Equilibrium and kinetic factors in protein crystal growth

by

Yuba Raj Dahal

Msc., Tribhuvan University, Nepal, 2004

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Abstract

Diseases such as Alzheimer's, Parkinson's, eye lens cataracts, and Type 2 diabetes are the results of protein aggregation. Protein aggregation is also a problem in pharmaceutical industry for designing protein based drugs for long term stability. Disordered states such as precipitates and gels and ordered states such as crystals, micro tubules and capsids are both possible outcomes of protein–protein interaction. To understand the outcomes of protein–protein interaction and to find the ways to control forces, it is required to study both kinetic and equilibrium factors in protein–protein interactions.

Salting in/salting out and Hofmeister effects are familiar terminologies used in protein science field from more than a century to represent the effects of salt on protein solubility, but they are yet to be understood theoretically. Here, we build a theory accounting both attractive and repulsive electrostatic interactions via the Poisson Boltzmann equation, ion–protein binding via grand cannonical partition function and implicit ion–water interaction using hydrated ion size, for describing salting in/salting out phenomena and Hofmeister and/or salt specific effect. Our model free energy includes Coulomb energy, salt entropy and ion–protein binding free energy. We find that the salting in behavior seen at low salt concentration near the isoelectric point of the protein is the output of Coulomb energy such that the addition of salt not only screens dipole attraction but also it enhances the monopole repulsion due to anion binding. The salting out behavior appearing after salting in at high salt concentration is due to a salt mediated depletion interaction. We also find that the salting out seen far from the isoelectric point of the protein is dominated by the salt entropy term. At low salt, the dominant effect comes from the entropic cost of confining ions within the aggregates and at high salt, the dominant effect comes from the entropy gain by ions in solution by enhancing the depletion attraction. The ion size has significant effects on the entropic term which leads to the salt specificity in the protein solubility.

Crystal growth of anisotropic and fragile molecules such as proteins is a challenging task because kinetics search for a molecule having the correct binding state from a large ensemble of molecules. In the search process, crystal growth might suffer from a kinetic trap called self–poisoning. Here, we use Monte Carlo simulation to show why protein crystallization is vulnerable to the poisoning and how one can avoid such trap or recover crystal growth from such trap during crystallization. We show that self–poisoning requires only three minimal ingredients and these are related to the binding affinity of a protein molecule and its probability of occurrence. If a molecule attaches to the crystal in the crystallographic state then its binding energy will be high but in protein system this happens with very low probability ($\approx 10^{-5}$). On the other hand, non–crystallographic binding is energetically weak, but it is highly probable to happen. If these things are realized, then it will not be surprising to encounter with self–poisoning during protein crystallization. The only way to recover or avoid poisoning is to alter the solution condition slightly such as by changing temperature or salt concentration or protein concentration etc.

Equilibrium and kinetic factors in protein crystal growth

by

Yuba Raj Dahal

Msc., Tribhuvan University, Nepal, 2004

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Jeremy Schmit

Copyright

Abstract

Diseases such as Alzheimer's, Parkinson's, eye lens cataracts, and Type 2 diabetes are the results of protein aggregation. Protein aggregation is also a problem in pharmaceutical industry for designing protein based drugs for long term stability. Disordered states such as precipitates and gels and ordered states such as crystals, micro tubules and capsids are both possible outcomes of protein–protein interaction. To understand the outcomes of protein–protein interaction and to find the ways to control forces, it is required to study both kinetic and equilibrium factors in protein–protein interactions.

Salting in/salting out and Hofmeister effects are familiar terminologies used in protein science field from more than a century to represent the effects of salt on protein solubility, but they are yet to be understood theoretically. Here, we build a theory accounting both attractive and repulsive electrostatic interactions via the Poisson Boltzmann equation, ion–protein binding via grand cannonical partition function and implicit ion–water interaction using hydrated ion size, for describing salting in/salting out phenomena and Hofmeister and/or salt specific effect. Our model free energy includes Coulomb energy, salt entropy and ion–protein binding free energy. We find that the salting in behavior seen at low salt concentration near the isoelectric point of the protein is the output of Coulomb energy such that the addition of salt not only screens dipole attraction but also it enhances the monopole repulsion due to anion binding. The salting out behavior appearing after salting in at high salt concentration is due to a salt mediated depletion interaction. We also find that the salting out seen far from the isoelectric point of the protein is dominated by the salt entropy term. At low salt, the dominant effect comes from the entropic cost of confining ions within the aggregates and at high salt, the dominant effect comes from the entropy gain by ions in solution by enhancing the depletion attraction. The ion size has significant effects on the entropic term which leads to the salt specificity in the protein solubility.

Crystal growth of anisotropic and fragile molecules such as proteins is a challenging task because kinetics search for a molecule having the correct binding state from a large ensemble of molecules. In the search process, crystal growth might suffer from a kinetic trap called self–poisoning. Here, we use Monte Carlo simulation to show why protein crystallization is vulnerable to the poisoning and how one can avoid such trap or recover crystal growth from such trap during crystallization. We show that self–poisoning requires only three minimal ingredients and these are related to the binding affinity of a protein molecule and its probability of occurrence. If a molecule attaches to the crystal in the crystallographic state then its binding energy will be high but in protein system this happens with very low probability ($\approx 10^{-5}$). On the other hand, non–crystallographic binding is energetically weak, but it is highly probable to happen. If these things are realized, then it will not be surprising to encounter with self–poisoning during protein crystallization. The only way to recover or avoid poisoning is to alter the solution condition slightly such as by changing temperature or salt concentration or protein concentration etc.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to acknowledge my major professor Dr. Jeremy Schmit. Without his continuous guidance, my research at Kansas State University would be impossible. I would like to thank Professors Chris Sorensen, Bret Flanders and Paul Smith for accepting my request to sit in my research committee and providing me valuable suggestions during my research. I would also like to thank Dr. Steve Whitelam, research scientist at the Molecular Foundry at LBNL, for working with us in crystal growth project. I would also like to thank Dr. Baron Peters, associate professor of chemical engineering department at UCSF, for working with us in electrostatic work.

I am very much thankful to department of Physics at Kansas State University for selecting me as a graduate student to pursue my PhD degree. Beocat at Kansas State University provided me the computer resources in one of my research project. I want to thank Beocat family. My relatives and friends always motivated me during my study so thank you all.

Last but not the least, I want to thank my parents, my sisters, my beautiful wife Samida and my cute daughter Nerika for their unconditional support.

Chapter 1

Introduction

1.1 Introduction

Protein–protein interaction results in several outcomes. One of them is protein aggregates. Often, protein aggregates are toxic and cause diseases. For example, neurodegenerative diseases such as Alzheimer's and Parkinson's are caused by aggregation of Aβ42, and alpha-synuclein proteins respectively[1]. Type 2 diabetes is caused by aggregation of IAPP poly–peptide[2]. Similarly, eye lens cataract and sickle cell anemia are caused by condensation of lens crystallin proteins and red blood cell haemoglobin proteins respectively[3]. Apart from these diseases, complexes like virus capsid, micro tubules, enzymes are also the outcomes of protein–protein self–assembly. Protein aggregation is also a great challenge and subject of interest in the pharmaceutical industry. They want to design protein–based drugs with long term stability, but as the drugs are transported and/or stored proteins aggregate over time.

Protein crystals are also an outcome of protein–protein interaction. Protein crystals are required for the X–ray crystallography. X-ray crystallography[4;5] is a highly precise technique which reveals the three dimensional structure of macromolecules including proteins. The number of crystal structures in the protein data bank are now around 125,000[6] and more than 90 % of the structures are determined by the X-ray crystallography. But, it requires protein crystal of sufficient size and quality suitable for the diffraction. Protein crystals have

well defined structures such as cubic, tetragonal, monoclinic etc. So from such structures, it is easy to extract important informations such as interaction sites, volume fraction of protein and solvent etc.

But, the task of obtaining protein crystal of sufficient size and quality from the solution of protein monomers is difficult. In protein crystallization experiments, people frequently encounter three kinds of unwanted results[7]. The first kind of crystallization failure is that the solution remains homogeneous at the end of experiment. It means no new phase is formed. The second kind of failure is that they achieve a dense phase but not a crystal. The solid phase obtained is a gel or disordered aggregates. Another undesireable result is the formation of crystals that lacks the quality to be used in crystallography.

The first published protein crystal was haemoglobin of earthworm blood and it was first observed by Hunefeld[8] in 1840. However, protein crystallization is still a challenging task because of the lack of rational guidelines. The guidelines made for the crystallization of one protein don't work for another protein because each protein has different physical and chemical properties. So, the trial and error approach is still the most common method used for crystallization. Many organic molecules crystallize easily compared to the proteins. This is because the crystallization condition of such molecules depends on few parameters such as temperature and crystallizing species. But, in the protein system other parameters such as pH of the solution, various buffer agents, precipitating agents, concentration of salt added and its type must also be monitored in addition to the temperature and crystallizing species[9].

George and Wilson measured the osmotic second virial coefficients (B_{22}) of many proteins under crystallizing condition by using static light scattering and their study demonstrated that the value of B_{22} to lie in very narrow range. The narrow range of B_{22} value is called the crystallization slot. The crystallization slot of proteins lies between (-1×10^{-4} mol ml g^{-2}) and (-8×10^{-4} mol ml g^{-2})[10]. This "slot" is explained as the competition between two effects. Too weak attraction can't initiate phase transition leaving the solution in homogeneous form while too strong attraction leads to the unwanted gel and amorphous aggregates. For the successful crystallization of protein, it requires the precise balance between attractive and repulsive interactions.

How do protein–protein interactions lead to different types of aggregates? What are the ways to control them? These questions remain still unknown. To understand the different outcomes of protein–protein interactions and find out the ways to control them, it is required to study both kinetic and thermodynamic factors affecting the protein–protein interaction. In the following section, we present a brief introduction of kinetics and then thermodynamics associated with protein–protein interaction.

1.2 Kinetics

Crystallization is a two step process; nucleation and crystal growth. At first, crystallizing molecules nucleate to form a nucleus. Then, the nucleus searches for a crystallizing molecule having correct binding energy required for the growth. A schematic diagram of crystal growth kinetics is shown in Fig.(1.1). It shows three critical protein concentrations indicating the location of solution phase, crystal growth regime, kinetic trap, and the amorphous growth.

At low protein concentration region (below $C1$ in Fig.(1.1)), crystallizing molecules have low driving force which is insufficient to cross the energy barrier required for the nucleation. The insufficient driving force leaves crystallizing molecules in the solution phase. If supersaturation lies between $C1$ and $C2$, then it is suitable for the crystal growth. In this region, the on–rate of molecules is greater than their off-rate in such a way that the defects, if any of them were created, get enough time to be removed. Experimental evidences suggest that the crystal growth window of proteins is very narrow[10]. Further increase of supersaturation causes the growth to fall into kinetic trap which means there is no growth because the rate of binding and unbinding of crystallizing molecules are equal. The kinetic trap region lies between $C2$ and $C3$ in Fig.(1.1). After concentration $C3$, the binding rate of molecules dominates their unbinding rate such that the molecules bound in incorrect ways don't get sufficient time for unbinding leading to the amorphous growth.

At higher concentration, crystal growth is prevented by kinetic traps, a well known kinetic trap is the appearance of amorphous aggregation at concentration above $C3$. In this work we study less well understood kinetic trapping at an intermediate range of concentration

(between $C2$ and $C3$) that don't show the obvious signature of large insoluble precipitates.

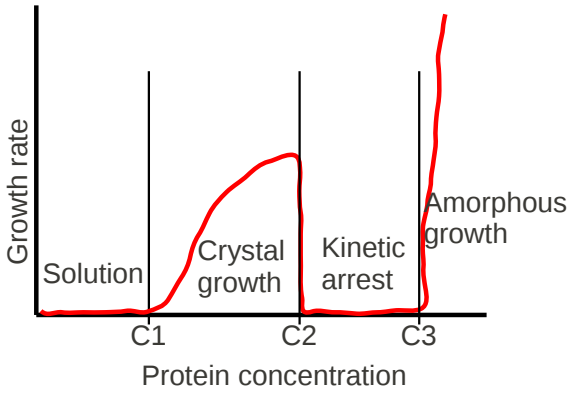

Figure 1.1: *Kinetics of crystal growth. There are three concentrations separating different stages of growth*[11]*. Below C1, crystallizing molecules remain in solution phase. Between C1 and C2, crystal grows. Crystal growth suffers from kinetic trap in concentration range from C2 to C3. After C3, there is uncontrolled aggregation leading to amorphous growth.*

1.3 Protein–protein interaction at equilibrium and "phase" diagram

The phase diagram of a protein provides the guidelines for different outcomes such as crystal, gel, amorphous structures. The study of the phase diagram also helps to narrow down the parameter space by identifying the less expensive tuning parameters required for experiments. A simple 2D phase diagram is shown in Fig. (1.2). In the figure, the precipitant concentration is varied and solubility line is drawn. The solubility line separates the two phases, namely, undersaturated and supersaturated. The undersaturated zone lies under the solubility curve whereas the supersaturated region lies above the solubility curve. Since supersaturation is the driving force for the phase transition, one should perform crystallization experiment maintaining the solution condition supersaturated. The supersaturated region is further divided into three regions. They are the metastable, labile, and precipitation zones[12]. Right above the solubility line, there exists a metastable zone where the nucleation

occurs spontaneously. But, if supersaturation is weak then it takes unreasonable time for molecules to nucleate in this zone. Right above the metastable zone, there is labile zone where nucleation occurs in experimental time scales and next to it there is the precipitation zone. In the precipitation zone, the favorable outcome is the disordered structures due to the stronger supersaturation.

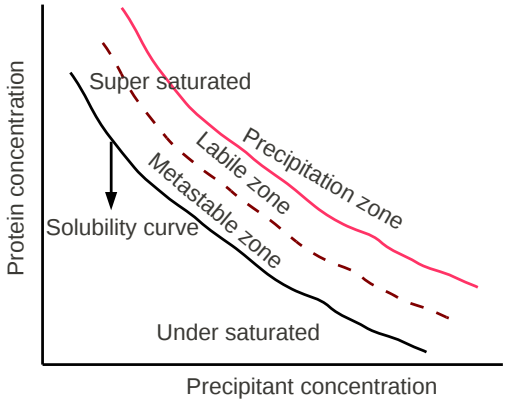

Figure 1.2: *Phase diagram showing solubility line. Solubility line separates the under-saturated and super-saturated region.*

Protein solubility is an essential thermodynamic quantity needed for the study of aggregates in different fields. For example, the first thermodynamic information needed for crystallographers for designing a crystallizing system is the solubility[13]. Unlike the second virial coefficient[14–17] which measures the two bodies interaction, solubility measures the strength of many body interaction. It is defined as the protein concentration in the soluble state when chemical potential of protein in the soluble state and the crystal state are equal[18;19]. Experimentally, the solubility at a given set of solution condition is measured either by dissolving protein crystal in the under–saturated solution till the saturation is reached or by leaving crystal in the over–saturated solution allowing the crystal growth to attain the equilibrium condition. The protein solubility curve, which is obtained by joining the protein concentration data at equilibrium at varying parameters, provides the information regarding the location of crystallizing conditions which are otherwise expensive to explore in terms of time and materials.

Experimentally, it has been shown that the solubility of protein relies on several factors. These factors include not only the properties of the protein itself but also they include the properties of the environment in which the proteins are put. The internal factors that affect protein solubility are shape and size of protein, type of amino acids and their location in the protein. Properties of the solvent such as pH[9;20], concentration and type[20–24] of salt, and temperature[9;23;25] are some external factors affecting the solubility of protein. There is no well known theory to explain these experimental measurements.

Protein–protein interaction is composed by both electrostatic[26–29] and non–electrostatic[30;31]. Below, we will discuss how solution pH and the type and concentration of salt affect electrostatic and non–electrostatic interactions.

1.3.1 Electrostatic contribution

It is known that the electrostatic free energy density of a system containing protein and mobile salt ions is composed by the Coulomb energy and salt entropy[26–28]. When a charged protein is placed in a solution having mobile salt ions then, counter ions form a screening layer around protein. While forming the screening layer, ions lose their entropy and in return, the electrostatic interaction gets strongly reduced. The competition of Coulomb energy and salt entropy can be accounted for using the Poisson Boltzmann equation.

Effect of solution pH

The net charge of a protein can be positive, zero or negative as determined by the protonation state of its charged amino acids. The contribution on protein charge by its charged amino acid depends on the pKa value of the side chain and pH of the solution. The Hendersen–Haselbach equation is used to calculate the charge of amino acid. The charge of positive and negative amino acids are $Q^+ = \frac{1}{1+10^{(pH-pka)}}$ and $Q^- = \frac{-1}{1+10^{-(pH-pka)}}$, respectively which show how the pKa of amino acid and solvent pH affect protein charge in combination. The net charge of protein is given by the summation of positive and negative residues ($Q = \sum Q^- + \sum Q^+$). The protein net charge is positive below its isoelectric point whereas above the isoelectric

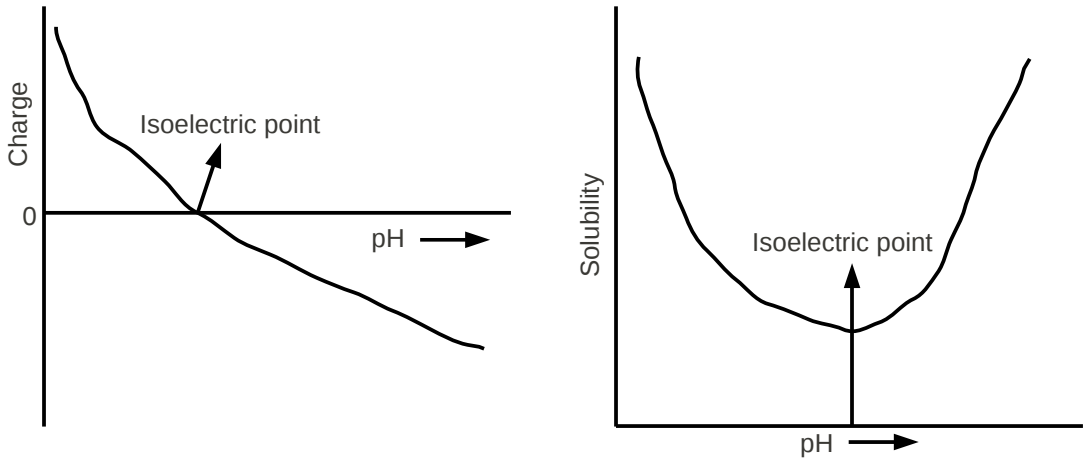

Figure 1.3: *(Left:) Variation of protein charge with respect to solution pH. The net charge of protein is zero at the isoelectric point (pI). Net charge of protein is positive when $pH < pI$ and it is negative when $pH > pI$. (Right:) Solubility of protein with respect to pH. Often, solubility is minimum at the isoelectric point.*

point it is negative. Isoelectric point (pI) of a protein is the pH at which the average charge of protein is zero. The value of pI varies from one protein to another protein depending on the population of charged amino acids on protein.

The charge state of a protein is extremely important in the study of its solubility. Generally, the solubility of protein is minimum at its isoelectric point and solubility increases while going away from isoelectric point in either direction giving "U" shape variation with pH[20]. The schematic diagram showing the variation of protein net charge and solubility with pH is shown in Fig.(1.3). Hence, the solution pH can be used as a tuning parameter to change the interaction.

Screening effect of salt

In addition to pH, salt concentration also strongly affects protein solubility by screening charges. This effect can be monotonic or non–monotonic with increase in the salt concentration (see Fig.1.4). When a protein is placed in a salt solution the salt ions, particularly counter ions build a screening layer around protein, which means the electric potential decays faster than the Coulomb potential (r^{-2}). The screening length for monovalent salt ions

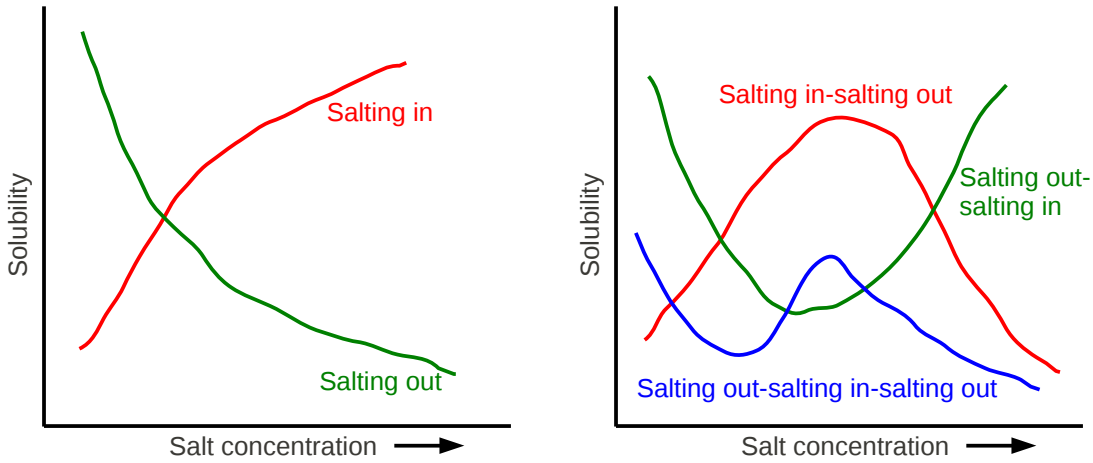

Figure 1.4: *(Left:) A diagram to show a monotonic behavior of solubility with salt concentration. If the solubility increases with the salt concentration, it is called salting in and if it decreases with the increase in salt then it is called salting out. (Right:) Non–monotonic dependence of solubility with salt concentration.*

is given by expression ($\kappa^{-1} = \sqrt{\frac{\epsilon k_B T}{2e^2 c_0}}$). The dielectric constant of the medium, temperature and salt concentration are the factors on which the screening length depends. If the dielectric constant of water is used at temperature $300K$, then the relation between screening length and salt concentration expressed in molar unit becomes ($\frac{3.05[\mathring{A}]}{\sqrt{c_0/[M]}}$). It shows that the screening length is inversly proportional to the square root of salt concentration. For an example, the value of screening length at salt concentration $1M$ is around $3\mathring{A}$.

Screening of protein charges softens the electrostatic interaction. For example, screening of monopole charge results the weakening of electrostatic repulsion and screening of higher order charge such as dipole weakens the electrostatic attraction. The net charge of the protein, the monopole, dominates the repulsive interaction between proteins. If we keep adding salt to the solution then the repulsive interaction becomes progressively weaker. Therefore, the screening of monopole repulsion leads to the salting out of proteins.

The salting out of protein is popularly used as a technique to isolate and purify protein. For this process, a salt is continuously added to the protein solution until the proteins precipitate. This method is used in number of experiments in different proteins[21;32]. At high salt concentration, salting out phenomenon is described by Cohn emperical formula[32].

Cohn formula relating the protein solubility with the salt concentration is ($\ln S = \beta - k_s m_s$). Where, S is protein solubility, m_s is salt concentration, k_s is salting out constant and β is the intercept of the straight line. This formula predicts a linear dependence of solubility with the salt concentration. But, there are many experiments that have measured the non–linear nature of salting out in the salt concentration range in which electrostatic interaction can't be neglected. Such non–linear salting out behavior of protein has been explained by Schmit and Dill by accounting for the effect of salt entropy on the repulsive electrostatic interaction[33;34].

Often, salting out of a protein is observed if the protein is far from its isoelectric point. At such pHs, effect of the higher order charges of protein are insignificant for the interaction. But, the result can be salting in if the higher order charges are significant in comparison to the monopole charge. This condition is satisfied near the isoelectric point. Experimentally, people have measured salting in behavior in number of proteins near their isoelectric point[14;35;36]. To explain salting in behavior, people have proposed two possible mechanisms. Near the isoelectric point, the net charge of protein is small allowing for attractive interactions through the alignment of patches with complementary charge. In such case, the screening of dipole attractions by salt ions leads to the increased solubility[36] (salting-in). Another proposed mechanism of salting in is the enhancement of monopole repulsion due to ion binding[37–39] to the protein surface.

Protein solubility can show non–monotonic behavior with the salt concentration too. For example, salting out can be followed by salting in and vice–versa. Often, at low salt concentration, protein solubility increases with the increase in salt concentration and after certain salt concentration it starts to decrease with the increase in salt concentration. For example, the salting in at low salt and salting out at high salt in chymosin protein have been measured experimentally[40]. Salting out followed by salting in have also been measured[41] in lysozyme protein experimentally. With the help of computer simulation by treating protein charges in a discrete way, people have shown non–monotonic solubility with the salt concentration[42;43]. At low salt, they have observed salting out. At the intermediate salt concentration, they have obtained salting in and at high salt, they have seen salting out.

The explanation of such non–monotonic solubility requires the consideration of other effects of salt such as ion–protein binding and the ion size effects in addition to the screening effect.

The screening effect of salt on charges is meaningful only up to $\approx 1M$ salt concentration because the thickness of screening layer above this salt concentration is less than $3\mathring{A}$ (comparable to the hydrated ion size). Above this concentration, the electrostatic interaction plays poor role in the interaction so the dominant interaction comes from non–electrostatic protein–ion interaction. In such high salt regime, usually proteins, follow the salting out behavior.

1.3.2 Non–electrostatic mechanism

Ion binding effect

In addition to the role of salt ions in screening electrostatic interactions, they can also bind on the protein surface by mostly non–electrostatic protein–ion interaction. There is plentiful experimental evidence showing the effect of ion binding on the protein charge[44–47]. The binding of ions not only alters the net charge of protein but also it changes the dipole moment of the protein. These both have immediate impact on the protein solubility. The number of ions that bind on the protein surface increases with the salt concentration until they find the available location to bind on the protein surface. The ion binding causes protein charge to change with the salt concentration.

In the absence of binding, the net charge of protein is constant so the screening effect of salt always weakens the electrostatic interaction. But, in the presence of ion binding, salt ions have two effects happening concurrently. One effect is to screen electrostatic interaction and another effect is to change charge of protein. So, the competition of the screening effect and the binding effect decides the nature of solubility variation with the salt concentration. Many theoretical works either ignore binding effect or use emperical formula to account for it[48]. One of the limitation of traditional Poisson Boltzmann equation is it doesn't account ion–protein association. In our model, we account for it with the help of a grand cannonical partition function.

Ion size effect

Salt specificity is important in many chemical and biological phenomena. For instances, electrolyte activities, buffers, viscosities, solubilities, protein cloud points, protein surface charges etc. Salt type can increase or decrease protein solubility. Hofmeister in 1888 classified different salt ions on the basis of their ability to precipitate protein[21]. The ranking of different salt ions on the basis of their ability to precipitate protein is known as the Hofmeister series. After Hofmeister, there are number of other solubility experiments measuring salt specific effect in various proteins[24;51;52]. Some experiments regarding the salt specific effect follow direct Hofmeister series while some follow a reverse Hofmeister series depending on the solution pH whether it is below or above the isoelectric point of protein and the observation of reverse and direct Hofmeister series also depends on the salt concentration used. For instance, people have measured the reverse Hofmeister series at pH below isoelectric point and they have measured the direct Hofmeister series above the isoelectric point in number of experiments[53;54;56;57;62]. Recently, at pH below the isoelectric point, people have measured the reverse Hofmeister series at low salt and direct series at high salt in the cloud point temperature measurement of lysozyme protein[48]. But, the underlying mechanism of salt specific effect is still poorly understood theoretically.

