**The Dissertation Committee for Pai-Chi Li Certifies that this is the approved version of the following dissertation:**

# Molecular Dynamics Simulations of the Mechanical Unfolding of Proteins

**Committee:**

Dmitrii E. Makarov, Supervisor

Peter J. Rossky

Jason B. Shear

Thomas M. Truskett

Robert E. Wyatt

# Molecular Dynamics Simulations of the Mechanical Unfolding of Proteins

by

**Pai-Chi Li, 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, 2006**

# Molecular Dynamics Simulations of the Mechanical Unfolding of Proteins

Publication No._____________

Pai-Chi Li, Ph.D
The University of Texas at Austin, 2006

Supervisor: Dmitrii E, Makarov

A number of proteins perform load-bearing functions in living organisms and often have unique mechanical properties. In recent years, there has been considerable effort to understand the relationships between the molecular structure of such proteins and their mechanical response. Several of them have been studied in great detail through single molecule mechanical pulling experiments. Interpretation of these experiments requires the use of atomistic simulations. However typical simulation time scales are many orders of magnitude shorter than relevant experimental and/or physiological time scales. In this dissertation, we have developed a simulation methodology that provides a direct link between experiments and simulations and is capable of predicting the outcome of single molecule pulling experiments. By using this methodology, we have been able to understand the relationships between the molecular structure and the mechanical properties of a number of proteins. I report on our studies of the mechanical unfolding of the I27 domain of the muscle protein titin, ubiquitin, and protein G and compare them with the existing experimental data. The distribution of the unfolding force as well as its dependence on the pulling rate predicted by our simulations is found to be in good agreement with AFM experiments. We demonstrate that the mechanical unfolding pathway can be altered by changing the pulling geometry and that the presence of a hydrogen bonded clamp between terminal parallel strands of these domains is the key property that is responsible for their high mechanical stability. We have also extended our studies of single protein domains to protein dimers. Our replica-exchange molecular dynamics simulation study of the mechanical unfolding of a segment-swapped protein G dimer suggests that the mechanical resistance of a protein complex may be controlled not only by the mechanical stability of individual domains but also by the inter-chain interactions between domains.

# Table of Contents

Chapter I. Introduction ........................................................................... 1

Chapter II. Mechanical unfolding of the muscle protein titin ................................. 7
    II.1 Introduction ........................................................................... 7
    II.2 Steered molecular dynamic simulation of I27 stretching ..................... 10
    II.3 The potential of mean force ................................................ 14
    II.4 Langevin equation model: comparison with MD simulations .............. 21
    II.5 Comparison with experimental data ..................................... 22
    II.6 Discussion ........................................................................... 27

Chapter III. Mechanical unfolding of segment ubiquitin-like protein domains ... 30
    III.1 Introduction ........................................................................... 30
    III.2 Methods ........................................................................... 32
    III.3 Results ........................................................................... 34
    III.4 Discussion ........................................................................... 38

Chapter IV. Mechanical unfolding of ubiquitin ................................................ 41
    IV.1 Introduction ........................................................................... 41
    IV.2 Methods ........................................................................... 45
    IV.3 Results ........................................................................... 46
    IV.4 Comparison with experimental data ..................................... 51
    IV.5 Discussion ........................................................................... 55

Chapter V. Mechanical unfolding of segment-swapped protein G dimer ............ 57
    V.1 Introduction ........................................................................... 57
    V.2 Methods ........................................................................... 60
    V.3 Results ........................................................................... 62
    V.4 Discussion and conclusions ................................................ 69

Chapter VI. Summary ....................................................................................72

References..............................................................................................................74

Vita ......................................................................................................................82

# Chapter I

# Introduction

In every living organism, there are proteins that perform "load-bearing" functions. Examples include the giant muscle protein titin[78,79,84,99,100], tenascin[90], fibronectin[88,92], spectrin[66,67,101], barnase[10], nacre protein perlucin[108], and spider silk[5,47,48,51,91,118]. These proteins are often unique materials that exhibit a combination of high mechanical strength and high elasticity. For instance, spider capture silk has tensile strength comparable to steel, but it is extremely elastic and can be stretched to nearly 5-10 times its original length[5,47,48,51,91,118]. In the pass ten years, there has been considerable effort to understand the relationships between the molecular structure of such proteins and their mechanical response. It has been understood that the mechanical unfolding of certain proteins is a key mechanisms that that accounts for the high toughness of natural materials[108]. Moreover, studies of the mechanical unfolding of proteins provide an opportunity to probe their energy landscapes. By using an understanding of the relationship between the structural and the mechanical properties of proteins, one can use certain proteins as building blocks for the construction of new nanomechanical materials or design biologically inspired polymeric materials. This dissertation reports on our studies of the mechanical unfolding of proteins by using molecular dynamic simulations.

Mechanical properties of individual protein molecules can be explored by performing single-molecule pulling experiments[5,9-11,27,32,33,42,43,59,65,69,70,84,88-90,94,99-101,108,111,112,114,116,117,124,125], in which proteins are stretched by mechanical forces and their mechanical resistance is measured. Atomic force microscopy (AFM) is one such single-molecule stretching technique and is illustrated in Figure I.1(a). In a typical AFM experiment, one end of a polyprotein chain, which is composed of multiple protein domains, is attached to a cantilever. The other end of the polyprotein chain is attached to a surface. The surface is moved away from the cantilever with constant velocity *v.* Then the force response of the protein, which is equal to the force response of the cantilever, is measured by

$$f = k_c(L - r) = k_c x \,. \qquad \text{(I.1)}$$

Here $k_c$ is the force constant of the cantilever, $L$ is the distance between the cantilever and the surface, $r$ is the end-to-end distance for the polyprotein chain, and $x$ is the displacement of the cantilever. Figure I.1(a) shows a typical result of an AFM experiment[100]. The x-axis is the elongation of the polyprotein chain which is equal to $vt$. Here $v$ is the stretching velocity, and $t$ is the stretching time. The y-axis is the force response of the proteins calculated by eq 1. The force-elongation curve often exhibits a saw-tooth pattern. Every time there is a drop of the force it reflects the unfolding event of an individual protein domain in the chain. The force can be measured at different pulling speeds and the average force is typically reported at a given pulling speed. The result of an AFM experiment for a 500 nm-long titin segment is shown in the Figure I.1(b)[100]. The inset at upper left is the resulting force extension curve. It shows nine peaks averaging 190 pN and 130 pN when pulled at a speed of 0.5 μm/s and 0.01 μm/s, respectively. The mean unfolding force often exhibits a logarithmic dependence on the pulling speed.

Molecular level understanding of these single-molecular pulling experiments requires the use of atomistic simulations. Steered molecular dynamics (SMD) is one computational method commonly used to study the mechanical response of proteins. The method of SMD is briefly described below. One end of the polypeptide chain is attached to a spring, which has a force constant $k_c$. Similar to the AFM experiment, the mechanical response of the protein can be measured by moving the spring. The force response of the protein can be calculated as the spring constant times the spring extension,

$$f = k_c(R_0 - R) \,. \qquad \text{(I.2)}$$

Here $R_0$ is the total extension, i.e., the distance between the two ends of the chain plus the extension of the spring, and $R$ is the simulated extension of the protein itself. The total extension is usually increased at a linear rate:

$$R_0 = R_{fold} + vt \,, \qquad \text{(I.3)}$$

where $R_{fold}$ is the initial extension in the folded state, $v$ is the stretching velocity, and $t$ is the stretching time. The outcome of a single SMD run is a force-extension curve f(R) of

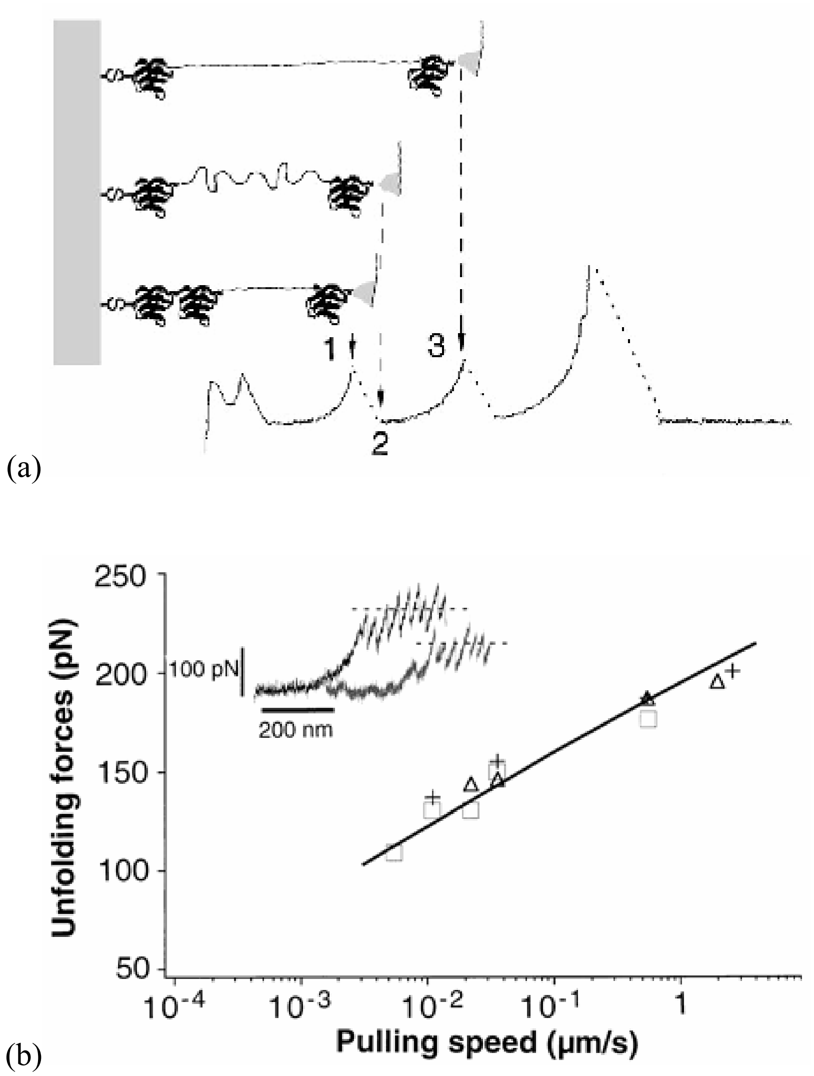


Figure I.1: (a) Mechanical stretching of a polyprotein chain using the Atomic force microscope (AFM) and a typical force-extension curve observed in a stretching experiment. (b) The mean domain unfolding forces exhibit a logarithmic dependence on the pulling speeds.[100]

the protein for a given stretching velocity, as calculated from eqs. I.2 and I.3. In Figure I.2 SMD force-vs.-extension curve for the I27 domain in titin shows a single force peak similar to the sawtooth-shaped force profile in AFM experiments. The force required to unfold a protein is defined as the peak of the force curve when the force drops. This atomic level simulation method provides access to the internal mechanical properties of proteins, such as the molecular mechanism of protein unfolding, which is often difficult to directly observe experimentally. However, typical simulation time scales in SMD are about six orders of magnitude shorter than relevant experimental and/or physiological time scales. Consequently, it is not feasible to perform a direct molecular dynamics simulation of a single molecule mechanical unfolding experiment at the experimental pulling speed. In addition, it is impossible to extrapolate the unfolding forces extracted from SMD simulations to the experimental region, since the mechanical response of the molecule calculated in SMD simulations is dominated by dissipative, friction forces[49,2,55,56,28] for pulling speeds as high as 0.1 – 10 Å/ps. Such a force that is assumed proportional to the velocity $u$ would be negligible in the experimental studies, where $u$ is some six orders of magnitude smaller.

The main goal of the research presented in this thesis is to develop simulation methodology that provides a direct link between experiments and simulations and is capable of predicting the outcome of single molecule pulling experiments and, by using this methodology, to understand the relationships between the molecular structure and the mechanical properties of a number of proteins.

Our general approach is to model the unfolding dynamics of a protein domain as diffusive motion along a single unfolding coordinate $R$ (equal to the domain extension) in the presence of an external driving potential and the equilibrium potential of mean force $G(R)$. The computation of $G(R)$ is by itself a challenging task, which is accomplished by using a combination of several methods that are necessary to improve sampling and to extract maximum information from molecular dynamics trajectories.

Once G(R) is known, we use transition state theory to compute force-dependent rates for the mechanical unfolding of domains. These rates are subsequently used to perform

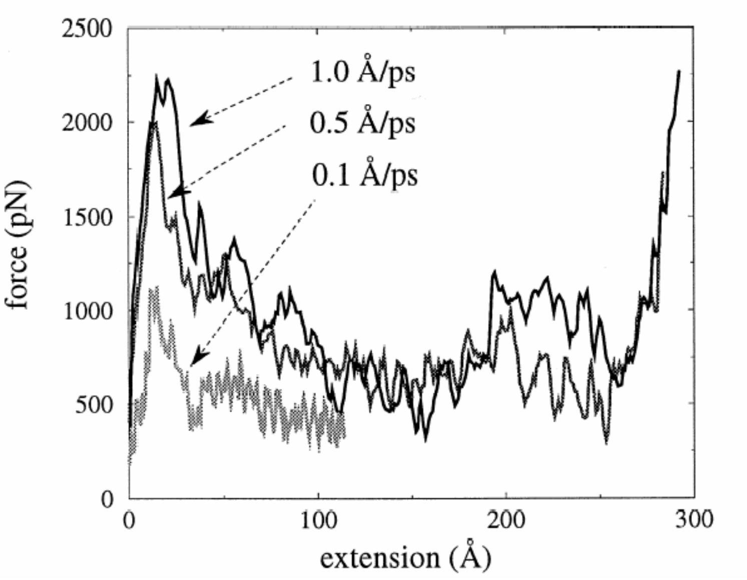


Figure I.2: The force-extension profile from the Steered Molecular dynamics (SMD) trajectory for an I27 domain in titin.[79]

kinetic Monte Carlo simulations and further to predict the outcome of single molecule AFM pulling experiments.

Some of the theoretical predictions made by using this approach have already been confirmed by subsequent experimental studies.[12,19]

The layout of this dissertation is as follows:

Chapter II: Mechanical unfolding of the muscle protein titin”. In this chapter, we describe the procedure we have developed to predict the outcomes of the AFM pulling experiments based on molecular dynamics simulations, and apply this procedure to the mechanical unfolding of a polyprotein chain which is composed of multiple I27 domains.

Chapter III: “Mechanical unfolding of segment ubiquitin-like protein domains.” It has been suggested in our paper [K. Eom, P.-C. Li, D.E. Makarov, and G.J. Rodin, J. Phys. Chem. B, **107** (2003) 8730] that that the “clamp” formed by the parallel strands in this domain represents an optimal topology maximizing the mechanical strength of cross-linked polymer chains. The goal of the work described in this chapter was to verify this hypothesis by studying the mechanical stability of non-mechanical proteins that have the same structural motif. To this end, we have studied the mechanical response of two small globular proteins, ubiquitin and protein G IgG-binding domain III, which do not have an obvious mechanical function, and demonstrated that they nevertheless exhibit considerable mechanical stability.

Chapter IV: "Mechanical unfolding of ubiquitin." Biological functions of polyubiquitin chains are related to different linkages between individual ubiquitin domains. Some of these functions are believed to be related to the different mechanical resistance of the ubiquitin domains with respect to forces that are applied differently for different linkages. Motivated by the recent experimental study of the linkage dependence of the mechanical response of polyubiquitin chains[D. J. Brockwell et al, *Nature Structural Biology* **10,** 7312003); M. Carrion-Vazquez et al, *Nature Structural Biology* **10,** 738 (2003)], we have used simulations to study the mechanical unfolding of ubiquitin and demonstrated that the mechanical unfolding pathway indeed can be altered by changing the pulling geometry. This study provides an excellent opportunity to probe different mechanical unfolding reaction pathways on the multidimensional free energy landscapes of proteins. Some of these pathways may potentially be relevant for chemical or thermal unfolding mechanisms for the same protein.

Chapter V: "Mechanical unfolding of segment-swapped protein G dimer." This chapter reports on our replica-exchange molecular dynamics simulation study of the mechanical unfolding of segment-swapped protein G dimer. This study provides insight into the mechanical properties of bulk materials, which usually involve complex assemblies of proteins, controlled not only by the mechanical stability of individual domains but also by the inter-chain interactions.

Chapter VI contains the summary of the main results of this dissertation.

The materials of Chapter II-IV have been published[71,72,73], and that of Chapter V has been accepted to be published[74].

# Chapter II

## Mechanical unfolding of the muscle protein titin[a]

### II. 1 INTRODUCTION

Single molecule experiments, in which proteins are stretched by mechanical forces, reveal a wealth of information about the mechanical properties of structural proteins[5,9-11,27,32,33,42,43,59,65,69,70,84,88-90,94,99-101,108,111,112,114,116,117,124,125], many of which are unique materials. Atomistic-level computer simulations[10,15,49,50,56,62,68,78,79,84,92,103,107] performed in conjunction with experiments often reveal the mechanisms of proteins' response to a tensile loading. Some of the proteins well characterized both experimentally and computationally include titin[78,79,84,99,100], fibronectin[88,92], and barnase[10].

Molecular dynamics (MD) simulations have provided valuable insights into the mechanical response of the I27 domain of the muscle protein titin. A combination of experimental[84,99,100,116,117] and computational[78,79,84] studies have revealed the structural changes experienced by the I27 domain as it unfolds when the ends of the polypeptide chain are pulled apart.

A direct comparison of experiments with simulations is however often impossible because of the disparity between their time scales. At a much faster rate of loading typical of MD simulations, proteins may unfold via mechanisms different from those explored in experimental studies[81].

In experiments, the mean domain unfolding force usually exhibits a weak logarithmic dependence on the pulling speed $u$. This dependence is generally expected when unfolding is driven by thermally activated barrier crossing[29]. Simulated unfolding forces are much higher and exhibit a much stronger dependence on $u$[78,79]. An attempt to extrapolate the unfolding forces extracted from MD simulations to the experimental pulling speeds leads to a nonsensical result, as shown in Fig. II.1. One reason why this happens is the fact that for pulling speeds as high as 0.1 – 10 Å/ps the mechanical

---

[a] Large portions of this chapter have been previously published as reference 71.

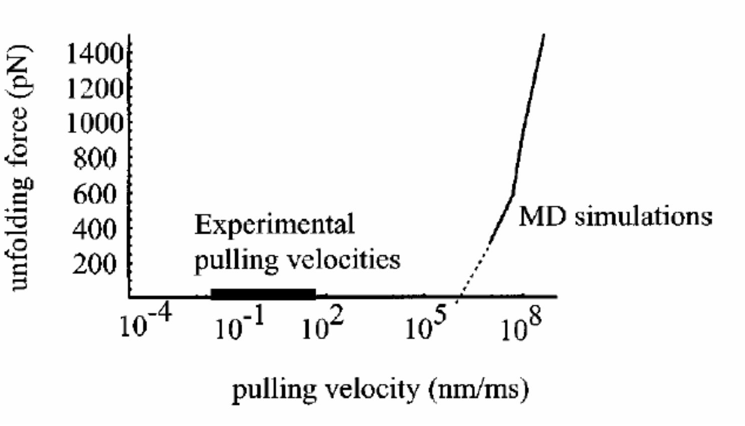


Figure II.1: The dependence of average unfolding force for the I27 domain of the muscle protein titin on the pulling velocity, as obtained from our steered MD simulations described in Section II.2 (solid line). An attempt to extrapolate these data to lower pulling velocities leads to an absurd result (dashed line).

response of the molecule is dominated by dissipative, friction forces[2,28,49,55,56]. Even if the simulation is performed in vacuum, the energy pumped into a single degree of freedom associated with the molecule's extension is dissipated into other degrees of freedom of the molecule itself, resulting in a Stokes-type force. Assuming that such a force is proportional to the velocity $u$, it would be negligible in the experimental studies, where $u$ is some six orders of magnitude smaller.

Several approaches have been developed recently, whose goal is to extrapolate single-molecule stretching data to lower pulling speeds. Jarzynski[57,58] and subsequently Hummer and Szabo[53] showed that equilibrium free energy dependences can in principle be reconstructed from single-molecule experiments or simulations even when those are performed far from equilibrium. This approach has been tested experimentally[76] and by calculations[102]. Since this method involves averaging over multiple single-molecule trajectories, it is computationally expensive.

In another approach[2,28,55,56], one explicitly subtracts the dissipative force to obtain the equilibrium potential of mean force from non-equilibrium steered molecular dynamics simulations by using the assumption that single-molecule trajectories are described by a Langevin equation. Instead of averaging over multiple trajectories, one smoothes single-molecule trajectories by using the moving average or another smoothing technique and then solves the inverse problem of reconstructing the Langevin equation from the simulated trajectories $<R>(t)$ that describe the dynamics of the expectation value of the molecule's extension $R$ as a function of time $t$. In practice, the statistical errors that result from approximating the actual $<R>(t)$ by a smoothed single-molecule trajectory are often large. Although these errors can be reduced by carefully choosing the smoothing procedure and simulation parameters (such as the pulling speed),[2] the optimal choice of the latter often conflicts the requirement of computational feasibility.

In this chapter, rather than extracting the molecule's potential of mean force $G(R)$ from non-equilibrium steered MD trajectories, we use an umbrella-sampling-type approach[87] to calculate $G(R)$ from a series of equilibrium molecular dynamics trajectories. We further show that with $G(R)$ calculated in this way, steered MD trajectories are indeed well described by a Langevin equation as suggested in[2,55,56], thus validating our procedure for the calculation of $G(R)$. With the computed $G(R)$, we then proceed to predict the outcome of single molecule AFM experiments and compare our $G(R)$ with that inferred from the experimental studies. We find that, while our simulation predicts unfolding force distributions and their dependence of the pulling rate that are consistent with the experimental results, the position of the transition state associated with the maximum of $G(R)$ is different from the value estimated from the experimental studies of titin[69]. This difference can be traced back to the assumption that the unfolding free energy barrier depends linearly on the force, which is commonly used in interpretation of experimental data and is inconsistent with the present study. However limitations of the reported simulations may also contribute to this discrepancy.

