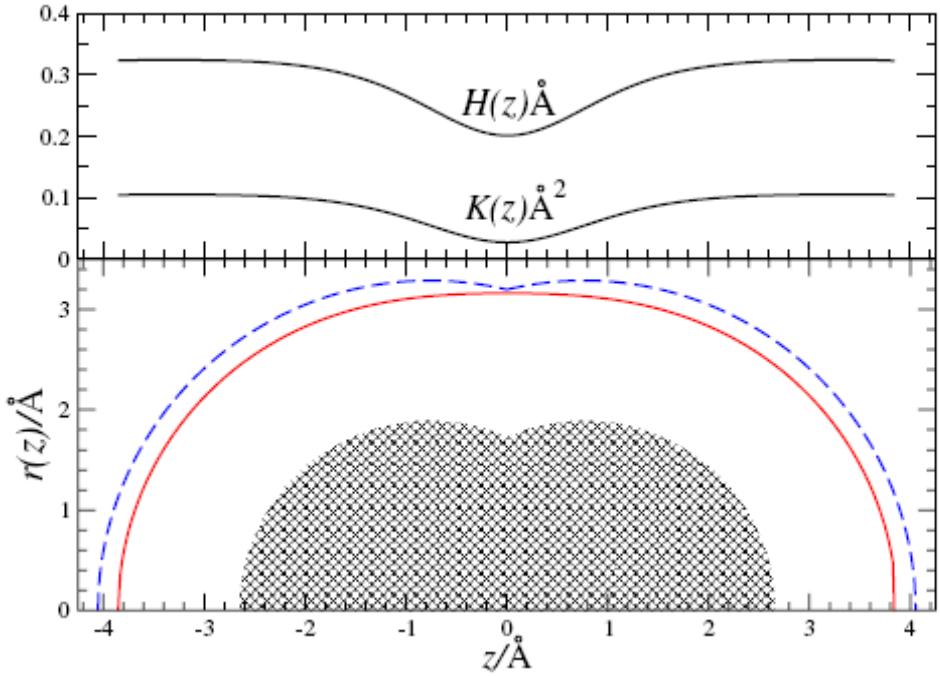

Figure 7.1: Mean $H(z)$ and Gaussian $K(z)$ curvature and shape function $r(z)$ (solid lines) for ethane. The canonical SAS (dashed line) from rolling a probe sphere with radius r_p 1.4 Å over the vdW surface (shaded region) is also shown.

7.3.2 Nano Length Scales

Let us consider two spherical solutes that we assume to be homogeneously assembled of CH_2 groups with a uniform density $\rho = 0.024$ Å^{-3} up to a radius $R = 15$ Å, defined by the maximal distance between a CH_2 center and the center of the solute [22]. The integration of the CH_2-water LJ interaction over the volume of a sphere yields a 9-3-like potential for the interaction between the center of the paraffin sphere and a water molecule [23]. The intrinsic, nonelectrostatic solute-solute interaction $U_{ss}(r_{12})$ in a center-to-center distance r_{12} can be obtained in a similar fashion. The solvation of the two solutes is studied for a fixed surface-to-surface distance, which we define as $s_0 = r_{12} - 2R_0$. We obtain an effective SAS radius of one

sphere of about $R \cong R_0 + 2.4$ Å so that the effective surface-to-surface distance is roughly $s \cong s_0 - 4.8$ Å. Since we are also interested in the effects of charging up the solutes we place opposite charges $\pm Ze$, where e is the elementary charge, in the center or on the edge of the two spheres.

In the following we focus on a separation distance of $s_0 = 8$ Å to investigate the influence of different contributions to the energy functional on the shape function, $r(z)$, and the curvatures, $K(z)$ and $H(z)$. For $s_0 = 8$ Å, it follows that $s \cong 3.2$ Å, such that two water molecules could fit between the solutes on the z axis. We systematically change the solute-solute and solute-solvent interactions, as summarized in Table 7.1. We begin with only the LJ repulsive interactions in system I and then add a curvature correction with $\delta = 0.75$ Å vdW attractions, and sphere centered charges $Z = 4$ and $Z = 5$ in systems II–V, respectively. To study the influence of charge location, we reduce the magnitude of each charge in system VI to $Z = 1$ and move them to the edge of the spheres on the symmetry axis such that they are 8 Å apart (indicated by arrows in Figure 7.2). The surface tension and dielectric constant of the vapor and liquid are fixed to $\gamma_{lv} = 72$ mJ/m^2, $\varepsilon_v = 1$, and $\varepsilon_v = 78$, respectively.

Table 7.1: Studied systems for two alkane-assembled spherical solutes. If $r\_z\_0\_\_0$ the system is ‘‘dewetted.’’ In VI the solutes’ charge is located off-center (OC) at the solute surface.

System	δ (Å)	vdW attraction	Z	$W(s_0)/k_BT$	Dewetted
I	0.00	no	0	−57.6	yes
II	0.75	no	0	−34.1	yes
III	0.75	yes	0	−6.3	yes
IV	0.75	yes	4	−9.2	yes
V	0.75	yes	5	−5.1	no
VI	0.75	yes	1 (OC)	−1.3	no

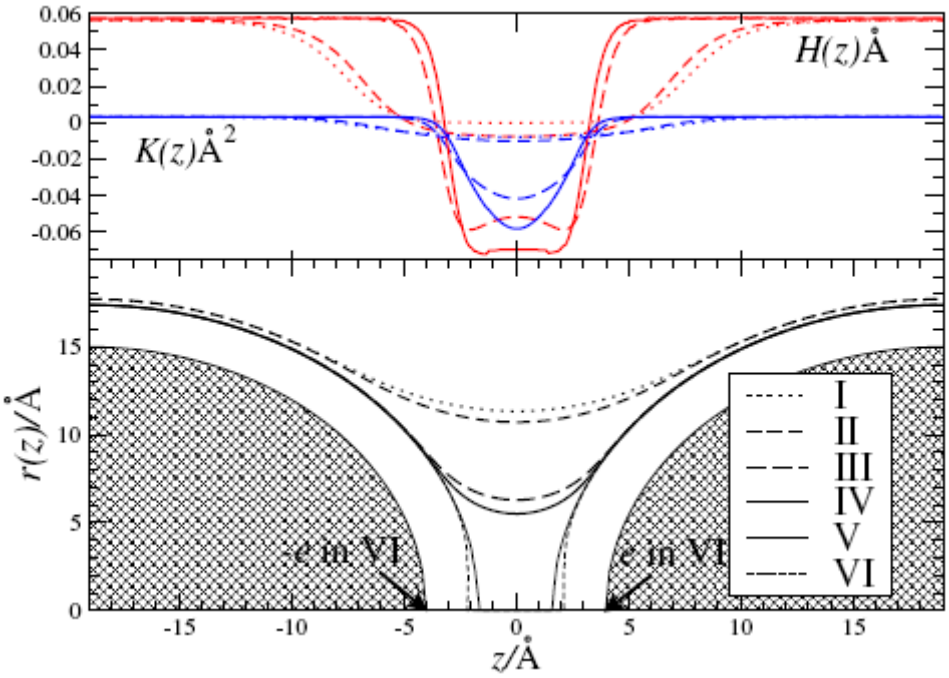

Figure 7.2: Mean $H(z)$ and Gaussian $K(z)$ curvatures and shape function $r(z)$ for two alkane-assembled solutes of radius $R_0 = 15$ Å (shaded region) for systems I–VI. The position of the charges $Z \pm 1$ in VI are indicated by arrows. Curvatures are not shown for the ''wet'' systems V and VI.

The results for the curvatures and $r(z)$ for systems I–VI are shown in Figure 7.2. Away from the center of mass ($|z| \geq 10$ Å), systems I–VI show very little difference. The curvatures are $H \cong 1/R$ and $K \cong 1/R^2$ with $R \cong 17.4$ Å. Close to the center of mass ($z \cong 0$), however, the influence of changing the parameters is considerable. In system I, equation 7.4 reduces to the minimum surface equation $H(z) = 0$ for $z \cong 0$. For two adjacent spheres the solution of this equation is the catenoid $r(z) \propto \cosh(z)$, which features zero mean curvature (κ_1 and κ_2 cancel each other) and negative Gaussian curvature. This leads to a vapor bubble bridging the solutes. When curvature correction is applied (system II), the mean curvature becomes nonzero and negative (concave) at $z \cong 0$, while the Gaussian curvature grows slightly more negative. As a consequence, the total enveloping surface area becomes larger and the solvent inaccessible

volume shrinks; i.e., the value of $r(z \cong 0)$ decreases. Turning on solute-solvent dispersion attraction amplifies this trend significantly as demonstrated by system III. Mean and Gaussian curvatures increase fivefold, showing strongly enhanced concavity, and the volume empty of water decreases considerably, expressed by $r(z = 0) \cong 10.7$ Å dropping to $r(z = 0) \cong 6.3$ Å. These trends continue with the addition of electrostatics in system IV. When the sphere charges are further increased from $Z = 4$ to $Z = 5$ (system IV → V), we observe a wetting transition: the bubble ruptures and the SAS jumps to the solution for two isolated solutes, where $r(z = 0) \cong 0$. The same holds when going from III to VI, when only one charge, $Z = 1$, is placed at each of the solutes' surfaces. Importantly, this demonstrates that the present formalism captures the sensitivity of dewetting phenomena to specific solvent-solute interactions as demonstrated in previous studies [4 –10]. Note that the SAS at $z \cong 2$ Å is closer to the solutes in VI compared to V due to the proximity of the charge to the interface. Clearly, the observed effects, in particular, the transition from III to VI, cannot be described by existing solvation models, which use the SAS [14], or effective surface tensions and macroscopic contact angles [12] as input. The significant change of the SAS with the solutesolvent interaction has a strong impact on the pmf, $W(s_0) = G(s_0) - G(\infty) + U_{ss}(s_0)$. Values of $W(s_0 = 8\,\text{Å})$ are given in Table 7.1. From system I to VI the total attraction between the solutes decreases almost 2 orders of magnitude. Interestingly, the curvature correction (I → II) lowers W by a large 23.5 k_BT, even though $R >> \delta$. A striking effect occurs when vdW contributions are introduced (II → III): the intersolute attraction decreases by ~ 28 k_BT while the dispersion solute-solute potential, $U_{ss}(s_0 = 8\,\text{Å})$, changes by only -0.44 k_BT. Similarly, adding charges of $Z = 5$ (III → V) at the solutes' centers or $Z = 1$ (III → VI) at the solutes' surfaces decreases the total attraction by 1.2 k_BT and 5 k_BT, respectively. Note that the total attraction decreases although electrostatic

attraction has been added between the solutes. The same trends have been observed in explicit water simulations of a similar system of charged hydrophobic nanosolutes [8].

7.4 Conclusions

These results clearly demonstrate that solvation effects and solvent mediated phenomena are not only strongly influenced by solute-solvent interactions, but that these interactions are inherently coupled. By including coupling, our formalism captures the balance between hydrophobic, dispersive, and electrostatic forces, which has been observed in previous studies [4–10] but never described in a single theoretical framework. Nonpolar and polar coupling is expected to be crucial for a complete characterization of biomolecular solvation. The present formalism is limited only by the crude curvature and dielectric descriptions currently employed. Future efforts to improve these approximations are critical to accurately describe solvation effects on multiple length scales and for more complicated geometries.

This chapter is a reprint in full of material that appeared in *Coupling Hydrophobicity, Dispersion, and Electrostatics in Continuum Solvent Models.* Dzubiella J., J.M.J. Swanson, J.A. McCammon, Phys. Rev. Lett.96, 087802 (2006). I was a secondary researcher and author of this work.

Chapter Eight

Coupling Nonpolar and Polar Solvation Free Energies in Implicit Solvent Models

ABSTRACT

Recent studies on the solvation of atomistic and nanoscale solutes indicate that a strong coupling exists between the hydrophobic, dispersion, and electrostatic contributions to the solvation free energy, a facet not considered in current implicit solvent models. In this chapter we extend our theoretical formalism which accounts for coupling by minimizing the Gibbs free energy of the solvent with respect to a solvent volume exclusion function. The resulting differential equation is similar to the Laplace-Young equation for the geometrical description of capillary interfaces but is extended to microscopic scales by explicitly considering curvature corrections as well as dispersion and electrostatic contributions. Unlike existing implicit solvent approaches, the solvent accessible surface is an output of our model. The presented formalism is illustrated on spherically or cylindrically symmetrical systems of neutral or charged solutes on different length scales. The results are in agreement with computer simulations and, most importantly, demonstrate that our method captures the strong sensitivity of solvent expulsion and dewetting to the particular form of the solvent-solute interactions.

8.1 Introduction

Implicit solvent models are widely used in theoretical chemistry to study the solvation of biomolecular systems, as well described in the review of Roux and Simonson.[23] They provide a more efficient, although generally less accurate, alternative to atomistically resolved explicit solvent simulations. The solvation free energy in these models is usually split into nonpolar (np) and polar (p) terms,

$$\Delta G = \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{p}} \qquad 8.1$$

which are treated in separate energetic evaluations. The nonpolar term includes the energetic cost of cavity formation, solvent rearrangement, and solute-solvent dispersion interactions introduced when the uncharged solute is brought from vacuum into the solvent environment. The polar term describes the free energy of charging the mono- or multi-polar solute in the dielectric medium.

