FORCE-INDUCED TOPOLOGICAL CHANGES OF PROTEINS STUDIED AT THE SINGLE-MOLECULE LEVEL

by

Chengzhi He

B.Sc., Jilin University, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2015

Abstract

The knotted polypeptide chain is one of the most surprising topological features found in certain proteins. Understanding how knotted proteins overcome the topological difficulty during the folding process has become a challenging problem. Theoretical studies suggest that a slipknotted structure can serve as an important intermediate in the transformation from unknotted structure to a knotted structure. This thesis mainly focuses on the mechanical folding/unfolding of a slipknotted protein, the ORF109 of the *Acidianus Filamentous Virus* 3 (AFV3-109), using single-molecule force spectroscopy (SMFS). To show the power of atomic force microscopy (AFM) in the study of SMFS, an α/β protein, NuG2, is stretched and relaxed at a single-molecule level using AFM and the mechanical unfolding and folding events are directly observed. By applying force onto AFV3-109 in different directions, we are able to untie the slipknot to a linear polypeptide chain, as well as tighten it into a trefoil knot involving $\sim$ 13 amino acid residues. Multiple pathways of untying and tightening are found by both SMFS experiments and Steered Molecular Dynamics (SMD) simulations, revealing that the kinetic partitioning mechanism governs the unfolding of the slipknotted protein. In addition, SMD simulations provide detailed molecular mechanisms of the unfolding of the protein and the topological changes from a slipknot to a linear chain, as well as from a slipknot to a trefoil knot. Moreover, the mechanical folding of AFV3-109 is directly observed using optical tweezers, providing new insight into the folding mechanism of knotted/slipknotted proteins.

Preface

Chapter 2 has been published as [He CZ], Hu CG, Hu XD, Hu XT, Xiao A, Perkins TT & Li HB. Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angew. Chem.*, **2015**, 127: 10059–10063. This chapter is incorporated in this thesis with permission of John Wiley and Sons, Copyright (2015). My supervisor Dr. Li HB and I designed the project together. I conducted all the experiments with the help from Xiao A. I analyzed all the data. Hu CG, Hu XD, Hu XT and Perkins TT provided insightful feedback on experimental design and data analysis. I wrote the manuscript together with Dr. Li HB.

Chapter 3 has been published as [He CZ], Genchev GZ, Lu H, & Li HB Mechanically Untying a Protein Slipknot: Multiple Pathways Revealed by Force Spectroscopy and Steered Molecular Dynamics Simulations. *J. Am. Chem. Soc.*, **2012**, 134 (25), 10428–10435. This chapter is incorporated in this thesis with permission of American Chemical Society, Copyright (2012). I designed the project together with Dr. Li HB. I conducted all the AFM experiments and analyzed all the data from the AFM experiments. I did part of simulations. Genchev conducted most of simulations and analyzed the simulations results. I wrote the manuscript together with Dr. Li HB. The part on the simulation results was partially written by Genchev GZ and Lu H and edited by me and Dr. Li HB.

Chapter 4 has been published as [He CZ], Lamour G, Xiao A, Gsponer J, & Li HB. Mechanically Tightening a Protein Slipknot into a Trefoil Knot. *J. Am. Chem. Soc.*, **2014**, 136(34), 11946-11955. This chapter is incorporated in this thesis with permission of American Chemical Society, Copyright (2014). I designed the project together with Dr. Li HB. I conducted all the AFM experiments with help from Xiao A. I did all the simulations with help

from Lamour G and Gsponer J. I analyzed all the results. I wrote the manuscript together with Dr. Li HB.

A version of Chapter 5 is in preparation for publication. [He CZ], Lyu CM, Hu CG, Hu XD, Hu XT, Zhang XH & Li HB. Tying a Protein Slipknot against Force. I designed the project together with Dr. Li HB. Lyu CM and I conducted all the experiments while Zhang XH helped with the preparation of the DNA handles. I analyzed all the data. Hu CG, Hu XD and Hu XT provided insightful feedback on experimental design and data analysis. I wrote this chapter.

Table of Contents

List of Tables

List of Figures

List of Symbols

α_0	unfolding rate at zero force
$\alpha(F)$	unfolding rate at the force of F
β_0	folding rate at zero force
$\beta(F)$	folding rate at the force of F
ΔG	Gibbs free energy
Δx_{u}	unfolding distance, the distance between native and transition state
Δx_{f}	folding distance, the distance between unfolded and transition state
ΔL_{C}	contour length change
F	force
I	intermediate state of protein
k_{B}	Boltzmann constant
k_{C}	spring constant
L_0	distance between the N- and C- termini of a protein in native state
L_{C}	contour length
N	native state of protein
T	temperature in Kelvin
TS	transition state
U	unfolded state of protein

List of Abbreviations

AFM	atomic force microscopy
AFV3-109	the ORF109 of the *Acidianus Filamentous Virus 3*
CD	circular dichroism
CFT	Crooks fluctuation theorem
DNA	deoxyribonucleic acid
DTT	dithiothreitol
FEC	force-extension curve
GB1	the B1 binding domain of protein G from *streptococcus*
GdmCl	guanidine hydrochloride
MC	Monte Carlo
NuG2	a designed variant of protein G
PBS	phosphate buffer saline
PDB	protein data bank
PCR	polymerase chain reaction
RNA	ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMD	steered molecular dynamics
SMFS	single-molecule force spectroscopy
WLC	worm-like chain

Acknowledgements

I would like to thank my supervisor, Dr. Hongbin Li, for his supervision and support. His enthusiasm for science greatly encouraged me when I was having a difficult time. I benefit a lot from his support in many ways. I very much appreciate the experience of studying in his group. Thank Dr. Suzana K. Straus, Dr. Alex Wang, Dr. Lawrence McIntosh and Dr. Elliott Burnell for being on my supervisory committee. I especially thank Dr. Suzana K. Straus for her critical review of the first draft of this thesis.

I would like to thank all the members in Li's group. Thank Dr. Qing Peng for teaching me the molecular biology and AFM experiments. Thank Dr. Yi Cao for the inspiring discussion about scientific research and career. Thank Dr. Shuling Zhuang for teaching me the molecular dynamics simulations. Thank Dr. Peng Zheng, Dr. Shanshan Lv, Dr. Tianjia Bu, Tao Shen, Dr. Jie Fang, Dr. Guillaume Lamour, Na Kong, Kailing Zhou, Jinliang Li, Dr. Yanyan Wang, Adam Xiao, Yabin Guo, Hai Lei, Bryce Laver, Han Wang and Linglan Fu for their help and friendship.

I would also like to thank Dr. Erkai Liu for being a best friend of mine.

Thank Qian Tao, Dr. Jinxin Yi, Yongchun Bu, Peng Zhang, Jie Liu, and Wei Liu for their concern and support.

Last but not least, thank my parents for their love and support.

Dedication

给

曾经的你

曾梦想仗剑走天涯

看一看世界的繁华

年少的心总有些轻狂

如今你四海为家

Chapter 1: Introduction

Biomacromolecules, including proteins, nucleic acids, lipids and polysaccharides, play their specific roles in different biological process. The structure of a molecule determines its function. Among these biomacromolecules, proteins have the broadest scope of functions due to their structural diversity. To effectively carry out their biological functions, most proteins fold spontaneously from linear polypeptide chains into their unique and well-defined three-dimensional structures in a physiological environment (*in vivo*) and even *in vitro*. Therefore, it is essential to understand how proteins fold into their native structures. In the last a few decades, efforts from both experimental and theoretical studies have not only greatly expanded the understanding of protein folding/unfolding, but raised a lot of new interesting questions in this field as well.[1-8] One of the most surprising findings is the knotted/slipknotted topology of the backbone of proteins, which was previously considered almost impossible because of the topological constraints imposed on such a conformation.[9-12] However, bioinformatics has suggested that, actually, many proteins might have knotted backbones and the knot type also varies.[10-20] The proteins with knotted/slipknotted topology and the recent studies on these proteins will be briefly introduced in this chapter.

1.1 Structure and Folding of Protein

Proteins are a type of biomacromolecule that are essential in almost all biological functions. The variety of protein structures determines the wide range of biological functions.[21,22] Thus, it is critical to understand the structure of a protein and the protein folding mechanism.

1.1.1 Protein Structure

Most proteins possess well-defined three dimensional structures. Four levels of structure are used to describe a protein architecture (Figure 1.1). The primary structure of a protein is

defined as the sequence of the amino acid residues within the main chain of the polypeptide. The primary structure determines the final three-dimensional structure of proteins.[23] Interactions between amino acid residues within the protein lead to regular patterns of local structure such as α-helix and β-sheet, which are the secondary structures of proteins.[24] The arrangement of multiple secondary structures result in the global three dimensional structure of a folded polypeptide, which is defined as the tertiary structure.[25] Many proteins are composed of multiple tertiary structures which are arranged in a particular manner. The quaternary structure refers to the three-dimensional structure of the complex of these tertiary structures.[26]

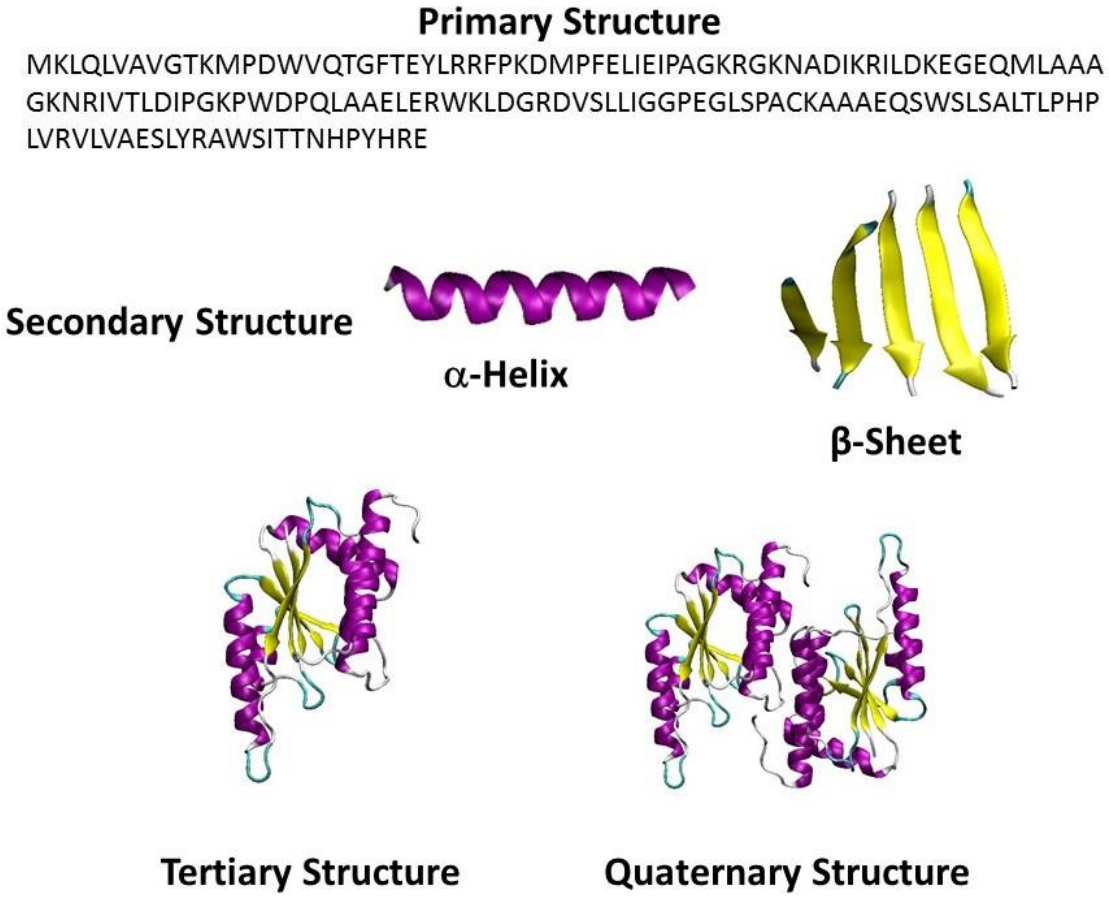

Figure 1.1 Four levels of protein structures.

α-helix and β-sheet are colored in purple and yellow, respectively. The schematics are generated by the Visual Molecular Dynamics (VMD) [27] software using the protein YibK (PDB code 1MXI).

Currently, over 100,000 protein structures have been solved and deposited into the protein data bank (PDB), a database of the three-dimensional structures of proteins and some related nucleic acids. [28] The protein structures are highly diverse due to the numerous possibilities of the arrangement of amino acid residues and secondary structures. According to the structures, proteins can be classified into the following classes as shown in Figure 1.2.[29,30]

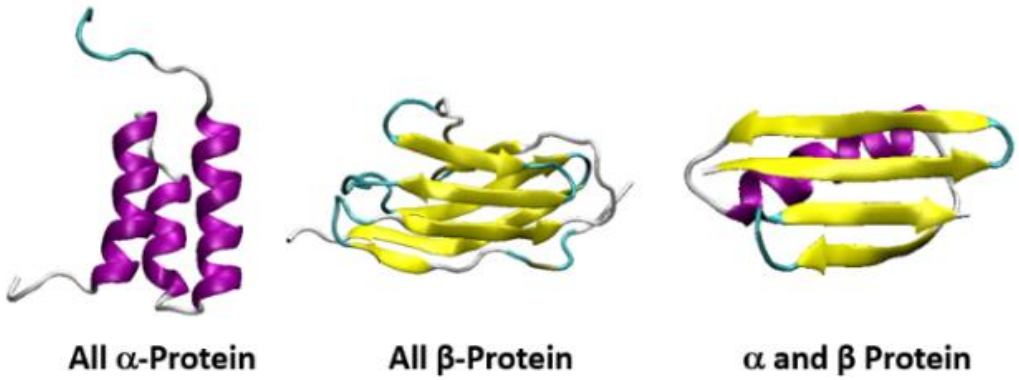

Figure 1.2 Structural classes of proteins.

The schematics are generated by the VMD software. All α-proteins are the proteins that contain predominantly α-helices (PDB code 2KDL). All β-proteins are the proteins that contain predominantly β-strands (sheets) (PDB code 1TIT). α- and β-proteins are the proteins containing both α-helices and β-strands (sheets) (PDB code 1PGA).

1.1.2 The Protein Folding Paradox

Despite being able to sample a range of conformations, proteins are capable of folding spontaneously from disordered polypeptide chains to their correct structures in a physiological environment within a surprisingly short time after being synthesized. How a protein folds is one of the greatest current challenges in biological sciences. The challenge of the protein folding problem is often demonstrated by the Levinthal Paradox, proposed by Cyrus Levinthal.[31] This paradox states that there are a very large number of possible conformations (3^{100}) for a protein of 100 amino acid residues, assuming three possible conformations for each residue. Consequently, it would take forever for the protein to search every possible conformation randomly during the folding process and find the correct structure, while it

actually only takes from less than a second to a few hours for most proteins to fold. Thus, the protein must have certain pathways to avoid the random search required to sample all possible conformations.[31] Indeed, theoretical and experimental efforts have been carried out to explain the folding pathways using several models[5,6,8,21,32-38]. In the diffusion-collision model, the secondary structures form independently before the tertiary structure. These secondary structures diffuse and collide with each other, resulting in the formation of a compact tertiary structure.[39] In the nucleation-condensation model, the hydrophobic polypeptide chain collapses and stabilizes the secondary structures, i.e. the secondary and tertiary structures form simultaneously.[40] Both of these models have been supported and some proteins combine both models during the folding process.[41] Furthermore, while many proteins can fold to the native structures themselves under physiological conditions, some proteins require assistance from a type of proteins called chaperones in order to fold correctly.[42-44]

1.1.3 Funnel Shaped-Energy Landscape

A concept called the folding funnel is often used to describe the rapid folding of proteins.[45,46] As shown in Figure 1.3, the folding funnel theory assumes that a disordered polypeptide chain folds along the funnel-shaped energy landscape towards a global energy minimum to achieve the well-defined native structure. The disordered polypeptide chain has numerous conformations and very high energy. During the folding process, along the free energy landscape, some proteins collapse into a compact folding intermediates called the "molten globule state", causing the secondary structures to form and the entropy of the water molecules around the proteins to increase. Then many native contacts between amino acid residues form to further stabilize the native structure at the global free energy minimum.[45,46]

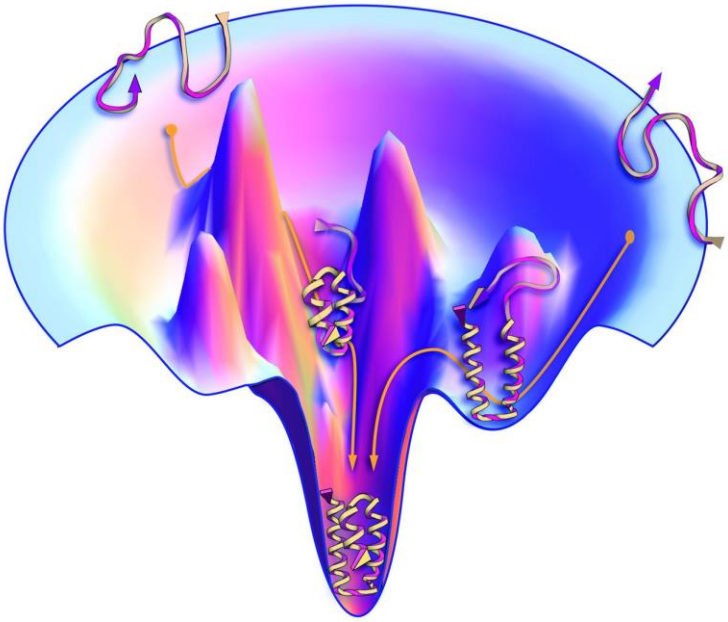

Figure 1.3 Funnel shaped-energy landscape of protein folding.

Proteins fold from disordered structures at high free energy to native structure at the bottom of the energy funnel. Reprinted from (Dill KA & MacCallum JL (2012), Science 338(6110):1042-1046) with permission of AAAS.[45]

In addition, the surface of the folding funnel is often described to be “rough" and many local energy minima, separated by energy barriers, could exist along the folding pathways. Folding intermediates could be kinetically trapped at these local energy minima.[6]

1.1.4 Two-State and Three-State Models for Protein Folding Pathway

Proteins can transit between the native and unfolded states upon environmental changes, such as changes in temperature and solvent. While folding is the transition from an unfolded state to the native state, the transition from native to unfolded state is called unfolding or denaturation. The folding/unfolding of many simple and small proteins is approximately described using two-state model, which assumes that it is all-or-none transition and only fully folded (native) and fully unfolded states are populated.[47] In the two-state model, there is only one energy barrier, where the transition state is located, between the native and unfolded states (Figure 1.4 left).

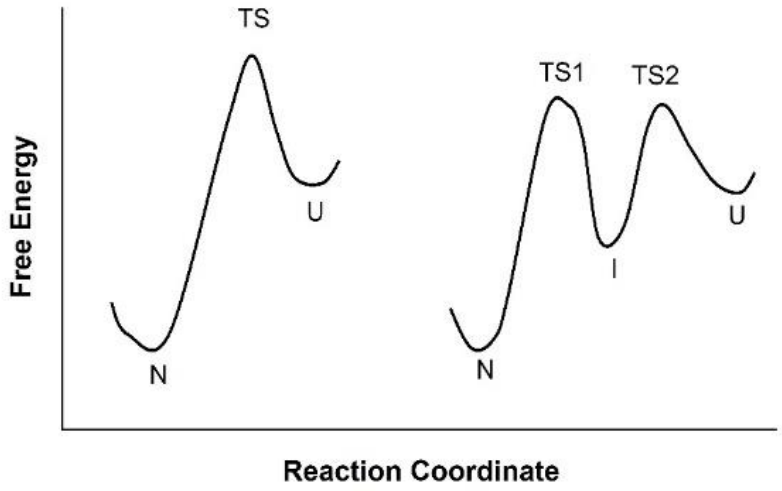

Figure 1.4 Protein folding/unfolding pathways.

Left: two-state model. Right: Three-state model with an intermediate state. N, TS, U and I represent native, transition, unfolded and intermediate states, respectively.

While using two-state model makes it easy to characterize the thermodynamics of protein folding/unfolding, protein folding can be far more complex and it is often necessary to adopt more sophisticated models. For this purpose the three-state model is often used as an alternative, when the two-state model is not sufficiently accurate to describe a folding/unfolding system. In the three-state model, in additional to the native and unfolded states, an intermediate state is trapped by two separate energy barriers and situated between the native and unfolded states (Figure 1.4 right).[48] In addition, many proteins fold in more complex ways that have more than three states.

1.1.5 Mechanical Folding/Unfolding of Proteins

In protein folding/unfolding studies, chemical denaturation is one of the most widely used methods. A chemical denaturant such as urea or guanidine hydrochloride (GdmCl) can perturb the native structure of a protein and stabilize the unfolded polypeptide chain. Changing the concentration of denaturant would change the folded and unfolded fractions of the protein.[48]

Similarly, a force can be applied to a single molecule protein and acts as a "denaturant" to trigger the transitions between native and unfolded states of the protein. This is called mechanical folding/unfolding.[49] Zhurkov-Bell-Evans model is often used to describe the mechanical unfolding of proteins.[50-52]

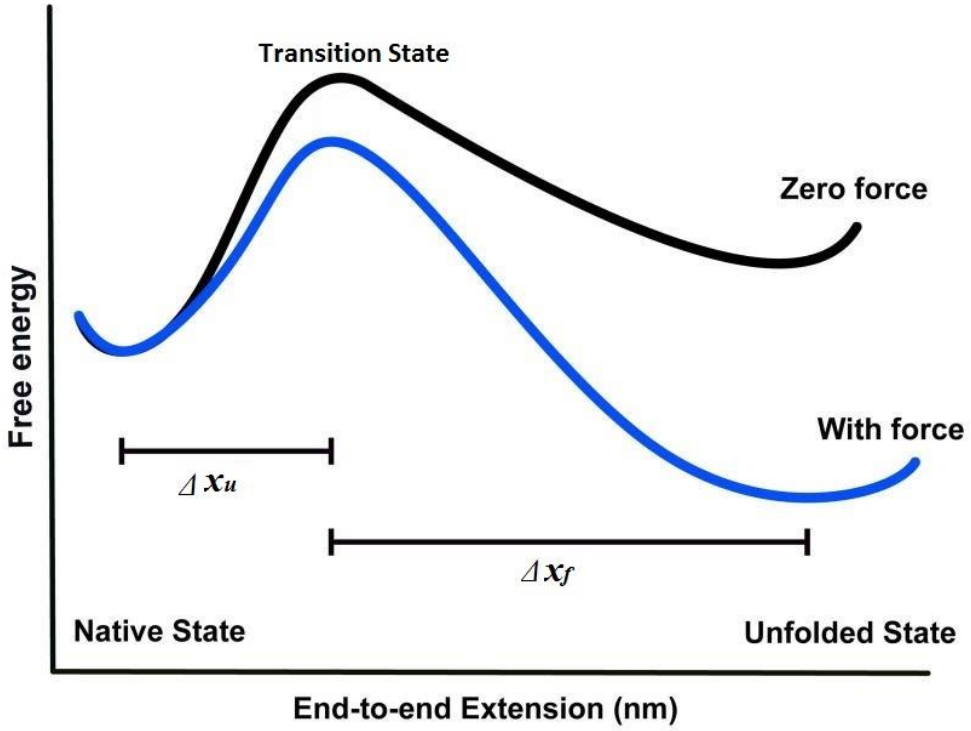

Figure 1.5 The effect of force on the energy landscape of protein folding/unfolding.
The blue and black curves are the energy landscapes with and without force, respectively. The unfolding distance and folding distance are shown by Δx_u and Δx_f, respectively. Adapted from (Jagannathan B & Marqusee S (2013) Biopolymers 99(11):860-869.) with permission of John Wiley and Sons.[53]

The native and unfolded states of a given protein are separated by an energy barrier (Figure 1.5). The distance between the unfolded and transition states is defined as the folding distance (Δx_f) and the distance between the native and transition states as the unfolding distance (Δx_u). In the absence of a stretching force, the native state thermodynamically dominates due to its lower free energy. In contrast, when a stretching force is applied to the protein, the free energy

barrier is reduced by the product (Fx) of the force (F) and the distance (x) along the direction of the force. If the force is high enough to make the free energy of the unfolded state lower than the free energy of the native state (blue curve in Figure 1.5), i.e. $Fx>\Delta G$, thermodynamically, the unfolded state would become more stable. What is more, the force lowers the energy barrier for unfolding by $F\Delta x_u$, which kinetically facilitates the unfolding reaction. [53] Zhurkov-Bell-Evans model assumes a single energy barrier between the native and unfolded state and the force applied on the molecule does not change the location of the transition state. In addition, the effect of the shape of the energy landscape is not considered. The effect of force on the kinetics of protein folding/unfolding is discussed later throughout this thesis.

1.2 Single-Molecule Force Spectroscopy

Single-molecule force spectroscopy (SMFS) has become a powerful tool for investigating mechanical folding/unfolding of proteins at the single-molecule level, providing new insights that are otherwise impossible to obtain using traditional methods. [2,35,49,54-61] SMFS can be performed using several techniques such as optical tweezers, magnetic tweezers, atomic force microscopy (AFM), etc. In this thesis, we mainly use AFM and optical tweezers to study the folding/unfolding of proteins.

1.2.1 SMFS Using AFM

Among the force spectroscopy techniques, AFM is the most widely used due to its ability to measure large force (10 to 10^4 pN) and given its excellent spatial (angstrom) and temporal resolution (ms).[35,62-72] Figure 1.6 shows the principle of a SMFS experiment using AFM. After being deposited on a substrate, a single-molecule polyprotein is randomly picked up and stretched by an AFM tip. The relationship between the restoring force and the extension of the

polyprotein chain can be well described by the worm-like chain (WLC) model of polymer elasticity[35,73]:

$$F(x) = \frac{k_B T}{p}[\frac{1}{4}(1 - \frac{x}{L_C})^{-2} - \frac{1}{4} + \frac{x}{L_C}]$$

where, $F(x)$ is the restoring force at the extension of x, k_B is the Boltzmann constant, T is the temperature in Kelvin, p is the persistence length and L_C is the contour length of the polymer. The WLC model treats the polymer chain as an isotropic rod with continuous flexibility that is smoothly curved at room temperature. The average cosine of the bending angle, θ, over all starting positions decays exponentially with the distance, L, at the scale of the persistence length (p):

$$\langle cos\theta \rangle = \exp(-L/p)$$

The persistence length characterizes the stiffness of a polymer. When the polyprotein is extended, the force increases non-linearly (Figure 1.6B (1)). As the force accelerate the unfolding rate of the proteins as previously introduced, one folded domain in the polyprotein unfolds (Figure 1.6B (2)), leading to an increase of the contour length of the polyprotein molecule and thus a force drop to a lower value. The force-extension relationship of unfolded polypeptide chain can be described by the WLC model using a persistence length of 0.4 nm.[35] Further stretching increases the force and will lead to the sequential unfolding of the domains in the polyprotein chain, leading to a FEC (FEC) with a saw-tooth pattern appearance. Each force peak represents an unfolding event of one domain within the polyprotein except the last force peak, which usually corresponds to the detachment of the polyprotein either from the substrate or the tip. By fitting the consecutive force peaks in the FEC, the contour length increment caused by the unfolding of each individual domain from the polyprotein can be measured.

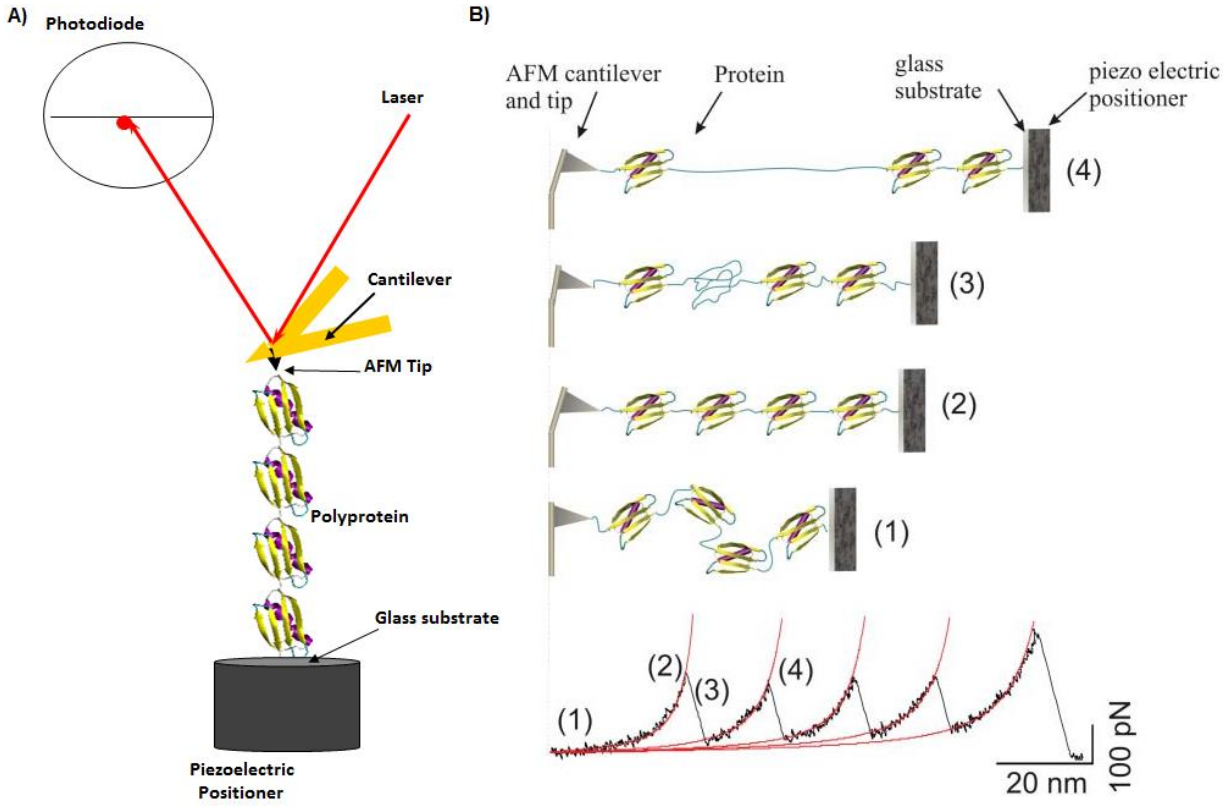

Figure 1.6 SMFS experiment based on AFM.

A): Schematic of experimental setup. Tandem polyprotein is deposited on a glass substrate. The movement of the glass substrate is controlled by a piezoelectric positioner. An AFM tip on a cantilever picks up a polyprotein by non-specific attachment. A laser beam is reflected on the back of the cantilever and shined onto a photodiode. The deflection of the cantilever can be measured by monitoring the movement of the laser on the photodiode, through which the force applied onto the protein is then calculated. B) Schematic of stretching a single-molecule polyprotein and the corresponding FEC with a saw-tooth pattern. (1) The AFM tip picks up the polyprotein. (2) As the substrate moves away from the tip, the force is increased according to the worm-like-chain model. (3) One of the multiple domains unfolds, resulting in the increase of the contour length and the decrease of force. (4) Force is increased upon further stretching. Reproduced in part from (Li H (2007) Org Biomol Chem 5(21):3399-3406) with permission of The Royal Society of Chemistry.[74]

1.2.2 SMFS Using Optical Tweezers

As another powerful force spectroscopy technique, the optical tweezers have been used extensively in single molecule studies, including protein folding and unfolding, due to their superior force resolution (0.1 pN) and force range (1 pN to ~100 pN). However, the

experimental setup is more complex and technically more demanding than AFM.[55,73] The optical tweezers are very suitable for monitoring the folding/unfolding of mechanically labile proteins.[55,75-85]

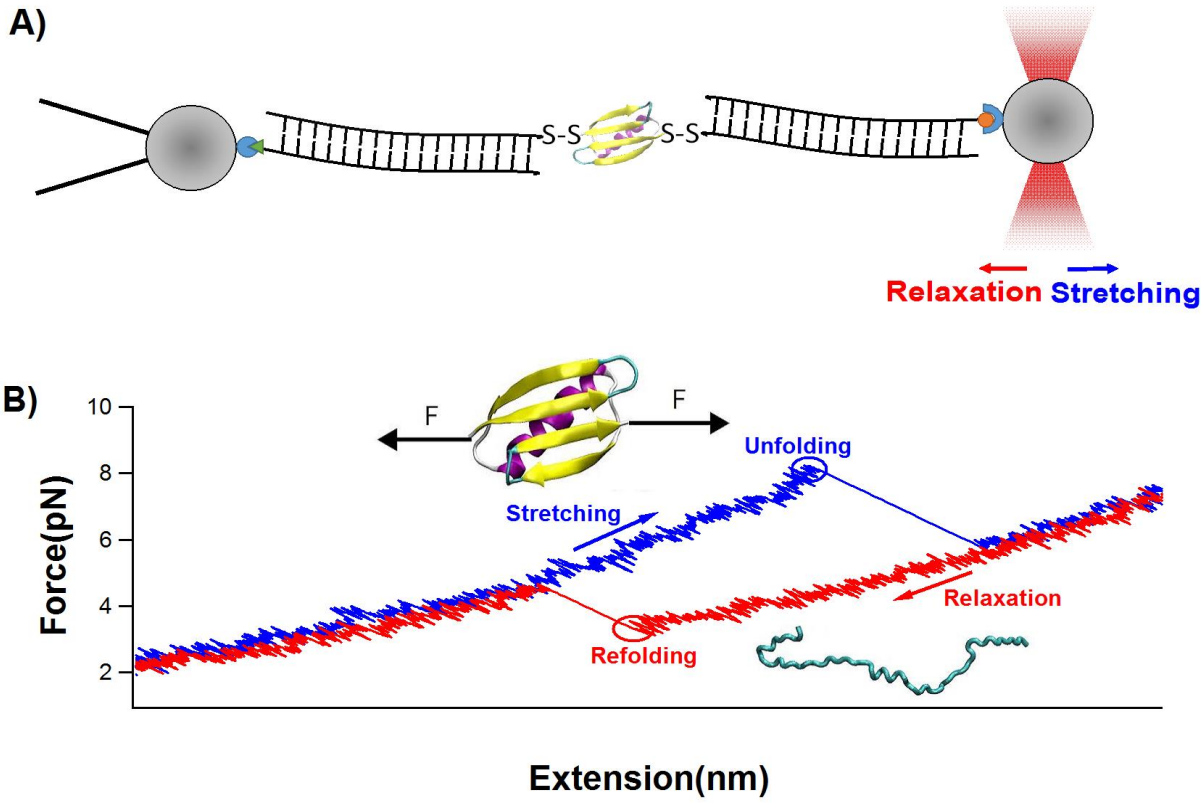

Figure 1.7 Protein folding/unfolding using optical tweezers.

A) Schematic of optical tweezers experiment. The protein is engineered to have two cysteine residues. Each thiol group of these two cysteine is connected to a double-strand DNA handle with a thiol group at one end through disulfide bonding. The other end of the DNA handles with antibodies are then coupled with specific antigens on two polystyrene or silica beads. One of the two beads is fixed on a micropipette and the other is trapped and controlled by a laser beam. The force in the direction of stretching or relaxation can be determined by the deviation of the laser beam from the center of the bead. B) Representative traces of stretching and relaxation. Similar as AFM experiment, when the protein-DNA complex is being stretched (stretching trace colored in blue), the force is increased. Upon the unfolding of the protein, the force is suddenly released. As the direction of the movement of the bead is reversed (relaxation trace colored in red), the force is decreasing. Upon the refolding of the protein, the force is suddenly increased.

As shown in Figure 1.7, in a typical optical tweezers experiment, the protein is coupled to DNA handles at each ends through disulfide bonding. The DNA is then coupled with polystyrene or silica beads that can be either fixed on a micropipette or trapped and moved by a laser beam. When the bead is trapped by a laser beam, the spatial information can be obtained by monitoring the movement of the laser beam and the force in the direction of stretching or relaxation can be determined by the deviation of the laser beam from the center of the bead. Fitting the FECs from optical tweezers experiments can be more sophisticated because the DNA handles have a different persistence length (~50 nm)[73] than polypeptide (~ 0.8 nm)[86]. In addition, the enthalpic compliance of DNA when being stretched must be accounted for. The extensible WLC model with an additional parameter of stretch modulus, K_0 (~1000 pN), has been used to describe the extension of DNA more precisely[73]:

$$\frac{FP}{k_BT} = \frac{1}{4}\left(1 - \frac{x}{L_C} + \frac{F}{K_0}\right)^{-2} - \frac{1}{4} + \frac{x}{L_C} - \frac{F}{K_0}$$

Thus fitting the FECs from optical tweezers experiments requires the combination of the WLC model on polypeptide and the extensible WLC model on DNA handles.

Both AFM and optical tweezers are useful in SMFS studies. They can complement each other and provide a lot of valuable information about protein folding/unfolding, which can be extracted through the simultaneously obtained parameters such as the force, distance and time.

