Rational Design of Additives For Inhibition of Protein Aggregation

by

Diwakar Shukla

B.Tech., Chemical Engg., Indian Institute of Technology Bombay, 2006
M.Tech., Chemical Engg., Indian Institute of Technology Bombay, 2006
M.S., Chemical Engg. Practice, Massachusetts Institute of Technology, 2009

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2011
(September 2011)

Author .....
Department of Chemical Engineering
May 20, 2011

Certified by .....
Bernhardt L. Trout
Professor of Chemical Engineering
Thesis Supervisor

Accepted by .....
William M. Deen
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Rational Design of Additives For Inhibition of Protein Aggregation

by

Diwakar Shukla

Submitted to the Department of Chemical Engineering
on May 20, 2011, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

Abstract

Protein based therapeutics hold great promise in the treatment of human diseases and disorders and subsequently, they have become the fastest growing sector of new drugs being developed. Proteins are, however, inherently unstable and the degraded form can be quite harmful if administered to a patient. Of the various degradation pathways, aggregation is one of the most common and a cause for great concern. Aggregation suppressing additives have long been used to stabilize proteins, and they still remain the most viable option for combating this problem. However, the mechanisms by which the most commonly used additives inhibit aggregation still remain a mystery for the most part. It is clear that additive selection and the development of better performing additives will benefit from a more refined understanding of how commonly used additives inhibit or enhance aggregation. Aqueous arginine solutions are widely used to suppress protein aggregation and protein-protein interactions. Attempts have been made to develop cosolvents that are similar to arginine, but more effective at inhibiting aggregation. Therefore, a clear picture of the mechanism by which arginine inhibits protein aggregation is desirable. Baynes and Trout have proposed the design of a novel class of additives called "Neutral Crowder", which does not affect the free energy of isolated protein molecules but selectively increases the free energy of the protein-protein encounter complex. They proposed that arginine can be a "Neutral crowder" as the magnitude of the observed aggregation suppression effect of arginine is quantitatively equivalent to a neutral crowder of its size.

On the basis of the results obtained in this thesis, we have been able to show that self-interaction of arginine plays a critical role in the mechanism by which it inhibits aggregation. The preferential interaction between protein and arginine is also influenced by the intra-solvent interactions in aqueous arginine solutions, something that is often overlooked and yet essential to understanding the effect of additives on aggregation. Furthermore, the linking together of arginine clusters into bigger clusters by hydrogen bond accepting counterions enhances its aggregation suppressing ability.

According to the "Neutral Crowder" theory, large molecules that have the same concen-

tration on the protein surface as the bulk solution should be effective at inhibiting protein association. However, large molecules naturally tend to be excluded from protein surfaces (e.g. polyethylene glycol) due to steric exclusion. We theorized, though, that if functional groups which tend to preferentially bind to proteins (e.g. guanidinium, urea, etc.) were added to the surface of a large, core structure that the resulting molecule could potentially behave as a neutral crowder. Therefore, creating a neutral crowder molecule requires a balance between attraction and repulsion with respect to the surface of a protein. Choosing a proper balance of interactions allowed us to produce compounds which have been shown to be potent aggregation suppressors, slowing aggregation by an order of magnitude more than the commonly used additives. Such potent aggregation suppressing additives might be useful during production and formulation, as they could improve yield and extend the shelf-life of protein therapeutics.

Thesis Supervisor: Bernhardt L. Trout
Title: Professor of Chemical Engineering

Acknowledgements

I would first of all like to thank Professor Bernhardt L. Trout for his constant support, help, and for giving me the freedom to pick and choose my projects during the course of this thesis. I would also like to thank my committee members, Professor Amy Keating, and Professor Daniel Blankschtein, their comments and discussions through out the course of the project were helpful. Outside the graduate research work, my tenure as a teaching assistant with Professor William M. Deen helped me develop interest in teaching and deepened my desire to pursue an academic career.

Dr. Curtiss P. Schneider, who worked along with me on the experimental aspect of this project has provided useful insights, which helped me in choosing correct research directions. He has always been very enthusiastic about conducting experiments related to the ideas that I proposed based on the simulation results. Without his helpful comments and experimental insights, this project would not have taken its present shape. There have been many present and former group members who have helped me get started with the project and helped me at various times during the past five years. I would like to thank the former Trout group members Dr. Naresh Chennamsetty, and Chetan Shinde for helping me start with the molecular simulations. I would like to thank the current Trout group member, Dr. Erik E. Santiso, who helped me get started with the implementation of the free energy calculation techniques in NAMD. I would also like to thank the rest of Trout group members for their helpful inputs at the group meetings and their company.

Professor Anurag Mehra, my master's thesis advisor at IIT Bombay, has been a great mentor and a friend. His philosophy of life and his practical suggestions have always been helpful, especially during the graduate school. Lastly and most importantly, I would like to thank with all my heart my parents, to whom this thesis is dedicated, for their advice, love, support, and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Protein aggregation is probably the most common, least understood and most problematic form of protein degradation.[12–14] It occurs in almost every phase of development[15] and essentially all proteins and polypeptides are prone, to some degree, to the formation of non-native aggregates.[16] The presence of aggregates in an injected solution, even in small doses, poses a grave threat of an immune response that not only can reduce the efficacy of the product over time, but more importantly, has the potential to elicit adverse reactions which can put the life of the patient in jeopardy.[17] With such dire and uncertain consequences, it is generally accepted that biopharmaceutical formulations must be substantially free of aggregates and their formation inhibited during storage,[18] which is a difficult challenge to achieve, especially considering that high protein concentration formulations are often desired.[19] With the past 25 years seeing an explosive growth in the number of protein based therapeutics developed, great interest has been sparked in developing improved formulation methods for preventing aggregation.[14,20]

It has long been known that the presence of small molecular weight species (e.g. sugars, polyols, salts, amino acids, etc.) can greatly influence the stability of a protein in solution.[21–23] Therefore, aggregation suppressing additives (often called cosolutes, cosolvents, and excipients) have long been utilized during production, purification, and formulation and they still remain the most viable option for dealing with this problem.[13,22] Most often, some

type of sugar or polyol (e.g. sucrose, trehalose, glycerol, sorbitol, etc.), in combination with other components that influence protein stability (buffers, salts, surfactants, etc.), is used to inhibit aggregation.[24] However, no one formulation recipe works well for all proteins and other excipients have been used and/or are gaining more attention (e.g. amino acids, polymers, proteins, etc.).[25] Furthermore, there is a desire to discover or create better performing excipients because for many proteins, a stable liquid formulation cannot be created, and the product must be lyophilized and reconstituted prior to injection, which for the most part is undesirable, and, in some cases, a difficult and costly challenge.[14] To accomplish such a task, a better understanding of how commonly used additives inhibit aggregation needs to be established. Such information will also help to improve the methodology by which additives are selected because selection is almost always made via an ad hoc trial-and-error process using empirically derived heuristics. This can be a lengthy process that can delay the release of a product, resulting in the loss of potential sales. By incorporating detailed mechanistic understandings of various excipients with the known structure of a protein, ideal formulation recipes will likely be developed more quickly and maybe predicted prior to the formulation phase. However, mechanistic inquiries tend to be narrow in focus in terms of the contributing factors that influence stability and often serve to promote previously proposed models. Here, we review fundamental concepts employed to understand the effect of additives on proteins, some recent molecular level inquiries on a few select additives to highlight overlooked contributions to the suppression of protein aggregation and to establish the importance of the key objectives of this thesis. Most notably, but not exclusively, we take an interest in intra-solvent interactions (i.e. interactions between additive molecules), which are often ignored or not considered when studying protein-additive interactions but is something that has been gaining attention lately due to the availability of molecular dynamics simulations. The need to study such interactions is evident because proteins are seldom formulated with just a single cosolute and stability might be improved if intra-solvent interactions are tuned appropriately to enhance stabilizing effects.

1.1 Preferential Interactions

Though the topic is well established and has been reviewed many times before,[26,27] it will be helpful to briefly address the concept of preferential interactions because the most dominant influence an additive may have on the physical properties of a protein stem from whether the additive is attracted or repelled from the surface of a protein[28] and this concept is fundamental in addressing protein-additive interactions on a molecular level.[1] In addition, as new insight is shed on the mechanisms of various aggregation suppressing additives, it is important to compare such results to what has generally been accepted as contributing factors in the stabilization. This is critical in developing a clear mechanistic picture because many current mechanistic explanations have been built on indirect evidence.

1.1.1 Binding and Exclusion

It is well established that if the concentration of an additive in the local domain around a protein differs from the concentration in the bulk solution, significant changes in the thermodynamic properties of the protein will arise that influence solubility and conformational stability.[26] The most often used method to quantify such behavior is via measuring or calculating the preferential interaction coefficient, Γ_{23}, which is a measure of the preference a cosolute has for the protein surface and is defined by the following expression,

$$\Gamma_{23} = \left(\frac{\partial m_3}{\partial m_2}\right)_{T,P,\mu_3} = -\left(\frac{\partial \mu_2}{\partial \mu_3}\right)_{T,P,m_2}, \tag{1.1}$$

where m, T, P and μ represent molal concentration, temperature, pressure, and chemical potential, respectively.[29] The subscripts used indicate solution components in Scatchard notation: water (subscript 1), the protein (subscript 2), and the cosolute (subscript 3). Additives with a positive Γ_{23} are typically described as being preferentially bound to the protein surface due to an increase in the concentration of the cosolute in the local domain and this favorable interaction, as indicated by Equation 1.1, lowers the chemical potential of the protein. The opposite is true for additives with a negative Γ_{23}, which are typically

described as being preferentially excluded from the surface of the protein.

Preferential binding or exclusion can arise from either indirect (often called nonspecific interactions) with the protein surface or direct (i.e. specific) interactions with individual amino acid residues or the peptide backbone.[26] Most notable of the nonspecific interactions are volume exclusion and perturbation of surface free energy. To elucidate, additives larger than water will tend to induce preferential hydration, which is equivalent to preferential exclusion, due to there being a solvation layer around the protein which is accessible to water but not the additive.[30] Furthermore, if the additive changes the surface tension of water, it will tend to be either depleted (higher surface tension) or accumulated (lower surface tension) at the protein-solvent interface in accordance to the Gibbs absorption isotherm given that the protein surface constitutes a liquid interface.[28] However, one is cautioned in drawing conclusions directly from surface tension measurements at an air-water interface considering that air is a poor model for a protein surface.[23]

Direct interactions with amino acid residues or the peptide backbone may arise due to electrostatic interactions,[31] hydrogen bonding,[32] cation-π interactions with aromatic residues,[33] hydrophobic interactions,[34] and solvophobic effects.[35] The last category of interaction is often used to describe additives which enhance the repulsion between the solvent and hydrophobic residues due to how the additive is solvated by water, giving rise to the term "kosmotrope" and the opposing term "chaotrope", which are additives that reduce repulsive interactions with hydrophobic residues.[35] Such interactions are often referred to as indirect given that the interaction arises from how the additive changes the structure of water. These terms have long been used to categorize additives and for the most part, have lost their original meaning. Though the mechanism is often used to explain how certain additives can enhance or disrupt the hydrophobic effect that cause proteins to fold and hydrophobic particles to aggregate,[35] it is still heavily debated to this date given that it is unclear to what extent an additive may change the structure of water.[36]

1.1.2 Intra-Solvent Interactions

As pointed out above in the above section, preferential interactions can arise from a number of direct and indirect interactions between the protein and additive. However, we would like to point out in this section, in prelude to the discussion later in the chapter, that intra-solvent interactions between additive molecules will influence how the molecules interact with the protein. This is a general theme throughout this thesis, as the implications of such interactions have only recently come to light in determining the aggregation suppression mechanisms of a variety of additives. Of particular interest is the effect a counterion may play in the interaction with a protein.

The Hofmeister Series is a well know and often cited empirical ranking of how commonly used ions influence protein solubility and stability. The behavior can be correlated with preferential interactions but it is still unclear what gives rise to the differing behavior between the ions. Recent studies, though, have discovered that the extent of hydrogen bond formation between the cation and anion contribute to this behavior, suggesting that strong attractive interactions will lead to clustering and will inhibit an ion that would otherwise bind to a protein from making such an interaction. This is of particular importance in explaining the behavior of guanidinium. A strikingly similar behavior is also observed for nonionic additive mixtures, such as urea-TMAO, suggesting that the behavior of a particular additive will be influenced by other cosolutes if the formulation has multiple components.

1.1.3 Thermodynamic and Kinetic Effects

1.1.3.1 Conformational and Colloidal Stability

It can easily be shown and understood that preferential interactions will influence any reaction the protein may undergo if the extent of the preferential interaction differs between the product (P) and reactant (R) states, as described by the Wyman Linkage Function:[37]

$$-\left(\frac{\partial \Delta G^{\circ}}{\partial \mu_3}\right) = \left(\frac{\partial ln K}{\partial ln a_3}\right) = \Gamma_{23}^{P} - \Gamma_{23}^{R}. \tag{1.2}$$

The two main reactions of interest, unfolding (denaturation) and precipitation (salting-out), result in changes in solvent accessible surface area (SAA), which is directly linked to the extent of preferential interactions,[38] whether the additive is bound or excluded. As a result, if the nature of the interaction with the protein does not change, bound additives will tend to induce a state with the most surface area exposed (unfolded and dissolved), while excluded additives will tend to induce a state with the least amount of surface area exposed (folded and precipitated). Though this trend is absolute for precipitation, it is only generally true for unfolding due to the exposure of hydrophobic residues and a reduction in the density of charged residues on the protein surface, both of which can change the nature of the protein-additive interaction. Therefore, other thermodynamic techniques which involve thermal unfolding (e.g. DSC, CD Spec., etc.) are required to confirm how a cosolute influences conformational stability.[26] Such effects play an important role in protein aggregation by two different means: conformational and colloidal stability.[21] The most dominant factor that causes protein aggregation is thought to be the reduced exposure of hydrophobic patches.[39] The unfolded or partially unfolded states are more prone to aggregation due the exposure of hydrophobic residues that are typically buried in the native state and the structure of a protein is not static, making it likely that the native state is in equilibrium with a number of different partially unfolded species.[40] Whether or not partial unfolding is the rate limiting step in aggregation, additives which tend to promote a compact, native state (e.g. sugars, polyols, etc.) will often inhibit aggregation by either slowing the unfolding step or reducing the equilibrium concentration of aggregation prone species.[16] On the other hand, if aggregation proceeds through a nucleated polymerization pathway, additives which increase protein solubility may inhibit aggregation by shifting the nucleation equilibrium toward the dissociated state.[41] And lastly, disrupting protein-protein interactions, or making proteins more colloidally stable, is often discussed as a major contributing factor towards the inhibition of aggregation.[21] This is something that is not always well captured by the overall preferential interaction coefficient, thus the need to study osmotic second virial coefficients and to perform MD simulations to elucidate molecular level interactions. In particular, the

disruption of protein-protein interactions which provide or reduce colloidal stability are often discussed, most notable are ones that are electrostatic or hydrophobic in origin.

1.1.3.2 Gap Effect

While the thermodynamic effects preferential interactions have on the solubility and conformational stability of a protein have long been established, up until recently, it was unclear how such interactions directly influence the rate of protein association. Baynes and Trout[42] investigated this through computational methods and developed a model which incorporates additive size and preferential interactions into the relative rate of protein association. Consistent with depletion forces, a well-established aspect of colloid science,[43] for a given additive size, the model predicts that the more excluded the additive, the more it enhances association, a fundamental phenomenon in crowded media[44] and something often observed for large additives, such as PEG and dextrin.[45,46] Likewise, bound additives were predicted to inhibit association in relation to the degree of preferential binding, consistent with how denaturants increase solubility and stabilize unfolded proteins against aggregating.

The key observation, though, was that as the additive size was increased while holding the preferential interaction constant, the rate of association decreased by several orders of magnitude for all types of additives via a "Gap Effect" mechanism (see Figure 1-1, which shows a drawing based on the original results). That is, as two protein molecules associated, a gap formed in which the additive was two large to solvate but still large enough for hydration, thus leading to an increase in the free energy of the encounter complex due to the exclusion of additives from this gap. This effect, of course, was more pronounced for larger additives, due to an increase in the size of the gap, and for bound additives, which had a greater affinity to solvate the protein. An important consequence of this phenomenon, though, is that additives which are neither preferentially bound nor excluded will slow association if they are much larger than water, thus exerting a purely kinetic effect on protein aggregation. Such a hypothetical molecule is referred to as a "neutral crowder and is comparable to the behavior of arginine, a unique additive with such behavior. It should be pointed out for

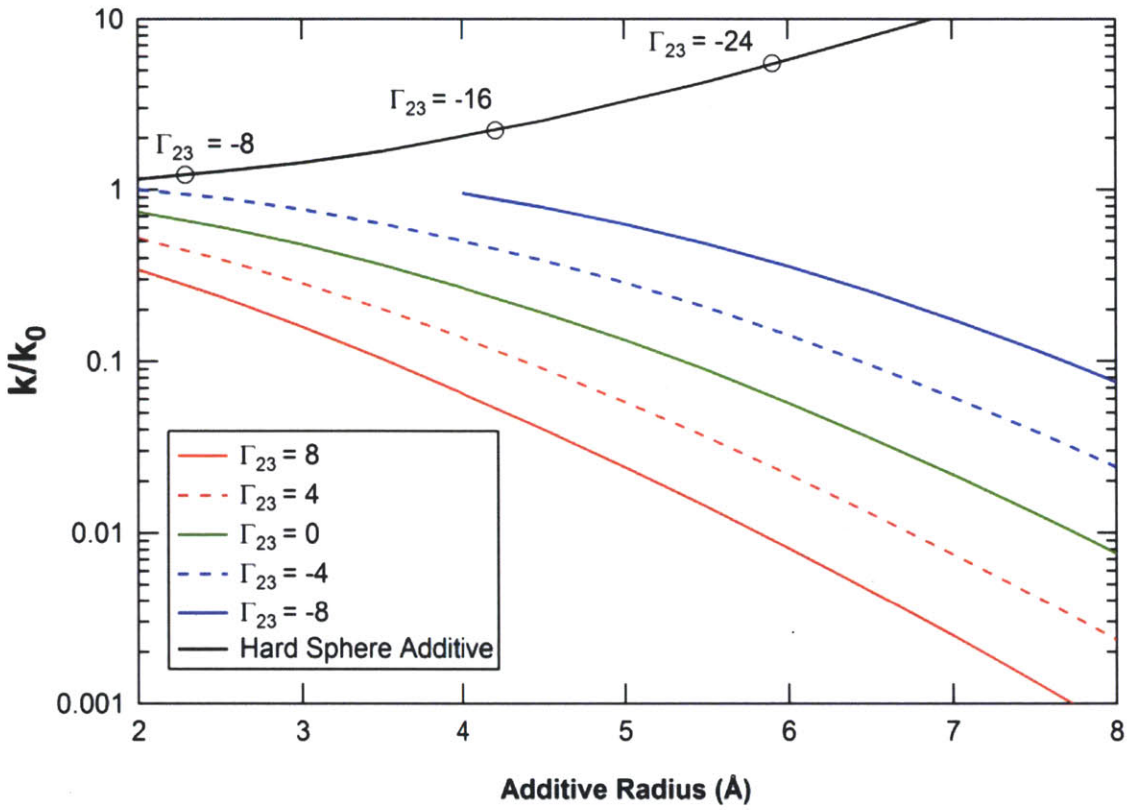

Figure 1-1: **"Gap Effect" predictions of the change in the rate of protein association, shown as the relative rate constant, as a function of additive size and preferential interaction. The model was applied to the association of two spherical protein molecules with a radius of 20 Å in the presence of a 1 M solution of spherical additives. The plot also includes predictions using a "hard sphere" potential, which serves as a limit for enhanced association via a depletion force resulting from preferential exclusion.**

clarity that the "Gap Effect" model is fundamentally different from how depletion forces arise when large colloidal particles (e.g. proteins) are immersed in a fluid of smaller colloidal particles (i.e. hard sphere additives). In such a case, only steric exclusion interactions exist (i.e. $\Gamma_{23} = \Gamma_{23}^{StericExclusion} \approx -C_3 V_S$, where C_3 is the molar concentration of the additive and V_S is the volume of the excluded shell) and thus, when the excluded volume shells overlap when two proteins approach each other, a gap of pure water forms at a separation distance less than the diameter of the "hard sphere" additive. This essentially creates an osmotic pressure force on the two protein molecules, forcing them together and thus, enhancing the rate of

association (see Figure 1-1, which includes a prediction for "Hard Sphere" additives, showing that the Baynes and Trout model incorporates depletion force effects). This phenomenon is entropic in nature since the association of the protein molecules decreases the volume of exclusion, giving the "hard sphere" additives more volume to solvate. However, the case when $\Gamma_{23} > \Gamma_{23}^{StericExclusion}$ (all values below the "Hard Sphere Additive" curve in Figure 1-1) implies that attractive interactions exist between the protein and the additive that counteract the steric exclusion. Since large additives cannot penetrate the excluded volume shell, such attractive interactions increase the concentration of the additive above that of the bulk concentration in the solvation layers immediately outside the excluded volume region (observed as peaks in radial distribution functions). Such an increase in concentration in the solvation layers around the protein will obviously counteract the depletion effect because of the loss of this more concentrated volume when two protein molecules associate. In other words, as Γ_{23} increases for a given additive size, there is an ever increasing concentration difference between the solvation layers and the gap of pure water, an entropically unfavorable condition. At some point this exclusion will exceed the depletion force effect and the additive will then inhibit association, giving rise to the Gap Effect mechanism described by Baynes and Trout. As shown in Figure 1-1, for large additives, this occurs at Γ_{23} values well below zero, making it possible for additives with a $\Gamma_{23} = 0$ capable of inhibiting association.

1.2 Molecular picture of Protein-Additive Interactions

Now that the general effects of solution additives have been discussed, the remainder of the chapter will focus on the current understanding of the mechanisms by which a select group of commonly used additives affect protein stability. Of course, these additives have been the subject of numerous reviews and research articles over the past three decades. The intent of this section is to report recent mechanistic insights obtained from a molecular interactions perspective. In particular, we have focused on the effects of additive-additive and direct protein-additive interactions, which up to this point have not been considered heavily in the pursuit of understanding how additives inhibit aggregation. In a typical protein formulation,

a variety of additives might be present because of the high cost of removing an additive added to the protein solution during processing. However, the presence of multiple additives in a formulation leads to a scenario where the interplay between multitudes of interactions determines the overall stability of the drug. Therefore, it is critical to understand the effect of a particular additive on all other molecular interactions in the formulation.

1.2.1 Urea

Even though urea is a widely used denaturant, understanding how the molecule forms attractive interactions with proteins will be useful in developing a complete understanding of protein-additive interactions. The mechanism of urea induced protein denaturation has been studied extensively from both theoretical and experimental viewpoints; however, no clear consensus has been established given that the mechanism is still actively debated.[47] The two basic lines of reasoning suggest that urea denatures proteins via either a direct (interaction of urea with the protein) or an indirect (effect of urea on water structure) mechanism, however, as new insights emerge, the two mechanisms are likely not mutually exclusive, though they are often treated as such.[47,48] The direct effect is gaining more attention as the main driving force for denaturation; however, the means by which this interaction arises is still disputed, with recent evidence suggesting stronger interactions with nonpolar groups rather than the often claimed hydrogen bonding to polar residues.[49]

Given the circumstances of this ongoing debate, we cannot effectively give a conclusive view of the urea mechanism. However, we would like to highlight an overlooked interactions in urea solutions, and emphasize the need to understand all molecular interactions while designing formulations. It has been speculated for more than a half a century[50] and now widely accepted, that urea has a tendency to reversibly self-associate in solution.[51] Today, however, this phenomenon is not often taken into consideration in explaining how urea denatures proteins and the implications of urea self-association are far from clear. Recently, Stumpe and Grubmuller[32] showed that the association of urea molecules contributes to the indirect effect but more importantly, their results indicate how urea might be able to pref-

erentially interact with apolar groups, contributing to the direct effect. Urea substitutes well for water, geometrically, in the hydrogen bond network but energetically, the strength of the hydrogen bonds are quite different. Water-water hydrogen bonds are stronger than water-urea or urea-urea hydrogen bonds. This difference in hydrogen-bond strength leads to urea self-interaction and strengthens the water structure, giving insight into how urea may interface between less polar residues and water. Other researchers have since incorporated urea association in mechanistic models that take into account both indirect and direct interactions, giving rise to the idea that urea association will induce concentration dependent behavior.[52]

Along the lines of additive-additive interactions, the counteracting effect of trimethylamine N-oxide (TMAO) on urea induced protein denaturation provides a classic and naturally occurring example of the effect intra-solvent interactions have on protein stability.[53] Early studies suggested that the two osmolytes acted independently of each other. Later on, MD simulations suggested that TMAO strengthens the urea-water interaction, thereby limiting urea-protein interactions by inhibiting urea-protein hydrogen bonds.[54] Recent studies indicate strong TMAO-urea interactions, with the TMAO-urea hydrogen bond stronger than the TMAO-water hydrogen bond, leading to the hypothesis that urea and water prefer to solvate TMAO, rather than the protein. These studies reiterate the need to understand all possible interactions in aqueous protein-additive solutions.[55]

1.2.2 Guanidinium

From a preferential interaction perspective, salts are typically treated as single components despite the presence of two or more ions in solution. The behavior of the cation and anion could differ widely, not only in terms of their interactions with the protein surface, but also in terms of their self-interaction. The role of intra-solvent interactions in protein-protein interactions is most obvious for the case of guanidinium (Gdm) salts.[56,57] GdmSCN and GdmCl are protein denaturants, whereas, $(Gdm)_2SO_4$ is effectively neutral in its effect on protein stability.[58,59]

The effect of the ions on proteins have been explained in terms of the changes induced by these salts on the water structure, with sulfate salts labeled as kosmotropes (structure makers) and SCN salts labeled as chaotropes (structure breakers).[59] However, changes in the water structure seems to be only limited to the first solvation layer around the ion and a relatively new picture of direct interactions of these ions with each other and with proteins has been gaining ground.[60] $(Gdm)_2SO_4$ has been shown to form mesoscopic clusters in solution and these clusters are formed due to the ability of Gdm and sulfate ions to form multiple hydrogen bonds with each other, which are stronger than the hydrogen bonds formed between ions and water.[59] GdmSCN shows a marked contrast in terms of ion pairing as compared to the sulfate salts, with limited or negligible interactions between Gdm and SCN ions.[56,59] The difference in the ion pairing behavior of these salts is likely a contributing factor in the reversal or enhancement of the denaturing ability of Gdm. For the sulfate salt, the binding of Gdm to the protein surface is limited due to the strong interaction between Gdm and sulfate ions, which make Gdm molecules unavailable for binding to the protein surface.[61,62] These conclusions were drawn based on not only molecular dynamics simulations but neutron scattering data of the Gdm salts as well. The presence of clusters could affect solution properties such as viscosity, protein diffusivity, etc., that could influence the rate of aggregation, however, the contribution from such changes have yet to be investigated.

The preferential interaction coefficient values at concentration of 1 M for BSA in the presence of Gdm salts show that the sulfate salt ($Gdm(SO_4)_{1/2}$) is excluded ($\Gamma_{23} = -8$) and the chloride salt is highly bound ($\Gamma_{23} = 18$).[63] The contrasting clustering behavior for these salts provides a more realistic explanation of these observed preferential interaction values. Before, it was believed that the effects from each ion were additive and the net contribution was responsible for the behavior. To elaborate, for $(Gdm)_2SO_4$ and other Hofmeister salts, the cation and anion were thought to act independently of each other and the resulting preferential interaction coefficient and influence on stability was simply an average of the effect the two solutes imposed. Such an explanation is satisfactory for a mixture of uncharged solutes, however, for electrolytic solutions, such a scenario would result

in an unfavorable charge separation for $(Gdm)_2SO_4$, with the Gdm molecules bound and the sulfate ions excluded.[62] From a molecular interaction perspective, cation-anion clustering would not result in such a charge separation and the interaction with sulfate would limit the hydrogen-bonding and cation-π interactions of Gdm group with the protein, thus eliminating its denaturing effect. It is likely that the intra-solvent interactions exhibited in the series of studies featuring Gdm extends to the whole Hofmeister Series, but to a lesser degree given that Gdm and sulfate are on the two ends of the spectrum of hydrogen bond donating and accepting ions, respectively.[59]

Strong support for the clustering model over the previous model comes from a case in which a protein was found to have the unique behavior of being sensitive to GdmCl denaturation but insensitive to stabilization from alkali metal sulfates.[61,62] For this particular protein, $(Gdm)_2SO_4$ had no destabilizing effect, which leads to only one possible conclusion, that sulfate inhibited the binding of Gdm due to clustering because sulfate had no stabilizing effect on the protein. These results confirmed that ions could affect the self-interaction of each other and their interaction with the protein. For a binary salt solution, there are ten possible binary interactions in a protein-additive mixture with the possibility of each interaction affecting the other. However, as seen in the case of Gdm salts, only a few of these possible interactions play a dominant role in protein stability.

1.2.3 Arginine

The amino acid arginine is a fascinating case study in terms of the multitude of interactions it may form that influence how it interacts with proteins. To summarize, it (1) is large relative to water (volume exclusion), (2) increases the surface tension of the solution,[64,65] (3) is a salt (chloride form most common), thus electrostatic and counterion interactions come into play (4) is zwitterionic, thus has two other ionic charge locations, (5) has the hydrogen bond donating and protein denaturing functional group guanidinium, which allows it to interact favorably with the protein surface or hydrogen bond accepting groups, (6) has a hydrogen bond accepting carboxylate moiety, (7) has an amine group, another location for donating

hydrogen bonds and (8) has a hydrophobic alkyl chain three carbons long.

Various experimental observations have been made about the effects of arginine on protein-protein association reactions. It has been shown that arginine reduces attractive protein-protein interactions, as indicated by a shift from a negative to positive osmotic second virial coefficient measured by light scattering experiments,[66] increases the solubility of unfolded species of hen egg white lysozyme,[67] and decreases the rate of association of unfolded and partially folded intermediates on the folding pathway during refolding as measured by native protein activity and size-exclusion chromatography.[68] Attempts have been made to develop cosolvents that are similar to arginine, but more effective at inhibiting aggregation.[69,70] Therefore, a clear picture of the mechanism by which arginine inhibits protein aggregation is desirable. Many theories have been proposed to explain the effect of cosolvents on proteins.[3,42,65,71–74] In 1888, Hofmeister[71] ordered cations and anions according to their ability to stabilize protein solutions. The origin of this series has been attributed to the structural changes that ions cause in the water network. However, this rationale only applies to the simple monoatomic ions. For complex molecular cosolvents like arginine, several functional groups in the molecule produce complex solvent structuring patterns that depend on the nature of the functional groups and their relative positions. Arginine is an aggregation suppressor, but it increases the surface tension of water on addition, which is similar to the behavior of the protein denaturants, like guanidinium hydrochloride(GdmHCl).[65] Experimental studies to determine the solubility of amino acids in aqueous arginine reveal that arginine, like GdmHCl, interacts favorably with all the amino acid side chains, with both compounds showing strong interaction with aromatic residues.[65] However, the interaction of arginine with the protein surface is limited due to its large size as compared to Gdm. It is speculated that the limited binding of arginine plays a major role in its ability to suppress aggregation. Recently, Schneider and Trout[3] have reported an interesting trend in the interaction of arginine with proteins as a function of concentration and protein size. They observed that at low concentrations, arginine is bound to the protein surface ($\Gamma_{23} \sim 0$) but as the concentration increases, arginine becomes increasingly excluded from the protein

surface. They suggested that the possible reason for this non-linear exclusion of arginine from the protein surface is that the protein surface becomes saturated with arginine as the concentration is increased. The current understanding of the mechanism by which arginine inhibits aggregation is limited. There are three proposed hypothesis for the effect of arginine on the stability of protein solutions.

1. Tsumoto et. al.[66] suggested that interactions between the guanidine group of arginine and tryptophan side chains on the protein surface may be responsible for suppression of protein aggregation. Solubility of tryptophan is significantly increased in GdmHCl solutions due to the cation-π interactions.[65]

2. Baynes and Trout[42] proposed the "Gap effect" theory[42] could explain the aginine induced protein aggregation suppression. Neutral crowders (like arginine) do not affect the free energy of unfolding, and are, hence, "neutral", but due to their larger size as compared to water molecules they "crowd out" the protein-protein interactions. The authors proposed that arginine can be a "neutral crowder" as the magnitude of the observed aggregation suppression matches the theoretical prediction for that of a neutral crowder of a size of arginine.

3. Arginine molecules stack in a head-to-tail fashion, exposing their methylene groups as a hydrophobic column along one crystallographic axis.[75] Das et. al.[74] proposed that the arginine clusters in solution also display a hydrophobic surface by a similar alignment of arginine's three methylene groups. This hydrophobic surface can interact with the hydrophobic residues on the protein surface, which could inhibit protein aggregation. They showed that arginine increases the solubility of pyrene in water and modulates the hydrophobic interaction of Alzheimer's amyloid beta by binding to its surface.

Based on the above survey, it can be seen that there is no agreed upon mechanistic picture of the arginine-induced aggregation suppression.

1.3 Objectives and Outline of Thesis

The main objectives of this thesis are to (a) understand the mechanism by which arginine inhibits protein aggregation, and (b) design new additives based on the "neutral crowder" theory and the acquired understanding of the arginine mechanism.

The methodology for the estimation of Preferential interaction coefficients is reported in Chapter 2. Structure and Interactions in aqueous arginine solutions has been investigated to understand the intra-solvent interactions and their role in determining the overall protein stability (Chapter 3). Aqueous arginine solutions are also very effective as an eluent in affinity chromatography. We have used a molecular level approach to study the mechanisms by which arginine affect protein-protein interactions between an antibody and Protein-A. (Chapter 4) Preferential Interaction coefficients of proteins in aqueous arginine solutions show a unique trend as compared to other commonly used additives. On the basis of this study, the molecular level interactions responsible for the non-linear exclusion of arginine from the protein surface are identified (Chapter 5). The implications of the arginine mechanism established on the basis of the understanding of the protein-arginine and arginine-arginine interactions for the design of cosolvents are discussed in Chapter 6. In Chapter 7, the rational design approach based on the "neutral crowder" theory and the accquired understanding of the arginine mechanism is presented. In Chapter 8, we will present some of the salient conclusions of the thesis and possible future directions.

Chapter 2

Estimating Preferential Interaction Coefficients of Proteins

The theory of preferential binding and the concept of preferential interactions between protein and cosolvents have been proposed to understand the effect of cosolvents in biomolecular systems.[76] The preferential interaction coefficient is a purely thermodynamic quantity which measures the excess number of cosolvent molecules in the vicinity of the protein molecule as compared to the bulk cosolvent. The connection between the thermodynamic definition and the intuitive notion of binding[77,78] comes from statistical mechanics[79–81]

$$\Gamma_{23} = \left\langle n_3^{II} - n_1^{II}\left(\frac{n_3^I}{n_1^I}\right)\right\rangle \tag{2.1}$$

where n_j^i denotes the number of molecules of species j in domain i, and the angled brackets <> stand for ensemble average. Subscripts 1, 2 and 3 stand for water, protein and cosolvent respectively. Superscripts I and II stand for bulk and local domain respectively. When the cosolvent concentration is higher in the local domain of the protein as compared to the bulk domain, Γ_{23} is positive, indicating a favorable interaction. On the other hand, a lower cosolvent concentration in the vicinity of protein leads to negative Γ_{23}, indicating an unfavorable interaction. Thus, the modifier "preferential" essentially indicates that the protein has higher affinity (preference) for one solvent over other.

For several decades, dialysis/densimetry was the only established method for experimentally measuring preferential interaction coefficient for protein-cosolvent systems.[5,64,82,83] Recently, Courtenay and coworkers have utilized a new methodology based on vapor pressure osmometry to calculate preferential interaction coefficient.[84–86] In order to predict preferential interaction coefficients, several cosolvent interaction models have been proposed.[37,86–88] The most general model of cosolvent interaction comes from considering equilibrium of all possible protein cosolvent complexes.[37] Although the model is general, it requires estimation of large number of equilibrium binding constants which cannot be determined experimentally. Schellman[88] proposed a site exchange model which reduced the number of unknown constant to one by assuming that exchange reactions take place on fixed number of binding sites on protein surface which are estimated based on calorimetry data.[89] The model assumes independent binding sites on the surface of the protein with the same binding constant. Record and coworkers[86] developed the local-bulk domain model which measures solution composition difference between protein surface and bulk solvent. This model also requires measurement of an unknown constant, the partition coefficient between bulk and protein surface. Kirkwood-buff theory has also been used to calculate preferential interaction of cosolvents with proteins.[90,91] Recently, Local chemical potential equalization model which uses Kirkwood-Buff theory to calculate changes in cosolvent and water concentration near the protein surface has been proposed.[92] The local chemical potential equalization model also requires unknown constants which are determined either experimentally or by fitting the experimental preferential interaction data.

Molecular dynamics simulations can be used to calculate preferential interaction coefficients. This method can be used to estimate preferential interaction coefficient for systems where no experimental data is available. Baynes and Trout[1] performed the first molecular dynamics simulations to estimate the preferential interaction parameters. They were able to accurately compute the parameters on selected systems with 2 ns simulations. Recently, Kang and Smith[93] have performed simulations to evaluate preferential interaction of urea with lysozyme using KBFF and OPLS urea force fields while applying positional constraints

on the C^{α} atoms of the protein to prevent unfolding. They found that preferential interaction coefficient varied significantly with change in force field. However, both the force fields failed to predict the experimental preferential interaction coefficient value. They averaged over 5 ns of simulation time to determine preferential interaction values.

This study aims to elucidate the effect of simulation time, force field parameters and protein structure fluctuations on the preferential interaction of cosolvents with proteins by extended molecular dynamics simulations. To this aim, simulations with proteins lysozyme, α-chymotrypsinogen A, and RNase T1 and cosolvents urea, glycerol, arginine hydrochloride, guanidinium hydrochloride and glucose are performed to study the predictive ability of the approach for cosolvents with very different effect on proteins. In order to elucidate the sensitivity of the preferential interaction to changes in force field parameters, simulations with two different force fields for cosolvents urea and glycerol are performed. The first set of force field parameters are obtained from the urea model proposed by Duffy et al.[94] and glycerol parameters are based on analogy from carbohydrate parameters.[95,96] Recently, new force field parameters are reported for both these cosolvents,[2,4] simulations are performed with these parameters and compared with results obtained using the older force field parameters.

2.1 Methods

2.1.1 Simulation Setup

All simulations employed the CHARMM22[95] force field parameters for proteins and TIP3P[97] model for water. Structures of RNase T1 (PDB code: 1ygw), Hen egg white Lysozyme (PDB code: 1e8l) and α-Cgn A (PDB code: 2cga) were obtained from the Protein Data Bank.[98] Force field parameters for arginine were taken from the CHARMM22 force field. The CHARMM force field parameters are for arginine molecules with N and C terminals bonded to the neighboring residues in the protein structure. For arginine molecules present in the solution, the N terminal is protonated ($-NH_3^+$), and the C terminal is deprotonated (at pH of the systems investigated here) according to the CTER and NTER parameters avail-

Table 2.1: **Setup of Simulation systems**

Protein	cosolvent	cosolvent molecules	water molecules
RNase T1	Urea	90	4544
RNase T1	glycerol	87	4596
RNase T1	ArgHCl	90	4110
RNase T1	-	-	2932
Lysozyme	Urea	157	8353
Lysozyme	glycerol	157	7538
Lysozyme	ArgHCl	158	8479
Lysozyme	GdnHCl	173	9657
Lysozyme	Glucose	166	9109
α-Cgn A	Urea	220	11973
α-Cgn A	glycerol	210	11592
α-Cgn A	ArgHCl	188	10411
α-Cgn A	GdnHCl	207	11541
α-Cgn A	Glucose	205	10857

able in CHARMM. The force field for glucose was taken from the CHARMM carbohydrate simulation force field.[99] Force field parameters for guanidinium were taken from CHARMM parameters for arginine with the atomic partial charges assigned symmetrically.[100] Two different force field parameters for glycerol and urea were used. The first set of parameters for these cosolvents were taken from the parameters used by Baynes and Trout.[1] The second set of parameters were taken from the recently published force field parameters for these cosolvents.[2,4] Simulations for RNase T1 and hen egg white lysozyme were performed using CHARMM[95] and simulations for α-chymotrypsinogen were performed using NAMD[101] software package. Details of the simulations are shown in Table 2.1. Simulations performed in CHARMM used a truncated octahedral box extending a minimum of 12 Å from the protein surface while simulations performed using NAMD used a cubic box of 75 Å extending a minimum of 12 Å from the protein surface. All simulations were performed at 298 K and 1 atm. The pH of each simulation was set up by changing the protonation states for each ionizable chain to the dominant form expected for each amino acid at the pH of interest. Simulations involving RNase T1 and lysozyme were performed at pH 7.0, and those involving α-Cgn A were performed at pH 4.75. The initial placement of water and cosolvent molecules were ran-

dom. The systems were equilibriated for 1 ns as it was reported to be sufficient equilibriation time.[1,93] Equilibriation time of 1 ns is also sufficient to remove any bias due to the initial random placement of additive molecules. The diffusivities for the set of additives studied in this paper lie in the range 1.31×10^{-9} (urea)- 5.90×10^{-10} (ArgHCl) m^2/s. Therefore, 1 ns is sufficient for the cosolvents to drift significantly from their original position. The configurations are then saved at an interval of 0.1 ps to compute properties of interest.

Simulations involving RNase T1 were repeated with harmonic restraints on the protein backbone atoms. The restraints were applied using the harmonic atom constraint in CHARMM. The program calculates the deviation of the backbone atoms with respect to the reference structure of the protein and adds an additional energy term, $\Sigma_i k_i m_i (r_i - r_{0i})^2$ for all atoms that are to be restrained. k_i is the force constant, m_i is the mass of the atom i, r_i is the position of the atom i, and r_{0i} is the position of the atom in reference structure. The reference structure is chosen to be the minimized crystal structure of RNase T1. The value of force constant used is 5. Simulation of RNase T1 and urea is also repeated with restraints on all the atoms in RNase T1, using the same procedure.

2.1.2 Calculation of Preferential Interaction Parameters

The preferential interaction parameter is calculated in the NTP ensemble. MD simulations use the description of preferential interaction parameter as a measure of how the cosolvent concentration changes when protein is added to the solution in order to keep the chemical potential of cosolvent constant. This preferential interaction coefficient cannot be measured directly using experiments. Therefore, approximate definitions of preferential interaction coefficient are used. Vapor Pressure Osmometry (VPO) measures the change in cosolvent molality when protein molality is changed at constant T,P and water chemical potential(μ_1). Dialysis/Densimetry measures the change in cosolvent molality when protein molality is changed at constant T, cosolvent(μ_3) and water chemical potential (μ_1). Schneider and Trout[3] have reported the exact relationship between the preferential binding coefficients measured using VPO and dialysis/densimetry. The quantitative differences between the

experimental preferential binding coefficients measured using the two techniques for a set of five commonly used cosolvents and three proteins was found to be less than the error in the individual measurements.

The method of calculating preferential interaction parameters, based on a statistical mechanical method applied to all-atom model with no adjustable parameter, was developed by Baynes and Trout.[1] This approach is used to calculate number of bound molecules to the protein without *a priori* information about any binding sites on the protein and yields a detailed description of interactions between proteins and cosolvents. The variation of concentration as a function of distance from protein surface is used to calculate Γ_{23} as a function of distance from protein until it approaches a constant value. The MD run is saved at periodic time intervals (0.1 ps) and these saved frames are used to find Γ_{23}. The following points elucidate the algorithm used in calculation of the preferential interaction parameter:

1. Every molecule (water and cosolvent) is treated as a point at its center of mass. Distance of the molecule from the surface of all protein atoms is calculated. Every atom is considered as a sphere with radius equal to its Van der Waal radius.

2. The minimum of all such distances is identified and is put into bins of size 0.1 Å. We found that an accuracy of 0.1 Å is required to capture details in the variation of Γ_{23} with distance. From the definition of instantaneous $\Gamma_{23}(t)$, Γ_{23} as a function of distance r from protein can be computed using,

$$\Gamma_{23}(r,t) = n_3^{II}(r,t) - n_1^{II}(r,t)\left(\frac{n_3^I(t)}{n_1^I(t)}\right). \tag{2.2}$$

The ratio n_3^I/n_1^I is the ratio of bulk cosolvent to solvent density. The distance r^* at which the number density ratio goes to a constant value is identified as bulk. The above formula can be modified to incorporate the effect of movement of water and cosolvent molecules in and out of the local domain of protein. Γ_{23} is estimated at each value of r assuming that it is the end of the local domain. The $\Gamma_{23}(r,t)$ for distances

r greater than $r*$ will be constant. With this modification, the expression is[102]

$$\Gamma_{23}(r,t) = n_3(r,t) - n_1(r,t)\left(\frac{n_3 - n_3(r,t)}{n_1 - n_1(r,t)}\right), \tag{2.3}$$

where n_3 is the total number of cosolvent molecules, and n_1 is the total number of water molecules. The motivation to modify Equation 2.2 comes from the fact that Equation 3 provides all of the information necessary to estimate the preferential interaction coefficient. The apparent Γ_{23} (Equation 2.3) as a function of distance r from protein provides value of the preferential binding coefficient for all possible values of the extent of the local domain. This gives the value of the preferential binding coefficient and how far we need to go away from the protein surface to get a reasonably constant value. Equation 2.3 is used to determine the value of r^*. The final value of Γ_{23} obtained from Equation 2.3 will be same as the value of Γ_{23} obtained from Equation 2.2 (n_3^I/n_1^I estimated using r^*). Once r^* is determined, either of the equations can be used to obtain the final prediction. $\Gamma_{23}(r^*,t)$ is the constant value reached beyond r^*, and it is the instantaneous value of the preferential interaction coefficient, $\Gamma_{23}(t)$.

3. The preferential interaction coefficient for the entire trajectory is defined as the time average of all these instantaneous values.

$$\Gamma_{23} = \frac{\sum_{i=0}^{\tau}\Gamma_{23}(t_i)}{\tau}, \tag{2.4}$$

where τ is the time period of the entire run, and $\Gamma_{23}(t_i)$ stands for the value of the preferential interaction coefficient at time t_i.

Another method to estimate Γ_{23}, which requires estimation of the radial distribution function $g(r)$ of the cosolvent and water around the protein surface, can be used. To estimate $g(r) = \rho(r)/\rho(\infty)$, a number density function is obtained as a function of distance r from protein surface: $\rho_1(r)$ for water and $\rho_3(r)$ for the cosolvent. The preferential interaction coefficient

is related to the radial distribution functions of the cosolvent (g_3) and water (g_1):

$$\Gamma_{23}(t) = n_2^{II}(t) - n_2^{II}(t)\left(\frac{n_3^I(t)}{n_1^I(t)}\right) \tag{2.5}$$

$$= \rho_3(\infty)\int g_3(r)dV - \left(\frac{\rho_3(\infty)}{\rho_1(\infty)}\right)\rho_1(\infty)\int g_1(r)dV \tag{2.6}$$

$$= \rho_3(\infty)\int (g_3(r) - g_1(r))dV. \tag{2.7}$$

The integral extends from $r = 0$ to ∞, but this integral is evaluated from protein surface to the box boundary ($\sim 10 - 15$ Å). It should be noted that the expression inside the integral is equal to zero in the bulk domain. r^* is the distance from the surface of protein at which bulk domain begins where $g_3(r)$ and $g_1(r)$ are both equal to 1. The box size is chosen to be greater than r^*. The two methods discussed here give the same estimate of Γ_{23}. However, the second method involves additional steps for estimating the radial distribution function and subsequent intergation of Equation 2.7 in the local domain of protein.

For a 1:1 electrolyte cosolvent in a solution, the excess number of solute molecules around the protein involves both the distribution of cations and anions. Therefore, preferential interaction coefficient calculations would require an estimate of Γ for both the cation and the anion. The preferential interaction coefficient for the cosolvent would then be given by[81]

$$\Gamma_{23} = \left(\Gamma_{2,-3} + \Gamma_{2,+3} - |Z_3|\right)/2, \tag{2.8}$$

where $\Gamma_{2,-3}$ is the preferential interaction coefficient for the anion, $\Gamma_{2,+3}$ is the preferential interaction coefficient for the cation, and $|Z_3|$ is the net charge of the protein. If the counter ions added to balance the protein charge are the same as the cation or anion then the charge on the protein is subtracted from Γ_{23} as the first Z_3 ions accumulate on the protein surface to satisfy the charge balance and do not contribute to the preferential interaction coefficient.

2.1.3 Estimation of statistical error

Computer simulations are subject to statistical errors due to finite sampling. For our long (but finite) simulations we compute averages and variances assuming Gaussian statistics. For simulation data that contains a total of τ_{run} time steps with statistically independent observations of $A(\tau)$, the variance in the mean is given by:[103]

$$\sigma^2(\langle A\rangle_{run}) = \sigma^2(A)/\tau_{run} \tag{2.9}$$

However, the data points in our simulation are not independent. Therefore, the entire run is broken down into blocks of length τ_b, such that there are n_b such intervals. Thus, the $\langle A\rangle_b$'s for n_b blocks are used to calculate variance $\sigma^2(\langle A\rangle_b)$. As the block length becomes large enough to be statistically uncorrelated, the variance in means of block averages is inversely proportional to the block length. This constant of proportionality is required in order to evaluate the statistical error in the run. This constant is defined as:[104]

$$s = \lim_{\tau_b\to\infty} p(\tau_b) = \lim_{\tau_b\to\infty} \frac{\tau_b\sigma^2(\langle A_b\rangle)}{\sigma^2(A)} \tag{2.10}$$

The quantity s is called the statistical inefficiency, and any technique that reduces s will allow the calculation of more accurate simulation averages. The statistical inefficiency s reaches a plateau value as the block length τ_b is increased. This asymptotic s value along with τ_b, and $\sigma^2(A)$ are substituted in Equation 2.10 to calculate $\sigma(\langle A_b\rangle)$.

