MECHANICAL UNFOLDING AND FOLDING STUDIES BY OPTICAL TWEEZERS

by

Yabin Guo

B.Sc., Nankai University, 2010

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)

April 2017

Abstract

As a single molecule technique, optical tweezers technique proves to be a powerful tool to investigate the physical and chemical properties of DNA/RNA and protein molecules. In this thesis, optical tweezers are applied to two studies. In the first study, we directly investigated the unfolding and folding pathways and kinetics of the wild-type Top7 with optical tweezers. The existence of a folding intermediate state is confirmed. The unfolding process also occasionally shows non-cooperative behavior which has not been observed before. To identify if the mechanical stability of an isolated fragment of Top7 is responsible for the non-cooperative unfolding and folding behavior of Top7, we purified the C-fragment of Top7 and found that it reaches equilibrium at low applied forces, which indicates that Top7's C-fragment could unfold and fold independently, but the unfolding and folding behavior of Top7 depends on the mutual assistance of both N-terminal and C-terminal residues. Illuminated by computational simulation methods, six residues were mutated aiming at improving the folding cooperativity of Top7. The results show that the folding cooperativity is improved significantly, while the unfolding intermediate appears more frequently. The possible influence of pathways on the frequency of occurrence of unfolding/folding intermediate state is discussed. In the second study, the two-step unfolding behavior of rubredoxin is revealed by optical tweezers. The reversible unfolding/folding behavior under force pressure and chemical pressure are further studied. Optical tweezers technique is proved to be well suited for mechanical unfolding/folding studies of metalloproteins.

Preface

Chapter 2 is based on unpublished work. My supervisor, Hongbin Li, designed the project. Dr. Chengzhi He helped me with the molecular biology. I prepared the sample, did the optical tweezers experiments, and wrote the chapter.

Chapter 3 is also based on unpublished work. My supervisor, Hongbin Li, designed the project. I performed the molecular biology, prepared the sample, did the optical tweezers experiments, and wrote the chapter.

Table of Contents

List of Figures

List of Symbols

λ	wavelength
d	diameter
F	force
D	half width of the detector
Ψ	silicon detector's responsivity
c	the speed of the light
t	time
k	unfolding/folding rate constant
k_u	unfolding rate constant
$k_u(F)$	unfolding rate constant at force F
$k_u(0)$	unfolding rate constant at zero force
$k_f(F)$	folding rate constant at force F
$k_f(0)$	folding rate constant at zero force
k_f	folding rate constant
Δx_u	unfolding distance
Δx_f	folding distance
k_B	Boltzmann's constant
p	persistence length
T	absolute temperature
x	extension
A	Arrhenius type pre-exponential coefficient

K	stretch modulus
$\Delta G_{FS,TS}$	energy barrier for unfolding process
$\Delta G_{US,TS}$	energy barrier for folding process
τ	average lifetime
κ_s	spring constant
v	pulling speed
$\dot{F}$	loading rate
$\bar{F}$	average force
pN	pico-Newton

List of Abbreviations

WLC	worm-like chain
eWLC	extensible worm-like chain
FS	folded state
US	unfolded state
TS	transition state
FEC	force-extension curve
AFM	atomic force microscopy
aa	amino acid
ssDNA	single strand DNA
dsDNA	double strand DNA
CD	circular dichroism
NMR	nuclear magnetic resonance
DTT	dithiothreitol
PCR	polymerase chain reaction
PDB	protein data bank
PBS	phosphate buffered saline
GB1	the B1 domain of protein G from *Streptococcus*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
wt-Top7	wild-type Top7
cm-Top7	cooperative mutant Top7
RD	rubredoxin

*pf*RD	rubredoxin from *clostridium pasteurianum*
CXXC	cysteine binding motif
TCEP	Tris(2-carboxyethyl)phosphine)

Acknowledgements

First, I would like to thank my supervisor Dr. Li who gave valuable guidance and supports on research throughout my three years of Master study. Without his tremendous help, I would have never been able to finish my project. At the same time, I would like to thank my committee members for their insightful comments and reviews.

I thank Dr. He for his guidance on molecular biology and significant suggestions on the project. I also thank my former and present colleagues for all the working-hard and fun time that we have. I feel lucky to work in such a friendly and vigorous group.

Special gratitude is owing to my parents and girlfriend who stand behind me and encourage me all the time.

Chapter 1: Introduction

In this chapter, an introduction of the instrument we mainly used will be presented. The principle, development and applications of optical tweezers will be discussed.

1.1 Principles of optical tweezers

The force exerted on the particles by the light was first suggested by Johannes Kepler in 1619. And an article named "First experimental evidence for pressure of the light on the solid bodies" was published in 1901 which first proved the pressure of the light on the object. Optical tweezers, also known as optical trap, were invented by Arthur Ashkin in 1970 and were used to manipulate the dielectric object with light[1]. The object with high indexes of refraction and small size can be attracted to the focal point of a continuous laser beam. In most optical tweezers experiments, the dielectric object used is polystyrene bead.

The trapping mechanism can be described in two regimes. Here we will only talk about the case when the wavelength (λ) of the light is much smaller than the size of the particle (d) since it is applied in most optical tweezers experiments including our setup. Cartoon representation of the mechanism is shown in Fig. 1.1. When the light hits the bead, the bead will gain momentum, resulting in a force with direction shown in the Fig. 1.1A. When the bead is in the center of a focused laser, the bead will be influenced by three types of force due to the reflection of the light, the refraction of the light, and scattering (Fig. 1.1B). Along the direction of light propagation, the momentum change of reflected light will result in a force pushing the bead away from the laser beam; the momentum change of the refracted light will lead to a force attracting the bead towards to the laser beam; the scattering force led by the photons will push the bead away. In a stable

optical trapping system, all the forces exerted on the bead will be balanced and keep the bead at a constant position. When the bead is moved away from the focus of the laser beam, the force F_1 resulted from the intense beam is larger than the force F_2 resulted from the weak beam (Fig. 1.1C). As a result, the bead will be under the attraction towards the focus of the laser beam. For small displacements (< 150 nm) of the trapped object from its equilibrium position the force is linearly proportional to the displacement[2]. With the development of optical tweezers, two laser beams are usually used to counteract the scattering force. The restoring force acts as Hookean force and is proportional to the displacement and the spring constant. The spring constant of the optical trap depends on how tightly the laser is focused, the laser power, and the polarizability of the trapped object[3]. It is possible to measure the change in the momentum flux of the light beam by measuring the angular intensity distribution of the laser light as it enters and leaves the trap such that the external force can be calculated directly[4]. With the signals collected by the detector, the applied force can be decomposed into two expressions,

$$F_x = \frac{D_x R_D}{c \Psi_{R_L}}, \tag{1.1}$$

$$F_y = \frac{D_y R_D}{c \Psi_{R_L}}, \tag{1.2}$$

where R_D is the detector's half-width, R_L is an effective radius for the lens, c is the speed of the light, D_x and D_y are two difference signals acquired by dual-axis detector, Ψ is the silicon detector's responsivity. Through the above expressions, the force on the two directions can be calculated directly from the signal obtained from position-sensitive photodetector. Alternatively, the fact that the overstretched state of DNA occurs at ~65 pN is usually used for the force calibration of optical tweezers[5]. Since the restoring force is proportional to the displacement of the bead, 65 pN and measured bead's position can be used to determine the spring constant of the

optical tweezers. In a typical optical tweezers experiment, two ends of the target protein or DNA molecule of interest are connected to the anti-digoxigenin coated bead and the streptavidin coated bead such that the force applied to the bead can be exerted on the molecule indirectly and the non-specific interactions between two beads are effectively avoided.

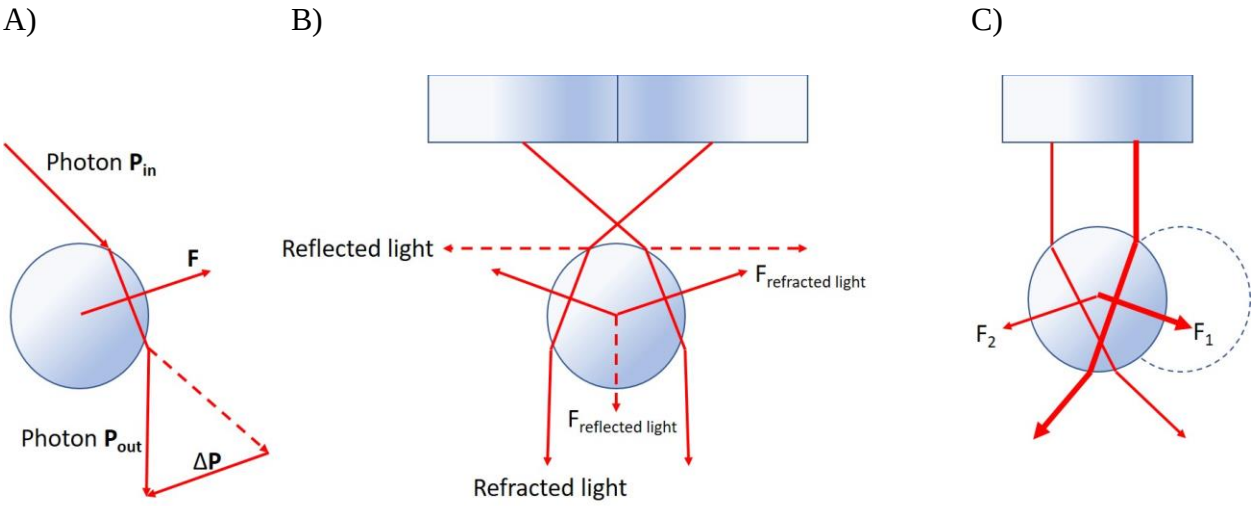


Figure 1.1 Diagram showing the ray optics to explain how the light exerts force on the bead. The intensity profile denotes the intensity of the light source from left to the right. A) When a photon hits the bead, the lost momentum of the photon will exert a force on the bead. B) When the bead is in the center of focused laser, all the forces exerted on the bead will be balanced and the bead will be kept in the center position. C) When the bead is moved away from the focus of the laser beam, the force F_1 resulted from the intense beam is larger than the force F_2 resulted from the weak beam. Therefore, the bead will be under the attraction towards the focus of the laser beam.

1.2 Data acquisition and processing

As the molecule of interest is captured between two beads, we are able to indirectly manipulate the molecule by changing the applied force. This is an outstanding advantage of optical tweezers. There are four modalities to manipulating the force: a) force clamp, where the applied force is kept constant and the extension of the molecule is either constant or fluctuates between two values; b) constant position, where the position of the trap is unchanged and the fluctuations of extension and

force are recorded; c) constant pulling velocity, where the position of the trap changes linearly d) force ramp, where the applied force changes linearly. The responses of the molecule can be categorized into two major behaviors, equilibrium and non-equilibrium. In this thesis, we will discuss more about mode a) and c) which were used in our optical tweezers experiments. The sampling frequency of the mini-Tweezers we used is 1000 Hz. With the help of high frequency acquisition, the sampling frequency can be up to 30000 Hz. Each raw data point contains the information of the position of the trap and the applied force.

1.2.1 Data acquisition and processing using constant force protocol

With the constant force protocol, the molecule of interest is maintained between two spherical beads at a constant applied force using a feedback loop. The extension of the molecule and force acting on the molecule are recorded as a function of time. In the simplest case, the molecule will fluctuate between two states, folded state and unfolded state (Fig. 1.2). The lifetimes at two state can be extracted. The distribution of the lifetimes can be fitted with a single exponential curve[6] (Fig. 1.2). The probability density (*P(t)*) of a given lifetime or dwell-time (*t*) can be fitted to the equation:

$$P(t) = k \exp(-kt), \tag{1.3}$$

where k is the unfolding/folding rate constant at the constant applied force. With optical tweezers, unfolding/folding rate constant at a given force can be obtained by observing hopping of the molecule at difference forces. The unfolding/folding rate constant at zero force is extrapolated with Bell-Evans model or other related models which will be discussed with more details in the following pages.

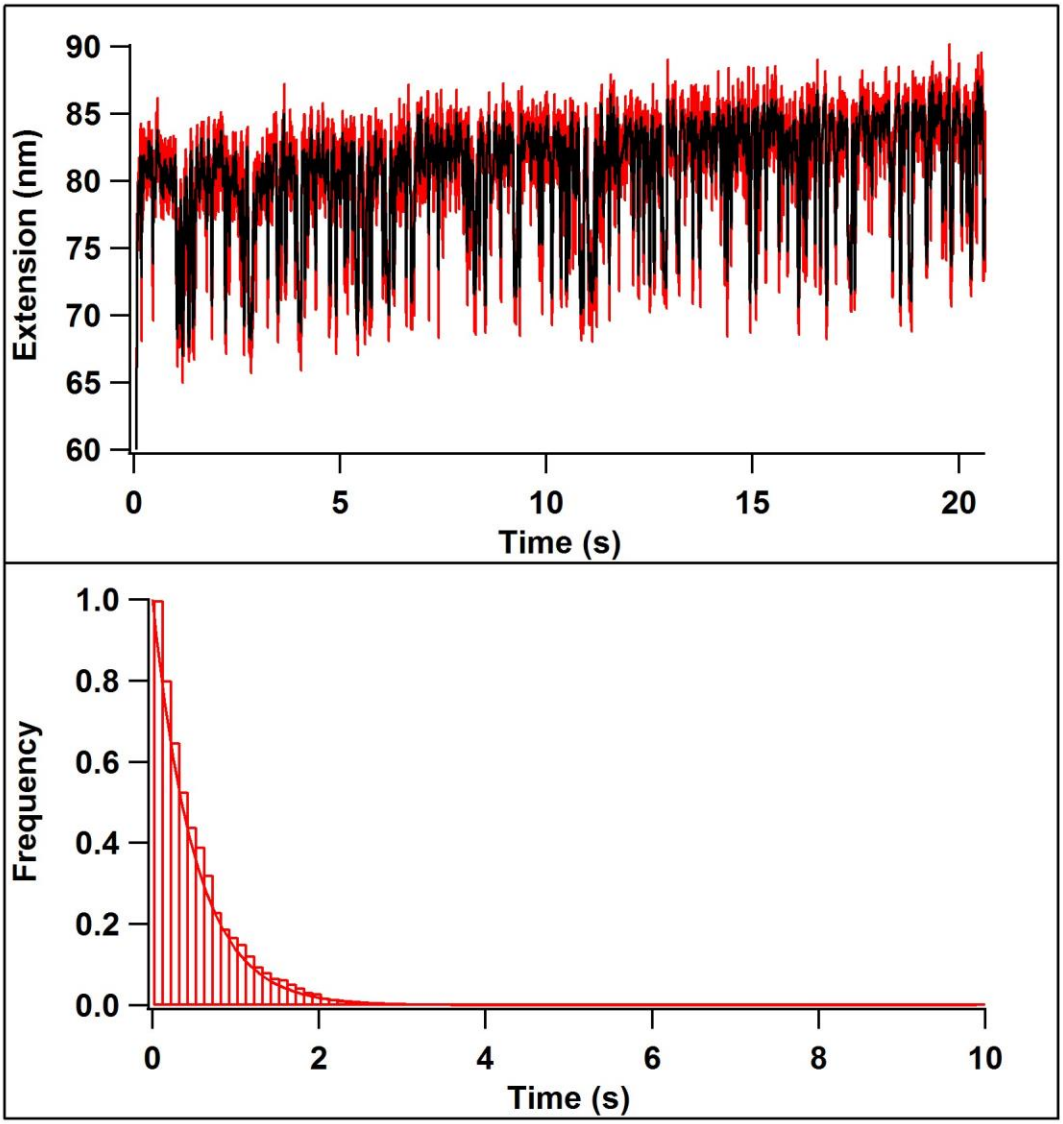


Figure 1.2 Sample extension-time trace (up) and distribution of lifetime at unfolded state (down). At a constant external force, the molecule is fluctuating between the unfolded state and the folded state. The distribution of the lifetimes at unfolded state conforms a single exponential fit.