1.3 Steered Molecular Dynamics Simulations

Although SMFS has been proven to be an excellent method for the studies on protein folding/unfolding, only limited details of the molecular mechanisms are available in SMFS experiments. Molecular dynamics (MD) simulations compliment the SMFS experiments and suggest more details of the mechanism of the mechanical folding/unfolding of proteins. A number of programs have been developed for MD simulations such as CHARMM, NAMD,

AMBER, etc.[87-89] In MD simulations, the physical movement of atoms is described using newton's equations of motion. In particular, steered molecular dynamics (SMD) simulations provide a way to directly apply forces on specific atoms, which can be used to mimic the SMFS experiments.

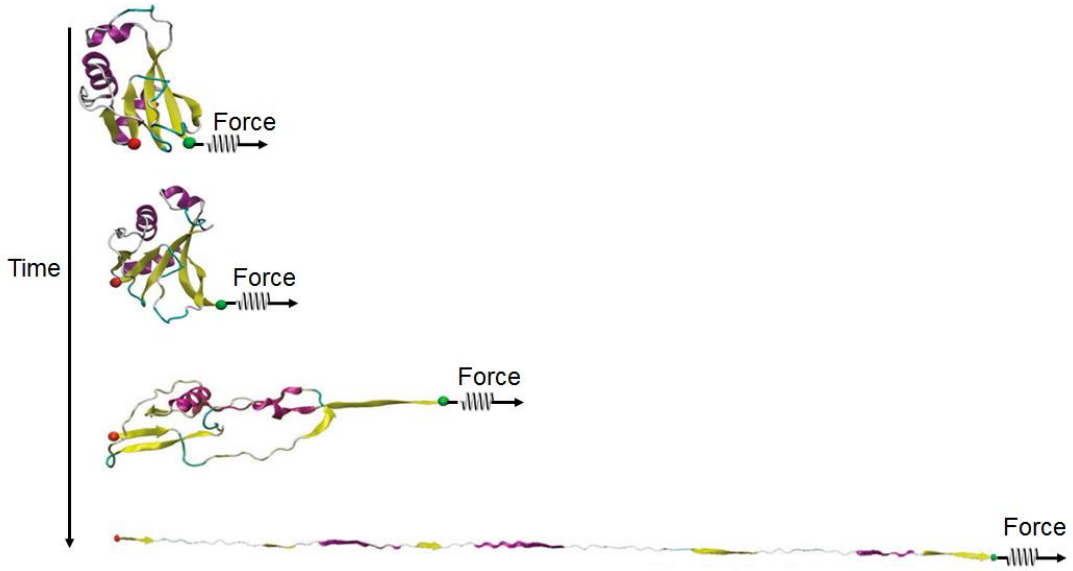

Figure 1.8 Snapshots from the trajectory of pulling a protein using SMD simulations.

The fixed atom is labeled as the red balls and the atom being stretched is labeled as the green balls. The springs and arrows represent the force applied on the protein. The schematics are generated by the Visual Molecular Dynamics (VMD) [27] software using the structure of AFV3-109 (PDB code 2J6B).

In a typical SMD simulation that simulates the pulling on a protein in SMFS experiments, the structure of protein must be known. Before the pulling, the protein is solvated in either an explicit solvent or an implicit solvent environment. The energy of the protein-solvent system is minimized and equilibrated. As shown in Figure 1.8, at the beginning of the pulling, the coordinate of one atom in the protein structure is fixed (red balls in Figure 1.8) and a stretching force is applied on another atom (green balls in Figure 1.8) along the opposite direction (the springs and arrows) of the fixed atom. The protein is then extended along the direction of the

force and the native structural elements are being broken as the time elapses. The simulation can be paused or stopped at any time. The coordinates of each individual atom in the protein structure at each time step are determined and recorded by the simulation program. A trajectory of the movement of each atom is generated for further analysis. Usually, constant velocity and constant force modes are used to simulate the stretching of a protein. In the constant velocity mode, the speed of the movement of the atom being stretched is constant and the force applied on the atom is calculated and recorded at each time step. Essentially, the SMD simulations in constant velocity mode mimic the pulling experiments in SMFS at constant pulling speeds and generate FECs with all the structural information, which can be very useful to estimate the mechanical stability of the proteins and the mechanism of the unfolding. In contrast, in the constant force mode, the force is kept constant and the extension of the atom being stretched is determined and recorded at each time step, leading to a length-time curve. The length-time curve is particularly useful to identify mechanically stable intermediate states during the pulling. Through the information provided by the simulation, detailed molecular mechanism can be drawn to compare with the results from SMFS experiments. Further details about SMD simulations will be given in Chapter 3 and 4.

1.4 Identifying Knotted Proteins

The complexity of protein folding makes it very challenging to fully understand the folding mechanism of proteins. More complex yet interesting challenges have been proposed. One of the most interesting challenges is that of the protein with knotted topologies.[9-12] Knots are very common in our daily life. Figure 1.9 shows a schematic of a simple trefoil knot. A knotting loop (yellow) is formed and two ends (red and blue) are threaded through the knotting loop. Sometimes we hate the knot and need to untie it. For example, our earphone cable often gets

knotted and it takes forever to untie it. But, more often, we need the knot, for example, a tie knot or shoelace. It also requires some effort to tie a knot. Similarly, as the protein is a polypeptide chain, it is, in principle, possible to form a knot. However, most proteins favor simple topologies as they need to fold efficiently to the native state. It is, therefore, very interesting to study the proteins that can adopt a knotted topology despite the associated topological complexity.

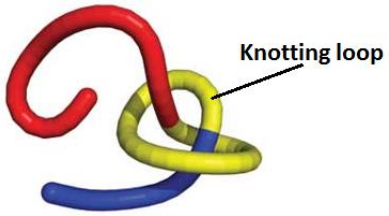

Figure 1.9 Schematic of a trefoil knot.

The knotting loop is colored in yellow. The two ends colored in red and blue are threaded through the knotting loop. Adapted by permission from Macmillan Publishers Ltd: [Nat Chem Biol] (Mallam AL & Jackson SE (2012), Nat Chem Biol 8(2):147-153), copyright (2012).[90]

About two decades ago, knots were not considered to exist in the native state of proteins due to the topological difficulty of the knotting process. In fact, in 1994, Mansfield published a paper that questioned the existence of knots in proteins.[9,91] By analyzing the structures of about 400 proteins, Mansfield determined that none of the 400 proteins have a knot except one, human carbonic anhydrase, which appears to have a very shallow knotted topology, though it has only a few amino acid residues entwined at one end in a loop. Almost at the same time, Liang and Mislow reported their findings that a few metalloproteins possess a trefoil knotted topology.[92] In those knotted metalloproteins, metal atoms and disulfide bonds are important parts of the knots. In other words, without those metal atoms and disulfide bonds, the knot would not exist. The absence of knots in protein structures was interpreted as meaning that the

folding mechanism of proteins only searches for conformations without knots, to avoid the high potential energy barrier that associated with the knots.

1.4.1 Discovery

Nevertheless, in 1996, Takusagawa and Kamitori reported that (S)-adenosylmethionine synthetase (MAT), a protein containing 380 amino acid residues, contains a trefoil knot in its native structure.[12] Residues from 239 to the C-terminus pass through the knotting loop formed by the residues from 11 to 154 as shown in Figure 1.10. To explain how the knot forms, they proposed that the β1 strand at the N-terminus of MAT was plugged through the knotting loop after the synthesis of β9 strand at the C-terminus.

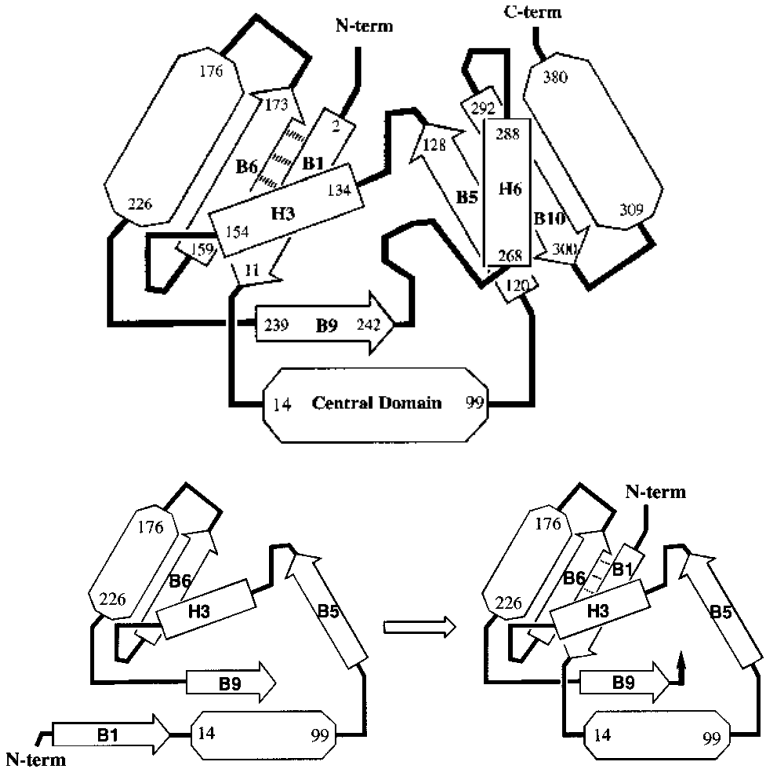

Figure 1.10 Trefoil knotted protein MAT.

Top: Schematic of the trefoil knot structure of MAT. The numbers of the start and end of amino acid residues are shown. Bottom: Proposed folding mechanism for knot formation. After synthesis of the β9, the N-terminus moves into the knotting loop, and the β1 forms hydrogen bonds (dotted lines) with the β6. Adapted with permission from (Takusagawa F & Kamitori S (1996), J Am Chem Soc 118(37):8945-8946). Copyright (1996) American Chemical Society.[12]

Although this is a good example of a knot identified in a native structure of protein, there are only ~ 10 amino acid residues at the N-terminus that pass through the knotting loop. In other words, it was still considered to be a very shallow knot. Thus, more effort in examining whether proteins contain knots were necessary to move this research area forward.

1.4.2 Expansion

Over time, the development of computer algorithms greatly improved the efficiency of the search for possible knots in proteins.[10,11,18,93] Usually, one could pull the two ends of a string to determine whether it is knotted. However, due to the complexity of protein structures, if the two ends of protein are not on the surface of the structure when being pulled, they may pass through a loop and it could untie a knot or even create a new knot. In 2000, Taylor reported a simple yet sufficient computer algorithm to determine the existence of a knot/knots formed by the polypeptide chain of proteins.[10] The two termini of the polypeptide chain are fixed and the remaining amino acid residues are made to shrink around the two termini. This method smooths the polypeptide chain while the two termini are fixed and two parts of the chain (N-terminus and C-terminus) cannot pass through each other. Ultimately the polypeptide chain can be reduced to a straight line between two termini if there is no knot. Otherwise, the polypeptide chain would be blocked by the knot(s). By using this algorithm, the knots in proteins could be characterized by the number of residues removed from N- and C-termini to reduce the protein to a straight line, i.e. to remove the knot.

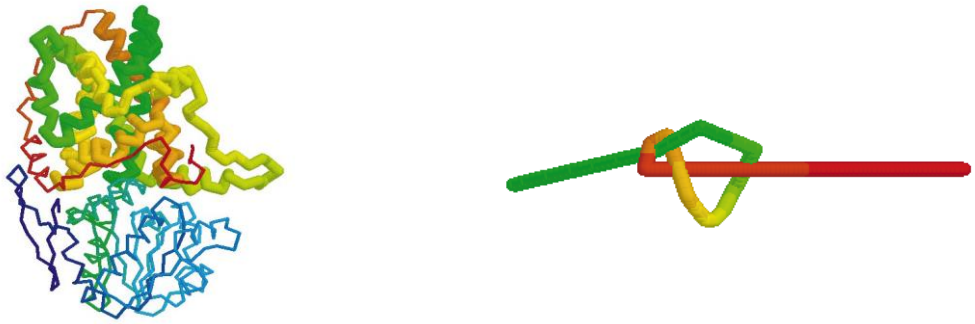

Figure 1.11 The figure-eight knotted protein (PDB code lyve).
Left: backbone representation of the native structure of 1yve. Right: the figure-eight knotted core in the smoothed representation of 1yve. Reprinted by permission from Macmillan Publishers Ltd: Nature (Taylor WR, Nature 406(6798):916-919.), copyright (2000).[10]

In this way, Taylor found that acetohydroxy acid isomeroreductase (PDB code: 1yve) contains a deep figure-eight knot (Figure 1.11). Meanwhile, he proposed that the duplicated domains in the knotted protein are important in the folding process. This discovery greatly promoted both experimental and theoretical studies on the knotted proteins.

As more protein structures are solved, the number of knotted proteins keeps increasing. In particular, Lim et al resolved the structure of the YibK methyltransferase, which contains a trefoil knot with a chain of ~ 40 amino acid residues knotted by a knotting loop.[94] Interestingly, YbeA-like proteins, another type of methyltransferase, were found to possess similar trefoil knotted topology as YibK does.[95] Because the trefoil knot is the simplest knot, YibK and YbeA have become very popular models for the studies on the knotted proteins (Figure 1.12).

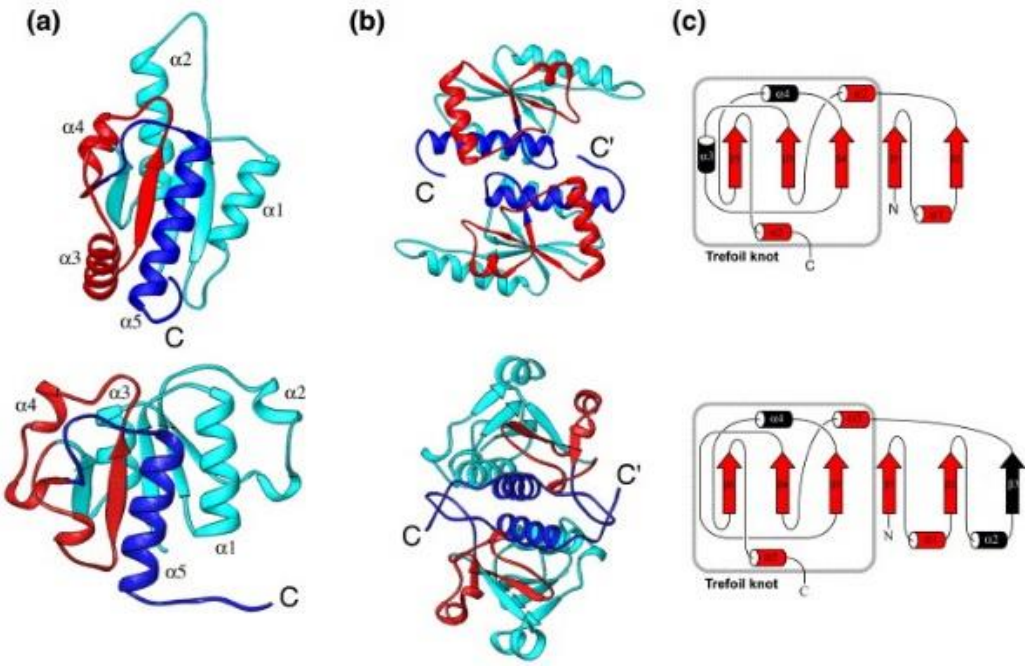

Figure 1.12 Structures of YbeA (top) and YibK (bottom).

Both proteins contain a trefoil knot. (a) Ribbon diagram of a monomer subunit, showing the deep trefoil knot at the C terminus. The knotting loop highlighted in red and the knotted chain highlighted in dark blue. (b) Dimeric structures of YbeA (top) and YibK (bottom). (c) Topological diagrams of YbeA (top) and YibK (bottom). Reprinted from Ref [95] with permission of Elsevier.

Beside the trefoil knot and figure-eight knot, more complex knots, such as the three-twist knot and the Stevedore knot, have been identified in the protein data bank. Until now, about 1% of the proteins in the PDB contain a knotted topology, and can be classified according to the types of knots. Figure 1.13 shows four representative proteins that possess four different types of knots. Accordingly, to systemically identify and classify the knotted proteins, researchers have built web servers that summarize all the knotted proteins, as well as automatically determine whether a protein is knotted or not.[17,18,96]

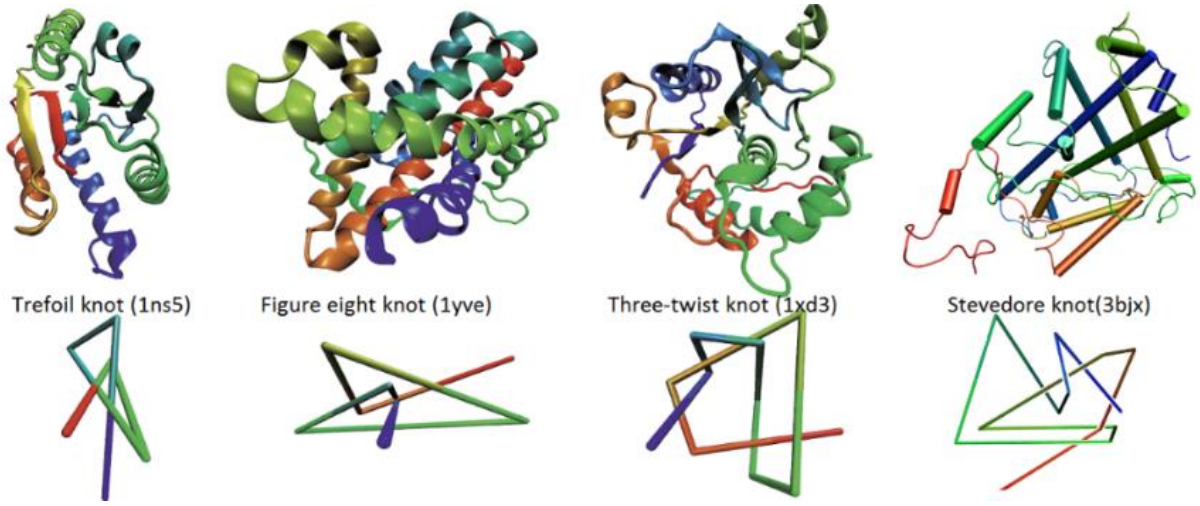

Figure 1.13 Three-dimensional structures (top) and topological schematics (Bottom) of proteins possessing different types of knots.

Adapted from Ref [20,97] with permissions.

1.5 Folding/Unfolding Mechanism of Knotted Proteins

Although it is now clear that many proteins are indeed knotted, determining the folding mechanism of knotted proteins remains quite challenging. Both experimental and theoretical studies have shed light on the folding/unfolding process of the knotted proteins.[10,11,13,16,19,90,94,97-112] Among all the knotted proteins, YibK and YbeA have been chosen as models due to the simplicity of the knot type (the simplest trefoil knot).

1.5.1 Experimental Studies on the Folding and Unfolding of Knotted Proteins YibK and YbeA

The folding pathways of YibK and YbeA were compared by Mallam and Jackson using chemical denaturation methods.[95] Both YibK and YbeA are dimers in their native states. Multiple pathways coupled with multiple intermediate states were found during the folding of YibK, while the folding and unfolding of YbeA only occurs as a simple three-state sequential

pathway with a single intermediate state. During the folding of both proteins, the dimerization from monomeric intermediates to dimeric native states is the slowest step.

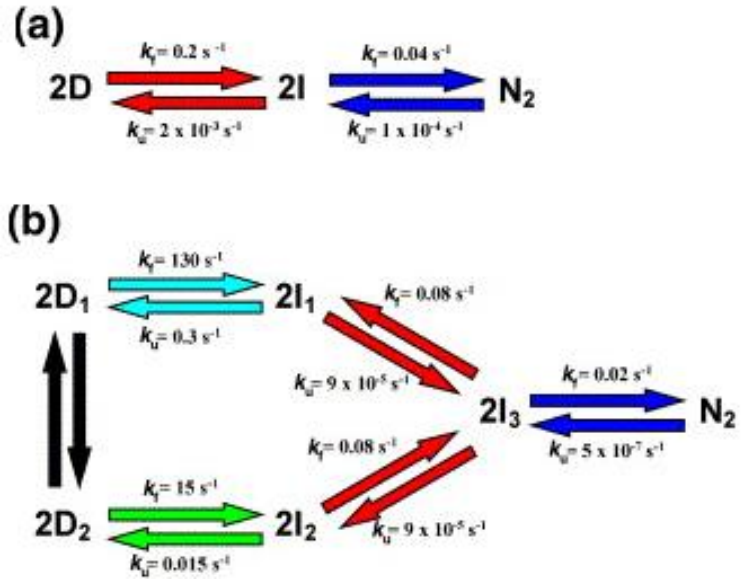

Figure 1.14 The folding pathway of (a) YbeA and (b) YibK.

D: monomeric denatured state; I: monomeric intermediate state; N_2: dimeric native state. Reprinted from Ref [95] with permission of Elsevier.

To further study when a protein forms a knotted topology during the folding, Mallam, Rogers and Jackson designed a clever experiment that covalently circularized YibK by forming a disulfide bond between the two residues mutated to cysteine at N- and C-termini.[109] When the disulfide bond was formed in the native state, the trefoil knot would remain even when the protein was then fully denatured by adding a chemical denaturant. They found that the knotted and denatured polypeptide can fold back to native state, which was not surprising because the knot had been already formed. Interestingly, when the disulfide bond was formed after the protein was fully denatured, it could still fold into its native trefoil knotted state. As the topology of the denatured polypeptide was fixed by the disulfide bond, it would never fold into its native state if it were unknotted. Thus the denatured state of YibK possesses the same trefoil

knotted topology as the native state does. YbeA was also studied and showed the same behavior. This finding successfully identified that the knotted topology exists in the denatured state when most secondary structure elements of the protein have not been formed.

1.5.2 The Importance of the Slipknotted Topology

Although experimental studies have revealed the folding pathway of a knotted protein, the protein structure during the folding process remains unclear. Theoretical studies, on the other hand, have provided invaluable structural information about the folding mechanism. Sułkowska et al have proposed a possible folding mechanism for YbeA using molecular dynamics simulations based on a coarse-grained model.[107] As shown in Figure 1.15, there are five key steps in the folding pathway. In the first step, by crossing two parts of itself, the polypeptide chain forms a loop, which is the knotting loop. In the second step, one end of the chain is bent towards the knotting loop to form a hook. Instead of plugging the terminus of the hook into the knotting loop, which would raise a substantial energy barrier, the middle of the hook is then threaded into the knotting loop and a slipknot is formed as circled in Figure 1.15. The threaded loop is then further passed through the knotting loop. The trefoil knot forms when the terminus of the threaded loop completely passes through the knotting loop. Molecular dynamics simulations revealed that the pathway with the slipknotted intermediate results in a greatly reduced the energy barrier, as compared to the pathway that involves plugging the terminus through the knotting loop.

Thus, the slipknotted topology appears to be important for the folding of knotted proteins. A slipknot is like a shoelace and it is, technically, not a real knot. Instead of completely threading one end of the chain through the knotting loop, a threaded loop partially passes through the knotting loop and the end of the chain remains at the other side of the knotting loop as shown

in Figure 1.15. King et al have found several rare examples of proteins that have slipknotted topology in their native conformations.[16] Due to its importance in protein structure and folding, slipknotted proteins have drawn particular attention and will be further discussed in this thesis.

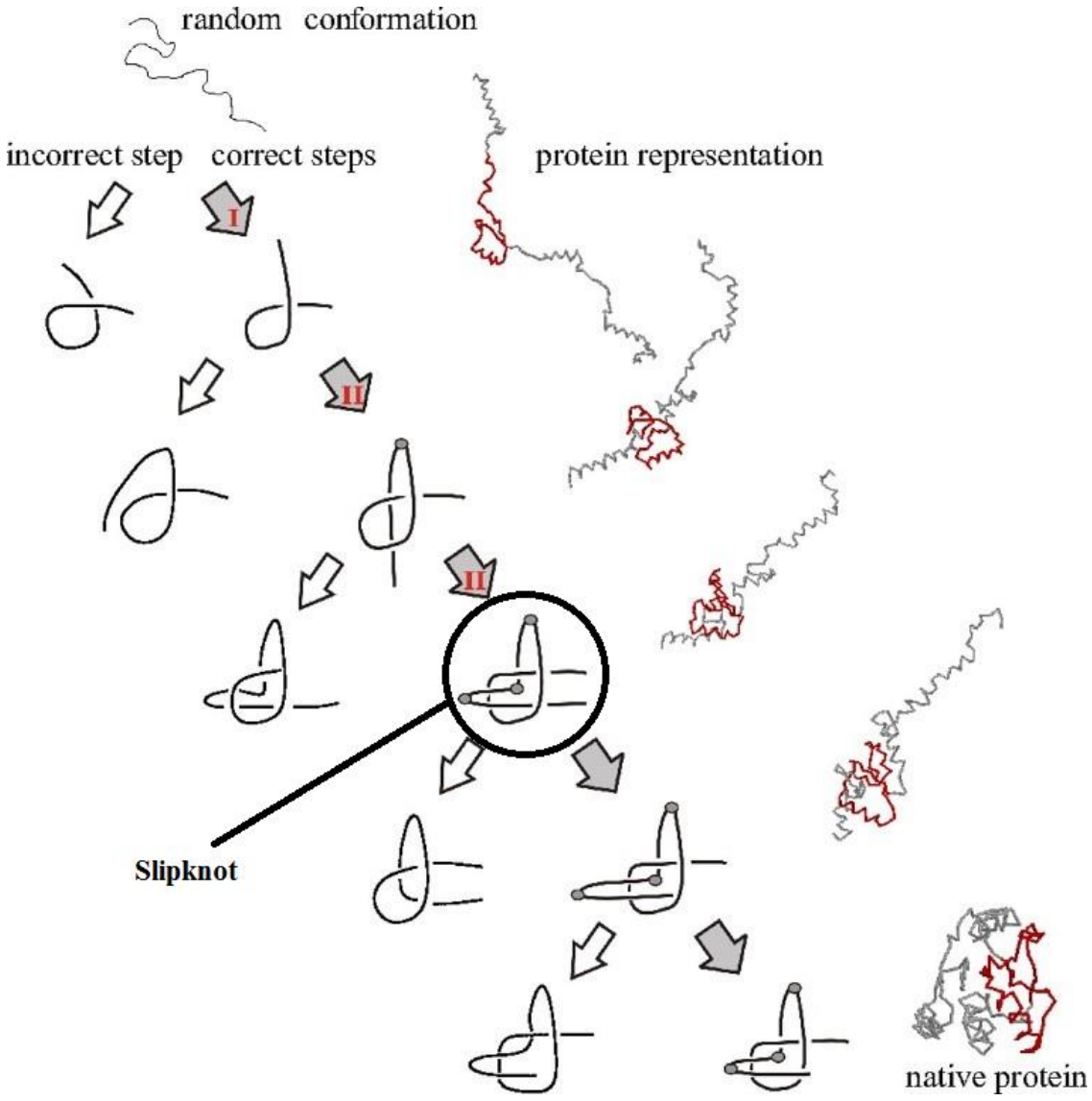

Figure 1.15 The folding pathway of a trefoil-knotted protein.

Schematics of topologies of polypeptide are shown in the middle. Representative protein conformations are shown at the right side and the locations of the knot are colored in red. Incorrect folding steps are shown at the left side. The slipknot conformation is circled. Adapted from Ref[107] with permission of Copyright (2009) National Academy of Sciences, USA.

1.6 Force-Induced Topological Change of Knotted/Slipknotted Protein

As mentioned earlier in this chapter, pulling on two ends of a protein is a simple way to test the existence of a knot. More importantly, the mechanism of pulling knotted/slipknotted proteins can provide useful insight into the formation of the knot/slipknot. Actually, quite a few knotted/slipknotted proteins have been pulled in both experimental and computational studies.[113-120]

1.6.1 Tightening the Knot

SMD simulation has been generally used to apply external force on certain atoms of proteins in specific pulling direction *in silico*. Upon being pulled on its two ends, a knotted protein will be unfolded and extended and, subsequently, the knot will be tightened. Sułkowska et al have stretched 20 proteins with a knotted topology by SMD simulations using a coarse-grained model.[115] When a stretching force is applied onto the two ends of a knotted protein, the knot shrinks and one end of the knot moves along the polypeptide chain with sudden jumps. This is quite different from the tightening of a homopolymer in which the knot slides off the chain as a diffusive motion. A possible reason is that some amino acid residues have larger side chains than others so that those large residues create steric hindrance when the knot moves.

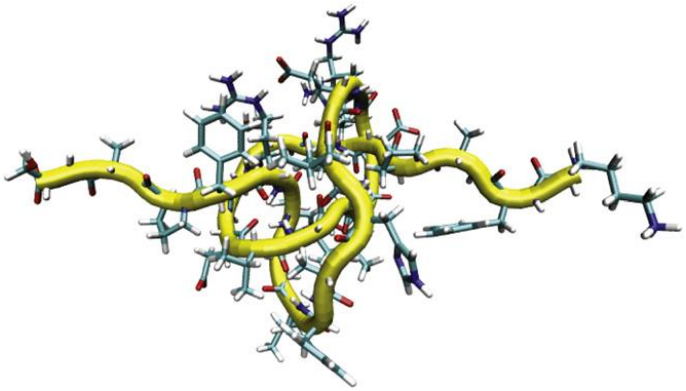

Figure 1.16 Tightened figure-eight knot in phytochrome stretched by SMD simulation.
Adapted from Ref [116] with permission of Elsevier.

Along with SMD simulations, SMFS is able to experimentally hold a protein and stretch it by its two ends. Wang et al performed SMFS on a shallow trefoil knot protein, bovine carbonic anhydrase B, and stretched it to a tightened knot.[120] A figure-eight knotted protein, phytochrome, was stretched by Bornschlogl et al using SMFS based on AFM, as well as SMD simulations.[116] The unfolding force of phytochrome was determined to be ~ 70 pN, which is not as high as many mechanically stable proteins such as the I27 domain of human titin (~200 pN).[49] In addition, both their experimental and simulation results revealed that the tightened figure-eight knot contains 17 to 19 amino acid residues as shown in Figure 1.16. Further discussion about the tightening of the knot can be found in chapter 4.

1.6.2 Untying the Knot/Slipknot

Almost everyone has the experience of untying a knot or slipknot. By being stretched on the middle of a knotting loop, the knot can be loosen and then untied. Similarly, SMD simulations are able to untie a protein knot by applying force on the knotting loop of the protein. Sułkowska et al used YbeA as a model to demonstrate the feasibility of untying a trefoil knot using SMD simulations.[117] Different anchor points to apply a stretching force on were chosen. Their results indicated that the untying of the knot depends on the length of the knotting and threaded loop and the location of the anchor point of force. As shown in Figure 1.17, in order to successfully untie the knot, at least one of the anchor points should be located within the knotting loop.

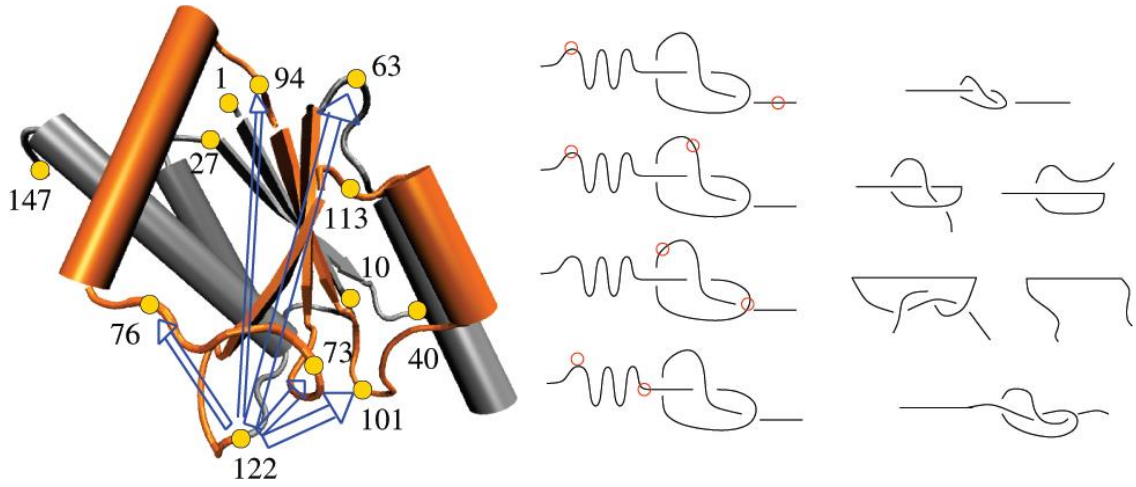

Figure 1.17 Pulling a protein knot.

Left: three dimensional structure of YbeA. Middle: Schematics of knotted topology and the force anchor points (red circles). Right: Schematics of possible resultant topologies. The knot can be either untied or not in the cases of the second and third from top to bottom. Adapted) with permission from (Sulkowska JI, Sulkowski P, Szymczak P, & Cieplak M (2010) Untying Knots in Proteins. J Am Chem Soc 132(40):13954-13956), Copyright (2010) American Chemical Society.[117]

On the other hand, one can easily untie a slipknot by pulling on its ends, just like when one unties a shoelace. Recently, SMD simulations on a slipknotted protein, thymidine kinase, illustrated the unfolding pathways upon stretching on both its ends (Figure 1.18).[121] Under a low pulling force, the shrinking rate of the threaded loop is faster than that of the knotting loop, meaning that the threaded loop can be pulled out of the knotting loop without the knotting loop being tightened. This pathway I is called the slip-bonds pathway as the slip-bonds become more labile when a force is applied. Under a high pulling force, the shrinking rate of knotting loop increases dramatically and exceeds the shrinking rate of threaded loop, which is then jammed by the tightened knotting loop. This pathway II is called the catch-bonds pathway as catch-bonds become more stable when a force is applied.

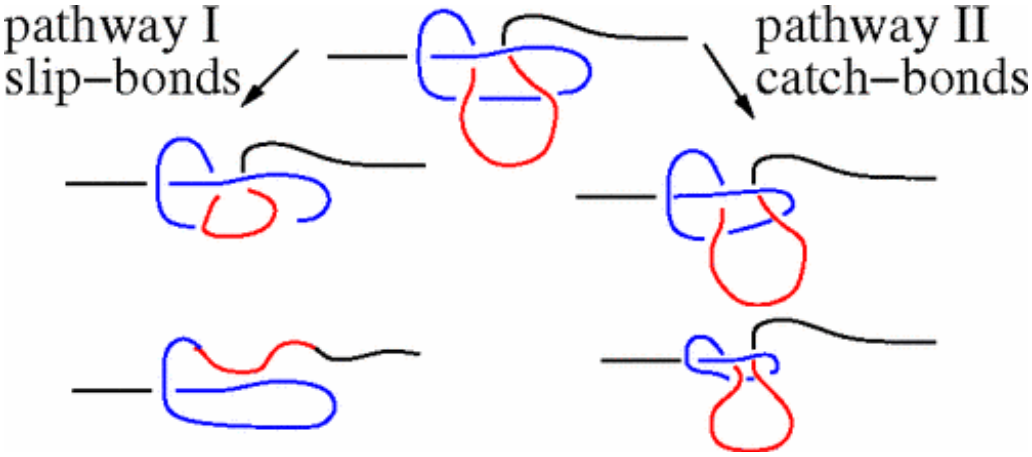

Figure 1.18 Two unfolding pathways of slipknotted protein, thymidine kinase.
The knotted loop and threaded loop are colored in blue and red respectively. At low force, the threaded loop shrinks faster than the knotting loop, and the slipknot is then loosen and untied (pathway I). At high force, the knotting loop shrinks faster than threaded loop, and the threaded is then jammed by the tightened knotting loop (pathway II). Figure adapted from Ref [121] with permission of American Physical Society.

1.7 Aim of This Thesis

As discussed above, the slipknotted topology is of great importance in terms of its structure and role in the folding of knotted proteins. Although there are very interesting studies the slipknotted protein using computational methods, there are comparatively few experimental studies. SMFS is one of the most suitable tools to investigate the folding and unfolding of this type of protein as it is able to directly stretch the protein and monitor the dynamics. Inspired by the pioneer studies introduced above, in this thesis, we mainly focus on the unfolding and folding of a slipknotted protein using SMFS base on both AFM and optical tweezers.

In order to demonstrate how to use AFM to study the mechanical unfolding and folding of proteins, in chapter 2, we use a small protein, NuG2, as a model, though it contains no knot or slipknot. The kinetics of mechanical unfolding and folding of NuG2 is extracted from SMFS experiments and Monte Carlo (MC) simulations. In chapter 3 and 4, we mainly study the mechanical unfolding of a slipknotted protein, AFV3-109 using AFM and SMD simulations.

The slipknotted topology of AFV3-109 is converted to either a linear chain or a trefoil knot upon stretching in different directions. Multiple pathways are found in the mechanical unfolding of AFV3-109. Although AFM can be used to study the folding of proteins, our results show its limitation in detecting the folding of AFV3-109. Hence, in chapter 5 we make use of the advantage of the optical tweezers, which is more suitable for the study of protein folding at very low forces, to revisit the mechanical folding of AFV3-109. As a result, a simple two-state folding is observed, suggesting that the slipknotted topology does not create a high energy barrier for AFV3-109. In chapter 6, we summarize the conclusions from this thesis and give our opinions about the future directions of the studies of knotted/slipknotted proteins.

