THERMODYNAMIC MODELING OF PROTEIN INTERACTIONS AND PHASE BEHAVIOR

by

Leigh Jian Quang

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Chemical Engineering

Spring 2012

THERMODYNAMIC MODELING OF PROTEIN INTERACTIONS AND PHASE BEHAVIOR

by
Leigh Jian Quang

Approved: _____
Abraham M. Lenhoff, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Stanley I. Sandler, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Norman J. Wagner, Ph.D.
Chair of the Department of Chemical Engineering

Approved: _____
Babatunde A. Ogunnaike, Ph.D.
Interim Dean of the College of Engineering

Approved: _____

Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

There were many people who have helped me during my time at the University of Delaware. My experience as a graduate student was perhaps the most humbling period of my life, but the support from my peers was instrumental in making this work possible. I am grateful and honored to have had the opportunity to collaborate with so many diligent and exceptional people in the chemical engineering department. I will always remember the unique camaraderie among everyone in the graduate student community.

First, I would like to thank the members of the Lenhoff and Sandler groups. I would like to acknowledge my advisers Professor Abraham Lenhoff and Professor Stan Sandler for their critique of this thesis. I would like to thank Dr. Rachael Lewus for her helpful experimental insights in protein phase behavior. She was able to answer many of my conceptual questions during the early stages of my work, which helped me tremendously in getting the big picture of my project. I also thank Dr. Jaeoon Chang for discussions on the patch-antipatch model for proteins and general aspects of molecular simulation. His down-to-earth personality and innate ability to explain complicated concepts in simple terms helped me tremendously in learning the technical aspects of simulation. I would like to thank Dr. Chris Gillespie who provided me his source code for the atomistic B_{22} calculations. I like to acknowledge the support from Steve Traylor, Anvar Samadzoda, and Dr. Kelley Kearns. They were always there to discuss not only the technical aspects of my research, but also a wide variety of nontechnical topics ranging from sports to politics.

It was always refreshing to be able to talk about things outside of research. Sometimes it is easy to forget that there is a world beyond graduate school. Russ Burnett and Vinit Choudhary were both helpful for their constructive feedback on my work. Although my project was outside of their area of expertise, they were still able to provide me with some of their perspectives for my work.

This work would not have been possible without support from outside of the chemical engineering department. I am grateful for the National Science Foundation for providing the funding for this project. I also thank Dr. Sandeep Patel in the Department of Chemistry for providing the computational resources that allowed me to perform the B_{22} calculations.

Finally, I would like to acknowledge the support from my fellow classmates at the University of Delaware. I would like to give special thanks to Nima Nikbin and Stephen Edie for putting up with me during my last two months in Delaware (thanks guys I really appreciate it!).

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Protein phase behavior encompasses the formation of dense phases, which include amorphous aggregates, gels, dense liquids, and crystals. The major solution variables that dictate the type of dense phase that is formed are pH, temperature, type of precipitant, precipitant concentration, and protein concentration. Because of the large parameter space and rich variety of phase transitions possible, protein phase behavior is a complex phenomenon. Fundamentally, macroscopic phase transitions are governed by the molecular interactions between proteins in solution. One promising way of quantifying protein-protein interactions and relating them to phase behavior is through the osmotic second virial coefficient B_{22}, a dilute-solution property that characterizes two-particle interactions. The relationship of B_{22} to overall phase behavior of proteins is explored in this work.

The goal of this thesis is to quantitatively relate protein-protein interactions to protein phase diagrams in order to develop predictive models of phase behavior under different solution conditions. A continuum-level approach is used initially to relate experimental B_{22} data and phase diagrams of proteins by appealing to existing thermodynamic models, with the expectation that a simple continuum model could provide a useful mechanistic framework for predicting protein phase behavior. The first approach attempted was to relate protein interactions and phase behavior within the Flory-Huggins theory of polymer solutions. The second approach utilized the model of Haas and Drenth, which is based on the free energy of mixing for hard spheres. Finally, phase equilibrium was predicted from virial coefficients using the

osmotic virial equation. A qualitative relationship was found between B_{22} and phase behavior from these continuum models; however, quantitative agreement could not be obtained. The isotropic assumption shared among these models in addition to the orientationally-averaged nature of B_{22} suggests that the anisotropic character of protein interactions cannot be neglected, demonstrating the need for more detailed molecular-level models.

The role of anisotropy in protein interactions was explored through analysis of "patch-antipatch" pairs in the computation of B_{22} in atomistic detail. Patch-antipatch pairs represent highly attractive orientations resulting from geometric complementarity between protein surfaces. Previous work used simple Monte Carlo integration for the calculation of B_{22} from atomistic models of proteins. However, the presence of patch-antipatch pairs led to significant numerical concerns. These concerns warranted a reexamination of the numerical methods for computing B_{22}.

A hybrid Monte Carlo/patch integration approach is utilized to calculate B_{22} for lysozyme and chymosin B. This method involves a combination of numerical integration techniques in an attempt to obtain better convergence in predicting B_{22}. The overall B_{22} for the proteins studied was separated into three components: contributions from the excluded volume, from the patch-antipatch pairs, and from background configurations. The excluded volume component was found to be adequately determined using simple Monte Carlo integration. The contributions from individual patch-antipatch pairs were accounted for by carefully integrating the subregions of the configuration space occupied by these pairs using a globally adaptive integration routine. The background component to B_{22} was also calculated by

simple Monte Carlo integration in which the regions of the configuration space occupied by the patch-antipatch pairs were excluded.

The calculations performed that account for the full protein structure emphasize the importance of several features of protein interactions. First, the difference in the interaction behavior of the two proteins studied was found to be largely attributed to the charge anisotropy of patch-antipatch pairs. However, the relation of the results to experimental data is limited by the omission of accounting for the specific hydration of proteins. Hydration effects are known to affect, and usually attenuate, patch-antipatch configurations, and therefore would be expected to significantly impact the accurate prediction of B_{22}. Classical colloidal as well as atomistic models that omit these important features are inadequate in providing a quantitative representation of protein interactions for a wide range of solution conditions.

Chapter 1

INTRODUCTION

Protein phase behavior refers to the appearance of various condensed phases that proteins can form in solution. It encompasses the formation of amorphous aggregates, gels, dense liquids, and crystals. The types of dense phases formed are sensitive to the solution conditions, which include pH, temperature, type of precipitant, precipitant concentration, and protein concentration. Because of the large parameter space and rich variety of phase transitions possible, protein phase behavior is a complex phenomenon.

The phase behavior of proteins plays a key role in many biopharmaceutical processes. Undesired phase separation of protein therapeutics can occur during processing and storage, which can raise serious efficacy and safety concerns. However, phase separation may be desired in other situations, and operations such as precipitation can be employed in downstream protein purification processes because of their low cost. One specific form of phase separation of wide interest is protein crystallization, which is a prerequisite step for determining protein structure by x-ray diffraction; however, the growth of high-quality crystals is often the bottleneck. Crystallization is also of interest in the pharmaceutical industry because it offers an advantageous way for delivery of doses of highly concentrated protein therapeutics. Phase separation of proteins can have serious medical implications as well. The onset of neurodegenerative diseases such as Alzheimer's disease has been attributed to the formation of fibrillar protein aggregates called amyloids.

Understanding the conditions and mechanisms that lead to formation of protein dense phases from solution is therefore crucial to predicting and ultimately controlling phase transitions in these diverse systems. Such knowledge will be invaluable in a wide variety of applications. This work aims to contribute towards this endeavor by utilizing and evaluating quantitative models that construct the pathway from molecular structure to the thermodynamic solutions properties of proteins.

1.1 The Protein Phase Diagram

Early experimental studies of liquid-liquid phase separation of globular proteins have led to interpretion of protein phase behavior within the framework of colloidal systems. Ishimoto and Tanaka first presented evidence of temperature-induced liquid-liquid phase separation of aqueous lysozyme solutions (1). These results were questioned by Phillies (2), but Taratuta et al. confirmed them by performing extensive cloud-point temperature measurements at different pH values, salt types, and salt concentrations (3). It was later observed that the liquid-liquid phase separation was actually metastable with respect to the solid-liquid transition (4), behavior characteristic of colloidal particles experiencing short-ranged attractions (5, 6). Similar phase behavior was also observed experimentally for several different γ-crystallins (7–10). The similarities in the metastability of the liquid-liquid phase separation for both proteins and colloids suggest that protein phase behavior follows the same physics as the phase transitions of colloids.

The phase diagram of proteins can be theoretically interpreted within the context of the theoretical phase diagram for colloidal systems experiencing short-ranged interactions (5, 6) (Figure 1.1). The colloidal phase diagram is most directly compared to experimental protein phase behavior results when presented in a two-

dimensional plane where the colloid concentration is plotted on the ordinate and the reduced temperature on the abscissa. The direction of decreasing temperature corresponds to increasing interparticle attraction, a trend that is qualitatively analogous to increasing precipitant concentration under salting-out conditions.

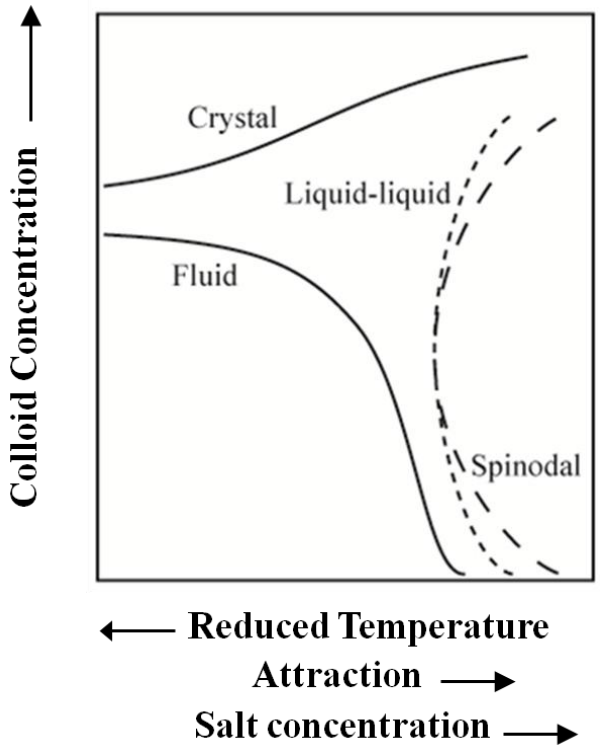

Figure 1.1: Theoretical colloidal phase diagram adapted from Foffi et al. (6).

The main feature of this phase diagram is a solid-liquid equilibrium region, which corresponds to protein crystals in equilibrium with a protein-lean supernatant fluid. Within this region lies the metastable liquid-liquid coexistence envelope (bounded by the binodal curve) in which the protein partitions into a dilute and more concentrated liquid phases. Inside this liquid-liquid coexistence region is

the spinodal boundary beyond which phase separation occurs instantaneously through spinodal decomposition. Thus, knowing the relative locations of the phase boundaries as a function of the solution conditions is important in navigating through the phase diagram.

The metastable nature of the liquid-liquid phase separation is an important feature of protein phase behavior. A correlation exists between the metastable liquid-liquid coexistence and crystallization (11). Liquid-liquid phase separation has been suggested to significantly change the kinetic pathway for crystal nucleation (12, 13). Specifically, the free energy barrier for crystal nucleation is drastically reduced near the liquid-liquid critical point due to critical density fluctuations. Therefore, knowing the location of the liquid-liquid coexistence region can have implications for selecting optimal solution conditions for protein crystallization.

1.2 Protein-Protein Interactions

Protein phase behavior is governed fundamentally by the molecular interactions between protein molecules, which are still not completely understood. What is known is that the interactions include contributions from van der Waals forces, electrostatics, hydration effects (including hydrophobic interactions), and depletion effects (where relevant). Attractive van der Waals forces arise from three different contributions: permanent dipole-induced dipole interactions (Debye interaction), permanent dipole-permanent dipole interactions (Keesom interaction), and induced dipole-induced dipole interactions (London dispersion interaction) (14, 15). These attractions are complemented by long-ranged electrostatic interactions due to the charges that some amino acid residues carry on the protein surface.

Solvation forces, which are associated with water structuring around the protein surface, include effects that may be classified as hydrophobic or as hydration effects (15, 16). The hydrophobic effect results from the presence of nonpolar patches on the surface, with which water molecules are unable to form hydrogen bonds. To minimize the free energy, the nonpolar regions associate with one another, driving water molecules away from the surface to more extensive formation of hydrogen bonds in the bulk of the solution. Hydration effects may occur in hydrophilic regions on the solute molecules, where hydrogen bonding with adjacent water molecules may result in a steric and hence effectively repulsive barrier to association with other solute molecules. Solvation forces are still poorly understood and a quantitative explanation is presently lacking.

Additional depletion attraction is induced due to the osmotic pressure gradient caused by the addition of nonadsorbing polymers such as polyethylene glycol (PEG) (17). Depletion interactions are a result of an entropic effect; when two protein molecules are in sufficient proximity, the polymer-excluded volumes of the molecules overlap. Consequently, the polymer molecules cannot penetrate into the space between the protein molecules, resulting in effective attraction due to the osmotic pressure of the polymer in the bulk solution.

In addition to the different contributions to protein-protein interactions, an important feature is also their anisotropy, which is a consequence of the nonspherical shape of the protein molecule and the heterogeneous properties of the protein surface. The shape anisotropy of proteins has a profound effect on the van der Waals attractions. Computations have shown that shape anisotropy has an appreciable effect

on the magnitude and orientational distribution of van der Waals interaction energies compared to calculations based on the ideal sphere approximation (18).

There is a complex interplay among the different forces that govern protein-protein interactions and solution conditions. For example, at low salt concentrations protein interactions are dominated by long-ranged electrostatic forces, which are usually repulsive. At high concentrations of salt, electrostatic forces are screened and short-ranged van der Waals forces and hydrophobic interactions tend to drive the precipitation of protein from solution. The phenomenon where proteins become less soluble as more salt is added is known as salting-out. One explanation for this behavior is that the salt ions alter the hydrogen bonding network of the layer of water that shields the protein surface hydration layer (19). Consequently, protein molecules interact less with water, resulting in an increase in the protein-protein interactions. Different ions have been found to have varying impacts on protein phase behavior. The salting-out effectiveness of different ions is reflected in the Hofmeister series (20), which is an empirical ranking of the ability of different ions to disrupt the hydration layer.

A quantitative understanding of the underlying mechanism of specific ion effects on phase behavior is still incomplete, but much progress has been made in the past fifteen years (21–26). It was long believed that the Hofmeister effect was due to the influence of the ion on the hydrogen bonding network of bulk water, but experimental results suggest that this influence does not extend beyond the first hydration shell (27–29). Rather, direct ion-protein interactions appear to contribute significantly to the ability of a specific ion to salt out proteins from solution.

Two recent developments have emerged that have led to the development of salt-specific models of protein interactions. The first cites significant contribution from dispersion forces between ions and protein molecules (24, 30–33). The polarizabiltity of the ion is the unique physical characteristic that determines the ion's specificity. Including ion dispersion contributions in calculating protein-protein interactions has been shown to qualitatively capture the reverse Hofmeister behavior exhibited by lysozyme (34, 35), where the salting-out trend follows the opposite order of the Hofmeister ranking. These findings are consistent with experimental observations (36).

The second development emphasizes the role of solvent-assisted ion binding to the hydrophobic regions of the protein surface. Efforts to model this effect have used molecular dynamics simulations of ions and hydrophobic interfaces to determine the effective interactions between proteins (37–40). Models of protein interactions incorporating effects of ion binding have also yielded predictions that are qualitatively consistent with the salting-out behavior for lysozyme (41).

1.3 Theoretical and Simulation Studies

The interactions between protein molecules play a central role in defining the macroscopic thermodynamic properties of protein solutions. The key to relating the microscopic and macroscopic properties is the potential of mean force (PMF). The PMF represents the effective interaction between two molecules in a system of *n* molecules as a combination of the direct interactions between the molecules and indirect forces from the other species, which include the solvent and ions (42, 43). Another way of interpreting the PMF is that it is the free energy required to bring two molecules in an *n* body system from infinite separation to a particular configuration in

solution. In principle, specifying the PMF fixes the thermodynamic properties of the system. However, the intermolecular interactions of proteins are quite complex and are strongly dependent on the solution conditions. Therefore, it is not possible at present to determine an exact PMF model of protein-protein interactions. An alternative approach is to start with simple idealized models with a few parameters based on the physics of the system, and fit the parameters to correlate experimental data.

Experimental evidence of a metastable liquid-liquid phase transition in proteins has led to theoretical efforts to relate the phase behavior and intermolecular interactions of proteins within the framework of model colloids (44, 45). It is well known from theory (46), experiment (47), and simulation that the phase behavior of colloidal dispersions is sensitive to the range of the interaction between the particles. Colloidal particles experiencing an attraction that is short-ranged with respect to its diameter exhibit a metastable liquid-liquid phase transition. Since proteins exhibit similar phase behavior, idealized colloidal models can serve as a starting point for describing the thermodynamics of protein solutions.

Early theoretical work attempted to approximate protein solutions as a one-component system and model the effective interactions between proteins with a simple form of the pair potential. Within this framework, the solvent is treated as a continuum background and the protein molecules are represented as hard spheres experiencing attraction that is a function only of the center-to-center distances (Figure 1.2).

Several simple potentials have been used to model protein interactions, including the adhesive hard sphere (48–51), square-well (52–55), Yukawa (6, 56, 57),

and modified Lennard-Jones (12, 58) potentials. While the forms of these potentials are different, each of these models has parameters for the interaction strength, interaction range, and particle diameter. In these models, the interaction range is the dominant parameter that controls the shape and location of the phase boundaries (58). Large values of the range parameter lead to phase diagrams that have a stable vapor-liquid phase separation, which is analogous to a liquid-liquid phase transition for the colloidal system. As the range parameter becomes sufficiently small, this transition becomes metastable with respect to the solid-liquid phase transition. In addition, it has been shown that these isotropic potential models follow an extended law of corresponding states (59). The consequence of this is that the thermodynamic phase behavior of systems that interact through short-ranged attraction becomes insensitive to the details of the interaction potential if they are scaled by the proper parameters.

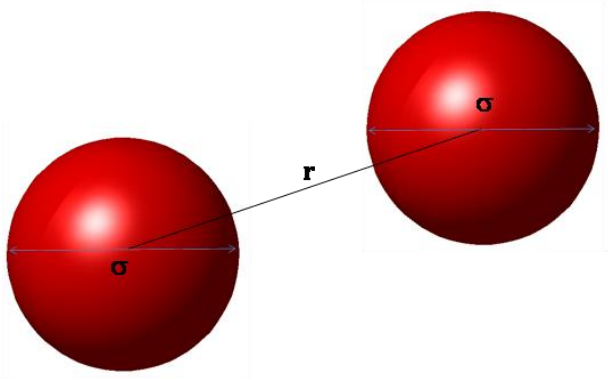

Figure 1.2: Cartoon of simple isotropic model of proteins. Protein molecules are represented as perfect spheres of diameter σ and the interactions depend only on the center-to-center distance *r*.

A more complex formulation that has been used to characterize colloidal interactions and has been applied to proteins comes from DLVO theory (14, 15, 60,

61). In this theory the particles are modeled as ideal spheres with a homogeneous charge distribution that interact via short-ranged van der Waals attraction and long-ranged Coulombic repulsion. The solvent is assumed to be a structureless continuum with a uniform dielectric constant. The potential function that reflects this framework is the sum of three contributions

$$u_{DLVO} = u_{HS} + u_{vdW} + u_{elec} \qquad 1.1$$

where u_{HS} is the hard-sphere excluded volume contribution to the potential, u_{vdW} is the contribution from the van der Waals attraction, and u_{elec} is the contribution from the electrostatic repulsion. Within the DLVO framework, the salt screens the electrostatic interaction between protein molecules, thereby reducing the repulsive electrostatics. The advantage of DLVO theory is that the model parameters can be directly related to the solution conditions and physical properties of the system. The van der Waals contribution is characterized by the Hamaker constant and the size and the separation distance between the two spheres. The Coulombic repulsion is governed by pH, which determines the net charge of the protein, and the solution ionic strength, as well as the size and separation distance.

Studies using simple isotropic models have provided qualitative insight into the relationship between interactions and the phase behavior of proteins. Specifically, the metastable liquid-liquid transition exhibited by protein solutions can be explained in terms of attraction that is short-ranged relative to the size of the protein. The phase diagrams from these simple models are in qualitative agreement with experimental observation; however, the model parameters cannot always be physically related to the solution conditions (pH, temperature, ionic strength), which are known to determine the protein-protein interactions and therefore phase behavior

(62, 63). Consequently, the parameters of these simple models cannot be determined *a priori* and can only be used to fit experimental data (62, 63). In addition, the metastable critical point for protein systems has been shown to be sensitive to the solution conditions, and this sensitivity cannot generally be captured by spherically symmetric potentials (64, 65). Thus, simple intermolecular potentials can only be used as empirical models and cannot be used to predict protein phase behavior for a wide range of solution conditions.

The phase diagrams predicted by the DLVO model qualitatively correlate the experimental phase behavior data for lysozyme and γ-crystallin (66–68), but the model is unable to quantitatively predict phase behavior that agrees with experiment. There are several problems with DLVO theory that limits its predictive capability for the phase behavior of proteins. First, the model does not account for other important solvation forces that are known to be significant, such as hydrophobic interactions and hydration effects. Omission of these forces is one of the reasons that DLVO theory failed to describe phase behavior for some proteins such as apoferritin (69) and hemoglobins (HbS and HbA) even qualitatively (70). In addition, DLVO theory does not properly take into account specific ion effects because the theory treats ions as point charges in solving the Poisson-Boltzmann equation (32, 60). Consequently, the model cannot explain the varying salting-out abilities of different ions at high salt concentrations. Another limitation of this theory is that it is not capable of explaining the salting-in behavior of some proteins. This discrepancy is due to the assumption of a uniform charge distribution, which inherently treats the electrostatics as always repulsive. However, the distribution of charges carried by the titratable amino acids can lead to attractive electrostatic interactions. The screening of the attractive

electrostatic interactions with increasing salt concentration leads to increasing stability of the protein solution, resulting in salting-in behavior. Therefore, DLVO theory provides an incomplete description of protein-protein interactions, and cannot be expected to provide quantitatively accurate predictions of protein phase behavior.

More complex models that go beyond spherically symmetric potentials have been used to predict the phase behavior of proteins. These models emphasize different features of protein-protein interactions and have predicted phase behavior with varying degrees of success. An embedded charge model has been used to account for the charge anisotropy of proteins (71), but the phase diagrams predicted from this representation were found to agree only qualitatively with experiment (72). One class of models that have been used for colloids and have been used to represent the anisotropy of the short-ranged attractions of proteins are patch models (73–78). Patch models represent protein molecules as spheres that carry attractive regions on the surface to account for orientationally local strong interactions. Patch models have been shown to provide more accurate quantitative representations of protein phase diagrams (73, 75, 79).

1.4 The Osmotic Second Virial Coefficient: B_{22}

One method for characterizing the effective protein-protein interactions is through the osmotic second virial coefficient, B_{22}. B_{22} is a dilute solution property that is a measure of effective two-body interactions in solution, and it provides a link to the PMF via the statistical mechanical expression for B_{22} which, accounting for orientation dependence is given as (42, 80)

$$B_{22} = -\frac{1}{16MW^2\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty}\left(e^{-W/kT}-1\right) \times r_{12}^2 dr_{12}\sin\theta d\theta d\phi d\alpha \sin\beta d\beta d\gamma \qquad 1.2$$

Here W is the PMF, r_{12} is the center-to-center distance, ϕ and θ are the spherical angles representing the location of the second molecule relative to the first, and α, β, γ are the Euler angles denoting the rotation of the second molecule (81).

Extensive computations explicitly accounting for the full structural details of proteins have shown that due to the Boltzmann weighting of the PMF, a few highly attractive complementary configurations contribute disproportionately to the calculation of B_{22} (82, 83). That protein-protein interactions can be dominated by a few highly attractive configurations has been demonstrated experimentally by a single point mutation of a crystal contact for T4 lysozyme (84). These attractive configurations that control bulk solution properties are characteristic of molecular recognition stemming from the geometric complementarity of apposing regions. These attractive regions can serve as contacts for crystal formation, which suggest a plausible correlation with crystallization (85) and protein phase behavior (45, 86–88). For example, when B_{22} is positive, the protein molecules on balance repel one another and remain stable in solution. As B_{22} becomes negative, the protein interactions are net attractive and may lead to the formation of condensed phases. The region of slightly negative B_{22} values known as the crystallization slot (-1×10^{-4} to -8×10^{-4} mol ml/g^2) was identified empirically by George and Wilson to be conducive to the formation of protein crystals (88). If B_{22} is too negative, the strong attractions may prevent the protein molecules from rearranging and forming the specific contacts that lead to a crystalline lattice, resulting in amorphous precipitates and gels.

1.5 Objective and Thesis Outline

The objective of this thesis is to quantitatively relate protein-protein interactions to protein phase diagrams in order to develop predictive models of protein phase behavior at different solution conditions. The motivation stems from the need for providing a rational methodology for the design and optimization of bioseparation processes. Developing such rational strategies can be significantly aided by knowledge of the phase diagram. Experimental measurements of phase diagrams for proteins is a nontrivial task; it can be expensive in terms of time, labor, and supply of protein due to the difficulty of characterizing the various dense phases, time to attain true equilibrium, and the wide range of possible solution conditions. In addition, if crystallization conditions are not known for a protein, measurement of the solid-liquid phase boundaries is not possible since crystals are needed. Consequently, complete phase diagrams have been measured for only a few proteins (89). Therefore, developing predictive models of protein phase behavior is essential and can have significant industrial and scientific benefits.

The following chapters aim to elucidate the path from molecular structure to the thermodynamic properties of proteins. Proteins can be represented at various levels of coarse-graining, from simple spheres to a full atomistic structure. The level of structural representation directly impacts the ability to model protein interactions (PMF), which ultimately allows the prediction of bulk solution properties such as B_{22} and phase behavior. The conceptual path from molecular structure to thermodynamic properties is illustrated in Figure 1.3. In this thesis, models that represent proteins at various levels are explored to evaluate their capability of providing the link between protein interactions and phase behavior.

In Chapter 2, an attempt is made to model protein solutions within the framework of existing continuum thermodynamic models that have been established for polymer and colloidal systems. This was done by quantitatively evaluating the relationship of measured B_{22} and phase behavior data for a model globular protein using these models. In addition, phase equilibrium is modeled from the osmotic virial equation derived from McMillan-Mayer solution theory. Chapter 3 focuses on the anisotropy of protein-protein interactions on the molecular level and how this feature impacts the prediction of B_{22}. The anisotropy arising from the shape complementarity between protein surfaces and the charge distribution is analyzed within the context of the "patch-antipatch" representation of protein interactions. Further, the numerical technique for computing B_{22} from atomistic descriptions of proteins is reexamined, and a new approach is proposed and outlined. This thesis is concluded by summarizing the findings from this work and recommendations are made for future directions.

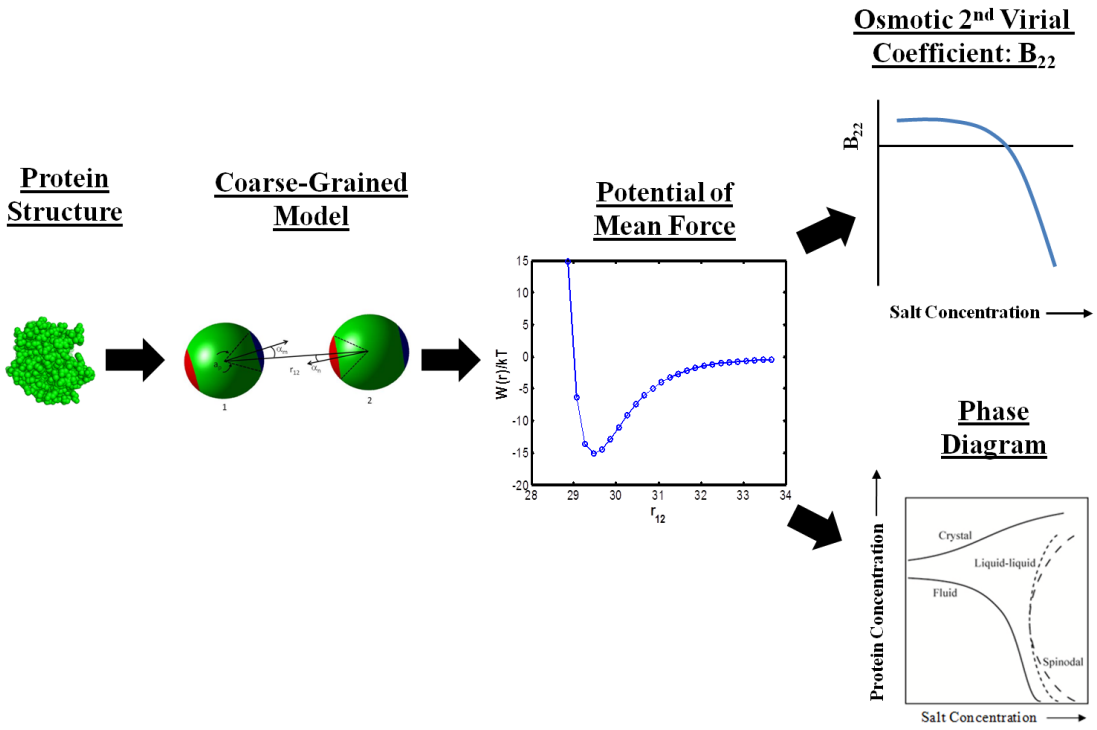

Figure 1.3: Schematic of the conceptual path from molecular structure to thermodynamic solution properties of proteins, which includes the osmotic second virial coefficient B_{22} and phase behavior.

Chapter 2

CONTINUUM THERMODYNAMIC MODELS OF PROTEIN INTERACTIONS AND PHASE BEHAVIOR

2.1 Motivation and Goal

Unlike previous studies that have taken a molecular approach to develop predictive models (12, 53, 56, 66–68), a continuum-level approach is explored here to relate protein-protein interactions and phase behavior. Previous work has demonstrated a correlation between the osmotic second virial coefficient, B_{22}, and phase diagrams found from experiments (45, 86, 87). Therefore, there is evidence to suggest that B_{22} may offer a reasonable quantitative measure of the effective interactions between protein molecules. That is, the evidence suggests that the overall effects of the solution conditions (pH, temperature, and precipitant concentration) on protein-protein interactions can collectively be represented by B_{22}. The aim of this chapter is to use experimental B_{22} data to quantitatively predict the phase diagrams of proteins by utilizing existing classical thermodynamic models. The hope is that a simple continuum model with few parameters can provide a useful mechanistic framework for predicting the phase behavior of protein solutions.

The models investigated in this work were the Flory-Huggins model (90), the Haas-Drenth model (91–93), and the osmotic virial equation (94). These models were used to calculate values of B_{22} and the phase behavior for ribonuclease A in ammonium sulfate solutions at pH 7 and 23°C and to compare the calculated values to experimental data obtained by Dumetz et al. (45, 86) (Figure 2.1). The B_{22} values for this system were measured using self-interaction chromatography (95, 96) and the

dilute boundary of the liquid-liquid coexistence region of the phase diagram was obtained by a microbatch technique. This particular system was chosen because the B_{22} values partially overlap the binodal curve over the range of 0.90 M to 1.25 M ammonium sulfate (Figure 2.1). It was therefore possible to make a direct correlation of B_{22} data with phase behavior within this salt range using the models listed above. It should be noted that B_{22} is often difficult to measure at salt concentrations in which phase separation is observed since the attractions are very strong. As a result, B_{22} data rarely overlap phase behavior data for proteins over the same range of salt concentration, if at all. This difficulty makes relating complementary sets of data with the above models challenging.

This chapter of the thesis is organized as follows. The modeling structure used for relating B_{22} and phase behavior within the framework of the continuum models is briefly described. For each model, the theoretical foundation is introduced and the equations that govern phase equilibrium are presented. These equations provide the modeling structure used for relating B_{22} and phase behavior. Next, results of the correlations between experimental B_{22} and phase behavior data for ribonuclease A for each of the models are presented and discussed. Finally, conclusions are drawn on the capability of these continuum models to relate protein interactions and phase behavior based on the results for ribonuclease A.

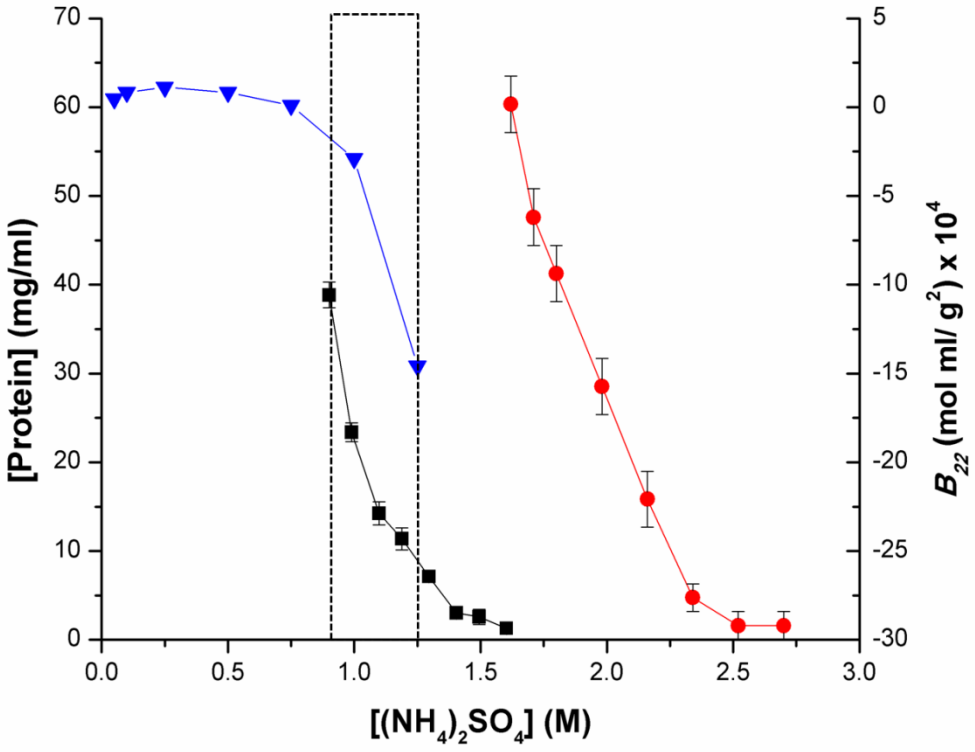

Figure 2.1: (■) Binodal, (●) spinodal, and (▼) B_{22} data for ribonuclease A in ammonium sulfate system at 23°C, pH 7. The dotted rectangle encloses the region where B_{22} and phase behavior data overlap. Results were taken from Dumetz et al. (45, 86)**.**

2.2 Modeling Structure

The physical parameters for ribonuclease A and the solvent (water) used for this study are presented in Table 2.1 and Table 2.2, respectively. The specific volume of ribonuclease A used in these calculations is a value generally used in the literature for globular proteins (91), which is v_2=0.735 ml/g.

Table 2.1: Physical properties of ribonuclease A used in the continuum models.