The popular $DLVO$ (Boris Derjaguin and Lev Landau, Evert Verwey and Theodoor Overbeek) theory[49;50] which accounts for long ranged electrostatic repulsion and van der Waals attraction fails to capture the salt specific effect. This theory works well for colloidal particles at relatively low salt concentration. But, the salt specific effect is more pronounced at high salt where electrostatic interaction is less effective. It doesn't predict distinct result if salts are made by ions of equal valency. For example, $NaCl$ and NaI salts produces the same result if $DLVO$ theory is applied though they are vastly different. In addition, globular proteins are small in size so the van der Waals interaction involved in $DLVO$ plays an insignificant role in the interaction.

The salt specific effect is being studied theoretically using different approaches that account for various salt specific interactions such as ion–protein dispersion interaction, im-

age charge interaction, hydrophobic interactions[30;57–62]. To capture salt specific effects on protein–protein interaction people have also considered the ion–water interaction. On the basis of ion–water interactions people have divided ions in to two catagories, namely, kosmotropes and chaotropes[31;63]. Kosmotrope ions have high charge density so they make thick hydration layer around them. On the other hand, chaotrope ions have low charge density so they make thin or no hydration shell around them. Most of the existing theoretical studies of salt specific effect are limited to two body interactions calculating second virial coefficient or pressure between two surfaces[57;64].

Aggregates involve many body interactions. In case of aggregates, solubility is a parameter to measure the strength of interaction. The role of salt entropy on the formation of protein aggregates has been shown before in Ref.[33;34]. In fact, the translational entropy of salt ion depends on the size making it an ion specific quantity. Here we show that if salt specific entropy is accounted for in the model then it will describe the physics undergoing in the protein aggregates in different salt solutions.

1.4 Our approach to model protein solubility and Hofmeister effect

Here, our purpose is to design a minimal analytical theory for describing protein solubility and the Hofmeister effect. In our model, we include Coulomb energy, salt entropy and protein–ion binding free energy as ingredients of total free energy. Among these three ingredients of total free energy, we show how the salt entropy term is essential for solubility and the Hofmeister effect.

The entropic cost of confining salt ions in the protein aggregate is the major contributor to the electrostatic free energy, as previously shown by Schmit and Dill[33;34]. But, in their isotropic model, only the net charge of protein was considered and higher order charges were left out. Their model also doesn't capture the salt specific effect. Here, we improve the model of Schmit and Dill by adding the electrostatic contribution from higher order charge.

In addition, we add other two effects of salt, namely, the ion binding effect and the depletion effect. With these modifications to the Schmit and Dill model, our model describes many effects of salt on the protein solubility.

Both the repulsive and attractive electrostatic interaction originate from the protein surface charge. To account for this, we use a multipole expansion of charge up to its first order. The zeroth order moment represents the monopole and it is resposible for the repulsive interaction while the first order moment represents the asymetric distribution of charged amino acid on the protein surface and captures the attractive interaction. The region where the dominant effect of zeroth order moment and first order moment depends on the pH scale of solution and salt concentration of salt used. Normally, away from isoelectric point zeroth moment has a dominant role whereas near or at isoelectric point higher order moments show their effect. The value of salt concentration also determines when zeroth moment and first order moment show their dominant behavior. The effect of zeroth order moment is observed at low salt and the effect of first order moment is observed at slightly higher salt concentration[43]. The screening effect of salt produces the salting out followed by salting in if both monopole and dipole charges have significant values[42;43] in proteins. In addition to the screening effect of salt[65], we account for binding[44] and the depletion effect[66;67] of salt too.

Salt ions are not point particles instead they have a finite size and their bare size is strongly effected by the solvation. The finite size of the ions is one source of ion specificity in the protein–protein interaction. Due to their finite sizes, there is some excluded region around proteins for ions. It affects both electrostatic and non–electrostatic interactions. When proteins make an aggregate or a crystal then it has to satisfy the neutrality condition. For the neutrality of aggregate, proteins have to trap mostly counter–ions in the solvent cavity. At low salt, ion size strongly affects the entropic loss during confinement in the cavity which is required to neutralize the aggregate. For example, for bigger ions, the entropic penalty to confine in a place is higher than the penalty for smaller ions because of the higher excluded volume. Ion size is important at high salt too. For example, at high salt, ions exert a force to bring proteins together so that the ions can minimize their excluded volume and

can gain entropy in solution. This effect of ion size is a non–electrostatic interaction known as the depletion attraction and plays a dominant role at high salt concentration.

There are numerous models developed before to describe salt specific effects such as protein–ion dispersion effect[30;46;68], image charge effect[59] and hydrophobic effect[58] but all of them are missing contribution of salt entropy. We believe, if new model is made by combining salt entropy effect from our model with other existing salt specific effects, then it will be more accurate to explain Hofmeister effect–one of the oldest problem in this field.

1.5 Our approach to model kinetics

We study the kinetics of crystal growth using Monte Carlo simulation. The kinetics of protein crystal growth look for the correct binding state because proteins are anisotropic and flexible molecules. We study crystal growth from a solution having two components. The two components are the representations of a molecule regarding its way of binding to the crystal. A molecule can attach to the crystal in two ways, namely, crystallographic and non–crystallographic ways. In our model, the crystallographic state of a molecule is represented by Blue color and the non–crystallographic state is represented by Red color. We account the binding affinity of a molecule and its probability to the crystal in following ways. The attachment of a molecule to the crystal in crystallographic way is strongly energetically favorable for the crystal growth. On the other hand, the attachment in a non–crystallographic way is weak energetically. But, the probability of crystallographic binding is very low in comparison to the non–crystallographic way of binding.

In our study, by accounting for the ways of a molecule binding to the crystal and their respective probability, we show the variation of growth rate with respect to the protein concentration. From our study, we also show how the quality of crystal changes with the protein concentration, why proteins are vulnerable for growth poisoning and how one can avoid self–poisoning or recover crystal growth.

1.6 Overview of chapters

In chapter 2, we describe our model which captures many effects of salts such as electrostatic screening, ion binding, and the depletion effect on the thermodynamics of protein crystal growth. We also break down interactions to electrostatic and non–electrostatic part to explore their individual contribution on protein solubility. In chapter 3, we study the kinetics with the help of simulation and explain the minimal requirements for self–poisoning in the crystal growth. In the final chapter, we summarize our current work and discuss potential future work.

1.7 Bibliography

Bibliography

[1] G Brent Irvine, Omar M El–Agnaf, Ganesh M Shankar, and Dominic M Walsh, *Protein Aggregation in the Brain: The Molecular Basis for Alzheimers and Parkinsons Diseases*, Mol. Med., **14(7-8)**, 451–464, 2008.

[2] Jed J. W. Wiltzius, Stuart A. Sievers, Michael R. Sawaya, and David Eisenberg, *Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process*, Protein Science, **18**, 1521–1530, 2009.

[3] Kate L. Moreau and Jonathan A. King, *Protein Misfolding and Aggregation in Cataract Disease and Prospects for Prevention*, Trends in Molecular Medicine, **18 (5)**, 2012

[4] Bernal JD, Crowfoot D, *X-ray photographs of crystalline pepsin*, Nature, **133**, 794–795, 1934.

[5] Feigin LA, Svergun DI, *Structure Analysis By Small-angle X-ray and Neutron Scattering*, New York: Plenum Press; 1987.

[6] *https://en.wikipedia.org/wiki/Protein*$_D$*ata*$_B$*ank.*

[7] N. Asherei, *Protein crystallization and phase diagrams*, Methods, **34**, 266, 2004.

[8] F. L. Hunefeld, *Der Chemismus in der thierischen Organisation. Physiologisch-chemische Untersuchungen der materiellen Veranderungen, oder des Bildungslebens im thierischen Organismus; insbesondere des Blutbildungsprocesses, der Natur der Blut Korperchen und ihrer Kernchen. Ein Beitrag zur Physiologie und Heilmittellehre. Brockhaus, Leipzig, (1840).*

[9] R. C. DeMattei and R. S. Feigelson, *Journal of Crystal Growth*, **110(34)**, (1991).

[10] A. George, W. Willium Wilson, *Predicting Protein Crystallization from a Dilute Solution Property*, Acta Cryst. D, **50**, 361–365, 1994.

[11] S. Whitelam, Y. R. Dahal and J. Schmit, *J. Chem. Phys.*, **144**, 064903, (2016).

[12] A. McPherson, newblock *Crystallization of Biological Macromolecules*, Cold Spring Harbor: CSHL Press, 1999.

[13] J. W. Mullin, *Crystallizatlon, 2nd ed. Butterworths, London*, (1972).

[14] Andre C. Dumetz, Aaron M. Chockla, Eric W. Kaler, Abraham M. Lenhoff, *Effects of pH on proteinprotein interactions and implications for protein phase behavior*, Biochimica et Biophysica Acta, **1784**, 600–610, 2008.

[15] B.L. Neal, D. Asthagiri, A.M. Lenhoff, *Molecular Origins of Osmotic Second Virial Coefficients of Proteins*, Biophysical Journal, **75**, 2469–2477, 1998.

[16] B.L. Neal, D. Asthagiri, O.D. Velev, A.M. Lenhoff, E.W. Kaler, *Why is the osmotic second virial coecient related to protein crystallization* , Journal of Crystal Growth, **196**, 377–387, 1999.

[17] Andre C. Dumetz, Ann M. Snellinger–O'Brien, Eric W. Kaler, Abraham M. Lenhoff, *Pattern of protein–protein interaction in salt solution and implication for protein crystallization*, Protein Science, **16**, 1867–1877, 2007.

[18] Cohn, E.J., J.T. Edsall, *Proteins, Amino Acids, and Peptides, Hafner Publishing Company New York*, 1943.

[19] Arakawa, T., and S. N. Timasheff, *Theory of protein solubility*, Methods Enzymol., **114**, 49–77, 1985.

[20] A. A. Green, *J. Biol. Chem.*, **93**, 517, 1931.

[21] F. Hofmeister, *Arch. Exp. Pathol. Pharmakol.*, **24**, 247, 1888.

[22] Vincent Mikol and Richard Giege, *Phase diagram of a crystalline protein: determination fo the solubility of concanavalin A by a microquantitation assay*, Journal of Crystal Growth, **97**, 324–332, 1989.

[23] Sylvaine Lafont, Stphane Veesler, Jean Pierre Astier, Roland Boistelle, *Journal of Crystal Growth*, **143**, 249–255 (1994).

[24] Claire Carbonnaux, Madeleine Ries—Kautt, and Arnaud Ducruix, *Protein Science*, **4**, 2123–2128 (1995).

[25] Monika Budayova, Jean-Pierre Astier, Stephane Veesler, Mirjam Czjzek, Anne Belaich, Roland Boistelle, *Journal of Crystal Growth*, **196**, 297–304 (1999).

[26] Kim A. Sharp and Barry Honig, *Calculating Total Electrostatic Energies with the Nonlinear Poisson-Bottzmann Equation*, J. Phys. Chem. **94**, 7684, (1990).

[27] D. Andelman, *Electrostatic Properties of Membranes: The PoissonBoltzmann Theory*, Handbook of Biological Physics (Elsevier Science), **1**, 603–641, (1995).

[28] Andelman, D., *Introduction to electrostatics in soft and biological matter*, In Proceedings of the Nato ASI and SUSSP on Soft Condensed Matter Physics in Molecular and Cell Biology, 97–122, Taylor Francis, New York, (2006).

[29] Jones S. Thornton JM, *Principle of protein–protein interactions*, PNAS, **93**, 13–20, (1996).

[30] Barry W. Ninham and Vassili Yaminsky, *Ion Binding and Ion Specificity: The Hofmeister Effect and Onsager and Lifshitz Theories*, Langmuir, **13**, 2097–2108, (1997).

[31] K. A. Dill, T. M. Truskett, V. Vlachy, and B. Hribar-Lee, *Annu. Rev. Biophys. Biomol. Struct.*, **34(173)**, (2005).

[32] E.J. Cohn, *The physical chemistry of the proteins*, Physiol. Rev., **5**, 349–427, (1925).

[33] Jeremy D. Schmit, S. Whitelam, and Ken A. Dill, *Electrostatics and aggregation: How charge can turn a crystal into a gel*, J. Chem. Phys., **135(085103)**, (2011).

[34] Jeremy D. Schmit, and Ken A. Dill, *The Stabilities of Protein Crystals*, J. Phys. Chem. B, **114**, 4020–4027, (2010).

[35] J. Mellanby, *Globulin*, J. Physiol., **33**, 338–373, (1905).

[36] C. Tanford, *Physical Chemistry of Macromolecules*, John Wiley and Sons Inc., (1966).

[37] T. Arakawa, SN Timasheff, *Biochemistry*, **21**, 6545–6552, (1982).

[38] P. Retailleau, M. Ries–Kautt, A. Ducruix, *Biophys. J.*, **73**, 2156–2163, (1997).

[39] AA Green, *J. Biol. Chem.*, **95**, 47–66, (1932).

[40] Ronald W. Maurer, Stanley I. Sandler, Abraham M. Lenhoff, *Salting-in characteristics of globular proteins*, Biophysical Chemistry, **156**, 7278, (2011).

[41] Philippe Benas, Laurent Legrand and Madeleine Ris-Kautt, *Strong and specific effects of cations on lysozyme chloride solubility*, Acta Cryst., **D58**, 1582–1587, (2002).

[42] E. Allahyarov, H. Lowen, J. P. Hansen, and A. A. Louis, *Nonmonotonic variation with salt concentration of the second virial coefficient in protein solutions*, Physical Review E, **67**, (2003).

[43] F. W. Tavares, D. Bratko, A. Striolo, H. W. Blanch and J. M. Prausnitz, *Phase behavior of aqueous solution containing dipolar proteins from second–order perturbation theory*, J. Chem. Phys., **120 (20)**, 9859–5869, (2004).

[44] Yatin R. Gokarn, R. Matthew Fesinmeyer, Atul Saluja, Vladimir Razinkov, Susan F. Chase, Thomas M. Laue, and David N. Brems, *Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions*, Protein Science, **20**, 580587, (2011).

[45] Manoj K. Menon and Andrew L. Zydney, *Measurement of Protein Charge and Ion Binding Using Capillary Electrophoresis*, Anal. Chem., **70**, 1581–1584, (1998).

[46] Luca Medda , B. Barse, F. Cugia, M. Bostrom, D. F. Parsons, B. W. Ninham, M. Monduzzi, and A. Salis, *Hofmeister Challenges: Ion Binding and Charge of the BSA Protein as Explicit Examples*, Langmuir, **28**, 16355–16363, (2012).

[47] Travis T. Waldron, Modestos A. Modestou, and Kenneth P. Murphy, *Anion binding to a proteinprotein complex lacks dependence on net charge*, Protein Science, **12**, 871–874, (2003).

[48] Yanjie Zhang and Paul S. Cremer, *The inverse and direct Hofmeister series for lysozyme*, PNAS , **106(36)**, 15251–15257, (2009).

[49] Derjaguin B., Landau L., *Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes*, Acta Physico Chemica URSS, **14**, 633, (1941).

[50] Verwey E., Overbeek T., *Theory of the stability of lyophobic colloids*, Amsterdam: Elsevier, (1948).

[51] Riess–Kautt MM, Ducruix AF, *Relative effectiveness of various ions on the solubility and crystal-growth of lysozyme*, J. Biol. Chem., **264**, 745–748, 1989.

[52] Stephane Veesler, Sylvaine Lafont, Sylvie Marcq, Jean Pierre Astier, Roland Boistelle, *Journal of Crystal Growth*, **168**, 124–129, 1996.

[53] Franks, G. V., *J. Colloid Interface Sci.*, **44**, 249, 2002.

[54] Franks, G. V., Meagher, L., *Colloids Surfaces A*, **99**, 214, 2003.

[55] D. F. Parsons, M. Bostrom,T. J. Maceina, A. Salis, and B. W. Ninham, *Why Direct or Reversed Hofmeister Series? Interplay of Hydration, Non-electrostatic Potentials, and Ion Size*, Langmuir, **26(5)**, 3323–3328, (2010).

[56] Finet S, Skouri-Panet F, Casselyn M, Bonnete F and Tardieu A, *Current Opinion in Colloid and Interface Science*, **9**, 112–116, 2004.

[57] M. Lund and P. Jungwirth, *J. Phys: Condens. Matter*, **20**, 494218 (2008).

[58] A. P. dos Santos and Y. Levin, *Phys. Rev. Lett.*, **106**, 1, (2011).

[59] Huan-Xiang Zhou, *Interactions of Macromolecules with Salt Ions: An Electrostatic Theory for the Hofmeister Effect*, PROTEINS: Structure, Function, and Bioinformatics, **61**, 69–78, (2005).

[60] M. Bostrom, F.W. Tavares , S. Finet , F. Skouri-Panet , A. Tardieu , B.W. Ninham, *Why forces between proteins follow different Hofmeister series for pH above and below pI*, Biophysical Chemistry, **117**, 217–224, (2005).

[61] K. D. Collins, G. W. Neilson, and J. E. Enderby, *Biophysical chemistry*, **128**, 95 (2007).

[62] Drew F. Parsons and Barry W. Ninham, *Importance of Accurate Dynamic Polarizabilities for the Ionic Dispersion Interactions of Alkali Halides*, Langmuir, **26(3)**, 1816–1823, (2010).

[63] B. Hribar, N. T. Southall, V. Vlachy, and K. A. Dill, *J. Am. Chem. Soc.*, **124**, 12302, (2002).

[64] S. A. Edwards and D. R. M. Williams, *Phys. Rev. Lett.*, **92 (24)**, (2004).

[65] P. Debye, E. Huckel, *Physikalische Zeitschrift.*, **24**, 185–206, (1923).

[66] Asakura, S., and F. Oosawa, *Interactions between particles suspended in solutions of macromolecules.*, J. Polym. Sci. [B], **33**, 183–192, (1958.

[67] Kim A. Sharp, *Analysis of the size dependence of macromolecular crowding shows that smaller is better*, PNAS, **112**, 7990–7995, (2015).

[68] Sarah C. Flores, Jaibir Kherb, and Paul S. Cremer, *Direct and Reverse Hofmeister Effects on Interfacial Water Structure*, J. Phys. Chem. C., **116**, 14408–14413, (2012).

Chapter 2

Protein–protein interaction at equilibrium

2.1 Introduction

Proteins are flexible and anisotropic biomolecules built from different types of amino acids such as hydrophobic, hydrophilic, charged etc. These amino acids play vastly different roles in intra– and inter–molecular interactions. A number of diseases such as type 2 diabetes[1], cataracts[2], neurodegenerative diseases[3] etc. are caused by protein aggregation. Also complexes such as virus capsid, microtubules and protein crystals are also possible outcomes of protein–protein interaction. To understand how protein–protein interaction lead to different outcomes, we study a thermodynamic quantity called "solubility" because it measures the strength of many body interactions. It has been measured experimentally that protein solubility is affected by several factors such as solution pH, salt concentration, salt type etc but it is poorly understood theoretically.

The electrostatic interactions between proteins has been studied previously[4–6] and it has been the subject of interest becasuse it plays a major role in protein–protein interactions. For simplicity in calculation, people often use only the zeroth order moment of charge (monopole) assuming charges are distributed isotropically on the protein surface. The higher order

charges are short ranged and they are less important when proteins are far from the isoelectric point. At those regions, the monopole charge is large and its repulsive strength dominates the attractive strength of higher order charges. But, near the isoelectric point, the net charge of protein is almost zero so the effect of higher order charge can be important.

In salt solution, the electrostatic interactions are screened by the counter-ions. To do this, ions have to lose some entropy when they form a screening layer around charged macromolecules. The importance of salt entropic contribution to the formation of protein aggregate has been shown by Schmit and Dill[7]. As a consequence of the screening effect, the repulsion of monopole charge as well as the attraction of higher order charge are weakened by the addition of salt. Screening effect of salt has the opposite effect on the monopole charge and higher order moments. By increasing the screening effect on the monopole charge, one can promote the aggregation of proteins which is a salting out mechanism, whereas in the case of higher order moments of charges the aggregation is supressed by the addition of salt which is a salting in phenomenon[4]. The Hofmeister effect (1888), or salt specific effect, was discovered more than a century ago but it is still not understood well theoretically. People have proposed number of salt specific interactions such as ion–protein dispersion interaction[10;13;14], solvophobic interaction[11], image charge interactions[12] as the cause of Hofmeister effect but none of them has accounted the salt specific entropy.

Here, our purpose is to design a theory that accounts many effects of salts and effect of pH on protein solubility. The free energy of a protein state, in our model, includes the Coulomb energy, salt entropy, and protein–ion binding free energy. From this model, we show that most of the free energy of protein aggregation is contributed by the salt entropy term. In addition to that we show that entropic term is salt specific if the excluded volume of salt is included. We model protein charge distribution using a first order spherical harmonic expansion which allows us to include both repulsive and attractive electrostatic interactions. The dominant region for repulsive or attractive electrostatic interaction relies on the values of solution pH and isoelectric point of model protein. In our model, we account for protein–ion association with the help of grand canonical partition function and also account for the effect of protein–ion binding on protein charge. We treat the ion as a finite sized entity and

use hydrated ion size to account the solvation effect implicitly. We also account for excluded volume effects in our model. Our model which includes screening and binding effect of salt and excluded volume effect is capable of capturing salting in/salting out and Hofmeister effect. This model is also capable of describing the reversal Hofmeister series seen while shifting salt concentration from low to high regime.

2.2 Model

Our model considers two states, namely, the solution state and the crystal state. At equilibrium, chemical potential in the solution state μ_s is equal to the chemical potential in the crystal state μ_c.

$$\begin{aligned} \mu_{\mathrm{c}} &= \mu_{\mathrm{s}} \\ F_{\mathrm{b}} + F_{\mathrm{c}} &= k_B T \ln c_0 + F_{\mathrm{s}} \\ F_{\mathrm{b}} + F_{\mathrm{c}} - F_{\mathrm{s}} &= k_B T \ln c_0 \end{aligned} \tag{2.1}$$

Where F_b is the salt independent protein–protein interaction in the crystal state, F_c and F_s are the salt dependent free energies in crystal state and solution state respectively. c_0 is the concentration of protein in solution state, and $k_B T \ln c_0$ represents the translational entropy of protein in solution state.

Solving for c_0, we get,

$$c_0 = A e^{\frac{(F_{\mathrm{c}} - F_{\mathrm{s}})}{(k_B T)}} \tag{2.2}$$

which is the concentration of protein in solution state and at equilibrium, it is defined as the protein solubility[20;21]. Here, $A = e^{F_{\mathrm{b}}/k_B T}$ is a constant that we use as a fitting parameter.

The salt dependent free energies are,

$$F_{\mathrm{s/c}} = E_{\mathrm{coul}} - TS_{\mathrm{salt}} + F_{\mathrm{bind}} \tag{2.3}$$

Where, E_{coul} is the Coulomb energy and TS_{salt} is the energy contributed by ion entropy at temperature T. F_{bind} is the free energy contributed by the binding of ions to the protein.

The Coulomb energy E_{coul} is,

$$E_{\mathrm{coul}} = \int \frac{\epsilon}{2} \mid \nabla\Psi \mid^2 d^3r \tag{2.4}$$

Where, ϵ is the position dependent permittivity and Ψ is the electric potential.

The free energy due to ion entropy is[8],

$$\begin{aligned} -TS_{\mathrm{salt}} &= k_BT[\int_V [c_+ \ln \frac{c_+}{c_s} - c_+ + c_s]d^3r + \\ &\quad \int_V [c_- \ln \frac{c_-}{c_s} - c_- + c_s]d^3r] \end{aligned} \tag{2.5}$$

Where $c_+ = c_s e^{-\frac{e\Psi}{k_BT}}$ and $c_- = c_s e^{\frac{e\Psi}{k_BT}}$ are the concentrations of positive and negative ions respectively. c_s is the concentration of ions at the bulk solution. The first term in the right hand side of equation (2.5) comes from the entropy of positive ions and the last term represents the entropy of negative ions.

When the total volume integration in the equation (2.5) is divided into ion excluded and accessible regions then it can be written as,

$$\begin{aligned} -TS_{\mathrm{salt}} &= k_BT[\int_{(V-V_{ex}^+)} [c_+ \ln \frac{c_+}{c_s} - c_+ + c_s]d^3r + \\ &\quad \int_{(V-V_{ex}^-)} [c_- \ln \frac{c_-}{c_s} - c_- + c_s]d^3r + \\ &\quad c_s(V_{\mathrm{ex}}^+ + V_{\mathrm{ex}}^-)] \end{aligned} \tag{2.6}$$

The last term in the right hand side of the equation (2.6) is contributed by excluded volumes, V_{ex}^{+} and V_{ex}^{-} where $c_{+/-} = 0$. The excluded volumes depend on the ion size requiring distinct values for each ion. The excluded volume is also a function of the protein size, shape, and the protein-protein separation, resulting in a depletion attraction. Importantly, we take the ion size to be an effective parameter that includes the effects of bound water, thereby implicitly capturing solvent structuring.

We compute F_{bind} term by writing the grand partition function

$$e^{-F_{\text{bind}}/k_BT} = \sum_{n_b} \frac{N_s! e^{(-E_s + n_b(\mu - E_{\text{bind}}))/k_BT}}{(N_s - n_b)! n_b!} \tag{2.7}$$

where N_s, n_b, E_s, $\mu = k_BT \ln c_s$, E_{bind} are the number of binding sites on the protein, number of bound ions, electrostatic protein-ion energy, ion chemical potential, and non-electrostatic protein-ion attraction, respectively. Due to the uncertainty in the location of the binding sites, we assume that the number of bound ions is equal in the soluble and aggregated states. Therefore, F_{bind} does not change upon aggregation and the contribution of this term is to modify the charge on the protein.