The nonpolar term is commonly approximated by surface area models, i.e., $\Delta G_{\mathrm{np}} \cong \gamma S$, where S is the solvent accessible surface area[119] and γ is an energy per surface area constant, which is *a priori* not known but fit to atomistic simulations. The deficiencies of this simple surface area approach have been recognized and a further decomposition of the nonpolar term into cavity (cav) and van der Waals dispersion (vdW) terms has been proposed,[120, 121] $\Delta G_{\mathrm{np}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}}$. This approach has shown improved results for the solvation of alkanes,[121, 122] the alanine peptide,[123] and nonpolar native and misfolded proteins.[124] The electrostatic (polar) contribution of the solvation free energy is often approximated by generalized Born[125] (GB) or Poisson-Boltzmann[12] (PB) models. Both methods use a position-dependent dielectric constant,[23] assigned on the basis of the solute surface, which can be defined in several ways[126] or defined implicitly by integration methods. It has been

emphasized that all three contributions, ΔG_{cav}, ΔG_{vdW}, and ΔG_{p}, depend critically on the location of the solvent-solute interface. It has *also* been shown that the effective location of the solvent-solute interface can vary according to the local electrostatic[70] and dispersion[127] potentials. This suggests that interfacial, dispersion, and electrostatic contributions should be coupled in implicit solvent approaches. The importance of capturing the right balance between nonpolar and electrostatic contributions in implicit solvation models was emphasized by Ashbaugh *et al.* in their study of amphiphiles.[128]

The significance of the coupling of nonpolar and polar solvation becomes even more evident when solvation is studied on length scales which are large compared to the solvent molecule (typically ≥ 1 nm for water), where solvent dewetting ("drying") can occur. In this mechanism, first envisioned by Stillinger,[129] the solvent molecules tend to move away from the surface of a large nonpolar solute forming a liquidgas- like interface parallel to the solute interface. When the surfaces of two large solutes come together dewetting can be amplified due to the gain of interfacial free energy (by decreasing the total liquid-vapor interface area) giving rise to a strong effective attraction.[130-132] Early evidence of confinement-induced dewetting was given only by explicit water simulations for smooth platelike solutes with a purely repulsive solute-solvent interaction.[133] More recently, however, it has been demonstrated in varying degrees in several systems with attractive solute-solvent interactions including smooth parallel platelike solutes,[134] atomistically resolved paraffin plates,[135] graphite plates,[136] carbon nanotubes,[137] and hydrophobic ion channels.[138-140]

Several of these studies indicated that the magnitude of dewetting is sensitive to the nature of the solute-solvent attractive dispersion interactions.[134-136] A similar sensitivity was found in systems where the solutes carry charges or are exposed to an external electric field, e.g., electrostatic interactions have been shown to strongly affect the dewetting behavior of hydrophobic channels[141-143] and hydrophobic spherical nanosolutes..[144, 145]

Furthermore, two recent simulations of proteins supported the importance of solvent dewetting and its sensitivity in realistic biomolecular systems. First, a simulation of the two-domain BphC enzyme showed that the region between the two domains was completely dewetted when vdW and electrostatic interactions were turned off, but accommodated 30% of the bulk density with the addition of vdW attraction (water was found mainly at the edges of the considered volume, while the central region was still empty), and 85%–90% with the addition of electrostatic interactions.[146] Second, Liu *et al.* observed a clear dewetting transition in the simulation of the collapse of the melittin tetramer, which was strongly sensitive to the type and location of the hydrophobic residues around the dewetted region.[33]

Considering the aforementioned studies, we postulate that coupling of the nonpolar and polar solvation contributions in implicit solvent models is crucial for an accurate determination of solvation free energies without too many system-dependent fit parameters. We suggest a general theoretical formalism in which the particular energetic contributions are coupled. Similar to the approach of Parker *et al.* in their study of bubble formation at hydrophobic surfaces,[147] we express the Gibbs free energy as a *functional* of the solvent volume exclusion function[148] and obtain the optimal solute surface via minimization. As we will show, this minimization leads to an expression which is similar to the Laplace-Young equation for the description of macroscopic capillary interfaces[149] but is generalized to explicitly include curvature corrections and solvent-solute interactions, i.e., short-range repulsion (excluded volume), dispersion, and electrostatics. This extension of the Laplace-Young theory allows a geometric description of solvation on mesoscopic and microscopic scales. Related approaches in other fields are the Helfrich description of vesicle and membrane surfaces,[150, 151] wetting in colloids and granular media,[149, 152] and functional treatments of electrowetting.[153]

While most implicit solvent approaches define the solute surface with a geometrical evaluation of the molecular surface, vdW surface, or canonical solvent accessible surface

(SAS),[119, 126] it is an output of our theory. The surface obtained by minimizing our free energy functional will, in general, be very different than the aforementioned established surface definitions. In particular, our solvent accessible surface should not be confused with the canonical SAS,[119] which is simply the envelope surrounding probe-inflated spheres. Similarly, phenomenological continuum theories applied to solvent dewetting always assume a certain, simplified geometry for the dry region, e.g., a cylindrical volume for systems such as hydrophobic ion channels,[139, 142, 154] platelike particles,[130, 134] or two hydrophobic spherical solutes.[155] For a few simple systems this might be a valid approximation but for more complicated solute geometries the shape of the dewetted volume is unknown and a different approach, as suggested in this work, is necessary. We expect our formalism to be particularly useful in solvation studies of large protein assemblies where the hydrophobic surfaces are highly irregular and laced with hydrophilic units,[156, 157] and for which a unified description of hydration on different length scales is important.[132] Another potential application is the solvation of superhydrophobic nanosolutes[158] and wetting/dewetting in near-critical colloidal mixtures.[152]

A brief summary of this work was presented in chapter 7. In this chapter we present more challenging test cases and an expanded discussion of the approximations and limitations of this model. For clarity, the theoretical formalism is presented again and the chosen approximations are expanded upon. The method is then verified to describe solvation on molecular scales with noble gases, ions, and small alkanes. It is then demonstrated to capture the strong sensitivity of dewetting and hydrophobic hydration to specific solute-solvent interactions on larger scales with two alkane-assembled spheres.

8.2 Theory

8.2.1 Basic Formalism

Let us consider an assembly of solutes with arbitrary shape and composition surrounded by a dielectric solvent in a volume W. Furthermore, we define a subvolume (or cavity) V empty of solvent for which we can assign a volume exclusion function given by

$$v(\vec{r}) = \begin{cases} 0 \text{for } \vec{r} \in V; \\ 1 \text{else.} \end{cases} \tag{8.2}$$

We assume that the surface surrounding the volume is continuous and closed, i.e. has no boundary. The absolute volume V and surface area S of V can then be expressed as functionals of $v(\vec{r})$ via

$$\begin{aligned} V[v] &= \int_W \mathrm{d}^3 r \, [1 - v(\vec{r})] \\ S[v] &= \int_W \mathrm{d}^3 r \, |\nabla v(\vec{r})|, \end{aligned} \tag{8.3}$$

where $\nabla \equiv \nabla_{\vec{r}}$ is the usual gradient operator with respect to the position vector $\vec{r}$ and $|\nabla v(\vec{r})|$ gives a δ-function-like contribution only at the volume boundary. The expression $\mathrm{d}^3 r \, |\nabla v(\vec{r})| \equiv \mathrm{d}S$ can thus be identified as the infinitesimal surface element. In this continuum solvent model, V defines the volume which is not accessible to solvent molecules and the solvent density distribution is simply $\rho(\vec{r}) = \rho_0 v(\vec{r})$, where ρ_0 is the bulk density of the solvent at the desired temperature and pressure. Local inhomogeneities of the solvent density, apart from the zero to ρ_0 transition at the volume boundaries, are neglected. The solutes' positions and conformations are fixed, so that the solutes can be considered as an external potential to the solvent without any degrees of freedom.

As motivated before, we suggest expressing the Gibbs free energy $G[v]$ as a functional of the volume exclusion function $v(\vec{r})$, and obtaining the optimal solute volume via minimization

$$\delta G[v]/\delta v[\vec{r}] = 0, \tag{8.4}$$

where δ denotes the functional derivative. We adopt following ansatz for the Gibbs free energy of the solvent:

$$\begin{aligned} G[v] &= G_{\text{pr}}[v] + G_{\text{int}}[v] + G_{\text{ne}}[v] + G_{\text{es}}[v] \\ &= PV[v] + \int_W \mathrm{d}^3 r\, \gamma(\vec{r};[v])\,|\nabla v(\vec{r})\,| \\ &\quad + \rho_0 \int_W \mathrm{d}^3 r\, v(\vec{r}) U(\vec{r}) \\ &\quad + \int_W \mathrm{d}^3 r \left\{ \frac{\varepsilon_0}{2} \varepsilon(\vec{r};[v])[\nabla\Psi(\vec{r})]^2 - \lambda(\vec{r})\Psi(\vec{r}) + v(\vec{r}) U_{\text{mi}}(\vec{r}) \right\} \end{aligned} \tag{8.5}$$

Let us discuss each of the terms in equation 8.5 in turn. The first term, $G_{\text{pr}}[v]$, proportional to the volume V, is the energy of creating a cavity in the solvent against the difference in bulk *pressure* between the liquid and vapor phase, $P = P_l - P_v$. For water in ambient conditions, which is close to the liquid-vapor transition, this term is relatively small and can generally be neglected for solutes on microscopic scales. The second term $G_{\text{int}}[v]$ describes the energetic cost due to solvent rearrangement around the cavity *interface* with area S in terms of a free energy/surface area functional $\gamma(\vec{r};[v])$. This interfacial energy penalty is thought to be the main driving force behind hydrophobic phenomena.[132] γ is a solvent specific quantity that also depends on the particular topology of the cavity-solvent interface , i.e. it varies locally in space and is a functional of the volume exclusion function $\gamma = \gamma(\vec{r};[v])$.[159] The exact form of this functional is not known.

The third term, $G_{\text{ne}}[v]$, is the total energy of the *non-electrostatic* solute-solvent interaction given a solvent density distribution $\rho_0 v(\vec{r})$. The potential $U(\vec{r}) = \sum_i U_i(\vec{r} - \vec{r}_i)$ is the sum of the (short-ranged) repulsive exclusion and attractive dispersion interaction between each solute atom i at position $\vec{r}_i$ and a solvent molecule at $\vec{r}$. Classical solvation studies typically represent the interaction between a solute atom and a solvent molecule, U_i, as an isotropic Lennard-Jones (LJ) potential,

$$U_{\text{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{8.6}$$

with an energy scale ε, length scale σ, and center-to-center distance r. Using the form of (8.6) implies that $v(\vec{r})$ is defined with respect to the LJ-centers of the solvent molecules.

The last term, $G_{\text{es}}[v]$, describes the total energy of the *electrostatic* field and the mobile ions in the system expressed by the local electrostatic potential $\Psi(\vec{r})$ assuming linear response of the dielectric solvent.[63] Similar to $\gamma(\vec{r};[v])$, the position-dependent dielectric constant depends on the geometry of $v(\vec{r})$, i.e. $\varepsilon(\vec{r}) = \varepsilon(\vec{r};[v])$ with an unknown functional form. $\lambda(\vec{r})$ is the fixed charge density distribution of the solutes and the local energy density of the *mobile ions* is[12, 160]

$$U_{\text{mi}}(\vec{r}) = k_B T \sum_{j=+,-} q_j \rho_j \{\exp[-\beta q_j \Psi(\vec{r})] - 1\}. \tag{8.7}$$

with the thermal energy $k_B T = \beta^{-1}$. Variation of (8.5) for a fixed $v(\vec{r})$ with respect to $\Psi(\vec{r})$ yields the Poisson-Boltzmann equation [12, 160]

$$\begin{aligned} \text{PB}(\vec{r}) = 0 = \nabla \cdot [\varepsilon_0 \varepsilon(\vec{r};[v]) \nabla \Psi(\vec{r})] + \lambda(\vec{r}) \\ + v(\vec{r}) \sum_{j=+,-} q_j \rho_j \exp[-\beta q_j \Psi(\vec{r})], \end{aligned} \tag{8.8}$$

where q_j and ρ_j are the charge and concentration of the mobile ion species $j = +, -$. Note that the ionic charge density in (8.8) is multiplied by $v(\vec{r})$ to account for the fact that ions usually cannot penetrate the volume empty of polar solvent due to a huge free energy penalty.

Let $v_{\min}(\vec{r})$ be the exclusion function which minimizes the functional (8.5). Then, the resulting Gibbs free energy of the system is given by $G[v_{\min}]$. The solvation free energy ΔG is the reversible work to solvate the solute and is given by

$$\Delta G = G[v_{\min}] - G_0, \tag{8.9}$$

where G_0 is a constant reference energy which can refer to the pure solvent state and an unsolvated solute. The potential of mean force (pmf) along some given reaction coordinate x (e.g. the distance between two solutes' centers of mass) is given, within a constant, by $G[v_{\min}]$, where $v_{\min}(\vec{r})$ must be evaluated for every x. In order to proceed we will need valid approximations for $\gamma(\vec{r};[v])$ and $\varepsilon(\vec{r};[v])$ with which $v_{\min}(\vec{r})$ can be calculated by explicitly minimizing our free energy functional (8.5) according to (8.4).

