

The Effect of Osmolytes on Protein Folding

by

Shirin Hadizadeh

B. Physics, Shahid Beheshti University, Tehran, 2000

M. Physics, Shahid Beheshti University, Tehran, 2003

M. Physics, University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Physics)

The University Of British Columbia

(Vancouver)

December 2010

© Shirin Hadizadeh, 2010

Abstract

Living cells are composed of a variety of biological macromolecules such as nucleic acid, metabolites, proteins and cytoskeletal filaments as well as other particles. The fraction of the cellular interior volume that is taken by these biomolecules is about 30%, leading to a highly crowded environment. Biomolecules present in an extremely dense environment inside a cell have a completely different set of kinetic and thermodynamic behavior than in a test tube [1–5]. Therefore comprehending the effect of crowding conditions on biological molecules is crucial to broad research fields such as biochemical, medical and pharmaceutical sciences. Experimentally, we are able to mimic such crowded environments; which are of more physiological relevance, by adding high concentrations of synthetic macromolecules into uncrowded buffers.

Theoretically, very little attention has been paid to the effects of the dense cellular cytoplasm on biological reactions. The purpose of this work is to investigate analytically the effects of crowding agents on protein folding and stability. We present a new parameter as the measure of the polymer size, which will substitute the traditional measurements of the radius of gyration of the polymer, R_G , and the end to end distance of a polymeric chain, R_{ee} . Using this quantity we derive an expression for the free energy of the polymer which can easily be generalized to provide the free energy of a protein. This mechanism enables us to study the effect of crowding on folding and stability of a protein. The stabilization effects of the crowding particles depend on the concentration and the size of the crowders and also the type of the crowding particles that are present in the system. In our calculations the type of the crowders is controlled by the parameter ϵ_{po} , that is the energetic parameter between the protein and surrounding macromolecules.

Preface

A version of chapter 4 and 5 is in preparation for publication. S. Hadizadeh and A. Linhananta and S. S. Plotkin, *Physical Review E*, 2010 ([6]). I conducted all the calculations, produced all the plots. The manuscript was written by me and Dr. Steven Plotkin.

Chapter 6 is based on our study on the simulations of the effect of osmolytes on protein folding which has been submitted for publication. A. Linhananta and S. Hadizadeh and S. S. Plotkin, *Biophys. J.*, 2010 ([7]). The manuscript was originally drafted by Apichart Linhananta and Dr. Steven Plotkin. Although the theory and simulations of this work have been progressing in parallel, my contribution to this manuscript is primarily in the general discussions about the choice of the problems that need to be simulated and the initial design of the simulation in order to have a reasonable harmony between my theoretical calculations and Dr. Apichart Linhananta's simulations, and I also wrote parts of the text. Initially there was a theory section in the manuscript with a few of my results for comparison, but we decided to move that section into the second manuscript ([6]) to present these two aspects of the problem (theory and simulations) separately. Therefore most of the plots that are currently presented in chapter 6 are produced by Dr. Apichart Linhananta and Dr. Steven Plotkin.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
Acknowledgments	x
1 Introduction	1
2 Biological Motivations	14
2.1 Reverse Proteolysis	14
2.2 Formation of Rod-like Protein Aggregates (Parkinson's Disease) . .	16
2.3 High Stability of the Crystallin Proteins in the Lens of the Eye . .	17
2.4 Assembly of Cell Division Protein FtsZ into One-monomer Thick Ribbons	18
2.5 The Effect of Macromolecular Crowding on Signal Transduction and Metabolite Channeling	20
2.6 Enhancement of Thermal Stability of Rabbit Muscle Creatine Kinase	21
2.7 Molecular Crowding Effect on an Alzheimer's β -amyloid Peptide	22
2.8 Effect of Macromolecular Crowding Agents on HIV Type 1 Capsid Protein Assembly	23

2.9	Molecular Crowding Creates an Essential Environment for the Formation of Stable G-quadruplexes in Long Double-stranded DNA .	25
3	Previous Approaches and Contributions	27
3.1	The Density Functional Model	27
3.2	The Gaussian Chain and Hard Sphere Model	31
3.2.1	Fundamental Measure Theory (FMT)	36
3.3	The Radial Distribution Function Model	37
3.4	The Chemical Potential Model	42
4	Thermodynamics of a Polymer	47
4.1	Measures and Statistics of Polymer Size	47
4.1.1	Self-Avoiding Random Walk	48
4.1.2	Off-lattice SAWs from Discontinuous Molecular Dynamics Simulations	51
4.2	Thermodynamics of a Polymer	58
4.2.1	The Effect of Osmolytes on the Thermodynamics of a Polymer	61
5	The Effect of Osmolytes on Protein Folding	69
5.1	Interactive Osmolytes	74
6	Discontinuous Molecular Dynamics Simulations	79
6.1	Free Energy and Entropy Functional	82
6.2	The Correspondence between Explicit and Implicit Osmolyte Models	83
6.3	Results	86
6.3.1	Simulations	86
6.3.2	A Hard-sphere Solvent Induces a Desolvation Barrier between Native Contacts, but Decreases Folding Cooperativity Relative to the Implicit Solvent Model	87
6.3.3	Effects of Osmolytes and Denaturants on Stability, Polymer Collapse, and Folding Cooperativity	89
6.3.4	Excluded Volume Stabilization Due to Hard Sphere Solvents	92

6.3.5	Calculation of the Free Energy, Enthalpy, and Entropy Changes for Osmolytes and Denaturants	92
7	Conclusion	98
	Bibliography	108
	Appendix: Discontinuous Molecular Dynamics	120

List of Tables

1.1	The volume of a few osmolytes in units of \AA^3	4
3.1	Parameter sets of the phase diagram.	30
3.2	The range of validity of equation (3.21).	35
4.1	Values of parameters A , θ and δ for (a) end to end probability distribution of an ideal random walk, (b) end to end probability distribution of an on-lattice SAW, (c) end to end probability distribution of an off-lattice SAW using the pivot algorithm, (d) end to end probability distribution of an off-lattice SAW using the naive growth algorithm, (e) probability distribution of the linear size of the polymer using DMD simulations (equation (4.5)), (f) volume probability distribution of the polymer using DMD simulations in equation (4.4) and (g) volume probability distribution of the polymer using equation (4.4) without the factor of v_0 to best fit the DMD simulations.	59

List of Figures

1.1	Three possible responses of organisms to osmotic dehydration. . .	2
4.1	A comparison of the end to end probability distribution of an on-lattice SAW.	51
4.2	On lattice saw for $N = 200$ using the theory and pivot algorithm. .	52
4.3	The end to end distance probability distribution of a continuum SAW.	54
4.4	The comparison between different ways of measuring the size of a chain.	55
4.5	Molecular volume (v_{mol}) versus principal box volume (v_{gb}).	56
4.6	The effective diameter probability distribution of a 100-mer. . . .	57
4.7	Volume probability distribution of a (a) 50-mer, (b) 100-mer and (c) 200-mer.	65
4.8	Volume probability distribution of a 50-mer, and the fits.	66
4.9	A schematic representation of the semi-grand canonical scheme. .	66
4.10	Cumulative probability.	67
4.11	Free energy of a 50-mer with and without osmolytes present. . . .	68
5.1	The largest number of contacts per residue of a protein.	71
5.2	Free energy of a coarse-grained protein with 50 amino acids. . . .	73
5.3	Free energy of a coarse-grained protein with 50 amino acids in the presence of osmolytes.	74
5.4	The stability of a protein with 50 residues as a function of ϵ_{po} in units of T	77
5.5	The heat capacity of a coarse-grained protein with 50 amino acids in a box of crowding agents with packing fraction 30%.	78

6.1	A: Ground-state native structure of the all-atom $G\bar{o}$ model of the Trp-cage protein; B: Denatured (unfolded) Trp-cage in urea solution; C: Folded Trp-cage in osmolyte solution.	87
6.2	(a) Reduced Heat capacity ($C^* = C/k_B$) versus reduced temperature T^* of Trp-cage. (b) Probability distributions of energy for protein-protein plus protein-solvent interactions.	88
6.3	(A) C^* versus T^* of Trp-cage. (B) Solvent quality temperature (T^*) versus solvent quality (ϵ_{ps}^*) phase diagram.	90
6.4	(A) Radius of gyration of the unfolded states. (B) Cooperativity of the folding transition.	95
6.5	C^* versus T^* of plots Trp-cage for several solvent models.	96
6.6	(a) Free energy versus protein foldedness (b) Protein internal energy $\langle E \rangle$ versus Q (c) Entropy (S) versus Q (d) Change in entropy (ΔS) versus Q (e) Changes in enthalpy between osmolyte and neutral solvent.	97

Acknowledgments

I feel most fortunate to have had the opportunity to study and broaden my knowledge in the University of British Columbia and to carry out my doctoral dissertation at the Department of Physics and Astronomy.

The writing of this dissertation would not have been possible without the help and encouragement of many people. It is a great pleasure to extend my sincere gratitude toward all of them.

First, I am highly indebted to my supervisor, Dr. Steven S. Plotkin, as this dissertation borrows from his help and effort. In particular I would like to thank Dr. Plotkin for motivating me to be an independent researcher.

I am deeply grateful to my PhD committee members, Dr. Andre Marziali, Dr. Ariel Zhitnitsky and Dr. Joerg Rottler, excellent teachers, whose support and advice I have relied on during my PhD program.

I have enjoyed my collaboration with Dr. Apichart Linhananta from Lakehead University and would like to acknowledge him for making the process of writing this dissertation a manageable task.

I appreciate the useful discussions I had with other members of Dr. Plotkin's group, Christopher Yearwood, Erik Abrahamsson, Ali Mohazzab, Alex Morriss-Andrews, Will Guest, Sebastian Ohse, Alvin Cheung, Gene Polovy, Philip Edgcumbe and Miguel Garcia, in our weekly group meetings.

I would like to thank Dr. Janis McKenna for her help and support, specially for editing my thesis and for the encouraging and motivating guidance during the discussions we had.

A special thanks goes to my husband, Kambiz Karbasi, for assisting me in typesetting, generating schematic plots and more importantly being a main source

of hope and encouragement. Without his help the completion of this dissertation would have been an unachievable goal.

At last but not least, I would like to express my heartfelt and everlasting appreciation to my parents and siblings for their endless love and support. I especially owe a great debt to my mother and father for sacrificing their own joy and comfort, and giving me moral support throughout all years of my study. Nothing would have ever been accomplished without their continuous love and encouragement. This dissertation is dedicated to my parents.

I would also like to thank NSERC for funding this project.

To my parents

Chapter 1

Introduction

Sit down before fact like a little child, and be prepared to give up every preconceived notion. Follow humbly wherever and to whatever abyss Nature leads, or you shall learn nothing. — Thomas Huxley (1860)

The main property of the intercellular medium is to sustain a liquid milieu for macromolecules to assemble and function properly. Water is the main constituent of this environment and lack of it, or dehydration, is the major habitat stress. There are two important factors that cause dehydration stresses: evaporation of internal water and osmosis into crowded aquatic environment. Osmotic dehydration happens mostly in saline waters, for example in the oceans which are the earth's largest habitat, or as a result of the diseases that produce high osmotic concentrations, for instance elevated blood glucose in diabetes, or in case of freezing.

When a mobile organism is threatened with dehydration, it will look for a new environment. But if the organism cannot move, it may choose one of the following responses (figure 1.1). (a) the endangered cell may shrink passively and consequently suffer from inhibited metabolic reactions and growth rates, and even death in case of severe shrinkage. (b) normally in a multicellular organism, each cell can sustain its own volume by accommodating organic particles. Organisms that use this treatment are called osmoconformers and range from single-celled archaea to some tissues of mammals. (c) also a multicellular organism called osmoregulator, can employ special organs such as kidney to compensate for changes in internal os-

otic pressure. Even in osmoregulators, the organ that is responsible for osmotic regulation should itself survive high osmotic pressure. Biologists used to think of mammals as osmoregulators, however it became evident later that their renal cells are exemplary osmoconformers. Osmoconformers are very important for survival, because even in extremely dehydrating conditions when mammalian osmoregulators fail to function properly, the cells in other organs such as brain and heart will use osmoconformers.

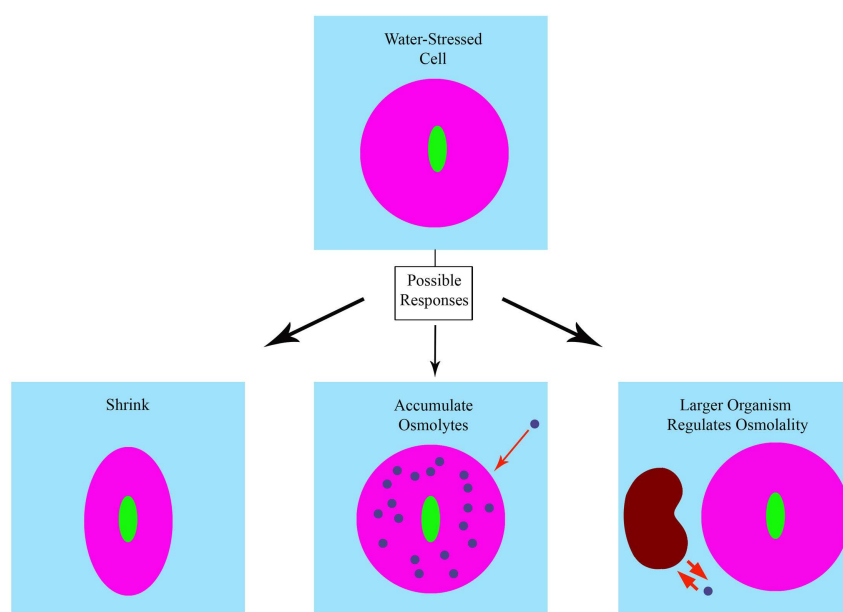


Figure 1.1: Three possible responses of organisms to osmotic dehydration.

The folding of a protein into its proper 3-D native structure, the assembly of complexes containing several macromolecules and macromolecular reactions are determined not only by the macromolecules' amino acid sequence, but also by the surrounding particles and solutes [8, 9]. There are a variety of small organic molecules that enable the cell to overcome various physical stresses by accumulating to high concentrations inside cells to protect it against stressing conditions while not perturbing the macromolecule's function [10]. These molecules that stabilize proteins from unfolding under high temperatures, desiccation, or chemical denaturants such as urea are called osmolytes [11].

Organic osmolytes fall into four major chemical categories:

- small carbohydrates including sugars (e.g. trehalose), polyols (glycerol, inositol, sorbitol, etc.) and derivatives (such as *o*-methyl-inositol)
- amino acids (glycine, proline, taurine, etc.) and derivatives (e.g. ectoine)
- methylamines (such as TMAO (Trimethylamine N-oxide) and glycine betaine) and methylsulfonium solutes including dimethylsulfonopropionate (DMSP)
- urea

We have calculated the molecular volume of a few osmolytes using mol-volume software¹ in Table 1.1. With an exception of urea which is found only in few types of animals' organs, other categories of osmolytes are widespread in nature. For example, glycine betaine occurs in every realm of life, and taurine is used vastly by marine animals and some mammalian organs. Carbohydrate osmolytes are found in archaea, fungi, algae, plants and mammalian kidneys, and possibly deep-sea invertebrates. Organs that tolerate or avoid freezing, such as terrestrial plants, insects, amphibians and some polar fishes usually use sugars and polyols. Moreover, there are many organs that use combinations of osmolyte categories, for instance, the mammalian kidney uses the polyols myo-inositol and sorbitol, the methylamines glycerophosphorylcholine (GPC) and glycine betaine, and the amino acid taurine (the organ also has high urea as both a waste product and an osmotic agent to concentrate the urine).

Biologists had noticed that when they isolate a biologically active substance such as an enzyme out of a cell, they have to add a high concentration of sucrose or glycerol to the system to maintain the stabilization of the activity. On the other hand, for isolated enzymes and proteins, biochemists have used high concentrations of salts, most commonly ammonium sulfate, to keep the proteins functional (basic solutes found in most cells have osmotic concentration of about 300 – 400

¹mol-volume software is a program for calculating the macromolecular volume by Alexander Balaeff. The Theoretical Biophysics Group, Beckman Institute, and The Board of Trustees of the University of Illinois

Table 1.1: The volume of a few osmolytes in units of \AA^3 .

osmolyte	volume
Urea	42.85
Betaine	108.80
Sorbitol	137.25
Taurine	84.78
Glycine	57.07
Glycerol	73.41
TMAO	67.12

milliosmoles per liter). This process is called “salting-out” [8]. In a similar manner, researchers have known that adding concentrated urea or guanidine hydrochloride (GuaHCl) to the medium leads to the denaturation of proteins (enzymes), that is the loss of proteins’ biochemical activity by an uncoiling of their native structure.

The effects of salts on protein’s activity and stabilization can be explained in terms of two opposing phenomena: (a) salting-in, that is at low salt concentrations the protein solubility increases because of the Debye-Huckel electric field screening² [12], and (b) salting-out at higher salt concentrations. In the latter process, the log of protein solubility depends linearly on the ionic strength with the slope K_s which is called the salting-out constant.

On the other hand, researchers have observed that denaturants such as urea and GuaHCl, cause a protein to unfold by the direct interaction (binding) to the protein molecules [13]. Tanford has done an extensive study on the thermodynamic mechanism of protein unfolding in the presence of urea or GuaHCl [14, 15]. He showed that the unfolding process can be explained in terms of the changes in the free energy of the native and denatured proteins when transferring from water to a denaturant solution. These changes include the free energy of interaction of the denaturants with protein groups that become exposed to solvent upon unfolding.

Finally, the process of protein stabilization by compounds such as sucrose and glycerol, has been the focus of interest in many fields. Scientists used to believe that these substances protect the protein from unfolding by forming a shell around

²The Debye-Huckel states that the ions in an electrolyte have a screening effect on the electric field from individual ions. The screening length is called the Debye length and varies as the inverse square root of the ionic strength.

the protein, however there was no experimental evidence for it. Around three decades ago, with studies of both sucrose [16] and glycerol [17] it became clear that in the process of protein stabilization by some crowding particles, the opposite happens. That is, in fact these compounds are excluded from the immediate zone of protein in an environment enriched in water.

These three discussed phenomena may appear unrelated at first, however; more detailed studies revealed that they share one basic fact. The solvent denaturation, native protein stabilization, and precipitation or salting-out process, all happen in the presence of high concentrations of the added compound. We can conclude that the dominant interactions in these processes can neither be strong nor specific, and in fact they should be categorized into the realm of weakly interacting systems. Systematic studies of interactions between soluble native proteins and molecules that stabilize their structure, have shown that the three processes can be formulated into the three-component thermodynamic theory, with water being the third constituent [18–20]. What happens in all these processes is that the structure-stabilizing, precipitating, and self-assembly-inducing agents cause the protein to be preferentially hydrated by being preferentially excluded from the protein domain [21–23]. Denaturants such as urea act in the opposite way, that is they are preferentially bound to proteins [24–26]. Both protein stabilizers and denaturants belong to a single class of molecules that affect the stabilization of proteins and ranges from strong enhancers of the protein's native structure to strong protein unfolders. In the same manner, the way an agent affects solubility depends only on the balance between the affinities of the protein for water and the added agent. In the former case, the protein has strong preference for water and in the latter case, for cosolvent. Therefore, in this kind of approach these various processes follow unified underlying principles and can be brought under one umbrella by using preferential interactions. These principles have been vastly employed by nature to protect life from freezing or osmotic shock. Amazingly, a small number of osmolyte types suffice to supply such important purposes that certain microorganisms, frogs, and desert plants have used to protect themselves from extinction.

Thermodynamic principles that govern all of these phenomena are based on three related parameters and their changes during the course of a reaction. These parameters are:

1. The transfer free energy of the protein, $\Delta\mu_{2,tr}$, which is the change in the free energy of the protein when transferred from pure water to the cosolvent solution.

$$\Delta\mu_{2,tr} = \mu_2(cosolvent) - \mu_2(water) \quad (1.1)$$

where $\mu_i = \mu_i^0(P, T) + RT \ln a_i$ is the chemical potential of the i th component with a_i being the activity of component i . Moreover, $a_i = m_i \gamma_i$, where m_i is the molal concentration and γ_i is the activity coefficient of component i . The index 2 refers to protein in the Scatchard notation of solution components³, and water and cosolvent are indicated as components 1 and 3 respectively [27].

2. The preferential interaction parameter, $(\partial\mu_2/\partial m_3)_{T,P,m_2} = (\partial\mu_3/\partial m_2)_{T,P,m_3}$, which is the changes in the transfer free energy with cosolvent concentration, and provides a measure of the thermodynamic interactions. Thus,

$$\Delta\mu_{2,tr} = \left(\int_0^{m_3} \frac{\partial\mu_2}{\partial m_3} \right)_{T,P,m_2} dm_3 \quad (1.2)$$

3. The preferential binding parameter, $(\partial m_3/\partial m_2)_{T,P,\mu_3}$. This is a measure of the amount of osmolytes that needs to be added to (or removed from) the solvent to restore thermodynamic equilibrium after we add the protein to the system.

$$\left(\frac{\partial m_3}{\partial m_2} \right)_{T,P,\mu_3} = - \frac{(\partial\mu_2/\partial m_3)_{T,P,m_2}}{(\partial\mu_3/\partial m_3)_{T,P,m_2}} \quad (1.3)$$

There are two main categories of osmolytes. In the first category, the preferential exclusion is caused by factors that are not dependent on the chemical nature of protein surface. That is, there is no interaction between osmolytes and protein and the protein only presents a surface to osmolytes. In the second category, on the other hand, the preferential exclusion is brought forth by factors that are sensitive to the chemical nature of the protein surface. In this case, the osmolytes can distinguish particular chemical features of the protein surface that they are interacting with, either by attraction or repulsion. The first category is ruled by two principal constituents, steric exclusion and perturbation of the surface free energy of water by osmolytes. Steric exclusion which was first introduced by Kauzmann [13], is

³Scatchard notation is named after George Scatchard, an American chemist.

the result of the difference in size of the osmolyte and water molecules. When a large osmolyte molecule is in contact with the protein, a shell is formed around the protein which cannot be further penetrated by the osmolytes, but in principle can be occupied by the smaller (water) molecules. This results in excess water around the protein surface which is thermodynamically interpreted as preferential hydration [28, 29]. Moreover, the protein is separated from the solvent by an interface, and changes to protein conformation cause perturbation in the surface tension or the surface free energy of water. Gibbs has shown that changes in the surface tension must cause changes in the concentration of osmolytes in the protein-solvent interface [30]. The Gibbs Adsorption Isotherm is:

$$\left(\frac{\partial m_3}{\partial m_2}\right)_{T,P,\mu_3}^{(\sigma)} = -\frac{S_2}{RT} \left(\frac{\partial \sigma}{\partial \ln a_3}\right)_{T,P,m_2} \quad (1.4)$$

where S_2 is the surface of the protein and σ is the surface tension. It can be pointed out from equation (1.4) that the increase in the surface tension of water by an osmolyte leads to its depletion from the surface layer. Sugars, nonhydrophobic amino acids, and most salts belong to the class of osmolytes that increase the surface tension of water and therefore their concentration in the interface is less than that in the bulk solvent. This is translated to preferential hydration, which means that the chemical potential perturbation is positive. The surface tension at the transition temperature at any cosolvent concentration, $\sigma_{T_m}^{m_3}$, is given by:

$$\sigma_{T_m}^{m_3} = \sigma_{T_m}^0 + \left(\frac{d\sigma}{dm_3}\right) m_3 + \left(\frac{d\sigma}{dT}\right) \Delta T_m \quad (1.5)$$

where $\sigma_{T_m}^0$ is the surface tension at the transition temperature in water and ΔT_m is the difference between the transition temperatures of a solvent containing osmolytes with concentration m_3 and in pure water. It is known that the surface tension of water decreases as temperature increases. The experimental studies show that at transition temperature, T_m , the changes in surface tension caused by the presence of osmolytes are compensated by the changes that are caused by the temperature increase from transition temperature in water to transition temperature in a solution with sugar. In this case σ_{T_m} remains constant which means that when the interaction between the protein and the osmolyte is dominated by the preferential exclusion

resulting from nonspecific solvent properties, transition occurs at a constant surface free energy.

In the second category of preferential exclusion, the chemical nature of the protein surface plays a major role in determining the interactions. In this case, the exclusion is dominantly driven by the solvophobic effect, which causes preferential hydration in case of glycerol [17] and polyols [31]. The structure of the polyols allows them to form the proper hydrogen bonds that fortify water interactions [32]. As a result, the nonpolar residues of the proteins have less tendency to form contacts with the polyol solution in comparison to that with water and therefore, polyol molecules are pushed away from the surface of the protein. An interesting agent in the second category is MPD (2-methyl-2,4-pentanediol which is the most popular chemical additive used for crystallization of biological macromolecules), which is strongly involved in electrostatic interactions [33] and moves away from the densely charged protein surface. Consequently, MPD causes high value of the preferential hydration and therefore is considered a good precipitant. This explains the high efficiency of MPD in protein-crystallizing⁴ [34].

This analysis of preferential hydration enables us to explain the reasons for a class of preferentially excluded osmolytes to act as stabilizers of the native proteins whereas others are destabilizers. In case of a totally nonspecific source of preferential exclusion, the unfolding process will be only affected by an increase in the interface which is the depletion zone, that results in an increase in the unfavorable free energy of interaction. This is the case for sucrose where for the native protein $\Delta\mu_{2,tr}$ increases linearly with sugar concentration. The stabilizing effects of glycerol and polyols can be explained in a more or less similar manner. In this case, the interaction between nonpolar residues of the protein and the agents is unfavorable and since the number of these unfavorable sites on the protein increases as it unfolds, the preferential hydration increases as well. This results in the stabilization of the native protein.

MPD however, has an opposite situation. Its interaction with charges is repul-

⁴Proteins, can be prompted to form crystals when placed in the appropriate conditions. In order to crystallize a protein, the purified protein undergoes slow precipitation from an aqueous solution. As a result, individual protein molecules align themselves in a repeating series of unit cells by adopting a consistent orientation. The crystalline lattice that forms is held together by noncovalent interactions.

sive while there is a strong attraction between nonpolar residues on proteins and MPD [35]. The surface charge density of the unfolded protein is less than that of the native protein, and therefore when the protein unfolds, MPD has a chance to interact favorably with newly exposed nonpolar residues, stabilize the unfolded structure. On the other hand, PEG (Polyethylene glycol) is both excluded non-specifically from native proteins [28, 36, 37], and is strongly nonpolar [38]. But since both MPD and PEG are strongly excluded from the native protein, they are excellent precipitants [36] and protein crystallizers [39, 40].

The non-specific steric repulsion, also known as excluded volume effect, is always present in cells. As R. John Ellis pointed out: “Crowding is similar to gravity, it cannot be avoided and organisms have to cope with its consequences”. Despite the fact that molecular crowding can in principle affect any biochemical process that is associated with major reduction in excluded volume, biochemists commonly investigate the properties of proteins in solutions with concentration of 110 g/liter or less, in which crowding effects are insignificant. The kinetic and thermodynamic effects of crowding are so important that many calculations of folding rates and stability done under uncrowded conditions of the test tube are different from those under crowded conditions in cellular environments by orders of magnitude. There are numerous experimental studies with biological applications that indicate the diversity and magnitude of the effect of osmolytes [41–47].

Experimental measurements of protein folding rates and equilibria are mostly done under idealized conditions where the effects of nonspecific interactions between the protein under study and other macromolecular particles present inside the cell are minimized. However; the ideal dilute solution is different from biological media in both the high concentration that a single macromolecule may have and in the variety of crowder species that exist in the solvent bath. For instance the concentration of RNA plus protein inside the cytoplasm of *E-coli* is 340 g/liter [48]. Biologically this medium is referred to as crowded. Moreover, there usually exist some structural macromolecules such as microtubules, intermediate filaments, membranous boundaries and F-actin that are soluble and are treated as background particle because they have no direct participation in biologically relevant reactions.

A general description of protecting and denaturing (urea) osmolytes has been proposed by Bolen [49], by developing Tanford’s transfer model [14] to account

for both protein side-chains and backbone. Essentially the same thermodynamic cycle approach has been applied to theories of protein stabilization due to macromolecular crowding agents [50, 51], which focus on the excluded volume (entropic) aspects of the transfer process. In fact, the physical origins of protein or polymer stabilization by steric osmolytes or crowders are essentially the same as those leading to phase separation in colloidal suspensions due to the addition of a non-adsorbing polymer for example [52–54]. We hope in this work to unify some of the concepts that have developed in parallel in the fields of osmolytes, crowding, and colloidal systems.

Bolen’s approach results in a solvent quality paradigm which classifies solvents as good or poor. In a poor solvent (solvophobic), protein intramolecular interactions dominate, which favors a compact folded native state that minimizes solvent exposed protein surface area. In a good solvent (solvophilic), protein-solvent interactions dominate, which favors an unfolded state that maximizes protein-solvent contacts. At the middle of the solvent quality scale is the neutral solvent that favors neither native nor unfolded states. Water is a poor solvent for proteins since the effective water-protein interactions lead to the hydrophobic effects, one of the major forces that fold proteins. Aqueous osmolytes and aqueous urea solution are poor and good solvents, respectively. The solvent quality paradigm has led to several molecular free energy transfer models, which are phenomenological models that utilize as input, experimentally measured change in free energy of proteins on transfer from pure water to aqueous osmolyte/urea solutions [55, 56]. A recent study combines protein conformations from simulation data of $G\bar{o}$ model with experimentally measured transfer free energy to infer the thermodynamic properties of proteins in solutions of osmolytes and urea [57]. The study predicts that solution of osmolyte and urea raises and lowers, respectively, the folding temperatures of the proteins. Since free energy transfer models utilize experimental data, the predicted enthalpy change due to osmolyte/urea should be accurate. However, the free energy transfer model does not take into account the size of solvents, and may lead to an incomplete description of the changes in protein conformational entropy due to solvents. Excluded volume due to solvents can reduce the amount of accessible protein conformations, changing an unstructured unfolded protein state to a more compact native-like unfolded state, as observed in several experiments [58–60]. In

addition, the binding of solvent to protein depends on the size of water, osmolytes and urea, and therefore the solvent size must be included in theoretical calculations and MD simulations to fully assess the enthalpy change due to solvents.

In chapter 2, we review a few biological applications of volume exclusion effects to indicate the diversity and magnitude of the effect of osmolytes. It is evident from the observations reported in this section that macromolecular crowding *in vivo* has effects on many different aspects of cellular function. The fact that biological macromolecules are evolutionary being adapted to function in such crowded environments, brings up a few biologically important questions: Why do cells have such a highly crowded interior? Is there any advantage in being crowded?

In chapter 3, we briefly review previous approaches to the problem of crowding. In particular we will discuss the contributions of Takada *et al.* [61, 62], Zhou [63], Bolen [56, 64, 65] and Minton [66] to settling different aspects of this problem. In the following chapters, we explain how our analytical method could contribute to further unraveling of polymer collapse and protein folding in dense systems. Takada *et al.* have shown that for a fixed packing fraction of osmolytes, small osmolytes have stronger stabilization effect than larger osmolytes, and as we will see our theoretical and simulation studies confirm this result. However, what is missing in their approaches is a model for the internal structure of the protein, that is, the protein as a whole is considered to be either in folded state (N) or in unfolded state (D), and the model is not flexible to the changes in the energetic parameters between residues. In our theoretical model by changing the energetic parameters of the primary, secondary and tertiary structure of the protein one can attain a more detailed description of a specific protein. Moreover, the model allows us to change the size of the protein residues. Another advantage of our theoretical work is that the protein can take any conformation that is consistent with the existing constraint in the system, and there is no need to be restricted to only two possible states of the protein (N and D).

The interaction parameter between osmolytes and the protein is taken to be zero in Takada's approach and therefore osmolytes are merely interacting with the protein via a hard-sphere potential, whereas in our model a general description of protecting and denaturing (urea) osmolytes has been provided by changing the interaction energy between the polymer and the osmolytes from negative to positive

values. Essentially the same thermodynamic approach can be applied to theories of protein stabilization due to macromolecular protecting crowding agents [50, 51], and protein destabilization due to the presence of denaturants. Therefore, our model enables us to unify the effects of protecting osmolytes and denaturants in the same theoretical scheme. One of the natural extensions of our theory is investigating the role of depletion interaction in aggregation resulted from protein-protein interactions and signal transduction. This effect has been partially studied in Takada's model by changing the interaction parameter between two unfolded proteins to model protein aggregation.