2.2.1 Protein in solution state is modeled as a sphere with monopole and dipole

In the solution state, we model the protein as charged sphere and surface charges are approximated by a first order multipole expansion $\rho(r, \theta) = (\sigma_0 + \sigma_1 \cos\theta)\delta(r - R)$ as shown in figure (2.1A). The zeroth order term represents the net charge $n_0 = 4\pi R^2 \sigma_0 / e$ of protein (monopole) in the unit of electronic charge and produces the repulsive interaction between proteins. The first order term represents the dipole $n_1 = 4\pi R^2 \sigma_1 / e = \frac{3p}{eR}$ charge in electronic units, which is the source of attractive electrostatic interaction. The dipole moment (p) of a protein is calculated based on how its charged amino acid are distributed from the centroid of protein. To evaluate the dipole charge (n_1), we match the dipole moment p to the first moment of charge distribution. The net charge (n_0) and dipole moment (p) at different pHs

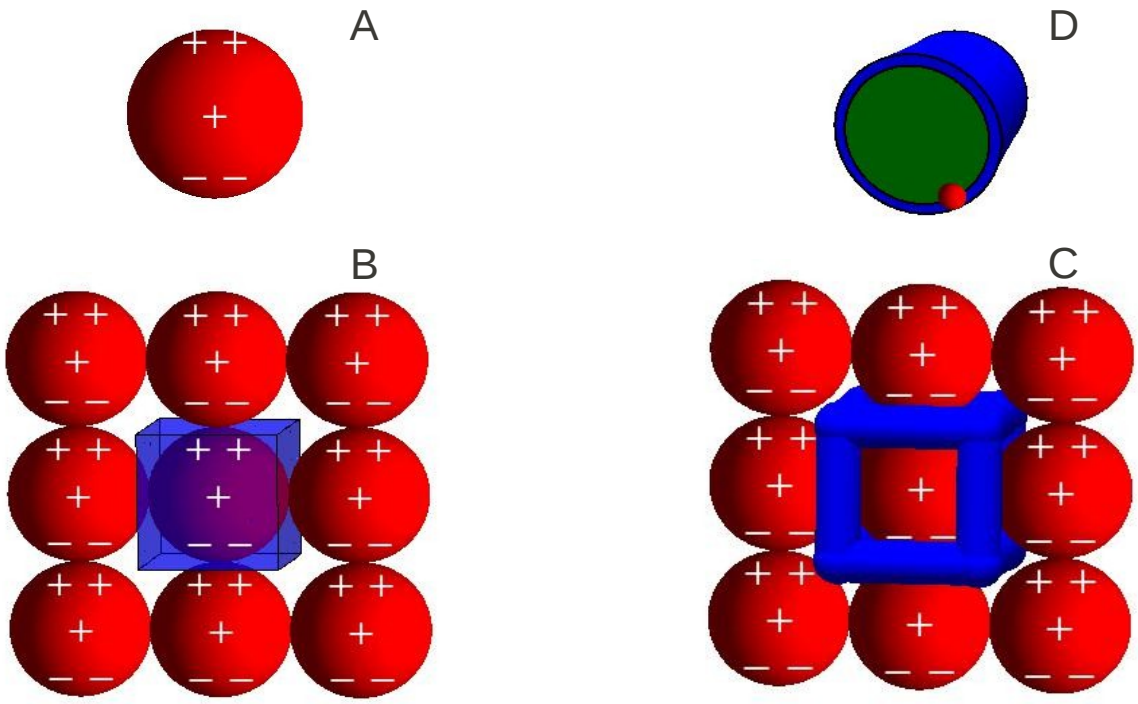

Figure 2.1: *Cartoon representation of the geometry used in our model. A) The protein is modeled as a sphere embedded in an aqueous environment. The charge distribution is described by a monopole and dipole, which is schematically shown as charges at the sphere center and poles, respectively. B) Each protein in the aggregate is surrounded by a Wigner cell consisting of the protein (red spheres) and surrounding water (blue). C) We approximate the surrounding water as cylindrical channels. D) The volume accessible to ions in the channels depends on the ionic radius. Smaller ions have a larger accessible volume (green).*

are calculated from the Henderson-Hasselbach equation using model pKa values of amino acids[44]. R is the radius of spherical protein which we calculate from the crystal structure of a protein using following relation,

$$R = \left[\frac{3abcp_c}{4\pi n_m}\right]^{\frac{1}{3}} \tag{2.8}$$

Where, a, b, c are the dimensions of a unit cell, n_m is number of protein molecules, p_c is the fraction of protein means $(1 - p_c)$ is the fraction of solvent in a unit cell.

We solve the linearized Poisson Boltzmann equation ($\nabla^2\Psi_{out} = \kappa^2\Psi_{out}$) and Laplace equation ($\nabla^2\Psi_{in} = 0$) in spherical geometry to find the electric potential outside Ψ_{out} and inside protein Ψ_{in} by using appropriate boundary condition. The first boundary condition says that the potentials at the surface of protein are equal.

$$\Psi_{in} \mid_{(r=R)} = \Psi_{out} \mid_{(r=R)} . \tag{2.9}$$

The second boundary condition is,

$$\epsilon_w \frac{\partial \Psi_{out}}{\partial r} \mid_{(r=R)} - \epsilon_p \frac{\partial \Psi_{in}}{\partial r} \mid_{(r=R)} = -(\sigma_0 + \sigma_1 \cos\theta) \tag{2.10}$$

Where, $\epsilon_w = 80\epsilon_0$ and $\epsilon_p = 4\epsilon_0$ are the permittivity of water and protein. ϵ_0 is the permittivity of vacuum. For the dielectric constants of water and protein we use 80 and 4, respectively.

By solving above equations, we get the electric potential outside the protein at distance r to be,

$$\begin{aligned} \Psi_{out}(r) = & \frac{n_0 e e^{-\kappa(r-R)}}{4\pi\epsilon_w r(1+\kappa R)} \\ & + \frac{n_1 e R(1+\kappa r)\cos\theta}{4\pi\epsilon_w r^2[\frac{\epsilon_p}{\epsilon_w}(1+\kappa R) + (2+2\kappa R+\kappa^2 R^2)]} \\ & \times e^{-\kappa(r-R)} \end{aligned} \tag{2.11}$$

The first term in Eq. (2.11) represents the potential due to monopole charge and the second term is the potential due to dipole, κ is the inverse Debye length and e is an electric charge.

Similarly, we find the electric potential inside the protein to be,

$$\begin{aligned}\Psi_{in} &= \frac{n_0 e}{4\pi\epsilon_w R(1+\kappa R)} \\ &+ \frac{n_1 e r(1+\kappa R)\cos\theta}{4\pi\epsilon_w R^2[\frac{\epsilon_p(1+\kappa R)}{\epsilon_w} + (2+2\kappa R+\kappa^2 R^2)]}\end{aligned} \tag{2.12}$$

These potentials are due to the bare charges which will get modified by ion binding events. We account for the ion binding events, particularly anion binding because many experiments and theoretical study have shown that anion interaction is much stronger with the protein surface than it is for cations[30–33;35].

The following expression, which utilizes Eq. (2.7), gives the total number of anions that bind to the protein surface.

$$n_b = \left.\frac{\partial F_{\text{bind}}}{\partial \mu}\right|_{\mu = k_B T \ln c_s} \tag{2.13}$$

with

$$E_s = \sum_{n=0}^{n_b - 1} \Psi(n_0 - n, n_1 - n)\,|_{(r=R,\theta=\frac{\pi}{3})}\,(-e) \tag{2.14}$$

Where $\Psi(n_0 - n, n_1 - n)\,|_{(r=R,\theta=\frac{\pi}{3})}$ is the electric potential which is the function of monopole and dipole charge modulated by the anion binding. The dipole correction is an approximation that assumes that anions bind primarily to the positive hemisphere, where $\theta = \pi/3$ gives a median value for the potential. $n = 0$ means there are no anions bound.

2.2.2 Crystal state is modeled as cylindrical channels surrounded by proteins

When proteins are in the crystal state, we assume that the dipoles align with each other cancelling their dipole charges and leaving only monopole contribution in the electrostatic interaction. The cancellation of asymmetric charge due to the protein alignment requires no calculation of the Couloumb energy inside the protein because the symmetric charge

distribution has no contribution to the energy inside the protein[9]. To neutralize the repulsive interaction of monopole charges and successfully form the electrically neutral protein crystal, couter–ions must be recruited in the cavities so that net system gets neutralized. The cavities in the protein crystal are in fact occupied by solvent and many salt ions. We need to calculate the energy and entropy in the cavities. For this purpose, We model each cavity as a cylindrical channel of solvent surrounded by the proteins. The surface to volume ratio in case of cylinder is better representation than the case of sphere in our model. As a rough approximation, we assume that each protein is surrounded by channels with a total length of $24R$, as would be expected in a crystal with cubic packing symmetry (see Fig. 2.1C). We further approximate these channels as cylinders with a radius, R_c, chosen to match the solvent content of the aggregate. Since each channel is surrounded by four proteins (see Fig. 2.1B,C), the solvent volume per protein is $6\pi R_c^2 R$. To obtain the ion accessible volume we use $R_c \to (R_c - R_{\text{ion}})$, where R_{ion} is the effective radius of the ion and its solvation shell. We estimate the radius of cavitiy (R_c) from the protein crystal data using following expression.

$$R_c = \left[\frac{abc(1-p_c)}{6\pi n_m R}\right]^{\frac{1}{2}} \tag{2.15}$$

We solve the linearized Poisson Boltzmann equation around a non-zero average potential (ϕ_0)[7] to find the potential in the cavity. The dimensionless potential within a protein aggregate, $\Phi = e\Psi/k_BT$, often exceeds the threshold $\Phi < 1$ for linearization of the PB equation. However, since the cavities are small, on the order of κ^{-1}, the variation in the potential is small. Under these conditions, it is an excellent approximation to linearize the PB equation around a nonzero potential ϕ_0. We assume that a cavity is made by two concentric cylinders trapping ionic solvent. The outside cylinder has radius R_c and the inside cylinder has radius R_{in}. To eliminate numerical issues at $r = 0$, the inner cylinder radius R_{in} is set to the small value 0.01Å in the numerical calculations. The linearized Poisson Boltzmann

equation around non-zero potential is,

$$\begin{aligned} \nabla_y^2 \Phi &= \sinh(\phi + \phi_0) \\ \nabla_x^2 \phi &= \phi + \tanh \phi_0 \end{aligned} \tag{2.16}$$

Where, $\Phi = \phi + \phi_0$, $y = \kappa r$ and $x = \sqrt{\cosh \phi_0} y$.

With the boundary conditions,

$$\begin{aligned} -\frac{d\phi}{dx}|_{x=\alpha} &= E_0 \\ \frac{d\phi}{dx}|_{x=\beta} &= 0 \end{aligned} \tag{2.17}$$

the solution of equation (2.16) in cylindrical coordinates is,

$$\Phi(x) = \frac{E_0(K_1[\beta]I_0[x] + I_1[\beta]K_0[x])}{[I_1[\alpha]K_1[\beta] - I_1[\beta]K_1[\alpha]]} - \tanh \phi_0 + \phi_0 \tag{2.18}$$

Where K and I are modified Bessel functions, $\alpha = \sqrt{\cosh \phi_0}\kappa R_c$, $\beta = \sqrt{\cosh \psi_0}\kappa R_{in}$ and the dimensionless electric field at the cylinder surface is

$$E_0 = \frac{(n_0 - n_b)e^2}{12\pi\epsilon_w R k_B T \alpha}. \tag{2.19}$$

The non-zero averaged potential is obtained by using the neutrality condition in the cavity as follows.

$$(n_0 - n_b) = -(\bar{c_+} v_+ - \bar{c_-} v_-) \tag{2.20}$$

Where v_+ and v_- are the accessible volume for positive and negative ions in the cavity respectively. The ion accessible volume in the cavity, $6\pi((R_c - R_i)^2 - R_{in}^2)R$, depends on the size of ion R_i. $\bar{c_+} = c_s e^{\frac{-e\phi_0}{k_B T}}$ and $\bar{c_-} = c_s e^{\frac{e\phi_0}{k_B T}}$ are the concentration of positive and negative ions inside the cavity in the presence of average potential.

After calculating energy and entropy in two states and finding the free energy there, we use equation (2.2) to find solubility of proteins at various salt concentration, pH and type of salts. Note that, due to the uncertainty in the location of the binding sites, F_{bind} term in solution state and crystal state are taken equal so its net contribution to the change in free energy is zero. To calculate the contribution of the ion excluded regions on the solubility, we use computer program Chimera[49] which evaluates the change in excluded volume in the soluble state and aggregate state. The excluded region of ions gets reduced upon the aggregation of protein by the amount which is equal to the value of overlapped volume $(V_{\text{ex}}^{s} - V_{\text{ex}}^{c})$.

To calculate the overlapped volume, we first load a protein of our interest in software using its PDB code (Protein Data Bank) then create crystallographic copies around $5\mathring{A}$ distance. Let the total number molecules be n_p. Then, we expand each molecule by the size of an ion of our interest and roll an ion over the surface of a molecule. This process measures the volume of an individual molecule (let it be V_1). After measuring the volume of individual molecules, we combine all the copies of molecules. Then, we again roll the same ion over the combined surface of molecules. This measures the total volume of complexes (let it be V_2). Finally, we calculate the volume difference in two states which provides us the value of overlapped volume $(V_{\text{ex}}^{s} - V_{\text{ex}}^{c} = n_p \times V_1 - V_2)$. This is the amount of volume that ions feel increased in the bulk solution and they gain entropy.

2.3 Result and Discussion

2.3.1 Theory compares well with experiment to describe salting in, salting out and ion specific effect

The solubility is a sensitive function of the solvent content within the aggregate state. To facilitate the comparison of our theory with experiments, we focus on crystalline aggregates where the solvent content is readily obtained from the atomic structure. Protein and cavity radii are chosen to match the protein and solvent volumes reported in the crystal structures of

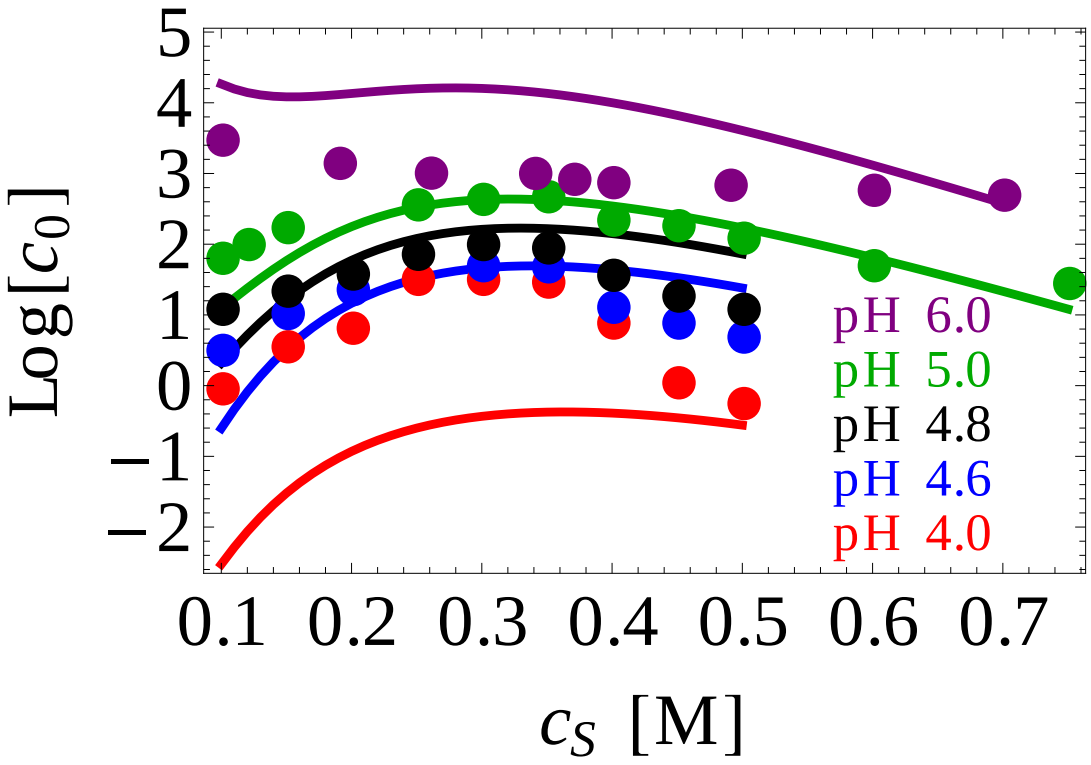

Figure 2.2: *Comparison of chymosin solubility (points)*[41] *to the theoretical model (lines) as a function of salt concentration c_s. The model captures the transition from pure salting-out at pH=6 to non-monotonic solubility at lower pH. $R = 23.3$Å, $R_c = 8.6$Å, $\Delta V_{\rm ex}^{\rm Na} = 4800$Å^3, and $\Delta V_{\rm ex}^{\rm Cl} = 3200$Å^3.*

Lysozyme[47] and Chymosin[48]. Following Eqs. 2.8 and 2.15, we found the radius of chymosin and lysozyme to be 23.3$\mathring{A}$ and 16.1$\mathring{A}$ respectively and the corresponding cylindrical cavity sizes to be 8.6$\mathring{A}$ and 6.0$\mathring{A}$.

We have shown the comparision between experiment[41] and our theory in figure (2.2). This is the solubility of chymosin protein near its isoelectric point in $NaCl$ solution. The hydrated ion radii used for sodium and cloride are $R_{Na^+} = 1.67\mathring{A}$[42] and $R_{Cl^-} = 1.50\mathring{A}$[43] respectively. For this comparision, we have selected anion binding sites to be $N_s = 36$ and binding affinity to be $E_b = -0.3k_BT$. We will describe the reason why we selected these values of N_s and E_b later when we compare our another result with experiment. The fitting parameter required in our main expression of solubility is prefactor A which is obtained to be $A_{\rm Chy} = 2.12$ mg/ml for chymosin protein. Theory compares well with experiment at all

pHs. Both experiment and theory predict salting in followed by salting out in pHs 4.0, 4.6, 4.8 and 5.0. In pH 6.0, experiment and our theory both predict salting out. The solubility strongly depends on the pH of the solution because charge of the protein changes with the pH.

The charge of chymosin protein at multiple pHs are tabulated in table (2.1) calculated using the Henderson-Hasselbach equation. For the calculation of charge, we use pKa values of amino acids from Ref.[44]. Since there is little variation in the dipole parameter n_1, we use the average value ($n_1 = 17.65$) at all pHs, which is equivalent to a moment of 653 Debye (137 eÅ).

Table 2.1: *Charge of chymosin vs pH*

pH	Monopole charge(n_0)	Dipole parameter(n_1)
4.0	4.77	17.49
4.6	−2.4	17.47
4.8	−3.9	17.62
5.0	−5.04	17.74
6.0	−9	17.92

We have also shown the comparison of lysozyme solubility at pH 4.5 between our theory and the experimental result by Ries-Kautt[45] in the Fig.(2.3). At pH 4.5, the net charge of protein is $n_0 = 10$ and dipole parameter (n_1) is 7. Here, the common anion is cloride and its radius is taken to be 1.50$\mathring{A}$. The cations used in the comparison are Na^+, K^+ and NH_4^+ with sizes 1.67$\mathring{A}$, 1.50$\mathring{A}$ and 1.25$\mathring{A}$[43] respectively. Note that, the hydration effect reverses the relative size rankings of sodium and potassium demonstrating the importance of water-ion interactions in ion specificity[46].

The solubility decreases in the order $NH_4^+ > K^+ > Na^+$ which agrees with the direct Hofmeister series and there is good aggreement between theory and experiment. Here, we can notice that the big co–ions are more favorable for aggregation than small co–ions. The binding sites used in lysozyme protein for anion is $N_s = 18$. We will explain the reason of using $N_s = 18$ in the ion binding subsection later. The parameters that require fitting are the prefactor A and E_{bind}, the non-electrostatic anion-protein interaction. $A_{\text{Lys}} = 15.64$

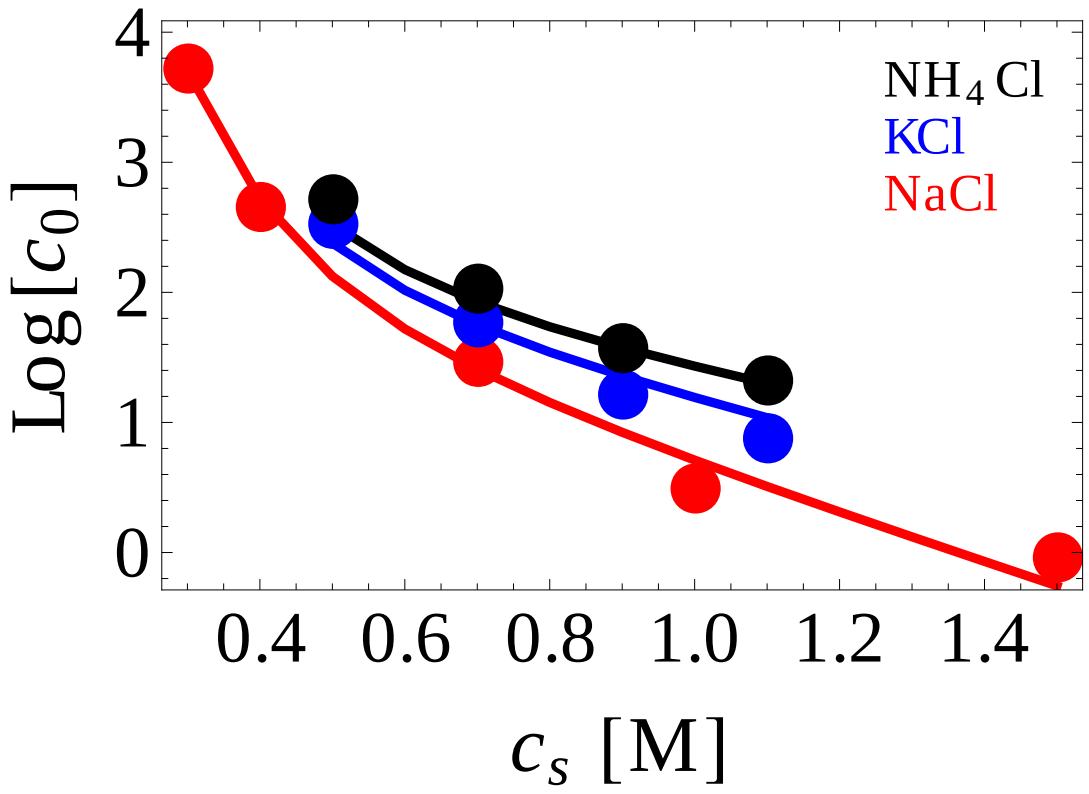

Figure 2.3: *Comparison of lysozyme solubility (points)*[45] *to the theoretical model (lines) at pH* 4.5. *The ion specificity comes from the preferential exclusion of large coions from the crystal interior.* $R = 16.1\text{Å}$ *and* $R_c = 6\text{Å}$.

mg/ml and $E_{\text{bind}} = -0.3k_BT$ are obtained by fitting to lysozyme solubility (Fig. 2.3). We used the same value $E_{\text{bind}} = -0.3k_BT$ in chymosin protein because in both proteins the binding species is Cl^-.

The salting in/salting out and salt specific effects are a result of several competing terms in the free energy. In the following sections we examine these contributions individually.

2.3.2 Electrostatic mechanism

2.3.3 Monopole repulsion, an electrostatic source of salting out, is dominated by counter-ion entropy

To simplify our analysis of the contributions to the free energy, we begin by examining a version of the model in which the ion-specific features of ion excluded volume and ion-protein binding have been removed. This model is equivalent to a Poisson-Boltzmann analysis and gives a view of the behavior expected from purely electrostatic interactions.

The first plot in Fig.(2.4) shows the variation of solubility with respect to zeroth order or monopole charge (n_0) at various salt concentrations with a negligible dipole term ($n_1 = 0$). The solubility of protein becomes minimum at a point when net charge of protein is zero. This point is known as the isoelectric point. While going away on either side of the isoelectric point, both solubility and charge of protein are increased. The solubility also depends on the salt concentration. The solubility decreases if we increase the salt concentration keeping the monopole charge constant. This means that the screening effect of salt on electrostatic repulsion is one cause of salting out. This figure reproduces the well known behavior of protein solubility with pH and salt concentration which tells that the solubility is minimum at the isoelectric point and screening effect of salt on net charge of protein causes salting out[25;26].

Coulomb energy and salt entropy are the electrostatic contributions to the free energy ($F = E_{coul} - TS_{salt}$) as shown in equation (2.3). In Fig.(2.4)B, we have investigated the contribution of Coulomb energy and salt entropy to the change in free energy between

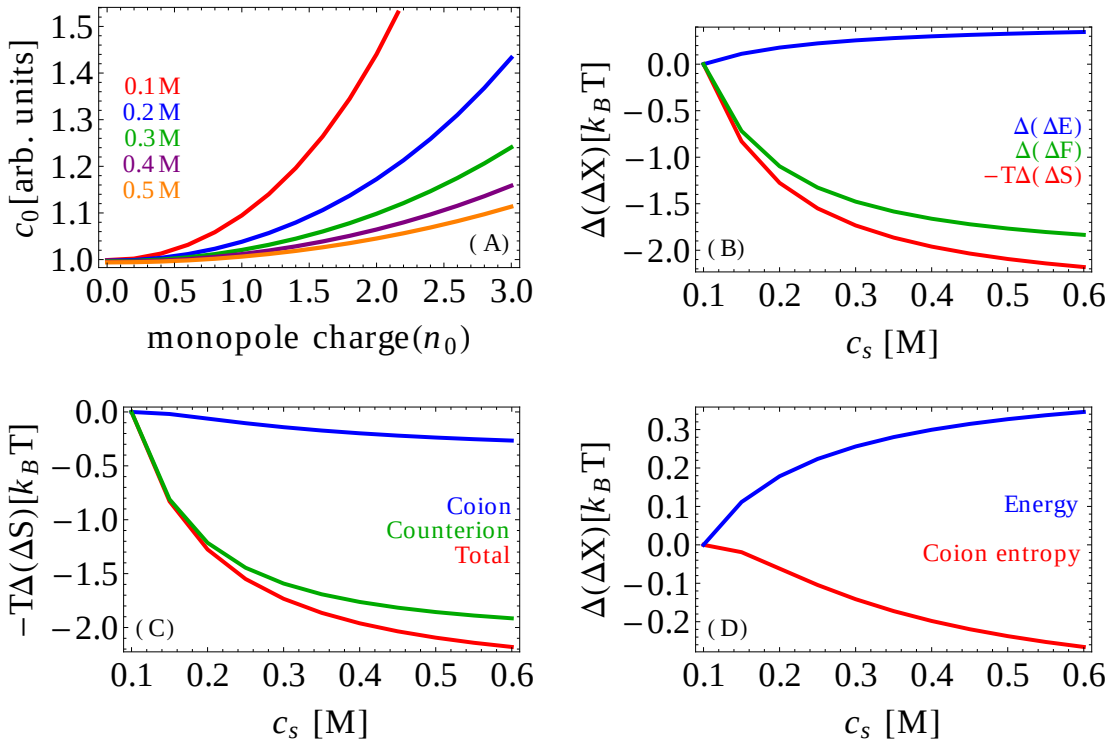

Figure 2.4: *A) Solubility of a charged spheres as a function of monopole charge n_0 and salt concentration. The dipole moment and non-electrostatic effects have been omitted ($\sigma_1 = N_s = E_{\text{bind}} = R_i = 0$). Repulsion between proteins stabilizes the solution state and increases the solubility. Adding salt screens the repulsion and leads to salting-out. B) Change in the Coulomb energy, entropy, and free energy of aggregation relative to 100 mM salt, $\Delta(\Delta X) = \Delta X(c_s) - \Delta X(0.1\text{ M})$, for spheres with charge $n_0 = 5$. The repulsive interaction is dominated by the ion entropy, so adding salt leads to a large decrease in the entropy penalty. C) The salt entropy can be further separated into coion and counterion terms demonstrating that the dominant contribution comes from the confinement of counterions. D) The salt entropy contributed by co–ions is energetically almost equal to the Coulomb energy of the sytem but their nature is just opposite.*

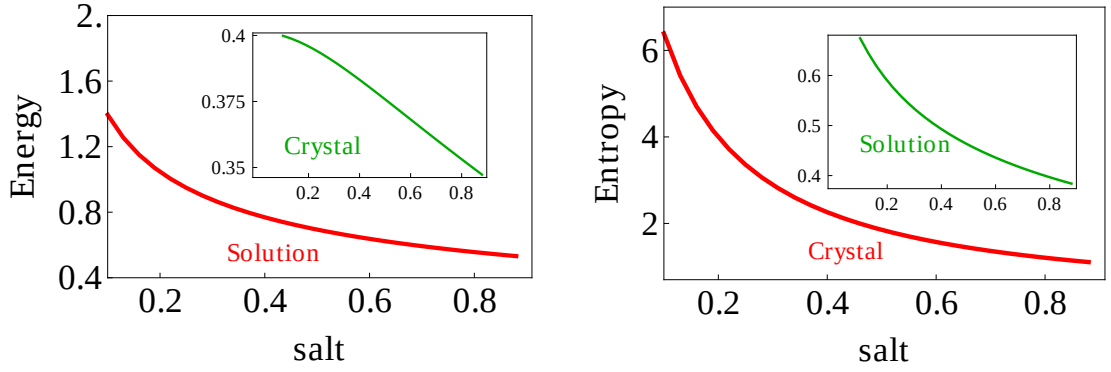

Figure 2.5: *(Left) Coulomb energy in solution state (inset: crystal state) as a function of salt concentration. Coulomb energy in solution state dominates its counterpart in crystal state. (Right) Entropy of salt ions in crystal state (inset: solution state) as a function of salt concentration. Entropy in crystal state dominates its counterpart in solution state. The dipole moment and non-electrostatic effects have been omitted ($\sigma_1 = N_s = E_{\text{bind}} = R_i = 0$). Charge in both cases are equal to $n_0 = 5$ and radius of protein and crystal cavity are taken to match with the lysozyme.*

the aggregate and solution states. It tells us that the electrostatic origin of salting out is dominated by the salt entropy because ions are required to be confined within the aggregates to form a crystal[7;9]. Since counter–ions are required for the crystal neutrality, the total salt entropy is mostly contributed by counter-ions entropy which is shown in Fig.(2.4)C. There is finite probability to find co–ions in the cavity which means co–ions also have to lose their entropy. The contribution of energy term in free energy is so small that its magnitude is comparable to the contribution of co–ion entropy (Fig.(2.4)D).

