Mechanical unfolding and folding studies on proteins with high sequence identity but different conformations

by

Jinliang Li

B.S., Nanjing University, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2015

ABSTRACT

Although a few mutations can radically shift the equilibrium between denatured state and native state of a protein, it is surprising that one mutation can switch one fold into another completely different fold. Two *Streptococcus* binding domains GA and GB could be mutated so that ultimately two completely different folds had only one different amino acid in their sequences. This experiment established a mutational pathway to switch a protein's fold and function. In order to further understand the mechanism underlying this pathway, single molecule force spectroscopy was carried out using optical tweezers to investigate certain proteins along the mutational pathway to determine their mechanical stability and unfolding/folding kinetics. In this dissertation, GB's homologous protein NuG2 was studied and demonstrated that the force spectroscopy was a robust and informative tool to determine the unfolding/folding kinetics and the free energy profile of protein unfolding. Additionally, the kinetics and free energy profiles of GA and other mutants including GA30, GA77, GA95 and GB30 were characterized. These results provide a clear tendency of free energy change along the mutational pathway.

PREFACE

The work presented in this thesis is based on original work by collaborative research. Contributions by other collaborators are stated below.

Chapter 3 is based on unpublished work. My supervisor, Hongbin Li, designed the research. Chunmei Lyu performed the molecular biology experiments. Chunmei Lyu and I performed the optical tweezers experiments. I performed data analysis and simulations. Hongbin Li, Chunmei Lyu, Chengzhi and I discussed the results. I wrote the chapter and Hongbin Li edited the chapter.

Chapter 4 is based on unpublished work. My supervisor, Hongbin Li, designed the research. Chengzhi He helped me with molecular biology work. I performed the rest of the experiments, data analysis and simulations. Hongbin Li and I discussed the results. I wrote the chapter and Hongbin Li edited the chapter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AFM Atomic Force Microscopy

CFT Crooks Fluctuation Theorem

d_{aa} the contribution of each amino acid to the protein contour length

$d_{i,j}$ the theoretical contour length from amino acid i to amino acid j

DS Denatured State

DTDP 2,2'-dithioldipyridine

DTT Dithiothreitol

eWLC extensible WLC

Fc constant region of IgG

FJC Free jointed Chain

G free energy

GA Streptococcal protein G-related albumin-binding module

GAXX/GBXX the mutant of GA/GB which has XX percentage of sequence identity with GBXX/GAXX

GB B1binding domain of Streptococcal protein G

HSA Human Serum Albumin

IgG Immunoglobulin G

k rate constant

K DNA stretching modulus

k_B Boltzmann constant

k_h spring constant of a hookean spring

L contour length

M mol/L

Mad2 Mitotic Arrest Deficient 2

nm nanometer

NMR nuclear magnetic resonance

NS Native State

NuG2 a de novo designed version of GB1 in which 11 residues in the first beta hairpin were mutated to increase its folding rate

p persistence length

P probability in a specific state, such as NS and DS

PBS phosphate buffer saline

pN pico-newton

s second

t time

T temperature

TS Transition State

WLC Worm Like Chain

x extension

x_u unfolding distance, the distance between native state and transition state

x_f folding distance, the distance between denatured state and transition state

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Hongbin Li for his guidance and support. I would also like to thank Chengzhi He and all members of the research group for their help. Finally, I would like to thank Dr. Matthias Rief and Alexander Mehlich from Technische Universität München for their guidance.

1. Introduction

1.1. Concepts of protein folding

Proteins are the main structural components in cells. They do not only serve as the building blocks of cells, but are also involved in most biological processes, including molecular transportation (e.g. hemoglobin), signal transduction (e.g. cellular receptors) and catalysis of chemical reactions (e.g. numerous enzymes). It is amazing that one kind of molecules can carry out such diversified functions. To perform their functions, globular proteins which are synthesized as linear chains of amino acids must fold rapidly and reliably to a specific structure designed by evolution for the particular task. The protein folding can be described into two requirements: thermodynamic and kinetic (Karplus, & Sali, 1995; Dobson, Sali, & Karplus, 1998). The thermodynamic requirement is that the polypeptide should adopt a particular and stable folded conformation (native state, or NS) under physiological conditions. The kinetic requirement is that the denatured protein can fold into the native state within a reasonable time scale.

In terms of the thermodynamics of protein folding, it is useful to consider the free energy of a protein under native state and denatured state (DS). The free energy of the native state of a protein is located at the minimal free energy of the polypeptide where it is only slightly lower than the denatured state under physiological conditions due to a near cancellation of the energetic and entropic contributions. Various interactions which favor the native state can contribute to the driving force of protein folding, including: (i) hydrogen bond, (ii) van der Waals interactions, (iii) backbone angle preferences, (iv) electrostatic interactions and (v) hydrophobic interactions (Dill, & MacCallum, 2012). An important feature of protein folding is that the overall structure is determined by the sequence of the protein. Much insights comes from the work of Anfinsen (1972) and many others and they have demonstrated that the amino acid sequence itself contains all the

information for most proteins to fold into the tertiary structure (the overall folding pattern). The difference in sequence gives rise to the difference in the secondary structure (alpha helix and beta sheet regions) and the tertiary structure. However, it has been estimated that only approximately 1,200 different folds exist based on the Protein Data Bank, a collection of more than 80,000 protein structures at atomic level which are determined by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (Murzin et al, 1995; Holm, & Sander, 1996). This number is relatively small compared to the number of human genome sequence. One particular fold can be adopted by many different sequences but only one fold can exist for most of the sequences. The uniqueness of the native state arises from the fact the interactions mentioned above stabilize the native state and destabilize alternative folds of the sequence (Finkelstein, Badretdinov, & Gutin, 1995). That is how the protein sequence eliminate the wrongly folded structures designed by evolution.

A polypeptide chain can adopt so many potential conformations so that there must be a way to reach the native state instead of a random search (Levinthal, 1969; Zwanzig, Szabo, & Bagchi, 1992). If a protein consists of 100 amino acids and each amino acid has three conformations, there is a total of about 10^{47} potential conformations. If converting conformation takes only 10^{-11} s, reaching the unique native state still require 10^{36} s or 10^{29} years by a random search. In terms of this needle-in-a-haystack conundrum, Levinthal (1968) suggested pathways for the protein folding meaning that there exist a well-defined sequence of events to carry the protein from the denatured state to the native state. Studies of the chain entropies in models of foldable polymers showed that conformational ensembles of lower energy have fewer conformations, implying that protein folding has a funnel shape in its conformation space (Bryngelson, & Wolynes, 1987; Leopold, Montal, & Onuchic, 1992; Bryngelson et al, 1995). An unstructured sequence takes random steps

that are mostly incrementally downhill in energy to fold rapidly (Zwanzig, Szabo, & Bagchi, 1992). However, this model is still not enough to inspire computer search methods and a more narrative mechanism that explains the evolution of protein folding derived from sequence and its surrounding conditions is of demand. A mechanism need to be applicable to a wide range of proteins and it is more than just the sequences of events followed by a given protein in experimental or computational trajectories. Despite the long-standing interest in protein folding, how the protein explore the conformation space rapidly is still an open question in biophysics.

1.2. A system of two small proteins that has almost identical sequences but completely different folds

As mentioned, one of the fundamental understandings of protein folding is Anfinsen's hypothesis, describing that the protein fold is the most thermodynamically stable conformation, which corresponds to one of a few thousand unique folds. There are interesting examples of natural or engineered proteins that can have drastic change in their conformations, such as Mad2 and Arc repressor (Luo et al, 2004; Cordes et al, 2000), but they cannot be considered as counter examples of Anfinsen's hypothesis. The different conformations of these examples can be described as the fluctuations or perturbations of the native state since those conformations have similar stabilizing interactions. Even though it had been proposed that there might be sequences that can adopt multiple unrelated structures, none of the natural protein was found to have multiple unrelated structures until recently.

The human chemokine lymphotactin was found to be metamorphic and two unrelated conformations have been signaled by NMR spectroscopy under physiological conditions (Tuinstra et al, 2008; Volkman, Liu, & Peterson, 2009). According to the structure characterized by NMR spectroscopy, those two conformations Ltn10 and Ltn40 not only have entirely different

hydrophobic cores, but also involve mutually exclusive sets of stabilizing interactions, i.e. the same residue in two different conformations favors different secondary structures. The two conformations can be switched to each other by changing temperature and salt concentration and they are nearly populated under physiological conditions with a slow interconversion at the time scale of NMR spectroscopy. At low temperature and high NaCl concentration, Ltn10 is predominant. At high temperature and in the absence of NaCl, Ltn10 is no longer visible. It is hypothesized that the function of this protein is achieved by adopting different native states to realize multiple binding sites. However, a mechanistic understanding of the conversion from state to state is still needed.

In addition to the natural example of multi native states of protein, protein design has been used to investigate how similar the amino acid sequence identities of two proteins can be while keeping the distinct original topologies. Earlier studies show that 50% or more sequence homology can be achieved but it is very difficult to get greater because of protein aggregation (Dalal, & Regan, 2000). More recently, however, a binary system that two proteins had more than 90% sequence homology was devised (Alexander et al, 2007; Alexander et al, 2009). In this system, two binding domains of Streptococcus GA and GB were chosen to study due to their completely different topologies but relatively small size. These two proteins consist of only 56 amino acids and GA has three alpha helix bundle (3α) conformation while GB has an alpha helix packing against a four-stranded beta sheet (α/β) conformation. Figure 1.1(a) shows that almost every amino acid forms different secondary structures between these two proteins. Amino acids 1-8 in 3α conformation are unstructured while those in α/β conformation form the central β1 strand. Amino acids 9-22 form helix α1 in 3α conformation but in α/β conformation form β2 strand, turn between β1 strand and β2 strand and turn between β2 strand and central helix. Amino acid 23 is the end of

helix α1 in 3α conformation and is the start of central helix in α/β conformation. In α/β conformation amino acids 24-36 form the central helix and in 3α conformation amino acids 27-33 form a shorter helix. Amino acids 39-51 form helix α3 in 3α conformation and in α/β conformation form β3 strand, part of β4 strand and the linker between them. Amino acids 52-56 are unstructured in 3α conformation and in α/β conformation form the end of the β4 strand. Besides different topologies, the biological activities of the two proteins are distinct as well which are recognizable under binding affinity test. The GA domain of 48 structured amino acid residues can bind to human serum albumin (HSA) while the GB domain of 56 structured amino acid residues can bind to the constant (Fc) region of IgG.

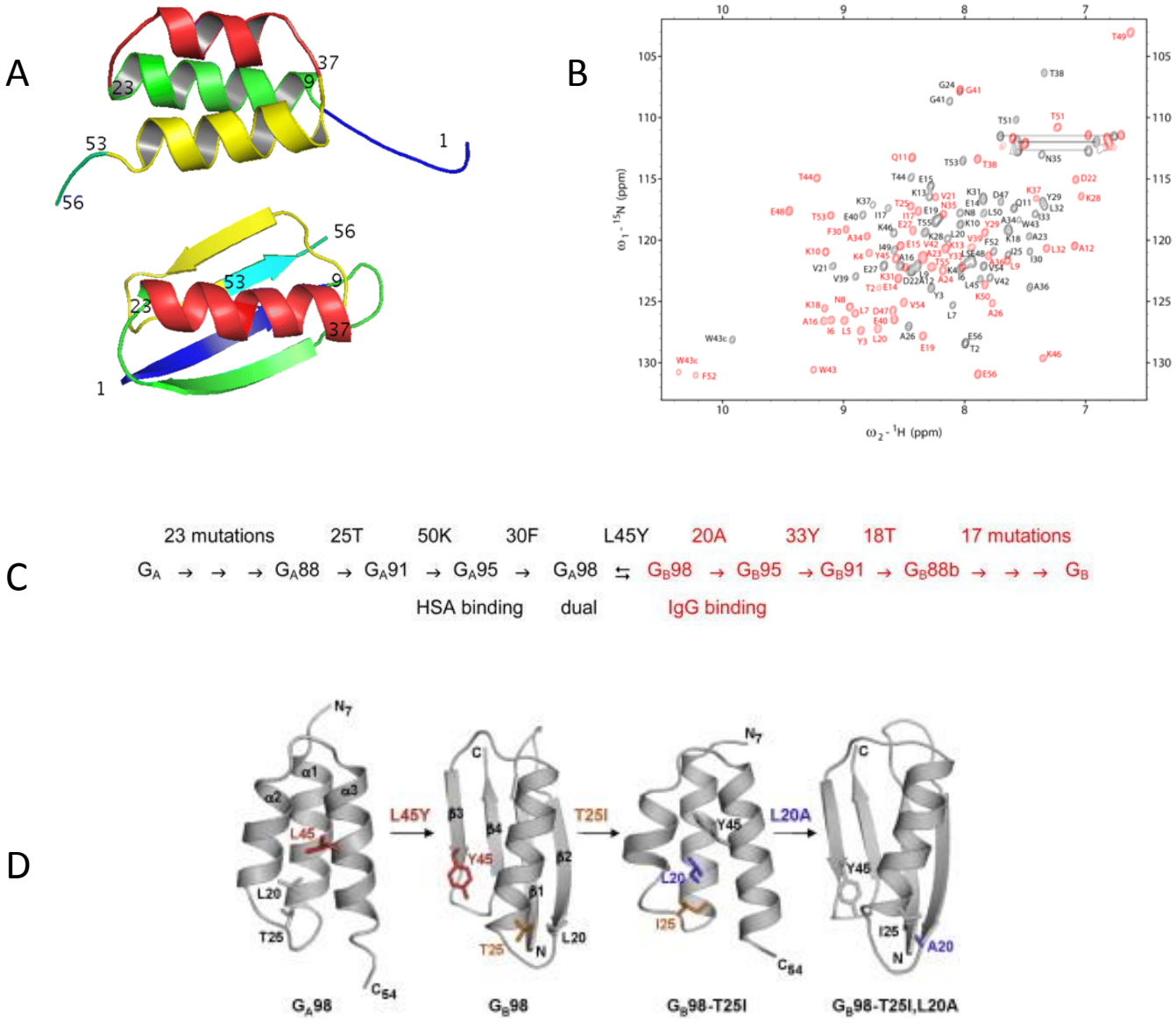

Figure 1.1. Illustrations of GA and GB. (a) Topologies of GA (top) and GB (bottom). (b) ^{15}NHQC spectra of GA88 and GB88 which display distinct patterns to recognize identity. (c) The designing of two mutants with only one amino acid difference but totally different conformations from two unrelated proteins. (d) Other mutational critical point in the same system.

The protein engineering started from a point where two protein sequences GA and GB have only 11 amino-acid residues in common, equivalent to 20% sequence homology. Proteins of GA

and GB that share great sequence homology are sought by introducing mutations based on the sequence of each other and sorting out mutated candidates which can keep the original fold and function. Protein topologies were tested by NMR spectroscopy (Figure 1.1(b) is the overlapped ^{15}N-HSQC NMR spectra of GA88 and GB88, where 88 denotes the percentage of sequence homology) and protein functions are tested by HAS and IgG binding affinity. The mutation route developed by Bryan et al (2010) can finally reach a critical point of 98% sequence homology between two structurally unrelated proteins which is shown in Figure 1.1(c). The 98% sequence homology means that the two different conformations GA98 and GB98 could be manipulated by only one specific amino acid. This amino acid (L45/Y45) plays a crucial role in controlling the protein conformation. Switching L45 to Y45 could destabilize 3α fold and stabilize α/β fold. Alternatively switching Y45 to L45 has the opposite effect. Another surprising phenomenon is that besides the ability to bind to HAS, GA98 can bind to IgG with a comparable binding constant, revealing that the protein of GA98 have a hidden propensity to switch into α/β topology stimulated by IgG (Alexander et al, 2009). The following research also showed that other critical points can exist in this system. Likewise, two other mutants GB98-T25I and GB98-T25I, L20A can respectively hold 3α and α/β fold (He et al, 2012), which are illustrated in Figure 1.1(d), representing two additional critical points.