Chapter 2: Directly Observing the Reversible Two-state Unfolding and Refolding of an α/β Protein by Single-Molecule Force Spectroscopy Based on Atomic Force Microscopy

In order to study the folding and unfolding of a slipknot protein, we choose AFM to perform SMFS experiments due to its convenience and capability in the study of mechanical unfolding and folding of proteins. Here we use a small protein, NuG2, as a model to demonstrate the use of AFM and some general methods used in force spectroscopy studies, though NuG2 contains no knot or slipknot.

2.1 Studying Protein Folding Using AFM

Protein folding-unfolding is a fundamental event in life, and of vital importance for essentially every single biological process. Understanding protein (un)folding and misfolding mechanisms remains a critically important task in life sciences.[21,22] Improvements in experimental and computational methods continue to cast new insights into this important problem.[2,37,54,55,122-124] The high spatial resolution and fast dynamic response of AFM make it especially appealing for investigating protein dynamics. However, the relatively poor long-term stability in force found in AFM makes it challenging to monitor protein folding-unfolding in real time. Although the development of a lock-in detection scheme has made it possible to observe protein folding,[125] force drift remains the limiting factor for using AFM in protein folding studies.[126-128]

Chapter 2 has been published as [He CZ], Hu CG, Hu XD, Hu XT, Xiao A, Perkins TT & Li HB. Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angew. Chem.*, **2015**, 127: 10059–10063. This chapter is incorporated in this thesis with permission of John Wiley and Sons, Copyright (2015).

Recent progress has led to significant improvements in the long-term stability of AFM, allowing for the folding of small proteins to be directly observed in real time and for the full characterization of the folding-unfolding energy landscape.[54,125,126,129-132] In such experiments, protein folding-unfolding at low stretching forces occur under conditions close to equilibrium, giving distinct (un)folding events accompanied with protein shortening (or lengthening). However, reported direct observations of real-time protein folding using AFM are limited to all-α proteins, such as Ankyrin[54,132] and calmodulin.[126,129,130,133] These proteins are mechanically labile and unfold at low forces, allowing protein folding and unfolding to occur near equilibrium. However, the direct, real-time observation of the folding of α/β or all-β proteins, which involves the formation of long range interactions so that their native structures are reached, remains challenging and thus underexplored. Hence, SMFS studies of the folding-unfolding of these proteins remain limited to non-equilibrium unfolding reactions, where information about their folding energy landscape remains inaccessible. By leveraging recent improvements in AFM's long-term force stability by eliminating cantilever drift caused by its metal coating,[127] we directly observed the folding-unfolding of a small α/β protein NuG2 in real time under near-equilibrium conditions. These experiments also enabled us to obtain the first equilibrium free energy information on the folding-unfolding transition of NuG2 from single-molecule AFM experiments.

2.2 Results and Discussion

To examine the feasibility of observing the real-rime folding of α/β or all-α proteins, we chose the fast folding protein NuG2 as a model system. NuG2 is a computationally designed variant of the protein GB1 and assumes a characteristic β-grasp fold in which an α-helix packs against a four-stranded β-sheet (Figure 2.1).[122,134]

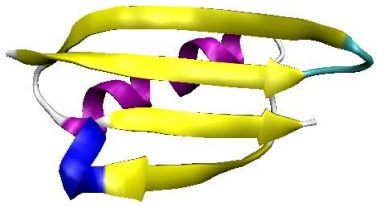

Figure 2.1 Three dimensional structure of NuG2 generated by VMD using PDB code 1MI0.

In our previous single-molecule AFM work, we constructed the polyprotein $(NuG2)_8$, which contains eight identical tandem repeats of NuG2.[135] Due to the poor long-term stability of force measurements arising from the cantilever drift, our previous studies on $(NuG2)_8$ were limited to force-induced unfolding reactions under non-equilibrium conditions.[135] In such experiments, the pulling speed had to be greater than 50 nm/s, as the force drift from the cantilever would overwhelm the experimental signal if the pulling speed were less than 50 nm/s. Figure 2.2 shows typical stretching and relaxation FECs of $(NuG2)_8$ obtained using a standard gold-coated Si_3N_4 cantilever (MLCT with a spring constant k_c = ~40 pN/nm). Stretching $(NuG2)_8$ resulted in FECs with a characteristic sawtooth-pattern appearance, where each sawtooth peak corresponds to the mechanical unfolding of one of the NuG2 domains in the polyprotein chain being stretched. Fitting the unfolding events of NuG2 using the worm-like-chain (WLC) model of polymer elasticity[73] measures a contour length increment (ΔL_c) of NuG2 of 18 nm upon complete unfolding, in good agreement with our previous measurements as well as the value calculated from NuG2's structure[135]. In contrast to the sawtooth-like FECs observed when NuG2 unfolds, refolding of this protein results in largely featureless spectra, with no clear refolding events at both pulling speeds of 400 nm/s and 50 nm/s. Large hysteresis exists between unfolding and refolding traces, suggesting that the unfolding of NuG2 at these pulling

speeds occurs under non-equilibrium conditions. In addition, drift in the force signal is evident in the FEC at 50 nm/s (Figure 2.2 bottom).

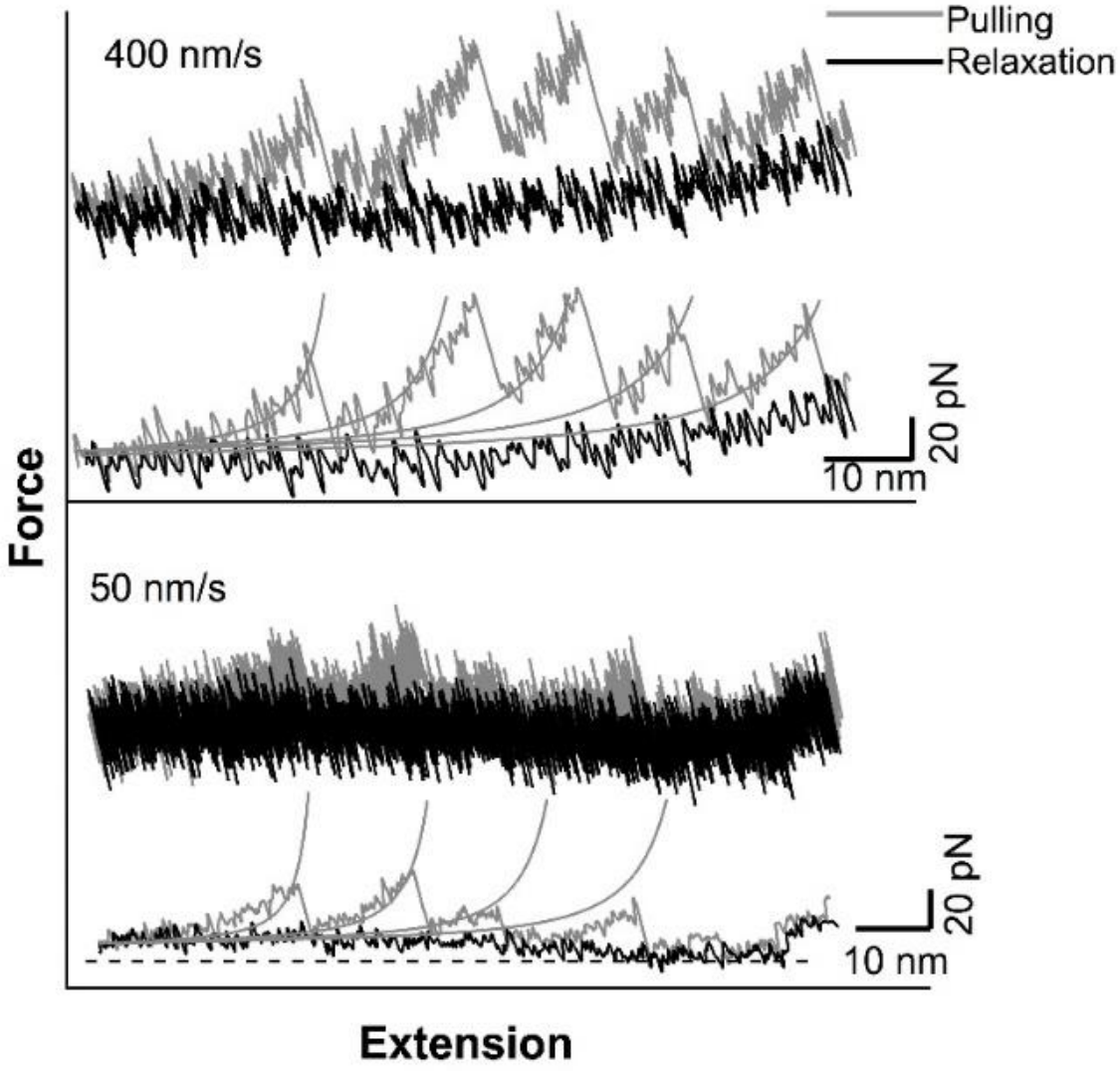

Figure 2.2 An FEC measured using conventional gold-coated Si_3N_4 cantilevers.
The sampling frequency for the top curves is 10 kHz and the bottom ones are low-pass filtered with a frequency of 500 Hz. Thin lines correspond to the global WLC fits with an ΔL_C of 18 nm. The force drift is evident in the bottom curves), as indicated by the dotted line and the deviation of the WLC fits from the data.

To improve the long-term stability of AFM force measurements, Churnside et al. developed a simple yet very effective method: they showed that removing the gold coating from the

cantilevers by wet etching can significantly reduce the force drift of the cantilever (Biolever, k_c = 6 pN/nm), and greatly improve the long-term stability of AFM without affecting the signal-to-noise ratio of force-extension measurements.[127] In this study, the gold coating of BioLever cantilevers from Olympus was removed using a wet-etching method as described previously[127] to increase the stability of the experimental system. SMFS experiments were performed on a Cypher AFM (Asylum Research). Figure 2.3A shows stretching and relaxation curves of NuG2 at a pulling speed of 400, 50 and 2 nm/s, respectively. In these spectra at 50 nm/s and 2nm/s, features slightly greater than the noise are clearly visible. After filtering the data to 500 Hz using a low-pass filter, sawtooth-like FECs are evident in both the stretching and relaxation traces. The sawtooth pattern observed during relaxation is very similar to that observed during stretching. This pattern clearly indicates that NuG2 displays refolding events during relaxation, which are manifested as the shortening of the taut polypeptide chain and a corresponding increase in force. Fitting refolding events using the WLC model revealed a ΔL_C of 18 nm, identical to that of NuG2 unfolding, suggesting that the refolding of NuG2 occurs in a two-state fashion. It is of note that there are several places in the FECs where the WLC fit deviates rather dramatically from the data. At such an exceedingly slow pulling speed (such as 2 nm/s), data does not always follow the WLC fit due to residual force drift and/or drift in the mechanical loop of the instrument. We occasionally observed rapid hopping events between single folded and unfolded NuG2 domains in both the stretching and relaxation curves (Figure 2.3B). During this hopping phenomenon, the protein transitions between two states, indicative of rapid unfolding-refolding transitions occurring close to equilibrium. Such a phenomenon has been observed in the folding-unfolding of calmodulin[126,130] and Ankyrin[54,131,132].

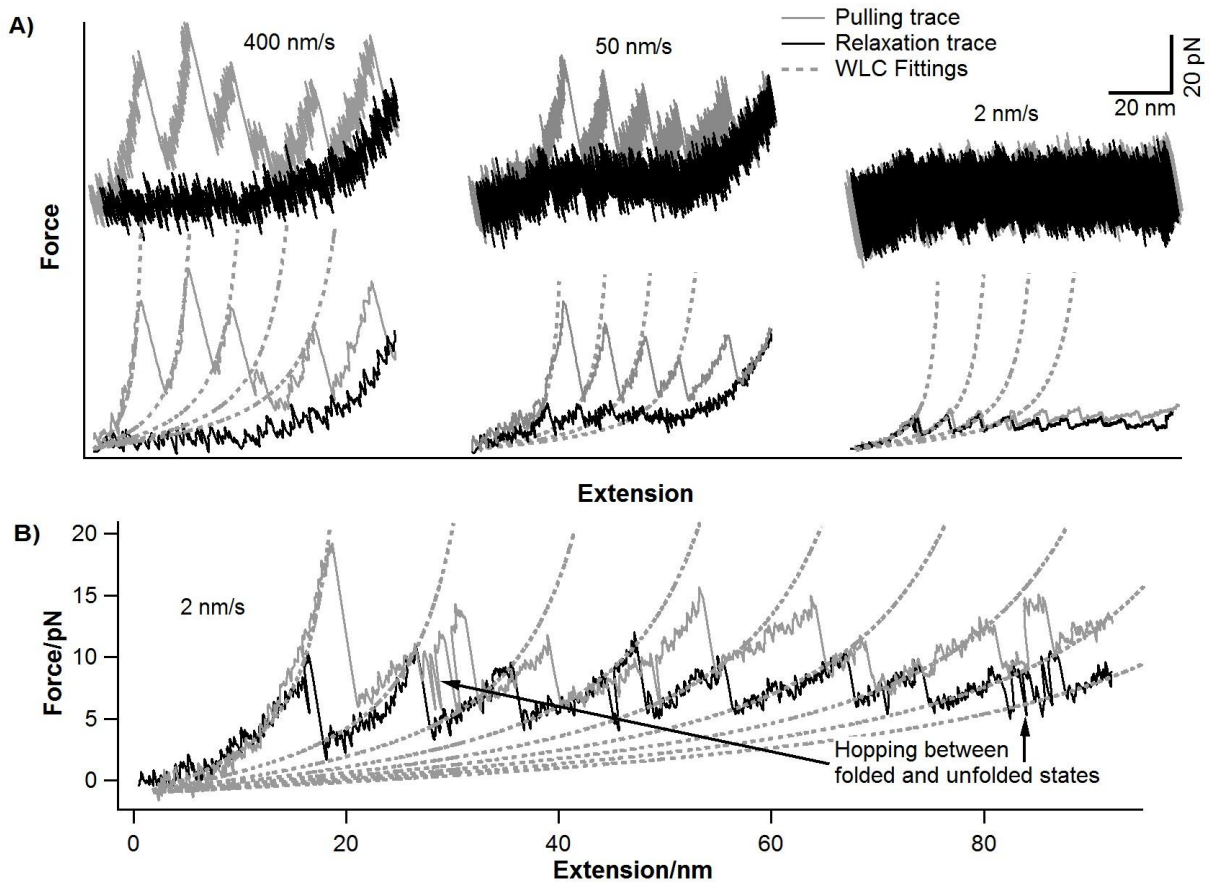

Figure 2.3 Improved SMFS experiments using uncoated cantilevers.

A) Comparison of pulling (solid grey) and relaxation (solid black) curves of $(NuG2)_8$ at different pulling speeds. Raw and filtered FECs are shown on the top and bottom, respectively. Dashed lines correspond to the global WLC fits with an ΔL_C of 18 nm. B) Typical low-pass-filtered pulling and relaxation traces of $(NuG2)_8$ at a pulling speed of 2 nm/s. Occurrences when the protein hops between folded and unfolded states are shown by arrows.

We carried out stretching-relaxation experiments at pulling speeds from 2 to 2000 nm/s, and measured the pulling speed dependence of the unfolding and refolding force, respectively (Figure 2.4). The hysteresis between unfolding and refolding forces (which represents the energy dissipation under the non-equilibrium conditions) becomes smaller upon a decrease in pulling speed. Additionally, the unfolding force of NuG2 decreases monotonically as the pulling speed decreases, while the refolding force of NuG2 increases as the relaxation speed decreases (Figure 2.4B). At the slowest pulling speeds, the unfolding and refolding forces of

NuG2 are close to each other, indicating that the unfolding and refolding are close to equilibrium. It is also noteworthy that the pulling speed dependence of unfolding force shows an inflection point at around 10 nm/s, indicating different pulling speed dependency of the unfolding forces at different pulling speeds. The unfolding force increases rapidly with increasing pulling speed at pulling speeds higher than >10 nm/s; however, the unfolding force becomes less sensitive to pulling speed at extremely slow pulling speeds (<10 nm/s). A similar change of the pulling speed dependence of unfolding (or unbinding) force has been reported in numerous previous studies and is often interpreted to be due to two different energy barriers in the unfolding/unbinding energy landscape or due to the shift of the unfolding pathway. [136-138] However, for NuG2, this change clearly results from the fact that unfolding and refolding of NuG2 occur close to equilibrium, rather than from two different energy barriers in the unfolding energy landscape or a switch of the pathway used for unfolding.[136,139]

In order to estimate the key parameters which characterize the unfolding/folding energy landscape of NuG2, Monte Carlo simulations that represent the exact experimental conditions were performed and the resultant data was used to "fit" the dependency between pulling speeds and unfolding/refolding forces.[51,75,136,140] Representative simulated FECs are shown in Figure 2.5. We found that an unfolding rate α_0 of 0.04 s^{-1} and the distance to the transition state Δx_u of 0.42 nm can describe the experimental data well, indicating that NuG2's mechanical unfolding transition state is very close to its native state. In contrast, a folding rate at zero force β_0 of 10000 s^{-1} and a folding distance Δx_f of 5 nm were found to be sufficient to describe the experimental data. This fast-folding rate makes it possible to observe how NuG2 hops between refolded and unfolded states in real-time. Analysis of these rates using $\Delta G^0 = k_B T \ln(\beta_0/\alpha_0)$

implies a free energy difference of ~12.4 k_BT between the native and unfolded states at zero force.

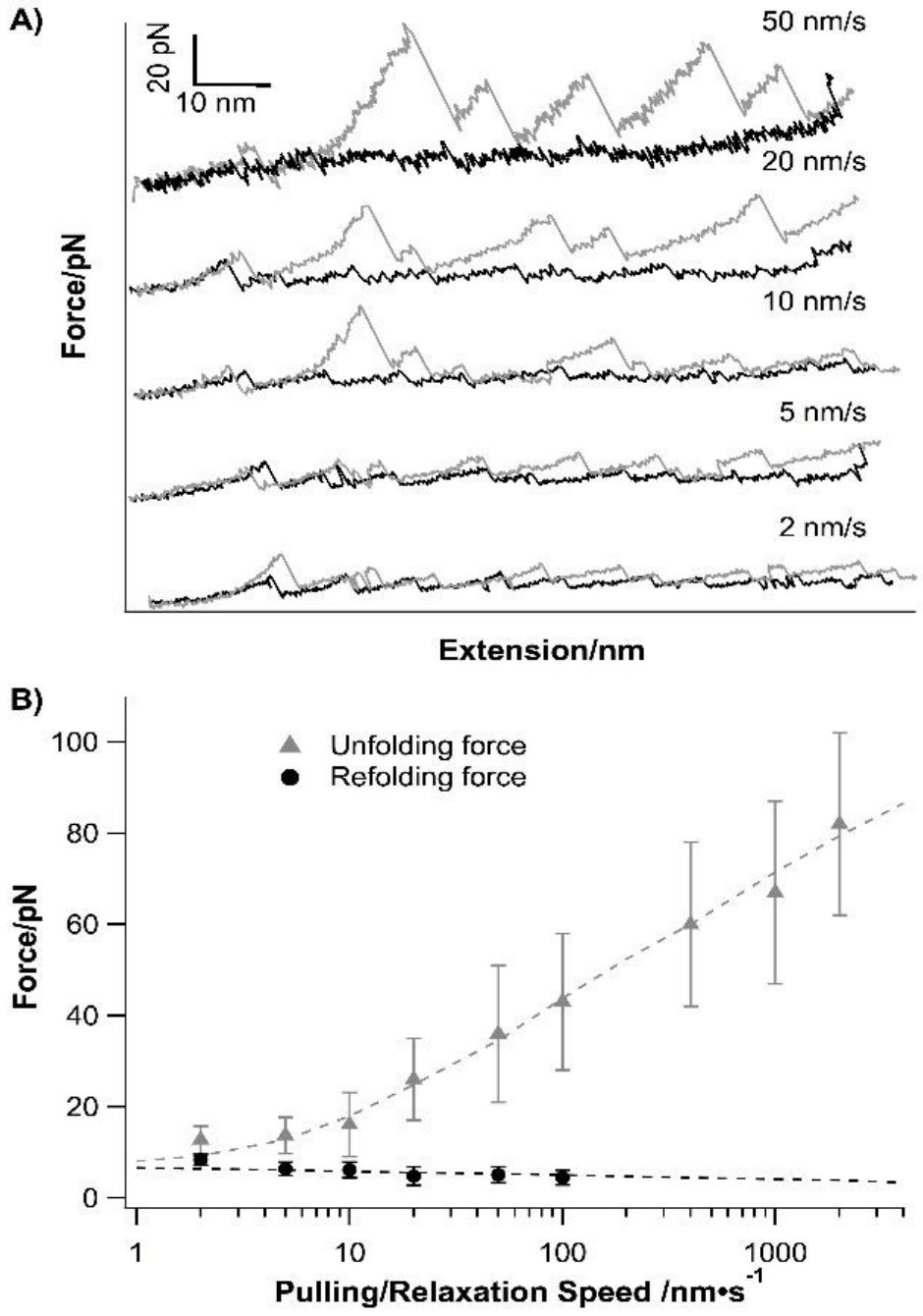

Figure 2.4 Pulling speed dependence of folding and unfolding.

A) FECs of $(NuG2)_8$ polyproteins obtained at different pulling speeds. Pulling curves and relaxation curves are in grey and black, respectively. B) Pulling/relaxation speed dependency of the mechanical unfolding/refolding force of NuG2. Dashed lines are Monte Carlo simulation results.

It is of note that Δx_u estimated here (0.42 nm) is slightly larger than that reported in our previous force-clamp spectroscopy study (0.25 nm), where stiffer cantilevers (with a spring constant of 40 pN/nm) were used than those in the current study (6 pN/nm). Previous force spectroscopy studies on calmodulin also showed similar discrepancies.[75,126] The reason underlying this difference is unclear and deserves a systematic investigation.

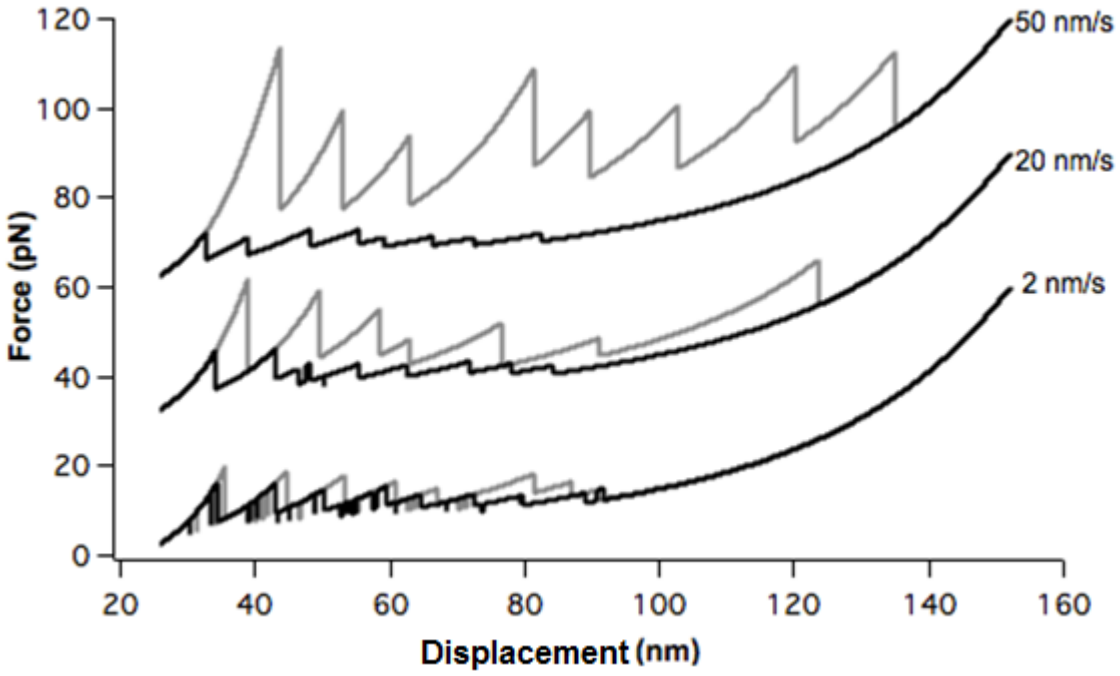

Figure 2.5 Representative force-displacement relationships obtained using Monte Carlo simulations. An unfolding rate α_0 of 0.04 s^{-1}, unfolding distance Δx_u of 0.42 nm and a folding rate at zero force β_0 of 10000 s^{-1} are used. The pulling speed is indicated in the figure.

Unfolding-refolding transitions of NuG2 observed in our AFM experiments were non-equilibrium in nature, and thus associated with hysteresis. Using fluctuation theorems, one can relate the work along non-equilibrium trajectories to thermodynamic free energy differences between states; these methods have been previously used in optical tweezers studies to obtain information on the free energy of proteins and nucleic acids secondary/tertiary structures.[141-143] Here we examine the feasibility of using Crooks fluctuation theorem (CFT) [141,142] to

estimate the equilibrium free energy difference ΔG^0 between folded and unfolded states from such non-equilibrium force vs. distance curves in AFM experiments. CFT allows one to relate this irreversible work done to unfold and refold a protein to ΔG^0 through the following equation:

$$\frac{P_{\mathrm{U}}(w)}{P_{\mathrm{R}}(w)} = \exp\left(\frac{w - \Delta G^0}{k_{\mathrm{B}}T}\right)$$

where $P_{\mathrm{U}}(w)$ and $P_{\mathrm{R}}(w)$ correspond to the work distribution associated with unfolding and refolding of the protein. The intersection of $P_{\mathrm{U}}(w)$ and $P_{\mathrm{R}}(w)$ yields the equilibrium free energy difference of ΔG^0 to be 15 $k_{\mathrm{B}}T$ between the folded and unfolded states of NuG2. Hence, CFT can be used in AFM-based force spectroscopy experiments that employ much stiffer probes (6 pN/nm) to determine the equilibrium free energy than those used in optical-trapping experiments (0.2 pN/nm), given sufficient instrumental stability.

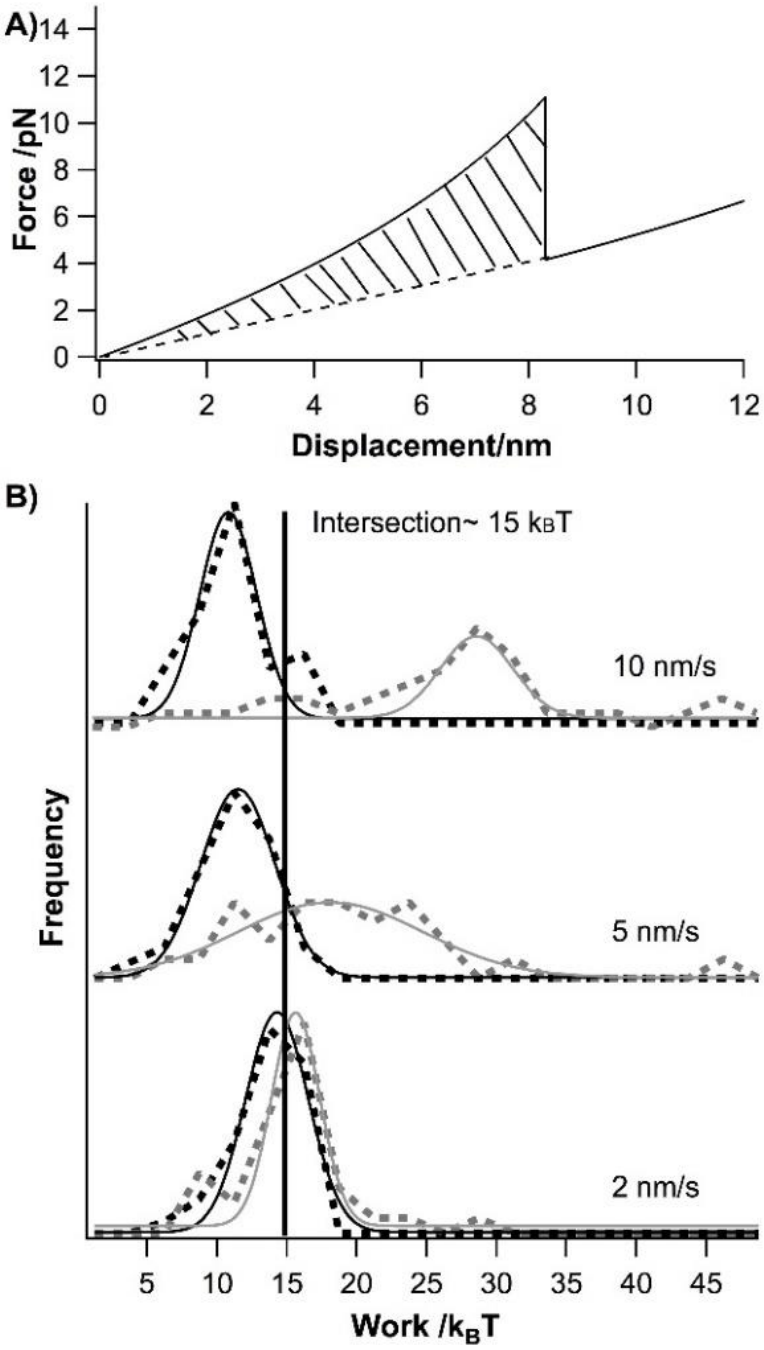

Figure 2.6 CFT analysis of NuG2 unfolding/folding.

A) A schematic showing the data acquisition process. The area enclosed by the experimental force peak (solid curve) and WLC fit (dashed curve) is shaded. B) Normalized histograms (dashed lines) represent unfolding (grey) and refolding (black) work. The intersection of unfolding and refolding work (~ 15 k_BT) is measured from the Gaussian fits (solid thin lines). Note: The Bennett acceptance ratio method[144] should be used to estimate the intersection more accurately. Here we use Gaussian fits for simplicity, although it may lead to certain errors.

In this study, we used to two different approaches (Monte Carlo and CFT) to determine the equilibrium free energy difference between the folded and unfolded NuG2, ΔG. The two approaches give similar results (12.4 k_BT from MC versus 15 k_BT from CFT). Both of these two different approaches have advantages and disadvantages. For Monte Carlo, the ΔG is essentially model-dependent, as the rate constants are estimated based on the Zhurkov-Bell-Evans model. If a different model is used to describe the force-dependent unfolding/folding reactions, the estimated rates will be slightly different, resulting in different ΔG. Thus, the error in estimating the unfolding/folding rate constant using MC approach is relatively large, [49] leading to a relatively large error in the estimated ΔG (a few k_BT). However, the advantage of Monte Carlo simulations is that one can estimate the ΔG even if the folding and unfolding are far from equilibrium. This has been the standard approach in AFM-based protein folding-unfolding studies.[126,140] On the other hand, the ΔG estimated using CFT is not model dependent. However, the CFT approach does require that the unfolding-folding are close to equilibrium so that more accurate ΔG can be estimated. The difference in the estimated ΔG using these two approaches may reflect the intrinsic differences of these two approaches and/or the error in estimating the rate constant using MC.

2.3 Experimental Section

2.3.1 Protein Engineering

The plasmid encoding NuG2 protein was a generous gift from Dr. David Baker from the University of Washington. The $(NuG2)_8$ polyprotein gene was constructed using a recombinant strategy, as previously described.[135] By digesting the pUC19/NuG2 plasmid with the restriction enzymes BamHI and KpnI, we obtained the NuG2 insert containing "sticky ends". The NuG2 insert was subsequently ligated by T4 DNA ligase to the corresponding "sticky

ends" of the pUC19/NuG2 vector, which was obtained by digesting the pUC19/NuG2 plasmid with BglII and KpnI. From this ligation, we obtained the pUC19/$(NuG2)_2$ plasmid. By repeating these steps, we obtained the $(NuG2)_8$ insert, which was ligated to pQE80L vector digested with BglII and KpnI to get the final pQE80L/$(NuG2)_8$ plasmid, which was then transformed into the expression host Escherichia coli DH5α. 300 mL Luria Broth (LB) medium , inoculated by 3 mL of the starter culture of the $(NuG2)_2$ construct, was incubated at 37 °C in a shaking incubator at 225 rpm. Proteins were over-expressed upon the induction by 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density (OD_{600}) of ~0.7 of the cell culture. Centrifugation was used to harvest the cells 4 hours after the induction. The proteins were extracted through the lysis of the freeze-thawed cells by lysozyme. Co^{2+} affinity chromatography (Clonetech) was used for protein purification. The purified proteins were stored at 4 °C in phosphate buffered saline (PBS) solution at a concentration of ~ 1 mg/mL.

2.3.2 SMFS using AFM

SMFS experiments were carried out on Cypher AFM from Asylum Research. Before each experiment, we calibrated the spring constant (~6 pN/nm) of each individual AFM cantilever (Biolever from Olympus) using the equipartition theorem.[145] In a typical AFM experiment, we deposited ~1.0 μL of protein solution (1.0 mg/mL) in PBS onto a clean glass slip covered by PBS buffer (~50 μL) and allowed the protein to adsorb onto the substrate for ~10 min before force spectroscopy experiments. Data analysis was accomplished using custom written codes in IGOR Pro 6. We used the worm-like chain (WLC) model of polymer elasticity [73] to fit consecutive unfolding force peaks to obtain the contour length increment upon domain unfolding. A persistence length of 0.4 nm, which is typical for unfolded polypeptide chains, was used in all WLC fittings.[35]

2.3.3 Monte Carlo Simulations

The mechanical unfolding of NuG2 was described using the Zhurkov-Bell-Evans model.[50-52,136] Force-dependent unfolding rate constants can be described as

$$\alpha(F) = \alpha_0 \exp(F\Delta x_u / k_B T)$$

where k_B is the Boltzmann constant, T is the temperature in Kelvin, $\alpha(F)$ is the unfolding rate constants at an external force of F, α_0 is the unfolding and refolding rate constants at zero force, and Δx_u is the distance from native to transition states. The refolding rate constant of NuG2 constant includes a non-exponential force dependence and is described as

$$\beta(F) = \beta_0 \exp(-\Delta G(F, F_T)/k_B T)$$

where $\beta(F)$ is the folding rate constant at an external force of F, β_0 is the folding rate at zero force, and $\Delta G(F, F_T)$ is the free energy change of the complete system (including the polypeptide chain and the cantilever), from the state prior to folding to the folding transition state.[75,146] The free energy of the polypeptide chain is calculated as integrals over the WLC, and the free energy of the cantilever is described as $F^2/(2k_C)$, where k_C is the spring constant of cantilever, F is the force and F_T is the force when the protein is at its folding transition state. We carried out Monte Carlo simulations according to previously described procedures[49,140] to reproduce the force-distance curves of stretching and relaxation of $(NuG2)_8$ at various pulling speeds and measured the average unfolding force and refolding force. In the calculations, both the elastic response of the polypeptide chain and the cantilever are included in the calculations. By comparing the MC data versus experimental data, the kinetic parameters α_0, β_0 and Δx_u, which define the mechanical unfolding and refolding of NuG2, are estimated.

Chapter 3: Mechanically Untying a Protein Slipknot

Protein structure is highly diverse when considering a wide range of protein types, helping to give rise to the multitude of functions that proteins perform. In particular, certain proteins are known to adopt a knotted or slipknotted fold. In chapter 2, we have demonstrated the power of AFM for studying protein folding/unfolding. In this chapter, how such proteins undergo mechanical unfolding was investigated utilizing a combination of single-molecule AFM, protein engineering, and SMD simulations to show the mechanical unfolding mechanism of the slipknotted protein, the ORF109 of the *Acidianus Filamentous Virus* 3 (AFV3-109). Our results reveal that the mechanical unfolding of AFV3-109 can proceed via multiple parallel unfolding pathways that all cause the protein slipknot to untie and the polypeptide chain to completely extend. These distinct unfolding pathways proceed via either a two- or three-state unfolding process involving the formation of a well-defined, stable intermediate state. SMD simulations predict the same contour length increments for different unfolding pathways as single molecule AFM results, thus providing a plausible molecular mechanism for the mechanical unfolding of AFV3-109. These SMD simulations also reveal that two-state unfolding is initiated from both the N- and C-termini, while three-state unfolding is initiated only from the C-terminus. In both pathways, the protein slipknot was untied during unfolding, and no tightened slipknot conformation was observed. Detailed analysis revealed that interactions between key structural elements lock the knotting loop in place, preventing it from shrinking and forming a tightened slipknot conformation.

Chapter 3 has been published as [He CZ], Genchev GZ, Lu H, & Li HB Mechanically Untying a Protein Slipknot: Multiple Pathways Revealed by Force Spectroscopy and Steered Molecular Dynamics Simulations. J. Am. Chem. Soc., 2012, 134 (25), 10428–10435. This chapter is incorporated in this thesis with permission of American Chemical Society, Copyright (2012).

Our results demonstrate the bifurcation of the mechanical unfolding pathway of AFV3-109 and point to the generality of a kinetic partitioning mechanism for protein folding/unfolding.

