Molecular Simulation of Biomaterials and Biomolecules at the Solid-Liquid Interface

by

Stephen Thomas Kottmann

B.S. Chemistry/Physics
Eckerd College (2001)

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Author .....
Department of Chemistry
December 14, 2007

Certified by .....
Angela M. Belcher
Professor of Materials Science and Engineering
and Biological Engineering
Thesis Supervisor

Accepted by .....
Robert W. Field
Professor of Chemistry
Chairman, Committee on Graduate Studies

Thesis Committee Certification of Approval

of

Molecular Simulation of Biomaterials and Biomolecules at the Solid-Liquid Interface

by

Stephen Thomas Kottmann

Graduate Thesis Commitee

Angela M. Belcher
Professor of Materials Science and
Biological Engineering

Moungi G. Bawendi
Professor of Chemistry
Thesis Committee Chair

Robert G. Griffin
Professor of Chemistry

Molecular Simulation of Biomaterials and Biomolecules at the Solid-Liquid Interface

by

Stephen Thomas Kottmann

Submitted to the Department of Chemistry
on December 14, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemistry

Abstract

Biomaterials and biomineralization have been successfully utilized in a broad variety of technical applications. Properties of natural biopolymers, such as the ability to control the nucleation, growth, and organization of crystals, have been extended to a much wider array of technologically applicable materials through combinatorial selection techniques. However, detailed mechanisms of peptide adsorption on inorganic surfaces have largely escaped characterization. This knowledge would open new routes for the rational design of nanostructures and composite biomaterials. The development of accurate and computationally efficient methods for the simulation of biopolymer-inorganic surface adsorption could provide a more detailed understanding of adsorption mechanisms. While simple models involving reduced solvent representations and polymer flexibility have found some success in limited applications, robust performance for systems of varying size and composition can generally be expected only through accurate inclusion of these key details. Fully atomistic representations of biopolymer and surface are necessary for detailed molecular recognition, while polymer flexibility is required to capture structural rearrangment and the resulting free energy contributions. Finally, electrostatic interactions between the adsorbing biopolymer and inorganic surface, as well as interactions of the polymer and surface with the surrounding solvent environment will play a dominant role in the adsorption process, and an accurate representation of the solvated system is inherently necessary. Computational efficiency can be increased through the application of implicit solvent models, which replace the numerous solvent molecules with a continuum dielectric, and seek to capture the average effects of the statistical solvent environment. The Poisson-Boltzmann model represents the most rigorous treatment of implicit solvent. This model, however, requires the relatively expensive solution of a second order eliptical differential equation over the space of the system. Here, a method is presented which reduces the scale at which the Poisson-Boltzmann equation must be solved. However, even when combined with an efficient multi-grid solver, the Poisson-Boltzmann model represents a significant computational cost. The modified Generalized Born model, GBr^6, based on an approximation to the Poisson-Boltzmann

model, offers a computationally efficient alternative. Generalized Born models, however, are often inaccurate in the case of charges positioned near an extended dielectric interface, which is precisely the system we wish to investigate. Here, an analytical integration of the approximate electric displacement is used to calculate Born radii, and tested in application to surface adsorption studies. Replica-exchange Monte Carlo simulations with modified Generalized Born implicit solvent environment is then used to study the adsorption mechanism of a set of rationally designed sapphire-binding peptides. Modulation of binding affinity is predicted to depend on multiple interactions between basic amino acids and the negatively charged sapphire surface. The proximity of charged residues to one another as well as the conformational ability of each peptide to present functional groups towards the surface are shown to control the relative binding affinities.

Thesis Supervisor: Angela M. Belcher
Title: Professor of Materials Science and Engineering
and Biological Engineering

Acknowledgments

I have had the honor and pleasure of being surrounded by many great mentors, colleagues, and friends throughout my graduate career, whom I am indebted to for their extensive support and encouragement. Particularly I would like to thank my advisor, Angela Belcher, for her guidance and support, as well as the freedom to pursue my interests in research. In starting my graduate career at the University of Texas, I was fortunate to establish a relationship with Peter Rossky, whose generous counsel continued thoughout my time at MIT. I also extend my appreciation to my colleagues in the Belcher laboratory. The invaluable friendships of Daniel Solis, Brian Reiss, Eric Krauland, Andrew Magyar and Ahmad Khalil have formed the basis of my graduate school community. I am indebted to them for their support and advice in all aspects of life. I am grateful also for the financial support of the Army Research Office Institute of Collaborative Biotechnologies.

I have also been fortunate to be surrounded by many friends and family outside of school. It is hard to imagine being where I am today without the unfailing love and encouragement of my parents, Robert and Elizabeth Kottmann, and the example set by my brother R. Matthew Kottmann. I am truly in awe of the tremendous love and support of my fiancée, Sherri Havas, who has brought so much joy to my life. The outpouring of support from my friends and community at Christ the King has been overwhelming. Finally, and most importantly, all of my thanks go to God, who has blessed me with this life and guided me always with His steadfast love.

Proverbs 3:5-6

"Trust in the LORD with all your heart and lean not on your own understanding; in all your ways acknowledge Him, and He will make your paths straight."

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Biomineralization

Natural biopolymers, such as antifreeze and marine shell proteins, exhibit specific recognition of inorganic materials as well as control over the nucleation, growth and orientation of crystals[84, 91, 145, 146]. Protein systems are used in nature to both encourage and inhibit the growth of crystalline inorganic materials. Antifreeze proteins allow organisms to live in sub-freezing temperatures by binding to the surface of intracellular ice, truncating the crystals and preventing further growth. Alternatively, many organisms utilize proteins to nucleate and control the growth of inorganic materials. This is exemplified in the bones and teeth of vertibrate animals, as well as the mineral-protein composite found in many marine organism's protective shells. Figure 1-1 illustrates the exquisite control over crystal growth elicited by natural biopolymers[50]. Proteins from the gastropod mollusc abalone control the crystal growth of calcium carbonate, inducing the formation of nanoscale tablets which then stack to form the composite structure.

The general transfer of *in vivo* natural biomineralization processes to technologically applicable *in vitro* biomimetic mineralization techniques has been demonstrated through isolation of these mollusc shell proteins[106]. Polyanionic proteins isolated from calcite and aragonite phases of the abalone shell were demonstrated to nucleate the corresponding crystal phase from a solution of mineral precursors. As for

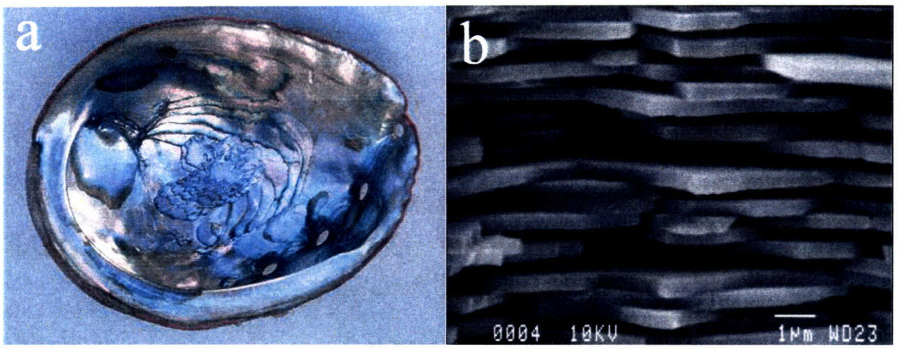

Figure 1-1: Biomineralization in (a) abalone shell. Proteins from the abalone control the crystal growth of calcium carbonate, forming a (b) "stack of coins" structure that is many times stronger than the geological mineral[30, 50]

technological applications, however, this calcium carbonate system represents only limited potential. In fact, the majority of naturual biomineralization systems consist of carbonates and phsophates of alkaline earth metals or oxides of iron and silicon. While iron and silicon oxides certainly have high technological applicability, it is desirable to extend this biomimetic process to encompass a wider variety of materials. To this end, a combinatorial selection process was developed[12] using commercially available phage display libraries. This process, illustrated in Figure 1-2, utilizes a repeated cycle of culling and amplification to select peptides which bind (specifically) to the material of interest.

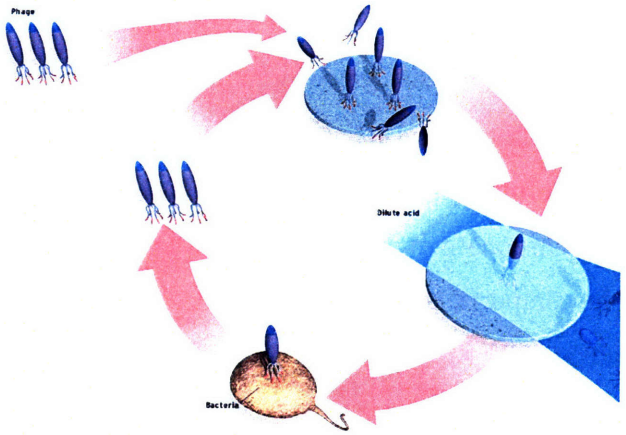

Figure 1-2: Biopanning for the selection of material specific peptides. Peptides are selected from an initial library of approximately 10^9 individual sequences through a repeated cycle of culling and amplification.

The applications of this scheme are extensive, ranging from the imobilization of proteins on crystalline substrates[10] to the further utilization of the self-assembly properties of bacteriophage to provide a scaffold for the synthesis of high aspect ratio crystalline structures (see Figure 1-3)[9]. However, detailed mechanisms of peptide adsorption on inorganic surfaces have largely escaped experimental characterization. Many surface sensitive techniques lack the flexibility to study a wide variety of inorganic materials and it is often difficult to differentiate between high binding affinities using traditional biochemical techniques. Detailed understanding of these interactions would facilitate rationally designed biomaterials and self-assembly methods.

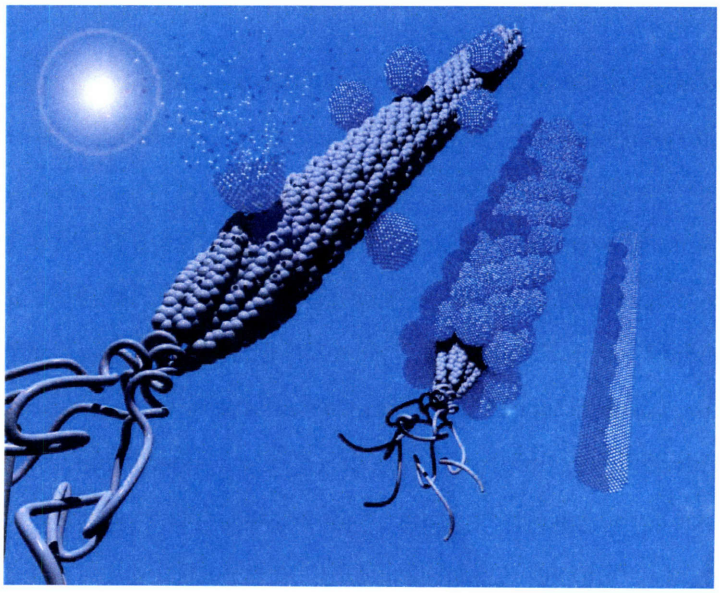

Figure 1-3: A biomineralization scheme which utilizes the self-assembled M13 bactereophage virus capsid as a scaffold for the nucleation and growth of crystalline nanowires from a wide array of materials. Nanoparticals are nucleated on the capsid by material specific peptides, and the wires are then annealed to remove organic material and fuse particles into single crystal wires.

The development of accurate and computationally efficient methods for the simulation of adsorption of biopolymers at inorganic surfaces would provide a more detailed understanding of the adsorption mechanisms which have largely escaped experimental characterization. This knowledge, in turn, would open new routes for the rational design of nanostructures and composite biomaterials. The enormous number of pos-

sible arrangements of amino acids in even a short peptide represents the power of biological systems and largely limits the experimenter to a small number of peptides obtained from evolutionary selection from a combinatorial display library. Rational design using detailed knowledge of adsorption mechanisms would eliminate much of this barrier and allow for the fine-tuning of peptide binding properties.

1.2 Molecular Simulation of Biological Molecules and Surface Adsorption

Biomolecule adsorption simulations range in detail from those focused on macroscopic, colloidal representations[73, 87] which neglect the atomistic detail of electrostatic and van der Waals interactions, to the common atomistic representation of rigid molecular structures in docking simulations[36, 37, 129, 137, 141], and more recently fully flexible biopolymer adsorption simulations[40, 127]. Ignoring atomistic detail has obvious implications for the fidelity of molecular recognition, while rigid molecular representations cannot account for protein adaptive conformational changes during adsorption as well as conformational entropic considerations. While each of these methods have found success in some applications, robust performance for systems of varying size and composition can generally be expected only through the inclusion of key details in the model.

Fully flexible, atomistically detailed simulation of protein-surface adsorption presents several difficulties. *Ab Initio* quantum chemical methods are computationally expensive and applicable only for small molecules coupled with limited surface sizes. Thus for larger peptide and protein systems molecular mechanics type force-fields must be employed. There are a number of alternative force-fields, and these have shown varying success in predicting peptide and protein structure[111]. There also exist inter-atomic force-fields for a wide variety of mineral systems[53, 76]. However, the interaction between organic and inorganic components may not follow a simple combination of force-fields[34]. Therefore, it is necessary that force-fields specifically

intended for the organic-inorganic interactions be developed and validated[136].

The aqueous environment in which protein-surface adsorption occurs also plays a critical role in the process and presents a distinct challenge to molecular simulation. Water molecules screen electrostatic interactions, reorganize to drive hydrophobic interactions, and must be displaced during the adsorption process. Detailed information related to water structure gained by explicit inclusion of solvent molecules comes at the cost of increasing system size, often by at least an order of magnitude. Because of the large computational cost of explicit water, the average effects of water are often included in molecular simulations through implicit solvent models[109]. These models replace the numerous atoms and corresponding degrees of freedom with a continuum dielectric. Solvation energies are then calculated as the difference in free energy for "charging" the molecule in solution and gas phase, plus non-polar van der Waals and cavity formation energies. This process is illustrated in Figure 1-4. Distance-dependent dielectric functions represent a very simple implementation of implicit solvent and have been used with some success in protein adsorption simulations[45, 46, 129]. However, these ad hoc functions likely oversimplify the solvent environment and a more accurate description is desired for general molecular recognition applications. The Poisson-Boltzmann model[62] represents a rigorous treatment of the electrostatic properties of charges in an inhomogeneous dielectric environment. However, solution methods for the Poisson-Boltzmann model are still relatively computationally demanding. The Generalized Born model[55] provides a computationally convenient alternative that can produce accurate solvation energies at a minimal increase in computational cost[70] compared to the distance-dependent dielectric model.

1.3 Scope of Work

In this project, a simulation package based on computationally efficient methods is developed and applied to the simulation of peptide adsorption at inorganic crystalline surfaces. The core of the simulation package is based on traditional Monte

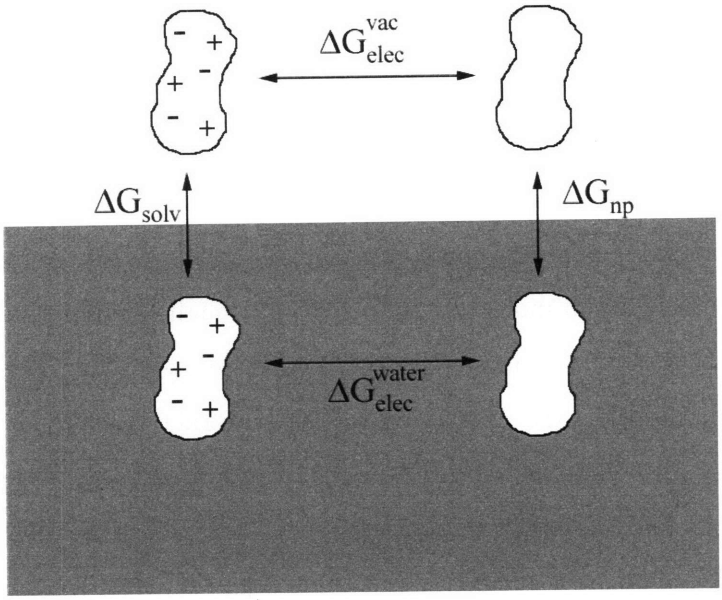

Figure 1-4: Schematic representation of the solvation process involving removal of charge, transfer of molecular cavity into solution, and recharging of solute.

Carlo molecular mechanics, which have proven effective in the simulation of biological molecules[140]. A parallel tempering algorithm[80] was added to the standard Metropolis[138] procedure in order to increase sampling efficiency. Advancements to the initialization procedures for Poisson-Boltzmann implicit solvent calculations[109] involving the definition and handling of the molecular dielectric cavity, as well as several alternatives for efficient solution of the Poisson-Boltzmann equation were explored for incorporation into molecular mechanics simulations. Ultimately, the Poisson-Boltzmann model was not considered computationally efficient enough for use in the current applications. The modified Generalized Born model[147], termed GBr^6, was adapted as an alternative to the more rigorous Poisson-Boltzmann calculations. An inconsistency with the handling of the molecular volume in the model was resolved and shown to improve the model's accuracy. The combined replica-exchange Monte Carlo/Generalized Born implicit solvent simulation was then utilized to study the adsorption mechanism of a set of rationally designed sapphire binding peptides.

Relative binding affinities of this set of peptides was shown to depend on the proximity of charged residues to one another, as well as the conformational ability of each peptide to present functional groups towards the surface.

Chapter 2

Poisson-Boltzmann Implicit Solvent

2.1 Introduction

Implicit solvent models have become a valuable resource in the characterization of biochemical and macromolecular systems[48, 109]. While explicit inclusion of solvent molecules is ostensibly the most accurate method, implicit solvent models have proven to reproduce the effects of solvent environments in many systems while increasing computational efficiency[3, 60, 61, 64, 65, 99, 131, 135]. These models replace the numerous solvent molecules with a continuum dielectric and seek to capture the average effects of the solution environment, greatly reducing the number of degrees of freedom involved in a molecular simulation.

The most robust and rigorous treatment of continuum electrostatics is the Poisson-Boltzmann equation (PBE)[62]. In the simplest case of charges in a uniform dielectric, the electrostatic potential is given by Gauss' law

$$\nabla^2\phi(\mathbf{r}) = \frac{-4\pi\rho(\mathbf{r})}{\epsilon} \tag{2.1}$$

where $\phi(\mathbf{r})$ is the electrostatic potential, $\rho(\mathbf{r})$ is the charge distribution, ϵ is the dielectric, and $\mathbf{r}$ is the position. In an inhomogeneous dielectric, Poisson's equation

must be used to calculate the electrostatic potential

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.2}$$

where the spatial dependence of the permittivity is now included in $\epsilon(\mathbf{r})$. In the presence of a mobile charge distribution, such as in an ionic solution environment, the total charge distribution is decomposed into fixed and mobile contributions. Solute charges are assumed to be at a fixed position, while secondary ion charges are free to react to the eletric field produced by the fixed solute charges.

$$\rho(\mathbf{r}) = \rho^f(\mathbf{r}) + \rho^m(\mathbf{r}) \tag{2.3}$$

At equilibrium, the chemical potential of each mobile ion species must be uniform throughout the solution,

$$\begin{aligned} \mu_i(\mathbf{r}) &= \mu_i^\circ + k_B T \ln a_i(\mathbf{r}) + z_i e \phi(\mathbf{r}) \\ &= \mu_i^o + k_B T \ln a_i^b \end{aligned} \tag{2.4}$$

where k_B is the Boltzmann constant, T the absolute temperature, μ_i is the chemical potential of species i, μ° is the standard chemical potential, a_i is the activity of species i, a^b is the activity in bulk solution where the electrostatic potential is zero, z_i is the charge of species i, and e is the elementary unit of charge.

Assuming that the activity coefficient is unity and independent of electrostatic potential and concentration, or equivalently neglecting the mutual Debye-Hückel interaction, the activity can be equated with the concentration, $a_i = c_i$, and the above relation can be rewritten as a Boltzmann expression

$$k_B T \ln c_i(\mathbf{r}) - k_B T \ln c_i^b = -z_i e \phi(\mathbf{r}) \tag{2.5}$$

Rearrangement yields

$$c_i(\mathbf{r}) = c_i^b \exp(-z_i e \phi(\mathbf{r}) / k_B T) \tag{2.6}$$

for the local concentration of a mobile ion species as a function of bulk concentration and local electrostatic potential. Finally, by assuming a one-to-one stoichiometric relationship between salt counterions and utilizing the relationship $\sinh(x) = \frac{1}{2}(e^{x} - e^{-x})$, the mobile charge distribution can be combined with the fixed solute charge distribution to yield the non-linear Poisson-Boltzmann equation

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) - \overline{\kappa}^2(\mathbf{r}) \sinh[\frac{e\phi(\mathbf{r})}{kT}] = -4\pi\rho(\mathbf{r}) \tag{2.7}$$

where the potential has been replaced by the unitless potential, $e\phi/k_BT$, for simplicity, and $\overline{\kappa}^2 = 8\pi e^2 I/kT$ is a dielectric independent Debye-Hückel parameter, with I the ionic strength of mobile charges. In practical applications, the PBE is often linearized by the assumption that the electrostatic potential is small in the ion-accessible region outside of the molecular volume and distant from fixed solute charges. In this case, the relationship, $\sinh(x) \approx x$, yeilds the linear Poisson-Boltzmann equation (LPBE)

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) - \overline{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.8}$$

Once the electrostatic potential is known, the electrostatic free energy of the system is obtained from the integral[62]

$$G_{elec} = \int (\rho^f \phi - \Delta\Pi - \mathbf{E} \cdot \mathbf{D}/2) dv \tag{2.9}$$

where the first term in the integral, $\rho^f\phi$, is the interaction of each fixed molecular charge with the electric field and represents the largest contribution to the integral. The remaining terms, $(\Delta\Pi)$ and $(\mathbf{E} \cdot \mathbf{D}/2)$, are the excess osmotic pressure and the electrostatic stress, respectively. In the case of the linearized PBE, the excess osmotic pressure reduces to $\rho^m\phi/2$, and the integral expression of Gauss' law, $\int \mathbf{E} \cdot \mathbf{D}/2dv = \int \rho\phi/2dv$, can be substituted into Equation 2.9 to yield

$$G_{elec} = \int \rho^f \phi/2dv \tag{2.10}$$

The difference in electrostatic free energy between the inhomogeneous dielectric and a reference calculation in homogeneous (usually vacuum) dielectric yields the electrostatic component of the free energy of hydration, also called the reaction field energy or solvation energy. This energy is the free energy change involved in transferring a molecule from vacuum into solution resulting from electrostatic contributions and can be recast from Equation 2.10 as a "charging" integral

$$\Delta G_{solv} = \int_0^Q (\phi_s(\mathbf{r}) - \phi_v(\mathbf{r}))dq \tag{2.11}$$

where ϕ_s and ϕ_v are the calculated solvent and vacuum electrostatic potentials, respectively. Assuming a linear charging response, this integral can be replaced by the sum over individual charges, Q_n,

$$\Delta G_{solv} = \frac{1}{2}\sum_n Q_n(\mathbf{r})(\phi_s(\mathbf{r}) - \phi_v(\mathbf{r})) \tag{2.12}$$

It is important to note that the calculated solvation energy is a *free energy*, as the bulk properties its derivation is based upon, e.g. the dielectric constant, include both the enthalpic interaction of charges and the entropic rearrangement of solvent molecules in response to the electric field. In fact, the calculated energy represents the solvation energy averaged over all possible configurations of solvent molecules. This calculation of pre-averaged energy gives implicit solvent models particular advantage in the simulation of large systems, eliminating the necessity for computationally expensive averaging of a multitude of solvent configurations.

Poisson-Boltzmann calculations, like all implicit solvent models, lack detail in the structure of water, particularly important at the solute-solvent boundary. However, they offer the most rigorous treatment of electrostatic effects in solvated systems, as well as the basis upon which other implicit solvent models are constructed and verified against[26, 47, 70].

In order to evaluate hydration free energies via the Poisson-Boltzmann model, the electrostatic potential at the location of each charge must be known. However, there

exist only a limited number of analytical solutions for symmetrically shaped dielectric cavities, such as a sphere, cylinder, or plane geometry. Therefore, the electrostatic potential must be evaluated over the entire system through numerical methods.

2.2 Finite-Difference Poisson-Boltzmann Equation

In order to solve the Poisson-Boltzmann equation over the entire problem space-domain, the molecular system is discretized onto a set of vertices spanning the volume of the molecule and a surrounding solvent volume. The electrostatic potential can then be calculated by a variety of methods, including finite-element[42] and finite-difference (FD) techniques[59, 100], as well as through integral formulations of the PBE[112]. In any discretized, numerical method utilized for the approximation of continuous functions, there will inevitably be error introduced by the discretization, particularly in quickly varying functions. Finite-element methods are attractive due to the ability to utilize non-uniform tetrahedral grids. Grid point density can be increased in areas of space where the electrostatic potential varies quickly, such as the molecular surface and charge centers. In areas of space where the potential varies slowly or is constant, grid density can be much lower, limiting the computational resources necessary for the calculation. This principle has been utilized effectively in both *a priori*[41] and adaptive[103] mesh generation. However, the computational algorithm used in finite-element calculations is significantly inefficient in comparison to other numerical methods, such as finite-difference[4].

Finite-difference methods are unable to effectively utilize a non-uniform grid and instead rely upon a uniform, rectangular grid. For a continuous function discretized on a set of evenly spaced vertices, the derivative of the function at vertex i is given by

$$\left(\frac{\partial f(x)}{\partial x}\right)_i \approx \frac{f(x_{i+1}) - f(x_{i-1})}{2h} \tag{2.13}$$

where h is the spacing of vertices, and x_i is the position of the ith vertex. Application

of this approximation to the PBE yields

$$\frac{\sum \epsilon_i (u_i - u_0)}{h^2} - \frac{\overline{\kappa}^2 \sinh(u_0)}{h} = \frac{q_0}{h^3} \tag{2.14}$$

where u is the potential, the sum is carried out over the six adjacent grid points, i, in the x, y, and z directions, h is the grid spacing, and q_0/h^3 is the charge density contained in the cube surrouding each grid point. Rearrangement of this equation leads directly to an iterative set of equations, which can be used to solve for the electrostatic potential

$$u_0^{k+1} = \frac{\sum \epsilon_i u_i^k + q_0/h}{\sum \epsilon_i + h\overline{\kappa}^2 g(u_0^k)} \tag{2.15}$$

where u^k is the kth estimate of the potential, and the function $g(u_0) = \sinh(u_0)/u_0$ is often approximated through a power series expansion as $g(u_0) = 1 + u_0^2/3! + u_0^4/5! + ...$, or simply $g(u_0) = 1$ for the linearized PBE.

2.2.1 Solution Methods for FDPBE

The finite-difference approximation to the PBE represents a very sparse, banded matrix and straightforward matrix inversion techniques offer a very inefficient solution method. Cubic grid sizes often exceed millions of variables in practical applications, the majority of which having zero direct influence on each other. Therefore, iterative solution techniques are employed for PBE solution. These methods start from an initial estimate of the solution (often $u = 0$), from which the error, or residual, is calculated from Equation 2.15 as $(u^k - u^{k-1})$. The residual is then used to update the current estimate, and a new residual is calculated. This process is repeated until a pre-defined convergence criteria is reached. Many iterative solution techniques exist for systems of linear equations.

Boundary conditions for the finite-difference grid must also be provided in order to reach a solution. The development of the Poisson-Boltzmann equation relies on the assumption that at large distances from the solute charges, the electrostatic potential decays to zero. This represents one possible choice for boundary conditions, provided

that a large enough grid can be constructed such that the zero potential approximation is valid. In practice this is not a viable option as the large grid results in exorbitant computational costs. Grid boundaries are therefore often truncated and values for the electrostatic potential are approximated as the Coulombic potential in zero ionic strength and by the Debye-Hückel potential for ionic solutions.

Jacobi Iterations

Jacobi iterations[108] are perhaps the most straightforward iterative technique and often provide the most stable convergent properties. For a system of linear equations

$$\mathbf{Ax} = \mathbf{b} \tag{2.16}$$

where $\mathbf{A}$ is an $n \times n$ matrix, and $\mathbf{x}$ and $\mathbf{b}$ are vectors of length n, the Jacobi iteration is defined as

$$\mathbf{x}^{k+1} = \mathbf{D}^{-1}[-(\mathbf{L} + \mathbf{U})\mathbf{x}^k + \mathbf{b}] \tag{2.17}$$

where $\mathbf{D}$,$\mathbf{L}$, and $\mathbf{U}$ are the matrix diagonal, lower, and upper decomposition of $\mathbf{A}$, respectively. Jacobi iterations are slowly, but stably convergent in application to the PBE. A major limitation however, is the necessity to calculated fully the next iteration of the estimated solution before updating the current estimate. This requires computer memory allocation of twice the number of variables.

Gauss-Seidel Iterations

A variation on Jacobi iterations is the Gauss-Seidel iteration[2], defined as

$$\mathbf{x}^{k+1} = (\mathbf{D} - \mathbf{L})^{-1}[-\mathbf{U}\mathbf{x}^k + \mathbf{b}] \tag{2.18}$$

Gauss-Seidel iterations utilize the updated estimate of each variable as they are produced. This results in faster convergence, as corrections to the current estimate and their effect on adjacent variables are immediately incorporated rather than waiting for the entire residual to be calculated. Computationally, this also eliminates the neces-

sity to allocate memory to a temporary array at each iteration and subsequently copy information back to the current estimation. The PBE offers a particularly efficient application of Gauss-Seidel iterations, implemented through so-called "red-black" iterations. Since the value of the electrostatic potential depends only on the six adjacent grid points directly, the current estimate of the potential can be effectively updated by two separate "sweeps" through the finite-difference grid on alternating points.

Successive Over-Relaxation

Successive over-relaxation (SOR)[123] increases the convergence speed of Gauss-Seidel iterations by over counting the residual error that is added back to the current estimation of the solution

$$\mathbf{x}^{k+1} = (\mathbf{D} - \omega\mathbf{L})^{-1}[(-\omega\mathbf{U} + (1-\omega)\mathbf{D})\mathbf{x}^k + \omega\mathbf{b}] \tag{2.19}$$

where ω is the spectral radius of convergence and must be adjusted for optimal convergence. SOR has been used effectively in PBE calculations, most notably in the popular commercially available implementation Delphi[66].

Direct Inversion of the Iterative Subspace

Direct inversion of the iterative subspace (DIIS)[116] is an iterative technique widely applied in quantum mechanical calculations[22, 23]. In this technique, iterations proceed by the Jacobi method. However, instead of replacing the current estimate and discarding the old, each estimate is saved and added to the so called "subspace". The residual error associated with each estimate is also saved. The DIIS method assumes that a good estimate to the solution, $\mathbf{x}'$, can be obtained from a linear combination of the jacobi estimates

$$\mathbf{x}' = \sum_{i}^{m} c_i \mathbf{x}^i \tag{2.20}$$

where m is the number of vectors in the subspace. The coefficients c_i are determined such that the linear combination of residuals, $\mathbf{r^i}$, approximates the zero vector,

$$\mathbf{r}' = \sum_i^m c_i \mathbf{r}^i \tag{2.21}$$

also subject to the requirement that the sum of the coefficients satisfies

$$\sum_i^m c_i = 1 \tag{2.22}$$

Thus we seek to minimize the norm of the residual vector

$$< \mathbf{r}'|\mathbf{r}' >= \sum_{ij}^m c_i^* c_j < \mathbf{r}^i|\mathbf{r}^j > \tag{2.23}$$

Here, we utilize the Lagrangian multiplier, λ, to define

$$\mathbf{\Gamma} = \mathbf{c}^\dagger \mathbf{B}\mathbf{c} - \lambda(1 - \sum_i^m c_i) \tag{2.24}$$

where $B_{ij} =< \mathbf{r}^i|\mathbf{r}^j >$. We can minimize Γ with respect to a coefficient c_k to obtain

$$\frac{\partial \Gamma}{\partial c_k} = \sum_i^m c_i B_{ki} - \lambda = 0 \tag{2.25}$$

Finally, this set of equations is solved for the coefficients c_i by inversion of the matrix $\mathbf{B}$. Matrix inversion typically scales poorly with matrix size. Therefore, the number of previous estimates that should be kept in memory and used in DIIS calculations must be optimized. Too few previous estimates does not provide an effective basis set for a linear combination, while storing too many previous estimates results in diminishing return on computational investment. Figure 2-1 demonstrates the optimization of DIIS application to PBE calculations. Electrostatic energy calculations were performed for blocked alanine (acetyl-alanine-methyl amide), with convergence assumed when the energy changed by less than 1×10^{-6} kcal/mol, with varying DIIS subspace sizes. Calculations were carried out on a 1.6 GHz AMD Athlon processor.

Jacobi iterations with no DIIS converged in 821 iterations and 45.063 seconds. Although the number of iterations required to reach convergence continues to decrease up to a subspace size of 20, the computational time reaches a minimum at a subspace size of 5. At this size, convergence is reached in 52 iterations and 4.566 seconds.

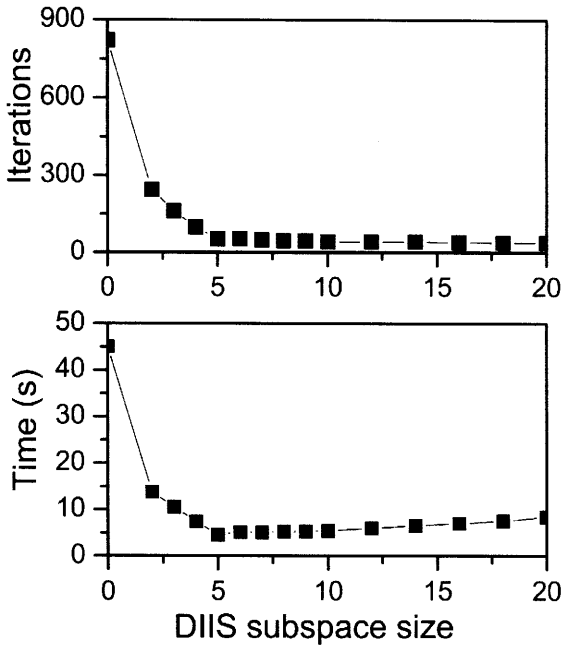

Figure 2-1: Optimization of DIIS subspace size. Matrix inversion scales poorly with size and eventually results in longer convergence time in spite of decreased iterations.

In general, larger DIIS subspace sizes produce a more rapidly converging iteration procedure. However, the decrease in iterations is offset by increased compuational cost involved in each iteration. This balance can be leveraged by a modified DIIS procedure in which the direct inversion of the retained subspace is carried out only periodically, rather than after each underlying Jacobi iteration. The process begins by following general Jacobi iterations until a given number of subspace vectors have been retained. The subspace inversion then supplies the next estimate of the solution vector, often with a significant decrease in error. The retained subspace is then discarded and the process repeated. Figure 2-2 demonstrates the effectiveness of this method. Periodic inversion of the subspace offers similar performace as a function

of iterations as the traditional DIIS procedure, but without the necessity of costly inversion at each iteration. This is not unexpected since the traditional DIIS method involves linear combinations of linear combinations and is therefore in some senses, redundant.

