The Dissertation Committee for Serdal Kirmizialtin Certifies that this is the approved version of the following dissertation:

Computer Simulations of Protein Translocation and Stretching

Committee:

Dmitrii E. Makarov, Supervisor

Peter J. Rossky

Venkat Ganesan

Rick Russel

Graeme Henkelman

Computer Simulations of Protein Translocation and Stretching

by

Serdal Kirmizialtin, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2007

Dedication

to my love

Acknowledgements

I would like to thank my thesis supervisor Dr. Dmitrii Makarov for his kind support, remarkable intuition, enthusiasm and inspiration; without him, this thesis would not be possible. Thank you for introducing me to the exciting field of biophysics and always being welcoming and accessible to discuss science with me.

Thank you to my examiners, Professor Peter Rossky, Dr. Graeme Henkelman, Dr. Rick Russell, and Dr. Venkat Ganesan for being on my thesis committee and reading my whole thesis despite your extremely busy schedules. I would like to specifically thank Professor Peter Rossky for being an endless support and encourager during my study. Especially, I would like to express my gratitude for assisting me in finding a postdoctoral position.

Many faculty members and students have contributed to my doctoral study, however, I take this opportunity to thank Dr. Venkat Ganesan, Dr. Liviu Movileanu and Lei Huang for their fruitful collaborations and for the great discussions that we had during the past years. I must also thank Dr. Gisela Kramer for proofreading some of my chapters and Sky Pike for his valuable friendship during the past years.

I am forever indebted to my parents for their support, understanding, and the endless patience that they showed during my education.

My final words go to my wife, Suphan, who endured this long process with me offering dedication, care, and support. Thank you for always being there for me with such a gigantic love…

Computer Simulations of Protein Translocation and Stretching

Publication No._____

Serdal Kirmizialtin, PhD

The University of Texas at Austin, 2007

Supervisor: Dmitrii E. Makarov

Many biomolecular processes involve mechanical force-induced reactions in the cell, such as translocation, and mechanical stretching of biopolymers. Recent advances in single molecule manipulation techniques make it possible to apply mechanical force to individual biomolecules and study their dynamics. To gain molecular level understanding of these processes and to interpret the single-molecule experiments, we used Langevin dynamics simulations of coarse-grained biopolymer models.

Our result show that the mechanism of translocation of proteins through pores depends on the pore diameter, on the magnitude of the pulling force and on whether the force is applied at the N- or the C-terminus of the chain. In addition, the translocation kinetics of peptides varies with their stability. The mechanism of protein translocation is found to be different from that of a structureless polypeptide of the same length. We further showed that unfolding mechanism of translocation process is different from when the same protein is stretched between its C- and N-termini.

We also studied the mechanical and chemical/thermal denaturation of proteins. We observed that the free energy profile along the mechanical reaction coordinate and the chemical reaction coordinate are different. In our protein model, the mechanical and chemical/thermal denaturation cannot be simply explained in terms of a simple one-dimensional free energy landscape. We further analyzed the spontaneous folding and

refolding under a constant force and found that refolding generally occurs via different mechanisms. Similarly, we investigated the protein unfolding/refolding under the applied force that varies with a constant loading rate. This study shows that unfolding/refolding pathways are generally similar for low loading/unloading rates while they become different for high loading/unloading rates.

Finally, we studied the dynamics of molecular friction knots formed by a pair of polymer strands. We examined different knot types, and different polymer sequences. Depending on the knot type and the nature of the polymer, we observed two different behaviors when the force F is exerted to separate the polymer strands. The knot between polymer strands can be strong (the time τ the knot stays tied increases with the force F applied to separate the strands) or weak (τ decreases with increasing F).

Table of Contents

List of Figures

Chapter 1

1. Introduction

1.1. PROTEIN TRANSLOCATION AND STRETCHING

Many cellular processes in biology involve translocation of biomolecules. Nucleic acid and protein translocation across channels or membranes, gene swapping through bacterial pili[1], translocation of nascent proteins inside the ribosomal tunnel [2-5], gene transfer by transduction [6], protein translocation across organelle [5, 7-9] and protein degradation by ATP- depended proteases[7, 10, 11] are only a few examples.

Most proteins are synthesized in the cytoplasm from nuclear encoded mRNAs in a process called translation, see Fig 1.1. mRNA contains the information and consists of codons that will encode specific amino acids. Eukaryotic translation begins when the small subunit of the ribosome attaches to the methylated cap found at the 5'end of the mRNA, eventually forming an initiation complex after the large ribosomal subunit joined. Polypeptide chain elongation starts from this complex. Another molecule, tRNA, that contains the anticodon complementary to the codon on the mRNA binds to the mRNA. The first tRNA occupies the P site and the second tRNA that is complementary to the second codon enters to the complex and occupies the A site. The amino acid on the P site of the tRNA is transferred to the A site amino acid and the first tRNA exits from the complex. This process continues as the ribosome moves along the mRNA until the stop codon is encountered. The nascent peptide that is synthesized this way moves through the ribosomal tunnel running through the large ribosomal subunit [3, 12-14].The crystal structure of *H. marismortuio* 50S ribosomal subunit has revealed that the length of the tunnel is ~100 Å and its width *D* varies between 10 and 28 Å at different points along the tunnel[12]. This tunnel diameter is too narrow to accommodate any three-dimensional protein structure with an exception of alpha-helix[4] in its narrowest constriction. The tunnel diameters in other ribosomal systems are assumed to be similar. After exiting the ribosomal tunnel, folding takes place for some polypeptides but chaperone activity is observed outside the cytosol for others [2, 4, 15, 16]; implying that the polypeptides transverse the ribosome as unfolded or partially folded molecules. It is also expected that

the mechanism by which protein folds in the course of translocation would be different than the spontaneous folding [17].

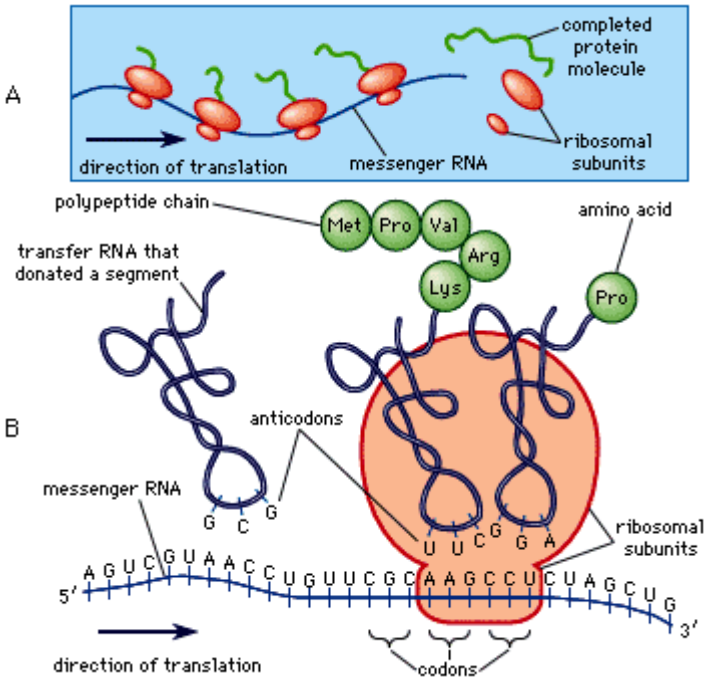

Fig. 1.1A-B. Polypeptide synthesis on ribosome (adapted from reference [18]).

Some polypeptides that exit the ribosome must be imported into the destined organelles such as mitochondria, endoplasmic reticulum and chloroplast, see Fig. 1.2. Proteins destined for transport into an organelle contain a signal sequence. This sequence binds to the passenger protein (chaperone) with ATP without affecting its folding[19]. It rather acts as a targeting mechanism to ensure that the protein is delivered to the proper organelle. Translocation takes place across the membrane channels with diameters that are too small to accommodate even a single domain protein in its folded state[7]. Thus, the protein must unfold. In the case of mitochondrial translocation, unfolding of the protein is induced by an electrical potential across the inner membrane because the targeting sequence and the membrane have opposite net charges[18]. A chaperone protein called mtHsp70 found in the inner membrane of mitochondria binds to the protein to be translocated and pulls it inside the membrane as it unfolds. The observed unfolding times in this translocation process are hundreds of times faster than spontaneous unfolding times. How does the mtHsp70 unfolding machinery works, how much force it generates and what is the mechanism by which proteins unfold during translocation are some of

the questions that have been studied both experimentally and theoretically [7, 10, 11, 20-22]. In particular, single molecule techniques and computer simulations have provided a wealth of information about these processes.

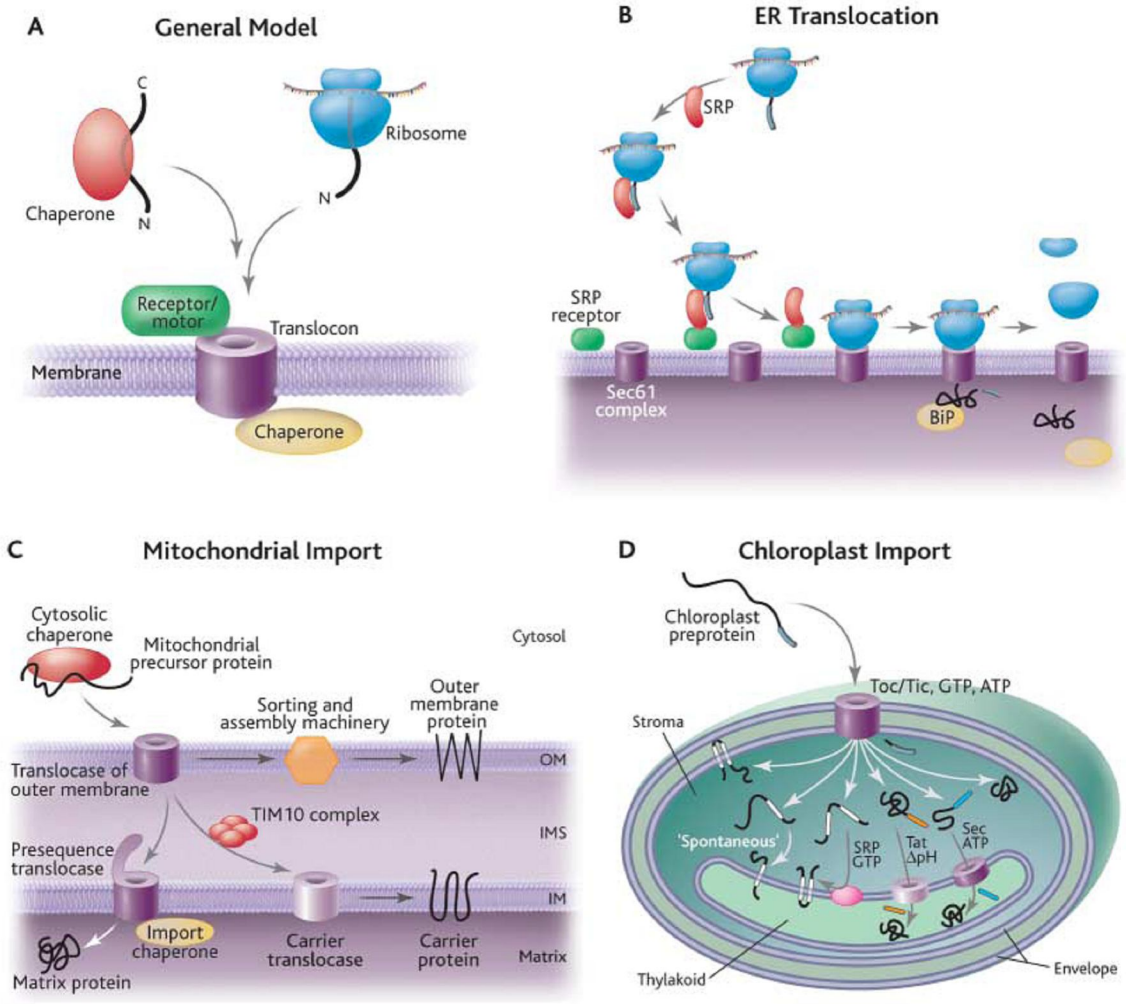

Fig. 1.2. Translocation mechanisms and pathways: (A) General model of translocation (B) Cotranslational translocation into the endoplasmic reticulum (C) Mitochondrial protein import OM, outer membrane, IMS, intermembrane space, IM, inner membrane. (D) Chloroplast protein import (adapted from reference [23]).

Protein translocation is an example of an *in vivo* process that involves a mechanical force. Many other examples exist. Biomolecules are often anchored in two or more locations and they are exposed to mechanical stretching that is due to tissue strain or sometimes is exerted by molecular motors. The application of a mechanical force induces conformational transitions and regulates or changes the function of biomolecules such as polysaccahrides, DNA and proteins[24-28]. For example, the extracellular matrix

(ECM) protein fibronectin contains tandem arrays of fibronectin type III domains. Experimental evidence suggests that[29] the force, transduced from the cell's actin cytoskeleton to the ECM, unfolds some of the domains of fibronectin type III domains. The resulting structural changes of the RGD loop of the domains regulate the binding between fibronectin and integrin[27, 28, 30, 31] . Similiarly, integrins change their conformations under force[32] . Application of force, induces conformational transitions and by this way their binding site to the ligands changes; and so does the interaction between the protein and the ligand.

There are also a number of proteins that perform "load-bearing" functions in living organisms. Example include titin [33-37], tenascin [29], spectrin [38, 39], and spider silk proteins[40-42]. These proteins often exhibit a remarkable combination of high elasticity and mechanical strength. Therefore, studying how the mechanisms through which proteins respond to mechanical forces depend on their structure is likely to provide new insights into their biological function as well as lead to discovery of new biologically inspired materials.

1.2. SINGLE MOLECULE MANIPULATION TECHNIQUES

Single molecule experiments have become an increasingly popular tool for studying biomolecular phenomena, and in particular, mechanical processes involving biomolecules. Examples of biological processes that have been studied through single-molecule techniques include but are not limited to DNA stretching, translocation of RNA polymerase, myosin VI walking, protein translocation, protein unfolding and folding under mechanical force, RNA folding and unfolding, as well as enzyme activity (reviewed in ref. [43-45]). Single molecule experiments (SMEs) allow one to access biomolecular processes by tracking individual trajectories in real time. Traditional bulk methods, on the other hand, only give ensemble averages, which may or may not reflect the characteristics of any individual trajectory. Additionally, biochemical processes are often complex and occur through multiple steps, alternative pathways, and multiple intermediates; their presence is often difficult to infer from bulk experiments.

Single-molecule techniques also provide a direct link to time-dependent computer simulations of individual molecules, since the latter are inherently performed at the single-molecule level

Three major experimental approaches have been developed for the investigation of protein translocation and folding/unfolding of proteins at the single-molecule level: (1) force spectroscopy, which is a direct application of force by using atomic force microscope (AFM), optical or magnetic tweezers; (2) electrophoresis which involves the movement of electrically charged molecules under the influence of an electric field; (3) single-molecule fluorescence spectroscopy, which involves the analysis of the fluorescence emission from individual, fluorescently labeled molecules. Particular attention in this thesis will be paid to the electrophoresis and force spectroscopy studies on biopoloymers.

1.2.1. Atomic Force Microscopy (AFM)

AFM was invented to image the topography of metallic surfaces with an Angstrom resolution, yet it is widely applied in biology to exert and manipulate mechanical forces on biomoleculer systems. AFM is composed of a piezomotor surface, a cantilever and a laser photodetector, see Fig 1.3A. The surface and the tip of the cantilever are coated with molecules that can bind to the biopolymer to be stretched. As a result, two ends of the biopolymer are attached to the AFM. When the piezomotor surface is moved away from the cantilever with a constant velocity, the molecule is stretched. If the biopolymer resists to conformational changes, the cantilever tip retracts and the deflection of the tip is measured by the photodetector. Deviation of the cantilever away from equilibrium dx is used to calculate the force acting on the molecule by $F = k_s dx$ where k_s is the stiffness of the cantilever and is typically in the range of 10-1000 pN/nm. This high stiffness makes spatial and force resolution to be poor in AFM experiments. This factor limits the use of AFM to study systems with strong interactions only [45]. Fig. 1.3B-C shows force extension curves of poly-protein I27 domains. As the distance between chain ends increases, the force increases due to the deflection on the cantilever. Upon the unfolding of a domain, restoring force drops down, generating a saw-tooth pattern on the force extension curves.

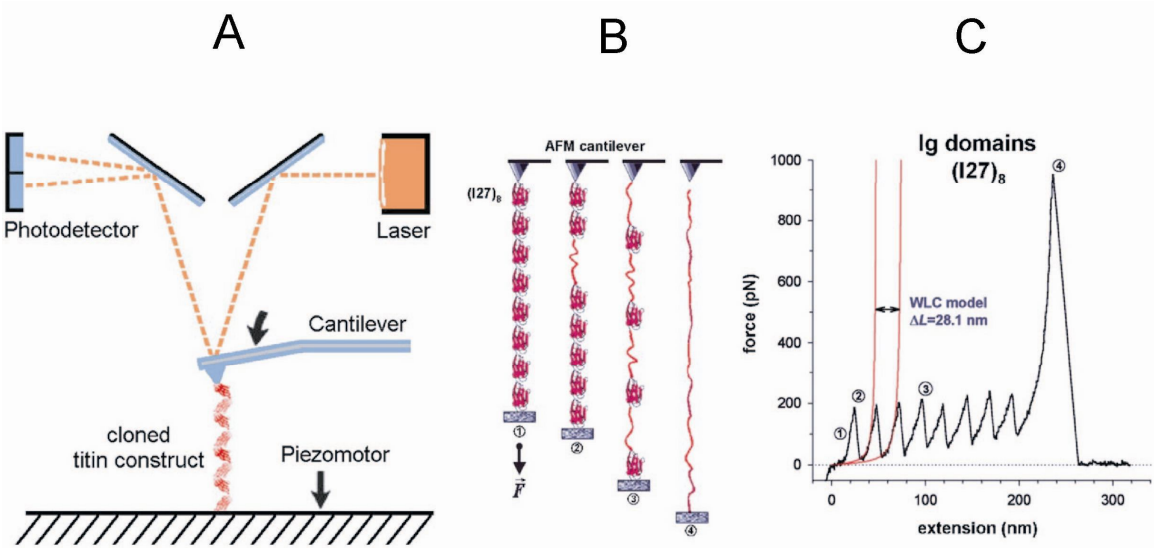

Fig. 1.3. (A) Schematic representation of an AFM set up, (B) Stretching of I27 domains, (C) Force extension curves measured by deflection of cantilever. Red solid lines are the fits to Worm like Chain Model (WLC) (adopted from reference [46]).

1.2.2. Laser Optical Tweezer (LOT)

Laser optical tweezer utilizes radiation pressure to exert mechanical force (see Fig. 1.4(A)). In the LOT experiments, the biopolymer is attached to DNA handles from the two ends. These handles are attached to beads such as polystyrene. The first bead is attached to a pipet tip which is moved with constant velocity by a piezoelectric actuator. The second bead is trapped by a focused laser beam. The displacement of the second bead from the trap center gives the effect of external force whose magnitude is linearly proportional to the magnitude of the bead displacement. Thus, conceptually LOT is identical to AFM with the major difference that the effective spring constant in LOT is in the range of 0.001-0.1pN/nm. This makes LOT very precise in weakly interacting systems but not for strong ones, since at high force LOT becomes insensitive to the change in extension [45]. Instead of pulling the substrate relative to a flexible force sensor, recent studies introduce a new technique called force clamp[47, 48]. In this technique, the force acting on the protein is controlled while extension is measured.

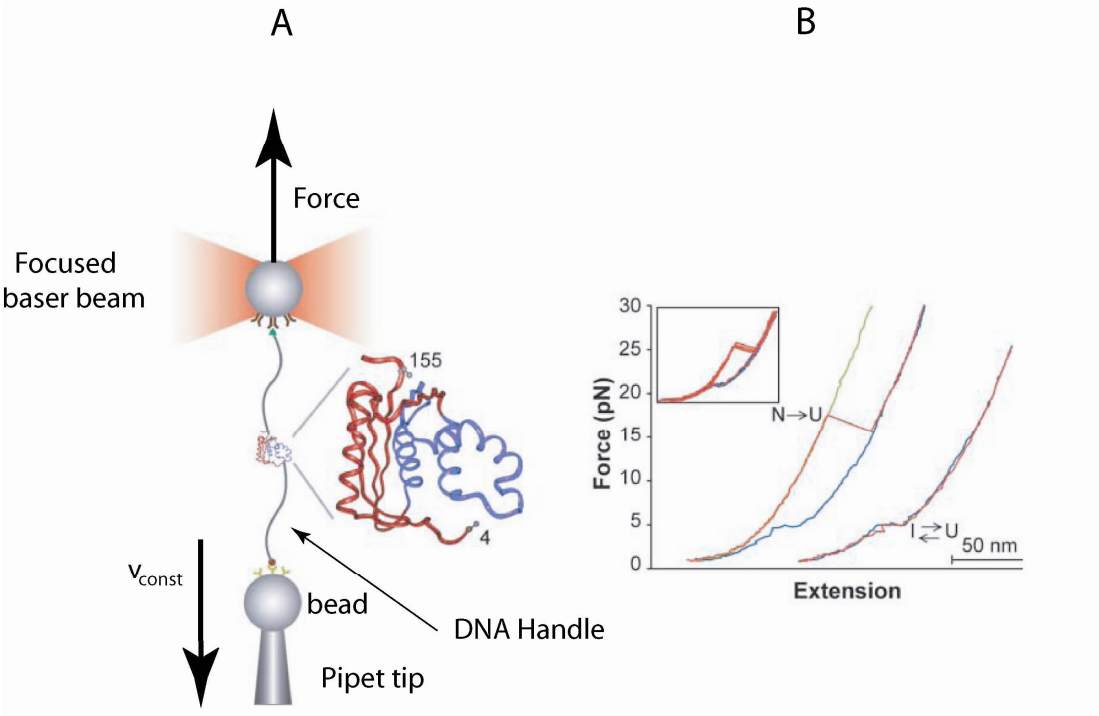

Fig. 1.4 (A) Schematic representation of LOT set up, (B) shows force extension curves of RNase H upon stretching (red) and relaxation (blue); it shows that there is a well defined on-pathway intermediate state (I) between native (N) and unfolded (U) states. (Taken from ref. [49])

1.2.3. Single-Molecule Electrophoresis

It is possible to measure experimentally the time it takes for a polymer to get across a nanometer-sized pore. One technique (see, [50-52]) involves the use of an electric field to drive a charged biopolymer across a transmembrane protein channel pore such as α-hemolysin (α-HL) while simultaneously measuring the ionic current across the channel. Whenever a single polymer is inside the pore, it partially or completely blocks the current. The duration of such blocking events directly reports the time of the polymer translocation. It is therefore of interest to examine the dependence of the translocation time on the driving force (that would be proportional to the electric field).

Electrically driven transport of biomolecules is widely utilized in the partition of polynucleotides through nanometer size pores. Recently, this method has attracted researchers who study protein dynamics and folding. α-HL, a β-barrel protein pore, produced by the bacterium *Staphylococcus aureus*, has several properties that make it ideal for studying protein translocation. (i) It can self-assemble in a lipid bilayer, in effect

creating a nanometer size pore across lipid membranes; (ii) under certain conditions the pore stays stable for tens of minutes and (iii) it shows chemical and structural similarities with mitochondrial and chloroplast membrane pores which allows elucidating the mechanism of protein import into these organelles [53].

In general, an equimolar concentration of protein is prepared at the two compartments of the lipid bilayer. Applying a potential at one side of the membrane causes the oppositely charged molecules to enter the pore and slide through the channel from one side of the membrane to the other. This process is driven by the electric field gradient across the membrane. The transport of an individual molecule through the pore causes transient ionic current blockades as shown in Fig 1.5. The distribution of the blockade times as well as the frequency with which the blockades occur, depend on the translocation mechanism which thus can be inferred from such measurements.

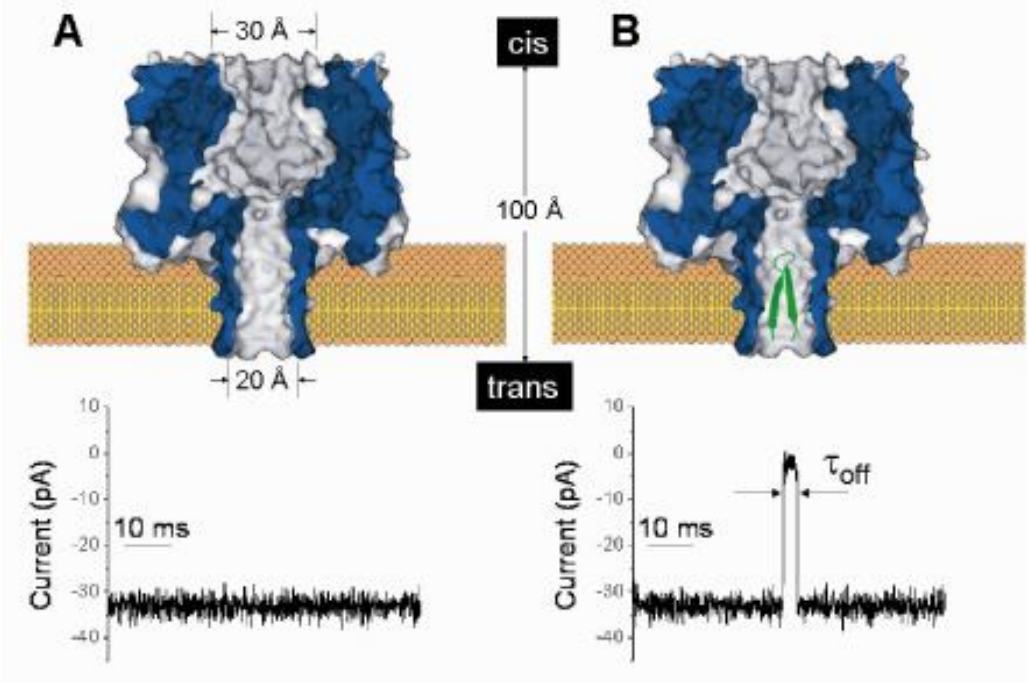

Fig. 1.5 (A) α-HL forms a pore that remains open for long periods; (B) The translocation of a β-hairpin through an α-HL pore produces a transient current blockade. (Taken from ref. [54])

1.3. COMPUTER SIMULATIONS

1.3.1. All atom molecular dynamic simulations

The Molecular Dynamics (MD) simulation technique simply solves $F = ma$ for a system of particles. The forces acting on each particle are calculated from an empirical potential energy function. This atomic level simulation method provides access to the molecular details of processes which are generally difficult to observe experimentally [33, 34, 55-61] .

Steered molecular dynamics (SMD) is the commonly used computational method to study mechanical unfolding of proteins [34, 59, 62, 63]. In SMD, one effectively attaches a spring to one end of the protein and pulls at that spring at a constant velocity. The force response of the protein is calculated as the spring constant times the spring extension,

$$f = k_s(R_{fold} - R + vt) \tag{1.1}$$

Here, k_s is the stiffness of the spring, R is the domain extension. The domain extension R is defined as the distance between the two residues between which the stretching force is measured. R_{fold} is the initial extension of the folded domain and v is the loading rate. This method mimics real AFM experiments. However, typical simulation time scales in SMD are about six orders of magnitude shorter than the experimentally relevant time scales. To simulate the unfolding within computationally accessible times, the protein domain is pulled up to six orders of magnitude faster with two orders of magnitude stiffer spring constants. When the dynamics studied with SMD under such conditions is extrapolated to the pulling rates of the experiments, the results are found to be very different [59].

Several approaches have been developed recently, whose goal is to provide a direct comparison between singe-molecule experiments and simulations. One method attempts to develop a simplified description of the dynamics along a one-dimensional reaction coordinate R. The system constrained at a given value of R, feels the potential of mean force $G(R)$. It is well known that a poor choice of a reaction coordinate in transition

state theory can lead to very poor estimates of reaction rates [64-66]. On the other hand, models assuming that the effect of all degrees of freedom other than the reaction coordinate amounts to a Stokes-type friction and a stochastic force and postulating a simple Langevin equation to describe one-dimensional dynamics along the reaction coordinate seem to agree quite well with steered Molecular Dynamics simulations of protein unfolding [34, 59, 62, 63].

In this method, the free energy as a function of the reaction coordinate R is calculated from equilibrium calculations by

$$G(R) = -k_B T \ln p(R) \tag{1.2}$$

where $p(R)$ is the probability distribution of R. To obtain the global shape of $G(R)$ far away from the predominantly populated native state, the weighted histogram/umbrella sampling method is used[67-69].

Despite the success of this method in studying unfolding dynamics of proteins under mechanical force, such an approach cannot be used to characterize the unfolded state of a protein for large extensions, because it requires prohibitively long simulation time scales. Particularly, recent SMEs such as protein unfolding at constant force and refolding [47-49] of fully stretched proteins under mechanical force necessitate a dynamics model that allows one to characterize the unfolded states and to capture the protein folding time scales.

1.3.2. Coarse grained models

One attempt to tackle the problem of long time scales is using a reduced, coarse grained representation of the molecular system of interest. In this method, parameters of the potential are tuned empirically by matching experimental data or all-atom simulations. Although the use of coarse-grained models may strip the results of explicit biological significance, such models have been shown to provide valuable insight into various biophysical phenomena such as protein folding, mechanical processes in proteins and RNA and DNA supercoiling.

1.4. OBJECTIVE AND OVERVIEW OF THESIS

The main goal of this thesis is to use a combination of theory and computer simulations to develop a molecular level understanding of biomolecular processes that involve mechanical forces and that occur *in vivo* and/or are studied by single-molecule experimental techniques. We focus on three classes of biomolecular processes: Protein translocation, mechanical stretching of individual proteins and the dynamics of molecular knots.

Our general approach is based on utilizing Langevin dynamics simulations of coarse-grained biopolymer models. Depending on the timescale of the process under study we have either directly simulated its dynamics or have developed a reduced representation based on a low-dimensional free energy landscape. Our work on protein translocation has been carried out in part in collaboration with the experimental group of Dr. Liviu Movileanu at Syracuse University, resulting in a molecular-level view of the translocation of peptides across the alpha-hemolysin pore.

The following is the overview of the thesis:

In Chapter 2 we study the translocation of a β-hairpin forming peptide through a tunnel that mimics the exit tunnel in a ribosome. We have computed the free energy of the peptide as a function of its position relative to the tunnel exit and also studied the properties of the conformational ensemble when the peptide's position is restricted at different points along the tunnel.

In Chapter 3, we study the translocation of a protein pulled through a pore and compare it with the mechanical unfolding of the same protein in an AFM-type experiment, where a stretching force is applied between the ends of the chain. We have computed the potential of mean force (PMF) experienced by the domain for both scenarios and identified the unfolding intermediates corresponding to the local minimum of the PMFs. We have shown that the observed unfolding mechanisms are different for translocation and stretching. We have further compared the free energy cost of squeezing an initially folded protein into the pore with that of a random-coil-like homopolymer and we have shown that the former case involves several partially folded intermediates.

Chapter 4 describes a joint experimental and theoretical study of electrophoretic translocation of various β-hairpin peptides across the α-hemolysin protein pore. The β-hairpins of varying stability corresponded to the C terminal residues of the B1 domain of protein G were electrically driven through the pore (experiments performed by the Movileanu group at Syracuse University). We have used Langevin dynamics simulations to study the same process. The dependence of translocation times on the electric force and on the stability of the peptides has been studied. We have compared the results of experiments and simulations and analyzed the molecular details of translocation events with the help of simulations. We have further studied the dependence of the average translocation time and the distribution of translocation times on the applied transmembrane potential and compared the cis and trans translocation kinetics.

In Chapter 5 we have simulated the mechanical unfolding and refolding of a minimalist off-lattice model of the protein ubiquitin to explore in detail the slice of the multidimensional free energy landscape that is accessible via mechanical pulling experiments. We have constructed a two-dimensional free energy surface as a function of both chemical and mechanical reaction coordinates and examined the coupling between the two. We have further studied the refolding trajectories after the protein has been pre-stretched by a large force, as well as the mechanical unfolding trajectories in the presence of a large stretching force. Finally, we have proposed a free energy model of a polyprotein chain consisting of multiple domains to explain the recently observed [47] "slow phase" in the refolding of proteins under mechanical tension.

In Chapter 6 we study the mechanical response of a protein domain in an experiment[49], where an increasing force was applied to unfold a protein and then lowered at the same rate until refolding was observed. The observed force-vs.-extension curves for stretching and refolding are generally different and depend on the loading/unloading rate. To gain a better insight on the dynamics of transitions and dependence of these transitions on the loading rate we have studied a computer experiment with the same method. We have further compared the outcome of two different experiments: force control and displacement control in loading.

Chapter 7 describes our computer simulation study of the dynamics of "friction knots" joining individual polymer strands. A friction knot splicing two ropes becomes jammed when the ropes are pulled apart. In contrast, molecular friction knots eventually become undone by thermal motion. We have shown that depending on the knot type and on the polymer structure, a microscopic friction knot can be strong (the time τ the knot stays tied increases with the force F applied to separate the strands) or weak (τ decreases with increasing F). We have further presented a simple model explaining these behaviors.

Finally, Chapter 8 summarizes the key results of this thesis.

Chapter 2

Exiting of β-hairpin-forming Peptide through a Cylindrical Tunnel*

2.1. INTRODUCTION

Biopolymers (RNA, DNA and proteins) translocate accross narrow constrictions during biological processes [1, 71] . One particularly important example involves nascent peptides, which exit ribosomal subunit that has a lentgth of ~100Å and width varies between 10-28 Å [3, 12] .

Confining the peptide within a sufficiently wide tunnel stabilizes the folded state[72-78] however, in the ribosomal subunit the tunnel is too narrow in some places (D ~ 10Å) to accommodate any protein structure other than an extended chain or a helix[12]. This suggests that the effect of confinement on the state of the protein inside the tunnel may be rather complex. Experiments and simulations also indicate that different translocation scenarios are possible, depending on the specific protein in question [2, 79]. In particular, there is experimental evidence that some peptides partially fold to form α-helices within the tunnel[80] and that co-translational folding takes place for some but not all peptides [2, 4, 15, 16]. Computer simulations also support the possibility of co-translational folding in some protein models [17, 79], and demonstrated that co-translational folding is faster than direct folding from the denatured state and that confinement can change the folding pathways[17].

In order to elucidate the functional role played by the ribosomal tunnel, and the state of peptide during translocation we have studied a minimalist model of a beta hairpin forming peptide inside a cylindrical tunnel. Our study is different from previous theoretical studies of polymer translocation[50, 81, 82] (in particular, DNA translocation), which were mostly limited to unstructured homopolymer models.

We assume that translocation is a quasi-equilibrium process at a typical rate of polypeptide synthesis, (~15 residues/sec in bacteria and ~3 residues/sec in

* Large portions of this chapter have been published in reference 70.

eukaryotes[83]) the polypeptide formed inside the tunnel moves slowly enough that it has sufficient time to reach thermal equilibrium at any location along the tunnel. The quasi-equilibrium assumption allows us to avoid dealing with the actual kinetics of translocation. Instead, we use umbrella sampling to map out the free energy profile for a protein as a function of its position relative to the tunnel exit. This free energy profile is compared with that of a homopolymer that has no structural preferences. For diameter of the pore larger than a critical value $D \geq D_c$ the peptide stays folded all the way through the tunnel while for $D < D_c$ we find that intermediate structures are formed that depend on the location of the peptide relative to the exit.

