Molecular Simulations of Chaperonins

A Thesis

Submitted to the Faculty

of

Drexel University

by

Yelena R. Sliozberg

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

September 2007

Dedications

This thesis is dedicated to my parents, Renald Sliozberg and Dr. Liya Lipovich, my husband Ilya Povolotskiy and my son Leonid Povolotskiy

Acknowledgments

It is a pleasure to thank the many people who have been a part of my graduate education, as teachers, colleagues and friends, and made this thesis possible.

Many thanks to my advisor, Dr. Cameron Abrams, for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences through out the work. He has excelled in training, motivation, inspiring and supporting me as a graduate student. He encouraged me to participate in conferences and write publications. He has read my numerous revisions and helped make some sense of the confusion. Above all, he has been always willing to listen and help me to understand any complicated scientific problems. Dr. Abrams is a truly scientist, he has passion in science and exceptional number of new ideas. He has inspired and enriched my growth as a researcher and a scientist.

I gratefully acknowledge the efforts of my dissertation committee members, Dr. Richard Cairncross, Dr. Nily Dan, Dr. Jeffery G. Saven and Dr. Jian-Min Yuan for their valuable suggestions and feedback. I would also like to thank all the chemical and biological engineering faculty members for being supportive and friendly.

I would like to specially thank members of our research group Dr. Ehsan Jabbarzadeh and Harish Vashisth for fruitful discussion of all research projects. Their supportive conversations have always provided me with new ideas and gave me the

endurance and patience to continue with my research. They also have gave me an emotional support for all these years.

This research would not have been possible without the financial support from National Science Foundation.

I would like to give my special thanks to all my family for their unconditional love and support which give me the strength, persistence and confidence to achieve my goals.

Table of Contents

List of Tables

List of Figures

Abstract

Molecular Simulations of Chaperonins

Yelena R. Sliozberg

Cameron F. Abrams, Ph.D.

Chaperonins are a class of cage-like molecular machines that assist the folding of polypeptides by binding and releasing non-native substrates into their inner cavity, where sequestered from the medium the substrate can fold to its native state. The main goal of this work is to understand the structure-function relationship of chaperonin. Here we study two major aspects of this relationship: (1) thermodynamics of protein folding inside the chaperonin cavity and (2) conformational changes of the *E. coli* chaperonin GroEL in its reaction cycle. We have studied the thermodynamics of protein folding, confined in the chaperonin cavity using the simple HP model of protein undergoing the coil-to-globule transition. Using the Wang-Landau method we have quantified the increase in thermal stability of a native state in a spherical confining geometry, measured as an increase in melting temperature with decreasing radius of confining sphere. We are the first to show that the $\mathbf{t} \rightarrow \mathbf{r}$ transition in GroEL subunit occurs spontaneously using unbiased molecular dynamics simulation. The transition pathway is shown to lie along low frequency quasi-harmonic modes of vibration. We are also the first to observe the spontaneous insertion of Ala480 into the empty nucleotide binding pocket, required for negative interring cooperativity.

Chapter 1: Introduction

1.1 Motivation and Research Objectives

The main goal of this work is understanding the structure-function relationship of chaperonins. In order to achieve this task we have studied the thermodynamics of protein folding inside the chaperonin central cavity and the mechanism of conformational changes of the *E coli* chaperonin GroEL through its reaction cycle.

1.1.1 Thermodynamics of Protein Folding Inside the Chaperonin Central Cavity

The first major part of this work is a study of the thermodynamics of protein folding inside the chaperonin central cavity.

The major property of the chaperonin cavity is a destabilization of the unfolded state due to confining geometry's restriction on the number of accessible "open" conformations. The consequent reduction in conformational entropy of the unfolded state shifts the equilibrium population distribution toward the native state [8, 9, 10, 11, 12]. The effect of conformational restrictions of the chaperonin cavity may smooth free energy landscape of folding for the larger substrate proteins by preventing formation kinetically stable intermediates [13, 14]. If it is true that confinement acts to reduce the conformational freedom of open states, then what is actually being probed is the effect of confinement on the density of states of the polypeptide. It is therefore desirable to show directly that confinement changes the density of states.

The main focus of this part of our work is to study the effect of mechanical confinement in a spherical cavity on the thermodynamics of a simple, "folding" bead-spring heteropolymer using the Wang-Landau MC method. We are particularly interested in how confinement can stabilize folded or partially folded states versus open states, manifest as an increase in the transition temperature. We are also interested in a possible difference of the free energy landscape of folding inside and outside of confining pore. We also are concerned with how best to adapt the Wang-Landau method for use in continuous-space simulations of single macromolecules in confinement.

1.1.2 Structural basis of the *E. coli* Chaperonin GroEL Reaction Cycle

The second major part of this work is study of the mechanism of conformational changes of the *E. coli* chaperonin GroEL through its reaction cycle.

The bacterial chaperonin GroEL is the most throughly studied chaperonin, which along with its cochaperonin GroES is involved in folding of a diverse range of substrate proteins [15, 16, 17]. Experimental evidence has shown that binding of ATP to GroEL triggers the ejection of substrate polypeptide into the chaperonin central cavity reducing affinity for non-native substrate [18, 19, 20, 21]. Although the importance of ATP binding to GroEL for substrate release is evident, a detailed picture of conformational changes introduced by binding and therefore the mechanism of the transition remains unclear.

Given both the need to predict the structure of the ATP-bound state of GroEL as well as to understand detailed transition mechanisms, molecular simulation is a powerful tool to clarify the effect of ATP binding on GroEL working cycle. In this

work we studied the effect of the presence of ATP in its binding pocket on the equilibrium structure and dynamics of GroEL chaperonin by means of atomistic molecular dynamics simulations in explicit water, comparing equilibrium structures and quasi-harmonic modes of *apo* and ATP-bound forms of the *E. coli* GroEL subunit.

1.2 Organization of Thesis

Chapter 2 includes a review of literature resources relevant to this thesis. Topics reviewed include principles of protein folding in cells, role of chaperoninins in protein folding, reaction cycle of bacterial chaperonin GroEL, folding of a substrate protein inside the chaperonin cavity and computer modeling of chaperonins.

In Chapter 3, we present the techniques that we have employed to perform simulations of chaperonins: molecular dynamics, Wang-Landau Monte Carlo and quasi-harmonic analysis.

In Chapter 4, we describe molecular dynamics simulation of phase transition simulation of generic comb-graft heteropolymer, which may be considered as a simple model of a protein folding. Understanding what topological parameters control rates of various mechanisms could be a further step toward a better general understanding of the relationship between primary and secondary structure in self-assembling biological macromolecules.

Chapter 5 investigates thermodynamics of protein folding inside the chaperonin cavity. The Wang-Landau method of direct density-of-states calculation has been used for off-lattice simulation of mechanical confinement of a substrate protein, modeled as a generic multiblock copolymer undergoing folding transition in the spherical

cavity. We propose a simple scaling relationship allowing the change in latent energy of the transition to scale independently of the change in latent entropy with the radius of the cavity resulting in a better fit to the simulation data, compared to the scaling relationships considered previously.

In Chapter 6 we study the mechanism of the reaction cycle in the *E. coli* chaperonin GroEL, using all-atom simulation of one subunit in the explicit solvent. We focus on the conformational transition of GroEL upon ATP binding. We are the first to observe conformational changes in GroEL subunit in unbiased molecular dynamics simulation.

In Chapter 7 we study the dynamic behavior of one ring of the bacterial chaperonin GroEL, using all-atom simulation in the explicit solvent. We are particularly interested in the mechanism of positive cooperativity ruling conformational changes of the chaperonin upon ATP binding.

Finally, we summarize our findings and outline research goals for the future in Chapter 8.

Chapter 2: Background

2.1 Fundamentals of Protein Folding

2.1.1 Main Principles and Dominant Forces of Protein Folding

Proteins are among the most abundant molecules within a cell and they are involved in all essential processes in living organisms. Proteins are linear heteropolymers made of amino acids. But a protein chain only becomes a functional protein when it folds into its characteristic three-dimensional structure, called a native state. Interest in the understanding of the protein folding process has raised an enormous amount of research. Several years of study of protein folding have shown that a protein in its native state possesses the lowest value of free energy. That postulates that the steric information of the correct folding of a newly synthesized protein into its native state is uniquely determined by the amino acid sequence [22].

The thermodynamic models of protein folding hold that free energy landscape of folding is dominated by competition between loss of conformational entropy and gain of attractive energy. That suggests the important role of molecular topology of a protein chain. For structurally complex proteins, there may be several possible free energy pathways for a given topology, where a protein would prefer the lowest free energy one, depending on local free energy of ordering of different parts of the protein [23].

The dominant force of protein folding is the hydrophobic force, which segregates nonpolar residues of a chain from water [24]. These interactions stabilize the otherwise weak tertiary structure of the chain and lead to the compact transition state, structurally close to the native state but more distorted. Steric constrains of this compact state are largely responsible for the internal architecture, including specific secondary structure arrangement, of the protein native state.

The hydrogen bonding and van der Waals interactions play a major role in stability of secondary structure of the native state of a protein. Solvated ion pairing or salt bridges, consisting of electrostatic interactions and some degree of hydrogen bonding, produce favorable contributions to stability, while isolated, buried salt bridges have destabilizing effect [24, 25, 26].

2.1.2 Protein Folding in Cellular Environments

Understanding protein folding *in vivo* remains a major challenge of a modern biology. In spite of the native state being the lowest thermodynamic state of a protein [22], there is kinetic partitioning between protein folding and aggregation [27]. Although folding *in vivo* and *in vitro* results in the same structure of the native state, success of the folding process under natural conditions is greatly challenged by cellular environment. In these natural cytoplasmic environments, proteins, protein complexes, lipid bilayer structures, and more, both crowd and confine the newly transcribed or thermally denaturated polypeptide.

According to theory these "crowding" and "confining" interactions play important roles in cell biology. The resulting excluded volume effect enhances all association

processes causing a protein chain to adopt the compact form: desirable self-assembly as well as undesirable aggregation, which may sufficiently reduce productivity of the protein folding and lead to accumulation of toxic aggregates [28, 9, 10, 29].

Another possible danger for a non-native protein with a complex domain structure is to become kinetically trapped in a local free energy minimum forming a physiologically non-functional state. These misfolded protein structures could be relatively stable and not achieve their native state on biologically relevant timescales [30].

To prevent aggregation and keep proteins out of kinetic traps, a specialized class of proteins, molecular chaperones, have evolved to help proteins to fold correctly in intracellular conditions [31]. In intracellular environments, newly synthesized proteins do not self-assemble spontaneously, but are assisted by transferring among a network of chaperones sequentially folding to their native state [29].

2.1.3 Roles of Chaperones and Chaperonins in Protein Folding

Chaperones are divided into two major classes based on the mechanism of the substrate folding: chaperones such as Hsp70s and trigger factor and large cylindrical chaperonins have been identified as protein helpers in the cytosol of eukaryotic and prokaryotic cells and also inside mitochondria and chloroplasts. Most chaperones use ATP binding and hydrolysis to regulate their reaction cycle. Small chaperones prevent aggregation of non-native proteins through multiple cycles of their binding and release, regulated by chaperone ATPase activity, which could be sufficient to achieve correct folding of some substrates, but could require help of the large chaperonin system for the other substrates [32, 33, 14].

The most interesting subclass of chaperones composed of the large double-ring structures are called the chaperonins. Chaperonins bind and release non-native polypeptide substrates into an inner cavity, where sequestered from the medium a substrate can fold to its native state [34, 35, 33]. Folding inside the chaperonin cavity is believed to be favorable because of three properties: solitary confinement removes danger of aggregation, hydrophilicity of the cavity lining leads to burial of hydrophic and exposure of hydrophilic surfaces of protein, and confinement reduces number of accessible non-folded states of the protein including kinetically trapped [31, 9, 10, 13].

Recent research has shown the great importance of chaperonins proper functioning in preventing human diseases. Mutations in the human chaperonins may cause some inherited diseases, such as McKusick-Kaufman syndrome and Bardet-Bieldi syndrome [36, 37, 38]. Several studies of bacterial chaperonin prototype have demonstrated that chaperonins may be modifiers of missense mutations causing metabolic diseases [38].

Considering the essential role of chaperonins in protein folding in cellular environment, the detailed understanding of mechanisms of chaperonin machinery will potentially impact on the biotechnology and medicine.

2.2 GroEL-GroES Chaperonin System

2.2.1 Architecture of GroEL

Substantial amount of research has been performed on the *E. coli* chaperonin GroEL, that assists folding of a diverse range of substrate proteins [39, 40, 41, 30, 16, 17]. The *E. coli* chaperonin GroEL is a cage-like molecular machine that assists the folding of non-native polypeptides in an ATP-dependent manner. GroEL is a large protein complex, which in its unliganded state has a cylindrical form 135 Å in diameter and

145 Å in height. This cylinder contains the central cavity large enough to accommodate a non-native protein of ∼50-60 kDa and composed of two rings of seven identical subunits of 57 kDa per ring . The cochaperonin GroES consists of a single ring with seven subunit of 10 kDa and possesses a small central cavity. GroES binds to and unbinds from GroEL during the chaperonin reaction cycle [42, 43, 44, 16, 17, 40].

A single GroEL ring exists in one of two states: a binding-active state and a folding-active state [31, 45]. In its binding-active state GroEL exposes its hydrophobic apical cavity to bind a substrate protein, rescuing it from the danger of aggregation in the medium. In the presence of ATP a protein-bound ring of GroEL binds cochaperonin GroES and undergoes major conformational changes, forming a folding-active complex containing an encapsulated chamber with hydrophilic walls, where the substrate protein is released to fold in isolation (Fig 2.1(a)).

Each subunit of GroEL may be divided at three distinctive domains : well-ordered, providing most inter-subunit, including all inter-ring, contacts equatorial domain, less ordered apical domain and relatively small intermediate domain connects the apical and equatorial domains [44, 39]. A protein substrate and cochaperonin GroES bind to the apical domain, which is able to adjust its conformation to accommodate various protein substrates [46, 47, 48, 49], whereas the equatorial domain contains the nucleotide binding site. The sequence of GroEL subunit with secondary structure elements is shown in Fig. 2.1(b).

2.2.2 Reaction Cycle

A single ring can bind seven molecules of ATP and at least one substrate polypeptide. GroEL functions by switching heptamer ring ligand affinities in an alternating manner between rings. Each GroEL ring is in equilibrium between two states: a tense ($\mathbf{t}_7$ or $\mathbf{T}$) state with low affinity for ATP and high affinity for the non-native substrate protein and a relaxed ($\mathbf{r}_7$ or $\mathbf{R}$) state with high affinity for ATP and low affinity for the non-native substrate. In the presence of ATP the GroEL double ring undergoes allosteric transition from the **TT** states (both rings in the **T** state) to the **TR** state (as result of the negative cooperativity between the rings, the adjoining ring simultaneously undergoes a transition in the opposite direction) [50, 18, 19, 51]. A ring can bind a nascent or misfolded substrate protein in the **T** state [52] and eject it by transitioning to the **R** state. [20, 21] Binding of GroES transfers the **R**-state ring to the $\mathbf{R}''$ state, forming a cage [2] into which polypeptide is ejected and left to fold without potentially aggregation-inducing interactions [53]. The encapsulated protein is contained for a time determined by ATP hydrolysis in this newly formed *cis* assembly and discharges with GroES into medium after binding of ATP to the unliganded *trans* ring [39, 6]. A schematic representation of GroEL-GroES-protein working cycle appears in Fig. 2.2.

Although a complete understanding of the structural basis of GroEL function remains unresolved, it necessarily involves large relative motion of three distinct domains. Binding of ATP and cochaperonin GroES to a ring of GroEL releases the substrate protein from its apical binding site into the cavity and causes major conformational changes in this ring: the intermediate domain of the each subunit rotates downward 25° along its pivot point connection to the equatorial domain and the apical domain undergoes an elevation of 60° in reference to the equatorial domain and

clockwise rotation outward on its own axis by 90°.

2.2.3 Importance of ATP Binding. R-State

Experimental evidences show that binding of ATP alone to GroEL triggers the ejection of substrate polypeptide into the chaperonin central cavity on account of the observation that the ATP-bound ring of GroEL has a lower affinity for non-native substrate than does ATP-free ring of GroEL [18, 19, 20, 21]. Although the importance of ATP binding to GroEL for substrate release is evident, a detailed picture of conformational changes introduced by this event and therefore the mechanism of **T**→**R** transition and structure of **R**-state remain unclear.

No crystal structure of an **R**-state ring exists because of crystal lattice prevention of conformational changes leading to formation of **R** state [54]. Asymmetric **TR** tetradecamers of hydrolysis-defective mutants have been visualized using cryoelectron microscopy [21]. According to these results, in transitioning from **T** to **R**, the intermediate domain closes down over the nucleotide binding pocket rotating downward by $\sim 20^\circ$, accompanied by a small counterclockwise rotation of the apical domain ($\sim 25^\circ$) in the plane of the ring [21]. The mechanism of this transition may be explained as a result of a normal range of the thermal motion of GroEL domains, where maximum extent of this motion is stabilized by interaction between the intermediate domain residues and the bound nucleotide [40].

T to **R** transition caused by ATP binding leads to breaking of Arg197-Glu386 salt bridges between the apical and the intermediate domains of neighboring subunits of the same ring and then formation of the new intersubunit salt bridges between

Glu386 and Lys80 of the equatorial domain. This salt bridge switching may play important role in positive cooperativity inside GroEL ring [21].

2.2.4 Positive and Negative Cooperativity

Both positive intra-ring and negative inter-ring cooperativity with respect to binding of ATP play central roles in the reaction cycle of GroEL and drives $\mathbf{t}_7$:$\mathbf{r}_7 \rightarrow \mathbf{r}_7$:$\mathbf{t}_7$. The positive intra-ring cooperativity in ATP binding of GroEL described by an equilibrium between two states: **T** state with the low affinity to ATP and **R** state with high affinity to ATP, where ATP induced conformational changes are concerted. The negative inter-ring cooperativity is believed to be sequential [50, 18, 19, 51].

Switching of intersubunit salt bridges [21, 1, 40] has been invoked to explain positive intra-ring and negative inter-ring cooperativity in ATP binding. Mediating the necessary communication of binding pocket occupancy from one subunit to its intra- and inter-ring neighbors must involve communication between intersubunit salt bridges, because mutants lacking these bridges display reduced or absent cooperativity [50, 40]. In particular, Glu386 links to Arg197, joining a subnunit's intermediate domain to the apical domain of its counterclockwise neighbor in the ring (viewed from above), and the Arg197 mutant displays no positive cooperativity [50]. Negative allosteric signaling is believed to be result of communication of two inter-ring salt bridges defined by pairing of Arg452 and Glu461 [21, 1]. In order to clarify the GroEL structure-fuction relashionship, one must understand the mechanisms by which salt bridges communicate and manage the binding pockets occupancy among the neighboring subunits. This task requires understanding the dynamics involved in

transitioning among the various allosteric states, since the proper transition of the allosteric signals is the imperative condition for the GroEL reaction cycle.

The concerted nature of **T**→**R** transition has been believed to be a result of the steric clashes arising when one subunit undergoes **t**→**r** transition when either neighbor of the same ring remains in **t** state [55, 56]. Concertedness among subunits, explained by Ma et al. as a result of steric interactions between neighbor subunits in **r** and **t** [56] have not been experimentally confirmed. Shiseki et al. [57] results have shown that GroEL mutant having two cross-linked subunits (the cross-linked subunts lack possibility of **t**→**r** transition) possesses ATPase activity. This important because that incomplete $\mathbf{t}_7 \rightarrow \mathbf{r}_7$ transition in the case of two cross-linked subunits disrupts the negative cooperativity between the rings [57].

2.2.5 Computational Studies

For all the information extracted from the previous studies some questions still remain unresolved, among them the structural basis of the **R** state as well the detailed mechanism of **t**→**r** transition. Computer simulation offers the best chance to probe the details of those processes.

The first detailed molecular model of the dynamics of the $\mathbf{t} \rightarrow \mathbf{r}''$ transition, using normal mode analysis was put forth by Ma et al [55]. The same authors continued to study the transition with targeted molecular dynamics, from which they explained that concertedness among subunits in the transition is a result of steric interactions. They also indicated that the **r** state has an apical domain rotated clockwise rather than counterclockwise [55, 56]. Significantly, their results are in disagreement with

the cryo-EM structures of **R**-state rings [21] with regard to the direction of rotation of the apical domains upon ATP binding. The explanation of this divergence is the following: Ma et al. simulated the entire **t**→**r**″ transition using targeted molecular dynamics (TMD) [58], in which fictitious forces on all C_α's drove the structure from **t** to **r**″ in under one nanosecond, in effect producing a non-dynamic pathway that interpolates between the endpoints. Because the net in-plane rotation of the apical domain is about 90° *clockwise* based on comparison of *apo* and ADP-bullet crystal structures [2], an initial 20° counterclockwise rotation of the apical domain in **t**→**r** does not appear along a TMD pathway.

In a much more recent and larger-scale coarse-grained molecular modeling study of the entire **T**→**R**″ transition, counterclockwise apical domain rotation in the **t**→**r** transition was enforced by using the cryo-EM structure as an intermediate target [59], and, as a result, the role of static interactions in concertedness was reduced. Both of these studies involved forcing the subunits to execute their allosteric transitions from a starting structure to a target structure in computationally feasible times using fictitious driving forces that bias motion toward the target. Although large-scale, detailed, targeted dynamic simulations are able to elucidate plausible transition mechanisms, because of the fictitious nature of the driving forces, it is in principle impossible to gauge their accuracy without reference to unbiased simulations, which have been avoided primarily because of computational expense.

2.3 Protein Folding inside the Chaperonin Cavity

2.3.1 Role of Molecular Confinement on Protein Thermodynamics

Generically, forcing any polypeptide into a confined space must affect its thermodynamic behavior as compared to its behavior in dilute solution. Quantifying the degree to which this generic effect is important for any particular protein and environment may one day prove crucial in engineering protein production pathways. In recent years, there has been a significant amount of research aiming to understand how confinement in general affects protein thermodynamics.

Experimental results show that confinement enhances the thermal stability of proteins in the native state. Eggers and Valentine observed a 25-30°C increase in the native-to-coil melting temperature of α-lactalbumin using a sol-gel entrapment system [60]. Bollis et al. observed a similar stabilization effect for four single domain proteins: titin 127 and bacterial, yeast and human frataxins. The largest increase in melting temperature, about 5°C in polyacrylamide gel, was observed for the least stable of the examined proteins, yeast frataxin [61]. Particular attention has been paid to the *E. coli* chaperonin complex GroEL/GroES, which is a main focus of this work [39, 41, 30]. Brinker and coworkers, using a rapid biotin-streptavidin-based inhibition of chaperonin function, showed that mere confinement of a range of synthetic polypeptides in the GroEL cage significantly enhances folding throughput comparing with spontaneous, unassisted folding [13, 30].

Partial explanations of this behavior from theory hold that the primary mechanism is a destabilization of the unfolded state due to confining geometry's restriction on the number of accessible "open" conformations. The consequent reduction in conformational entropy of the unfolded state shifts the equilibrium population distribution toward the native state. The effect of conformational restrictions may

also smooth the free energy landscape of folding for the larger substrate proteins by preventing formation kinetically stable intermediates [13, 14, 31, 62]. Using a statistical-thermodynamic model, Minton estimates the effect of excluded volume due to macromolecular crowding or confining media as an increase of unfolding temperature by 5-20°C [9, 10]. Similarly, Zhou and Dill showed that mere mechanical confinement in a small inert cage would stabilize a protein against reversible unfolding by excluding some expanded conformations of the unfolded chain [11]. They predicted an increase in the melting temperature of 31°C for α-lactalbumin, in agreement with measurements of Eggers and Valentine.