3.1 Slipknotted Topology of AFV3-109

Over the last two decades, theoretical and experimental work has significantly advanced our understanding of how proteins fold.[1-4,6,7] Proteins navigate themselves on a high-dimensional funnel-shaped energy landscape to achieve robust and fast folding. Most proteins do not possess complex topology, where unfolding most proteins from its N- and C-termini will simply result in a simple linear polypeptide chain.[9] However, a small proportion of proteins have been observed to exhibit complex topologies involving the formation of knots and slipknots.[106] Upon being stretched from N- and C-termini, such proteins would form a tightened knot instead of a linear polypeptide chain. Despite their complex topology, these knotted or slipknotted proteins can spontaneously fold into their native conformations in order to carry out their designated biological functions.[107] Understanding how a linear polypeptide chain is able to overcome the topological difficulty inherent within these proteins to fold into such complex topologies has attracted considerable research interest over the past few years. Both experimental and theoretical work has been used to shed light on the folding mechanism of such proteins.[16,98,101,103,105,107,108,116,117,121,147] In particular, computational studies have provided certain mechanistic insights into the folding mechanisms of these proteins. For example, the formation of a slipknot conformation was suggested as a key intermediate state to reduce the topological complexity during the spontaneous folding of a knotted protein, revealing the importance of the slipknot conformation in the folding of knotted proteins.[107,108] Technically, a slipknot is not a real knot. In a slipknot, the end of the polypeptide chain is threaded back after being threaded through the knotting loop, forming a threaded loop akin to

a shoelace (Figure 3.1). In this case, pulling on the N- and C-termini of a slipknotted protein will untie the slipknot and lead to a fully extended linear polypeptide chain. Due to its structural importance and its role in the folding of knotted proteins, we combined SMFS and SMD simulations to investigate the unfolding mechanism of the slipknotted protein AFV3-109.

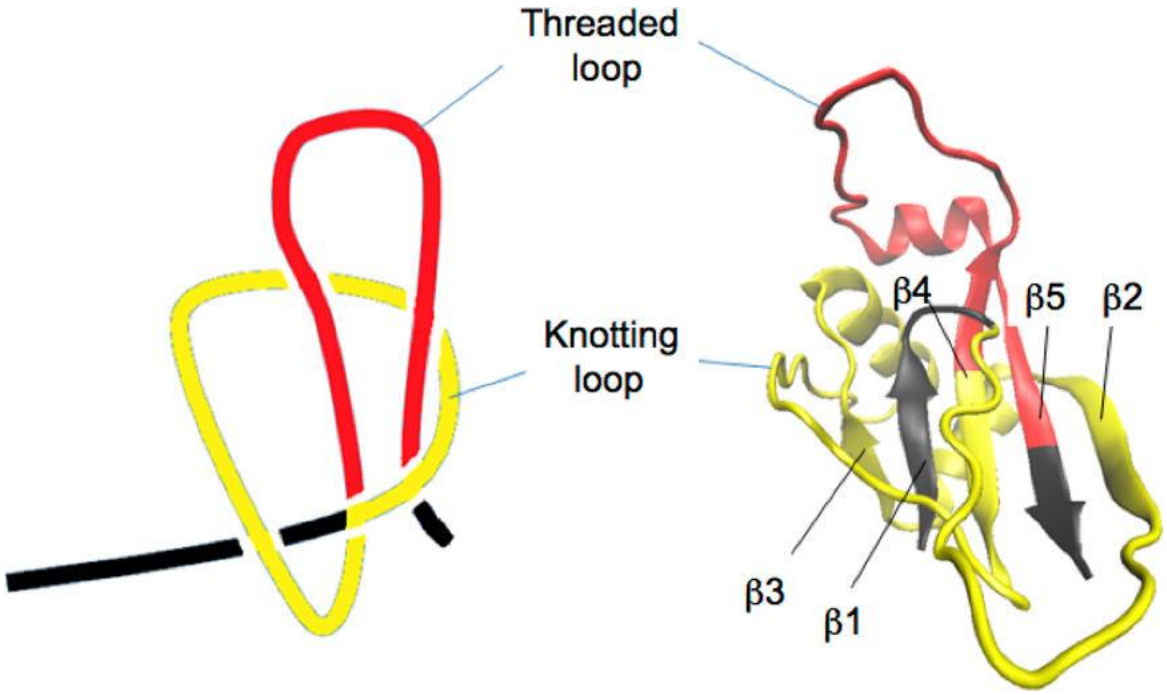

Figure 3.1 The three-dimensional structure of AFV3-109.

AFV3-109 is an α/β protein. A five stranded β-sheet connected by helices and loops forms the core of AFV3-109. The order of the five strands is β3β1β4β5β2, where β5 is antiparallel to the others. The knotting loop (colored in yellow) encompasses amino acid residues 8−77, including β2, α1, β3, α2, and part of β4. Amino acid residues from 78 to 106, including part of β4, α3, and β5, are threaded through the knotting loop twice and form a threaded loop (colored in red).

AFV3-109 is a protein from *acidianus filamentous virus 3* with unknown biological function.[148] Among the known slipknotted proteins, AFV3-109 has a relatively simple structure. As shown in Figure 3.1, the core of AFV3-109 is a β-sheet with five β-strands connected by helices and loops. A clear slipknot conformation is present in AFV3-109 with a knotting loop (colored in yellow, Figure 3.1) near the N- terminus and a threaded loop (colored

in red, Figure 3.1) near the C- terminus. As the "reverse" process of folding, the unfolding of slipknotted proteins will involve untying the slipknot. Understanding this molecular mechanism of untying a slipknot will help to understand how the protein folded in the first place. Applying a mechanical stretching force to the slipknotted protein, either through SMFS experiments or SMD simulations, is an ideal approach to investigating this intricate process.

SMFS has evolved into a powerful tool to probe the mechanical unfolding and folding dynamics of proteins at the single molecule level.[2,60,149-153] SMD simulations complement SMFS experiments and provide invaluable insights into describing atomic level molecular events that occur during the mechanical unfolding process, while helping to guide new experimental designs.[151-157] Both single-molecule AFM and simulation have been used to stretch knotted proteins in order to study their mechanical unfolding mechanism. These simulations suggest that the presence of a knot and slipknot in protein structures stabilizes the protein and that the stretching of slipknotted proteins may involve the formation of an intermediate state comprising a temporarily tightened slipknot during the mechanical unfolding process.[121] Single molecule AFM experiments showed that the mechanical stretching of phytochrome, a protein with a figure-eight knot, leads to a tightened knot that involves 17-19 residues.[116] Here we apply a stretching force to the N- and C-termini of the slipknotted protein AFV3-109 in order to investigate its mechanical unfolding and folding mechanism. Due to its slipknot structure, we believe that AFV3-109 can be unfolded and untied just as a shoelace is untied. Our investigations using single molecule AFM clearly show that the mechanical unfolding of AFV3-109 unties the slipknot structure, leading to a fully extended polypeptide chain. The mechanical untying of the slipknot can occur through three distinct mechanical unfolding pathways: one in a simple two-state fashion, and the other two

in a three-state fashion that involves the formation of a well-defined intermediate state. SMD simulations of the mechanical unfolding of AFV3-109 reveal a similar bifurcation phenomenon, providing a plausible molecular explanation of the mechanical unfolding process observed in single molecule AFM experiments. These simulations show that the two-state unfolding process initiates simultaneously from the N- and C-termini, while the three-state unfolding process initiates only from the C-terminus. In the three-state unfolding mechanism, the formation of the unfolding intermediate results from the unraveling of the C-terminal threaded β-hairpin loop, while final unfolding corresponds to a complete unraveling of the intermediate state. The agreement between single molecule AFM and SMD simulation results provides a detailed molecular picture of how the slipknotted protein AFV3-109 unfolds; these results contribute toward a clearer understanding of how slipknotted and knotted proteins unfold and fold.

3.2 Results

3.2.1 The Slipknot in AFV3-109 Can Be Mechanically Untied.

To investigate the mechanical unfolding of AFV3-109 using single molecule AFM, we constructed two different AFV3-109-containing polyprotein chimeras $(GB1)_4$-AFV3-109-$(GB1)_4$, referred as G4-A-G4, and $(GB1\text{-}AFV3\text{-}109)_n$ (n typically varies from 1 to 6), referred as $(G\text{-}A)_n$. In both polyprotein chimeras, mechanically well characterized GB1 domains are used as fingerprint domains to facilitate the identification of single molecule stretching events and discerning the mechanical unfolding signatures of AFV3-109, as GB1 unfolds at a characteristic force of ∼180 pN with a contour length increment (ΔL_C) of 18 nm at a pulling speed of 400 nm/s.[158]

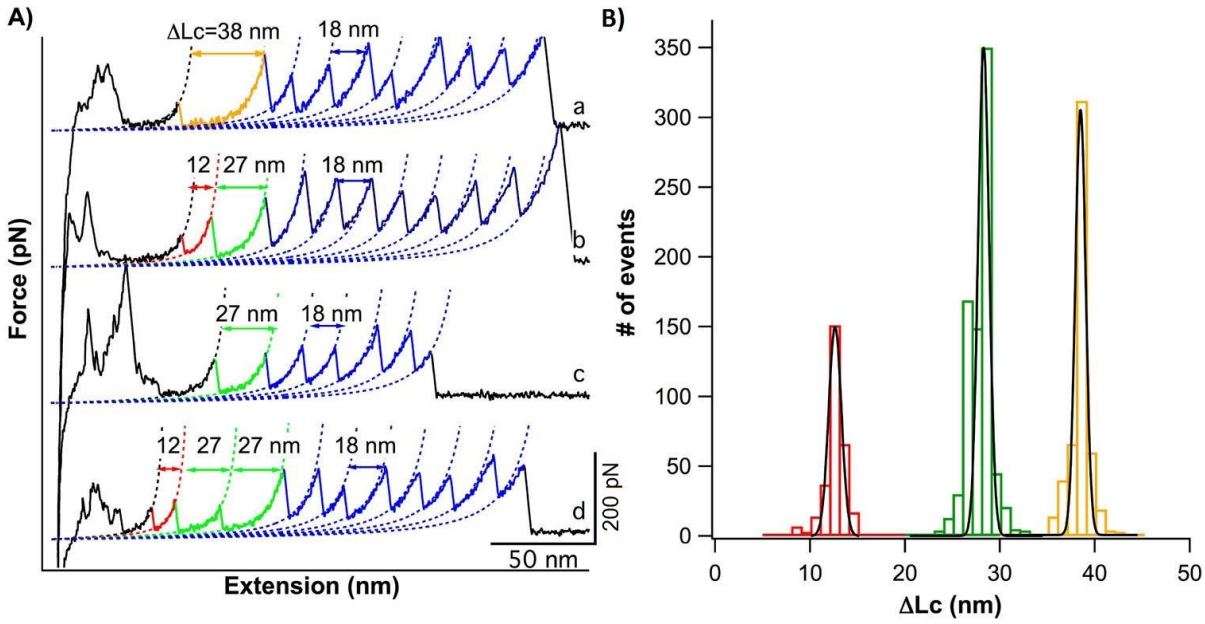

Figure 3.2 The mechanical unfolding of AFV3-109.

(A) Force−extension curves of G4-A-G4. Dotted lines are WLC fits to the force−extension curves. The unfolding events of the GB1 fingerprint domains are characterized by a ΔL_C of ~18 nm (colored in blue). The mechanical unfolding of AFV3-109 can proceed through multiple pathways: in curve (a), the unfolding of AFV3-109 is an apparent two-state unfolding event with a ΔL_C of 38 nm; and in curve (b), the unfolding of AFV3-109 proceeds in a three-state unfolding pathway involving a stable intermediate state, with a $\Delta L_{C(N-I)}$ of 12 nm and a $\Delta L_{C(I-U)}$ of 27 nm. In some cases, the N−I unfolding event was not observed, and only the I−U unfolding was observed (curve c). For clarity, the two-state unfolding event is colored in orange, the N−I event is colored in red and the I−U event is colored in green. (B) ΔL_C histograms for mechanical unfolding events of AFV3-109. Gaussian fits to the experimental data show a ΔL_C of 12 ± 1 nm (n = 298, red), 27 ± 1 nm (n = 790, green), and 38 ± 1 nm (n = 521, orange).

Stretching the polyprotein $(GB1)_4$-AFV3−109-$(GB1)_4$ yielded force−extension curves with a characteristic saw-tooth pattern, where the force peaks correspond to the mechanical unfolding of individual domains in the polyprotein chain, and the last peak generally corresponds to the stretching and subsequent detachment of the unfolded polypeptide chain. Figure 3.2A shows representative curves from the unfolding of the G4-A-G4 polyprotein. The unfolding of GB1 fingerprint domains (colored in blue) can be readily identified by their unique unfolding signatures: unfolding force of ~180 pN and ΔL_C of ~18 nm measured by fitting the worm-

like-chain model of polymer elasticity to consecutive unfolding force peaks. [73,159] As AFV3-109 is flanked by two $(GB1)_4$ at both N- and C-termini, the observation of five or more GB1 unfolding events in a force−extension curve ensures that the unfolding signatures of AFV3-109 are included within the force−extension curve. Therefore, unfolding force peak(s) occurring before GB1 domain unfolding can be attributed to the mechanical unfolding of the AFV3-109 domain within the polyprotein chain.

In Figure 3.2A, curve (a), the unfolding of AFV3-109 occurs in an apparent two-state fashion with a ΔL_C of ∼38 nm (colored in orange). As there are 109 residues in AFV3-109, the contour length of the unfolded and fully extended AFV3-109 is 39.2 nm when there is no knot or tightened slipknot. The distance between the N- and C- termini of AFV3-109 in the native state, L_0, is 1.2 nm (PDB code 2j6b). Therefore, the complete unfolding of AFV3-109, from its slipknotted structure to a fully extended polypeptide chain, would result in a ΔL_C of ∼38 nm, which is in excellent agreement with the experimentally measured ΔL_C. This result suggests that upon stretching, AFV3-109 can be mechanically unraveled, and the slipknot structure is fully untied in a manner akin to how a shoelace is untied.

3.2.2 Untying the Slipknot in Multiple Ways: Bifurcation of the Mechanical Unfolding Pathways of AFV3-109

In addition to two-state unfolding, we also observed that the mechanical unfolding of AFV3-109 can proceed via different pathways. Figure 3.2A, curve (b) shows an example in which the mechanical unfolding of AFV3-109 proceeds in a three-state fashion, with the formation of an intermediate state that gives rise to two unfolding force peaks for one AFV3-109 domain. The first unfolding force peak of the three-state unfolding occurs at ∼90 pN with a $\Delta L_{C(N\rightarrow I)}$ of ∼12 nm, corresponding to the transition from the native state to an unfolding intermediate state

(N→I) (colored in red). The second step occurs at ~100 pN with $\Delta L_{C(I\rightarrow U)}$ of ~27 nm, corresponding to unfolding of the intermediate state (I→U) (colored in green). The sum of the two ΔL_C is 39 nm, which is in agreement with the predicted ΔL_C of the complete unfolding of AFV3-109. This result again suggests that AFV3-109 has been completely unraveled, and the slipknot has been untied. It is of note that the unfolding intermediate state is mechanically stable and unfolds at ~100 pN.

In addition to two- and three-state unfolding, we also observed AFV3-109 unfolding events that show a $\Delta L_{C(I\rightarrow U)}$ of only ~27 nm (Figure 3.2A, curves c and d), suggesting that the unfolding of AFV3-109 in this pathway begins from the same unfolding intermediate state as in the three-state unfolding pathway. It is likely that this unfolding scenario is a special case of the three-state unfolding pathway, in which part of AFV3-109 in the native state is unfolded prior to stretching or is unfolded at forces that are below our 20 pN detection limit. The ΔL_C histogram of all mechanical AFV3-109 unfolding events displays three well-defined peaks (Figure 3.2B). Fitting this data using a Gaussian fit shows three ΔL_C of 12 ± 1 nm (n = 298), 27 ± 1 nm (n = 790), and 37 ± 1 nm (n =521), corresponding to the contour length increments deriving from the N→I, I→U, and N→U unfolding steps. These differences in unfolding behavior clearly indicate that the mechanical unfolding of AFV3-109 can proceed through multiple unfolding pathways, specifically, and that such pathways are bifurcated. Since AFV3-109 is completely unfolded in all three pathways, our results suggest that there are multiple ways to untie the slipknot and fully extend AFV3-109. The three distinct unfolding pathways occur roughly at a frequency of 20% (N→I→U):40% (I→U):40% (N→U) (Figure 3.2B), revealing the kinetic partitioning that occurs as AFV3-109 is mechanically unfolded.

3.2.3 The Mechanical Unfolding of AFV3-109 Does Not Involve the Formation of a Tightened Slipknot.

Coarse grained MD simulations[121] on slipknotted protein thymidine kinase suggested that during the mechanical unfolding of a slipknotted protein, a high stretching force could result in molecular jamming and lead to the formation of a tightened slipknot as an unfolding intermediate state. This possibility raises the question whether the unfolding intermediate state (with a $\Delta L_{C(I\rightarrow U)}$ of 27 nm) observed in our AFM experiments on AFV3-109 could correspond to a tightened slipknot conformation.

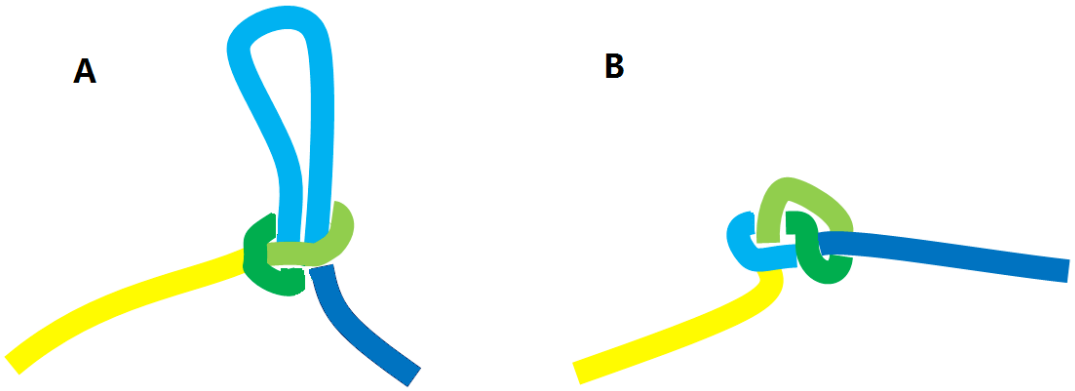

Figure 3.3 Comparison between the tightened slipknot of AFV3-109 (A) and tightened figure-eight knot in phytochrome (B).

The tightened figure-eight knot is more complex and bigger than the tightened slipknot. The tightened figure-eight knot contains 17 residues and is ~6.2 nm in size.

If a tightened slipknot could form in the mechanical unfolding of AFV3-109, $\Delta L_{C(I\rightarrow U)}$ of 27 nm would correspond to the length increment due to untying the tightened slipknot, which should be the sum of the length of the polypeptide chain trapped in the tightened slipknot and the size of the tightened slipknot itself. The polypeptide chain trapped in the tightened slipknot would include residues 73−106 (β4, α3, and part of β5) and would be ∼12 nm long. Thus, the

size of the tightened slipknot would be ~15 nm (27−12 nm). Considering the complexity of this tightened slipknot and tightened figure-eight knot (Figure 3.3), the size of a tightened slipknot should be no bigger than a tightened figure-eight knot, which has been estimated to be 6.2 nm using single molecule AFM and SMD simulations. [116] Therefore, 15 nm is unphysically large for a tightened slipknot conformation. This result strongly suggests that the unfolding intermediate state of AFV3-109 observed in our single molecule AFM experiments does not correspond to the tightened slipknot, and the unfolding of AFV3-109 does not involve the formation of a tightened slipknot conformation.

3.2.4 The Fusion of $(GB1)_4$ to the C-Terminus of AFV3-109 Does Not Impede the Folding of AFV3-109.

In the slipknot structure of AFV3-109, the threaded loop is at the C-terminus. Since the G4-A-G4 polyprotein is expressed as a continuous polypeptide chain, it is unknown whether the fusion of $(GB1)_4$ at the C-terminus of AFV3-109 interferes with the intrinsic folding of AFV3-109. To ensure that the fusion of $(GB1)_4$ to the C-terminus of AFV3-109 does not adversely affect the intrinsic folding of AFV3-109, we constructed the polyprotein $(GB1\text{-}AFV3\text{-}109)_n$ (n typically varies from 3 to 6) based on the polymerization of Cys-GB1-AFV3-109-Cys using thiol-maleimide coupling chemistry (see experimental section).[160] In this construct, the C-terminus of AFV3-109 is free, and the folding of AFV3-109 should not be affected as the polyprotein $(GB1\text{-}AFV3\text{-}109)_n$ is formed only after AFV3-109 has folded properly.

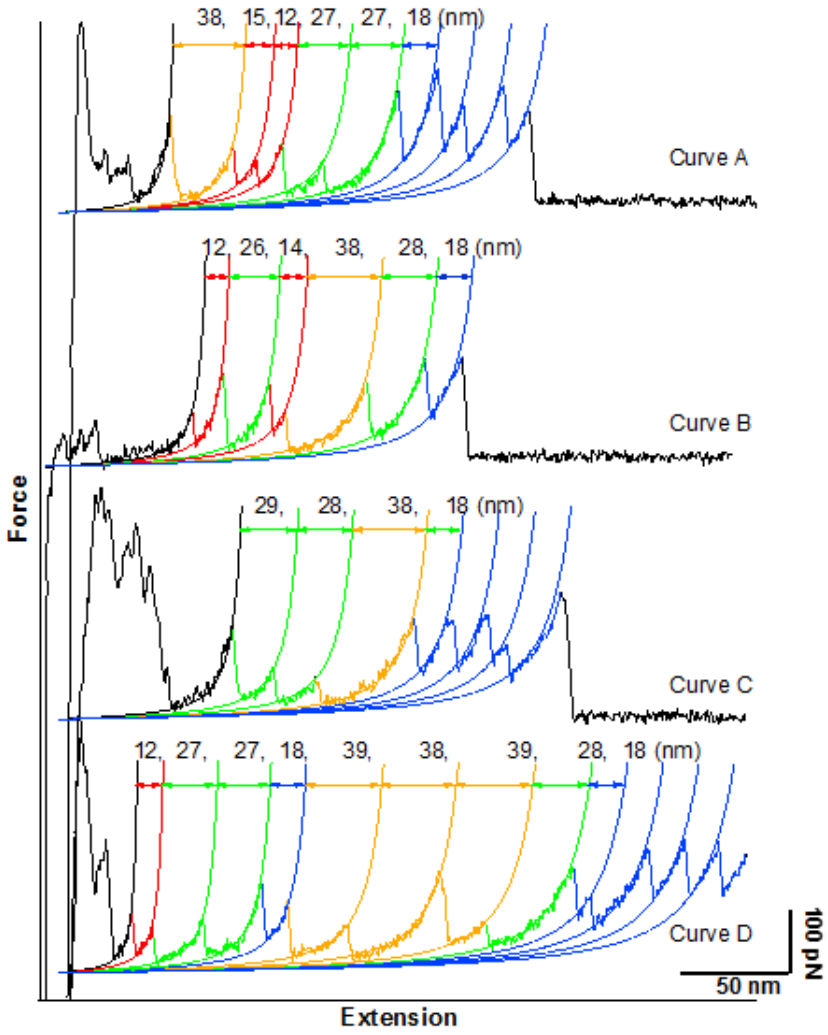

Figure 3.4 FECs of (GB1-AFV3-109)$_n$.

The mechanical unfolding of GB1 fingerprint domains is colored in blue and characterized by a ΔL_C of 18 nm. The unfolding events of AFV3-109 are colored according to their unfolding pathways: two-state unfolding is colored in orange, N-I in red and I-U in green, respectively.

Force−extension curves of (GB1-AFV3-109)$_n$ are shown in Figure 3.4. We observed that the unfolding signatures exhibited by AFV3-109 in (GB1-AFV3-109)$_n$ are identical to those observed in the original (GB1)$_4$- AFV3-109-(GB1)$_4$ polyprotein; specifically, that in multiple unfolding pathways, the same average unfolding force and contour length increment are all demonstrated. In addition, we noticed that in force−extension curves containing more than one

unfolding event of AFV3-109, a N→I unfolding event is not necessarily followed by an I→U unfolding event. The N→I unfolding of a different AFV3-109 domain can occur between the N→I and I→U events of the preceding AFV3-109 domain, indicating that the mechanical stability of the intermediate state is comparable to that of the native state. This result strongly suggests that the C-terminal fusion protein containing $(GB1)_4$ does not adversely influence the formation of the slipknot structure in AFV3-109 and that the folding of AFV3-109 does not involve complex chain crossing.

3.2.5 Signatures of the Mechanical Unfolding of AFV3-109

Having identified the mechanical unfolding pathways and mechanical unfolding intermediate state, we can now characterize these unfolding pathways in detail. The average unfolding force is 110 ± 40 pN for N→U (n = 521), 90 ± 30 pN for N→I (n = 298), and 100 ± 40 pN for I→U (n = 790) at a pulling speed of 400 nm/s (Figure 3.5A, C and E). It is evident that the three unfolding steps have similar mechanical resistance. To characterize the underlying energy profiles of these distinct unfolding pathways, we carried out pulling experiments at different pulling velocities (Figure 3.5B, D and F). The pulling speed dependence of the unfolding force is similar for N→U and I→U, but that the N→I transition differs significantly, which may be suggestive of the large difference in transition states in these unfolding steps.

We carried out Monte Carlo simulations[49,161] to simulate the unfolding behaviors of AFV3-109 in order to estimate the distance to the transition state, Δx_u, and the spontaneous unfolding rate constant at zero force, α_0, for each individual unfolding step (Figure 3.5B, D and F). These results are shown in Table 3.1. It is worth noting that Δx_u for N→U (0.24 nm) is significantly smaller than that of N→I (0.59 nm), suggesting that the N→U pathway is fundamentally different from that of N→I, eliminating the possibility that the N→U pathway is a special case

of the N→I→U pathway, where the intermediate state was not detected due to insufficient temporal resolution of this method.

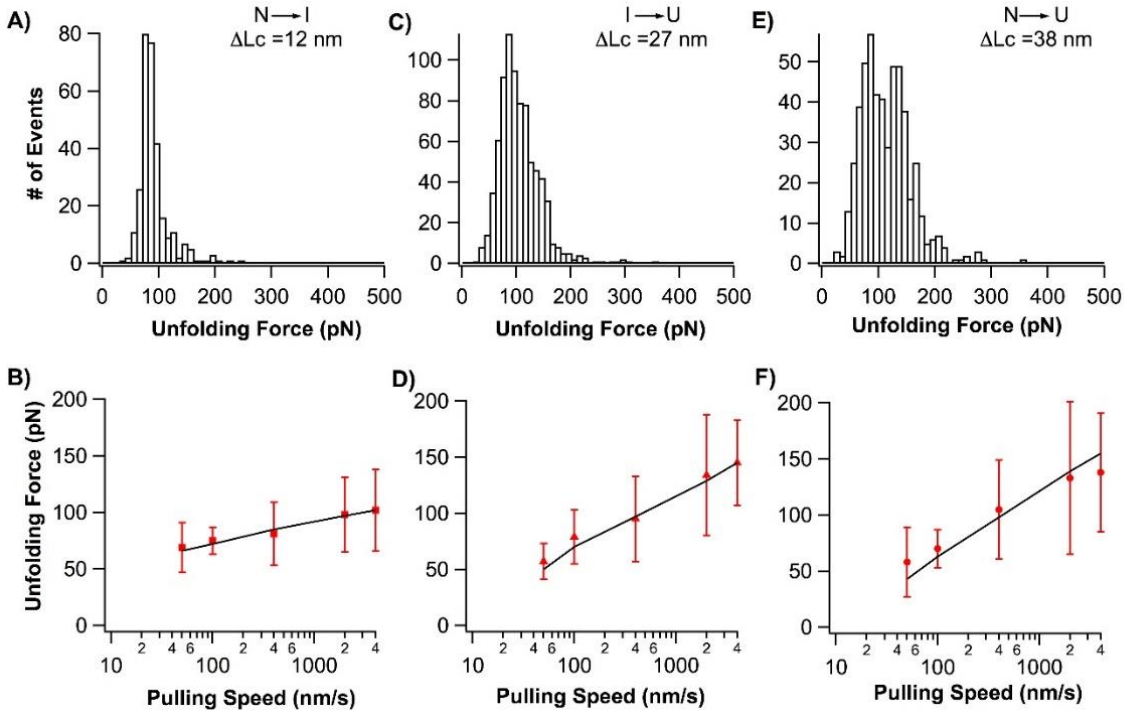

Figure 3.5 Signatures of the mechanical unfolding of the slipknotted protein AFV3-109.
(A,C,E) Unfolding force histograms for the unfolding step N→I (A), I→U (B), and N→U (C). (B,D,F) Pulling speed dependence of the unfolding force for each individual unfolding step. Solid lines correspond to the Monte Carlo simulation results using α_0 and Δx_u tabulated in Table 1. The red bars represent the standard deviation of unfolding forces.

Table 3.1. Kinetic parameters characterizing the mechanical unfolding of AFV3-109

Unfolding Event	ΔL_C (nm)	α_0 (s^{-1})	Δx_u (nm)
N→U pathway	38	1.8	0.24
N→I→U pathway			
N→I	12	0.011	0.59
I→U	27	0.72	0.27
N···I→U pathway			
N→I	N/A	N/A	N/A
I→U	27	0.72	0.27

3.2.6 SMD Simulations Uncover the Molecular Mechanism Behind Bifurcation of Mechanical Unfolding Pathways of AFV3-109.

To understand the molecular mechanism underlying the bifurcation of mechanical unfolding pathways and the nature of the mechanical unfolding intermediate state of AFV3-109, we carried out all-atom SMD simulations to simulate the mechanical unfolding of AFV3-109. In total, 23 simulations were carried out with a total duration of 387 ns.

The first event that occurs when a steering force is applied is the unzipping of antiparallel strands β4β5, where the five-strand-β sheet is separated into two parts consisting of β3β1β4 and β2β5 strands respectively, forming a quasi-native state (Figure 3.6). The extension gain from unzipping is approximately 3 nm (where the unzipping of β4β5 is completed when a total N-to-C extension of 4.5 nm is reached). Upon continued application of stretching force, the unfolding of AFV3-109 exhibits two distinct types of pathways for all unfolding trajectories: one with an unfolding intermediate state and one without. These results are summarized in Figure 3.6, and a description of these pathways is included below.

In the first pathway, unfolding initiates from the C-terminus. The applied steering force leads to unzipping of antiparallel strands β2β5 as well as the sliding of α3 through the loop formed by β2 and α1 (snapshot 1 in Figure 3.6). This causes the β5α3β4 arm to become fully extended, fully exposing the β1β4 parallel strands to the steering force. During this process, the threaded loop is pulled out of the knotting loop, leading to the untying of the slipknot in AFV3-109. The resultant state is mechanically resistant and serves as a well-defined intermediate state (shown as I in Figure 3.6). Further stretching AFV3-109 leads to the unraveling of this intermediate state. The major barrier during this process corresponds to the mechanical unraveling of the

β1β4 (snapshot 2 in Figure 3.6); following the breakage of the β1β4 barrier, there is no further mechanical barrier, and the protein stretches to a fully extended state.

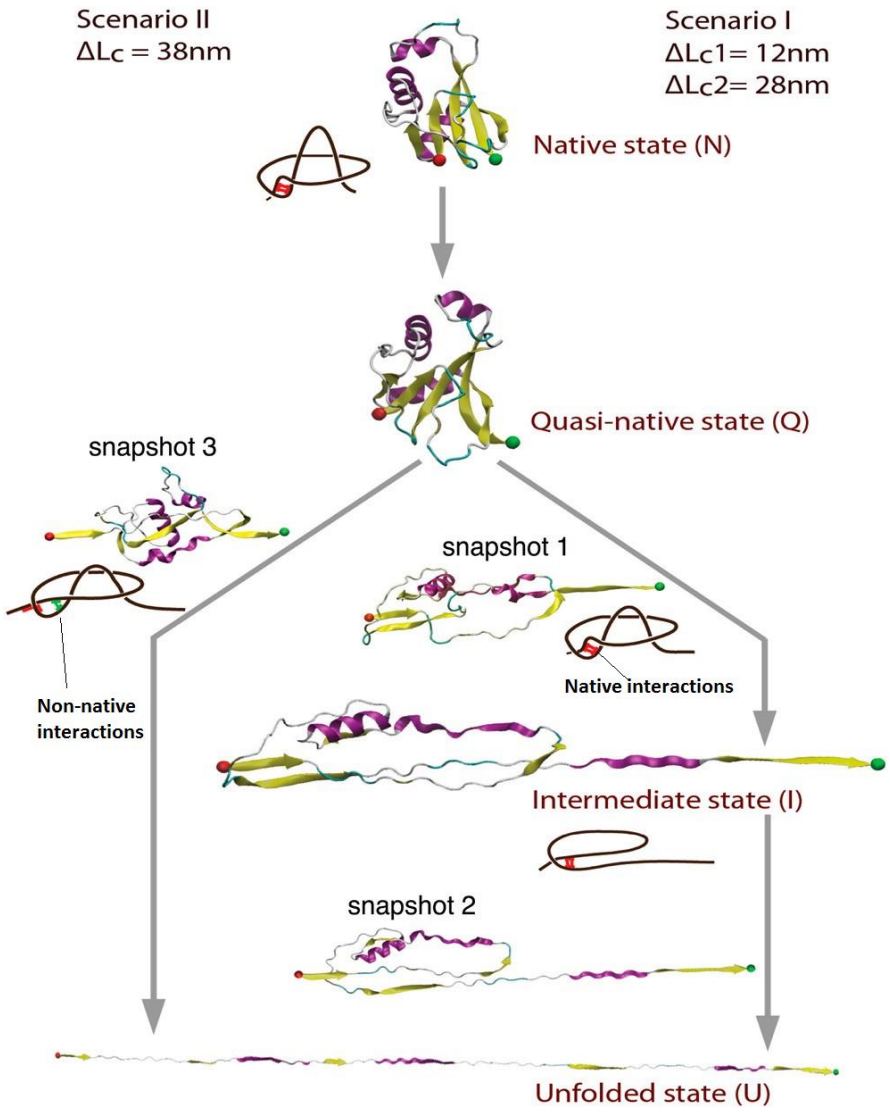

Figure 3.6 SMD simulations reveal two distinct mechanical unfolding pathways of AFV3-109.

Scenario I corresponds to the unfolding pathway involving an unfolding intermediate state, and scenario II corresponds to the two-state unfolding pathway. Along each step of the unfolding process, snapshots of AFV3-109 are shown (as indicated by snapshots 1−3) to indicate the structural changes of AFV3-109 during the unfolding process. In both scenarios, the slipknot structure was untied, and no tightened slipknot conformation observed. Interactions (specific or nonspecific) between key structural elements are responsible for preventing the formation of the "tightened knot" structure. Such interactions, which are indicated in the schematic drawing of AFV3-109 beside the snapshots, prevent the shrinking of the knotting loop and help untie the slipknot. Red bars indicate specific interactions between strands β1 and β4, while green bars indicate the non-native interactions formed after the sliding of strands β1 and β4.

Detailed analysis of trajectories following this type of pathway reveals that the intermediate state I always occurs at around 13 nm. However, the initial peak occurs at a much lower frequency, and in some trajectories, this step does not lead to an observable unfolding force peak (Figure 3.7). Comparing Figure 3.7 with the contour length figure of the AFM data, we can conclude the two peaks at 3 and 13 nm correspond to the N→I and I→U transitions, respectively. Additionally, there are typically more peaks than accounted for by the initial and intermediate peaks. Figure 3.7 shows a typical trajectory with three extra peaks at 4.5, 5.5, and 7 nm. To further study these peaks, we collected peaks from all trajectories (inset of Figure 3.7). We found that these extra peaks are mostly caused by friction that occurs when pulling the threaded loop out of the knotting loop; friction originates from nonspecific interactions, which are different in different trajectories, and thus occurs at different locations during different unfolding events. In contrast, the initial (N→I) and intermediate (I→U) peaks were due to specific interactions.