2.2. MODEL AND METHODS

2.2.1. The protein model

Our goal here is to elucidate the interplay between folding and translocation, we necessarily need a model of a protein whose folding dynamics can be studied over reasonably long times. To ensure this, we adopt the off-lattice minimalist coarse-grained protein model[84] that consists of N C_α atoms each connected to the adjacent one by a virtual bond of length σ = 3.8 Å, which is the average distance between consecutive C_α atoms in peptides. These residues can be of three types: hydrophobic (B), hydrophilic (L), and neutral (N). The potential energy of a given conformation of a peptide, with the i-th C_α atom located at r_i, is given by:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V_{bond} + V_{bend} + V_{dih} + V_{non\text{-}bonded} + V_{w\text{-}c} \quad (2.1)$$

Here, the potential Vbond accounts for the connectivity of the chain and assumes that each bond is a stiff harmonic spring,

$$V_{bond} = \sum_{i=2}^{N} k_b (|\mathbf{u}_i| - \sigma)^2 / 2 \quad (2.2),$$

where $\mathbf{u}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$ is the bond vector. The force constant of the spring was chosen to be $k_b = 100\ \varepsilon_h/\sigma^2$, where ε_h is a unit of energy that represents a typical energy scale of the hydrophobic interaction ($\varepsilon_h \cong 1$ kcal/mol) in proteins. The bending potential imposes the constraints inherent to the peptide bond geometry:

$$V_{\text{bend}} = \sum_{i=2}^{N-1} k_\theta (\theta_i - \theta_0)^2 / 2 \qquad (2.3)$$

where $\theta_0 = 105^0$ is the equilibrium bending angle, θ_i is the angle between $\mathbf{u}_i$ and $\mathbf{u}_{i+1}$, and $k_\theta = 20\ \varepsilon_h/(rad)^2$ is the spring constant. The dependence of the energy on the dihedral angles φ_i formed between the $\mathbf{d_i} = (\mathbf{u_i} \times \mathbf{u_{i+1}})$ and $\mathbf{d_{i+1}} = (\mathbf{u_{i+1}} \times \mathbf{u_{i+2}})$ vectors is incorporated in V_{dih}:

$$V_{\text{dih}} = \sum_{i=2}^{N-2} A_i(1+\cos\varphi_i) + B_i(1+\cos 3\varphi_i) \qquad (2.4)$$

Depending on the parameters A_i and B_i, each term in Eq. 2.4 has several minima corresponding to *gauche* ± and *trans* states. The parameters A_i and B_i depend on the residue type. If at most one of the residues among the four that define a dihedral is neutral (N) then, $A_i = B_i = 1.2\varepsilon_h$ so that the *trans* state is preferred over the *gauche* states resulting in an extended conformation. Otherwise $A_i = 0$ and $B_i = 0.2\varepsilon_h$, resulting in *trans* and *gauche* states of equal probability.

The energy $V_{\text{non-bonded}}$ describes the interaction between sequence-distant residues that are not covalently bonded. We assume that it can be represented as a sum of pairwise potentials:

$$V_{\text{non-bonded}} = \sum_{|i-j|\geq 3} V_{ij}(|\mathbf{r}_j - \mathbf{r}_i|) \qquad (2.5)$$

The interaction potential V_{ij} between residues *i* and *j* depends on the type of each of them. The interaction of a neutral residue with another residue is described by a short ranged repulsion term that arises from the volume exclusion:

$$V_{N\alpha}(r) = 4\varepsilon_h \left(\frac{\sigma}{r}\right)^{12} \ (\alpha = \text{N, L, B}). \qquad (2.6)$$

The interaction between a pair of hydrophobic residues is given by a Lennard-Jones potential

$$V_{BB}(r) = 4\varepsilon_h \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (2.7)$$

The interaction between L-B and L-L pairs is taken to be repulsive and long ranged:

$$V_{L\alpha} = 4\varepsilon_L \left[\left(\frac{\sigma}{r} \right)^{12} + \left(\frac{\sigma}{r} \right)^{6} \right] \tag{2.8}$$

with $\varepsilon_L = 2/3\varepsilon_h$.

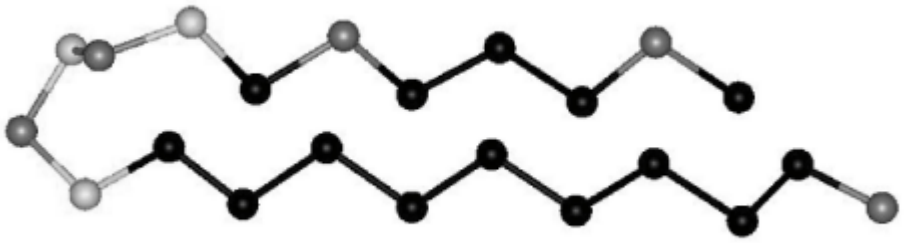

Fig. 2.1 Native state of $LB_9(NL)_2NBLB_3LB$ peptide (*peptide B*). Each bead represents an amino acid residue. The gray beads correspond to hydrophilic (*L*), white beads neutral (*N*) and black beads to hydrophobic (*B*) residues.

We have studied two peptide models, each consisting of 22 residues; The first one, from here on referred to as *peptide B*, has the sequence $LB_9(NL)_2NBLB_3LB$ and it folds to a β-hairpin structure see Fig 2.1. The second model represents a simple poly-amino acid that has no structural preferences; this is accomplished by choosing a sequence of neutral residues (N_{22}). In addition, we have modified the dihedral interaction potential for this peptide and used

$$V_{dih} = \sum_{i=2}^{N-2} B(1 + \cos 3\varphi_i) \tag{2.9}$$

where, $B = 1.2\varepsilon_h$, instead of Eq. 2.4. From now on, we will refer to this model as *peptide R*.

To describe the interaction of the peptide with the walls of the tunnel we introduce the confinement potential $V_{w\text{-}c}$ in Eq. 2.1. We assume that the peptide does not stick to the walls. The interaction between each C_α atom and the cylinder wall is purely short range and repulsive:

$$V_{w\text{-}c}(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = \sum_{i=1}^{N} 4\varepsilon_h \left(\frac{\sigma}{\tilde{r}(\mathbf{r}_i)} \right)^{12}, \tag{2.10}$$

where $\tilde{r}(\mathbf{r}_i)$ is the shortest distance between the wall and the atom located at $\mathbf{r} = \mathbf{r}_i$. The geometry of the tunnel is depicted in Fig 2.1. Its diameter is D and the total length is $2L$. The cylinder axis coincides with the z-axis such that $-L \le z \le L$ for points inside the tunnel. The shortest distance $\tilde{r}(\mathbf{r})$ from a point $\mathbf{r} = (x, y, z)$ to the cylinder is given by:

$$\tilde{r}(\mathbf{r}) = \begin{cases} \sqrt{(x^2+y^2)\left(1 - \dfrac{D}{2(\frac{D}{2} - \sqrt{x^2+y^2})}\right)^2 + (z-L)^2}, & |z| \ge |L| \\ \sqrt{(x^2+y^2)\left(1 - \dfrac{D}{2(\frac{D}{2} - \sqrt{x^2+y^2})}\right)^2}, & |z| \le |L| \end{cases} \tag{2.11}$$

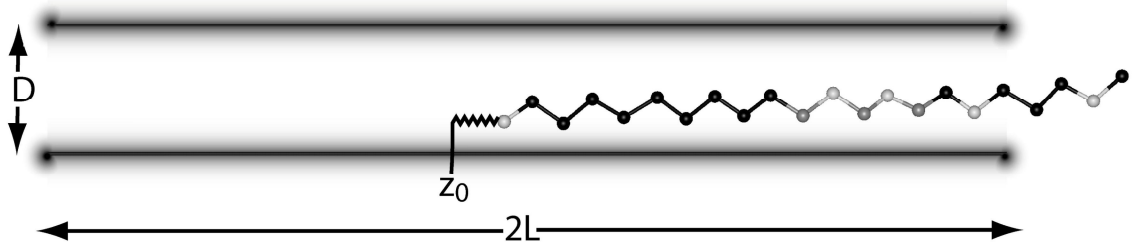

Fig. 2.2 *Peptide B* constrained at $z_0 = 4\sigma$ inside a cylindrical tunnel. The cylinder axis coincides with the z-axis such that $-L \le z \le L$ for points inside the tunnel; the total length of the tunnel is $2L = 40\sigma$.

The cylinder walls are shown to be fuzzy in Fig. 2.2 to represent the fact that the interaction between the peptide and the wall is not hard-wall repulsion but a continuous function that decays on a length scale on the order of σ. Also shown in Fig. 2.2 is a snapshot from a simulation showing the dimensions of the polypeptide when it assumes a

stretched conformation. Here the horizontal displacement z_0 of the leftmost C_α atom is constrained to be near $z = 4\sigma$ by attaching to it a harmonic spring. Applying such constraints is useful in free energy calculations and will be discussed in detail.

2.2.2. Simulation of the dynamics.

Following [84-86], we assume that the peptide's dynamics are governed by the Langevin equation,

$$m\ddot{\mathbf{r}}_i(t) = -\zeta \dot{\mathbf{r}}_i(t) - \frac{\partial V}{\partial \mathbf{r}_i} + \boldsymbol{R}_i(t)\,, \tag{2.12}$$

where m is the mass of a C_α atom, $\mathbf{r}_i(t)$ is the position of the i-th atom at time t, ζ is a friction coefficient, and $\boldsymbol{R}_i(t)$ is a stochastic force on the i-th atom. This force satisfies the fluctuation-dissipation relationship

$$< R_{i\alpha}(t) R_{j\beta}(t') >= 2\zeta k_B T \delta_{ij}\delta_{\alpha\beta}\delta(t-t') \tag{2.13}$$

where the Greek letter subscripts refer to the x, y and z components of the force, T is the temperature and $\delta(t)$ is the delta function. We use $\zeta = 0.05\ (\sigma^2/m\varepsilon_h)^{-1/2}$ in all of our simulations. The equation of motion in eq (2.11) is solved by using the velocity Verlet algorithm described in reference [87].

2.2.3 Free energy calculations

As noted in Introduction, we assume that the synthesis rate is slow enough that the protein is in thermal equilibrium at every point along the tunnel. To monitor the progress of the peptide as it moves along the tunnel, we use the "reaction coordinate" equal to the position of one end of the chain z_0 and calculate the free energy profile $G(z_0)$ as a function of this coordinate.

We start by generating a fully extended zigzag structure, which has an end-to-end distance of ~17.5σ. We then restrain the motion of the leftmost end C_α atom near a given position z along the tunnel by adding a harmonic spring term $V_s(z) = k_s(z - z_0)^2/2$ to the energy of Eq. 2.1. To improve the sampling, we use the parallel tempering method [88, 89]. The free energy profile $G(z_0)$ along the tunnel is computed from the average value of the force that is experienced by the peptide inside the pore and is given by $< f > =$ -

dG/dz_0. The average force can be measured directly in the simulation $< f >= k_s(< z_0 > - z)$. By performing the simulation for $z = z^{(1)}, z^{(2)}, \ldots z^{(N)}$ one finds the dependence $< f > (< z_0 >)$. The free energy $G(z)$ can be calculated by integrating this dependence:

$$G(z) = -\int_0^z < f > (z)dz \tag{2.14}$$

Limitations of this procedure are discussed in reference [59].

2.3 RESULTS

2.3.1 Thermodynamics of folding in the bulk and inside the tunnel

In coarse grained protein models thermodynamics of folding is characterized by collapse temperature T_θ and folding temperature T_f [73, 74, 84, 86, 90]. The temperature, at which the protein undergoes a transition from extended conformations to compact is called the collapse temperature T_θ. At this temperature, the heat capacity $c_V(T)$ as a function of temperature exhibits a peak. As temperature is further decreased, the folded state becomes predominantly populated. This is referred to as the folding transition. An easy way to determine folding temperature T_f is to monitor an order parameter that describes how close a given structure is to the native state.

$$\chi = \frac{2}{(N-2)(N-3)} \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \Theta(\varepsilon - \left| r_{ij} - r_{ij}^{N} \right|), \tag{2.15}$$

where, r_{ij} is the distance between the i-th and j-th atoms and r_{ij}^{N} is the corresponding distance in native state. The parameter ε allows for fluctuations around the native state and is taken to be 0.3; $\Theta(x)$ is the step function thus $\chi \approx 1$ in native ensemble. The folding temperature T_f is the temperature at which the susceptibility $\left\langle \Delta\chi^2 \right\rangle = < \chi^2 > - < \chi >^2$, exhibits a peak.

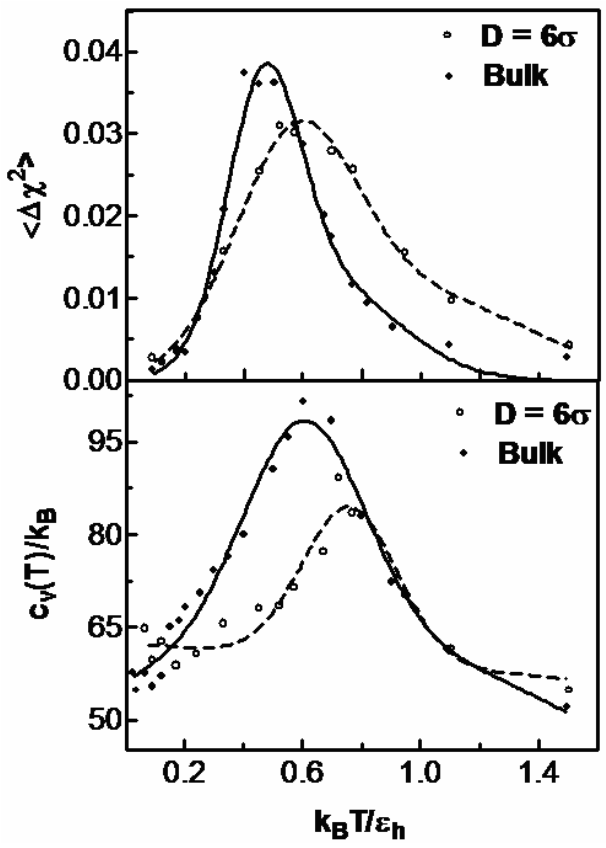

Fig. 2.3 (A) Susceptibility $\left\langle \Delta\chi^2 \right\rangle = <\chi^2> - <\chi>^2$ and (B) the heat capacity $c_V(T)$ as a function of temperature for *peptide B* in bulk (filled circles) and inside an infinitely long tunnel with $D = 6\sigma$ (open circles). The lines are fit to the data.

It has been previously pointed out that confinement should stabilize the more compact native state relative to the more extended, denatured state of the protein [72-78]. Indeed, depending on the tunnel dimensions we have observed both stabilization and destabilization of the native state in our simulations. Confining *peptide B* inside an infinitely long tunnel with $D = 6\sigma$ results in an increase in its folding temperature. This is demonstrated in Fig. 2.3, where the temperature dependences of the heat capacity $c_V(T)$ and the susceptibility $<\Delta\chi^2>$ are plotted for the peptide inside a tunnel and in the absence of the tunnel. Shift of T and t to higher temperatures shows that confinement indeed stabilizes folding.

However a sufficiently narrow tunnel destabilizes the native state. This was observed in a series of simulations that were performed for different values of D at a temperature $T = 0.2\ \varepsilon_h/k_B$ well below the folding transition point. We have found that there is a critical diameter $D_c \cong 3.6\sigma$ such that for $D \geq D_c$ the peptide assumes the folded conformation inside the tunnel while for $D < D_c$ the peptide unfolds an assumes extended conformations.

2.3.2. Free energy change in the course of translocation

From the results of Section 2.3.1 we expect that translocation of a peptide through a narrow tunnel with $D < D_c$ would be quite different from the case of a wider tunnel, $D \geq D_c$. To understand the structural changes that the peptide undergoes as it emerges from the tunnel, we have restrained it at various positions along the tunnel and computed the free energy $G(z_0)$ as a function of the coordinate z_0 of the left end atom of the chain (see Fig. 2.2). The results are reported below for $T = 0.2\ \varepsilon_h/k_B$ and for two different values of the tunnel diameter, one below D_c, $D = 3\sigma$, and the other above Dc, $D = 6\sigma$. In each case, we have compared the calculated free energy profiles $G(z_0)$ with those for the unstructured *peptide R*. We have further studied the ensembles sampled by the peptide conformations when its end atom is constrained along different positions along the tunnel.

A. Narrow tunnel ($D = 3\ \sigma$)

When confined to a narrow tunnel, the peptide does not have enough space to form any structure other than extended chain. As a result, both the structureless *peptide R* and the β-hairpin forming *peptide B* exhibit similar trends in $G(z_0)$ for $z_0 \leq 6\sigma$, as seen in Fig. 2.4a. The two dependences diverge for larger values of z_0 and the free energy of *peptide B* is found to be lower than that of *peptide R* at the tunnel exit. Figs. 2.4b and 2.4c display the average potential energy $\langle V \rangle(z_0)$ of *peptides B* and *R* whose end atoms were restrained at different points z_0 along the tunnel. The total potential energy V of the chain is the sum of the energy of the chain itself $V_{c\text{-}c} = V - V_{w\text{-}c}$ and the interaction energy of the chain with the wall, $V_{w\text{-}c}$. Both of these contributions are shown in Figs. 2.4b, c. $\langle V_{w\text{-}}$

$_c$>(z_0) decreases monotonically for both peptides because the repulsive interaction between the wall and the chain is roughly proportional to the number of C_α atoms inside the tunnel. The chain energy <V_{c-c}> behaves differently: For *peptide B* it starts to drop at $z_0 \approx 12\sigma$ while for *peptide R* (Fig. 2.4c) it stays nearly constant throughout the translocation process. Fig. 2.4d shows the chain entropy $S(z_0)$ calculated for each peptide from the equation $G = \langle E \rangle - TS$, where $\langle E \rangle$ is the average total energy. Here, we observe different behavior for the two peptides. Each C_α that exits

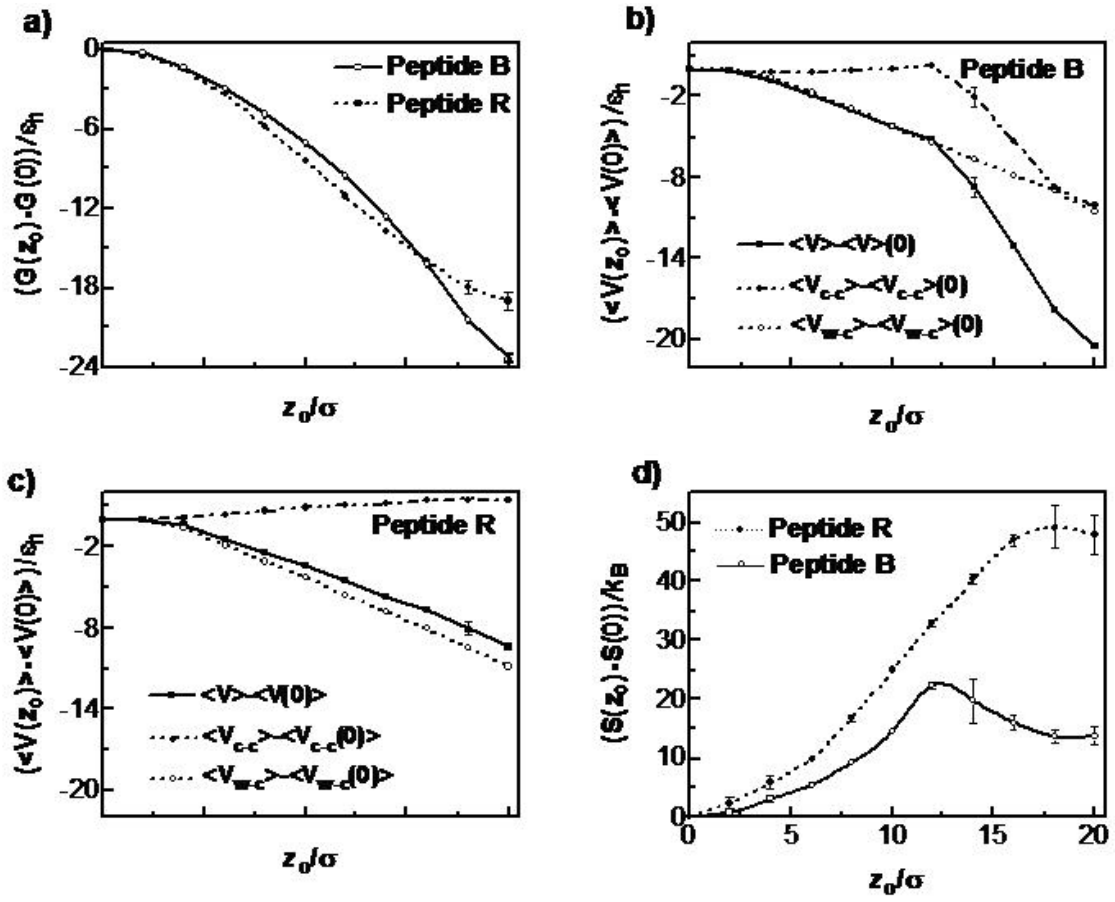

Fig. 2.4 (a) Free energy, (b,c) average potential energy of *peptides B* and *R*, respectively, and (d) entropy of *peptides B* and *R* as a function of the translocation coordinate z_0 for the narrow tunnel, $D = 3\ \sigma$. The lines are to guide the eye. The total potential energy consists of the chain energy <V_{c-c}>(z_0) and the wall-chain interaction, <V_{w-c}>(z_0), which are also shown for *peptides B* and *R*.

the tunnel has more space to sample in the case of *peptide R* as it has no structural preference. The drop in both the chain energy and the entropy at $z_0 \approx 12\sigma$ indicates that the *peptide B* undergoes a structural change for $12\sigma \leq z_0 \leq 20\sigma$ so that its free energy drops below that of the random coil denatured state modeled by *peptide R*. Further, *peptide B* is folded when it emerges at the tunnel exit ($z_0 = 20\sigma$). Indeed, the decrease in the chain energy $\langle V_{c\text{-}c} \rangle(20\sigma) - \langle V_{c\text{-}c} \rangle(12\sigma) = (-10.34 \pm 0.05)\varepsilon_h$ is close to the minimum energy of the native state $V_{MIN} = -10.75\varepsilon_h$. Similarly, the drop in entropy that *peptide B* exhibits upon exiting the tunnel (between $z_0 = 12\sigma$ and 20σ) can be explained by its folding.

The gradual changes in *G*, *S*, and <V> that one sees midway between $z_0 = 12\sigma$ and the exit point $z_0 = 20\sigma$ can be due to two reasons: Either the peptide undergoes structural changes that involve perhaps some misfolded or partially folded conformations or the ensemble consists of unfolded and folded conformations and the population of the native state gradually increases as z_0 is increased. To find out which of the two possibilities apply to our case, we have studied in more detail the conformational ensemble at different values z_0. To this end, we have calculated the free energy profiles $G(R_g)$ and $G(\chi)$ for the chain with the coordinate z_0 of its end atom harmonically restrained to be around z (see Section 2.2). Here, the free energy profile along a coordinate Γ is defined as $G(\Gamma) = -k_B T \ln(P(\Gamma))$ where $P(\Gamma)$ is the probability distribution function of Γ. These are shown in Fig. 2.5 for different values of *z*. Representative chain conformations corresponding to these values of *z* are shown in Fig. 2.6. The part of the peptide that is confined inside the tunnel remains in the extended conformation, while the part that is outside the tunnel undergoes structural changes as *z* is increased.

For $0 \leq z \leq 10\sigma$, the middle part of *peptide B* is inside the tunnel thereby precluding the formation of the beta turn. Indeed, we find that both $G(R_g)$ and $G(\chi)$ have a single minimum ($R_g \sim 4.4\sigma$, $\chi \sim 0.35$) corresponding to the extended chain (not shown in Fig. 2.5). For $z \geq 12\sigma$ the profiles $G(R_g)$ and $G(\chi)$ depend on the translocation

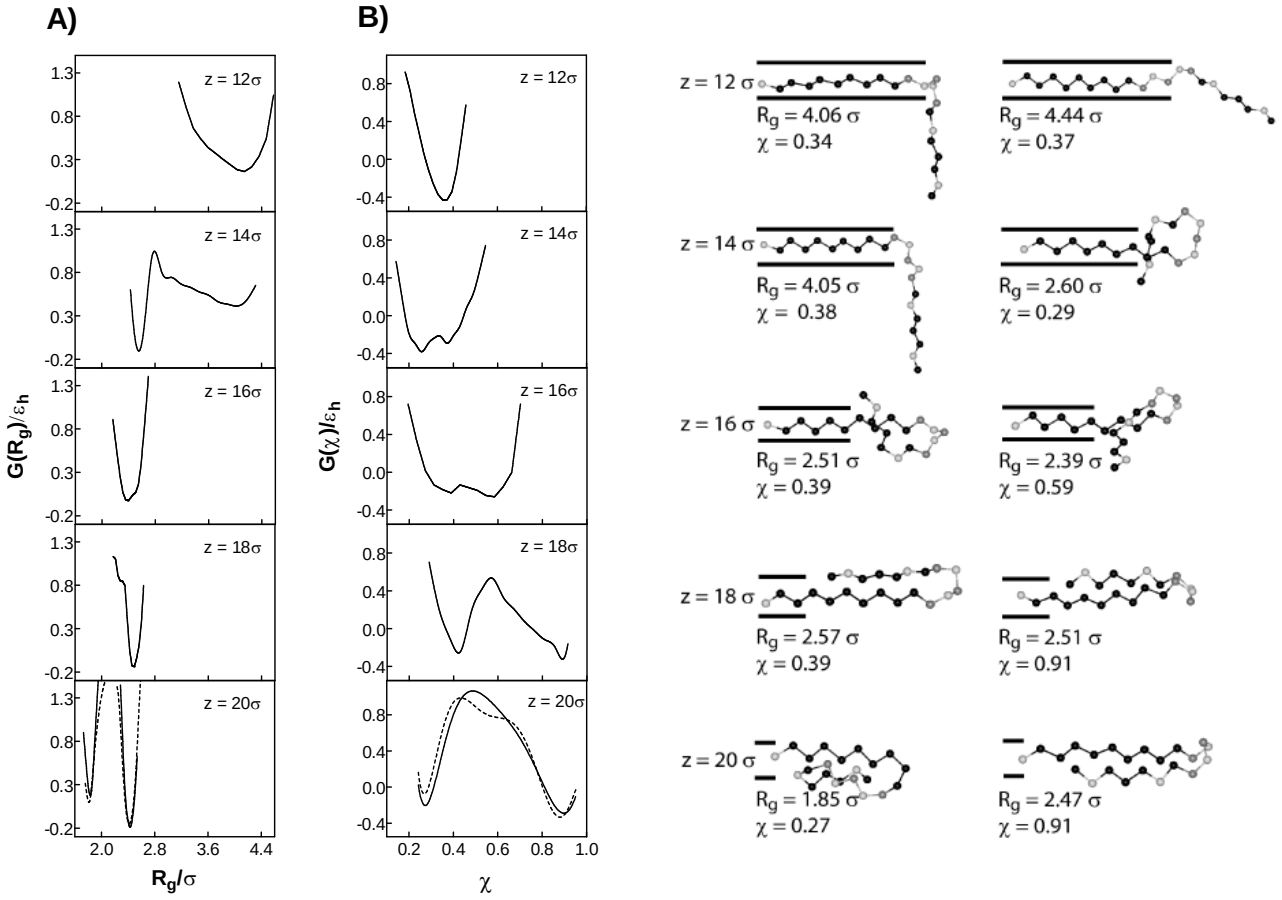

Fig. 2.5 The free energy as a function of (A) the radius of gyration, $G(R_g) = -k_BT\ ln\ P(R_g)$, and (B) the order parameter, $G(\chi) = -k_BT\ ln\ P(\chi)$, for *peptide B* with its end atom coordinate z_0 harmonically constrained near different points z along the tunnel. The diameter of the tunnel is $D = 3\sigma$. Also the free energy profiles in the absence of the tunnel (Case $z = 20\sigma$, dashed lines) are plotted.

Fig. 2.6 Snapshots of typical peptide configurations observed when the peptide's translocation coordinate z_0 is constrained at different positions along the tunnel.

coordinate z. For $z = 12\sigma$ we still see one minimum for both $G(R_g)$ and $G(\chi)$; Fig. 2.6 shows two representative chain conformations located in the vicinity of these minima. As z is further increased ($z = 14\sigma$), two minima appear in both $G(R_g)$ and $G(\chi)$. One minimum still corresponds to extended conformations similar to those found for $z = 12\sigma$ while the other minimum ($R_g \sim 2.6\sigma$, $\chi \sim 0.29$) corresponds to more compact, misfolded conformations where the chain has folded back onto itself (see Fig. 2.6).

For $z \geq 16\sigma$ the minimum corresponding to the extended state disappears from $G(R_g)$ and $G(\chi)$. The only conformations present in this case are the more compact misfolded structures with $R_g \sim 2.5\sigma$ close to that of the folded state; The structural order parameter χ in this case has a bimodal distribution peaked around two dominant values (corresponding to two minima in $G(\chi)$). One of these dominant values approaches $\chi \sim 1$ corresponding to the β-hairpin while the other remains small and corresponds to non-native compact conformations. When the peptide is near the exit ($z = 20\sigma$), the profiles $G(R_g)$ and $G(\chi)$ are nearly identical to the corresponding free energy profiles found in the bulk; The latter are shown as dashed lines in Fig. 2.5. The ensemble in this case is dominated by the fully folded and collapsed state (this state is known to be the lowest energy structure in zero temperature), which appear as two minima in both $G(R_g)$ and $G(\chi)$. The effect of the tunnel on the peptide conformation in this case is weak and manifests itself in small differences of the free energy profiles from those in the bulk.

B. Wide tunnel ($D = 6\ \sigma$)

The free energy profile $G(z_0)$ is shown in Fig. 2.7 for both *peptide B* and *peptide R*. Since the diameter of the tunnel is larger then D_c we expect *peptide B* to be folded inside the tunnel. Indeed, its mean radius of gyration stays nearly constant and equal to the value expected for the folded protein, $\langle R_g \rangle \sim 2.4\sigma$, at each position in the course of translocation. Similarly, the mean value of the order parameter stays nearly constant, $\langle \chi \rangle \sim 0.8$.

The free energy of *peptide B* stays constant for $z_0 \leq 12\sigma$. At this point, the peptide starts emerging from the tunnel. Since the atoms outside the tunnel have a larger volume

to explore, this results in an increase in entropy and thus a decrease in the free energy seen in Fig. 2.7. The free energy of *peptide R* exhibits similar behavior. The chain energy and the energy of interaction with the wall, $\langle V_{w\text{-}c}\rangle(z_0) - \langle V_{w\text{-}c}\rangle(0)$ and $\langle V_{c\text{-}c}\rangle(z_0) - \langle V_{c\text{-}c}(0)\rangle$, (not shown) do not exhibit any abrupt changes and stay close to zero for both peptides so that the free energy decrease seen in Fig. 2.7 is a purely entropic effect.

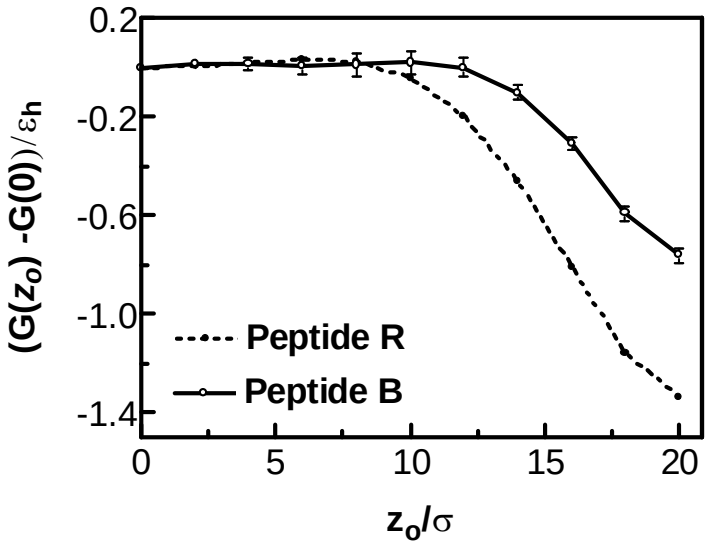

Fig. 2.7. Free energy for *peptides B* and *R* as a function of the translocation coordinate z_0 for the wide tunnel with $D = 6\ \sigma$. The lines are to guide the eye

Note that Figure 2.7 does not imply, as it may appear, that the free energy of the native conformation of *peptide B* is higher than that of a denatured, random-coil ensemble mimicked by *peptide R*. Each point in Fig. 2.7 represents the free energy relative to that *inside the tunnel* (i.e., at $z_0 = 0$). For *peptide B* this initial state is a chain that is *already folded* so the free energy decrease seen in Fig. 2.7 for *peptide B* (as well as for *peptide R*) is the free energy of confinement. This is in contrast to Fig. 2.4a, in which both *peptide B* and *peptide R* are initially unfolded and the free energy of *peptide B* changes due to both confinement and folding.

Since *peptide B* is folded and thus more compact than *peptide R* ($\langle R_g\rangle \sim 2.9\sigma$ for *peptide R* in the bulk), the free energy cost of confining *peptide B* within a tunnel is

expected to be smaller than that for *peptide R*[78] and this manifests in the lower free energy change as seen in Fig. 2.7.

2.4. DISCUSSION

The state of a peptide inside a tunnel depends on two parameters: the diameter of the tunnel D and the proximity of the peptide to the tunnel exit (that can be quantified as the position z_0 of one of the end α-carbons of the chain). We found here that this parameter space is divided into two distinct domains: For wide tunnels, $D \geq D_c$, the protein is folded regardless of its location inside the tunnel. For narrow tunnels, $D < D_c$, the peptide goes through a series of intermediate, misfolded (or partially folded) compact conformations as it approaches the tunnel exit.

The critical tunnel diameter D_c should depend on the specific peptide inside the channel. It is instructive that for the model studied here our estimate for this diameter, D_c ~ 13.7 Å, is comparable with the dimensions of a real ribosomal tunnel. Because the width varies along a ribosomal tunnel and because D_c would be different for different peptides, our findings suggest that folding starts on the ribosome and leaves the exit either folded or partially folded state depending on the specific peptide in question.

As the peptide progresses towards the exit of the tunnel, its free energy $G(z_0)$ decreases for two reasons: First, it becomes less confined so its entropy increases. Second, it partially folds so that its energy decreases. This results in a force $f = -dF / dz_0$ that expels the peptide from the tunnel. Based on the free energy curves calculated here, this force reaches a maximum value of ~ 46 pN for the narrow tunnel (D = 11.4 Å) and ~ 3 pN for the wide one (D = 22.8 Å). The magnitude of this expulsion force is comparable with typical forces generated by molecular motors[91, 92]. In our case this force "helps" the translocation process. However it is conceivable that in other cases (e.g., when a peptide has to enter a translocation pore) this kind of forces may have the opposite effect. It may be possible to measure mechanical forces in this range by single molecule force probe spectroscopy techniques [93, 94].

Chapter 3

Unfolding of a Protein Pulled Mechanically through a Pore*

3.1 INTRODUCTION

Protein degradation by ATP-dependent proteases and protein import into the mitochondrial matrix involve the threading of proteins though narrow constrictions whose dimensions are too small to accommodate folded proteins[9, 11]. The ensuing unfolding process is sometimes orders of magnitude faster than chemical denaturation of the same proteins[10, 11, 96, 97], suggesting that unfolding in the cell may occur via pathways different from those probed in chemical/thermal denaturation studies[98]. It has been hypothesized that the cell machinery accomplishes unfolding by mechanically pulling at the end of the polypeptide chain that is labeled for degradation or translocation[9, 96, 98-100].

Very little is known about the molecular details of such a process. Most current insights into the unfolding-via-translocation mechanisms[98, 100] are inferred from the single molecule pulling experiments (reviewed in refs. [43, 93]), which probe the mechanical unfolding process induced by stretching polypeptide chains.

One however expects significant differences between these two cases. It is known[61, 98, 100-102] that the direction and the geometry of the applied force may dramatically affect the mechanical unfolding mechanism. Since the mechanical pulling experiments involve the application of a force at the ends of the chain while translocation involves a distributed force arising from the interaction between the protein and the pore, the resulting unfolding mechanisms should generally be different.

In this chapter we discuss the mechanism of protein unfolding induced by translocation along a long cylindrical pore and we compare them with the mechanical unfolding as observed in mechanical unfolding.

* Large portions of this chapter have been published in references 50, 95.

We have resorted to the use of a minimalist off-lattice model of a ubiquitin-like domain with *N*=68 residues[103-106].

Even after the computational savings provided by the use of minimalist models, slow, biologically relevant time scales associated with barrier crossing events are rarely accessible via direct simulations of protein dynamics[59]. In a simulation, one can speed up translocation by applying a very large force, in the spirit of the steered molecular dynamics (SMD) method[63]. However extrapolation of the SMD results to the lower force regime is a very difficult task[59, 62, 63, 107, 108]. Instead, here we are following the approach that has been used and extensively tested in our previous work on the mechanical unfolding of proteins[59-61, 109].

Specifically, when a constant external force **f** is applied, the protein experiences a potential of mean force equal to $G_f(z) = G_0(z) - fz$. Here z is the reaction coordinate coupled to the force and $G_0(z)$ is the protein's free energy as a function of this reaction coordinate (in the absence of the force). The choice of the reaction coordinate depends on the process considered, as shown in Fig. 3.1. In the case of translocation, $z \equiv z_1$ (or z_N) is the displacement of the N- (or C-) terminus, whichever is pulled mechanically, along the direction of pulling. In the case of mechanical stretching, $z \equiv | z_N - z_1 |$ is the protein extension along the direction of the stretching force **f.**

We then adopt the simplified view that the translocation dynamics can be viewed as one-dimensional diffusive motion along z in the potential $G_f(z)$. The validity of this view in the context of mechanical unfolding has been critically assessed previously[59, 62, 63, 107, 109]. The key quantity in our theory is therefore the free energy profile $G_0(z)$. In the following, we describe our calculation of $G_0(z)$ for the translocation process and compare it with that for the mechanical stretching case. Once we know $G_0(z)$, we examine the translocation mechanism at different values of the driving force f applied at either N- or C-terminus of the chain. In addition, we examine how the translocation mechanism depends on the pore diameter.