The chapter is organized as follows: In Section II.2 we report on steered Molecular Dynamics (SMD) simulations of the stretching of the I27 domain and describe an attempt to approximate our simulation results by a Langevin equation and to reconstruct the parameters of the latter directly from SMD trajectories. In Section II.3 we describe our procedure to compute the potential of mean force for the stretched I27 domain. In Section II.4 we return to the Langevin equation model and show that, using the potential of mean force computed in Section II.3, one reproduces the SMD simulation results thus validating both the use of the Langevin equation and the computed potential of mean force. In Section II.5 we use transition state theory to calculate the force-dependent unfolding rate for the potential of mean force of Section II.3. We then perform kinetic Monte Carlo simulations to compute the mean unfolding force and the distribution of the unfolding force as a function of the pulling velocity and compare our results with experimental data. In Section II.6 we compare the computed force dependence of the unfolding rate with that deduced from AFM experiments. We show that, while the two dependences are close to one another in the experimental range of forces, the computed value of the unfolding rate extrapolated to zero force is much smaller than the one previously estimated from experimental data (and also smaller than the unfolding rate measured in chemical denaturation experiments). This discrepancy is due to the assumption that the unfolding barrier depends linearly on the force, which is inconsistent with the results of the present study.

## II. 2 STEERED MOLECULAR DYNAMIC SIMULATION OF I27 STRETCHING

Our molecular dynamics simulations of the stretching of the I27 domain (pdb code 1TIT) were performed with Tinker molecular dynamics software[126] using the GB/SA continuum model for solvation[98] and the Charm27 force field[80]. To "stretch" the molecule, one adds a penalty term

$$(1/2)\ k_c\ (R-R_0(t))^2 \qquad \text{(II.1)}$$

to the molecule's energy[10,49,50,78,79]. Here $R$ is the distance between the outermost $\alpha$-carbon atoms of the protein molecule and is the measure of the molecule's extension in our studies. The penalty energy (1) results in a harmonic restraint that plays a role similar to that of a cantilever in AFM protein pulling experiments[42,43,99,100,117]. The force constant $k_c$ used in our calculations was chosen to be $k_c = 1.38$ N/m. The time dependence of $R_0(t)$

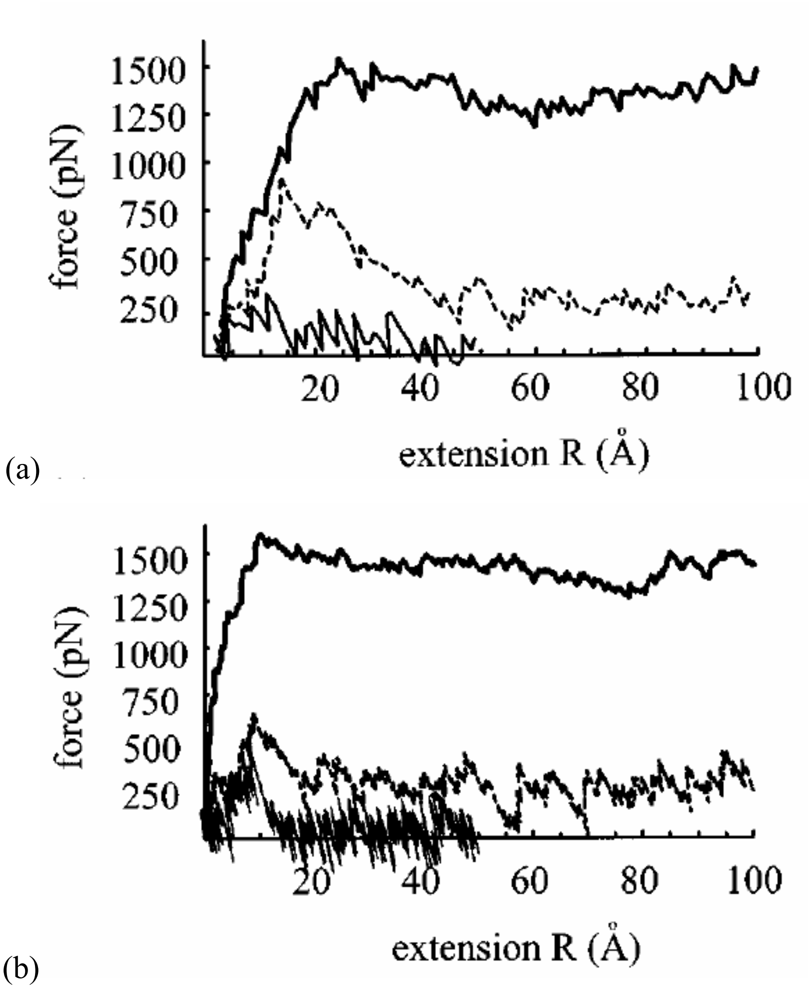


Figure II.2: (a) Force-vs.-extension curves of the I27 domain observed in steered MD simulations for different pulling velocities $u$. Heavy solid line: $u$ = 5 Å/ps. Dashed line: $u$ = 1 Å/ps. Thin solid line: $u$ = 0.1 Å/ps. (b) Force-vs.-extension curves for the same pulling velocities computed by using the Langevin equation (II.4) with the potential of mean force calculated in Section II.3.

describes the external driving and is chosen to be linear:

$$R_0(t) = ut \tag{II.2}$$

where $u$ is the pulling velocity. In practice, $R_0(t)$ was incremented by $\Delta R$=1Å every $\Delta t$= $\Delta R / u$ picoseconds. The instantaneous stretching force is given by:

$$f = k_c\,(R_0(t)-R) \tag{II.3}$$

Prior to the stretching simulation, the minimum energy structure of I27 was determined via steepest descent minimization and it was equilibrated by running 100 ps of unconstrained molecular dynamics simulation. Force-extension curves $f(R)$ obtained in this way for different values of the pulling velocity $u$ are shown in Fig. II.2(a) As the pulling velocity is increased, the force tends to increase. For low pulling speeds the dependence $f(R)$ shows a pronounced peak followed by a drop. This behavior is similar to that observed in experimental force-extension curves of titin[99,100] and is due to the unfolding of the I27 domain[78,79,99,100]. However when the pulling velocity $u$ becomes larger, the unfolding peak disappears. This, together with the fact that the observed forces in the limit of high pulling velocities tend to be proportional to $u$, suggests that the domain's response to the pulling is dominated by a dissipative Stokes-type force that is much larger than equilibrium stretching forces.

Following previous work[2,28,55,56], we attempted to model the observed single-molecule trajectories in terms of a Langevin equation:

$$m\ddot{R} = -\eta\dot{R} - G'(R) + k_c(ut - R) + f(t) \tag{II.4}$$

where η is a friction coefficient and $f(t)$ is a random force obeying the fluctuation-dissipation theorem:

$$\langle f(t)f(t')\rangle = 2\eta k_B T\delta(t-t') \tag{II.5}$$

We further assume that the motion is overdamped such that the intertial term in the Langevin equation can be neglected (i.e., the lhs of Eq. II.4 is set to zero).

In principle, both $G(R)$ and the friction coefficient η can be determined by analyzing single-molecule trajectories $R(t)$, provided that Eq. II.4 is accurate enough. By averaging over an ensemble of trajectories (and neglecting the lhs) in Eq. II.4 one finds:

$$0 = -\eta \left\langle \dot{R} \right\rangle - \left\langle G'(R) \right\rangle + k_c (ut - \left\langle R \right\rangle) \qquad \text{(II.6)}$$

If we write $R(t) = \langle R(t) \rangle + \delta R(t)$, then, to second order in $\delta R$,

$$\left\langle G'(R) \right\rangle \approx G'(\left\langle R \right\rangle) + \left\langle \delta R^2 \right\rangle G'''(\left\langle R \right\rangle)/2$$

The magnitude of $\left\langle \delta R^2 \right\rangle$ can be reduced by increasing the force constant $k_c$ albeit at the expense of increasing the corresponding magnitude of the fluctuations of the force (II.3).[2]Therefore one can choose $k_c$ such that $\left\langle G'(R) \right\rangle$ can be replaced by $G'(\left\langle R \right\rangle)$, resulting in an equation of motion for $\bar{R}(t) \equiv \left\langle R(t) \right\rangle$:

$$0 = -\eta \dot{\bar{R}}(t) - G'(\bar{R}(t)) + k_c (ut - \bar{R}(t)) \qquad \text{(II.7)}$$

If one knows η and $\bar{R}(t)$ then $G(R)$ is obtained from Eq. II.7:

$$G'(\bar{R}(t)) = -\eta \dot{\bar{R}} + k_c (ut - \bar{R}(t)) \qquad \text{(II.8)}$$

Unfortunately, calculating $\bar{R}(t)$ from steered MD trajectories would imply averaging over a large number of single-molecule trajectories and would be computationally prohibitive. Instead, one can use $\bar{R}(t)$ obtained via an appropriate smoothing procedure such as averaging *R(t)* over a time window centered around t.[2] The errors introduced by using a smoothed *R(t)* instead of the actual $\bar{R}(t)$ depend on the type of averaging. Useful guidelines for choosing the smoothing procedure and minimizing the errors have been established and theoretically justified in ref.[2] Unfortunately, in the particular range of pulling speeds used in our MD simulations, we have not been able to apply this procedure successfully: the statistical errors in the determination of $G'(R)$ were always comparable with the force itself.

We now turn to the determination of the friction coefficient η. It is possible to obtain it by analyzing the force-force correlation function[2]. Alternatively, one can obtain the friction coefficient by comparing single-molecule trajectories recorded for two different values of the pulling speed *u*. Let $\bar{R}_1(t)$ and $\bar{R}_2(t)$ be the trajectories obtained for

$u = u_1$ and $u_2$, respectively. Define functions $t_1(R)$ and $t_2(R)$ such that $\bar{R}_1(t_1(R)) = R$ and $\bar{R}_2(t_2(R)) = R$. Then, by using Eq. II.8, we have

$$G'(\bar{R}_1(t_1(R))) = -\eta \dot{\bar{R}}_1(t_1(R)) + k_c(u_1 t_1(R) - \bar{R}_1(t_1(R))) \tag{II.9a}$$

$$G'(\bar{R}_2(t_2(R))) = -\eta \dot{\bar{R}}_2(t_2(R)) + k_c(u_2 t_2(R) - \bar{R}_2(t_2(R))) \tag{II.9b}$$

and subtracting Eq. (II.9a) from Eq. (II.9b) one finds:

$$\eta = k_c \frac{u_1 t_1(R) - u_2 t_2(R)}{\dot{\bar{R}}_1(t_1(R)) - \dot{\bar{R}}_2(t_2(R))} \tag{II.10}$$

Since the result of Eq. II.10 should be independent of the choice of $R$, then – in the hope to reduce the errors resulting from the smoothing procedure – one may write

$$\eta = \frac{k_c}{R_b - R_a} \int_{R_a}^{R_b} dR \frac{u_1 t_1(R) - u_2 t_2(R)}{\dot{\bar{R}}_1(t_1(R)) - \dot{\bar{R}}_2(t_2(R))} \tag{II.11}$$

where $R_b$ and $R_a$ can be arbitrarily chosen such that they lie within the range of extensions spanned by $\bar{R}_1(t)$ and $\bar{R}_2(t)$.

By choosing any two trajectories plotted in Fig. II.2(a) and applying Eq. II.11 we have calculated the friction coefficient to be $\eta \approx 2.8 \times 10^{-12}$ (N ×s/m). Because of the averaging over R performed in Eq. II.11, the value of friction coefficient thus obtained is less sensitive to the smoothing errors than the potential $G(R)$, which we have not been able to reliably determine from the data plotted in Fig. II.2(a).

## II. 3 THE POTENTIAL OF MEAN FORCE

Having not been able to accurately determine $G(R)$ from the SMD trajectories, we resort to an umbrella sampling type of approach[87]. Suppose a harmonic energy term (cf. Eq. II.1), $V(R_0) = k_c(R-R_0)^2/2$, is added to the molecule's energy and the value $R_0$ is fixed. The total free energy of the molecule (including the constraint energy) is equal to:

$$\tilde{G}_{R_0}(R) = G(R) + (1/2)k_c(R - R_0)^2 \tag{II.12}$$

and the equilibrium extension $R$ of the molecule is found from $d\tilde{G}_{R_0}/dR = 0$, which gives

$$f_{eq} = k_c(R_0 - R_{eq}) = G'(R_{eq}), \tag{II.13}$$

where $f_{eq}$ is the equilibrium stretching force on the molecule. Strictly speaking, $R_{eq}$ and $f_{eq}$ are the *most probable* extension and force, which have to be distinguished from the *average* values $<R>$ and $<f>$. [82]In practice, the difference between the most probable and the average values is small provided that the fluctuations of the extension described by the standard deviation $<\delta R^2> = <(R - <R>)^2>$ are sufficiently small. One can always suppress fluctuations (i.e., reduce $<\delta R^2>$) by choosing a sufficiently large value of $k_c$. For the value of the spring constant $k_c$ = 1.38 N/m used in our calculations, the above difference turns to be immaterial and for this reason in the following discussion we do not differentiate between the most probable and the average values of the force and the extension, referring to them as the equilibrium values.

In the vicinity of $R_{eq}$, the probability distribution of $R$ is given by

$$p_{R_0}(R) \propto \exp[-\tilde{G}_{R_0}(R)/k_BT] \propto \exp\left[-\frac{k_c + G''(R_{eq})}{2k_BT}(R - R_{eq})^2\right] \tag{II.14}$$

so that the standard deviation from the equilibrium extension is given by:

$$\left\langle \delta R^2 \right\rangle = \frac{k_BT}{k_c + G''(R_{eq})} \tag{II.15}$$

We have reconstructed the potential of mean force $G(R)$ by using the two methods below:

**Method 1:** *G(R) is equal to the equilibrium extension work.*

(i) Perform a series of equilibrium molecular dynamics simulations with different values of $R_0 = R_0(i)$, $i = 1, 2, \ldots, N$

(ii) For each $i$ compute $R_{eq}(i) = <R>$ and $f_{eq}(i) = k_c(R_0(i) - R_{eq}(i))$

(iii) Interpolate between the points $(R_{eq}(i),\ f_{eq}(i))$ by using, e.g., a polynomial fit, to obtain the dependence $f_{eq}(R_{eq})$

(iv) Obtain $G(R)$ by integrating Eq. II.13:

$$G(R) = \int_0^R f_{eq}(R_{eq})dR_{eq} \tag{II.16}$$

**Method 2:** *(Self-consistent histogram method[31,36]).*

From Eq. II.12 we have

$$G(R) = \tilde{G}_{R_0}(R) - (1/2)k_c(R - R_0)^2 = -k_B T \ln p_{R_0}(R) - (1/2)k_c(R - R_0)^2 , \qquad \text{(II.17)}$$

so that *G(R)* can in principle be found from the histogram of the extension *R* recorded for any given value of *R₀*. In practice, this procedure is accurate only for the values of R in the vicinity of *R₀* thus necessitating performing the simulation for different values of *R₀* as in step (i) of Method 1. In the self-consistent histogram method[31,36] one combines the information obtained from each of these simulations and finds an optimal estimate for $p(R) \propto \exp(-G(R)/k_B T)$ by writing this as a linear combination of the estimates obtained for each *R₀* and minimizing the error. This approach is superior to Method 1 because it utilizes all information contained in each distribution $p_{R_0}(R)$ rather than only its moments.

We have computed G(R) by using both of the above methods and obtained identical results. The discussion below assumes using Method 1, as it is somewhat more intuitive.

In performing biased MD simulations for each $R_0$ ( step (i) ) one has to ensure that the molecule has achieved thermal equilibrium in each of the simulations. The equilibration time $\tau_{eq}$ can be roughly estimated by assuming that *G(R)* is harmonic such that $G''(R) = G''(0)$ is independent of R. Then one finds[2]

$$\tau_{eq} = \eta/(k_c + G'') < \eta/k_c \qquad \text{(II.18)}$$

For the system studied here we find $\eta/k_c \sim 2$ ps so that we expect equilibration in a few picoseconds. In practice, we have run each simulation for a considerably longer time (50 ps) to ensure proper equilibration and to reduce statistical errors. If one performs a 50 ps calculation for N=20 different values of $R_0$ then the total trajectory time is 1 ns.

The choice of the initial configurations of the molecule used to start each of the N trajectories may strongly affect the performance of the method. If, for a given *R₀(i)*, an initial configuration differs considerably from equilibrium configurations, such poor choice of the initial condition may lead to an anomalously long equilibration time and, possibly, induce conformational changes in the molecule that are absent when the distance between the ends of the domain is changed continuously. One may choose to increment $R_0$ by the same amount in each equilibration run, i.e., set $R_0(i) = i \Delta R$ and to use, as a starting point in the i-th run, the molecule's configuration obtained at the end of the ($i$-1)-th run. This procedure would be equivalent to performing a steered MD calculation according to Eq. II.1 with $R_0(t)$ being a stepwise function. Again, one fears that such stepwise changes may lead to computational artifacts.

In an effort to avoid these difficulties, we chose a different approach. First, we perform adiabatic mapping[87,103] by starting from the minimum energy structure of the molecule, increasing $R_0$ in small ($\Delta R$=0.1Å) increments and re-minimizing the structure in each step. This can be thought as a zero-temperature pulling simulation performed at infinitely slow pulling velocity. Because energy minimization takes much less CPU time than a 50 ps MD trajectory, it is not difficult to ensure that the distance increment $\Delta R$ is small enough that $R_0$ is increased nearly continuously. As a result of the adiabatic mapping, we have (locally) minimal energy structures for each value $R_0$; we use those as initial configurations for each of the MD trajectory performed in step (i) of the above procedure. Fig. II.3 displays MD trajectories obtained for different values of $R_0$. We observe that for $R_0 \leq 11.8$ Å the average force $<f>$ increases as $R_0$ is increased (Fig. II.3(a)). However for $R_0 = 12.8$ Å and higher, the equilibrium value of the force drops and becomes much lower (see Fig. II.3(b)). We associate this behavior with the unfolding of the I27 domain. Indeed, by examining the configurations of the molecule for these values of $R_0$ we find that the parallel strands A and G' become separated when $R_0 = 12.8$ Å (Fig. II.4) This behavior has been previously shown[78,79] to correspond to the unfolding of the I27 domain, lowering its mechanical resistance and resulting in a sharp drop in its force-extension curves $f(R)$.

These observations are consistent with the picture that $G(R)$ has a maximum at $R = R^\dagger$ and the force $f = G'(R)$ drops abruptly for $R > R^\dagger$. The precise location of the top of the barrier $R^\dagger$ is hard to determine from our calculation because once the free energy $G(R_{eq})$ is within the thermal energy $k_BT$ from the transition state barrier $G(R^\dagger)$ , unfolding via thermally activated barrier crossing can take place during the simulation. In other

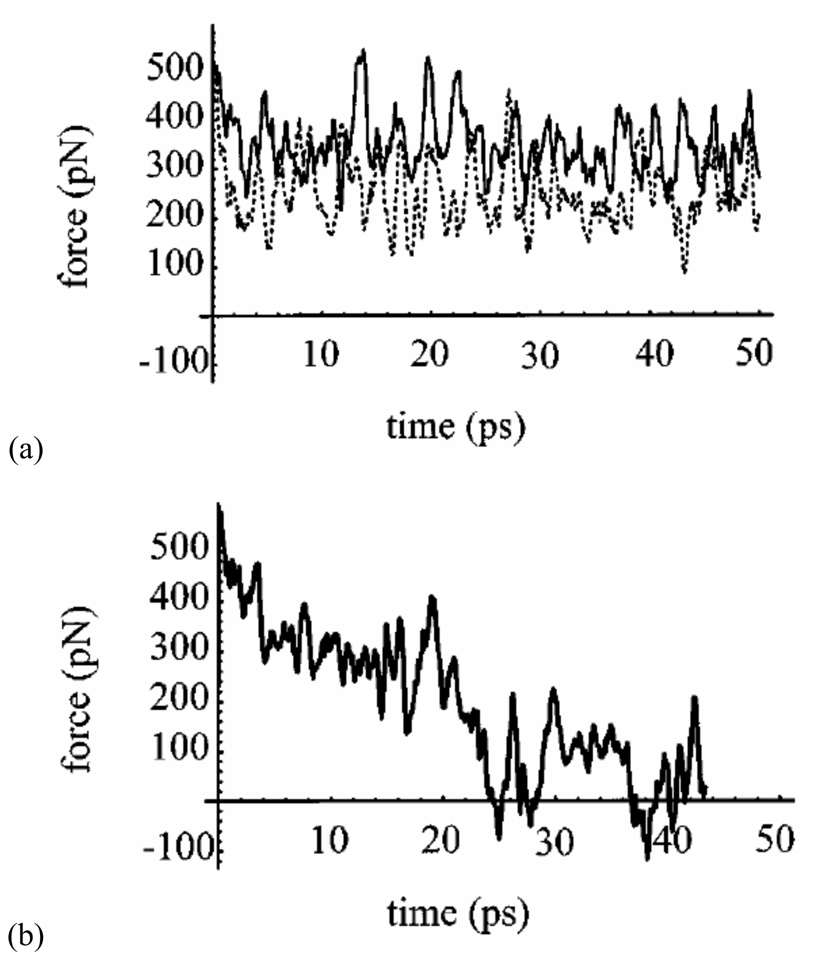


Figure II.3: Stretching force as a function of time measured in MD simulations, in which a harmonic constraint was imposed on the molecule's extension R such that the latter was constrained to be close to $R_0$. (a) Dashed line: $R_0 = 8.8$ Å. Solid line: $R_0 = 11.8$ Å. (b) $R_0 = 12.8$ Å.