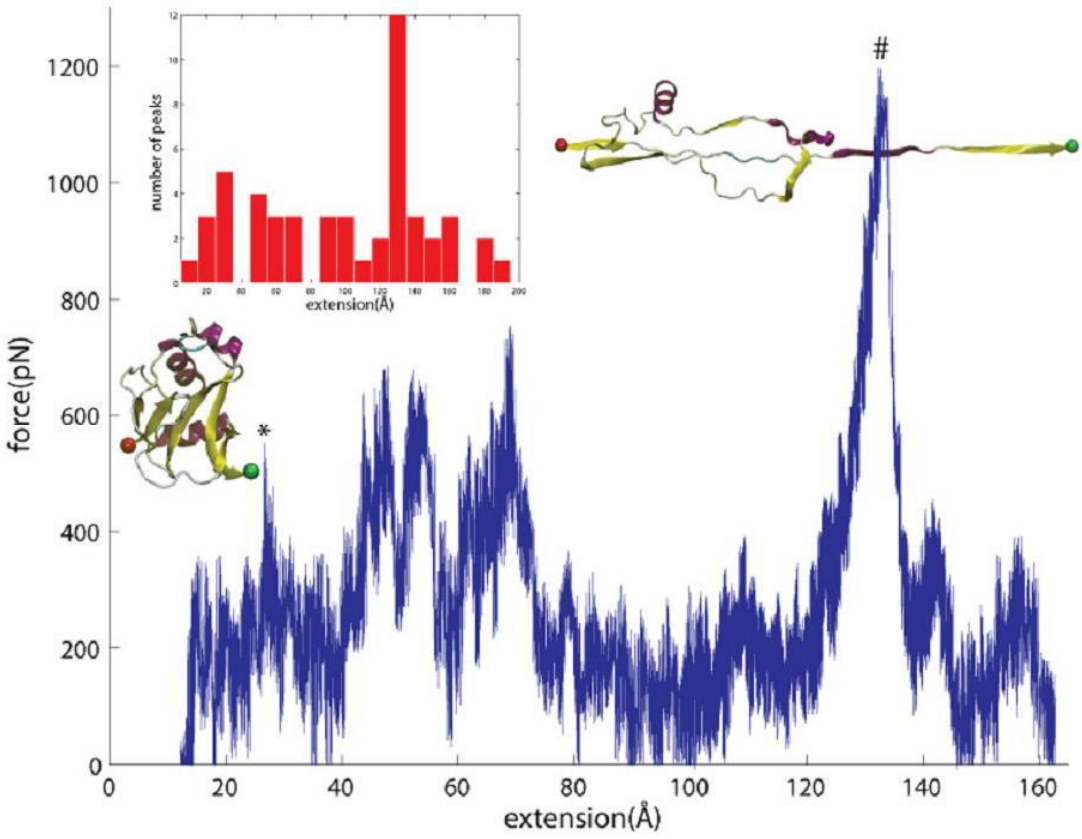

Figure 3.7 Non-specific interactions can lead to extra unfolding force peaks in SMD simulations.
Inset: histogram of the occurrences of the unfolding force peaks with different contour length increments. For unfolding trajectories following this type of pathway, the intermediate state I always occurs. However, the initial peak occurs at a much lower frequency, and some extra peaks were observed due to friction when pulling the threaded loop out of the knotting loop.

In contrast to this, such additional peaks were not observed within AFM experiments. This is likely due to the pulling speed at which SMD simulations were performed (0.25−50 nm/ns), which is at least 6 orders of magnitude faster than the pulling speed used in the AFM experiments; the decreased speed at which AFM experiments are conducted could eliminate or reduce the friction that causes additional peaks.

In the second type of unfolding pathway, unfolding is initiated from both the N- and C-termini, where it seems that there is only one major unfolding barrier to unfolding. From the native

state, the five-strand β sheet breaks into β2β5 and β3β1β4 beta sheets (Figure 3.6). This is immediately followed by breakage of parallel strands β3β1β4. Continued application of steering force leads to the unzipping of anti-parallel strands β2β5 (snapshot 3 in Figure 3.6, where β2 and β5 are shown in the process of breaking). Following the continuous application of the steering force, the protein unfolds to full extension. The second pathway corresponds to two-state unfolding behavior. Unfolding pathways predicted by SMD simulations yield contour length increments ΔL_C that are identical to those measured by single molecule AFM experiments, suggesting that the unfolding pathways observed in SMD simulations, likely correspond to the ones observed in single molecule AFM experiments. Thus, SMD simulations provide a plausible mechanistic description of the mechanical barrier crossing by AFV3-109 as it is steered along its mechanical unfolding pathways.

For the N→I transition, the major barrier corresponds to the breakage of the interface between strands β4 and β5 (snapshot 1 in Figure 3.6), allowing for an extension of 12 nm up to the β1β4 barrier. During this process, the threaded loop unravels, and the slipknot structure is effectively untied, leading to a mechanically stable intermediate state. The subsequent breakage of the β1β4β strands results in an additional 28 nm extension, corresponding to the I→U transition. For the N→U pathway, the major barrier is the unzipping of β4β5 and the concurrent breakage of β1β4.

In both pathways, it seems that the unzipping of β4β5 is a crucial event in the mechanical unfolding of AFV3-109. After this event, the unfolding pathways bifurcate into two macroscopically and microscopically distinct pathways. It is important to note that in both pathways, the threaded loop is pulled out of the knotting loop, and the tightened knot intermediate, which was predicted for a different slipknot protein thymidine kinase in coarse

grain MD simulations, [121] is not observed. This result is in good agreement with our single molecule AFM results, and thus SMD simulations provide a complementary and plausible molecular mechanism for the mechanical unfolding of AFV3-109.

3.3 Discussion

3.3.1 A Kinetic Partitioning Mechanism for the Mechanical Unfolding.

A kinetic partitioning mechanism has been predicted as a general mechanism that governs protein folding and unfolding dynamics.[6-8,162] Kinetic traps on the folding/unfolding energy landscape lead to the formation of intermediate states and the bifurcation of folding and unfolding pathways. A growing number of proteins have shown clear evidence that a kinetic partitioning mechanism defines their folding/unfolding dynamics.[36,163-165] Here, we combine single molecule AFM and SMD simulations to investigate the unfolding mechanism of AFV3-109 in detail. Our results revealed that AFV3-109 unfolds via three distinct unfolding pathways. If the unfolding is initiated at the C-terminus of AFV3-109, a stretching force pulls the threaded loop out of the knotting loop and an intermediate state forms, where β1 and β4 constitute a shear topology that provides mechanical stability; if the unfolding is initiated from both termini, the potential intermediate state is avoided, and the protein unfolds in a two-state fashion. In the latter case, the threaded loop must be pulled out of the knotting loop before the threaded loop is tightened. This requires that the unraveling of β4β5 and β2β5 to be faster than the unraveling of β1 and β4. These results demonstrate that multiple parallel pathways are available as AFV3-109 unfolds and unties the slipknot under mechanical force, providing a clear example of how the kinetic partitioning mechanism determines protein folding/unfolding mechanisms in general. Since the folding of a slipknotted protein from a completely unfolded polypeptide chain involves complex chain movement, it is likely that more kinetic traps will

form along the folding pathway. Thus, kinetic partitioning may be more pronounced in the folding pathway. SMD simulations provide a detailed description of the molecular events leading to the unfolding of AFV3-109, revealing that the interactions between β strands β4 and β5 are critical for the mechanical integrity of AFV3-109. It is of note that though the disruption of β4β5 is obligatory for AFV3-109 to unfold, the two pathways bifurcate from this point on, leading to very different means of untying the slipknot. That these two unfolding pathways are fundamentally different is shown by the large difference in mechanical unfolding distance between the two pathways (0.59 nm for N−I−U versus 0.24 nm for N−U). It seems that the relative stability of the β1β4 sheet may serve as a "gear box" to steer the mechanical unfolding into one pathway or the other; however, the detailed mechanism in which this "gear box" operates remains to be elucidated.

Although SMD simulations provide a plausible mechanism for the mechanical unfolding of AFV3-109, it is important to note that the pulling speed used in SMD simulations (∼0.25 m/s) is orders of magnitude faster than that used in single molecule AFM experiments (∼400 nm/s). Such differences could partially account for the differences between AFM experiments and SMD simulations. For example, the friction seen in SMD simulations is largely absent in AFM results. The relative population of different pathways is quite different between AFM experiments and SMD simulations: in AFM experiments, a large fraction (40%) of AFV3-109 was observed to unfold in a two-state manner, while only a small fraction (∼10%) of SMD trajectories correspond to two-state unfolding.

3.3.2 Comparing All-Atom SMD to Coarse Grain MD Simulations.

The mechanical unfolding of the slipknotted protein thymidine kinase has been previously studied using coarse grain MD simulations.[121] These simulations suggested two different and

general mechanical unfolding pathways depending on the relative shrinking rates of the two loops, which would vary depending on the stretching force. Under low stretching forces, the threaded loop shrinks faster than the knotting loop and the slipknot loosens. Under high stretching force, the knotting loop shrinks faster, and the stretching force temporarily tightens the slipknot. As a result, a tightened slipknot is formed as the intermediate state in this unfolding pathway.[121] However, in our all-atom SMD simulations on AFV3-109, we did not observe such a "tightened knot" intermediate state. Instead, we observed that the slipknot loosens, and an intermediate formed in which the slipknot has been untied and the protein is in a deformed state, where a shearing geometry provides the necessary mechanical resistance (β strands 1 and 4). Moreover, in both single molecule AFM experiments and SMD simulations, we observed that the pulling velocity did not affect unfolding pathways of AFV3-109 in two vastly different pulling speed regimes (50−4000 nm/s for AFM experiments and 0.25−50 nm/ns for SMD simulations), as the relative ratio between the two pathways remains roughly the same at different pulling velocities. However, it remains worth investigating whether pathway switching can occur in the relevant pulling speed regime predicted by coarse grained MD simulations.[166] Coarse grained MD and SMD simulations on two different model proteins revealed that different slipknotted proteins can unravel following different pathways, suggesting that detailed interactions in different slipknotted proteins play important roles in determining their specific unfolding pathways and kinetics. In principle, AFV3-109 could form a tightened slipknot under a stretching force. The absence of a tightened slipknot intermediate in our experiments suggests that forming a tightened slipknot may not be energetically favorable. The formation of a tightened knot requires the rupture of interactions between β-strands 1 and 4 and 1 and 3 as well as interactions between β-strands 2 and 5. In contrast,

unknotting would require the rupture of fewer interactions: only those between β-strands 2 and 5 as well as 4 and 5. Intuitively, unknotting should be favorable. In order to understand why the "tightened knot" did not form in AFV3-109 more thoroughly, we carried out a detailed analysis of our SMD trajectories. We found that in both pathways, interactions (specific or nonspecific) between key structural elements prevent the formation of the "tightened knot" structure. In scenario I (Figure 3.6), specific hydrogen bonds between β1 and β4 strands are at the two ends of the knotting loop and form a lock to prevent the loop from shrinking (highlighted in red in snapshot 1 in Figure 3.6). Thus, no jamming would appear in this case, and the β1− β4 ruptures after the threaded loop is pulled out. In scenario II, although the β1−β4 breaks earlier after the sliding of the two strands, a non-native (weaker) patch forms which again prevents the knotting loop from shrinking (highlighted in green in snapshot 3 in Figure 3.6). These insights reflect the importance of detailed interactions within protein structure in determining the unfolding pathways and unfolding mechanism of knotted or slipknotted proteins.

3.3.3 Implication for the Folding Mechanism of Slipknotted and Knotted Proteins.

Our results on the unfolding of AFV3-109 have interesting implications concerning the folding mechanism of slipknotted and knotted proteins. Coarse grained MD simulations on the folding of knotted proteins suggested two different mechanisms for the formation of knotted structures during the folding of knotted proteins: where the slipknot forms as an intermediate state or through plug movement of the threaded loop.[107] In the N−I−U pathway of AFV3-109, the formation of the intermediate state involves the unthreading of the threaded loop in the knotting loop. The reverse of this step would be the formation of the slipknot structure. Our preliminary refolding experiments by single molecule AFM showed that the unfolded AFV3-109 can refold

back to its native state when it is relaxed to zero force (Figure 3.8), suggesting that the folding of AFV3-109 is a spontaneous process. In addition, AFV3-109 can refold into the unfolded intermediate state prior to successfully refolding into the native state (Figure 3.8), suggesting that threading the threaded loop in the knotting loop may represent a major barrier for AFV3-109 folding. In this sense, the folding of AFV3-109 can serve as a model system to study the mechanism of slipknot formation in knotted proteins.

Furthermore, our preliminary results show that the folding process of AFV3-109 is slow and shows large variation from molecule to molecule. For some "efficient" cases, around 10% AFV3-109 could refold back to its native state in 15 s. For most cases, AFV3-109 does not fold within our instrument observation time window (~30 s). How nature solves this knotting problem to make slipknotted and knotted proteins fold efficiently remains an open question. Jackson and coworkers recently discovered that chaperones can accelerate the folding of knotted proteins,[90] raising an interesting question whether chaperones can also facilitate the folding of slipknotted proteins in a similar fashion. Future experiments will be needed to examine the molecular mechanism of slipknot formation and whether chaperones can speed up the folding of slipknotted proteins.

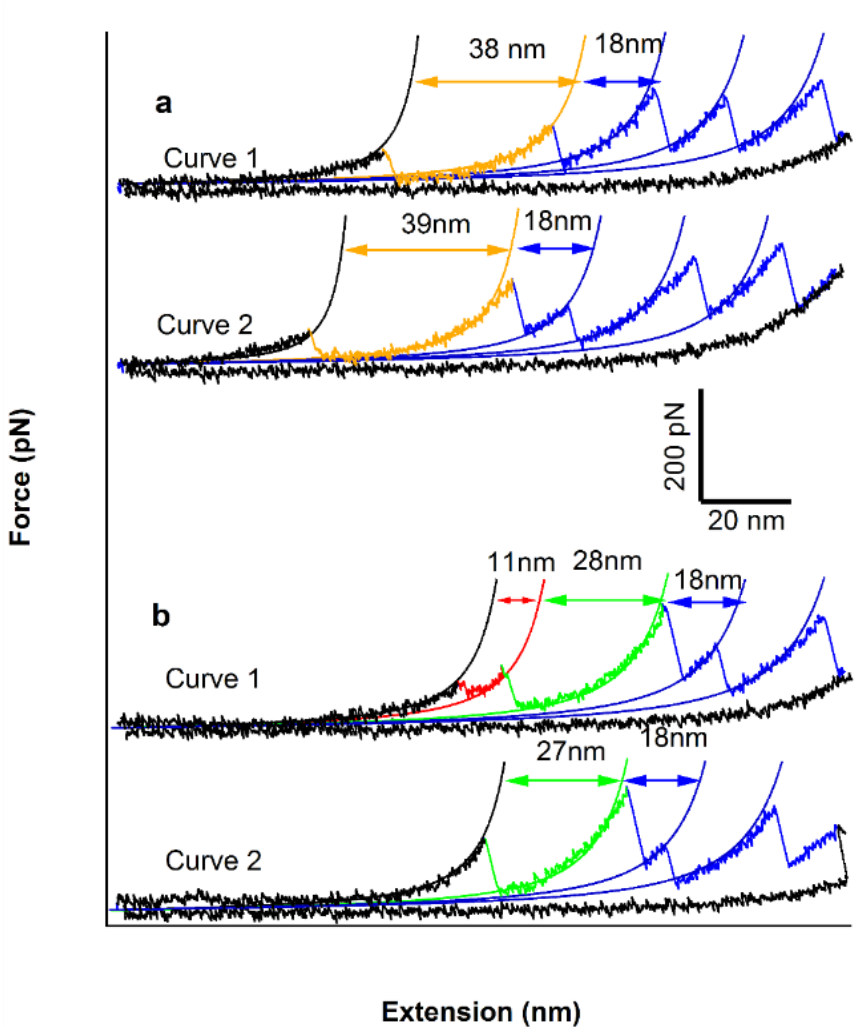

Figure 3.8 Mechanical folding of AFV3-109.

Unfolded AFV3-109 can refold back to its native state (a) as well as unfolding intermediate state prior to its complete folding (b). The mechanical unfolding of GB1 fingerprint domains is colored in blue and characterized by a ΔLC of 18 nm. The unfolding events of AFV3-109 are colored according to their unfolding pathways: two-state unfolding is colored in orange, N-I in red and I-U in green, respectively.

3.4 Experimental Section

3.4.1 Protein Engineering

The plasmid containing the genes that encode the AFV3-109 was purchased from GeneScript. The protein constructs in this thesis were built using recombinant strategies similar as described in Chapter 2. By digesting the pUC57/AFV3-109 plasmid with the restriction

enzymes BamHI and KpnI, we obtained the AFV3-109 insert containing “sticky ends”. The AFV3-109 insert was subsequently ligated by T4 DNA ligase to the corresponding “sticky ends” of the pQE80L/$(GB1)_4$ vector, which was obtained by digesting the pQE80L/$(GB1)_4$ plasmid with BglII and KpnI. From this ligation, we obtained the pQE80L/$(GB1)_4$-AFV3-109 plasmid, which was digested with BglII and KpnI to get the pQE80L/$(GB1)_4$-AFV3-109 vector. By digesting the pQE80L/$(GB1)_4$ plasmid with BamHI and KpnI, we obtained a $(GB1)_4$ insert, which was ligated to the pQE80L/$(GB1)_4$-AFV3-109 vector in order to construct the final pQE80L/$(GB1)_4$-AFV3-109-$(GB1)_4$ plasmid. To get the pQE80L/Cys-GB1-AFV3-109-Cys plasmid, we digested the pQE80L/Cys-GB1 plasmid with BglII and KpnI to obtain the pQE80L/Cys-GB1 vector, to which the AFV3-109 insert with a cysteine at the C- terminus was ligated.

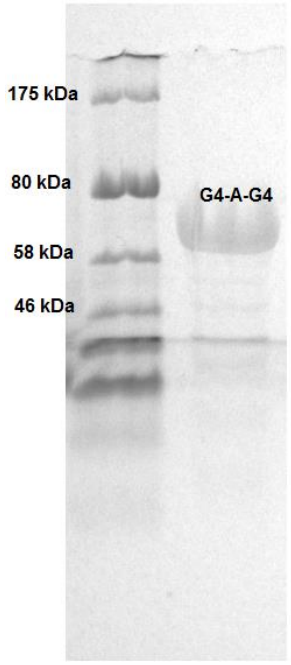

Figure 3.9 Coomassie blue stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) photograph of protein G4-A-G4

Left lane: pre-stained protein molecular weight marker. Right lane: Protein G4-A-G4

300 mL Luria Broth (LB) medium, inoculated by 3 mL of the starter culture, was incubated at 37 °C in a shaking incubator at 225 rpm. Proteins were over-expressed upon the induction by 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density (OD_{600}) of ~0.7 of the cell culture. Centrifugation was used to harvest the cells 4 hours after the induction. The proteins were extracted through the lysis of the freeze-thawed cells by lysozyme. Co^{2+} affinity chromatography (Clonetech) was used for protein purification. The purified proteins were stored at 4 °C in phosphate buffered saline (PBS) solution at a concentration of ~ 1 mg/mL.

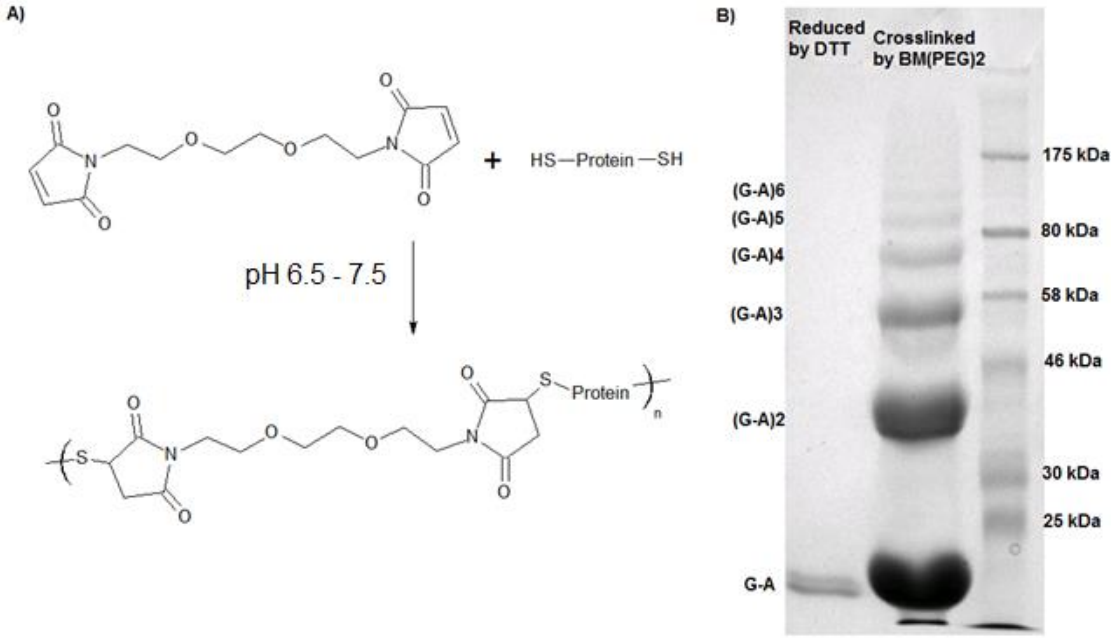

Figure 3.10 The Crosslinking of Cys-GB1-AFV3-109-Cys.

A) The schematic of the crosslinking reaction. B) SDS-PAGE of Cys-GB1-AFV3-109-Cys reduced by DTT (left lane) and crosslinked by $BM(PEG)_2$ (middle lane). The right lane is a protein molecular weight marker.

Thiol groups from the cysteines residues at both the N- and C- termini of the heterodimer protein Cys-GB1-AFV3-109-Cys (G-A) were reduced by 5 mM dithiothreitol (DTT). The Zeba desalting columns were used to remove the DTT. Then 100 μL of the solution of 50 μM Cys-GB1-AFV3-109-Cys was allowed to react with 1 μL of the aqueous solution of 5 mM of

1,8-bis-maleimidotriethyleneglycol, referred as $BM(PEG)_2$, for 3 hours at room temperature. By forming a thioether bond between thiol groups in the heterodimer proteins and maleimide groups in $BM(PEG)_2$, polyproteins $(G\text{-}A)_n$ (n ranges from 2 to 6) were formed as shown in Figure 3.10.[160]

3.4.2 SMFS

SMFS experiments were carried out on a home-made AFM[167] as well as commercially available AFMs (Cypher and MFP3D AFM from Asylum Research) the same way as described in chapter 2 (2.3.2) except using MLCT Si_3N_4 cantilevers with a spring constant of ~40 pN/nm from Bruker. In a typical AFM experiment, we deposited ~1.0 μL of protein solution (1.0 mg/mL) in PBS onto a clean glass slip covered by PBS buffer (~50 μL) and allowed the protein to adsorb onto the substrate for ~10 min before force spectroscopy experiments. Data analysis was accomplished using custom written codes in IGOR Pro 6. We used the worm-like chain (WLC) model of polymer elasticity[73] to fit consecutive unfolding force peaks to obtain the contour length increment upon domain unfolding. A persistence length of 0.4 nm, which is typical for unfolded polypeptide chains, was used in all WLC fittings.[35]

3.4.3 Monte Carlo Simulations

The mechanical unfolding of protein was described using the Zhurkov-Bell-Evans model. [50-52] [51,52,136] Force-dependent unfolding rate constants can be described as

$$\alpha(F) = \alpha_0 \exp(F\Delta x_u / k_B T)$$

where k_B is the Boltzmann constant, T is the temperature in Kelvin, $\alpha(F)$ is the unfolding rate constants at an external force of F, α_0 is the unfolding and refolding rate constants at zero force, and Δx_u is the distance from native to transition states. We carried out Monte Carlo simulations according to previously described procedures[49,140] to reproduce the force-distance curves of

stretching at various pulling speeds and measured the average unfolding force. By comparing the MC data versus experimental data, the kinetic parameters α_0, and Δx_u, which define the mechanical unfolding, are estimated.

3.4.4 SMD Simulations

Both constant velocity and constant force SMD simulations of mechanical unfolding of AFV3-109 were carried out. The coordinate file for the AVF3−109 slipknotted protein was obtained from the PDB (PDB code: 2J6B) and a protein structure file (psf) created using the modeling software Visual Molecular Dynamics (VMD)[27] and the psfgen plug-in. The protein was solvated in an explicit solvent environment. The CHARMM[168] force field topology parameters were used for the protein, and water was considered using the TIP3P model.[169] A total of three protein-water systems were created containing 45 589, 82 903, 136 947 atoms. Solvent boxes with length, width, and height dimensions of (10, 6.7, 7.2 nm), (24, 6, 6 nm), and (40.5, 6, 6 nm) were utilized. The system was energy minimized and equilibrated, and the resulting coordinates and velocities used as starting point for constant velocity and constant force SMD simulations. The simulations were performed with periodic boundary conditions in the isobaric−isothermic (NPT) ensemble; electrostatic interactions were computed by the particle mesh Ewald (PME) method,[170] non-bonded interactions were treated with a cutoff using a switching function beginning at 1.0 nm and reaching zero at 1.4 nm, and the uniform dielectric constant was set to 1. Pulling velocities used for the constant velocity simulations ranged from 0.25 to 50 nm/ns, and applied forces in the constant force simulations were 500 pN and 1 nN. The spring constant used in SMD simulations was 7 kcal·mol^{-1}·Å^{-2} (or 486 pN/Å). The simulation program was NAMD. [88] Data extraction and analysis were performed in VMD and MATLAB.

Chapter 4: Mechanically Tightening a Protein Slipknot into a Trefoil Knot

The knotted/slipknotted polypeptide chain is one of the most surprising topological features found in certain proteins. Understanding how knotted/slipknotted proteins overcome the topological difficulty during the folding process has become a challenging problem. In chapter 3, we have successfully untied a slipknotted protein into an unknotted polypeptide chain by stretching its two ends. In principle, by applying force in a proper direction, it is possible to convert a slipknot to a true knot. Here, we use SMFS as well as SMD simulations to investigate how the slipknotted protein AFV3-109 is transformed into a tightened trefoil knot by applying a pulling force. Our results show that by pulling the N-terminus and the threaded loop of AFV3-109, the protein can be unfolded via multiple pathways and the slipknot can be transformed into a tightened trefoil knot involving ~13 amino acid residues as the polypeptide chain is apparently shortened by ~4.7 nm. The SMD simulation results are largely consistent with our experimental findings, providing a plausible and detailed molecular mechanism of mechanical unfolding and knot tightening of AFV3-109. These simulations reveal that interactions between shearing β-strands on the threaded and knotting loops provide high mechanical resistance during mechanical unfolding.

Chapter 4 has been published as [He CZ], Lamour G, Xiao A, Gsponer J, & Li HB. Mechanically Tightening a Protein Slipknot into a Trefoil Knot. *J. Am. Chem. Soc.*, **2014**, 136(34), 11946-11955. This chapter is incorporated in this thesis with permission of American Chemical Society, Copyright (2014).

4.1 Changing the Pulling Direction

Proteins exhibit the remarkable ability of navigating themselves on a complex energy landscape to achieve robust and fast folding.[1,6,8,38] Knotted structures within folded proteins was previously considered impossible due to the enormous topological difficulty during the protein folding process.[9] Indeed, most proteins do not possess knotted topologies. Stretching such proteins will fully extend them to linear polypeptide chains without any knots. However, the development of bioinformatics tools has suggested that proteins do exhibit these more complex topologies, where approximately 1% of proteins in the protein data bank contain knotted or slipknotted backbones.[9,10,16,20,107] The knot types vary from the simplest trefoil knot to the complex Stevedore's knot. [171] Web servers have been established to detect knots in protein structures and we can expect that the number of knotted proteins will continue to grow.[17,18,96] Although it is still debated as to whether and how the knotted topology relates to biological functions, recent studies have shown that knotted regions are important to both ligand binding and enzyme activity.[14,97,102,107,108] Despite their complex topology, these knotted or slipknotted proteins can spontaneously fold into their native conformations in order to carry out their specific biological functions.[1,10,13,16,102,107,108,112,172,173] Understanding how proteins fold into such complex knotted or slipknotted structures has attracted considerable interests over the last few years. Experimental and simulation efforts have started to offer insights into the molecular mechanisms of these complex folding processes.[10,14,90,95,98,102-105,107,108,110,112,173-176] Experimental studies have shown that although knotted proteins can acquire their native knotted topology spontaneously, molecular chaperones inside the cell can speed up the folding process of such knotted proteins.[90] Simulation studies suggested that the formation of a slipknot structure can serve as an important intermediate step to reduce the topological

difficulty as knotted proteins fold.[107,108,173] However, it is difficult to test this prediction experimentally, because such a slipknot intermediate lacks most tertiary structure of a folded knotted protein and it is difficult to detect its formation and conversion into a true knot in experiments. In contrast to such slipknot intermediate states, slipknot proteins assume well-defined slipknotted structures with fully formed secondary and tertiary structure in their native state, which can be readily detected in experiments. Simulation studies revealed that by pulling a knotted protein from different directions, it is possible to untie a knot or tighten a knot. [117] Similarly, it should be feasible to untie a slipknot or convert a slipknot into a true knot by pulling a slipknot from the appropriate directions. Tying a slipknot into a true knot could offer insights that are relevant for understanding the mechanism and energetics for the conversion of a slipknotted intermediate to a knotted structure, despite that slipknotted proteins are not ideal model systems for testing the proposed slipknot intermediate mechanism. Moreover, tying a slipknot into a tightened knot will provide an invaluable opportunity to study the refolding process of a slipknotted protein starting from a tightened knot state, which is similar to the reverse process for the knot-forming process from a slipknot intermediate state. Here, we use a small slipknotted protein AFV3-109 as a model system[148] to study the conversion of a slipknotted protein into a trefoil knotted structure upon stretching by combining protein engineering, AFM-based SMFS, and SMD simulations techniques.

AFV3-109 (PDB code 2J6B) is a small protein with a slipknot topology. As shown in Figure 3.1 and 4.1, a knotting loop is formed near the N-terminus (colored in yellow) of AFV3-109, where a threaded loop near the C-terminus (colored in red) is inserted into the knotting loop. Just like a shoelace, the protein slipknot can be untied or converted into a tightened knot depending on where pulling force is applied. Chapter 3 confirmed that slipknotted AFV3-109

can be untied and fully extended to a linear polypeptide chain by pulling the protein from its N- and C- termini.[113] If one pulls the threaded loop of AFV3-109, it is possible to pull the threaded loop through and into the knotting loop to convert a slipknot into a true knot. Using protein engineering techniques, we engineered a cysteine variant Lys98Cys of AFV3-109 that allows us to stretch AFV3-109 from theN-terminus and residue Cys98.

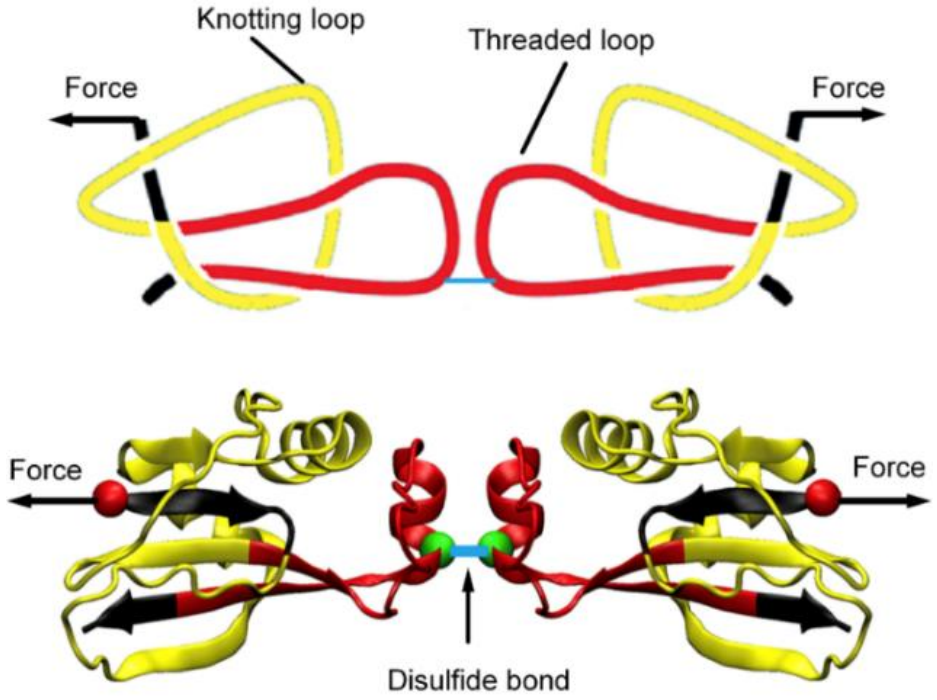

Figure 4.1 Schematic showing K98C as it is pulled from its N-terminus and residue 98.
The threaded loop is colored in red and the knotting loop in yellow. A disulfide bond can be formed by oxidizing residue Cys98, making it possible to stretch AFV3-109 from its N-terminus and residue Cys98.

To study the mechanical tightening of the protein slipknot located in AFV3-109, we mutated residue 98 located in the threaded loop (colored in red, Figure 4.1) into a cysteine residue and fused K98C to the C-terminus of the polyprotein $(GB1)_4$. We found that residue Cys98 can be readily oxidized to form a disulfide bond between two neighboring $(GB1)_4$-K98C molecules to form the dimer $(GB1)_4$-K98C-K98C-$(GB1)_4$ (Figure 4.12). The formation of $(GB1)_4$-K98C-K98C-$(GB1)_4$ made it possible to stretch the slipknotted protein AFV3-109 from its N-

terminus and residue Cys98 to convert the slipknot structure into a tightened trefoil structure. In this construct, well-characterized GB1 domains, which unfold at ~180 pN at a pulling speed of 400 nm/s with a contour length increment (ΔL_C) of 18 nm,[158,159] serve as fingerprint domains for identifying single molecule stretching events and the mechanical unfolding signature of K98C.

4.2 Results and Discussion

4.2.1 Tightening the Slipknot into a Trefoil Knot.

Stretching $(GB1)_4$-K98C-K98C-$(GB1)_4$ resulted in FECs that exhibited a characteristic saw-tooth pattern appearance (Figure 4.2A), where each saw-tooth corresponds to the mechanical unfolding of one domain in the polyprotein chain; the last peak typically corresponds to the detachment of the polypeptide chain from either the AFM tip or the substrate. Fitting these unfolding force peaks using the worm-like chain model of polymer elasticity revealed the ΔL_C upon domain unfolding. [73] Unfolding events that occur at ~180 pN with a ΔL_C of ~18 nm corresponds to the unfolding of GB1 fingerprint domains (colored in black).[158,159] As two K98C domains are flanked by $(GB1)_4$ repeats in the disulfide bonded dimer $(GB1)_4$-K98C-K98C-$(GB1)_4$, we can ensure that the two K98C domains have been stretched and extended if five or more GB1 unfolding events are observed.[113,165] As shown in Figure 4.2A, we observed two additional unfolding events with a ΔL_C of ~28 nm and an unfolding force of ~240 pN (colored in blue) in addition to the GB1 unfolding events (seven for the top trace and eight for the bottom trace). These two unfolding events can thus be readily attributed to the unfolding of the slipknotted protein K98C. Most K98C unfolding events occur after GB1 domains have been unfolded (top trace, Figure 4.2A), suggesting that K98C is mechanically more stable than GB1. Indeed, the average unfolding force of K98C is ~240 pN at a pulling speed of 400 nm/s,

higher than that of GB1 (~180 pN) (Figure 4.2B). It is of note that due to the stochastic nature of mechanical unfolding, it is possible that some K98C unfolding events occur before all GB1 domains have unfolded (bottom trace, Figure 4.2A).

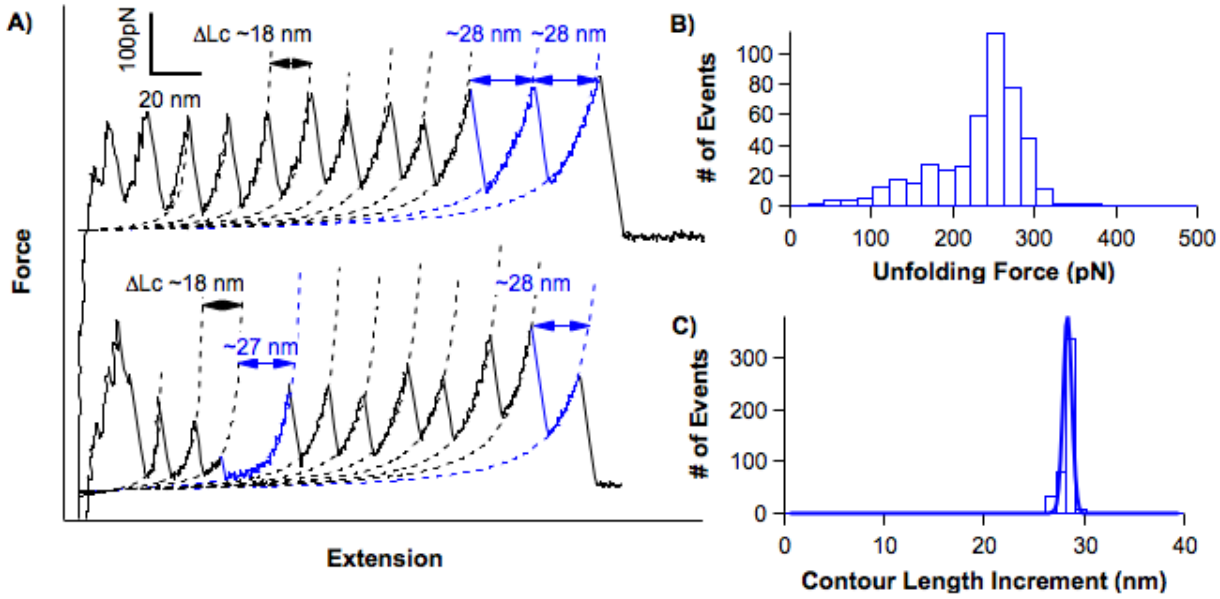

Figure 4.2 The majority of K98C unfold in a two-state fashion via pathway I.