3.2. MODEL AND METHODS

3.2.1. The protein model

We used the off-lattice model of Sorenson and Head-Gordon[103-106], which in turn builds on the earlier work of Thirumalai's group[84, 86] that has been utilized in the previous chapter. In this new model a protein is represented by a bead sequence and a dihedral sequence. Note that in the earlier work protein is represented by bead sequence only.

Similiar to the previous model (see section 2.2.2) the protein is represented with beads of three type: hydropphobic (B), hydrophilic (L), and neutral (N). The interaction potential, as a function of the position $\mathbf{r}_i$, i=1, …, N, of each residue, is given by:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V_{\text{bond}} + V_{\text{bend}} + V_{\text{dih}} + V_{\text{non-bonded}} + V_{\text{pore}} \qquad (3.1)$$

The details of the potential V_{bond} that accounts for bond stretching and V_{bend} that imposes bending angle constraints inherent to the peptide bond geometry is summarized in section 2.2.2.

The dependence of the energy on the dihedral angles φ_i formed between the $\mathbf{d_i} = (\mathbf{u_i} \times \mathbf{u_{i+1}})$ and $\mathbf{d_{i+1}} = (\mathbf{u_{i+1}} \times \mathbf{u_{i+2}})$ vectors is incorporated in V_{dih}:

$$V_{\text{dih}} = \sum_{i=2}^{N-2} A_i(1+\cos\varphi_i) + B_i(1-\cos\varphi_i) + C_i(1+\cos 3\varphi_i) + D_i\left[1+\cos\left(\varphi_i + \frac{\pi}{4}\right)\right] \qquad (3.2)$$

The parameters A_i, B_i , C_i and D_i are determined by the *dihedral sequence* of the chain. There are three dihedral conformations: helical (H) with A=0, B=C=D=1.2ε_H, extended (E) with A=0.9 ε_H, B=D=0, C=1.2 ε_H, and turn (T) with A=B=D=0, C=1.2 ε_H.

The energy $V_{\text{non-bonded}}$ describes the interaction between sequence-distant residues that are not covalently bonded. This term accounts for excluded volume interactions as well as the attractive forces between hydrophobic residues and is taken to be a sum of pairwise potentials:

$$V_{\text{non-bonded}} = \sum_{|i-j|\geq 3} 4\varepsilon_H S_1\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - S_2\left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \qquad (3.3)$$

where the value of the parameters S_1 and S_2 depends on the type of the residues *i* and *j*: $S_1 = S_2 = 1$ for BB interactions; $S_1 = 1/3$, $S_2 = -1$ for LL and LB interactions, and $S_1 = 1$, $S_2 = 0$ for NN, NB, and NL interactions.

Following the work of Head-Gordon's group[104-106], the bead sequence used for the ubiquitin-like model is LBLBLBLBLBNNLNBBBBBBBBNNLLBBLBBLLBN NLBBBBBNLBLBLLBNLBBLBBLNBBLBLBLBL and its dihedral sequence is EEEEEEEEHTHEHTEEEEHHEHHHHHHHHHHHEHTEEEEETHEEEETEETHHHHH HHEHHEEEEE. For sufficiently low temperature, this 68-residue chain reproducibly folds in the course of a Langevin dynamics simulation attaining a ubiquitin-like conformation.

Within our model, the interaction between the protein with a cylindrical pore whose radius is r_{pore} is given by:

$$V_{pore} = \sum_{C_\alpha} \frac{a}{1+e^{-z_i/l}} \left[1.0 - \frac{1}{1+\left(\frac{x_i^2 + y_i^2}{r_{pore}^2} \right)^{10}} \right] \tag{3.4}$$

where $a = 20\varepsilon_h$ and $l = \sigma/20$. Since this is a continuous function rather than a hard-wall repulsive potential, the parameter r_{pore} only roughly describes the pore dimension.

3.2.2. Simulating the protein dynamics

Following the previous study it was assumed that the dynamics of each atom in the chain are governed by the Langevin equation as details are provided in the section 2.2.2.

3.2.3. Calculation of free energy profiles

When a constant external force **f** is applied, the protein experiences a potential of mean force equal to $G_f(z) = G_0(z) - fz$. Here *z* is the reaction coordinate coupled to the force and $G_0(z)$ is the protein's free energy as a function of this reaction coordinate (in the absence of the force). The choice of the reaction coordinate depends on the process

considered, as shown in Fig. 3.1. In the case of translocation, $z \equiv z_1$ (or z_N) is the displacement of the N- (or C-) terminus, whichever is pulled mechanically, along the direction of pulling. In the case of mechanical stretching, $z \equiv | z_N - z_1 |$ is the protein extension along the direction of the stretching force **f.**

The replica exchange method[88, 89] was used to improve sampling statistics and avoid the trapping of the system in metastable states over the simulation timescale. The free energy as a function of the reaction coordinate *z* is given by

$$G_0(z) = -k_B T \ln p(z) \qquad (3.5)$$

where $p(z)$ is the sampled probability distribution of *z*. To obtain the global shape of $G_0(z)$ far away from the predominantly populated native state, the weighted histogram/umbrella sampling method was used[67-69].

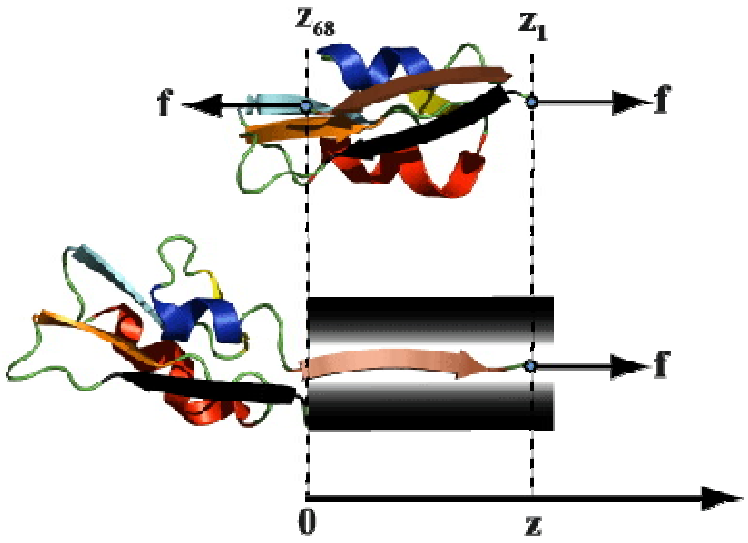

Fig. 3.1. The mechanical stretching reaction coordinate is the component of the end-to-end distance vector in the direction of the stretching force. The translocation coordinate is the displacement of the chain end along the axis of the cylindrical pore, relative to the pore entrance. The force **f** applied to this end acts along the axis.

3.3. RESULTS

3.3.1. Comparison of translocation and stretching

Free energy profiles $G_f(z)$ are shown in Figs. 3.2-4 for the case of mechanical stretching (Fig. 3.2) and for pulling the same protein through a pore by applying the force either at its N- (Fig. 3.3) or C-terminus (Fig. 3.4).

When one end of the chain is moved along the pore, $G_0(z)$ increases monotonically, reflecting the entropic and enthalpic costs to squeeze the protein, until the entire protein is inside the pore, after which $G_0(z)$ remains constant. At this point, the protein has achieved its maximally extended state attainable for a given pore diameter. When it first happens, the leftmost end of the chain is located at the tunnel entrance (Fig. 1); Therefore the value of the translocation reaction coordinate $z \equiv z_{1(N)} \approx 45\sigma$ measured relative to the tunnel entrance is the same as the protein extension $z \equiv | z_N - z_1 |$. It is then not surprising that the free energy cost $G0(z \approx 45\sigma)$ of achieving the same extension is similar for both translocation and stretching.

In the case of stretching, $G_0(z)$ rises abruptly if the domain is extended past $z \approx 45\sigma$ because at this point further chain extension involves a high enthalpic cost associated with a deformation of molecular bonds.

While the overall free energy cost of extending the protein is similar in each of the three cases presented in Figs. 3.2-3.4, the *shapes* of $G_0(z)$ and, consequently, the force induced unfolding mechanisms reflected in the shape of $G_f(z)$, are different in each case. For the case of translocation, the free energy profile $G_f(z)$ is a "downhill ramp" for large values of *z*, favoring translocation thermodynamically (See Figs. 3,4). The kinetics of translocation however depends on the applied force. For modest forces, squeezing the protein into the pore requires surmounting one or several free energy barriers. Similarly, the native-like (small *z*) and the extended (large *z*) conformations of protein mechanically stretched by a force are separated by one or more barriers (Fig. 3.2).

Local minima of $G_f(z)$ correspond to unfolding intermediates, which are different in all three cases.

When the protein is pulled at its N-terminus (Fig. 3.3), the resulting translocation proceeds via three distinct intermediates represented by structures 1, 2, and 3 in Fig. 3.3. For a sufficiently high pulling force, the rate limiting step corresponds to the barrier encountered between structures 1 and 2 and involves the peeling of a β-strand from the rest of the structure and its entrance into the pore. At lower force however the rate limiting step corresponds to the transition between the last intermediate, structure 3, and the extended state.

Similarly, when the C-end of the protein is pulled (Fig. 3.4), several intermediates are observed. As the force is increased, the last surviving barrier is that occurring early in the translocation process (the one between structures 1 and 2 in Fig. 3.4) while at low forces the rate limiting step takes place late in the translocation process and involves the squeezing of the tail of the chain into the pore.

Inspection of the intermediate structures encountered in the translocation process reveals that in each case the protein unravels from the end at which is pulled. This is easy to observe in the contact maps of the intermediate structures. A contact map here is a plot containing the points $\{i,j\}$ for each pair of residues i and j such that $|i-j|>3$ and $\left|\mathbf{r}_i-\mathbf{r}_j\right|<d$, where $d = 1.97\sigma = 7.5$ Å. In the case where the N-terminus (corresponding to the residue i=1) is pulled, the contacts $\{i,j\}$ with low values of i or j are destroyed first and the ones in the upper right corner of the map (corresponding to i and j close to N=68) survive last, indicating that the residues that are close the N-terminus are first to become separated from the rest of the domain. Similarly, when the C-terminus (i.e., the 68-th residue) is pulled, this end becomes separated from the rest of the domain first and the local structure involving the residues close to the N-terminus (i. e., with low values of i and j) is the last to disappear. These findings support the view that, in the course of translocation, the protein unfolds sequentially from the end containing the targeting sequence[10, 11, 97].

Certain similarities exist between the mechanical unfolding mechanism shown in Fig. 3.2 and the translocation pathways of Figs. 3.3-4. In particular, in each case the first step of unfolding involves separation of the two terminal parallel strands (the transition between the native-like structure 0 and structure 1). However in the case of translocation this step is followed by the Structure 1 to Structure 2 transition that involves a substantial barrier. This barrier is not found in the mechanical unfolding case. The physical origin of the Structure 1 to Structure 2 transition depends on which chain end is being pulled. For example, when the force is applied at the N-terminus, the local structure destroyed in this transition involves one of the α-helices; the same structure survives until late in the unfolding process in both the mechanical unfolding case and in the case of the C-terminus driven translocation. For sufficiently high forces, these observations are consistent with the view that the local stability of the part of the protein that enters first the translocation pore determines the overall resistance of the protein to mechanical unfolding[11, 96].

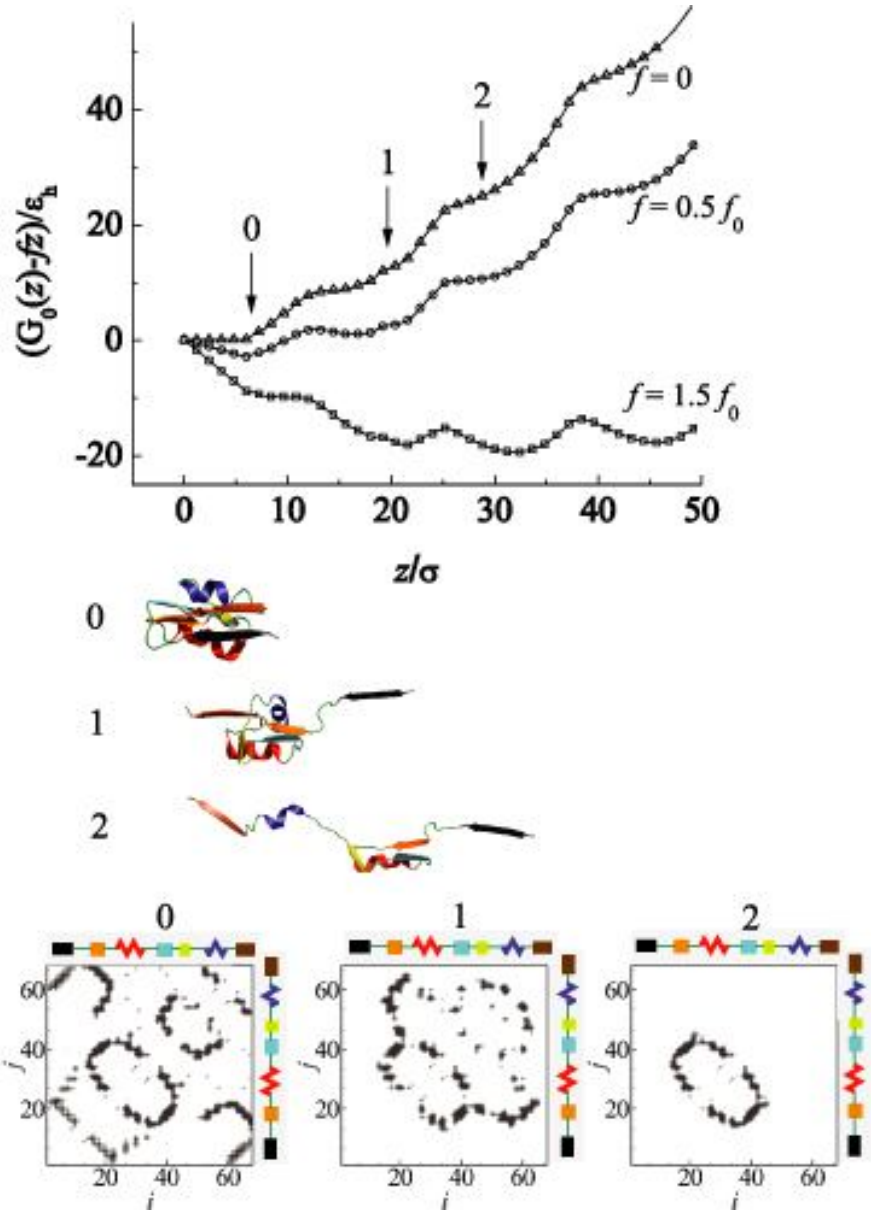

Fig. 3.2 The potential of mean force $G_f(z)$ is plotted as a function of the mechanical stretching reaction coordinate at different values of the stretching force. The force is measured in dimensionless units of $f_0 = \varepsilon_h / \sigma$. The darkness of each point reflects the probability of observing the corresponding contact in the equilibrium ensemble of conformations corresponding to the given extension *z*. Secondary structures (helices and strands), to which residues i and j belong, are shown along the i and j axes so that clusters of contacts on the map correspond to the proximity of secondary structure elements.

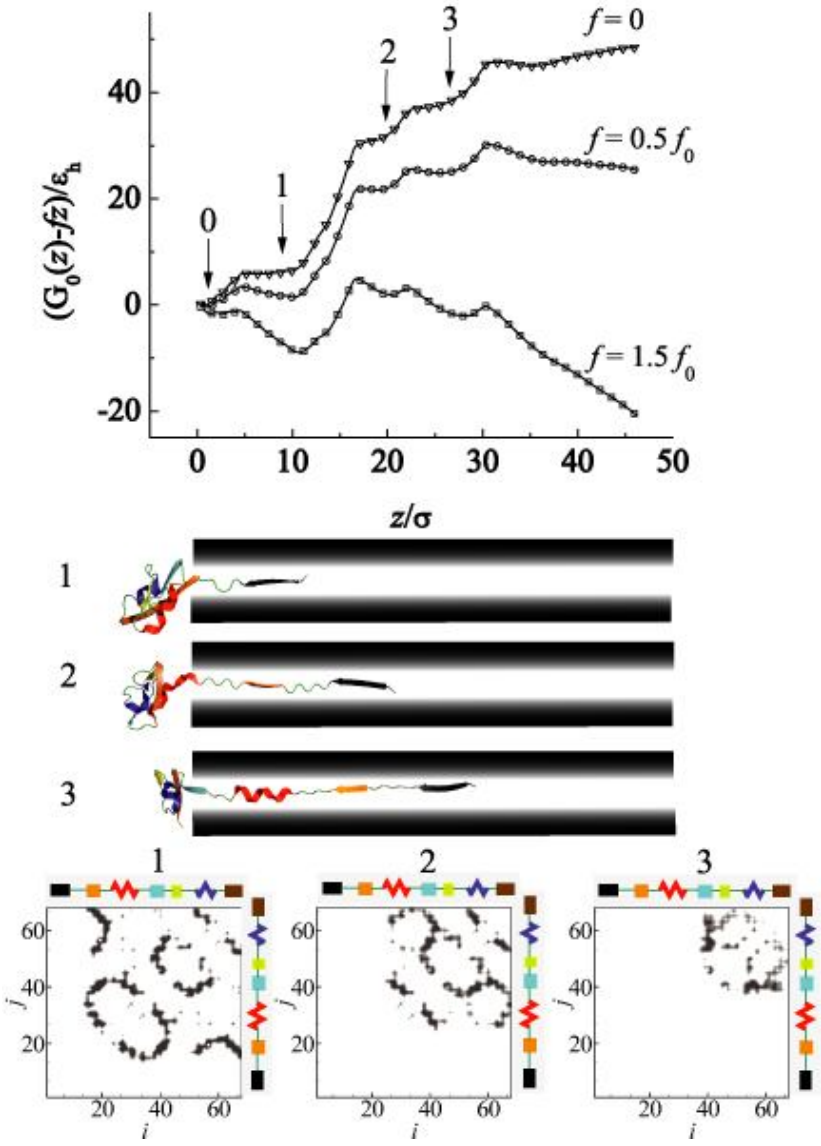

Fig. 3.3 The potential of mean force $G_f(z)$ is plotted as a function of the translocation coordinate equal to the position of the N-end of the chain along the pore at different values of the stretching force. The pore radius is $r_{pore} = \sigma$. The free energy is measured in units of ε_h and the force is measured in units of $f_0 = \varepsilon_h / \sigma$ (see the Methods Section) The minima of $G_f(z)$ correspond to translocation intermediates 1-3, whose structure is shown along with the corresponding contact maps. The native-like structure 0 not shown here is similar to structure 0 shown in Fig. 3.2.

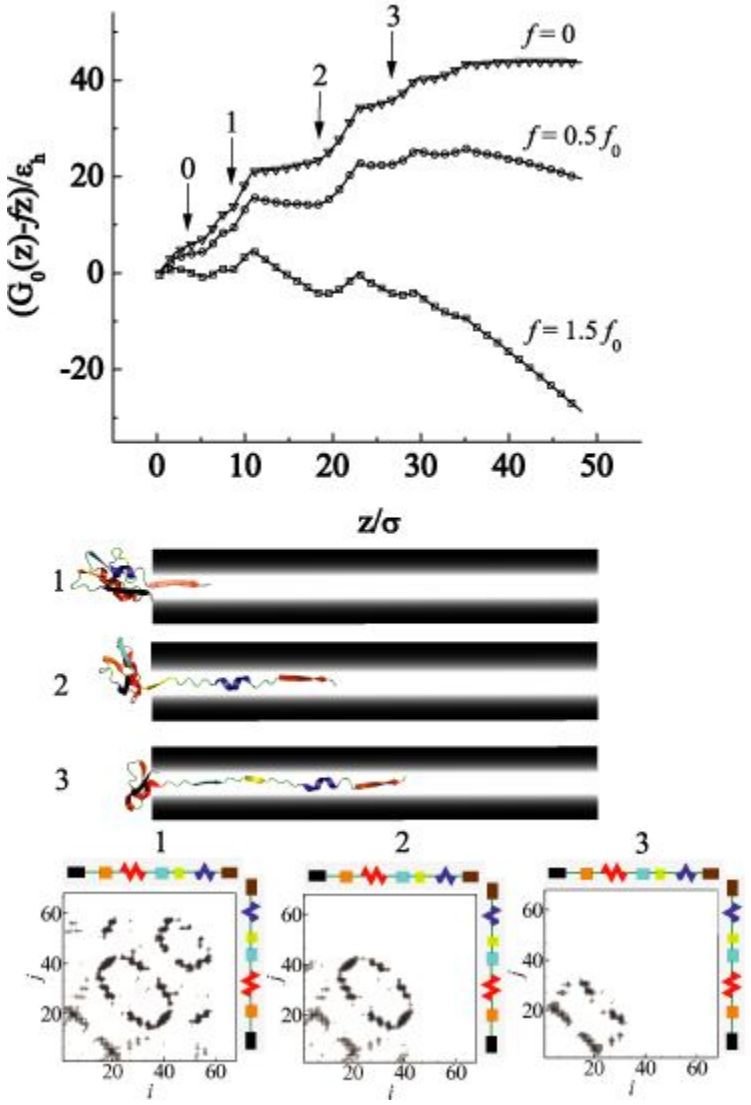

Fig. 3.4 The potential of mean force $G_f(z)$ is plotted as a function of the translocation coordinate equal to the position of the C-end of the chain along the pore at three different values of the stretching force. The pore radius is $r_{pore} = \sigma$. The free energy is measured in units of ε_h and the force is measured in units of $f_0 = \varepsilon_h / \sigma$ (see the Methods Section). The structure of representative translocation intermediates 1-3 corresponding to local minima of $G_f(z)$ is depicted and contact maps corresponding to those intermediates are plotted. The native-like structure 0 is shown in Fig. 3.2.

3.3.2. The translocation time as a function of the pulling force

Despite the fact that the free energy cost of translocation is the same regardless of which chain end is pulled, at finite values of the force *f* the time that it takes to translocate the protein would be different for whether the protein pulled at the N or C terminus. Time of translocation can be crudely estimated as $\tau_{translocation} \approx (\nu e^{-\Delta G_u(f)/k_B T})$ where $\Delta G_u(f) = \max G_f(z) - \min G_f(z)$ is the overall barrier encountered along the reaction coordinate and ν is a prefactor. The dependence of this translocation barrier on the applied force is shown in Fig. 3.5.

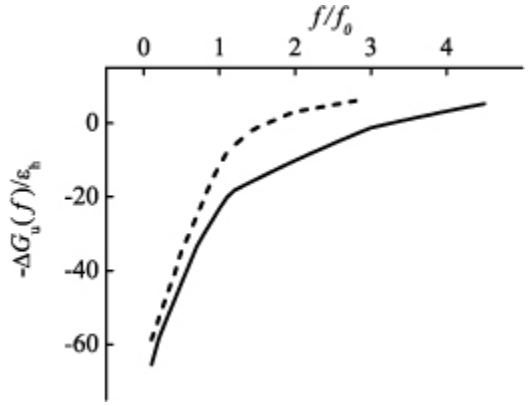

Fig. 3.5.The unfolding barrier $\Delta G_u(f)$ is plotted as a function of the pulling force applied to the N-terminus (solid line) and C-terminus (dashed line). The free energy is measured in units of ε_h and the force is measured in units of $f_0 = \varepsilon_h / \sigma$.

In mechanical pulling studies the unfolding free energy barrier is often assumed to be a linear function of the force[36, 110, 111]:

$$\Delta G_u(f) = \Delta G_0(0) - f\Delta z \tag{3.6}$$

The coefficient Δz has the simple meaning of the extension corresponding to the transition state relative to the native state. This approximation does not hold in Fig. 3.5. Instead, the slope of the function $\Delta G_u(f)$ undergoes an abrupt change corresponding to the transition between the "late" transition state (low force) to the "early" transition state (high force) scenario, which is accompanied by an abrupt change in Δz.

3.3.3. The pore size effect

The dimensions of various translocation channels span a considerable range and the width of each individual channel also often varies along the channel. Fig. 3.6 shows the free energy profile $G_0(z)$ for three different values of the pore radius, $r_{\text{pore}} = \sigma$ (Case 1), 1.5σ (Case 2), and 2σ (Case 3), for the protein that is pulled at its N-terminus. The difference between cases 1 and 2 is quantitative – the narrower the pore the higher the free energy barrier – but not qualitative.

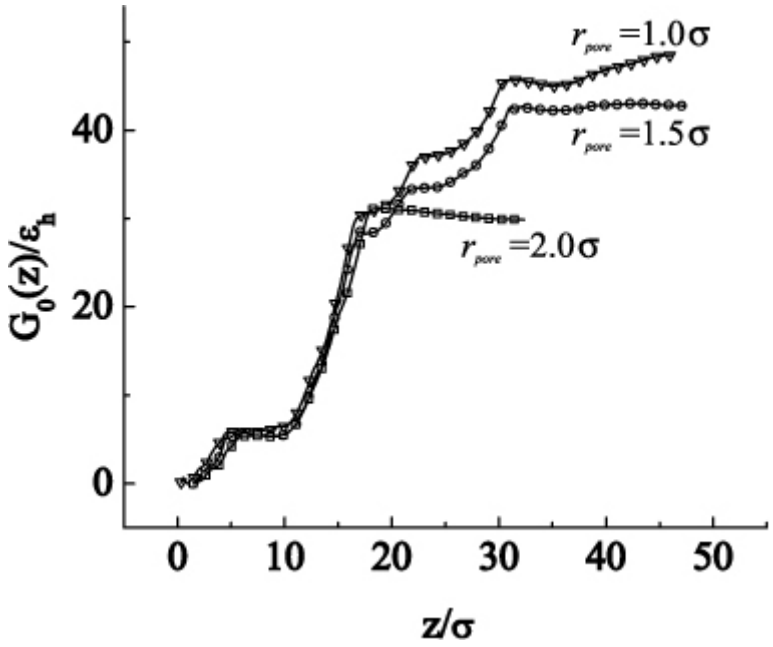

Fig. 3.6. The potential of mean force $G_0(z)$ for different values of the pore diameter

Case 3 is different. The pore is now too wide for the protein to attain a linear, extended conformation without any tertiary structure. In Fig. 3.7, which shows $G_f(z)$ for this case at different values of the force, one finds only one intermediate (Structure 1), which is very similar to Structure 1 in Fig. 3.3. In other words, the initial step of the translocation process, in which a β-strand becomes separated from the rest of the domain that still remains outside the pore, is the same for wide (Case 3) and narrow (Case 1) pores. However, further stages of translocation in Case 3 are different. As seen from Fig. 3.7, the largest barrier associated with translocation involves the squeezing of a partially

folded protein into the pore (Structure 1 to Structure 2 transition in Fig. 3.7). Once the protein inside the pore, it assumes the partially folded Structure 2 and moves along the pore without further unfolding. Notice that this final Structure 2 is different from any of the intermediates observed for the narrow pore (Cases 1 and 2).

A remarkable feature observed in Fig. 3.6 is that for $z \leq 17\sigma$ the shape of $G_0(z)$ is independent of the pore size. As a consequence, in the large force limit (corresponding to the "early" transition state in Fig. 3.3, see the case of $f = 1.5\varepsilon_h / \sigma$) the shape, the height and the location of the translocation barrier is the same regardless of the pore size (cf. Figs. 3.3 and 3.7 for $f = 1.5\varepsilon_h / \sigma$). We conclude that the translocation time will be independent of the pore size (in the range of the pore sizes studied) in this regime. The translocation barrier and consequently the translocation time will become dependent on the pore size when the force is lower (cf. Figs. 3.3 and 3.7 for $f = 0.5\varepsilon_h / \sigma$).

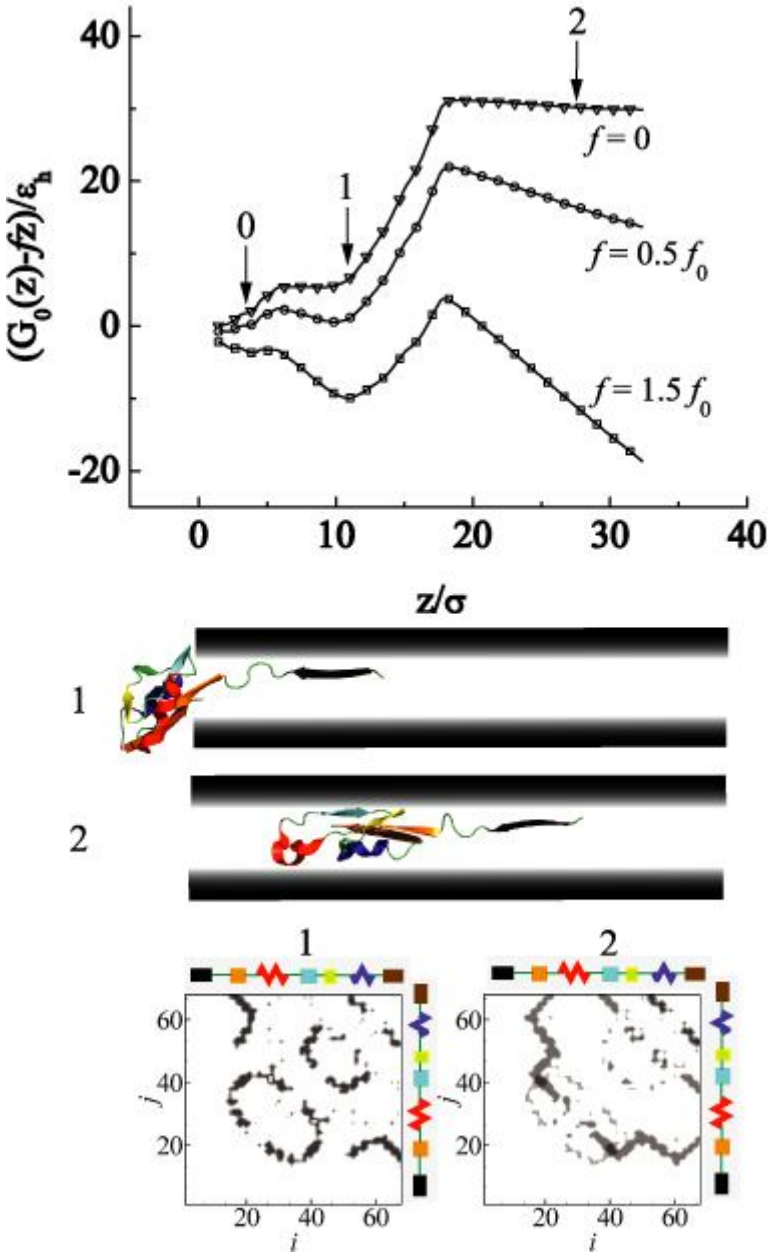

Fig. 7. The potential of mean force $G_f(z)$ is plotted as a function of the translocation coordinate equal to the position of the N-end of the chain along the pore at different values of the stretching force for pore size r=2σ. Also shown are representative structures encountered in the course of translocation, along with the corresponding contact maps.

3.3.4. Comparison of translocation of the protein and a homopolymer

Given that most theoretical work on translocation has previously focused on homopolymers[78, 81, 82, 112-117], it is instructive to compare the translocation of ubiquitin with that of an unstructured, random-coil-like homopolymer of the same length. This comparison is shown in Fig. 3.8. Fig. 3.8A presents the translocation free energy profiles $G_0(z)$ for both cases. The free energy cost of squeezing a homopolymer into the pore is much lower than that for ubiquitin. In the homopolymer case, the barrier in the potential $G_0(z) - fz$ (not shown) disappears at a low force $f \approx 0.4\varepsilon_h / \sigma \approx 7.2\,pN$, while in the case of ubiquitin, a substantial barrier exists even at a much higher force of $f \approx 1.5\varepsilon_h / \sigma$.

The difference between the two cases is qualitative rather than just quantitative. Our homopolymer essentially behaves as a random coil and the free energy cost required to accommodate it inside the pore has largely an entropic origin since the entropy of the polymer constrained by the pore is lower than that of the free random coil. The situation is different in the case of a folded domain. As it enters the pore, the resulting change in the entropy is a result of two opposite trends: Unfolding of the domain is associated with an entropy increase; however, confinement of the unfolded domain within the pore results in a decrease in entropy. Depending on the pore radius, the entropy reduction due to confinement may or may not be larger than the entropy of unfolding. Consequently, the protein inside the pore may have entropy that is higher than its entropy outside the pore.

This is indeed the case here. Fig. 3.8B shows the entropy $S(z)$ of both ubiquitin and the homopolymer as a function of the translocation coordinate. The entropy was calculated by using the relationship $S(z) = \left(\left\langle V(\mathbf{r}_1,...,\mathbf{r}_N)\right\rangle_{z_1=z} - G_0(z)\right)/T$, where $\left\langle V(\mathbf{r}_1,...,\mathbf{r}_N)\right\rangle_{z_1=z}$ is the polymer's energy with the z-displacement of its first bead constrained at z. We see that, unlike the case of the homopolymer, the entropy of the domain inside the pore is higher than that of the domain outside (for the value of the pore

radius used). The main origin of the free energy barrier in the case of ubiquitin comes from the energetic cost of denaturing the protein.

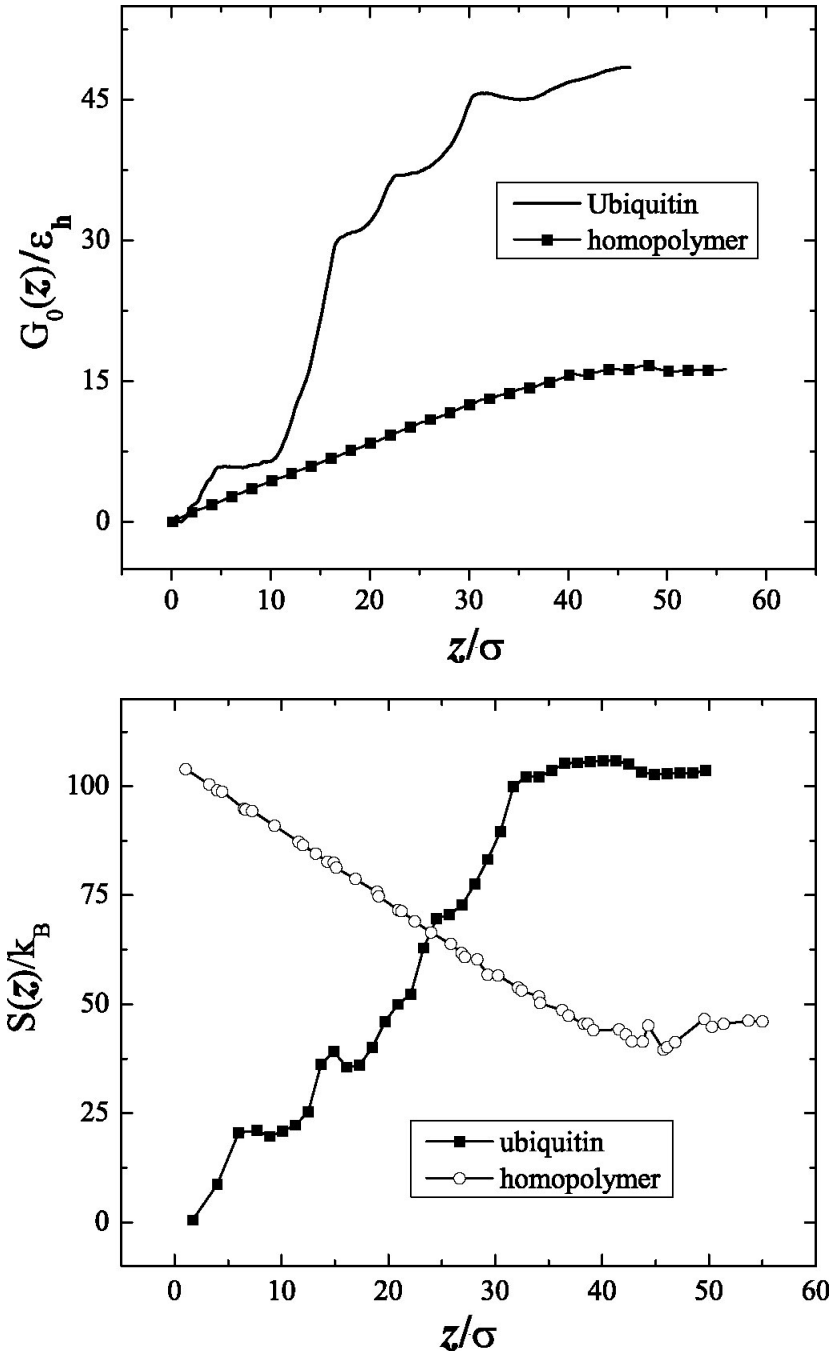

Figure 3.8 (A) Free energy $G_0(z)$ and (B) entropy $S(z)$ as a function of the translocation coordinate z for ubiquitin and for a homopolymer model that consists of neutral beads.