From the Fig.(2.4)B, we knew that the entropy change and Coulomb energy change show distinct variation with the salt concentration. The former becomes more attractive and the later becomes more repulsive with the addition of salt. To analyze it, we have plotted Fig. (2.5) which compares the contribution of energy and entropy in each of solution state and crystal state when protein charge is modeled up to monopole term only. The magnitude of energy term (first figure) in solution state dominates the counterpart in the crystal state. This is because the potential in the crystal state varies slowly with the distance due to the lack of enough space in comparison to the variation of potential with distance in the solution state. This produces the slow variation of energy with respect to salt in crystal state and

fast variation of energy in the solution state. So, the difference between the crystal state energy and solution state energy is attractive at low salt and becomes less attractive with the addition of salt. Due to this reason, we see the repulsive behavior of energy term with salt in Fig.(2.4)B.

On the other hand, the entropic penalty is greater in the crystal state than it is in the solution state. It is because ions have to lose more entropy while being confined in a small cavity in crystal state than the amount of entropy they lose while forming the screening layer around protein in soluble state. Thus, the entropy difference between the crystal state and solution state is dominated by the entropy in the crystal state. Furthermore, it is repulsive. With the addition of salt, it becomes less repulsive and more favorable for protein aggregation. Also, from the figure, it is seen that the entropy in the crystal state dominates other components of the free energy such as the entropy in solution state, the Coulomb energy in the aggregate state, and the Coulomb energy in the solution state. For this reason, the solubility of a protein having only monopole charge is dominated by the translational entropy of salt ions (Fig.(2.4)B).

2.3.4 Dipole attraction leads to salting in and it is dominated by energetic term

In Fig.(2.6)A, we have plotted the solubility with respect to dipole charge (n_1) at various salt concentrations. In this case, the monopole charge is negligible ($n_0 = -0.01$). The dipole has the opposite effect of the monopole on the solubility. The solubility decreases with the increase in dipole charge at a fixed salt concentration. The increased dipole charge means the increased attractive strength and the increased attractive strength is favorable for the aggregation. If the dipole moment is kept constant and the salt concentration is increased then the solubility increases due to the screening effect of salt on the dipole attraction. The electrostatic origin of salting in is screened dipole attraction.

The second plot shows the contribution of energy (blue line) and entropy (red line) to the change in free energy (green line). Another difference between the dipole and the monopole

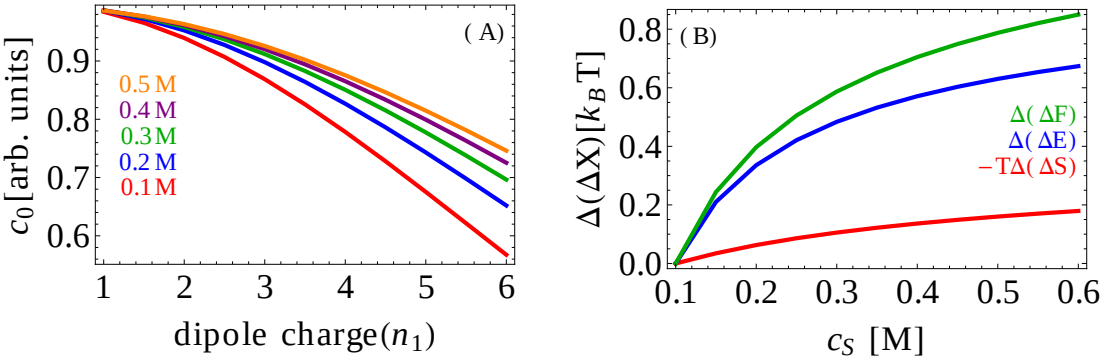

Figure 2.6: *Electrostatics-only model ($N_s = E_{\text{bind}} = R_i = 0$) showing the solubility of a nearly ideal dipole ($n_0 = -0.01$). A) Dipole attraction decreases the solubility and the addition of salt increases solubility. B) Variation of the energy, entropy and free energy of aggregation ($\Delta(\Delta X) = \Delta X(c_s) - \Delta X(0.1\text{ M})$) of a pure dipole ($n_1 = 10$). Both the energy and entropy are favorable for aggregation and become less favorable with the addition of salt.*

interaction is that both energy and entropy have same type of contribution to the free energy. This is because the association of proteins leads to both the release of counterions and the close association of complementary charges between proteins. In the dipole only model, the free energy is dominated by the Coulomb energy which is shown in Fig.(2.6)B unlike in monopole only model in which entropic effect of counter–ions dominanated change in free energy.

2.3.5 Competition of monopole repulsion and dipole attraction

In our study, we have expanded the charge distribution of the protein up to first order. Remember that the zeroth order charge represents the net charge of protein and the first order represents the dipole moment. We intend to apply our theory to explain salting in and salting out of protein. Since the protein monopole and dipole moments have opposite effects on the solubility, it is important to determine whether salting-in or salting-out or combination of them will occur. The combined effect of zeroth and first order charge is plotted in Fig.(2.7).

Fig.(2.7) shows the solubility of protein with respect to the mixed charge distribution at different salt concentration. For this purpose, we have kept dipole charge fixed at $n_1 = 5$

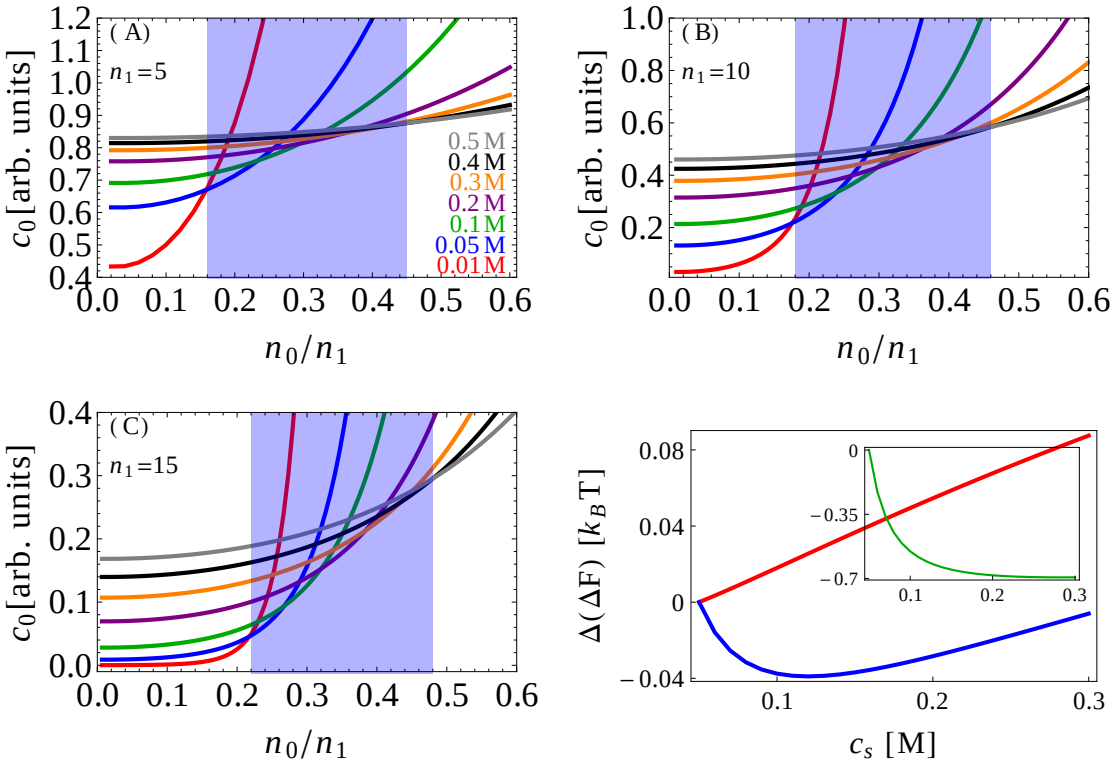

Figure 2.7: *Competition between the monopole and dipole leads to salting-in and salting-out in an electrostatics only model ($N_s = E_{\text{bind}} = R_i = 0$). (A) Solubility vs ($\frac{n_0}{n_1}$) at $n_1 = 5$. (B) Solubility vs ($\frac{n_0}{n_1}$) at $n_1 = 10$. (C) Solubility vs ($\frac{n_0}{n_1}$) at $n_1 = 15$. We see pure salting in (left to the shaded region), pure salting out (right to the shaded region) and salting out–salting in trends (shaded region). These crossover points depend weakly on the magnitude of the dipole. In the last figure, three different outcomes are shown by choosing different values of $\frac{n_0}{n_1}$. When $\frac{n_0}{n_1}$ is 0.2 then it gives pure salting in (red colored line). The values of $\frac{n_0}{n_1}$ equal to 0.3 and 0.4 give non–monotonic result (salting out followed by salting in shown in blue color) and monotonic salting out (shown in inset) respectively. The radius of protein is 23.3Å and radius of solvent cavity is 8.6Å.*

(Fig. A), $n_1 = 10$ (Fig. B) and $n_1 = 15$ (Fig. C) and changed the monopole charge to express the mixed charge distribution in the ratio of monopole charge to the dipole charge n_0/n_1 at different salt concentrations. In all graphs, we observe three distinct regions, namely, only salting in, only salting out and salting out followed by salting in. At lower values of $|n_0/n_1|$ roughly below 0.19 ± 0.03, the solubility of protein increases with the increase in salt concentration which is the salting in region. In this region, the monopole charge is weak so the dipole attraction is stronger than the monopole repulsion and we observe pure salting in. This means that the dipole charge has to be substantially bigger than the monopole for salting-in to occur. Also note that, the cross–over point of salting out from salting in depends on the salt concentration. Here, the lowest salt concentration we have studied is $0.01M$ and the highest salt concentration is $0.5M$ which is reasonable for the electrostatic interaction.

When the value of n_0/n_1 is more than 0.46 ± 0.02 then we see only salting out region. It means the monopole charges are strong enough that the repulsive interaction dominates the attractive interaction produced by the dipole charge. But, the solubility is non–monotonic when the value of n_0/n_1 lies in the region between 0.2 and 0.45. In this region, both monopole and dipole are not strong enough to show dominant effect. In this case, we observe salting out of protein at low salt and it is followed by the salting in at high salt. The salting out lasts for narrow range of low salt concentration than the salting in but it becomes wider as the value of n_0/n_1 is increased and finally yielding pure salting out.

In the plots, we have used three values of n_1 which are 5, 10 and 15 and changed the monopole charge to get different values of ratio n_0/n_1. The value of ratio (n_0/n_1) up to which we get pure salting in, salting out–salting in and pure salting out depends weakly on the value of dipole charge. For example, if $n_1 = 5$ then we see salting in up to $n_0/n_1 = 0.16$. This means that the monopole charges up to 0.8 are insufficient to dominate over a dipole charge equal to 5. If we increase the dipole charge by three fold then we see pure salting in up to the value of ratio $n_0/n_1 = 0.22$ which means monopole charge is 3.3 which is greater than the three fold value of previous monopole charge 2.4 but the difference in monopole charge is less than a unit charge. So, the ratio of n_0/n_1 can be useful to predict the conditions to

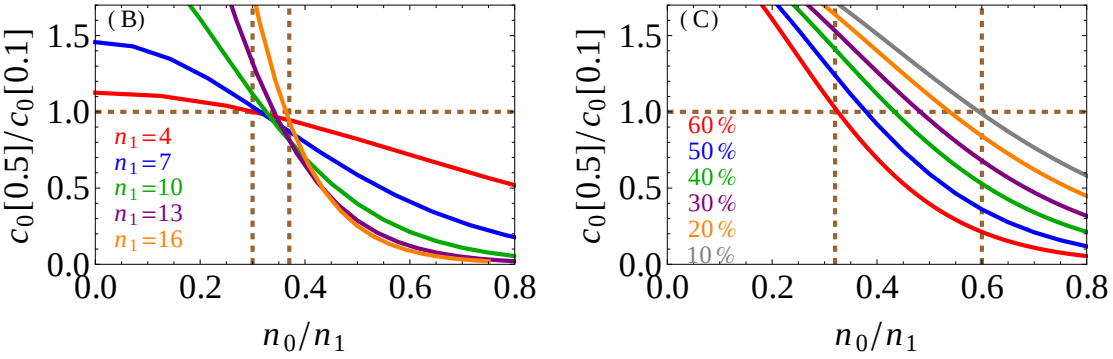

Figure 2.8: *(B) Variation of* $\frac{c_0[0.5M]}{c_0[0.1M]}$ *with respect to* $\frac{n_0}{n_1}$ *at various dipoles which shows that the cross-over of salting out from salting in at different cases are almost at the same point around* 0.3. *(C) Variation of* $\frac{c_0[0.5M]}{c_0[0.1M]}$ *with respect to* $\frac{n_0}{n_1}$ *at various protein volume fraction choosing dipole charge to be* 10 *which shows that the cross-over of salting out from salting in shifts towards the larger value of* n_0/n_1 *if the volume fraction of protein is decreased. Ratios greater than unity indicate salting-in over this range of salt concentrations. As the protein volume fraction increases from 10% to 60%, smaller monopoles are required for salting out to dominate. The radius of protein is* $23.3\mathring{A}$ *and radius of solvent cavity is* $8.6\mathring{A}$.

achieve a monotonic and a non–monotonic behavior of solubility.

In the last plot of Fig.(2.7), we have selected three values of n_0/n_1 from three distinct regions, namely, from left of the shaded region, shaded region, and right of the shaded region to show three different results of solubility with respect to salt concentration. As expected, the lower value of the ratio ($n_0/n_1 = 0.2$) choosen from the left side of shaded region shows only salting in. The intermediate value (0.3) selected from the shaded part shows the non–monotonic solubility showing salting out followed by salting in and the higher value of n_0/n_1 (0.4) picked from right side of the shaded part in the figure shows the pure salting out. This result indicates that the electrostatic interactions can lead to either monotonic or non–monotonic behavior of solubility depending on the strengh of terms in the expansion of charge.

2.3.6 Shift of salting out cross–over point with dipole charge and volume fraction of protein

For the further study of salting in and salting out with respect to the mixed charge distribution of protein, we have plotted Fig.(2.8). In Fig.(2.8B), we have plotted the ratio of two solubilities with respect to the ratio of monopole to dipole at various dipole charges. To test the cross-over point of salting out from salting in, we have taken the ratio of two solubilities at salt concentrations $500mM$ and $100mM$ and observe the variation with respect to the ratio of n_0 to n_1 for various dipoles ranging from 4 to 16 which are indicated by different colors in the figure. Here, the value of ratio $\frac{c_0[0.5M]}{c_0[0.1M]}$ on the vertical axis greater than 1 indicates salting in and less than 1 indicates salting out. The result is either salting in or salting out depending upon the values of n_0 and n_1. Near the isoelectric point, n_0 is small resulting small n_0/n_1. In this situation, the monopole charges are small so the dipole interaction plays dominant role giving salting in. On the other hand, far from the isoelectric point, n_0 is large so the ratio n_0/n_1 is also large. In this case, the monopole interaction dominates producing salting out. This plot indicates that the cross–over point of salting out from salting in falls around ($n_0/n_1 = 0.3$) at 60% protein volume fraction. This means that the dipole charge has to be substantially larger than the monopole for salting-in to occur. This plot also shows that the cross–over point depends weakly on the magnitute of dipole.

In Fig.(2.8C), we have shown the effect of volume fraction on the salting in/salting out cross–over point by changing the volume fraction from 10% to 60% choosing dipole parameter (n_1) to be 10. Here we observed that the salting out cross–over point shifts towards higher values of n_0/n_1 if the volume fraction of protein is decreased. The value of cross-over point increases from 0.33 to 0.60 when the volume fraction of protein is reduced from 60% to 10%. The salting out transition point depends strongly on the volume fraction of protein in the aggregate. The salting in/salting out cross–over point varies due to the distinct mechanisms underlying the dipole and monopole interactions. The dipole interactions are driven by the energetic gain of pairing charged patches while the monopole repulsion is dominated by counterion confinement entropy. Therefore, increasing the solvent content of the aggregate

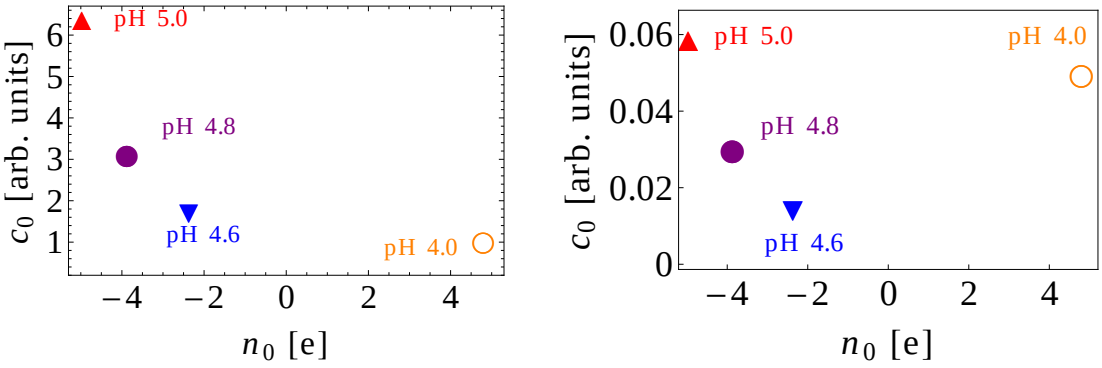

Figure 2.9: *Measurement and calculation show the different values of pH related to solubility minima. (Left:) Measured solubility minima is at pH* 4.0[41]*. (Right:) Calculated solubility minima is at pH* 4.6*. In both cases salt concentration is* $0.1M$*. These plots suggest that non–electrostatic effects are needed to explain the shift in the solubility minima to lower pH value.*

(means decreasing the protein volume fraction) will reduce the monopole repulsion with minimal effect on the dipole attraction giving higher values of the cross–over point.

2.3.7 Non-electrostatic mechanism

Protein-ion interaction shifts isoelectric point

Salt ions, mostly counter ions, build the screening layers around the charged protein which help to weaken electrostatic interactions. In addition to building the screening effect, salt ions have significant probability to bind to the protein. This effect is important for our modeling of chymosin, which has a calculated isoelectric point closer to pH of 4.6 while the measured solubility minimum is closer to 4.0[41]. The solubility minimum is at different pH in experiment and theory which is shown in Fig. (2.9).

The result of ions binding to the surface of protein will be directly seen on the electrostatic interaction because ion binding to the protein shifts its isolectric point. The shift of the isoelectric point due to anion binding is shown in Fig.(2.10). For the purpose of showing the effect of anion binding on the protein charge, we have selected binding sites on the protein surface to be 15 and varied the binding affinity. The stronger the binding affinity the more bound ions there are.

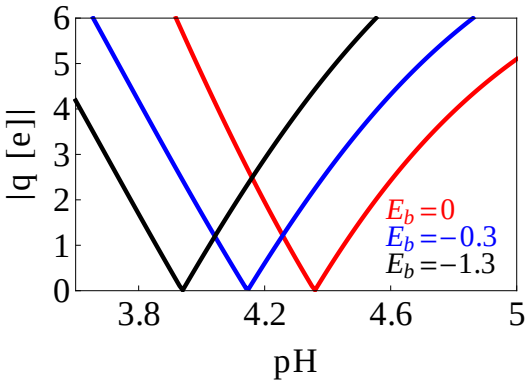

Figure 2.10: *Anion binding shifts the isoelectric point of protein. Magnitude of protein charge with respect to solution pH using different values of anion binding affinity. The binding affinities are shown in the figure. The strong binding affinity increases the number of bound anions as a result of which pI shifts towards lower values of pH.*

In fact, the amount of charge perturbation depends on the number of ions bound to the protein. The dipole moment perturbation depends on both the number of anion bound and the location of binding. There are no such standard results which suggest the number of bound ions on the protein surface at certain condition. However, there are multiple experiments on lysozyme protein reporting the number of binding sites with varying results such as 35 ± 7[56], 9 ± 4[57], 7 ± 2[58], 22 ± 1[59].

For the comparision of theory to experiment, we calculated the average number of binding sites from these varying results and used it in the study of lysozyme protein. The average value of binding sites is found to be ($N_s = 18$) from experiments cited above for lysozyme. For the anion binding study of chymosin protein, we used ($N_s = 36$). For chymosin protein, in our knowledge, there are no such experimental evidences to give binding sites on its surface. Due to unavailability of binding sites data for this protein, we estimated the number of binding sites on its surface. The chymosin protein has nearly double surface area than the lysozyme protein. So, by comparing the surface area of lysozyme and chymosin, we used ($N_s = 36$) for chymosin which is double the binding sites of lysozyme.

The inclusion of anion binding to our theory helped us to shift the isoelectric point from pH 4.6 to pH 4.0. The solubility trends with respect to pHs before and after anion binding inclusion are shown in Fig. (2.11).

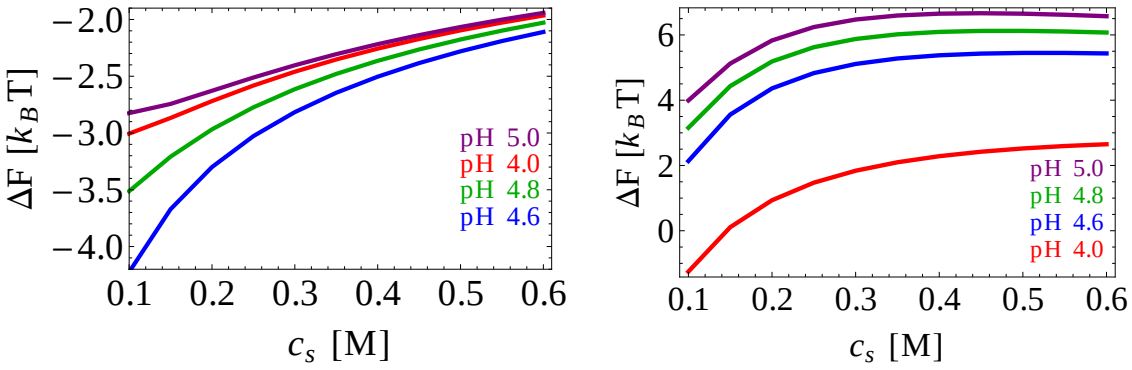

Figure 2.11: *Effect of anion binding on protein solubility. (Left:) Calculated solubility without accounting anion binding. The solubility minimum is at pH* 4.6*. (Right:) Calculated solubility accounting anion binding. The solubility minima is changed from pH* 4.6 *to pH* 4.0*. Salt concentration used in both cases is* $0.1M$*.*

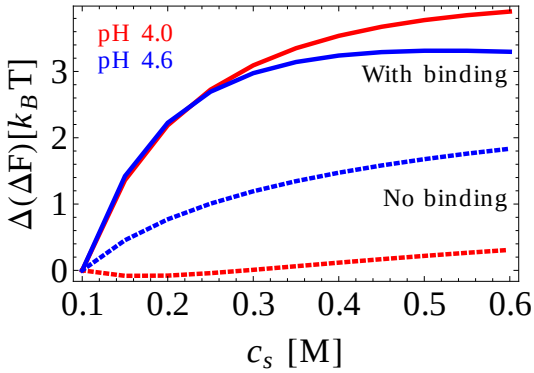

Figure 2.12: *Free energy change of chymosin protein* ($\Delta(\Delta F) = \Delta F(c_s) - \Delta F(0.1\text{ M})$) *with respect to salt at pHs* 4.6 *and* 4.8 *considering (solid line) and without considering (dashed line) anion binding. For anion binding* $N_s = 36$ *and* $E_b = -0.3k_BT$*. Anion binding promotes salting in in these pHs.*

Anion binding affects electrostatic interaction by making charge dynamic with salt concentration

Here, we perform the case study of chymosin's solubility with and without anion binding. Figure (2.12) shows the effect of anion binding on the electrostatic interaction. Solid lines in the plot shows the electrostatic interactions at various pH with respect to the salt concentration after anion binding whereas the dashed line shows the the variation of free energy in the absence of anion binding in the respective pH. Without anion binding, there is weak

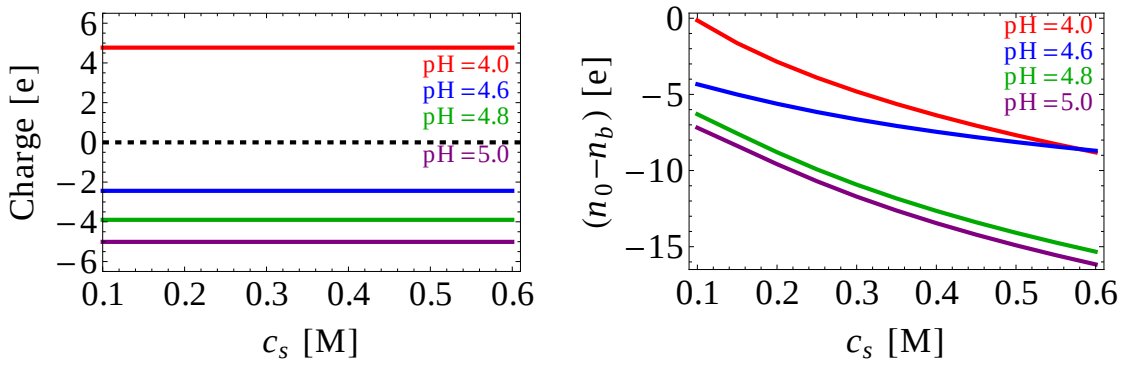

Figure 2.13: *Charge of chymosin protein vs salt concentration. Figures represent charges vs salt with and without ion binding at different pHs. Charge remains uniform with changing salt concentration if no binding is happened (left panel). With ion binding, the charges change with the salt concentration leading to the charge reversal or over–charging (right panel).* $N_s = 36$, $E_b = -0.3k_BT$.

salting in effect but in the presence of anion binding the salting in effect is enhanced. This demonstrates that anion binding can enhance the monopole repulsion.