2.2 Results and Discussion

2.2.1 Extent of the local domain

In order to decide the value of r* which separates the local and bulk domain, a plot of Γ_{23} as a function of distance r from the protein surface is used. This quantity Γ_{23} is called the apparent preferential interaction coefficient by Baynes and Trout[1] and has been plotted in

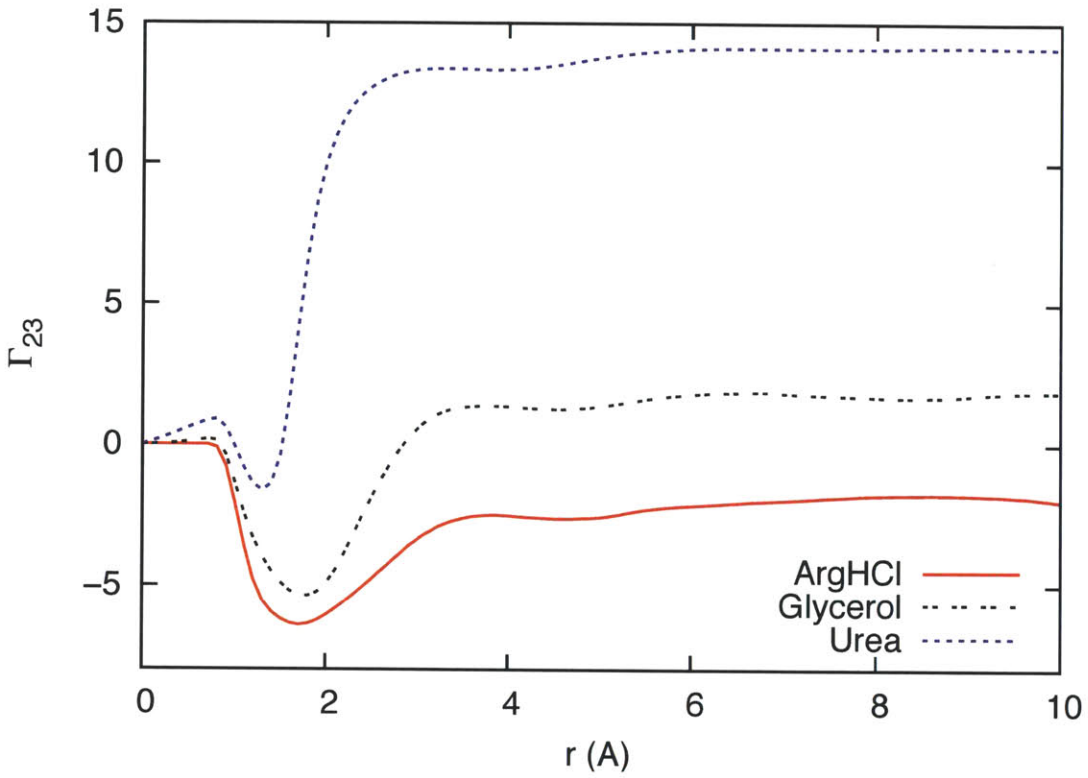

Figure 2-1: Γ_{23} **plotted as a function of distance from the protein surface for the systems comprising protein RNase T1 in aqueous argHCl, glycerol and urea solutions.The corresponding preferential interaction coefficients are reported in Table 2.2. The simulation results shown in the figure are for unrestrained protein simulations.**

Figure 2-1 for RNase T1 with three different additives. The apparent preferential interaction coefficient at any given distance r from the protein gives information about the excess number of cosolvent molecules inside the region defined by r. Water molecules being smaller than cosolvents have a higher presence in the vicinity of protein, which is apparent from the negative dip in preferential interaction coefficient between 1-2 Å. At farther distances from the protein, the larger cosolvent molecules are no longer excluded and the preferential interaction parameter increases and subsequently attains a constant value. In all the three cases, the values remain constant after 6 Å, and this distance can be taken as the location of boundary separating local and bulk domain for these cosolvents. For other simulations involving proteins Lysozyme and α-Cgn A, the value of r^* is found to be also around 6 Å. The same

information can also be obtained from the plot of bulk density $((n_3 - n_3(r))/(n_1 - n_1(r)))$ as a function of distance r from the protein as shown in the Figure 2-2. It can be seen from the figure that bulk density values are constant beyond 6 Å. Therefore, 6 Å can be chosen as r^*. In the set of simulations reported in this paper, a small divergence of apparent Γ_{23} and bulk density values beyond 6 Å is observed. These are caused by small fluctuations in the bulk density and apparent Γ_{23} in the region beyond r^*. These fluctuations could change the final predicted value. The r^* value reported here is applicable only to the systems studied in this paper and depends on the protein and cosolvents.

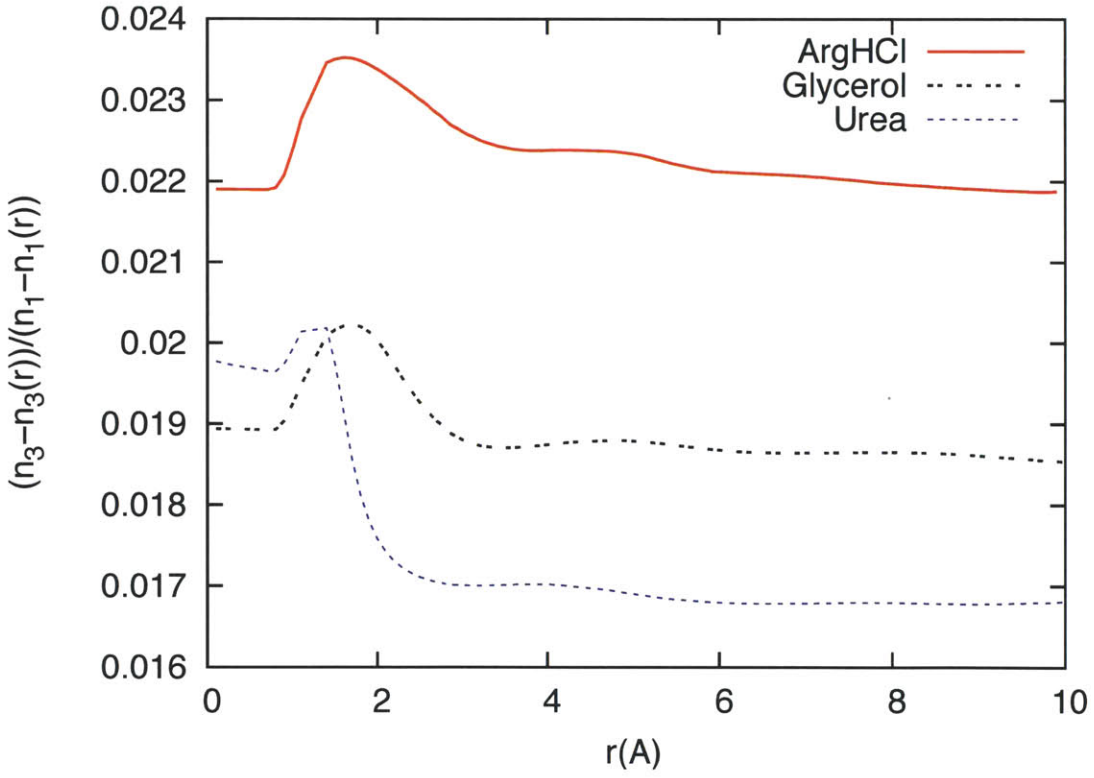

Figure 2-2: **Ratio $(n_3\text{-}n_3(r))/(n_1\text{-}n_1(r))$ as function of distance r from the protein surface for the systems comprising RNase T1 in aqueous argHCl, glycerol and urea solutions.**

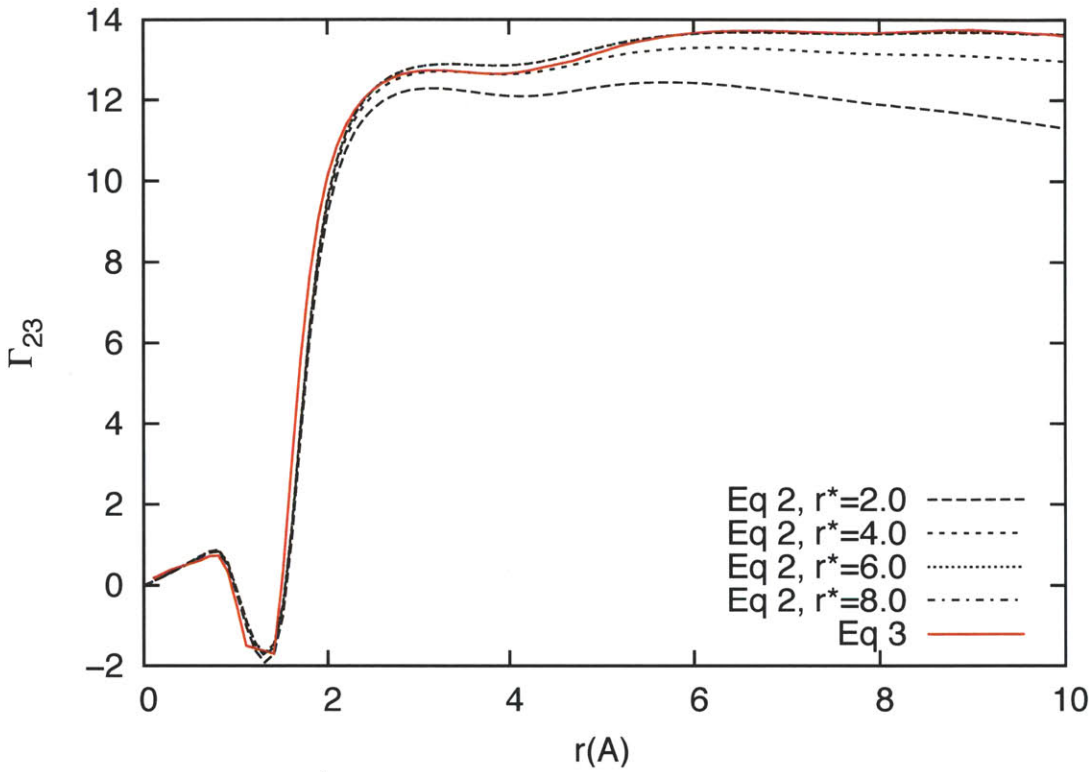

Figure 2-3: **Apparent Γ_{23} obtained using equations 2.2 and 2.3 for the system RNase T1 in aqueous urea solution.**

2.2.2 Preferential Interaction Coefficients

Γ_{23} values for the system RNase T1 and urea computed using equations 2.2 (for different r* values) and 2.3 are shown in Figure 2-3. The Γ_{23} values obtained using equation 2.2 for r* values greater than 6 Å are equal to the Γ_{23} values obtained using equation 2.3. Computed values of Γ_{23} for RNase T1 and cosolvents urea, glycerol and arginine hydrochloride along with their statistical error are shown in Table 2.2. The confidence intervals on Γ_{23} are an estimate of the statistical error resulting from the use of a finite trajectory. For comparison with the computed values, the experimental values spanning the molality of interest are interpolated to the bulk molality calculated from the simulations. Baynes and Trout[1] estimated the Γ_{23} values for RNase T1 and cosolvents urea and glycerol based on only 2 ns of dynamics. The calculated average values for the first two nanoseconds and for the entire run obtained from MD simulations are compared with the experimental values obtained

from literature in Table 2.2. Experimental data for RNase T1 is available for only one of the cosolvents studied in this paper, urea. The computed value of Γ_{23} compares favorably with the experimental data for the first 2 ns. However, for extended runs the computed values differ from the values obtained from the first 2 ns of dynamics. It is expected that the Γ_{23} values for extended simulations would give a better match with experimental data, but the extended simulation value for urea is 13.8, which is much higher than the experimental value of 6.4. The Γ_{23} values for extended simulations of glycerol and ArgHCl also change significantly from the values computed over 2 ns. Furthermore, glycerol is a stabilizer which is expected to have a negative preferential interaction coefficient, but the predicted value from simulations is 1.9.

Table 2.2: Γ_{23} **for RNase T1 computed from MD simulations and available experimental values extrapolated to the concentration of interest.**

Cosolvent	Simulation time (ns)	Γ_{23} (First 2 ns)	Γ_{23} (Entire run)	Experimental Γ_{23} (Densimetry)[83]	m_{bulk}
ArgHCl	15	-5.2±2.0	-2.1±0.9		1.20
glycerol	15	-2.6±2.3	1.9±1.0		1.04
Urea	19	5.2±1.6	13.8±0.9	6.40	0.94

Table 2.3: Γ_{23} **for α-Chymotrypsinogen A computed from MD simulations and available experimental values extrapolated to the concentration of interest.**

Cosolvent	Simulation time (ns)	Γ_{23}	Experimental Γ_{23} (VPO)[3]	m_{bulk}
ArgHCl	15	-8.0±2.5	-8.7±2.9	1.05
GdnHCl	15	2.2±1.5	3.7±0.9	0.98
Glucose	15	-3.4±1.9	-4.6±0.4	1.03
glycerol	15	1.7±1.6	-14.4±0.7	1.02
Urea	15	10.4±1.7	8.8±1.0	0.96

The Γ_{23} values for α-Cgn A and Lysozyme in aqueous solutions of cosolvents argHCl, gdnHCl, glucose, glycerol and urea are shown in Table 2.3 and 2.4 respectively. Schneider and Trout[3] have performed experiments with proteins Bovine Serum Albumin (BSA), α-Cgn A & Lysozyme, and a variety of cosolvents employing Vapor Pressure Osmometry (VPO) to measure the Γ_{23} values. Experimental values from the literature using Dialysis/Densimetry

Table 2.4: Γ_{23} **for Lysozyme computed from MD simulations and available experimental values extrapolated to the concentration of interest.**

Cosolvent	Simulation time (ns)	Γ_{23}	Experimental Γ_{23} (Densimetry)[5,64]	Experimental Γ_{23} (VPO)[3]	m_{bulk}
ArgHCl	15	-3.7±1.5	-3.9	-4.2±1.6	1.05
GdnHCl	15	3.7±0.8		3.4±0.9	1.02
Glucose	15	-2.6±1.1		-2.7±0.3	1.03
glycerol	10	-2.0±0.9		-6.3±0.3	1.17
Urea	15	8.8±1.2	6.3±1.0	3.1±0.9	0.99

to measure Γ_{23} were available for Lysozyme with the cosolvents argHCl & urea. Our computed values of Γ_{23} agree quite favorably with the experimental data except for the cosolvent glycerol. The predicted Γ_{23} values for glycerol are much higher than the experimental values for both α-Cgn A and lysozyme. For lysozyme-urea, the experimental values based on VPO and Dialysis & Densimetry are different from each other. The computed value of 8.8 compares well with the Dialysis/Densimetry value of 6.3, but it is much higher than the value of 3.1 measured using VPO. Computed values for the cosolvent argHCl, with lysozyme match well both the VPO and densimetric measurements at 1 molal total argHCl comcentration. However, Schneider and Trout[3] have shown that the Γ_{23} values for the lysozyme-argHCl reported by Kita et. al.[64] differ significantly from the VPO and the Dialysis/Densimetry measurements performed by them at lower concentrations. For the cosolvents gdnHCl and glucose, the computed Γ_{23} values match well with the experimental VPO values. It is important to establish the validity of the methodology due to the difference in the predicted and experimental values observed for some protein-cosolvent studied here. Therefore, these extended runs are investigated in to understand why there are deviations from the observed experimental values.

2.2.3 RMSD Analysis

Root mean square deviation (RMSD) represents the average deviation of protein structure from a reference structure, the reference structure in our case was the protein structure obtained after the minimization of the system. It was found that RMSD of RNase T1 is

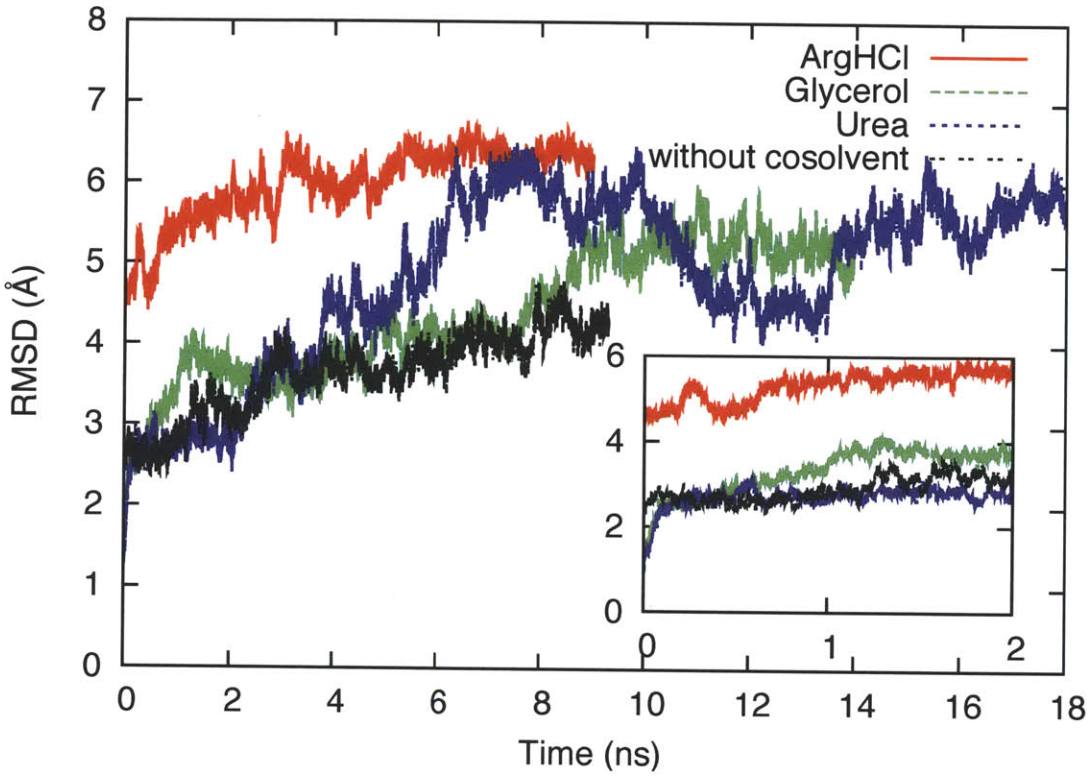

Figure 2-4: **RMSD variation with time from simulations of protein RNase T1 in 1 m aqueous solution of cosolvents argHCl(Red), glycerol(Green) and urea(Blue). The black curve shows the time variation of RMSD from simulation of RNase T1 in water box without any cosolvent. Inset shows the same plot for the first 2 ns. The simulations results shown above are for unrestrained protein simulations.**

related to the large variation in Γ_{23}. Note that, during the simulations, the protein was not fixed in any way and was free to rotate or translate through the box. Therefore for purposes of calculating the RMSD, the protein is re-centered and reoriented so that the RMSD captures only structural changes like partial unfolding and not changes due translation and rotation. A plot of RMSD versus time is shown in Figure 2-4. For the first 2 ns of dynamics, the RMSD values are stable, and therefore, Baynes and Trout [1] did not observe such behavior. For the longer trajectory, the RMSD values for RNase T1 even without any cosolvent reached a peak value of 4.0 Å and RNase T1 has a very high RMSD for all three additives, the highest being 6.6 Å for arginine, 6.2 Å for urea and 5.8 Å for glycerol. Thus, the presence of cosolvent seems to have an effect on the protein conformation resulting in unfolding to some extent.

Table 2.5: **RMSD of proteins Lysozyme and α-Chymotrypsinogen A from the minimized structure computed from simulations of the respective proteins in aqueous cosolvent solutions.**

Protein	cosolvent	RMSD (Å)
Lysozyme	ArgHCl	1.7±0.2
Lysozyme	GdnHCl	1.8±0.2
Lysozyme	Glucose	1.7±0.2
Lysozyme	glycerol	3.0±0.3
Lysozyme	Urea	2.8±0.2
α-Cgn A	ArgHCl	2.1±0.2
α-Cgn A	GdnHCl	2.0±0.2
α-Cgn A	Glucose	2.3±0.3
α-Cgn A	glycerol	2.1±0.3
α-Cgn A	Urea	2.2±0.4

Experimental data on the unfolding of RNase T1 in aqueous urea solutions show negligible unfolding up to 3 molar urea, and the time scales for our simulations are very small for such transformation to take place.[105] The average RMSD value for RNase T1 in a water box without any cosolvent is 3.5 Å which is comparable to fluctuations observed in the presence of cosolvents. Structural fluctuations are high even for highly excluded cosolvents which tend to stabilize the protein structure. Table 2.5 shows the RMSD of Lysozyme and α-Cgn A in the presence of cosolvents urea, gdnHCl, argHCl, glucose and glycerol at 1 molal total concentration. These values are in the range 1-3 Å, and the standard deviations are less than 0.4 Å, which indicates that after initial adjustment in the proteins solvent accessible area due to presence of cosolvents, only minor structural fluctuations take place during simulation. This indicates that these high RMSD values observed for RNase T1 are not related to cosolvents. The fluctuations in protein conformations are correlated with the fluctuations observed in the preferential interaction parameter. Due to structure fluctuations, the protein is not in the native state. This leads to sampling of configurations which would not be sampled in a real experiment. The uncertainties in the Γ_{23} are low but the wrong set of configurations have been sampled. Therefore, the experimental Γ_{23} values do not match the theoretical prediction. If experimental data is not available, Figure 2-4 can be used to decide whether restricted simulations are required or not. For example, experimental data

is not available for RNaseT1-ArgHCl system but large unphysical fluctuations observed in this case would significantly affect the estimated Γ_{23}. While evaluating Γ_{23}, it should be ensured that the protein remains in the native state. Therefore, external restraints need to be applied to keep the protein in a stable state during simulations.

2.2.4 Restrained simulations

In order to study the effects of protein dynamics on the preferential interaction parameters, the simulations were repeated for RNase T1 and cosolvents urea, glycerol and argHCl, but with the protein backbone atoms constrained to their minimized structure positions as described in the simulation setup section. Another system was studied comprising of RNase T1 and cosolvent urea with the entire protein resstrained to its minimized structure. The simulation time for the system with the entire protein restrained in the aqueous urea solution is restricted to 7 nanoseconds, as the system is found to be equilibrated with respect to Γ_{23}. The cumulative average reaches a constant value of 8 within 1 nanosecond and is within ± 0.5 of that value for further times. Similar behavior is observed for the backbone restrained simulation as well. Comparison between the RMSD observed during the restrained and unrestrained simulations is shown in Table 2.6. Restrained simulations show a small average deviation from the minimized structure as compared to the unrestrained simulations as deviation from minimized structure is only due to fluctuations of side chain residues. Table 2.7 compares the Γ_{23} values from restrained and unrestrained simulations, showing significant differences. For the cosolvent urea, both the backbone and the entire protein restrained simulation values compare favorably with the experimental data.

Thus, constraining the protein greatly reduces the fluctuations in the measured preferential interaction parameter and moreover, leads to a values much closer to the experimental values. In which cases should artificial restraints be applied on proteins? For RNase T1, experimental data on unfolding and conformational stability does not provide any reason for applying restraints. At 1 molal urea concentration, negligible protein unfolding is observed experimentally.[105] $\Delta G(H_2O)$, which is obtained by extrapolating the free energy of unfold-

ing, ΔG, as a function of urea concentration to zero concentration is used as a measure of stability of native state. RNase T1 and lysozyme have $\Delta G(H_2O)$ values of 8.9 Kcal/mol and 8.8 Kcal/mol respectively.[105] Lysozyme shows stable RMSD values during simulation whereas RNase T1 shows large structure fluctuations. Similarly, α-Cgn A has a $\Delta G(H_2O)$ value of 14.2 kcal/mol (higher than RNase T1) but it still shows stable RMSD values. The only reason which can explain high RMSD values for RNase T1 during simulation is that the force field is not able to replicate stable protein structure in solution. Thus, a rule of thumb could be that restraints should be applied when the RMSD of the unrestrained simulation is greater than 3-4 Å.

Table 2.6: **Root Mean Square deviation (RMSD) of protein RNase T1 from the minimized structure computed from MD simulations in which protein is restrained. The last column lists the RMSD values from unrestrained simulations.**

Restraint	Cosolvent	RMSD (Å) Restrained	RMSD (Å) Unrestrained
Backbone	ArgHCl	0.5±0.1	6.0±0.5
Backbone	glycerol	1.1±0.1	4.4±0.8
Protein	Urea	0.4±0.1	4.9±1.1
Backbone	Urea	0.6±0.1	4.9±1.1

Table 2.7: **Comparison between Γ_{23} computed from restrained and non-restrained MD simulations, and available experimental values extrapolated to the concentration of interest.**

Restraint	Cosolvent	Simulation time (ns)	Γ_{23} Restrained	Γ_{23} Unrestrained	Γ_{23} Experimental[83]
Backbone	ArgHCl	10	-4.4±1.3	-2.1±0.9	
Backbone	glycerol	10	-0.1±1.1	1.9±1.0	
Protein	Urea	7	8.0±0.8	13.8±0.9	6.4
Backbone	Urea	10	8.1±1.0	13.8±0.9	6.4

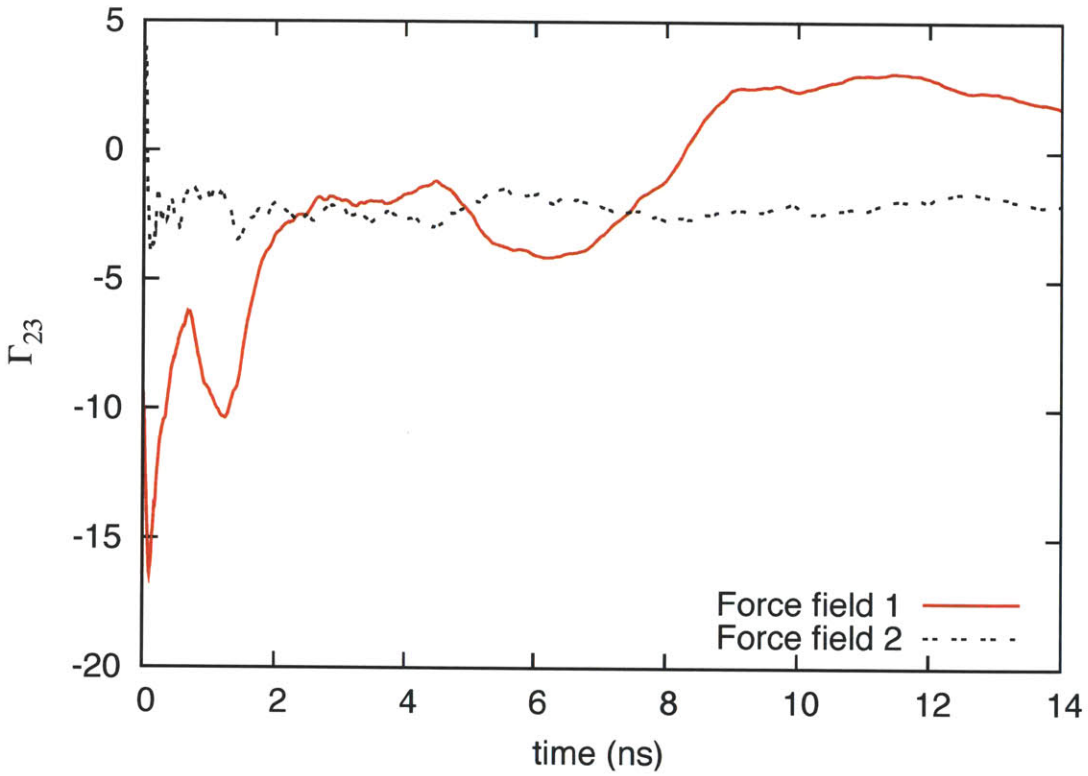

Figure 2-5: **Comparison between simulated preferential interaction parameter Γ_{23} for glycerol and α-Cgn A at 300 K and pH 4.75 as a function of time obtained using glycerol force field parameters taken from Baynes & Trout[1](Force field 1), and Kamath et. al.[2](Force field 2). The experimental value is -14.39.[3]**

2.2.5 Effect of force field parameters

2.2.5.1 Glycerol

The force field used in the simulation involving cosolvent glycerol was constructed using the standard CHARMM geometries and partial charges for the atoms in a -CHOH unit as done previously.[1] The computed Γ_{23} values for glycerol are much higher than the experimental values. Therefore, the current force field underpredicts the extent of exclusion of glycerol from the protein surface. Recently, Kamath et. al.[2] have reported new force field parameters for glycerol. The density computed using the new force field matches well with the experimental density data in the range 0.5-5.0 molal glycerol concentration. In order to test the new force field parameters, a simulation for the system α-Cgn A-glycerol is performed with new force

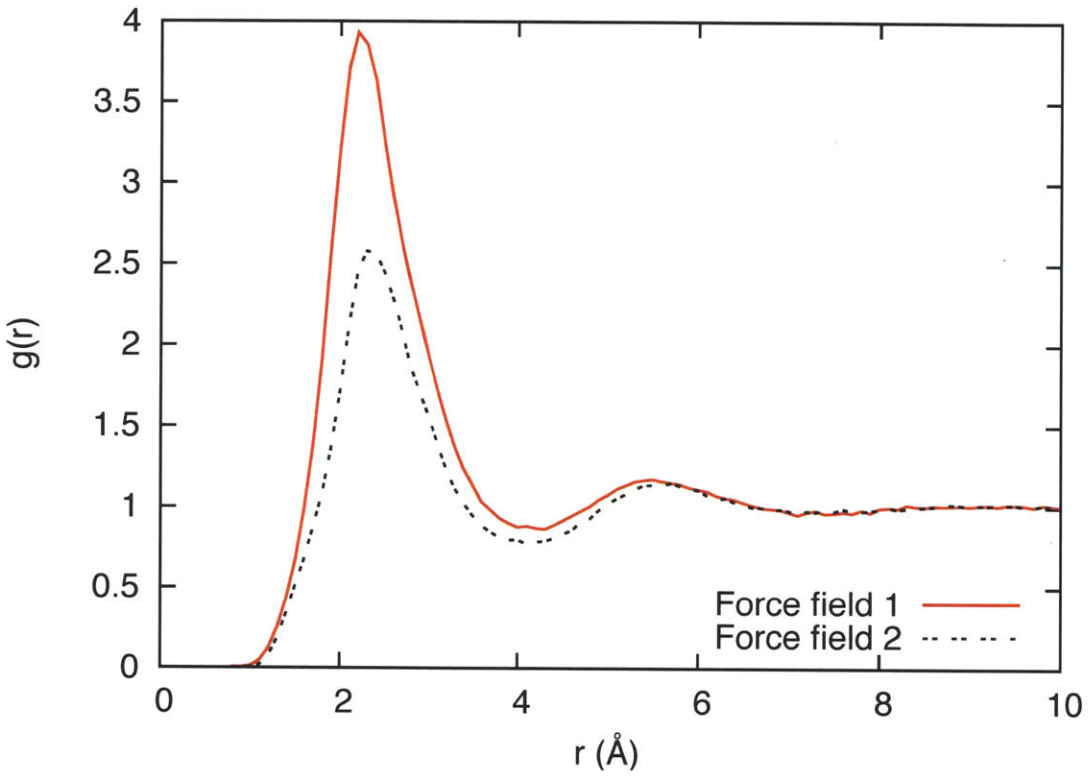

Figure 2-6: **Comparison between RDF's of glycerol as a function of the closest distance(r) to any protein (α-Cgn A) atom obtained using glycerol force field parameters are taken from Baynes & Trout[1] (Force field 1), and Kamath et. al.[2] (Force field 2)**

field parameters as this system shows the maximum deviation from the experimental value. A plot of cumulative Γ_{23} as a function of simulation time for the two force field parameters is shown in Figure 2-5. The Γ_{23} value obtained using the new force field parameters is -2.0 as compared to 1.7 computed using the old force field. However, both these values are much higher than the experimental value of -14.4. Comparison between the radial distribution function of glycerol around the protein molecule for both the force fields is shown in Figure 2-6. The RDF's are calculated from the surface of the protein. Therefore, the peaks in the RDF's are located at a shorter distance as compared to RDF's calculated from the center of the closest protein atom. The new force field from Kamath et. al.[2] shows a smaller first peak signifying more exclusion of glycerol molecules from the local domain of protein. Exclusion from the protein surface requires glycerol molecules to have more

favorable interactions with water versus the protein surface. In order to test whether these force fields replicate the interaction with water molecules correctly, their interaction with water molecules is studied in detail. To obtain partial charges for a molecule in CHARMM, QM minimum interaction energies and geometries between molecule and water along with dipole moments are used as a target function. This approach is used for inter-molecular force field parameterization involving electrostatic interactions.[106]All quantum calculations were performed using Gaussian03.[107] The QM level of theory used is the standard HF/6-31G* and MP2/6-31G*. No corrections for basis-set superposition error(BSSE) were made for the HF calculations in accordance with the standard charmm force field development procedure. Comparison between the interaction energy of the -OH groups in glycerol with water, the minimum energy distance and the dipole moment of glycerol measured for both the force fields and QM data is shown in Table 2.8. The HF/6-31G* interaction energies match well with the empirical interaction energies computed from the force fields, if a scaling factor of 1.16 is applied to overcome the underestimation of the interaction energy using HF/6-31G* level of theory.[106] The minimum energy distance is expected to be ~0.2 Å shorter than the QM value. Hartree-Fock model overestimates the minimum interaction distances due to the absence of the dispersion contribution and neglect of many body effects.[106] For the middle OH group, the minimum energy distance is much higher than the QM value of 2.83 Å. Empirical dipole moments are expected to be ~10 % larger than the QM value. For both the force fields, the match is poor. Furthermore, if we move to a higher level of theory(MP2/6-31G*) and account for the BSSE, the match between empirical values and QM data is even poorer. BSSE is calculated using the counterpoise method.[108] Therefore, these force fields do not represent the interaction of glycerol with water correctly, the most likely reason why predicted Γ_{23} values do not match well with the experimental values.

2.2.5.2 Urea

Force field parameters used in our simulations are based on the urea model by Duffy et. al.[94] Cabballero-Herrera and Nilsson[4] have proposed a new set of urea parameters which

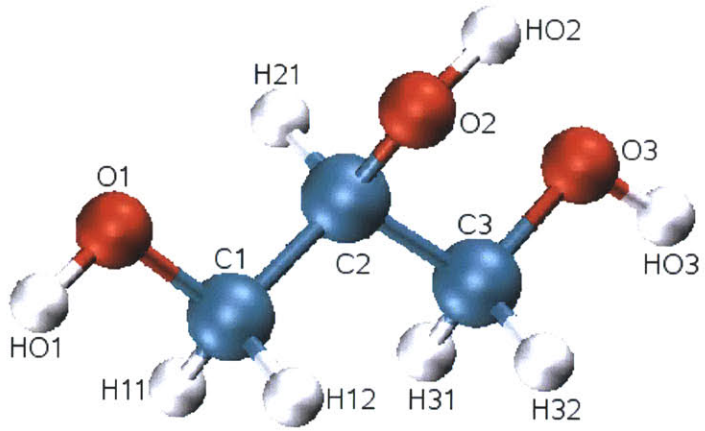

Figure 2-7: **A glycerol molecule with atoms labeled according to the atom names used in Table 2.8.**

Table 2.8: **Comparison of energies, minimum energy distances for glycerol-water interaction and dipole moment obtained using two glycerol force fields with QM data. The units of interaction energy, dipole moment and minimum energy distance are kcal/mol, Debye and Angstrom respectively. HF energies are reported without BSSE correction. Glycerol force field parameters are taken from Baynes & Trout[1](Force field 1), and Kamath et. al.[2](Force field 2).**

	HF/6-31G*	MP2/6-31G*	Force field 1	Force field 2
Interaction energy HO1-OHH	-6.11	-7.80	-7.51	-6.80
Interaction energy HO2-OHH	-0.13	-3.03	0.15	0.01
Dipole moment	3.56	3.75	4.47	2.69
Min. energy distance HO1-OHH	1.98	1.91	1.81	1.83
Min. energy distance HO2-OHH	2.83	2.29	8.95	8.00

takes into account urea dimer formation, while being consistent with the protein parameters used with the CHARMM[95] force field and the TIP3P[97] water model. A simulation involving lysozyme-urea system is repeated with the new force field parameters for urea. The plot of cumulative Γ_{23} as a function of simulation time for the two force field parameters is shown in Figure 2-8. Final Γ_{23} values for the new force field is 6.6 which is lower than the Γ_{23} value of 8.8 computed using the old force field. Furthermore, the value compares quite favorably with the experimental value of 6.3.[5] Lin and Timasheff[5] have not reported error bars on the experimental data, but in other papers dealing with the estimation of Γ_{23} experimentally using Dialysis/Densimetry by Timasheff and coworkers[5,64,82] in which

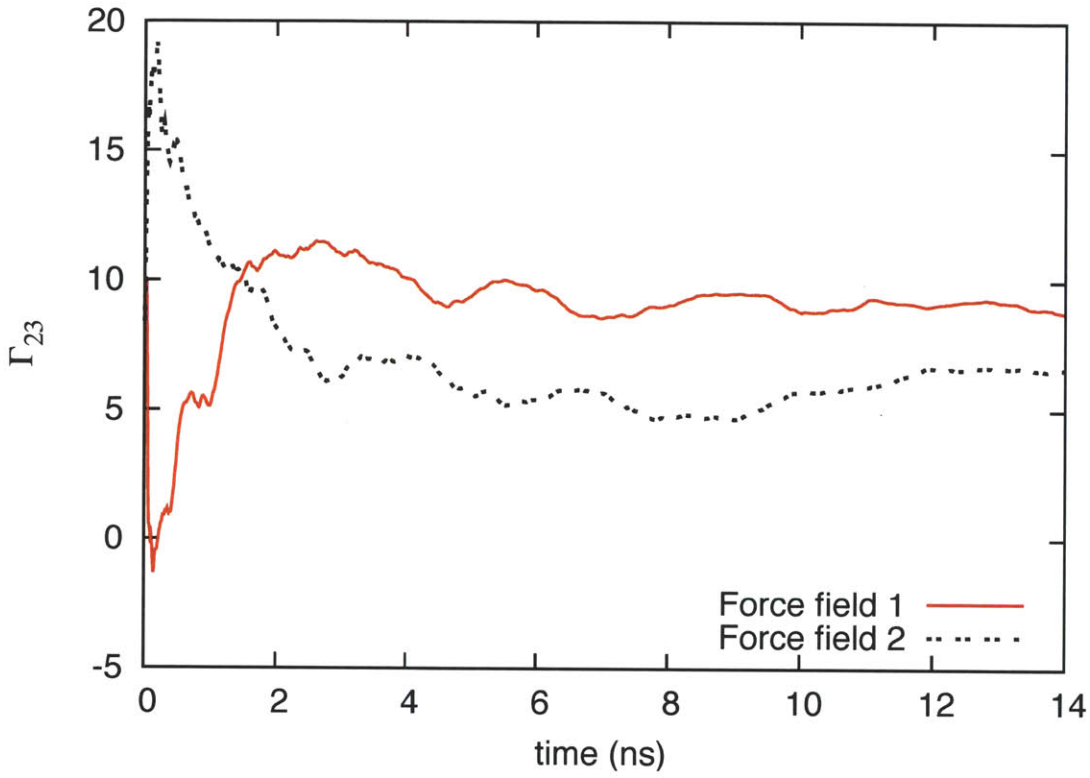

Figure 2-8: **Comparison between simulated preferential interaction parameter Γ_{23} for urea and lysozyme at 300 K and pH 7.0 as a function of time obtained using urea force field parameters taken from Baynes & Trout[1](Force field 1), and Caballero-Herrera & Nilsson[4](Force field 2).The experimental value reported using dialysis and densimetry is 6.3.[5]**

they have reported error bars, the error is in the range 2-4. Therefore, both the force fields give a good match for Γ_{23} with experimental data. Analysis of the interaction of urea with water (similar to the analysis performed with the glycerol models) is not performed for the urea models as the partial charges for both urea models are not estimated using the standard CHARMM[95] empirical force field parameter estimation procedure. Caballero-Herrera and Nilsson[4] have estimated partial charges based on the interaction of urea in a complex with several water molecules, whereas Duffy et. al.[94] optimized partial charges based on interaction of water molecule with each individual atom in urea. However, the important issue here is the sensitivity of Γ_{23} to changes in the force field parameters. The urea models used here match their respective experimental target data, like free energy of

hydration[94] of urea, QM interaction energies[4] with water etc. but due to the differences in the chosen target experimental data set, the force fields give different Γ_{23} values. The Γ_{23} value changes from 8.8 to 6.6, which is approximately a 25% change, due to differences in the parameter estimation procedures for two models. Therefore, the preferential interaction of cosolvents is very sensitive to force field parameters.

The comparison betwen cosolvent models for glycerol and urea clearly highlights the importance of correct force field parameters in estimating preferential interaction coefficients.

2.2.6 Minimum Simulation time

In order to get accurate estimate of Γ_{23}, sufficient sampling of position-space configurations in time is required. The parameters which affect the time required for good sampling of cosolvent position space are diffusivity of cosolvent, concentration of cosolvent and protein dynamics. The total simulation time must be much larger than the average time between cosolvent-cosolvent contact, as cosolvent dynamics is the most important time scale in this system.[1] The average time between cosolvent contacts can be approximated as

$$t_{contact} = \frac{1}{6D}\left(\frac{3V_{solv}}{4\pi n_X}\right)^{2/3} \approx \frac{1}{12D}\left(\frac{V_{solv}}{n_X}\right)^{2/3} \tag{2.11}$$

where D is the cosolvent diffusivity, V_{solv} is the solvent volume, and n_X is the number of cosolvent molecules. Urea is the smallest cosolvent studied in this paper with a radius of gyration equal to 1.4 Å, and argHCl is the largest with a radius of gyration equal to 3.2 Å. The diffusion coefficients for urea and arginine in a cosolvent-water system at 1 m cosolvent concentration are 1.31 x 10^{-9} m^2/s and 5.90 x 10^{-10} m^2/s respectively.[109,110] The $t_{contact}$ value for urea is about 90 ps and the value for argHCl is 200 ps. Considering that simulation time should be 100 times longer than the cosolvent contact time, simulation time for urea, and argHCl should be 9 ns, and 20 ns respectively. These estimates should only be used as a rough guideline before the start of simulation as the diffusion coefficient varies significantly from the experimental value for most cosolvent models. Diffusion coefficient for the urea model

proposed by Caballo-Herrera and Nilsson[4] is twice as large as the experimental value.[110] Similarly, the glycerol model proposed by Kamath et. al.[2] has an error of 15% in the predicted diffusion coefficient value. Furthermore, diffusion of cosolvents in the local domain is affected by the interaction with the protein.

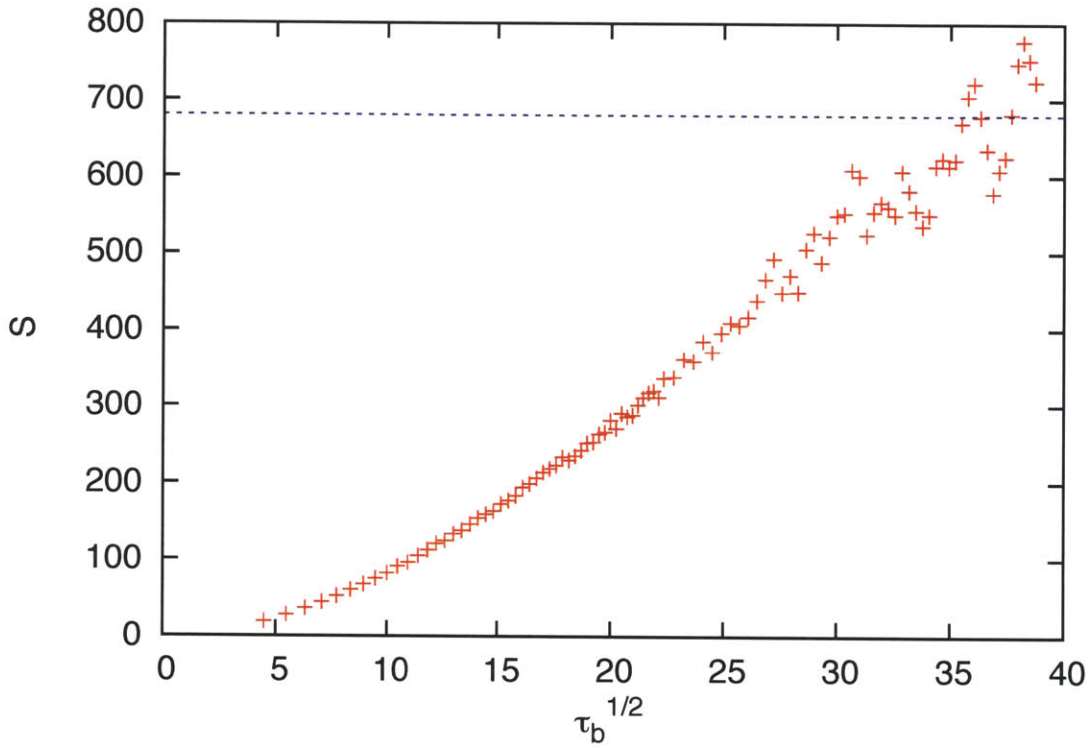

Figure 2-9: **Statistical inefficiency as a function of the square root of the block size for preferential interaction coefficient data from simulations of Lysozyme and cosolvent ArgHCl. The dotted line shows the value where a plateau is approached.**

Statistical error estimation based on section 3 is performed for all the systems investigated in this paper. A representative plot of statistical inefficiency against $\tau_b^{1/2}$ (square root of the block size) for the preferential interaction coefficient in a simulation of lysozyme-ArgHCl system is shown in Figure 2-9. Preferential interaction coefficient values are estimated for every picosecond of simulation. A plateau value of $S = 680$ is reached. This implies that only one configuration in every 680 ps contributes a completely new information to the average value. In other words, this is the measure of autocorrelation in data. ArgHCl has

Table 2.9: **Statistical inefficiency(S) values for preferential interaction coefficient calculated from simulation of cosolvents with proteins Lysozyme, and α-Cgn A**

Protein	cosolvent	S	Protein	cosolvent	S
Lysozyme	ArgHCl	680	α-Cgn A	ArgHCl	590
Lysozyme	GdnHCl	260	α-Cgn A	GdnHCl	220
Lysozyme	Glucose	265	α-Cgn A	Glucose	410
Lysozyme	glycerol	240	α-Cgn A	glycerol[1]	230
Lysozyme	Urea[1]	350		glycerol[2]	330
	Urea[4]	300	α-Cgn A	Urea	340

the highest statistical inefficiency as it samples cosolvent position space slowly due to its large size as compared to other cosolvents studied here. Statistical inefficiency values for all the simulations performed with proteins lysozyme and α-Cgn A are shown in Table 2.9. It can be seen that S values depend not only on the cosolvent but also on the protein, which highlights the importance of both solvent and protein dynamics. The statistical inefficiency values are used to calculate the observed standard deviation of the average Γ_{23} values. The observed standard deviation assuming different total simulation time can be compared to gauge the effect of simulation time on error bar. For the lysozyme-argHCl system with statistical inefficiency of 680, for simulations of length 5, 10, 15, 20, and 30 ns, the standard deviation in Γ_{23} would be 2.47, 1.74, 1.43, 1.23, 1.01 respectively. The standard deviation in experimental data are typically in the range 1-2, therefore simulation time of 15 ns would give good confidence interval. For other cosolvents, the statistical inefficiency values are smaller which would mean even smaller error. Therefore, for the cosolvents studied here, the simulation time should be in the range 10-20 ns to get a confidence interval in the range 1-2.

2.3 Conclusions

A quantitative method based on single trajectory MD simulations, with all atom potential models was tested and validated for proteins RNase T1, lysozyme and α-Cgn A, and for cosolvents covering a range of positive and negative binding behavior. MD simulations were performed over extended time of about 10-20 nanoseconds, which revealed some com-

putational properties of the system not seen earlier.[1] The extended simulations without constraints do not match the experimental data and the simulation results reported earlier for RNase T1. Unphysical protein structure fluctuations were observed which lead to changes in the predicted Γ_{23} values. A constrained protein trajectory successfully overcame this problem yielding good agreement with experimental values for extended runs. Protein dynamics as well as solvent dynamics play an important role in determining Γ_{23}. In order to sample all conformations in a system, longer runs are needed. Cosolvent dynamics defines the most important time scale to be captured in the simulation, and this time scale depends on the diffusivity and thereby on the size of the cosolvents. For the cosolvents studied here and all other cosolvents with similar sizes, the simulation time should be in the range 10-20 ns in order to get a reliable estimate of Γ_{23}. MD simulations used standard CHARMM force field and the results are found to be sensitive to these force field parameters. Preferential interaction is measure of preference of protein surface for cosolvent as compared to the water molecules. Therefore, cosolvent-water interactions should be given more weight when evaluating and optimizing the force field parameters for cosolvents. Force field parameters taken from non-compatible force fields or built from analogy can be relied only after testing their match with extensive experimental target data or quantum mechanical data. Having said all of that, the approach presented here, allows the calculation of accurate values of Γ_{23} from molecular simulations without any ad-hoc parameter fitting.

Chapter 3

Interactions in Aqueous Arginine Solutions

When attempting to explain the arginine mechanism, many researchers start by addressing the effect of the guanidinium moiety. Indeed, the guanidinium functional group must play an important role, because not only is arginine the only amino acid with a guanidinium moiety, no other amino acid exhibits aggregation suppression characteristics quite like arginine, though other amino acids, particularly proline, have been used as stabilizers. However, without a clear understanding at the molecular level, not only of the interaction of arginine with protein molecules, but also of how arginine behaves in solution, the arginine mechanism will remain a mystery. Recently, aqueous urea[32] and Gdm[57] solutions have been investigated to obtain a comprehensive picture of the interaction of water and cosolvent molecules in solution, which has implications for the mechanism by which these cosolvents denature proteins. In view of the importance of these interactions in aqueous cosolvent solutions, it is of interest to examine more closely the structure of and interactions in aqueous arginine solutions, the interactions between the arginine and the protein surface residues, and the implications of these interactions for the mechanism by which arginine inhibits aggregation. There have been no reported simulation studies of aqueous arginine solutions to the best of the authors' knowledge. Here, a molecular dynamics study of this binary system is described. The main

focus is on the structural and energetic (hydrogen bonding) properties of aqueous arginine solutions. The temperature and arginine concentration dependence is covered here for the full experimentally accessible range. Simulations of proteins in aqueous arginine solutions were also performed in order to understand the interaction of arginine with the protein surface.