3.4 DISCUSSION

We have studied how a protein unfolds when being pulled across a long narrow cylindrical pore. The observed unfolding mechanism is found to be different than that probed by the single molecule experiments. We found that it depends on the diameter of the pore as well as on whether it is pulled from the C- or N- terminus. The translocation kinetics exibit a pulling force dependence that is more complex than a simple exponential function expected on the basis of simple phenomenological models of translocation. The models that can succesfully model the force dependence of unfolding kinetics in single molecule unfolding fails to explain translocation kinetics. Also this abrubt change of transition state on force is specific to protein translocation and can not be captured in polymer translocation.

It should be noted that our study has several limitations that may prevent its direct comparison with experiments. First of all, we assume that peptide-pore interactions are only repulsive, while most of the protein pores have electrostatic and hydrophobic sites that binds to the protein[53] may favor domain unfolding it even the pore size is larger. Also very long pore with constant diameter would be an incorrect representation of the pores in biology. Secondly, the assumption that translocation is a slow barrier crossing type project may not be correct if timescale of translocation dynamics is comparable to the time scale of internal dynamics of protein. In that situaion one-dimensional diffusion in the equilibrium free energy is no longer applicable and full-blown simulation of translocation dynamics may be required.

Chapter 4

Single Molecule Electrophoresis of β-Hairpin Peptides across α-Hemolysin Pore*

4.1. INTRODUCTION

Electrophoretically driven translocation of proteins across certain biological or synthetic pores can be studied experimentally at the single molecule level. The technique has originally been developed for DNA and RNA[118-120] and for other polymers[51, 121-123], and recently progress has been made towards studying proteins inside transmembrane pores[53]. Transmembrane pores such as α-hemolysin[124] (α-HL) are convenient model translocation systems and they can potentially be used as biosensors. Whenever a single protein is inside the pore, it partially or completely blocks the current accross the pore; The duration of such blocking events directly reports on the time of translocation. It is therefore of interest to examine the dependence of the translocation time on the driving force (that would be proportional to the electric field).

In this chapter, we will summarize a joint work of single-channel electrical recordings and Langevin molecular dynamics simulations that were used to probe the translocation of β-hairpin peptides[125] of the B1 domain of protein G through an α-HL protein pore at the single-molecule level. The α-HL pore is a heptameric mushroom-shaped protein of known crystal structure[124] and remains open for long periods in an applied transmembrane potential[122]. Experimentally, three β-hairpin peptides were selected for this preliminary study: (i) the wild-type peptide (G41) has the sequence GEWTYDDATKTFTVTE (residues 41-56) [125], (ii) Ac-G40 has acetylated Gly at the N-terminus (residues 40-56), and (iii) K41 has a Lys residue in position 41 (residues 41-56). The three β-hairpins explored in this work, vary by the number of terminal ion pairs, and have 30%, 5% and 50% stabilities in aqueous phase. The β-hairpin peptides, when

* Large portions of this chapter has been published in references 54, 95.

added to the *trans* side of the bilayer at low micromolar concentrations, produced transient current blockades, the nature of which depended on the features of the β-hairpins.

In parallel to these single-channel studies of β-hairpin peptides, we have performed Langevin dynamics simulations of peptide translocation using a minimalist off-lattice model[104-106] for both the β-hairpins and the β-barrel part of the α-HL pore. We have designed sequences of four β-hairpin peptides that differ in their stability and their hydrophobic content. Residues in each of the peptides were assigned charges corresponding to those in the experiment. The results show that our minimalistic model of protein translocation is very succeful to predict the outcome of experiments and it also can provide molecular level detail of protein translocation process.

To study the translocation rates of proteins accross the α-HL pore we resort to a completely different approach than the previous chapter. In Chapter 3 we report our study on a protein domain that is pulled mechanically through a neutral pore by applying a force to one end of the polypeptide chain. The pore is narrow enough that it requires the protein to unfold to allow its passage. This results in a large free energy barrier. The translocation is then a rare event that is unlikely to be observed in a molecular simulation of a realistic duration. Thus we estimated the translocation rate indirectly by considering the free energy barriers encountered by the protein as it moves along the pore and using transition state theory for the rate of barrier crossing.

In this chapter we will report our simulation of electric-field-driven translocation of a short charged peptides across the α-HL pore. The pore is large enough to accommodate the folded b-hairpin geometrically[70]. When an electric field is applied, the translocation process is expected to be relatively fast and its time scale can be captured directly in a computer simulation. Therefore, the distribution of the translocation times was directly computed for this case.

4.2. MODEL AND METHODS

4.2.1. The protein and the pore model

A minimalist off-lattice model[104-106] was used to describe the energy of the peptide and its interaction with the pore. The peptide potential, as a function of the position $\mathbf{r}_i$, i=1, …, N, of each residue, is given by:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V_{\text{bond}} + V_{\text{bend}} + V_{\text{dih}} + V_{\text{non-bonded}} + V_{\text{pore}} \quad (4.1)$$

The details of the potential V_{bond} and V_{bend} that accounts for bond stretching and bond bending inherent to the peptide bond geometry is summarized in section 2.2.2. and the potential terms for dihedral angle constraint V_{dih} and, non-bonded interactions $V_{\text{non-bonded}}$ are detailed in section 3.2.2 of this dissertation.

Four β-hairpin forming peptide models were constructed. The bead sequences for each peptide were as follows:

Peptide 1: BBLBLBNBNNBNBLBB (82% folded, 56% hydrophobic)

Peptide 2: NBLBNBNLLNBLBNBN (70% folded, 37% hydrophobic)

Peptide 3: NBLBLBNNNNBLBLBN (54% folded, 37% hydrophobic)

Peptide 4: BNLBLBNNNNBLBLBN (32% folded, 37% hydrophobic)

The same dihedral sequence EEEEETTTEEEEE was used for each peptide. Each of the four peptides formed a hairpin structure similar to that of the β-hairpin fragment of the B1 domain of protein G (pdb file 1PGB) at low enough temperatures. We have further assumed that the beads 7 and 16 are positively charged and the beads 1, 10, 11, 15 are negatively charged so that the hairpin carries an overall charge of -2. The electrostatic interactions between these charged residues were not explicitly included in the model since they were considered to be part of the total potential energy $V(\mathbf{r}_1,...,\mathbf{r}_N)$ described above.

The model of the α-HL pore used here was obtained by placing beads at the positions of the α-carbon atoms from the X-ray structure of the α-HL pore (pdb file 7αHL) [126]. The beads were assigned to be hydrophobic, neutral, or hydrophilic depending on the hydrophobicity of the residues in the pdb file (see Figure 4.1). The pore geometry is fixed (i.e., the pore beads do not move). The bead model only includes the β-

barrel region (i.e., the narrow part) of the α-HL pore. The potential V_{pore} describing the peptide-pore interactions is of the same form as $V_{\text{non-bonded}}$ above.

In the beginning of the simulation, the peptide was placed outside the narrow part of the α-HL pore. By artificially creating a "vestibule" (shown by dashed lines in Fig. 4.1) with an external potential that confines the peptide within a cylinder of radius $R=6\sigma$, we have ensured that the peptide cannot wander too far from the pore entrance and so we are likely to see translocation to occur in the course of a simulation. However R was chosen large enough to allow equilibration before translocation takes place and to ensure that the resulting translocation time does not depend on the precise manner in which the peptide was initially positioned. The presence of this potential affects the frequency with which the peptide attempts to translocate across the narrow part of pore, but not the translocation time itself. A force $0.5f$ ($-0.5f$) in the direction of the pore axis is applied to the positively (negatively) charged residues so that the interaction between the peptide and the electric field is described by the term

$$V_{pull} = -f(z_7 + z_{16} - z_1 - z_{10} - z_{11} - z_{15})/2 \qquad (4.2)$$

and the total force acting on the peptide is f. Here z_i is the z-coordinate of the i-th bead, where z direction coincides with the pore axis.

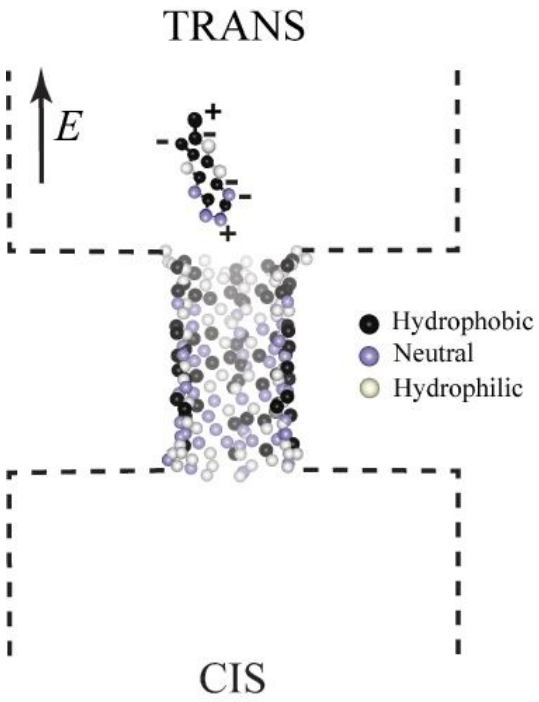

Fig. 4.1. β-hairpin peptide at the entrance of the narrow part of the αHL pore. Translocation is driven by the electric field that acts in the direction along the pore. Both the β-hairpin and the pore consist of beads that can be hydrophobic, hydrophilic, or neutral. In addition, 6 residues of the hairpin carry charges, as indicated. The dashed lines show the additional confinement potential that prevents the peptide to wander too far from the pore entrance.

4.2.2 Simulating the protein dynamics

It was assumed that the dynamics of each atom in the chain are governed by the Langevin equation of the form

$$m\ddot{r}_i(t) = -\xi\dot{r}(t) - \frac{\partial(V + V_{\text{pull}})}{\partial r_i} + R(t) \qquad (4.3)$$

where $\mathbf{r}_i$ is the position of the *i*-th atom, *m* is its effective mass, $\xi = 2.0\left(\sigma^2 / m\varepsilon_{\text{h}}\right)^{-1/2}$ is the friction coefficient, and $\mathbf{R}(t)$ is a random δ–correlated force satisfying the fluctuation-dissipation theorem. This equation was solved by using the velocity Verlet algorithm as described by Paterlini et al[87]. We use dimensionless units of energy, distance, time, and

force respectively equal to ε_h, σ, $\tau = (m\sigma^2/\varepsilon_h)^{1/2} \cong 2.5\,\text{ps}$, and $f_0 = \varepsilon_h/\sigma \cong 18\,\text{pN}$. All simulations were performed at temperature T=0.28 ε_h/k_B. We note that the above value of the friction coefficient is 25 times lower than that expected for water[86, 127]. However because it is high enough to ensure overdamped regime of the dynamics (where the acceleration term in the Langevin equation can be neglected), the time scales of the dynamics in water are recovered by simply rescaling all the times by a factor of 25.

4.3 RESULTS

4.3.1. Electrical recordings of Translocation of β-hairpin Forming Peptides

Electrical recordings of single molecule experiments revealed an exponential decrease of the event duration with the applied transmembrane potential (Figure 4.2). Importantly, for similar electric forces, events produced by the unfolded Ac-G40 peptide were shorter than those values measured for the folded G41 and K41 peptides (Figure 4.2).

4.3.2. Langevin Dynamics Simulation of Translocation of β-hairpin Forming Peptides

Similar to the experimental data, the electric field dependence of the translocation time is close to exponential at very low fields. Such an exponential dependence can be understood within the framework of Bell's model[110], where escape from the pore is controlled by crossing a free energy barrier that is lowered by the electric force. This description becomes invalid at high fields where the barrier vanishes, as seen in Fig. 4.3. The translocation time is correlated with the stability of the peptide and decreases monotonically from peptide 1 to peptide 4. The somewhat counterintuitive observation that translocation is faster for less stable peptides can be understood if one considers the three typical translocation trajectories observed in the simulations (Figure 4.3B). In the trajectories of type 1, the peptide remains in an extended conformation and the residues are threaded through the pore in a single file. This type of trajectories corresponds to the fastest observed translocation times. In type 2 trajectories, corresponding to longer dwell times, the peptide remains folded as it goes through the pore. Type 3 trajectories, which were only observed for the most hydrophobic peptide 1, correspond to the slowest translocation events, in which a peptide-pore complex is formed and the peptide remains

trapped inside the pore in a misfolded conformation. Less stable peptides are more likely to enter the pore in an open conformation, resulting in a fast type 1 translocation event. In

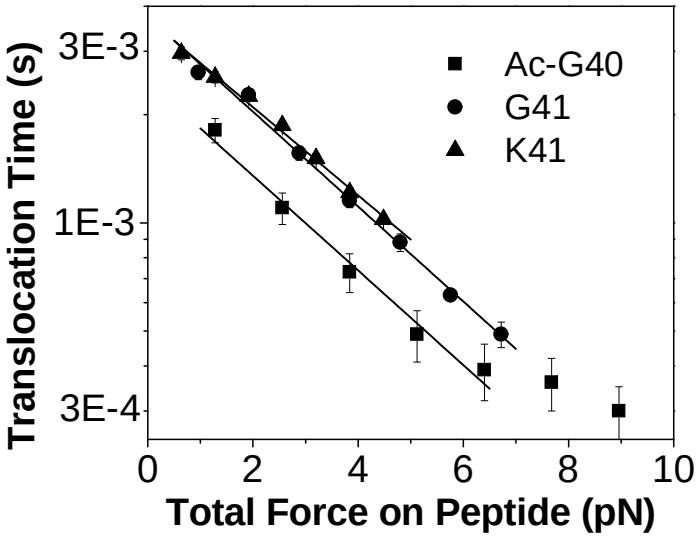

Fig. 4.2. The dependence of the translocation time of β-hairpin peptides on the electric force. The electric force was derived as $F=qV/l$, where the effective charge q is half of the peptide charge, V is the transmembrane potential, and l=50 Å is the length of the β-barrel lumen.

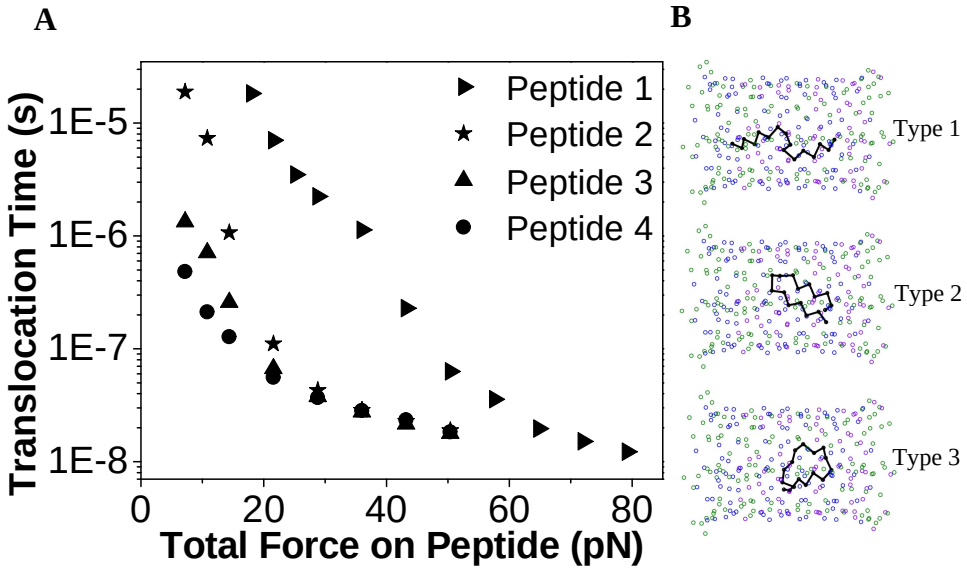

Figure 4.3. (A) The computed mean translocation time as a function of the total force on the β-hairpin. (B) Typical conformations of the peptide inside the β-barrel part of the α-HL pore corresponding to three types of observed translocation trajectories.

contrast, translocation of the more stable peptides is more likely to occur via the type 2 or 3 trajectories.

To gain better insight into the origin of the translocation time dependence on the peptide stability, we have studied the statistics of individual translocation events (Figure 4.4). To automatically distinguish between the open, folded and misfolded conformations of the peptide, we have used the method of Baysal et al. [128], in which the overall shape of a molecule is characterized by calculating the projection of its configuration onto the principal axes. If the conformation of the chain remains the same (open, folded, or misfolded) during at least 50% of the total translocation time then the trajectory is considered to be type 1, 2, or 3, respectively.

Some trajectories cannot be classified as one of the above three types because the peptide undergoes multiple transitions among different conformational types in the course of a single translocation event. However, the relative number of such trajectories tends to be small (<12%).

Shown in Figure 4.4 are histograms of the translocation times that correspond to each type of translocation event. Each histogram is normalized such that the sum of the bar heights gives the total fraction of the observed trajectories corresponding to the given type. Also shown are the histograms of translocation times corresponding to any type of event (the bar heights add up to one). Type 1 translocation events are the fastest and type 3 the slowest. Type 3 trajectories are only observed for the most stable and the most hydrophobic peptide 1, for which the number of translocation events of type 1 is negligible. As the stability is decreased from peptide 2 to peptide 4, the relative number of fast type 1 trajectories goes up and the fraction of the slower type 2 trajectories decreases.

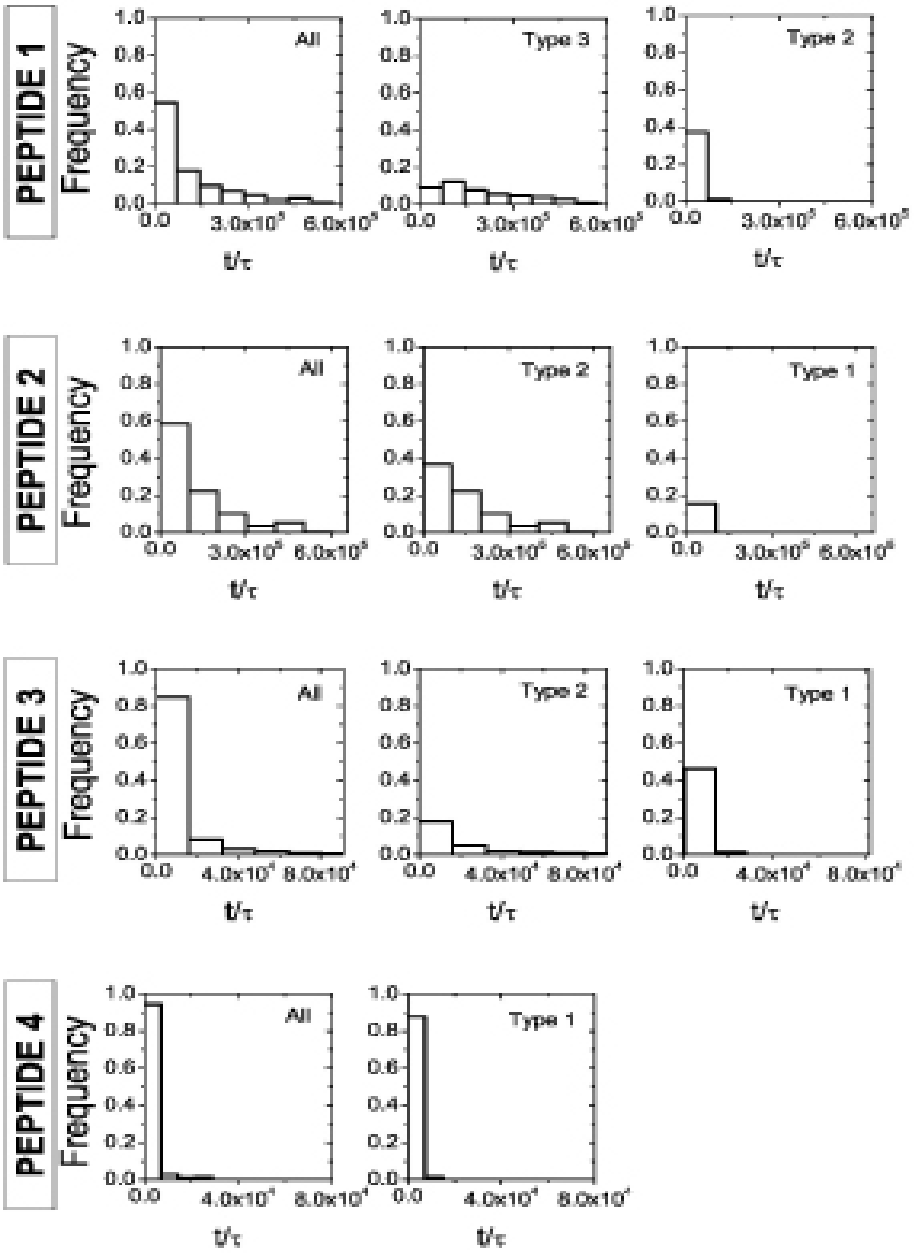

Fig. 4.4. Histograms of translocation times for the four peptides. The value of the total force on the peptide is f=1.2f_0 in the case of Peptide 1 and f=0.8f_0 for peptides 2-4.

4.3.3. Force dependence of probability distribution of the translocation times

We also studied the translocation times and probabilty distributions for different values of the driving force *f*. Histograms of the observed translocation times for peptide 1 are shown in Fig. 4.5 For low values of the force *f*, the probability distribution of the translocation time *t* is exponential,

$$p(t) = \tau_{tr}^{-1} \exp(-t/\tau_{tr}), \tag{4.4}$$

where $\tau_{tr} = \langle t \rangle$ is the average translocation time. Such Poisson statistics of translocation times is usually a signature of a barrier crossing process, whose time scale is determined by a single rate limiting step involving overcoming a barrier that is higher than the thermal energy [129].

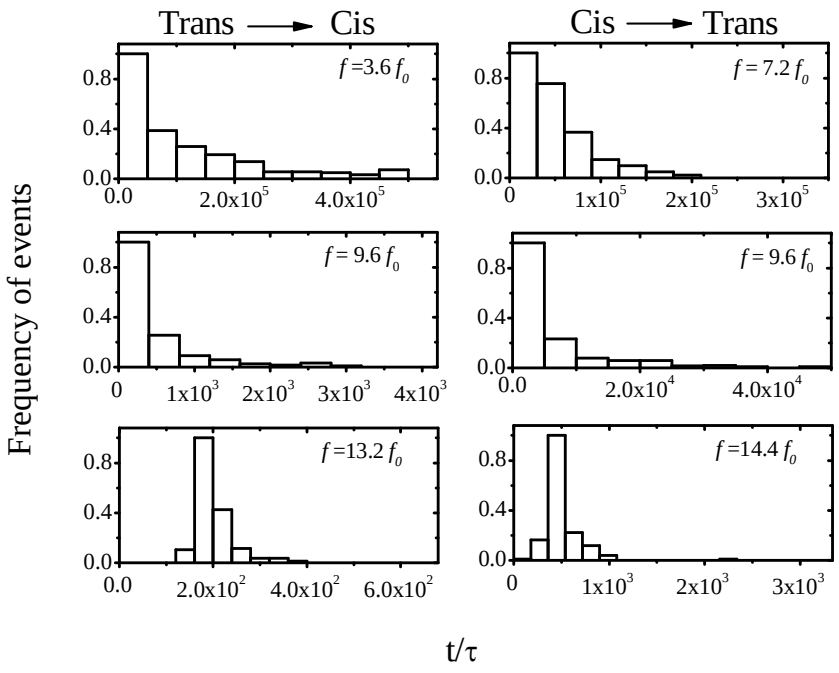

Figure 4.5. Histograms of the time spent by the peptide inside the narrow part of the α-HL pore for different values of the driving force. Left: trans-to-cis translocation Right: cis-to-trans translocation.

At higher values of the force $p(t)$ is no longer exponential but rather exhibits a single maximum. This situation is expected in the case of a "downhill" process that involves no barriers. To illustrate this, consider a simple one-dimensional model [113, 130], in which we view translocation as a diffusive motion of a particle in a pore-induced potential $V(x)$, where x is the translocation coordinate. In the presence of a driving force f, the dynamics is described by a one-dimensional Langevin equation of the form,

$$m\ddot{x} = -\frac{d}{dx}(V(x) - fx) - \gamma\dot{x} + F_r(t), \tag{4.5}$$

where $F_r(t)$ is a random δ-correlated force satisfying the fluctuation-dissipation theorem. Even if $V(x)$ has barriers, for a sufficiently high force f the potential $V(x)$-fx will be a monotonic function decreasing in the direction of the force. The average time for such downhill diffusion between the pore entrance and its exit can then be estimated by averaging over thermal noise in the above equation. If we assume $\langle V(x)\rangle = V\left(\langle x\rangle\right)$, which is valid to 1st order in fluctuations ($x - \langle x\rangle$), then averaging gives the deterministic equation

$$md^2\langle x\rangle / dt^2 = -\frac{d}{dx}(V(\langle x\rangle) - f\langle x\rangle) - \gamma d\langle x\rangle / dt \tag{4.6}$$

The effect of the thermal noise will be to broaden the distribution of the translocation times around the mean value obtained from this deterministic equation of motion, producing a translocation time distribution similar to the ones seen in Fig. 4.5 for high values of the force.

In an overdamped regime (high value of the friction coefficient γ) the lhs of the above equation is close to zero. Furthermore, at high values of the force the term $V\left(\langle x\rangle\right)$ can be neglected (as compared to $-f\langle x\rangle$), which gives motion at a constant velocity that is proportional to the force $d\langle x\rangle / dt = f/\gamma$. We therefore expect the mean translocation time to be inversely proportional to the force in this limit.

The dependence of the translocation time on the applied force is shown in Fig. 4.6. Trans-to-cis translocation is faster than that in the cis-to-trans case. As seen from

Fig. 4.6A, at low forces (corresponding to the case where the translocation time distribution is exponential in Fig. 4.6) the average translocation time is well described by the equation

$$\langle t \rangle = \tau \exp(-af / f_0 + b) \ , \qquad (4.7)$$

consistent with Bells' phenomenological model [110], which assumes a free energy barrier that depends linearly on the force and is used to describe many force-dependent molecular processes [129]. The straight lines in Fig. 4.6A are described by the following parameters: a = 0.783, b=16.35 for the cis-to-trans case and a= 0.990, b= 15.42 for the trans-to-cis case. Note that the value of b, which describes the translocation time in the limit of zero force, $\langle t \rangle = \tau e^b$, is the same (to within extrapolation uncertainties/statistical errors) for the cis-to-trans and the trans-to-cis translocation.

At higher forces, the above exponential dependence breaks down. In this regime the inverse translocation time is a linear function of the force, $\langle t \rangle^{-1} = Af - B$, as seen from Fig. 4.6B, which plots $\langle t \rangle^{-1}$ vs. f for the same data. This behavior is consistent with the above arguments based on a one-dimensional model.

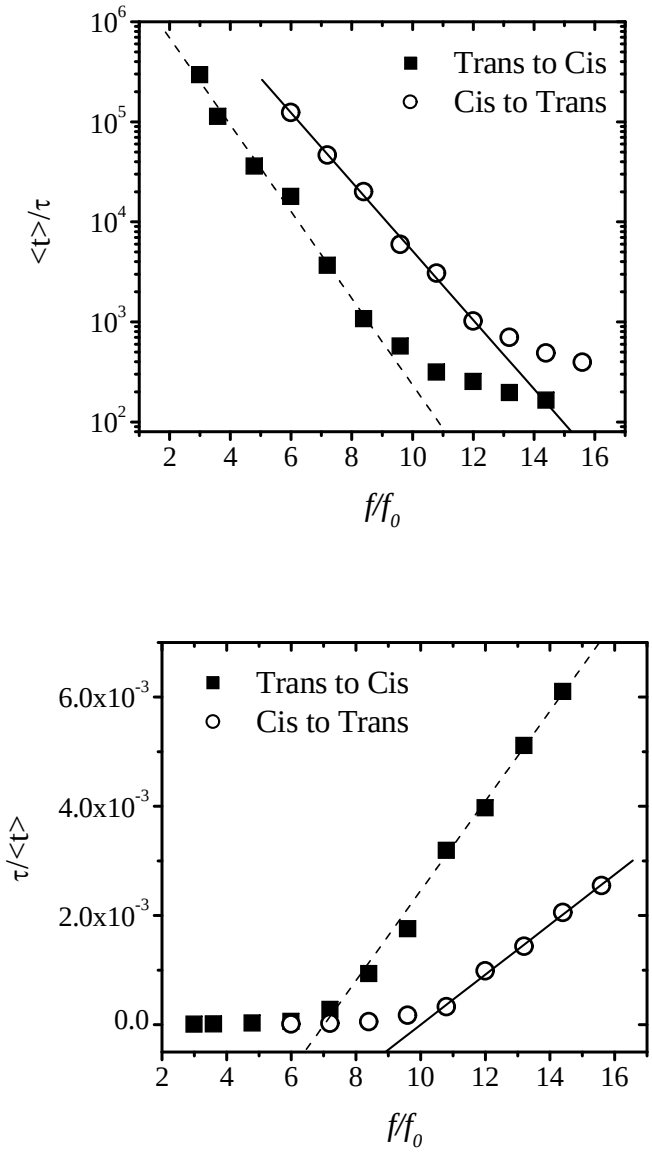

Figure 4.6. (A) The dependence of the average time $\langle t \rangle$ spent by the peptide 1 inside the narrow part of the α-HL pore on the driving force. The solid and the dashed lines are linear fits of the form $\ln \langle t \rangle / \tau = -af / f_0 + b$. (B). Same data as in Fig. 4.6(A) but replotted in the form $\tau / \langle t \rangle$ vs. *f*. At high forces, the data are well described by the linear relationship $\langle t \rangle^{-1} = Af - B$, where $A>0$ and $B>0$.

4.4. DISCUSION

The translocation trajectories observed in the simulations provide a plausible explanation of the experimental correlation between the β-hairpin stability and the translocation time. Coarse grained peptide models are not realistic enough to allow a direct quantitative comparison with experimental data and CPU limitations commonly necessitate using electric forces that are higher than those employed experimentally. Nevertheless simulated translocation times and the experimental translocation times extrapolated to higher forces are within an order of magnitude from one another and show similar force dependence, suggesting that our model may correctly capture the translocation mechanism.

These observations, combined with recent single-channel studies on the translocation of α-helical peptides through the α-HL pore [53, 131], reveal the complexity of the translocation mechanism and its sensitivity to the details of the peptide-pore interactions as well as to the secondary structure of the peptide. It is also conceivable that the kinetics of the translocation of polypeptides through protein pores is dependent on the features of the pore lumen [79].

We finally note that unlike the neutral pore that was studied in the case of ubiquitin translocation, the inner surface of the α-HL pore is rather "sticky" as it contains about 36% of hydrophobic residues that have attractive interactions with the peptide, which is 56% hydrophobic. This substantially slows down the translocation process. Specifically, when we replaced all the pore residues by neutral ones, this has sped up the translocation process by two orders of magnitude. This suggests that more realistic molecular models are desirable in order to obtain a complete physical picture of protein translocation.

Chapter 5

Topography of the Free-energy Landscape Probed via Mechanical Unfolding of Proteins*

5.1. INTRODUCTION

In recent years, there have been numerous experimental studies of the mechanical unfolding of single protein molecules [29, 35-37, 39, 40, 47, 48, 93, 94, 101, 102, 111, 132-146]. One of the goals of these studies is to understand the mechanisms through which some proteins perform mechanical, "load-bearing" functions in living organisms [40, 132, 147, 148] and to elucidate the relationship between their structure and mechanical function. The other goal is to shed new light on the fundamental protein folding problem. However it is not yet entirely clear what kind of information about folding and/or chemical denaturation of proteins can be extracted from the mechanical unfolding experiments. Mechanical stretching can generally alter protein unfolding pathways by biasing free energy landscapes [43]. Experimentally, the relationship between the chemical and the mechanical unfolding mechanisms has been addressed through a comparison of the unfolding rate constant k_u^{chem} measured in chemical denaturation experiments with the mechanical unfolding rate $k_u(f)$ extrapolated to zero pulling force $f \to 0$. Coincidence of these two rates was taken as evidence that the mechanical and chemical unfolding transition states are the same. While k_u^{chem} and $k_u(0)$ were indeed found to nearly coincide for the immunoglobulin domain I27 [111], other studies involving mutant domains suggested that these rates can be different and therefore the mechanical and chemical unfolding pathways may not be the same [133, 145]. Furthermore, molecular dynamics studies carried in the references [59, 60] suggest that the similarity between k_u^{chem} and $k_u(0)$ may be accidental: The experimental estimates of $k_u(0)$ are based on the assumption that $\log k_u(f)$ is a linear function of the force f [48,

* Large portions of this chapter have been published in reference 109.

110]; however this is not the case for more realistic free energy profiles resulting in values of $k_u(0)$ that may be orders of magnitude different from those originally estimated.

Interpretation of the experimental data provided by force probe spectroscopy thus depends on the model chosen for the free energy profile of the protein along the mechanical reaction coordinate. There are two key assumptions that are commonly made to facilitate interpretation of pulling experiments. (i) Firstly, one assumes that the unfolding free energy barrier has a linear force dependence [110], $\Delta G_u(f) = \Delta G_u(0) - f\Delta z_u$, where Δz_u is the protein extension corresponding to the transition state. This assumption leads to the exponential dependence of $k_u(f)$ mentioned above and also allows one to estimate Δz_u.

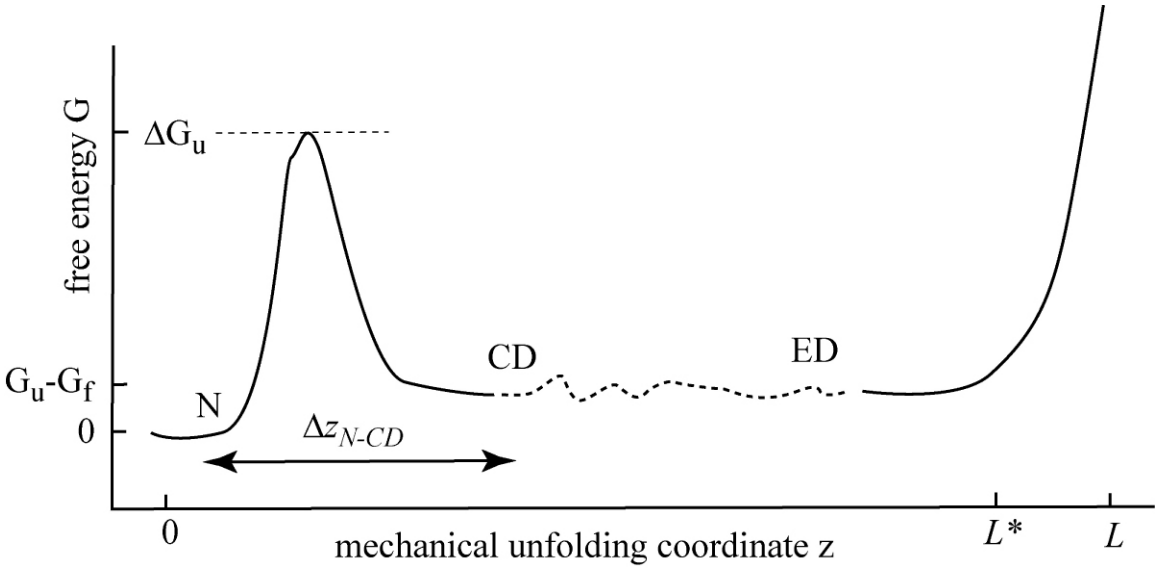

Figure 5.1. A cartoon of the often assumed dependence of the protein free energy on its extension[111, 135] (i.e. the mechanical unfolding reaction coordinate). The native state (N) is separated from the condensed denatured state (CD) by a barrier. The extended denatured state (ED) corresponds to larger extensions.