Anion binding to the protein enhances salting in which is true only in certain conditions. One suitable condition to enhance salting in due to anion binding is that a protein should be at or below its isoelectric point at the start of binding event. For the enhancement of the salting in, the net monopole charge should increase due to the ion binding. For this purpose, at least one of the following mechanisms are essential. If protein has net charge zero (isoelectric point) before binding then the addition of few anions on the protein surface make it negatively charged and with the further addition of salt make it more negatively charged which can lead to the salting in phenomenon. In another case, if protein is initially negatively charged (above isoelectric point) then its net charge increases due to the anion binding and the overcharging happens with the further addition of salt. The overcharging can promote salting in but only at low salt regime. For instance, at high salt concentration (≈ 0.5), the screening effect of monopole charge is so strong that it can lead to the salting out even if the monopole charge is increased. In addition to monopole role, if the dipole moment of protein gets reduced by the binding process then it also enhances salting in.

The effect of anion binding on the net charge of chymosin protein at various pHs are

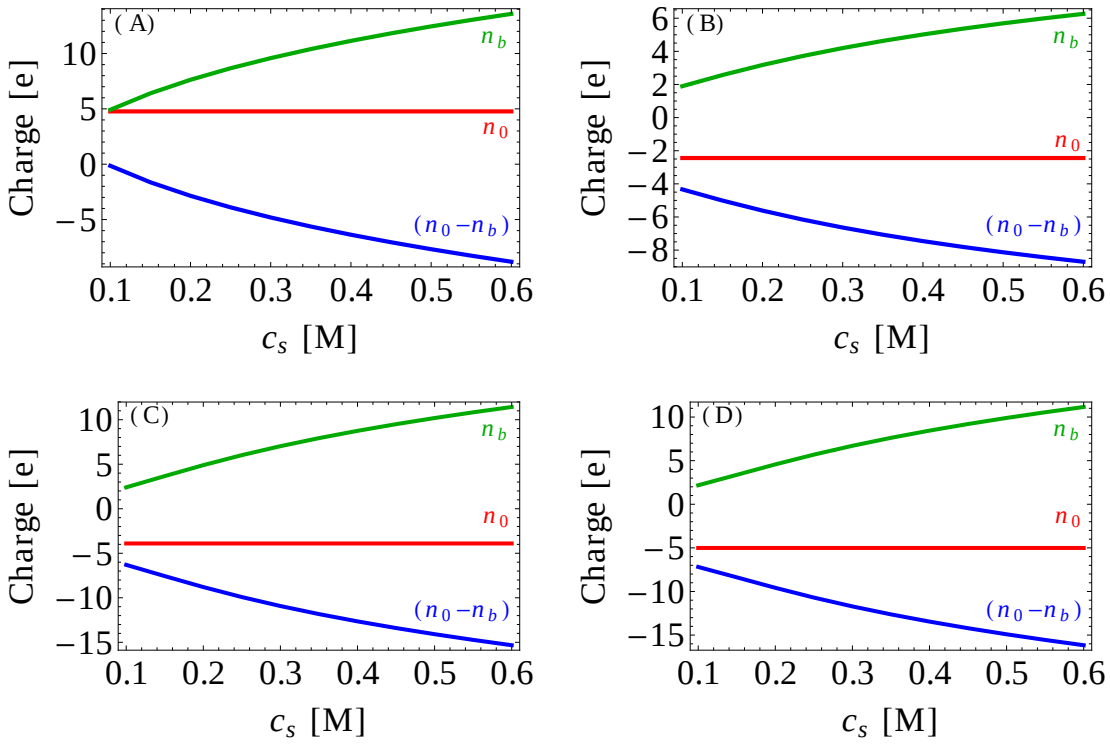

Figure 2.14: *Charge of chymosin protein vs salt concentration. Figures represent pHs* 4.0 *and* 4.6, 4.8 *and* 5.0. *The red horizontal line in each figure is the monopole charge of protein. The green line is the number of anion bound on protein and the blue line is the net charge of protein after anion binding.* $N = 36$, $E_b = -0.3 k_B T$.

shown in Fig.(2.13). Before considering the anion binding, the protein charges are constant at a pH in entire salt concentration range which is shown in the first graph. At this situation, the addition of salt has only screening effect on the charge. But, the consideration of anion binding adds one more additional effect of salt which is to make charges of protein dynamic with the salt concentration at a pH in addition to its screening effect.

The addition of salt will not always weaken electrostatic interaction if the effect of the increased protein charge is stronger than the screening effect. In chymosin protein, anion binding is enhancing salting in effect because at low salt regime, net charges of protein are increasing in such a way that the screening effect isn't sufficient to keep up with the increased charge. Due to the increased electrostatic repulsion with the increase in salt concentration, the salting in behavior is enhanced. In chymosin protein, the electrostatic repulsion is increased with salt concentration because of two phenomenons happening together due to anion binding. The first is the enhanced monopole repulsion and the second is the weakening dipole attraction.

The binding of ions to the protein upon the addition of salt alters the protein monopole charge as well as its dipole moment, which can have dramatic effects on the solubility. Since anions are the binding species, the net positive charge of the protein will decrease with added salt. This means that the magnitude of the protein charge will increase if the pH is above the isoelectric point. Fig. 2.14 shows this increase in the protein charge along with the beginnings of a saturation at high salt as the sites become occupied. Between the saturation of binding sites and the enhanced screening at high salt, the effects of ion binding on the solubility are confined to the low salt regime (< 0.3 M). This is demonstrated in Fig. 2.12 which shows that the inclusion of ion binding dramatically enhances the salting-in effect below 0.3 M salt. This is because the addition of charge to the protein happens faster than the enhancement of the screening effect. At higher salt the effect dissipates as the salt becomes concentrated enough to screen the addition of further charges.

Effect of ion size on depletion attraction and monopole repulsion

Electrostatic interaction between proteins is important at low salt concentrations. But at high salt, it becomes weak due to the screening effect of salt. In such high salt concentrations or crowded environments, the dominant effect originates from non-electrostatic interactions. One non-electrostatic interaction is the depletion interaction which is attractive in nature and is one source of salting out. So, as we change salt concentration from low to high or vice versa then we need to make sure to include both electrostatic and non-electrostatic interactions.

At high salt concentration, salt ions occupy a significant amount of volume in the system and the dominant contribution to the free energy comes from the salt-mediated depletion interaction. The depletion effect arises from the final term of Eq. 2.6. The attraction is proportional to the salt concentration and to the change in the ion accessible volume upon aggregation. The accessible volume for different ions is different so the depletion interaction is ion specific. The depletion effect starts to play a role once the ion excluded volumes overlap with each other due to the close proximity of proteins. The amount of overlapped volume for different sized ions for different proteins are shown in table 2.2.

Table 2.2: *Change in ion accessible volume per protein in lysozyme and chymosin crystals.*

Size of ions (Å)	Lysozyme($Å^3$)	Chymosin($Å^3$)
1.25 (NH_4^+)	800	1820
1.50 (Cl^-, K^+)	1200	3200
1.67 (Na^+)	2000	4800

The variation of free energy contributed by the depletion effect with respect to salt concentration is shown in Fig.(2.15). It becomes more important as the salt concentration is increased. To show its ion specificity, we have shown its effect for different salts. This effect also depends on the size of the protein. Here, we have shown this effect for chymosin and lysozyme protein. Fig.(2.16) shows the importance of depletion interaction at high salt where the electrostatic effect is minimal.

The size of an ion is correlated with its population in the cavity. To show the effect of

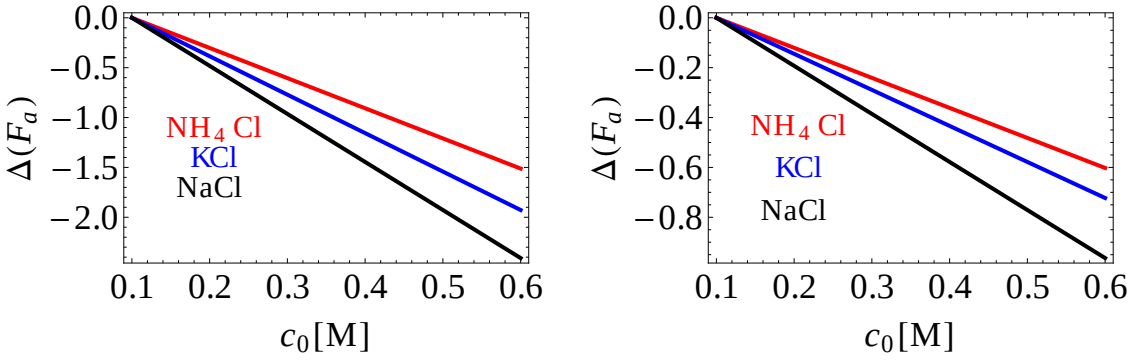

Figure 2.15: *Free energy contributed by different salts at different concentration via depletion attraction. (Left:) Chymosin protein. (Right:) Lysozyme protein.*

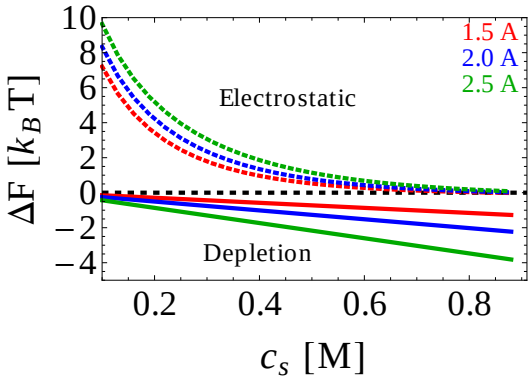

Figure 2.16: *Variation of electrostatic and non–electrostatic components of free energy with respect to salt concentration. The size of anions are changed by keeping cation size fixed at* $1.5\mathring{A}$.

ion size on the electrostatic interaction, we have choosen the lysozyme as our model protein. The radius of lysozyme in the spherical model is $16.1\mathring{A}$ and the radius of cylindrical cavity formed in the aggregate state is $6.0\mathring{A}$ based on the crystal structure data[47]. We have studied lysozyme far from its isoelectric point at pH 4.5 where its net charge is $n_0 = 10$ and dipole charge is $n_1 = 7$.

Figure (2.17) shows the free energy change with respect to salt concentration. The free energy change decreases with the increase in salt concentration leading to salting out. In the left hand side figure, we have changed the size of anions keeping the cation size fixed. Notice that the bigger anions produce a higher solubility than the smaller ones. In this case, protein is positively charged ($n_0 = 10$) so anions are counter-ions. The bigger counter-ions are unfavorable to be accomodated in the crystal cavity. The population of bigger counter-ions in the cavity makes it difficult to satisfy the neutrality condition in the crystal. So, due to the exclusion of bigger counter-ions, the resulting solubility is higher. On the other hand, the smaller counter-ions are easily accomodated in the cavity and make it easier for protein crystal satisfy the neutrality condition resulting the lower solubility.

In the right hand side figure of figure (2.17), the size of cations are changed by keeping the common anion size fixed. The size effect of cation is just opposite than anions. In this case, cations are coions. The use of bigger coions decreases the solubility whereas the smaller coions increase the solubility.

The crystal cavities are supposed to be filled mostly by counter-ions. But, there are some co-ions too. So, if we compare two figures then we can see that the size of anions effect is larger than the cation effect. The size of ions used here are shown in figure. The size of common cation in left figure and the size of common anion in right figure are equal to $1.5\mathring{A}$. In both figures, we haven't considered the depletion interaction so the whole effect is coming from purely electrostatic origin.

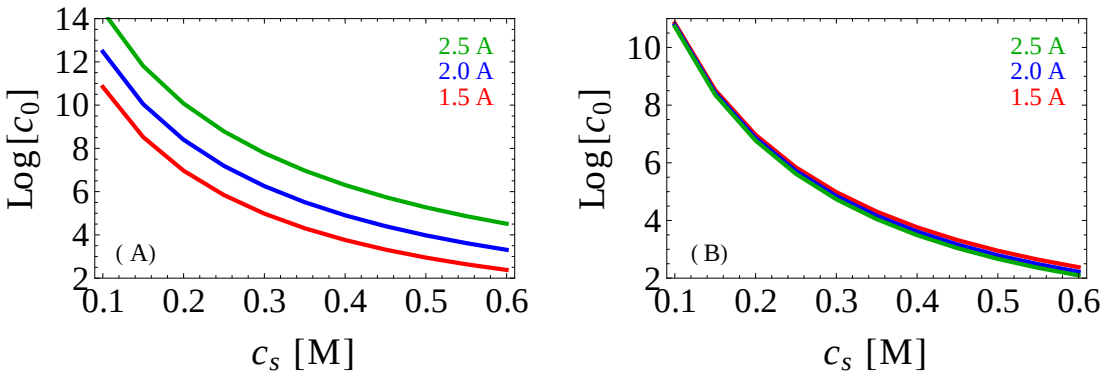

Figure 2.17: *Electrostatic effect on solubility due to various anion and cation size. (Left) Size of anions are changed. (Right) Size of cations are changed. In both figures, the charge of protein is taken to be* 10 *and size of common ion is taken to be* 1.50$\mathring{A}$*, size of protein and cavity are* 16.1$\mathring{A}$*, and* 6.0$\mathring{A}$ *respectively. Note that the anion effect is more pronounced than cation effect.*

Salting out–salting in–salting out trends

In the previous section, we have mentioned the role of anion binding on the salting in enhancement. The enhancement can be observed at or above isoelectric point at salt concentrations where electrostatic interactions are significant. The outcome of anion binding at low salt can be salting out too if the protein is below isoelectric point. In such case, both screening effect and anion binding effect are favorable for the salting out. To observe salting out at low salt, the binding should decrease the net monopole charge before its charge get reversed. If the magnitude of protein charge gets decreased at low salt and then increases at high salt after charge reversal then in such cases, the solubility can be non–monotonic function of salt concentration producing salting out at low salt, salting in at intermediate salt, and salting out at high salt. The first salting out is the result of reduced monopole charge plus the screening effect of salt. The follow up salting in at intermediate salt concentration is the outcome of increasing charge due to charge reversal and the final salting out at high salt is the result of the screening effect (if salt concentration isn't high enough already to neutralize electrostatic interaction) plus salt mediated depletion interaction. The multiple non–monotonic solubility of a model protein with lysozyme like parameters at pH 7.8 is shown in Fig.(2.18). The origin of such non–monotonicity in solubility is due to the

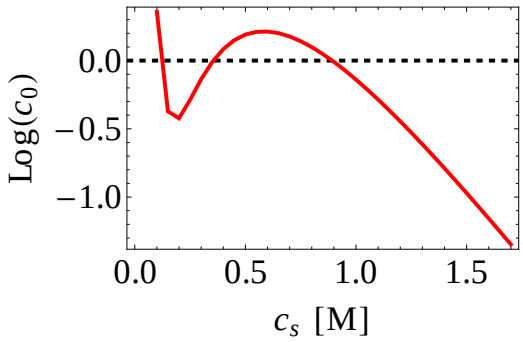

Figure 2.18: *Non–monotonic solubility (Salting out–salting in–salting out) of lysozyme protein at pH* 7.8*. Salting out at low salt is followed up by salting in at intermediate salt which is followed up by salting out at high salt. The anion binding sites on protein surface are taken to be* 25*.*

salt dependence of protein net charge. A schemetic diagram showing the variation of protein charge and its dramatic effect on solubility with respect to salt concentration is presented in Fig.(2.19).

This kind of non–monotonic behavior in lysozyme has been measured in the second virial coefficient[53], in cloud point temperature[54] and in solubility[55]. Allahyarov et al. (2003) has performed simulation capturing the discrete charge of protein predicting the salting out at low salt followed by salting in and salting out. Similarly, Tavares et al (2004) modeled protein charge by monopole and dipole and showed that the monopole screening happens in faster rate than the dipole screening resulting salting out behavior at lower salt regime which is followed by salting in at relatively high salt concentration. Broide et al (1996) has measured salting out–salting in behavior of lysozyme experimentally with transition happening at around $1M$ for various ions but notable effect is seen on ions with valency more than one. But, the begining of salting in at such high salt concentration ($> 1M$) cast doubt that it is caused by the electrostatic effect.

Size of counter-ion competes for depletion and neutralization entropy

For the illustration of ion size effect on solubility at low and high salt concentration, we have plotted Fig.(2.20). In the left hand side figure, size of anions (counter-ions) are varied from

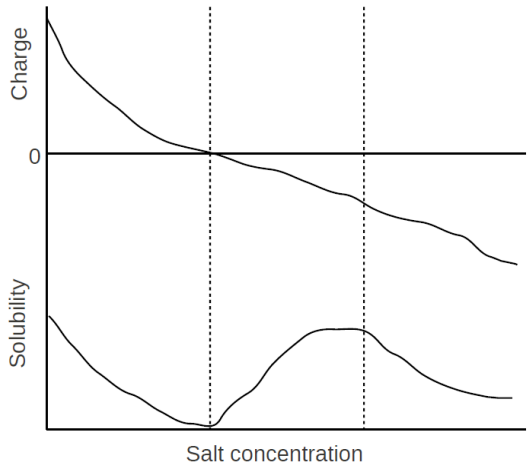

Figure 2.19: *A schematic diagram demonstrating the effect of anion binding on protein net charge and solubility.*

1.50$\mathring{A}$ to 2.50$\mathring{A}$ in the step of 0.5$\mathring{A}$ keeping cation size constant at 1.50$\mathring{A}$. In this plot, we observe the reversal effect on solubility at low salt and high salt. At low salt, big counter-ions result the strong monopole repulsion giving large solubility. But, at high salt, we see the reversal effect on solubility as a outcome of larger depletion attraction caused by the bigger ions.

In the right hand side figure in Fig.(2.20), size of cations (co-ions) are varied from 1.50$\mathring{A}$ to 2.50$\mathring{A}$ in the step of 0.5$\mathring{A}$ keeping anion size constant at 1.50$\mathring{A}$. Here, we don't see solubility reversal with respect to the salt concentration because cations are co-ions and their size effect at low salt is hard to observe due to the low probability of finding them in the cavity. Even if co-ions are present, the probability of finding smaller ions is more than the probability of finding bigger ions. Due to this reason, bigger co-ions are bad to achieve higher solubility at low salt. But at high salt, they play similar role as counter–ions play in non-electrostatic interaction leading to salting out with the nonlinear screening behavior giving way to a linear depletion effect at high salt.

In Fig. (2.21), we have plotted the effect of two anions–floride and iodide on the solubility. In this plot, floride ion makes protein more soluble at low salt and it makes protein less soluble at high salt than the iodide ion does. The size of floride and iodide ions are 3.52$\mathring{A}$ and 2.16$\mathring{A}$[51] respectively. The binding sites is taken to be $N_s = 18$ and $E_b = -0.3k_BT$

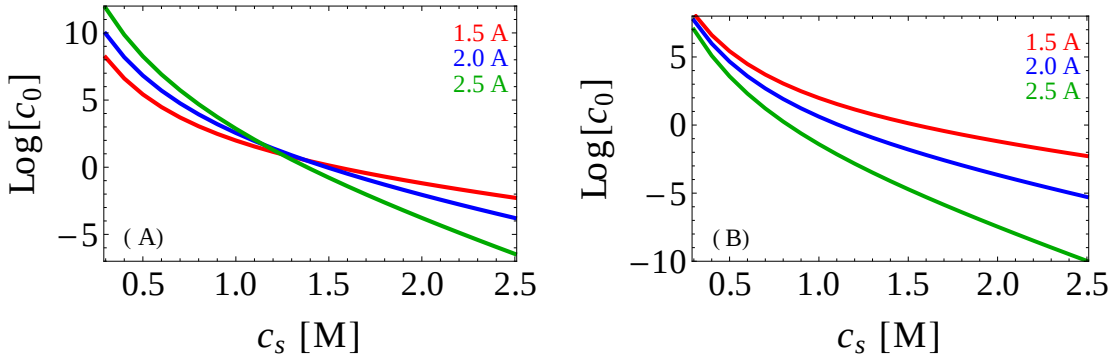

Figure 2.20: *Comparison of the effect of excluded volume on counterions and coions. A) Variation in the counterion size leads to a reversal in the Hofmeister series with large ions more effective at salting-out at high salt and small ions more effective at low salt. B) Coions do not show a reversal since the exclusion of large ions and the depletion effect both favor aggregation. In both panels, the protein charge is* $n_0 = 10$*, the common ion size is* 1.5Å*, and ion binding effects have been removed* ($N_s = E_{\text{bind}} = 0$) *to highlight the excluded volume effect.*

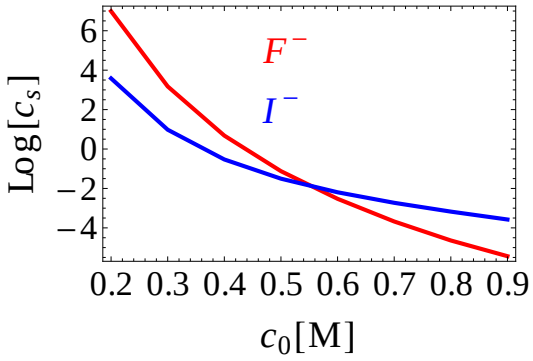

Figure 2.21: *Reversal in the Hofmeister series for a positively charged protein with* $n_0 = 10$*. At low salt, there is a reverse Hofmeister effect dominated by the entropy of neutralizing the aggregate, whereas at the high salt there is direct Hofmeister effect due to the depletion interaction. The ion radii are* $R_{\text{F}} = 3.52$Å*,* $R_{\text{I}} = 2.16$Å*?* *, and* $R_{\text{Na}} = 1.67$Å*.* $R = 16.1$Å*,* $R_c = 6$Å*,* $N_s = 15$*,* $E_{\text{bind}} = -0.3k_BT$*. Charge of protein is taken to be* 10*.*

respectively in all cases to minimize the parameter in the theory though it changes from one ion to another.

From the effectiveness of these two anions to salt out protein, we see that they follow the reverse Hofmeister series at low salt regime and follow the direct Hofmesiter series at high salt concentration. This kind of reverse Hofmeister series at low salt and direct Hofmeister series at high salt is observed in the experimental measurement of cloud point temperature of lysozyme[52]. Previous work has attributed the reversal to the non-electrostatic association of anions with the protein surface[60]. However, Boström et al. did not consider the ionic excluded volume, suggesting that ion-protein association and excluded volume effects both contribute to the salt concentration dependent reversal.

2.4 Conclusion

We have designed a theoretical model that captures salting in, salting out and Hofmeister effects. Many effects of salt such as the screening effect, the binding effect, and the depletion effect have been accounted in the theory. By accounting for the monopole and dipole of the charge distribution, we have shown when to expect salting in/salting out of protein. In monopole dominant case (away from the isoelectric point), the screening of electrostatic interaction gives the salting out result. In this case, we have also found that the majority of free energy is contributed by the entropic loss of counter ions while being confined in the cavities of the aggregate as earlier shown by Schmit and Dill[7;9]. In dipole dominant region (near the isoelectric point), we found that the screening effect of salt results in salting in. In this case, the free energy is mostly contributed by the Coulomb energy term because due to the dipole alignment the salt ions trapped in screening layers escape thereby gaining entropy. These two results are purely electrostatic and represents the screening effect of salt.

But, the straightforward results mentioned above change when other effects of salt are accounted for. If anion binding to the protein surface is considered in addition to the screening effect then the competition of weakening and enhancing of electrostatic interaction determines the results. At the isoelectric point or above it, anion binding always enhances

monopole repulsion whereas below the isoelectric point anion binding can reduce monopole repulsion. The salt dependent charge caused by anion binding strongly effects the electrostatic interaction. The salting in behaviors seen in chymosin protein at low salt regimes are due to the combined effect of dipole screening and monopole enhancement. The salting out behavior seen in that protein at high salt following salting in is due to the non–electrostatic depletion attraction. At high salt, electrostatics are negligible so the non–electrostatic depletion effect plays a dominant role. At high pH, the monopole charge becomes large so we obtain only salting out behavior which is caused by monopole screening (at low salt) and depletion effect (at high salt).

To capture the salt specific effect in the theory, we treat salt ions not being point particles instead they have a finite size and their bare size is strongly effected by the solvation. The finite size of the ions has many effects that must be accounted for. First, the ion size correlates with polarizability, which affects the protein-ion interaction[35]. Secondly, the ion radius determines its affinity to water, resulting in non-trivial corrections to the effective ion volume[34]. This effective size, in turn, determines the entropic cost of trapping ions within the aggregate and the strength of the salt-mediated depletion attraction.

Our theory is capable of explaining the reversal Hofmeister series seen for counter-ions at low salt and high salt. According to our model, at low salt regime, big counter–ions are unfavorable for protein aggregation because they exclude from the aggregate cavities and make it difficult for the aggregate to satisfy the neutrality condition. The exclusion of counter–ions from the cavities leads to the increase in solubility. On the other hand, at high salt, big counter–ions are favorable for protein aggregation because they enhance the depletion attraction. The enhancement of depletion attraction at high salt decreases protein solubility. Our theory is also capable of describing the co–ions specific effect seen in lysozyme protein. We found that big co–ions are favorable for protein aggregation both in low salt and high salt regimes. The exclusion of big ions makes aggregate easy to satisfy neutrality condition at low salt and this effect decreases the protein solubility. At high salt, big co–ions enhance depletion attraction like counter–ions and also decrease solubility. Our theory of co–ion specific effect compares well with the experimental result obtained by Kautt et al

(1989) in lysozyme protein at pH 4.5.

2.5 Bibliography

Bibliography

[1] Jed J. W. Wiltzius, Stuart A. Sievers, Michael R. Sawaya, and David Eisenberg, *Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process*, Protein Science, **18**, 1521–1530, 2009.

[2] Kate L. Moreau and Jonathan A. King, *Protein Misfolding and Aggregation in Cataract Disease and Prospects for Prevention*, Trends in Molecular Medicine, **18 (5)**, 2012.