3.1 Methods

3.1.1 Simulation Setup

Table 3.1: **Setup of the simulation system. The number of argHCl and water molecules in the system, and molal concentrations are listed. The saturation limit at 298 K is at a molal concentration of 2.81.**[3]

Number of Water molecules	Number of ArgHCl molecules	Molality
4000	18	0.25
3778	34	0.50
3556	48	0.75
3333	60	1.00
3111	70	1.25
2963	80	1.50
2825	89	1.75
2694	97	2.00
2568	104	2.25
2444	110	2.50
2343	116	2.75

All simulations were performed using the NAMD[101] package, with the CHARMM22[95] force field. The TIP3P[97] water model was used. The pKa values for the C-terminal, the N-terminal, and the side chain in an arginine molecule are 1.8, 9.0, and 12.5 respectively.[111] The N-terminal and the side chain are protonated, whereas the C-terminal is deprotonated in the pH range 1.8-9.0. This pH range is of interest as proteins are observed to be highly unstable at low and high pH.[112,113] Therefore, force field parameters for arginine were taken from the CHARMM force filed with the protonated N terminal, and the side chain, and

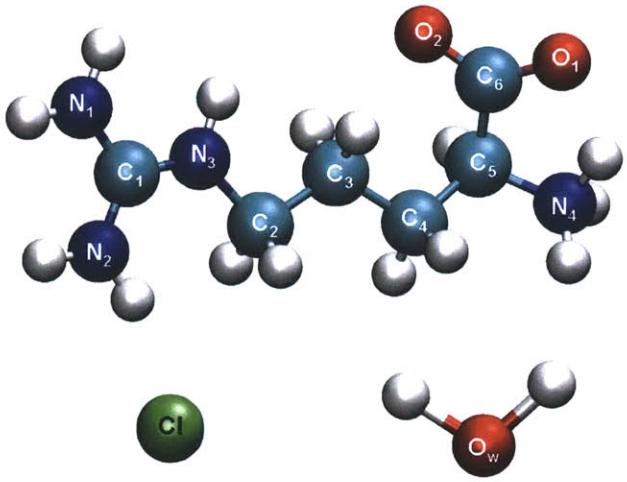

Figure 3-1: **Labeling scheme for the atoms in arginine, and water molecule and chloride ion. In the subsequent figures, all nitrogen atoms are shown in blue, oxygen in red, carbon in cyan, chloride in green, and hydrogen in white.**

deprotonated C terminal. The parameters for the N and C terminal were taken from the CTER and NTER parameters available in CHARMM. Mass densities were compared to the experimental density data,[3] and the error was found to be $\sim$ 1 %. All simulations were performed in the NpT ensemble with periodic boundary conditions, and full electrostatics were computed using the particle mesh Ewald (PME) method,[114] with a grid spacing of 1 Å or less. The pressure was maintained at 1 atm using the Langevin piston method,[115] with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature of 298 K. An integration step of 1 fs was used. The initial size of the periodic rectangular box was set to (50 Å)^3 in all of the simulations. To set up the simulation systems for various arginine hydrochloride concentrations ranging from 0.25 to 2.75 molal, arginine and chloride ions were randomly placed within the simulation box (while assuring no overlap), and subsequently overlapping water molecules were removed. The system was then equilibriated for 1 ns at constant pressure and temperature. All 11 systems described in Table 3.1 were simulated for 10 ns each. Simulations with 2.50 molal concentration were also performed at five different temperatures ranging from 278 to 358 K for 10 ns. We have also analyzed the interaction of arginine (1 molal concentration, 298 K, 1 atm, pH 4.75) with the surface groups of the protein α-Chymotripsinogen A. A 50 ns simulation of α-Chymotripsinogen A was performed

in a 75 Å box containing 188 arginine and 10411 water molecules. In order to study the interaction of arginine with hydrophobic residues which are not normally accessible in a folded protein, a 25 ns simulation of a helical peptide, melittin in a 2.75 molal ArgHCl solution was performed. Melittin is a 26 amino acid peptide with 15 hydrophobic residues and has a +6 charge at pH 7.0.[116,117] The simulation box with sides of 60 Å contained 1 melittin, 204 Arginine, 210 chloride ions and 4121 water molecules.

3.1.2 Clustering

Self-aggregation or clustering of arginine molecules was quantified in terms of the reduction of total solvent accessible surface area of arginine with respect to a monomeric arginine molecule. The solvent accessible area was estimated using standard CHARMM commands with a probe sphere of 1.4 Å radius. The solvent accessible surface area was used as a measure, since minimization of exposed surface area is one of the main driving forces for clustering. Furthermore, it has been reported that this measure is more sensitive than other measures like Kirkwood-Buff integrals.[32]

3.1.3 Effect of cosolvent on the protein association reaction

In order to quantify the effect of cosolvents on protein-protein association reactions, we study a model system for the association of proteins. The reaction of two parallel, planar plates and the reaction of two spheres are used as the extreme cases of the geometry of two associating proteins. The two spheres, each 20 Å in radius are used a model for the association reaction involving two spherical proteins. The distance between the centre of the proteins is defined as the reaction coordinate, x. For planes, the distance between the faces of the planes acts as the reaction coordinate. The surface area of the plates (400 Å^2) is selected to make the change in the protein solvent accessible area of reaction same as for the case of the two spheres. The thermodynamic properties of these plates are obtained by calculating the property per unit surface area of a pair of infinite plates and then multiplied by the area. The free energy of this protein complex, $\mu_{2,0}(x)$ can be modeled as described

previously,[42]

$$Spheres: \quad \mu_{2,0}(x) = \left(\tfrac{15}{x}\right)^6 - 8.21e^{-\left(\frac{x-20}{10}\right)^2} + 2.12e^{-\left(\frac{x-40}{10}\right)^2}, \tag{3.1}$$

$$Planes: \quad \mu_{2,0}(x) = \left(\tfrac{1}{x}\right)^6 - 8.51e^{-(x-1.5)^2} + 2.02e^{-\left(\frac{x-4}{2}\right)^2}. \tag{3.2}$$

The above equation places a 2 kcal/mol free energy barrier for the association between the two proteins. The dimer state is chosen to be 8 kcal/mol more stable than the monomer state. The free energy in the presence of cosolvent is computed by adding the free energy in the absence of additive with the transfer free energy. The transfer free energy, $\Delta\mu_2^{tr}$ is computed via:[42,68]

$$\Delta\mu_2^{tr} = -\int_0^{m_3} \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \Gamma_{23} dm_3 \tag{3.3}$$

The expression for Γ_{23} from equation 2.7 is substituted into the above equation.

$$\Delta\mu_2^{tr} = -\int_0^{m_3} c_3 \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \left(\int (g_3 - g_1) dV\right) dm_3 \tag{3.4}$$

The above equation can be simplified using the following assumptions. (1) The water cosolvent interactions are ideal, which makes the derivative of μ_3 with respect to m_3 equal to RT/m_3. (2) The concentration of cosolvent is low so that molal and molar concentrations are equal. (3) The RDF of the cosolvent and water with respect to the protein are represented using a three parameter Exp-6 potential. This function was fitted to the radial distribution functions obtained using molecular dynamics simulations.[42,68] Applying the above assumptions equation 3.4 is simplified to

$$\Delta\mu_2^{tr} = -RTc_3 \int (e^{-\langle U_{23}\rangle/RT} - e^{-\langle U_{21}\rangle/RT}) dV(x), \tag{3.5}$$

where R is the gas constant, T is the absolute temperature, c_3 is the cosolvent concentration, $\langle U_{23}\rangle$ is the protein-cosolvent potential of mean force, $\langle U_{21}\rangle$ is the protein-water potential of mean force. The intergral is over the system volume which is a function of the reaction

coordinate. The relative change in the association rate can be calculated using:

$$\frac{k_a}{k_0} = e^{-\Delta\Delta\mu_2^{tr}/k_bT}, \tag{3.6}$$

where k_a is the rate constant in the presence of cosolvent, k_0 is the rate constant in the absence of cosolvent, $\Delta\Delta\mu_2^{tr}$ is the change in the activation free energy, k_b is the Boltzmann's constant, and T is the absolute temperature.

3.2 Results and Discussion

3.2.1 Structural properties

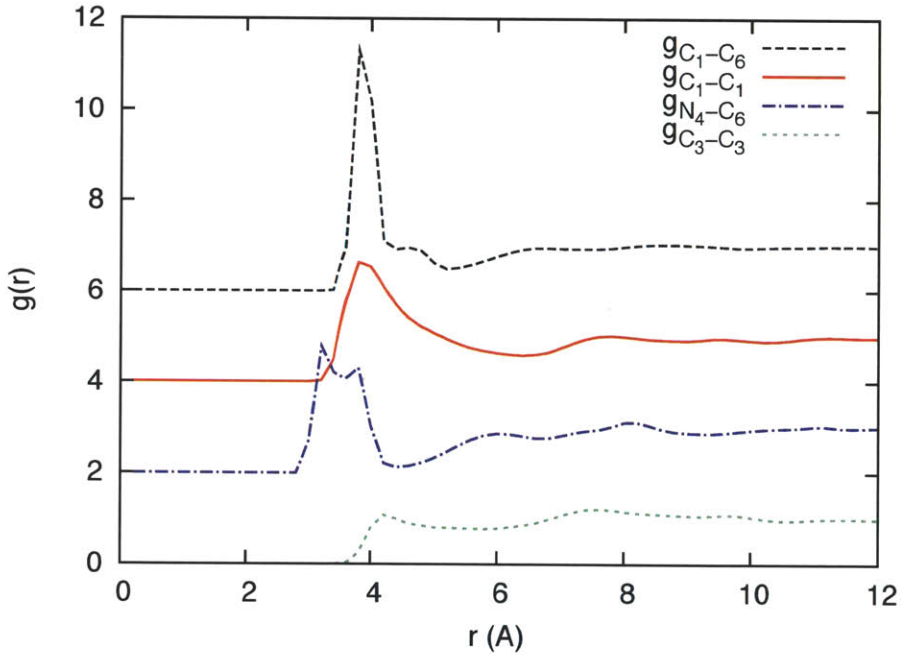

Figure 3-2: **Radial distribution functions between Gdm and carboxylate carbon ($g_{C_1-C_6}$), the Gdm carbon atoms ($g_{C_1-C_1}$), and between the N-terminal and the C-terminal ($g_{N_4-C_6}$) of the arginine molecule at a concentration of 2.75 molal.**

Spatial density distributions, including the translational and the rotational degrees of freedom, would capture details of the solution order. Such an analysis is possible for planar

and rigid molecules like urea and Gdm.[32,57] A conformation analysis of arginine in the gas phase shows a large number of local minima with comparable energies, due to the many easily rotatable single bonds.[118,119] Due to the flexibility of arginine, spatial density distributions were found to be smeared in the vicinity of the molecule. Therefore, site-site radial distribution functions (RDF) were used to characterize the geometry and short-range order of arginine-water solutions.

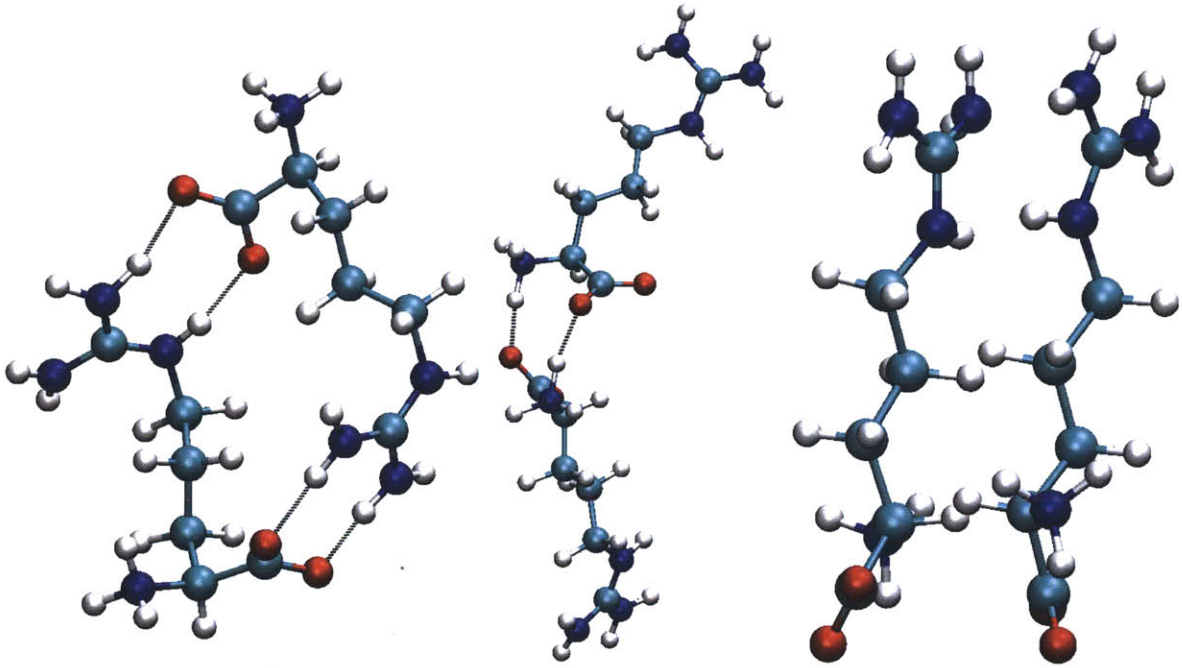

Figure 3-3: **Snapshots of arginine molecules from the MD simulation illustrating possible configurations in which the two arginine molecules can interact to from a dimer. A dimer formed due to the hydrogen bonding between the Gdm and carboxylate groups (left). A dimer formed due to hydrogen bonding between the N-terminal and the C-terminal of adjacent arginine molecules (center). A dimer with the stacked Gdm groups (right).**

Figure 3-2 shows the RDF between the middle carbon atoms ($g_{C_3-C_3}$)of the arginine molecules at 2.75 molal concentration. The first distinct peak is at 4.2 Å, and the second peak is at 7.6 Å, with a shoulder at 9.6 Å. The first peak corresponds to the two adjacent arginine molecules bonded together by two or more hydrogen bonds in a head-to-tail fashion due to the strong interaction between the Gdm and carboxylate groups as shown in the left panel of Figure 3-3. The second peak corresponds to adjacent arginine molecules with

their N-terminal and carboxylate groups hydrogen bonded (center panel Figure 3-3), and Gdm groups stacked on top of each other as shown in the right panel of Figure 3-3. Gdm ion stacking in aqueous solutions has been reported by Mason et. al.[57] based on the MD simulation and neutron diffraction experiment of 3 molal GdmHCl solution. Vondrášek et. al.[120] have reported that the cavitation energy, dispersion interactions, and the reduction in electrostatic repulsion due to the flat geometry and non-homogeneous charge distribution are the main factors responsible for the favorable association of like charge Gdm-Gdm pairs. The distance between Gdm groups is observed to be around 4 Å. This distance is larger than the Van der Waals contact distance, but it is not sufficiently large for water molecules to occupy the gap. Gdm-Gdm pair can also be stabilized by the tendency of the chloride ions to occupy equatorial positions in the Gdm plane. The presence of the chloride ion diminishes the inter-ion repulsion due to like-charged Gdm groups.[121] The distance between Gdm carbons in this case is observed to be around 7.5 Å, which is the same as the value reported for hetero-ion complexes in GdmHCl solutions.[57] Some arginine-arginine dimers were found to be in fully extended conformations in which the carboxylate and N-terminal of the adjacent arginine molecules were held together by two hydrogen bonds as shown in Figure 3-3. The shoulder in the second peak corresponds to such dimers. The dimers formed due to hydrogen bonding between Gdm and carboxylate groups of adjacent arginine molecules also contribute to the same peak. It was observed that such tail-to-tail or head-to-tail hydrogen bonding increases with the increase in concentration as arginine dimers can interact with other dimers or free arginine molecules. The guanidinium carbon-carbon ($g_{C_1-C_1}$) and the Gdm carbon-carboxylate carbon ($g_{C_1-C_6}$) RDF's are shown in Figure 3-2. The C_1-C_1 RDF shows two peaks around 4 Å, and 7.5 Å. These peaks correspond to the two different types of Gdm-Gdm stacking discussed above. The C_1-C_6 RDF shows a peak at 3.8 Å, with a shoulder at 4.2 Å. The maximum corresponds to the hydrogen bonding between Gdm and carboxylate groups of adjacent arginine molecules. The shoulder corresponds to the carboxylate group hydrogen bonded to the N-terminal group of the adjacent arginine molecule. These RDF's also give a measure of the population of two different types of dimers

shown in Figure 3-3. It can be seen from the figures and the coordination number obtained after integrating these curves (not shown here) that the population of the first dimer which has two arginine molecules bonded with two to four hydrogen bonds and is higher than the other dimers. The N_4-C_6 RDF exhibits a peak around 3.5 Å, which correspond to the the hydrogen bond between the the N-terminal and C-terminal of adjacent arginine molecules. The C-terminal carbon within the same molecule is not included in the calculation of the RDF.

The radial distribution functions for the N-terminal nitrogen and the chloride ion are shown in the Figure 3-4a. The prominent first peak in this function is at 3.2 Å, with a significant shoulder at 5.4 Å. The maximum corresponds to the chloride ion occupying the position between hydrogen atoms of the N-terminal nitrogen (see Figure 3-5). The shoulder corresponds to the chloride ions coordinated to a water molecule hydrogen bonded to the N terminal nitrogen. There is a competition among the chloride, water and the arginine atoms to hydrogen bond with the N-terminal nitrogen. The peak height decreases with increasing concentration due to the hydrogen bonding interaction between the N-terminal and the carboxylate group of arginine. The chloride-chloride radial distribution function (Figure 3-4b) shows three peaks. The first maximum is at 4.8 Å, which corresponds to the two chloride ions mutually coordinated to an intervening water molecule. The second peak falls at 7.4 Å and corresponds to the two chloride ions occupying adjacent positions in the first hydration shell around the Gdm group and the N-terminal (see Figure 3-5). The third peak at 9.8 Å corresponds to the two chloride ions coordinated with the Gdm group and the N-terminal each. In agreement with the previous MD simulations of chloride salts,[57,122] no significant direct chloride ion pairing was observed under these conditions. The Gdm carbon-chloride RDF (Figure 3-4c) shows two prominent peaks. The first peak is at 3.8 Å, with a shoulder at 4.5 Å. The first peak corresponds to the chloride ion occupying the position in the first hydration shell between the hydrogen atoms of the Gdm, hydrogen bonding to each at angle of 140°. The shoulder corresponds to the chloride ions making a single linear hydrogen bond to a single hydrogen of the Gdm (see Figure 3-5). The second peak in the RDF is at a

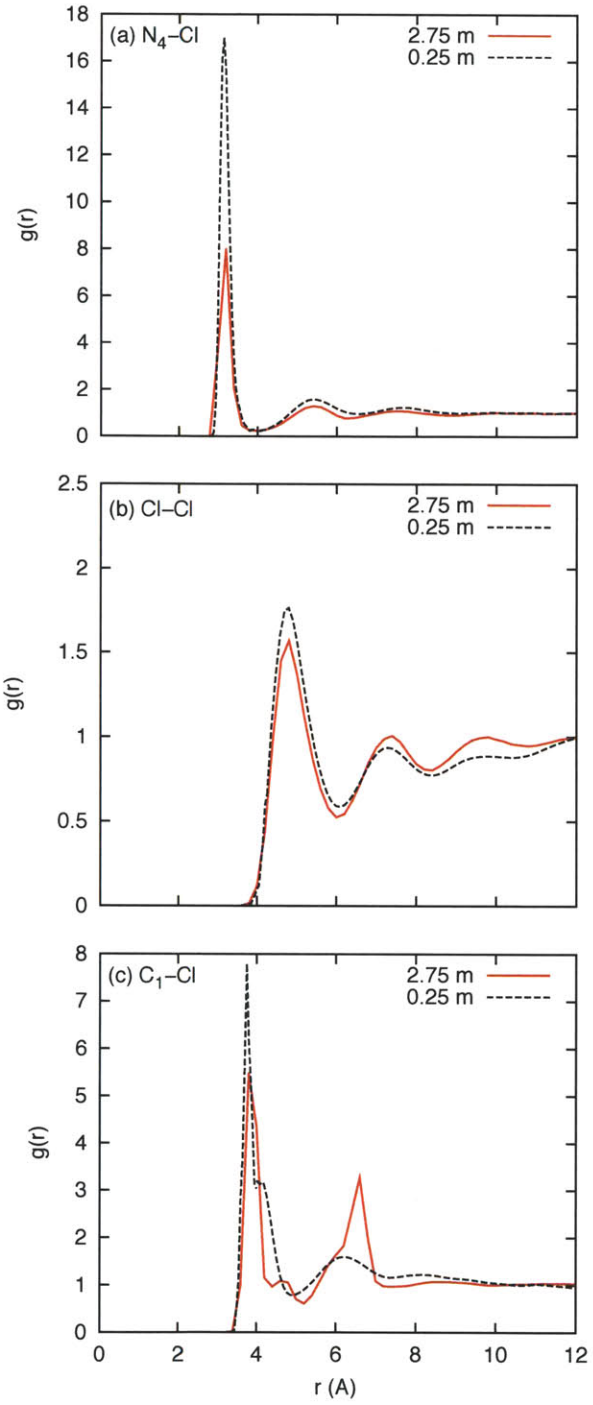

Figure 3-4: **Radial distribution functions between chloride ions and (a) the N-terminal Nitrogen, (b) the chloride ion, and (c) the Gdm carbon.**

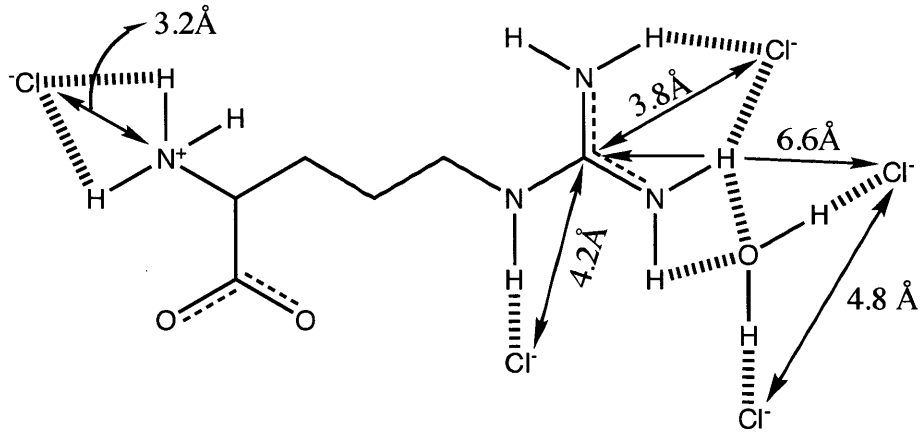

Figure 3-5: **Arginine-Chloride ion pairing found in the simulations. The distances reported in the figure correspond to the peaks in radial distribution functions shown in the Figure 3-4. Hashed lines denote a hydrogen bond and dotted lines denote a partial double bond.**

distance of 6.6 Å and corresponds to the chloride hydrogen bonded to hydrogen atoms of the Gdm with a water molecule acting as a bridge. The first peak decreases in height and second peak height increases with increase in concentration due to the hydrogen bonding interaction between arginine molecules which restricts the direct interaction of chloride ion.

RDF's between water-oxygen and arginine atoms are shown in Figure 3-6. Water molecules can be seen to be strongly coordinated with the three charged groups in the molecule as shown by the prominent peaks in the RDF's (N_4-O_w,C_6-O_w,C_1-O_w). However, the middle carbon atom shows no prominent peak due to the hydrophobic nature of the three methylene groups in the center of the molecule. The water structure around arginine is not significantly perturbed due to the self-interaction of arginine. The water molecules form linear hydrogen bonds with the Gdm group and are therefore, constrained to remain in the plane of the Gdm group. When two arginine molecules form a cluster with Gdm-Gdm stacking, the equatorial positions are still unoccupied. Furthermore, these n-mers are held together by bonding with only a part of the group. For example, in a dimer with head-to-tail bonding, only one of the nitrogen atoms in the Gdm group is involved. The water-water radial distribution functions (not shown here) also supports this observation. It is seen that arginine only slightly per-

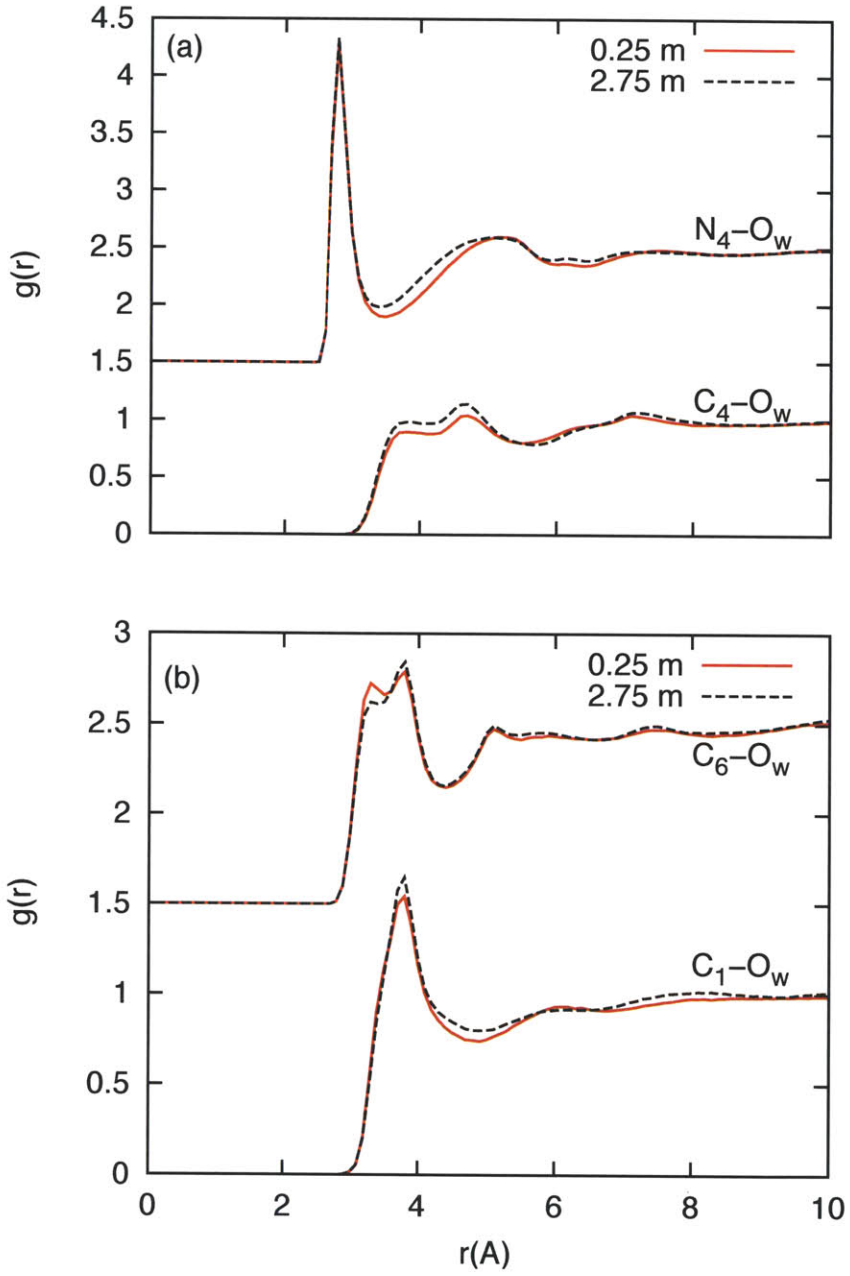

Figure 3-6: **Arginine-water, site-site radial distribution functions. RDF's for N-terminal nitrogen, and carboxylate carbon (C_6) are offset 1.5 unit along the ordinate.**

turbs the water structure. The population of the first solvation shell, which can be used as a measure of the short-range order in the water structure does not change significantly with

an increase in argHCl concentration. This observation indicates that the indirect effect of arginine on proteins through its effect on the structure of water is not significant.

3.2.2 Hydrogen bonds

The strength and the number of hydrogen bonds in the aqueous arginine solutions give further insight into the nature of the interactions. Figure 3-7 shows the number of hydrogen bonds per molecule for different arginine concentrations. The number of arginine-arginine hydrogen bonds per arginine molecule increases with increasing concentration. The number of arginine-water bonds decreases with the increasing concentration. The total number of bonds per arginine molecule decreases slightly as fewer arginine-arginine bonds are formed than arginine-water bonds are lost. For water, similar behavior is observed in terms of total number of hydrogen bonds per water molecule. This marginal decrease in the number of hydrogen bonds per arginine or water molecule again confirm the observation that arginine perturbs the water structure only slightly.

Figure 3-8 shows the hydrogen-bond energies for all donor-acceptor combinations. The strongest hydrogen bond is formed between the carboxylate oxygens (acceptor) and the N-terminal nitrogen (donor) pair with an average energy of 33 kJ/mol (298 K). The strongest hydrogen bonds between arginine and water were those formed between the water oxygen (acceptor) and the carboxylate oxygen (donor) with a mean energy of ~30 kJ/mol (298 K). The water-water hydrogen bonds (20.5 KJ/mol) are weak as compared to arginine-arginine or arginine-water hydrogen bonds. The self-aggregation of arginine is therefore enthalpically favorable. From the entropic viewpoint, a single arginine molecule would free about 10 water molecules. The number of water molecules replaced by an arginine is calculated by placing an arginine molecule in a water box and counting the number of water molecules within 2.2 Å of arginine molecule. Therefore, the self-interaction is promoted as the water molecules solvating the individual arginine molecules are released, which increases the translational and rotational entropy. The lifetime of individual arginine-arginine hydrogen bonds is calculated using the procedure of Rapaport.[123] The average bond lifetime is found to vary between

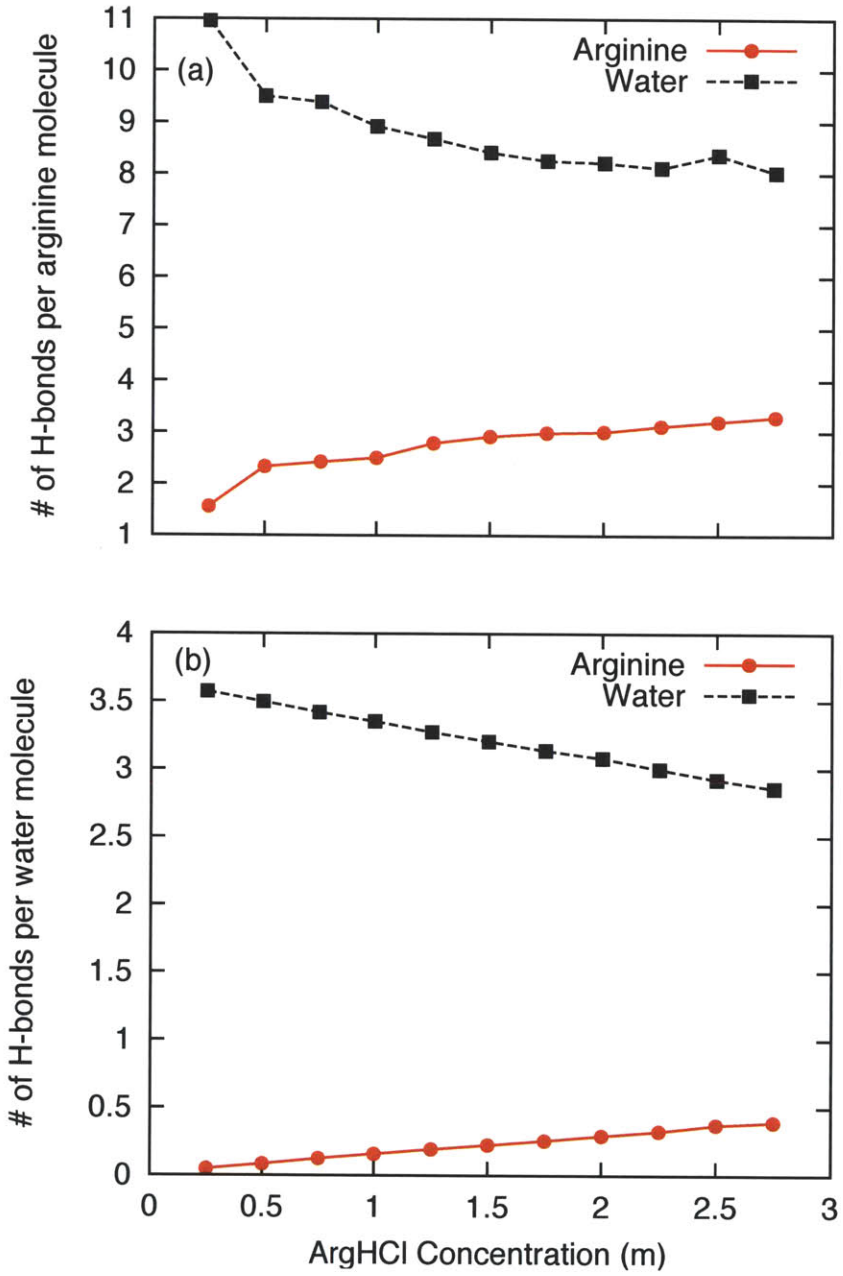

Figure 3-7: **Average number of hydrogen bonds (a) per arginine molecule (b) per water molecule.**

7-9 ps, depending on the solution concentration. However, the time for which two arginine molecules in a cluster stay connected is found to lie between 90-110 ps. The clusters are held together by multiple hydrogen bonds, the individual hydrogen bonds breaking after every

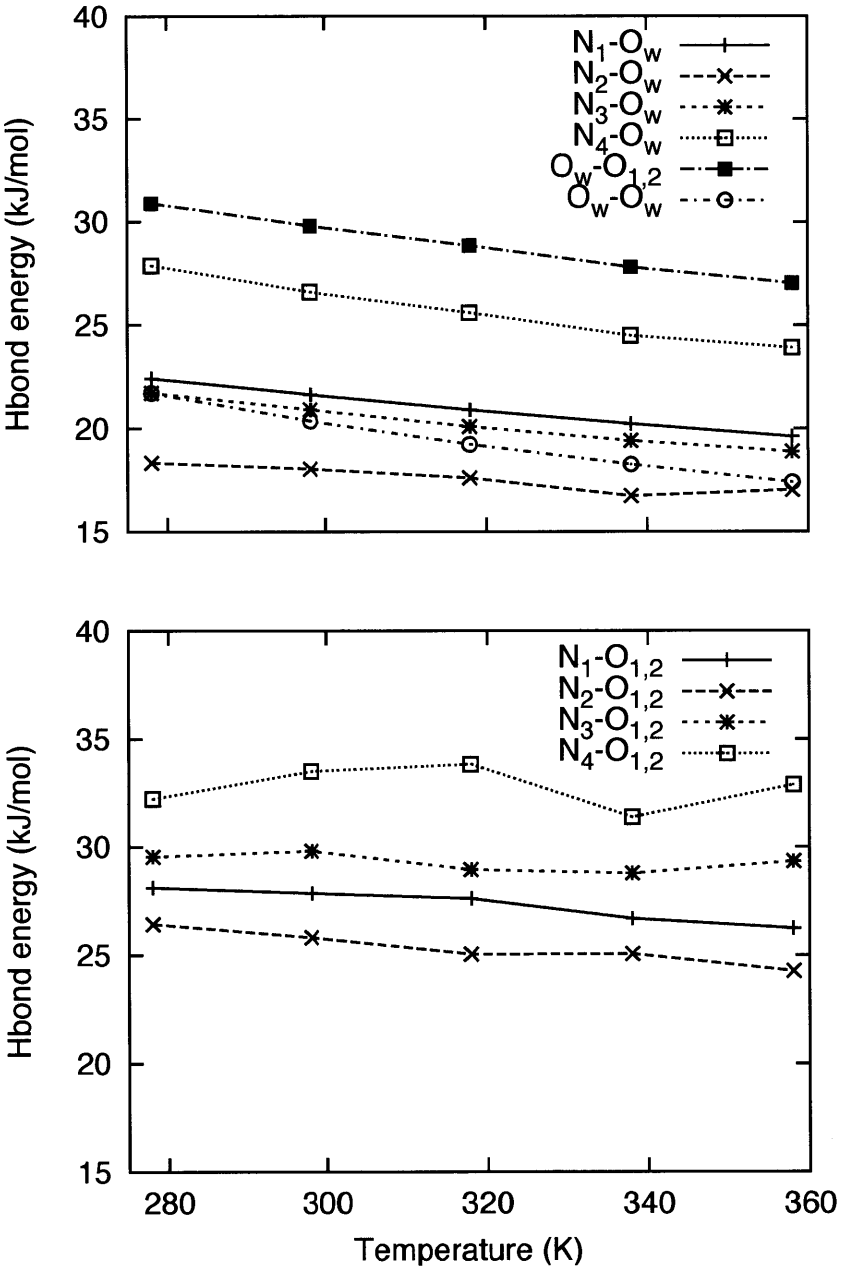

Figure 3-8: **Mean energy per hydrogen bond for all donor-acceptor combinations at 2.50 molal concentration.**

7-9 ps, but the cluster breaks after a much longer time.

3.2.3 Clustering

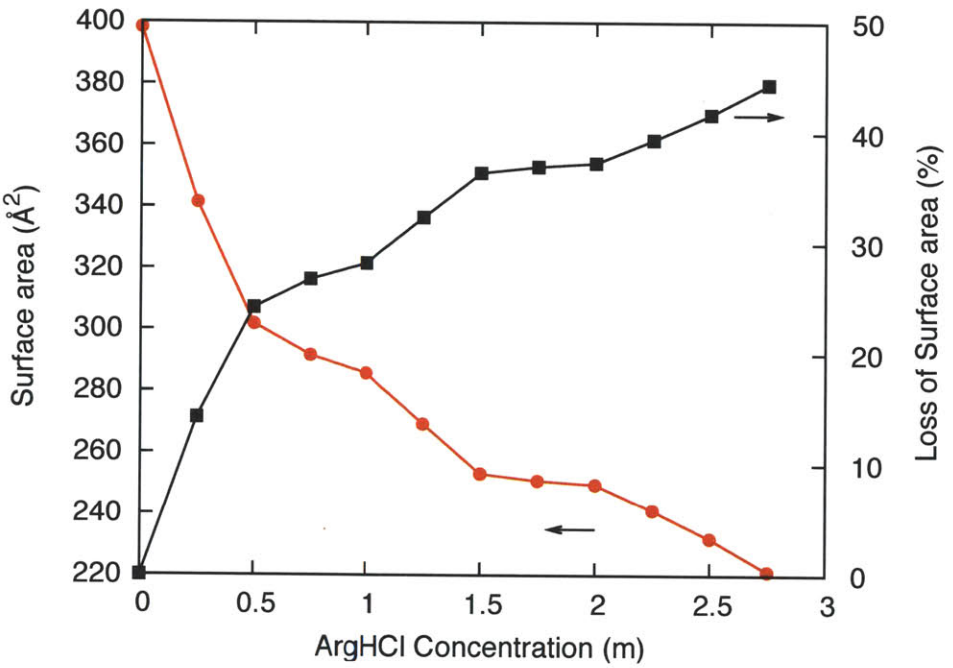

Figure 3-9: **Solvent accessible surface area as a function of argHCl molality normalized by the total number of arginine molecules in the system (left). Percentage loss in surface area as a function of argHCl molality. (right)**

Figure 3-9 shows the mean surface area per arginine molecule and the percentage loss in surface area compared to that of a single arginine molecule in a water box as a function of concentration. The maximum percentage loss in area is 50%. The solubility of argHCl at 298 K is 2.81 molal. For the highest concentration (2.75 molal) studied here, the percentage loss in surface area is ~45%. With increaseing temperature, the extent of clustering (for 2.50 molal) is observed to decrease, due to a higher solubility and increase in the overall volume of box. However, this approach overestimates the extent of clustering due to the fact that, in the absence of any interaction, random contacts between arginine molecules would reduce this solvent accessible surface area. Stumpe and Grubmüller[32] have proposed a procedure to distinguish this from real self-aggregation for the clustering in aqueous urea solutions. Such area loss due to random contacts is expected to be small in case of arginine due to the low diffusivity and strong interaction between arginine molecules.

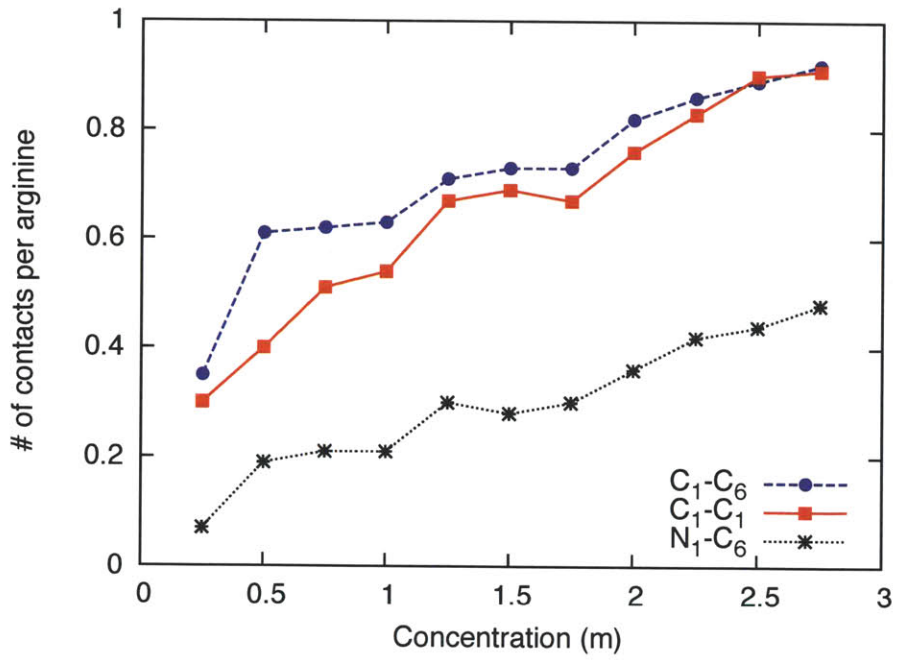

Figure 3-10: **Number of contacts per arginine molecule between the Gdm and Carboxylate groups (C_1-C_6), Gdm groups (C_1-C_1), and the N and C-terminal of arginine molecules in solution.**

The populations of clusters of various sizes were calculated where the criteria for a molecule to be in a cluster are (a) connected by one or more hydrogen bonds (b) the specified atoms are within a given distance range, which is defined by the position of the first minimum between the atoms in the appropriate g(r). The second constraint applies only to the clusters with Gdm-Gdm stacking. The cluster size is calculated by counting all the molecules that are connected to one other molecule in a cluster. The number of contacts between the groups in the interacting arginine molecules is calculated using the same criteria. The number of Gdm-Carboxylate, Gdm-Gdm, and N-terminal-C terminal contacts per arginine molecule as a function of concentration is shown in Figure 3-10. The Gdm-Gdm and Gdm-Carboxylate are the dominant modes of contact between arginine molecules in solution. A single arginine molecule can form up to four contacts. The number of contacts is observed to increase with increasing concentration. The total number of contacts at 2.75 molal concentration is 2.3. The population of clusters as a function of size and the probability to find an arginine molecule in a cluster of a given size is shown in Figure 3-11. It is found that for the 0.25

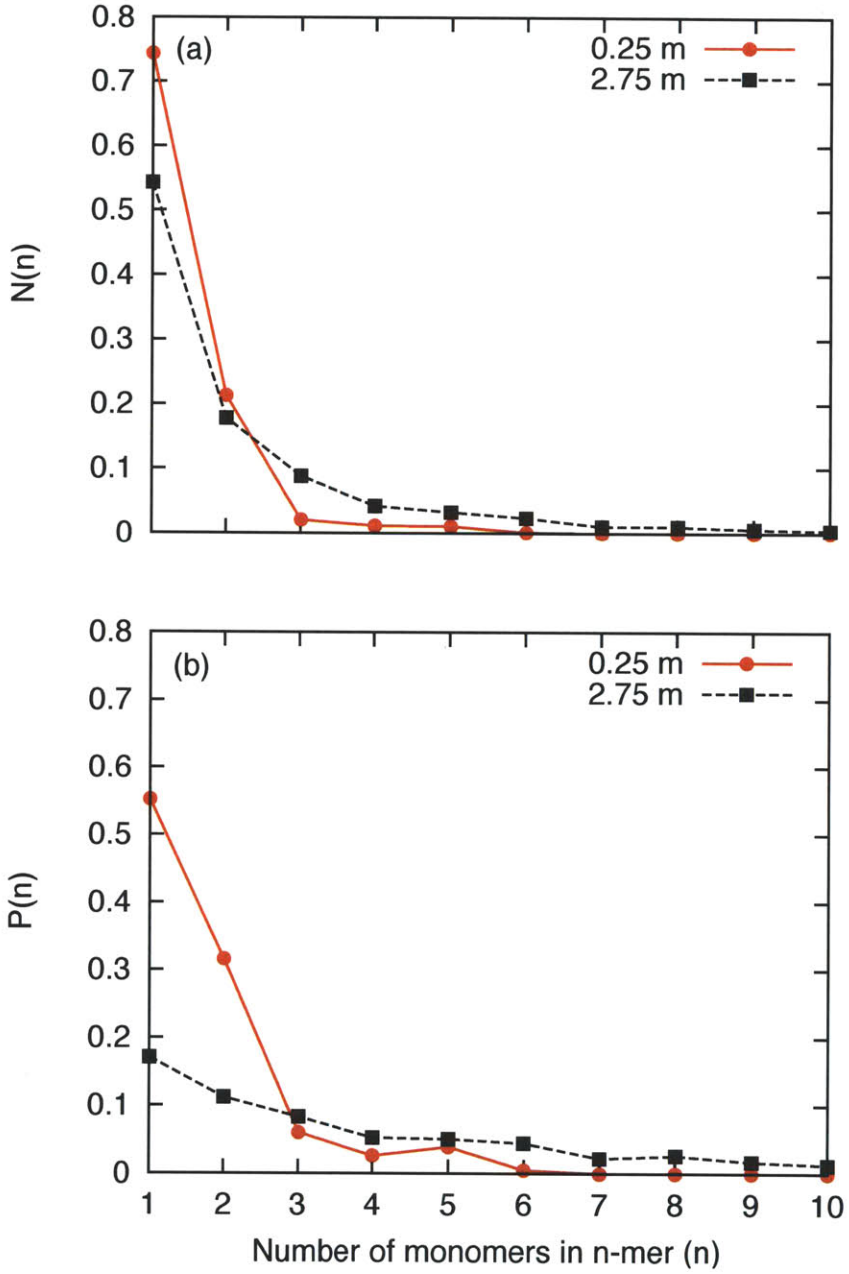

Figure 3-11: **Distribution of population of arginine clusters. (a) Normalized population of arginine clusters of various sizes. (b) Probability of finding an arginine molecule in a cluster of a particular size.**

molal system (Figure 3-11b), 45% of the arginine are in dimers or higher order n-mers. On an average, only 13% of the arginine was found to be in n-mers larger than a dimer. If a number

of clusters of various sizes are considered, only 25% of the clusters are dimers or higher order n-mers. For the 2.75 molal system, the number of arginine molecules present as monomers is as low as 18%. The n-mers from monomer to decamer account for ~60% of the total arginine molecules. From the plot of number of clusters of various sizes, it can be seen that there are few very large clusters present in the solution at 2.75 molal. For low concentrations, the solution is dominated by the arginine monomers and dimers but as the concentration increases the n-mers of larger sizes are formed. These n-mers are nanoscale clusters held together by hydrogen bonds, hetero-ion pairing and Gdm-Gdm stacking. These clusters are significantly larger in size as compared to a single arginine molecule, and as such, they are expected to play a significant role in stabilizing the protein molecules in mixed solvents.