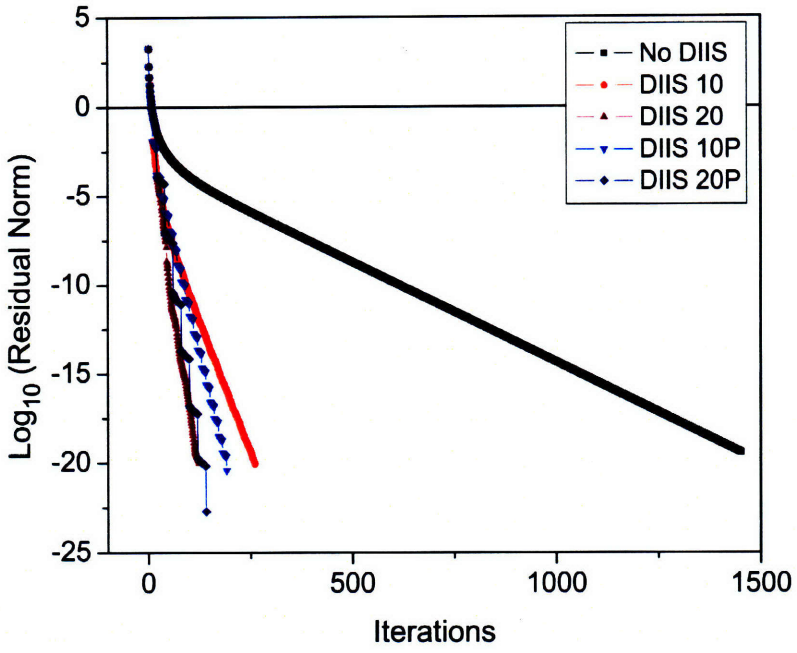

Figure 2-2: Larger DIIS subspace size generally produces convergence in fewer iterations, but at the cost of greater computational time per iteration. This balance can be leveraged by maintaining a large DIIS subspace without performing the direct inversion at each iteration. The subspace is built over a set of Jacobi iterations until a given subspace size is reached. The inversion routine produces the next approximate solution as the linear combination of the subspace vectors. The subspace is then discarded and rebuilt following more Jacobi iterations. Performace of periodic inversions (DIIS 10P, DIIS 20P) is similar to inversions at each step (DIIS 10, DIIS 20) for subspace sizes of 10 and 20 vectors.

Multigrid Solution of the Poisson-Boltzmann Equation

The application of Poisson-Boltzmann calculations to molecular simulations is an enticing prospect. Explicit inclusion of solvent in a biomolecule simulation can increase

the number of atoms and degrees of freedom which must be sampled by an order of magnitude[130]. Not only does implicit representation of solvent eliminate costly averaging calculations, but these models also have the effect of increasing sampling efficiency in the biomolecule itself by removing viscocity related impedimants to the biomolecule's motion[105]. However, PBE calculations still represent a significant computational investment and significant efforts have been aimed at accelerating solvation energy calculations. Perhaps the most successful of these efforts in application to finite-difference techniques is the development of multi-grid methods[124].

Given some initial estimate to the Poisson-Boltzmann equation, $\mathbf{u}$,

$$\mathbf{Lu} + \sum_i K_i e^{-q_i \mathbf{u}/kT} + \mathbf{f} \approx 0 \tag{2.26}$$

where we have written the finite-difference operator defined in Equation 2.13 as $\mathbf{L}$, and the source term, $-4\pi\rho$, as $\mathbf{f}$ for clarity. There exists some correction vector, $\mathbf{v}$, to $\mathbf{u}$ that solves the equation, i.e.

$$\begin{aligned} 0 &= \mathbf{L}(\mathbf{u}+\mathbf{v}) + \sum_i K_i e^{-q_i(\mathbf{u}+\mathbf{v})/kT} + \mathbf{f} \\ &= \mathbf{Lv} + \sum_i K_i e^{-q_i \mathbf{u}/kT} e^{-q_i \mathbf{v}/kT} + (\mathbf{Lu}+\mathbf{f}) \end{aligned} \tag{2.27}$$

which has the same form as the original PBE, namely

$$\begin{aligned} \mathbf{Lv} + \sum_i J_i e^{-q_i \mathbf{v}/kT} + \mathbf{r} &= 0 \\ J_i = K_i e^{-q_i \mathbf{u}/kT}, \quad \mathbf{r} &= \mathbf{Lu}+\mathbf{f} \end{aligned} \tag{2.28}$$

The multigrid method seeks to solve the correction term on a courser grid, interpolate the solution back to the fine grid, and add the correction term back to the current estimate, $\mathbf{u}$. This method is not simply limited to two grids, one coarse and one fine. The correction term can be treated analagously to the current estimate, and we can thus calculate a correction to the correction at an even courser grid. In practice, this process is repeated until the number of variables is small enough for direct inversion of

the finite-difference operator, **L**. Figure 2-3 demonstrates the improved convergence of the multigrid method in comparison to Gauss-Seidel iterations for calculation of blocked alanine electrostatic energy at 1 Å grid spacing.

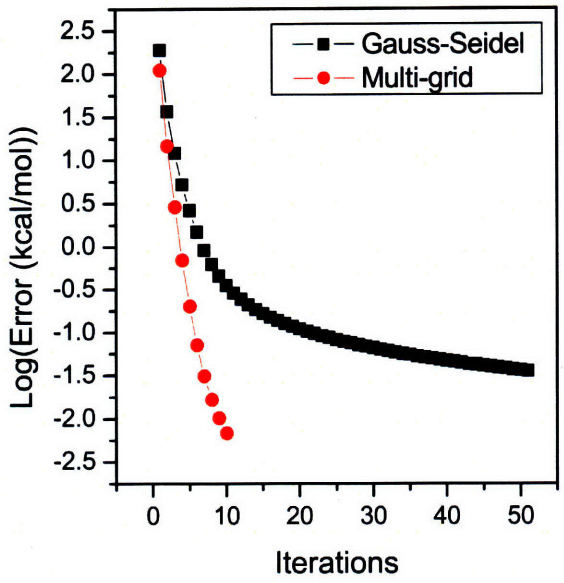

Figure 2-3: Convergence of multigrid solution of the Poisson-Boltzmann equation in comparison to Gauss-Seidel iterations

We can understand the effectiveness of the multigrid method by imagining the decomposition of the solution of the PBE into a linear combination of sine and cosine functions. At a given grid spacing, high frequency errors propogate quickly through the grid, while the low frequency error is slow to converge. Frequency in the error, however, is related to grid spacing, ie. low frequency error at a fine grid spacing is high frequency error at a coarse grid spacing. By "restricting" the correction calculation onto a courser grid where the low frequency errors propogate quickly, this slow convergence is eliminated from iterations at the fine grid level. However, the high frequency errors at the fine grid spacing are unresolvable at coarse grid scales, and information is inevitably lost through multigrid correction cycles. Therefore, multigrid cycles in practical application do not proceed straight from one grid scale to the next, but are separated by smoothing iterations at each grid scale, by one of

the methods listed above, most often Gauss-Seidel.

We attempted to combine the improved convergence properties of the DIIS method with those of the multigrid technique. It was hoped that by applying DIIS to the smoothing iterations that separate restriction and prolongation to coarser and finer grids better current estimates of the solution would be transfered to the next grid. Figure 2-4 illustrates the application of DIIS to multigrid calculations of Born ion electrostatic energy at 0.5 Å grid spacing. Initially, DIIS offers some improvement on the rate of convergence. However, as the error is reduced, the DIIS subspace for the coarse grids approches a linearly dependent set. Inversion of the subspace challenges the limits of machine precision in standard double-precision computations and results in the introduction of large error and an oscilating convergence pattern. A switching from DIIS to Gauss-Seidel iterations at a predefined error criteria was considered. However, the marginal improvement to convergence was not considered sufficient to pursue such a solution, and further development focused only on Gauss-Seidel iterations.

Poisson-Boltzmann Homogeneous Dielectric Reference Calculation

The ultimate goal of developing a Poisson-Boltzmann equation solver is incorporation of the model into molecular simulation software. The OPLS-AA molecular mechanics force-field[139] contains scaling factors utilized in the calculation of non-bonded interactions. Therefore, in order to remain consistent with the existing force-field, the solvation energy must be isolated from the total electrostatic free energy. This total free energy calculated by solution of the PBE contains the coulombic interaction of all charges, molecular solvation energy, and the self energy. The self energy is the interaction of each charge with its own electric field. This term is analytically infinite, but can be extracted computationally as finite. Thus a reference calculation must also be carried out in uniform vacuum dielectric, as discussed in Section 2.1.

It has been demonstrated that this reference calculation can be replaced by a direct coulombic interaction by accounting for effective grid distances[96]. More recently, it was shown that solvation energies could be computed directly from a single

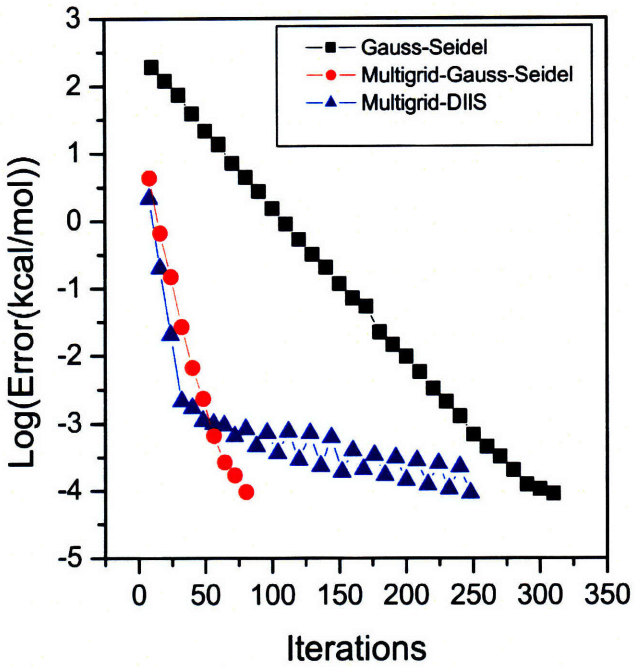

Figure 2-4: Convergence properties of Gauss-Seidel, multigrid Gauss-Seidel, and multigrid DIIS iterations. DIIS causes instability in multigrid calculations due to nearly linear dependent subsequent estimates.

inhomogeneous dielectric calculation[67]. Solute charges induce polarization charges at the dielectric boundary of the molecular surface. Reformulation of the PBE to evaluate this charge and subsequent calculation of the coulombic interaction between solute charges and polarization charges results in the solvation energy. This method, however, was developed in the context of a discontinuous permittivity function at the molecular surface and has not been shown to extend generally to smooth permittivity functions discussed below in Section 2.3.2. Effective grid distances are discussed further in Section 2.5.

2.3 Error in Solvation Energy Calculations

Finite-difference Poisson-Boltzmann calculations are subject to several sources of error[64]. Any iterative technique is of course reliant upon the convergence criteria

employed. The convergence of iterative computational methods is often monitored by the norm of the residual error vector or the maximum element of the residual vector. In practical applications of PBE calculations, however, this source of error can be better controlled by monitoring the convergence of the solvation energy. The energy calculation is computationally simple, and thus offers a direct measure of the error introduced into a molecular simulation. Limiting the convergence criteria to 10^{-2} kcal/mol certainly maintains solvation energy error within the uncertainty of typical molecular mechanics force-fields.

A second source of error in implicit solvent models is based upon the loss of atomic detail in the structure of water. When a molecular system is discretized for finite-difference calculations, the molecular volume must be mapped onto the cartesian grid. The region of space spanned by the molecular volume is assigned a molecular dielectric (typically 1-4), while the solvent volume is assigned a solvent dielctric ($\approx$ 80 for water). Large protein molecules often contain small regions of space embedded in the molecular interior which fall outside of the molecular surface, but are not large enough to accomadate a water molecule. These areas, called microdielectrics, are thus assigned the highly polarizable solvent dielectric even though water is not present. In fact, even areas large enough to accomodate a single or small number of water molecules should not be considered so highly polarizable. However, it is not clear at what size the transition to bulk dielectric properties should occur. This effect has led to the development of so-called "re-entrant" molecular surfaces[67, 119], constructed by rolling a probe sphere over the surface of the solvated molecule. This probe sphere is usually defined with a radius of 1.4 Å, half the average oxygen-oxygen separation in liquid water.

The difficulty in describing the molecular surface and discretization of the molecular volume leads to a further complication in solvation energy calculation, particularly in the application of PBE to molecular simulations. Imagine a molecular system with an embedded microdielectric cavity just large enough to enclose the probe sphere and thus assigned the solvent polarizability. A small conformational change which results in constriction of the microdielctric and expulsion of the probe sphere causes

a drastic change in the dielectric of the region, and in turn, large changes in the electrostatic potential and solvation energies. This effect turns out to have consequences in nearly all translational, rotational, and conformational changes to the molecular structure as charges and molecular surfaces move in relation to the discretized grid. Stabilization of solvation energy with respect to molecular grid position turns out to be one of the most challenging aspects of PBE calculations[18, 101, 110, 122, 132]. This complication has obvious implications for the application of PBE calculations in physics-based molecular simulations. Significant effort has been aimed at model definitions for charge and molecular volume discretization that decrease the grid positional and orientational dependence of solvation energy calculations.

2.3.1 Discretization of Charge

The simplest representation of atomic charge discretization would be to assign the atomic charge to the grid point closest to the atom center. This model, however, is obviously not continuously varying with respect to atom position and linear interpolation methods have traditionally been employed to spread the charge over the eight surrounding grid points. Improved stability has been achieved by uniform charge distribution[18] and antialiasing[132] methods which distribute the charge over all grid points contained within the molecular volume. The primary strength of these methods is related to the self energy of each charge. If charge is distributed over only eight points, the Coulombic interaction of these charges with each other as grid scale is reduced becomes very large and challanges machine precision. If instead, the charge is spread over the molecular volume, more grid points are incorporated as the grid scale is reduced; the charge density and self energy remains constant. However, distribution of charge over the whole of the molecular volume, especially at large grid scales, results in the assignment of charge to grid points at or outside of the molecular surface.

In the current efforts, a charge assignment method based on inverse quadratic interpolation was used[110]. Charges are spread over the $3\times3\times3$ cube of grid points,

partitioned by

$$\begin{aligned} q_{i-1} &= Q[\frac{1}{8} - \frac{1}{2}(x-i) + \frac{1}{2}(x-i)^2] \\ q_i &= Q[\frac{3}{4} - (x-i)^2] \\ q_{i+1} &= Q[\frac{1}{8} + \frac{1}{2}(x-i) + \frac{1}{2}(x-i)^2] \end{aligned} \tag{2.29}$$

for $\frac{1}{2} \leq x \leq i + \frac{1}{2}$ for cell i, and Q is the atom charge. This method is continuously varying over all grid translations while maintaining a localized charge density at the atomic center as well as conserving dipole moment. The charge at a given node in three dimensions is the product of the fractional partition for each dimension, i.e.

$$q_{i(\pm1)j(\pm1)k(\pm1)} = Q \times q_{i(\pm1)} \times q_{j(\pm1)} \times q_{k(\pm1)} \tag{2.30}$$

2.3.2 Smooth Permittivity Functions

In addition to the discretization of charge, electrostatic potential and solvation energy are highly dependent on the discretized map of the molecular volume. The traditional representation of the molecular volume produces a discontinuous step in the dielectric at the molecular boundary. As grid vertices pass through the boundary, the abrupt change in dielectric causes large fluctuations in the calculated solvation energy. Davis and McCammon showed that the errors associated with the precipitous change in dielectric could be alleviated by harmonically averaging the permittivity over the grid line connecting two vertices, rather than solely taking the value at the midpoint[101]. This conclusion was inspired by matching finite-difference theory to the analytical solution for the electrostatic potential in a parallel plate capacitor. The result can also be obtained by the subdivision of a single grid line followed by application of one-dimensional finite-difference approximations and elimination of variables. Ignoring charge and non-linear terms and examining one dimension for clarity, the finite

difference approximation to the PBE yields

$$u_0 = \frac{\epsilon_{-\frac{1}{2}} u_{-1} + \epsilon_{+\frac{1}{2}} u_{+1}}{\epsilon_{-\frac{1}{2}} + \epsilon_{+\frac{1}{2}}} \tag{2.31}$$

where u_{-1} and u_{+1} are the potential at neighboring grid points, and $\epsilon_{-\frac{1}{2}}$ and $\epsilon_{+\frac{1}{2}}$ are the dielectric at the midpoint of the connecting grid lines, illustrated in Figure 2-5. Extending the system to include two "virual" grid points at the midpoints of each

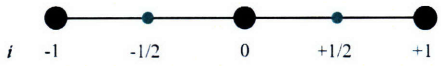

Figure 2-5: 1D FD grid line, with grid points $i = -1, 0, 1$

grid line for five grid points total, the three internal grid points lead to the set of equations

$$\begin{aligned} u_{-1} &= \frac{\epsilon_{-\frac{3}{2}} u_{-2} + \epsilon_{-\frac{1}{2}} u_0}{\epsilon_{-\frac{3}{2}} + \epsilon_{-\frac{1}{2}}} \\ u_0 &= \frac{\epsilon_{-\frac{1}{2}} u_{-1} + \epsilon_{+\frac{1}{2}} u_{+1}}{\epsilon_{-\frac{1}{2}} + \epsilon_{+\frac{1}{2}}} \\ u_{+1} &= \frac{\epsilon_{+\frac{1}{2}} u_0 + \epsilon_{+\frac{3}{2}} u_{+2}}{\epsilon_{+\frac{1}{2}} + \epsilon_{+\frac{3}{2}}} \end{aligned} \tag{2.32}$$

as illustrated in Figure 2-6. Combining equations and elimination of the variables

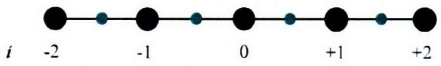

Figure 2-6: 1D FD grid line, with grid points $i = -2, -1, 0, 1, 2$

u_{-1} and u_{+1} yields

$$u_0 = \frac{(\epsilon'_-) u_{-2} + (\epsilon'_+) u_{+2}}{(\epsilon'_- + \epsilon'_+)} \tag{2.33}$$

where $\epsilon'_{+/-}$ is the effective dielectric over the grid line

$$\epsilon'_{-} = \left(\frac{2}{\frac{1}{\epsilon_{-\frac{3}{2}}} + \frac{1}{\epsilon_{-\frac{1}{2}}}} \right), \quad \epsilon'_{+} = \left(\frac{2}{\frac{1}{\epsilon_{+\frac{1}{2}}} + \frac{1}{\epsilon_{+\frac{3}{2}}}} \right) \tag{2.34}$$

Extending to n grid line subdivisions, the effective dielectric is the harmonic average

$$\epsilon' = \left(\frac{n}{\sum_n 1/\epsilon_n} \right) \tag{2.35}$$

As illustrated in the example of a parallel plate capacitor, this method becomes exact in the limit of zero curvature in the dielectric boundary or equivalently, zero grid spacing.

This averaging technique can be interpreted as an increase in the precision with which the location of the dielectric boundary is defined, as more detailed information about the permittivity function has been included. The traditional, binary representation provides no further information than between which two grid points the molecular surface lies. By averaging along the grid line, we gain more precise information of where the boundary falls between two grid points. This also produces the effect of a smoothly varying dielectric at the molecular surface, resulting in improved computational stability and convergence, although the fundamental model of the molecular surface is unchanged.

Averaging has also been employed in a slightly altered manner to produce smooth permittivity effects by weighted averaging of the dielectric over surrounding grid points[132]. In contrast to the above described averaging method, this averaging destroys information about the molecular surface. Averaging over grid lines includes more information than is already present on the finite-difference grid. However, averaging local grid values blurs the molecular surface and although a smoothly varying dielectric is produced, it is not clear that this should result in improved accuracy.

Alternatively, the model itself can be adapted to include a smoothly varying definition of the permittivity[110, 122]. Perhaps the most attractive and elegant of these models is the Gaussian based atomic volume function. In this model the molecular

volume is described by a set of overlapping Gaussian functions. The Gaussian-based density of an atom of nominal radius σ_A is given by

$$\rho_A^g(\mathbf{r}) = p_A \exp(-\kappa r_A^2/\sigma_A^2) \tag{2.36}$$

where p_A is a hieght factor, r_A is the radial distance $(\mathbf{r} - \mathbf{r_A})$ from atom A, and κ is a dimensionless exponent. The volume integral is

$$V_A = \int dr^3 \rho_A^g = p_A \left(\frac{\pi}{\kappa}\right)^{3/2} \sigma_A^3 \tag{2.37}$$

where dr^3 is the volume element and the integral is taken over all space. Consistency with a physically realistic description of atomic volume is maintained by requiring that the Gaussian volume equal that of a solid sphere

$$p_A \left(\frac{\pi}{\kappa}\right)^{3/2} \sigma_A^3 = \frac{4}{3}\pi\sigma_A^3 \tag{2.38}$$

This relationship reduces the parameterization of the Gaussian function to a single variable. Adapting the total molecular volume encompassed by a set of overlapping spheres[128] to the set of overlapping Gaussians, the molecular volume is defined by the Poincaré sum, as

$$\begin{aligned} \rho_{mol}(\mathbf{r}) &= 1 - \prod_A (1 - \rho_A^g) \\ &= \sum_A \rho_A^g - \sum_{A>B} \rho_A^g \rho_B^g + \sum_{A>B>C} \rho_A^g \rho_B^g \rho_C^g - \ldots \end{aligned} \tag{2.39}$$

This equation represents the sum of each atom contribution to the molecular volume, $\sum_A \rho_A^g$, plus correction terms to account for overcounting of overlapping volumes. This formulation has been previously shown to produce excellent results for molecular volumes and surface areas[115]. Linear mapping of this molecular volume function to dielectric values produces a dielectric that increases far too rapidly towards solvent

values with distance from atomic centers. Instead, the permittivity is described as

$$\epsilon(\mathbf{r}) = \epsilon_{solute} + (\epsilon_{solvent} - \epsilon_{solute})e^{-A\rho_{sum}(\mathbf{r})} \tag{2.40}$$

$$\rho_{sum}(\mathbf{r}) = \sum_i p_A e^{-\kappa r_i^2/\sigma_i^2} \tag{2.41}$$

where the sum is carried out over the atoms of the molecule and does not include the overlap terms of Equation 2.39. The exponential in Equation 2.40 serves to smooth out the dielectric in the molecular interior and provide a quicker transition to solvent at the molecular surface. The dimensionless parameter, A, is determined for each combination of dielectrics, $(\epsilon_{solute}, \epsilon_{solvent})$, by fitting PB results to traditional definitions of the molecular surface.

The Gaussian model imparts several benefits. Besides offering an arguably more physically realistic basis, the atomic Gaussian molecular volume provides simpler construction than the molecular surface. Differentiability with respect to atomic position allows for the direct calculation of solvent forces and subsequent incorporation into molecular dynamics simulations. Also, similarly to the averaging technique, this smoothly varying permittivity function provides computational stability and improved convergence. Finally, there has been considerable debate of the proper value for the polarizability of protein molecules. While internal molecular dielectrics can be argued to be optimally set to a value of 2[63], accurate results in protein systems often require a molecular dielectric between 4-20, which can result from the microdielectric effects discussed above. Interestingly, the Gaussian based model achieves accurate results for these systems with internal dielectrics of 1-2. The tails of the Gaussian functions overlap in the microdielectric regions to yield dielectric values intermediate to the internal and external values.

2.3.3 Combining Smooth Permittivity Functions and Local Dielectric Smoothing

Although smooth permittivity functions and local averaging both have the broader result of smoothly varying dielectric, it is important to mark the distinction between the atom-centered Gaussian function as a fundamental model of the solute-solvent boundary and harmonic averaging as a technique to obtain more detailed information of the discretized molecular surface at a given grid scale. Because of this distinction, it is potentially applicable to combine these methods. To this end, the Gaussian volume function is first evaluated at adjacent grid points i and $i+1$ by Equation 2.40 to give ρ_i and ρ_{i+1}, respectively. With a continuous definition of the permittivity, the harmonic average in Equation 2.35 is replaced by the analagous integral equation, written as[82]

$$\begin{aligned}\epsilon(\mathbf{r}) &= \frac{(\rho_{i+1}-\rho_i)}{\int_{\rho_i}^{\rho_{i+1}} d\rho[\epsilon_{solute}+(\epsilon_{solvent}-\epsilon_{solute})exp(-A\rho)]^{-1}} \\ &= \frac{\epsilon_{solute}(\rho_{i+1}-\rho_i)}{(\rho_{i+1}-\rho_i)-A^{-1}\ln(\epsilon_{i+1}/\epsilon_i)}\end{aligned} \tag{2.42}$$

where ϵ_i is the dielectric evaluated by Equation 2.40 at the grid vertex. This method retains the benefits of the physically appealing smooth permittivity function, while also capitalizing on the increased positional stability of the harmonic averaging technique.

This method was tested through solvation energy calculations for a variety of systems ranging from a single Born ion to large proteins. In all calculations, solute and solvent dielectric constants were set to 1 and 80, respectively. Zero ionic strength was assumed and convergence set as $1\text{x}10^{-6}$ kcal/mol. Finite-difference grid boundary conditions were set according to the Coulombic potential.

The Born ion provides a simple, clear testing ground as the solvation energy of a single ion with charge, q, and radius, σ, is available analytically for comparison as

$$\Delta G_{Born} = \frac{q^2}{8\pi\epsilon_0\sigma}\left(\frac{1}{\epsilon_{solvent}}-\frac{1}{\epsilon_{solute}}\right) \tag{2.43}$$

In order to illustrate practical application of this method, solvation energy calculations were also performed on a set of small molecules and proteins from the Protein Data Bank (http://www.rcsb.org/pdb). Finally, the solvation contribution to the binding energy of a thrombin-NAPAP complex is used to demonstrate application to computational binding simulations. In order to assess positional error, the solvation energy of each ion or small molecule was calculated at 100 random positions relative to the FD grid. Protein solvation energies were calculated at 20 random positions relative to the FD grid. The standard deviation and range, defined as the difference between the maximum and minimum values, of calculated solvation energies were used to evaluate the positional stability of four dielectric models: the traditional discontinuous molecular surface (MS), the harmonically averaged molecular surface (MS-HA), the Gaussian based permittivity (GAUSS), and the harmonically averaged Gaussian model (GAUSS-HA).

The solvation energy of a single ion with radius $\sigma = 2$Å and unit charge, $q = 1$ was evaluated at 100 grid positions for grid spacings in the range of 0.1-1.0 Å. Figure 2-7a illustrates the average error in the Born ion solvation energy calculation over the range of grid scales. Solvation energies are known to be highly sensitive to the description of the molecular surface and therefore force-fields are often re-parameterized for optimum accuracy[70, 97, 98]. This is not a surprising result considering that solvation energy in the continuum representation is equated with the build-up of induced polarization charge at the dielectric interface. Smooth permittivity models represent the region of induced polarization charge as a three-dimensional volume, whereas in the discrete molecular surface model polarization charges exist only on a two-dimensional surface. This difference in the definition of the dielectric boundary and position of induced charges will considerably alter the solvation energy. It is apparent that this is the case with the Gaussian based models, as both converge at fine grid scale to a solvation energy of $\Delta G_{solv} = -88.47$ kcal/mol rather than the analytical value of $\Delta G_{Born} = -81.95$ kcal/mol. Optimal re-parameterization has been demonstrated for several popular force-fields by Swanson *et al.*[97, 98] by both the rescaling of current force-field parameters as well as development of new parameter sets. In the absence

of this re-parameterization focus should instead be placed on the error relative to the fine grid value for the Gaussian based models. The increased accuracy achieved through harmonic averaging is very apparent in the MS models, and can also be seen in the Gaussian based models at larger grid scales. The stabilizing benefits of harmonic averaging are illustrated in Figures 2-7b and 2-7c by examining the standard deviation and range of solvation energies, respectively, for the 100 repeated calculations at each grid scale. For a desired stability characterized by a standard deviation of 10^{-2} kcal/mol, the MS-HA model requires a grid scale of 0.2 Å, the GAUSS model requires a grid scale of slightly less than 0.4 Å, while the GAUSS-HA model achieves this level of stability at a grid scale of 0.6 Å, with an absolute error of less than 0.3 kcal/mol. This represents a powerful, yet straightforward means for accelerating Poisson-Boltzmann calculations as cubic grid based methods typically scale as the cube of the grid size.

The second test of combining Gaussian based permittivity functions and harmonic averaging was the calculation of molecular solvation energies. A set of eight small molecules and four proteins were used to test the accuracy and stability of the three reference dielectric models and the new combined method. Atom charge and radius parameters were taken from the Optimized Parameters for Liquid Simulations (OPLS)[140], with the exception of charged hydrogens whose radius was set as 0.8 Å rather than 0.0, as such a radius is inappropriate for a molecular volume based solvation energy calculation. Interior and exterior dielectric constants were set to 1 and 80, respectively, as consistent with the OPLS force-field. Table 2.1 shows the solvation energy, standard deviation, and range of energies for each molecule calculated at a relatively large grid spacing of 1 Å. Solvation energies are similar between all methods, although it is again apparent, particularly for the protein energies, that the Gaussian models should be re-parameterized for optimal agreement with the molecular surface models.

In these calculations, a large grid scale was chosen to emphasize the stability imparted by the combination of a smooth dielectric model and the averaging technique. Accurate work normally requires a grid scale of 0.5 Å or less. Even at a large grid

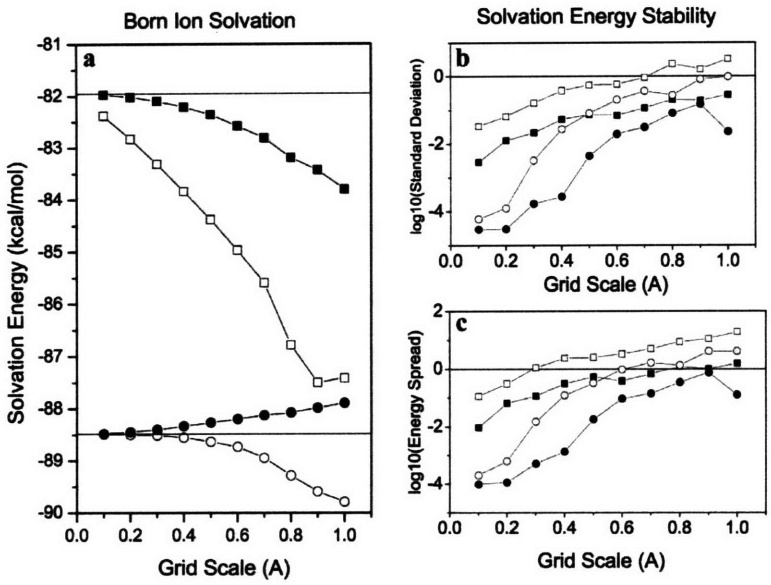

Figure 2-7: Calculation of Born ion solvation energies as a function of grid spacing for each of the four dielectric models: (□) Traditional molecular surface, (■) Harmonically averaged molecular surface, (○) Gaussian atomic volume function, and (●) Harmonically averaged Gaussian atomic volume function. (a) Ion solvation energy. Each point represents the average solvation energy calculated at 100 random positions relative to the finite-difference grid. The horizontal line at -81.95 kcal/mol represents the theoretical solvation energy of a 2 Å Born ion in water (ϵ_{solv}=80), while the horizontal line at -88.47 kcal/mol represents the fine grid solvation energy for the Gaussian model and serves as a guide to the eye. (b) Standard deviation of the calculated solvation energies. (c) Energy spread is the difference between the maximum and minimum calculated solvation energies. Note that (b) and (c) are on a logarithmic scale.

scale, the GAUSS-HA model produces standard deviations in small molecule solvation energies of less than 0.35 kcal/mol. Also, the range of calculated energies for each of the small molecules is comparable to the thermal energy, k_BT; an important comparison when considering Monte Carlo simulations, for example. Protein calculations are similarly stabilized, exemplified by the relative standard deviation of a ferrodoxin protein (PDB 2FDN) solvation energy, which is limited to 0.03%.

Poisson-Boltzmann calculations have long been used to determine solvation forces[102, 104] and there have been significant recent efforts aimed at incorporation of PB mod-

els into molecular dynamics simulation programs[105, 122, 130, 131]. Although the majority of work contained here focuses on Monte Carlo simulation methods in which only the solvation energy must be calculated, these methods are equally valid and applicable to molecular mechanics force based simulations. A representative subset of atoms from the list of small molecules in Table 2.1 was used to examine the accuracy and stability of solvation forces evaluated through the GAUSS-HA method. As demonstrated in Figure 2-8, harmonic averaging improves the accuracy at large grid scales and dramatically stabilizes the Gaussian model even at very large grid spacings.

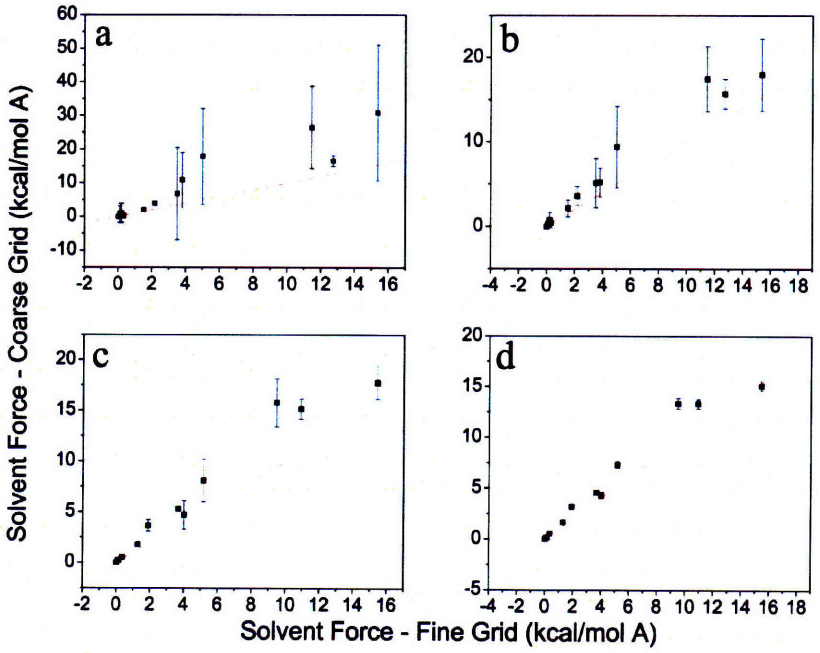

Figure 2-8: Solvation forces for a subset of atoms from the test structures listed in Table 2.1 comparing coarse grid scale calculations (1.0 Å) to forces calculated at a fine grid scale (0.1 Å). Data points represent the average force for each atom calculated at 25 random molecular positions relative to the finite-difference grid, with standard deviations represented by error bars, and the line ($y = x$) as a guide for the eye. (a) MS model, (b) MS-HA model, (c) GAUSS model, (d) GAUSS-HA model.