[3] G Brent Irvine, Omar M El–Agnaf, Ganesh M Shankar, and Dominic M Walsh, *Protein Aggregation in the Brain: The Molecular Basis for Alzheimers and Parkinsons Diseases*, Mol. Med., **14(7-8)**, 451–464, 2008.

[4] C. Tanford, *Physical Chemistry of Macromolecules*, John Wiley and Sons Inc., (1966).

[5] Kim A. Sharp and Barry Honig, *Calculating Total Electrostatic Energies with the Nonlinear Poisson-Bottzmann Equation*, J. Phys. Chem., **94**, 7684, (1990).

[6] Andelman D, *Handbook of Biological Physics*, Elsevier Science, Amsterdam, 1 (1995) Chapter 12 (Electrostatic Properties of Membranes: The Poisson-Boltzmann Theory).

[7] J. D. Schmit, S. Whitelam, and K. A. Dill, *The Journal of Chemical Physics*, **135**, 085103 (2011).

[8] J. Theodoor G. Overbeek, *The role of energy and entropy in the electrical double layer*, Colloids and Surfaces, **51**, 6175, 1990.

[9] Jeremy D. Schmit, and Ken A. Dill, *The Stabilities of Protein Crystals*, J. Phys. Chem. B, **114**, 40204027, 2010.

[10] B. W. Ninham and V. Yaminsky, *Langmuir*, **13**, 2097, (1997).

[11] A. P. dos Santos and Y. Levin, *Phys. Rev. Lett.*, **106**, 1 (2011).

[12] Huan-Xiang Zhou, *Interactions of Macromolecules with Salt Ions: An Electrostatic Theory for the Hofmeister Effect*, PROTEINS: Structure, Function, and Bioinformatics, **61**, 69–78 (2005)

[13] S. C. Flores, J. Kherb, and P. S. Cremer, *J. Phys. Chem. C*, **116**, 14408 (2012).

[14] L. Medda, B. Barse, F. Cugia, M. Bostrom, D. F. Parsons, B. W. Ninham, M. Monduzzi, and A. Salis, *Langmuir*, **28**, 16355 (2012).

[15] M. Lund and P. Jungwirth, *J. Phys: Condens. Matter*, **20**, 494218 (2008).

[16] X. Tadeo, B. Lopez-Mendez, D. Casta no, T. Trigueros, and O. Millet, *Biophys. J.*, **97**, 2595 (2009).

[17] J. W. Bye and R. J. Falconer, *Protein Science*, **22**, 1563 (2013).

[18] E. Sedlak, L. Stagg, and P. Wittung-Stafshede, *Arch. Biochem. Biophys*, **479**, 69 (2008).

[19] M. Bostrom, F.W. Tavares , S. Finet , F. Skouri-Panet , A. Tardieu , B.W. Ninham, *Why forces between proteins follow different Hofmeister series for pH above and below pI*, Biophysical Chemistry, **117**, 217–224, (2005).

[20] E. J. Cohn and J. T. Edsall, *Proteins, Amino Acids, and Peptides*, Hafner Publishing Company, New York, 1943.

[21] Arakawa, T., and S. N. Timasheff, *Theory of protein solubility*, Methods Enzymol., **114**, 49–77, 1985.

[22] R. C. DeMattei and R. S. Feigelson, *Journal of Crystal Growth*, **110**, 34 (1991).

[23] F. Hofmeister, *Arch. Exp. Pathol. Pharmakol.*, **24**, 247 (1888).

[24] E. J. Cohn and J. T. Edsall, *Proteins, amino acids and peptides as ions and dipolar ions (American Chemical Society. Monograph series)*, Reinhold Pub. Corp, 1943.

[25] E.J. Cohn, *The physical chemistry of the proteins*, Physiol. Rev., **5**, 349–437, (1925).

[26] A. A. Green, *J. Biol. Chem.*, **93**, 517 (1931).

[27] A.A. Green, *Studies in the physical chemistry of the proteins. X. The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentration*, J. Biol.Chem., **95**, 47–66, (1932).

[28] J. Mellanby, *Globulin*, J. Physiol., **33**, 338–373, (1905).

[29] A.C. Dumetz, A.M. Chockla, E.W. Kaler, A.M. Lenhoff, *Effects of pH on proteinprotein interactions and implications for protein phase behavior*, Biochim. Biophys. Acta, **1784**, 600–610, (2008).

[30] Yatin R. Gokarn, R. Matthew Fesinmeyer, Atul Saluja, Vladimir Razinkov, Susan F. Chase, Thomas M. Laue, and David N. Brems, *Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions*, Protein Science, **20**, 580–587, 2011.

[31] Manoj K. Menon and Andrew L. Zydney, *Measurement of Protein Charge and Ion Binding Using Capillary Electrophoresis*, Anal. Chem., **70**, 1581–1584, 1998.

[32] Luca Medda, Brajesh Barse, Francesca Cugia, Mathias Bostrom, Drew F. Parsons, Barry W. Ninham, Maura Monduzzi, and Andrea Salis, *Hofmeister Challenges: Ion Binding and Charge of the BSA Protein as Explicit Examples*, Langmuir, **28**, 16355–16363, 2012.

[33] Travis T. Waldron, Modestos A. Modestou, and Kenneth P. Murphy, *Anion binding to a proteinprotein complex lacks dependence on net charge*, Protein Science, **12**, 871–874, (2003).

[34] K. D. Collins, G. W. Neilson, and J. E. Enderby, *Biophysical chemistry*, **128**, 95 (2007).

[35] Drew F. Parsons and Barry W. Ninham, *Importance of Accurate Dynamic Polarizabilities for the Ionic Dispersion Interactions of Alkali Halides*, Langmuir, **26(3)**, 1816–1823, 2010.

[36] Y. Zhang and P. S. Cremer, *Curr. Opin. Chem. Biol.*, **10**, 658 (2006).

[37] V. Vlachy, H. W. Blanch, and J. M. Prausnitz, *AIChE Journal*, **39**, 215 (1993).

[38] B. Hribar-Lee, V. Vlachy, and K. A. Dill, *Acta chimica Slovenica*, **56**, 196 (2009).

[39] Asakura, S., and F. Oosawa, *Interactions between particles suspended in solutions of macromolecules*, J. Polym. Sci. [B], **33**, 183–192, 1958.

[40] Kim A. Sharp, *Analysis of the size dependence of macromolecular crowding shows that smaller is better*, PNAS, **112**, 7990–7995, 2015.

[41] Ronald W. Maurer, Stanley I. Sandler, Abraham M. Lenhoff, *Salting-in characteristics of globular proteins*, Biophysical Chemistry, **156**, 72–78, (2011).

[42] D. F. Parsons, M. Bostrom,T. J. Maceina, A. Salis, and B. W. Ninham, *Why Direct or Reversed Hofmeister Series? Interplay of Hydration, Non-electrostatic Potentials, and Ion Size*, Langmuir, **26(5)**, 3323–3328, 2010.

[43] Jacob Kielland, *Individual Activity Coefficients of Ions in Aqueous Solutions*, J. Am. Chem. Soc., **59**, 1675–1678, 1937.

[44] Gerald R. Grimsley, J. Martin Scholtz, and C. Nick Pace, *A summary of the measured pK values of the ionizable groups in folded proteins*, protein science, **18**, 247–251, 2009.

[45] Madeleine M. Ries-Kautt and Arnaud F. Ducruix, *Relative Effectiveness of Various Ions on the Solubility and Crystal Growth of Lysozyme*, The Journal of Biological Chemistry, **264(2)**, 745–748, 1989.

[46] Matjaz Boncina, Jurij Rescic and Vojko Vlachy, *Solubility of Lysozyme in Polyethylene Glycol-Electrolyte Mixtures: The Depletion Interaction and Ion-Specific Effects*, Biophysical Journal, **95**, 1285–1294, 2008.

[47] M. C. Vaney *et al.*, *Acta Crystallographica. Section D, Biological Crystallography.*, **52**, 506 (1996).

[48] M. Newman *et al.*, *Journal of Molecular Biology*, **221**, 1295 (1991).

[49] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, *J. Comput. Chem.*, **25**, 1605, 2004.

[50] E. F. Pettersen *et al.*, *Journal of computational chemistry*, **25**, 1605 (2004).

[51] R. Nightingale, *Journal of Physical Chemistry*, **63**, 1381 (1959).

[52] Yanjie Zhang and Paul S. Cremer, *The inverse and direct Hofmeister series for lysozyme*, PNAS , **106(36)**, 15251 (2009).

[53] E. Allahyarov, H. Lowen, J. P. Hansen, and A. A. Louis, *Nonmonotonic variation with salt concentration of the second virial coefficient in protein solutions*, Physical Review E, **67**, 2003.

[54] F. W. Tavares, D. Bratko, A. Striolo, H. W. Blanch and J. M. Prausnitz, *Phase behavior of aqueous solution containing dipolar proteins from second–order perturbation theory*, J. Chem. Phys., **120 (20)**, 9859–5869, 2004.

[55] Michael L. Broide, Tina M. Tominc, and Marc D. Saxowsky, *Using phase transitions to investigate the effect of salts on protein interactions*, Phys. Rev. E, **53(6)**, 1996.

[56] J. Poznański, M. Wszelaka-Rylik, and W. Zielenkiewicz, *Thermochimica Acta*, **409**, 25 (2004).

[57] J. Poznański, M. Wszelaka-Rylik, and W. Zielenkiewicz, *Biophysical Chemistry*, **113**, 137 (2005).

[58] M. Boncina, J. Lah, J. Rescic, and V. Vlachy, *The Journal of Physical Chemistry B*, **114**, 4313 (2010).

[59] L. Sibille and M. L. Pusey, *Acta crystallographica. Section D, Biological crystallography*, **50**, 396 (1994).

[60] Mathias Bostrom, Drew F. Parsons, Andrea Salis, Barry W. Ninham, and Maura Monduzzi, *Possible Origin of the Inverse and Direct Hofmeister Series for Lysozyme at Low and High Salt Concentrations*, Langmuir, **27**, 9504–9511, 2011.

Chapter 3

Kinetics of crystal growth

We have published some contents of this chapter in **J. Chem. Phys., 144, 064903 (2016).**

3.1 Introduction

Crystallization in a solution is a two step process. In the first step, crystallizing particles make a nucleus through the process of nucleation. Then in the next step, other crystallizing particles attach to the nucleus to advance the crystal growth. For the advancement of growth, it is required to have molecules with greater chemical potential in solution than in the crystal. This condition is fulfilled when the solution is supersaturated. But, to figure out the supersaturation suitable for the growth of high quality crystals in a reasonable time is a challenging task. Too low supersaturation (metastable) isn't a favorable condition because it is insufficient to form crystal in reasonable time whereas too high supersaturation leads to gelation and unwanted aggregation so it is also not preferable. People prefer to perform crystallization experiments at intermediate supersaturation (labile)[1]. But, even in that regime, often crystallization of many small molecules such as polymers and proteins end up with no crystal due to kinetic traps.

Self–poisoning is one of the kinetic traps that can prevent the crystallization of molecules from solution[2;3]. To hamper crystal growth, molecules possessing internal degrees of freedom

attach to the crystal in such a way that they don't follow the existing pattern in terms of orientations and conformations. If crystal growth has encountered self–poisoning then the growth rate doesn't increase monotonically with the thermodynamic driving force. Instead, it shows non–monotonic behavior. At first, growth rate increases with the driving force, then it drops which is the characteristic of self–poisoning. Finally it increases sharply after it has passed the disordered precipitation limit.

Schmit and Dill (2012) have shown the adverse effect of non–productive binding of molecules on the growth rate of protein crystal in their analytical model[4]. The non–productive binding of molecule on the crystal growth site doesn't allow a productive molecule to advance crystallization. People have also seen self–poisoning in hard rod liquids[5], in the assembly of polymers[2;6;7], and proteins[8;9]. In hard rods, orientational degree of freedom determines crystallization and in polymer and protein both orientational and conformational degrees of freedom plays role in crystallization.

Here, we use Monte Carlo simulation to study self–poisoning in crystal growth. Anisotropic and flexible molecules such as proteins are prime candidates for growth poisoning. Crystal growth of anisotropic and flexible molecules look for the correct binding state. Such molecules may fall into kinetic traps while searching for the crystallographic state. Our simulation study shows that three minimal requirements are sufficient for the emergence of self–poisoning in crystal growth. One required ingredient is that a molecule can bind in two ways to a crystal, namely, crystallographic and non–crystallographic. Both kinds of binding are energetically attractive but crystallographic way of binding is more energetically favorable than the non–crystallographic way of binding. The next requirement is that the energetically favorable binding events are less probable to happen than non–crystallographic binding events. If these minimal requirements are fulfilled then the crystal growth rate can be non–monotonic with the thermodynamic force. The growth rate is also closely related to the quality of the crystal. As the growth rate increases the quality of crystal declines. This study also shows the guidelines for avoiding or recovering from poisoning. Our simulation results are consistant with the results obtained from mean field theory of crystal growth[13].

3.2 Monte Carlo simulations

To carry out our simulation, we represent crystallizing molecule in the solution by either blue or red color. Similar type of molecular representations are used in previous works[10–12]. Blue colored particles represent the crystallographic conformation and orientation of a molecule whereas the red color represents the non–crystallographic conformation and/or orientation of the same molecule. We assume that blue and red particles are in $p/(1-p)$ proportion in the solution and p is less than $1/2$, which means that a molecule possessing a non-crystallographic state is more probable than the crystallographic one. The interaction energy between these particles is attractive but the strength of the interaction depends on the type of particles involving in the interaction. For example, if the interaction is between crystallographic particles (blue–blue) then it is stronger. Otherwise (blue–red or red–red), the interaction energy is weak. We have shown the representation and interaction energy between particles in Fig.(3.1).

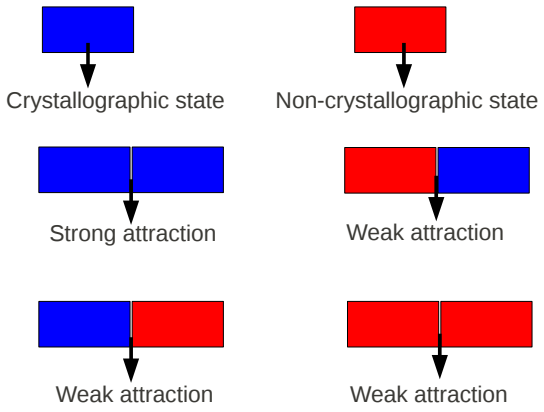

Figure 3.1: *Cartoon representation of crystallographic and non–crystallographic state. Interaction energies between crystallographic states are strong and all remaining interactions are weak.*

We use lattice Monte Carlo simulations to see the growth of structure from two component system. Here, the simulation consists of a 3D cubic lattice with system dimensions $15 \times 15 \times 100$. We apply periodic boundary conditions along the two short directions. At each time step we choose a site randomly. If the choosen site is empty then we propose with probability

(p) to fill it by blue particle and with probability $(1-p)$ to fill by red particle. If the randomly selected site is occupied with a blue or red particle, we attempt to remove that particle and make site empty. We don't allow the interchange between blue and red particles. Also, we don't allow the change of state of any lattice site that has been surrounded by 6 nearest neighbours particles. We accept the move according to the following Metropolis probabilities. Here, W represents the empty site, B and R represent the blue and red particles, respectively.

$$\begin{aligned} W \to B &: \min(1, p^{-1}e^{-\beta\Delta E}) \\ B \to W &: \min(1, pe^{-\beta\Delta E}) \\ W \to R &: \min(1, (1-p)^{-1}e^{-\beta\Delta E}) \\ R \to W &: \min(1, (1-p)e^{-\beta\Delta E}) \end{aligned}$$

Where, β is the reciprocal of thermal energy and ΔE is the energy change due to the insertion or removal of particle in a site. To evaluate the energy change, it is required to identify the status of nearest neighbour site whether it is occupied or not and if it is occupied then it is further required to know whether the particle is blue or red. We use the lattice energy function to calculate the energy change as follows:

$$E = \sum_{<i,j>} \epsilon_{C(i)C(j)} + \sum_{i} \mu_{C(i)} \tag{3.1}$$

Where, $\epsilon_{C(i)C(j)}$ is the interaction energy between colors $C(i)$ and $C(j)$ and the first sum in equation (3.1) runs over all distinct nearest neighbours to find out the total interaction energy. The second sum in equation (3.1) runs over all sites. The chemical potential $\mu_{C(i)}$ is $\mu k_B T$, $-k_B T \ln p$ and $-k_B T \ln(1-p)$ for W, B and R respectively. Note that, the blue–blue interaction is strong and blue–red or red–red interaction is weak and these pairwise interaction energies are represented by following expression.

$$\epsilon_{BB} = -\epsilon_b k_B T \tag{3.2}$$

$$\epsilon_{BR} = \epsilon_{RB} = \epsilon_{RR} = -\epsilon_d k_B T \tag{3.3}$$

If the pairwise interaction energies are turned off then the probability of a given site to be empty, blue and red is

$$p_W = \frac{1}{(1+e^{\mu})} \tag{3.4}$$

$$p_B = \frac{p}{(1+e^{-\mu})} \tag{3.5}$$

$$p_R = \frac{(1-p)}{(1+e^{-\mu})} \tag{3.6}$$

respectively.

To begin simulation, we put three complete layers of blue particles at one end of the simulation box. This eliminates the need for spontaneous nucleation. Then, we choose the interaction energies and p parameter and let simulation happen for fixed time. A layer in the box is considered to be grown only if at least 50% of sites in that layer are occupied by blue or red particles. We measure growth rate and composition of the structure at different values of the parameter $c = e^{\mu}$.

3.3 Results

3.3.1 Growth rate with respect to driving force

Fig.(3.2) shows simulation snapshots taken at the end of 5×10^9 MC sweeps for different values of the driving force (concentration). The energy parameters for crystallographic binding and non–crystallographic binding are $E_b = -3.5k_BT$ and $E_d = -1.4k_BT$ respectively, which suggests that crystallographic binding is energetically favorable. We have selected the p parameter to be 10^{-2} which means the probability of finding the correct binding state for

crystallization is 1%. The driving force has been increased from left to right. Low value of concentration means molecules approach the crystal for binding at a slow rate. This also means they get enough time to unbind from the crystal to remove defects. On the other hand, high concentration means molecules come at faster rate for binding so that the bound molecules don't have enough time for unbinding. From the unequal height of structure shown in figure, we can say that the growth rate of crystal is non–monotonic function of concentration. Also, it can be observed that the color of the structure is turning from blue to red as we go from left to right. The more blue color in the structure, the higher the purity of crystal and as the population of red color increases the purity of the crystal declines. The growth rate first increases with concentration but this trend doesn't last long because the growth rate start to decrease with the further increase in concentration. This is the signature of self–poisoning. Self–poisoning happens because the crystal becomes less pure which means the effective driving force for the growth decreases due to the fact that interaction energy is weak. This trend is also not a permanent because the growth rate increases sharply with the further increase in concentration which is the indication that the driving force has surpassed the disordered precipitation limit.

In left hand side figure in Fig.(3.3), the growth rate with respect to concentration for different values of weak interaction energies is shown. The strong interaction energy and p parameters are kept constant at $3.5k_BT$ and 10^{-2} respectively. The weak interaction energies are shown in the plot legends. The growth rate is measured by the number of layers added to the aggregate at the end of 5×10^9 MC sweeps. We consider a layer to have been added if greater than 50% of the sites are occupied by red or blue particles. Here, we have normalized the layers grown by the total number of layers in the simulation box. At concentrations just above the blue solubility limit the aggregate grows linearly with increasing concentration. At higher concentrations the growth rate reaches a maximum and then drops sharply. After crossing the red solubility limit, the growth rate increases fast. The location of red solubility limit and self–poisoning concentration strongly relies on the value of weak interaction energy such that both of them shift towards lower value of concentration if the non–crystallographic binding energy is increased. This makes sense

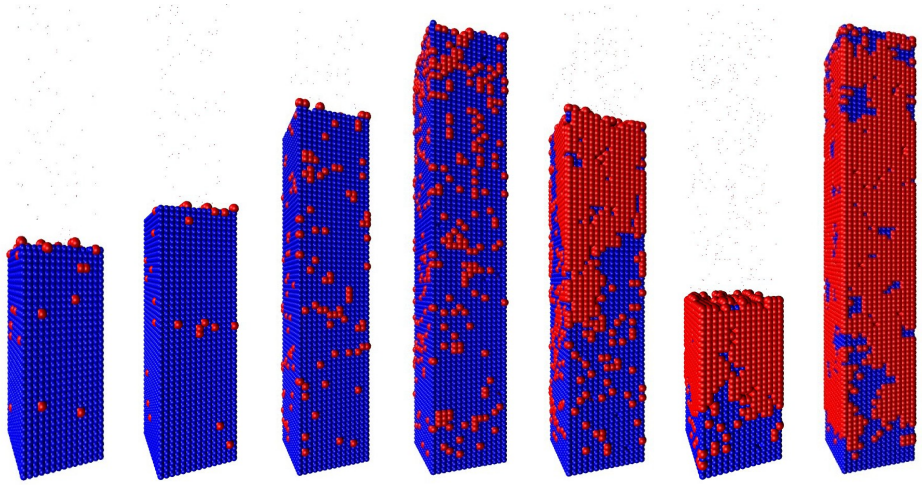

Figure 3.2: *Snapshots at the end of equal simulation time for a range of concentration (driving force). The concentration is increasing from left to right. The height of structure first increases and then decreases with concentration and finally it increases beyond the precipitation line. Meanwhile, the quality of structure changes continuously with more blue color at low driving force and the color turns in to more red as the driving force increases. Growth rate shows the non–monotonic behavior with concentration. Parameters are* $E_b = 3.5k_BT$, $E_d = 1.4k_BT$ *and* $p = 10^{-2}$. *From left to right, values of* c *are* 0.008, 0.0083, 0.009875, 0.0119, 0.014225, 0.0149, 0.01512

because the increased value of non–crystallographic energy reaches close to the value of crystallographic energy so that it is expected to reduce the crystal quality and the gap between the blue and red solubility line. To investigate the effect of the strong interaction energy to the onset of blue and red growth, we have plotted right hand side figure in Fig.(3.3) keeping $E_d = 1.4$ and $p = 0.01$ parameter fixed. In this plot, we don't see the shift of concentrations associated with the self–poisoning and red solubility limit but there is slight increase in the onset of blue growth when the crystallographic interaction energy is decreased. From these observation, we can expect that the energy difference between crystallographic and non–crystallographic interactions determine the range of concentration in which crystal growth can happen. Generally, the more the energy difference between these two ways of binding, the longer the range of concentration suitable for high quality crystal growth.

The simulation doesn't predict an arrested concentration regime where the growth rate is zero as predicted a the mean field theory developed by Whitelam and Schmit[13]. The mean field result representing growth rate and the quality of crystal is shown in Fig.(3.4). However, the simulation confirms the qualitative prediction that the growth rate is a non-monotonic function of the concentration. The growth rate in the simulation doesn't vanish because it satisfies detailed balance and must evolve to equilibrium at long times. We have shown the evolution of the structure grown at red solubility region towards the equilibrium with time in Fig.(3.5).

We have also measured the crystal quality by an order parameter (m) which is shown in Fig.(3.6). It is measured by taking the difference between total blue particles and red particles in the structure grown at the end of simulation time. We have normalized it by the total number of particles in the structure. So, the positive value indicates that the structure is mostly built by blue particles and the negative value means the structure is dominated by red particles. As we can see in the figure, the crystal quality is high (m value closer to 1) at low concentration but it declines as the concentration is increased. The decline in crystal quality works as a driving force for poisoning because unbinding of molecules is favorable due to the weak interaction energy.