In another case, protein molecules, such as immunoglobulin modules[7,8] and ubiquitin modules[9], may exist and function in tandem. To investigate the unfolding and folding dynamics of protein, polyprotein with tandem repeated domains is usually constructed. While maintaining force exerted on the polyprotein molecule constant, the molecule's responses result in staircase-like extension-

time curve (Fig 1.3) assuming each domain behaves independently. Every stair denotes the unfolding event of a single domain. As for polyprotein molecules, both dwell-time analysis[10] and pseudo-dwell-time method[11] can be used to extract the dynamic information.

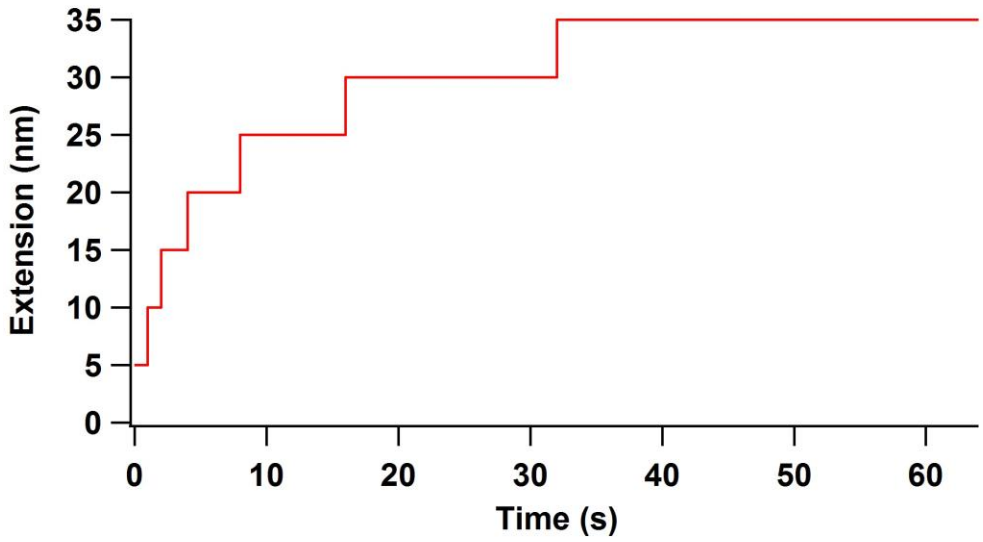


Figure 1.3 Sample extension-time curve for demonstration of force-clamp experiment. Shown is an example of extension-time curve for a polyprotein with six repeated domains. Each stair denotes the unfolding event of a single domain.

1.2.2 Data acquisition and processing using constant pulling speed protocol

With the constant pulling speed protocol, the molecule of interest is pulled by moving the trap away from the micropipette at constant pulling speed. The force is recorded as a function of extension of the molecule or the position of the trap. A typical force-extension curve and a force-position curve are shown in Fig. 1.4. Trap position represents the position of the optical trap and extension represents the distance between two ends of the attached molecule. The difference between trap position and extension is the displacement of the bead between equilibrium position and actual position. As the trap moving away from the micropipette, the force starts from zero and increases smoothly. A sudden drop of the force denotes the unfolding event of the molecule. The

single molecule attachment is checked by using the force of overstretching state (~ 65pN). When the force increases to a pre-set value, the trap is moved towards to the micropipette at the same pulling speed. A sudden increase of the force denotes the folding event of the molecule. A large number of cycles of pulling and relaxing are recorded for extracting kinetic information of the molecule. Various methods and models have been developed to analyze and interpret the single molecule experiments.

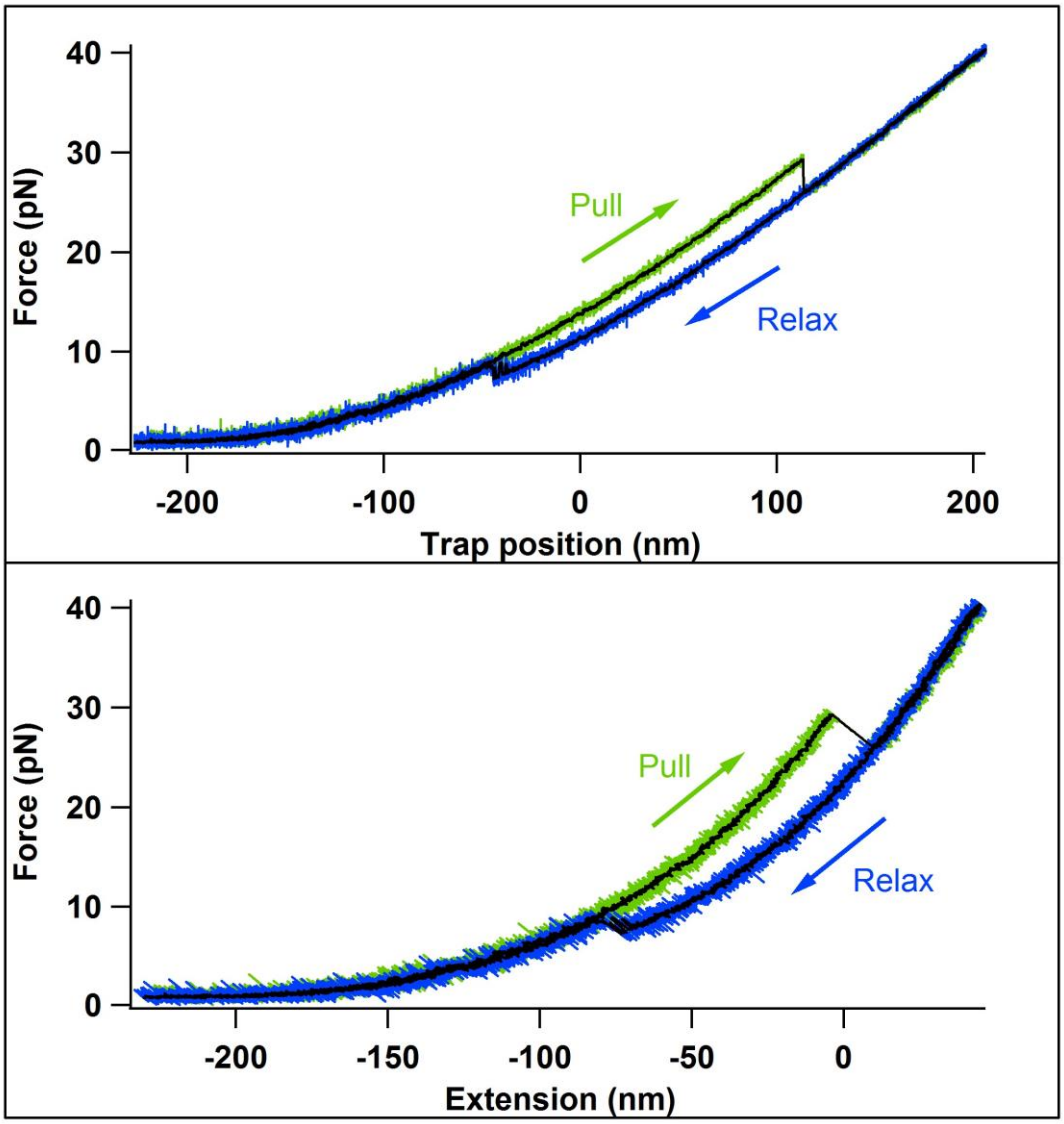


Figure 1.4 Sample trace of force-position curve (up) and force-extension curve (down) in a complete pulling-relaxing cycle. The molecule of interest in this example is Top7 molecule.

1.2.3 Models and methods used to analyze the single molecule experiments

1.2.3.1 Force extension models

In constant pulling speed mode, the force as a function of extension or trap position is recorded. The force-extension behavior of protein molecule and double-strand DNA is well described by the worm-like chain (WLC) model in polymer physics. The WLC model is suitable for describing the

manners of semi-flexible polymers. The relationship between force (F) and extension (x) can be expressed with the formula,

$$F = \frac{k_B T}{p}\left[\frac{1}{4*(1-\frac{x}{L})^2} - \frac{1}{4} + \frac{x}{L}\right], \quad (1.4)$$

where k_B is Boltzmann's constant, p is persistence length of the polymer, T is temperature in Kelvin, L is contour length of the polymer. The contour length of the molecule is the maximum possible distance between two ends of the molecule. At zero force, the molecule exists with stable conformation. When the external force is applied and keeps increasing, the molecule is extended following the above formula. Bustamante et al. successfully interpreted the entropic elasticity of λ-phage DNA with WLC model in 1994[5]. The WLC model is widely used in atomic force microscope and optical tweezers to determine the integrity of the protein's structure and the fit of protein's force-extension curves.

In a typical optical tweezers experiment, the protein molecule of interest is connected to two DNA molecules. The persistence length of DNA molecule is ~50 nm and that of protein molecule is ~0.8 nm[12,13]. Although the contour length of the molecule keeps unchanged, the elastic response of DNA cannot be ignored. The stretch modulus, K[1], is introduced to account for the elasticity of the DNA[14]. After several approximations, the interpretation formula of force-extension curve is given by,

$$F = \frac{k_B T}{p}\left[\frac{1}{4*(1-\frac{x}{L}+\frac{F}{K})^2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K}\right]. \quad (1.5)$$

[1] The stretch modulus, K, has a unit of pN.

Thus, a combination of WLC model and eWLC model is used to fit force-extension curves obtained by optical tweezers.

1.2.3.2 Bell-Evans model

The most widely used analysis of the single molecule experiments is based on Bell-Evans model[15,16]. In this model, the molecule is in a two-state force-induced conformational change process and a one-dimensional free energy profile along the reaction coordinate can be built. The unfolding rate constant and folding rate constant at zero force can be given by,

$$k_u(0) = A * \exp\ (\frac{-\Delta G^0_{FS,TS}}{k_B T}), \tag{1.6}$$

$$k_f(0) = A * \exp\ (\frac{-\Delta G^0_{US,TS}}{k_B T}), \tag{1.7}$$

where *A* is an Arrhenius type pre-exponential coefficient, $\Delta G^0_{FS,TS}$ and $\Delta G^0_{US,TS}$ denote the free energy change from the folded state to the transition state and the free energy change from the unfolded state to the transition state at standard condition, k_B is Boltzmann's constant, and *T* is temperature in Kelvin. The force is applied along the reaction coordinate and add an energy term to the energy barrier. The unfolding rate constant at a given force F can be expressed by,

$$k_u(F) = A * \exp\left(\frac{-(\Delta G^0_{FS,TS} - F*\Delta x_u)}{k_B T}\right) = k_u(0) * \exp(\frac{F*\Delta x_u}{k_B T}). \tag{1.8}$$

Likewise, the folding rate constant at a given force F is given by,

$$k_f(F) = k_f(0) * \exp(\frac{F*\Delta x_f}{k_B T}), \tag{1.9}$$

where $k_u(0)$ and $k_f(0)$ are unfolding rate constant and folding rate constant at zero force, unfolding distance and folding distance, Δx_u and Δx_f, represent the distance between the minimum and

maximum of the barrier along the reaction coordinate. For most proteins with two-state conformational change, the unfolding rate constant and folding rate constant at zero force can be extrapolated by plotting $ln(k_u(F))$ versus F and $ln(k_f(F))$ versus F. The unfolding distance and folding distance can be calculated via the slope of the fitted linear line. However, the assumption that unfolding distance and folding distance are independent of the external force doesn't always hold true as more experiments were carried out. In other words, the relationship between the unfolding/folding rates and force is not strict linear. Evans et al. extends Bell-Evans model with Kramers' theory to account for the disobedience and limitation of Bell-Evans model[16]. Eq. (1.2) is implemented by,

$$k_u(F) = k_u(0)g(F)\exp(\frac{F*\Delta x_u}{k_BT}), \tag{1.10}$$

where $g(F)$ is in terms of applied force F and depends on the deformation of energy landscape by F.

1.2.3.3 Theoretical calculations underlying data interpretation for force-ramp experiments

For force-clamp mode, it is straightforward to extract force-dependent unfolding rate constant and folding rate constant from the distribution of dwell-time intervals. As for force-ramp mode, underlying dynamic information consists in force-extension curves. How to extract kinetic information from force-ramp experiments is of vital importance. Since the theory applied to unfolding force analysis is also applicable to folding force, we will only discuss analysis of unfolding force distribution for simplicity. Folding kinetic information can be deducted by simply

changing the subscript of the variables from u to f (such as $\Delta x_u \rightarrow \Delta x_f$). The unfolding probability, *P*, can be expressed by,

$$dP_u = k_u(t) * [(1 - P_u(t)]dt. \tag{1.11}$$

For force ramp mode, the applied force F changes linearly with time *t*. Therefore, the equation after integrating will be,

$$P_u(\mathrm{F}) = 1 - e^{-\frac{k_u(0)k_BT}{a\Delta x_u}(e^{\frac{F\Delta x_u}{k_BT}}-1)}, \tag{1.12}$$

where *a* is the loading rate. Hence, the probability density can be deducted as follow,

$$\frac{\mathrm{d}P_u}{dF} = \frac{k_u(0)}{a} * e^{\frac{F\Delta x_u}{k_BT}} * e^{-\frac{k_u(0)k_BT}{a\Delta x_u}(e^{\frac{F\Delta x_u}{k_BT}}-1)}. \tag{1.13}$$

Through the raw data, the rupture force distribution can be built. After integrating, the unfolding probability distribution (P_u) is obtained. Eventually, the unfolding distance and spontaneous unfolding rate constant can be determined through curve fitting.

