Theses and Dissertations

2014

Protein folding and Protein-Protein interactions in a Cell-like environment

Apratim Bhattacharya
Lehigh University

Recommended Citation

Protein folding and protein-protein interactions in a cell-like environment

by

Apratim Bhattacharya

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Chemical Engineering

September 2014

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Date

_____

Prof. Jeetain Mittal (Thesis Adviser)

_____

Prof. Anand Jagota

_____

Prof. Mark A. Snyder

_____

Prof. Dimitrios Vavylonis

_____

Dr. Young C. Kim

iii

Acknowledgements

At the very onset I would like to thank my adviser Prof. Jeetain Mittal for providing me with an excellent environment for research. Without his technical and editorial advice this dissertation would not have been possible. His organizational skills, attention to details, critical appreciation of research topics and the ability to simplify complex ideas have taught me innumerable lessons and insights on the workings of research in general. His immense patience and encouragement have been extremely beneficial to me personally.

Next, I would like to thank my committee members for their constant support and critical inputs in shaping up my dissertation. I am extremely grateful to Dr. Youngchan Kim for an immensely fruitful collaboration in understanding the effects of macromolecular crowding on protein-protein interactions.

I consider myself extremely fortunate to have attended some of the lectures of the excellent faculty members at Lehigh University. Prof. Anand Jagota and Prof. Jeetain Mittal taught me the basics of Statistical Physics and Numerical Methods necessary to generate almost all the data in this thesis. I am immensely grateful to Prof. Mark A. Snyder and Prof. James F. Gilchrist for making the Transport Phenomena lectures extremely lucid. I would like to thank Prof. Michael Behe for allowing me to audit his excellent Biochemistry course that made me acquainted to the complexities of biomolecules. Finally, I would like to thank Prof. Manoj K. Chaudhury allowing me to audit the Colloids and Surface Chemistry course in which I came to appreciate the limitations of molecular simulations.

I would like to thank my current and former group members Yajun Ding, Minseok Song, Bilge Uz, Gül Zerze, Alexander Bourque, Jonathan Rosen, Matthew Melillo,

My undergraduate advisers, Prof. M.N. Biswas, Prof. Chiranjib Bhattacharjee, Prof. Dulal C. Mukherjee, Prof. Debasish Sarkar and Prof. B.B. Paira have motivated me to pursue graduate studies. I would like to thank them for their constant support and encouragement. Thank you Mr. Anirban Roy for introducing me to the possibility of undergraduate research and guiding me immensely during the graduate school application process. Thanks to Prof. Dwaipayan Sen for your constant support during my tenure at Jadavpur University. I would also like to thank Profs. Nirman Ganguly, Amitava Hatial, and Indrajit Das at Heritage Institute of Technology for their encouragement. Thank you Suvrajit, and Debadrito for the never ending telephonic conversations.

Finally, I cannot thank my family enough for their constant love and support. Thank you Apramit, for your support. Thank you mom and dad for encouraging my diverse interests. This would not have been possible without your constant encouragement.

Contents

List of Figures

Abstract

The cellular environment is highly crowded owing to the presence of several kinds of macromolecules such as lipids, sugars, nucleic acids, and proteins, along with large organized macromolecular arrays such as the Chaperones. Thus, most of the *in vivo* processes such as the interaction between proteins and folding of proteins into their compact three-dimensional structures, that are of paramount biological importance, occur in restricted spaces. The primary focus of this dissertation is to broaden the understanding of protein folding and protein-protein interactions in a cell mimetic environment.

We have reviewed the important developments that have furthered our understanding of the effects of macromolecular crowding on protein-protein interactions. We have outlined the development of a comprehensive crowding theory that can predict the binding thermodynamics considering both repulsive and attractive protein-crowder interactions. It has been observed that favorable weak interactions between proteins and crowders destabilize the protein complex formation in contrast to the traditional understanding that owing to the excluded volume effects, the primary effect of macromolecular crowding is to stabilize the bound complex.

Additionally, we have performed replica exchange molecular dynamics (REMD) simulations to study the binding thermodynamics of proteins modeled as conformationally 'flexible' entities under the influence of macromolecular crowding. We observe similar destabilization due to attractive interactions between proteins and crowders. Interestingly, we observe that the Scaled Particle Theory, originally developed for hard-sphere particles, can be utilized to predict the binding thermodynamics of flexible proteins.

1

In order to study the effects of confinement, we have performed extensive REMD simulations using all-atom explicit solvent models to study the conformational stability of the sixteen residue GB1-hairpin confined between planar Lennard-Jones walls. We observe that confinement significantly alters the free-energy landscape. Under confinement, the misfolded state of the peptide is completely absent owing to the preferential adsorption of the hydrophobic residues on the confinement walls.

Finally, in order to study the effect of cylindrical confinement, we have performed extensive REMD simulations of polyalanine helices confined within carbon nanotubes using all-atom explicit solvent models. We observe that the α-helix propensity of polyalanine is significantly reduced under confinement and that the effect is independent of the lengthscale of confinement.

Chapter 1

Introduction

Protein folding and protein-protein interactions *in vivo* occur in a highly crowded and heterogeneous environment, owing to the presence of biomolecules such as DNA, RNA and proteins in high concentrations. Biomolecules other than the protein(s) under study are referred to as 'crowders' and the crowded cellular environment is expected to alter both the thermodynamics and kinetics of biophysical processes relative to what is observed in bulk. Experimental and theoretical studies probing the effects of *macromolecular crowding* [142, 155] on protein folding and protein-protein interactions have shown that the effect of the presence of crowder particles range from volume exclusion to non-specific attractive interactions between the proteins and crowders.

In the limit where the dimensions of the crowder particles are significantly larger than the proteins, thereby static at the protein's relevant timescales, macromolecular crowding effects can be approximated by confinement effects. Moreover, protein folding *in vivo* occurs in the presence of cellular machinery, such as the approximately spherical chaperonin cavity, that can assist folding by preventing misfolding

and aggregation [56, 152]. The exact mechanism of these chaperones is still unknown. Folding of proteins may also be coupled directly to their synthesis, commonly referred to as co-translational folding, during which newly synthesized proteins emerge out of the narrow, roughly cylindrical, ribosome exit tunnel [55, 43]. Additionally, protein folding may also occur near the cellular membrane surfaces which provide the basis of cellular compartmentalization [40]. Thus, *confinement* refers to the volume excluded by a fixed (or confining) boundary to the protein(s) under study.

The steric excluded volume effects of 'inert' crowding agents on the formation of complexes had been the prime focus of most of the studies [98, 153, 78]. However, several additional effects arising due to electrostatic interactions, non-specific attractive interactions (like hydrophobic interactions), and hydrogen bonding maybe important in a crowded environment. In the early models of crowding developed by Minton, attractive interaction potentials were approximated by an effective hard-sphere potential [98]. However, more serious consideration to the effects of attractive interactions between proteins and crowders on protein-protein association has been given quite recently [31, 71, 124]. These studies have highlighted the importance of accounting for enthalpic effects arising from attractive interactions in addition to commonly invoked excluded volume effects. It was found that the enthalpic effects can actually increase the binding free energy, thereby destabilizing the bound complex in contrast to predictions based on theoretical models capturing solely the entropic effects.

It has been shown using theoretical arguments, that confining a protein in an 'inert' space would stabilize the protein against reversible unfolding [154]. This stabilization could be explained to be arising from entropic forces, where the confining entities are expected to exclude an increased number of conformational states of the

unfolded protein as compared to the folded state. A similar argument was put forward by Ziv *et al.* [159] while elucidating the role of the ribosome exit tunnel in stabilizing α-helices. The ribosome exit tunnel was modeled as a cylindrical cavity in their studies. Entropic stabilization of the folded state has also been suggested in the simulation studies of Takagi *et al.* [135]. Using a coarse-grained model they have shown that the shift in the folding temperature with respect to bulk scaled with the radius of the confining cage. Baumketner *et al.* have shown that repulsive confinement raises the collapse temperature of peptides [10]. Another study by Rathore *et al.* showed that the stabilization of proteins under confinement varies with the nature of the confining potentials [119]. They showed that the effect of increased confinement of a soft-repulsive potential on a protein is the destabilization of the folded state. Mittal and Best have shown that the effect of the confining entity on the folding thermodynamics and kinetics is nearly independent on the geometry of confinement. Additionally, they have shown that the effects of confinement are specific to the protein under consideration [101]. Using simple lattice models it has been shown that confinement enhances the rate of protein folding [21]. Hayer-Hartl *et al.* have studied the dependence of the rate of folding of a polypeptide confined within a spherical cavity upon the size of the cavity of confinement. They showed using simple analytical expressions that the rate of two-state folding is maximized at an intermediate size of the spherical cavity and this optimal size increases with increasing molecular weights of polypeptides [58]. Maximization of folding rates at an optimum cavity size has been also been observed in simulation studies of Cheung *et al.* [25]. They also show that the yield of folding can be enhanced by repeated switching between favorable (hydrophobic) and unfavorable (hydrophilic) interactions between the peptide and the confining cavity.

However, most of these studies represented proteins using polymer models and solvent effects were not considered directly. The seminal role of water molecules in driving various biomolecular processes is widely recognized now [68]. The first study probing the behavior of proteins under confinement, in the presence of explicit solvent was performed by Lucent *et al.* [92]. They found, that the primary effect of repulsive confinement on a protein immersed in a bath of water molecules is the destabilization of the folded state. Repulsive confinement is a natural choice as it acts as the basic simplification to represent physical boundaries in a system without any effect on the biomolecule under study. Their study highlighted that the solvent-mediated effect was more dominant than the excluded volume effect of the confining entity on the folding of the peptide. In a later simulation study, Sorin and Pande have shown that confining the α-helical polyalanine peptide with explicit solvent inside carbon nanotubes denatures peptide helicity [129]. They have shown that α-helicity decreases monotonically with increasing confinement and correlates with the decrease in solvent entropy due to confinement. Tian *et al.* have shown that for Trp cage, non-polar confinement stabilizes the folded state due to the effects of volume reduction destabilizing the unfolded state. However, polar confinement has a net destabilizing effect arising from the competitive entropic forces stabilizing the folded state and the enhanced interactions between the charged side chains of the peptide and the surface of the confinement [138]. Another study on the Trp cage peptide under hydrophobic confinement showed that adsorption of the peptide on the hydrophobic walls of confinement stabilizes intermediate structures not observed in bulk [95]. Trp cage has been a popular model for studying protein folding as it has been found to fold rapidly and spontaneously into a globular structure with well-defined secondary and tertiary structure elements. Moreover, it has been studied

extensively by both experiments and simulations [97] and thus acts as a natural choice for studying protein folding in perturbed environments.

Several experimental studies have been conducted attempting to understand the folding stability of peptides in cell-like milieus. Generally, encapsulation is either achieved by porous silica gels or by bis(2-ethylhexyl) sulfosuccinate (AOT) reverse micelles. In both cases, the cage sizes can be controlled. Ravindra *et al.* have shown that the stability of the protein RNase A confined in the mesoporous silicate system is significantly stabilized with the melting temperature of the protein being raised by $\approx 30^o$C. Peterson *et al.* utilized a three-helix bundle protein of de novo design, α_3W, and mutated some of the residues such that it becomes unfolded in a dilute solution. When the variant α_3W was encapsulated in AOT reverse micelles with a low (smaller cage) water loading (ratio of water to AOT), NMR experiments revealed helical propensities corresponding to that of the non-mutated α_3W [113] suggesting stabilization of folded states due to confinement. Mukherjee *et al.* observed increased helix formation of alanine-rich peptides in AOT reverse micelles [107]. They attributed the increased helix formation to decreased hydration of backbone amide and carbonyl groups.

Thesis Organization

The purpose of this thesis is to broaden the understanding the effects of macromolecular crowding on protein-protein interactions, and study the effect of simple confinement entities on protein folding using molecular dynamics simulations coupled with enhanced sampling methods. Although there has been significant progress in the experimental forefront [26, 67, 114], comprehensive understanding of protein

folding and protein-protein interactions relies, to a great extent, on molecular dynamics simulations that allow for characterization of the dynamics of proteins at spatial and temporal scales difficult to access by experiments [75, 82]. Chapter 2 describes the standard molecular dynamics simulation protocols along with enhanced sampling methods. In chapter 3, important developments towards a comprehensive understanding of protein-protein interactions under the effect of macromolecular crowding are reviewed. Understanding of macromolecular crowding effects on protein-protein interactions usually considers proteins as 'rigid'-entities. Chapter 4 investigates the effects of macromolecular crowding on protein-protein interactions considering the interacting proteins as flexible entities. Chapters 5 and 6 investigate the effects of confinement using simplistic confining entities such as planar walls and cylindrical carbon nanotubes respectively. Below is an elaborate outline of the latter chapters in the thesis:

Chapter 2: Molecular Dynamics Simulations of Biomolecules

In this chapter we describe in detail the standard protocols usually employed while performing molecular dynamics simulations. Additionally, two methods employed to enhance sampling of the conformational space of proteins are described. We have performed extensive replica exchange molecular dynamics (REMD) [134] simulations with alanine dipeptide and the GB1-hairpin in explicit solvent. Additionally, we have also performed replica exchange with solute tempering [145] (REST2) simulations using the same assemblies. We show that although REST2 appears to be an excellent alternative in terms of improvement of computational efficiency, there are some differences between the results obtained using REMD and REST2. We conclude that REST2 requires further investigation using various forcefields and peptide models.

Chapter 3: Effects of Macromolecular Crowding on Protein-Protein interactions

In this chapter, we describe the important developments in recent years that have furthered our understanding and even allowed prediction of the consequences of macromolecular crowding on proteinprotein interactions. We outline the development of a crowding theory developed in our group that can predict the change in binding free energy due to crowding quantitatively for both repulsive and attractive proteincrowder interactions. One of the most important findings from our recent work is that weak attractive interactions between crowders and proteins can actually destabilize protein complex formation as opposed to the commonly assumed stabilizing effect predicted based on traditional crowding theories that only account for the entropic-excluded volume effects. We also discuss the implications of macromolecular crowding on the population of encounter versus specific native complex.

Chapter 4: Effects of Macromolecular Crowding on the binding thermodynamics of flexible proteins

In chapter 3, the effects of macromolecular crowding on protein-protein interactions has been discussed with the assumption that proteins are 'rigid'-bodies interacting with each other. However, in reality, proteins are flexible entities. Morever, several well-characterized protein-protein interactions involve disorder-to-order transitions upon binding, which is called coupled binding and folding. One or even both proteins can be disordered prior to the interaction. In such cases, it is imperative to consider flexibility. We have studied the effects of macromolecular crowding on the thermodynamics of binding proteins in which the proteins are represented as 'flexible' bodies using a coarse-grained representation. We have studied the effects of both repulsive and attractive crowders in order to study the effects of volume exclusion and

9

non-specific interactions between the crowders and binding proteins. Additionally, we also show, how our comprehensive understanding of binding of 'rigid' proteins can be utilized to study the binding of 'flexible' proteins.

Chapter 5: Planar confinement significantly modulates the Free-Energy landscape of GB1 hairpin

In this chapter, we study the effects of confinement between planar walls on the folding thermodynamics of a β-hairpin, using large-scale replica-exchange molecular-dynamics simulations with an all-atom model and explicit solvent. We find that the folding free-energy landscape of this peptide observed in bulk is significantly modified when the peptide is confined between the walls. Most notably, the propensity of the peptide to form a misfolded state observed in the bulk solution becomes negligible under confinement. The absence of the misfolded state under confinement can be explained by an increased tendency of hydrophobic aromatic side chains to stay near the walls, because the misfolded state is characterized by a nonnative arrangement of aromatic side chains. These results from a simple confinement model may provide clues about the role of chaperonin confinement in smoothing folding landscapes by avoiding trapped intermediates.

Chapter 6: Nanotube confinement de-stabilizes helical propensities of peptides

In this chapter we study the effects of confinement within the cylindrical cavity of carbon nanotubes on α-helical peptides. We have performed extensive replica exchange molecular dynamics simulations in explicit solvent, with all-atom representations of the carbon nanotube and peptide. Various degrees of confinement has been studied by confining the peptide within nanotubes of varying diameters. We observe that the primary effect of nanotube confinement is the destabilization of α helices

and that this effect is nearly independent of the diameter of the cylindrical cavity. Additionally, we also observe that extended conformations of peptides are stabilized under such conditions.

Chapter 2

Molecular Dynamics Simulations of Biomolecules

2.1 Introduction

Seminal research in the past have provided atomic resolution models of biomolecules including various proteins, DNA and the RNA. Although information regarding the static structures of even single molecules can be obtained with great accuracy using a wide variety of experimental methods [104, 130, 48, 6], the molecules are, in reality, dynamic and their motions are critical to function. Additionally, it might be extremely challenging to devise experimental protocols probing the behavior of these molecules under conditions resembling that of the living cell. Moreover, existence of functional, yet intrinsically disordered protein domains [35] posits significant challenges in characterizing conformations experimentally. Molecular Dynamics (MD) Simulations provide insight into the workings of biomolecular systems at spatio-temporal scales usually difficult to probe by experimental methods.

Recent developments hold significant promise to increase the utility of MD simulations in the study of biomolecular systems [82]. First, accurate simulations of biological systems to the atomistic detail at physiologically relevant timescales, are possible owing to the recent advances in parallelization schemes achieved through general purpose computer-chips, graphical processing units and even development of special-purpose parallel architectures [32]. Second, utilization of enhanced sampling methods, in combination with the improvement of force fields (or energy functions) has made it possible to extensively sample the conformational space of proteins and other biomolecules [161]. Improved sampling, reveals shortcomings in force fields usually not elucidated by shorter simulations. Third, development of simplified models, based on implicit solvation and coarse-graining can be used to access even longer timescales. Finally, concurrent progresses in laboratory methods offer opportunities to compare computational results to experimental data gathered on nearly similar timescales thereby validating models and methods underlying molecular simulations [125].

In this chapter, we describe the fundamentals of molecular dynamics simulations followed by a description of enhanced sampling methods.

2.2 Fundamentals of Molecular Dynamics Simulations

Molecular simulations are based on the validity of a few assumptions. The first of these is the Born-Oppenheimer approximation which allows the wavefunction of a molecule to be broken into its electronic and nuclear components. The Born-Oppenheimer approximation stems from the fact that the inertia of the electrons are negligible as

compared to the inertia of the nucleus and thus, the potential energy is calculated as a function of the nuclear positions only. Molecular dynamics simulations are based upon a model of interactions within the system which comprises of stretching of bonds, opening and closing of angles, the rotations about single bonds and the inter-molecular interactions between atoms which are not involved in any bonded interaction. The set of parameters, and mathematical functions, which describe these interactions, collectively constitute the *force field*. One of the key attributes of a force field is transferability, for it enables a set of parameters developed and tested on relatively small number of cases to be applied to a much wider range of problems. Additionally, parameters developed from data on small molecules can be used to study larger molecules such as proteins and DNA molecules.

Many of the molecular modeling force fields are interpreted in terms of the intra- and inter- molecular forces within the system under study. There are energetic penalties that are associated with the deviation of bonds and angles away from their *reference* or *equilibrium* values. Additionally there is a function that describes how the energy changes as bonds are rotated and finally the force field contains terms that describe the interaction between non-bonded parts of the system. This representation allows various terms to be ascribed to changes in specific internal coordinates such as bond lengths, angles, the rotation of bonds or movements of atoms relative to each other. A functional form for such a force field that can be used to model single molecules or assemblies of atoms and/or molecules is:

$$U(\mathbf{r}^N) = \begin{aligned} &\sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right). \end{aligned}$$

(2.1)

$U(\mathbf{r}^N)$ denotes the potential energy, which is a function of the positions $\mathbf{r}$ of N particles (usually atoms). The various contributions are schematically represented in figure 2.1. The top row of eq. 2.1 refers to the bonded interactions within the system of atoms. The first term models the interaction between pairs of bonded atoms, modeled here by a harmonic potential that gives the increase in energy as the bond length l_i deviates from the reference value $l_{i,0}$. The second term is a summation over all valence angles in the molecule, again modeled using a harmonic potential (a valence angle is the angle formed between three atoms A-B-C in which A and C are both bonded to B). The third term is a torsional potential that models how the energy changes as the bond rotates. The fourth term represents the contributions from the non-bonded interactions. This is calculated between all pairs of atoms (i and j) that are in different molecules or that are in the same molecule but separated by at least three bonds (i.e. have a 1,n relationship where $n \geq 4$). In a simple force field the non-bonded terms are usually modeled using a Lennard-Jones potential for van der Waals interactions (the first term in the bottom row of eq. 2.1) and a Coulomb potential term for electrostatic interactions (the second term in the bottom row of eq. 2.1).

To define a force field not only does the functional form needs to be specified but also the parameters (i.e. the various constants such as $k_i, V_n, \sigma_{ij}, \epsilon_{ij}$ in eq. 2.1). Two force fields may use an identical functional form but different parameters. It should

be noted that the force fields are usually *empirical* since most of the parameters are derived from different types of experimental data [94]. Some of the most popular force fields used are AMBER [28], CHARMM [23] and GROMOS [127]. To study a particular biomolecular system, one needs to perform molecular simulations with various force fields. The one that satisfies most of the important experimental characteristics of a control system consisting of a particular biomolecule is usually chosen to study more sophisticated assemblies with the same. An example of such an approach can be found in the works of Best *et al.* [18].

Having chosen a particular force field to describe a biomolecular system, two principal aspects constitute a molecular dynamics calculation: a. the algorithm used to integrate the equations of motion, b. the calculation of forces using information from the force field, which should be done accurately and efficiently. Various molecular simulation engines with emphasis on biomolecular simulations have been developed over the past few years [111, 116, 61]. All the simulations in this work has been carried out using the Gromacs Simulation package [61]. We focus on each of the above mentioned aspects below:

2.3 Algorithms used to integrate the equations of motion

The position of a particle at a time $t + \Delta t$ is expressed in terms of its position, velocity, and acceleration at time t according to:

$$\mathbf{r}_i(t + \Delta t) \approx \mathbf{r}_i(t) + \Delta t \dot{\mathbf{r}}_i(t) + \frac{1}{2} \Delta t^2 \ddot{\mathbf{r}}_i(t), \tag{2.2}$$
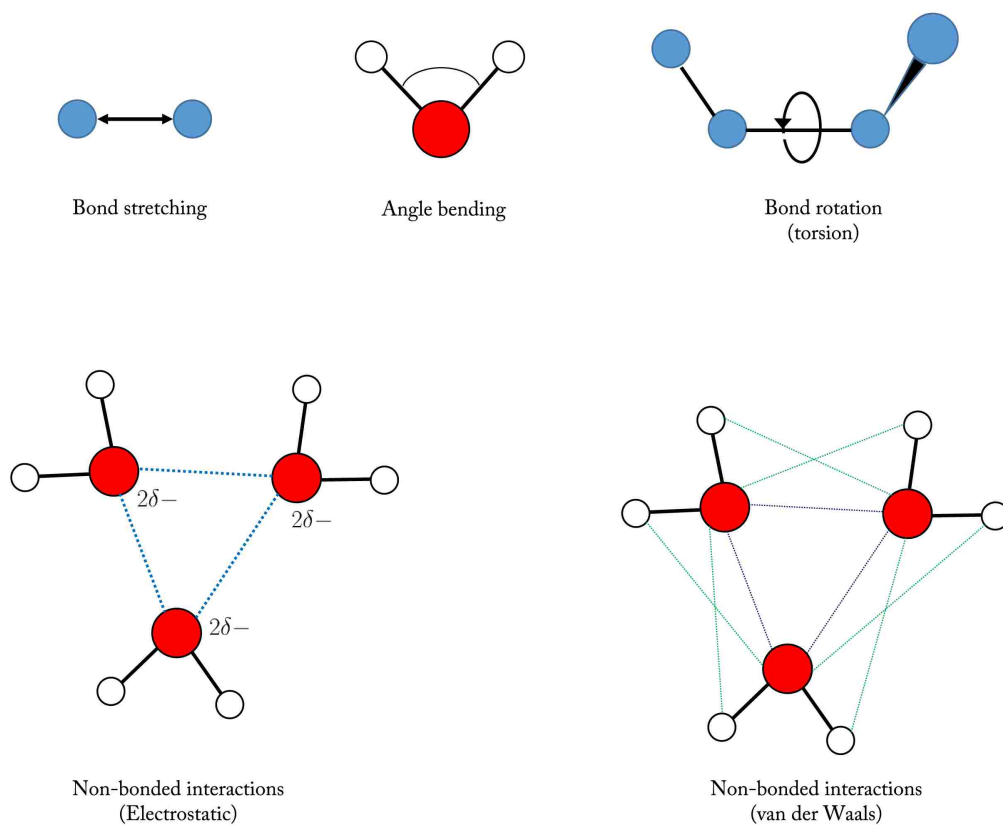
Figure 2.1: Schematic representation of the key contributions to a molecular mechanics force field.

where all terms higher than second order in Δt have been dropped. Since $\dot{\mathbf{r}}_i(t) = \mathbf{v}_i(t)$ and $\ddot{\mathbf{r}}_i(t) = \mathbf{F}_i(t)/m_i$ by Newton's second law, eq. 2.2 can be written as

$$\mathbf{r}_i(t + \Delta t) \approx \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t). \tag{2.3}$$

Now a velocity independent scheme is obtained by writing a similar expansion for $\mathbf{r}_i(t - \Delta t)$:

$$\mathbf{r}_i(t - \Delta t) \approx \mathbf{r}_i(t) - \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i} \mathbf{F}_i(t). \tag{2.4}$$

Adding eqs. 2.3 and 2.4 and rearranging,

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \mathbf{F}_i(t). \tag{2.5}$$

Eq. 2.5 is known as the *Verlet algorithm* [143]. Given initial positions and velocities eq. 2.3 can be used to obtain positions at time $t + \Delta t$ after which eq. 2.5 can be used to generate a trajectory of an arbitrary length. Velocities are constructed at any point in the trajectory using:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)}{2\Delta t}. \tag{2.6}$$

A variant of the Verlet integrator, known as the *Velocity Verlet* algorithm, explicitly evolves positions and velocities. This approach is more elegant, since the phase space is composed of both positions and velocities (or momenta). In this method, the initial condition is considered to be $\mathbf{r}_i(t + \Delta t)$ and $\mathbf{v}_i(t + \Delta t)$. Also, instead of $\mathbf{F}_i(t)$, $\mathbf{F}_i(t + \Delta t)$ are computed and the trajectory is evolved backwards in time to $\mathbf{r}_i(t)$ according to:

$$\mathbf{r}_i(t) = \mathbf{r}_i(t + \Delta t) - \Delta t \mathbf{v}_i(t + \Delta t) + \frac{\Delta t^2}{2m_i} \mathbf{F}_i(t + \Delta t). \tag{2.7}$$

Substituting eq. 2.2 for $\mathbf{r}_i(t + \Delta t)$ into eq. 2.7 and solving for $\mathbf{v}_i(t + \Delta t)$ yields:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{2m_i} [\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)]. \tag{2.8}$$

Thus, the velocity Verlet algorithm uses both eqs. 2.3 and 2.8 to evolve positions and velocities simultaneously.

Another similar scheme, *Leapfrog* integration algorithm has been used to generate most of the data in the subsequent chapters. The *Leapfrog* scheme, conserves angular momentum unlike the Velocity Verlet algorithm and calculates more accurate velocities, using velocities at half time steps.

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}_i(t)}{m_i}\Delta t. \tag{2.9}$$

The velocities at time t can be computed from

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t + \frac{\Delta t}{2}) + \mathbf{v}_i(t - \frac{\Delta t}{2})}{2}, \tag{2.10}$$

and the atomic positions are then obtained from

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right)\Delta t. \tag{2.11}$$

The main advantages of these integrators are that they are time reversible and symplectic in nature. Time reversibility ensures that the motion in the phase space can be traced back in time. The symplectic characteristic ensures that the area in the phase space is preserved along the trajectory. Additionally, these algorithms conserve energy and thus generate the microcanonical ensemble (constant N, V, E where N are the number of particles, V is the volume of the system and E is the total energy).

However, most of the laboratory experiments performed correspond to either a canonical (constant N, V, and T where T is the temperature of the system) or a isothermal-isobaric (constant N, P, T with P being the constant pressure acting on the system) ensemble.

Various thermostats have been developed [64] and Berendsen, Nosé-Hoover, Langevin, velocity rescaling thermostates are the commonly used among others. Two thermostats that have been extensively used to generate the results in the subsequent chapters are discussed below.

2.3.1   Langevin Thermostat

In the implementation of this type of thermostat, $\ddot{\mathbf{r}}_i$ in eq. 2.2 is expressed as

$$\ddot{\mathbf{r}}_i(t) = m_i^{-1}\mathbf{F_i}(t) - \gamma_i(t)\dot{\mathbf{r}}_i(t) + m_i^{-1}\mathbf{R}_i(t), \tag{2.12}$$

where $\mathbf{R}_i$ (t) is a stochastic force and γ_i a positive friction coefficient. Using this thermostat, each particle i is considered to be moving as if it is immersed in a bath of much smaller particles which continuously "jostle" the particle, giving rise to the stochastic noise $\mathbf{R}_i(t)$ term in the force and provide a viscous drag proportional to the velocity $-\gamma(t)\mathbf{v}_i(t)$. The drag and the noise terms balance each other over time and a canonical distribution is thus generated.

2.3.2   Nosé - Hoover Thermostat

In this formulation, the system Hamiltonian is extended by introducing a thermal reservoir and a friction term in the equations of motion. The friction force is proportional to the product of each particle's velocity and a friction parameter ξ. This friction parameter (or "heat bath" variable) is a fully dynamic quantity with its own momentum (p_ξ) and equation of motion; the time derivative is calculated from the difference between the current kinetic energy and the reference temperature.

$\ddot{\mathbf{r}}_i$ in eq. 2.2 is expressed as

$$\ddot{\mathbf{r}}_i(t) = m_i^{-1}\mathbf{F_i}(t) - Q^{-1}p_\xi \dot{\mathbf{r}_i}, \tag{2.13}$$

and the equation of motion for the heat bath is :

$$\dot{p}_\xi = (T - T_0), \tag{2.14}$$

where T_0 is the reference temperature while T is the current instantaneous temperature of the system. The strength of the coupling is determined by the constant Q usually called the "mass parameter" of the reservoir in combination with the reference temperature. The conserved quantity for the Nosé-Hoover equations of motion is not the total energy, but rather

$$H = \sum_{i=1}^{N} \frac{\mathbf{P}_i}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) + \frac{p_\xi^2}{2Q} + N_f k_B T \xi, \tag{2.15}$$

where N_f is the total number of degrees of freedom.