Property	Value	Ref
pI	9.6	(45)
MW (g/mol)	13700	(45)
σ (nm)	3.1	(86)
v_2 (ml/g)	0.735	(91)

Table 2.2: Physical properties of water used in the continuum models.

Property	Value
MW (g/mol)	18.02
ρ (g/ml)	0.998
$\underline{V}$ (ml/mol)	18.06

A schematic of the modeling structure that was followed for relating B_{22} and phase behavior using the three continuum models is shown in Figure 2.2. With the Flory-Huggins and Haas-Drenth models, the phase behavior can be predicted directly from B_{22} values at each salt concentration, or vice versa. In using phase behavior to calculate B_{22}, the larger salt range available in the phase behavior data can be taken advantage of to predict B_{22}. By following this path, B_{22} predictions can be made for salt concentrations beyond the experimental range of the B_{22} data. However, using measured B_{22} data to find phase behavior predictions is restricted to the window

of conditions for which both B_{22} and phase diagrams are available. Consequently, phase behavior calculations cannot be made at higher salt concentrations due to the lack of B_{22} data.

For the osmotic virial equation, only one approach was followed, in which experimental B_{22} data were used as inputs to predict the corresponding phase behavior. The reason this path was chosen is that B_{22} data are needed to compute higher virial coefficients necessary for the osmotic virial equation to relate interactions and phase behavior. The reverse approach of using experimental phase behavior data to calculate virial coefficients could not be utilized because of the need to calculate both B_{22} as well as at least one higher virial coefficient.

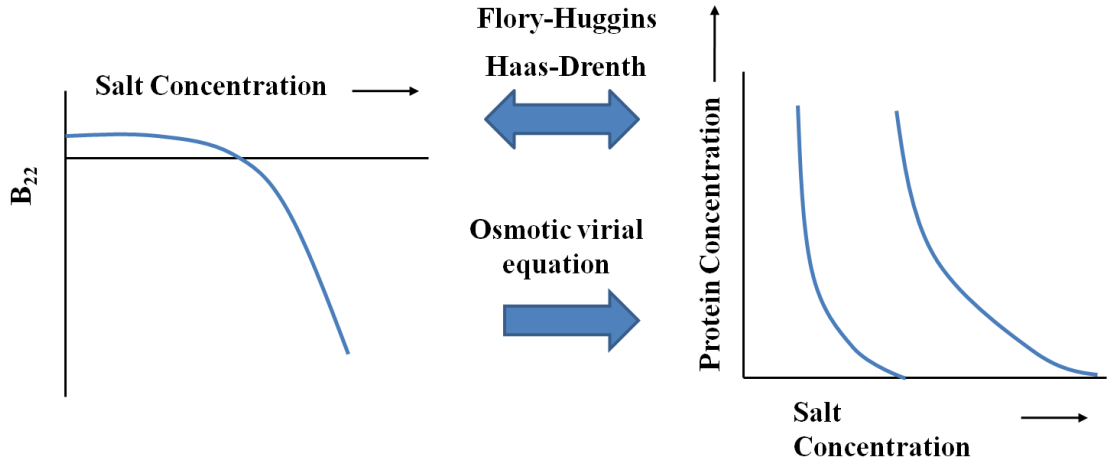

Figure 2.2: Schematic of the modeling pathways used to relate B_{22} and phase behavior with the continuum models.

2.3 Flory-Huggins Model

2.3.1 Theory

The Flory-Huggins model of polymer solutions is derived from simple lattice theory for fluids and has historically been used to predict the phase behavior of polymer-solvent systems (90). In this model, the system is considered to be composed of uniform lattice sites that can be occupied by either one solvent molecule or a polymer subunit (monomer). The polymer is represented as a linear, flexible chain of interconnected subunits in which the chain is free to adopt any configuration. Each monomer is allowed to occupy any one lattice site as long as the monomers remain interconnected. Because of its simplicity and few parameters, the Flory-Huggins model was deemed worth exploring. The excess free energy of the system has been derived as (97)

$$\frac{\underline{G}^{ex}}{RT} = \left[x_1 \ln\frac{\phi_1}{x_1} + x_2 \ln\frac{\phi_2}{x_2}\right] + \chi(x_1 + mx_2)\phi_1\phi_2 \qquad 2.2$$

where 1 and 2 refer to water and protein, respectively, x_i is the mole fraction of species i, and ϕ_i is the volume fraction of species i. The m parameter reflects the degree of polymerization of the polymer relative to the size of the solvent molecule and can be calculated as the ratio of the molar volumes of the polymer to the solvent

$$m = \frac{\underline{V_2}}{\underline{V_1}} \qquad 2.3$$

where $\underline{V_i}$ is the molar volume of species i. Based on the physical parameters listed in Tables 2.1 and 2.2, a value of 559 was set for the m parameter. The χ parameter is an adjustable parameter that represents the effective interaction between the solute and solvent.

For the Flory-Huggins model, it can be shown that liquid-liquid phase equilibrium is modeled by

$$\ln\left(\frac{1-\phi_2^I}{1-\phi_2^{II}}\right)+\left(1-\frac{1}{m}\right)(\phi_2^I-\phi_2^{II})+\chi[(\phi_2^I)^2-(\phi_2^{II})^2]=0 \qquad 2.4$$

$$\ln\left(\frac{\phi_2^I}{\phi_2^{II}}\right)+(m-1)[(1-\phi_2^I)-(1-\phi_2^{II})]+ m\chi[(1-\phi_2^I)^2-(1-\phi_2^{II})^2]=0 \qquad 2.5$$

The interaction parameter χ is linearly related to B_{22} by (14)

$$B_{22}=\frac{v_2^2}{\underline{V_1}}\left(\frac{1}{2}-\chi\right) \qquad 2.6$$

where v_2 is the specific volume of the protein in units of volume/mass. The spinodal region of the phase diagram can be determined by the criterion for thermodynamic stability given as (97)

$$\frac{\partial^2 \Delta \underline{G}_{mix}}{\partial {\phi_2}^2}=0 \qquad 2.7$$

From the criterion given by equation 2.7, the spinodal curve from the Flory-Huggins model is

$$\chi=\frac{\phi_2(m-1)+1}{2m\phi_2(1-\phi_2)} \qquad 2.8$$

The critical point of the phase diagram occurs at the maximum of the spinodal curve and therefore can be determined by setting the derivative of equation 2.8 with respect to the volume fraction ϕ_2 equal to zero. From this procedure it can be shown that the critical volume fraction and χ are related to the m parameter by

$$\phi_{2,critical}=\frac{1}{1+\sqrt{m}} \qquad 2.9$$

$$\chi_{critical} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{m}}\right)^2 \qquad 2.10$$

The critical B_{22} is obtained by substituting equation 2.10 into equation 2.6, which yields

$$B_{22,critical} = \frac{v_2^2}{2V_1}\left(\frac{2\sqrt{m}+1}{m}\right) \qquad 2.11$$

2.3.2 Results

The B_{22} predictions made from experimental phase behavior data using the Flory-Huggins model are presented in Figure 2.3. The model predictions correctly capture the qualitative trend of decreasing B_{22} values with increasing salt concentration, but the results differ quantitatively from the experimental values. The Flory-Huggins model predicts stronger protein-protein attractions than those reflected by the experimental measurements for the entire B_{22}-phase behavior overlap region (0.90 M to 1.25 M). In addition, the steep slope observed in the experimental B_{22} data is an important feature not captured by the model.

The reverse path in which phase behavior was predicted from experimental B_{22} data was also followed using the Flory-Huggins model. The model predicts phase separation to occur at higher salt concentrations, as suggested by its critical point, which is located at 1.22 M ammonium sulfate (Figure 2.4). The critical point represents the threshold for phase coexistence, with phase separation not observed at salt concentrations less than that at the critical point. The critical point predicted by the model indicates that phase separation occurs only at salt concentrations beyond the overlap region for which both B_{22} and phase behavior data are not available. As a result, binodal calculations could not be performed using the

model due to the lack of experimental B_{22} values at the higher salt concentrations. Despite this limitation, the location of the critical point suggests that the Flory-Huggins model does not predict phase equilibrium over the same salt range as the experimental data, and hence does not adequately describe the phase behavior.

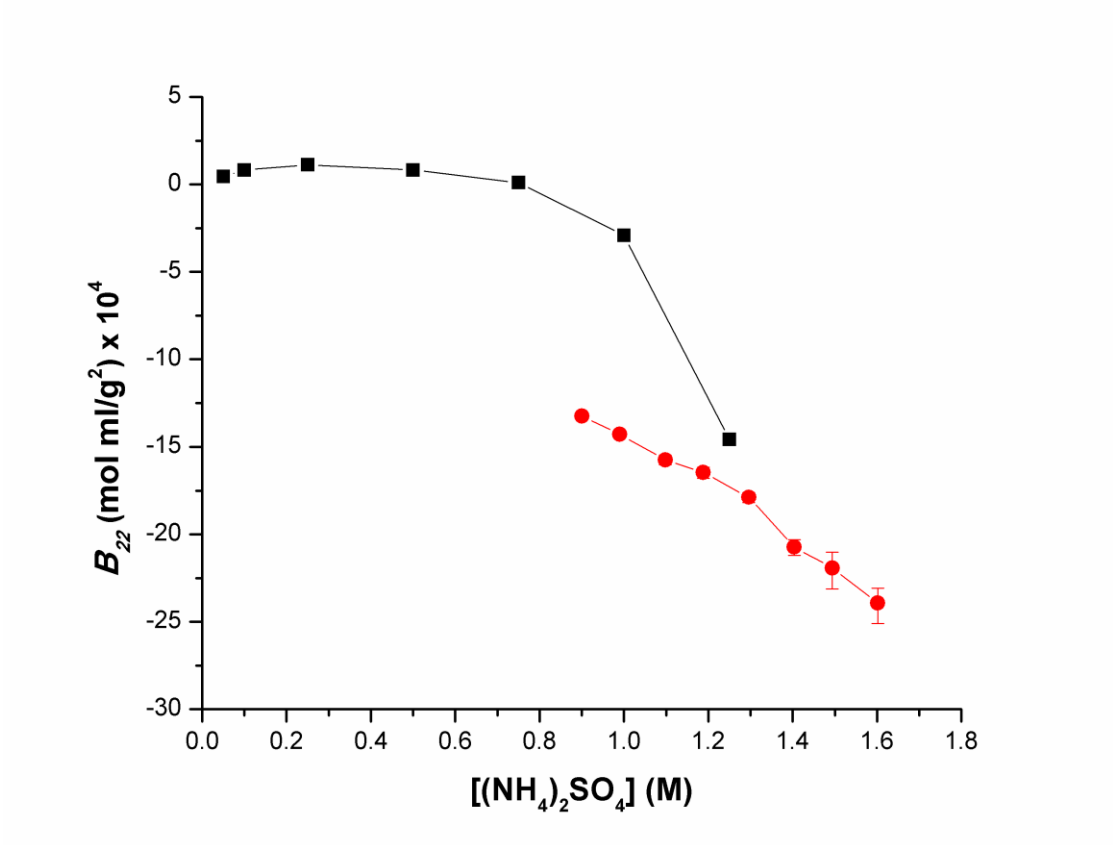

Figure 2.3: Comparison of B_{22} predictions from the (●) Flory-Huggins model with (■) experimental B_{22} data.

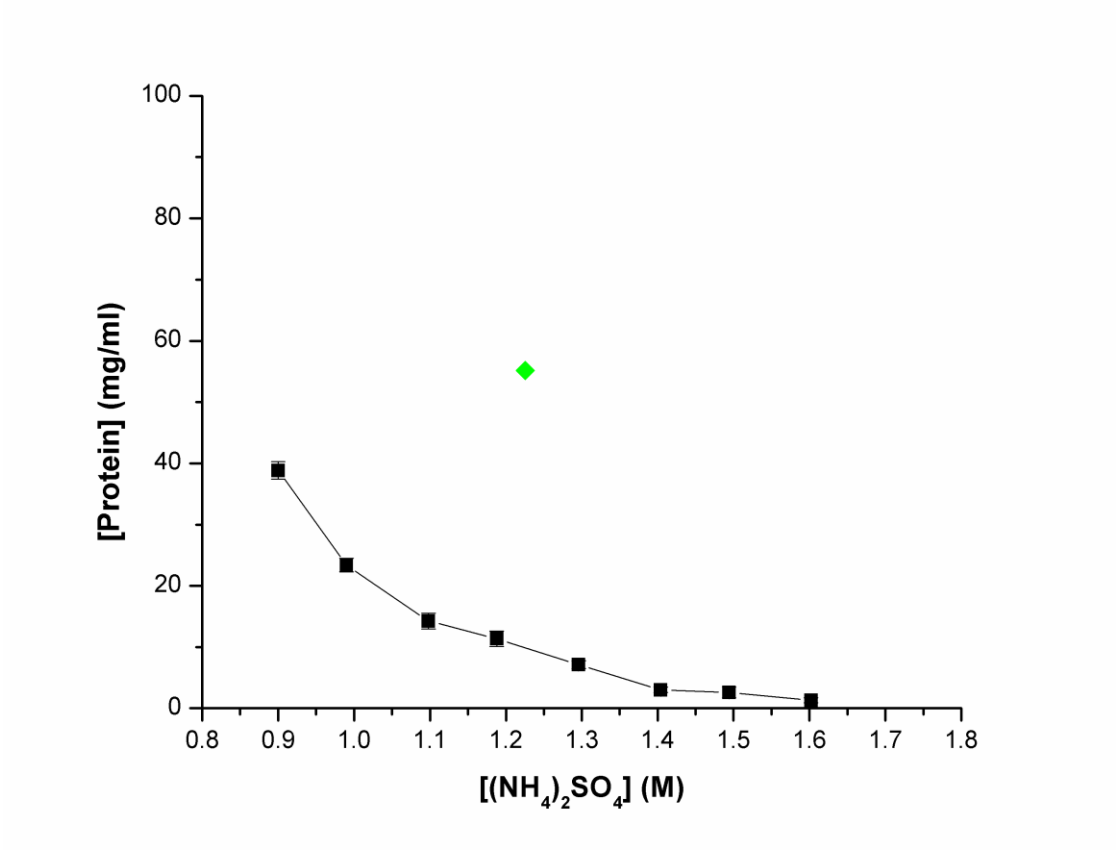

Figure 2.4: Critical point predicted by the Flory-Huggins model compared with (■) experimental binodal data. The critical point is located at a salt concentration of 1.22 M. The location of the critical point demonstrates that the equilibrium phase boundary is located at higher salt concentrations.

A sensitivity analysis was performed to probe the response of the B_{22} predictions to small perturbations in the Flory-Huggins model parameters. The parameter that was perturbed was the *m* parameter, which characterizes the size of the protein, which is directly related to the specific volume of the protein ν_2. The *m* parameter was adjusted by ±10% from the original value of *m*=559 used for ribonuclease A and B_{22} values were then recalculated from the model. The predicted

B_{22} values for the small adjustments in m are presented in Figure 2.5. Perturbing the m parameter causes a shift in the predictions; decreasing m causes the B_{22} predictions to shift to higher values whereas increasing m causes them to shift to lower values. However, tuning the m parameter does not lead to a significant change in the slope of the B_{22} predictions. Thus, based on the results from the sensitivity analysis, there does not appear to be a value of m that would lead to B_{22} predictions that match the experimental data more convincingly.

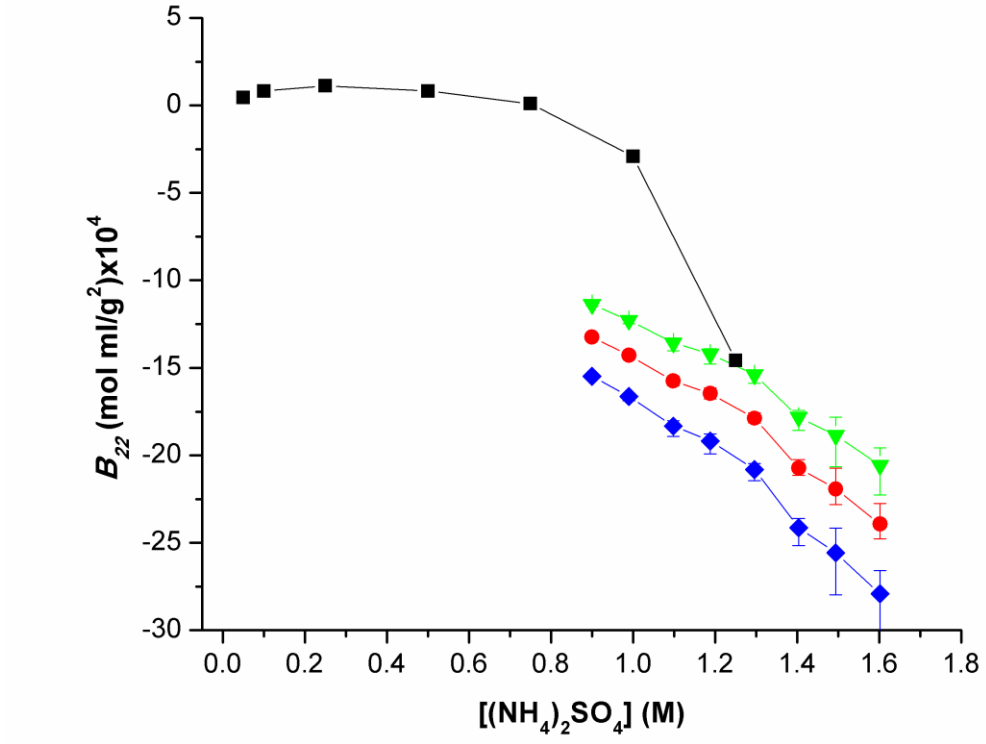

Figure 2.5: B_{22} predictions from Flory-Huggins model calculated from values of (▼) m=503 and (◆) m=615 compared with original predictions from (●) m= 559 and (■) experimental B_{22} values.

The effect on the phase behavior predicted from B_{22} data using adjusted values of the *m* parameter was also investigated. As previously mentioned, the model yielded a critical point that was located at higher salt concentrations than the experimental data suggest. For the Flory-Huggins model, the critical point location depends solely on the value of the *m* parameter, as shown in equations 2.9 and 2.10. Figure 2.6 shows the effect of changing *m* by $\pm$10% from the original value (*m*=559) on the location of the critical point. Increasing *m* shifts the critical point towards higher salt concentrations and lower protein concentrations whereas decreasing *m* shifts the critical point towards lower salt concentrations and higher protein concentrations. Based on this trend, *m* should be decreased in order for phase separation to be predicted at the lower salt concentrations found in the experimental data. The *m* parameter was subsequently decreased by 50% of the original value used, to a value of 279. The predicted binodal curve was calculated since phase coexistence shifted to lower salt concentrations, for which B_{22} interaction data were available (Figure 2.7). However, it does not appear that there is a physically reasonable value of *m* that would lead to predicted phase behavior from B_{22} data that matches the experimental results. Consequently, extension of this analysis to lower *m* values was not performed.

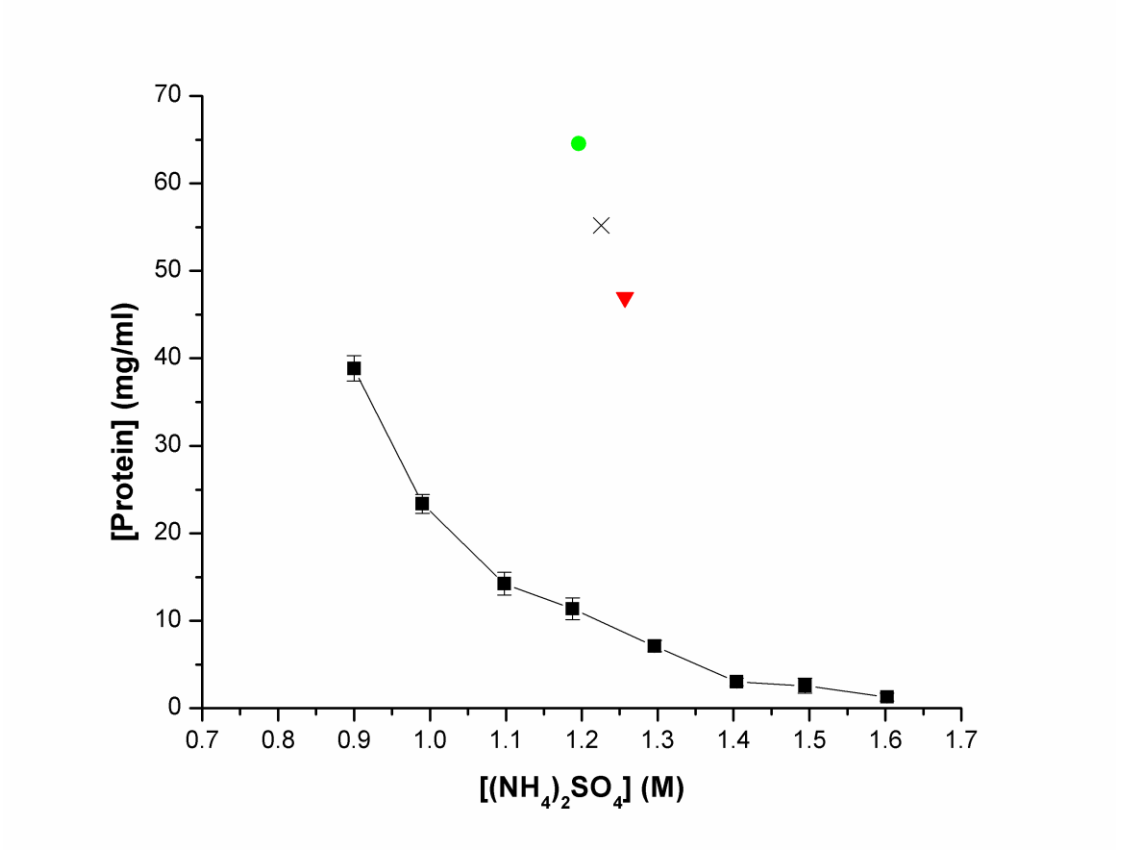

Figure 2.6: Predicted critical point from the Flory-Huggins model for (●) *m*=503, (×) *m*=559, and (▼) *m*=603 relative to the (■) experimental binodal boundary.

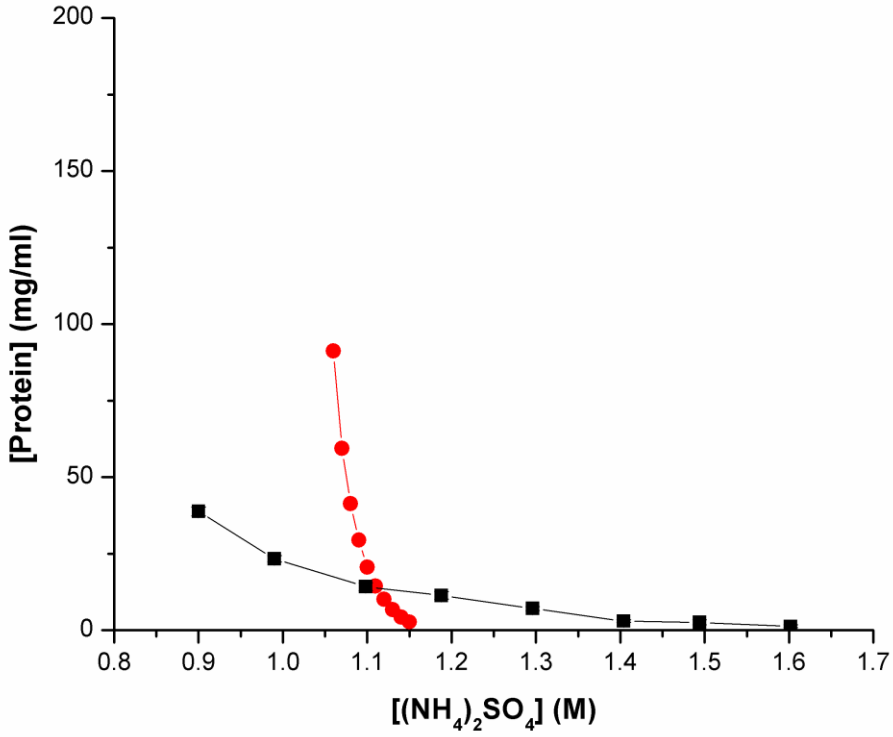

Figure 2.7: Predicted binodal boundaries from the Flory-Huggins model for (●) *m*=279 compared with (■) experimental results.

2.4 Haas-Drenth Model

2.4.1 Theory

The Haas-Drenth model is based on the free energy of mixing for hard spheres in a solvent and has been used to describe the protein-water phase diagram. The free energy per unit volume in this model is given as (91–93, 98)

$$G(\phi) = \frac{1}{\Omega}\left[\left(\frac{\phi_2^2}{\phi_c}\right) g_\lambda + kT\phi_2 \ln\frac{\phi_2}{m} - kT\left\{\frac{\phi_2 - 6\phi_2^2 + 4\phi_2^3}{(1-\phi_2)^2}\right\}\right] \qquad 2.12$$

where Ω is the molecular volume of the protein, *m* represents the size of the protein relative to the solvent molecule, *k* is the Boltzmann constant, *T* is the absolute temperature, g_λ represents the interaction between protein molecules in solution, ϕ_2 is the volume fraction of protein, and ϕ_c is the protein volume fraction in the crystal (usually taken to be 0.50). Based on the physical parameters listed in Tables 2.1 and 2.2, a value of 559 was set for the *m* parameter. The first term of equation 2.12 represents the enthalpic contribution from protein-protein interactions. The second and third terms together represent the contribution from the entropy of mixing for hard spheres (14, 15). Liquid-liquid coexistence for this model is obtained from the two equilibrium conditions

$$G^I(\phi_2^I) - G^{II}(\phi_2^{II}) = \phi_2^I\left(\frac{\partial G^I}{\partial \phi_2^I}\right)_{\phi_2^{II}} - \phi_2^{II}\left(\frac{\partial G^{II}}{\partial \phi_2^{II}}\right)_{\phi_2^I} \qquad 2.13$$

$$\left(\frac{\partial G^I}{\partial \phi_2^I}\right)_{\phi_2^{II}} = \left(\frac{\partial G^{II}}{\partial \phi_2^{II}}\right)_{\phi_2^I} \qquad 2.14$$

The $g_\lambda$interaction parameter is linearly related to B_{22} by

$$B_{22} = \frac{\nu_2}{MW}\left(4 + \frac{g_\lambda}{kT\phi_c}\right) \qquad 2.15$$

where *MW* is the molecular weight of the protein.

The spinodal curve can be derived by applying the condition for thermodynamic stability given by equation 2.7 to the free energy model in equation 2.12. The critical point occurs at the maximum of the spinodal curve and therefore, by taking the derivative of the equation for the spinodal curve, it can be shown that the Haas-Drenth free energy model predicts a critical volume fraction and g_λ as

$$\phi_{critical} = 0.130 \tag{2.16}$$

$$g_{\lambda,critical} = -10.601kT\phi_c \tag{2.17}$$

The critical B_{22} is obtained by substituting equation 2.17 into equation 2.15, which yields

$$B_{22,critical} = \frac{6.601\nu_2}{MW} \tag{2.18}$$

2.4.2 Results

The B_{22} predictions made from experimental phase behavior data using the Haas-Drenth model are presented in Figure 2.8. Similar to the predictions from the Flory-Huggins model, the Haas-Drenth model predictions correctly capture the qualitative trend of decreasing B_{22} values with increasing salt concentration. However, the predictions are still quantitatively different from the experimental values. The model predicts values of B_{22} that are on the same order of magnitude as the data, but seems to underpredict the attractions at higher salt concentrations. In addition, the steep slope observed in the experimental B_{22} data is an important feature not captured by the model.

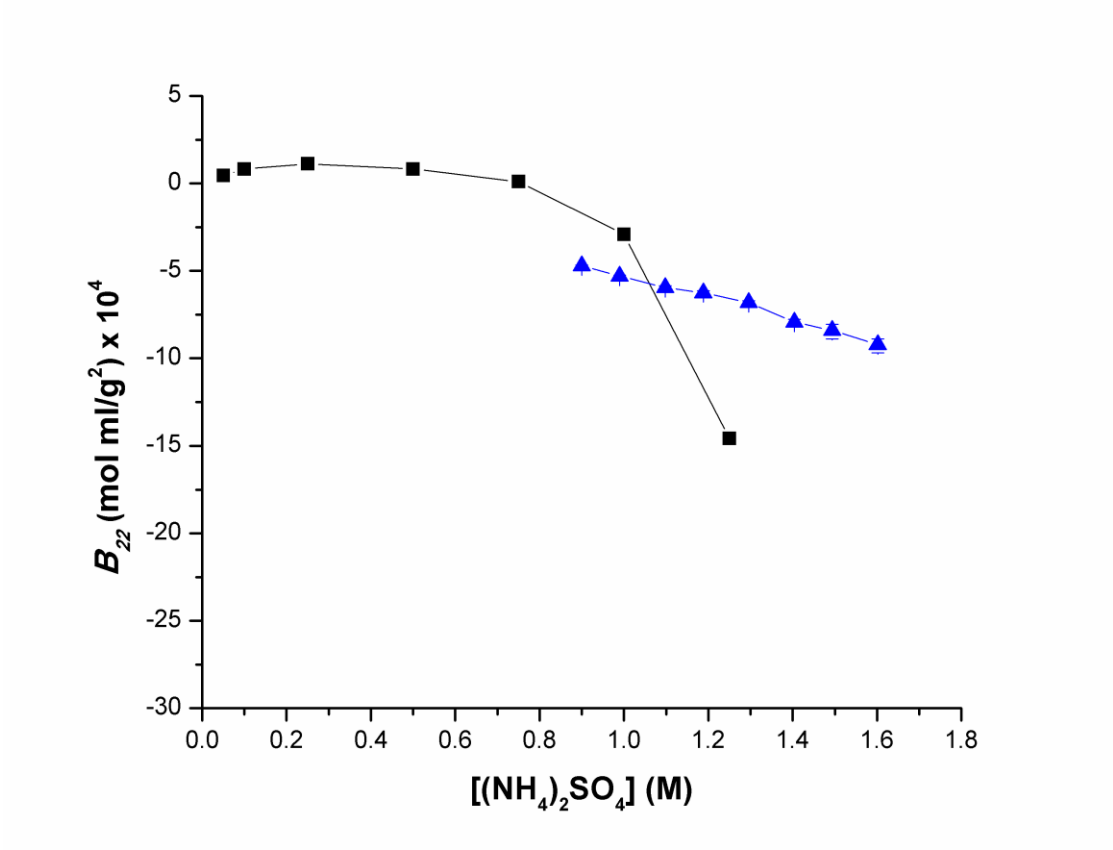

Figure 2.8: Comparison of B_{22} predictions from the (▲) Haas-Drenth model with (■) experimental B_{22} data.

To fully explore the prediction capability of the Haas-Drenth model, the reverse path in which phase behavior was predicted from experimental B_{22} data was also followed. The predicted equilibrium binodal boundary and critical point are presented in Figure 2.9. The Haas-Drenth model appears to provide a better description of the phase behavior than the Flory-Huggins model. Phase separation is predicted to occur in the small overlap salt range of the experimental data. The predicted binodal phase boundary also decays more sharply when compared with the experimental data. While the phase behavior results appear to be reasonable, the

experimental data show that the actual critical point occurs at a lower salt concentration than that predicted by the model, which is at 1.02 M ammonium sulfate. Therefore, it would seem that the Haas-Drenth model does not predict phase equilibrium over the same salt range as the experimental results.

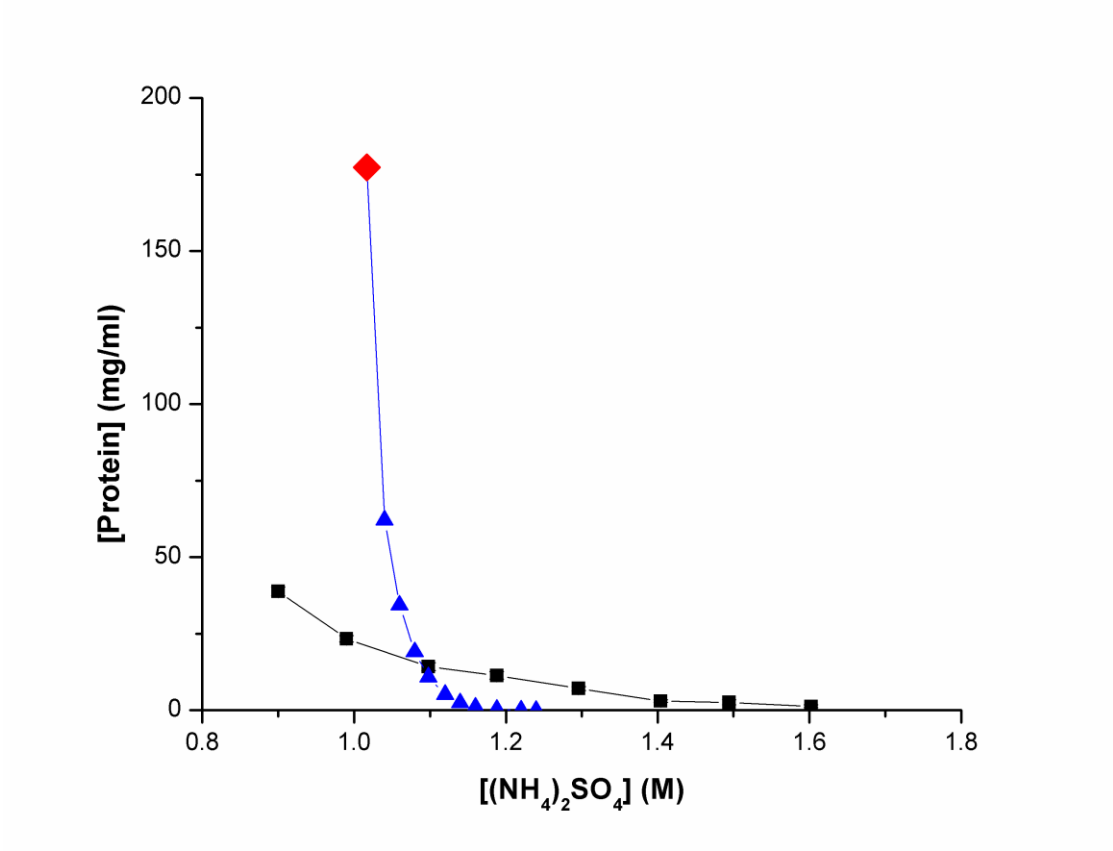

Figure 2.9: Phase behavior predictions from experimental B_{22} values with the Haas-Drenth model compared with (■) experimental binodal data. The Haas-Drenth model does predict (▲) the binodal boundary to exist within the overlap region. The predicted (◆) critical point occurs at an ammonium sulfate concentration of 1.02 M and a protein concentration of 177 mg/ml.

The sensitivity of the B_{22} predictions to perturbations in the model parameters was explored. The *m* parameter was adjusted by ±10% from the original

value of m=559 used for ribonuclease A and B_{22} values were then recalculated. The predicted B_{22} values for the adjusted values of m are presented in Figure 2.10. Similar to the Flory-Huggins model, decreasing m causes the B_{22} predictions to shift to higher values whereas increasing m causes the predictions to shift to lower values. However, the B_{22} predictions seem to be insensitive to small adjustments in m. In addition, tuning the m parameter does not lead to a change in the slope of the B_{22} predictions. Thus, there does not appear to be a value of m that would lead to B_{22} predictions that match the experimental data.