8.2.2 Approximations for $\gamma(\vec{r};[v])$ and $\varepsilon(\vec{r};[v])$

Let us start with a possible description of $\gamma(\vec{r};[v])$. For a planar macroscopic interface the parameter γ is usually identified by the surface tension of the solvent adjacent to the second medium. This surface tension obviously depends on the microscopic interactions between the medium and the solvent and is generally decreased by attractive dispersion or electrostatic contributions. It seems that microscopic interactions are adequately represented by a macroscopic quantity like γ *if* their range is much smaller than the investigated length scales, such as the radii of curvature or mean particle distances. The effect of the microscopic

interactions are then absorbed in γ. This has been exemplified with large, neutral plate-like or spherical alkane-assembled solutes.[127, 134] For the description of solvation on smaller length scales, however, it seems important to separate the free energy into a part which accounts for the formation of a cavity and a part which describes the dispersion interactions explicitly.[120] Furthermore, it has been shown that the water liquid-vapor surface tension γ_{lv} is the asymptotic value of the solvation free energy per surface area for hard spherical cavities in water in the limit of large radii.[130, 161] These considerations motivate our choice of the second and third term in the functional (8.5) and lead to the assumption $\gamma = \gamma_{\mathrm{lv}}$ in the limit of vanishing curvatures.

The surfaces of realistic (bio)molecules, however, display highly curved shapes, so $\gamma(\vec{r};[v])$ will strongly depend on the interface geometry around $\vec{r}$ in a complicated fashion. In the following we make a *local curvature approximation*, i.e. we assume that $\gamma(\vec{r};[v])$ can be expressed solely as a function of the local mean curvature

$$H(\vec{r}) = \left(\kappa_1(\vec{r}) + \kappa_2(\vec{r})\right)/2 = R^{-1}(\vec{r}), \qquad 8.10$$

where R is the radius of mean curvature and $\kappa_1(\vec{r})$ and $\kappa_2(\vec{r})$ are the local principal curvatures of the interface. The latter are formally given by the eigenvectors of the Hessian (or shape operator) of v, $\hat{S}$, which can be expressed by the vector gradient of the unit normal vector field $\vec{n}(\vec{r})$ of the surface, $\hat{S} = \nabla\vec{n}(\vec{r}) = \nabla\frac{\nabla v(\vec{r})}{|\nabla v(\vec{r})|}$. The mean curvature H is only defined at the boundary of $v(\vec{r})$. We have chosen the convention in which the curvatures are negative for convex surfaces (e.g. a spherical cavity) and positive for concave surfaces (e.g. a spherical droplet).

The curvature dependence of the liquid-vapor surface tension is a long standing subject of research but still under steady discussion.[152, 162, 163] For water, which is close to the critical point under ambient conditions, the curvature correction is argued to be nonanalytical in

curvature.[163] The first order correction term, however, is likely to be linear in curvature as predicted by scaled-particle theory,[164] the commonly used ansatz to study the solvation of hard spherical cavities. Although this result is only strictly valid for the case of spherical particles, we assume that it can be applied to local mean curvatures such that $\gamma(\vec{r};[v])$ reduces to the function

$$\gamma(\vec{r}) = \gamma_{\mathrm{lv}}(1 + 2\delta H(\vec{r})), \tag{8.11}$$

where δ is the Tolman length, which is expected to be of molecular size.[165] In our study we assume δ is constant and positive, while the curvature can be positive or negative as defined above. Note that this leads to an increase of surface tension for concave surfaces. It has been shown by computer simulations of growing a hard spherical cavity in water that (8.11) predicts the interfacial energy rather well for radii 3 Å.[160] A major drawback of approximation (8.11) is that it gives unphysical results if the radius of mean curvature is smaller than twice the Tolman length, $|R| < 2\delta$. It yields negative and diverging surface tensions for convex and concave surfaces, respectively, the latter not possible due to the finite size of the solvent molecules. Thus, with approximation (8.11) care has to be taken when investigating systems which can exhibit radii of curvature $|R| < 2\delta$.

Let us now turn to electrostatics. The most common approximation for the position-dependent dielectric constant is proportional to the volume exclusion function $v(\vec{r})$,1 and the functional $\varepsilon(\vec{r};[v])$ reduces to the function

$$\varepsilon(\vec{r}) = \varepsilon_v + v(\vec{r})(\varepsilon_l - \varepsilon_v), \tag{8.12}$$

where ε_v and ε_l are the dielectric constants inside and outside the volume V, respectively. Equation 8.12 is valid only in the limit large solute sizes when the molecular size of the solvent is negligible. For charged solutes on a molecular scale, let's say mono- or polyvalent ions, two difficulties arise. First the electric field close to the highly curved solutes can be very strong and

the dielectric constant becomes field dependent. This formally affects the form of the electrostatic term in the free energy functional which assumes a linear response of the solvent. An improvement for continuum models along these lines has been proposed by Luo and Tucker.[166] Second, the effective position of the dielectric boundary is known to depend on the sign of the charge carried by the solute for asymmetric solvent molecules like water. This expresses itself for instance in different Born radii for two equally charged ions which have exact the same LJ parameters but a different sign of charge. A reasonable improvement of (8.12) would be to shift the dielectric boundary at $\vec{r}$ *parallel* to the volume boundary by a potential dependent amount $\xi(\Psi(\vec{r}))$:

$$\varepsilon(\vec{r}) = \varepsilon_v + v\left(\vec{r} - \xi(\Psi)\vec{n}(\vec{r})\right)(\varepsilon_l - \varepsilon_v). \qquad 8.13$$

We do not attempt however, to find an approximation for the function $\xi(\Psi)$ in this work and postpone this investigation to later studies. For further illustration of our approach we content ourselves with the approximations (8.11) for $\gamma(\vec{r};[v])$ and (8.12) for $\varepsilon(\vec{r};[v])$.

8.2.3 Minimization of the Free Energy Functional

For the functional derivative of the interfacial term, $G_{\text{int}}[v]$, we utilize

$$\frac{\delta}{\delta v}\int_W \mathrm{d}^3 r \,|\nabla v(\vec{r})| = \frac{\delta}{\delta v}\int_{\partial W} \mathrm{d}S = -2H \qquad 8.14$$

and

$$\frac{\delta}{\delta v}\int_W \mathrm{d}^3 r \; H\,|\nabla v(\vec{r})| = \frac{\delta}{\delta v}\int_{\partial W} \mathrm{d}S \; H = -K \qquad 8.15$$

which has been derived in detail by Zhong-can and Helfrich by means of differential geometry.[150] The variable $K(\vec{r}) = \kappa_1(\vec{r})\kappa_2(\vec{r})$ is the Gaussian curvature of the interface,

which is an intrinsic geometric property of v. The other derivatives follow from functional analysis and the total result of the minimization is

$$0 = \mathrm{de}(\vec{r}) = P - 2\gamma_{\mathrm{lv}}\left[H(\vec{r}) + \delta K(\vec{r})\right] - \rho_0 U(\vec{r}) - \frac{\varepsilon_0}{2}[\nabla\Psi(\vec{r})\varepsilon(\vec{r})]^2\left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_v}\right) - U_{\mathrm{mi}}(\vec{r}). \qquad 8.16$$

Equation 8.16 is a partial second order differential equation (de) for the optimal exclusion function $v_{\min}(\vec{r})$ expressed in terms of pressure, curvatures, short-range repulsion, dispersion, and electrostatic terms, all of which have dimensions of energy density. It can also be interpreted as a mechanical balance between the forces per surface area generated by each of the particular contributions. Thus, in our approach the surface shape and geometry, expressed by H and K, are directly related to the inhomogeneous potential contributions. Note that the constant solute charge density $\lambda(\vec{r})$ does not appear explicitly in (8.16) but is implicitly considered in the PB equation (8.8), which must be solved simultaneously. If curvature correction (K-term) and the last three energetic terms are neglected one obtains the Laplace-Young equation,

$$P = 2\gamma_{lv}H, \qquad 8.17$$

which is exclusively used for the shape description of macroscopic capillary and interfacial phenomena in conjunction with appropriate boundary conditions, e.g. prescribed liquid-solid contact angles at the solid surfaces.[149] In our description the boundary conditions are provided by the constraints given by the short-ranged repulsive term in $U(\vec{r})$, and the distribution of dispersion and electrostatics, allowing an extrapolation of the Laplace-Young description to mesoscopic and microscopic scales. Notice that in our approach the solvent is treated as a continuum while the solute is explicitly resolved. One could use a coarse-grained treatment for the solute by including the appropriate non-electrostatic and electrostatic interactions in (8.5).

The solution of (8.16) requires an appropriate parametrization, i.e. coordinate representation, for the curvatures H and K, such that the equation is expressed as a function of the vector $\vec{r}$ and its first and second derivatives in space. Analytical solutions to the much simpler (8.17) and

thus to (8.16) are only available for systems with very simple geometries.[149] Thus we use numerical solutions of (8.16) in the following to further illustrate our theory.

8.3 Applications

First, we will consider the solvation of microscopic solutes such as noble gases, simple alkanes, and ions which can be treated as neutral or charged Lennard-Jones spheres. Then, we will investigate alkane assemblies on a larger scale, where interfacial and dewetting effects are much more dominant. For simplicity, mobile ions (counterions, salt) will be neglected in these illustrations.

8.3.1 One Lennard-Jones Sphere

In this section we compare our approach to published simulation results using explicit water models, such as SPC and SPC/E.[167] We refrain from comparing to real experiments for two reasons: first, approximations in computer experiments are more easily controlled and second, the LJ parameters of the solutes are commonly parametrized to yield accurate results in classical computer simulations.

For a spherical solute with a charge Q homogeneously distributed over its surface, the functional (8.5) with approximations (8.11) and (8.12) and no mobile ions reduces to a function of R, the radius of the sphere empty of solvent. The solvation free energy is

$$\begin{aligned}\Delta G(R) &= \Delta G_{\mathrm{pr}}(R) + \Delta G_{\mathrm{int}}(R) + \Delta G_{\mathrm{ne}}(R) + \Delta G_{\mathrm{es}}(R) \\ &= \frac{4}{3}\pi R^3 P + 4\pi R^2 \gamma_{\mathrm{lv}}\left(1 - \frac{2\delta}{R}\right) \\ &\quad + \int_R^\infty 4\pi r^2 \mathrm{d}r\ \rho_0 U_{\mathrm{LJ}}(r) \\ &\quad + \frac{Q^2}{8\pi\varepsilon_0 R}\left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_v}\right).\end{aligned} \tag{8.18}$$

Note that the last term in (8.18) is equivalent to the Born electrostatic solvation free energy. Differentiation of (8.18) with respect to R and subsequent division by $4\pi R^2$ yields

$$0 = P + \frac{2\gamma_{\mathrm{lv}}}{R}\left(1 - \frac{\delta}{R}\right) - \rho_0 U_{\mathrm{LJ}}(R) - \frac{Q^2}{32\pi^2 \varepsilon_0 R^4}\left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_v}\right), \tag{8.19}$$

which is in accord with Eq. (8.16) given sphere-like curvatures, $H = -1/R$ and $K = 1/R^2$. We can now calculate the solvation free energies of simple spherical solutes, such as noble gases or ions, by finding the radius $R_{\min}$ which minimizes (8.19). The free parameters are the pressure P, Tolman length δ, liquid-vapor surface tension γ_{lv}, and dielectric constants ε_v and ε_l.

8.3.2 One Neutral LJ Sphere

First, let us focus on uncharged spheres, for which the electrostatic term in (8.18) can be neglected. We compare the results from our theory to those calculated by Hummer *et al.* [168] for neutral LJ spheres in SPC water, and those calculated by Paschek[169] for noble gases in SPC and SPC/E water. The solute-water LJ parameters σ and ε are summarized in Table 8.1. The surface tension γ_{lv} was set to that estimated for SPC and SPC/E water at 300K, $\gamma_{\mathrm{lv}} = 65\,\mathrm{mJ/m}^2$ and $\gamma_{\mathrm{lv}} = 72\,\mathrm{mJ/m}^2$, respectively.[161, 170] The pressure is fixed to 1atm. Finally, the remaining free parameter δ was fit to reproduce the simulation solvation free energies. The solvation free energies from simulation ΔG_{sim} and best fit Tolman lengths δ_{bf} are shown in tables 8.1 and 8.2 for the SPC and SPC/E models, respectively.