The ability to switch folds of protein is relatively uncommon and few have been characterized structurally, thermodynamically or kinetically. The design of this simple but representative system can provide an opportunity to examine the mutational requirements for changing folds and functions. Despite the structures and thermal stabilities of this binary conformation system having been measured, other properties that are intrinsically related to the sequence, such as folding and unfolding kinetics, still remain unclear. In this dissertation, optical tweezers were utilized to

investigate some proteins having 3α or α/β fold, shedding light into the mechanical stabilities of these two conformations and extrapolating their folding and unfolding rate constants to zero force which are closely linked to the free energy profile of the protein unfolding.

1.3. Mechanical stabilities of different protein folds

The biological molecules' resistance to unfolding in response to an applied mechanical force is a defined mechanical stability. Mechanical stability is a critically important physiological function in biological systems since many cellular processes involve mechanical work including replication, transcription, translation, cell adhesion, protein and nucleic acid unfolding and protein and nucleic acid splicing (Bustamante et al., 2004). Until the advent of force spectroscopy techniques recently, the only way to measure the stability of a protein was to use some denaturants including heat, pressure and chemical to change the physical or chemical environment of the protein and monitor its loss of conformation. Force spectroscopy techniques, which uses force as a denaturant to perturb the protein conformation, may be very different. Research on comparison between classic ensemble and mechanical denaturation had showed that the mechanical stability does not necessarily correlate with thermodynamic stability. Due to the local action of the force, the unfolding force is related to the local structure rather than the global domain structure (Bustamante et al., 2004). In other words, mechanical stability is strongly correlated to the local secondary structure where resists to unfolding and the pulling direction. Proteins that are predominant an alpha helical structure [Figure 1.2 (c)] are less mechanical stable than the proteins containing beta sheets. Also, if the beta sheet is stretched in opposing directions which forms shear and zipper configurations (Carrion-Vazquez et al., 2000), mechanical stability can be very different.

The shear topology means that the mechanical force is applied on opposite sides of two beta strands and the hydrogen bonds between this two beta strands are perpendicular to the applied force [Figure 1.2(a)]. Unfolding this pattern requires rupturing this cluster of hydrogen bonds simultaneously. Each hydrogen bond contributes to the resistance of the whole construct, so this pattern is expected to have high mechanical stability and accordingly high rupture force. Stretching the two termini of titin I27 domain is of this pattern, demonstrating a rather high rupture force ranged from 150 pN to 300 pN (Rief et al., 1997; Carrion-Vazquez et al., 1999). Another example is GB protein (Cao, Lam, Wang, & Li, 2006), which has the α/β conformation previously mentioned in chapter 1.2. Though this protein is not involved in any mechanical processes in the cell, this protein has high mechanical resistance when stretching its two ends with rupture forces of ~180 pN.

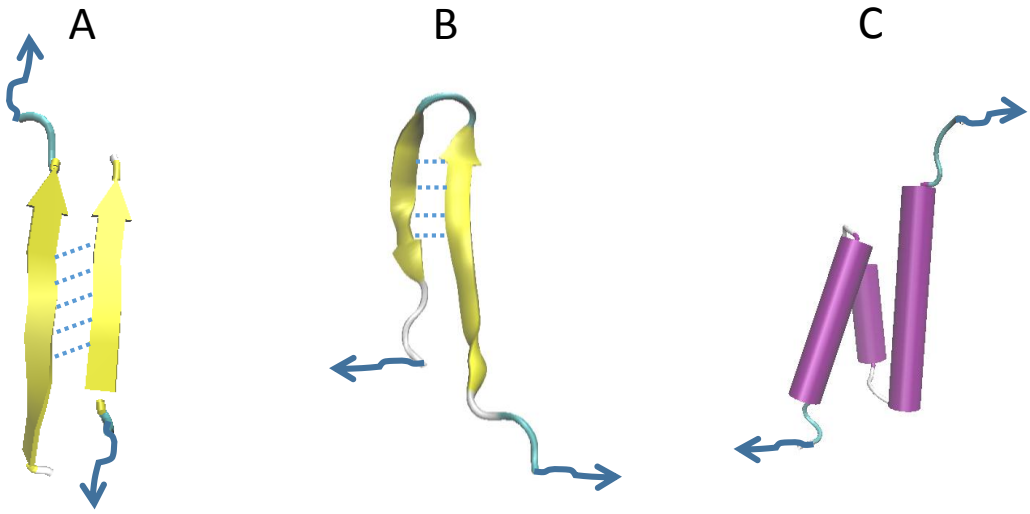

Figure 1.2. Different pulling topology can result in different mechanical stability. (a)shear topology, (b) zipper topology and (c) alpha helix bundle.

In contrast to the shear model, mechanical force is applied on the same side of two strands in zipper topology, so the hydrogen bonds between these two strands are ruptured sequentially. Due

to the fact that only one hydrogen bond contributes to the mechanical resistance at a time, the rupture force of this model is much lower than the sheer model predicted by molecular dynamics simulation. For example, the C2 domain of synaptotagmin I has rupture force of ~60 pN (Carrion-Vazquez et al., 2000).

The predominantly alpha-helical protein has very low mechanical stability compared to proteins that contain beta sheets (Lu, & Schulten, 1999). For example, the rupture force of calmodulin is distributed at 10-15 pN (Junker, Ziegler, & Rief, 2009; Stigler et al, 2011). Spectrin, another protein of alpha helical structure, has rupture forces of about 25-35 pN (Rief et al, 1999). In order to explain the relatively low mechanical stability of alpha helical bundles, Rief et al (1999) have suggested that alpha helical bundles are of hydrophobic interaction which are more delocalized and have associated with a longer distances to their transition state. Since the transition barrier of protein unfolding is tilted by a term of $-Fx$ according to Bell model where F is the mechanical force and x is the distance from native state to transition state, A longer distance requires a lower force to eliminate the same barrier. These terms have clear definitions in two-state model in next chapter. Therefore, proteins having 3α helical bundles, are predicted to have a relatively low mechanical stability and thus low rupture force.

1.4. Optical tweezers and miniTweezers

Over the past two decades single molecule force spectroscopy has been established as a powerful, bulk-complementary and accurate method of characterizing the thermodynamics and kinetics of biomolecules including proteins and DNAs. Thanks to these single molecule force spectroscopic techniques it is possible to observe the mechanical biological processes on a single molecule and measure the instantaneous deviations from the average behavior on bulk level now. Single molecule force spectroscopic techniques, including atomic force spectroscopy (AFM),

optical tweezers and magnetic tweezers, extend the detection limit with accuracy to pico-Newtons force and Angstroms distance. A comparison between optical tweezers and other techniques is shown in table 2.1 (Greenleaf, Woodside, & Block, 2007). The force ranges, displacement sensitivity and probe stiffness vary significantly among them, exhibiting a powerful toolbox for a wide range of biological processes.

Table 1.1. Comparison of single-molecule force spectroscopy techniques

	Optical tweezers	Magnetic tweezers	AFM
Spatial resolution (nm)	0.1-2	2-10	0.5-1
Temporal resolution (s)	10^{-4}	10^{-2}	10^{-3}
Stiffness (pN/nm)	0.005-1	10^{-4}	$10\text{-}10^{5}$
Force range (pN)	0.1-100	$0.01\text{-}10^{4}$	$10\text{-}10^{4}$
Displacement range (nm)	$0.1\text{-}10^{5}$	$5\text{-}10^{4}$	$0.5\text{-}10^{4}$
Probe size (micron)	0.25-5	0.5-5	100-200
Features	Low noise and low drift	Force clamp; Bead rotation	High resolution imaging
Limitations	Photodamage; Sample heating; nonspecific	No manipulation	Large high stiffness probe; Large minimal force; nonspecific

1.4.1. Principles of optical tweezers

Even though the first observation of radiation pressure can date back to 1609 when a German astronomer found that the tails of comets always point away the sun, the first application of radiation pressure using a laser beam to trap and manipulate tiny particles wasn't realized until the 1970s by Arthur Ashkin (1973; 1978; 1986). Later in 1987, the first biological application of radiation trapping was performed on Escherichia coli bacteria, Tobacco mosaic virus and protozoa

(Ashkin, & Dziedzic, 1987). The following development in instrumentation and applications made it possible to explore a broad variety of biological systems, including molecular motors (Svoboda et al., 1993; Finer, Simmons, & Spudich, 1994), unfolding and folding of proteins and nucleic acids (Cecconi et al., 2005; Baumann et al., 2000).

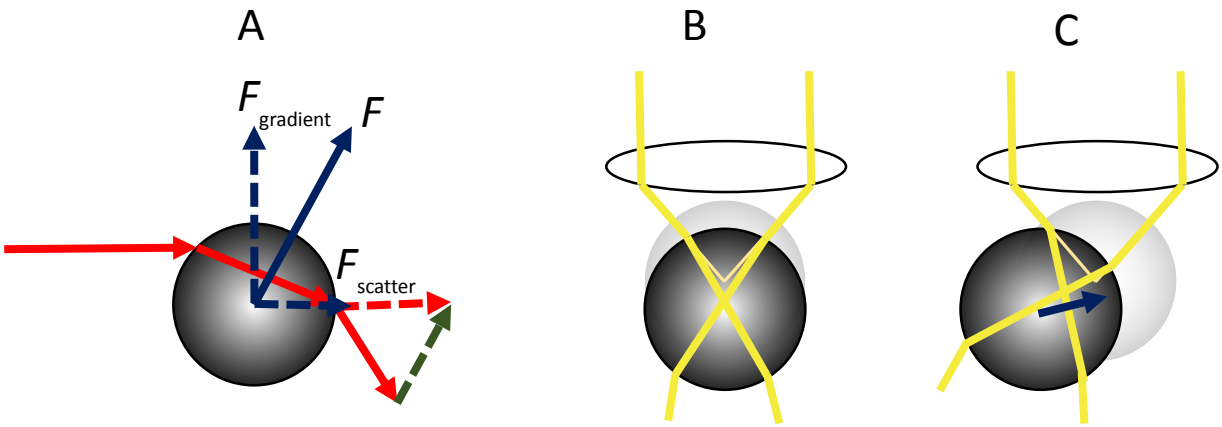

Figure 1.3. Ray-optic diagrams on the principles of optical tweezers. (a) The light (red arrow) propagating through bead could have a momentum change (green arrow) resulting in a force (blue arrow), which can be split into two components: gradient force (blue dash arrow) and scattering force (blue dash arrow). The laser light, shown as two light beams (yellow), is focused by an objective lens and then diffracted by the bead. When the bead is in balance (b), its position is slightly below the focus due to an equilibrium between gradient force and scattering force. For a horizontal displacement (c), the restoring force (blue arrow) points toward the balance point acting like a spring.

The basic principle of optical tweezers can be explained by the ray-optic diagram (Figure 1.3). Due to the conservation of light momentum, the light ray reflected or refracted by an object will transfer the momentum reversely on the object in the opposite direction. In most optical tweezers experiments, polystyrene or silica spheres are trapped by laser beams. Figure 1.3 shows ray-optic diagrams of how a light ray exerts force on a spherical particle. The light passing the particle gets

refracted at two interfaces resulting in momentum transfer to the bead. The generated force can be split into two perpendicular components: the scattering and gradient forces.

The scattering force arises by the photons pushing the particle, so it has the same direction as the light propagating direction.

The gradient force arises from the maximum region of light intensity and pulls the particle towards the region of maximum light intensity.

Near the focus of the laser, the scattering forces exerting on the spherical particle will cancel out in all directions but the direction of the light. If the scattering force and gradient force equilibrate, the particle can be stably trapped by the laser. Figure 1.3(c) shows that when the particle deviates from the equilibrium position by an external force, a restoring force will be exerted on the particle towards the equilibrium position. The restoring force acts as a hookean force $\vec{F}$ with a spring constant k_{h} for small displacement $\vec{x}$ out of the equilibrium position, given by:

$$\overrightarrow{F} = -k_{\mathrm{h}}\vec{x}. \qquad (1.1)$$

The restoring force has an opposite direction against the displacement towards the equilibrium position.

1.4.2. MiniTweezers

The miniTweezers is the latest and most mature generation of optical tweezers instruments (Figure 1.4), which is exclusively designed for high accuracy and precision force spectroscopy. One of the biggest advantages of this setup is that two counter laser beams are used to form a single trap so that the scattering forces along the light directions cancel out each other but the gradient

force is increased with a greater light intensity. A stiff trap up to 200 pN can be easily achieved with only 5 Watt lasers. Its trapping force can be calibrated by conservation of light momentum, Stokes' law or the overstretching state of DNA (Smith, CUI, & Bustamante, 2002; Smith, 2009).

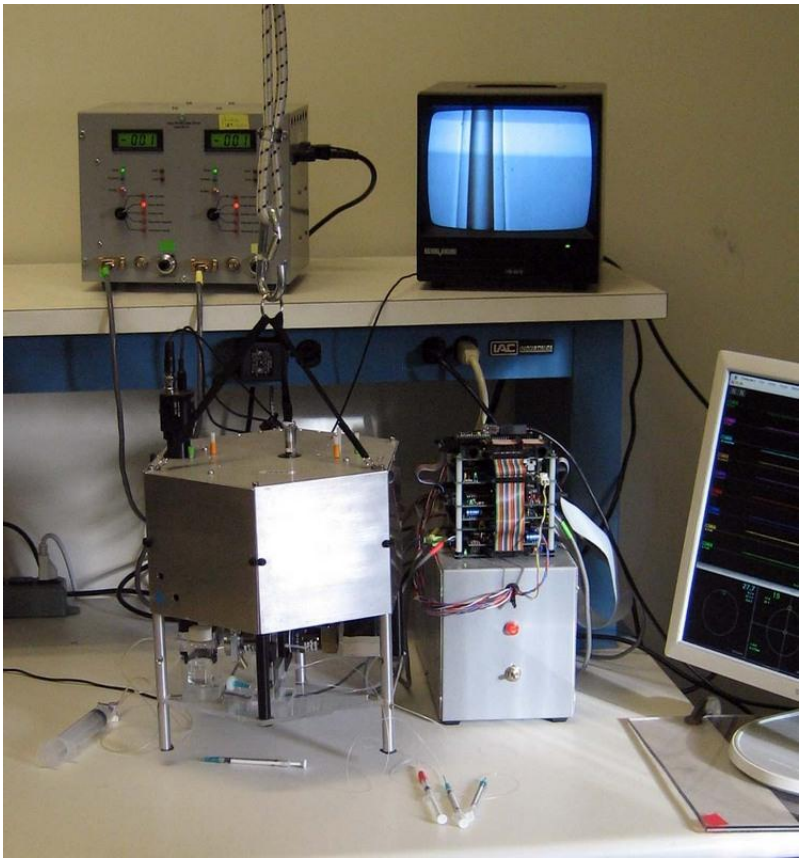

Figure 1.4. A picture of the miniTweezers.