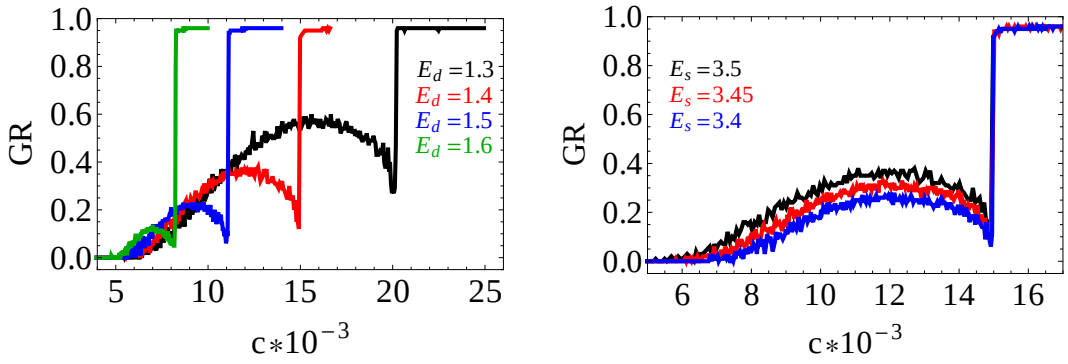

Figure 3.3: *(Left): Normalized growth rate vs concentration at $E_b = 3.5k_BT$, $p = 0.01$ and different E_d. Growth rate shows non–monotonic behavior with concentration at different values of weak interaction energies. The onset of blue growth barely depends on the weak interaction energy but the concentrations associated with both self–poisoning and onset of red growth show strong dependence on it by shifting their locations towards low concentrations with the increase in the interaction energy (E_d). (Right): Growth rate vs concentration at $E_d = 1.4k_BT$, $p = 0.01$ and different strong interaction energy. Growth rate shows non–monotonic behavior with concentration but the position change of self–poisoning and red precipitation line are hard to notice but the onset of blue growth increases with the decrease in interaction energy.*

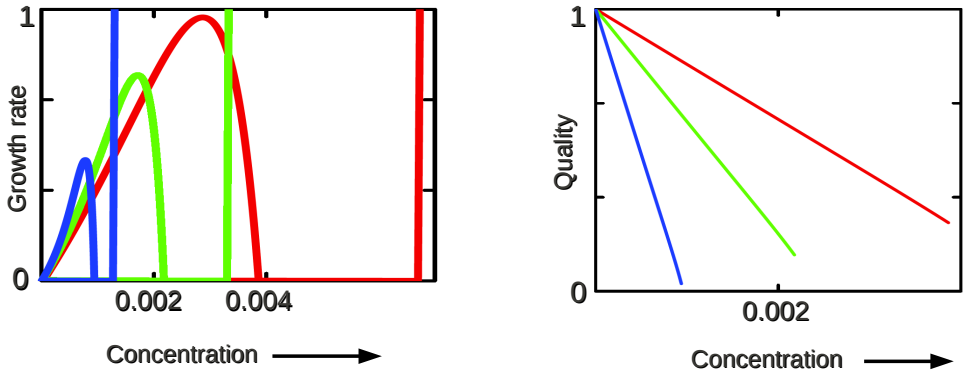

Figure 3.4: *Prediction of a dynamic mean field theory[13] for self–poisoning. It shows that the growth rate vanishes as it is suffered by the poisoning (Left). It also shows that the quality of the crystal deteriorates with the driving force. Different colors represent the binding energy of a molecule in non–crystallographic manner.*

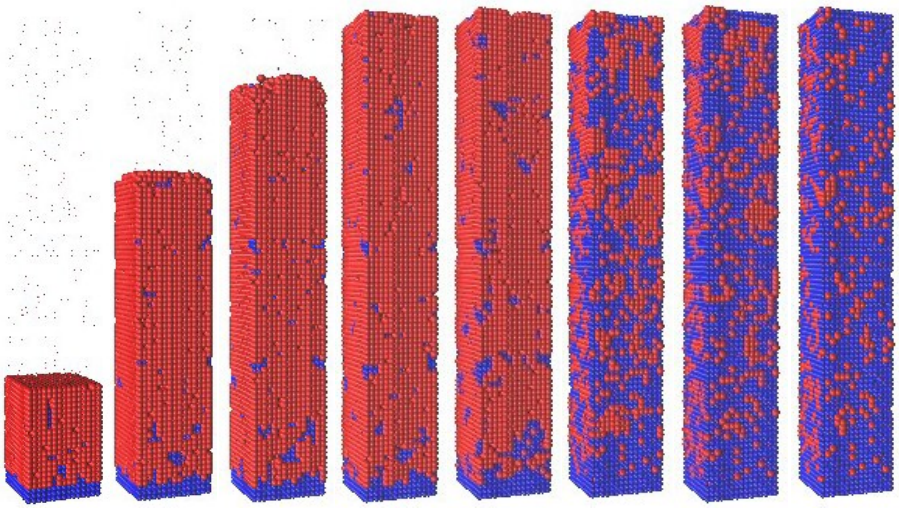

Figure 3.5: *Structure eventually evolves to equilibrium because simulation satisfies the detailed balance. Here, we show snapshots at increasing time from left to right and this simulation is done in the precipitation regime ($c = 0.0274$). $E_b = 3.5k_BT$, $p = 0.01$ and $E_d = 1.2$.*

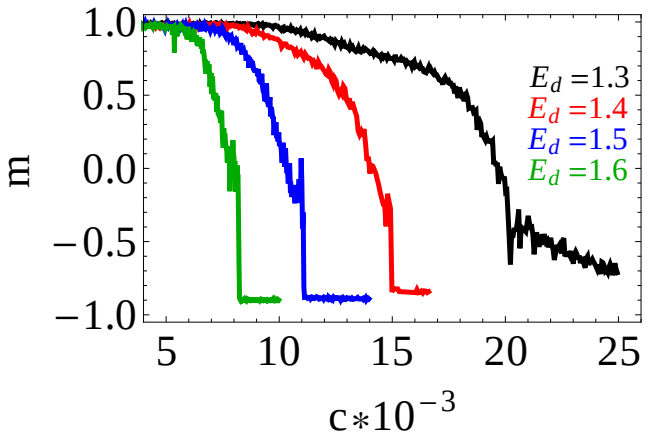

Figure 3.6: *Quality of structure vs concentration at* $E_b = 3.5k_BT$, $p = 0.01$ *and different* E_d. *The value of* m *closer to* 1 *suggest that the quality of crystal is better (blue phase) while the value closer to* -1 *means crystal is enriched by defects (red phase). The value around* 0 *indicates the mixed state which is also not a good crystal.*

3.3.2 Growth of blue and red particles with respect to driving force

In the growth rate plot shown in Fig.(3.3), the contribution of both blue and red particles are included so it is impossible to identify how these blue and red particles are growing with the concentration. To see the growth of blue and red particles, we have plotted Fig.(3.7). In Fig.(3.7), we see that the blue particles grow roughly linearly with the concentration unless their growth is poisoned. The reason behind the poisoning of blue growth is the growth of red particles. Once the blue growth is hit by poisoning, red particles grow so fast that blue growth can't recover. This figure also clearly shows how the red solubility line shifts towards lower concentration with the increase in their interaction energy. It shifts towards low concentration because the bound molecules are less probable to unbind due to increased energy and other molecules are coming in faster rate for binding. In some of the simulations (lower values of weak interaction energy) there is a significant increase in the blue particles at concentrations just above the red solubility limit. This is a result of annealing occurring within the aggregate in the neighborhood of voids and may not be physical for certain systems (i.e. protein crystals). These void are less common at higher concentrations and we observe

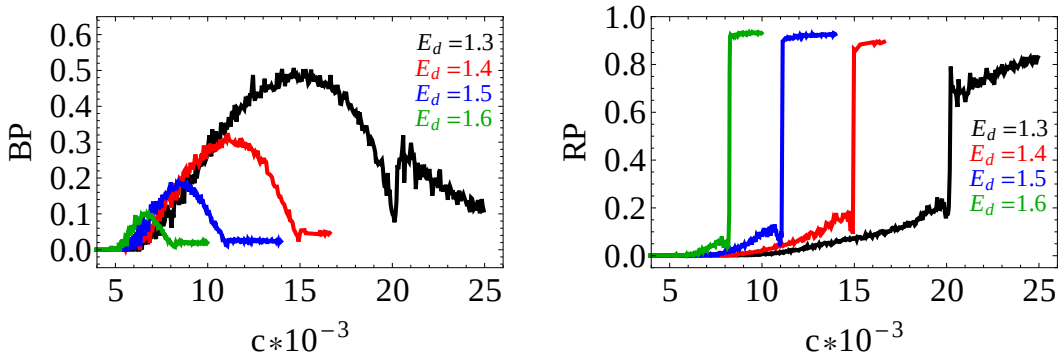

Figure 3.7: *Growth of blue (Left) and red (Right) particles vs concentration at* $E_b = 3.5k_BT$, $p = 0.01$ *and different* E_d. *Blue particles grows linearly at low concentration while red particles grows negligibly. Once red particles start to grow then the growth of blue particles declines resulting the crystal growth self–poisoning. After poisoning, only red particles grow which is just the precipitation not a useful crystal.*

a correspondingly lower amount of annealing.

In Fig.(3.8), we have shown the number of snapshots taken at different time during simulation. Snapshots are taken at concentration $c = 0.0149$ in which growth was poisoned for energy parameters $E_b = 3.5k_BT$ and $E_d = 1.4$. We see that the growth front of all structures are occupied by the red molecules which are unstable and fluctuates with time. The fluctuation of molecules with time is also shown in Fig.(3.9). The blue growth is free from fluctuation because of the strong binding energy. Blue particles are slowly growing from lower end of structure which shows the sign of structure evolving towards equilibrium.

3.4 Conclusion

With the help of computer simulation, we have presented the minimal requirements for the crystal growth self–poisoning and none of the requirements needed molecular details. The minimal requirements include the binding energy of molecule and its probability such that a molecule (monomer of a crystal) can bind either in crystallographic or non–crystallographic way and these ways of binding can happen with sufficiently unequal probability. The crystallographic way of binding should be energetically stronger than the non–crystallographic

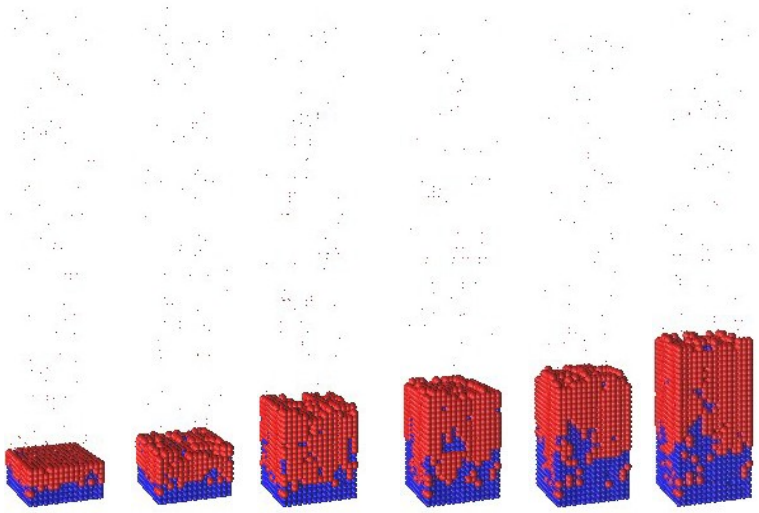

Figure 3.8: *Snapshots to represent the fluctuation of red particles at poisoned concentration.* $E_b = 3.5k_BT$, $p = 0.01$, $c = 0.0149$ *and* $E_d = 1.4$. *The red particles are surrounded by empty sites so they are unstable. The population of blue particles increases with time and there are fewer empty sites in their neighborhood. Blue particles are slowly growing from lower end of structure which shows the sign of structure evolving towards equilibrium.*

way of binding but the probability of crystallographic binding is much smaller than the probability of non–crystallographic binding. If these three requirements are fulfilled then crystal growth will suffer from self–poisoning and the growth becomes a non–monotonic function of the thermodynamic driving force. Self–poisoning is seen in many molecular systems[2;5;8;9] because these systems satisfy the minimal requirements presented here. For instance, protein crystallization is a challenging task for crystallographers and would benefit from rational guidance[4;15–19]. Our simulation parameters suggest that proteins are strong candidates to suffer from self–poisoning during their crystallization because they have much smaller p parameter than other small rigid molecules. The low value of p parameter means that the rate of attachment of non–crystallographic molecule increases and it also increases the wait time for the growth of thermodynamically stable crystals. Proteins are anisotropic and conformationally flexible biomolecules so their non–crystallographic modes of binding dominate their crystallographic mode of binding by a factor of order of 10^4 or 10^5 [4;8;14]. In

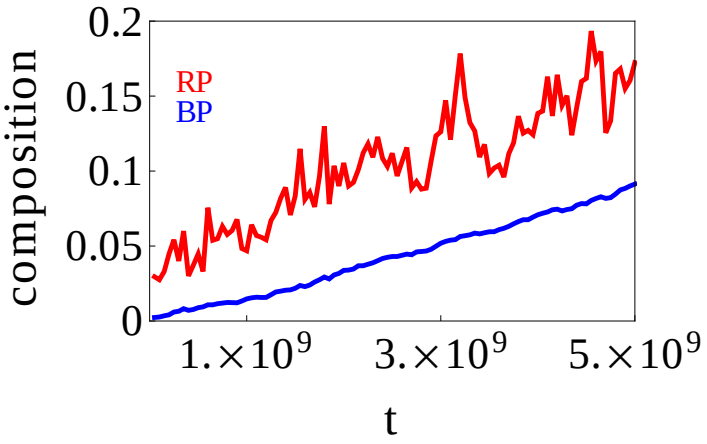

Figure 3.9: *Blue and red particle fluctuation vs time at poisoned concentration ($c = 0.0149$). $E_b = 3.5k_BT$, $p = 0.01$ and $E_d = 1.4$. Blue particles varies with time monotonically and red particles show the fluctuation with time.*

protein crystallization experiments, people frequently don't achieve a new phase such that a solution remains homogeneous at the end of experiment[20]. This failure may be due to the self–poisoning phenomenon.

The present model can also suggest the ways of avoiding self–poisoning. The small change in parameters can be used as a technique to recover crystal growth. For instance, Fig.(3.3) shows the growth rates suffering from poisoning for varying values of interaction energies. To avoid poisoning, one could change the driving force of crystallization or change the solution conditions, for instance, by adding salt to change the interaction energy of molecules so that the existing location of poisoning can be relocated or permanantly disappeared.

3.5 Bibliography

Bibliography

[1] A. McPherson, *Crystallization of Biological Macromolecules, CSHL Press, Cold Spring Harbor*, 1999.

[2] Ungar G., Putra E. G. R., de Silva D. S. M., Shcherbina M. A., and Waddon A. J. *Interphases and Mesophases in Polymer Crystallization I*, 45–87, (2005).

[3] J. J. De Yoreo and P. G. Vekolov, *Rev. Mineral. Geochem.*, **54**, 57, (2003).

[4] Schmit J. D. and Dill K. A., *Journal of the American Chemical Society*, **134(9)**, 3934, (2012).

[5] Schilling T. and Frenkel D., *Physical Review Letters*, **92 (8)**, 085505 (2004).

[6] P. G. Higgs and G. Ungar, *J. Chem. Phys.*, **100**, 640, (1994).

[7] G. Ungar, P. Mandal, P. Higgs, D. De Silva, E. Boda, and C. Chen, *Phys. Rev. Lett.*, **85**, 4397, (2000).

[8] Asthagiri, D., Lenhoff, A., and Gallagher, D. *Journal of Crystal Growth*, **212(3-4)**, 543–554, (2000).

[9] Schmit, J. D., *The Journal of Chemical Physics*, **138(18)**, 185102 (2013).

[10] S. Whitelam, L. O. Hedges, and J. D. Schmit, *Phys. Rev. Lett.*, **112**, 155504, (2014).

[11] A. C. –H Sue, R. V. Mannige, H. Deng, D. Cao, C. Wang, F. Gandara, F. Stoddart, S. Whitelam, and O. M. Yaghi, *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 5591, (2015).

[12] L. O. Hedges, R. V. Mannige and S. Whitelam, *Soft Matter*, **10**, 6404, (2014).

[13] S. Whitelam, Y. R. Dahal and J. Schmit, *J. Chem. Phys.*, **144**, 064903, (2016).

[14] Kierzek, A. M., Wolf, W. M., and Zielenkiewicz, P. *Biophysical journal*, **73(2)**, 571–580, (1997).

[15] George, A. and Wilson, W. W., *Acta crystallographica. Section D, Biological crystallography*, **50(4)**, 361–365, (1994).

[16] P. R. ten Wolde and D. Frenkel, *Science*, **277**, 1975, (1997).

[17] J.-u. Shim, G. Cristobal, D. R. Link, T. Thorsen, abd S. Fradel, *Cryst. Growth Des.*, **7**, 2192, (2007).

[18] T. K. Haxton and S. Whitelam, *Soft Matter*, **8**, 3558, (2012).

[19] D. Fusco and P. Carbonneau, *Colloids Surf., B*, **137**, 22, (2016).

[20] Luft, J. R., Wolfley, J. R., and Snell, E. H., *Crystal growth & design*, **11(3)**, 651–663, (2011).

Chapter 4

Conclusion and future work

4.1 Conclusion

Protein–protein interactions are a subject of interest due for multiple reasons. The varying assembly products such as protein crystals and protein aggregates are the outcomes of protein–protein interaction. Crystals are required in x–ray crystallography for structural study. Protein aggregates often cause diseases such as Alzheimer's, Parkinson's, eye lens cataracts etc. The outcomes of protein assembly change significantly when the interaction strength are modified. For instances, proteins dissolve or precipitate in solution depending on the interaction strength. So, to control outcomes it is necessary to know the role of different parameters on protein–protein interaction.

Using analytical theory and computer simulation, we have studied the kinetic and equilibrium factors in protein crystal growth. Since proteins are anisotropic and flexible molecules, their crystal growth is a delicate process. Supersaturated solution is required for crystal growth but the supersaturation of solution doesn't assure protein crystallization because there must also be accessible kinetic pathways to the correct binding state. Additionally, proteins are often part of a heterogeneous systems, (for instance, the cell), so the interaction between proteins not only depend on themselves but also it depends on neighbouring molecules and the solution conditions. There could be so many interactions happening in

the system concurrently such as electrostatic, hydrophobic, hydrogen bonding, salt bridging, van der Waals and they all make some fraction of contribution to the net interaction. In our equilibrium study, we emphasized the role of pH, salt concentration and its type in the protein–protein interaction by developing an analytical theory. For the kinetics, we used computer simulation to identify the associated difficulties during the course of crystal growth.

Salting in/salting out and the Hofmeister effect are familiar words in this field for over a century. They describe the effect of salt on protein solubility. They are measured experimentally but understood very little theoretically. Protein solubility is a thermodynamic quantity that measures the strength of interaction in many body system. Our main ingredient of capturing salting in/salting out and Hofmeister effect is salt entropy[1;2]. In their isotropic model, Schmit et al (2010, 2011) accounted for monopole charge and designed an analytical theory to model the salting out of proteins. In their model, they demonstrated that the screening of monopole repulsion is the cause for non–linear salting out. More importantly, their study also demonstrated the importance of salt entropy such that the free energy associated with the protein aggregate or crystal is dominated by the entropy of mobile salt ions.

Knowing that solubility increases with salt concentration in salting in phenomenon, we primarily modeled it accounting the dipole term in the charge distribution which was absent in the Refs.[1;2]. The idea to account for the dipole in describing salting in is inspired by Tanford (1966) who qualitatively proposed that salting in is resulted by the screening of dipole attraction[3]. The reality with dipole is that it becomes noticeable only if monopole charge is weak. It happens at or close to isoelectric point of protein where the average charge of protein is zero or small. In such condition the screening of dipole attraction by adding salt to the solution make protein–protein interaction less attractive giving salting in. Salting in is unfavorable for protein aggregation. One way to avoid possible protein aggregation is to shift solution pH towards the isoelectric point of protein. This explanation is valid for proteins which show salting in behavior near their isoelectric point at low salt regime. There are number of proteins showing salting in behavior near their isoelectric point[3–6].

Often, salting in is observed at low salt and it is followed by salting out as the salt

concentration increases. In salting out, the protein solubility decreases with the increase in salt concentration. It is observed in number of different solution conditions. For instances, if solution pH is away from the isoelectric point then the dipole interaction is negligible compared to monopole repulsion so the addition of salt screens electrostatics resulting in salting out. This example is true only when salt concentration isn't high enough to screen all the electrostatic interaction. Once the electrostatic interaction is highly screened at high salt then the salt mediated hydrophobic interaction could cause the salting out. This could be true in high salt without depending on the pH of the solution. The Cohn emperical formula describes salting out, but this formula predicts only the linear solubility[7]. Many experiments have measured the non–linear solubility in number of protein which can't be described by using the Cohn formula[1;8;9]. The non–linear salting out seen at electrostatic regime of salt concentration has been modeled by Schmit et al. using the Poisson Boltzmann equation by accounting the screening of monopole repulsion[1] and our model has generalized and expanded this approach.

We accounted for effect of anion binding on the protein–protein interaction in addition to the screening effect. A number of experiments have shown that anions are the primary binding species to proteins regardless of the charge state of protein[10–12]. These experiments have also revealed how the charge of the protein changes with the number of bound ions. Keeping this effect in our model, we evaluated the number of bound anions to the protein using a grand cannonical partition function and used it to calculate the effective charge and dipole moment. The inclusion of anion binding in the theory has multiple effects on the solubility of protein depending on the solution conditions such as pH and salt concentration. Near the isoelectric point or above it, anion binding causes salting in enhancement by increasing the monopole repulsion whereas below from isoelectric point, it causes salting out by reducing the monopole repulsion. Of course, the number of anions binding to the protein surface depends on the binding affinity, binding sites available, and salt concentration.

We have also accounted for the depletion effect. This effect plays significant role once the salt particles occupy a significant amount of volume in the system[13;14]. This is the case at high concentration and in such high salt almost all the electrostatic effect gets screened.

In our model of chymosin protein, the salting out observed following the salting in at high salt concentration ($\approx 0.3M$) was caused by the depletion interaction. In the same protein, we observed only salting out at pH 6 which was due to the combined effect of screening monopole repulsion and depletion interaction but the dominant cause of that salting out came from the screening effect of salt. pH 6 is far enough from the isoelectric point that the dipole effect is unnoticeable. We modeled this effect by considering the excluded volume for finite size of salt in the solution state and in the aggregate state. This interaction depends on the identity of both protein and salt as well as on the concentration of salt. As the salt concentration increased we observed more salting out behavior and the non–linearity of salting out observed at low salt regime turned to a linear behavior because of the weak electrostatic effect at such high salt.

Our model captured salt specific effects popularly known as the Hofmeister effect. We included Coulomb energy, salt entropy and ion–protein interaction terms in our master free energy equation. Besides the ion–protein interaction term, the entropic term also contained salt specific effects. For this purpose, we treated salt ions as a finite size entities such that the size of one ion differs from the size of another ion. We took hydrated ion size to include the ion–water interaction implicitly. Note that, the salt specific effect can't be explained solely by using $DLVO$ theory because it treats ions with equal valency as the same. Our model of protein aggregate required to confine salt ions in the aggregate cavities to make the whole aggregate electrically neutral. The ion accessible volume in aggregate cavity differs from one ion to another because salt ions have different hydrated radii.

Ion size has immediate impact on entropy and hence on protein solubility in entire range of salt concentration. At low salt, entropy of salt ions affects electrostatic interactions and at high salt, entropy affects depletion interaction. Counter–ions size shows opposite effect on protein solubility at low salt and high salt producing a reversal in the Hofmeister series. Bigger counter–ions make the protein more soluble at low salt than smaller counter–ions do. On the other hand, at high salt, smaller counter–ions are less favorable for protein aggregation than bigger counter–ions. This means in order to precipitate protein at low salt (below $\approx 1M$), it is wise to use counter–ions which have a small hydrated radii so that they

can easily fit to the aggregate cavity and make it easier for the protein aggregate to satisfy the neutrality condition. But, at high salt (above $\approx 1M$), it is better to select counter–ions with high charge density (so their hydrated size is large) so that they enhance depletion attraction and force proteins to precipitate. But, co–ions size have similar effect on protein solubility in entire salt concentration. Big hydrated co–ions are always favorable for protein aggregation. At low salt, their exclusion from the cavity make aggregate simple to attain neutrality condition as a result of which solubility gets reduced and at high salt, they enhance depletion attraction as done by big counter–ions which also decreases solubility.

To sum up the equilibrium study of crystal growth, our model successfully described the salting in/salting out and the Hofmeister effect. The salting in effect observed in chymosin protein near its isoelectric point was due to the combined effect of dipole screening and monopole enhancement and the follow up salting out effect was due to the enhanced depletion effect. The pure salting out observed in the same protein far from isoelectric point was due to the combination of monopole screening and salt mediated depletion attraction. The salt specific effect seen in salting out of lysozyme protein was the combined effect come from the exclusion of bigger ions from the aggregate and the enhanced salt specific depletion effect. In existing studies of salt specific effects people considered ion–protein dispersion interaction (Ninham et al)[15], image charge effect (Zhou et al)[16], hydrophobic effect (Levin et al)[17] as a source of specificity but the contribution of salt specific ion entropy has been neglected. But, these existing studies dealt mostly with two body interactions, such as the second virial coefficient so the neglect of ion entropy might have been a reasonable approximation. But, when dealing with the protein aggregates, entropy must not be ignored because it plays the dominant role in the free energy. Further, entropy is a source of salt specific interaction. So, here, we emphasize that if we could design a unified theory by including existing salt specific term plus salt entropy term then it will be more accurate and will be valuable contribution to this field.

The kinetics of crystal growth is difficult especially for large and fragile molecules like proteins. Crystal growth requires a search for the rare correct binding event. For an example, in protein crystal growth, the probability of a correct binding event is very low (10^{-4}

or 10^{-5})[18]. Due to the rare nature of correct binding events, kinetics often favor disordered assemblies because the probability of molecules to bind in incorrect ways is high. Alternatively, kinetic factors could result in no assembly at all. One reason behind the last type of possibility is a phenomenon called self–poisoning. If crystal growth is suffered by the self–poisoning then the growth rate shows the non–monotonic behavior with the thermodynamic driving force. Our kinetic study showed that the crystal growth from molecules possessing many orientational and conformational degree of freedoms is vulnerable to be suffered by self–poisoning.

We used Monte Carlo simulation and predicted minimal physical requirements for crystal growth self–poisoning. We found that self–poisoning phenomenon required only three ingredients to happen. In crystal growth, the binding energy should be realized in two ways such as if a molecule binds in a crystallographic way then the associated binding energy should be stronger whereas if it binds in non–crystallographic way then the energy of interaction should be weaker. Further, these two binding energies should come with sufficiently unequal probability such that the non–crystallographic binding is more probable than crystallographic binding. In our simulations, we used 10^{-2} as a probability of events associated with crystallographic binding and we selected different values of binding energies. By using these conditions, we showed that the growth rate suffered from self–poisoning. These minimal requirements can easily be satisfied by proteins and other flexible polymers so the lack of crystal growth from such molecules in experiments could be due to the self–poisoning. Our work also showed how the small changes in the energy parameters can be used to avoid self–poisoning or to recover crystal growth. Experimentally, binding parameters could be increased or decreased by changing the pH of the solution, salt concentration or salt type etc.

In our studies, we used simple coarse–grained model to capture the important interactions. In reality, atomic details are important in the interactions. However, high resolution models are impossible to solve analytically. All atom simulations are the alternative but such simulations are computationally expensive becasue of the inclusion of the huge number of atoms in the system. In addition, in all atom simulation model it is difficult to figure

out the role of different interactions in certain phenomenon unlike in analytical model where we can easily realize the consequence of changing certain parameters. Existing biomolecular simulation packages such as CHARMM are designed to study all atom interactions but they don't account the entropic effect of salt. This means they are ignoring this important effect in their force field because previous study and our current study have proven that the entropy of salt plays major role in protein–protein interaction particularly while forming the aggregates. So, our claim is that such simple analytical model should get more attention while designing the higher resolution force fields needed for capturing complex biological systems.

4.2 Future direction

Our current model compared well with the experiments[6] showing the non–monotonic solubility with salt concentration at number of pHs. In this work, non–monotonicity of solubility means the salting in behavior at low salt was followed by salting out at high salt. Further, the current model captured the monotonic experimental results quantitatively which included only salting out in entire salt concentration. For example, salting out of chymosin[6] at pH 6.0, salt specific salting out of lysozyme[8] at pH 4.5. But, there are some unique experimental observations in which salting out at low salt is followed by salting in at high salt such as the experiments of Broide et al (1996)[19] and Ries–Kautt et al (2002)[20]. In these experiments, the observed salting in lies outside the range of concentration in which electrostatic is believed to be significant ($\approx 1M$). These experiments were performed far from isoelectric point of proteins which means monopole interaction is playing a dominant role. Furthermore, the clear salting out–salting in behavior were seen when one of the salt ion was not monovalent. We can observe salting out followed by salting in below ($1M$) in lysozyme protein by using our current model but for this purpose we have to increase our binding site parameter $N_s = 18$ by almost 50%. Note that, we were using only salts made by monovalent ions in the current model. So, we believe that a new model can be developed in future for higher valency ions keeping the heart of the current model the same to capture

salting out–salting in quantitatively.

There are also many experiments measuring the salt specific cloud point[19;21] of proteins and these measurements have shown the reversal Hofmeister series at low salt and at high salt. Our current model measured the reversal Hofmeister series of protein solubility qualitatively. This model showed that the volume fraction of solvent in the protein aggregate can shift the monopole and dipole dominant region significantly. The solvent volume fraction is a key parameter that is connected with the cloud point of proteins. So, in future, we can test our current model by comparing with the experimental results for the cloud point temperature.

4.3 Bibliography

Bibliography

[1] Jeremy D. Schmit, and Ken A. Dill, *The Stabilities of Protein Crystals*, J. Phys. Chem. B, **114**, 4020–4027, 2010

[2] J. D. Schmit, S. Whitelam, and K. A. Dill, *The Journal of Chemical Physics*, **135**, 085103 (2011).

[3] C. Tanford, *Physical Chemistry of Macromolecules [Hardcover]*, John Wiley & Sons, Inc, 1966.

[4] J. Mellanby, *Globulin*, J. Physiol., **33**, 338–373, 1905.