In addition to thermostats, various barostats are also used in practice to generate the Isothermal-isobaric ensemble using molecular simulations. For an extensive treatise the reader is referred to chapter 5 of [139]. In practice, there is no widely agreed-upon theory of selecting and parameterizing the correct thermostat and barostat that guarantees a correct and reliable simulation. General rules of thumb, such as Berendsen thermostat/barostat for relaxation, and Nose-Hoover thermostat/ Parinello-Rahman barostat for production simulations, have been established as a reasonably good practice.

One question that we have set apart thus far is: How do we calculate the forces in the equations of motion?
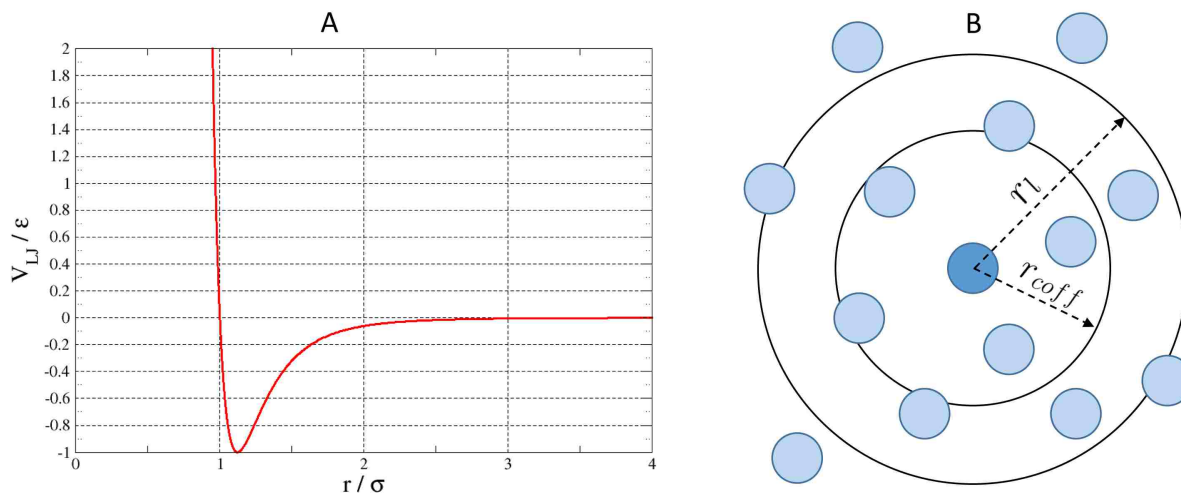
Figure 2.2: **A.** Lennard-Jones potential to model the van der Waals interactions between atoms. **B.** The close neighbors of a molecule.

2.4 Calculation of forces acting on all molecules

The calculation of forces acting on all molecules in a system forms the kernel of a molecular dynamics simulation program. The forces at a particular time-step t are obtained from the negative gradient of the force field (eq. 2.1) using the coordinates at t.

$$\mathbf{F}_i(t) = -\nabla U(\mathbf{r}^N). \tag{2.16}$$

For the simulation of conformationally flexible molecules such as proteins and DNA molecules, the behavior is usually a complex superposition of different motions. The high frequency (such as bond vibrations) are usually of less interest than the lower frequency modes, which often correspond to major conformational changes. In molecular dynamics simulations, the time step is dictated by the highest frequency motion present in the system and thus, it would be of considerable benefit to be

able to increase the time step without prejudicing the accuracy of the simulations. Constraint dynamics enables individual internal coordinates or combinations of specified coordinates to be constrained, or 'fixed' during the simulations without affecting other internal degrees of freedom. Various methods exist for applying constraints in molecular dynamics simulations. The underlying principle of all these methods is the minimization of constraint forces by the technique of lagrange multipliers or projection methods. Common methods used are SETTLE, SHAKE, RATTLE and LINCS [87, 60].

During the computation of forces using eq. 2.16, calculation of the non-bonded interactions is the most computationally intensive. Since this part of the code is executed countless number of times, great care needs to be taken to make it as efficient as possible. Various methodologies such as avoiding square roots methods while evaluating odd-exponent potentials, preparing tables for evaluation of more sophisticated potentials, usage of shifted potentials, etc. are discussed in Chapter 5 of [4]. The van der Waals forces are of the short-range order while the electrostatic forces are long-range. We discuss a few methods for calculations of both short-range and long-range forces below:

2.4.1   Calculation of short-range forces

Cutoff distance

For an N particle system, taking Newton's third law into account, $N(N-1)/2$ calculations of forces are necessary per unit time step. In the short-range order, the interaction energies between the particles rapidly decreases with the particle-particle separation over a distance of only a few times the particle diameter.

The most commonly used potential to calculate short-range forces is given by the Lennard-Jones potential $U_{LJ}(r)$ which is expressed as:

$$U_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{2.17}$$

where σ is the quantity corresponding to the particle diameter, and r is the particle-particle separation. Fig. 2.2A shows the curve of the Lennard-Jones potential. A steep potential barrier in the range $r \leq \sigma$ is illustrated, which induces such a repulsive interaction that the particles are prevented from overlapping, and an attractive interaction in the range $r \geq \sigma$, which rapidly decreases to zero. These characteristics of the potential indicate that the interaction energy after a distance of approximately $r = 3\sigma$ can be assumed to be negligible. Hence forces do not need to be calculated in the range $r > 3\sigma$ in simulations. The distance for cutting off the calculation of forces is the cutoff distance r_{coff}.

Verlet Neighbor List Method

In the Verlet neighbor list method [137], a distance r_l, which is longer than the cutoff radius, is adopted, and for each particle i, a list is created of all the particles that are within the range r_l from its center. Referring to Fig. 2.2B, it is clear that particles within range of $r < r_{coff}$ are certainly within range of $r < r_l$. Now the list of particles within range of $r < r_l$ is renewed with a frequency such that the particles outside $r = r_l$ cannot penetrate the shell $r < r_{coff}$. Hence it is sufficient to calculate the forces between i and particles $r < r_{coff}$ resulting in a significant reduction in the amount of computational time.
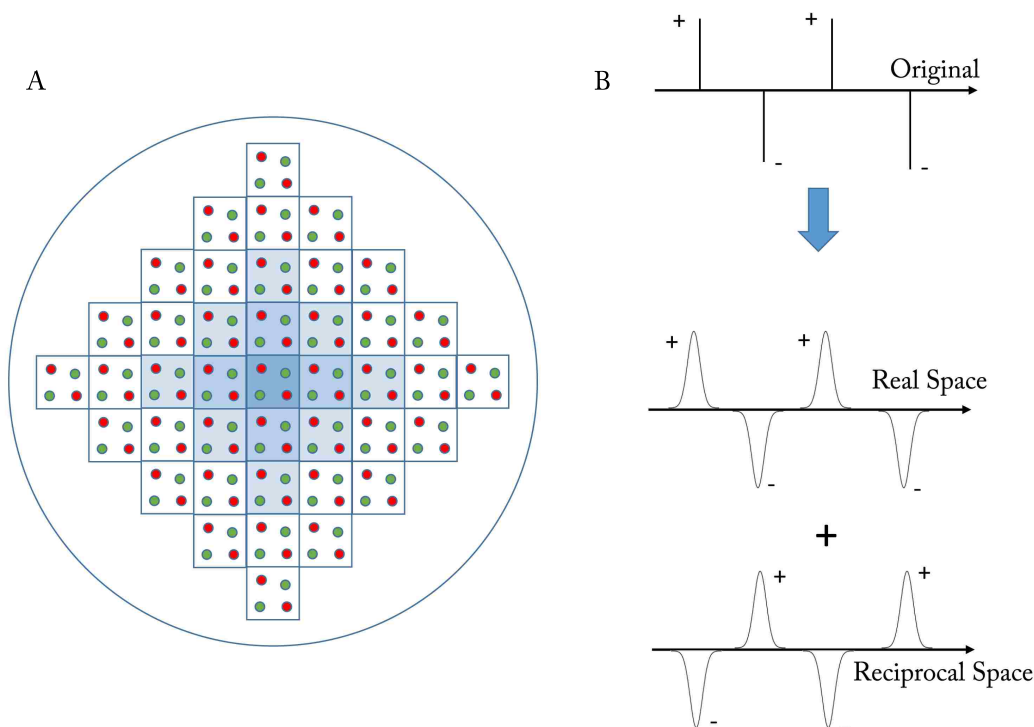
Figure 2.3: **A.** Construction of a system of periodic cells in the Ewald method. **B.** In the Ewald summation method, the initial set of charges are replaced by a Gaussian distribution (calculated in real space) to which a canceling charge distribution must be added (calculated in the reciprocal space).

2.4.2 Calculation of long-range Forces

Interactions which decay no faster than r^{-n}, n being the dimensionality of the system, are problematic as their range is often greater than half the box length. The charge-charge interaction, which decays as r^{-1}, is particularly problematic in molecular simulations. With the increase in computational power, various rigorous ways of dealing with long-range forces can be considered, even in simulations of large systems. A variety of methods have been developed to handle long-range forces [87]. The method of Ewald summation, is discussed below:

The Ewald Summation Method

In this method, a particle interacts with all the other particles in the simulation box and with all of their images in an infinite array of periodic cells. Fig. 2.3A illustrates how the array of the simulation cells is constructed; in the limit, the cell array is considered to have a spherical shape. The position of each image box (assumed to be cube of side L containing N charges) can be related to the central box by specifying a vector, each of whose components is an integral multiple of the length of the box, ($\pm iL, \pm jL, \pm kL$) ; $i, j, k = 0, 1, 2, 3$,etc. The charge-charge contribution to the potential energy due to all pairs of charges in the central simulation box can be written:

$$U(r_{ij}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{2.18}$$

where r_{ij} is the minimum distance between the charges i and j. There are six boxes at a distance L from the central box with coordinates ($\mathbf{r}_{box}$) given by $(0, 0, L)$, $(0, 0, -L)$, $(0, L, 0)$, $(0, -L, 0)$, $(L, 0, 0)$ and $(-L, 0, 0)$ (only four of these are shown in the two-dimensional picture in Figure 2.3A). The contribution of the charge-charge interaction between the charges in the central box and all images of all particles in these six surrounding boxes is given by:

$$U(r_{ij}) = \frac{1}{2} \sum_{\text{n box} = 1}^{6} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 |\mathbf{r}_{ij} + \mathbf{r}_{box}|}. \tag{2.19}$$

In general, for a box which is positioned at a cubic lattice point $\mathbf{n}$ ($= (n_x L, n_y L, n_z L)$ with n_x, n_y, n_z being integers):

$$U(r_{ij}) = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 |\mathbf{r}_{ij} + \mathbf{n}|}. \tag{2.20}$$

$| \mathbf{n} |$ thus takes values $1, \sqrt{2}, \ldots$. This expression is often written in such a way as to incorporate the interactions between pairs of charges in the central box (for which $| \mathbf{n} |= 0$):

$$U(r_{ij}) = \frac{1}{2} \sum_{|\mathbf{n}|=0}^{\prime} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 |\mathbf{r}_{ij} + \mathbf{n}|}. \tag{2.21}$$

The prime on the first summation indicates that the series does not include the interaction $i = j$ for $\mathbf{n} = 0$.

There is, thus, a contribution to the total energy from the interactions in the central box together with the interactions between the central box and all image boxes. Also, there is a contribution from the interaction between the spherical array of boxes and the surrounding medium. However, eq. 2.21 converges extremely slowly and is *conditionally convergent*. A conditionally convergent series is a mixture of positive and negative terms such that the positive terms alone form a divergent series (i.e. a series which does not have a finite sum) as do the negative terms when taken alone. The sum of a conditionally convergent series depends on the order in which its terms are considered. The rapid variation at small distances also poses a problem while considering the Coulomb interaction.

An elegant way while calculating the Ewald sum is to convert the summation into two series, each of which converges much more rapidly. The mathematical foundation for this is the following identity:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}. \tag{2.22}$$

The aim is now to choose an appropriate function $f(r)$ which will deal with the rapid variation of $\frac{1}{r}$ at small r and the slow decay at long r. In the Ewald method each charge is considered to be surrounded by a neutralizing charge distribution of equal magnitude but of opposite sign as shown in figure 2.3B . A Gaussian charge distribution of the following functional form is commonly used

$$\rho_i(\mathbf{r}) = \frac{q_i\alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2). \qquad (2.23)$$

The sum over point charges is now converted to a sum of the interactions between the charges *plus* the neutralizing distributions. The dual summation (the 'real space' summation) is given by:

$$U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{|\mathbf{n}|=0}' \frac{q_i q_j}{4\pi\epsilon_0} \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|}, \qquad (2.24)$$

where erfc is the complementary error function, defined as

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt. \qquad (2.25)$$

The Ewald method uses $\mathrm{erfc}(r)$ for the function $f(r)$ in eq. 2.22. The crucial point is that this new summation involving the error function converges rapidly and beyond some cutoff distance its value can be considered negligible. The rate of convergence depends upon the width of the cancelling Gaussian distributions; the wider the Gaussian, the faster the series converges. Specifically, α should be chosen so that the only terms in the series 2.24 are those for which $| \mathbf{n} |= 0$ (i.e. only pairwise interactions involving charges in the central box, or if a cutoff is used α is chosen so that only interactions with other charges within the cutoff are included). A second charge

distribution is added to the system which exactly counteracts the first neutralizing (Fig. 2.3B). The contribution from this second charge distribution is

$$U = \frac{1}{2} \sum_{k \neq 0}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\epsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}). \tag{2.26}$$

This summation is performed in *reciprocal space*. The vectors $\mathbf{k}$ are reciprocal vectors and are given by $\mathbf{k} = 2\pi\mathbf{n}/L$. This reciprocal sum which corresponds to the second term in eq. 2.23, too, converges much more rapidly than the original point-charge sum. However, the number of terms that must be included increases with the width of the Gaussians. There is thus a clear need to balance the real-space and reciprocal-space summations; the former converges more rapidly for large α, whereas the latter converges rapidly for small α. The Fourier transform of this summation can be represented by a small number of reciprocal vectors. The sum of Gaussian functions in real space includes the interaction of each Gaussian with itself. A third self-term must therefore be subtracted:

$$U_{subtract} = -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\epsilon_0}. \tag{2.27}$$

A fourth correction term may also be required, depending upon the medium that surrounds the sphere of simulation boxes. If the surrounding medium has an infinite relative permittivity (as in the case of a conductor) then no correction term is required. However, if the surrounding medium is a vacuum (with relative permittivity of 1) then the following energy must be added:

$$U_{correction} = \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\epsilon_0} \mathbf{r}_i \right|^2. \tag{2.28}$$

The final expression is thus:

$$
U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \begin{array}{l} \displaystyle\sum_{|\mathbf{n}|=0}^{\infty}{}' \frac{q_i q_j}{4\pi\epsilon_0} \frac{\mathrm{erfc}(\alpha|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} \\[2ex] \displaystyle+\sum_{k\neq 0}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\epsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(\mathbf{k}\cdot\mathbf{r}_{ij}) \\[2ex] \displaystyle-\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\epsilon_0} + \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\epsilon_0} \mathbf{r}_i \right|^2 \end{array} \right. \tag{2.29}
$$

The Ewald sum is the most 'correct' way so far to accurately include all the effects of long-range forces in a computer simulation. It has been extensively used [151, 50] in simulations involving highly charged systems and is increasingly being applied to other systems where electrostatic effects are important, such as proteins, DNA and lipid bilayers. However, the Ewald summation is computationally expensive. Several approaches have been proposed such as the use of fast Fourier transform (FFT) to compute the reciprocal space summation. However, the FFT method requires that the data are not continuous but are discrete values. In order to employ FFT in the Ewald summation, the point charges with their continuous coordinates must be replaced by a grid-based charge distribution. Each of the atomic charges are distributed among the surrounding grid points in some fashion so as to reproduce the potential of the charge at the original location. A number of variants on this general idea exist, such as the particle-mesh Ewald (PME) method [30] and the particle-particle particle-mesh Ewald summation [93]. The PME summation method has been extensively used to generate most of the results in the subsequent chapters.

2.5   Enhanced Sampling Methods

For a comprehensive understanding of molecular functions, it is imperative to adequately sample the conformational space of biomolecules. However, most molecular dynamics trajectories involving complex biomolecular systems are not ergodic and leave many relevant regions of the conformational space unexplored. Typically, these systems have *frustrated* or *rugged energy landscapes*. At low temperatures, it is difficult for the system to move between the relevant low-energy regions and thus the probabilities of the system surmounting the energy barriers are almost negligible. Surmounting energy barriers, are thus, *rare events* at low temperatures.

A wide range of methods [161, 1] exist which are designed to accelerate equilibration in simulations of systems with rugged energy landscapes. Two such enhanced sampling techniques are discussed below:

2.5.1   Replica Exchange Molecular Dynamics

The replica exchange molecular dynamics simulation involves running MD simulations at constant temperatures in parallel on a set of replica systems, each at different temperatures $\{T_0, T_1, T_2, ..., T_k\}$ where the temperatures are ordered from the lowest T_0 to the highest T_k. After every certain number of time steps, an attempt is made to exchange configuration of a pair of neighboring replicas, and this exchange is accepted satisfying the condition of detailed balance. In general, different replicas can have not only different temperatures but also different potential functions $\{U_0(\mathbf{r}_0), U_1(\mathbf{r}_1), U_2(\mathbf{r}_2), ..., U_k(\mathbf{r}_k)\}$ where $\mathbf{r}_i$ represents the configurational coordinates of the ith replica system. It is assumed that the probability of configuration $\mathbf{r}_i$ in the ith replica obeys the Boltzmann distribution. Therefore:

$$P(\mathbf{r}_i) = \frac{\exp(-\beta_i U_i(\mathbf{r}_i))}{Z_i}, \tag{2.30}$$

where $\beta = (k_B T_i)^{-1}$ and Z_i is the partition function. The joint probability distribution of the extended K-replica system is thus given by:

$$P = \prod_i^K P_i(\mathbf{r}_i) = \prod_i^K \frac{\exp\{-\beta_i U_i(\mathbf{r}_i)\}}{Z_i}. \tag{2.31}$$

At this point, the transition probability that the configuration $\mathbf{r}_m$ of replica m (with potential energy function U_m) is exchanged with the configuration $\mathbf{r}_n$ of replica n (with potential energy function U_n) is introduced. This probability is denoted by $w(\mathbf{r}_m, \beta_m, U_m; \mathbf{r}_n, \beta_n, U_n)$ and the probability of the reverse process is denoted by $w(\mathbf{r}_n, \beta_n, U_n; \mathbf{r}_m, \beta_m, U_m)$.

At equilibrium ,the detailed balance condition for the system is satisfied. That is, before and after exchange:

$$P_{m\to n} \times w(\mathbf{r}_m, \beta_m, U_m; \mathbf{r}_n, \beta_n, U_n) = P_{n\to m} \times w(\mathbf{r}_n, \beta_n, U_n; \mathbf{r}_m, \beta_m, U_m), \tag{2.32}$$

holds true. Here, $P_{m\to n}$ $(P_{n\to m})$ is the joint probability of the extended K-replica system to be in a state such that the configuration $\mathbf{r}_m$ $(\mathbf{r}_n)$ of replica $m(n)$ is exchanged with configuration $\mathbf{r}_n$ $(\mathbf{r}_m)$ of replica n (m).

This leads to

$$\frac{w(\mathbf{r}_m, \beta_m, U_m; \mathbf{r}_n, \beta_n, U_n)}{w(\mathbf{r}_n, \beta_n, U_n; \mathbf{r}_m, \beta_m, U_m)} = \frac{P_{n\to m}}{P_{m\to n}}. \tag{2.33}$$

The probabilities on the right hand side of eq.2.33 can be expressed as:

$$P_{m \to n} = \frac{\exp\{-\beta_m U_m(\mathbf{r}_m)\}}{Z_m} \times \frac{\exp\{-\beta_n U_n(\mathbf{r}_n)\}}{Z_n} \times \prod_{i \neq m,n}^{K} \frac{\exp\{-\beta_i U_i(\mathbf{r}_i)\}}{Z_i},$$

$$P_{n \to m} = \frac{\exp\{-\beta_m U_m(\mathbf{r}_n)\}}{Z_m} \times \frac{\exp\{-\beta_n U_n(\mathbf{r}_m)\}}{Z_n} \times \prod_{i \neq m,n}^{K} \frac{\exp\{-\beta_i U_i(\mathbf{r}_i)\}}{Z_i}.$$

$$(2.34)$$

For temperature replica exchange, all the replicas have the same potential energy function and thus the subscripts of U_i are dropped. Also the transition probabilities are expressed as $w(\mathbf{r}_m, \beta_m; \mathbf{r}_n, \beta_n)$ instead of $w(\mathbf{r}_m, \beta_m, U_m; \mathbf{r}_n, \beta_n, U_n)$

Hence from eqs. 2.33 and 2.34:

$$\frac{w(\mathbf{r}_m, \beta_m; \mathbf{r}_n, \beta_n)}{w(\mathbf{r}_n, \beta_n; \mathbf{r}_m, \beta_m)} = \exp(-\Delta_{mn}), \tag{2.35}$$

where

$$\Delta_{mn} = (\beta_m - \beta_n)\{U(\mathbf{r}_m) - U(\mathbf{r}_n)\}. \tag{2.36}$$

For the exchange between the adjacent replicas, the Metropolis acceptance criterion is adopted. Thus:

$$w(\mathbf{r}_m, \beta_m; \mathbf{r}_n, \beta_n) = \begin{cases} 1 & \text{for } \Delta_{mn} \leq 0 \\ \exp(-\Delta_{mn}) & \text{for } \Delta_{mn} > 0 \end{cases} \tag{2.37}$$

In order to perform replica exchange molecular dynamics simulations usually a *parallel computing scheme* is used, in which a specific number of computer nodes (or processors) is assigned to each replica.

How do we choose the number of replicas? Let us consider a swap between two adjacent temperatures T_m and T_n and that $T_n > T_m$. We expect that the configuration drawn from the higher temperature will have a higher energy. Thus, $U(\mathbf{r}_n) > U(\mathbf{r}_m)$ and $\beta_m > \beta_n$. Hence $\Delta_{mn} > 0$ and thus from eq. 2.39, the exchange probability will decrease, with increasing difference between the adjacent temperatures. Hence, the temperatures must be placed close enough in order to achieve a good rate of swaps. The usual protocol is to choose $T_{\min}$ as the temperature at which we intend to exhaustively sample the conformational space of a biomolecule. $T_{\max}$ should be chosen high enough such that free energy barriers can be crossed, but not so high as to require many intermediate temperatures.

2.5.2   Replica Exchange with Solute Tempering (REST)

The replica exchange method (REM) has been extremely popular in the sampling of biomolecular systems [18, 138, 53, 24]. However, the REM method is limited to small systems, since the number of replicas required scales as $O(\sqrt{N})$ where N is the number of degrees of freedom of the system. A large number of replicas are required to simulate systems with large systems of physiological relevance. The main reason for this is that the overall Hamiltonian grows with the system size. The acceptance probability for the exchange between two adjacent replicas is $\exp(\Delta\beta\Delta E)$, a quantity that depends exponentially on the change in energy. Thus for a large system a smaller $\Delta\beta$ must be chosen to obtain viable acceptance probabilities, which is obtained by choosing replica systems spaced more closely in temperature and, thus, more replicas to cover a given range of upper and lower temperatures. Replica exchange with solute tempering (REST) is a method in which the exchange depends on the change in energy of a small part of the system, achieving a sufficiently large acceptance

probability for replicas with widely separated temperatures, thus reducing the need for a large number of replicas.

For simulations of biomolecular assemblies, it is desirable to explicitly include water molecules primarily because water plays an important role in various physiological processes [122, 70] and also, implicit solvent models cannot quite capture the thermodynamics as comprehensively as explicit solvent models [156, 51]. Usually in solute-solvent systems, the primary interest is the conformational ensemble of the solute. REST originally proposed by Liu *et al.* in [91] and later improved by Wang *et al.* in [145] is based on the idea that the potential energy of the system scales with temperature in such a way that the molecule of interest appears to get hotter, "but the water stays cold as one climbs the replica ladder". The acceptance probability for replica exchange scales only with the number of degrees of freedom of the solute molecule and not the water molecules. Since water molecules contribute significantly to the Hamiltonian, REST holds promise for simulation of large, physiologically relevant biomolecular systems.

Methodology

For the more general case where every replica has a different potential energy function, eqs. 2.33, 2.34 and 2.35 yields

$$\Delta_{mn} = -\beta_m \{U_m(\mathbf{r}_m) - U_m(\mathbf{r}_n)\} - \beta_n \{U_n(\mathbf{r}_n) - U_n(\mathbf{r}_m)\} \tag{2.38}$$

A biomolecular system such as a protein solution consists of a protein molecule (labeled p) dissolved in a large number of water molecules (labeled w), the system is decomposed into two parts. The protein is taken as the central part(p) dissolved in a bath of water molecules(w). The potential energy of the system can be written as

$$U_0(\mathbf{r}) = U_{\mathrm{p}}(\mathbf{r}) + U_{\mathrm{pw}}(\mathbf{r}) + U_{\mathrm{ww}}(\mathbf{r}). \tag{2.39}$$

where U_{p}, U_{pw} and U_{ww} are, respectively, the intramolecular energy of the protein, the interaction energy between the protein and water, and the interaction of water molecules with each other. Usually, the first two terms depend on only a relatively small set of coordinates compared with the last term.

Now, the lowest replica is considered to be the replica of interest with the potential energy given by eq. 2.39 at temperature T_0. This replica is labeled by the index 0. For the replicas at higher temperatures the potential energy surface is rescaled. According to the original implementation (hereafter referred to as REST1), [91], the replica at temperature T_m has the following potential energy:

$$U_{\mathrm{m}}^{\mathrm{REST1}}(\mathbf{r}) = U_{\mathrm{p}}(\mathbf{r}) + \frac{\beta_0 + \beta_m}{2\beta_m} U_{\mathrm{pw}}(\mathbf{r}) + \frac{\beta_0}{\beta_m} U_{\mathrm{ww}}(\mathbf{r}). \tag{2.40}$$

Substituting eq. 2.40 in 2.38:

$$\Delta_{mn}(\mathrm{REST1}) = (\beta_m - \beta_n)\left[\left\{U_{\mathrm{p}}(\mathbf{r}_n) + \frac{1}{2}U_{\mathrm{pw}}(\mathbf{r}_n)\right\} - \left\{U_{\mathrm{p}}(\mathbf{r}_m) + \frac{1}{2}U_{\mathrm{pw}}(\mathbf{r}_m)\right\}\right]. \tag{2.41}$$

It is to be noted that U_{ww}, interaction energy between the water molecules does not appear in eq. 2.41 and this is the reason only a relatively small number of replicas are sufficient to achieve good exchange probabilities in REST1.

In REST1 both the potential energy and the temperature are different for different replicas. According to the law of corresponding states, the thermodynamic properties of a system with potential energy E_m at temperature T_m. are the same as those for a system with potential energy $(T_0/T_m)E_m$ at temperature T_0. Using this concept in

the variant REST [145] method (hereafter referred to as REST2), all of the replicas are run at the same temperature T_0, but the potential energy for replica m is scaled differently:

$$U_{\mathrm{m}}^{\mathrm{REST2}}(\mathbf{r}) = \frac{\beta_m}{\beta_0} U_{\mathrm{p}}(\mathbf{r}) + \sqrt{\frac{\beta_m}{\beta_0}} U_{\mathrm{pw}}(\mathbf{r}) + U_{\mathrm{ww}}(\mathbf{r}). \tag{2.42}$$

The interaction energies in eq. 2.42 is achieved by scaling the bonded interaction energy terms, the Lennard-Jones (LJ) ϵ parameters and the charges of the solute atoms by (β_m/β_0), (β_m/β_0) and $(\beta_m/\beta_0)^{1/2}$ respectively, and the scaling factor for the U_{pw} term is a consequence of the combination rules used to calculate the cross-interaction parameters for LJ interactions. The different scaling factor for the U_{pw} term in REST2 as compared to REST1 is reported to increase the efficiency of the REST method [145]. The acceptance ratio in REST2 between replicas m and n is determined by:

$$\Delta_{mn}(\mathrm{REST2}) = (\beta_m - \beta_n) \Big[\big\{ U_{\mathrm{p}}(\mathbf{r}_n) - U_{\mathrm{p}}(\mathbf{r}_m) \big\} + \frac{\sqrt{\beta_0}}{\sqrt{\beta_m} + \sqrt{\beta_n}} \big\{ U_{\mathrm{pw}}(\mathbf{r}_n) - U_{\mathrm{pw}}(\mathbf{r}_m) \big\} \Big]. \tag{2.43}$$

REST2 has been implemented in GROMACS [61] as described in [136]. The implementation involves defining two end potentials U_L and U_H as:

$$U_L = \frac{\beta_L}{\beta} U_p + \sqrt{\frac{\beta_L}{\beta}} U_{pw} + U_{ww}$$

$$U_H = \frac{\beta_H}{\beta} U_p + \sqrt{\frac{\beta_H}{\beta}} U_{pw} + U_{ww} \tag{2.44}$$

where β is the inverse of the system temperature, β_L is the inverse of the lowest effective solute temperature, which is usually the same as the system temperature (thus $\beta_L = \beta$), and β_H is the inverse of the highest effective solute temperature. Using the above potentials as two-ends, we now introduce many potentials as the linear combination of these two end-potentials. The potential energy function for the ith replica is defined by:

$$U_i = (1-\lambda_i)U_L + \lambda_i U_H = \frac{\beta_L(1-\lambda_i) + \beta_H}{\beta}U_p + \left(\sqrt{\frac{\beta_L}{\beta}}(1-\lambda_i) + \sqrt{\frac{\beta_H}{\beta}}\lambda_i\right)U_{pw} + U_{ww}.$$
(2.45)

The replica with $\lambda_i = 0$ corresponds to the Hamiltonian of our interest and all the other replicas are artificial ones introduced solely for the purpose of solute tempering.

To study the sampling efficiency of REST2 we have performed simulations with two different peptides in explicit solvent: a. Alanine dipeptide which is the simplest model peptide that has been extensively studied by both experimental and theoretical methods [59]. For proof of concept, we have compared the results against standard molecular dynamics and REMD simulations. b. To study the β-sheet propensity, the GB1-hairpin has been simulated and the results are compared against REMD simulations.