(ii) Second assumption is that the free energy profile of a protein domain $G(z)$ along the one-dimensional mechanical pulling coordinate (equal to the protein extension z) has a shape that consists of a "native" minimum separated from the unfolded basin of attraction

by a barrier, see Fig. 5.1. In particular, it has been proposed [111, 135] that $G(z)$ involves a native minimum N separated from a condensed denatured state CD by a barrier ΔG_u; when further stretched, the protein attains an extended denatured state ED. Finally, as the polypeptide chain approaches its fully extended conformation the free energy rises abruptly; This part of the $G(z)$ curve is usually described in terms of polymer extension models such as the wormlike chain model [149-151]. The location, the width, and the height of the barrier have been estimated [111] – based on the above two assumptions - from the measured force dependence of the mechanical unfolding and refolding rates.

First assumption has been questioned by recent theoretical studies and simulations [59, 60, 152, 153], which indicate that, while the linear approximation may be adequate in the relatively narrow range of force f probed in pulling experiments, it cannot be used to extrapolate $k_u(f)$ to zero force. However, another question arises in connection with assumption (i), to what extent can the actual multidimensional folding/refolding dynamics of a protein be understood in terms of its projection onto a one- or low-dimensional space? It is well known that a poor choice of a reaction coordinate in transition state theory can lead to very poor estimates of reaction rates [64-66]. On the other hand, models assuming that the effect of all degrees of freedom other than the reaction coordinate amounts to a Stokes-type friction and a stochastic force and postulating a simple Langevin equation to describe one-dimensional dynamics along the reaction coordinate seem to agree quite well with steered Molecular Dynamics simulations of protein unfolding [34, 59, 62, 63].

Second assumption is hard to reconcile with the recent experimental data on protein *refolding* under mechanical tension [47, 48]. Specifically, Fernandez and Li report on an experimental observation of a folding trajectory of a single protein molecule by using force-clamp spectroscopy [47]. In their study, a chain containing several tandem repeats of the ubiquitin domain was first extended by applying a high force f_h between the chain ends. Then the force was reduced to a lower value f_l. By monitoring the chain extension $z(t)$ as a function of time they find that the protein refolds and attains a value of z corresponding to the chain in which all the domains are folded.

The protein is seen to refold even if a stretching force as high as f_l = 40 pN is applied to the protein. In other words, the compact native state of each domain remains thermodynamically stable relative to the extended state at this value of the stretching force. However assuming a free energy profile depicted in Fig.5.1, even a much lower force should destabilize the native state and favor the extended state thermodynamically. To see this, note that when a stretching force $\boldsymbol{f}$ is applied between the ends of the chain, where the direction of the force is taken to be along the z-axis, the domain free energy becomes

$$G^{(f)}(z) = G(z) - fz \tag{5.1}$$

Assuming that the dependence $G(z)$ stays relatively flat in the region between states CD and ED, as suggested in Fig. 5.1, application of the force f will turn this flat dependence into a downhill ramp until the extension reaches a value L^* comparable with the contour length L. The extended state will have free energy lower than that of the native state if

$$G_u - G_f - fL^* < 0 \tag{5.2}$$

where the $G_u - G_f$ is the free energy difference between the denatured and the native states in the absence of the force. Using the values $G_u - G_f \sim 5$ kcal/mol and $L^* \sim 280$Å from ref. [111], we find that this corresponds to a force $f \geq f_m = 1.24\text{pN}$! Although we have used the values of energy and extension reported in[111] for the I27 domain while the mechanical refolding experiments were performed with ubiquitin, these two domains are known to have very similar mechanical properties [60, 61, 102] and it is very unlikely that the differences between these two domains would account for the over an order of magnitude difference in the force f_m.

Perhaps the estimated value for the maximum extension L^* is too high because we have assumed a flat free energy profile thus ignoring the entropic elasticity of the unfolded chain. To account for this effect let us assume that the unfolded protein can be described as a Gaussian coil whose mean square end-to-end distance is given by

$\left\langle r^2 \right\rangle = Mb^2$ where M+1 is the number of residues in the domain. The minimum free energy of the extended state can then be written as

$$\min_z \left(G_u + \frac{3z^2}{2Mb^2} k_B T - fz \right) = G_u - \frac{f^2 Mb^2}{6k_B T}, \tag{5.3}$$

where the 2nd term in the lhs of this expression is the free energy contribution arising from the entropy of the Gaussian chain[112, 154, 155]. The extended state becomes more stable than the native state if this energy is less than G_f, leading to the condition

$$f \geq f_m = \left[\frac{6k_B T (G_u - G_f)}{Mb^2} \right]^{1/2} \tag{5.4}$$

Using M=75 (for ubiquitin) and a typical value b=8 Å(see, e.g., [156, 157]), Eq. 5.4 gives $f_m = 4\,pN$, still an order of magnitude lower than the force observed experimentally.

These estimates suggest that the common notion of a bistable free energy profile $G(z)$ represented by the cartoon in Fig. 5.1 may be incorrect. Mechanical unfolding experiments do not measure $G(z)$ directly but instead indirectly probe the tilted free energy surface $G^{(f)}(z)$ (Eq. 5.1) in the presence of a pulling force f. $G^{(f)}(z)$ is indeed a bistable (or possibly multistable) function. Recent single molecule RNA stretching experiments[158] have demonstrated that by tilting $G^{(f)}(z)$ with an appropriate force one can observe, in real time, the switching of a single molecule between its unfolded and folded conformations corresponding to two minima of $G^{(f)}(z)$. The mechanically unfolded state in this case is different from the chemically denatured state although the free energies of the two can be compared if one corrects for the reduced entropy of the mechanically unfolded state[158].

While $G^{(f)}(z)$ is bi- or multi-stable, $G(z)$ does not have to be. Recent theoretical studies[59, 153] showed that different models for $G(z)$ can comparably well reproduce the available experimental data.

We use simulations to investigate the curve $G(z)$ without making any preconceived assumptions about its shape. Unfortunately, computing $G(z)$ from atomistic

simulations is not possible because characterizing the unfolded state of a protein (corresponding to large extensions), would require prohibitively long simulation timescales. Here we resort to Langevin dynamics simulations of a minimalist model of a ubiquitin-like domain to compute the free energy profile $G(z)$ and to compare it with the free energy profiles along other reaction coordinates that are commonly used to monitor the protein folding progress. We find that while the latter can exhibit a typical double-well shape with the two wells corresponding to the native and denatured ensembles, the free energy profile $G(z)$ along the mechanical unfolding coordinate is a monotonically increasing function in the absence of force. When a finite force is applied, the tilted free energy $G_f(z) = G(z) - fz$ develops well defined minima corresponding to the native ensemble and extended state(s). For this reason there is no contradiction between our present findings and previous experiments that probe $G_f(z)$. However we will demonstrate in this chapter that when the assumption that $G(z)$ is bistable as in Fig. 5.1 is abandoned, the resulting estimate for the force f_m necessary to destabilize the native protein turns out to be considerably higher and is consistent with the exeriments[47, 48].

Also, we address the issue of adequacy of reduced low-dimensional models to some extent by studying trajectories of folding and refolding in the two-dimensional space that consists of the mechanical and "chemical" coordinates. We find that these trajectories tend to stay close to the steepest descent "gutters" on the two-dimensional free energy surfaces, suggesting that, reduced representations of mechanical folding/refolding dynamics are meaningful.

Most single molecule protein pulling studies employing AFM involve chains of identical domains rather than a single domain. In Section 5.3.4 of this chapter we will consider a simple model of a polyprotein chain that consists of N domains and study its free energy $G_N(z)$ as a function of the chain extension z. We will demonstrate that, in a certain range of temperatures and for a sufficiently large N, this free energy is simply a linear function of z:

$$G_N(z) = f_m z \tag{5.5}$$

We propose that the "slow refolding phase" observed in the mechanical refolding of polyubiquitin chains[47] can be viewed as one-dimensional overdamped motion in the linear potential of Eq. (5.5).

5.2 MODEL AND METHODS

5.2.1 The protein model

The minimalist off-lattice model[104-106] of ubiquitin-type protein, that has already utilized in Chapter 3, is used. In this model, the peptide potential, as a function of the position $\mathbf{r}_i$, i=1, …, N, of each residue, is given by:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V_{\text{bond}} + V_{\text{bend}} + V_{\text{dih}} + V_{\text{non-bonded}} \quad (5.6)$$

where, V_{bond}, V_{bend}, V_{dih} and, $V_{\text{non-bonded}}$ are bond stretching, bond bending, torsion bending and non-bonded potential terms detailed in 3.2.2 of this dissertation.

To study the kinetics of unfolding and refolding, we apply a mechanical force f to the first and last bead of the chain (N=75) by adding the term $V_{pull} = -f(z_1 - z_{75})$ to the equation (5.6).

Within this model, the peptide was found to reproducibly assume a ubiquitin-like fold (shown in Fig. 5.2) when the temperature is below the folding temperature $T_f = 0.56\,\varepsilon_h / k_B$.

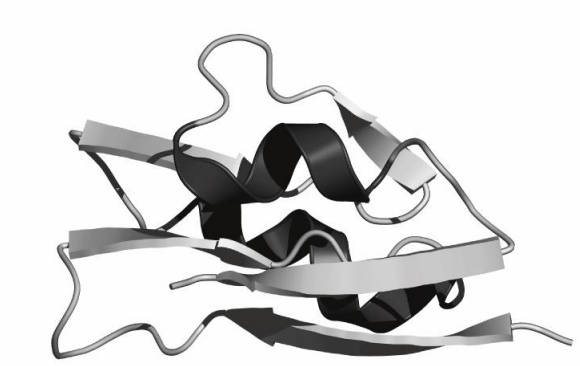

Figure 5.2. The native configuration of the ubiquitin-like domain.

5.2.2. Simulating the protein dynamics

It was assumed that the dynamics of each atom in the chain are governed by the Langevin equation of the form

$$m\ddot{r}_i(t) = -\xi\dot{r}(t) - \frac{\partial V}{\partial r_i} + R(t) \tag{5.7}$$

where $\mathbf{r}_i$ is the position of the *i*-th atom, *m* is its effective mass, $\xi = 0.05(\sigma^2 / m\varepsilon_h)^{-1/2}$ is the friction coefficient, and $\mathbf{R}(t)$ is a random δ–correlated force satisfying the fluctuation-dissipation theorem. This equation was solved by using the velocity Verlet algorithm as described by Paterlini et al.[11] We use dimensionless units of energy, distance, time, and force respectively equal to ε_h, σ, $(m\sigma^2/\varepsilon_h)^{1/2} \cong 2.5\,\text{ps}$, and $\varepsilon_h/\sigma \cong 18\,\text{pN}$.

5.2.3. Obtaining free energy profiles

To improve sampling in the temperatures lower than folding temperature we had to resort to the replica exchange method[88, 89], in which Langevin trajectories were run in parallel for a set of different temperatures and configurations were exchanged according to an algorithm that satisfies detailed balance.

The free energy profile of any structural property χ can be obtained as

$$G(\chi) = -k_B T \ln p(\chi) \tag{5.8}$$

where $p(\chi)$ is the sampled probability distribution of χ. The quantity χ may either be a scalar or a vector, each component of which corresponds to a different reaction coordinate or order parameter. To obtain the global shape of $G(\chi)$ far away from minimum of $G(\chi)$, the standard weighted histogram/umbrella sampling method was used as described in[67-69].

5.3 RESULTS

5.3.1 Free energy profile along the mechanical reaction coordinate of single domain protein

Fig. 5.3 depicts the computed free energy profile $G(z)$ for different temperatures. We find that, in contrast to the hypothetic $G(z)$ depicted in Fig. 5.1, the computed profile has no minima except the one at z=0. For $z \le 6\sigma$ the free energy is nearly constant. This plateau is anticipated for a protein in its native state (assuming that the protein shape fluctuations are small so that its end-to end vector has a constant absolute value and arbitrary direction) . Indeed, the distance between the ends of the chain in the native state is $r_0 \approx 6.3\sigma$.

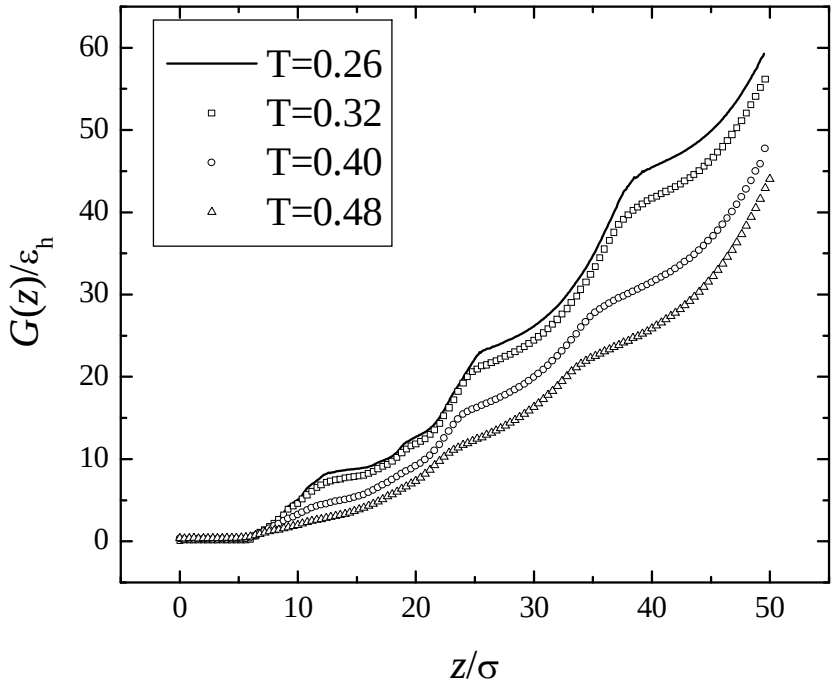

Figure 5.3. The computed free energy as a function of the domain extension at different temperatures.

As z is increased, $G(z)$ rises in a series of steps. Each of these steps turns into a barrier when a stretching force f is applied. This is seen in Fig. 5.4, which shows the free energy $G_f(z) = G(z) - fz$ for different forces at T=$0.32\,\varepsilon_h / k_B$. Correspondingly, as it is being stretched, the protein assumes a series of partially folded conformations. Three representative conformations corresponding to the extensions z = 6.6σ, 17.8σ, 29.8σ (points I, II, III in Fig. 5.4) are shown in Fig. 5.5. In addition, average contact maps corresponding to these values of z are also shown. [The contact map here is a plot containing points (i,j) for every pair of residues i and j that fall within a distance of $|r_i - r_j| < 1.97\sigma$. The darkness of each point in this map increases with the probability of this contact to occur in the ensemble of stretched protein configurations, black points corresponding to contacts that occur with a probability close to 100%]. The conformations corresponding to the point IV in Fig. 5.4 (z=41.8σ) are predominantly extended, contain no residual structure and no contacts.

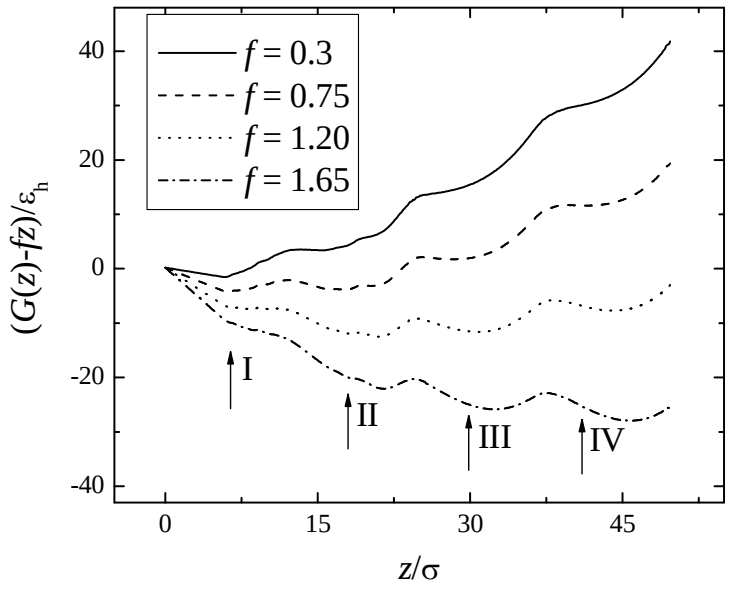

Figure 5.4. The free energy $G(z)$-fz of the domain in the presence of a force f at T=$0.32\,\varepsilon_h / k_B$ and at different values of f.

The first barrier in Fig. 5.4 is encountered at $z = 12\sigma$. By comparing the structures with $z < 12\sigma$ and $z > 12\sigma$ (exemplified by structures I and II in Figure 5.5), one finds that the former are native-like while the latter preserve much of the native structure except that the two terminal parallel strands of the native protein become separated.

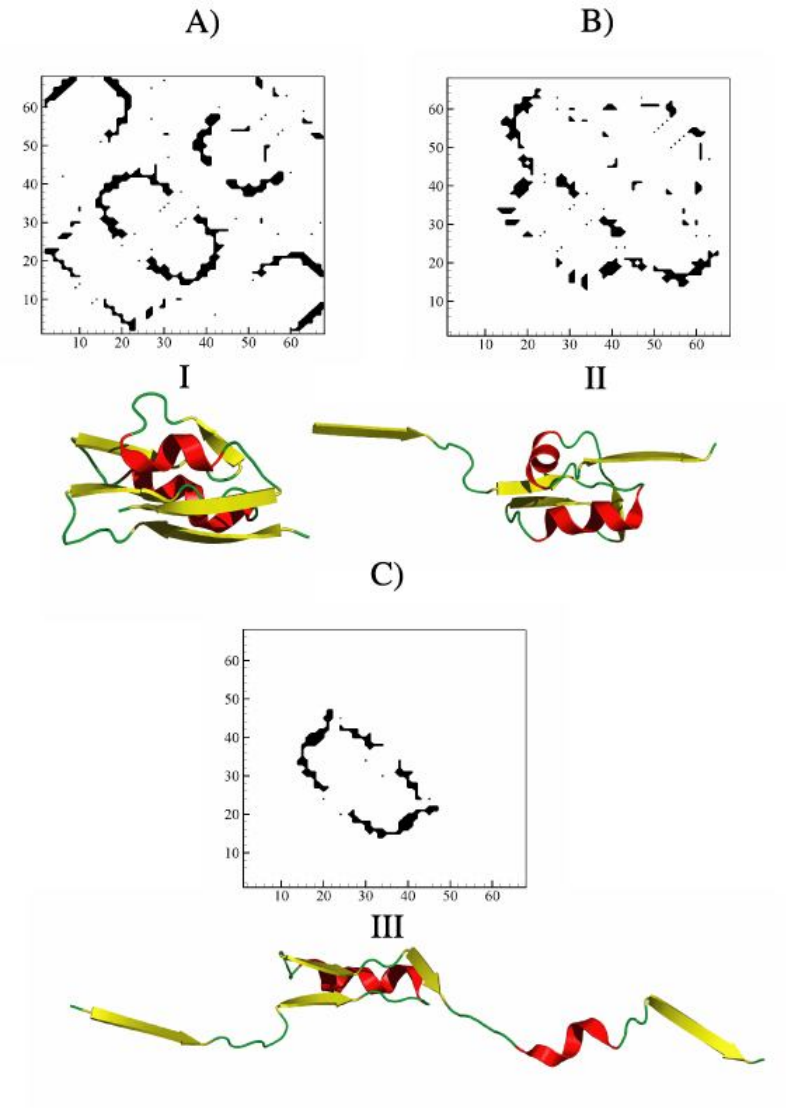

Figure 5.5 Representative configurations and average contact maps for the equilibrium ensemble of conformations of the ubiquitin domain at T=0.32 ε_h / k_B constrained to have its extension equal to (a) 6.6σ, (b) 17.8σ, and (c) 29.8σ corresponding to the points I, II, III in Fig. 5.4 The z value was constrained by a spring to be equal to those values.

Atomistic simulations of ubiquitin[48, 60, 61] showed similar behavior: the separation of the terminal parallel strands results in a free energy barrier that accounts for the mechanical stability of ubiquitin-like domains. Similarly, simulations of other protein domains containing terminal parallel strands indicated that their mechanical stability is largely accounted for by these strands forming a "clamp" that resists mechanical force[33-35, 60, 63]. Further, it has been argued[159, 160] that such terminal parallel strands represent an optimal topology maximizing the mechanical resistance of a protein .

The height of this first free energy barrier corresponding to the separation of the parallel strands depends on temperature. As T is increased and approaches the folding temperature from below, the clamp formed by the terminal strands becomes less stable resulting in a lower free energy barrier. For the lowest temperature for which our calculations were performed, the barrier is $\approx 8\varepsilon_h$. If we assume $\varepsilon_h = 1-2$ kcal/mol [84], the resulting barrier is 8 -16 kcal/mol, which is lower than the free energy barrier $\Delta G_u = 29$ kcal/mol estimated for ubiquitin from Molecular Dynamics simulations[60]. This is not surprising: The coarse-grained model used here does not include hydrogen bonds, which stabilize the parallel strands[34]. The fact that this model still captures the unfolding scenario observed in atomistic simulations supports the notion that native topology largely determines the mechanical unfolding mechanism[159-161].

From Fig. 5.4, the barrier separating the conformations I and II disappears at a force $f \approx 1.2\varepsilon_h / \sigma$ (at T=0.32 ε_h). For ε_h in the range 1-2 kcal/mol, this corresponds to a force $f = 22-44\,pN$, which is substantially lower than that found in simulations[59, 60] or inferred from experiments[102]. Again, this discrepancy is not surprising given the fact that interactions in the model used here are not necessarily quantitative.

Despite the fact that the present model underestimates the unfolding barriers and forces, it predicts that the force required to destabilize the native state and render an extended state thermodynamically favorable is larger than that estimated by assuming the hypothetic free energy profile shown in Fig. 5.1. Indeed, one sees in Fig. 5.4 that at a force $f = 0.75\varepsilon_h / \sigma \approx 14-28\,pN$ (again assuming $\varepsilon_h = 1-2$ kcal/mol) the native state

still remains stable (i.e. $G(z)$ is uphill as z is increased). For $f \geq 1.65\varepsilon_h / \sigma \approx 30-60\,pN$ the extended state (conformation IV in Fig. 5.4) becomes thermodynamically favorable, while at intermediate forces $0.75\varepsilon_h / \sigma < f < 1.65\varepsilon_h / \sigma$ partially unfolded states II and III shown in Fig. 5.5 may be stabilized by the force.

Despite the fact that the model significantly underestimates the unfolding free energy barriers, it predicts that the protein should refold to a native-like conformation under a force ~14-28 pN that is comparable with the forces used in the mechanical refolding experiments [47]. We anticipate even a larger force required to destabilize the native protein if more realistic interactions are used. This is in contrast with the free energy of Fig. 5.1, for which a stretching force of a few piconewtons would inhibit protein refolding.

5.3.2. The free energy profile along a "chemical" reaction coordinate

Our finding that $G(z)$ does not have two minima corresponding to the unfolded and folded ensembles but rather is a monotonic function appears to contradict common intuition about free energy landscapes of proteins. We have experimented with a number of different protein models and in each case we found $G(z)$ to be a monotonic function regardless of temperature and details of the model. This type of behavior is also found in atomistic simulations of mechanical unfolding [59-61] although those cannot be trusted to adequately sample the denatured ensemble. We believe that this common property of $G(z)$ has to do with the inability of the extension, as a reaction coordinate, to distinguish between the folded and unfolded state.

In studies of protein folding, a number of other reaction coordinates or order parameters have been used, all of which can distinguish between the folded and unfolded states. Examples include the number of native contacts, the root mean square deviation from the native conformation (*rmsd*), and the radius of gyration (R_g). The free energy profile, plotted as a function of one of these order parameters, is commonly expected to exhibit two minima. Indeed, we have found this to be the case for our model.

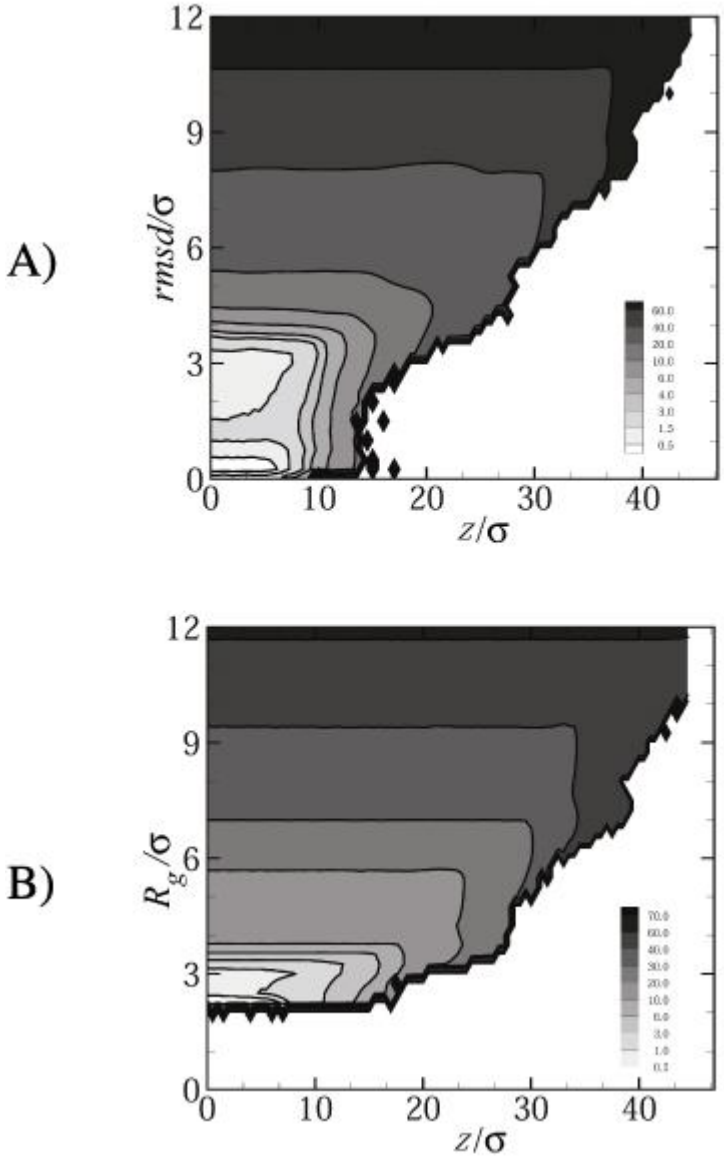

Figure 5.6. Two-dimensional contour maps for the free energy as a function of A) the radius of gyration and the protein extension and B) the *rmsd* from the native conformation and the protein extension at T=0.48 ε_h / k_B.

To gain further insight into the free energy landscape of our model protein, we have constructed two-dimensional free energy surfaces $G(z, \chi_{chem})$, where the "chemical" reaction coordinate χ_{chem} is R_g, or *rmsd* . Our results for T=0.48ε_h / k_B are

shown in Fig. 5.6. Fig. 5.7 shows slices of this two-dimensional free energy surface when it is dissected along the direction of the z axis. For small z $G(z, rmsd)$ shows two minima, as a function of *rmsd*. The minimum corresponding to the native state is lower than that corresponding to the denatured ensemble. As z is increased, the minimum corresponding to the unfolded state becomes lower (relative to the native state) and eventually the native state minimum disappears. Interestingly, this happens at a fairly small value of z that roughly corresponds to the first transition state associated with the separation of the terminal parallel strands.

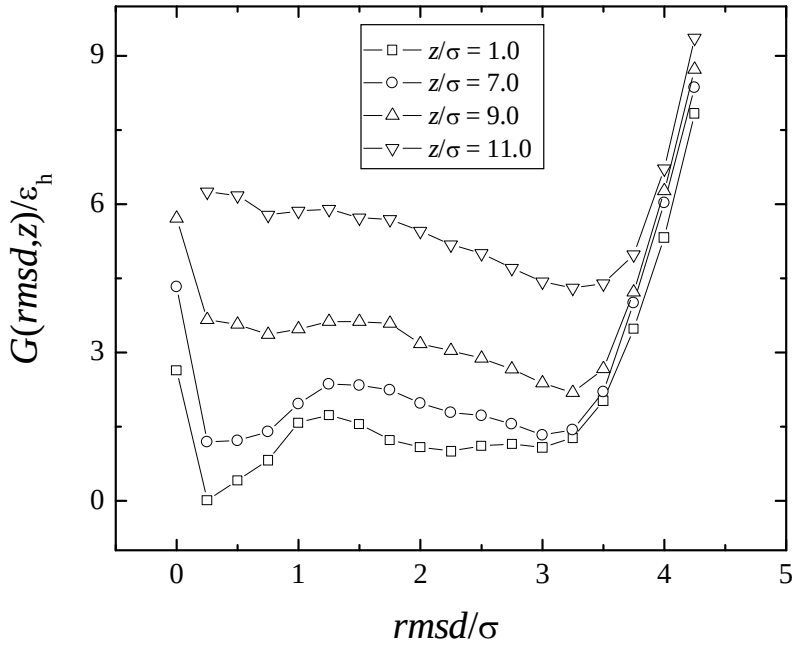

Figure 5.7. Slices of the two-dimensional free surface $G(rmsd, z)$ plotted in Fig. 5.6B taken at different values of the protein extension z.

The mechanical and the chemical reaction coordinates are therefore coupled. The stretching of the protein biases the conformational ensemble towards the unfolded state – in this sense one can talk about *mechanical unfolding*. However for sufficiently small

extensions both the denatured and the native ensembles can coexist (for temperatures that are close to the folding temperature) and their coexistence is not reflected in the shape of the one-dimensional free energy curve $G(z)$ in an obvious way.

5.3.3. The kinetics of unfolding and refolding

Here we consider unfolding and refolding trajectories in the two-dimensional space (z, *rmsd*) that consists of the mechanical and a chemical coordinates. In our refolding simulations, the protein was pre-stretched with a high force, after which the force was relaxed to a smaller value f_l. Then the refolding trajectory ($z(t)$, *rmsd*(t)) was monitored as a function of time. A typical refolding trajectory for the case $f_l = 0$ is shown in Fig. 5.8A. This trajectory wanders quite a bit from the steepest descent path on the free energy surface $G(z, rmsd)$ and gets trapped near the minimum corresponding to an unfolded compact state. Eventually, it crosses over the barrier that exists along the chemical reaction coordinate (cf. Fig. 5.7) and arrives at the native state. In Fig. 5.8B, we plot *rmsd*, R_g , and z as functions of time for this trajectory. Interestingly, both R_g and z attain their native values before *rmsd* does suggesting that, before folding, the protein first reaches a collapsed non-native state that cannot be distinguished from the native state by monitoring either R_g and z. The time at which the protein finally folds thereby achieving its native *rmsd* value is indicated by an arrow in Fig. 5.8B.

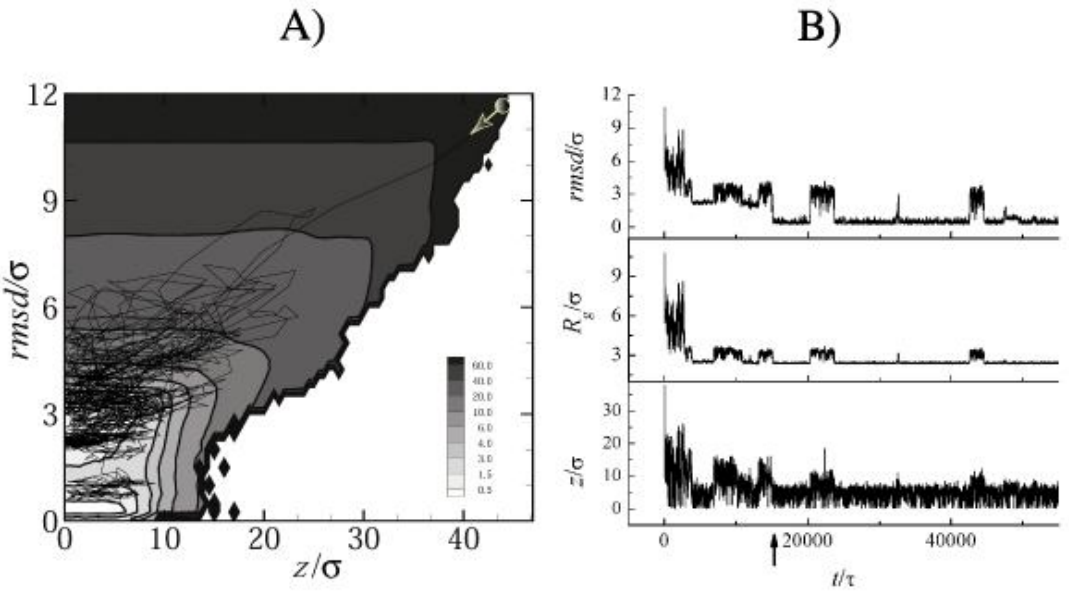

Figure 5.8. A typical refolding trajectory observed when the domain is initially stretched by a high force and the force is then removed at t=0. A): *rmsd* vs z for this trajectory plotted against the free energy contour map $G(rmsd, z)$. The trajectory originates at a large value of z corresponding to an extended conformation. B) The time dependence of *rmsd*(t), $R_g(t)$ and $z(t)$ for this trajectory. The arrow indicates the time at which *rmsd* attains its native value. Only the part of the trajectory proceeding this moment is shown in Fig. 5.8A.

Figure 5.9 shows typical trajectories in the presence of a stretching force. When f_l is sufficiently high, this force prevents the domain from completely folding; Instead, it ends up in a collapsed unfolded state corresponding to $rmsd \approx 3\sigma$. This situation is illustrated in Fig. 5.9A. The force f_l also causes the free energy landscape $G(z, rmsd) - f_l z$ to become more rough. As a result, the folding trajectory becomes temporarily trapped in the regions of the free energy surface where its gradient is small. Note also that the stretching force generally tends to align the trajectories along the steepest descent path on the $G(z, rmsd) - f_l z$ surface.

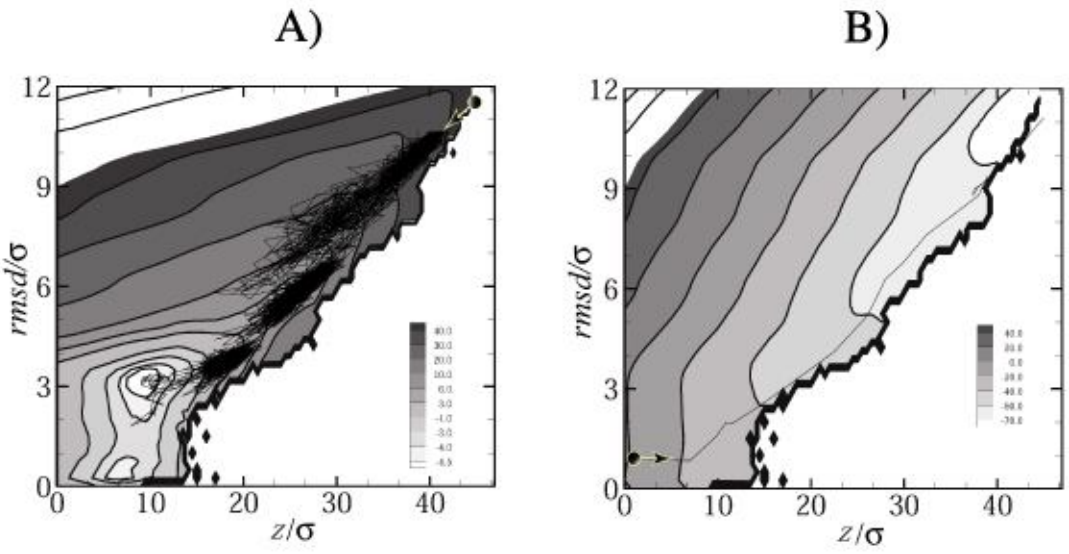

Fig. 5.9: A) The trajectory observed when the domain is initially stretched by a high force and the force is then reduced to $f_l = 0.8\varepsilon_h / \sigma$ at t=0. This trajectory is plotted against the free energy contour plot $G(rmsd, z) - f_l z$ (b): A typical unfolding trajectory in the presence of a high value of the unfolding force f_h=3.5ε_h / σ .

Fig. 5.9B illustrates the unfolding process occurring when a much higher force f_h is applied. The unfolding trajectories in this case are generally very close to the steepest descent path on $G(z, rmsd) - f_h z$, which is consistent with the expectation that application of a sufficiently large force should impose the mechanical reaction coordinate on the system.