Before we discuss the results let us show the particular energy contributions $\Delta G_i(R)$ with $i = \mathrm{pr,int,ne}$ for one chosen example, a neutral LJ-sphere with sodium-water LJ-

parameters, Na^0, plotted in Figure. 8.1. As anticipated, the pressure term $\Delta G_{pr}(R)$ with $P = 1$ atm is negligible compared to the other contributions. The interfacial term $\Delta G_{int}(R)$ is an increasing function of the cavity radius R trying to decrease cavity size. The integrated LJ-interaction term $\Delta G_{ne}(R)$ shows long-range attraction and a steep short-ranged repulsion with a minimum at $R = \sigma(Na^0) = 2.85$ Å. The total solvation free energy for the Na^0 shows a single minimum at $R_{min} = 2.32$ Å with $\Delta G = 9.2$ kJ/mol for a δ_{bf} =0.79Å.

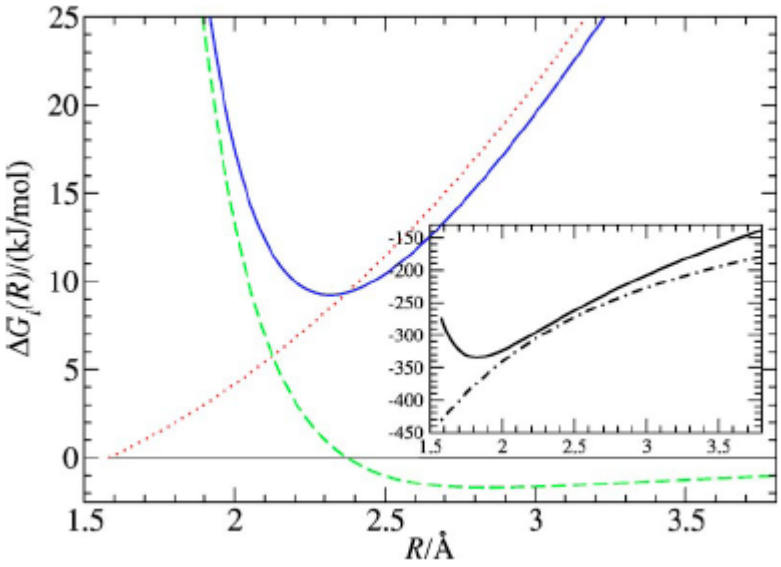

Figure 8.1: The particular solvation energy contributions $\Delta G_i(R)$ with $i = p, int, ne$ in equation 8.8 for one LJ sphere with Na^0 parameters given in Table 8.1. The pressure term ΔG_{pr} (thin solid line) with $P = 1$ atm is basically zero on this scale. The interfacial term $\Delta G_{int}(R)$ (dotted line) with $\gamma_{lv} = 65$ mJ/m^2 increases with radius R. The LJ term $\Delta G_{ne}(R)$ is given by the dashed line. The sum of the three contribution gives the total $\Delta G(R)$ (solid line) with a minimum at $R_{min} = 2.32$ Å for the uncharged sodium Na^0. The inset shows the electrostatic contribution $\Delta G_{es}(R)$ (dot-dashed line) and the total $\Delta G(R)$ for the charged Na^+ with a minimum at $R_{min} = 1.83$ Å. The best fit Tolman length is $\delta = 0.79$ Å in this example.

The results for the other LJ-spheres, summarized in tables 8.1 and 8.2, reveal several noteworthy observations. First, the best fit Tolman lengths δ_{bf} range from 0.76Å to 1.00Å;

they are not only of molecular size, as expected, but are approximately half the LJ-radius of a SPC or SPC/E water molecule. Moreover, the δ_{bf} values for noble gases in SPC/E water (Table 8.2) are approximately 10% larger than those in SPC water (Table 8.1). This is in qualitative agreement with Huang *et al.* who measured $\delta = 0.76 \pm 0.05$ Å and $\delta = 0.90 \pm 0.03$ Å for SPC and SPC/E, respectively, by fitting equation 8.11 to the hydration free energy of hard spheres with varying radii.

Second, it seems that on average δ_{bf} also differs between the works of Hummer and Paschek for the same water model, showing that δ may be a sensitive quantity. When further inspecting the quite accurate data of Paschek, we observe a systematic increase of δ_{bf} with solute size. The inability of our theory to be fit by one fixed constant δ_{bf} points to the already expected fact that equation 8.11 can not capture strong curvature effects accurately and will have to be refined for small solutes. Despite this shortcoming, these results show surprisingly good agreement; if we assume a fixed delta, for instance $\delta = 0.91$ Å for all noble gases in the SPC data of Paschek, the theoretical prediction would give results within 15% of the simulation data, thereby providing a quick and reasonable prediction of the solvation free energy. Finally, we observe that the effective optimal sphere radius $R_{\min}$ is always smaller than the radius of the canonical SAS with a typical probe radius of 1.4Å,[171] $R_{\min} < (\sigma_{ss}/2 + 1.4\text{Å}) \simeq \sigma$, but larger than the vdW surface, $R_{\min} > \sigma_{ss}/2$, where σ_{ss} is the solute-solute LJ-length.

Table 8.1: Solute-water LJ parameters and solvation free energy ΔG_{sim} for neutral Lennard-Jones spheres from the SPC water simulations performed by Hummer *et al.[168]* and Paschek.[169] δ_{bf} is the Tolman length best fit to ΔG_{sim} (rounded to two digits after the decimal point). R_{min} is the resulting optimal radius excluded of solvent. Also shown are the values for the simple alkanes methane (Me), ethane, propane, and butane from the study of Ashbaugh *et al.[128]* Simulation errors are given in parentheses.

Solute	ϵ (kJ/mol)	σ (Å)	ΔG_{sim} (kJ/mol)	δ_{bf} (Å)	R_{min} (Å)
SPC	0.65	3.17	⋯	⋯	⋯
SPC/E	0.65	3.17	⋯	⋯	⋯
Na^0	0.2005	2.85	9.2 (1)	0.79	2.32
K^0	0.0061	4.52	23.7 (5)	0.76	2.83
Ca^0	0.6380	3.17	10.2 (3)	0.85	2.80
F^0	0.5538	3.05	9.7 (2)	0.85	2.68
Cl^0	0.5380	3.75	21 (3)	0.80	3.30
Br^0	0.4945	3.83	24 (3)	0.77	3.35
Ne	0.3156	3.10	11.41 (0.05)	0.84	2.61
Ar	0.8176	3.29	8.68 (0.08)	0.90	2.96
Kr	0.9518	3.42	8.12 (0.1)	0.91	3.10
Xe	1.0710	3.57	7.65 (0.15)	0.92	3.27
Me	0.8941	3.44	10.96 (0.46)	0.85	3.11
Ethane	⋯	⋯	10.75 (0.50)	0.87	⋯
CH_3	0.7503	3.46	⋯	⋯	⋯
Propane	⋯	⋯	13.81 (0.54)	0.94	⋯
Butane	⋯	⋯	14.69 (0.54)	0.96	⋯
CH_2	0.5665	3.52	⋯	⋯	⋯
CH_3	0.6900	3.52	⋯	⋯	⋯

Table 8.2: Solvation free energies for neutral Lennard-Jones spheres in SPC/E water from the simulations of Paschek.[169] δ_{bf} and R_{min} are defined as in Tab. I.

Atom	ΔG_{sim} (kJ mol^{-1})	δ_{bf} (Å)	R_{min} (Å)
Ne	11.65 (0.05)	0.88	2.60
Ar	8.83 (0.08)	0.96	2.94
Kr	8.20 (0.1)	0.98	3.09
Xe	7.58 (0.15)	1.00	3.25

8.3.3 One Charged LJ Sphere

Let us now turn to charged Lennard-Jones spheres (ions) also examined in the paper by Hummer *et al.* with SPC water simulations. We assume δ to be the previously obtained δ_{bf} values for uncharged spheres. The dielectric constants are set to $\varepsilon_v = 1$ and $\varepsilon_l = 65$ in accord with SPC water.[172] The electrostatic contribution $\Delta G_{es}(R)$ and the total $\Delta G(R)$ are exemplified for Na^+ in the inset of Figure 8.1 The electrostatic contribution shifts the optimal radius to a smaller $R_{min} = 1.83$ Å compared to the uncharged case giving a solvation free energy of $\Delta G = -334$ kJ/mol. In fact, the optimal sphere radius R_{min} is always considerably smaller for the charged solutes, shown in Table 8.3, than for their neutral counterparts (Table 8.1). This is caused by the strong compressing force of the polar solvent attempting to penetrate the low dielectric cavity. The results for ΔG from theory are compared to those from simulation for all ions investigated by Hummer *et al.* (also shown in Table 8.3). While the theory describes the hydration free energies for positively charged ions within 15%, it considerably underestimates those of the negative ions. This qualitative disagreement between positive and negative ions could be anticipated considering that the fitted Born radii for anions are always smaller than those for cations, a consequence of the different solvation structure around charged solutes with

opposite signs. As mentioned in the previous section, the position of the dielectric boundary has to be refined for accurate estimates of the electrostatic contribution to the hydration free energy. If we apply the correction (8.13) to the dielectric boundary with a crude potential-independent shift $\xi_+ = -0.25\,\text{Å}$ for positive and $\xi_- = -1.05\,\text{Å}$ for negative spheres, implying that the dielectric boundary sphere has a radius $R_b = R + \xi_\pm < R$, all simulation values can be reproduced within 10% by our approach!

Table 8.3. Solvation free energies for charged LJ spheres in SPC water from the simulations of Hummer et al. compared to the theoretical result ΔG. For δ we use the best fits δ_{bf} to the solvation of neutral spheres as shown in Table.8.1

Ion	q	ΔG_{sim} (kJ mol^{-1})	ΔG (kJ mol^{-1})	ΔG_ξ (kJ mol^{-1})	R_{min} (Å)
Na^+	1	−398	−334	−394	1.83
K^+	1	−271	−246	−282	2.35
Ca^{2+}	2	−1306	−1181	−1364	2.09
F^-	−1	−580	−274	−630	2.25
Cl^-	−1	−371	−198	−342	2.97
Br^-	−1	−358	−192	−328	3.02

8.3.4 Linear Alkanes

Let us now consider simple polyatomic molecules, such as ethane, propane, or butane in a one-dimensional chain conformation, which can be treated as cylindrically symmetric geometries. Other conformations will be neglected. The cylindrical symmetry allows us to express the volume exclusion function $v(\vec{r})$ of the enveloping surface by a one dimensional shape function $r(z)$, where z is the coordinate on the symmetry axis and r the radial distance to it. The full surface in three-dimensional space is obtained by revolving the shape function

$r(z)$ around the symmetry axis. In our parametrization we express $r = r(l)$ and $z = z(l)$ as functions of the parameter l. The principal curvatures are then given by[173]

$$\kappa_1(r,z) = \frac{-z'}{r\sqrt{z'^2 + r'^2}}, \kappa_2(r,z) = \frac{z'r'' - z''r'}{(z'^2 + r'^2)^{3/2}}, \quad 8.20$$

where the primes indicate the partial derivative with respect to l. Additionally, the unit normal vector reads

$$\vec{n}(r,z) = \frac{1}{\sqrt{z'^2 + r'^2}} \begin{pmatrix} z'(2) \\ -r' \end{pmatrix}. \quad 8.21$$

The differential equation (8.16) is then solved by a forward relaxation scheme in time t

$$\begin{pmatrix} r(t+\Delta t)(4) \\ z(t+\Delta t) \end{pmatrix} = \begin{pmatrix} r(t)(5) \\ z(t) \end{pmatrix} - \Delta t\, \vec{n}(r,z)\mathrm{de}(r,z), \quad 8.22$$

where the steady-state solution $\partial(r,z)/\partial t = 0$ is the solution of $\mathrm{de}(r,z) = 0$ we are looking for. In the numerical calculations we use a grid of 500 bins and an integration time step of $\Delta t = 0.001$. The first and second derivatives are approximated using a symmetric two and three-step finite difference equation, respectively. Convergence is usually reached after 10^5 time steps. The result is observed to be independent of the initial choice of $r(z)$ at $t = 0$.

The LJ parameters for ethane and methane are the same as those used by Ashbaugh *et al.* [128] in their SPC simulation of linear alkanes (see Table 8.1). The simulation solvation energy of the spherical methane, $\Delta G = 10.96$ kJ/mol, can be reproduced with a fit $\delta = 0.85$ Å. Solving the cylindrically symmetric problem for ethane using the same δ, we obtain a fit-parameter-free $\Delta G = 11.40$ kJ/mol, which is only 7% larger than the simulation results. Alternatively, the best fit $\delta = 0.87$ Å reproduces the simulation energy exactly. This is surprisingly good agreement given the crude curvature correction we apply and the fact that the large curvature of the system varies locally in space. The curvature and shape functions are plotted in Figure 8.2 together with

the vdW surface and the canonical SAS obtained from rolling a probe sphere with the typically chosen radius $r_p = 1.4$ Å over the vdW surface.[171] Away from the center of mass $|z| 1$ Å the curvatures follow the expected trends for the spherical surfaces: $H \simeq -1/R_{min}$ and $K \simeq 1/R_{min}^2$ with $R_{min} \simeq 3.1$ Å. The optimal surface resulting from our theory is smaller than the canonical SAS and smooth at the center of mass ($z = 0$) where the canonical SAS has a kink. Thus our surface has a smaller mean curvature at $z = 0$ and an almost zero Gaussian curvature, which is typical for a cylinder geometry with one of the principal curvatures equal to zero. These results may justify the use of smooth surfaces in coarse-grained models of closely-packed hydrocarbons, a possibility we will explore in the following section with solvation on larger length scales where dewetting effects can occur. If we repeat the above calculation for propane and butane (three and four LJ-spheres, see also Table 8.1 for parameters) we need a best fit of $\delta = 0.94$ Å and $\delta = 0.96$ Å, respectively, to reproduce the simulation results exactly. The growing difference to the best fits for methane and ethane can be probably explained by the fact that we perform our calculation only for the cylindrically symmetric conformation of the linear alkanes; other conformations are neglected which will change the results.