The advantage of our theoretical framework to Zhou's models is that again the protein has no internal structure in his approach, that is, the protein as a whole is considered to be either a globular object or a chain whereas our model has more details about the structure of the protein. Moreover, in Zhou's approach one cannot change the concentration of the protein in the system, and in our model we introduce a few characteristics of the protein (number of residues and the size of each residue) that the change in any of those results in the change in protein concentration. However; the result of Zhou's calculations which shows an increase in the folded state stability as the packing fraction of osmolytes increases is in qualitative agreement with our results.

The purpose of this work is to find a theoretical framework to investigate the effects of osmolytes on protein folding and stability. In the chapter 4, we introduce a new parameter that is more adequate as the size of a polymer chain, instead of the old notions of the linear dimension of a chain, the radius of gyration of the polymer, R_G , and the end to end distance of a polymeric chain, R_{ete} . In terms of this new parameter, we find an expression for the free energy of the polymer which enables us to investigate the thermodynamics and kinetics of the effect of osmolytes on its collapse. In chapter 5, we derive the free energy of a protein by introducing energetic parameters for primary, secondary and tertiary structure of the protein and using the same algorithm as in the case of polymers we study the effect of crowding on protein's folding and stability. Furthermore, we show that this effect is a function of the concentration and the size of the crowding agents as well as the type of the crowding particles that are present in the system. The latter case is accounted for by the parameter ϵ_{po} which is the interaction energy between

the protein and surrounding osmolytes.

In chapter 6, we discuss the discontinuous molecular dynamics (DMD) simulations of an all-atom $G\bar{o}$ model Trp-cage protein (PDB 1L2Y) immersed in explicit solvent molecules to investigate how osmolytes or urea stabilize or destabilize proteins [7]. The Trp-cage DMD model is immersed in spherical solvent molecules, in which binding of solvent to protein is controlled by a solvent-protein contact energy, ϵ_{ps}^* . Protein stabilization or destabilization in the solvent model used here arises from a change in solvent-protein interactions which implicitly accounts for the presence of osmolytes or denaturants.

Chapter 2

Biological Motivations

In a closed system and in the presence of obstacles a polymer chain will contract. Sasahara *et al.* [67], Tokuriki *et al.* [68], Baumgartner and Muthukumar [69], Honeycutt and Thirumalai [70], Dayantis *et al.* [71] have reported direct evidence of the compaction of the unfolded protein by crowding. At high concentrations of crowding macromolecules, the unoccupied interstitial spaces become too small to accommodate the globular native protein which is modeled as a hard sphere. On the other hand, the unfolded protein which can be modeled as a Gaussian chain is flexible in changing its conformation and therefore can be partially accommodated in narrow spaces. This causes different restrictions on the folded and unfolded protein by crowding agents. Therefore, the native protein is forced to maintain its native structure and remain functional. This phenomenon has vast biological applications.

In this section, we review a few biological applications of volume exclusion effects to indicate the diversity and magnitude of the effect of osmolytes. The goal is to encourage researchers to bear in mind the crowding effects when modeling the protein functions inside living cells.

2.1 Reverse Proteolysis

Proteolysis is the digestion of proteins by specific enzymes called proteases and it happens inside living cells for different purposes, such as digesting of proteins in

foods to produce amino acids, removing the signal peptides after their transport and removing the N-terminal methionine residues after translation. Proteases used to be thought of as critical elements in progression of catabolic reactions *in vivo*, but it is shown [41] that in the presence of crowding agents, subunits that build the protein are converted to the native protein. This shift of the peptide bond equilibrium is due to the fact that although the energetic barrier of ionization of the carboxyl is lowered by osmolytes, but the entropic barrier is higher since the reacting ends are brought closer to each other and the net difference in free energy increases which means that volume exclusion effect plays the role of a driving force in reverse proteolysis.

Somalinga and Roy have studied the (V8) protease-catalyzed syntheses of three 20-residue peptides [41]. The organic co-solvents that are usually used in protein and osmolyte mixture experiments are dextran and PEG (polyethylene glycol). Dextran is a complex polysaccharide composed of several glucose molecules. Dextran can be synthesized in different lengths (from 10 to 150 kilodaltons) and usually is used as an antithrombotic to reduce blood viscosity. PEG, on the other hand is an oligomer or polymer of ethylene oxide in liquid or low-melting solid state depending on its molecular weight and is usually considered to have a molecular mass below 20000 g/mol. PEG is soluble in water and is coupled to hydrophobic molecules to lower the surface tension of water and allow easier spreading.

The synthesis of these three proteins has been done *in vitro* with different concentrations of osmolytes and the results have confirmed the hypothesis of reverse-proteolytic condensation of complementary fragments. In their paper, Somalinga and Roy have shown that the formation of LIAA (20-residue designer peptides with the consensus sequence DIAQALKQIAEALQKIAGGY) increases with increase in dextran concentration which verifies the shift of the bias from bond hydrolysis to synthesis by adding osmolytes [41].

2.2 Formation of Rod-like Protein Aggregates (Parkinson's Disease)

There are a few factors through which macromolecular crowding can affect protein aggregation. Destabilizing the individual-molecule states and stabilizing the compound states, decreasing protein solubility in water and as a result inducing protein self-interaction by changing water activity, decreasing diffusion rate and therefore increasing the kinetics of aggregation by increasing viscosity, are among these factors. Aggregation appears after the formation of partially folded intermediates rich in beta structure which may have hydrophobic side-chains on their surface and therefore can bring about non-polar interactions between molecules, resulting in aggregation. Initially there is a lag time, which corresponds to the formation of the nucleus and then the growth rate of the aggregation increases exponentially. Both of these steps are affected by adding osmolytes into the system.

Munishkina *et al.* have investigated the aggregation of α -synuclein to explain some of the effects of molecular crowding on proteins [42]. α -synuclein is a 14 kDa protein, primarily found in neural tissue but also there are traces of it in glial cells. Purified α -synuclein is normally an unstructured soluble protein, but under crowded circumstances can aggregate to form insoluble fibrils, known as Lewy bodies which are responsible for degenerative diseases such as Parkinson's and dementia. In addition, an α -synuclein fragment, known as the non-Abeta component (NAC), is found in amyloid plaques in Alzheimer's disease.

The result of Munishkina's experiment indicates that high concentration of dextran/PEG increases the rate of α -synuclein fibrillation. In [42], they have shown that the addition of more and more PEGs results in larger acceleration of the rate of fibril formation. Their results also indicated that high enough concentrations of PEG 3350 have almost similar accelerating effects on α -synuclein fibril formation, which can be explained in terms of viscosity. When the system reaches a critical density, viscosity that works in the opposition of volume exclusion becomes important.

2.3 High Stability of the Crystallin Proteins in the Lens of the Eye

The cytoplasm of vertebrate eye lens cells consists of about 40% crystallin proteins, α -, β - and γ -crystallins, whose molecular weight differ for various peptide compositions. A degree of short-range order between these proteins is needed for the eye lens to be transparent because if they scatter light randomly, the lens would appear as an opaque object. Benedek showed that the observed transparency of the system can be explained by a limited degree of short-range order which means that it is not necessary to have a crystalline or paracrystalline state [72]. The opacity of the eye lens, such as in cataract, can also be explained by scattering of light by larger molecules which are the result of aggregation of crystallins, because the size of these aggregations is comparable to the wavelength of the light.

Having the highest molecular weight and concentration in the cytoplasm, α -crystallin is the main protein of the eye lens and its structural function is to sustain a suitable refractive index of the lens. The solubility of this crystallin is high (up to 250 mg/ml) and its physiochemical changes are important in senile cataract formation. Another function of α -crystallin in the eye lens is to act as a molecular chaperone to prevent the aggregation of intermediates by interacting with them in the early stages of their unfolding and capturing them inside its hollow interior. To function as a chaperone properly, α -crystallin needs to remain in its native state, because due to its complex assembly, the chaperone properties of refolded α -crystallin is different from those of its native structure [44]. α -crystallin can become unfolded easily and there are several factors that cause the unfolding of this protein, such as heat, pH of the solution and the presence of denaturants such as GdmCL. On the other hand, there are different ways to avoid denaturation of α -crystallin through crowding. Due to high concentration in the cytoplasm in the eye lens fiber cells, α -crystallin owes its stability partially to friction. Moreover, as a result of volume exclusion, macromolecular crowding is responsible for α -crystallin stability in high protein concentration, which is what naturally happens in the eye.

Andries *et al.* have investigated the effect of high protein concentrations on the scattered intensity of the lens [43]. At protein volume fraction of 0.035 the

diffusive behavior of the α -crystallin solutions undergoes drastic changes. They measured the scattered light in photon correlation spectroscopy. Photon correlation spectroscopy technique is normally used to study the behavior of complex fluids such as concentrated polymer solutions. The light scatters in all directions when it hits small particles (Rayleigh scattering). For a monochromatic and coherent light source such as a laser beam, the scattering intensity will have time-dependent fluctuations. These fluctuations caused by the Brownian motion of the small molecules in solutions are used to evaluate the changes in the distance between the scatterers in the solution. The information about the time scale of movement of the scatterers can be derived from the constructive or destructive interference of the scattered light by the surrounding particles.

In their paper, Andries *et al.* have measured the second order cumulant¹ as well as the correlation function of the eye lens at different protein volume fraction of ϕ [43]. They have shown that the normalized second order cumulant deviates drastically from 0 as they increase ϕ and moreover the correlation function has slowly decaying components. The cumulant analysis can be used to calculate an effective diffusion coefficient, D_{eff} . The osmolyte concentration of α -crystallin from bovine lenses at different angles and at a lower ($\omega = 0.08$) and a higher ($\omega = 0.32$) ionic strength was calculated. At higher ω the diffusion coefficient decreases for α -crystallin from bovine lens cortex and for calf cortex there is a small increase followed by a decrease.

2.4 Assembly of Cell Division Protein FtsZ into One-monomer Thick Ribbons

The FtsZ² is a protein that exists in most of the prokaryotic organisms and has a crucial role in microbial and organelle divisions by constructing a ring around the division site [73, 74]. The biological activities of FtsZ are the ability to efficiently

¹Second order cumulant which provides an indication of the variance of the system, is the coefficient of the second term in the expansion of the auto-correlation function in terms of time.

²FtsZ is named after “filamenting temperature-sensitive mutant Z”. The hypothesis was that due to the inability of the daughter cells to separate from one another, cell division mutants of *E. coli* would grow as filaments.

polymerize into filaments *in vitro* in the presence of potassium and GTP, to readily hydrolyze GTP to GDP and to involve in a reversible magnesium-linked assembly to form linear oligomers in the presence of GDP.

Gonzalez *et al.* have studied the effect of osmolytes on FtsZ assembly in the presence of the most abundant form of the nucleotide inside the cell, GTP [46]. They have found that the effect of high packing fractions of molecular crowdings which resemble the crowded *E. coli* cytoplasm is the formation of FtsZ polymer ribbons. The GTPase activity of FtsZ which is the polymer assembly/disassembly dynamics and nucleotide exchange within the polymer is significantly retarded in crowded environments in comparison to the same parameters of the FtsZ filaments that are formed in dilute solutions. Therefore they conclude that the self-organization process of FtsZ spontaneous arrangement into ribbons only happens in the bacterial interior and suggest that in the non-dividing cells the regulation of Z-ring assembly is modulated by other mechanisms.

They have analyzed the effect of inert macromolecular crowders such as Ficoll and dextran on the polymerization of FtsZ. In the presence of potassium ions in KGA buffer (25 mM Hepes/acetate, pH 7.4, plus 100 mM potassium glutamate, 300 mM potassium acetate and 5 mM magnesium acetate) by adding GTP the FtsZ filaments are formed. In the presence of high concentrations of osmolytes the FtsZ polymers form. Addition of 200 g/liter of Ficoll causes FtsZ polymers with width of 40-100 nm to form. Since the size of a single FtsZ monomer is 4-5 nm, these polymers correspond to a width of 8-25 protofilaments. The addition of 200 g/liter of dextran gives similar results. When they use incubation in KGA buffer with 0.1 nM of GMPCPP³ instead of GTP the polymers have been still observed. However if the incubation time was more than one hour, the polymers disappeared unless GTP-regenerating system regenerated GTP. This suggests that the polymers are dynamic structures.

Moreover; the atomic force microscopy (AFM) images confirm that the width of the FtsZ polymers that are induced by osmolytes is 40-100 nm and their thick-

³guanylyl-(α , β)-methylene-diphosphonate or GMPCPP hydrolyzes more slowly than GTP. The polymerization rate with GMPCPP is very similar to that of GTP, however, in contrast to polymerization with GTP, polymers formed with GMPCPP do not depolymerize rapidly after isothermal dilution.

ness is 3-4 nm. Considering that this size is compatible with 2-D structures with thickness of one protofilament, they concluded that these polymers are ribbons. These ribbons can be observed by optical microscopes when fluorescently FtsZ is used.

2.5 The Effect of Macromolecular Crowding on Signal Transduction and Metabolite Channeling

The effect of high concentrations of macromolecular crowding on equilibrium phenomena such as protein binding to DNA is a well-established problem [75]. However; the mechanism on the effect of osmolytes on nonequilibrium phenomena such as signal transduction and metabolic fluxes (the rate at which a metabolite is produced during a biological process) requires more study and attention. In biology, signal transduction is referred to any process in a cell which results in conversion of one kind of signal or stimulus into another. These processes normally involve ordered sequences of biochemical reactions inside the cell performed by enzymes and activated by second messengers that generate a signal transduction pathway. The signal transduction processes usually last only for a few milliseconds in the case of ion flux, or minutes for the activation of protein and lipid-mediated kinase cascades.

The metabolic chemical reactions are organized into metabolic pathways in which a principal chemical is converted into another chemical through a series of steps catalyzed by a sequence of enzymes. Enzymes are essential to metabolism because they assist organisms to have desirable reactions that take energy and do not occur by themselves, by adjoining them to spontaneous reactions that release energy. Moreover, enzymes are in charge of regulation of metabolic pathways against the changes in the cell's environment or signals from other cells. In some metabolic pathways, the product of an enzyme converts into the next enzyme in the pathway without equilibrating with the bulk solution. This direct enzyme-enzyme interaction is known as metabolite channeling.

Rohwer *et al.* have investigated the effect of different enzyme concentrations on the flux through the bacterial phosphotransferase system (PTS), the major car-

bohydrate transport system in bacteria, *in vitro* [45]. They have measured the PTS-mediated carbohydrate phosphorylation at different dilutions of *E. coli* extract. Their results show that the flux J is proportional to c^α where c is the protein concentration and $1 < \alpha < 2$.

Rohwer *et al.* have shown that at lower protein concentrations the addition of 9% PEG 6000 stimulates the PTS flux and inhibits the flux at higher protein concentrations. At lower PEG 6000 concentrations, the stimulation to inhibition of the PTS flux transition point occurs at higher protein concentrations. This suggests that the presence of osmolytes cause a decrease in the dissociation rate constants of enzyme complexes. Moreover, the addition of PEG 35000 inhibits the PTS flux in all conditions. These results indicate that in the crowded environment of the cell, on the time-scale of their turnover the PTS enzyme complexes live longer.

2.6 Enhancement of Thermal Stability of Rabbit Muscle Creatine Kinase

Jiang *et al.* have investigated the effect of osmolytes on the native conformation and thermal stability of creatine kinase (CK) using the far-ultraviolet circular dichroism spectra to reflect the secondary structure of the native state [76]. CK also known as creatine phosphokinase (CPK) or phospho-creatine kinase is an enzyme belonging to guanidino kinase family which is expressed by various tissues and cell types. This enzyme catalyzes the reversible transfer of a phosphoryl group from $ATP - Mg^{2+}$ to creatine which produces phosphocreatine and $ADP - Mg^{2+}$. CK is important in regeneration of ATP inside the cells that consume ATP rapidly, such as skeletal muscle, brain, photoreceptor cells of the retina, hair cells of the inner ear, spermatozoa and smooth muscle. Moreover, CK is an important diagnostic indicator of nervous system diseases, the heart muscle diseases and etc. In clinics, CK is assayed in blood tests as a marker of heart attack, severe muscle breakdown, muscular dystrophy, and in acute renal failure.

In their work, Jiang *et al.* have used the purified rabbit muscle CK which is an oligomeric two-domain protein and investigated the changes to its conformation and stability over an extended range of dextran 70 concentration. They

calculated the fraction of unfolded protein, f_u , at different temperatures from CD spectra results. By making the assumption that each spectrum at any given dextran concentration can be described by a weighted linear combination of only the native state and the unfolded state components, the dextran concentration dependence of ellipticity was analyzed in a two-state fashion. The observed ellipticity is taken as the average of the native state ellipticity and the fully unfolded ellipticity. In this two-state model the protein is assumed to be in its native state at $T = 25^\circ\text{C}$ and at $T = 80^\circ\text{C}$ it is fully unfolded at all dextran concentrations. In thermally induced unfolding, the behavior of transition temperature, $T_{1/2}$, shifts to higher temperatures as osmolyte concentration increases. As they increased dextran concentration from 0 g/l to 260 g/l, $T_{1/2}$ was raised by up to 8°C .

2.7 Molecular Crowding Effect on an Alzheimer's β -amyloid Peptide

Alzheimer's disease is a result of the fibril formation by the Alzheimer's β -amyloid ($\text{A}\beta$) peptide in brain tissue. Amyloids are insoluble proteinaceous fibrillar assemblies which are formed by aggregation of misfolded states of normally soluble proteins [77, 78]. The structure of amyloid fibrils has been demonstrated by X-ray fiber diffraction studies of a wide range of fibril types which clarified that these amyloids have no 3-D structural homology to their native state. The cleavage of a large amyloid precursor protein, APP, produces the $\text{A}\beta_{1-42}$ peptide which is present in unaffected individuals and has a normal physiological role. In Alzheimer's disease however, this peptide forms ordered aggregates that are deposited extracellularly as amyloid plaques or senile plaques in the neuropil and in vascular deposits. To design effective therapeutic agents against Alzheimer's disease, we have to understand the conformational properties and the mechanisms triggering aggregation of the $\text{A}\beta$ peptides at an atomic level.

Because amyloid fibrils are insoluble it is difficult to study their molecular structure, mechanisms of conformational transition, and the formation of the fibril aggregate. However, recently techniques such as NMR, electron paramagnetic resonance, and electron microscopy have revealed that $\text{A}\beta_{1-42}$ undergoes a confor-

mational transition from coil to α -helix to β -strand during amyloidogenesis [79].

From investigations of the refolding of reduced hen lysozyme in the presence of different crowding agents (CA), we know that the presence of osmolytes affects the aggregation of refolding protein molecules. Based on these consideration, Li *et al.* have constructed a computational model that consists of an atomistic description of a peptide, inert macromolecules of about 70 KD molecular weight to model osmolytes and a continuum solvent model [80]. They have applied this model to monomeric Alzheimer's β -amyloid peptide segment ($A\beta_{10-35}$).

Li *et al.* have started the simulations with one completely extended structure, one β -strand structure, and four NMR structures both in dilute and crowded (with CA packing fractions of $\phi = 30\%$ and $\phi = 40\%$) solutions. For two of the NMR structures, an additional simulations with CA packing fractions of $\phi = 35\%$ were done as well. In all simulations $A\beta_{10-35}$ adopted a collapsed coil conformation. For dilute solutions the results were in reasonable qualitative agreement with experimental and other simulation results. However; the results of the simulations in crowded environment showed some distinct changes in properties of the $A\beta_{10-35}$. For example, the appearance of transient β -structure increases and diffusivity decreases with increasing CA concentration. Moreover, the internal properties of $A\beta_{10-35}$ such as order parameter or atomic root mean square (RMSD) fluctuations are less sensitive to the changes in CA concentrations.

2.8 Effect of Macromolecular Crowding Agents on HIV Type 1 Capsid Protein Assembly

Human immunodeficiency virus (HIV) is a member of the retrovirus family, i.e. RNA viruses that are replicated in a host cell and produce DNA from its RNA genome via the reverse transcriptase. HIV causes acquired immunodeficiency syndrome (AIDS), which is the failure of the proper operation of the immune system in humans. The immature HIV particle has a spherical shell composed of a few thousand copies of the Gag polyprotein surrounded by the membrane. After the virus particle buds from the cell using the viral protease, Gag is cleaved and releases the matrix, capsid and nucleocapsid polypeptides (viral capsid proteins that

are associated with viral nucleic acid). This phenomenon initiates a structural rearrangement which is called maturation and is mostly involved nucleocapsid and viral RNA assembly into the mature capsid. Recent experimental studies reveal that the assembly of the cone-shaped mature capsid requires a *de novo* nucleocapsid assembly which is feasible at very high protein concentrations inside the extracellular virion.

in vitro assembly of capsid of HIV-1 strain BH10 were first explored in solutions with protein concentrations of 1 to 100 μM range at different pHs and ionic strengths without crowding agents [81]. Lanman *et al.* observed that after many hours of incubation no assembly is detected at physiologic ionic strength (150 mM NaCl) even though the protein concentrations were high. Assembly was still very slow at 1.75 M NaCl, however at 2.25 M NaCl, it was relatively fast. Moreover, the assembly rate depends on pH as well, that is at pH 7.8 it is a few times higher than at pH 7.0. As expected by increasing the concentration of capsid both the rate and the amount of polymer formed increased. The critical capsid concentration was 5.6 μM at ionic strength 2.25 M NaCl and pH 7.8.

Later the effects of Ficoll 70 and dextran 10 on the kinetics of capsid assembly were studied [82]. At first the capsid concentration was set to 15 μM and other conditions were highly advantageous for *in vitro* polymerization, i.e. pH 7.8 and ionic strength 2.25 M NaCl. The addition of Ficoll 70 accelerated capsid assembly significantly. Replacing Ficoll 70 with dextran 10 gives almost the same amount of enhancement in capsid assembly. In the presence of 100 g/liter Ficoll or dextran the critical capsid concentration was reduced from about 5.6 μM to 3.1 μM or 2.3 μM respectively.

The significant enhancement in the capsid assembly rate in the presence of osmolytes facilitates the possibility of attaining rapid and efficient polymerization *in vitro* under authentic conditions, i.e. high capsid concentration and low (physiologic) ionic strength. In the absence of osmolytes, at 150 mM NaCl and 600 μM capsid concentration, no polymerization was observed even after 65 hours of incubation. However; addition of 250 g/liter Ficoll 70 causes efficient capsid assembly in less than 2 min. In these conditions the cylindrical structures with 37 ± 9 nm diameter were being formed. The shape and dimensions of these structures were indistinguishable from those obtained in standard conditions, that is

high ionic strength and low capsid concentrations, in the absence of osmolytes. Conical structures with sizes close to those of authentic mature capsids of HIV-1 were also observed.

2.9 Molecular Crowding Creates an Essential Environment for the Formation of Stable G-quadruplexes in Long Double-stranded DNA

Zheng *et al.* have prepared long dsDNA from human genome that carry G-quadruplex-forming sequences with flanking duplex at both sides [83]. They then have investigated the effect of the molecular crowding on the formation of G-quadruplex during the process of *in vitro* transcription and heat denaturation/renaturation. Their data showed that osmolytes create an essential environment for stable G-quadruplex formation in dsDNA.

In their experiment, Zheng *et al.* constructed two dsDNAs carrying the core G-rich sequence from the C-MYC and NRAS gene, respectively using overlap polymerase chain reaction (PCR) and genomic DNA from HeLa cells as template. The G-rich core sequence was on the nontemplate strand and connected with a flanking promoter sequence (short DNA sequences that border a transcription unit) for the T7 RNA polymerase (an RNA polymerase that catalyzes the formation of RNA in the 5' → 3' direction) at its 5' side. The two dsDNAs were subjected to heat denaturation/renaturation or transcription with T7 RNA polymerase. The conditions under which the experiments were carried out were similar to authentic intracellular environment, i.e. at neutral pH in 150 mM K⁺ solution with (40% PEG 200) and without osmolytes. G-quadruplex formation was then detected by coupling the N7 of guanines to dimethyl sulfate (DMS). DMS can effect the base-specific cleavage of guanine in DNA and therefore it can be used to determine base sequencing, cleavage on the DNA chain, and other applications. Although the N7 in DNA duplex has a tendency to methylation by DMS and subsequently cleaved with piperidine, but the N7 in the G-quartet of G-quadruplex structure cannot be methylated. They labeled two dsDNAs at the 5'-end of the G-rich strand with a fluorescent dye.

Zheng *et al.* have marked the distinct bands corresponding to the cleavage of the four runs of guanines with a circle in the core G-rich sequence for the dsDNAs that were imposed to transcription or heat denaturation/renaturation. These bands become protected from cleavage when the treatment was done in the presence of 40% PEG. This reveals that in these DNAs G-quadruplex is consisted of three G-quartets that were formed during the process of RNA transcription and heat denaturation/renaturation. Interestingly, the guanines in the flanking sequences were always similarly attacked in the presence and absence of osmolytes, which indicates that these sequences were in the duplex form.

Other processes that are directly influenced by molecular crowding are:

- Roles of cytoplasmic osmolytes, water, and crowding in the response of *Escherichia coli* to osmotic stress [47].
- Osmolytes stabilize ribonuclease S by stabilizing its fragments S protein and S peptide to compact folding-competent states [60].
- Salt-induced stabilization of pair and many-body hydrophobic interactions [84].
- Stabilization of Cutinase by Trehalose [58].
- Stabilization of the ribosomal protein S6 by Trehalose is counterbalanced by the formation of a putative off-pathway species [59].
- Molecular crowding in the *Escherichia coli* periplasm maintains α -synuclein disorder [85].

Chapter 3

Previous Approaches and Contributions

In this section we briefly review previous approaches to the problem of crowding environments. Takada *et al.* [61, 62], Zhou [63], Bolen [56], [64, 65] and Minton [66] had significant contributions to settling different aspects of the effects of crowding agents on protein folding theoretically.

3.1 The Density Functional Model

In [61] the authors present an analytical framework to describe nonuniform system of proteins in crowded media, which is a system with infinite degrees of freedom. Density functional theories are particularly appropriate to describe interfacial phenomena such as phase separations of colloidal particles and polymer mixtures. Therefore, their theory provides a unified description of protein stability and aggregation in crowded environments. In this model the crowded system is coarse-grained, and proteins are represented by density fields. The native and denatured states of the protein are characterized by their intrinsic free energies, η_N and η_D , respectively and protein can transform from a native state to a denatured state and vice versa. In principle, the intrinsic free energies can be further broken into energetic and entropic parts, but in the following calculations the decomposition is not

necessary because the temperature is fixed. They set $\eta_C = 0$ in their calculations because the intrinsic free energy of the crowding agent does not play any role in their theory.

Further, they assume that the interaction ($u_{\alpha,\beta}$) between two molecules is the hard-core square-well potential:

$$u_{\alpha,\beta}(r) = \begin{cases} \infty, & r \leq R_\alpha + R_\beta \\ \epsilon_{\alpha,\beta}, & R_\alpha + R_\beta < r \leq c(R_\alpha + R_\beta) \\ 0, & c(R_\alpha + R_\beta) < r \end{cases} \quad (3.1)$$

where r is the intermolecular distance, R_α and R_β are the radii of molecular species α and $\beta (= N, D, C)$, and c is a constant factor greater than one and is set to 3 in this work. The excluded volume interaction is held by the uppermost line in the above equation and $\epsilon_{\alpha,\beta}$ represents other longer range interactions. Since in this calculation they were only concerned with the excluded volume interaction, they set $\epsilon_{\alpha,\beta} = 0$ except for $\epsilon_{D,D}$. Denatured proteins tend to attract each other because hydrophobic residues and isolated hydrogen bond donors and acceptors are exposed in unfolded states. Therefore, $\epsilon_{D,D}$ is usually set to a negative value. The length scale in the present theory is set to ~ 10 nm, which is taken as the unit length. The radii of the protein and crowding agent are of order of ~ 0.1 unit length.

The system is represented by $\phi^\alpha(r)$ where α specifies different species as D for denatured protein, N for native protein and C for crowding agents. The density field of solvent, $\phi^S(r)$, is defined as $\phi^S(r) = \rho_0 - \sum_\alpha \phi^\alpha(r)$, where ρ_0 is the total bulk density of the system. Then the free energy of the system can be written as:

$$F[\phi^\alpha(r)] = F_i + F_n \quad (3.2)$$

with F_i being the ideal part of the free energy and F_n the nonideal part. F_i is given by:

$$F_i = \int dr [\sum_\alpha \eta_\alpha \Phi^\alpha + T \sum_\alpha \Phi^\alpha \log \Phi^\alpha + T \Phi^S \log \Phi^S] \quad (3.3)$$

where Φ^S is the density field of the solvent and η_D and η_N are the intrinsic free energies of the protein in denatured and native states respectively. The nonideal part of the free energy is given by:

$$F_n = \frac{1}{2} \Sigma_{\alpha,\beta} \int \int dr_1 dr_2 \Phi^\alpha(r_1) \Phi^\alpha(r_2) T (1 - \exp(-u_{\alpha,\beta}/T)) \quad (3.4)$$

where $u_{\alpha,\beta}$ is the hard sphere potential between different species. Next, they define local chemical potential $\mu_\alpha(r)$:

$$\mu_\alpha(r) = \frac{\delta F}{\delta \Phi^\alpha(r)} \quad (3.5)$$

which in addition to the equilibrium condition $\mu_\alpha(r) = \mu_\alpha^0$ in which μ_α^0 is the equilibrium chemical potential, leads to a set of self-consistent equations of state:

$$\Phi^\alpha(r) = \frac{\rho_0 \exp[-(\eta_\alpha + W_\alpha(r) - \mu_\alpha^0)/T]}{1 + \Sigma_\beta \exp[-(\eta_\beta + W_\beta(r) - \mu_\beta^0)/T]} \quad (3.6)$$

where $W_\alpha(r)$ can be regarded as the potential of mean force for the species α . Solving this set of equations gives density fields $\Phi^\alpha(r)$ at equilibrium. The calculations of equilibrium states were done with various parameter values for $\rho_P = \phi^D + \phi^N$, $\rho_C = \phi^C$, $\varepsilon_{D,D}$ and different initial conditions. They found that there are two different phases, namely the uniform phase (U phase) and the phase with aggregates of the denatured proteins in it and the remaining region is mostly uniform (A_D phase). The lowest free energy state of the A_D phase has one spherical aggregate. They determine the phase boundary between the U and A_D phases by the relative spatial deviation of the native protein density field $\phi^N(r)$:

$$D_N = \frac{\sqrt{\langle (\phi^N - \langle \phi^N \rangle)^2 \rangle}}{\langle \phi^N \rangle} \quad (3.7)$$

When $D_N < 10^{-8}$ the system is in the U phase and in the A_D phase otherwise. Further, they introduce the bulk densities of protein, ρ_P , and crowding agents, ρ_C . Then the fraction of the native protein, $f_N = \phi^N / \rho_P$, determined the stability of the native protein.

To generate phase diagrams, they used $T = 1$, $\eta_D = 0$, $R_N = 0.4$, $\varepsilon_{\alpha,\beta} = 0$ except for $\varepsilon_{D,D}$ and ρ_C is between 0 to 0.8. Other parameters are shown in table

3.1.

Table 3.1: Parameter sets of the phase diagram.

Plane	ρ_P	η_N	$\varepsilon_{D,D}$	R_C	R_D
$\rho_C - \varepsilon_{D,D}$	0.1	0		0.4	0.6
$\rho_C - \rho_P$		0	-0.058	0.4	0.6
$\rho_C - \eta_N$	0.1		-0.058	0.4	0.6
$\rho_C - R_C$	0.1	0	-0.058		0.6
$\rho_C - R_D$	0.1	0	-0.058	0.4	

Moreover, the bulk volume fraction of the protein is between 0.0268 and 0.0905 when $R_N = 0.4$, $R_D = 0.6$ and $\rho_P = 0.1$ and the bulk volume fraction of the crowding agents is 0.215 when $R_C = 0.4$ and $\rho_C = 0.8$. Therefore the total volume fraction of molecules which is between 0.2 and 0.3 corresponds approximately to that of the living cells.

The phase diagram on the $\rho_C - \varepsilon_{D,D}$ plane indicates that aggregation increases as ρ_C becomes larger and also as long as the system is in the U phase, the native protein becomes more stabilized as ρ_C increases. We can derive some essential features of the crowded environments in the $\rho_C - \varepsilon_{D,D}$ phase diagram, however; experimentally it is not easy to manipulate the parameter $\varepsilon_{D,D}$ and therefore they tried other variables for phase diagrams. The $\rho_C - \rho_P$ phase diagram indicates that as ρ_C increases, the aggregation of the denatured protein and the stabilization of the native protein in the uniform phase are enhanced.