A) Representative FECs involving two-state unfolding of K98C. Dotted lines correspond to WLC fits to the experimental data. A persistence length of 0.4 nm was used in the WLC fitting. GB1 unfolding events are colored in black. Unfolding events of K98C are colored in blue. B) Unfolding force histogram of K98C at a pulling speed of 400 nm/s in the two-state unfolding pathway. The average unfolding force of K98C at 400 nm/s is 240 ± 40 pN (average ± standard deviation, n=467). C) ΔL_C histogram of K98C in the two-state unfolding pathway. Gaussian fit to the experimental data measures an average ΔL_C of 27.8±0.5 nm (n=467).

The unfolding of K98C occurs as a single step with a ΔL_C of ~28 nm, suggesting that the unfolding of K98C occurs in an all-or-none (two-state) fashion (Figure 4.2C). We term this pathway as pathway I. In addition, the large ΔL_C observed during the mechanical unfolding of K98C suggests that most of the tertiary and secondary structure of K98C unravels during this process. ~72% of the unfolding events of K98C follow this two-state pathway.

As the force is applied on residues 1 and 98 in K98C, the fully extended length of the polypeptide chain would be 35.8 nm if there were no knotted structure (98 aa×0.365 nm/aa). The distance between residue 1 and 98 in the native state is 3.1 nm. Thus, the complete unfolding of K98C would result in a ΔL_C of ~32.7 nm if there were no knot formation. The experimentally observed ΔL_C of ~28 nm is ~4.7 nm shorter than the expected ΔL_C without a knot. This ~4.7 nm shortening, which corresponds to ~13 residues, can be attributed to the formation of a tightened knot. This result clearly indicates that upon stretching from residues 1 and 98, the slipknot structure in AFV3-109 does not get untied. Instead, the slipknot is pulled into a tightened knot, corresponding to the simplest trefoil knot.

4.2.2 Slipknot Tightening Can Be Accomplished through Multiple Pathways, While the Size of the Tightened Knot Remains the Same.

In addition to the predominant two-state unfolding of K98C, we also observed that the unfolding of K98C and the tightening of the trefoil knot can occur following multiple different pathways, involving the formation of intermediate states. As shown in Figure 4.3A, the unfolding of K98C domains can occur in two steps, giving rise to unfolding events with a ΔL_{C1} of ~21 nm for the first step, and a ΔL_{C2} of ~7 nm for the second step (Figure 4.3B, red and green histograms). This result suggests that unfolding of K98C involves the formation of an intermediate state.

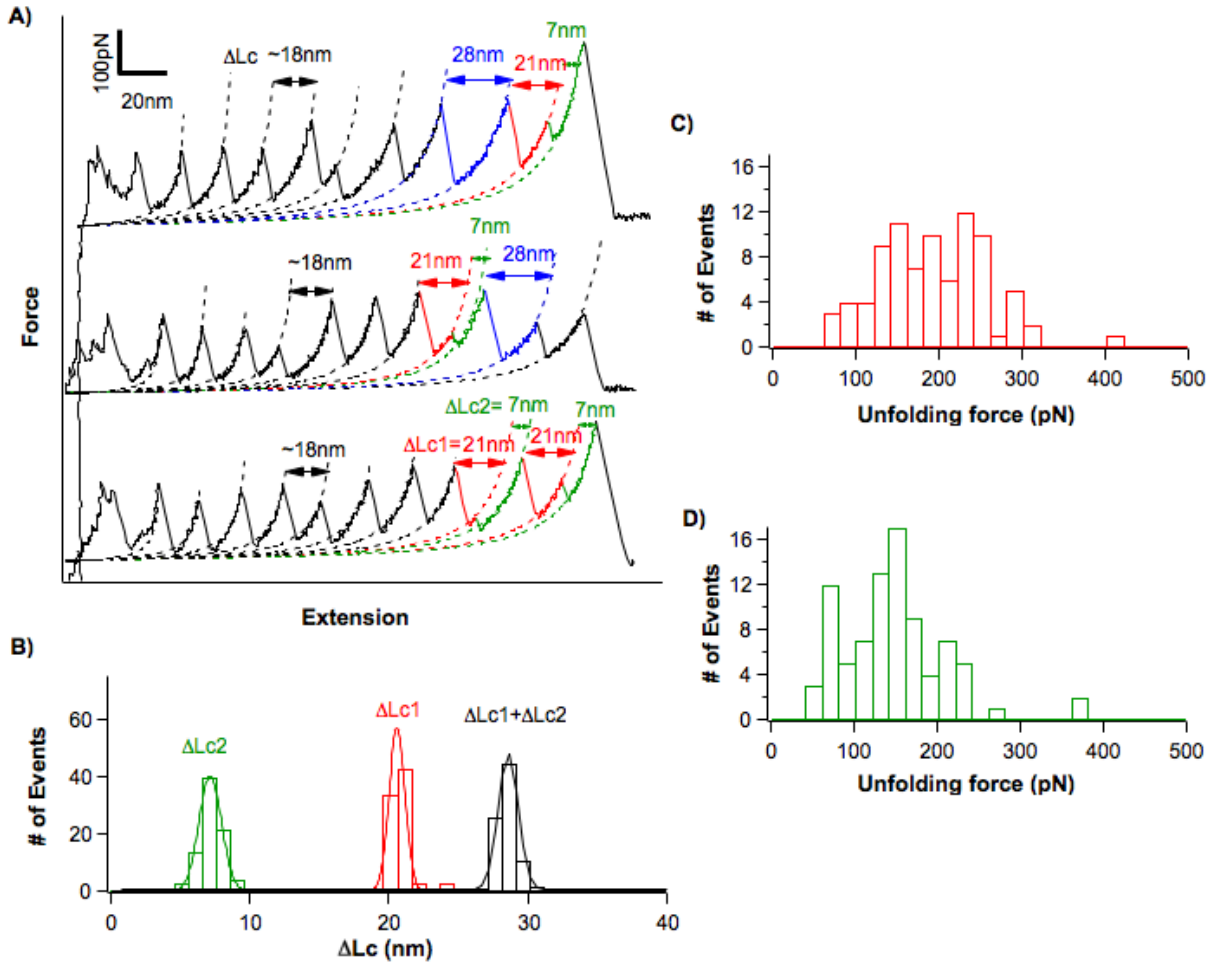

Figure 4.3 K98C can unfold via a three-state pathway (pathway II) involving an unfolding intermediate I. A) Representative FECs involving the formation of the intermediate I for K98C. GB1 unfolding events are colored in black, and two-state unfolding events are colored in blue. Three-state unfolding events of N-I and I-U in K98C are colored in red and green, respectively. Dotted lines show WLC fits to the experimental data. C) The contour length increment histogram for the three-state unfolding of K98C through intermediate I. Gaussian fits (solid lines) to the experimental data measure ΔL_{C1} of 20.8±0.7 nm (n=85), ΔL_{C2} of 7 ± 1 nm and the sum ΔL_{C1} + ΔL_{C2} of 27.8±0.7 nm, respectively. C) and D) Unfolding force histogram of N-I (red) and I-U (green) in three-state unfolding of K98C. The average unfolding force for N-I is 200±60 pN, and 160 ± 60 pN for I-U (n=85).

In this pathway (termed as pathway II), the native state is more mechanically resistant than the intermediate state, and unfolds with an average force of ~200 pN (Figure 4.3C), while the unfolding of the intermediate state, which is termed as 'intermediate I', occurs at ~160 pN (Figure 4.3D). This timing of a higher unfolding force peak followed by a lower force peak is

indicative of a reverse mechanical hierarchy, suggesting that the mechanical unfolding intermediate state is protected by the native structure and is subject to the stretching force only after the native state has been partially unraveled. In addition, it is noteworthy that the sum of ΔL_{C1} and ΔL_{C2} (~28 nm, Figure 4.3B) is the same as that for the two-state unfolding pathway, suggesting that this three-state unfolding pathway also leads to the formation of a similar tightened trefoil knot structure when K98C is unfolded and extended to higher extensions. This three-state unfolding pathway occurs at a frequency of ~13%.

Interestingly, a small percentage of K98C domains (~2%) unfold through a different three-state unfolding pathway (termed as pathway III, Figure 4.4A). Initial unfolding of K98C results in unfolding events with a ΔL_C of ~7 nm, followed by a second unfolding step of ΔL_C of ~21 nm, which corresponds to the unfolding of a different intermediate state. We termed this intermediate as 'intermediate II'. The sum of the ΔL_{C1} and ΔL_{C2} is 28 nm (Figure 4.4B), suggesting that this unfolding pathway also leads to the formation of a tightened trefoil knot structure that is similar in size to that encountered in the first two pathways. Unfolding forces for the two steps in this pathway are shown in Figure 4.4C and D.

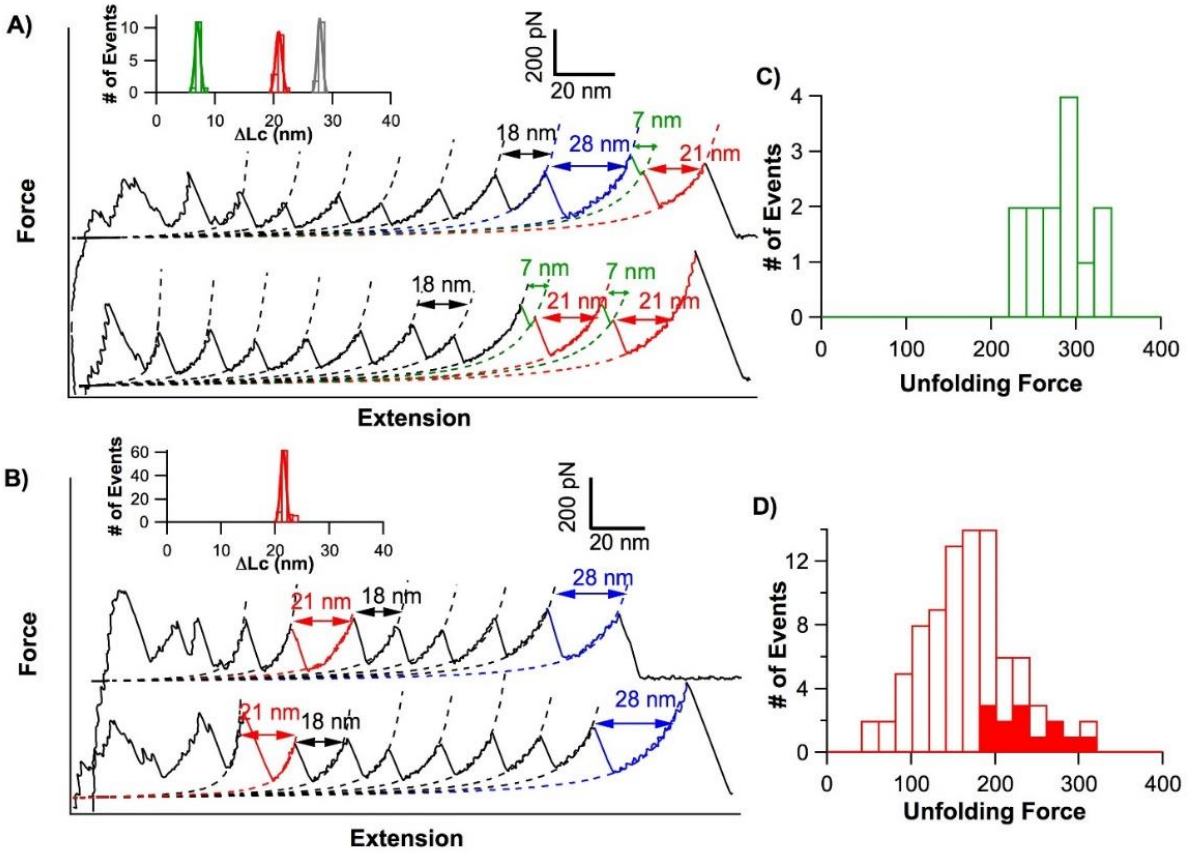

Figure 4.4 A small percentage of K98C unfolds in three-state fashion involving the formation of an unfolding intermediate II.

A) Representative FECs of $(GB1)_4$-K98C-K98C-$(GB1)_4$ with N-II-U three state unfolding of K98C. GB1 unfolding events are colored in black, and two-state K98C unfolding events are colored in blue. Unfolding events of N-II and II-U during the three-state unfolding of K98C are colored in green and red, respectively. The inset shows ΔL_C histograms for the three-state unfolding events. Gaussian fits (solid lines) to the experimental data measure ΔL_{C1} of 7.0±0.5 nm (n=13), ΔL_{C2} of 20.8±0.6 nm and the sum $\Delta L_{C1} + \Delta L_{C2}$ of 27.8±0.5 nm, respectively. B) Representative FECs showing unfolding events of II-U only. The inset shows the ΔL_C histogram. Gaussian fit measures an average ΔL_{C2} of 21.0±0.7 nm (n=87). C) and D) Unfolding force distribution histograms of N-II (green) and II-U (red) in three-state unfolding of K98C. The Red solid bars are from II-U events in complete N-II-U pathway and red open bars are from II-U only events. The average unfolding force for N-II is 280±40 pN (n=13), 180±60 pN for all II-U events (n=100).

In addition to the observed pathway I-III, we also observed a significant fraction of K98C unfolding events (~13%) that show only ΔL_C of ~21 nm (Figure 4.4B), suggesting that the

unfolding of K98C began from the unfolding intermediate state II as in the pathway III. It is likely that part of K98C in the native state is unfolded at low forces, making the initial unfolding undetectable because the force is below the force detection limit of our AFM or such low force events are buried in the region of FECs dominated by non-specific interactions (Figure 4.4B). Thus, the unfolding events displaying only ΔL_C of ~21 nm likely follow a pathway similar to unfolding pathway III, except that the initial unfolding step of ΔL_C of ~7 nm occurs at much lower forces. It is of note that the unfolding force histogram of II-U in the complete N-II-U pathway matches the high force part of the distribution from the II-U only events (Figure 4.4D). For simplicity, we consider both types of unfolding events (7 nm event followed by 21 nm, and 21 nm only events) as events following pathway III, which amount to 15% of all unfolding events of K98C. The ΔL_C of pathway III is indistinguishable from that ascribed to pathway II; however, it is unknown whether similar structural elements are responsible for the similar ΔL_C exhibited by the two unique three-state unfolding pathways.

4.2.3 Multiple Pathways Reveal a Kinetic Partitioning Mechanism for Mechanical Unfolding.

Our results demonstrate that the mechanical unfolding of the slipknotted protein variant K98C proceeds via multiple parallel pathways, which all result in the tightening of the slipknot into a trefoil structure. Approximately 72% of K98C unfolded through a two-state unfolding pathway (pathway I), with ~13% occurring through intermediate I (pathway II) and ~15% through intermediate II (pathway III). These results are in agreement with the kinetic partitioning mechanism [8,113,165] for protein folding and unfolding, which suggests that kinetic traps on the energy landscape can lead to the bifurcation of folding/unfolding pathways and the formation of intermediate states.

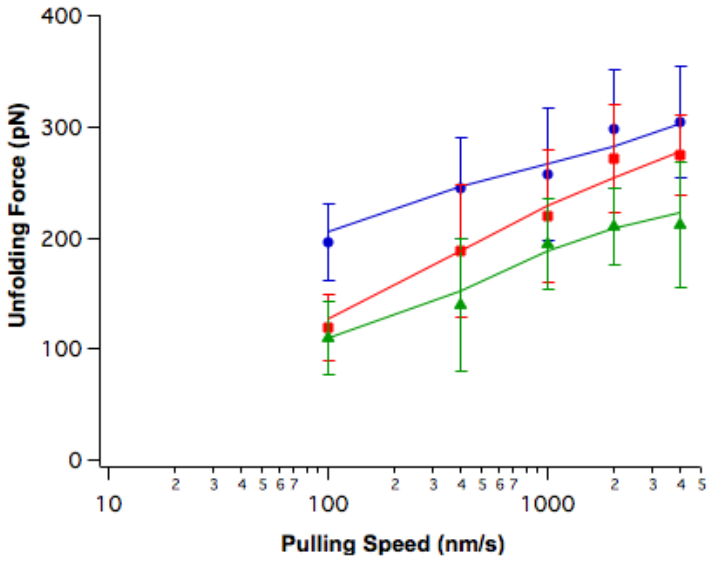

Figure 4.5 Pulling speed dependence of K98C unfolding.

Two-state unfolding events are colored in blue, and three-state unfolding events (with intermediate I) are colored in green and red. Solid lines are Monte Carlo simulation results using kinetic parameters tabulated in Table 4.1.

Table 4.1 Kinetic parameters of unfolding events from different pathways[a]

unfolding event		unfolding force (pN)	ΔL_C(nm)	Δx_u (nm)	α_0 (s^{-1})	energy barrier (k_BT)[b]
Pathway I	N-U	240±40	27.7±0.5	0.19	0.0012	20.5
Pathway II	N-I	200±60	20.8±0.7	0.14	0.05	16.8
	I-U	160±60	7 ± 1	0.16	0.5	14.5
Pathway III	N-II[c]	280±40[d]	7	N/A	N/A	N/A
	II-U	180±60[e]	21.0±0.7	0.14	0.05	16.8

a: ΔL_C, contour length increment upon unfolding; Δx_u, distance between the native state and mechanical unfolding transition state; α0, unfolding rate constant at zero force. Data is represented as average ± standard deviation.Data is represented as average ± standard deviation. **b**: Unfolding energy barrier at zero force was estimated based on α_0 assuming a prefactor of 10^6 s^{-1}. [177] **c**: Kinetic parameters were not estimated for N−II step in Pathway III due to the small number of events. **d**: The average unfolding force was calculated based on unfolding events with clearly identifiable ΔL_{C1} of 7 nm. The number of events is 13. **e**: The average unfolding force was calculated based on all unfolding events with ΔL_{C2} of 21 nm following Pathway III. The number of events is 100.

To further investigate the unfolding kinetics of K98C, we performed force-extension experiments at different pulling speeds. In our experiments, we did not see any obvious effects of pulling speed on the partitioning of different unfolding pathways. Figure 4.5 shows the effect that pulling speed has on unfolding forces exhibited by pathways I and II. To estimate the kinetic parameters for different unfolding pathways, we carried out Monte Carlo simulations to reproduce these AFM experiments.[49,159] We found that the kinetic parameters shown in Table 1 can accurately describe our experimental data. Specifically, simulation results show that the unfolding distance to the transition state Δx_u for pathways I and II is small, indicative of how brittle the unfolding transition is.[49] Due to the low frequency of the unfolding step N-II (native to intermediate II) in pathway III, we were not able to obtain a pulling speed dependency or the kinetic parameters for this particular pathway.

4.2.4 K98C can Refold From a Tightened Trefoil Knot Conformation to Its Native Slipknot Conformation.

By stretching AFV3-109 from its N-terminus and residue 98, we stretched a slipknot protein into a tightened trefoil knot. Such a tightened trefoil knot provides an invaluable opportunity to investigate whether a tightened trefoil knot can loosen up to allow the polypeptide to refold into its native slipknot conformation, which involves the conversion of a trefoil knot into a slipknot. To investigate this possibility, we carried out refolding experiments on K98C. First, we stretched K98C to unfold K98C and convert it into a tightened trefoil knot; then we relaxed the unfolded polypeptide chain quickly to zero force and waited for 10 seconds to allow the unfolded protein to refold. We then stretched the protein again to examine whether the K98C has refolded to obtain its native slipknot conformation. Figure 4.6 shows some FECs from the same molecule during such an experiment. We found that the tightened trefoil knot can refold

to the native slipknot conformation of K98C upon the relaxation of the pulling force as judged by ΔL_C (indicated by blue stars). However, the probability of successful refolding of K98C to its slipknot native state from its trefoil knot conformation is low, as most non-GB1 unfolding force peaks do not show ΔL_C of ~28 nm (indicated by red stars), suggesting that K98C misfolds into conformations other than the native slipknot conformation. These results suggest that the tightened knot does not prevent refolding, but may make the energy landscape more rugged and complex and the protein more prone to misfolding.

However, a complete characterization of the refolding behaviors of a tightened knot has not been possible, as such experiments are challenging and require one to hold onto a single molecule for an extended period of time to carry out unfolding and refolding experiments, which is currently beyond the capability of our AFM instruments. More robust and specific attachment chemistry will be required for such refolding experiments and the results will be reported in the future.

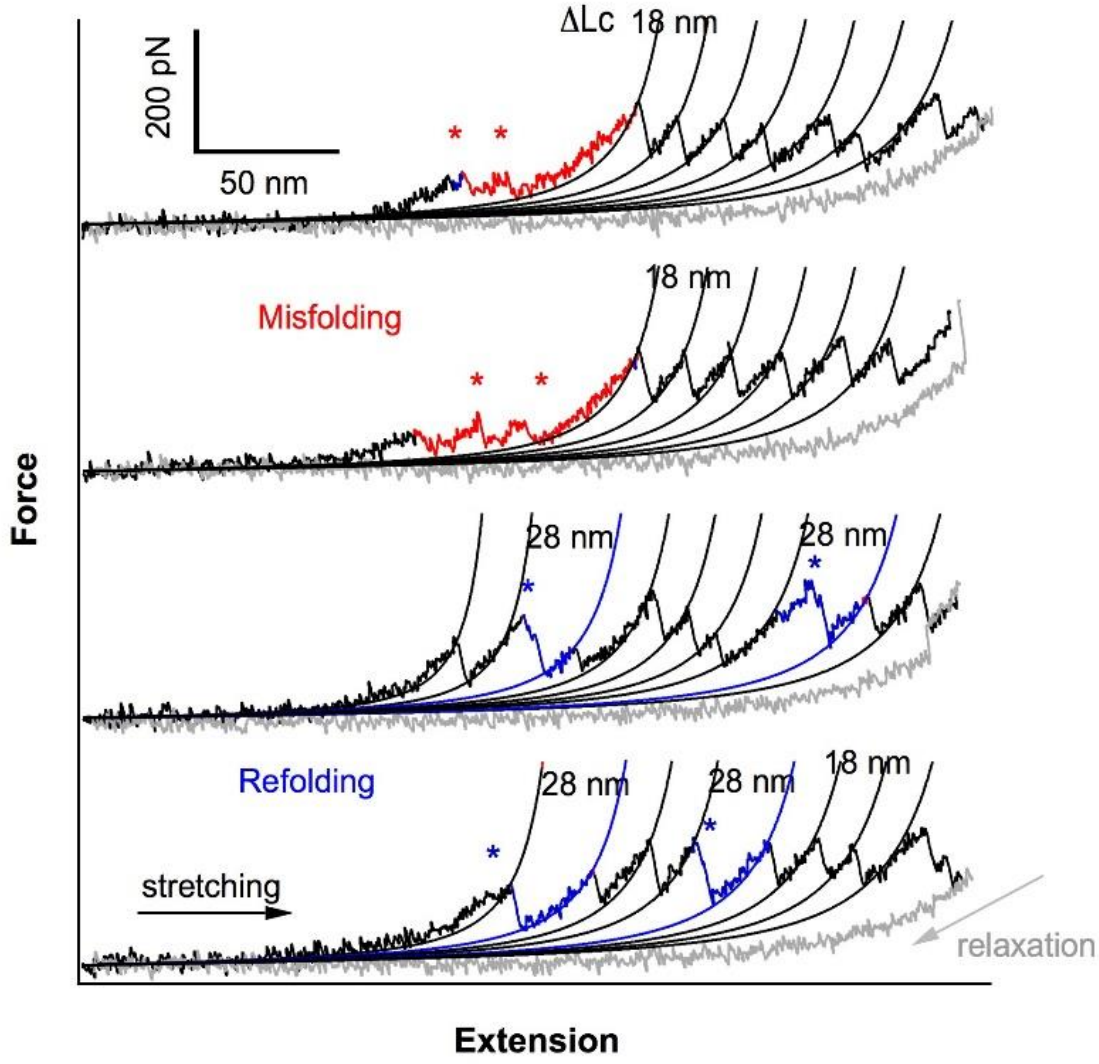

Figure 4.6 Representative refolding traces of $(GB1)_4$-K98C-K98C-$(GB1)_4$.

During the refolding experiment, the same polyprotein molecule was stretched to unfold all the domains in the polyprotein chain, and then quickly relaxed to zero force. After waiting for 10 seconds at zero force to allow the unfolded protein to refold, the protein was stretched again. Unfolding events with ΔL_C of 28 nm (colored in blue and indicated by blue stars) correspond to the unfolding of refolded slipknot protein K98C, unfolding events with ΔL_C of 18 nm correspond to the GB1 unfolding events. The unfolding events that give rise to irregular ΔL_C (indicated by red stars) are assigned to the unfolding of misfolded K98C domains.

4.2.5 SMD Simulations Reveal Molecular Mechanisms for Slipknot Tightening.

SMFS experimental results reveal complex unfolding behaviors of K98C when its slipknot is tightened into a trefoil knot. To extend K98C, it is necessary to disrupt interactions between

different secondary/tertiary structural elements as well as the hydrophobic core; the ΔL_C measured during mechanical unfolding of K98C could provide a glimpse of structural elements ruptured during unfolding. To further understand the molecular mechanism underlying multiple unfolding pathways and identify structural elements/interactions critical to the mechanical unfolding of K98C, we carried out SMD simulations under constant pulling velocities as well as constant pulling forces.[88,152,153] Recognizing that the AFM experiments and SMD simulations are carried out at time scales that differ by more than six orders of magnitude, we intend to use SMD simulations to obtain a plausible molecular level explanation of our experimental observations and help design new AFM experiments.

When the force is applied between the N-terminus and residue 98 in SMD simulations, β strands β1, β4 and β5 are subject to the applied stretching force, resulting in a bifurcation of unfolding pathways. In the pathway colored in blue (Figure 4.7), the applied steering force caused simultaneous unravelling of parallel β strands β1β3β4 and anti-parallel strands β4β5 (as evidenced by the disruption of the backbone hydrogen bonds between these β-strands), the only major energy barrier between the native and extended tightened knotted state (Figure 4.8A). During this process, the threaded loop (β4β5) is pulled through the knotting loop, converting the slipknot topology to a trefoil knotted structure. Further stretching causes the loss of remaining secondary structures and tightening of the trefoil knot.

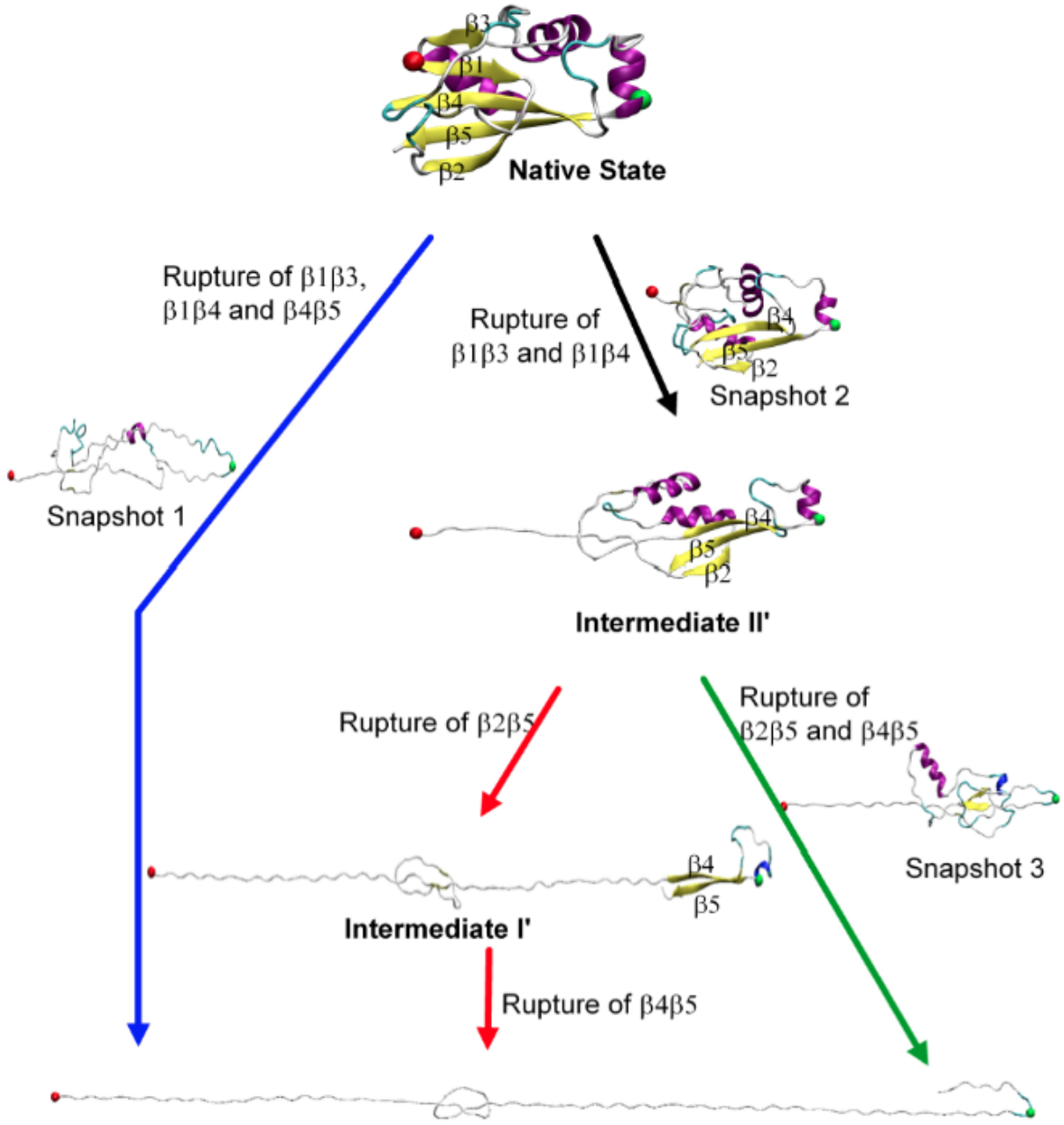

Figure 4.7 Distinct mechanical unfolding pathways of K98C as observed through SMD simulations.

Upon stretching from its N-terminus and residue 98, K98C variant of AFV3-109 unfolds via three distinct unfolding pathways, all of which lead to the conversion of the slipknot into a tightened trefoil knot. Pathway I (colored in blue) corresponds to the two-state unfolding pathway, and Pathways II and III (colored in red and green) correspond to unfolding pathways involving distinct unfolding intermediate states. Along each individual unfolding pathway, snapshots of K98C are shown to indicate structural changes occurring for K98C during the unfolding process. Structural elements/interactions that are ruptured during each step of the unfolding process are also indicated.

It is likely that this trajectory corresponds to the two-state unfolding pathway (pathway I) observed experimentally. In other trajectories (Figure 4.7, colored in green and red), the first step corresponds to the unraveling of β1β3β4, while interactions between the β4β5 strands remain intact, leading to an extension of ~7 nm and the formation of an intermediate state (termed as intermediate II'). After this initial energy barrier is surmounted, β2β5 in intermediate II' is exposed to the steering force directly, and unfolding proceeds via two pathways. In the pathway illustrated in red (Figure 4.8B), intermediate state II' is short-lived. The β2β5 is ruptured and the threaded loop containing the intact β4β5 is pulled through the knotting loop. Further stretching leads to straightening of the helices and unstructured coils, as well as the formation of a tightened trefoil knot while β4β5 remains intact. This intact β4β5 is mechanically resistant, and serves as another intermediate state (we term it intermediate I'). Further stretching resulted in the unraveling of β4β5 and a further extension of the protein by ~6 nm.

Figure 4.8B indicates that energy barriers for this pathway are located at extensions of 1 nm, 7 nm and 20 nm, corresponding to the unravelling of β3β1β4, β2β5 and β4β5 respectively. The intermediate I' formed at 20 nm extension likely corresponds to the intermediate I observed at ~21 nm in pathway II in single molecule AFM experiments (Figure 4.2, colored in red), while the short-lived intermediate II' is not found in pathway II from experiments. To confirm the link between AFM experiments and SMD simulations, it will be necessary to use loop elongation variants to change the ΔL_C of the intermediate state. Such experiments are currently underway and will be reported in the future.

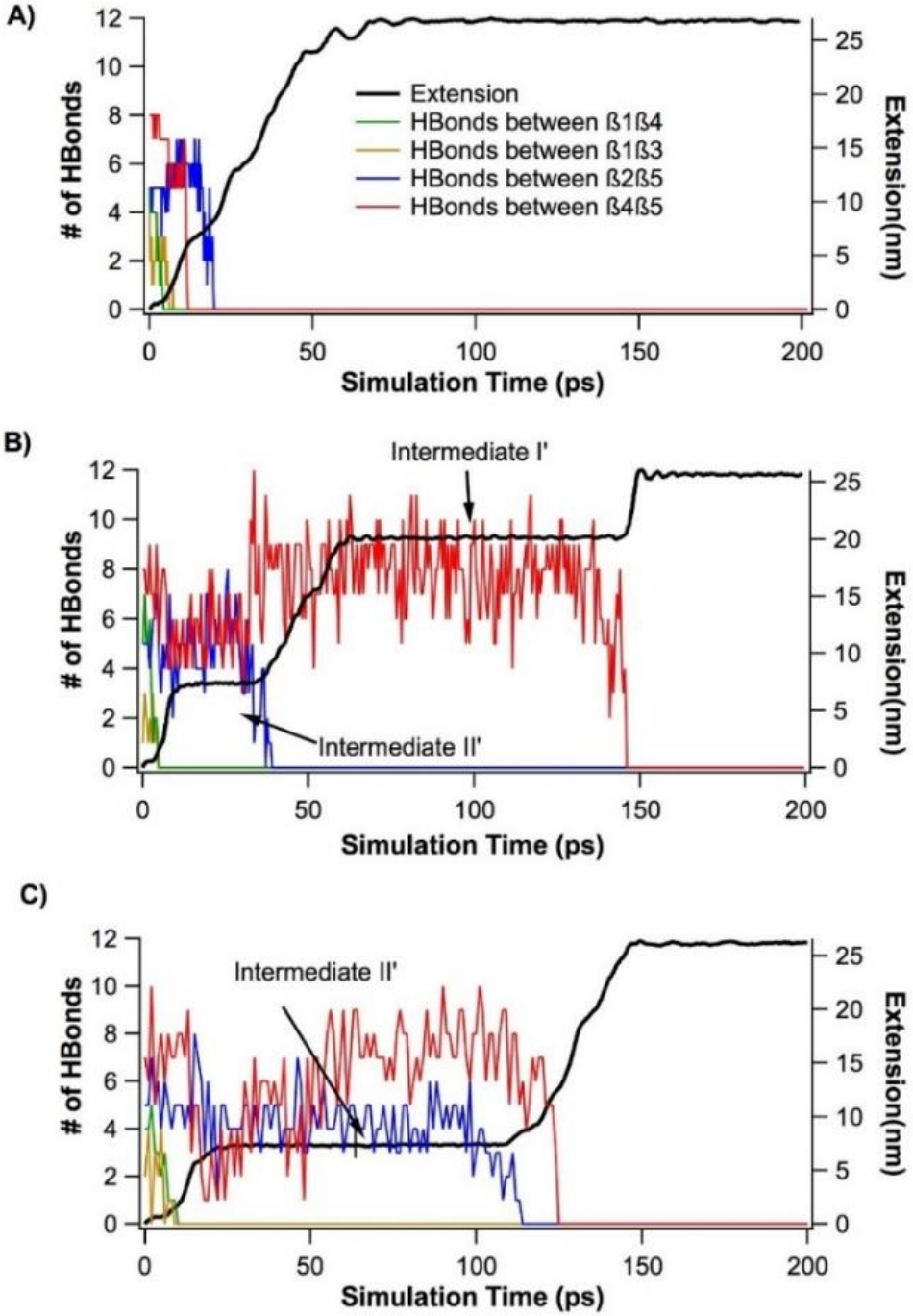

Figure 4.8 Results of constant force SMD simulations, showing extension and number of hydrogen bonds versus simulation time.

A) Two-state unfolding without intermediate state. B) Unfolding with two intermediate states. C) Unfolding with one intermediate state. Unfolding intermediate states manifest themselves as long plateaus in the Extension-time curves. The plateau at the extension of ~26 nm corresponds to the yet to be fully tightened trefoil knot. The three trajectories shown in this figure were run at a constant force of 1200 pN for 200 ps.

In the pathway colored in green, intermediate state II' is long-lived (Figure 4.8C). β4β5 and β2β5 provide the major resistance for intermediate II' to stretching and are broken almost simultaneously, where further stretching straightens helices and unstructured coils, and tightens the trefoil knot formed by the polypeptide chain. Figure 4.8C indicates that the energy barriers, located at extension of 1 nm and 7 nm, correspond to the unravelling of β3β1β4 and β4β5β2. Intermediate II' at 7 nm likely corresponds to the intermediate state II observed experimentally in single molecule AFM experiments (Figure 4.3).

SMD simulation results revealed the number of unfolding pathways and the location of intermediate states, which are largely in agreement with our AFM results, thus providing a plausible molecular mechanism for mechanical unfolding pathways observed experimentally. However, it is important to note that despite the similarity between pathways found in SMD simulations and AFM experiments, the frequency of trajectories in each pathway found in SMD simulations is different from those observed experimentally. Among constant force SMD simulations at 1200 pN, the pathway with intermediate II was encountered ~85% of the time, while two-state unfolding pathway and the pathway with intermediate I were observed ~10% and ~5% of the time, respectively. In contrast, in AFM experiments ~72% of K98C unfolded through a two-state unfolding pathway, with ~13% occurring through intermediate I and ~15% through intermediate II. This difference is likely because SMD simulations are carried out at a time scale that is ~10^6 to 10^7 times shorter than that of AFM experiments. At the SMD simulation time scale, friction within and between polypeptide chains in the protein structure become significant while it is less important for the pulling velocity at which single molecule AFM experiments are performed. It has been suggested that protein unfolding may occur via different mechanisms depending on the pulling velocity.[138] It is uncertain whether the time

scale difference between the two techniques is sufficient to explain the difference in trajectory appearance frequency. Ultrafast AFM pulling experiments[138] could possibly bridge the divide between AFM experiments and SMD simulations.