1.2.3.4 Dudko method

For a two-state conformational change, the survival probability *S(t)* represents the probability that the system is still in folded state (FS). The survival probability *S(t)* can be related to the unfolding probability *P(t)* by, *S(t)* = 1 – *P(t)*. Assuming denaturation of the molecule is a first order reaction and ignoring the fluctuation of the force[17]. *S(t)* is given by,

$$-\frac{dS}{dt} = \frac{S(t)}{\tau(F(t))} = k(t)S(t), \tag{1.14}$$

where *F(t)* is the time-dependent force, and *τ(F(t))* is the average lifetime at a given force *F(t)*. Since *S(t)* can be related to the probability distribution of rupture force *p(F)* by,

$$-dS = p(F)dF = \frac{S(t)*dt}{\tau(F(t))}, \tag{1.15}$$

the relationship between probability distribution of rupture force and distribution of average lifetime can be expressed by,

$$p(F) = \frac{\exp(-\int_0^F [\dot{F}(f)\,\tau(f)]^{-1} df)}{\dot{F}(F)\tau(F)}, \tag{1.16}$$

where $\dot{F}$ is loading rate of the force. Similarly, the distribution of lifetimes can be expressed in terms of probability distribution of rupture force and the loading rate of the force by,

$$\tau(F) = \int_F^\infty p(f)df/[\dot{F}(F)p(F)]. \tag{1.17}$$

Due to the fact that the molecule system includes not only protein molecule but also flexible linker and DNA handles, the loading rate of the force is not independent of force F. Dudko et al. described a method to transform rupture force histograms into average lifetimes at force F according to Eq. (1.17) with the assumption that the rupture can be regarded as escape from a deep one-dimensional free-energy well[18]. Consider a rupture force histogram starts from F_0 and ends at F_N. F_N is given by,

$$F_N = F_0 + N\Delta F, \tag{1.18}$$

where N is the number of the bins and ΔF is the width of each bin. The interpretation of rupture force histograms into $\tau(F)$ is given by,

$$\tau\left(F_0 + \left(k - \frac{1}{2}\right)\Delta F\right) = \frac{(\frac{h_k}{2} + \sum_{i=k+1}^{N} h_i)\Delta F}{h_k \dot{F}(F_0 + \left(k - \frac{1}{2}\right)\Delta F)}, \tag{1.19}$$

where k is a positive integer (1, 2, …, N), $h_i = C_i/(N_{tot}\Delta F)$ with the number of count in the ith bin C_i and total number of counts N_{tot}. With resulted $\tau(F)$, the unfolding rate at F can be concluded

and used to derive the kinetic parameters of the protein molecule. They validated this method with the AFM data of immunoglobulin-like domain 4, ddFLN4[18].

1.2.3.5 Monte Carlo simulation method

In addition to exact the kinetic information directly from the data, the Monte Carlo simulation based on two-state conformational change reaction has been widely used to explore the dynamic parameters of single protein molecule. A huge advantage of the Monte Carlo simulation method over the direct analysis is that many assumptions in direct fitting of the data can be avoided. A comprehensive explanation, procedure, and discussion of the Monte Carlo simulation designed for single molecule AFM experiments was presented by Yang et al. in 2010[19]. To make the procedure applicable for optical tweezers experiments as well, some minor modifications need to be made. Consider a pulling experiment for a single domain molecule using constant pulling velocity mode of optical tweezers and the extension of the molecule starts from zero. For a time interval Δt, the corresponding extension change of the molecule is $\Delta x = v*\Delta t$. The applied force $F(t)$ at time t can be calculated with Eq. (1.5) which describes the relationship between the force and the extension and unfolding rate constant at force $F(t)$ can be calculated with Eq. (1.8). The probability of occurrence of an unfolding event is given by,

$$P_u = k_u(F(t))\Delta t, \tag{1.20}$$

This probability is compared to a random number r generated from a continuous uniform distribution on (0,1) to determine whether the unfolding event would happen. Once the unfolding event occurs, a new cycle will be operated. With repeated cycles, a histogram of unfolding force distribution can be drawn. However, the resulted histogram is based on two input value, unfolding rate constant at zero force $k_u(0)$ and unfolding distance Δx_u, which are used to calculate $F(t)$ in Eq.

(1.8). Therefore, those two dynamic parameters can be found by varying them until the experimental unfolding force distribution is fully recovered by the simulated histogram.

1.2.3.6 Oberbarscheidt method

A simple and model-free method to extract the force-dependent unfolding/folding rates was presented by Oberbarnscheidt et al. in 2009[20]. This method is used in the data analysis involved in this thesis. Recall that the distribution of the lifetimes follow an exponential decay which is shown in Eq. (1.3). The probability of getting N survival times is given by,

$$\mathrm{P}\left(\{\mathrm{t}_{1,}, \dots, \mathrm{t}_{\mathrm{N}}\}, \mathrm{k}_{\mathrm{u}}\right) = \prod_i P(\mathrm{t}_{\mathrm{i}}, \mathrm{k}_{\mathrm{u}}), \tag{1.21}$$

Consider the probability is in terms of k_u, the most possible unfolding rate constant is given by,

$$\mathrm{k}_{\mathrm{u}} = N/\sum t_i, \tag{1.22}$$

Assuming M curves have been collected using force-ramp mode of optical tweezers, the force extension curve (FEC) is divided into tiny pieces and each piece can be regarded as operating at constant force mode. Eq. (1.22) can be related to Eq. (1.20) by,

$$P_u = k_u(F)\Delta t = N(F)/M(F), \tag{1.23}$$

where *N(F)* is the number of unfolding events observed, *M(F)* is the total number of the chose curves, Δt is the amount of time that each tiny piece takes. Thus, the unfolding rate constant at force F can be extracted directly with the equation,

$$k_u(F) = N(F)/(M(F) * \Delta t), \tag{1.24}$$

1.3 Advantages and drawbacks of optical tweezers

Since the stiffness of the optical tweezers system (0.005-1 pN/nm) is typically lower than that of AFM cantilevers (10-10^5 pN/nm), optical tweezers can generate much lower forces (0.1 pN to 150 pN) on the particle. Besides, optical tweezers with high stability and accurate motion measurements make it extremely sensitive to the changes of force and position. Directly manipulating the force and the position makes the optical tweezers well suited for investigating the biophysical characteristics of both DNA/RNA and protein molecules. For single molecule AFM experiments, a molecule with tandem repeated domains is usually constructed to obtain FECs with multiple sawtooth-like force peaks. Compared with AFM, it is convenient to precisely manipulate the extension of the molecule with optical tweezers such that a molecule with a single domain is also able to be studied. The ability of precisely manipulating the extension of the molecule also makes it possible to use the single molecule optical tweezers technique to investigate the folding process of both DNA/RNA and protein molecules.

However, advantages of optical tweezers also bring drawbacks. Although a force between 0.1 to 150 pN can be achieved, the likelihood of detachment of the molecule would increase when the force is higher than 50 pN. An approach to overcome this limitation was proposed in a recent study[21]. With AFM, He et al. directly observe the reversible two-state unfolding and folding of an α/β protein by lowering the pulling velocity and increasing the stability of the system by using gold-free cantilevers[22]. Same biophysical results for this α/β protein were extracted with the optical tweezers by Lei et al.[21] In that study, a trajectory of a polyprotein is presented to make it possible to investigate proteins with high stability with optical tweezers. Nevertheless, investigating the protein with complex unfolding/folding behaviors and high stability by using optical tweezers still

need more effort. Another drawback of the optical tweezers technique lies in that the optical trap would easily trap tiny particles other than target bead. Thus, the requirement on the high purity of the sample needs to be satisfied. Otherwise, the FECs with sawtooth-like force peak or staircase-like extension-time curves are difficult to acquire. In addition, there are also other drawbacks in terms of the instrument such as influences of supply power fluctuations and potential optical damages to biomacromolecules[3].

1.4 Applications of optical tweezers

There are mainly three types of assays based on optical tweezers (Fig. 1.5), including interaction assays, tethered assays, and dumbbell assays. The experiments referred in this thesis is based on dumbbell assays. The only difference between our assays and dumbbell assays is that one end of the molecule is attached to a bead held by a micropipette instead of an optical trap. As a single molecule force spectroscopy technique, optical tweezers technique has been widely used in extracting kinetic information[23-26], establishing profiles of energy landscape[25,27] and exploring unfolding/folding pathways of DNA/RNA and protein molecules[24,28-32]. Many proteins with relatively small size (less than 200 amino acids) have been well studied by optical tweezers technique, such as calmodulin[24,33] and the prion protein[30]. Although investigating the large protein is believed to be difficult because of the aggregation of domains and existence of multiple intermediate states, optical tweezers experiments have been carried out to study biophysical properties of large protein molecules such as molecular machine Hsp90 with 1418 amino acids (aa)[29]. The development of the instruments of single molecule technique over the past decade allowed researchers to investigate not only the unfolding processes of nucleic acids or protein molecules but also their folding processes at single molecule level[34]. Due to the fact that an

intermediate state often exists with a short average lifetime, it is difficult to directly detect the intermediate state during conformational changes using traditional bioanalytical methods. However, optical tweezers have proved to be ideally suitable to directly observe the single or multiple intermediate states during transitions between native and denatured states[29,33,35]. In summary, optical tweezers technique has shown prominent advantages in exploring the structure information of nucleic acids and proteins because of its high sensitivity and resolution.

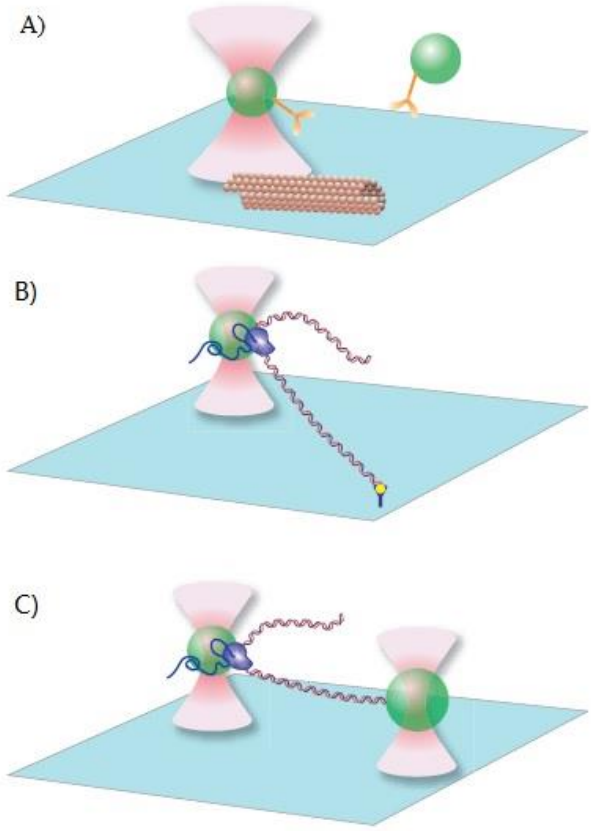


Figure 1.5 Schematics of optical tweezers-based assays. A) Interaction assays: one kinesin-coated polystyrene bead (green) is trapped by an optical trap; the bead is moved towards to the microtubule attached on the surface of the mini-chamber; as the kinesin moves along the microtubule, the generated force is recorded. B) Tethered assays: one RNA polymerase-coated bead (green) is trapped by an optical trap and a DNA template is attached to the surface of the mini-chamber; the force generated by transcription process can be recorded. C) Dumbbell assays: similar to tethered assays, the free end of the DNA template is attached to a trapped bead instead of the surface of the mini-chamber. Reprinted from [3], Copyright © 2008, with permission from Nature Publishing Group.

Chapter 2: Mechanical unfolding/folding of Top7 revealed by optical tweezers

2.1 Introduction

2.1.1 Development on study of Top7

Top7 is a 92-residue protein *de novo* designed by Brian Kuhlman and Gautam Dantas in David Baker's laboratory in 2003[36]. This computationally designed protein, consisting of five paralleled β strands and two parallel α helices, folds with the same structure as designed (Fig. 2.1). Its novel globular protein folding pattern not yet observed in nature is extremely stable. In native state, the non-covalent interactions (hydrogen bonds) between two adjacent β strands make them directly connected and form the shear topology[37]. These hydrogen bonds supply a high resistance to mechanical stretching and believes to be the source of Top's high mechanical stability. Classical ensemble techniques (Stop-flow kinetics, NMR spectroscopy and circular dichroism) have revealed that folding of Top7 is not in a simple all-or-none manner and three phases exist in the folding process[38]. Coarse-grained, native-centric chain model-based computational simulations predicted that Top7 has a stable folding intermediate state at which C-terminal half has completely folded and N-terminal half remains unordered[39]. Compared with other natural proteins of the similar size, folding cooperativity of Top7 is obviously lower. The natural selection might be an underlying cause of highly-cooperative folding since misfolding and aggregation of protein structure may lead to malfunction of the protein *in vivo* and even diseases[38-40]. With the intention of making it fold cooperatively, some modifications on the amino acid sequence were proposed recently based on structure-based models[41] with a lack of experimental data.

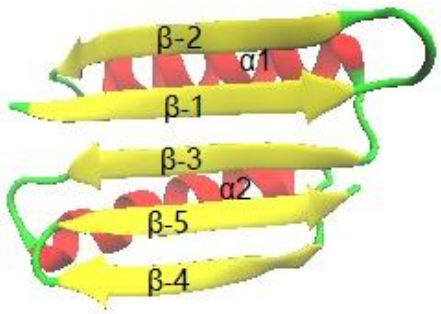


Figure 2.1 Cartoon representation of Top7 structure (PDB: 1QYS). It is composed of two α-helices (red: α1, α2) and five β-strands (yellow: β1-β5). The figure is generated with VMD[42].

2.1.2 Single molecule force spectroscopy of Top7

Using atomic-force microscopy, Li's group investigated the mechanical unfolding of wild-type Top7 (wt-Top7), as well as its high mechanical stability and the origin of the mechanical stability[43]. In that study, a polyprotein of $(GB1)_4$-$(Top7)_2$-$(GB1)_4$ was constructed (Fig. 2.2A). Single molecule experiment using constant pulling speed protocol resulted in FECs with multiple sawtooth-like peaks (Fig. 2.2B). Based on steered molecular dynamics (SMD) simulations, two possible mechanical unfolding pathways[43] were proposed, which will be discussed in details later in this thesis. The previous results paved the road to fully understand the biophysical properties of Top7.