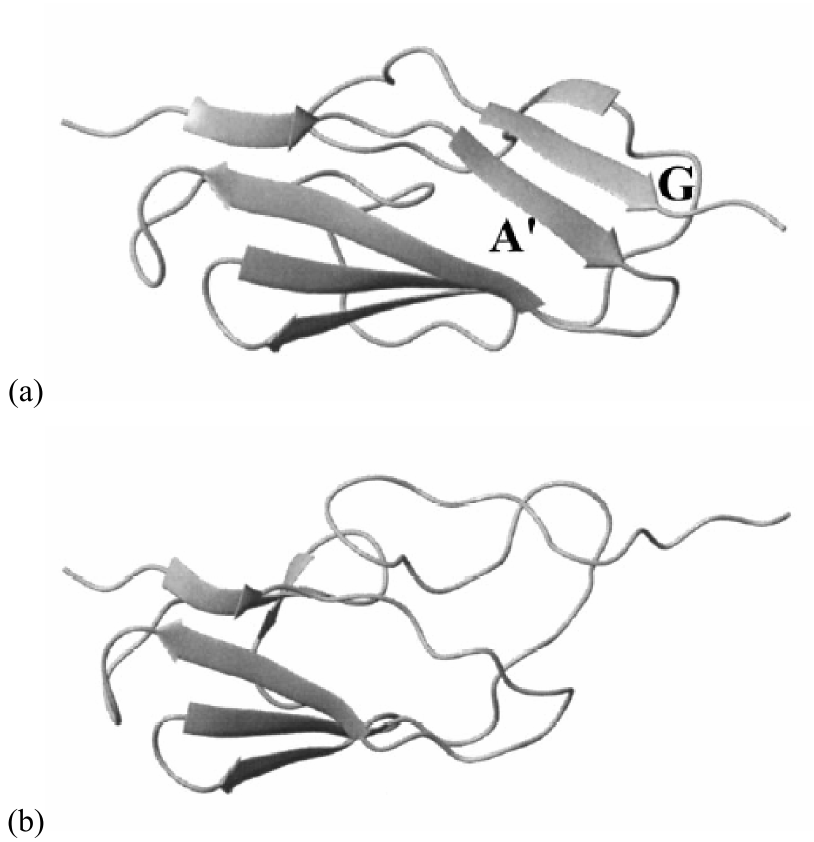


Figure II.4: Snapshots of the I27 domain obtained in the course of MD simulations for (a) $R_0$ = 11.8 Å and (b) $R_0$ = 12.8 Å. The parallel strands A' and G shown in Fig.II.4(a) become separated in the case (b). This picture was generated with the MOLMOL software[63].

words, $R_{eq}$ represents a local free energy minimum of the free energy $\tilde{G}(R)$ and can only be determined if it is well separated from other minima, a condition that is violated once $R_{eq}$ is close to $R^{\dagger}$. Thus our approach cannot yield the precise shape of the free energy barrier in the vicinity of $R^{\dagger}$ and the magnitude of the barrier can only be determined to within ~$k_BT$. In view of this and of the observation that the force drops abruptly once $R_0$ exceeds a critical value we model the free energy $G(R)$ and the force $f(R) = G'(R)$ as discontinuous functions so that $f(R) = 0$ for $R > R^{\dagger}$. Note that the lack of knowledge of the precise shape of $G(R)$ to the right of the transition state $R^{\dagger}$ does not affect the transition-state theory analysis of unfolding described in subsequent sections. Assuming that the critical value of the constraint, for which unfolding takes place, is $R_0$ = 12.8 Å, we find $R^{\dagger}$ from the equilibrium condition: $k_c(R_0 - R^{\dagger}) = G'(R^{\dagger})$, which gives $R^{\dagger} \cong 10.0$ Å. The equilibrium forces $f_{eq}$ and extensions $R_{eq}$ obtained from equilibrium MD trajectories for different values of $R_0$ as well as a polynomial fit $f(R)$ of these data points are shown in Fig. II.5. The dependence $G(R)$ is obtained by integrating $f(R)$ (see Eq. II.16). In particular, the magnitude of the unfolding free energy barrier is given by:

$$\Delta G_u \equiv G(R^{\dagger}) = \int_0^{R^{\dagger}} f(R)dR \simeq 23 \text{ kcal/mol}, \qquad \text{(II.19)}$$

close to the value $G_{exp}$ = 22.2 kcal/mol inferred from experiments[69].

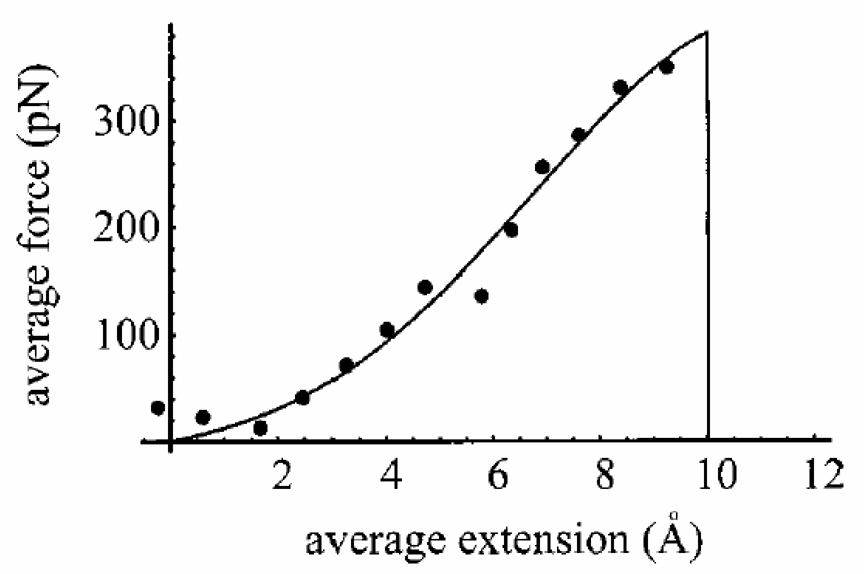


Figure II.5: Average stretching force $f_{eq}$ as a function of the average extension $R_{eq}$ of the I27 domain.

## II. 4 LANGEVIN EQUATION MODEL: COMPARISON WITH MD SIMULATIONS

To verify that the free energy surface *G(R)* obtained in Section II.3 describes the same unfolding pathway as the one observed in SMD simulations described in Section II.2, we have simulated the dynamics described by the Langevin equation (II.4) for the same values of the pulling speed *u* as the ones used in our steered MD simulations in Section II.2. The algorithm used to solve the Langevin equation is the same as the one described in ref.. [2]The results are shown in Fig. II.2(b) and are to be compared with Fig. II.2(a). This comparison suggests that the steered MD results are well described by the Langevin equation model and that our inability to obtain *G(R)* directly from the SMD trajectories is not due to the Langevin equation's failure. This further validates the procedure we used to compute *G(R)*.

## II. 5 COMPARISON WITH AFM EXPERIMENTAL DATA

Equipped with the potential of mean force *G(R)* we can now theoretically predict AFM force-extension curves of titin. When the pulling speed *u* is within the experimental range of 1 – $10^5$ nm/s, the time scale associated with the changes in the pulling force is much slower than that of molecular motions. In this case one can assume that the molecule experiencing the potential

$$G_f(R) = G(R) - fR \tag{II.20}$$

is (nearly) equilibrated in this potential at any given force $f$. The unfolding is then a thermally activated process described by a rate that can be calculated from transition state theory:

$$k_u(f) = \nu(f)\exp[-\Delta G_u(f)/k_B T] \tag{II.21}$$

Here the free energy barrier is determined as the difference between the maximum and minimum values of the free energy in the presence of the force, as illustrated in Fig. II.6(a):

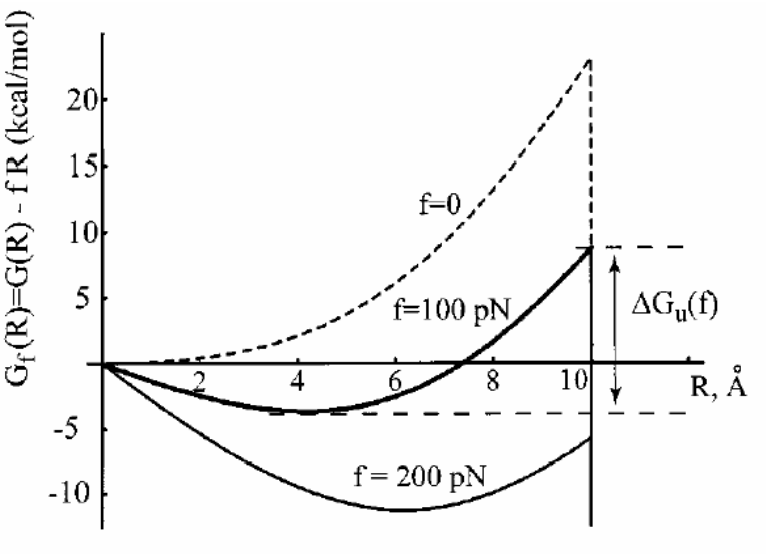
$$\begin{aligned}\Delta G_u(f) &= \max_R G_f(R) - \min_R G_f(R) \\ &= G(R_{\max}(f)) - G(R_{\min}(f)) - f[R_{\max}(f) - R_{\min}(f)]\end{aligned}, \tag{II.22}$$

where $R_{min}(f)$ and $R_{max}(f)$ are the positions of the minimum and the maximum of $G_f(R)$, respectively. The force dependence of the free energy barrier, as calculated from the the minimum and the maximum of $G_f(R)$ are only weakly shifted by the force *f*)

The prefactor $\nu(f)$ can in principle be calculated from Kramers' theory[41]:

$$\nu(f) = \frac{\sqrt{(\partial^2 G_f(R)/\partial R^2)\big|_{R=R_{\min}(f)}(-\partial^2 G_f(R)/\partial R^2)\big|_{R=R_{\max}(f)}}}{2\pi\eta}, \tag{II.23}$$

Unfortunately, as explained above, the curvature of $G_f(R)$ in the vicinity of its maximum $R_{max}$ could not be extracted from the simulation . Using our fit for *G(R)*, the curvature $G''_{RR}$ should be in the range ~10 – 50 pN/Å and thus one estimates the prefactor to be ν ~ $10^{10}$ $s^{-1}$. However because we used an implicit solvation model, the only source of energy dissipation in our case is that into degrees of freedom of the molecule itself. The friction due to the interaction with water molecules is not present in our case although it can be taken into account in an ad hoc manner by performing Brownian dynamics simulations[87]



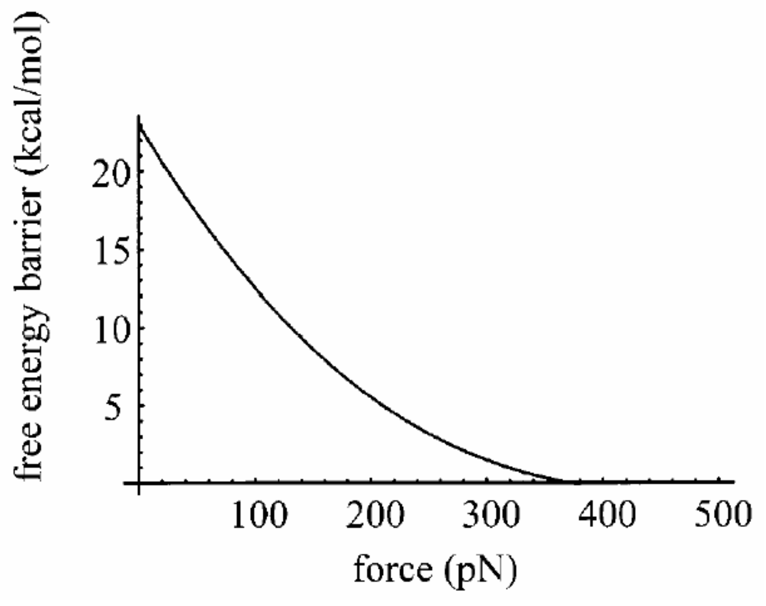


Figure. II.6. (a) The free energy $G_f(R) = G(R) - fR$ for f = 0, 100, and 200 pN. The freeenergy barrier $\Delta G_u(f)$ is indicated for f = 100 pN. (b). The free energy barrier $\Delta G_u(f)$ as a function of force.

instead of molecular dynamics. The actual friction coefficient is then expected to be larger than the value of η inferred from our MD trajectories. The above estimate for ν only provides an upper limit for ν, although a more accurate first-principles estimate can in principle be obtained by including water molecules explicitly in the calculation.

Suppose that the I27 domain is acted on by a force *f(t)* that is increased as a function of time. The probability *p(t)dt* that it unfolds between *t* and *t* + *dt* is given by[81,99,28,29]:

$$p^{(t)}(t)dt = k_u(f(t))\exp\left[-\int_0^t k_u(f(t'))dt'\right]dt \tag{II.24}$$

and the probability distribution of the force *F*, at which the domain unfolds, is given by

$$p^{(F)}(F)dF = \left.\frac{p(t)}{|df/dt|}\right|_{f(t)=F} dF \tag{II.25}$$

The unfolding force distribution $p^{(F)}(F)$, as well as the mean unfolding force $\langle F\rangle = \int dF\, F\, p^{(F)}(F)$ are the quantities usually measured in experiments and will be calculated here from our model. To do this, we need to know how the stretching force *f(t)* changes as a function of time. In AFM studies, the unfolded protein domain is part of the titin chain that is attached to a cantilever and includes unstructured chain segments as well as other domains. The effective force constant of an unfolded domain is typically much higher than that of unstructured polypeptides and of the cantilever. Thus for the purpose of calculating the force *f(t)* acting on a domain one may neglect the domain extension itself. Following refs.[99,100] we model the overall elastic response of the chain by using the wormlike chain model[34,64]. In this model, the relationship between the chain extension x and the force $g(x)$ is given by

$$g(x) = \frac{k_B T}{l_p}\left[(1/4)(1-x/L)^{-2} - 0.25 + x/L\right], \tag{II.26}$$

where $l_p$ is the persistence length and L is the contour length of the chain. In our calculations, we used L = 580 Å, $l_p$ = 4Å[99,100]. In an AFM experiment, one end of the molecule is attached to a substrate while the other end is attached to a cantilever so that the total displacement of the substrate is equal to the sum of the cantilever displacement $x_c$ and the molecule's extension x:

$$ut = x_c + x = f(t)/\gamma_c + x \qquad \text{(II.27)}$$

where $\gamma_c$ is the cantilever force constant, taken in our calculations equal to 0.06 N/m[115,117]. The time dependence of the force is then determined by numerically solving the equation

$$f(t) = g(x) = g(ut - f(t)/\gamma_c) \qquad \text{(II.28)}$$

Once the dependence $f(t)$ is determined, one can compute the distributions of the unfolding force and time either by using Eqs. II.24, II.25 or by performing kinetic Monte Carlo simulations[81], in which the outcome of an unfolding experiment is simulated according to the time-dependent unfolding probabilities determined by $k_u(f(t))$. Simulated distributions of the unfolding force for different pulling velocities are shown in Fig. II.7.(a). The mean unfolding force, as a function of $u$, is presented in Fig. II.7.(b). Since we do not have a first principle estimate for the prefactor ν, we chose it to be independent of the force and equal to $\nu = 10^8$ $s^{-1}$, a value, for which the estimated unfolding forces fall within experimental range. This value is consistent with the above estimate for the upper limit for ν. While the choice of ν determines the magnitude of the unfolding force, the resulting slope in the nearly linear dependence of the mean unfolding force on the pulling speed crucially depends on the properties of $G(R)$ and is found in Fig. II.8 to be close to that observed experimentally[69,99,100].

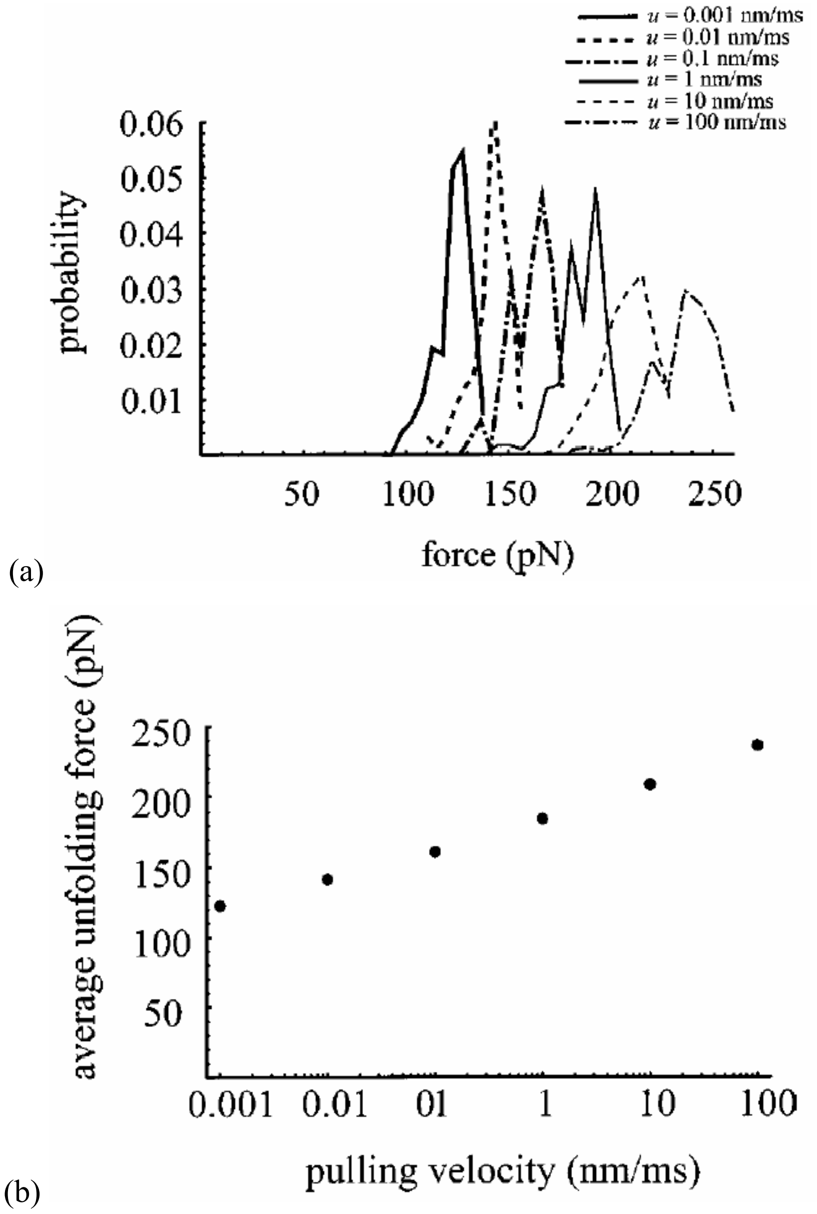


Figure. II.7. (a) The predicted unfolding force distribution for different values of the pulling velocity *u*. (b) Average unfolding force as a function of the pulling velocity *u*.

## II. 6 DISCUSSION

The unfolding free energy profile *G(R)* is often deduced from experiments[9,11,21,69,99] by comparing the experimental data with the results of kinetic Monte Carlo simulations similar to that reported in Section II.5. In doing so, one often assumes that the free energy barrier $\Delta G_u(f)$ changes linearly with the force. Fig. II.6 suggests that this approximation is inaccurate and this may affect some of the conclusions drawn on the basis of AFM measurements. This is demonstrated below.

In the standard two-state model that is commonly used to interpret the experimental unfolding data[9,11,21,69,99] the unfolding rate depends on the force according to the equation:

$$k_u(f) = \alpha_0 \exp(f \Delta x_u / k_B T), \tag{II.29}$$

where the prefactor $\alpha_0$ is identified with the "intrinsic" unfolding rate $k_u(0) = \alpha_0$ and $\Delta x_u$ is the extension of the molecule in the transition state relative to the folded state; In other words, $\Delta x_u$ should be the same as $R^{\dagger}$. These two parameters have been estimated, based on experimental data[21], to be $\Delta x_u$ =2.5Å and $\alpha_0 = 3.3\times 10^{-4}$ s$^{-1}$. These estimates appear at first glance to disagree with our results. Specifically, in Section II.3 we have found $R^{\dagger}$=10Å. Further, using Eq. II.21, we estimate the "intrinsic unfolding rate" to be $k_u(0)$ = $1.4 \times 10^{-9}$ s$^{-1}$, five orders of magnitude lower than the value of $\alpha_0$ reported experimentally. There is however no contradiction. Experimental measurements of $k_u(f)$ probe a limited range of unfolding forces. In particular, the regime of low unfolding forces is hard to access as it corresponds to very low unfolding probabilities. For a relatively narrow range of forces f, say between $f_0 - \Delta f$ and $f_0 + \Delta f$ one can replace ln $k_u(f)$ by a linear dependence,

$$\ln k_u(f) \approx \ln k_u(f_0) - \frac{\left[\Delta G_u(f)\right]'_{f=f_0}}{k_B T}(f - f_0), \tag{II.30}$$

which is of the form of Eq. II.29. This linearized dependence is plotted in Fig. II.8 for $f_0$ = 200 pN as the solid line, together with the actual $k_u(f)$ shown as the dashed line. Fig. II.8 suggests that for forces $f$ within a range 150 pN < $f$ < 250 pN it would be hard to distinguish between the true dependence of ln $k_u(f)$ and its linearized version described by

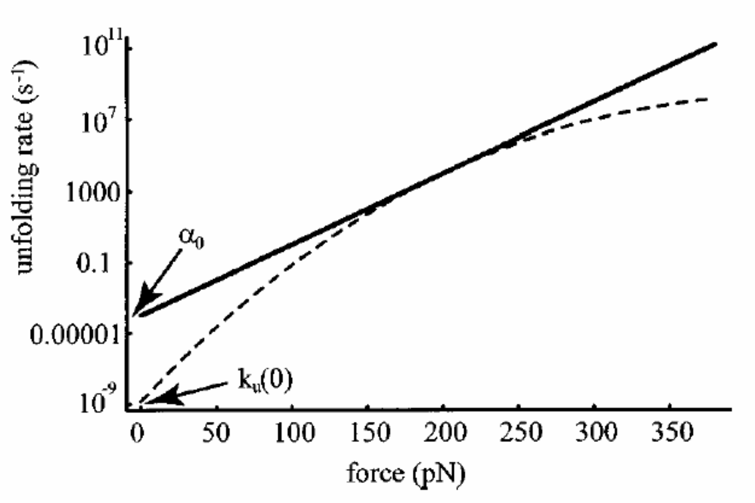


Figure. II.8. The computed force dependence of the unfolding rate constant $k_u(f)$ (dashed line). The solid line is obtained by linearizing ln $k_u(f)$ near $f = f_0 = 200$ pN. The extrapolated value $\alpha_0 = k_u(0)$ obtained from this linearized dependence is different from the actual value $k_u(0)$.

Eq. II.29. By comparing Eqs. II.29 and II.30, we find that the straight line in Fig. II.8 is described by the parameters $\alpha_0 = 1.02\times10^{-4}$ $s^{-1}$ and $\Delta x_u$ = 3.77 Å, which are not too far from the above experimental values[21]. Thus for typical experimental unfolding forces $f$, the apparent values of $\alpha_0$ and $\Delta x_u$ as estimated from our theory are in reasonable agreement with experimental estimates. However the "true" value of $k_u(0)$, as well as the location of the transition state $R^{\dagger}$, as predicted by our theory, are quite different from and $\Delta x_u$ as the zero-force limit of the unfolding rate $k_u(0)$ and the true transition-state extension $R^{\dagger}$, respectively.

One often views the equality of $k_u(0)$ and of the unfolding rate $k_{chem}$ measured in chemical denaturation experiments as an indication of equivalence of chemical and force-induced protein unfolding pathways[21]. In view of the above considerations, comparison between the experimental values of $\alpha_0$ and $k_{chem}$ may be inconclusive. The intrinsic unfolding rate at zero force, $k_u(0)$, estimated here, is much lower than the chemical denaturation rate ($k_{chem} = 4.9\times 10^{-4}$ $s^{-1}$).[21] This is consistent with the idea that at small forces the unfolding molecule finds a pathway with a lower free energy barrier and does not follow the mechanical reaction coordinate[11]. Of course, given the limitations of our MD calculation, the above numerical value of $k_u(0)$ should be taken with a grain of salt.