It has to be mentioned that considering the chaperonin inner cavity as a simple isolation device preventing aggregation and limiting size of the protein chain would be an oversimplification. Experiments have showed that folding rate of substrate proteins depend not only on relative size of the chain, but also of inner charge of the chaperonin cavity. Negatively charged inner surface of the cage repels the folding change because most of proteins folding with GroEL help have net negative charge [63, 64].

2.3.2 Computational Studies

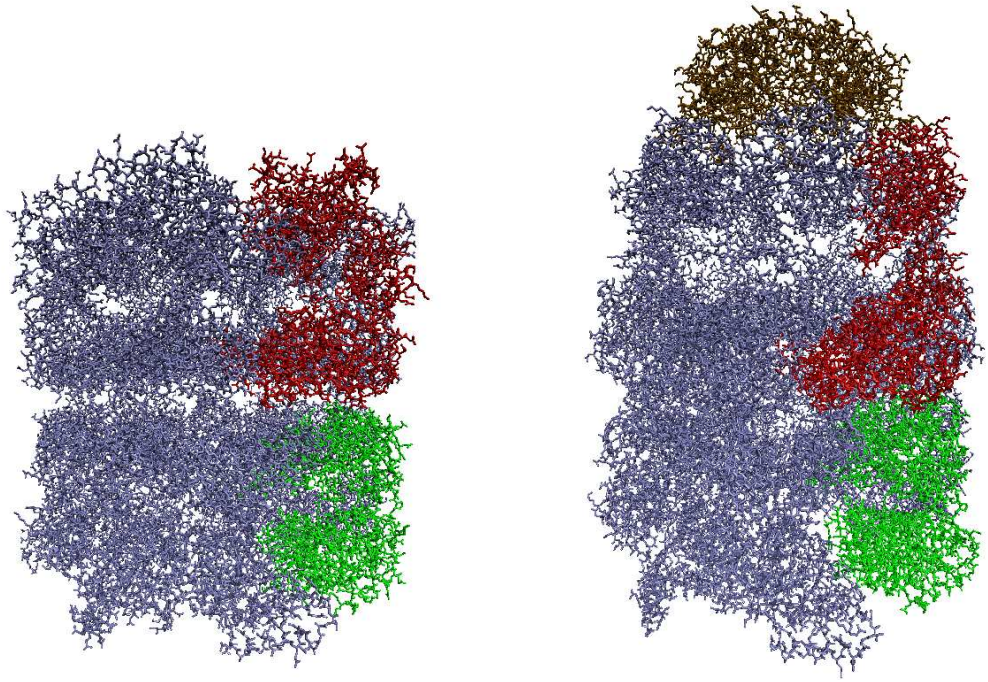
Several studies using direct simulation have corroborated and expanded the theoretical picture of confinement's effect on protein stability. Klimov et al. studied the effect of confinement in an inert spherical cavity on thermodynamic behavior of an off-lattice coarse-grained model protein using Langevin dynamics [65]. They observed that even "modest" confinement increases the stability of native states.

Takagi et al., also using Langevin dynamics, expanded on the work of Klimov et al. by considering several coarse-grained model proteins in inert cylindrical cages. They found that a shift in transition temperature caused by confinement in a cage of size L scales as $\Delta T_f \sim (R_0/L)^\nu$, where $\nu \approx 3$ and R_0 is the characteristic size of the polypeptide [66]. Though both of these authors claim that their results show that confinement enhances the rate of folding, evidence supporting such a claim is obscured by the non-conservative nature of the temperature control scheme used. Ping et al. performed standard Boltzmann Monte Carlo of HP lattice proteins in order to investigate two kinds of confinement: inside a rigid box and tethered to a solid surface [67]. They found that confinement indeed stabilizes folded conformations relative to unfolded, and they found that melting temperatures increase with decreasing confinement dimension. These results agreed with their experimental observation that thermal stability of horseradish peroxidase and acid phosphatase, encapsulated in mesoporous silica, increases as the mean pore size of the silica matrix decreases.

It should be appreciated that the simulation results described above are synthesized from many individual simulations at various temperatures to reconstruct the full thermodynamic behavior over a temperature range of interest. If it is true that confinement acts to reduce the conformational freedom of open states, then what is actually being probed is the effect of confinement on the *density of states* of the polypeptide, $g(U)$, where U is potential energy, which is itself *independent* of temperature. If $g(U)$ is known, thermodynamic quantities can be calculated for any temperature. It is therefore desirable to show directly that confinement changes the density of states. We will consider density of states in detail in Chapter "Effect of

Confinement on the Thermodynamics of a Collapsing Heteropolymer"

(a)



(b)

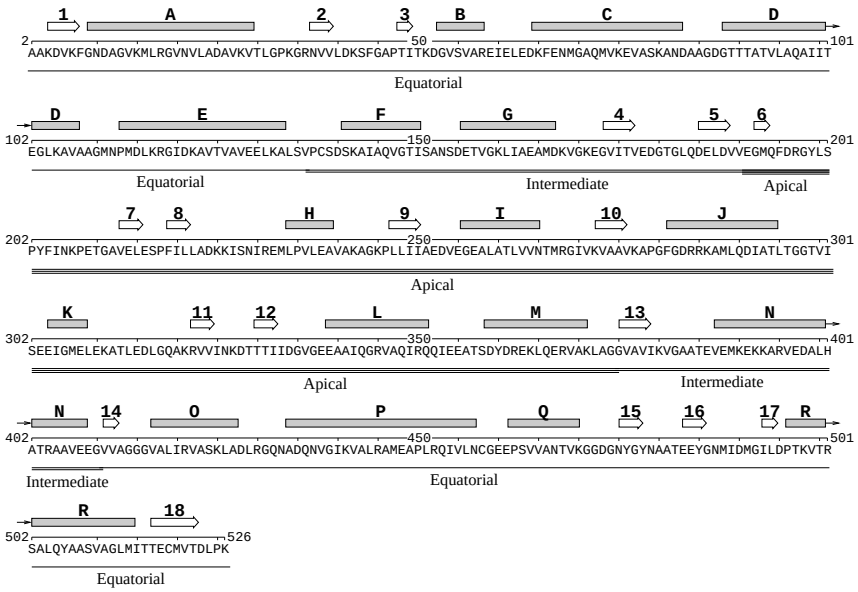


Figure 2.1: Architecture of GroEL chaperonin. (a) Binding and folding active states of GroEL. All 14 subunits are identical in the binding-active state. ATP and GroES binding to one ring of GroEL transforms it structural arrangement, forming assymetrical "bullet" [2]. Red and green colors represent individual subunits from the upper and lower ring. For this figure we used crystallographic data of Wang and Boisvert [PDB code 1KP8 [3]] and Xu, Horwich and Sigler [PDB code 1AON [2]].This figure and others with protein renderings produced using VMD [4, 5]. (b) The sequence of GroEL subunit shown with secondary and tertuary structure elements. Secondary structure elements are shown by rectangles (α-helices) or arrows (β-strands)

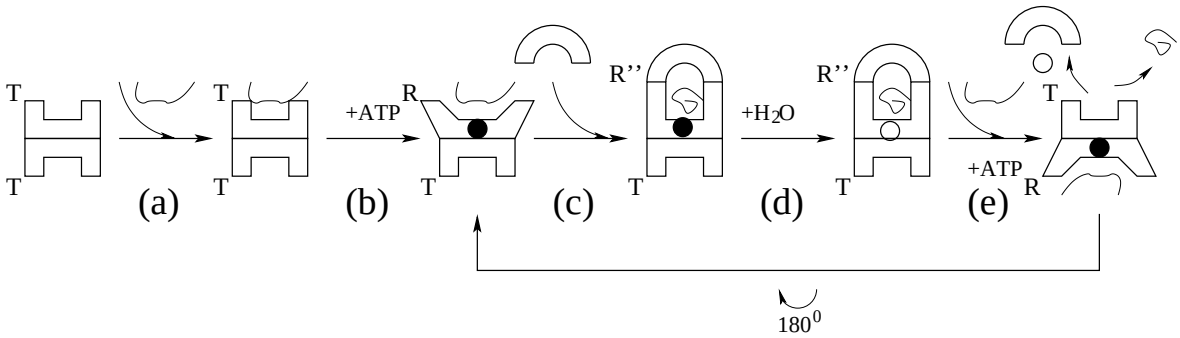


Figure 2.2: Schematic representation of GroEL-GroES-protein reaction cycle. (a) GroEL is in its binding-active state binds the substrate protein. (b) **TT**→**TR** transition: Binding of ATP to the one ring reduces the affinity to substrate for the same ring and reduces the affinity to ATP for the opposite ring. (c) **TR**→**TR**″ transition: Binding of cochaperonin GroES forms folding chamber for substrate (folding-active state). (d) Protein folds inside the cavity for a time determined by ATP hydrolysis. ATP hydrolysis into ADP primes GroES and substrate to release into solution [6]. (e) **TR**″→**RT** transition: ATP binding to the **T** ring makes transit into the **R** ring and leads to **R**″→**T** transition for the opposite ring and therefore GroES and the substrate are released into solution. Additional binding of the substrate protein facilitates the release [7]

Chapter 3: Models and Methods

3.1 Molecular Simulations of Biomolecules

3.1.1 Introduction

Computer simulation is a powerful approach to extract dynamical and statistical information from molecular systems. Simulation techniques are not only an additional tool for experimentalists to predict and interpret characteristic behavior of the different biosystems, but are capable of explaining the mechanism of many processes better than experimental methods, because they allow one to detect and describe important metastable structures inaccessible with experimental techniques.

The most important element of a molecular simulation is the often complicated model that computes the total potential energy, U, as a function of all atom Cartesian positions. Since many of the problems that one would like to address in biological systems involve many atoms, it is not yet feasible to treat these systems using quantum mechanics. The solution of this problem is use of empirical potential energy functions, which are much less computationally demanding than quantum mechanics. We will discuss this approach in detail in Section "Models"

There are two major approaches to perform molecular simulations: stochastic and deterministic. The stochastic approach, called Monte Carlo, is based on exploring the energy surface randomly probing the conformational space of the molecular system. As a result, Monte Carlo simulation accumulates data which is used for calculations

of different thermodynamic properties of the system. The deterministic approach, called Molecular Dynamics, simulates the time evolution of a molecular system and can be used to extract the kinetic and thermodynamic properties of the system.

3.1.2 Thermodynamic Properties of Molecular Systems

Molecular simulations generate information at the microscopic level. Observable or macroscopic properties of the molecular system are calculated from statistical averages over a large number of individual system snapshots, called ensemble averages. The statistical ensemble is defined as a collection of all possible microscopic systems with different microscopic states but having an identical macroscopic state.

There exist different kinds of statistical ensembles characterized by fixed values of thermodynamic variables such as particle number, N, volume, V, temperature, T, energy, E or chemical potential, μ. The Microcanonical Ensemble is characterized by fixed values of particle number, volume and energy and denoted as NVE. The Canonical Ensemble or NVT has constant values of particle number, volume and temperature. Other examples of statistical ensembles are Isobaric-Isothermal Ensemble (NPT) and Grand Canonical Ensemble (μVT).

The observable property of the molecular system is calculated as ensemble average $\langle A \rangle_{ensemble}$ and defined by

$$\langle A \rangle_{ensemble} = \int\int dpdr A\left(p, r\right) P\left(p, r\right), \tag{3.1}$$

where p and r are generalized momenta and internal coordinates of the system and P is the probability density of observing a system in microstate (p, r) and is given in

the Canonical ensemble by

$$P(p,r) = \frac{1}{Z} \exp\left[-H(p,r)/k_B T\right], \tag{3.2}$$

where H is the Hamiltonian of the system: $H = K + U$, where K is kinetic energy and U is potential energy of the system. Z is partition function and defined as

$$Z = \int\int dpdr \exp\left[-H(p,r)/k_B T\right], \tag{3.3}$$

Knowing partition function for the system of constant temperature and volume (NVT) the thermodynamic properties can be calculated, for example

$$A = -k_B T \log Z, \tag{3.4}$$

where A is Helmholtz free energy.

3.2 Methods

In this section we present the techniques we have used in our studies of the chaperonins action in protein folding, these included molecular dynamics, Wang-Landau Monte Carlo and quasi-harmonic analysis.

3.2.1 Molecular Dynamics Simulation

Molecular dynamics (MD) simulation is a one of the major computational tools used to study structure and dynamics of biological systems. MD simulations can lead to detailed understanding of the mechanisms of the basic biological processes as a result

of the inner motions of biomolecules.

MD calculates the time dependent behavior of a system of discrete particles under known laws of physics. The motion of particles obeys classical Newtonian equations, which adequately describe a wide range of systems, including biomolecules [68].

For particle i

$$\vec{F_i} = m\vec{a_i} \tag{3.5}$$

Interactions between particles of the system are described by the potential energy, where the potential energy is a function of the coordinates of all particles. In the case of biomolecule simulations, the information about the potential energy is typically encoded in empirical potential functions.

The force acting on particle i is the negative gradient of the total potential energy of the system.

$$\vec{F_i} = -\nabla_{\vec{r_i}} U\left(\vec{r^N}\right) \tag{3.6}$$

Due to the complicated nature of potential energy, Newton's equations of motions are solved numerically. From many possible integrated schemes, the velocity Verlet algorithm is the most often used. This algorithm is relatively fast, accurate and time reversible [69]. Update of coordinates and velocities is calculated as

$$r\left(t + \Delta t\right) = r\left(t\right) + v\left(t\right) + \frac{f\left(t\right)}{2m}\Delta t^2 \tag{3.7}$$

$$v\left(t + \Delta t\right) = v\left(t\right) + \frac{f\left(t + \Delta t\right) + f\left(t\right)}{2m}\Delta t \tag{3.8}$$

The initial conditions for MD simulation are a set of particle positions and ve-

locities. The velocities are assigned randomly for each particle from the Maxwell distribution for the desirable temperature, where probability, P over the velocity, v range $v \in [0, \infty)$ is given by

$$P(v) = \sqrt{\frac{2}{\pi}} \frac{v^2 \exp^{-v^2/2a^2}}{a^3}, \tag{3.9}$$

defining $a = \sqrt{k_B T/m}$, where m is a mass of a particle

The instantaneous absolute temperature, T and particle velocities are related through the mean kinetic energy of the system:

$$T = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f}, \tag{3.10}$$

where v_i is velocity of an atom i and $N_f = 3N - 6$ is the number of degrees of freedom for the three-dimensional system of N particles.

The limitations of MD are the approximate functions for the potential energy and length of simulations [68]. Being approximate, MD potentials are in the user's control and could be changed toward more accurate ones, which would lead to extension of the simulation time. The length of simulation depends on two quantities: the physiological time of the simulated process and the size and detailed description of the studied system. For speed up the simulation introducing methods based on MD trajectories as free energy calculations, quasi-harmonic analysis, targeted molecular dynamics (TMD) may be employed. The coarse-graining the system or using generic potential functions also reduce the computer time of simulation.

The new generation of supercomputers using parallel molecular dynamics helps to extend the simulation for the large biosystems allowing us to calculate processes in nanosecond or even microsecond timescale. The current record is the all-atom simulation of the ribosome, containing approximately 2.64 millions atoms for 4 ns [70].

MD could be used as a trial move within an otherwise standard Boltzmann Monte Carlo simulation has acquired the name "hybrid Monte Carlo." [71] or incorporated in the Wang-Landau Monte Carlo simulation [72].

We have used MD as a primary tool for the study of role of chaperonins action in protein folding. In order to achieve this task we have performed MD simulation of model coarse-grained generic comb-graft heteropolymer to understand generally the relationship between primary and secondary structure in self-assembling macromolecules. The second step was detailed all-atom MD simulation of bacterial chaperonin GroEL to investigate mechanism of conformational changes of the chaperonin through its reaction cycle. We have also incorporated short NVE molecular dynamics trajectories in our library of trial moves for Wang-Landau simulation of confining collapsing heteropolymer.

Langevin and Dissipative Particle Dynamics Thermostats

Equation 3.6 describes NVE ensemble, where a microcanonical dynamics trajectory may be seen as an exchange of potential and kinetic energy, with the total energy being conserved (the Hamiltonian, H of the system: $H = K + U = const$). For canonical or NVT ensemble the energy is in exchange with a thermostat. A variety of thermostat methods are required to add or remove energy from the system. In our work we have used two popular techniques to control temperature include Langevin

and Dissipative Particle Dynamics (DPD) thermostats.

For the Langevin thermostat [73], each particle of the system is coupled with a heat bath and equation of motion becomes:

$$m\ddot{\mathbf{r}} = -\nabla_i U - m\Gamma\dot{\mathbf{r}} + W_i(t), \tag{3.11}$$

where Γ is a constant friction, $W_i(t)$ is a random force (Gaussian white noise), received by particle i and given by

$$\langle W_i(t) \cdot W_j(t') \rangle = \delta_{ij}\delta(t - t')\, 6k_B T\Gamma, \tag{3.12}$$

The advantage of using the Langevin thermostat is that the larger time step permitted and therefore simulations of large and detailed systems, such as biomolecules, are facilitated. The disadvantages are the lack of momentum conservation and inaccurate representation of thermodynamics. The Dissipative Particle Dynamics (DPD) thermostat avoids these disadvantages and keeps the large simulation step [74].

For the DPD thermostat the equation of motion is given by

$$\dot{\vec{p_i}} = \vec{F_i} + \vec{F_i^D} + \vec{F_i^R}, \tag{3.13}$$

where $\vec{p_i}$ is momentum of particle i, $\vec{F_i^D}$ and $\vec{F_i^R}$ are dissipative and random forces respectively. The dissipative force and the random force are pairwise and given by

$$\vec{F_{ij}^D} = -\zeta\omega^D(r_{ij})(\hat{r}_{ij} \cdot \vec{v}_{ij})\,\hat{r}_{ij} \tag{3.14}$$

$$\vec{F_{ij}^{R}} = -\sigma\omega^{R}\left(r_{ij}\right)\theta_{ij}\hat{r}_{ij}, \tag{3.15}$$

where $\vec{v_{ij}} = \vec{v_i} - \vec{v_j}$ is the relative velocity between particles i and j, $\hat{r}_{ij}$ is the unit vector of the interatomic axis $\vec{r_{ij}} = \vec{r_i} - \vec{r_j}$, ζ is the friction constant and σ is the noise strength,

ω^D and ω^R are r-dependent weight functions, defined as

$$\omega^{D}\left(r\right) = \left[\omega^{R}\left(r\right)\right]^{2}, \tag{3.16}$$

where $\omega\left(r\right) = \begin{cases} \left(r_c - r\right)^2 & \left(r < r_c\right) \\ 0 & \left(r \geq r_c\right) \end{cases}$ and r_c is cutoff distance.

3.2.2 Wang-Landau Monte Carlo Simulation of Off-Lattice Chain

One of the most important quantities in the statistical physics is density of states, $g(U)$, defined as number of all possible states of the system for a given energy level U. All thermodynamic quantities can be calculated for any temperature from $g(U)$.

$$Z\left(N,V,T\right) = \sum_{U} g\left(N,V,U\right)\exp^{-\beta U}, \tag{3.17}$$

where Z is a partition function and N, V, U, T are number of particles, volume, potential energy and temperature of the system respectively and $\beta = \frac{1}{k_B T}$.

A relatively new direct simulation method of Monte Carlo simulation, developed by Wang and Landau [75], allows calculation of $g(U)$ in a single simulation and could be used for a simple, off-lattice polypeptide in dilute solution conditions [72, 76].

The Wang and Landau method computes $g(U)$ in an iterative fashion carrying out a random walk in energy space. This algorithm is designed to produce the flat energy histogram, based on the observation that the probability of visiting a given energy level U is inversely proportional the density of states [75].

$$P(U) = \frac{1}{g(U)} \tag{3.18}$$

The following is a brief description of the Wang-Landau algorithm. A finite domain of potential energy, $U_{\text{low}} < U < U_{\text{high}}$, divided into a finite number of bins of uniform width ΔU, is considered. In the course of the simulation, it is nessesary to keep track of two histograms: (a) the number of times each bin is visited, denoted $H(U)$, and (b) the density of states value of each bin, stored as $\ln g(U)$. To begin a simulation, all bins are initialized such that $H(U) = 0$ and $\ln g(U) = 0$. Then a random walk in energy space is initiated by attempting trial moves that change the system potential energy. A move taking the system from bin U_1 to U_2 is accepted according to the probability

$$p_{\text{acc}}(U_1 \to U_2) = \min\left[\frac{g(U_1)}{g(U_2)}, 1\right] \tag{3.19}$$

A move is rejected outright if $U_2 < U_{\text{low}}$ or $U_2 > U_{\text{high}}$. After a move regardless of whether it is accepted or rejected, then the histograms are updated according to $H(U) \leftarrow H(U) + 1$ and $\ln g \leftarrow \ln g + \ln f$, This random walk simulation is continued until the histogram $H(U)$ is "sufficiently flat." At this point, H is re-zeroed, $\ln g$ is maintained, and f is reduced according to $f \leftarrow \sqrt{f}$. Iterations in this fashion are chained together until $\ln f < 10^{-8}$, at which time $g(U)$ is taken to be the final density

of states.

It is also possible to use short NVE molecular dynamics trajectories in the library of trial moves for Wang-Landau simulation. The use of NVE MD as a trial move within an otherwise standard Boltzmann Monte Carlo simulation has acquired the name "hybrid Monte Carlo." [71] and has been employed by Rathore and de Pablo, [72] for Wang-Landau simulation of a small protein.

The energy distribution function at a particular temperature, $P(U)$, can be calculated by

$$P\left(U_i\right) \equiv P_i = \frac{g\left(U_i\right) \exp^{-\beta U_i}}{\sum_k g\left(U_k\right) \exp^{-\beta U_k}}, \tag{3.20}$$

where the summation is over all energy bins. We can calculate thermodynamic quantities such as free energy, internal energy, entropy and specific heat:

$$\langle U\left(T\right)\rangle = \sum_i U_i P_i \tag{3.21}$$

$$C\left(T\right) = \frac{\partial \langle U\left(T\right)\rangle}{\partial T} = \frac{\langle U^2\rangle_T - \langle U\rangle_T^2}{T^2} \tag{3.22}$$

$$F = -k_B T \log Z \tag{3.23}$$

$$S\left(T\right) = \frac{U\left(T\right) - F\left(T\right)}{T} \tag{3.24}$$

Some of these quantities, such as entropy and free energy, are not directly accessible from conventional Monte-Carlo simulations.

It should be mentioned that, without a reference state for which $g(U)$ is known independently, this algorithm only provides $g(U)$ to within an undetermined multi-

plicative constant. This constant bears no relevance in the computed thermodynamic quantities listed above, including the melting temperature, because it divides out of the probability distribution. It is, however, not strictly allowed that we quantitatively compare absolute $g(U)$'s from the simulations among the various runs.

The Wang-Landau simulation is preferential to use for studying systems undergoing a phase transition, such as folding of protein chain, because conventional Monte-Carlo requires long time for a system to travel from one phase to another. Because Wang-Landau algorithm allows the system to visit all energy level with equal probability the tunneling barrier between coexisting phases of conventional Monte-Carlo is avoided [75].

We have employed the Wang-Landau method to study thermodynamics of protein folding in the chaperonin center cavity, as will be presented in Chapter 5.

3.2.3 Quasi-Harmonic Analysis of Conformational Changes of Proteins

Multidomain proteins undergo conformational changes via large-amplitude collective motion of a great number of atoms. This kinds of motion often happens on timescales on the order of microseconds, which are unaccessible by straightforward molecular dynamics simulation. The harmonic approximation of protein dynamics, called normal mode analysis, models atoms as point masses connected by springs and may be used for calculating protein flexibility using much less computer time than conventional MD [77, 78]. It has been shown that the highly collective motion of a protein during its conformational change is well described by a single low-frequency normal mode [79].

Being a fast and simple method, the harmonic approximation of proteins dynamics is unable to describe accurately the flexibility of a protein above 100 K and therefore in order to describe low-frequency normal mode corresponding to the protein conformational change anharmonic character of the protein motion should be incorporated in the applying force field [80, 81]. The quasi-harmonic approximation uses anharmonic MD trajectories to construct the force field and normal mode eigenvectors and frequencies are evaluated [81].