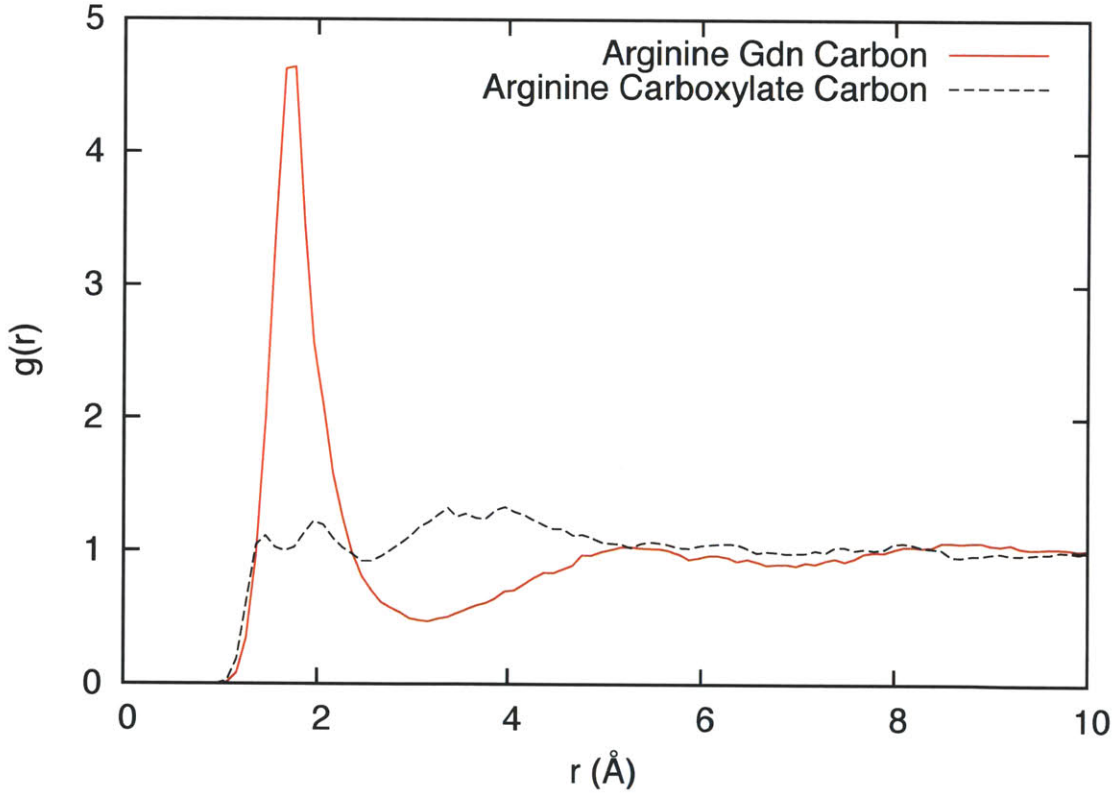

Figure 3-12: **Radial distribution functions of the guanidinium carbon(C_1) and the carboxylate carbon(C_6) in the arginine molecule around the protein α-Cgn A.**

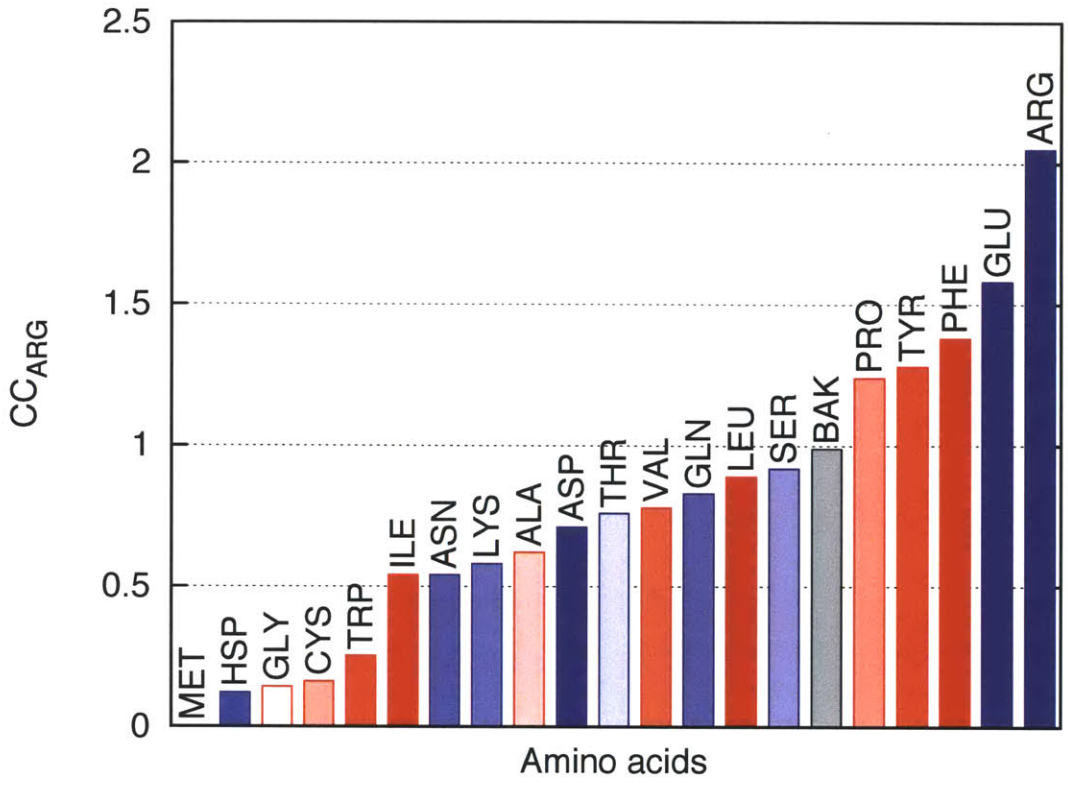

Figure 3-13: **Contact coefficient, CC_{ARG} for each amino acid in α-Cgn A. The colors represent the hydrophobicity of amino acids. Red Hydrophobic; blue Hydrophilic; The intensity of the bars depends on the normalized hydrophobicity values[6] of each amino acid. The label BAK denotes the protein backbone. It can be seen that there is no trend between the CC values and the hydrophobicity of the residues.**

3.3 Protein-Arginine Interaction

In order to understand the role that interactions between the protein surface and the arginine play in the mechanism by which arginine inhibits aggregation, a molecular dynamics simulation of the protein α-Chymotripsinogen A in a 1 molal aqueous arginine solution was performed. Arginine is found to interact with the protein surface mainly via the guanidinium group. Radial distribution functions of the guanidinium and the carboxylate carbon in the arginine molecule around the protein α-Cgn A are shown in Figure 3-12. It can be seen that arginine is preferentially oriented with respect to the protein surface. The preferential binding behavior of arginine with each type of amino acid and the backbone in α-Cgn A is shown

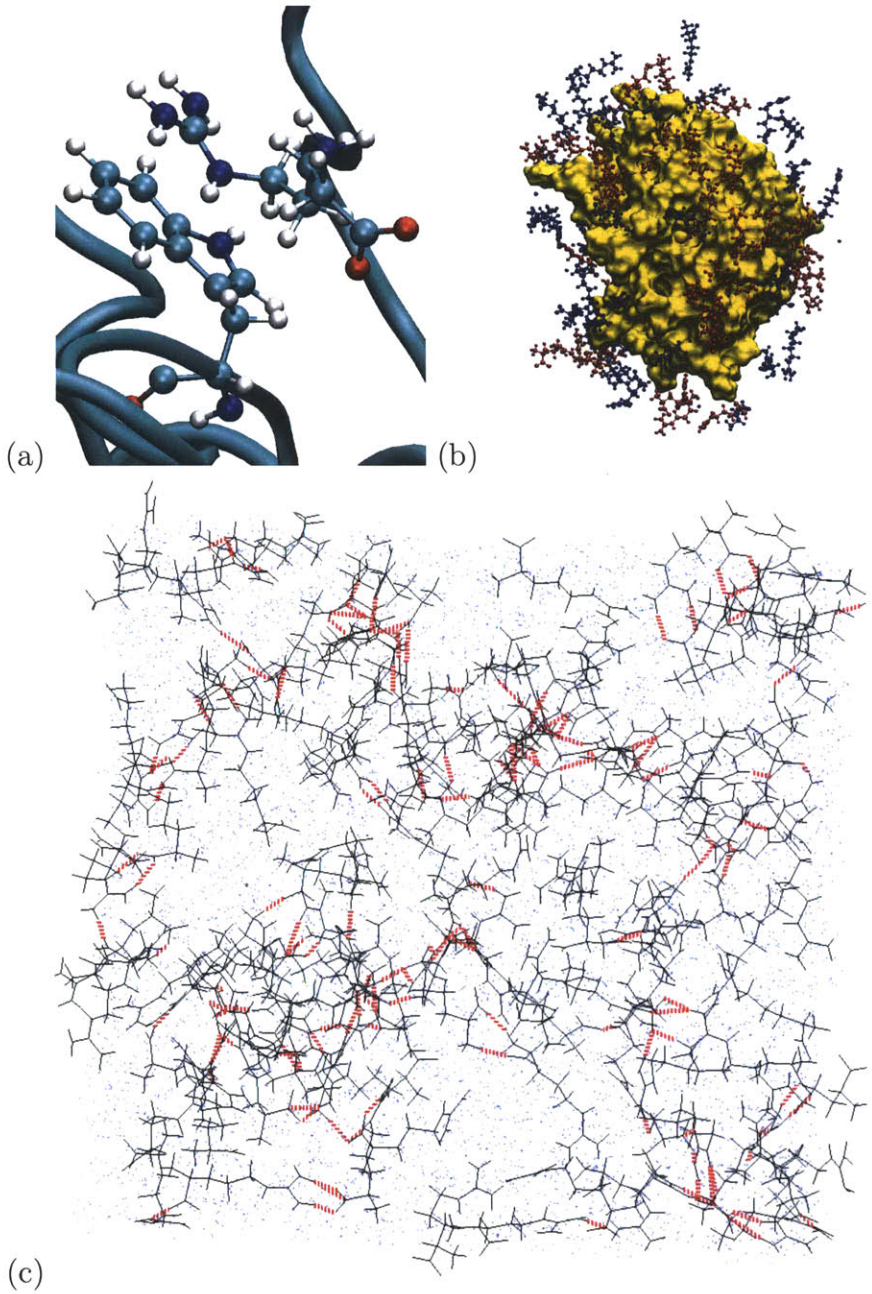

Figure 3-14: **(a) Cation-π interaction between a tryptophan residue on the surface of α-Cgn A and Gdm group of an arginine molecule in solution. (b) Snapshot of arginine molecules present in the local domain (6.0 Å) of α-Chymotripsinogen A. Arginine molecules present as dimers are shown in red. (c) Snapshot of the MD simulation box containing aqueous arginine solutions at 1 molal concentration (right). Water molecules are shown as blue dots for clarity. Hydrogen bonds between arginine molecules are shown as hashed red lines.**

in Figure 3-13. The contact coefficient is defined as the ratio of the arginine concentration around a particular amino acid and the bulk concentration. Stumpe and Grubmüller[34] have reported contact coefficient values for interaction of urea with glycine capped tri-peptides. The normalization with the bulk concentration gives a better idea of the tendency of the cosolvent to interact preferentially with the surface of a particular amino acid. The number of arginine molecules and chloride ions bound to each amino acid on the protein surface is included in the supporting information. Arginine is observed to interact strongly with the aromatic and charged residues on the protein surface. The charged groups in arginine can interact with both the positively and negatively charged amino acid side chains. Arginine is observed to interact with charged residues via hydrogen bonding, similar to the behavior of arginine in aqueous solution. The average concentration around the arginine and glutamic acid residues on the protein surface is found to be 2.1 and 1.6 times the bulk concentration respectively. The Gdm group interacts with aromatic side chains residues via cation-π interactions. Figure 3-14a shows an arginine molecule interacting with a tryptophan residue on the protein surface. The average surface concentration of arginine around tyrosine and phenylalanine side chains in α-Cgn A is found to be 1.4 and 1.3 times the bulk concentration respectively. The backbone atoms can hydrogen bond with the arginine molecules. The average number of arginine molecules bound to the backbone atoms is 5.8, which corresponds to a contact coefficient value of 1.0. In our simulations, interactions between hydrophobic residues and arginine is observed. The methylene groups in the arginine molecule can interact with the hydrophobic residues on the protein surface. If we consider all the non-aromatic hydrophobic residues (Val, Ile, Leu, Met, Cys, Ala, Pro, and Gly) the average local concentration around these residues is 0.5 times the bulk concentration which indicates that arginine does not interact strongly with the hydrophobic resides. In the native structure of the protein, the hydrophobic resides are not accessible to the solvent. Methionine, Histidine, and Tryptophan have low contact coefficient values due to the limited exposure of these residues to the solvent. Water molecules can interact with these partially exposed residues, but large arginine molecules cannot. Therefore, a simulation of the polypeptide with exposed

hydrophobic groups (melittin) was performed to assess the interactions of arginine with the hydrophobic regions. The total number of arginine coordinated with the protein is 9.0. The number of arginine coordinated with the hydrophobic residues (Val, Ile, Leu, Pro, and Gly) is 2.9, corresponding to a local concentration of 0.54 relative to the bulk concentration. Although the number of aromatic and charged amino acids in melittin is less than the number of hydrophobic amino acids, these amino acids account for 6.1 molecules out of 9 molecules coordinated with the protein. Furthermore, arginine clusters were not found to interact with the hydrophobic residues by stacking the methylene groups to form a surface as suggested by Das et. al.[74] Therefore, simulation of a polypeptide with exposed hydrophobic groups also provides similar results for the interaction of arginine with protein surface residues.

The preferential interaction coefficient, Γ_{23}, measures the excess number of cosolvent molecules within the local domain of the protein as compared to that in bulk solution. Recently, Schneider and Trout[3] have reported preferential binding coefficient data for α-Chymotripsinogen A. The experimental preferential binding coefficient value at 1 molal concentration is -8.7±2.9. The computed value of -8.3±0.7 is in good agreement with the experimental data. Arginine solutions are comprised of monomers and higher order n-mers. It would be interesting to look at the preferential binding data for the monomers ($\Gamma_{23,m}$) and higher order n-mers. $\Gamma_{23,m}$ can be calculated by counting the number of monomers in the local and bulk domain. Similar calculations can also be performed for the higher order n-mers. However, the standard deviation of these coefficients would be large due to the small population of these n-mers. Therefore, longer simulations would be required to calculate these coefficients. However, the difference between the overall Γ_{23} and $\Gamma_{23,m}$ would give an average value for all n-mers (n ¿= 2). For α-Cgn A, the $\Gamma_{23,m}$ was found to be -5.08. The value for $\Gamma_{23,m}$ is higher than the overall Γ_{23}. This implies that the concentration of monomers within the local domain is higher that the n-mer (n¿=2) concentration. This is expected because the local concentration around protein is less than the bulk concentration and the self-interaction between the arginine molecules increases with the increasing concentration.

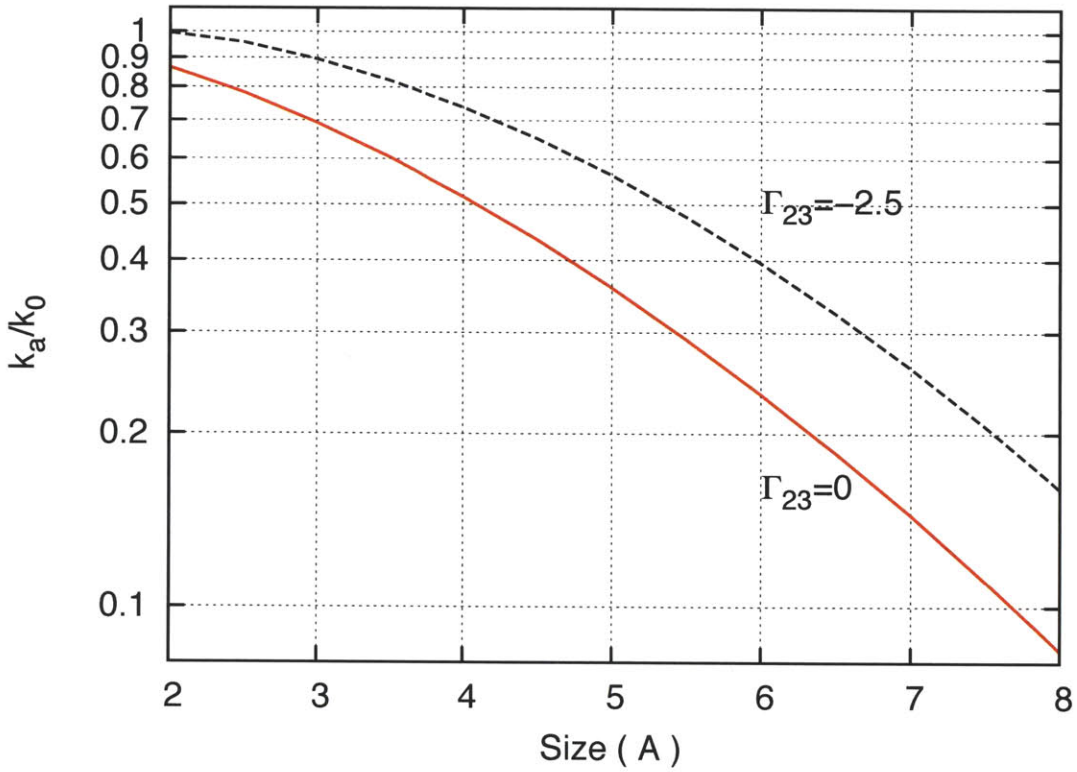

Figure 3-15: **Relative change in the association rate for 20 Å spherical proteins caused by a 0.5 M cosolvent solution as a function of cosolvent size.**

3.3.1 Mechanism

The three proposed mechanisms by which arginine inhibits protein aggregation should be reassessed based on the molecular level insights obtained from the simulations performed in this paper.

Neutral crowders are cosolvents which have a preferential binding coefficient value of zero (or slightly positive) and are larger than water in size. The preferential binding data of arginine from Schneider and Trout[3] clearly shows that arginine is highly excluded from the protein surface and therefore, it cannot be classified as neutral. However, arginine can still be classified as crowder. Arginine molecules which accumulate on the protein surface are responsible for crowding out the protein-protein interactions. According to the "gap effect" theory, the rate of association decreases with the increasing additive size and preferential interaction coefficient. Arginine has a negative preferential binding coefficient, which would

increase the association rate of proteins. On the basis of the simulations of aqueous arginine solutions, it is concluded that arginine forms clusters in solution and these clusters have sizes larger than the arginine molecule. This effective increase in the size of arginine due to self interaction counters the change in the association rate due to exclusion of arginine from the protein surface. Therefore, the observed aggregation suppression would be equivalent to a neutral crowder of the size of an arginine molecule. The aggregation suppression effect of arginine on the proteins can be illustrated clearly using the association model described in the methods section. The relative change in the association rate as a function of the additive size for two different values of Γ_{23} is shown in Figure 3-15. Arginine has a radius of gyration of 3.6 Å. If arginine is assumed to be neutral ($\Gamma_{23} = 0$) then relative change in the association rate is 0.57. If the arginine is excluded from the protein surface ($\Gamma_{23} < 0$), then the ratio k_a/k_0 would be greater than 0.57. However, the size of arginine is most likely enhanced due to the self-interaction which compensates for the negative contribution due to preferential interaction. Experimental or theoretical preferential interaction coefficients for arginine hydrochloride are reported for four proteins.[3] The values at 0.5 molal argHCl concentration lie in the range -4 to -1. Therefore, a value of -2.5 is chosen as the preferential interaction coefficient for the model protein. For $\Gamma_{23} = -2.5$, the value of k_a/k_0 is 0.8. From the $\Gamma_{23} = -2.5$ plot in the Figure 3-15, it can be seen that the effective size of the arginine should be about 4.8 Å for the k_a/k_0 to be equal to that of the neutral crowder($\Gamma_{23} = 0$) with the size of an arginine molecule. On the basis of the loss of the surface area per arginine molecule, the population of monomers and higher n-mers can be calculated. At 0.5 molal, the percentage of monomers is 68%. Therefore, the effective size of the arginine in the solution is calculated to be 4.77 Å, which is same as the effective size of the arginine required to compensate for the exclusion of arginine from the protein surface. The effective size is calculated based on the population of monomers and dimers in solution. At low concentrations, the solution predominately contains monomers and dimers. The size of the dimer is assumed to be equal to the size of the two arginine molecules separated by the distance equal to the location of the first minimum in the RDF between arginine molecules

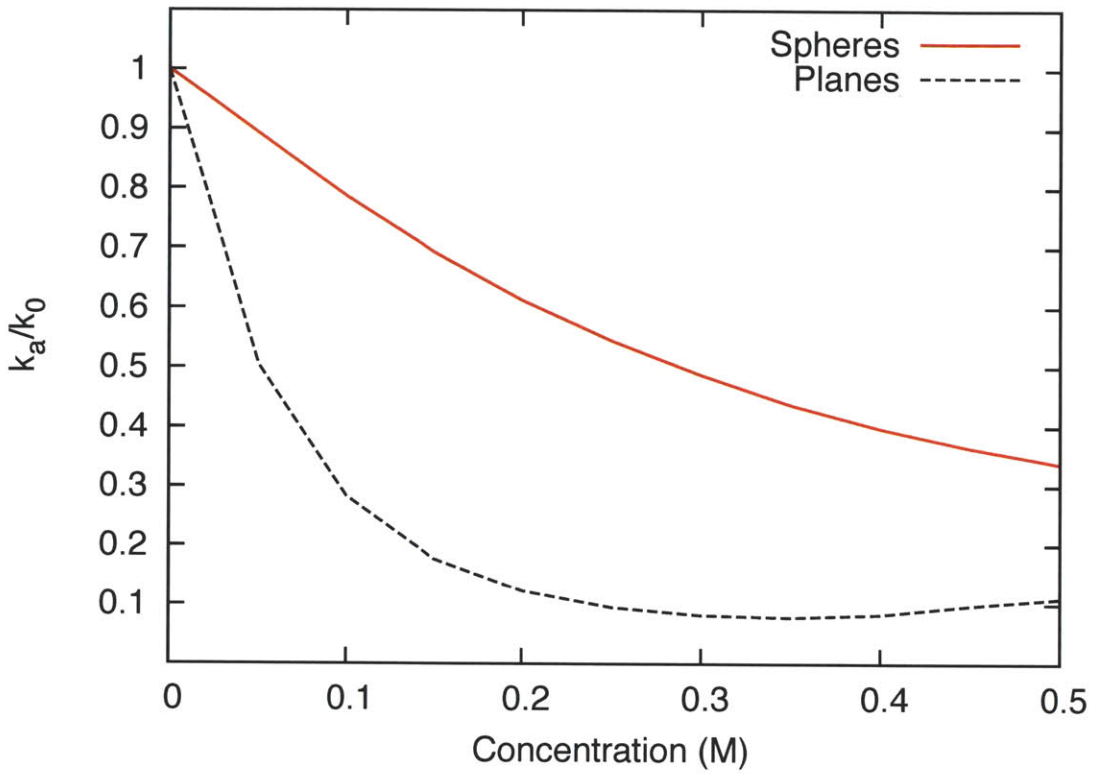

Figure 3-16: **Relative change in the association rate for the spherical and planar proteins due to arginine solution as a function of concentration.**

in solution. Arginine is preferentially excluded from the protein surface. Therefore, the local concentration around protein is less than the bulk concentration. The extent of clustering of arginine increases with increasing concentration, which implies that the $\Gamma_{23,monomer}$ will always be higher than the Γ_{23}. The $\Gamma_{23,monomer}$ value calculated for the simulation of α-Cgn A in 1 m arginine solution supports this point. From the viewpoint of aggregation suppression, monomers have a small size but a higher Γ, and n-mers have low Γ but are large in size. This observation shows the trade-off between Γ and the size of the cosolvent in terms of their effect on the association rate.

The relative change in the association rate as a function of arginine concentration for the two extreme cases of spherical and planar proteins is shown in Figure 3-16. In order to calculate the rate of association in the presence of arginine, experimental preferential interaction data[3] for lysozyme in arginine solution and the effective size of the arginine

molecule in solution are used. The relative change in the association rate is calculated using equation 3.6. Hirano et. al.[124] have reported the aggregation rate constant for lysozyme with and without arginine. The experimental value at 600 mM is found to be 0.14, which is in agreement with our predictions of 0.11-0.30.

On the basis of the simulation results, it can be concluded that arginine interacts strongly with aromatic residues. The number of aromatic amino acids in any protein are typically less than 10% of the total number of amino acids.[125] Therefore, the number of arginine molecules bound to the aromatic residues would be a small fraction of the total number of arginine molecules associated with the protein. If we consider a situation that all the aromatic residues in α-Cgn A are exposed then the number of arginine molecules bound to them can be calculated on the basis of the number of arginine molecules coordinated per residue for a particular amino acid. The number of arginine associated with the aromatic residues would be 12. However, if the same analysis is done for non-aromatic residues. The number of arginine associated with the non-aromatic residues is found to be 77. The number of arginine associated with the aromatic residues is a small fraction of the total arginine associated with the protein. Therefore, the interactions between arginine and aromatic residues alone cannot account for the aggregative suppression behavior of arginine.

Simulation results and experimental light scattering data suggest that arginine forms molecular clusters in solutions.[74] However, the hypothesis that these clusters present a large hydrophobic surface by the alignment of their methylene groups is not supported by the simulations. The simulation involving Melittin which has an exposed hydrophobic surface does not show any such interactions. No large hydrophobic surfaces similar to the hydrophobic columns present in L-arginine crystals is observed in simulations of aqueous arginine solutions. A typical snapshot of the MD simulation box containing aqueous arginine solution with and without protein is shown in Figure 3-14b and c. It can be seen that n-mers (¿ n=2) of arginine are formed by hydrogen bonding between charged groups without any alignment of methylene groups to form a large hydrophobic surface. In order to support the hypothesis that the interactions between methylene groups in arginine clusters and the

protein surface are responsible for the aggregation suppression, Das et. al.[74] measured the solubility of pyrene and performed ANS fluorescence emission intensity studies of aqueous arginine solutions. The solubility of pyrene was observed to increase in the presence of arginine. The solubility of pyrene increases with the decrease in the polarity of the solvent. The ANS fluorescence intensity also increased, and a blue shift in the maximum wavelength was observed. These changes in ANS fluorescence are observed when the fluorophore is in hydrophobic environment. On the basis of these observations, the authors suggested that methylene groups in arginine are responsible for these effects on non-polar compounds. However, pyrene is a polycyclic aromatic hydrocarbon, which forms a large flat aromatic system, and ANS is a charged molecule with phenyl and napthalene rings. It has been shown that the Gdm group interacts strongly with aromatic residues due to cation-π interactions.[33,65,126] Recently, Mason et. al.[33] have reported that GdmCl significantly suppresses aromatic interactions between pyridine molecules but has no effect on the isopropanol aggregation. Therefore, the Gdm group is more likely to be responsible for both the phenomenon observed by the authors. Das et. al. also showed that arginine increases the solubility of the Aβ_{1-42} peptide and decreases fibrillar formation. They suggested that arginine clusters with aligned methylene groups mask the hydrophobic surface of Aβ_{1-42}. However, Aβ_{1-42} has 4 aromatic and 12 charged residues which can interact with the arginine molecules via cation-π interaction and salt-bridge formation respectively. The authors have shown the effect of arginine on various compounds, but there is no conclusive evidence for the presence of a hydrophobic surface formed due to stacked methylene groups or relating the observed effects to such a hydrophobic surface. Furthermore, the simulations results reported in this paper show that there is no such surface formed due to the stacking of methylene groups. Therefore, the anti-aggregative property of arginine is most likely not due to hydrophobic interactions between the clusters and the protein surface.

The proposed effect of the arginine on the free energy of protein along the refolding/aggregation reaction coordinate is shown in Figure 3-17. In the absence of arginine (solid line), the small aggregate (A_2) is formed from an unfolded or partially unfolded inter-

mediate (I). These unfolded intermediates are formed due to the exposure of hydrophobic residues which are not exposed in the native state (N). In the presence of arginine (dotted line), the interaction of arginine with aromatic groups of the partially unfolded intermediates stabilizes these intermediates. Crowding around the macromolecule leads to an increase in the height of the barrier corresponding to the association of the partially unfolded proteins to form an aggregate. The free energy of the encounter complex increases due to the exclusion of the arginine from the gap between the associating macromolecules. The free energy of the native and the aggregated states also change due to the presence of arginine. The interaction between the hydrophilic groups on the protein surface and arginine (enthalpic) lowers the free energy whereas the exclusion of arginine from the protein surface (entropic) increases the free energy. However, arginine is not a highly excluded cosolvent (as compared to sucrose and other sugars). Therefore, it is expected that the free energy of the native and the aggregated states will decrease in the presence of arginine.

It has been shown that the charge based interactions between peptides are the dominant force for association in aqueous solution.[127] The self-interaction of arginine is also due to the hydrogen bonding between the oppositely charged groups. It raises an interesting question: Why do other charged amino acids not show such strong self-interaction? The possible reason for such a behavior can be the absence of Gdm like side chain in other amino acids. The Gdm side chain can have strong, multiple interactions and forms a planar structure, which helps in Gdm-Gdm stacking and interaction with neighboring molecules in aqueous solution.[74] Lysine has a positively charged side chain but it does not form clusters in solution.[128] Another possible reason for a such a behavior could be the presence of a highly flexible methylene chain which would significantly reduce the stability of the dimers. Recently, Schulund et. al.[118,119] have shown that molecular rigidity is critical for the stability of the dimers formed by arginine in the gas-phase. They showed that 2-(guanidino carbonyl)-1H-pyrrole-5-carboxylate forms more stable clusters than arginine due to the rigidity of the groups holding the terminal carboxylate and Gdm groups. Therefore, making the chain less flexible might lead to better inhibition of protein aggregation. On the basis of the mechanism proposed in this paper,

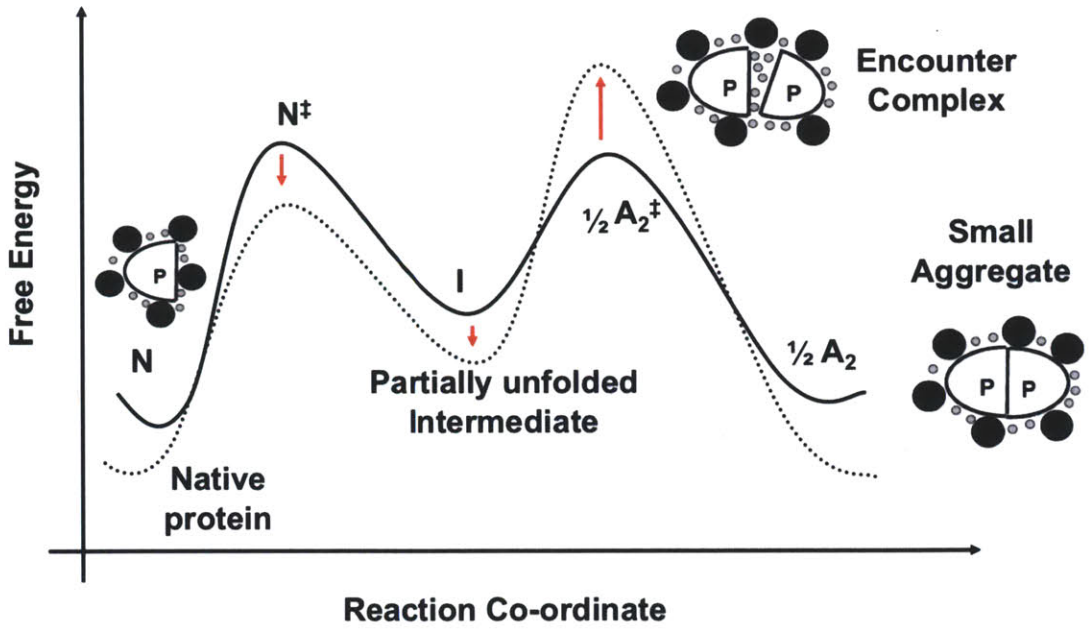

Figure 3-17: **The effect of arginine on the free energy protein states along the refolding/aggregation reaction coordinate. The solid line represents the free energy in the absence of cosolvent and, as a dotted line in the presence of cosolvent. Additive molecules are shown as large black circles and water molecules as small grey circles.**

any change in the cosolvent structure which promotes self-interaction keeping the preferential interaction same would improve the aggregation inhibition ability of the cosolvent.

3.4 Conclusions

This study aimed at understanding the structure and interactions in aqueous arginine solutions. The radial distribution functions obtained from the simulations suggested a tendency for the arginine molecules to self-associate. There are several ways in which arginine molecules can self-associate. The hydrogen bonding between the Gdm and carboxylate group and the stacking of Gdm groups in adjacent arginine molecules are the two important contacts between arginine molecules. The weak hydrogen bonds between arginine and water are replaced by stronger arginine-arginine hydrogen bonds on self-association. The total

number of hydrogen bonds per water molecule remains almost constant, which shows that arginine substitutes well for water in the hydrogen bond network. The population of the first solvation shell (within 0.5Å) around water molecules decreases by a small amount with increases in the arginine concentration. This shows that the water structure is not perturbed significantly by the presence of arginine molecules.

From our results, it can be seen that the self-association of arginine could play an important role in its binding and inhibition of protein aggregation. These results also highlight the role of the carboxylate group in the arginine molecule. Guanidinium salts (e.g. GdmHCl) interact too strongly with the protein and thereby unfold the protein which promotes aggregation. The presence of carboxylate group in arginine would limit the interaction of Gdm group with protein surface. Therefore, the interaction is strong enough to stabilize partially unfolded intermediates but not strong enough to unfold the protein. From the simulation of proteins in an aqueous arginine solution, the aromatic and charged residues were found to interact with arginine via cation-π interactions and hydrogen bonding respectively. These results demonstrate that to explain how cosolvents work, all the possible interactions between the components of a mixed solvent have to be analyzed. The existing mechanisms for the effect of arginine on protein aggregation are analyzed on the basis of the information available from the simulation results and recent experimental preferential binding data of arginine. A detailed analysis of all the existing mechanism shows that none of them by themselves are completely consistent with the simulation results. The mechanism proposed in this paper takes into account the intra-solvent and solvent-protein interactions. The crowding due to the arginine molecules within the local domain of the protein and the interactions due to the Gdm group are mainly responsible for the aggregation suppression by arginine. The crowding effect of arginine is enhanced due to its self-association.

Chapter 4

Arginine as an Eluent in Affinity Chromatography

Monoclonal antibodies (mAbs) are currently used for several diagnostic and therapeutic applications, including treatment of diseases like Alzheimer's and cancer.[7,129–131] Due to the enhanced demand for these antibodies, large-scale manufacturing processes have been developed. However, the presence of several different compounds in the production medium makes purification of antibodies the most critical step in the entire production process.[132–135]

Affinity chromatography is a widely used separation process that makes use of the binding sites on the antibody.[136–138] It involves reversible binding of an antibody to a resin containing a ligand, which interacts with specific binding sites on an antibody. Antibodies are grouped into five classes based on the sequence of their heavy chain constant regions: IgM, IgD, IgG, IgE and IgA. IgG is the most commonly used antibody for therapeutic purposes and is the focus of this paper. Antibodies are further subdivided structurally into two variable domains (Fab) linked to a constant domain (Fc), as shown in Figure 4-1a. The Fab domain binds the antigen, whereas the Fc domain has binding sites for several proteins such as Protein A, Protein G, etc. Typically, Fc domain binding sites are used for purification purposes because Fab is the active part of IgG. Protein A is one of the most widely used natural ligands for the antibody purification process, allowing rapid and selective separation of antibodies.[133]

Protein A is a constituent of the cell wall of *Staphylococcus aureus* and it consists of four homologous regions that can bind the Fc part of IgG from various species. Deisenhofer[8] obtained the crystal structure of the complex formed by a human IgG Fc fragment and a fragment B of Protein A. Recently, Cavallotti and coworkers[139] have performed the molecular modeling of Protein A affinity chromatography to understand the properties of the complex between fragment B of Protein A and the Fc domain of IgG. They have identified the key amino acids involved in the complex formation by replacing the amino acids in Protein A with alanine and estimating the variation in the binding free energy. The authors have also estimated theoretical binding free energy (-9.2 kcal/mol) of the IgG-Protein A complex, which matches well with the experimental binding free energy (-9.8 kcal/mol).[140] Protein A mimetics have been build with improved IgG binding properties.[140] This high affinity leads to effective separation of the antibody from the production medium but makes the elution of the antibody difficult. Highly acidic solutions are required to elute antibodies from Protein A columns. These harsh conditions lead to denaturation and subsequent aggregation of antibodies, which reduces the yield of monomeric antibodies. Several different approaches have been used to counter this problem, e.g., using low affinity ligands[141] or eluents that work under milder conditions, etc.[142]

Arginine has been used in various chromatographic techniques to elute antibodies and other proteins under mildly acidic conditions.[10,11,142,143] Experimental studies have been performed to test arginine derivatives, and other amino acids such as glycine, proline, lysine and histidine; but none of them are as effective as arginine under similar conditions.[142] Guanidinium Hydrochloride was found to be an effective eluent; but the eluted antibodies showed more aggregates than arginine due to the denaturing effect of guanidinium hydrochloride. The ability of arginine to elute antibodies increases with increasing concentration, with almost 100% recovery of IgG4 at 1 M concentration and pH 4.3. Sodium citrate is commonly used as an eluent and gives ~90% recovery at pH of 2.7 and 0.1 M concentration.[10,11,142] At pH 4.3, sodium citrate has ¡ 10% recovery at the 0.1 M concenration. The recovery of antibody from the column is dependent on several factors including the type of feedstock

used, its composition, the type of antibody, and experimental conditions. Therefore, the dramatic change in recovery with increased pH may not be observed for all antibodies at the same experimental conditions. However, the general trend that for citrate solutions, the recovery decreases with increasing pH, will still hold. Arginine solutions are widely used for suppressing aggregation and for enhancing the refolding yield of proteins.[65,66] However, for a molecule to be an effective eluent, it should not only inhibit aggregation of the eluted antibodies but also facilitate dissociation of the IgG-Protein A complex. Although arginine has been widely used as an eluent in various chromatographic techniques, no computational studies have been performed to elucidate the mechanism of arginine elution. Molecular dynamics simulations can give us insight into the interactions between arginine and surface groups of the Fc domain and Protein A involved in complex formation. Furthermore, the effect of cosolute molecules on the free energy of the binding of proteins can also be estimated. In this study, simulations of the Fc domain of IgG, the B fragment of Protein A and the IgG-Protein A complex are performed in the presence of arginine at two different concentrations. The effect of arginine on elution is compared to that of sodium citrate by performing simulations in similar conditions. The results obtained from this study provide a molecular-level understanding of the mechanism of arginine elution, which is useful for the design and search of effective eluents for affinity chromatography.

4.1 Simulation setup

4.1.1 Initial structure

Crystallographic coordinates of the IgG-Protein A complex were obtained from Protein Data Bank (PDB) entry 1FC2 (Figure 4-1c), which includes half of the FC domain of IgG1 and the B domain of Protein A.[8] The region of the antibody involved in the complex formation is usually referred to as the "consensus binding site" (CBS), as this site is involved in the interaction of IgG with several ligands, both natural and synthetic.[8,144–148] In order to avoid the relative bending of the CH2-CH3 sub-domains of the antibody, the whole FC domain of

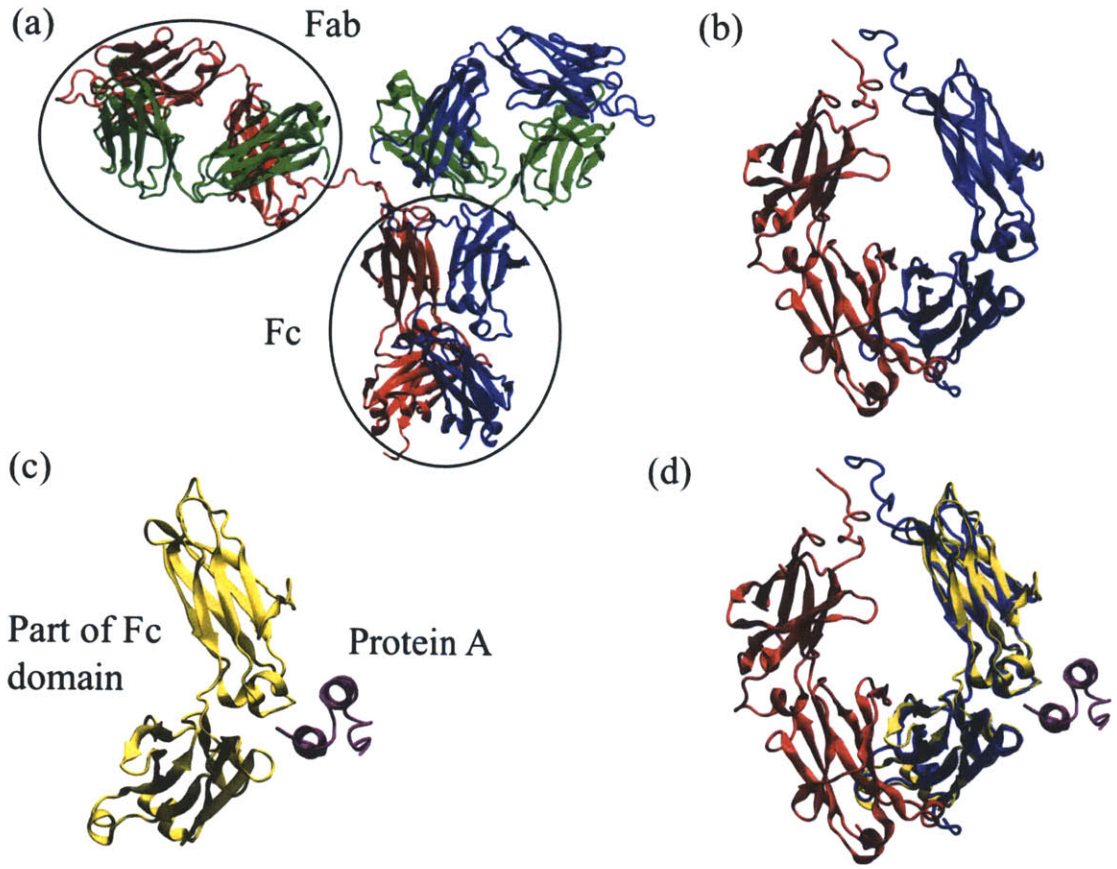

Figure 4-1: **(a) Structure of the IgG obtained from pdb 1HZH.[7] The heavy chains are shown in red and blue and the light chains are shown in green. (b) Side view of the Fc domain. (c) Structure of B fragment of Protein A bound to a part of the Fc domain of IgG obtained from pdb 1FC2.[8] (d) The overlap between the two structures (1FC2 and 1HZH) obtained after transforming the coordinates of 1FC2.**

IgG was reconstructed. This was accomplished by matching the crystallographic structure of the complex (Figure 4-1c) with the X-ray structure of the full antibody obtained from PDB entry 1HZH[7] (Figure 4-1b). The CYS residues forming disulfides bonds in the hinge region of entry 1FC2 were selected and translated to overlap with the corresponding cysteine residues in entry 1HZH. The overlap was optimal in the consensus binding site (CBS), but not as good at both ends of the half-FC domain of IgG; therefore only the coordinates of the CBS binding site of 1FC2 (obtained using the matching) were introduced in the 1HZH entry, to replace the original coordinates. This resulting structure (Figure 4-1d) maintains

the crystallographic geometry of the IgG-Protein A complex and at the same time it is more stable than the original structure. The Fab domains were not included in the system to reduce the system size and thereby the computational cost associated with the simulations. The IgG-Protein A complex, was subsequently minimized in vacuum to remove any unfavorable contacts between the FC residues at the sites where 1FC2 and 1HZH structures were merged.

4.1.2 Simulation details

All simulations were performed using the NAMD[101] package, with the CHARMM22[95] force field. The TIP3P[97] water model was used. All simulations were performed in the NpT ensemble with periodic boundary conditions and full electrostatics computed using the particle mesh Ewald (PME) method,[114] with the grid spacing on the order of 1 Å or less. Pressure was maintained at 1 atm using the Langevin piston method,[115] with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature of 298 K. An integration step of 1 fs was used. The simulation box size was chosen such that the minimum distance between the protein and the edge of the box was 15 Å. All protein residues and cosolute molecules were protonated in order to account for an experimental[66] pH value of 4.3. Cosolute molecules and counter-ions were added to obtain a total neutral charge. In order to set up the simulation systems for various cosolute concentrations, cosolute molecules were randomly placed within the simulation box, and subsequently overlapping water molecules were removed. The system was then equilibrated for 1 ns at constant pressure and temperature. The simulation details are shown in Table 4.1. Snapshots of the simulation box were saved every picosecond.

4.1.3 Force Field Parameters

Structures of cosolutes considered in the present study, citrate and arginine, are shown in Figure 4-2. The force field parameters for arginine were taken from the CHARMM[95] force field with the N terminal and the side chain protonated and C terminal de-protonated. The parameters for the N and C terminal were taken from the CTER and NTER parameters

Table 4.1: **Setup of Simulation systems**

Simulation	System	Cosolute	# of cosolute molecules	# of water molecules	Concentration mol/l	length (ns)
S1	Fc	Arginine	311	31975	0.5	35
S2	Fc	Arginine	1052	20652	2.0	35
S3	Fc	Citrate	307	32143	0.5	35
S4	Fc	Citrate	1021	21429	2.0	35
S5	Protein A	Arginine	61	6258	0.5	50
S6	Protein A	Arginine	204	3999	2.0	50
S7	Protein A	Citrate	60	6331	0.5	50
S8	Protein A	Citrate	198	4148	2.0	50
S9	Fc-Protein A	Arginine	309	31789	0.5	35
S10	Fc-Protein A	Arginine	1047	20554	2.0	35
S11	Fc-Protein A	Citrate	304	31973	0.5	35
S12	Fc-Protein A	Citrate	1016	21324	2.0	35

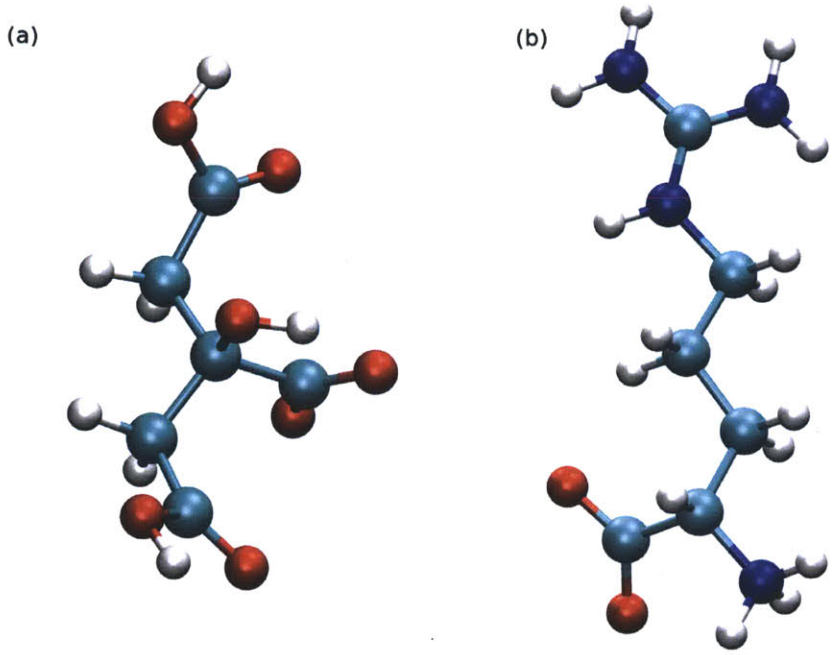

Figure 4-2: **Structures of (a) dihydrogen citrate, and (b) arginine molecule.**

available in CHARMM. For citrate, the first step was to determine determine its protonation state at pH 4.3. The dissociation constants of the carboxylic groups in citric acid at 25°C are 3.128, 4.761, 6.396[149] in a solution at pH 4.3. The percentage population of the possible protonation states is as follows: 5% citrate (Cit^{3-}), 24% hydrogen citrate ($HCit^{2-}$) and 71%

dihydrogen citrate (H_2Cit^-) ions. Thus, all citrate molecules were modeled as the predominant anionic form, H_2Cit^- (Figure 4-2a). Data reported in the literature indicates that the central carboxylic group of citric acid is the first to be ionized, so that dihydrogen citrate is present mainly in its symmetric isomer form.[150] Force field parameters for dihydrogen citrate were developed following the procedure defined by MacKerell and co-workers,[2] according to which parameters are iteratively adjusted to reproduce both intramolecular and intermolecular target data. Intermolecular data are used in the first step of optimization procedure to evaluate partial atomic charges of the molecule. The intermolecular data consists of the interaction energy and the minimum energy distance between the dihydrogen citrate and water molecules. The hydrogen bonding of each polar group (both acceptors and donors) of dihydrogen citrate with a single water molecule was calculated using CHARMM. Target values for intermolecular data were obtained by means of ab initio simulations at the HF/6-31g* level. Although H_2Cit^- is a symmetric molecule, evaluation of its parameters is hampered by the relatively high number of atoms in the molecule. As a consequence, to obtain a reliable first guess for partial atomic charges of H_2Cit^-, the analysis of intermolecular interactions and determination of partial charges was initially limited to the central core of the molecule ($(CH_3)_2COHCOO^-$). Initial values of all other parameters were derived by analogy, using the parameters of the molecules with similar structure and groups.[2,95] Intramolecular data includes geometry, dipole moment, vibrational frequencies of the minimum energy structure and potential energy surfaces (PES) of dihedral angular scans. Vibrational frequencies in CHARMM were computed with the MOLVIB program; PES fitting was carried out using the program recently developed by Guvench and MacKerell.[151] Target values for intramolecular data were obtained through QM computations at the MP2/6-31+g* level of theory. Optimized partial charges, comparison between optimized and QM target data, and potential energy surfaces for key dihedrals are included in the supporting information.

4.2 Methodology

4.2.1 Effect of cosolutes on free energy of binding

The change in binding free energy($\Delta\Delta G_b$) between IgG and Protein A in the presence of cosolute as compared to water can be calculated using following expression:

$$\Delta\Delta G_b = \Delta G_b^{cosolute} - \Delta G_b^{water} \tag{4.1}$$

$$= \Delta\mu_{complex}^{tr} - \Delta\mu_{IgG}^{tr} - \Delta\mu_{protein-A}^{tr} \tag{4.2}$$

$\Delta\mu^{tr}$ is the transfer free energy that captures the change in interaction energy of protein with solvent when it is transferred from pure water to a cosolute solution. In Scatchard notation,[152] water, protein, and cosolute are designated as components 1, 2 and 3 respectively. The transfer free energy ($\Delta\mu_2^{tr}$) is given by:

$$\Delta\mu_2^{tr} = \mu_2^{cosolute} - \mu_2^{water} \tag{4.3}$$

where μ_2 is the chemical potential of protein. If we examine the situation in a reciprocal manner, when a protein is added to a water-cosolute mixture, the chemical potential of the cosolute is disturbed by the protein, $(\partial\mu_3/\partial m_2)_{T,P,m_3}$, and that of the protein by cosolute $(\partial\mu_2/\partial m_3)_{T,P,m_2}$. Here m stands for concentration in terms of molality, and T and P have their usual meaning of temperature and pressure. The total change in chemical potential of protein due to transfer to a cosolute solution is given by the following integral:

$$\Delta\mu_2^{tr} = \int_0^{m_3} \left(\frac{\partial\mu_2}{\partial m_3}\right)_{T,P,m_2} dm_3 \tag{4.4}$$

$$= -\int_0^{m_3} \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \left(\frac{\partial m_3}{\partial m_2}\right)_{T,P,\mu_3} dm_3 \tag{4.5}$$

$$= -\int_0^{m_3} \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \Gamma_{23} dm_3 \tag{4.6}$$

where Γ_{23} is the preferential interaction coefficient. This preferential interaction coefficient is the number of cosolute molecules added to the solution per protein molecule to keep the chemical potential of the cosolute constant. It is also a measure of the excess local concentration around the protein as compared to the bulk cosolute solution. It can be measured experimentally using dialysis/densimetry[79] and vapor-pressure osmometry.[86] The first partial derivative $(\partial\mu_3/\partial m_3)_{T,P,m_2}$ in Equation 4.6 is the variation in the chemical potential of the cosolute as a function of molality, which can be obtained from the experimental activity coefficient or osmolality data.[3] Using the expression for chemical potential as a function of activity and the Gibbs-Duhem relationship to approximate the change in the activity of cosolute in terms of the change in the activity of water:

$$\left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} = RT\frac{\partial ln(a_3)}{\partial m_3} \approx -RT\frac{m_1}{m_3}\frac{\partial ln(a_1)}{\partial m_3} = -\frac{RT}{m_3}\frac{\partial Osm}{\partial m_3} \tag{4.7}$$

where Osm stands for the osmolality of the binary water-cosolute solution. Assuming that preferential interaction coefficient is a linear function of cosolute molality, Equation 4.6 can be approximated as:

$$\Delta\mu_2^{tr} = \int_0^{m_3}\frac{RT}{m_3}\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\left(\frac{\Gamma_{23}}{m_3}\right)m_3 dm_3 \tag{4.8}$$

$$= RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\left(\frac{\Gamma_{23}}{m_3}\right)\int_0^{m_3} dm_3 \tag{4.9}$$

$$= RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\Gamma_{23} \tag{4.10}$$

On substituting the above expression for transfer free energy in Equation 4.2,

$$\Delta\Delta G_b = RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}(\Gamma_{23}^{complex} - \Gamma_{23}^{protein-A} - \Gamma_{23}^{IgG}) \tag{4.11}$$

$$\Delta\Delta G_b = RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\Delta\Gamma_{23} \tag{4.12}$$

Therefore, in order to evaluate change in the free energy of binding, the experimental activity or osmolality data for the cosolute solution and the preferential interaction coefficients for the IgG, Protein A, and the complex in the cosolute solution are required. Lee and Kim[153] have reported the experimental activity of water in the presence of L-Arginine. The water activity data for sodium dihydrogen citrate as a function of salt molality is not available. However, experimental data for potassium dihydrogen citrate are available.[154] It is expected that the behavior of potassium and sodium salt will not be too different. Therefore, the activity coefficient data for KH_2Cit has been used for calculations.