The intrinsic free energy η_N of the native protein can be experimentally manipulated by mutations. The $\rho_C - \eta_N$ phase diagram indicates the tendency of the crowding agents to enhance aggregation and stabilizing the native protein in the uniform phase. When $\eta_N > 1$ which means that the native protein is intrinsically highly destabilized, the phase boundary on the $\rho_C - \eta_N$ plane is almost vertical. This shows that when the native protein is unstable, the critical value of ρ_C for aggregation behaves almost independently of η_N .

Takada *et al.* have also investigated the effects of crowding environments with different sizes of the crowding agents and unfolded proteins. Experimentally, we can easily change the size of the crowding agent replacing the crowding agent

to a different kind. In this calculations, R_C was varied from 0.1 to 0.45. The $\rho_C - R_C$ phase diagram indicates that, for fixed ρ_C the larger crowding agents are better stabilizers, that is, the increase in R_C has similar effects to the increase in the number density ρ_C . This is expected, because at a constant number density ρ_C , a larger crowding agent takes more space. However, they defined $\tilde{\rho}_C = (4\pi/3)\rho_C R_C^3$ as the volume fraction of the crowding agent. The $\tilde{\rho}_C - R_C$ phase diagram shows that within the range of R_C that they studied, for a fixed $\tilde{\rho}_C$, smaller crowding agents have more significant effects on the stabilization of the native protein in the U phase and on the aggregation of the denatured proteins. At fixed $\tilde{\rho}_C$ the crowding agents with larger ρ_C reduce the native protein stability in the U phase and prevent aggregation. Therefore they concluded that the crowding effects on protein stabilization and aggregation are both R_C and $\tilde{\rho}_C$ dependent.

At the end, they studied the dependence of the crowding effects on the size of the denatured proteins, R_D . The $\rho_C - R_D$ phase diagram indicates that for large R_D , that is for $R_D > 0.5$ the increase in ρ_C enhances the aggregation. For $R_D > R_N = 0.4$, larger ρ_C values stabilize the native protein in the U phase as in previous cases. The crowding effects are more significant for larger R_D .

3.2 The Gaussian Chain and Hard Sphere Model

In his first paper on crowding effects, Zhou uses theoretical models that are intentionally simple [63]. These models are suggested to capture the essence of the essential effects of crowding, however they lack realistic details. In his work, Zhou treats an unfolded protein as a Gaussian chain and a folded protein as a hard sphere and assumes that these two states are separated from each other by a difference in free energy ΔG and then he studies the changes that the presence of osmolytes makes in ΔG .

The protein folding is the accumulation of native contacts. Let us consider the formation of a contact between two residues of the unfolded chain. The probability density for finding two monomers of an unfolded protein at a distance r from each other is:

$$P(r) = \left(\frac{3}{2\pi nb^2} \right)^{3/2} \exp\left(-\frac{3r^2}{2nb^2}\right) \quad (3.8)$$

where n is the number of monomers between the two residues and b is the bond length. The potential that gives rise to the above probability distribution is:

$$P(r) = \exp(-\beta U(r)) \quad (3.9)$$

If we assume that every close approach between monomers leads to productive contact formation then the problem of contact formation becomes equivalent to the problem of a Brownian particle moving in potential $U(r)$ and being absorbed at the contact distance $r = a$. The Kramers rate equation [86, 87] for a Gaussian chain results in:

$$k_{f0} = \frac{3\sqrt{\frac{6}{\pi}}Da}{(nb^2)^{3/2}} \quad (3.10)$$

where D is the relative diffusion constant and k_{f0} is the rate constant.

However, the formation of a native contact happens when the two residues are in their native states. If we assume that the transitions into and out of the native state are rate processes and the transitions of individual residues take place independently of other residues, then the rate constant of native contact formation is smaller than that of total contact formation by:

$$\frac{k_{f0}}{k_f} = \frac{\omega_{A-}\omega_{B-}}{\omega_{A+}\omega_{B+} [aD^{-1/2}(\omega_{A+} + \omega_{B+} + \omega_{A-} + \omega_{B-})^{1/2} + 1]} \quad (3.11)$$

where ω_- and ω_+ are residue (A or B) transitions rate into and out of the native state.

If the two proteins are held together by a strong short-range potential $U(r)$, the binding constant is given by [88]:

$$K_s = \int_a^{r^*} \exp[-\beta U(r)] 4\pi r^2 dr \quad (3.12)$$

where r^* is the upper limit that defines the bound state. The subscript 's' stands for

spherical geometry.

However; a protein complex is stereospecific which means that the complex forms only if specific translational and rotational constraints between two proteins are held. The relative displacement vector between two proteins is taken to be r and the rotational angles of the two proteins are summarized in Ω_A and Ω_B . Then the binding constant is given by:

$$K = \int_{\Gamma} \exp[-\beta U(\mathbf{r}, \Omega_A, \Omega_B)] d\mathbf{r} d^3\Omega_A d^3\Omega_B / (8\pi^2)^2 \quad (3.13)$$

where Γ represents the configurational space of the bound state.

Using [89]:

$$k_s = 4\pi D a \quad (3.14)$$

one can find the rate constant for the binding of two spherical particles at distance a at steady state. If we assume that initially there is a uniform distribution, then at time t the Smoluchowski rate constant at steady state becomes,

$$k_s(t) = 4\pi D a \left[1 + \frac{a}{(\pi D t)^{1/2}} \right] \quad (3.15)$$

In the presence of a potential, $U(r)$, the steady-state binding rate constant is given by [90]:

$$k_s = \frac{D}{\int_a^\infty \exp[\beta U(r)] (4\pi r^2)^{-1/2} dr} \quad (3.16)$$

In general, it is difficult to solve equation (3.16), however; for a long-range potential a simple approximation has been obtained [91]:

$$k = k_0 \langle \exp(-\beta U) \rangle^* \quad (3.17)$$

where k_0 is the rate constant in the absence of the potential and $\langle \rangle^*$ means average over the outer boundary of the bound state.

Shift of folding equilibrium by crowding

Crowding agents limit the motional freedom of a protein and because of the size difference, the denatured and native proteins are affected to different extents and therefore there is a shift in the folding equilibrium. For an unfolded (or Gaussian) chain, the probability density $G(\mathbf{x}, \mathbf{x}_0, n)$ that the chain with its origin at x_0 will end at x after n steps satisfies a diffusion equation [92, 93]:

$$\frac{\partial G(\mathbf{x}, \mathbf{x}_0, n)}{\partial n} = \frac{b^2}{6} \nabla^2 G(\mathbf{x}, \mathbf{x}_0, n) \quad (3.18)$$

When the chain length is large enough $N \gg 1$ the discrete variable n can be treated as continuous to a good approximation. n is replaced with time in the diffusion of a Brownian particle with a diffusion constant $D = b^2/6$. In fact, since the probability densities of both an unfolded chain and a Brownian particle are Gaussian (with $\langle r^2 \rangle$ proportional to the bond length in the former and the lapsed time in the latter), we can use them interchangeably. Moreover, since a physical chain cannot cross any obstacle such as a crowding macromolecule, an absorbing boundary condition should be used, i.e. $G(\mathbf{x}, \mathbf{x}_0, n) = 0$ at the positions that are taken by crowding agents. On the other hand, the crowding macromolecules eliminate conformations that were available to the unfolded chain in the absence of these agents. The fraction f_u of chain conformations that do not cross any boundaries is given by:

$$f_u = \int d\mathbf{x}^3 \int d\mathbf{x}_0^3 G(\mathbf{x}, \mathbf{x}_0, N) / V \quad (3.19)$$

where the integration is over the whole volume V of the solution. This problem is equivalent to the problem of a Brownian particle in a field of static absorbing traps with $\int d\mathbf{x}^3 G(\mathbf{x}, \mathbf{x}_0, N)$ being the survival probability of the Brownian particle. The additional integration over x_0 in f_u is equivalent to averaging over positions of traps.

They assume that the traps are spheres with radius a_c and concentration c and therefore they can use the Smoluchowski results. For short to moderate times, the survival probability $S(t)$ of a Brownian particle is given by (equation (3.15)):

$$-\ln S(t) = c \int_0^t dt' k_s(t') = 4\pi D a_c t [1 + 2a_c(\pi D t)]^{-1/2} \quad (3.20)$$

Equation (3.20) for extremely small values of $S(t)$ is accurate [94, 95]. However, at around $S(t) = \exp(-6.68/\phi^{1/2})$ where $\phi = 4\pi a_c^3 c/3$ is the volume fraction of traps, this theory starts to fail, e.g. at $\phi = 0.25$ equation (3.20) gives $-\ln S(t) = 13.4$.

By mapping the problem of a Brownian particle in the presence of traps to the problem of a Gaussian chain in a crowded environment, we find:

$$-\ln f_u = 3\phi y^2 \left(1 + \frac{2}{y\pi^{1/2}}\right) \quad (3.21)$$

where $y = R_g/a_c$. Table 3.2 shows the range of validity of equation (3.21) for different values of y . It is seen from this table that equation (3.21) is reliable under biological conditions.

Table 3.2: The range of validity of equation (3.21).

y	ϕ
1	no limit
2	0.50
3	0.32
4	0.23

On the other hand, because of the presence of crowding agents many positions of the protein will not be allowed. In physical chemistry, the inverse of the fraction of the folded conformations that do not overlap with the crowding molecules, f_f , is called the activity coefficient of the folded protein. In a model that both the folded protein and the crowding agents are hard spheres, the scaled-particle theory predicts [96]:

$$\begin{aligned}
-\ln f_f = & -\ln(1 - \phi) + (3z + 3z^2 + z^3) \frac{\phi}{1 - \phi} \\
& + \left(\frac{9}{2}z^2 + 3z^3\right) \left(\frac{\phi}{1 - \phi}\right)^2 + 3z^3 \left(\frac{\phi}{1 - \phi}\right)^3
\end{aligned} \tag{3.22}$$

where $z = d_f/2ac$ with d_f being the effective diameter of the folded protein. Zhou specifically considers the effect of crowding by ribonuclease A on the folding free energy of α -lactalbumin with parameters: $R_g = 30\text{\AA}$, $d_f/2 = 17.2\text{\AA}$, $a_c = 15.4\text{\AA}$ and $M_c = 17,000$ where M_c is the molecular weight of the crowding agents. The difference between $-\ln f_u$ and $-\ln f_f$ increases as c increases.

3.2.1 Fundamental Measure Theory (FMT)

In his latest work, Zhou does not make any assumption about the shape of the proteins [97], and therefore his new theory is based on the fundamental measure theory (FMT) [98–100], which is a density functional theory for fluids of convex hard particles. In the case of spherical particles, FMT reduces to SPT (scaled particle theory) [99]. FMT predicts that by placing a test protein in a sea of convex crowders, the increase in the chemical potential of the protein is [99]:

$$\Delta\mu = \Pi_c v_p + \gamma_c s_p + \kappa_c l_p - k_B T \ln(1 - \phi) \tag{3.23}$$

where k_B is Boltzmann's constant and T is the absolute temperature; v_p , s_p , and l_p are the volume, surface area, and linear size of the test protein; Π_c is the osmotic pressure of the crowders, γ_c and κ_c are the corresponding quantities for surface tension and bending rigidity; and ϕ is the total volume fraction of the crowders. If there are more than one crowder species in the system, then the crowder quantities will be expressed in terms of the weighted number densities of the different species:

$$c = \sum_{\alpha} c_{\alpha}; c_l = \sum_{\alpha} l_{\alpha} c_{\alpha}; c_s = \sum_{\alpha} s_{\alpha} c_{\alpha}; \phi = \sum_{\alpha} v_{\alpha} c_{\alpha} \tag{3.24}$$

where c_α is the number density of species α , and l_α , s_α , and v_α are their linear size, surface area, and volume, respectively. The results are:

$$\begin{aligned}\frac{\Pi_c}{k_B T} &= \frac{c}{1-\phi} + \frac{c_l c_s}{(1-\phi)^2} + \frac{c_s^3}{12\pi(1-\phi)^3} \\ \frac{\gamma_c}{k_B T} &= \frac{c_l}{1-\phi} + \frac{c_s^2}{8\pi(1-\phi)^2} \\ \frac{\kappa_c}{k_B T} &= \frac{c_s}{1-\phi}\end{aligned}\tag{3.25}$$

The FMT has been applied to convex particles with relatively simple geometric shapes [99]. Zhou generalizes the FMT to test proteins represented at the atomic level and refers to his method as the generalized fundamental measure theory (GFMT). Moreover, Zhou found that $\Delta\mu$ obtained in his previous simulation work [101] by the insertion procedure could be fitted to equation (3.23). The fitting did not produce a predictive method however, because it was not clear how v_p , s_p , and l_p could be calculated since an all-atom protein is not a convex particle.

Zhou showed that there is a good agreement between the GFMT predictions on $\Delta\mu$ and the data that obtained from simulations in Zhou's previous work [101].

3.3 The Radial Distribution Function Model

Bolen, on the other hand, has a different theoretical approach to the problem of the effects of crowding on protein folding [56, 64, 65]. He believes that the nonideal solution behavior has a structural origin. MacMillan and Mayer [102] and later Kirkwood and Buff [103] showed that one can express the thermodynamic properties of an isotropic solution as a function of the structure of the solution. The structure of a solution is reflected in the radial distribution functions $g_{\alpha,\beta}(r)$ between species α and β . One can interpret the radial distribution function as a measure of the deviation of particles of type- β from the random distribution around a central particle of type- α . The radial distribution function is a function of the distance from the central particle. Type- α and type- β particles can in principle be atoms,

or molecules such as proteins, water, or cosolutes like osmolytes.

When there is no correlation between particles type- α and type- β , $g_{\alpha,\beta}(r) = 1$. However, when there is an excess or deficit of type- β at a distance r from the central particle type- α , $g_{\alpha,\beta}$ deviates from unity positively or negatively, depending on the positive or negative correlation of α and β at r . They define the overall correlation $\mathcal{G}_{\alpha,\beta}$ as the excess or deficit of type- β particles around type- α in the whole volume that is occupied by these particles. One can obtain $\mathcal{G}_{\alpha,\beta}$ as a function of the packing fraction using the Kirkwood-Buff integrals defined as:

$$\mathcal{G} = 4\pi \int_0^\infty [g_{\alpha,\beta}(r) - 1] r^2 dr \quad (3.26)$$

Now let us consider a solution with osmolytes. Kirkwood and Buff [103] have also developed the dependence of the osmolyte's chemical potential μ_{os} on the osmolyte concentration c_{os} :

$$\frac{1}{RT} \left(\frac{\partial \mu_{os}}{\partial c_{os}} \right)_{T,p} = \frac{1}{c_{os}} + \frac{\mathcal{G}_{WO} - \mathcal{G}_{OO}}{1 - (\mathcal{G}_{WO} - \mathcal{G}_{OO})c_{os}} \quad (3.27)$$

where W and O stand for water and osmolyte respectively. The solvation features strongly depend on the osmolyte concentration, and therefore \mathcal{G}_{WO} and \mathcal{G}_{OO} will also depend on osmolyte concentration, leading to a complicated concentration dependence of $(\mathcal{G}_{WO} - \mathcal{G}_{OO})$ in general. A comparison between equation (3.27) and experimental data confirms this statement.

Recently Rosgen *et al.* have built a statistical mechanical theory to study the nonideal behavior of solutions with different concentrations of solutes such as salts and osmolytes [104, 105]. The chemical potential μ_{os} of the osmolyte (to the first approximation) is given by Rosgen *et al.* [105]:

$$\mu_{os} = \mu_{os}^o + RT \ln \left(\frac{c_{os}}{1 - V_1 c_{os}} \right) \quad (3.28)$$

where the constant V_1 is the effective molar volume of the osmolyte and μ_{os}^o is the standard chemical potential. By taking the derivative of the chemical potential (equation (3.28)) with respect to osmolyte concentration c_{os} we find:

$$\frac{1}{RT} \left(\frac{\partial \mu_{os}}{\partial c_{os}} \right)_{T,p} = \frac{1}{c_{os}} + \frac{V_1}{1 - V_1 c_{os}} \quad (3.29)$$

By comparing (3.29) with equation (3.27) one finds that the apparent hydrated volume of the osmolyte V_1 to first order is equal to $(\mathcal{G}_{WO} - \mathcal{G}_{OO})$. This conclusion provides a simple first order interpretation of solution behavior for osmolytes.

Equation (3.29) is applicable to a wide range of osmolytes and V_1 is a constant in this first order expression of the chemical potential [105]. Consequently, the difference between osmolyte hydration and osmolyte self-correlation, $(\mathcal{G}_{WO} - \mathcal{G}_{OO})$, must not depend on the concentration as well. That is, although osmolyte hydration \mathcal{G}_{WO} and self-solvation \mathcal{G}_{OO} can individually be a function of concentration, nevertheless their concentration dependence should cancel out in $(\mathcal{G}_{WO} - \mathcal{G}_{OO})$ for solutions that obey equation (3.28). This means despite the fact that the individual hydration and solvation correlations between osmolytes are nontrivial, osmolyte molecules can in principle behave thermodynamically as independent particles.

As mentioned above most of the studied osmolytes follow this first order behavior. The rest of the osmolytes are well described by the second order approximation. However; urea has a trivial behavior, that is:

$$\frac{1}{RT} \left(\frac{\partial \mu_{os}}{\partial c_{os}} \right)_{T,p} \approx \frac{1}{c_{os}} \quad (3.30)$$

When we compare equation (3.30) with equation (3.27) we see that independent of the urea concentration, hydration and self-solvation of urea are almost equal, i.e. $\mathcal{G}_{WO} \approx \mathcal{G}_{OO}$. This ideal behavior is a special case of the first order behavior.

So far, we discussed that the Kirkwood-Buff theory along with Bolen's theory of solutions provide information about structural features of osmolyte solutions. Now, let us consider three-component solutions, including water, osmolytes and proteins. If we assume that the protein concentration is low, then we can write its chemical potential, μ_{prot} , as a function of the osmolyte concentration c_{os} as the following [106, 107]:

$$\frac{1}{RT} \left(\frac{\partial \mu_{prot}}{\partial c_{os}} \right)_{T,p} = \frac{\mathcal{G}_{PW} - \mathcal{G}_{PO}}{1 - (\mathcal{G}_{WO} - \mathcal{G}_{OO})c_{os}} \quad (3.31)$$

The denominator in equation (3.31) contains merely information about the structure of the bulk solution, i.e. the Kirkwood-Buff integrals for osmolyte self-solvation \mathcal{G}_{OO} and osmolyte hydration \mathcal{G}_{WO} (equation (3.27))

$$\frac{1}{1 - (\mathcal{G}_{WO} - \mathcal{G}_{OO})c_{os}} = \frac{1}{RT} \left(\frac{\partial \mu_{os}}{\partial c_{os}} \right)_{T,p} \quad (3.32)$$

However; the numerator contains the Kirkwood-Buff integrals for the protein hydration, \mathcal{G}_{PW} , and osmolyte solvation of the protein, \mathcal{G}_{PO} . In three-component solutions the differences between solute-macromolecule and water-macromolecule interactions have thermodynamic effects that depend on solute concentration and are characterized quantitatively by preferential interaction coefficients. The preferential interaction parameters Γ_{μ_i} with $i = 1, 2, 3$ (1 for water, 2 for protein and 3 for osmolyte) are defined as partial derivatives that specify the dependence of the molality of the smaller solute on the macromolecular molality at fixed temperature and pressure, that is:

$$-\Gamma_{\mu_3} = \left(\frac{\partial \mu_2}{\partial m_3} \right)_{T,P,m_2} \quad (3.33)$$

where m_2 is the moles of the protein per Kg of solvent. Moreover, the product of the difference $\mathcal{G}_{PW} - \mathcal{G}_{PO}$ and the osmolyte concentration c_{os} , gives the preferential interaction parameter $-\Gamma_{\mu_3} = c_{os}(\mathcal{G}_{PW} - \mathcal{G}_{PO})$ [108–110]. To determine whether a cosolute is stabilizing, one has to evaluate the protein's preference to have positive correlations with water or with osmolyte. On the other hand the preference of the protein determines the sign of the solvation expression $(\mathcal{G}_{PW} - \mathcal{G}_{PO})$ or equivalently, the sign of Γ_{μ_3} . The denominator in equation (3.31) is always positive, and it can only modulate (up or down) the sensitivity of the protein chemical potential with respect to c_{os} .

The inverse Kirkwood-Buff theory [111] allows for a numerical determination of the Kirkwood-Buff integrals $\mathcal{G}_{\alpha,\beta}$ from experimental data. Recent MD simulations on the preferential interaction parameters of RNaseA and RNaseT1 in aqueous urea and glycerol provide important information about the thermodynamics of the protein solvation [112]. However, one should bear in mind that these results were all under the limiting assumption of ideal solution conditions. This limitation

can be serious in some cases which can have a major impact on the solvation of proteins. In the case of the denaturant urea however, since it has an almost ideal behavior, the simulation results [112] are valid over a larger range of concentration.

However one can still derive general, system-independent concepts about the impact of the structure of nonideal solutions on protein stability. The protein stability can be determined by the Gibbs free energy of unfolding, $\Delta_N^D G = RT \ln K$, where K is the equilibrium constant. Moreover, $\Delta_N^D G = RT \ln K$ can be derived using the differences of chemical potentials of the native and the denatured state. By taking the difference between the native and the denatured state, Δ_N^D , in (3.31) one obtains the derivative, m , of $\Delta_N^D G$ with respect to osmolyte concentration:

$$-\left(\frac{\partial \ln K}{\partial c_{os}}\right)_{T,P} = \frac{m}{RT} = \frac{\Delta_N^D(\mathcal{G}_{PW} - \mathcal{G}_{PO})}{1 - (\mathcal{G}_{WO} - \mathcal{G}_{OO})c_{os}} \quad (3.34)$$

Experimental results show that the m -value of protein unfolding is constant and negative in sign for urea, and does not depend on the urea concentration [113–115]. Moreover, the m -values for protecting osmolytes are positive in sign, and constant [116, 117]. By combining (3.32) and (3.34), one can find the dependence of the solvation preference of the native state compared to that of the denatured state, $\Delta_N^D(\mathcal{G}_{PW} - \mathcal{G}_{PO})$, on the m -value:

$$\Delta_N^D(\mathcal{G}_{PW} - \mathcal{G}_{PO}) = \frac{m}{\left(\frac{\partial \mu_{os}}{\partial \ln c_{os}}\right)_{T,P}} \quad (3.35)$$

The solvation preference relative to that at 0M osmolyte is given by:

$$\frac{\Delta_N^D(\mathcal{G}_{PW} - \mathcal{G}_{PO})}{\Delta_N^D(\mathcal{G}_{PW} - \mathcal{G}_{PO})_{c_{os}=0}} = \frac{RT}{\left(\frac{\partial \mu_{os}}{\partial \ln c_{os}}\right)_{T,P}} \quad (3.36)$$

where $\left(\frac{\partial \mu_{os}}{\partial \ln c_{os}}\right)_{T,P,c_{os}=0} = RT$. In the above equation, the derivative is given by (3.29). Bolen compared the change in solvation preference of proteins for several osmolytes as a function of osmolyte concentration and concluded that in the case of stabilizing osmolytes, except glycine, the slope is exceedingly steep which means that the concentration dependence is larger in comparison to urea. Although in case of urea the solvation preferences does not vary much with the changes of

urea concentration, in case of protecting osmolytes, however, the protein solvation preferences change significantly as the osmolyte concentration is increased.

3.4 The Chemical Potential Model

In this section we discuss Minton's approach to the problem of crowding [66]. In this treatment, the native state of the protein is denoted by N , and the i th nonnative conformational state by D_i . The main characteristic of nonnative states is their radii of gyration, that is, each nonnative state belongs to a set of nonnative states with the same radius of gyration, denoted by $R_{G,i}$. The chemical potential of the native state is given by:

$$\mu_N = \mu_N^o + RT \ln c_N + RT \ln \gamma_N \quad (3.37)$$

where μ_N^o is the chemical potential of the native state at concentration of unity in ideal solution where there is no solute-solute interaction. c_N is the molar concentration of the native state, γ_N the thermodynamic activity coefficient of the native state in real solution, R the molar gas constant, and T the absolute temperature. To write a similar expression for nonnative states, we need to choose a reference nonnative state, namely D_0 . In this model, the reference nonnative state is the state which is most abundant in ideal solution and under the selected environmental conditions. The chemical potential of an arbitrary nonnative state D_i is then given by

$$\mu_i = \mu_D^o + RT \ln c_i + RT \ln \gamma_i \quad (3.38)$$

where μ_D^o is the chemical potential of the reference nonnative state at concentration of unity in ideal solution, and c_i and γ_i are the concentration and thermodynamic activity coefficient of state D_i in real solution respectively. The fact that we can assign a single standard state chemical potential to all D_i allows us to take the differences between the chemical potential of any two nonnative states at equal concentrations as the differences in the activity coefficients. For instance when $c_i = c_j = c$:

$$\Delta\mu_{ij}(c) \equiv \mu_j(c) - \mu_i(c) = RT \ln \frac{\gamma_j}{\gamma_i} \quad (3.39)$$

However, we know that at equilibrium, the chemical potential of all species should be equal, therefore from equation (3.38) we find that for any two species we have:

$$\frac{c_i}{c_0} = \frac{\gamma_0}{\gamma_i} \quad (3.40)$$

The fraction of state i among all denatured states at equilibrium is then:

$$f_i \equiv \frac{c_i}{\sum_j c_j} = \frac{\gamma_i^{-1}}{\sum_j \gamma_j^{-1}} \quad (3.41)$$

Moreover, since $\langle X \rangle = \sum_j f_j X_j$, where $\langle \rangle$ denotes the mean value and X is any state property, from (3.41) we find:

$$\langle \gamma_D \rangle = \frac{\Lambda}{\sum_j \gamma_j^{-1}} \quad (3.42)$$

where $\langle \gamma_D \rangle$ is the mean activity coefficient of all denatured states and Λ is the total number of denatured states.

At constant temperature, pressure, and solvent conditions the difference between the chemical potential of the two standard states N and D_0 should be constant and therefore a thermodynamic constant is defined as the following:

$$K^* \equiv \exp \left(-\frac{\mu_D^o - \mu_N^o}{RT} \right) = \frac{\gamma_i c_i}{\gamma_N c_N} \quad (3.43)$$

Let f_N be the fraction of protein in the native state, therefore the fraction of denatured protein, f_D , is then equal to $1 - f_N$. We can then define the equilibrium unfolding constant of the native state as:

$$K_{ND} = \frac{f_D}{f_N} = \frac{\sum_i c_i}{c_N} = K^* \Lambda \frac{\gamma_N}{\langle \gamma_D \rangle} \quad (3.44)$$

and also a root mean-square radius of gyration of the denatured state as:

$$R_G^{RMS} = \langle R_G^2 \rangle^{1/2} = (\sum_j f_j R_{G,j}^2)^{1/2} \quad (3.45)$$

For the rest of this work, the value of quantities in the dilute (ideal) solution are indicated with a superscript “o”. Since $\gamma_N^o = 1$ by definition, it follows from equations (3.42) and (3.44) that:

$$\frac{K_{ND}}{K_{ND}^o} = \gamma_N \frac{\sum_i \gamma_i^{-1}}{\sum_i (\gamma_i^o)^{-1}} \quad (3.46)$$

Minton treats a denatured protein as a random coil or freely jointed chain [118]. According to Flory and Fisk [119], the probability distribution of the radius of gyration, R_G , of an ideal chain is given approximately by:

$$P(R_G) = A \left(\frac{R_G^2}{\langle R_G^2 \rangle} \right)^3 \exp \left(-\frac{7}{2} \frac{R_G^2}{\langle R_G^2 \rangle} \right) \quad (3.47)$$

where $\langle R_G^2 \rangle$ is the mean-squared radius of gyration and A is a normalization constant. More recently, Lhuillier [120] proposed the following probability distribution of the radius of gyration of a nonself-intersecting polymer chain in three dimensions:

$$P(R_G) = P(R_G^*) \exp \left[-B \left(\frac{4r^{-15/4}}{5} + \frac{6r^{5/2}}{5} - 2 \right) \right] \quad (3.48)$$

where R_G^* is the radius of gyration that maximizes $P(R_G)$, $r \equiv R_G/R_G^*$ and B is the scaling parameter.

From equation (3.40) we have:

$$\gamma_i^o = \frac{c_0^o}{c_i^o} = \frac{P(R_{G,0})}{P(R_{G,i})} \quad (3.49)$$

because $\gamma_0^o = 1$. By combining equations (3.48) and (3.49) we get:

$$\ln \gamma_i^o = B \left(\frac{4r_i^{-15/4}}{5} + \frac{6r_i^{5/2}}{5} - 2 \right) \quad (3.50)$$

where $r_i \equiv R_{G,i}/R_{G,0}$. From equation (3.39) we have (the authors must have assumed that the concentration of state i is the same before and after transfer):

$$\gamma_i(\phi) = \gamma_i^o \exp\left(\frac{\Delta\mu_i(\phi)}{RT}\right) \quad (3.51)$$

where $\Delta\mu_i(\phi)$ is the transfer free energy of a polypeptide chain with a radius of gyration $R_{G,i}$ to move from a solution with no cosolute to a solution containing inert macromolecular cosolutes with volume fraction ϕ .

Minton then made estimates of the contribution of the excluded volume interaction to the transfer free energy, using two different models for the polypeptide chain. The first model is called the Gaussian cloud. In this model, the polypeptide chain is taken to be a spherically symmetric cloud of residues. At any given point inside this sphere with a distance r_p from the c.o.m. of a polymer chain, the average density of residues is given by the Gaussian function [118]:

$$\rho = A \exp(-B^2 r_p^2) \quad (3.52)$$

where $B^2 = 3/2R_G^2$, $A = n[3/2\pi R_G^2]^{3/2}$ and n is the total number of residues in the chain. For polymer chains in a good solvent the semiempirical function in equation (3.52) provides a reasonable description of the density of the chain. To obtain the contribution of the excluded volume interaction in the transfer free energy, Minton calculates the probability for a rigid cosolute to penetrate to a certain distance in the cloud without intersecting any residue, and then integrates over the whole volume of the cloud.

The second model is called the equivalent hard sphere model. Here, the polypeptide chain is taken to be an equivalent rigid sphere having the corresponding radius of gyration. The covolume of this rigid sphere and the hard particle cosolute are calculated in the conventional manner [121].

Minton calculated the properties of the denatured states for 30 states of four proteins with R_G values spaced logarithmically in the range between 0.4 and 1.5 times $\langle R_G^2 \rangle^{1/2}$. Calculations show that increasing the density of states and/or the range of R_G does not change the final results significantly. γ_i^0 were calculated using equations (3.49) and (3.50), γ_i were then calculated as a function of ϕ for each of the discussed models, $\langle \gamma_D \rangle$ was calculated using equation (3.42) with $\Lambda = 30$ and γ_N was calculated using equation (3.43).

Furthermore, the results indicated that negligence of steric exclusion between

non-adjacent residues results in a reduction of $\sim 50\%$ in the difference between the free energies of the unfolded and native states at all $\phi \neq 0$. In the absence of this long-range steric repulsion the equilibrium average conformation is more compact. When long-range intramolecular steric exclusion is neglected, at higher ϕ a significant fraction of the equilibrium ensemble of unfolded protein have R_G that is close to or even less than that of a sphere consisting of an equal number of close packed residues [122], which is not realistic.

Moreover, the calculations in both models discussed above indicate that high concentrations of hard sphere or a hard rod cosolutes substantially stabilizes the native state of a dilute protein in comparison to its denatured state. For a given ϕ of cosolute, the magnitude of the stabilization depends significantly on the size of the protein relative to the solute.

It should be emphasized that neither of these models provides a realistic picture of the excluded volume interaction between the polymer and a rigid cosolute for all accessible radii of gyration. However, the Gaussian cloud model is a good approximation of excluded volume interaction for large radii of gyration, that is when the protein is in nonnative states. Moreover, the equivalent hard sphere model provides a more realistic picture when the radius of gyration is small which means when protein is in the native state. Therefore the combination of these two models gives the limiting estimates of the magnitude of the excluded volume interaction.

Chapter 4

Thermodynamics of a Polymer

4.1 Measures and Statistics of Polymer Size

In this chapter I will investigate the statistics of an isolated polymer in a theoretical framework, which facilitates the study of the effects of osmolytes on polymer collapse.

Our focal point is to find an appropriate theoretical framework for describing the statistics of polymers.¹ The development of a theory detailing the thermodynamics of conformational ensembles of homopolymers has been the topic of research interests. As mentioned by Chan and Dill in their review, scaling of polymer size with polymer length provides a suitable probe of the nature of interactions between polymer and its environment [122–124]. In an extensive study Flory has shown that for a chain with length N the average radius of gyration, R_g , scales according to $R_g = lN^\nu$ [125, 126], where l and ν are determined by the characteristics of the solution and dimensionality. For instance in three dimensions when polymer is in a good solvent $\nu \approx 0.6$ and for a polymer in a bad solvent $\nu \approx 0.34$.