Finally, the GAUSS-HA model was demonstrated in comparison to the other three models in calculating the solvation contribution to the binding energy of the bovine

thrombin-NAPAP (N^{α}-(2-naphthyl-sulphonyl-glycyl)-D-p-amidino-phynylalanyl-piperidine) complex[16]. The input structures were prepared fromt he Protein Data Bank file (PDB 1ETS) by removing waters and ensuring neutrality of the thrombin protein. This coagulation protein-inhibitor complex consists of 2652 atoms, and was chosen as representative of general protein-ligand binding experiments. Atom and radius parameters were again taken from the OPLS force-field, with interior and exterior dielectric constants of 1 and 80. Solvation energy calculations were carried out at 20 random positions relative to the finite difference grid. The contribution to binding energy was calculated as the difference in mean solvation energies between the complex and its component parts, and standard deviations were combined to yield the standard deviation of the binding energy. Figure 2-9 shows the difference in solvation energies over a range of grid spacings from 0.3 to 1.4 Å for each dielectric model. Once again, it is evident that the absolute energy calculated with the Gaussian model differs from the molecular surface model when using the same atomic parameters, and comparison should be made to the energies calculated at fine grid spacing. The GAUSS model alone offers similar accuracy and stability at large grid scales to the MS-HA model. Application of the averaging technique to the Gaussian based permittivity functions further stabilizes the calculation, reducing the standard deviation of the computed binding energies by a factor of 3-5 and absolute errors by a factor of 2-3 over the range of grid scales.

Harmonic averaging in the MS-HA model adds a significant computational cost to the grid initialization routine. For the thrombin-NAPAP complex on a 1.0 Å grid, the initialization CPU time for the MS model was 2.1 seconds on a single 2.4-GHz Intel Xeon porcessor. Ten point subdivision of molecular surface-spanning grid lines and harmonic averaging increased the initialization time to 14 seconds. However, in the case of the Gaussian model, there was no increase in initialization time. Initialization for the GAUSS model involves evaluation of the Gaussian volume and dielectric, Equation 2.40, at the midpoint of each grid line. However, Equation 2.42 includes the Gaussian volume and dielectric at the grid vertices only, and explicit subdivision of grid lines is not necessary. Therefore, the minimal computational cost of evaluating

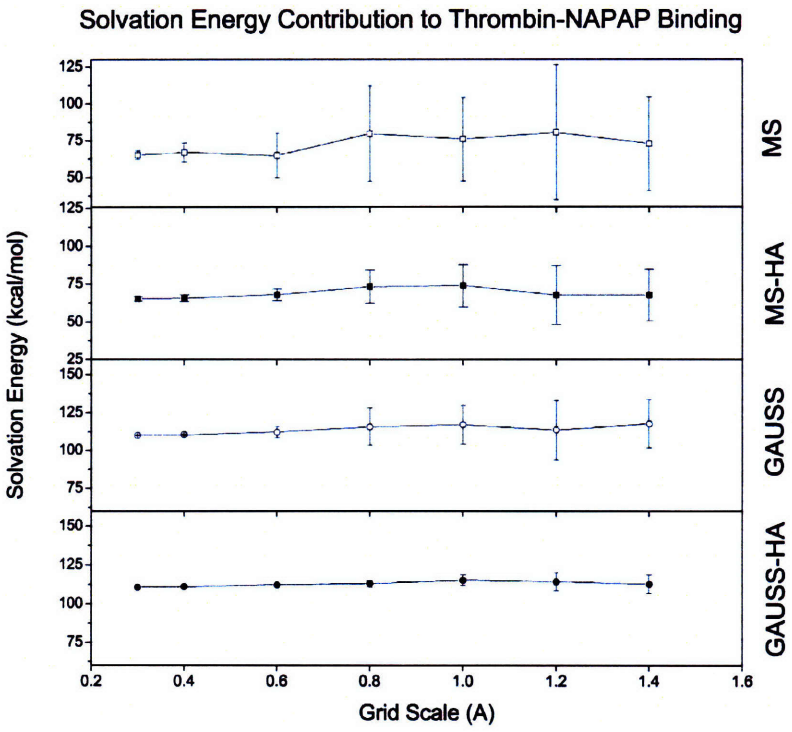

Figure 2-9: Solvation energy contribution to bovine thrombin-NAPAP[16] binding energy. Each point represents the difference in solvation energy between the complex and its individual componenets, each averaged over 20 random positions relative to the finite-difference grid. Error bars represent the standard deviation. From top to bottom, (MS) the traditional molecular surface model, (MS-HA) harmonically averaged molecular surface, (Gauss) Atomic Gaussian volume descriptors, (Gauss-HA) harmonically averaged Gaussian volume descriptors.

the harmonic average is recovered by evaluating the Gaussian density and dielectric at the grid vertices only, rather than at the midpoints of each grid line. Initialization times were 2.2 and 2.1 seconds for the GAUSS and GAUSS-HA models, respectively.

2.4 Verification of Implementation

For this project, a Poisson-Boltzmann solver was independently developed rather than application of existing software. The reason for this was two-fold: flexibility in development of model definitions and computational algorithms and the necessity

for an extremely fast implementation in order to feasibly accommodate the large system size to be studied. Prior to implementation of the newly developed PBE solver, the accuracy of solvation energy calculations must be verified against an established version of the model. For this purpose, the commercially available and widely considered industry standard software Delphi was used. A series of peptide solvation energies were calculated using the developed software, termed PBD, and Delphi. A set of 60 random peptide sequences were generated ranging from 5 to 160 amino acids, and atomic coordinates were generated using the *pepz* program distributed with MCPRO[29]. This program builds full peptide structures from a predefined database of atomic coordinates. Solvation energies from linearized PBE calculations in implicit water, modeled as a continuum dielectric of $\epsilon = 80$ and 0 ionic strength are compared against Delphi in Figure 2-10 for a grid spacing of 0.5 Å, and a convergence criteria of 10^{-3} kcal/mol. PBD solvation energies are nearly identical to Delphi results over a wide range of solvation energies.

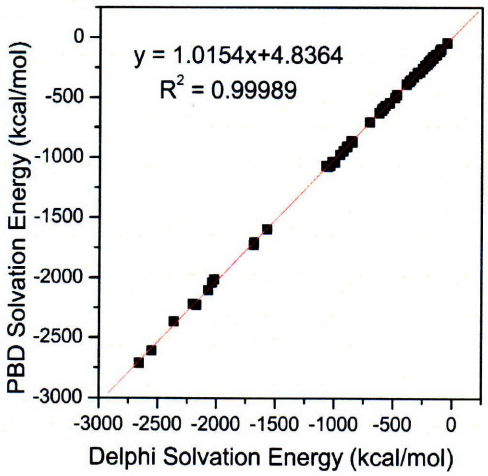

Figure 2-10: Accuracy comparison of newly developed PBE solver in comparison to the industry standard software Delphi for linear calculations with zero ionic strength solution

In Figure 2-11, the accuracy of non-linear calculations in ionic solution environments is illustrated. In these plots, the solvation energy of each peptide used in the linear comparison was calculated using the non-linear PBE at each ionic strength. The

reaction field energy in zero ionic strength water has been subtracted from each solvation energy in order to more accurately compare the effects of secondary ions. PBD again produces nearly identical results as Delphi over the range of ionic strengths.

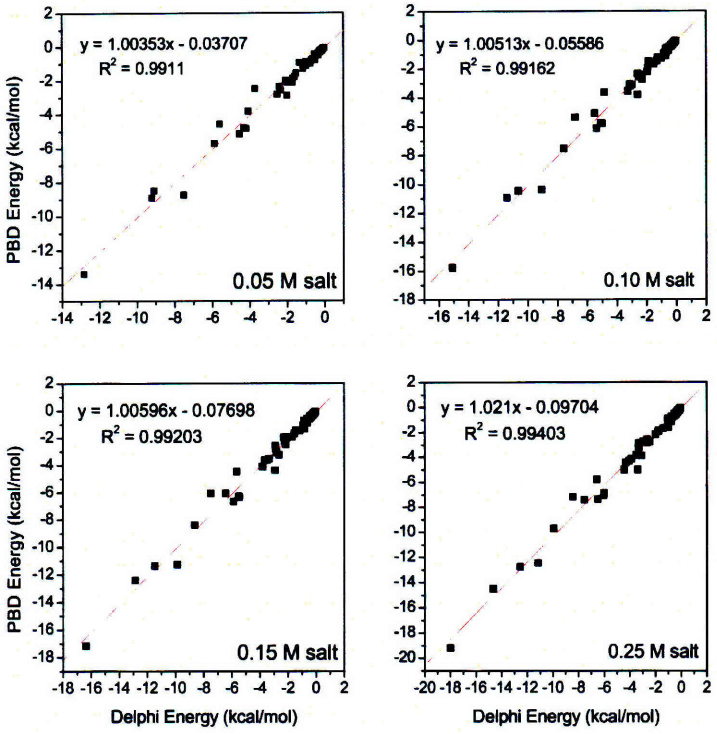

Figure 2-11: Accuracy comparison of newly developed PBE solver in comparison to the industry standard software Delphi for non-linear calculations in varying salt concentrations. Plotted energies are the difference in solvation energy from the zero ionic strength calculation, representing the contribution of the secondary ions to the total solvation energy.

Next, the computational efficiency of the new PBE implementation was assessed in comparison to Delphi. In Figure 2-12, the average calculation time is examined as a function of system size. Calculation times depend directly on the size of the finite-difference grid, and therefore indirectly on the molecule size. Converged solvation energies require approximately six times longer for linear and ten times longer for non-linear calculations for Delphi in comparison to PBD.

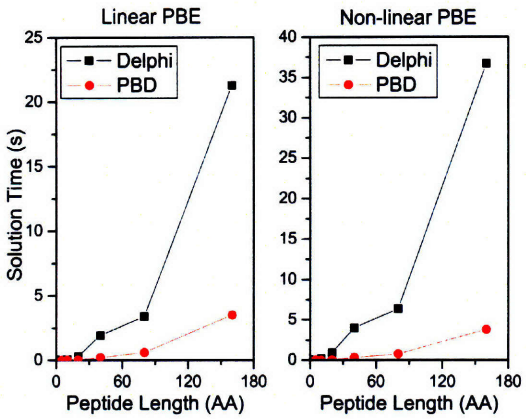

Figure 2-12: Poisson-Boltzmann calculation efficiency for PBD and Delphi. The multigrid method of PBD significantly outperforms the SOR method used in Delphi

2.5 Effective Grid Deformation Error

Examining a two atom system rotating relative to the finite-difference grid reveals an interesting consequence of the FD approximations. Luty *et al.* showed that the reference solution of the PBE could be substituted with a direct sum of Coulombic interactions by accounting for the effective distances imposed by the finite-difference grid[96]. They demonstrated the Green's function for the electrostatic potential on a cubic finite-difference grid in a uniform dielectric is

$$\begin{aligned}\phi(i,j,k,i',j',k') &= \frac{2q}{\epsilon h N^3} \sum_{k_x,k_y,k_z=1}^{N-1} \mathrm{CS}\left(\frac{k_x\pi i}{N}\right)\mathrm{CS}\left(\frac{k_x\pi i'}{N}\right)\mathrm{CS}\left(\frac{k_y\pi j}{N}\right) \\ &\times \mathrm{CS}\left(\frac{k_y\pi j'}{N}\right)\mathrm{CS}\left(\frac{k_z\pi k}{N}\right)\mathrm{CS}\left(\frac{k_z\pi k'}{N}\right) \\ &\times \left[\sin^2\left(\frac{k_x\pi}{2N}\right)+\sin^2\left(\frac{k_y\pi}{2N}\right)+\sin^2\left(\frac{k_z\pi}{2N}\right)\right]^{-1} \end{aligned} \tag{2.44}$$

where (i',j',k') is the grid point of the source charge $(-N/2 \leq i',j',k' \leq N/2)$, (i,j,k) is the grid point of observation, q is the charge, ϵ is the dielectric constant, h is the grid spacing, and N is the number of grid points. The function $\mathrm{CS}(k_\alpha\theta)$ is

given by

$$\mathrm{CS}(k_\alpha\theta) = \begin{cases} \cos(k_\alpha\theta) & k_\alpha \in \text{odd} \\ \sin(k_\alpha\theta) & k_\alpha \in \text{even} \end{cases} \tag{2.45}$$

Without loss of generallity, placing the charge at the origin yields

$$\begin{aligned} \phi(\Delta i, \Delta j, \Delta k) = \tfrac{2q}{\epsilon h N^3} \textstyle\sum_{k_x,k_y,k_z=1}^{N/2} \\ \times \cos\left[\tfrac{(2k_x-1)\pi\Delta i}{N}\right] \cos\left[\tfrac{(2k_y-1)\pi\Delta j}{N}\right] \cos\left[\tfrac{(2k_z-1)\pi\Delta k}{N}\right] \\ \times \left(\sin^2\left[\tfrac{(2k_x-1)\pi}{2N}\right] + \sin^2\left[\tfrac{(2k_y-1)\pi}{2N}\right] + \sin^2\left[\tfrac{(2k_z-1)\pi}{2N}\right]\right)^{-1} \end{aligned} \tag{2.46}$$

and then letting the grid become infinite, i.e., $\lim_{N\to\infty}$

$$\begin{aligned} \phi(\Delta i, \Delta j, \Delta k) = \frac{q}{\epsilon 4\pi^3 h} \int_0^\pi \int_0^\pi \int_0^\pi dx\, dy\, dz \\ \times \frac{\cos(x\Delta i)\cos(y\Delta j)\cos(z\Delta k)}{\sin^2\left(\frac{x}{2}\right) + \sin^2\left(\frac{y}{2}\right) + \sin^2\left(\frac{z}{2}\right)} \end{aligned} \tag{2.47}$$

Writing this equation in a form which maintains the functional relationahip of Coulomb's law gives

$$\phi(\Delta i, \Delta j, \Delta k) = \frac{q}{4\pi\epsilon h d_{eff}(\Delta i, \Delta j, \Delta k)} \tag{2.48}$$

where,

$$\begin{aligned} d_{eff}(\Delta i, \Delta j, \Delta k)^{-1} = \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \int_0^\pi dx\, dy\, dz \\ \times \frac{\cos(x\Delta i)\cos(y\Delta j)\cos(z\Delta k)}{\sin^2\left(\frac{x}{2}\right) + \sin^2\left(\frac{y}{2}\right) + \sin^2\left(\frac{z}{2}\right)} \end{aligned} \tag{2.49}$$

The integrals in Equation 2.49 were evaluated by Gaussian quadrature and are listed in Table 2.2. Distances parallel to the grid axes are effectively contracted, while distances along the grid diagonals are effectively dilated. Figure 2-13 illustrates this effective grid deformation in two dimensions. As long as spherical symmetry or rotational orientation to the grid is maintained, this effective deformation of space is not expected to have important consequences on the calculated solvation energy. However, for an asymmetric system rotating relative to the FD grid, as might occur

in docking, for example, distances between atoms, charges and dielectric boundaries are effectively changing and therefore altering the solvation energy.

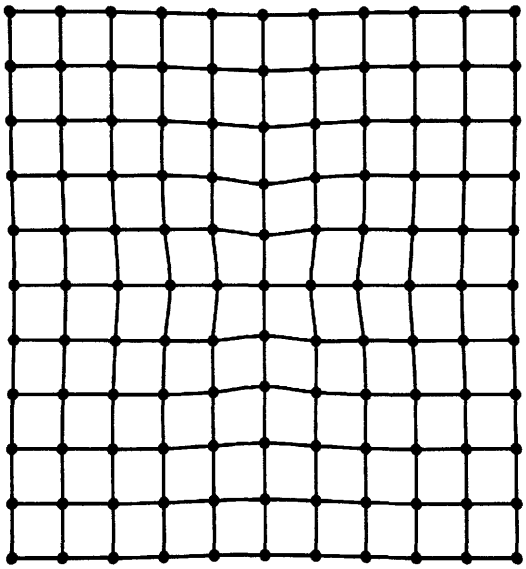

Figure 2-13: Illustration of the effective grid deformation imposed by finite-difference approximations. Distances parallel to the grid axes are effectively contracted, while distances along the grid diagonal are effectively dilated.

In order to explore the size of this effect, a charged atom with $q = 0.5e$, was placed at the origin of the FD grid and a second, uncharged atom was rotated around it in the x-y plane. This system was chosen in order to isolate the effects of the grid deformation from any changes in the mapping of charge on the grid. The finite-difference grid was constructed with a spacing of 0.3 Å and the atomic volumes were mapped by Equation 2.42 with radii of 1.25 Å. Figure 2-14 shows the deviation of the solvation energy as a function of the rotation angle. The solvation energy varies smoothly as the pair rotates, showing a period of 90° due to the symmetry of the rectangular grid. Figure 2-15A shows the amplitude of the energy fluctuation as a function of atomic separation for grid spacings of 0.40, 0.35, 0.30, and 0.25 Å and atomic radii of 2.25 Å. Here, the amplitude is considered to be positive if the solvation energy in the diagonal alignment is higher (more positive) than the parallel

alignment and negative if it is lower (more negative). At small atomic separations, the diagonal alignment is higher in energy, peaking at a separation approximately equal to the atomic radius. At larger separations, the diagonal alignment is lower in energy, peaking at a separation of approximately twice the atomic radius, or the point of atomic contact. Finally, at very large separations, the amplitude returns to zero as the effective deformation of the FD grid becomes small. Inset, the value of the maximum at separation $r = 2.25$ Å (marked by 3C) is a linear function of the difference in effective distance $\Delta d_{eff}(r = 2.25\text{Å})$ between the parallel and diagonal directions for grid spacings 0.25-0.50 Å.

Figures 2-15B and 2-15C show an overlay of the dielectric cavities formed by the pair of atoms in the parallel (black) and diagonal (red) alignments for the separations marked in Firure 2-15A. The difference between effective distances parallel and diagonally aligned to the grid causes some areas of the dielectric boundary to be closer to the central atomic charge than others. At small atomic separation (2-15B), it is apparent that the portion of the boundary belonging to the charged atom is on average closer to the charge center in the parallel alignment than in the diagonal alignment. Thus, molecular rotation results in an effective increase in radius and corresponding increase in solvation energy. However, at larger separations the reverse is true. The diagonal alignment shows a significantly higher penetration of water into the neck region between the two atoms, and therefore a decrease in solvation energy as the pair is rotated.

The deformation of the finite-difference grid is subtle, applying a systematic error to energy calculations. The error is small in magnitude and was previously masked by the more irregular errors associated with the discretized representation of the dielectric boundary. In experimental applications, the case of highly overlapping atoms in Figure 2-15B is representative of bonded pairs, while the case in Figure 2-15C is near the minimum in non-bonded van der Waals energy. Rotation of a molecule produces relative rotations amongst each pair of its atoms, with effective distance errors combining to produce a cumulative error of the molecule. One would expect minimual implications for systems with roughly spherical symmetry, while

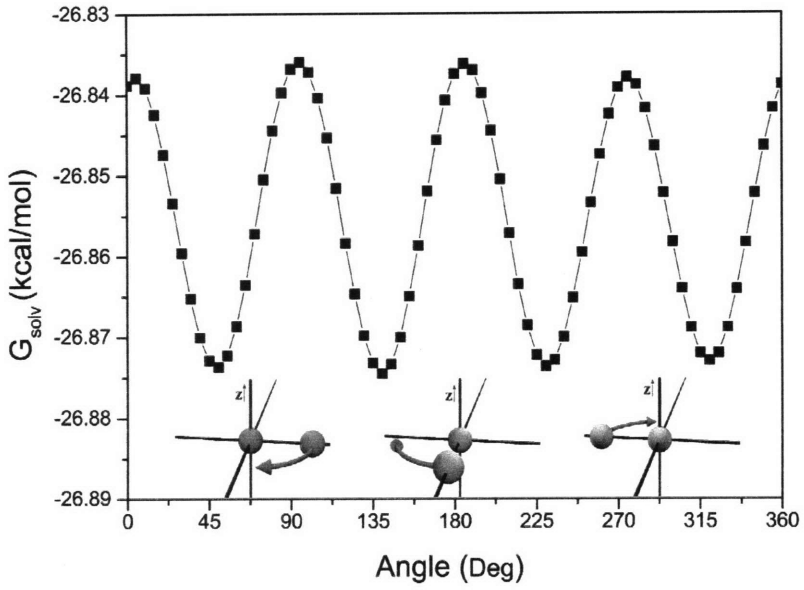

Figure 2-14: Solvation energy fluctuation as an atom pair is rotated relative to the finite-difference grid.

systems that are largely extended in one dimension, such as an extended peptide conformation, may have larger errors. In a sample calculation with a grid spacing of 0.4 Å, the dodecamer peptide $(GK)_6$ in a fully extended conformation ($\phi = \psi = 180$) shows a preferred diagonal alignment by approximately 1 kcal/mol, while at a grid spacing of 1.0 Å, the diagonal alignment is favored by 5 kcal/mol. These errors are similar in magnitude to thermal fluctuations at small grid spacings, but may be important at larger grid scales.

Except for the special circumstance of a linear arrangement, the effective distances cannot be corrected for more than two points without altering their geometry. The deformation of the FD grid is local, directional, and relative to a central point. Thus, while the effective position of charges and dielectric boundaries could be corrected relative to one atom, the same adjustments will not apply arbitrarily to a second atom. This is demonstrated in Figure 2-16 for a three point system. Attempting to correct for the effective distances by shortening distances along the diagonal and lengthening distances parallel to grid axes cannot, in general, be accomplished without altering

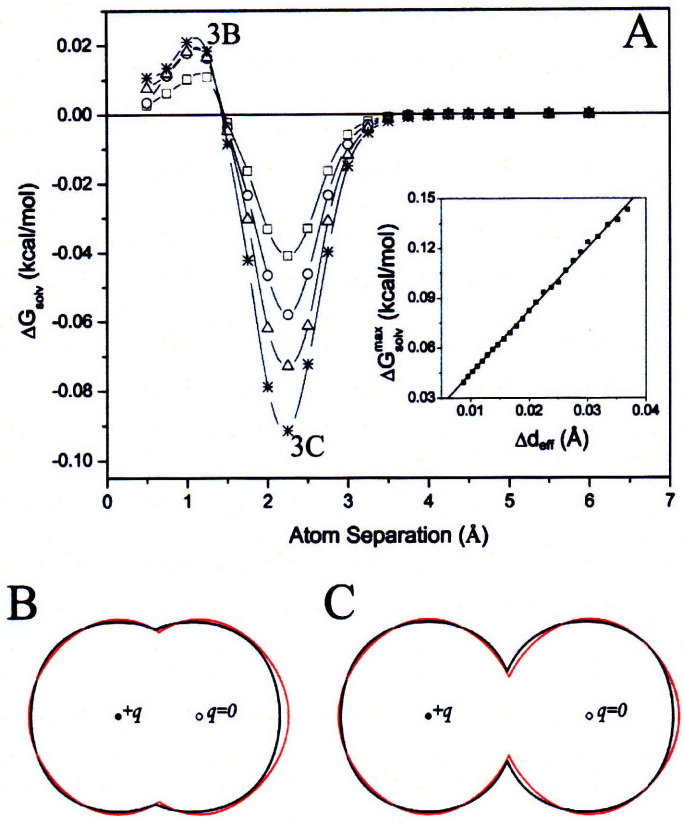

Figure 2-15: (A) Dependence of the solvation energy fluctuation on grid size for spacings of 0.40 (∗), 0.35 (△), 0.30 (○), and 0.25 (□) Å shows a linear relation to the difference in effective distance between the parallel and diagonal directions (inset). (B&C) Overlay of dielectric boundaries based on effective distances for the parallel (black) and diagonal (red) alignments at the separations marked in 3A.

the geometry of the system. Reduction of grid scale remains as a viable solution to eliminate orientational error.

2.6 Conclusions

Poisson-Boltzmann solvent models represent the most rigorous of the implicit solvent models. The ultimate goal of this project is the incorporation of the implicit solvent model into a molecular mechanics simulation package. To this end, there are two primary concerns; accuracy of the solvation model and speed of calculation.

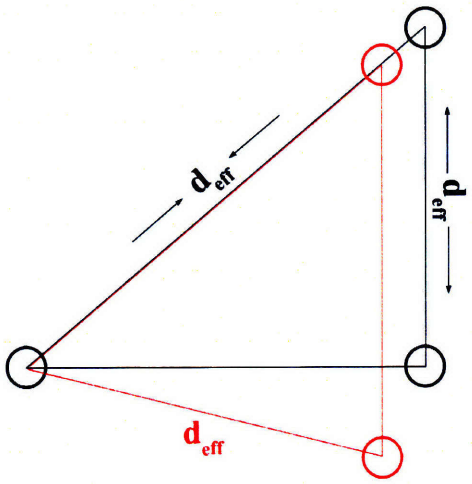

Figure 2-16: In general the effective grid distances cannot be corrected in a simple fashion. Considering the three point system shown here, the distance along the diagonal must be shortened to correct for of the effective dilation. However, the vertical distance must be lengthened to correct for the effective contraction, which combined with the previous correction alters the overall geometry of the three point system. Only the linear case presents a straightforward scheme for approximate correction of effective grid distances.

Through efficient multigrid calculations and the application of local averaging to smooth permittivity functions, a Poisson-Boltzmann solver was developed with exceptional speed and accuracy. The implementation developed through this project is considerably faster than commercially available software packages. Solvation energy calculations were also shown to be more stable at larger grid scales further increasing the computational efficiency by extending the range of grid scales which can be used. However, in the process of removing errors associated with the discretization of molecular volume, an underlying orientational error fundamental to the the finite-difference approximations was uncovered.

This orientational error is incompattible with a molecular simulation. It is particularly damaging due to the deterministic nature of the orientational dependence. Tests incorporating this model into Monte Carlo simulations revealed that molecular position and orientation, as well as conformation, were affected by this error. Short peptide systems preferrentially aligned themselves with the grid diagonal and were

prevented from conformational changes that would bring portions of the peptide out of this alignment.

Grid scale reduction remains as an option for eliminating error in the solvation energy calculations. However, limiting the error to a level appropriate within the framework of a molecular mechanics simulation requires a grid scale that is not feasible in the sense of computational efficiency. Monte Carlo simulations must average properties over many conformations, each of which accompanied by a solvation energy calculation. At even one second per solvation energy calculation, a one million step Monte Carlo simulation will take on the order of 10 days. Limiting the diagonally aligned preference to $\sim$0.5 kcal/mol results in a computational wall time for a single solvation energy calculation of roughly 4 seconds. In the next chapter, we turn to approximations to the Poisson-Boltzmann equation which lead to greater computational efficiency.

Table 2.1: Grid Stability in Solvation Energy Calculations. Solvation energies for each small molecule were calculated at 100 random positions relative to the finite-difference grid, while each protein was sampled at 20 random positions. All energies are reported in kcal/mol and given as the mean +/- the standard deviation. Atomic parameters were taken from the Optimized Parameters for Liquid Simulations (OPLS)[140], except for charged hydrogens which have a radius of zero in the OPLS force field. Such a radius is inappropriate for Poisson electrostatic calculations, and therefore has been reset to 0.8 Å. A grid spacing of 1.0 Å and relative dielectric constants of 1 and 80 were used for the interior and exterior values, respectively, for all calculations. The range is the difference between the maximum and minimum calculated solvation energies. $C7^{eq}$-Ala, C5-Ala, and α_R-Ala represent different conformations of the alanine dipeptide, as described elsewhere[24].

	MS		MS-HA		Gauss		Gauss-HA	
Molecule	E_{solv}	Range	E_{solv}	Range	E_{solv}	Range	E_{solv}	Range
Methanol	-10.26 ± 1.52	8.49	-8.70 ± 0.55	1.94	-8.47 ± 0.60	2.24	-7.15 ± 0.17	0.66
Ethanol	-9.86 ± 1.46	5.79	-8.37 ± 0.30	1.43	-7.89 ± 0.48	1.75	-6.49 ± 0.12	0.59
2-Propanol	-9.80 ± 1.63	5.96	-8.25 ± 0.29	1.48	-7.78 ± 0.62	2.68	-6.32 ± 0.15	0.65
Acetone	-6.08 ± 0.90	4.16	-5.29 ± 0.11	0.55	-4.94 ± 0.37	1.64	-4.33 ± 0.06	0.23
Methyl Acetate	-5.41 ± 0.79	3.44	-4.52 ± 0.18	0.78	-4.18 ± 0.29	1.35	-3.53 ± 0.12	0.43
Acetic Acid	-96.75 ± 6.57	23.91	-90.77 ± 0.66	3.26	-90.54 ± 3.05	12.24	-85.09 ± 0.34	1.31
Acetamide	-15.25 ± 1.18	5.59	-13.02 ± 0.32	1.51	-12.96 ± 0.53	2.41	-10.86 ± 0.17	0.64
$C7^{eq}$-Ala	-20.84 ± 1.84	7.82	-17.82 ± 0.39	1.87	-16.83 ± 0.71	3.11	-14.07 ± 0.16	0.67
C5-Ala	-23.74 ± 1.93	7.24	-20.72 ± 0.31	1.15	-19.44 ± 0.84	3.54	-16.95 ± 0.22	0.81
α_R-Ala	-24.50 ± 1.57	6.21	-21.76 ± 0.37	1.95	-20.62 ± 0.70	2.93	-18.22 ± 0.13	0.48
1GQV	-3587.89 ± 23.89	98.97	-3273.0 ± 9.27	48.92	-2837.55 ± 6.59	27.71	-2550.87 ± 1.60	6.33
1HJE	-160.34 ± 4.95	22.69	-140.02 ± 1.91	8.85	-119.00 ± 1.96	9.02	-103.01 ± 0.41	1.81
1KCH	-651.35 ± 9.71	46.10	-567.69 ± 4.47	19.92	-432.16 ± 1.74	7.91	-364.98 ± 0.59	2.43
2FDN	-6641.30 ± 24.39	99.88	-6497.97 ± 4.93	23.10	-6367.31 ± 8.78	36.00	-6238.17 ± 1.84	7.25

Table 2.2: Values of effective grid distances $d_{eff}(\Delta i, \Delta j, \Delta k)$.

$\Delta i, \Delta j \Delta k$	Value	Distance
000	0.31488	0.00000
100	0.92464	1.00000
110	1.44187	1.41421
111	1.82612	1.73205
200	1.85546	2.00000
210	2.21476	2.23607
211	2.48984	2.44949
220	2.83690	2.82843
221	3.04260	3.00000
300	2.88944	3.00000
310	3.11799	3.16228
311	3.31138	3.31662
222	3.51331	3.46410
320	3.60502	3.60555
321	3.76136	3.74166
400	3.92243	4.00000
322	4.15702	4.12311
410	4.07658	4.12311
330	4.25236	4.24264
411	4.21765	4.24264
331	4.37981	4.35890
420	4.46023	4.47214
421	4.58287	4.58258
332	4.72241	4.69042
422	4.91463	4.89898
430	5.00555	5.00000
500	4.94270	5.00000

Chapter 3

Generalized Born Implicit Solvent Model

3.1 Introduction

Dealing with solvent in molecular simulations has been a perpetual problem in computational chemistry. Explicit inclusion of solvent requires computationally expensive averaging if one is to achieve converged results. Implicit solvent models have become a popular alternative[48, 109] with the Poisson-Boltzmann equation, discussed in Chapter 2, providing the standard of reference. However, Poisson-Boltzmann calculations still represent a significant computational investment when grid-position stability is required, as in the case of incorporation into a molecular mechanics simulation. Instability of the discretized calculation can preclude comparison of molecular solvation energies following translational, rotational, or conformational changes. Therefore, fine grid scales are required to stabilize the calculation which, in turn, greatly increases the computational demands.

A popular approximation to the Poisson-Boltzmann model is the so called "Generalized Born" method[55]. The total electrostatic free energy, G_{es}, of a system of separated, charged atoms in a medium of dielectric ϵ is given by the sum of Coulombic

interactions and Born solvation energies

$$G_{es} = 332 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{r_{ij}\epsilon} - 166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i}^{n} \frac{q_i^2}{\sigma_i} \tag{3.1}$$

where q is the atomic charge, r_{ij} is the interatomic separation, and σ_i is the atomic radius. Expanding the Coulombic interaction energy into the vacuum Coulombic energy and a term which accounts for the effect of the dielectric yields

$$G_{es} = 332 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{r_{ij}} - 332 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{r_{ij}} - 166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i}^{n} \frac{q_i^2}{\sigma_i} \tag{3.2}$$

Finally, the similar form of the second and third terms prompts their combination to give the polarization free energy of solvation

$$G_{pol} = -166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{q_i q_j}{f_{GB}} \tag{3.3}$$

where the function f_{GB} is a non-uniquely defined function of σ_i and r_{ij}. The commonly used form of this expression is

$$f_{GB} = (r_{ij}^2 + \sigma_{ij}^2 e^{-r_{ij}^2/4\sigma_{ij}^2})^{1/2} \tag{3.4}$$

where $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$. At zero separation, this function reduces to the Born radius, while at large distances the function reduces to the separation distance. For a molecule consisting of a set of partially overlapping atoms, the radius used in this model is not simply the atomic radius, but is instead replaced by an effective radius which accounts for the displacement of solvent by all other atoms in the molecule. Specifically, the Born radius of atom i is the effective radius of a hypothetical spherical particle with exactly the solvation energy of the molecule when only atom i is charged. In contrast to distance dependent dielectric models, the function in Equation 3.4 takes into account the degree of solvent exposure in addition to atom separation. Given an accurate Born radius for each atom in a given molecule, the empirical formula for f_{GB}

produces very good agreement with more rigorous treatments[5, 43]. Strictly speaking, the accurate computation of effective Born radii therefore requires the solution of the Poisson-Boltzmann equation for each atom in the molecule. This, of course, does not lead to any computational advantage, and methods for approximating these Born radii has been the major focus of continuing efforts in the development of Generalized Born methods. Many attempts have been made to parameterize Born radii, as any other element of a molecular mechanics force-field, by fitting solvation energies to experimental and explicit solvent simulations[69, 121]. However, this approach has not produced sufficiently accurate results and it is apparent that a method taking into account the effect of the molecular shape on Born radii is necessary[93].