4.3 Results and Discussion

4.3.1 Convergence of Preferential Interaction Coefficient

Preferential Interaction coefficient values as a function of distance from the protein surface are shown in Figure 4-3. The distance beyond which Γ_{23} values are constant is the extent of the local domain around the protein. The extent of local domain is a measure of the size of the region around the protein in which cosolute concentration is different from the bulk concentration. The extent of the local domain is found to be ~6-7 Å in all our simulations. However, for some cases, the Γ_{23} values are not constant but increase with the distance from the protein surface. This is the case of simulations S2 (FC domain in 2.0 M arginine solution), S4 (FC domain in 2.0 M citrate solution), and S12 (complex in 2.0 M citrate solution). One of the three simulations with divergence in Γ_{23} values is analyzed to understand the source of this behavior. Change in the preferential interaction coefficient as a function of distance (r) from the protein is dependent on the number of water and cosolute molecules within the shell of thickness r around the protein. Although the numbers of water (n_1) and cosolute (n_3) molecules at a particular distance from the protein and the bulk density (ρ_{bulk}) are related to each other, the contribution of the variation in each of these variables to the change in

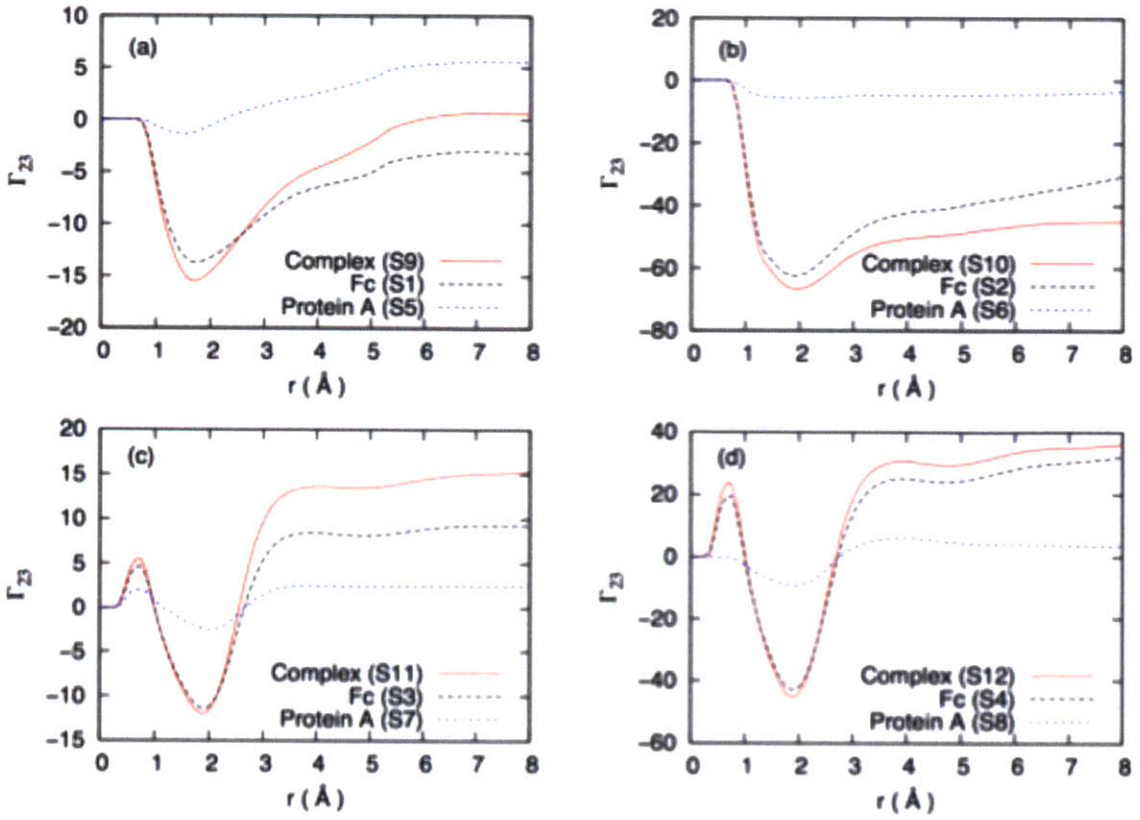

Figure 4-3: **Variation in preferential interaction coefficients as a function of distance from the protein surface for simulations involving Protein A, the Fc domain, and the complex in the (a) 0.5 M arginine, (b) 2 M arginine, (c) 0.5 M sodium citrate, and (d) 2.0 M sodium citrate solution.**

Γ_{23} can be easily estimated by differentiating Equation 2.1.

$$\frac{d\Gamma_{23}}{dr} = \frac{dn_3}{dr} - \rho_{bulk}\frac{dn_1}{dr} - n_1\frac{d\rho_{bulk}}{dr} \tag{4.13}$$

The gradient of Γ_{23} and the contribution of each term on the right-hand side of equation 4.13 for simulation S4 (FC domain in 2.0 M citrate solution) are shown in Figure 4-4. These gradients show large fluctuations within the first solvation shell. For $r > 6$ Å, $d\Gamma_{23}/dr$ is expected to be zero. When the small deviation observed beyond 6 Å is integrated over the distance from the protein leads to a large deviation in the Γ_{23} value. From the inset in Figure 4-4, it can be seen that the gradient of Γ_{23} is $\sim$2 for $6 \leq r \leq 10$. On integrating $d\Gamma_{23}/dr$, the deviation in the Γ_{23} value is 8. The bulk density remains constant beyond 6 Å.

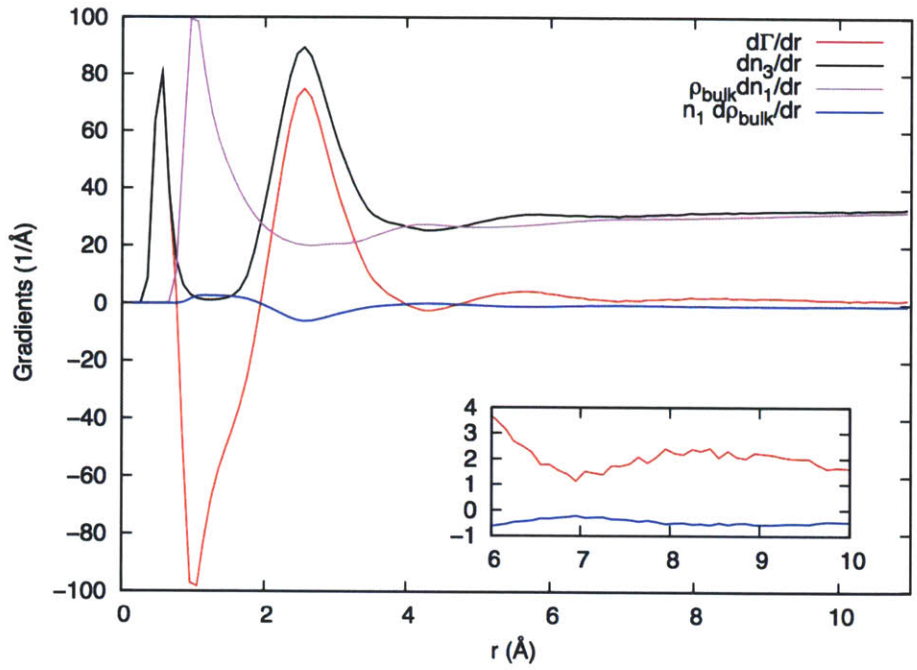

Figure 4-4: **Gradient of Γ_{23} and the contribution of each term on the right-hand side of Equation 4.13 to the gradient for the simulation of Fc in 2.0 M citrate solution.**

Therefore, the main contribution to $d\Gamma_{23}/dr$ is from the difference of the first two terms in Equation 4.13. The source of this deviation is the fluctuations in the distribution of water and cosolute molecules around the protein. Rossky and coworkers[155,156] have analyzed the effect of fluctuation in solvent distribution around a solute on the calculation of volumetric properties of the solution like excess volume and isothermal compressibility that, like Γ_{23}, involve integral truncation at a finite distance from the solute. Due to the finite size of the simulation box, the packing forces lead to oscillations in the solute (protein) - solvent (cosolute and water) distribution functions. Matubayasi and Levy[157] have shown that the solute-water distribution functions have significant long-range oscillations. It is not possible to achieve complete convergence within a feasible computational time because it requires either a large simulation box or a very long (in μs) trajectory. Vagenende et al.[158] have shown that there is no statistical difference between the Γ_{23} values beyond 6 Å. Therefore, for the purposes of the Γ_{23} calculation, it is sufficient that the number of cosolute and

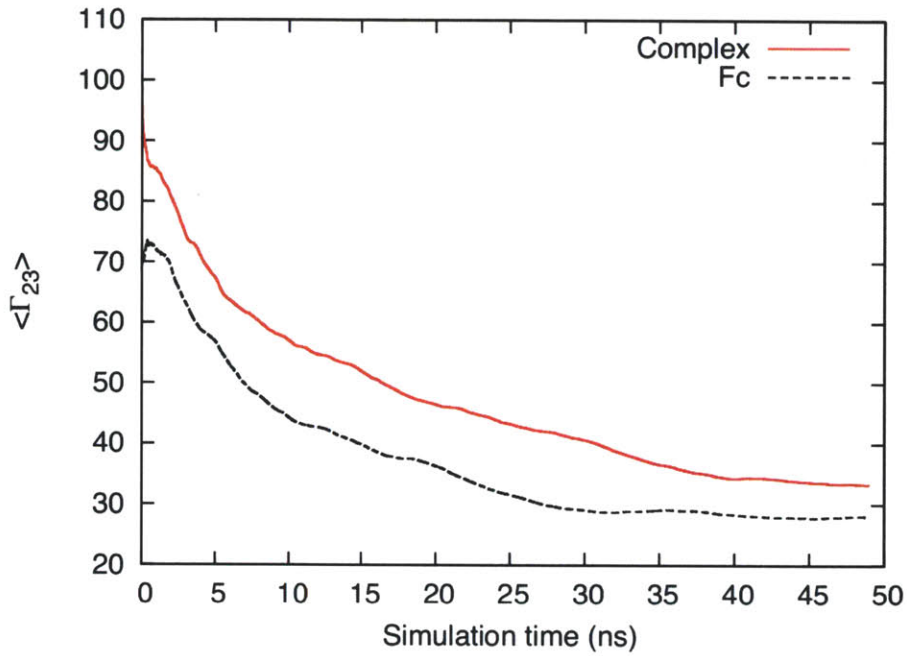

Figure 4-5: **Convergence of Γ_{23}(r = 6 Å) as a function of simulation time for simulations (S4 and S12) involving the complex, and Fc domain in 2.0 M citrate solution.**

water molecules around protein within the local domain and the bulk density of the solution outside the local domain are converged. The convergence of these quantities implies that time-averaged Γ_{23} value calculated at a fixed value of the extent of the local domain (r = 6 Å) will not change with the simulation time. The time-averaged Γ_{23} values for simulations S4 (FC domain in 2.0 M citrate solution) and S12 (complex in 2.0 M citrate solutin) as a function of simulation time are shown in Figure 4-5.

4.3.2 Free energy of binding

Changes in free energy of binding ($\Delta\Delta G_b$) of antibody and Protein A in the presence of arginine and citrate solution are shown in Table 4.2. Arginine increases the free energy of binding, whereas citrate decreases the free energy of binding. The free energy of binding increases with increasing arginine concentration, whereas it decreases with increasing citrate concentration. The trend in $\Delta\Delta G_b$ agrees well with the experimental data for the percentage

Table 4.2: **Change in free energy of binding, $\Delta\Delta G_b$ between IgG and Protein A in the presence of arginine and citrate, and the preferential interaction coefficient (Γ_{23}) values for all the simulated systems. The error bars on the preferential interaction coefficient are ~1 and the error bars on $\Delta\Delta G_b$ are ~1.8 kcal/mol. The experimental data[10,11] for the recovery of the antibody IgG1 at pH 4.3 is also included for comparison.**

Cosolute	Concentration mol/L	Γ_{23} Complex	Γ_{23} Fc	Γ_{23} Protein A	$\Delta\Gamma_{23}$	$\Delta\Delta G_b$ kcal/mol	Recovery % of loaded protein
Arginine	0.5	0	-3	6	-3	1.5	38
	2.0	-47	-37	-3	-7	3.6	77
Citrate	0.5	15	9	3	3	-2.7	¡ 10
	2.0	34	27	3	4	-3.3	

recovery of the loaded antibody.[10,11] Experimental results are found to be sensitive to the pH value. For IgG1, experimental results show that recovery of antibody decreases with increasing pH. For IgG4, a similar experimental trend is observed for both arginine and citrate. The simulations are performed with a fixed ionization state for all the charged residues in protein and cosolute molecules, given a pH of 4.3. Therefore, the simulation results can only provide trends in $\Delta\Delta G_b$ for different cosolutes and for the same cosolute at different concentrations.

From these results, it can be concluded that the arginine facilitates the elution of antibody from the Protein A column by increasing the free energy of binding, whereas citrate enhances the binding of antibody to Protein A. Such results for arginine agree well with the behavior of an IgG synthetic affinity ligand characterized by a dendrimer structure (4 arms) and 4 Arg as terminal residues. The synthetic ligand was discovered through combinatorial libraries, determining its affinity for IgG through direct competition with protein A.[147] A recent work by Fassina and coworkers[159] has confirmed that the Arg residues of the ligand can significantly interact with IgG. Thus, the competitive binding of arginine to IgG, whether as a cosolute or as a key residue in an affinity ligand, explains why its presence in solution hinders the interaction between protein A and IgG. The $\Delta\Gamma_{23}$ values from the Table 4.2 show that for arginine, the $\Delta\Gamma_{23}$ values are less than zero, whereas for citrate $\Delta\Gamma_{23}$ values are greater than zero for both 0.5 and 2.0 M concentration. A $\Delta\Gamma_{23}$ value less than zero implies that

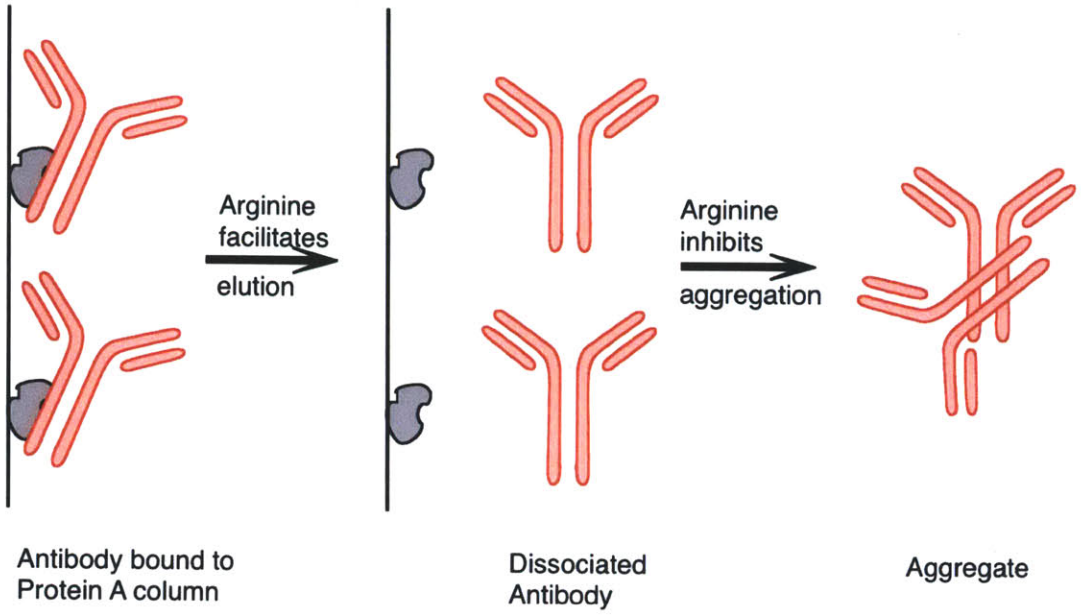

Figure 4-6: **Mechanism by which arginine promotes elution of antibodies from Protein A column.**

the concentration of arginine around the proteins in the dissociated state is higher than in the associated state. For citrate, the concentration around the complex is higher than the concentration around the dissociated Fc and Protein A. The higher concentration around the dissociated proteins implies that the dissociated state is preferred over the associated state. However, this preference is only one of the reasons for the enhanced performance of column chromatography in the presence of arginine and the decrease in the performance in the presence of citrate. A higher concentration around proteins leads to a higher barrier for protein aggregation because there are more cosolute molecules around proteins to crowd out protein-protein interactions. Arginine is extensively used for refolding of proteins and suppression of protein aggregation. The eluent is present in the solution after the antibodies are eluted from the column. In the presence of arginine, unfolded eluted antibodies can refold back to the native state and eluted antibodies do not form aggregates. Therefore, there is a two-step mechanism which explains the role of arginine as an effective eluent. This two-step mechanism is illustrated in Figure 4-6. Citrate has a positive preferential interaction coefficient, which is typically true for denaturants like urea and guanidinium hydrochloride.[3]

Urea and guanidinium hydrochloride enhance aggregation by unfolding the proteins.[160] The Γ_{23} values for citrate are positive and these values increase with increasing concentration. Therefore, it is expected that citrate will not act as an aggregation suppressor but could enhance aggregation by unfolding proteins.

4.3.3 Residues involved in complex formation

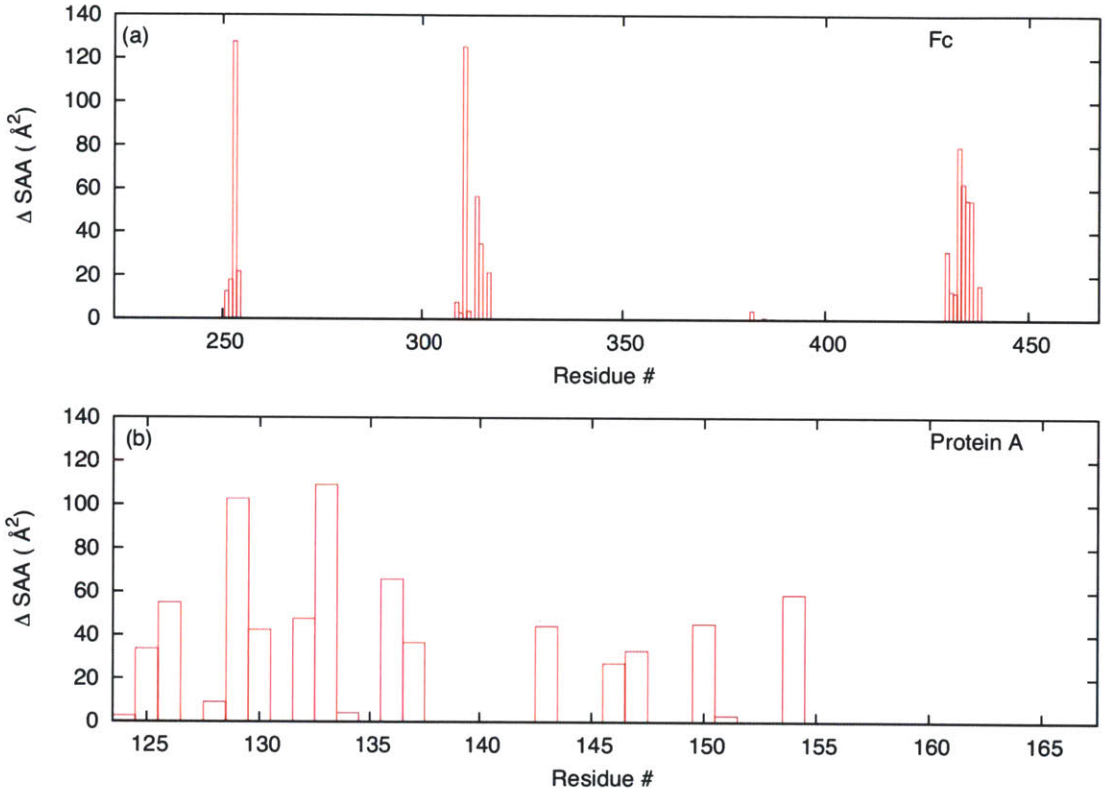

Figure 4-7: **Change in the solvent accessible area (Δ SAA) of the residues in (a) Fc and (b) Protein A when they are in the dissociated state as compared to the associated state.**

In order to understand the interactions between the proteins and cosolutes responsible for the higher concentration of arginine and lower concentration of citrate around dissociated proteins, the nature of residues on the Fc-protein A interface and their interactions with cosolutes needs to be identified. The change in the solvent accessible area (SAA) of the protein residues when they are in the dissociated state as compared to the associated state is

used to identify the residues on the Fc-Protein A interface. The SAA values are determined using the Lee and Richard[161] surface area calculation method implemented in the CHARMM package. SAA is the area of the surface obtained from rolling a probe sphere on the Van der-Waals spheres of a CPK model of the molecule. A probe sphere of radius 1.4 Å is used, which is equivalent to that of a water molecule. The surface areas used to calculate ΔSAA are the average values obtained from the 50 ns simulation of the complex, Fc, and Protein A in 0.5 M arginine solution. The ΔSAA values for residues in Fc and Protein A are shown in Figure 4-7. In the Fc domain, three distinct regions on the surface are involved in complex formation. These three regions are the β-turns present at the junction of CH_2 and CH_3 chains (Figure 4-8a), which form the consensus binding site. In Protein A, several residues are present at the interface. These residues are located at the two helices of the B-fragment of Protein A. The residues with ΔSAA greater than 20 $Å^2$ are shown in Figure 4-8. These residues are the main contributors to the interface area. Li et al.[140] have characterized the Fc-Protein A binding site using the x-ray crystallographic structure of the complex. Although the exact surface area values are different, the key residues involved in the formation of complex identified using ΔSAA are same as those reported by Li et al.[140] In Protein A, the interface is formed by the hydrophobic core Phe132-Tyr133 with polar and charged residues (Asn130, Asn147, Arg146, Glu143, Gln129, Asn125, Lys154, Lys126 and His137) around the core. In the Fc, Ile253 and Leu314 provide the hydrophobic effect and several charged and polar residues (Ser254, Asn434, Gln311, Asp315, Glu430, His433, His435, Lys317) on the three β-turns form salt bridges and hydrogen bonds at the interface. The contribution of different types of amino acids to the solvent accessible area at the interface are listed in Table 4.3; such contributions are expressed as percentages, normalized to the sum of ΔSAA values of all amino acids of the same type in Fc and Protein A located at the interface. From the nature of the residues in Fc and Protein A found at the interface, it can be concluded that the highest proportion of residues is polar (29.7%) and charged (28.4%) with aliphatic (24.9%) and a few aromatic (17%) residues providing the hydrophobic effect.

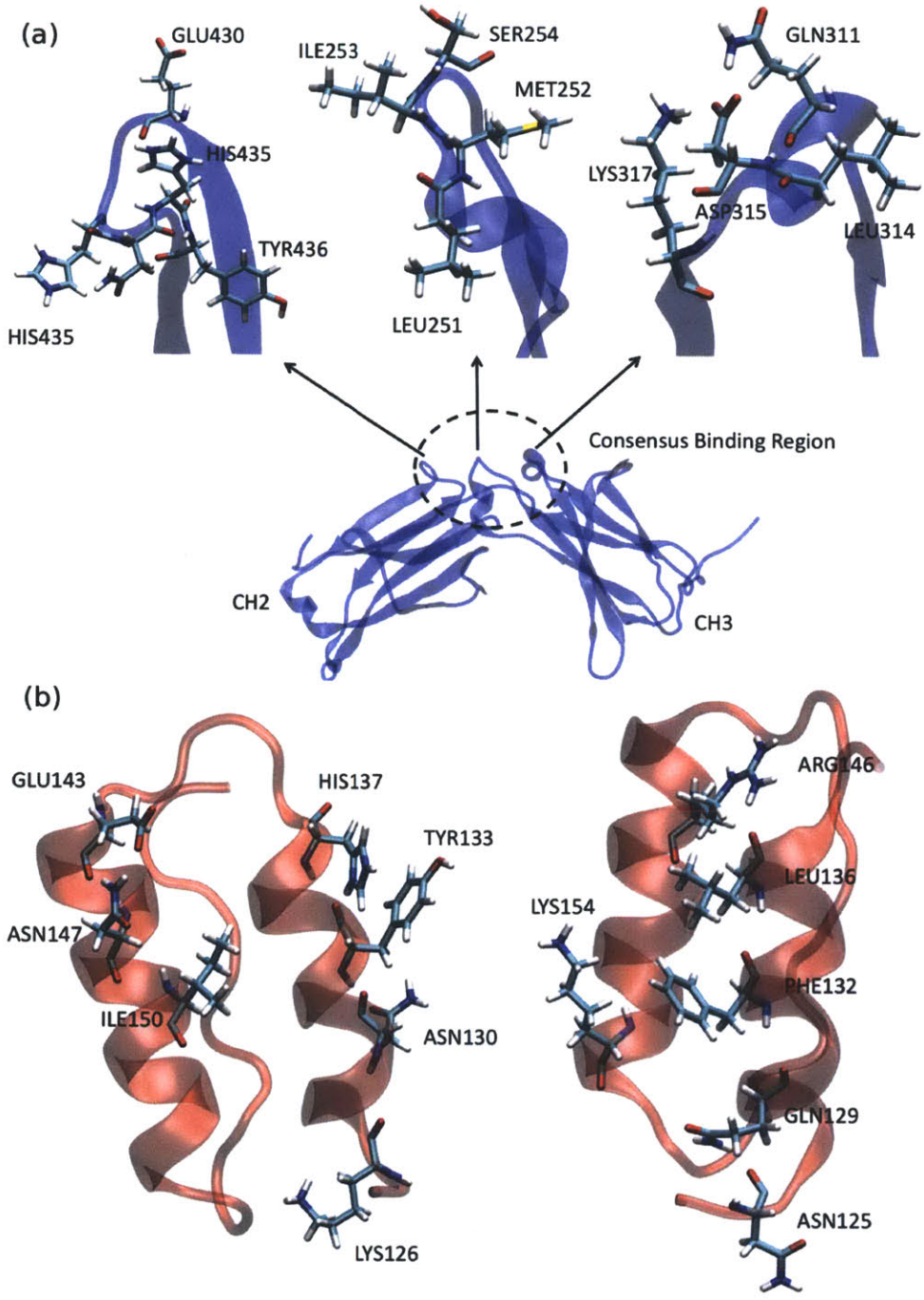

Figure 4-8: **Residues in (a) Fc domain, and (b) Protein A with Δ SAA $\geq 20\text{Å}^2$.**

4.3.4 Contact Coefficient

The total interaction of a cosolute with a protein is conceived as a large number of small interactions involving individual exposed amino acids. In order to calculate the local concen-

Table 4.3: **Contributions of different types of amino acids to the solvent accessible area at the interface between Fc and Protein A when the proteins are in a dissociated state.**

Type of amino acid	Protein A Å^2	Fc Å^2	%
Aliphatic	111	256	24.9
Positively charged	141	79	14.9
Negatively charged	49	150	13.5
Aromatic	196	55	17.0
Polar	223	214	29.7

tration around each amino acid, the following procedure is used. For every cosolute molecule in solution, the nearest amino acid is determined by computing the distance between the center of mass of an cosolute molecule and the Van der Waals surface of the protein, is computed. The protein backbone is defined as NH-CH-CO- as well as the extra proton at the N-terminal and the extra OH at the C-terminal. The definition of the local domain is the same as that used for the estimation of the preferential interaction coefficient. The contact coefficient is the ratio of the local cosolute concentration to the bulk cosolute concentration around each amino acid.[34] The average number of cosolute and water molecules (within the local domain) coordinated with a particular type of amino acid is used to calculate the ratio of cosolute and water molecules around that kind of residue. This local number ratio is normalized with the bulk ratio of cosolute and water molecules to get the contact coefficient.

Contact coefficient values for each type of exposed amino acid in the Fc domain and cosolutes arginine and citrate are shown in Figure 4-9. There is a clear difference between citrate and arginine in terms of their preference for different types of amino acids. Citrate molecules have a higher concentration around aliphatic and positively charged amino acids, whereas their local concentration is low around polar (including backbone), aromatic and negatively charged amino acids. Arginine molecules tend to accumulate near aromatic, polar, and positively, and negatively charged amino acids. The local arginine concentration around aliphatic amino acids is lowest among all exposed amino acids. Due to the presence of the versatile guanidinium group and both negatively and positively charged groups in arginine,

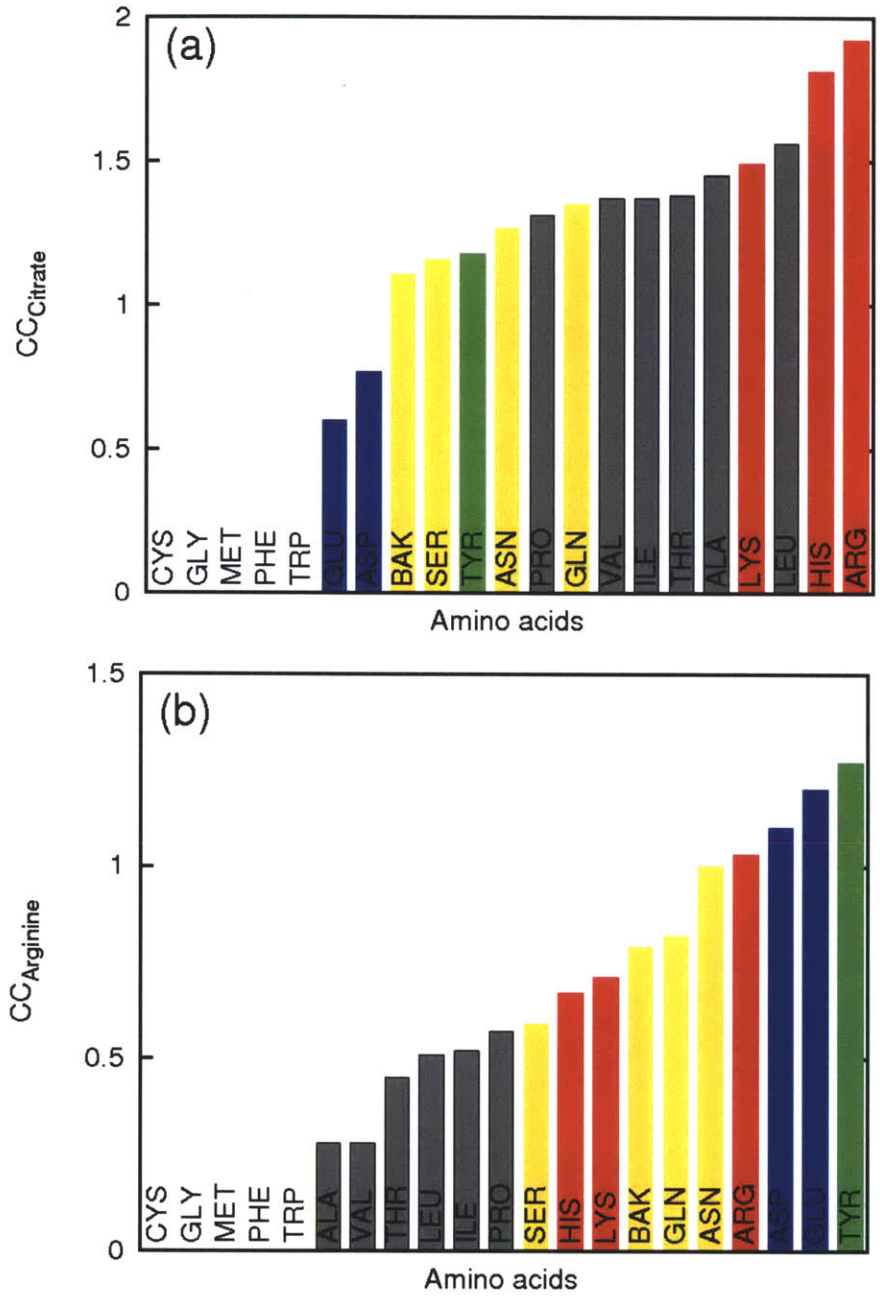

Figure 4-9: **Contact coefficient for each amino acid in presence of (a) 2.0 molar citrate and (b) 2.0 molar arginine solution. The color characterizes the amino acids. Red, positively charged; blue, negatively charged; yellow, polar; gray, aliphatic; green, aromatic. The backbone is denote as BAK (shown in yellow along with polar groups). The residues with negligible exposed solvent-accessible area (Cys, Gly, Met, Phe and Trp) are shown with zero contact coefficient values.**

it can interact with a wider variety of amino acids. Arginine prefers interacting with polar, aromatic, and charged residues due to its ability to hydrogen-bond with both negatively and positively charged amino acids and to form cation-π interactions with aromatic residues. The Fc-Protein A interface is dominated by polar and charged residues. Therefore, the contact coefficient values for arginine and citrate explain the difference in the concentration around proteins in the dissociated state. According to Li et al.,[140] Phe132 and Tyr133 are the key amino acids in Protein A which provide the largest surface area to the interface. Citrate molecules do not interact strongly with aromatic residues. Tyr436 (Fc) and Phe124 (Protein A) do not provide a large surface area to the interface but their solvent accessibility is limited in the complex due to their location at the edge of the interface. All these aromatic residues interact strongly with arginine and are responsible for the higher concentration of arginine at the interface in the dissociated state.

The preference of citrate molecules for aliphatic residues further strengthens the argument that citrate molecules could denature the proteins, especially at high concentration. This preference of citrate molecules for aliphatic residues is similar to that of urea (a commonly used denaturant), which also accumulates at the surface of aliphatic residues.[34]

4.3.5 Conclusions

On the basis of this study, the following conclusions can be drawn about the role of arginine and citrate as eluents in affinity chromatography:

1. Arginine improves the performance of column chromatography because it facilitates the dissociation of IgG-Protein A complex and inhibits aggregation of eluted antibodies.

2. Citrate does not perform well as an eluent at high pH because it strengthens the binding of antibody to Protein A column and could potentially enhance the aggregation of eluted antibodies.

3. The number of arginine molecules around the proteins in the dissociated state is higher than in the associated state, which explains the preference for the dissociated state in

the presence of arginine. The behavior of arginine is in complete contrast with that of citrate, which has a higher concentration around the associated state as compared to the dissociated state.

4. From a molecular perspective, arginine due to its charged N and C terminal and guanidinium group can preferentially interact with charged (both +ve and -ve), aromatic, and polar amino acids, whereas citrate prefers negatively charged, and aliphatic amino acids. The interface between antibody and Protein A is dominated by polar, aromatic, and charged residues.

Chapter 5

Preferential Interaction of Proteins in Aqueous Arginine Solutions

For several decades, the only established method for determining the preferential interaction coefficient has been through approximating Γ_{23} via equilibrium dialysis techniques at constant temperature (T), chemical potential of water (μ_1) and cosolvent (μ_3), while allowing pressure to vary. Such techniques involve placing a protein-cosolute solution under dialysis against the solvent and using various techniques (predominately high precision densimetry) to determine the slight change in cosolute concentration upon reaching dialysis equilibrium.[162] Though widely used, such techniques are troublesome since a tremendous amount of cosolute, protein, and time are required to obtain just a single data point, often with poor precision. Therefore, the relatively new technique of vapor pressure osmometry (VPO), which requires only a relatively minor amount of the solutes and offers somewhat of a rapid turnaround, has come into play in recent years.[3,86] This technique measures the change in the cosolute molality (m_3) with respect to protein molality (m_2) at constant T, P, and water chemical potential (μ_1). However, the VPO technique is not without its disadvantages, namely the accuracy of osmolality readings above 1200 mmol/kg. In these studies, a Wescor vapor pressure osmometer (model 5520) was utilized to obtain preferential interaction coefficient data. Even though the instrument is designed to measure osmolality readings

as high as 3200 mmol/kg, the accuracy of measurements above the last calibration point of 1000 mmol/kg is significantly impaired from extrapolating the calibration curve and from increased system noise at high solute concentrations. Though conducting extra measurements on the solutions containing a high cosolute concentration would improve the accuracy, it was found that the best results were obtained when the osmolality readings were kept below 1200 mmol/kg.[3] Such restrictions greatly limit the concentration range of the cosolute, especially for cosolute compounds comprised of ionic species. Due to these experimental limitations, the preferential interaction coefficient measurement of arginine have not been reported at high cocentrations (beyond 0.7 molal).

Schneider and Trout[3] have used VPO to measure preferential interaction coefficients for arginine using three different model proteins (BSA, lysozyme, and α-Chymotripsinogen A (α-Cgn A)). The authors observed a trend in the preferential interaction of arginine with proteins as a function of concentration. Arginine was found to become increasingly excluded from the protein surface as its concentration increased. This trend was correlated with the protein size and differs from other cosolutes in terms of its non-linear dependence on concentration. The authors suggested that such behavior might be due to the protein surface becoming saturated with arginine, thus causing any additional arginine to be excluded from the protein surface.

In this chapter, we report results from the molecular dynamics simulations of arginine with two common proteins, to determine the cosolute-protein preferential interaction coefficients. These preferential interaction measurements are performed at high arginine concentrations (0.25-2.75 molal) for which experimental data is not reported due to the experimental limitations. These simulations answer questions about the preferential interaction of arginine with proteins raised by the experimental data reported by Schneider and Trout.[3] In particular, the simulations answer the following questions: Why are the preferential interaction coefficients of arginine non-linear? Is the protein surface saturated with arginine? Is the preferential interaction of arginine dependent on the protein size? What is the effect of the self-interaction of arginine on its interaction with the protein?

On the basis of the preferential interaction measurements for the entire experimentally accessible concentration range, a detailed picture of the molecular level interactions between the protein surface and arginine and new insights into the mechanism by which arginine inhibits aggregation are obtained.

5.1 Methodology

Table 5.1: **Setup of Simulation systems**

Simulation	Protein	Molality	Number of ArgHCl molecules	Number of water molecules
L1	Lysozyme	0.25	59	12851
L2	Lysozyme	0.50	89	9784
L3	Lysozyme	1.00	158	8479
L4	Lysozyme	1.50	262	9385
L5	Lysozyme	2.00	315	8750
L6	Lysozyme	2.50	360	8000
C1	α-Cgn A	0.25	57	12296
C2	α-Cgn A	0.50	107	11648
C3	α-Cgn A	1.00	188	10411
C4	α-Cgn A	1.50	253	9488
C5	α-Cgn A	2.00	302	8397
C6	α-Cgn A	2.50	348	7734

All simulations were performed using the NAMD[101] package, with the CHARMM22[95] force field. The TIP3P[97] water model was used. The force field parameters for arginine were taken from the CHARMM force field with the N terminal and the side chain protonated, and C terminal deprotonated. The parameters for the N and C terminal were taken from the CTER and NTER parameters available in CHARMM. Mass densities were compared to experimental density data[3] and the error was found to be $\sim$ 1 %. All simulations were performed in the NpT ensemble with periodic boundary conditions and full electrostatics computed using the particle mesh Ewald (PME) method,[114] with the grid spacing on the order of 1 Å or less. Pressure was maintained at 1 atm using the Langevin piston method,[115] with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature

of 298 K. An integration step of 1 fs was used. The initial size of the periodic rectangular box containing protein and water was set to $(75\ \text{Å})^3$ in all simulations. To set up the simulation systems for various arginine hydrochloride concentrations ranging from 0.25 to 2.5 molal, arginine and chloride ions were randomly placed within the simulation box, and subsequently overlapping water molecules were removed. The system was then equilibriated for 1 ns at constant pressure and temperature. All the simulation systems shown in Table 5.1 were simulated for 50 ns each. A total of 0.6 μs of simulations were performed for the systems comprising of ~40000 atoms.

5.1.1 Orientation of Arginine

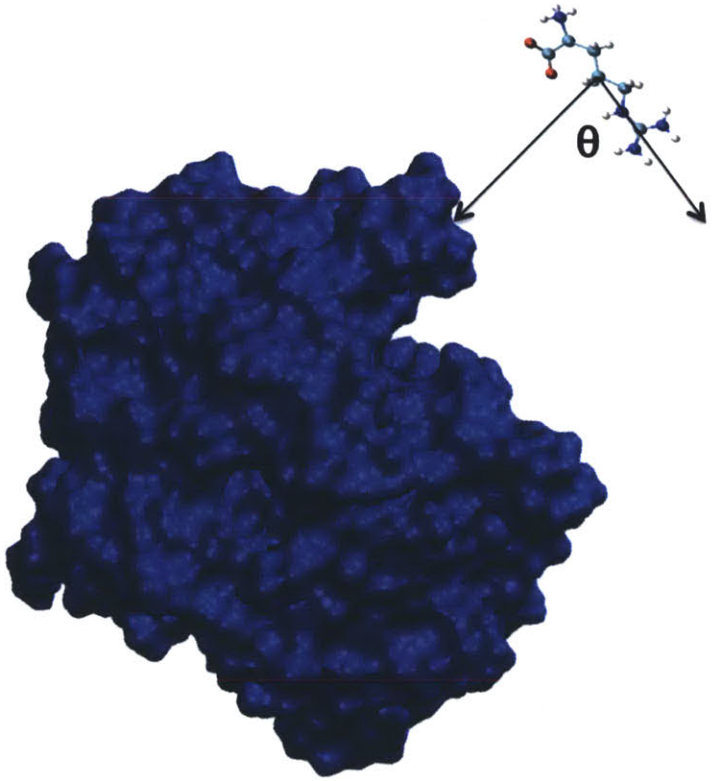

Figure 5-1: **Orientation angle definition relative to the protein surface. Orientation angle is defined as the angle between the vector joining the center of mass of arginine and the guanidinium carbon and the vector normal to the protein surface and passing through the centre of mass of arginine.**

The definition of orientation of arginine with respect to the protein surface is defined in the Figure 5-1. When $\theta < 90°$, the guanidinium group is facing the protein surface and when $\theta > 90°$, the guanidinium group is facing the bulk solution. In order to determine the minimum distance between an arginine molecule and the protein at which the orientation of an arginine is not affected by the protein surface, the average of arginine orientation as a function of distance from the protein surface was computed. The ratio of the populations where guanidinium group in the arginine molecule is facing towards or away from the protein is used to define an orientation free energy, ΔG_θ:

$$\Delta G_\theta = -RTln\left(\frac{\int_0^{90} f_\theta d\theta}{\int_{90}^{180} f_\theta d\theta}\right) \tag{5.1}$$

where f_θ is the normalized orientation probability density, R is the universal gas constant, T is the absolute temperature.The distance beyond which there is no significant statistical preference for either orientation is classified as "bulk".

5.1.2 Local Arginine Concentration

The total interaction of a cosolvent with a protein is conceived as a large number of small interactions involving individual exposed amino acids. In order to calculate the local concentration around each amino acid, the following procedure is used. For every arginine molecule in solution, the nearest amino acid, in terms of the distance of the arginine molecule from the Van der Waals surface of the protein, is computed. Here, the protein backbone is defined as NH-CH-CO- as well as the extra proton at the N-terminal and the extra OH at the C-terminal. The definition of the local domain is the same as that used for the estimation of the preferential interaction coefficient. In order to analyze the binding of arginine to the protein surface, the residence time distribution of arginine molecules within the local domain of the arginine is also estimated.

5.1.3 Survival Functions

Survival functions represent the average number of molecules of a cosolvent with a residence time (within the local domain of protein) longer than t.[158] The residence time of a cosolvent molecules in the local domain is defined as the difference between the exit and entrance time in the local domain. The survival functions for the cosolvent around a protein are computed using the following expression:[163]

$$N_{23}(t) = \frac{1}{N_t}\Sigma_{n=1}^{N_t}\Sigma_j P_j(t_n, t), \tag{5.2}$$

where N_t is the number of simulation time frames, and the function $P_j(t_n, t)$ takes the value of 1 if the j-th cosolvent molecule is within a certain cut off (5 Å) of the protein molecule between time t_n and $t_n + t$, and zero otherwise. Residence times are monitored for all solvent molecules and the survival functions are calculated based on this data for all simulations. For arginine, the sum of two exponential functions gives a good fit with residuals less than 1% for t > 50 ps.

$$N_{23}(t) \approx n_1 e^{-t/\tau_1} + n_2 e^{-t/\tau_2}, \tag{5.3}$$

where n_i and τ_i represent the number and characteristic residence times of cosolvent molecules.

5.2 Results and Discussion

5.2.1 Preferential Interaction Coefficient

From Γ_{23} computations, the extent of the local domain is found to be around 5 Å, which is similar to the value reported previously.[1] Experimental preferential interaction data reported by our laboratory[3] indicates that for lysozyme, arginine is slightly bound ($0 < \Gamma_{23} < 1$) for concentrations less than 0.5 m and then becomes excluded ($\Gamma_{23} < 0$) as the concentration increases. The terms bound and excluded mean that the preferential interaction coefficient is greater than zero and less than zero respectively. For α-Cgn A, it was found that arginine is excluded even at low concentrations. Preferential interaction coefficient values as a func-

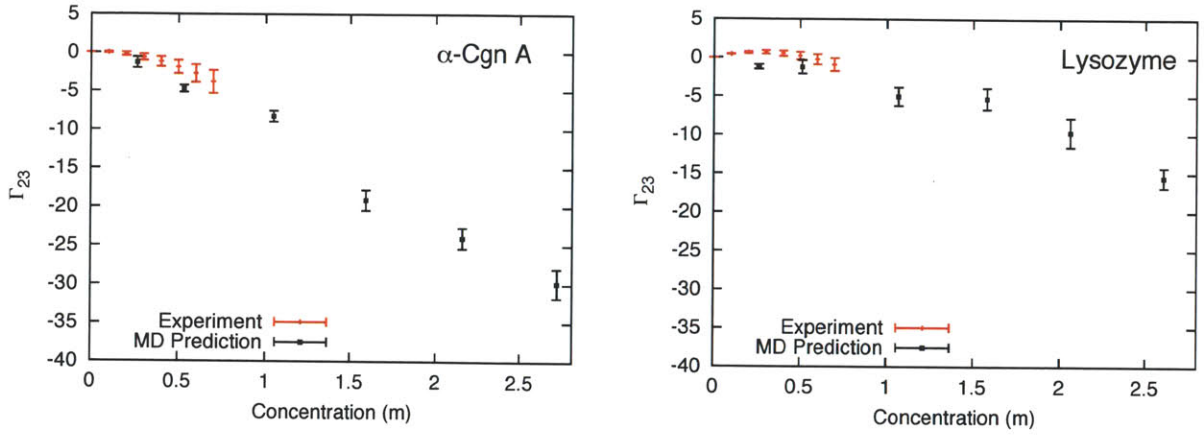

Figure 5-2: **Comparison of arginine preferential binding coefficient values calculated from simulations and experimental trend for α-Chymotripsinogen A and Lysosyme. The experimental data is available up to 0.7 molal.[3] Thus experimental trend is extrapolated to higher concentrations.**

tion of concentration are shown in Figure 5-2. For both α-Cgn A and lysozyme, predicted Γ_{23} values match well with the experimental trend. Γ_{23} values do not follow a linear trend with concentration as compared to other commonly used cosolutes.[3] The computed preferential interaction coefficient varies with the square of the bulk cosolute concentration for low concentrations and tends to be linear at higher concentration. This non-linear variation of the preferential interaction coefficient as a function of concentration can be clearly illustrated by estimating the local-bulk partition coefficient (K_p) as a function of bulk cosolute concentration. K_p is calculated using the following expression:

$$K_p = \frac{(n_3/n_1)^{local}}{(n_3/n_1)^{bulk}} \tag{5.4}$$

The ratio of the number of cosolvent to water molecules in the local and bulk regions can be estimated directly from the simulation. K_p is related to the Γ_{23}, it is proportional to the slope of the Γ_{23} versus concentration plot shown in Figure 5-2. Equations 2.2 and 5.4 can be used to derive the exact relationship between Γ_{23} and K_p.

$$\frac{\Gamma_{23}}{(n_3/n_1)^{bulk}} = (K_p - 1)n_1^{local} \tag{5.5}$$

For cosolutes with a linear relationship between the preferential interaction coefficient and the bulk concentration, the value of K_p is constant. Partition coefficients as a function of bulk argHCl concentration for lysozyme and α-Cgn A are shown in Figure 5-3. It can be

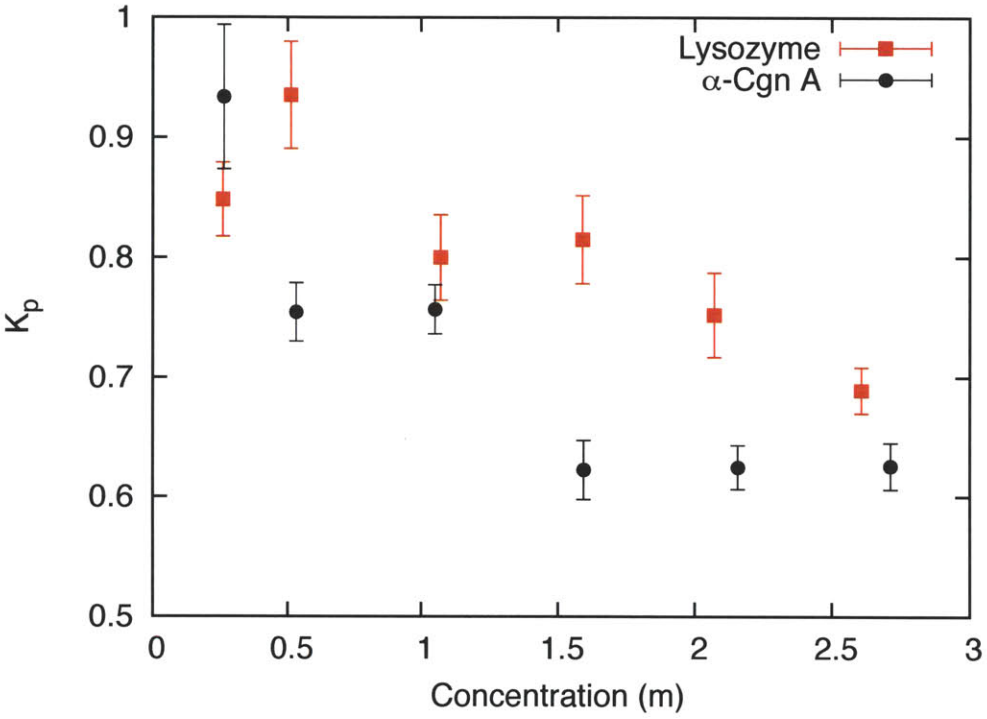

Figure 5-3: **Local-bulk partition coefficient, K_p, as a function of arginine concentration for Lysozyme and α-Cgn A.**

seen that the partition coefficient for both Lysozyme and α-Cgn A is not constant. At high concentrations, the partition coefficient becomes constant for both the proteins, which implies that the preferential interaction coefficient varies linearly with concentration. At intermediate concentrations, K_p varies linearly with concentration. Therefore, Γ_{23} should vary as a quadratic function of concentration as shown below.

$$\Gamma_{23} = (n_3/n_1)^{bulk}(K_p - 1)n_1^{local} = (n_3/n_1)^{bulk}(C * (n_3/n_1)^{bulk} - 1)n_1^{local} \tag{5.6}$$

For lysozyme, the partition coefficient increases initially and then drops to a constant value. This trend is also observed in the experimental and theoretical preferential interaction data

for lysozyme.[3]

On the basis of our previous results on the structure and interaction in aqueous arginine solutions, it can be argued that the self-interaction of arginine is responsible for such a behavior. Simulations of aqueous arginine solutions show a marked tendency for arginine molecules to form clusters via hydrogen bonding and Gdn-Gdn stacking. The non-linear variation of the preferential interaction coefficient can be explained in terms of competition to attract arginine molecules between the bulk arginine solution and the protein surface. The two driving forces, which determine the fraction of molecules that accumulate near the protein surface, are the interaction of arginine with the sites on the protein and the self-interaction of arginine within the bulk solution. The sites on the protein surface get filled as the concentration increases. The extent of arginine clustering or the driving force for arginine interaction with other arginine in the bulk increases with concentration. It changes the fraction of arginine molecules that remain in bulk. Therefore, the partition coefficient decreases with increase in occupancy of arginine binding sites and increase in the interaction of arginine with bulk solution. However, when all the arginine-binding sites get filled and the arginine clustering in the bulk becomes constant, partition coefficient becomes constant and the preferential interaction coefficient varies linearly with concentration. At high concentration, arginine behavior is similar to that of cosolutes like sucrose, glycerol etc, whose exclusion (constant $K_p < 1$) is due to forces totally independent of the chemical nature of the protein surface. Timasheff and coworkers have discussed the behavior of protein-cosolute systems that are chemically inert toward each other.[64] The reasons for the exclusion of such cosolute are the steric excluded volume effect and the perturbation of surface tension of water. This behavior can also be understood in terms of nature of sites on the protein surface. There are three types of sites on protein surface 1) indifferent to cosolute and water ($K_p = 1$) 2) prefer water (constant $K_p < 1$) and 3) prefer cosolute ($K_p > 1$). At high concentration, all the sites of type 3 are occupied. The type 2 sites still repel arginine with a constant $K_p < 1$, and type 1 sites partition arginine with constant $K_p = 1$. Therefore, the sites of type 1 and 2 together partition arginine at a constant average $K_p < 1$. At

low concentrations, protein surface residues, which interact favorably with arginine are not occupied and the number of arginine molecules in the bulk is small. Therefore, the partition coefficient is close to 1. With an increase in argHCl concentration, the opportunities for an arginine molecule to interact with other arginine molecules in the bulk increases along with the decrease in the available arginine binding sites on the protein surface, which leads to a decrease in the partition coefficient. However, at high concentrations, the extent of arginine self-interaction becomes constant and the sites on the protein surface (residues which interact favorably with arginine) are occupied. Therefore, the partition coefficient becomes constant.