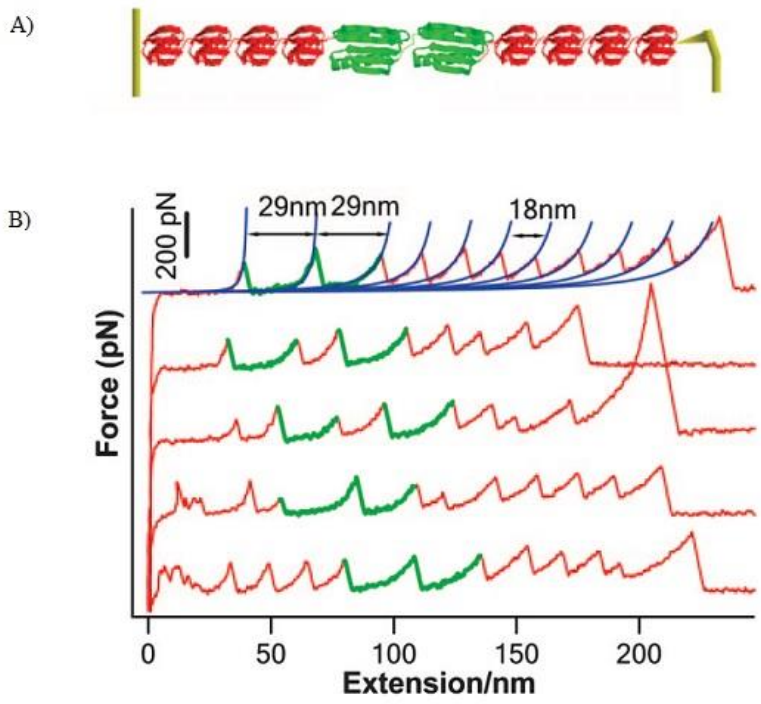


Figure 2.2 Schematic of polyprotein construct and FECs obtained by AFM. A) Shown is a schematic of polyprotein construct $(GB1)_4$-$(Top7)_2$-$(GB1)_4$. B) Five sample FECs captured by AFM single molecule experiment using constant pulling speed protocol. Mechanical unfolding of Top7 results in ΔL_c of ~27 nm (green). Mechanical unfolding of Top7 results in ΔL_c of ~18 nm (red). WLC fitted curves are overlaid on the FECs. The figure is reprinted from [43], Copyright © 2007, with permission from National Academy of Sciences, U.S.A.

2.1.3 Aim of research studies in this thesis

Although the unfolding of Top7 has been well studied by AFM, the folding process of Top7 remains unclear to us. The high spatial and time resolution of the optical tweezers makes it possible to probe the folding behavior. In the study of this thesis, we combined optical tweezers and protein engineering to further investigate the unfolding/folding pathways and kinetic information of wt-Top7. The main goal is to ensure whether the folding intermediate exists. The truncated C-fragment was also constructed and pulled using constant pulling speed mode. The folding cooperativity of a cooperative mutant (cm-Top7) was tested using optical tweezers. Both constant

force protocol and constant pulling velocity protocol with optical tweezers are applied depending on the properties of protein molecules.

2.2 Methods and material

2.2.1 Protein Engineering

The gene of wt-Top7 was received as a kind gift from David Baker, and the restriction sites of BamHI (GGATCC), BglII (AGATCT) and KpnI (GGTACC) were added using standard PCR. The insert of wt-Top7 was obtained through digestion with BamHI and KpnI. The synthetic gene of Top7 cooperative mutant was purchased from GenScript, Inc, which contains the restriction sites of BamHI and BglII followed by KpnI. Gene of Top7's C-fragment was acquired via regular PCR technique using designed forward primer (IDT) carrying the restriction site of BamHI and the reverse primer containing the restriction site of KpnI. The vector of pQE80L (QIAGEN, Valencia, CA) was modified to pQECC vector with DNA hybridization and ligation. The pQE80L vector was digested with BamHI and HindIII to obtain the vector. Two complementary ssDNA were purchased from IDT and formed the dsDNA by using DNA hybridization. The dsDNA was designed to be BamHI-Cys-BglII-Gly-KpnI-Cys-Stop-SacI-Stop-HindIII. The dsDNA sample was digested with BamHI and HindIII to obtain the insert. After ligation, pQE80L was successfully modified to pQECC vector. All the inserts of the target gene, digested with BamHI and KpnI, were sub-cloned into pQECC expression vector (digested with BglII and KpnI). All proteins were overexpressed in Escherichia coli strain DH5α and purified by using Co^{2+} affinity chromatography with TALON His-Tag purification kit. The purity of the protein is confirmed by running sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The purified protein was kept

in phosphate buffered saline (PBS, pH 7.4) at 4 ºC. The protein sequence of the three constructs can be found in Appendix **B**.

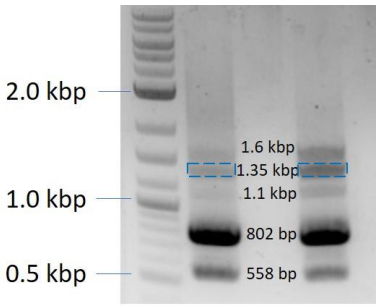
2.2.2 Synthesis of oligonucleotide-coupled protein

Two types of modified GC-rich ds-DNA were synthesized using regular PCR protocol. One of them contains 5' biotin moiety and 5' amino group with a length of 802 base pairs, and the other one carries 5' digoxigenin moiety and 5' amino group with a length of 558 base pairs. The sequence of the primer can be found in Appendix **A**. Plasmid of pGEMEX-1, widely used in the preparation of the DNA handles, was used as the template in the PCR. Two types of DNA were mixed with the 1:1 ratio and reacted with SMCC (succinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate, SIGMA-ALDRICH) overnight at room temperature. In this step, amino group from the DNA reacted with the ester group from SMCC via NHS ester reaction. The purified DNA handles samples (~3 μM) were kept in the Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4) at -20 ºC.

The target protein (20 μL, 1 mg/mL) was reduced with excess DTT (2 μL, 100 mM) before coupling to cleave the inter-molecular disulfide bond. Removal of DTT was carried out by letting the solution run through acetate saturated desalting columns (7 kMW, Zeba). Two cysteine residues at both terminal ends allow the protein molecule to react with the connected DNA-SMCC molecule via maleimide thiol reaction. In typical coupling experiments, protein solution (0.5 μL, 12 μM) was added to 2 μL of DNA (3 μM) handles with the 1:1 molar ratio. The mixture was incubated overnight at room temperature. The coupled DNA-protein sample may be composed of five components, including unreacted 558 bp DNA handles, unreacted 802 bp DNA handles,

resulted DNA(558)-protein-DNA(558), resulted DNA(558)-protein-DNA(802), and resulted DNA(802)-protein-DNA(802). The DNA-coupled protein sample was further purified by conducting agarose gel electrophoresis. Five corresponding bands can be found on the gel image (Fig 2.3A). The desired construct of protein with two DNA handles was screened and cut from agarose gel followed by purifying with Gel Extraction Kit (QIAquick). The biotin and digoxigenin at two ends of the coupled DNA-protein molecule make it be able to connect to the streptavidin-coated bead and anti-digoxigenin-coated bead through strong non-covalent interactions. The cartoon representation of the whole experimental setup is shown in Fig. 2.3B.

A)



B)

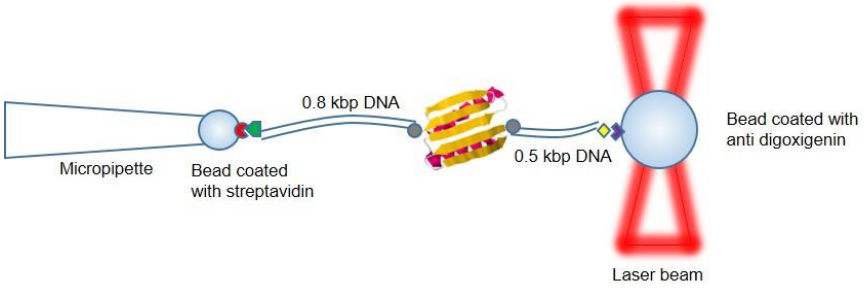


Figure 2.3 Agarose gel image of coupled DNA-protein product and the experimental setup used in optical tweezers experiments. A) Five bands can be seen on the agarose gel image. The desired band is circled with blue dash box. B) One cysteine is added to both N- and C- terminal of Top7. DNA handle is attached to the polystyrene bead with non-covalent interactions. (Red: streptavidin; green: biotin; yellow: digoxigenin; purple: anti-digoxigenin.) DNA handles modified with amino acid groups and proteins are connected with SMCC (grey) through NHS ester reactions and maleimide reactions.

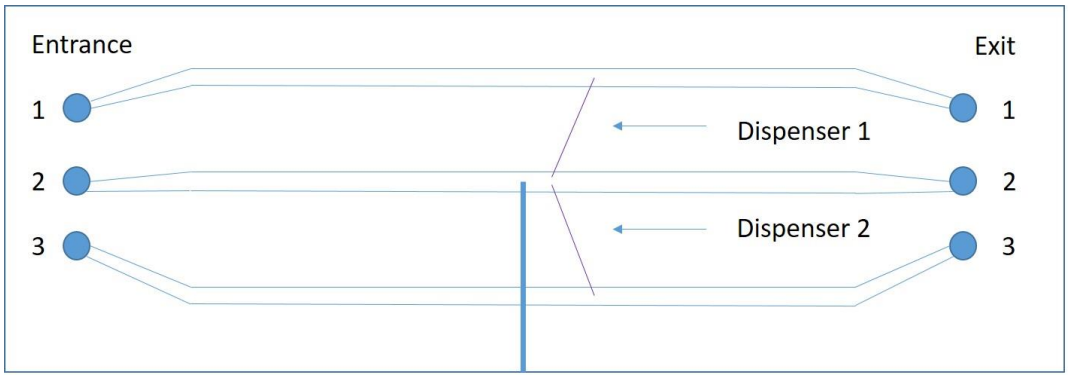
2.2.3 Single molecule optical tweezers experiments

Prior to the optical tweezers experiment, the DNA concentration of the DNA-coupled protein solution was measured with NanoDrop-1000 UV-Vis spectrometer (Thermo Fisher Scientific, Wilmington, DE) and diluted to 10 nM with deionized water. 2 µL of the DNA-coupled protein sample was coupled with 4 µL suspension of anti-digoxigenin coated beads (0.5% w/v, 2 microns, Spherotech Inc.) for 30 minutes and diluted into a syringe with 3 mL Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4). A suspension of 0.5 µL of streptavidin coated beads (1% w/v, 1 micron, Spherotech Inc.) was diluted into a syringe with 3 mL Tris buffer as well. Single molecule optical tweezers experiments were performed with a miniTweezers (Steve Smith, University of California, Berkeley). Two bead solutions were injected into separate channels in the miniTweezers chamber (Fig. 2.4A).

During the optical tweezers experiment, the streptavidin beads solution was injected into the chamber via Entrance 3 and the bead would enter the middle channel through Dispenser 2. The laser beams were manipulated to trap a single streptavidin bead and moved it onto the micropipette in the chamber. A syringe connected to the micropipette was used to ensure the immobilization of the streptavidin bead. The anti-digoxigenin beads solution was injected into the chamber via Entrance 1 and the anti-digoxigenin bead could be found at the end of Dispenser 1. The laser beams were used to trap the single anti-digoxigenin bead. The anti-digoxigenin bead was moved towards the streptavidin bead to ensure intact connection. The single molecule attachment was confirmed by observing the existence of the force plateau at ~65 pN and the length of the plateau in force-position curves. The whole experiment was performed in Tris buffer (20 mM Tris, 150 mM NaCl,

pH 7.4). A CCD camera was used to monitor the chamber. An image (Fig. 2.4B) was captured which shows the experimental condition inside the chamber.

A)



B)

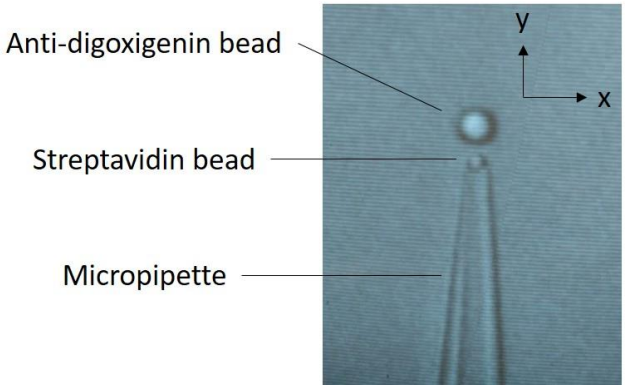


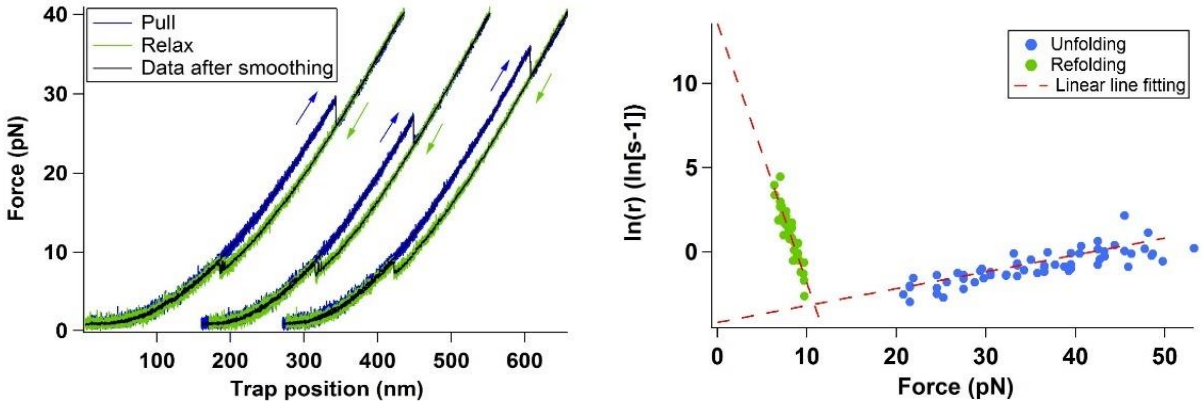
Figure 2.4 Cartoon representation of the microfluid chamber and image of the screen captured during the optical tweezers experiment. A) The anti-digoxigenin bead solution is injected into channel 1through the entrance and enter the channel 2 through dispenser 1. The streptavidin bead solution is injected into channel 3 through the entrance and enter the channel 2 through dispenser 2. Channel 2 is loaded with Tris buffer. The micropipette is in the middle of the microfluid chamber and used to suck the streptavidin bead. B) In typical optical tweezers experiments, a streptavidin bead is sucked onto the micropipette. The anti-digoxigenin bead is aligned well with the micropipette in the Y direction.

2.3 Results and discussion

2.3.1 Unfolding and folding of wt-Top7

To investigate the mechanical unfolding and folding of wt-Top7, we stretched the monomer of Top7 from its two termini at a constant pulling velocity of 20 nm/s with optical tweezers. Cycles of pulling and relaxing the DNA-coupled wt-Top7 molecule at single molecule level resulted in force-position curves shown in Fig. 2.5A. To further investigate the unfolding and folding kinetics, unfolding and folding rates were extracted using the Oberbarnscheidt method[20] and the fit of Bell's model was applied while assuming that it was a simplified two-state process (Fig. 2.5B). The resulting unfolding rate, folding rate at zero force, unfolding distance and folding distance were in great agreement with values reported in previous work of another research group[44]. The equilibrium force was determined to be ~12 pN which was also consistent with previously published work[44].