3.2 Coulomb Field Approximation

The traditional method for approximation of Born radii relies on the Coulomb Field Approximation (CFA)[25, 77]. The electrostatic energy of a solute consisting of N atoms with charges, $q_1 \dots q_N$, can be evaluated by integration of the energy density of the eletric field over all space, i.e.

$$E^{el} = \frac{1}{8\pi\epsilon_s} \int_{solvent} \mathbf{D}^2(\mathbf{r}) d\mathbf{r} + \frac{1}{8\pi\epsilon_i} \int_{solute} \mathbf{D}^2(\mathbf{r}) d\mathbf{r} \tag{3.5}$$

where $\mathbf{D}$ is the dielectric displacement, $\mathbf{r}$ is the position, ϵ_s and ϵ_i are the solvent and solute dielectrics, respectively, and the integration over all space has been split into the solvent and solute regions of space. By adding and subtracting the integral of $\mathbf{D}^2/(8\pi\epsilon_s)$ over the solute volume

$$E^{el} = \frac{1}{8\pi\epsilon_s} \int_{R^3} \mathbf{D}^2(\mathbf{r}) d\mathbf{r} + \frac{\tau}{8\pi} \int_{solute} \mathbf{D}^2(\mathbf{r}) d\mathbf{r} \tag{3.6}$$

where R^3 represents the integration over all space, and $\tau = 1/\epsilon_i - 1/\epsilon_s$. In the first integral, the dielectric displacement is approximated by the Coulomb field, introducing a relative error of only a few percent in the electrostatic energy[44]. This is justified by observing that for a small solute, charges are highly exposed and deviation from the

Coulombic field is small, while for large solutes the integration over the solute volume in the second integral will dominate the total electrostatic energy[77]. Integrating the Coulomb field over all space yields the Coulomb interaction energy, $q_i q_j/\epsilon_s r_{ij}$, in the off-diagonal terms $2\mathbf{D}_i \cdot \mathbf{D}_j$, and the Born self energy, $q_i^2/2\epsilon_s\sigma_i$, in the diagonal terms $\mathbf{D}_i^2$. Distinguishing the self energy and Coulombic interaction energy terms in the total electrostatic energy yields

$$E^{el} = \sum_i E_i^{self} + \sum_{i<j} E_{ij}^{int} \tag{3.7}$$

$$E_i^{self} = \frac{q_i^2}{2\epsilon_s\sigma_i} + \frac{\tau}{8\pi}\int_{solute} \mathbf{D}^2(\mathbf{r})d\mathbf{r} \tag{3.8}$$

$$E_{ij}^{int} = \frac{q_i q_j}{\epsilon_s r_{ij}} + \frac{\tau}{4\pi}\int_{solute} \mathbf{D}_i \cdot \mathbf{D}_j(\mathbf{r})d\mathbf{r} \tag{3.9}$$

leaving only an integral over the finite volume of the solute. In order to evaluate this integral over the solute volume, the Coulomb Field Approximation is again applied. The dielectric displacement at point $\mathbf{r}$ due to charge i is approximated by the Coulomb field, $\mathbf{D}_i(\mathbf{r}) = q_i/\mathbf{r}$. This allows the integration of the dielectric displacement without prior knowledge of the electrostatic potential. This approximation in effect ignores the reaction field contribution to the dielectric displacement. Therefore, the CFA can be expected to lead to an overestimation of the electrostatic energy.

The inverse relationship of the electrostatic energy to the Born radius leads to

$$\frac{1}{B_i} = \frac{1}{\sigma_i} - \frac{1}{4\pi}\int_{solute, r'>a_i} \frac{d\mathbf{r}'}{r^4} \tag{3.10}$$

where B_i is the effective Born radius of atom i, and the integration of r^{-4} is taken over the solute volume outside of atom i. The integration in Equation 3.10 can be carried out numerically[71] or analytically[89], by considering the molecular volume as the sum of overlapping atomic spheres.

3.3 Deficiency in the CFA

The Coulomb Field Approximation is exact for single charge located at the center of a spherical cavity in a homogeneous dielectric. However, for nonspherical molecular geometries and charge distributions, the CFA must be justified by the short range nature of the electric field. The first solvation shell around an average atom accounts for 58% of the self-energy and solvation energy[77]. The maximum error of the CFA can be exemplified by a single charge positioned at an infinite planar dielectric boundary. For this geometry, the electric field can be solved analytically by the image charge solution, and the self energy integral is overestimated by 59%. When the charge is separated from the dielectric boundary by one atom layer, the error is reduced to 9.4%.

Since this type of system represents exactly the system of interest in application to surface adsorption simulations, the traditional GB formulation based on the CFA is not appropriate. Even the separated error estimate of 9.4% would produce a significant influence on surface adsorption of biopolymers at an inorganic-solvent interface. Therefore, we must seek a correction to the GB-CFA model which better handles charges at an extended dielectric boundary.

3.4 Alternative Calculations of Born Radii

Most corrections to the Coulomb Field Approximation take the form of a higher order integration of the distance[49, 70, 71]. These methods add a corrective term to the Coulomb integration, generally of the form

$$A_{3+n} = \left(\frac{1}{\sigma^n} - \frac{1}{4\pi} \int_{solute, r>\sigma} \frac{1}{r^{3+n}} d\mathbf{r} \right)^{1/n} \tag{3.11}$$

The Born radii are then calculated as

$$B_i = \frac{S}{C_0 A_4 + C_1 A_{3+n}} + D \tag{3.12}$$

where S, C_0, C_1, and D are adjustable parameters, and A_4 is the CFA integral. These corrective terms are empirically designed and seek to correct the overestimation of the long range effects of the dielectric displacement, hence the use of more quickly decaying functions. The adjustable parameters in Equation 3.12 have no physical meaning and must be optimized by comparison of solvation energies against the Poisson-Boltzmann calculations we seek to approximate.

As a first implementation, the corrective integration term was included as the integration of $1/r^5$ as suggested by Lee *et al.*, ($n = 1$ in Equation 3.11). Equation 3.12 was parameterized as $C_0 = -1$, $C_1 = 2\sqrt{2}$, $S = 1$, and $D = -0.38$. Integration over the molecular volume was carried out numerically using a spherically symmetric grid centered at each atom. The integration grid was constructed by equal division of the radial coordinate up to a cut-off distance, and an on-the-fly adjustable division of the angular coordinates. A maximum subdivision for the azimuthal angle is predefined, and the zenith coordinate subdivision was calculated as $(2N_\phi - 1)|\sin\phi| + 1$. The parameter N_ϕ is adjusted to give some spatial uniformity to the grid points. Romberg integration was used over the grid to give the value of the integrals in Equation 3.12.

The computational cost and relative error of the numerical integration were investigated to determine the optimal balance of grid density. Figure 3-1 illustrates the time and error scaling with grid subdivision in the numerical integration routine for a solvation energy calculation of blocked alanine. In Figure 3-1a, the time scaling properties for increasing subdivision in the radial coordinate, N_r, for a constant subdivision of the angular coordinate, N_ϕ. Figure 3-1c shows a similar plot for increased subdivision of the angular coordinate ϕ with constant radial subdivision, N_r. Computational costs scale exponentially with each coordinate, however, the error calculated solvation energies are significantly more dependent on the subdivision of the radial coordinate, illustrated in Figure 3-1b and d. Therefore, all further Generalized Born calculations using this numerical integration method utilize a fine radial grid spacing of $N_r = 125$ grid points and a moderate angular distribution of $N_\phi = 16$.

The accuracy of this model was further assessed by solvation energy calculations for amino acids and short peptide molecules. Poisson-Boltzmann reaction field en-

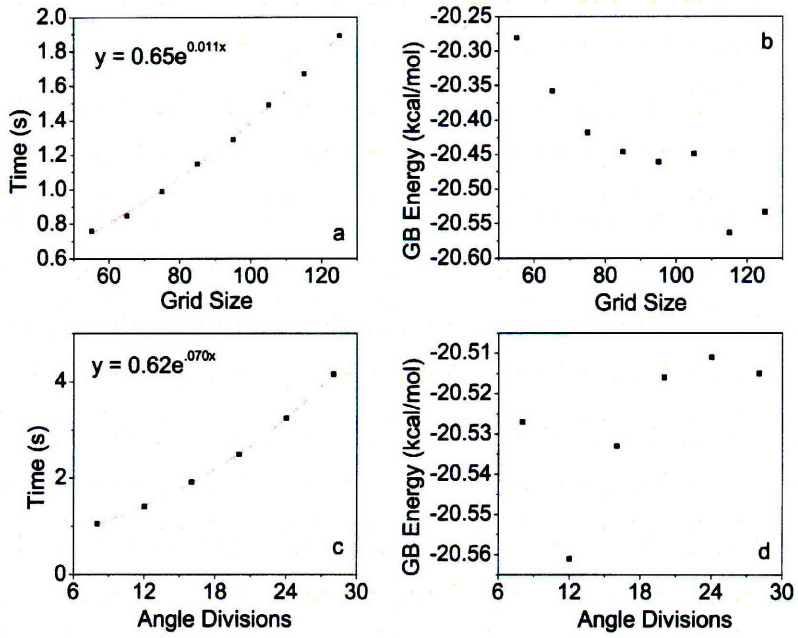

Figure 3-1: Time and error scaling with grid subdivision in numerical integration for evaluation of Born radii (a) Time scaling for radial subdivision ($N_\phi = 16$), (b) Error assessment for radial subdivsion, (c) Time scaling for angular subdivision ($N_r = 125$), and (d) Error assessment for angular subdivision

ergies were calculated with Delphi[66] using atomic parameters from the OPLS-AA force field[140], with the exception of charged hydrogen atoms which have a radius of 0.0 in the OPLS force field. Such a radius is inappropriate in the context of a volume based solvation energy calculation and hydrogen radii were set to 1.0 Å. Grid spacing for PB calculations was set as 0.5 Å and salt concentration was set to zero. Correlation with PB solvation energies is excellent in the GB method for the twenty naturally occuring amino acids, illustrated in Figure 3-2. The average error in these calculations was 0.96%.

A similar analysis was carried out for a set of ten random dodecamer peptide sequences in order to verify applicability to the simulation of short peptide sequences. The peptide sequences listed in Table 3.1 were generated using the *pepz* utility distributed with MCPRO[29]. This utility generates atomic coordinates from input

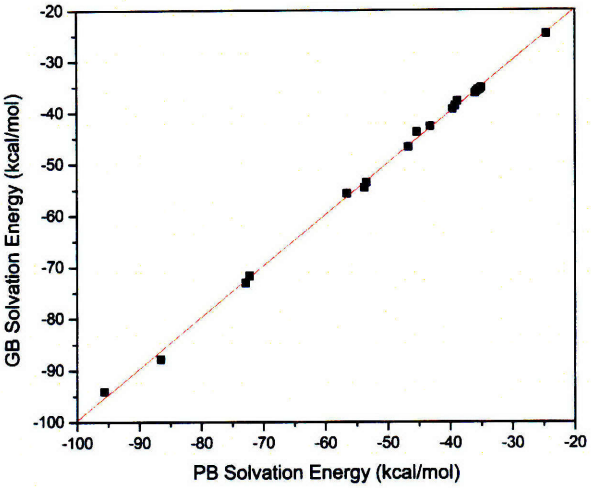

Figure 3-2: Solvation energy calculation for the 20 naturally occuring amino acids, comparing Generalized Born results with Delphi Poisson-Boltzmann reaction field energies.

peptide sequences and a predefined library of residue coordinates in a fully extended conformation ($\phi = \psi = 180$). The peptides were then energy minimized in vacuo using the conjugate gradient method of MCPRO. Solvation energies were calculated for the energy minimized conformations using Delphi and the Generalized Born implementation. Figure 3-3 illustrates the agreement between the GB implementation and Poisson-Boltzmann reaction field energies. Average relative error for peptide solvation energies was 1.30%.

Finally, the applicability of the GB implementation to surface adsorption simulations was tested by positioning charged atoms near an extended dielectric boundary. First, a dipole system consisting of a negative half-unit charge atom with a radius of 1.5 Å embedded in the surface of a 4 nm^2 dielectric slab, 1 nm in thickness and a positive half-unit charge positioned outside of the slab in the solvent region. The relative solvent dielectric was set as 80 while the slab and atom relative dielectric was set as 1. The solvation energy of the system was calculated for varying distances of the positively charged atom from the surface, illustrated in Figure 3-4. Solvation energy calculations reach a relative error of 30.4% at an atom separation of 3 Å compared

Table 3.1: Randomly generated peptide sequences used to verify applicability of Generalized Born solvation energies to the simulation of short peptide sequences.

Peptide Sequence	Charge
GAVLISTCMFYW	0
DNEQRHKPGAVL	+1
ISTCMFYWDNEQ	-2
WLACPHFSWQAC	+1
RHCIVNSCPQYS	+2
AVGILMFVVPGA	0
DNDHQTTYSREQ	-1
CIETQHGHPPCY	+1
TLGSYDCTEPIV	-2
TFFMEPHGVTDR	0

to Poisson-Boltzmann reaction field energies.

This test was repeated for a quadrupole system in order to determine how the interaction of more charges would effect the solvation energy calculation. In this test, a dipole was embedded in the surface, while a second dipole was moved toward the dielectric slab. Increasing the system charge only worsens the error in the Generalized Born calculations. For a charged peptide consisting of hundreds of atoms and an ionic crystalline surface, these errors would produce an unacceptable lack of precision and a further refined Generalized Born model must be developed.

3.5 Analytical Integration of GBr^6 Model

An alternative modified Generalized Born method, termed GBr^6[147], also relies on a higher order integration of the distance. However, the GBr^6 method uses the higher order function itself, rather than as a corrective term to the CFA. The Kirkwood model[79] of biopolymer electrostatics gives the electrostatic free energy of a charge, q, in a spherical dielectric cavity of radius a as

$$\Delta G^{KM} = \frac{-q^2}{8\pi\epsilon_0}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_s}\right)\frac{1}{a(1-p^2)} \tag{3.13}$$

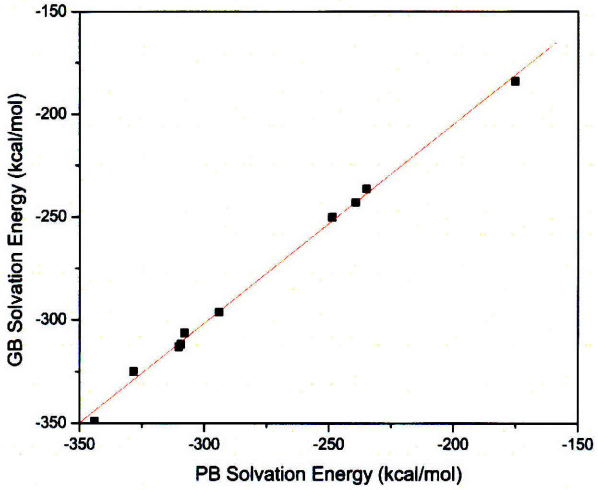

Figure 3-3: Solvation energy calculation for the 10 random dodecamer peptides in a gas phase energy minimized conformation, comparing Generalized Born results with Delphi Poisson-Boltzmann reaction field energies.

where ϵ_i and ϵ_s are the molecular and solvent dielectrics, respectively, and $p = d/a$ is a dimensionless factor with d the distance of the charge from the center of the spherical cavity. Comparison of this equation with the Born ion solvation energy gives the effective Born radius as

$$B = a(1 - p^2) \tag{3.14}$$

The general form of the above relationship can be achieved through a single integration, and we see that

$$\begin{aligned}
\int_{solvent} \frac{1}{r^6} d\mathbf{r} &= 2\pi \int_a^\infty dr\, r^2 \int_{-1}^{1} \frac{d\cos\theta}{(r^2 + d^2 - 2dr\cos\theta)^3} \\
&= \frac{\pi}{2d} \int_a^\infty dr \left(\frac{r}{(r-d)^4} - \frac{r}{(r+d)^4} \right) \\
&= \frac{\pi a}{3} \left(\frac{3}{(a^2 - d^2)^2} + \frac{a^2 + 3d^2}{(a^2 - d^2)^3} \right) \\
&= \frac{4\pi}{3} \left(\frac{1}{a(1-p^2)} \right)^3 = \frac{4\pi}{3} \frac{1}{B^3}
\end{aligned} \tag{3.15}$$

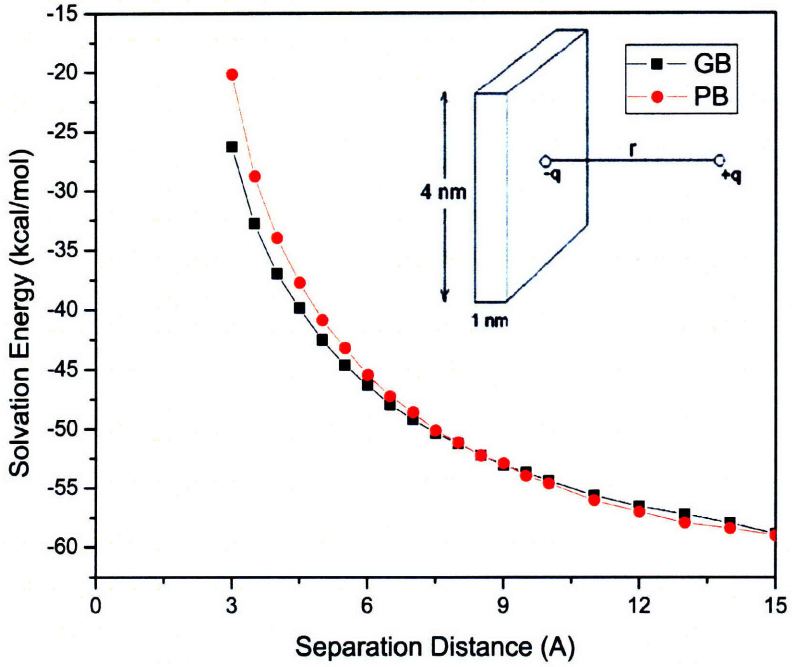

Figure 3-4: Testing the applicability of the Generalized Born solvent model for surface adsorption simulations. A dipole system with one charge embedded in a low dielectric surface. The Generalized Born implementation reaches an error of 30% at short separation distances.

and therefore,

$$\frac{1}{B} = \left(\frac{3}{4\pi} \int_{solvent} \frac{1}{r^6} d\mathbf{r} \right)^{1/3} \tag{3.16}$$

which can then be converted into an integration over the volume of the molecule as in Equation 3.10. This integration over the volume of the molecule can then be approximated by the sum of contributions from each atom. In order to calculate the Born radius of atom i, the integral is considered individually for each atom j over the portion of the van der Waals sphere which does not overlap atom i. By this decomposition, the integral Z_{ji} can be evaluated analytically following the method of Gallicchio and Levy[89]. There are four possible cases[147]:

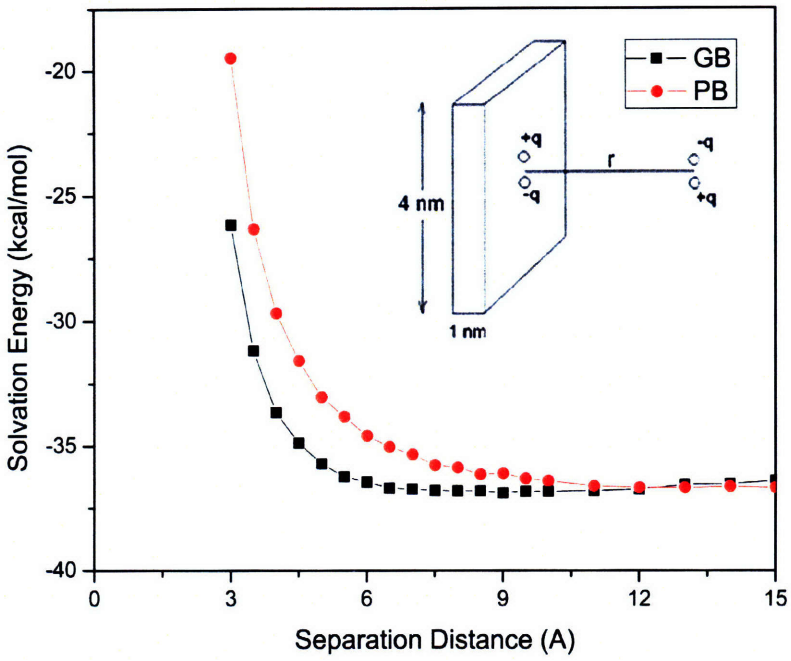

Figure 3-5: Testing the applicability of the Generalized Born solvent model for surface adsorption simulations. A quadrupole system with one dipole embedded in a low dielectric surface.

1. Atoms i and j do not intersect, $r_{ij} \geq \sigma_i + \sigma_j$

$$Z_{ji} = \frac{a_j^3}{(r_{ij}^2 - a_j^2)^3} \tag{3.17}$$

2. Atoms i and j intersect, but neither is completely occluded by the other, $(\sigma_i + \sigma_j)^2 > r_{ij}^2 \geq (\sigma_i - \sigma_j)^2$

$$Z_{ji} = -\frac{3}{8r_{ij}}\left(\frac{1}{a_i^2} - \frac{1}{(r_{ij}+a_j)^2}\right) + \frac{1}{2}\left(\frac{1}{a_i^3} - \frac{1}{(r_{ij}+a_j)^3}\right) - \frac{3(r_{ij}^2 - a_j^2)}{16r_{ij}}\left(\frac{1}{a_i^4} - \frac{1}{(r_{ij}+a_j)^4}\right) \tag{3.18}$$

3. Atom i is completely inside atom j, $r_{ij}^2 < (\sigma_i - \sigma_j)^2$ and $\sigma_j \geq \sigma_i$

$$Z_{ji} = \frac{1}{a_i^3} - \frac{a_j^3}{(a_j^2 - r_{ij}^2)^3} \tag{3.19}$$

4. Atom j is completely inside atom i, $r_{ij}^2 < (\sigma_i - \sigma_j)^2$ and $\sigma_j < \sigma_i$. In this case, atom j does not contribute to the descreening of atom i, and $Z_{ji} = 0$.

The sum of integrals over the van der Waals volume for each atom overestimates the total integral due to overcounting of overlapping regions of space. This effect is accounted for by scaling the volume integrals by the fractional self-volume, the portion of the volume occupied exclusively by the atom in question. The volume of a set of overlapping spheres is given by the Poincaré inclusion-exclusion principle[115]

$$V = \sum_i V_i - \sum_{j>i} V_{ij} + \sum_{k>j>i} V_{ijk} - \ldots \tag{3.20}$$

where V_i is the van der Waals volume of atom i, V_{ij} is the intersection volume of atoms i and j, and so forth. It follows that the volume belonging exclusively to atom i, the self-volume is

$$V_i^* = V_i - \frac{1}{2}\sum_j V_{ij} + \frac{1}{3}\sum_{k>j} V_{ijk} - \ldots \tag{3.21}$$

The self-volume of atom j calculated by Equation 3.21 includes *all* other atoms, and therefore, the scaling of the integral over atom j outside of atom i must be modified by adding back the overlap of atoms i and j. The scaling factor is therefore defined as

$$s_{ji} = \frac{V_j^* + \frac{1}{2}V_{ji}}{V_j} \tag{3.22}$$

This gives the fractional volume associated exclusively with atom j outside of atom i and the correct result of $s_{ji} = 1$ when no other atoms intersect atom j. The Born radius of atom i can then be evaluated by the sum of atomic integrals, each

appropriately scaled by the fractional self-volume

$$\frac{1}{B_i} = \left(\frac{1}{\sigma_i^3} - \sum_{j \neq i} s_{ji} Z_{ji} \right)^{1/3} \tag{3.23}$$

It remains only to calculate the overlap volumes for the set of atoms. Describing the molecular volume by a set of overlapping atomic Gaussian density functions provides an efficient approach to calculating these volumes[89, 115]. According to this model, the volume of each atom is described by

$$\rho_i(\mathbf{r}) = p\, e^{-c_i(\mathbf{r}-\mathbf{r}_i)^2} \tag{3.24}$$

where p and c_i are adjustable parameters. The overlap volume of n spheres is then approximated by the integral of the product of the n Gaussian functions

$$V^g_{12\ldots n} = \int d^3\mathbf{r} \rho_1(\mathbf{r}) \rho_2(\mathbf{r}) \ldots \rho_n(\mathbf{r}) \tag{3.25}$$

which can be evaluated analytically as

$$V^g_{12\ldots n} = p_{12\ldots n} e^{-K_{12\ldots n}} \left(\frac{\pi}{\Delta_{12\ldots n}} \right)^{3/2} \tag{3.26}$$

where

$$p_{12\ldots n} = p^n \tag{3.27}$$

$$K_{12\ldots n} = \frac{1}{\Delta_{12\ldots n}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} c_i c_j r_{ij}^2 \tag{3.28}$$

$$\Delta_{12\ldots n} = \sum_{i=1}^{n} c_i \tag{3.29}$$

The Gaussian parameter is defined as $c_i = \kappa/\sigma_i^2$, where κ is a dimensionless parameter which defines the diffuseness of the atomic volume. The parameters κ and p are related by the equation

$$p = \frac{4\pi}{3} \left(\frac{\kappa}{\pi} \right)^{3/2} \tag{3.30}$$

which satisfies the requirement that the integral of the atomic volume over all space produces the hard sphere volume $4\pi\sigma_i^3/3$. The original model development sets the value of $\kappa = 2.227$, and therefore $p = 2.5$[115].

This approximation works well for the overlap volumes of heavy atoms[147]. However, hydrogen atoms are deeply buried into attached heavy atoms and the volume overlap integral incurs significant error. The solution proposed by Gallicchio and Levy is the neglect of hydrogen contributions to overlap volumes[89]. Extra care must also be taken when calculating Born radii as occasional zero or negative values can be produced. The empirical solution to this problem was the application of a switching filter of the form

$$\frac{1}{B_i} = f_b(B_i) = \begin{cases} \sqrt{1/b^2 + 1/B_i^2} & B_i > 0 \\ 1/b & B_i \leq 0 \end{cases} \tag{3.31}$$

with the constant, b, set to some maximum cutoff for the Born radii. While these empirical corrections effectively displace the integral error, as well as physically unrealistic Born radii, and manage to produce negligable error in large protein solvation energy calculations, energies for small molecules can have large errors. By neglecting the contribution of hydrogens to overlap volume, yet giving them their own total van der Waals volume, the total calculated molecular volume is significantly overestimated. Disregarding certain overlap volumes also results ultimately in arbitrarily setting Born radii to some large cutoff value due to a lack of a physically meaningful radius. This may not be of much importance for an atom buried deep within a protein interior and shielded from direct interaction with the solvent, but these inaccuracies are critically important for small molecule solvation energies. For example, the calculated solvation energy of lysine by the ignored hydrogen overlap volume approximation is -118.84 kcal/mol, compared to the Poisson equation value of -149.27 kcal/mol.

The inconsistency in the treatment of hydrogens, and all atoms as it turns out, can be corrected more appropriately by normallization of the volume overlap integrals. With fixed parameters p and κ in Equation 3.24, overlap integrals do not produce the required result that two coincident spheres of equal radius have the overlap volume

of the hard sphere equivalent $4\pi\sigma^3/3$. Setting the overlap integral of n coincident identical Gaussians of radius σ equal to the hard sphere volume

$$\int p^n e^{-n\kappa(\mathbf{r}-\mathbf{r}_i)/\sigma^2} = \frac{4}{3}\pi\sigma^3 \tag{3.32}$$

and solving for κ for each n gives

$$\begin{aligned}
\kappa_2 &= \pi^{1/3}\left(\frac{48\sqrt{2}}{32}\right)^{2/3} \\
\kappa_3 &= \pi^{1/3}\left(\frac{108\sqrt{3}}{64}\right)^{2/6} \\
\kappa_4 &= \pi^{1/3}\left(\frac{216\sqrt{4}}{128}\right)^{2/9} \\
\kappa_5 &= \pi^{1/3}\left(\frac{405\sqrt{5}}{256}\right)^{2/12} \\
\kappa_6 &= \pi^{1/3}\left(\frac{729\sqrt{6}}{512}\right)^{2/15}
\end{aligned} \tag{3.33}$$

up to $n = 6$. Generally, the form of κ_n follows

$$\kappa_n = \pi^{1/3}\left(\frac{I_n\sqrt{n}}{2^{n+3}}\right)^{2/(3(n-1))} \tag{3.34}$$

where the integer I_n follows the sequence $\{48, 108, 216, 405, 729, \ldots\}$. The values of the constants κ and p of the Gaussian volume function in Equation 3.24 evaluated by Equations 3.34 and 3.30 such that the overlap volume of n atoms is normalized to the hard sphere volume are listed in Table 3.2. Interestingly, the optimized values for κ and p using a single value for all overlap volumes are 2.227 and 2.5, respectively[115]. These values fall intermediate to the normalized values for overlaps of order two and three, which would represent the largest contributions to the total overlap volume. Higher order overlaps are both less common and usually represent smaller volumes. It is therefore sensible that a single parameter optimzed for the total calculation would be intermediate to the values of overlap orders two and three. Values of κ and p were

not calculated for overlap order higher than 6, as this situation has not been observed in any real molecular system studied to date.

Table 3.2: List of Gaussian volume parameters which normalize the overlap of two to six coincident atoms to the hard sphere volume

Overlap order (n)	κ_n	p_n
2	2.418	2.828
3	2.094	2.279
4	1.919	2.000
5	1.808	1.829
6	1.730	1.712

This implementation was first tested through solvation energy calculations for a set of proteins from the Protein Data Bank (http://www.rcsb.org/pdb). The solvation energy for the proteins listed in Table 3.3 were calculated by Poisson-Boltzmann, GB-CFA, GBr^6 exluding hydrogens in the self volume calculations, and GBr^6(mod) with hydrogens included in the self volume calculation under the normalized overlap integrals. The inaccuracy of the CFA method is apparent as solvation energies are significantly underestimated and the correlation coefficient is approximately 0.9. The GBr^6 model produces a stronger correlation with a fitted slope of 0.93, and a correlation coefficient of 0.9988. The effects of neglecting hydrogens in protein solvation energy calculations are small due to the large number of hydrogens buried within the molecule having little direct interaction with the solvent. The normalized volume overlap integrals produce a slightly better fit, with a slope of 1.002 and a correlation coefficient of 0.9997. Small molecule solvation energy also correlates much more strongly. The electrostatic hydration free energy of lysine calculated with the modified self volume calculations was -147.23 kcal/mol, compared to -118.84 kcal/mol for the unmodified self-volume calculation, and -149.27 kcal/mol by the Poisson equation.

Incorporation of the solvation model into a molecular mechanics simulation requires that the model be able to differentiate not only between solvation energies for different molecules, but also between different conformations of the same molecule. This was tested for two cases. First a stringent test of 100 conformations of blocked

Table 3.3: List of Proteins used for Generalized Born Model Verification. Solvation Energies were computed by the traditional Coulomb Field Approximation and the GBr^6 model with and without inclusion of hydrogens in self-volume calculations

1AHO	1HJE	1P9G	1W6Z
1C75	1IJ4	1PQ7	1W71
1CEW	1KCH	1R01	1WNU
1EJG	1KCJ	1SSW	1WY3
1ETL	1L9L	1SUP	1YK4
1F94	1MCA	1TQG	2BF9
1G66	1NA9	1TT8	2ERL
1GQV	1OK5	1UCS	2FDN

glycine (acetyl-glycine-methyl amide) saved from a short Monte Carlo simulation. This small flexible molecule does not contain large partial charges and conformational changes should not result in large changes in solvation energy. Figure 3-7a illustrates the correlation between the GBr^6 model, with and without the inclusion of hydrogens in the self-volume calculation, to Poisson solvation energies. Neglecting hydrogens again underestimates the solvation energy, although a strong linear correlation is obtained. The normalized volume overlap integrals produce excellent correlation to Poisson results between conformations of blocked glycine over the total range of approximately 0.2 kcal/mol. In Figure 3-7b, a similar analysis of the GBr^6 model with the modified self-volume calculation is carried out for the dodecamer peptide $(GK)_6$. Again, strong linear correlation, with a slope of 1.063 and a correlation coefficient of 0.9956 is obtained.

The modified Generalized Born method GBr^6 agrees well with the more rigorous Poisson-Boltzmann model for proteins, small molecules, and differing conformations of the same molecule. By these criteria, the model can be incorporated into the molecular mechanics simulation software with some confidence. However, it remains to be tested whether accurate results can be obtained in the application to surface adsorption simulations. Comparisons of implicit solvent models to density functional theory/self consistent reaction field calculations in application to surface adsorption simulations have revealed that the models vary significantly in energy calculations[83]

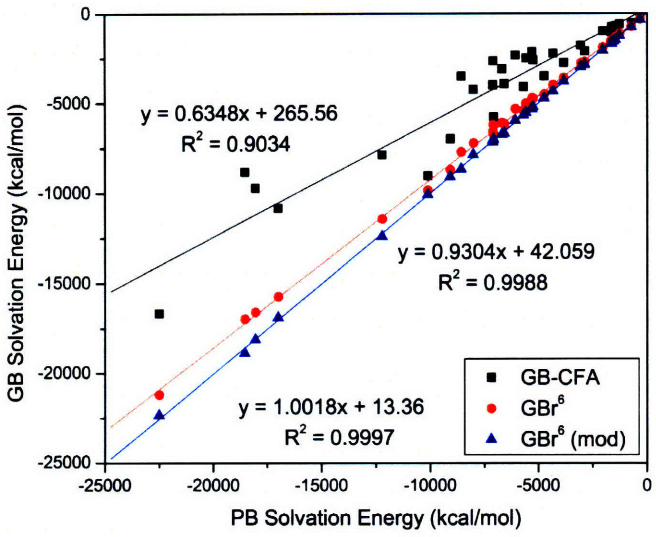

Figure 3-6: Comparison of Generalized Born with Coulomb Field Approximation (GB-CFA), modified Generalized Born GBr^6 without hydrogens in self-volume calculations (GBr^6), and GBr^6 with hydrogens included in self-volume calculations through normalized volume overlap integrals (GBr^6 mod).

although they may each produce acceptable results in isolated molecule calculations. The GBr^6 model must therefore be verified in this specific application before implementation into the surface adsorption molecular mechanics simulation.

The desolvation effects of an extended surface on charges positioned near the dielectric boundary were tested by monitoring the change in solvation energy of a single lysine residue as the surface-lysine distance was decreased. The lysine residue was positioned above a surface of uncharged atoms in the sapphire crystal structure. Surface atoms were not charged in order to isolate the effects of the dielectric boundary from any Coulombic interaction with the surface. The lysine residue was oriented with the side chain axis parallel to the surface normal, as illustrate in Figure 3-8a, in order to ensure optimal approach of the charged functional group to the surface. The traditional Coulomb Field Approximation to the Generalized Born model significantly overestimates the long range effects of the surface on the solvation energy of nearby charges. However, the GBr^6 model reproduces Poisson-Boltzmann solvation energies

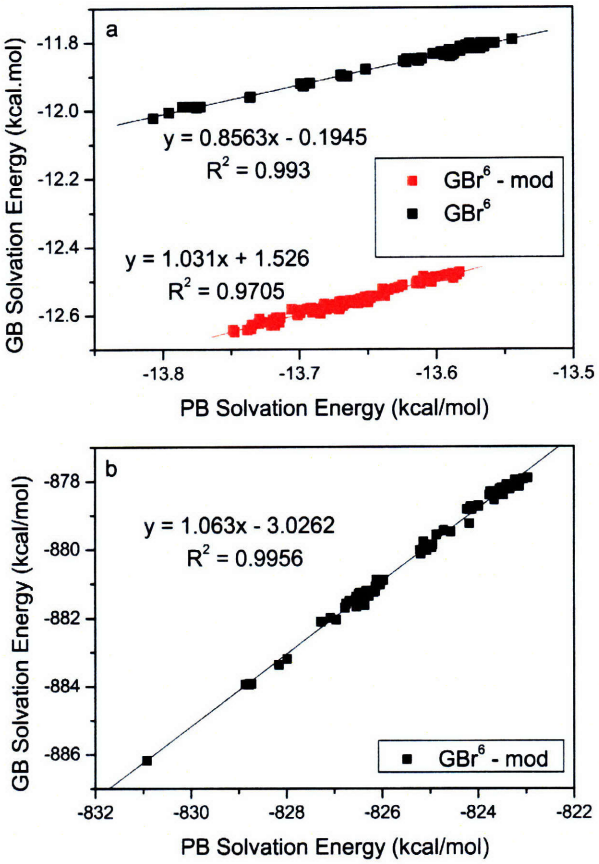

Figure 3-7: Correlation of Generalized Born to Poisson-Boltzmann solvation energies for difference conformations of the same molecule. (a) Solvation energy for 100 conformations of blocked glycine for GBr^6 with (GBr^6-mod) and without (GBr^6) inclusion of hydrogens in self-volume calculations. (b) Solvation energy for 100 conformations of the peptide $(GK)_6$ saved from a short Monte Carlo simulation.

reasonably well.