Pipette-tethered dumbbell assay is used in miniTweezers. As shown in Figure 1.5, one bead is tethered to a micro glass pipette tip with a caliber smaller than 1 micron and the other bead is trapped by two counter laser beam in this assay. These polystyrene or silica beads coated with streptavidin or anti-digoxigenin are commercially available. Functionalized DNA molecules that have thiol group and biotin/digoxigenin moiety are synthesized using polymerase chain reaction (PCR). The target protein can be coupled to the DNA molecules by thiol chemistry (see details in

2.2). In this chimera, the DNA molecules, namely “molecular handles”, function as spacers to prevent bead-bead nonspecific interaction. In order to couple the protein between two DNA handles, two cysteine residues must be engineered on the positions where the external force is applied on the molecule. In addition, other solvent-accessible cysteine residues must be mutated to avoid unwanted coupling reactions. The disadvantage of introducing DNA handles is that there is an overstretching state of DNA molecule at ~65 pN which limits the maximum force.

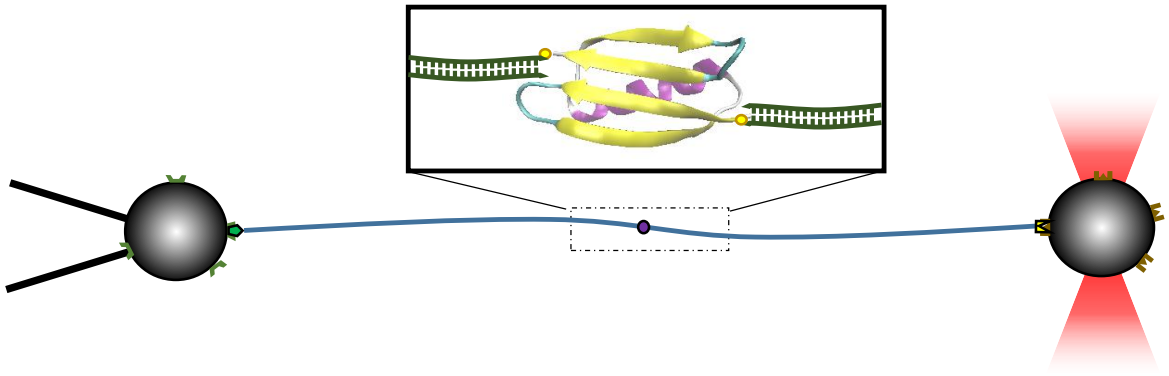

Figure 1.5. Sketch of the pipette-tethered dumbbell assay for MiniTweezers. The beads (dark) are functionalized with Strepatvidin (orange) and anti-digoxigenin (brown). The streptavidin bead is fixed by the suction of the pipette tip while the anti-digoxigenin is trapped by the laser. The 802 bp DNA handles (blue) containing biotin (green) and digoxigenin (black) are connected via disulfide bond, which is shown in inset.

1.4.3. Two common force operation modes of miniTweezers

The constant pulling velocity mode can give an overall response how the protein reacts to force. The schematic of constant velocity mode is shown in Figure 1.6. In this mode, the two beads are brought close enough to form the dumbbell assay as mentioned. Then the trapped bead is pulled apart by moving the optical trap at a constant rate ranging from 1 nm/s to 100 nm/s until reaching the set position and simultaneously a restoring force of the trap can be exerted on the protein-DNA construct along the pulling direction. The force will be transmitted to the protein and stretch the

protein along the direction of the two cysteine residues connected to DNA handles. When approaching to the DNA contour length, the force rises high enough and triggers the unfolding of the protein, resulting in a sudden relaxation in force caused by an increase in the contour length of the molecule. If the trapped bead moves back towards the tethered bead, the force on the protein will be relaxed. Then the protein has a chance to refold against the force. The refolding of protein can be indicated by a decrease in contour length and a sudden increase of force.

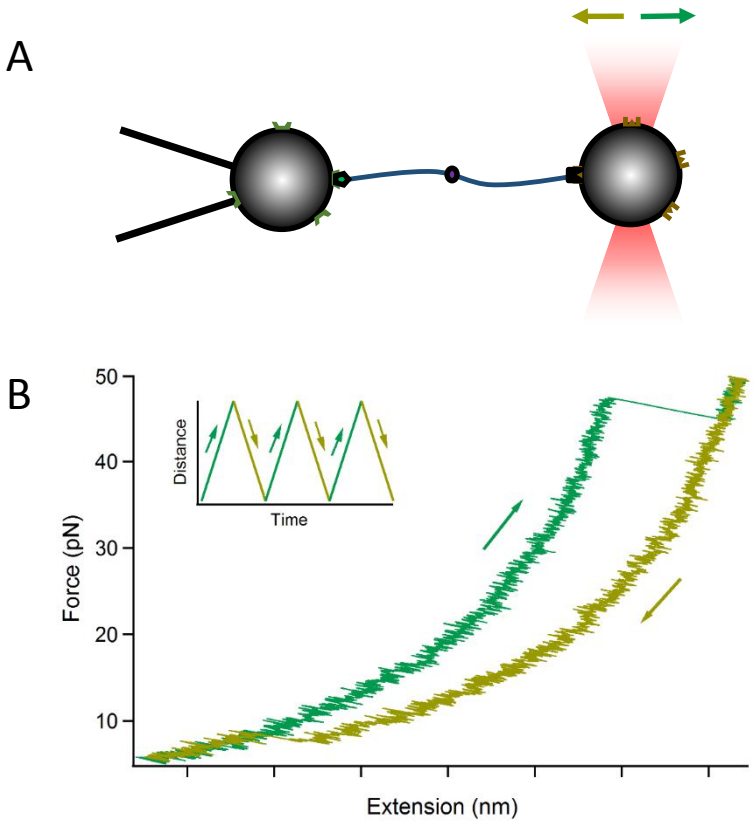

Figure 1.6. Schematic of constant velocity protocol. (a) The trap is moved at a constant velocity in the constant velocity protocol. Moving outwards (green) and backwards (brown) is a cycle of stretching and relaxation. Figure (b) is the force extension curves of a typical stretching and relaxation cycle. The protein may unfold once the force is high enough and it may fold back during the relaxation.

The constant force mode is generally performed to assess how long the protein needs to fold or unfold under a specific force. This mode provides a more straightforward profile on the protein kinetics. In this scheme, the laser trap is controlled by a feedback loop to ensure the constant force condition. Once the protein transits to another state, the contour length of the construct may have a sudden change resulting in a force change. Immediately the feedback loop adjusts the trap position to recover the set force within several milliseconds depending on the feedback response time. As Figure 1.7, the trap position has distinct levels indicating different states of the protein. The level that has the greatest separation between the trapped bead and the tethered bead represents the denatured state and the level that has least separation represents the native state. Due to the stochastic nature of protein folding and unfolding, the distribution of dwell time at each state should follow the Markov model. In the two state model of protein folding, proteins only have native state and denatured state without any intermediate state. In this case, due to the stochastic nature of protein folding and rupture, their probabilities *P* can be described by a force-dependent rate constant *k* as first-order reaction, so the dwell time distribution at each state is exponential, given by:

$$P = \exp(-kt), \tag{1.2}$$

where *t* denotes the dwell time under the stretching force.

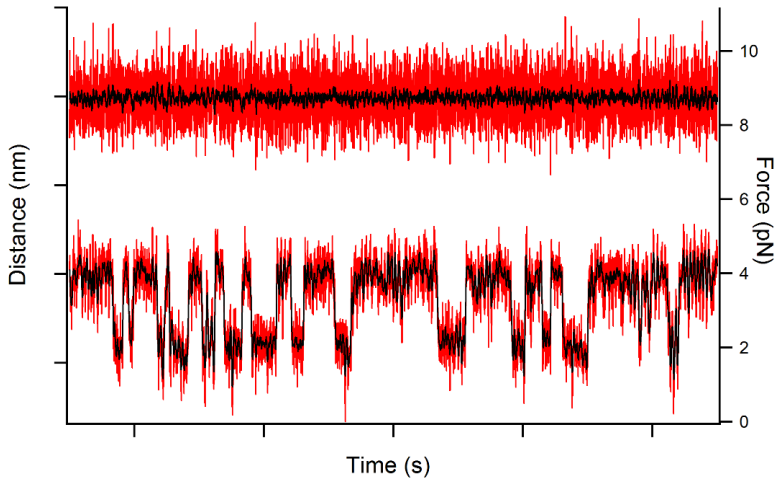

Figure 1.7. Example of constant force protocol results. The top curve is the force over this period and the bottom curve is the displacement of the bead indicating the status of the protein. Longer displacement represents the unfolded protein and shorter displacement represent native protein.

In fact, dwell time analysis can be also used to extract unfolding/folding rate constant *k* from constant velocity data. This scheme was first devised by Oberbarnscheidt et al (2009). In this scheme, a force-extension curve can be split into many small time windows, each of which is short and may be regarded as a single constant force experiment. Since these pulling force curves are completely independent of each other, the data within time windows of the same force can be merged and be described as one constant force experiment.

1.5. Models for force spectroscopy

1.5.1. Force extension models

The force-extension behavior of proteins and DNA can be well described by some polymer models, e.g. Worm-like Chain (WLC) model. The WLC model assumes that the molecule is a

polymer chain and it describes the entropic force behavior of the unstructured protein well. An interpolation formula can be used to approximate the force-extension relationship (Bustamante et al., 1994):

$$F_{\mathrm{WLC}}(x) = \frac{k_{\mathrm{B}}T}{p}\left(\frac{1}{4\left(1-\frac{x}{L}\right)^2} - \frac{1}{4} + \frac{x}{L}\right) \tag{1.3}$$

with persistence length *p*, contour length *L* and extension *x*. Contour length is a feature of the polymer defined by the physical maximum extension. In WLC model, contour length is considered to be constant but the extensibility of contour length cannot be negated for DNA. Extensible WLC (eWLC) model is a better model to describe DNA's force-extension relationship (Wang et al., 1997). An elastic modulus *K* is introduced to account for the extension of contour length:

$$F_{\mathrm{eWLC}}(x) = \frac{k_{\mathrm{B}}T}{p}\left(\frac{1}{4\left(1-\frac{x}{L}-\frac{F}{K}\right)^2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K}\right). \tag{1.4}$$

Since the globular structure of the protein is much stiffer in persistent length and shorter in contour length than the peptide chain, the native state protein can be treated as an integrative rod and its extension can be related to the external force by the monomeric force-extension relationship in the freely-jointed chain (FJC) model (Yao, Chen, & Yan, 2015). Each globular protein domain serves as an independent rigid segment and the force-extension relationship is given by:

$$\frac{x}{L} = \left(\coth\frac{FL}{k_{\mathrm{B}}T} - \frac{k_{\mathrm{B}}T}{FL}\right)\left(1 + \frac{F}{F_0}\right) \tag{1.5}$$

with extension *x*, contour length *L* and external force *F*. The contour length *L* is the distance between two stretched cysteines $L_{i,j}$ of the structured protein which can be obtained from the crystal

structure of the protein. The factor of $(1+F/F_0)$ accounts for the segment extensibility under mechanical force, though the mechanical deformability of protein domain is negligible.

The extension of the whole construct can be described by the combination of the extension of DNA and protein, given by:

$$x_{\text{protein-DNA}}(F) = x_{\text{eWLC}}(F) + x_{\text{protein}}(F) \tag{1.6}$$

with:

$$x_{\text{protein}}(F) = \begin{cases} x_{\text{WLC,DS}}(F), & \text{denatured state,} \\ x_{\text{FJC,NS}}(F), & \text{native state.} \end{cases} \tag{1.7}$$

The theoretical contour length increment between amino acid *i* and *j* for a specific protein is given by:

$$d_{i,j} = (j-i)d_{\text{aa}} - L_{i,j} \tag{1.8}$$

with each unstructured amino acid contribution d_{aa} = 0.36 nm (Dietz, & Rief, 2006).

In order to determine the contour length increment of protein unfolding, each force-extension curve is split into different parts according to the state of the protein. The part containing native-state protein can be fitted to the eWLC model. These fitted parameters remain unchanged and be applied to a WLC curve to account for the additional contour length of denatured protein. However, an alternative method which can extract the contour length increment of protein unfolding without any influence of DNA's extension can take advance of the force extension data near the unfolding event. According to formulae 1.7, the extension of DNA can be cancelled out between native state and denatured state as long as the pulling force is the same. As a result, the difference of protein extension between native state and denatured state at a specific force is given by:

$$\Delta x_{\text{protein}} = x_{\text{WLC,DS}}(F) - x_{\text{FJC,NS}}(F). \quad (1.9)$$

After compensating the extension of native state protein, these extension values can fit to WLC model with the persistence length and contour length of the protein *per se*.

1.5.2. Two-state kinetic model and force dependent transition rate constants

Many proteins fold in a highly cooperative fashion which is well described by the two-state kinetic model. Thus, this model is often used to describe the protein unfolding and folding introducing the effect of force. In this model, the inter-conversion between native state and denatured state need to surmount a high energy barrier which is called transition state (TS). The unfolding rate constant for transiting from native state to denatured state and the folding rate constant for transiting from denatured state to native state are given by:

$$k = A\exp\left(\frac{-\Delta G_{i,\text{TS}}}{k_{\text{B}}T}\right), \quad (1.10)$$

where *i* denotes the initial state, such as native state and denatured state, and the respective energy barrier is $\Delta G_{i,\text{TS}}$. Figure 1.8 shows the free energy profile of a two-state model. In the absence of force proteins favor the native state because of the lower free energy. If force is exerted on the protein the free energy will be tilted by the external force. The model developed by Bell (1978), which well describes the dissociation kinetics in ligand-receptor systems under force, can also describe the protein unfolding and folding under force. Force is defined as the reaction coordinate and tilt free energy by a factor of Fx where x denotes the distance between transition state and initial state. The unfolding and folding rate constants depending on force are given by:

$$k_{\text{u}}(F) = A\exp\left(\frac{-\left(\Delta G^{0}_{\text{NS,TS}} - Fx_{\text{u}}\right)}{k_{\text{B}}T}\right) = k^{0}_{\text{u}}\exp\left(\frac{Fx_{\text{u}}}{k_{\text{B}}T}\right), \quad (1.11)$$

$$k_{\mathrm{f}}(F) = A\exp\left(\frac{-(\Delta G^{0}_{\mathrm{DS,TS}} + Fx_{\mathrm{f}})}{k_{\mathrm{B}}T}\right) = k_{\mathrm{f}}^{0}\exp\left(\frac{-Fx_{\mathrm{f}}}{k_{\mathrm{B}}T}\right), \quad (1.12)$$

where unfolding distance x_{u} denotes the distance between transition state and native state, folding distance x_{f} denotes the distance between transition state and denatured state and k_{u}^{0} and k_{f}^{0} are the unfolding and folding rate constants under zero force.