[5] Andre C. Dumetz, Aaron M. Chockla, Eric W. Kaler, Abraham M. Lenhoff, *Effects of pH on proteinprotein interactions and implications for protein phase behavior*, Biochimica et Biophysica Acta, **1784**, 600–610, 2008.

[6] Ronald W. Maurer, Stanley I. Sandler, Abraham M. Lenhoff, *Salting-in characteristics of globular proteins*, Biophysical Chemistry, **156**, 72–78, (2011).

[7] E.J. Cohn, *The physical chemistry of the proteins*, Physiol. Rev., **5**, 349–427, 1925.

[8] Riess–Kautt MM, Ducruix AF, *Relative effectiveness of various ions on the solubility and crystal-growth of lysozyme*, J. Biol. Chem., **264**, 745–748, 1989.

[9] Forsythe, E., Judge, R., Pusey, M, *J. Chem. Eng. Data*, **44**, 637640, (1999).

[10] Yatin R. Gokarn, R. Matthew Fesinmeyer, Atul Saluja, Vladimir Razinkov, Susan F. Chase, Thomas M. Laue, and David N. Brems, *Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions*, Protein Science, **20**, 580–587, 2011.

[11] Manoj K. Menon and Andrew L. Zydney, *Measurement of Protein Charge and Ion Binding Using Capillary Electrophoresis*, Anal. Chem., **70**, 1581–1584, 1998.

[12] Luca Medda, Brajesh Barse, Francesca Cugia, Mathias Bostrom, Drew F. Parsons, Barry W. Ninham, Maura Monduzzi, and Andrea Salis, *Hofmeister Challenges: Ion Binding and Charge of the BSA Protein as Explicit Examples*, Langmuir, **28**, 16355–16363, 2012.

[13] Asakura, S., and F. Oosawa, *Interactions between particles suspended in solutions of macromolecules*, J. Polym. Sci. [B], **33**, 183–192, 1958.

[14] Kim A. Sharp, *Analysis of the size dependence of macromolecular crowding shows that smaller is better*, PNAS, **112**, 7990–7995, 2015.

[15] B. W. Ninham and V. Yaminsky, *Langmuir*, **13**, 2097 (1997).

[16] Huan-Xiang Zhou, *Interactions of Macromolecules with Salt Ions: An Electrostatic Theory for the Hofmeister Effect*, PROTEINS: Structure, Function, and Bioinformatics, **61**, 69–78, (2005).

[17] A. P. dos Santos and Y. Levin, *Phys. Rev. Lett.*, **106**, 1 (2011).

[18] Schmit J. D. and Dill K. A, *Journal of the American Chemical Society*, **134(9)**, 3934–3937 ,(2012).

[19] Michael L. Broide, Tina M. Tominc, and Marc D. Saxowsky, *Using phase transitions to investigate the effect of salts on protein interactions*, Phys. Rev. E, **53(6)**, 1996.

[20] Philippe Bnas, Laurent Legrand and Madeleine Ris-Kautt, *Strong and specific effects of cations on lysozyme chloride solubility*, Acta Cryst., **D58**, 1582–1587, 2002.

[21] Yanjie Zhang and Paul S. Cremer, *The inverse and direct Hofmeister series for lysozyme*, PNAS , **106(36)**, 15251, 2009.

Appendix A

Solution state calculation

A.1 Electric potential inside and outside the protein

The linearized Poisson Boltzmann equation is,

$$\begin{aligned} \nabla^2\Phi &= \frac{2e^2c_0}{\epsilon k_B T}\Phi \\ &= \kappa^2\Phi \end{aligned} \tag{A.1}$$

Where, $\kappa^2 = \frac{2e^2c_0}{\epsilon k_B T}$ and $\kappa^{-1} = \sqrt{\frac{\epsilon k_B T}{2e^2c_0}}$ is the Debye screening length.

(∇^2) term in spherical polar coordinate with azimuthal symmetry can be written as,

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial}{\partial r}) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta}) \tag{A.2}$$

Using equation (A.2), equation (A.1) can be written as,

$$\frac{\partial^2\Phi}{\partial r^2} + \frac{2}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial\Phi}{\partial\theta} - \kappa^2\Phi = 0 \tag{A.3}$$

The electric potential (Φ) is space(r) and angle dependent(θ). Using the separation of variables, we can write,

$$\Phi(r,\theta) = R(r)\Theta(\theta) \tag{A.4}$$

using (A.4), equation (A.3) becomes,

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} - \kappa^2 r^2 + \frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{d\Theta}{d\theta} = 0 \tag{A.5}$$

Let us solve angular part first,

Let,

$$\begin{aligned} \frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} + \frac{1}{\Theta}\frac{\cos\theta}{\sin\theta}\frac{d\Theta}{d\theta} &= -l(l+1) \\ \frac{d}{d\theta}(\sin\theta\frac{d\Theta}{d\theta}) + l(l+1)\sin\theta\Theta &= 0 \end{aligned} \tag{A.6}$$

Where l is the constant of separation.

To convert equation (A.6) in to standard Legendre polynomial, let us change variables,

$$\begin{aligned} x &= \cos\theta \\ dx &= -\sin\theta d\theta \\ \frac{d}{d\theta} &= -\sin\theta\frac{d}{dx} \end{aligned} \tag{A.7}$$

Also,

$$\Theta(\theta) = Y(x) \tag{A.8}$$

Using equations (A.7) and (A.8) in equation (A.6) and solving, we get,

$$(1-x^2)\frac{d^2Y}{dx^2} - 2x\frac{dY}{dx} + l(l+1)Y = 0 \tag{A.9}$$

The solution of (A.9) is the Legendre polynomial of order l and its solution is [Arfken and

Weber].

$$Y(x) = \Theta(\theta) = \sum_l P_l(\cos\theta) \tag{A.10}$$

It gives the solution of angular part. To get the solution of radial part, we go back to equation (A.5),

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} - \kappa^2 r^2 - l(l+1) = 0 \tag{A.11}$$

$$\frac{d^2R}{d^2r} + \frac{2}{r}\frac{dR}{dr} - (\kappa^2 + \frac{l(l+1)}{r^2})R = 0 \tag{A.12}$$

let,

$$R(r) = \frac{u(r)}{\sqrt{\kappa r}} \tag{A.13}$$

Now, the equation (A.12), with the help of equation (A.13) becomes,

$$r^2\frac{d^2u}{dr^2} + r\frac{du}{dr} - (\kappa^2 r^2 + (l+\frac{1}{2})^2)u(r) = 0 \tag{A.14}$$

Equation (A.14) is the modified Bessel equation [Arfken and Weber] and its solution is

$$u(r) = AI_{l+\frac{1}{2}}(\kappa r) + BK_{l+\frac{1}{2}}(\kappa r) \tag{A.15}$$

Using equation (A.15) in equation (A.13) we get,

$$R(r) = A_l i_l(\kappa r) + B_l k_l(\kappa r) \tag{A.16}$$

where A_l and B_l are constants and

$$i_l(\kappa r) = \sqrt{\frac{\pi}{2}}\frac{I_{l+\frac{1}{2}}(\kappa r)}{\sqrt{\kappa r}} \tag{A.17}$$

$$k_l(\kappa r) = \sqrt{\frac{2}{\pi}}\frac{K_{l+\frac{1}{2}}(\kappa r)}{\sqrt{\kappa r}} \tag{A.18}$$

are spherical modified Bessel functions first and second kind respectively [Arfken and Weber].

Now, for the complete solution, we use equations (A.10) and (A.16) in equation (A.4).

$$\Phi(r,\theta) = \sum_l [A_l i_l(\kappa r) + B_l k_l(\kappa r)] P_l(\cos\theta) \tag{A.19}$$

Equation (A.19) is the potential due to a charged sphere. According to the properties of spherical Bessel function, its first part, i.e. $i_l(\kappa r)$ which is modified spherical bessel function of first kind, diverges at infinity so we drop it from the potential expression and we keep only the second kind of modified spherical bessel function ($k_l(\kappa r)$) in the expression. The electric potential containing second kind of modified spherical bessel function is,

$$\begin{aligned} \Phi_{out}(r,\theta) &= \sum_l B_l k_l(\kappa r) P_l(\cos\theta) && \text{(A.20)} \\ &= B_0 k_0(\kappa r) + B_1 k_1(\kappa r)\cos\theta && \text{(A.21)} \end{aligned}$$

To calculate the electric potential inside protein, we use Laplace equation which is,

$$\nabla^2 \Phi_{in} = 0 \tag{A.22}$$

The solution of equation (A.22) is

$$\begin{aligned} \Phi_{in} &= \sum_l A_l r^l P_l(\cos\theta) \\ &= A_0 + A_1 r\cos\theta && \text{(A.23)} \end{aligned}$$

Now, at the surface of the protein (boundary condition),

$$\Phi_{in} \mid (r = R) = \Phi_{out} \mid (r = R)$$

$$A_0 + A_1 R\cos\theta = B_0 k_0(\kappa R) + B_1 k_1(\kappa R)\cos\theta \tag{A.24}$$

Equating the coefficients we get,

$$A_0 = B_0 k_0(\kappa R) \tag{A.25}$$

and

$$B_1 = \frac{A_1 R}{k_1(\kappa R)} \tag{A.26}$$

Another boundary condition is,

$$\epsilon_w E_2 - \epsilon_p E_1 = \sigma_0 + \sigma_1 \cos\theta \tag{A.27}$$

Where,

$$\begin{aligned} E_2 &= -\frac{\partial \Phi_{out}}{\partial r} \mid (r = R) \\ &= -B_0 k_0'(\kappa R) - B_1 k_1'(\kappa R) \cos\theta \end{aligned} \tag{A.28}$$

Also,

$$\begin{aligned} E_1 &= -\frac{\partial \Phi_{in}}{\partial r} \mid (r = R) \\ &= -A_1 \cos\theta \end{aligned} \tag{A.29}$$

Using equations (A.28) and (A.29), equation (A.27) becomes,

$$\begin{aligned} \epsilon_w(-B_0 k_0'(\kappa R) - B_1 k_1'(\kappa R)\cos\theta) & \\ -\epsilon_p(-A_1 \cos\theta) & \\ = \sigma_0 + \sigma_1 \cos\theta & \end{aligned} \tag{A.30}$$

By solving equation (A.30), we will get the coefficients,

$$B_0 = -\frac{\sigma_0}{\epsilon_w k_0'(\kappa R)} \tag{A.31}$$

$$A_0 = -\frac{\sigma_0 k_0(\kappa R)}{\epsilon_w k_0'(\kappa R)} \tag{A.32}$$

$$A_1 = \frac{\sigma_1}{[\epsilon_p - \epsilon_w R\frac{k_1'(\kappa R)}{k_1(\kappa R)}]} \tag{A.33}$$

and

$$B_1 = \frac{\sigma_1 R}{k_1(\kappa R)[\epsilon_p - \epsilon_w R\frac{k_1'(\kappa R)}{k_1(\kappa R)}]} \tag{A.34}$$

Now, equation (A.23) becomes,

$$\Phi_{in} = -\frac{\sigma_0 k_0(\kappa R)}{\epsilon_w k_0'(\kappa R)} + \frac{\sigma_1}{[\epsilon_p - \epsilon_w R\frac{k_1'(\kappa R)}{k_1(\kappa R)}]} r\cos\theta \tag{A.35}$$

We can replace modified spherical bessel function of second kind k_l by [Arfken and Weber],

$$k_l(x) = (-1)^l x^l (\frac{1}{x}\frac{d}{dx})^l \frac{e^{-x}}{x} \tag{A.36}$$

Here, κr is replaced by x for simplicity.

For $l = 0$,

$$k_0(\kappa r) = \frac{e^{-\kappa r}}{\kappa r} \tag{A.37}$$

$$k_0'(\kappa R) = -\frac{e^{-\kappa R}}{\kappa R^2}(1 + \kappa R) \tag{A.38}$$

$$\frac{k_0(\kappa r)}{k_0'(\kappa R)} = -\frac{R^2}{r}\frac{e^{-\kappa(r-R)}}{(1+\kappa R)} \tag{A.39}$$

Similarly, for $l = 1$,

$$k_1(\kappa r) = \frac{e^{-\kappa r}}{\kappa^2 r^2}(1 + \kappa r) \tag{A.40}$$

$$k_1'(\kappa R) = -\frac{e^{-\kappa R}}{\kappa^2 R^3}(2 + 2\kappa R + \kappa^2 R^2) \tag{A.41}$$

$$\frac{k_1(\kappa r)}{k_1'(\kappa R)} = -\frac{R^3}{r^2}\frac{e^{-\kappa(r-R)}(1+\kappa r)}{(2+2\kappa R+\kappa^2 R^2)} \tag{A.42}$$

Now, using

$$\frac{k_0(\kappa R)}{k_0'(\kappa R)} = -\frac{R}{(1+\kappa R)} \tag{A.43}$$

and

$$\frac{k_1'(\kappa R)}{k_1(\kappa R)} = -\frac{2+2\kappa R+\kappa^2 R^2}{R(1+\kappa R)} \tag{A.44}$$

equation (A.35) becomes,

$$\Phi_{in} = \frac{\sigma_0 R}{\epsilon_w(1+\kappa R)} + \frac{\sigma_1 r\cos\theta}{[\epsilon_p + \epsilon_w \frac{(2+2\kappa R+\kappa^2 R^2)}{(1+\kappa R)}]} \tag{A.45}$$

This equation (A.45) gives the potential inside the protein.

Now, equation (A.21) becomes,

$$\Phi_{out} = -\frac{\sigma_0 k_0(\kappa r)}{\epsilon_w k_0'(\kappa R)} + \frac{\sigma_1 R k_1(\kappa r)\cos\theta}{k_1(\kappa R)[\epsilon_p - \epsilon_w R\frac{k_1'(\kappa R)}{k_1(\kappa R)}]} \tag{A.46}$$

Using

$$\frac{k_0(\kappa r)}{k_0'(\kappa R)} = -\frac{R^2}{r}\frac{\exp[-\kappa(r-R)]}{(1+\kappa R)} \tag{A.47}$$

$$\frac{k_1(\kappa r)}{k_1(\kappa R)} = \frac{R^2}{r^2}\frac{(1+\kappa r)}{(1+\kappa R)}\exp[-\kappa(r-R)] \tag{A.48}$$

and

$$\frac{k_1'(\kappa r)}{k_1(\kappa R)} = -\frac{(2+2\kappa R+\kappa^2 R^2)}{R(1+\kappa R)} \tag{A.49}$$

equation (A.46) becomes,

$$\begin{aligned}\Phi_{out} \;=\; & \frac{\sigma_0}{\epsilon_w}\frac{R^2}{r}\frac{\exp[-\kappa(r-R)]}{(1+\kappa R)} \\ & + \frac{\sigma_1 R\cos\theta}{[\epsilon_p + \epsilon_w \frac{(2+2\kappa R+\kappa^2 R^2)}{(1+\kappa R)}]} \times \\ & \frac{R^2}{r^2}\frac{(1+\kappa r)}{(1+\kappa R)}\exp[-\kappa(r-R)]\end{aligned} \tag{A.50}$$

This gives the electric potential distribution in the salt solution.

A.2 Ion binding to the protein

We account for only anion binding event. We do this using a grand cannonical partition function.

$$Z = \sum_{n_b} \frac{N!}{(N-n_b)!n_b!} \exp[-ES + n_b(\mu - E_b)] \tag{A.51}$$

where, N, n_b, ES , μ, E_b are number of binding sites in protein, number of anion bound in protein, electrostatic energy between protein and anion that bind to the protein, chemical potential and binding energy of anion respectively. The electrostatic potential due to a spherically modeled protein on its surface is given by equation (A.52).

$$\begin{aligned} \Phi_{r=R} &= \frac{n_0 e}{4\pi\epsilon_w(\kappa R+1)R} \\ &\quad + \frac{n_1 e}{4\pi\epsilon_w R} \frac{(\kappa R+1)\cos\theta}{(2+2\kappa R+\kappa^2R^2+\frac{\epsilon_p}{\epsilon_w}(\kappa R+1))}. \end{aligned} \tag{A.52}$$

In short,

$$\Phi_{r=R} = An_0 + ABn_1$$

where,

$$A = \frac{e}{4\pi\epsilon_w(\kappa R+1)R}$$

$$B = \frac{(\kappa R+1)^2\cos\theta}{2+2\kappa R+\kappa^2R^2+\frac{\epsilon_p}{\epsilon_w}(\kappa R+1)}$$

The electrostatic energy between protein and anion that bind to the protein is,

$$ES = \sum_{N'=0}^{n_b-1} [A(n_0 - N') + AB(n_1 - N')](-e) \tag{A.53}$$

For simplicity, let us insert e into A then

$$ES = \sum_{N'=0}^{n_b-1} [-A(n_0 - N') - AB(n_1 - N')] \tag{A.54}$$

Now, A is modified to

$$A = \frac{e^2}{4\pi\epsilon_w(\kappa R + 1)R}$$

To find the sum, we used the following formula.

$$\sum_{N'=0}^{n_b-1}(n_0 - N') = \frac{1}{2}n_b(1 + 2n_0 - n_b) \tag{A.55}$$

$$\sum_{N'=0}^{n_b-1}(n_1 - N') = \frac{1}{2}n_b(1 + 2n_1 - n_b) \tag{A.56}$$

Now, the electrostatic energy becomes,

$$\begin{aligned} ES &= \frac{-An_b}{2}(1 + B + 2n_0 + 2n_1B) + \frac{n_b^2}{2}A(1 + B) \\ -ES &= \frac{An_b}{2}(1 + B + 2n_0 + 2n_1B) - \frac{n_b^2}{2}A(1 + B) \end{aligned} \tag{A.57}$$

Now, partition function becomes,

$$Z = \sum_{n_b}\frac{N!}{(N - n_b)!n_b!}\exp[cn_b]\exp[-d^2n_b^2] \tag{A.58}$$

Where,

$$c = \frac{A}{2}(1 + B + 2n_0 + 2n_1B) + \mu - E_b \tag{A.59}$$

$$d = \sqrt{\frac{A(1 + B)}{2}} \tag{A.60}$$

Let, $n_b = N/2 - s$ then equation (A.58) becomes,

$$\begin{aligned} Z = &\sum_{s=-\frac{N}{2}}^{\frac{N}{2}}\frac{N!}{(\frac{N}{2} - s)!(\frac{N}{2} + s)!}\exp[c(\frac{N}{2} - s)] \\ &\exp[-d^2(\frac{N}{2} - s)^2] \end{aligned} \tag{A.61}$$

Using an approximation,

$$\frac{N!}{(\frac{N}{2}-s)!(\frac{N}{2}+s)!} = g(N,0)\exp[-\frac{2s^2}{N}] \tag{A.62}$$

equation (A.61) becomes,

$$\begin{aligned} Z &= \sum_{s=-\frac{N}{2}}^{\frac{N}{2}} g(N,0)\exp[-\frac{2s^2}{N}]\exp\left[c(\frac{N}{2}-s)\right] \\ &\quad \exp\left[-d^2(\frac{N}{2}-s)^2\right] \end{aligned} \tag{A.63}$$

Where,

$$g(N,0) = \sqrt{\frac{2}{\pi N}}2^N \tag{A.64}$$

After the simplification of equation (A.63) and changing the summation into integration, we get

$$Z = T\int_{-\frac{N}{2}}^{\frac{N}{2}} \exp[-a(s-b)^2]ds \tag{A.65}$$

Where,

$$T = g(N,0)\exp[\frac{Nc}{2} - \frac{N^2d^2}{4} + \frac{(Nd^2-c)^2}{4(\frac{2}{N}+d^2)}] \tag{A.66}$$

$$a = (\frac{2}{N}+d^2) \tag{A.67}$$

$$b = \frac{(Nd^2-c)}{2(\frac{2}{N}+d^2)} \tag{A.68}$$

From equation (A.65), we get the number of anion binding to the protein,

$$n_b = \left.\frac{\partial\ \log Z}{\partial\mu}\right|_{\mu=k_BT\ln[c_s]} \tag{A.69}$$

A.3 Dipole moment calculation

From the distribution amino acids, we calculate the dipole moment of a protein. The dipole moment is given by following expression,

$$\begin{aligned} p &= \int_v r\rho(r)dv \\ &= \int_{r=R}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} r\sigma_1\cos\theta\delta(r-R)r^2dr\sin\theta d\theta \\ & \quad d\phi \end{aligned} \tag{A.70}$$

We calculate the component of dipole along x, y, and z direction. Along $x-$ direction,

$$\begin{aligned} p_x &= \int_{r=R}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}(r)_x\sigma_1\cos\theta\delta(r-R) \\ & \quad r^2dr\sin\theta d\theta d\phi \\ &= \int_{r=R}^{\infty}\sigma_1 r^3\delta(r-R)dr\int_{\theta=0}^{\pi}\cos\theta\sin^2\theta d\theta \\ & \quad \int_{\phi=0}^{2\pi}\cos\phi d\phi \\ &= 0 \end{aligned} \tag{A.71}$$

Along $y-$ direction,

$$\begin{aligned} p_y &= \int_{r=R}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}(r)_y\sigma_1\cos\theta\delta(r-R) \\ & \quad r^2dr\sin\theta d\theta d\phi \\ &= \int_{r=R}^{\infty}\sigma_1 r^3\delta(r-R)dr\int_{\theta=0}^{\pi}\cos\theta\sin^2\theta d\theta\int_{\phi=0}^{2\pi}\sin\phi d\phi \\ &= 0 \end{aligned} \tag{A.72}$$

Along $z-$ direction,

$$\begin{aligned} p_z &= \int_{r=R}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}(r)_z\sigma_1\cos\theta\delta(r-R)r^2dr\sin\theta d\theta d\phi \\ &= \int_{r=R}^{\infty}\sigma_1 r^3\delta(r-R)dr\int_{\theta=0}^{\pi}\cos^2\theta\sin\theta d\theta\int_{\phi=0}^{2\pi}d\phi \\ &= \int_{r=R}^{\infty}\frac{n_1 e}{4\pi r^2}r^3\delta(r-R)dr\int_{\theta=0}^{\pi}\cos^2\theta\sin\theta d\theta\int_{\phi=0}^{2\pi}d\phi \\ &= \frac{n_1 e}{4\pi}\int_{r=R}^{\infty}r\delta(r-R)dr(\frac{2}{3})2\pi \\ &= \frac{n_1 e}{3}R \\ &= \frac{n_1 eR}{3} \end{aligned} \tag{A.73}$$

The resultant theoritical dipole moment is,

$$\begin{aligned} p &= \sqrt{p_x^2+p_y^2+p_z^2} \\ &= \frac{n_1 eR}{3} \\ n_1 &= \frac{3p}{eR} \end{aligned} \tag{A.74}$$

In equation (A.74), p is dipole moment in unit eA which can be obtained by knowing the distribution of charged amino acids, R is the radius of protein and n_1 is the dipole charge.

Appendix B

Aggregate state calculation

The dimensionless non–linear PB equation is,

$$\nabla_y^2 \Phi = \sinh \Phi \tag{B.1}$$

Where,

$$\Phi = \frac{e\Psi}{k_B T}$$

is dimensionless potential and

$$y = \kappa r$$

is dimensionless length.

Now, we linearize the potential (Φ) around the average potential (ϕ_0) such that $\Phi(y) = \phi(y) + \phi_0$, so we have,

$$\begin{aligned} \nabla_y^2 \Phi &= \sinh(\phi + \phi_0) \\ \nabla_x^2 \phi &= \phi + \tanh \phi_0 \end{aligned} \tag{B.2}$$

Where,

$$x = \sqrt{\cosh \phi_0} y \tag{B.3}$$

and the non-zero averaged potential (ϕ_0) is obtained by using the neutrality condition in the cavity as follows.

$$(n_0 - n_b) \quad = \quad -(\bar{c_+}v_+ - \bar{c_-}v_-) \tag{B.4}$$

where v_+ and v_- are the accessible volume for positive and negative ions in the cavity respectively. The ion accessible volume in the cavity, $6\pi((R_c - R_i)^2 - R_{in}^2)R$, depends on the size of ion R_i. $\bar{c_+} = c_s e^{\frac{-e\phi_0}{k_B T}}$ and $\bar{c_-} = c_s e^{\frac{e\phi_0}{k_B T}}$ are the concentration of positive and negative ions inside the cavity in the presence of average potential.

To make problem calculation friendly, let us consider that the solvent is trapped between two concentric cylinders. Let R_c and R_{in} be the radii of outside and inside cylinder respectively. Now, the solution of equation (B.2) in cylindrical coordinate is,

$$\phi(x) = AI_0[x] + BK_0[x] - \tanh\phi_0 \tag{B.5}$$

Let,

$$\alpha = \sqrt{\cosh\phi_0}\kappa R_c \tag{B.6}$$

$$\beta = \sqrt{\cosh\phi_0}\kappa R_{in} \tag{B.7}$$

To find the constants, we use the boundary conditions. Electric field at the outer bound-

ary of the cavity is,

$$\begin{aligned} E_c &= -\frac{d\psi}{dr} \\ \frac{qe}{2\pi\epsilon_w rl} &= \frac{k_B T}{e}\frac{d\phi}{dr} \\ \frac{qe\kappa}{2\pi\epsilon_w yl} &= \frac{k_B T\kappa}{e}\frac{d\phi}{dy} \\ \frac{qe\kappa\sqrt{\cosh\phi_0}}{2\pi\epsilon_w xl} &= \frac{k_B T\sqrt{\cosh\phi_0}\kappa}{e}\frac{d\phi}{dx} \\ \frac{qe^2}{2\pi\epsilon_w k_B T\alpha l} &= \frac{d\phi}{dx}|_{x=\alpha} \\ E_0 &= AI_1[\alpha] - BK_1[\alpha] \end{aligned} \tag{B.8}$$

Where,

$$E_0 = \frac{qe^2}{2\pi\epsilon_w k_B T\alpha l} \tag{B.9}$$

The second boundary condition tells that the electric field vanishes at the inner boundary of the cavity due to the charge symmetry,

$$\begin{aligned} \frac{d\phi}{dx}|_{x=\beta} &= 0 \\ AI_1[\beta] - BK_1[\beta] &= 0 \\ B = A\frac{I_1[\beta]}{K_1[\beta]} & \end{aligned} \tag{B.10}$$

Solving equation (B.8) with the help of equation (B.10), we get,

$$A = \frac{E_0 K_1[\beta]}{[I_1[\alpha]K_1[\beta] - I_1[\beta]K_1[\alpha]]} \tag{B.11}$$

Now, equation (B.10) becomes,

$$B = \frac{E_0 I_1[\beta]}{[I_1[\alpha]K_1[\beta] - I_1[\beta]K_1[\alpha]]} \tag{B.12}$$

Now, equation (B.5) becomes,

$$\phi(x) = \frac{E_0 K_1[\beta] I_0[x] + E_0 I_1[\beta] K_0[x]}{[I_1[\alpha] K_1[\beta] - I_1[\beta] K_1[\alpha]]} - \tanh \phi_0 \tag{B.13}$$

Putting Eq.(B.13) in $\Phi = \phi + \phi_0$ gives the electric potential in the cavity.