In the quasi-harmonic approximation potential energy function U is expressed in a form of quadratic potential:

$$U = \frac{1}{2} \sum_i K_{ij} \left(q^i - {q^i}_0\right) \left(q^j - {q^j}_0\right), \tag{3.25}$$

where $(q^i - {q^i}_0)$ is the displacement of the i th internal coordinate from its equilibrium value and K_{ij} is a force constant.

Considering only a diagonal force constant ${K_q}^i$ one may calculate it from

$${K_q}^i = \frac{k_B T}{\left\langle \left(q^i - {q^i}_0\right)^2 \right\rangle}, \tag{3.26}$$

where k_B is Boltzmann's constant, T is the absolute temperature, and $\left\langle \left(q^i - {q^i}_0\right)^2 \right\rangle$ is the second moment of the virtual internal coordinate fluctuations calculate from the time average of MD trajectories:

$$\left\langle \left(q^i - {q^i}_0\right)^2 \right\rangle = \frac{1}{M} \sum_{t=1}^{M} \left({q^i}_t - {q^i}_0\right)^2, \tag{3.27}$$

where M is number of selected samples in the MD trajectory. Thus anharmonic effects are incorporated in the harmonic force field.

The normal modes are obtained from

$$|F - \omega M| = 0, \tag{3.28}$$

where $F_{ij} = d^2U/dx_i dx_j$ is a matrix of the second derivatives of the potential energy and M is the mass weighted kinetic energy matrix. Obtained effective modes can be used for calculations of root mean-square fluctuations and displacements along mode directions.

Quasi-harmonic analysis is a relatively accurate method to describe conformational transitions of the proteins and according to Tama et al [79] yields a single low-frequency normal mode corresponding the collective motion of the conformational change. We have employed quasi-harmonic analysis for the study of domain motion of bacterial chaperonin GroEL during its conformational transitions and identified the modes driving the transitions.

3.3 Models

3.3.1 Atomistic Representation of the Biomolecular System

A detailed description of the structural changes occurring during dynamical changes in the conformations of biomolecules requires taking into account interactions among all atoms of the biological system. Emperical potentials provide all necessary information for interactions of protein and solvent atoms.

The energy functions used for proteins are generally composed of the bonding

component such as bond length, bond angles and torsional angles and the non-bonding component, which treats van de Waals and electrostatic interactions. At normal temperatures, bond lengths and angles in proteins are close to their equilibrium values and therefore the potential energy of bonding is approximated as harmonic [68]. These empirical potential functions and set of parameters are called the "force field". One of the most widely employed force field for simulating biomolecules is CHARMM [82, 83]. The potential energy function used in CHARMM is

$$\begin{aligned} U(r) &= \sum_{bonds} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{angle} K_\theta (\theta - \theta_0)^2 \\ &+ \sum_{dihedrals} K_\chi (1 + \cos(n_\chi - \delta)) + \sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 \\ &+ \sum_{nonbond} \varepsilon_{ij} \left[\frac{Rmin_{ij}}{r_{ij}}^{12} - \frac{Rmin_{ij}}{r_{ij}}^{6} \right] + \frac{q_i q_j}{e r_{ij}}, \end{aligned}$$

where b is a bond length, S is the distance between atoms separated by two covalent bonds, θ is a valence angle, χ is a dihedral angle, φ is an improper angle and r_{ij} is a distance between non-bonded atoms i and j. Parameters are force constant K_b, Urey-Bradley force constant K_{UB}, valence angle force constant K_θ, dihedral force constant K_χ, improper force constant K_{imp}, equilibrium bond distances b_0 and S_0, equilibrium valence angle θ_0 multiplicity n, phase angle δ, equilibrium improper angle φ_0, Lennard-Jones well depth ε_{ij}, minimum interaction radius $Rmin_{ij}$, partial charge q_i and dielectric constant e.

Parameters used by CHARMM can be evaluated from experiment or quantum me-

chanical calculations. It has to be mentioned that CHARMM employs partial atomic charges and therefore does not have special term to represent hydrogen bonding, which are treated as sum of van der Waals and electrostatic interactions.

The required coordinates for the atomistic simulation may be obtained from X-ray crystallographic or NMR structure data.

3.3.2 Generic Protein Model

The most accurate representation of the bio-system would be solution of the Newtonian equations for the all atoms of the system including explicitly modeling solvent molecules. With all its benefits, atomistic representation of the large bio-systems makes it difficult to perform simulations comparable with timescale of the large conformational changes. Generic representation of protein could be used for understanding general kinetic and thermodynamic features of the protein undergoing conformational changes. Because the dominant force of protein folding is hydrophobic force [24] protein could be represented as a bead-spring chain with polar and hydrophobic particles, called"HP" model undergoing coil-to-globule transition in a polar solvent.

The total potential energy U of a system configuration $\mathbf{r}$ is given by a sum of Lennard-Jones and bonded interactions:

$$U\left(\{\mathbf{r}\}\right) = \sum_{i} \sum_{j>i} U_{\mathrm{LJ}}\left(r_{ij}\right) + \sum_{kl} U_{\mathrm{FENE/LJ}}\left(r_{kl}\right), \tag{3.29}$$

where the double sum in the first term involves only nonbonded pairs of particles, while the summation in the second term involves only bonded pairs of particles.

Generally, the pair interaction between topologically nonconnected particles is

described by the standard truncated Lennard-Jones pair potential:

$$U_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} - \left(\frac{\sigma}{r_c}\right)^{12} + \left(\frac{\sigma}{r_c}\right)^{6} \right], \tag{3.30}$$

where ϵ is the depth of the potential well and σ is the distance where interparticle force is zero.

r_c represents the cutoff distance, if distance between a pair of particles larger than r_c the potential energy U equals to 0. For the fully repulsive interactions, we choose $r_c = 2^{1/6}\sigma$, yielding so-called Weeks-Chandler-Andersen excluded volume potential, [84] denoted U_{WCA}. For attractive interactions, we choose $r_c = 4.0\sigma$.

Topologically bound monomers interact according to the standard FENE/Lennard-Jones bonded potential, $U_{\mathrm{FENE/LJ}}$:

$$U_{\mathrm{FENE/LJ}}\left(r\right) = U_{\mathrm{WCA}}\left(r\right) + U_{\mathrm{FENE}}\left(r\right), \tag{3.31}$$

where

$$U_{\mathrm{FENE}}\left(r\right) = -\frac{a}{2} R_0^2 \ln\left[1 - \left(\frac{r}{R_0}\right)^2\right]. \tag{3.32}$$

Chapter 4: Kinetics of "Folding" of Comb-Graft Heteropolymer

4.1 Introduction

Molecular dynamics simulations of model coarse-grained comb-graft heteropolymer were performed to understand general mechanistic features of coil-to-globule relaxation. Microphase separation undergone by comb-graft heteropolymer is among the simplest model representations of protein folding. It has been shown that molecule of polymer with well separated blocks of hydrophilic and hydrophobic groups aggregates in selective solvent into a single micelle due to intrachain micellization [85, 86]. The driving force for amphiphilic polymers to collapse is lowering the free energy of polymer-solvent system.

In our work we discuss classical molecular dynamics simulation of generic comb-graft heteropolymer in both non-selective and selective solvent. The main focus of this study is to understand the effect of side-chain length on the relaxation kinetics that leads to intramolecular micellization, in the topological regime in which side chain are equal or longer of the length of backbone between two grafting points ("spacer"). We are particularly interested in comparing the relaxation time in selective solvent, t_r, and the characteristic Zimm time, t_z, in non-selective solvent.

4.2 Simulation Description

4.2.1 Model System

Our model system is depicted schematically in Fig. 4.1. We consider a single comb heteropolymer in which the monomers in the side chains are type "A" and monomers in the backbone are type "B," immersed in a dense solvent whose particles are type "C". Employing standard notation for the heteropolymer topological variables, we considered backbones of length $N_b = 30$ monomers, side chains length n varies as 5, 10, 20 monomers respectively, where these side chains are found branching from the backbone at equally spaced intervals of $m = 5$ monomers. Hence, our heteropolymers contain 5 total side chains, giving 25, 50, 100 total side chain monomers, or roughly more than one-half of the total mass of the heteropolymer, $N = 55$, 80, 130 respectively. The total system of heteropolymer plus solvent is confined in a cubic domain with periodic boundaries in all directions. We considered relatively dense systems with an overall particle number density of $\rho = 0.85$.

In order to drive a phase separation, we require a fully repulsive interaction energy for any pair of particles in which one is a side chain member and the other is not (AB and AC interactions), while all other interactions are identically attractive (AA, BB, CC, and BC); in other words, we enforce "poor" solvent conditions with respect to the backbone monomers. We encode this information in the convenient form of the Lennard-Jones pair potential:

$$U_{\mathrm{LJ}}^{\alpha\beta}(r) = \begin{cases} 4\left[\left(\frac{1}{r}\right)^{12} - \left(\frac{1}{r}\right)^{6} - \left(\frac{1}{r_c^{\alpha\beta}}\right)^{12} + \left(\frac{1}{r_c^{\alpha\beta}}\right)^{6}\right] & r < r_c^{\alpha\beta}; \\ 0 & r \geq r_c^{\alpha\beta}. \end{cases} \tag{4.1}$$

where $\alpha\beta \in (\mathrm{AA}, \mathrm{BB}, \mathrm{CC}, \mathrm{AB}, \mathrm{AC}, \mathrm{BC})$. For the fully repulsive interactions, we

choose $r_c^{\mathrm{AB}} = r_c^{\mathrm{AC}} = 2^{1/6}$, yielding so-called Weeks-Chandler-Andersen excluded volume potential, [84] denoted U_{WCA}. For all other r_c's, we choose 2.5, giving a well depth of about ϵ per pair at a distance of $r = 2^{1/6}$. Such a prescription is a crude model of the hydrophobic interaction, and has been used in several previous studies of homopolymer collapse [87, 88, 89].

Topologically bound monomers interact according to the standard FENE/Lennard-Jones bonded potential, $U_{\mathrm{FENE/LJ}}$ from relation 3.31. We choose standard parameter values for of $R_0 = 1.5$ and a =30 [90].

Reduced unit system is adopted for convenient expression of parameters: length is measured in particle diameters σ, energy in units of ϵ, and mass particle of mass, m. The unit of time is $\sigma\sqrt{m/\epsilon}$ (denoted τ) and the unit of temperature is k_b/ϵ, where k_b is Boltzmann's constant.

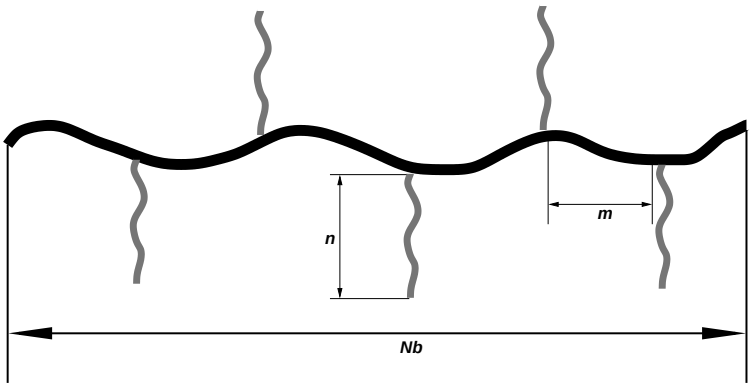


Figure 4.1: Comb-heteropolymer topological parameters: N_b is the backbone length, m is a spacer length and n is side chain length n

4.2.2 Sample Generation and Initial Conditions

The initial configuration of a comb-graft heteropolymer molecule is constructed as a random walk of beads with step equal to 0.97 (where U_{FENE} is minimal). Solvent particles are put on a lattice and their number are chosen such that length of simulation box sufficiently exceeds the size of the simulated molecule. Velocity components are randomly selected to achieve the desired temperature ($T = 1.0$). In order to guarantee separation between particles, a pre-equilibrium "warm-up" is performed. It consists of 4000 time steps and uses a time-dependent potential that incrementally increases from 0 to fully repulsive U_{WCA}. When the system transits from the random walk to "excluded volume" model all resulting velocities are scaled to adjust to the desired temperature.

4.2.3 Collapse Simulation

During the main part of simulation positions of particles are stored in regular time intervals, which is 25 τ for a polymer with side-chain of length 5 and 2 τ for a polysoap with side-chain of length 10 and 20 monomers. Visual monitoring of molecular state is performed. 50 independent configuration have been selected for each polysoap topology. The simulation runs until a stable radius of gyration of the side-chain agglomerate R_g^2 is reached, indicating that a heteropolymer molecule is fully collapsed:

$$R_g^2 = \frac{1}{N^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\mathbf{r}_i - \mathbf{r}_j)^2 \tag{4.2}$$

In the simulation we have used so-called velocity Verlet algoritm from relations 3.7 and 3.8 [69], Time step of $\Delta t = 0.004$ is chosen for stability reason. In order to reduced time of energy evaluation that scales as N^2 in the regular MD simulation,

the linked-list method is used. The linked-list method scales as N [91].

4.2.4 Simulation in Nonselective Solvent

We have performed series of MD simulations, where all pair interactions have been set attractive.

τ_Z is defined as the average time required for a chain to move its radius of gyration [92] or in the other words, the mean-squared displacement of monomers relative to the molecule center of mass, $g1$ and the mean-squared displacement of the molecule center of mass, $g2$ are equal. $g1$ and $g2$ are computed from

$$g_1(t) = \frac{1}{N}\langle \sum_i \{[r_i(t) - r_{cm}(t)] - [r_i(0) - r_{cm}(0)]\}^2 \rangle \tag{4.3}$$

$$g_2(t) = \langle [r_{cm}(t) - r_{cm}(0)]^2 \rangle, \tag{4.4}$$

where r_{cm} is a position of center of mass.

4.3 Result and Discussion

The main concern of this study is investigation of the influence of molecular topology on the coil-to-globule relaxation time in the backbone-selective solvent, τ_r. For this purpose we have run 50 independent trajectories starting from the different initial conditions for each side chain. In order to compare the coil-to-globule relaxation time with the Zimm relaxation time in nonselective solvent τ_Z, we have performed ten individual runs have been performed each topology, each run was several thousand τ in length.

The coil-to-globule relaxation time is seen to decrease with increasing side-chain length in a topological regime in which the side chain are longer than the backbone spacer between grafting points. This decrease is due to the higher probability of initial contacts among monomers in distinct side chains at the moment of quench when side chains are longer. A two stage relaxation is generally observed, in which rapid coalescence into one or more micelles occurs in the first stage, and slow coalescence of micelles into single globule occurs in the second stage. The second stage is likely governed by a free energy barrier, as evidenced by visual inspection of metastable states and logarithmic scaling of the number of clusters during this stage.

The relaxation into a globular state is many times faster than characteristic Zimm relaxation time in non-selective solvent. We also observed that this Zimm relaxation time strongly increases with side chain length. It has to be mentioned that our Zimm time does not obey the standard Zimm scaling $\tau_Z \propto R^3$, but close to $\tau_Z \propto R^6$. This effect could be explained by relatively small size of simulated chains. We therefore can conclude that for the small molecules micellization is a fast process on molecular time scale, even if some metastable intermediate form along the pathway.

The dependence of relaxation time in the backbone-selective solvent, τ_r and the Zimm relaxation time in non-selective solvent τ_Z of molecular topology are given in Table 4.1.

Although the simple comb-like model heteropolymers cannot accurately mimic the much more complicated proteins, our observation of their dynamic behavior could enrich the knowledge of the protein folding pathway. We have demonstrated that

Table 4.1: Effect of side-chain length, M, on initial radius of gyration of side-chain monomers, $\left(R_g{}^2\right)(0)$. Zimm relaxation time is in non-selective solvent, τ_Z; initial number of clusters, $N_c\,(0)$; and coil-to-globule relaxation time, τ_r. N_t is the total number of monomers. All measured quantities are averaged over 50 independent trajectories. Reduced Lennard-Jones units are used throughout. (number of backbone monomers, N, is 30 for all cases.)

M	N_t	$\left(R_g{}^2\right)(0)$	τ_Z	$N_c\,(0)$	τ_r
5	55	9.17	466	3.68	198
10	80	12.5	928	2.98	156
20	155	17.6	3230	1.84	105

generic comb-like heteropolymer with longer side chains would undergo complete coil-to-globule transition in deterministic fashion, but a heteropolymer with shorter side chains would collapsed via many intermediate states.

4.4 Conclusion

We have studied comb-like collapsing heteropolymers in selective and nonselective solvents. The coil-to-globule relaxation time decreases with increasing side-chain length in a topological regime in which the side chains are longer than the backbone spacer between the grafting points. A two stage relaxation is generally observed, composed of a deterministic coalescence into one or more micelles, as the first stage, and slow coalescence of micelles into single globule in the second stage. The relaxation time is many times faster than Zimm relaxation time, which lead to the conclusion that relaxation of the small size generic comb-like heteropolymer is a relatively fast process. Understanding what topological parameters control rates of various mechanisms could be a further step toward a better general understanding of the relationship between primary and secondary structure in self-assembling macromolecules.

Chapter 5: Effect of Confinement on the Thermodynamics of a Collapsing Heteropolymer

5.1 Introduction

This chapter describes the thermodynamics of protein folding inside the chaperonin central cavity. We have focused on the favorable effect of mechanical confinement inside the chaperonin cavity on the stability of native state of the substrate protein due to confining geometry's restriction on the number of accessible non-native conformations.

We have employed the Wang-Landau method of direct density-of-states calculation to study thermodynamics of protein folding to demonstrate that confinement changes the density of states of the substrate protein. We have modeled the substrate protein as a simple collapsing generic heteropolymer and the chaperonin cavity as a spherical pore with repulsive walls.

5.2 Simulation Techniques

5.2.1 The Model Heteropolymer

A three-dimensional system composed of a single chain of a generic multiblock-copolymer is encapsulated in a spherical pore of variable size with repulsive walls. The chain has length $N = 109$ monomers, represented by the palindromic sequence $(A_9B)_{10}A_9$. Type "A" monomers may be thought of as hydrophilic, and type B as hydrophobic. A snapshot of such a heteropolymer in an open conformation appears

in Fig. 5.1a.

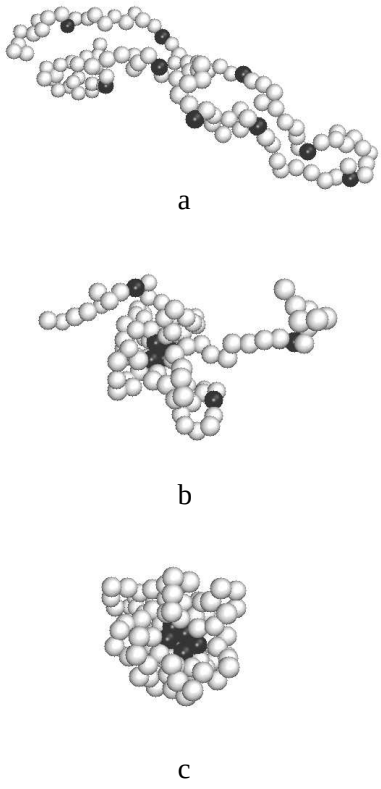


Figure 5.1: Representative snapshots of the A_9B heteropolymer. (a) an "open" conformation; (b) a "partially collapsed" conformation; (c) a "fully collapsed" conformation.

Particle interactions are pairwise additive and depend on the particle type and existence of chemical bonds between the monomers. A reduced unit system is adopted for convenient expression of parameters: length is measured in units of the particle diameter σ, energy in units of ϵ. The unit of temperature is ϵ/k_B, where k_B is Boltzmann's constant. The total potential energy U of a system configuration $\mathbf{r}^N$ is

given by a sum of Lennard-Jones, bonded, and wall interactions:

$$U\left(\mathbf{r}^N\right) = \sum_i \sum_{j>i} U_{\mathrm{LJ}}\left(r_{ij}\right) + \sum_{kl} U_{\mathrm{FENE/LJ}}\left(r_{kl}\right) + \sum_i U_w\left(r_i\right), \tag{5.1}$$

where the double sum in the first term involves only nonbonded pairs of particles, while the summation in the second term involves only bonded pairs of particles. The third term accounts for repulsive interactions with the wall of the confining sphere. Generally, the pair interaction between topologically nonconnected particles is described by the standard Lennard-Jones pair potential:

$$U_{\mathrm{LJ}}^{\alpha\beta}(r) = \begin{cases} 4n\left[\left(\frac{1}{r}\right)^{12} - \left(\frac{1}{r}\right)^{6} - \left(\frac{1}{r_c^{\alpha\beta}}\right)^{12} + \left(\frac{1}{r_c^{\alpha\beta}}\right)^{6}\right] & r < r_c^{\alpha\beta}; \\ 0 & r \geq r_c^{\alpha\beta}., \end{cases} \tag{5.2}$$

where $\alpha\beta \in (\mathrm{AA}, \mathrm{BB}, \mathrm{AB})$. BB pairs interact with an full Lennard-Jones potential with an attractive well; AB and BB pairs interact with a fully repulsive truncated Lennard-Jones potential. For the fully repulsive interactions, we choose $r_c^{\mathrm{AA}} = r_c^{\mathrm{AB}} = 2^{1/6}$, yielding so-called Weeks-Chandler-Andersen excluded volume potential, [84] denoted U_{WCA}. For attractive interactions, we choose $r_c = 4.0$. In order to ensure phase separation we choose the factor n equal to 10.0 for attractive interactions, while it is equal to unity for any repulsive interactions. For modest temperatures near unity, we know these strong attractive interactions drive a phase separation to a collapsed, core-shell structure, as shown in Fig. 5.1c.

Topologically bound monomers interact according to the standard FENE/Lennard-Jones bonded potential, $U_{\mathrm{FENE/LJ}}$ (eq.3.31). We choose standard parameter values of

$R_0 = 1.5$ and $a = 30.0$ [90]

Repulsive interactions between the spherical wall and any particle are given by a $10-4$ Lennard-Jones potential, U_w: [93]

$$U_w(r) = \begin{cases} 2\pi\left(\frac{2}{5}r^{10} - r^4 + \frac{3}{5}\right) & r < r_c; \\ 0 & r \geq r_c. \end{cases} \tag{5.3}$$

where r is the distance between the wall and particle,and r_c is equal to 1.0. We have studied confining radii of size $R = 5$, 7, 10, 20 σ and bulk (i.e, "infinite" radius).

5.2.2 Methods

We have employed the Wang-Landau method of direct density-of-states calculation in continuous three dimensional space, described in the chapter "Methods" (section "Wang-Landau Monte Carlo Simulation of Off-Lattice Chain").

To determine the appropriate energy range, we conducted Langevin dynamics simulations at $T = 1.0$ beginning from open conformations and noted the values of U accessed. We require a range from $U_{\text{low}} = 1950$ to $U_{\text{high}} = 2450$ to observe our chain in all relevant conformations (from open chain to a single-core globule). We divide the energy range into bins of size $\Delta U = 0.1$. This is a smaller bin size than that used by Rathore [72], but we found that this small bin size results in acceptance of a larger number possible conformations and should consequently lead to more accurate density of states. The scale factor f is initially set at 1.064. We explored several flatness criteria of various severities and settled on the convention used by Shell et al., namely that every bin in H has at least 20 hits [94]. We perform five independent

runs for each pore size to obtain better statistics.