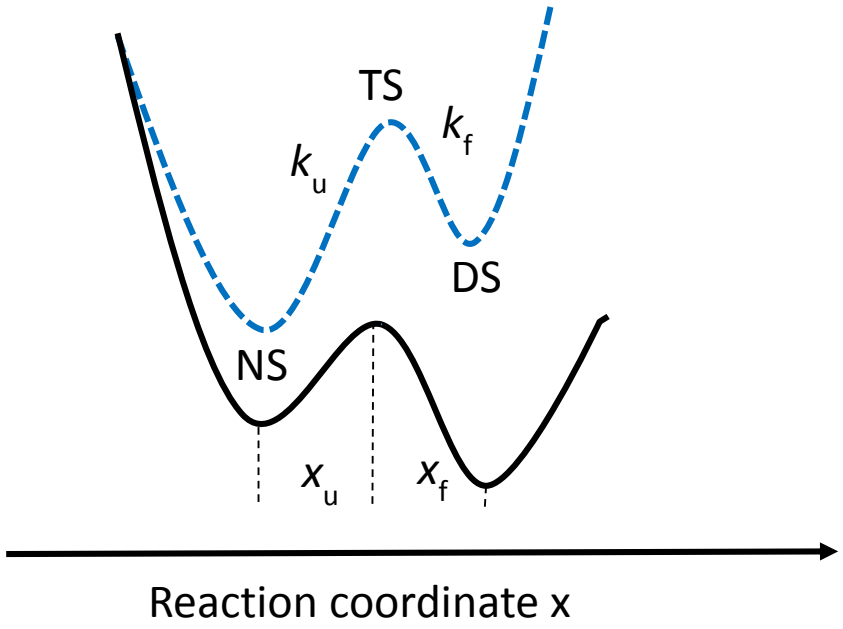

Figure 1.8. The effect of force on the free energy profile of a two-state system. The dash curve denotes no applied force and solid curve denotes positive applied force. The application of force lowers the free energy of both the transition state and denatured state relative to native state according to the reaction coordinate, which increases the unfolding rate and the population of denatured state. Since the free energy profile is curved, the positions of the free energy minima (NS and DS) and maximum (TS) shift a bit to longer x_{f} and shorter x_{u} respectively, with a positive applied force. Their relative shifts in position depend on the local curvature of the free energy surface.

However, Bell model neglects the potential force dependence of *x* and *k* especially if the free energy profile of protein unfolding is curved. Besides, the nonlinear effect of the DNA linker and unstructured protein chain are negated. A different model has been developed by Schlierf (2007) and adapted to optical tweezers system by Gebhardt (2010). This model accounts for the additional

energy of the energy barrier due to the energy changes of DNA linkers and spring. The free energy of DNA linker, unstructured protein and the bead in state *i* are given by:

$$G_{i,\mathrm{bead}}(F_i) = {F_i x_i}/{2}, \tag{1.13}$$

$$G_{i,\mathrm{DNA}}(F_i) = \int_0^{x_{DNA}(F_i)} F_{\mathrm{eWLC}}(x_{\mathrm{DNA}})\mathrm{d}x_{\mathrm{DNA}}, \tag{1.14}$$

$$G_{i,\mathrm{protein}}(F_i) = \int_0^{x_{\mathrm{protein}}(F_i)} F_{\mathrm{protein}}\left(x_{\mathrm{protein}}\right)\mathrm{d}x_{\mathrm{protein}}. \tag{1.15}$$

The free energy of protein switch from state *i* to state *j* is given by:

$$\Delta G_{i,j}\left(F_i, F_j\right) = \Delta G_{i,j}^0 + \Delta G_{i,j,\mathrm{bead}}\left(F_i, F_j\right) + \Delta G_{i,j,\mathrm{DNA}}\left(F_i, F_j\right) + \Delta G_{i,j,\mathrm{protein}}\left(F_i, F_j\right). \tag{1.16}$$

Therefore, the rate constant is given by:

$$k_{i,j} = A\exp\left(-\frac{\Delta G_{i,j}^0 + \Delta G_{i,j,\mathrm{bead}}\left(F_i, F_j\right) + \Delta G_{i,j,\mathrm{DNA}}\left(F_i, F_j\right) + \Delta G_{i,j,\mathrm{protein}}\left(F_i, F_j\right)}{k_{\mathrm{B}}T}\right)$$

$$= k_{i,j}^0\exp\left(-\frac{\Delta G_{i,j,\mathrm{bead}}\left(F_i,F_j\right)+\Delta G_{i,j,\mathrm{DNA}}\left(F_i,F_j\right)+\Delta G_{i,j,\mathrm{protein}}\left(F_i,F_j\right)}{k_{\mathrm{B}}T}\right). \tag{1.17}$$

1.5.3. Thermodynamic free energy difference between native state and denatured state

Thermodynamic free energy difference between native state and denatured state can be measured by bulk denaturation methods including thermal and chemical denaturation. It is also of great significance to extract this information from the force spectroscopy measurement. If the unfolding and folding rate constants under zero force are known, the free energy difference $\Delta G_{\mathrm{NS-DS}}^0$ can be calculated by:

$$\Delta G_{\mathrm{NS-DS}}^0 = k_{\mathrm{B}}T\ln\left(\frac{k_{\mathrm{f}}^0}{k_{\mathrm{u}}^0}\right). \tag{1.18}$$

Using fluctuation theorems such as Crooks fluctuation theorem (CFT) can relate the work along the transition trajectories of the protein to the free energy difference between states. CFT has been previously used in optical tweezers studies to determine the free energy difference between native state and denatured state of proteins. It quantifies the work *w* done to fold or unfold protein and associate the work to the free energy difference $\Delta G^0_{\mathrm{NS-DS}}$ by:

$$\frac{P_{\mathrm{U}}(w)}{P_{\mathrm{F}}(w)} = \exp\left(\frac{w-\Delta G^0_{\mathrm{NS-DS}}}{k_{\mathrm{B}}T}\right), \tag{1.19}$$

where $P_{\mathrm{U}}(w)$ and $P_{\mathrm{F}}(w)$ defines the protein unfolding and folding distribution associated with the applied work. It is of note that $\Delta G^0_{\mathrm{NS-DS}} = w$ once $P_{\mathrm{U}}(w) = P_{\mathrm{F}}(w)$.

2. Materials and methods

2.1. Protein expression and purification

The gene of a given protein was engineered into a modified PQE80L expression vector, which contains a cysteine codon at each terminus of the target gene. Then the protein was overexpressed in Escherichia coli strain DH5a and purified using Co^{2+} affinity chromatography with TALON His-Tag purification kit. After the purification, the protein was kept in a phosphate-buffered saline (PBS, 300 mM NaCl, pH 7.4, 150 mM imidazole) at 4 ^{0}C.

2.2. Cross-linking reaction between protein and DNA

2.2.1. Principle

In order to obtain the ability to bind to two differently functionally coated polystyrene/silica beads (generally one bead is coated with anti-digoxigenin antibodies and the other is coated with streptavidin), two modified ds-DNA are attached to the protein molecules prior to forming the dumbbell assay. One handle contained a 5'-thiol group and a 5'-biotin moiety and the other contained a 5'-thiol group and a 5'-digoxigenin moiety. This reaction can be facilitated by 2,2'-dithioldipyridine (DTDP) activation (Cecconi et al., 2008; Zhang et al., 2009).

After DNA handles are activated by DTDP, the attachment of the two handles can be carried out via a sequential strategy. First, five to ten fold of protein is added to one of the activated handles at room temperature for 2-4 hours. Then, the construct of protein with one handle can be separated by size-exclusive chromatography. Next, five to ten fold of the other handle is added under room temperature. The attachment of the second handle is typically much slower than the first handle

due to electrostatic repulsion, and generally takes more than one day. This sequential route can maximize the likelihood of the construct of protein with two different handles.

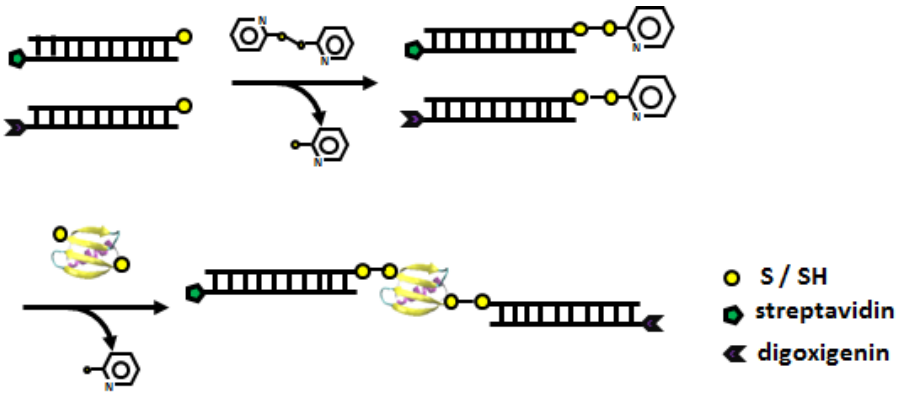

Figure 2.1. The scheme of one-step protein DNA coupling. In this scheme, the DNA molecules are activated by DTDP and then reacted with protein molecule. The protein sandwiched by DNA handles can be formed ultimately.

The alternative way which is the predominant attachment method in this dissertation is a simpler, less time-consuming one-step reaction (Figure 2.1). In this case, an 1 to 4 fold excess of protein is reacted with an 1:1 mixture of the two DTDP activated DNA handles. The final result is a mixture of unreacted molecules, protein with one handle, protein with two same handles and protein with two different handles. The yield of the required configuration is less than 50% but is sufficient for optical tweezers experimentation. This method does not require further purification, as only correctly labelled proteins can function in the dumbbell assay. Other species present in the sample are unable to bind to two differently functionalized beads simultaneously.

2.2.2. Experimental conditions

The protein solution was added with Dithiothreitol (DTT) was to 1 mM to reduce the protein-protein disulfide bonds for 1 hour before the protein-DNA coupling reaction. The removal of DTT and buffer exchange were achieved by passing the solution twice through two acetate buffer (200 mM sodium acetate, pH 5.6) saturated desalting columns (Zeba, 7KMW). These two desalting columns are saturated by passing through acetate buffer 4 times. The eluted protein solution is immediately mixed by ~2 μM DTDP-activated 802 bp DNA handle mixture (equal concentration of biotin modified handle and digoxigenin modified handle; handle is activated by ~30 times moles of DTDP) with an 1:1 molar ratio. The molar concentration of protein is two times higher than the molar concentration of each DNA handle in order to optimize the coupling efficiency. The reaction mixture was incubated under room temperature for 14-20 hours and afterwards kept at -80 ^{0}C for long term storage.

2.3. Optical tweezers experiment conditions

An aliquot of the DNA-coupled protein sample was diluted into ~5 nM protein concentration prior to optical tweezers experiments. 1 μl diluted sample was mixed with 2 μl 0.5% wv 2 micron anti-Digoxigenin bead suspension for 15 minutes and then diluted with 3 ml Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.5). 1 μl 1% wv 1 micron Streptavidin bead suspension was diluted with 3 ml Tris buffer directly. These two diluted bead suspensions were respectively injected into the two side channels of the miniTweezers chamber (Figure 2.2). The laser beams can trap and bring two different beads sequentially to the vicinity of the pipette tip to build up the dumbbell assay.

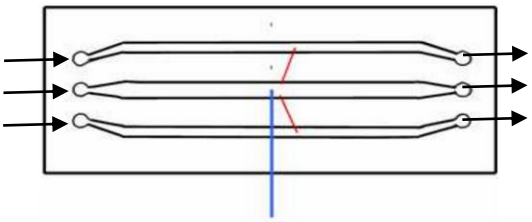

Figure 2.2. The schematic of miniTweezers chamber. There are three channels in this chamber. The laser and dumbbell assay are located at the middle channel. The blue line denotes the pipette tip which can fix a bead by suction. Two kinds of beads are injected into the side channels. Beads can diffuse into the middle channel through a bypass (red line). The black arrows denote the fluidic direction.

2.4. Monte Carlo simulation

Monte Carlo simulation is widely used to reproduce the unfolding and folding events in constant-velocity experiments of AFM and optical tweezers. The kinetic properties of protein folding and unfolding can be estimated by reproducing experimental force distributions computationally. Even if the kinetic properties are known, Monte Carlo simulation is still very useful for result verification. In this scheme, the protein-DNA construct is extended by a spring with a spring constant K at a constant velocity v from zero extension ($x = 0$, $t = 0$). At each discrete time step (Δt), the spring position vt consists of spring deformation F/K, DNA extension $x_{\mathrm{eWLC}}(F)$ and protein extension $x_{\mathrm{FJC,NS}}(F)$ or $x_{\mathrm{WLC,DS}}(F)$. The corresponding force F can be searched by Bisection method to obtain a satisfying approximate. The probability of observing an unfolding or folding event is $P = k(F)\Delta t$, where $k(F)$ is calculated by Bell model or Schlierf model mentioned in chapter 2.3.2. Then this probability value is compared to a random number uniformly distributing on (0, 1). If P is greater than the random number, the transition event happens at the corresponding time step. This scheme continues until reaching the set force then the spring starts

to move back to zero extension in the opposite direction. If this cycle of stretching and relaxing is repeated the statistics of rupture force and folding force can be collected on these simulations.

3. Mechanical unfolding and folding a protein of α/β conformation

3.1. Synopsis

Streptococcus GB protein, namely the B1 binding domain of protein G, is a wild type protein in the binary conformation system as mentioned in chapter 1.2. GB holds the α/β conformation with an alpha helix packing against a four-stranded beta sheet and it has been reported to possess high mechanical stability with a rupture force ~180 pN at a pulling speed of 400 nm/s using AFM. Even though the pulling speed of optical tweezers is much lower as mentioned in chapter 2.1.4, which resulted in a relatively lower rupture force. Accordingly, this protein is still too mechanically stable to unfold within the force upper limit of 65 pN. Therefore, a homologous protein NuG2 is selected instead. NuG2, holding the same α/β conformation as GB, is a de novo designed variant of the protein GB in which 11 residues in the first beta hairpin were mutated to increase the folding rate. This redesigned protein can fold ~100 times faster than its wild type protein. In terms of the mechanical stability of NuG2, this protein was also well documented by previous AFM studies and its rupture forces distributed around 105 pN at the 400 nm/s pulling speed which were significantly lower than GB. Thus, NuG2 is a rational replacement for investigating α/β proteins.

In this chapter, NuG2 monomer and tetramer were studied by the optical tweezers. The polyprotein (NuG2)4 was previously constructed by Cao and the original gene of NuG2 is a generous gift from Dr. Baker. The monomer was studied with constant velocity mode while the tetramer was studied with constant velocity mode along with constant force mode. All these

experiment closely agreed on the same values of kinetics, confirming the free energy profile of protein unfolding of NuG2.

3.2. Results and discussions

To investigate the mechanical unfolding and folding of NuG2, we stretched NuG2 monomer from its two termini using the constant velocity protocol. Stretching the DNA-coupled NuG2 molecule at single molecule level resulted in force-extension curves. Force-extension curves of a typical constant velocity cycle is shown in Figure 3.1. In these trajectories NuG2 unfolded at ~30 pN and folded at ~8 pN. To further ascertain if the globular protein is unfolded completely, the contour length increment upon unfolding is useful and serves as an important characteristic for a specific protein. Here, based on the state of the protein, we fitted the native state part of force-extension curve to eWLC model with best-fit parameters of persistence length p = 13 nm, contour length L = 545 nm and elastic modulus K = 500 pN, comparable to values reported by other groups (Gebhardt, Bornschlögl, & Rief, 2010). Also, this contour length of DNA was equivalent to the length of ~1600 bp, namely the length of two DNA handles. Then, the protein's contour length increment was estimated to be 17 nm by fitting the denatured state part of the force-extension curve, assuming that the persistence length of unstructured protein is 0.7 nm. This contour length increment value was in good agreement with the theoretical and experimental values for the complete unfolding of its homological protein GB. Thus, this contour length increment and the fact that no intermediate state was observed indicated that the folding and unfolding of NuG2 were in an all-or-none fashion.