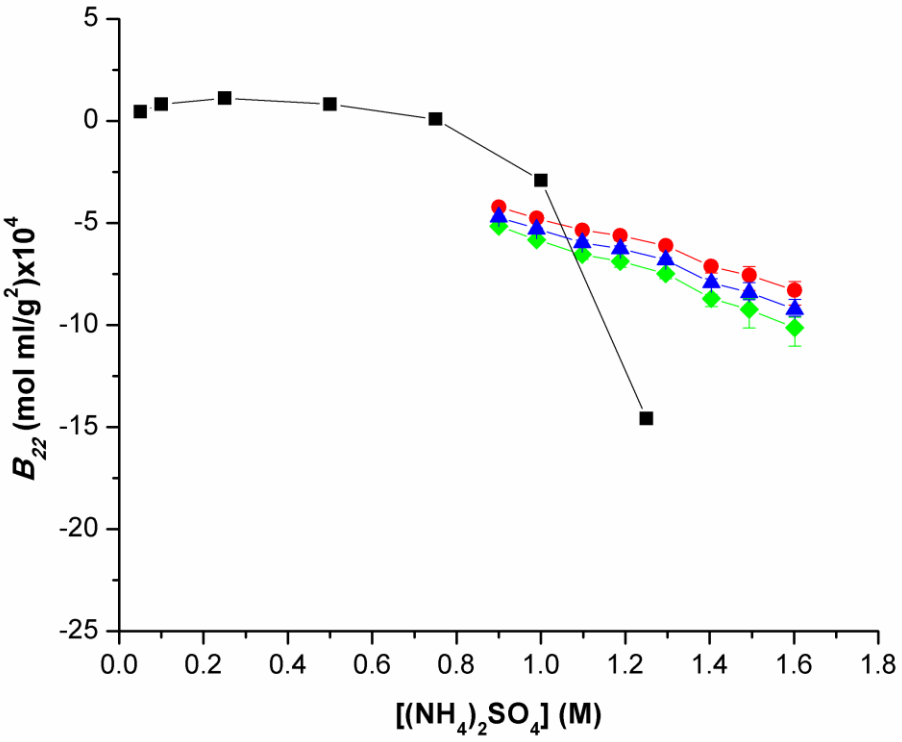

Figure 2.10: B_{22} predictions from the Haas-Drenth model calculated from values of (●) m=503 and (◆) m=615 compared with original predictions from (▲) m= 559 and (■) experimental B_{22} values.

The effect on the phase behavior predicted by the Haas-Drenth model for adjustments in the value of the *m* parameter was also explored. The location of the critical point of this model depends primarily on the specific volume of the protein v_2, which in turn is directly related to the *m* parameter. Increasing *m* shifts the critical point towards higher salt concentrations and lower protein concentrations, whereas decreasing *m* shifts the critical point towards lower salt concentrations and higher protein concentrations. Since the original calculations did not predict phase separation to occur at the lower salt concentrations seen experimentally, *m* had to be decreased in order to shift the predicted boundary in the direction of lower salt concentration. The *m* parameter was subsequently decreased by 50% of the original value used to a value of 279. The resulting binodal curve from this adjustment compared with the original prediction is shown in Figure 2.11. The adjustment did not shift the boundary significantly enough to capture the correct phase behavior over the entire salt range. Thus, it appears that there is no physically realistic value of *m* that would lead to phase behavior predictions consistent with the experimental data.

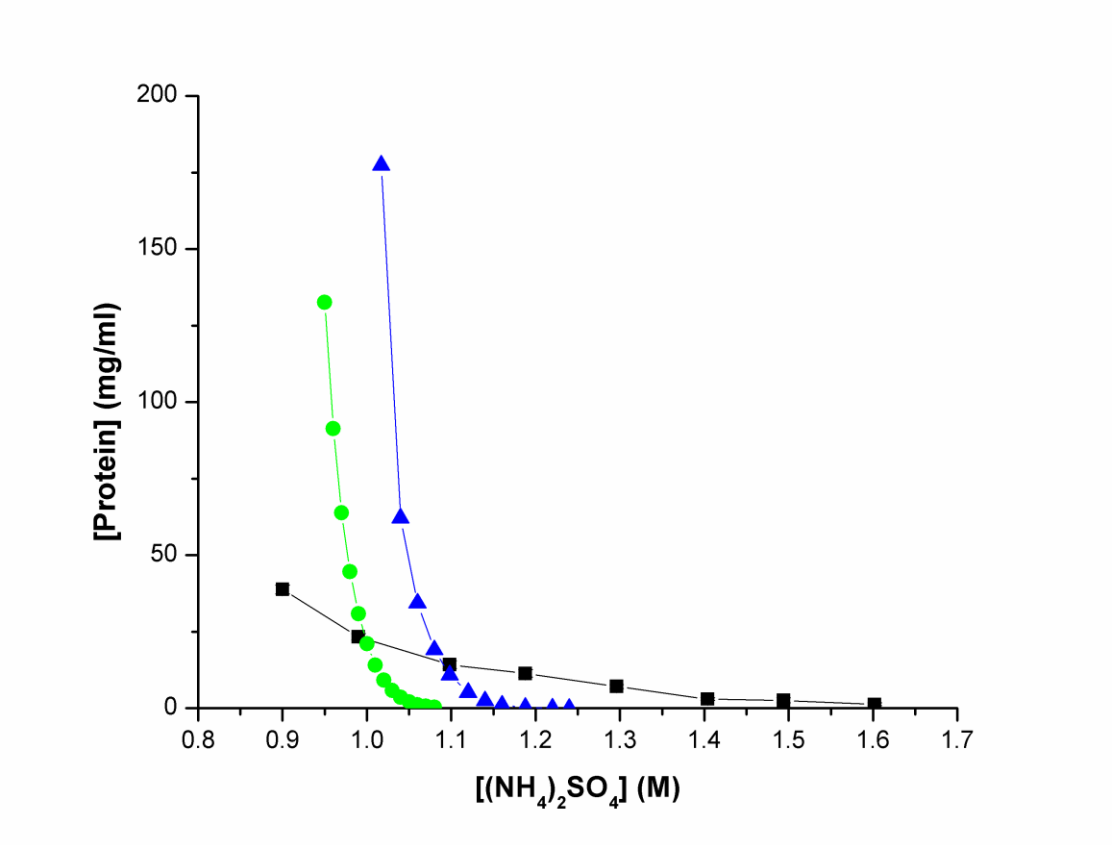

Figure 2.11: Predicted binodal boundaries from the Haas-Drenth model for (▲) (●) *m*=279 compared with the original predictions from *m*=559 and the (■) experimental results.

2.5 Osmotic Virial Equation

2.5.1 Theory

The osmotic virial equation was derived by McMillan and Mayer to describe the nonideality of dilute solutions and is given by (94)

$$\frac{\Pi}{RT} = \frac{c}{MW} + B_{22}c^2 + B_{222}c^3 + \cdots \qquad 2.19$$

where Π is the osmotic pressure, R is the molar gas constant, and c is the protein concentration in units of mass/volume. Equation 2.19 is frequently truncated at the B_{22} term; however, in order for the model to predict phase separation, it must include at least the term in the osmotic third virial coefficient B_{222}, which represents three-body interactions in solution. The governing equations for liquid-liquid coexistence for the osmotic virial equation can be derived as (see Appendix A)

$$\frac{\phi_I - \phi_{II}}{\underline{V_2}} + B_{22}\left(\frac{MW}{\underline{V_2}}\right)^2(\phi_I^2 - \phi_{II}^2) + B_{222}\left(\frac{MW}{\underline{V_2}}\right)^3(\phi_I^3 - \phi_{II}^3) = 0 \qquad 2.20$$

$$\ln\left(\frac{\phi_I}{\phi_{II}}\right) + \left(\frac{2B_{22}MW^2}{\underline{V_2}} - 1\right)(\phi_I - \phi_{II}) + \left(\frac{3B_{222}MW^3}{2\underline{V_2}^2} - \frac{B_{22}MW^2}{\underline{V_2}}\right)(\phi_I^2 - \phi_{II}^2) - \left(\frac{B_{222}MW^3}{\underline{V_2}^2}\right)(\phi_I^3 - \phi_{II}^3) = 0 \qquad 2.21$$

Predicting fluid phase equilibrium requires that the third virial coefficient B_{222} be specified. B_{222} can be theoretically calculated using a simple model of the potential of mean force (PMF) for protein-protein interactions. The hard-core attractive Yukawa potential was chosen as the PMF model because it has been used to describe colloidal interactions (6, 56). The Yukawa potential was originally derived as a screened Coulombic PMF model to capture long-ranged repulsive electrostatic interactions (101), but it has been modified to model the short-ranged attractions that

dominate phase separation in protein solutions. This Yukawa potential consists of a hard-sphere contribution and an attractive tail and is given by (2, 17)

$$u(r) = \begin{cases} \infty, & r < \sigma \\ -\frac{\varepsilon\sigma}{r} e^{-b(r-\sigma)}, & r \geq \sigma \end{cases} \qquad 2.22$$

where r is the center-to-center intermolecular distance, ε is the interaction well depth, σ is the particle diameter, and b is a parameter that characterizes the range over which the attraction occurs in units of inverse length. The potential can be rewritten in terms of reduced variables, in which the parameters of the potential are scaled by characteristic values. If reduced variables are defined as $b^*=b\sigma$, $r^*=r/\sigma$, and $u^*=u/\varepsilon$, the Yukawa potential can be expressed as

$$u^*(r^*) = \begin{cases} \infty, & r^* < 1 \\ -\frac{e^{-b^*(r^*-1)}}{r^*}, & r^* \geq 1 \end{cases} \qquad 2.23$$

A plot of the Yukawa potential for different values of the range parameter b^* is shown in Figure 2.12.

B_{222} was calculated from experimental B_{22} data for different values of b^*. To determine a starting value for these calculations, b^* was chosen such that the Yukawa potential approximately overlapped with the 140-35 Lennard-Jones potential. The 140-35 Lennard-Jones potential was empirically found to adequately describe short-ranged non-electrostatic interactions between protein molecules based on more extensive atomistic simulations of protein-protein interactions (78). From this procedure, a value of $b^*=22$ for the Yukawa potential was found to approximately overlay with the 140-35 Lennard-Jones potential (Figure 2.13).

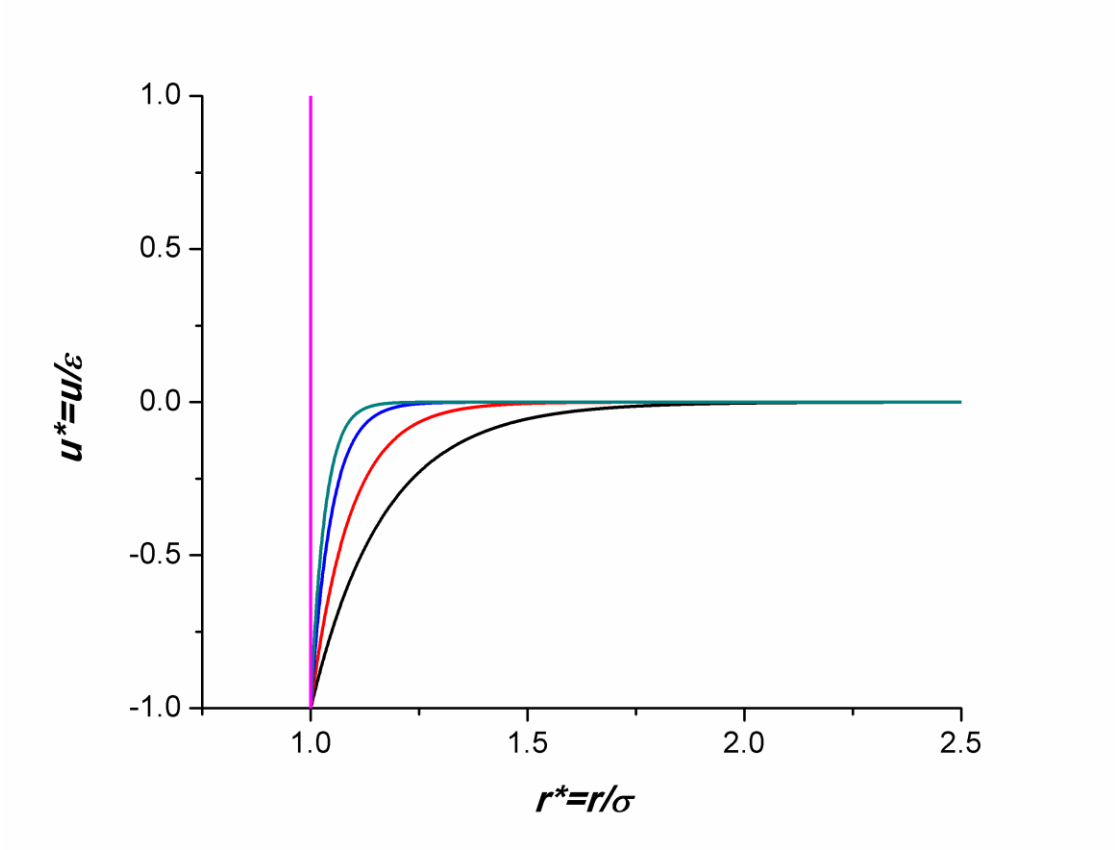

Figure 2.12: Plot of the Yukawa potential in reduced units for *b values of (–) 5, (–) 10, (–) 20, and (–) 30. The (–) hard-sphere repulsion occurs at *r**=1.0. Increasing *b** corresponds to a decrease in the range of attraction.**

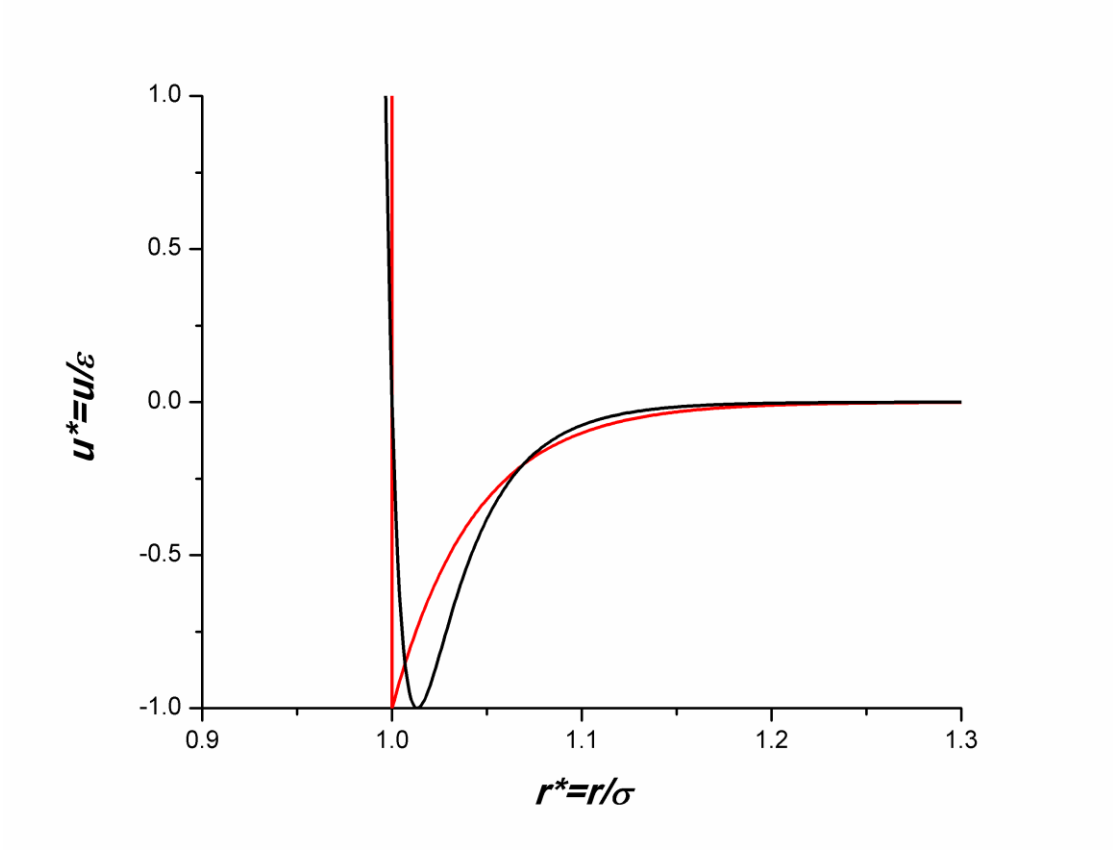

Figure 2.13: Comparison of the (–) 140-35 Lennard-Jones potential with the (–) Yukawa potential for b^* = 22.

The procedure for calculating B_{222} from B_{22} involved the following steps:

1) The reduced interaction range parameter b^* was fixed and the molecular diameter, σ, of ribonuclease A was taken to be the sphere-equivalent value of 3.1 nm (86).

2) The ε parameter was determined from the isotropic model for the second virial coefficient, which is given by (42)

$$B_{22} = -\frac{2\pi N_A}{MW^2}\int_0^\infty \left[exp\left(-\frac{u(r)}{kT}\right) - 1\right] r^2\, dr \qquad 2.24$$

where N_A is Avogadro's number and $u(r)$ is the Yukawa potential. The experimental B_{22} values were used as inputs into equation 2.24 and the corresponding ε values were calculated at each salt concentration.

3) Values of the ε parameter from step 2) were used to calculate B_{222} values from the equation for third virial coefficients (102–104)

$$B_{222} = \frac{-8N_A^2\pi^2}{3MW^3}\int_0^\infty\int_0^\infty\int_{|r_{12}-r_{23}|}^{r_{12}+r_{23}} f(r_{12})r_{12}f(r_{13})r_{13} \times f(r_{23})r_{23}\, dr_{12}dr_{13}dr_{23} \qquad 2.25$$

where r_{ij} is the intermolecular separation between particle i and j and $f(r_{ij})$ is the Mayer cluster function defined as (42)

$$f(r_{ij}) = exp\left(-\frac{u_{ij}}{kT}\right) - 1 \qquad 2.26$$

The model for the third virial coefficient represented by equation 2.25 assumes that the molecules are spherically symmetric, the interactions are pairwise additive and multibody interactions are neglected. To compute B_{222} using the Yukawa potential in equation 2.25, the method of Alder and Pople (103) for calculating third virial coefficients for potentials with hard-sphere cores was utilized. This method was previously used by Graben and Present to calculate third virial coefficients for the Sutherland potential (104). The computed third virial coefficients were compared with the results of Naresh and Singh (105), who utilized the Mayer sampling technique to calculate the virial coefficients for the Yukawa potential. All calculations and data analysis were performed using appropriate numerical tools in Matlab (see Appendix B for actual code).

2.5.2 Results

The computed B_{222} values are presented in Figure 2.14. Different trends were observed in the behavior of B_{222} for different b^* values as a function of salt. For b^*=25, B_{222} initially increases with increasing salt concentration but then sharply decreases to negative values. This trend has been observed in the behavior of third virial coefficients for other potential models (102). For b^*=35, B_{222} is positive throughout the salt range of interest, but the values appear to plateau around 1.25 M, which suggests that the values would begin to decrease at higher salt concentrations. When b^* is further increased to a value of 45, B_{222} increases over the entire salt range. Thus, it can be inferred that as b^* increases, B_{222} predicted from the Yukawa potential increases over a wider range of salt concentration.

The phase diagram predictions from the experimental B_{22} data using the osmotic virial equation are presented in Figure 2.15. Phase separation was predicted by the model for B_{222} values calculated using b^* values of 22 or greater in the Yukawa potential, which corresponds to an interaction range that is 1/22 of the particle diameter σ, or less than 1.5 Å. It was determined from these calculations that positive B_{222} values are needed in order for the osmotic virial equation to predict phase separation. Phase behavior calculations with the model were made for b^* values of 25, 35, and 45, corresponding to decreasing interaction distances. Calculations were not performed for higher values of b^* because such short interaction ranges were considered physically unrealistic. The phase behavior predictions follow the correct qualitative trend of decreasing solubility with increasing salt concentrations; however, none of the values of b^* yield phase behavior that matches the experimental data. In each case, phase separation is not predicted at lower salt concentrations, for which it is observed experimentally.

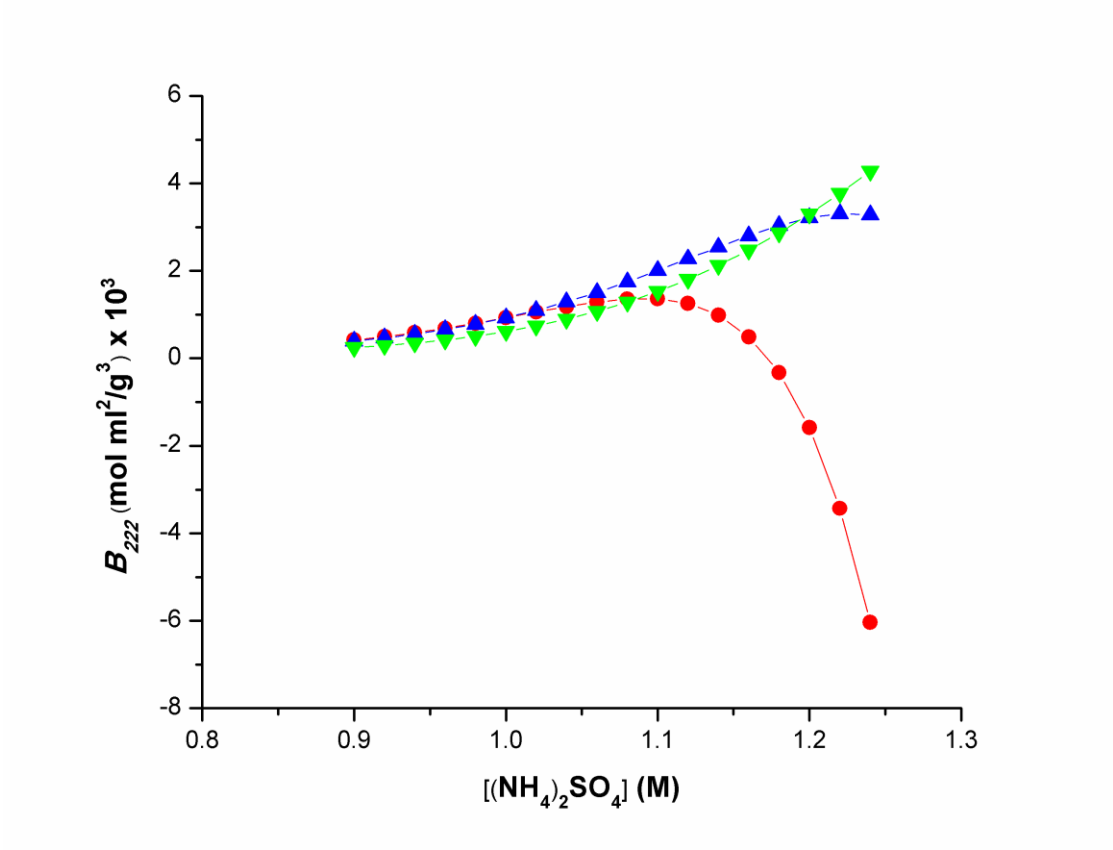

Figure 2.14: Computed B_{222} from the Yukawa potential for b^* values of (●) 25, (▲) 35, and (▼) 45.

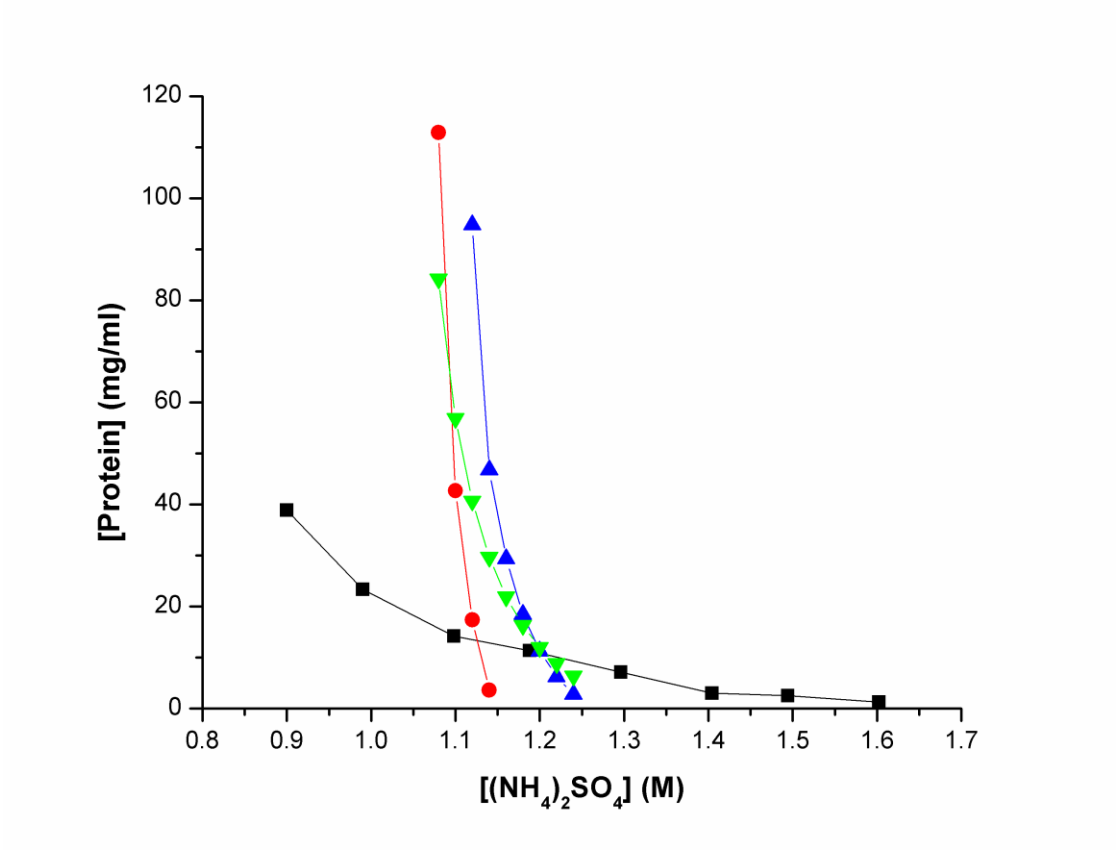

Figure 2.15: Phase behavior predictions from the osmotic virial equation based on B_{222} calculated from the Yukawa potential. Predictions were made for b^* values of (●) 25, (▲) 35, and (▼) 45 and compared with the (■) experimental binodal data.

The sensitivity of the phase equilibrium predictions from the osmotic virial equation to experimental errors in B_{22} was also explored. An approximate estimate of $\pm 2\times10^{-4}$ mol ml/g^2 was assumed for the error in the experimental B_{22} data. The original experimental B_{22} values were adjusted by either $+2\times10^{-4}$ mol ml/g^2 or -2×10^{-4} mol ml/g^2 and corresponding B_{222} values were recalculated for fixed values b^*. The phase diagrams were subsequently recalculated from the osmotic virial model. The resulting phase diagram predictions for b^* values of 35 and 45 are presented in Figures 2.16 and 2.17, respectively. Decreasing the B_{22} values by -2×10^{-4} mol ml/g^2 causes the predicted equilibrium phase boundary to fall back to lower protein concentrations, which essentially means that the binodal curve becomes broader. This trend is to be expected because decreasing B_{22}, i.e., increasing the strength of attraction, would lead to the protein being less soluble in solution and therefore would push the equilibrium phase boundary to lower concentrations. However, it does not appear that the predictions are sufficiently sensitive to B_{22} changes that errors in B_{22} would account for the discrepancy between the predicted binodal and the experimental phase behavior results.

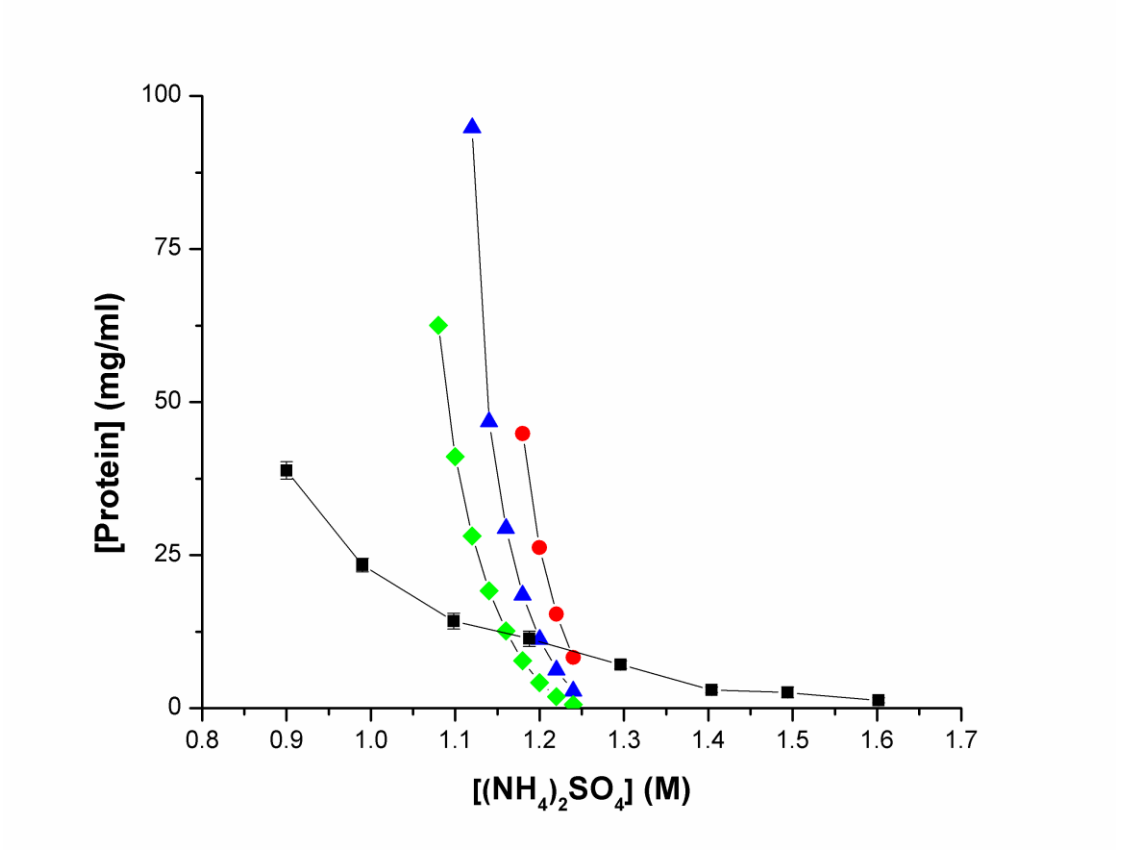

Figure 2.16: Binodal predictions based on b^*=35 compared with the (■) experimental binodal data. Osmotic virial predictions were made from the experimental B_{22} values with an assumed error of (●) $+2\times10^{-4}$ mol ml/g^2 and (◆) -2×10^{-4} mol ml/g^2. The results are compared with the predictions from the (▲) original B_{22} data set.

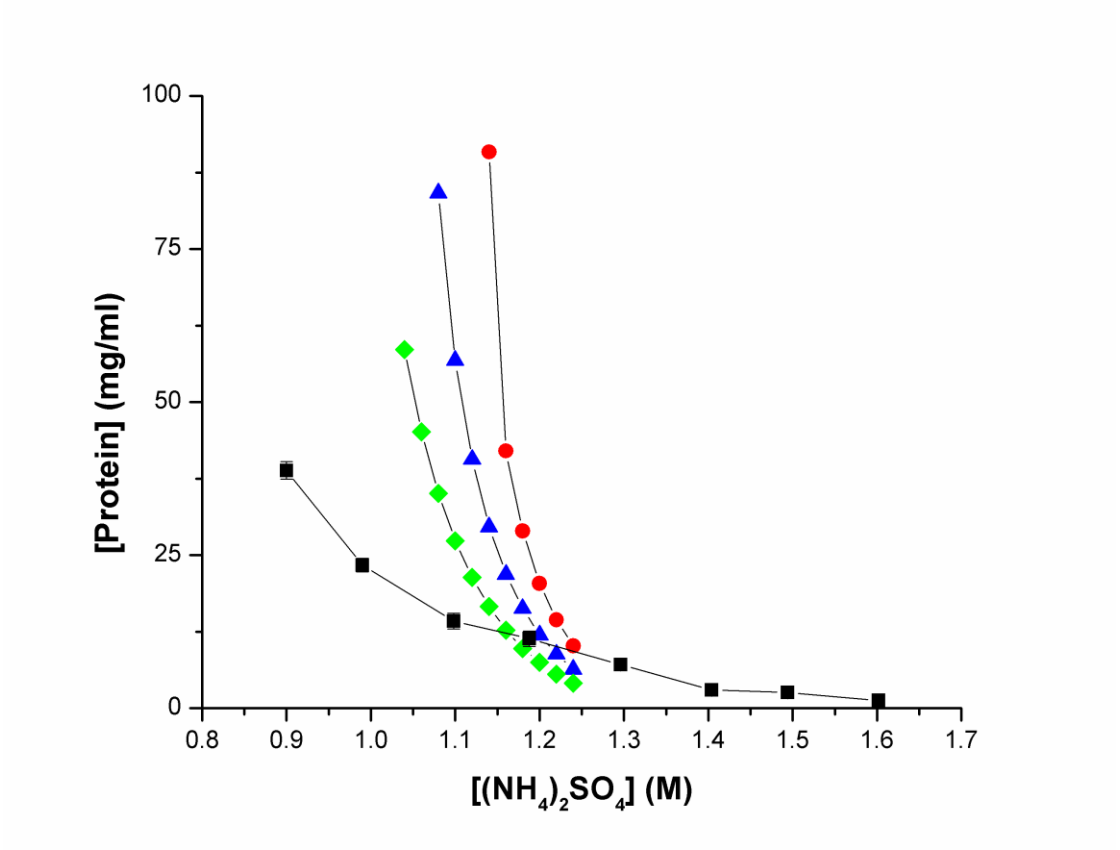

Figure 2.17: Binodal predictions based on b^*=45 compared with the (■) experimental binodal data. Osmotic virial predictions were made from the experimental B_{22} values with an assumed error of (●) $+2\times10^{-4}$ mol ml/g^2 and (◆) -2×10^{-4} mol ml/g^2. The results are compared with the predictions from the (▲) original B_{22} data set.

Because the predicted phase coexistence computed using the Yukawa potential did not match experiment, other potential of mean force models were explored to determine if they could lead to better phase behavior results. These models included the square-well potential, 140-35 Lennard-Jones potential, and the ten Wolde-Frenkel potential. The square-well potential was chosen because of its simplicity and the fact that since the interactions between proteins are very short-

ranged, a potential of mean force model of this form may be adequate to describe the interaction. The 140-35 Lennard-Jones potential (78) was used because this isotropic potential provides a reasonable approximation to detailed atomistic simulations of protein-protein interactions. The ten Wolde-Frenkel potential was also investigated because it has been used to account for both direct and for solvent-induced interactions between globular proteins (12). The same methodology as that for calculating B_{222} from the Yukawa potential was utilized for these potential models.