5.2.3 System Initialization

The initial configuration of a heteropolymer is constructed as a random walk of N beads in 3D continuous space with a step equal to 0.97 (where $U_{FENE}(r)$ is minimal). In order to guarantee separation between particles, a two-step pre-equilibrium "warm-up" simulation is performed to gently introduce excluded volume and place the heteropolymer in a spherical pore. In step 1, all pairwise non-bonded forces during the warm-up are derived from the fully repulsive WCA potential (i.e., no attraction), capped at a value of 1.963 ϵ/σ. The step-1 simulation consists of 50,000 time steps of length of 0.01 and employs the dissipative particle dynamics (DPD) thermostat [74], with a friction of 0.5 and a temperature of 1.0. Bound monomers interact according to the standard FENE/Lennard-Jones bonded potential, $U_{\mathrm{FENE/LJ}}$. In step 2, in order to insert the chain into a spherical pore of a certain size, we create a pore, centered on the molecule center of mass, with diameter equal to the maximum end-to-end distance. We then perform constant temperature molecular dynamics simulation by reducing the sphere radius by 0.01 σ each time step until the desired pore size is reached. This second molecular dynamics warm-up employs U_{WCA}, $U_{\mathrm{FENE/LJ}}$ for bound monomers and the repulsive wall potential U_w. Both warm-up steps are performed using the standard velocity-Verlet integrator (relations 3.7 and 3.8) [69].

5.2.4 Off-Lattice Trial Moves

One challenge in conducting Wang-Landau simulations for continuous space systems is developing a suitable library of trial moves. We use a mixture of moves including

translation of center of mass, rotation of chain subsections around various axes defined by the chain geometry, and short NVE molecular dynamics trajectories. Trial moves acting on small segment of the chain are reffered to as "local", while trial moves operating on large chain segments or the whole chain are "global". Both local and global trial moves are equally important. The global moves are good for generation of unfolded states, but at the same time they fail to sample folded conformations in the low energy regime. The local moves continuously drive the system through energy space, but they are not sufficiently effective to explore the high energy regime. In detail, the trial moves employed are as follows, and their weights are reported in Table 5.1.

Table 5.1: Weighting of the Trial Moves

Move	Weight	in %
	bulk	confinement
rotational:		
-crankshaft	13.33	8.33
-simple hinge	13.33	8.33
-compound hinge	13.33	8.33
-global rotation	-	8.33
translational	-	29.17
MD	60.00	37.50

Translation. The center of mass of the molecule is shifted by a random vector $(\delta x, \delta y, \delta z)$, where each displacement is chosen randomly and uniformly between 0.1 and r_c. Clearly, this trial move is only employed for runs in the confining sphere and not in the bulk simulations for which there is no way for translation to produce a new value of U.

Global Rotation. A global coordinate system is centered on the molecule center of mass. The entire molecule is rotated by a randomly chosen angle between $-\pi$ and π around the z-axis, then by another angle, its cosine chosen between -1 and 1 around the y-axis, and yet a third time around the z-axis with an angle chosen between $-\pi$ and π. As with translation, this move is only applied for runs in the confining sphere.

Compound Hinge. We employ an off-lattice version of the traditional lattice hinge move, with an important generalization. Roughly speaking, a hinge move requires a defined subchain to be rotated through space as a rigid body with reference to a single monomer (the "hinge point"), while the remainder of the chain beyond the subchain is translated as a rigid body without rotation to maintain connectivity. In effect, this means that "two" intramolecular angles are potentially altered; hence we term this a "compound" hinge and depicted it schematically in Fig. 5.2. The details of the compound hinge are as follows. A hinge point monomer is selected at random from internal monomers, and designated "a". One of the two neighbors of "a" is chosen randomly, and it is then placed at a random polar angle θ and azimuthal position ϕ, both chosen uniformly, at a fixed distance from "a". (Strictly speaking, ϕ is chosen randomly from $-\pi < \phi < \pi$, and $\cos\theta$ between $-1 < \cos\theta < 1$, to ensure a uniform probability density on a surface of a sphere.) The same displacement in angle is then applied to all monomers between "b" and a third monomer "c" selected between "b" and the chain end on the other side of "b" from "a", rotation this group as a rigid body, preserving all internal angles. Monomers beyond "c" are translated without rotation to ensure connectivity. If monomer "c" is a chain-end, then this move is referred to as a "simple" hinge.

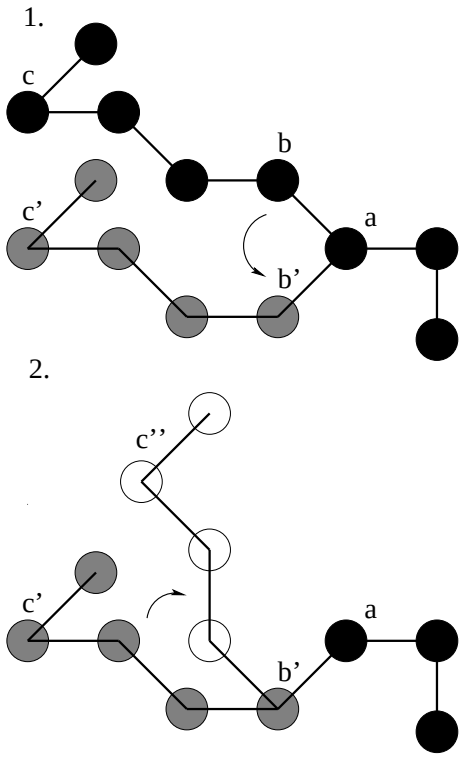


Figure 5.2: Schematic representation of the compound hinge trial move. Step 1 involves a hinge wchich swings the bond between "a" and "b"; step 2 swings the chain segment between "b" and "c" as a rigid body.

Crankshaft. An off-lattice version of the traditional lattice crankshaft move is also employed. Here, two monomers separated by at least one monomer along the chain are chosen, and designated a and b. The vector $\mathbf{r}_{ab}$ forms the axis around which monomers between a and b are rotated as a single rigid body by an random angle between 0 and 2π.

NVE Molecular Dynamics. While the moves listed above produce both large and small conformational changes, they are insufficient in allowing the system to visit compact configurations in reasonable time. As can be appreciated from Fig. 5.1c, the low-energy conformations our molecule adopts have a relatively densely packed core of B-type monomers. For this reason, it is advantageous to include a trial move

which allows *collective* rearrangement of a large number of monomers. Following the example of Rathore and de Pablo, [72] we therefore incorporate short NVE molecular dynamics trajectories in our library of trial moves.

The use of NVE MD as a trial move within an otherwise standard Boltzmann Monte Carlo simulation has acquired the name "hybrid Monte Carlo." [71] For an NVE MD trial move to satisfy detailed balance, the acceptance probability applied to other trial moves (Eq. 3.19) must be modified to

$$p_{\mathrm{acc}}\left(U_1 \rightarrow U_2\right) = \min\left[\exp\left(-\beta\Delta K\right)\frac{g\left(U_1\right)}{g\left(U_2\right)}, 1\right]. \tag{5.4}$$

The Boltzmann factor involving the change in kinetic energy, $\Delta K \equiv K_2 - K_1$, arises due to the *a priori* probabilities of (a) finding the system configuration with kinetic energy K_1 prior to attempting the forward move and (b) finding the system with kinetic energy K_2 prior to attempting the hypothetical reverse move. It is important to note, however, that temperature is not an inherent thermodynamic variable in the Wang-Landau method, yet the MD trial move introduces it in *two* places: (1) the temperature characterizing the Boltzmann distribution of velocities assigned to the particles to initialize the MD trial move, which we denote T_{MD}, ($K_1 = \frac{3}{2}Nk_{\mathrm{B}}T_{\mathrm{MD}}$), and (2) the temperature in the Boltzmann factor in the acceptance criterion, which we denote T_{B} ($\beta = 1/k_{\mathrm{B}}T_{\mathrm{B}}$). In traditional hybrid MC at constant T, these two temperatures are identical; yet this is not a requirement here.

For each global MD move, we randomly select number of MD steps between 1 and 30 and T_{MD} between 0.1 and 3.0 ϵ/k_B for bulk and between 1.0 and 4.0 ϵ/k_B for

confined runs. Higher values for T_{MD} help in sampling the high energy regime and with overcoming local minima in any energy regime. Lower values for T_{MD} ensure MD trial moves work in the low energy regime, favoring folding of the chain. (These values were arrived at by trial and error; confined runs were slower to converge if the maximum temperature was set at 3.0, for example.) Velocity components are randomly selected for the chosen segment to achieve the selected temperature while constraining the total momentum to zero. Molecular dynamics integration is performed with the standard velocity Verlet algorithm with a time-step equal to 0.01.

Another important distinction between our NVE MD trial moves and those typically used in hybrid Boltzmann MC is that we only apply the MD integration to a randomly selected subset of monomers while others are held fixed. The "MD-active" segment is chosen by randomly selecting two monomers, a and b, and requiring randomly that either (i) all monomers between a and b are MD-active

The strategy behind these "partial" MD trial moves is to allow collective rearrangement on a variety of length scales throughout the simulation. We are allowed to use such moves here because we are effectively tuning the masses of our monomers to make the MD moves more efficient (infinite mass means a monomer does not move), and because masses are not part of the total potential energy (Eq. 5.1), no balance criterion is violated. We do not systematically compare the use of "partial" MD to "total" MD trial moves in this work, but our anecdotal observations so far are that their use greatly speeds up convergence in simulations of bulk chains or chains in large pores.

During the simulation, we measure cluster size distributions at each energy bin. A

cluster is composed of one or more hydrophobes, and if two hydrophobes are separated by less 1.5 σ, they belong to the same cluster. Cluster size is denoted m, and number of clusters of size m is denoted n_m. Our heteropolymer has 10 hydrophobic beads, so the fully open chain would possess $n = 10$ clusters of size $m = 1$. Likewise the fully folded chain would possess $n = 1$ cluster of size $m = 10$. The clustering analysis is helpful for visualization of a macrostate, corresponding to each energy bin, as an average of microscopically different partially folded chains.

5.3 Results and Discussion

5.3.1 Simulation Results

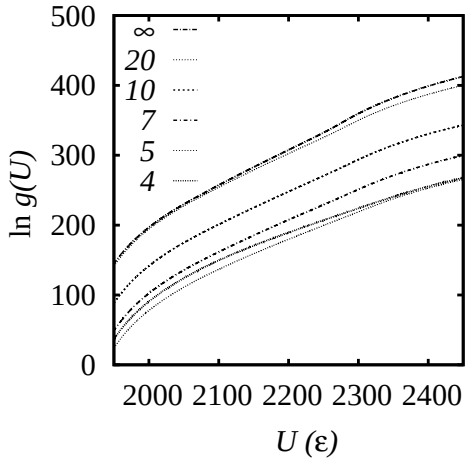


Figure 5.3: Natural log of the density of states vs. total potential energy for different pore sizes.

In Fig. 5.3, we show the logarithm of densities-of-state, $\ln g(U)$, for different confinement radii, R, and for bulk as a function of total potential energy. These are "raw" simulation results in that no shifting or scaling for reasons of comparison have been made. Each curve is the average of five independently calculated densities-of-

state. As discussed previously, we are not strictly allowed to compare the absolute magnitudes of one $g(U)$ to another. However, because each simulation ran for the same number of iterations, we can infer that the higher $g(U)$'s, corresponding to larger R, required many more histogram hits than did the lower $g(U)$'s, corresponding to lower R, to achieve flatness. This is an indirect and slightly imprecise measure of the degree to which confinement reduces $g(U)$. Thus from Fig. 5.3 we infer that the number of accessible states in the high energy regime decreases with decreasing pore radius, because confinement reduces number of possible conformations of the open chain. Some expanded conformations of the unfolded chain cannot exist in the confined space. [11]

Fortunately, the $g(U)$ relationships give a direct and unambiguous view of the thermodynamics of our confined heteropolymers. That the heteropolymer undergoes a first-order "melting" transition in bulk can easily be seen from the internal energy, $\langle U(T) \rangle$ (Eq. 3.21), shown in Fig. 5.4. The melting temperature, T_m, is identifiable where $\langle U \rangle$ undergoes a sharp jump; for bulk, $T_m^b = 1.96$. For the confined heteropolymers, T_m is seen generally to increase as R decreases, although it is somewhat hard to pinpoint a value for T_m for the lowest R due to the rather shallow nature of the jump. It is more straightforward to ascribe T_m as the temperature at which there is a peak in the heat capacity, $C(T)$ (Eq. 3.22), as depicted in Fig. 5.5.

We instead followed the original method of Wang and Landau [75] choosing to ascribe T_m to that temperature at which the two peaks in the probability distribution are of equal height, meaning the most probable microstate in both states have the same probability. Sweeping T in increments of 0.001 from 1.0 to 4.0, we are able to

extract $T_m(R)$ from probability distributions. The probability distributions at the various T_m for each R are plotted in Fig. 5.6 and values for T_m are reported in Table 5.2. We see that $T_m(R)$ increases as R decreases. That is, confinement seems to stabilize the low energy state.

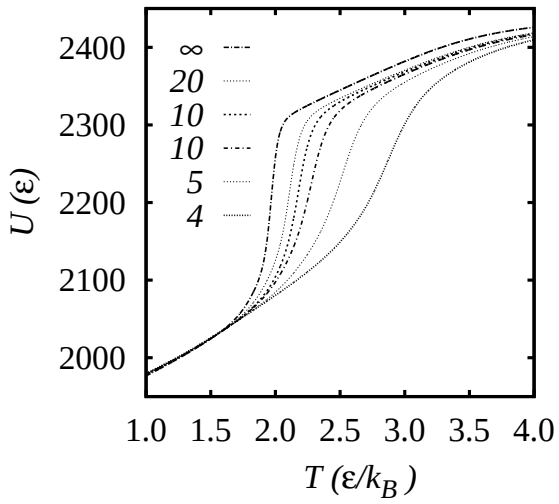


Figure 5.4: Internal energy vs. temperature for different pore sizes.

The transition detected in $\langle U \rangle$ is the melting of a core-shell "folded" conformation (e.g., Fig. 5.1c) to form an open conformation (e.g., Fig. 5.1a). Not only is this apparent from snapshots, but also from the clustering analysis. As an example, we consider the case for confinement at $R = 20$. Probability distributions of potential energy for various temperatures are shown in Fig. 5.7a, and clearly show the first-order nature of the transition. Below this, in Fig. 5.7b, we show the cluster size distributions as a function of potential energy. We see that, at low energies, the states are mostly comprised of conformations with a large cluster of size seven or

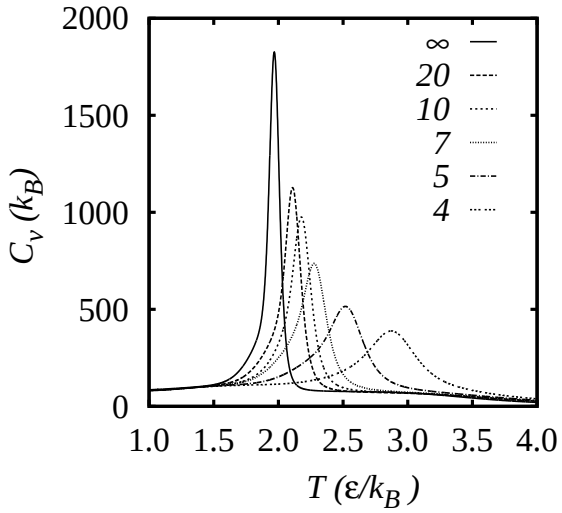


Figure 5.5: Heat capacity vs. temperature for different pore sizes.

higher with the balance made up by single clusters, while at high energies, the states are mostly comprised of conformations of many single hydrophobes. At the lowest energies, we observe conformations with a single cluster of ten hydrophobes, while at the highest energies, we observe conformations of ten free hydrophobes. Interestingly, the transition state, identifiable from the probability distribution as the minimum between the two peaks at a temperature of about 2.1, corresponds to conformations with exactly five free hydrophobes.

5.3.2 A Simple Thermodynamic Scaling Model for the Shift in the Melting Temperature

The natural log of the density of states is the entropy of the chain for constant energy according to our adopted reduced unit system:

$$S\left(U\right) = k_B \ln g \tag{5.5}$$

At an equilibrium between two phases, the free energy of the two phases is equal.

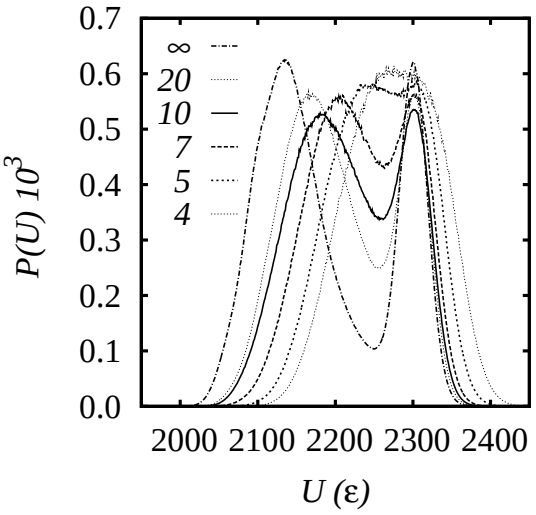


Figure 5.6: Probability distributions (Eq. 3.20)at respective melting temperatures for various pore sizes.

-

Table 5.2: Thermodynamic Properties of the Model Heteropolymer Confined in Spherical Pores of Various Sizes, R. All values shown are in reduced Lennard-Jones units

R	T_m	ΔU	ΔS
4	2.93	40.0	13.6
5	2.56	63.0	24.6
7	2.29	93.0	40.5
10	2.18	120.0	54.9
20	2.11	137.0	65.0
∞	1.96	165.0	84.2

Hence,

$$\Delta U = T_m \Delta S, \tag{5.6}$$

where ΔU is the change in internal energy upon "melting," (a "latent heat") and ΔS is the corresponding change in entropy (a "latent entropy"), and T_m is the transition, or "melting," temperature. where the superscript b denotes the cases for bulk (i.e., no confinement). From here, it is straightforward algebra to show change in melting

temperature upon confinement is given by

$$\frac{\Delta T_m}{T_m^b} = \frac{\frac{\Delta U}{\Delta U^b}}{\frac{\Delta S}{\Delta S^b}} - 1, \tag{5.7}$$

where the superscript "b" denotes the cases for bulk (i.e., no confinement).

With densities of state in hand, we can now address how the latent heat and the latent entropy individually depend on the size of the confining sphere. We first note that there is some minimal radius, denoted R_{min}, below which is impossible to observe a first order transition. This trend is apparent from the probability distributions in Fig. 5.6: as R decreases, the two distinct peaks merge. Second, we require that, as $R \to \infty$, $\Delta U(R) \to \Delta U^b$ and $\Delta S(R) \to \Delta S^b$. With the properties in mind, we now suppose that, in dimensionless forms, these latent quantities can be expressed as

$$\frac{\Delta U(R)}{\Delta U^b} = \left(1 - \frac{R_{min}}{R}\right)^p \tag{5.8}$$

$$\frac{\Delta S(R)}{\Delta S^b} = \left(1 - \frac{R_{min}}{R}\right)^q, \tag{5.9}$$

where p and q are positive scaling exponents. Combining this relations, we obtain

$$\frac{\Delta T_m}{T_m^b} = \left(1 - \frac{R_{min}}{R}\right)^\gamma - 1, \tag{5.10}$$

where $\gamma = p - q$.

By allowing R_{min} and γ to vary, we fit the relation in eq 5.10 to our simulation data

for $\frac{\Delta T_m}{T_m^b}$ and obtained R_{min}=2.923 and γ=-0.306. This value of R_{min} is close to the mean radius of gyration of the collapsed state, which is about 3. We then used the resulting value of R_{min} in fits of the relations 5.8 and 5.9 to obtain p=1.060 and q=1.368. The simulation data and fits for ΔU and ΔS are shown at Figure 5.8 a. In Figure 5.8b we plot our scaled shift in melting temperature, with the fit model of eq 5.10. First, both the simulation and the model predict that confinement increases the melting temperature, so long as there is a first-order transition to observe. Furthermore, the model states that the melting temperature increases with reduction of the pore size for any $R > R_{min}$ if $\gamma < 0$ or $q > p$. This means that in order for this stabilization to occur the latent entropy must decrease with R faster than does latent heat. In other words, stabilization upon confinement is guaranteed if entropy rather than energy is more sensitive to the confinement volume. In contrast, in the case of positive γ ($q > p$), for $R > R_{min}$, the shift in T_m upon confinement is always negative. This reveals a possible route to using confinement to destabilize folded proteins: the cavity must be designed in such a way that the latent heat is more sensitive to cavity size than is the latent entropy. We might speculate that such a cavity would have a certain density of reactive sites on its inner walls, then we might expect the latent heat to be sensitive to cavity size by means of the internal surface area. Entropy is generally sensitive to the volume of the cavity, so a cavity that destabilizes folded proteins would be unlikely to be a perfect sphere with smooth walls.

We now consider whether our form gives a better fit to the simulation data than previous scaling relationships. Also shown in Figure 5.8b are curves corresponding

to the simple scaling relationships of the form

$$\frac{\Delta T_m}{T_m^b} = (R_g/R)^\nu \tag{5.11}$$

with ν values of 2 and 3.25 [66] with proportionality constants extracted from fits, for comparison. Our model outperforms both simple scaling laws, especially for large degree of confinement, where "large" is defined as confining sphere radii almost 10 times as large a the radius of gyration of the collapsed state, R_{min}.

A word of caution is warranted at this point. Our simulations were performed for a single chain length, topology (number of hydrophobes), and Hamiltonian, so it is premature to state that we have discovered a "new" scaling behavior for the thermodynamics of intra-molecular self-assembly(i.e., "collapse" and "folding") in confined geometry. However, because the Wang-Landau method gives us a detailed view of the thermodynamics, we are able to delineate the effect of confinement on both latent heat and latent entropy, which is difficult if not impossible using other direct simulation techniques so far considered.

5.3.3 Free Energy Barrier

Confinement of a chain inside a spherical pore associates with the free energy lost, proportional to the relative sizes of the chain and pore:

$$\Delta G = RT \log \left(1 - \frac{R_c}{R_s}\right)^3, \tag{5.12}$$

where ΔG is access of the Gibbs free energy, R_c and R_s are size of the chain and pore respectively [9].

We have computed free energy as a function of the temperature using eq.3.23 and presented at Fig 5.9. This loss of free energy during placing a chain into the cavity has to be compensated by some outside energy sources. For the cavity of chaperonin GroEL, compaction of a substrate protein chain could be result from either partial aggregation of the chain itself outside of the cavity or direct compressive action of chaperonin system upon lodging the substrate into the cavity [62]. We think that ATP binding, as a part of the chaperonin systems working cycle, may provide the necessary amount of free energy required for compression of a protein chain in order to place it into the cavity.

As been mentioned earlier the confinement could change the free energy landscape of folding with comparison to bulk. The free energy barrier of folding (melting) transition could be calculated at the melting temperature for each pore size as

$$\Delta F = F_t - F_u = F_t - F_f = -k_B T_m \log \frac{P(U_t)}{P(U_f)}, \tag{5.13}$$

where u, f and t denote unfolded, folded and transition states of the heteropolymer.

From Fig. 5.10 one can see that the confinement reduces the free energy barrier, allowing a greater population of unfolded states to attain energy needed to overcome the transition energy and proceed to the folded conformation. We explain this effect that the confinement helps to synchronize energy gain and entropy lost making transition state less different from the "pre-reacted" unfolded conformation, which

becomes more compact with decreasing the pore size. As result of this action the free energy landscape getting smoother and for the strong confinement $R = 4$ the energy barrier is almost nonexistent.

For larger and structurally much more complex proteins, the studied process of folding of the collapsing heteropolymer may describe only one step of multistep protein folding and therefore the unfolded state of our model heteropolymer may be one of many folding intermediates. So, considering the effect of the confinement on reduction of the free energy barriers, the confinement will help proteins avoid kinetically trapped states, making folding pathway to occur only among compact native-like intermediates. This effect of confinement on free energy landscape on protein folding inside the inner cavity of the bacterial chaperonin GroEL has been suggested by Brinker and coo-workers [13] and explains the facilitation of folding of the larger substrate protein assisted by the chaperonins.