4.2.6 The Size of Tightened Trefoil Knot is Independent of Unfolding Pathways.

Our results from SMFS experiments showed that the size of the tightened trefoil knot is the same between all unfolding pathways (~13 aa residues in SMFS experiments). The mechanical tightening of knotted proteins has been studied previously both experimentally and computationally.[115,116,119,120] In the study of the mechanical tightening of phytochrome, a protein with a figure-eight knot topology, the size of the tightened figure-eight knot was found to involve ~17 aa residues. [116] It is reasonable to assume that a tightened trefoil knot is smaller, considering the difference in knot complexity between the figure-eight knot and the simplest trefoil knot.

SMD simulations also provide key information about the size and location of the tightened trefoil knot within the extended polypeptide chain. In constant velocity SMD simulations (Figure 4.9), we found that after overcoming frictions along the pulling process, the trefoil knot became fully tightened at high force (~3000 pN), and the size of the tightened trefoil was ~14 residues, which is in close agreement with that found in the AFM experiments (~13 residues). The location of the tightened trefoil knot varies in different trajectories. In addition, constant velocity SMD simulations carried out at different pulling velocities do not show a noticeable difference in the molecular events along the unfolding trajectories, but show increased friction experienced by the polypeptide chain at higher pulling velocity.

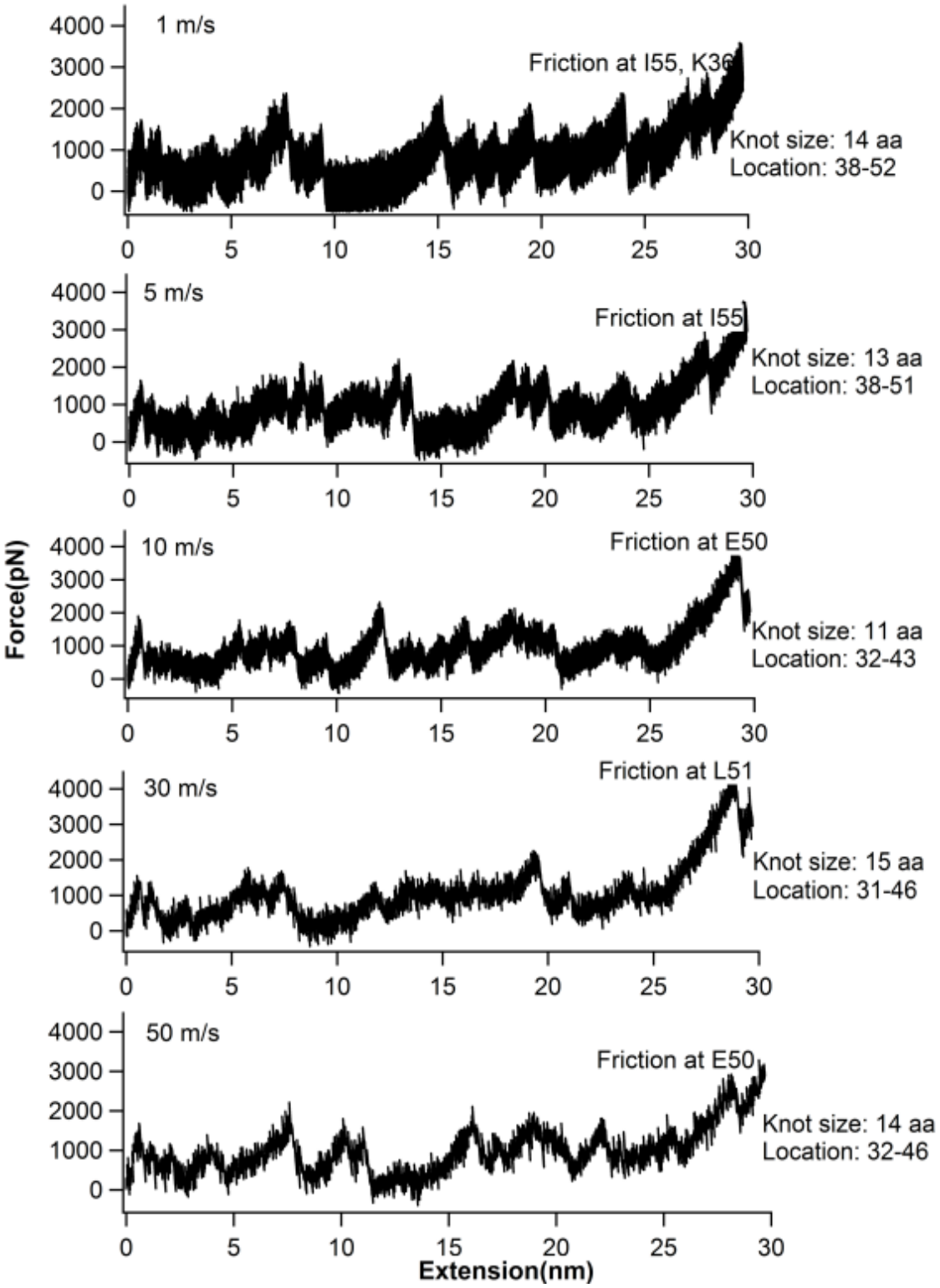

Figure 4.9 FECs from SMD simulations in constant velocity mode.

The knot sizes and locations are labeled at the right of each curve. The locations of significant frictions that resist the tightening of knot at high forces are labeled above the force peaks.

In contrast to constant velocity SMD simulations, we found that most of the trefoil knots are not fully tightened due to steric hindrance/friction as shown in Figure 4.10A in constant force SMD simulations at 1200 pN. The size of these knots ranges from 14 to 24 amino acid residues (Figure 4.10B). It is of note that almost all of these knots start and end at residues that have bulky side-chains, such as lysine (K), isoleucine (I), glutamic acid (E), and leucine (L) (Figure 4.10C). Evidently, frictions between side-chain of these residues and the knotting loop prevent further tightening of the knot. In addition, half of our simulations show that, on the knotting loop and threaded loop, the tightened knot forms non-native β-strands, between which the interactions provide further resistance to stretching and prevent further shrinking of the trefoil knot (Figure 4.10A, right). Previous simulation studies on tightening protein knots showed similar frictions.[115,116,119]

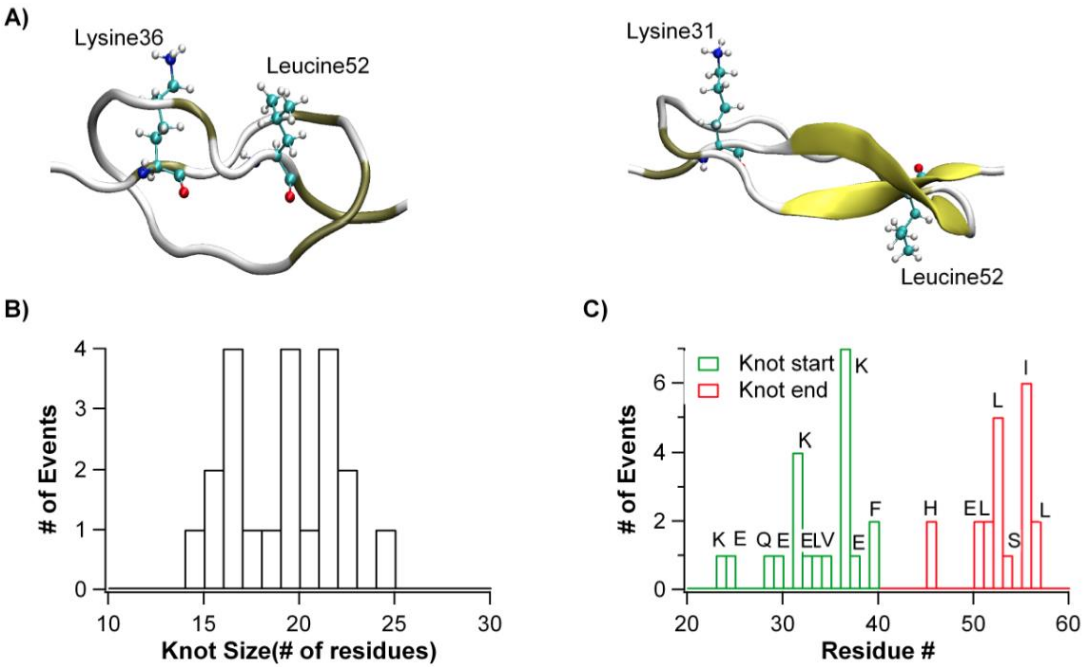

Figure 4.10 Location and size of tightened knots elucidated through SMD simulations.

A) Snapshots of two tightened knots. We found that the knot described in SMD simulations is not completely tightened. The start and end points of the tightened knot are often residues with bulky side chains. B) The tightened trefoil knot size histogram. The average size of the tightened trefoil knot in SMD simulations is 20 residues. C) A histogram of the start and end residues of the tightened trefoil knot.

To further examine the effect of bulky side chain on knot tightening, we mutated K31, K36 and I55, three bulky residues that prevent knot shrinking/sliding in the simulations, with a smaller residue, alanine. The distribution of knot locations becomes broader (Figure 4.11), which indicates that losing the bulky side chain facilitates the sliding of the knot. Thus, it is reasonable to conclude that the bulkier side chains are responsible for preventing the sliding of the knot in the simulation. It is of note that such effect does not exist ont the time scale at which the AFM experiments are performed, because K98C is stretched to a fully tightened trefoil knot (~13 residues) in the AFM experiments.

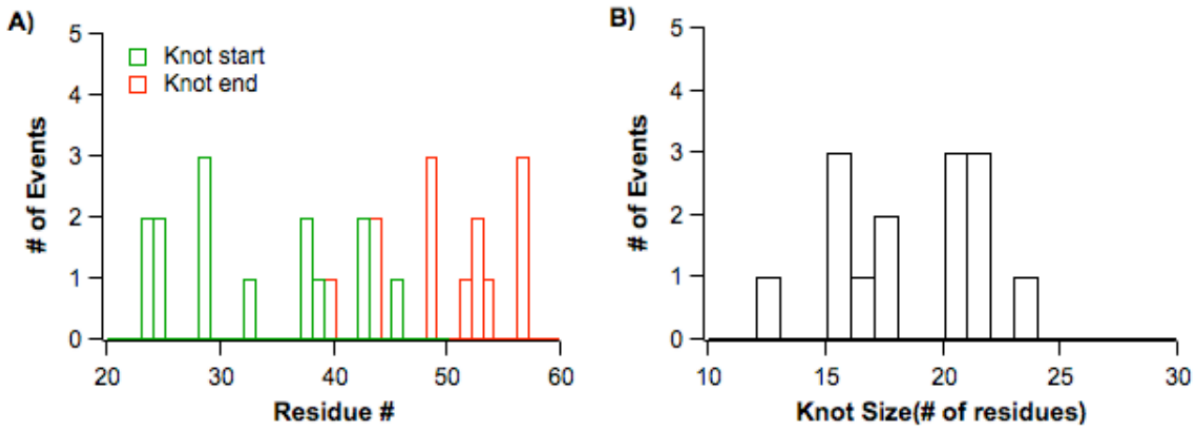

Figure 4.11 Residues with less bulky side chains facilitate the sliding of the knot.
A) Histogram of the start and end residues of the tightened trefoil knot, where residues K31, K36 and I55 were replaced with smaller residue alanine. B) The tightened trefoil knot size histogram. The average size of the tightened trefoil knot in SMD simulations is 20 residues.

4.2.7 Major Energy Barriers Do Not Necessarily Arise as a Trefoil Knot is Formed from a Slipknot.

By correlating our SMFS experimental results and SMD simulations, we identified three different pathways (N-U, N-I-U, N-II-U), and the likely location of energy barriers/intermediate states along each pathway. As shown in Figure 4.8, unraveling of β1β3β4

is involved in the unfolding of the native state of K98C within all pathways, demonstrating their importance in maintaining the mechanical stability of the native state of K98C.

In contrast, the key necessary step towards transforming the slipknot to the trefoil knot is to pull the threaded loop and the C-terminus into the knotting loop, which requires the rupture of β2β5. In the two-state pathway (Pathway I) and the three-state pathway (Pathway II with ΔL_{C1} of 21 nm and ΔL_{C2} of 7 nm), the conversion of the slipknot to a trefoil knot occurs after the major barrier for the unfolding of the native state of K98C (that is the unraveling of β1β3β4) has been circumvented, and does not result in any experimentally observable event. Therefore, converting the slipknot to a trefoil knot in these two unfolding pathways (Pathways I and II) in the AFM experiments most likely does not involve a significant energy barrier. In comparison, in most trajectories following Pathway III, the energy barrier for the unfolding of intermediate state II' corresponds to the concurrent rupture of β4β5 and β2β5. Thus, the conversion of the slipknot to a trefoil knot likely contributes to the experimentally observed unfolding energy barrier. Since it is the concurrent rupture of β4β5 and β2β5 that gives rise to the observed energy barrier, which is estimated to ~16.8 k_BT assuming a prefactor of 10^6 s^{-1},[177] the barrier caused by rupturing β2β5 alone should be significantly smaller, although we are not able to separate the contributions of rupturing β4β5 from those required for rupturing β2β5. These results suggest that it is possible that topologically converting the slipknot to a trefoil knot does not involve a significant energy barrier.

Simulations reported by Sułkowska et al proposed that forming a slipknot conformation and then threading the end of loop can reduce the topological difficulty of folding into a knotted structure, thus the slipknot can serve as an important intermediate to facilitate the folding of knotted proteins.[107,108,173] It has been challenging to experimentally test this prediction, because

such a slipknot intermediate lacks most of the tertiary structure of a folded knotted protein and is difficult to detect its formation and conversion into a true knot in experiments. Although AFV3-109 is not an ideal model system for testing this prediction, our results on K98C may nonetheless provide some relevant insights. Compared with the fully structured slipknot of K98C, the proposed slipknot intermediate lacks most tertiary and secondary structure. It can be anticipated that the energy required to convert the "unstructured" slipknot intermediate state to a true knot should be lower than that required for converting the fully structured slipknot in K98C to a trefoil knot. Thus, our results suggest that converting the slipknot in K98C to a trefoil knot does not involve significant energy barrier, making the slipknot intermediate mechanism plausible for the folding of knotted proteins. However, it is important to note that if significant contacts were to form in the slipknot intermediate state between the threading and knotting loops (as in the case of intermediate state II of K98C) during the folding of a knotted protein, a significant energy barrier might arise for converting the slipknot intermediate state to the knotted conformation, leading to a deep kinetic trap that may significantly slow down the folding of the knotted protein.

Furthermore, our AFM experiments also suggest that knot tightening does not increase the mechanical resistance of proteins as compared to unknotted proteins at the time scale the AFM experiments are performed; this may be an important distinguishing feature between AFM experiments and SMD simulations, where friction becomes much more important and tightening a knot could significantly increase the mechanical resistance.

4.3 Experimental Section

4.3.1 Protein Engineering

The cysteine variant of AFV3-109 (K98C) was constructed using standard site-directed mutagenesis methods. Protein $(GB1)_4$-K98C was constructed and expressed the same way as described in Chapter 3 (3.4.1). The residue Cys98 of K98C was oxidized by air at room temperature for 2 hours to form a disulfide bond between two neighboring $(GB1)_4$-K98C molecules to form the dimer $(GB1)_4$-K98C-K98C-$(GB1)_4$ (Figure 4.12).

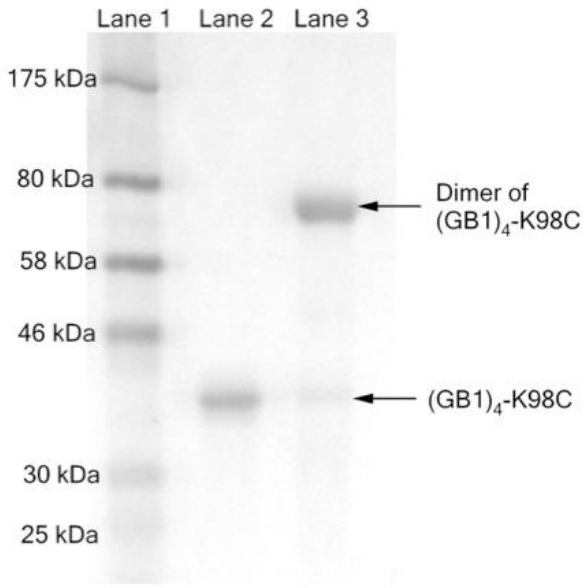

Figure 4.12 Dimerization of $(GB1)_4$-K98C.

Coomassie blue stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) photograph of $(GB1)_4$-K98C. SDS-PAGE analysis shows that $(GB1)_4$-K98C can be readily oxidized into a dimer $(GB1)_4$-K98C-K98C-$(GB1)_4$. Lane 1: pre-stained protein molecular weight marker; Lane 2: $(GB1)_4$-K98C reduced by DTT; Lane 3: air oxidized $(GB1)_4$-K98C. Under oxidized condition, most $(GB1)_4$-K98C exists as a dimer $(GB1)_4$-K98C-K98C-$(GB1)_4$.

4.3.2 SMFS Experiments

SMFS experiments were carried out the same way as described in chapter 3 (3.4.2).

4.3.3 Monte Carlo Simulations

Monte Carlo simulations were done using the same method as described in chapter 3 (3.4.3).

4.3.4 SMD Simulations

Constant force and constant velocity SMD simulations were performed using NAMD[88] and CHARMM force field with CMAP correction. [168,178] The structure of K98C was obtained by mutating the wild-type AFV3-109 (PDB code: 2J6B) using Visual Molecular Dynamics (VMD).[27] The protein was solvated using FACTS, an implicit solvent environment, and the ensemble was NVT. [156,179] The system was energetically minimized and equilibrated for 1 ns before applying pulling force on the N-terminus and residue 98. Pulling forces varied from 800 to 1500 pN in constant force mode, and pulling velocity ranged from 1 to 50 nm/ns in constant velocity mode. Three trajectories were obtained at each constant pulling velocity and constant force, with the exception that at constant force of 1200 pN, 20 trajectories were obtained. VMD and IGOR Pro were used for data analysis.

Chapter 5: Tying a Protein Slipknot against Force

The folding mechanism of knotted/slipknotted proteins has become a challenging question over the last two decades. Understanding how these proteins fold would be a great achievement in the field of protein folding. Recent advances in protein folding shed light on this question using both experimental and computational methods.[10,100,101,106,110,111,180-182] Most experimental studies on the folding of knotted proteins have been limited to bulk experiments.[90,95,98,101,102,104-106,109] Single-molecule fluorescence resonance energy transfer has been used to study the location, size and movement of the knot in a denatured protein. [183] Experimental study of the folding of knotted protein remains a mystery at the single-molecule level. The results from AFM-based SMFS in chapters 3 and 4 suggest that it is possible to observe the refolding of the slipknotted protein AFV3-109.[113,114,148] However, due to the limitation of force resolution and poor long term stability of AFM, features of folding are still unclear. Optical tweezers, developed for SMFS experiments, have been proven to be a powerful tool for protein folding studies due to its superior force resolution and long term stability.[55,75-85] It is possible to work on one single molecule protein for hours using optical tweezers with very low force drift of the system. Hence, here we make use of the advantages associated with the optical tweezers to revisit the folding of slipknotted protein AFV3-109 at single molecule level. Our results from optical tweezers suggest a simple two-state folding for AFV3-109, which is fairly common for many small proteins. This suggests that tying a protein slipknot does not necessarily create a topological barrier that is difficult to overcome.

5.1 Results

5.1.1 Direct Observation of Refolding of AFV3-109

To investigate the folding of AFV3-109 by optical tweezers, we use the previously constructed Cys-GB1-AFV3-Cys to form a protein-DNA chimera. DNA handles with biotin and digoxin are coupled to the N- and C-termini cysteine residues of Cys-GB1-AFV3-Cys using the methods described in the experimental section 5.3.1. GB1 is mechanically stable[158,159] and should not be unfolded before the unfolding of AFV3-109, which we have proved in our AFM experiments in chapter 3 when we are pulling from N- and C-termini of AFV3-109. Only the unfolding and refolding of AFV3-109 could be observed if the force is limited at a low level (below 15 pN) and thus the mechanical unfolding and refolding of GB1 will not affect the analysis of our results.

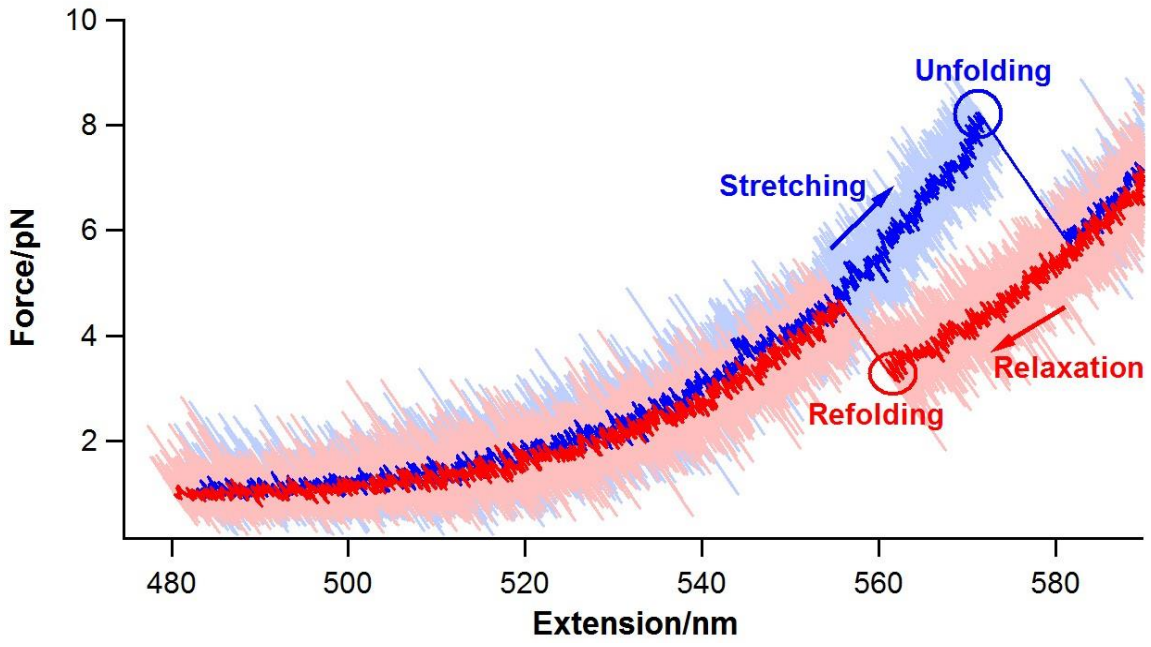

Figure 5.1 Representative FEC of stretching and relaxation of AFV3-109 using optical tweezers.
The pulling speed is 20 nm/s. Light blue and light red traces are the raw data for stretching and relaxation at a sampling rate of 1 kHz, respectively. Traces filtered to 100 Hz are shown in dark blue and red. The pulling directions are indicated by arrows. The unfolding and refolding events are circled in blue and red, respectively.

In Figure 5.1, the force peak (blue circle) in the stretching trace, which is similar to the protein unfolding signature in AFM experiments, corresponds to the unfolding of AFV3-109. In the relaxation trace, the sudden force increase indicated by the red circle implies a shortening of the polypeptide chain due to the refolding of AFV3-109. The extension and shortening of the polypeptide chain during unfolding and refolding transitions are consistent with theoretical calculations using the WLC model[73] with a persistence length of 0.8 nm[75] ($\Delta L_{C(measured)}$~38 nm). Figure 5.2 shows a narrow distribution of folding force centered at ~4 pN and a broader distribution of unfolding force ranging from 4 to 12 pN. As applied onto the N- and C-termini of AFV3-109, the stretching force is unzipping the β1, β4 and β5, which is typically predicted to be mechanically labile. The low unfolding force (~10 pN) of AFV3-109 obtained from optical tweezers experiments agrees with this prediction.

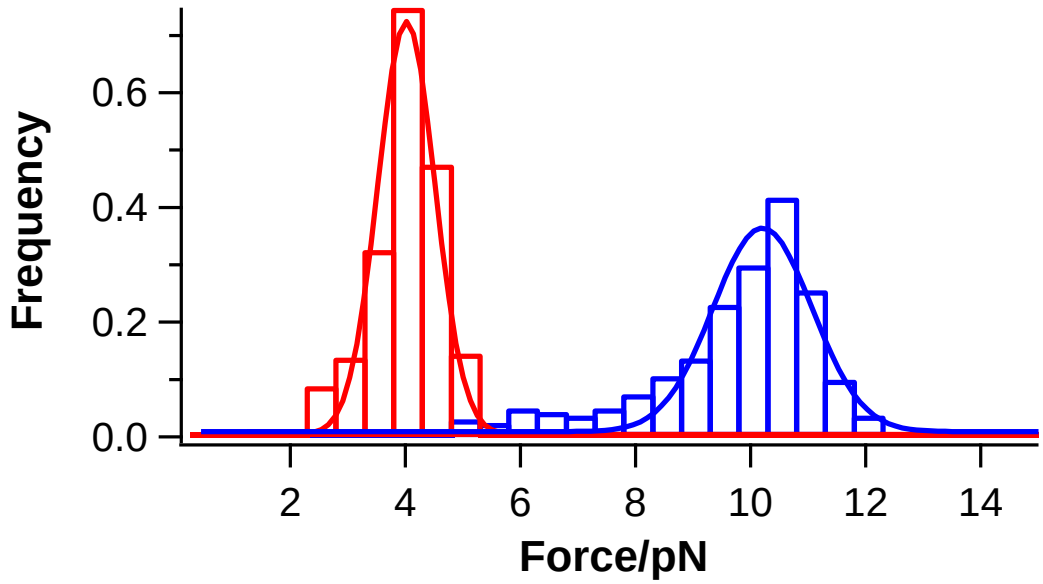

Figure 5.2 Normalized histograms of unfolding and refolding forces.

Histograms are experimental results and solid lines are Gaussian fits of unfolding (blue) and refolding (red) forces, respectively. The unfolding (blue) and refolding forces are 10±1 pN and 3.8±0.7 pN, respectively. The total number of unfolding and refolding events is 882.

To further characterize the kinetics of the folding and unfolding of AFV3-109, we extracted the folding and unfolding rates at different forces from the force-extension data using the methods reported by Oberbarnscheidt et al[184] (see experimental section 5.3.3), which are plotted in Figure 5.3. As the force is decreased, the unfolding rate is decreased, whereas the folding rate is increased. When force is below ~4 pN, the folding rate is much faster than unfolding rate. The histogram of unfolding force in Figure 5.2 also indicates that the unfolding hardly occured below 4 pN.

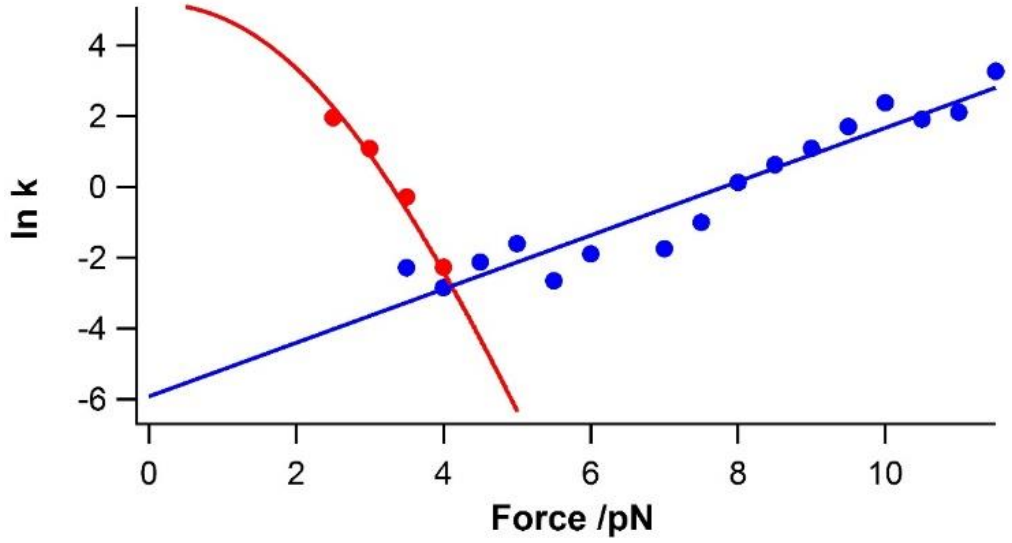

Figure 5.3 Force-dependent unfolding rate (blue) and folding rate (red) of AFV3-109.
Filled circles are experimental results and solid lines are fittings using Zhurkov-Bell-Evans model for unfolding and modified model for folding.

Similar as described in chapter 2 for NuG2, the mechanical unfolding of AFV3-109 can be described as $\alpha(F) = \alpha_0 \exp(F\Delta x_u/k_BT)$ using the Zhurkov-Bell-Evans model and the refolding rate can be described as $\beta(F)=\beta_0 \exp(-\Delta G(F,F_T)/k_BT)$. After fitting our results using these models, we have measured an unfolding rate at zero force α_0 of 0.005 s^{-1}, an unfolding distance of 3.5 nm and a folding rate at zero force β_0 of 180 s^{-1}.

Force-extension experiments are non-equilibrium experiments, as we discussed in chapter 2. The unfolding and refolding of AFV3-109 clearly shows an energetic hysteresis as shown in Figure 5.1. Thus, it enables us to determine the equilibrium free energy using Crooks fluctuation theorem (CFT).[141,142] The histograms of unfolding and refolding work are plotted in Figure 5.4 using the method introduced in chapter 2. The free energy of AFV3-109 is measured to be ~ 13.5 k_BT by CFT.

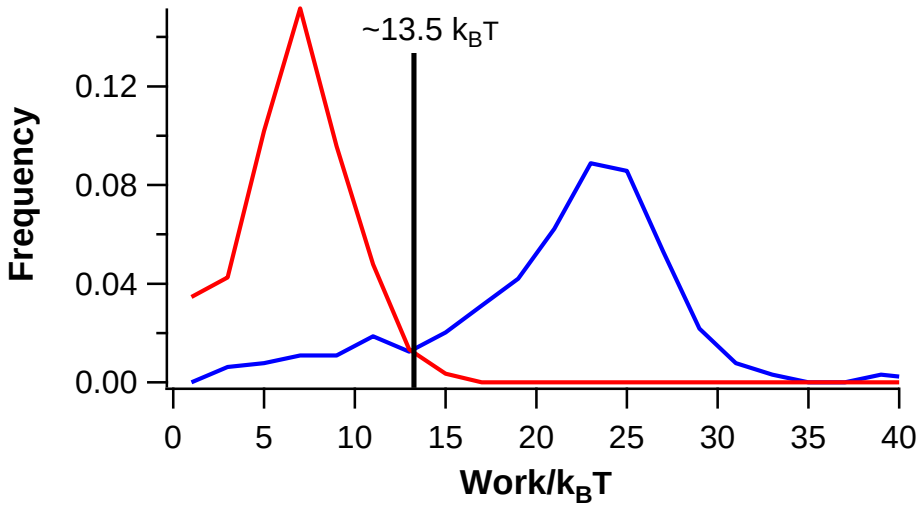

Figure 5.4 CFT analysis of AFV3-109 unfolding/folding.

Normalized histograms (solid lines) represent unfolding work (blue) and refolding work (red). The intersection of unfolding and refolding work (~ 13.5 k_BT) is measured.

To verify the results from CFT analysis, we carried out bulk experiment to determine the free energy of the unfolding of AFV3-109 to be ~ 7.4 kcal/mol, equivalent to 12.5 k_BT by measuring tryptophan fluorescence at different concentrations of guanidine (see experimental section 5.3.4). It is in close agreement with our result from CFT analysis on the data from SMFS. Thus, the CFT analysis and chemical denaturation indicate that AFV3-109 has successfully refolded to its native structure with the slipknotted topology in our SMFS experiments.

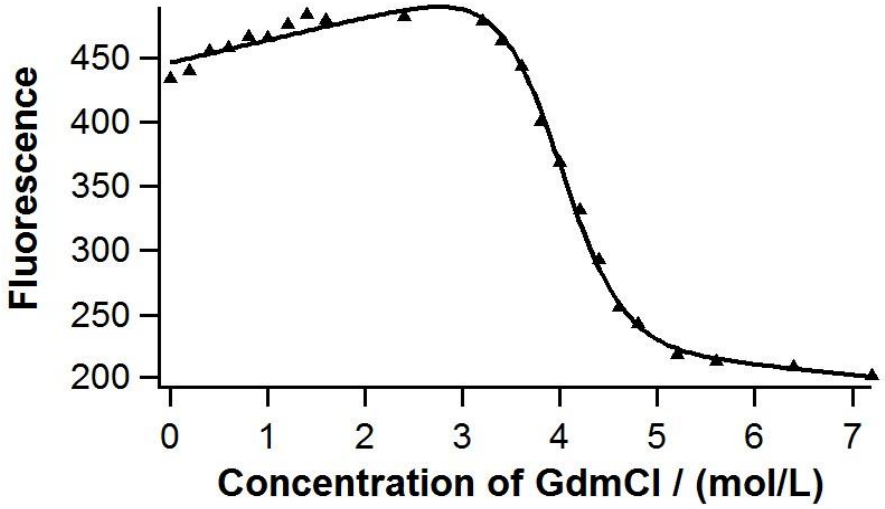

Figure 5.5 Chemical denaturation of AFV3-109 by Tryptophan fluorescence.

The experimental results of fluorescence are shown by the solid triangles and fitted by the solid line using a global fitting method described in experimental section 5.3.3. The free energy is measured to be ~ 7.4 kcal/mol, which is equivalent to 12.5 k_BT.

5.2 Discussion

5.2.1 Comparison of the Results from AFM and Optical Tweezers Experiments

In chapter 3, we studied AFV3-109 using single-molecule AFM. The results from optical tweezers are different from the results from AFM.

5.2.1.1 Different Unfolding Behaviors

The unfolding force obtained by optical tweezers is much smaller than the force obtained by AFM. This can be partially explained by the slower pulling speed used in optical tweezers experiments. A similar phenomenon was observed in chapter 2 for the unfolding of NuG2. The unfolding force of NuG2 can be as low as 10 pN at low pulling speed, while it is much higher at high pulling speeds. More importantly, the unfolding distance of AFV3-109 measured by optical tweezers (3.5 nm) is dramatically larger than the value measured by AFM (0.24 nm), which could lead to a much lower unfolding force. For such a long distance from native state

to transition state, the Zhurkov-Bell-Evans may not be suitable for describing the unfolding. More sophisticated models such as Dudko-Hummer-Szabo model[185] may provide more accurate estimation of the kinetics parameters of unfolding. In addition, in our previous study using AFM, the unfolding of AFV3-109 occurs along two different pathways, two-state and three-state with an intermediate state at ~ 12 nm. However, in the results from optical tweezers, we did not observe any intermediate state. This raises the question of whether the same protein behaves differently in similar but different stretching experiments. Optical tweezers are different from AFM in terms of the working principle. First, the stiffness of the optical trap (~0.25 pN/nm) is much less than the stiffness of AFM cantilever (~40 pN/nm). Second, the molecular linker, i.e. the DNA handle, is much longer in optical tweezers experiment than it in AFM experiment. Theoretical studies have demonstrated that the stiffness of the cantilever or optical trap can affect the free energy landscape of the whole system composed of the protein and the spring and thus affect the kinetics of protein unfolding/folding.[186] However, there is no systematic experimental investigation of this issue and so the reason behind this difference remains unclear.