3.6 Conclusions

Due to deficiencies relating to grid stability in finite-difference Poisson-Boltzmann calculations and the resulting computational costs of limiting grid scale and hence, grid related errors, methods approximating Poisson electrostatics must be employed. Generalized Born based models offer an attractive alternative computationally due to

the pairwise sum in Equation 3.3. However, calculated solvation energies are highly dependent on accurate evaluation of effective Born radii.

This is particularly true in the application to surface adsorption simulations. A critical approximation in the derivation of the Generalized Born model relies on the spherical symmetry of the molecular system. In practical application to isolated molecular systems, this approximation often does not impact calculated energies to an unacceptable degree. However, in the presence of an extended surface, the spherical symmetry is distinctly broken. The traditional Coulomb field approximation used to calculate effective Born radii significanly overestimates the screening of a charge positioned near a dielectric boundary. In this case, the extended range of the surface effectively descreens the charge in all directions, when in reality, the screening only occurs in one direction.

By replacing the CFA with a higher order approximation to the electric field, r^{-6}, based on the Kirkwood model of electrostatics[79], more accurate solvation energies can be obtained. An analytical treatment of the integration over the molecular volume, based on separation of the volume into atomic contributions[147], has been modified by normalization of the Gaussian overlap volumes. This model was verified by comparison to more rigorous Poisson electrostatics for solvation energies of proteins, peptide conformational changes, and charges positioned near an extended atomically modeled surface.

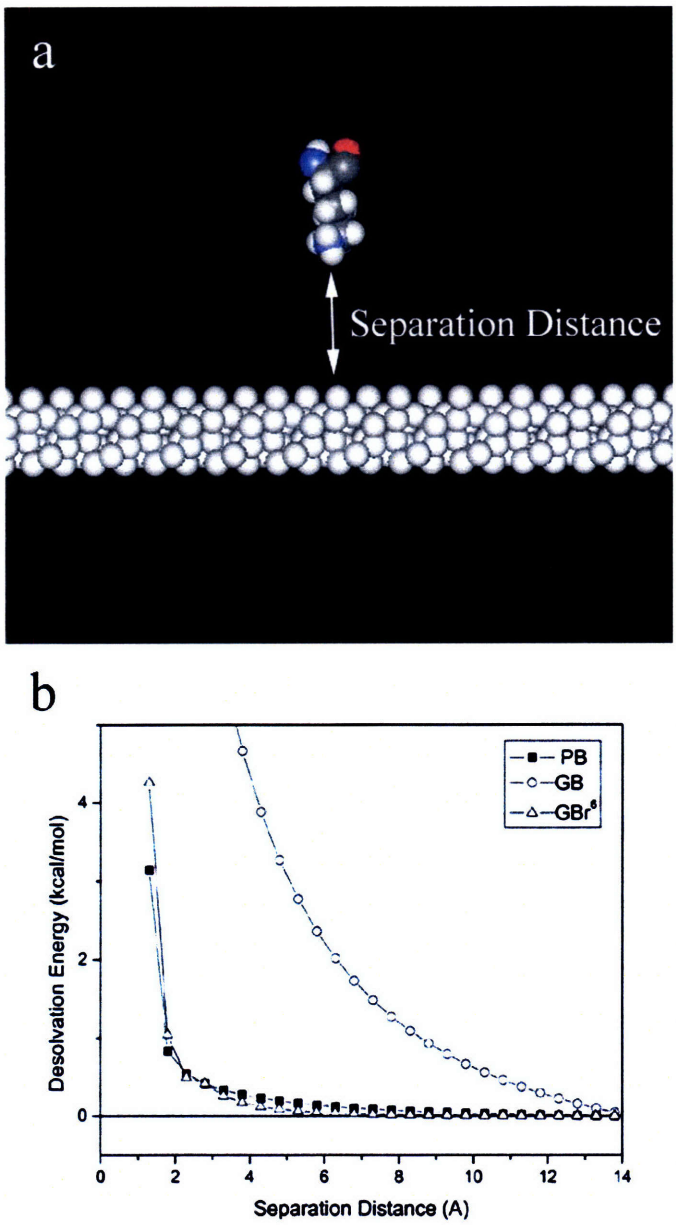

Figure 3-8: Comparison of Generalized Born-CFA, GBr^6, and Poisson-Boltzmann solvation models for the desolvation penalty of a lysine residue at an uncharged dielectric surface. (a) Position and orientation of lysine residue, and (b) the CFA model significanly overestimates the long range effects of the surface, while the GBr^6 model predicts PB solvation energies well.

Chapter 4

Simulation of Supramolecular Assemblies

4.1 Incorporation of Peptides into Virus Capsid

Experimental Motivation (adapted from [9])

The reliance of future technologies on developing scalable and economic methods for the fabrication of one-dimensional (1D) systems has spurred intense and rapid progress in the area of materials synthesis. In particular, 1D materials have been enthusiastically pursued for their applications in the study of electrical transport[31], optical phenomena[113], and as functional units in nanocircuitry[90]. Pursuit of "bottom-up" methods for the synthesis of semiconducting, metallic, and magnetic nanowires has yielded strategies including, but not limited to, vapor liquid solid (VLS), chemical, solvothermal, vapor phase, and template-directed fabrication. Although each method developed for the production of nanowires has had success in achieving high-quality materials, no distinct strategy to date has yielded monodisperse, crystalline nanowires of radically different compositions. The realization of such a system would require the combination of substrate-specific ligands with the predictability of self-assembly that is commonly found in nature. Biological systems offer a high degree of organization, efficient chemical modifications, and a wide variety

of naturally occuring self-assembly motifs.

The ability to store information about a material, including composition, phase, and crystallographic detail, within the genetic code of the M13 bacteriophage virus DNA has proven to be a viable means of synthesizing and organizing materials on the nanometer scale. The use of phage display techniques has led to the discovery of material-specific peptides that have preferrential binding[12], control over nanoparticle nucleation[8], and the ability to order on the basis of the inherent shape anisotropy of the filamentous M13 virus[13]. Because the protein sequences responsible for these attributes are gene linked and contained within the capsid of the virus, exact genetic copies of the virus scaffold are easily reproduced by infection into its bacterial host.

Screening of ZnS, CdS, FePt, CoPt systems with commercially available bacteriophage libraries (New England Biolabs) expressing either a disulfide constrained (Cys-Cys) heptapeptide or a linear dodecapeptide as a fusion to the gene product (gP) 3 protein located at the proximal tip of the virus has yielded nucleating peptides with the sequences: CNNPMHQNC (termed A7; ZnS), SLTPLTTSHLRS (termed J140; CdS), HNKHLPSTQPLA (termed FP12; FePt), and CNAGDHANC (termed CP7; CoPt)[9]. The incorporation of these peptides into the highly ordered, self-assembled capsid of the M13 bacteriophage virus provides a linear template that can simultaneously control particle phase and composition, while maintaining adaptability through genetic tuning of the basic protein building blocks.

Capsid Structural Analysis

The M13 bacteriophage is comprised of five genetically modifiable proteins, termed gP3, gP6, gP7, gP8, and gP9[144]. About 2700 copies of the gP8 protein, a 50 amino acid alpha-helical protein, form the capsid of the wild-type virus. The gP8 protein was genetically modified and expressed using a phagemid system, resulting in the fusion of the substrate specific peptides to the N terminus of the protein, which is displayed on the exterior of the assembled virus capsid. During assembly, stacking of the gP8 unit cell results in a five-fold symmetry down the length (c axis) of the virus. Figure 4-1a demonstrates the assembled bacteriophage capsid with phagemid-altered

peptide fusion proteins incorporated at 20% of all gP8 copies. Although the symmetry is apparently 10-fold in Figure 4-1b, there are in fact two fivefold symmetric unit cells rotated 36 degrees relative to one another, as well as translated along the c-axis. The phagemid system results in two distinct versions of the gene which encodes for the gP8 protein; a wild type version included in the "helper-phage" as well as an altered peptide-fusion version. The phagemid system encodes only the altered gP8, and thus an inital stock of bacteriophage is necessary to provide the genes encoding the remaining proteins. As a result, the assembled bacteriophage do not include a peptide fusion in each copy of the gP8 capsid protein, but rather some (unknown) percentage.

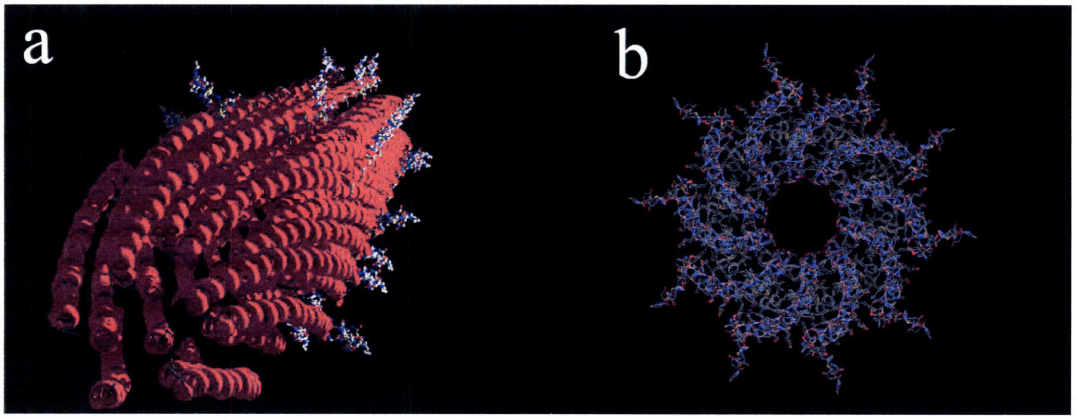

Figure 4-1: Visualization of the peptide fusion in the gP8 capsid protein of M13 bacteriophage. The virus assembly was reconstructed from the gP8 protein structure obtained from the Protein Data Bank (number 1IFJ) by application of the appropriate translation and rotation operations. A random number generator was used to realistically mimic the phagemid system and incorporate peptides at a given percentage of the total assembly.

The formation of single crystal nanowires through an annealing reaction that removes the organic virus material from the wire is facilitated by continuous coverage of the virus capsid by nucleated material. As the organic material is removed, the adjacent nanoparticles are able to fuse together into a continuous wire. With both wild-type and modified peptide-fusion gP8 proteins expressed by the host bacteria, one should not expect greater than 50% incorporation rate of the nucleating peptide. In reality, limitations of the modified gP8 proteins to assemble into the virus cap-

sid likely restrict the incorporation rate considerably. Analysis of nearest neighbor peptide separation along the virus capsid reveals that high incorporation rates are not necessary for complete mineralization of the virus to occur. The average nearest-neighbor distance between peptides decreases rapidly as incorporation is increased at low rates. However, the distance quickly stabilizes to less than 4 nm at incorporation rates above 20% (see Figure 4-2a). The density of incorporated peptides on the surface of the virus capsid increases linearly (see Figure 4-2b), as one would expect and serves to verify the randomly generated capsid assemblies.

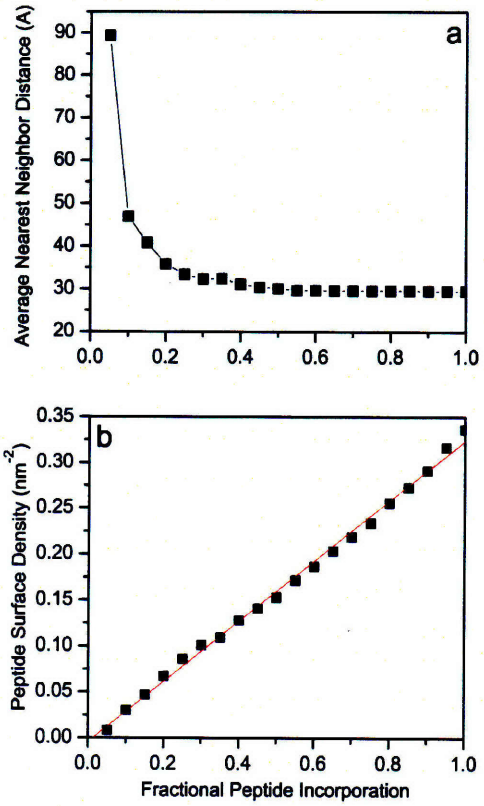

Figure 4-2: An analysis of the peptide expression on the M13 virus capsid. (a) Average nearest neighbor distances stabilize at approximately 3 nm at and above 20% incorporation. Assuming a nanoparticle size of 3-4 nm, continuous mineralization of the virus capsid can be achieved at incorporation rates much lower than 50%. (b) Density of peptides increases linearly with incorporation.

The formation of single crystalline nanowires is also facilitated by the alignment

of individual crystal orientations in the unannealed assembly. In order to explore this ordering of nucleated particles, Monte Carlo simulations of the A7 peptide were carried out in the presence of the capsid environment. A section of the virus capsid surrounding a single A7 peptide fusion was isolated from a complete capsid assembly by applying an inclusion cutoff at 30 Å from the center of the A7 conformational loop. Atoms further than the distance cutoff from the geometric center of the peptide were excluded from the simulation in order to obtain a computationally efficient simulation. The A7 peptide sits in a groove on the virus capsid created by parallel copies of the gP8 protein (see Figure 4-3). Simulations were carried out using the Monte Carlo software MCPRO[29], with implicit solvent included through the Poisson-Boltzmann model described in Chapter 2. Conformational freedom of the peptide in the capsid environment was compared to the solution phase free peptide by calculating the average standard deviation of the ensemble distributions for each of the peptide backbone dihedrals (ϕ,ψ). Transfer of the peptide from isolation to the capsid environment resulted in a decrease in conformational freedom of 21.2%. For comparison, breaking the disulfide bond in the isolated peptide increases conformational freedom by 33.7%. The restriction of conformational freedom imposed when the peptide is seated in the groove between capsid proteins is similar in magnitude to the loop-constraint in the heptamer peptide. The average standard deviation of backbone dihedrals is limited to 14.27 degrees for the capsid incorporated peptide.

The ordering of the nucleated particles with regard to preferred crystallographic orientation along the length of the virus is thus believed to be a result of the stability of the peptide fusion and the symmetry of the virus coat. This nanocrystal ordering promoted the single-crystal nature of annealed nanowires by satisfying the orientation requirements of the aggregation-based crystal growth mechanism[114]. Although particles exhibiting orientations that are not coherent with that of the majority are expected, these minority nanocrystals should rotate to adopt the preferred crystallographic orientation and merge with the majority during annealing to minimize interfacial and grain-boundary energies. Thus, the exploitation of the self-assembly motifs employed by the M13 bacteriophage to produce a biological scaffold provides a means

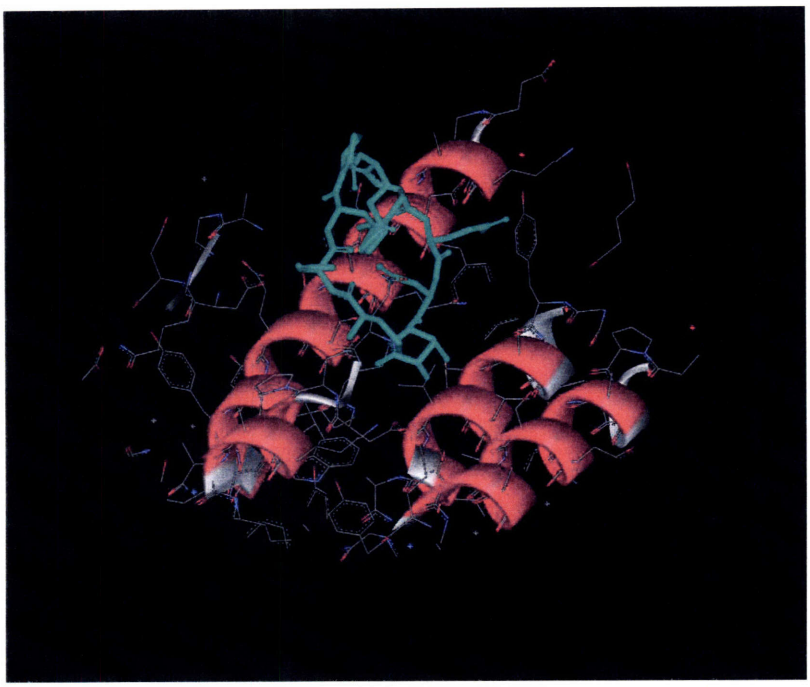

Figure 4-3: Section of the virus capsid with peptide fusion used to explore the conformational flexibility of the peptide in the capsid environment. Transfer of the peptide from solution into the capsid environment results in a 21% decrease in conformational flexibility measured by the dihedral distributions of the peptide backbone (ϕ, ψ). Peptide shown in green, gP8 proteins shown as ribbons.

of generating complex and highly ordered templates for the synthesis of single-crystal nanowires.

4.2 Mechanical Properties of Viral Assembly

Experimental Motivation (adapted from [6])

The Ff class of filamentous bacteriophage, composed of the structureally akin species f1, fd, and M13, has elicited the interest of many wide-ranging scientific communities because of its self assembling nature. Protected and transported within the highly organized, protein-based capsid is the structural and assembly information necessary for its own production. This structural feature provides a direct and accessible link between phenotype and genotype, which particularly in the case of M13 bacterio-

phage, has proven advantageous for numerous studies and applications. For instance, combinatorial libraries of polypeptides can be fused to M13 coat proteins, in a technique known as phage display, as a means of screening binding candidates against targets[144]. In addition to serving as the vehicle for displaying these ligands, the unique structure of M13 itself has been exploited as a biological template for nanotechnology, such as in the directed synthesis of semiconducting and magnetic nanowires and lithium ion battery electrodes[8, 9, 11]. Considering its utility as both a genetic blueprint and stuctural backbone for materials and device architecture, a better understanding of its mechanical behavior and a novel means of actively assembling M13 can greatly advance the design of future M13-based materials.

Heterobifunctional phages were designed by displaying hexahistidine epitopes at the remote tips and biotin molecules linked through selenocysteine residues at the proximal tips. The modified phage particles were then suspended between antibody-functionalized coverslips and streptavidin-coated polystyrene microspheres (see Figure 4-4). The polystyrene beads were trapped by the optical gradient forces of a tightly focused laser beam and positioned a set height above the coverslip surface. The piezo-electric stage was then translated laterally while bead displacements from the trap center were recorded. With the necessary calibrations, these results were then converted to force-displacement (F-x) measurements. Despite its hierarchical structure, M13 F-x behavior was reminiscent of typical worm-like-chain (WLC) biopolymer stretching.

Modeling of Mechanical Properties

The mechanical properties and fluctuations of semiflexible polymers are well described by the WLC theory[133]. Here, the configuration of a polymer is represented by a space curve of fixed, zero tension contour length, L_0, with a bending energy that is quadratic in the chain curvature. External forces stretching WLC polymers, therefore, do work against the conformational entropy of the chain. With space curve, $\mathbf{r}(s)$, parameterized by the polymer's arc length s, the chain's curvature is simply $\kappa = |\partial^2\mathbf{r}(s)/\partial s^2| = |\partial\mathbf{t}(s)/\partial s|$, where $\mathbf{t}(s)$ is the unit vector tangent to the chain. The

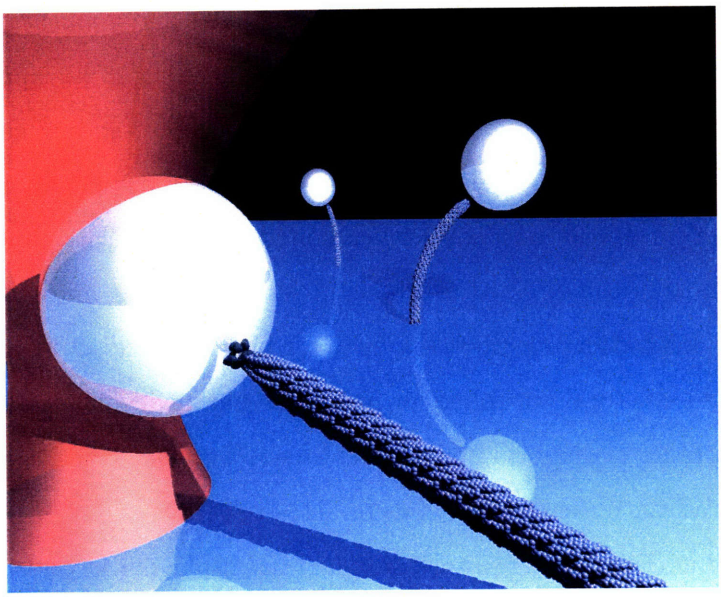

Figure 4-4: Experimental measurement of the mechanical properties of M13 filamentous bacteriophage. Heterobifunctional phage were designed by displaying hexahistidine epitomes at the remote tip and biotin molecules at the proximal tip. These modified phage particles were then suspended between antibody-functionalized coverslips and streptavidin-coated polystyrene beads for laser trapping experiments

resulting elastic energy, E, of a WLC polymer being mechanically stretched by a uniaxial force is

$$\frac{E}{k_B T} = \int_0^L \frac{l_p}{2}\kappa^2 ds - \frac{F}{k_B T}x \tag{4.1}$$

where x is the total extension of the chain, l_p the persistence length, k_B the Boltzmann constant, T the absolute temperature, and F the force. The persistence length is the characteristic scale over which thermal fluctuations begin to dominate the orientation of the chain's tangent vectors, is material specific, and is independent of the contour length, L_0.

One can imagine the experimental realization of single molecule force-extension experiments taking two distinct forms[20]. First, the isometric experiment, in which all points of the force-extension curve are characterized by the end-to-end separation distance of the molecule being exactly constant. In this case, the location of the trap center is adjusted by a feedback loop so as to cancel all fluctuations of the bead position, modulating the force to maintain constant extension. At each distinct dis-

tance, x, the force, F, must be averaged for some appropriate period of time, resulting in the mean force as a function of extension, $\langle F \rangle(x)$. Alternatively, the isotensional experiment holds the force fixed while the extension is allowed to fluctuate. In this experiment, the trap center is adjusted in order to keep the bead position fixed at some distance from the trap center, and thus the force constant. Fluctuations in extension are then averaged to yield a function of the applied force, $\langle x \rangle(F)$. Inverting the experimental results for this second experiment results in the function $F(\langle x \rangle)$, closely related to the isometric function $\langle F \rangle(x)$. These two alternative experiments represent the application of the Gibbs (isometric) and Helmholtz (isotensional) ensembles in evaluation of thermodynamic properties[125].

In the isometric experiment, the work performed on the molecule during extension from 0 to x, often called the potential of mean force, $V(x)$, is given by

$$V(x) = \int_0^x \langle F \rangle(x')dx' \tag{4.2}$$

In the presence of a fixed, external force, the isotensional energy of the molecule is given by

$$W(x, F) = V(x) - Fx \tag{4.3}$$

where $V(x)$ is the isometric potential of mean force, and Fx is the work done by the external force. The measured mean extension is then given statistically as

$$\langle x \rangle(F) = \Xi(F)^{-1} \int_0^\infty x e^{-(V(x)-Fx)/k_BT} dx \tag{4.4}$$

where Ξ is the partition function, $\Xi = \int_0^\infty e^{-(V(x)-Fx)/k_BT} dx$. The free energy, $U(F)$, relative to the isometric experiment can then be defined as

$$\begin{aligned} U(F) &= -k_BT \ln \frac{\Xi(F)}{q} \\ &= \int_0^\infty e^{-(V(x)-Fx)/k_BT} dx \Big/ \int_0^\infty e^{-V(x)/k_BT} dx = \langle e^{Fx/k_BT} \rangle \end{aligned} \tag{4.5}$$

From Equations 4.4 and 4.5, we can then relate the free energy to the measured mean

extension as,

$$U(F) = -\int_0^F \langle x \rangle(F')dF' \tag{4.6}$$

The Gibbs free energy, $V(x)$, and Helmholtz free energy, $U(F)$, are thus related by a Laplace transform analagously to the relationship between canonical and grand canonical ensemble partition functions, expressed as

$$e^{-U(F)/k_BT} = \int_0^\infty e^{-V(x)/k_BT} e^{Fx/k_BT} dx \tag{4.7}$$

In general, the WLC model is difficult to solve for arbitrary boundary conditions. However, an analytical solution is available for the equilibrium extension of polymers with contour length on the order of, or shorter than, the persistence length ($L_0 \leq 2l_p$). In this fluctuating rod limit, the rod tangent vectors make only small deviations away from the direction of the force and a harmonic approximation can be taken and the generating functional method used to obtain the average extension[81]. This solution was modified to include a stretching term that allows the modeling of the full range of bacteriophage extensions. An effective stretching energy that is quadratic in the polymer's elongation, $E_e = \int_0^L 1/2K(s/s_0 - 1)^2 ds$, was added to Equation 4.1. In the case of small elongations, the resulting average extension is

$$\langle x \rangle(F) = L_0 - \frac{k_BT}{2F}\left[L_0\sqrt{\frac{F}{A}}\coth\left(L_0\sqrt{\frac{F}{A}}\right) - 1\right] + \frac{FL_0}{K} \tag{4.8}$$

where $A = l_p k_B T$ and K is an elastic stretching modulus[20]. Here the end tangent vectors are assumed to be collinear with the force, consistent with the experimental setup, where linkages were engineered from the proximal and remote tips (i.e., from small, pivoting proteins as opposed to the crystalline gP8 capsid).

However, it is not clear whether the experimental setup is more accurately modeled as an isometric or isotensional experiment, and thus whether the experimental data should be fit to $F(\langle x \rangle)$ or $\langle F \rangle(x)$. In the thermodynamic limit, $L_0 \gg l_p$, these two quantities are equal, but in general will differ to some extent. Thus we wish to predict the magnitude of this difference in order to determine whether it can be exper-

imentally resolved, and further, which model represents the appropriate relationship for the experimental setup. To this end, the isotensional free energy was evaluated by Equation 4.6 for the mean extension given by Equation 4.8. The inverse Laplace transform, Equation 4.7, was then applied numerically to give the potential of mean force, Equation 4.2. Finally, the derivative of the potential of mean force gives the isometric force-extension curve. The isometric and isotensional force-extension curves are plotted for $L_0/l_p = 0.5$, 1.0, and 5.0 in Figure 4-5. Experimentally measured values of L_0 and l_p for the M13 bacteriophage from single molecule stretching experiments were 939.7 $\pm$ 46.1 nm and 1,265.7 $\pm$ 220.4 nm. This L_0/l_p ratio is intermediate to the 0.5 and 1.0 plots in Figure 4-5.

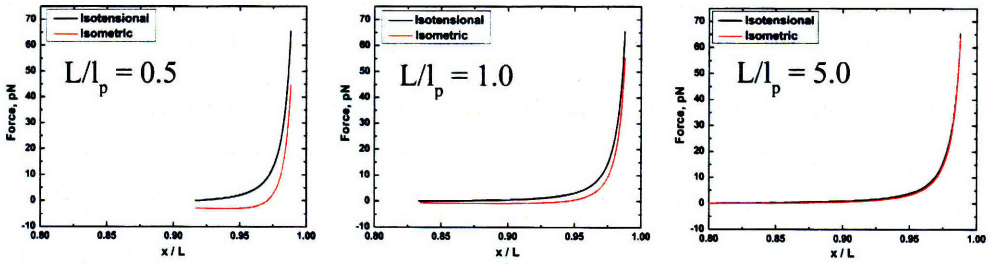

Figure 4-5: Isometric and isotensional force-extension curves for WLC biopolymer stretching experiments. In the non-thermodynamic limit, ($L_0 \gg l_p$), the ensembles result in different force-extension curves.

The actual experimental setup was designed to reproduce neither isometric, nor isotensional results. Rather, the experimental setup is likely intermediate to the two extreme cases. However, the difference between the predicted isometric and isotensional force-extension curves is not substantial enough to give a clear indication as to the nature of the experimental setup. In fact, the differnce between the predicted force-extension curves is not likely experimentally resolvable. Thus, the isotensional mean extension model, Equation 4.8 can be used with confidence to fit data from molecular stretching experiments.

Chapter 5

Sapphire-Binding Peptides

5.1 Introduction

Recently, a set of dodecamer peptides was identified froma yeast surface display library[10] with binding affinity for sapphire (α-Al_2O_3). These peptides were shown to interact with the sapphire surface through multiple basic amino acids. Further investigation of these interactions was carried out by the construction of designer peptides, each of identical composition but differing in sequence. Three peptides, termed K1, K2, and K3, each composed of six glycine and six lysine residues varying in sequence order were used to demonstrate the importance of peptide sequence in binding affinity. Binding assays of two additional peptides, cK1, and cysteine constrained circular version of K1, and K1P, a version of K1 with three glycine residues replaced by proline, demonstrated the importance of conformational flexibility in the adsorption process. Molecular simulations of these peptides reveal the basis for sequence and structural dependence of binding affinities.

Material Selection

Metal oxides have found increasing technological applications in sensitive gas sensors[28] and promising new biosensors[92]. Materials such as alumina (Al_2O_3) and silica (SiO_2) are often used as substrates for biological assays due to their compatibility with aqueous environments and lack of cytotoxicity. Single crystal alumina, or syn-

thetic sapphire (α-Al_2O_3), is commonly used as a substrate for the epitaxial growth of semiconductors[68, 94], and is widely available commercially. This material is often used as a model metal oxide in the study of environmental adsorbents[1, 17], posessing excellent chemical resistence and durability. Understanding the mechanism of amino acid and peptide adsorption at this model metal oxide surface could facilitate the development of many novel biological applications.

Experimental Identification of Peptides

Yeast surface display libraries[7] were used for biopanning experiments against three synthetic sapphire crystal faces (C, A, and R)[10]. Peptides were selected from a library of approximately 10^7 unique sequences. Although a concensus binding motif was not identified from this selection process, comparison of sequence composition for selected peptides against the naïve library gives some insight into the adsorption process. Basic amino acids were over-represented in selected peptides, populating approximately 40% of the peptide compared to 10% of the naïve library. Acidic amino acids were under-represented, populating 2% of the peptide, compared to 7% in the naïve library. Finally, hydrophobic residues were also under-represented, populating 15% of the peptide compared to 40% in the naïve library. This compositional analysis reveals the importance of highly positively charged peptides in adsorption to sapphire surfaces, with most of the selected peptides carrying a charge of +4 to +6, but is not particularly informative regarding peptide sequence and structure dependence.

Rational Design of High Affinity Peptides

With the importance of basic amino acids established, the role of spacing of the charges was investigated through the design and cloning of a set of identically composed peptides based on lysine-glycine repeat units[10]. These peptides are listed in Table 5.1 and form the basis of the computational simulations. A simple set of peptides (K1, K2, and K3) efficiently explore the role of charge grouping in +6 charged peptides. The peptide R1 explores the specific dependence on amino acid, while the peptides, cK1 and K1P, explore the effects of structural limitations in otherwise

identically composed peptides.

Experimental Interrogation of Peptides

Peptides were interrogated experimentally through both yeast surface display and ELISA assays of peptide-protein fusions. Yeast surface display binding was characterized optically by examining the crystalline surface and observing the cell binding as the percent area coverage (PAC) of yeast cells. This was measured as the ratio of cell area to total image area using image analysis software. Modified ELISA experiments provide a slightly more quantitative measurement spectroscopically, and are described in detail below.

5.2 Molecular Simulation of Peptides

5.2.1 Simulation Details

Thermodynamically favorable peptide adsorption occurs when the change in Gibbs free energy, ΔG_{ads}, of the system is negative for the adsorption process[57]. Therefore, computational prediction of peptide-surface binding requires the calculation of this change in free energy,

$$\Delta G_{ads} = \Delta H_{ads} - T\Delta S_{ads} \tag{5.1}$$

where ΔH_{ads} is the change in enthalpy, ΔS_{ads} is the change in entropy, and T is the absolute temperature. Within the model used here, the enthalpy and entropy of adsorption each can be considered to be composed of two separable components: the peptide-surface (P-S) internal and interaction contributions ($\Delta H_{P-S}, \Delta S_{P-S}$) and the change in interaction with the surrounding solvent environment ($\Delta H_{water}, \Delta S_{water}$), which includes solvent reorganization.

$$\Delta H_{ads} = \Delta H_{P-S} + \Delta H_{water} \tag{5.2}$$

$$\Delta S_{ads} = \Delta S_{P-S} + \Delta S_{water} \tag{5.3}$$

Peptide internal energy and peptide-surface interaction energy were calculated using molecular mechanics with the OPLS-AA force-field in the program MCPRO[29], capturing the enthalpic contribution to the peptide-surface interaction. Peptide degrees of freedom are sampled through Monte Carlo molecular mechanics following the Metropolis algorithm[138], which aims to reproduce a Boltzmann weighting of sampled conformational states. Generally, the Monte Carlo algorithm as applied to molecular simulations relies on a computer's pseudo-random number generator to produce a Markovian chain of configurational states[54]. The necessary requirement that the limiting distribution is reached can be enforced through the requirement of detailed balance

$$P(x)T(x \rightarrow y) = P(y)T(y \rightarrow x) \tag{5.4}$$

where $T(x \rightarrow y)$ is the transition probability of reaching state y from state x, and $P(x)$ is the probability of realizing state x in the final distribution. The requirement of detailed balance defines the ratio of transition probabilities for a desired Boltzmann distribution non-uniquely as

$$\frac{T(x \rightarrow y)}{T(y \rightarrow x)} = e^{-\beta(E_y - E_x)} \tag{5.5}$$

Finally, accepting all moves to lower energy states, one arrives at the acceptance criteria which drives the Metropolis Monte Carlo molecular simulation

$$\mathrm{acc}(x \rightarrow y) = \min[1, e^{-\beta \Delta E(x \rightarrow y)}] \tag{5.6}$$

OPLS-AA Force-Field

The OPLS force-field[139, 140] is a molecular mechanics force-field developed with a simple, efficient computational form and optimized to directly reproduce experimental thermodynamic and structural data on fluids. Therefore, this force-field represents an ideal parameter set for the solution phase simulation of peptides. Investigation of the propensity of differing force-fields to form secondary structures in short peptides has shown that the OPLS-AA force-field produces good agreement with experimental

results[111]. The force-field consists of an all-atom molecular representation, where the energy is calculated as

$$E = E_{bnd} + E_{ang} + E_{dih} + E_{nb} \tag{5.7}$$

where E_{bnd} and E_{ang} are spring-like bond stretching and angle bending energies

$$E_{bnd} = \sum_{bonds} K_r(r - r_{eq})^2 \tag{5.8}$$

$$E_{ang} = \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 \tag{5.9}$$

where K_r and K_Θ are atom-type specific parameters and r_{eq} and Θ_{eq} are the experimentally observed equilibrium bond lenths and angles. The torsional energy component is evaluated by the Fourier series

$$E_{dih} = \sum_{dihedrals} \frac{V_1}{2}[1 + \cos(\phi)] + \frac{V_2}{2}[1 - \cos(2\phi)] + \frac{V_3}{2}[1 + \cos(3\phi)] \tag{5.10}$$

where V_1, V_2, V_3 are atom-type specific parameters and ϕ is the dihedral angle. Finally, the non-bonded energy is evaluated by a pairwise sum over Coulombic interactions between charged atoms and Lennard-Jones interactions

$$E_{nb} = \sum_i \sum_j \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] f_{ij} \tag{5.11}$$

where q_i is the charge on atom i, and standard mixing rules are used such that $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$, and σ_i, ϵ_i are the atom-type specific Lennard-Jones parameters. The function f_{ij} is used to correct the non-bonded interactions in bonded and angle or dihedral connected atoms, set as $f_{ij} = 1.0$ in general, but $f_{ij} = 0.5$ for atoms separated by three or fewer bonds.