Simulation Details

a. Alanine Dipeptide

The Amber03w forcefield[16] was used to model the peptide along with the TIP4P/2005 model for the 708 water molecules used to solvate the peptide. For the solute tempering simulations, the two end-potentials are chosen such that $T_L =$

300 K and $T_H = 595$K. The two intermediate 'λ-points' chosen are $\lambda = \{0.3, 0.65\}$. Thus in all we have simulated four replicas in parallel. For the REMD simulations, we have used 32 replicas spanning 300 - 650 K. Using eq. 2.45, we have scaled the Hamiltonians such that the 300 K replica corresponds to the 'λ-points' chosen for the solute tempering simulations. Using these scaled Hamiltonians, we have performed three additional REMD simulations with 32 replicas spanning 300 - 650 K for comparison. Finally, we have performed four standard molecular dynamics simulations using the scaled Hamiltonians (corresponding to the 'λ-points') at 300 K. All the simulations were performed in cubic box with side 2.8 $\mathring{A}$ using periodic boundary conditions in the x, y,and z directions. The Langevin thermostat has been used with a friction coefficient of 1 ps^{-1}. The Particle Mesh Ewald method [41] has been used to calculate electrostatic interactions with a real space cutoff of 1.0 nm. The cutoff for van der Waals interactions was also taken to be 1.0 nm. All the simulations were performed for sufficient time to attain convergence.

b. GB1-hairpin

For the GB1-hairpin simulations the Amber03* [17] forcefield has been used along with TIP3P model [72]for water molecules. For the solute tempering simulations using the GB1-hairpin, the 'λ-points' chosen are $\lambda = \{0, 0.137, 0.273, 0.409, 0.545, 0.682, 0.841, 1.0\}$. The end potentials are chosen such that $T_L =$300 K and $T_H =$520 K. Other details of the simulations can be found in [18].

Results and Discussion

Figure 2.4 shows the probability distribution functions of the potential energies obtained from molecular dynamics (MD), replica exchange with solute tempering (REST2) and replica exchange molecular dynamics (REMD) simulations. At the
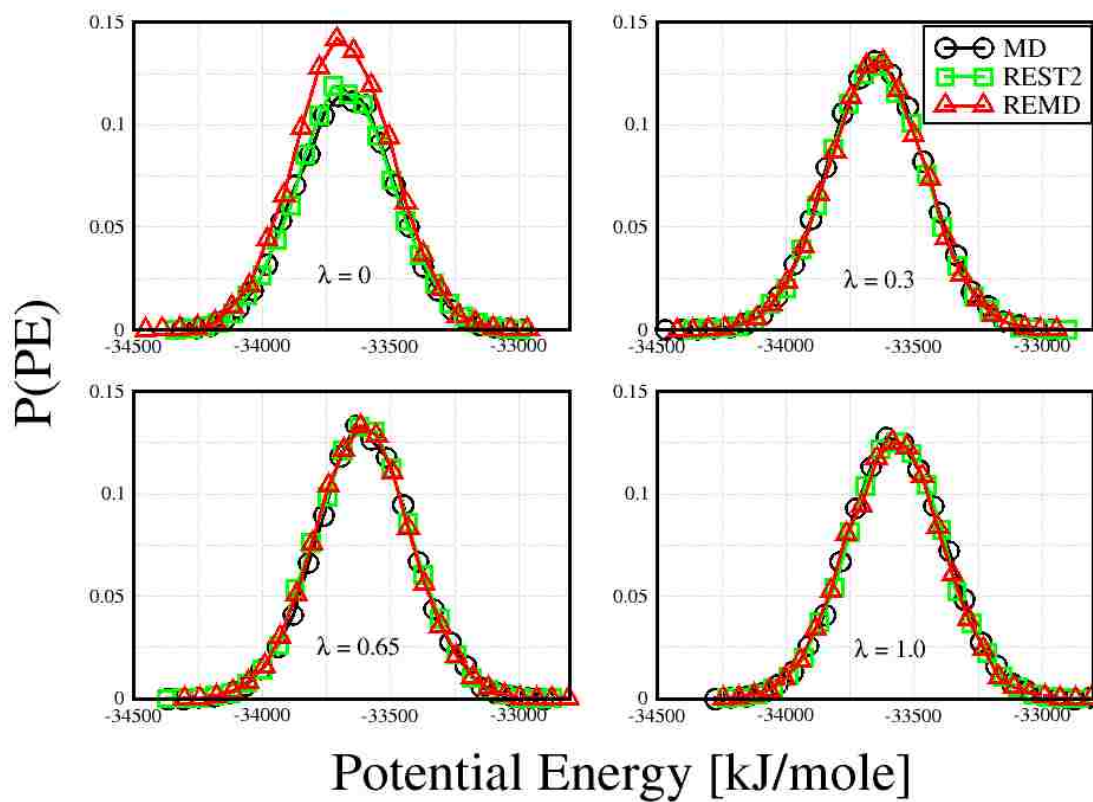
Figure 2.4: The probability distributions of the potential energies obtained from Molecular Dynamics, Replica Exchange with Solute Tempering-2 and Replica Exchange Molecular Dynamics at different 'λ'-points
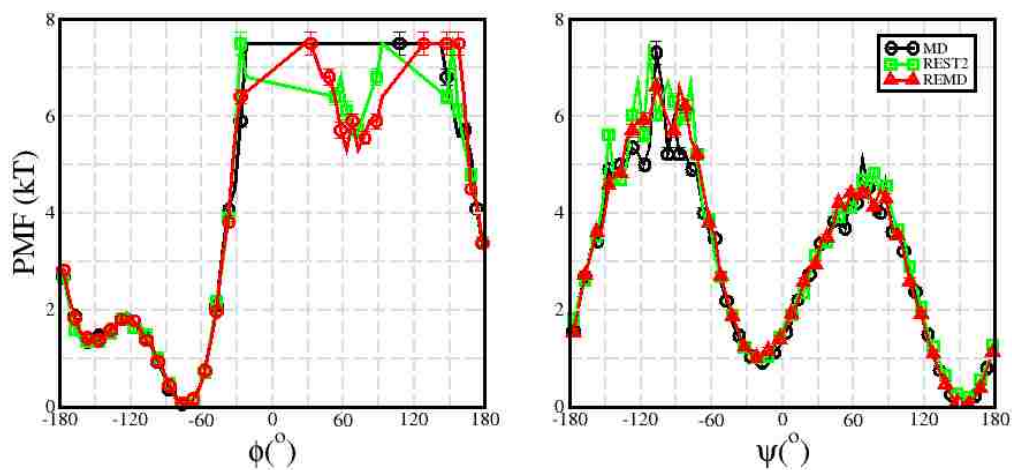
Figure 2.5: One dimensional free energy based on the ϕ (left) and ψ (right) angles of the alanine dipeptide
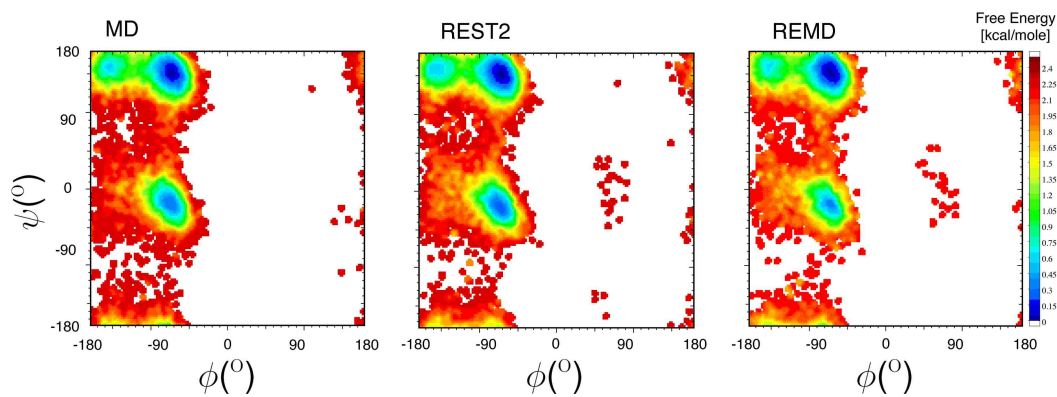
Figure 2.6: Conformational free energies at 300 K in the Ramachandran space for alanine dipeptide obtained from Molecular Dynamics, Replica Exchange with Solute Tempering-2 and Replica Exchange Molecular Dynamics simulations.
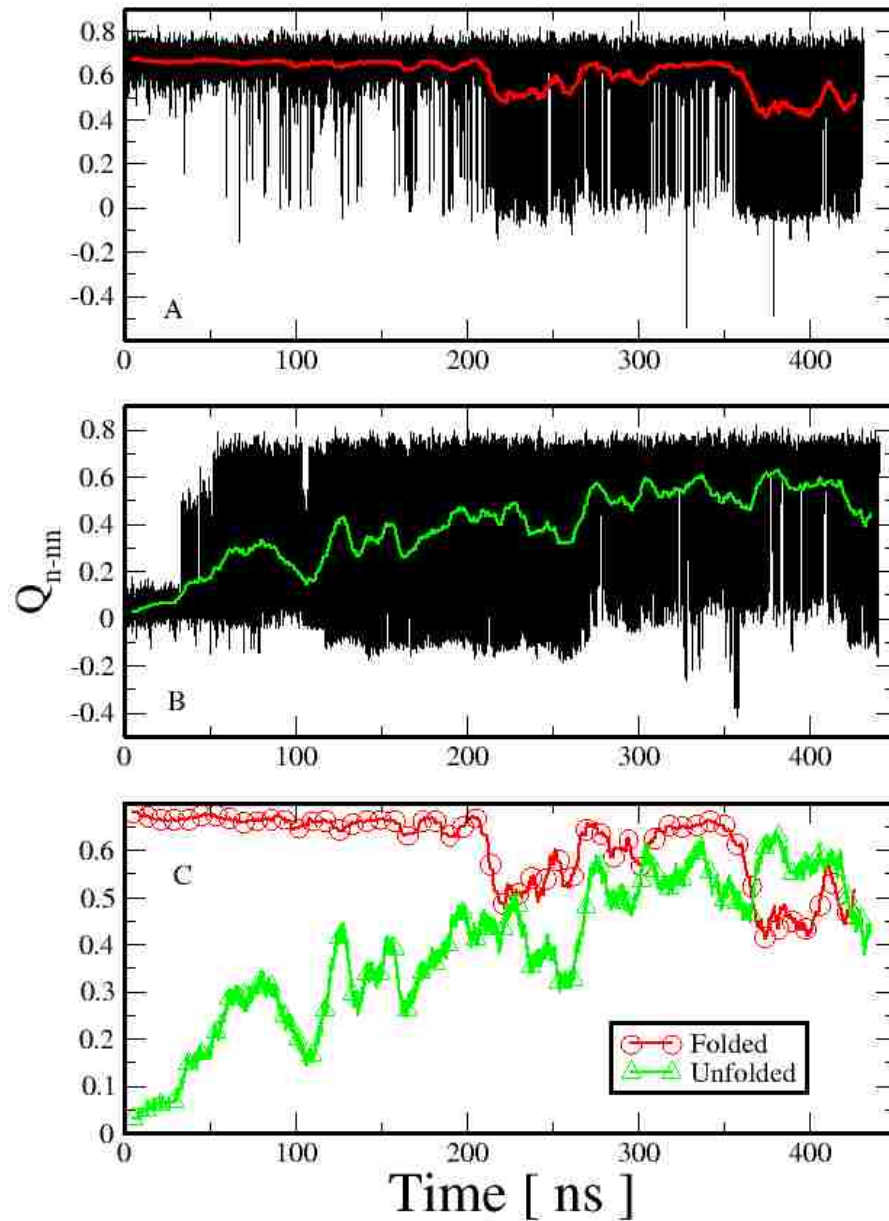
Figure 2.7: Time series of fraction of native minus non-native contacts Q_{n-nn} of the GB1 hairpin obtained from Replica Exchange with Solute Tempering - 2 simulations initiated from **A.** the folded state and **B.** the unfolded state of the hairpin. The guides are window averages obtained with a window size of 10ns. **C.** The window averages are re-plot for comparison.
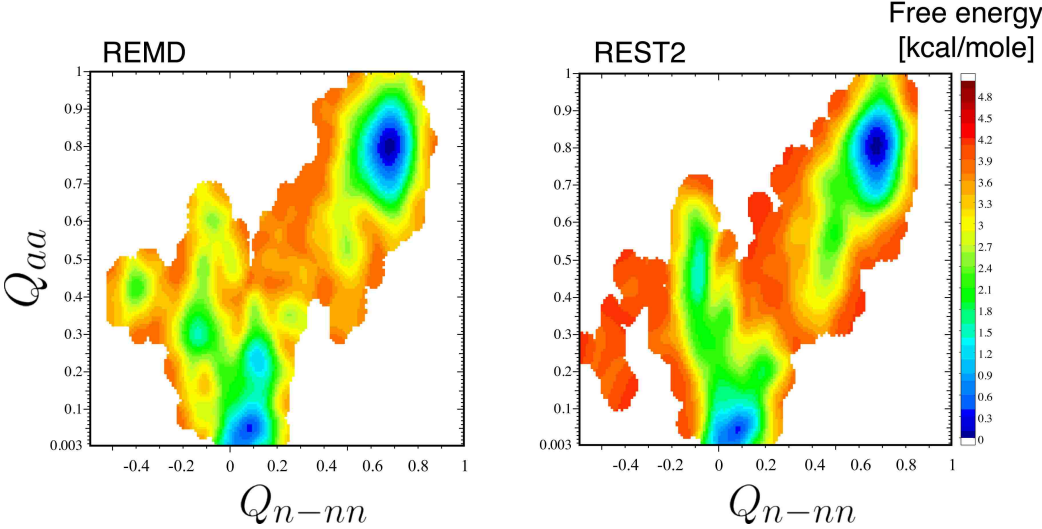
Figure 2.8: The two-dimensional potential of mean force based on Q_{aa} and Q_{n-nn} obtained from Replica Exchange Molecular Dynamics (left) and Replica Exchange with Solute Tempering - 2 (right) simulations.

temperature of interest (300 K / $\lambda = 0$), it can be observed that the distributions obtained from the MD and the REST2 simulations are almost exactly same. The peak of the distribution obtained from the REMD simulation is slightly shifted owing to a better sampling obtained. However, the mean values obtained from the distributions are the same. It should be noted that at the other λ-points the distributions exactly overlap with each other.

Figure 2.5 shows the free energies based on ϕ (left) and ψ (right) angles of alanine dipeptide at 300 K. The characteristic minima of the free energy profile obtained from REMD simulations based on ϕ are $\phi \approx -75º$, $\phi \approx -150º$ and $\phi \approx 75º$. As can be observed, the profiles obtained from REST2 and MD simulations compare well with the one obtained using REMD simulation at $\phi \approx -75º$, and $\phi \approx -150º$. Moreover, the barrier around $\phi \approx -120º$ between the two states overlap with each other. However, the region around $\phi = 75º$ is not sampled using MD simulation.

The characteristic minima based on ψ are $\psi \approx -30^{\circ}$ and $\psi \approx 150^{\circ}$. As is observed, the free energy profiles obtained from all three simulations overlap with each other around these minima. However, the barrier around $\psi \approx 110^{\circ}$ is not sampled well enough using MD simulation.

Figure 2.6 shows the Ramachandran map for the alanine dipeptide. It is observed that the characteristic minima near the polyprolene-II helix region ($-100^{\circ} < \phi < -50^{\circ}, 120^{\circ} < \psi < 170^{\circ}$), the right-handed α-helical region, α_R, ($-80^{\circ} < \phi < -60^{\circ}, -40^{\circ} < \psi < -10^{\circ}$) and the β-sheet region ($-170^{\circ} < \phi < -150^{\circ}, 150^{\circ} < \psi < 170^{\circ}$) are sampled with nearly equal propensities from all the three simulations. However, there are subtle differences near the α_L-region ($60^{\circ} < \phi < 80^{\circ}, -30^{\circ} < \psi < 30^{\circ}$). REST2 is found to sample the conformational space with nearly equal propensity as REMD simulations.

To check convergence we have performed two sets of REST2 simulations with the GB1-hairpin, one initiated from the unfolded state ($Q_{n-nn} = 0$) and the other initiated from the folded state ($Q_{n-nn} = 1$) of the peptide. Figure 2.7 shows the timeseries of Q_{n-nn} for the $\lambda = 0$ replica. The red and the green guides in figure 2.7 A and B respectively, show the window averaged value with a window size of 10ns. In figure 2.7 C, the two window averaged data from the timeseries are plot for comparison. It appears that using REST2 the system converges after ≈ 200 ns of simulation time per replica. Using REMD, the simulations were performed for 0.5 μs per replica [18], which is the same as the length of REST2 simulations performed here. The computational efficiency of REST2 in which just 8 replicas have been used is thus quite significant as compared to REMD in which simulations have been performed with 32 replicas.

The free energy landscape of the GB1-hairpin is characterized by two parameters the fraction of heavy atom native contacts, Q_{aa}, [12, 15, 16] and an alternative coordinate, the fraction of native minus non-native contacts, Q_{n-nn}. These coordinates are defined in section 5.2.2. It is observed in in figure 2.8 that the folded $Q_{n-nn} = 0.7$ and the unfolded states $Q_{n-nn} \approx 0$ are sampled sufficiently using REST2. However, when compared to REMD simulations, the misfolded state characterized by $Q_{n-nn} \approx -0.4$ is sampled quite poorly. However, quite remarkably, REST2 can be used to fold complex peptides such as

Conclusion

Molecular dynamics simulations coupled with enhanced sampling methods enables us to comprehensively sample the conformational space of the biomolecules under study. Replica exchange molecular dynamics has been used quite extensively to study the behavior of proteins [82], under physiologically relevant conditions. One of the major drawbacks of REMD is the need for a large number of replicas to effectively sample the conformational space of proteins. As the number of replicas required increases with larger systems, REMD can be somewhat preventive while studying large systems of physiological relevance. Replica exchange with solute tempering, provides an excellent alternative. Although there are apparent differences in the estimates of free energies obtained from REMD and REST2, REST2 can be used to obtain the correct folded structure of the GB1 hairpin in close agreement with the native structure of the peptide. Future work should be directed towards validation of REST2 using various peptides and forcefields before it can be used to study larger systems of physiological relevance.

Chapter 3

Effects of Macromolecular Crowding on Protein-Protein interactions

3.1 Introduction

Many biological processes rely on protein-protein interactions in a highly crowded cellular environment; a cell is typically occupied up to 40% of its volume by various macromolecules, such as DNA, RNA, proteins, sugars and other organelles. Consequently, it is necessary for a protein to move around crowding macromolecules to find its way to bind its interaction partner(s) to carry out some specific biological function. Therefore it is important to understand the effects of macromolecular crowding on protein-protein interactions to better understand biological processes in a living cell.

In recent years, significant progress has been made in understanding protein-protein interactions at the molecular level by both experimentation and simulation

[27]. However, most studies have been performed in dilute solutions - *in vitro* or *in silico*, which are not representative of *in vivo* conditions. This raises some important following questions - (i) how can we understand protein-protein interactions in a living cell given the difficulties in designing experiments and simulations that attempt to mimic the *in vivo* environment? (ii) Can we utilize the enormous information collected over the years on protein-protein interactions in dilute solutions and relate it to protein-protein interactions in a crowded cellular environment?

Studying protein-protein interactions via *in vivo* experiments is extremely challenging though in-cell spectroscopic techniques are making great progress in terms of achieving this goal [146]. Thus, to answer the above questions, many experimental studies have been performed using synthetic polymers or specific proteins as crowding agents. These studies have addressed both the thermodynamics and kinetics of protein-protein interactions in a crowded environment [99, 98, 69, 142, 148, 106, 110, 85, 160, 115, 146, 44, 114]. Using "inert" crowding agents, the primary focus in most of these studies had been to understand the non-specific excluded-volume effects of crowding agents on the formation of protein complexes [98, 78]. As expected based on simple theoretical models, these experiments have shown that the consequence of excluded-volume effects (entropic in nature) is to force proteins to form a stable complex thereby increasing the volume available to the crowding agents. However, few experiments have shown an unexpected trend, i.e., destabilization of protein complexes in the presence of crowding agents [115, 114, 71]. This observation can be explained by attractive interactions between proteins and crowding agents, which can inevitably arise from a combination of electrostatic interactions, hydrogen bonding, hydrophobic interactions, and van der Waals interactions. As opposed to the stablizing effect of entropic-excluded volume interactions, the attractive protein-crowders

will actually increase the binding free energy due to the enthalpic penalty in breaking favorable protein-crowder contacts to form a protein complex. The separation of these competing effects in an experimental setup is quite challenging due to unknown interaction parameters and presents a major barrier in developing a theoretical model to interpret (and even predict *a* priori) experimental observations.

Computational models can help form a basis of our understanding in separating these various effects as the interactions between proteins and crowding agents can be tuned precisely. Earlier computational studies based on purely repulsive protein-crowder interactions (such as hard-sphere as well as soft repulsive interactions) have shown that the repulsive interactions stabilize the formation of the protein complex by lowering the binding free energy. However, the extent of such stabilization has been found to be rather modest [78]. Interestingly, the population of nonspecific encounter complexes, which are now believed to play an important role in forming the native functional complex, may well be decreased by the presence of repulsive crowders. Rosen *et al.* have considered attractive protein-crowder interactions in their protein binding simulations [124]. The work of these researchers has highlighted the importance of accounting for enthalpic effects arising from the attractive protein-crowder interactions (if present) in addition to the commonly invoked excluded volume effects [31]. Modest protein-crowder attractions can actually increase the binding free energy with respect to the crowder-free solution. Even if the binding free energy is decreased with respect to the crowder-free solution (weak protein-crowder attractions), the extent of this decrease is overestimated by theoretical models based solely on repulsive protein-crowder interactions.

Most analytical theories of macromolecular crowding are based on the scaled particle theory (SPT) of hard-sphere fluids developed almost half a century ago [88].

The theory is outlined in A. The SPT provides an analytical expression for the free energy cost of creating a spherical cavity in a bath of hard-sphere particles. By approximating proteins as spherical particles, these theories have been applied to interpret experimental [110, 128] and computational data with varying degrees of success [155, 78, 102]. However, these theories fail to account for any kind of attractive protein-crowder interactions and need to be modified before they can be applied to more realistic scenarios.

Modifications to account for protein-crowder attractions, have been proposed by Jiao *et al.* [71] and Rosen *et al.* [124]. These researchers added a phenomenological mean-field term (proportional to the protein surface area) to their respective SPT-based crowding theories to interpret experimental and simulation data respectively.

More recently, a microscopic theory based on statistical mechanics of simple liquids has been introduced to describe the protein-binding simulation data in the presence of attractive protein-crowder interactions [80]. Without any adjustable parameter in the model, this theory has been able to predict the change in binding free energy from the molecular simulation data remarkably well over a wide range of parameters, such as crowder size, packing fraction and the protein-crowder attraction strength. A brief summary of the efforts to understand thermodynamics of protein-protein interactions in a crowded environment using computational methods has been presented.

3.2 Thermodynamics of protein-protein interactions in a crowder solution

The thermodynamics of protein-protein interactions is based on Fig.3.1 for the case of a simple dimerization reaction, $A + B \rightleftharpoons AB$, in the presence of spherical crowders.
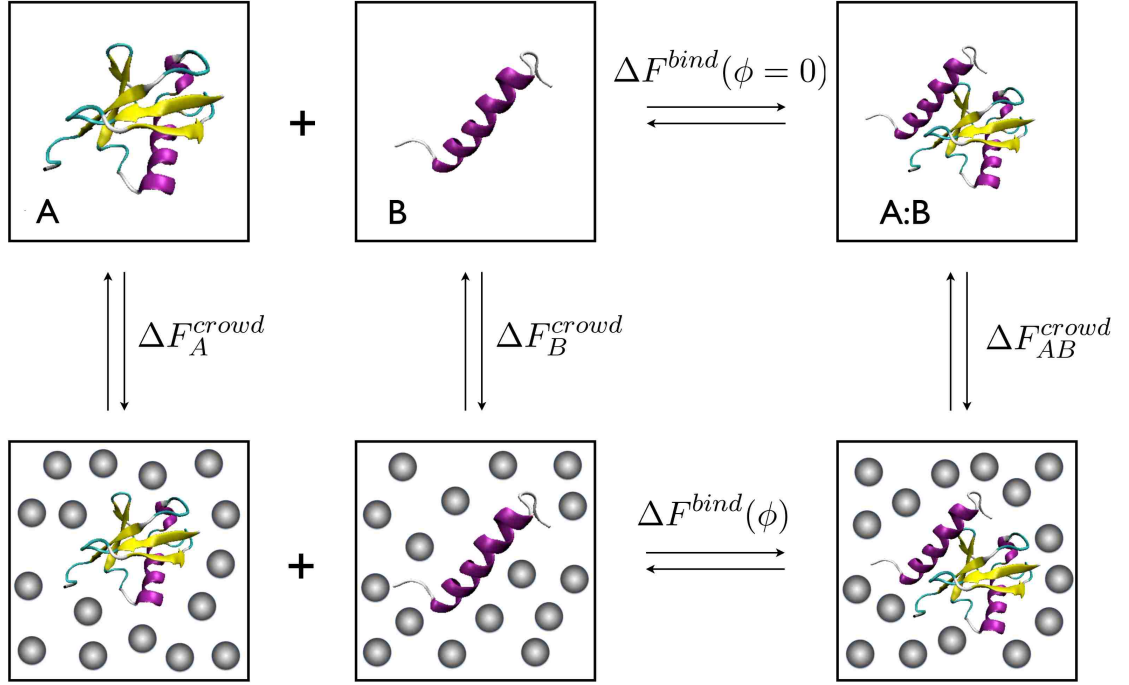
Figure 3.1: Thermodynamic cycle for the formation of a protein complex (between ubiquitin and UIM1) in bulk and in a crowded solution.

The thermodynamic quantity that is mostly sought in both experiments and simulations is the change in the binding free energy relative to that of the crowder-free solution ($\phi = 0$), $\Delta\Delta F^{\text{bind}}(\phi)$, in a crowded solution with crowder packing fraction ϕ (bottom horizontal reaction in Fig.3.1). Theoretically, given the various solvation (or crowding) free energies, $\Delta F_X^{\text{crowd}}$ ($X = A, B, AB$) - the free energy of inserting X in a crowded solution, $\Delta\Delta F^{\text{bind}}(\phi)$ is given by (using a thermodynamic cycle),

$$
\begin{aligned}
\Delta\Delta F^{\text{bind}}(\phi) &= \Delta F^{\text{bind}}(\phi) - \Delta F^{\text{bind}}(\phi = 0), \quad (3.1)\\
&= \Delta F_{AB}^{\text{crowd}} - \Delta F_A^{\text{crowd}} - \Delta F_B^{\text{crowd}}.
\end{aligned}
$$

A major challenge is thus to calculate the crowding free energy $\Delta F_X^{\text{crowd}}$ for a protein or a complex X.

The procedure for calculating the crowding free energy is described next. Without loss of generality, let $U_{\text{int}, X}(\Omega)$ be the overall interaction energy between a protein X and a crowder, where Ω denotes the collective variables of protein atoms and the crowders. Thus the crowding free energy, $\Delta F_X^{\text{crowd}}$, can be obtained by,

$$\exp\{-\beta \Delta F_X^{\text{crowd}}\} = \langle\langle \exp\{-\beta U_{\text{int, x}}\}\rangle\rangle, \tag{3.2}$$

where $\beta = 1/k_B T$ and $\langle\langle \cdot \rangle\rangle$ refers to the canonical ensemble average over protein and crowder configurations. For weakly-to-moderately interacting proteins, structural changes are minimal during binding events. Thus the double ensemble average in eq. 2 can be further approximated to,

$$\exp\{-\beta \Delta F_X^{\text{crowd}}\} \approx \langle \exp\{-\beta U_{\text{int, x}}\}\rangle_{\text{native}}, \tag{3.3}$$

where $\langle\cdot\rangle_{\text{native}}$ is the ensemble average over crowder configurations using the native structures of individual proteins and complexes.

Following the Weeks-Chandler-Andersen (WCA) theory [147], we decompose $U_{\text{int}, X}$ into the repulsive and attractive parts. The crowding free energy, $\Delta F_X^{\text{crowd}}(\phi)$, can thus be divided into two separable contributions as,

$$\Delta F_X^{\text{crowd}}(\phi) = \Delta F_{\text{X,rep}}^{\text{crowd}}(\phi) + \Delta F_{\text{X,att}}^{\text{crowd}}(\phi), \tag{3.4}$$

where $\Delta F_{X,\text{rep(att)}}^{\text{crowd}}$ is the contribution to the crowding free energy from the repulsive (attractive) protein-crowder interactions.

Now, for any generic type of interaction between proteins and crowders, the task of calculating crowding free energy is divided into calculating the repulsive and attractive contributions for isolated proteins and the bound complex.

It is to be noted that most of the currently used theoretical models aiming to describe experimental data ignore the attractive contribution to the change in the binding free energy, assuming that the interactions between inert crowding agents and proteins can be approximated effectively by repulsive potentials.

A recent theory [80] takes into account both the attractive and repulsive protein-crowder interactions. It is outlined below:

3.3 Repulsive contribution to the crowding free energy

To calculate the repulsive contribution to the crowding free energy, scaled particle theory (SPT) [121] has been adopted. The SPT provides an analytical expression for the free energy of creating a spherical cavity of radius R in a hard-sphere fluid (particle radius R_c) at packing fraction ϕ which is given by,

$$\Delta F_{\mathrm{SPT}}(\phi) = (3y + 3y^2 + y^3)\tilde{\phi} + (4.5y^2 + 3y^3)\tilde{\phi}^2 + 3y^3\tilde{\phi}^3 - \ln(1 - \phi). \qquad (3.5)$$

where $\tilde{\phi} = \phi/(1 - \phi)$ and $y = R/R_c$. The most important question that may arise from the use of SPT to predict first term in eq. 3.4 is : can anisometric proteins and commonly-used crowding agents be represented as hard spheres to form the physical basis of excluded volume interactions?

It can be argued that for the cases in which long-range electrostatic interactions are negligible, the interactions between proteins and crowders and between crowders themselves can be approximated by effective hard-sphere potentials with contact distances adjusted to fit the experimental/simulation data. Previous successes in this direction indicates that such an approximation is acceptable in several cases [98, 38, 57, 39, 132] and in fact one can even come up with rational ways to define effective hard-sphere diameter for proteins [102, 78]. On the other hand, the failure of the SPT to explain the crowding data does not necessarily invalidate the SPT or the underlying spherical approximation. In fact, it highlights the importance of attractive interactions that must be present between proteins and crowders.