We finally note that statistical errors did not allow us to resolve finer features of the potential of mean force $G(R)$ such as the previously reported unfolding intermediate, which would manifest itself as a dip in $G(R)$. Investigation of those finer details of the mechanical unfolding mechanisms may require more substantial computational effort.

# Chapter III

# Mechanical unfolding of segment ubiquitin-like protein domains[b]

## III. 1 INTRODUCTION

Single molecule pulling experiments, in which proteins are unfolded by mechanical forces, provide a wealth of information about the mechanical properties of proteins that have load-bearing functions in living organisms[5,9-11,27,32,33,42,43,59,65,69,70,84,88-90,94,99-101,108,111,112,114,116,117,124,125]. Certain proteins, such as titin, can sustain large forces and dissipate large amounts of energy in the course of their mechanical unfolding[78,79,84,99,100,108]. This dissipated energy is often considerably larger than the free energy of folding[106]. This property is believed to account for the remarkable toughness of many natural materials[108]. On the other hand, non-mechanical proteins such as barnase often exhibit very little resistance to mechanical unfolding[10].

Naturally occurring load-bearing proteins often contain sequences of tandemly repeated domains. It is possible to incorporate domains that are not commonly found in natural mechanical proteins into genetically engineered "polyprotein" chains[9,10,122]. Such chains may have novel mechanical properties and have potential applications in fiber and tissue engineering. The overall mechanical response of natural or engineered polyproteins is controlled by the mechanical properties of the constituent individual domains.

Recently we have studied[25] the relationship between the topology and the mechanical unfolding mechanisms of cross-linked polymer chains; Viewing those as caricatures of proteins, we discovered that the maximum resistance to unfolding (measured either as the peak force or the energy dissipated in the course of unfolding) is achieved if the cross-links are organized into a "clamp" formed by parallel strands. It had been previously noticed that the high mechanical strength of the immunoglobulin-like domain I27 of titin is related to the presence of such a clamp in the domain[11,78,79,84], suggesting that its optimality may indeed be utilized in some structural proteins. One may then wonder whether other protein domains, not necessarily having any mechanical function and exhibiting folds different from the β-sandwich fold of I27 but having a similar parallel strand arrangement, would also be highly mechanically resistant.

---

[b] Large portions of this chapter have been previously published as reference 72.

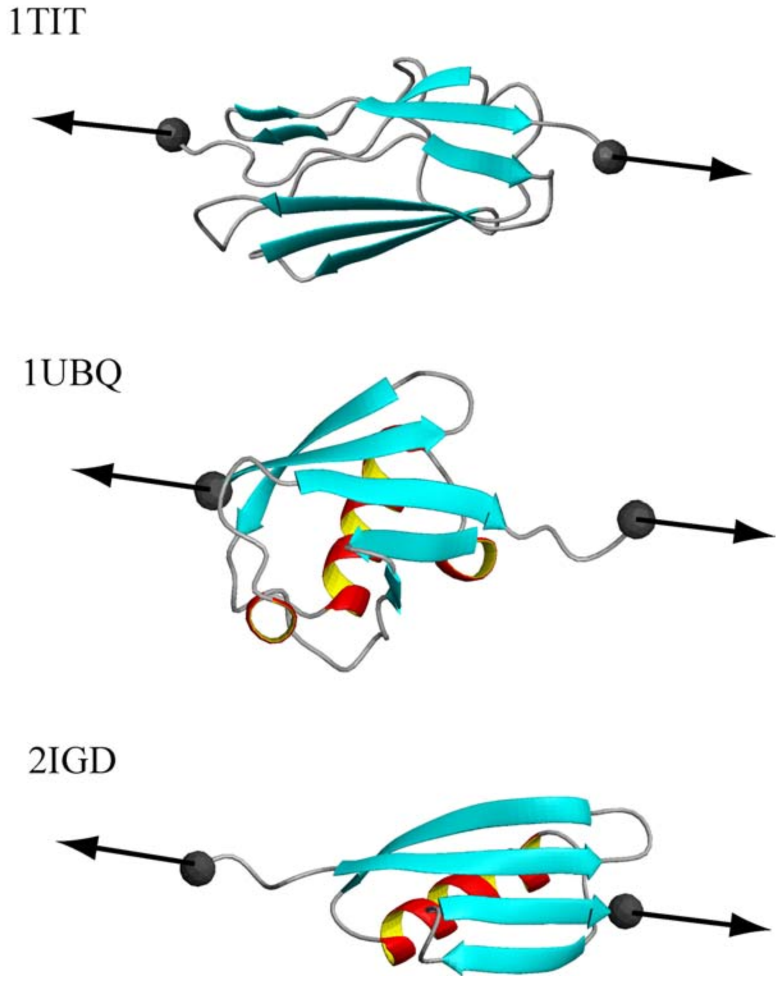


Figure. III.1. The structures of the I27, ubiquitin and protein G domains (pdb codes 1TIT, 1UBQ, and 2IGD, respectively). Arrows indicate at which points the force was applied in the simulation. This figure was generated with the MOLMOL software[63].

In this chapter we present evidence that this is indeed the case. We study the mechanical unfolding of two mixed $\alpha$+$\beta$ domains, streptococcal protein G IgG-binding domain III (pdb code 2IGD) and ubiquitin (pdb code 1UBQ), both having a $\beta$-grasp fold, and show that they display high resistance to unfolding similar to that of I27 and that, as in the case of I27, their mechanical unfolding mechanisms involve separation of parallel strands. See Fig. III.1 for the structures of these two domains along with the I27 domain. We further predict the outcome of an AFM pulling experiment, in which an I27 domain is replaced by one of the above two domains. One of the two proteins, ubiquitin, has recently been studied experimentally[20] , and the unfolding forces found in our simulation are close to those measured experimentally. To our knowledge, the second protein has not been studied experimentally so far.

## III. 2 METHODS

A force $f$ applied at the ends of a protein domain may cause it to unfold. If $f$ is not too high then the domain unfolding is a thermally activated barrier crossing process that can be described by a force-dependent unfolding rate constant $k_u(f)$ (assuming that unfolding is a first order process, which was found to be the case, e.g., for the I27 domain in titin[99]). To calculate $k_u(f)$ one can use transition state theory. We describe the state of the domain by using a single reaction coordinate, the domain extension $R$, defined as the distance between its first and last $\alpha$-carbon atoms. The rate constant $k_u(f)$ can be calculated if the free energy of the domain $G(R)$ is known as a function of $R$. Further, the validity of the assumption that unfolding is a two-state process characterized by a single rate constant is related to the shape of $G(R)$: multiple minima in $G(R)$ would be indicative of unfolding intermediates.

To compute $G(R)$ we used the procedure described in our earlier paper[71]. Here we give a brief summary of our method:

For a set of values $R_0 = R_0(i)$, $i = 1, 2, \ldots, M$, spanning the range of extensions of interest, we perform molecular dynamics (MD) simulations of the domain with a penalty term[10,49,50,78,79]

$$V(R, R_0) = k_c(R-R_0)^2/2 \qquad \text{(III.1)}$$

added to its energy. The force constant $k_c$ used in the calculations was equal to 1.38 N/m. The penalty term introduces a bias that ensures that the MD simulation efficiently samples extensions $R$ in the vicinity of $R_0$. For each value of the distance restraint $R_0$, we compute the equilibrium distribution of the extension $R$, which is related to the potential of mean force $G(R)$ according to the equation

$$p_{R_0}(R) \propto \exp\left[-\left(G(R) + k_c(R-R_0)^2/2\right)/k_BT\right], \qquad \text{(III.2)}$$

allowing accurate reconstruction of $G(R)$ in the vicinity of $R_0$:

$$G(R) = -k_BT \ln p_{R_0}(R) - (1/2)k_c(R-R_0)^2 + \text{constant} \qquad \text{(III.3)}$$

The optimum estimate for $G(R)$ is obtained by using the self-consistent histogram method[31,36], in which it is constructed as a linear combination of the estimates obtained for each value of $R_0$:

$$\exp(-G_{opt}(R)/k_BT) = \sum_i w_i(R) z_i \exp(-G_i(R)/k_BT), \qquad \text{(III.4)}$$

where $\sum_i w_i(R) = 1$ and $G_i(R)$ is the estimate for $G(R)$ obtained from Eq. III.3 with $R_0 = R_0(i)$. The normalization factors $z_i$ and the weights $w_i(R)$ are obtained by minimizing the error, leading to self-consistent equations

$$\exp[-G_{opt}(R)/k_BT] = \sum_i p_{R_0(i)}(R) / \sum_i z_i^{-1} \exp[-(1/2)k_c(R-R_0(i))^2/k_BT] \qquad \text{(III.5)}$$

$$z_n = \int dR \exp[-(1/2)k_c(R-R_0(n))^2/k_BT] \sum_i p_{R_0(i)}(R) / \sum_i z_i^{-1} \exp[-(1/2)k_c(R-R_0(i))^2/k_BT]$$

, which are solved iteratively.

The initial structures for the MD runs for different $R_0$'s were generated by using adiabatic mapping[87,103]: We have started with the X-ray structure of the molecule and minimized its energy. Then we added the penalty term (1) to the energy. Starting with $R_0$ corresponding to the extension of the minimum energy structure, $R_0$ was increased in small ($\Delta R$ = 0.1Å) increments and the energy of the structure was re-minimized in each step. This yielded locally minimal energy structures for each value $R_0(i)$, which were used to start each MD run. This method helps avoid potential problems caused by starting each simulation too far from the mechanical equilibrium. With this choice of initial conditions, the structures were found to equilibrate in a few picoseconds in each MD run. For each $R_0$, we have run 50 ps of MD simulation at 298K, with a time step of 1fs. Both adiabatic mapping and MD simulations were performed with Tinker software[126] using the GB/SA solvation model[98] and the CHARMM27 force field[80].

### III. 3 RESULTS

The potential of mean force $G(R)$ reconstructed from MD trajectories as described in Section III.2 is shown in Fig. III.2 for protein G IgG-binding domain III (2IGD) and ubiquitin (1UBQ). In each case, after an initial rise, $G(R)$ levels out so that the corresponding equilibrium force $f_{\text{eq}} = G'(R)$ drops. This feature indicates that the domain no longer resists force and thus unfolds. Examination of the molecules'snapshots corresponding to the onset of this flat section of the $G(R)$ plot (Fig. III.2) shows than in each case the two parallel strands become separated. This unfolding scenario has previously been observed in the mechanical unfolding of the I27 domain of titin[71,78,79,84]. The unfolding free energy barrier corresponding to the maximum of $G(R)$ is ≈29 kcal/mol for 1UBQ and ≈17 kcal/mol for 2IGD. For comparison, the unfolding barrier for the I27 domain was estimated to be ≈22 kcal/mol both from experimental data[21] and simulation[71].

In a typical AFM pulling experiment, the stretching force $f$ applied between the ends of the molecule varies slowly compared to the time scale of molecular motions. For each $f$ the domain then sees a (nearly) static potential $G_f(R)$ given by:

$$G_f(R) = G(R) - fR \tag{III.6}$$

This potential is shown in Fig. III.3 for each protein for different values of the force $f$. As seen in Fig. III.3, application of a force lowers the unfolding barrier, which is defined as the difference between the maximum and the minimum values of the free energy $G_f(R)$.

$$\Delta G_u(f) = \max[G_f(R)] - \min[G_f(R)] \tag{III.7}$$

Once the force is large enough, the free energy barrier disappears and the protein unfolds. For example, the barrier for 2IGD disappears when the applied force is ~250 pN.

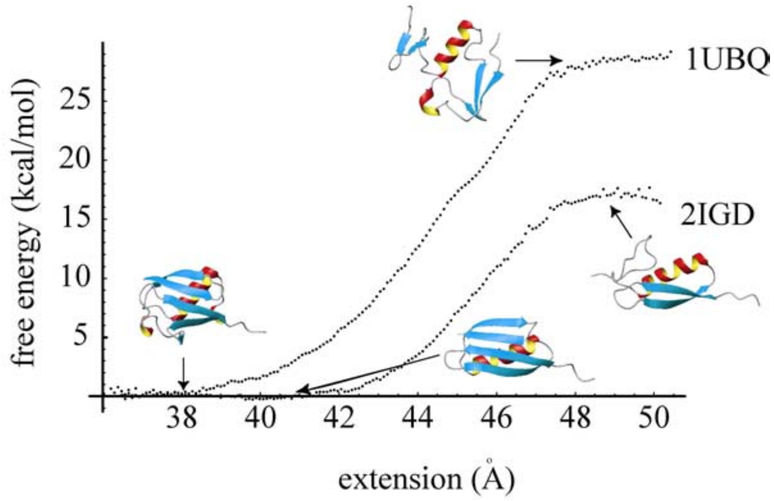


Figure. III.2. The potential of mean force $G(R)$ for 1UBQ and 2IGD. Representative structures corresponding to the folded state and to the unfolding transition state are also shown. The protein pictures were generated with the MOLMOL program[63].

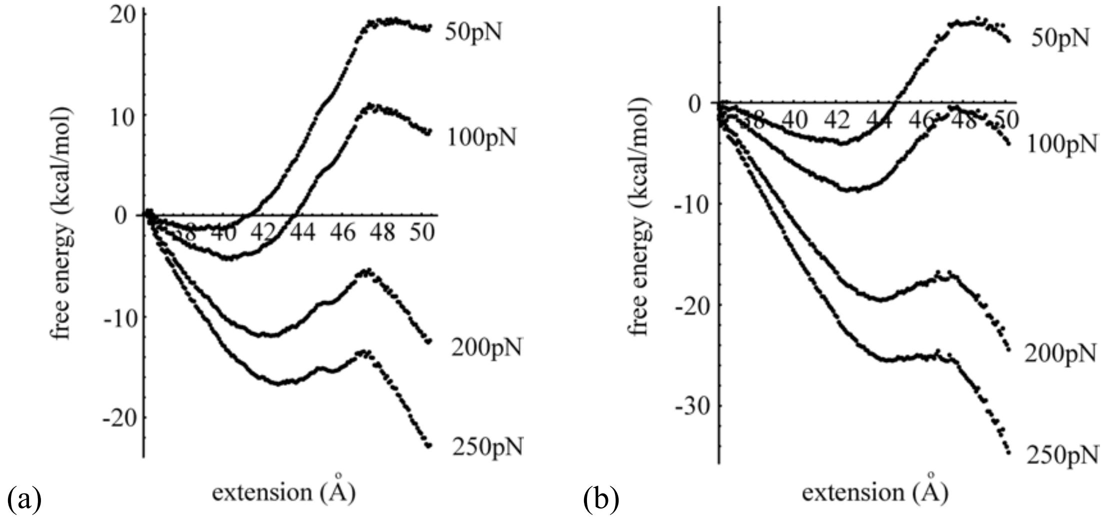


Figure. III.3. The unfolding free energy profile, Eq. III.6, for (a) 1UBQ and (b) 2IGD for different values of the applied force: 50pN, 100pN, 200pN and 250pN.

However under typical experimental conditions domains unfold at forces that are lower than those required to completely wipe out the barrier $\Delta G_u(f)$ because it can be surmounted via activated barrier crossing. The force-dependent rate of barrier crossing is given by transition state theory:

$$k_u(f) = \nu(f)\exp[-\Delta G_u(f)/k_B T] \tag{III.8}$$

If one assumes[2,55,56] that the dynamics along the reaction coordinate $R(t)$ can be viewed as one-dimensional diffusive motion in the effective potential $G_f(R)$ obeying a Langevin equation[2,55,56] then the unfolding prefactor $\nu(f)$ can be estimated from

Kramers' theory[41]. In the absence of memory effects and in the overdamped limit it is given by[41] $\nu(f) = (2\pi\eta)^{-1}\sqrt{G''(R_N(f))G''(R_{TS}(f))}$, where $\eta$ is the friction coefficient and $R_N(f)$ and $R_{TS}(f)$ are the positions of the "native" well and the transition state (i.e., the minimum and the maximum of $G_f(R)$). A number of methods have been proposed to extract the friction coefficient from MD simulations[2,55,56,71]; However because the present simulation uses an implicit solvation model it cannot provide a first principles estimate for $\eta$. Here we simply chose the unfolding prefactor $\nu$ to be the same as that estimated for the I27 domain in our earlier study[71] , $\nu = 10^8$ s$^{-1}$. Figure III.4 illustrates how the unfolding rate constant depends on the force for each protein.

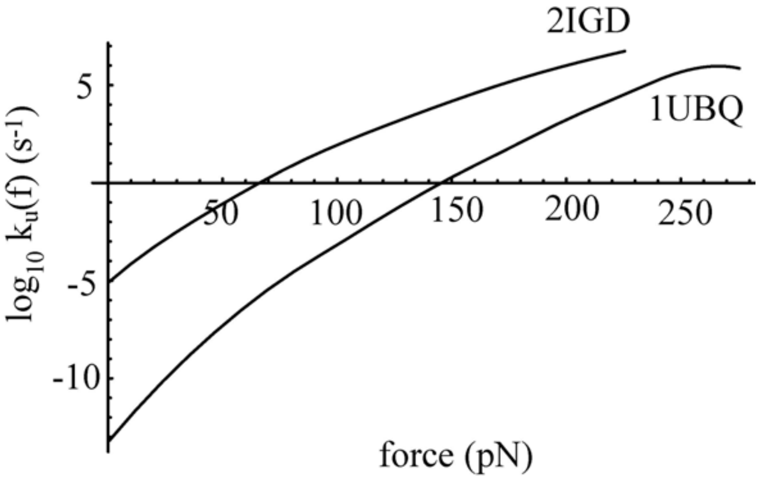


Figure. III.4. Unfolding rate constant as a function of the applied force.

The mechanical unfolding of ubiquitin has been studied in single molecule pulling AFM experiments[20]. To our knowledge, there is no experimental data on the mechanical unfolding of 2IGD. As previously shown[71,81,99], one can use the computed unfolding rates $k_u(f)$ to predict the outcome of such AFM experiments. Imagine a hypothetic experiment, in which one of the I27 domains in the titin molecule is replaced by either 2IGD or 1UBQ. If the molecule is stretched at a constant velocity $u$, the probability $p(t)dt$ that the domain unfolds between time $t$ and $t + dt$ is given by:

$$p(t)dt = k_u(f(t))\exp[-\int_0^t k_u(f(t))dt]dt \tag{III.9}$$

and the probability distribution for the unfolding force $F$ is found from Eq. III.9:

$$p(F)dF = \left.\frac{p(t)}{|df(t)/dt|}\right|_{f(t)=F} dF \tag{III.10}$$

In AFM experiments, one end of the chain is attached to a cantilever and the other end is fastened on a substrate that moves at a constant velocity $u$. To calculate the stretching force $f(t)$ as a function of time one needs to know the compliance of the entire chain and the cantilever. If the cantilever is modeled as a harmonic spring with a force constant $\gamma_c$ and the wormlike chain model[34,64,83] is used to describe the overall compliance of the chain then one finds[71,81,99] $f(t)$ to be the solution of the equation

$$f(t) = \frac{k_B T}{l_p}[\frac{1}{4}(1-\frac{(ut - f(t)/\gamma_c)}{L})^{-2} - \frac{1}{4} + \frac{(ut - f(t)/\gamma_c)}{L}] \tag{III.11}$$

The persistence length $l_p$ = 4Å, the contour length $L$ = 580 Å[99,100], and the cantilever force constant $\gamma_c$ = 0.06 N/m[117,115] are chosen here to be the same as those in our previous study[71]. The predicted average unfolding force $\langle F \rangle = \int dFp(F)F$ is plotted as a function of the pulling speed $u$ in Figure III.5, showing that 2IGD would unfold at somewhat lower forces than ubiquitin and it would also exhibit a stronger dependence of $\langle F \rangle$ on the pulling rate. For typical AFM pulling speeds $u$ = 0.1 – 10 nm/ms the unfolding force is in the range 100-150 pN for 2IGD and 190-220 pN for 1UBQ. For ubiquitin, the measured unfolding forces[20] are close to the values calculated here.

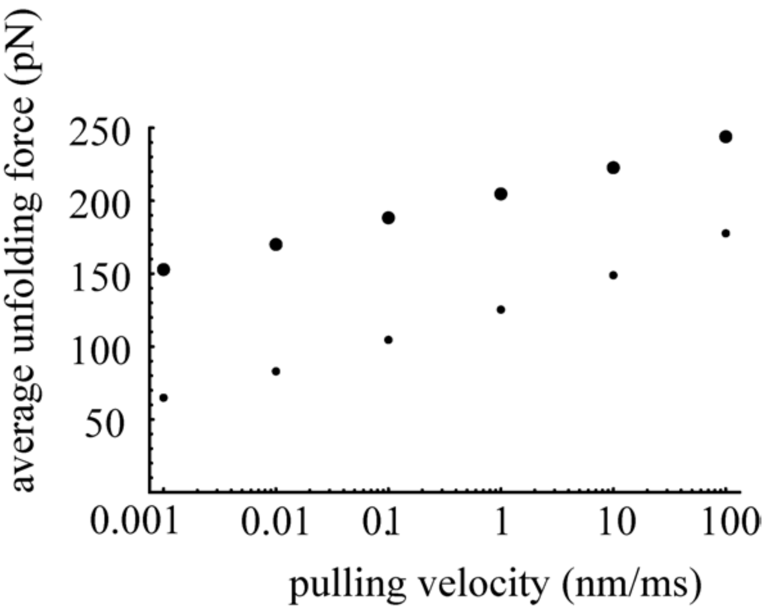


Figure. III.5. Predicted average unfolding force as a function of the pulling speed for 1UBQ(●) and 2IGD (•).

## III. 4 DISCUSSION

The high mechanical resistance of ubiquitin-like domains appears to be accidental rather than necessitated by their function. These $\alpha+\beta$ proteins display mechanical strength comparable to that of the $\beta$-sandwich domains implicated in load-bearing tasks in living organisms. Their mechanical unfolding mechanism is also similar to that of the I27 domain in the muscle protein titin and involves separation of two parallel strands, which are seen to be the key to their high resistance to unfolding.