In order to validate the hypothesis proposed above, molecular level interactions between all the species in solution are studied to conclude the following:

1. Self-interaction of arginine increases with concentration and becomes constant at high concentrations ($>$ 1.5 m).

2. The occupancy of arginine-binding sites on the protein surface increases with concentration and saturates at high concentration.

In the previous chapter on aqueous arginine solutions, it has been shown that the loss of surface area per arginine molecule in a water box as function of argHCl concentration increases with increasing concentration and becomes constant for concentrations above 1.5 m. Solvent-accessible surface area measurements of the arginine molecules in the bulk aqueous arginine solution show similar trend. The loss of surface area is a good measure of the extent of arginine self-interaction in solution. The hydrogen bonds between arginine molecules were found to be stronger than those between arginine and water. These observations corroborate the conclusion that self-interaction of arginine increases with concentration and becomes constant at high concentrations.

The first step involved in measuring the occupancy of arginine-binding sites on the protein surface is to identify the surface residues, which interact strongly with the arginine molecules and the groups in the arginine molecule, which interact with protein. Dynamics of protein solvation has been characterized using survival functions in the following section to identify the regions on the protein surface, which are solvated by arginine molecules.

5.2.2 Survival Functions

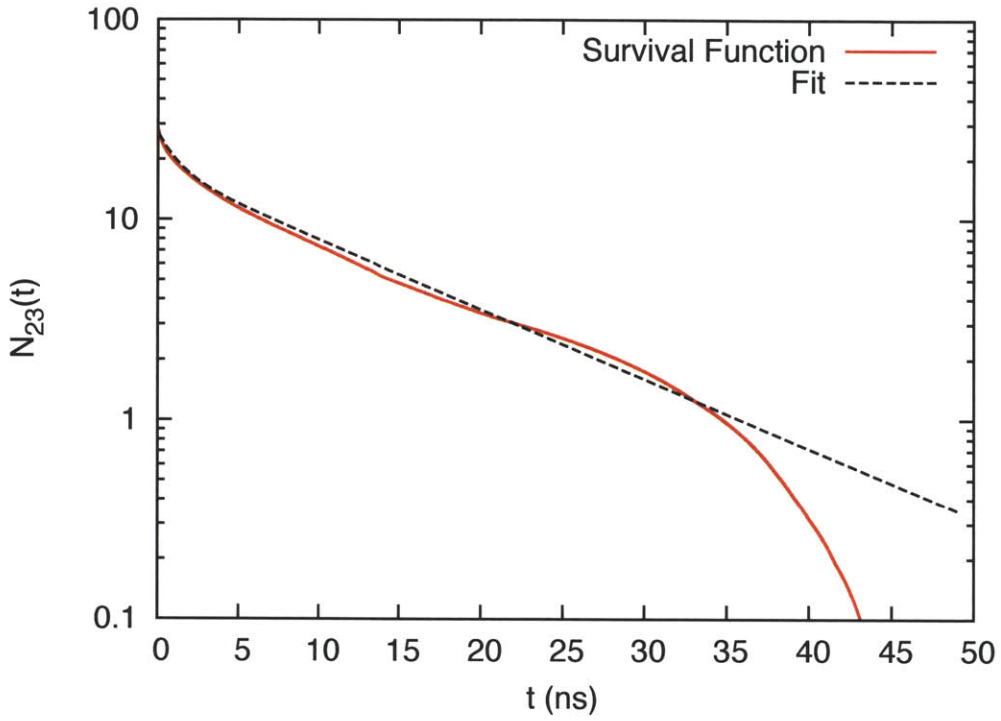

Figure 5-4: **Survival function and the fitted function for simulation L6.**

The survival function data for arginine molecules from simulation L6 and the fitted function are shown in Figure 5-4. Characteristic residence times obtained from fitting equation 6 to the survival function data for all simulations are listed in Table 5.2. The residence time of cosolute molecules is directly related to the average number of molecules in the local domain around protein.

$$n_3^{local} = \Sigma_{i=1}^{n_3} \tau_i / T \tag{5.7}$$

where, T is the simulation time, τ_i is the residence time for the i-th cosolute molecule, n_3 is the total number of cosolute molecules. Therefore, arginine molecules with high residence time values significantly contribute to the preferential interaction coefficient. Arginine molecules can be classified into two different classes based on the characteristic residence times. Class 1 arginine molecules have τ_1 values between 5.8-14.2 ns (lysozyme) and 6.6-8.4 ns (α-Cgn A), and class 2 arginine molecules have τ_2 values between 0.4-2.0 ns (lysozyme) and 0.6-1.3

ns (α-Cgn A). If three exponential function are used to fit the survival function data, then another class of arginine molecules with a very short residence times (< 0.1 ns) is found and that does not improve the fitting significantly. Therefore, only two exponential functions are used.[158]

Table 5.2: **Characteristic residence times and the number of molecules per cosolute class. c is the number of arginine molecules which remain inside the local domain for the entire length of simulation.**

Simulation	n_1	τ_1(ns)	n_2	τ_2(ns)	c
L1	4.7	14.2	4.9	1.2	0
L2	4.7	9.2	7.8	2.0	0
L3	11.0	5.8	6.2	0.8	0
L4	12.9	10.7	8.5	1.4	2
L5	19.7	7.2	4.8	0.4	1
L6	18.7	12.5	9.4	1.2	0
C1	10.4	7.3	3.1	0.7	0
C2	12.6	6.6	7.1	1.3	0
C3	11.1	8.4	11.8	1.3	0
C4	15.8	7.6	10.5	0.8	0
C5	12.6	7.4	18.3	1.3	2
C6	24.5	7.2	16.3	0.6	1

To gain further insight into the type of interactions, which lead to different classes of cosolute molecules, simulation trajectories were analyzed to identify the residues that interact strongly with arginine molecules. Strong interaction implies a long characteristic residence time during the course of the simulations. It was found that arginine molecules belonging to Class 2 interact only with one residue during the course of an event when it moved in and out of the local domain, where as Class 1 molecules interacted with several residues. Class 2 molecules typically interacted with residues, which do not interact strongly with arginine. Class 2 molecules were also found to interact with residues that strongly bind with arginine but are located in a region of the protein surface dominated by non-interacting residues. Class 1 molecules were typically found to be present in a region on the protein surface where these molecules can interact with several different residues via hydrogen bonding, cation- or Gdn-Gdn stacking. In this paper, these regions are referred to as arginine-binding

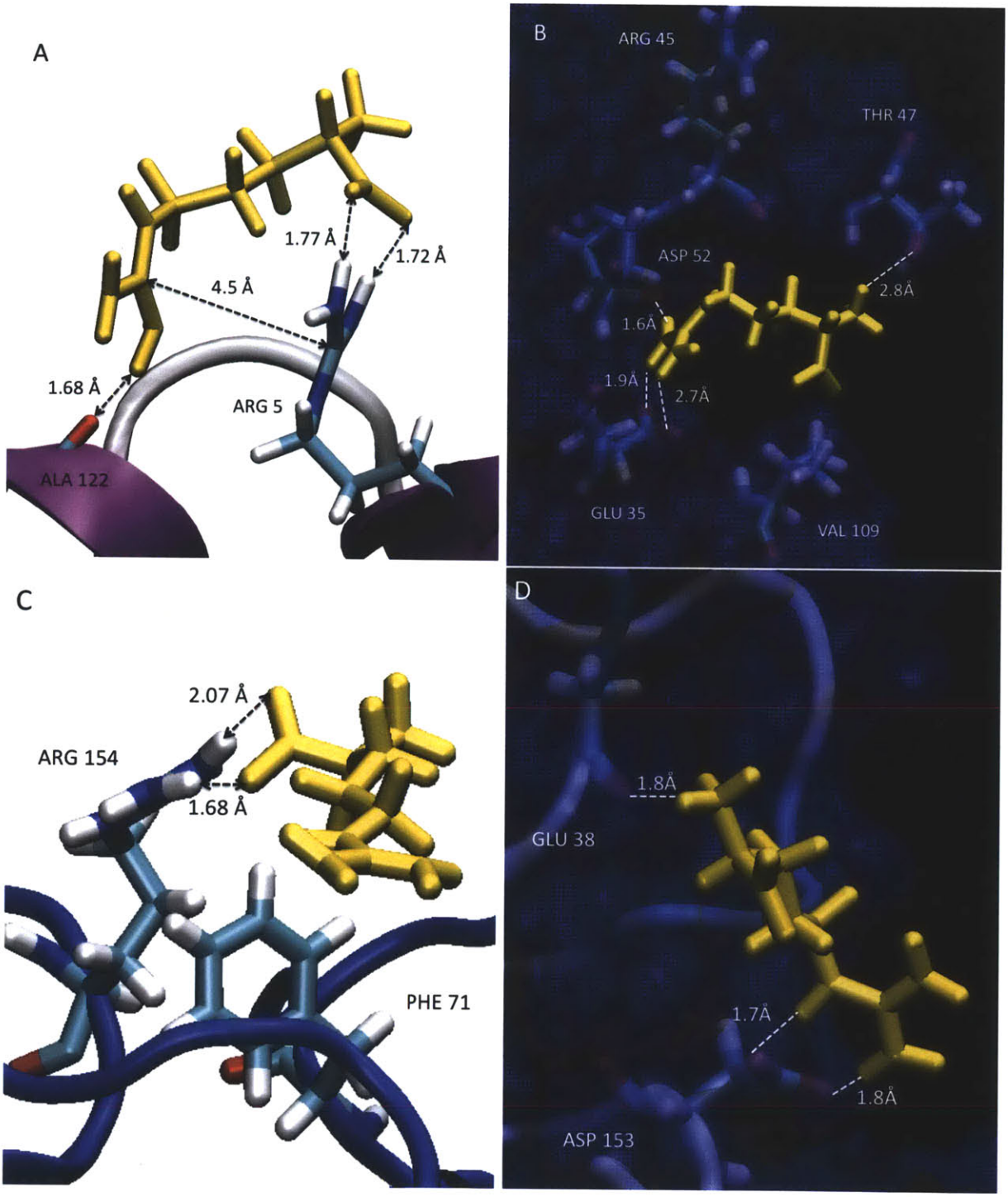

Figure 5-5: **Snapshots of arginine molecules with long residence times and the arginine-binding regions on the protein surface which interact with these molecules in simulation A) L4 B) L5 C) C5 D) C6.**

regions. For simulations (C5, C6, L4, and L5), one or two arginine molecules stayed within the local domain for the entire simulation. The arginine-binding regions in which these molecules reside are shown in Figure 5-5. Figure 5-5a shows the region formed by ARG5 and ALA122, which trapped an arginine molecule for 50 ns in simulation L4. The trapped arginine molecule (carboxylate group) formed h-bonds with ARG5 (Gdn Group) and ALA122 (backbone) residues. Furthermore, the Gdn groups in the trapped arginine molecule (in solution) and ARG5 form a stack on top of each other with a distance of 4.5 Å between them. In the chapter on aqueous arginine solutions, it was shown that the RDF between central carbon atoms in the Gdn groups has a peak between 4-4.5 Å. Figure 5-5b shows an arginine molecule trapped inside a cleft formed by THR47, ARG45, ASP52, GLU35 and VAL109 in simulations L4 and L5. The trapped arginine molecule is shown to form multiple h-bonds with ASP52 (carboxylate group), GLU35 (carboxylate group) and THR47 (backbone). Similarly, Figure 5-5c and d show arginine-binding regions on the surface of α-Cgn A. In Figure 5-5c, the trapped arginine molecule forms h-bonds with ARG154 and interacts with PHE71 via cation- interactions. The nature of the amino acids forming the arginine-binding sites associated with long residence time suggests that arginine interacts preferentially with charged, aromatic side chains and the protein backbone. The presence of arginine molecules with large residence times also raises an interesting question. What is the error introduced in our estimate of preferential interaction coefficients due to the presence of such molecules? These molecules with large residence time are Class 1 molecules that are present in a region on the protein surface, which has several arginine binding groups. It is expected that the probability to find an arginine molecule associated with such sites will be very high. In the current simulation, the probability is one because the site is occupied for entire simulation time. The error introduced in the gamma simulation due to n such molecules with probability p of occupying the site is given by

$$\Delta\Gamma_{23} = n\left(1 + \frac{n_1(r,t)}{n_1 - n_1(r,t)}\right)(1-p) \approx 1.25n(1-p) \tag{5.8}$$

Therefore, it can be seen that the error introduced is negligible (0-0.25), due to high value

of p (assumed to be 0.9) and small value of n (0-2).

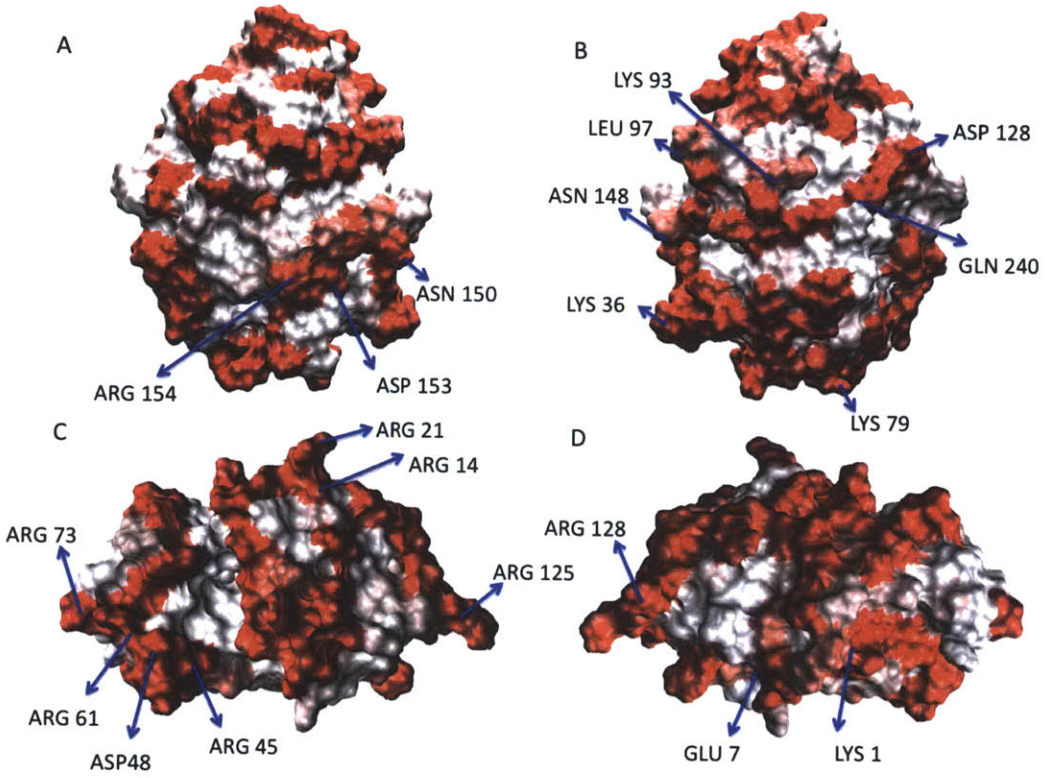

Figure 5-6: **Number of arginine molecules coordinated with residues on the surfaces of α-Cgn A ((A) z-axis pointing out, (B) z-axis pointing in) and Lysozyme ((C) z-axis pointing out, (D) z-axis pointing in) in 2.5 molal ArgHCl (Simulations C6 and L6 respectively). Ten residues with the highest number of arginine molecules coordinated with them are labeled for each protein. The structures of the proteins shown in the figure are obtained after minimizing and heating the protein in water to 298 K.**

In order to support this observation, surface residues with highest number of arginine molecules coordinated with them are identified. Images of Lysozyme and α-Cgn A with surfaces colored according to the number of arginine molecules coordinated with a particular residue are shown in Figure 5-6. Ten residues with the highest number of coordinated molecules are also marked. These residues include Arginine (ARG), Glutamic Acid (GLU), Aspartic Acid (ASP), Lysine (LYS), Asparagine (ASN), and Glutamine (GLN). ARG, GLU, ASP, and LYS contain charged groups, and ASN and GLN both contain -CO-NH2 group, which can form hydrogen bonds with the charged groups in arginine. Lysozyme contains 11

ARG residues and 7 of them appear in the top ten residues. The lysozyme surface interacts strongly (high Γ_{23}) with arginine molecules in solution due to the presence of a large number of exposed charged groups, where as, α-Cgn A has a relatively weaker interaction due to a smaller fraction of exposed charged groups. Arginine also interacts favorably with aromatic residues but the solvent-accessible surface area for aromatic residues is small. Therefore, they do not contribute significantly to the preferential interaction coefficient for proteins in the native state. The large number of charged groups on the surface of lysozyme can also explain why arginine has a slightly positive preferential interaction coefficient for concentrations less than 0.5 m. Arginine molecules in solution are more attracted to the protein surface as compared to the bulk at such low concentrations because the interaction between the charged residues on the lysozyme surface and the arginine molecules is favorable as compared to the interactions in the bulk.

On the basis of the survival function and local concentration analysis, it can be inferred that arginine interacts with charged and aromatic sites on the protein surface. However, the exact groups within the arginine molecule, which are responsible for interactions with charged and aromatic residues on the protein surface are still not identified. The preferred orientation of arginine molecule with respect to the protein surface can be used to identify the group pointing towards the protein surface.

5.2.3 Orientation of Arginine

The orientation free energy for an arginine molecule as a function of distance from the protein is shown in Figure 5-7. It can be seen that for $r \leq 3$ Å, the guanidinium group is preferentially oriented towards the protein surface at all concentrations studied. These arginine molecules are the ones which are influenced by the protein. Furthermore, it can be seen that beyond 3 Å, there is no orientation preference.The distribution of orientation angle observed during the simulations of the arginine molecules within 3 Å of the protein surface is shown in the Figure 5-8. The probability distribution is normalized such that the integral over the entire range of orientations is equal to 1. The random probability density is shown

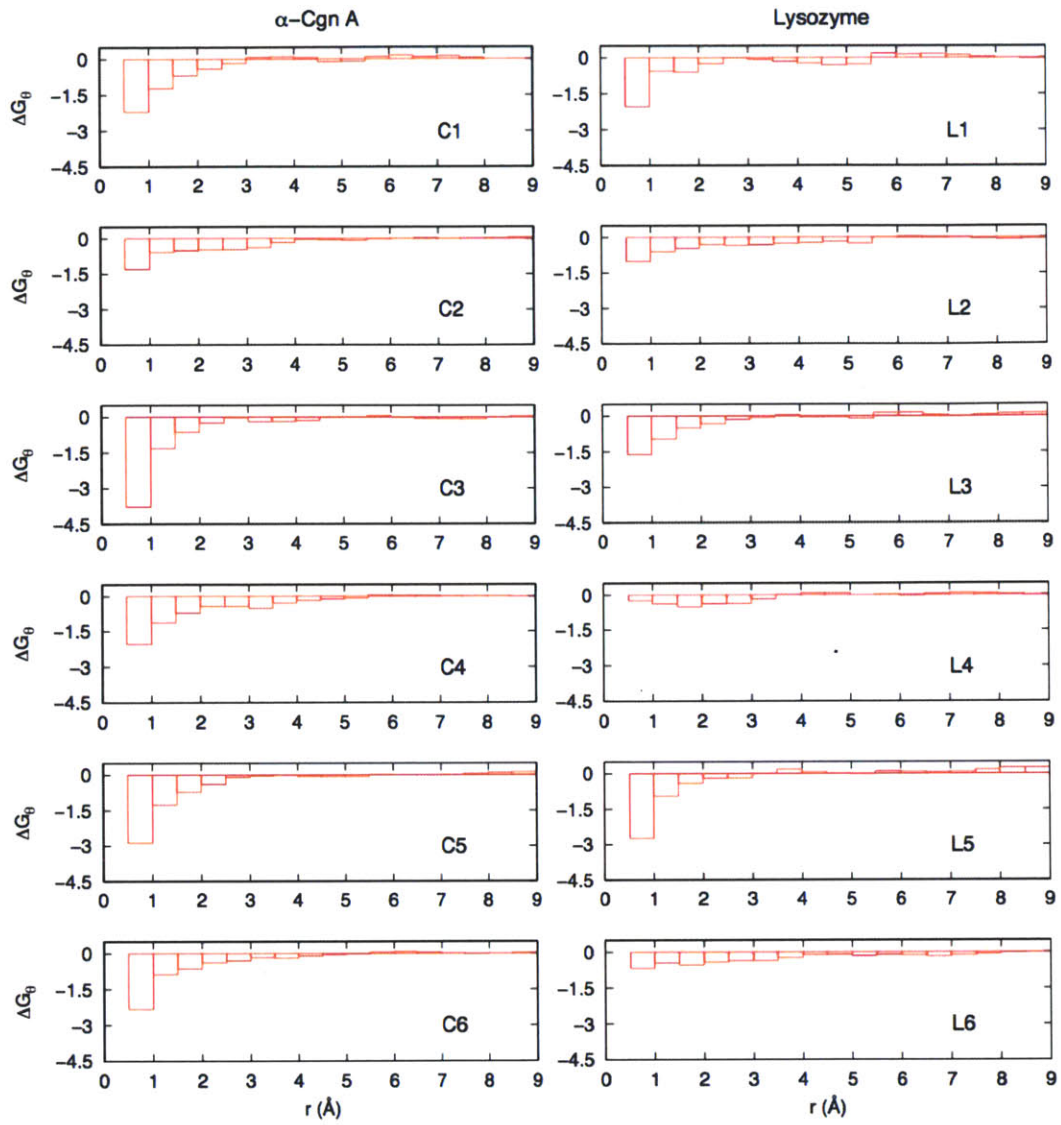

Figure 5-7: **Orientation Free energy of arginine (ΔG_θ in kcal/mol) as a function of distance from the protein (r).**

in the figure for comparison. The random probability distribution is calculated using:

$$f_\theta = \frac{d}{d\theta}\left(\int_0^\theta \int_0^{2\pi} sin(\theta) d\phi d\theta\right) = sin(\theta) \tag{5.9}$$

which gives the orientation of a randomly oriented vector with respect to x-axis. The result after proper normalization is $f_\theta = (\pi/360)sin(\theta)$. Arginine molecules far from the protein surface have a random probability distribution. The deviation from the random distribution

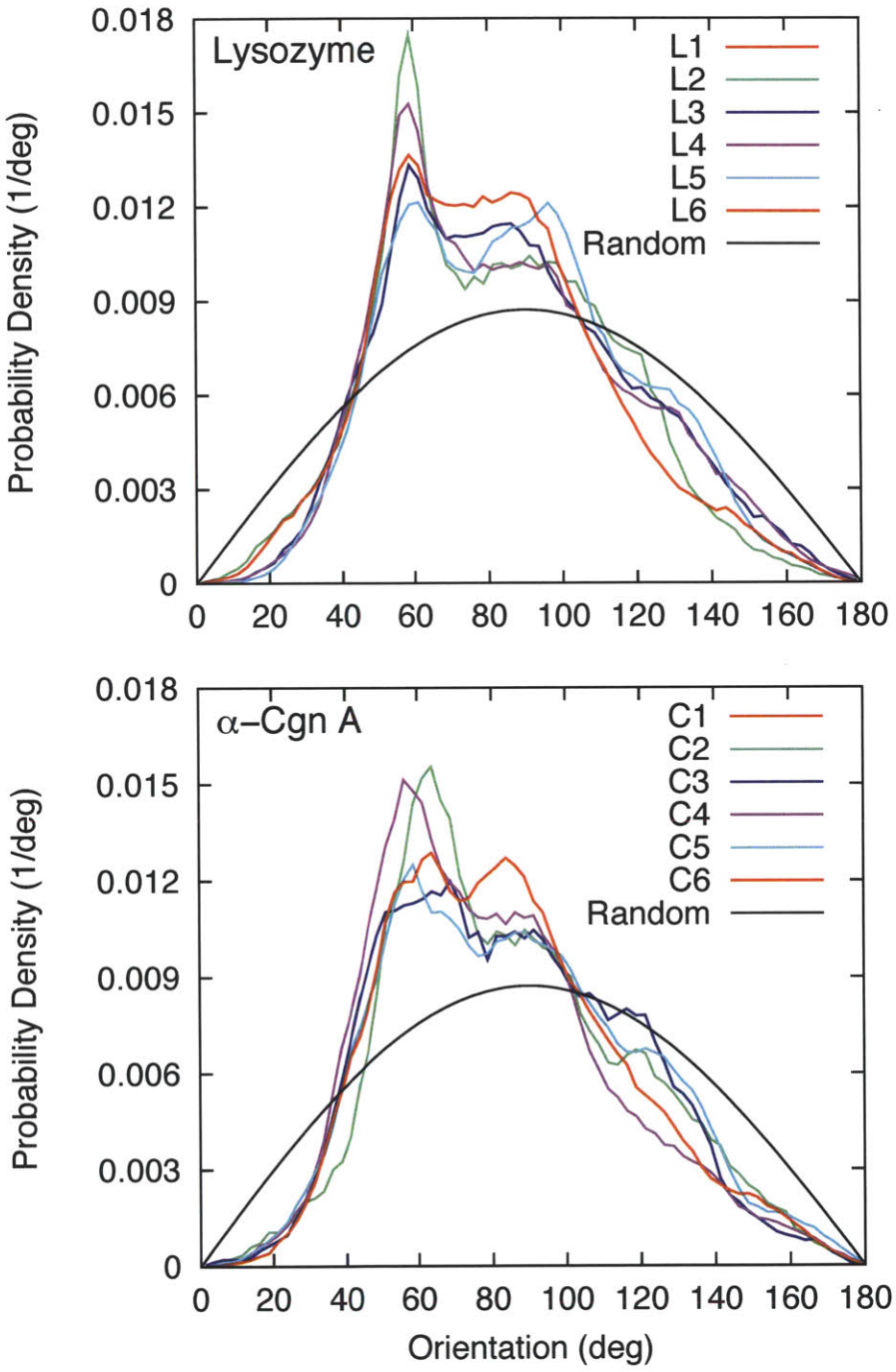

Figure 5-8: **Probability density of arginine orientation relative to protein. The random probability density is shown for comparison.**

in the region $r \leq 3$ Å shows that arginine molecules have a preferential orientation angle in the range of 40-90°, and avoids the orientation angles outside this range. There are slight deviations in the range with changes in arginine concentration. When the concentration of arginine increases, the deviation from a random probability distribution becomes less pronounced as seen from the disappearance of the peak around 60°. The loss of peak can be attributed to the strong self-interaction of arginine at high concentrations, which makes the orientation more random.

A guanidinium group can interact with charged amino acid side chains (and the protein backbone) via hydrogen bonding and aromatic residues via cation- interactions. The number of exposed aromatic residues on the protein surface is small. Therefore, the arginine molecules in solution interact with the protein surface mainly via hydrogen bonding. In order to show that the arginine-binding sites on the protein surface are saturated at high concentration, the number of hydrogen bonds formed between arginine molecules and the protein are measured.

5.3 Hydrogen bonds with surface residues

The number of hydrogen bonds between the protein surface and arginine molecules in solution as a function of bulk argHCl concentration is shown in Figure 5-9. The number of hydrogen bonds between the protein and arginine molecules were calculated using the standard CHARMM tools with a cut off of 3.0 Å between the hydrogen atom and the acceptor atom and a cutoff angle of 30°. The number of hydrogen bonds increases with increasing cosolute concentration. At higher concentrations (> 1.5 m), the number of hydrogen bonds approaches a constant value between 35-40 for both lysozyme and α-Cgn A. On the basis of this observation, it can be concluded that the arginine-binding sites get saturated at the high concentrations. This also verifies our hypothesis proposed to explain the non-linear exclusion of arginine from the protein surface.

Lysozyme (a 14.3 kDa protein with 129 residues) forms the same number of hydrogen bonds as α-Cgn A (a 25.7 kDa protein with 245 residues). This observation implies that

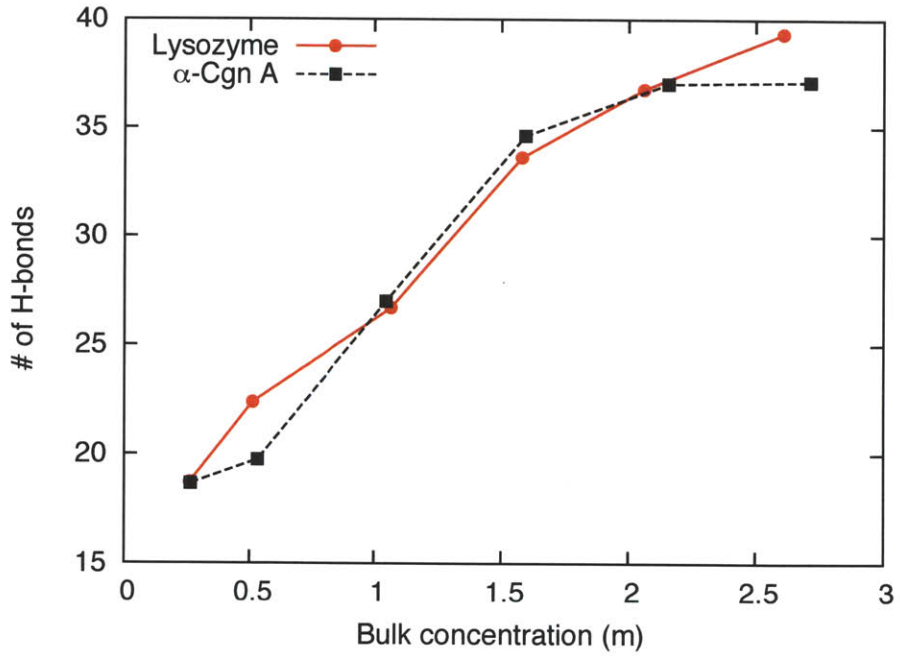

Figure 5-9: **Number of hydrogen bonds between protein and arginine as a function of bulk argHCl concentration for Lysozyme and α-Cgn A.**

lysozyme and α-Cgn A have the same number of surface residues, which can h-bond with arginine, even though α-Cgn A is a much larger protein. The fraction of titratable residues in lysozyme (0.21) is larger than that in α-Cgn A (0.14). Therefore, the preferential interaction coefficient does not depend on the protein size but on the nature of the protein surface.

5.4 Conclusions

The results reported in this paper show a molecular level picture of the interaction of arginine with proteins in the native state and insight into the molecular level mechanisms responsible for the exclusion of arginine from the protein surface. The following conclusions can be made about the variation of the preferential interaction coefficient as a function of bulk concentration. 1) At low concentrations (< 0.5 m), surface residues, which interact favorably with arginine, are not completely coordinated with arginine, and the extent of self-interaction of arginine is low, which leads to Γ_{23} values close to zero. 2) At intermediate concentrations

(0.5−1.5 m), the extent of self-interaction increases with increasing concentration, which draws arginine from the protein surface into the bulk. Therefore, the partition coefficient decreases with increasing concentration, and the preferential interaction coefficient varies as a square of the concentration. 3) At high concentration (> 1.5 m), the extent of self-interaction becomes constant which also makes the partition coefficient constant and therefore, Γ_{23} varies linearly with concentration.

From the results reported in this paper, it can be seen that preferential interaction is a function of the nature of the protein surface. For lysozyme, the protein surface is comprised of a large number of charged groups, which interact favorably with arginine. Therefore, the value of the preferential interaction coefficient for lysozyme is higher than the value for α-Cgn A.

In this chapter, we have shown that self-interaction of arginine also leads to its exclusion from the protein surface. According to the "Neutral-Crowder" theory,[42] size and concentration of the crowder are directly proportional to its ability to inhibit protein-protein association reactions. However, as the size of the cosolute increases, its preferential interaction coefficient goes down due to the excluded volume effect. In the case of arginine, these two opposing effects of arginine self-interaction (i.e. decrease in preferential interaction coefficient and increase in effective size) on inhibition of protein aggregation tend to balance each other out. Experimental data on the effect of arginine on aggregation kinetics, protein and small molecule solubility, and refolding yield show a plateau in the performance of arginine above 1-1.5 mol/l concentration.[164–166] In this chapter, we have shown that effective size of the arginine and the local concentration of arginine around proteins also reach plateau around the same concentration.

The current study supports the idea that in order to study the effect cosolutes have on protein stability; interactions between all the components of the protein-mixture should be investigated. Molecular dynamics simulations, along with preferential interaction theory, provide a framework for the molecular-level understanding of the interactions between the chemical species in the protein-mixture.

Chapter 6

Arginine and the Hofmeister Series

The molecular dynamics simulations of aqueous arginine solutions reported in Chapter 3 reveal self-association of arginine. These simulations show that the guanidinium functional group does indeed interact with the surface of a protein molecule in a fashion similar to free guanidinium; however, this functional group also forms attractive interactions with other arginine molecules. This self-association of arginine molecules likely prevents arginine from binding too strongly to the surface of protein molecules, preventing it from denaturing the protein but at the same time preventing protein-protein interactions. This self-association also explains the unique trend observed for preferential interaction coefficient measurements (i.e. slightly bound at low concentrations but highly excluded at high concentrations) and explains why the arginine effect diminishes at high concentrations. (Chapter 4) Therefore, we were motivated to explore other arginine salt forms, not only with the aim of discovering a better performing excipient, but to further our understanding of this mechanism and how guanidinium-anion interactions may contribute to the observed behavior.

The Hofmeister Series, first published in 1888, is an empirical ranking of how various ions influence protein solubility (i.e. ions to the left have a greater ability for precipitating proteins), which in turn is related to the thermodynamic stability of the native state.[71] The series has been studied extensively since the time of Hofmeister and serves as the foundation for selecting salts that will influence protein solubility, crystallization, denaturation,

and aggregation, though the mechanism for this behavior is not well understood.[167] It has been clearly demonstrated that the choice of anion has a more pronounced effect on these properties than the choice of cation. Furthermore, chloride is often placed in the middle of the series, thus other arginine salt forms are likely to exhibit a drastically different behavior, both positively and negatively, in regards to protein stability.[36,168] To the best of the authors' knowledge, the only other form of arginine studied in this context has been $ArgH(SO_4)_{1/2}$, though the extent of those studies is limited. Therefore, it is clear that little attention has been devoted towards other salt forms besides chloride. The chloride form has been studied so exclusively that ArgHCl is often referred to simply as arginine.[169,170] In a recently published review, Lange and Rudolph[171] (Rudolph being one of the researchers who discovered the arginine effect) postulated, based on existing data, about how other arginine salt forms would behave as excipients and also commented on the refolding behavior of the sulfate and phosphate salt forms, stating that those two forms of arginine perform poorly as refolding excipients even though sulfate and phosphate often provide conformational stabilization. However, no citations or experimental results were provided to support those claims. Even though sulfate and phosphate may reduce the refolding abilities of arginine, it was found that they enhance its aggregation suppression ability when α-Chymotripsinogen A is aggregated under accelerated conditions.[9] Moreover, switching to a thiocyanate salt induces rapid aggregation under similar conditions rather than enhancing aggregation suppression, opposite of what was hypothesized.[9] However, testing on other proteins is required before one could generalize these observations.

Bonner, using osmotic coefficient data and Raman spectra, gave valuable insight into the hydrogen bond interactions between guanidinium and halide ions over three decades ago.[172] Bonner showed that guanidinium and fluoride form a strong hydrogen bond, possibly even a partial H-F bond, in solution and this interaction is weaker for other halides in the order of $Cl^- > Br^- > I^-$, due to ion size and charge density. At the time, these results were used to speculate on the hydrogen bond formation between guanidinium and proteins in an attempt to explain the denaturing effect of guanidinium. Mason and coworkers[57] revealed that the

structure of the poorly hydrated guanidinium molecule allows the positively charged ions to associate, with the planar molecules stacked on top of each other. These results contradict long held beliefs about how certain cosolutes influence the structure of water (solutes have long been labeled either a "chaotrope" or a "kosmotrope" because of this water structure viewpoint) and how the perturbed water structure influences protein stability.[36] Their results are compelling and have strong implications for explaining the arginine mechanism, not only with how arginine will interact with various counterions.

Experimental data[9] for the relative aggregation rate suppression in presence of various arginine salts shows that at the same concentration, ArgH(Citrate)$_{1/2}$ and ArgH(H_2PO_4) slow the rate by a factor of approximately 3 and 8, respectively. On the other hand, ArgHSCN at 150 mM increases the rate by a factor of 6. Other salts such as Acetate and fluoride are essentially as effective as chloride. The thermodynamic stability of aCgn in the presence of the arginine salts shows that the anions have a pronounced effect on the thermodynamic stability of the protein and they correspond directly to the ordering of the relative rate constant values. ArgH(H_2PO_4), ArgH(SO_4)$_{1/2}$ and ArgH(Citrate)$_{1/2}$ increase the melting temperature of the protein at a rate of 6.3, 5.4, and 3.3°C*M^{-1}, respectively. Clearly, these salts are able to stabilize the native structure of the protein and inhibit the formation of partially unfolded species. ArgH(Acetate), ArgHF, and ArgHCl exhibit no thermodynamic stabilization, which is consistent with previous reports for ArgHCl. ArgHSCN lower the melting temperature significantly, both with a rate greater than 20°C*M^{-1}. This denaturing effect is the cause of the increase the rate of aggregation at all concentrations. These results, along with aggregation suppression data, support the claim that Arginine cation acts as an association suppressor and shows that the choice of anion will either enhance or counteract this aggregation suppression by influencing the stability of the native structure.

In this chapter, we studied the influence a variety of arginine salt forms have on the interaction between arginine and α-chymotrypsinogen A. Trends in the preferential interaction coefficient, as determined by MD simulations (and verified experimentally[9]) show a key difference between the anions that enhance aggregation suppression and the anions that di-

minish it. Osmotic second virial coefficient measurements, as determined by MD simulations and VPO, give a clear indication for the observed behavior and support previous reports of ion-ion interactions. Phosphate, sulfate, and citrate seem to interact favorably with the guanidinium moiety, with phosphate interacting quite strongly, which reduces the solubility of arginine considerably. These ion-ion interactions seem to limit interactions with protein molecules, while the anions that have a weaker interaction with guanidinium allow arginine molecules to self-associate and the counterion to freely interact with the protein surface, with iodide and thiocyanate forming a very strong interaction, leading to denaturation. These results not only help to elucidate the arginine mechanism, they also have large implications for interpreting the mechanism behind the Hofmeister Series (i.e. ion-ion interactions and other ion specific behaviors are important considerations and the mechanism cannot be completely generalized). In addition, the results will help to improve the production and formulation of protein therapeutics and aid in the development of novel excipients due to a stabilization method not yet considered.

6.1 Simulation Setup

Molecular dynamic simulations of protein-cosolute solutions and cosolute only solutions were performed to give a molecular perspective on how arginine interacts with aCgn, how arginine interacts with itself, and how the counterion influences these interactions. All simulations were performed using the NAMD[101] package, with the CHARMM22[95] force field. The TIP3P water model was used.[97] The force field parameters for arginine were taken from the CHARMM force field with the N terminal and the side chain protonated, and C terminal deprotonated. The parameters for the N and C termini were taken from the CTER and NTER parameters available in CHARMM. Force field parameters for sulfate, citrate and thiocyanate ions were taken from the literature.[173,174] Force field parameters for acetate, chloride and phosphate ions were available in CHARMM. Fluoride, bromide and iodide were not computationally studied due to the absence of reliable force field parameters in CHARMM. For the purposes of understanding the ordering of anions in the Hofmeister series, the two opposite cases of sulfate and thiocyanate, with few anions in between the two extremes, were studied. It has been shown that non-polarizable force fields do not capture the polarization of water in the vicinity of the highly charged ions, which enhances the ion-pairing in solution.[175] In the current study, we have used non-polarizable force fields. Therefore, for the results reported in this study, the order of the strength of ion-pairing for different salts will be exact but the simulation results may not match the experimental activity data for real solutions. All simulations were performed in the NpT ensemble with periodic boundary conditions and full electrostatics computed using the particle mesh Ewald (PME) method, with the grid spacing on the order of 1 Å or less.[114] Pressure was maintained at 1 atm using the Langevin piston method, with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature of 298 K.[115] An integration step of 1 fs was used. The initial size of the periodic rectangular box containing protein and water was set to (75 Å)^3 in all simulations. The box size for simulations involving aqueous solution of arginine salts was set to (50 Å)^3. Arginine and counter ions were randomly placed (without overlap with protein) within the simulation box, and subsequently overlapping water molecules were

removed. The system was then equilibrated for 1 ns at constant pressure and temperature. The simulation time for each aqueous arginine salt solution (without protein) was 100 ns and for simulations involving protein in presence of arginine salts was 50 ns. A total of ~1 microsecond of simulation time was used. The simulation snapshots were saved every picoseconds. See Table 6.1 for a summary of all of the simulation systems.

Table 6.1: **Setup of simulation systems.**

Protein	Arg Salt	C(mol/L)	No. of arginine	No. of counter-ions	Length of simulation
α-Cgn A	$(SO_4)_{1/2}$	0.5	90	48	50ns
α-Cgn A	$(Citrate)_{1/2}$	0.5	90	48	50ns
α-Cgn A	(H_2PO_4)	0.2	36	42	50ns
α-Cgn A	Acetate	0.5	90	96	50ns
α-Cgn A	Cl	0.5	90	96	50ns
α-Cgn A	SCN	0.5	90	96	50ns
	$(SO_4)_{1/2}$	0.5	34	17	100ns
	$(Citrate)_{1/2}$	0.5	34	17	100ns
	(H_2PO_4)	0.5	34	34	100ns
	Acetate	0.5	34	34	100ns
	Cl	0.5	34	34	100ns
	SCN	0.5	34	34	100ns

6.1.1 Osmotic Virial Coefficient Values from MD Simulations

To characterize the nature of interaction between specific ion pairs, the osmotic second virial coefficient was calculated from MD simulations by integrating the Mayer function:

$$B_{22} = -2\pi \int_0^\infty (g_{ij} - 1) r^2 dr. \tag{6.1}$$

The McMillan-Mayer B_{22} parameter is obviously not the same as the Pitzer B^ϕ parameter. However, both parameters describe binary intermolecular interactions that lead to nonideal osmotic coefficient behavior, thus it is expected that the trends in both parameters will be identical. The finite size of the simulation box leads to erroneous normalization of the simulated radial distribution functions (RDF).[176] Therefore, the asymptotic behavior of the

simulated RDF's are corrected by introducing a normalization factor, $f_{ij}(\rho)$,

$$g_{ij}(\rho) = f_{ij}(\rho) g_{ij}^{sim}(r, \rho). \tag{6.2}$$

The procedure used for the normalization of RDFs is included in the supporting information. The same technique has been used for estimating the osmotic second virial coefficient of monoatomic salts such as NaCl, KF, etc. Once the correct normalization factor is obtained, equation 6.1 is used to estimate the osmotic second virial coefficient for individual ion pairs using the normalized RDF. The upper limit of the integral was set to a finite cut-off (dependent on the system). The cut off was chosen as the distance around which osmotic second virial coefficient becomes constant or the corresponding RDF reach a value close to 1. The cut-off values lie in the range 1.2-2 nm, with the highest cut-off used for sulfate salt. The variation of the osmotic second virial coefficient calculation as the function of cut-off value for the sulfate salt is included in the Figure 6-1.