In most simulations, crowders are represented by hard spheres or spheres with purely repulsive potentials. In addition, the details of pairwise interactions between protein atoms (or collection of atoms in a coarse-grained protein model) and crowders are given explicitly. Thus to validate the SPT for the effects of repulsive protein-crowder interactions on protein binding thermodynamics, it is essential to provide an unambiguous way to determine the effective sphere radius, R_X, for a protein or complex X. One approach suggested, is to use the Boltzmann factor criteria [11] and define the effective radius for a crowder size R_c as,

$$\frac{4\pi}{3}(R_X + R_c)^3 = \int_{U_{\mathrm{int,X,rep}}(\boldsymbol{r}) \geq f k_B T} d\boldsymbol{r}, \tag{3.6}$$

where $U_{\mathrm{int,X,rep}}$ is the repulsive part of the interaction energy between protein X and a spherical crowder, while $\boldsymbol{r}$ is the three-dimensional vector pointing from the center of mass of the protein to the center of the crowder. We use $f = 2$ that has been successful in describing the thermodynamic and dynamic behavior of Lennard-Jones

fluid based on hard-sphere fluid [100]. Further justification of this f value can be found in the work of Speedy *et al.* [131]. Moreover, a slightly different value of f (e.g., $f = 1$) doesn't result in significantly different results. Note that R_X will then depend weakly on R_c as well as on a given protein conformation X.

The success of the SPT eq. 3.5 using the above prescription for the spherical approximation for anisometric proteins has been demonstrated in simulation studies of two distinct protein complexes, ubiquitin/UIM1 (Ubq/UIM1) and cytochrome c/cytochrome c peroxidase (Cc/CcP). Here, Ubq, Cc, CcP are globular proteins close to spherical shape with 76, 108 and 294 residues, respectively. On the other hand, UIM1 is the 24-residue rod-like protein. It was observed that $\Delta\Delta F^{\mathrm{bind}}(\phi, R_c)$ from the SPT based theory for repulsive protein-crowder interactions agreed remarkably well with simulation data for a wide range of ϕ and R_c even for Ubq/UIM1 complex [78].

3.4 Attractive contribution to the crowding free energy

Calculating the attractive contribution, $\Delta F_{X,\mathrm{att}}^{\mathrm{crowd}}$, to the crowding free energy is more challenging even for a simple spherical solute. In principle, one can use thermodynamic perturbation theory approach to obtain an approximate analytical expression for $\Delta F_{X,\mathrm{att}}^{\mathrm{crowd}}$. Validity of this approach will then depend on the strength of the attractive interactions proteins and crowders and if the reference repulsive interactions can describe the structure (protein-crowder, crowder-crowder) well. Thus, up to the first order in the attractive part of the interaction energy, $U_{\mathrm{int},X,\mathrm{att}}$, $\Delta F_{X,\mathrm{att}}$ can be

expressed as [47],

$$\Delta F_{X,\text{att}}^{\text{crowd}} \approx \langle U_{\text{int},X,\text{att}} \rangle_{\text{rep}} = \int \rho U_{\text{int},X,\text{att}}(r,\omega) g_0(r) r^2 dr d\omega, \tag{3.7}$$

where ρ is the crowder number density related to ϕ via $\rho = \phi/(4\pi R_c^3/3)$, and $g_0(r)$ is the radial distribution function between a protein and a crowder in the reference repulsive ensemble $\langle \cdot \rangle_{\text{rep}}$, while ω denotes the angular degree of freedom. We approximate $g_0(r)$ to be a stepwise function, i.e., $g_0(r) = 0$ for $r < r_0$, $g_0(r) = g_{\text{max}}$ for $r_0 \leq r < r_1$, and $g_0(r) = 1$ for $r > r_1$. Employing accurate Carnahan-Starling equation of state for a hard sphere fluid for g_{max}, we thereby obtain crowding free energy up to linear order in ϕ as,

$$\Delta F_{X,\text{att}}^{\text{crowd}} \approx -\kappa(R_c) u_X S_X \phi, \tag{3.8}$$

where S_X is the surface area of a protein or complex encompassed by the center of a crowder, u_X is the average strength of the attractive protein-crowder interaction on the surface S_X, while κ depends only on R_c.

It was observed that u_X depends weakly on the type of protein X [80]. We then obtain the attractive contribution to the change in the binding free energy using Eqs. 3.2, 3.4, and 3.8 as,

$$\Delta\Delta F_{\text{att}}^{\text{bind}} \approx -\kappa(R_c)\bar{u}\Delta S\phi, \tag{3.9}$$

where $\bar{u} = (u_A + u_B + u_{AB})/3$ and $\Delta S = S_{AB} - S_A - S_B$ is the change in the surface area upon binding. The theory outlined above, provides a microscopic foundation for

the phenomenological expressions proposed earlier by Jiao *et al.* [71] and Rosen *et al.* [124].

From eq. 3.9 it is evident that the attractive protein-crowder interactions destabilize complex formation or stabilize isolated proteins in a crowded solution ($\Delta\Delta F_{att}^{bind} > 0$) because $\Delta S < 0$ in general. This enthalpic effect (dependent on protein surface areas) is in contrast to the entropic effect caused by repulsive interactions (dependent on protein volumes) that favors compact structures that occupy lesser volume. This competition can then give rise to numerous different scenarios in experiments or simulations depending on their individual contributions.

3.5 Computational model of protein-protein interactions in a crowded solution

It is quite challenging to test the theory presented above experimentally, since the exact nature of the microscopic interactions between proteins and crowders is unknown in most cases. In case of experimental data (if available), one can consider the physical parameters in the crowding theory outlined above, as fitting parameters to describe the experimental data. The resulting fit parameters will have a sound physical basis and can therefore provide information on unknown microscopic interactions. However, such an approach is still ambiguous as pointed out by Elcock [37]; for dextran, a commonly used crowding agent, three different spherical approximations have been used to fit SPT equation in three different experiments [8, 7].

Computational models provide a concrete platform to test a theory without much difficulty and can also provide new physical insight. Although there have been computational studies mimicking the cellular environment by employing various molecules

measured experimentally inside the cytoplasm of *Escheria coli* [96], such an approach can be computationally prohibitive to sample binding thermodynamics and kinetics accurately. Thus most simulations have been performed using spherical crowders and can still provide valuable insights [144]. Though atomistic description is desirable for protein simulations, computational cost is again prohibitive in sampling protein-protein interactions. Thus a coarse-grained protein model is often used in simulating protein-protein interactions [79]. In particular, the residue-level coarse-grained model developed by Rosen *et al.* has been quite successful in yielding binding free energies for weak-to-moderately binding protein complexes and in describing the non-specific complexes in good agreement with NMR experiments [81].

For interaction between a protein residue and a crowder, modified Lennard-Jones (LJ) potential has been found to be suitable and is given by,

$$V(r) = 4\left[\epsilon_r\left(\frac{\sigma_r}{r - \sigma + \sigma_r}\right)^{12} - \epsilon_a\left(\frac{\sigma_r}{r - \sigma + \sigma_r}\right)^{6}\right], \qquad (3.10)$$

where σ is the contact radius (i.e., $V(r = \sigma) = 0$) between a protein residue and a crowder, while σ_r is the interaction range (set equal to 6 Å). The potential

3.6 Effects of crowding on the thermodynamic stability of a protein complex

The crowding theory described here and its variants proposed earlier by others predict that excluded volume interactions due to repulsive protein-crowder interactions will lower the binding free energy thereby favoring complex formation in a crowded solution. The extent of this binding free energy change with respect to crowder-

free solution will depend on various factors like the crowder packing fraction, crowder size, and size of the protein molecules. Simulation studies employing repulsive protein-crowder interactions have essentially validated this qualitative expectation and showed strengthened protein complex formation [78]. In a recent work from our group [78], using LJ-type repulsive interaction potential, we also observed a similar trend as shown here in Fig. 2a. for the Cc/CcP complexes. Again, these results are anticipated, since these protein complexes occupy less volume as compared to the total volume occupied by the isolated proteins. With increasing crowder packing fraction, the binding free energy is lowered in a non-linear fashion. Also, for a given crowder packing fraction, smaller crowders have a more stabilizing influence on the complex formation — a prediction borne out by the SPT. The predictions of our crowding theory presented in the previous section, for which proteins are mapped onto spheres, agrees remarkably well with the simulation data over a wide range of crowder sizes and crowder packing fraction. This theory can provide quantitative predictions for the change in binding free energy in the presence of repulsive spherical crowders with minimal information (protein structures). Note, however, that whether the theory can be still valid for highly anisometric proteins, in general, requires further tests in future work.

Thus, thermodynamically the excluded volume effect due to crowding favors the association of macromolecules. Various experimental studies that are consistent with this expectation are reviewed by Zimmerman and Minton [158] (Table 2) and Zhou [155] (Table 1). Adding to the repertoire, a recent study [2] probed the effect of crowding (using Ficoll 70) on a heptameric protein (human cpn10 or GroES in E. *coli*) consisting of seven identical $\beta-$barrel subunits assembling into a ring. Using tyrosine fluorescence, it was observed that the monomer-heptamer dissociation constant value

is lower in the Ficoll 70 solution than in the buffer thereby suggesting a stabilization of the heptameric complex due to crowding.

Although crowding effects by hemoglobin, serum albumin and dextran can be quantitatively accounted for [123], excluded-volume based models fail to account for the crowding effects exerted by another commonly used crowding agent, i.e., polyethylene glycol (PEG). Phillip *et al.* [115] found negligible impact due to increased PEG 1000 crowding varying up to 30% packing fraction, on the binding affinity of TEM1-β-lactamase with its inhibitor β-lactamase and barnase with barstar. Experimental studies [29, 115] suggest the presence of an attractive interaction between PEG molecules and proteins. Based on our earlier discussion, attractive protein-crowder interactions will actually counteract the stabilizing effect of excluded volume on complex formation and can help explain this trend.

To probe the effect of attractive interactions in addition to the excluded volume effects, we conducted simulations over a wide range of parameters [124]. As shown in fig. 3.4, after a critical threshold, increasing the protein-crowder attraction strength results in the destabilization of the protein complex relative to the crowder-free solution (i.e., $\Delta\Delta F^{\mathrm{bind}} > 0$). This effect is found to be more pronounced for crowders of smaller sizes. At low protein-crowder attraction strengths, we find that the stabilizing entropic effect is dominant over the destabilizing enthalpic effect. In fact, the critical attraction strength for which the binding free energy exhibits no change in a crowded solution as compared to the crowder-free solution (i.e., $\Delta\Delta F^{\mathrm{bind}} = 0$) is approximately independent of the crowder packing fraction ϕ as shown in Fig. 3.4. This observation is a reflection of the approximately linear dependence of the binding free energy with packing fraction up to modest packing fractions. Importantly, the

agreement between our crowding theory (solid curves in fig. 3.4) and simulation data is quite remarkable.

Probing the effects of attractive interactions experimentally also have begun only quite recently. Jiao *et al.* [71] have recently studied catalase and superoxide dismutase interactions in the presence of dextran 70, ficoll 70 and polyethylene glycol 2000 at various concentrations and temperatures. They found that above a particular temperature, denoted by T_θ, the primary effect of addition of crowders was the enhancement of protein association, whereas, at a temperature below T_θ, attractive interactions between the proteins and polymers predominate, inhibiting protein association. At a temperature T approximately equal to T_θ the two effects cancel each other thereby showing no effect on protein association (similar to our critical protein-crowder interaction strength) upon increasing the concentration of the crowding polymers.

3.7   Effects of mixed macromolecular crowding

So far we have discussed the development of a theory and associated simulation results for a single type of crowder in a crowded solution only. On the other hand, macromolecules present in a cell are quite diverse in their sizes and interactions. Therefore, mixed crowding with different types of crowder particles will be a more realistic description of cellular crowding. One then may ask if there is any difference in treating the effects of mixed macromolecular crowding compared to the single-component crowders and if one may observe qualitatively different results? A few recent studies, experimental as well as theoretical, have proposed that the mixed crowding solution may actually enhance the effect of crowding and may result in non-monotonic effects unlike single-component results [33, 8]. However, our simulation
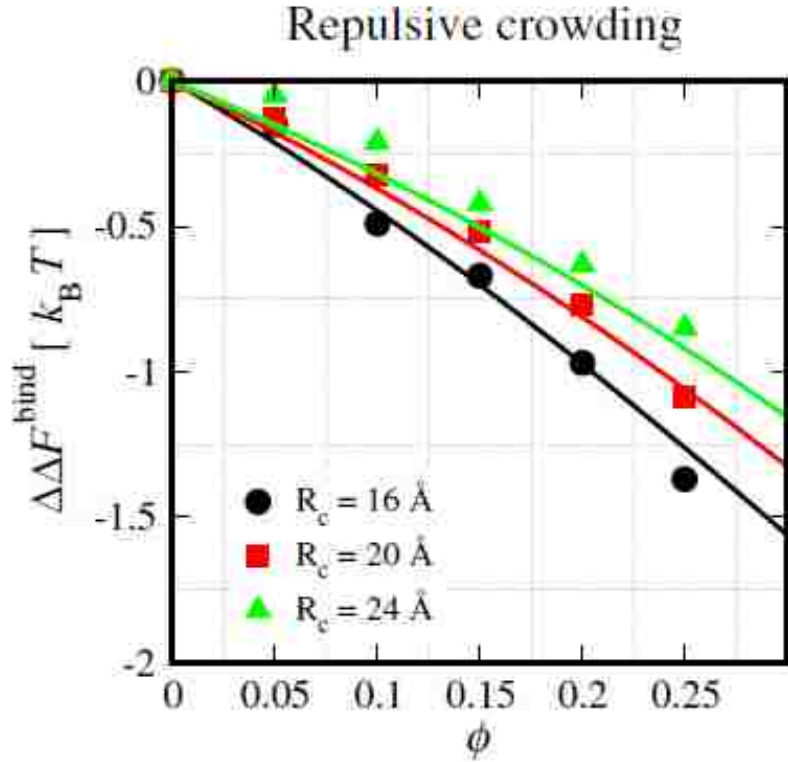
Figure 3.2: The change in binding free energy $(\Delta\Delta F^{bind}(\phi)$) data obtained from replica exchange Monte Carlo simulations (symbols) are shown as a function of the crowder packing fraction (ϕ) for different crowder sizes for a purely repulsive protein-crowder interaction strength of $\epsilon_r = 1.69$ k$_B$T [80]

data for a binary crowder mixture (repulsive protein-crowder interactions) did not show anything qualitatively different from a single-component crowder solution [78]. Moreover, within the SPT-like approach the effects of mixed repulsive crowding on the binding free energy are actually additive and we had proposed the following *ansatz*,

$$\Delta\Delta F^{\text{bind}}(\phi_1, \phi_2, \cdots, \phi_N) = \sum_i x_i \Delta\Delta F_i^{\text{bind}}(\phi), \qquad (3.11)$$

where $i = 1$ to N, ϕ_i is the packing fraction of component i in a crowding mixture, $x_i = \phi_i/\phi$ is the relative fraction of component i, $\phi = \sum_i \phi_i$, and $\Delta\Delta F_i^{\text{bind}}$ is the change
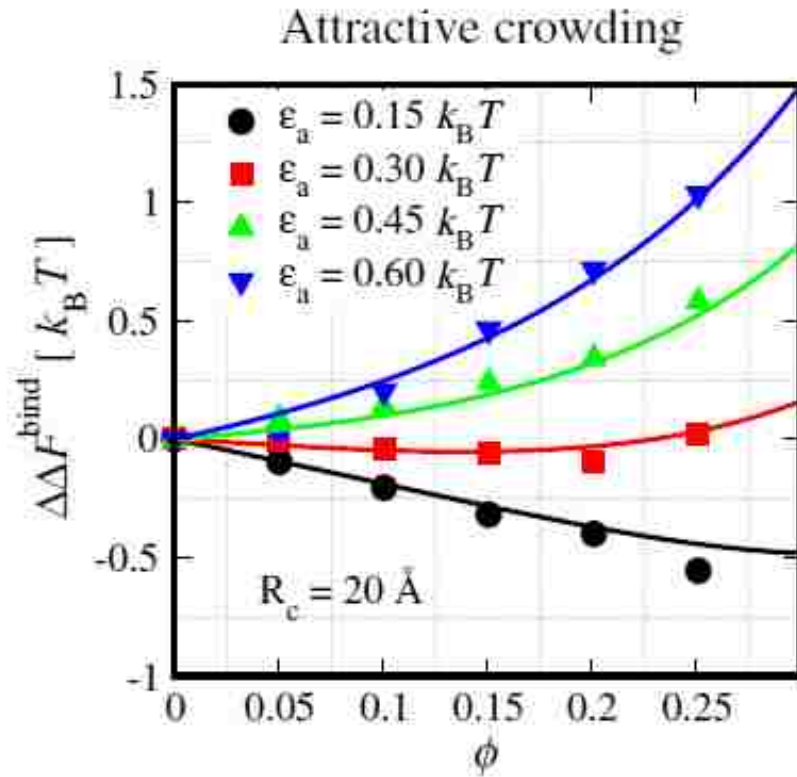
Figure 3.3: The change in binding free energy ($\Delta\Delta F^{bind}(\phi)$) data obtained from replica exchange Monte Carlo simulations (symbols) are shown as a function of the crowder packing fraction (ϕ) for different protein-crowder interaction strengths for a particular crowder of radius 20 Å [80]

in the binding free energy of a pure component i at total crowder packing fraction ϕ.

Simulations for repulsive protein-crowder interactions with a binary crowder mixture

for the sizes between 12 Å and 20 Å showed an excellent agreement with the prediction

of eq. 11 [78]. In case of attractive protein-crowder interactions, the additivity ansatz

eq. 11 still holds for a range of crowder sizes, 12 Å- 20 Å and for weak attractive

interactions. However mutual attraction between different crowder components can

affect the behavior in a way that cannot be captured in an additive sense. More work
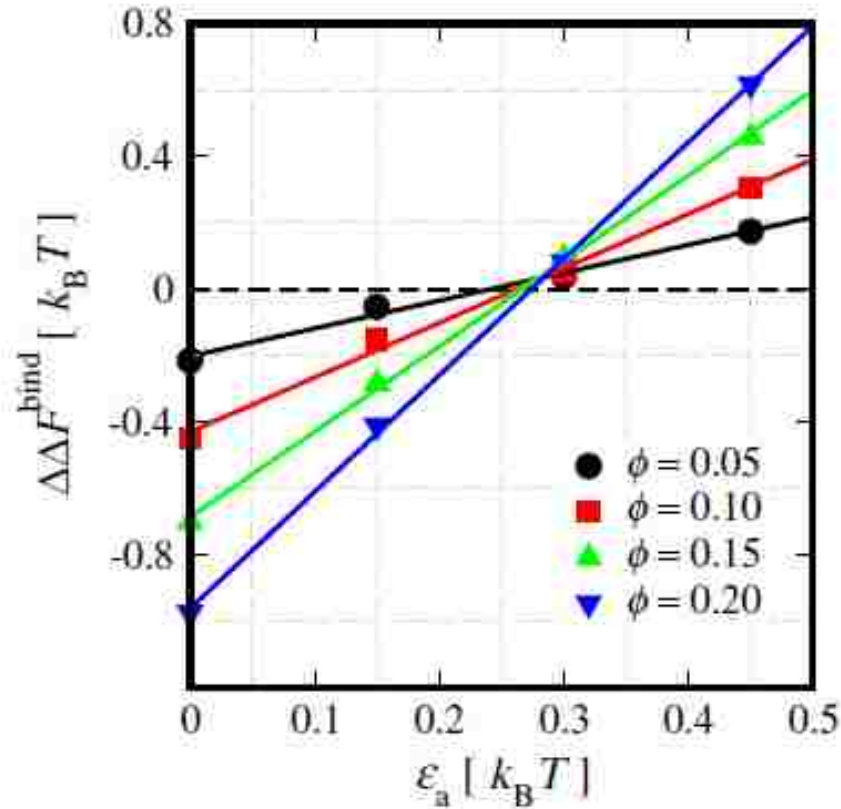
Figure 3.4: The change in binding free energy data calculated from simulations (symbols) are shown as a function of crowder-protein attraction strengths (ϵ_a) for various crowder packing fractions ϕ for the Cc/CcP complex. The fit curves for different ϕ converge around the point where $\Delta\Delta F^{bind} \approx 0$ [80]

is needed to identify the parameter range for which eq. 11 is a good approximation to yield accurate estimate of $\Delta\Delta F^{\text{bind}}$ for a crowder mixture.

3.8 Effects of crowding on protein complex specificity: native vs encounter complex

Recent experiments and simulation have shown the existence of nonspecific complexes in solution, *albeit* in low population ($< 10\%$), which may play an important role in
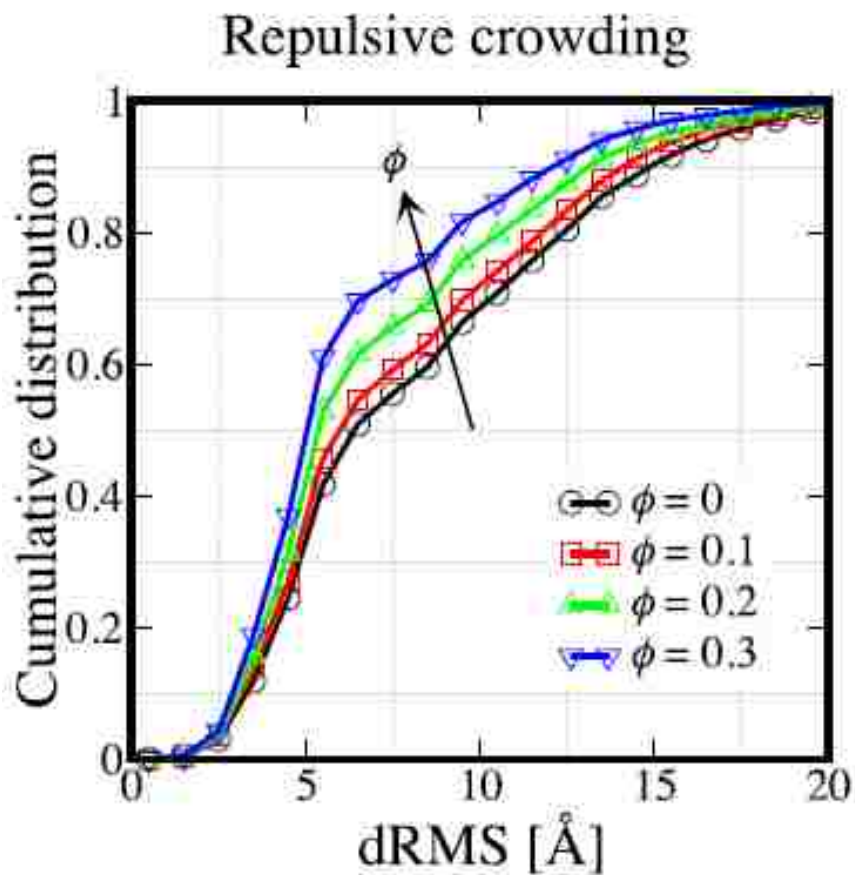
Figure 3.5: Cumulative distribution of bound complexes in the presence of repulsive crowders are shown as a function of dRMS calculated based on the experimental native structure for the Cc/CcP. [80]
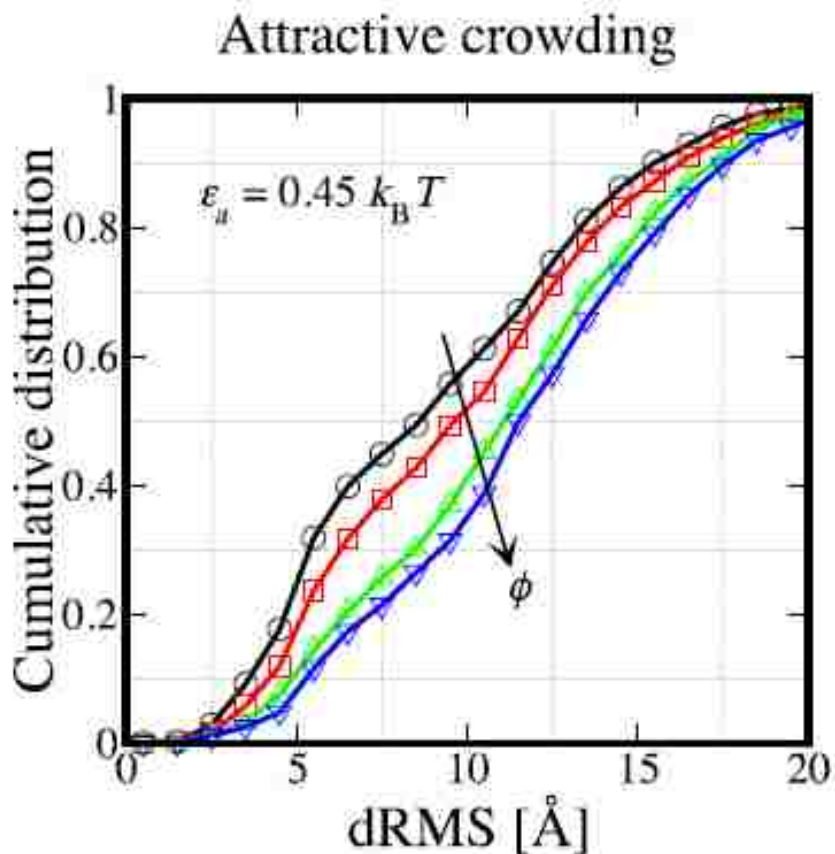
Figure 3.6: Cumulative distribution of bound complexes in the presence of attractive crowders are shown as a function of dRMS calculated based on the experimental native structure for the Cc/CcP. [80]

the complex formation by reducing the degree of freedom during the binding target search process [126, 90, 146]. What is the effect of macromolecular crowding on the presence and stability of such nonspecific complexes? Fig. 3.5 shows the cumulative distributions of dRMS with increasing crowder packing fraction ϕ for the Cc/CcP complex in the presence of repulsive crowders. Here dRMS is a measure of similarity between the experimental native complex and simulated structures. Structures with dRMS less than 5 $\mathring{A}$ are very similar to the native complex structure. It was observed that as the crowder packing fraction increases, the population of native-like structures
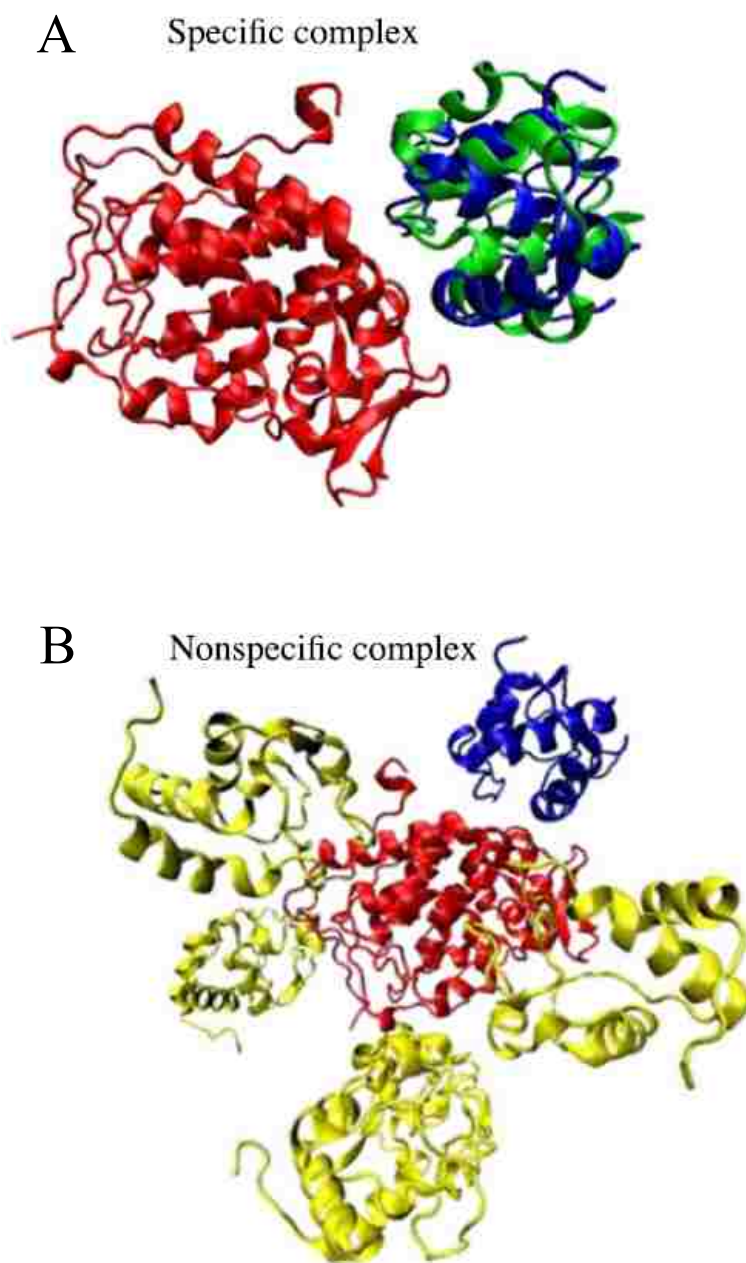
Figure 3.7: **A.** The specific bound complexes are shown where the *red-blue* combination is used for the experimental structure (PDB:2pcc) and the *red-green* combination is used for the complex structure obtained from simulation. **B.** Several instances of the nonspecific bound complexes (*red-yellow* combination) are shown and the experimental native structure *red-blue* is also shown for reference. [80]

with dRMS less than 5 $\mathring{A}$ increases, while the populations of transient encounter complexes (dRMS greater than 5 $\mathring{A}$) decrease [78]. This suggests that proteins in the presence of repulsive crowders are more likely to form more compact native complexes (fig. 3.7A) than metastable intermediate states (fig. 3.7B).

On the contrary, as the attractive protein-crowder interactions tend to maximize contact between a protein and crowders, one would expect that the nonspecific complexes with larger surface area exposed to the crowders pare stabilized in this case. This expectation is confirmed in our simulations as shown in fig. 3.6; the population of nonspecific complexes (dRMS greater than 5 $\mathring{A}$) is enhanced as the protein-crowder attraction strength increases. These findings are also consistent with recent experimental results [114].

3.9 Conclusions

We have presented a review of the development of a general theory to describe the effects of macromolecular crowding on protein-protein interactions. The theory accounts for both repulsive and attractive protein-crowder interactions. The change in the binding free energy due to crowding can be separated into repulsive and attractive contributions. The repulsive contribution is well described by the scaled particle theory (SPT) by approximating proteins as spherical objects. An approximate analytical expression, meanwhile, is obtained for weak protein-crowder interactions by using the statistical mechanics of hard-sphere fluid and the first-order perturbation theory. To validate the theory, simulations were performed on two distinct protein complexes, Ubq/UIM1 and Cc/CcP, using a residue-level coarse-grained protein model and spherical crowders. It was shown that the theory can describe the simulation

data remarkably well for both repulsive and attractive protein-crowder interactions over a wide range of parameters. In addition, simulations show that the repulsive protein-crowder interactions increase the population of the native complex at the expense of transient encounter complexes, while the opposite trend was observed for the attractive protein-crowder interactions.