5.3.4 Acceptance Ratios

In Fig. 5.11, we show acceptance ratios for the various trial moves as functions of potential energy of the originating state for any move. Fig. 5.11a shows the acceptance ratios for the bulk simulations, while Fig. 5.11b shows them for the case of confinement at $R = 5$. We observe a good rate of acceptance of the MD moves, about 40%, over a broad span of the energy domain for both bulk and confined simulations. Clearly, MD moves are quite valuable to this method and relatively more important for strongly confined heteropolymers as opposed to more weakly confined hetropolymers. Also, the bulk simulations benefit greatly from the conformational

moves (crankshaft and hinges) only at relatively high energies for which the conformations are predominantly open, as one would expect. For confined simulations, we see that global rotations and translations are crucially important, while they are by construction not needed in the bulk runs. Also in the confined runs, the conformation shifting moves are almost unimportant. It seems that NVE MD, global rotation, and global translation are most important in covering potential energy space, though we noticed that completely neglecting the conformational moves in the bulk runs results in much slower convergence.

5.4 Conclusions

The Wang-Landau method of direct density-of-states calculation has been used for off-lattice simulation of mechanical confinement of a generic multiblock copolymer undergoing first-order collapse transition. We have shown that folding inside the spherical cavity is favorable because of destabilization of the unfolded state of a protein due to confining geometry's restriction on the number of accessible "open" conformations. The conformational restriction of the chaperonin cavity leads to reduction of the free energy barrier of the folding process. We have developed a suitable library of trial moves, including a "partial NVE MD" trial move, which have been shown useful for both free and confined bead-spring heteropolymer. Using this technique we have quantified the increase in thermal stability of a collapsed globular state in a spherical confining geometry, measured as an increase in melting temperature with decreasing radius of confining sphere. We have shown that a simple scaling relationship that allows the change in latent energy of the transition to scale independently of the

change in latent entropy with the radius results in a better fit to the simulation data, compared to simpler scaling relationships considered previously. This indicates a new direction in pursuing a theoretical understanding of both the thermodynamics and kinetics of confined proteins. Moreover, our work supports the use of flat-histogram Monte Carlo methods to study heteropolymer thermodynamics in various settings.

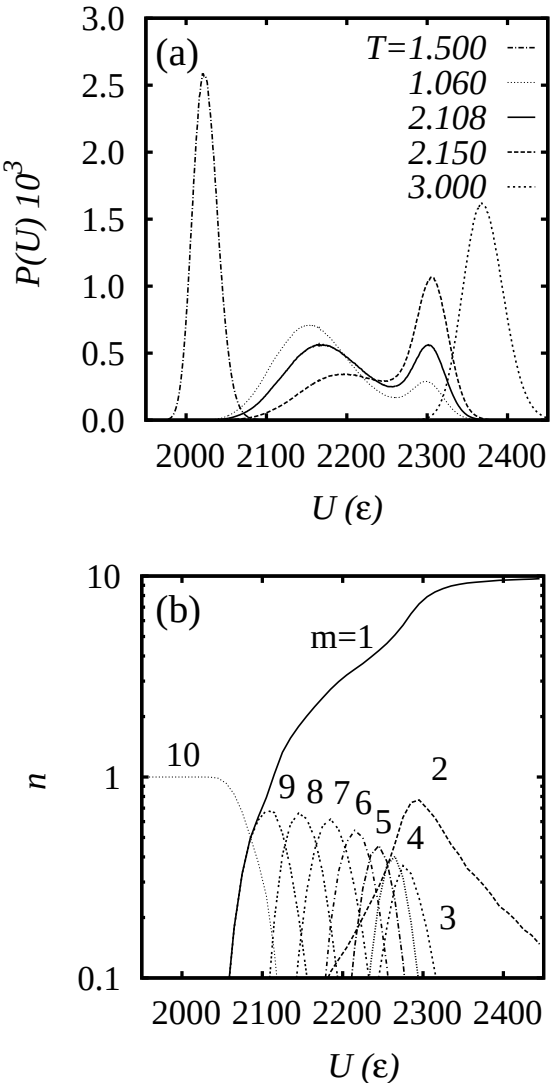


Figure 5.7: (a) Normalized probability distributions for various temperatures for the case of confinement at $R = 20$. (b) Cluster size distributions as functions of potential energy for $R = 20$; each curve labeled m corresponds to clusters of m hydrophobes and the y-axis measures the number of clusters, n.

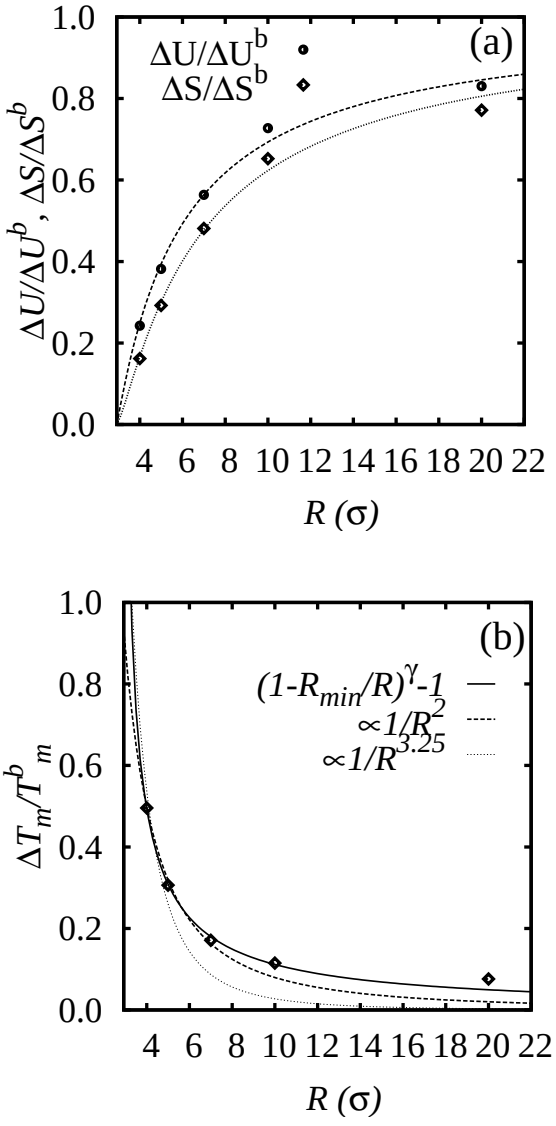


Figure 5.8: Increase in thermal stability of native states with increasing confinement. (a) Scaled latent heat, ΔU, and latent entropy, ΔS, vs pore size, R. Curves are fits to eqs 5.8 and 5.9 with p=1.060, q=1.368, and R_{min}=2.923. (b) Relative shift in melting temperature upon confinement, $\Delta T_m(R)/T_m$, as a function of the pore size. Solid curve is Eq. 5.10 with the scaling exponents noted; Dashed curves show scaling relationships with a single exponent of 2 and 3.25, respectively.

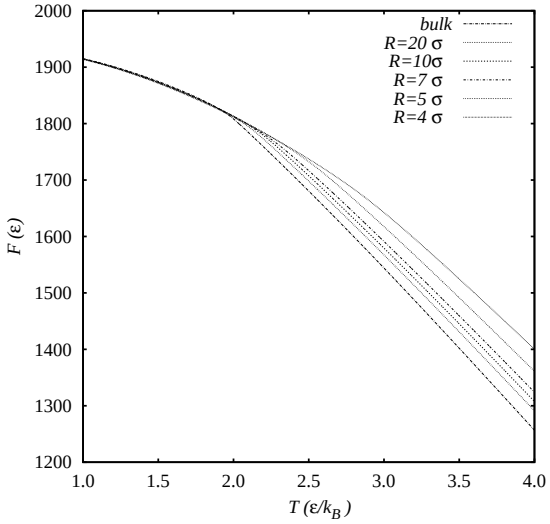


Figure 5.9: Free energy vs. temperature for different pore sizes.

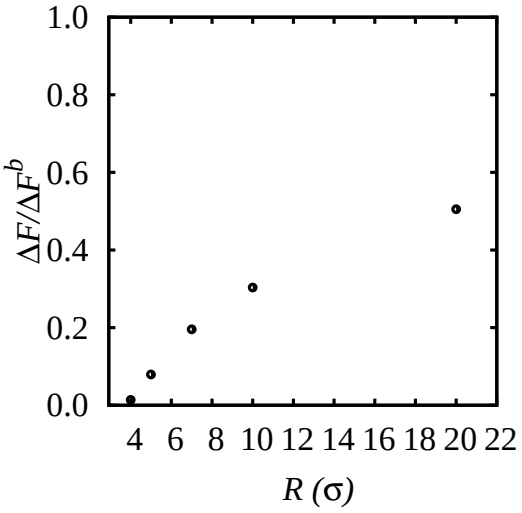


Figure 5.10: Scaled free energy barrier ΔF for folding-melting transition (Eq. 5.13) for various pore sizes.

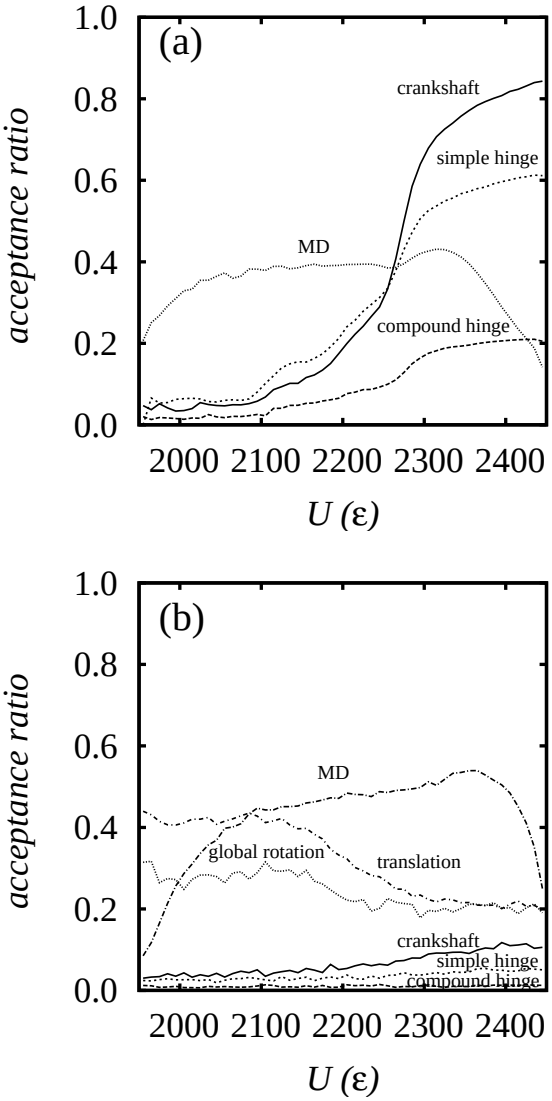


Figure 5.11: Acceptance ratios vs. potential energy, for the various trial moves: ((a) bulk; (b) confined at $R = 5$.

Chapter 6: Conformational changes in the *E. coli* GroEL subunit

6.1 Introduction

This chapter investigates the working cycle of the most thoroughly studied chaperonin the *E. coli* chaperonin GroEL, concentrating on the conformational transition of the GroEL subunit upon ATP binding.

We present a comparison of equilibrium structures and quasi-harmonic modes of apo and ATP-bound forms of the *E. coli* chaperonin GroEL subunit protein (57 kDa) employing all-atom simulation of a single subunit of GroEL in the explicit solvent we also directly demonstrate the spontaneous nature of the several important onformational changes. We also considered the mechanisms of positive and negative cooperativity, responsible for much of GroEL function. The major questions we seek to answer are the structure-functional relationship of ATP-bound and ligand-free states of GroEL and interpolation of these properties of a single subunit of GroEL to the working cycle of the whole chaperonin complex.

6.2 Model and Methods

We generated two 20-ns MD trajectories: one with KMgATP, designated the *holo* simulation, and one without, designated the *apo* simulation, using NAMD v2.5 [95] and the CHARMM force field [82, 83]. Initial coordinates for both the *holo* and *apo* forms of the subunit were taken from the crystallographic data of the $(\text{GroEL-KMgATP})_{14}$

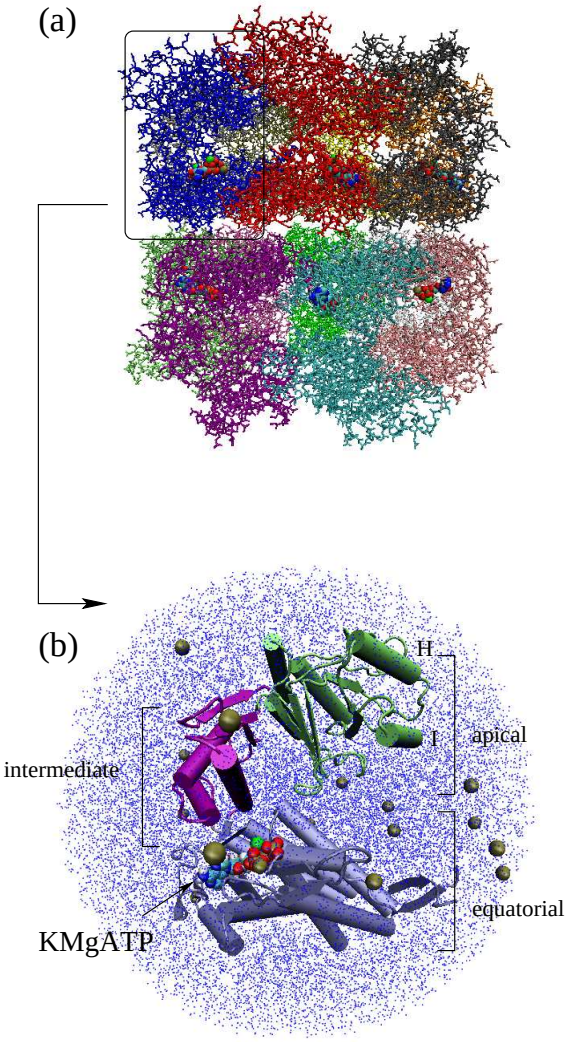


Figure 6.1: Schematic representation of construction of the simulation system for the *holo* (ATP-bound) simulation. (a) The tetradecamer, $(\text{GroEL-KMgATP})_{14}$, from crystallographic data of Wang and Boisvert [PDB code 1KP8 [3]] (b) Single GroEL subunit with bound KMgATP and counterions in water solution. Protein is shown with simplified secondary structure representation, nucleotide atoms and ions as van der Waals spheres, part of the explicit solvent shell is schematically represented by blue dots.

complex [3] (PDB code 1KP8). Although ATP is present in the binding pocket in this structure, the intermediate domain is not closed, perhaps due to crystal packing effects. For the *holo* simulation, the Mg^{2+} and K^{+} ions and six crystallographic waters were retained; these along with ATP were deleted to generate initial coordinates for the *apo* simulation. Hydrogen atoms were inserted assuming pH 7.0. Each subunit was solvated by random placement of water molecules within a sphere of radius 56 Å, maintained using harmonic boundary conditions (k = 10 kcal/mol/Å^2). Potassium ions were added at random locations for charge neutrality. Each system has about 69,000 atoms. No holonomic bonded constraints were used, full electrostatics were calculated every second step, and van der Waals interactions were cut off beyond 12 Å. Each system was energy-minimized via 500 cycles of conjugate-gradient optimization. MD was then run for 20 ns with a 1 fs time step. The temperature held at 37 °C using a Langevin thermostat on all heavy atoms with a coupling coefficient of 5 ps^{-1}. A schematic explaining the construction of the simulation system appears in Fig. 6.1.

As will be detailed in the next section, our *holo* and *apo* subunits access three major tertiary states, for which we assign the following designations: (*i*) **a** designates the *apo* state, which we consider the final 10 ns of the *apo* simulation. (*ii*) **t** designates the state with ATP bound with low affinity, which we take as the state accessed between 2 and 6 ns in the *holo* simulation. (*iii*) **r** designates the ATP-bound state with the closed binding pocket, which we take as the state accessed between 10 and 20 ns in the *holo* simulation.

Using the appropriate average coordinates from the simulations, we constructed

model intra-ring trimers from among various permutations of each state by aligning equatorial domain backbone α-helix atoms against those atoms in six distinct reference heptamers: (1) either ring in $(\text{GroEL-KMgATP})_{14}$ [1KP8 [3]], (2) either ring in $(\text{GroEL})_{14}$ [1CXK [1]], the ring (3) *cis* or (4) *trans* to GroES in the ADP bullet [1AON [2]], the (5) *holo* or (6) *apo* ring in $(\text{GroEL-KMgATP})_7$-$(\text{GroEL})_7$ [2C7E [96]]. Each trimer is labeled with a three-letter string denoting the order of the subunit tertiary states (e.g., **aaa**) when viewed transverse to the plane the ring from outside the heptamer ring. We generated the following trimers against all reference structures (averaging among all unique trimers within each reference ring): **aaa**, **ata**, **ara**, **ttt**, **trt**, and **rrr**.

C_α quasi-harmonic analysis was performed for each state (**a**, **t**, and **r**) using at least 4 nanoseconds of integration sampled at 10-ps intervals, using built-in routines in the AMBER software package [97]. In standard quasi-harmonic analysis, one diagonalizes a mass-weighted cross-correlation matrix of coordinate fluctuations about a mean structure. Quasi-harmonic analysis is relatively accurate method to describe conformational transitions of the proteins and according to Tama et al [79] yields a single low-frequency normal mode corresponding the collective motion of the conformational change. In the analyses that follow, we describe the our approach of finding a single mode responsible for the conformational change.

In order to identify a mode that describes the conformational change of a protein we define this mode as a mode which contributes the most to the coordinate difference between average structures of two conformational states. The difference between

averaged equilibrated structures of states A and B for each i atom is given by

$$\vec{x_i} = \vec{x_i^B} - \vec{x_i^A} \tag{6.1}$$

Cartesian displacement of structure A by a mode of the interest is given by $\alpha M^{-1/2}\vec{y^A}$, where α is an amplitude, M is kinetic energy diagonal matrix and $\vec{y^A}$ is a mode of the interest [78]. M is constant for our analysis and therefore Cartesian displacement may be defined for an atom i as $a\vec{y_i^A}$, where a is new amplitude.

Mean square deviation (msd) between equilibrium structure B obtained from molecular simulation and one resulted from quasi-harmonic motion of structure A by a mode of the interest is define as

$$msd = \frac{1}{N}\sum_{i=1}^{N}\left(\vec{x_i} - a\vec{y_i^A}\right)^2, \tag{6.2}$$

The opmimal amplitude is found from condition of the minimum msd

$$\frac{d\,(msd)}{da} = 0 \tag{6.3}$$

$$a = \frac{\sum_{i=1}^{N}\vec{x_i}\vec{y_i^A}}{\sum_{i=1}^{N}\vec{y_i^A}^2} \tag{6.4}$$

Then substituting a from 6.4 to 6.2 we may define a mode with the lowest value of msd as a one responsible for conformational transition between two states A and B.

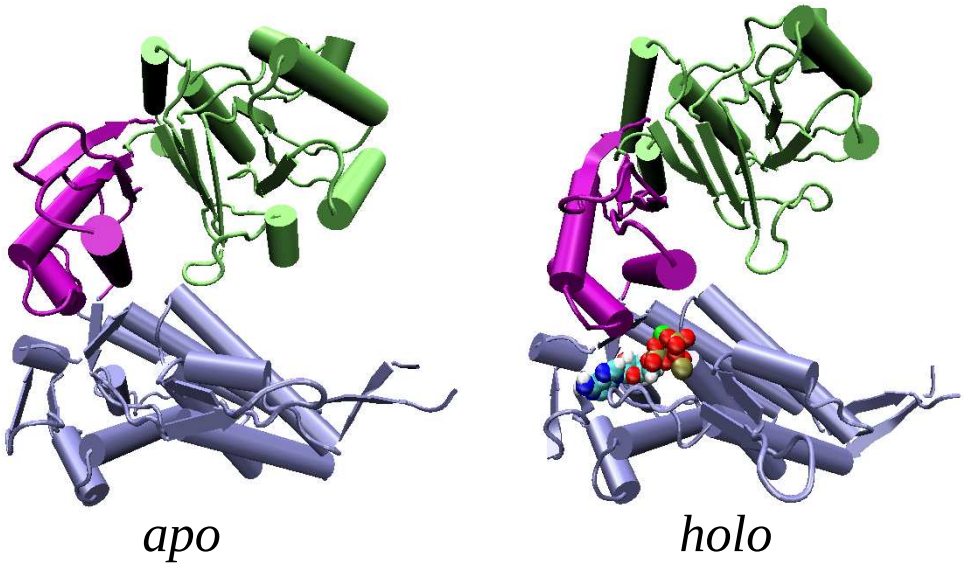


Figure 6.2: *apo* and *holo* subunits of GroEL: snapshot from molecular dynamics simulation. The structures are fitted by the equtorial domain. Protein is shown with simplified secondary structure representation. The equatorial, intermediate and apical domains are blue, purple and green respectively. ATP and ions are shown as van der Waals spheres.

6.3 Results and Discussion

6.3.1 The t→r Transition. Closing of Nucleotide Binding Pocket

We conducted two 20-ns MD simulations of the GroEL subunit with initially open binding pockets: the first with ATP in the pocket (*holo*) and the second without (*apo*). The most important result of our simulation is spontaneous occurrence of the **t→r** transition with KMgATP in the nucleotide pocket. We have observed the closing of the located on the equatorial domain of GroEL nucleotide binding pocket with the intermediate domain for ATP-bound GroEL subunit. Fig. 6.2 shows the conformational difference between *apo* and *holo* subunits of GroEL.

The sequence of steps in the transition and their abruptness can be appreciated

from Fig. 6.3, in which we show interatomic distance traces from the *holo* and *apo* simulations, with reference measurements from the appropriate crystal structures. The Asp155-Arg395 salt bridge ruptures at about 6 ns, as shown by the trace of the distance between the C_γ atom of Asp155 and the C_ζ atom of Arg395 in Fig. 6.3a. In the *holo* simulation, we see that this distance closely matches that observed in the *cis* ring of the "ADP bullet" crystal structure [2], which is the structure we assume as a reference displaying unambiguously closed binding pockets. We observe that, although this salt bridge breaks once early in the *apo* run, it reforms quickly and is remarkably stable for the remainder of the 20 ns simulation. This bridge is intact in all subunits of the *apo* tetradecamer [1]. Next, Ile150 is pulled over Gly492 and settles into a stable hydrophobic interaction with Ile493 and the adenosine group of ATP after another nanosecond, as shown by the trace of the distance between the C_δ atom of Ile150 and the $C_{\gamma 2}$ atom of Ile493 in Fig. 6.3b. This hydrophobic contact is absent in the *apo* crystal structure and present in the *cis* ring of the ADP bullet.