A) B)



C)

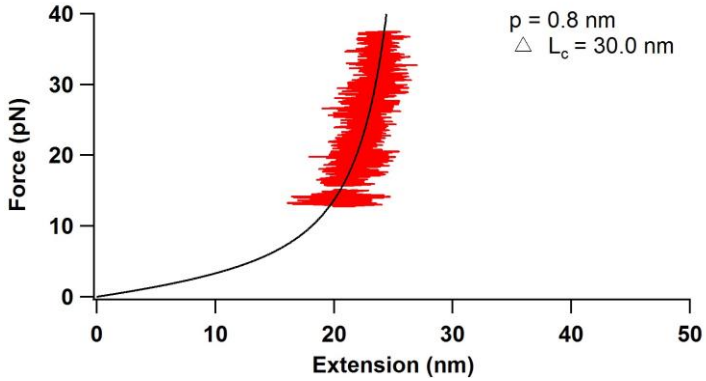
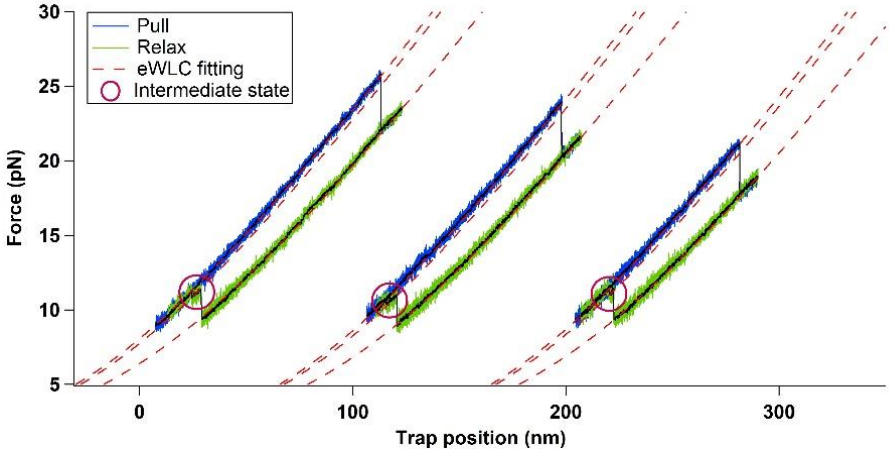


Figure 2.5 Wt-Top7 unfolds and folds at the constant pulling speed mode. A) Sample traces obtained by pulling with a constant velocity (20 nm/s). B) Force-dependent unfolding rates and folding rates under the assumption of a two-state unfolding/folding process. Δx_u and Δx_f were determined to be (0.384 ± 0.035) nm and (6.345 ± 0.440) nm, respectively. Unfolding rate and folding rate were determined to be (0.0151 ± 0.0047) s^{-1} and (7.78 ± 6.82) x 10^5 s^{-1}, respectively. C) WLC fitting of the extension differences between native state and denatured state. The parameter of persistence length (p) used in the fitting is 0.8 nm. The contour length increment is determined to be ~30.0 nm.

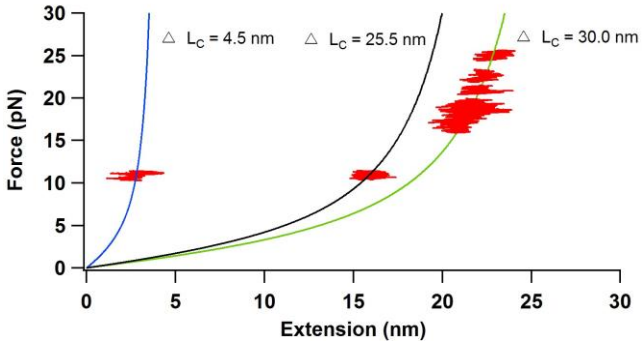
We found that fast transitions occur before the folding process. Although the unfolding force and folding force distributions became close and even overlapping as the pulling speed decreased, it was difficult to evaluate whether the near-equilibrium caused those transitions, since it was hard to ensure whether every transition reached the native state and then went back to unfolded state.

Therefore, we tried to pull the molecule at low velocities such as 5 nm/s, 2 nm/s and 1 nm/s. We were able to observe the folding intermediate directly with the traces pulling at a low velocity (Fig. 2.6A). This folding intermediate could be observed during each pulling and relaxing cycle, meaning the molecule must undergo this intermediate state before folding back to the native state. Wt-Top7 is 33.1 nm long when it is fully unfolded and extended since it has 92 residues and each residue is ~0.36 nm long. Complete unfolding of wt-Top7 should lead to ~30.1 nm of contour length increment ΔLc since the distance between two termini at the native state is ~3.0 nm. Through extensible worm-like chain (eWLC) model fitting, ΔLc between the native state and the complete unfolded state of the wt-Top7 is determined to be 30.0 nm which is in agreement with the calculated value above. Similarly, ΔLc between the native state and the intermediate state was determined to be 4.5 nm. This number is further proved by WLC fitting with the method introduced by Zhang et al.[13] (Fig. 2.6B). Small distance gap between the native state and the intermediate state supports the argument that Top7 has the state with high native-like secondary structures during folding process[45]. Although we still cannot determine the structure of the intermediate state merely with this value, it can give us some guidance on prediction about the structure of the intermediate state in the future. With high sampling frequency (30 kHz), we found that a small portion of unfolding events also showed intermediate states which had never been observed and expected before (Fig. 2.6C).

A)


Pull
Relax
eWLC fitting
Intermediate state
Force (pN)
Trap position (nm)

B)


L_C = 4.5 nm
L_C = 25.5 nm
L_C = 30.0 nm
Force (pN)
Extension (nm)

C)

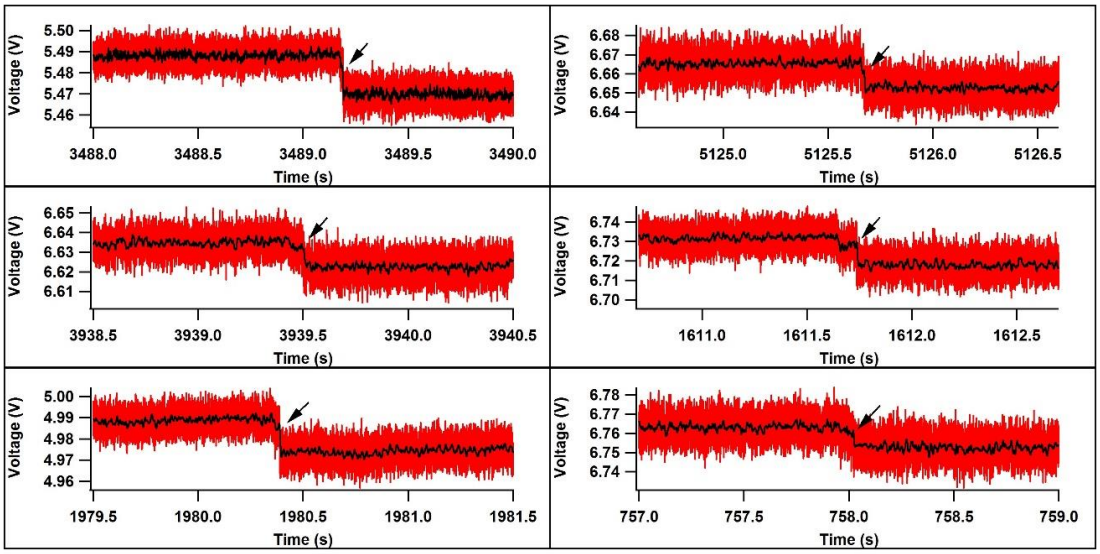

Voltage (V)
Time (s)

Figure 2.6 Wt-Top7 shows clear folding intermediate at low pulling speed. A) Sample traces obtained by pulling with a constant velocity (5 nm/s). Each folding pathway shows clear and arbitrary intermediate states. ΔL_c between the folded state and the intermediate state is fitted to be 4.5 nm with eWLC model while ΔL_c between the intermediate state and the unfolded state is fitted to be 25.5 nm. B) WLC fitting (blue, black, and green curves) of the extension differences between native state and intermediate state (left), intermediate state and denatured state (middle), native state and denatured state (right). The parameter of persistence length (p) used in the fitting is 0.8 nm. C) Six unfolding curves recorded with high sampling frequency (30 kHz) show transient unfolding intermediate state. The arrow points to the intermediate state in each trace.

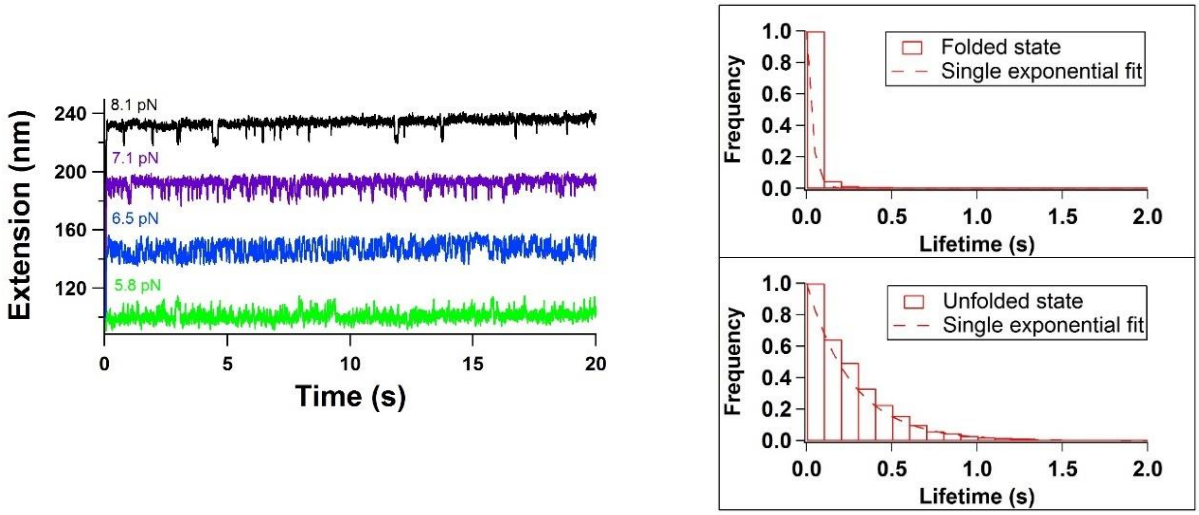
2.3.2 Unfolding and folding of Top7's C-fragment

To investigate the unfolding and folding manner of Top7's C-fragment, we constructed a GB1-Top7's C-fragment-GB1 monomer flanked by two cysteine residues. GB1 works as a fingerprint to pinpoint the existence of the protein molecule between two DNA handles. Its unfolding behavior has been well characterized by single molecule AFM experiments[43]. In optical tweezers experiments, GB1 unfolds with contour length increments ΔLc of ~18 nm and at forces higher than 20 pN. Although the C-fragment of Top7 has a similar contour length increment (49 aa x 0.36 nm/aa – 0.20 nm = 17.44 nm), the unfolding and folding processes occur at forces lower than 10 pN, which can be differentiated from the unfolding events of GB1.

Mechanically stretching resulted in two patterns of force-position curves shown in Fig. 2.8, where each scenario existed independently, can be observed successively but cannot be converted into the other one. In one pattern, two force peaks at forces higher than 20 pN represented mechanical unfolding of GB1, one equilibrium at forces between 4 pN and 8 pN represented the unfolding and folding of Top7's C-fragment (Fig. 2.8A). Due to the fact that Top7's C-fragment reached equilibrium at relatively low forces, the constant force protocol could be applied to study the kinetics of the protein. We applied the constant pulling velocity protocol first to ensure the

tethering of single molecule of monomer, and then the molecule was stretched under different forces at which GB1 domains are supposed to remain folded. As the force decreased, the displacement distribution shifted from the unfolded state to the folded state (Fig. 2.7A). Unfolding and folding rate constants at specific forces could be derived from dwell-time distribution analysis (Fig. 2.7B, C). The fact that the C-fragment of Top7 can unfold and fold independently supported the assumption regarding the structure of wt-Top7's intermediate state, at which C-terminal half (β3-α2-β4-β5) has folded and very little of N-terminal half of the protein (β1-β2-α1) has folded[38,46]. Remarkably, mechanical unfolding scenario of Top7's C-fragment has never been observed in wt-Top7. Fast transitions between the folded state and the unfolded state indicated that without the assistance of the N-fragment of Top7, the mechanical stability dramatically decreases.

A) B)



C)

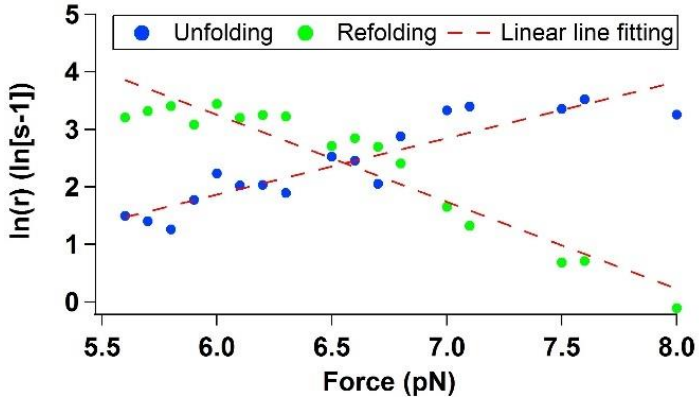
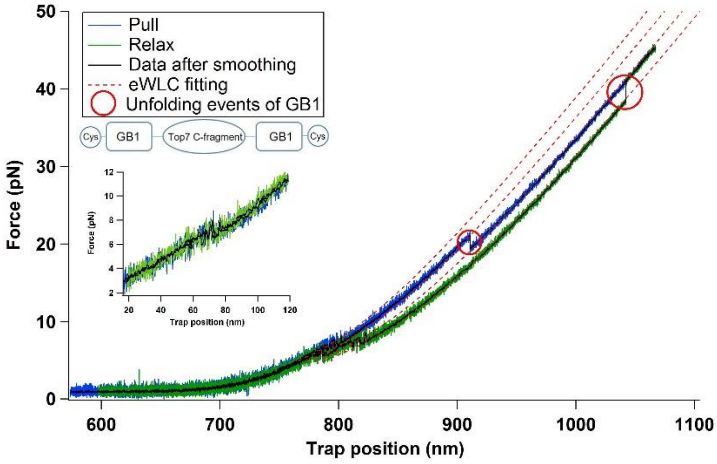


Figure 2.7 Kinetic information of Top7's C-fragment obtained from the constant force mode. A) Extension changes with time at different constant forces. B) The distribution of the lifetimes of folded/unfolded state is single exponential. C) Force-dependent unfolding rates and folding rates derived from the constant force strategy. Δx_u and Δx_f are determined to be (4.03 ± 0.46) nm and (6.22 ± 0.53) nm, respectively, and unfolding rate and folding rate are determined to be (0.0180 ± 0.0132) s^{-1} and (2.27 ± 1.93) x 10^5 s^{-1}, respectively.