Another quantity that is usually used in the literature as the linear size of a chain is its end to end distance, which is the distance between the first and last monomers. However; neither R_g nor the end to end distance can precisely capture the right features of a chain. Although when the chain is crammed into a compact

¹A version of this chapter is in preparation for publication [6].

conformation the radius of gyration might be a suitable quantity to measure the linear size of the chain, but as the polymer expands to more stretched out conformations the R_g description fails to represent the right size of the chain. For instance in the limit of a completely expanded rod-like conformation R_g estimates $\frac{L}{\sqrt{12}}$ as the linear size of the chain instead of L which is the contour length of the chain. On the other hand, the end to end distance description provides reasonable measurement of the linear size of an expanded chain whereas in case of a compact chain many conformations lie outside the end to end sphere, which is a sphere with its center at the center of mass of the chain with radius R_{ete} . There are numerous conformations of the chain with $R_{ete} \approx 0$ whereas even in the most collapsed state a polymer in principle cannot get smaller than a certain size, usually given by $R_g \propto N^{1/3}$.

Furthermore, the likelihood of close approaches between parts of the polymer non-local in sequence depends on the effective polymer density or packing fraction. The density is given by the number of monomers divided by the volume of the polymer. The volume of the polymer may be estimated by a sphere with radius given by the end to end distance R_{ete} or the radius of gyration R_g , however many configurations will have substantial amounts of polymer outside of these spherical approximations. We thus seek improved measures of the volume of the polymer.

Accurate measures of polymer size are important because the volume occupied by a protein changes dramatically during folding or collapse. Such improved measures facilitate an accurate statistical mechanical description of polymer-osmolyte and protein-osmolyte mixtures. One of our intents in this work is to inspect through various possible quantities that represent the correct size of the polymer and hence provide a more reasonable description of the polymer statistics.

4.1.1 Self-Avoiding Random Walk

The end to end distance probability distribution, $P(R_{ete})$, of a freely jointed ideal chain converges to a Gaussian even for chains as small as 10 residues [125, 127]. However for polymer chains undergoing random walks in dimensions less than 4 the end to end distance probability distribution of a chain is no longer Gaussian when excluded volume effects are present [128–131]. The statistics of a polymer is then often approximated by that of a self avoiding random walk (SAW)

on a hypercubic lattice. The functional form of $P(R_{ete})$ for an excluded volume chain has been investigated using both Monte Carlo simulations [132] and analytical approaches [133, 134]. Here the results from Monte Carlo simulations and Lagrangian theory approaches are in excellent agreement for two and three-dimensional chains. A scaling law for the distribution $P(R_{ete})$ has been derived by des Cloizeaux for an on-lattice SAW [134]

$$P(N) = \frac{1}{\xi^d} C \rho^\theta e^{-A \rho^\delta} \quad (4.1)$$

where $\rho = \frac{r}{\xi}$ in which r is the end to end distance of the chain and $\xi^2 = \frac{\langle r^2 \rangle}{2d}$. The mean-squared end to end distance is given by $\langle r^2 \rangle = B^2 \ell^2 N^{2\nu}$ where ℓ is the bond length (taken to be unity), the exponent $\nu = 0.59$ and the prefactor $B = 1.2$ is a number that is important for quantifying return probabilities. We take this value from end to end distance data in Monte-Carlo simulations of off-lattice self-avoiding chains of various lengths [135]. C is a non-universal number of order unity.

For an ideal random walk, $\nu = 0.5$ and $B = 1.0$; values of A , θ , and δ are given in table 4.1 for comparison. The parameter A and exponents θ and δ are universal, that is they are the same for all on-lattice SAW's in a given number of dimensions regardless of chain length. The universal parameters for a 3-D SAW are summarized in table 4.1.

The configurations of real polymers are better described by off-lattice self-avoiding random walks rather than on-lattice models, which allow the angle between three consecutive monomers to have any value consistent with steric volume constraints. We have written a MATLAB² algorithm to generate off-lattice SAW conformations by the well-known pivot algorithm, which has been shown to deal effectively with the attrition problem for SAWs [136, 137]. Other generating algorithms which solve the attrition problem are also widely used [138, 139]. The attrition problem corresponds to the fact that for a polymer to avoid a sterically-excluded region, the appropriate boundary condition corresponds to that of an absorbing boundary. Any generated conformation which penetrates into the bound-

²The MathWorks, Natick, MA.

ary must then be removed from the statistics. Thus in generating conformations of a self-avoiding polymer, any conformation of the polymer that by chance wanders into the sterically excluded region corresponding to the previously generated monomer positions must be eliminated, and the walk must be either re-initiated or appropriately reweighted [140]. Walks that survive this process become exponentially rare as the chain length increases.

The pivot algorithm is implemented for off-lattice SAWs by first generating a viable N -step SAW as follows. The $i + 1$ th residue is placed a distance ℓ from the i th residue at random angle; if the distance $r_{i+1,j}$ for any $j < i$ is less than 2σ , where σ is the monomer radius taken here to be $\sigma = \ell/2$, the walk is canceled and restarted until a SAW of N steps is generated. New conformations are then obtained by performing random symmetry operations at random positions along the chain. These symmetry operations include rotations around the pivot point with arbitrary Euler angles. Because the pivot algorithm generates radically different conformations, after $\sim N^{0.19}$ moves a globally different conformation is achieved [137].

To see the effects of a self-avoiding walk versus a reflecting boundary condition on SAW statistics, we have also generated random walks using a naive growth algorithm that corresponds to a reflecting boundary condition as follows. Walks are generated as above, with the $i + 1$ th residue placed a distance ℓ from the i th residue at random angle; but now if the distance $r_{i+1,j}$ for any $j < i$ is less than 2σ , only the last step is canceled and a new step is attempted, until a walk of N steps is generated. The process is then repeated from the first step to generate a new conformation.

Using the above generating methods (for $N = 50, 100, 200, 500, 700, 900$), we have found that the end to end distribution for an off-lattice SAW has the same functional form as the on-lattice SAW in (4.1), however with different universal exponents. These are summarized in table 4.1.

Figure 4.1 compares the end to end probability distribution of an on-lattice SAW and an off-lattice SAW for a 100-mer. As a check, figure 4.2 shows that on-lattice SAWs that we generated using the pivot-algorithm recover exactly the same set of universal parameters for an on-lattice SAW as derived in previous studies [134].

The quantity A and exponents δ and θ are universal, that is they are indepen-

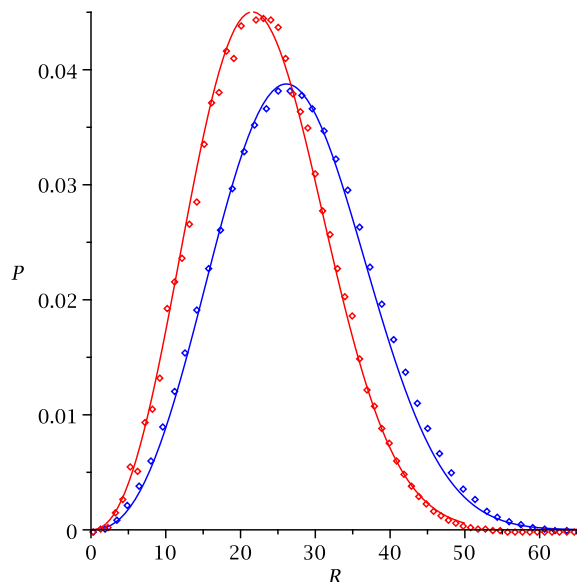


Figure 4.1: A comparison of the end to end probability distribution of an on-lattice SAW with $N = 200$ (red curve is obtained by using the des Cloizeaux functional form in equation (4.1) using the well-known exponents in table 4.1 and red circles are data from the on-lattice pivot algorithm) and an off-lattice SAW (blue curve is taken again from equation (4.1) using the exponents in table 4.1 and blue circles are data from the off-lattice pivot algorithm). This fitting is done for several values of N to give the exponents in table 4.1. R is in units of angstrom.

dent of chain length, however they differ for on- and off-lattice walks and are less universal than the Flory exponent ν , which is the same for both on- and off-lattice walks (0.59 ± 0.01).

4.1.2 Off-lattice SAWs from Discontinuous Molecular Dynamics Simulations

Discontinuous Molecular Dynamics (DMD) is an efficient method that has been used to study protein folding and *ab initio* protein structure prediction [141–146], protein aggregation [147, 148], and the effects of osmolyte-protein interactions [145]. A brief description of the DMD method for freely jointed polymer chains is pre-

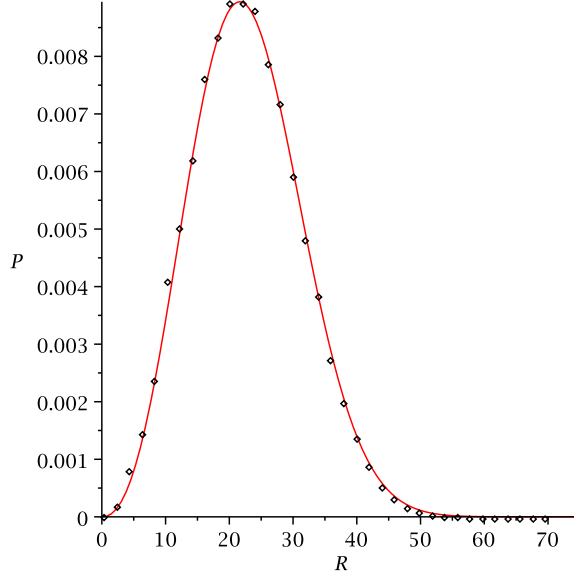


Figure 4.2: On lattice saw for $N = 200$ using the theory and pivot algorithm.

The parameters of the fitting curve are as the following: $A = 0.144$, $\theta = 2.269$, $\delta = 2.45$ and the value of these parameters in the literature are: $A = 0.144$, $\theta = 2.269$, $\delta = 2.43$. R is in units of angstrom.

sented here (more complete descriptions of DMD methods for polymers and proteins are contained in the above references and in the appendix). We used DMD simulations for the purpose of generating polymer configurations. In our model the polymer is a freely jointed chain of N beads or monomers and $N - 1$ joints, wherein each monomer is represented as a hard sphere. Two bonded beads i and j are constrained to be within 10% of an average distance, ℓ , by an infinite square-well potential:

$$u_{i,i+1}^{bond} = \begin{cases} \infty, & r \leq 0.9\ell \\ 0, & 0.9\ell < r \leq 1.1\ell \\ \infty, & 1.1\ell < r \end{cases} \quad (4.2)$$

where $\ell = 1$. Two non-bonded atoms i, j may interact by hard-sphere potential (purely repulsive) with a hard-core radius:

$$u_{ij}^{non-bond} = \begin{cases} \infty, & r \leq \sigma_{HS} \\ 0, & \sigma_{HS} < r \end{cases} \quad (4.3)$$

where σ_{HS} is the hard-sphere diameter of a monomer. In this work we set $\sigma_{HS} = \ell$.

Using these parameters we have performed DMD simulations on homopolymers of length 50, 100, and 200 with the above potentials, as a confirmation of SAW statistics generated by the pivot algorithm mentioned above.

The system is simulated at a finite temperature (in practice T was set to $10K$ here, however since there is no interaction energy-scale for a purely self-avoiding polymer, any finite temperature will generate the same equilibrium ensemble). Temperature equilibration is achieved by the ghost particle methods. Moves are generated by integrating Newton's equations and conserving momentum and energy for inter-particle collisions. Statistics such as the probability distribution of the end to end distance are calculated by sampling 1,000,000 conformations for each polymer chain length.

A plot of the end to end distribution for a walk of length $N = 200$, as generated by the pivot algorithm, naive growth algorithm, and DMD simulations, is shown in figure 4.3. As expected, the DMD simulations reproduce the same statistics as those of the pivot algorithm for a continuum SAW.

To determine an accurate measure of the volume occupied by a polymer, we considered four different ways to enclose a specific conformation of polymer. Figure 4.4 depicts the four different models: (a) The end to end distance model, which approximates polymer size by a sphere centered at the center of mass of the polymer, having radius R_{ete} ; (b) The embedding sphere model, which is a sphere with its center at the c.o.m. of the polymer and radius $R = |\vec{r}_f - \vec{r}_{com}|$ where \vec{r}_f is the position of the farthest monomer from c.o.m.; (c) The Cartesian box model, which is a box oriented in a fixed Cartesian frame of reference, with volume v given by $|x_{max} - x_{min}| \times |y_{max} - y_{min}| \times |z_{max} - z_{min}|$ where x_{min} and x_{max} are the

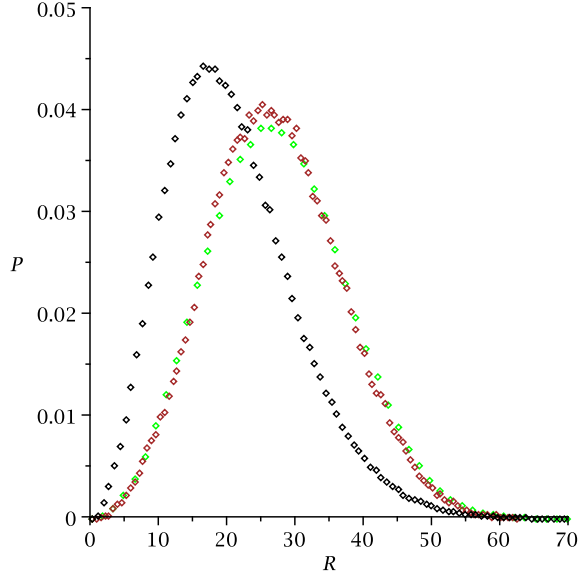


Figure 4.3: The end to end distance probability distribution of a continuum SAW with $N = 200$ using the naive growth algorithm (black diamonds), the pivot algorithm (green diamonds) and DMD simulations (brown diamonds). The naive chain growth algorithm with effective reflecting boundary condition when dead-ends are encountered does not provide the correct statistics of a SAW. The effective absorbing boundary condition of a self-avoiding walk shifts the distribution to larger values of end to end distance. The pivot algorithm and DMD simulations show essentially the same statistics. R is in units of angstrom.

x-components of the position vector of the monomers with smallest and largest components along the x-axis correspondingly (the same definition applies in y and z directions); (d) The principal box model, which determines polymer volume using a box aligned along the principal axes of the polymer in each conformation, having volume $v = |r_{1max} - r_{1min}| \times |r_{2max} - r_{2min}| \times |r_{3max} - r_{3min}|$ where r_{1min} and r_{1max} are the components of the position vector of the monomers with smallest and largest components along the first principal axis respectively (the same definition applies in second and third principal axis directions).

The principal box volume correlates with the steric volume of the polymer, but is always larger than the steric polymer volume (see figure 4.5). The steric

molecular volume may be calculated by measuring how many vertices of a dense cubic lattice happen to be within a region bounded by a surface determined by tracing a probe of given radius over all of the monomers constituting the polymer.

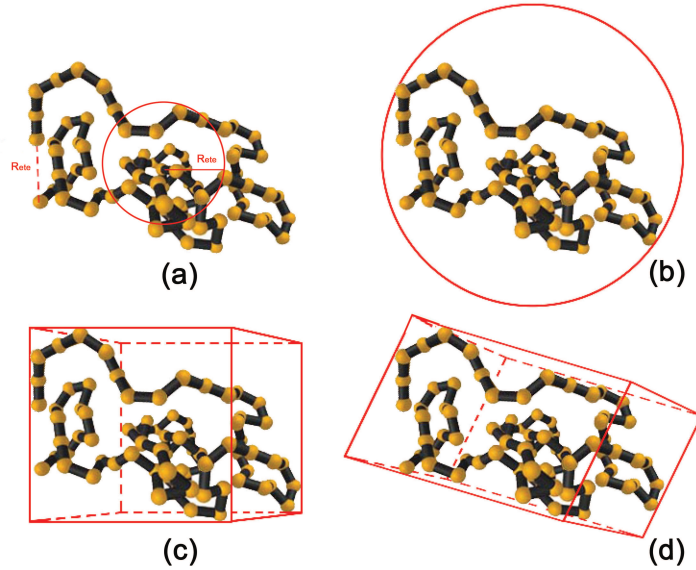


Figure 4.4: The comparison between different ways of measuring the size of a chain. (a) the end to end distance model, (b) the embedding sphere model, (c) the Cartesian box model and (d) the principal box model.

Figure 4.6 compares the effective diameter probability distributions using different models of the size of a 100-mer. Employing the letter notation above describing the four size measures, the effective diameter representing the linear size of the polymer for (a) is given by $d_{eff} = 2R_{eff}$, for (b) by $d_{eff} = 2R$ and for (c) and (d) by $d_{eff} = (\frac{6v}{\pi})^{1/3}$. We choose this last equation rather than $d_{eff} = v^{1/3}$ in order to compare diameters of effective spheres for all measures. To compare spherical measures for all four methods shown in figure 4.6, we use an effective sphere for methods (c) and (d) which has the same volume as the fixed reference frame or principal axes box respectively. However it is the volume rather than the linear size which enters into the analysis below.

For many conformations the end to end distance description does not accurately represent the size statistics of a polymer, since the end to end distance of a

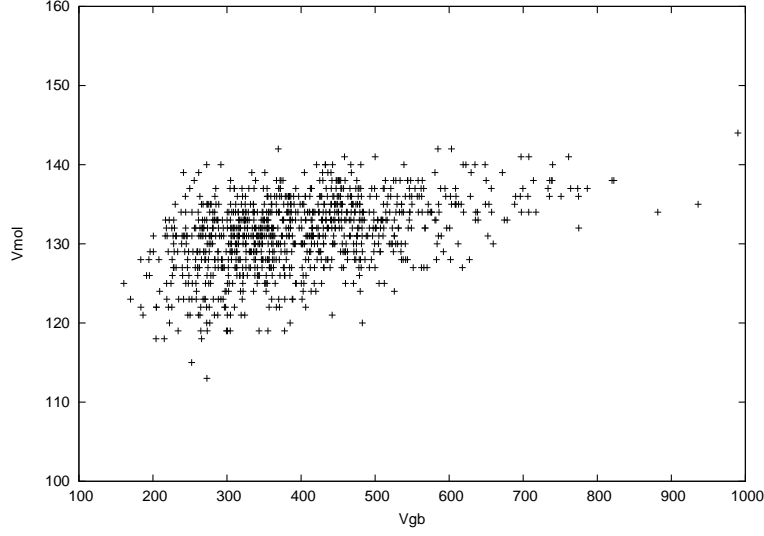


Figure 4.5: Molecular volume (v_{mol}) versus principal box volume (v_{gb}). The principal box volume correlates with the steric volume of the polymer, but is always larger than the steric polymer volume. The steric molecular volume may be calculated by measuring how many vertices of a dense cubic lattice happen to be within a region bounded by a surface determined by the tracing a probe of given radius over all of the monomers constituting the polymer. v_{gb} and v_{mol} are in units of \AA^3 .

polymer can be zero, whereas even in the most collapsed conformation the real size of a polymer cannot be smaller than $\approx \ell N^{1/3}$. The embedding sphere and Cartesian box models overestimate the size of the polymer, as can be seen in figure 4.6 and therefore cannot be considered as a good measure of the size of the polymer. Moreover, we have calculated the probability distribution of the effective diameter of the gyration tensor volume which is $d_{eff} = 2(\sqrt{e_1^2 + e_2^2 + e_3^2})$ where e_1 , e_2 and e_3 are the eigenvalues of the gyration tensor of the polymer. As it can be seen from figure 4.6 numerous conformations of the polymer have larger principal box diameter than the effective gyration tensor diameter, which means that the gyration tensor model underestimates the volume of the polymer. Therefore, for the rest of this work we use $d_{eff} = (\frac{6v_{pb}}{\pi})^{1/3}$ as the suitable linear size of the polymer, where v_{pb} is the volume of the principal box enclosing the polymer, and henceforth we replace v_{pb} with v .

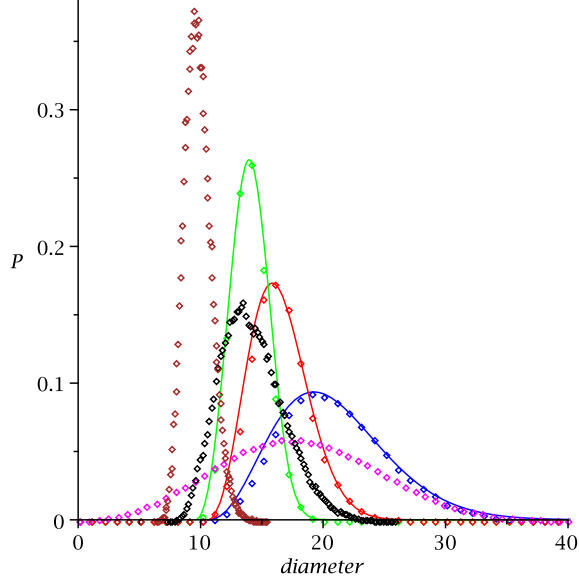


Figure 4.6: The effective diameter probability distribution of a 100-mer. Green: effective principal diameter probability distribution, Orange: effective random box diameter probability distribution, Blue: embedding sphere diameter probability distribution, Magenta: the end to end distance probability distribution using the pivot algorithm, Black: radius of gyration probability distribution and Brown: gyration tensor probability distribution. Diameter is in units of angstrom.

The effective volume containing the polymer is then the volume of a box aligned with its principal axes, that encloses all the monomers. We have performed DMD simulations for homopolymers with different lengths and calculated the volume of the polymer in each conformation. Figure 4.7 displays the volume probability distribution of a 50-mer, 100-mer and 200-mer. By curve fitting the DMD data, we have found that a universal expression that describes the volume probability distribution of homopolymers $P(v)$ is analogous to $P_{SAW}(R_{ete})$ except for a shift in the volume variable $v - v_0$:

$$P(v, N) = C \left(\frac{v - v_0}{\delta v} \right)^{\theta'} e^{-A' \left(\frac{v - v_0}{\delta v} \right)^{\delta'}} \quad (4.4)$$

The values of A' , θ' and δ' in equation (4.4) are summarized in table 4.1. The

quantity v_0 is the principal volume of the collapsed conformation of the polymer which is related to the number of monomers, N , and the bond length, ℓ , as $v_0 = c\ell^3 N^\alpha$ where $\alpha = 1.92 \pm 0.02$ and c is a constant of order unity. The quantity δv is given by $\sqrt{\langle v^2 \rangle} = c'\ell^3 N^\gamma$ with $\gamma = 2.0 \pm 0.04$.

Equation (4.4) includes a nonzero offset v_0 to account for the fact that the polymer volume can never be smaller than the fully collapsed volume. This term is more appropriate for a metric accounting for polymer volume than for end to end distance because the end to end distance can in principle be zero while the polymer volume cannot. The offset v_0 improves the fit to simulated data. Taking equation (4.4) without the factor of v_0 results in the fit to the data given in figure 4.8. The degrees of freedom in the equation increases from 134 (140 data points minus 6 degrees of freedom $A', \theta', \delta', \gamma$, and v_0 or c) to 135. This gives an F-value of 122, so that the probability of expression (4.4) without the factor of v_0 being correct is $\frac{1}{\text{F-value}} = 0.008$.

To compare with the previous distributions such as the end to end distance distribution, we define $r = (\frac{3v}{4\pi})^{1/3}$ to obtain the probability distribution of the linear size of the polymer as $P(r) = P(v(r)) |dv/dr|$, which is a function of the form:

$$P(r, N) = 4\pi r^2 C \left(\frac{r - r_0}{\zeta} \right)^\theta \exp \left(-A \left(\frac{r - r_0}{\zeta} \right)^\delta \right) \quad (4.5)$$

where r_0 is the linear size of the collapsed polymer and is given by $r_0 = c\ell N^\Delta$ with $\Delta = 0.64 \pm 0.02$ and $\zeta = \sqrt{\langle r^2 \rangle}$. The values of A , θ and δ are summarized in table 4.1.

4.2 Thermodynamics of a Polymer

The grand potential is the natural free energy for describing mixtures of polymer and osmolytes. We consider a box of fixed volume enclosing a protein and osmolytes system, and a subsystem with a permeable boundary to osmolytes, but impermeable to the monomers of the polymer, as shown schematically in figure 4.9.

Entropic considerations are very important in investigating the effects of the solvent on protein folding and polymer collapse. The folding dynamics involve a

conformational search, guided by energetic bias, of all allowed states. The conformational entropy as a function of the polymer volume v is given by

$$S(v, N) = S_0 - k_B \log(P(v, N)) \quad (4.6)$$

where S_0 is the total entropy of the polymer and k_B is Boltzmann's constant. Now consider a polymer in good solvent where the monomers are attracted to each other such that two non-bonded monomers in contact have contact energy $-\varepsilon$. Two monomers i and j are in contact if $|i - j| > 3$ and $|\vec{r}_i - \vec{r}_j| < r_c$ where r_c is an interaction cutoff distance. In a mean field approximation the number of contacts, n , of a N -mer with volume v is given by:

$$n(v) = zN\eta(v) \quad (4.7)$$

where the constant z is the coordination number in a maximally compact configuration, and $\eta(v)$ is the packing fraction of the chain defined as:

$$\eta(v) = Nv_m/v \quad (4.8)$$

where v_m is the volume taken up by each monomer, and v is the volume of the

Table 4.1: Values of parameters A , θ and δ for (a) end to end probability distribution of an ideal random walk, (b) end to end probability distribution of an on-lattice SAW, (c) end to end probability distribution of an off-lattice SAW using the pivot algorithm, (d) end to end probability distribution of an off-lattice SAW using the naive growth algorithm, (e) probability distribution of the linear size of the polymer using DMD simulations (equation (4.5)), (f) volume probability distribution of the polymer using DMD simulations in equation (4.4) and (g) volume probability distribution of the polymer using equation (4.4) without the factor of v_0 to best fit the DMD simulations.

	a	b	c	d	e	f	g
A	1.5	0.144	0.057 ± 0.002	0.46 ± 0.01	2.70 ± 0.01	4.00 ± 0.01	6.90 ± 0.03
θ	0.0	2.269	2.40 ± 0.01	2.1 ± 0.01	2.38 ± 0.02	2.50 ± 0.02	7.50 ± 0.02
δ	2.0	2.43	2.85 ± 0.05	1.7 ± 0.02	2.70 ± 0.01	1.35 ± 0.01	1.20 ± 0.03

principal box found previously. The interest in accurate representations of polymer volume is justified by the fact that the number of pairwise contacts in a polymer depends on the packing fraction η , which itself depends on the volume that the polymer takes up.

In the mean-field approximation, the internal energy of the polymer is given by:

$$E(v) = -\epsilon z N \eta(v) = -\epsilon z N^2 v_m / v \quad (4.9)$$

The free energy $F(v, T)$ of a homopolymer as a function of the polymer volume v at a given temperature is given by $E(v) - TS(v)$, where the entropy $S(v)$ is given in equation (4.6). In our model, the internal energy of the polymer $E(v)$ is given by the number of contacts times the energy per contact, thus:

$$F(v, T) = -\epsilon n(r < r_c, v) - TS(v). \quad (4.10)$$

Here the quantity $n(r < r_c, v)$ is the average number of residue pairs within a cutoff distance r_c , given the polymer takes up a volume v . A contact occurs when $|\mathbf{r}_i - \mathbf{r}_j| < r_c$.

In what follows, we find the mean number of contacts that would occur by chance for a polymer having volume v , $n(r < r_c, v)$. Because the number of contacts varies between conformations, we found that the best statistics were obtained by first finding the cumulative probability distribution for the number of contacts of all conformations that had a volume *less than* a cutoff volume v_c . The number of contacts for conformations having volume v_c was then found as

$$n(r_c, v) = \frac{N(N-3)}{2} \left. \frac{\partial}{\partial v_c} P(r < r_c, v < v_c) \right|_{v_c=v} \quad (4.11)$$

where $P(r < r_c, v < v_c)$ is the cumulative probability of having a contact within r_c for all conformations with $v < v_c$.

We enumerated the number of contacts of a polymer using DMD simulations. Figure 4.10 shows the cumulative probability for two non-bonded monomers with $|i - j| > 3$ to be within a distance r_c , i.e. $P(|\mathbf{r}_i - \mathbf{r}_j| < r_c)$. The curves were obtained by calculating the distance between all pairs separated by three or more bonded

monomers ($|i - j| > 3$) for all conformations with volume v less than the cutoff volume v_c and then normalizing to obtain the probability distribution.

Each curve in figure 4.10 corresponds to a given cutoff volume. An expression that best describes the DMD simulation data can be written in terms of complete and incomplete gamma functions:

$$P(r < r_c, v < v_c) = C \frac{\xi A^{\frac{-(1+\theta)}{\delta}}}{\delta} \left(\Gamma\left(\frac{1+\theta}{\delta}\right) - \Gamma\left(\frac{1+\theta}{\delta}, f(v) \left(\frac{r}{\xi}\right)^\delta \right) \right)$$

where $\xi = \sqrt{\langle r^2 \rangle}$ and the function $f(v)$ is:

$$f(v) = \frac{B}{\left(\frac{v}{\sqrt{\langle v^2 \rangle}}\right)^\gamma} + D \quad (4.12)$$

where constants γ , B and D are equal to 1.21, 1.15 and 4.78 respectively. Since $P(r < r_c, v < v_c)$ is the cumulative probability, the average number of contacts for conformations having volume v is then given by equation (4.11).

4.2.1 The Effect of Osmolytes on the Thermodynamics of a Polymer

In this section, we examine the statistical mechanics of mixtures of polymer and neutral osmolytes, which are represented as hard sphere particles. The polymer and osmolytes interact only by hard-sphere (HS) potential resulting in excluded-volume effects, which we calculated using the free volume theory. HS systems are efficient models that have been used to study dense gases and liquids, and complex liquid systems [149]. By studying the HS system we neglect interaction enthalpy effects between polymer or protein and osmolyte [150], but we focus on the often neglected excluded-volume features of osmolyte-induced stability. We begin with the Percus-Yevick HS integral equation to determine the compressibility of a dense hard-sphere gas as a function of the packing fraction. Thiele [151], Wertheim [152], and Carnahan and Starling [153] have found an approximate closed-form solution of the HS Percus-Yevick equation for the compressibility of a dense hard-sphere gas:

$$Z = \frac{PV}{NRT} = \frac{1 + \phi + \phi^2 - \phi^3}{(1 - \phi)^3} \quad (4.13)$$

Here Z is the compressibility and $\phi = N_o 4\pi r_o^3 / 3V$ is the packing fraction of hard spheres with radius r_o . On the other hand from the hard-sphere equation of state, the free energy is given by:

$$F = \int dF = k_B T N \int \frac{Z}{\phi} d\phi \quad (4.14)$$

By substituting equation (4.13) into equation (4.14) we obtain the free energy of a system of hard sphere fluid:

$$F = TN \left(\log(\phi) + \frac{2}{1 - \phi} + \frac{1}{(1 - \phi)^2} \right) \quad (4.15)$$

The first term in this expression is the free energy of an ideal gas. In the low packing fraction limit, $\phi \rightarrow 0$, the second and third terms in equation (4.15) become negligible.

To apply the above expression into our model of mixture of polymer and osmolytes it is important to note that $\phi = N_o 4\pi r_o^3 / 3V_a$ where V_a is the volume that is available to osmolytes. Depending on the polymer configuration, V_a varies as a function of the number of contacts, n , which is itself taken to be a function of v . The volume of the system that is excluded to the osmolytes decreases when a contact forms, because the total depletion zone of two monomers decreases when they are closer than $2(r_m + r_o)$ [154]. In our model we take the following form for the available volume:

$$V_a(v) = V_{system} - N \frac{4\pi}{3} (r_m + r_o)^3 + v_{dz} n(v) \quad (4.16)$$

where V_{system} is the total volume of the system (the outer box in figure 4.9), v is the principal volume of the polymer, $n(v)$ is given in equation (4.11), and r_m and r_o are the radii of monomers and osmolytes, respectively. v_{dz} is a purely geometrical parameter that counts the gain in the available volume when the depletion zones of two monomers in contact overlap, and is given by [154]:

$$v_{dz} = \begin{cases} \pi(2r(r_m + r_o)^2 - \frac{r^3}{6}), & 2r_m \leq r \leq 2(r_m + r_o) \\ 0, & 2(r_m + r_o) < r \end{cases} \quad (4.17)$$

where r is the distance between two (spherical) monomers.