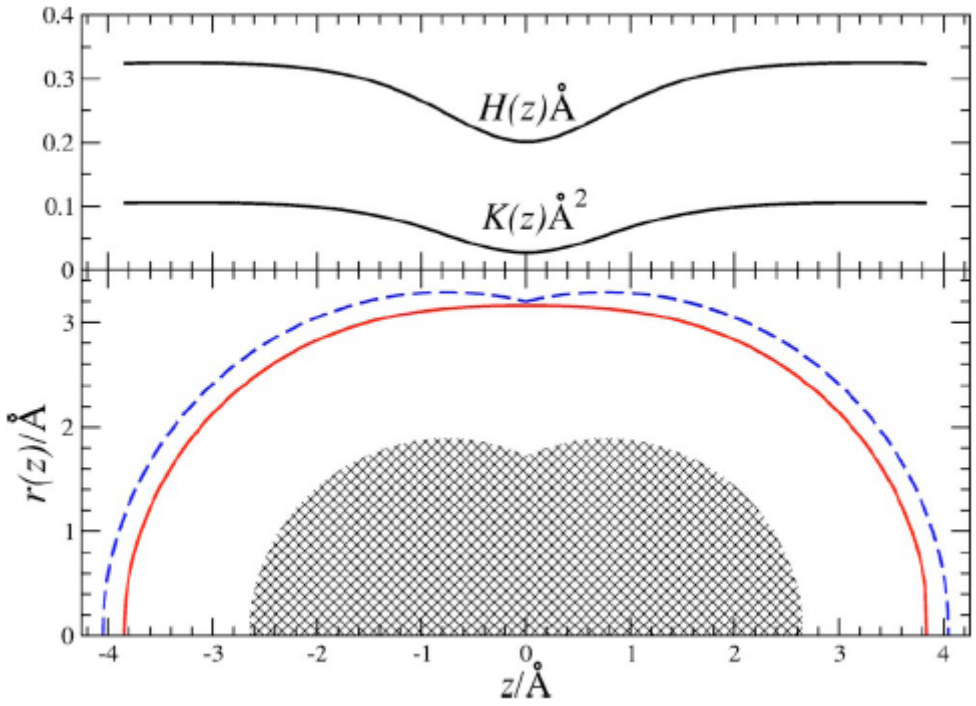

Figure 8.2: Mean $H(z)$ and Gaussian $K(z)$ curvature and shape function $r(z)$ (solid lines) for ethane. The canonical SAS (dashed line) from rolling a probe sphere with radius $r_p = 1.4$ Å over the vdW surface (shaded region) is also shown. The vdW surface is defined by the solute-solute LJ-radius $\sigma_{ss}/2 = 1.73$ Å.

8.3.5 Two Spherical Nanosolutes

Let us now consider two spherical solutes which we assume to be homogeneously assembled CH_2 groups with a uniform density ρ=0.024Å^{-3} up to a radius $R_0 = 15$ Å, defined by the maximal distance between a CH_2 center and the center of the solute. Integration of the CH_2-water LJ interaction over the entire sphere yields a 9-3 like potential $U_i(r)$ for the interaction between the center of the solute ($i = 1,2$) and a water molecule.[127] The intrinsic, nonelectrostatic solute-solute interaction at a surface-to-surface distance s_0, $U_{ss}(s_0)$, can be obtained in a similar fashion. The CH_2-water LJ parameters, $\varepsilon = 0.5665$ kJ/mol and $\sigma = 3.52$ Å, are taken from the OPLSUA force-field[174] and are similar to those used by

Huang et al. in their study on dewetting between paraffin plates.[135] Minimizing equation 8.18 for just one sphere we obtain an optimal radius excluded for water $R_{\min} \simeq R_0 + 2.4 = 17.4$ Å. Since we are also interested in the effects of electrostatic interactions we place opposite charges $\pm Ze$, where e is the elementary charge, in the center or on the edge of the two spheres.

Since we neglect mobile ions in our work, the PB equation reduces to Poisson's equation. It is solved on a two dimensional grid in cylindrical coordinates r and z with a finite difference method. The gradient and Laplacian are given then by $\nabla = (\partial r, \partial z)$ and $\Delta = \partial_r + \partial_r / r + \partial_z^2$, respectively. The first and second derivatives are approximated using symmetric two or three-step finite-difference equations. An explicit, forward time relaxation scheme is used to find the solution of Poisson's equation:

$$\Psi(t + \Delta t; \vec{r}) = \Psi(t; \vec{r}) - \Delta t \mathrm{PB}(\Psi(t; \vec{r})). \qquad 8.23$$

In most cases we use a lattice spacing of $\Delta r = \Delta z = 0.4$ Å on a $n_r \times n_z = 100 \times 200$ grid, and an integration time step $\Delta t = 0.05$. Convergence takes approximately 10^5 time steps. For the charged particles which are buried in the nanosolutes we use homogeneously charged spheres with a radius of 2Å. Instead of a sharp transition for the dielectric boundary [127], we use a smoothing function for reasons of numerical stability:

$$\varepsilon(\vec{r}) = \frac{\varepsilon_l - \varepsilon_v}{\exp(\kappa d(\vec{r})) + 1} + \varepsilon_v, \qquad 8.24$$

where the absolute value of the length $d(\vec{r})$ is given by the nearest distance to the boundary of the volume exclusion function $v(\vec{r})$. d is defined to be positive when $\vec{r} \in V$ and negative elsewhere. The inverse length κ defines the width of the boundary region and in the limit $\kappa \to \infty$ we recover the sharp transition [127]. We choose a value $\kappa 3 \text{Å}^{-1}$ for which the solution of Poisson's equation becomes basically independent of the choice of κ. An example for the

dielectric boundary is shown in Figure.8. 3 for two partly dewetted nanosolutes of radius $R_0 = 15\,\text{Å}$ at a distance $s_0 = 7\,\text{Å}$ carrying a charge $Q = 5e$ (system V in sec. III.C).

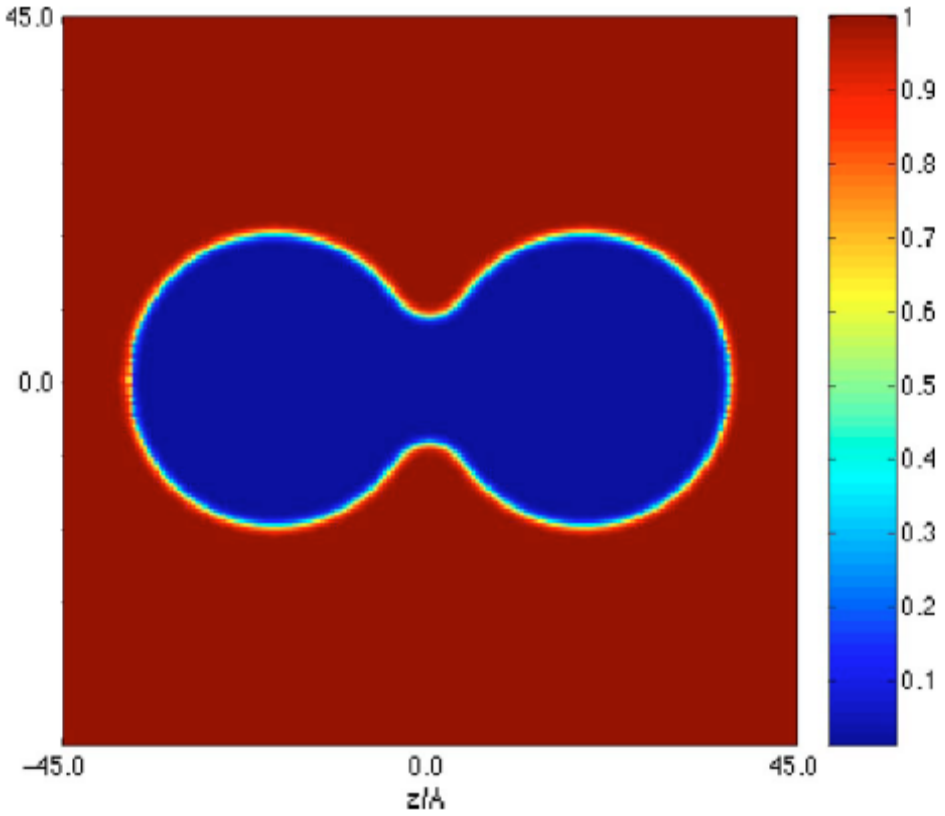

Figure 8.3: Distribution of the dielectric constant in space for two nanosolutes with $R_0 = 15\,\text{Å}$ at a distance $s_0 = 7\,\text{Å}$ carrying a charge $Q = 5e$ (system V). The region between the spheres is dewetted. The distribution is scaled by $\varepsilon_l = 78$.

In order to obtain the optimal shape function $v_{\min}(\vec{r})$ the shape equation (8.16) has to be solved simultaneously with Poisson's equation when the solutes are charged. In practice, we first solve (8.16) without any electrostatic contributions. In the second step, we solve Poisson's equation with the dielectric boundary (8.24) given by the volume exclusion function of the former solution. The result for the electric energy density is then plugged back in the shape equation in the third step. The last two steps are repeated until the solution for $v_{\min}(\vec{r})$ is fully

converged. Since the results for $r(z)$ excluding and including electrostatics are quite similar for our systems, full convergence takes usually only 6 to 7 repetitions of the described iteration steps.

The solvation of the two solutes is studied for a fixed surface-to-surface distance which we define as $s_0 = r_{12} - 2R_0$, where r_{12} is the solute center-to-center distance. The effective surface-to-surface distance defined by the accessibility of the solvent centers is thus $s \simeq r_{12} - 2R_{\min} = s_0 - 4.8$ Å. In the following we focus on a separation distance of $s_0 = 8$ Å to investigate the influence of different energetic contributions on the shape function, $r(z)$, and the curvatures, $K(z)$ and $H(z)$. For $s_0 = 8$ Å, it follows that $s \simeq 3.2$ Å, such that two water molecules could fit between the solutes on the z-axis. We systematically change the solute-solute and solute-solvent interactions, as summarized in Table 8.1. We begin with only the repulsive part of the nonelectrostatic interaction $U_i(r)$ in system I, and then adding a curvature correction with $\delta = 0.75$ Å, vdW attractions, and sphere-centered charges $Z = 4$ and $Z = 5$ in systems II-V, respectively. To study the influence of charge location, we shift each charge to the edge of the spheres such that they are 8 Å apart and reduce their magnitude to $Z = 1$ in system VI. The surface tension and dielectric constant of the vapor and liquid are fixed to $\gamma_{\mathrm{lv}} = 72$ mJ/m^2, $\varepsilon_v = 1$, and $\varepsilon_l = 78$, respectively.

8.3.6 Behavior of Shape Function

Table 8.4: Studied systems for two alkane-assembled spherical solutes. $W(s_0)$ is the inter-solute pmf. If $r(z=0)\neq 0$ the system is 'dewetted'. In system VI the solutes' charge is located off-center (oc) at the solute surface.

System	δ (Å)	vdW attraction	Z	$W(s_0)/k_BT$	dewetted
I	0.00	no	0	−57.6	yes
II	0.75	no	0	−34.1	yes
III	0.75	yes	0	−6.3	yes
IV	0.75	yes	4	−9.2	yes
V	0.75	yes	5	−5.1	no
VI	0.75	yes	1 (oc)	−1.3	no

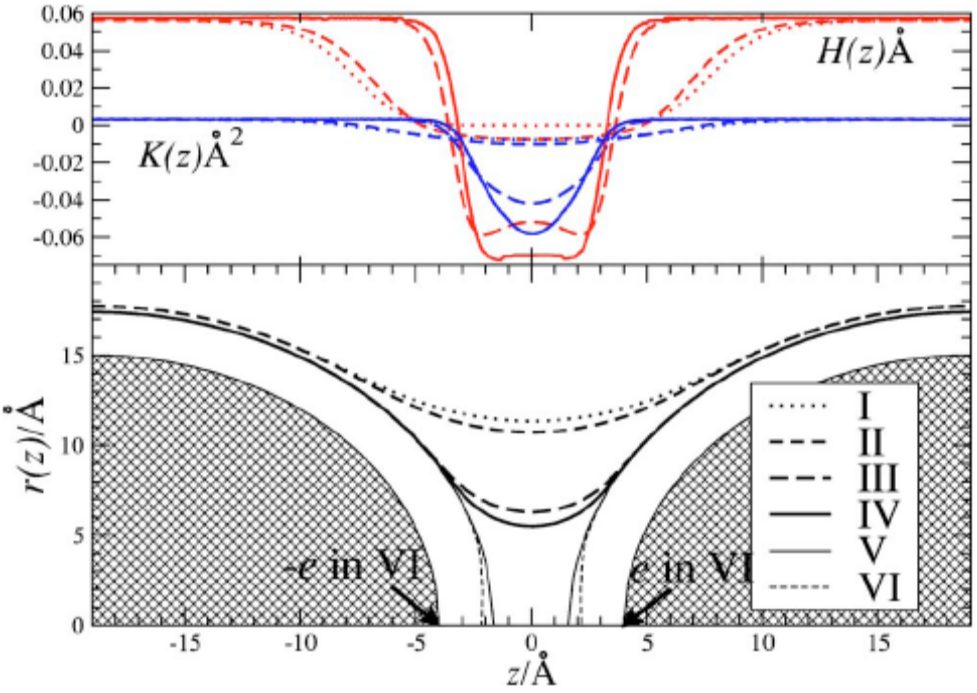

Figure 8.4: Mean $H(z)$ and Gaussian $K(z)$ curvatures and shape function $r(z)$ for two alkane-assembled solutes of radius $R_0 = 15$ Å (shaded region) at a distance $s_0 = 8$ Å for systems I-VI. Curvatures are not shown for the 'wet' systems V and VI.