The square-well potential is the simplest attractive potential and is given in terms of reduced variables as

$$u^*(r^*) = \begin{cases} \infty, & r^* < 1 \\ -1, & 1 < r^* < \gamma^* \\ 0, & \gamma^* < r^* < \infty \end{cases} \qquad 2.27$$

where γ is the parameter that characterizes the range of attraction. B_{222} for the square-well potential has the analytical form (106, 107)

$$\begin{aligned} B_{222} = \frac{1}{8MW^3}[&(\gamma^6 - 18\gamma^4 + 32\gamma^3 - 15)x \\ &+(-2\gamma^6 + 36\gamma^4 - 32\gamma^3 - 32\gamma^2 + 16)x^2 \\ &-(6\gamma^6 - 18\gamma^4 - 18\gamma^2 - 6)x^3] \ for\ \gamma \le 2 \end{aligned} \qquad 2.28$$

$$\begin{aligned} = \frac{1}{8MW^3}[&5 - 17x + (32\gamma^3 - 18\gamma^2 - 48)x^2 \\ &-(5\gamma^6 - 32\gamma^3 + 18\gamma^2 + 26)x^3] \ for\ \gamma \ge 2 \end{aligned}$$

where x is defined as

$$x = \exp\left(\frac{\varepsilon}{kT}\right) - 1 \qquad 2.29$$

A plot of B_{222} values calculated from equation 2.28 for different values of γ is presented in Figure 2.18. For low values of γ, B_{222} initially remains fairly flat but then decreases sharply as the salt concentration increases. When γ is sufficiently large, B_{222} instead increases monotonically within the salt range of interest. The results suggest

that the B_{222} values determined from higher values of γ would be more suitable for calculating phase equilibrium since positive values of B_{222} are needed for the osmotic virial equation to predict phase separation.

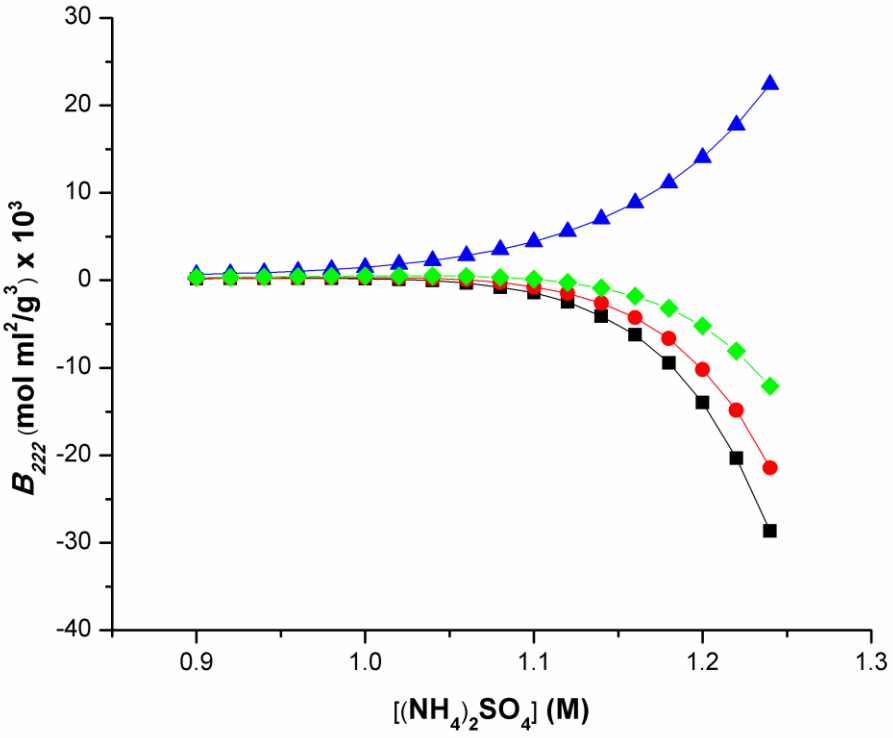

Figure 2.18: B_{222} computed from the square well potential for γ values of (■) 1.05, (●) 1.20, (◆) 1.50, and (▲) 2.10.

B_{222} was also computed from the 140-35 Lennard-Jones potential, which is given by

$$u^*(r^*) = 2.1165\left[\left(\frac{1}{r^*}\right)^{140} - \left(\frac{1}{r^*}\right)^{35}\right] \qquad 2.30$$

A plot of the resulting B_{222} values is presented in Figure 2.19. The computed B_{222} values are mostly negative over the range of salt concentration and decrease sharply at higher salt concentrations.

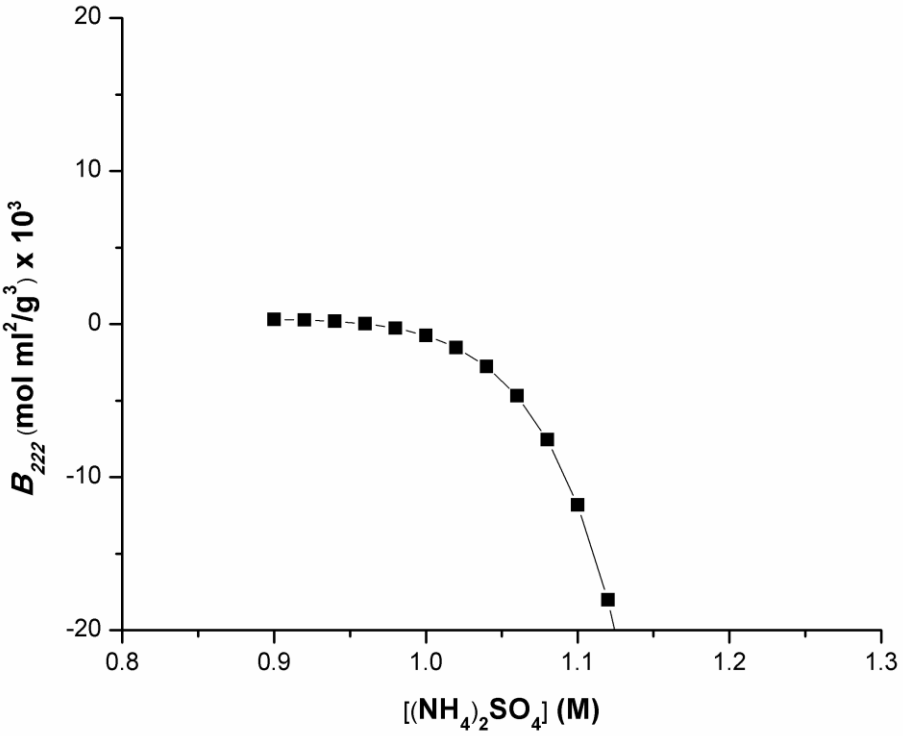

Figure 2.19: B_{222} computed from the 140-35 Lennard-Jones potential.

The ten Wolde-Frenkel potential is a generalized Lennard-Jones potential and is given as

$$u^*(r^*) = \begin{cases} \infty, & r^* < 1 \\ -\frac{4}{\alpha^2}\left\{\frac{1}{\left[r^{*2}-1\right]^6} - \alpha\frac{1}{\left[r^{*2}-1\right]^3}\right\}, & r^* \geq 1 \end{cases} \qquad 2.31$$

where α is the parameter that controls the range of attraction. As α increases, the interaction range decreases and vice versa. A plot of the potential for different values of α is shown in Figure 2.20. An α value of 50 was determined previously to qualitatively reproduce the phase behavior of proteins (12), so this value was used as a starting point in the B_{222} calculations. The α parameter was then adjusted and the corresponding B_{222} values were computed, as shown in Figure 2.21.

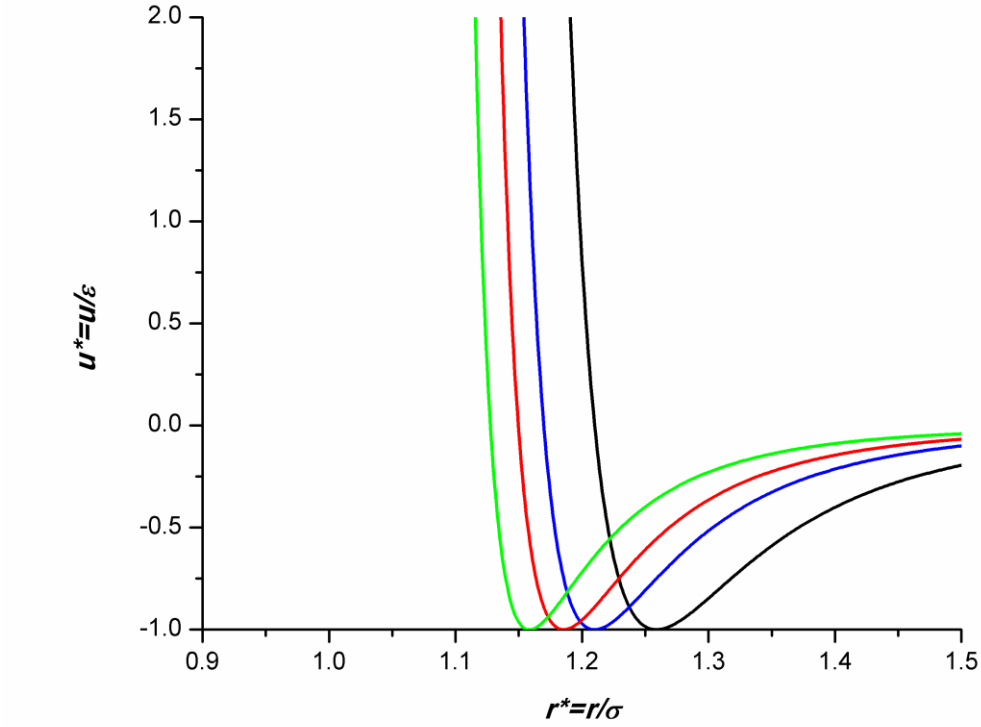

Figure 2.20: Plot of the ten Wolde-Frenkel potential for α values of (–) 10, (–) 20, (–) 30, and (–) 50.

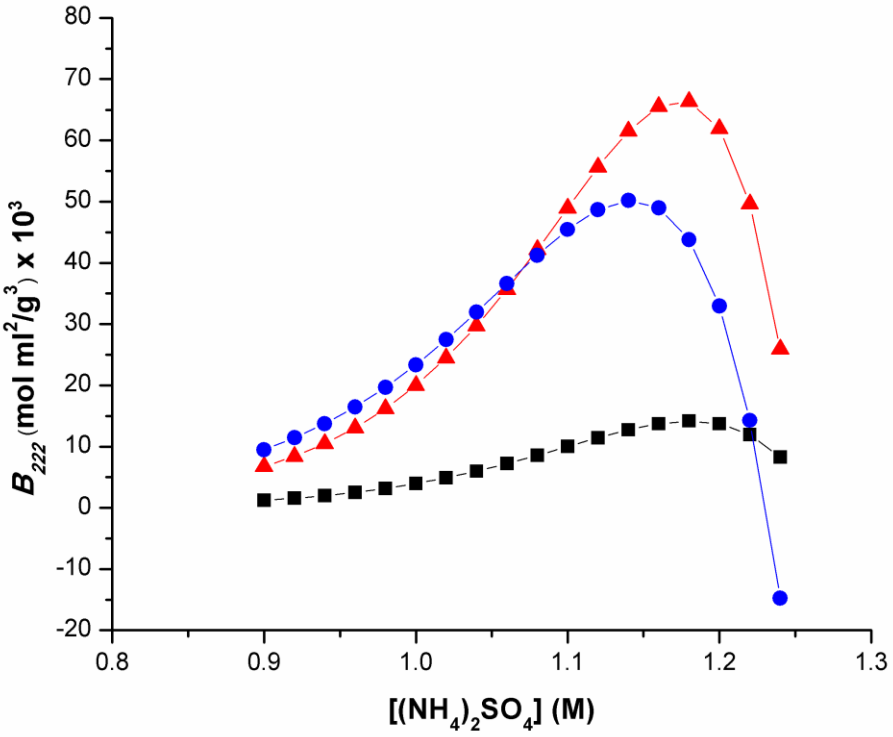

Figure 2.21: B_{222} computed from the ten Wolde-Frenkel potential for α values of (■) 10, (▲) 20, and (●) 30.

Phase behavior calculations were subsequently performed using B_{222} values calculated from the above potential of mean force models. However, the resulting B_{222} values did not lead to phase separation with the osmotic virial equation. The inability to predict phase coexistence suggests that these potential models are not adequate for describing the interactions for ribonuclease A. However, the problem may arise from the need to include higher-order virial coefficients to obtain better phase coexistence predictions.

2.6 Discussion

The discrepancies between the model predictions and experimental data may be the result of the simplifying assumptions inherent in the continuum models. The lattice representation used by Flory-Huggins theory allows the polymer chain to adopt any random configuration, which is unrealistic for proteins since they are known to have a preferred native conformation. In addition, the interaction parameter assumes that each monomer interacts equally with the solvent molecules and that the magnitude is dependent on the number of contacts. This assumption effectively represents the interactions as being isotropic since the theory does not account for strongly attractive regions of the polymer chain. The Haas-Drenth model treats protein molecules as interacting hard spheres where with no regions that display strong attractions. The framework of hard spheres also implies the assumption of isotropic interactions. The inability of the osmotic virial equation predictions to match experimental phase behavior could be partially due to the assumptions made in calculating B_{222} from B_{22} data. B_{222} was determined by assuming pairwise additivity and neglecting multibody interactions in addition to the isotropic assumption.

Another possible reason for the inability to predict phase behavior could be the form of the potential of mean force used to calculate B_{222}. The Yukawa potential is a simple isotropic potential model that would not be expected to realistically capture the complex anisotropic interactions of proteins in solution. The other potential models that were utilized to calculate B_{222}, which included the square-well potential, 140-35 Lennard-Jones potential, and ten Wolde-Frenkel potential, did not result in B_{222} values that led to prediction of phase equilibrium. Furthermore, improvements in predicting phase behavior with the osmotic virial equation may require inclusion of higher-order virial coefficient terms; however, calculating such

virial coefficients from PMF models is a difficult task and therefore may not be an efficient path to follow for future work.

Another issue that may explain the quantitative disagreement between prediction and experiment could be the orientation-averaged nature of B_{22}. While B_{22} is dominated by a few attractive configurations due to the Boltzmann weighting of the PMF (82), the orientational averaging essentially washes out the molecular details of the interactions. Because of this averaging, B_{22} provides an incomplete representation of protein-protein interactions and thus would be expected to be limited in its ability to quantitatively predict phase behavior.

2.7 Conclusions

Isotropic interactions are commonly assumed in the models used, and may be the reason for the limited quantitative capabilities of the models to predict protein phase behavior. Therefore, based on the work that has been done with the continuum models, it can be concluded that the anisotropic character of protein-protein interactions should be taken into account to quantitatively predict protein phase behavior. In order to account for anisotropy of protein interactions to predict phase behavior, molecular-level modeling methods will be needed.

One class of models that has been used to account for the anisotropy of colloidal particle interactions and has recently been applied to proteins are patch models (74). These models have been used to simulate colloidal phase behavior and the results from these models have been shown to be different from those resulting from isotropic models. The use of patch models to model protein-protein interactions is the subject of the next chapter.

Chapter 3

PATCH-ANTIPATCH MODEL OF PROTEINS AND THE CALCULATION OF B_{22}

3.1 Introduction

3.1.1 Review of Patch Models

The interactions of protein molecules are inherently anisotropic. The physical basis of this anisotropy stems from the nonuniform charge distribution, nonspherical shape, rough local topography, and heterogeneous functionality on the protein surface. Anisotropic interactions are responsible for the wide range of solution phenomena observed in proteins, which include the formation of clusters, gels, glasses, and crystal nucleation. Incorporating this feature in modeling protein-protein interactions is important in the simulation of protein phase behavior. However, direct molecular simulations of phase behavior using models of proteins represented in full atomistic detail with explicit solvent are not presently computationally tractable. A more feasible approach entails utilizing a simplified coarse-grained representation of protein molecules that captures the essential physics of protein interactions. One such approach is the use of patch models.

Several classes of patch models have been proposed to account for the orientation dependence in protein-protein interactions. One of the earliest patch models applied for proteins was the aeolotopic model developed by Lomakin et al. for describing the phase behavior of γ-crystallin (75). In this model, the protein molecule

is represented as a sphere of which the surface includes a number of attractive spots. Neighboring protein molecules are said to make contact when the interactions are between these spots, with the interaction modeled by a square-well potential. Lomakin et al. found that by including this directionality in modeling the interactions, the predicted fluid-fluid coexistence curve broadens and more closely matches experimental measurements than does the isotropic square-well model. Sear (76) later developed a conical site model for globular proteins in which the particle is modeled as a hard sphere and interactions occur between paired sites. Using Wertheim perturbation theory to predict phase behavior, he found that this model was capable of predicting a metastable fluid-fluid phase transition. Kern and Frenkel (73) proposed a patch model for colloidal particles in which the specific directional interactions between patches depend on the relative orientations of two interacting protein molecules. Their model offered greater flexibility in the number of patches, patch coverage, and range of patch-patch interactions. Liu et al. (79) later extended the Kern and Frenkel model by adding a background isotropic square-well attraction in addition to the patch interactions to simulate protein phase behavior.

The various patch models that have been used to represent proteins have different characteristics, but they do share common features. They generally represent each protein molecule as a hard sphere with specific attractive regions on the surface. In most of the models studied, each patch on one sphere can interact equally with all other patches on surrounding spheres. The interactions between these regions are strongly attractive and short-ranged with respect to the particle size. Most work with patch models has described the patch-patch attractions using the square-well potential, but there have been studies that have used different potential models (77, 78, 108).

Patch models involve a larger parameter set, which includes the number of patches, patch size, patch arrangement, and range and strength of patch-patch interactions. Thus, patch models offer greater flexibility and potential for describing the rich variety of dense phases and their phase boundaries observed experimentally for proteins.

The phase behavior predicted by patch models is both qualitatively and quantitatively different from the behavior predicted by isotropic potential models. Similar to isotropic models, patch models are able predict a metastable fluid-fluid transition region for short-ranged interactions (73, 76, 77, 109, 110). However, the number of patches is one of the key parameters that controls the phase diagram (79, 111–115). For patch models, it has been shown by theory and simulation (79, 109, 113, 114) that decreasing the number of patches shifts the critical point to lower temperatures and densities. When compared in reduced units, the fluid-fluid coexistence curve is broader and the form is in quantitative agreement with experimental data for lysozyme and γ-crystallin (73, 75, 109). This result indicates the importance of patchiness in modeling phase behavior for proteins. Patch models also open the possibility of describing competing crystalline phases that can be orientationally ordered or disordered (108, 110, 116–120). The types of stable crystal structures predicted depend on the compatibility of the patch arrangement on the surface. Most studies of patch models so far have focused on symmetric arrangements of patches, although the distribution of attractive patches on actual protein surfaces is certainly non-uniform. Incorporating anisotropy through patch models has been useful in describing the rich phase behavior known experimentally for proteins.

Current patch models provide a useful framework for incorporating anisotropy; however, they are inadequate in representing some of the unique

molecular details of proteins. First, these models assume that the protein molecules are spheres, which is a major simplification since even globular proteins are not perfectly spherical. In addition, most studies of patch models have assumed symmetric patch distributions where each patch can interact equally with all other patches. In reality, the attractive regions are non-uniformly distributed on the protein surface. Also, specific interactions between geometrically complementary regions are known to be an intrinsic feature of protein-protein interactions (121). It is these highly specific attractions that play an important part in determining solution properties, such as the osmotic second virial coefficient B_{22}, and that control protein crystallization (82). Therefore, current patch-patch models cannot truly describe the structure of any real protein solid phase. So far, there have been no studies that have explored the effect of non-uniform patch distribution on the structure of crystals and other dense phases. Theoretical examination of protein phase behavior with patch models could be improved by realistically representing the structural details of protein shape, patch distribution, and specific interactions between geometrically complementary regions. One model that takes a step in this direction is the patch-antipatch model developed by Hloucha et al. (78), which is explored in this chapter.

3.1.2 The "Patch-Antipatch" Model

The patch-antipatch model explicitly accounts for highly specific interactions that arise from complementary regions on the protein surface. In this model, the protein is represented as a sphere decorated with patches and corresponding antipatches on the surface (Figure 3.1). In addition to the weak, distance-dependent isotropic interaction, there are strong interactions that occur only between patch-

antipatch pairs, which reflect the specific pairwise attractions between geometrically complementary regions. The interaction potential that describes this framework is

$$u(r) = u_{iso}(r) + \sum_{m,n}^{N_P} S(\alpha_m, \alpha_n) u_{PA_{m,n}}(r) \qquad 3.1$$

in which

$$S(\alpha_m, \alpha_n) = \begin{cases} \left(1 - \frac{2\alpha_m}{\alpha_p}\right)^2 \left(1 - \frac{2\alpha_n}{\alpha_p}\right)^2, if\ \alpha_m < {}^{\alpha_p}/_2, \alpha_n < {}^{\alpha_p}/_2 \\ 0, \qquad otherwise \end{cases} \qquad 3.2$$

where N_p is the number of patch-antipatch pairs and u_{iso} and $u_{PAm,n}$ are the isotropic and patch-antipatch contributions to the interaction potential, respectively. Each unique patch *m* interacts only with its corresponding antipatch *n*. The patch-antipatch interactions are modulated by the scaling function $S(\alpha_m, \alpha_n)$, which is dependent on the relative orientations of the patch and antipatch α_m and α_n, respectively, and on the size of the patches α_p. The virtue of this model over other patch models is that it explicitly represents the molecular recognition phenomenon characteristic of protein-protein interactions.

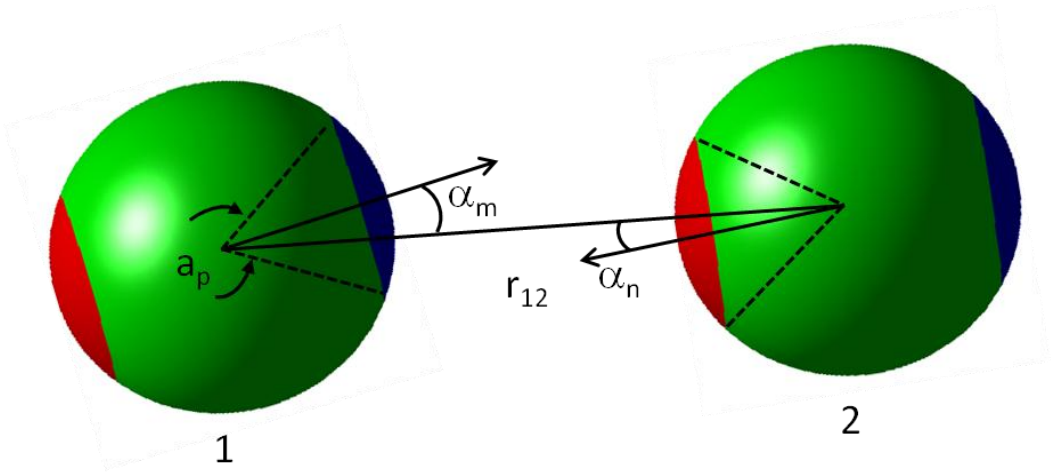

Figure 3.1: A cartoon of a "patch-antipatch" model of protein molecules. The "patch" is colored in blue and the corresponding "antipatch" is colored in red. Specific interactions occur only between unique blue and red colored regions, which depend on the angles of alignment α_m and α_n.

3.1.3 Objective

The objectives of this work are to 1) identify patch-antipatch pairs for specific model proteins and determine the physical patch-antipatch parameters and 2) analyze the impact of these highly attractive patch-antipatch pairs on the computation of B_{22} at the atomistic level. Understanding the role of these patch-antipatch pairs will give insight into the anisotropic nature of protein-protein interactions and the influence it has on the solution properties of proteins, which includes phase behavior.

The unique patch-antipatch parameters for individual proteins are determined using a hybrid atomistic/continuum methodology for calculating interaction energies between protein molecules (82, 122). This method involves simulation of two protein molecules modeled in full in atomistic detail and calculating the energy of interactions for different angular configurations. This method is capable

of capturing the effects of geometric complementarity between the surfaces of protein molecules. The parameters are:

1) The number of patch-antipatch pairs N_p
2) The location and arrangement of the patch-antipatch pairs specified by the translation angles ϕ, θ and rotation angles α, β, γ
3) The size of the patch-antipatch pairs α_p
4) The strength of patch-antipatch interactions ε_p

These patch-antipatch parameters are determined for two model proteins that exhibit different solution behavior: lysozyme and chymosin B. Lysozyme has been extensively studied in the literature and is known to exhibit salting-out behavior for a wide range of pH values (123). Chymosin B, on the other hand, exhibits salting-in behavior at pH values near its pI, which is thought to be due to the anisotropy of its charge distribution (124). The structural data from the PDB files and physical properties of these model proteins are presented in Table 3.1.

Table 3.1: Proteins studied for patch-antipatch analysis and their physical properties.

Protein	PDB ID	MW (g/mol)	Residues	pI
Lysozyme	4LYZ	14300	129	11
Chymosin B	1CMS	35673	323	4.6

3.2 Theory and Methods

3.2.1 Determining "Patch-Antipatch" Pairs

Patch-antipatch pairs were determined from calculations of short-ranged attractions (primarily van der Waals interactions) between two protein molecules

represented in full atomistic detail. Protein structures were obtained from the atomic coordinates contained in PDB files from the RCSB Protein Data Bank (www.pdb.org). The protein molecules were assumed to be rigid bodies in their native conformation. Because of the short-ranged nature of van der Waals attractions, these interactions are sensitive to the local geometry of the surfaces. In fact, the level of detail used for the protein structure has a profound effect on the magnitude of the van der Waals attraction (18). Thus, regions where there is geometric complementarity of apposing surfaces lead to stronger attraction. Shape complementarity of the surfaces plays an important role in the "lock and key" mechanism that is intrinsic in the biological specificity of protein-protein interactions.

The attraction between two protein molecules was quantified by calculation of the interaction energies, which depends on the relative orientations of the molecules. The interactions for a unique angular configuration were determined by fixing one protein molecule at the origin and translating the second molecule towards the first in fixed steps, with the interaction energy calculated at each step. For orientations where attractions are strong, this leads to a potential with a larger well depth. Thus, a "patch-antipatch" pair is characterized by angular orientations that lead to interaction potentials with particularly deep wells.

3.2.2 Interaction Energies

3.2.2.1 Short-Range Interactions

A hybrid atomistic/continuum method was used to calculate the short-ranged interactions (82, 122); a brief review of the model formulation is given here. In this formulation, interactions between two protein molecules are calculated as a

sum of the pairwise interactions between the atoms of the proteins, with solvent effects implicitly taken into account. The interactions between atom pairs are modeled based on the separation distances between the atoms. For atom pairs separated by a center-to-center distance of more than 6 Å, the interactions are determined by the continuum Lifshitz-Hamaker formulation of van der Waals interactions given by

$$U_{ij}^{LH} = -\frac{A_H}{\pi^2}\int_{V_i}\int_{V_j}\frac{1}{r_{ij}^6}dV_i dV_j \qquad 3.3$$

where r_{ij} is the center-to-center separation distance between two volume elements dV_i and dV_j, and A_H is the Hamaker constant, which for protein-water-protein interactions has been determined to be 3.1 *kT* (125). If two atoms are represented as spheres of radii R_i and R_j, it can be shown from equation 3.3 that the interaction potential is (126)

$$U_{ij}^{LH} = \frac{-A_H}{6}\left[\frac{2R_iR_j}{r_{ij}^2-\left(R_i+R_j\right)^2}+\frac{2R_iR_j}{r_{ij}^2-\left(R_i-R_j\right)^2} + \ln\left(\frac{r_{ij}^2-\left(R_i+R_j\right)^2}{r_{ij}^2-\left(R_i-R_j\right)^2}\right)\right], \quad if\ r_{ij} > 6\text{Å} \qquad 3.4$$

The total free energy of interaction in the Lifshitz-Hamaker approach, W_{LH}, is the sum of all atom-atom pair interactions described by equation 3.4

$$W_{LH} = \sum_i\sum_j U_{ij}^{LH} \qquad 3.5$$

For atom pairs separated by a center-to-center distance of less than 6 Å, the continuum approximation for the solvent breaks down, and therefore equation 3.4 cannot be used. For this situation, the atomistic Lennard-Jones formulation is used to determine the short-range interaction

$$U_{ij}^{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \qquad if\ r_{ij} < 6\text{Å} \tag{3.6}$$

where U_{ij} is the interaction energy, r_{ij} is the center-to-center distance between two atoms, σ_{ij} is the size parameter, and ε_{ij} is the strength of interaction parameter. The total free energy of interaction in the Lennard-Jones approach, W_{LJ}, is the sum of all atom-atom pair interactions described by equation 3.6

$$W_{LJ} = \sum_i \sum_j U_{ij}^{LJ} \tag{3.7}$$

The parameters for the Lennard-Jones model were taken from the OPLS-AA force field (127). The issue with utilizing equation 3.6 to model the dispersion interaction is that the solvent molecules need to be included explicitly. To take into account effects from solvation forces, an empirical parameter α is introduced to correct the magnitude of the Lennard-Jones contribution. It has been shown that a value of 0.50 for α provides a reasonable adjustment when compared with experimental binding free energies for proteins (128). The total non-electrostatic interaction W_{ne} is the sum of both free energy contributions from the Lifshitz-Hamaker and Lennard-Jones approaches

$$W_{ne} = W_{LH} + \alpha W_{LJ} \tag{3.8}$$

in which the Lennard-Jones contribution W_{LJ} is scaled by the empirical factor α = 0.50.

3.2.2.2 Electrostatic Interactions

The electrostatic interactions are a result of the charges carried by titratable amino acid residues and partial charges of the atoms. A pairwise screened

Coulombic potential was utilized in computing the electrostatic contribution to the interactions between protein molecules W_{elec}

$$W_{elec} = \sum_i \sum_j \frac{q_i q_j}{\varepsilon_0 \varepsilon_r r} e^{-\kappa r} \qquad 3.9$$

where q_i and q_j are the charges on the two interacting atoms, ε_0 is the permittivity of free space, ε_r is the dielectric constant of the solvent, and κ is the Debye parameter that characterizes the length scale for screening of electrostatic interactions by the free ions in solution. The Debye parameter is related to the ionic strength by

$$\kappa = \sqrt{\frac{\frac{1}{2}\sum_i (z_i e)^2 c_i}{\varepsilon_0 \varepsilon_r kT}} \qquad 3.10$$

where c_i is the concentration of the ion i, e is the elementary charge, and z_i is the valence of the ion i. Within this framework, the solvent is treated as a structureless continuum in which its effects are characterized solely by its dielectric constant ε_r, which for water was taken to be approximately 80.

The partial charges carried by each atom were taken from the OPLS-AA force field. The effects of pH are reflected in the distribution of charges that are assigned to ionizable amino acid residues. The protonation state of these residues for a given pH depends on the pK_a values of the titratable groups on the amino acids. Because the folding of the protein places ionizable residues in environments different from the solvent-exposed one typical of free amino acids, the local electrochemical environment of such residues may be altered. These effects can alter the pK_a values of the side chains relative to those for the corresponding free amino acid. To address these effects, pK_a values were determined from the web server propKa (129, 130)

(http://propka.ki.ku.dk). From these pK_a values and the known pH, the magnitudes of the charges were computed.

Previous methods of accounting for electrostatics involved solving the Poisson-Boltzmann equation using a finite-difference method or a boundary-element approach (131, 132). The advantage of the method used in this work is that it provides a simple and computationally faster method to account for the effect of protein shape on the electrostatic interactions.

3.2.3 Calculation of B_{22}

The calculation of B_{22} for proteins involves sampling angular configurations between two protein molecules, calculating the interaction energies using the models described above, and integrating over all possible configurations. Similar calculations using atomistic models of proteins have been made with more elaborate approaches (133, 134). However, the emphasis in this work is on simulating two protein molecules because B_{22} is by definition a dilute solution property that characterizes the interactions between two molecules. Allowing for the relative orientation of two anisotropic molecules, B_{22} is given as (42, 80)

$$B_{22} = -\frac{1}{16MW^2\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty}\left(e^{-W/kT}-1\right) \times r_{12}^2 dr_{12}\sin\theta d\theta d\phi d\alpha \sin\beta d\beta d\gamma \qquad 3.11$$

where W is the PMF, r_{12} is the center-to-center distance, ϕ and θ are the spherical angles representing the translation of the second molecule relative to the first molecule, and α, β, γ are the Euler angles denoting the rotation of the second molecule. The potential of mean force W represents the interactions between the molecules and is modeled as the sum of the contributions from the non-electrostatic

(van der Waals and solvation forces) W_{ne} and electrostatic W_{elec} contributions. It is a function of both the relative orientation and center-to-center distance between two molecules. Equation 3.11 can be rearranged by decomposing the radial integral and consolidating the angular integrals to give

$$B_{22} = \frac{1}{16MW^2\pi^2} \int_{\Omega} \left[\frac{1}{3} r_c^3 - \int_{r_c}^{\infty} \left(e^{-W/kT} - 1 \right) r_{12}^2 dr_{12} \right] d\Omega \qquad 3.12$$

where the orientation angles are collectively represented by Ω. The first term in brackets in Equation 3.12 represents the excluded volume contribution to B_{22} and is dependent on r_c, the center-to-center distance at contact. The center-to-center distance at contact r_c was approximated by linearly interpolating between the points in the PMF in which the interaction energy transitioned from negative (attraction) to positive (repulsion due to overlap). The second term is the distance-dependent integral, which is referred to as the inner integral I_{in}, and is a direct measure of the energetic contribution to B_{22} due to the interactions between molecules for a specific set of orientation angles Ω. To calculate I_{in} for a given Ω, the cubic splines method was used to interpolate the points of the PMF and then a one-dimensional Gaussian quadrature from the Fortran subroutine library QUADPACK was used to perform the integration. By designating the distance integral as I_{in}, Equation 3.12 can be represented as

$$B_{22} = \frac{1}{16MW^2\pi^2} \left[\int_{\Omega} \frac{1}{3} r_c^3 d\Omega - \int_{\Omega} I_{in} d\Omega \right] \qquad 3.13$$

The key challenge is evaluating the two configuration integrals in equation 3.13. Previous work (82, 84, 135–137) utilized Monte Carlo integration to compute these integrals due to the irregular nature of the integrands. In this approach, *N*

different orientations are randomly sampled from the global configuration space. For each of these configurations, unique values r_c and I_{in} are computed. B_{22} is calculated as the arithmetic average of all r_c and I_{in} computed for all N sampled configurations (138)

$$B_{22} = \frac{1}{16MW^2\pi^2}\frac{\mathrm{V}}{N}\left[\sum_{i=1}^{N}\left(\frac{1}{3}r_c^3\right)_i - \sum_{i=1}^{N}(I_{in})_i\right] \tag{3.14}$$

where V is the hypervolume of the configuration space that is explored. The full angular space for the configuration integral in equation 3.12 is

$$\begin{aligned} V &= \int_{\Omega} d\Omega \\ &= \int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} \sin\theta d\theta d\phi d\alpha \sin\beta d\beta d\gamma = 32\pi^3 \end{aligned} \tag{3.15}$$

The error in B_{22} is determined by the variance of the integrand f as

$$\Delta B_{22} = \pm\frac{1}{16MW^2\pi^2}V\sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N}} \tag{3.16}$$

The angle brackets in equation 3.16 denote the arithmetic mean of the integrand over the N sample points.