Since there are no interactions between the polymer and the osmolytes, the final expression for the free energy of the mixture of a N -mer and N_o osmolytes in a volume V_{system} can be written as

$$F = F_{poly}(N, v) + F_{osm}(N_o, V_a(v)) \quad (4.18)$$

The free energy expression decouples into two separate terms corresponding to pure polymer subsystem and pure hard sphere fluid subsystem in a volume V_a . The connection between these two terms is held merely in the dependence of V_a on v . Figure 4.11 shows how the presence of osmolytes stabilizes the collapsed conformations of a 50-mer. The contact energy ε between monomers in the homopolymer is taken to be $3k_B T$, and the osmolyte packing fraction $\phi = 0.3$.

In figure 4.11 the polymer in implicit solution without osmolytes (magenta plot) shows a free energy minimum at high principal volume of the polymer, v , corresponding to an extended state. For osmolyte to polymer monomer size ratio $r_o/r_m = 0.7$ (navy), the free energy has shifted to lower volume $v = 120\text{\AA}^3$, corresponding to a stable collapsed polymer state. This shows that the presence of pure HS osmolytes can induce polymer collapse by excluded volume effects. Even more remarkable behavior is observed when the size of the monomer is decreased, while fixing the volume fraction at $\phi = 0.3$. The free energy profiles in figure 4.11 for small osmolytes, $r_o/r_m = 0.6$ (black) and $r_o/r_m = 0.5$ (maroon) show that the collapsed configurations are even more strongly favored. Our results show that for a given packing fraction of osmolytes, smaller osmolytes are better stabilizers. This is in agreement with previous theoretical results, for example calculations done by Takada *et al.* confirms this conclusion [61, 62] and model calculations [66]. Moreover, Zhou has done both experimental and analytical studies on the size effects of crowders on protein folding and his results confirm that smaller osmolytes (Dextran 6KD, 10KD and 20KD) have stronger effect than larger osmolytes (Ficoll 70

and Dextran 70KD, 100KD and 150KD) [155–157]. Tsao *et al.* have also shown that for hard sphere osmolytes, the smaller osmolyte have a larger protecting effect per unit weight/volume concentration [158]. Saunders *et al.* have further investigated the effect of polyol osmolytes with different molecular weights (80 to 504 g/mol) on two conformational equilibria of iso-1-ferricytochrome c and showed that smaller polyols induce larger protein stabilization [159].

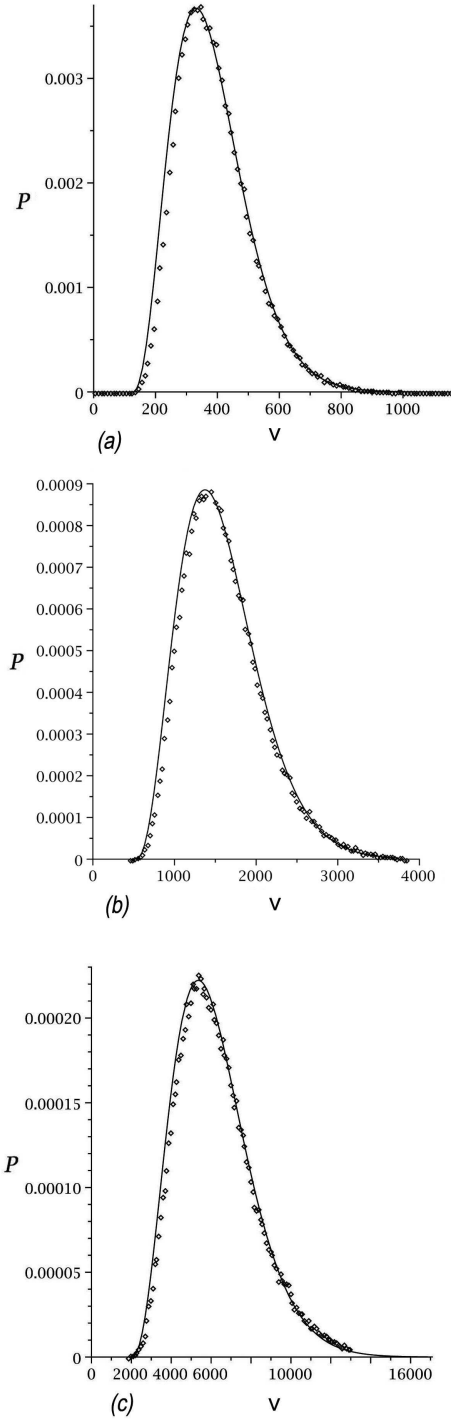


Figure 4.7: Volume probability distribution of a (a) 50-mer, (b) 100-mer and (c) 200-mer. DMD simulations are performed for homopolymers with $N = 50, 100, 200$ and the volume of the polymer in each conformation is calculated (circles). By curve fitting the DMD data, it is found that a universal expression that describes the volume probability distribution of homopolymers $P(v)$ is given by 4.4. v is in units of \AA^3 .

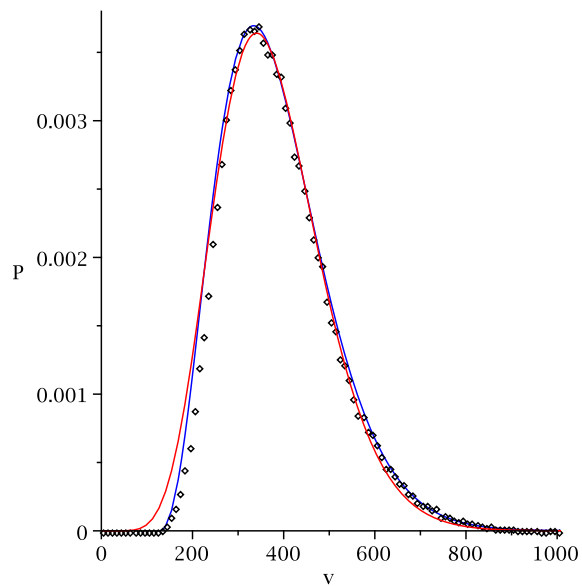


Figure 4.8: Volume probability distribution of a 50-mer, and the fits using 4.4 with (blue) and without (red) the factor of v_0 . It can be seen that the model with more variables (the shift factor v_0) provides a more accurate description. The F-value is 122, with associated cumulative probability of 0.008 by random chance. v is in units of \AA^3 .

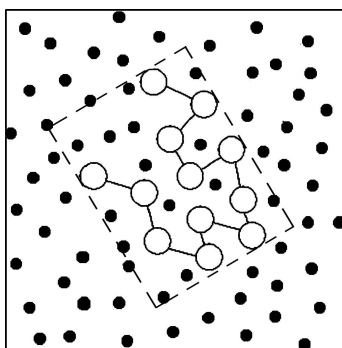


Figure 4.9: A schematic representation of the semi-grand canonical scheme.

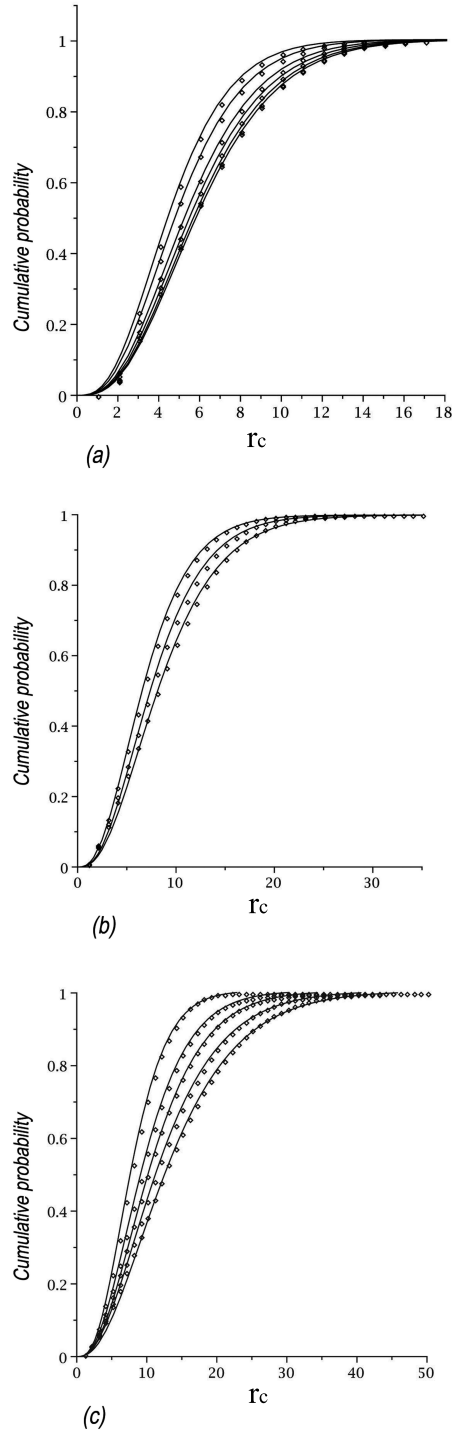


Figure 4.10: Cumulative probability of a (a) 50-mer, different curves from top to bottom correspond to $v_c = 200, 250, 350, 450, 650, 950 \text{ \AA}^3$, (b) 100-mer, different curves from top to bottom correspond to $v_c = 1000, 1500, 3500 \text{ \AA}^3$ and (c) 200-mer, different curves from top to bottom correspond to $v_c = 2000, 3000, 4000, 6000, 13000 \text{ \AA}^3$. r_c is in units of angstrom.

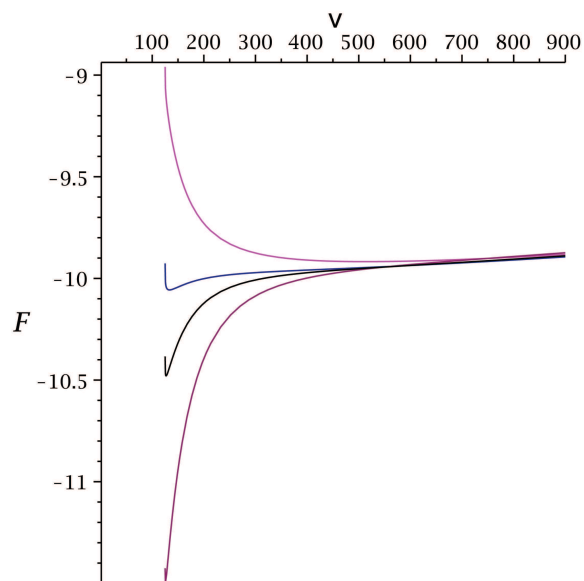


Figure 4.11: Free energy of a 50-mer with and without osmolytes present, as a function of polymer volume as defined by the volume of the principal box. Magenta curve: without osmolyte; other curves have osmolytes present with fixed packing fraction $\phi = 0.3$, but varying osmolyte sizes, and thus differing concentrations of osmolytes. Navy: 50-mer and osmolytes with $r_o/r_m = 0.7$; Black: 50-mer and osmolytes with $r_o/r_m = 0.6$; Maroon: 50-mer and osmolytes with $r_o/r_m = 0.5$. v is in units of \AA^3 .

Chapter 5

The Effect of Osmolytes on Protein Folding

In this chapter I will derive the free energy of an isolated protein and investigate the changes in the protein free energy when osmolytes, which are modeled as hard spheres are added to the system.

Proteins are polymer chains with selected attractive interactions between their residues so that they fold into one or a few specific conformations [160–162]; minimal frustration enables proteins to acquire [163] and sustain [164] their native structure, resulting in a globally funneled energy landscape guiding the conformational search [80, 165–172].¹ The starting point of many theoretical models of protein folding is the derivation of a free energy profile [173–178].

Our polymer model can be generalized to a protein model by adding a few energetic terms to the polymeric energy. Following the formulation of Bryngelson and Wolynes [164], [75] we assign an energy of $-\epsilon_p$ to a pair of amino acids when they are in their native state and zero when they are not, which naively will resemble the secondary structure energy. By entering explicit cooperativity into our model, we mimic the idea that only formed secondary structure units can couple [177, 179]. Moreover, three-body interactions are added into the energetic contributions, by including terms in the energy functional that are proportional to higher powers of the number of native contacts than simply linear terms.

¹A version of this chapter is in preparation for publication [6].

In the previous section we found the total number of contacts as a function of the size of the protein. The largest number of contacts corresponds to the native protein or fully-collapsed polymer, where the protein or polymer has its smallest volume, that is $n_{max} = n(v_0)$.

In Figure 5.1 we compare the largest number of contacts per residue of a protein suggested in our model and the number of bonds made per monomer in a Hamiltonian walk of N steps (on-lattice) which was studied by Douglas and Ishinabe [180]. For the most collapsed walk, the dependence of the number of bonds made per monomer, $z(N)$, in a Hamiltonian (dense) walk of N steps, on N is clearly due to the fact that monomers on the surface have less contacts than in the bulk. For three-dimensional systems, $z(N)$ is given approximately by:

$$z(N) \equiv \frac{1}{N} \text{Int}[2N - 3(N+1)^{2/3} + 3] \quad (5.1)$$

where $\text{Int}[\]$ means the integer part. From equation (5.1) we see that the effect of the surface on the number of contacts is quite important even for large macromolecules, as $z(N)$ approaches its bulk value of two contacts per monomer rather slowly, as $\sim 2 - 3N^{-1/3}$.

We define $Z = n(v)/n_{max}$ as the fraction of total contacts per monomer in any conformation. As mentioned above there is an attractive energy attributed to each bond that is made between two non-local residues that takes care of the polymeric energy. Moreover, we define $q = Q(v)/n_{max}$ as the fraction of native contacts present in an arbitrary conformation of the protein where $Q(v)$ is the number of native contacts in a given conformation and n_{max} is the maximum number of native contacts. Finally the internal energy of the protein can be written in the mean-field approximation as [181]:

$$\begin{aligned} E(q, Z) &= -n_{max}\epsilon_p q \\ &- n_{max}\epsilon_s [(1-\alpha)q + \alpha q^2] \\ &- n_{max}\epsilon_t [(1-\alpha)Z + \alpha Z^2] \\ &- n_{max}\frac{\epsilon_z^2}{2T} \{1 - [(1-\alpha)q + \alpha q^2]^2\} \end{aligned} \quad (5.2)$$

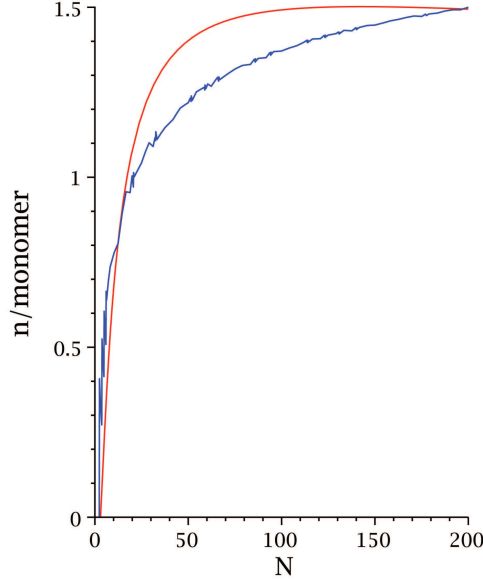


Figure 5.1: The largest number of contacts per residue of a protein calculated in our model (red) and the number of bonds made per monomer in a Hamiltonian walk of N steps (blue) which was studied by Douglas and Ishinabe [180]. For the most collapsed walk, the dependence of the number of bonds made per monomer, $z(N)$, in a Hamiltonian (dense) walk of N steps, on N is clearly due to the fact that monomers on the surface have less contacts than in the bulk. For three-dimensional systems, $z(N)$ is given approximately by (5.1).

where ϵ_s are the energetic parameters of the protein, ϵ_s is an energy scale for native contacts, ϵ_t is an energy scale for contacts of any kind (native or non-native), and ϵ_z is an energy scale for non-native contacts. α is a measure of the amount of three-body force present, when $\alpha = 0$ there are purely two-body forces, and $\alpha = 1$ indicates purely many-body (higher than two-body) forces. . We assume that each amino acid residue can take v discrete states. The parameter v resembles the coordination number of a lattice. We take 10 available states per residue consisting of one native state and 9 unfolded states. The polymeric part of the entropy of a protein is consistent with the polymer entropy $S(Z)$ which was derived in previous chapter. In order to find S as a function of Z we need to substitute $v(Z)$ into $S(v)$

where $v(Z)$ is the inverse function of $Z(v)$ derived previously), and again $S_0 = N \log(v)$.

Furthermore, we have to count for changes in entropy S_{mix} due to different ways of choosing $n_{max}q$ native contacts from $n_{max}Z$ total contacts.

$$S_{mix} = n_{max}(Z \log Z - q \log q - (Z - q) \log(Z - q)) \quad (5.3)$$

As discussed by Plotkin *et al.* [177, 182] the reduction in entropy S_b accounts for the fact that a particular set of native contacts $n_{max}q$ should be formed and finally the entropy decrease S_{ab} of having a set of native contacts $n_{max}Z - n_{max}q$ that should not be formed.

$$S_b = \frac{3}{2} n_{max} q (\log C + \log q) \quad (5.4)$$

and

$$S_{ab} = n_{max} \frac{1}{C} \int_{Cq}^{CZ} dx \log(1 - x^{3/2}) \quad (5.5)$$

where C is a constant of order 1. The free energy of the protein is then given by:

$$F(Z, q) = E(q, Z) - T(S(Z) + S_0 + S_{mix} + S_b + S_{ab}) \quad (5.6)$$

Figure 5.2 shows the free energy of a coarse-grained protein with 50 amino acids and $C = 0.6$, $\alpha = 0.1$, $\epsilon_p = 2.24$, $\epsilon_t = 0.6$, $\epsilon_z = 0.9$, $T = 1.51K$. We can see that the free energy has two minima at folded and unfolded states of the protein.

We can now study the effect of the presence of neutral osmolytes on the protein folding. By employing the Carnahan-Starling free energy for osmolytes as mentioned formerly, the free energy of the system consisting of the protein and osmolytes is given by:

$$F = F_{prot}(Z, q) + F_{osm}(N_o, V_a) \quad (5.7)$$

where $V_a = V_{system} - N \frac{4\pi}{3} (r_m + r_o)^3 + n_{max}(\bar{v}_{dzz}Z + \bar{v}_{dzq}q)$.

The parameters \bar{v}_{dzz} and \bar{v}_{dzq} are average depletion zone volume for total and native contacts correspondingly which are given by equation (4.17). For $r_m =$

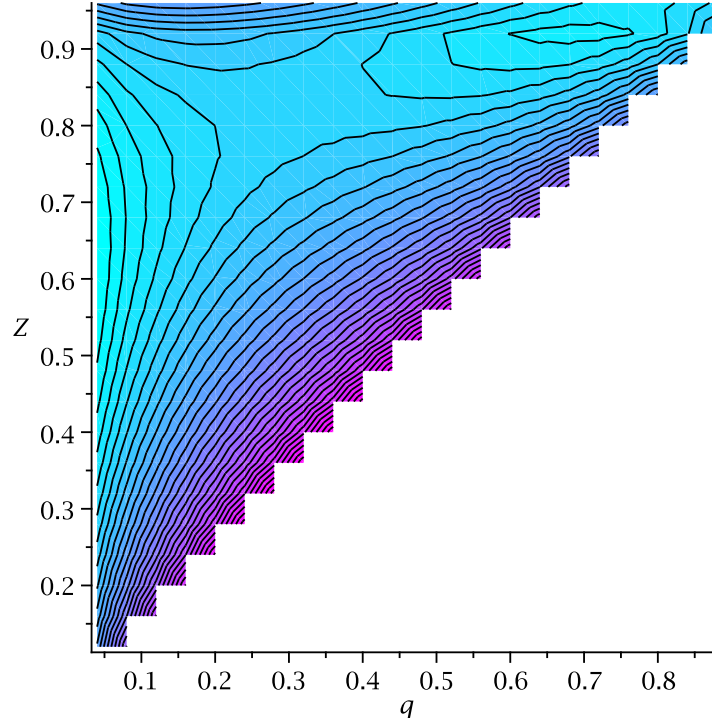


Figure 5.2: Free energy of a coarse-grained protein with 50 amino acids using equation (5.6) with $C = 0.6, \alpha = 0.1, \epsilon_p = 2.24, \epsilon_t = 0.6, \epsilon_s = 0.5, \epsilon_z = 0.9, T = 1.51K$. We can see that the free energy has two minima at folded ($q = 0.75, Z = 0.9$) and unfolded ($q = 0, Z = 0.65$) states of the protein.

0.48\AA and $r_o = 0.24\text{\AA}$ we get $\bar{v}_{dzz} = 1\text{\AA}^3$ and $\bar{v}_{dzq} = 2.2\text{\AA}^3$ (the values for r_m and r_o are unphysical values and are used purely in the context of the model). Figure 5.3 shows that by adding neutral osmolytes with packing fraction $\phi = 0.12$, the free energy minimum shifts to high Z and q , and the folded state of the protein becomes the stable state.

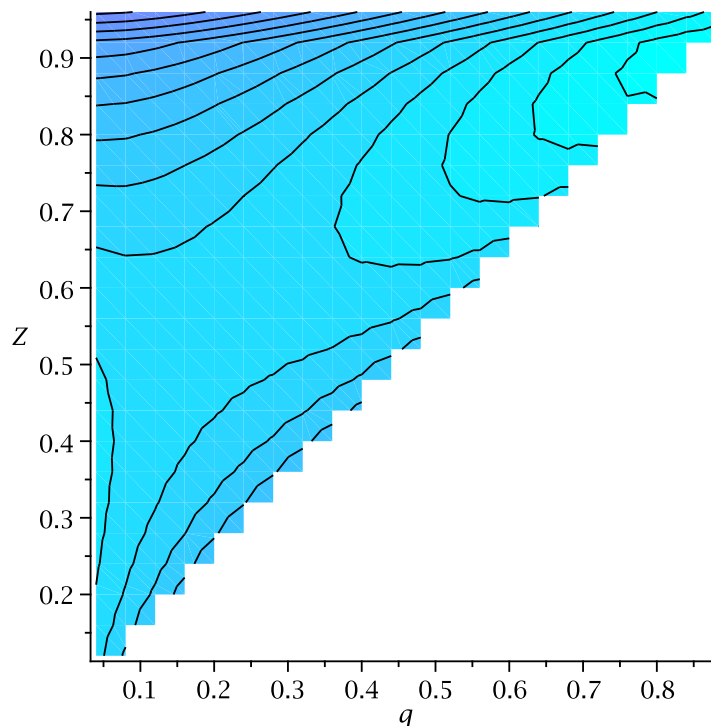


Figure 5.3: Free energy of a coarse-grained protein with 50 amino acids with $C = 0.6, \alpha = 0.1, \varepsilon_p = 2.24, \varepsilon_t = 0.6, \varepsilon_s = 0.5, \varepsilon_z = 0.9, T = 1.51K$ in the presence of neutral osmolytes with packing fraction $\phi = 0.12$ using equation (5.7). We can see that the free energy minimum shifts to high $Z = 0.9$ and $q = 0.85$, and the folded state of the protein becomes the stable state.

5.1 Interactive Osmolytes

The stabilizing property of osmolytes has been shown to correlate with the preferential exclusion of osmolytes from unfolded protein domains, resulting in the preferential accumulation of water (preferential hydration) near an unfolded protein [183, 184]. This implies a net repulsive interaction between stabilizing osmolytes and protein, and indeed preferential exclusion has been shown to arise from repulsive interactions between osmolytes and the backbone of proteins [49, 185, 186]. Repulsive osmolyte-backbone interactions would raise the enthalpy of a protein,

and the increase in enthalpy would be larger for the unfolded state due to its larger solvent exposed backbone area. Consequently the unfolded state would be more destabilized, stabilizing the folded native state.

In contrast to protective osmolytes that protect cells against environmental stresses such as high temperature, desiccation and pressure, denaturants, such as urea and guanidinium chloride (GdmCl), destabilize proteins. The interaction between urea and the protein is attractive leading to the preferential accumulation of urea in the vicinity of proteins [187]. The attractive urea-protein interactions lower the free energy of both the native and unfolded states, but due to its larger solvent exposed surface area, the free energy of the unfolded state is lowered more compare to folded state [49]. Consequently, the addition of urea to protein solutions shifts the equilibrium to the unfolded state. The attractive interactions between urea and protein must overcome the entropic stabilization of the protein mentioned above. In the process of denaturation proteins or nucleic acids fail to preserve their tertiary structure and secondary structure by application of some external compound, such as a strong acid or base, a concentrated inorganic salt, an organic solvent such as alcohol and chloroform. Denaturation of the proteins in a living cell cause disruption of cell activity whose characteristics range from loss of solubility to communal aggregation and possibly cell death.

So far we have studied the effect of neutral osmolytes, however; as we discussed above crowding agents in general can interact with amino acids of a protein. To calculate the effect of these interactions on protein stability, we need to measure the surface area of the protein as a function of its total and native contacts. In the completely extended conformation, the surface area of a N -mer protein is given by:

$$A = 4\pi N(r_m + r_c)^2 \quad (5.8)$$

where as mentioned above r_m and r_c are the radius of monomers and crowding agents respectively. When a bond forms between two monomers, the surface area of the protein reduces by:

$$\Delta A = \pi r^2 \quad (5.9)$$

for $2r_m \leq r \leq 2(r_m + r_c)$ where r is the distance between two monomers in contact

[188]. Therefore, for an arbitrary conformation of the protein with $n_{max}Z$ number of total contacts and $n_{max}q$ number of native contacts the surface area is given by:

$$A = \pi N \left(4(r_m + r_c)^2 - n_{max}(\overline{r_Z^2}Z + \overline{r_q^2}q) \right) \quad (5.10)$$

where $\overline{r_Z^2}$ and $\overline{r_q^2}$ are average r^2 for total contacts and native contacts. The difference between these two parameters comes from the fact that non-native contacts have a larger cut-off distance.

The interaction free energy here is very similar to the free energy of the adhesion of gas molecules to a 2D lattice sites with binding energy ϵ_{po} . However; the number of adsorbing sites available on the surface of a single protein that is embedded in a sea of osmolytes with packing fraction of more than 30% is much less than the number of crowding particles, and therefore to a good approximation we can assume that all the adsorbing sites on the protein surface are always occupied by an osmolyte. This means that the number of states is 1 and the adsorption entropy is zero. Hence we have:

$$F_{int} = \langle E_{ads} \rangle = \epsilon_{po} N_s \frac{e^{-\epsilon_{po}/k_B T}}{1 + e^{-\epsilon_{po}/k_B T}} \quad (5.11)$$

where N_s is the number of the adsorbing sites on the protein surface and is given by $\frac{A}{\pi r_c^2}$. The free energy of the mixture is:

$$F = F_{prot}(Z, q) + F_{osm}(N_o, V_a(Z, q)) + F_{int}(Z, q) \quad (5.12)$$

Figure 5.4 shows the stability ($= F_{folded} - F_{unfolded}$) as a function of ϵ_{po} .

The reduction in entropy of the unfolded state due to osmolytes results in a loss of folding cooperativity. This can be seen by investigating the heat capacity as a function of temperature as osmophobicity of the solvent, ϵ_{po} , is varied. It is clearly seen that there is a loss in folding cooperativity when osmolytes are stabilizing the native state, again due to the reduction in conformational entropy of the unfolded state. Conversely, there is an increase in folding cooperativity in the presence of denaturant due to swelling of the unfolded protein. Figure 5.5 shows the heat capacity plot of a protein with 50 residues in a box of crowding agents with packing fraction 30% and $\epsilon_{po} = -0.2, 0, 0.2$ using:

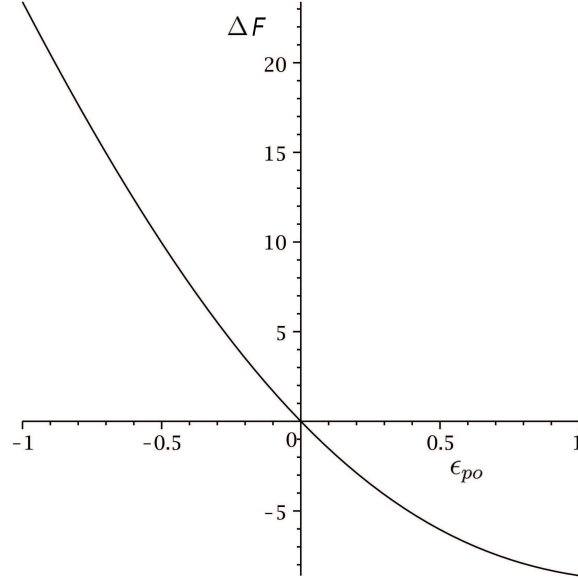


Figure 5.4: The stability of a protein with 50 residues as a function of ϵ_{po} in units of T .

$$C_v = \frac{d\langle E \rangle}{dT} = \frac{d}{dT} \frac{E_f \exp(-F_f/k_B T) + E_u \exp(-F_u/k_B T)}{\exp(-F_f/k_B T) + \exp(-F_u/k_B T)} \quad (5.13)$$

where E_f (equation (5.2) with $q = 0.1$ and $Z = 0.6$) and E_u (equation (5.2) with $q = 0.7$ and $Z = 0.9$) are the internal energy of the protein in folded and unfolded state correspondingly and F_f (equation (5.12) with $q = 0.1$ and $Z = 0.6$) and F_u (equation (5.12) with $q = 0.7$ and $Z = 0.9$) are the free energy of the folded and unfolded protein correspondingly.

We can compare our analytical result with the DMD all-atom simulation of the Trp-cage protein, a designed 20-residue protein. The initial heavy-atom positions were obtained from the NMR structure (structure 1 of PDB 1L2Y34) and the missing polar hydrogen molecules are constructed as in [133]. The resulting structure is comprised of 189 heavy atoms and polar hydrogen atoms (non-polar hydrogen atoms are not represented), which in the discontinuous model are represented as hard spheres. The energetic parameters that are used in the simulation are explained in next chapter.

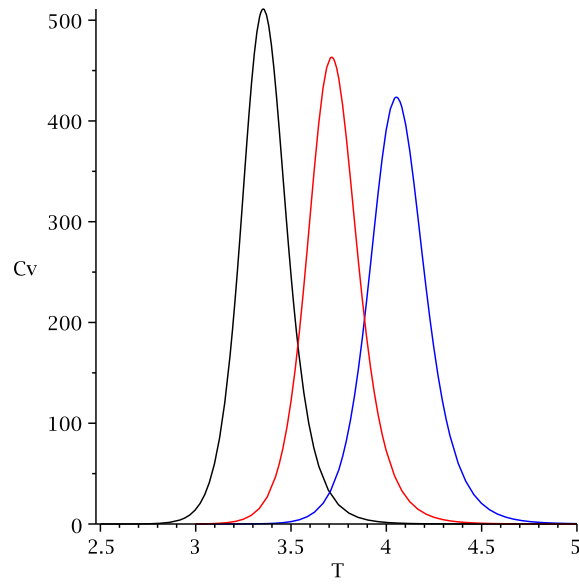


Figure 5.5: The heat capacity of a coarse-grained protein with 50 amino acids in a box of crowding agents with packing fraction 30% with $r_o/r_m = 0.9$ and $C = 1, \alpha = 0.1, \epsilon_p = 2.24, \epsilon_t = 0.6, \epsilon_z = 0.9$. Different curves correspond to $\epsilon_{po} = -0.2$ (black), $\epsilon_{po} = 0$ (red) and $\epsilon_{po} = 0.2$ (blue).

Chapter 6

Discontinuous Molecular Dynamics Simulations

In this chapter an all-atom $G\bar{O}$ simulation of Trp-cage protein is discussed employing discontinuous molecular dynamics in an explicit minimal solvent, using a single, contact-based interaction energy between protein and solvent particles.

Molecular dynamics (MD) is a very popular technique for simulating the properties of large systems of molecules.¹ By recurrently solving Newton's equations of motion, the trajectories of a group of atoms ($10^2 - 10^6$) are computed. As the system evolves in time, its macroscopic properties can be calculated by taking averages of instantaneous properties. In general molecular dynamics simulations are categorized as continuous molecular dynamics and discontinuous molecular dynamics (DMD). In the former, a continuous potential energy field such as Lennard-Jones potential generates the force that acts on molecules whereas in DMD, the potential is discontinuous, for instance the hard-sphere potential and the force is impulsive. Since the time integration in DMD is very simple, it provides the opportunity to explore phenomena that require long time scales using relatively inexpensive hardware.