Most of the work so far, however, has been focused on the complex formation between well-structured proteins, thus ignoring any protein conformational change. In such cases it was found that a single conformation of a protein or complex was sufficient to calculate the crowding free energy. However, there exist many eukaryotic proteins which are disordered in isolation under physiological conditions, but fold into their native conformations upon binding to target proteins [150]. No theoretical and computational studies have focused on the effects of macromolecular crowding on such protein complexes. Since the folding of a protein is tightly coupled to the binding event and crowding agents may exert a different level of effect on the stability of the folded state and the association equilibria, one may encounter complex scenarios in the presence of crowders. Although it can easily be predicted that the binding of these complexes would be enhanced by the excluded-volume effects, it is still unclear whether the theory introduced here will be adequate to describe their behavior. In addition, one may ask whether the kinetics [77], and, in particular, underlying mechanism remain the same in the presence of generic crowding agents. Thus further theoretical as well as computational studies are warranted along these directions.

Chapter 4

Effects of Macromolecular Crowding on the binding thermodynamics of proteins modeled as flexible entities

4.1 Introduction

Proteins are conformationally 'flexible' entities and it is necessary to take into account their dynamical properties to understand their function. Binding of one protein to another, thus, is more involved than what is envisaged using the 'rigid-body' assumption commonly made while studying protein-protein interactions [89].

In general, in the timescale of binding, proteins fall onto a structural continuum, from tightly folded single domains all the way to highly flexible heterogeneous unstructured states [35]. Significant research is being carried out to understand the role

of these intrinsically unstructured (or disordered) proteins in various physiological processes such as the regulation of transcription and translation [49], cellular signal transduction [65, 34], protein phosphorylation [66] and macromolecular self-assembly [108]. It is hypothesized that the lack of intrinsic structure in IDPs plays an important role in their varied functions in these processes. It is, thus, imperative to venture beyond the 'rigid-body' assumption while studying protein-protein interactions.

However, prior to understanding the binding of flexible proteins to their targets under physiologically relevant conditions (such as macromolecular crowding), it is necessary to identify a model, that comprehensively captures important details of the process. Sugase *et al.*, using NMR titration and ^{15}N relaxation dispersion have shown that the unstructured phosphorylated kinase-inducible domain (pKID) of transcription factor CREB binds to the KIX domain of a CREB binding protein (CBP), inducing the folding of pKID into two α-helical domains [133]. Previous experimental studies [118, 117] have also revealed that the KID domain is largely unstructured. This model has been used in various simulation studies [140, 46] to understand the microscopic details of the binding mechanism.

Having identified an appropriate model for study, we intend to study the effects of macromolecular crowding on the thermodynamics of binding of flexible proteins. Secondly, are the theories described in chapter 3 adequate enough to describe the behavior of flexible proteins? To address these issues, in this chapter, we study the effects of macromolecular crowding on the pKID:KIX complex considering the proteins as flexible bodies. A C_α based coarse grain model was used to represent the pKID:KIX complex. To account for the varying crowder-protein interactions we have used Lennard-Jones type potentials representing both repulsive and attractive protein-crowder interactions. For repulsive crowding, we study the effects over a

range of crowder sizes and the effects of attractive crowding were studied by varying the interaction strength for a representative particle size. We observe that repulsive crowding stabilizes the bound complex. Smaller crowders have a more stabilizing effect towards the bound complex as compared to the larger ones. In the scenario where there is an additional attractive interaction between the crowders and the peptides, we observe a destabilizing effect towards the bound complex. Additionally, we observe that the scaled particle theory (eq. 3.5) which was originally developed for hard-sphere particles, remarkably predicts the simulation data for a wide range of ϕ and R_c. In order to predict the simulation data using attractive crowders, we have modified the SPT by adding a phenomenological term as described in eq. 3.9.

4.2   Models and Methods

4.2.1   Model

We used a minimalist representation of the protein where each amino acid residue is described by one bead located at the C_α position. The mass of each bead is considered to be the mass of the entire amino acid residue it represents. The beads are connected by a coarse-grained representation of the peptide bond by a harmonic potential along the protein backbone [74] and the angle between the three adjacent beads were also represented by a harmonic potential. A knowledge-based sequence specific, but topology-independent torsional potential has been used to represent the virtual torsional angles between four adjacent beads along the C_α chain. Such sequence specificity and native structure independence has been effective in capturing the subtle differences in folding mechanisms and kinetics arising from sequence differences in topologically similar peptides [74]. All the force constants are proportional

to an energy scale defined as the average native contact per residue, which is directly proportional to the folding temperature of a particular protein.

The non-bonded interaction terms include favorable interactions for residues that are in contact in the native state and a repulsive interaction between all other pairs of residues. For residues involved in intra-molecular native contacts, separated in sequence by two or more bonds, the potential is of the form

$$V_{ij}(r_{ij}) = \epsilon_{ij} \left[13 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 18 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} + 4 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \quad (4.1)$$

where r_{ij} is the distance between residues i and j, σ_{ij} is the distance between i and j at which the interaction energy is a minimum (equal to $-\epsilon_{ij}$). The σ_{ij} was set to the α-carbon separation distance of the pair i and j in the native state. Different values of ϵ_{ij} have been used accounting for varying types of native contacts. Residues which are hydrogen-bonded in the native state interact with an ϵ_{ij} value of unity. This particular assignment takes into account the backbone hydrogen bonds between residues involved in α-helices. To maintain backbone orientation for residues involved in β-strands, four additional native contacts each with $\epsilon_{ij} = 0.25$ were defined in the vicinity of a hydrogen bonded pair, i.e., for residues i and j involved in a hydrogen bond, the additional weak native contacts were defined for residues $(i-1, j), (i, j-1), (i, j+1)$ and $(i+1, j)$. Incorporating the variegated chemistries of the sidechains, for side chains involved in native contacts, the contact energies were obtained from knowledge based values reported by Miyazawa and Jernigan [105] and scaled appropriately [74]. This scaling was necessary because of the introduction of a new energy scale relative to the hydrogen bond native contact energy.

For residues involved in inter-molecular native contacts between pKID and KIX, a 12-6 Lennard Jones type potential was used.

$$V_{ij}(r_{ij}) = \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \tag{4.2}$$

The advantage of this particular form of the potential is two-fold. The value of the binding dissociation constant K_d obtained in bulk using replica exchange molecular dynamics simulations at 300K was just an order of magnitude different from the experimental value. This was achieved without the introduction of any further scaling of the interaction parameters. A faster convergence was also observed. Owing to the lack of structural experimental evidence of the transition transition complex ensemble, K_d emerges to be the sole parameter bridging experimental and simulation studies.

For pairs of residues not involved in native contacts, a purely repulsive potential of the form,

$$V_{ij}(r_{ij}) = \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right], \tag{4.3}$$

have been used where ϵ_{ij} is set to a very low value proportional to the energy scale discussed above.

For protein-crowder interactions of the repulsive form, the following potential has been used:

$$V_{ij}(r_{ij}) = \epsilon\left[\left(\frac{\sigma_{ref}}{r_{ij} - \sigma_{ij} + \sigma_{ref}}\right)^{12}\right]. \tag{4.4}$$

For attractive protein-crowder interactions, the potential used is of the following form:

$$V_{ij}(r_{ij}) = 4\epsilon_r\left(\frac{\sigma_{ref}}{r_{ij} - \sigma_{ij} + \sigma_{ref}}\right)^{12} - 4\epsilon_a\left(\frac{\sigma_{ref}}{r_{ij} - \sigma_{ij} + \sigma_{ref}}\right)^{6} \tag{4.5}$$
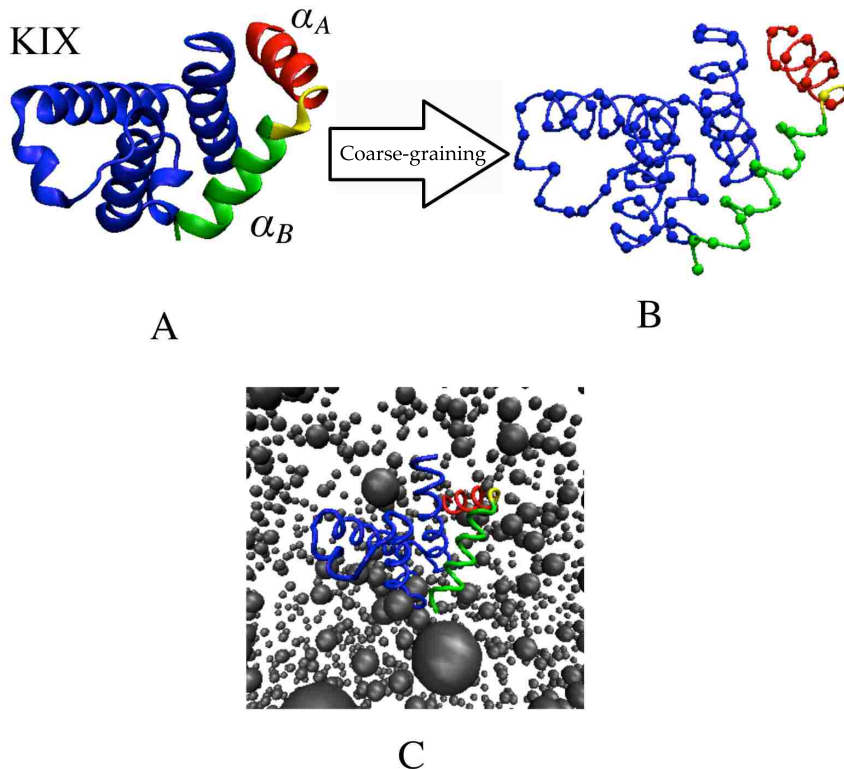
Figure 4.1: **A.** Solution structure of the pKID-KIX complex. KIX contains three α -helical domains. pKID contains two helical domains - α_A (colored red) and α_B (colored green). **B.** C_α based coarse-grained representation of the pKID-KIX complex. **C.** Schematic of our model system in the presence of crowders.

For both repulsive and attractive protein-crowder interactions, σ_{ref} is the interaction range set equal to $6\mathring{A}$ and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ is the interaction diameter calculated from crowder radius $r_c = \sigma_i/2$ and van der Waals radius $\sigma_j/2$ for a given residue . For crowder-crowder interactions, we used eq. 4.4 with $\sigma_{ij} = 2r_c$. For repulsive protein-crowder and crowder-crowder interactions, ϵ in eq. 4.4 was set to $1.67k_BT$. For repulsive protein-crowder interactions, we used crowders of varying sizes ranging $6\mathring{A}$, $9\mathring{A}$ and $12\mathring{A}$. For attractive protein-crowder interactions, simulations were performed using a single crowder size of $9\mathring{A}$ and protein-crowder interaction strengths of $\epsilon_r(= \epsilon_a) = \{0.45, 0.90\} \ k_BT$.

4.2.2 Simulation Details

To obtain the equilibrium ensemble of protein complexes, we performed replica exchange molecular dynamics [134] (REMD) in the NVT ensemble using the langevin thermostat with a friction coefficient of 0.2 ps^{-1} which is smaller than that used to mimic water (50-100ps^{-1}). Cubic boxes of varying sides were used with periodic boundary conditions. The box sizes were so varied as to vary the concentration of the complex. Sufficiently long REMD simulations were run for each of the 16 replicas spanning 276 to 665K with a timestep of 10fs with an exchange frequency of 5ps ensuring adequate sampling as shown from negligible standard errors in fig.4.2 and 4.3. The simulation times varied between 4-6 μs. The non bonded interactions were cutoff at 30 Å . We discarded the first 500ns in lieu of equilibration. All the simulations were performed using Gromacs 4.0.5 modified to incorporate the nonbonded interactions.

4.2.3 Free Energy Calculations

In order to determine the binding free energy ΔF_b, we measured the dissociation constant K_d of the complex which is related to the binding free energy via $\Delta F_b = k_B T(\ln K_d/K_0)$ where K_0=1M. We calculated the fraction of bound complexes by assigning complexes as bound if the inter-molecular interaction energies between pKID and KIX is less than -1 $k_B T$. The choice of the cutoff was based on the bimodal distribution of the interaction energies obtained from the simulations. The binding dissociation constant was obtained by fitting the fraction of bound complexes as a function of protein concentration via the titration formula,

$$y = \frac{[P]}{[P] + K_d} \tag{4.6}$$

where [P] is the protein concentration and y is the fraction of bound complexes. Eq. 4.6 is a consequence of the Langmuir isotherm of adsorption.

4.3   Results and Discussion

The top panel of figure 4.2 shows data for the fraction of bound complexes for the pKID-KIX system as a function of the protein concentration at different crowder packing fraction ϕ and crowder sizes. The top panel shows the data for crowders of radius $r_C = 6\mathring{A}$, the middle panel shows the data for $r_C = 9\mathring{A}$ and the bottom panel shows the data for $r_C = 12\mathring{A}$. The error bars were obtained by dividing the equilibrated trajectory into four equal parts and calculating the standard deviations of the four observations. The lines were fits obtained using eq. 4.6. The dissociation constant (K_d) obtained from the simulations without any crowders (black curves in fig. 4.2) is 273 μM. K_d (in μM)obtained from simulations over various crowder sizes and packing fractions are:

$r_C(\mathring{A})$ ϕ	6	9	12
0.05 (red)	132.5	185.7	195.1
0.15 (green)	59	64	102
0.25 (blue)	12.4	23.2	47

It is observed that for a particular crowder size, the bound complex is stabilized as the packing fraction of the crowders is increased. Also, for a particular packing fraction, smaller crowders stabilize the complex more than the larger ones.

In figure 4.3, we plot a representative case where for the attractive protein-crowder interaction (magenta). For the plot, the crowder packing fraction is $\phi = 0.15$ and
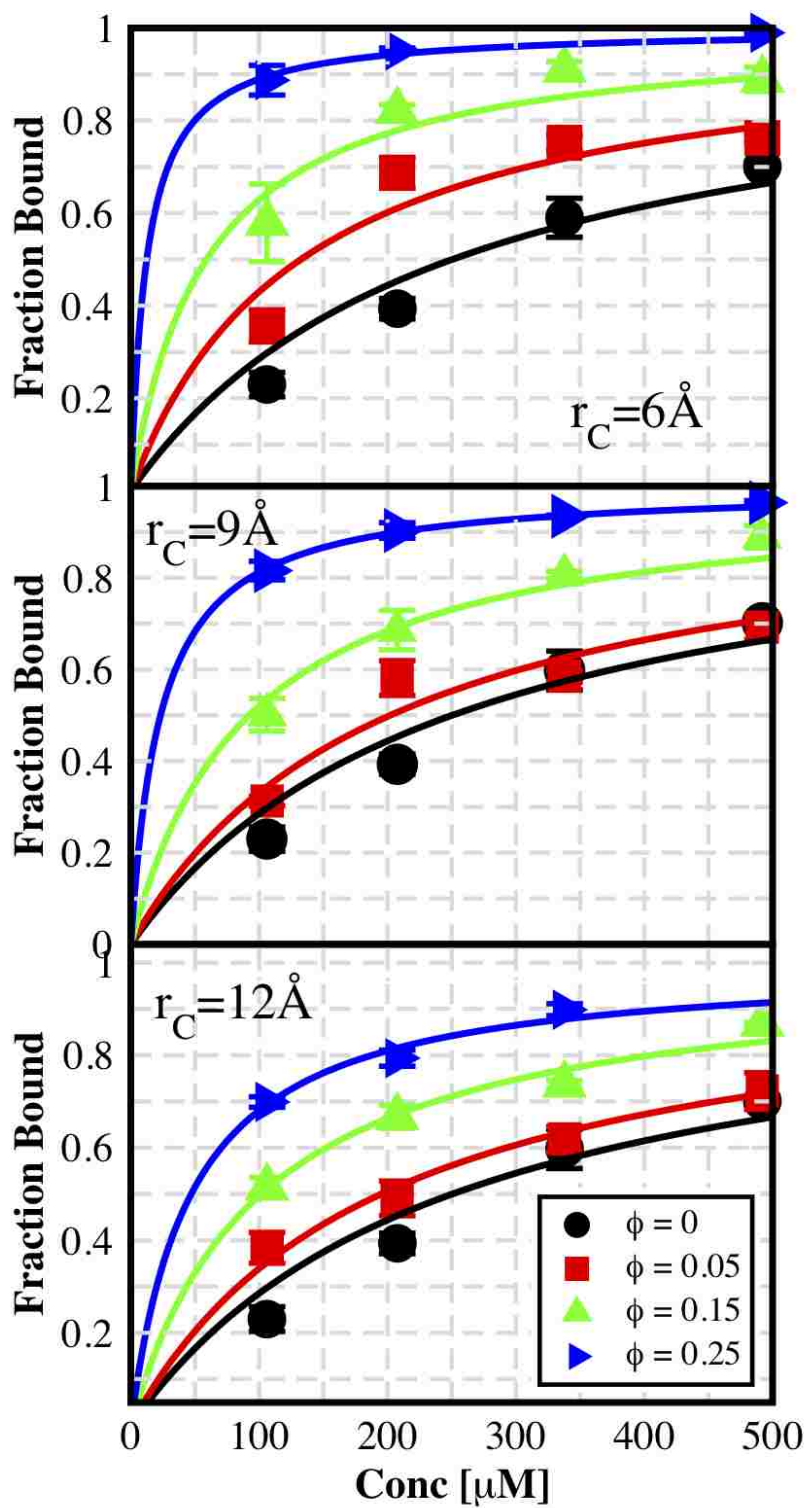
Figure 4.2: Titration curves obtained from simulations over a wide range of crowder packing fractions and crowder sizes. The black data points represents the bulk data.
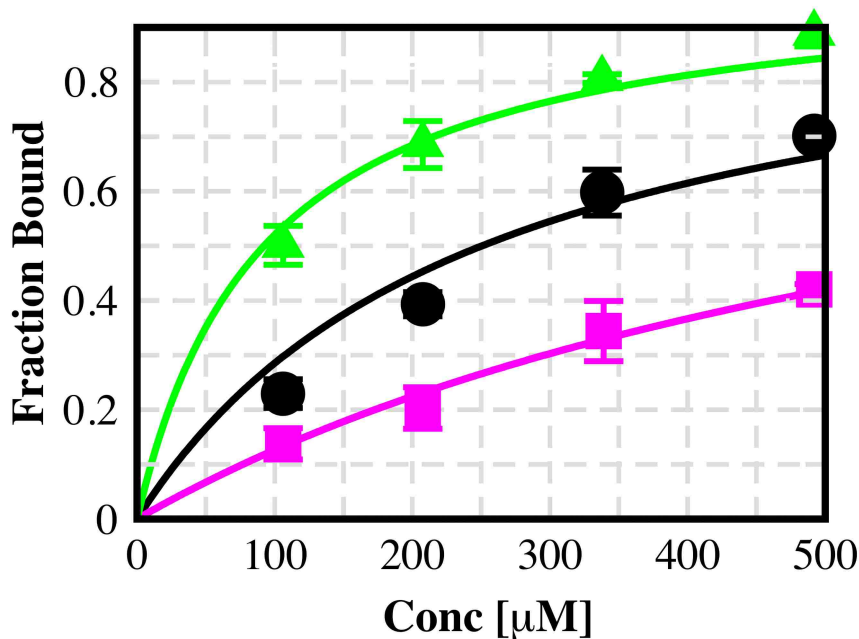
Figure 4.3: Titration curve (magenta line) obtained for a protein-crowder attraction strength of $\epsilon_a = 0.9 k_B T$ (magenta squares) and crowder size $r_C = 9 \mathring{A}$. Data obtained from bulk simulations and in the presence of repulsive protein crowder interaction are shown for comparison

the crowder size is $r_C = 9 \mathring{A}$. For comparison, we plot the data obtained from the repulsive protein-crowder interaction with the same parameters (green) along with the bulk data. The protein-crowder attraction in this particular case is 0.9 $k_B T$. It is observed that attractive crowding destabilizes the protein complex with respect to bulk.

For a quantitative treatment, the effects of crowding on the protein binding are described by calculating the difference in the binding free energy between the cases in bulk ($\phi = 0$) and in the presence of crowders. Figure 4.4 shows the binding free energy difference $\Delta\Delta F_b = \Delta F_b(\phi) - \Delta F_b(\phi = 0)$, as a function of crowder packing fraction ϕ for the pKID-KIX system. As discussed above, increased crowding stabilizes the bound complex by lowering the binding free energy $\Delta F(\phi)$. The binding
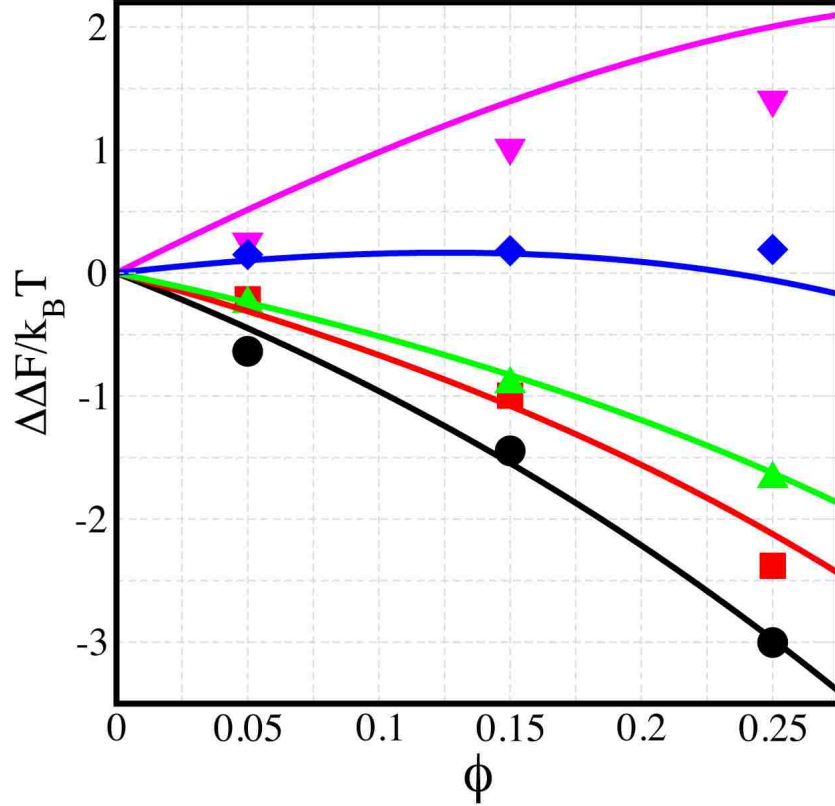
Figure 4.4: The change in binding free energy $\Delta\Delta F_b(\phi) = \Delta F_b(\phi) - \Delta F_b(\phi = 0)$ data obtained from REMD simulations (symbols) are shown as a function of crowder packing fraction ϕ for various cases. Black circles, red squares and green triangles denote the value obtained for repulsive crowding using $r_c = 6$Å, 9Å, and 12Å respectively. The blue diamonds and magenta triangles (pointed down) indicate the values obtained for attractive protein crowder interactions at ϵ_a/k_BT values of 0.45 and 0.90 respectively with $r_c = 9$Å for both cases. The black, red and green lines are the predictions of our theoretical model using SPT for the respective crowder sizes and the blue and magenta lines are the predictions using the modified SPT model incorporating attractive crowder protein interactions for the respective attractive strengths.

free energy changes up to 3 $k_B T$ at the highest crowder packing fraction simulated here ($\phi = 0.25$). This behavior is consistent with the previous studies as shown in chapter 3[78].

To what extent can we use the theories described in chapter 3 for 'rigid' proteins to this particular case where the proteins are considered to be 'flexible' entities ? To answer this, we use the scaled particle theory (reiterated here for convenience) where the free energy required to form a spherical cavity of radius R in a hard sphere fluid with particle radius r_c and density ϕ is given by

$$\Delta F^{SPT}/k_B T = (3y + 3y^2 + y^3)\rho + \left(\frac{9}{2}y^2 + 3y^3\right)\rho^2 + 3y^3\rho^3 - \ln(1 - \phi) \qquad (4.7)$$

Here, $\rho = \phi/(1 - \phi)$ and $y = R/r_c$. The underlying idea is that the free energy difference between the bound and unbound states in the presence of crowders can be calculated by considering a thermodynamic cycle of the complex formation as described in section 3.2. Specifically, we calculate the effect of crowding on the free energy of bound and unbound complex separately. The difference in the two quantities determined by SPT provides us with an analytic estimate of crowding effects on binding equilibrium, i.e., $\Delta\Delta F_b$ while the simulation estimate is obtained along the other edges of the cycle.

Although scaled particle theory is defined in terms of spherical objects, we have shown previously that a simple spherical approximation of aspherical proteins can be incorporated into eq. 4.7 to describe the excluded volume effect on protein binding. We obtained the effective radius of a protein, R_{eff} by considering the protein as a set of solvated spheres of radii $\sigma_i/2 + r_c$ where $\sigma_i/2$ is the radius of residue i, and

by equating the total volume occupied by these spheres to the volume of a sphere of radius $R_{eff} + r_c$. It was observed that the above mapping in conjunction with SPT model was able to predict quantitatively the effect of crowding on the binding equilibria [78]. This result asserts the robustness of the SPT model in predicting the binding thermodynamics of proteins.

In order to capture the effect of attractive protein-crowder interactions explicitly, we have augmented the SPT equation with a mean-field free energy term due to attractive interactions. Thus:

$$\Delta\Delta F(r_c, \phi, \epsilon_a) = \Delta\Delta F^{SPT}(r_c, \phi) + \Delta\Delta F^{att}(r_c, \phi, \epsilon_a) \tag{4.8}$$

where

$$\Delta\Delta F^{att}(r_c, \phi, \epsilon_a) = -\kappa(r_c)\Delta S\epsilon_a\phi/r_c^3 \tag{4.9}$$

In eq. 4.9, κ is a phenomenological constant which depends on r_c and ΔS is the change in accessible protein surface area to crowder particles in going from the unbound proteins to a bound complex. The accessible surface area was obtained by calculating the overall surface area represented by the spheres solvated with crowders. The predictions of this modified SPT model are shown in fig. 4.4, with fit parameter $\kappa/\text{Å} = 3.15$ ($r_c = 9\text{Å}$). The blue diamonds along with the curve corresponds to $\epsilon_a = 0.15$ kT and the magenta triangles along the curve corresponds to $\epsilon_a = 0.90$ kT

Previous studies as described in chapter 3 focused on protein-protein interactions in the presence of crowders where the proteins were represented as 'rigid' bodies. We hypothesized that the spherical representation of proteins may only work well for those that are modestly anisometric rigid objects and for crowders of size similar

to proteins. In plotting, although the data points in figure 4.4 were obtained using flexible proteins, we calculated the R_{eff} values for eq. 4.8 based on the coarse-grained structure of the complex obtained from the PDB. In other words, we have obtained the ΔF_X values in the vertical edges of the thermodynamic cycle of 3.1 by considering the individual peptides as 'rigid' bodies and fit the data obtained from simulation of 'flexible' bodies along the horizontal edges. What happens to the fit if we calculate the ΔF_X values by considering the horizontal edges in the thermodynamic cycle of fig. 3.1?

To answer the question, we estimated the R_{eff} values of the proteins and the complex averaging over the ensemble in the absence of crowders. The distribution of the R_{eff} values were observed to be unimodal. We used the mean values of R_{eff} and re-plot 4.4. From figure 4.5 the nature of the fits obtained are comparable to what was obtained previously. We used the same κ value used to plot 4.4 in this case. This further asserts the robustness of the SPT in predicting the thermodynamics of binding of proteins in a crowded environment.

4.4   Conclusion

Using a Go like C_{α} based coarse grain model, we have performed extensive replica exchange molecular dynamics simulations of the pKID:KIX complex in the presence of crowders to understand how macromolecular crowding affects the thermodynamics of binding by considering proteins as 'flexible' bodies. In order to account for the effects of volume exclusion by the crowders and possible non-specific attractive interactions, we studies both attractive and repulsive protein-crowder interactions. We show that repulsive crowding stabilizes the bound complex whereas attractive crowd-
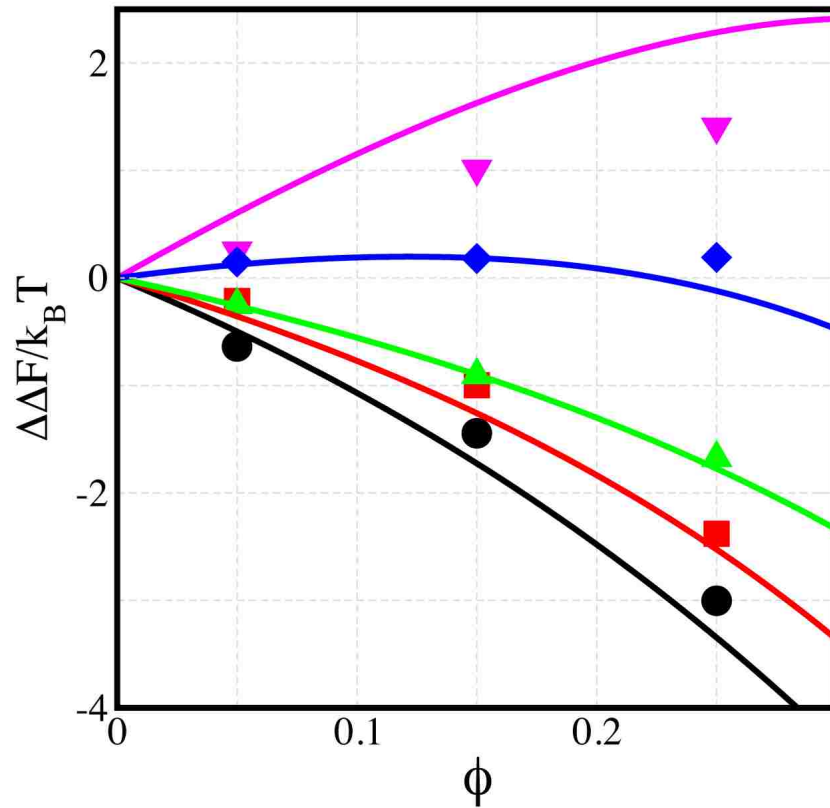
Figure 4.5: Figure 4.4 re-plot with the curves obtained using R_{eff} values from ensemble averages of the simulation data.

ing destabilized the bound complex. This observation is consistent with what one would expect by considering proteins as 'rigid' bodies. The most remarkable observation of this study was the robustness of the scaled particle theory in predicting the binding thermodynamics of proteins in the presence of crowders. It should be noted that scaled particle theory was originally developed to estimate the free energy of cavity formation in a hard-sphere fluid. SPT thus has been quite successful in studying the behavior of proteins considered as 'rigid' effective spheres under the influence of macromolecular crowding. In this study, we have asserted that the theory can be used to study the thermodynamics of proteins which are considered as 'flexible'

entities. Studies should be directed to further strengthen our assertion with various protein complexes and improved coarse-grained models.