We think that the driving force of the motion of helix M comes from both electrostatic attraction between Mg^{2+} and the oxygens of Asp398 and a water-mediated hydrogen bond network joining Mg^{2+} and Asp398. There is a single water tightly coordinating the ion that donates an H-bond to an intermediate water that donates an H-bond to either oxygen of Asp398. This intermediate water is not static during first 10 ns of the simulation, but this site is almost always occupied by a water molecule. During the transition, this water is squeezed out and Asp398 accepts an H-bond from the water directly bound to the metal ion, which is thereafter a stable arrangement,

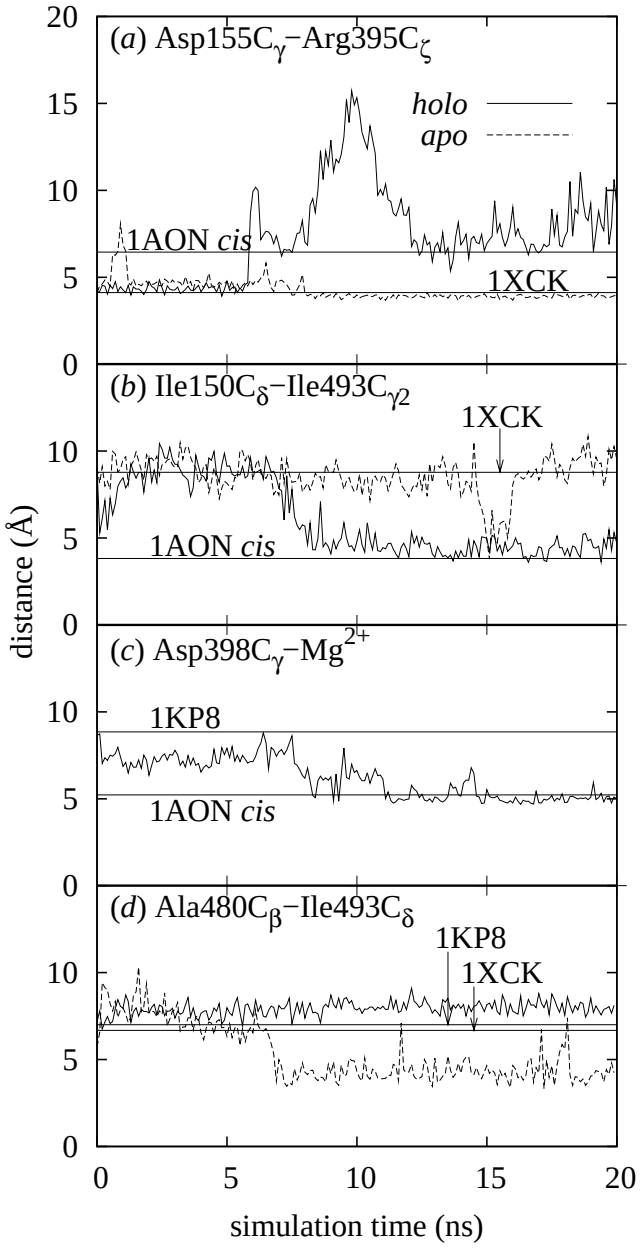


Figure 6.3: Interatomic distance traces from the MD trajectories reveal important events in the **t**→**r**(a-c) and Ala480-insertion (d) transitions. Distances shown are between (a) C_γ of Asp155 and C_ζ of Arg 395; (b) C_δ of Ile150 and $C_{\gamma 2}$ of Ile493; (c) C_γ of Asp398 and the Mg^{2+} ion of KMgATP; and (d) C_β of Ala480 and C_δ of Ile493. Horizontal lines denote measurements from reference structures indicated by their PDB accession codes (see text).

as can be seen in the distance trace between the C_γ atom of Asp398 and Mg^{2+} in Fig. 6.3c.

After the major motion of the transition is complete, the distance between Asp398 and the Mg^{2+} agrees with that found in the closed binding pockets in the *cis* ring of the ADP bullet. The stability of the closed nucleotide binding pocket of ATP-bound GroEL may be explained by of interactions between the intermediate domain and bound nucleotide. The nucleotide interacts with the intermediate domain at two locations: one is hydrophobic and van der Waals interactions between Ile 150 of helix F (residues 141 to 152) and adenosine of ATP and another is between carboxylate oxygen of Asp 398 of helix M (residues 386 to 409) and bound magnesium ion through a water molecule(Fig. 6.4). There is also a potassium ion bound to the backbone oxygen of Leu 494, the oxygen of carboxyl group Asp 495 and the oxygen of ribose hydroxyl of ATP, which helps to make the hydrophobic-polar arrangement more stable. The last interaction may be responsible for hydrolytic cleavage of the γ phosphate bond, however, such a reaction is strictly impossible to observe in our simulation because the CHARMM force field preserves bonds. Based on the comparison between the distance traces in Fig. 6.3a-c and the respective reference distances, we submit that the final state of our *holo* simulation is the **r** state.

These bonds were not observed for $(\text{GroEl-KMgATP})_{14}$ crystal structure [3], which could be explained by certain limitation of this method. The partial closing of the binding pocket upon ATP binding was experimentally observed from cryo-electron microscopic representation of ATP-bound states of GroEL [21]. The interactions

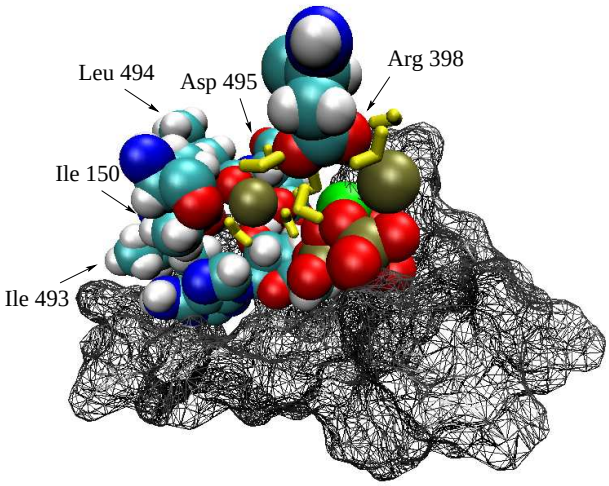


Figure 6.4: Ile 150 and Asp 398 closing the nucleotide binding pocket. Snapshot from molecular dynamics simulation. The residues of the nucleotide binding pocket are shown as a wireframe surface model. ATP, ions, Ile 150, Asp 398, Ile 493, Leu 494 and Asp 495 are shown as van der Waals spheres. Water molecules appear as yellow van der Waals sticks

between Ile 150 and Asp 398 of the intermediate domain and ATP and magnesium ion are present in GroEL-GroES-ADP_7 crystal structure [2], but the binding pocket is in a more closed position and therefore the positions of Ile 150 and Asp 398 are different. We propose that GroES binding is necessary for further downward rotation of the intermediate domain.

The major role in stabilizing the intermediate domain in downward position is played by interactions between residues Ile 150 and adenosine base of ATP. The importance of interactions between Ile 150 and ATP was mentioned by Sigler al et. [39] and confirmed by mutation studies. The I150E mutant is devoid of ATPase activity, nevertheless could bind the substrate protein, but cannot released it and bind GroES [46]. At the same time the D398A mutant has reduced ATPase activity, but

otherwise functions in the same fashion as wild chaperonin [6, 21]. These observations suggest that interaction between Ile 150 residues and the bound nucleotide is a key for successful closing of the nucleotide pocket and, at the same time, reorientation of the apical domain.

Helices F, G, and M are the only helices in the intermediate domain, and F and M are connected via different hinges directly to the equatorial domain. F and G rotate with M to maintain the hydrophobic core of the intermediate domain, but because of the distinct positions of the two hinges, the rotation involves significant internal structural rearrangement of this domain. In fact, the Asp155-Arg395 bridge, because it anchors M to G, opposes the transition because it opposes the rearrangement that must occur to allow the rotation. This explains why the **t** state remains apparently stable for 6 ns, during which time the Asp155-Arg395 bridge is constantly challenged by the force of attraction between Arg398 and Mg^{2+}. When it finally ruptures, the transition is rapid, reminiscent of uncoiling of a coiled spring. It is particularly interesting that the single underlying structural feature of GroEL that necessitates relative motion of helices G and M during the transition is that the entire protein is one giant hairpin with opposing strands in the intermediate domain with four distinct hinge points.

The Asp155-Arg395 bridge is known to be important for positive cooperativity. In a recent study of the Asp155→Ala mutant, Danziger et al. [98] showed that removing the Asp155-Arg395 salt bridges led to a sequential rather than concerted **T**→**R** transition. Our simulation shows that Asp155-Arg395 exerts a force on helix M which is opposed by Asp398-Mg^{2+} attraction. The former is "weaker" than the latter because

it eventually yields catastrophically. If Glu386-Asp197(R) ("R" denotes that Asp197 belongs to the right-hand neighboring subunit) also exerts a force on helix M, then, working together, both bridges offer greater resistance to the motion demanded by the Asp398-Mg^{2+} interaction than is offered by either individually, consistent qualitatively with the interpretation of Danziger et al. [98].

Interestingly, it does not appear necessary that the Asp155-Arg395 salt bridge break in order for Ile150 to sample the empty hydrophobic nucleotide binding cavity, as shown by the data from the *apo* run in Fig. 6.3b in the vicinity of 15 ns. But detailed examination of the *apo* trajectory revealed that this contact is short-lived and quickly reverted to the original conformation. In order for it to be stable ATP in the binding pocket is nessesary. Also for the hydrophobic interactions to be steady Ile150 has to protrude futher into hydrophobic cavity. This action is prevented by Asp155-Arg395 bridge. One may conclude that breakage of Asp155-Arg395 salt bridge is necessary for **t→r** transition.

6.3.2 During t→r the apical domain rotates counterclockwise

Fig. 6.5a shows the change in angular orientation of three representative helices with time, projected into the plane of the GroEL ring for the *holo* simulation, where helix H and L are in the apical domain and helix M is in intermediate domain, directly closing the binding pocket. The data display that the apical domain rotates in-plane between 15 and 30 degrees counterclockwise, moving along with the intermediate domain as a single unit. Fig. 6.5b shows the analogous data from the *apo* simulation. Therefore, closing of a binding pocket occurs as result of global motion of the intermediate and

apical domain with respect to the equatorial domain of GroEL subunit and connects the downward motion of the intermediate domain with simultaneous rotation of the apical domain.

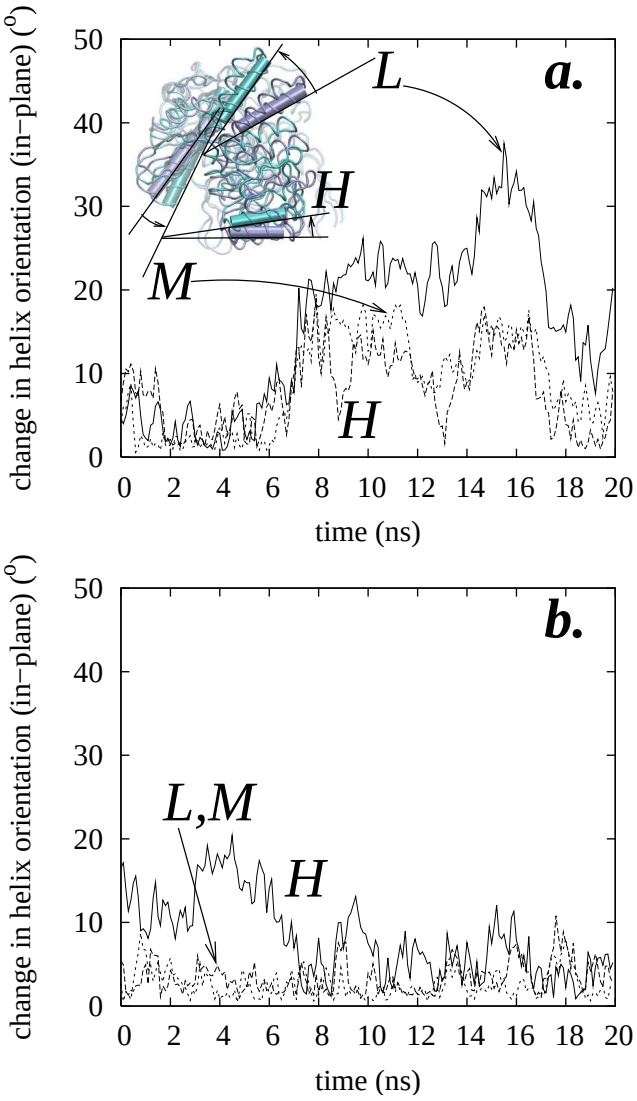


Figure 6.5: During the **t**→**r** transition, the apical domain rotates counterclockwise and the upper wing (helices K and L) swings away. (a) Change in helix orientation, projected into the GroEL ring plane, vs. time for helices H, L, and M from the *holo* simulation. Inset: overlay of **t** (grey) and **r** (blue) trace structures, top view, with helices H, L, and M denoted. (b) Same as (a) for the *apo* simulation.

Significantly, our simulation results agree with the cryo-EM structures of **R**-state rings [21] with regard to both the direction and magnitude of rotation of the apical domains upon ATP binding, in disagreement with structures from a previous simulation study [56]. Ma et al. simulated the entire **t**→**r** transition using targeted molecular dynamics (TMD) [58], in which fictitious forces on all C_α's drove the structure from **t** to **r** in under one nanosecond, in effect producing a non-dynamic pathway that interpolates between the endpoints. Because the net in-plane rotation of the apical domain is about 90° *clockwise* based on comparison of *apo* and ADP-bullet crystal structures [2], an initial 20° counterclockwise rotation of the apical domain does not appear along a TMD pathway. Our simulation supports the view that the GroEL subunit naturally shifts from **t** to **r** in response to the presence of ATP in the binding pocket.

6.3.3 T→R transition

$\mathbf{t}_7 \rightarrow \mathbf{r}_7$ transition of one GroEL ring is positive with respect to binding of ATP [52, 50, 51] with concerted nature of the transition [99, 100]. Switching of intersubunit salt bridges [21, 40] has been invoked to explain positive intraring cooperativity in ATP binding and concerted nature of **T**→**R** transition has been explained by steric clashes arising when one subunit undergoes **t**→**r** transition and either neighbor does not [55, 56].

To determine both the extent to which steric clashes arise and the disposition of intersubunit salt bridges when one subunit in a heptamer exists in a state different from either its left-hand or right-hand neighbor, we analyzed model trimers constructed

as described in Methods. Models were constructed using average structures from our simulations replicated and aligned along equatorial domains of reference structures.

We estimate the disposition of Glu386-Arg197 intersubunit salt bridges from the C_α-C_α distances reported in Table 6.1. We observe that both Glu386-Arg197 bridges are likely intact in **aaa** and **ata** trimers, and that the Glu386-Arg197 bridge is intact in an **rt** pair, based on alignments against the *apo* structure 1XCK [1]. Both appear broken when the central subunit is in the **r** state with **a** or **r** neighbors, and, interestingly, when all three subunits are in the **t** state. This suggests that what we label the **t** state is more accurately thought of as a short-lived intermediate on the *apo*→**t** transition. Indeed, Yifrach and Horovitz were the first to show that the intermediate state of any subunit in the heptamer **T**→**R** transition is **t**-like in structure but already has broken Glu386-Arg197 salt bridges [99].

We also considered the Glu386-Lys80 salt bridge, which has been shown to be intact for neighboring subunits in **R** state rings [21, 96], and is thought to play an important role in governing positive intraring cooperativity because it presumably stabilizes the multisubunit **R** state. We do not observe formation of this bridge in any alignments, most likely because there is no driving force for motion of Glu386 toward K80 of a second subunit in our single-subunit simulations. We noted that, when any central subunit in a trimer is **r**, the Glu386-Lys80 C_α-C_α distance is about 2.0 Å lower than in the *apo* crystal structure (data not shown).

We cannot claim to have observed breakage of the Glu386-Arg197 bridge because we conducted a simulation of only a single subunit. We therefore cannot claim that Asp398-Mg^{2+} attraction is sufficient to break both the Asp155-Arg395 and Glu386-

Asp197(R) bridges, but we are able to show that the observed **r** state requires that these bridges be broken, as a piece of additional evidence that the state reached by the simulation is indeed the **r** state.

No steric clashes involving atoms in the apical and/or intermediate domains are found in any trimers except those for which equatorial domains of the *cis* ring of the ADP bullet [2] were used as references. This was expected because of the slight inward tilt of the bullet equatorial domains. We disagree with previous modeling studies [55, 56] regarding steric clashes. Trimers constructed by replicating average structures with equatorial domains aligned along equatorial domains in several crystal structures revealed no clashes when one subunit was in our **r** state with **t**-state or *apo* neighbors. We note in passing that aligning a subunit from the **R** state ring of the cryo-EM structure [21] into an *apo* ring [1] also resulted in no such clashes as additional support for our observations. This is presumably because of an important difference between clockwise and counterclockwise in-plane rotation of apical domains in a heptamer ring. Counterclockwise rotation is allowed for a single apical domain, while clockwise rotation, at least to the degree reported previously, requires concerted rotation among all seven subunits [55, 56]. Given that counterclockwise rotation was apparent in the cryo-EM structures of $(\text{GroEL-ATP})_7$-$(\text{GroEL})_7$ of Ranson et al. [21], and that counterclockwise rotation naturally occurs in the simulation, it is unlikely that concertedness among subunit transitions in the overall **T**→**R** transition has anything to do with steric clashes.

6.3.4 Quasi-harmonic analysis

Table 6.1: C_α-C_α interatomic distances in the Glu386(L)-Arg197(C) ("l→c") and Glu386(C)-Arg197(R) ("c→r") bridges in trimers aligned against the 1XCK reference structure [1]

	Glu386-Arg197 distances (Å)	
Trimer	l→c	c→r
aaa	12.4	12.4
ata	13.6	12.6
ara	15.0	14.1
ttt	14.8	14.8
trt	15.5	13.4
rrr	16.2	16.2
1XCK	*11.9*	

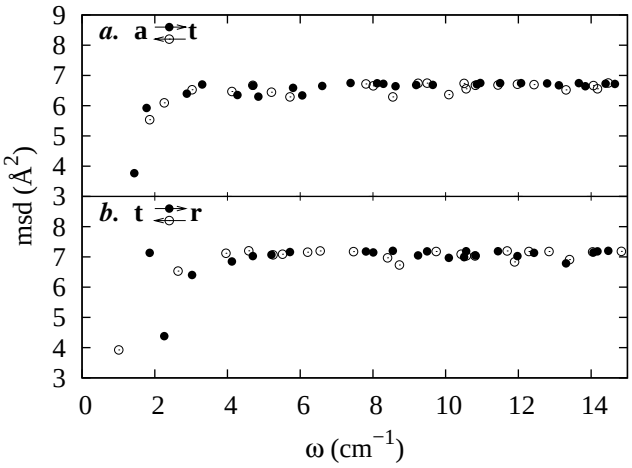


Figure 6.6: Minimum mean-squared differences between initial structures displaced along each quasi-harmonic mode and a target structure vs. mode frequency. (a) Transition from the **a** to **t** structure and back. (b) Transition from the **t** to **r** structure and back.

Quasi-harmonic analysis have showed that natural modes of vibration facilitate tertiary structural changes.

Testing the hypothesis of Tama and cooworkers [79] that one unique low frequency mode is responsible for a conformational change of in particular proteins we sought those modes whose motion contribute most strongly to the observed transitions. Because we observe a large-scale, spontaneous transition between two stable states directly, we have fully self-consistent data sets on which to test the notion that

natural modes of vibrations drive conformational transitions [55]. We computed all quasi-harmonic modes with frequencies of 50 cm^{-1} or lower for the **a**, **t**, and **r** states, as described in Methods, where the **t** state is a state with ATP bound prior to closure of the pocket, while **a** denotes the strict *apo* state. From a set of modes, the mode which, when used to project the initial structure, results in the structure most similar to the target, measured by mean-squared deviation, is the "most important mode" in driving the transition between the initial and target structure. We decribed this method in details in the section 6.2. Fig. 6.6 shows the results of this analysis in terms of the minimum mean-squared-deviation of initial structures projected along each mode relative to a target structure versus mode frequency.

Fig. 6.6a. indicates that there is one single normal mode ($\omega = 1.44$ cm^{-1}), which is the lowest frequency mode, responsible for **a**→**t** transition and there are no modes to drive the transition in the opposite direction. That may be explanation of irreversibility of ATP binding to GroEL causing, as we mentioned before, the short-lived **t**-like intermediate [99], which will then undergo reversible **t**↔**r** transition. In fact, both **t**→**r** and **r**→**t** transitions have a single mode driving each transition (Fig. 6.6b.). Values of the frequencies of these modes are close: $1.86cm^{-1}$ and $1.01cm^{-1}$ for **t**→**r** and **r**→**t** respectively.

Structures obtained by quasi-harmonic analysis using the decsribed method, are relatively accurate approximation of the structures of GroEL subunit conformational states via MD simulations (Fig. 6.7) . Mean-squared deviation between predicted and obtained form MD trajectory **a**, **t**, **r** structures is $3.77\mathring{A}^2$, $4.38\mathring{A}^2$ and $3.24\mathring{A}^2$ respectively. The difference between quasi-harmonic predicted and "real" (obtained

via MD) may be explained by the interatomic interations occuring close to transition point, which are not described by quasi-harmonic approach.

One should be cautious not to overinterpret the quasi-harmonic modes seen for the single subunit by extrapolating what these motions may accomplish in a complex, because the natural modes of the complex might be quite different in structure due to intersubunit contacts. However, it reasonable to expect that similar motion exists when a subunit is embedded in a complex.

6.3.5 Ala480 insertion into the empty nucleotide pocket

The triad Ile493, Ala480, and Pro33 defines the hydrophobic part of the nucleotide pocket: Ile493 and Pro33 sandwich adenosine while Ala480 contacts its N1. In the *apo* simulation, after about 7 ns, we observe a shift of residue Ala480 which places the methyl of Ala480 directly between Pro33 and Ile493 (Fig. 6.8). We include the distance trace between C_β of Ala480 and C_δ of Ile493 in Fig. 6.3d. The Ala480 insertion remains stable for the remaining 12 ns with occasional direct contact between Pro33 and Ile493. With Ala480 in this position, the accessible hydrophobic surface area owned by Pro33 and Ile493 together drops from about 100 Å^2 to 85 Å^2 and would lead indicate to change in exposed hydrophic area for this transition. This arrangement of this hydrophobic area is significant, and a bound ATP, were it present, would have to shift more than one Å along its long axis to allow a clash-free contact of Ala480 with N1. Because Asp87 and Thr89 anchor the phosphate-side of the binding pocket, such a collapse of the pocket would like render it incapable of accomodating ATP. The same reduction of hydrophobic area would give even more dramatic effect

for bound ADP, because of the smaller potential area of polar interactions.

That Ala480 insertion occurs spontaneously in the *apo* simulation could be expected given that the binding pocket is empty and the transition results in no significant net change in the amount of exposed hydrophobic surface area. The loop containing Ala480 is on the outer surface of the complex and no intersubunit interactions would easily prevent the motion of this loop that accompanies Ala480 insertion, suggesting this transition can occur easily in *apo* complexes. However, no *apo* crystal structure with Ala480 in the binding pocket has been observed, possibly because crystal forces may suppress exposure of Ala480 or more likely because no wild-type ATP-containing crystal structure exists. Nevertheless, it is reasonable to expect that Ala480 insertion competes with nucleotide binding for hydrophobic contacts with Pro33 and Ile493, as a pocket into and out of which Ala480 can place its methyl spontaneously will necessarily have a lower affinity for ATP than one for which Ala480 remains in place below the pocket.

This leads us to suggest a novel mechanism for negative interring allostery, shown schematically in Fig. 6.8. Ma and Karplus suppose that negative allostery requires that bound ATP in ring 1 lowers affinity for ATP in ring 2, requiring a state of "super" low ATP affinity, designated the $\mathbf{t}'$ state [55]. Here, we present a mechanism for nucleotide-depending shifting of an *apo* subunit toward the $\mathbf{t}'$ state that involves a network of steric and electrostatic interactions. Based on the most recent *apo* crystal structure of the tetradecamer [1], it was asserted that the only major interring contacts are the Arg452-Glu461 salt bridge. Arg452 lies one turn from the C terminus of helix P, at whose C terminus lies Cys458 which maintains a steric contact with Ala480.

Glu461 lies at the N terminus of helix Q and at the C terminus of a short loop connected to Cys458. We propose that the presence of ATP in the binding pocket a subunit of ring 1 forces Ala480(1) into Cys458(1), forcing the planar ring to rotate and shift such that Cys458(2) is driven into Ala480(2), a change in conformation that shifts the equilibrium in the *apo* subunit toward an Ala480-filled binding pocket.

6.3.6 Conclusions

We are the first to observe the **t**→**r** transition in an isolated GroEL subunit which is a large-scale allosteric transition directly in an unbiased, all-atom molecular dynamics simulation. The simulations show that the presence of KMgATP in the binding pocket causes by this transition due to both electrostatic attraction between Mg^{2+} and the oxygens of Asp398 and a water-mediated hydrogen bond network joining Mg^{2+} and Asp398. The newly formed **r** structure is stabilized by hydrophobic and polar interactions between residues of intermediate and equatorial domains and KMgATP.