5.2.1.2 Different Folding Behaviors

On the other hand, our results from the optical tweezers clearly show that AFV3-109 can always successfully refold to its native state at ~4 pN. However, in the results from the AFM experiment, AFV3-109 does not always refold to its native state. As previously discussed in chapters 3 and 4, AFM has its limitation in protein folding studies due to its relatively poor force resolution and long term stability. In the AFM experiment, when the probe is brought back to the substrate, it is possible that there is still a residual force that can prevent the folding of the proteins. As long as the residual force is larger than 4 pN, which is too low to be detected

in AFM experiments, the folding of AFV3-109 may not be observed. In addition, the lower stiffness of optical tweezers could also affect the folding kinetics but, again, the effect of the stiffness on the folding experiment is still unclear. In addition, it is also of note that a small fraction of the relaxation traces in optical tweezers experiments show that the folding events of AFV3-109 are not clear two-state transitions as the unfolding events are. Figure 5.6 showed a different folding pathway with a continuous transition between the unfolded and native states instead of a sharp transition as shown in Figure 5.1. Currently, we do not have enough data to reveal the mechanism of this folding pathway. But our results suggest that more effort should be made to explore the details about this continuous transition. It is possible that in this pathway, the protein refolds through an equilibrium-like transition similar to the B-S transitions between double-strand (B-form) and single-strand (S-form) forms of DNA in stretching and relaxation of a double-strand DNA.[187]

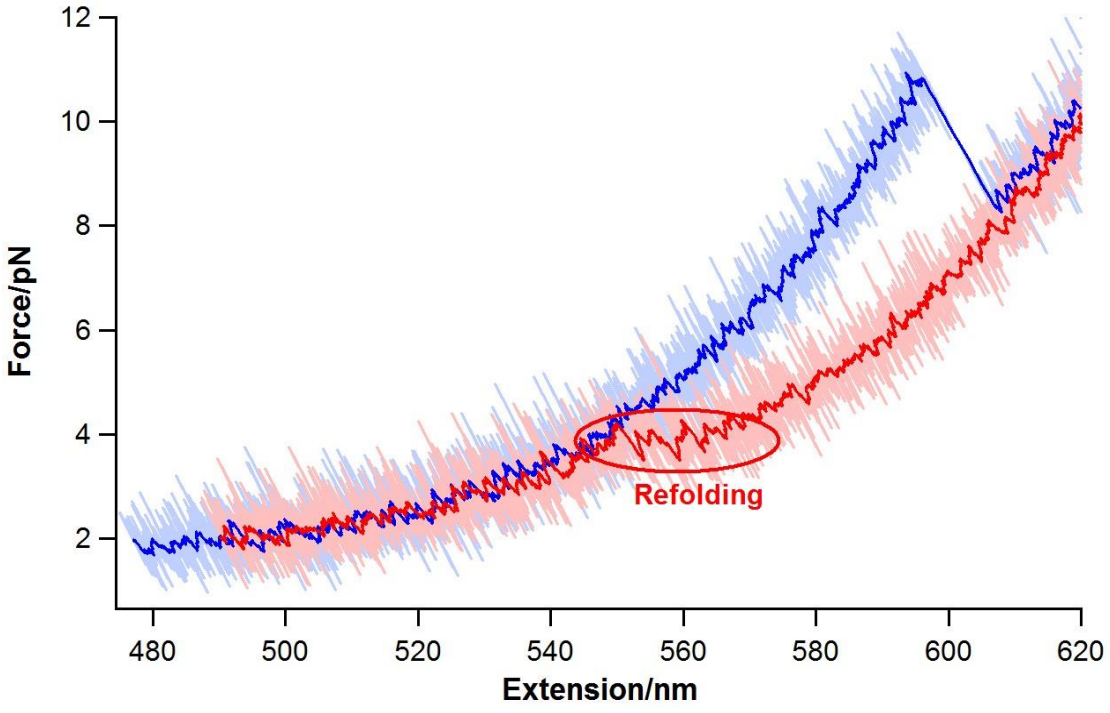

Figure 5.6 Refolding of AFV3-109 through a continuous transition.

The coloring is the same as Figure 5.1. The continuous folding transition is circled in red.

5.2.2 The Slipknotted Topology Does Not Prevent the Folding of AFV3-109

Theoretical studies have revealed that the formation of a slipknot can greatly reduce the topological difficulty of the folding from a linear polypeptide chain into a real knotted structure, which means the energy barrier associated with slipknotted topology during the folding process is low. Indeed, our results from optical tweezers has demonstrated that AFV3-109 can successfully fold into a slipknotted structure, even against external force. As we described in chapter 3, AFV3-109 has ~ 30 amino acid residues including β4 and β5 in threaded loop. The most difficult part of forming a slipknot is to insert the threaded loop into the knotting loop. The β1 and β2 stabilizes β4 and β5, respectively, which facilitates the process of loop insertion and reduces the energy barrier. However, AFV3-109 is a small protein and the topological hindrance can increase dramatically as the size of the threaded loop increases. Further study on much larger slipknotted proteins could provide more insights into the mechanism of the slipknot formation.

5.3 Experimental Section

5.3.1 Preparation of DNA-protein Chimera

The method for preparing protein Cys-GB1-AFV3-109-Cys is described in Chapter 3 (3.4.1). Two 802-bp DNA handles are prepared by regular PCR reactions with modified primers as previously reported. [86] IDT company synthesized the primers, 5'thiolmodifier C6-SS- CGA-CGA-TAA-ACG-TAA-GGA-CAT-C, 5'biotin-CAA-AAA-ACC-CCT-CAA-GAC-CC and 5'digoxigenin-CAA-AAA-ACC-CCT-CAA-GAC-CC. The template, pGEMEX-1 plasmid DNA, was a kind gift from Professor Xiaohui Frank Zhang at Lehigh University. The Onetaq DNA polymerase was purchased from New England Biolabs. PCR reactions followed the

regular protocol with 20 mM DTT added. Hispeed Plasmid Maxi kit from Qiagen was used to purify the PCR products. After purification, the DNA handles were allowed to react with 1.5 mM 2,2'-dithio-dipyridine (DTDP) to activate the 5'thiol. Excess DTDP was removed by precipitation of DNA using isopropanol. Before being coupled with DNA handles, 10 μM of cys-GB1-AFV3-109-cys protein was treated with 2 mM DTT for 2 hours at room temperature. Zeba desalting columns were used to remove the DTT. Then 1 μL of 1 μM cys-GB1-AFV3-109-cys was allowed to react with 1 μL of 2 μM DTDP-activated DNA handles in Tris buffer at pH of 7 overnight at room temperature. Coupled DNA-protein chimera is then diluted to ~ 5 nM and stored at -80 C°.

5.3.2 SMFS Using Optical Tweezers

The optical tweezers experiments were carried out using MiniTweezers purchased from Smith's Lab according to previously reported method. In each experiment, 1 μL of 0.5% streptavidin modified polystyrene beads (diameter of ~ 2 um, purchased from Spherotech) was diluted to 3 mL and injected to the liquid chamber of optical tweezers. A single streptavidin modified polystyrene bead was captured by a laser beam and fixed by a micro pipette tip within the chamber. 1 μL of 5 nM DNA-protein chimera was allowed to react with 1 μL of 0.5% anti-digoxigenin modified polystyrene beads (diameter of ~ 2 um, purchased from Spherotech) for 15 min at room temperature. The mixture was then diluted to 3 mL and injected into the chamber. A single anti-digoxigenin modified polystyrene bead was captured by laser beam. The laser beam controlled the movement of anti-digoxigenin modified polystyrene bead against the streptavidin modified polystyrene bead fixed on the pipette tip to carry out the force-extension experiments.

5.3.3 Extracting the Kinetics of Folding and Unfolding

We used previously reported method[184,188] to analyze the FECs. The FECs were divided into many small time windows (Δt) that is small enough so that the force can be considered constant within a time window. The probability of protein folding/unfolding within within one time window Δt can be calculated as $P(F)=N(F)/M(F)$, where the $N(F)$ is the total number of all the folding or unfolding events at the force of F and $M(F)$ is the total number of time windows at the force of F. The rate constant of protein folding/unfolding at the force of F can be calculated as $k(F)= P(F)/\Delta t$.

5.3.4 Chemical Denaturation

Chemical denaturation experiments were carried out on a Cary Eclipse fluorescence spectrophotometer. The concentration of protein in fluorescence measurement was 0.2 mg/mL. The tryptophan fluorescence of protein at 350 nm was monitored at different concentration of guanidine hydrochloride. The result of fluorescence is fitted using the following equation:

$$f(D) = ((aN+bN\times D)+(aD+bD\times D) \times\exp(m\times (D-D_{1/2})/0.592))/(1+\exp(m\times (D-D_{1/2})/0.592))$$

where D is the concentration of guanidine hydrochloride, aN and aD are the y-intercepts of native and denatured sections, respectively, bN and bD are the slopes of native and denatured sections, respectively, m is the slope of transition section and $D_{1/2}$ is the concentration of guanidine hydrochloride at which half of the protein is in unfolded state.

Chapter 6: Conclusions

6.1 Summary

Protein folding/unfolding has been investigated for years, but there are always new and notable topics that are interesting. Proteins with knotted/slipknotted topology are certainly one of the most interesting and challenging problems. This thesis uses a small protein with slipknotted topology as a model to study the folding/unfolding of a knotted/slipknotted protein. With the help from force spectroscopy based on AFM and optical tweezers, we have demonstrated how to change the topology of the slipknotted protein and revealed the folding/unfolding mechanism.

In chapter 2, we demonstrated the power of AFM in studying the protein folding/unfolding using modified cantilevers with improved long-term stability and force resolution. We directly observed the real-time folding and unfolding of a small protein NuG2 near equilibrium, allowing us to directly determine the equilibrium folding and unfolding free energy of proteins from single-molecule experiments for the first time. Our studies expand the range of proteins that can be examined using single-molecule AFM under conditions that approach equilibrium, which may allow for the examination of elastomeric proteins, such as the giant muscle protein titin and extracellular matrix protein fibronectin, at the forces that are physiologically relevant. Such experiments would greatly provide further understanding of the elastic behaviours of these elastomeric proteins in their biological settings.

In chapters 3 and 4, by combining single molecule AFM and molecular dynamics simulations, we detailed the complex mechanical unfolding of a small slipknotted protein AFV3-109, as well as the conversion of its slipknot conformation into a linear chain or a tightened trefoil knot. By stretching AFV3-109 across its N- and C-termini, we have successfully pulled the

threaded loop out of the knotting loop and thus untied the slipknot. By stretching AFV3-109 across its N-terminus and residue 98 on the threaded loop, we were able to pull the threaded loop through and into the knotting loop, thus converting a slipknot into a true knot. The mechanical unfolding of AFV3-109 can proceed via multiple parallel pathways: AFV3-109 can unfold in a two-state fashion as well as through three-state pathways, involving the formation of distinct intermediate states. Upon being pulled from its N-terminus and residue 98, the slipknot is stretched into a tightened trefoil knot, which involves ~13 amino acid residues and leads to a shortening of the fully extended polypeptide chain by ~4.7 nm. SMD simulation results confirm unfolding pathways that are consistent with our experimental results, providing a plausible molecular mechanism describing the mechanical unfolding that either unties the slipknot or tightens the trefoil structure. Our study demonstrates that force spectroscopy methods are a powerful tool to manipulate the knotted protein structure, paving the way towards investigating the folding mechanism of highly complex knotted and slipknotted proteins.

In chapter 5, we revisit the mechanical unfolding and folding of AFV3-109 by applying force on N- and C-termini using optical tweezers, which provides much better stability and force resolution than AFM. In contrast with the results from AFM experiments, AFV3-109 is able to efficiently refold to its native structure with slipknotted topology from a linear extended polypeptide chain at ~ 4 pN, which is very difficult to detect in AFM experiments. In addition, the simple two-state folding of AFV3-109 implicates that the energy barrier of forming the slipknotted topology is not difficult to overcome even against an external pulling force. This result supports the hypothesis that forming a slipknot can reduce the topological difficulty during the folding of knotted protein.

6.2 Future Directions

It is clear that knots exist in proteins and the number of knotted proteins is increasing as more protein structures are solved. However, only ~1% of the proteins with known structures have knotted topology.[93] It is still unclear whether the knot is essential for their biological function or it is just an evolutional mistake.[97,102,189,190]

A lot of efforts have been put into understanding the relationship between knotted topology and biological functions of knotted proteins.[97,102,189,190] This is a very challenging task because the whole structure of a knotted protein would be broken if the knot is untied. A possible way is to locate the knot in a protein and determine whether it is related to the active site of the protein. Recent studies have shown that the active sites of proteins can be modified and the enzymatic activity can be subsequently changed by the presence of a knot[15]. Another important finding is that the knotted topology is able to increase thermal stability[11] or resistance against degradation[15] of the protein. More efforts are expected to unveil the biological role of the knot in proteins.

On the other hand, many of the studies on knotted proteins are theoretical studies using computational methods, e.g. molecular dynamics simulations, which provide a lot of valuable insights about the folding/unfolding mechanisms.[108,110,112,173,181,191,192] It is important to verify these simulation results with convincing experimental evidence. Our study on the unfolding of AFV3-109 has shown a good way of combining both experimental and theoretical methods.

Furthermore, this thesis only studies the simplest slipknotted protein using force spectroscopy. However, folding/unfolding could be more complicated when the threaded loop is much larger. Thus, a slipknotted protein with large threaded loop such as thymidine kinase would be a good

candidate for studying the topological complexity of folding/unfolding of slipknotted proteins.[121]

Although the folding/unfolding of a slipknotted protein can help us understand the folding of knotted proteins, studying on knotted protein would be a more direct way for this purpose. There are many knotted proteins with more complex structures and knot types.[20,97] Each individual type of knot should have different topological complexity during the folding process. In the bigger picture, other biomolecules such as nucleic acids also possess knotted structures. For example, pseudo-knots exist widely in RNA[193] and many DNA molecules have knotted structures.[194] An important conference about the knotted structures of protein and nucleic acids was held in 2014 to discuss the significance of the knot. Interdisciplinary studies, involving not only physics, chemistry and biology, but mathematics as well, are necessary to expand the understanding of the knotted problem in biomolecules.

In summary, this thesis has demonstrated that force spectroscopy is a great technique to study knotted/slipknotted protein experimentally at a single-molecule level. The topology of a protein slipknot can be altered between linear chain, slipknot and trefoil knot. Future inspiring studies are expected to further reveal the mystery of knotted/slipknotted proteins.

Bibliography

(1) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. *Annu Rev Phys Chem* **1997**, *48*, 545.

(2) Borgia, A.; Williams, P. M.; Clarke, J. *Annual Review of Biochemistry* **2008**, *77*, 101.

(3) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.

(4) Wolynes, P. G. *Quarterly Reviews of Biophysics* **2005**, *38*, 405.

(5) Thirumalai, D. *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions* **1994**, *325*, 115.

(6) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. *Theoretical Chemistry Accounts* **1997**, *96*, 14.

(7) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annual Review of Biophysics, Vol 39* **2010**, *39*, 159.

(8) Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.

(9) Mansfield, M. L. *Nat Struct Biol* **1994**, *1*, 213.

(10) Taylor, W. R. *Nature* **2000**, *406*, 916.

(11) Taylor, W. R.; Lin, K. *Nature* **2003**, *421*, 25.

(12) Takusagawa, F.; Kamitori, S. *J Am Chem Soc* **1996**, *118*, 8945.

(13) Alam, M. T.; Yamada, T.; Carlsson, U.; Ikai, A. *Febs Letters* **2002**, *519*, 35.

(14) Lua, R. C.; Grosberg, A. Y. *Plos Computational Biology* **2006**, *2*, 350.

(15) Virnau, P.; Mirny, L. A.; Kardar, M. C. P. *PLoS Comput Biol* **2006**, *2*, e122.

(16) King, N. P.; Yeates, E. O.; Yeates, T. O. *Journal of Molecular Biology* **2007**, *373*, 153.

(17) Kolesov, G.; Virnau, P.; Kardar, M.; Mirny, L. A. *Nucleic Acids Res* **2007**, *35*, W425.

(18) Lai, Y. L.; Yen, S. C.; Yu, S. H.; Hwang, J. K. *Nucleic Acids Res* **2007**, *35*, W420.

(19) Taylor, W. R. *Computational Biology and Chemistry* **2007**, *31*, 151.

(20) Bolinger, D.; Sulkowska, J. I.; Hsu, H. P.; Mirny, L. A.; Kardar, M.; Onuchic, J. N.; Virnau, P. *Plos Comput Biol* **2010**, *6*.

(21) Dobson, C. M.; Sali, A.; Karplus, M. *Angewandte Chemie-International Edition* **1998**, *37*, 868.

(22) Luque, I.; Leavitt, S. A.; Freire, E. *Annual Review of Biophysics and Biomolecular Structure* **2002**, *31*, 235.

(23) Sanger, F. *Adv Protein Chem* **1952**, *7*, 1.

(24) Schellman, J. A.; Schellman, C. G. *Protein Sci* **1997**, *6*, 1092.

(25) Moss, G. P. *Pure Appl Chem* **1996**, *68*, 2193.

(26) Chou, K. C.; Cai, Y. D. *Proteins* **2003**, *53*, 282.

(27) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph Model* **1996**, *14*, 33.

(28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* **2000**, *28*, 235.

(29) Andreeva, A.; Howorth, D.; Chothia, C.; Kulesha, E.; Murzin, A. G. *Nucleic Acids Res* **2014**, *42*, D310.

(30) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J Mol Biol* **1995**, *247*, 536.

(31) Debrunner, P.; Tsibris, J. C. M.; Münck, E.; University of Illinois at Urbana-Champaign. Dept. of Chemistry.; University of Illinois at Urbana-Champaign. Dept. of Physics. *Mössbauer spectroscopy in biological systems; proceedings of a meeting held at Allerton House, March 17 and 18, 1969, Monticello, Ill., organized by the University of Illinois at Urbana-Champaign, Departments of Chemistry and Physics*; University of Illinois: Urbana, 1969.

(32) Nymeyer, H.; Onuchic, J. N. *Abstr Pap Am Chem S* **1997**, *214*, 244.

(33) Socci, N. D.; Nymeyer, H.; Onuchic, J. N. *Physica D* **1997**, *107*, 366.

(34) Onuchic, J. N. *P Natl Acad Sci USA* **1997**, *94*, 7129.

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

(36) Wright, C. F.; Lindorff-Larsen, K.; Randles, L. G.; Clarke, J. *Nature Structural Biology* **2003**, *10*, 658.

(37) Zhuang, X. W.; Rief, M. *Curr Opin Struc Biol* **2003**, *13*, 88.

(38) Onuchic, J. N.; Wolynes, P. G. *Current Opinion in Structural Biology* **2004**, *14*, 70.

(39) Karplus, M.; Weaver, D. L. *Protein Sci* **1994**, *3*, 650.

(40) Nolting, B.; Agard, D. A. *Proteins-Structure Function and Bioinformatics* **2008**, *73*, 754.

(41) Travaglini-Allocatelli, C.; Ivarsson, Y.; Jemth, P.; Gianni, S. *Curr Opin Struc Biol* **2009**, *19*, 3.

(42) Ivanyushina, V. A.; Meriin, A. B.; Kiselev, O. I. *Mol Biol+* **1991**, *25*, 679.

(43) Ellis, R. J.; Vandervies, S. M. *Annu Rev Biochem* **1991**, *60*, 321.

(44) Durand, R. *M S-Med Sci* **1991**, *7*, 496.

(45) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042.

(46) Onuchic, J. N.; Socci, N. D.; LutheySchulten, Z.; Wolynes, P. G. *Fold Des* **1996**, *1*, 441.

(47) Anson, M. L. *Adv Protein Chem* **1945**, *2*, 361.

(48) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat Struct Biol* **1996**, *3*, 193.

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

(50) Zhurkov, S. N. *Int J Fracture* **1984**, *26*, 295.

(51) Bell, G. I. *Science* **1978**, *200*, 618.

(52) Evans, E.; Ritchie, K. *Biophys J* **1997**, *72*, 1541.

(53) Jagannathan, B.; Marqusee, S. *Biopolymers* **2013**, *99*, 860.

(54) Lee, W.; Zeng, X. C.; Zhou, H. X.; Bennett, V.; Yang, W. T.; Marszalek, P. E. *Journal of Biological Chemistry* **2010**, *285*, 38167.

(55) Zoldak, G.; Rief, M. *Curr Opin Struc Biol* **2013**, *23*, 48.

(56) Fernandez, J. M.; Li, H. B. *Science* **2004**, *303*, 1674.

(57) Javadi, Y.; Fernandez, J. M.; Perez-Jimenez, R. *Physiology* **2013**, *28*, 9.

(58) Marszalek, P. E.; Dufrene, Y. F. *Chem Soc Rev* **2012**, *41*, 3523.

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

(60) Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H. B.; Fernandez, J. M. *Prog Biophys Mol Bio* **2000**, *74*, 63.

(61) Chen, H.; Yuan, G. H.; Winardhi, R. S.; Yao, M. X.; Popa, I.; Fernandez, J. M.; Yan, J. *J Am Chem Soc* **2015**, *137*, 3540.

(62) Moy, V. T.; Florin, E. L.; Rief, M.; Lehmann, H.; Ludwig, M.; Gaub, H. E.; Dornmair, K. *Scanning Probe Microscopies Iii* **1995**, *2384*, 2.

(63) Radmacher, M.; Rotsch, C.; Fritz, M.; Kacher, C. M.; Hofmann, U.; Gaub, H. E.; Hansma, P. K. *Abstr Pap Am Chem S* **1996**, *212*, 324.

(64) Jaschke, M.; Butt, H. J.; Manne, S.; Gaub, H. E.; Hasemann, O.; Krimphove, F.; Wolff, E. K. *Biosens Bioelectron* **1996**, *11*, 601.

(65) Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. *Science* **1997**, *275*, 1295.

(66) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys J* **1998**, *75*, 3008.

(67) Fernandez, J. M.; Gautel, M.; Rief, M.; Gaub, H. E. *Nato Adv Sci I C-Mat* **1999**, *519*, 319.

(68) Gaub, H. E. *Abstr Pap Am Chem S* **2001**, *221*, U324.

(69) Kuhner, F.; Morfill, J.; Neher, R. A.; Blank, K.; Gaub, H. E. *Biophys J* **2007**, *92*, 2491.

(70) Puchner, E. M.; Alexandrovich, A.; Kho, A. L.; Hensen, U.; Schafer, L. V.; Brandmeier, B.; Grater, F.; Grubmuller, H.; Gaub, H. E.; Gautel, M. *P Natl Acad Sci USA* **2008**, *105*, 13385.

(71) Gaub, B. M.; Kaul, C.; Zimmermann, J. L.; Carell, T.; Gaub, H. E. *Nanotechnology* **2009**, *20*.

(72) Gaub, H. *Biophys J* **2010**, *98*, 410a.

(73) Marko, J. F.; Siggia, E. D. *Macromolecules* **1995**, *28*, 8759.

(74) Li, H. *Org Biomol Chem* **2007**, *5*, 3399.

(75) Stigler, J.; Ziegler, F.; Gieseke, A.; Gebhardt, J. C. M.; Rief, M. *Science* **2011**, *334*, 512.

(76) Rief, M. *Biophys J* **2011**, *100*, 373.

(77) Fang, J.; Mehlich, A.; Koga, N.; Huang, J. Q.; Koga, R.; Gao, X. Y.; Hu, C. G.; Jin, C.; Rief, M.; Kast, J.; Baker, D.; Li, H. B. *Nat Commun* **2013**, *4*.

(78) Zoldak, G.; Stigler, J.; Pelz, B.; Li, H. B.; Rief, M. *P Natl Acad Sci USA* **2013**, *110*, 18156.

(79) Rognoni, L.; Most, T.; Zoldak, G.; Rief, M. *P Natl Acad Sci USA* **2014**, *111*, 5568.

(80) Stigler, J.; Rief, M. *Biophys J* **2015**, *109*, 365.

(81) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309*, 2057.

(82) Cecconi, C.; Shank, E. A.; Dahlquist, F. W.; Marqusee, S.; Bustamante, C. *Eur Biophys J Biophy* **2008**, *37*, 729.

(83) Kaiser, C.; Goldman, D.; Tinoco, I.; Bustamante, C. *Biophys J* **2012**, *102*, 68a.

(84) Leachman, S. M.; Wilson, C. A. M.; Marqusee, S.; Bustamante, C. *Biophys J* **2014**, *106*, 469a.

(85) Lee, A.; Rosenbloom, A. B.; Lee, S. H.; Bustamante, C. *Biophys J* **2014**, *106*, 25a.

(86) Zhang, X. H.; Halvorsen, K.; Zhang, C. Z.; Wong, W. P.; Springer, T. A. *Science* **2009**, *324*, 1330.

(87) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.;

Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J Comput Chem* **2009**, *30*, 1545.

(88) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J Comput Chem* **2005**, *26*, 1781.

(89) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wires Comput Mol Sci* **2013**, *3*, 198.

(90) Mallam, A. L.; Jackson, S. E. *Nat Chem Biol* **2012**, *8*, 147.

(91) Mansfield, M. L. *Nature Structural Biology* **1994**, *1*, 213.

(92) Liang, C. Z.; Mislow, K. *J Am Chem Soc* **1994**, *116*, 11189.

(93) Millett, K. C.; Rawdon, E. J.; Stasiak, A.; Sulkowska, J. I. *Biochem Soc T* **2013**, *41*, 533.

(94) Lim, K.; Zhang, H.; Tempczyk, A.; Krajewski, W.; Bonander, N.; Toedt, J.; Howard, A.; Eisenstein, E.; Herzberg, O. *Proteins* **2003**, *51*, 56.

(95) Mallam, A. L.; Jackson, S. E. *J Mol Biol* **2007**, *366*, 650.

(96) Lai, Y. L.; Chen, C. C.; Hwang, J. K. *Nucleic Acids Res* **2012**, *40*, W228.

(97) Virnau, P.; Mirny, L. A.; Kardar, M. *Plos Comput Biol* **2006**, *2*, 1074.

(98) Mallam, A. L.; Jackson, S. E. *J Mol Biol* **2005**, *346*, 1409.

(99) Erdmann, M. A. *J Comput Biol* **2005**, *12*, 609.

(100) Mallam, A. L.; Jackson, S. E. *J Mol Biol* **2006**, *359*, 1420.

(101) Wallin, S.; Zeldovich, K. B.; Shakhnovich, E. I. *J Mol Biol* **2007**, *368*, 884.

(102) Mallam, A. L.; Jackson, S. E. *Structure* **2007**, *15*, 111.

(103) Yeates, T.; Norcross, T. S.; King, N. P. *Current Opinion in Chemical Biology* **2007**, *11*, 595.

(104) Mallam, A. L.; Onuoha, S. C.; Grossmann, J. G.; Jackson, S. E. *Mol Cell* **2008**, *30*, 642.

(105) Mallam, A. L.; Morris, E. R.; Jackson, S. E. *P Natl Acad Sci USA* **2008**, *105*, 18740.

(106) Mallam, A. L. *Febs J* **2009**, *276*, 365.

(107) Sulkowska, J. I.; Sulkowski, P.; Onuchic, J. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 3119.

(108) Noel, J. K.; Sulkowska, J. I.; Onuchic, J. N. *P Natl Acad Sci USA* **2010**, *107*, 15403.

(109) Mallam, A. L.; Rogers, J. M.; Jackson, S. E. *P Natl Acad Sci USA* **2010**, *107*, 8189.

(110) Faisca, P. F. N.; Travasso, R. D. M.; Charters, T.; Nunes, A.; Cieplak, M. *Physical Biology* **2010**, *7*.

(111) Shakhnovich, E. *Nat Mater* **2011**, *10*, 84.

(112) Sulkowska, J. I.; Noel, J. K.; Onuchic, J. N. *P Natl Acad Sci USA* **2012**, *109*, 17783.

(113) He, C. Z.; Genchev, G. Z.; Lu, H.; Li, H. B. *J Am Chem Soc* **2012**, *134*, 10428.

(114) He, C. Z.; Lamour, G.; Xiao, A.; Gsponer, J.; Li, H. B. *J Am Chem Soc* **2014**, *136*, 11946.

(115) Sulkowska, J. I.; Sulkowski, P.; Szymczak, P.; Cieplak, M. *Phys Rev Lett* **2008**, *100*.

(116) Bornschlogl, T.; Anstrom, D. M.; Mey, E.; Dzubiella, J.; Rief, M.; Forest, K. T. *Biophys J* **2009**, *96*, 1508.

(117) Sulkowska, J. I.; Sulkowski, P.; Szymczak, P.; Cieplak, M. *J Am Chem Soc* **2010**, *132*, 13954.

(118) Szymczak, P. *Biochem Soc T* **2013**, *41*, 620.

(119) Dzubiella, J. *Journal of Physical Chemistry Letters* **2013**, *4*, 1829.

(120) Wang, T.; Ikai, A. *Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers* **1999**, *38*, 3912.

(121) Sulkowska, J. I.; Sulkowski, P.; Onuchic, J. N. *Phys Rev Lett* **2009**, *103*.

(122) Nauli, S.; Kuhlman, B.; Baker, D. *Nature Structural Biology* **2001**, *8*, 602.

(123) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. *Science* **2010**, *330*, 341.

(124) Schuler, B.; Hofmann, H. *Curr Opin Struc Biol* **2013**, *23*, 36.

(125) Schlierf, M.; Berkemeier, F.; Rief, M. *Biophys J* **2007**, *93*, 3989.

(126) Junker, J. P.; Ziegler, F.; Rief, M. *Science* **2009**, *323*, 633.

(127) Churnside, A. B.; Sullan, R. M. A.; Nguyen, D. M.; Case, S. O.; Bull, M. S.; King, G. M.; Perkins, T. T. *Nano Letters* **2012**, *12*, 3557.

(128) Churnside, A. B.; Perkins, T. T. *Febs Lett* **2014**, *588*, 3621.

(129) Yew, Z. T.; Schlierf, M.; Rief, M.; Paci, E. *Phys Rev E* **2010**, *81*.

(130) Junker, J. P.; Rief, M. *Angew Chem Int Edit* **2010**, *49*, 3306.

(131) Lee, W.; Zeng, X. C.; Rotolo, K.; Yang, M.; Schofield, C. J.; Bennett, V.; Yang, W. T.; Marszalek, P. E. *Biophysical Journal* **2012**, *102*, 1118.

(132) Lee, G.; Abdi, K.; Jiang, Y.; Michaely, P.; Bennett, V.; Marszalek, P. E. *Nature* **2006**, *440*, 246.

(133) Junker, J. P.; Rief, M. *P Natl Acad Sci USA* **2009**, *106*, 14361.

(134) Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D. *Protein Science* **2002**, *11*, 2924.

(135) Cao, Y.; Kuske, R.; Li, H. B. *Biophysical Journal* **2008**, *95*, 782.

(136) Evans, E. *Annual Review of Biophysics and Biomolecular Structure* **2001**, *30*, 105.

(137) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *Proc Natl Acad Sci U S A* **2012**, *109*, 17820.

(138) Rico, F.; Gonzalez, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. *Science* **2013**, *342*, 741.

(139) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *P Natl Acad Sci USA* **2012**, *109*, 17820.

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

(141) Crooks, G. E. *Phys Rev E* **1999**, *60*, 2721.

(142) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231.

(143) Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C. *Nature* **2010**, *465*, 637.

(144) Bennett, C. H. *J Comput Phys* **1976**, *22*, 245.

(145) Florin, E. L.; Rief, M.; Lehmann, H.; Ludwig, M.; Dornmair, C.; Moy, V. T.; Gaub, H. E. *Biosens Bioelectron* **1995**, *10*, 895.

(146) Rognoni, L.; Stigler, J.; Pelz, B.; Ylanne, J.; Rief, M. *P Natl Acad Sci USA* **2012**, *109*, 19679.

(147) King, N. P.; Jacobitz, A. W.; Sawaya, M. R.; Goldschmidt, L.; Yeates, T. O. *P Natl Acad Sci USA* **2010**, *107*, 20732.

(148) Keller, J.; Leulliot, N.; Cambillau, C.; Campanacci, V.; Porciero, S.; Prangishvili, D.; Forterre, P.; Cortez, D.; Quevillon-Cheruel, S.; Tilbeurgh, H. *Virology Journal* **2007**, *4*, 10.

(149) Puchner, E. M.; Gaub, H. E. *Curr Opin Struc Biol* **2009**, *19*, 605.

(150) Brockwell, D. J. *Curr Nanosci* **2007**, *3*, 3.

(151) Krammer, A.; Lu, H.; Isralewitz, B.; Schulten, K.; Vogel, V. *P Natl Acad Sci USA* **1999**, *96*, 1351.

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

(153) Lu, H.; Schulten, K. *Proteins-Structure Function and Genetics* **1999**, *35*, 453.

(154) Sharma, D.; Feng, G.; Khor, D.; Genchev, G. Z.; Lu, H.; Li, H. B. *Biophys J* **2008**, *95*, 3935.

(155) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144.

(156) Paci, E.; Karplus, M. *P Natl Acad Sci USA* **2000**, *97*, 6521.

(157) Genchev, G. Z.; Kallberg, M.; Gursoy, G.; Mittal, A.; Dubey, L.; Perisic, O.; Feng, G.; Langlois, R.; Lu, H. *Cell Biochem Biophys* **2009**, *55*, 141.

(158) Cao, Y.; Li, H. B. *Nat Mater* **2007**, *6*, 109.

(159) Cao, Y.; Lam, C.; Wang, M. J.; Li, H. B. *Angew Chem Int Edit* **2006**, *45*, 642.

(160) Zheng, P.; Cao, Y.; Li, H. B. *Langmuir* **2011**, *27*, 5713.

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

(162) Thirumalai, D.; Hyeon, C. *Biochemistry-Us* **2005**, *44*, 4957.

(163) Kiefhaber, T. *P Natl Acad Sci USA* **1995**, *92*, 9029.

(164) Mickler, M.; Dima, R. I.; Dietz, H.; Hyeon, C.; Thirumalai, D.; Rief, M. *P Natl Acad Sci USA* **2007**, *104*, 20268.

(165) Peng, Q.; Li, H. B. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 1885.

(166) Hyeon, C.; Dima, R. I.; Thirumalai, D. *Structure* **2006**, *14*, 1633.

(167) Cao, Y.; Li, H. B. *Nature Materials* **2007**, *6*, 109.

(168) 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.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J Phys Chem B* **1998**, *102*, 3586.

(169) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926.

(170) Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089.

(171) Taylor, W. R. *Nature* **2000**, *406*, 916.

(172) Sulkowska, J. I.; Sulkowski, P.; Szymczak, P.; Cieplak, M. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 19714.

(173) Sulkowska, J. I.; Rawdon, E. J.; Millett, K. C.; Onuchic, J. N.; Stasiak, A. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, E1715.

(174) Li, W. F.; Terakawa, T.; Wang, W.; Takada, S. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 18625.

(175) Beccara, S. A.; Skrbic, T.; Covino, R.; Micheletti, C.; Faccioli, P. *Plos Comput Biol* **2013**, *9*.

(176) Soler, M. A.; Faisca, P. F. N. *Plos One* **2013**, *8*.

(177) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *P Natl Acad Sci USA* **1996**, *93*, 11615.

(178) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J Comput Chem* **2004**, *25*, 1400.

(179) Haberthur, U.; Caflisch, A. *J Comput Chem* **2008**, *29*, 701.

(180) Yeh, Y. C.; Chen, T. H.; Chu, H. L.; Tsai, C. M.; Chang, C. C. *Faseb J* **2008**, *22*.

(181) Sulkowska, J. I.; Noel, J. K.; Onuchic, J. N. *Biophys J* **2013**, *104*, 370a.

(182) Wang, I.; Chen, S. Y.; Hsu, S. T. D. *J Phys Chem B* **2015**, *119*, 4359.

(183) Wang, P.; Yang, L. J.; Liu, P. C.; Gao, Y. Q.; Zhao, X. S. *Chem-Eur J* **2013**, *19*, 5909.

(184) Oberbarnscheidt, L.; Janissen, R.; Oesterhelt, F. *Biophys J* **2009**, *97*, L19.

(185) Dudko, O. K.; Hummer, G.; Szabo, A. *P Natl Acad Sci USA* **2008**, *105*, 15755.

(186) Yan, M. Y.; Hu, C.; Jie *Integrative Biology* **2015**.

(187) Cluzel, P.; Lebrun, A.; Heller, C.; Lavery, R.; Viovy, J. L.; Chatenay, D.; Caron, F. *Science* **1996**, *271*, 792.

(188) Schlierf, M.; Yew, Z. T.; Rief, M.; Paci, E. *Biophys J* **2010**, *99*, 1620.

(189) Zarembinski, T. I.; Kim, Y.; Peterson, K.; Christendat, D.; Dharamsi, A.; Arrowsmith, C. H.; Edwards, A. M.; Joachimiak, A. *Proteins* **2003**, *50*, 177.

(190) Stepanenko, O. V.; Bublikov, G. S.; Stepanenko, O. V.; Shcherbakova, D. M.; Verkhusha, V. V.; Turoverov, K. K.; Kuznetsova, I. M. *Febs J* **2014**, *281*, 2284.

(191) Tuszynska, I.; Bujnicki, J. M. *J Biomol Struct Dyn* **2010**, *27*, 511.

(192) Noel, J.; Sulkowska, J.; Onuchic, J. *Biophys J* **2012**, *102*, 458a.

(193) Staple, D. W.; Butcher, S. E. *Plos Biol* **2005**, *3*, 956.

(194) Liu, L. F.; Depew, R. E.; Wang, J. C. *J Mol Biol* **1976**, *106*, 439.

Appendix

Protein Sequences

NuG2

MDTYKLVIVL NGTTFTYTTE AVDAATAEKV FKQYANDNGV DGEWTYADAT KTFTVTE

GB1

MDTYKLILNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTE

AFV3-109

MLYILNSAIL PLKPGEEYTV KAKEITIQEA KELVTKEQFT SAIGHQATAE LLSSILGVNV PMNRVQIKVT HGDRILAFML KQRLPEGVVV KTTEELEKIG YELWLFEIQ

K98C (Mutant of AFV3-109)

MLYILNSAIL PLKPGEEYTV KAKEITIQEA KELVTKEQFT SAIGHQATAE LLSSILGVNV PMNRVQIKVT HGDRILAFML KQRLPEGVVV KTTEELECIG YELWLFEIQ