Chapter 5

Planar confinement significantly modulates the Free-Energy landscape of GB1 hairpin

5.1   Introduction

In vivo protein folding, like protein-protein interactions, is expected to be affected by the crowded cellular milieu, [155] especially, due to the presence of cellular machinery such as chaperones that can assist folding by preventing misfolding and aggregation.[152]. Several plausible mechanisms of chaperone function have been proposed including the suggestion that extreme protein confinement inside a chaperone cavity could be an important factor. Brinker *et al.* hypothesized that confinement of an unfolded protein inside GroEL/GroES chaperonin system may smoothen the energy landscape by preventing population of kinetically trapped intermediates [22]. Folding of proteins may also be coupled directly to their synthesis, commonly referred

to as *co-translational* folding, where newly synthesized proteins emerge out of a narrow ribosome tunnel [55, 43]. The potential importance of confinement on protein folding is thus well recognized in the protein folding literature.

Confinement of a polymer-like protein chain to an inert space has been theoretically shown to entropically stabilize the compact folded state with respect to the expanded unfolded state [21, 154, 101]. A more recent simulation study by Ziv *et al.* shows that confinement in a cylindrical cavity mimicking ribosome exit tunnel can also entropically stabilize α helices [159]. Protein stabilization was also reported for a β-protein in a spherical pore [83] by Klimov and Thirumalai. Friedel *et al.* [45] and Baumketner *et al.* [10] found that long-lived intermediates present in bulk have relatively shorter lifetimes when the protein in confined. Similar protein stabilization have been observed for proteins confined in silica pores in laboratory experiments [36, 120]. Mukherjee *et al.* [107] showed that the helicity of alanine-rich peptides confined within AOT reverse micelles, can be tuned by varying the degree of hydration.

It is worth mentioning that most previous simulation studies employed coarse-grained protein models in which solvent effects are included in a very indirect manner. For protein folding in bulk such models have been reasonably successful but they may not be able to capture all of the effects of confinement on folding. This is partly due to the fact that water behavior and hydrophobic effects under confinement are not well understood [68]. Hence, it becomes imperative to consider the solvent explicitly in a simulation model in order to provide an accurate description of the effects of confinement on protein folding. Lucent *et al.* [92] performed the first such study using a distributed computing environment (Folding@home) on an α-helical protein, the villin headpiece domain. They found that protein confinement alone stabilized

the folded state. However, when both protein and solvent were confined in a repulsive spherical cavity, the folded state was actually destabilized. This observed destabilization is in contrast to most previous work on protein confinement in a repulsive pore and clearly highlights solvent's role in protein folding under confinement.

The repulsive interactions between a protein and confining entities are a simplification to represent physical boundaries in a system without any direct affinity of the protein with the confining entity. It is known that attractive interactions between confinement boundaries and protein molecules can further affect protein behavior significantly. For example, a comprehensive study utilizing a coarse-grained protein model has shown that a protein populates new structural conformations under attractive confinement [25]. The modified protein conformations near an attractive surface might be useful in surface assisted peptide folding to achieve a specific functional state [157]. Mukherjee *et al.* [107] also found that the folding rate of a β-hairpin peptide can decrease in the presence of common crowding agents, a result that cannot be explained solely on the basis of excluded volume effects of crowding agents. This study was the first of its kind, in which the influence of attractive confining boundaries has been studied by an all-atom protein folding model, including explicit solvent. Any biological surface (chaperonin surface or large protein assemblies), if modeled using simple confining boundaries, must include attractive interactions in some form.

The folding of a β hairpin under the influence of an attractive planar confinement (Figure 5.1) has been investigated in this chapter using an all-atom representation of the peptide and water molecules. The planar confinement between rectangular walls serves as a basic model for future studies of more complex confinement entities. Extensive studies on the folding thermodynamics and kinetics of this hairpin in bulk [17, 18, 19] provides a comprehensive understanding of the otherwise unperturbed
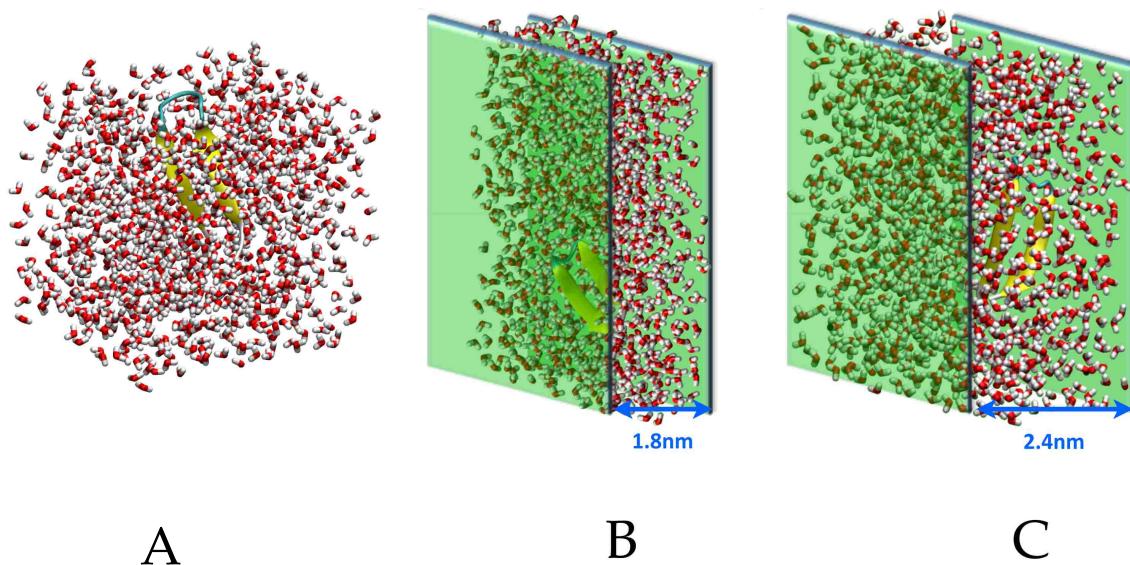
Figure 5.1: Model systems. Schematics of the GB1 hairpin solvated in (A) bulk water and (B) , (C) confined between planar Lennard Jones walls with two different separation distances of 1.8nm abd 2.4nm respectively.

assembly in bulk. Although typical Chaperonin substrates maybe significantly large, the size of the current system permits a thorough equilibrium sampling. Moreover, the system includes representative features of the folding of larger proteins, i.e., it folds in an approximately two-state fashion while populating stable nonnative states in the phase space. Specifically, in previous studies [18, 19], it had been found that this peptide populates a misfolded intermediate state, in which one strand of the hairpin is "flipped" and the hydrophobic side-chains on the two strands are found are on opposite faces of the hairpin. In this study it was observed that when this peptide is confined between two planar attractive walls, this misfolded state completely disappears from the free-energy surface. The misfolded state was found to be destabilized because the hydrophobic side-chains are most likely to be found on the same side due to their interactions with the confining walls. Although these results

were obtained from a very simple confinement model, they may provide insights into the role of chaperonin confinement in tackling protein misfolding by nonspecific weak interactions.

5.2 Methods

5.2.1 Simulation Details

The Amber03* forcefield [14, 17] was used for the peptide because it has been shown to alleviate biases toward a particular secondary structure [14], and the TIP3P model [72] for water, consistent with earlier studies on this peptide in bulk. The structure of the 16-residue GB1 hairpin was taken from residues 4156 of the full-length GB1 protein (Protein Data Bank ID: 1GB1). The protein was solvated using 3234 water molecules, and eight sodium and five chloride ions were added to neutralize the charge. To simulate confined systems with explicit solvent, we chose the separation between the walls to be 1.8 nm and 2.4 nm, respectively. Simulation boxes of dimensions $5.0233 \times 5.0233 \times 1.8$ nm^3 and $4.1645 \times 4.1645 \times 2.4$ nm^3 were used for the two cases, respectively, and periodic boundary conditions were applied in the x and y directions only. To enhance equilibrium sampling, replica exchange molecular dynamics (REMD) [134] simulations in the NVT ensemble has been performed using 32 replicas spanning a temperature range 278-595K for 500 ns per replica. The following temperatures (in K) has been used: 278, 287, 295, 303, 312, 321, 329, 338, 346, 355, 365, 375, 385, 396, 406, 416, 427, 437, 448, 459, 470, 482, 493, 505, 517, 528, 539, 551, 562, 573, 584, and 595. The Particle Mesh Ewald method [41] has been used to calculate electrostatic interactions with a real space cutoff of 0.9 nm. The cutoff for van der Waals interactions was taken to be 1.4 nm and the parameters

of the Lennard-Jones potential for the cross interactions between nonbonded atoms were obtained from Lorentz-Berthelot combination rules. The attractive confining walls are modeled using the Lennard-Jones 9-3 potential V given by,

$$V(z) = \epsilon \left[\frac{2}{15} \left(\frac{\sigma}{z} \right)^9 - \left(\frac{\sigma}{z} \right)^3 \right],$$ (5.1)

where z is the distance of an atom (protein and solvent) from the wall, σ is the characteristic length scale, and ϵ is the characteristic energy scale. To simulate stable liquid water between planar walls, [103] the values of $\sigma_{\text{wall}-\text{OW}} = 0.346$ nm and $\epsilon_{\text{wall}-\text{OW}} = 1.14$ kcal/mol. All simulations have been performed using GROMACS 4.0.4 [61].

5.2.2 Reaction Coordinates

The fraction of ordered contacts, Q_s, relative to a given structure, s (not necessarily the native state), is defined as

$$Q_s = N_s^{-1} \sum_{i,j} \frac{1}{1 + \exp(\gamma(r_{ij} - \lambda r_{ij}^0))}$$ (5.2)

where the sum runs over the N_s pairs (i, j) of native atomic contacts that are separated by distances r_{ij} in the configuration of interest and by r_{ij}^0 in s ($\gamma = 5\text{Å}^-1$; $\lambda = 1.5$). The parameter λ accounts for the fluctuations in distance between the residues in contact in the native state, and γ controls the steepness of the contact step function [18].

One of the reaction coordinates chosen, the fraction of all-atom (excluding hydrogen atoms) native contact, Q_{aa}, is defined using the native structure (Fig. 5.7F). Q_{aa} has been shown to be a very good reaction coordinate [13, 18]. Q_{aa} has a low value when the protein is unfolded and a value near unity for a folded state. However, this
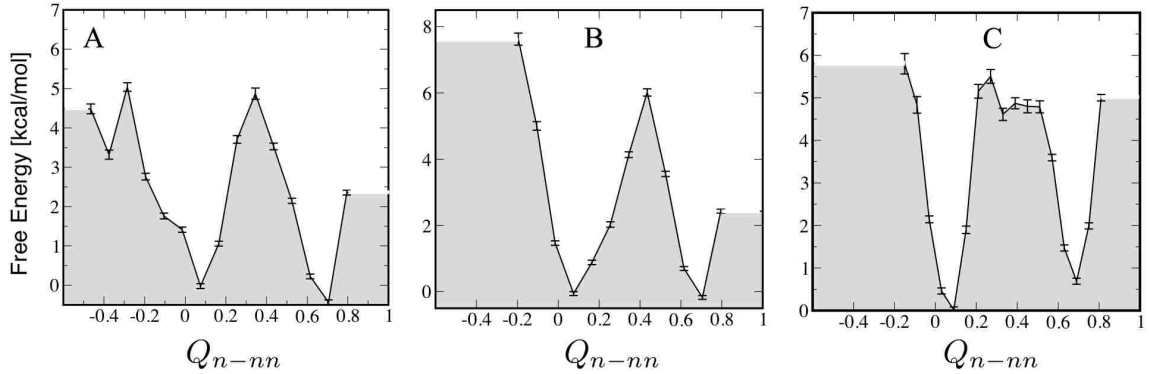
Figure 5.2: Free energy based on Q_{n-nn} of the GB1-harpin for **A.** Bulk **B.** confined between planar walls of distance 1.8 nm between them and **C.** confined between planar walls of distance 2.4 nm between them.

coordinate is not suitable for distinguishing between folded and misfolded states. In the misfolded structure of the hairpin, one of the strands of the hairpin is flipped [20] (Fig. 5.7 M). To separate this native-like off-pathway intermediate, which would otherwise stay close to the folding transition state on Q_{aa}, an alternative coordinate, $Q_{n-nn} = Q_n - Q_{nn}$, where Q_n and Q_{nn} are defined as above using the native and misfolded structures, respectively. For both Q_n and Q_{nn} atom pairs closer than 4.5 $\mathring{A}$ in s, belonging to residues separated by > 3 in the sequence, are considered.

5.2.3 Free-energy Calculations

From Fig. 5.2, the free energy of folding ΔF_{N-U} based on Q_{n-nn} where

$$\Delta F_{N-U} = -k_B T \ln \frac{\int_{Q^{\ddagger}_{n-nn}}^{1} e^{-\beta F(Q_{n-nn})} dQ_{n-nn}}{\int_{-0.2}^{Q^{\ddagger}_{n-nn}} e^{-\beta F(Q_{n-nn})} dQ_{n-nn}} \tag{5.3}$$

is the difference between the native F_N and unfolded F_U free energies, and $Q^{\ddagger}_{n-nn}(= 0.4)$ is the location of the barrier along Q_{n-nn}.

91

Figure 5.3: Two dimensional free-energy surfaces at 303 K are shown as a function of the order parameters Q_{aa} and Q_{n-nn} for the GB1-haripin in bulk. The three distinguishable states are labeled as M: misfolded, U: unfolded and F: folded

5.3   Results and Discussion

To characterize the folding thermodynamics in bulk and under confinement, we calculate the two-dimensional potential of mean force (PMF) as a function of two suitable reaction coordinates. The PMF plot is a convenient way to look at free energy projected onto various order parameters. Moreover, good order parameters are expected to provide an accurate description of system's thermodynamics and kinetics. Here, we use the fraction of heavy atom native contacts, Q_{aa}, [12, 15, 16] and an alternative coordinate, the fraction of native minus non-native contacts, Q_{n-nn}, as two order parameters to project folding free energy onto a low-dimensional surface. Q_{n-nn} was proposed earlier [20] to resolve a native-like off-pathway intermediate in which one strand of the hairpin is "flipped" (fig. 5.7M), which would otherwise lie close to the folding transition state on Q_{aa}. A combination of Q_{aa} and Q_{n-nn} was found to
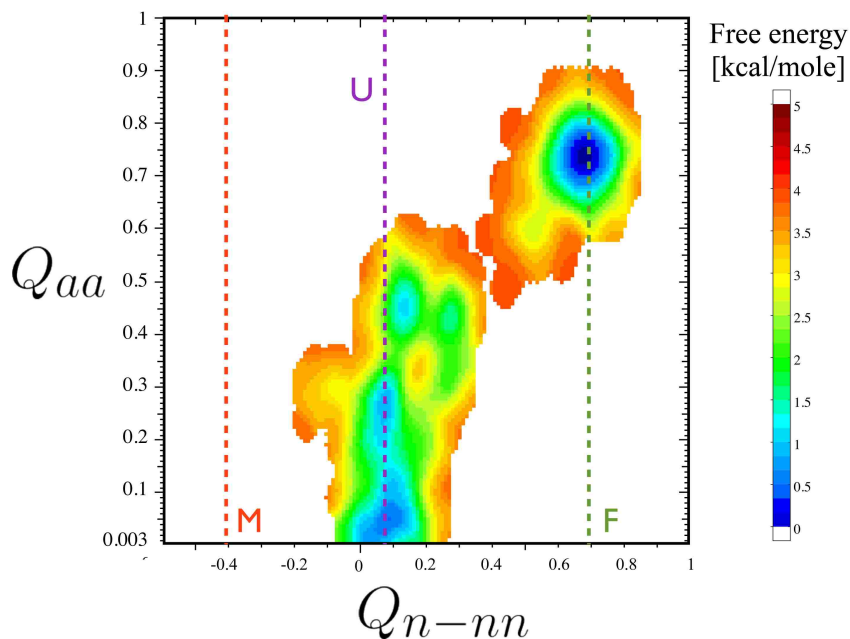
Figure 5.4: Two dimensional free-energy surfaces at 303 K are shown as a function of the order parameters Q_{aa} and Q_{n-nn} for the GB1-haripin confined between LJ walls separated by 1.8 nm . The three distinguishable states are labeled as M: misfolded, U: unfolded and F: folded

provide a good description of the hairpin folding landscape in bulk, and the folding free energy barrier estimated from Q_{n-nn} is consistent with experimental and detailed simulation data [20].

Figures 5.3 shows the folding free energy surfaces (FES) in bulk and figs. 5.4 and 5.5 show the FES under two different types of confinement (1.8 nm and 2.4 nm respectively). The most significant difference between the free energy surfaces in bulk and under confinement is the complete absence of misfolded state ($Q_{n-nn} \approx -0.4$) under confinement. This disappearance of the misfolded basin is observed independent of the confinement size which suggests that it is an interfacial effect presumably due to peptide proximity near one of the walls. To rule out insufficient simulation sampling time as a cause for the observation, addition REMD simulations under confinement
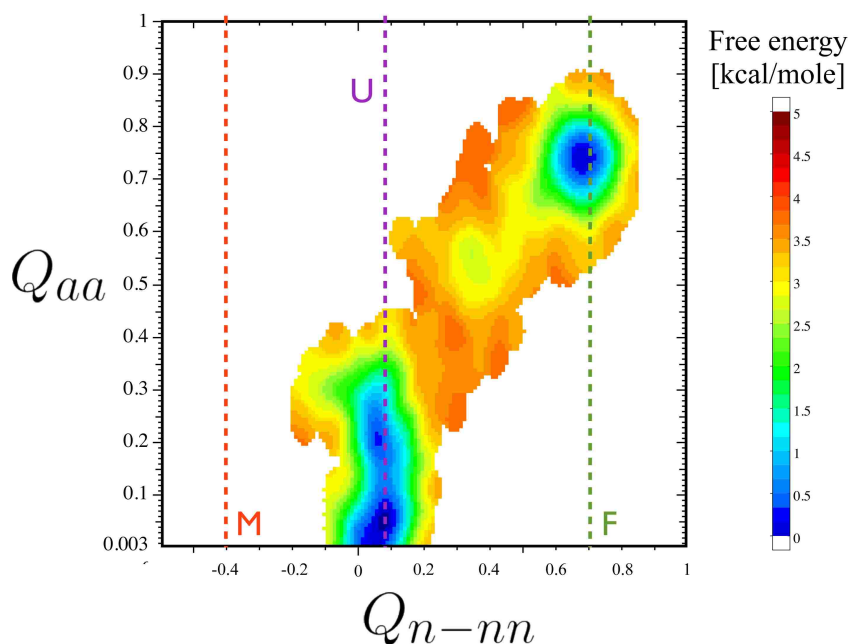
Figure 5.5: Two dimensional free-energy surfaces at 303 K are shown as a function of the order parameters Q_{aa} and Q_{n-nn} for the GB1-haripin confined between LJ walls separated by 2.4 nm . The three distinguishable states are labeled as M: misfolded, U: unfolded and F: folded

were performed in which all the replicas are initiated from the misfolded structure obtained from the bulk simulation. As shown in Fig. 5.10, the misfolded state under confinement is unstable and therefore disappears after a relatively short simulation time.

What is the rationale behind the disappearance of the misfolded basin under confinement? To address this question, we plot the density of the protein (5.8A) and of each residue's side-chain atoms along the z-direction (perpendicular to the walls) (5.9). Because of the attractive nature of the wall surface, the peptide backbone is more likely to be found near one of the walls as shown in (5.8A). More importantly, we find that the hydrophobic side-chain atoms of residues TRP3, TYR5, PHE12, and VAL14 are mostly found near the walls as opposed to other polar and neutral residues.
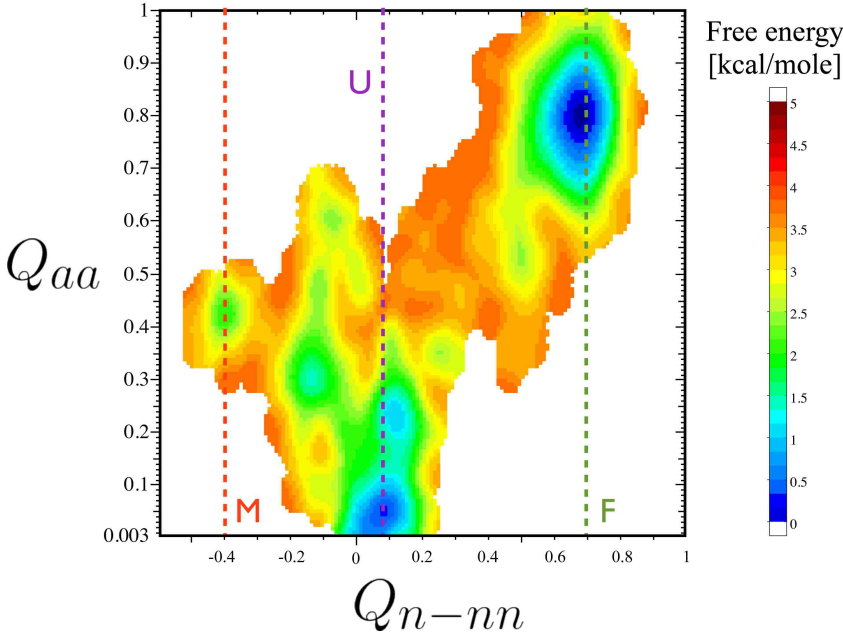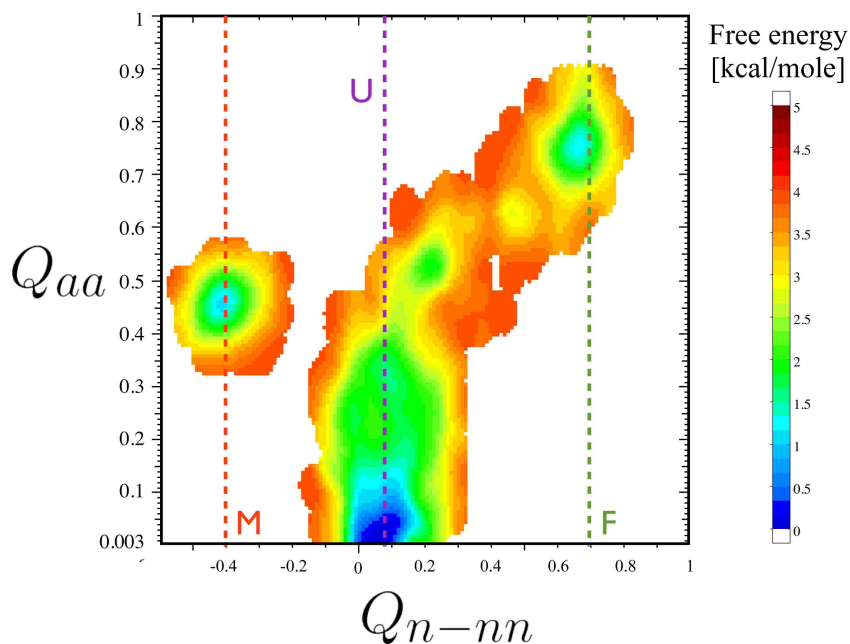
Figure 5.6: Two dimensional free-energy surfaces at 303 K are shown as a function of the order parameters Q_{aa} and Q_{n-nn} for the GB1-haripin confined between LJ walls separated by 1.8 nm . The interaction between the peptide and the walls were set to be repulsive in this case. The three distinguishable states are labeled as M: misfolded, U: unfolded and F: folded
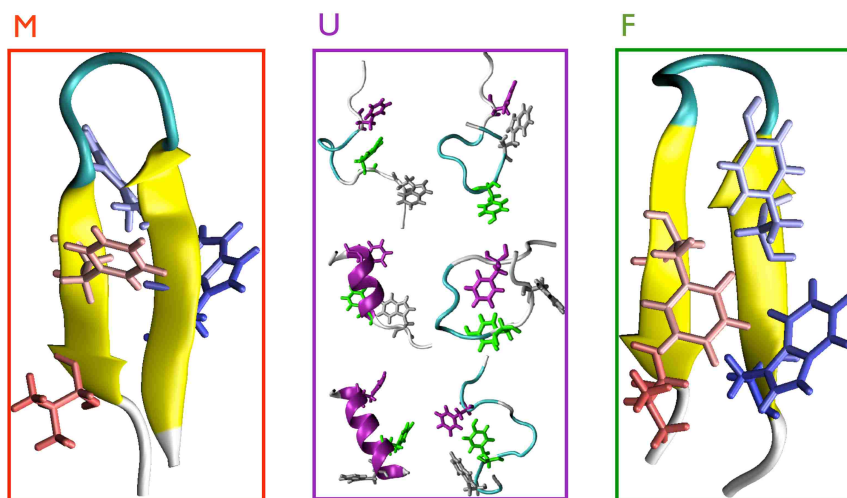
Figure 5.7: The structures of the peptide used in this study **M** misfolded peptide in which one of the strands of the hairpin is 'flipped' **U** a representation of the unfolded ensemble peptide **F** the folded hairpin in the native state.

95
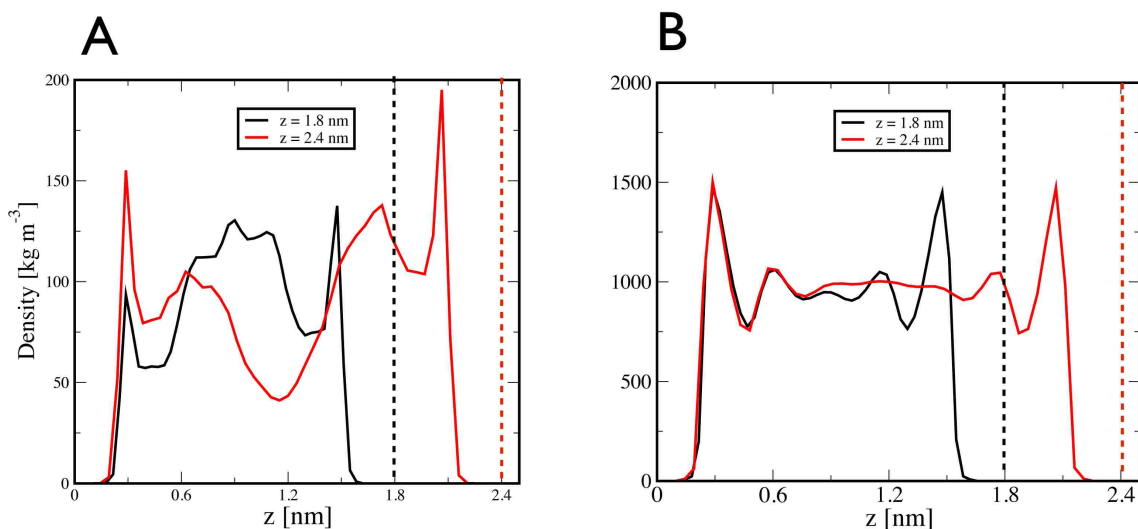
Figure 5.8: **A.** Density of the peptide confined between LJ-walls. **B.** The density of water molecules confined between the LJ-walls. The black and red curves are for cases in which the separation between the walls is 1.8 and 2.4 nm respectively.

Vaitheeswaran and Thirumalai [141] observed a similar preference for hydrophobic side chains to stay near the surface in hydrophobic nanopores, and found it to be related to the peptides stability under confinement. Therefore, configurations with hydrophobic residues on the same side near a wall are preferred due to the favorable association of hydrophobic residues with the wall, and hence the misfolded state, with hydrophobic residues on the opposite side of the backbone, is avoided. We also ran a control simulation in which the peptide-wall interactions were made repulsive (achieved by eliminating the second term in 5.1) but all other parameters were the same as in the attractive-wall simulation for a confinement size of 1.8 nm. As shown in fig. 5.6, the misfolded state in this case is still populated as in bulk. This provides further evidence in favor of our reasoning that the misfolded state in the case of attractive walls is eliminated due to preferential interactions with hydrophobic side chains.
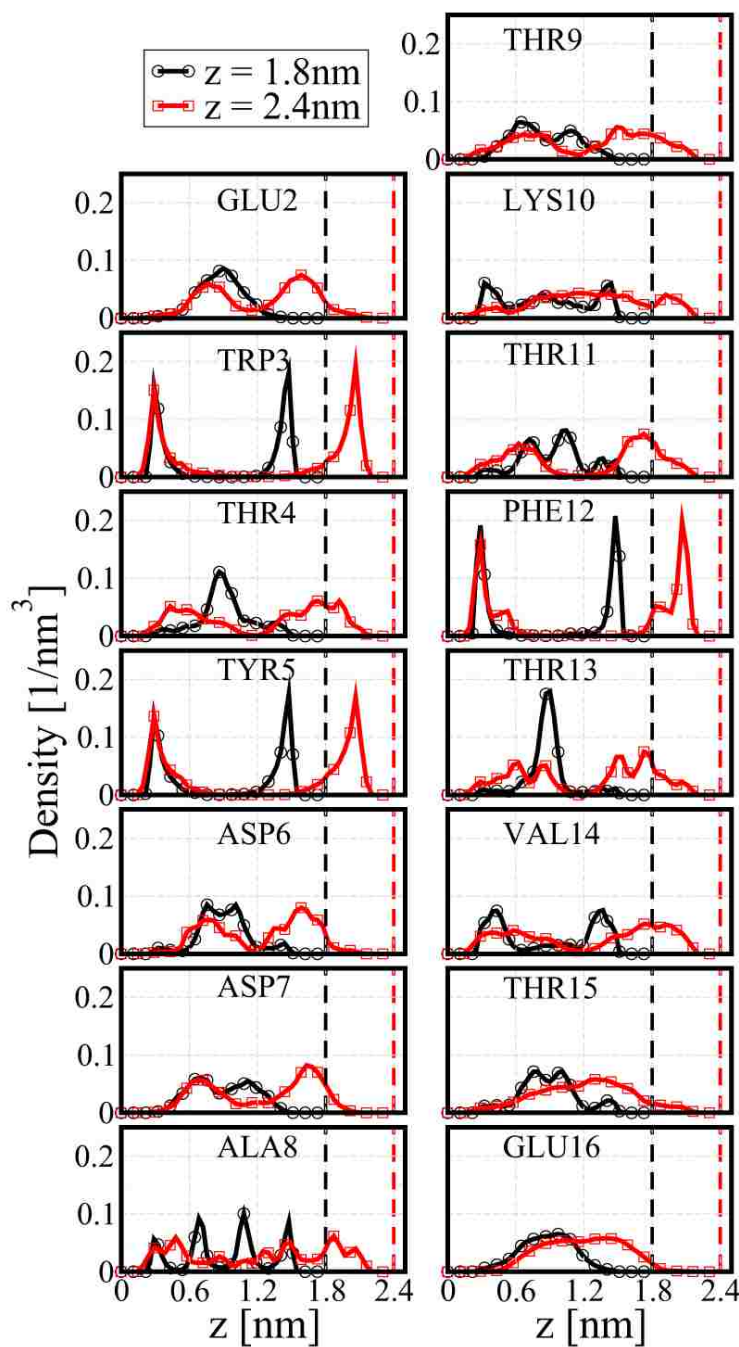
Figure 5.9: Protein side-chain number density profile (normalized by number of side-chain atoms) normal to the confining walls. The left wall is located at z = 0 nm and the right wall is located at z = 1.8 or 2.4 nm (shown by dashed vertical lines). Note that the slight asymmetry results only from statistical error.
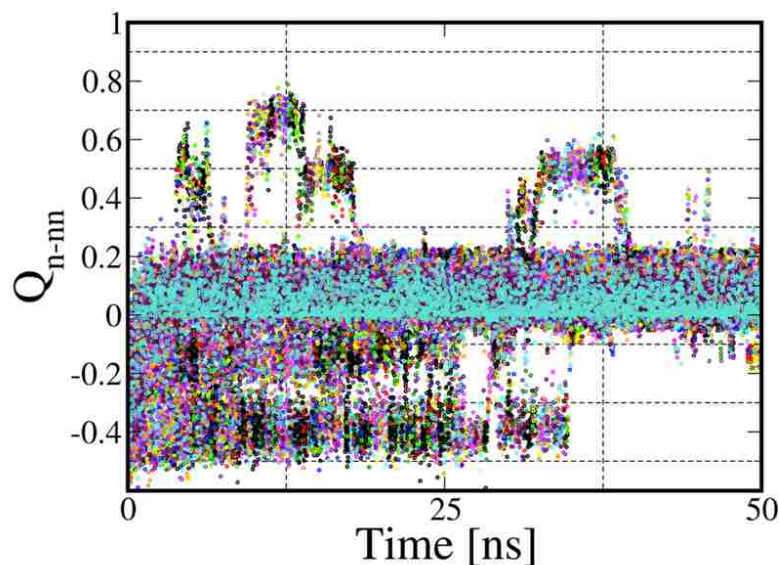
Figure 5.10: Q_{n-nn} as a function of time for all the 32 replicas for simulations initiated from the misfolded state. It is observed that the misfolded state $Q_{n-nn} = 0.45$ is unstable and disappears completely in a relatively short simulation time.