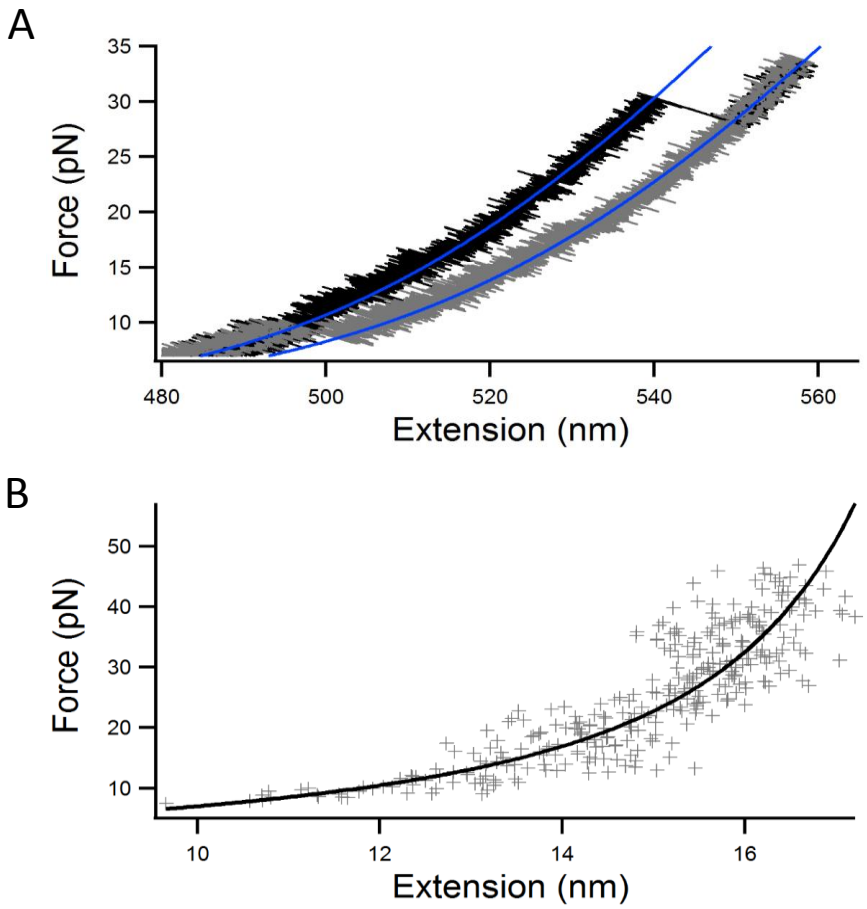

Figure 3.1. Contour length increment fitting for NuG2. (a) eWLC chain fitting (blue curves) for the molecule force-extension curve of a typical stretching (black plot) and relaxation (grey plot) with parameters: persistence length p_{DNA}=13 nm and $p_{protein}$=0.7 nm; contour length L=545 nm and increment of protein ΔL=17 nm; modulus of DNA K=500. (b) WLC fitting (black curve) for the extension of unstructured protein (grey plots) extracted from the extension difference between native state and denatured state. The fitting parameters are $p_{protein}$=0.68 and L=20.7 nm.

The contour length increment was also evaluated by analyzing the extension difference between native state and denatured state. The extension differences in relation to force from 20 constant velocity cycles were shown in Figure 3.1. To recover the extension of unstructured protein, the extension of structured protein should be considered and its value was calculated using FJC

model based on its corresponding force and NuG2's folded length (the distance between its two termini in crystal structure) of 2.6 nm. The WLC fitting on these force-extension relationships estimated the persistence length to be 0.68 nm and the contour length to be 20.7 nm. These results also supported the previous conclusion.

The force-extension trajectories suggested that the mechanical unfolding and folding under our experimental conditions is a non-equilibrium process. The typical rupture force and folding force were distinct indicating that energy dissipated during the stretching and relaxation. Although the protein folding and unfolding of NuG2 is non-equilibrium the unfolding of the protein tend to equilibrate with its folding if the pulling speed is low enough, resulting in close values of rupture force and folding force. Hence, the rupture and folding forces were measured at different pulling speeds including 2, 5, 10, 20, 50 and 100 nm/s, shown as distributions in Figure 3.2(a). As the pulling speed decreased, the rupture force and folding force distributions became close and even overlapped. It was of note that the histogram for 100 nm/s sharply dropped to nearly 0 at ~60 pN seemingly truncated by the force detection limit using DNA handles (chapter 2.2, ~65 pN). The dependency of force on the pulling velocity was also included in Figure 3.2(b). Referring to the previous AFM study on NuG2 in our lab, the pulling speed dependence of rupture force of NuG2 could have two regions: low speed region and high speed region. The slope of the dependency at low speed region was much flatter than the one at high speed region since the unfolding and folding was close to equilibrium at slow pulling velocity. Although the range of pulling speed was not wide here, the dependency obviously became flatter as the pulling speed decreased. Therefore both the force distributions and the pulling speed dependence suggested that protein folding and unfolding could be brought to equilibrium by slowing down the pulling speed.

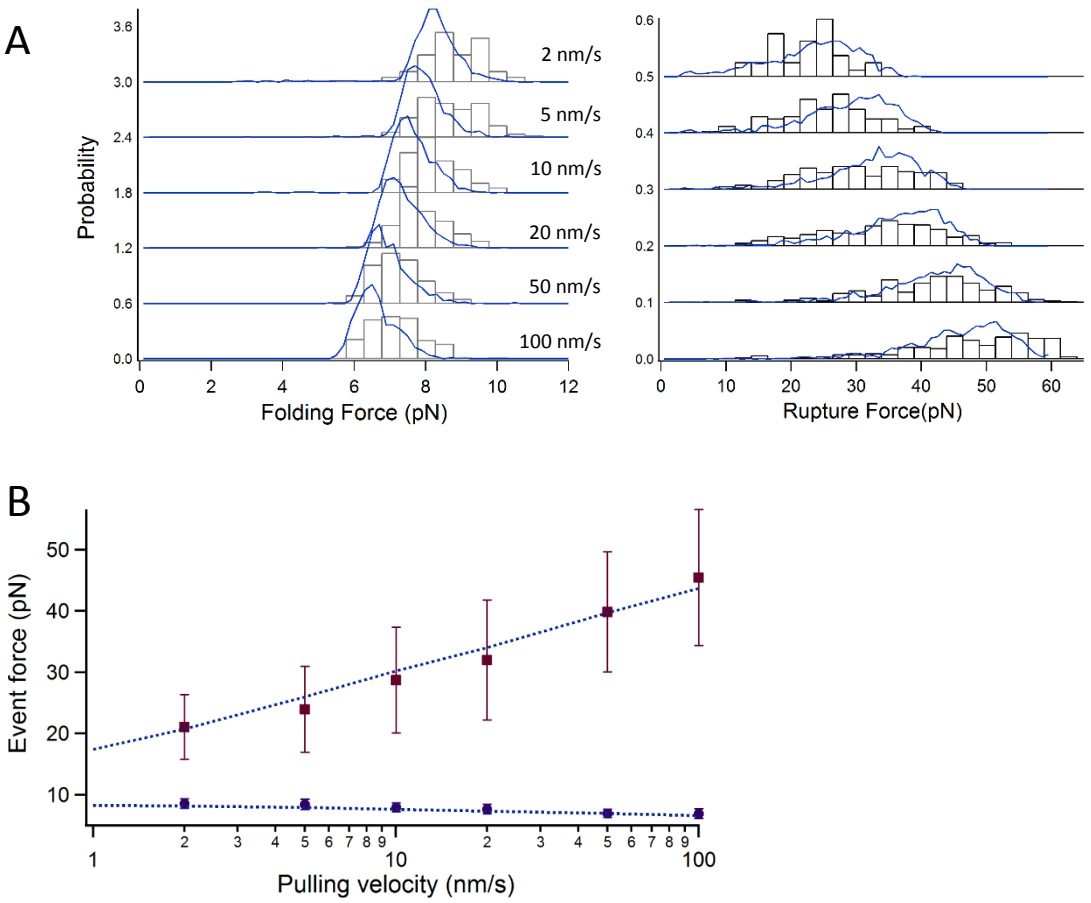

Figure 3.2. Unfolding and folding force distributions of NuG2 and the dependency of forces on pulling velocity. (a) Folding force and rupture force distributions for NuG2. From top to bottom, the pulling speed are 2, 5, 10, 20, 50 and 100 nm/s. It is obvious that the folding force and rupture force tend to get closer as pulling speed decreases. (b) The dependency of forces on pulling velocity. For both Figure (a) and (b), the blue curves are the results from Monte-Carlo simulation using parameters k_u^0=0.0015 s^{-1}, k_f^0=28000 s^{-1} and *x*u=0.7nm.

The value of $\Delta G^0_{\mathrm{NS-DS}}$ was estimated using fluctuation theorems, which could relate the nonequilibrium work done by an external perturbation to the equilibrium free energy difference between states. The Crooks fluctuation theorem (CFT) has been verified to apply in the folding and unfolding of RNA and proteins carried out by AFM and optical tweezers (Crooks,1999; Collin et al., 2005; He et al., 2015). The irreversible work done to unfold and fold a protein is equal to the free energy difference upon the transition, which was obtained by numerical integration of

WLC and eWLC equations with the values of persistence lengths and contour lengths of protein and DNA from previous fittings. As shown in Figure 3.3(a), the probability distributions of folding and unfolding associated with the work were intersected and defined the value of $\Delta G^0_{\mathrm{NS-DS}}$ is equal to ~16.5$k_{\mathrm{B}}T$. This value of $\Delta G^0_{\mathrm{NS-DS}}$ was close to the previous AFM studies of NuG2 (He et al., 2015).

In order to estimate the key parameters for the free energy profile of protein unfolding, the force-dependent rate constants for protein folding and unfolding need to be obtained and the relationship should be extrapolated to zero force. The Oberbarnscheidt method provided a chance to extract rate constants indirectly from the constant velocity data. As mentioned in chapter 2.1.4, each force extension trajectory could be considered as many small constant force experiments each of which was restricted in a small window. Though the length for each force window from a single trajectory is small, a large number of independent force-extension trajectories, generally >100, was sufficient to give accurate estimates for the rate constants. This method was applied to force-extension data of NuG2 and the rate constants depending on force were shown in Figure 3.3(b). The unfolding rate constants were fitted to the Bell model while the folding ones were fitted to Schlierf model. By the Bell model fitting, the kinetic parameters of unfolding rate constant at zero force k_{u}^0 and unfolding distance x_{u} were found to be 0.0015 s^{-1} and 0.7nm respectively. In contrast, the Schlierf fitting took the nonlinear factors into account, so the free energy associated to the protein transition was obtained by numerical integration as mentioned in last paragraph. The folding rate constant of NuG2 at zero force k_{f}^0 was extrapolated to be 28000 s^{-1}. These two rate constants together defined $\Delta G^0_{\mathrm{NS-DS}}$, the free energy difference between native state and denatured state at zero force, to be 16.7$k_{\mathrm{B}}T$ according to equation 2.19.

However, GB proteins (including GB, GB30, GB77 and GB88), according to some bulk chemical denaturation reports, was considered to be three-state folding mechanism in the environment of chemical denaturant involving two transition states since a pronounced curvature occurred in the unfolding arm of Chevron plot. In contrast, our mechanical study on NuG2 and previous AFM experiments on NuG2 and GB1 concluded that the logarithm of unfolding rate constant had a clear linear relationship against the force which indicated a two-state kinetic model. The reason underlying the difference between these two mechanisms might arise from different folding pathways along different reaction coordinate.

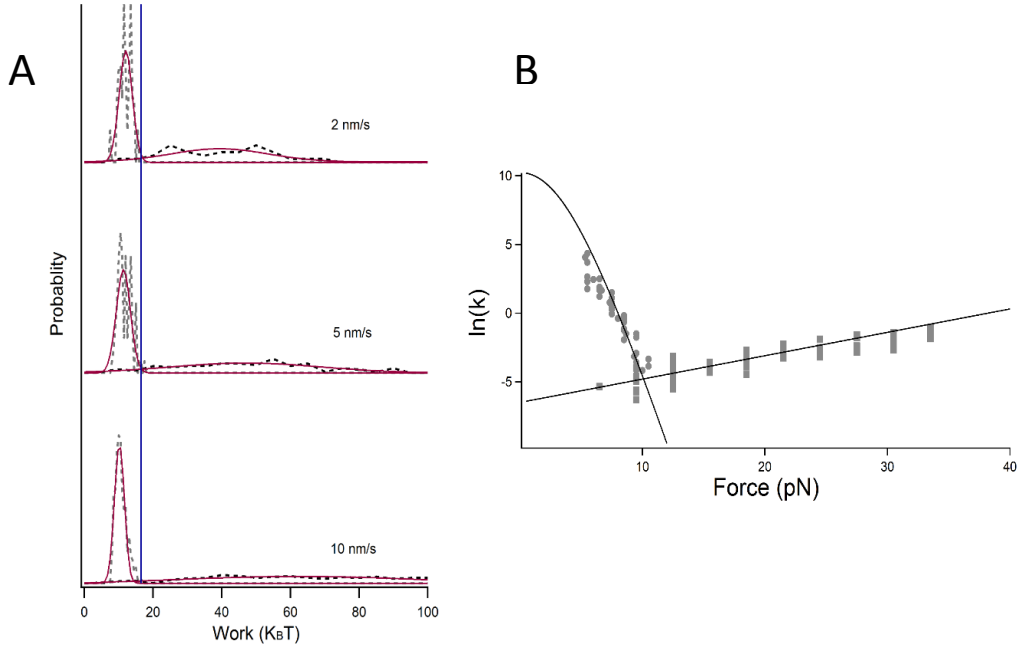

Figure 3.3. Crooks fluctuation theorem and unfolding and folding rate constants extrapolation to determine free energy difference between native state and denatured state at zero force. (a) The probability of protein unfolding and folding associated the free energy. CFT estimates that the intersection is the free energy difference between native state and denatured state. For NuG2, the free energy difference is ~16.5k_BT. (b) The dependency of rate constants on the force is fitted with Bell model and Schlierf model. The folding rate constant k_f^0 =28000 s^{-1}, unfolding rate constant k_u^0=0.0015 s^{-1} and unfolding distance x_u =0.7 nm. The free energy difference is 16.7k_BT.

Constructing polyprotein is a typical protocol in force spectroscopy techniques, especially AFM, which can yield unique sawtooth pattern trajectory of protein sequential unfolding using constant velocity protocol. Such characteristic sawtooth patterns were also used to identify the single molecule stretching. Although the DNA handles used in optical tweezers are sufficient to distinguish single molecule event through its well-characterized properties, the polyprotein protocol is still useful to improve data throughput by increasing the amount of rupture events in each trajectory. Here, besides the mechanical studies on the NuG2 monomer, the polyprotein of $(NuG2)_4$, which consisted of four NuG2 repeats, was also studied by optical tweezers. Figure 3.4 shows a representative trajectory yielded by constant velocity protocol. In this stretching and relaxation cycle, four protein modules were involved and produced four rupture and folding events with uniform spacing. The comparison of force distributions between NuG2 monomer and tetramer were built with two typical pulling speeds of 20 and 50 nm/s. Due to the homology of each module, the unfolding and folding events of $(NuG2)_4$ could be considered as 4 independent events of NuG2 monomer, so the force distributions of $(NuG2)_4$ were very close to the distributions of monomer.