5.3.4. The free energy of a polyprotein chain

In recent polyprotein stretching-refolding experiments [47, 48], Fernandez and coworkers observed a polyprotein fold under mechanical tension. Here we will assume that the observed refolding trajectory in the presence of a mechanical force f can be described as a one-dimensional stochastic motion in the potential of mean force

$$G_N^{(f)}(z) = G_N(z) - fz, \qquad (5.9)$$

where z is now the extension of the entire chain and $G_N(z)$ is the potential of mean force for the chain that consists of N domains. If the domains are statistically independent, (i.e. there are no inter-domain excluded volume effects) then this potential of mean force can be calculated from the equation:

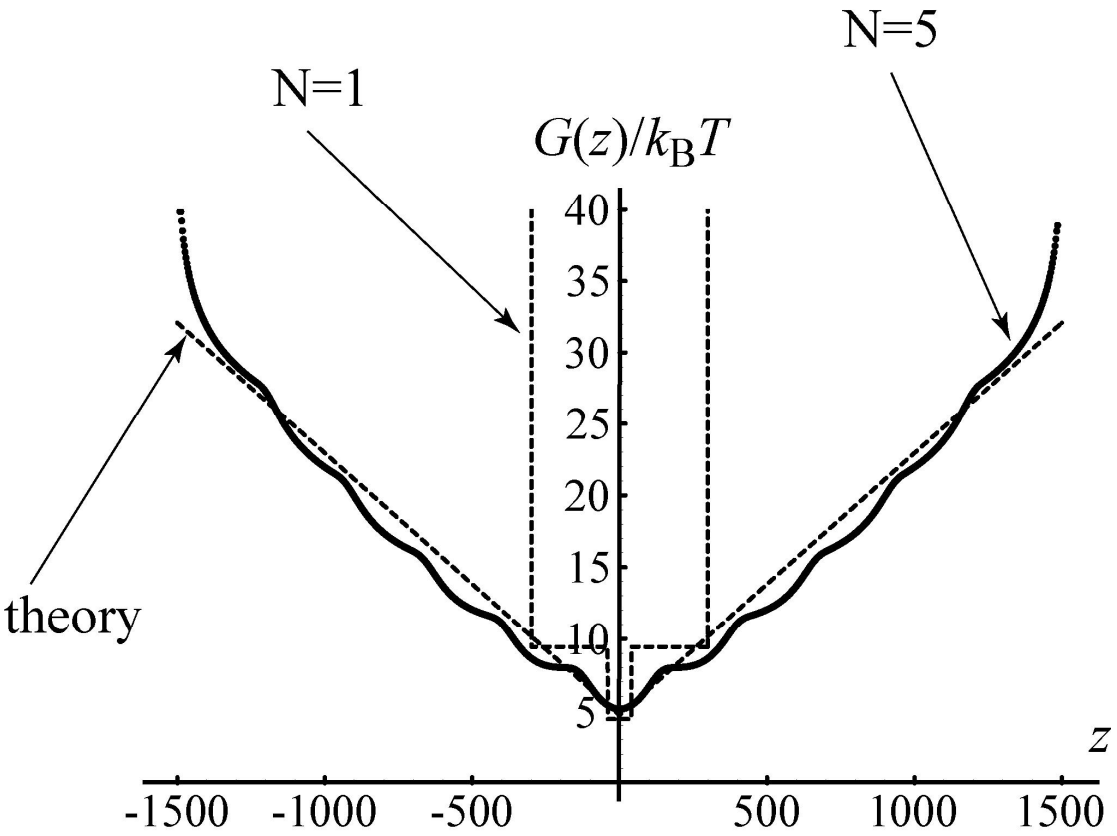

Figure 5.10. The free energy of the polyprotein chain $G_N(z)$ that consists of N domains as a function of its extension z. The free energy of a single domain is described by Eqs. 5.11-5.12. The theoretical curve is given by Eqs. 5.13, 5.23 with $b=L$ and $M=1$.

$$p_N(z) = \exp[-G_N(z)/k_BT] = \int \ldots \int dz_1 dz_2 \ldots dz_N \, p_1(z_1) \ldots p_1(z_N) \delta(z - z_1 - z_2 - \ldots z_N) \quad (5.10)$$

where

$$p_1(z) = \exp(-G(z)/k_BT) \quad (5.11)$$

is the probability distribution of the end-to-end distance for a single domain. Fig. 5.10 shows $G_N(z)$ for the case where the extension of a single domain z has the following normalized probability distribution:

$$p_1(z) = \left[2l + 2(L-l)\exp(-\varepsilon / k_B T)\right]^{-1} \begin{cases} 1, & |z| < l \\ \exp(-\varepsilon / k_B T), & l < |z| < L, \\ 0, & |z| > L \end{cases} \qquad (5.12)$$

and, consequently, the free energy profile $G_1(z) = -k_B T \ln p_1(z)$ for a single domain is given by a singe-step staircase potential also shown in Fig. 5.10. Here the maximum extension of the domain is L and l roughly corresponds to the equilibrium extension of the folded protein. By computing $G_N(z)$ numerically using Eq. 5.10 we find:

(i) $G_N(z)$ approaches a linear potential

$$G_N(z) = f_m z + \text{constant} \qquad (5.13)$$

when N is large enough (in practice, $N \geq 3$), except in the limits $z \to 0$ and $z \to NL$.

(ii) The slope f_m is independent of N.

These observations are somewhat surprising. Indeed, the extension $z = z_1 + z_2 + \ldots + z_N$ of the chain comprised of N domains can be viewed as a result of a one-dimensional random walk with steps z_1, z_2, …, z_N (cf. Eq. 5.10). The function $p_1(z)$ describes the probability distribution of a step length. According to Eq. 5.12, there are two types of steps: Short steps with a length $|z|<l$ that occur frequently and long steps that occur with a probability proportional to $\exp(-\varepsilon / k_B T)$. According to Eq. 5.13, the resulting probability distribution $p(z)$ is not Gaussian but exponential, contrary to what one would expect based on the central limit theorem. The Gaussian distribution is recovered in both low- and high temperature limits. In the former case, the probability of a long step is negligible (or, equivalently, the probability for a domain to unfold can be neglected) and the length of a step is a uniform number between $-l$ and l. In the latter, the step size is uniformly distributed between $-L$ and $+L$. In both cases $p(z)$ is Gaussian for N large enough, as is well known from theory of random walks.

The origin of the linear potential of Eq. 5.13 can be understood as follows: Suppose $n > 0$ domains (out of the total of N domains) are unfolded in the polyprotein chain. If we neglect the extension of the folded domain as compared to that of unfolded domains then the probability distribution of the chain extension z is that of the denatured, random-flight chain that consists of nM links, where M is the number of links in the single domain chain. Let us call this quantity $p^{(nM)}(z)$. Since the probability to have n particular domains unfolded is proportional to $\exp(-\varepsilon / k_B T)$ and there are $\binom{N}{n}$ ways to have n domains unfolded, then

$$p_N(z) \propto \sum_n \binom{N}{n} p^{(nM)}(z) e^{-n\varepsilon / k_B T} \tag{5.14}$$

To be more specific, let us model the unfolded chain as a random coil; For sufficiently large z one then can write[112, 154]:

$$p^{(Mn)}(z) \propto \exp\left[-g(z / R_F(Mn))\right] \tag{5.15}$$

where

$$R_F = a(Mn)^{\nu} \tag{5.16}$$

is Flory's chain dimension and

$$g(x) = Dx^{\delta},\ \delta = (1-\nu)^{-1} \tag{5.17}$$

The case $\nu = 3/5$ corresponds to random-flight chains with excluded volume constraints; Thermally and chemically denatured proteins are often well described by this approximation [157]. The case $\nu = 1/2$ corresponds to ideal polymer chains that obey Gaussian statistics.

The dominating contribution in the sum of Eq. 5.14 comes from the term $n = n^*$ corresponding to the minimum of the quantity

$$S(n) = D\left(\frac{z}{a(nM)^{\nu}}\right) + n\varepsilon / k_B T \tag{5.18}$$

(we have neglected the weaker n dependence of the combinatorial coefficients). This corresponds to the condition $dS/dn = 0$, which gives

$$n* = z\left(D\frac{\nu\delta k_B T}{a^{\delta} M^{\nu\delta}\varepsilon}\right)^{1/\delta} \tag{5.19}$$

so that

$$p_N(z) \propto \exp\left[-S(n*)\right] = \exp\left[-\frac{z}{aM^{\nu}} D^{1-\nu}\left(\frac{\varepsilon}{k_B T}\right)^{\nu} \frac{(1-\nu)^{\nu-1}}{\nu^{\nu}}\right] \tag{5.20}$$

And

$$G_N(z) = -k_B T \ln p_N(z) = f_m z + \text{constant}, \tag{5.21}$$

where

$$f_m = \frac{k_B T}{aM^{\nu}} D^{1-\nu}\left(\frac{\varepsilon}{k_B T}\right)^{\nu} \frac{(1-\nu)^{\nu-1}}{\nu^{\nu}} \tag{5.22}$$

Remarkably, the potential of mean force is linear regardless of the chain statistics; The Flory chain and the ideal chain both give qualitatively the same dependence. In particular, for ideal chains the last equation becomes

$$f_m = (2/a) D^{1/2}\left(\varepsilon k_B T / M\right)^{1/2} = \frac{1}{b}\sqrt{\frac{6\varepsilon k_B T}{M}} \tag{5.23}$$

This precisely the result of Eq. 5.4 if we replace $G_u - G_f$ by ε. Recalling how we have derived Eq. 5.4, the meaning of the force f_m as the force that destabilizes a single domain becomes apparent. Further extension of the chain leads to the unfolding of more domains such that the number of unfolded domains is proportional to the extension according to Eq. 5.19 and the force in the chain remains constant. This is reminiscent of the plateau force found for the mechanical unfolding of collapsed polymers (see, e.g., [162-164]).

Although the above reasoning seems to only apply to chains with very large *N*, where *n* effectively becomes a continuous variable, Eqs. 5.21-5.23 are remarkably accurate even for a small number of domains *N* ~5, as seen from Fig. 5.10, which compares $G_N(z)$ calculated numerically for the double-square-well potential of eq. 5.12 with Eqs. 5.21-5.22. To make this comparison, note that *b*=*L* and M=1 for an unfolded chain that consists of segments of length *L*.

Eqs. 5.19-5.22 are valid provided that $1 < n* < N$, which means that the temperature and the extension should be such that

$$1 < z\left(D\frac{\nu\delta k_B T}{a^{\delta} M^{\nu\delta}\varepsilon}\right)^{1/\delta} < N \qquad (5.24)$$

As noted above, when temperature is too high or too low, Gaussian chain statistics is recovered (for $z < NL$). We have also verified that $G_N(z)$ remains linear when a more general staircase-type potential $G(z)$ is assumed.

5.4. DISCUSSION

The above observations may shed light on the recent force-clamp stretching-refolding experiments by the Fernandez group [47]. In their study, a chain containing several tandem repeats of the ubiquitin domain was first extended by applying a high force f_h between the chain ends. Then the force was reduced to a smaller value f_l, which was low enough to allow the protein to refold. To study this refolding process, the chain extension $z(t)$ was monitored as a function of time t. For a sufficiently weak force f_l, the protein was seen to contract and to eventually attain an extension z that corresponded to the peptide length in its folded state. The recorded folding trajectories $z(t)$ depended on the magnitude of the force f_l and showed several remarkable features:

(i) When f_l is zero then folding is very fast so that $z(t)$ is essentially a step function given the temporal resolution of the experiment. The time scale of the chain contraction however was found to strongly depend on the applied residual force f_l and this force can be chosen such that the folding trajectory is slowed down enough that its details can be resolved experimentally.

(ii) The refolding trajectory $z(t)$ showed several distinct refolding stages. Remarkably, it included a slow stage, in which $z(t)$ is almost a linear function of time. Because during the stage $z(t)$ changes continuously rather than in a stepwise fashion expected in the case where each domain folds independently, the authors of ref. [47]proposed that this slow stage may involve cooperative collapse of ubiquitin domains.

(iii) The slow stage was typically followed by a much faster, abrupt event, in which the chain retracted completing the folding process

The present model can account for these observations if one assumes that folding can be viewed as a one-dimensional overdamped motion in the potential $G_N(z) \approx f_m z$:

(i) In the presence of the force $f_l < f_m$ one can view the chain contraction as downhill overdamped motion in the linear potential $G_f(z) = (f_m - f_l)z$. During this motion, the speed of contraction, $v = dz/dt$, should stay roughly constant, in accord with the experimental findings for the slow phase for refolding [47]. Further, the contraction rate is proportional to slope of the potential and therefore the refolding time should therefore depend on the force according to

$$t_f \propto \frac{1}{f_m - f_l} \qquad (5.25)$$

Eq. 5.25 will break down as f_l approaches f_m . In this limit the potential becomes flat and the folding time should approach the value corresponding to free diffusion. [However the effective diffusion coefficient may be modified because of the roughness of the free energy landscape]. It should be possible to verify these predictions experimentally.

(ii) The absence of distinct steps in $z(t)$ during the slow phase is a consequence of the smoothness of $G_N(z)$. The folding of a domain does not manifest itself as an abrupt change of $z(t)$ because the domain is connected sequentially to a soft-spring-like chain containing "slack" that is provided by the already unfolded domains.

(iii) When only one unfolded domain is left in the chain there is no longer any slack created by other domains. Thus once this domain unfolds, the chain should contract abruptly. This may explain the abrupt contraction of the chain observed in [47] at the end of each folding trajectory.

The above predictions are consistent with the experimental findings [47]. Are there any predictions that are inconsistent with the experiment? Our model essentially assumes independent folding of each domain such that the number of folded domains

depends roughly linearly on the chain extension (cf. Eq. 5.19). Fernandez and Li however point out [47] that when they stop the refolding process midway at a time $t_{mid} < t_f$ by applying a high unfolding force f_h again, they often (but not always) observe no steps in the ensuing stretching trajectory, which suggests that none of the domains has folded yet. [The state of each domain at this stage may correspond to the collapsed non-native state observed in Fig. 5.9A, which has the values of z and R_g that are very close to the native ones]. In our model, the probability of finding one or more folded domains depends on the waiting time t_{mid}. To unambiguously establish whether or not there actually is a disagreement between the experiments and our model, more detailed experimental data on the likelihood of finding folded domains as a function of t_{mid} may be necessary.

We finally note that $G(z)$ for a collapsed homopolymer may also exhibit a linear dependence on z [165, 166]. Therefore one expects that a polymer undergoing non-specific hydrophobic collapse may also exhibit the behavior similar to that observed in ref. [47]. Further experimental studies are required before the mechanism of the slow refolding kinetics can be unambiguously established.

Chapter 6

Hysteresis in the mechanical unfolding and refolding of proteins

6.1. INTRODUCTION

Recently, several experimental studies have been reported, in which a single protein molecule was first stretched mechanically, after which the stretching force was quenched, resulting in protein refolding[47-49]. Both protein unfolding and refolding were thus monitored by observing the time evolution of the stretching force and the protein extension. Consistent with the general observation that protein stretching is a nonequlibrium process[141], the protein refolding was seen to follow a path different from that of its unfolding. This hysteresis seen in the behavior of proteins that have load-bearing functions in living organisms is believed to have biological significance[147] allowing those proteins to dissipate large amounts of energy when subjected to mechanical forces.

Cecconi et al[49] observed that when RNAase H was stretched by using optical tweezers, it would undergo a distinct unfolding transition manifested as a drop in the measured force. Upon subsequent retraction of the molecular handles attached to the protein, the latter underwent a different type of conformational change, which the authors of ref.[49] interpreted as a transition to an intermediate conformation that is different from the native fold.

In this paper we report on a simulation "experiment" analogous to that of Cecconi et al[49] with the goal to gain better insight into the dynamics of the transitions that can be seen in the course of unfolding and refolding. The main drawback of the pulling experiments as a method of probing protein folding is the fact that the protein extension is often a poor measure of its "nativeness" [109]. In a simulation study, we have the luxury of simultaneously probing both the extension and the degree of the domain's nativeness, thereby providing a structural interpretation of the transitions observed in the

single-molecule pulling curves. We therefore hope that our results will help experimentalists better understand their data.

Although fully atomistic simulations of mechanical unfolding have been performed[33-35, 59-61, 63], minimalist off-lattice models have recently gained popularity as a tool to study mechanical processes in proteins[109, 161, 167-173]. Since for the present purpose we need a dynamics model that captures the protein folding time scale, we resort to such a minimalist model here.

6.2. MODEL AND METHODS

6.2.1. The protein model

We use the off-lattice model of protein G developed in refs. [104-106]. This model is the same model that we used in Chapter 3-5 of this dissertation. Each aminoacid residue is represented as a single bead that can be hydrophobic, hydrophilic, or neutral. The structure of the protein is then determined by the aminoacid sequence and by the dihedral sequence that specifies the dihedral preferences of the chain.The bead sequence used for the protein G model is LBLBLBLBBNNNLBBLBLBBBNNNLLBLLLBLLB NBBBLBBBBNNNLBBLBLBLBL and its dihedral sequence is EEEEEET EHTHEEEE EEEEHHEHHHHHHHH HHEHTEEEEEEETTTEEEEEEEE

The potential energy of the protein includes contributions from stretching, bending, and torsion potentials, which preserve the peptide bond geometry, and a Lennard-Jones-type potential that accounts for excluded volume interactions as well as for the effective attractive interaction between hydrophobic residues[104-106]. The details of the potential used can be found in Chapter3-5.

6.2.2. Simulating the protein dynamics

Similar to Section 5.2.2 the dynamics of the protein is asumed to be described by the Langevin equation. We used $\zeta = 2.0\ (\sigma^2/m\varepsilon_h)^{-1/2}$ in our simulations. This value is about 25 times lower than the friction coefficient estimated for water[86]. However it is high enough to ensure overdamped regime, where the times scales of dynamics are proportional to the friction coefficient; Therefore our results can be simply rescaled if one

wishes to be more realistic. We will use dimensionless units, in which the energy scale is set by the typical energy of hydrophobic interactions $\varepsilon_h \sim 1$ kcal/mol and the distance is measured in units of $\sigma = 3.8$ Å, which is the average distance between two neighboring α-carbon atoms. The unit of time is then $\tau = (m\sigma^2/\varepsilon_h)^{1/2} \cong 2.5 ps$, where *m* is the average mass of a residue, and the unit of force is $f_0 = \varepsilon_h / \sigma = 18 pN$.

The potential energy of the chain

$$V(\{\mathbf{r}_i\}) = V_c(\{\mathbf{r}_i\}) + V_{pull} \qquad (6.1)$$

consists of the potential V_c of the free chain and a term that describes the constraints that are used to pull the chain.

6.2.2. Pulling methods

Fig. 6.1 shows the native structure of the protein within this model and the mechanical pulling coordinate. In the simulations of the unfolding/refolding dynamics, we used two methods to stretch the protein:

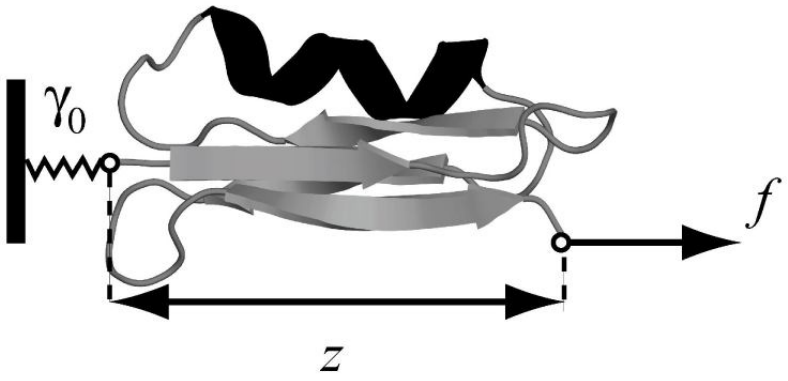

Fig. 6.1. The native structure of the protein G domain model. The *z*-component of one end of the polypeptide chain is constrained by a stiff spring. The protein extension is defined as the z-component of the end-to-end vector of the protein, $z = z_N - z_1$.

Pulling Method 1

In this method the position of the 1st bead is constrained in the *z*-direction by attaching it to a stiff spring with a force constant $\gamma_0 = 100\varepsilon_h / \sigma^2$. Then a softer spring with a spring constant γ is attached to the other end of the chain (the bead number *N*). This spring is used as a "handle" to pull on the molecule. The stiffness of this spring $\gamma = 10^{-2}\varepsilon_h / \sigma^2$ was chosen to roughly correspond to the stiffness of the double-stranded DNA "molecular handles" used in ref. [49]. The pulling is described by the potential:

$$V_{pull} = \gamma_0 z_1^2 / 2 + \gamma (z_N - z(t))^2 / 2 , \tag{6.2}$$

where the end of the soft spring was moved at a constant velocity *u* to stretch the protein and then retracted at the same velocity during the relaxation stage:

$$\begin{aligned} &z(t) = z_0 + ut, 0 \le t \le t_0 \\ &z(t) = z_0 - u(t - 2t_0), t_0 < t \le 2t_0 , \\ &z(t) = z_0, 2t_0 < t \le t_1 \end{aligned} \tag{6.3}$$

where t_0 is the duration of the stretching experiment, t_1 is the total duration of the stretching/relaxation experiment and z_0 is the protein's average extension in the absence of the force. Instead of specifying t_0 beforehand, in each pull we switch from stretching to retraction as soon as the protein extension reaches the value $z_N(t_0) - z_1(t_0) = 43\sigma$ corresponding to the fully extended polypeptide chain. The total experiment duration t_1 is the longer of the two times, $2t_0$ and t_f, where t_f is the time at which the protein first achieves the folded conformation and its root-mean-square deviation from the native conformation satisfies the criterion $RMSD(t_f) = 0.5\sigma$. Our choice of the time t_1 ensures that we monitor the protein until it folds even if it fails to achieve the native structure by the time 2 t_0 when the spring returns to its original position.

Pulling Method 2

In this method, we directly controlled the force that acted on last bead of the chain:

$$V_{pull} = \gamma_0 z_1^2 / 2 - f(t) z_N \tag{6.4}$$

where the force was first increased and then decreased at a constant rate:

$$f(t) = \alpha t, 0 \le t \le t_0$$
$$f(t) = \alpha(2t_0 - t), t_0 < t \le 2t_0 , \quad (6.5)$$
$$f(t) = 0, 2t_0 < t \le t_1$$

where the times t_0 and t_1 are defined as in Pulling Method 1. This situation corresponds to the force-clamp experiments, where the pulling force is controlled directly[47, 48]. Equilibrium free energy surfaces were computed by using the weighted histogram method as implemented in Chapters 3,5.

All simulations were performed at T=0.38 ε_h / k_B , which is below the temperature is T_f=0.41 ε_h / k_B, at which the folded protein conformation becomes thermodynamically favorable.

6.3. RESULTS

6.3.1 The free energy landscape of protein G

We calculate the free energy of our protein as a function of its extension $z = z_N - z_1$ (i.e. the component of the end-to-end vector in the direction of pulling) and the root-mean-square deviation (RMSD) from the native configuration. Our definition of the extension as the *z*-component of the end-to-end distance is more convenient than the absolute value of this distance because it is the former coordinate that couples to the force in our model. In the regime when the force is large enough that the protein is aligned along the force, the difference between the two is immaterial. RMSD is a convenient coordinate to measure the progress of folding since it is close to zero in the native state. The free energy surface *G*(*z*,RMSD) was computed by using umbrella sampling/weighted histogram method and is shown in Fig. 6.2. This surface has two distinct minima, one corresponding to the native state N and the other to a compact denatured state CD that has a nonzero RMSD but an extension comparable to that of the native state.

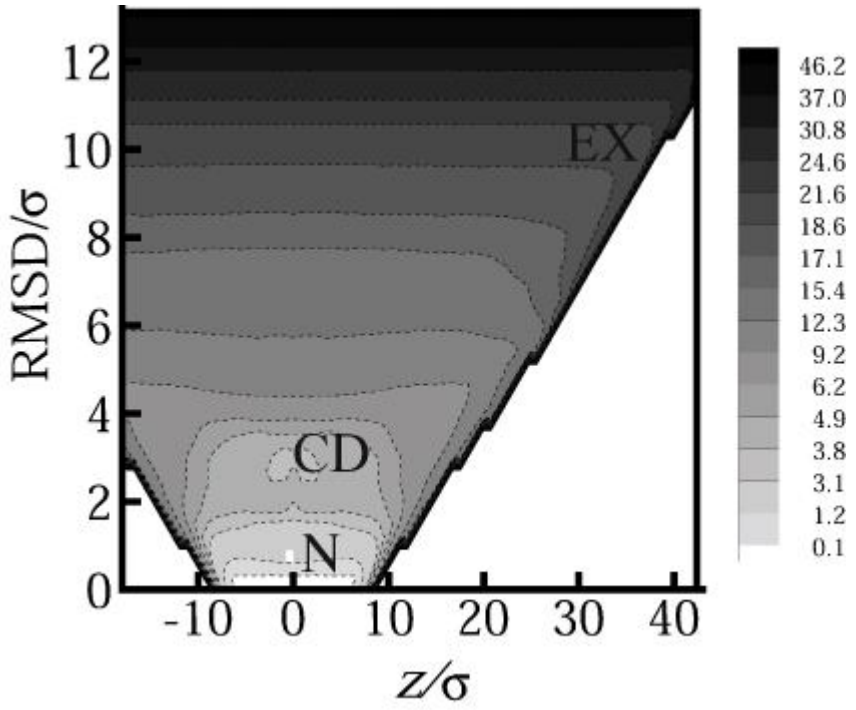

Fig. 6.2. Free energy of the protein G domain as a function of the z-component of the end-to-end vector and the root mean square deviation from the native conformation. The native (N) and the compact denatured (CD) basins of attraction are shown, as well as the extended state EX. .

6.3.2. Details of the hysteresis curve.

In Fig. 6.3a we have plotted the total extension $z(t)-z_1$, (i.e. the sum of the extensions of the soft spring and of the protein) versus the force *f* using the pulling method 1. The force is measured from the deflection of the soft spring, $f=\gamma(z(t)-z_N)$. In Fig. 6.3b we plot the extension of the protein alone, $z=z_N-z_1$, versus the force. Fig. 6.3a roughly corresponds to the experimental situation of ref. [49] where the total extension including that of the molecular handles is measured.

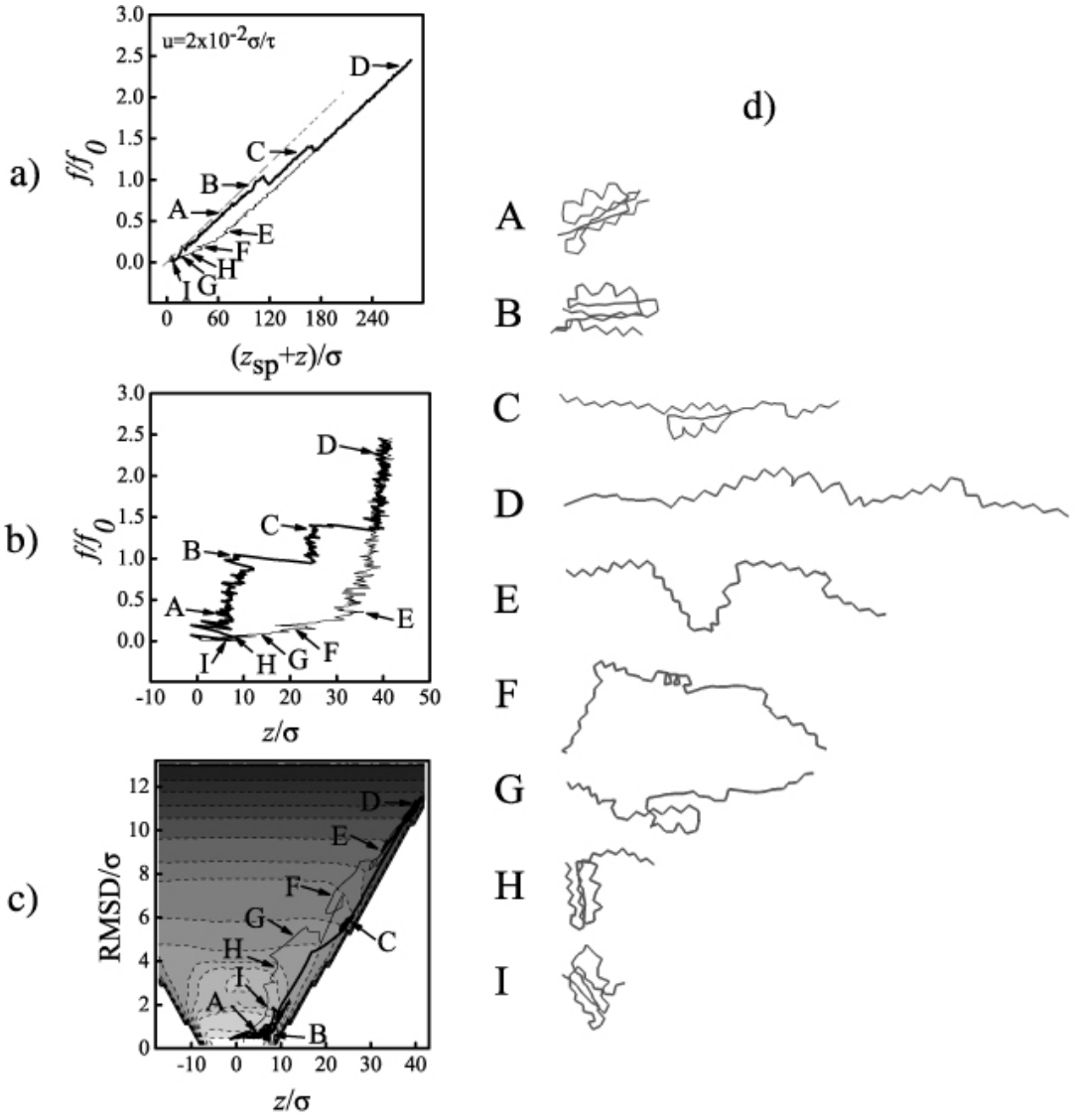

Figure 6.3. The force extension curves corresponding to the stretching and the relaxation of the protein G domain using pulling method 1. a) The total chain extension (i.e. the protein extension plus the spring extension) vs. the measured force. The dashed line shows the force-extension curve ($f = \gamma z$) of the spring itself. b) The protein extension vs. force. c) The trajectories of unfolding and refolding in the two-dimensional space ($z(t)$, RMSD(t)) plotted against the equilibrium free energy surface of the domain. Different stages (A-I) of the unfolding/refolding cycle are indicated. d) Domain structures corresponding to the stages A-I in the unfolding/refolding cycle

From Figs. 6.3a-b, the stretching and the relaxation curves are different. To characterize this hysteresis in more detail, we have placed the milestones A –I at different stages of the stretching/relaxation process. In Fig. 6.3c, we have further plotted the two-dimensional trajectory ($z(t)$, RMSD(t)) of the domain in this process against the equilibrium free energy surface $G(z, RMSD)$. Protein configurations corresponding to the milestones A-I are shown in Fig. 6.3d.

Upon initial stretching the domain stays nearly intact (A-B) while the force rises steeply with the increasing protein extension (see Fig. 6.3b). The stiffness of the domain is much higher than the spring constant γ so that the total extension is mostly comprised of that of the spring itself and the overall stiffness of the chain is close to γ, as seen in Fig. 6.3a.

Once the protein extension reaches $\approx 10\sigma$, a transition B→C takes place, in which the terminal parallel strands of the domain (cf. Fig. 6.1) become separated and the extension increases abruptly. Due to the extra contour length of the chain that is released in this transition, the force drops to a somewhat lower value. Upon further extension, a second transition occurs, in which the rest of the protein's secondary structure is lost (cf. structures C and D in Fig. 6.3d) and the protein attains an extended conformation D. As one proceeds to stretch the protein, the overall stiffness of the extended chain is high so that the force rises abruptly with the protein extension.

Relaxation of the same system follows a different path (D – I). The important distinction between the stretching and the relaxation curves is that the protein chain remains "soft" throughout the stages E – I and that the "clamp" formed by the parallel terminal parallel strands (which is destroyed in the unfolding transition B-C) does not reform until late in the relaxation process. This clamp (see Fig. 6.1) is responsible for the high mechanical stability of this domain with respect to unfolding[33-35, 59, 101, 133, 159, 174], as manifested by the steep rise A → B in the unfolding curve of Fig. 6.2b. At any given value of the domain extension, the protein is less native during relaxation than it is during stretching (cf. Fig. 6.2c) and the structures encountered along the relaxation curve tend to be more random than those during unfolding. However, the relaxation – at this pulling rate – cannot be viewed as an entirely nonspecific collapse: Instead we have

reproducibly observed a distinct structural transition around F for slow or intermediate loading rates.

As a consequence of the difference between the unfolding and refolding pathways, large amount of energy is dissipated during the irreversible stretching process. This dissipation mechanism is known to be important for the function of many natural fibers, adhesives, and composites[147].

6.3.3. The effect of the loading rate

Fig. 6.4 shows the effect of the pulling speed on the unfolding/relaxation curves. As expected, for a slower pulling speed the two curves become closer to one another (Fig. 6.4A). The forces corresponding to the two unfolding transitions are lower than in the case shown in Fig. 6.3. Moreover, the relaxation curve exhibits two transitions that are similar to those seen in the unfolding curve except that they occur at lower forces. Examining the two-dimensional trajectories of unfolding and refolding ($z(t)$, RMSD(t)), we observe that these two pathways become closer to one another as the pulling speed is decreased (Fig. 6.4B).

For a much higher pulling rate (the case shown in Figs. 6.4C,D) the unfolding pathway remains similar to that observed at lower pulling speeds. However the refolding pathway is much different. The refolding trajectory seen in Fig. 6.4D first visits the compact denatured state CD (the collapse stage), where it gets trapped for a while, and then crosses the barrier and attains the native state (the folding stage). If one only monitors the protein extension z then the second stage is hard to discern because it is not manifested by an appreciable change in z. The refolding trajectory in this regime is similar to that found in the case where the force is instantaneously switched off[109, 172]. In particular, Li et al[172] report observation of distinct collapse and refolding stages in their simulations of the mechanical unfolding and refolding of the I27 domain of the muscle protein titin.

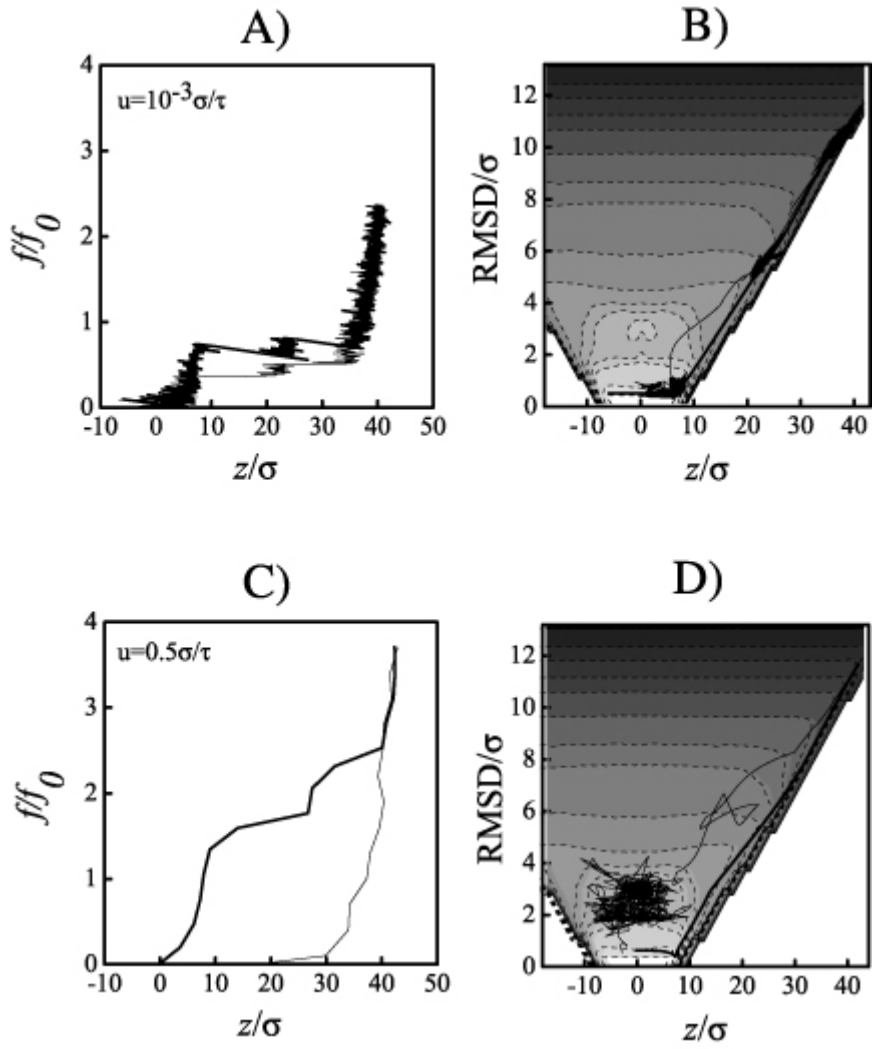

Figure 6.4. Force vs. protein extension and the two-dimensional trajectory ($z(t)$, RMSD(t)) plotted against the equilibrium free energy surface using the pulling method 1. The unfolding and refolding curves are shown as thick and thin lines, respectively. The pulling rate is (A,B) u=$10^{-3}\ \sigma/\tau$ and (C,D) u=$0.5\ \sigma/\tau$.