An effective denaturant or osmolyte solution can be constructed by making the interaction energy attractive or repulsive. A statistical mechanical equivalence is demonstrated between this effective solvent model and models in which proteins

¹A version of this chapter has been submitted for publication for publication [7].

are immersed in solutions consisting of water and osmolytes or denaturants. Analysis of these studies yields the following conclusions: 1) Osmolytes impart extra stability to the protein by reducing the entropy of the unfolded state. This result is in agreement with previous experimental studies, for example Gahl *et al* have shown that the physiological osmolyte TMAO stabilizes onconase (ONC) by decreasing the entropy of the unfolded state through a solvophobic effect [189] and Milev *et al* have shown that both glycerol and sorbitol lead to a large entropy loss in sequence-specific protein-DNA association of a complex composed of the integrase Tn916 DNA-binding domain and its target DNA duplex [190]. Zhou also employed a very simple model to calculate the effect of macromolecular crowding on folding entropy [155]. 2) Unfolded states in the presence of osmolyte are more collapsed than in water. Dhar *et al* have combined experiment and simulation to the effect of Ficoll 70 on the structure and function of phosphoglycerate kinase (PGK) [191]. In their paper Dhar *et al* show that the radius of gyration of PGK decreases as the volume fraction of Ficoll 70 increases. Moreover Stagg *et al* have investigated the consequences of Ficoll 70 on the 148-residue single-domain α/β protein, *Desulfovibrio desulfuricans* apoflavodoxin, and observed a decrease on the size of the protein (radius of gyration) as the concentration of Ficoll 70 increases [192]. 3) The folding transition in osmolyte solutions tends to be less cooperative than in water, as determined by the ratio of van't Hoff to calorimetric enthalpy changes. The decrease in cooperativity arises from an increase in native structure in the unfolded state, and thus a lower thermodynamic barrier at the transition midpoint. Auton *et al* have theoretically and experimentally calculated the m value of RCAM-T1 induced by TMAO, proline and urea [193]. The m value is a measure of the folding transition cooperativity and has units of $kcal/mol.M^{-1}$. They have confirmed that the folding cooperativity increases in the presence of urea and decreases in the presence of TMAO and proline. Moreover, Baskakov *et al* have shown that TMAO induces RCAM-T1 to fold in a cooperative manner, reaching a maximum in folding at TMAO concentrations above 2.5 M [194]. Henkels *et al* also have shown that *Bacillus subtilis* RNase P (P protein) which is predominantly unfolded in 10 mM sodium cacodylate at neutral pH undergoes a cooperative folding transition upon addition of TMAO [195]. 4) Weak denaturants destabilize small proteins not by lowering the unfolded enthalpy, but primarily by swelling the unfolded state

and raising its entropy. 5) The folding transition in denaturant-containing solutions is more cooperative than in water. This result is in agreement with previous experimental results, for example Auton *et al* have shown that RCAM-T1 folding is more cooperative in the presence of urea [193]. 6) Transfer to a concentrated osmolyte solution with purely hard-sphere steric repulsion significantly stabilizes the protein, due to excluded volume interactions not present in the canonical Tanford transfer model. 7) While a solution with hard-sphere interactions adds a solvation barrier to native contacts, the folding is nevertheless less cooperative for reasons 1-3 above, because a hard-sphere solvent acts as a protecting osmolyte.

The representative protein used in this work to illustrate the effects of solvent quality is a DMD all-atom model of the Trp-cage protein [133], a designed, 20-residue, truncated construct exhibiting cooperative folding to a stable structure at physiological pH. Initial heavy-atom positions were obtained from the NMR structure (structure 1 of PDB 1L2Y(35)) and the missing polar hydrogen molecules are constructed as in reference [133]. A $G\bar{o}$ model potential [196] is implemented by setting the non-bonded square-well depth ϵ_{pp} to $-\epsilon_{Go}$ for all ij pairs in the equilibrated structure with van der Waals overlap $r < 1.2\sigma_{ij}^{vdW}$, where σ_{ij}^{vdW} is the sum of the van der Waals (vdW) radii $r_i + r_j$ for each atom; for all other non-bonded ij pairs, the square-well depth is set to 0, so that these atom pairs are purely repulsive. The energy scale is set by the $G\bar{o}$ contact energy as in previous DMD studies [133, 197]; thus simulations are performed with the $G\bar{o}$ contact energy ϵ_{pp}^* set to $-\epsilon_{Go}^* = -1$, and all energies and temperatures are scaled in units of ϵ_{Go} ($E^* = E/\epsilon_{Go}$ and $T^* = k_B T/\epsilon_{Go}$).

The $G\bar{o}$ model protein in explicit solvent is implemented by placing the Trp-cage protein in a $40\text{\AA} \times 40\text{\AA} \times 40\text{\AA}$ box, along with a variable number of spherical solvent molecules randomly inserted without hardcore overlaps. Standard periodic boundary conditions are implemented. A typical simulation consisted of 1000 spherical solvent particles of radius 1.5\AA (the approximate radius of water). This is about half the number of water molecules in a 55M solution for a $(40\text{\AA})^3$ box. We employ such a dilute concentration for computational convenience; physical concentrations have collision times sufficiently short as to make such simulations prohibitively slow. Diluting the concentration weakens the effects that would be observed by varying solvent qualities from those at 55M, that is, this simplification

effectively places lower bounds on any trends that we predict would be observed. For this reason we find the approximation acceptable as it only strengthens the conclusions of this study.

Solvent molecules interact with both protein moieties and with each other by a square-well potential with well position given by $0.8\sigma_{ij}^{xs} < r < 1.2\sigma_{ij}^{xs}$ and well depth given by the parameter ϵ_{xs} or $\epsilon_{xs}^* = \epsilon_{xs}/\epsilon_{Go}$ where x may be either a protein atom p or another solvent residue s . If $r < 0.8\sigma_{ij}^{xs}$ the potential is infinity. For solvent-solvent interactions, σ_{ij}^{ss} is the vdW diameter of the solvent, which we generally set to 3.0\AA , roughly the size of a water molecule. ϵ_{ss}^* is the solvent-solvent square-well depth in units of the $G\bar{o}$ contact energy. For solvent-protein interactions, $\sigma_{ij}^{ps} = (\sigma_i^{vdW} + \sigma_j^{water})/2$ is the average vdW diameter of the protein-solvent ij pair, where σ_i^{vdW} is the vdW diameter of the i th atom of the protein according to CHARMM potential set 19 [147], and $\sigma_j^{water} = 3.0\text{\AA}$ is the vdW diameter of the j th water molecule. ϵ_{ps}^* is the protein-solvent square-well depth in units of the $G\bar{o}$ contact energy.

The quantity ϵ_{ps}^* is a measure of the solvent quality. As shown below through a correspondence between explicit and implicit solute, it is a well-defined function of solute interactions with the protein. In this study the solvent quality is varied from a minimum value of -0.8 representing a strongly denaturing aqueous urea solution, to a maximum value of $+1.0$ representing a strongly protein-stabilizing aqueous osmolyte solution. We may consider $\epsilon_{ps}^* = 0$ as representing a reference solution of “pure-water” (this is still stabilizing for the protein since $\epsilon_{pp}^* = -1$). The solvent-solvent square-well depth is generally taken to be $\epsilon_{ss}^* = -1$, reflecting an overall preference for solvent particles to interact with each other at least as favourably as with the protein ($\epsilon_{ss}^* < \epsilon_{ps}^*$).

6.1 Free Energy and Entropy Functional

The free energy calculations used here are based on standard multiple-histogram method [198–201], which calculates thermodynamic quantities by approximating the density of states $g(\epsilon)$ (that is, the number of states with energy ϵ) from simulation data. This gives the partition function $Z(x) = \Sigma^* g(\epsilon, x, \Delta x) \exp(-\beta\epsilon)$ where $*$

indicates a summation over all energy states with the order parameter constrained to values ranging from x to $x + \Delta x$, and $g(\epsilon, x, \Delta x)$ is the density of state of energy with the order parameter range x to $x + \Delta x$. Thermodynamic quantities only depend on the bin size Δx by an additive constant, so long as Δx is sufficiently small in the traditional coarse-graining sense. The free energy then is given by:

$$F(x) = -k_B T \log Z(x) = -k_B T \log P(x) + F \quad (6.1)$$

where $P(x)$ is the probability the system is within the order parameter range x to $x + \Delta x$, and $F = -k_B T \log Z$ is the total free energy. The internal energy is:

$$U(x) = \langle E(x) \rangle = \sum \frac{\epsilon g(\epsilon, x, \Delta x) \exp(-\beta \epsilon)}{Z(x)} \quad (6.2)$$

and the entropy $S(x) = (U(x) - F(x))/T$.

In this work we use the non-local native fraction of contacts Q as an order parameter, defined by first counting all atom pairs in the native structure that are within 1.2 times the sum of their hard-core radii, and between residues ij , such that $|i - j| > 3$. This gives 276 contacts in the native state. The value of Q in a given (partly folded) configuration is the fraction of these contacts present, and varies from 0 (completely unfolded) to 1 (completely folded). A bin size $\Delta Q = 0.02$ was used. As mentioned earlier, in here the protein internal energy includes intra-protein energy and protein-solvent energy, but not solvent-solvent interaction energy. The computed energy $u^*(Q)$, free energy $f^*(Q)$, and entropy $s^*(Q)$ functions are in units of the $G\bar{o}$ contact energy.

6.2 The Correspondence between Explicit and Implicit Osmolyte Models

We are interested in the effects of osmolytes on polymeric species in solution, including homopolymers and proteins. The system consists of protein or polymer (p), solvent such as water (s), and osmolyte solute (o). There are 6 relevant interaction enthalpies corresponding to the pair interactions between the above species: $\epsilon_{ss}, \epsilon_{pp}, \epsilon_{oo}, \epsilon_{po}, \epsilon_{ps}, \epsilon_{so}$. These interaction energies along with temperature T are

the relevant energy scales in the problem. There are also 3 length scales in the problem corresponding to the sizes (radii) of the various species or their constituents: r_p, r_s, r_o corresponding to the monomer, solvent, and osmolyte radius. In the approach described below, these radii simply correspond to different coordination numbers for each of the species.

A solution of water and osmolyte may be considered as an effective medium with modified solvent interactions so that the new system can be treated with just 3 energy scales: $\epsilon_{ss}^*, \epsilon_{pp}^*, \epsilon_{ps}^*$. To compare the simulation results to an analytic theory that treats osmolyte particles explicitly, we make the following formal correspondence between explicit and implicit osmolyte models. Given the six explicit-system (ES) parameters $\epsilon_{ss}, \epsilon_{pp}, \epsilon_{oo}, \epsilon_{po}, \epsilon_{ps}, \epsilon_{so}$, we seek the three implicit-system (IS) parameters $\epsilon_{ss}^*, \epsilon_{pp}^*, \epsilon_{ps}^*$ that would give the same average interaction probabilities for the system. The mean number of interactions N_{ij} between species i and j for the explicit system can be found from the sum rules for the total number of nearest neighbors:

$$\begin{aligned} q_p N_p &= 2N_{pp} + N_{ps} + N_{po} \\ q_s N_s &= 2N_{ss} + N_{ps} + N_{so} \\ q_o N_o &= 2N_{oo} + N_{po} + N_{so} \end{aligned} \tag{6.3}$$

where q_i is the coordination number of species i . The 3 quasichemical equations of mass balance for the reactions $po + so \rightleftharpoons ps + oo$, $so + so \rightleftharpoons ss + oo$, $po + po \rightleftharpoons pp + oo$, which give equations of the form:

$$\frac{N_{ps} N_{oo}}{N_{po} N_{so}} = e^{-\beta(\epsilon_{ps} + \epsilon_{oo} - \epsilon_{po} - \epsilon_{so})} \tag{6.4}$$

There are 3 independent rate equations, giving along with a total of 6 equations for the 6 N_{ij} . Several approximations are made in the above heuristic approach. The different sizes of the particles are only manifested in the different coordination numbers. More accurate models would modify the sum rules to more accurately account for the geometry of nearest neighbors, which would also result in modified stoichiometric coefficients in the mass balance equations. The parameter ϵ_{pp}

should be interpreted as a chemical potential, because the conformational entropy for protein monomers is smaller than that for free osmolyte or solvent particles.

To map to the implicit system, first let the IS particle numbers be given by: $N_p^* = N_p$, and $N_s^* = N_s + N_o$. Adding the solvent and solute equations in the sum rules for the total number of nearest neighbors (equation (6.3)) gives a relation between the coordination numbers and solvent particles which can be mapped to the implicit system:

$$q_s N_s + q_o N_o = 2(N_{ss} + N_{oo} + N_{so}) + N_{ps} + N_{po} \quad (6.5)$$

and

$$q_s^* N_s^* = 2N_{ss}^* + N_{ps}^* \quad (6.6)$$

Equating the numbers on the left hand side of (6.5) and (6.6) gives the coordination number for effective solvent particle as $q_s x_s + q_o x_o$, i.e. the solvent and osmolyte coordination numbers weighted by their respective mol fractions x_i . Equating protein-protein, protein-bath, and bath-bath contacts in both models gives the correspondence between N_{ij} and N_{ij}^* :

$$\begin{aligned} N_{pp}^* &= N_{pp} \\ N_{ss}^* &= N_{ss} + N_{oo} + N_{so} \\ N_{oo}^* &= N_{ps} + N_{po} \end{aligned} \quad (6.7)$$

The N_{ij}^* then determine the transfer energy in the IS model through:

$$\frac{N_{pp}^* N_{ss}^*}{(N_{ps}^*)^2} = e^{-\beta(\epsilon_{pp}^* + \epsilon_{ss}^* - 2\epsilon_{ps}^*)} \quad (6.8)$$

The other 2 equations determining the 3 energy scales in the IS model may be found from the conservation of total energy between the two models:

$$N_{pp}^* \epsilon_{pp}^* + N_{ss}^* \epsilon_{ss}^* + N_{ps}^* \epsilon_{ps}^* = \sum_{i,j} N_{ij} \epsilon_{ij} \quad (6.9)$$

And by the ansatz that the total bath interaction energy (or equivalently protein-bath interaction energy) be the same in both models:

$$N_{ss}^* \epsilon_{ss}^* = N_{ss} \epsilon_{ss} + N_{oo} \epsilon_{oo} + N_{so} \epsilon_{so} \quad (6.10)$$

Equations (6.8)-(6.10) then determine $\epsilon_{ss}^*, \epsilon_{pp}^*, \epsilon_{ps}^*$. For example, a system with $\epsilon_{ss}, \epsilon_{pp}, \epsilon_{oo}, \epsilon_{po}, \epsilon_{ps}, \epsilon_{so} = -1.25, -1.1, -0.4, 1.1, 0.25, -0.8$ in units of $k_B T$, with $q_p, q_o, q_s = 8, 6, 4$ and mol fractions $x_p, x_o, x_s = 0.05, 0.2, 0.75$ has effective energy scales $\epsilon_{ss}^*, \epsilon_{pp}^*, \epsilon_{ps}^* = -1, -1, 0.4$; a set of parameters that we often use in our simulations. We have thus shown that for any explicit osmolyte-solvent system, there exists an effective solvent model that captures the thermodynamics of the original system.

6.3 Results

6.3.1 Simulations

In this work, the Trp-cage $G\bar{O}$ model in implicit solvent, where the effects of water is incorporated into the intra-protein interaction energy, is used as a reference protein system (6.1A). The system is simulated by discontinuous molecular dynamics, with the initial Maxwell-Boltzmann distribution of particles defining the temperature T^* . The system exhibits two-state behaviour as shown by the heat capacity plot of figure 6.2a, which has a single first-order-like phase transition peak. The folding temperature $T_f^* = 4.0$, at the heat capacity maximum, separates the low temperate native state (N) from the high temperature unfolded (U) state. The plot was obtained by equilibrium simulations at temperatures from $T^* = 2.0$ to $T^* = 6.0$. The error bars were obtained by performing five $t^* = 60000$ (scaled time duration) independent runs at each temperature. To improve the accuracy near the folding temperature $T_f^* = 4.0$, where the fluctuation is high, $t^* = 300000$ five runs were performed at $T^* = 3.8, 4.0, 4.2$, which reduced the size of the error bar near the folding temperature. However, C^* versus T^* plots obtained from the multiple-histogram method vary negligibly with simulation length as long as a lower time limit (for the Trp-cage model $t^* = 60000$) is exceeded, thus the results are statisti-

cally reliable (we used $C^* = \frac{\langle E^{*2} \rangle - \langle E^* \rangle^2}{T^{*2}}$).

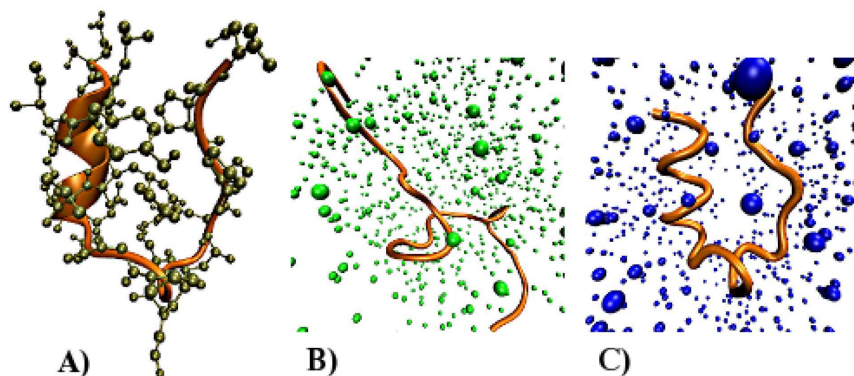


Figure 6.1: A: Ground-state native structure of the all-atom $G\bar{o}$ model of the Trp-cage protein; B: Denatured (unfolded) Trp-cage in urea solution; C: Folded Trp-cage in osmolyte solution.

Figure 6.2b shows the probability distribution versus energy at the folding temperature. The probability distribution is bimodal, with peaks at the unfolded and folded energy, no detectable specific intermediate state, and attenuated population between folded and unfolded states indicating a weakly two-state transition. We show below that an effective solvent with osmolyte present increases the folding temperature of the Trp-cage, and decreases the cooperativity of the transition, while an effective solvent modeling the presence of denaturant decreases the folding temperature and increases the cooperativity.

6.3.2 A Hard-sphere Solvent Induces a Desolvation Barrier between Native Contacts, but Decreases Folding Cooperativity Relative to the Implicit Solvent Model

Figure 6.2b shows the distribution of energy in the implicit solvent model at the folding temperature (solid), as well as distribution of energy of a “reference solvent” model at its own transition midpoint (dotted). The reference solvent is self-attractive ($\epsilon_{ss}^* = -1$), but has hard-sphere interactions with the protein ($\epsilon_{ps}^* = 0$). One can see here that the distribution of energy of the hard-sphere solvent is less cooperative than that of the implicit solvent model. However the protein in hard-

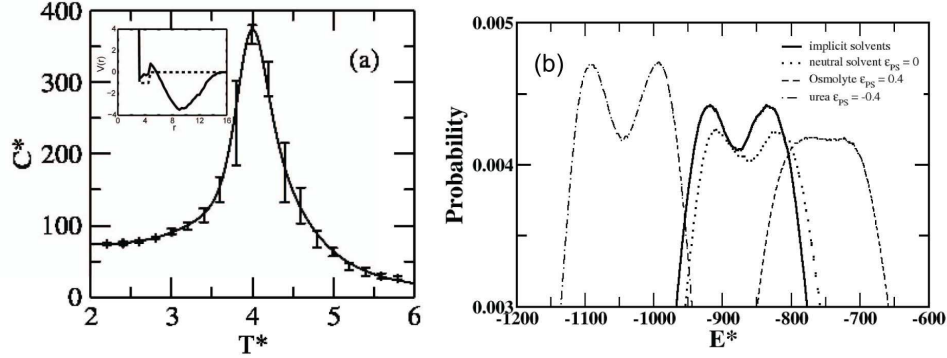


Figure 6.2: (a) Reduced Heat capacity ($C^* = C/k_B$) versus reduced temperature T^* of Trp-cage $G\ddot{o}$ model with implicit solvent. The plot is obtained by using the multiple-histogram method. The data points and error bars are averages taken from the five independent runs at each temperature. (b) Probability distributions of energy for protein-protein plus protein-solvent interactions, obtained by the histogram method, for several solvents at their respective folding temperatures (implicit solvent: $T_f^* = 4.0$, neutral solvent with $\epsilon_{ps}^* = 0$: $T_f^* = 4.55$, protective osmolyte solvent with $\epsilon_{ps}^* = 0.4$: $T_f^* = 5.08$, denaturing osmolyte solvent with $\epsilon_{ps}^* = -0.4$: $T_f^* = 3.913$). For explicit solvents, the energy generally includes protein-solvent interaction energy, however for the implicit and neutral hard-sphere solvent this contribution to the energy is zero. Protective osmolytes shift to the higher energies and show less cooperative transition while denaturing osmolytes shift to lower energies and show more cooperative transition. Comparing the neutral and implicit solvent histograms, the native ensemble shifts to higher energy because it has less overall native structure due to the less cooperative folding transition. The unfolded ensemble also shifts to the right because even though there is a tendency to have more native long range contacts, there are less local contacts.

sphere solvent still exhibits solvation barriers for native protein contacts (see figure 6.2a inset). Solvation barriers generally increase folding cooperativity, essentially by reducing the conformational space sampled by the protein in partially unfolded states, however native-centric modeling alone neglects other potential effects of the solvent on the unfolded ensemble. One important effect is the reduction of

polymer entropy in the unfolded state by induced polymer collapse, which results in an increased propensity for native and non-native contact formation. On the other hand a denaturing osmolyte such as GdHCl has attractive interactions with the protein, which results in expansion of the unfolded state, and increased folding cooperativity. We describe these effects in more detail below.

6.3.3 Effects of Osmolytes and Denaturants on Stability, Polymer Collapse, and Folding Cooperativity

Figure 6.3A shows C^* versus T^* plots of Trp-cage $G\overline{\sigma}$ model immersed in 1000 spherical solvent molecules with hard-sphere radius $r_{s,hs} = 1.5\text{\AA}$ (about the size of a water molecule) confined in a $(40\text{\AA})^3$ periodic box. For this system the solvent-solvent contact energy is fixed at $\epsilon_{ss}^* = -1$, with the solvent-protein contact energy varying from $\epsilon_{ps}^* = -0.6$ (strong urea solution) to $\epsilon_{ps}^* = 0.8$ (strong osmolyte solution). For the neutral solvent ($\epsilon_{ps}^* = 0$) the folding temperature increases to $T_f^* = 4.5$ and the heat capacity peak has decreased to $C_{peak}^* = 300$ in comparison to $C_{peak}^* = 390$ for the Trp-cage in implicit solvent model (figure 6.2a). The increase in the folding temperature as compared to the implicit solvent model in figure 6.3A, despite zero solvent-protein contact energy suggests that the change in thermodynamic property is due to excluded volume effects. That is, this significant change in stability is due to volume effects not accounted for in the Tanford transfer model ([14]), which is based upon interactions at the protein-solvent interface and thus has free energies scaling with solvent-accessible surface area, and energy of solvent-protein interaction. For $\epsilon_{ps}^* = 0$ there is no energy scale in the problem. For the osmolyte solvents (poor solvent), in which solvent molecules are repulsive to the protein ($\epsilon_{ps}^* > 0$), the folding temperature increases progressively to $T_f^* = 5.3$ for $\epsilon_{ps}^* = 0.8$, and the heat capacity peak also decreases progressively to $C_{peak}^* = 200$ for $\epsilon_{ps}^* = 0.8$. The addition of repulsive osmolyte solvents (bad solvents) stabilizes the Trp-cage, since the shift of the heat capacity peak indicates the native state is stable up to a higher temperature. For urea-like solvents (good solvent), in which solvent molecules are attracted to the protein ($\epsilon_{ps}^* < 0$), the folding temperature decreases progressively to $T_f^* = 3.5$ for $\epsilon_{ps}^* = -0.6$. The attractive urea-solvent interactions destabilize the native structure so that the Trp-cage unfolds at lower temperatures than for the reference (water-like) solvent with $\epsilon_{ps}^* = 0$.

The set of temperatures T_f^* of heat capacity peaks in figure 6.3A and their corresponding interaction energies ϵ_{ps}^* define a phase boundary between native (N) and unfolded (U) states. A solvent quality phase diagram can thus be plotted in figure 6.3B. This phase diagram illustrates that our protein in explicit solvent model reproduces protein stabilization/destabilization by osmolyte/urea, since the N state becomes more stable as the quality index ϵ_{ps}^* increases (that is becomes more like an osmolytic solvent).

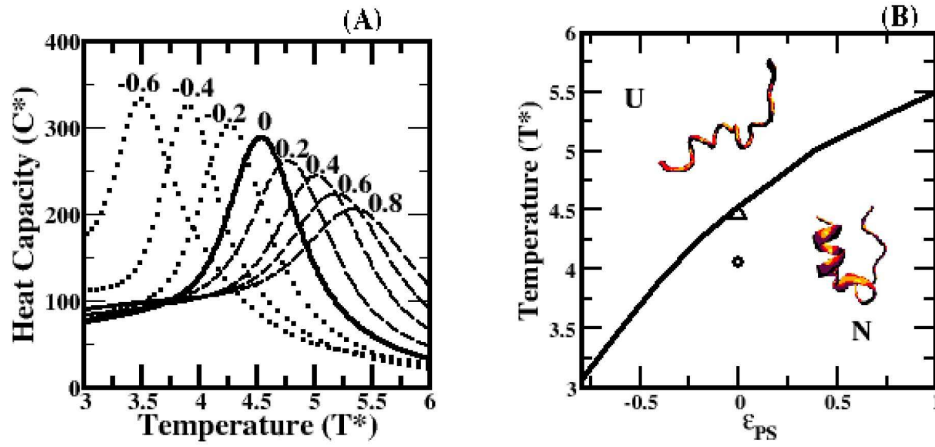


Figure 6.3: (A) C^* versus T^* of Trp-cage in solvent for protein-solvent contact energy $\epsilon_{ps}^* = 0$ (solid line), $\epsilon_{ps}^* = 0.2, 0.4, 0.6, 0.8$ (dashed lines), $\epsilon_{ps}^* = -0.2, -0.4, -0.6$ (dotted lines), $\epsilon_{ss}^* = -1$ for all solvents. (B) Solvent quality temperature (T^*) versus solvent quality (ϵ_{ps}^*) phase diagram. N denotes region where the native is stable, and U, region where the unfolded state is stable. The phase boundary (solid line) is determined by folding temperature T_f^* obtained from histogram analysis of the simulation data. The two symbols in the figure indicate two systems with different solvent conditions: O, location in the phase diagram of the Gō model in implicit solvent; and Δ, location in the phase diagram of the Gō model in hard sphere solvent ($\epsilon_{ps}^* = \epsilon_{ss}^* = 0$).

For each of the solutions plotted in figure 6.3A, the thermal average radius of gyration R_{GY} of the unfolded states with $Q < 0.2$ was recorded at the corresponding folding temperature; the results are plotted in figure 6.4A. The plot clearly shows that unfolded states become more collapsed as the solvent models one containing

osmolyte (i.e. as ϵ_{ps}^* increases to positive values), and more expanded or swollen as the solvent models one containing denaturant (more negative ϵ_{ps}^*). Inset images of figure 6.4A show representative snapshots illustrating that unfolded states become more collapsed as solvent quality decreases (as ϵ_{ps}^* increases). Error bars in the plot are obtained from the standard deviation of the R_{GY} values from simulations of half the total length of those used to obtain the plotted data points. In contrast, folded configurations with $Q > 0.6$ do not show significant variation with solvent quality parameter ϵ_{ps}^* (data not shown). The trend in figure 6.4A is also consistent with experimental evidence that in the presence of osmolytes, unfolded conformations of proteins become more compact, while folded conformations are unaffected ([58–60]).

We also plot the cooperativity of the transition versus ϵ_{ps}^* in figure 6.4B. The cooperativity is defined by the ratio of the van’t Hoff enthalpy over calorimetric enthalpy of the folding transition ([202–205]). We show in figure 6.6 that the internal enthalpy of protein-protein plus protein-solvent interactions is a nearly linear function of Q , hence we use Q here as a proxy for the enthalpy and calculate the various enthalpies as follows. The van’t Hoff enthalpy change corresponds to twice the standard deviation of the enthalpy (or Q) at the transition midpoint, while the calorimetric enthalpy corresponds to the difference in enthalpy of the unfolded and folded states well above and below the transition respectively ([203]). We found these values by taking the average values of Q both well below and well above the transition. The corresponding values of Q_f and Q_u did not vary significantly as ϵ_{ps}^* varied. In contrast, the van’t Hoff enthalpy varies significantly, as can be seen from the insets in figure 6.4B which show histograms of Q at the transition midpoints for $\epsilon_{ps}^* = -0.6$ and $\epsilon_{ps}^* = 0.8$. Denaturant solutions show enhanced two-state behaviour of the transition, with larger cooperativity and corresponding bimodal distribution of Q values at the transition midpoint. Osmolyte solutions show reduced two-state behaviour of the transition, with smaller cooperativity and unimodal (at least for the Trp-cage model) distribution of Q values at the transition midpoint.

6.3.4 Excluded Volume Stabilization Due to Hard Sphere Solvents

Figure 6.5 compares the Trp-cage in implicit solvent (thick solid) with Trp-cage immersed in 1000 (dashed) and 1500 (dashed-dotted) pure hard-sphere solvent (HSS) molecules with $\epsilon_{ps}^* = 0$ and $\epsilon_{ss}^* = 0$, where the hard-sphere radius of a solvent molecule is 1.2\AA , roughly 80% of the vdW radius of a water molecule (see model section). The HSS systems differ from the implicit solvent system: The folding temperature increases from $T_f^* = 4.0$ (implicit solvent) to $T_f^* = 4.45$ (1000 HSS) to $T_f^* = 4.7$ (1500 HSS). This is despite the fact that the interactions between the HSS molecules with themselves and with the protein are purely steric so there is no additional energy scale in the problem. This indicates that protein stabilization is due to an excluded volume effect. The addition of explicit solvent molecules into the $G\overline{7}$ model protein introduces an excluded volume effect that reduces the configuration/conformation space of the protein, to different extents depending on whether the protein is folded or unfolded.

Even for weak urea solution where the protein is weakly attractive to the solvent, such as $\epsilon_{ps}^* = -0.2$ of figure 6.3A, the folding temperature is higher than that of the Trp-cage implicit solvent model. In order to compensate for the stabilizing effects of excluded volume, a urea solution must have $\epsilon_{ps}^* = -0.3$ (and $\epsilon_{ss}^* = -1$), as in the dotted line of figure 6.5, which has the same folding temperature as the implicit-solvent model $T_f^* = 4.0$. This indicates that there is a critical net attractive energy between solute and protein for the solute to function as a denaturant.

Another observation that can be made from figure 6.5 is that the 1000 neutral solvent molecules model (the “reference” solvent with $\epsilon_{ps}^* = 0$ and $\epsilon_{ss}^* = -1$) has a folding temperature of $T_f^* = 4.57$ (light solid line of figure 6.5), which is higher than the 1000 HSS model, suggesting that the attractive solvent-solvent interactions enhance protein stabilization. This can be thought of as a minimal model of the hydrophobic effect.

6.3.5 Calculation of the Free Energy, Enthalpy, and Entropy Changes for Osmolytes and Denaturants

Figure 6.6a plots the free energy versus Q at $T_f^* = 4.8$ for solvents with $\epsilon_{ps}^* = -0.2, 0, 0.4$. For the neutral solvent ($\epsilon_{ps}^* = 0$), where $T_f^* = 4.57$, the U state is stable

as evident from the minimum at $Q_{min} = 0.1$. The osmolyte solvents ($\epsilon_{ps}^* = 0.4$) stabilize the protein so that the free energy minimum is now at $Q_{min} = 0.5$. Adding an osmolytic solvent with effective energy $\epsilon_{ps}^* = 0.4$ stabilizes the native state by about $5k_B T$. The relatively low value of Q is because the folding temperature of the protein-osmolyte system is at $T_f^* = 5.0$ (figure 6.3A), which is only slightly higher than $T_f^* = 4.8$, and also because a protein of such small size as the Trp-cage exhibits substantial fluctuations in the native basin. In the case of denaturing solvents ($\epsilon_{ps}^* = -0.2$) where $T_f^* = 4.2$, the protein is even more destabilized than the neutral solvent with a further change of stability of about $-4k_B T$, and a shift of the free energy towards lower Q values in figure 6.6a.

Figure 6.6b plots the internal thermal energy $\langle E \rangle$ versus Q at $T_f^* = 4.8$ for $\epsilon_{ps}^* = -0.2, 0, 0.4$. This readily shows that the osmolyte solvent decreases the energy of the protein uniformly (even though the solvent-protein interaction is repulsive), while the effects of the urea solvents are negligible (even though the solvent-protein interaction is attractive). An osmolytic solvent with repulsive ϵ_{ps}^* lowers the energy by inducing collapse in the protein in a manner that 1L2Y is not sensitive to Q over the range spanning the unfolded and folded states. Hence the stabilization of protein in our model is not due to the change in enthalpy since the osmolyte solvents decrease the energy of the native (high Q) and unfolded (low Q) essentially equally (figure 6.6e).