It is interesting to compare the unfolding rates calculated here with those for thermal or chemical denaturation. For ubiquitin, we have calculated $k_u(0)= 5.5\times10^{-14}s^{-1}$, 10 orders of magnitude smaller than the unfolding rate constant under native conditions, $k_{chem} \sim 4.3\times10^{-4}$ $s^{-1}$, which was extrapolated to zero denaturant concentration from ubiquitin chemical denaturing experiments[60]. We have seen a similar situation in our earlier study of the unfolding of the I27 domain in titin[71]. Linear extrapolation that assumes that the unfolding barrier is a linear function of the force $f$ and the unfolding rate constant obeys the equation

$$\ln k_u(\mathrm{f}) = \ln k_u(0) - f\,\Delta x/\mathrm{k_B T} \qquad \text{(III.12)}$$

is routinely used for interpretation of the experimental data. [Note however the recent experimental evidence[121] that Eq. III.12 breaks down in the case of the I27 domain for forces below 100 pN]. Eq. III.12 implies that the plots in Fig. III.4 are straight lines, an obviously incorrect assumption in our case. Because AFM probes a limited range of unfolding forces, Eq. III.12 may be valid locally in this range. The value of $k_u(0)$ extrapolated from the experimental data by using Eq. III.12 would then be higher than $k_u(0)$ calculated here.

A more intriguing question is why the calculated $k_u(0)$ is so much different from the intrinsic unfolding rate (which we identify with the chemical denaturation rate): In fact, the calculated mechanical unfolding rate $k_u(f)$ is lower than $k_{\mathrm{chem}}$ for $f < 90$ pN. Given a good agreement of our simulations with experimental data in the experimental range of forces, it seems unlikely that simulation errors could account for such a large discrepancy. It is more likely that the reason is a poor choice of the unfolding coordinate. Our procedure here is essentially a variational transition state theory[113] calculation of the rate, in which a reaction coordinate measuring the unfolding progress is selected and the free energy barrier is computed along this coordinate. Imposing a large force between the C- and N-termini of the chain selects a natural unfolding coordinate equal to the extension $R$. However $R$ may be a poor reaction coordinate when the force $f$ is small or zero. In a system that has a single transition state, a poor choice of the reaction coordinate can only lead to an overestimate for the rate constant because transition state theory ignores barrier recrossing. Thus one should expect that the true rate should be even smaller than estimated. However if the system has multiple transition states, a poor choice of the reaction coordinate may lead to the sampling of the neighborhood of a "wrong" transition state (i.e., not the one with the lowest free energy barrier), since a 50 ps MD trajectory will be unlikely to ergodically sample the entire available conformational space. In other words, driven along the selected reaction coordinate, our system does not have enough time to “discover” the true (i.e., the lowest) unfolding transition state during the course of a short-time simulation. Thus the computed barrier will correspond to the wrong transition state and will be higher than the correct free energy barrier, resulting in an underestimate for the intrinsic unfolding rate. We thus expect that the unfolding pathway should change as the force is lowered and that the low-force unfolding pathway cannot be probed by the present simulation because we do not have a good guess for the reaction coordinate in this case.

# Chapter IV

# Mechanical unfolding of ubiquitin[c]

## IV. 1 INTRODUCTION

In naturally occurring polyubiquitin chains, individual ubiquitin domains are linked either by peptide bonds (N-C termini linkage) or by isopeptide bonds formed between the C-terminus of one ubiquitin domain and the Lys sidechain of the adjacent domain (Lys-C terminus linkage).[22,93,95,96] There are 7 lysine residues in the ubiquitin domain and each can potentially be involved in an isopeptide bond. Isopeptide bonds frequently occur at the Lys11, Lys48, and Lys63 positions, Lys48-C linked polyubiquitin chains being the most abundant form[93]. Different ubiquitin linkages are believed to be associated with different biological functions[95,96]. For example, Lys48-C linked polyubiquitin targets protein substrates for degradation by the proteasome[23] while Lys63-C linked polyubiquitin is involved in DNA repair[52], ribosome function[109], and endocytosis of yeast plasma membrane proteins[37].

Carrion-Vazquez et al[20] used atomic force microscopy (AFM) to study the mechanical response of N-C and Lys48-C linked polyubiquitin chains. They found that the ubiquitin domain shows a considerably lower mechanical resistance when a stretching force is applied between its Lys48 and C-terminus, as compared to the case when the force is applied between the N- and C- termini of the same domain. The average domain unfolding force for the N-C linked polyubiquitin also exhibits a stronger dependence of the pulling rate, as compared to that for the Lys48-C linked chain.

In order to understand these observations, the authors of ref. [20] performed steered molecular dynamics simulations of the stretching of the ubiquitin domain in each case. They found that the unfolding of N-C ubiquitin (cf. Fig. IV.1(a)) involves separation of terminal parallel β–strands, a mechanism similar to that found in the much better studied mechanical unfolding of the I27 domain in the muscle protein titin[11,55,78,79,84]. We have found similar behavior in our own study of the mechanical unfolding of ubiquitin[73]. Stretching Lys48-C ubiquitin (see Fig. IV.1(b)), on the other hand, involves separation of a pair of antiparallel strands[20]. Even though the same number of hydrogen bonds are broken in each unfolding scenario[20] and therefore one expects the energetic cost of unfolding to be comparable, the forces required to unfold the domains are quite different.

---

[c] Large portions of this chapter have been previously published as reference 73.

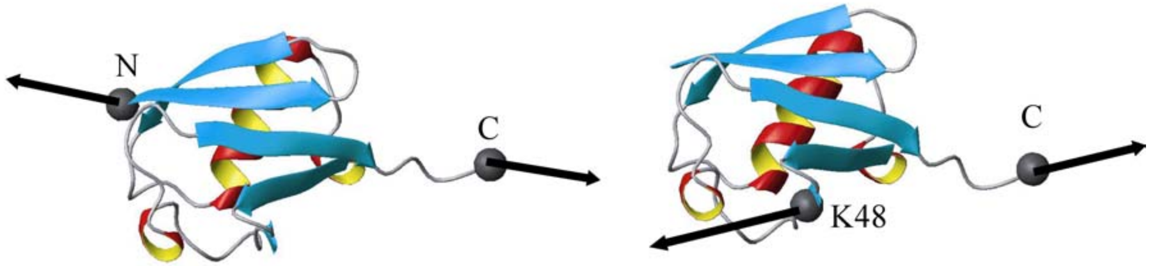
Brockwell et al. [13] studied – both experimentally and computationally – the mechanical unfolding of another protein, E2lip3, and demonstrated that the mechanical resistance of this beta-sheet protein crucially depends on the direction of the applied force.

The purpose of this paper is two-fold. First, we would like to better understand how mechanical unfolding mechanisms depend on where the pulling forces are applied and thereby to gain better insight into the results of the two AFM studies[13,20]. Although both refs.[13,20] reported on simulations of their pulling experiments, those simulations utilized the steered molecular dynamics (SMD) method. SMD simulations provide invaluable insights into the mechanical unfolding mechanisms but they require proteins to be stretched at a rate that is several orders of magnitude higher than that in a typical AFM experiment. As a result the unfolding forces are much higher than those seen experimentally, contain large contributions from dissipative, hydrodynamic-type forces that are typically negligible in an AFM experimental setup[2,38,40,49,50,55,56,71], exhibit a pulling rate dependence that is very different from that observed in AFM experiments, and often cannot be straightforwardly extrapolated to the experimental regime. In the present study, we use umbrella sampling in combination with transition state theory in order to calculate the unfolding forces that can be directly compared to experimental measurements. We have already applied this methodology to the I27[71] and ubiquitin[72] domains.

Secondly, forcing a protein to unfold mechanically by applying forces at different parts of the chain provides an intriguing opportunity to alter the unfolding reaction pathway. It has been previously pointed out that the mechanical unfolding coordinate associated with the distance between the ends of the polypeptide chain may be quite different from that of chemical or thermal denaturation[11,71,72,121] although the rates of chemical unfolding are often comparable in magnitude to those extrapolated to zero force from AFM experiments[21]. By varying the reaction coordinate we can potentially probe different unfolding transition states and drive the molecules along different unfolding pathways thus exploring different slices of multidimensional free energy landscapes of proteins. Some of these pathways may potentially be relevant for chemical or thermal unfolding mechanisms.

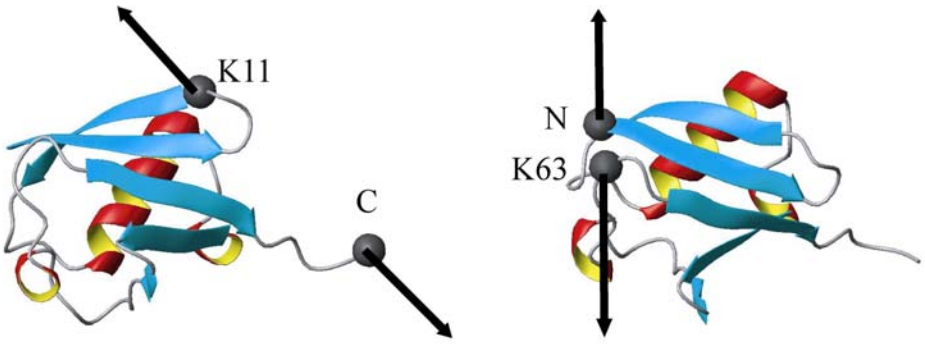
To achieve these goals, we have computed the free energy of the ubiquitin domain as a function of the distance between the points at which the force is applied. We have "pulled" the ubiquitin domain in four different ways shown in Fig. IV.1, in which the force is applied between (a) the N- and C-termini, (b) Lys48 and the C-terminus, (c) Lys11 and the C-terminus, and (d) Lys63 and the N terminus. Cases A and B have been studied experimentally and via steered molecular dynamics simulations[20]. Our results for case A have been previously reported in ref. [72]. Here we include those previous results in order to provide a detailed comparison among all four cases. Lys11-C linked polyubiquitin chains are abundant naturally[93] and so case C can in principle be studied experimentally with the AFM techniques as described in ref. [20] Case D represents a thought experiment. We demonstrate that these 4 different unfolding experiments result in different unfolding scenarios: Sliding of parallel (case a) or antiparallel (case b) strands and unzipping of parallel strands (cases c and d). In each case we predict the outcome of a hypothetic AFM experiment involving the stretching of the ubiquitin domain incorporated within a polyprotein chain.

This chapter is organized as follows: Section IV.2 describes the calculation methods. Our simulation results are described in Section IV.3. These results are compared with the existing experimental data in Section IV.4. Section IV.5 concludes with a discussion of the relevance of our results for chemical/thermal unfolding of proteins.



(a) N-C ubiquitin

(b) K48-C ubiquitin



(c) K11-C ubiquitin

(d) N-K63 ubiquitin

Figure. IV.1. Different pulling geometries used in the simulation: Ubiquitin domain stretched between its (a) N- and C- termini, (b) Lys48 and C terminus, (c) Lys11 and C-terminus, and (d) N-terminus and Lys63. This Figure was generated with the MOLMOL software[63].

## IV. 2 METHODS

To characterize the mechanical response of an ubiquitin domain we compute its free energy $G(R)$ as a function of its extension $R$. The domain extension $R$ is defined as the distance between the α-carbon atoms of the two residues, between which the stretching force is applied. For example, in the case of Lys48-C polyubiquitin $R$ is the distance between the α-carbons of Lys48 and Gly76. The free energy $G(R)$ is then computed by the umbrella sampling[36,87] method as described in our earlier papers[71,72]. Specifically, by adding a penalty term of the form $V(R, R_0) = \gamma_c(R-R_0)^2/2$ to the energy of the molecule, we can accurately sample the free energy in the vicinity of the point $R_0$:

$$G(R) = -k_B T \ln p_{R_0}(R) - (1/2)\gamma_c (R - R_0)^2 + \text{constant}\,, \qquad \text{(IV.1)}$$

where $p_{R_0}(R)$ is the equilibrium probability distribution of the extension obtained from a molecular dynamics simulation. Our simulations were performed with Tinker[126] software using the GB/SA continuum solvation model[98] and the CHARMM27 force field[80].

By repeating this procedure for a set of constraints $R_0$, we get a set of overlapping distributions $p_{R_0}(R)$ and combine the data by using the method of weighted histograms[31,36] to obtain the optimal estimate for the curve $G(R)$.

In a typical AFM stretching experiment the force $f$ acting on the protein is changing slowly enough that one can view the domain's dynamics as occurring in the static potential $G_f(R) = G(R)-fR$. For forces not too high, this potential typically exhibits a minimum corresponding to the structure of the folded domain and a maximum corresponding to the unfolding transition state. If there is only a single unfolding barrier then the unfolding kinetics can be described by a single force-dependent unfolding rate constant $k_u(f)$: [6]

$$k_u(f) = \nu(f)\exp[-\Delta G_u(f)/k_B T] \qquad \text{(IV.2)}$$

where the unfolding barrier is given by $\Delta G_u(f) = \max[G_f(R)] - \min[G_f(R)]$ and the prefactor can in principle be calculated from Kramers' theory[41] . In this paper we have simply assumed the value $\nu=10^8 s^{-1}$ previously estimated for the I27 domain.[71] The probability distribution of the unfolding time is then given by

$$p(t)dt = k_u(f(t))\exp[-\int_0^t k_u(f(t))dt]dt \qquad \text{(IV.3)}$$

and the probability distribution of the unfolding force is

$$p(F)dF = \frac{p(t)}{|df(t)/dt|}\Bigg|_{f(t)=F} dF\ , \qquad \text{(IV.4)}$$

where $f(t)$ is the time-dependent force in the polyprotein chain that incorporates the domain under study. In an AFM setup, one end of the chain is attached to a cantilever and the other to a substrate, and the cantilever can be moved relative to the substrate at a constant speed $u$. To find $f(t)$, one needs to know the compliance of the entire polyprotein chain and of the cantilever. Following previous work[17,99,100], we assume that the elasticity of the chain can be described by the wormlike chain model[34,64]. For a chain with the persistence length $l_p$ and the contour length $L$ and for a cantilever with a force constant $\gamma_c$, this gives a self-consistent equation for $f(t)$: [81]

$$f(t) = \frac{k_B T}{l_p}\left[\frac{1}{4}\left(1-\frac{(ut - f(t)/\gamma_c)}{L}\right)^{-2} - \frac{1}{4} + \frac{(ut - f(t)/\gamma_c)}{L}\right]. \qquad \text{(IV.5)}$$

This equation can be solved numerically. In our calculations, we used $l_p$ = 4Å[99] and $\gamma_c$ = 0.06 N/m[116,117]. The contour length of the chain depends on the polyprotein construct and the number and the type of the domains that are already unfolded; In addition, it changes in the course of the AFM unfolding experiment[11,81,88,125]. However the dependence of the unfolding force on $L$ is relatively week. Somewhat arbitrarily, we chose $L$ = 580 Å here.

## IV. 3 RESULTS

Figure IV.2 shows the computed potential of mean force $G(R)$ for ubiquitin stretched between its (a) N- and C- termini, (b) Lys48 and C-terminus, (c) Lys11 and C-terminus and (d) N-terminus and Lys63. For each of these four cases shown in Fig. IV.3 is the free energy profile in the presence of a stretching force $f$, $G_f(R) = G(R) - fR$, for different values of $f$.

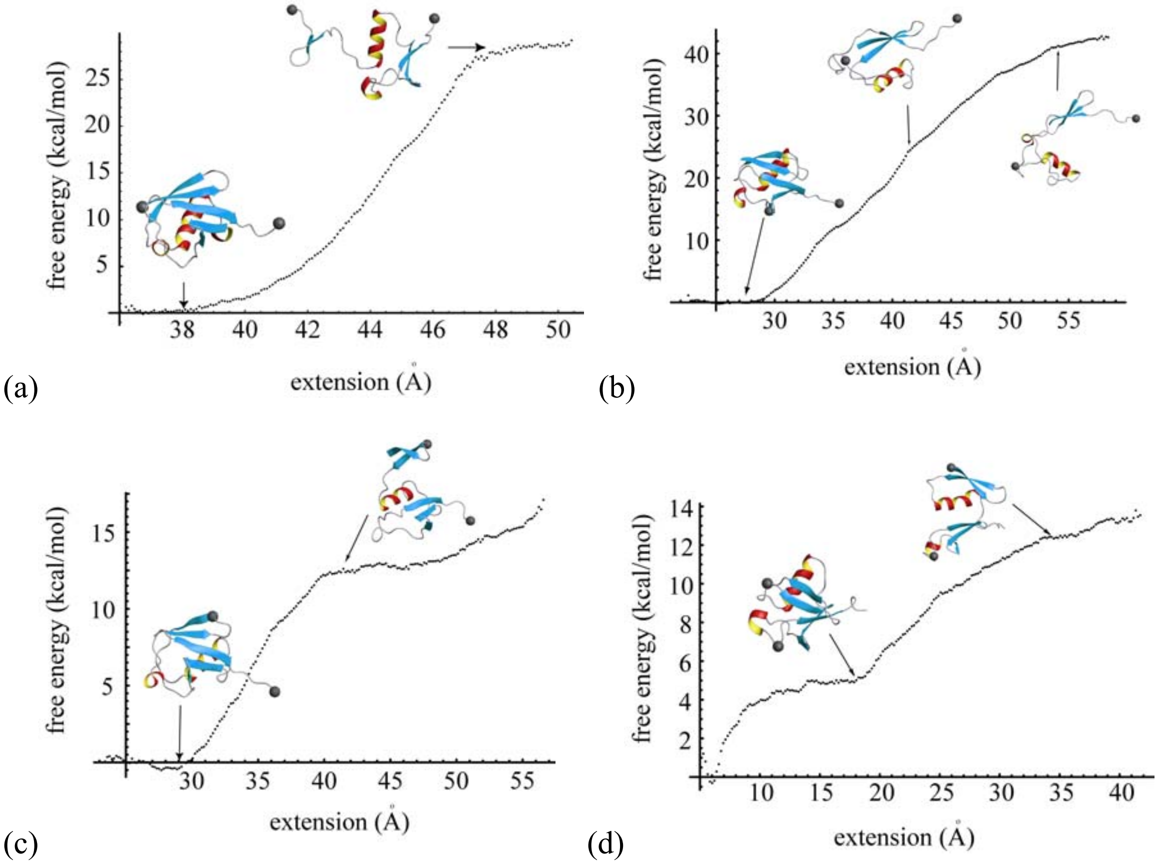


Figure. IV.2. Free energy $G(R)$ as a function of the domain extension for ubiquitin stretched between its (a) N- and C- termini, (b) Lys48 and C terminus, (c) Lys11 and C-terminus, and (d) N-terminus and Lys63. In each case $R$ is the distance between the residues at which the pulling force is applied. Representative snapshots of the domain structure are shown for different values of $R$. These protein representations were generated with the MOLMOL software[63].

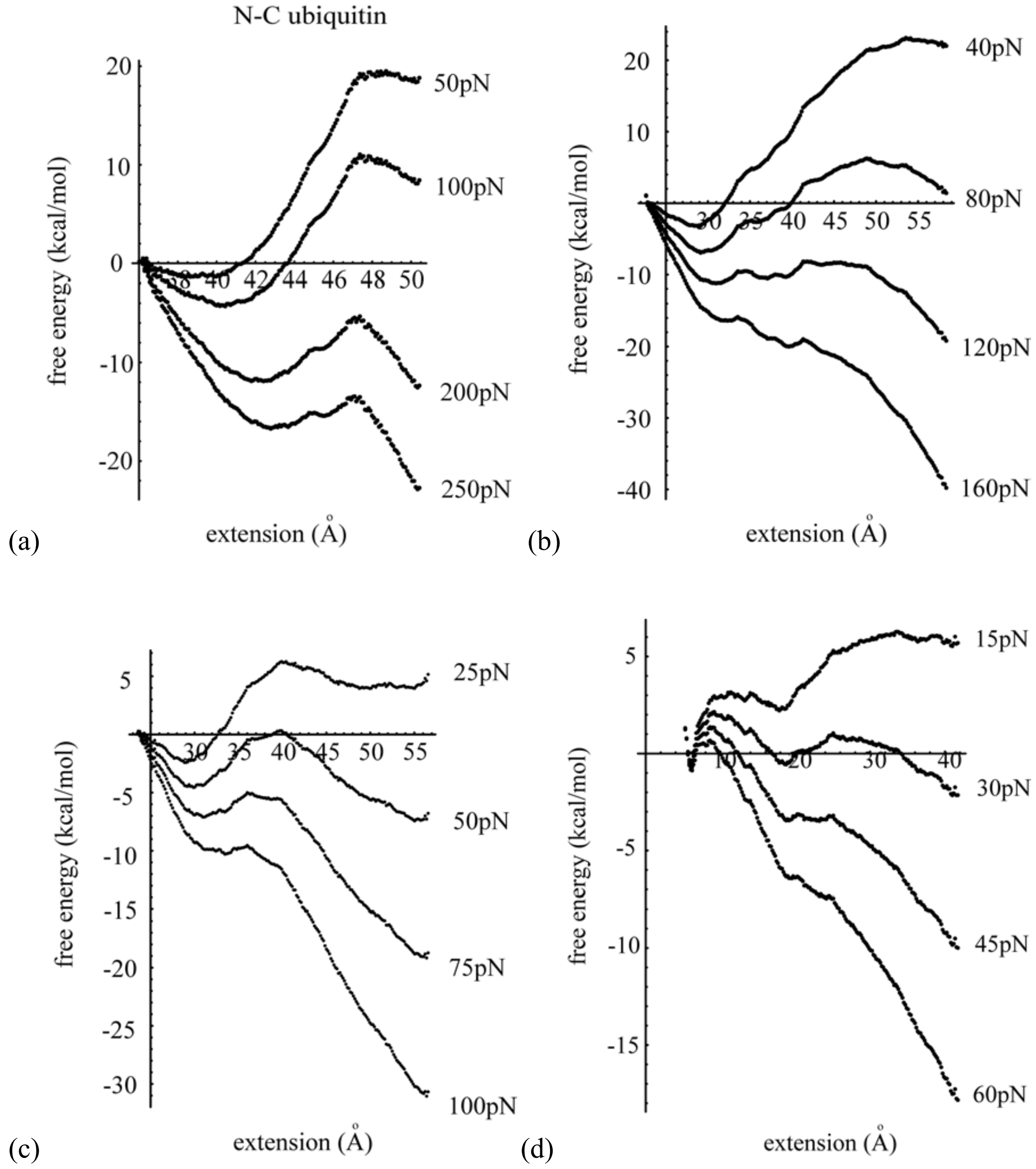


Figure. IV.3. Free energy $G_f(R) = G(R)-fR$ as a function of the domain extension for different values of the pulling force $f$. A: N-C- ubiquitin. B: Lys48-C ubiquitin. C: Lys11-C ubiquitin. D: N-Lys63 ubiquitin

**Case a** (see also ref.).[72] The free energy curve $G(R)$ exhibits a plateau for large values of $R$. This indicates that for large enough $R$ $G'(R)$ becomes close to zero: The domain no longer resists force. This can be understood more clearly by examining Fig. IV.3(a), which shows the unfolding free energy profile $G_f(R)$ for different values of the force $f$. The function $G_f(R)$ exhibits a maximum (i.e., a transition state), which is located at $R = R^{\dagger} \approx 47$ Å. The location of this transition state is nearly independent of the force. Independence of the position of the transition state of the pulling force was previously assumed by Hummer and Szabo[54] and by us[71] to introduce a simple, cusp-shaped model potential for $G(R)$ for the I27 domain in titin. A snapshot of the domain structure near the transition state is shown in Fig. IV.2(a) indicating that crossing the unfolding barrier involves separation of the terminal parallel strands. This is consistent with the previous simulations of ubiquitin[20,72]. The unfolding mechanism of N-C ubiquitin is thus similar to that of the I27 domain in titin[71,78,79,84].