Such a simple mechanism may also be plausible in the case of chaperonin function, as misfolded states appear as traps with nonnative arrangement of hydrophobic residues. Thus, in addition to providing an inert cage (and thereby preventing protein aggregation), protein-chaperonin wall interactions may also facilitate the correct formation of hydrophobic contacts and thus protein folding. In addition, some of the misfolded states may be depopulated within a chaperonin cavity due to these specific protein-chaperonin wall interactions. Our observations here therefore provide support for the hypothesis that confinement can prevent the population of kinetically trapped misfolded states, with the caveat that chaperone systems are more complicated than the one considered here.

As a secondary effect, we find that attractive confinement *destabilizes* the folded state with respect to the unfolded state. The folding free energy $\Delta F_{\mathrm{N-U}}$, based on one-dimensional reaction coordinate Q_{n-nn}, is -0.43 kcal/mol in bulk. Under confinement $\Delta F_{\mathrm{N-U}} = -0.08$ and 0.74 kcal/mol for wall separations 1.8 and 2.4 nm, respectively. The unfolded state stabilization near an attractive wall is expected as larger protein conformations can interact favorably with the wall surface as compared to the compact folded state. This effect is expected to be stronger in case of longer proteins as the size ratio of unfolded to folded state increases.

5.4   Conclusion

In conclusion, extensive molecular simulations has been carried out of the GB1 hairpin using an explicit solvent model in bulk solution as well as confined between planar attractive walls. It is observed that a misfolded state, which is otherwise observed in bulk is completely absent under confinement. It has been shown that hydrophobic residues tend to populate the region near the walls thereby avoiding the misfolded state. The current chapter provides an important insight into the role of direct interactions between protein and chaperonin walls in "smoothing" protein's folding landscape by eliminating off-pathway intermediate states.

Chapter 6

Nanotube confinement de-stabilizes helical propensities of peptides

6.1 Introduction

Besides serving as an excellent model for cylindrical confinement, owing to its remarkable properties such as large surface areas, high mechanical strengths yet ultra-light weights, rich electronic properties and outstanding chemical and thermal stabilities [3], carbon nanotubes continue to be the primary interest for several researchers, who have been exploring their potential in bioapplications [149, 9, 84]. Particularly, there has been significant interest in the ability of CNTs to serve as biocompatible transporters for drugs, genes, proteins and other biomolecules into the cells by penetrating the cellular membrane.

Many diseases are linked to the alterations in the functions of intracellular proteins which are engaged in performing essential biological functions such as enzyme catalysis, signal transduction, gene regulation, to name a few. Therefore, a robust

method of delivering active forms of proteins to specific cells and organs is an important goal in many medical applications, including cancer therapy, vaccination, regenerative medicine, treating loss-of-function genetic diseases and imaging. Target protein cargoes can be loaded into the CNTs using various strategies, such as physical adsorption and covalent/non-covalent encapsulation [52]. It is expected that one of the key functions of CNTs or any nano-carrier for that matter would be to protect the proteins from premature degradation and shield them from various denaturing interactions in the cellular milieu [112, 42]. There have been insightful works on the encapsulation of proteins inside carbon nanotubes [5]. Moreover, quite recently, there has been significant approaches in purifying [76] and synthesizing [54] single walled carbon nanotubes of large diameters. Strategies have been developed for the production of more than 90% population of vertically aligned SWCNTs with diameters ¿ 3nm. The large internal diameters can be utilized for the encapsulation of proteins and other biomolecules.

An obvious question thus arises. How stable are the functional proteins inside single walled carbon nanotubes? Insightful studies have been carried out to understand the stability of secondary structures under nanotube confinement. Modeling the proteins simply as random flight Gaussian chains, it has been theoretically shown [154], that confinement eliminates expanded unfolded conformations, thereby stabilizing the folded state. Ziv *et al.* [159] used a coarse grained representation of a sixteen residue polypeptide chain containing either hydrophobic or hydrophilic beads in a cylindrical confining potential. Using a self-avoiding random walk model of a helical polymer, they showed that α-helices are stabilized under cylindrical confinement and the stabilization is entropic.

However, Sorin and Pande [129] have shown that when polyalanine - an α-helical peptide is confined to a single-walled carbon nanotube with explicit solvent, the results were contrary. They observed that the 23-residue helical polyalanine denatured under confinement. The secondary structure propensity of the polyalanine was shown to be decreasing monotonically with increasing confinement. The observation was explained in terms of solvent entropy. They argued that in bulk, protein folding is expected to maximize solvent entropy which becomes limited under confinement, and thereby intra-peptide interactions, stabilizing secondary structures experience reduced entropic stabilization relative to protein-water interactions. Another study by Lucent *et al.* [92] highlights the importance of solvent in protein folding under confinement.

Another insightful study [109] on the confinement of different helix forming sequences inside a carbon nanotube using a coarse-grained model has elucidated the interplay of various factors such as the strength of nanotube peptide interactions, peptide-sequence and the tube diameter on the stability of helices. They predict that confinement-induced helix stability can be significantly altered by varying the intra-peptide interactions or by changing the interaction strength between the peptide and the nanotube.

In this current chapter, we explore the effects of nanotube confinement, on the stability of peptides. We have performed extensive replica exchange molecular dynamics simulations using all-atom representations of a sixteen residue polyalanine peptide in explicit solvent inside carbon nanotubes of various diameters. The general trend observed from our simulations was that polyalanine, which otherwise possesses a high propensity of α-helical structures in bulk, denatures significantly under the effect of nanotube confinement. We hypothesize that this denaturation is primarily

102

due to the presence of inner-surface of the carbon nanotubes, as the same trend was observed from simulations of polyalanine near a graphite surface.

6.2 Models and Methods

The Amber03* forcefield [14, 17] has been used to model the peptide along with the TIP3P [72] model for water. The forcefield water model combination has been shown to alleviate biases toward a particular secondary structure. We constructed sixteen residue polyalanine sequences and the C and the N terminals were capped using Acyl and amide groups respectively. Single walled armchair Carbon nanotubes with chiral indices ($\{m = n\} = 10, 11, 12, 13, 14, 15, 17, 19, 22$) were constructed using the nanotube builder plugin of Visual Molecular Dynamics package [63]. The length of the carbon nanotubes were set to 4nm. The interaction parameters of the carbon atoms were so chosen such that the carbon-water Lennard-Jones were $\sigma_{CO} = 3.2751 \mathring{A}$ and $\epsilon_{CO} = 0.11433 \, \text{kcal mole}^{-1}$ consistent with [62]. The carbon nanotubes were held fixed in space and was aligned along the z-axis of the simulation box. The peptide in the α-helical ($\phi = -57^o, \psi = -47^o$) configuration was inserted inside the carbon nanotube and an energy minimization was performed for the peptide-nanotube assembly using the conjugate gradient algorithm. The nanotube peptide assembly was centered in a periodic box such that water molecules can diffuse freely in and out of the carbon nanotubes. The peptide-nanotube assembly was solvated such that the P_{zz} element of the pressure tensor was within $\pm$ 100 bars during the production NVT runs. This was ensured by performing short simulations before production runs. All the simulations were carried out in the NVT ensemble using the Nosé-Hoover thermostat. All the simulations were performed using Replica Exchange Molecular Dynamics, using 40
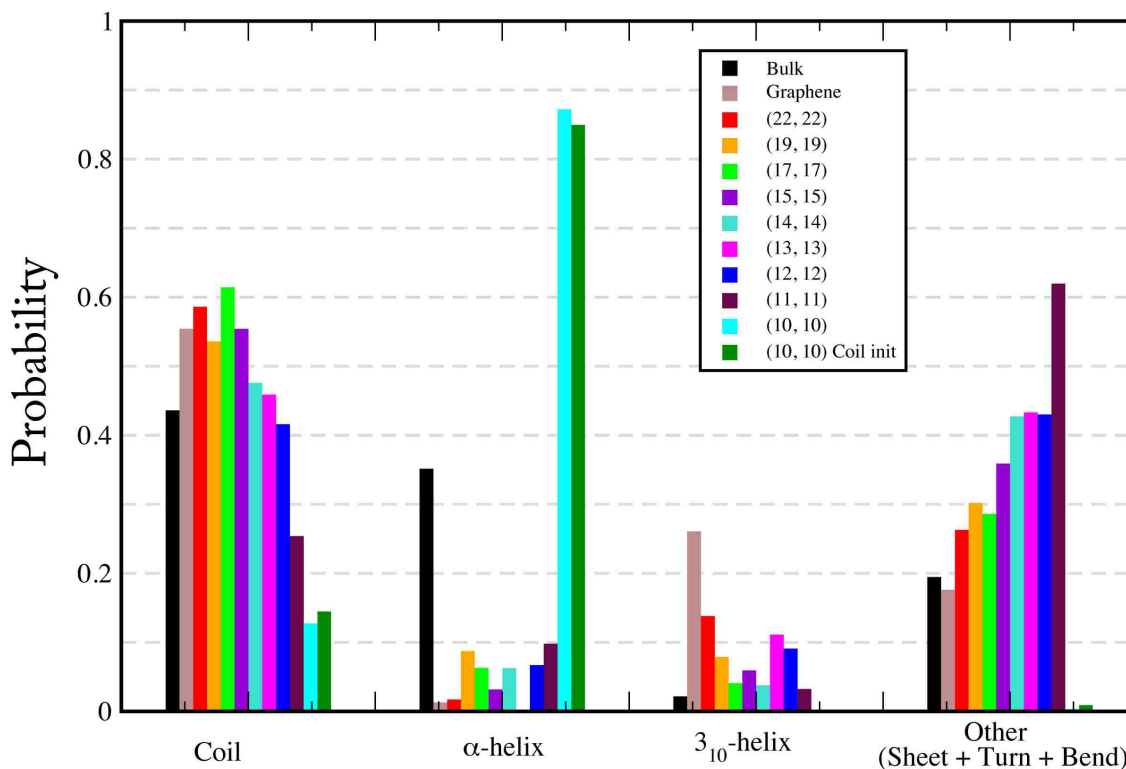
Figure 6.1: Secondary structure propensities of polyalanine (obtained using DSSP analysis) in bulk, near Graphene and under the effects of confinement of various carbon nanotubes simulated. All the simulations were initiated from the α-helical configuration of polyalanine. For confinement inside (10,10) nanotube, additionally, a simulation initiated from the coil state of the peptide was performed.

replicas spanning a temperature range of 300 to 517 K. The Particle Mesh Ewald method [41] has been used to calculate electrostatic interactions with a real space cutoff of 0.9 nm. The cutoff for van der Waals interactions was taken to be 1.2 nm. All the simulations were performed for 20 ns per replica with a total simulation time of 800 ns. The final 10ns of the data was used for analysis. The DSSP algorithm [73] was used to assign secondary structures to the peptides.
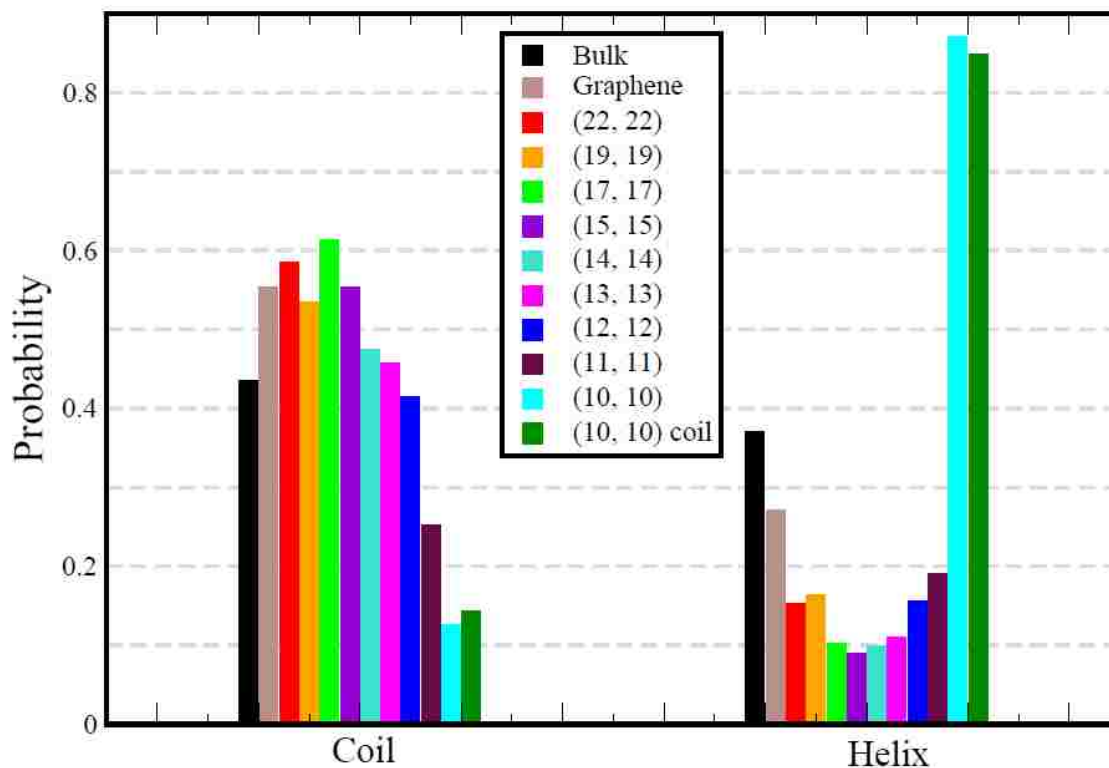
Figure 6.2: econdary structure propensities of polyalanine (obtained using DSSP analysis) in bulk, near Graphene and under the effects of confinement of various carbon nanotubes simulated. Helix refers to the sum of α and 3_{10}-helix.

6.3   Results and Discussion

Figure 6.1 shows the secondary structure propensities from DSSP analysis of the equilibrated trajectory. The propensities of β-sheets, Turns and Bends have been added and represented as Other. It is observed that polyalanine under bulk conditions, has nearly 35% helix propensity to form α-helices. This observed propensity is significantly reduced to 10 % when simulated near graphene or inside carbon nanotubes. An outlier in the series is the (10, 10) nanotube, which stabilizes the α-helix with respect to bulk. To rule insufficient simulation time as a cause of the observation, we performed another REMD simulation initiated with polyalanine in the random coil
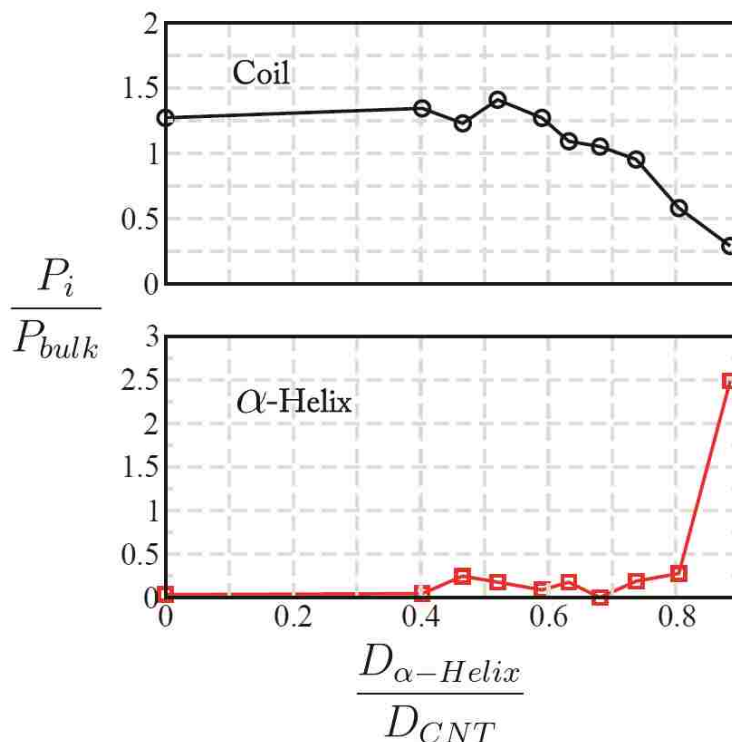
Figure 6.3: The coil and α-helix propensities as obtained from DSSP analysis. The values on the y-axes are normalized by the propensities observed in Bulk whereas the values on the x-axis are normalized by the diameter of an α-helix (1.2 nm).

state under (10, 10) confinement. The observations were consistent. From fig. 6.1, it should also be noted that the 3_{10}-helix propensities for polyalanine is maximum near graphene and, for the cases of confinement, it decreases with increasing confinement. Distinguishing between the two types of helices, experimentally had been quite challenging. Hence, we re-plot figure 6.1 adding the two helicities in fig 6.2. As can be observed, the combined helicity of polyalanine first decreases with increasing confinement (15,15) and seems to increase with increasing confinement upto (10, 10). This observation is contradictory to a previous study by Sorin *et al.* in [129] in which both the helicities were lumped and was shown that the combined helicity decreased with

increasing confinement. The disparity might be attributed to insufficient sampling of standard molecular dynamics simulations over REMD.

In figure 6.3, we plot only the α-helical propensies of polyalanine under confinement normalized by the propensity observed in bulk against the inverse of nanotube diameters multiplied by the diameter of an ideal α-helix (1.2 nm). As can be observed, the ratio of the α-helicities are almost negligible compared to bulk. The singular case of (10, 10) confinement stabilizes helix owing to the fact that the diameter of helix is nearly equal (0.9 times) to the diameter of (10, 10) nanotube. Another interesting observation from figure 6.3 is that the coil propensity decreases with increasing confinement. It should be noted that the β-sheet propensies of polyalanine were negligible, and although there has been work directed towards characterizing 3_{10}-Helices [86], the turn and bend structures characterized by DSSP cannot be probed by traditional experimental methods. The loss of coil propensity, thus, does not imply an increase in helical propensity.

Fig. 6.4 shows conformational free energies of polyalanine in the Ramachandran space. We have shown some representative plots from our analyses. In bulk, polyalanine has a distinct free energy minimum around (ϕ, ψ)=(-57°,-47°) corresponding to α-helices. Also, another distinguishable minimum is around (ϕ, ψ)=(-75°,150°) which corresponds to the polyproline-II helix region. Near a graphene surface, the free energy minimum corresponding to (ϕ, ψ)=(-57°,-47°) in bulk, appears to be shifted slightly towards increasing values of ϕ and ψ. Additionally, a basin corresponding to (ϕ, ψ)=(-75°,-160°) corresponding to extended conformations is observed. These three notable minima appear to be even more prominent under confinement inside carbon nanotubes. This appears to be a common feature under other cases of confinement in our study (data not shown) upto (13, 13). Under (12, 12), however, the
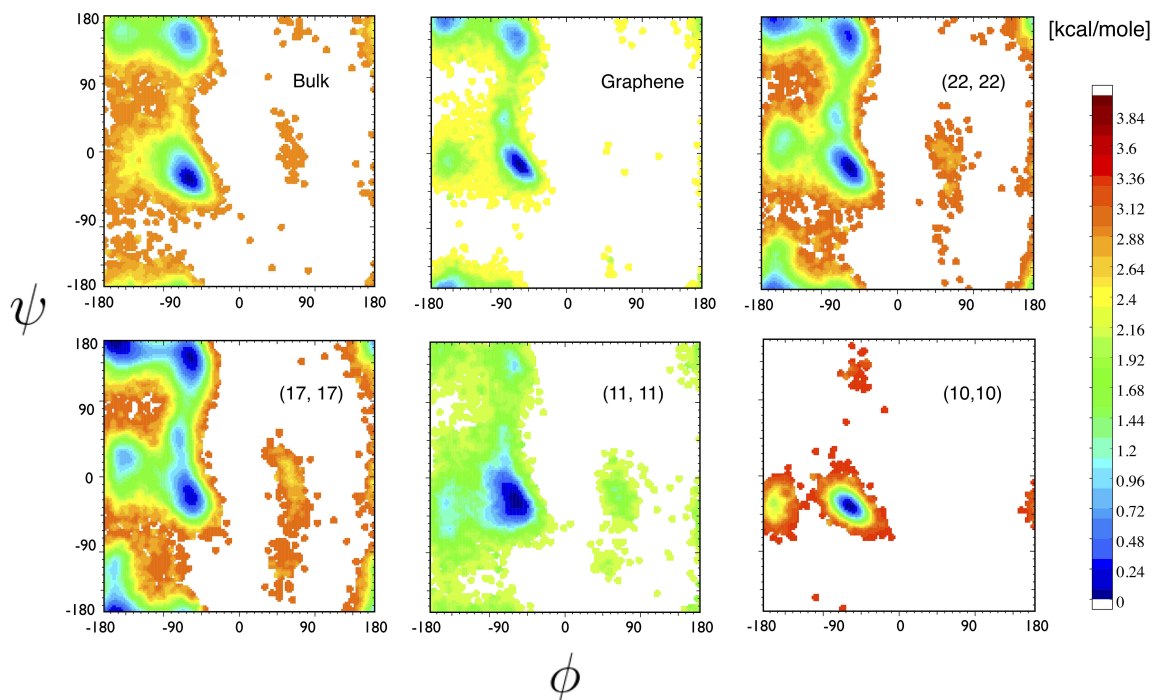
Figure 6.4: Conformational free energies at 300K in the Ramachandran space for polyalanine in bulk, near Graphene and confinement inside (22, 22), (17, 17) (11, 11) and (10, 10) carbon nanotubes.

polyprolene-II helix region is destabilized compared to bulk and the extended conformations observed near graphene surfaces and other cases of nanotube confinement, appear to be destabilized. A broad minimum appears around $-90^o<\phi<-50^o$ and $-50^o<\psi<0^o$. For the case of extreme confinement, (10, 10) nanotube, the only notable free energy minimum, is the α-helical basin corresponding to $(\phi, \psi) = (-57^o,-47^o)$. This observation was consistent from the DSSP analysis as discussed previously. This plot shows the same trend that α-helices denature near graphene surfaces or under nanotube confinement.

6.4   Conclusion

In this chapter, we have performed extensive all-atom replica exchange molecular dynamics simulations with polyalanine in bulk and under the effects of nanotube confinement of varying diameters. The common trend that appears is that carbon nanotubes denature α-helices significantly and that the denaturation is independent of the size of the cylindrical cavity. These results suggest that Carbon nanotubes cannot be effectively used to deliver peptide-based drugs. Although proteins can be easily encapsulated within Carbon nanotubes, the encapsulated protein would denature due to the effects of confinement. However, for a comprehensive understanding of the cause of this denaturation, systematic studies need to be performed varying the hydration of the backbone, and the interaction strengths between the protein and the inner walls of the Carbon nanotube.

Appendix A

Scaled-Particle Theory

The Scaled-particle theory [121] is a statistical mechanical theory of fluids which yields an approximate expression for the reversible work required to introduce a spherical particle into a fluid of spherical particles.

Let us consider a fluid of N hard spheres of diameter $d = 2R$ at a number density ρ. Let $W(R_0)$ be the reversible work required to create a spherical cavity of R_0 centered on a point $\mathbf{r}$ within the fluid. The probability that such a cavity will appear as an outcome of spontaneous thermodynamic fluctuations within the system is

$$p_0(R_0) = \exp[-\beta W(R_0)] \tag{A.1}$$

This is the same as the probability that there are no spheres whose centers lie within the spherical region of radius $R_0 + R$ around $\mathbf{r}$. This interpretation can be extended to negative values of R_0 in the range $-R \leq R_0 \leq 0$, in which case the radius of the region of interest is $0 \leq R_0 + R \leq R$. Since overlap of hard spheres is not possible, the region of interest might have *at most* one particle in such a region. The probability of such a situation is

$$p_1(R_0) = \frac{4}{3}\pi\rho(R_0 + R)^3 = 1 - p_0(R_0) \tag{A.2}$$

Combination of $A.1$ and $A.2$ gives

$$W(R_0) = -k_BT \ln[1 - \frac{4}{3}\pi\rho(R_0 + R)^3], \quad R_0 \leq 0 \tag{A.3}$$

In the other limit for $R_0 \gg R$, the reversible work is given by,

$$W(R_0) = P\Delta V_0 = P\frac{4}{3}\pi(R_0)^3, \quad R_0 \gg R \tag{A.4}$$

Now an assumption is made that for $R_0 > 0, W(R_0)$ is given by a cubic polynomial in R_0 and the term in R_0^3 (the dominant contribution for large cavities) is given by $A.4$.

Expanding $W(R_0)$ around R_0, we have,

$$W(R_0) = W(0) + W'(0)R_0 + \frac{1}{2}W''(0)R_0^2 + \frac{4}{3}\pi P(R_0)^3, \quad R_0 \geq 0 \tag{A.5}$$

Now, at $R_0 = 0$, $W(R_0)$ and its derivatives are continuous. Hence W(0) is given by $A.3$. Consequently,

$$\beta W(0) = -\ln(1 - \eta)$$
$$\beta W'(0) = \frac{4\pi\rho R^2}{1 - \eta}$$
$$\beta W''(0) = \frac{8\pi\rho R}{1 - \eta} + \frac{(4\pi\rho R^2)^2}{(1 - \eta)^2} \tag{A.6}$$

where η is the hard-sphere packing fraction.

The excess chemical potential of the fluid is the reversible work required to insert a hard sphere of radius $R_0 = R$, thus from $A.4$ and $A.6$

$$
\begin{aligned}
\beta \mu^{ex} &= \beta W(R_0) \\
&= -\ln(1-\eta) + \frac{6\eta}{1-\eta} + \frac{9\eta^2}{2(1-\eta)^2} + \frac{\beta P \eta}{\rho}
\end{aligned}
\tag{A.7}
$$

Using, $\frac{\partial P}{\partial \rho} = \rho(\frac{\partial \mu}{\partial \rho})$,

$$
\frac{\beta P}{\rho} = \frac{1+\eta+\eta^2}{(1-\eta)^3}
\tag{A.8}
$$

Putting $A.8$ in $A.7$

$$
W(R_0) = -\ln(1-\eta) + \frac{6\eta}{1-\eta} + \frac{9\eta^2}{2(1-\eta)^2} + \eta \frac{1+\eta+\eta^2}{(1-\eta)^3}
\tag{A.9}
$$

Bibliography

[1] Cameron Abrams and Giovanni Bussi. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy*, 16(1):163–199, December 2013.

[2] Ximena Aguilar, Christoph F Weise, Tobias Sparrman, Magnus Wolf-watz, and Pernilla Wittung-stafshede. Macromolecular Crowding Extended to a Heptameric System : The Co-chaperonin Protein 10. *Biochemistry*, 50:3034–3044, 2011.

[3] P. M. Ajayan. Nanotubes from Carbon. *Chemical Reviews*, 99(7):1787–1800, July 1999.

[4] P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford science publications. Clarendon Press, 1987.

[5] R. Ansari, E. Kazemi, E. Mahmoudinezhad, and F. Sadeghi. Preferred Position and Orientation of Anticancer Drug Cisplatin During Encapsulation Into Single-Walled Carbon Nanotubes. *Journal of Nanotechnology in Engineering and Medicine*, 3(1):010903, 2012.

[6] Alice I Bartlett and Sheena E Radford. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nature Structural and Molecular Biology*, 16(6):582–8, June 2009.

[7] Jyotica Batra, Ke Xu, Sanbo Qin, and Huan-Xiang Zhou. Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophysical Journal*, 97(3):906–11, August 2009.

[8] Jyotica Batra, Ke Xu, and Huan-Xiang Zhou. Nonadditive effects of mixed crowding on protein stability. *Proteins*, 77(1):133–138, October 2009.

[9] R. H. Baughman. Carbon Nanotube Actuators. *Science*, 284(5418):1340–1344, May 1999.

[10] A. Baumketner, A. Jewett, and J.E. Shea. Effects of Confinement in Chaperonin Assisted Protein Folding: Rate Enhancement by Decreasing the Roughness of the Folding Energy Landscape. *Journal of Molecular Biology*, 332(3):701–713, September 2003.