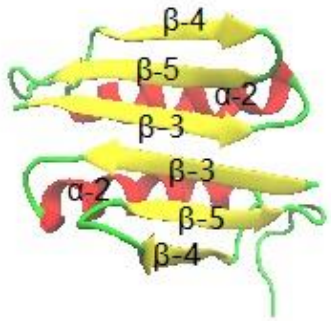
In the other pattern, the unfolding trace has one large peak with contour length increments ΔLc of ~54 nm followed by three small peaks with contour length increments ΔLc of ~18 nm (Fig. 2.8B). The possible explanation lies in the formation of homodimer since the total contour length

increment of the full molecule is ~108nm which is identical to that of a homodimer. Due to the highly symmetric sequence of wt-Top7, two single monomers of Top7's C-fragment are able to cooperate to pack mimicking wt-Top7 with the interactions between β3 stands of each monomer (Fig. 2.8C). One GB1 was introduced to both termini of Top7's C-fragment to avoid the formation of homodimer. This attempt succeeded for most of molecules which resulted in the first pattern. The reason led to the second pattern remains unknown. A possible explanation lies in two single molecules connecting with each other by disulfide bond, in which two Top7's C-fragment domains somehow pack with one GB1 domain between them with a non-native folded structure. To further avoid the formation of homodimer, we dealt the DNA-coupled protein sample with reducing agent dithiothreitol (DTT) and found out that all single molecules captured were monomers as expected and unfolding/folding pattern belonged to the first type demonstrated above.

A)



C)



B)

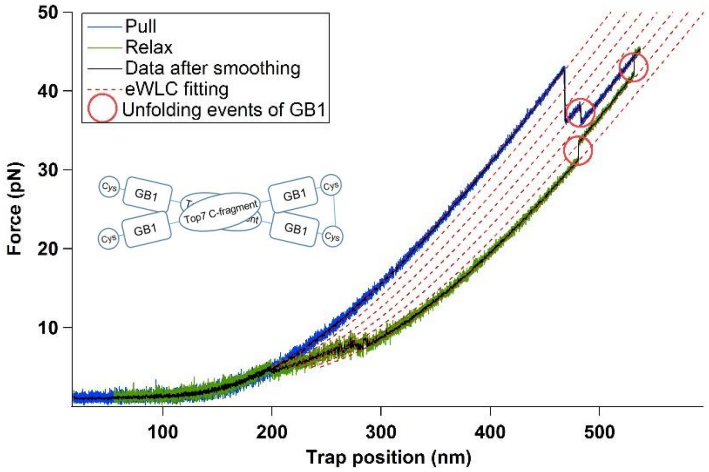


Figure 2.8 Monomer and dimer of GB1-Top7 C-fragment-GB1 shows different folding and unfolding behavior. A) Monomer of GB1-Top7 C-fragment-GB1 has three major unfolding/folding events. When the molecule is pulled, Top7's C-fragment unfolds first at around 8 pN and two GB1 domains unfold at relatively high force regions. When the molecule is relaxed, Top7's C-fragment folds first at around 7 pN and two GB1 domains fold at relatively low force (5 pN). Unfolding and folding of Top7's C-fragment show fast fluctuations between native and unfolded state at 6-8 pN (inset). B) Dimer of GB1-Top7 C-fragment has four major unfolding events, one with ΔL_c of 54 nm and three with ΔL_c of 18 nm. ΔL_c between every two adjacent eWLC fitted curves is 18 nm. C) Cartoon representation of Top7's C-fragment colored by secondary structure (PDB: 2GJH). It contains three β-sheets (yellow: β3-β5) and one α-helices (α2). Two molecules of C-fragment are able to form an exceptionally stable homodimer. The figure is generated with VMD[42].

2.3.3 Unfolding and folding of Top7 cooperative mutant (cm-Top7)

There is a difference in stability between N-terminal half and C-terminal half of Top7 (C-terminal half is more stable) though the sequences and structures are highly symmetric, which is suspected to cause the existence of the intermediate state. Two possible unfolding pathways were revealed by single molecule AFM experiments[43], either β1 slides away from β3 first (Pathway I) or β5 slides away from β3 first (Pathway II). Enlightened by the simulation results with molecular dynamics of coarse-grained structure-based models in previously published study[41], we mutated six amino acids (V6I-L29F-I40F-F71V-I79V-W83V) aiming at optimizing packing interactions in the N-terminal half and destabilizing the C-terminal half of Top7 to make it fold cooperatively without changing Top7's topology. Stretching the single molecule of cm-Top7 resulted in reversible unfolding and folding events (Fig. 2.9A). It is of note that both unfolding and folding intermediate states can be observed in some of the cycles, but occurrences of them have no correlations. By comparing the frequency of occurrence of folding intermediate states between wt-Top7 and cm-Top7 (Fig. 2.9B), folding cooperativity has been improved dramatically though non-cooperative folding processes still existed. The ratio of cooperative folding events has increased from 0 to 40% (the number of total events is 309) compared to wt-Top7. The unfolding and folding force-position traces could be well superimposed with those of wt-Top7 including the relative position of folding intermediate in the trace, which made us believe that the structure of the intermediate state of cm-Top7 is same as that of wt-Top7 and structured part of the whole molecule remained unchanged. Additionally, the frequency of occurrence of unfolding intermediate states was increased. The ratio of non-cooperative unfolding events has increased from 3% to 44% (the number of total events is 330) compared to wt-Top7. We deducted that Pathway I without any intermediate state dominated in wt-Top7 because of C-terminal half's higher stability while Pathway II with obligatory

intermediate state occurs less. As for cm-Top7, unfolding pathway may shift from Pathway I to Pathway II since the difference in stability between N/C-terminal halves has been lessened. As a result, unfolding cooperativity was disrupted. We did the same kinetic analysis for cm-Top7 as wt-Top7 assuming that cm-Top7 has simplified two-state unfolding/folding behavior. The free energy change of cm-Top7 was determined to be 19.51 k_BT compared with 17.76 k_BT of wt-Top7, which showed that both constructs have high mechanical stability since the overall topology kept constant. Besides, cm-Top7's higher unfolding distance resulted in lower rupture force though it has lower spontaneous unfolding rate constant.

A) B)

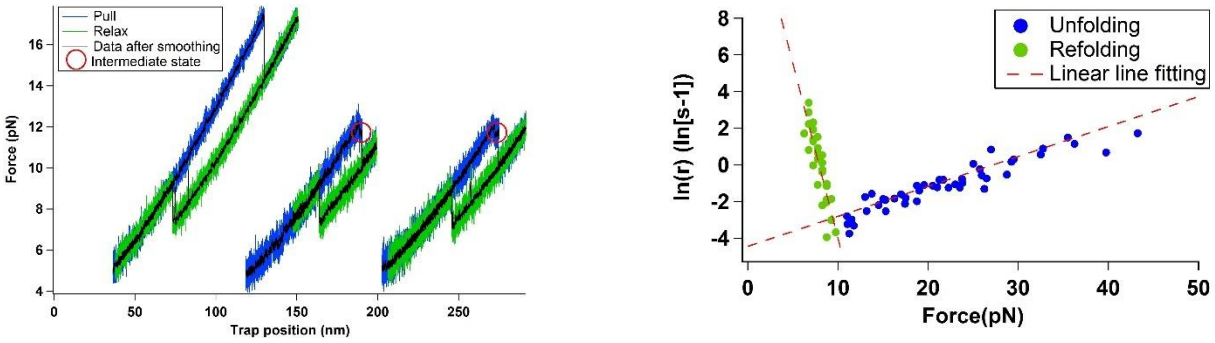


Figure 2.9 Cm-Top7s shows more cooperative folding events and non-cooperative unfolding events. A) Sample traces obtained by pulling with a constant velocity (20 nm/s). B) Force-dependent unfolding rates and folding rates under the assumption of a two-state unfolding/folding process. ΔX_u and ΔX_f are determined to be (0.627 ± 0.036) nm and (7.84 ± 0.86) nm, respectively, and unfolding rate and folding rate are determined to be (0.0117 ± 0.0024) s^{-1} and (3.49 ± 5.82) x 10^6 s^{-1}, respectively.

2.4 Conclusion

Top7 is a typical α/β protein with high mechanical stability. Here we start from investigating the unfolding and folding of wt-Top7 with the constant pulling velocity protocol to extract the kinetic properties. Top7 was shown to fold with intermediate states via both chemical denaturation

methods[38] and computer simulation methods[39,45,46]. The fast transitions at low forces in force-position curves drew our attention and made us realize that it might be possible to observe the folding intermediate state directly by slowing down the pulling velocity which can help us acquire more data points within the same change of the trap position. Here, we show direct evidence of the existence of one obligatory intermediate state during folding process by pulling/relaxing the molecule at a lower speed. Some possible reasons have been proposed previously[38,39,41,46]. Moreover, with the help of high sampling frequency (30 kHz) data acquisition, we discovered that a minor ratio of unfolding processes also showed non-two-state behavior, which had never been observed. During the folding process, the C-terminal half folds completely and N-terminal half partially folds at the intermediate state. This hypothesis was supported by our results which demonstrated the independency of unfolding/folding ability of Top7's C-fragment. Based on this hypothesis, we mutated three residues in N-terminal half to fill the hydrophobic core and stabilize it. At the same time, three residues in C-terminal half were mutated to destabilize it. Our results showed that the folding cooperativity was improved by altering the stability of two terminal halves, but unfolding cooperativity was disrupted. Although some questions such as what the structures of unfolding/ folding intermediate states of wt-Top7 are and the influence of sequence-dependent stability of N/C-terminal half on the choice of unfolding pathway remain open, our results prove that single molecule optical tweezers technique is a powerful tool to investigate proteins containing single or multiple intermediate states. Single molecule force spectroscopy may aid the process of understanding the structure and unfolding/folding pathways of the proteins.

Chapter 3: Mechanical unfolding/folding of rubredoxin revealed by optical tweezers

3.1 Introduction

3.1.1 General introduction

Metalloprotein is a class of proteins which contain a metal ion cofactor.[47] The metal ion cofactor is of vital importance in carrying out the biological functions of the protein. These functions, including oxygen transport, oxygen storage, and electron transfer, are ensured by metalloproteins such as heme iron proteins, non-heme iron proteins, zinc fingers, and copper proteins. Hemoglobin, myoglobin and hemerythrin are three common iron containing metalloproteins which can bind oxygen to participate in the transport and storage of oxygen. Hemocyanin is a typical copper protein which is involved in the transport and storage of oxygen. In addition, many other metalloproteins carrying metal ion work as oxidases and oxygenases in the cells such as iron-containing cytochromes P450[48] and cobalt-containing vitamin B12[49]. Iron sulfur sites, along with blue copper sites and cytochromes, are three main biological electron transfer sites. Iron sulfur electron transfer sites can be divided into four classes. Among the four classes, rubredoxin is the one with simplest structure. Reversible redox reaction of the iron center allows efficient short range electron transfer. Rubredoxin has been found in both anaerobes and aerobes.[50] The archaebacterium, *Pyrococcus furiosus*, and the mesophilic eubacterium, *Clostridium pasteurianum*, are two popular sources of rubredoxin and widely used in biophysical studies. Since the recombinant protein has proved to be able to overexpressed in Escherichia coli[50], we chose rubredoxin from *Pyrococcus furiosus* (*pf*RD) as the investigating target.

3.1.2 Structure of *pf*RD

The structures of rubredoxin from various sources have been investigated in depth with X-ray crystallography method and NMR technique[51-53]. So far, there have been more than 100 structures of rubredoxin in PDB. In a native *pf*RD molecule, a ferric ion is coordinated with four sulfur atoms from cysteine residues in a distorted tetrahedral manner (Fig. 3.1A). A simplified model was first prosed in 1967[54]. In that model, two cysteine coordination motifs (CXXC, residues 5-8 and residues 38-41) divide the whole protein into two segments. The metal center part (residues 5-41) is inside and wrapped by the outer part (residues 1-4 and residues 42-53) (Fig. 3.2B). In terms of the property of the bonding, the ferric-thiolate bond involved in the metal center is believed to be highly covalent[55,56] which allows it to be an electron transfer active site.

A) B)

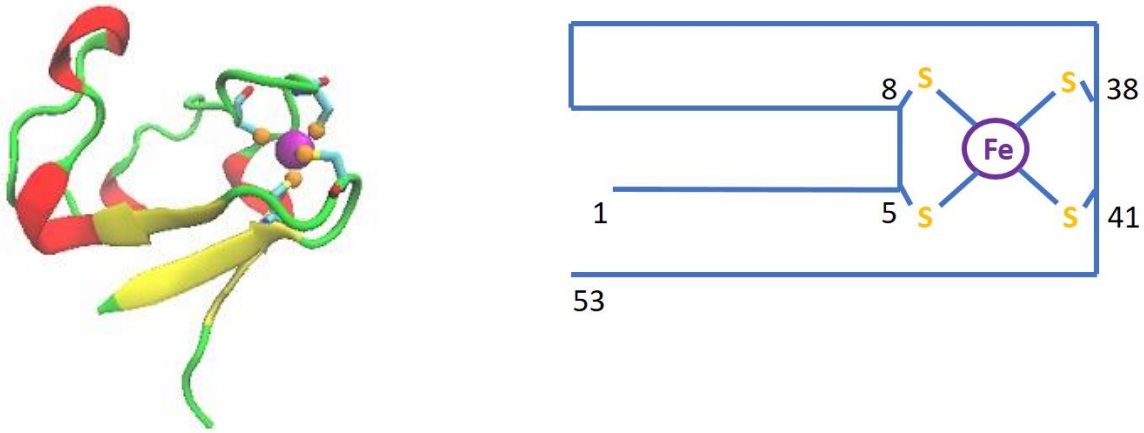


Figure 3.1 Structure of *pf*RD. A) Cartoon representation of *pf*RD's structure (PDB: 1BRF). The backbone of the protein is colored according to the secondary structure (yellow: β-sheet; red: α-helix; green: unstructured coil). The metal center (purple) is coordinated with four sulfur (orange) atoms from four cysteine residues. The figure is generated with VMD[42]. B) Schematic of *pf*RD's structure.