Counterclockwise apical domain rotation in agreement with that inferred from cryoelectron microscopy structures [21] is observed. An important aspect of our work is that we show that the flexibility and forces required to effect the **t**→**r** transition are encoded in the subunit sequence and do not depend upon interactions in the complex. As pointed out by Ma and Karplus [55], this indicates that the **T**→**R** transition in a GroEL ring is better thought of as coupled tertiary changes rather than concerted changes in quaternary structure. Based on model trimers, no intersubunit steric clashes arise during the **t**→**r** transition.

We are also the first to observe the spontaneous insertion of Ala480 into the empty

nucleotide pocket, a transition that may give rise to the "super low ATP-affinity" $\mathbf{t}'$ state, and we postulated a testable mechanism of inter-ring negative cooperativity in ATP binding that relies on this transition.

Quasi-harmonic analysis clearly shows that natural modes of vibrations drive conformational transitions, revealing particular modes responsible for $\mathbf{t}\leftrightarrow\mathbf{r}$ and $\mathbf{t}\leftrightarrow\ \mathbf{t}'$ transitions. This finding is in agreement with theoretical statement of Horwich et al. that $\mathbf{t}\rightarrow\mathbf{r}$ transition may result from normal range of the thermal motion of the intermediate domain, where maximum extent of the motion is stabilized by the bound nucleotide [40].

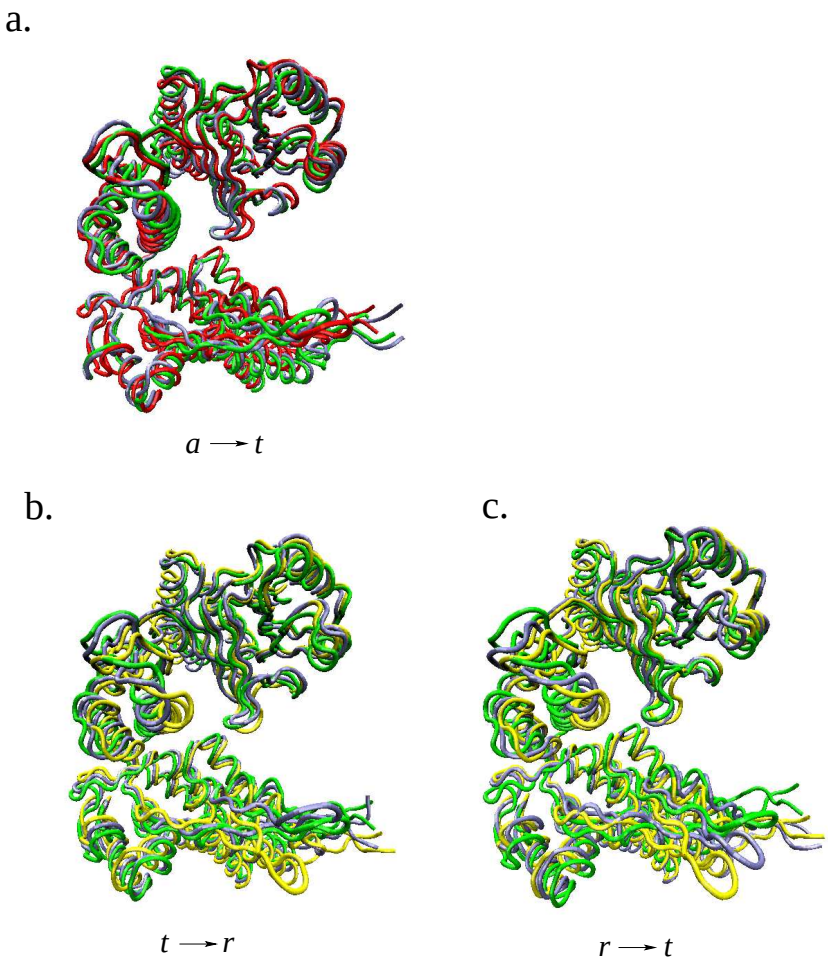


Figure 6.7: "Real" (obtained via MD) and quasi-harmonic predicted states of GroEL. Real structure are the time average structures obtained from MD trajectories, where **a**, **t**, **r** states are represented by red, green and yellow colors respectively. Quasi-harmonic structures have light blue color. (a) Transition from the **a** to **t** structure (b) Transition from the **t** to **r** structure (c) Transition from the **r** to **t** structure

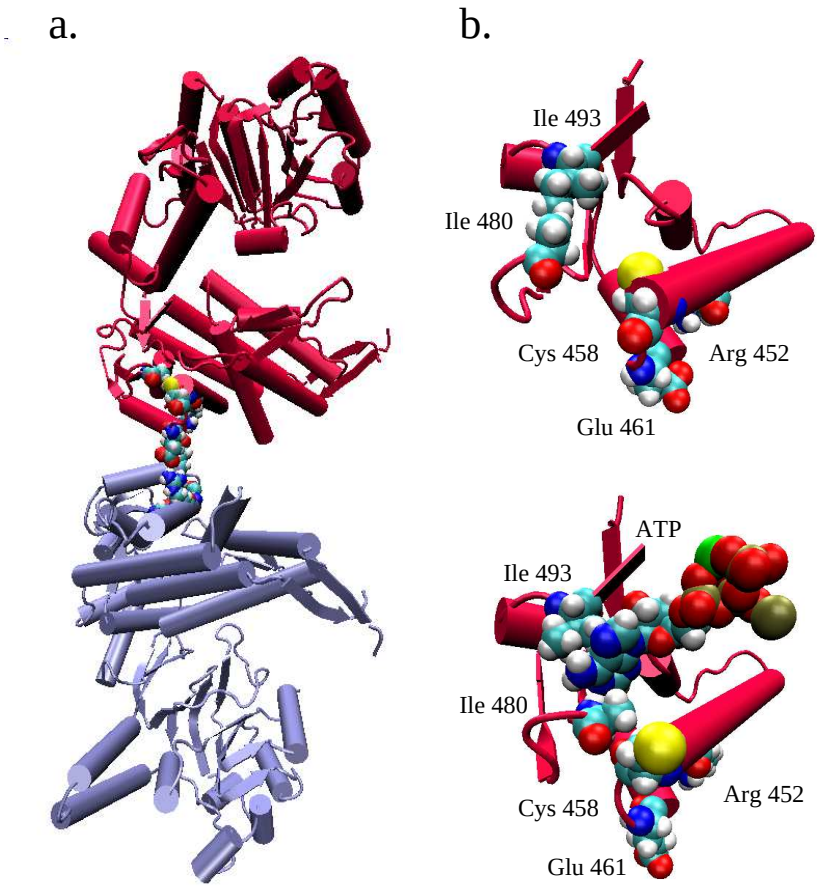


Figure 6.8: A speculation on the mechanism of interring cooperativity. (a) Select subunits in ring 1 and 2 from the *apo* crystal structure [1]. (b) Detailed view of the fragment of the equatorial domain of *apo* and *holo* GroEL subunit responsible for interring communications. The important residues are shown.

Chapter 7: All-atom molecular dynamics simulations of the *E. coli* GroEL ring

7.1 Introduction

This chapter describes communication between the nucleotide binding pockets of neighboring subunits of a ring of the *E. coli* chaperonin GroEL.

We present equilibrium structures of ATP-bound form of the *E. coli* chaperonin GroEL ring using all-atom simulation of one ring of GroEL in explicit solvent. We are particularly interested in the mechanisms of positive intraring cooperativity. The main focus of this part of our work is to study communication between subunits depending on the nucleotide binding pockets occupancy and understand the mechanism of interplay between structural changes of an individual subunit and the entire ring of GroEL. Although a direct tennanosecond computer simulation with atomistic resolution of the chaperonin ring cannot show any major conformational changes happening on the millisecond timescale, it can reveal interactions responsible for setting a favorable background for these changes.

7.2 Methods

We generated 12-ns MD trajectory using NAMD v2.5 [95] and the CHARMM force field [82, 83]. Initial coordinates of one ring, containing seven identical subunits were taken from the crystallographic data of the $(\text{GroEL-KMgATP})_{14}$ complex [3]. We retained the Mg^{2+} and K^{+} ions and six crystallographic water per subunit. The ring

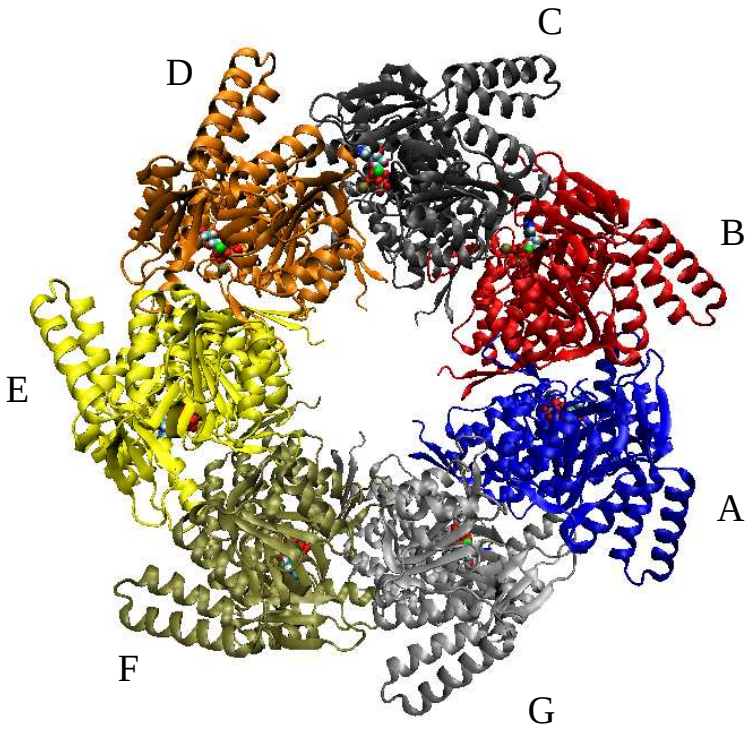


Figure 7.1: Schematic representation of the simulation system: ATP-bound ring of GroEL from crystallographic data of Wang and Boisvert [PDB code 1KP8 [3])]. Subunits are denoted as A, B, C, D, E, F, and G. Protein is shown with simplified secondary structure representation, nucleotide atoms and ions as van der Waals spheres. Water and counterions are not shown.

was solvated by random placement of water molecules within a sphere of radius 90 Å, maintained using harmonic boundary conditions (k = 10 kcal/mol/Å^2). Potassium ions were added at random locations for charge neutrality. The system has more than 285,000 atoms. No bonded constraints were used, full electrostatics were calculated every second step, and van der Waals interactions were cut off beyond 12 Å. The system was energy-minimized via 500 cycles of conjugate-gradient optimization. MD was then run for ∼11.5 ns with a 1 fs time step. The temperature held at 37 °C using a Langevin thermostat on all heavy atoms with a coupling coefficient of 5 ps^{-1}. A schematic explaining the construction of the simulation system appears in Fig. 7.1, where subunits are denoted as A, B, C, D, E, F, and G.

7.3 Results and Discussion

7.3.1 Asymmetry among GroEL subunits

One important result our simulation is an observation of significant asymmetry among the subunits of GroEL ring. Although having seven bound molecules of ATP (one molecule for each subunit) GroEL ring adopts a conformation where some of the subunits are in the **t** state while others, subunits C and G, are structurally closer to the **r** state (Fig. 7.2 and 7.2). For the definitions of **t** and **r** states we have used the same methodology as presented in Chapter 6, measuring distance between hydrophobic and charged parts of KMgATP and corresponding residues of the intermediate domain of GroEL.

This asymmetry may be due to the high flexibility of GroEL subunits, capable of adjusting their structure in response to structural changes undergone by other subunits, where each subunit is in close contact with its neighbors. The subunits do

not have interlinking covalent bonds, but they are joined via the complicated network of charged and polar interactions. One of the most important of them is the Glu386 and Arg197 salt bridge, which is somehow essential for intersubunit communication [21, 40]. During the course of the simulation, we have discovered the that Glu386-Arg197 salt bridge can switch to Glu386-Lys277 and sometimes even a three residue charged network Lys277-Glu386-Arg197 is formed for a short time (Fig. 7.4). Formation of a network involving more than two charged and polar residues has stronger cooperative effect for stability than do formation of isolated, buried salt bridges [25, 26], therefore connection among subunits of GroEL via charged interaction is present during the entire course of the simulation.

7.3.2 Mechanism of positive cooperativity

In the act of **t**→**r** transition the intermediate and apical domains rotate counterclockwise as one body movement as we have shown in Chapter 6. Therefore, if one subunit is undergoing this transition, the intermediate and apical domains of its right-hand neighbor will rotate in the opposite (clockwise) direction. If the counterclockwise rotation closes the nucleotide binding pocket, which makes GroEL subunit have higher affinity for ATP, the clockwise rotation opens the pocket making ATP affinity lower.

During the simulation we have observed an attempt of the subunit G of closing of its nucleotide binding pocket, which immediately induced the opening of the nucleotide binding pocket of its right neighbor - subunit A. The left neighbor subunit F did not undergo any major changes in time of this event (Fig. 7.5).

Analyzing intersubunit interactions, we have observed switch of Glu386-Arg197

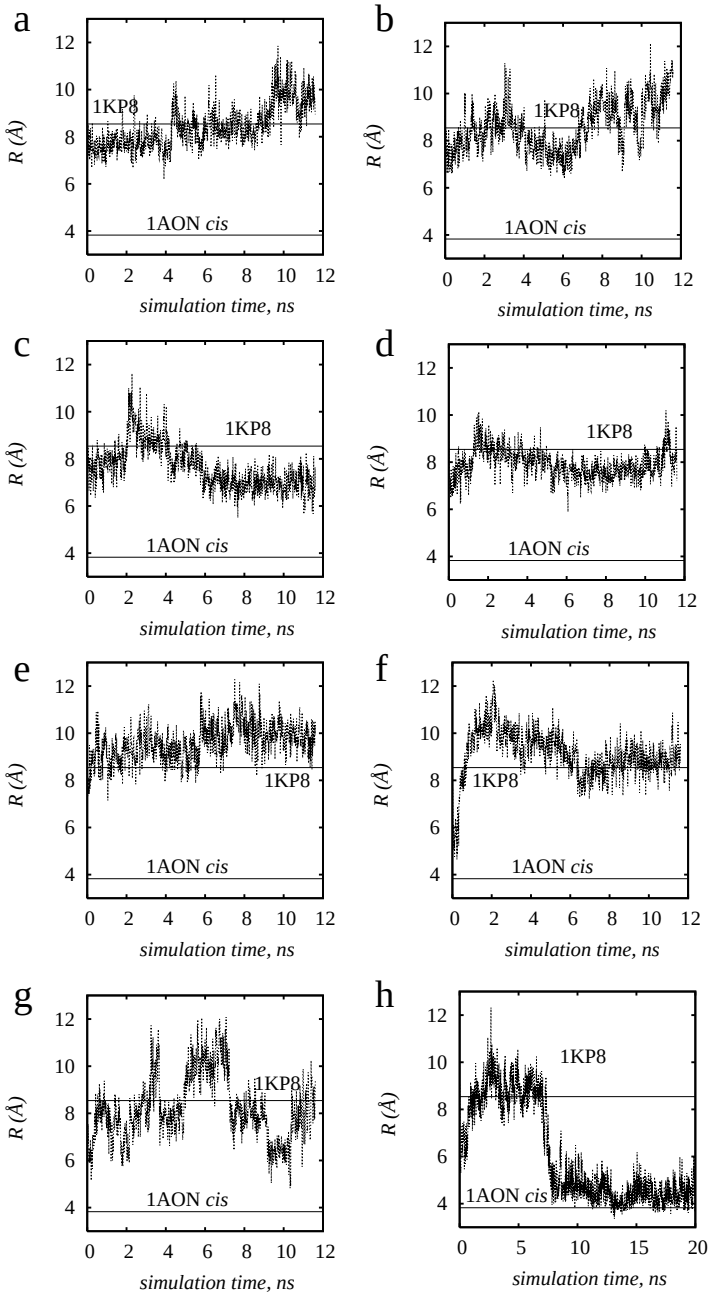


Figure 7.2: Interatomic distance traces between C_δ of Ile150 and $C_{\gamma 2}$ of Ile493 (a-g) from the MD trajectories for each subunit of GroEL ring from A to G , (h) from the MD trajectories for *holo* single subunit simulations. Horizontal lines denote measurements from reference structures indicated by their PDB accession codes (see text).

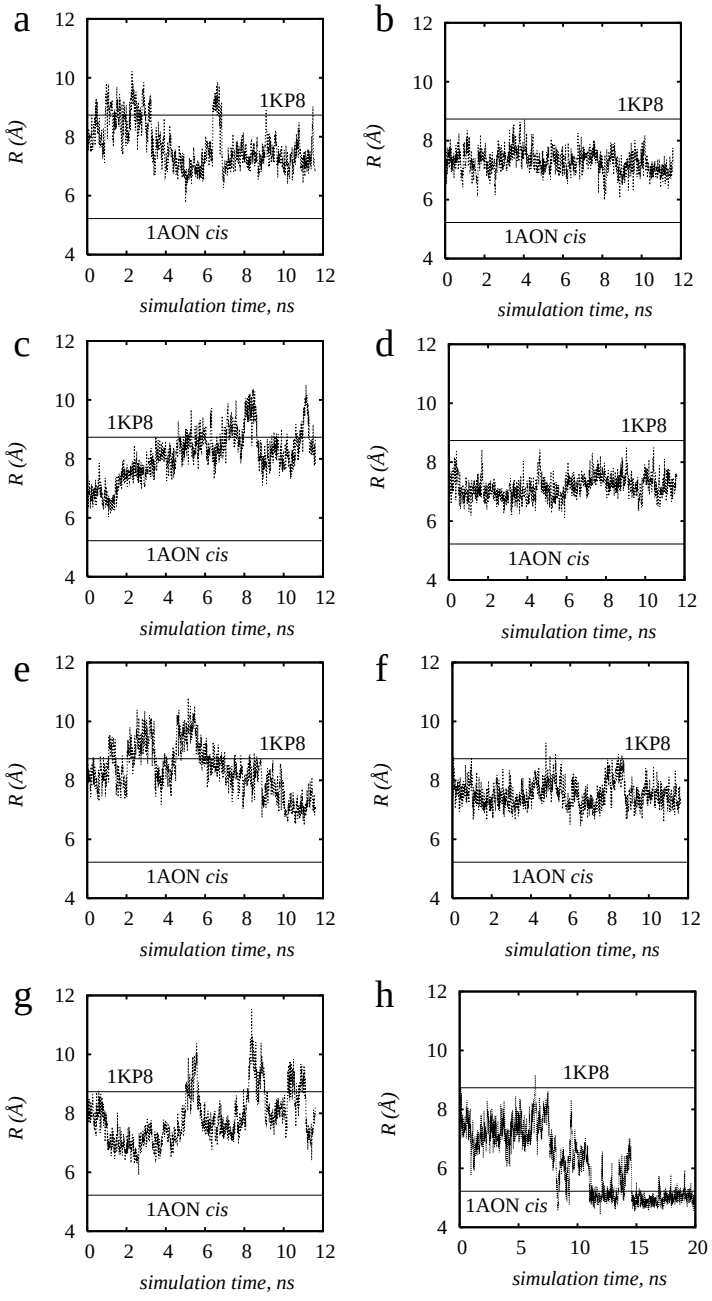


Figure 7.3: Interatomic distance traces between C_γ of Asp398 and the Mg^{2+} ion of KM-gATP (a-g) from the MD trajectories for each subunit of GroEL ring from A to G , (h) from the MD trajectories for *holo* single subunit simulations. Horizontal lines denote measurements from reference structures indicated by their PDB accession codes (see text).

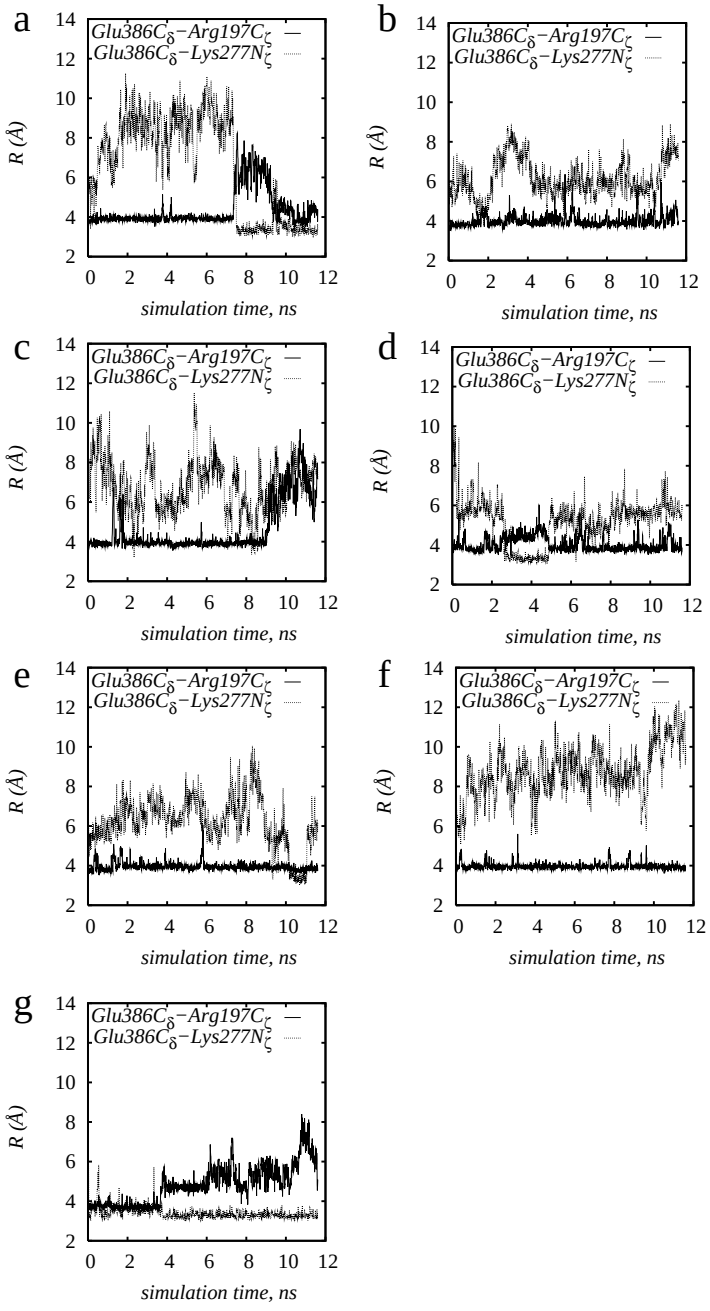


Figure 7.4: Formation and breakage of intersubunit salt bridges between Glu386-Arg197 and Glu386-Lys277 from the MD trajectories for each pair of subunits of GroEL ring: (a) A-B (b) B-C (c) C-D (d) D-E (e) E-F (f) F-G (g) G-A

to Glu386-Lys277 between two pairs of subunits G-A and A-B (Fig. 7.4). According to this observation, we can speculate that during **t**→**r** transition a contact mediating cooperativity among subunits could be the Glu386-Lys277 salt bridge.

This observation leads us to suggest a new mechanism for positive intraring cooperativity, schematically shown in Fig. 7.6. ATP weakly binds to one subunit of GroEL, denoted as central, making it undergo **t**→**r** transition. This event evokes the clockwise rotation of the intermediate and apical domains of the right-hand neighbor causing opening of its ATP binding pocket. As a result the central subunit strongly binds ATP and the right neighbor has the more room for the next molecule of ATP to bind, presumably exhibiting a lower barrier to binding allowing it to happen quickly. Therefore, binding of ATP facilitates counterclockwise through the ring, viewed transverse to the plane from outside the ring.