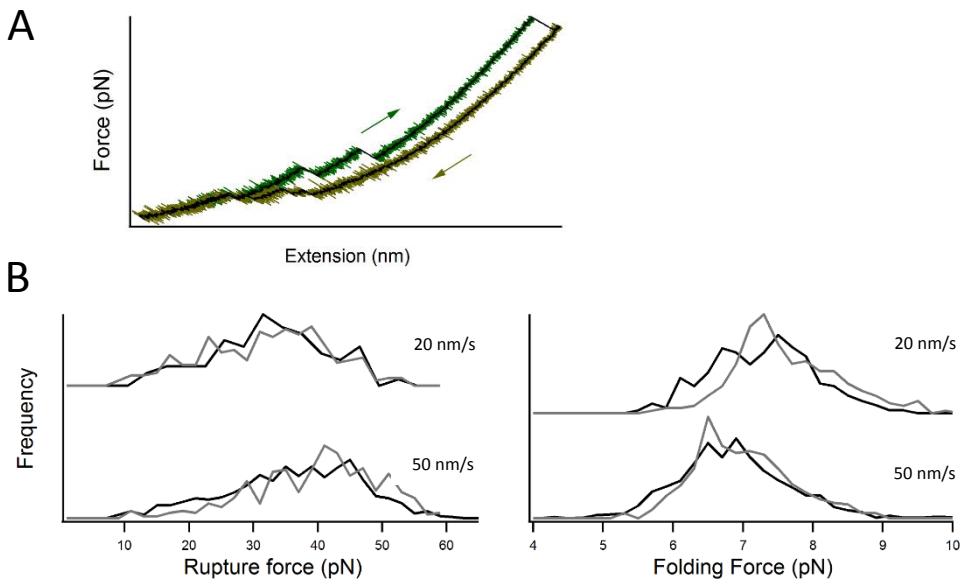

Figure 3.4. Characteristic pulling trajectories of NuG2 tetramer and rupture force and folding force distributions. (a) A characteristic trajectory for $(NuG2)_4$ of a stretching (green) and relaxation (brown) cycle. The trajectory may have up to 4 rupture or folding events. (b) The comparison on rupture force (left) and folding force (right) distributions at pulling speeds of 20 nm/s (top) and 50 nm/s (bottom) between monomer (grey) and tetramer (black). The distributions between monomer and tetramer do not have significant difference.

Due to the nonequilibrium nature in NuG2 unfolding and folding, dual constant force protocol was also carried out for $(NuG2)_4$. The stretching force was raised to relatively high, e.g. 20 pN, to unfold all the protein modules shortly. In fact, force suitable for unfolding the protein was too high for the denatured protein to fold. Therefore, once all the modules unfolded, the force dropped sharply down to a reasonable value, e.g. 8.5 pN, to facilitate protein folding. A typical trajectory yielded by this protocol was shown in Figure 3.5(a). It was clear that the molecule unfolded 4 times at a high force and then folded 4 times at a low force. Dwell time analysis can be applied to extract the rate constants of protein unfolding and folding under these forces. For polyprotein, the dwell time for each module at native state or denatured state was counted from the start of this

constant force to each unfolding event or folding event. The rate constants can be obtained by fitting the dwell time distribution to formula 1.2. A direct comparison between the rate constants obtained from monomer and the values obtained from tetramer was shown in Figure 3.5(b). The close values of rate constants indicated the consistency between different methods. Thus the fitting of Bell model and Schlierf model yielded close values for k_{u}^{0} (22,000 s^{-1}), Δx_{u} (0.6 nm), k_{f}^{0} (0.002 s^{-1}) and $\Delta G_{\mathrm{NS-DS}}^{0}$ (16.2 $k_{\mathrm{B}}T$). Those close kinetic values yielded from different methods and observations implied that the free energy profile of protein unfolding was an intrinsic property, independent of observing methods. These two different methods of obtaining the kinetic properties have their own advantages: using Oberbarnscheidt method can yield the rate constants in a longer range of force; the constant force protocol can yield rate constants without approximation, so the rate constants might be more accurate.

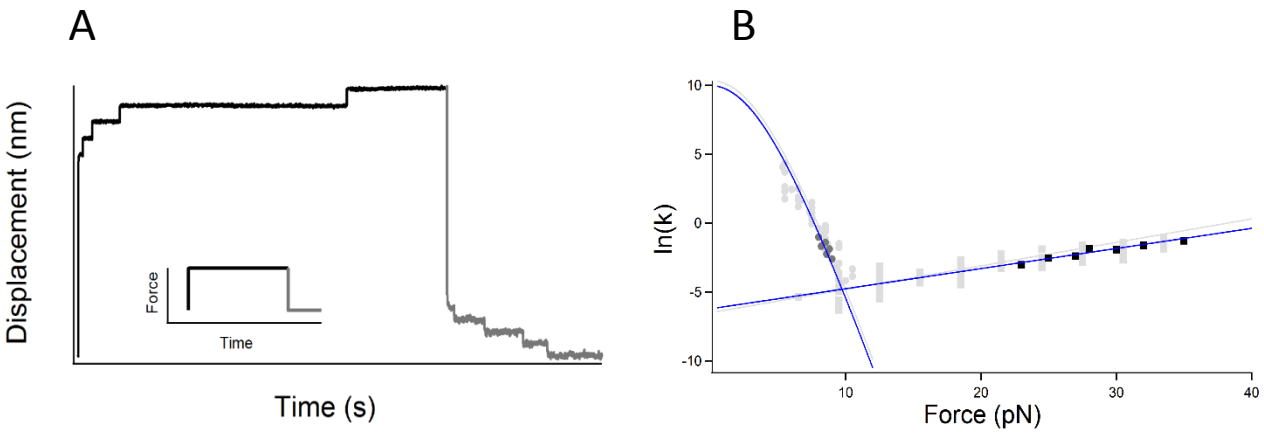

Figure 3.5. The displacement trajectory of NuG2 tetramer under dual constant force protocol and the extracted rate constants. (a) A typical trajectory of dual constant force protocol. The stretching force is first raised up to and kept at 20 pN, and then the protein moduli are sequentially unfolded resulting in step shaped increase in displacement. After 4 rupture events happened, the force is dropped down to 8.5 pN. 4 folding events will happen afterwards. The rate constants extracted from dual constant force protocol was shown in Figure (b). The folding rate constants (grey circle) are fitted with Schlierf model (blue curve) and the unfolding rate constants (black square) are fitted with Bell model (blue curve). The light grey plots and curves are the results from the constant velocity protocol of NuG2 monomer, serving as a reference. The unfolding rate constant k_u^0 =22000 s^{-1}, folding rate constant k_f^0 =0.002 s^{-1} and unfolding distance x_u =0.6 nm. The free energy difference is 16.2k_BT.

4. Mechanical unfolding and folding of GA protein and mutants of 3α and α/β conformations

4.1. Synopsis

In this chapter, we continued to investigate the mechanical unfolding and folding for the other wild type protein GA in the binary conformation system, which had not been studied by force spectroscopy before. From its conformation it was predicted that this protein had relatively low mechanical stability and was suitable for study by optical tweezers. The force spectroscopy results, obtained by using same methods as NuG2, reflected that the unfolding and folding of this protein were near equilibrium. Besides the studies on GA, other mutants including GA30, GA77, GA95 and GB30 were also investigated with constant velocity protocol, in order to understand how the free energy profile of protein unfolding changes as the sequence homology increased.

4.2. Results and discussions

Previous force spectroscopy studies about proteins of α-helix bundles, such as calmodulin and spectrin, showed that this class of protein had weak mechanical stabilities. GA, having a 3α-helix bundle conformation, was predicted to have similar mechanical stability. Typical force-extension curves of a constant velocity cycle were shown in Figure 4.1. This protein tended to unfold and fold at values close to ~10 pN. Close values of rupture force and folding force suggested that GA had very small hysteresis between its unfolding and folding, implying that the unfolding and folding of this protein were close to equilibrium in contrast to the non-equilibrium nature of NuG2 unfolding and folding. The folding force was so close to the unfolding force that this protein had a great chance to fold shortly after the protein unfolding. As a result, this protein might transit

rapidly between native state and denatured state in a proper force range and yielded rapid hopping events reflecting in force extension curves. Such near equilibrium phenomenon has been also observed in the mechanical study of calmodulin (Stigler, 2011).

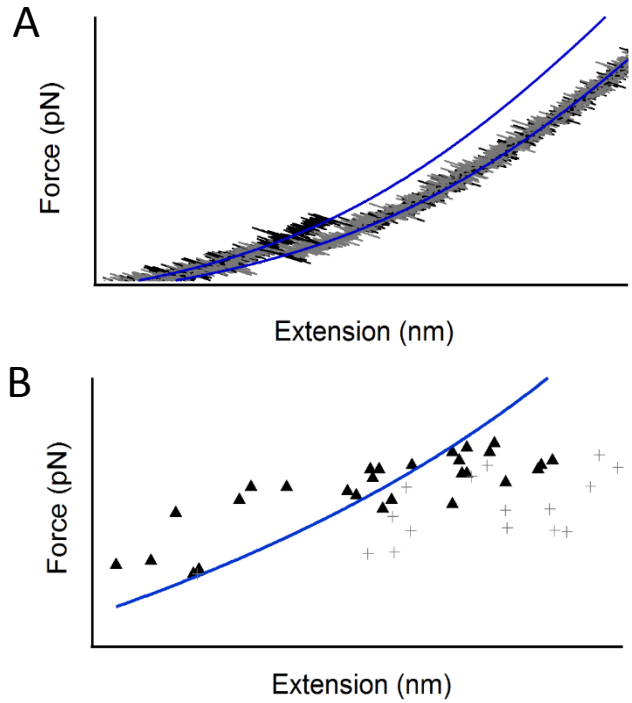

Figure 4.1. Contour length increment fitting for GA. (a) eWLC chain fitting (blue curves) for the molecule force-extension curve of a typical stretching (black plot) and relaxation (grey plot) with parameters: persistence length p_{DNA}=16 nm and $p_{protein}$=0.7 nm; contour length L=548 nm and increment of protein ΔL=13.5 nm; modulus of DNA K=480. (b) WLC fitting (black curve) for the extension of unstructured protein (black triangle) extracted from the extension difference between native state and denatured state. The fitting parameters are $p_{protein}$=0.8 and L=18 nm. The grey cross denotes the plot of extension of NuG2 serving as a reference.

The force-extension curves were fitted to the eWLC model to obtain the characteristic value of contour length increment. Using the same procedure mentioned in chapter 3, we could estimate the contour length increment to be ~14 nm. This value was apparently smaller than the contour length increment of NuG2, although they were synthesized of 56 amino acids. The conformation of GA could account for this difference. Its conformation consisted of a globular domain of 48

structured amino acid residues plus two unstructured peptide chains on its two ends. Since the contour length increment only reflected the conformation change upon unfolding, only the structured amino acid residues should be taken into consideration. Therefore, the theoretical value of the contour length increment for this protein was calculated to be 48 * 0.36 – 3.16 = 14.1 nm where the contour length of native state was estimated using the NMR spectroscopy from protein data bank. The extension difference upon protein transition was also analyzed and relationship between extension and force of denatured GA was included in Figure 4.1 (b). The extension of NuG2 was introduced as a reference, imdicating that the contour length increment of GA was less than NuG2.

The rupture force and folding force distributions also revealed the near equilibrium nature of the unfolding and folding of GA. The force distributions, which were yielded with pulling speeds of 5, 10, 20, 50 and 100 nm/s, were shown in Figure 4.2(a). The rupture force and folding force distributions at each pulling speed were largely overlapped, indicating that the unfolding and folding were close to equilibrium under these pulling speeds. Even for the pulling speed of 100 nm/s which was the largest perturbation among these pulling speeds and drove the system to the most non-equilibrium, the rupture force and folding force distributions were still largely overlapped, showing a very different phenomenon from force distributions of NuG2. The near equilibrium nature was also reflected by the pulling speed dependence of average forces, which was shown in Figure 4.2(b). The average rupture force and folding force only changed insignificantly as pulling speed changed exponentially, appearing as two flat curves over this range of pulling speeds. This flat pulling speed dependence only occurred in very slow pulling speed (slower than 10 nm/s) for NuG2 where its unfolding and folding were close to equilibrium (He et al., 2015). CFT was used to extract the free energy difference between native state and denatured

state at zero force $\Delta G^0_{\mathrm{NS-DS}}$ with similar procedures but different parameters for WLC model. The $\Delta G^0_{\mathrm{NS-DS}}$ was ultimately estimated to be $7.9k_{\mathrm{B}}T$.

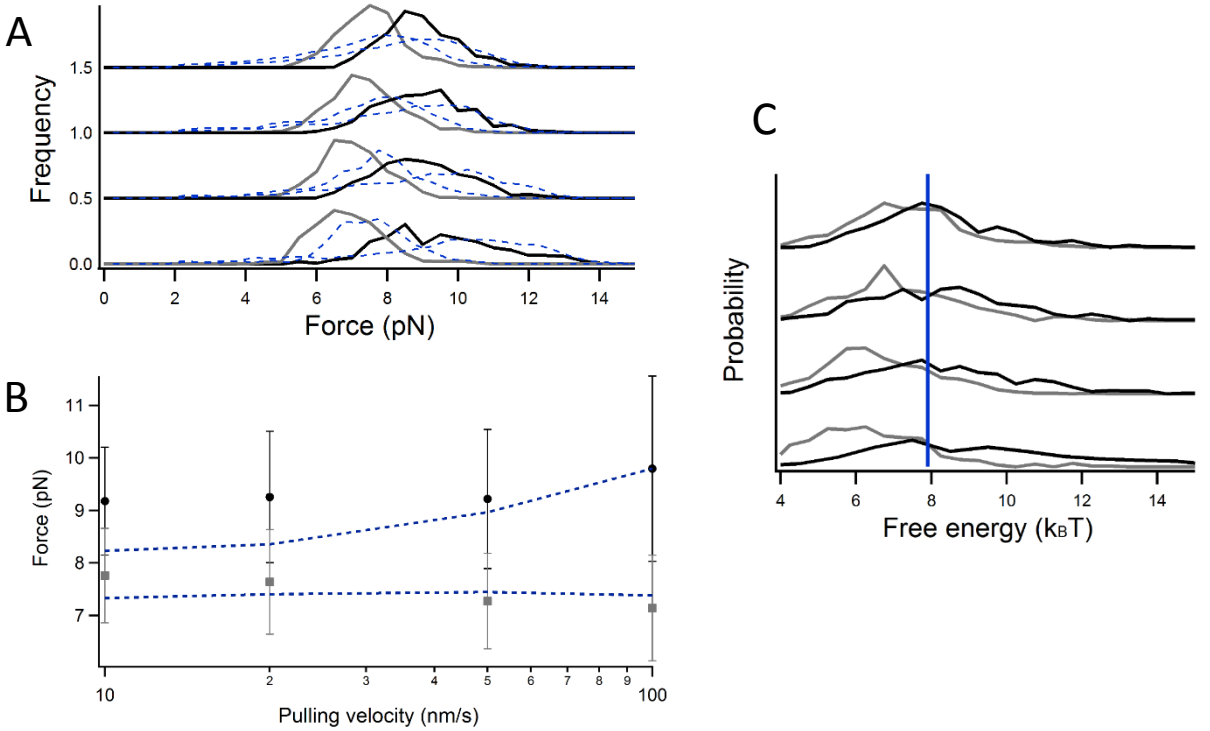

Figure 4.2. The rupture force and folding force distributions for GA, its dependency of forces on pulling velocity and Crooks fluctuation theorem fitting. (a) The rupture force (grey) and folding force (black) distributions of GA under different pulling velocities. The pulling velocities from top to bottom are 10, 20, 50 and 100 nm/s. Under each velocity, the rupture force and folding force distributions are largely overlapped, implying near equilibrium between folding and unfolding of GA. (b) the dependency of forces on the pulling velocity. For both of the rupture force and folding force, their dependency are flat. The blue dash line in Figure (a) and (b) are the Monte Carlo simulation results assuming k_{f}^0=450 s^{-1}, k_{u}^0=0.04 s^{-1} and *x*u=2.0 nm. (c) The probability of folding and unfolding depending on free energy. CFT estimates the free energy difference between native state and denatured state is 7.9 $k_{\mathrm{B}}T$.