Solvation Energy

As peptide adsorption occurs, water molecules are displaced from the region between the peptide and surface into bulk solution, reducing the solvation of exposed molecular surfaces. While explicit inclusion of water molecules would ostensibly be the most accurate and detailed representation, the addition of explicit water greatly increases the number of atoms and degrees of freedom to be sampled. For example, adding explicit water molecules to the system composed of a dodecamer peptide and six nanometer sapphire surface results in approximately 45,000 water molecules. The computational resources necessary to equilibrate and achieve thorough averaging of such as system are not currently available. Alternatively, implicit solvent models have proven to effectivly reproduce solvation effects in a number of systems, as discussed in Chapters 2 and 3. These models replace the numerous solvent molecules with a continuum dielectric, plus interfacial terms, and seek to capture the effects of the statistical solvent environment. Hydration energies calculated by continuum methods therefore contain both enthalpic and entropic contributions. These hydration free energy models are often decomposed into electrostatic (elec) contributions resulting from the polarization of the solvent by solute charges, and non-polar (np) interfacial contributions resulting from the changes in contact of water with the solute surface[109]. Electrostatic solvation energies were calculated by the modified Generalized Born method discussed in Chapter 3.

In additions to the description in Chapter 3, the solvation energy was refined by scaling[148] with the function

$$f(\epsilon_{in}, \epsilon_{ex}) = \frac{A + 2B\epsilon_{in}/\epsilon_{ex}}{1 + 2\epsilon_{in}/\epsilon_{ex}} \tag{5.12}$$

where $(A = -1.63 \times 10^{-3}|Q|^{0.65} + 2.18 \times 10^{-6}N_{atom} + 1.016)$, and $(B = 3.31 \times 10^{-2}|Q|^{0.65} - 4.77 \times 10^{-5}N_{atom} + 0.683)$, Q is the net charge of the molecule, and N_{atom} is the number of atoms in the molecule. This formula is an empirical fit to the observation that solvation energy calculations do not scale generally for all combinations of $\epsilon_{in}, \epsilon_{ex}$, and improves the accuracy of the energy calculations.

It is attractive and common to include the total electrostatic free energy from an implicit solvent calculation in the molecular mechanics force-field, as this eliminates the need to perform a reference calculation of the Coulombic electrostatic energy in gas phase. However, in the case of the OPLS force-field, we must be careful due to the scaling factor included in Equation 5.11. In addition to careful attention to the non-bonded scaling factor, particular attention must be paid to the choice of molecular dielectric constant. Although the common agreement of a physically realistic molecular dielectric constant calls for a relative value of two, the OPLS-AA force-field was developed with the assumption of a molecular relative dielectric of one. The effects of the increased polarization implicit with a dielectric of two are instead included in the parameterization of Lennard-Jones factors, and the incorporated implicit solvent model should remain consistent with this parameterization.

In order to verify the accuracy of an included implicit solvent model in MCPRO, Monte Carlo simulations were carried out on a dichloroethane-like hypothetical molecule in explicit TIP4P water, as well as with implicit solvent[21] (see Figure 5-1). The implicit solvent model was included with molecular dielectric constants of one and two, and with and without the scaling function f_{ij} in Equation 5.11 applied to the Coulombic interactions. Explicit solvent simulations were equilibrated for two million, and averaged over ten million Monte Carlo steps. All simulations were carried out under the NPT ensemble at 298 K. Implicit solvent simulations were equilibrated for one hundred thousand, and averaged for two million Monte Carlo steps. Implicit solvent simulations generally require far fewer steps due to the lack of solvent equilibration time, as well as the lack of viscosity effects in sampling the molecular configuration space. A molecular dielectric constant of one, and consistent inclusion of the OPLS scaling function in Coulombic energy calculations results in excellent agreement with explicit water simulations.

Non-polar hydration free energies are often modeled as the product of the molecular surface area and a phenomenological surface tension. However, it has been observed that more accurate correlation to experimental results can be obtained by decomposition of the non-polar hydration free energy into cavity formation and van

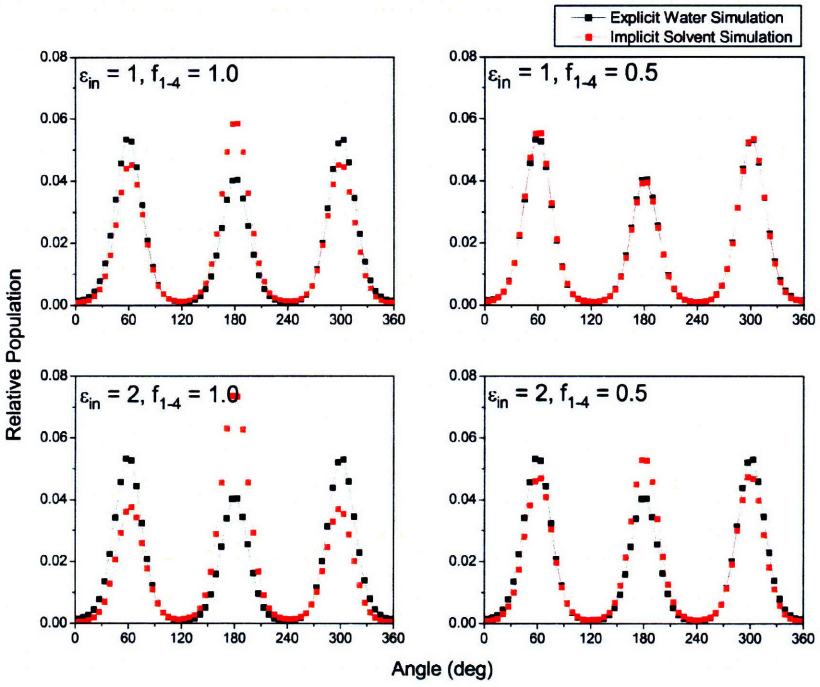

Figure 5-1: Comparison of explicit and implicit solvent in OPLS-AA Monte Carlo simulation of a dichloroethane-like hypothetical molecule. As expected, the best match to explicit water simulation is achieved by including the OPLS scaling function in the Coulombic interaction energy and with a molecular dielectric constant of $\epsilon_{in} = 1$.

der Waals interaction terms[39, 89]

$$\Delta G_{np} = \Delta G_{cav} + \Delta G_{vdW} \tag{5.13}$$

The cavity formation energy represents displacement of water molecules from the molecular volume and the accompanying reorganization, and is calculated by the surface area model,

$$\Delta G_{cav} = \sum_i \gamma_i A_i \tag{5.14}$$

where γ_i can be specific to each atom type, but in the current efforts is set as $\gamma_i = \gamma = 72$ cal mol^{-1} Å^{-2}[88], a value somewhat larger than other reported implementations[19, 56]. However, these implementations are taken to include the

van der Waals contribution to the non-polar solvation energy, which is generally favorable. A_i is the exposed surface area of each atom i. Exposed atomic surface areas can be calculated efficiently as the derivative of the atomic volume with atomic radius by employing the Gaussian molecular volume functions described in Chapters 2 and 3.

$$A_i = \frac{\partial V}{\partial R_i} = 4\pi R_i^2 - \sum_j \frac{\partial V_{ij}}{\partial R_i} + \sum_{j<k} \frac{\partial V_{ijk}}{\partial R_i} + \cdots \tag{5.15}$$

where

$$\frac{\partial V_{12\ldots n}}{\partial R_i} = \frac{2\kappa}{R_i^3}\left(\frac{3}{2\Delta_{12\ldots n}} + |r_i - r_{12\ldots n}|^2\right) V_{12\ldots n} \tag{5.16}$$

$$\Delta_{12\ldots n} = \sum_{i=1}^{n} \frac{\kappa}{R_i^2} \tag{5.17}$$

$$r_{12\ldots n} = \frac{1}{\Delta_{12\ldots n}} \sum_{j=1}^{n} \frac{\kappa}{R_i^2} r_i \tag{5.18}$$

The van der Waals energy term attmpts of capture the average interaction with all the surrounding water molecules, a generally favorable energetic contribution. Here, the Born radii, B_i, calculated for the electrostatic contribution to the solvation energy are utilized again. The van der Waals energy, in the absense of critical overlaps, is dominated by the same r^{-6} functional form used in Born radii evaluation. The van der Waals energy is decomposed into contributions from each atom of the solute, and can be approximated by integrating the attractive portion of the Lennard-Jones potential over the solvent volume. Assuming a constant solvent density of $\rho_w = 0.33428$ Å^{-3}, the van der Waals energy is

$$U_{vdW}(i) \approx -4\epsilon_{iw}\sigma_{iw}^6 \rho_w \int_{solvent} d^3r \frac{1}{(r - r_i)^6} \tag{5.19}$$

and it follows that,

$$\Delta G_{vdW} \approx \sum_i \alpha_i \frac{a_i}{(B_i + R_w)^3} \tag{5.20}$$

where α_i is an adjustable dimensionless, atom-type specific parameter on the order

of 1, R_w is the radius of water, here set to 1.4 Å, and

$$a_i = -\frac{16}{3}\pi\rho_w\epsilon_{iw}\sigma_{iw}^6 \tag{5.21}$$

where $\sigma_{iw} = \sqrt{\sigma_i\sigma_w}$ and $\epsilon_{iw} = \sqrt{\epsilon_i\epsilon_w}$ are the OPLS-AA force-field Lennard-Jones interaction parameters for atom i with the oxygen of TIP4P water ($\sigma_w = 3.15365$ Å, $\epsilon_w = 0.155$ kcal/mol).

Lekner Summation

In order to accurately represent the crystal surface as "infinite" in comparison to the peptide, the simulation cell was modeled as a slab geometry. Periodic boundary conditions were applied in the dimensions parallel to the crystal surface, while the dimension parallel to the surface normal was considered to be of finite size. Coulombic interactions in a periodic system are slowly converging and are often described by a decomposition into multiple, quickly converging sums. Perhaps the most popular example of this is the Ewald sum[38] which describes a three dimensional periodic system, but has been extended to two dimensional cases[117]. However, the two dimensional Ewald sum is not particularly fast to compute and other methods have been developed for simulating periodic conditions in one[58] and two[85] dimensional systems, although the error in such methods is not always optimal[33, 134]. The Lekner summation method[85, 86] is particularly effective in both its efficiency and accuracy. This method has been extended to arbitrary two dimensional systems[95, 143] and applied to molecular simulations of protein-membrane binding[142]. The electrostatic interaction energy (excluding the factor $1/4\pi\epsilon_0$) of a charge, q_i, with a second charge, q_j, and all periodic images of q_j in the x-y plane is given in the Lekner

sum form by U_{ij}^{lek} as,

$$\begin{aligned} U_{ij}^{lek} &= 4\frac{q_i q_j}{L_x}\sum_{n=1}^{\infty}\cos(2\pi\frac{\Delta x}{L_x}n) \\ &\times \sum_{k=-\infty}^{\infty} K_0(2\pi n[f^2(\frac{\Delta y}{L_y}+k)^2+(\frac{\Delta z}{L_x})^2]^{1/2}) \\ &- \frac{q_i q_j}{L_x}\ln(\cosh(2\pi\frac{\Delta z}{L_y})-\cos(2\pi\frac{\Delta y}{L_y}))-\frac{q_i q_j}{L_x}\ln 2 \end{aligned} \tag{5.22}$$

where $f = L_y/L_x$ and $\Delta x = x_i - x_j$, $\Delta y = y_i - y_j$, $\Delta z = z_i - z_j$, L_x and L_y are the repeat lengths in the x and y dimensions, and K_0 is the modified Bessel function

$$\begin{aligned} K_0(\alpha) &= \frac{1}{2}\int_{-\infty}^{\infty} e^{-\alpha\cosh(t)}dt \\ &\approx \sqrt{\frac{\pi}{2\alpha}}e^{-\alpha} \end{aligned} \tag{5.23}$$

At large z-separations the Lekner summation has the desired consequence in reducing to the interaction energy of a charge, q_i, with an infinite flat surface with a charge density of $\sigma = q_j/L_xL_y$

$$\frac{1}{4\pi\epsilon_0}U_{ij}^{lek} \longrightarrow -q_i\frac{\sigma}{2\epsilon_0}d \quad , \quad d \longrightarrow \infty \tag{5.24}$$

The Coulombic peptide-surface interaction energy is evaluated as the sum of the Lekner interaction energy for each peptide atom i and each crystal atom j. Since no peptide atoms are considered bonded to the crystal surface, the electrostatic energy can be evaluated directly without concern for the scaling factor in Equation 5.11. Crystal atoms are fixed throughout the simulation, and therefore crystal surface internal Coulombic energy is constant and plays no role here. Single peptides were considered in this work, and there are no periodic images of the peptides.

The Lekner sum is also utilized in the calculation of the GB electrostatic component of the solvation energy. The distance dependent function, f_{ij}, in Equation 3.4 reduces to the inter-atomic separation at large distances. If the repeat lengths L_x and L_y are large, for only one instance of each periodic surface atom will the function

f_{ij} differ significantly from the separation distance. Hence, the Lekner summation was modified by replacing this nearest surface interaction with the GB interaction to correctly account for the effect of the quasi-infinite surface on the peptide solvation energy. Thus, we obtain

$$\Delta G_{elec}^{GB} = -\frac{1}{4\pi\epsilon_0}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{ex}}\right)\left[\sum_{j>i}\left(U_{ij}^{lek} - \frac{q_i q_j}{r'_{ij}} + \frac{q_i q_j}{f'_{ij}}\right) + \sum_i \frac{q_i^2}{2f_{ii}}\right] \tag{5.25}$$

where r'_{ij} and f'_{ij} are the separation distance and distance dependent function, Equation 3.4, for atom i and the nearest instance of atom j. The crystal atom Born radii were only considered to vary in the central (nearest neighbor) simulation cell, and pariwise contributions to the surface GB electrostatic solvation energy were only considered with the nearest instance of each pair. Interactions of surface atoms with their own periodic images again fall under the approximation that $f_{ij} \approx r_{ij}$ and these contributions to solvation free energy were therefore constant throughout the simulation.

For comparative calculations, finite-difference Poisson-Boltzmann implicit solvent calculations were also incorporated into MCPRO. The PB calculations were carried out using the methods described in Chapter 2 with a grid spacing of 0.3 Å, interior and exterior relative dielectric constants of 1 and 80, respectively, and zero ionic strength.

Expanded Ensemble Simulations

A necessary condition for efficient averaging in Monte Carlo molecular simulations is that of ergodicity[118]. In systems characterized by local energy minima separated by large potential energy barriers, Monte Carlo simulations can become frustrated, or trapped in a local energy minimum. Average properties are then invalidated by the lack of proper sampling of equilibrium populated states. For the simulation of peptides bound tightly to inorganic surfaces, we expect precisely this situation. The large binding energies are likely to prevent translational, rotational and many conformational changes of surface bound peptides. This inefficiency can be circumvented

by a variety of expanded ensemble sampling methods. These methods generally alter the standard Metropolis Monte Carlo method in a manner which helps trapped systems escape the local energy minimum. Popular alternatives include entropic sampling[27, 107] and simulated tempering[15] methods. These methods, however, have considerable computational expense as entropic sampling requires initial simulations to determine the entropy landscape as a function of conformational space and simulated tempering methods require re-equilibration following each change in temperature, during which averaging cannot be conducted.

An efficient alternative sampling method can be achieved through replica exchange simulations[80], also known as parallel tempering. In this method, several copies of the system of interest are sampled independently at differing temperatures. Occasionally, the current configuration of a pair of adjacent temperature systems are exchanged. The rigorous acceptance criteria for the exchange move is developed analagously to the Metropolis algorithm, and is given by

$$\mathrm{acc}(x_\beta y_{\beta'} \rightarrow y_\beta x_{\beta'}) = \min[1, e^{\Delta\beta\Delta E}] \tag{5.26}$$

where β is the reciprical temperature, $1/k_BT$, and E is the energy. The replica exchange method provides efficient sampling of rough energy landscapes. The high temperature replicas escape local energy wells, while the low temperature replicas efficiently explore the well minima. At the same time, the ensemble of configurations at each temperature represents equilibrium throughout the simulation and therefore extra computational resources are not required for expensive re-equilibrations. Also, unlike some other alternative sampling methods, replica exchage *rigorously* satisfies the condition of detailed balance and therefore gaurantees eventual convergence to the desired distribution. No temperature dependence was included in our solvation parameters, and thus only the ensemble at 298.15 K is of final interest.

The set of temperatures to be used must still be determined. The set of temperatures should be chosen such that the exchange rate is optimized. Temperatures too close together results in a high exchange rate, but little tempering effect, while temper-

atures too far apart results in good tempering, but low exchange rates. Generally, the temperature set depends on the accessable conformational space at each temperature and follows an exponential pattern[32]. Simulations of dodecamer peptides were optimized to produce temperature exchange rates of approximately 40% in an eight replica simulation, resulting in a temperature set of 298, 323, 350, 379, 410, 444, 481, 521 K. The rate of *attempted* moves also must be optimizzed. Too short of a time between attempted exchange moves, and the high temperature system is unable to significantly move away from the low energy well. Too long between exchange attempts and the tempering effects are diminished. In the current application, exchange moves were attempted between two randomly chosen replicas every 100 Monte Carlo steps. In the eight temperature setup, this averages an exchange attempt for temperatures one and eight every 700 steps, and temperatures 2-7 every 350 steps since these temperatures can exchange conformations with both higher and lower temperatures. At a 40% acceptance rate for exchange moves, temperatures one and eight are able to exchange conformations every 1750 steps on average, while temperatures 2-7 are involved in an accepted exchange every 875 steps on average. Replica exchange was implemented in MCPRO through the Message Passing Interface (MPI) and executed on an eight processor computational cluster.

Peptides were constructed using the *pepz*[29] program included in the MCPRO distribution, which builds peptides in a fully extended conformation ($\phi = \psi = 180°$). Each peptide was capped with acetyl and methyl amide groups at the N and C termini, respectively, to ensure only amino acid side chains interact with the crystal surface. Each simulation consisted of energy minimization in the gas phase, followed by equilibration for 100,000 MC steps, at which point stable energies were confirmed. Properties were then averaged for 10^6 MC steps.

Crystal Surface Construction

Atomic coordinates for the R-face $(1\bar{1}02)$ of crystalline α-Al_2O_3 were generated for simulation of peptide adsorption. Bulk stoichiometric termination of the R-face displays a surface of mixed aluminum and oxygen composition. However, diffraction

studies[120, 126] of hydrated alumina surfaces suggest significant relaxations in the surface layers. The hydrated R-face is characterized by a relaxed surface with zero occupancy for the first layer of aluminum atoms, and significant rearrangement of atom layers near the surface (cf. Figure 5-2 and Figure 5-3). Rearrangement of the surface layers results in a negatively charged surface, observed experimentally[35]. The relaxed surface coordinates were used to generate an extended crystalline surface eight atom layers thick and extending approximately six nanometers in the x and y dimensions by repeating the unit cell in the surface plane. The constructed surface consisted of 2208 atoms total and stoichiometrically balanced charge.

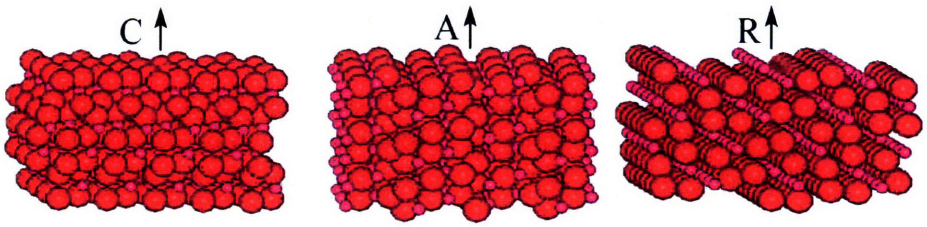

Figure 5-2: Crystal structure of α-Al_2O_3. The three unique crystal faces, termed C (0001), A ($11\bar{2}0$), and R ($1\bar{1}02$), shown here as the bulk stoichiometric truncation (Oxygen atoms shown in red, Aluminum atoms in pink)

There exists a wide variety of force-fields for inorganic materials, developed under an array of potential applications[51, 76]. However, the interaction between organic and inorganic components often does not follow simple application or combination of force-fields[34]. Therefore, it is necessary that force-fields specifically designed for the organic-inorganic interaction be constructed and validated for each system. For the current simulations, a set of non-bonded parameters were adapted from a previous study involving alkane adsorption to alumina clusters[52, 53]. This force-field was developed in relation to the OPLS united atom force-field, and is easily adapted to the current simulations.

Crystal atoms were assumed spatially fixed throughout the simulation eliminating the necessity for intra-crystal bonded atom potential energy parameters. While diffraction studies indicate some degree of mobility in the terminal hydroxyl surface layer, the lack of higher level density functional theory considerations prevents the

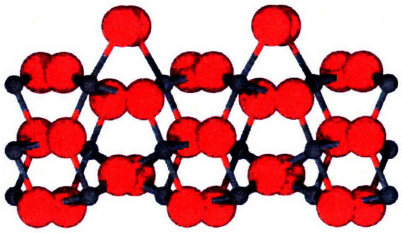

Figure 5-3: Rearrangement of the hydrated R-face ($1\bar{1}02$) of sapphire. The surface is characterized by a zero occupancy aluminum (gray) surface layer and relaxation of oxygen (red) layers

accurate inclusion of this mobility, and the surface is approximated as fixed.

5.3 Results

Monte Carlo simulations using the OPLS force-field have been shown to produce reliable results in peptide and protein structure and protein-ligand binding experiments[74]. Incorporation of implicit solvent models into molecular mechanics simulations has also been demonstrated in test cases to reproduce protein structural properties in explicit solvent simulations[131]. However, it has recently been demonstrated that implicit solvent models vary significantly in their application to surface adsorption studies, but that high quality results can be achieved through the choice of an accurate model[83]. As described in Chapter 3, the modified Generalized Born implicit solvent model used in the current simulations was verified for discrimination between conformations of model peptides, as well as in the direct application to surface adsorption simulations.

5.3.1 Validation of Model

The practicality of the simulation as a predictive method was examined through the simulation of six designer dodecamer peptides listed in Table 5.1[14]. These peptides were previously interrogated through a series of experiments which revealed distinct patterns of binding affinity[10]. Each peptide was engineered into the C-terminus of the surface displayed protein Aga2 in *S. cerevisiae* strain EBY100 downstream from a galactose based promoter. Regulated expression of this surface display was shown to induce yeast cell binding on crystalline sapphire surfaces. Surfaces were examined optically and differences in the percent area coverage (PAC) of yeast cells on the surface were used to infer relative binding affinities of the peptides. We note that the conversion of PAC numbers to binding energies is only approximate for several reasons. Most importantly, we lack detailed knowledge of how many copies of the peptide are interacting with the surface and what effect the yeast cell itself has on the binding properties. Since these PAC numbers only represent relative populations, we have normalized each relative affinity to the K1 peptide. The relative free energy of adsorption of peptide j compared to K1, $\Delta G_j^{ads} - \Delta G_{K1}^{ads}$, is related to the populations, P_j and P_{K1}, by

$$\begin{aligned} \Delta G_J - \Delta G_{K1} &= -k_B T \ln\left(\frac{P_j}{P_j^*}\right) + k_B T \ln\left(\frac{P_{K1}}{P_{K1}^*}\right) \\ &= -k_B T \ln\left(\frac{P_j}{P_{K1}}\right) - k_B T \ln\left(\frac{P_{K1}^*}{P_j^*}\right) \end{aligned} \tag{5.27}$$

where P^* is the bulk solution phase population. Since each yeast binding experiment began with a standardization of yeast cell concentration, the bulk concentration of each peptide is a constant and the second term in Equation 5.27 is zero. For validation of our simulation method against experiment, each calculated peptide-surface interaction free energy, $< G_{int} >=< G_{surf+pep} > - < G_{pep} > - < G_{surf} >$, was first differenced against K1. $< G_{surf} >$ and $< G_{pep} >$ are the simulation average free energies of the isolated surface and peptide, respectively, and $< G_{surf+pep} >$ is the simulation average free energy of the surface-peptide complex (see Table 5.1). The

number of peptides bound to the experimental surface in each case is unknown. This number would, ideally, scale the free energy difference in Equation 5.27. Hence, the relative free energies were further normalized against the total of the data set, (ie. $n\Delta\Delta G_j^{ads} / \sum_k n\Delta\Delta G_k^{ads}$), where n is the number of peptides. For simulated peptides, $n = 1$, while for experimental results n is unknown. We can then compare directly to experimentally determined binding affinities.

Table 5.1: List of designer dodecamer peptides, and average interaction free energy with the R-face of crystalline Sapphire. [a](G=glycine, K=lysine, C=cysteine, P=proline), [b]All sequences were capped by acetyl and methyl amide groups, [c]Energies in kcal/mol.

Name	Sequence[a,b]	$< G_{int} >^c$
K1	G*K*G*K*G*K*G*K*G*K*	-8.754
K2	GG*KK*GG*KK*GG*KK*	-5.590
K3	GGG*KKK*GGG*KKK*	-2.494
R1	G*R*G*R*G*R*G*R*G*R*	-0.895
cK1	CG*K*G*K*G*K*G*K*G*K*C	-9.016
K1P	G*K*P*K*G*K*P*K*G*K*P*K*	-3.004

This direct comparison is illustrated in Figure 5-4 for the six designer peptides in Table 5.1. The first three peptides, (K1, K2, and K3), show an interesting trend that is quite effectively captured by the simulations. As the charged lysine residues are grouped together, there is a significant reduction in binding affinity. This trend is well predicted by the simulations. The peptide R1 is a mutation of the K1 peptide in which all lysine residues have been altered to arginine. Experiments show that arginine binds less strongly than lysine, and it is speculated that delocalization of charge over the guanidinium group of arginine may reduce the interaction strength in comparison to the primary amine of lysine. From the plot in Figure 5-4, it appears the oligopeptide R1 is an outlier in its agreement between simulation and experiment. However, its place among those simulated is consistent with the model, compared to the other peptides. The OPLS-AA force-field represents the charge of the arginine guanidinium as fully delocallized and evenly distributed between two NH_2 functional groups. The delocalized charge would lessen the Coulombic interaction with an external charge

source. In addition, the guanidinium functional group requires the displacement of a larger volume of water from the hydrophillic sapphire surface than the smaller primary amine of lysine. This results in a free energy penalty for the surface desolvation by arginine relative to lysine. It is possible that the interaction of the guanidinium group with the negative surface could induce a (partial) localization of the positive charge on the arginine, resulting in increased Coulombic interactions and a smaller surface approach volume. Since the OPLS force-field contains no electronic structure calculations, this effect can not be investigated here, and the determination of the validity of this or other possible explanations will need to be the topic of future work. Higher level *ab initio* quantum chemical methods could be used to explore this possibility for single conformations, but would be too compuationally intensive for application in dynamical simulations. Thus, the examples below focus only on the lysine based peptides, where direct comparisons based on sequence and structure can be made without regard for functional group type.

Next, we consider two variations of the K1 peptide which can experimentally demonstrate the importance of peptide structure and flexibility. The peptide cK1 is a disulfide constrained, circular version of K1, while K1P has been altered by replacing three glycine residues with proline to introduce rigid kinks in the peptide structure. The constrained cK1 peptide shows (see Figure 5-4 and Table 5.1) little change in binding affinity in both yeast surface display experiments and molecular simulations. Analysis of the simulated (unconstrained) K1 end-to-end distance histogram reveals that a loop-like conformation is highly populated (see below and Figure 5-8). Thus the observed insensitivity in binding is sensible; constraining the peptide conformation in this way would not have a large effect on structure. Hence, peptide cK1 is not included in further discussion of structural differences between peptides. However, introduction of proline kinks has a significant effect on peptide adsorption (see Figure 5-4 and Table 5.1). The K1, K2, and K3 peptides are fifty percent glycine by composition and analysis of conformations from Monte Carlo simulations reveals high flexibility. This reduction in binding affinity in the proline variant thus suggests that this flexiblity is the key variable in binding.

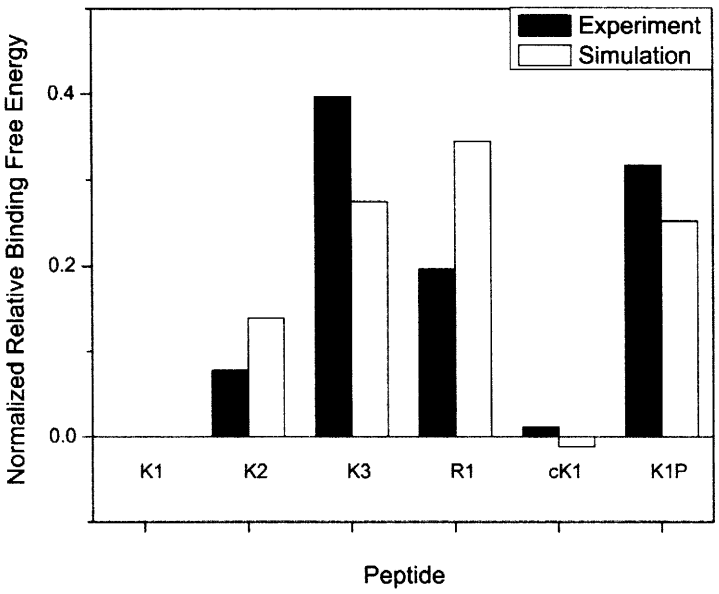

Figure 5-4: Comparison of adsorption simulation and experimental binding assay for designer dodecamer peptides. Experimental binding free energies relative to the K1 peptide are evaluated by Equation 5.27 from yeast cell populations bound to crystalline sapphire. Each binding free energy is normalized against the whole data set (ie. $\times 1/\sum G_j$), since the number of peptides participating experimentally in the yeast surface display is unknown. For reference, if the binding were driven by a single dodecamer polypeptide, the value of 0.1 on this scale would correspond to 2.28 kcal/mole of peptide

5.3.2 Sequence Dependence of Binding Affinity

The relationship between the peptides K1, K2, and K3 is interesting to analyze in more depth. Each peptide is a flexible, linear combination of exactly identical composition. Each peptide contains six glycine residues, an amino acid noted for structural flexibility due to a lack of steric hindrance from side-chain functional groups. This high degree of flexibility suggests that binding does not occur as a result of a matching of a well defined solution-phase structure, such as an alpha helix, with surface site arrangements. There must be some other dependence on peptide sequence able to modulate binding affinity. We are safe to assume that glycine residues do not interact with the negatively charged surface in an appreciable way and that all binding

is due to positively charged lysine residues. Therefore, sequence variation in this two component system can reasonably be described by two closely contributing factors: the space between adjacent lysines, and the grouping of lysines. It is convenient for elucidation to consider these separately.

Three series of peptides, listed in Table 5.2, were modeled to directly interrogate the effects of spacing and grouping of residues on the binding affinity of lysine based peptides. First, a set of decamer peptides, each consisting of two lysine and eight glycine residues were constructed with varying inter-lysine separations. A reference peptide with a single lysine residue was also included and used to normalize the binding free energies. Figure 5-5a shows the normalized (G_j/G_{ref}) peptide-surface interaction free energy for these di-lysine peptides as glycine spacers are inserted between the lysine residues. For closely grouped lysines, there is little cooperativity, with the interaction free energy approximately twice that of a single lysine. As glycine spacers are inserted, there is anti-cooperativity, with the interaction decreasing to only 1.4 times the interaction free energy of a single lysine. The increased separation of two lysine residues reduces the peptide-surface interaction towards that of a single residue, as one might expect due to the entropy cost of binding the peptide chain.

Table 5.2: List of peptides for testing lysine cooperativity in binding. [a](G=glycine, K=lysine), [b]All sequences were capped by acetyl and methyl amide groups.

di-Lysines[a,b]	Spaced Lysines[a,b]	Grouped Lysines[a,b]
GGGGG*K*GGGG	GGGGGG*K*GGGGG	GGGGGG*K*GGGGG
GGGG*KK*GGGG	GGGGG*K*G*K*GGGG	GGGGG*KK*GGGGG
GGGG*K*G*K*GGG	GGGG*K*G*K*G*K*GGG	GGGGG*KKK*GGGG
GGG*K*GG*K*GGG	GGG*K*G*K*G*K*G*K*GG	GGGG*KKKK*GGGG
GGG*K*GGG*K*GG	GG*K*G*K*G*K*G*K*G	GGGG*KKKKK*GGG
GG*K*GGGG*K*GG	G*K*G*K*G*K*G*K*G*K*	GGG*KKKKKK*GGG
GG*K*GGGGG*K*G		
G*K*GGGGGG*K*G		

The remaining two series of peptides were used together to investigate the role of grouping of charged residues by comparing the interaction free energy for alternating lysine-glycine patterns, as exemplified by the K1 peptide, with closely grouped lysine

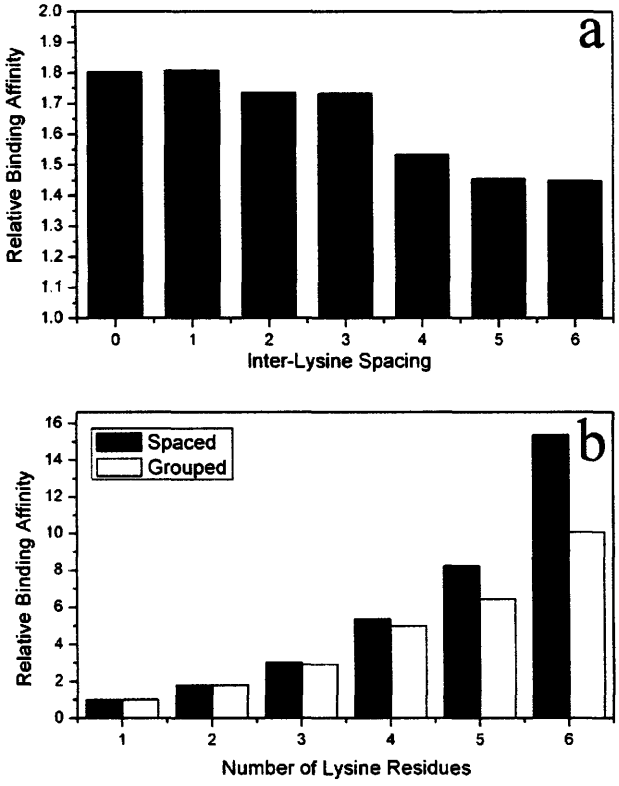

Figure 5-5: Cooperative binding properties of lysine based peptides. (a) Decrease in binding affinity as inter-lysine distance (in units of residue number) is increased, and (b) Cooperativity for grouped vs. spaced lysines. Binding energies are normalized to that of a single lysine residue in order to make clear the (anti-) cooperative nature of multiple lysine binding.

residues. In Figure 5-5b the interaction free energy is again normalized to that of a single lysine in order to make clear the cooperative nature of the binding process. Although there is clear cooperativity in each series, the effect is much more pronounced in the alternating sequences. For six lysine residues, the alternating sequence has a 53% higher interaction free energy than the grouped sequence. The binding of six grouped lysines is clearly cooperative, with the effective binding strength of 10 individual residues, while the alternating sequence produces with equivalent of over 15 times the binding free energy of a single lysine.