Figure 6.6d plots the entropy S versus Q , which shows a decrease/increase in the entropy of the unfolded protein for the osmolyte/urea solution, compared to the neutral solvent. The entropy difference between osmolyte/urea solvents and neutral solvent is plotted in figure 6.6d, that is, this is the change in entropy $\Delta S = S(\epsilon_{ps}^* = 0.4) - S(\epsilon_{ps}^* = 0)$ and $\Delta S = S(\epsilon_{ps}^* = -0.2) - S(\epsilon_{ps}^* = 0)$ versus Q . For the osmolyte solvents ($\epsilon_{ps}^* = 0.4$) the decrease in entropy, compare to the neutral solvents ($\epsilon_{ps}^* = 0$), is due to the excluded volume effects, which is expected to reduce the protein conformational/spatial configurations. The key point is that the reduction in entropy decreases as Q decreases (see dashed line of figure 6.6d), which means that the entropy of unfolded conformations is reduced more than folded conformations. This is consistent with the hypothesis that osmolytes stabilize proteins by introducing an entropic bias for unfolded conformations.

For the urea solvents ($\epsilon_{ps}^* = -0.2$) the increase in entropy, compare to the neu-

tral solvents ($\epsilon_{ps}^* = 0$), increases as Q decreases (dotted line of figure 6.6d), which means that the entropy of unfolded conformations is increased more than folded conformations. This suggests that urea destabilization of protein is also driven by entropy change. In fact, the denaturing solvent raised the enthalpy of the unfolded state relative to the neutral solvent, rather than lowering it. This is due to the fact that as favourable solvent-protein contacts are made, favourable protein-protein contacts are lost, so that the net result is a modest stabilization of the native state change due to enthalpy changes (figure 6.6e). The dominant factor contributing to the destabilization of the native state for weak denaturing solvents is the increase in entropy due to an expanded unfolded state. For sufficiently strong denaturants (ϵ_{ps}^* large and negative) we expect a crossover to a regime where enthalpic stabilization of the unfolded state becomes significant.

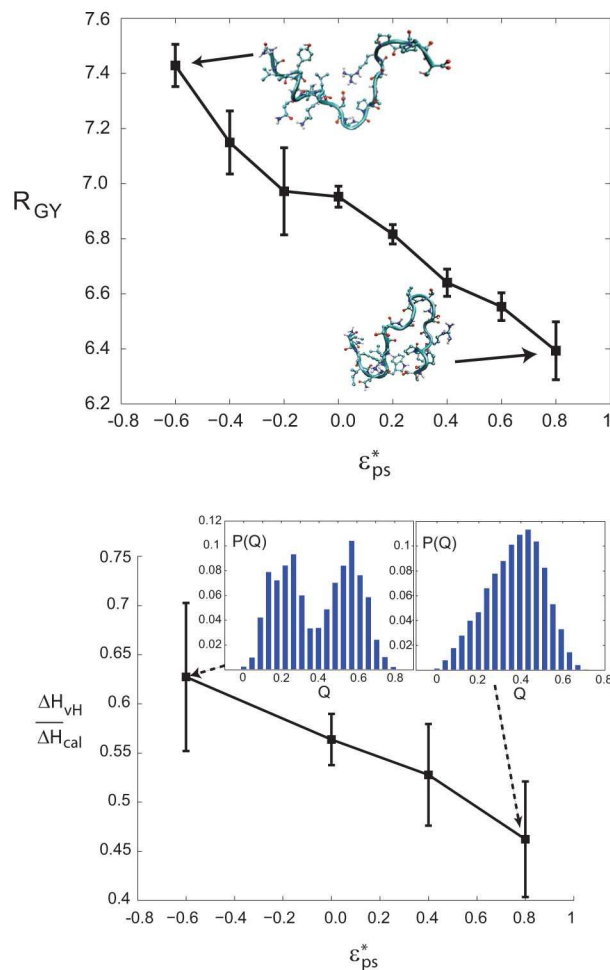


Figure 6.4: (A) Radius of gyration of the unfolded states with $Q < 0.2$ taken at the temperatures of the heat capacity peaks in figure 6.3, plotted as a function of the protein-solvent interaction energy ϵ_{ps}^* . Snapshots of representative unfolded states for $\epsilon_{ps}^* = -0.6$ and $\epsilon_{ps}^* = 0.8$ are shown in the insets. These snapshots are obtained by taking the first sampled conformation that had a R_{GY} within 2% of the average value given by the plotted data point. (B) Cooperativity of the folding transition, defined by the ratio of the van't Hoff enthalpy over calorimetric enthalpy, as a function of ϵ_{ps}^* . Histograms of the values of Q at the midpoints of the transition for $\epsilon_{ps}^* = -0.6$ and $\epsilon_{ps}^* = 0.8$ are shown in the insets, which are strongly bimodal for a denaturant-containing solvent, and unimodal (for the Trp-cage model) for a strong osmolyte-containing solvent.

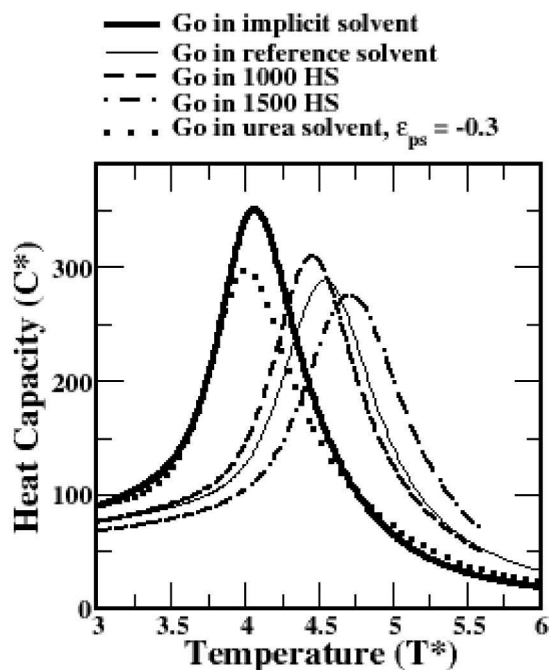


Figure 6.5: C^* versus T^* of plots Trp-cage for several solvent models: In implicit solvent (thick solid line); in 1000 spherical reference solvent molecules with $\epsilon_{ps}^* = 0, \epsilon_{ss}^* = -1$ (thin solid line, this curve is identical to the $\epsilon_{ps}^* = 0$ curve in figure 6.3A); in 1000 pure hard-sphere spherical solvents with $\epsilon_{ps}^* = \epsilon_{ss}^* = 0$ (dashed line); in 1500 pure hard-sphere spherical solvents $\epsilon_{ps}^* = \epsilon_{ss}^* = 0$ (dashed-dotted line); in 1000 urea-like spherical solvents with $\epsilon_{ps}^* = -0.3, \epsilon_{ss}^* = -1$ (dotted line).

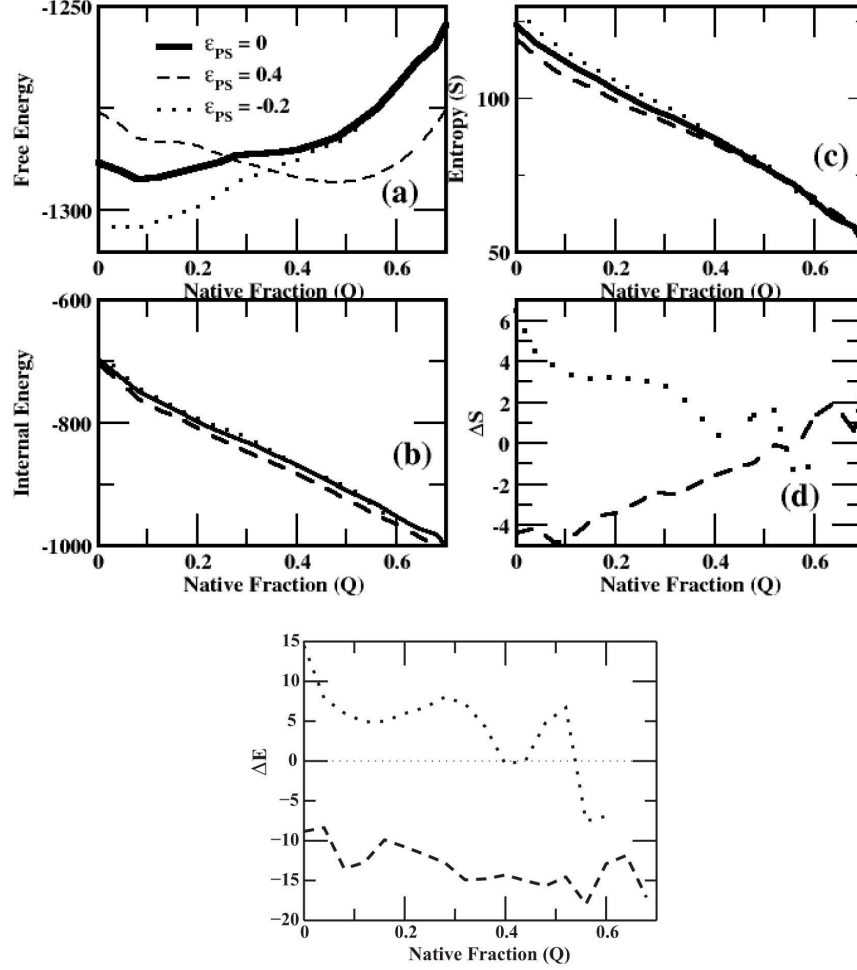


Figure 6.6: (a) Free energy versus protein foldedness (Q) at $T_f^* = 4.8$ for $\epsilon_{ps}^* = 0$ (solid line), $\epsilon_{ps}^* = 0.4$ (dashed line), and $\epsilon_{ps}^* = -0.2$ (dotted line), with solvent-solvent contact energy fixed at $\epsilon_{ss}^* = 0$ (solid line) (b) Protein internal energy $\langle E \rangle$ versus Q (c) Entropy (S) versus Q (d) Change in entropy (ΔS) versus Q , for osmolyte solvent ($\epsilon_{ps}^* = 0.4$) compared to neutral solvent ($\epsilon_{ps}^* = 0$), that is, $\Delta S = S(\epsilon_{ps}^* = 0.4) - S(\epsilon_{ps}^* = 0)$ (dashed line), and for urea ($\epsilon_{ps}^* = -0.2$) solvent compare to neutral ($\epsilon_{ps}^* = 0$) solvent $\Delta S = S(\epsilon_{ps}^* = -0.2) - S(\epsilon_{ps}^* = 0)$ (dotted line) (e) Changes in enthalpy between osmolyte and neutral solvent $\Delta E = E(\epsilon_{ps}^* = 0.4) - E(\epsilon_{ps}^* = 0)$ (dashed line), and between denaturant solvent and neutral solvent $\Delta E = E(\epsilon_{ps}^* = -0.2) - E(\epsilon_{ps}^* = 0)$ (dotted line).

Chapter 7

Conclusion

Our understanding of protein folding has been greatly acquired from studies under ideal conditions in dilute solutions. However, potentially macromolecular crowding inside the cell can significantly affect the thermodynamic stability and alter the folding/unfolding kinetics of proteins. A growing number of experimental, theoretical, and computational investigations on the consequences and magnitudes of macromolecular crowding predict significant stabilization [66, 67, 206]. This implied that proteins perform their functions in environments that are crowded with various macromolecules and macromolecular assemblies, as visualized recently by high-resolution cryoelectron tomography [207] and other techniques. The goal of the present work was to understand the effects of macromolecular crowding on the mechanical stability and unfolding kinetics of proteins at the single-molecule level using analytical calculations and computer simulations. We hope that this work can encourage researchers to consider the effect of crowdedness in their experiments and use proper approximations in order to avoid presenting the data and results that might be reflecting a picture that is either far from or the opposite of reality, as we have discussed in some of the examples in chapter 2.

Essentially all taxa utilize osmolytes which are ubiquitous, small organic molecules to cope with environmental, extracellular, or intracellular stress. The origin of the native protein stabilization by crowding is the steric repulsion between the amino acids of the protein and different kinds of osmolytes. The entropy induced depletion interaction, which is responsible for all the energetic consequences, was first

described by Asakura and Oosawa in 1954 [52] and later by Vrij [53]. The theory implies that there is an osmotic pressure pushing hard spheres toward each other when they are close enough. Moreover, the instant effect of the high packing fraction of macromolecules inside a living cell, which on the average is about 20% to 30%, implicates that a large fraction of the interior volume is excluded to anything whose size is comparable to cell constituent molecules, and this in turn, means that any reactions that are accompanied by dramatic change in volume accessible will be significantly affected in dense systems. The collapse of a heteropolymer and folding of proteins are partly driven by this effect. Furthermore, the thermodynamic effects of crowding agents vary from reduction of the diffusion rate to hindering degradative reactions [3].

The stabilizing property of osmolytes has been shown to correlate with the preferential exclusion of osmolytes from unfolded protein domains, resulting in the preferential accumulation of water (preferential hydration) near an unfolded protein [183, 184]. This implies a net repulsive interaction between stabilizing osmolytes and protein, and indeed preferential exclusion has been shown to arise from repulsive interactions between osmolytes and the backbone of proteins [49, 185, 186]. Repulsive osmolyte-backbone interactions would raise the enthalpy of a protein, and the increase in enthalpy would be larger for the unfolded state due to its larger solvent exposed backbone area and consequently the unfolded state would be more destabilized. Another possible stabilization mechanism is the osmolyte-induced loss of protein conformational entropy, with the greater entropic loss by the unfolded state, leading to an overall shift in equilibrium towards the native state. The entropy loss mechanism is consistent with experimental works that observed increased compactness in unfolded states of cutinase [58], protein S6 [59], and Rnase S [60] due to osmolytes. This would imply that even if an osmolyte interacts with a protein with the same energetics as water, it would still stabilize the protein for entropic reasons. In reference [60], which examines the thermal and chemical stabilization of Rnase S by osmolytes, protein stabilization by an increase in enthalpy that destabilizes the unfolded state is ruled out.

In contrast to protecting osmolytes, urea is a nonprotecting osmolyte. The interaction between urea and the protein is attractive leading to the preferential accumulation of urea in the vicinity of proteins [187]. The attractive urea-protein

interactions lower the free energy of both the native and unfolded states, but due to its larger solvent exposed surface area, the free energy of the unfolded state is lowered more compare to folded state [49]. Consequently, the addition of urea to protein solutions shifts the equilibrium to the unfolded state. The attractive interactions between urea and protein must overcome the entropic stabilization of the protein mentioned above.

Moreover, macromolecular crowding restricts the folded protein and the unfolded protein to different extents because the unfolded chain can penetrate into the interstitial spaces. When the concentration of crowding macromolecules is high, the interstitial voids are too small to serve as routes of translocation for the globular native protein. However, in the same osmolyte concentrations, the interstitial spaces still will allow the unfolded chain to leak. This is similar to confining the protein in a cage with holes. Although the native protein is fully restricted but the holes will compensate for the excess restriction of the cage on the unfolded chain.

To investigate the effect of molecular crowding on polymer collapse and protein folding, we need to formulate the thermodynamics of a polymer (or protein) in terms of its size. The simplest model for a protein is an ideal chain [118]. However, this model provides an unrealistic picture of proteins because there is no steric repulsion between monomers of the chain to avoid self-crossing. The intramolecular steric repulsion is usually accounted for by introducing an empirical factor that causes expansion of the chain [124]. We showed that using computer simulation algorithms we can calculate the conformational statistics of both on-lattice and off-lattice self-avoiding walks [208–210]. Goldenberg has done off-lattice Monte Carlo simulations which gave the radius of gyration probability distributions of four model protein chains, accounting for local steric interactions between adjacent amino acid residues by restricting angles [211]. Goldenberg has performed two sets of calculations, in which the long-range steric interactions between non-adjacent residues were ignored in one set, and in the other set they were present. The result of the calculations showed that when long-range steric interaction was present, for all four proteins the dependence of the root mean-square radius of gyration on the chain length was in good agreement with experimental data obtained for a large set of denatured proteins.

Real polymers can be modeled as off-lattice self-avoiding random walks, which

allow the angle between three consecutive monomers to have any value consistent with steric volume constraints. We have created a MATLAB pivot algorithm to generate off-lattice SAW conformations by the well-known pivot algorithm, to get around the attrition problem which corresponds to the fact that an absorbing boundary should be used for a polymer to avoid a sterically-excluded region. Any generated conformation which penetrates into the boundary must then be removed from the statistics. Thus in generating conformations of a self-avoiding polymer, any conformation of the polymer that by chance wanders into the sterically excluded region corresponding to the previously generated monomer positions must be eliminated, and the walk must be re-initiated.

To see the effects of a reflecting boundary condition on SAW statistics, we have also generated random walks using a naive growth algorithm that corresponds to a reflecting boundary condition. Walks are generated with the $i + 1$ th residue placed a distance ℓ from the i th residue at random angle; and if the distance $r_{i+1,j}$ for any $j < i$ is less than 2σ , only the last step is canceled and a new step is attempted, until a walk of N steps is generated. As we expected there is an obvious distinction between the end to end distance probability distribution of an off-lattice SAW using the naive growth algorithm and that using the pivot algorithm, and the distinction become more and more clear as N increases.

Moreover, we noticed that the difference between the probability distribution functions of an on-lattice SAW and its improved replica, the off-lattice SAW model (which is widely used in the literature), are not negligible. Therefore we were inspired to further look for a method that can generate a more realistic model of a polymer which led to employing DMD simulations in our work. In our model the polymer is a freely jointed chain of N beads or monomers and $N - 1$ joints, wherein each monomer is represented as a hard sphere. Two bonded beads i and j are constrained to be within 10% of an average distance, $\ell = 1\text{\AA}$, by an infinite square-well potential. Two non-bonded atoms i and j may interact by hard-sphere potential (purely repulsive). As expected, the DMD simulations reproduced the same statistics (end to end distance probability distribution) as that of the pivot algorithm for a continuum SAW.

We have introduced v_{pb} which is the volume of the smallest box along the principal axes of the polymer in each conformation, as the new measure of the size of

a polymer. We showed that v_{pb} or equivalently $r_{pb} = (\frac{3v_{pb}}{4\pi})^{1/3}$ provides a better measure of the size of a polymer chain. A comparison of the effective diameter probability distributions using different models of the size of a 100-mer, that is $d_{eff} = 2R_{eff}$ for the end to end distance model, $d_{eff} = 2R$ for the embedding sphere model and $d_{eff} = (\frac{6v}{\pi})^{1/3}$ for the Cartesian box model and the principal box model, shows that: 1) for many conformations the end to end distance description does not accurately represent the size statistics of a polymer, since the end to end distance of a polymer can be zero, whereas even in the most collapsed conformation the real size of a polymer cannot be smaller than $\approx \ell N^{1/3}$, 2) the embedding sphere and Cartesian box models overestimate the size of the polymer and therefore cannot be considered as a good measure of the size of the polymer. Moreover, we have calculated the probability distribution of the effective diameter of the gyration tensor volume which is $d_{eff} = 2(\sqrt{e_1^2 + e_2^2 + e_3^2})$ where e_1 , e_2 and e_3 are the eigenvalues of the gyration tensor of the polymer. We noticed that numerous conformations of the polymer have larger principal box diameter than the effective gyration tensor diameter, which means that the gyration tensor model underestimates the volume of the polymer.

In terms of this new variable, v_{pb} , we derived expressions for the entropy by finding the fits to the data obtained from DMD simulations. Furthermore, using the same method we found an expression that describes the number of close approaches of a polymer chain (a close approach happens when two non-neighbor monomers come closer to each other than a certain cut off distance) and we used this expression to find the internal energy of a polymer. The polymer free energy is then simply given by $F = E - TS$. The effect of neutral osmolytes were studied by adding the free energy of a hard sphere liquid system to the free energy of the polymer. The generalization to a protein is easily done by introducing energetic parameters that account for the primary, secondary and tertiary structures of the protein. The stabilization of a collapsed polymer or a folded protein by osmolytes depends on the size and concentration, as well as the type of the osmolyte which is represented by ϵ_{po} . At the same overall volume fraction ϕ , smaller osmolytes are more effective at stabilizing proteins.

In chapter 3 we discussed examples of different approaches to the problem of polymer (protein) and osmolyte mixtures which have provided powerful theories

to explain both thermodynamics and kinetics of the crowding effects on protein folding. However, what is commonly missing in these approaches is a model for the internal structure of the protein, that is, the protein as a whole is considered to be either a globular object or a chain, and often the model is not flexible to the changes in the energetic parameters between residues. In the model that we presented here, by changing the energetic parameters of the primary, secondary and tertiary structure of the protein one can attain a more detailed description of a specific protein. The three distinct aspects of a protein's structure are as the following:

- *Primary structure*: the amino acid sequence.
- *Secondary structure*: general three-dimensional form of local segments of the protein that are stabilized by hydrogen bonds. Common examples are the alpha helix, beta sheet and turns. Secondary structure does not, however, describe specific atomic coordinates in three-dimensional space.
- *Tertiary structure*: the overall 3-D shape of a single protein molecule; the spatial order of the secondary structures. The stabilization of tertiary structure is generally done by nonlocal interactions, such as the formation of a hydrophobic core, salt bridges, hydrogen bonds and disulfide bonds. The tertiary structure is what controls the function of the protein and therefore we can use the term “tertiary structure” and the word “fold” interchangeably.

Using the appropriate size of the protein molecule, and equipped with energetic parameters of the protein structure, our theoretical results provide a thorough description of the thermodynamics of the polymers and proteins that allows us to readily investigate the effect of osmolytes on their behavior. Furthermore, in this framework a general description of protecting and denaturing (urea) osmolytes has been provided by changing the interaction energy between the polymer and the osmolytes from negative to positive values. Essentially the same thermodynamic approach can be applied to theories of protein stabilization due to macromolecular protecting crowding agents [50, 51], and protein destabilization due to the present

of denaturants. Therefore, this model enables us to unify the effects of protecting osmolytes and denaturants in the same theoretical scheme.

In chapter 6 an all-atom $G\bar{o}$ simulation of Trp-cage protein was discussed employing discontinuous molecular dynamics simulations in an explicit minimal solvent, using a single, contact-based interaction energy between protein and solvent particles. An effective denaturant or osmolyte solution was constructed by making the interaction energy attractive or repulsive. Analysis of the results confirms the main findings of our theoretical work, which can be summarized as follows: 1) The comparison between figures 5.2 and 5.3 shows that by adding osmolytes to the system, the free energy minimum of the protein shifts to high values of Z (the fraction of total contacts per monomer in any conformation) and q (the fraction of native contacts per monomer in any conformation), and the folded state of the protein becomes the stable state. These osmolytes are neutral because the interaction energy between protein and osmolytes, ϵ_{po} , has considered to be zero in calculations, however it is argued in this work that negative values of ϵ_{po} strengthen the effect of neutral osmolytes and positive values of ϵ_{po} weaken the stability effect of neutral osmolytes. Figure 6.6a is the result of simulations and plots the free energy of the protein versus foldedness. We can see a comparison between the effect of neutral osmolytes ($\epsilon_{po} = 0$), protecting osmolytes ($\epsilon_{po} > 0$) and denaturants ($\epsilon_{po} < 0$). As the theory predicts, the protecting osmolyte solvents stabilize the protein so that the free energy minimum is at higher value than that for neutral osmolytes. Adding an osmolytic solvent with effective interaction energy of 0.4 stabilizes the native state by about $5k_B T$. In the case of denaturing solvents however, the protein is more destabilized than the neutral solvent with a further change of stability of about $-4k_B T$, and there is a shift of the free energy towards lower foldedness values, 2) The reduction in entropy of the unfolded state due to osmolytes results in a loss of folding cooperativity. This can be seen by investigating the heat capacity as a function of temperature as osmophobicity of the solvent, ϵ_{po} , is varied. There is a loss in folding cooperativity when osmolytes are stabilizing the native state, again due to the reduction in conformational entropy of the unfolded state. Conversely, there is an increase in folding cooperativity in the presence of denaturant due to swelling of the unfolded protein. Less strongly peaked and larger width of heat capacity indicates less cooperativity, because denaturant solutions show enhanced two-state

behaviour of the transition, with larger cooperativity and corresponding bimodal distribution of Q values at the transition midpoint. Osmolyte solutions show reduced two-state behaviour of the transition, with smaller cooperativity. We showed the heat capacity plot of a protein with 50 residues in a box of crowding agents with packing fraction 30% and $\epsilon_{po} = -0.2, 0, 0.2$ in figure 5.5 which was derived using theoretical results (equation (5.13)). Moreover, figure 6.3A showed the simulation results of the heat capacity of Trp-cage $G\bar{O}$ model immersed in 1000 spherical solvent molecules with hard-sphere radius $r_{s,hs} = 1.5\text{\AA}$ confined in a $(40\text{\AA})^3$ periodic box as a function of temperature, in the presence of osmolytes with different osmophobicity values. The increase in the folding temperature as compared to the implicit solvent model despite zero solvent-protein contact energy suggests that the change in thermodynamic property is due to excluded volume effects. When the interaction energy between protein and solvent, ϵ_{ps}^* , is zero there is no energy scale in the problem. For the osmolyte solvents (poor solvent), in which solvent molecules are repulsive to the protein, the folding temperature increases progressively and the heat capacity peak decreases progressively. The addition of repulsive osmolyte solvents (bad solvents) stabilizes the Trp-cage, since the shift of the heat capacity peak indicates the native state is stable up to a higher temperature. For urea-like solvents (good solvent), in which solvent molecules are attracted to the protein ($\epsilon_{ps}^* < 0$), the folding temperature decreases progressively. The attractive urea-solvent interactions destabilize the native structure so that the Trp-cage unfolds at lower temperatures than for the reference (water-like) solvent with $\epsilon_{ps}^* = 0$.

We also discussed the correspondence between explicit and implicit osmolyte models, to be able to compare our analytical results with simulation results, because in our theory there could be in principle 6 relevant interaction enthalpies corresponding to the pair interactions between the above species: ϵ_{ss} , ϵ_{pp} , ϵ_{oo} , ϵ_{po} , ϵ_{ps} , ϵ_{so} (explicit osmolyte model), whereas in the simulations a solution of water and osmolyte may be considered as an effective medium with modified solvent interactions so that the new system can be treated with just 3 energy scales: ϵ_{ss}^* , ϵ_{pp}^* , ϵ_{ps}^* (implicit osmolyte model).

At last, it is worth mentioning that one can use similar ideas that were used in this work to explain the effect of confinement on protein stability, because a macromolecular confiner is essentially a stabilizing osmolyte of larger effective

size. In fact, the physical origins of protein or polymer stabilization by steric osmolytes or crowders are essentially the same as those leading to phase separation in colloidal suspensions due to the addition of a non-adsorbing polymer for example [52–54]. The next step of generalizing our model is to change the effective size of the osmolytes to dimensions comparable to the size of confiner molecules and to investigate the difference of the free energy of the protein.

The confinement effect has significant biological applications, e.g. only a few proteins can fold unassisted into their unique 3-D structure that enables them to serve the purpose for which they are evolutionary designed, through the chemical properties of their amino acids. However; the majority of protein molecules require the aid of molecular chaperones that assist the non-covalent folding or assembly of macromolecular structure by confining them to fold into their native states. While both confinement and crowding effects are caused by reduction in the number of configurations that are available to a macromolecule because of the high-volume fraction of osmolytes or static barriers to movement, there is an important qualitative difference between these two phenomena. Although the free energy cost of crowding is minimum for protein conformations that are globally the most compact with the smallest radius of gyration, confinement favors conformations that have a complementary shape to the shape of the confining volume. For example, in a quasispherical cavity the globular conformation of the protein is favored, however; in a cylindrical pore the preferred conformation is rod-like, and when the protein is confined by two parallel hard walls, plate-like conformations are preferred. Therefore, numerical calculations of the magnitude of confinement effects depend both on the shape of the confining space and confined macromolecular species [212].

Other natural extensions of our theory are:

- studying the effect of charged osmolytes on protein stability.
- analyzing the effect of ellipsoidal crowding agents instead of spherical osmolytes (which could resemble peptides in *in vivo* experiments). This study requires calculating the difference between the excluded volume for various orientations of ellipsoidal osmolytes which can lead to their alignment in crystal phase.

- investigating the role of depletion interaction in aggregation resulted from protein-protein interactions and signal transduction.
- examining the dewetting transition and drying induced collapse and its effects on protein folding by changing the interaction energy between protein and osmolytes, ϵ_{po} , as well as the structural energetic parameters of the protein.