The location $R_{\mathrm{m}}$ of the minimum of $G_f(R)$ (i.e., the most probable extension of the folded domain) increases linearly as the force is increased. Taking the derivative of this we can estimate the stiffness of the folded domain

$$\gamma_{N-C} = df / dR_m \approx 0.5 \text{ N/m}. \tag{IV.6}$$

The same quantity, of course, could also be estimated from the curvature of $G(R)$ near the minimum. This value of stiffness is almost an order of magnitude higher than the cantilever force constant $\gamma_c$. This justifies the assumption made in Eq. IV.5 (and also routinely made in kinetic Monte Carlo simulations of polyprotein unfolding) that the extension of a folded domain is negligible in comparison with the overall chain extension and the cantilever displacement; The domain has to be much softer to make a noticeable contribution into the overall chain elasticity. Compliance of much softer, unfolded domains should of course be taken into account when considering the chain elasticity.

In the model of Hummer and Szabo[54] the domain is viewed as a linear "molecular spring" that ruptures when its extension exceeds a critical value $R^{\dagger}$. By fitting the experimental data for the I27 domain[21], they have been able to estimate the force constant of their molecular spring; the value $\gamma$=0.9 N/m[54] they report is comparable to $\gamma_{N-C}$

computed here for N-C ubiquitin, which is not surprising since the two domains have very similar mechanical properties[72]. Describing the elasticity of a domain by a single stiffness value is only useful if this value stays nearly constant (or, equivalently, $G(R)$ remains quadratic) in an appreciable range of extensions. For this reason, we do not report below the domain stiffness in cases where nonlinearity of $G(R)$ is significant (cases b and d).

As the force $f$ is increased, the unfolding barrier $\Delta G_u(f) = G_f(R^\dagger) - G_f(R_m)$ becomes smaller and finally disappears for $f_c \simeq 320$ pN, indicating that at forces higher than $f_c$ the domain unfolding would no longer involve barrier crossing and thus would be very fast. The unfolding free energy barrier in the limit $f \to 0$ is $\Delta G_u(0) \approx 29$ kcal/mol.

**Case b**. The free energy increases monotonically as $R$ is increased from ~25Å to ~60Å. The location of the transition state in the presence of a force is determined by the condition $G'(R^\dagger) = f$. According to this equation (and as seen from Fig. IV.3(b)) the transition state is shifted to the left as $f$ is increased. Unlike Case a, where the location of the transition state was virtually force-independent and associated with the separation of two parallel strands, the barrier in Case b does not seem to be associated with any specific molecular event. Examination of snapshots of the domain for different extensions reveals that two antiparallel strands come apart at $R \sim 42$ Å (see Fig. IV.2(b)); However this event is not associated with a well defined free energy barrier in $G_f(R)$.

The barrier between the folded and unfolded states disappears at a force $f_c \simeq 160$ pN, which is much lower than the force at which the domain loses its mechanical stability when the force is applied between its N- and C-termini.

**Case c**. In the case of Lys11-C ubiquitin, the applied force "unzips" the terminal parallel strands (see Fig. IV.2(c)). This process manifests itself as an abrupt change in the slope of $G(R)$ and as a maximum in $G_f(R)$. As seen from Figs IV.2(c) and IV.3(c), the transition state associated with this process is located at $R^\dagger \simeq 40$ Å. The free energy barrier associated with this process at low forces is ~13 kcal/mol, which is much lower than the free energy barrier associated with separation of the same strands under shear in Case a. This finding is further discussed in Section 5.

The overall "stiffness" of the folded domain with respect to the stretching force applied between Lys11 and the C-terminus is lower than in Case a:

$$\gamma_{K11-C} = df / dR_m \approx 0.25 \text{ N/m} \tag{IV.7}$$

The folded domain becomes mechanically unstable (i.e., the folding barrier disappears) for $f \geq f_c \approx 100\,\text{pN}$.

**Case d.** Unzipping of the terminal parallel strands in this case proceeds via an intermediate at $R \simeq 20\text{Å}$ (Fig. IV.2(d)), which is stabilized when a weak force ($f \geq 30\text{pN}$) is applied. The unfolding barrier vanishes for a force $f$ as low as 45 pN (see Fig. IV.3(d)).

## IV. 4 COMPARISON WITH EXPERIMENTAL DATA.

***Average unfolding force***. For N-C, Lys48-C and Lys11-C ubiquitin, we have computed the average unfolding force that would be observed in an AFM experiment, where the domain is incorporated within a polyprotein chain, see Fig. IV.4. In polyubiquitin pulling studies[20], the effective contour length $L$ of the chain depends on the type of linkage and on the number of domains that are already unfolded. Further, in a chain composed of several identical domains the probability of unfolding of one of them is proportional to the number of domains, resulting in an additional dependence of the average unfolding force on the number of unfolded domains[81,125]. These more subtle effects that have nothing to do with the mechanical resistance of individual domains are not considered here: Instead we assume a simpler, hypothetical AFM setup, in which a single ubiquitin domain is incorporated within a chain that has the same, constant contour length $L$ = 58 nm in each of the 3 cases.

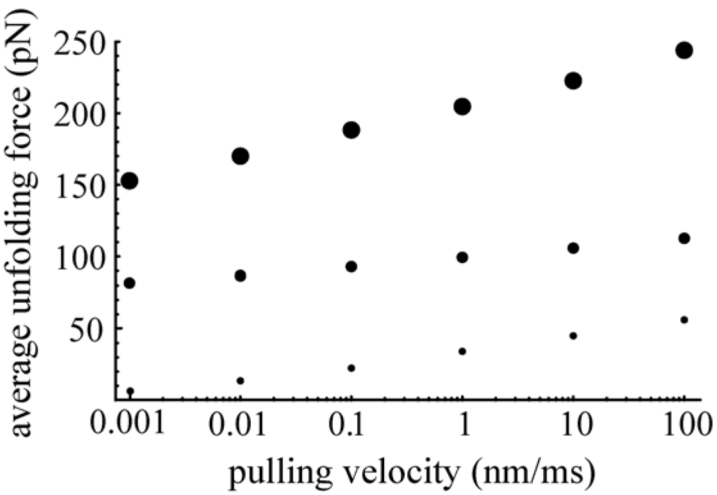


Figure. IV.4. The mean unfolding force $\langle f \rangle$ as a function of the pulling velocity $u$ for N-C ubiquitin (large symbols), Lys48-C ubiquitin (medium symbols) and Lys11-C ubiquitin (small symbols).

For the pulling speed varying within a typical range of $u$ = 0.1 – 10 nm/ms, we predict the average unfolding force $<f>$ for the N-C, Lys48-C, and Lys11-C ubiquitin to be in the range 190-220 pN, 90-110 pN, and 20-45pN, respectively. N-C ubiquitin has the strongest pulling rate dependence of the unfolding force and Lys48-C ubiquitin, the weakest. Based on Fig. IV.3(d), we predict Lys63-N ubiquitin to have very low mechanical stability and to unfold at forces that may not be detectable by AFM. Because of the more complicated shape of $G_f(R)$ in this case, the unfolding mechanism in this case would be more complex and Eqs. IV.2-3 would not be applicable.

The unfolding forces predicted here for the N-C and Lys48-C ubiquitin are close to those observed experimentally.[20] Further, in accord with the experimental data[20], the slope of the curve of $<f>$ vs. $\ln u$ is lower for Lys48-C than it is for the N-C linkage. The absolute value of the slope is however somewhat different, being lower than the experimental slope in each case, possibly in part because we have ignored the variability of the contour length in each case and assumed a single domain in the chain.

Although Lys11-C ubiquitin has not been pulled experimentally, Brockwell et al[13] studied the effect of the pulling geometry on the mechanical unfolding of a different β-sheet protein, E2lip3, and demonstrated that unzipping of the terminal β-strands in this protein takes place at a very low force not detectable by AFM, an observation qualitatively consistent with our prediction of a low unfolding force for the unzipping of two strands in ubiquitin.

***The location of the transition state***. By comparing the experimental data to a kinetic Monte Carlo simulation of mechanical unfolding[21], the authors of ref. [20] have been able to infer the position $\Delta R = R^{\dagger} - R_m$ of the unfolding transition state relative to the folded state. The values of $\Delta R$ estimated in ref. [20] are 2.5Å and 6.3 Å for N-C and Lys48-C ubiquitin, respectively. To estimate $\Delta R$ from experimental data, one customarily assumes that the unfolding rate constant depends exponentially on the pulling force[6] (or, equivalently, that the unfolding free energy barrier depends linearly on the force):

$$k_u(f) = \alpha_0 \exp(f \Delta R / k_B T) . \qquad \text{(IV.8)}$$

In a more recent study[104], the unfolding rate $k_u(f)$ for the N-C ubiquitin has been measured directly by using single molecule force-clamp techniques. Measurements of this type are extremely valuable because they allow one to monitor the unfolding kinetics in real time and to eliminate the potential errors that arise from the need to reconstruct the underlying kinetics from experimental observables such as the mean unfolding force and the unfolding force probability distribution. The authors of ref.[104] have found their data to be reasonably well fitted by Eq. IV.8 in the range $f$ = 70 – 200 pN and, based on this fit, estimated $\Delta R \simeq 1.4 - 1.7$Å, somewhat shorter than in the previous study[20].

Here, we have estimated $\Delta R \approx$ 10Å for N-C ubiquitin at low forces while for Lys48-C ubiquitin we have found $\Delta R$ to have a very strong force dependence and – for low forces – to be considerably larger than that for N-C ubiquitin (see Fig. IV.3(a, b). At first glance, these results appear to contradict the experimental data of refs[20,104]. However we have argued in ref.[71] that this "contradiction" is largely a consequence of the fact that $\Delta R$ inferred from Eq. IV.8 is not the true transition-state extension. It would be if the unfolding barrier were a linear function of the force $f$, $\Delta G_u(f) = \Delta G_u(0) - f\Delta R$ . Eq. IV.8

further implies that the transition state displacement $\Delta R$ is independent of the force. These assumptions are not true for a general function $G(R)$. For $G(R)$ computed here the unfolding rate constant does not satisfy Eq. IV.8 and $\Delta R$ is force dependent. For example, for N-C ubiquitin, $\Delta R \simeq 7$Å for $f = 100$ pN and $\Delta R \simeq 4$Å for $f = 250$ pN (see Fig. IV.3(a)). Because AFM experiments probe a limited range of forces $f$, within this range the unfolding rate force dependence $k_u(f)$ may be indistinguishable from that given by Eq. IV.8 (see Fig. IV.9 in ref.);[71] the *apparent* value of $\Delta R$ can be obtained by linearizing the function $\ln k_u(f)$ and is quite different from the actual transition-state displacement. In ref.[71] we showed for the I27 domain that, although the calculated transition-state displacement $\Delta R \approx 10$Å is 4 times longer than the experimental estimate based on Eq. IV.8, the calculated dependence $k_u(f)$ is close to the experimental result in the *experimentally accessible range of forces*; In this range of forces the apparent value of $\Delta R$ is also much closer to the experimental estimate although of course there still may be a discrepancy because of simulation errors. Similarly, for N-C ubiquitin, our simulation predicts the apparent transition state extension (estimated from the slope of $\ln k_u(f)$ vs. $f$ near $f = 200$ pN) to be $\Delta R_{app} \simeq 4$Å, which is smaller than our estimate $\Delta R \approx 10$Å for the true transition state extension. Still, our $\Delta R_{app}$ is larger than the values reported in both ref. [20]and ref.; [104]Thus our simulation predicts a somewhat stronger force dependence of $k_u(f)$ than that seen experimentally. This remaining discrepancy between our simulation and the experiment may be due to simulation errors.

The computed $k_u(f)$ deviates very strongly from Eq. IV.8 at low forces. This low force regime is discussed in more detail in the next Section.

We also note that the value of $\Delta R$ inferred from the fit of experimental data to a model is strongly model dependent, i.e., sensitive to the specific choice of the shape of $G(R)$. For example, Hummer and Szabo[54] used a harmonic cusp-shaped potential to describe the unfolding of I27. Like in our case, their $k_u(f)$ does not obey Eq. IV.8. While their model reproduces all of the experimental data for titin, their estimate for the transition state displacement, $\Delta R = 4.2Å$, is also noticeably different from that obtained on the basis of Eq. IV.8.

## IV. 5 DISCUSSION

The same two parallel strands show very different mechanical resistance depending on whether they are unzipped (Lys11-C ubiquitin) or slide with respect to each other under shear (N-C ubiquitin). The fact that the unzipping of two strands would require lower force is not surprising and it has been previously confirmed by molecular mechanics simulations[103]. In the case of two strands sliding with respect to one another, the hydrogen bonds between the strands are loaded in parallel, resulting in a high overall stiffness (cf. Eq. IV.6) and a high force[103]. This property was noted as the key to the mechanical stability of titin[11,13,25,71,78,79] and other domains with parallel strands[72].

However if the domain is extended in a truly equilibrium fashion then the free energy cost of separating the two strands in the two different ways should be the same[106]. While the unfolding force in the case of unzipping is lower, the overall displacement (i.e., the change in $R$) is larger and the work done in a quasistatic process of separating the strands should be the same. This expectation is in stark contrast with our finding that the zero-force unfolding free energy barrier for N-C ubiquitin (29 kcal/mol) is much higher than that for Lys11-C ubiquitin (13 kcal/mol). This suggests that the distance between the N- and C-termini is a poor choice of the reaction coordinate for the domain unfolding in the absence of a stretching force. By poor choice we mean that trajectories computed by molecular dynamics in our umbrella sampling procedure may remain trapped in the neighborhood of a “wrong” transition state (i.e. saddle point) on the protein’s multidimensional energy landscape. Although the time duration of molecular dynamics runs used in our simulations is chosen to be long enough that trajectories exhibit stationary behavior, it may be too short for the system to discover other, lower-energy basins of attraction on its multidimensional potential energy surface. Thus our sampling may not be truly ergodic and the domain may be driven along different unfolding pathways depending on the chosen reaction coordinate, leading to different unfolding free energy barriers. Provided that sampling of the folded ensemble does not suffer from this ergodicity problem, our procedure then yields an upper bound on the lowest unfolding free energy barrier. It is therefore not surprising that using Eq. IV.2 with the zero-force unfolding barrier $\Delta G_{\mathrm{u}}(0)$ = 29 kcal/mol estimated for the N-C ubiquitin, one finds the unfolding rate to be $k_{\mathrm{u}}(0) = 5.5\times 10^{-14}\mathrm{s}^{-1}$, ~10 orders of magnitude smaller than the reported chemical unfolding rate constant under native conditions, $k_{\mathrm{chem}} \sim 4.3 \times 10^{-4}$ $\mathrm{s}^{-1}$ [60]. Interestingly, using the free energy barrier found for Lys11-C case, we estimate the zero-force unfolding rate to be $k_{\mathrm{u}}(0) = 3.4 \times 10^{-2}$ $\mathrm{s}^{-1}$ , which is about 2 orders of magnitude higher than the experimental chemical denaturation rate. Given the uncertainties introduced by extrapolation of the chemical unfolding rates to zero denaturant condition[60], possible simulation errors and the uncertainty in the prefactor, the similarity between $k_{\mathrm{u}}(0)$ and $k_{\mathrm{chem}}$ is quite remarkable, suggesting that the mechanical unfolding of ubiquitin via unzipping of its terminal parallel strands may occur via the same transition state as its chemical unfolding.

When a force $f$ is applied, adding the term $-fR$ to the domain energy changes the relative heights of different barriers on the domain's multidimensional potential energy landscape. While the extension $R$ may be a poor unfolding reaction coordinate at $f$=0, for a sufficiently high force $f$ unfolding pathways that involve large extensions $R$ would become favored energetically. This is why computations of the type reported here are found to be in good agreement with AFM experiments[71,72], which typically probe a range of forces high enough to ensure that $R$ is a good reaction coordinate.

# Chapter V

# Mechanical unfolding of segment-swapped protein G dimer[d]

## V. 1 INTRODUCTION

Unique mechanical properties of natural fibers, adhesives, and composites can often be linked to the mechanical response of individual proteins they are composed of[108]. The mechanical properties of a number of individual "load-bearing" protein domains, as well as other domains that have no apparent mechanical function, have been well characterized both experimentally (via single-molecule pulling experiments) and computationally (via molecular dynamics studies), see[9,16,65] and the refs. therein. At the same time, there is a gap between our understanding of the mechanical response of proteins at the single-molecule level and that of the bulk materials they form. Since the latter usually involve complex assemblies of proteins, it is not clear to which extent their overall mechanical properties are controlled by the mechanical stability of individual domains as opposed to inter-chain interactions and the overall architecture of the material. Here, we study the mechanical response of the simplest possible protein complex, a protein dimer, when the C-terminus of one of its subunits and the N-terminus of the other are pulled apart. Being a dimer, in which segments of secondary structure are interchanged between the two monomers, its mechanical properties cannot be simply derived from those of its monomeric units.

We are particularly interested in finding out whether the unique mechanical stability achieved by Nature in some of its load-bearing proteins[10,12,13,20,25,26,71] can be retained when such proteins are assembled non-covalently to form complexes. There are very limited experimental data regarding the mechanical properties of protein complexes in general and dimers in particular[66,105] so we have to rely on simulations. Fortunately, previous studies have shown that simulations can lead to quantitative predictions regarding the mechanical strength of proteins.[55,71-73,78,79,92] Related simulation studies of forced dissociation of adhesion protein complex have recently been reported by others, [3,4]providing a different example of the mechanical response of a protein complex.

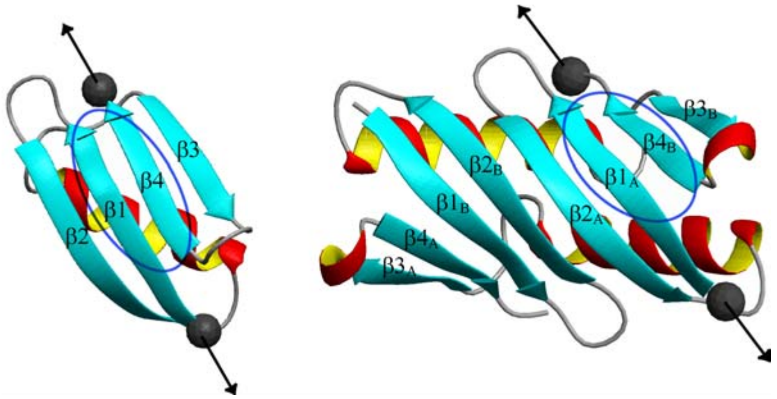
---

[d] Large portions of this chapter have been previously published as reference 74.

The present study may also have implications for several other topics of current interest. Segment swapped dimers are often viewed as models of protein aggregation[77]. While it is now understood[11,13,20,61,71,73,106] that mechanical unfolding experiments do not necessarily probe proteins' thermodynamic stability and their folding and/or chemical/thermal denaturation pathways, they may still offer insight into the stability of protein aggregates. Furthermore, mechanical response of protein aggregates is of interest since it is believed that the cellular machinery responsible for protein degradation unfolds them by pulling them mechanically through narrow constrictions[85,86,97], and since there is evidence that this machinery attempts to unfold protein aggregates[85].

Many naturally occurring protein complexes, or multimers, are assemblies of two to thousands of protein domains. They perform a variety of tasks including mechanical functions (e.g. in the case of collagen or cadherins). In segment-swapped dimers, each domain exchanges part of its secondary structure with neighboring domains.

The specific case of segment swapped dimer studied here is the dimeric mutant (Q10) of the B1 domain of streptococcal protein G[18,35], which has been engineered by introducing mutations in the core amino acids of the B1 domain of streptococcal protein G (GB1).[39] The structure of this dimer (Fig. V.1) involves terminal parallel strands formed by exchanging the terminal β-strands of the two monomeric subunits. If a stretching force is applied between the N-terminus of chain A and the C-terminus of chain B as shown in Fig. V.1(b), the two parallel strands become loaded in a manner similar to that found in a pulling experiment where a single GB1 domain is stretched between its C- and N- termini (as shown in Fig. V.1(a)). It has been recently understood that this type of arrangement of parallel strands is the key structural feature responsible for the high mechanical stability of the I27 domain of the muscle protein titin[11,71,79]. Furthermore, it was predicted theoretically[25,26,71] and established experimentally[12,13,19,20] that other domains (e.g., protein L, ubiquitin), which have no apparent mechanical function but feature the same terminal strand arrangement may exhibit high mechanical stability comparable to that of I27. Therefore if the high mechanical resistance with respect to stretching typical of some individual domains can be retained by protein complexes, the above dimer structure appears to be a good candidate. In what follows, we present a comparative study of the mechanical stability of the protein G dimer stretched between the C-terminus of one subunit and the N-terminus of the other (Fig. V.1(b)) with that of the monomeric protein G.



(a) (b)

Figure V.1. (a) The NMR structure of the Immunoglobulin-binding domain B1 of streptococcal protein G, (PDB code 1GB1). (b) The structure of the segment-swapped dimeric mutant of the B1 domain of streptococcal protein G (PDB code 1Q10). In our simulations, the stretching force is applied between the α-carbons of the terminal residues shown as dark spheres. Labeled elements of secondary structure are the strands β1(residue 2-8), β2 (residue 13-19), β3 (residue 42-46) and β4 (residue 51-55) for the monomer and the strands β1(residue 2-8), β2 (residue 13-19), β3 (residue 43-46) and β4 (residue 51-54) for each monomer of the dimer; the subscript labels the chains A/B. The location of the terminal parallel strands in each protein is circled. The plots were generated by the MOLMOL software[63].

The rest of this chapter is organized as follows: In section V.2 we describe the details of the methods used. Our results are presented in Section V.3 and Section V.4 concludes.