6.3.4. Force control vs. displacement control in loading

In many single molecule mechanical studies, pulling is achieved by moving a substrate, to which one end of the protein is attached, relative to a flexible force sensor (e.g., an AFM cantilever). However it is also possible to perform a force-clamp experiment, in

which the force that acts on the protein is controlled while its extension is measured[140]. The difference in pulling methods is often reflected in the observed force-extension curves[156]. To account for such differences, here we examine the force-vs.-extension curves and the refolding trajectories obtained in the second method of pulling where the loading force is directly controlled.

To compare the two pulling methods quantitatively, we need the loading rates in each scenario to be similar. In the 2nd pulling scenario, the forces rises linearly at the rate $df/dt = \alpha$. In the first scenario the loading rate can be roughly estimated as $df/dt \approx \gamma u$ assuming that the domain itself is much stiffer than the spring attached to its N-th bead. Fig. 6.5 shows the stretching/relaxation force-vs.-extension curves and the corresponding trajectories in the case of the 2nd pulling method. The loading rate is taken to be $\alpha = \gamma u$, with the values of *u* used in Figs. 6.4 A-D for Figs. 6.5A-D, respectively. The observed force-vs.-extension curves and the folding/refolding trajectories are very similar for both pulling scenarios both in the fast and the slow pulling case. One difference is that, in the slow pulling regime, a drop in force is observed after each unfolding transition for the 1st but not the 2nd pulling method. This drop is due to the slack that is created in the chain when it loses secondary structure and is commonly seen in single-molecule pulling experiments[36, 37, 49, 135]. Clearly, since the applied force is controlled in the 2nd pulling method, no such drop is observed there.

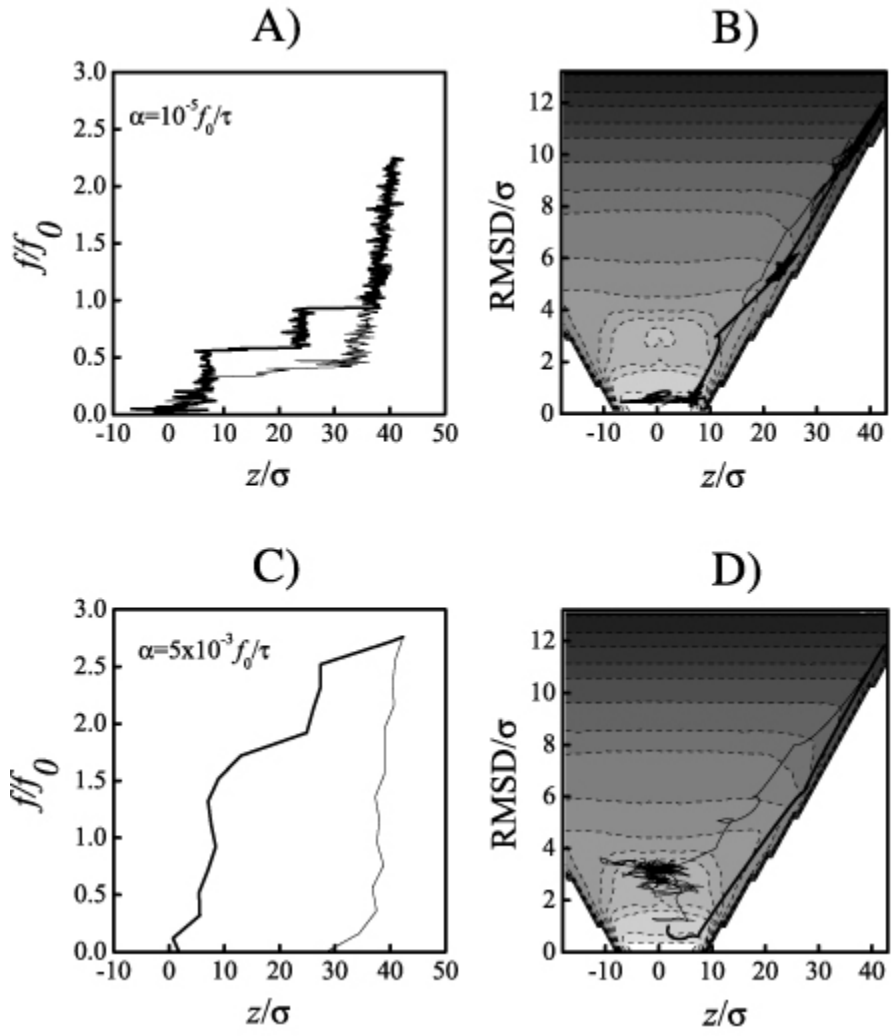

Figure 6.5. The force vs. protein extension curve and the two-dimensional trajectory ($z(t)$, RMSD(t)) plotted against the equilibrium free energy surface using the pulling method 2. The unfolding and refolding curves are shown as thick and thin lines, respectively. The loading rate is (A,B) $\alpha = 10^{-5} f_0 / \tau$ and (C,D) $\alpha = 5 \times 10^{-3} f_0 / \tau$, roughly corresponding to the pulling speeds in the cases (A,B) and (C,D) in Fig. 6.4.

6.4. DISCUSSION

Whether or not a protein subjected to a mechanical loading responds to it in an equilibrium fashion depends on the time scales of the protein's dynamics relative to the loading rate. Since protein folding/unfolding time scales vary within a wide range[175-

177], it is not surprising that under similar experimental (or physiological) conditions different proteins may behave differently. For example, while titin exhibits strong hysteresis in the AFM stretching/relaxation experiments, the myosin coiled coil responds to pulling in an equilibrium fashion at comparable pulling rates[141]. Nonequilibrium response of the load-bearing proteins that constitute building blocks of natural adhesives, composites, and fibers is believed to be an essential mechanism, through which those proteins dissipate large amounts of energy[147, 178].

Hysteresis observed in single-molecule stretching/relaxation curves can also provide information about the protein's free energy landscape and its dynamics. For example, observation of such hysteresis curves has led Cecconi et al[49] to the conclusion that RNAse H has a long-lived intermediate state and allowed them to characterize the kinetics of the transitions to and from this state.

Inferring the underlying free energy landscapes from mechanical experiments is not straightforward, since such experiments monitor only a one-dimensional projection (onto the coordinate equal to the protein extension) of highly multidimensional dynamics. Therefore, certain transitions can be "hidden" in mechanical experiments.

Consider, for example, the unfolding/refolding force-vs.-extension curves computed in our study in the cases of fast and slow loading. When the loading is sufficiently slow, we found that both unfolding and relaxation curves exhibit similar transitions (cf. Fig. 6.4A), although the unfolding transitions tend to take place at a higher force. When the loading rate is fast, we then found that the relaxation curves appear to correspond to continuous collapse, with no distinct transitions evident (cf. Figs. 6.3, 6.4C, 6.5C) . This behavior is readily understood if one considers two-dimensional trajectories ($z(t)$, RMSD(t)) during the refolding stage ($t>t_0$). If u (or α) is very high then the force is quenched to zero almost instantaneously. In this case, the protein undergoes a collapse transition (the transition EX → CD using the notation of Fig. 6.2) followed by a barrier crossing (CD → N) to the folded state (cf. Fig. 6.4D, 6.5D). However since the values of the protein extension are similar in both the collapsed and the native states, the latter transition is not evident from the time-dependence of the protein extension $z(t)$.

In many experimental studies one seeks an interpretation of the observed data in terms of a one-dimensional free energy landscape $G(z)$ as a function of the extension z. Is it possible to rationalize the observed behavior within such one-dimensional picture?

In what follows we show that although this appears possible, there are important constraints on the shape of $G(z)$. To see this, let us focus on the pulling scenario 2 (the force $f(t)$ is controlled in the experiment). We stretch the protein until the force reaches a value $f_{\max}$ and then reduce the force at a rate $df/dt = -\alpha$ (cf. Eq. 6.5). We view our system as a one-dimensional particle moving in an instantaneous potential $G(z)$-fz . Snapshots of this potential at different values of the force are shown in Fig. 6.6 The force tilts the potential $G(z)$ making extended states of the protein more thermodynamically favorable. We consider two different situations. On the left, we schematically depict the commonly used two-state model that has two minima separated by a barrier. At zero force, the native minimum has a lower energy than the minimum corresponding to the denatured state. At $f_{\max}$, the denatured minimum is lower. As the force is lowered during the relaxation stage of the experiment, the shape of the potential changes; However, a barrier between the denatured and the native minima always exists in this case. Therefore, we expect a barrier crossing refolding event to take place regardless of the rate at which the force is changed.

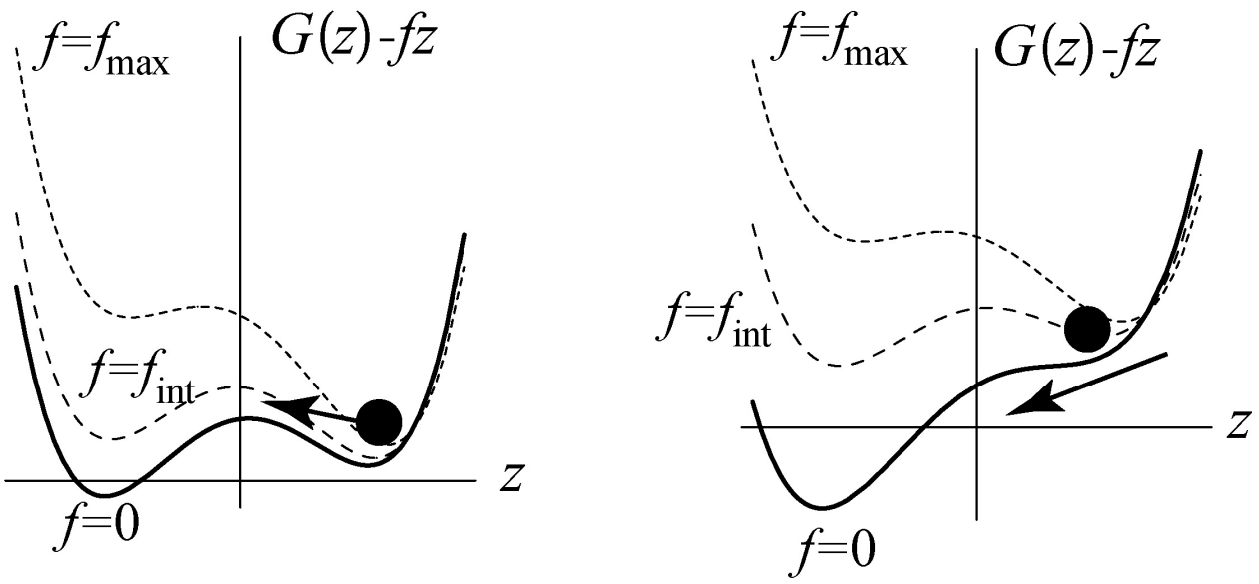

Figure 6.6. Hypothetic one-dimensional free energy profile (as a function of the protein extension) experienced by the protein at different values of the stretching force ($0 < f_{int} < f_{max}$). Left: the two-state model includes native and denatured minima separated by a barrier. The native minimum is lower at f=0 and the denatured state is thermodynamically favored when the force is high (the case $f=f_{max}$). The refolding transition always involves crossing a barrier between the denatured and the native states. Right: The free energy landscape (at f=0) has a single minimum[109]. When a stretching force is applied, two minima exist. If the force is lowered slowly, barrier crossing back to the native minimum will take place at some intermediate value of the force, $f = f_{int}$. However if the force is quenched instantaneously (or very fast) then the transition back to the native state is downhill in free energy.

The other scenario shown in Fig. 6.6 (right) corresponds to the case where the potential $G(z)$ has only a single minimum. At finite values of the force f two minima exist, separated by a barrier. If the force is lowered slowly then we expect a barrier-crossing-type transition back to the native minimum to take place at some intermediate value of the force f_{int}. The situation is however, different if the force is lowered quickly (high values of α). For example, in the limit where the force is switched off

instantaneously, our particle will experience a downhill potential corresponding to f=0, resulting in a continuous downhill relaxation process.

The downhill scenario of Fig. 6.6 is therefore consistent with the general rate dependence of the stretching/relaxation force-vs.-extension curves observed in our studies. In the case of intermediate loading/unloading rates distinct transitions are observed and for fast rates the refolding curve is continuous showing no transitions. [Notice that two, rather than one, transitions are generally seen in Figs. 6.3-6.5. This feature can be reproduced in a one-dimensional model that has two barriers at intermediate values of the force]. At the same time, the two-state picture shown on the left is inconsistent with our simulations as it fails to explain the continuous refolding curves observed in the case of fast unloading rate.

We conclude that the rate dependence of the observed force-vs.-extension curves can provide useful links between the mechanical response and the underlying free energy landscape.

Chapter 7

Simulations of the untying of molecular friction knots between individual polymer strands

7.1. INTRODUCTION

Molecular knots tied in individual polymer strands have fascinated researchers from many fields, see, e.g., [179-188]. Recent progress in single molecule manipulation techniques (reviewed in [189-191]) has enabled several experimentalists to tie a variety of knots in single biopolymer strands by using optical tweezers [192, 193]. With these techniques, it is possible to create individual polymeric structures of complex topology and to study their dynamics under controlled mechanical tension. Such structures may prove useful in nanotechnology applications. In addition, knotted DNA structures are common in biology; Studies of the intra-strand interactions in molecular knots may provide new insights into the molecular forces that control the DNA dynamics and the organization of the chromatin fiber [179].

Motivated by the experimental advances, this chapter discusses the dynamics of a friction knot formed by a pair of polymer molecules. Friction knots, such as the square knot shown in Fig. 7.1, are commonly used by sailors and climbers to join a pair of ropes. Pulling at the ends of the ropes in Fig. 7.1 jams the knot so that the ropes remain connected regardless of the applied force. An elegant theory exists [194], which explains this behavior and shows that if the friction coefficient between the ropes exceeds a certain knot dependent critical value then the two ropes will not come apart no matter how hard one pulls on them. This theory also explains why the modification of the square knot known as granny knot (also shown in Fig. 7.1) is a poor way of splicing two ropes that fails at a low force.

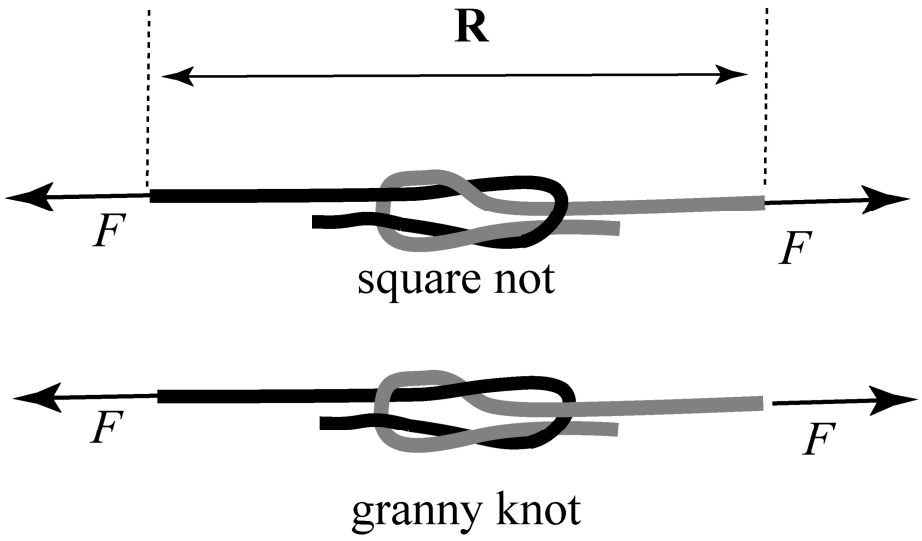

Fig. 7.1. The square knot and the granny knot.

In view of the advances reported in refs. [192, 193], an analogous microscopic arrangement involving a pair of interwoven polymer strands that are being separated by a mechanical force can be devised experimentally. In addition, this type of scenario has been suggested as a possible mechanism for slow relaxation in polymer melts[180, 185]. A friction knot scaled down to molecular dimensions will no longer hold indefinitely under applied tension, the two separated strands being always favored thermodynamically. Microscopically, eventual failure of the knot is caused by thermal fluctuations – a macroscopic analog of this would be to pull on the ropes joined by the knot while shaking them vigorously, which would obviously facilitate their separation.

In this chapter, we show that a signature of the knot jamming effect can be found when one examines the *dynamics* of molecular knots. Similarly, to the macroscopic case, this behavior is caused by the interaction between the two strands. Specifically, bumpiness of the energy landscape of the two interacting strands may be viewed as leading to a tension-dependent effective friction, which is however no longer static[195]. The sliding of one strand relative to the other is accomplished via thermally activated transitions from one local energy minimum to the next so that the roughness of the energy landscape determines the rate at which the slippage of the knot takes place. If this

roughness increases with the force F exerted to separate the strands then it may take *longer* to unravel a molecular friction knot when the force is *higher*. We will refer to this as "strong knot" behavior as opposed to "weak knots" that untie faster when higher forces are applied. Strong knots are reminiscent of molecular "catch-bonds" observed in forced dissociation of some biomolecular complexes (see, e.g., [196, 197] and refs. therein).

The results of our simulations described below show that both strong and weak knot behavior can be observed, depending on the knot type and the nature of the polymers.

6.2. MODEL AND METHODS

6.2.1. Polymer model and simulation method

We used a polymer model, in which monomers were represented as single beads. The potential energy of a strand, as a function of the position $\mathbf{r}_i$, i=1, …, N, of each bead, is given by:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V_{\text{bond}} + V_{\text{bend}} + V_{\text{non-bonded}} \qquad (7.1)$$

The potential V_{bond} accounts for the connectivity of the chain and assumes that each bond is a stiff harmonic spring,

$$V_{\text{bond}} = \sum_{i=2}^{N} k_b (|\mathbf{u}_i| - l_{i,i-1})^2 / 2 . \qquad (7.2)$$

Here $\mathbf{u}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$ is the bond vector and $l_{i,i-1}$ is the equilibrium bond length given by: $l_{i,i-1} = \rho_i + \rho_{i-1}$, where ρ_i, ρ_{i-1} are the effective sizes (i.e., the van der Waals radii) of the i-th and (i-1)-th monomers. We have constructed polymer chains consisting of two types of beads (see below), bead A and bead B with $\rho_A = \sigma/2$ and $\rho_B = 5\sigma/4$, where σ is the equilibrium A-A bond length. The spring constant is taken to be k_b = 500 ε/σ^2, where ε sets the energy scale. The bending potential is:

$$V_{\text{bend}} = \sum_{i=2}^{N-1} k_\theta (\theta_i - \theta_0)^2 / 2 \qquad (7.3)$$

where $\theta_0 = \pi$ is the equilibrium bending angle, θ_i is the angle between $\mathbf{u}_i$ and $\mathbf{u}_{i+1}$, and k_θ is the bending spring constant. The value $k_\theta = 5\varepsilon/(rad)^2$ used in our simulations corresponds to a persistence length of 15 monomers at temperature T=0.4 ε/k_B.

The energy $V_{\text{non-bonded}}$ describes the interaction between pairs of monomers that are not covalently bonded. We took this interaction to be purely repulsive:

$$V_{\text{non-bonded}} = \sum_{|i-j|\geq 2} \varepsilon \left[\left(\frac{\rho_i + \rho_j}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{12} \right] . \qquad (7.4)$$

In addition to interactions among non-bonded monomers within each chain, the same pairwise potential was used to describe the interactions between pairs of monomers belonging to different chains.

We further assumed that the dynamics of the chains were governed by the Langevin equation of the form $m\ddot{\mathbf{r}}_i = -\xi\dot{\mathbf{r}}_i - \partial V/\partial \mathbf{r}_i + \mathbf{f}_r(t)$, where $\mathbf{r}_i$ is the position of the i-th bead, m is its effective mass, ξ is the friction coefficient, for which we chose the value $\xi = 2.0\left(\sigma^2/m\varepsilon\right)^{-1/2}$, and $\mathbf{f}_r(t)$ is a random δ-correlated force satisfying the fluctuation-dissipation theorem. This equation was solved by using the velocity Verlet algorithm as described in [87]. In reporting our data below, we use dimensionless units of energy, distance, time, and force respectively equal to ε, σ, $\tau_0 = (m\sigma^2/\varepsilon)^{1/2}$, and $F_0 = \varepsilon/\sigma$.

In the beginning of each simulation, we connect the two strands by a square or granny knot positioned such that the contour length of the polymer chain between the knot and the end of each strand is the same. A force $F_p = 4.0\ F_0$ is then applied to the ends of one strand and $-F_p$ to the ends of the other strand, for an initial time of $t_p = 2000\ \tau_0$. This force pre-tensions the knot without considerably affecting its initial location relative to the ends of each polymer. After preparing the initial state of the knot this way, we start simulation at $t = 0$, with a force F applied to the first bead (i=1) of one chain and the opposite force acting on the last bead (i=N) of the other one. We monitor the presence of the knot by projecting the polymers' configuration onto a plane that is parallel to the

direction of the force and computing the chain intersections in this plane [188]. The knot disappears when the number of intersections falls below 6. This allows us to measure the time τ before the knot disappears.

7.3. RESULTS

7.3.1. Dynamics of knot untying

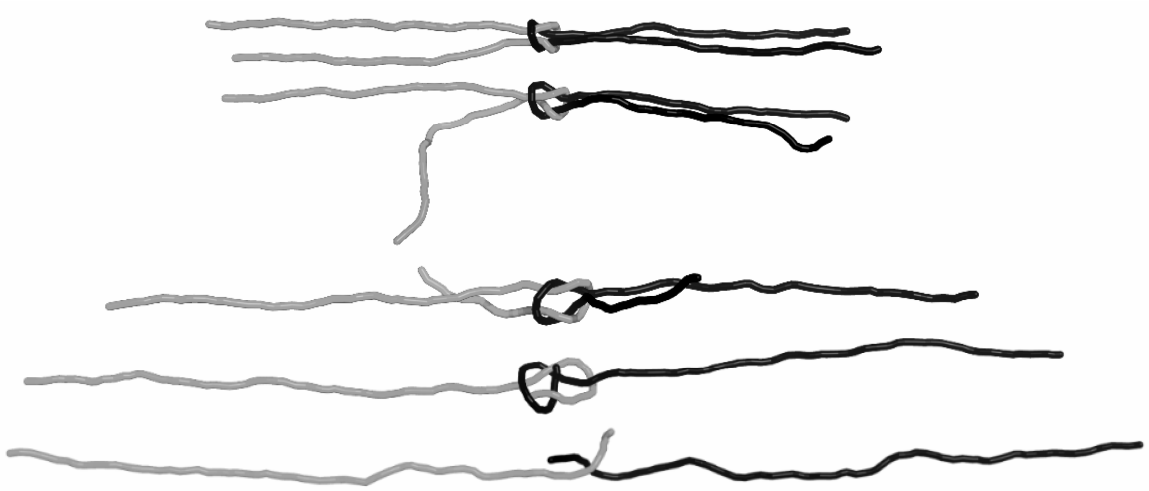

Fig. 7.2. Snapshots of two polymer strands joined by the square knot and subsequently pulled apart, as observed in a Langevin Dynamics simulation. The time increases from top to bottom. The snapshots were generated with the help of the PyMol software [198].

Figure 7.2 illustrates the dynamics of knots as observed in the simulations. At the beginning of each simulation, the two strands are connected by a square or a granny knot positioned such that the contour length of the polymer chain between the knot and the end of each strand is the same. The first monomer of one strand and the last monomer of the other are subjected to forces F and $-F$, respectively (cf. Fig. 7.1) and the knot is monitored as a function of time until it disappears and the two strands become separated. To describe the knot's response to the pulling force, we measure the mean time τ before the knot disappears and monitor the distance R between the monomers at which the

forces are applied. The observed trajectories $R(t)$ typically display an initial transient behavior that has to do with the particular way the knot is prepared followed by an approximately linear increase in the distance R. When reporting the average strand separation rate, $\langle dR/dt \rangle$, we discard the transient part.

7.3.2. Effect of the polymer sequence and of the knot type on the untying time

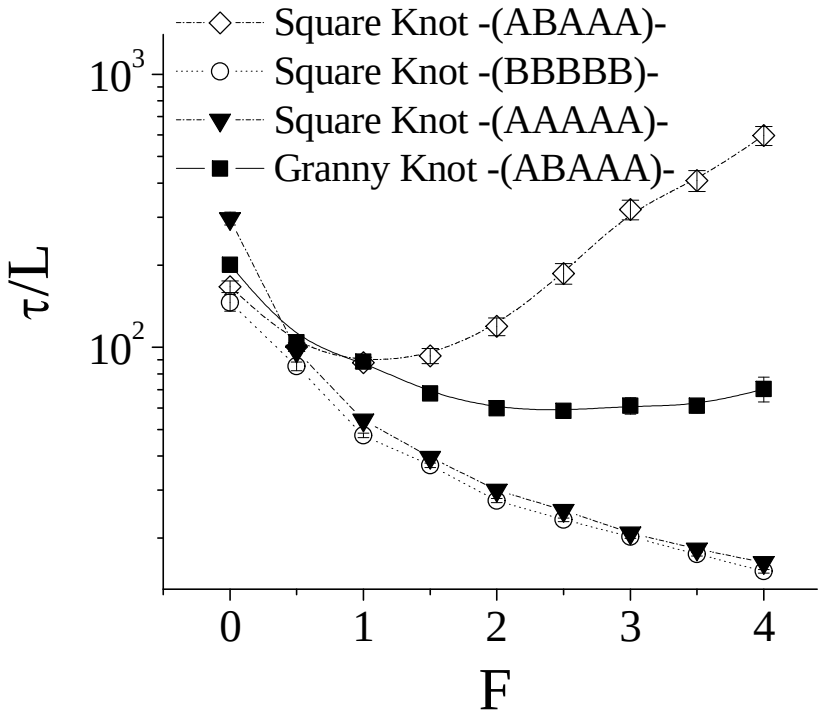

Fig. 7.3. The mean time τ (normalized by the polymer contour length L) for the untying of the square and the granny knots as a function of the force pulling the polymer strands apart for different polymer chains and different knots. The dimensionless units of force, time, and temperature used here and in other figures are explained in section 6.2.1.

Figure 7.3 shows the average separation time $\tau(F)$ as a function of the pulling force. When both strands consisted of 88 monomers of the same type (A_{88} or B_{88}), this time decreased monotonically with the increasing force F thus showing that our homopolymers formed weak knots. However when each strand was a heteropolymer with the sequence $AAA(ABAAA)_{17}$, the separation time initially decreased and then increased

with the increasing force thus exhibiting the strong knot behavior at high forces. This difference between the knot strength for homopolymers and heteropolymers can be rationalized based on the above view that the knot slippage rate is determined by the roughness of the energy landscape. The energy landscape of two intertwined chains of beads of variable size tends to be rougher and involves multiple configurations where the two chains can temporarily "snag", as the reader can establish by experimenting with two suitable pieces of jewelry.

Like its macroscopic counterpart, the molecular version of the granny knot fails much more easily than the square knot: When the same two heteropolymer strands were joined by the granny knot, the time $\tau(F)$ first decreased with the increasing force and then became nearly force-independent, as also shown in Fig. 7.3.

7.3.3. Effect of the temperature on the untying time

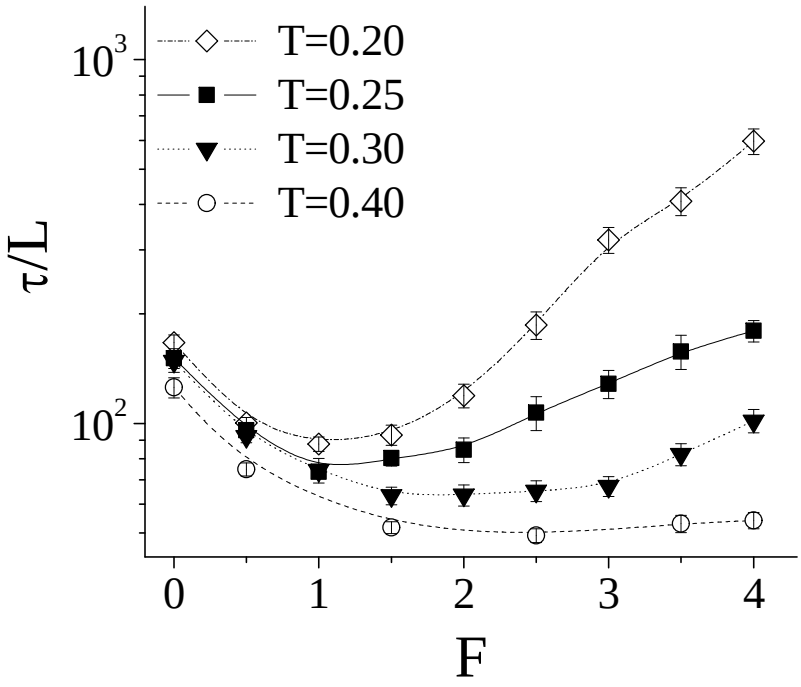

Fig. 7.4. The mean time τ for the untying of the square knot as a function of the force pulling the polymer strands apart at different temperatures.

It is reasonable to expect that the slowdown in the untying dynamics of molecular friction knots would be more pronounced at low temperatures, when there is less thermal motion. Indeed, this is what we see in Fig. 7.4, which explores the dependence of the mean strand separation time $\tau(F)$ on temperature.

7.3.5. The tilted periodic potential model

To rationalize the above findings and to understand how forces can influence the knot dynamics, consider the simplest model that relates effective friction to the features of the energy landscape [195]. Suppose the relative sliding of the two strands can be viewed as one-dimensional diffusive motion along the coordinate *R*; The Brownian dynamics along *R* is described by the stochastic equation $\eta\dot{R} = F - dV_F(R)/dR + f_r(t)$, where η is a friction coefficient and $f_r(t)$ is a random force that satisfies the standard fluctuation-dissipation relationship. The potential $V_F(R)$ is our model for the corrugated energy landscape for inter-strand interaction. We will assume it to be periodic, $V_F(R) = v(F)\sin(2\pi R/a)$. [A random potential may be a better model; however it will not qualitatively change our conclusions]. The effect of the force *F* is to tilt the overall potential, $V_F(R) \to V_F(R) - FR$, and also to change the degree of corrugation of the inter-strand potential, which is described by the parameter $v(F)$.

The average velocity of diffusion along *R* can be evaluated exactly[199]:

$$\langle dR/dt \rangle = \frac{k_B T}{\eta a}(1 - e^{-aF/k_B T})\left\{ \int_{x_0}^{x_0+1} dx \int_{x-1}^{x} dy e^{(v(F)\sin(2\pi x) - v(F)\sin(2\pi y) + Fay - Fax)/k_B T} \right\}^{-1}, \quad (7.5)$$

where the result does not depend on x_0. The amplitude $v(F)$ should increase with *F* to describe the tendency of the potential to become more corrugated. For low enough forces we can assume this to be a linear function: $v(F) = Fd$, where the coupling parameter *d* has the units of length. Depending on the value of *d*, there are two regimes illustrated in Fig. 7.5.

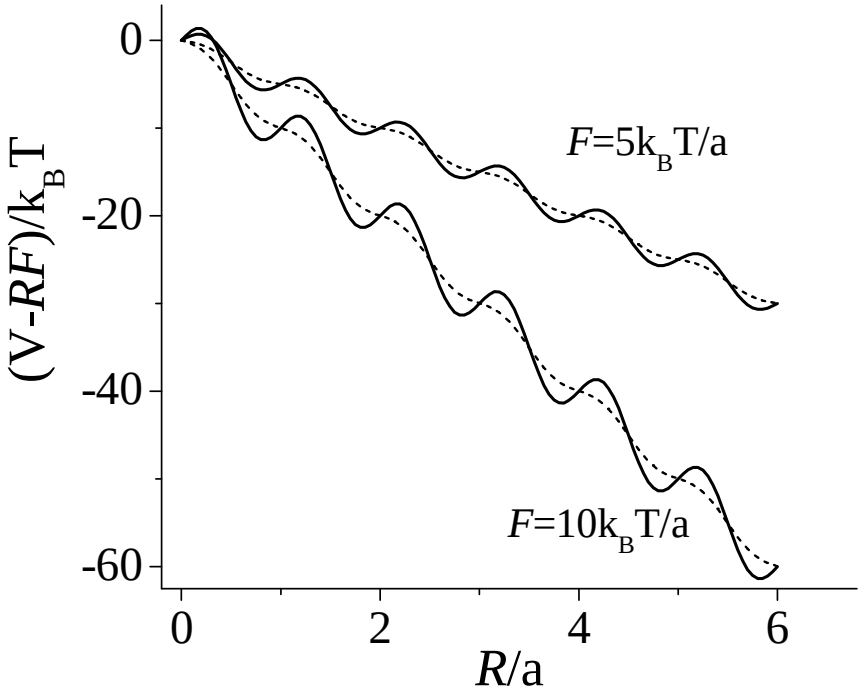

Figure 7.5. Periodic potential tilted by a force F, $V(R) = Fd\sin(2\pi R/a) - FR$. The dashed lines correspond to d=0.1a and the solid lines correspond to d=0.35a.

(1). If $d < d_c = a/2\pi$ then the potential $V_F(R) - FR$ is barrierless and decreases monotonically with F. In this case the sliding speed $\langle dR/dt \rangle$ should increase with the increasing force and the strand separation time should decrease monotonically. This is the weak knot behavior.

(2) However if $d > a/2\pi$ then the barriers in $V_F(R) - RF$ will become higher when F is increased. When they are higher than k_BT we expect this to lead to a decrease in $\langle dR/dt \rangle$. This is the strong knot regime. At low forces, evaluating Eq. 7.5 analytically to 1st order in F we see that it approaches the free drift limit $\langle dR/dt \rangle = \langle dR/dt \rangle_{free} = F/\eta$. The average sliding speed $\langle dR/dt \rangle$ thus first increases and then decreases with F, which explains the minimum of $\tau(F)$ seen in Figs. 7.3-7.4.

From Eq. 7.5, inter-strand interaction slows down the strand separation by the factor:

$$\frac{\langle dR/dt\rangle_{free}}{\langle dR/dt\rangle} = \frac{\tilde{F}}{1-e^{-\tilde{F}}}\left\{\int_{x_0}^{x_0+1} dx \int_{x-1}^{x} dy e^{(d/a)\tilde{F}[\sin(2\pi x)-\sin(2\pi y)]+\tilde{F}(y-x)}\right\}, \qquad (7.6)$$

which only depends on two parameters, the dimensionless force $\tilde{F} = Fa/k_BT$ and the dimensionless coupling strength d/a. We therefore expect that if we plot the drift velocity (normalized by $\langle dR/dt\rangle_{free}$) vs. F/T, all the data obtained at different temperatures and different forces should collapse onto the same curve. Fig. 7.6 shows that this is indeed the case: When the data used to generate Fig. 7.4 are represented in this manner, a single curve (to within statistical errors) is obtained, suggesting that our simple one-dimensional model captures the right physics.

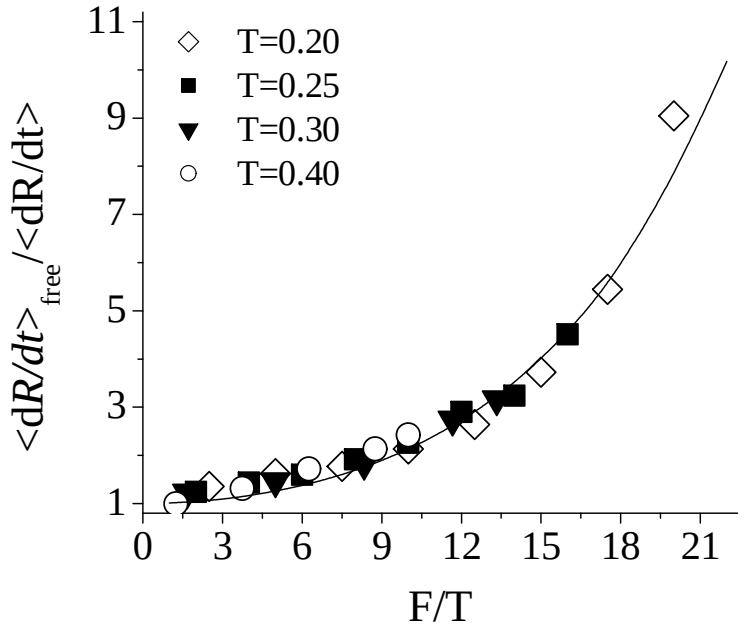

Fig. 7.6. Interaction between two polymer strands within the square knot slows down their separation by a factor $\langle dR/dt\rangle_{free}/\langle dR/dt\rangle$, which is plotted here as a function of *F*/*T* for different temperatures. The data used are the same as that for Fig. 7.5. The solid line is the best fit of the data obtained by using Eq. 7.6, where *a*=0.315 and *d*=0.141.