Bibliography

- [1] A. P. Minton. *Curr. Opin. Struct. Biol.*, 10:34, 2000. → pages ii
- [2] A. P. Minton. *Curr. Biol.*, 10:97, 2004. → pages
- [3] R. J. Ellis. *Curr. Opin. Struct. Biol.*, 11:114, 2001. → pages 99
- [4] R. J. Ellis and F. U. Hartl. *Curr. Opin. Struct. Biol.*, 9:102, 1999. → pages
- [5] A. P. Minton. *Biophys. J.*, 78:101, 2000. → pages ii
- [6] S. Hadizadeh, A. Linhananta, and S. S. Plotkin. *in preparation*, 2010. → pages iii, 47, 69
- [7] A. Linhananta, S. Hadizadeh, and S. S. Plotkin. *In press*, 2010. → pages iii, 13, 79
- [8] P. H. Yancey. *Biologist*, 50:126, 2004. → pages 2, 4
- [9] P. H. Yancey. *American Zoologist*, 41:699, 2001. → pages 2
- [10] P. H. Yancey. *J. Exp. Biol.*, 208:2819, 2005. → pages 2
- [11] P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, and G. N. Somero. *Science*, 217:1214, 1982. → pages 2
- [12] P. Debye and E. Huckel. *Physikalische Zeitschrift*, 24:185, 1923. → pages 4
- [13] R. B. Simpson and W. Kauzmann. *J. Am. Chem. Soc.*, 75:5139, 1953. → pages 4, 6
- [14] C. Tanford. *J. Amer. Chem. Soci.*, 86:2050, 1964. → pages 4, 9, 89
- [15] C. Tanford. *Adv. Protein Chem.*, 24:1, 1970. → pages 4

- [16] J. C. Lee and S. N. Timasheff. *J. Biol. Chem.*, 256:7193, 1981. → pages 5
- [17] K. Gekko and S. N. Timasheff. *Biochemistry*, 20:4667, 1981. → pages 5, 8
- [18] E. F. Casassa and H. Eisenberg. *Adv. Protein Chem.*, 9:287, 1964. → pages 5
- [19] J. G. Kirkwood and R. J. Goldberg. *J. Chem. Phys.*, 18:54, 1950. → pages
- [20] W. H. Stockmayer. *J. Chem. Phys.*, 18:1600, 1950. → pages 5
- [21] T. Arakawa, R. Bhat, and S. N. Timasheff. *Biochemistry*, 29:1914, 1990. → pages 5
- [22] T. Arakawa, R. Bhat, and S. N. Timasheff. *Biochemistry*, 29:1924, 1990. → pages
- [23] S. N. Timasheff and T. Arakawa. *In Protein Structure and Function: A Practical Approach*, ed T. E. Creighton. Oxford; IRL Press, 1988. → pages 5
- [24] E. P. K. Hade and C. Tanford. *J. Am. Chem. Soc.*, 89:5034, 1967. → pages 5
- [25] H. Inoue and S. N. Timasheff. *J. Am. Chem. Soc.*, 90:1890, 1968. → pages
- [26] M. E. Noelken and S. N. Timasheff. *J. Biol. Chem.*, 242:5080, 1967. → pages 5
- [27] G. Scatchard. *J. Am. Chem. Soc.*, 68:2315, 1946. → pages 6
- [28] T. Arakawa and S. N. Timasheff. *Biochemistry*, 24:6756, 1985. → pages 7, 9
- [29] R. Bhat and S. N. Timasheff. *Protein Sci.*, 1:1133, 1992. → pages 7
- [30] J. W. Gibbs. *Trans. Conn. Acad.*, 3:343, 1878. → pages 7
- [31] K. Gekko and T. Morikawa. *J. Biochem.*, 90:51, 1981. → pages 8
- [32] D. T. Warner. *Nature*, 196:1055, 1962. → pages 8
- [33] E. P. Pittz and S. N. Timasheff. *Biochemistry*, 17:615, 1978. → pages 8
- [34] M. V. King, B. S. Magdoff, M. B. Adelman, and D. Harker. *Acta Crystallogr.*, 9:460, 1956. → pages 8

- [35] E. P. Pittz and J. Bello. *Arch. Biochem. Biophys.*, 146:513, 1971. → pages 9
- [36] C. J. Lee and L. L. Y. Lee. *Biochemistry*, 18:5518, 1979. → pages 9
- [37] C. J. Lee and L. L. Y. Lee. *J. Biol. Chem.*, 256:625, 1981. → pages 9
- [38] G. G. Hammes and P. R. Schimmel. *J. Am. Chem. Soc.*, 89:442, 1967. → pages 9
- [39] A. Jr. McPherson. *J. Biol. Chem.*, 251:6300, 1976. → pages 9
- [40] A. Jr. McPherson. *Meth. Enzym.*, 114:120, 1985. → pages 9
- [41] B. R. Somalinga and R. P. Roy. *J. Biol. Chem.*, 277:43253, 2002. → pages 9, 15
- [42] L. A. Munishkina, E. M. Cooper, V. N. Uversky, and A. L. Fink. *J. Mol. Recognit.*, 17:456, 2004. → pages 16
- [43] C. Andries and J. Clauwaert. *Biophys. J.*, 47:591, 1985. → pages 17, 18
- [44] E. Doss-Pepe. *Experimental Eye Research*, 67:657, 1998. → pages 17
- [45] J. M. Rohwer, P. W. Postma, B. N. Kholodenko, and H. V. Westerhoff. *Proc. Natl. Acad. Sci.*, 95:10547, 1998. → pages 21
- [46] J. M. Gonzalez, M. Jimenez, M. Velez, J. Mingorance, J. M. Andreu, M. Vicente, and G. Rivas. *J. Biol. Chem.*, 278:37664, 2003. → pages 19
- [47] S. Cayley and M. T. Record. *Biochemistry*, 42:12596, 2003. → pages 9, 26
- [48] P. L. Altman. *Blood and Other Body Fluids*. Washington, DC: Faseb, 1961. → pages 9
- [49] D. W. Bolen and G. D. Rose. *Annu. Rev. Biochem.*, 77:339, 2008. → pages 9, 74, 75, 99, 100
- [50] H. X. Zhou, G. N. Rivas, and A. P. Minton. *Annual Review of Biophysics*, 37:375, 2008. → pages 10, 12, 103
- [51] H. X. Zhou. *Proteins-Structure Function and Bioinformatics*, 72:1109, 2008. → pages 10, 12, 103
- [52] S. Asakura and F. Oosawa. *J. Chem. Phys.*, 22:1255, 1954. → pages 10, 99, 106

- [53] A. Vrij. *Pure. Appl. Chem.*, 48:471, 1976. → pages 99
- [54] H. N. W. Lekkerkerker, W. C. K. Poon, P. N. Pusey, A. Stroobants, and P. B. Warren. *Europhys. Lett.*, 20:559, 1992. → pages 10, 106
- [55] T. O. Street, D. W. Bolen, and G. D. Rose. *Proc. Natl. Acad. Sci. USA*, 103:13997, 2006. → pages 10
- [56] J. Rosgen, B. M. Pettitt, and D. W. Bolen. *Biophys. J.*, 89:2988, 2005. → pages 10, 11, 27, 37
- [57] E. O'Brien, G. Zuv, G. Haran, B. R. Brooks, and D. Thirulamai. *Proc. Natl. Acad. Sci. USA*, 105:13403, 2008. → pages 10
- [58] R. P. Baptista, S. Pedersen, G. J. M. Cabrita, D. E. Otzen, J. M. S. Cabral, and E. P. Melo. *Biopolymers*, 89:538, 2008. → pages 10, 26, 91, 99
- [59] L. Y. Chen, G. J. M. Cabrita, D. E. Otzen, and E. P. Melo. *J. Mol. Biol.*, 351:402, 2005. → pages 26, 99
- [60] G. S. Ratnaparkhi and R. Varadarajan. *J. Biophys. Chem.*, 276:28789, 2001. → pages 10, 26, 91, 99
- [61] A. R. Kinjo and S. Takada. *Phys. Rev. E*, 66:031911, 2002. → pages 11, 27, 63
- [62] A. R. Kinjo and S. Takada. *Phys. Rev. E*, 66:051902, 2002. → pages 11, 27, 63
- [63] H. X. Zhou. *J. Mol. Recognit.*, 17:368, 2004. → pages 11, 27, 31
- [64] J. Rosgen, B. M. Pettitt, and D. W. Bolen. *Protein Sci.*, 16:733, 2007. → pages 11, 27, 37
- [65] A. C. M. Ferreón, J. C. Ferreón, D. W. Bolen, and J. Rosgen. *Biophys. J.*, 92:245, 2007. → pages 11, 27, 37
- [66] A. P. Minton. *Biophys. J.*, 88:971, 2005. → pages 11, 27, 42, 63, 98
- [67] K. Sasahara, P. McPhie, and A. P. Minton. *J. Mol. Biol.*, 326:1227, 2003. → pages 14, 98
- [68] N. Tokuriki, M. Kinjo, S. Negi, M. Hoshino, Y. Goto, I. Urabe, and T. Yomo. *Protein Sci.*, 13:125, 2004. → pages 14

- [69] A. Baumgartner and M. Muthukumar. *J. Chem. Phys.*, 87:3082, 1987. → pages 14
- [70] J. D. Honeycutt and D. Thirumalai. *J. Chem. Phys.*, 90:4542, 1989. → pages 14
- [71] J. Dayantis, M. J. M. Abadie, and M. R. L. Abadie. *Comp. Theor. Polym. Sci.*, 8:273, 1998. → pages 14
- [72] G. B. Benedek. *Appl. Opt.*, 10:459, 1971. → pages 17
- [73] E. Bi and J. Lutkenhaus. *Nature*, 354:161, 1991. → pages 18
- [74] S. Vitha, R. S. McAndrew, and K. W. Osteryoung. *J. Cell Biol.*, 153:111, 2001. → pages 18
- [75] A. P. Minton. *Mol. Cell. Biochem.*, 55:119, 1983. → pages 20, 69
- [76] Z. Jiang, H. Huawei, and L. Sen. *Tsinghua Science and Technology*, 13:454, 2008. → pages 21
- [77] J. W. Kelly. *Curr. Opin. Struct. Biol.*, 8:101, 1998. → pages 22
- [78] D. Thirumalai, D. K. Klimov, and R. I. Dima. *Curr. Opin. Struct. Biol.*, 13:146, 2003. → pages 22
- [79] L. C. Serpell. *Biochim. Biophys. Acta*, 1502:16, 2000. → pages 23
- [80] X. Li and E. L. Mehler. *Biopolymers*, 30:177, 1990. → pages 23, 69
- [81] J. Lanman, J. Sexton, M. Sakalian, and P. E. Prevelgie. *J. Virol.*, 76:6900, 2002. → pages 24
- [82] M. del Alamo, G. Rivas, and M. G. Mateu. *J. Virol.*, 79:14271, 2005. → pages 24
- [83] K. Zheng, C. Zhao, Y. Hao, and Z. Tan. *Nucleic Acids Research*, 38:327, 2010. → pages 25
- [84] T. Ghosh, A. Kalra, and S. Grade. *J. Phys. Chem.*, 109:642, 2004. → pages 26
- [85] B. C. McNulty, G. B. Young, and G. J. Pielak. *J. Mol. Biol.*, 355:893, 2006. → pages 26
- [86] H. A. Kramers. *Physica*, 7:284, 1940. → pages 32

- [87] A. Szabo, K. Schulten, and Z. Schulten. *J. Chem. Phys.*, 72:4350, 1980. → pages 32
- [88] D. Shoup and A. Szabo. *Biophys. J.*, 40:33, 1982. → pages 32
- [89] M. V. Smoluchowski. *Phys. Chem.*, 92:129, 1917. → pages 33
- [90] P. Debye. *Trans. Am. Electrochem. Soc. Sci.*, 82:265, 1942. → pages 33
- [91] H. X. Zhou. *Biophys. J.*, 73:2441, 1997. → pages 33
- [92] E. F. Casassa. *J. Poly. Sci. B*, 5:773, 1967. → pages 34
- [93] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, 1986. → pages 34
- [94] T. M. Nieuwenhuizen. *Phys. Rev. Lett.*, 62:357, 1989. → pages 35
- [95] G. T. Barkema, P. Biswas, and H. van Beijeren. *Phys. Rev. Lett.*, 87:170601, 2001. → pages 35
- [96] J. L. Lebowitz and J. S. Rowlinson. *J. Chem. Phys.*, 41:133, 1964. → pages 35
- [97] S. Qin and H. X. Zhou. *Phys. Rev. E*, 81:031919, 2010. → pages 36
- [98] Y. Rosenfeld. *Phys. Rev. Lett.*, 63:980, 1989. → pages 36
- [99] S. M. Oversteegen and R. Roth. *J. Chem. Phys.*, 122:214502, 2005. → pages 36, 37
- [100] J. Wu and Z. Li. *Annu. Rev. Phys. Chem.*, 58:85, 2007. → pages 36
- [101] S. Qin and H. X. Zhou. *Biophys. J.*, 97:12, 2009. → pages 37
- [102] W. G. McMillan and J. E. Mayer. *J. Chem. Phys.*, 13:276, 1945. → pages 37
- [103] J. G. Kirkwood and F. P. Buff. *J. Chem. Phys.*, 19:774, 1951. → pages 37, 38
- [104] J. Rosgen, B. M. Pettitt, J. Perkyuns, and D. W. Bolen. *J. Chem. Phys.*, 108:2048, 2004. → pages 38
- [105] J. Rosgen, B. M. Pettitt, and D. W. Bolen. *Biochem.*, 43:14472, 2004. → pages 38, 39

- [106] M. Aburi and P. E. Smith. *J. Phys. Chem. B*, 108:7382, 2004. → pages 39
- [107] P. E. Smith. *J. Phys. Chem. B*, 108:18716, 2004. → pages 39
- [108] A. Ben Naim. *Statistical Thermodynamics for Chemists and Biochemists*. Plenum Press, New York, NY, 1992. → pages 40
- [109] S. Shimizu and C. L. Boon. *J. Chem. Phys.*, 121:9147, 2004. → pages
- [110] C. F. Anderson, E. S. Courtenay, and M. T. Record. *J. Phys. Chem. B*, 106:418, 2002. → pages 40
- [111] A. Ben Naim. *J. Chem. Phys.*, 67:4884, 1977. → pages 40
- [112] B. M. Baynes and B. L. Trout. *J. Phys. Chem. B*, 107:14058, 2003. → pages 40, 41
- [113] R. F. Greene and C. N. Pace. *J. Biol. Chem.*, 249:5388, 1974. → pages 41
- [114] G. Makhatadze. *J. Phys. Chem. B*, 103:4781, 1999. → pages
- [115] S. N. Timasheff and G. Xie. *Biophys. Chem.*, 105:421, 2003. → pages 41
- [116] C. C. Mello and D. Barrick. *Protein Sci.*, 12:1522, 2003. → pages 41
- [117] D. J. Felitsky and M. T. Record. *Biochemistry*, 43:9276, 2004. → pages 41
- [118] C. Tanford. *Physical Chemistry of Macromolecules*. Wiley and Sons, New York, 1961. → pages 44, 45, 100
- [119] P. J. Flory and S. Fisk. *J. Chem. Phys.*, 44:2243, 1966. → pages 44
- [120] D. Lhuillier. *J. Phys.*, 49:705, 1988. → pages 44
- [121] A. P. Minton. *Meth. Enzym.*, 295:127, 1998. → pages 45
- [122] H. S. Chan and K. A. Dill. *Annu. Rev. Biophys. Biophys. Chem.*, 20:447, 1991. → pages 46, 47
- [123] P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, Ny, 1953. → pages
- [124] P. J. Flory. *Statistical Mechanics of Chain Molecules*. Hanser Publishers, Munich, 1969. → pages 47, 100
- [125] H. Yamakawa. *Modern Theory of Polymer Solutions*. Harper and Row, New York, 1971. → pages 47, 48

- [126] E. Teramoto, K. Kurata, R. Chujo, C. Suzuki, K. Tani, and T. Kajikawa. *J. Phys. Soc. Jpn*, 10:953, 1955. → pages 47
- [127] M. E. Fisher. *J. Chem. Phys.*, 44:616, 1966. → pages 48
- [128] A. Sali, E. I. Shakhnovich, and M. Karplus. *Nature*, 369:248, 1994. → pages 48
- [129] T. Lazaridis and M. Karplus. *Science*, 278:1928, 1997. → pages
- [130] V. S. Pande and D. S. Rokhsar. *Proc. Natl. Acad. Sci. USA*, 95:1490, 1998. → pages
- [131] S. Takada, Z. Luthey-Schulten, and P. G. Wolynes. *J. Chem. Phys.*, 110:11616, 1999. → pages 48
- [132] M. Bishop, J. H. R. Clarke, A. Rey, and J. J. Freire. *J. Chem. Phys.*, 95:4589, 1991. → pages 49
- [133] A. Linhananta, J. Boer, and I. MacKay. *J. Chem. Phys.*, 122:114901, 2005. → pages 49, 77, 81, 122
- [134] Y. Q. Zhou, , and M. Karplus. *J. Mol. Biol.*, 293:917, 1999. → pages 49, 50
- [135] A. J. Barrett, M. Mansfield, and B. C. Benesch. *Macromolecules*, 24:1615, 1991. → pages 49
- [136] M. Lal. *Mol. Phys.*, 17:57, 1969. → pages 49
- [137] N. Madras and A. D. Sokal. *J. Stat. Phys.*, 50:109, 1988. → pages 49, 50
- [138] K. Kremer and K. Binder. *Phys. Rep.*, 7:259, 1988. → pages 49
- [139] P. Grassberger. *Phys. Rev. E*, 56:3682, 1997. → pages 49
- [140] M. N. Rosenbluth and A. W. Rosenbluth. *J. Chem. Phys.*, 23:356, 1955. → pages 50
- [141] Y. Q. Zhou, C. Zhang, G. Stell, and J. Wang J. *J. Am. Chem. Soc.*, 125:6300, 2003. → pages 51
- [142] J. M. Borreguero, N. V. Dokholyan, S. V. Buldyrev, E. I. Shakhnovich, and H. E. Stanley. *J. Mol. Biol.*, 318:863, 2002. → pages
- [143] H. B. Jang, C. K. Hall, and Y. Q. Zhou. *Biophys. J.*, 86:31, 2004. → pages

- [144] F. Ding and J. J. LaRocque. *J. Biol. Chem.*, 280:40235, 2005. → pages
- [145] F. Ding, D. Tsao, H. F. Nie, and N. V. Dokholyan. *Structure*, 16:1010, 2008. → pages 51
- [146] M. S. Wertheim. *Phys. Rev. Lett.*, 10:321, 1963. → pages 51
- [147] E. Neria, S. Fischer, and M. Karplus. *J. Chem. Phys.*, 105:1902, 1996. → pages 51, 82, 122
- [148] E. Thiele. *J. Chem. Phys.*, 39:474, 1963. → pages 51
- [149] H. N. W. Lekkerkerker, W. C. K. Poon, P. N. Pusey, A. Stroobants, and P. B. Warren. *Europhysics Letters*, 20:559, 1992. → pages 61
- [150] D. G. A. L. Aarts, R. Tuinier, and H. N. W. Lekkerkerker. *J. Phys.: Condens. Matter*, 14:7551, 2002. → pages 61
- [151] C. Levinthal. *J. Chem. Phys.*, 65:44, 1968. → pages 61
- [152] C. Levinthal, P. Debrunner, J. C. M. Tsibris, and E. Munck. *Mossbauer Spectroscopy in Biological Systems*. Eds., University of Illinois Press, 1969. → pages 61
- [153] D. B. Wetlaufer. *Proc. Natl. Acad. Sci. USA*, 70:697, 1973. → pages 61
- [154] C. B. Anfinsen. *Science*, 181:223, 1973. → pages 62
- [155] H. X. Zhou. *Acc. Chem. Res.*, 37:123, 2004. → pages 64, 80
- [156] Ke Xu Jyotica Batra and Huan-Xiang Zhou. *Proteins*, 77:133, 2009. → pages
- [157] H. Dong, S. Qin, and H. X. Zhou. *PLoS Comp. Biol.*, 6:1, 2010. → pages 64
- [158] D. Tsao, A. P. Minton, and N. V. Dokholyan. *PLoS One*, 5:1, 2010. → pages 64
- [159] A. J. Saunders, P. R. Davis-Searles, D. L. Allen, G. J. Pielak, and D. A. Erie. → pages 64
- [160] C. B. Anfinsen, E. Haber, M. Sela, and Jr. F.H. White. *Proc. Natl. Acad. Sci.*, 47:1309, 1961. → pages 69

- [161] J. N. Onuchic and P. G. Wolynes. *Curr. Opin. Struct. Biol.*, 14:70, 2004. → pages
- [162] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. *Proteins*, 21:167, 1995. → pages 69
- [163] J. D. Bryngelson and P. G. Wolynes. *J. Phys. Chem.*, 93:6902, 1989. → pages 69
- [164] C. Clementi, H. Nymeyer, and J. N. Onuchic. *J. Mol. Biol.*, 298:937, 2000. → pages 69
- [165] P. E. Leopold, M. Montal, and J. N. Onuchic. *Proc. Natl. Acad. Sci.*, 89:8721, 1992. → pages 69
- [166] S. S. Plotkin, J. Wang, and P. G. Wolynes. *J. Chem. Phys.*, 106:2932, 1997. → pages
- [167] K. A. Dill and H. S. Chan. *Nat. Struct. Biol.*, 4:10, 1997. → pages
- [168] J. G. Lyubovitsky, H. B. Gray, and J. R. Winkler. *J. Am. Chem. Soc.*, 124:5481, 2002. → pages
- [169] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. *Science*, 254:1598, 1991. → pages
- [170] J. D. Bryngelson and P. G. Wolynes. *Cell. Biochem. Biophys.*, 46:123, 2006. → pages
- [171] D. Poland and H. A. Scheraga. *Theory of Helix-Coil Transitions in Biopolymers*. Academic, New York, 1970. → pages
- [172] S. S. Plotkin, J. Wang, and P. G. Wolynes. *Phys. Rev. E*, 53:6271, 1996. → pages 69
- [173] P. H. Verdier and W. H. Stockmayer. *J. Chem. Phys.*, 36:227, 1962. → pages 69
- [174] T. F. Schatzki. *J. Poly. Sci.*, 57:337, 1962. → pages
- [175] F. T. Wall, S. Windwer, and P. J. Gans. *J. Chem. Phys.*, 38:2220, 1963. → pages
- [176] J. des Cloizeaux. *Phys. Rev. A*, 10:1665, 1974. → pages

- [177] A. P. Minton. *Biopolymers*, 20:2093, 1981. → pages 69, 72
- [178] J. D. Bryngelson and P. G. Wolynes. *J. Phys. Chem.*, 93(19):6902, 1989.
→ pages 69
- [179] J. J. Portman. *J. Chem. Phys.*, 72:4350, 1980. → pages 69
- [180] J. F. Douglas and T. Ishinabe. *Phys. Rev. E*, 51:1791, 1997. → pages 70, 71
- [181] S. S. Plotkin, J. Wang, and P. G. Wolynes. *J. Chem. Phys.*, 106:2932, 1997.
→ pages 70
- [182] B. van den Berg, R. Wain, D. M. Dobson, and R. J. Ellis. *EMBO J.*,
19:3870, 2000. → pages 72
- [183] G. F. Xie and S. N. Timasheff. *Biophys. Chem.*, 64:25, 1997. → pages 74,
99
- [184] G. F. Xie and S. N. Timasheff. *Protein Sci.*, 6:211, 1997. → pages 74, 99
- [185] D. W. Bolen and L. V. J. Baskakov. *J. Mol. Biol.*, 310:955, 2001. → pages
74, 99
- [186] M. Auton and D. W. Bolen. *Proc. Natl. Acad. Sci. USA*, 102:15065, 2005.
→ pages 74, 99
- [187] B. J. Bennion and V. Daggett. *Proc. Natl. Acad. Sci. USA*, 100:5142, 2003.
→ pages 75, 99
- [188] A. P. Gast, C. K. Hall, and W. B. Russel. *Faraday Discuss Chem. Soc.*,
76:189, 1983. → pages 76
- [189] R. F. Gahl, M. Narayan, G. Xu, and H. A. Scheraga. *Biochemical and
Biophysical Research Communications*, 325:707, 2004. → pages 80
- [190] S. Milev, H. R. Bosshard, and I. Jelesarov. *Biochemistry*, 44:285, 2005. →
pages 80
- [191] A. Dhar, A. Samiotakis, S. Ebbinghaus, L. Nienhaus, D. Homouz,
M. Gruebele, and M. S. Cheung. *Proc. Natl. Acad. Sci.*, 107:17586, 2010.
→ pages 80
- [192] L. Stagg, S. Q. Zhang, M. S. Cheung, and P. Wittung-Stafshede. *Proc.
Natl. Acad. Sci.*, 104:18976, 2007. → pages 80

- [193] M. Auton and D. W. Bolen. *Proc. Natl. Acad. Sci.*, 102:15065, 2005. → pages 80, 81
- [194] I. Baskakov and D. W. Bolen. *J. Biol. Chem.*, 273:4831, 1998. → pages 80
- [195] C.H. Henkels, J. C. Kurz, C. A. Fierke, and T. G. Oas. → pages 80
- [196] N. Go. *Annual Review of Biophysics and Bioengineering*, 12:183, 1983. → pages 81, 122
- [197] Y. Q. Zhou and A. Linhananta. *Proteins-Structure Function and Genetics*, 47:154, 2002. → pages 81, 122
- [198] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. *J. Comput. Chem.*, 13:1011, 1992. → pages 82
- [199] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. *J. Comput. Chem.*, 16:1339, 1995. → pages
- [200] N. D. Socci and J. N. Onuchic. *J. Chem. Phys.*, 103:4732, 1995. → pages
- [201] A. Sali, E. Shakhnovich, and M. Karplus. *Nature*, 369:248, 1994. → pages 82
- [202] Y. Q. Zhou, C. K. Hall, and M. Karplus. *Protein Sci.*, 8:1064, 1999. → pages 91
- [203] H. Kaya and H. S. Chan. *Proteins-Structure Function and Genetics*, 40:637, 2000. → pages 91
- [204] C. Tanford. *Adv. Protein Chem.*, 23:121, 1968. → pages
- [205] P. L. Privalov. *Adv. Protein Chem.*, 33:167, 1979. → pages 91
- [206] N. Tokuriki, M. Kinjo, S. Negi, M. Hoshino, Y. Goto, I. Urabe, and T. Yomo. → pages 98
- [207] O. Medalia, L. Weber, A. S. Frangakis, D. Nicastro, G. Gerisch, and W. Baumeister. → pages 98
- [208] J. M. Victor, J. B. Imbert, and D. Lhuillier. *J.Chem. Phys.*, 100:5372, 1994. → pages 100
- [209] J. Dautenhahn and C. K. Hall. *Macromolecules*, 27:5399, 1994. → pages
- [210] M. Bishop and C. J. Saltiel. *J.Chem. Phys.*, 95:606, 1991. → pages 100

- [211] D. P. Goldenberg. *J. Mol. Biol.*, 326:1615, 2003. → pages 100
- [212] A. P. Minton. *Biophys. J.*, 63:1090, 1992. → pages 106

Discontinuous Molecular Dynamics

A brief description of the DMD simulation method is presented here. The initial heavy-atom positions were obtained from the NMR structure. The resulting structure is comprised of 189 heavy atoms and polar hydrogen atoms, which in the discontinuous model are represented as hard spheres. Two bonded atoms i and j , as well any 1,3 angle-constrained pair and 1,4 aromatic pair, are constrained to be within 10% of their distance in the NMR structure by an infinite square-well potential:

$$u_{ij}^{bond} = \begin{cases} \infty, & r \leq 0.9\sigma_{ij} \\ 0, & 0.9\sigma_{ij} < r \leq 1.1\sigma_{ij} \\ \infty, & 1.1\sigma_{ij} < r \end{cases} \quad (1)$$

where σ_{ij} is the separation distance of the bonded i, j pair in the equilibrium structure. The model also includes a discontinuous improper dihedral potential, to maintain chirality about tetrahedral heavy atoms (e.g. $C\alpha, C\beta, C, N$), and planar moieties such as tryptophan rings.

$$u_{ij}^{improp} = \begin{cases} \infty, & \omega < \omega_0 - 20^\circ \\ 0, & \omega_0 - 20^\circ < \omega < \omega_0 + 20^\circ \\ \infty, & \omega_0 + 20^\circ < \omega \end{cases} \quad (2)$$

Here ω represents the dihedral angles of the constrained atoms, which are restricted to values near $\omega_0 = 32.26439^\circ$ for tetrahedral heavy atoms, and $\omega_0 = 0^\circ$ for planar atoms.

As well, two non-bonded atoms i, j may interact by hard-sphere potential (purely repulsive) with a hard-core radius:

$$u_{ij}^{non-bond} = \begin{cases} \infty, & r < 0.8\sigma_{ij}^{vdW} \\ \varepsilon_{pp} = B_{ij}\varepsilon_{G\overline{o}}, & 0.8\sigma_{ij}^{vdW} < r < 1.2\sigma_{ij}^{vdW} \\ 0, & 1.2\sigma_{ij}^{vdW} < r \end{cases} \quad (3)$$

where σ_{ij}^{vdW} is the sum of the van der Waals (vdW) radii $r_i + r_j$ for each atom pair, as given by the CHARMM potential set 19 [147] and B_{ij} and $\varepsilon_{G\overline{o}}$ are $G\overline{o}$ interaction strength parameters giving the depth of the square well potential, which may depend on the identities of atoms i and j .

Using these parameters we performed a short DMD simulation with fixed native contacts to remove interactions violating equations (2)-(4) from the initial experimental structure, to produce an equilibrium structure of the model Trp-cage consistent with the above potentials. The $G\overline{o}$ model potential [196] is implemented by setting the non-bonded square-well depth ε_{pp} to $\varepsilon_{G\overline{o}}$ ($B_{ij} = -1$) for all ij pairs in the equilibrated structure with van der Waals overlap $r < 1.2\sigma_{ij}^{vdW}$; for all other non-bonded ij pairs, the square-well depth is set to 0 ($B_{ij} = 0$), so that these atom pairs are purely repulsive. As in previous DMD studies [133, 197], the energy scale is set by the $G\overline{o}$ contact energy; thus simulations are performed with the $G\overline{o}$ contact energy ε_{pp}^* set to $-\varepsilon_{G\overline{o}}^* = -1$, and all energies and temperatures are scaled in units

of $\epsilon_{G\bar{o}}$ ($E^* = E/\epsilon_{G\bar{o}}$ and $T^* = k_B T/\epsilon_{G\bar{o}}$). A reduced time unit $t^* = t\sqrt{\epsilon_{G\bar{o}}/m\sigma_L^2}$ is also used (m can be taken to be the average atomic mass of the atoms comprising the protein and $\sigma_L = 1\text{\AA}$).

The $G\bar{o}$ model protein in explicit solvent is implemented by placing the Trp-cage protein in a $40\text{\AA} \times 40\text{\AA} \times 40\text{\AA}$ box, along with a variable number of spherical solvent molecules randomly inserted without hardcore overlaps. A typical simulation consisted of 1000 spherical solvent particles of radius 1.5\AA (the approximate radius of water). This is about half the number of water molecules in a $55M$ solution for a 40\AA^3 box. We employ such a dilute concentration for computational convenience; physical concentrations have collision times sufficiently short as to make such simulations prohibitively slow. Diluting the concentration weakens the effects that would be observed by varying solvent qualities from those at $55M$, i.e. this simplification effectively places lower bounds on any trends that we predict would be observed. For this reason we find the approximation acceptable as it only strengthens the conclusions of this study. Standard periodic boundary conditions are implemented.

Solvent molecules interact with both protein moieties and with each other by a square-well potential plus hard core radii, having the form:

$$\frac{u_{ij}^{x-s}}{\epsilon_{G\bar{o}}} = \begin{cases} \infty, & r < 0.8\sigma_{ij}^{xs} \\ \epsilon_{xs}^* = \epsilon_{xs}/\epsilon_{G\bar{o}}, & 0.8\sigma_{ij}^{xs} < r < 1.2\sigma_{ij}^{xs} \\ 0, & 1.2\sigma_{ij}^{xs} < r \end{cases} \quad (4)$$

In above equation x may be either a protein atom p or another solvent residue s . For solvent-solvent interactions, σ_{ij}^{ss} is the vdW diameter of the solvent, which we generally set to $\sigma_{ij}^{ss} = 3.0\text{\AA}$: roughly the size of a water molecule. ϵ_{ss}^* is the solvent-solvent square-well depth in units of the $G\bar{o}$ contact energy $\epsilon_{G\bar{o}}$. For solvent-protein interactions, $\sigma_{ij}^{ps} = (\sigma_i^{vdW} + \sigma_j^{water})/2$ is the average vdW diameter of the protein-solvent (i, j) pair, where σ_i^{vdW} is the vdW diameter of the i th atom of the protein, and $\sigma_j^{water} = 3.0\text{\AA}$ is the vdW diameter of the j th water molecule. ϵ_{ps}^* is the protein-solvent square-well depth in units of the $G\bar{o}$ contact energy $\epsilon_{G\bar{o}}$.

To demonstrate the extent of improvements in efficiency by using discontinuous potential simulations, we shortly review the differences between DMD simulations and continuous potential simulations. In the latter case, generally a fixed time step is being used for all the atoms in the simulation. For each time step, the forces acting on each site are summed over all the sites in the simulation (a summation of length N). When performed over all N sites, the total summation time is of order N^2 . Simulations via discontinuous potentials basically deal with looking for future collisions and scheduling them. The interacting force between non-colliding atoms do not change because the potential is flat and consequently these atoms simply get closer together or farther apart. Therefore one needs to only calculate and schedule the future prospective collisions of the colliding atoms. This means that at the time of the collision the positions of all the interaction sites should be updated in time since the last collision. This process requires a loop over all N sites while the coordinates of only two sites have changed significantly. One way to get around this problem is to compute false positions of the two sites that were just colliding, at some previous update of all positions, such that the velocities and positions of sites are calculated correctly after the collision which in turn guarantees that all future events will be correct. Yet another trick is to compare the simulation box into cells with the size of the largest repulsive site. Therefore cell crossings can be tracked in much the same way as collisions and the only possible collisions, are generated by sites that are in adjacent cells. This will narrow down the domain of possible future collisions from all sites in the fluid to only a few. A similar cell structure may be used in continuous potential simulations to reduce their computation times to order N . However, the range of a continuous potential (e.g. Lennard-Jones potential) is much larger than that of a hard sphere potential and therefore the size of cells also should be very large. In general, the number of sites that are involved in summing the forces for a continuous potential model are about one to two orders of magnitude more than that in a discontinuous potential model leading to proportional penalties in computation time.

The general idea is, given the positions, velocities and other dynamic information of molecules at time t , we attempt to find the positions, velocities and other information of molecules at a later time $t + \delta t$ to a reasonable degree of accuracy. Consider two spheres with diameter σ whose position at time t is \mathbf{r}_i and \mathbf{r}_j and

their velocities at time t is \mathbf{v}_i and \mathbf{v}_j . If at time $t + \delta t$ these spheres collide then the following equation should be satisfied:

$$|\mathbf{r}_{ij}(t + \delta t)| = |\mathbf{r}_{ij} + \mathbf{v}_{ij}\delta t| = \sigma \quad (5)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. If we define $b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{v}_{ij}$ equation (5) becomes:

$$v_{ij}^2 \delta t^2 + 2b_{ij} \delta t + r_{ij}^2 - \sigma^2 = 0 \quad (6)$$

If $b_{ij} > 0$ then spheres are moving away from each other and will not collide. If $b_{ij} < 0$ and $b_{ij}^2 - v_{ij}^2(r_{ij}^2 - \sigma^2) < 0$ again equation (6) will have complex root and no collision will happen. Otherwise, the spheres will collide and equation (6) will have two positive roots, the smaller of which corresponds to impact.

Using conservation of total linear momentum and kinetic energy, the changes in velocity such that:

$$v_i(after) = v_i(before) + \delta v_i \quad (7)$$

is given by:

$$\delta v_i = -\frac{b_{ij}}{\sigma^2} r_{ij} = -v_{ij,v} \quad (8)$$

with b_{ij} now being calculated at the moment of impact. δv_i is then just simply negative of the projection of \mathbf{v}_{ij} along the \mathbf{r}_{ij} direction which is denoted by $v_{ij,v}$.