## V. 2 METHODS

Recent studies[71] showed that mechanical unfolding of proteins in many cases can be understood, quantitatively, by using a simple one-dimensional picture, where it is viewed as Langevin dynamics governed by the potential of mean force $G(R)$ that is a function of the "mechanical" reaction coordinate $R$[55]. The latter is usually chosen to be equal to the protein extension, i.e. the distance between the residues (or, more precisely, between their α−carbons) between which the pulling force is applied. The potential of mean force $G(R)$ can be computed using the standard umbrella sampling method[36,87], where a penalty term constraining the extension to a neighborhood of $R$ is added to the system energy; The shape of $G(R)$ is then obtained from a series of equilibrium molecular dynamics (MD) simulations performed with different constraints. By combining the data from such simulations via the weighted histogram method[36], one can recover the global shape of $G(R)$. The technical details of this procedure as applied to protein stretching are described in ref. [72].

Whenever one attempts to calculate an equilibrium property of a protein such as $G(R)$, there is a concern as to whether or not the configuration space of the protein is sampled ergodically within the time-scale constraints of the simulation. Although it impossible to guarantee complete sampling in any simulation, dependence of the simulation result on the length of the simulation would be a clear indication of trouble. In our previous studies of the mechanical stretching of small domains, the results appeared converged with respect to the length of each equilibrium MD simulation[71-73]. However for the protein G dimer we could not obtain converged results with the simulation times feasible given the computer resources available to us. This urged us to resort the replica exchange method (REM), which provides better sampling. The method is reviewed in ref. [36], and its specific implementation in MD simulations has been developed in ref. [110]. REM/MD involves running MD trajectories for a set of $n$ replicas of the system, each at different temperature: $T_i$ ($i$=1, 2,…,$n$), such as $T_1 < T_2 <\ldots< T_n$.

The algorithm further includes Monte Carlo trial moves that attempt to "swap" neighboring replicas. A trial move attempting to swap the configurations $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ of the replicas running at temperature $T_i$ and $T_{i+1}$ is accepted with a probability equal to $\min\{1, \exp[-(\beta_{i+1}-\beta_i)(U(\mathbf{r}_{i+1})-U(\mathbf{r}_i))]\}$, where $U$ is the potential energy of the system and $\beta_i = 1/(k_B T_i)$, where $k_B$ is Boltzmann's constant. The velocities of each atom are rescaled after the trial move such that the new velocities satisfy the equipartition theorem[110]. Such trial moves satisfy detailed balance and help the low-temperature replicas that are trapped near local minima on the system's potential energy landscape overcome barriers separating those from other regions of the configuration space.

Using REM combined with umbrella sampling, we have reconstructed the potential of mean force $G(R)$ in the cases of the stretching of protein G (PDB code 1GB1) and the protein G dimer (PDB code 1Q10). The reaction coordinate $R$ is defined as the distance between the α-carbons of the two terminal residues (Met1 and Glu56) in the case of the protein G monomer and the distance between the α-carbon of the Met1 in the chain A and the α-carbon of the Glu56 in the chain B in the case of the dimer. A harmonic penalty term $\gamma(R-R_0)^2/2$ was added to the energy to constrain the protein extension near a specified value $R_0$ in umbrella sampling simulations[71-73], with the spring constant equal to $\gamma$=1.38 N/m. The values of $R_0$ used were $R_0$ = 26.7, 27.7, 28.7, 29.7, 30.7, 31.7, 32.7 Å for protein G, and $R_0$=25.8, 26.8, 27.8, 28.8, 29.8, 30.8, 31.8 Å for the dimer. For each $R_0$, we have used $n$=10 replicas run at the temperatures (273, 285, 298, 312, 327, 343, 361, 379, 399K, and 422 K). The results reported below are for $T$=298K. Simulations were performed with Tinker[126] software using the GB/SA continuum solvation model[98], and the AMBER99 force field[119]. Both the monomeric and the dimeric protein G remain native-like in the course of a 0.5 ns unconstrained simulation at 298K, using this force field/solvation model. The Rattle algorithm[1] was used to fix all the bond lengths between hydrogen atoms and heavy atoms. The MD time step was 2.0 fs and a replica-swapping move was applied every 1.0 ps of the simulation. Protein configurations were saved every 0.5 ps. The total simulation time for each replica was 0.5 ns for each constraint and each temperature.

## V. 3 RESULTS

In Figure V.2, we show the free energy profile $G(\triangle R)$ as a function of the protein extension for the protein G monomer (GB1) and the dimer (Q10). The protein extension, $\Delta R = R - < R_{fold} >$, is defined as the distance R between the α-carbons of the terminal residues that are being pulled apart relative to the average extension $< R > \equiv < R_{fold} >$ measured in the equilibrium MD simulation in the absence of force. $< R_{fold} >$ is equal to 27.61 and 27.73 Å for GB1 and Q10, respectively. The free energy of protein reconstructed from the REM/MD with the smallest constrain $R_o$ is taken as the reference.

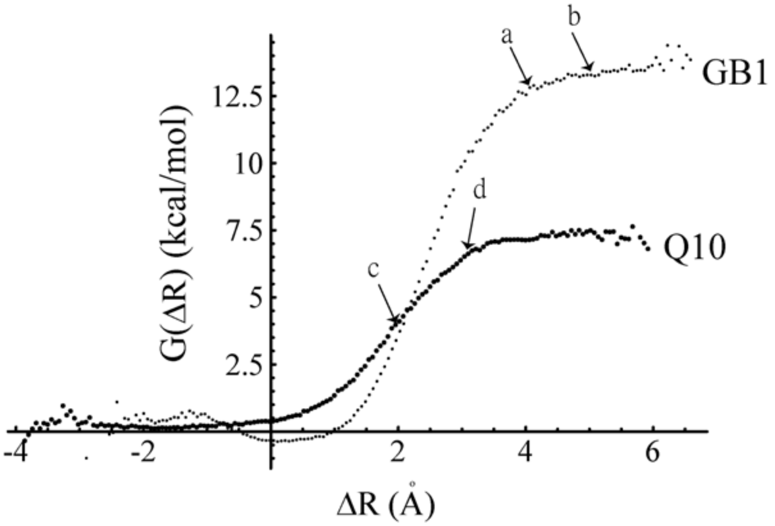


Figure V.2. The free energy of GB1 and Q10 as a function extension. The two arrows near each curve roughly delineate the region where the separation of the terminal parallel strands takes place. Domain configurations corresponding to these values of extension are further analyzed in Figs. V5-6. The positions $\Delta R$ of the arrows are: (a) 4.1 Å, (b) 5.1 Å for GB1 and (c) 2.1 Å, (d) 3.1 Å for Q10.

The negative value of extension shown in the Fig. V.2 is simply due to the fact that $<R_{fold}>$ is larger than the smallest constrain distances $R_o$ (i.e. 26.7 Å and 25.8 Å for GB1 and Q10), which are measured in the crystal structures of GB1 and Q10. The shapes of the two curves are very similar. They are also similar to the free energy profiles found in our previous studies of ubiquitin-like domains and I27[72]. The free energy $G(R)$ rises monotonically until it reaches a plateau value $G_{\mathrm{pl}}$. This value is $G_{\mathrm{pl}} \approx 14$ kcal/mol for the monomer and ≈ 7 kcal/mol for the dimer. The value of $G_{\mathrm{pl}} \approx 14$ kcal/mol is similar to the value $G_{\mathrm{pl}} \approx 17$ kcal/mol found in our previous study of the protein G IgG-binding domain III, which has similar structure.[72]

When the stretching force $f$ applied to the protein is not too high, the unfolding dynamics can be viewed as barrier crossing[6,28,29,99]. Specifically, the potential of mean force experienced by the protein is equal to

$$G_f(R) = G(R) - fR \tag{V.1}$$

This force-dependent free energy profile is shown in Fig. V.3 for both proteins at different values of the stretching force. Application of a force turns the global minimum corresponding to the native fold into a metastable state separated from the extended states with large $R$ by a barrier. The height of this barrier determines the rate constant of mechanical unfolding, which can be estimated by using transition state theory:

$$k_u(f) = \nu \exp(-\Delta G_u / k_B T) \,. \tag{V.2}$$

Here

$$\Delta G_u(f) = G(R^{\neq}) - G(R_N) \tag{V.3}$$

is the free energy barrier, i.e., the difference between $G(R)$ at the minimum $R_N$ corresponding to the native metastable state and the maximum $R^{\neq}$ corresponding to the transition state. The transition state extension $\Delta R^{\neq} = R^{\neq} - R_N$ determines how fast $k_u(f)$ changes as a function of the stretching force. It consequently impacts the pulling rate dependence of the average unfolding force measured in AFM pulling experiments[28,29].

As seen from Fig. V.3, the values of $\Delta R^{\neq}$ are similar for both the dimer and the monomer. For example, for $f$ = 100 pN, $\Delta R^{\neq} \simeq 3$ Å in each case. As a result, the force dependence of the barrier $\Delta G_u(f)$ is similar in the two cases, as shown in Fig. V.4. At a sufficiently high force, the unfolding free energy barrier disappears altogether, rendering the native conformation of the protein mechanically unstable. This critical force required to destabilize the native state of the protein mechanically is on the order of 200 pN for the dimer and ~400 pN for the monomer.

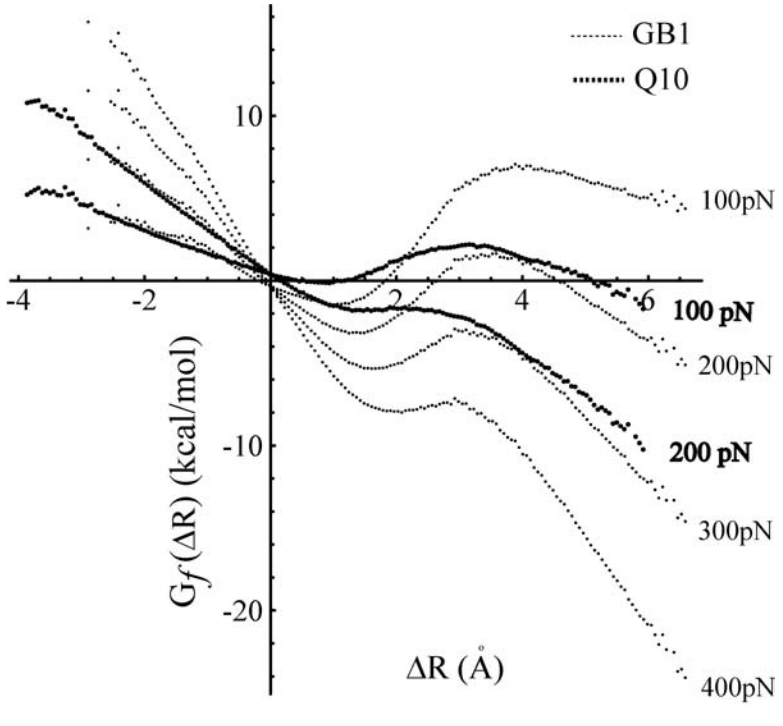


Figure V.3. The free energy $G_f(\Delta R) = G(\Delta R) - f\Delta R$ for different values of the stretching force f.

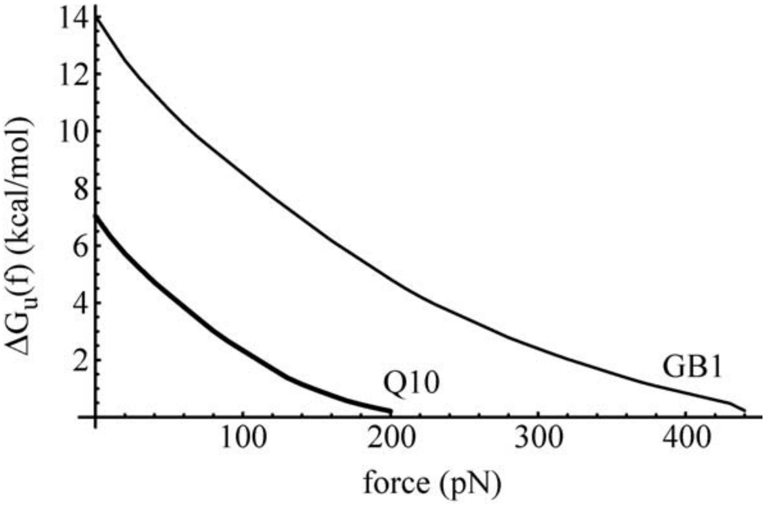


Figure V.4. The unfolding free energy barrier $\Delta G_u(f)$ as a function of the applied force for GB1 and Q10. The barriers were calculated starting from f = 0 pN with an increment of 10 pN; The continuous lines were obtained by interpolation.

To understand the unfolding mechanism better, we next consider the protein configurations at different stages of the unfolding process. A convenient way to represent an ensemble of configurations sampled by a protein is its contact map. For a given configuration we define the contact map as the plot of residue pairs $\{i,j\}$ such that the distance between their respective α-carbon atoms is less than a given cutoff distance *d*. We used the value *d*= 7.5Å here. To see how the protein structure depends on its extension, we have considered equilibrium ensembles of protein structures obtained via the constrained equilibrium REM/MD runs as described in section V.2. For a given $R_0$, for each contacting pair $\{i,j\}$ we then consider the probability of observing this contact in the equilibrium ensemble (i.e., the fraction of configurations for which their α-carbons are found within the distance *d*). We can color-code the $\{i,j\}$ map according to these probabilities. This results in a contact map that describes an ensemble of structures rather than a single structure.

Fig. V.5(a,b) displays the contact maps for the monomeric protein G obtained for two different values of extension $R_0$, which are indicated by arrows in Fig. V.2. It is tempting to identify these as "pre-transition-state" and "post-transition-state" ensembles; However as seen in Fig. V.3, the location of the transition state generally depends on the applied force so the transition-state extension is not well defined. Instead, we simply consider the changes in the protein structure as its extension is increased. Initially, the contact map remains virtually unchanged and similar to the native contact map (not shown). In particular, the contact map shown in Fig. V.5(a) is essentially the native contact map indicating that the protein native structure remains intact upon initial stretching. As the domain extension is increased further, as in Fig. V.5(b), and the free energy $G(R)$ approaches its plateau value $G_{\mathrm{pl}}$, the contacts formed by the terminal parallel strands (encircled in Fig. V.5) become partially destroyed. There is no considerable change in the rest of the contact map. An example of a structure belonging to the partially unfolded ensemble corresponding to Fig. V.5(b) is shown in Fig. 6(a). Note that the entire ensemble cannot be generally represented by this single structure.

Similar behavior is found in the case of the dimer (Fig. V.5(c,d)). In this case the contact map consists of two diagonal blocks corresponding to the contacts among residues within the same monomeric unit and the off-diagonal blocks, which correspond to the contacts formed between different monomers. Notably, the parallel strands $\beta 1$ and $\beta 4$ of protein G monomer become replaced by the parallel strands $\beta 1_{\mathrm{A(B)}}$ and $\beta 4_{\mathrm{B(A)}}$ formed across the monomers within the dimer. In addition to this parallel strand swapping, an additional pair of parallel strands ($\beta 2_{\mathrm{A}}$ and $\beta 2_{\mathrm{B}}$) is formed (Fig. V.1). Similarly to the case of the protein G monomer, as the dimer is stretched initially, the native structure remains intact (case c in Fig. V.5). As $R$ is further increased and $G(R)$ approaches its plateau value, the terminal parallel strands become disrupted as seen in Fig. V.5(d) and Fig. V.6(b).

The above unfolding mechanism proceeding via separation of the terminal parallel stands is similar to that observed in the simulations of I27[71,78,79,84], ubiquitin[72,73], and protein L[12]. This mechanism was predicted to maximize the mechanical strength of both proteins and toy models of cross-linked Gaussian polymers[25,26]

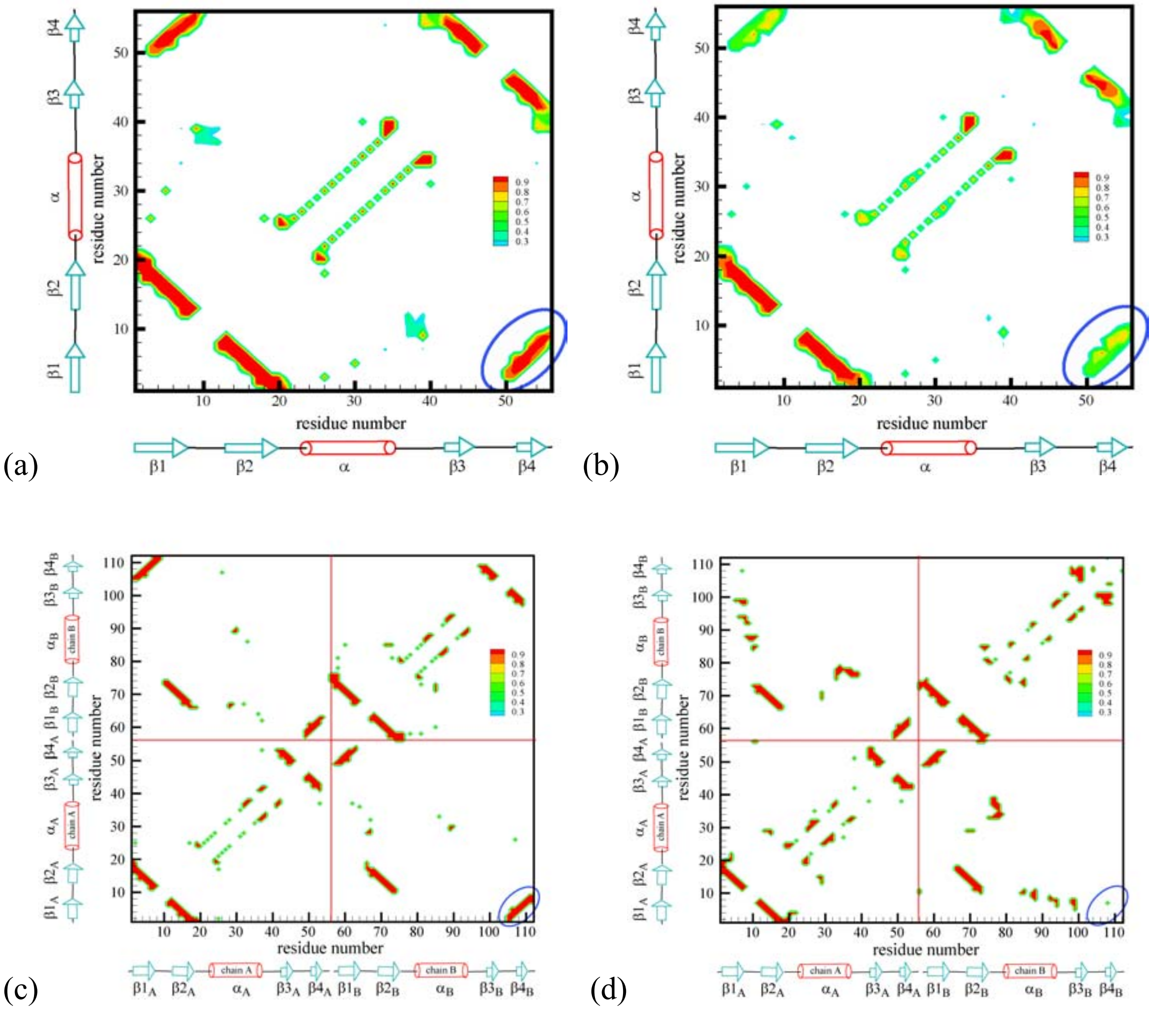
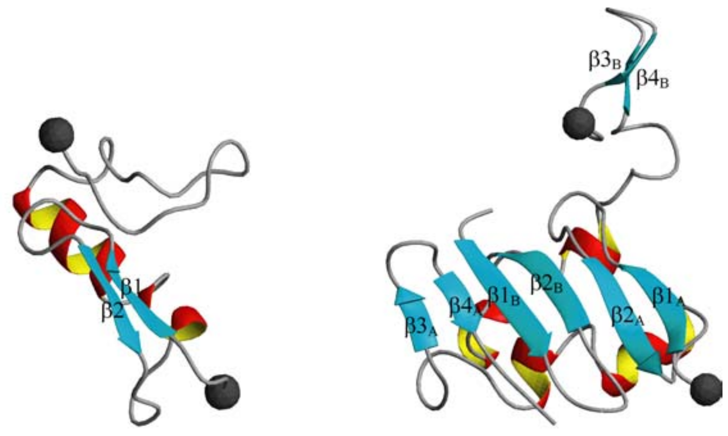


Figure V.5. The contact maps of (a,b) GB1 and (c,d) Q10 obtained from equilibrium MD runs with the protein extension $\Delta R$ constrained near the values indicated as arrows in Fig. V.2. (a) $\Delta R_a \approx 4.1$ Å; (b) $\Delta R_b \approx 5.1$ Å; (c) $\Delta R_c \approx 2.1$ Å; (d) $\Delta R_d \approx 3.1$ Å. The terminal parallel strands shown in Fig. V.1 are indicated by the blue circles on each contact map. See text for further details.



(a) (b)

Figure V.6. Two representative configurations of partially unfolded (a) GB1 and (b) Q10 taken from equilibrium ensembles corresponding to the situations depicted in Fig. V.5(b) and Fig. V.5(d), respectively. The plots were generated by the MOLMOL software[63].

.

We note that the shape of $G(R)$ for extensions much higher than those in Figs. V.2-3 is not readily accessible by molecular dynamics simulations. The problem is that once the protein is unfolded or partially unfolded, the configuration space sampled by it becomes much larger while the time scale of its motion becomes slower. This necessitates much longer simulation times that are no feasible even when enhanced sampling methods such as REM are used. For these reasons, on the basis of our simulations alone we cannot claim that the mechanical unfolding of each of the above proteins is an all-or-none process with a rate limiting step being the separation of terminal strands. However this picture seems to be consistent with the experimental studies of titin[11,21,89,99,100], ubiquitin[20,30,104], protein L[12]. It is further consistent with Langevin dynamics simulations of a minimalist foldable off-lattice model of ubiquitin, which captures the folding time scale and allows one to map out the free energy landscape of this domain far from equilibrium[61].