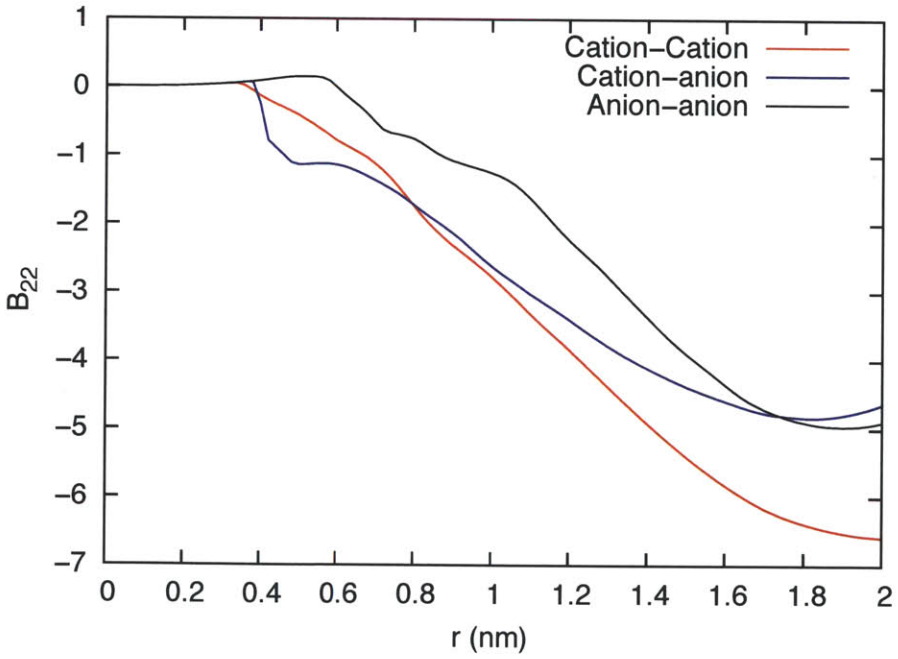

Figure 6-1: **The osmotic second virial coefficient values calculated using the normalized RDF between ion pairs in aqueous arginine sulfate solution.**

6.2 Results and Discussion

6.2.1 Ion-Ion Interactions

As mentioned in the introduction, ion-ion interactions may be contributing to the observed behavior of the arginine salts. To quantify this behavior, osmotic 2^{nd} virial coefficient values were determined for aqueous solutions of the arginine salts MD simulations. MD simulations were analyzed using the McMillan-Mayer model and compared with the experimental data obtained using the Pitzer model.[9] During the simulations, significant ion pairing was observed in the aqueous arginine salt solutions, which supports the above experimental results. The ion pairing was also observed to be dependent on the choice of anion. Ion-pairs in thiocyanate, chloride and acetate solutions are observed to be randomly distributed throughout the solution while citrate, phosphate and sulfate show a marked tendency to form hydrogen bonded clusters. Snapshots of the MD simulation boxes of arginine salts are shown in Figure 6-2. It can be seen that the structures formed by sulfate, citrate and phosphate salts are worm-like chains of sizes comparable to the box-size. Clustering observed in chloride, thiocyanate and acetate solutions is mainly due to Arg-Arg interaction as reported in the chapter on aqueous arginine hydrochloride solutions. Similar ionic clusters were also observed in aqueous guanidinium salt solutions. Brady and coworkers have shown that clustering in Gdm salts is not dependent on the choice of the water model and is not a simulation artifact.[59] Due to the presence of Gdm group as a side-chain in arginine, the clustering behavior of arginine salts is expected to be similar to Gdm salts. The only difference between Gdm and arginine salts is the presence of additional charged groups and a carbon chain in arginine. These additional charged groups (N-terminal amino group and C-terminal carboxylate group) further enhance the clustering in solution. Radial distribution functions (see Figure 6-3) between ion pairs help to further characterize the solution structure. Arg-Arg pairing, as illustrated by the RDF between the Gdm carbon atoms in the arginine molecule (see Figure 6-3a), shows a tendency of the Gdm side chain to stack with each other or form ion-pairs separated by the anion acting as a bridge between two Gdm side chains. The presence of an

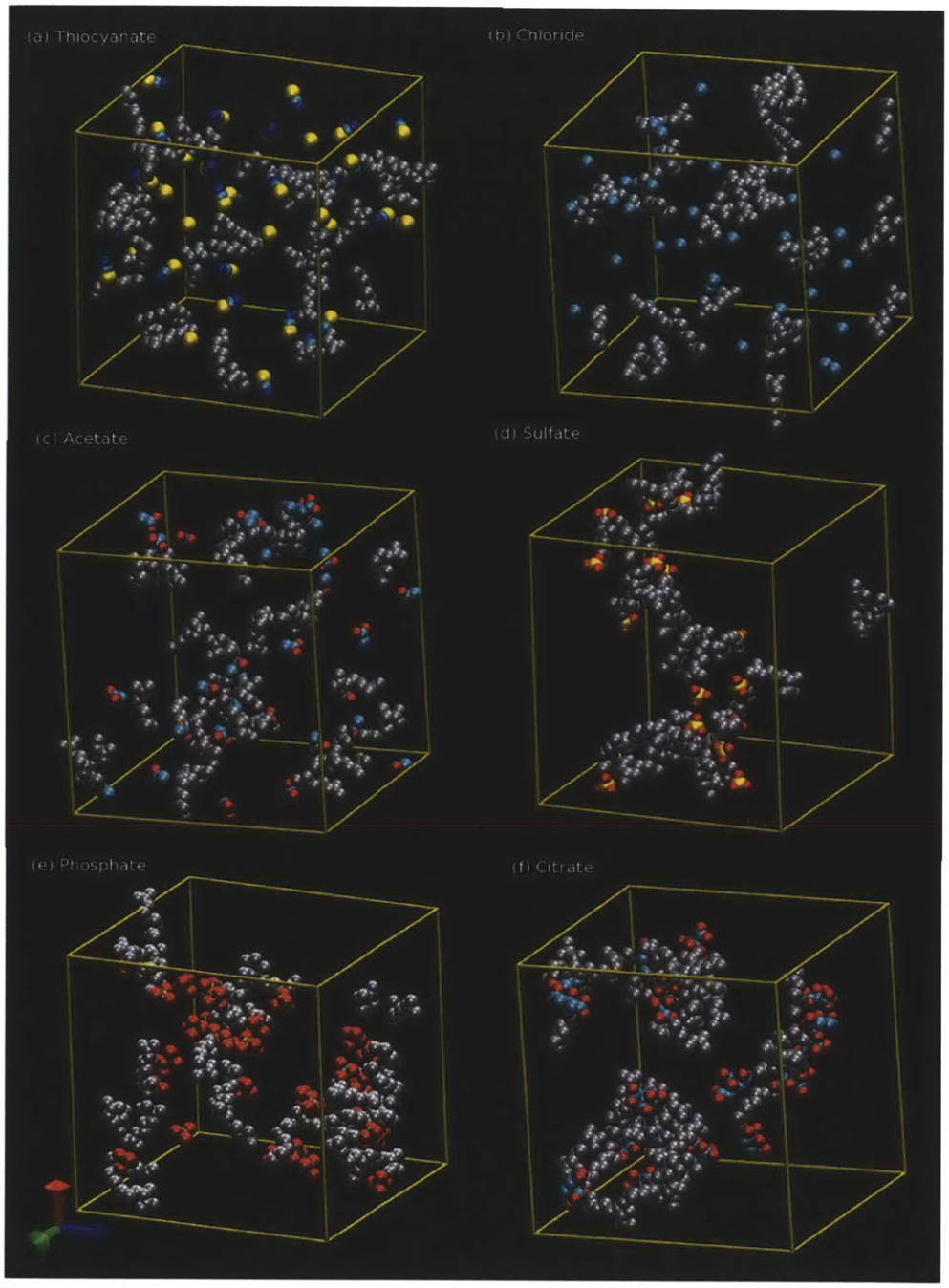

Figure 6-2: **Snapshots of the MD simulation box containing arginine salts at a molal concentration of 0.5 mol/kg. To improve the clarity of the image, water molecules are not shown and only heavy atoms (all atoms excluding hydrogen) in the arginine molecules and counter-ions are shown. The following color code is used to represent atoms: C (cyan), O (red), N (blue), S (yellow), Cl (light blue), and P (brown). Arginine molecules are shown in silver.**

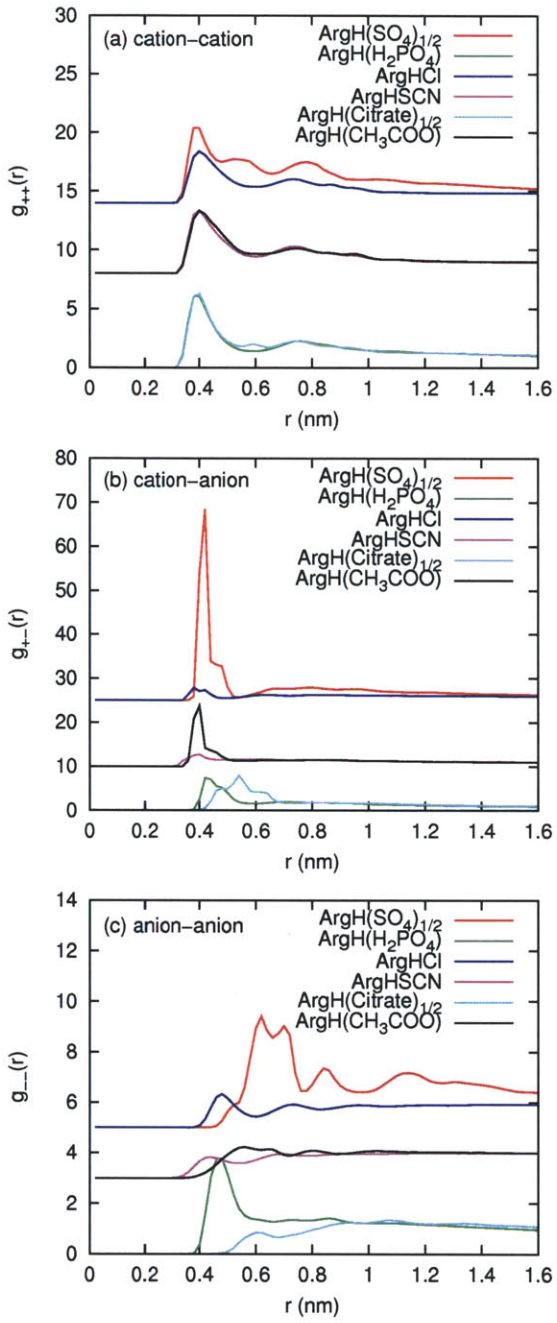

Figure 6-3: **Radial distribution functions (RDF) between ion-pairs in aqueous arginine salt solutions. The Cation-Cation RDF is the RDF between guanidinium carbon atoms of arginine. For the counter ions, the atoms used as centers for estimating the RDF's are: Sulfate:Sulfur atom, Phosphate:Phosphorus atom, Citrate:Central carbon atom, Thiocyanate:Nitrogen atom and Acetate:Carboxylate carbon atom.**

anion does not significantly affect the Arg-Arg pairing except in the case of citrate, where the large size of the anion crowds out the formation of Arg pairs. The pairing between cation and anion is strongest for the sulfate ion due to the exceptionally strong interaction between the Gdm group and sulfate (see Figure 6-3b). The acetate ion also has a strong interaction with arginine, however, the strong binding of sulfate and acetate is in stark contrast to the weak binding of thiocyanate and chloride. The pairing between anions (see Figure 6-3c) is significantly weak, as compared to other ion pairs in solution. Phosphate ion pairing exhibits a peak around 0.5 nm due to the presence of hydrogen bonding acceptors and donors in the same molecular anion. Sulfate also shows a peak around 0.6-0.7 nm but this is not due to direct interaction between sulfate ions; rather, the peak is due to the presence of multiple sulfate anions associating with adjacent or the same arginine cation. The peak heights for sulfate and chloride salts are comparable to the peak heights observed by Brady and coworkers for guanidinium salts. They also observe rugged RDF's due to the extensive clustering in guanidinium carbonate and sulfate salt solutions. (cite Brady papers for sulfate, carbonate and chloride)The RDF's between ions give an idea about their clustering behavior but these plots cannot be used as a direct measure of observed clustering in different salt solutions. Peak heights and positions are dependent on the charge on the ion, the number and type of hydrogen-bonding groups, the choice of reference atom for calculation of RDF and bulk densities. Therefore, second osmotic virial coefficient provides a better measure to compare different salts.

Table 6.2: **McMillan-Mayer Second Virial Coefficient, B_{22} (L/mol), values for ion pairs in aqueous arginine salt solutions.**

Arg Salt	C(mol/L)	B_{22}^{++}	$_{22}^{+-}$	B_{22}^{--}	Σ_{ij}
$(SO_4)_{1/2}$	0.49	-6.58	-4.63	-4.89	-20.73
$(Citrate)_{1/2}$	0.48	-2.13	-3.24	-0.35	-8.96
(H_2PO_4)	0.48	-3.20	-2.00	-1.60	-8.80
Acetate	0.48	-1.80	-1.25	-0.20	-4.50
Cl^-	0.48	-0.99	-0.90	0.45	-2.35
SCN^-	0.48	-1.40	-0.90	0.40	-2.80

The individual contribution of specific ion-pairs to the overall clustering in the solution

Table 6.3: **Number of hydrogen bonds between different ions in aqueous arginine salt solutions.**

Arg Salt	C(mol/L)	Cation-Cation	Cation-Anion	Anion-Anion	Arg-Anion-Arg
$(SO_4)_{1/2}$	0.49	42.7	51.3	0	67.3
$(Citrate)_{1/2}$	0.48	35.8	31.7	0.4	28.2
(H_2PO_4)	0.48	40.6	30.1	11.7	19.5
Acetate	0.48	29.3	30.6	0	8.3
Cl^-	0.48	49.4	6.3	0	0
SCN^-	0.48	38.2	4.0	0	0

can be estimated by calculating the osmotic second virial coefficient between ion-pairs in solution. B_{22} values for ion-pairs in each salt solution are reported in Table 6.2. It can be seen that the contribution of cation-cation interaction is similar in magnitude to cation-anion interaction. Σ_{ij} denotes the sum of interactions between all ion pairs. Sulfate, phosphate and citrate form a group with strong overall interactions whereas chloride, thiocyanate and acetate form a group with weak overall interactions. The difference between the collective structures of these two groups of anions apparently results from the differences in the hydrogen bonding of the anions to the arginine cation. The numbers of hydrogen bonds between different ion-pairs in solution are reported in Table 6.3. It can be seen that the highest number of hydrogen bonds between arginine molecules are formed in chloride and thiocyanate solutions due to the minimal interaction of these ions with the cation. The loss of hydrogen bonds between arginine molecules in acetate, sulfate, phosphate and citrate solutions is compensated by the formation of large number of cation-anion hydrogen bonds. Acetate also interacts strongly with the arginine, forming hydrogen bonds with the Gdm group as shown in Figure 6-4a, but due to the limited number of hydrogen bond acceptors, the acetate anion cannot act as a bridge between multiple cations. On the other hand, citrate, sulfate and phosphate can form bridged structures, as shown in Figure 6-4b, 6-4c, and 6-4d, respectively. The number of hydrogen bonds between arginine ions (with an anion acting as a bridge) is high for sulfate, phosphate and citrate as compared to acetate, chloride and thiocyanate, which have limited to negligible capacity to form such bridged structures. These observations also explain why sulfate, phosphate and citrate form large hydrogen bonded clusters in

solution as compared to acetate, chloride and thiocyanate. These results are in accordance to our previous reports, in which we showed that ArgHCl likely forms Arg-Arg clusters in solution at high concentration, which leads to a higher preferential exclusion.

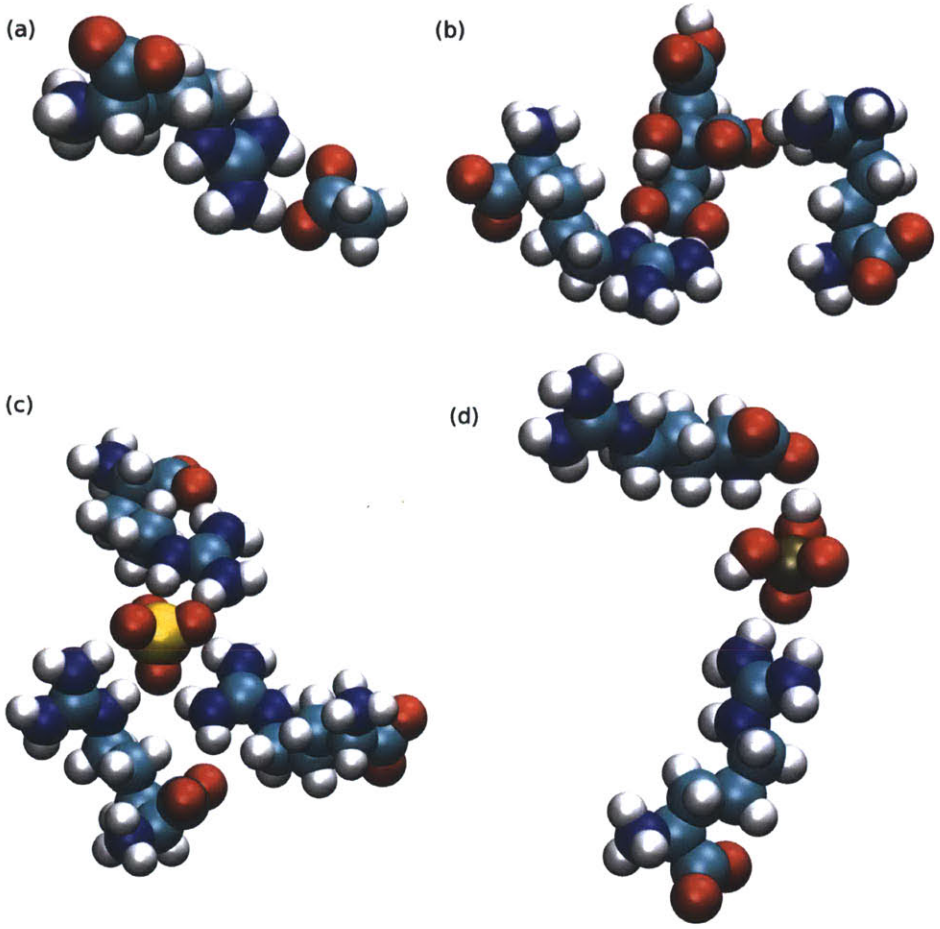

Figure 6-4: **Hydrogen bonding interaction between arginine and (a) acetate, (b) citrate, (c) sulfate and (d) phosphate anions. It can be seen that sulfate, phosphate and citrate can interact with multiple arginine molecules forming large hydrogen-bonded structures. The following color code is used to represent atoms: C (cyan), O (red), N (blue), S (yellow), and P (brown).**

6.2.2 Preferential Interactions

To gain insight into how the arginine salts inhibit protein-protein interactions, the preferential interaction coefficient, Γ_{23}, at various concentrations was determined computationally

via MD simulations and compared to the experimental VPO measurements. Theoretical preferential interaction coefficient values computed from the MD simulation are reported in Table 6.4. The estimated preferential interaction values match well with the trend in the experimental values. The experimental and computational estimates for sulfate, phosphate (if linearly extrapolated to 0.5m) and citrate indicate large, negative preferential interaction coefficient values, while thiocyanate exhibits a large, positive preferential interaction coefficient value. The experimental and computational estimates do not match exactly, which is likely due to the inability of the force field to capture the interactions of these ions with the protein surface. Force fields are optimized for interaction of ions with water molecules. However, as shown above, for arginine salts, interactions with water is only one of the key interactions. In order to obtain an accurate estimate of preferential interaction coefficient values, force fields for electrolytes should also be benchmarked to reproduce bulk solution properties and interactions between individual ions.[177] For the purposes of understanding the interesting experimental trends observed in this study, the current force fields provide reasonable estimates of the trends in the data. The Γ_{23} results for the salts show a clear difference between the anions to the left and to the right of chloride in the Hofmeister Series. The chloride, and thiocyanate salts have preferential interaction coefficient values that are positive, which is identical to GdmCl, whereas the values for the citrate, sulfate, and phosphate salts are negative. It should be noted that preferential interaction coefficient measurements and the interactions studied via MD simulations are for interactions with the native protein structure only and are not necessarily indicative of the propensity to denature protein molecules. However, most cosolutes that are attracted to the native state are typically more attracted to the unfolded state and vice versa for preferentially excluded cosolutes, due to most interactions being nonspecific in nature. It is this difference in the preferential interaction between the two states that leads to a change in the thermodynamic stability of the protein, which can be represented by the Wyman linkage relationship. Therefore, even though for most cosolutes there is a direct relationship between the preferential interaction coefficient and the folding stability of the protein, there are, however, some exceptions.[26] The mechanism

for each exception varies but all are the result of preferential interactions being repulsive in one state and attractive in the other. For compounds containing guanidinium and halide anions, exceptions can arise due to the exposure of hydrophobic residues upon unfolding.

Guanidinium, and thiocyanate are poorly hydrated ions that have a strong preference for hydrophobic residues which causes them to interact more strongly with the unfolded state and as a result, they denature proteins. That is, if they are not inhibited from interacting with the protein surface, as in the case for arginine. The preferential interaction coefficient values obtained experimentally[9] and MD simulations seem to indicate that arginine is inhibited from interacting strongly with the protein (either with the native state or the unfolded state) even though it has a guanidinium moiety, but the thiocyanate ion is free to interact with the protein. The magnitude of the interaction with the protein for these salts not only arises from a difference in the protein-anion interaction but also from a difference in the guanidinium-anion interaction, as discussed above, with thiocyanate exhibiting a weaker interaction with guanidinium.

The salts to the left of chloride in the Hofmeister series exhibit similar interactions but of differing magnitude. The phosphate, sulfate, and citrate salts have Γ_{23} values are negative, with values similar to other highly excluded solutes (e.g. glycerol), indicating that the salts have a strong repulsive interaction with the protein at all concentrations. It is well documented that salts with these anions typically stabilize the folded state, as is the case for the data shown here. Both this and the preferential exclusion can be contributed, in part, to the nonspecific repulsive interaction of the anions with protein molecules that result from the ions being well hydrated and preferring the bulk solution. But as discussed above, phosphate, sulfate and citrate have been shown to exhibit a strong attractive interaction with guanidinium, which further reduces the interaction with the protein. On the basis of the preferential interaction coefficient of individual ions, it can be concluded that the number of cations on the protein surface is related to the number of anions. Thiocyanate ions have high positive preferential interaction coefficient values but it does not interact strongly with arginine cations. Sulfate ions have a negative preferential interaction coefficient but it

Table 6.4: **Theoretical preferential interaction coefficient values for α-Cgn A in aqueous arginine salt solutions.The error bars on the preferential coefficient values are on the order of ± 1.**

Arg Salt	C(mol/L)	Γ_{VPO}	Γ_{MD}	Γ^+	Γ^-	Protein-Arg h-bonds
$(SO_4)_{1/2}$	0.50	-5.2±1.7	-4.5	-4	-3	15.3
$(Citrate)_{1/2}$	0.50	-7.7±1.5	-4.5	-4	-3	16.1
(H_2PO_4)	0.20	-5.7±0.4	-3.0	-2	-4	10.0
Acetate	0.50	N.A.	0.0	1	-1	19.7
Cl^-	0.50	-2.6±0.3	-2.5	-1	-4	18.9
SCN^-	0.48	0.2±1.2	4	4	4	26.5

interacts strongly with arginine. Therefore, there are two driving forces that push and pull arginine towards and away from the protein surface. In order to conserve the local charge near the protein surface, more arginine molecules are drawn towards the protein surface if there are a high number of anions near the protein surface. If the interactions between ions in the bulk solution are not as strong as compared to the interaction between the ion and the protein surface, the ions accumulate near the protein surface. Therefore, sulfate, phosphate and citrate salts have a large negative Γ_{23} because of their exclusion from the protein surface and the strong interaction with arginine molecules and vice-versa for thiocyanate, acetate and chloride salts. This interaction between arginine and the anion further reduces the preferential interaction of the arginine salt with the surface of the protein and further limits interactions with the unfolded state, thus enhancing thermodynamic stabilization. This cation-anion interaction does not necessarily reduce Arg-Arg interactions though, because the anion bridges together two arginine molecules, similar to guanidinium-sulfate clusters discovered by Mason and coworkers, causing the nonlinear trend observed.[59] Therefore, the ordering of the Γ_{23} and Tm data is not only ranked by how the individual anions interact with protein molecules but also, with how those anions interact with the guanidinium functional group on the arginine molecule.

6.3 Mechanistic Insight

The trends discussed here are related to the impact these salts have on protein aggregation. As discussed in the previous section, both changes in the folding equilibrium of the protein and changes in protein-protein interactions must be taken into consideration. The halide and thiocyanate salts exhibit an attractive interaction with the protein at low concentrations. This, according to the theory developed by Baynes and Trout, should reduce protein-protein interactions, much more so for arginine than for guanidinium due to the larger size of arginine.[42,68] However, as discussed before, thiocyanate also denature the protein, which counteracts this effect. This is not the case for chloride, thus ArgHCl reduces aggregation solely by reducing protein-protein interactions. However, at high concentrations, the ArgHCl shows a net exclusion and the aggregation suppression ability of ArgHCl plateaus because adding additional ArgHCl molecules at that point does not contribute any to protein-protein interaction suppression due to the additional arginine molecules being partitioned into the bulk solution. Even though ArgHCl shows a net exclusion at high concentrations, there are enough arginine molecules in the local domain of the protein that it significantly reduces protein-protein interactions, thus lowering the relative rate constant to about 0.2. This indicates that the effect from the arginine molecule should be significant for the highly excluded salt forms as well, though the effect of arginine might be reduced some. Furthermore, the aggregation suppression is further enhanced by the stabilizing effect the citrate, sulfate, and phosphate salt forms have on the folding equilibrium, leading to the relative rate constant being further reduced beyond that for ArgHCl. Moreover, another contributing factor may be an increased diffusional barrier. The cluster formation induced by the phosphate, sulfate, and citrate ions likely increases the viscosity of the solution in addition to forming a network of bridged ions around the protein, thus inhibiting the movement of protein molecules.

The clustering due to the presence of hydrogen-bonding anions leads to the formation of large clusters, which are either present within the local domain surrounding the protein or attached to the protein surface via hydrogen-bonds. These clusters of hydrogen-bonded molecules would be difficult to remove from the surface as compared to single molecules

within the local domain. In earlier chapters, we have shown that arginine self-interaction limits the binding of guanidinium group to the protein surface, which explained why arginine is not a protein destabilizer despite the presence of a denaturing guanidinium group. In this chapter, we have shown the effect of enhanced self-interaction in aqueous arginine salt solutions (due to the use of hydrogen-bonding anions) on protein-protein association.

6.4 Conclusions

For arginine salt solutions, we have shown that the interaction between ions and the interaction of the anion with the protein surface influences the preferential interaction coefficient value, which is directly related to the conformational stability of the native state. Attractive ion-ion interactions in solution lead to the formation of clusters that are larger in size as compared to the individual ions. These large clusters would be more effective at decreasing the protein-protein association than small molecules. Furthermore, clustering should lead to increase in the viscosity of the solution, which lowers the diffusion of proteins in solution. Therefore, ion-ion interactions affect the rate of aggregation via three mechanisms: 1) enhancing the native state conformational stability, 2) increasing the barrier for protein-protein association, and 3) decrease the protein-protein encounters by increasing the viscosity of the solution medium. These results not only help to elucidate the arginine mechanism, they also have large implications for interpreting the mechanism behind the Hofmeister Series, in that changes in the water structure did not seem to come into play. Rather, ion-ion interactions and other ion specific behavior seem to dominate the behavior of each arginine salt. On a last note, for a more definitive picture of how arginine influences aggregation, the research conducted here should be extended to other proteins and possibly a wider range of pH to better understand what other biophysical effects contribute to the arginine mechanism.

Chapter 7

Rational Design of Additives

Developing new additives which are more potent at inhibiting aggregation might be a means for improving biopharmaceutical formulations. However, there have only been a limited number of attempts at designing novel additives based on the current mechanistic understanding of commonly used additives. Several new compounds, which are derivatives of a particular additive, have been shown to enhance aggregation suppression.[178,179] The selection of these compounds is done mainly on the basis of their structural similarity with some commonly used additive. For example, L-Argininamide has been shown to increase the refolding of lysozyme more than arginine.[179] From a molecular interaction perspective, the additive design involves tuning various interactions in the mixed solvents in order to reduce protein-protein interactions. According to the Gap Effect theory, large molecules that have the same concentration on the protein surface as the bulk solution should be effective at inhibiting protein association. Such additives are called "neutral crowders". Creating a "neutral crowder" molecule obviously requires a balance between attractive and repulsive interactions with respect to the surface of a protein. Large molecules naturally tend to be sterically excluded from protein surfaces (e.g. polyethylene glycol). However, if functional groups that tend to preferentially bind to proteins (e.g. guanidinium, urea, etc.) were added to the surface of a large, core structure, the resulting molecule could potentially behave as a "neutral crowder". In other words, the attractive interaction between the functional

groups and the protein surface would balance out the repulsion generated by the large core structure not being able to penetrate the excluded volume region. Figure 7-1 summarizes the key parameters involved in the design of a neutral crowder molecule, which are: (i) the core structure (determines the size of the molecule), (ii) the surface functional group (determines the strength of the attractive interaction between the additive and a protein), and (iii) the counterion (only applicable if the molecule is charged but it can be used to tune the strength of the attractive interaction between the additive and a protein). Dendrimers

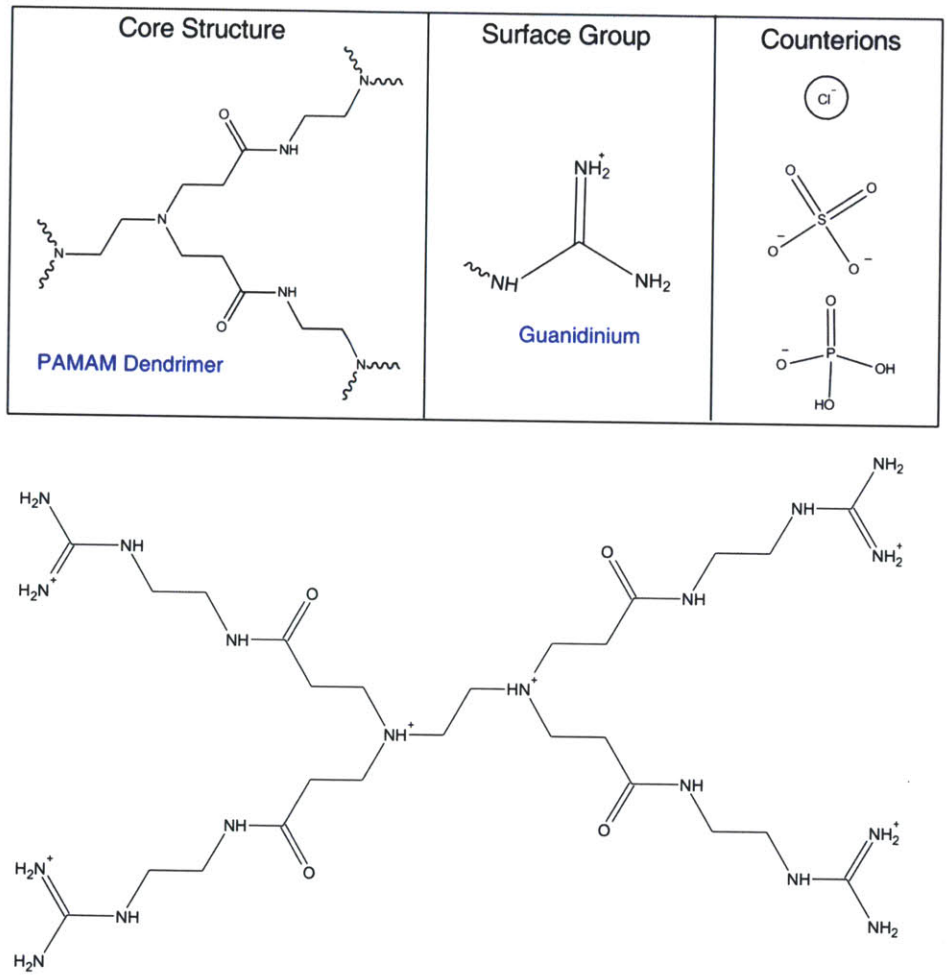

Figure 7-1: **Key structural elements of a Neutral Crowder compound: (i) Core Structure, (ii) Protein Binding Surface Group and (iii) Counterions. (top) Example Neutral Crowder compound. G0 PAMAM dendrimer with surface modified to guanidinium (bottom)**

were chosen for one of the core structures because many dendrimer forms are water-soluble polymers that are spherical in shape, vary in size, and have many surface groups that can be easily modified.[180] Guanidinium (Gdm) is a commonly used protein denaturant that forms attractive interactions with the surface of proteins and PAMAM dendrimers, which have amino surfaces, can easily be modified to a structure with a guanidino surface, making their combination an obvious choice (see Figure 7-1). Gdm is a positively charged group and the number of counterions per additive molecule is proportional to the number of surface groups. Therefore, there are a large number of counterions per additive molecule and these counterions are expected to play a critical role in the protein-additive interaction as indicated by how different arginine salts interact with proteins and inhibit aggregation.

In the previous chapter, we provided evidence that phosphate and sulfate are able to form strong hydrogen bonds with the guanidinium moiety of arginine, which seems to inhibit its binding to the surface of a protein, making the compound a conformational stabilizer. Moreover, it seems that this behavior extends to the dendrimer additives reported in this chapter. Initially, both types of additives were developed as chloride salts. However, these additives behaved in a fashion similar to denaturants, such as guanidinium chloride and urea, in that they slowed aggregation at low concentrations but enhanced aggregation at high concentrations. However, switching the chloride ion to either sulfate or phosphate eliminated this behavior, making both types of additives potent stabilizers at all concentrations.

7.1 Simulation Setup

Molecular dynamics (MD) simulations of aqueous dendrimer (generation 0) salt solutions were performed using the NAMD 2.7 program,[101] with the CHARMM27[95] force field. The TIP3P water model was used.[97] The force field parameters for the counterions were taken from the literature.[173] The force-field parameters for surface modified generation 0 dendrimer were developed using the CHARMM force field development procedure.[2,151] The partial charges and parameters for the arms of the dendrimers with guanidinium terminal groups were taken from the force field parameters for arginine. Due to the unavailability of

the experimental data related to these compounds, data obtained from quantum calculations was used to validate the force field. For the adjustment of the partial atomic charges, the QM interactions with water were calculated at the HF/6-31G* level of theory to ensure compatibility with the current CHARMM force fields. For the purpose of force field development, the dendrimer molecule was decomposed into molecular fragments, with ethylene diamine core as one fragment and the dendrimer arm as another fragment. It has been shown that non-polarizable force fields for polyvalent ions (e.g. sulfate) enhance the observed ion pairing in solution.[175] It is not possible to perform long simulations of large systems investigated in this study using polarizable force fields. Therefore, only the trends in data reported in this study will hold for sulfate. It would not make thermodynamic sense to compute preferential interaction coefficient for systems in which permanent precipitates are formed in solution. In the simulations, we have not observed permanent precipitate formation in solution but due to the significant ion-pairing observed in sulfate and phosphate solutions, the preferential interaction coefficient values do not reach a constant value within the simulation box. We have calculated the preferential interaction coefficient at a fixed distance from the protein surface. The extent of the local domain was chosen to be 6-8 Å as reported in the earlier chapters. Periodic boundary conditions were applied with long-range electrostatic interactions beyond the non-bonded cutoff of 10 Å accounted for using the Particle Mesh Ewald (PME) method[114] with the grid spacing on the order of 1 Å or less. Pressure was maintained at 1 atm using the Langevin piston method, with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature of 298 K.[115] The initial size of the periodic rectangular box containing dendrimer, counter-ion and water molecules was set to (50 Å)^3 in all simulations. To set up the simulation systems for various dendrimer salts, dendrimer molecules and counter-ions were placed randomly (without overlap) in the box, and subsequently the overlapping water molecules were removed such that the molality of the solution is 0.2 mol/kg. The total simulation time for each run was 100 ns, with a time step of 1 fs and sample collection every 1 ps. MD simulations of α-Chymotrypsinogen A (PDB Id: 2CGA) were also performed in presence of aqueous dendrimer salt solutions. The initial size of the

periodic rectangular box containing protein and water was set such that there was at least a 15 Å thick layer of water around the protein in each direction. The total simulation time for each run was 100 ns, with a time step of 1 fs. The details of the simulation setup are included in Table 7.1. A total of 600 ns of simulations were performed for these systems. The simulated trajectories were analyzed mainly in terms of radial distribution functions and the number of hydrogen bonds between protein, dendrimer and counter-ions. Theoretical preferential interaction coefficients are also calculated to assess the validity of the developed force field and to understand the nature of interactions between protein and dendrimer salts. The extent of clustering in dendrimer salts solutions was quantified in terms of the loss of solvent accessible surface area of dendrimer molecules.

Table 7.1: **Setup of Simulation systems. The molality of the dendrimer salt solutions were set to 0.2 mol/kg in all simulations.**

Simulation	Surface	Protein	No. of dendrimer molecules	No. of counter ions	No. of water molecules
S1	GdmCl		15	90	3889
S2	$Gdm(SO_4)_{1/2}$		15	45	3889
S3	$Gdm(H_2PO_4)$		15	90	3889
S4	GdmCl	α-Cgn A	36	222	10000
S5	$Gdm(SO_4)_{1/2}$	α-Cgn A	36	111	10000
S6	$Gdm(H_2PO_4)$	α-Cgn A	36	222	10000

7.2 Results and Discussion

7.2.1 Experimental results

In the experimental part of this project,[9] we showed that when the surface of PAMAM dendrimers are modified to either a guanidinium sulfate or a phosphate salt, the compounds become potent anti-aggregation additives. The sulfate and dihydrogen phosphate salt forms were shown to slow the rate of aggregation of model proteins (α-Chymotrypsinogen A, BSA and Concanavalin A) to about 2% of the original aggregation rate, which is around 10 times slower than when in the presence of arginine HCl or other aggregation suppressing excipients

(e.g. sucrose, glycerol, etc.). However, the dendrimers with guanidinium chloride surface only inhibited association at low concentrations and induced aggregation at high concentrations. We have shown that additive self-interaction could play a critical role in the mechanism by which it affects protein stability. Therefore, we studied such interactions in the modified dendrimer salts solutions to establish the exact mechansim by which these additives inhibit aggregation.

7.2.2 Intra-solvent Interactions

MD simulations were conducted on aqueous solutions of the generation 0 dendrimers to quantify how ion-ion interactions may be influencing the behavior of the additives. In Figure 7-2a, the Radial Distribution Functions (RDF) between the dendrimer and the counterions show that the sulfate and dihydrogen phosphate (H_2PO_4) ions interact strongly with the dendrimer molecules, as shown by the height of the peaks relative to chloride. The sulfate ion, which has a -2 charge on four oxygens, forms a much stronger hydrogen bond with guanidinium (gdm) surface groups as compared to H_2PO_4, which has -1 charge. The RDF peak height for sulfate is three times greater to that for H_2PO_4. In Figure 7-2b, the RDF's between dendrimer molecules indicate that in the presence of chloride ions the dendrimer molecules do not interact with each other, however, the presence of sulfate and H_2PO_4 ions tends to bring dendrimer molecules together. Clustering of dendrimers has been studied extensively but the self-clustering behavior is only observed for amphiphilic dendrimers with solvophobic core and a solvophilic shell.[181,182] PAMAM dendrimers with a charged core and surface groups can form aggregates only in the presence of an external bridging molecule, like an oppositely charged dendrimer[183] or peptide.[184] The RDF's between dendrimer molecules in the presence of different counterions show that sulfate and H_2PO_4 ions act as a bridge between dendrimer molecules leading to the formation of large clusters in solution. To verify and quantify this observation, the numbers of hydrogen bonds formed between different ion-pairs in aqueous dendrimer salt solutions were calculated (see Table 7.2). The number of hydrogen bonds was calculated using the standard CHARMM tools with a cutoff radius

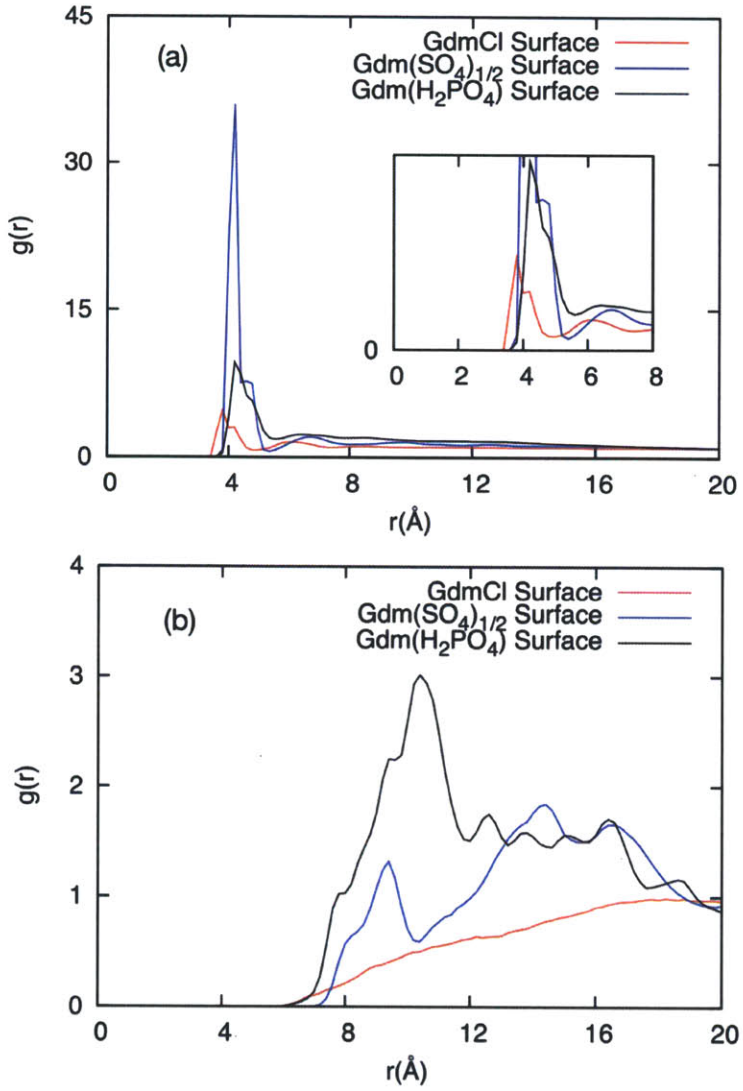

Figure 7-2: **RDF's for the interaction between the dendrimer and counterions (a) and the interaction between dendrimer molecules (b) in different dendrimer salt solutions. The distance between the centers of mass of the dendrimers is used for calculation of the RDF's. For the counterions, the sulfur atom in sulfate and the phosphorus atom in dihydrogen phosphate were utilized.**

of 2.5 between hydrogen atom and acceptor atom and a cutoff angle of 30°. Sulfate and H_2PO_4 ions, due to the presence of multiple hydrogen bond donors and acceptors, indeed act as a bridge joining dendrimer molecules together. On the other hand, chloride ions, due to the absence of multiple hydrogen-bonding groups, cannot form such bridged structures in solutions.

Table 7.2: **Number of hydrogen bonds per simulation frame formed between different species in aqueous solutions of surface modified Generation 0 PAMAM dendrimers and aCgn. [3]-molar concentration of the additive, D-Dendrimer, A-Anion, and P-Protein (aCgn)**

Surface	[3] (mol/l)	D-D	D-A	D-A-D	P-D	P-A	P-A-D
GdmCl	0.2	1	21	1	22	5	0
$Gdm(SO_4)_{1/2}$	0.2	1	146	124	13	13	35
$Gdm(H_2PO_4)$	0.2	5	150	73	14	18	32

Table 7.3: **Loss of the solvent-accessible surface area of generation 0 PAMAM dendrimers due to clustering in aqueous dendrimer salt solutions. The SAA of a dendrimer molecule in a water is 1260 Å^2.**

Surface	SAA (Å^2)	ΔSAA(Å^2)	ΔSAA(Å^2) dendrimer overlap	ΔSAA(Å^2) counter-ion overlap
GdmCl	993	267	107	160
$Gdm(SO_4)_{1/2}$	760	500	342	158
$Gdm(H_2PO_4)$	533	727	435	292

The extent of clustering in these solutions can be quantified in terms of the loss of the solvent-accessible area (SAA) of the dendrimer molecules (see Table 7.3). The solvent accessible area was estimated using standard CHARMM commands with a probe sphere of 1.4 Å radius. The SAA was used as a measure, since minimization of exposed surface area is one of the main consequences of clustering. The loss of SAA due to clustering is greatest for H_2PO_4 (60%), followed by sulfate (40%) and chloride (20%). In the case of chloride, the loss of SAA is mainly due to the presence of counterions near the dendrimer. For sulfate and H_2PO_4, the dominant component in the loss of SAA is due to the overlap of dendrimer molecules. The number of H_2PO_4 ions is twice the number of sulfate ions per dendrimer molecule, which contributes to the higher loss of SAA as compared to sulfate. MD snapshots of the simulation box containing aqueous dendrimer salt solutions confirms that significant ion pairing exists in the sulfate and H_2PO_4 salt solutions, providing a visual evidence for the presence of dendrimer aggregates formed due to the interaction between the guanidinium (Gdm) group on the dendrimer surface and the sulfate and H_2PO_4 ions (see Figure 7-3). For the chloride salt, no such dendrimer aggregates were observed. There are three key

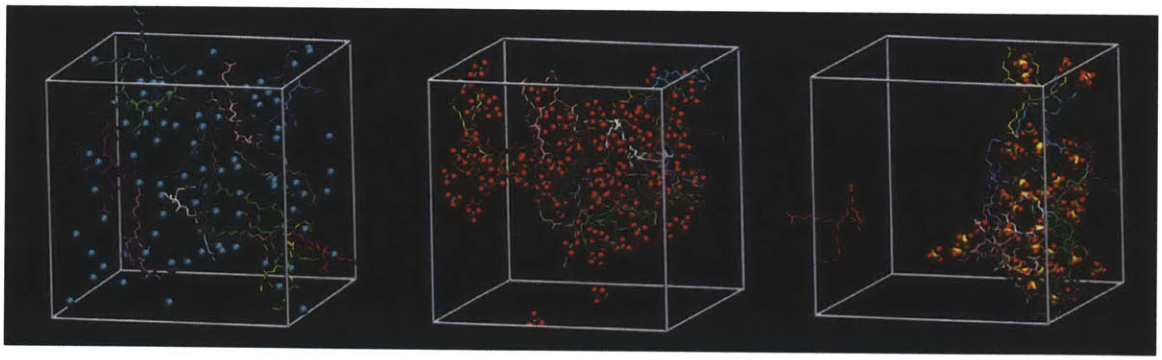

Figure 7-3: **Snapshots of aqueous generation 0 PAMAM dendrimer salt solutions obtained from MD simulations with the counterion either chloride (left), H_2PO_4 (middle) or sulfate (right). The dendrimer molecules are shown in Licorice style and counterions are shown as VdW spheres. The hydrogen atoms are not shown to improve the clarity.**

implications of this observed clustering. (i) It enhances the effective size, thus enhances the crowding ability, of the additives in solution. Furthermore, clusters are attached to the protein surface via hydrogen bonds, which could interfere with protein-protein interaction, (ii) the clusters are expected to reduce the mobility of the proteins in solutions due to a network of large hydrogen-bonded clusters around them, which could reduce the rate of protein-protein encounters[185] and (iii) the formation of these clusters influences the interaction between protein and dendrimer molecules. Therefore, from a design perspective, it is critical to understand the nature and the effect of clustering on the interactions between protein and dendrimer salts.

7.2.3 Preferential Interactions

To gain insight into how the modified dendrimer salts inhibit protein-protein interactions, the preferential interaction coefficient, Γ_{23}, at various concentrations was determined, both experimentally via vapor pressure osmometry (VPO) measurements and computationally via MD simulations. The experimental results for the preferential interaction between generation 0 PAMAM dendrimers with surfaces modified to various Gdm salts and aCgn are depicted in Figure 7-4, which shows values at various additive concentrations.[9] The first thing to note is that preferential interaction coefficient values for the chloride salt form are negative at all

concentrations, with the values exhibiting a linear trend with concentration. This indicates that the compound is excluded from the surface of the protein in its native state, which is in contradiction to the influence the compound has on stability. Typically, denaturing additives have positive preferential interaction coefficient values. Γ_{23} values for salts are a weighted average of the Γ_{23} values for individual ions. Theoretical preferential interaction coefficient values computed using the MD simulation can provide Γ_{23} values for both the cation and the anion separately. At a concentration of 0.18 mol/l, the theoretical preferential

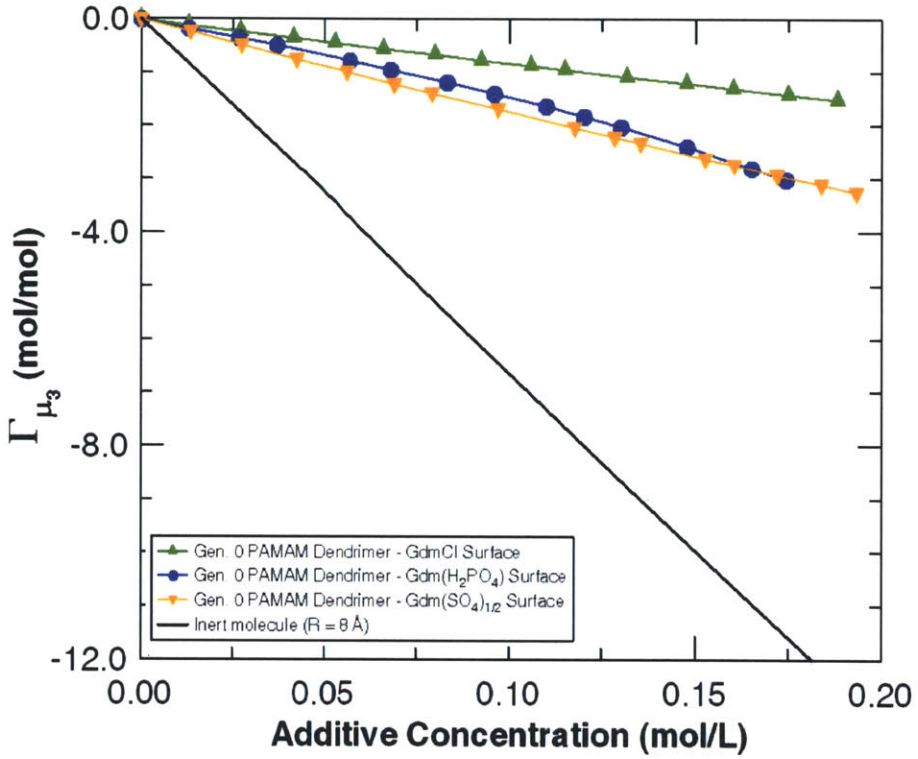

Figure 7-4: **Experimental Preferential Interaction Coefficient, Γ_{μ_3}, values versus additive concentration for the interaction between generation 0 PAMAM dendrimers, with surfaces modified to guanidinium, and aCgn.[9] Error bars left off for clarity and curves drawn through the plots to aid the eye.**

interaction coefficient for the chloride salt is found to be -0.2±1, which matches well with the experimental value of -1.5±0.7. The Γ_{23} values converge within 100 ns of simulation time. (see Figure 7-5) The size of the local domain used for the calculation was 7 Å (see Figure 7-6). The Γ_{23} value for the dendrimer cation was found to be 1, which shows that

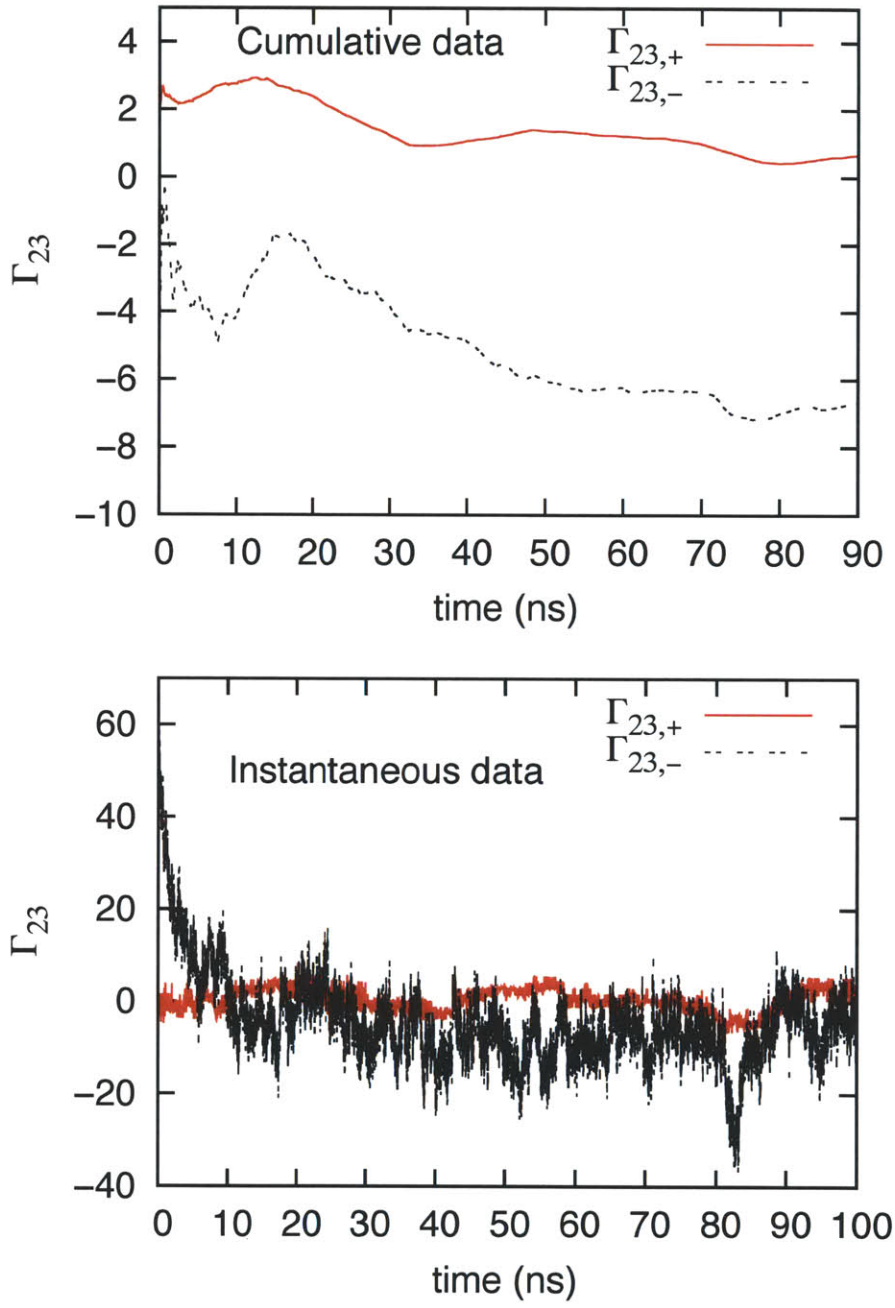

Figure 7-5: **Convergence of preferential interaction coefficient (Γ_{23}) of α-Cgn A in aqueous dendrimer (GdmCl surface) salt solution. The first 10 ns of instantaneous data are not used for calculation of cumulative averages.**

the local concentration of dendrimer molecules around the protein is higher than the bulk concentration. However, due to the negative preferential interaction value for the chloride