3.1.3 Single molecule studies on metalloprotein

As the simplest iron-sulfur protein, rubredoxin is an ideal model to investigate the folding/unfolding dynamics of metalloprotein. Rubredoxin was found to be highly thermally stable[52,57,58] which makes single molecule technique a promising approach to explore the kinetic information of rubredoxin. Despite of the high covalency of the ferric-thiolate coordination bond in *pf*RD, the mechanical stability of ferric-thiolate bond revealed by AFM experiments is surprisingly low[59]. AFM studies revealed that the unfolding process of rubredoxin adopts a two steps strategy. The outer part of protein unfolds first and then the metal center part is subjected to the force such that it unfolds subsequently. The mechanical rupture of the metal-ligand bonds in rubredoxin is further studied by Zheng et al.[60] They proposed three stochastic mechanical rupture mechanisms of rubredoxin metal center according to the results of AFM experiments. Apart from mechanical unfolding of rubredoxin, the folding study of rubredoxin has also been attempted by Zheng et al.[61] The AFM experiments reveal that rubredoxin is able to refold and rebuild the native metal center structure, whereas the pressure of force may lead to the loss of the metal ion.

In addition to rubredoxin, mechanical unfolding and folding study of a more complex electron transfer protein, ferredoxin, which contains a [2Fe-2S] center by AFM was published recently[62]. Similar to rubredoxin, ferredoxin adopts a two-step mechanical unfolding mechanism. And the [2Fe-2S] center can be fully reconstituted upon pulling and relaxing. The only difference in terms of the results is that the unfolding of ferredoxin's outer part is detectable. Additionally, the mechanical strength of coordination gold-sulfur bond in a gold-specific binding protein, GolB, was investigated using AFM[63]. In summary, the AFM spectroscopy has been proved to be a powerful tool to investigate the mechanical strength of the coordination bond in metalloprotein

and the mechanical unfolding/folding pathways of metalloprotein. However, the application of optical tweezers on metalloprotein is lacking.

3.1.4 Aim of research studies in this thesis

The overexpressed protein was found to be a mixture of iron-containing rubredoxin-GB1 protein and non-native zinc-substituted rubredoxin-GB1 protein[64]. The UV-Vis spectrum of the protein solution shows three absorption maxima (280nm, 390nm, and 494nm) (Fig. 3.2). The extinction coefficient of 9.22 mM^{-1} cm^{-1} at wavelength of 494 nm was used to determine the concentration of iron-containing protein and further ensure the ratio between iron-containing protein and zinc-substituted protein. Our expressed protein was found to contain similar content of Fe-form and Zn-form. For the samples used in the optical tweezers experiments, we didn't do further purification to separate two forms. The reasons lie in several aspects. First, zinc-substituted rubredoxin has similar tetrahedral geometry with ferric-rubredoxin except that the coordination center is a ZnS_4 instead of FeS_4[51,53]. Secondly, despite the difference between the covalency of zinc-thiolate bond and ferric-thiolate bond, the mechanical stabilities of zinc-rubredoxin and ferric-rubredoxin are similar which has been demonstrated with AFM experiments[65]. Additionally, although ZnS_4 is not a classic zinc finger motif, it is the most common one in zinc-containing protein family[66] and can serve as structural sites and active sites[67] which makes it worth studying. In this study, we combined protein engineering and optical tweezers single molecule experiment to study the mechanical unfolding of *pf*RD. Moreover, whether *pf*RD can mechanically unfold and fold reversibly was investigated. Unlike AFM, optical tweezers can be used to study the molecule with a single domain which makes it well suited for investigating the unfolding/folding pathways of protein. Besides, the folding study is more straightforward with optical tweezers. The high

resolution of the optical tweezers allow us to probe the unfolding/folding behavior of rubredoxin more meticulously. To see whether chelation can achieve and accelerate the loss of the metal ion, we also performed optical tweezers experiments in the EDTA-containing buffer. In the whole series of experiments, the reconstructed protein cys- *pf*RD-GB1-cys was used, where GB1 worked as a fingerprint.

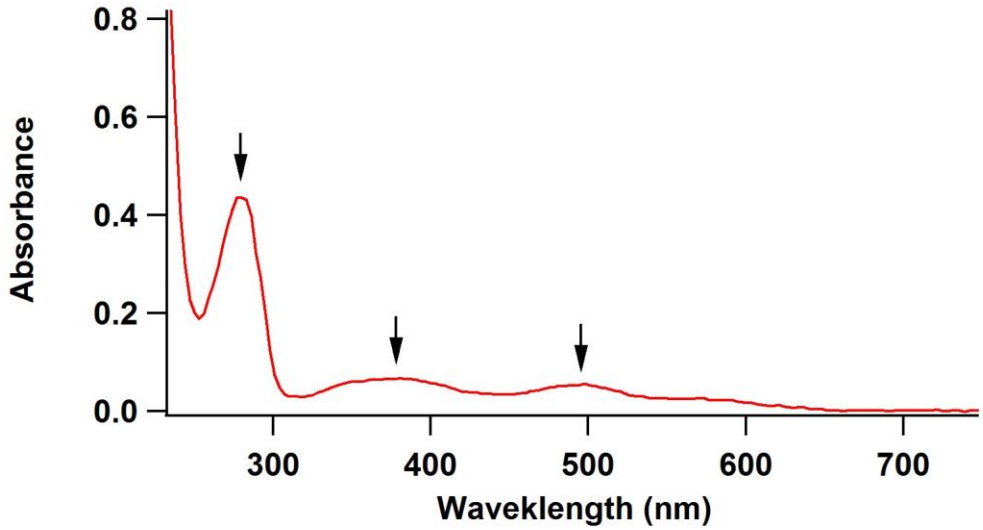


Figure 3.2 UV-Vis spectrum of cys-*pf*RD-GB1-cys. There are three absorption maxima at wavelength of 280nm, 390nm, and 494nm.

3.2 Methods and material

3.2.1 Protein engineering

The gene of *pf*RD was received as a kind gift from Dr. Eidsness and kept in pUC19 (New England Biolabs, Ipswich, MA) vector. The restriction sites of BamHI (GGATCC), BglII (AGATCT) and KpnI (GGTACC) were added using standard PCR. The product gene was nucleated with enzymes of BglII and KpnI to produce the vector. The gene of GB1 insert was obtained from digestion of pUC19/GB1 with BamHI and KpnI. The insert was ligated into the vector such that the recombinant pUC19/RD-GB1 gene was successfully constructed (Fig. 3.3). The cysteine codon

was added to both terminal ends of the gene of RD-GB1 by regular PCR. Finally, the recombinant gene was digested with BamHI and KpnI, and then sub-cloned into expression vector pQE80L. The protein was overexpressed in Escherichia coli strain DH5α and purified using Co^{2+} affinity chromatography with TALON His-Tag purification kit. The size of the protein is confirmed by running SDS-PAGE. The purified protein was kept in PBS (pH 7.4) at 4 ºC. The resulted protein solution usually had a concentration of ~4 mg/mL.

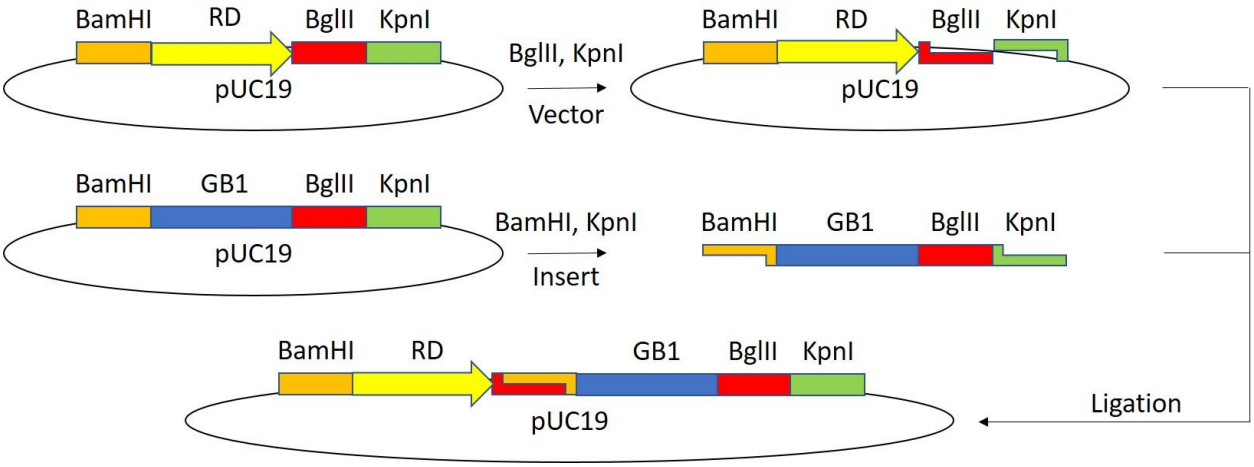


Figure 3.3 Schematic of the construction of pUC19/RD-GB1. The vector of pUC19/RD and the insert of GB1 were obtained through digestion. The resulted vector and insert were ligated to construct recombinant plasmid of pUC19/RD-GB1.

3.2.2 Sample preparation and single molecule optical tweezers experiments

Since the protein construct carried one cysteine at both side, the oligonucleotide-coupled protein was prepared with the method mentioned in sub-section **2.2.2**. Before an optical tweezers experiment, 2 µL of the oligonucleotide-coupled protein solution (20 nM) was reacted with 5 µL suspension of anti-digoxigenin coated beads (0.5% w/v, 2 microns, Spherotech Inc.) for 30 minutes and then diluted into 3 mL with Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4). The volume ratio between the DNA-coupled protein solution and the beads solution may vary depending on

the actual outcome of the experiments. 0.5 µL of streptavidin beads (1.0% w/v, 1 microns, Spherotech Inc.) was directly diluted to 3 mL with Tris buffer. During an optical tweezers experiment, an anti-digoxigenin coated bead was trapped by the laser trap and a streptavidin coated bead was sucked onto the micropipette. The relevant schematics or images can be found in sub-section **2.2.3**. The overstretching state of DNA and the unfolding/folding event of GB1 are used to confirm the single molecule tethering.

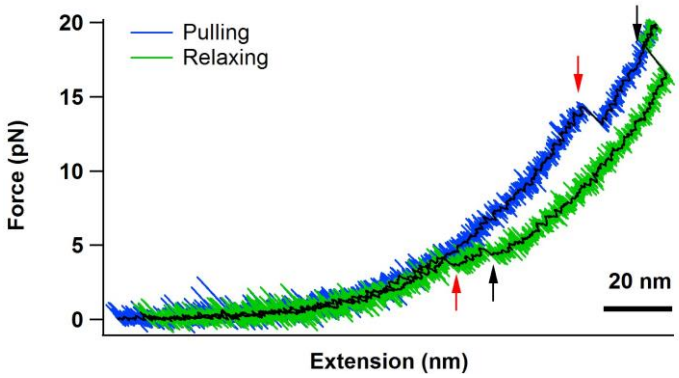
3.3 Results and discussion

3.3.1 Reversible unfolding and folding of rubredoxin

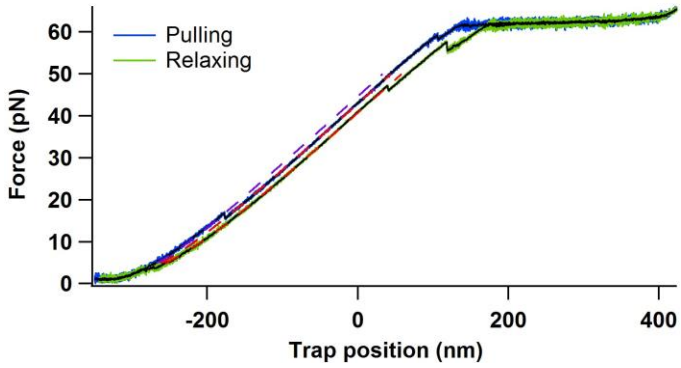
As shown in Fig. 3.1, rubredoxin is a 53-residue protein which is made of two segments, the metal center part and the outer part. Four metal-ligand bonds consititute the metal center which is surrounded by the outer part. We applied optical tweezers to single cys-*pf*RD-GB1-cys molecule and stretched the molecule with a constant pulling speed. Two major force peaks appeared on the FEC obtained from a single pulling-relaxing cycle (Fig. 3.4A). According to calculation, the metal center part corresponds to a ΔL_c of ~12.4 nm (0.36 nm/aa * 37 aa – 0.9 nm = 12.42 nm) and the outer part corresponds to a ΔL_c of ~4.6 nm (0.36 nm/aa * (4+12) aa – 1.2 nm = 4.56). We used eWLC model to fit the force-position curve (Fig. 3.4B). These two contour length increments were further confirmed by WLC fitting (Fig. 3.4C). Those two major peaks have ΔL_c of ~12 nm and ~18 nm, respectively. The force peak with the ΔL_c of 18 nm is believed to result from the unfolding of GB1, which serves as a fingerprint. And the force peak with the ΔL_c of 12 nm is resulted from the unfolding of the metal center part of rubredoxin. Due to the fact that the force cannot take effect on the metal center part of the protein if the surrounding outer part is still intact, the unfolding of the outer part must happen before the unraveling of the metal center part of the protein.

This means that the unfolding of the outer part occurs at an extremely low force, likely so low that the unfolding event cannot be distinguished from the noise. This observation is in agreement with the observation from AFM experiments[59]. Apart from the mechanical unfolding of rubredoxin, the folding of rubredoxin was also directly observed. Our results revealed that the single rubredoxin protein molecule adopts a two-state folding behavior (Fig. 3.4). We will discuss about the folding with more details in next sub-section. Upon further pulling and relaxing of the same molecule, the unfolding and folding ability of rubredoxin is maintained. Although our protein was a mixture of Zn-rubredoxin and Fe-rubredoxin, the mechanical unfolding/folding mechanisms were not expected to be different since the AFM results of them seemed to be similar except that the mechanical stability of ZnS_4 was slightly lower. The fact that we only discovered only one pattern of FEC instead of various patterns from multiple trials on different molecules (more than ten) actually supports this hypothesis.