The results for the curvatures and shape function, defined by $r(z)$, for systems I-VI are shown in Figure 8.3. Away from the center of mass ($|z|10$ Å) systems I-VI show very little difference. The curvatures are $H \simeq -1/R_{\min}$ and $K \simeq 1/R_{\min}^2$ with $R_{\min} \simeq 17.4$ Å. Close to the center of mass ($z \simeq 0$), however, the influence of changing the parameters is considerable. In system I, equation 8.16 reduces to the minimum surface equation $H(z) = 0$ for $z \simeq 0$. For two adjacent spheres the solution of this equation is the catenoid $r(z) \simeq \cosh(z)$, which features zero mean curvature (κ_1 and κ_2 cancel each other) and negative Gaussian curvature. As a consequence, the system exhibits a vapor bubble bridging the solutes. Water is removed from the surfaces between the spheres, although it fits in there. This dewetting is captured in our theory by the interfacial term G_{int} which always tries to minimize the liquid-vapor interface.

When curvature correction is applied (system II) the mean curvature becomes nonzero and positive (concave) at $z \simeq 0$, while the Gaussian curvature grows slightly more negative. Thus the total enveloping surface area becomes larger and the solvent inaccessible volume shrinks, i.e. the value of the shape function at $z \simeq 0$ decreases. Turning on solute-solvent dispersion attraction amplifies this trend significantly as demonstrated by system III. Mean and Gaussian curvatures increase fivefold, showing strongly enhanced concavity, and the volume empty of water decreases considerably, expressed by $r(z=0) \simeq 10.7$ Å dropping to $r(z=0) \simeq 6.3$ Å. These trends continue with the addition of electrostatics in system IV. When the sphere charges are further increased from $Z = 4$ to $Z = 5$ (system IV $\rightarrow$ V), we observe a wetting transition: the bubble ruptures and the shape function jumps to the solution for two isolated solutes, where $r(z \simeq 0) = 0$. The same holds when going from III to VI, when only one charge unit, $Z = 1$, is placed at each of the solutes' surfaces. Importantly, this demonstrates that the present formalism captures the sensitivity of dewetting phenomena to specific solvent-solute

interactions as reported in previous studies.[33, 135, 136, 141, 143, 144, 146] Note that the optimal shape function at $|z| \simeq \pm 2$ Å is closer to the solutes in VI compared to V due to the proximity of the charge to the interface. Clearly, the observed effects, particularly the transition from III to VI, cannot be described by existing solvation models which use the canonical SAS[119] or effective surface tensions and macroscopic solvent-solute contact angles[147, 149] as input.

8.3.7 Potential of Mean Force

The significant change of the shape function with the solute-solvent interaction has a strong impact on the potential of mean force (pmf) (or effective interaction) between the solutes

$$W(s_0) = G(s_0) - G(\infty) + U_{ss}(s_0). \qquad 8.25$$

Values of $W(s_0 = 8\text{Å})$ are given in Table 8.4. From system I to VI the total attraction between the solutes decreases almost two orders of magnitude. Interestingly, the curvature correction (I $\rightarrow$ II) lowers W by a large $23.5\,k_BT$, even though $R_{\min} \gg \delta$. The reason is that the mean radii of curvature between the spheres can assume values $\simeq \delta$, implying that curvature correction is also important for large solutes. A striking effect occurs when vdW contributions are introduced (II $\rightarrow$ III): the inter solute attraction decreases by approximately $28k_BT$ while the dispersion solute-solute potential $U_{ss}(s_0 = 8\text{Å})$ changes by only $-0.44k_BT$. Similarly, adding charges of $Z = 5$ (III $\rightarrow$ V) at the solutes' centers or $Z = 1$ (III $\rightarrow$ VI) at the solutes' surfaces decreases the total attraction by $1.2\,k_BT$ and 5k$_B$T, respectively. Note that the total attraction *decreases* even though electrostatic attraction has been added between the solutes. The same trend has been observed recently in explicit water simulations of a similar system of charged hydrophobic nanosolutes.[144, 145]

Now we turn our attention to varying the intersolute distance. The pmfs and mean force $F = \partial W(s_0)/\partial s_0$ between the solutes versus a whole range of s_0 for systems I,II,III, and VI are shown in Figure 8.5. System I, with purely repulsive solute-solvent interactions, displays a strong attraction ($W \simeq -150 k_B T$) at $s_0 = 2\,\text{Å}$ which decreases, almost linearly, to zero at a distance $s_0 = 13.5\,\text{Å}$ where the system shows a wetting transition. The corresponding force is discontinuous at this critical distance. The steep repulsion at short intersolute distances ($s_0 \simeq 1.5\,\text{Å}$) stems from the repulsive term of the LJ interaction between the solutes. The intrinsic solute-solute interaction $U_{ss}(s_0)$, shown in Figure 8.5, is almost two orders of magnitude smaller than the hydrophobic attraction. Adding the curvature correction in system II decreases the range and strength of the pmf by approximately 20%, which is significant and unexpected since $R_{\min} \gg \delta$.

Adding dispersion attractions in system III decreases the range and strength of the hydrophobic attraction considerably, but it is still much stronger than the inter solute dispersion attraction U_{ss} alone. When surface charges ($Z = 1$) are added in system VI, the range of hydrophobic attraction further decreases but the total attraction increases at short intersolute distances. This is due to the increasing size of the bridging bubble ($r(z = 0)$ increases) as the two solutes approach each other, which decreases the high dielectric screening of the solute-solute electrostatic attraction. This again underlines the importance of coupling electrostatics and dewetting effects, as the electrostatic attraction (or repulsion) may be magnified by more than an order of magnitude when dewetting occurs. For charges with opposite sign this could be interpreted as the stabilization of a salt bridge due to dehydration.[175] Systems IV and V, not shown in Figure 8.5, exhibit the same qualitative behavior as system VI.

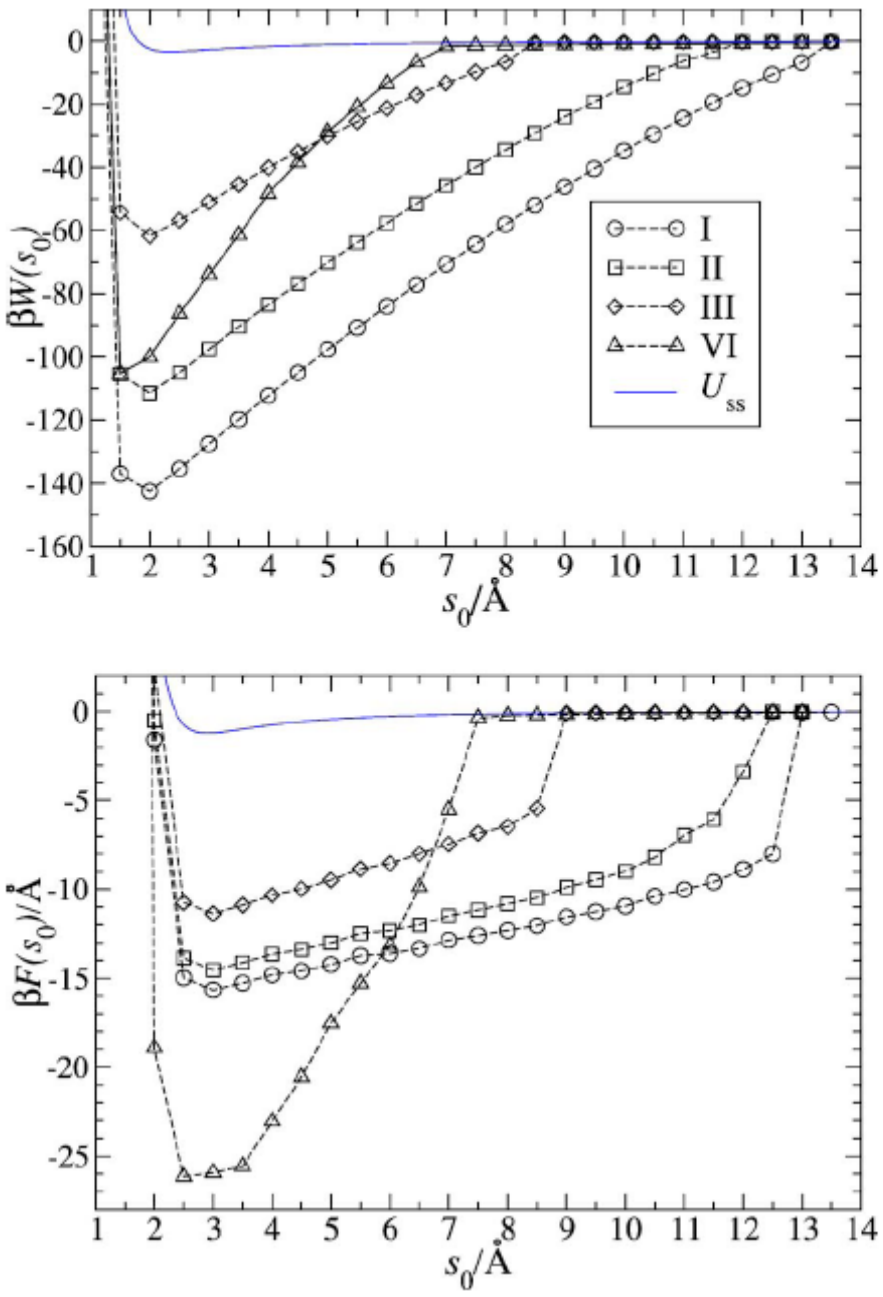

Figure 8.5: Top frame: pmfs for the systems I-III, and VI versus the solute distance s_0. Bottom frame: corresponding mean forces.

8.4 Conclusions

In summary, we have presented a novel implicit solvent model which couples polar and nonpolar solvation contributions by employing a variational formalism in which the Gibbs free

energy of the system is expressed as a functional of the solvent volume exclusion function. Minimization of the free energy leads to a Laplace-Young like equation for the solvent excluded cavity around the solutes, which is extended to describe solvation on mesoscopic and microscopic scales. We have shown that the theory gives a reasonable description of the solvation of microscopic solutes, such as ions and alkanes. Improved accuracy will require further refinement of the curvature dependence of the surface tension $\gamma(\vec{r};[v])$ and the definition of the position-dependent dielectric constant $\varepsilon(\vec{r};[v])$. Extensions based on physical rational and further empirical refinements could lead to an accurate implicit solvent description with only a few or none fit-parameters.

We have further demonstrated that on larger scales, where solvent dewetting can play an important role in solvation, our formalism captures the delicate balance between hydrophobic, dispersive and electrostatic forces which has been observed in previous systems.[130, 148, 159, 168, 176] The dewetting in our model is captured by the interfacial term which always tries to minimize the solvent interface. A comment must be made here regarding the sensitivity of dewetting to the particular solvent-solute interactions. As already argued by Chandler[132] extended fluid interfaces near phase coexistence are often referred to as 'soft' because they can be deformed with only little or no free-energy change.[177] In our approach this sensitivity seems to be considered, because little changes of the constraints in the differential equation 8.16 for the shape function, given e.g. by the dispersion potential close to the solute surface, can lead to a major deformation or even rupture (wetting transition) of the dewetted region. As we have shown this can significantly change the pmf for the solutes. Thus we anticipate that slight changes in the geometry of the considered system, e.g. a slight concave or convex bending of two plate-like solutes,[135, 136] can lead to very different results for the dewetting magnitude and the pmf.

The current illustrations utilized spherical and cylindrical symmetries. More complex molecules, such as proteins, will require solving the full three dimensional problem. Numerical algorithms for the calculation of interface evolution for more complicated geometries are provided by efficient level-set methods or fast marching methods.[178] We believe that even in the full three-dimensional (3D) case, our method will be much more efficient than other mesoscopic approaches which partly resolve the water structure (e.g. the Lum-Chandler-Weeks (LCW) theory[130] or information theory[168, 176] as only a two-dimensional surface is sought rather than a 3D density distribution.