7.4. DISCUSSION

Maddocks & Keller theory[194] predicts that the friction coefficient between two ropes must exceed a knot-type dependent critical value for the knot to hold. Our model's prediction for molecular friction knots is similar: The value of the coupling parameter d/a depends on both the knot type (which determines how the tension in the polymer strands is transmitted into the intra-strand effective friction[194]) and the nature of the polymer strands. As noted above, in order for a knot to be strong, this parameter must exceed a certain critical value. The weakness of the granny knot and of the square knot between two homopolymer strands observed here can be interpreted as a consequence of the coupling being too low.

Chapter 8

Conclusion

The aim of this thesis was to obtain theoretical insights into the dynamics of mechanically driven biomolecular processes. For this purpose we used Langevin dynamics simulations of a minimalistic off-lattice models of biomolecules.

1. A pore unfolds a protein when its size falls below a critical value. The protein stability inside a cylindric tunnel crucially depends on the tunnel diameter D. When the tunnel is wider than a critical value, $D \geq D_c$, we found that the native state of the peptide is stabilized. However for $D < D_c$, the protein unfolds inside the tunnel. For the peptide studied by us, the value of the critical diameter D_c separating these two regimes is found to be ~13.7 Å, which is comparable to the average diameter of the ribosomal tunnel (~15 Å). This finding has important implications for the process, in which peptides exit the ribosome tunnel: Given that different peptides might have different D_c, proteins may or may not exit the tunnel as folded. When $D< D_c$, the protein undergoes a series of partially folded or misfolded states during the translocation process and the dynamics of translocation is very different than those of an unstructured peptide both in the narrow and the wide tunnel cases.

2. The protein unfolding mechanisms in two mechanical process, mechanically driven translocation and mechanical AFM-type stretching, are different. In the case of translocation, the unfolding mechanism depends on the pore diameter, on the magnitude of the pulling force and on whether the force is applied at the N- or the C-terminus of the chain, and involves intermediates that are not observed when the same protein is stretched between its C- and N-termini. The magnitude of the unfolding rate and its force dependence are different in both cases. This finding is important because an analogy between the two types of processes is often drawn in the literature [9, 11, 98]. Given that single-molecule studies of protein translocation are only beginning to emerge while the AFM studies of mechanically stretched proteins are widespread, it is tempting

to extend the insights from the latter to protein translocation. Our results, however, suggest that this analogy may be limited.

3. The mechanism of protein translocation is different from that of a homopolymer, and force-driven translocation kinetics of proteins cannot be well decribed by simple phenomenological theories. Protein translocation mechanisms involve intermediates and multiple barriers. Consequently, the translocation rate exhibits a pulling force dependence that is more complex than a simple exponential function expected on the basis of simple phenomenological models of translocation.

The free energy cost of squeezing an initially folded protein into the pore is observed to be different than that in the case of translocation of a random-coil-like homopolymer. In homopoymer translocation, the free energy cost required to accommodate a homopolymer is entropic in origin, since the entropy of the polymer constrained by the pore is lower than that of the free random coil. In the case of protein translocation, on the other hand, unfolding of the domain is associated with an entropy increase, while the increase in the free energy comes from the enthalpic costs of unfolding protein .

4. Translocation kinetics of β-hairpin peptides vary with their stability. Our simulations provide an explanation of the experimental finding that more stable β-hairpin forming peptides translocate more slowly. Highly unfolded peptides enter the pore in an extended conformation, resulting a fast translocation event. In contrast, folded β-hairpins translocate more slowly.

5. Translocation into a narrow pore is a barrier-crossing event but this view of translocation is not applicable for high forces. We observe that first-order kinetics, which is a sign of barrier crossing process, becomes invalid at high forces where the dynamics can be described as a downhill diffusive motion along a linear potential of mean force.

6. Mechanical and chemical/thermal denaturation of proteins generally occur via different mechanisms. In order to understand the difference between the chemical and mechanical unfolding of proteins, we have explored in detail the slice of the multidimensional free energy landscape that is accessible via mechanical pulling

experiments and chemical denaturation. We observed that the free energy profile along the mechanical reaction coordinate has a staircase shape while the free energy along a chemical reaction coordinte has a double well shape. We demonstrate that the stretching forces required to destabilize the native state thermodynamically are larger than those expected on the basis of previous experimental estimates of the free energy landscapes of proteins because mechanical and chemical/thermal denaturation cannot be simply explained in terms of a simple one-dimensional free energy landscape . This finding is consistent with the recent experimental studies [47] indicating that proteins may refold even in the presence of a substantial stretching force.

7. Spontenous folding and refolding under applied force of proteins generally occur via different mechanisms. We further found that refolding trajectories of i) spontenous folding and ii) refolding under applied force follow different paths on the two-dimensional free energy profile. These paths cannot be discerned from the force-vs.-extension curves because the protein extension has similar values in both the folded and the collapsed states. However, unfolding trajectories are always close to the steepest descent path on the multidimensional surface.

8. Refolding of a multidomain protein chain under an applied force can be viewed as a downhill diffusion along a linear potential. We have shown that for certain temperatures the free energy of a protein chain consisting of multiple domains is a linear function of the chain extension. We propose that the recently observed “slow phase” in the refolding of proteins under mechanical tension [47] may be viewed as downhill diffusion in such a linear potential.

9. Mechanical unfolding and refolding paths are different and depend on the loading rate. We study the dynamics of transitions seen in the course of unfolding and refolding on pulling the experiment in ref.[49]. We observed that force-vs.-extension curves for stretching and refolding are generally different and depend on the loading/unloading rate: When the rate is slow, both stretching and relaxation curves exhibit similar structural transitions, and both the unfolding and the refolding pathways resemble one another. However, when the loading/unloading rate is fast, the observed relaxation curve is continuous and displays no distinct transitions. We show that in this

case the domain relaxation process involves a collapse followed by a folding transition. The observed behavior cannot be modeled within a simple one-dimensional two-state model of mechanical unfolding.

10. Signatures of the knot jamming effect can be found when in molecular knots tied on polymer strands. We studied the dynamics of molecular friction knots formed by a pair of polymer strands. Depending on the knot type and the nature of the polymer, we observed two different behaviors, when the force F is exerted to separate the polymer strands. The time τ during which the knot stays tied, increases with the force F applied to separate the strands in strong knot behavior, and the opposite is seen in weak knot behavior as the force F is applied. Certain types of heteropolymers tied into a square knot, show strong knot behavior, while the same heteropolymers forming a granny knot as well as homopolymers forming a square knot show weak knot behavior in the simulations. Our finding that a polymer strand can show strong knot behavior is important, because it has not been observed in experiments.

We have shown that the relative sliding of strands depends on the interchain friction, which is tension-dependent and comes from the roughness of the energy landscape. When the roughness of the energy landscape increases with the force F exerted to separate the strands, then we observed the strong knot behavior. A corrugated energy landscape of interstrand interactions succesfully predicts the outcome of the dynamics of the square knot in simulations.

References

1. Kaemmli, U.K. and M. Favre, J. Mol. Biol., 1973. **80**: p. 575.
2. Choi, K.M. and R. Brimacombe, Nucl. Ac. Res. Mol. Biol., 1998. **26**(4): p. 887-895.
3. Gabashvili, I.S., S.T. Gregory, M. Valle, R.A. Grassucci, M. Worbs, M.C. Wahl, A.E. Dahlberg, and J. Frank, Mol. Cell., 2001. **8**: p. 181.
4. Hardesty, B. and G. Kramer, Prog. Nuc. Ac. Res. Mol. Biol., 2001. **65**: p. 41-66.
5. Lingappa, V.R., J. Chaidez, C.S. Yost, and J. Hedgpeth, Proc. Natl. Acad. Sci USA, 1984. **81**(2): p. 456-460.
6. Driselkelmann, B., Microbiol. Rev., 1994. **58** p. 293–316.
7. Matouschek, A., A. Azem, K. Ratliff, B. Glick, K. Schmid, and G. Schatz, EMBO J, 1997. **16**(22): p. 6727-36.
8. Palade, G.E., Science, 1975. **189**: p. 347-358.
9. Prakash, S. and A. Matouschek, TIBS, 2004. **29**(11): p. 593-600.
10. Huang, S.H., K.S. Ratliff, M.P. Schwartz, J.M. Spenner, and A. Matouschek, Nat. Struct. Biol., 1999. **6**(12): p. 1132-1138.
11. Matouschek, A., Curr. Opin. in Struct. Biol., 2003. **13**(1): p. 98-109.
12. Nissen, P., J. Hansen, N. Ban, P.B. Moore, and T.A. Steitz, Science, 2000. **289**(5481): p. 920-930.
13. Verschoor, J.R., S. Warner, S. Srivastava, R.A. Grassucci, and J. Frank, Nuc. Ac. Res., 1998. **15**: p. 655-661.
14. Yonath, A., K.R. Leonard, and H.G. Wittmann, Science, 1987. **236**(4803): p. 813-816.
15. Netzer, W.J. and F.U. Hartl, Nature, 1997. **388**(6640): p. 343-349.
16. Whitley, P., I. Nilsson, and G. vonHeijne, J. Biol. Chem., 1996. **271**(11): p. 6241-6244.

17. Morrissey, M.P., Z. Ahmed, and E.I. Shakhnovich, Polymer, 2003. **45**: p. 557.
18. Huang, S., K.S. Ratliff, and A. Matouschek, Nat. Struct. Biol., 2002. **9**: p. 301-307.
19. Matouschek, A., A. Azem, K. Ratliff, B. Glick, K. Schmid, and G. Schatz, EMBO J, 1997. **16**(22): p. 6727-36.
20. Masatoshi Esaki, Takashi Kanamori, Shuh-ichi Nishikawa, and T. Endo, Proc. Natl. Acad. Sci USA, 1999. **96**: p. 11770-11775.
21. Tian, P. and I. Andricioaei, J. Mol. Biol., 2005. **350**(5): p. 1017-1034.
22. West, D.K., D.J. Brockwell, and E. Paci, Biophysical J., 2006. **91**: p. L51-L53.
23. Wickner, W. and R. Schekman, Science, 2005. **310**: p. 1452-1456.
24. Erickson, H.P., Proc. Natl. Acad. Sci USA, 1994. **91**: p. 10114-10118.
25. Hegner, M., S.B. Smith, and C. Bustamante, Proc. Natl. Acad. Sci USA, 1999. **10109-10114**.
26. Marszalek, P.E., Y.-P. Pang, H. LI, J. Yazal, A.F. Oberhauser, and J.M. Fernandez, Proc. Natl. Acad. Sci USA, 1999. **96**: p. 7894-7898.
27. Shaub, A., Nature Cell. Biol, 1999. **1**: p. E173-E175.
28. Zhong, C., M. Chrzanowska-Wodnicka, J. Brown, A. Shaub, A.M. Belkin, and K. Burridge, J. Cell Biol., 1998. **141**: p. 539-551.
29. Oberhauser, A.F., P.E. Marszalek, H.P. Erickson, and J.M. Fernandez, Nature, 1998. **393**(6681): p. 181-185.
30. Craig, D., A. Krammer, K. Schulten, and V. Vogel, Proc. Natl. Acad. Sci USA, 2001. **98**: p. 5590–5595.
31. Krammer, A., H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, Proc. Natl. Acad. Sci USA, 1999. **96**: p. 1351-1356.
32. Bao, G., J. Mech. and Phys. Sol., 2002. **50**: p. 2237-2274.
33. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophysical J., 1998. **75**(2): p. 662-671.
34. Lu, H. and K. Schulten, J. Chem. Phys., 1999. **247**(1): p. 141-153.

35. Marszalek, P.E., H. Lu, H.B. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, and J.M. Fernandez, Nature, 1999. **402**(6757): p. 100-103.
36. Rief, M., J.M. Fernandez, and H.E. Gaub, Phys. Rev. Lett., 1998. **81**(21): p. 4764-4767.
37. Rief, M., M. Gautel, F. Oesterhelt, J.M. Fernandez, and H.E. Gaub, Science, 1997. **276**(5315): p. 1109-1112.
38. Khorasanizadeh, S., I.D. Peters, T.R. Butt, and H. Roder, Biochemistry, 1993. **32**: p. 7054-7063.
39. Rief, M., J. Pascual, M. Saraste, and H.E. Gaub, J. Mol. Biol., 1999. **286**(2): p. 553-561.
40. Becker, N., E. Oroudjev, S. Mutz, J.P. Cleveland, P.K. Hansma, C.Y. Hayashi, D.E. Makarov, and H.G. Hansma, Nat. Mat., 2003. **2**(4): p. 278-283.
41. Hayashi, C.Y. and R.V. Lewis, J. Mol. Biol., 1998. **275**: p. 773-780.
42. Oroudjev, E., J. Soares, S. Arcdiacono, J.B. Thompson, J.B. Fossey, and H.G. Hansma, Proc. Natl. Acad. Sci USA, 2002. **99**: p. 6460-6465.
43. Bustamante, C., Y.R. Chemla, N.R. Forde, and D. Izhaky, Ann. Rev. Biochem., 2004. **73**: p. 705-748.
44. Cornish, P.V. and T. Ha, ACS Chem. Biol., 2007. **2**: p. 53-61.
45. Ritort, F., J. Phys. : Condens. Matt., 2006. **18**: p. R513-R583.
46. Linke, W.A., M. Kulke, H. Li, S. Fujita-Becker, C. Neagoe, D.J. Manstein, M. Gautel, and J.M. Fernandez, J. Struct. Biol. **137**: p. 194-205.
47. Fernandez, J.M. and H.B. Li, Science, 2004. **303**(5664): p. 1674-1678.
48. Schlierf, M., H.B. Li, and J.M. Fernandez, Proc. Natl. Acad. Sci USA, 2004. **101**(19): p. 7299-7304.
49. Cecconi, C., E.A. Shank, C. Bustamante, and M. S., Science, 2005. **309**: p. 2057.
50. Huang, L., S. Kirmizialtin, and D.E. Makarov, J. Chem. Phys., 2005. **123**(12).
51. Howorka, S., L. Movileanu, X.F. Lu, M. Magnon, S. Cheley, O. Braha, and H. Bayley, J. Am. Chem. Soc., 2000. **122**(11): p. 2411-2416.

52. Hinnah, S.C., R. Wagner, N. Sveshnikova, R. Harrer, and J. Soll, Biophysical J., 2002. **83**(2): p. 899-911.
53. Movileanu, L., J.P. Schmittschmitt, J.M. Scholtz, and H. Bayley, Biophysical J., 2005. **89**: p. 1030-1045.
54. Goodrich, C.P., S. Kirmizialtin, B.M. Huyghues-Despointes, A.P. Zhu, J.M. Scholtz, D.E. Makarov, and L. Movileanu, J. Phys. Chem. B, 2007. **111**(13): p. 3332-3335.
55. Aksimentiev, A., J.B. Heng, G. Timp, and K. Schulten, Biophys. J., 2004. **87**: p. 2086-2097.
56. Aksimentiev, A. and K. Schulten, Biophys. J., 2005. **88**: p. 3745-3761.
57. Karplus, M., Physica A, 1999. **263**(1-4): p. 389-391.
58. Karplus, M. and A. Sali, Curr. Opin. in Struct. Biol., 1995. **5**(1): p. 58-73.
59. Li, P.C. and D.E. Makarov, J. Chem. Phys., 2003. **119**(17): p. 9260-9268.
60. Li, P.C. and D.E. Makarov, J. Phys. Chem. B, 2004. **108**(2): p. 745-749.
61. Li, P.C. and D.E. Makarov, J. Chem. Phys., 2004. **121**(10): p. 4826-4832.
62. Balsera, M., S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten, Biophys. J., 1997. **73**(3): p. 1281-1287.
63. Isralewitz, B., M. Gao, and K. Schulten, Curr. Opin. in Struct. Biol., 2001. **11**(2): p. 224-230.
64. Johannesson, G.H. and H. Jonsson, J. Chem. Phys., 2001. **115**(21): p. 9644-9656.
65. Makarov, D.E. and H. Metiu, J. Chem. Phys., 1997. **107**(19): p. 7787-7799.
66. Truhlar, D.G., B.C. Garrett, and S.J. Klippenstein, J. Phys. Chem., 1996. **100**(31): p. 12771-12800.
67. Ferrenberg, A.M. and R.H. Swendsen, Phys. Rev. Lett., 1989. **63**(12): p. 1195-1198.
68. Frenkel, D. and B. Smit, *Understanding Molecular Simulation*. 2nd ed. Computational Science Series. 2002, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo: Academic Press. 638.

69. Kumar, S., J.M. Rosenberg, D. Bouzida, R.H. Swendsen, and P.A. Kollman, J. Comp. Chem., 1995. **16**(11): p. 1339-1350.
70. Kirmizialtin, S., V. Ganesan, and D.E. Makarov, J. Chem. Phys., 2004. **121**(20): p. 10268-10277.
71. Jahnig, F., Proc. Natl. Acad. Sci USA, 1983. **80**(12): p. 3691-3695.
72. Eggers, D.K. and J.S. Valentine, Protein Sci., 2001. **10**(2): p. 250-261.
73. Friedel, M., D.J. Sheeler, and J.E. Shea, J. Chem. Phys., 2003. **118**(17): p. 8106-8113.
74. Klimov, D.K., D. Newfield, and D. Thirumalai, Proc. Natl. Acad. Sci USA, 2002. **99**(12): p. 8019-8024.
75. Ping, G., J.M. Yuan, M. Vallieres, H. Dong, Z. Sun, Y. Wei, F.Y. Li, and S.H. Lin, J. Chem. Phys., 2003. **118**(17): p. 8042-8048.
76. Takagi, F., N. Koga, and S. Takada, Proc. Natl. Acad. Sci USA, 2003. **100**(20): p. 11367-11372.
77. Wei, Y., J.G. Xu, Q.W. Feng, H. Dong, and M.D. Lin, Mat. Lett., 2000. **44**(1): p. 6-11.
78. Zhou, H.X. and K.A. Dill, Biochemistry, 2001. **40**(38): p. 11289-11293.
79. Martinez, L.M.C., F.J. Martinez-Veracoechea, P. Pohkarel, A.D. Stroock, F.A. Escobedo, and M.P. DeLisa, Biotech. Bioeng., 2006. **94**(1): p. 105-117.
80. Mingarro, I., I. Nilson, P. Whitley, and G. Heijine, Cell Biology, 2000. **1:3**.
81. Chern, S.S., A.E. Cardenas, and R.D. Coalson, J. Chem. Phys., 2001. **115**(16): p. 7772-7782.
82. DiMarzio, E.A. and A.J. Mandell, J. Chem. Phys., 1997. **107**(14): p. 5510-5514.
83. Pain, R.H., *Mechanisms of Protein Folding.* 2. ed. 2000: Oxford University Press.
84. Honeycutt, J.D. and D. Thirumalai, Biopolymers, 1992. **32**(6): p. 695-709.
85. Guo, Z.Y. and D. Thirumalai, Biopolymers, 1995. **36**(1): p. 83-102.
86. Veitshans, T., D. Klimov, and D. Thirumalai, Folding & Design, 1997. **2**(1): p. 1-22.
87. Paterlini, M.G. and D.M. Ferguson, Chem. Phys., 1998. **236**(1-3): p. 243-252.

88. Hansmann, U.H.E. and Y. Okamoto, J. Comp. Chem., 1997. **18**(7): p. 920-933.
89. Sugita, Y. and Y. Okamoto, Chem. Phys. Lett., 1999. **314**(1-2): p. 141-151.
90. Guo, Z.Y. and C.L. Brooks, Biopolymers, 1997. **42**(7): p. 745-757.
91. Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger Principles of Biochemistry*. 2000, New York Worth.
92. Maier, B., L. Potter, M. So, H.S. Seifert, and M.P. Sheetz, Proc. Natl. Acad. Sci USA, 2002. **99**(25): p. 16012-16017.
93. Best, R.B., D.J. Brockwell, J.L. Toca-Herrera, A.W. Blake, D.A. Smith, S.E. Radford, and J. Clarke, Anal. Chim. Acta, 2003. **479**(1): p. 87-105.
94. Lavery, R., A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette, J. Phys. : Condens. Matt., 2002. **14**: p. R383 - R414.
95. Kirmizialtin, S., L. Huang, and D.E. Makarov, Phys. Stat. Sol. B, 2006. **243**(9): p. 2038-2047.
96. Kenniston, J.A., T.A. Baker, J.M. Fernandez, and R.T. Sauer, Cell, 2003. **114**(4): p. 511-520.
97. Lee, C., M.P. Schwartz, S. Prakash, M. Iwakura, and A. Matouschek, Mol. Cell, 2001. **7**(3): p. 627-637.
98. Matouschek, A. and C. Bustamante, Nat. Struct. Biol., 2003. **10**(9): p. 674-676.
99. Matouschek, A., Curr. Opin. in Struct. Biol., 2003. **13**: p. 98-109
100. Shariff, K., S. Ghosal, and A. Matouschek, Biophysical J., 2004. **86**(6): p. 3647-3652.
101. Brockwell, D.J., E. Paci, R.C. Zinober, G.S. Beddard, P.D. Olmsted, D.A. Smith, R.N. Perham, and S.E. Radford, Nat. Struct. Biol., 2003. **10**(9): p. 731-737.
102. Carrion-Vazquez, M., H.B. Li, H. Lu, P.E. Marszalek, A.F. Oberhauser, and J.M. Fernandez, Nat. Struct. Biol., 2003. **10**(9): p. 738-743.
103. Brown, S. and T. Head-Gordon, Protein Sci., 2004. **13**(4): p. 958-970.
104. Sorensen, J.M. and T. Head-Gordon, Proteins: Struct. Funct. Gen., 2002. **46**(4): p. 368-379.
105. Sorenson, J.M. and T. Head-Gordon, J. Comp. Biol., 2000. **7**(3-4): p. 469-481.

106. Sorenson, J.M. and T. Head-Gordon, J. Comp. Biol., 2002. **9**(1): p. 35-54.
107. Izrailev, S., S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, Biophysical J., 1997. **72**(4): p. 1568-1581.
108. Makarov, D.E., P.K. Hansma, and H. Metiu, J. Chem. Phys., 2001. **114**(21): p. 9663-9673.
109. Kirmizialtin, S., L. Huang, and D.E. Makarov, J. Chem. Phys., 2005. **122**(23).
110. Bell, G.I., Science, 1978. **200**(4342): p. 618-627.
111. Carrion-Vazquez, M., A.F. Oberhauser, S.B. Fowler, P.E. Marszalek, S.E. Broedel, J. Clarke, and J.M. Fernandez, Proc. Natl. Acad. Sci USA, 1999. **96**(7): p. 3694-3699.
112. De Gennes, P.G., *Scaling concepts in polymer physics*. 1979, Ithaca, N.Y.: Cornell University Press.
113. Lubensky, D.K. and D.R. Nelson, Biophysical J., 1999. **77**(4): p. 1824-1838.
114. Muthukumar, M., Phys. Rev. Lett., 2001. **86**(14): p. 3188-3191.
115. Muthukumar, M., Electrophoresis, 2002. **23**(10): p. 1417-1420.
116. Muthukumar, M., J. Chem. Phys., 2003. **118**(11): p. 5174-5184.
117. Sung, W. and P.J. Park, Phys. Rev. Lett., 1996. **77**(4): p. 783-786.
118. Kasianovicz, J.J., E. Brandin, D. Branton, and D.W. Deamer, Proc. Natl. Acad. Sci USA, 1996. **93**(24): p. 13770-13773.
119. Kasianowicz, J.J., Disease Markers, 2002. **18**(4): p. 185-191.
120. Meller, A., L. Nivon, and D. Branton, Phys. Rev. Lett., 2001. **86**(15): p. 3435-3438.
121. Movileanu, L. and H. Bayley, Proc. Natl. Acad. Sci USA, 2001. **98**(18): p. 10137-10141.
122. Movileanu, L., S. Cheley, and H. Bayley, Biophysical J., 2003. **85**(2): p. 897-910.
123. Movileanu, L., S. Howorka, O. Braha, and H. Bayley, Nat. Biotech., 2000. **18**(10): p. 1091-1095.
124. Song, L.Z., M.R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J.E. Gouaux, Science, 1996. **274**(5294): p. 1859-1866.

125. Gallagher, T., P. Alexander, P. Bryan, and G.L. Gilliland, Biochemistry, 1994. **33**(15): p. 4721-4729.
126. Kong, C.Y. and M. Muthukumar, Electrophoresis, 2002. **23**(16): p. 2697-2703.
127. Klimov, D.K. and D. Thirumalai, Phys. Rev. Lett., 1997. **79**(2): p. 317-320.
128. Baysal, C., B. Erman, and B. Chu, J. Chem. Phys., 2001. **114**: p. 5444-5449.
129. Hanggi, P., P. Talkner, and M. Borkovec, Rev. Mod. Phys., 1990. **62**(2): p. 251-341.
130. Berezhkovskii, A.M., M.A. Pustovoit, and S.M. Bezrukov, J. Chem. Phys., 2002. **116**(22): p. 9952-9956.
131. Stefureac, R., Y.T. Long, H.B. Kraatz, P. Howard, and J.S. Lee, Biochemistry, 2006. **45**(30): p. 9172-9179.
132. Best, R.B., B. Li, A. Steward, V. Daggett, and J. Clarke, Biophys. J., 2001. **81**(4): p. 2344-2356.
133. Brockwell, D.J., G.S. Beddard, J. Clarkson, R.C. Zinober, A.W. Blake, J. Trinick, P.D. Olmsted, D.A. Smith, and S.E. Radford, Biophys. J., 2002. **83**(1): p. 458-472.
134. Fisher, T.E., P.E. Marszalek, and J.M. Fernandez, Nat. Struct. Biol., 2000. **7**(9): p. 719-724.
135. Fisher, T.E., A.F. Oberhauser, M. Carrion-Vazquez, P.E. Marszalek, and J.M. Fernandez, TIBS, 1999. **24**(10): p. 379-384.
136. Hansma, H.G. and L. Pietrasanta, Curr. Opin. in Struct. Biol., 1998. **2**(5): p. 579-584.
137. Hansma, H.G., L.I. Pietrasanta, I.D. Auerbach, C. Sorenson, R. Golan, and P.A. Holden, J. Biomat. Sci.-Polymer Ed., 2000. **11**(7): p. 675-683.
138. Li, H.B., M. Carrion-Vazquez, A.F. Oberhauser, P.E. Marszalek, and J.M. Fernandez, Nat. Struct. Biol., 2000. **7**(12): p. 1117-1120.
139. Oberhauser, A.F., C. Badilla-Fernandez, M. Carrion-Vazquez, and J.M. Fernandez, J. Mol. Biol., 2002. **319**(2): p. 433-447.

140. Oberhauser, A.F., P.K. Hansma, M. Carrion-Vazquez, and J.M. Fernandez, Proc. Natl. Acad. Sci USA, 2001. **98**(2): p. 468-472.
141. Schwaiger, I., C. Sattler, D.R. Hostetter, and M. Rief, Nat. Mat., 2002. **1**(4): p. 232-235.
142. Thompson, J.B., H.G. Hansma, P.K. Hansma, and K.W. Plaxco, J. Mol. Biol., 2002. **322**(3): p. 645-652.
143. Tskhovrebova, L., J. Trinick, J.A. Sleep, and R.M. Simmons, Nature, 1997. **387**(6630): p. 308-312.
144. Viani, M.B., T.E. Schaffer, A. Chand, M. Rief, H.E. Gaub, and P.K. Hansma, J. Appl. Phys., 1999. **86**(4): p. 2258-2262.
145. Williams, P.M., S.B. Fowler, R.B. Best, J.L. Toca-Herrera, K.A. Scott, A. Steward, and J. Clarke, Nature, 2003. **422**(6930): p. 446-449.
146. Zinober, R.C., D.J. Brockwell, G.S. Beddard, A.W. Blake, P.D. Olmsted, S.E. Radford, and D.A. Smith, Protein Science, 2002. **11**(12): p. 2759-2765.
147. Smith, B.L., T.E. Schaffer, M. Viani, J.B. Thompson, N.A. Frederick, J. Kindt, A. Belcher, G.D. Stucky, D.E. Morse, and P.K. Hansma, Nature, 1999. **399**(6738): p. 761-763.
148. Zhuang, X.W. and M. Rief, Curr. Opin. in Struct. Biol., 2003. **13**(1): p. 88-97.
149. Bouchiat, C., M.D. Wang, J.F. Allemand, T. Strick, S.M. Block, and V. Croquette, Biophys. J., 1999. **76**(1): p. 409-413.
150. Fixman, M. and J. Kovac, J. Chem. Phys., 1973. **58**: p. 1564.
151. Kratky, O. and G. Porod, Recl. Trav. Chim. Pays-Bas, 1949. **68**: p. 1106.
152. Dudko, O.K., J. Mathe, A. Szabo, A. Meller, and G. Hummer, Biophys. Chem., 2007. **92**: p. 4188-4195.
153. Hummer, G. and A. Szabo, Biophysical J., 2003. **85**(1): p. 5-15.
154. Des Cloizeaux, J. and G. Jannink, *Polymers in Solution: Their Modelling and Structure*. 1990, Oxford: Clarendon Press.
155. Flory, P.J., *Principles of Polymer Chemistry*. 1953, Itacha: Cornell University Press.

156. Makarov, D.E., C.A. Keller, K.W. Plaxco, and H. Metiu, Proc. Natl. Acad. Sci USA, 2002. **99**(6): p. 3535-3539.
157. Millet, I.S., S. Doniach, and K.W. Plaxco, Adv. Protein. Chem., 2002. **62**: p. 241-262.
158. Liphardt, J., B. Onoa, S.B. Smith, I. Tinoco, and C. Bustamante, Science, 2001. **292**(5517): p. 733-737.
159. Eom, K., P.C. Li, D.E. Makarov, and G.J. Rodin, J. Phys. Chem. B, 2003. **107**(34): p. 8730-8733.
160. Eom, K., D.E. Makarov, and G.J. Rodin, Phys. Rev. E, 2005. **71**: p. 021904.
161. Klimov, D.K. and D. Thirumalai, Proc. Natl. Acad. Sci USA, 2000. **97**(13): p. 7254-7259.
162. Haupt, B.J., T.J. Senden, and E.M. Sevick, Langmuir, 2002. **18**(6): p. 2174-2182.
163. Lee, N.K., A. Johner, and T.A. Vilgis, Macromolecules, 2002. **35**(15): p. 6043-6054.
164. Marenduzzo, D., A. Maritan, A. Rosa, and F. Seno, Phys. Rev. Lett., 2003. **90**(8).
165. Halperin, A. and E.B. Zhulina, Europhys. Lett., 1991. **15**(4): p. 417-421.
166. Geissler, P.L. and E.I. Shakhnovich, Macromolecules, 2002. **35**(11): p. 4429-4436.
167. Cieplak, M., T.X. Hoang, and M.O. Robbins, Proteins: Struct., Func., Gen. , 2002. **49**: p. 114-121
168. Cieplak, M., T.X. Hoang, and M.O. Robbins, Proteins: Struct., Func., Gen. , 2004. **56**: p. 285-297.
169. Cieplak, M. and P.E. Marsalek, J. Chem. Phys., 2005. **123**: p. 194903
170. Irback, A., S. Mitternacht, and S. Mohanty, Proc. Natl. Acad. Sci USA 2005. **102**: p. 13427-13432.
171. Lacks, D.J., Biophysical J., 2005. **88**: p. 3493-3501.
172. Li, M.S., M. Kouza, and C.K. Hu, Biophysical J., 2007. **92**(2): p. 547-561.
173. West, D.K., D.J. Brockwell, P.D. Olmsted, S.E. Radford, and E. Paci, Biophysical J., 2006. **90**: p. 287-297.

174. Cao, Y., C. Lam, M.D. Wang, and H. Li, Angew. Chem., Int. Ed. Engl., 2006. **45**: p. 642-645.
175. Makarov, D.E. and K.W. Plaxco, Protein Sci. , 2003. **12**: p. 17-26.
176. Plaxco, K.W., K.T. Simons, and D. Baker, J. Mol. Biol. , 1998. **277**: p. 985.
177. Plaxco, K.W., K.T. Simons, I. Ruczinski, and B. David, Biochemistry, 2000. **39**: p. 11177.
178. Fantner, G.E., E. Oroudjev, G. Schitter, I.S. Golde, P. Thumer, M.M. Finch, P. Turner, T. Gutsmann, D.E. Morse, H. Hansma, and P.K. Hansma, Biophysical J. , 2006. **90**: p. 1411-1418.
179. Bates, A.D. and A. Maxwell, *DNA Topology*. 2005, Oxford: Oxford University Press.
180. De Gennes, P.G., Macromolecules, 1984. **17**: p. 703-704.
181. Frank-Kamenetskii, M.D., *Unraveling DNA*. 1997, Massachusetts: Perseus Books, Reading.
182. Grosberg, A.Y., Phys. Rev. Lett., 2000. **85**: p. 003858.
183. Kauffman, L.H., *Knots and physics*. 2001, New Jersey, London, Hong Kong: World Scientific.
184. Pieranski, P., S. Przybyl, and A. Stasiak, Eur. Phys. J. E, 2001. **6**: p. 123-128.
185. Saitta, A.M. and M. Klein, J. Phys. Chem., 2001. **28**: p. 6495-6499.
186. Lobovkina, T., P. Dommersnes, J.-F. Joanny, P. Bassereau, M. Karlson, and O. Owar, Proc. Natl. Acad. Sci USA, 2004. **101**: p. 7949-7953.
187. Taylor, W.R., Nature, 2000. **406**: p. 916-919.
188. Vologodskii, A., Biophysical J., 2006. **90**: p. 1594-1597.
189. Bockelman, U., Curr. Opin. in Struct. Biol., 2004. **14**: p. 368-373.
190. Bustamante, C., S.B. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struct. Biol., 2000. **10**: p. 279.
191. Bustamante, C., V. Chemla, N.R. Forde, and D. Izhaky, Ann. Rev. Biochem., 2004. **73** p. 705-748.

192. Arai, Y., R. Yasuda, K.-i. Akashi, Y. Harada, H. Miyata, K. Kinosita, and H. Itoh, Nature, 1999. **399**: p. 446.
193. Bao, X.R., H.J. Lee, and S.R. Quake, Phys. Rev. Lett., 2003. **91**: p. 265506-1.
194. Maddocks, J.H. and J.B. Keller, SIAM J. Appl. Math, 1987. **47**: p. 1185.
195. Persson, B.N.J., *Sliding Friction. Physical Principles and Applications*. 1998, Berlin Heidelberg Springer-Verlag.
196. Barsegov, V. and D. Thirumalai, Proc. Natl. Acad. Sci USA 2005. **102**: p. 1835-1839.
197. Pereverzev, Y.V., O.V. Prezhdo, M. Forero, E.V. Sokurenko, and W.E. Thomas, Biophysical J., 2005. **89**: p. 1446-1454.
198. DeLano, W.L. 2002, DeLano Scientific: San Carlos, CA,.
199. Reinmann, P., C. Van den Vroeck, H. Linke, P. Hanggi, J.M. Rubi, and A. Perez-Madrid, Phys. Rev. E., 2002. **65**: p. 031104.

VITA

Serdal Kirmizialtin was born in Van, Turkey, on April 3, 1975. After graduating from the high school, Nisantasi Anadolu Lisesi, in Istanbul, he entered Koc University in 1995. He completed his BS in the Chemistry Department of Koc University in 2000 and got his MS from the Material Science and Engineering Department of Sabanci University in 2002. In his MS, he studied computer simulations of conformational and dynamical properties of high performance polymers.

In 2002 he started his PhD in the University of Texas at Austin, Department of Chemistry and Biochemistry. During his PhD work he studied various problems on the biomolecular processes that involve mechanical forces such as protein translocation, mechanical stretching of proteins and dynamics of molecular friction knots. This work is compiled under the thesis entitled: “Computer Simulations of Protein Translocation and Stretching”.

Permanent address: Gumuspala Mah. Bekar Sokak No: 26, D:5, K.Cekmece, Istanbul, TURKEY

This dissertation was typed by Serdal Kirmizialtin