To further characterize the unfolding and folding rate constants under zero force, the Oberbarnscheidt method was introduced to obtain the rate constants under different forces from force-extension curves. The relationship between rate constant and force was shown in Figure 4.3 which was fitted with the Bell model and the Schlierf model. These two models extrapolated that

the unfolding rate constant at zero force k_u^0 =0.04 s^{-1}, unfolding distance Δx_u =2.0 nm and folding rate constant at zero force k_f^0 = 450 s^{-1}. From the values of k_u^0 and k_f^0, ΔG_{NS-DS}^0 could be calculated to be 9.4k_BT (equivalent to 5.6 kcal/mol, which was very close to the value of 6.0 kcal/mol extracted from thermal denaturation). Using these kinetic parameters, we could use the Monte Carlo simulation to reproduce force-extension curves, rupture forces and folding forces of GA. The simulation rupture force and folding force distributions and speed dependence were compared with experimental data in Figure 4.2. Additionally, the reason underlying the lower rupture force of GA could be explained by its longer unfolding distance, which required lower force to eliminate the same energy barrier according to the Bell model. Therefore, although the unfolding rate constant of GA (0.04 s^{-1}) was less than the value of NuG2 (0.14 s^{-1}), GA tended to unfold at lower force.

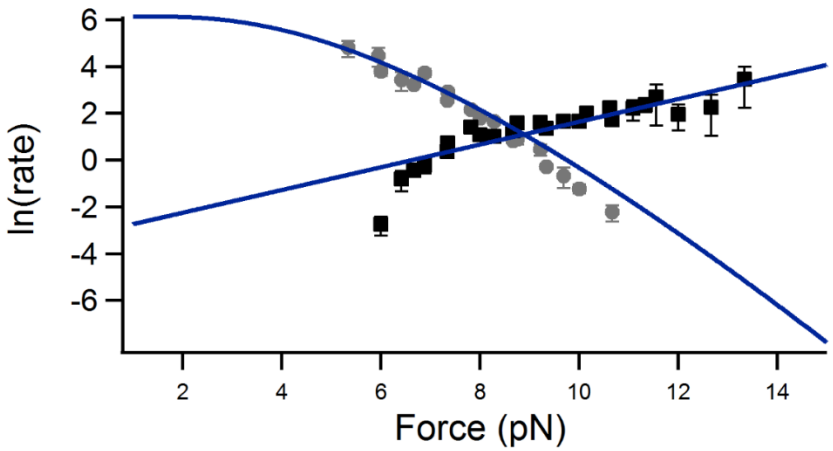

Figure 4.3. The dependency of rate constants of GA on the force fitted with the Bell model and the Schlierf model. Black squares denote unfolding rate constants and grey circles denote folding rate constants. The unfolding rate constant at zero force k_u^0 =0.04 s^{-1}, folding rate constant at zero force k_f^0=450 s^{-1} and unfolding distance x_u =2.0 nm. The free energy difference is 9.4k_BT.

The constant force protocol for GA monomer was slightly different from the one for NuG2 tetramer, since unfolding and folding of GA were close to equilibrium. The rupture force and folding force distributions of GA were largely overlapped as mentioned before, so this protein could possess comparable unfolding rate and folding rate at a force within this overlapped range. Therefore, the regular constant force protocol could be applied to study this protein. The distance trajectories under different forces yielded by this protocol were shown in Figure 4.4. The displacement distribution along the trajectory could be considered as the combination of two normal distributions, indicating that the protein unfolding and folding \was two-state behavior without any intermediate state. As the stretching force increased, the probability of observations at native state decreased while the probability of observation at denatured state increased. Accordingly, there was a tendency that the equilibrium shifted gradually from native state to denatured state. Dwell time analysis could estimate the folding and unfolding rate constants under these forces. These results were shown in Figure 4.4 compared with the rate constants extracted from constant velocity protocol. The kinetic properties from constant force protocol included: k_{u}^{0} = 0.05 s^{-1}; Δx_{u} =2.0 nm; and k_{f}^{0} = 400 s^{-1}, which were very close to the results from constant velocity protocol. The corresponding $\Delta G_{\mathrm{NS-DS}}^{0}$ was $9.0k_{\mathrm{B}}T$.

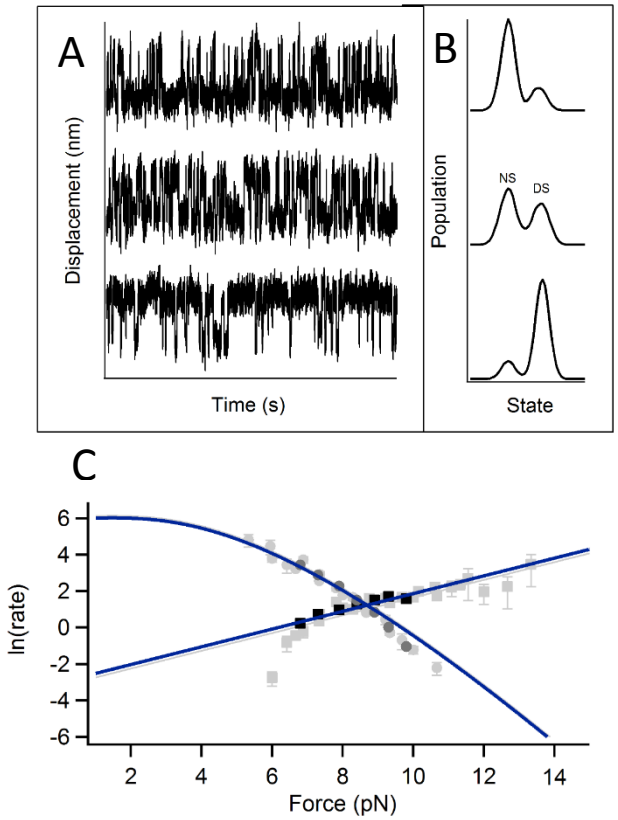

Figure 4.4. The displacement trajectory of GA under constant force protocol and the extracted rate constants. (a) The typical trajectories yielded from GA with constant force protocol. The set forces are descending from top to bottom with values of 7.8, 8.3 and 9.3 pN. As the stretching force increases, the protein tend to populate more on the denatured state. The tendency is very clear in Figure (b), which are histograms of displacement for each trajectory. At 8.3 pN, the population at native state and denatured state are close to equilibrium. (c) The unfolding (black squares) and folding (grey circles) rate constants extracted from the constant force trajectories are fitted to the Bell model and the Schlierf model with the following parameters: k_u^0 = 0.05 s^{-1}; Δx_u =2.0 nm; and k_f^0 = 400 s^{-1}. The results from constant velocity protocol (light grey) were introduced for comparison.

Besides the mechanical studies on GA, we continued to investigate other 3α-helix bundle conformation proteins including GA30, GA77 and GA95. Here, the constant velocity protocol associated with Oberbarnscheidt method (which can yield accurate estimates without introducing Monte Carlo simulations in traditional methods) was carried out to estimate the kinetic parameters for these proteins. GA30 was a mutant of GA with only 2 mutations on the second helix and 1 mutation on the third helix. It was reported to have a free energy difference between native state

and denatured state ($\Delta G^0_{\mathrm{NS-DS}}$) of 6 kCal/mol according to the thermal denaturation, as thermaldynamically stable as GA. From Bell model and Schlierf fitting, GA30's kinetic properties were found that: k^0_{u} = 0.035 s^{-1}; Δx_{u} =2.0 nm; and k^0_{f} =500 s^{-1}, defining $\Delta G^0_{\mathrm{NS-DS}}$ to be 9.6$k_{\mathrm{B}}T$. GA77 had 10 more mutations than GA30 involving all three helices. The thermal denaturation revealed that $\Delta G^0_{\mathrm{NS-DS}}$ of GA77 was 5 kcal/mol, less thermaldynamic stability than GA and GA30. The kinetic properties of GA77 were found that: k^0_{u} =0.15 s^{-1}; Δx_{u} =2.0 nm; and k^0_{f} =180 s^{-1}, defining $\Delta G^0_{\mathrm{NS-DS}}$ to be 7.1$k_{\mathrm{B}}T$. GA95, which was had only 3 different residues with GB95, had about 20 mutations distributing in all three helices compared wild type GA. The thermal stability of this protein was even lower than GA77 with a reported $\Delta G^0_{\mathrm{NS-DS}}$ value of ~3 kcal/mol. The kinetic properties for this protein were found to be: k^0_{u} =0.18 s^{-1}; Δx_{u} =2.9 nm; and k^0_{f} =100 s^{-1}. The corresponding $\Delta G^0_{\mathrm{NS-DS}}$ was 6.3$k_{\mathrm{B}}T$.

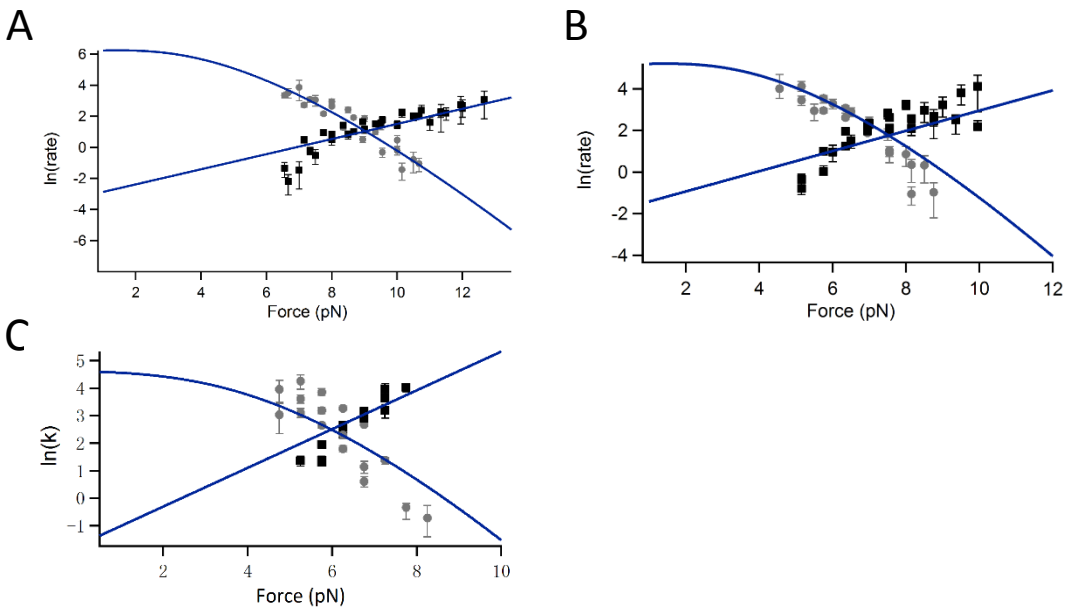

Figure 4.5. The dependency of rate constants of GA30 (a), GA77 (b) and GA95 (c) on the force. The unfolding rate constants are shown in black squares and the folding rate constants are shown in grey circles. The properties of kinetics of these proteins are fitted to be: for GA30, k^0_{u} = 0.035 s^{-1}, Δx_{u} =2.0 nm and k^0_{f} =500 s^{-1}; for GA77, k^0_{u} =0.15 s^{-1}, Δx_{u} =2.0 nm, and k^0_{f} =180 s^{-1}; for GA95, k^0_{u} =0.18 s^{-1}, Δx_{u} =2.9 nm, and k^0_{f} =100 s^{-1}.

Given kinetic parameters of this series of proteins, we want to emphasize that the folding rate constants k_f^0 of GA77 and GA95 decreased while the unfolding rate constants k_u^0 increased as homology percentage increased except GA30 which had only 3 mutation sites, implying that the free energy of both the native state and transition state have changed. On the other hand, how the mechanical stability changed for mutants of α/β conformation was equally important. Among those mutants, only GB30 have been investigated so far. In this mutant, although only 5 mutation sites were introduced into the wild type GB, the thermaldynamic stability of GB30 dropped from 7 kcal/mol down to 4.5 kcal/mol. Figure 4.6 included the rate constant dependency on force extracted from constant velocity data. The kinetic parameters of this protein was extrapolated, given by: k_u^0 =0.05 s^{-1}; Δx_u =0.5 nm; and k_f^0 =500 s^{-1}. The corresponding ΔG^0_{NS-DS} was 9.2k_BT.

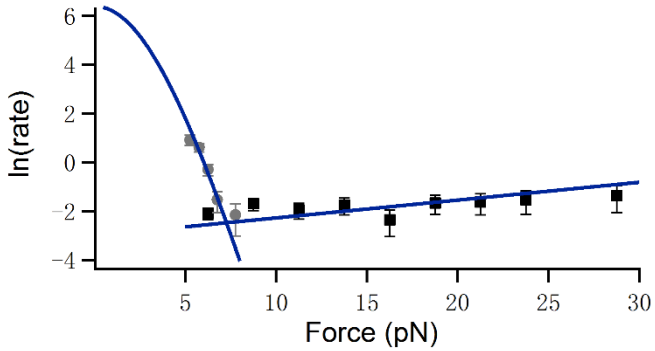

Figure 4.6. The dependency of rate constants of GB30 on the stretching force. The rate constants are fitted to Bell model and Schlierf model with parameters: k_f^0 = 500 s^{-1}, Δx_u =0.5 nm and k_u^0 = 0.05 s^{-1}.

Our force spectroscopy study on these 3α proteins has revealed that the folding and unfolding rate constants were associated with mutations. The kinetic analysis showed that the folding rate constant decreased and the unfolding rate constant increased along with the mutations but the unfolding distances for these mutants were approximately the same value of 2.0 nm. Rate constants were strongly associated with the transition barrier. Figure 4.7 was a sketch of free energy profile

under two-state model assuming that denatured proteins, namely polypeptides, have identical free energy level. The increase of unfolding rate constant implied that the barrier from native state to denatured state is diminished which was quantified by the unfolding rate constant. This result was consistent with thermal denaturation experiments which revealed that the conformation stability decreased as the sequence identity increased.

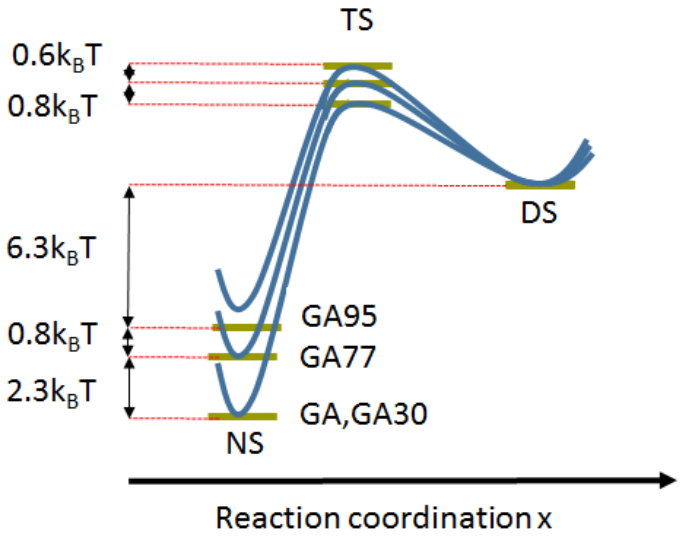

Figure 4.7. Sketch of free energy profiles for the GA proteins including GA, GA30, GA77 and GA95 based on the change in kinetics of unfolding and folding. The mutations done on the GA not only change the free energy of the native state, but also slightly shift the free energy of the transition state.