According to our results we propose that the $\mathbf{t}_7 \rightarrow \mathbf{r}_7$ transition is indeed concerted on the macroscale, may be actually sequential on the microscale: each following subunit in one ring undergoes **t**→**r** transition easier than preceding one. The microsequential transition is in agreement of Shiseki et al. [57] results showing that GroEL mutant having two cross-linked subunits (lacking possibility of **t**→**r** trasition) has ATPase activity with inhibited chaperonine activity due to impaired negative interring cooperativity.

All these findings have to be clarified by simulation of GroEL ring with a single bound ATP molecule. We defer the simulation of this system to a later work.

7.3.3 Conclusion

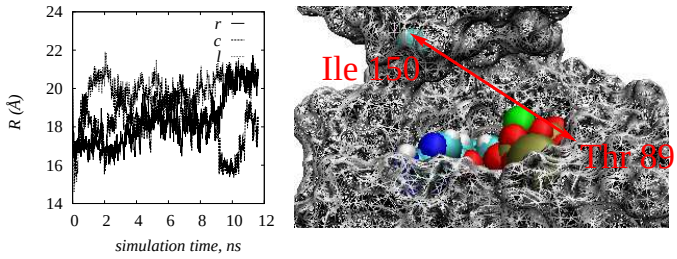


Figure 7.5: Closing the nucleotide binding pocket of one subunit induces opening of the nucleotide binding pocket of its right neighbor. Interatomic distance traces between C_α of Ile150 and Thr89 from the MD trajectories for subunits A, G and F of GroEL, where G is the central subunit, F is its left-hand and A is its right-hand neighbor

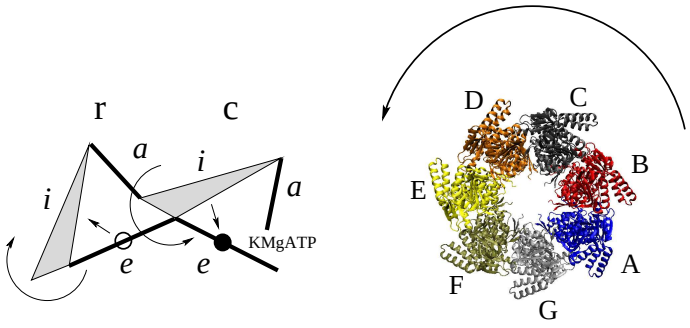


Figure 7.6: Schematic representation of the possible mechanism of intraring positive cooperativity. Small letter c and r denote central and right-hand neighbor subunits; e,i,a mark the equatorial, intermediate and apical domains.

The simulation of one ring of GroEL with seven bound ATP molecules clearly shows significant asymmetry among the subunits of GroEL ring. We are also have observed how counterclockwise rotation of the intermediate and apical domains of one subunit induces the clockwise rotation of its right hand neighbor, leading to the closing of the nucleotide binding pocket of one subunit and opening the pocket for another.

These observation make possible to propose that the $\mathbf{t}_7 \rightarrow \mathbf{r}_7$ transition being concerted on the macroscale, is sequential on the microscale, propogating counterclockwise through the ring.

Chapter 8: Findings and Future Directions

8.1 Findings

Several years of research have demonstrated that proteins require assistance of chaperonin systems for correct and fast folding into their native states. Chaperonins are large protein structures, possessing an inner cavity capable to accommodate a diverse range of substrate polypeptides. Chaperonins bind and encapsulate a non-native polypeptide chain, which folds to its native state isolated from the other protein chains of the crowded intracellular environment. The main goal of this work is a study of chaperonins action in protein folding using different computational approaches. In the first major part of this study we have modeled a chaperonin inner cavity as a spherical pore with repulsive wall and in the second part, we have examined the structure-functional relashionship of a chaperonin system considering detailed all-atom representation of the bacterial chaperonin GroEL.

We have shown that a substrate protein aims to adopt a compact form inside the chaperonin cavity due to confining restriction on the number of accessible extended conformations. With absence of any danger of aggregation inside the isolated chaperonin cavity, the reduction of the accessible volume leads to enhancing of association processes and therefore produce a shift of the equilibrium population distribution of the protein toward its native state. We have proved that confinement inside the chaperonin central cavity reduces the conformational entropy of the unfolded state

directly demonstrating that confinement changes the density of states of the substrate protein.

The stabilizing effect of the chaperonin cavity on a protein native state versus non-native open state is manifested by increase in melting temperature (temperature of native-nonnative transition) with a decreasing size of confinement. We have measured the increase in the melting temperature and proposed a scaling relationship for the change of the melting temperature with the size of confinement resulting in a better fit to the simulation data, compared to simpler scaling relationships considered previously.

We have shown that protein folding inside the chaperonin cavity accompanied with smoothing of the free energy barriers in account of confinement makes both unfolded and transition states of the protein more compact and therefore leads to progression of the substrate protein toward its native state.

We have investigated the conformational changes of an isolated subunit of the bacterial chaperonin GroEL. We are the first to demonstrate that two conformational transitions in the isolated GroEL subunit defined as $\mathbf{t}\rightarrow\mathbf{r}$ and $\mathbf{t}\rightarrow\mathbf{a}$ occur naturally as a result of presence or absence of KMgATP in the binding pocket respectively. We have described three states of the GroEL subunit: a single state of the unliganded subunit defined as **a** state and two possible states of ATP-bound subunit: **t** and **r**. We have examined the structure of the nucleotide binding pocket in details, demonstrating that bound KMgATP stabilizes hydrophobic and polar interactions between residues of the intermediate and equatorial domains. We have observed the counterclockwise apical domain rotation during $\mathbf{t}\rightarrow\mathbf{r}$ transition, which agrees with cryoelectron microscopy

evidence.

We have demonstrated there are no intersubunit steric clashes arise during the **T→R** transition, which indicates a new direction in pursuing a theoretical understanding of intraring positive cooperativity. We have also postulated a testable mechanism of interring negative cooperativity in ATP binding.

Quasi-harmonic analysis clearly shows that natural modes of vibrations drive conformational transitions, revealing a single particular mode responsible for each of **t↔r** and **a↔ t** transitions. This finding is in agreement with theoretical studies considering the conformational transition of the chaperonins being a result of normal range of the thermal motion, where maximum extent of the motion is stabilized by the bound nucleotide.

All these findings have clarified some important aspects of the mechanisms of chaperonin machinery and demonstrated the necessity of chaperonin assistance in the protein folding. Our results support the computational approach in studying of biomolecules as a valuable addition to the experimental investigations.

8.2 Future Directions

We propose the following suggestions for developing the knowledge of the mechanism of chaperonin action in protein folding.

We have modeled the chaperonin cavity as a spherical pore with repulsive walls considering only spacial effect of confining. This approach may be altered considering the cavity which possesses selectively attractive or repulsive features depending on the polar or hydrophobic nature of the substrate protein atoms. The new proposed

model of chaperonin central cavity will demonstrate not only influence of mechanical confinement, but how the hydrophilicity of the cavity lining leads to burial of hydrophobic and exposure of hydrophilic surfaces of protein [31].

We have studied dynamics of the conformational changes of a single subunit and one ring of the bacterial chaperonin GroEL upon ATP binding defined as the allosteric **T**→**R** transition, concentrating on the positive and negative cooperativity of the transition. The next step may be simulation of two subunits belonging to the different rings of GroEL, which interact via inter-ring surface. Three choices of the simulated system may be considered: both rings have the nucleotide binding pockets empty, one ring has its binding pocket occupied by ATP and two rings have ATP-bound binding pockets. This approach will be helpful to clarify the mechanism of negative cooperativity.

Experimental evidence has demonstrated that binding of ATP alone to GroEL reduces affinity for non-native substrate with comparison with ATP-free ring [18, 19, 20, 21]. In order to build the knowledge about the allosteric signaling between two binding pockets of GroEL (substrate protein and nucleotide), we suggest including the substrate protein in the simulated system. Free energy measurements of ligand binding have to be done at this step.

We have investigated only **T**→**R** transition of GroEL chaperonin, which is only one part of the chaperonin machinery. To understand the detailed picture of GroEL reaction cycle, we consider the simulation of the whole chaperonin system will be the best solution.

Bibliography

[1] C. Bartolucci, D. Lamba, S. Grazulis, E. Manakova, and H. Heumann. Crystal structure of wild-type chaperonin GroEL. *J. Mol. Biol.*, 354:940–951, 2005.

[2] Z. Xu, A. L. Horwich, and P. B. Sigler. The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. *Nature*, 388:741–750, 1997.

[3] J. Wang and D. C. Boisvert. Structural basis for GroEL-assisted protein folding from the crystal structure of $(GroEL\text{-}KMgATP)_{14}$ at 2.0 Å resolution. *J. Mol. Biol.*, 327:843–855, 2003.

[4] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[5] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23:566–579, 1995.

[6] H. S. Rye, S. G. Burston, W. A. Fenton, J. M. Beechem, Z. Xu, P. B. Sigler, and A. L. Horwich. Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. *Nature*, 388:792–798, 1997.

[7] H. S. Rye, A. M. Rosenman, S. Chen, K. Furtak, W. A. Fenton, H. R. Saibil, and A. L. Horwitch. GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell*, 97:325–338, 1999.

[8] K. A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501–1509, 1985.

[9] A. P. Minton. Confinement as a determinant of macromolecular structure and reactivity. *Biophys. J.*, 63:1090–1100, 1992.

[10] A. P. Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.*, 276:10577–10580, 2001.

[11] H. X. Zhou and K. A. Dill. Stabilization of proteins in confined spaces. *Biochemistry*, 40:11289–11293, 2001.

[12] D. Thirumalai, D. K. Klimov, and G. H. Lorimer. Caging helps proteins fold. *Proc. Natl. Acad. Sci. USA*, 100:11195–11197, 2003.

[13] 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*, 107:223–233, 2001.

[14] F. U. Hartl and M. Hayer-Hartl. Molecular chaperones in cytosol: from nascent chain to folded protein. *Science*, 295:1852–1858, 2002.

[15] A. L. Horwich, K. B. Low, W. A. Fenton, I. N. Hirshfield, and K. Furtak. Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. *Cell*, 74:909–917, 1993.

[16] K. L. Ewalt, J. P. Hendrick, W. A. Houry, and F. U. Hartl. In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell*, 90:491–500, 1997.

[17] W. A. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, and F. U. Hartl. Identification of *in vivo* substrates of the chaperonin GroEL. *Nature*, 402:147–154, 1999.

[18] O. Yifrach and A. Horovitz. Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry*, 34:5303–5308, 1995.

[19] O. Yifrach and A. Horovitz. Allosteric control by ATP of non-folded protein binding to GroEL. *J. Mol. Biol.*, 255:356–361, 1996.

[20] R. A. Staniforth, S. G. Burston, T. Atkinson, and A. R. Clarke. Affinity of chaperonin-60 for a protein substrate and its modulation by nucleotides and chaperonin-10. *Biochem J.*, 300:651–658, 1994.

[21] N. A. Ranson, G. W. Farr, A. M. Rosenman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, 107:869–879, 2001.

[22] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[23] V. Grantcharova, E. J. Alm, D. Baker, and A. L. Horwich. Mechanisms of protein folding. *Curr. Opin. Struct. Biol.*, 11:70–82, 2001.

[24] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133–7155, 1990.

[25] S. Albek, R. Unger, and G. Schreiber. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J. Mol. Biol.*, 298:503–520, 2000.

[26] F. B. Sheinerman and B. Honig. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.*, 318:161–177, 2002.

[27] R. Jaenicke. Protein self-organization in *in vitro* and *in vivo*: partitioning between physical biochemistry and cell biology. *Biol. Chem.*, 379:237–243, 1998.

[28] A. P. Minton. Excluded volume as determinant of macromolecular structure and reactivity. *Biopolymers*, 20:2093–2120, 1981.

[29] R. J. Ellis and A. P. Minton. Protein aggregation in crowded environments. *Biol. Chem.*, 387:485–497, 2006.

[30] R. J. Ellis. Molecular chaperones: Inside and outside the Anfinsen cage. *Curr. Biol.*, 11:R1038–R1040, 2001.

[31] W. A. Fenton and A. L. Horwich. Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q. Rev. Biophys.*, 36:229–256, 2003.

[32] R. J. Ellis. Roles of molecular chaperones in protein folding. *Curr. Biol.*, 4:117–122, 1994.

[33] J. Frydman. Folding on newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.*, 70:603–647, 2001.

[34] W. A. Fenton and A. L. Horwich. GroEL-mediated protein folding. *Protein Sci.*, 6:743–760, 1997.

[35] B. Bukau and A. L. Horwich. The Hsp70 and Hsp60 chaperone machines. *Cell*, 92:351–366, 1998.

[36] D. L. Stone, A. Slavotinek, G. G. Bouffard, S. Banerjee-Basu, A. D. Baxevanis, M. Barr, and L. G. Biesecker. Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nat. Genet.*, 25:79–82, 2000.

[37] A. M. Slavotinek, E. M. Stone, K. Mykytyn, J. R. Heckenlively, J. S. Green, E. Heon, M. A. Musarella, P. S. Parfrey, V. C. Sheffield, and L. G. Biesecker. Mutations in MKKS cause Bardet-Bieldl syndrome. *Nat. Genet.*, 26:15–16, 2000.

[38] A. M. Slavotinek and L. G. Biesecker. Unfolding the role of chaperones and chaperonins in human disease. *Trends Genet.*, 17:528–535, 2001.

[39] P. B. Sigler, Z. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich. Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.*, 67:581–608, 1998.

[40] A. L. Horwich, F. W. Farr, and W. A. Fenton. GroEL-GroES-mediated protein folding. *Chem. Rev.*, 106:1917–1930, 2006.

[41] R. J. Ellis and F. U. Hartl. Principles of protein folding in the cellular enviroment. *Curr. Opin. Struct. Biol.*, 9:102–110, 1999.

[42] T. Hohn, B. Hohn, A. Engel, M. Wurtz, and P. R. Smith. Isolation and characterization of the host protein groE involved in bacteriophage lambda assembly. *J. Mol. Biol.*, 129:359–373, 1979.

[43] R. W. Hendrix. Purification and properties of groE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.*, 129:375–392, 1979.

[44] K. Braig, Z. Otwinovski, R. Hedge, D. C. Boisvert, A. Joachimiak, A. L. Horwich, and P. B. Sigler. The crystal structure of bacterial chaperonin GroEL at 2.8 Å. *Nature*, 371:578–586, 1994.

[45] F. Motojima, C. Chaudhry, W. A. Fenton, G. W. Farr, and A. L. Horwich. Substrate polypetide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl. Acad. Sci. USA*, 101:15005–15012, 2004.

[46] W. A. Fenton, Y. Kashi, K. Furtak, and A. L. Horwich. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, 371:614–619, 1994.

[47] J. Wang and L. Chen. Domain motions in GroEL upon binding of an oligopeptide. *J. Mol. Biol.*, 334:489–499, 2003.

[48] S. Falke, F. Tama, C. L. Brooks , III, E. P. Gogol, and M. T. Fisher. The 13 Å structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.*, 348:219–230, 2005.

[49] L. Chen and P. B. Sigler. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell*, 99:757–768, 1999.

[50] O. Yifrach and A. Horovitz. Two lines of allosteric communication in the oligomeric chaperonin GroEL are revealed by the single mutation ARG196→ALA. *J. Mol. Biol*, 243:397–401, 1994.

[51] A. Horowitz and K. R. Willison. Allosteric regulation of chaperonins. *Curr. Opin. Struct. Biol.*, 15:646–651, 2005.

[52] E. S. Bochkareva, N. M. Lissin, G. C. Flynn, J. E. Rothman, and A. S. Girshovich. Positive cooperativity in the functioning of molecular chaperone GroEL. *J. Biol. Chem.*, 267:6796–6800, 1992.

[53] W. A. Fenton, J. S. Weissman, and A. L. Horwich. Putting a lid on protein folding: structure and function of the co-chaperonin, GroES. *Chem. Biol.*, 3:157–161, 1996.

[54] H. R. Saibil and N. A. Ranson. The chaperonin folding machine. *Trends Biochem. Sci.*, 27:627–632, 2002.

[55] J. Ma and M. Karplus. The allosteric mechanism of the chaperonin GroEL: A dynamic analysis. *Proc. Natl. Acad. Sci. USA*, 95:8502–8507, 1998.

[56] J. Ma, P. B. Sigler, Z. Xu, and M. Karplus. A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.*, 302:303–313, 2000.

[57] K. Shiseki, N. Murai, F. Motojima, T. Hisabori, M. Yoshida, and H. Taguchi. Synchronized domain opening motion of GroEL is essential for communication between the two rings. *J Biol Chem.*, 276:11335–11338, 2001.

[58] J. Schlitter, M. Engels, P. Kruger, E. Jacoby, and W. Wollmer. Targeted molecular dynamics simulation of conformational change - Application to the T-R transition in insulin. *Mol. Sim.*, 10:291–308, 1993.

[59] C. Hyeon, G. H. Lorimer, and D. Thirumalai. Dynamics of allosteric transitions in GroEL. *Proc. Natl. Acad. Sci. USA*, 103:18939–18944, 2006.

[60] D. K. Eggers and J. S. Valentine. Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Sci.*, 10:250–261, 2000.

[61] D. Bolis, A. S. Politou, G. Kelly, A. Pastore, and P. A. Temussi. Protein stability in nanocages: a novel approach for influencing protein stability by molecular confinement. *J. Mol. Biol.*, 336:203–212, 2004.

[62] Z. Lin and H. S. Rye. Expansion and compression of a protein folding intermediate by GroEL. *Mol. Cell*, 16:23–34, 2004.

[63] R. J. Ellis. Protein folding: Inside the cage. *Nature*, 442:360–362, 2006.

[64] Y.-C. Tang, H.-C.Chang, A. Roeben, D. Wischnewski, N. Wischnewski, M. J. Kerner, F. U. Hartl, and M. Hayer-Hartl. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell*, 125:903–914, 2006.

[65] D. K. Klimov, D. Newfield, and D. Thirumalai. Simulations of β-hairpin folding confined to spherical pores using distributed computing. *PNAS*, 99:8019–8024, 2002.

[66] F. Takagi, N. Koga, and S. Takada. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proc. Natl. Acad. Sci. USA*, 100:11367–11372, 2003.

[67] G. Ping, J. M. Yuan, M. Vallieres, H. Dong, Z. Sun, Y. Wei, F. Y. Li, and S. H. Lin. Effects confinement on protein folding and protein stability. *J. Chem. Phys.*, 118:8042–8048, 2003.

[68] M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347:631–639, 1990.

[69] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Applications to small water clusters. *J. Chem. Phys.*, 76(1):637–649, 1982.

[70] K. U. Sanbonmatsu and C.-S. Tung. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *J. Struct Biol.*, 157:470–480, 2007.

[71] B. Mehlig, D. W. Heermann, and B. M. Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B*, 45:679–685, 1992.

[72] N. Rathore, T. A. Knotts IV, and J. J. de Pablo. Density of states simulations of proteins. *J. Chem. Phys.*, 118(9):4285–4290, 2003.

[73] G. S. Grest and K. Kremer. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A*, 33(5):3628–3631, 1986.

[74] T. Soddemann, B. Dunweg, and K. Kremer. Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. *Phys. Rev. E*, 68(046702), 2003.

[75] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050, 2001.

[76] N. Rathore, Q. Yan, and J. J. de Pablo. Molecular simulation of the reversible mechanical unfolding of proteins. *J. Chem. Phys.*, 120:5781–5788, 2004.

[77] R. M. Levy and M. Karplus. Vibrational approach to the dynamics of an α-helix. *Biopolymers*, 18:2465–2495, 1979.

[78] B. R. Brooks, D. Janezik, and M. Karplus. Harmonic analisys of large systems. i. Methodology. *J. Comput. Chem.*, 16:1522–1542, 1995.

[79] P. Durand and Y.-H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Prot. Engineering*, 14:1–6, 2001.

[80] R. M. Levy, D. Perahia, and M. Karplus. Molecular dynamics of an α-helical polypeptide: Temperature dependence and deviation from harmonic behavior. *Proc. Natl. Acad. Sci. USA*, 79:1346–1350, 1982.

[81] R. M. Levy, A. R. Srinivasan, W. K. Olson, and J. A. McCammon. Quasi-harmonic method for studying very low-frequency modes in proteins. *Biopolymers*, 23:1099–1112, 1984.

[82] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack , Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher , III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[83] A. D. MacKerell, Jr. and N. K. Banavali. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.*, 21:105–120, 2000.

[84] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54(12):5237–5247, 1971.

[85] O. V. Borisov and A. Halperin. Micelles of polysoaps. *Langmuir*, 11:2911–2919, 1995.

[86] A. Laschewsky. Molecular concepts, self-organisation and properties of polysoaps. *Adv. Polym. Sci.*, 124:1–86, 1995.

[87] R. Chang and J. Yethiraj A. Solvent effects on the collapse dynamics of polymers. *J. Chem. Phys.*, 114:7688–7699, 2001.

[88] J. M. Polson and M. J. Zuckermann. Simulation of short-chain polymer collapse with an explicit solvent. *J. Chem. Phys.*, 116:7244–7254, 2002.

[89] C. F. Abrams, N. Lee, , and S. Obukhov. Collapse dynamics of a polymer chain: Theory and simulation. *Europhys. Lett.*, 59:391–397, 2002.

[90] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.*, 92(8):5057–5086, 1990.

[91] D. Frenkel and B. Smit. *Understanding molecular simulation: From algorithms to applications*. Academic Press, New York, 1996.

[92] B. Dunweg and K. Kremer. Molecular-dynamics simulation of a polymer-chain in solution. *J. Chem. Phys.*, 99:6983–6997, 1993.

[93] C. F. Abrams and K. Kremer. The effect of bond length on the structure of dense bead-spring polymer melts. *J. Chem. Phys.*, 115:2776–2785, 2001.

[94] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos. Generalization of the wang-landau method for off-lattice simulations. *Phys. Rev. Lett.E*, 66:056703–1–056703–9, 2002.

[95] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.*, 151:283–312, 1999.

[96] N. A. Ranson, D. K. Clare, G. W. Farr, D. Houldershaw, A. L. Horwich, and H. R. Saibil. Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nat. Struct. Mol. Biol.*, 13:147–152, 2006.

[97] D. A. Pearlman, D. A. Case, J. W. Caldwell, W.S. Ross, III T.E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. 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. *Comp. Phys. Commun.*, 91:1–41, 1995.

[98] O. Danziger, D. Rivenzon-Segal, S. G. Wolf, and A. Horowitz. Conversion of the allosteric transition of groel from cocerted to sequential by the single mutation Asp-155 → Ala. *Proc. Natl. Acad. Sci. USA*, 100:13797–13802, 2003.

[99] O. Yifrach and A. Horovitz. Mapping the transition state of the allosteric pathway of GroEL by protein engineering. *J. Am. Chem. Soc.*, 120:13262–13263, 1998.

[100] A. Horovitz and O. Yifrach. On the relationship between the Hill coefficients for the steady-state and transient kinetic data: a criterion for concerted transition in allosteric proteins. *Bul. Math. Biol.*, 62:241–246, 2000.

Vita

Yelena R Sliozberg was born in Moscow, Russia on October 1,1971. She attended D. Mendeleev University of Chemical Technology of Russia, Moscow, Russia and received the Bachelor and Master of Technology degree in Chemical Engineering with a minor in Cybernetics of Chemical Technological Processes in 1994. Yelena worked as a research assistant in Research Institute of Textile Restoration, Moscow, Russia, from 1994 to 1997. Yelena joined the Ph.D. program in Chemical Engineering at Drexel University in Fall 2001, where she worked as a graduate student under Dr. Cameron F Abrams. In 2002 and 2004, Yelena was awarded a Mollye and Mitchel Glick scholarship in recognition of research and other graduate work. During her graduate study, Yelena was a teacher assistant for two semesters (Fluid Mechanics; Instr. Charles Weinberger and Gennaro Maffia).