## V. 4 DISCUSSION AND CONCLUSIONS

All-atom studies of protein unfolding unavoidably have limitations. For example, the energy landscape and the native structure found from equilibrium REM simulations of the monomeric protein G show significant force field dependence[123]. The use of an implicit solvent model imposes another important limitation. When explicit solvent degrees of freedom are eliminated from the system, this results in additional effective solvent-induced forces[7,8,14,24]. These forces are not present in the standard continuum solvation model used here (see, e.g.[45,75]) Solvent-induced forces are known to be important when considering protein-protein interactions and protein folding/unfolding[44]. In particular, correct description of hydrogen bonding interactions within a continuum model requires a very careful treatment of those effects[44,46]. Since the rupturing of the hydrogen bonds between the terminal strands arguably contributes large part of the free energy barrier to unfolding[78,79], the ability of the GB/SA solvation model to provide a quantitative estimate for this barrier should be questioned. An explicit solvent representation would overcome this limitation. Unfortunately its computational cost would be prohibitive for the present case.

Our previous experience with using the same solvation model indicates that this type of calculation correctly reproduces relative mechanical stability of many proteins. In particular, the relative mechanical stabilities of the I27 domain of the muscle protein titin[71], ubiquitin[72], and protein G[72] obtained from our simulations are in close agreement with experiments[12,19]. Moreover, the computed relative mechanical stabilities of the same domain with respect to different ways of pulling it[73] are also in close agreement with experimental studies[20]. We believe that the success of the method is partly due to the fact the mechanical stability of a protein is largely determined by its native topology[25,62] (and pulling geometry), which is why even cruder, minimalist off-lattice models do reasonably good job predicting the mechanical response of proteins[61,120].

With the above caveats in mind, our main findings discussed below still await their experimental verification. Our calculations suggest that the protein G dimer is less resistant to mechanical pulling than its monomeric counterpart. Nevertheless, it exhibits considerable mechanical stability. Moreover, the mechanical unfolding mechanism for this dimer resembles closely that observed in the I27 domain of the muscle protein titin, ubiquitin, and the protein G monomer. All these proteins share the same structural feature, the presence of terminal parallel strands, which is responsible for their high mechanical resistance with respect to the stretching between their C- and N-termini. The mechanical unfolding pathway then involves the separation of the parallel strands accompanied by the breaking of several hydrogen bonds. As those bonds are loaded in parallel, their concerted rupture requires a high force.

The significant strength of the "clamp" formed by the terminal parallel strands results in the rest of the protein structure being "protected" from the stretching force until the strands become separated. In other words, no domain unfolding (or partial unfolding) occurs prior to the separation of the terminal strands. Interestingly, similar behavior was found in a recent simulation study of the forced dissociation of an adhesion protein complex in the regime where the applied force was ramped slowly enough[3].

It is not inconceivable that the mechanical response of natural fibers, adhesives and composites could be optimized both at the single chain level and at the mesoscopic level through interactions among individual chains.

In contrast to the plethora of experimental data on the mechanical response of individual protein domains, single molecule pulling studies of non-covalently bound protein complexes have begun only recently[66,105]. Although recent AFM pulling studies of spider silk[5] and other natural materials[108], in principle, locally probe one or several polypeptide chains that are non-covalently crosslinked with other proteins, the precise system that is being mechanically loaded and the manner in which the loading is done is not well defined (and is likely to vary from pull to pull). Schwaiger et al[105] have studied the mechanical response of the much better characterized ddFLN dimer. While their study mainly focused on the unfolding of individual domains within each monomer, they point out that the dimer bond is strong enough to stay intact while those domains are unfolding. They further estimate the lower bound on the average dimer separation force to be ~200pN at the pulling rate used in the experiments. Unfortunately, direct measurement of the dimer bond strength was precluded by the fact that it is hard to distinguish separation of the two monomers from desorption of the entire molecule from ether the AFM tip or the substrate. We hope that computational studies like this one will provide estimates of the unfolding forces that could be anticipated in such experiments and stimulate future experimental studies.

# Chapter VI

# Summary

In this dissertation, we have developed computational methodology that allows one to predict the mechanical resistance of proteins at loading rates that are comparable with those in single molecule pulling experiments. The computational procedure involves calculating the free energy profile of a protein along the mechanical unfolding coordinate from a series of equilibrium molecular dynamics simulations by using umbrella sampling/weighted histogram analysis. Our approach assumes that the mechanical unfolding can be viewed as one-dimensional Langevin dynamics governed by this free energy profile and it has been validated by comparing the computed Langevin dynamics with the results of steered molecular dynamics simulations. With the calculated free energy profile, transition state theory is used to calculate force-dependent mechanical unfolding rates, and kinetic Monte Carlo simulations are subsequently performed to predict the outcome of single molecule AFM pulling experiments. To test our approach, we have studied the mechanical unfolding of the I27 domain in the muscle protein titin. The predicted average unfolding force as well as its dependence on the pulling rate is found to be in agreement with AFM experiments. Moreover, the simulated unfolding free energy barrier for the I27 domain is very close to that inferred from experiments. In addition, we have confirmed the earlier observation that the unfolding mechanism involves the separation of two terminal parallel strands.[71,78,79,84]

In order to understand the relationships between structure and the mechanical resistance of the proteins, we have studied the mechanical unfolding of two non-mechanical proteins, ubiquitin and protein G IgG-binding domain III, which both have the structural motif involving terminal parallel strands, similar to those found in the I27 domain of the muscle protein titin. We found that these two proteins also exhibit high force resistance comparable to that of titin. Our finding suggests that non-mechanical proteins with the desired structure and topology can have high mechanical stability, an observation that has indeed been confirmed in recent experimental studies.[12,19]

We have further examined the effect of the pulling geometry on the mechanical unfolding rates and demonstrated that the mechanical unfolding pathways can indeed be altered by changing the pulling geometry. These are in close agreement with single molecule pulling studies of polyubiquitin chains.

While the mechanical stability of individual proteins has been linked to the overall mechanical strength and toughness of the bulk materials they comprise, the effect of the inter-chain interactions is poorly understood. As a step towards the understanding of those, we have studied the mechanical unfolding and separation of the segment-swapped protein G dimer and compared its mechanical stability with that of the single protein G domain. While the dimer is found to be less resistant to mechanical unfolding than its monomeric counterpart, the two proteins exhibit the same mechanical unfolding mechanism involving separation of the terminal parallel strands. Our results suggest that the mechanical properties of natural materials may be optimized not only at a single molecule level but also at the mesoscopic level through the interactions among individual chains.

Our simulations confirm the hypothesis that the mechanical stability of individual protein domains can be maximized by the presence of a hydrogen-bond clamp between the terminal parallel strands[71,78,79,84]. Since this motif is found in a number of proteins, both mechanical and non-mechanical, we expect that our finding may prove useful in the area of functional protein engineering to identify suitable building blocks for new materials with superior mechanical properties.

# References

(1) Andersen, H. C. *Journal of Computational Physics* **1983**, *52*, 24-34.

(2) Balsera, M.; Stepaniants, S.; Izrailev, S.; Oono, Y.; Schulten, K. *Biophysical Journal* **1997**, *73*, 1281-1287.

(3) Bayas, M. V.; Schulten, K.; Leckband, D. *Biophysical Journal* **2003**, *84*, 2223-2233.

(4) Bayas, M. V.; Schulten, K.; Leckband, D. *Mech. Chem. Biosys* **2004**, *1*, 101-111.

(5) Becker, N.; Oroudjev, E.; Mutz, S.; Cleveland, J. P.; Hansma, P. K.; Hayashi, C. Y.; Makarov, D. E.; Hansma, H. G. *Nature Materials* **2003**, *2*, 278.

(6) Bell, G. I. *Science* **1978**, *200*, 618-627.

(7) Ben-Naim, A. *Journal of Physical Chemistry* **1990**, *94*, 6893 - 6895.

(8) Ben-Naim, A. *The Journal of Chemical Physics* **1990**, *93*, 8196-8210.

(9) Best, R. B.; Brockwell, D. J.; Toca-Herrera, J. L.; Blake, A. W.; Smith, D. A.; Radford, S. E.; Clarke, J. *Analytica Chimica Acta* **2003**, *479*, 87-105.

(10) Best, R. B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. *Biophysical Journal* **2001**, *81*, 2344.

(11) Brockwell, D. J.; Beddard, G. S.; Clarkson, J.; Zinober, R. C.; Blake, A.; Trinick, J.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophysical Journal* **2002**, *83*, 458.

(12) Brockwell, D. J.; Beddard, G. S.; Paci, E.; West, D. K.; Olmsted, P. D.; Smith, D. A. M.; Radford, S. E. *Biophysical Journal* **2005**, *89*, 506 - 519.

(13) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nature Structural Biology* **2003**, *10*, 731.

(14) Bruge, F.; Fornili, S. L.; Malenkov, G. G.; Palma-Vittorelli, M. B.; Palma, M. U. *Chemical Physcis Letters* **1996**, *254*, 283-291.

(15) Bryant, Z.; Pande, V. S.; Rokhsar, D. S. *Biophysical Journal* **2000**, *78*, 584.

(16) Bustamante, C.; Chemla, Y. R.; Forde, N. R.; Izhaky, D. *Ann. Rev. Biochem.* **2004**, *73*, 705-748.

(17) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science* **1994**, *271*, 1599.

(18) Byeon, I.-J. L.; Louis, J. M.; Gronenborn, A. M. *J Mol Biol.* **2003**, *333*, 141-152.

(19) Cao, Y.; Lam, C.; Wang, M.; Li, H. *Angew. Chem. Int. Ed* **2006**, *45*, 642-645.

(20) Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez., J. M. *Nature Structural Biology* **2003**, *10*, 738-743.

(21) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marsalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci USA* **1999**, *96*, 3694-3699.

(22) Chan, N.-L.; Hill, C. P. *Nat Struct Biol* **2001**, *8*, 650.

(23) Chau, V.; Tobias, J. W.; Bachmair, A.; Marriott, D.; Ecker, D. J.; Gonda, D. K.; Varshavsky, A. *Science* **1989**, *243*, 1576-1583.

(24) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198 - 2202.

(25) Eom, K.; Li, P.-C.; Makarov, D. E.; Rodin, G. J. *J. Phys. Chem. B* **2003**, *107*, 8730.

(26) Eom, K.; Makarov, D. E.; Rodin, G. J. *Phys. Rev. E.* **2005**, *71*, 021904.

(27) Erickson, H. P. *Science* **1997**, *276*, 1090.

(28) Evans, E.; Ritchie, K. *Biophysical Journal* **1997**, *72*, 1541-1555.

(29) Evans, E.; Ritchie, K. *Biophysical Journal* **1999**, *76*, 2439.

(30) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674-8.

(31) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Letters* **1989**, *63*, 1195.

(32) Fisher, T. E.; Marsalek, P. E.; Fernandez, J. M. *Nature Structural Biology* **2000**, *7*, 719.

(33) Fisher, T. E.; Oberhauser, A. F.; Vezquez, M. C.; Marsalek, P. E.; Fernandez, J. *TIBS* **1999**, *24*, 379.

(34) Fixman, M.; Kovac, J. *J. Chem. Phys.* **1973**, *58*, 1564.

(35) Frank, M. K.; Dyda, F.; Dobrodumov, A.; Gronenborn, A. M. *Nature Structural Biology* **2002**, *9*, 877 - 885.

(36) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, secon edition ed.; Academic Press: San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo, 2002.

(37) Galan, J.-M.; Haguenauer-Tsapis, R. *The EMBO Journal* **1997**, *16*, 5847-5854.

(38) Galligan, E.; Roberts, C. J.; Davies, M. C.; Tendler, S. J. B.; Williams, P. M. *J. Chem. Phys.* **2001**, *114*, 3208-3214.

(39) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657-61.

(40) Grubmuller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997-999.

(41) Hanggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251.

(42) Hansma, H. G.; Pietrasanta, L. *Current Opinion in Chemical Biology* **1998**, *2*, 579.

(43) Hansma, H. G.; Pietrasanta, L. I.; Auerbach, I. D.; Sorenson, C.; Golan, R.; Holden, P. A. *Journal of Biomaterials Science. Polymer Edition* **2000**, *11*, 675.

(44) Hassan, S. A. *J. Phys. Chem. B* **2004**, *108*, 19501 -19509.

(45) Hassan, S. A. *J. Phys. Chem. B* **2005**, *109*, 21989-21996.

(46) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6478 -6489.

(47) Hayashi, C. Y.; Lewis, R. V. *J. Mol. Biol.* **1998**, *275*, 773.

(48) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *International Journal of Biological Macromolecules* **1999**, *24*, 271.

(49) Heymann, B.; Grubmuller, H. *Chemical Physics Letters* **1999**, *303*, 1-9.

(50) Heymann, B.; Grubmuller, H. *Biophysical Journal* **2001**, *81*, 1295-1313.

(51) Hinman, M. B.; Jones, J. A.; Lewis, R. V. *Trends in Biotechnology* **2000**, *18*, 374.

(52) Hofmann, R. M.; Pickart, C. M. *Cell* **1999**, *96*, 645-653.

(53) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3658.

(54) Hummer, G.; Szabo, A. *Biophysical Journal* **2003**, *85*, 5.

(55) Isralewitz, B.; Gao, M.; Schulten, K. *Current Opinion in Structural Biology* **2001**, *11*, 224-230.

(56) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophysical Journal* **1997**, *72*, 1568-1581.

(57) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.

(58) Jarzynski, C. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3636.

(59) Kellermayer, M. S. Z.; Smith, S. B.; Granzier, H. L.; Bustamante, C. *Science* **1997**, *276*, 1112.

(60) Khorasanizadeh, S.; Peters, I. D.; Butt, T. R.; Roder, H. *Biochemistry* **1993**, *32*, 7054-7063.

(61) Kirmizialtin, S.; Huang, L.; Makarov, D. E. *J. Chem. Phys* **2005**, *122*, 234915.

(62) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7254.

(63) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51-55.

(64) Kratky, O.; Porod, G. *Revc. Trav.Chim.Pays-Bas.* **1949**, *68*, 1106.

(65) Lavery, R.; Lebrun, A.; Allemand, J.-F.; Bensimon, D.; Croquette, V. *J. Phys. : Condens. Matter* **2002**, *14*, R383 - R414.

(66) Law, R.; Carl, P.; Harper, S.; Dalhaimer, P.; Speicher, D. W.; Discher, D. E. *Biophysical Journal* **2004**, *84*, 533-544.

(67) Lenne, P. F.; Raae, A. J.; Altmann, S. M.; Saraste, M.; Horber, J. K. *FEBS Lett.* **2000**, *476*, 124-128.

(68) Li, F.-Y.; Yuan, J.-M.; Mou, C.-Y. *Physical Review E* **2001**, *63*, 021905-1.

(69) Li, H.; Carrion-Vazquez, M.; Oberhauser, A. F.; Marsalek, P. E.; Fernandez, J. M. *Nature Structural Biology* **2000**, *7*, 1117.

(70) Li, H.; Oberhauser, A. F.; Fowler, S. B.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6527.

(71) Li, P.-C.; Makarov, D. E. *J. Chem. Phys.* **2003**, *119*, 9260.

(72) Li, P.-C.; Makarov, D. E. *J. Phys. Chem. B* **2004**, *108*, 745-749.

(73) Li, P.-C.; Makarov, D. E. *J. Chem. Phys* **2004**, *121*, 4826.

(74) Li, P.-C.; Huang, L.; Makarov, D. E. *J. Phys. Chem. B* **2006**, accepted.

(75) Li, X.; Hassan, S. A.; Mehler, E. L. *Proteins* **2005**, *60*, 464-484.

(76) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Science* **2002**, *296*, 1832.

(77) Liu, Y.; Eisenberg, D. *Protein Science* **2002**, *11*, 1285-1299.

(78) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophysical Journal* **1998**, *75*, 662.

(79) Lu, H.; Schulten, K. *Chemical Physics* **1999**, *247*, 141.

(80) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; W. E. Reiher, I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(81) Makarov, D. E.; Hansma, P. K.; Metiu, H. *J. Chem. Physics* **2001**, *114*, 9663.

(82) Makarov, D. E.; Wang, Z.; Thompson, J.; Hansma, H. G. *J. Chem. Phys.* **2002**, *116*, 7760.

(83) Marko, J. F.; Siggia, E. *Macromolecules* **1995**, *28*, 209.

(84) Marsalek, P. E.; Lu, H.; Li, H.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. *Nature* **1999**, *402*, 100.

(85) Matouschek, A. *Current Opinion in Structural Biology* **2003**, *13*, 98-109.

(86) Matouschek, A.; Bustamante, C. *Nature Struct. Biol* **2003**, *10*, 674-6.

(87) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Canbridge University Press: Cambridge, 1987.

(88) Oberhauser, A. F.; Badilla-Fernandez, C.; Carrion-Vazquez, M.; Fernandez, J. M. *J. Mol. Biol.* **2002**, *319*, 433.

(89) Oberhauser, A. F.; Hansma, P. K.; Carrion-Vazquez, M.; Fernandez, J. M. *Proc. Natl. Acad. Sci USA* **2001**, *98*, 468-472.

(90) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H.; Fernandez, J. M. *Nature* **1998**, *393*, 181.

(91) Oroudjev, E.; Soares, J.; Arcdiacono, S.; Thompson, J. B.; Fossey, J. B.; Hansma, H. G. *PNAS* **2002**, *99*.

(92) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441-459.

(93) Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. *Nature Biotechnology* **2003**, *21*, 921-926.

(94) Pennisi, M. E. *Science* **1999**, *283*, 168.

(95) Pickart, C. M. *Trends in Biochemical Science* **2000**, *25*, 544-548.

(96) Pickart, C. M. *Annu. Rev. Biochem.* **2001**, *70*, 503-533.

(97) Prakash, S.; Matouschek, A. *Trends Biochem. Sci* **2004**, *29*, 593-600.

(98) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.

(99) Rief, M.; Fernandez, J. M.; Gaub, H. E. *Phys. Rev. Lett.* **1998**, *81*, 4764.

(100) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109-1112.

(101) Rief, M.; Pascual, J.; Saraste, M.; Gaub, H. E. *Journal of molecular biology* **1999**, *286*, 553-561.

(102) Ritort, F.; Bustamante, C.; Tinoco, I. *Proc. Natl. Acad. Sci USA* **2002**, *99*, 13544-13548.

(103) Rohs, R.; Etchebest, C.; Lavery, R. *Biophysical Journal* **1999**, *76*, 2760-2768.

(104) Schlierf, M.; Li, H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7299.

(105) Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat Struct Mol Biol.* **2004**, *11*, 81-85.

(106) Schwaiger, I.; Sattler, C.; Hostetter, D. R.; Rief, M. *Nature Materials* **2002**, *1*, 232.

(107) Shen, T.; Canino, L. S.; McCammon, J. A. *Phys. Rev. Lett.* **2002**, *89*, 068103-1.

(108) Smith, B. L.; Schaffer, T. E.; Viani, M.; Thompson, J. B.; Frederick, N. A.; Kindt, J.; Belcher, A.; Stucky, G. D.; Morse, D. E.; Hansma, P. K. *Nature* **1999**, *399*, 761.

(109) Spence, J.; Gali, R. R.; Dittmar, G.; Sherman, F.; Karin, M.; Finley, D. *Cell* **2000**, *102*, 67-76.

(110) Sugita, Y.; Okamoto, Y. *CHEMICAL PHYSICS LETTERS* **1999**, *314*, 141-151.

(111) Thompson, J. B.; Hansma, H. G.; Hansma, P. K.; Plaxco, K. W. *Journal of Molecular biology* **2002**, *322*, 645-652.

(112) Thompson, J. B.; Kindt, J. H.; Drake, B.; Hansma, H. G.; Morse, D. E.; Hansma, P. K. *Nature* **2001**, *414*, 773.

(113) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771-12800.

(114) Tskhovrebova, L.; Trinic, J. A.; Sleep, J. A.; Simmons, R. M. *Nature* **1997**, *387*, 308.

(115) Viani, M. B.; Pietrasanta, L. I.; Thompson, J. B.; Chand, A.; Gebeshuber, I. C.; Kindt, J. H.; Richter, M.; Hansma, H. G.; Hansma, P. K. *Nat Struct Biol* **2000**, *7*, 644.

(116) Viani, M. B.; Schaffer, T. E.; Chand, A.; Rief, M.; Gaub, H. E.; Hansma, P. K. *Journal of Applied Physics* **1999**, *86*, 2258-2262.

(117) Viani, M. B.; Schaffer, T. E.; Paloczi, G. T.; Pietrasanta, I.; Smith, B. L.; Thompson, J. B.; Richter, M.; Rief, M.; Gaub, H. E.; Plaxco, K. W.; Cleland, A.

N.; Hansma, H. G.; Hansma, P. K. *Review of Scientific Instruments* **1999**, *70*, 4300.

(118) Vollrath, F. *Scientific American* **1992**, *March*, 70.

(119) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(120) West, D. K.; Brockwell, D. J.; Olmsted, P. D.; Radford, S. E.; Paci, E. *Biophys. J.* **2006**, *90*, 287-297.

(121) Williams, P. M.; Fowler, S. B.; Best, R. B.; Toca-Herrera, J. L.; Scott, K. A.; Steward, A.; Clarke, J. *Nature* **2003**, *422*, 446.

(122) Yang, G.; Cecconi, C.; Baase, W. A.; Vetter, I. R.; Breyer, W. A.; Haack, J. A.; Matthews, B. W.; Dahlquist, F. W.; Bustamante, C. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 139.

(123) Zhou, R. *Proteins* **2003**, *53*, 148 - 161.

(124) Zhuang, X.; Rief, M. *Current Opinion in Structural Biology* **2003**, *13*, 88-97.

(125) Zinober, R. C.; Brockwell, D. J.; Beddard, G. S.; Blake, A. W.; Olmsted, P. D.; Radford, S. E.; Smith, D. A. *Protein Science* **2002**, *11*, 2759-2765.

(126) Tinker: http://dasher.wustl.edu/tinker/

# Vita

Pai-Chi Li was born in Taipei, Taiwan, on November 25, 1976, the daughter of Tsung-Ming Li and Shiou-Shing Lin. After completing her work at National Taichung girls' senior high school, Taichuang, Taiwan, she entered National Taiwan University, Taipei, Taiwan in 1996. When she completed her work for Bachelor's degree in Chemistry in 1999, she received the Outstanding Undergraduate Award from the College of Science, National Taiwan University and the Research Creativity Award for undergraduate from the National Science Council of Taiwan. In 2001, she received her Master degree in Chemistry from National Taiwan University. In September 2002, she entered the department of Chemistry and Biochemistry, University of Texas at Austin as a Ph.D student. She was awarded Ph.D degree in August 2006. During her PhD study, she published five papers, which are part of this dissertation, in the Journal of Physical Chemistry B, and the Journal of Chemical Physics.

Permanent address: No.1, Shueiwei Ln, Shiashi, Shioushuei, Changhua, Taiwan 504
This dissertation was typed by Pai-Chi Li.