It is tempting to hypothesize that peptide adsorption in the sequence (K1, K2,

K3) is modulated by the ability of these peptides to simultaneously present their multiple positively charged lysine residues to the negatively charged sapphire surface. Alternating sequences have all lysine residues on the same side of the linear peptide chain, whereas gropued sequences will present some of their charged groups in opposite directions. A cursory analysis based on this simple principle is illustrated in Figure 5-6 and would (incorrectly) predict the order of binding as K1>K3>K2.

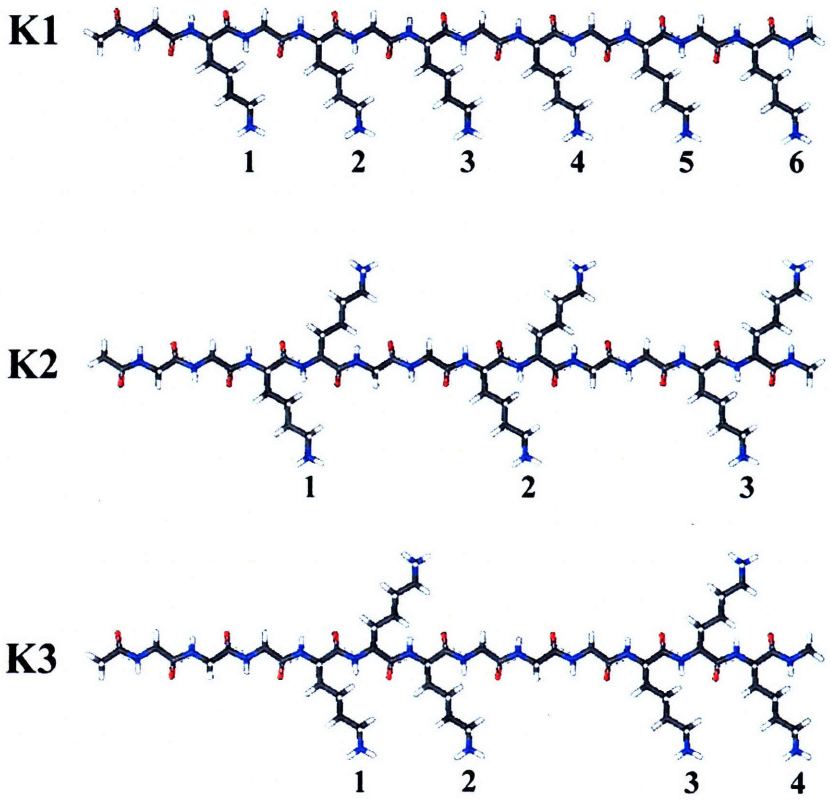

Figure 5-6: Structure of designed dodecamer peptides, indicating the naïve expectations for an extended linear peptide's ability to present positively charged functional groups to the negatively charged surface.

In Figure 5-7, the ability of each peptide to bind multiple lysine residues is examined by defining a pair-correlated density profile near the crystal surface as the vertical position of the amine functional group of residue i when residue $j \neq i$ is bound to the surface. For this purpose, we define "bound" operationally as a separation distance

of less than 4 Å between the primary amine nitrogen of lysine and the plane through the first surface layer of atoms in the sapphire crystal. This profile is plotted, along with a schematic representation of the relavent inter-lysine spacings, for K1, K2, K3, and K1P in Figure 5-7a-d, respectively. The legends in insets are numbered by amino acid separation. For example, the sequence (KGK) represents a 1-3 spacing and (KGGK) represents a 1-4 spacing. For a linear conformation, odd numbered interlysine spacings, such as 1-3, 1-5, and 1-7, have lysines facing the same direction from the peptide backbone, while even numbered spacings 1-2, 1-4, and 1-6 face opposite directions. The odd numbered spacings are highly localized at the surface in K1 and K3, indicating a strong cooperativity, as expected from the analysis shown in Figure 5-6. Large inter-lysine spacings in K1 are influenced by the intermediate lysines and do not show the separation dependence of Figure 5-5. However, this dependence is apparent for K3, where the 1-7 separation is considerable less localized at the surface than shorter inter-lysine spacings (see Figure 5-7c). The 1-2 spaced lysines in K3 are localized away from the surface, uninvolved in surface binding, indicating a primary reason for the decrease in affinity compared to K1. Analysis of these correlations for K2 reveal that there is a small propensity for the peptide to turn on its side and bind both residues in 1-2 spaced arrangements. For well separated residues, the peptide apparently is increasingly able to twist to allow such even-spaced residues to interact with the surface (see Figure 5-7b). While the even numbered spacings are largely eliminated from binding in K3 (see Figure 5-7c), they are still able to interact significantly with the surface in K2, leading to stronger binding than the naïve analysis in Figure 5-6 predicts. Finally, the profile for K1P (Figure 5-7d) demonstrates that the rigid structural kinks imposed by the replacement of glycine residues with proline results in the prevention of cooperative interaction with the surface. Although this peptide contains the same sequential arrangement of lysine residues as K1, the lack of flexibility in the peptide backbone precludes simultaneous presentation of these residues to the surface (cf. Figure 5-7a and Figure 5-7d).

In Figure 5-8a-d, the peptide end-to-end distance is used to evaluate structural change imposed by surface adsorption for peptides K1, K2, K3, and K1P, respectively.

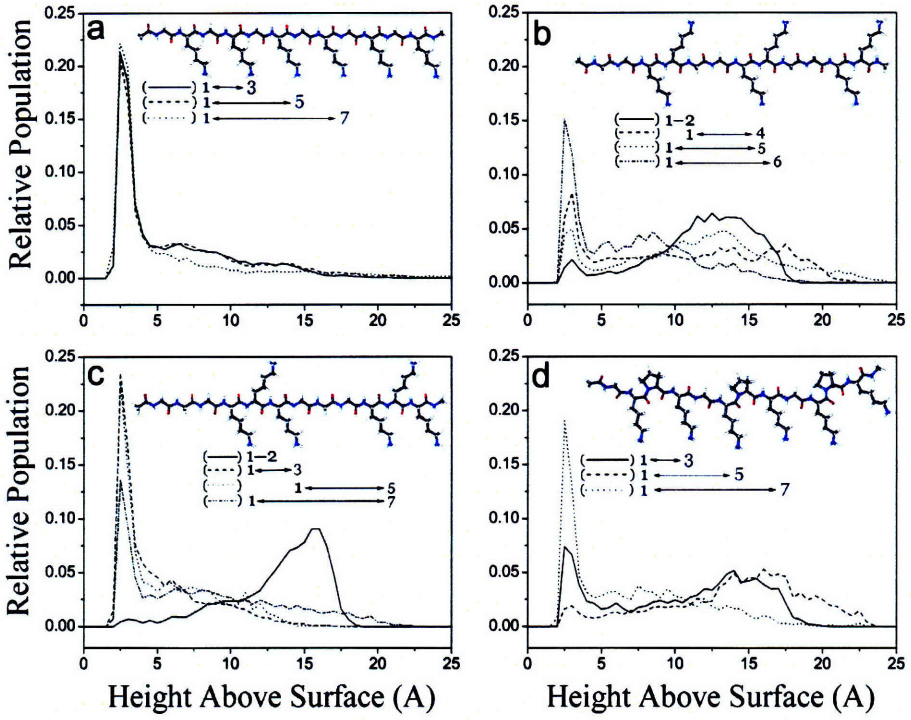

Figure 5-7: Pair correlation density profile for cooperative binding, and schematic representation of important inter-lysine spacings (inset) for peptides (a) K1, (b) K2, (c) K3, and (d) K1P. This function represents the relative density of the amine functional group of residue i above the surface when residue $j \neq i$ is bound to the surface. "Bound" is defined opperationally here as a distance of less than 4 Å between the amine nitrogen and the first plane of crystal surface atoms. Binding is indicated by localization near the surface, while structural limitations for multiple lysine binding is manifest by localization away from the surface.

The K1 and K3 peptides present the lysine residues responsible for surface binding from one side of the peptide backbone and undergo little change in end-to-end distance upon adsorption (see Figure 5-8a,c). However, the K2 peptide shows a pronounced structural change upon adsorption (Figure 5-8b). Twisting of the peptide to lay flat at the surface causes an increase in the end-to-end distance. This effect is also apparent in the K1P peptide, where structural changes must be accomodated for maximal interaction with the surface (Figure 5-8d).

Representative peptide structures are presented in Figure 5-9. While previous computational studies of material binding peptides and adsorption at solid surfaces

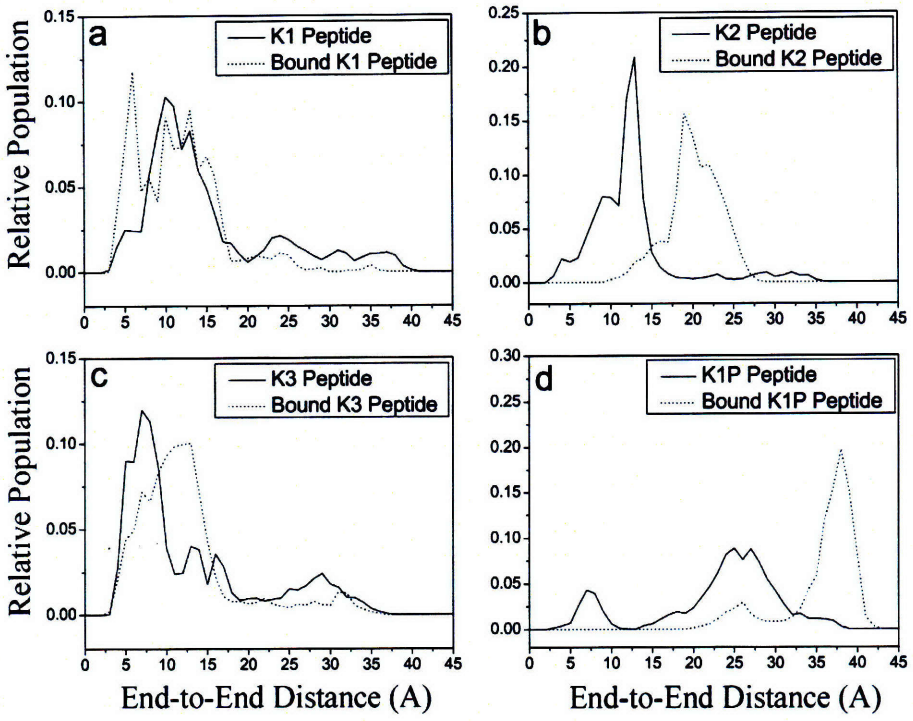

Figure 5-8: End-to-end distance histograms for sapphire binding peptides (a) K1, (b) K2, (c) K3, and (d) K1P. Differences in end-to-end distance indicate conformational changes imposed by surface adsorption

have predicted a dependence on secondary structure[40, 72], the peptides studied here generally form only random coil conformations. This is consistent with the lack of secondary strucutre in poly(L-lysine) at physiological conditions[78] as well as the large compositional fraction of glycine, an amino acid noted for its structural flexibility. In this random coil conformation, the K1 peptide is able to efficiently present many lysine residues to the sapphire surface (see Figure 5-9a) and forms a highly populated loop-like structure in both the solution phase and surface bound states (see Figure 5-8a and Figure 5-9b). The peptide K2 (Figure 5-9c), in contrast, must twist around its backbone in order to present multiple residues to the sapphire surface (see Figure 5-8b and Figure 5-9). The surface bound conformation of K2 shown here is largely linear along the peptide backbone, with only the glycine-glycine terminal section curling back, and away from the surface. Finally, Figure 5-9e clearly

shows the even-numbered-spacing lysine residue held away from the surface, while the two lysines on the opposite side of the peptide backbone cooperatively bind to the surface.

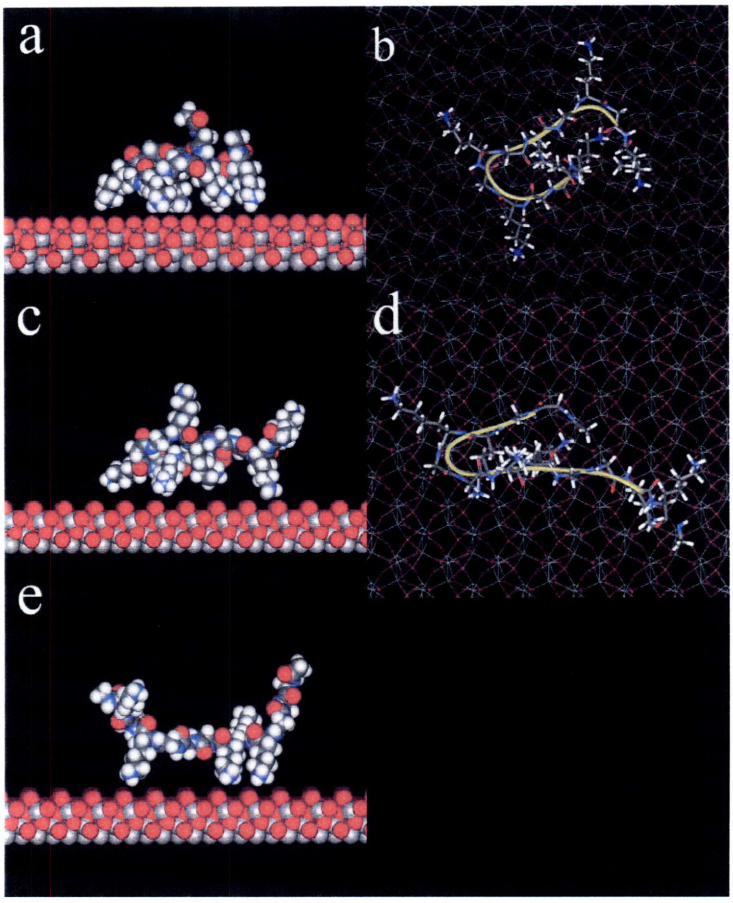

Figure 5-9: Representative conformations of bound peptides. (a,b) The surface bound K1 peptide is able to simultaneously present many lysine residues to the surface and forms a highly populated loop-like structure. (c,d) The surface bound K2 peptide twists to present residues to the surface, resulting in a straightened backbone compared to the solution phase ensemble. (e) The even spaced lysine residues are localized away from the surface and do not participate in surface binding

In the following discussion, the set of rationally designed peptides (K1, K2, K3, cK1, and K1P) will be referred to as the "Kx" series.

5.3.3 Predictive Screening of Peptides

With the developed model demonstrating the ability to reproduce experimentally observed differences in binding affinity, it is then desirable to use the model to make predictions about experimentally unobserved peptide systems. In order to explore the predictive capabilities of the model, a new set of identically composed peptides were designed. The original application of the model was the binding of highly basic, positively charged peptides to a negatively charged surface. While the differentiation of binding affinities is the ultimate goal, and is reasonably well predicted, the binding of highly and oppositely charged molecules does not represent the most challenging system. Consistency is maintained in the choice of material in the crystal surface, but to increase the level of challenge in the system, the new set of peptides were constructed to be net neutral in charge. Each peptide consists of a random sequence of four lysine, four glycine, and four glutamic acid residues. There are approximately 32,000 unique sequences (excluding reversed sequences) which can be composed from this set of amino acids. A subset of 24 sequences were selected at random from this list of possible sequences (see Table 5.3)

Table 5.3: Set of random, net neutrally charged peptides used in predictive screening experiments. [a](G=glycine, K=lysine, E=glutamic acid), [b]All sequences were capped by acetyl and methyl amide groups

Name	Sequence[a,b]	Name	Sequence[a,b]
s01	EGKEGGGKKKEE	s13	GKKGEGKKEEGE
s02	EKGKEKEKGGGE	s14	KEGKGGGKEEKE
s03	EKGKKEKGEGGE	s15	KGEEGKKEGGKE
s04	EKKEKGGKGEGE	s16	KGEKKGGEEKEG
s05	GEGGKGKEEKKE	s17	KGGEKEEGKKGE
s06	GEKKKGKEGEEG	s18	KGGEKEKEGEGK
s07	GEKKKKEGEGEG	s19	KGKEGKEKGEGE
s08	GGEGEKEGKKKE	s20	KKEEGGEGKGKE
s09	GGKGKEKGEEKE	s21	KKEKGKEEEGGG
s10	GGKKEEGGKKEE	s22	KKGEEEKGKGEG
s11	GKEEKEKGGGKE	s23	KKGGGEGKEEEK
s12	GKGEKKEGKGEE	s24	KKKGGGEEEEKG

Each peptide was computationally screened against the R-face ($1\bar{1}02$) face of hydrated sapphire. Calculated binding energies for each peptide are shown in Figure 5-10. Binding energies generally follow lysine spacing and grouping rules that were discovered through the simulations of the Kx peptides. For example, the peptide s02 has optimal spacing of lysine residues, and is again predicted to be a strong binder relative to the other peptides in the subset. However, the magnitude of the binding free energy is significantly reduced through the limitation of four lysine residues, compared to six in the Kx peptides, as well as by the introduction of negatively charged glutamic acid residues. Similarly, the peptide s07, with lysine residues grouped at positions three through six, represents a relatively weak binder.

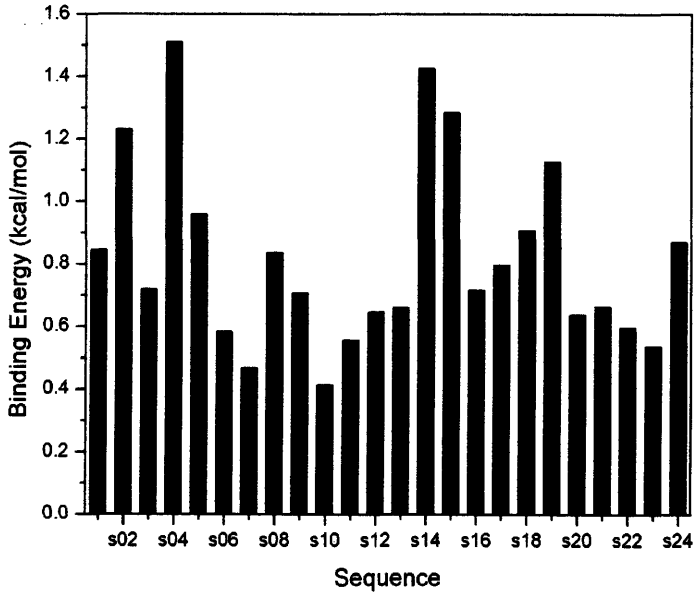

Figure 5-10: Calculated binding energies for the predictive screening of the neutral peptides listed in Table 5.3

The consistency of the general binding rules based on grouping and spacing of lysine residues was then further explored through sequence mutations to three of the peptides from the list in Table 5.3. The peptide s02 has optimal spacing of lysines and represents a strong binder. Mutations based on carriage shifts, and swapping of

a single pair of amino acids were used to create eight new variants of the s02 peptide (see Table 5.4). Each of these peptides was simulated under the same procedure outlined for the Kx peptides. Binding energies for each of the s02 variants are shown in Figure 5-11. Carriage shift mutations that position lysine residues towards the ends of the peptide result in a slight increase in binding affinity (s02.1-s02.3). This can be attributed to the ability to bind lysine residues while maintaining a higher entropic contribution of the unbound end, as well as the freedom of glutamic acid residues to be localized away from the surface at the unbound end of the peptide. Deliberate amino acid pair swap mutations (s02.4-s02.8) aimed at increasing the grouping of lysine residues has the effect of decreasing the predicted binding affinity, again following the spacing/grouping rules discovered through the Kx peptides.

Table 5.4: Set of variants on the s02 peptide (see Table 5.3) based on carriage shift and amino acid pair swap mutations used to test the ability to make deliberate mutations to peptides. [a](G=glycine, K=lysine, E=glutamic acid), [b]All sequences were capped by acetyl and methyl amide groups

Name	Sequence[a,b]	Name	Sequence[a,b]
s02.1	GEEKGKEKEKGG	s02.5	EKEKEKGKGGGE
s02.2	GGGEEKGKEKEK	s02.6	EKGGEKEKKGGE
s02.3	EKGGGEEKGKEK	s02.7	EKKKEGEKGGGE
s02.4	EKGKEGEKGKGE	s02.8	EKKKEKEGGGGE

This experiment was repeated for a second strong binding peptide, s14, and a weak binding peptide, s07. Sequece mutated variants for s07 and s14 are listed in Table 5.5 and Table 5.6, respectively. Binding free energies for s07 variants and s14 variants are shown in Figure 5-12 and Figure 5-13, respectively. Deliberate mutations of the strong binding peptide s14 that increase the grouping of lysine residues, as well as mutations which place lysine residues amongst groups of glutamic acid residues have the effect of decreasing binding affinity. In contrast, mutations of the s07 relatively weak binding peptide that decrease the grouping of lysine residues generally increase the binding affinity.

With a general set of predictive rules based simply on lysine spacing patterns, it

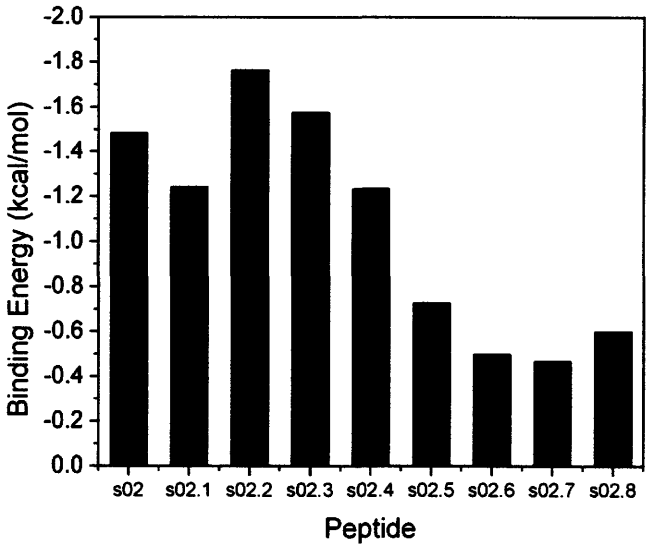

Figure 5-11: Binding free energies for the mutated variants of peptide s02 (see Table 5.3 and Table 5.4). Deliberate mutations resulting in increased grouping of lysine residues generally produce a decrease in predicted binding affinity

is desirable to develop a simple scoring function that could forego the computational cost of molecular simulations. A similar set of mutated variant peptides was produced for the s04, s10, s20, and s24 peptides, and a multiple regression fit was calculated to the occurance of certain subsequence patterns for the entire set of 80 polypeptides. Specifically, we wish to score peptide binding based on the lysine subsequence patterns identified through the Kx rationally designed sequences. Each peptide was examined for the occurance of "KK", "KXK", "KXXK", and "KXXXK" subsequences, hereafter referred to as K(1-2), K(1-3), K(1-4), and K(1-5), respectively. The variants explored in the previous discussion also demonstrate a dependence on the proximity of negatively charged glutamic acid residues to the positively charged lysine residues. The sequences were therefore also examined for the occurance of the sequences "KE", "KXE", "KXXE", and "KXXXE", hereafter referred to as E(1-2), E(1-3), E(1-4), and E(1-5), respectively. The multiple regression fit gives parameters, in units of energy (kcal/mol), that each of these subsequences contributes to the free energy of adsorption, $F_{K(1-2)}$ for the K(1-2) subsequence, and the scored free energy is calculated

Table 5.5: Set of variants on the s07 peptide (see Table 5.3) based on carriage shift and amino acid pair swap mutations used to test the ability to make deliberate mutations to peptides. [a](G=glycine, K=lysine, E=glutamic acid), [b]All sequences were capped by acetyl and methyl amide groups

Name	Sequence[a,b]	Name	Sequence[a,b]
s07.1	EGGEKKKKEGEG	s07.5	KEKGKKEGEGEG
s07.2	EGEGGEKKKKEG	s07.6	GGKKKKEGEGEE
s07.3	EGEGEGGEKKKK	s07.7	GEKKKGEKEGEG
s07.4	GEKKGKEKEGEG	s07.8	GEKGKKEGEGEK

Table 5.6: Set of variants on the s14 peptide (see Table 5.3) based on carriage shift and amino acid pair swap mutations used to test the ability to make deliberate mutations to peptides. [a](G=glycine, K=lysine, E=glutamic acid), [b]All sequences were capped by acetyl and methyl amide groups

Name	Sequence[a,b]	Name	Sequence[a,b]
s14.1	KEKEGKGGGKEE	s14.5	KEKKGGGKEEGE
s14.2	EEKEKEGKGGGK	s14.6	KEGKGGKKEEGE
s14.3	GKEEKEKEGKGG	s14.7	KEKGGGGKEEKE
s14.4	KEGKGKGGEEKE	s14.8	KEGGKGGKEEKE

as

$$\begin{aligned} \underset{\sim}{\Delta G_{ads}} &= F_{K(1-2)}N_{K(1-2)} + F_{K(1-3)}N_{K(1-3)} + F_{K(1-4)}N_{K(1-4)} \\ &+ F_{K(1-5)}N_{K(1-5)} + F_{E(1-2)}N_{E(1-2)} + F_{E(1-3)}N_{E(1-3)} \\ &+ F_{E(1-4)}N_{E(1-4)} + F_{E(1-5)}N_{E(1-5)} \end{aligned} \tag{5.28}$$

where $N_{X(i-j)}$ is the number of occurances of the subsequence $X(i-j)$. Figure 5-14 shows a comparison between binding free energies calculated from molecular simulation and those calculated using this simple scoring function. The scoring function has only weak correlation to simulation results (R^2=0.41), indicating that the factors involved in modulation of the binding affinity have been oversimplified. However, the energy parameters are generally consistent with the rules observed for lysine sequences in the Kx series and demonstrate a clear dependence on the proximity of

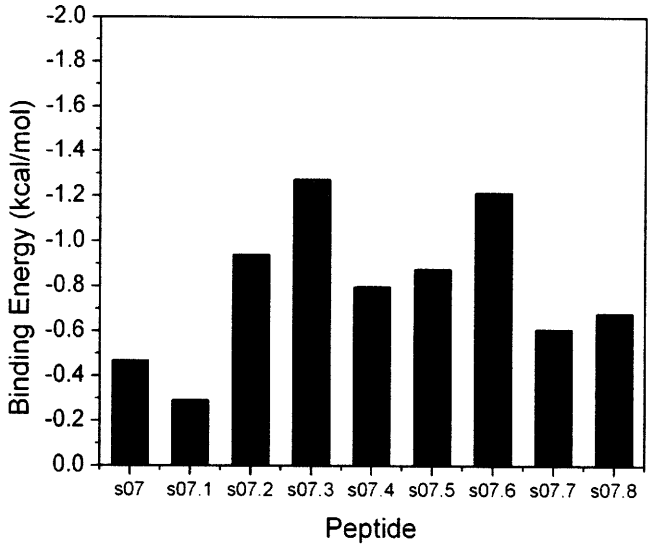

Figure 5-12: Binding free energies for the mutated variants of peptide s07 (see Table 5.3 and Table 5.5). Deliberate mutations resulting in increased spacing of lysine residues generally produce an increase in predicted binding affinity

glutamic acid residues to the binding lysines (see Table 5.7). Positive values are unfavorable to binding, while negative values represent a favorable contribution. The grouped lysine pattern, K(1-2) is highly unfavorable, consistent with the hypothesis that this prevents interaction of one of the lysine residues with the crystal surface. The optimally spaced lysine pattern, K(1-3), is the most favorable pattern, again consistent with structural predictions. The K(1-4) pattern is small, but favorable, while the K(1-5) pattern has a larger favorable impact. The proximity of glutamic acid to binding lysine residues is always an unfavorable contribution, decreasing as the distance between residues increases.

Maltose Binding Protein-Peptide Fusion

Finally, in order to explore the predictive capabilities of the computational model for the binding of neutral peptides to crystalline sapphire, an alternative binding assay was developed. Yeast display libraries showed that a high positive charge was required for peptides to overcome the repulsion between the negatively charged yeast

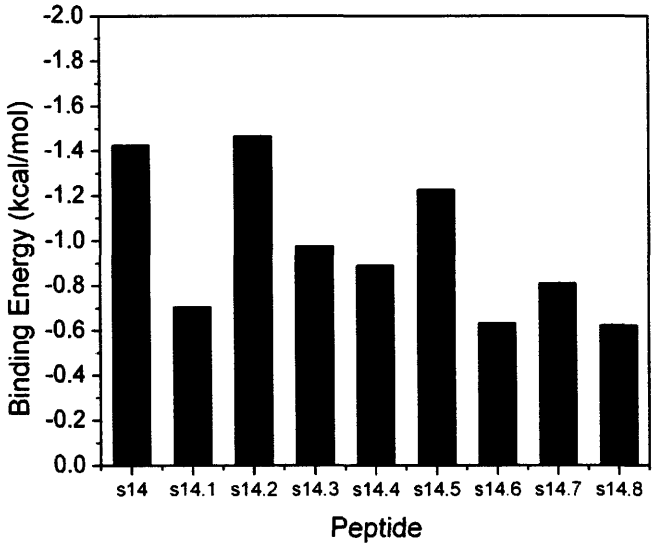

Figure 5-13: Binding free energies for the mutated variants of peptide s14 (see Table 5.3 and Table 5.6). Deliberate mutations resulting in increased grouping of lysine residues generally produce a decrease in predicted binding affinity

surface and the hydrated sapphire surface. Thus, the neutral peptides explored here are unlikely to induce yeast cell adsorption in the surface display experiments used in the Kx series.

Each peptide was cloned onto the c-terminus of maltose binding protein (MBP) in order to build a modified ELISA assay for peptide adsorption, following a previously developed protocol[10]. A multiple digestion of the pMAL-c2x vector (New England Biolabs, Beverly, MA), using EcoRI and HindIII allowed for the insertion of oligonucleotides on the c-terminus end of the gene encoding cytoplasmically expressed MBP. Complementray oligos with EcoRI and HindIII-compatible ends, encoding the s02 (MBP-s02), s07 (MBP-s07), s14 (MBP-s14), K1 (MBP-K1), R1 (MBP-R1), and stop codon (MBP*), were annealed and ligated into the digested pMAL-c2x vector. The vector was then transformed into chemically competent TOP10 *E. coli* (Invitrogen) and cloning success was verified through sequencing. DNA from successful clones was then transformed into chemically competent TB1 *E. coli* for protein expression.

TB1 *E. coli* harboring the modified pMAL vectors were grown to mid-log phase in

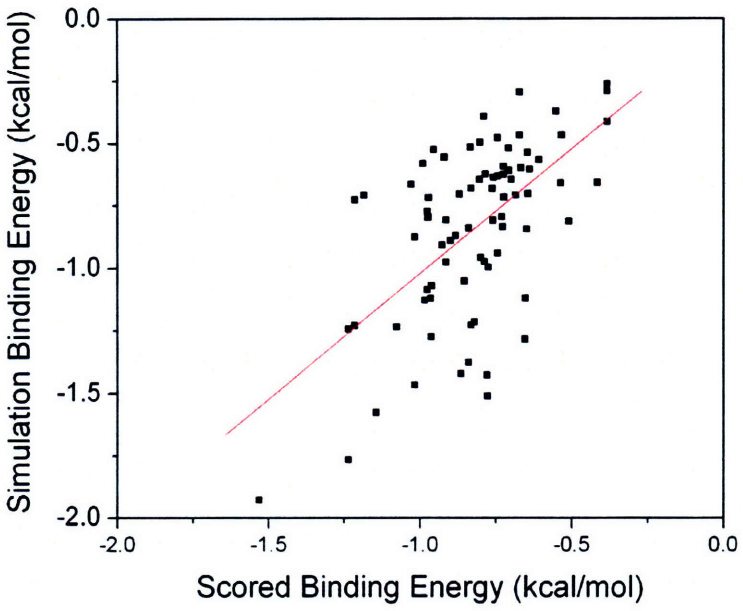

Figure 5-14: Comparison of binding free energies calculated from molecular simulations and free energies calculated from the simple scoring function given in Equation 5.28. The scoring function has only weak correlation to simulation results, indicating an oversimplification of involved parameters. However, the parameters used are consistent with observed rules regarding the spacing of lysine and glutamic acid residues (see Table 5.7).

Glucose-Rich Media plus ampicillin before induction with IPTG to a final concentration of 0.3 mM. After two hours of induction, cells were harvested by centrifugation and frozen overnight at -20 degrees Celsius. The cells were then thawed in cold water and lysed by probe sonication. The crude extract was separated from the insoluble cell matter by centrifugation and applied to an amylose resin column. The bound MBP constructs were then eluted from the column with 20 mM maltose in 1× column buffer (20 mM Tris HCl, 1mM EDTA, 200 mM NaCl) and concentrated in 10,000 MWCO Centricon Plus-20 centrifugal filtration devices (Millipore, Billerica, MA). Purification steps were monitored by SDS-PAGE and the final concentration of protein was calculated by absorbance at 280 nm and referenced with a known MBP standard from New England Biolabs.

MBP construct stocks were diluted down to the appropriate concentration in 1×

Table 5.7: Multiple regression fit parameters for scoring binding free energies based on the occurance of subsequences of lysines

Subsequence	Parameter (kcal/mol)
K(1-2)	1.694
K(1-3)	-1.407
K(1-4)	-0.159
K(1-5)	-0.790
E(1-2)	0.934
E(1-3)	0.584
E(1-4)	0.523
E(1-5)	0.215

PBS containing 0.1% Tween-20 (PBST). Two hundred and fifty microliters of protein solution were added to clean sapphire substrates in 48-well plates and incubated for 3 h under constant agitation on an orbital shaker. Substrates were washed twice, each time transferring to new wells containing 400 μL PBST and agitating for 15 min. Substrates were then transferred to wells containing a 2,000-fold dilution of stock HRP conjugated anti-MBP monoclonal antibody (New England Biolabs) in PBS containing 5 mg/mL BSA (PBS-BSA) and agitated for 30 min. The substrates were washed two times as before with PBS-BSA then transferred to a clean well containing 200 μL of chromogen solution (0.5 mg/mL ABTS (2,2'-Azino-di-(3-ethylbenz-thiazoline Sulfonic Acid)), 0.03% hydrogen peroxide in 0.1 M citrate buffer, pH4.2) and agitated. The absorbance of each well at 405 nm (A_{405}) was then monitored on a UV/Vis plate reader (SpectraMAX 250, Molecular Devices, Sunnyvale, CA).