[11] Dor Ben-Amotz and George Stell. Reformulation of WeeksChandlerAndersen Perturbation Theory Directly in Terms of a Hard-Sphere Reference System . *The Journal of Physical Chemistry B*, 108(21):6877–6882, May 2004.

[12] Robert Best and Gerhard Hummer. Diffusive Model of Protein Folding Dynamics with Kramers Turnover in Rate. *Physical Review Letters*, 96(22):2–5, June 2006.

[13] Robert B Best and Gerhard Hummer. Reaction coordinates and rates from transition paths. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6732–6737, May 2005.

[14] Robert B Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The Journal of Physical Chemistry. B*, 113(26):9004–15, July 2009.

[15] Robert B Best and Gerhard Hummer. Coordinate-dependent diffusion in protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3):1088–93, January 2010.

[16] Robert B Best and Jeetain Mittal. Balance between α and β Structures in Ab Initio Protein Folding. *The Journal of Physical Chemistry B.*, 114:8790–8798, 2010.

[17] Robert B Best and Jeetain Mittal. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *The Journal of Physical Chemistry. B*, 114(46):14916–23, November 2010.

[18] Robert B Best and Jeetain Mittal. Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins*, 79(4):1318–28, April 2011.

[19] Robert B Best and Jeetain Mittal. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, 108(27):11087–11092, July 2011.

[20] Robert B Best and Jeetain Mittal. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, 108(27):11087–11092, July 2011.

[21] M R Betancourt and D Thirumalai. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *Journal of Molecular Biology*, 287(3):627–44, April 1999.

[22] A Brinker, G Pfeifer, M J Kerner, D J Naylor, F U Hartl, and M Hayer-Hartl. Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, 107(2):223–33, October 2001.

[23] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*, 30(10, Sp. Iss. SI):1545–1614, JUL 30 2009.

[24] Alan a Chen and Angel E García. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):16820–5, October 2013.

[25] Margaret S Cheung and D Thirumalai. Nanopore-protein interactions dramatically alter stability and yield of the native state in restricted spaces. *Journal of Molecular Biology*, 357(2):632–643, March 2006.

[26] HS Chung, JM Louis, and WA Eaton. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proceedings of the National Academy of Sciences of the United States of America*, 106:11837, 2009.

[27] GM Clore and J Iwahara. and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews*, 109(9):4108–4139, 2009.

[28] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter a. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, May 1995.

[29] Peter B Crowley, Keith Brett, and Jimmy Muldoon. NMR spectroscopy reveals cytochrome c-poly(ethylene glycol) interactions. *ChemBioChem*, 9(5):685–8, March 2008.

[30] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089, 1993.

[31] Jack Douglas, Jacek Dudowicz, and Karl Freed. Crowding Induced Self-Assembly and Enthalpy-Entropy Compensation. *Physical Review Letters*, 103(13):135701, September 2009.

[32] Ron O Dror, Robert M Dirks, J P Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–52, January 2012.

[33] Fen Du, Zheng Zhou, Zhong-Ying Mo, Jun-Zhi Shi, Jie Chen, and Yi Liang. Mixed macromolecular crowding accelerates the refolding of rabbit muscle creatine kinase: implications for protein folding in physiological environments. *Journal of Molecular Biology*, 364(3):469–482, December 2006.

[34] A K Dunker, I Silman, V N Uversky, and J L Sussman. Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6):756–764, 2008.

[35] H Jane Dyson and Peter E Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews. Molecular Cell Biology*, 6(3):197–208, March 2005.

[36] D.K. Eggers and J.S. Valentine. Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Science*, 10(2):250–261, 2001.

[37] Adrian H Elcock. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Current Opinion in Structural Biology*, 20(2):196–206, April 2010.

[38] R J Ellis. Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences*, 26(10):597–604, October 2001.

[39] R John Ellis and Allen P Minton. Protein aggregation in crowded environments. *Biological chemistry*, 387(5):485–97, May 2006.

[40] Donald M Engelman. Membranes are more mosaic than fluid. *Nature*, 438(7068):578–80, December 2005.

[41] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577, 1995.

[42] Amir H Faraji and Peter Wipf. Nanoparticles in cellular drug delivery. *Bioorganic & medicinal chemistry*, 17(8):2950–62, April 2009.

[43] Daria V Fedyukina and Silvia Cavagnero. Protein folding at the exit tunnel. *Annual Review of Biophysics*, 40:337–59, June 2011.

[44] Adedayo a Fodeke and Allen P Minton. Quantitative characterization of temperature-independent and temperature-dependent protein-protein interactions in highly nonideal solutions. *The Journal of Physical Chemistry B.*, 115(38):11261–8, September 2011.

[45] Miriam Friedel, Daniel J. Sheeler, and Joan-Emma Shea. Effects of confinement and crowding on the thermodynamics and kinetics of folding of a minimalist β-barrel protein. *The Journal of Chemical Physics*, 118(17):8106, 2003.

[46] Debabani Ganguly and Jianhan Chen. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. *Proteins: Structure, Function, Bioinformatics*, 79(4):1251–66, April 2011.

[47] Shekhar Garde, Angel E. Garca, Lawrence R. Pratt, and Gerhard Hummer. Temperature dependence of the solubility of non-polar gases in water. *Biophysical Chemistry*, 78(12):21 – 32, 1999.

[48] Alexandra K Gardino, Janice Villali, Aleksandr Kivenson, Ming Lei, Ce Feng Liu, Phillip Steindel, Elan Z Eisenmesser, Wladimir Labeikovsky, Magnus Wolf-Watz, Michael W Clarkson, and Dorothee Kern. Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell*, 139(6):1109–18, December 2009.

[49] AS Garza, N Ahmad, and R Kumar. Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sciences*, 84(7-8), 02 2009.

[50] Todd R. Gingrich and Mark Wilson. On the Ewald summation of Gaussian charges for the simulation of metallic surfaces. *Chemical Physics Letters*, 500(1-3):178–183, November 2010.

[51] Frithjof Godschalk, Samuel Genheden, Pär Söderhjelm, and Ulf Ryde. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Physical chemistry chemical physics : PCCP*, 15(20):7731–9, May 2013.

[52] Zhen Gu, Anuradha Biswas, Muxun Zhao, and Yi Tang. Tailoring nanocarriers for intracellular protein delivery. *Chemical Society Reviews*, 40(7):3638–55, July 2011.

[53] Ming Han and Ulrich H E Hansmann. Replica exchange molecular dynamics of the thermodynamics of fibril growth of Alzheimer's Aβ42 peptide. *The Journal of Chemical Physics*, 135(6):065101, August 2011.

[54] Zhao Jun Han and Kostya Ostrikov. Uniform, dense arrays of vertically aligned, large-diameter single-walled carbon nanotubes. *Journal of the American Chemical Society*, 134(13):6018–24, April 2012.

[55] B Hardesty, T Tsalkova, and G Kramer. Co-translational folding. *Current Opinion in Structural Biology*, 9(1):111–114, February 1999.

[56] F Ulrich Hartl and Manajit Hayer-Hartl. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science (New York, N.Y.)*, 295(5561):1852–8, March 2002.

[57] Danny M Hatters, Allen P Minton, and Geoffrey J Howlett. Macromolecular crowding accelerates amyloid formation by human apolipoprotein C-II. *The Journal of Biological Chemistry*, 277(10):7824–30, March 2002.

[58] Manajit Hayer-Hartl and Allen P Minton. A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. *Biochemistry*, 45(44):13356–60, November 2006.

[59] Jan Hermans. The amino acid dipeptide: small but still influential after 50 years. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8):3095–6, February 2011.

[60] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, September 1997.

[61] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, March 2008.

[62] Gerhard Hummer, J C Rasaiah, and J P Noworyta. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature*, 414(6860):188–90, November 2001.

[63] W Humphrey, A Dalke, and K. Humphrey, W., Dalke, A. and Schulten. Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[64] PH Hünenberger. Thermostat algorithms for molecular dynamics simulations. *Advanced Polymer Science*, 173:105–149, 2005.

[65] Lilia M. Iakoucheva, Celeste J. Brown, J.David Lawson, Zoran Obradović, and A.Keith Dunker. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology*, 323(3):573–584, October 2002.

[66] Lilia M Iakoucheva, Predrag Radivojac, Celeste J Brown, Timothy R O'Connor, Jason G Sikes, Zoran Obradovic, and a Keith Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, 32(3):1037–49, January 2004.

[67] Ankur Jain, Ruijie Liu, Biswarathan Ramani, Edwin Arauz, Yuji Ishitsuka, Kaushik Ragunathan, Jeehae Park, Jie Chen, Yang K Xiang, and Taekjip Ha. Probing cellular protein complexes using single-molecule pull-down. *Nature*, 473(7348):484–8, May 2011.

[68] Sumanth N. Jamadagni, Rahul Godawat, and Shekhar Garde. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annual Review of Chemical and Biomolecular Engineering*, 2(1):147–171, July 2011.

[69] TC C Jarvis and DM M Ring. Macromolecular Crowding: Thermodynamic Consequences for Protein-Protein Interactions within the T4 DNA Replication Complex. *Journal of Biological Chemistry*, 265(25):15160–15167, 1990.

[70] B Jayaram and Tarun Jain. The role of water in protein-DNA recognition. *Annual review of biophysics and biomolecular structure*, 33:343–61, January 2004.

[71] Ming Jiao, Hong-Tao Li, Jie Chen, Allen P Minton, and Yi Liang. Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophysical Journal*, 99(3):914–23, August 2010.

[72] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926, 1983.

[73] Wolfgang;Sander Christian Kabsch and Christian Sander. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geomtrical Features. *Biopolymers*, 22(12):2577–2637, 1983.

[74] John Karanicolas and C.L. Brooks III. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Science*, 11(10):2351–2361, 2002.

[75] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–52, September 2002.

[76] Masatoshi Kawai, Haruka Kyakuno, Takuya Suzuki, Toru Igarashi, Hironori Suzuki, Toshiya Okazaki, Hiromichi Kataura, Yutaka Maniwa, and Kazuhiro Yanagi. Single chirality extraction of single-wall carbon nanotubes for the encapsulation of organic molecules. *Journal of the American Chemical Society*, 134(23):9545–8, June 2012.

[77] Jun Soo Kim and Arun Yethiraj. Effect of macromolecular crowding on reaction rates: a computational and theoretical study. *Biophysical Journal*, 96(4):1333–40, February 2009.

[78] Young C Kim, Robert B Best, and Jeetain Mittal. Macromolecular crowding effects on protein-protein binding affinity and specificity. *The Journal of Chemical Physics*, 133(20):205101, November 2010.

[79] Young C Kim and Gerhard Hummer. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *Journal of Molecular Biology*, 375(5):1416–33, February 2008.

[80] Young C. Kim and Jeetain Mittal. Crowding Induced Entropy-Enthalpy Compensation in Protein Association Equilibria. *Physical Review Letters*, 110(20):208102, May 2013.

[81] Young C Kim, Chun Tang, G Marius Clore, and Gerhard Hummer. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proceedings of the National Academy of Sciences of the United States of America*, 105:12855–12860, 2008.

[82] John L Klepeis, Kresten Lindorff-Larsen, Ron O Dror, and David E Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, 19(2):120–7, April 2009.

[83] D K Klimov, D Newfield, and D Thirumalai. Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):8019–24, June 2002.

[84] K Kostarelos, A Bianco, and M Prato. Promises, facts and challenges for carbon nanotubes in imaging and therapeutics. *Nature Nanotechnology*, 4(10):627–33, October 2009.

[85] Noga Kozer and Gideon Schreiber. Effect of crowding on protein-protein association rates: fundamental differences between low and high mass crowding agents. *Journal of Molecular Biology*, 336(3):763–74, February 2004.

[86] Ahmed Lakhani, Anjan Roy, Matteo De Poli, Marcelo Nakaema, Fernando Formaggio, Claudio Toniolo, and Timothy a Keiderling. Experimental and theoretical spectroscopic study of 3(10)-helical peptides using isotopic labeling to evaluate vibrational coupling. *The Journal of Physical Chemistry. B*, 115(19):6252–64, May 2011.

[87] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education. Prentice Hall, 2001.

[88] J. L. Lebowitz and J. S. Rowlinson. Thermodynamic Properties of Mixtures of Hard Spheres. *The Journal of Chemical Physics*, 41(1):133, 1964.

[89] T Lengauer and M Rarey. Computational methods for biomolecular docking. *Current Opinion in Structural Biology*, 6(3):402–6, June 1996.

[90] Xiaofan Li, Iain H Moal, and Paul a Bates. Detection and refinement of encounter complexes for protein-protein docking: taking account of macromolecular crowding. *Proteins*, 78(15):3189–96, November 2010.

[91] Pu Liu, Byungchan Kim, Richard a Friesner, and B J Berne. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13749–13754, September 2005.

[92] Del Lucent, V Vishal, and Vijay S Pande. Protein folding under confinement: a role for solvent. *Proceedings of the National Academy of Sciences of the United States of America*, 104(25):10430–4, June 2007.

[93] BA Luty and WF van Gunsteren. Calculating electrostatic interactions using the particle-particle particle-mesh method with nonperiodic long-range interactions. *The Journal of Physical Chemistry*, 100(7):2581–2587, 1996.

[94] Alexander D Mackerell. Empirical force fields for biological macromolecules: overview and issues. *Journal of Computational Chemistry*, 25(13):1584–604, October 2004.

[95] Kristen a Marino and Peter G Bolhuis. Confinement-induced states in the folding landscape of the Trp-cage miniprotein. *The Journal of Physical Chemistry B.*, 116(39):11872–80, October 2012.

[96] Sean R McGuffee and Adrian H Elcock. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Computational Biology*, 6(3):e1000694, March 2010.

[97] Heleen Meuzelaar, Kristen a Marino, Adriana Huerta-Viga, Matthijs R Panman, Linde E J Smeenk, Albert J Kettelarij, Jan H van Maarseveen, Peter Timmerman, Peter G Bolhuis, and Sander Woutersen. Folding dynamics of the Trp-cage miniprotein: evidence for a native-like intermediate from combined time-resolved vibrational spectroscopy and molecular dynamics simulations. *The Journal of Physical Chemistry B.*, 117(39):11490–501, October 2013.

[98] Allen P Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Molecular and Cellular Biochemistry*, 55:119–140, 1983.

[99] Allen P. Minton and J Wilf. Effect of Macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*, 20(17):4821–4826, August 1981.

[100] J Mittal, JR R Errington, and TM M Truskett. Relationships between self-diffusivity, packing fraction, and excess entropy in simple bulk and confined fluids. *The Journal of Physical . . .*, 111:10054–10063, 2007.

[101] Jeetain Mittal and Robert B Best. Thermodynamics and kinetics of protein folding under confinement. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51):20233–20238, December 2008.

[102] Jeetain Mittal and Robert B Best. Dependence of protein folding stability and dynamics on the density and composition of Macromolecular crowders. *Biophysical Journal*, 98(2):315–320, January 2010.

[103] Jeetain Mittal and Gerhard Hummer. Interfacial thermodynamics of confined water near molecularly rough surfaces. *Faraday Discussions*, 146:341–352, 2010.

[104] Anthony Mittermaier and Lewis E Kay. New tools provide new insights in NMR studies of protein dynamics. *Science (New York, N.Y.)*, 312(5771):224–8, April 2006.

[105] S Miyazawa and R L Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of Molecular Biology*, 256(3):623–44, March 1996.

[106] A S Morar, X Wang, and G J Pielak. Effects of crowding by mono-, di-, and tetrasaccharides on cytochrome c-cytochrome c peroxidase binding: comparing experiment to theory. *Biochemistry*, 40(1):281–5, January 2001.

[107] Smita Mukherjee, Pramit Chowdhury, and Feng Gai. Tuning the cooperativity of the helix-coil transition by aqueous reverse micelles. *The Journal of Physical Chemistry B.*, 110(24):11615–9, June 2006.

[108] K Namba. Roles of partly unfolded conformations in macromolecular self-assembly. *Genes to Cells*, 6(1):1–12, January 2001.

[109] Edward P O'Brien, George Stan, D Thirumalai, and Bernard R Brooks. Factors governing helix formation in peptides confined to carbon nanotubes. *Nano Letters*, 8(11):3702–8, November 2008.

[110] Chetan N Patel, Schroeder M Noble, Gresham T Weatherly, Ashutosh Tripathy, Donald J Winzor, and Gary J Pielak. Effects of molecular crowding by saccharides on alpha-chymotrypsin dimerization. *Protein Science*, 11:997–1003, 2002.

[111] David a. Pearlman, David a. Case, James W. Caldwell, Wilson S. Ross, Thomas E. Cheatham, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, September 1995.

[112] Dan Peer, Jeffrey M Karp, Seungpyo Hong, Omid C Farokhzad, Rimona Margalit, and Robert Langer. Nanocarriers as an emerging platform for cancer therapy. *Nature Nanotechnology*, 2(12):751–60, December 2007.

[113] Ronald W Peterson, Karthik Anbalagan, Cecilia Tommos, and a Joshua Wand. Forced folding and structural analysis of metastable proteins. *Journal of the American Chemical Society*, 126(31):9498–9, August 2004.

[114] Yael Phillip, Vladimir Kiss, and Gideon Schreiber. Protein-binding dynamics imaged in a living cell. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5):1461–6, January 2012.

[115] Yael Phillip, Eilon Sherman, Gilad Haran, and Gideon Schreiber. Common crowding agents have only a small effect on protein-protein interactions. *Biophysical Journal*, 97(3):875–85, August 2009.

[116] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–802, December 2005.

[117] I Radhakrishnan, G C Pérez-Alvarado, H J Dyson, and P E Wright. Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS letters*, 430(3):317–22, July 1998.

[118] I Radhakrishnan, G C Pérez-Alvarado, D Parker, H J Dyson, M R Montminy, and P E Wright. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell*, 91(6):741–52, December 1997.

[119] Nitin Rathore, Thomas a Knotts, and Juan J de Pablo. Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations. *Biophysical Journal*, 90(5):1767–73, March 2006.

[120] Revanur Ravindra, Shuang Zhao, Hermann Gies, and Roland Winter. Protein encapsulation in mesoporous silicate: the effects of confinement on protein stability, hydration, and volumetric properties. *Journal of the American Chemical Society*, 126(39):12224–12225, October 2004.

[121] H. Reiss, H. L. Frisch, and J. L. Lebowitz. Statistical Mechanics of Rigid Spheres. *The Journal of Chemical Physics*, 31(2):369, 1959.

[122] Young Min Rhee, Eric J Sorin, Guha Jayachandran, Erik Lindahl, and Vijay S Pande. Simulations of the role of water in the protein-folding mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6456–61, April 2004.

[123] G Rivas, J A Fernández, and a P Minton. Direct observation of the enhancement of noncooperative protein self-assembly by macromolecular crowding: indefinite linear self-association of bacterial cell division protein FtsZ. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6):3150–5, March 2001.

[124] Jonathan Rosen, Young C Kim, and Jeetain Mittal. Modest protein-crowder attractive interactions can counteract enhancement of protein association by intermolecular excluded volume interactions. *The Journal of Physical Chemistry B.*, 115(11):2683–2689, March 2011.

[125] R Dustin Schaeffer, Alan Fersht, and Valerie Daggett. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Current Opinion in Structural Biology*, 18(1):4–9, February 2008.

[126] G Schreiber, G Haran, and H-X Zhou. Fundamental aspects of protein-protein association kinetics. *Chemical Reviews*, 109(3):839–860, March 2009.

[127] Lukas D. Schuler, Xavier Daura, and Wilfred F. van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205–1218, 2001.

[128] Karim Snoussi and Bertil Halle. Protein Self-Association Induced by Macro-molecular Crowding: A Quantitative Analysis by Magnetic Relaxation Dispersion. *Biophysical Journal*, 88(4):2855–2866, December 2008.

[129] Eric J Sorin and Vijay S Pande. Nanotube confinement denatures protein helices. *Journal of the American Chemical Society*, 128(19):6316–6317, May 2006.

[130] Marcos Sotomayor and Klaus Schulten. Single-molecule experiments in vitro and in silico. *Science (New York, N.Y.)*, 316(5828):1144–8, May 2007.

[131] R.J. Speedy, F.X. Prielmeier, T. Vardag, E.W. Lang, and H.-D. Ldemann. Diffusion in simple fluids. *Molecular Physics*, 66(3):577–590, 1989.

[132] Loren Stagg, Shao-Qing Zhang, Margaret S Cheung, and Pernilla Wittung-Stafshede. Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proceedings of the National Academy of Sciences of the United States of America*, 104(48):18976–81, November 2007.

[133] Kenji Sugase, H Jane Dyson, and Peter E Wright. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–5, June 2007.

[134] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(November):141–151, January 1999.

[135] Fumiko Takagi, Nobuyasu Koga, and Shoji Takada. How protein thermody-namics and folding mechanisms are altered by the chaperonin cage : Molecular Simulations. *Proceedings of the National Academy of Sciences*, 100(20):11367–11372, 2003.

[136] Tsuyoshi Terakawa, Tomoshi Kameda, and Shoji Takada. On Easy Implementa-tion of a Variant of the Replica Exchange with Solute Tempering in GROMACS. *Journal of Computational Chemistry*, 2010.

[137] S Thompson. Use of neighbor lists in molecular dynamics. *Information Quaterly, CCP5*, 1983.

[138] Jianhui Tian and Angel E Garcia. Simulation studies of protein fold-ing/unfolding equilibrium under polar and non-polar confinement. *Journal of the American Chemical Society*, August 2011.

[139] M. Tuckerman. *Statistical Mechanics and Molecular Simulations*. Oxford Grad-uate Texts. Oxford University Press, UK, 2008.

[140] Adrian Gustavo Turjanski, J Silvio Gutkind, Robert B Best, and Gerhard Hummer. Binding-induced folding of a natively unstructured transcription factor. *PLoS Computational Biology*, 4(4):e1000060, April 2008.

[141] S Vaitheeswaran and D Thirumalai. Interactions between amino acid side chains in cylindrical hydrophobic nanopores with applications to peptide stability. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):17636–41, November 2008.

[142] B van den Berg, R J Ellis, and C M Dobson. Effects of macromolecular crowding on protein folding and aggregation. *The EMBO journal*, 18(24):6927–33, December 1999.

[143] L Verlet. Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98–103, 1967.

[144] Lingle Wang, B J Berne, and Richard a Friesner. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6):1937–42, February 2012.

[145] Lingle Wang, Richard A RA Friesner, and BJ J Berne. Replica Exchange with Solute Scaling : A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B.*, 115:9431–9438, 2011.

[146] Qinghua Wang, Anastasia Zhuravleva, and Lila M Gierasch. Exploring weak, transient protein-protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry*, 50(43):9225–36, November 2011.

[147] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247, 1971.

[148] J R Wenner and V a Bloomfield. Crowding effects on EcoRV kinetics and binding. *Biophysical Journal*, 77(6):3234–41, December 1999.

[149] Stanislaus S. Wong, Ernesto Joselevich, Adam T. Woolley, Chin Li Cheung, and Charles M. Lieber. Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology. *Nature*, 394:52–55, 1998.

[150] Peter E Wright and H Jane Dyson. Linking folding and binding. *Current Opinion in Structural Biology*, 19(1):31–8, March 2009.

[151] In-Chul Yeh and Max L. Berkowitz. Ewald summation for systems with slab geometry. *The Journal of Chemical Physics*, 111(7):3155, 1999.

[152] Jason C. Young, Vishwas R. Agashe, Katja Siegers, and F. Ulrich Hartl. Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol*, 5(10):781–791, 10 2004.

[153] Huan-Xiang Zhou. Loops, linkages, rings, catenanes, cages, and crowders: entropy-based strategies for stabilizing proteins. *Accounts of Chemical Research*, 37(2):123–30, February 2004.

[154] Huan-Xiang Zhou and Ken A Dill. New Concepts Stabilization of Proteins in Confined Spaces. *Biochemistry*, 40(38):11289–11293, 2001.

[155] Huan-Xiang Zhou, Germán Rivas, and Allen P Minton. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual review of biophysics*, 37:375–97, January 2008.

[156] Ruhong Zhou. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins*, 53(2):148–61, November 2003.

[157] Zhuoyun Zhuang, Andrew I Jewett, Silvan Kuttimalai, Giovanni Bellesia, S Gnanakaran, and Joan-Emma Shea. Assisted peptide folding by surface pattern recognition. *Biophysical Journal*, 100(5):1306–15, March 2011.

[158] SB B Zimmerman and Allen P AP Minton. Macromolecular Crowding: Biochemical, Biophysical, and Physiological consequences. *Annual review of biophysics and biomolecular structure*, 22:27–65, 1993.

[159] Guy Ziv, Gilad Haran, and D Thirumalai. Ribosome exit tunnel can entropically stabilize alpha-helices. *Proceedings of the National Academy of Sciences*, 102:18956–18961, 2005.

[160] Silvia Zorrilla, G Rivas, AU U Acuña, and MP P Lillo. Protein self-association in crowded protein solutions : A time-resolved fluorescence polarization study. *Protein Science*, 13:2960–2969, 2004.

[161] Daniel M Zuckerman. Equilibrium sampling in biomolecular simulations. *Annual Review of Biophysics*, 40:41–62, January 2011.

Biography

Apratim Bhattacharya was born on 18 September, 1987 to Swapan Kumar Bhattacharyya and Gayatri Bhattacharyya in Calcutta, West Bengal, India. He attended St. Xavier's Collegiate School, Calcutta. He has received his Bachelor of Technology in Chemical Engineering from Heritage Institute of Technology, Kolkata in 2010. Upon completion of his undergraduate studies, he joined Lehigh University to pursue his PhD in Chemical Engineering. During the course of his undergraduate studies, he was selected to participate in a Research Experience for Undergraduates program at the New Jersey Institute of Technology, Newark, NJ, USA. Subsequently, he had worked at the Membrane Separations Laboratory at Jadavpur University, Kolkata which motivated him to pursue graduate school. He has worked for M.N. Dastur and Co. Consulting Engineers during the summer of 2009, and the Computational Modeling Center at Air Products and Chemicals, during the Spring of 2014. He is also an avid Philatelist and an accomplished vocalist.

Curriculum Vitae

Education

Ph. D. in Chemical Engineering
Graduation Date: September, 2014
Lehigh University, Bethlehem, PA, USA. GPA: 3.89/4.0

B. Tech in Chemical Engineering
Graduation Date: July, 2010
Heritage Institute of Technology, Kolkata, West Bengal, India. GPA: 9.18/10.0

Research Experience

PhD Intern, Computational Modeling Center,
Air Products and Chemicals, Trexlertown, PA, USA. Spring 2014

Graduate Research Assistant, Dept. of Chemical Engineering,
Lehigh University, Bethlehem, PA, USA. 2010 - 2014

Research Experience for Undergraduates Program
Dept. of Chemical Engineering
New Jersey Institute of Technology, Newark, NJ, USA. Summer 2008

Independent Research, Dept. of Chemical Engineering
Jadavpur University, Kolkata, West Bengal, India. 2008 - 2010

Other Experience

Summer Intern
M.N. Dastur and Co. Pvt. Ltd. Consulting Engineers, Kolkata, West Bengal, India
Summer 2009.

Publications in peer-reviewed journals

- **A. Bhattacharya**, Y.C. Kim, J. Mittal. Protein-protein interactions in a crowded environment. *Biophysical Reviews*, 5, 99, 2013.

- **A. Bhattacharya**, R.B. Best, J. Mittal. Smoothing of the GB1 Hairpin Folding Landscape by Interfacial Confinement, *Biophysical Journal*, 103(3), 596-600, 2012.

- D. Sen, A. Roy, **A. Bhattacharya**, D. Banerjee, C. Bhattacharjee, Development of a knowledge based hybrid neural network (KBHNN) for studying the effect of diafiltration during ultrafiltration of whey, *Desalination*, 273 (1), 168, 2011.

- D. Sarkar, **A. Bhattacharya**, C. Bhattacharjee, Modeling the performance of a standard single stirred ultrafiltration cell using variable velocity back transport flux, *Desalination*, 261, 89, 2010.

- Debasish Sarkar, **A. Bhattacharya**, Chiranjib Bhattacharjee, Dulal C. Mukherjee. Modeling and Simulation of Unstirred Dead End Ultrafiltration of Macromolecules. *Journal of the Indian Chemical Society*, 87(1),10511, 2010.

Oral Presentations

- **A. Bhattacharya**, J. Mittal, Protein folding and protein-protein interactions in a cell-like environment, Graduate Student Seminar, Dept. of Chemical Engineering, Lehigh University, PA, 2013.

- **A. Bhattacharya**, Y. C. Kim, J. Mittal, Macromolecular Crowding Effects On Coupled Folding and Binding. AIChE Annual Meeting, Pittsburgh, PA, 2012.

- **A. Bhattacharya**, R. B. Best, J. Mittal, Confinement and Protein-Wall Interactions Can Significantly Alter Protein Folding Landscape. AIChE Annual Meeting, Pittsburgh, PA, 2011.

- **A. Bhattacharya**, Anirban Roy, Debasish Sarkar, Chiranjib Bhattacharjee, Modeling of absorption of ammonia from coke oven gas by water in a falling liquid film. Chemcon 2009, Visakhapatnam.

Poster Presentations

A. Bhattacharya, R. B. Best, J. Mittal, Transformations in the Protein Folding Landscape by Interfacial Confinement. Presented in:

- 244th National Meeting of the American Chem. Soc., Philadelphia, PA, 2012.

- Pennsylvania Network Meeting, Physics and Chemistry of Biological Systems, Lehigh University, PA, 2012.

- NSF Workshop on Multiscale Modeling and Simulation at Institut Pasteur de Montevideo, Uruguay, Sept. 2012.

Awards and Honors

- **Baldwin Fellowship**, Lehigh University, Spring 2011, Spring 2014.

- **Gotshall Fellowship**, Lehigh University, Spring 2012, Fall 2012.

- **Travel grant** to participate at the **NSF Workshop on Multiscale Modeling and Simulation at Institut Pasteur de Montevideo, Uruguay**, Sept. 2012

Leadership and other activities

- Led a team to implement Google Apps at Heritage Institute of Technology.

- Avid vocalist specializing in Indian music with solo and choral experience at various events.

- Philatelist by hobby with exhibits earning acclaims at International Philatelic Exhibitions.

- Served as Treasurer of India Club (Indian Graduate Student association) at Lehigh University.

- Served as Historian and Webmaster of Chemical Engg. Graduate association at Lehigh University.

- Served as the Secretary of Rotaract Club of Heritage Institute of Technology.

- hello world