Equation 3.16 demonstrates that the rate of convergence using Monte Carlo integration is proportional to $1/\sqrt{N}$. However, the irregular, highly peaked nature of I_{in} due to the Boltzmann weighting of the PMF combined with its multidimensionality make the reliability of Monte Carlo integration questionable. Because there are highly peaked regions in the I_{in} landscape due to highly attractive patch-antipatch interactions, much of the sampling should focus on these regions since they make the most significant contribution to B_{22}. However, Monte Carlo integration

approximates the integral by determining the mean value of I_{in} and multiplying it by the domain of integration. The linear averaging involved in this scheme weights the contributions from each sampled individual configuration equally. Thus, the few tall peaks that occupy a small fraction of the global configuration space skew the linear average, and consequently the integrated value is overestimated. In addition, the configuration space of I_{in} is very large due to its high dimensionality, and therefore there is always uncertainty in the identification of all the peaked regions. The numerical concerns associated with Monte Carlo integration warrant a reexamination of the numerical method for computing the configuration integral, and ultimately B_{22}.

To address these numerical concerns, a hybrid Monte Carlo/patch integration method is proposed and utilized to compute B_{22}. In this approach, B_{22} is broken into the sum of three contributions

$$B_{22} = B_{22}^{Ex} + B_{22}^{PA} + B_{22}^{Background} \tag{3.17}$$

where B_{22}^{Ex} is the excluded volume contribution, B_{22}^{PA} is the contribution from the patch-antipatch interactions, and $B_{22}^{Background}$ is the contribution from the non-patch-antipatch interactions. Configurations with well depths more attractive than -20 kT were considered to be patch-antipatch pairs and were included in B_{22}^{PA}. Conversely, configurations with less attractive wells were incorporated in $B_{22}^{Background}$.

The excluded volume contribution is computed using the Monte Carlo integration method described above

$$B_{22}^{Ex} = \frac{1}{16MW^2\pi^2}\frac{\mathrm{V}}{N}\sum_{i=1}^{N}\left(\frac{1}{3}r_c^3\right)_i \tag{3.18}$$

The landscape of the integrand r_c is expected to be relatively flat since physically the center-to-center distance at contact is expected to have a limited range. Therefore,

random sampling of the global space is sufficient to yield adequate convergence using Monte Carlo integration.

The energetic contribution to B_{22} from the patch-antipatch interactions requires a more careful and detailed integration procedure. Suppose there are N_p known unique patch-antipatch pairs and the central orientation for the i^{th} patch-antipatch pair is $\Omega_i = \{\phi_i, \theta_i, \alpha_i, \beta_i, \gamma_i\}$. Furthermore, suppose that the boundary of the subregion that each patch-antipatch pair occupies is $\pm\Delta$ around its respective central orientation, which represents the size of the patch. The localized patch integration

$$B_{22}^{PA_i} = -\frac{1}{16MW^2\pi^2} \int_{\Omega_{\mathrm{PA}}} I_{in} d\Omega_{\mathrm{PA}} \qquad 3.19$$

is then performed where the domain of local integration is $\Omega_{PA} = [\phi_i\text{-}\Delta, \phi_i\text{+}\Delta]\times[\theta_i\text{-}\Delta, \theta_i\text{+}\Delta]\times[\alpha_i\text{-}\Delta, \alpha_i\text{+}\Delta]\times[\beta_i\text{-}\Delta, \beta_i\text{+}\Delta]\times[\gamma_i\text{-}\Delta, \gamma_i\text{+}\Delta]$. The integration over this subregion is computed using the globally adaptive multidimensional integration routine DCUHRE (139, 140).

Equation 3.19 represents the contribution to B_{22} from an individual patch-antipatch configuration, and the integral in the equation is referred to in what follows as the configurational integral I_{config}. If all patch-antipatch pairs occupy distinct, non-overlapping subregions, then the total contribution to B_{22} of all N_p patch-antipatch pairs is obtained simply by summation

$$B_{22}^{PA} = \sum_{i=1}^{N_p} B_{22}^{PA_i} \qquad 3.20$$

To complete the calculation of the overall B_{22}, the background contributions from the non-patch-antipatch configurations are accounted for by Monte Carlo integration. However, instead of sampling the entire global configuration space,

the subregions that the patch-antipatch pairs occupy are excluded from the sampling. By excluding the configurations that fall in the peaked regions and retaining the ones that are low to moderately peaked, the I_{in} landscape is presumably flatter and therefore Monte Carlo integration would be suitable for calculating this contribution

$$B_{22}^{Background} = \frac{1}{16MW^2\pi^2}\frac{T}{N}\sum_{i=1}^{N}(I_{in})_i \qquad 3.21$$

where T is the size of the hypervolume occupied by the non patch-antipatch interactions.

3.3 Results

3.3.1 Identification of Patch-Antipatch Pairs

The histograms representing the distributions of the well depths for non-electrostatic interactions from 10^4 randomly sampled configurations for lysozyme and chymosin B are shown in Figure 3.2. For each protein, most of the configurations sampled are weakly to moderately attractive, with the modes of the histograms occurring between -1 kT and -2 kT. The linear average of the well depths for each protein is about -3 kT. However, for each of the proteins, relatively few strongly interacting configurations were found, with the strongest configuration identified for chymosin B having a well minimum of -16.6 kT. However, the question that arises is whether all high complementary configurations have been identified from the initial orientation sampling.

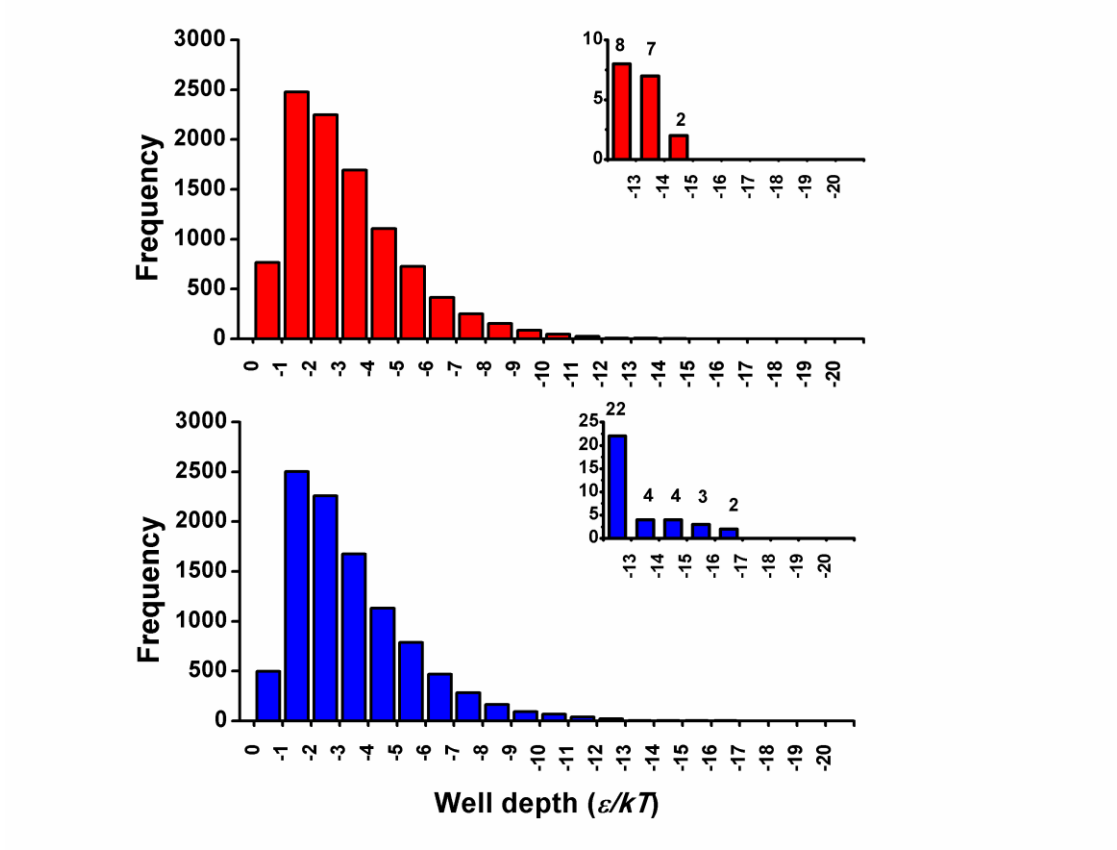

Figure 3.2: Histograms of the distribution of the short-ranged interaction well minima for 10^4 randomly sampled configurations for (■) lysozyme and (■) chymosin B. The inset histograms for each protein are meant to magnify the tails of the distributions.

The effect of more extensive orientation sampling on patch-antipatch identification was subsequently explored. The number of random configurations sampled for each protein was increased by an order of magnitude to 10^5 configurations. The histograms showing the distributions of the well depths for those sampled orientations are presented in Figure 3.3. The shapes of the distributions are similar to those for the 10^4 configurations initially sampled for each protein; however, upon closer inspection of the tails of the histograms, more highly attractive

configurations were identified as a result of the increase in sampling. The results indicate that not all patch-antipatch pairs were identified from the initial 10^4 random configurations sampled. A further increase in the number of orientations sampled was therefore needed to adequately explore the configuration space and identify patch-antipatch pairs. A sampling of 10^6 random configurations was performed and the distributions of the well depths are shown in Figure 3.4. Once again more attractive angular configurations were detected, with the largest well depths being -22.3 *kT* for lysozyme and -27.1 *kT* for chymosin B. Thus, the challenge in the orientation sampling is to be able to properly and effectively sample the tails of the well depth distribution. It is interesting to note that more configurations with ε < -20 *kT* were identified for chymosin B than for lysozyme even though the number of configurations sampled was the same for both proteins. The configurations with ε < -20 *kT* for each protein are presented in Tables 3.5 and 3.6.

The relative frequencies of the well depth distribution for both proteins appear to be independent of the degree of sampling, with the exception of the histogram tails (Tables 3.3 and 3.4). The shapes of the histograms suggest that the well depth variable for both proteins follows approximately a log normal probability distribution. The probability distribution function that describes such a variable is given by (141)

$$f(x) = \frac{1}{x\beta\sqrt{2\pi}} exp\left\{\frac{-(\ln x - \alpha)^2}{2\beta^2}\right\}; 0 < x < \infty \qquad 3.22$$

where *f* is the probability of observation, *x* is the random variable, α is the location parameter, and β is the scale parameter. The α and β parameters characterize the log normal distribution and are computed from the unbiased estimators (141)

$$\alpha = \frac{\sum_{i=1}^{N} \ln x_i}{N}$$
$$\beta = \frac{\sum_{i=1}^{N} (\ln x_i - \alpha)^2}{N - 1} \qquad 3.23$$

Because the log normal distribution is defined only for positive values of *x*, the well depth random variable is defined as the absolute value of ε/kT

$$x \equiv \left| \frac{\varepsilon}{kT} \right| \qquad 3.24$$

The parameters for both proteins were determined from well depths computed from the 10^6 sampled configurations and are shown in Table 3.2. The fits of the distributions from these parameters are shown in Figure 3.5. For both proteins, the log normal probability distribution provides a reasonable description of the relative frequencies of the well depths in the peaked region of the histograms. However, this distribution does not provide a good distribution of the tails of the histograms since the tails do not provide a representative sample of the population of very strongly attractive configurations.

Table 3.2: Log normal probability distribution function parameters for lysozyme and chymosin B estimated from the 10^6 sampled configurations.

Protein	α	β
LYZ	0.975	0.652
CMS	1.067	0.628

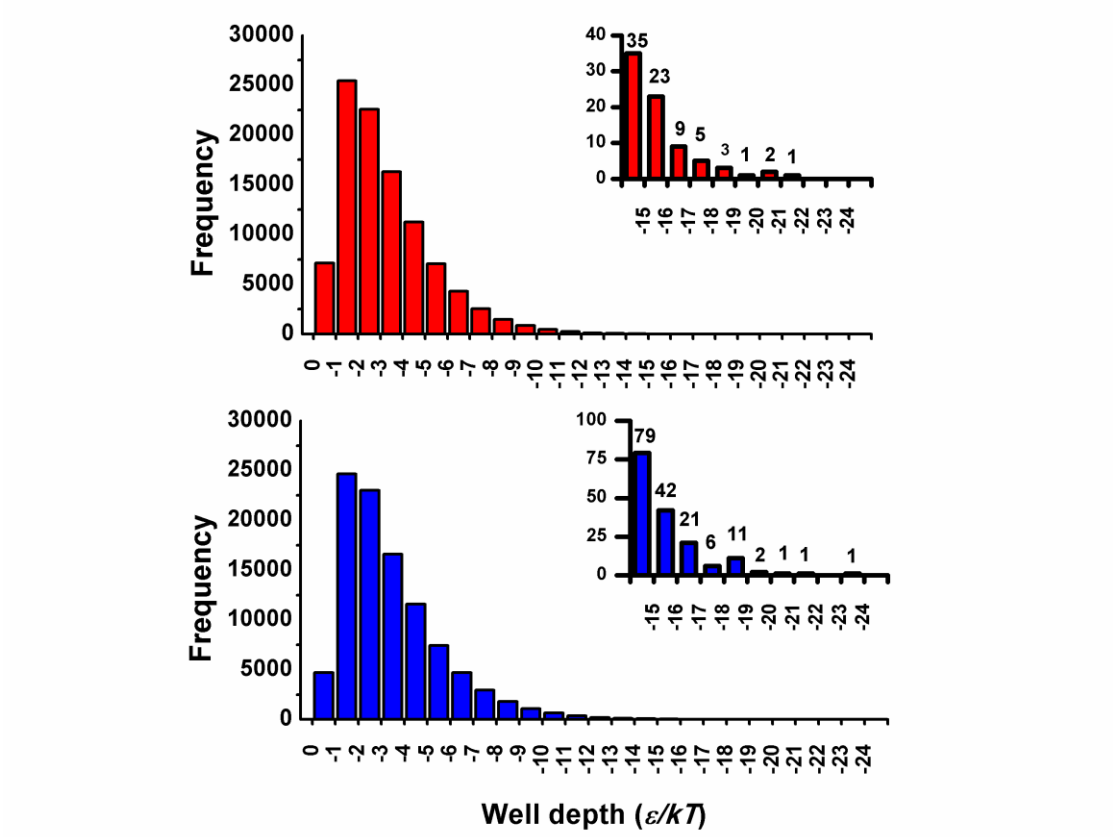

Figure 3.3: Histograms of the distribution of the short-ranged interaction well minima for 10^5 randomly sampled configurations for (■) lysozyme and (■) chymosin B.

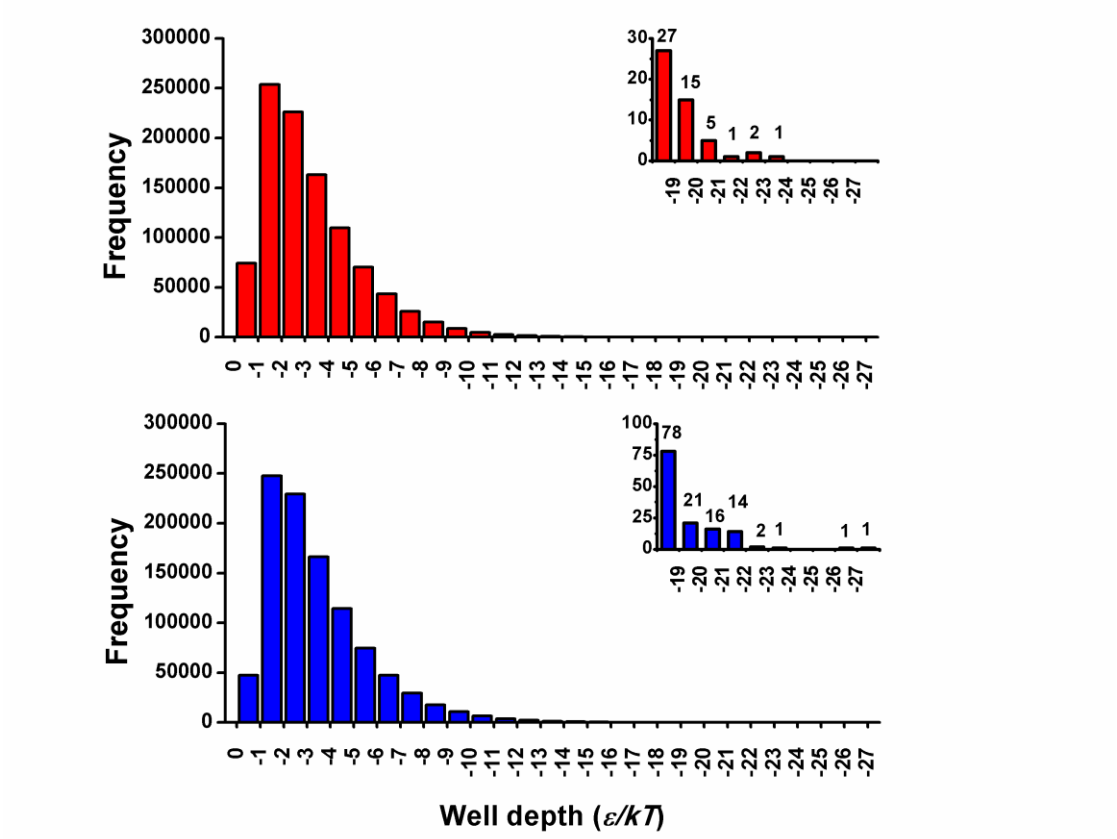

Figure 3.4: Histograms of the distribution of the short-ranged interaction well minima for 10^6 randomly sampled configurations for (■) lysozyme and (■) chymosin B. The largest well depth identified was on the order of -20 *kT*.

Table 3.3: Absolute and relative frequencies of lysozyme well depths for different sampling.

	10^4 Configurations		10^5 Configurations		10^6 Configurations	
Bin	**Frequency**	**Rel. Freq.**	**Frequency**	**Rel. Freq.**	**Frequency**	**Rel. Freq.**
(0,-1)	766	7.65×10^{-2}	7155	7.15×10^{-2}	74269	7.43×10^{-2}
(-1,-2)	2479	2.48×10^{-1}	25450	2.55×10^{-1}	253823	2.54×10^{-1}
(-2,-3)	2248	2.25×10^{-1}	22552	2.26×10^{-1}	225918	2.26×10^{-1}
(-3,-4)	1694	1.69×10^{-1}	16298	1.63×10^{-1}	163073	1.63×10^{-1}
(-4,-5)	1106	1.11×10^{-1}	11249	1.12×10^{-1}	109809	1.10×10^{-1}
(-5,-6)	724	7.24×10^{-2}	7042	7.04×10^{-2}	70141	7.01×10^{-2}
(-6,-7)	415	4.15×10^{-2}	4327	4.33×10^{-2}	43472	4.35×10^{-2}
(-7,-8)	249	2.49×10^{-2}	2543	2.54×10^{-2}	25850	2.59×10^{-2}
(-8,-9)	152	1.52×10^{-2}	1482	1.48×10^{-2}	15077	1.51×10^{-2}
(-9,-10)	83	8.30×10^{-3}	882	8.82×10^{-3}	8567	8.57×10^{-3}
(-10,-11)	44	4.40×10^{-3}	477	4.77×10^{-3}	4665	4.67×10^{-3}
(-11,-12)	23	2.30×10^{-3}	266	2.66×10^{-3}	2482	2.48×10^{-3}
(-12,-13)	8	8.00×10^{-4}	127	1.27×10^{-3}	1360	1.36×10^{-3}
(-13,-14)	7	7.00×10^{-4}	71	7.10×10^{-4}	723	7.23×10^{-4}
(-14,-15)	2	2.00×10^{-4}	35	3.50×10^{-4}	394	3.94×10^{-4}
(-15,-16)	0	0	23	2.30×10^{-4}	188	1.88×10^{-4}
(-16,-17)	0	0	9	9.00×10^{-5}	88	8.80×10^{-5}
(-17,-18)	0	0	5	5.00×10^{-5}	50	5.00×10^{-5}
(-18,-19)	0	0	3	3.00×10^{-5}	27	2.70×10^{-5}
(-19,-20)	0	0	1	1.00×10^{-5}	15	1.50×10^{-5}
(-20,-21)	0	0	2	2.00×10^{-5}	5	5.00×10^{-6}
(-21,-22)	0	0	1	1.00×10^{-5}	1	1.00×10^{-6}
(-22,-23)	0	0	0	0	2	2.00×10^{-6}
(-23,-24)	0	0	0	0	1	1.00×10^{-6}

Table 3.4: Absolute and relative frequencies of chymosin B well depths for different sampling.

	10^4 Configurations		10^5 Configurations		10^6 Configurations	
Bin	**Frequency**	**Rel. Freq.**	**Frequency**	**Rel. Freq.**	**Frequency**	**Rel. Freq.**
(0,-1)	498	4.98×10^{-2}	4700	4.70×10^{-2}	47522	4.75×10^{-2}
(-1,-2)	2502	2.50×10^{-1}	24660	2.47×10^{-1}	247585	2.48×10^{-1}
(-2,-3)	2261	2.26×10^{-1}	23024	2.30×10^{-1}	229386	2.29×10^{-1}
(-3,-4)	1675	1.68×10^{-1}	16587	1.66×10^{-1}	166303	1.66×10^{-1}
(-4,-5)	1129	1.13×10^{-1}	11564	1.16×10^{-1}	114250	1.14×10^{-1}
(-5,-6)	787	7.87×10^{-2}	7420	7.42×10^{-2}	74522	7.45×10^{-2}
(-6,-7)	467	4.67×10^{-2}	4717	4.72×10^{-2}	47298	4.73×10^{-2}
(-7,-8)	284	2.84×10^{-2}	2967	2.97×10^{-2}	29520	2.95×10^{-2}
(-8,-9)	164	1.64×10^{-2}	1802	1.80×10^{-2}	17750	1.78×10^{-2}
(-9,-10)	92	9.20×10^{-3}	1066	1.07×10^{-3}	10703	1.07×10^{-2}
(-10,-11)	68	6.80×10^{-3}	637	6.37×10^{-3}	6611	6.61×10^{-3}
(-11,-12)	37	3.70×10^{-3}	373	3.73×10^{-3}	3783	3.78×10^{-3}
(-12,-13)	22	2.20×10^{-3}	190	1.90×10^{-3}	2077	2.08×10^{-3}
(-13,-14)	4	4.00×10^{-3}	103	1.03×10^{-3}	1217	1.22×10^{-3}
(-14,-15)	4	4.00×10^{-4}	79	7.90×10^{-4}	644	6.44×10^{-4}
(-15,-16)	3	3.00×10^{-4}	42	4.20×10^{-4}	381	3.81×10^{-4}
(-16,-17)	0	0	21	2.10×10^{-4}	202	2.02×10^{-4}
(-17,-18)	0	0	6	6.00×10^{-5}	112	1.12×10^{-4}
(-18,-19)	0	0	11	1.10×10^{-4}	78	7.80×10^{-5}
(-19,-20)	0	0	2	2.00×10^{-5}	21	2.10×10^{-5}
(-20,-21)	0	0	1	1.00×10^{-5}	16	1.60×10^{-5}
(-21,-22)	0	0	1	1.00×10^{-5}	14	1.40×10^{-5}
(-22,-23)	0	0	0	0	2	2.00×10^{-6}
(-23,-24)	0	0	1	1.00×10^{-5}	1	1.00×10^{-6}
(-24,-25)	0	0	0	0	0	0
(-25,-26)	0	0	0	0	0	0
(-26,-27)	0	0	0	0	1	1.00×10^{-6}
(-27,-28)	0	0	0	0	1	1.00×10^{-6}

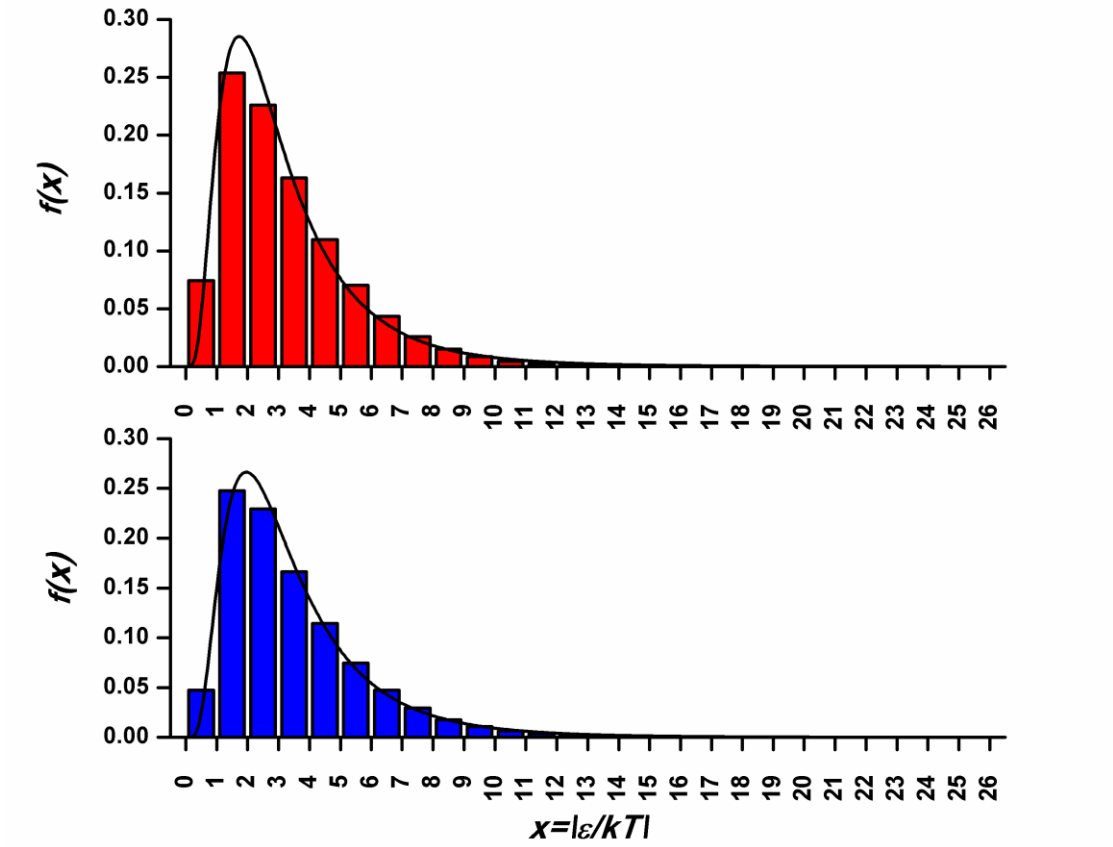

Figure 3.5: Comparison of relative frequencies of (■) lysozyme and (■) chymosin B absolute well depths with respective fits from (–) log normal probability distribution function. The fits are based on parameter values of $\alpha = 0.975$, $\beta = 0.652$ for lysozyme and $\alpha = 1.067$, $\beta = 0.628$ for chymosin B.

Table 3.5: Ten most attractive angular configurations for short-range interactions of lysozyme identified from 10^6 randomly sampled orientations.

#	ϕ	θ	α	β	γ	ε/kT
1	1.191	0.975	5.309	1.505	4.201	-22.28
2	2.213	1.292	2.358	0.763	1.862	-20.70
3	3.513	1.002	4.286	0.681	1.800	-20.05
4	3.301	1.228	0.607	1.424	2.635	-23.32
5	5.211	0.575	4.674	1.283	5.123	-20.62
6	1.110	1.145	4.566	1.696	4.923	-21.09
7	3.276	1.010	2.392	0.831	2.138	-20.18
8	2.338	1.019	0.422	0.110	2.240	-22.36
9	1.776	2.791	4.207	1.594	4.237	-20.94
10	1.565	2.830	2.183	2.465	0.545	-19.98

Table 3.6: Thirty-five most attractive angular configurations for short-range interactions of chymosin B identified from 10^6 randomly sampled orientations.

#	ϕ	θ	α	β	γ	ε/kT
1	4.298	2.105	1.138	2.030	1.845	-23.63
2	4.074	1.347	4.930	2.051	2.096	-26.12
3	4.299	0.977	5.746	0.751	0.217	-20.97
4	3.909	1.157	5.566	0.448	5.185	-20.43
5	4.399	2.511	0.031	1.757	5.561	-20.94
6	4.084	0.746	0.395	2.584	5.010	-20.35
7	3.607	1.295	2.076	1.657	0.177	-21.27
8	5.884	1.458	3.317	0.997	5.614	-20.03
9	5.888	1.565	1.271	2.518	2.597	-20.58
10	3.599	2.140	5.977	0.252	0.706	-21.38
11	6.025	1.461	0.227	1.556	0.962	-20.77
12	6.233	1.303	4.536	1.702	1.704	-20.68
13	0.048	2.491	2.023	0.158	0.812	-20.59
14	3.026	1.342	2.839	2.825	0.116	-27.36
15	3.016	1.126	6.121	2.775	3.114	-22.30
16	1.168	2.841	3.861	1.119	2.824	-20.16
17	2.722	2.274	0.056	2.153	0.649	-21.12
18	2.694	0.945	6.057	2.047	5.737	-20.93
19	0.567	0.818	0.489	1.655	5.215	-21.06
20	0.994	2.640	3.606	2.428	4.930	-21.95
21	0.447	1.413	3.793	1.225	5.585	-21.55
22	0.512	1.594	2.939	1.809	0.381	-21.99
23	0.577	1.249	2.706	1.075	5.117	-21.63
24	1.122	2.518	1.715	1.153	0.248	-21.99
25	2.476	2.103	1.461	0.973	5.029	-20.36
26	0.870	2.331	3.964	1.604	3.330	-21.08
27	2.096	2.440	3.645	1.933	2.479	-21.26
28	2.495	1.400	0.534	1.085	4.471	-20.62
29	1.164	2.290	6.260	2.271	3.240	-20.56
30	1.802	2.312	4.456	1.532	3.731	-20.83
31	2.204	2.035	1.788	1.511	0.062	-20.00
32	1.644	2.268	0.733	1.815	2.108	-21.42
33	0.969	1.238	4.187	2.865	5.352	-21.18
34	1.084	1.130	2.555	1.397	3.987	-22.94
35	1.744	2.180	1.114	1.999	4.534	-21.30

Another issue that arises in the random orientation sampling is whether the patch-antipatch pairs identified are sampled at the optimally aligned orientation that represents the true well depth. To explore this, the first highly attractive configuration for lysozyme in Table 3.5 (entry 1) was selected as a test case. Each individual orientation angle for this configuration was varied within ± 0.20 radian around the central orientation while holding the other angles fixed at their respective central values. The response in the well depth to these variations is shown in Figure 3.6. If the central orientation were at the true optimum alignment, the minimum in the interaction energy profile would occur at a variation of 0 radian. However, Figure 3.6 indicates that this configuration is in fact not the optimally aligned one. A similar result was obtained by Hloucha et al. (see Figure 2 in reference (78)). Thus, it can be inferred that the orientations that are identified in the random sampling may not, in general, be at their energy minima.

To approximate the optimal alignment that leads to the energy minimum, a local sampling around the central orientation for each patch-antipatch configuration was performed. This local sampling entailed sampling 10^5 random orientations confined within the limits of ± 0.10 radian around each angle of the central orientation. This local sampling procedure was performed for several of the patch-antipatch pairs for lysozyme and chymosin B and the fined-tuned orientations are shown in Tables 3.7 and 3.8, respectively. When compared with the initial sampled configurations, the differences in each of the angles are small, yet these small variations can lead to substantial changes in the well depths. The most significant case is for the first patch-antipatch pair for lysozyme, in which a configuration with a well depth on the order of -40 *kT* was identified. Similar results can be seen for the chymosin B patch-antipatch

configurations in which the refinement led to a well-depth as high as -39 *kT*. The interactions between patch-antipatch pairs for proteins are quite sensitive to small perturbations in orientation; a slight change in the alignment can lead to a significant change in the attraction. This sensitivity indicates that the region representing a given patch-antipatch interaction is only a minute fraction of the global angular space and that these patch regions are very small.

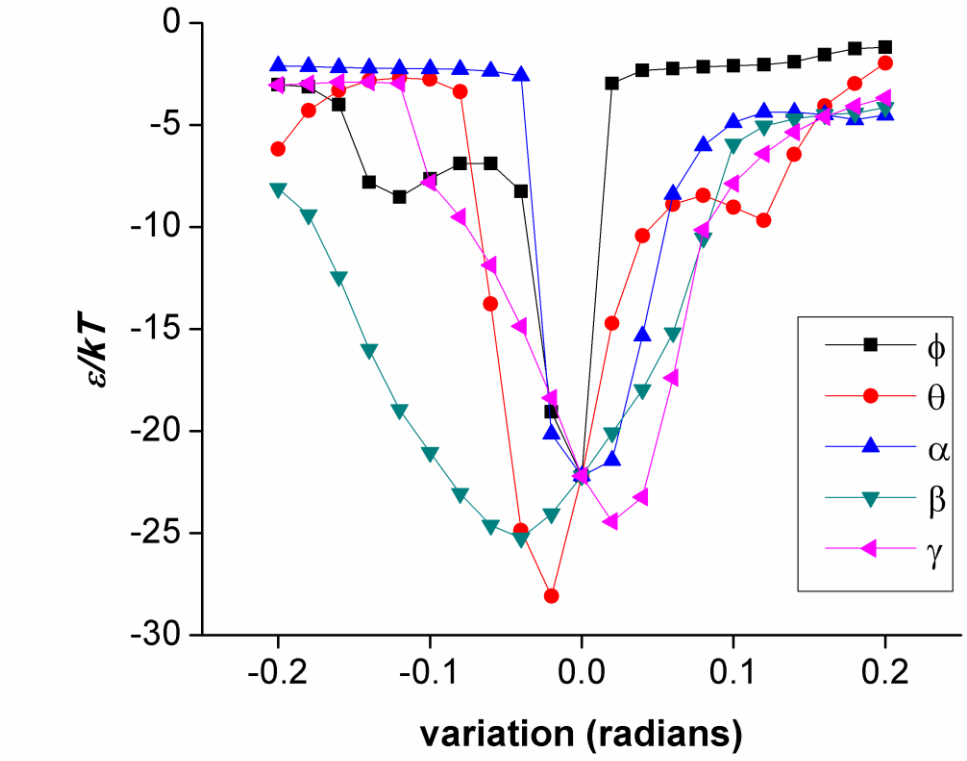

Figure 3.6: Well depth as a function of the angles for the orientation listed in entry 1 of Table 3.5. The largest change occurs when θ is decreased by -0.02 radian, which indicates that the originally sampled orientation is not the optimum alignment.

Table 3.7: Refined orientations for lysozyme identified from the local sampling in ±0.10 radian around the central orientation in Table 3.4. The resulting angular configurations are significantly more attractive than the originally sampled orientations.