MBP-K1 and MBP-R1 were first used to verify assay results consistent with the yeast surface display experiments. Figure 5-15 shows the binding of the K1- and R1-MBP fusion proteins in relation to the control stop codon MBP*. Binding strength is generally consistent with yeast surface display experiments. Protein concentrations were 1.0 μg/mL. The ELISA was then used to explore the binding strength of the three neutral peptides s02, s07, and s14. Figure 5-16 demonstrates the relative binding affinities for these peptides. The higher concentration necessary for the much weaker binding neutral peptides generally increases background signal, decreasing

reproducability. However, the general relation (s02, s14 strong binding - s07 weak binding) is observed in the MBP modified ELISA assay.

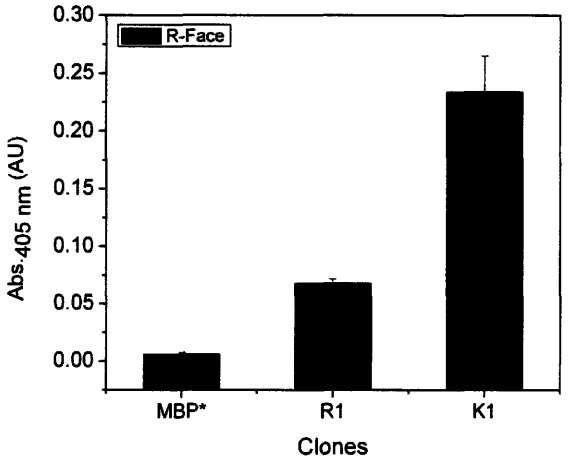

Figure 5-15: Modified ELISA assay for K1 and R1-MBP fusion proteins. Results are generally consistent with yeast surface display experiments. Protein concentration during incubation was 1 μg/mL.

5.4 Conclusions

In the preceding discussion we have focused on the interaction enthalpy and changes in the free energy of hydration. The remaining term in Equation 5.2 representing peptide entropic contributions is difficult to calculate. For these relatively large flexible molecules, it can be argued that entropic change due to the loss of translational and rotational degrees of freedom are offset in part by the introduction of new vibrational degrees of freedom, and relatively small compared to internal peptide conformational entropy. The adsorption process forces a restriction of the peptide conformational space in comparison to the non-adsorbed state. Basalyga and Latour, Jr. [75] present the argument that residue functional groups in the non-adsorbed state already are subject to a considerable degree of restriction due to the presence of adjacent functional groups. They propose the example of a five to ten-fold restriction in conformational freedom, leading to a value of $-T\Delta S_{ads}$ of 1.0 to 1.5 kcal/mol. In

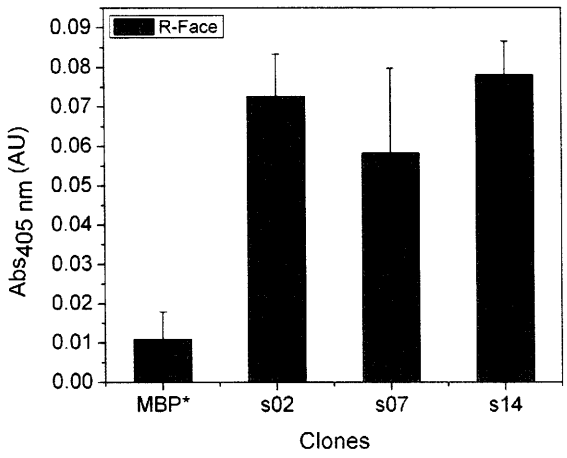

Figure 5-16: Modified ELISA assay for the neutral peptides s02, s07, and s14. Protein concentration during incubation was 1 mg/mL, 1000× the concentration used for K1, while the adsorbance signal is significantly weaker. Binding strength is reduced due to the presence of negatively charged residues and higher protein concentrations must be used. The higher concentration of protein increases background signal and generally reduces the reproducability of the experiment. However, the binding order of these three peptides (s02, s14 strong binding - s07 weak binding) is predicted.

contrast, Mungikar and Forciniti[40] predict quite small, (and favorable) changes in entropy for small alpha-helical peptides adsorbed to solid surfaces. For the purposes of this study, the identically composed peptides can be assumed to have similar conformational freedom in the non-adsorbed state. Changes on binding should be reflected in conformational distributions. Analysis of the end-to-end distance for the K1 and K3 peptides show little change in conformation on binding. While the K2 peptide demonstrates a difference in average end-to-end distance upon adsorption, the width of the distribution, and thus the conformational range, is similar. This suggests a similarity between the conformational entropy of each of the three peptides, generally consistent with the previous discussion in that experimentally observed differences in binding affinity are reasonably well predicted by changes in adsorption enthalpy and hydration free energies. In less favorable cases, a more complex simulation including population exchange between bound and free states could directly account for this term.

Computational simulations of peptide and protein adsorption to inorganic surfaces offers the potential to provide detailed mechanisms which have largely escaped experimental characterization. This work builds upon previous experimental studies in rationally designed metal-oxide binding peptides, as well as previous computational studies of amino acid and peptide adsorption to synthetic surfaces. The adsorption properties of the peptide system demonstrated here are dependent on the adsorption enthalpy and the interaction of the peptides with the polarizable solvent.

Experimental methods based on combinatorial libraries of peptides offer an efficiency that cannot be approached in the forseeable future by computational screening methods. For example, biopanning experiments often begin with libraries in excess of 10^9 individual sequences and can reach a concensus binding motif in five or fewer rounds of selection. For comparison, the binding simulations in this work took $\sim$12 days each running on 2.4 GHz Intel Xeon processors. Sequentially screening a library of only 1000 peptides would thus take 32 years of processor time, and it is not suggested to make such an application of the currently developed computational model. However, experimental biopanning methods are unable to provide information about the surrounding sequence space, those sequences differing only slightly from that selected, or the precise role of individual amino acids. This information is efficiently provided by simulation, and it is here that the computational model can be applied in providing guidance for deliberate and judicious mutations of experimentally selected peptides.

Bibliography

[1] G. Ma; D.F. Liu; H.C. Allen. Piperidine adsorption on hydrated alpha-alumina (0001) surface studied by vibrational sum frequency generation spectroscopy. *Langmuir*, 20(26):11620–11629, 2004.

[2] Roberto Bagnara. A unified proof for the convergence of Jacobi and Gauss-Seidel methods. *Society for Industrial and Applied Mathematics Review*, 37(1):93–97, 1995.

[3] Jason Wagoner; Nathan A. Baker. Solvation forces on biomolecular structures: A comparison of explicit and Poisson-Boltzmann models. *Journal of Computational Chemistry*, 25(13):1623–1629, 2004.

[4] Nathan A. Baker. Improving implicit solvent simulations: a Poisson-centric view. *Current Opinion in Structural Biology*, 15(2):137–143, 2005.

[5] Alexey Onufriev; David A. Case; Donald Bashford. Effective born radii in the Generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry*, 23(14):1297–1304, 2002.

[6] Ahmad S. Khalil; Jorge M. Ferrer; Ricardo R. Brau; Stephen T. Kottmann; Christopher J. Noren; Matthew J. Lang; Angela M. Belcher. Single M13 bacteriophage tethering and stretching. *Proceedings of the National Academy of Sciences*, 104(12):4892–4897, 2007.

[7] Beau R. Peele; Eric M. Krauland; K. Dane Wittrup; Angela M. Belcher. Probing the interface between biomolecules and inorganic materials using yeast surface display and genetic engineering. *Acta Biomaterialia*, 1(2):145–154, 2005.

[8] Chuanbin Mao; Christine E. Flynn; Andrew Hayhurst; Rozamond Sweeney; Jifa Qi; George Georgiou; Brent Iverson; Angela M. Belcher. Viral assembly of oriented quantum dot nanowires. *Proceedings of the National Academy of Sciences*, 100(12):6946–6951, 2003.

[9] Chuanbin Mao; Daniel J. Solis; Brian D. Reiss; Stephen T. Kottmann; Rozamond Y. Sweeney; Andrew Hayhurst; George Georgiou; Brent Iverson; Angela M. Belcher. Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. *Science*, 303(5655):213–217, 2004.

[10] Eric M. Krauland; Beau R. Peelle; K. Dane Wittrup; Angela M. Belcher. Peptide tags for enhanced cellular and protein adhesion to single-crystalline sapphire. *Biotechnology and Bioengineering*, 97(5):1009–1020, 2007.

[11] Ki Tae Nam; Dong-Wan Kim; Pil J. Yoo; Chung-Yi Chiang; Nonglak Meethong; Paula T. Hammond; Yet-Ming Chiang; Angela M. Belcher. Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science*, 312(5775):885–888, 2006.

[12] Sandra R. Whaley; D.S. English; Evelyn L. Hu; Paul F. Barbara; Angela M. Belcher. Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature*, 405(6787):665–668, 2000.

[13] Seung-Wuk Lee; Chuanbin Mao; Christine E. Flynn; Angela M. Belcher. Ordering of quantum dots using genetically engineered viruses. *Science*, 296(5569):892–895, 2002.

[14] Stephen T. Kottmann; Eric M. Krauland; Peter J. Rossky; Angela M. Belcher. Computational modeling of sapphire-binding peptides. *In Preparation*, 2007.

[15] Gustavo Stolovitzky; B.J. Berne. Catalytic tempering: A method for sampling rough energy landscapes by Monte Carlo. *Proceedings of the National Academy of Sciences*, 97(21):11164–11169, 2000.

[16] H. Brandstetter; D. Turk; H.W. Hoeffken; D. Grosse; J. Sturzebecher; P.D. Martin; B.F.P. Edwards; W. Bode. Refined 2.3å X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA: A starting point for improving antithrombotics. *Journal of Molecular Biology*, 226(4):1085–1099, 1992.

[17] A.S. Templeton; T.P. Trainor; S.J. Traina; A.M. Spormann; G.E. Brown. Pb(II) distributions at biofilm-metal oxide interfaces. *Proceedings of the National Academy of Sciences*, 98(21):11897–11902, 2001.

[18] Robert E. Bruccoleri. Grid positioning independence and the reduction of self-energy in the solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 14(12):1417–1422, 1993.

[19] Thomas Simonson; Axel T. Brünger. Solvation free energies estimated from macroscopic continuum theory: An accuracy assessment. *Journal of Physical Chemistry*, 98(17):4683–4694, 1994.

[20] David Keller; David Swigon; Carlos Bustamante. Relating single-molecule measurements to thermodynamics. *Biophysical Journal*, 84(2):733–738, 2003.

[21] Marco Scarsi; Joannis Apostolakis; Amedeo Caflisch. Comparison of a GB solvation model with explicit solvent simulations: Potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. *Journal of Physical Chemistry B*, 102(18):3637–3641, 1998.

[22] Irina V. Ionova; Emily A. Carter. Direct inversion in the iterative subspace-induced acceleration of the ridge method for finding transition states. *Journal of Chemical Physics*, 103(13):5437–5441, 1995.

[23] Irina V. Ionova; Emily A. Carter. Orbital-based direct inversion in the iterative subspace for the generalized valence bond method. *Journal of Chemical Physics*, 102(3):1251–1256, 1995.

[24] Charles L. Brooks III; David A. Case. Simulations of peptide conformational dynamics and thermodynamics. *Chemical Reviews*, 93(7):2487–2502, 1993.

[25] Donald Bashford; David A. Case. Generalized Born models of macromolecular solvation effects. *Annual Review of Physical Chemistry*, 51:129–152, 2000.

[26] Holger Gohlke; David A. Case. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *Journal of Computational Chemistry*, 25(2):238–250, 2004.

[27] Jooyoung Lee; M.Y. Choi. Optimization by multicanonical annealing and the traveling salesman problem. *Physical Review E*, 50(2):R651–R654, 1994.

[28] E. Comini. Metal oxide nano-crystals for gas sensing. *Analytica Chimica Acta*, 568(1):28–40, 2006.

[29] William L. Jorgensen; MCPRO; Version 1.68; Yale University; New Haven; CT.

[30] J.D. Currey. Mechanical properties of mother of pearl in tension. *Proceedings of the Royal Society of London B*, 196(1125):443–463, 1977.

[31] R. de Picciotto; H.L. Stormer; L.N. Pfeiffer; K.W. Baldwin; K.W.West. Four-terminal resistance of a ballistic quantum wire. *Nature*, 411(6833):51–54, 2001.

[32] David J. Earl; Michael W. Deem. Optimal allocation of replicas to processors in parallel tempering simulations. *Journal of Physical Chemistry B*, 108:6844–6849, 2004.

[33] B. Wood; W.M.C. Foulkes; M.D. Towler; N.D. Drummond. Coulomb finite-size effects in quasi-two-dimensional systems. *Journal of Physics: Condensed Matter*, 16(6):891–902, 2004.

[34] J.H. Harding; D.M. Duffy. The challenge of biominerals to simulations. *Journal of Materials Chemistry*, 16(12):1105–1112, 2006.

[35] Jeffrey P. Fitts; Xiaoming Shang; George W. Flynn; Tony F. Heinz; Kenneth B. Eisenthal. Electrostatic surface charge at aqueous/α-Al_2O_3 single crystal interfaces as probed by optical second-harmonic generation. *Journal of Physical Chemistry B*, 109(16):7981–7986, 2005.

[36] Nicholas J. Reeves; John Spencer Evans. Polypeptide interactions at ice and biomineral interfaces are defined by secondary structure-dependent chain orientations. *Journal of Physical Chemistry B*, 101(34):6665–6669, 1997.

[37] Yaohua Dai; John Spencer Evans. An energy based mapping method for identifying the in-plane orientations of polypeptides and other macromolecules at crystalline interfaces. *Journal of Chemical Physics*, 112(11):5144–5157, 2000.

[38] Paul Peter Ewald. Die berchnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 64:253–287, 1921.

[39] Ronald M. Levy; Linda Y. Zhang; Emilio Gallicchio; Anthony K. Felts. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *Journal of the American Chemical Society*, 125(31):9523–9530, 2003.

[40] Amol A. Mungikar; Daniel Forciniti. Conformational changes of peptides at solid/liquid interfaces: A Monte Carlo study. *Biomacromolecules*, 5(6):2147–2159, 2004.

[41] Christian M. Cortis; Richard A. Friesner. An automatic three-dimensional finite-element mesh generation system for the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 18(13):1570–1590, 1997.

[42] Christian M. Cortis; Richard A. Friesner. Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *Journal of Computational Chemistry*, 18(13):1591–1608, 1997.

[43] Shlomit R. Edinger; Christian Cortis; Peter S. Shenkin; Richard A. Friesner. Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *Journal of Physical Chemistry B*, 101(21):1190–1197, 1997.

[44] Michael Schaefer; C. Froemmel. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *Journal of Molecular Biology*, 216(4):1045–1066, 1990.

[45] Giuseppina Raffaini; Fabio Ganazzoli. Simulation study of the interaction of some albumin subdomains with a flat graphite surface. *Langmuir*, 19(8):3403–3412, 2003.

[46] Giuseppina Raffaini; Fabio Ganazzoli. Molecular dynamics simulation of the adsorption of a fibronectin module on a graphite surface. *Langmuir*, 20(8):3371–3378, 2004.

[47] Laurent David; Ray Luo; Michael K. Gilson. Comparison of Generalized Born and Poisson models: Energetics and dynamics of HIV protease. *Journal of Computational Chemistry*, 21(4):295–309, 2000.

[48] Michael K. Gilson. Theory of electrostatic interactions in macromolecules. *Current Opinion in Structural Biology*, 5(2):216–223, 1995.

[49] Tomasz Grycuk. Deficiency of the Coulomb field approximation in the Generalized Born model: An improved formula for Born radii evaluation. *Journal of Chemical Physics*, 119(9):4817–4826, 2003.

[50] D.E. Morse; M.A. Cariolou; G.D. Stucky; C.M. Zaremba; P.K. Hansma. Genetic coding in biomineralization of microlaminate composites. *Materials Research Society Proceedings*, 292:59–67, 1993.

[51] Ellen F. Sawilowsky; Oussama Meroueh; H. Bernhard Schlegel; William L. Hase. Structures, energies, and electrostatics for methane complexed with alumina clusters. *Journal of Physical Chemistry A*, 104(21):4920–4927, 2000.

[52] Ryan Y. Jin; Kihyung Song; William L. Hase. Molecular dynamics simulations of the structures of alkane/hydroxylated α-Al_2O_3(0001) interfaces. *Journal of Physical Chemistry B*, 104(12):2692–2701, 2000.

[53] Kim Bolton; Sylvie B.M. Bosio; William L. Hase; William F. Schneider; Kenneth C. Hass. Comparison of explicit and united atom models for alkane chains physisorbed on α-Al_2O_3 (0001). *Journal of Physical Chemistry B*, 103(19):3885–3895, 1999.

[54] Scott Brown; Teresa Head-Gordon. Cool walking: A new Markov chain Monte Carlo sampling method. *Journal of Computational Chemistry*, 24(1):68–76, 2003.

[55] W. Clark Still; Anna Tempczyk; Ronald C. Hawley; Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16):6127–6129, 1990.

[56] Robert B. Hermann. Theory of hydrophobic bonding. II. the correlation of hydrocarbon solubility in water with solvent cavity surface area. *Journal of Physical Chemistry*, 76(19):2754–2759, 1972.

[57] Terrell L. Hill. *An Introduction to Statistical Thermodynamics.* Dover Publications, Inc., New York, New York, 1986.

[58] Axel Arnold; Christian Holm. MMM1D: A method for calculating electrostatic interactions in one-dimensional periodic geometries. *Journal of Chemical Physics*, 123(14):144103, 2005.

[59] Anthony Nicholls; Barry Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 12(4):435–445, 1991.

[60] Doree Sitkoff; Kim A. Sharp; Barry Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *Journal of Physical Chemistry*, 98(7):1978–1988, 1994.

[61] Doree Sitkoff; Nir Ben-Tal; Barry Honig. Calculation of alkane to water solvation free energies using continuum solvent models. *Journal of Physical Chemistry*, 100(7):2744–2752, 1996.

[62] Kim A. Sharp; Barry Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *Journal of Physical Chemistry*, 94(19):7684–7692, 1990.

[63] Kim A. Sharp; Barry Honig. Electrostatic interactions in macromolecules: Theory and applications. *Annual Review of Biophysics and Biophysical Chemistry*, 19:301–332, 1990.

[64] Machael K. Gilson; Kim A. Sharp; Barry H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*, 9(4):327–335, 1987.

[65] Vinod K. Misra; Kim A. Sharp; Richard A. Friedman; Barry Honig. Salt effects on ligand-DNA binding: Minor groove binding antibiotics. *Journal of Molecular Biology*, 238(2):245–263, 1994.

[66] Walter Rocchia; Emil Alexov; Barry Honig. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *Journal of Physical Chemistry B*, 105(28):6507–6514, 2001.

[67] Water Rocchia; Sundaram Sridharan; Anthony Nicholls; Emil Alexov; Alessandro Chiabrera; Barry Honig. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Application to the molecular systems and geometric objects. *Journal of Computational Chemistry*, 23(1):128–137, 2002.

[68] Z.C. Feng; T.R. Yang; Y.T. Hou. Infrared reflectance analysis of GaN epitaxial layers grown on sapphire and silicon substrates. *Materials Science in Semiconductor Processing*, 4(6):571–576, 2001.

[69] Brian N. Dominy; Charles L. Brooks III. Development of a Generalized Born model parameterization for proteins and nucleic acids. *Journal of Physical Chemistry B*, 103(18):3765–3773, 1999.

[70] Michael Feig; Alexey Onufriev; Michael S. Lee; Wonpil Im; David A. Case; Charles L. Brooks III. Performance comparison of Generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry*, 25(2):265–284, 2004.

[71] Michael S. Lee; Freddie S. Salsbury Jr.; Charles L. Brooks III. Novel Generalized Born methods. *Journal of Chemical Physics*, 116(24):10606–10614, 2002.

[72] Simon Mitternacht; Stefan Schnabel; Michael Bachmann; Wolfhard Janke; Anders Irbäck. Differences in solution behavior among four semiconductor-binding peptides. *Journal of Physical Chemistry B*, 111(17):4355–4360, 2007.

[73] Yu-Jane Sheng; Heng-Kwong Tsao; Jian Zhou; Shaoyi Jiang. Orientation of a Y-shaped biomolecule adsorbed on a charged surface. *Physical Review E*, 66(1):011911, 2002.

[74] Jonathan W. Essex; Daniel L. Severance; Julian Tirado-Rives; William L. Jorgensen. Monte Carlo simulations for proteins: Binding affinities for trypsin-benzamidine complexes via free-energy perturbations. *Journal of Physical Chemistry B*, 101(46):9663–9669, 1997.

[75] D.M. Basalyga; R.A. Latour Jr. Theoretical analysis of adsorption thermodynamics for charged peptide residues on SAM surfaces of varying functionality. *Journal of Biomedical Materials Research*, 64(1):120–130, 2003.

[76] Randall T. Cygan; Jian-Jie Liang; Andrey G. Kalinichev. Molecular models of hydroxide, oxyhydride, and clay phases and the development of a general force field. *Journal of Physical Chemistry B*, 108(4):1255–1266, 2004.

[77] Michael Schaefer; Martin Karplus. A comprehensive analytical treatment of continuum electrostatics. *Journal of Physical Chemistry*, 100(5):1578–1599, 1996.

[78] Atsushi Harada; Sandrine Cammas; Kazunori Kataoka. Stabilized α-helix structure of poly(L-lysine)-block-poly(ethylene glycol) in aqueous medium through supramolecular assembly. *Macromolecules*, 29(19):6183–6188, 1996.

[79] John G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *Journal of Chemical Physics*, 2(7):351–361, 1934.

[80] David A. Kofke. On the acceptance probability of replica-exchange Monte Carlo trials. *Journal of Chemical Physics*, 117(15):6911–6914, 2002.

[81] Yuko Hori; Ashok Prasad; Jane Kondev. Stretching short biopolymers by fields and forces. *Physical Review E*, 75(4):041904, 2007.

[82] Stephen T. Kottmann. Harmonic averaging of smooth permittivity functions in finite-difference Poisson-Boltzmann electrostatics. *Theoretical Chemistry Accounts - In press*, 2007.

[83] Yu Sun; Robert A. Latour. Comparison of implicit solvent models for the simulation of protein-surface interactions. *Journal of Computational Chemistry*, 27(16):1908–1922, 2006.

[84] D. Wen; R.A. Laursen. A model for binding of an antifreeze polypeptide to ice. *Biophysical Journal*, 63(6):1659–1662, 1992.

[85] John Lekner. Summation of dipolar fields in simulated liquid-vapour interfaces. *Physica A*, 157(2):826–838, 1989.

[86] John Lekner. Summation of Coulomb fields in computer-simulated disordered systems. *Physica A*, 176(3):485–498, 1991.

[87] Charles M. Roth; Abraham M. Lenhoff. Electrostatic and van der Waals contributions to protein adsorption: Computation of equilibrium constants. *Langmuir*, 9(4):962–972, 1993.

[88] E. Gallicchio; M. Kubo; R.M. Levy. Enthalpy-entropy and cavity decomposition in alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *Journal of Physical Chemistry B*, 104(26):6271–6285, 2000.

[89] Emilio Gallicchio; Ronald M. Levy. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *Journal of Computational Chemistry*, 25(4):479–499, 2004.

[90] Yu Huang; Xiangfeng Duan; Yi Cui; Lincoln J. Lauhon; Kyoung-Ha Kim; Charles M. Lieber. Logic gates and computation from assembled nanowire building blocks. *Science*, 294(5545):1313–1317, 2001.

[91] Andrey V. Kajava; Steven E. Lindow. A model of the three-dimensional structure of ice nucleation proteins. *Journal of Molecular Biology*, 232(3):709–717, 1993.

[92] H. Tang; W. Zhang; P. Geng; Q.J. Wang; L.T. Jin; Z.R. Wu; M. Lou. A new amperometric method for rapid detection of *Escherichia coli* density using a self-assembled monolayer-based bienzzyme biosensor. *Analytica Chimica Acta*, 562(2):190–196, 2006.

[93] D. Erion; Atul Agarwal; Vellarkad N. Viswanadhan; D. Quentin McDonald; W. Clark Still M. Rami Reddy; Mark. Solvation free energies calculated using the GB/SA model: Sensitivity of results on charge sets, protocols, and force-fields. *Journal of Computational Chemistry*, 19(7):769–780, 1998.

[94] B.M. Ataev; V.V. Mamedov; A.K. Omaev; B.A. Magomedov. Epitaxial ZnO films grown by RF-assisted low-temperature CVD method. *Materials Science in Semiconductor Processing*, 6(5):535–537, 2003.

[95] Martial Mazars. Lekner summations. *Journal of Chemical Physics*, 115(7):2955–2965, 2001.

[96] Brock A. Luty; Malcolm E. Davis; J. Andrew McCammon. Electrostatic energy calculations by a finite-difference method: Rapid calculation of charge-solvent interaction energies. *Journal of Computational Chemistry*, 13(6):768–771, 1992.

[97] Jessica M.J. Swanson; Jason A. Wagoner; Nathan A. Baker; J. Andrew McCammon. Optimizing the Poisson dielectric boundary with explicit solvent forces and energies: Lessons learned with atom-centered dielectric functions. *Journal of Chemical Theory and Computation*, 3(1):170–183, 2007.

[98] Jessica M.J. Swanson; Stewart A. Adcock; J. Andrew McCammon. Optimized radii for Poisson-Boltzmann calculations with the AMBER force-field. *Journal of Chemical Theory and Computation*, 1(3):484–493, 2005.

[99] M. Zacharias; Brock A. Luty; Malcolm E. Davis; J. Andrew McCammon. Combined conformational search and finite-difference Poisson-Boltzmann approach for flexible docking. *Journal of Molecular Biology*, 238(3):455–465, 1994.

[100] Malcolm E. Davis; J. Andrew McCammon. Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods. *Journal of Computational Chemistry*, 10(3):386–391, 1989.

[101] Malcolm E. Davis; J. Andrew McCammon. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: An approach to improve accuracy and convergence. *Journal of Computational Chemistry*, 12(7):909–912, 1991.

[102] Michael K. Gilson; Malcolm E. Davis; Brock A. Luty; J. Andrew McCammon. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *Journal of Physical Chemistry*, 97(14):3591–3600, 1993.

[103] Nathan A. Baker; Michael J. Holst; J. Andrew McCammon. The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM Journal of Research and Development*, 45(3):427–438, 2001.

[104] F. Fogolari; G. Esposito; P. Viglino; H. Molinari. Molecular mechanics and dynamics of biomolecules using a solvent continuum model. *Journal of Computational Chemistry*, 22(15):1830–1842, 2001.

[105] Federico Fogolari; Alessandro Brigo; Henriette Molinari. Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophysical Journal*, 85(1):159–166, 2003.

[106] A.M. Belcher; X.H. Wu; R.J. Christensen; P.K. Hansma; G.D. Stucky; D.E. Morse. Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature*, 381(6577):56–58, 1996.

[107] Bernd A. Berg; Thomas Neuhas. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68(1):9–12, 1992.

[108] H.H. Goldstine; F.J. Murray; J.von Neumann. The Jacobi method for real symmetric matrices. *Journal of the Association for Computing Machinery*, 6(1):59–96, 1959.

[109] Barry Honig; Anthony Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–1149, 1995.

[110] J.Andrew Grant; Barry T. Pickup; Anthony Nicholls. A smooth permittivity function for Poisson-Boltzmann solvation methods. *Journal of Computational Chemistry*, 22(6):608–640, 2001.

[111] Takao Yoda; Yuji Sugita; Yuko Okamoto. Secondary-structure preferences of force-fields for proteins evaluated by generalized-ensemble simulations. *Chemical Physics Letters*, 307(2):269–283, 2004.

[112] Abdallah Sayyed-Ahmad; Kagan Tuncay; Peter J. Ortoleva. Efficient solution technique for solving the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 25(8):1068–1074, 2004.

[113] Y. Wang; L. Zhang; C. Liang; G. Wang; X. Peng. Catalytic growth and photoluminescence properties of semiconductor single-crystal ZnS nanowires. *Chemical Physics Letters*, 357(3):314–318, 2002.

[114] Jillian F. Banfield; Susan A. Welch; Hengzhong Zhang; Tamara Thomsen Ebert; R. Lee Penn. Aggregation-based crystal growth and microstructure development in natural iron oxyhydroxide biomineralization products. *Science*, 289(5480):751–754, 2000.

[115] J. Andrew Grant; Barry T. Pickup. A Gaussian description of molecular shape. *Journal of Physical Chemistry*, 100(6):3503–3510, 1996.

[116] Péter Pulay. Convergence acceleration of iterative sequences. the case of SCF iterations. *Journal of Chemical Physics*, 73(2):393–398, 1980.

[117] R.E. Johnson; S. Ranganathan. Ewald sum for electronic bilayer systems. *Physical Review E*, 63(5):056703, 2001.

[118] Jeremy Schofield; Mark A. Ratner. Monte Carlo methods for short polypeptides. *Journal of Chemical Physics*, 109(20):9177–9191, 1998.

[119] B. Lee; Frederic M. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55(3):379–400, 1971.

[120] Peter J. Eng; Thomas P. Trainor; Gordon E. Brown Jr.; Glenn A. Waychunas; Matthew Newville; Stephen R. Sutton; Mark L. Rivers. Structure of the hydrated α-Al_2O_3 (0001) surface. *Science*, 288(5468):1029–1033, 2000.

[121] Mafalda Nina; Dmitri Beglov; Benoit Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *Journal of Physical Chemistry B*, 101(26):5239–5248, 1997.

[122] Wonpil Im; Dmitri Beglov; Benoit Roux. Continuum solvation model: Computation of electrostatic forces from numerical solution to the Poisson-Boltzmann equation. *Computer Physics Communications*, 111(1):59–75, 1998.

[123] Axel Ruhe. SOR-methods for the Eigenvalue problem with large sparse matrices. *Mathematics of Computation*, 28(127):695–710, 1974.

[124] Michael Holst; Faisal Saied. Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 14(1):105–113, 1993.

[125] Supuma Sinha; Joseph Samuel. Inequivalence of statistical ensembles in single molecule measurements. *Physical Review E*, 71(2):021104, 2005.

[126] Thomas P. Trainor; Peter J. Eng; Gordon E. Brown Jr.; Ian K. Robinson; Maurizio De Santis. Crystal truncation rod difraction study of the α-Al_2O_3 $(1\bar{1}02)$ surface. *Surface Science*, 496(3):238–250, 2002.

[127] Ersin Emre Oren; Candan Taerler; Mehmet Sarikaya. Metal recognition of septapeptides via polypod molecular architecture. *Nanoletters*, 5(3):415–419, 2005.

[128] Kenneth D. Gibson; Harold A. Scheraga. Exact calculation of the volume and surface area of fused hard-sphere molecules with unequal atomic radii. *Molecular Physics*, 62(5):1247–1265, 1987.

[129] Vincent Noinville; Claire Vidal-Madjar; Bernard Sébille. Modeling of protein adsorption on polymer surfaces. computation of adsorption potential. *Journal of Physical Chemistry*, 99(5):1516–1522, 1995.

[130] Kim A. Sharp. Incorporating solvent and ion screening into molecular dynamics using the finite-difference Poisson-Boltzmann method. *Journal of Computational Chemistry*, 12(4):454–468, 1991.

[131] Ninad V. Prabhu; Peijuan Zhu; Kim A. Sharp. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *Journal of Computational Chemistry*, 25(16):2049–2064, 2004.

[132] Robert E. Bruccoleri; Jiri Novotny; Malcolm E. Davis; Kim A. Sharp. Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *Journal of Computational Chemistry*, 18(2):268–276, 1997.

[133] John F. Marko; Eric D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995.

[134] Henrik G. Petersen; D. Soelvason; J.W. Perram; E.R. Smith. Error estimates for the fast multipole method. I. the two-dimensional case. *Proceedings of the Royal Society of London A*, 448(1934):389–400, 1995.

[135] Arald Jean-Charles; Anthony Nicholls; Kim Sharp; Barry Honig; Anna Tempczyk; Thomas F. Hendrickson; W. Clark Still. Electrostatic contributions to solvation energies: Comparison of free energy perturbation and continuum calculations. *Journal of the American Chemical Society*, 113(4):1454–1455, 1991.

[136] H. Sun. COMPASS: An *ab Initio* force-field optimized for condensed-phase applications-overview with details on alkane and benzene compounds. *Journal of Physical Chemistry B*, 102(38):7338–7364, 1998.

[137] S. Ravichandran; J.D. Madura; J. Talbot. A Brownian dynamics study of the initial stages of hen egg-white lysozyme adsorption at a solid interface. *Journal of Physical Chemistry B*, 105(17):3610–3613, 2001.

[138] Nicholas Metropolis; Arianna W. Rosenbluth; Marshall N. Rosenbluth; Augusta H. Teller; Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[139] William L. Jorgensen; David S. Maxwell; Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[140] William L. Jorgensen; Julian Tirado-Rives. The OPLS potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[141] A.H. Juffer; P. Argos; J. De Vlieg. Adsorption of proteins onto charged surfaces: A Monte Carlo approach with explicit ions. *Journal of Computational Chemistry*, 17(16):1783–1803, 1996.

[142] Andre H. Juffer; Craig M. Shepherd; Hans J. Vogel. Protein-membrane electrostatic interactions: Application of the Lekner summation technique. *Journal of Chemical Physics*, 114(4):1892–1905, 2001.

[143] A.T. Clark; T.J. Madden; P.B. Warren. Summation of electrostatic interactions in quasi-two-dimensional simulations. *Molecular Physics*, 87(5):1063–1069, 1996.

[144] R. Webster. *Phage Display: A Laboratory Manual.* Cold Spring Harbor Lab Press, Cold Spring, NY, 2001.

[145] H.A. Lowenstam; S. Weiner. *On Biomineralization.* Oxford University Press, New York, New York, 1989.

[146] Charles A. Knight; Andrzej Wierbicki; Richard A. Laursen; Wei Zhang. Adsorption of biomolecules to ice and their effects upon ice growth. 1. measuring adsorption orientations and initial results. *Crystal Growth and Design*, 1(6):429–438, 2001.

[147] Harianto Tjong; Huan-Xiang Zhou. GBr^6: A parameterization free, accurate, analytical Generalized Born method. *Journal of Physical Chemistry B*, 111(11):3055–3061, 2007.

[148] Harianto Tjong; Huang-Xiang Zhou. The dependence of electrostatic solvation energy on dielectric constants in Poisson-Boltzmann calculations. *Journal of Chemical Physics*, 125(20):206101, 2006.