A protein switch that can adopt multiple native conformations needs to fulfill requirements on thermodynamics and kinetics. The thermodynamic requirement defines close free energies of different native states to guarantee close populations at different states. The kinetic requirement means that the protein should have adequate transiting rate to other conformations in reasonable time scale. For this potential protein switch involving two unrelated conformations, a drastic overall change should take place in global conformation. Therefore, it is very likely that the

transformation between different conformations is associated with the state of unstructured protein. If so, the decrease of unfolding barrier from GA to GA95 may accelerate the protein unfolding and thus convert into the other conformation.

However, it is too difficult to fully understand the mechanism underlying the mutation-controlled conformational change only with the free energy profiles of GA proteins. Hence, we studied the mutant of GB30, which has 4 mutation sites compared with GB. This mutant has similar folding and unfolding rate constants but much smaller unfolding distance compared to GA proteins. If compared to NuG2, their unfolding distances and unfolding rate constants are close but GB30 has much smaller folding rate. Although it is unlikely to conclude the change in kinetics for the following mutants, GB30 would open an avenue to the following investigations for other GB mutants.

5. Conclusions and future prospects

Chapter 3 described the feasibility of studying an α/β protein NuG2 using optical tweezers. Characteristic patterns of protein folding and unfolding could be observed in both constant velocity and constant force protocols. No matter what observation and data analysis, same values of properties associated to the free energy profile of protein unfolding could be obtained. In terms of the kinetics of NuG2 under zero force, the folding rate constant (k_u^0), the unfolding rate constant (k_f^0), the unfolding distance (x_u) and the energy difference between native state and denatured state (ΔG_{NS-DS}^0) were measured to be ~25,000 s^{-1}, ~0.002 s^{-1}, 0.7 nm and 16.5k_BT respectively which were consistent with previous studies. That the unfolding and folding of NuG2 was non-equilibrium also indicated that other α/β proteins (GB) were very likely to have the same feature.

Chapter 4 mainly described the unfolding and folding features of 3α proteins. This conformation had low mechanical stability and its unfolding and folding were close to equilibrium. The kinetic parameters of GA including k_u^0, k_f^0, x_u and ΔG_{NS-DS}^0 were measured to be 0.04 s^{-1}, 450 s^{-1}, 2.0 nm and 9.1k_BT. Other mutants including GA30, GA77 and GA95 were also studied. For GA30, k_u^0, k_f^0, x_u and ΔG_{NS-DS}^0 were 0.035 s^{-1}, 500 s^{-1}, 2.0 nm and 9.6k_BT. For GA77, k_u^0, k_f^0, x_u and ΔG_{NS-DS}^0 were 0.15 s^{-1}, 180 s^{-1}, 2.0 nm and 7.1k_BT. For GA95, k_u^0, k_f^0, x_u and ΔG_{NS-DS}^0 were 0.18 s^{-1}, 100 s^{-1}, 2.9 nm and 6.3k_BT. The change in unfolding and folding rate constants revealed that the unfolding barrier decreased as sequence identity increased. The decrease of unfolding barrier might facilitate the transformation between different conformations.

In this thesis, only one α/β protein GB30 on the mutational pathway was investigated. In order to understand the change in unfolding and folding kinetics for α/β proteins as sequence identity

increases, other mutants, such as GB77 and GB95, should be studied in the future. Also, different conformations can be distinguished by force spectroscopy according to their mechanical stability. Thus, it is very helpful to dynamically investigate the transformation between different native conformations in equilibrium, such as the dual binding affinity of GA98.

References:

Alexander, P. A., He, Y., Chen, Y., Orban, J., & Bryan, P. N. (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proceedings of the National Academy of Sciences*, *104*(29), 11963-11968.

Alexander, P. A., He, Y., Chen, Y., Orban, J., & Bryan, P. N. (2009). A minimal sequence code for switching protein structure and function. *Proceedings of the National Academy of Sciences*, *106*(50), 21149-21154.

Anfinsen, C. B. (1972). Studies on the principles that govern the folding of protein chains.

Ashkin, A., & Dziedzic, J. M. (1973). Radiation pressure on a free liquid surface. *Physical Review Letters*, *30*(4), 139.

Ashkin, A. (1978). Trapping of atoms by resonance radiation pressure. *Physical Review Letters*, *40*(12), 729.

Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E., & Chu, S. (1986). Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters*,*11*(5), 288-290.

Ashkin, A., & Dziedzic, J. M. (1987). Optical trapping and manipulation of viruses and bacteria. *Science*, *235*(4795), 1517-1520.

Baumann, C. G., Bloomfield, V. A., Smith, S. B., Bustamante, C., Wang, M. D., & Block, S. M. (2000). Stretching of single collapsed DNA molecules.*Biophysical journal*, *78*(4), 1965-1978.

Bell, G. I. (1978). Models for the specific adhesion of cells to cells. *Science*,200(4342):618–27.

Bryan, P. N., & Orban, J. (2010). Proteins that switch folds. *Current opinion in structural biology*, *20*(4), 482-488.

Bryngelson, J. D., & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, *84*(21), 7524-7528.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis.*Proteins: Structure, Function, and Bioinformatics*, *21*(3), 167-195.

Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–600.

Bustamante, C., Chemla, Y. R., Forde, N. R., & Izhaky, D. (2004). Mechanical processes in biochemistry. *Annual review of biochemistry*, *73*(1), 705-748.

Cao, Y., Lam, C., Wang, M., & Li, H. (2006). Nonmechanical protein can have significant mechanical stability. *Angewandte Chemie*, *118*(4), 658-661.

Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., & Fernandez, J. M. (1999). Mechanical and chemical unfolding of a single protein: a comparison. *Proceedings of the National Academy of Sciences*, *96*(7), 3694-3699.

Carrion-Vazquez, M., Oberhauser, A. F., Fisher, T. E., Marszalek, P. E., Li, H., & Fernandez, J. M. (2000). Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Progress in biophysics and molecular biology*, *74*(1), 63-91.

Cecconi, C., Shank, E. A., Bustamante, C., & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science*,*309*(5743), 2057-2060.

Cecconi, C., Shank, E. A., Dahlquist, F. W., Marqusee, S., & Bustamante, C. (2008). Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *European Biophysics Journal*, *37*(6), 729-738.

Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., & Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, *437*(7056), 231-234.

Cordes, M. H., Burton, R. E., Walsh, N. P., McKnight, C. J., & Sauer, R. T. (2000). An evolutionary bridge to a new protein fold. *Nature Structural & Molecular Biology*, *7*(12), 1129-1132.

Crooks, G. E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*,*60*(3), 2721.

Dalal, S., & Regan, L. (2000). Understanding the sequence determinants of conformational switching using protein design. *Protein Science*, *9*(09), 1651-1659.

Dietz, H. and Rief, M. (2006). Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1244–7.

Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on.*Science*, *338*(6110), 1042-1046.

Dobson, C. M., Šali, A., & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angewandte Chemie International Edition*, *37*(7), 868-893.

Finkelstein, A. V., Badretdinov, A. Y., & Gutin, A. M. (1995). Why do protein architectures have boltzmann-like statistics?. *Proteins: Structure, Function, and Bioinformatics*, *23*(2), 142-150.

Finer, J. T., Simmons, R. M., & Spudich, J. A. (1994). Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature*, *368*(6467), 113-119.

Gebhardt, J. C. M., Bornschlögl, T., & Rief, M. (2010). Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences*, *107*(5), 2013-2018.

Greenleaf, W. J., Woodside, M. T., & Block, S. M. (2007). High-resolution, single-molecule measurements of biomolecular motion. *Annual review of biophysics and biomolecular structure*, *36*, 171.

He, C., Hu, C., Hu, X., Hu, X., Xiao, A., Perkins, T. T., & Li, H. (2015). Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angewandte Chemie*.

He, Y., Chen, Y., Alexander, P. A., Bryan, P. N., & Orban, J. (2012). Mutational tipping points for switching protein folds and functions. *Structure*,*20*(2), 283-291.

Holm, L., & Sander, C. (1996). Mapping the protein universe. *Science*,*273*(5275), 595-602.

Junker, J. P., Ziegler, F., & Rief, M. (2009). Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science*, *323*(5914), 633-637.

Karplus, M., & Šali, A. (1995). Theoretical studies of protein folding and unfolding. *Current opinion in structural biology*, *5*(1), 58-73.

Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, *89*(18), 8721-8725.

Levinthal, C. (1968). Are there pathways for protein folding. *J. Chim. phys*,*65*(1), 44-45.

Levinthal, C. (1969). Mossbauer Spectroscopy in Biological Systems Proceedings of a meeting held at Allerton House Monticello Illinois. *P. Debrunner, JCM Tsibris, and E. M nck*.

Lu, H., & Schulten, K. (1999). Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Structure, Function, and Bioinformatics*, *35*(4), 453-463.

Luo, X., Tang, Z., Xia, G., Wassmann, K., Matsumoto, T., Rizo, J., & Yu, H. (2004). The Mad2 spindle checkpoint protein has two distinct natively folded states. *Nature structural & molecular biology*, *11*(4), 338-345.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*, *247*(4), 536-540.

Oberbarnscheidt, L., Janissen, R., & Oesterhelt, F. (2009). Direct and model free calculation of force-dependent dissociation rates from force spectroscopic data. *Biophysical journal*, *97*(9), L19-L21.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., & Gaub, H. E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM.*science*, *276*(5315), 1109-1112.

Rief, M., Pascual, J., Saraste, M., & Gaub, H. E. (1999). Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *Journal of molecular biology*, *286*(2), 553-561.

Schlierf, M., Berkemeier, F., & Rief, M. (2007). Direct observation of active protein folding using lock-in force spectroscopy. *Biophysical journal*, *93*(11), 3989-3998.

Smith, S. B., CUI, Y., & Bustamante, C. (2002). Optical-trap force transducer that operates by direct measurement of light momentum. *Methods in enzymology*, *361*, 134-162.

Smith, S. B., (2009). How to calibrate the mini-tweezers. Retrieved from http://tweezerslab.unipr.it/cgi-bin/mt/documents.pl

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. M., & Rief, M. (2011). The complex folding network of single calmodulin molecules. *Science*, *334*(6055), 512-516.

Svoboda, K., Schmidt, C. F., Schnapp, B. J., & Block, S. M. (1993). Direct observation of kinesin stepping by optical trapping interferometry. *Nature*,*365*(6448), 721-727.

Tuinstra, R. L., Peterson, F. C., Kutlesa, S., Elgin, E. S., Kron, M. A., & Volkman, B. F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. *Proceedings of the National Academy of Sciences*, *105*(13), 5057-5062.

Volkman, B. F., Liu, T. Y., & Peterson, F. C. (2009). Lymphotactin structural dynamics. *Methods in enzymology*, *461*, 51-70.

Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophysical journal*, 72(3):1335–46.

Yao, M., Chen, H., & Yan, J. (2015). Thermodynamics of force-dependent folding and unfolding of small protein and nucleic acid structures. *Integrative Biology*.

Zhang, X., Halvorsen, K., Zhang, C. Z., Wong, W. P., & Springer, T. A. (2009). Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science*, *324*(5932), 1330-1334.

Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences*, *89*(1), 20-22.

Appendix A

Sequences of GA, GB and other mutants

GA
Protein sequence

MEAVDANSLA QAKEAAIKEL KQYGIGDYYI KLINNAKTVE GVESLKNEIL KALPTE

Nucleotide sequence

ATGGAAGCGGTGGACGCAAACTCTCTGGCACAAGCGAAAGAAGCGGCTATCAAAGA
ACTGAAACAGTATGGTATTGGCGATTATTATATTAAACTGATCAACAATGCGAAAAC
GGTTGAAGGTGTTGAATCGCTGAAAAATGAAATCCTGAAAGCCCTGCCGACGGAA

GB
Protein sequence

MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE

Nucleotide sequence

ATGACCTACAAACTGATCCTGAACGGTAAAACCCTGAAAGGTGAAACCACCACCGA
AGCTGTAGACGCTGCTACTGCAGAAAAAGTTTTCAAACAGTACGCTAACGACAACG
GTGTCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACGGTTACCGAA

GA30
Protein sequence

MEAVDANSLA QAKEAAIKEL KQYGIGEKYI KLINNAKTVE GVWSLKNEIL KALPTE

Nucleotide sequence

ATGGAGGCGGTGGACGCCAACAGCCTGGCGCAGGCGAAGGAGGCGGCGATCAAGG
AGTTGAAGCAATACGGCATCGGGGAAAAATACATCAAACTGATCAACAACGCCAAA
ACCGTGGAAGGCGTGTGGTCGTTAAAAAACGAGATCTTGAAGGCCTTGCCGACGGA
A

GB30
Protein sequence

MTYKLILNGK TLKGETTTEA VDAATAEKYF KLYANDKTVE GEWTYDDATK TFTVTE

Nucleotide sequence

ATGACCTACAAACTGATTCTGAACGGTAAAACGCTGAAAGGCGAAACGACCACGGA
AGCTGTTGATGCGGCGACGGCTGAAAAATACTTTAAACTGTATGCGAACGATAAAA
CGGTGGAAGGCGAATGGACCTATGACGATGCGACGAAAACCTTTACGGTGACGGAA

GA77
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL KATVTE

Nucleotide sequence

ACCACTTACAAATTAATCCTGAACCTGAAACAGGCGAAGGAGGAGGCGATCAAGGA
GTTGGTGGACGCGGGCATCGCGGAAAAATACATCAAACTGATCGCCAACGCCAAAA
CCGTGGAAGGCGTGTGGACCTTAAAAGACGAGATCTTGAAGGCCACCGTGACGGAA

GB77
Protein sequence

TTYKLILNGK QLKEEAITEA VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

Nucleotide sequence

ACCACTTACAAATTAATCCTTAATGGTAAACAGTTGAAAGAAGAAGCGATCACTGA
AGCTGTTGATGCTGCTACTGCAGAAAAATACTTCAAACTGTACGCTAACGCCAAAAC
TGTTGAGGGTGTGTGGACTTACAAAGATGAAACTAAGACCTTTACAGTTACTGAA

GA88
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL TFTVTE

GB88
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

GB88b
Protein sequence

TTYKLILNLK QAKEEAITEA VDAGTAEKYF KLYANAKTVE GVWTYKDEIK TFTVTE

GA95
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIK TFTVTE

Nucleotide sequence

ACCACTTACAAATTAATCCTGAACCTGAAACAGGCGAAGGAGGAGGCGATCAAGGA
GTTGGTGGACGCGGGCACCGCGGAAAAATACATCAAACTGATCGCCAACGCCAAAA
CCGTGGAAGGCGTGTGGACCTTAAAAGACGAGATCAAGACCTTCACCGTGACGGAA

GB95
Protein sequence

TTYKLILNLK QAKEEAIKEA VDAGTAEKYF KLIANAKTVE GVWTYKDEIK TFTVTE

Nucleotide sequence

ACCACTTACAAATTAATCCTTAATCTTAAACAGGCGAAAGAAGAAGCGATCAAGGA
AGCTGTTGATGCTGGTACTGCAGAAAAATACTTCAAACTGATCGCTAACGCCAAAAC
TGTTGAGGGTGTGTGGACTTACAAAGATGAAATTAAGACCTTTACAGTTACTGAA

GA98
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAGTAEKYF KLIANAKTVE GVWTLKDEIK TFTVTE

GB98
Protein sequence

TTYKLILNLK QAKEEAIKEL VDAGTAEKYF KLIANAKTVE GVWTYKDEIK TFTVTE

NuG2
Protein sequence

DTYKLVIVLN GTTFTYTTEA VDAATAEKVF KQYANDNGVD GEWTYADATK TFTVTE