This chapter is a reprint in full of material that appeared in *Coupling nonpolar and polar solvation free energies in implicit solvent models.* Dzubiella J., J.M.J. Swanson, J.A. McCammon, J. Chem. Phys.12, 084905 (2006). I was a secondary researcher and author of this work.

Bibliography

1. McCammon, J.A., *Theory of biomolecular recognition.* Current Opinion in Structural Biology, 1998. **8**(2): p. 245-249.

2. Gilson, M.K., et al., *The statistical-thermodynamic basis for computation of binding affinities: a critical review.* Biophysical Journal, 1997. **72**(3): p. 1047-1069.

3. McQuarrie, D.A., *Statistical Mechanics*. 1976, London: Harper and Row.

4. Ajay and M.A. Murcko, *Computational methods to predict binding free energy in ligand-receptor complexes.* Journal of Medicinal Chemistry, 1995. **38**(26): p. 4953-4967.

5. Simonson, T., G. Archontis, and M. Karplus, *Free energy simulations come of age: protein-ligand recognition.* Accounts of Chemical Research, 2002. **35**(6): p. 430-437.

6. Straatsma, T.P. and J.A. McCammon, *Computational alchemy.* Annual Review of Physical Chemistry, 1992. **43**: p. 407-435.

7. Jorgensen, W.L., *The Many Roles of Computation in Drug Discovery.* Science, 2004. **303**(5665): p. 1813-1818.

8. Hermans, J. and L. Wang, *Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme.* Journal of the American Chemical Society, 1997. **119**(11): p. 2707-2714.

9. Shirts, M.R., et al., *Computing the free energy of binding of ligands to fkbp12 using worldwide distributed computing.* Journal of Chemical Theory and Computation, 2006. **(submitted)**.

10. Woo, H.-J. and B. Roux, *Calculation of absolute protein-ligand binding free energy from computer simulations.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6825-6830.

11. Woo, H.-J. and B. Roux.

12. Sharp, K.A. and B. Honig, J. Phys. Chem., 1990. **94**: p. 7684.

13. Baker, N.A., *Improving implicit solvent simulations: a Poisson-centric view.* Curr Opin Struct Biol, 2005. **15**(2): p. 137-43.

14. Bashford, D. and D. Case, Ann. Rev. Phys. Chem., 2000. **51**: p. 129-152.

15. Feig, M. and C.L. Brooks, 3rd, *Recent advances in the development and application of implicit solvent models in biomolecule simulations.* Curr Opin Struct Biol, 2004. **14**(2): p. 217-24.

16. Tanford, C. and J.G. Kirkwood, *Theory of protein titration curces. I. General equations for impenetratable spheres.* Journal of American Chemical Society, 1957. **79**: p. 5333-5339.

17. Chothia, C., *Hydrophobic bonding and accessible surface area in proteins.* Nature, 1974. **248**(446): p. 338-9.

18. Lee, M.S. and M.A. Olson, *Evaluation of Poisson Solvation Models Using a Hybrid Explicit/Implicit Solvent Method.* Journal of Physical Chemistry B, 2005. **109**: p. 5223-5236.

19. Massova, I. and P.A. Kollman, *Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding.* Perspectives in Drug Discovery and Design, 2000. **18**: p. 113-135.

20. Lin, J.H., et al., *Computational drug design accomodating receptor flexibility-the relaxed complex scheme.* Journal of the American Chemical Society, 2002. **124**: p. 5632-5633.

21. Steinberg, I. and H. Scheraga, *Entropy Changes Accompanying Association Reactions of Proteins.* Journal of Biological Chemistry, 1963. **238**: p. 172-181.

22. Petrone, P.M. and A.E. Garcia, *MHC-peptide binding is assisted by bound water molecules.* Journal of Molecular Biology, 2004. **338**(2): p. 419-435.

23. Roux, B. and T. Simonson, *Implicit Solvent Models.* Biophysical Chemistry, 1999. **78**: p. 1-20.

24. Rashin, A.A. and B. Honig, *Reevaluation of the Born Model of Ion Hydration.* Journal of Physical Chemistry, 1985. **89**(26): p. 5588-5593.

25. Roux, B., H.A. Yu, and M. Karplus, *Molecular-Basis for the Born Model of Ion Solvation.* Journal of Physical Chemistry, 1990. **94**(11): p. 4683-4688.

26. Lee, B. and F.M. Richards, *Interpretation of Protein Structures - Estimation of Static Accessibility.* Journal of Molecular Biology, 1971. **55**(3): p. 379-&.

27. Connolly, M.L., *Computation of Molecular Volume.* Journal of the American Chemical Society, 1985. **107**(5): p. 1118-1124.

28. Grant, J.A., B.T. Pickup, and A. Nicholls, *A smooth permittivity function for Poisson-Boltzmann solvation methods.* Journal of Computational Chemistry, 2001. **22**(6): p. 608-640.

29. Im, W., D. Beglov, and B. Roux, *Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation.* Computer Physics Communications, 1998. **111**(1-3): p. 59-75.

30. Gallicchio, E., 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, 2000. **104**: p. 6271-6285.

31. Wagoner, J.A. and N.A. Baker, *Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms.* 2006. **in press**.

32. Spolar, R.S., J.H. Ha, and M.T. Record, Jr., *Hydrophobic effect in protein folding and other noncovalent processes involving proteins.* Proc Natl Acad Sci U S A, 1989. **86**(21): p. 8382-5.

33. Liu, P., et al., Nature, 2005. **437**: p. 159.

34. Ashbaugh, H.S. and L.R. Pratt, Arxiv preprint physics/0307109, - arxiv.org, 2003.

35. Beveridge, D.L. and F.M. DiCapua, *Free-Energy via molecular simulation - applications to chemical and biomolecular systems.* Annual Review of Biophysics and Biophysical Chemistry, 1989. **18**: p. 431-492.

36. Aqvist, J., V.B. Luzhkov, and B.O. Brandsdal, *Ligand binding affinities from MD simulations.* Accounts of Chemical Research, 2002. **35**(6): p. 358-365.

37. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices.* Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.

38. Vorobjev, Y.N. and J. Hermans, *ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model.* Biophysical Chemistry, 1999. **78**(1-2): p. 195-205.

39. Luo, H.B. and K. Sharp, *On the calculation of absolute macromolecular binding free energies.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(16): p. 10399-10404.

40. Lazaridis, T., A. Masunov, and F. Gandolfo, *Contributions to the binding free energy of ligands to avidin and streptavidin.* Proteins-Structure Function and Genetics, 2002. **47**(2): p. 194-208.

41. Luo, R. and M.K. Gilson, *Synthetic adenine receptors: Direct calculation of binding affinity and entropy.* Journal of the American Chemical Society, 2000. **122**(12): p. 2934-2937.

42. Burkhard, P., P. Taylor, and M.D. Walkinshaw, *X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies.* Journal of Molecular Biology, 2000. **295**(4): p. 953-962.

43. Go, N. and H. Scheraga, *Analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules.* Journal of Chemical Physics, 1969. **51**: p. 4751-4767.

44. Frisch, M.J., et al. *Gaussian 98*. 1998 [cited.

45. Wang, J.M., W. Wang, and P.A. Kollman, *Antechamber: An accessory software package for molecular mechanical calculations.* Abstracts of Papers of the American Chemical Society, 2001. **222**: p. U403-U403.

46. Goodford, P.J., *A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.* J Med Chem, 1985. **28**(7): p. 849-57.

47. Vriend, G., *WHAT IF: A molecular modeling and drug design program.* J Mol Graph, 1990. **8**(1): p. 52-6, 29.

48. Case, D.A., et al. *AMBER 7*. 2002 [cited.

49. Baker, N.A., et al., *Electrostatics of nanosystems: Application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(18): p. 10037-10041.

50. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces.* Biopolymers, 1996. **38**(3): p. 305-20.

51. Steinberg, I. and H. Scheraga, *Entropy changes accompanying association reactions of proteins.* The Journal of Biological Chemistry, 1963. **238**(1): p. 172-181.

52. Finkelstein, A.V. and J. Janin, *The price of lost freedom - entropy of biomolecular complex-formation.* Protein Engineering, 1989. **3**(1): p. 1-3.

53. Tidor, B. and M. Karplus, *The contribution of vibrational entropy to molecular association - the dimerization of insulin.* Journal of Molecular Biology, 1994. **238**(3): p. 405-414.

54. Head, M.S., J.A. Given, and M.K. Gilson, *"Mining minima": direct computation of conformational free energy.* Journal of Physical Chemistry A, 1997. **101**(8): p. 1609-1618.

55. Finkelstein, A.V. and J. Janin, *The Price of Lost Freedom - Entropy of Biomolecular Complex-Formation.* Protein Engineering, 1989. **3**: p. 1-3.

56. Carlsson, J. and J. Aqvist, *Absolute and relative entropies from computer simulation with applications to ligand binding.* Journal of Physical Chemistry B, 2005. **109**: p. 6448-6456.

57. Swanson, J.M.J., 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, 2004. **86**(1): p. 67-74.

58. Tai, K., et al., *Mechanism of acetylcholinesterase inhibition by fasiculin: a 5-ns molecular dynamics simulation.* Journal of the American Chemical Society, 2002. **124**: p. 6153-6161.

59. Bui, J.M., K. Tai, and J.A. McCammon, *Acetylcholinesterase: enhanced fluctiations and alternative routes to the active site in the complex with fasiculin-2.* Journal of the American Chemical Society, 2004. **126**: p. 7198-7205.

60. Karplus, M. and J. Kuschick, *Method for Estimating the Configurational Entropy of Macromolecules.* Macromolecules, 1981. **14**: p. 325-332.

61. Chang, C.E., W. Chen, and M.K. Gilson, *Evaluating the Accuracy of the Quasiharmonic Approximation.* Journal of Chemical Theory and Computation, 2005. **1**: p. 1017-1028.

62. Dunitz, J., *The entropic cost of bound water in crystals and biomolecules.* Science, 1994. **264**: p. 670.

63. Jackson, J.D., *Classical Electrodynamics*. 1962: John Wiley and Sons, Inc.

64. Warwicker, J. and H.C. Watson, *Calculations of the electric protential in the active site cleft due to alpha-helix-dipoles.* Journal of Molecular Biology, 1982. **157**: p. 671-679.

65. Gouy, M., *Sur la constitution de la charge electrique a la surface d'un electrolyte.* Journal of Physics, 1910. **9**: p. 457-468.

66. Hirata, F., P. Redfern, and R.M. Levy, *Viewing the Born Model for Ion Hydration through a Microscope.* International Journal of Quantum Chemistry, 1988: p. 179-190.

67. Jayaram, B., et al., *Free-Energy Calculations of Ion Hydration - an Analysis of the Born Model in Terms of Microscopic Simulations.* Journal of Physical Chemistry, 1989. **93**(10): p. 4320-4327.

68. Sitkoff, D., K. Sharp, and B. Honig, *Accurate calculation of hydration free energies using macroscopic solbent models.* Journal of Physical Chemistry, 1994. **98**: p. 1978-1988.

69. Bordner, A.J., C.N. Cavasotto, and R.A. Anagyan, *Accurate Transferable Model for Water, n-Octanol, and n-Hexadecane Solcation Free Energies.* Journal of Physical Chemistry, 2002. **106**: p. 11009-11015.

70. Nina, M., D. Beglov, and B. Roux, *Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations.* Journal of Physical Chemistry B, 1997. **101**(26): p. 5239-5248.

71. Nina, M., W. Im, and B. Roux, *Optimized atomic radii for protein continuum electrostatics solvation forces.* Biophysical Chemistry, 1999. **78**(1-2): p. 89-96.

72. Shirts, M.R., et al., *Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins.* Journal of Chemical Physics, 2003. **119**(11): p. 5740-5761.

73. Deng, Y.Q. and B. Roux, *Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules.* Journal of Physical Chemistry B, 2004. **108**(42): p. 16567-16576.

74. Chipman, D.M., *Computation of pK(a) from dielectric continuum theory.* Journal of Physical Chemistry A, 2002. **106**(32): p. 7413-7422.

75. Beroza, P. and D.A. Case, *Calculations of proton-binding thermodynamics in proteins.* Energetics of Biological Macromolecules, Pt B, 1998. **295**: p. 170-189.

76. Lu, Q. and R. Luo, *A Poisson-Boltzmann dynamics method with nonperiodic boundary condition.* Journal of Chemical Physics, 2003. **119**(21): p. 11035-11047.

77. Mongan, J. and D.A. Case, *Biomolecular Simulations at Constant pH.* Current Opinions in Structural Biology, 2004.

78. Wagoner, J. and N.A. Baker, *Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson-Boltzmann models.* Journal of Computational Chemistry, 2004. **25**(13): p. 1623-1629.

79. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

80. Luo, R., L. David, and M.K. Gilson, *Accelerated Poisson-Boltzmann calculations for static and dynamic systems.* Journal of Computational Chemistry, 2002. **23**(13): p. 1244-1253.