#	ϕ	θ	α	β	γ	ε/kT	Initial ε/kT
1	1.184	0.907	5.274	1.540	4.158	-40.48	-22.28
2	2.203	1.273	2.269	0.747	1.897	-24.30	-20.70
3	3.478	1.050	4.245	0.773	1.856	-26.34	-20.05
4	3.244	1.190	0.612	1.369	2.537	-29.06	-23.32
5	5.260	0.597	4.640	1.189	5.123	-25.05	-20.62
6	1.085	1.125	4.504	1.754	4.905	-24.97	-21.09
7	3.244	0.963	2.332	0.919	2.235	-26.04	-20.18
8	2.355	0.987	0.505	0.134	2.163	-22.88	-22.36
9	1.808	2.754	4.295	1.621	4.184	-24.76	-20.94
10	1.533	2.818	2.122	2.469	0.608	-24.60	-19.98

Table 3.8: Refined orientations for patch-antipatch pairs for chymosin B identified from the local sampling in ±0.10 radian around the central orientation in Table 3.5.

#	ϕ	θ	α	β	γ	ε/kT	Initial ε/kT
1	4.322	2.148	1.212	2.105	1.919	-26.88	-23.63
2	4.028	1.379	4.833	2.064	2.040	-39.11	-26.12
3	4.311	0.940	5.784	0.708	0.192	-26.90	-20.97
4	3.921	1.144	5.648	0.446	5.110	-25.38	-20.43
5	4.480	2.548	0.106	1.806	5.660	-25.22	-20.94
6	4.110	0.764	0.494	2.622	5.109	-24.98	-20.35
7	3.555	1.243	1.987	1.715	0.193	-25.83	-21.27
8	5.853	1.452	3.268	1.022	5.663	-28.84	-20.03
9	5.930	1.563	1.355	2.525	2.589	-22.89	-20.58
10	3.593	2.122	5.890	0.179	0.769	-24.69	-21.38
11	5.994	1.448	0.177	1.526	0.999	-25.66	-20.77
12	6.214	1.347	4.517	1.614	1.754	-27.06	-20.68
13	0.036	2.515	1.924	0.195	0.861	-25.32	-20.59
14	2.972	1.340	2.831	2.771	0.208	-35.80	-27.36
15	2.954	1.126	6.109	2.861	3.213	-33.06	-22.30
16	1.267	2.860	3.836	1.106	2.823	-24.05	-20.16
17	2.691	2.312	-0.007	2.146	0.650	-25.77	-21.12
18	2.756	0.896	6.132	2.146	5.686	-25.16	-20.93
19	0.580	0.805	0.538	1.705	5.241	-23.76	-21.06
20	0.950	2.620	3.512	2.428	4.868	-28.02	-21.95
21	0.453	1.426	3.781	1.324	5.642	-28.02	-21.55
22	0.483	1.567	2.881	1.763	0.360	-26.92	-21.99
23	0.536	1.249	2.618	1.087	5.129	-27.48	-21.63
24	1.073	2.518	1.677	1.091	0.186	-27.09	-21.99
25	2.451	2.091	1.411	0.955	5.128	-22.51	-20.36
26	0.872	2.310	3.968	1.562	3.338	-28.43	-21.08
27	2.090	2.402	3.629	1.895	2.504	-28.19	-21.26
28	2.495	1.432	0.521	1.022	4.446	-23.66	-20.62
29	1.065	2.278	6.198	2.259	3.325	-23.99	-20.56
30	1.727	2.350	4.358	1.544	3.830	-25.02	-20.83
31	2.230	1.979	1.713	1.610	-0.037	-24.60	-20.00
32	1.743	2.293	0.832	1.828	2.027	-27.89	-21.42
33	0.918	1.250	4.125	2.766	5.301	-25.59	-21.18
34	1.097	1.074	2.557	1.297	3.937	-27.08	-22.94
35	1.744	2.137	1.120	2.057	4.596	-38.82	-21.30

3.3.2 Calculation of B_{22}

The excluded volume contribution to B_{22} was first calculated using the Monte Carlo integration approach, where the global configuration space was randomly sampled and the center-to-center contact distance was calculated for each configuration sampled. The final numerical values are presented and compared with the theoretical hard sphere B_{22} values based on the equivalent sphere diameter of each protein (Table 3.9). The values are comparable in magnitude, but the value is higher when the protein is modeled atomistically than when it is represented as a hard sphere. This result is consistent with the findings of Neal and Lenhoff (142), who demonstrated that the magnitude of the excluded volume contribution depends on the level of structural detail of the protein considered. They found that when protein molecules are resolved atomistically, the excluded volume contribution can be as much as 40% greater than the result obtained when the protein is modeled as an ideal sphere. The disparity in the excluded volume B_{22} predicted by the two representations is attributed to the roughness of the protein surface. In the atomistic case, groups of atoms that protrude on the surface limit the closeness of approach when two protein molecules come in contact. Consequently, there is a greater effective center-to-center distance leading to a greater excluded volume contribution to B_{22} than for the ideal sphere representation.

The contributions from the molecular interactions were subsequently included in the B_{22} calculations. As a starting point, only the short-ranged non-electrostatic interactions were considered in the calculation. The B_{22} values based on the excluded volume and van der Waals interactions calculated using Monte Carlo integration are presented in Table 3.9. There are two symptoms of computational challenges evident in these results. First, the B_{22} values computed for each protein do

not appear to converge even after 10^6 configurations are sampled and second, the final computed values are orders of magnitude greater than typical experimentally measured values. To understand why the magnitudes of the computed B_{22} value are so large, the histograms of the computed I_{in} values for the 10^6 configurations sampled for lysozyme and chymosin B are shown in Figures 3.7 and 3.8, respectively. The range of I_{in} computed spans many orders of magnitude, from as low as 1×10^1 Å^3 to as high as 1×10^{14} Å^3. This wide range is due to the nature of the integrand in I_{in}; configurations with large well depths are magnified as a result of the Boltzmann weighting of the PMF. Because the Monte Carlo integration method takes the linear average of all I_{in} values, the few configurations that have extremely large I_{in} values contribute disproportionately to the mean value of I_{in} and consequently the integral is overestimated. Thus, the linear averaging used in Monte Carlo integration cannot provide an accurate numerical estimate of the configurational integral to compute B_{22}.

Table 3.9: B_{22} calculated from 10^6 randomly sampled configurations based on excluded volume contribution and both excluded volume and short-range attraction. The σ value is the sphere equivalent diameter determined from the empirical correlation of Neal and Lenhoff (142)**. The error in the Monte Carlo estimate is calculated from equation 3.16.**

Protein	B_{22}^{Ex} ($\times10^4$ ml mol/g^2)	σ (Å)	B_{22}^{HS} ($\times10^4$ ml mol/g^2)	Total B_{22} (ml mol/g^2)
LYZ	2.997 ± 0.001	32.0	2.02	-0.24 ± 0.07
CMS	1.3990 ± 0.0005	42.9	0.783	-0.95 ± 0.59

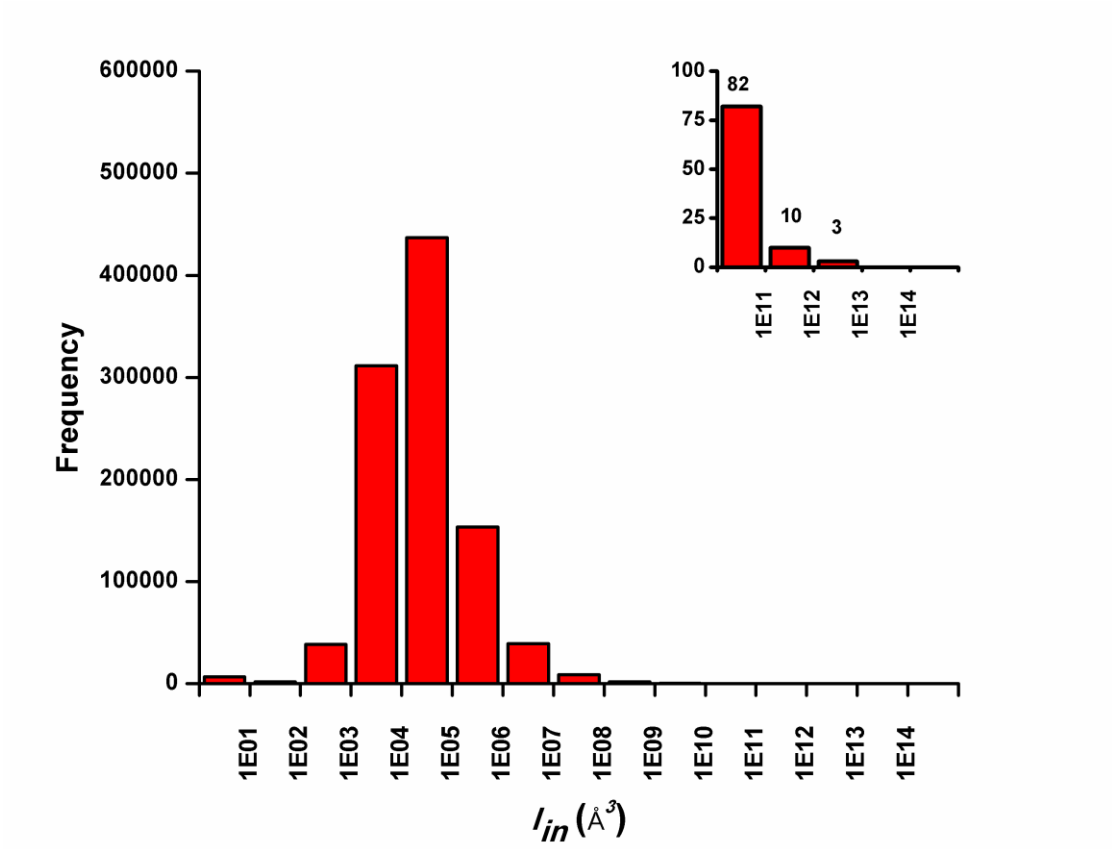

Figure 3.7: Histogram of the computed I_{in} for the 10^6 randomly sampled configurations for lysozyme. The inset shows an enlarged view of the high-I_{in} tail of the distribution.

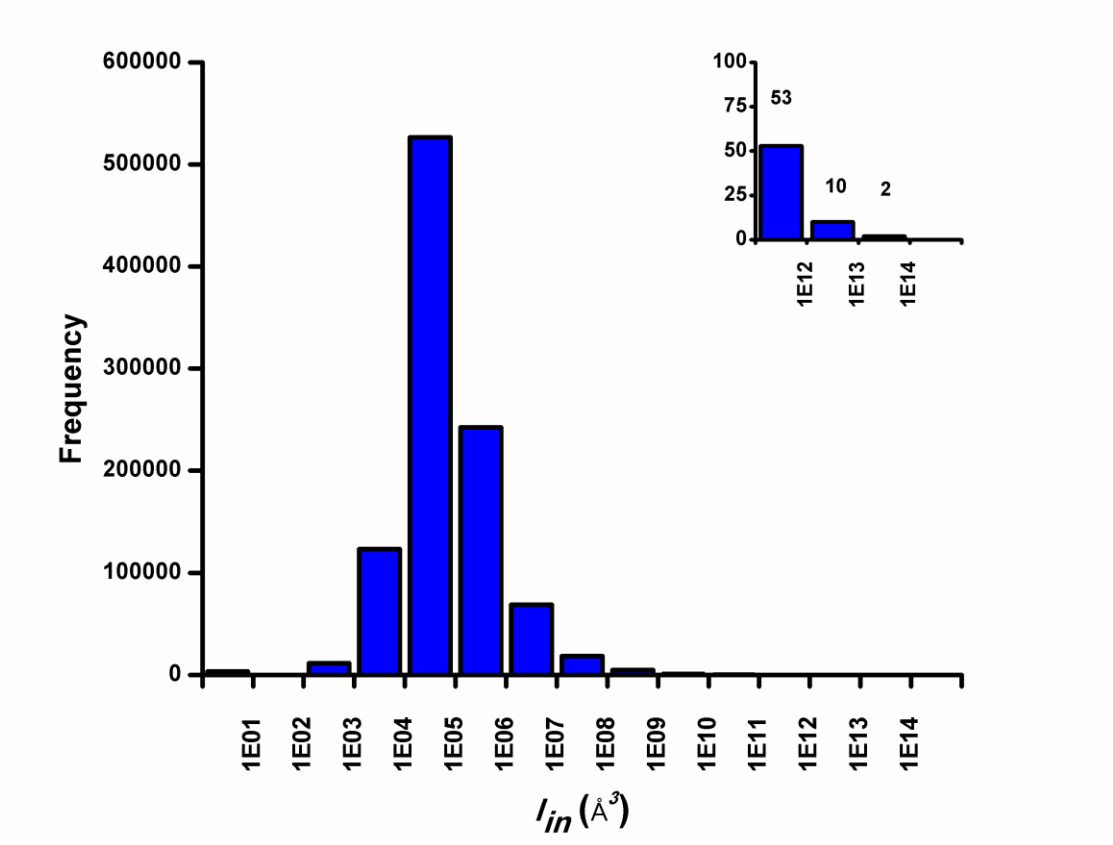

Figure 3.8: Histogram of the computed I_{in} for the 10^6 randomly sampled configurations for chymosin B. The inset shows an enlarged view of the high-I_{in} tail of the distribution.

To address the numerical issues associated with Monte Carlo integration, a hybrid Monte Carlo/patch integration method was utilized to compute B_{22}. Different configurations of varying strengths for each protein, which included the highly attractive patch-antipatch pairs, were selected and integrated within the limits of $\pm\Delta$ radians around the central orientations using the DCUHRE integration routine. The Δ parameter directly determines the size of the subregion over which the interaction energy is strongly attractive before it decays, and therefore represents the patch size. However, increasing Δ enlarges the hypervolume of the integration and consequently I_{config} increases monotonically. This can be seen in Figure 3.9 in which I_{config} is computed using the integration routine for the fourth lysozyme patch-antipatch pair listed in Table 3.7. Thus, the size of a patch-antipatch pair cannot be determined directly from the dependence of I_{config} on Δ. To estimate an appropriate value for Δ, I_{config} was normalized by the hypervolume of integration v_0 and the result was plotted against Δ. The integration hypervolume v_0 is directly related to Δ through equation 3.15 with the appropriate integration limits. This normalization filters the effect of the hypervolume size in the integration and gives some indication of how the rate of growth of the integral decreases due to the decay of the interaction potential. Figure 3.10 shows that at approximately $\Delta = 0.10$ radian the normalized integral decreases by a factor of about 10 from its highest value. This significant decay would indicate that the effects of the attraction on I_{config} dissipates and therefore $\Delta = 0.10$ radian characterizes the patch size for this particular orientation. Although the patch size for each patch-antipatch pair may be different, $\Delta = 0.10$ radian was chosen as a universal value for all integration of patch-antipatch pairs in order to simplify further computations.

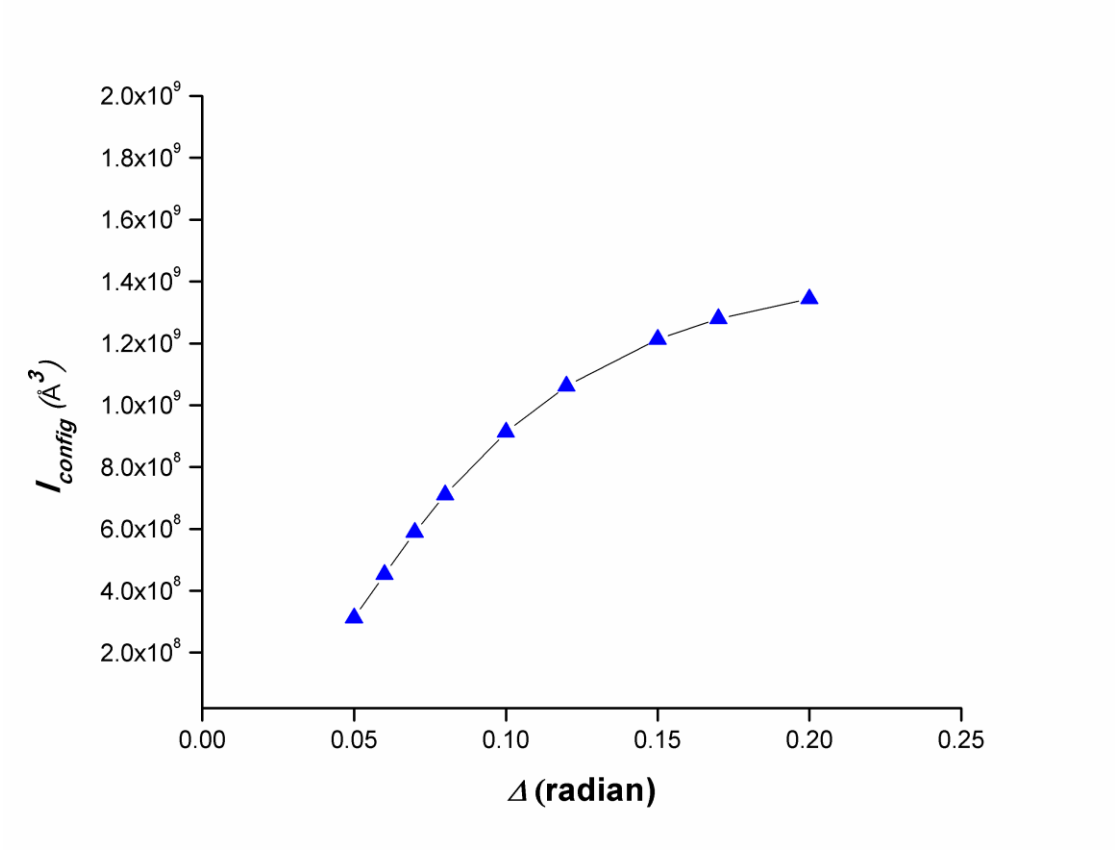

Figure 3.9: I_{config} computed from the DCUHRE routine as a function of Δ for lysozyme patch-antipatch pair 4 in Table 3.7. I_{config} increases monotonically as Δ increases due to the increase in the hypervolume of the integration.

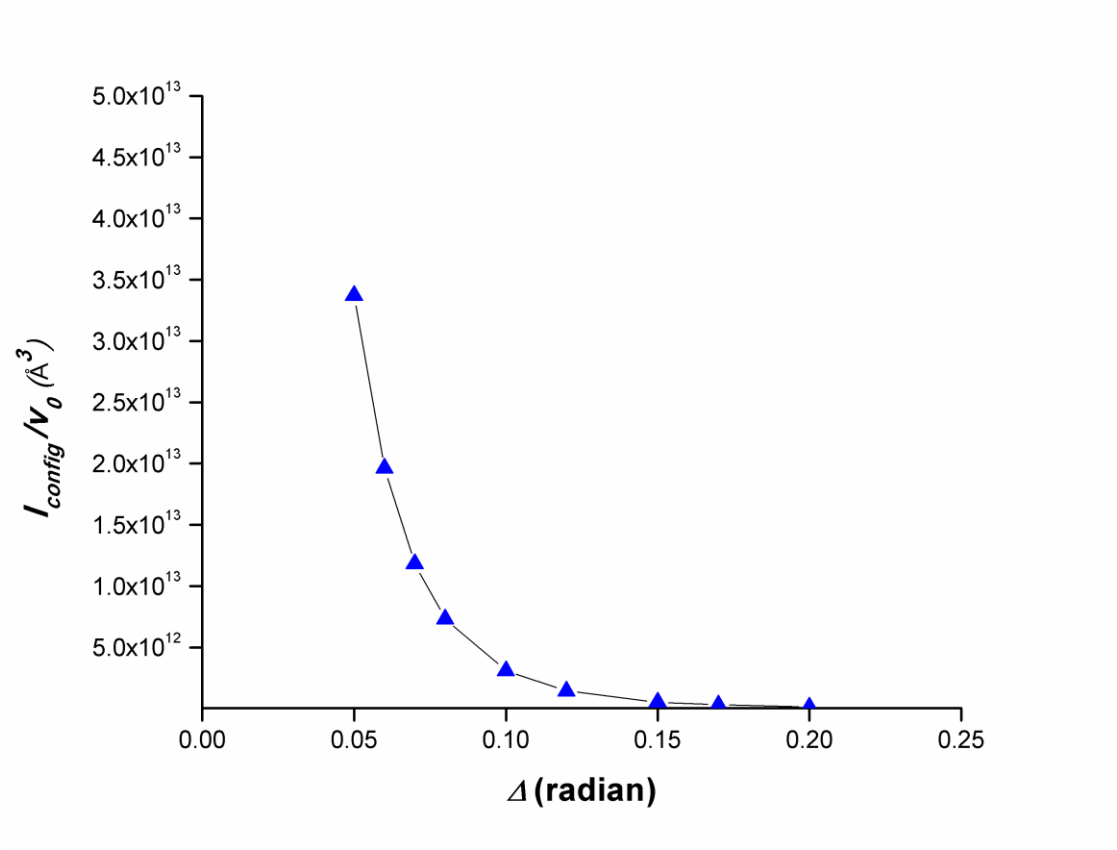

Figure 3.10: I_{config} normalized by the volume of integration v_0 as a function of Δ for lysozyme patch-antipatch pair 4 in Table 3.7. The normalized integral decreases as Δ is increased.

The localized integration was performed for the cases where only short-range interactions were considered and where both short-range and electrostatic interactions were included. The local integration represents the contribution of these individual patch-antipatch regions to B_{22}. The results of the integration, in units of $Å^3$, are plotted as a function of the total well depth (short-range and electrostatics) of the configuration with the lowest energy minimum for each patch-antipatch pair in Figure 3.11 for lysozyme and in Figure 3.12 for chymosin B. The results were fitted using an

exponential regression model. There is some scatter in the results, which is to be expected since there is variability in the shapes of the energy profiles and uncertainty in the energy minima, and not all patch-antipatch interactions may be confined completely within ± 0.10 radians. Despite this scatter, there is a consistent, well-behaved relationship in which the computed localized configurational integral I_{config} increases as the strength of interaction increases. The correlations for both proteins are similar; both regressions have comparable preexponential factors, but the regressed fit for chymosin B possesses a less negative exponent. The trends demonstrate that configurations with stronger attractions make a greater contribution to B_{22}. Electrostatics can tune a configuration's contribution to B_{22} by either reducing its attraction through addition of repulsion or enhancing its attraction. Repulsive electrostatics result in decreasing I_{config} whereas attractive electrostatics corresponds to increasing I_{config}. By knowing a patch-antipatch pair's well depth, Figures 3.11 and 3.12 can be used to empirically estimate its contribution to B_{22} without the need for performing a detailed integration using the DCUHRE routine.

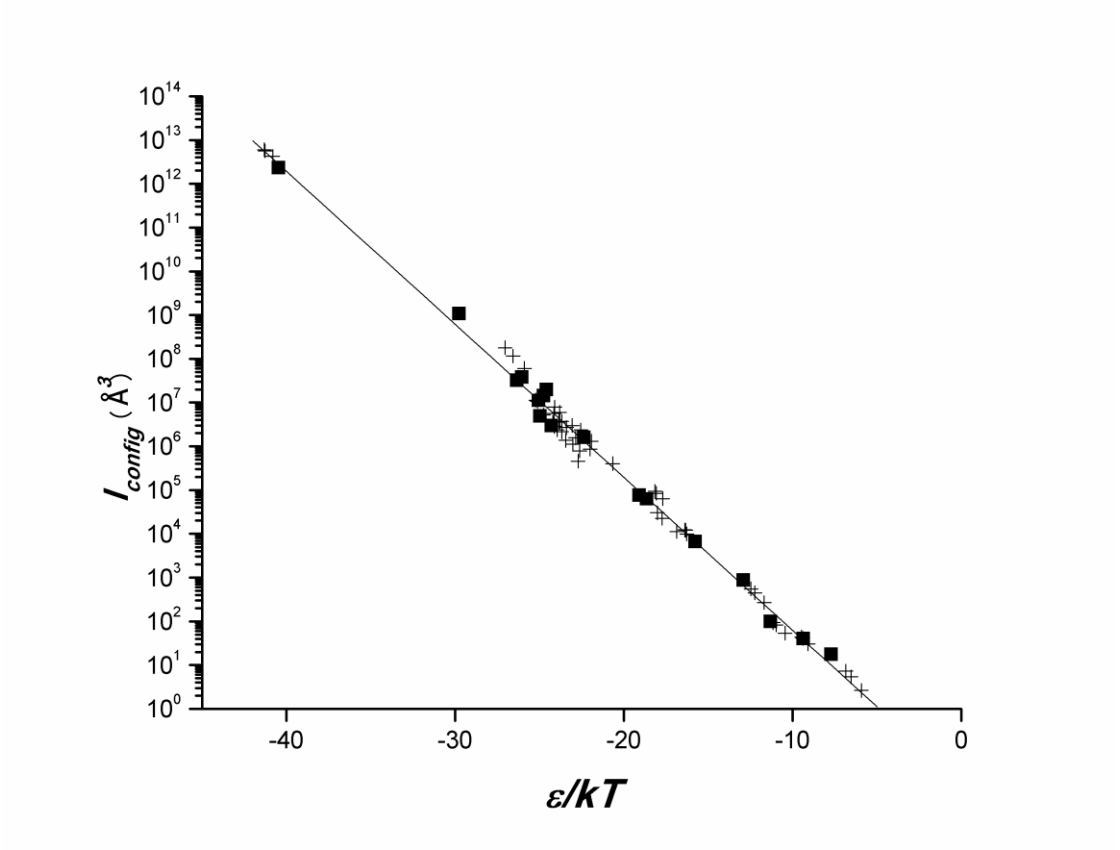

Figure 3.11: Plot of the localized configuration integration for lysozyme as a function of the total well depth. I_{config} was computed based on (+) short-range interactions alone and (■) short-range interactions with electrostatics. Integration was performed within the limits of ±0.10 radian around the central orientation of each patch-antipatch configuration using the DCUHRE routine. The regressed curve is $F(x)=0.0200\ exp(-0.805x)$, $R^2=0.9941$.

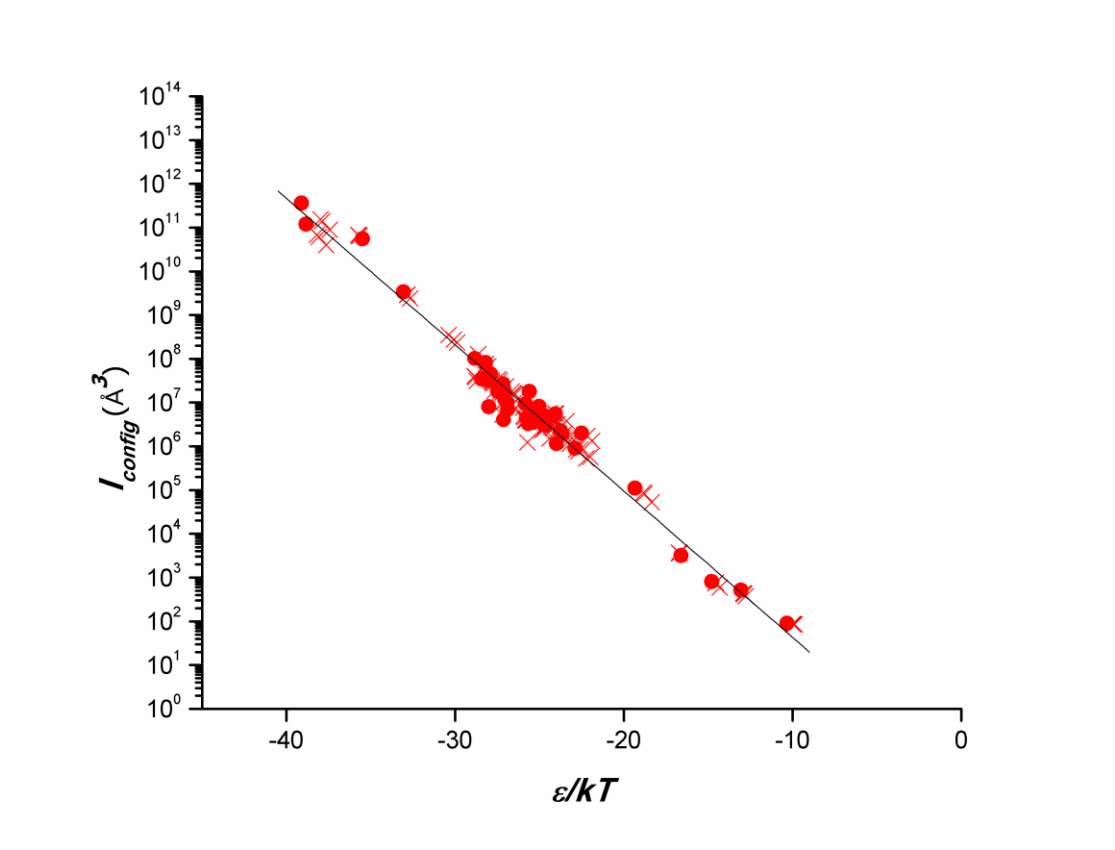

Figure 3.12: Plot of the localized configuration integration for chymosin B as a function of the total well depth. I_{config} was computed based on (×) short-range interactions alone and (●) short-range interactions with electrostatics. Integration was performed within the limits of ±0.10 radian around the central orientation of the configurations. The regressed curve is $F(x)=0.0195\ exp(-0.770x)$, $R^2=0.9806$.

The B_{22} contributions for the individual lysozyme pairs are presented in Table 3.10, which include contributions resulting from the short-range interactions alone and from both short-range and electrostatic interactions. The results show that inclusion of electrostatics significantly reduces the attractions for most of the patch-antipatch pairs, as indicated by the increased contribution to B_{22} as the ionic strength increases. This indicates the electrostatic interactions in the configurations

represented by the patch-antipatch pairs for lysozyme are predominantly repulsive. However, two of the ten patch-antipatch pairs with $\varepsilon < -20\ kT$ (entries 1 and 8 in Table 3.10) appear to have an increased B_{22} contribution when electrostatics are incorporated, which suggests that this particular pair possesses attractive electrostatics. It is interesting to note that the configuration in entry 1 of Table 3.10, which is the -40 *kT* patch identified after local sampling refinement, has a contribution that is orders of magnitude greater than those of the other pairs. The contribution by this lone patch-antipatch pair would therefore overwhelm the contributions from the other patches if they were summed together.

The trends in the B_{22} contributions from the chymosin B patch-antipatch pairs appear to show different behavior. Although some of the pairs exhibit repulsive electrostatics, some of the configurations show the opposite trend. In fact, almost one-third of the patch-antipatch pairs with $\varepsilon < -20\ kT$ were found to have increased B_{22} contributions when electrostatics are incorporated. Thus, chymosin B was not only found to have a larger number of strongly attractive patch-antipatch pairs than lysozyme, but also a greater percentage of the attractive patch-antipatch pairs possess attractive electrostatics. For a few configurations, the differences are within the errors of integration and may be statistically insignificant, but the results indicate that in general there are smaller repulsive electrostatic effects for chymosin B than for lysozyme. Given that the local integration for each pair occurs over a very small portion of the global configuration space, the B_{22} contributions from these patch-antipatch pairs are quite remarkable.

Table 3.10: Contributions to B_{22} from individual patch-antipatch pairs for lysozyme based on short-range non-electrostatic interaction energies alone and with electrostatics at pH 7. The contributions were determined by integrating within ±0.10 radian around the central orientation using the DCUHRE integration routine.

Non-electrostatics		pH 7, 0.10M		pH 7, 0.20M		pH 7, 0.30M	
ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)
-40.48	-4.37×10^{1}	-40.82	-8.00×10^{1}	-41.27	-1.06×10^{2}	-41.31	-1.13×10^{2}
-24.31	-5.54×10^{-5}	-22.00	-8.51×10^{-6}	-22.71	-1.60×10^{-5}	-23.01	-2.10×10^{-5}
-26.34	-6.06×10^{-4}	-22.62	-1.48×10^{-5}	-23.69	-4.03×10^{-5}	-24.21	-6.65×10^{-5}
-29.77	-2.01×10^{-2}	-25.89	-1.12×10^{-3}	-26.57	-2.17×10^{-3}	-27.05	-3.35×10^{-3}
-25.05	-2.09×10^{-4}	-24.77	-9.72×10^{-5}	-25.13	-2.01×10^{-4}	-25.21	-2.13×10^{-4}
-24.97	-9.14×10^{-5}	-23.45	-2.58×10^{-5}	-23.95	-4.45×10^{-5}	-24.12	-5.32×10^{-5}
-26.05	-7.08×10^{-4}	-22.86	-2.92×10^{-5}	-23.70	-6.90×10^{-5}	-24.14	-1.05×10^{-4}
-22.36	-3.02×10^{-5}	-23.39	-5.02×10^{-5}	-23.82	-6.68×10^{-5}	-23.85	-6.90×10^{-5}
-24.76	-2.69×10^{-4}	-23.05	-5.60×10^{-5}	-23.82	-1.11×10^{-4}	-24.11	-1.47×10^{-4}
-24.60	-3.69×10^{-4}	-20.66	-7.46×10^{-6}	-21.93	-2.45×10^{-5}	-22.55	-4.40×10^{-5}
-19.09	-1.43×10^{-6}	-17.70	-1.19×10^{-6}	-18.09	-1.58×10^{-6}	-18.16	-1.74×10^{-6}
-18.65	-1.17×10^{-6}	-16.87	-2.12×10^{-7}	-17.73	-4.24×10^{-7}	-18.02	-5.67×10^{-7}
-15.77	-1.24×10^{-7}	-16.27	-1.85×10^{-7}	-16.33	-2.26×10^{-7}	-16.39	-2.30×10^{-7}
-12.91	-1.62×10^{-8}	-11.69	-5.02×10^{-9}	-12.24	-8.39×10^{-9}	-12.46	-1.03×10^{-8}
-11.32	-1.84×10^{-9}	-10.46	-1.00×10^{-9}	-10.98	-1.54×10^{-9}	-11.16	-1.74×10^{-9}
-9.37	-7.58×10^{-10}	-9.09	-5.72×10^{-10}	-9.39	-7.64×10^{-10}	-9.46	-8.16×10^{-10}
-7.72	-3.30×10^{-10}	-5.92	-4.95×10^{-11}	-6.54	-1.01×10^{-10}	-6.84	-1.35×10^{-11}

Table 3.11: B_{22} contributions from individual patch-antipatch pairs for chymosin B based on short-range non-electrostatic interaction energies alone and with electrostatic interactions at pH 5. The contributions were determined by integrating within ±0.10 radian around the central orientation using the DCUHRE integration routine.