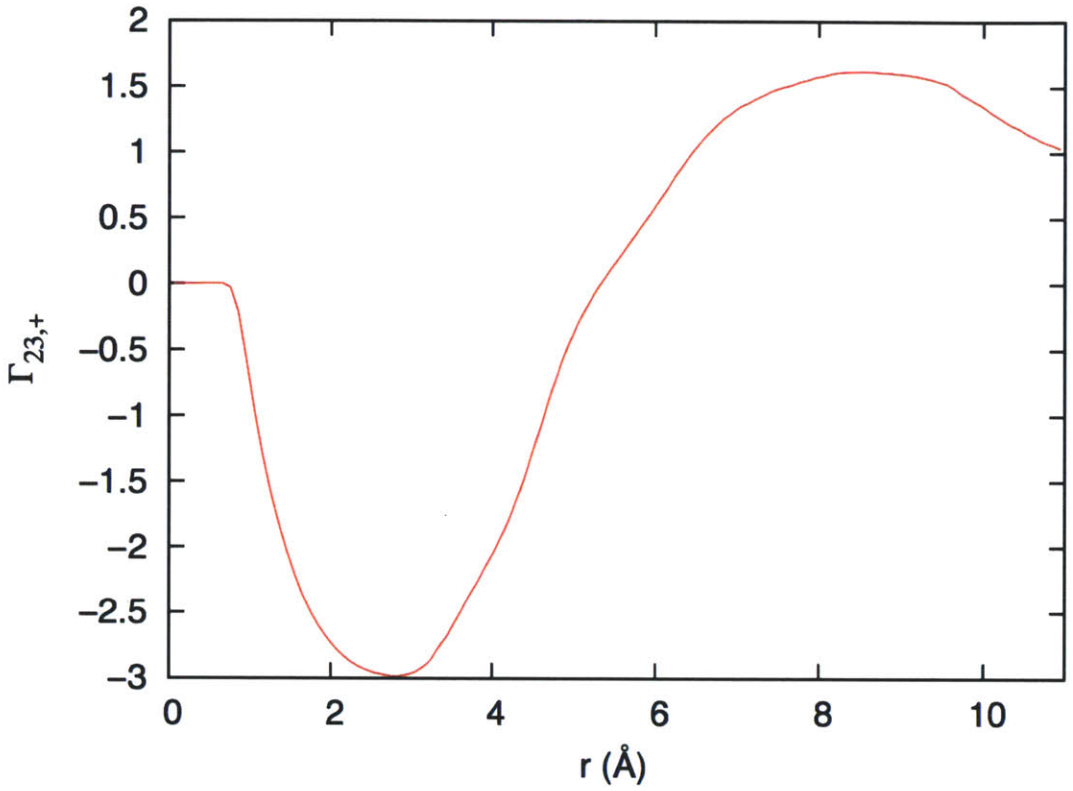

Figure 7-6: **Variation of the preferential interaction coefficient as a function of distance from the protein surface for the surface modified PAMAM dendrimer. The plot is only for the dendrimer and does not include the preferential interaction coefficient contribution due to the chloride ion.**

ion (-7), the overall preferential interaction coefficient was found to be negative. From the aggregation suppression data reported in our earlier work, we showed that generation 0 dendrimers doubles the aggregation rate at a concentration of 0.2 mol/l. Modified dendrimers can interact favorably with a variety of amino acids on the protein surface due to the presence of the Gdm group, which can form hydrogen bonds with negatively charged amino acids and the protein backbone and can also interact with aromatic amino acids via cation-π interactions. Furthermore, dendrimer molecules can bind cooperatively with the protein surface due to multiple Gdm surface groups simultaneously interacting with the protein surface (See Figure 7-7). This ability to interact favorably with a variety of amino acids explains the strong attractive protein-dendrimer interaction, which leads to enhanced protein

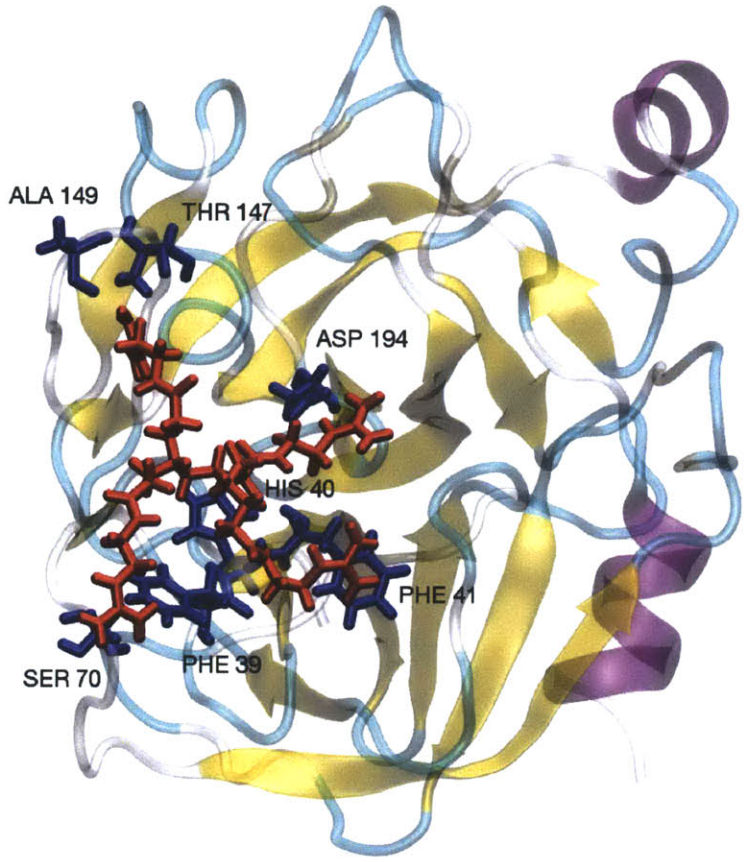

Figure 7-7: **PAMAM dendrimer with guanidinium chloride surface interacting with multiple groups on the protein surface. The guanidinium group can hydrogen bond with negatively charged amino acids and the protein backbone. It can also form cation-π interaction with aromatic amino acids.**

aggregation due to conformational destabilization, similar to the behavior of guanidinium chloride, which can also inhibit association at low concentrations but enhances aggregation at higher concentration. The concentration dependent behavior of the modified dendrimer mirrors that of ordinary guanidinium chloride when scaled by the number of surface groups. Experimental preferential interaction coefficient values for the sulfate form also exhibits a linear trend with concentration, but values for the dihydrogen phosphate form has a slight nonlinear trend with concentration, becoming more and more excluded as the concentration is increased. Therefore, when the chloride ion is exchanged with either dihydrogen phosphate or sulfate, the dendrimer becomes only slightly more excluded from the native state but is

inhibited from binding to the unfolded state, shifting the folding equilibrium to the native state (supported by the DSC experiments conducted in our lab). On the basis of the strong interaction between the dendrimer and these counterions, it can be argued that sulfate and H_2PO_4 pull the dendrimer away from the protein surface, thereby reducing the strong attractive interaction between protein and dendrimer molecules. Theoretical preferential interaction coefficients for dendrimer cation in the sulfate and dihydrogen phosphate salts is found to be -3, which shows that the dendrimer molecule is indeed slightly more exclude from the native state. The overall preferential interaction coefficients for these salts also match well with their experimental values. However, these theoretical values should be used with caution because due to the formation of clusters because the diffusivity of individual dendrimer molecules is significantly reduced for sulfate and H_2PO_4 salts. Therefore, the preferential interaction coefficient values appear to be constant for a 100 ns simulation but could change for a long simulation. Snapshots of the MD simulation boxes of aCgn in the presence of generation 0 dendrimer salts are shown in Figure 7-8. It can be seen that there are large clusters formed around the protein for the sulfate and H_2PO_4 salts. The effect of the cluster formation on the interaction between protein and dendrimers can be understood by measuring the change in number of hydrogen bonds formed between them. The number of hydrogen bonds between the protein and the dendrimer salts is reported in Table 7.2. For the chloride salt, there are 22 h-bonds between the protein and the dendrimer, which reduces by half when the ion is exchanged with sulfate or H_2PO_4. This loss in the number of direct h-bonds is compensated by the increase in the number of indirect h-bonds formed between the protein and the dendrimer, in which the counterion acts as a bridge. This observation provides direct evidence of the interference in the protein-dendrimer interaction due to the presence of a particular counterion. Another observation that supports the above hypothesis is obtained by measuring the RDF's between guanidinium carbon atoms in the four dendrimer arms with respect to the protein surface (see Figure 7-9). The RDF for the closest arm remains almost the same for all dendrimer salts but the RDF's for the remaining arms show a sharp decrease in peak height and increased distance from the surface of the

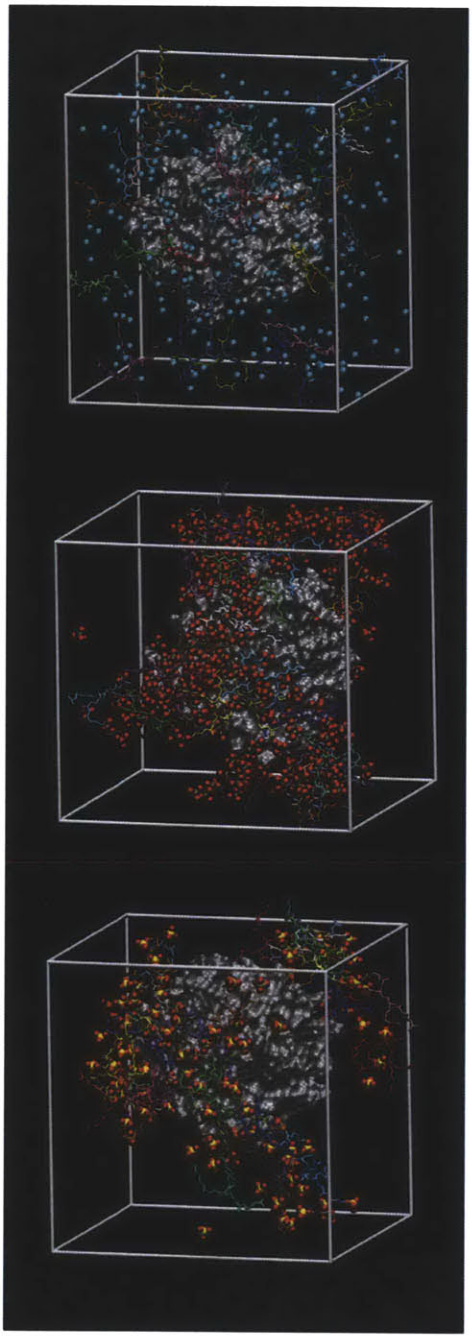

Figure 7-8: **Snapshots of simulation box from MD simulations of protein (α-Chymotrypsinogen A) in presence of aqueous dendrimer salt solutions. Chloride salt (top), H_2PO_4 (middle) or sulfate (bottom). Water molecules and hydrogen atoms are not shown to improve clarity. Dendrimer molecules are shown in Licorice representation and counter-ions are shown in van der Waals representation.**

protein for the sulfate and H_2PO_4 salt forms. This observation again supports the physical picture that the counterion limits the interaction between the terminal Gdm groups and the protein surface by interacting with the Gdm ion. These results are similar to the interaction of ArgHCl with proteins, where the carboxylate group limited the interaction between a protein and the Gdm group in arginine.21

7.2.4 Mechanistic Insight

From the results reported in this chapter, it is obvious that placing multiple guanidinium chloride functional groups on the surface of a single compound enhances the ability of the compound to bind to unfolded proteins, likely in a cooperative fashion, even though the large size of the compound should hinder binding from steric exclusion effects. This is because preferential exclusion resulting from volume exclusion effects increases in proportion to the radius of the additive raised to the third power, while preferential binding should change in proportion to the number of guanidinium groups per surface area.[68] Therefore, when only considering a balance of these two effects, large compounds, like dendrimers, should be highly excluded from the surface of the protein, even with a surface modified to a protein binding functional group, because the excluded volume effect should dominate over the preferential binding. However, the larger and the more flexible the compound, the stronger it binds to proteins,[186] indicating that the large and flexible nature of the surface modified dendrimers allows for a cooperative interaction of the guanidinium groups with the surface of the model protein, more so for the unfolded state. This enhances the amount of preferential binding per surface area, something that was not considered before, however, as demonstrated by the stabilizing effect of the phosphate and sulfate salt forms, ion-ion interactions between the guanidinium functional groups and the counterion influences how surface modified dendrimers interact with proteins, in addition to how a dendrimer molecule interacts with other dendrimers species in solution.

From the design perspective, the most effective "neutral crowder" compound would be one with the highest preferential interaction coefficient such that the conformational stabil-

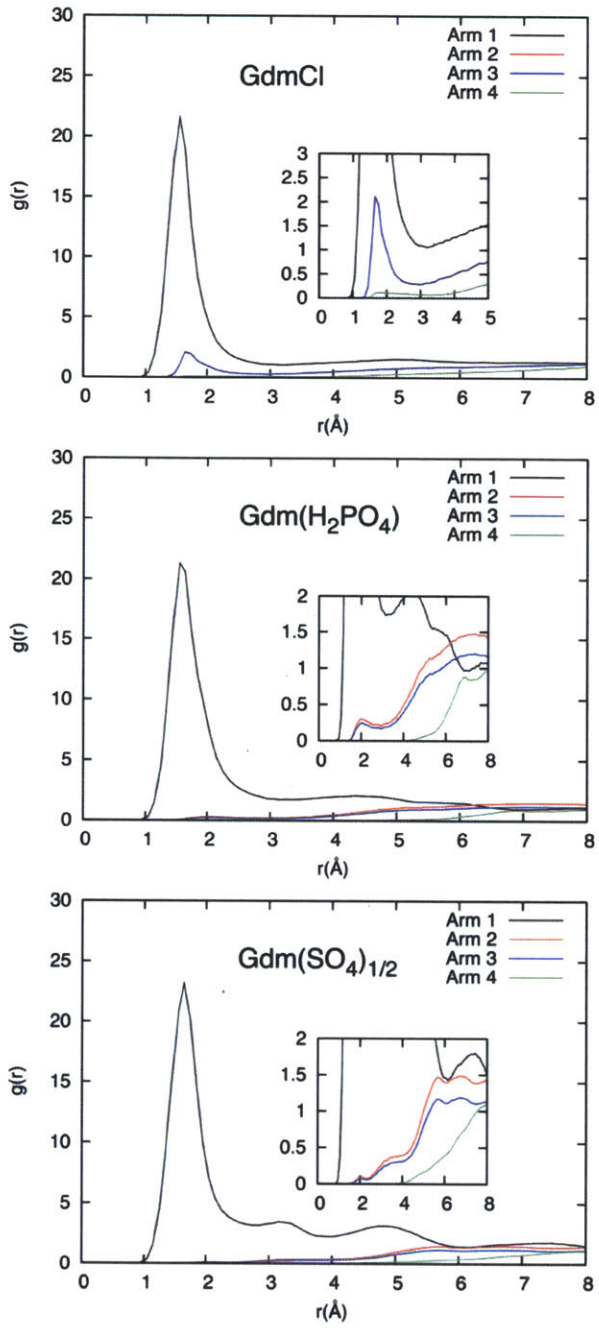

Figure 7-9: **RDF's between the model protein, aCgn, and the surface guanidinium groups on PAMAM dendrimer with the counterion either chloride (top), H_2PO_4 (middle) or sulfate (bottom). The arms of the dendrimer are labeled 1-4 depending on their distance from the protein surface, with 1 denoting the closest arm. The distance of the central carbon atom in the guanidinium group from the protein surface is used for the calculations.**

ity of the native state is not compromised. Typically, compounds with positive Γ_{23} values destabilize the native state, thereby enhancing nonnative aggregation, whereas compounds with negative Γ_{23} values could enhance the rate of native aggregation even though they stabilize the native structure. Therefore, a compound with a Γ_{23} value of zero is considered an ideal choice. Preferential interaction data reported above raises an interesting question about why the designed compounds with negative preferential interaction coefficient are considered "neutral". Although the sulfate and H_2PO_4 salts have a negative value of preferential interaction coefficient, if we look at the theoretical preferential interaction coefficient values for an inert compound the same size as the modified generation 0 dendrimer but lacking any ability to form attractive interactions, it would have a preferential exclusion over four times greater than the modified dendrimers (see Figure 7-4). It is clear from these results that the surface modified dendrimers can be considered to be only slightly excluded when compared to how excluded they would be without any protein binding surface groups. From this perspective, the preferential interaction of the surface modified dendrimers can be considered nearly "net-neutral". Also considering that all the salt forms inhibit protein aggregation at low concentrations, such results support the hypothesis that surface modified dendrimers are able to inhibit aggregation, in part, by slowing protein association through a disruption of protein-protein interactions. A highly excluded compound would not exhibit much of an effect on association due to a depletion of additive molecules in the local domain of the protein. In fact, large and highly excluded compounds often induce association due to a colloidal depletion force.[45] These results indicate that the counterion plays a critical role in fine tuning the attraction between protein and additive molecules, such that the extent of binding of the modified dendrimer molecule is different among the different salt forms and in certain cases, the attractive interaction between the additives and a protein is strong enough to inhibit protein-protein interactions but not strong enough to denature the protein. Counterions also play a critical role as a bridge joining the dendrimer molecules to form clusters in solution. This clustering effect also provides a key insight, from a design perspective, that the size of the additive can be enhanced in solution by inducing self-interaction. Clustering reduces

the effective number density of the additives in solution but it would still be important for cases where size has a more dominant effect as compared to concentration. The large size of the surface modified dendrimers inhibits protein association due to the "gap effect". However, the "gap effect" alone cannot account for the observed aggregation suppression. For an additive of the size of the dendrimer (8 Å) at a concentration of 0.2 mol/L, the reduction factor would be of the order of 0.1, which is an order of magnitude smaller than the observed aggregation suppression.

Conformational stabilization is a critical stabilizing mechanism for large excluded additives and it increases as the size of the additive increases. The excluded volume is the region around the protein that is accessible to water but is inaccessible to the additive. The thickness of the region is equal to the effective radius of the additive.[26] The preferential exclusion for the additives is not as high as an inert compound of the same size as the modified generation 0, as reported above (see Figure 7-4). However, the effective size of the designed molecules is expected to be much larger than the size of the individual additive molecules due to the formation of clusters, however, as the size increases due to clustering, the effective concentration of molecules also decreases. Therefore, it is likely that these two opposite effect cancel each other and clustering does not affect the conformational stability. The sulfate and phosphate salts are more excluded than the chloride, which is the source of observed conformational stabilization.

However, conformational stabilization is not the only thing to consider when assessing how a cosolute influences aggregation. Additives deter aggregation via conformational stabilization or association suppression. A conformational stabilizer shifts the folding equilibrium from the partially unfolded state toward the native state whereas an association suppressor specifically inhibits association without influencing the folding equilibrium.[68] The "neutral crowder" theory developed by Baynes and Trout[42] describes a class of additives, which behave solely as association suppressors, and as described in the introduction, is the motivation for modifying dendrimer compounds. An ideal "neutral crowder" would not change the conformational stability of the protein. It is difficult to isolate the contribution of the con-

formational stabilization and association suppression to the overall aggregation suppression. However, it can still argue which effect is dominant for the designed additives. The sulfate and phosphate salt forms increase the melting temperature of the protein at a rate of 15.2 and 37.4°C*M^{-1}, which implies a ΔT_m value of 3.04°C and 7.48°C at a concentration of 0.2 mol/L respectively. The ΔT_m values for sucrose and other sugars for a variety of proteins lie in the range of 3.0-17.0°C for a concentration range 0.2-1.0 mol/L.[187] These additives outperform sucrose and other sugars by more than an order of magnitude, by the amount predicted by the "gap effect". Therefore, it can be concluded that conformational stabilization is not the dominant stabilization mechanism and cannot alone explain the observed aggregation suppression. Rather, it is clear combination of the "gap effect" and conformational stabilization, which each alone would reduce aggregation to only 10-20% of the original value. Combined, they reduce aggregation to about 1-2% of the original value. Figure 7-10 shows this proposed mechanism responsible for the effect of designed additives on the free energy of protein states along the refolding/aggregation reaction coordinate.

7.3 Conclusions

In summary, we have established a likely mechanism by which designed additives suppress aggregation. In this particular case, the modified PAMAM dendrimers interact strongly with the protein due to the presence of protein binding moiety. Selecting a counterion which forms an attractive interaction with the guanidinium functional group, controls the strength of the interaction. This attractive interaction causes dendrimer molecules to form clusters in solution and inhibit the additive from binding too strongly to the surface of the protein, producing a scenario in which the additive is capable of disrupting protein-protein interactions by slightly solvating the surface of the protein but also providing conformational stabilization. The neutral crowder was used as a guiding theory for the design of these additives. However, it can be clearly seen that it is difficult to design association suppressors without affecting the conformational stability of protein. Furthermore, the effect of these additives on the diffusional barrier to protein association and the formation of large additive

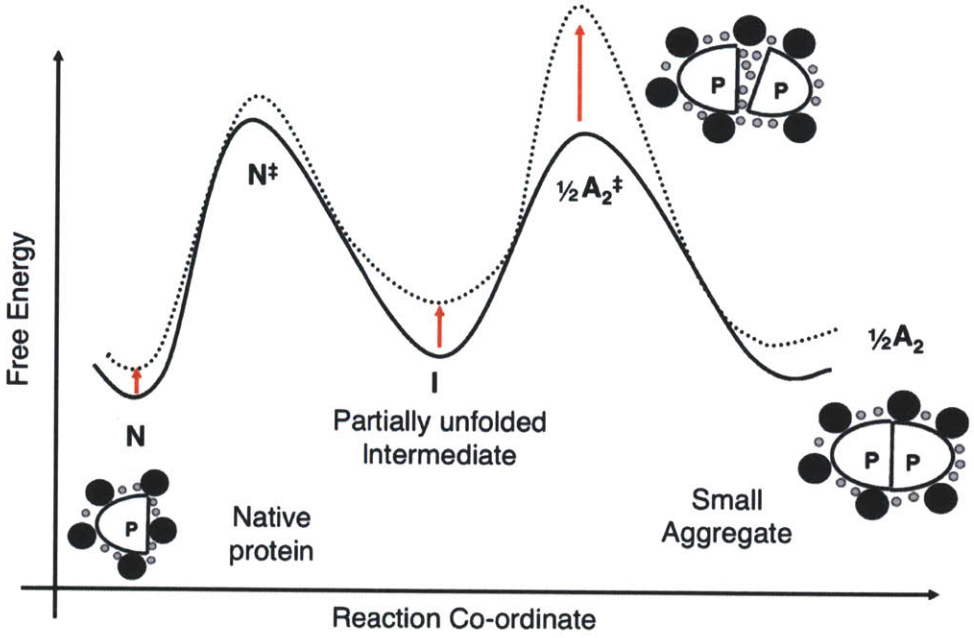

Figure 7-10: **Proposed mechanism by which the surface modified PAMAM dendrimers inhibit protein aggregation. The solid line represents the free energy in the absence of additive and the dotted line that in the presence of additive. Additive molecules are shown as large black circles and water molecules as small gray circles.**

clusters attached to the protein surface are ideas that are derived from the understanding of the mechanism by which arginine inhibits protein aggregation. A key insight obtained from this study is that additive-additive interactions play a critical role in protein-additive interactions.

Chapter 8

Conclusions and Future Work

The two main objectives of this thesis were to: (1) Elucidate the mechanism by which arginine inhibits protein aggregation. (2) Design new additives based on the acquired mechanistic understanding of the arginine and the "Neutral Crowder" theory. The main conclusion reached with respect to each of the objectives are included below.

8.0.1 Arginine Mechanism

A key insight obtained from this study is that additive-additive interactions play a critical role in protein-additive interactions. Arginine forms clusters in solution due to the self-interaction. The self-interaction of arginine limits the binding of the guanidinium group to the protein. This results explains why arginine is not a protein denaturant despite the presence of a strongly denaturing guanidinium side chain. The interaction between protein and arginine is tuned by arginine self-interaction such that it inhibits aggregation but does not alter the stability of native protein. The other effect of additive-additive interaction on protein-additive interaction on the variation of the preferential interaction coefficient values of proteins in aqueous arginine solutions. It was explained that due to the favorable arginine self-interaction in the bulk, arginine molecules are excluded from the protein surface leading to a high negative Γ_{23} values. On the basis of the mechanistic understanding of the role of arginine as an eluent in chromatography, it was found that arginine due to its charged N

and C terminal and guanidinium group can preferentially interact with a variety of amino acids (charged, aromatic, and polar), which is the molecular basis for the ability of arginine to interfere with the protein-protein association and binding. We have also shown that enhanced clustering in the aqueous arginine salts like sulfate, dihydrogen phosphate and citrate could be responsible for their high aggregation suppression ability as compared to the chloride salts. The understanding of the arginine mechanism has allowed for a rational design of novel additives that are more potent at suppressing aggregation.

8.0.2 Design of New Additives

The rational design approach presented in this thesis used the ideas from the "Neutral Crowder" theory and the understanding of the arginine mechanism. We have designed various salts of the PAMAM dendrimers with guanidinium surface groups to act as potent aggregation suppressors. The designed compounds perform an order of magnitude better than the commonly used additives such as sugars, polyols, arginine etc. We have shown that intra-solvent interactions between cations and anions in these salt solutions determine the overall strength of protein-addditive interaction. This is due to such molecules having multiple functional groups that could either cooperatively bind to the protein surface or form extended networks in solution depending on how intra-solvent interactions are tuned with the addition of other solutes or counterions. Choosing a proper balance of interactions allowed us to produce compounds which have been shown to be potent aggregation suppressors, slowing aggregation significantly. Such potent aggregation suppressing additives might be useful during production and formulation, as they could improve yield and extend the shelf-life of protein therapeutics.

8.1 Future Work

1. **Improved Additives:** The designed additives can be further optimized to improve their performance for a particular type of protein. For example, dendrimer molecules

have a large number of positively charged groups which would bind strongly with negatively charged protein, thereby denaturing it. Therefore, dendrimers with negatively charged groups (glutamic acid functional surface group) could be investigated for their potential use in suppressing aggregation of proteins. Similarly, other amino acids and functional groups could be incorporated in the design to tune the additive performance on a case by case basis. The other key functional group that can be utilized in the place of guanidinium is urea.

2. **Design of eluents for protein purification:** From the molecular viewpoint, design of eluents for separation of antibodies using affinity chromatography requires molecules which are polar so that they reduce the free energy of binding between antibosy and column, and the eluent should inhibit aggregation of eluted antibodies. The designed additives satisfy both these criterion. The change in the free energy of binding of antibody to a protein-A column in presence of additive solution could be estimated computationally.

3. **Group contribution model for preferential interaction:** Preferential Interaction coefficient can be viewed as a sum of contributions from individual exposed amino acids. A model based on a group contribution approach could be build by obtaining the preferential interaction data for individual amino acids in aqueous additive solution. The model would be able to predict the value of Γ_{23} for a particular additive for all proteins for which structural data is available.

4. **Computational procedure for predicting the aggregation rate:** In order to computationally predict the influence of additives on the free energy of association between proteins, association of two proteins in solution have to be simulated. However, the required computational time for such a system would be huge. Therefore, association of small peptides could be used as a model system. If potential aggregation prone motifs in the protein can be identified and the peptides form an aggregate structure similar to the protein aggregates such peptide sequences could be used as the model

system for the association of the protein. For example, LVEALYL peptide from insulin forms an aggregate in solution, which has a structure similar to the insulin fibrils.[188]

Bibliography

(1) Baynes, B.; Trout, B. *J. Phys. Chem. B* **2003**, *107*, 14058–14067.

(2) Kamath, G.; Guvench, O.; MacKerell Jr, A. *J. Chem. Theo. Comput.* **2008**, *4*, 765–778.

(3) Schneider, C. P.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 2050–2058.

(4) Caballero-Herrera, A.; Nilsson, L. *J. Mol. Struc.-Theochem* **2006**, *758*, 139–148.

(5) Timasheff, S. N.; Xie, G. *Biophys. Chem.* **2003**, *105*, 421–48.

(6) Black, S.; Mould, D. *Anal. Biochem* **1991**, *193*, 72–82.

(7) Saphire, E. O.; Parren, P. W. H. I.; Pantophlet, R.; Zwick, M. B.; Morris, G. M.; Rudd, P. M.; Dwek, R. A.; Stanfield, R. L.; Burton, D. R.; Wilson, I. A. *Science* **2001**, *293*, 1155.

(8) Deisenhofer, J. *Biochemistry* **1981**, *20*, 2361.

(9) Schneider, C. Ph.D. thesis, Massachusetts Institute of Technology, 2010.

(10) Arakawa, T.; Philo, J. S.; Tsumoto, K.; Yumioka, R.; Ejima, D. *Protein Expres. Purif.* **2004**, *36*, 244–248.

(11) Tsumoto, K.; Umetsu, M.; Kumagai, I.; Ejima, D.; Philo, J. S.; Arakawa, T. *Biotechnol. Prog.* **2004**, *20*, 1301–1308.

(12) Wang, W. *Int. J. Pharm.* **2005**, *289*, 1–30.

(13) Manning, M.; Chou, D.; Murphy, B.; Payne, R.; Katayama, D. *Pharm. Res.* **2010**, *27*, 544–575.

(14) Frokjaer, S.; Otzen, D. *Nat. Rev. Drug Discovery* **2005**, *4*, 298–306.

(15) Cromwell, M.; Hilario, E.; Jacobson, F. *AAPS J.* **2006**, *8*, 572–579.

(16) Roberts, C. *Biotechnol. Bioeng.* **2007**, *98*, 927–938.

(17) Rosenberg, A. *AAPS J.* **2006**, *8*, 501–507.

(18) Carpenter, J.; Kendrick, B.; Chang, B.; Manning, M.; Randolph, T. *Methods Enzymol.* **1999**, *309*, 236.

(19) Shire, S.; Shahrokh, Z.; Liu, J. *J. Pharm. Sci.* **2004**, *93*, 1390–1402.

(20) Leader, B.; Baca, Q.; Golan, D. *Nat. Rev. Drug Discovery* **2008**, *7*, 21–39.

(21) Chi, E.; Krishnan, S.; Randolph, T.; Carpenter, J. *Pharm. Res.* **2003**, *20*, 1325–1336.

(22) Wang, W. *Int. J. Pharm.* **1999**, *185*, 129–188.

(23) Davis-Searles, P.; Saunders, A.; Erie, D.; Winzor, D.; Pielak, G. *Annu. Rev. Biophys. Biomolec. Struct.* **2001**, *30*, 271–306.

(24) Wang, W.; Singh, S.; Zeng, D.; King, K.; Nema, S. *J. Pharm. Sci.* **2007**, *96*, 1–26.

(25) Jorgensen, L.; Hostrup, S.; Moeller, E.; Grohganz, H. *Exp. Opin. Drug Deliv.* **2009**, *6*, 1219–1230.

(26) Timasheff, S. *Annu. Rev. Biophys. Biomolec. Struct.* **1993**, *22*, 67–97.

(27) Timasheff, S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721.

(28) Arakawa, T.; Timasheff, S. *Biochemistry* **1982**, *21*, 6536–6544.

(29) Casassa, E.; Eisenberg, H. *Adv. Protein. Chem.* **1964**, *19*, 287–395.

(30) Bhat, R.; Timasheff, S. *Protein Sci.* **1992**, *1*, 1133.

(31) Vagenende, V.; Yap, M.; Trout, B. *Biochemistry* **2009**, *48*, 11084–11096.

(32) Stumpe, M. C.; Grubmuller, H. *J. Phys. Chem. B* **2007**, *111*, 6220.

(33) Mason, P.; Dempsey, C.; Neilson, G.; Kline, S.; Brady, J. *J. Am. Chem. Soc.* **2009**, *131*, 16689–16696.

(34) Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126–16131.

(35) Moelbert, S.; Normand, B.; De Los Rios, P. *Biophys. Chem.* **2004**, *112*, 45–57.

(36) Zhang, Y.; Cremer, P. *Curr. Opin. Chem. Biol.* **2006**, *10*, 658–663.

(37) Wyman, J.; Gill, S. J. *Binding and Linkage: Functional Chemistry of Biological Macromolecules*; University Science Books, 1990.

(38) Courtenay, E.; Capp, M.; Anderson, C.; Record Jr, M. *Biochemistry* **2000**, *39*, 4455–4471.

(39) Chi, E.; Krishnan, S.; Kendrick, B.; Chang, B.; Carpenter, J.; Randolph, T. *Protein Sci.* **2003**, *12*, 903–913.

(40) Lumry, R.; Eyring, H. *J. Phys. Chem.* **1954**, *58*, 110–120.

(41) Tischer, A.; Lilie, H.; Rudolph, R.; Lange, C. *Protein Sci.* **2010**, *19*, 1783–1795.

(42) Baynes, B.; Trout, B. *Biophys. J.* **2004**, *87*, 1631–1639.

(43) Asakura, S.; Oosawa, F. *J. Polym. Sci.* **1958**, *33*, 183–192.

(44) Minton, A. *Curr. Opin. Biotechnol.* **1997**, *8*, 65–69.

(45) Kozer, N.; Kuttner, Y.; Haran, G.; Schreiber, G. *Biophys. J.* **2007**, *92*, 2139–2149.

(46) Lindner, R.; Ralston, G. *Biophys. Chem.* **1995**, *57*, 15–25.

(47) Bennion, B.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 5142.

(48) Paul, S.; Patey, G. *J. Am. Chem. Soc.* **2007**, *129*, 4476–4482.

(49) Zangi, R.; Zhou, R.; Berne, B. *J. Am. Chem. Soc.* **2009**, *131*, 1535–1541.

(50) Schellman, J. *Biophys. Chem.* **2002**, *96*, 91–101.

(51) Idrissi, A.; Cinar, E.; Longelin, S.; Damay, P. *J. Mol. Liq.* **2004**, *110*, 201–208.

(52) Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2009**, *113*, 12816–12824.

(53) Yancey, P.; Clark, M.; Hand, S.; Bowlus, R.; Somero, G. *Science* **1982**, *217*, 1214.

(54) Bennion, B.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A* **2004**, *101*, 6433.

(55) Paul, S.; Patey, G. N. *J. Am. Chem. Soc.* **2007**, *129*, 4476–4482.

(56) Mason, P.; Neilson, G.; Dempsey, C.; Barnes, A.; Cruickshank, J. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 4557.

(57) Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M. L.; Dempsey, C. E.; MacKerell, A. D.; Brady, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 11462–11470.

(58) Courtenay, E.; Capp, M.; Record Jr, M. *Protein Sci.* **2001**, *10*, 2485–2497.

(59) Mason, P.; Dempsey, C.; Neilson, G.; Brady, J. *J. Phys. Chem. B* **2005**, *109*, 24185–24196.

(60) Omta, A.; Kropman, M.; Woutersen, S.; Bakker, H. *Science* **2003**, *301*, 347.

(61) Mason, P.; Dempsey, C.; Vrbka, L.; Heyda, J.; Brady, J.; Jungwirth, P. *J. Phys. Chem. B* **2009**, *113*, 3227–3234.

(62) Dempsey, C.; Mason, P.; Brady, J.; Neilson, G. *J. Am. Chem. Soc.* **2007**, *129*, 15895–15902.

(63) Arakawa, T.; Timasheff, S. *Biochemistry* **1984**, *23*, 5924–5929.

(64) Kita, Y.; Arakawa, T.; Lin, T. Y.; Timasheff, S. N. *Biochemistry* **1994**, *33*, 15178–15189.

(65) Arakawa, T.; Ejima, D.; Tsumoto, K.; Obeyama, N.; Tanaka, Y.; Kita, Y.; Timasheff, S. *Biophys. Chem.* **2007**, *127*, 1–8.

(66) Tsumoto, K.; Umetsu, M.; Kumagai, I.; Ejima, D.; Philo, J. S.; Arakawa, T. *Biotechnol. Progr.* **2004**, *20*, 1301–1308.

(67) Reddy, K. R. C.; Lilie, H.; Rudolph, R.; Lange, C. *Protein Sci.* **2005**, *14*, 929.

(68) Baynes, B. M.; Wang, D. I. C.; Trout, B. L. *Biochemistry* **2005**, *44*, 4919–4925.

(69) Okanojo, M.; Shiraki, K.; Kudou, M.; Nishikori, S.; Takagi, M. *J. Biosci. Bioeng.* **2005**, *100*, 556–561.

(70) Shiraki, K.; Kudou, M.; Nishikori, S.; Kitagawa, H.; Imanaka, T.; Takagi, M. *Eur. J. Biochem.* **2004**, *271*, 3242–3247.

(71) Hofmeister, F. *Arch. Exp. Pathol. Pharmakol* **1888**, *24*, 247.

(72) Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

(73) Timasheff, S. N. *Adv. Protein. Chem.* **1998**, *51*, 355–432.

(74) Das, U.; Hariprasad, G.; Ethayathulla, A. S.; Manral, P.; Das, T. K.; Pasha, S.; Mann, A.; Ganguli, M.; Verma, A. K.; Bhat, R. et al. *PLoS ONE* **2007**, *2*, e1176.

(75) Karle, I. L.; Karle, J. *Acta Cryst.* **1964**, *17*, 835–841.

(76) Schellman, J. A. *Biophys. Chem.* **2002**, *96*, 91–101.

(77) Schellman, J. A. *Biopolymers* **1994**, *34*, 1015–26.

(78) Kirkwood,; Goldberg, R. J. *J. Chem. Phys.* **1950**, *18*, 54–57.

(79) Inoue, H.; Timasheff, S. N. *Biopolymers* **1972**, *11*, 737–43.

(80) Reisler, E.; Haik, Y.; Eisenberg, H. *Biochemistry* **1977**, *16*, 197–203.

(81) Record Jr, M. T.; Anderson, C. F. *Biophysical Journal* **1995**, *68*, 786–794.

(82) Gekko, K.; Timasheff, S. N. *Biochemistry* **1981**, *20*, 4667–4676.

(83) Lin, T. Y.; Timasheff, S. N. *Biochemistry* **1994**, *33*, 12695–12701.

(84) Zhang, W.; Capp, M. W.; Bond, J. P.; Anderson, C. F.; Record Jr, M. T. *Biochemistry* **1996**, *35*, 10506–10516.

(85) Courtenay, E. S.; Capp, M. W.; Record Jr, M. T. *Protein Sci.* **2001**, *10*, 2485.

(86) Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record Jr, M. T. *Biochemistry* **2000**, *39*, 4455–71.

(87) Tanford, C. *J. Am. Chem. Soc.* **1964**, *86*, 2050–2059.

(88) Schellman, J. A. *Biopolymers* **1978**, *17*, 1305–1322.

(89) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1992**, *226*, 491–505.

(90) Smith, P. E. *Biophys. J.* **2006**, *91*, 849.

(91) Shulgin, I. L.; Ruckenstein, E. *Biophys. J.* **2006**, *90*, 704–707.

(92) Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 16271–16278.

(93) Kang, M.; Smith, P. E. *Fluid Phase Equilib.* **2007**, *256*, 14–19.

(94) Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *Israel J. Chem.* **1993**, *33*, 323–330.

(95) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(96) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207–21.

(97) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(98) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(99) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.

(100) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc* **2007**, *129*, 7346–7353.

(101) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem* **2005**, *26*, 1781–1802.

(102) Tang, K. E. S.; Bloomfield, V. A. *Biophys. J.* **2002**, *82*, 2876–2891.

(103) Jacucci, G.; Rahman, A. *Il Nuovo Cimento D* **1984**, *4*, 341–356.

(104) Friedberg, R.; Cameron, J. E. *J. Chem. Phys.* **2003**, *52*, 6049.

(105) Pace, C. N.; Heinemann, U.; Hahn, U.; Saenger, W. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 343–360.

(106) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(107) Frisch, M. J. et al. *Gaussian 03, Revision C.02*; Gaussian, Inc., Wallingford, CT, 2004.

(108) Frisch, A. E.; Frisch, M. J.; Gary, W. *Gaussian 03 User's Reference*; Gaussian, Inc., 2003.

(109) Wu, Y.; Ma, P.; Liu, Y.; Li, S. *Fluid Phase Equilibr.* **2001**, *186*, 27–38.

(110) Gosting, L. J.; Akeley, D. F. *J. Am. Chem. Soc.* **1952**, *74*, 2058–2060.

(111) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part I: The Conformation of Biological Macromolecules*; WH Freeman, San Francisco, 1980.

(112) Pfeil, W.; Privalov, P. L. *Biophys Chem* **1976**, *4*, 41–50.

(113) Schaefer, M.; Sommer, M.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 1663–1683.

(114) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(115) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.

(116) Mason, P. E.; Brady, J. W.; Neilson, G. W.; Dempsey, C. E. *Biophys. J.* **2007**, *93*, L04.

(117) Terwilliger, T. C.; D., E. *J. Biol. Chem.* **1982**, *257*, 6010–6015.

(118) Schlund, S.; Schmuck, C.; Engels, B. *Chem. Eur. J.* **2007**, *13*, 6644.

(119) Schlund, S.; Muller, R.; Grabmann, C.; Engels, B. *J. Comput. Chem.* **2008**, *29*, 407.

(120) Vondrášek, J.; Mason, P. E.; Heyda, J.; Collins, K. D.; Jungwirth, P. *J. Phys. Chem. B* **2009**, *113*, 9041–9045.

(121) Angelini, T. E.; Liang, H.; Wriggers, W.; Wong, G. C. L. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 8634–8637.

(122) Guardia, E.; Rey, R.; Padró, J. A. *Chem. Phys* **1991**, *155*, 187–195.

(123) Rapaport, D. *Mol. Phys.* **1983**, *50*, 1151–1162.

(124) Hirano, A.; Hamada, H.; Okubo, T.; Noguchi, T.; Higashibata, H.; Shiraki, K. *Protein J.* **2007**, *26*, 423–433.

(125) Barnes, M. R. *Bioinformatics for geneticists: A bioinformatics primer for the analysis of genetic data*; John Wiley and Sons Ltd., West Sussex, England, 2007.

(126) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709–17.

(127) McLain, S. E.; Soper, A. K.; Daidone, I.; Smith, J. C.; Watts, A. *Angew. Chem. Int. Ed.* **2008**, *120*, 9199–9202.

(128) Julian, R. R.; Beauchamp, J. L.; Goddard III, W. A. *J. Phys. Chem. A* **2002**, *106*, 32–34.

(129) Waldmann, T. A.; Levy, R.; Coller, B. S. *Hematology* **2000**, *2000*, 394.

(130) Scallon, B. J.; Snyder, L. A.; Mark Anderson, G.; Chen, Q.; Yan, L.; Weiner, L. M.; Nakada, M. T. *J. Immunother.* **2006**, *29*, 351.

(131) Weiner, M. L.; Surana, R.; Wang, S. *Nat. Rev. Immunol.* **2010**, *10*, 317.

(132) Maggon, K. *Curr. Med. Chem.* **2007**, *14*, 1978–1987.

(133) Shukla, A. A.; Hubbard, B.; Tressel, T.; Guhan, S.; Low, D. *J. Chromatogr. B* **2007**, *848*, 28–39.

(134) Roque, A. C. A.; Lowe, C. R.; Taipa, M. Â. *Biotechnol. Prog.* **2008**, *20*, 639–654.

(135) Keller, K.; Friedmann, T.; Boxman, A. *Trends Biotechnol.* **2001**, *19*, 438–441.

(136) Lowe, C. *Curr. Opin. Chem. Biol.* **2001**, *5*, 248–256.

(137) Labrou, N. *J. Chromatogr. B* **2003**, *790*, 67–78.

(138) Clonis, Y. *J. Chromatogr. A* **2006**, *1101*, 1–24.

(139) Salvalaglio, M.; Zamolo, L.; Busini, V.; Moscatelli, D.; Cavallotti, C. *J. Chromatog. A* **2009**, *1216*, 8678–8686.

(140) Li, R.; Dowd, V.; Stewart, D. J.; Burton, S. J.; Lowe, C. R. *Nature Biotechnol.* **1998**, *16*, 190–195.

(141) Gülich, S.; Uhlén, M.; Hober, S. *J. Biotechnol.* **2000**, *76*, 233–243.

(142) Ejima, D.; Yumioka, R.; Tsumoto, K.; Arakawa, T. *Anal. Biochem.* **2005**, *345*, 250–257.

(143) Arakawa, T.; Ejima, D.; Tsumoto, K.; Ishibashi, M.; Tokunaga, M. *Protein Expres. Purif.* **2007**, *52*, 410–414.

(144) Sauer-Eriksson, A.; Kleywegt, G.; Uhlén, M.; Jones, T. *Structure* **1995**, *3*, 265–278.

(145) Corper, A.; Sohi, M.; Bonagura, V.; Steinitz, M.; Jefferis, R.; Feinstein, A.; Beale, D.; Taussig, M.; Sutton, B. *Nat. Struct. Mol. Biol.* **1997**, *4*, 374–381.

(146) Burmeister, W.; Gastinel, L.; Simister, N.; Blum, M.; Bjorkman, P. *Nature* **1994**, *372*, 336 – 343.

(147) Fassina, G.; Verdoliva, A.; Odierna, M.; Ruvo, M.; Cassini, G. *J. Mol. Recognit.* **1996**, *9*, 564–569.

(148) Newcombe, A.; Cresswell, C.; Davies, S.; Watson, K.; Harris, G.; O'Donovan, K.; Francis, R. *J. Chromatogr. B* **2005**, *814*, 209–215.

(149) Bates, R. G.; Pinching, G. D. *J. Am. Chem. Soc.* **1949**, *71*, 1274–1283.

(150) Glusker, J. P.; van der Helm, D.; Love, W. E.; Dornberg, M. L.; Patterson, A. L. *J. Am. Chem. Soc.* **1960**, *82*, 2964–2965.

(151) Guvench, O.; MacKerell, A. *J. Mol. Model.* **2008**, *14*, 667–679.

(152) Scatchard, G. *J. Am. Chem. Soc.* **1946**, *68*, 2315–2319.

(153) Lee, B.; Kim, K. *Korean J. Chem. Eng.* **2010**, *27*, 267–277.

(154) Sadeghi, R.; Goodarzi, B. *Biophys. Chem.* **2008**, *135*, 116–124.

(155) Lockwood, D.; Rossky, P. *J. Phys. Chem. B* **1999**, *103*, 1982–1990.

(156) Lockwood, D.; Rossky, P.; Levy, R. *J. Phys. Chem. B* **2000**, *104*, 4210–4217.

(157) Matubayasi, N.; Levy, R. *J. Phys. Chem* **1996**, *100*, 2681–2688.

(158) Vagenende, V.; Yap, M.; Trout, B. *J. Phys. Chem. B* **2009**, *113*, 11743–11753.

(159) Moiani, D.; Salvalaglio, M.; Cavallotti, C.; Bujacz, A.; Redzynia, I.; Bujacz, G.; Dinon, F.; Pengo, P.; Fassina, G. *J. Phys. Chem. B* **2009**, *113*, 16268–16275.

(160) Makhatadze, G.; Privalov, P. *J. Mol. Biol.* **1992**, *226*, 491–505.

(161) Lee, B.; Richards, F. *J. Mol. Bio.* **1971**, *55*, 379–400.

(162) Casassa, E.; Eisenberg, H. *J. Phys. Chem. B* **1961**, *65*, 427–433.

(163) Sterpone, M., F. Ceccarelli; M., M. *J. Mol. Bio.* **2001**, *311*, 409–419.

(164) Arakawa, T.; Tsumoto, K.; Nagase, K.; Ejima, D. *Protein Expres. Purif.* **2007**, *54*, 110–116.

(165) Arakawa, T.; Tsumoto, K.; Kita, Y.; Chang, B.; Ejima, D. *Amino acids* **2007**, *33*, 587–605.

(166) Ghosh, R.; Sharma, S.; Chattopadhyay, K. *Biochemistry* **2009**, *48*, 1135–1143.

(167) Baldwin, R. *Biophys. J.* **1996**, *71*, 2056–2063.

(168) Tadeo, X.; Pons, M.; Millet, O. *Biochemistry* **2007**, *46*, 917–923.

(169) Lodderstedt, G.; Sachs, R.; Faust, J.; Bordusa, F.; Kühn, U.; Golbik, R.; Kerth, A.; Wahle, E.; Balbach, J.; Schwarz, E. *Biochemistry* **2008**, *47*, 2181–2189.

(170) Maity, H.; Karkaria, C.; Davagnino, J. *Curr. Pharm. Biotechnol.* **2009**, *10*, 609–625.

(171) Lange, C.; Rudolph, R. *Curr. Pharm. Biotechnol.* **2009**, *10*, 408–414.

(172) Bonner, O.; Cerutti, P. *J. Chem. Thermodyn.* **1976**, *8*, 105–112.

(173) Cannon, W.; Pettitt, B.; McCammon, J. *J. Phys. Chem.* **1994**, *98*, 6225–6230.

(174) Sansone, R.; Ebner, C.; Probst, M. *J. Mol. Liq.* **2000**, *88*, 129–150.

(175) Wernersson, E.; Jungwirth, P. *J. Chem. Theo. Comput.* **2010**, *6*, 1549–9618.

(176) Lyubartsev, A.; Marčelja, S. *Phys. Rev. E* **2002**, *65*, 041202.

(177) Kalcher, I.; Dzubiella, J. *J. Chem. Phys.* **2009**, *130*, 134507.

(178) Shiraki, K.; Kudou, M.; Nishikori, S.; Kitagawa, H.; Imanaka, T.; Takagi, M. *Eur. J. Biochem.* **2004**, *271*, 3242–3247.

(179) Hamada, H.; Shiraki, K. *J. Biotechnol.* **2007**, *130*, 153–160.

(180) Lee, C.; MacKay, J.; Fréchet, J.; Szoka, F. *Nat. Biotechnol.* **2005**, *23*, 1517–1526.

(181) Mladek, B.; Gottwald, D.; Kahl, G.; Neumann, M.; Likos, C. *Phys. Rev. Lett.* **2006**, *96*, 45701.

(182) Mladek, B.; Kahl, G.; Likos, C. *Phys. Rev. Lett.* **2008**, *100*, 28301.

(183) Jasmine, M.; Prasad, E. *J. Phys. Chem. B* **2010**, *114*, 7735–7742.

(184) Leisner, D.; Imae, T. *J. Phys. Chem. B* **2004**, *108*, 1798–1804.

(185) Minton, A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 34–39.

(186) Okuro, K.; Kinbara, K.; Tsumoto, K.; Ishii, N.; Aida, T. *J. Am. Chem. Soc.* **2009**, *131*, 1626–1627.

(187) Back, J.; Oakenfull, D.; Smith, M. *Biochemistry* **1979**, *18*, 5191–5196.

(188) Ivanova, M.; Sievers, S.; Sawaya, M.; Wall, J.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A* **2009**, *106*, 18990.