81. Michael, F., J. Karanicolas, and C.L. Brooks, III, *MMTSB Toolset*. 2001, MMTSB NIH Research Resource: The Scripps Research Institute.

82. Dunbrack, R.L. and F.E. Cohen, *Bayesian statistical analysis of protein sidechain rotamer preferences* . Protein Science, 1997. **6**: p. 1661-1681.

83. Laskowski, R.A., et al., *PDBsum: a Web-based database of summaries and analyses of all PDB structures.* Trends in Biochemical Sciences, 1997. **22**(12): p. 488-490.

84. Beglov, D. and B. Roux, *Finite Representation of an Infinite Bulk System - Solvent Boundary Potential for Computer-Simulations.* Journal of Chemical Physics, 1994. **100**(12): p. 9050-9063.

85. Nina, M. and T. Simonson, *Molecular dynamics of the tRNA(Ala) acceptor stem: Comparison between continuum reaction field and particle-mesh Ewald electrostatic treatments.* Journal of Physical Chemistry B, 2002. **106**(14): p. 3696-3705.

86. Kumar, S., et al., *The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method.* Journal of Computational Chemistry, 1992. **13**(8): p. 1011-1021.

87. Dong, F., M. Vijayakumar, and H.X. Zhou, *Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar.* Biophysical Journal, 2003. **85**(1): p. 49-60.

88. Baker, J.E., *Reducing bias and inefficiency in the selection algorithm*, in *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*. 1987, Lawrence Erlbaum Associates, Inc: Mahwah. p. 14-21.

89. Tsui, V. and D.A. Case, *Molecular dynamics simulations of nucleic acids with a generalized born solvation model.* Journal of the American Chemical Society, 2000. **122**(11): p. 2489-2498.

90. Bondi, A., Journal of Chemical Physics, 1964. **64**: p. 441.

91. Roux, B. and D. Case, J.M.J. Swanson, Editor. 2004.

92. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, *Designing a 20-residue protein.* Nature Structural Biology, 2002. **9**(6): p. 425-430.

93. Gronenborn, A.M., et al., *A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein-G.* Science, 1991. **253**(5020): p. 657-661.

94. Tilton, R.F., J.C. Dewan, and G.A. Petsko, *Effects of Temperature on Protein-Structure and Dynamics - X-Ray Crystallographic Studies of the Protein Ribonuclease-a at 9 Different Temperatures from 98-K to 320-K.* Biochemistry, 1992. **31**(9): p. 2469-2481.

95. Lee, M.S., et al., *New analytic approximation to the standard molecular volume definition and its application to generalized born calculations (vol 24, pg 1348, 2003).* Journal of Computational Chemistry, 2003. **24**(14): p. 1348-1356.

96. Friedrichs, M., et al., *Poisson-Boltzmann analytical gradients for molecular modeling calculations.* Journal of Physical Chemistry B, 1999. **103**(16): p. 3057-3061.

97. Im, W.P., M.S. Lee, and C.L. Brooks, *Generalized born model with a simple smoothing function.* Journal of Computational Chemistry, 2003. **24**(14): p. 1691-1702.

98. Prabhu, N.V., P.J. 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, 2004. **25**(16): p. 2049-2064.

99. Swanson, J.M.J., S.A. Adcock, and J.A. McCammon, *Optimized Radii for Poisson-Boltzmann Calculations*

with the AMBER Force Field. Journal of Chemical Theory and Computation, 2005. **1**(3): p. 484-493.

100. Masunov, A. and T. Lazaridis, *Potentials of mean force between ionizable amino acid side chains in water.* Journal of the American Chemical Society, 2003. **125**(7): p. 1722-1730.

101. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium.* Journal of Physical Chemistry, 1996. **100**(51): p. 19824-19839.

102. Davis, M.E. and J.A. McCammon, *Electrostatics in biomolecular structure and dynamics.* Chemical Reviews, 1990. **90**: p. 509-521.

103. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**(5214): p. 1144-9.

104. Levy, R.M., et al., *On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy.* J Am Chem Soc, 2003. **125**(31): p. 9523-30.

105. Wang, J.M., P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* Journal of Computational Chemistry, 2000. **21**(12): p. 1049-1074.

106. Lwin, T.Z., R.H. Zhou, and R. Luo, *Is Poisson-Boltzmann theory insufficient for protein folding simulations?* Journal of Chemical Physics, 2006. **124**(3): p. -.

107. Wagoner, J.A. and N.A. Baker, Proceedings of the National Academy of Sciences of the United States of America, 2006. **editorial approval**.

108. Izvekov, S., et al., *Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching.* Journal of Chemical Physics, 2004. **120**(23): p. 10896-913.

109. Izvekov, S. and G.A. Voth, *A Multiscale Coarse-Graining Method for Biomolecular Systems.* Journal of Physical Chemistry B, 2005. **109**(7): p. 2469-73.

110. Cramer, C.J. and D.G. Truhlar, *General Parameterized Scf Model for Free-Energies of Solvation in Aqueous-Solution.* Journal of the American Chemical Society, 1991. **113**(22): p. 8305-8311.

111. Kelly, C.P., C.J. Cramer, and D.G. Truhlar, *SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters.* Journal of Chemical Theory and Computation, 2005. **1**(6): p. 1133-1152.

112. Wagoner, J.A. 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, editorial approval.

113. Dzubiella, J., J.M.J. Swanson, and J.A. McCammon, *Coupling nonpolar and polar solvation free energies in implicit solvent models.* Journal of Chemical Physics, 2006. **124**(8): p. -.

114. Dzubiella, J., J.M.J. Swanson, and J.A. McCammon, *Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models.* Physical Review Letters, 2006. **96**(8): p. -.

115. Pearlman, D.A., et al., *Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules.* Computer Physics Communications, 1995. **91**(1-3): p. 1-41.

116. Hodsdon, M.E. and D.P. Cistola, *Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by N-15 NMR relaxation and H-1 exchange.* Biochemistry, 1997. **36**(8): p. 2278-2290.

117. Onufriev, A., D.A. Case, and D. Bashford, *Effective Born radii in the generalized Born approximation: the importance of being perfect.* J Comput Chem, 2002. **23**(14): p. 1297-304.

118. Swanson, J.M.J., J. Mongan, and J.A. McCammon, *Limitations of atom-centered dielectric functions in implicit solvent models.* Journal of Physical Chemistry B, 2005. **109**(31): p. 14769-14772.

119. Richards, F.M., *Areas, volumes, packing and protein structure.* Annu Rev Biophys Bioeng, 1977. **6**: p. 151-76.

120. Gallicchio, E., M.M. Kubo, and R.M. Levy, J. Phys. Chem. B, 2000: p. 6271-6285.

121. Gallicchio, E., L.Y. Zhang, and R.M. Levy, J. Comput. Chem., 2002: p. 517-529.

122. Zacharias, M., J. Chem. Phys. A, 2003: p. 3000-3004.

123. Su, Y. and E. Gallicchio, Biophys. Chem., 2004: p. 251-260.

124. Levy, R.M., et al., Journal of American Chemical Society, 2003: p. 9523-9530.

125. Bashford, D. and D.A. Case, Ann. Rev. Phys. Chem, 2000. **51**: p. 129.

126. Baker, D., Curr. Opin. Struct. Bio, 2005. **15**: p. 137.

127. Huang, D.M. and D. Chandler, J. Phys. Chem. B, 2002. **106**: p. 2047.

128. Ashbaugh, H.S., E.W. Kaler, and M.E. Paulaitis, Biophys. J., 1998: p. 755-768.

129. Stillinger, F.H., J. Solution Chem., 1973: p. 141.

130. Lum, K., D. Chandler, and J.D. Weeks, J. Phys. Chem. B, 1999: p. 4570.

131. Hummer, G., L.R. Pratt, and A.E. Garcia, J. Chem. Phys, 1997: p. 9275.

132. Chandler, D., Nature, 2005. **437**: p. 640.

133. Wallquist, A. and B.J. Berne, J. Phys. Chem., 1995: p. 2893.

134. Huang, X., C.J. Margulis, and B.J. Berne, PNAS, 2003: p. 11953-11958.

135. Huang, X., R. Zhou, and B.J. Berne, J. Phys. Chem. B, 2005. **109**: p. 3546.

136. Choudhury, N. and B.M. Pettitt, J. Am. Chem. Soc., 2005. **127**: p. 3556.

137. Hummer, G., J.C. Rasaiah, and J.P. Nowortya, Nature, 2001: p. 188.

138. Allen, R., S. Melchionna, and J.P. Hansen, Phys. Rev. Lett., 2002: p. 175502.

139. Beckstein, O., P.C. Biggin, and M.S.P. Sansom, J. Phys. Chem., 2001: p. 12902.

140. Anishkin, A. and S. Sukharev, Biophys. J., 2004: p. 2883-2895.

141. Dzubiella, J., R.J. Allen, and J.P. Hansen, *Electric field-controlled water permeation coupled to ion transport through a nanopore.* J. Chem. Phys., 2004. **120**: p. 5001.

142. Dzubiella, J. and J.P. Hansen, *Electric-field-controlled water and ion permeation of a hydrophobic nanopore.* J. Chem. Phys., 2005. **122**: p. 234706.

143. Vaitheesvaran, S., J.C. Rasaiah, and G. Hummer, J. Chem. Phys., 2004: p. 7955-7965.

144. Dzubiella, J. and J.P. Hansen, *Reduction of hydrophoic attraction between charged solutes in water.* Journal of Chemical Physics, 2003. **119**: p. 12049-12052.

145. Dzubiella, J. and J.P. Hansen, *Competition of hydrophobic and Coulombinc interactions between nanosized solutes.* J. Chem. Phys., 2004. **121**: p. 5514.

146. Zhou, R., et al., Science, 2004. **305**: p. 1605.

147. Parker, J.L., P.M. Claesson, and P. Attard, J. Phys. Chem., 1994. **98**: p. 8468.

148. Beglov, D. and B. Roux, J. Chem. Phys., 1996: p. 8678.

149. Kralchevsky, P. and K. Nagayama, *Particles at Fluid Interfaces and Membranes*. 2001, Amsterdam: Elsevier.

150. Helfrich, W., Z. Naturforsch. C, 1973. **28**: p. 693.

151. Zhong-can, O.-Y. and W. Helfrich, Phys. Rev. A, 1989. **39**: p. 5280.

152. Bieker, T. and S. Dietrich, Physica A, 1998. **252**: p. 85.

153. Chou, T., Phys. Rev. Lett., 2001. **87**: p. 106101.

154. Allen, R., J.P. Hansen, and S. Melchionna, J. Phys.: Condens. Matt., 2003: p. 5297.

155. Attard, P., Langmuir, 2000. **16**: p. 4455.

156. Cheng, Y.K. and P.J. Rossky, Nature, 1998. **392**: p. 696.

157. Gerstein, M. and C. Chothia, Proc. Natl. Acad. Sci. USA, 1996. **93**: p. 10167.

158. Lau, K.K.S., et al., Nano Letters, 2003. **3**: p. 1701.

159. Triezenberg, D.G. and R. Zwanzig, Phys. Rev. Lett., 1972. **28**: p. 1183.

160. Gilson, M.K., et al., J. Phys. Chem, 1993. **97**: p. 3591.

161. Huang, D.M., P.L. Geissler, and D. Chandler, J. Phys. Chem. B, 2001: p. 6704.

162. Henderson, J.R., *Fundamentals of Inhomogeneous Fluids*, ed. D. Henderson. 1992, New York: Marcel Dekker.

163. Stewart, M.C. and R. Evans, Phys. Rev. E, 2005. **71**: p. 011602.

164. Reiss, H., Adv. Chem. Phys., 1965. **9**: p. 1.

165. Tolman, R.C., J. Chem. Phys., 1949: p. 333.

166. Luo, H. and C. Tucker, J. Phys. Chem., 1996. **100**: p. 11165.

167. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, J. Phys. Chem., 1987: p. 6269.

168. Hummer, G., L.R. Pratt, and A.E. Garcia, J. Phys. Chem, 1996: p. 1206.

169. Paschek, D., J. Chem. Phys., 2004. **120**: p. 6674.

170. Alejandre, J., D.J. Tildesley, and G.A. Chapela, J. Chem. Phys., 1995. **102**: p. 4574.

171. Connolly, M.L., J. Mol. Graphics, 1993. **11**: p. 139.

172. van der Spoel, D., P.J.v. Maaren, and H.J.C. Berendsen, J. Chem. Phys., 1998. **108**: p. 10220.

173. Frankel, T., *The Geometry of Physics*. 1997, Cambridge: Cambridge University Press.

174. Jorgensen, W., J.D. Madura, and C.J. Swenson, J. Am. Chem. Soc., 1984. **106**: p. 6638.

175. Fernandez, A., Nature Biotech., 2004. **22**: p. 1081.

176. Garde, S., et al., Phys. Rev. Lett., 1996. **77**: p. 4966.

177. Safran, S.A., *Statistical Thermodynamics of Surfaces, Interfaces and Membranes* 1994, Reading, MA: Addison-Wesley.

178. Sethian, J.A., *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. 1999, Cambridge: Cambridge University Press.