Non-electrostatics		pH 5, 0.10M		pH 5, 0.30M		pH 5, 0.40M	
ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)	ε/kT	B_{22}^{PA} (mol ml/g^2)
-26.88	-2.10×10^{-5}	-27.66	-5.65×10^{-5}	-27.12	-3.11×10^{-5}	-27.00	-2.93×10^{-5}
-39.11	-1.08×10^{0}	-37.43	-2.67×10^{-1}	-37.87	-4.17×10^{-1}	-37.98	-4.57×10^{-1}
-26.90	-2.97×10^{-5}	-25.74	-1.23×10^{-5}	-25.87	-1.39×10^{-5}	-25.94	-1.49×10^{-5}
-25.38	-1.05×10^{-5}	-24.41	-4.56×10^{-6}	-24.85	-6.87×10^{-6}	-24.92	-7.33×10^{-6}
-25.22	-1.99×10^{-5}	-23.83	-5.72×10^{-6}	-24.64	-8.89×10^{-6}	-24.68	-1.18×10^{-5}
-24.89	-1.04×10^{-5}	-25.15	-1.63×10^{-5}	-25.14	-1.68×10^{-5}	-25.09	-1.49×10^{-5}
-25.83	-2.80×10^{-5}	-24.56	-1.03×10^{-5}	-25.15	-1.63×10^{-5}	-25.23	-1.75×10^{-5}
-28.84	-3.06×10^{-4}	-28.71	-9.31×10^{-5}	-28.80	-1.14×10^{-4}	-28.83	-1.21×10^{-4}
-22.89	-2.71×10^{-6}	-23.30	-3.77×10^{-6}	-23.19	-3.45×10^{-6}	-23.19	-3.45×10^{-6}
-24.69	-1.49×10^{-5}	-22.58	-2.57×10^{-6}	-23.76	-8.13×10^{-6}	-23.94	-9.64×10^{-6}
-25.66	-9.85×10^{-6}	-25.78	-1.19×10^{-5}	-25.84	-1.21×10^{-5}	-25.80	-1.17×10^{-5}
-27.06	-3.58×10^{-5}	-23.60	-4.77×10^{-6}	-25.27	-1.79×10^{-5}	-25.59	-2.32×10^{-5}
-25.32	-1.86×10^{-5}	-25.27	-1.79×10^{-5}	-25.53	-2.20×10^{-5}	-25.54	-2.23×10^{-5}
-35.50	-1.66×10^{-1}	-35.73	-2.06×10^{-1}	-35.72	-2.03×10^{-1}	-35.68	-1.93×10^{-1}
-33.06	-1.02×10^{-2}	-32.66	-7.20×10^{-3}	-32.83	-8.67×10^{-3}	-32.84	-8.84×10^{-3}
-24.05	-1.63×10^{-5}	-23.40	-1.14×10^{-5}	-24.02	-1.65×10^{-5}	-24.05	-1.69×10^{-5}
-25.77	-1.35×10^{-5}	-25.96	-2.29×10^{-5}	-25.74	-1.84×10^{-5}	-25.65	-1.72×10^{-5}
-25.16	-1.54×10^{-5}	-24.65	-1.48×10^{-5}	-24.72	-1.46×10^{-5}	-24.71	-1.43×10^{-5}
-23.76	-6.76×10^{-6}	-22.68	-2.30×10^{-6}	-23.42	-4.11×10^{-6}	-23.49	-4.38×10^{-6}
-28.02	-2.40×10^{-5}	-25.70	-3.67×10^{-6}	-26.99	-7.07×10^{-5}	-27.20	-1.59×10^{-5}
-28.02	-9.45×10^{-5}	-26.44	-2.24×10^{-5}	-27.22	-4.96×10^{-5}	-27.32	-5.43×10^{-5}
-27.12	-7.85×10^{-5}	-25.34	-1.91×10^{-5}	-26.46	-4.62×10^{-5}	-26.65	-5.39×10^{-5}
-27.48	-1.04×10^{-4}	-26.47	-2.07×10^{-5}	-27.37	-9.54×10^{-5}	-27.45	-1.02×10^{-4}
-27.09	-5.72×10^{-5}	-28.20	-1.33×10^{-4}	-27.94	-1.05×10^{-4}	-27.83	-9.44×10^{-5}
-22.51	-5.98×10^{-6}	-21.88	-3.98×10^{-6}	-22.17	-5.17×10^{-6}	-22.21	-5.40×10^{-6}
-28.43	-1.04×10^{-4}	-27.31	-5.51×10^{-5}	-27.73	-8.20×10^{-5}	-27.79	-8.60×10^{-5}
-28.19	-2.44×10^{-4}	-30.39	-1.05×10^{-3}	-30.08	-8.20×10^{-4}	-29.89	-7.09×10^{-4}
-23.66	-5.84×10^{-6}	-21.99	-1.71×10^{-6}	-22.67	-3.29×10^{-6}	-22.81	-3.73×10^{-6}
-23.99	-3.46×10^{-6}	-24.60	-1.05×10^{-5}	-24.30	-8.28×10^{-6}	-24.22	-7.79×10^{-6}
-25.02	-2.44×10^{-5}	-23.55	-5.71×10^{-6}	-24.22	-1.15×10^{-5}	-24.35	-1.31×10^{-5}
-24.60	-9.13×10^{-6}	-24.72	-7.61×10^{-6}	-24.75	-7.72×10^{-6}	-24.73	-7.70×10^{-6}

-27.89	-1.36×10^{-4}	-28.63	-3.81×10^{-4}	-28.16	-2.27×10^{-4}	-28.04	-2.01×10^{-4}
-25.59	-5.35×10^{-5}	-23.72	-8.40×10^{-6}	-24.32	-1.50×10^{-5}	-24.41	-1.72×10^{-5}
-27.15	-8.03×10^{-5}	-26.21	-3.81×10^{-5}	-26.52	-4.84×10^{-5}	-26.54	-4.94×10^{-5}
-38.82	-3.58×10^{-1}	-37.64	-1.19×10^{-1}	-38.08	-1.86×10^{-1}	-38.19	-2.07×10^{-1}
-19.34	-3.31×10^{-7}	-18.35	-1.59×10^{-7}	-18.78	-2.50×10^{-7}	-18.83	-2.61×10^{-7}
-16.16	-9.54×10^{-9}	-16.71	-1.09×10^{-8}	-16.70	-1.07×10^{-8}	-16.73	-1.11×10^{-8}
-14.78	-2.40×10^{-9}	-14.32	-1.78×10^{-9}	-14.54	-2.29×10^{-9}	-14.56	-2.24×10^{-9}
-13.06	-1.54×10^{-9}	-12.76	-1.13×10^{-9}	-12.88	-1.27×10^{-9}	-12.91	-1.33×10^{-9}
-10.33	-2.69×10^{-10}	-9.88	-2.56×10^{-10}	-9.93	-2.58×10^{-10}	-9.98	-2.62×10^{-10}

The background contribution from the non-patch-antipatch configurations was accounted for using Monte Carlo integration. The subregions occupied by the patch-antipatch pairs with $\varepsilon < -20\ kT$ were excluded in the orientation sampling. The calculated background contributions to B_{22} are shown in Table 3.12 and 3.13. The results show that that the two proteins have differing background B_{22} trends. The background B_{22} is reduced when electrostatics are incorporated for lysozyme, but this contribution increases as the electrostatics are screened. On the other hand, the background B_{22} for chymosin B is enhanced when electrostatics are included. Typical experimentally measured B_{22} values are on the order of 10^{-5} to 10^{-3} mol ml/g^2. However, the magnitude of the background contribution is still beyond experimentally measured values despite the exclusion of the strongly attractive patch-antipatch configurations. This is due to the cumulative contribution from the configurations that display moderately attractive well depths. These configurations still disproportionately impact the mean value I_{in} in the Monte Carlo averaging. It suggests that Monte Carlo integration may not be suitable even for these moderate configurations and that the same detailed integration procedure used for the patch-antipatch pairs is needed. To obtain a better numerical estimate, the background contribution to B_{22} was broken down into two components: weak interactions and moderate interactions.

Table 3.12: Background B_{22} for lysozyme based on short-range non-electrostatic interactions alone and with addition of electrostatics at pH 7. Patch-antipatch pairs with $\varepsilon < -20\ kT$ were excluded in the Monte Carlo integration.

	$B_{22}^{Background}$ (mol ml/g^2)
Non-electrostatics	-0.10 ± 0.01
0.10M	-0.046 ± 0.003
0.20M	-0.060 ± 0.003
0.30M	-0.078 ± 0.003

Table 3.13: Background B_{22} for chymosin B based on short-range interactions alone and with addition of electrostatics at pH 5. Patch-antipatch pairs with $\varepsilon < -20\ kT$ were excluded in the Monte Carlo integration.

	$B_{22}^{Background}$ (mol ml/g^2)
Non-electrostatics	-0.055 ± 0.003
0.10M	-0.070 ± 0.005
0.30M	-0.062 ± 0.004
0.40M	-0.061 ± 0.004

The configurations that are considered to display weak interactions were identified by specifying a cutoff well depth ε_{cutoff} and determining the contributions to B_{22} from the configurations with $\varepsilon > \varepsilon_{cutoff}$ by the Monte Carlo method. To estimate ε_{cutoff}, a useful benchmark is the unweighted average well depth from all the sampled 10^6 configurations, i.e. $\varepsilon \approx -3\ kT$, which corresponds to $I_{in} \approx 3.9\times10^4$ Å^3. This value can be seen as the baseline in the I_{in} landscape around which most of the peaks from the weakly attractive configurations are scattered. Monte Carlo integration estimates an integral I by the product of the mean value of a function $\langle f \rangle$ and the region of integration v

$$I = v\langle f \rangle \qquad 3.25$$

For a function that is characterized as a peak within a domain, the mean value is the average height that is located somewhere between the apex and base of the peak. The I_{in} landscape can be viewed as being composed of discrete peaks in a multidimensional domain, with each peak having its own mean value $\langle I_{in} \rangle$. Configurations with stronger attractions will have taller peaks in the I_{in} landscape and therefore will have a greater $\langle I_{in} \rangle$. Because there are peaks that disproportionately influence the overall mean value of the I_{in} function, those peaks should be excluded in the Monte Carlo averaging. Furthermore, because the baseline I_{in} defines the boundary between weak and moderate configurations in the I_{in} landscape, it is postulated that configurations with $\langle I_{in} \rangle$ that is equal to or less than the baseline I_{in} are appropriate for inclusion in the Monte Carlo integration.

From the general expression in equation 3.25, $\langle I_{in} \rangle$ for a configuration is determined by scaling its local integrated peak I_{config} by the subdomain of integration v_0

$$\langle I_{in} \rangle = \frac{I_{config}}{v_0} \qquad 3.26$$

For $\Delta = 0.10$ radian, the average hypervolume occupied by a patch-antipatch pair is $v_0 = 1.97 \times 10^{-4}$. Furthermore, the empirical correlations of I_{config} from Figures 3.11 and 3.12 provide the functional relationship between I_{config} and ε/kT for both proteins

$$I_{config} = 0.0200\, exp\left(-0.805 \frac{\varepsilon}{kT}\right);\ for\ lysozyme \qquad 3.27$$

$$I_{config} = 0.0195\, exp\left(-0.770 \frac{\varepsilon}{kT}\right);\ for\ chymosin\ B \qquad 3.28$$

Substituting equations 3.27 and 3.28 into equation 3.26 gives the relationship between a configuration's mean value in the I_{in} landscape and ε/kT for a configuration. Using

the postulated condition that the weak configurations are ones that have $\langle I_{in} \rangle \leq 3.9 \times 10^4$ Å^3, it was determined that $\varepsilon \approx$ -7 kT satisfies this criterion. Therefore, it was concluded that configurations that have $\varepsilon > \varepsilon_{cutoff}$ =-7 kT were appropriate for inclusion in the Monte Carlo averaging.

The contribution from configurations with stronger interaction $\varepsilon < \varepsilon_{cutoff}$ = -7 kT had to be determined by the patch integration method. However, there is a large number of such configurations and performing integration for individual patches would be computationally expensive. Rather than perform a detailed local integration for each orientation, therefore, this explicit approach was reserved for the strongest patch-antipatch pairs with $\varepsilon <$ -20 kT (Tables 3.10 and 3.11) and the remaining integrals with moderate interactions were evaluated using the empirical correlations for I_{config} for both proteins from equations 3.27 and 3.28. The individual contributions were summed together, achieving a significant savings in computing time.

The number of moderate configurations was determined using the observation that the relative frequency of the well depths as a function of ε is relatively consistent even as the degree of sampling increases (Figures 3.2 to 3.4). It was postulated that the relative frequency of the well depth distribution f_ε provides an estimate of the fraction of the configuration space that is occupied by configurations in a particular range of well depths $d\varepsilon$. Therefore, the size of the hypervolume V_ε occupied by configurations with well depth ε can be represented as

$$V_\varepsilon = V f_\varepsilon(\varepsilon) \qquad 3.29$$

such that the total hypervolume is conserved

$$V = \int_0^{-\infty} V_\varepsilon(\varepsilon) d\varepsilon \qquad 3.30$$

where V is the total hypervolume of the global space equal to $32\pi^3$. As mentioned previously, the relative frequency of ε/kT for both proteins is adequately described by a log normal probability distribution function (Figure 3.5). By knowing the hypervolume occupied by configurations in a particular range of ε and the hypervolume of a single patch (which is set by the Δ parameter), the absolute number of moderate configurations N_ε in this range can be approximated by

$$N_\varepsilon = \frac{V_\varepsilon}{v_0} \tag{3.31}$$

The total B_{22} contribution from the moderate configurations that have $\varepsilon < \varepsilon_{cutoff}$ is

$$B_{22}^{Mod} = -\frac{1}{16MW^2\pi^2} \sum_{\varepsilon<\varepsilon_{cutoff}} N_\varepsilon \left(I_{config}\right)_\varepsilon \tag{3.32}$$

$B_{22}^{Background}$ calculated using the procedure outlined above at different ionic strengths for lysozyme at pH 7 and chymosin B at pH 5 is shown in Table 3.14 and Table 3.15, respectively. For lysozyme, $B_{22}^{Background}$ decreases as the ionic strength increases. The background B_{22} contributions from the weak and moderate configurations are similar when only short-range interactions are considered. However, the addition of electrostatics leads to decreased contributions from both components and the overall background contribution decreases. At each level of ionic strength, configurations with weak interactions actually contribute more than the moderate interactions. In other words, the moderate configurations display more repulsive electrostatics and these electrostatic effects have slightly more impact on B_{22} for lysozyme.

$B_{22}^{Background}$ for chymosin B exhibits different behavior. The addition of electrostatics increases the overall attraction. Although the weak contribution shows a trend of increasing attraction as the ionic strength increases, the magnitude of the

moderate contribution increases when electrostatics are incorporated, indicating that it is the moderately attractive configurations that are most responsible for this effect. The net result is increased attraction due to offsetting the effect of repulsion by the moderate configurations with attractive electrostatics.

Table 3.14: Background B_{22} for lysozyme based on short-range non-electrostatics interactions alone and with addition of electrostatics at pH 7. Patch-antipatch pairs with $\varepsilon < -7\ kT$ were excluded in the Monte Carlo integration.

	$B_{22}^{Background}$ ($\times 10^4$mol ml/g^2)		
	Weak	**Moderate**	**Total**
Non-electrostatics	-10.3	-10.4	-20.7
0.10M	-5.62	-5.45	-11.1
0.20M	-7.76	-6.99	-14.8
0.30M	-8.66	-8.58	-17.2

Table 3.15: Background B_{22} for chymosin B based on short-range non-electrostatics interactions alone and with addition of electrostatics at pH 5. Patch-antipatch pairs with $\varepsilon < -7\ kT$ were excluded in the Monte Carlo integration.

	$B_{22}^{Background}$ ($\times 10^4$mol ml/g^2)		
	Weak	**Moderate**	**Total**
Non-electrostatics	-3.66	-1.73	-5.39
0.10M	-3.58	-2.04	-5.62
0.30M	-3.65	-1.98	-5.63
0.40M	-3.65	-1.76	-5.41

3.4 Discussion

The identification of patch-antipatch pairs for proteins is strongly dependent on the density of orientation sampling of the two protein molecules.

However, the degree of sampling is limited by the available computational power. Present computing speeds readily allow sampling of order 10^6 angular configurations to be performed for identifying patch-antipatch pairs, which is two orders of magnitude greater than the sampling reported previously by Hloucha et al. for bovine chymotrypsinogen (78). In their work, the strongest patch-antipatch pairs that were identified possessed well depths between -13 *kT* and -15 *kT*. From this work, most of the strongest configurations were on the order of at least -20 *kT*.

The presence of strongly attractive patch-antipatch configurations with well depths on the order of -20 *kT* led to significant numerical issues when Monte Carlo integration was used. While this integration approach was found to be appropriate for determining the excluded volume contribution to B_{22}, it fails when the energetic contributions are included. The convergence issues were addressed using a hybrid Monte Carlo/patch integration method. By performing a detailed integration over the patch-antipatch subregions using a globally adaptive integration routine, a functional relationship between the strength of attraction for a configuration and its contribution to B_{22} was obtained. However, calculation of B_{22} could not be completed for both proteins due to the skewed contributions from anomalously strong patch-antipatch pairs that possessed well minima that were on the order of -30 *kT* to -40 *kT* (Tables 3.10 and 3.11). This indicates that other effects need to be accounted for in the calculation of B_{22}.

To make complete and accurate predictions of B_{22}, the effect of hydration must be accounted for explicitly. In these interaction calculations, an implicit solvent assumption was made where the effects of the solvent were accounted for by the Hamaker constant in the Lifshitz-Hamaker model and by an empirical factor in the

Lennard-Jones model. The hybrid method is capable of capturing surface complementarity well, but the effect of strongly bound water molecules is lost. Explicit inclusion of hydration effects can lead to elimination of high complementarity protein-protein configurations (patch-antipatch pairs) (135, 136). Such an effect would be expected to play an important role in accurately calculating B_{22}. This is illustrated by the contribution of a -40 kT patch-antipatch pair for lysozyme, which was found to have a dominant B_{22} contribution that was orders of magnitude greater than what is experimentally measured. Such an anomalously attractive configuration could very well be eliminated by one or more strongly bound water molecules. Thus, hydration effects cannot be neglected and are crucial in the accurate prediction of B_{22}.

The calculations of B_{22} using the proposed hybrid method include additional potential sources of uncertainty. First, there is the issue of whether all patch-antipatch pairs can be identified by a random sampling. The configuration space is large due to the degrees of freedom in defining the relative orientation of two protein molecules. Another uncertainty is whether the identified patch-antipatch pairs are in the optimally aligned orientation. While refinement was attempted by a local sampling within ± 0.10 radian around the central orientations of the strongest configurations, a different energy minimum may or may not be detected if the limits are expanded. Once again, this issue is due to the size of the configuration space. The uncertainties in the optimal alignment and absolute population of patch-antipatch pairs are a cause for concern in calculating B_{22} using atomistic models. However, the statistical distribution of well depths from a finite sampling provides valuable information on the fraction of the global angular space occupied by the different configurations. This is because the relative frequency of orientations with various

levels of attractions is independent of the sample size taken, as shown by the consistent distribution of well depths for different levels of sampling performed. Furthermore, the distribution was found to be adequately described by an ideal log normal probability distribution function. This information significantly aided in the computation of the background contribution to B_{22}. Another significant uncertainty is the size of the patch-antipatch pairs. A simplification was made in which all configurations were assumed to be confined within ± 0.10 radian around their respective central orientations. The localized patch integration calculations were performed using this assumption. Although the B_{22} contributions from the patch integration approach are very much dependent on what is chosen for Δ, it is has been shown that the interactions of the patch-antipatch configurations may be quite sensitive to perturbations in orientation. Therefore, it can be inferred that these patch-antipatch pairs occupy only a small fraction of the global configuration space. For Δ = 0.10 radian, the subregion of the configurational hypervolume occupied by each patch-antipatch pair is on average approximately 1.97×10^{-4}, which is a very small portion of the global configurational space.

3.5 Conclusions

A detailed numerical approach for computing B_{22} from atomistic models of proteins was proposed and carried out. The issues this work has attempted to address illustrate the difficulty and uncertainties of computing B_{22} at the atomistic scale. The results highlight the influence of structure on the anisotropy of protein-protein interactions and therefore the solution properties of proteins. The heterogeneity of the I_{in} function makes calculation of the B_{22} integral particularly challenging. The proposed hybrid approach appears to provide a better method of

calculating B_{22} at the atomistic scale when compared to the Monte Carlo approach. However, the identification of anomalously attractive configurations with well depths ~ -30 *kT* to -40 *kT* strongly emphasizes the importance and necessity of incorporating hydration effects to make accurate predictions of B_{22}, which will require the use of molecular dynamics simulations.

A particular aspect of this work is accounting for the charge distribution of the protein surface in representing protein interactions. The effect of electrostatics on the B_{22} trends for lysozyme and chymosin B was captured by incorporating a screened Coulombic potential contribution in the pairwise atomistic interaction calculations. This approach provided a simple and computationally efficient way of accounting not only for the charge distribution, but also the effect of shape on the electrostatic interactions. By incorporating the effects of electrostatics, the qualitative trends in B_{22} were distinguished for the two proteins studied that agreed with experimental observations. Simple colloidal models that do not account for this important feature are unable to qualitatively predict differences in the solution behavior of proteins.

Chapter 4

CONCLUSIONS AND RECOMMENDATIONS

The central theme of this thesis is elucidation of the relation between the molecular structure of proteins and their continuum thermodynamic properties. The structural properties of proteins are complex, and it is this complexity that has a profound impact on the molecular interactions that ultimately dictate their macroscopic solution properties. The osmotic second virial coefficient, B_{22}, provides a promising qualitative link between the protein-protein interactions and phase behavior of proteins. This relationship was explored quantitatively, which yielded significant insights into which essential aspects of protein interactions must be incorporated in thermodynamic models in order to lead to accurate predictions of the solution behavior for proteins. In this chapter, the findings and conclusions from this work are summarized and recommendations for future investigations are put forth.

4.1 Conclusions

4.1.1 Continuum Thermodynamic Models

In Chapter 2, an attempt was made to quantitatively relate experimental B_{22} values with the phase diagrams for the model protein ribonuclease A. Several continuum models derived from classical theories for polymers and colloids were explored in an effort find a mechanistic framework for protein solutions. While a qualitative correlation between B_{22} and phase behavior was found, quantitative agreement could not be obtained using the continuum models. Phase equilibrium was

also predicted from osmotic virial coefficients using the osmotic virial equation derived from McMillan-Mayer solution theory. Although theoretically calculated third virial coefficients along with B_{22} values were used in the phase equilibrium calculations, the phase diagram predicted from this model only qualitatively agreed with experimental results.

The discrepancy of the results from the continuum models may be due to the isotropic assumption inherent in each model and also the molecular nature of B_{22}. The orientationally-averaged character of B_{22} provides an incomplete description of protein-protein interactions and therefore B_{22} is limited in its ability to predict phase behavior. It is clear that the anisotropic character of protein-protein interactions cannot be neglected, providing the most likely explanation for the inadequacy of the continuum models studied. This finding justified the need for exploring molecular-level models that incorporate the anisotropy as an essential feature.

4.1.2 Patch-Antipatch Model of Proteins and the Calculation of B_{22}

Patch models have proven to be useful in providing a coarse-grained representation of proteins to model their anisotropic interactions. They have been shown to provide a better description of protein phase behavior and much progress has been made in understanding the phase diagrams predicted from these models. Accounting for the specific interactions through incorporation of patches can explain features of protein phase behavior not possible with isotropic models. However, because of the large parameter space that is possible for different combinations of patch parameters, much of the work in the literature has focused on simplified patch representations that provide only a caricature of proteins. The parameters of patch models should be faithful to the physical and structural attributes of proteins.

The patch-antipatch parameters for two proteins were determined from extensive atomistic simulations. Several new findings are reported in this thesis beyond the work of Neal et al. (82) and Hloucha et al. (78). First, the orientational sampling for identifying patch-antipatch pairs was increased from 10^4 configurations to 10^6 configurations. As a consequence of this increase in sampling, a larger number of patch-antipatch pairs with deeper attractive wells were identified. Local refinements were performed to find the approximate energy minima, which in some cases led to configurations with well depths as a high as -40 kT. The large size of the global configurational space leads to uncertainties in identifying all patch-antipatch pairs. However, the relative frequency distribution of interaction well depths was found to be independent of the degree of orientation sampling and thus provides an estimate of the fraction of the configuration space that is occupied by the different patch-antipatch pairs.

It has been shown in this work that the patch-antipatch pairs contribute significantly to the calculation of B_{22}. Due to these patch-antipatch configurations, the Monte Carlo integration approach was found to be unsuitable for evaluating the multidimensional integral necessary for computing B_{22}. The hybrid Monte Carlo/patch integration approach that is proposed in Chapter 3 addresses some of the numerical issues in calculating B_{22}. However, the presence of anomalously strong patch-antipatch pairs gave rise to B_{22} contributions that were orders of magnitude greater than typical experimental values. This discrepancy may largely be attributed to the fact that hydration effects were neglected in the calculations. Hydration effects are known to attenuate the interactions of patch-antipatch pairs and will have a significant impact on the calculation of B_{22}. Thus, further refinements are needed in

order to make accurate predictions of B_{22}, which will allow for meaningful comparisons with experimentally measured values.

4.2 Recommendations and Future Directions

4.2.1 The Calculation of B_{22}

Accurate predictions of B_{22} from atomistic models will require accounting for the specific hydration of proteins, which is known to affect high-complementarity configurations (patch-antipatch pairs) (135, 136). The hybrid continuum/atomistic approach used to calculate the short-range non-electrostatic interaction energies captures the effect of surface complementarity between proteins, but the effect of strongly bound water molecules is not generally accounted for. These water molecules strongly bound to the protein provide additional steric hindrance that can effectively inhibit the interactions between patch-antipatch pairs. Accounting for such effects will require the use of such methods as molecular dynamics simulations where water molecules are explicitly included.

Another direction that should be explored is the calculation of electrostatic interactions. The addition of a pairwise screened Coulomb contribution to the interaction energies provided a simple way of accounting for the shape anisotropy of the charge distribution in the calculation of B_{22}. This approach was able to qualitatively reflect the impact of this anisotropy on B_{22} for lysozyme and chymosin B. However, this model is simplistic in its treatment of the electrostatics since it does not explicitly account for the local dielectric boundaries. It would be necessary to compare how the magnitude of the electrostatic energies computed from the pairwise screened Coulomb potential approach compares with the energies determined from

solving the Poisson-Boltzmann equation for the full protein geometry. Several software packages that utilize finite-difference or boundary-element approaches for solving the Poisson-Boltzmann equation can be used for such a comparison.

Once these further refinements in the calculation procedure of B_{22} are performed and affirmed, the methodology can be extended to determining cross-interaction virial coefficients B_{23}, which characterize the association of two different protein molecules. Experimental B_{23} data for protein mixtures have previously been measured and reported in the literature (123, 143). Quantitative predictions for the cross-interactions of proteins can have relevant applications in industrial protein separation processes.

4.2.2 Molecular Simulation of Protein Phase Behavior

Another avenue that should be explored further is predicting phase behavior from the "patch-antipatch" representation of proteins. The patch-antipatch parameters determined from the atomistic simulations performed in this work should be utilized in the simulation of the phase behavior for these proteins. Although atomistic models of the kind used in Chapter 3 are, in principle, suitable for describing crystalline phases (144), such computations would be very challenging based on current computational capabilities. However, patch models based on spheres may be appropriate for simulating phases such as liquids since these disordered states do not require a very specific 3-D molecular packing structure.

The logical route towards this endeavor is through the use of molecular simulation (145, 146). The patch-antipatch parameters specific to the protein of interest can be determined *a priori* from atomistic calculation methods described in Chapter 3. There are various standard molecular simulation techniques for predicting

phase behavior, which include the Gibbs ensemble Monte Carlo (147–150) and grand canonical Monte Carlo/histogram reweighting techniques (150, 151). An in-depth discussion of these techniques is beyond the scope of this work, but it is clear that because of the strong attractions due to the specific patch interactions, particles may be trapped by the very strong attractions in some configurations (152). As a result, sampling of the entire configuration space becomes prohibitively long and the system becomes nonergodic. Advanced simulation methods for strongly associating fluids will need to be implemented. Several biased methods have been reported to address such issues, which include the bond-biased Monte Carlo method (153), association-biased Monte Carlo method (154), and aggregation volume bias Monte Carlo method (155). These methods improve upon standard simulation methods by biasing the acceptance criterion for accepting trial moves that allow for more efficient sampling of the configuration space. These biased methods should be utilized to simulate the equilibrium phases of proteins.

REFERENCES

1. Ishimoto, C., and T. Tanaka. 1977. Critical behavior of a binary mixture of protein and salt water. Physical Review Letters. 39: 474–477.

2. Phillies, G. 1985. Comment on "Critical behavior of a binary mixture of protein and salt water."Physical Review Letters. 55: 1341-1341.

3. Taratuta, V.G., A. Holschbach, G.M. Thurston, D. Blankschtein, and G.B. Benedek. 1990. Liquid-liquid phase separation of aqueous lysozyme solutions: effects of pH and salt identity. Journal of Physical Chemistry. 94: 2140–2144.

4. Broide, M., T. Tominc, and M. Saxowsky. 1996. Using phase transitions to investigate the effect of salts on protein interactions. Physical Review E. 53: 6325-6335.

5. Asherie, N., A. Lomakin, and G. Benedek. 1996. Phase diagram of colloidal solutions. Physical Review Letters. 77: 4832-4835.

6. Foffi, G., G. McCullagh, A. Lawlor, E. Zaccarelli, K. Dawson, F. Sciortino, P. Tartaglia, D. Pini, and G. Stell. 2002. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: application to protein crystallization. Physical Review E. 65: 1-17.

7. Broide, M.L., C.R. Berland, J. Pande, O.O. Ogun, and G.B. Benedek. 1991. Binary-liquid phase separation of lens protein solutions. Proceedings of the National Academy of Sciences of the United States of America. 88: 5660-5664.

8. Schurtenberger, P., R. Chamberlin, G. Thurston, J. Thomson, and G. Benedek. 1989. Observation of critical phenomena in a protein-water solution. Physical Review Letters. 63: 2064-2067.

9. Thomson, J.A., P. Schurtenberger, G.M. Thurston, and G.B. Benedek. 1987. Binary liquid phase separation and critical phenomena in a protein/water solution. Proceedings of the National Academy of Sciences of the United States of America. 84: 7079-7083.

10. Berland, C.R., G.M. Thurston, M. Kondo, M.L. Broide, J. Pande, O. Ogun, and G.B. Benedek. 1992. Solid-liquid phase boundaries of lens protein solutions. Proceedings of the National Academy of Sciences of the United States of America. 89: 1214-1218.

11. Muschol, M., and F. Rosenberger. 1997. Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. The Journal of Chemical Physics. 107: 1953-1962.

12. ten Wolde, P.R., and D. Frenkel. 1997. Enhancement of protein crystal nucleation by critical density fluctuations. Science. 277: 1975-1978.

13. ten Wolde, P.R., and D. Frenkel. 1999. Enhanced protein crystallization around the metastable critical point. Theoretical Chemistry Accounts: Theory, Computation, and Modeling. 101: 205–208.

14. Hiemenz, P., and R. Rjagopalan. 1997. Principles of colloid and surface chemistry. 3rd ed. New York: Marcel Dekker.

15. Israelachvili, J. 1992. Intermolecular surfaces and forces. 2nd ed. Academic Press.

16. Leckband, D., and J. Israelachvili. 2001. Intermolecular forces in biology. Quarterly Reviews of Biophysics. 34: 105-267.

17. Kulkarni, A.M., A.P. Chatterjee, K.S. Schweizer, and C.F. Zukoski. 2000. Effects of polyethylene glycol on protein interactions. The Journal of Chemical Physics. 113: 9863-9873.

18. Roth, C.M., B.L. Neal, and A.M. Lenhoff. 1996. Van der Waals interactions involving proteins. Biophysical Journal. 70: 977-987.

19. Collins, K.D. 2004. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. Methods. 34: 300-311.

20. Hofmeister, F. 1888. Zur Lehre von der Wirkung der Salze. Arch. Exp. Pathol. Pharmakol. 24: 247-260.

21. Kunz, W. 2010. Specific ion effects. 1st ed. World Scientific.

22. Kunz, W. 2010. Specific ion effects in colloidal and biological systems. Current Opinion in Colloid & Interface Science. 15: 34-39.

23. Wilson, E.K. 2007. A renaissance for Hofmeister. Chemical & Engineering News. 85: 47–49.

24. Kunz, W., P. Lo Nostro, and B.W. Ninham. 2004. The present state of affairs with Hofmeister effects. Current Opinion in Colloid & Interface Science. 9: 1-18.

25. Tavares, F.W., M. Boström, E.R.A. Lima, and E.C. Biscaia Jr. 2010. Ion-specific thermodynamic properties of colloids and proteins. Fluid Phase Equilibria. 296: 99-105.

26. Lima, E.R.A., E.C. Biscaia Jr., M. Boström, and F.W. Tavares. 2010. Ion-specific thermodynamical properties of aqueous proteins. Anais da Academia Brasileira de Ciências. 82: 109-126.

27. Omta, A.W., M.F. Kropman, S. Woutersen, and H.J. Bakker. 2003. Influence of ions on the hydrogen-bond structure in liquid water. The Journal of Chemical Physics. 119: 12457-12461.

28. Omta, A.W., M.F. Kropman, S. Woutersen, and H.J. Bakker. 2003. Negligible effect of ions on the hydrogen-bond structure in liquid water. Science. 301: 347-349.

29. Zhang, Y., and P.S. Cremer. 2006. Interactions between macromolecules and ions: The Hofmeister series. Current Opinion in Chemical Biology. 10: 658-663.

30. Ninham, B.W., and V. Yaminsky. 1997. Ion binding and ion specificity: the Hofmeister effect and Onsager and Lifshitz theories. Langmuir. 13: 2097-2108.

31. Boström, M., D.R.M. Williams, and B.W. Ninham. 2003. Specific ion effects: why the properties of lysozyme in salt solutions follow a Hofmeister series. Biophysical Journal. 85: 686-694.

32. Boström, M., D. Williams, and B. Ninham. 2001. Specific ion effects: why DLVO theory fails for biology and colloid systems. Physical Review Letters. 87: 1-4.

33. Lima, E.R.A., E.C. Biscaia, M. Bostrom, F.W. Tavares, and J.M. Prausnitz. 2007. Osmotic second virial coefficients and phase diagrams for aqueous proteins from a much-improved Poisson-Boltzmann equation. Journal of Physical Chemistry C. 111: 16055-16059.

34. Lettieri, S. 2010. Specific ion effects and the phase diagram of lysozyme. Physics Procedia. 6: 46-51.

35. Lettieri, S., X. Li, and J. Gunton. 2008. Hofmeister effect and the phase diagram of lysozyme. Physical Review E. 78: 1-6.

36. Ries-Kautt, M.M., and A.F. Ducruix. 1989. Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. The Journal of Biological Chemistry. 264: 745-748.

37. Lund, M., and P. Jungwirth. 2008. Patchy proteins, anions and the Hofmeister series. Journal of Physics: Condensed Matter. 20: 494218.

38. Lund, M., and P. Jungwirth. 2008. Ion specific protein assembly and hydrophobic surface forces. Physical Review Letters. 100: 1-4.

39. Lund, M., R. Vacha, and P. Jungwirth. 2008. Specific ion binding to macromolecules: effects of hydrophobicity and ion pairing. Langmuir. 24: 3387-3391.

40. Horinek, D., and R.R. Netz. 2007. Specific ion adsorption at hydrophobic solid surfaces. Physical Review Letters. 99: 48-51.

41. Lettieri, S., X. Li, and J. Gunton. 2009. Ion specific effects on phase transitions in protein solutions. Physical Review E. 79: 1-6.

42. McQuarrie, D.A. 2000. Statistical mechanics. Mill Valley, CA: University Science Books.

43. Curtis, R., and L. Lue. 2006. A molecular approach to bioseparations: protein–protein and protein–salt interactions. Chemical Engineering Science. 61: 907-923.

44. Piazza, R. 2000. Interactions and phase transitions in protein solutions. Current Opinion in Colloid & Interface Science. 5: 38-43.

45. Dumetz, A.C., A.M. Chockla, E.W. Kaler, and A.M. Lenhoff. 2008. Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. Biophysical Journal. 94: 570-583.

46. Gast, A.P., C.K. Hall, and W.B. Russel. 1983. Polymer-induced phase separations in nonaqueous colloidal suspensions. Journal of Colloid and Interface Science. 96: 251-267.

47. Ilett, S., A. Orrock, W. Poon, and P. Pusey. 1995. Phase behavior of a model colloid-polymer mixture. Physical Review E. 51: 1344-1352.

48. Rosenbaum, D., P. Zamora, and C. Zukoski. 1996. Phase behavior of small attractive colloidal particles. Physical Review Letters. 76: 150-153.

49. Rosenbaum, D., and C. Zukoski. 1996. Protein interactions and crystallization. Journal of Crystal Growth. 169: 752-758.

50. Miller, M.A., and D. Frenkel. 2004. Simulating colloids with Baxter's adhesive hard sphere model. Journal of Physics: Condensed Matter. 16: S4901-S4912.

51. Miller, M.A., and D. Frenkel. 2004. Phase diagram of the adhesive hard sphere fluid. The Journal of Chemical Physics. 121: 535-545.

52. Liu, H., S. Garde, and S. Kumar. 2005. Direct determination of phase behavior of square-well fluids. The Journal of Chemical Physics. 123: 174505.

53. Lomakin, A., N. Asherie, and G.B. Benedek. 1996. Monte Carlo study of phase separation in aqueous protein solutions. The Journal of Chemical Physics. 104: 1646-1656.

54. Pagan, D.L., and J.D. Gunton. 2005. Phase behavior of short-range square-well model. The Journal of Chemical Physics. 122: 184515.

55. Duda, Y. 2009. Square-well fluid modeling of protein liquid-vapor coexistence. The Journal of Chemical Physics. 130: 116101.

56. Hagen, M.H.J., and D. Frenkel. 1994. Determination of phase diagrams for the hard-core attractive Yukawa system. The Journal of Chemical Physics. 101: 4093.

57. Shukla, K.P. 2000. Phase equilibria and thermodynamic properties of hard core Yukawa fluids of variable range from simulations and an analytical theory. The Journal of Chemical Physics. 112: 10358.

58. Lutsko, J.F., and G. Nicolis. 2005. The effect of the range of interaction on the phase diagram of a globular protein. The Journal of Chemical Physics. 122: 244907.

59. Noro, M.G., and D. Frenkel. 2000. Extended corresponding-states behavior for particles with variable range attractions. The Journal of Chemical Physics. 113: 2941.

60. Curtis, R.A., J.M. Prausnitz, and H.W. Blanch. 1998. Protein-protein and protein-salt interactions in aqueous protein solutions containing concentrated electrolytes. Biotechnology and Bioengineering. 57: 11-21.

61. Gunton, J.D., A. Shiryayev, and D.L. Pagan. 2007. Protein condensation. 1st ed. New York: Cambridge University Press.

62. Li, J., R. Rajagopalan, and J. Jiang. 2010. Molecular modeling and simulation for phase behavior of protein solutions. Biomolecular Engineering. 661: 75-130.

63. Li, J. 2008. Phase separations in protein solutions: a Monte Carlo simulation study. Ph.D. Thesis, National University of Singapore.

64. Rosenbaum, D.F., A. Kulkarni, S. Ramakrishnan, and C.F. Zukoski. 1999. Protein interactions and phase behavior: sensitivity to the form of the pair potential. The Journal of Chemical Physics. 111: 9882.

65. Vliegenthart, G.A., and H.N.W. Lekkerkerker. 2000. Predicting the gas–liquid critical point from the second virial coefficient. The Journal of Chemical Physics. 112: 5364.

66. Pellicane, G., D. Costa, and C. Caccamo. 2004. Theory and simulation of short-range models of globular protein solutions. Journal of Physics: Condensed Matter. 16: S4923-S4936.

67. Pellicane, G., D. Costa, and C. Caccamo. 2003. Phase coexistence in a DLVO model of globular protein solutions. Journal of Physics: Condensed Matter. 15: 375-384.

68. Pellicane, G., D. Costa, and C. Caccamo. 2004. Microscopic determination of the phase diagrams of lysozyme and gamma-crystallin solutions. The Journal of Physical Chemistry B. 108: 7538–7541.

69. Petsev, D.N., and P.G. Vekilov. 2000. Evidence for non-DLVO hydration interactions in solutions of the protein apoferritin. Physical Review Letters. 84: 1339-1342.

70. San Biagio, P.L., and M.U. Palma. 1991. Spinodal lines and Flory-Huggins free-energies for solutions of human hemoglobins HbS and HbA. Biophysical Journal. 60: 508-512.

71. Carlsson, F., M. Malmsten, and P. Linse. 2001. Monte Carlo simulations of lysozyme self-association in aqueous solution. Journal of Physical Chemistry B. 105: 12189-12195.

72. Rosch, T.W., and J.R. Errington. 2007. Investigation of the phase behavior of an embedded charge protein model through molecular simulation. The Journal of Physical Chemistry B. 111: 12591-12598.

73. Kern, N., and D. Frenkel. 2003. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction. The Journal of Chemical Physics. 118: 9882.

74. Bianchi, E., R. Blaak, and C.N. Likos. 2011. Patchy colloids: state of the art and perspectives. Physical Chemistry Chemical Physics. 13: 6397-6410.

75. Lomakin, A., N. Asherie, and G.B. Benedek. 1999. Aeolotopic interactions of globular proteins. Proceedings of the National Academy of Sciences of the United States of America. 96: 9465-9468.

76. Sear, R.P. 1999. Phase behavior of a simple model of globular proteins. The Journal of Chemical Physics. 111: 4800.

77. Gögelein, C., G. Nägele, R. Tuinier, T. Gibaud, A. Stradner, and P. Schurtenberger. 2008. A simple patchy colloid model for the phase behavior of lysozyme dispersions. The Journal of Chemical Physics. 129: 085102.

78. Hloucha, M., J.F.M. Lodge, A.M. Lenhoff, and S.I. Sandler. 2001. A patch–antipatch representation of specific protein interactions. Journal of Crystal Growth. 232: 195-203.

79. Liu, H., S.K. Kumar, F. Sciortino, and G.T. Evans. 2009. Vapor-liquid coexistence of fluids with attractive patches: an application of Wertheim's theory of association. The Journal of Chemical Physics. 130: 044902.

80. Zimm, B.H. 1946. Application of the methods of molecular distribution to solutions of large molecules. The Journal of Chemical Physics. 14: 164-179.

81. Goldstein, H. 1950. Classical mechanics. Cambridge: Addison-Wesley Press Inc.

82. Neal, B.L., D. Asthagiri, and A.M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. Biophysical Journal. 75: 2469-2477.

83. Neal, B., D. Asthagiri, O.D. Velev, A.M. Lenhoff, and E.W. Kaler. 1999. Why is the osmotic second virial coefficient related to protein crystallization? Journal of Crystal Growth. 196: 377-387.

84. Chang, R.C., D. Asthagiri, and A.M. Lenhoff. 2000. Measured and calculated effects of mutations in bacteriophage T4 lysozyme on interactions in solution. Proteins. 41: 123-132.

85. Guo, B. 1999. Correlation of second virial coefficients and solubilities useful in protein crystal growth. Journal of Crystal Growth. 196: 424-433.

86. Dumetz, A.C., A.M. Snellinger-O'Brien, E.W. Kaler, and A.M. Lenhoff. 2007. Patterns of protein protein interactions in salt solutions and implications for protein crystallization. Protein Science. 16: 1867-1877.

87. Lewus, R.A., P.A. Darcy, A.M. Lenhoff, and S.I. Sandler. 2011. Interactions and phase behavior of a monoclonal antibody. Biotechnology Progress. 27: 280-289.

88. George, A., and W.W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. Acta Crystallographica. Section D, Biological Crystallography. 50: 361-365.

89. Asherie, N. 2004. Protein crystallization and phase diagrams. Methods. 34: 266-272.

90. Flory, P.J. 1953. Principles of polymer chemistry. 1st ed. Ithaca, NY: Cornell University Press.

91. Haas, C., and J. Drenth. 1998. The protein−water phase diagram and the growth of protein crystals from aqueous solution. The Journal of Physical Chemistry B. 102: 4226-4232.

92. Haas, C., and J. Drenth. 2000. The interface between a protein crystal and an aqueous solution and its effects on nucleation and crystal growth. The Journal of Physical Chemistry B. 104: 368-377.

93. Haas, C., and J. Drenth. 1999. Understanding protein crystallization on the basis of the phase diagram. Journal of Crystal Growth. 196: 388-394.

94. McMillan Jr, W.G., and J.E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. The Journal of Chemical Physics. 13: 276-305.

95. Tessier, P.M., S.D. Vandrey, B.W. Berger, R. Pazhianur, S.I. Sandler, and A.M. Lenhoff. 2002. Self-interaction chromatography: a novel screening method for rational protein crystallization. Acta Crystallographica Section D Biological Crystallography. 58: 1531-1535.

96. Tessier, P.M., A.M. Lenhoff, and S.I. Sandler. 2002. Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. Biophysical Journal. 82: 1620-1631.

97. Sandler, S.I. 2006. Introduction to chemical, biochemical, and engineering thermodynamics. 4th ed. John Wiley & Son.

98. Haas, C., and J. Drenth. 1995. The interaction energy between two protein molecules related to physical properties of their solution and their crystals and implications for crystal growth. Journal of Crystal Growth. 154: 126-135.

99. Overbeek, J.T.G. 1978. The first Rideal lecture. Microemulsions, a field at the border between lyophobic and lyophilic colloids. Faraday Discussions of the Chemical Society. 65: 7–19.

100. Carnahan, N.F., and K.E. Starling. 1970. Thermodynamic properties of a rigid-sphere fluid. The Journal of Chemical Physics. 53: 600-603.

101. Rowlinson, J.S. 1989. The Yukawa potential. Physica A: Statistical Mechanics and its Applications. 156: 15-34.

102. Sherwood, A., and J. Prausnitz. 1964. Third virial coefficient for the Kihara, exp-6, and square-well potentials. The Journal of Chemical Physics. 41: 413-428.

103. Alder, B.J., and J.A. Pople. 1957. Third virial coefficient for intermolecular potentials with hard sphere cores. The Journal of Chemical Physics. 26: 325-328.

104. Graben, H., and R. Present. 1964. Third virial coefficient for the Sutherland (∞, ν) potential. Reviews of Modern Physics. 36: 1025-1033.

105. Naresh, D.J., and J.K. Singh. 2009. Virial coefficients of hard-core attractive Yukawa fluids. Fluid Phase Equilibria. 285: 36-43.

106. Kihara, T. 1953. Virial coefficients and models of molecules in gases. Reviews of Modern Physics. 25: 831-843.

107. Hirschfelder, J.O., C.F. Curtiss, and R.B. Bird. 1954. Molecular theory of gases and liquids. New York: John Wiley & Sons Inc.

108. Chang, J., A.M. Lenhoff, and S.I. Sandler. 2004. Determination of fluid--solid transitions in model protein solutions using the histogram reweighting method and expanded ensemble simulations. The Journal of Chemical Physics. 120: 3003-3014.

109. Liu, H., S.K. Kumar, and F. Sciortino. 2007. Vapor-liquid coexistence of patchy models: relevance to protein phase behavior. The Journal of Chemical Physics. 127: 084902.

110. Li, X., J.D. Gunton, and A. Chakrabarti. 2009. A simple model of directional interactions for proteins. The Journal of Chemical Physics. 131: 115101.

111. Bianchi, E., J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino. 2006. Phase diagram of patchy colloids: towards empty liquids. Physical Review Letters. 97: 1-4.

112. Fantoni, R., D. Gazzillo, A. Giacometti, M.A. Miller, and G. Pastore. 2007. Patchy sticky hard spheres: analytical study and Monte Carlo simulations. The Journal of Chemical Physics. 127: 234507.

113. Bianchi, E., P. Tartaglia, E. Zaccarelli, and F. Sciortino. 2008. Theoretical and numerical study of the phase diagram of patchy colloids: ordered and disordered patch arrangements. The Journal of Chemical Physics. 128: 144504.

114. Giacometti, A., F. Lado, J. Largo, G. Pastore, and F. Sciortino. 2010. Effects of patch size and number within a simple model of patchy colloids. The Journal of Chemical Physics. 132: 174110.

115. Giacometti, A., F. Lado, J. Largo, G. Pastore, and F. Sciortino. 2009. Phase diagram and structural properties of a simple model for one-patch particles. The Journal of Chemical Physics. 131: 174114.

116. Song, X. 2002. Role of anisotropic interactions in protein crystallization. Physical Review E. 66: 1-4.

117. Romano, F., E. Sanz, and F. Sciortino. 2009. Role of the range in the fluid-crystal coexistence for a patchy particle model. The Journal of Physical Chemistry B. 113: 15133-15136.

118. Romano, F., E. Sanz, and F. Sciortino. 2010. Phase diagram of a tetrahedral patchy particle model for different interaction ranges. The Journal of Chemical Physics. 132: 184501.

119. Noya, E.G., C. Vega, J.P.K. Doye, and A.A. Louis. 2010. The stability of a crystal with diamond structure for patchy particles with tetrahedral symmetry. The Journal of Chemical Physics. 132: 234511.

120. Noya, E.G., C. Vega, J.P.K. Doye, and A.A. Louis. 2007. Phase diagram of model anisotropic particles with octahedral symmetry. The Journal of Chemical Physics. 127: 054501.

121. Jones, S., and J.M. Thornton. 1996. Principles of protein-protein interactions. Proceedings of the National Academy of Sciences. 93: 13-20.

122. Asthagiri, D., B. Neal, and A. Lenhoff. 1999. Calculation of short-range interactions between proteins. Biophysical Chemistry. 78: 219–231.

123. Moon, Y.U., R.A. Curtis, C.O. Anderson, H.W. Blanch, and J.M. Prausnitz. 2000. Protein – protein interactions in aqueous ammonium sulfate solutions. Lysozyme and bovine serum albumin (BSA). Journal of Solution Chemistry. 29: 699-717.

124. Maurer, R.W., S.I. Sandler, and A.M. Lenhoff. 2011. Salting-in characteristics of globular proteins. Biophysical Chemistry. 156: 72-78.

125. Roth, C., and A.M. Lenhoff. 1996. Improved parametric representation of water dielectric data for Lifshitz theory calculations. Journal of Colloid and Interface Science. 179: 637-639.

126. Hamaker, H. 1937. The London—van der Waals attraction between spherical particles. Physica. 4: 1058-1072.

127. Jorgensen, W.L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society. 110: 1657–1666.

128. Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. Protein Science. 1: 169-181.

129. Li, H., A.D. Robertson, and J.H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 61: 704-721.

130. Olsson, M.H.M., C.R. Søndergaard, M. Rostkowski, and J.H. Jensen. 2011. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. Journal of Chemical Theory and Computation. 7: 525-537.

131. Yoon, B.J., and A.M. Lenhoff. 1992. Computation of the electrostatic interaction energy between a protein and a charged surface. The Journal of Physical Chemistry. 96: 3130–3134.

132. Yoon, B.J., and A.M. Lenhoff. 1990. A boundary element method for molecular electrostatics with electrolyte effects. Journal of Computational Chemistry. 11: 1080-1086.

133. McGuffee, S.R., and A.H. Elcock. 2006. Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. Journal of the American Chemical Society. 128: 12098-110.

134. Mereghetti, P., R.R. Gabdoulline, and R.C. Wade. 2010. Brownian dynamics simulation of protein solutions: structural and dynamical properties. Biophysical Journal. 99: 3782-3791.

135. Paliwal, A., D. Asthagiri, D. Abras, A.M. Lenhoff, and M.E. Paulaitis. 2005. Light-scattering studies of protein solutions: role of hydration in weak protein-protein interactions. Biophysical Journal. 89: 1564-1573.

136. Asthagiri, D., A. Paliwal, D. Abras, A.M. Lenhoff, and M.E. Paulaitis. 2005. A consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution. Biophysical Journal. 88: 3300-3309.

137. Gillespie, C. 2011. Protein solution thermodynamics: polymorphic crystallization, liquid-liquid phase separation and osmotic second virial coefficients. Ph.D. Thesis, University of Delaware.

138. Press, W.H., B.P. Flannery, S.A. Teukolosky, and W.T. Vetterling. 1986. Numerical recipes: the art of scientific computing. Cambridge: Cambridge University Press.

139. Berntsen, J., T.O. Espelid, and A. Genz. 1991. Algorithm 698; DCUHRE: an adaptive multidemensional integration routine for a vector of integrals. ACM Transactions on Mathematical Software. 17: 452-456.

140. Berntsen, J., T.O. Espelid, and A. Genz. 1991. An adaptive algorithm for the approximate calculation of multiple integrals. ACM Transactions on Mathematical Software. 17: 437-451.

141. Ogunnaike, B.A. 2010. Random phenomena: fundamentals of probability and statistics for engineers. Boca Raton: CRC Press.

142. Neal, B.L., and A.M. Lenhoff. 1995. Excluded volume contribution to the osmotic second virial coefficient for proteins. AIChE Journal. 41: 1010-1014.

143. Cheng, Y.-C., C.L. Bianco, S.I. Sandler, and A.M. Lenhoff. 2008. Salting-out of lsozyme and ovalbumin from mixtures: predicting precipitation performance from protein−protein interactions. Industrial & Engineering Chemistry Research. 47: 5203-5213.

144. Chang, J., A.M. Lenhoff, and S.I. Sandler. 2005. The combined simulation approach of atomistic and continuum models for the thermodynamics of lysozyme crystals. Journal of Physical Chemistry B. 109: 19507-19515.

145. Allen, M.P., and D.J. Tildesley. 1987. Computer simulation of liquids. 1st ed. Oxford University Press.

146. Frenkel, D., and B. Smit. 2002. Understanding molecular simulation: from algorithms to applications. 2nd ed. Academic Press.

147. Panagiotopoulos, A.Z., V. Wong, and M.A. Floriano. 1998. Phase equilibria of lattice polymers from histogram reweighting Monte Carlo simulations. Macromolecules. 31: 912-918.

148. Panagiotopoulos, A. 1987. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. Molecular Physics. 100: 237-246.

149. Panagiotopoulos, A.Z. 1995. Gibbs ensemble techniques. NATO ASI Series C Mathematical and Physical Sciences-Advanced Study Institute. 460: 463–502.

150. Panagiotopoulos, A. 2000. Monte Carlo methods for phase equilibria of fluids. Journal of Physics: Condensed Matter. 25: R25-R52.

151. Ferrenberg, A.M., and R.H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. Physical Review Letters. 61: 2635-2638.

152. Hloucha, M. 1999. Theoretical investigation of protein interactions. Unpublished Report, University of Delaware.

153. Tsangaris, D.M., and J.J. de Pablo. 1994. Bond-bias simulation of phase equilibria for strongly associating fluids. The Journal of Chemical Physics. 101: 1477.

154. Busch, N.A., M.S. Wertheim, Y.C. Chiew, and M.L. Yarmush. 1994. A Monte Carlo method for simulating associating fluids. The Journal of Chemical Physics. 101: 3147.

155. Chen, B., and J.I. Siepmann. 2000. A novel Monte Carlo algorithm for simulating strongly associating fluids: applications to water, hydrogen fluoride, and acetic acid. The Journal of Physical Chemistry B. 104: 8725-8734.

156. Dobert, F., A. Pfennig, and M. Stumpf. 1995. Derivation of the consistent osmotic virial equation and its application to aqueous poly(ethylene glycol)-dextran two-phase systems. Macromolecules. 28: 7860-7868.

Appendix A

DERIVATION OF LIQUID-LIQUID EQUILBIRIUM FROM THE OSMOTIC VIRIAL EQUATION

The derivation of the equations for describing the liquid-liquid equilibrium of protein solutions from the osmotic virial equation is presented in this appendix. This derivation is consistent with the one reported by Döbert et al. (156). The protein solution is modeled as a binary mixture consisting of solvent (1) + protein (2) species. It is assumed that there is no repartitioning of the salt species in the mixture. The criterion for liquid-liquid equilibrium is the equality of the chemical potentials in the mixture of both the solvent, μ_1, and the protein, μ_2, in each phase:

$$\begin{aligned} \mu_1^I &= \mu_1^{II} \\ \mu_2^I &= \mu_2^{II} \end{aligned} \tag{B.1}$$

where I and II designate the light and dense phases, respectively. To determine the chemical potentials for the species, it can be shown that the chemical potential of the solvent is directly related to the osmotic pressure π (14)

$$\mu_1 - \mu_1^o = -\pi \bar{V}_1 \tag{B.2}$$

where μ_1^o is the chemical potential of the pure solvent, and $\overline{V}_1$ is the partial molar volume of the solvent. The osmotic virial equation in terms of protein mass concentration *c* and truncated at the third virial term is

$$\frac{\pi}{RT} = \frac{c}{MW} + B_2\left(\frac{c}{MW}\right)^2 + B_3\left(\frac{c}{MW}\right)^3 \tag{B.3}$$

where R is the molar gas constant, MW is the molecular weight of the protein, and B_2 and B_3 are the second and third virial coefficients, respectively. Substituting equation (B.3) into equation (B.2) yields the expression for the chemical potential of the solvent

$$-\frac{\mu_1 - \mu_1^o}{RT\bar{V}_1} = \frac{c}{MW} + B_2\left(\frac{c}{MW}\right)^2 + B_3\left(\frac{c}{MW}\right)^3 \tag{B.4}$$

To determine the chemical potential of the protein species, the Gibbs-Duhem relation (97) is invoked

$$N_1\left(\frac{\partial \mu_1}{\partial N_2}\right)_{T,P,N_1} + N_2\left(\frac{\partial \mu_2}{\partial N_2}\right)_{T,P,N_1} = 0 \tag{B.5}$$

where N_1 and N_2 are the number of moles of the solvent and protein species, respectively. Rearranging equation (B.5) in terms of μ_2 leads to

$$\mu_2 = \int -\frac{N_1}{N_2}\left(\frac{\partial \mu_1}{\partial N_2}\right)_{T,P,N_1} dN_2 + C \tag{B.6}$$

where C is the constant of integration. To evaluate the derivative inside the integral of equation (B.6), it is convenient to write μ_1 in equation (B.4) in terms of N_1 and N_2. The mass concentration of protein c is related to its mole fraction x_2 by

$$c = \frac{MW x_2}{\underline{V}} \tag{B.7}$$

where $\underline{V}$ is the molar volume of the solution. For a binary system, $\underline{V}$ is the sum of the partial molar volumes of the solvent and protein weighted by their respective mole fractions x_1 and x_2

$$\underline{V} = x_1\overline{V}_1 + x_2\overline{V}_2 \tag{B.8}$$

Furthermore, a species mole fractions is the ratio of the number of moles of the species to the total number of moles of the system

$$x_1 = \frac{N_1}{N_1 + N_2}, \qquad x_2 = \frac{N_2}{N_1 + N_2} \tag{B.9}$$

By appropriate substitutions of equations (B.7), (B.8), and (B.9) into equation (B.4), the chemical potential of the solvent μ_1 can be rewritten in terms of N_1 and N_2

$$-\frac{\mu_1 - \mu_1^o}{RT\bar{V}_1} = \frac{N_2}{N_1\overline{V}_1 + N_2\overline{V}_2} + B_2\left(\frac{N_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^2 + B_3\left(\frac{N_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^3 \tag{B.10}$$

Inserting equation (B.10) into equation (B.6), differentiating μ_1 with respect to N_1, and integrating the result with respect to N_2 yields the expression for μ_2

$$\mu_2 = RT\left[\ln\left(\frac{N_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right) + \left(1 - \frac{2B_2}{\overline{V}_2}\right)\left(\frac{N_1\overline{V}_1}{N_1\overline{V}_1 + N_2\overline{V}_2}\right) + \left(\frac{3B_3}{2\overline{V}_2^2} - \frac{B_2}{\overline{V}_2}\right)\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^2 - \frac{B_3}{\overline{V}_2{}^2}\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^3\right] + C \tag{B.11}$$

The initial condition that $N_1 = 0$ when $\mu_2 = \mu_2^0$ is used to find the integration constant C

$$C = \mu_2^0 - RT\left[\ln\left(\frac{1}{\overline{V}_2}\right) - \frac{B_2}{\overline{V}_2} + \frac{B_3}{2\overline{V}_2{}^2}\right] \tag{B.12}$$

where μ_2^0 is the chemical potential of the pure protein species. The integration constant C is inserted back into equation (B.11), and μ_2 becomes

$$\frac{\mu_2 - \mu_2^0}{RT} = \ln\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right) + \left(\frac{2B_2}{\overline{V}_2} - 1\right)\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right) + \left(\frac{3B_3}{2\overline{V}_2^2} - \frac{B_2}{\overline{V}_2}\right)\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^2 - \frac{B_3}{\overline{V}_2{}^2}\left(\frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2}\right)^3 + 1 - \frac{B_2}{\overline{V}_2} - \frac{B_3}{2\overline{V}_2{}^2} \tag{B.13}$$

The chemical potentials can be expressed in terms of the protein volume fraction ϕ and the osmotic virial coefficients. The protein volume fraction is

$$\phi = \frac{N_2\overline{V}_2}{N_1\overline{V}_1 + N_2\overline{V}_2} \tag{B.14}$$

The osmotic virial coefficients are defined as

$$\begin{aligned} B_{22} &= \frac{B_2}{MW^2} \\ B_{222} &= \frac{B_3}{MW^3} \end{aligned} \tag{B.15}$$

Using these relationships, the chemical potential of the solvent in equation (B.10) and the chemical potential of the protein in equation (B.13) can be rewritten in terms of ϕ, B_{22}, and B_{222}

$$-\frac{\mu_1 - \mu_1^o}{RT\bar{V}_1} = \frac{\phi}{\overline{V}_2} + B_{22}\left(\frac{\phi MW}{\overline{V}_2}\right)^2 + B_{222}\left(\frac{\phi MW}{\overline{V}_2}\right)^3 \tag{B.16}$$

$$\begin{aligned} \frac{\mu_2 - \mu_2^0}{RT} = \ln\phi &+ \left(\frac{2B_{22}{MW_p}^2}{\bar{V}_2} - 1\right)\phi \\ &+ \left(\frac{3B_{222}{MW_p}^3}{2\overline{V}_2^2} - \frac{B_{22}{MW_p}^2}{\bar{V}_2}\right)\phi^2 \\ &- \left(\frac{B_{222}{MW_p}^3}{\overline{V}_2^2}\right)\phi^3 + 1 - \frac{B_{22}{MW_p}^2}{\overline{V}_2} - \frac{B_{222}{MW_p}^3}{2{\overline{V}_2}^2} \end{aligned} \tag{B.17}$$

For dilute concentrations of protein, it can be assumed that the mixture is ideal enough that the partial molar volume of the protein is equal to its molar volume, $\overline{V}_2 = \underline{V_2}$. Applying the conditions represented in equation (B.1) yields the model for liquid-liquid equilibrium

$$\frac{\phi_I - \phi_{II}}{\underline{V_2}} + B_{22}\left(\frac{MW}{\underline{V_2}}\right)^2(\phi_I^2 - \phi_{II}^2) + B_{222}\left(\frac{MW}{\underline{V_2}}\right)^3(\phi_I^3 - \phi_{II}^3) = 0 \tag{B.18}$$

$$\ln\left(\frac{\phi_I}{\phi_{II}}\right)+\left(\frac{2B_{22}MW^2}{\underline{V_2}}-1\right)(\phi_I-\phi_{II}) \\ +\left(\frac{3B_{222}MW^3}{2\underline{V_2}^2}-\frac{B_{22}MW^2}{\underline{V_2}}\right)(\phi_I^2-\phi_{II}^2)-\left(\frac{B_{222}MW^3}{\underline{V_2}^2}\right)(\phi_I^3-\phi_{II}^3)=0 \quad \text{(B.19)}$$

in which ϕ_I and ϕ_{II} are the concentration of protein in the light and dense liquid phases, respectively. Equations (B.18) and (B.19) can be solved simultaneously to obtain the equilibrium concentrations in the light and dense liquid phases, provided that the physical properties of the protein (MW and $\underline{V_2}$) are specified and the virial coefficients B_{22} and B_{222} at known solution conditions are used as inputs

.

Appendix B

CALCULATION OF THE OSMOTIC THIRD VIRIAL COEFFICIENT B_{222}

This appendix contains the MATLAB source code for the calculation of B_{222} from the Yukawa potential that is discussed in Chapter 2. The program utilizes MATLAB's numerical integration toolbox. In order to run this code, the numerical integration toolbox needs to be in the same directory as the program files during execution. This toolbox contains the functions implemented in the code for the numerical integration of the third virial coefficient, which can be obtained from the website http://www2.math.umd.edu/~jmr241/mfiles/nit/.

The file names and descriptions are given below. The range parameter *b* is set at a value 35 in line 65 of the main driver file B222_Yukawa.m; however, this value can be adjusted by the user. The user is also free to adjust the values of the parameters in the file parameters.inp. Attached is a copy of the input files that are needed. In principle, this code can be modified for any potential that possesses hard-sphere repulsion with an attractive tail. It is left to the user to make such modifications.

File: *parameters.inp* – Input file which contains physical parameters

```
Temp    296
R       8.314
k       1.38E-23
Na      6.023E23
e       1.60E-19
epsr    78.54
eps0    8.85E-12
MW1     18.02
rho1    0.998
MW2     13700
rho2    4.5
sigma   3.1E-9
phi_c   0.5
```

File: *B22_Data.dat* – Input file which contains the experimental B_{22} data

```
-1.27E-04
-1.55E-04
-1.88E-04
-2.24E-04
-2.65E-04
-3.10E-04
-3.61E-04
-4.17E-04
-4.79E-04
-5.48E-04
-6.24E-04
-7.07E-04
-7.98E-04
-8.97E-04
-1.01E-03
-1.12E-03
-1.25E-03
-1.39E-03
```

File: *B222_Yukawa.m* – Main driver file for B_{222} calculation

```
% This MATLAB program is used to calculate the third virial
% coefficient from the Yukawa potential using the method of
% Alder and Pople, J. Chem Phys., 1957. To run this file,
% the numerical integration toolbox needs to be in the directory.
%
% INPUT
%      parameters.inp
%           File with necessary physical parameters for calculations
%
%      B22_Data
%           File with the experimental B22 data. This data is used to
%           fit the epsilon parameter of the Yukawa potential for a
%           fixed value of the range parameter b*.
%
% OUTPUT
%      T*
%           Reduced temperature T/epsilon
%      B222
%           Osmotic third virial coefficient
%
% Authors: Leigh J. Quang
%          Abraham Lenhoff
%          Stanley Sandler
%
% Last Modified: 11/29/2011
% University of Delaware
% Department of Chemical Engineering

clc
clear all
close all
clear global
addpath nit
global T k Na e epsr eps0 MW1 V1 rho1 phi_c MW2 m s

%% Experimental Data
% Read in data
fid = importdata('parameters.inp');

% Constants
T = fid.data(1);        % Temperature (K)
R = fid.data(2);        % Molar gas constant (J/molK)
k = fid.data(3);        % Boltzmann constant (J/K)
Na = fid.data(4);       % Avogadro's number (1/mole)
e = fid.data(5);        % Elementary charge (C)
```

```
epsr = fid.data(6);      % Dielectric constant of solvent
eps0 = fid.data(7);      % Permitivity constant

% Water Properties
MW1 = fid.data(8);       % Molecular weight (g/mol)
rho1 = fid.data(9);      % Density (g/ml)
V1 = MW1/rho1;           % Molar volume (ml/mol)

% Protein Properties
MW2 = fid.data(10);      % Molecular weight (g/mol)
rho2 = fid.data(11);     % g/cm^3
V2 = MW2/rho2;           % Molar volume (ml/mol)
s = fid.data(12);        % Diameter (m)
m = (rho1*MW2)/(rho2*MW1);
phi_c = fid.data(13);

%% B22 Data
ydata = importdata('B22_Data.dat');
n = length(ydata);

%% Range Parameter
b = [35];
m = length(b);

%% B222 Calculation
syms x y r z
epsilon = zeros(n,m);
B3 = zeros(n,m);
Temp = zeros(n,m);
rmax = 5;
for j = 1:1:m
    for i = 1:1:n
        % Yukawa Epsilon Parameter Optimization
        resid = @(eps) norm(B2(eps,b(j))- ydata(i));
        epsilon(i,j) = fminsearch(resid,500);

        % Third Virial Coefficient from Yukawa
        Tr = T/epsilon(i,j);
        wx = -exp(-b(j)*(x-1))/x;
        wy = -exp(-b(j)*(y-1))/y;
        wz = -exp(-b(j)*(z-1))/z;
        fx = exp(-wx/Tr)-1;
        fy = exp(-wy/Tr)-1;
        fz = exp(-wz/Tr)-1;

        C1 = 5/8;
        C2 = -12*numint2((1-0.75*x+0.0625*x^3)*x^2*fx,x,1,2,y,0,1);
        C3 = 36*[numint3(x*y*z*fx*fy,z,x-y,1,y,1,x,x,1,2) + ...
                 numint3(x*y*z*fx*fy,z,x-y,1,y,x-1,x,x,2,rmax)];
        C4 = -12*[numint3(x*y*z*fx*fy*fz,z,1,x+y,y,1,x,x,1,2) + ...
             numint3(x*y*z*fx*fy*fz,z,x-y,x+y,y,1,x-1,x,2,rmax) + ...
```

```
                    numint3(x*y*z*fx*fy*fz,z,1,x+y,y,x-1,x,x,2,rmax)];
        B3(i,j) = C1+C2+C3+C4;
        Temp(i,j) = Tr;
    end
end

%% Output results
format short
disp('T*:')
disp(Temp)
disp(' ')
format short eng
disp('3rd Virial Coefficient B222 (cm^6*mol/g^3):')
disp(' ')
B222 = B3.*[((2*pi*Na*s^3)/3).^2.*100^6.*(1/MW2^3)];
disp(B222)
```

File: *B2.m* – Integral equation for B_{22} in terms of reduced variables

```
function F = B2(eps,b)

% Function file for solving the B22 integral equation to determine
% epsilon. The range parameter b must be passed to this function.

global T k Na e epsr eps0 MW1 V1 rho1 phi_c MW2 m s

F = ((2.*pi.*Na.*s.^3)./3).*(1 + quadgk(@(r) f(r,eps,b),1,Inf));

end
```

File: *f.m* – Mayer cluster function for the Yukawa potential

```
function F = f(r,eps,b)

% Function file that contains the Mayer cluster function for the
% Yukawa potential. This function is using in the function file B2.m
% for MATLAB's quadgk integration function.

global T k Na e epsr eps0 MW1 V1 rho1 phi_c MW2 m s

w = -eps.*exp(-b.*(r.^(1/3)-1))./r.^(1/3);
F = 1 - exp(-w/T);

end
```