A)



B)



C)

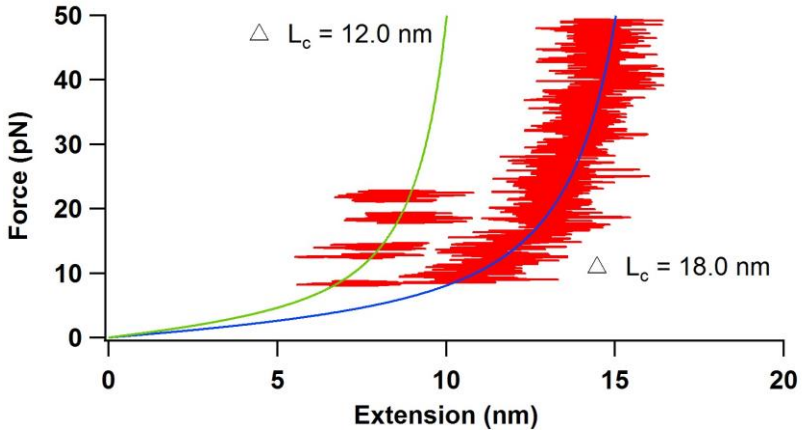


Figure 3.4 Mechanically unfolding and folding of cys- *pf*RD-GB1-cys is reversible. A) In a typical FEC obtained by pulling at 50 nm/s, the outer part of rubredoxin protein unfolds first which cannot be observed; the metal center part unfolds at forces between 4 to 20 pN and folds at forces between 1 to 6 pN (red arrow); GB1 serves as a fingerprint which unfolds at much higher force and folds before the folding of the metal center part (black arrow). B) The fitted eWLC curve (dash line) is overlaid onto a typical force-position curve obtained by pulling at 50 nm/s. The first force peak corresponds to a ΔL_c of ~12 nm. Unfolding of GB1 usually occurs at force larger than 65 pN which cannot be directly detected. The ΔL_c of GB1 is fitted to be ~18nm. C) WLC fitting (blue and green curves) of the extension differences from rubredoxin (left) and GB1 (right). The parameter of persistence length (p) used in the fitting is 0.8 nm.

3.3.2 Force-induced loss of the metal ion is not observed

In a single pulling-relaxing cycle, the folding of rubredoxin was directly observed (Fig. 3.4). As shown, the red arrows in Fig. 3.4 denote the unfolding and folding of rubredoxin. However, two

questions need to be answered. Does rubredoxin fold back to the native structure? Under the high tension, does dissociation of metal ion happen? For the first question, through FEC we found that the mechanical unfolding/folding behavior of the same molecule can be reproduced and lasts for many cycles without showing fatigue until the breakage of the tethering between the bead and the molecule. The apo-rubredoxin has proved to not show any characteristic force peaks in FEC[59] which is also confirmed and will be discussed in the following sub-section **3.3.4**. Combining this fact and our results, we can conclude that the structure of the metal center can be completely reconstructed. A schematic of the mechanism is shown in Fig. 3.5. Since there are two CXXC motifs in the metal center, the unfolding of the protein must lead to the rupture of at least one CXXC motif. And this rupture can be fixed upon the relaxation of the molecule.

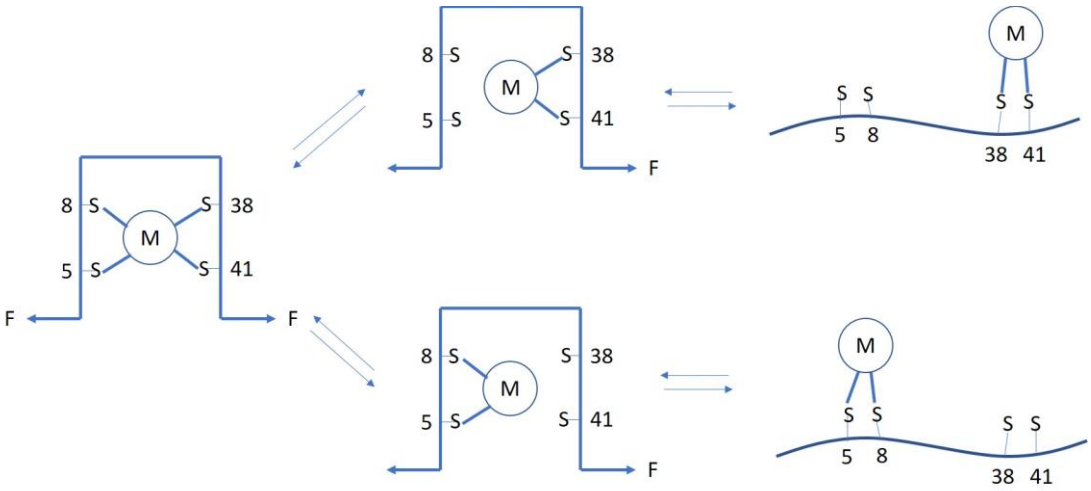


Figure 3.5 Schematic of mechanical unfolding/folding mechanism of FeS_4 center. When the molecule is pulled, one of the two CXXC motif is deconstructed. The metal ion is still attached through two remaining metal-thiolate bonds upon further pulling. When the molecule is relaxed, the broken metal-thiolate bonds are reconstructed. Finally, the FeS_4 center folds back to the native structure. (M = Zn^{2+} or Fe^{3+})

To answer the second question mentioned above, we designed a pulling strategy to test whether the metal ion would be released under the high tension of optical tweezers. The molecule was first

pulled with a constant pulling speed (50 nm/s) at a relatively broad force range (0 – 45 pN) for 50 cycles. And for each cycle, the molecule was held at 45 pN for five seconds. After those 50 cycles, the same molecule was stretched with same pulling speed (50 nm/s) at a relatively narrow force range (0 – 20 pN) for another 90 cycles. Five consecutive FECs from those 90 cycles were shown in Fig. 3.6A. Two-state mechanical unfolding and folding behavior was maintained. The unfolding/folding force distribution showed no obvious difference with the unfolding/folding force distribution of a molecule which hadn't been stretched to higher forces (Fig. 3.6B). It indicates that the mechanical stability of rubredoxin doesn't change and the refolded rubredoxin adopts a native folding structure. Moreover, the metal ion and ferric-thiolate bonds of refolded rubredoxin were proved to be intact even the molecule was exposed to the high applied forces. This finding is different from the results obtained from AFM experiments by Zheng. et al[61,65]. They found that the metal center could not be reformed after the single rubredoxin molecule was pulled and relaxed for more than several cycles no matter whether the metal ion is Zn^{2+} or Fe^{3+}. We think that might be caused by the difference of the applied forces. Our experiments were carried out using optical tweezers with a constant pulling speed of 50 nm/s and the AFM experiments were conducted with a constant pulling speed of 400 nm/s. Correspondingly, the unfolding of rubredoxin's metal center occurred at ~9 pN using optical tweezers compared with ~210 pN using AFM[59]. Thus, the loss of the metal ion might need higher applied force to be achieved.

A) B)

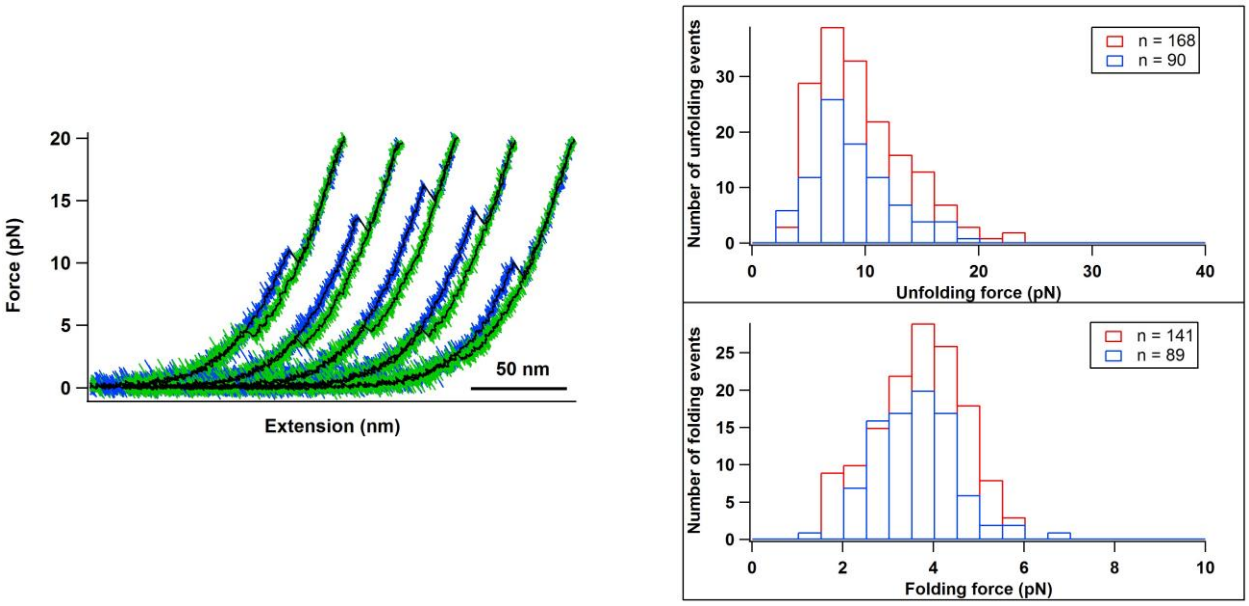


Figure 3.6 The metal center part of rubredoxin protein keeps intact and possesses the same mechanical stability after 50 cycles of pulling/relaxing. A) Five consecutive FECs obtained from the molecule which has been pulled for 50 cycles are shown. The metal center part of the rubredoxin protein still shows the reversible unfolding/folding behavior with the ΔL_c of ~13nm (fitting not shown). B) The unfolding/folding force distribution obtained from the molecule which has been pulled for 50 cycles is graphed (blue) against the unfolding/folding force distribution obtained by simple pulling experiment. The accordance indicates the mechanical stability of the metal center part keeps constant.

3.3.3 Unfolding/folding kinetics revealed by optical tweezers

To extract the kinetic information of rubredoxin, we used the Oberbarscheidt method mentioned in subsection **1.2.3.6** to analyze the data. The spontaneous unfolding rate constant and folding rate constant are determined to be 0.314 s^{-1} and 148.41 s^{-1} using Bell-Evans linear fitting (Fig. 3.7). It is of note that those numbers are much smaller than that of Top7, a fast unfolding/folding protein introduced in Chapter 2. The mechanical stability of rubredoxin's metal center is quite low though it is composed with four high covalent ferric-thiolate bonds. The kinetic information reflects the unique property of FeS_4 center which is worthy of further studying.

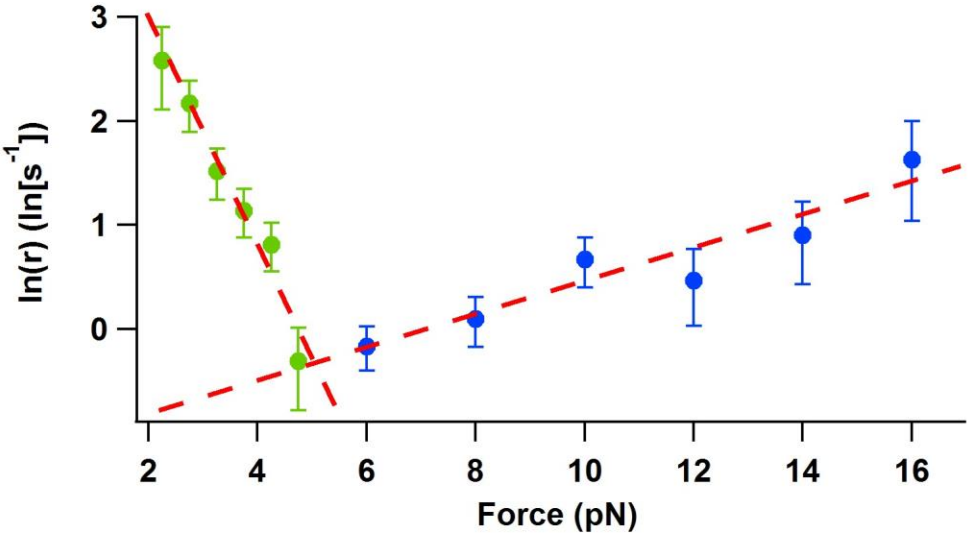


Figure 3.7 Unfolding/folding kinetics revealed by optical tweezers experiments. The unfolding rate constants (blue) and folding rate constants (green) at different forces are plotted. The unfolding distance and folding distance are determined to be 0.66 nm and 4.44 nm. The spontaneous unfolding rate constant and folding rate constant are determined to be 0.314 s^{-1} and 148.41 s^{-1}.

3.3.4 Loss of the metal ion induced by introduction of EDTA

To investigate whether EDTA could chelate the metal ion from the metal center of rubredoxin, we performed optical tweezers experiments for the reconstructed protein cys-*pf*RD-GB1-cys in the presence of 50 mM EDTA (pH = 8.0). There are two possible scenarios. Either the metal center is directly disassembled because of the loss of the metal ion, or the metal ion is attached and the loss of the metal ion requires an external force such that the metal ion is fully exposed to the buffer. We captured ~10 different molecules. For most of the molecules (8/10), FECs only showed one force peak with a ΔL_c of ~ 18 nm which was caused by the unfolding/folding of GB1 (Fig. 3.8). And consecutive FECs show that the force increases smoothly with the increasing extension at the forces below 20 pN, indicating the disruption of the FeS_4 center. This observation also proves that the apo-form of rubredoxin doesn't have any detectable force peaks. For the other two molecules, FECs showed some unresolved force peaks which couldn't be explained (data not shown).

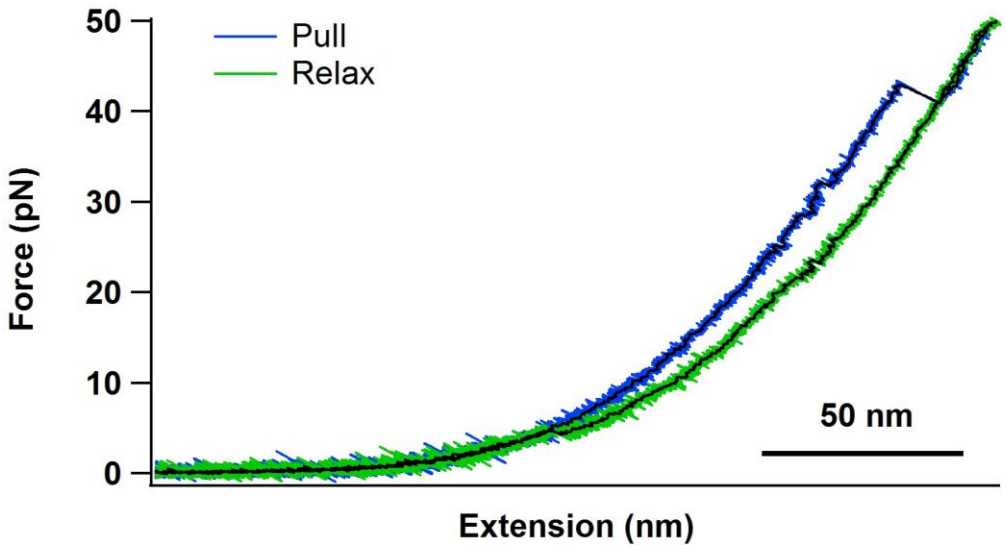


Figure 3.8 The metal will be released in presence of EDTA (50 mM) and apo-rubredoxin shows no detectable force peaks. A typical FEC obtained by keeping the whole system in presence of 50mM EDTA only shows the reversible unfolding/folding event of GB1. The metal center has been fully destructed and apo-form of rubredoxin shows no detectable characteristic force peak.

3.4 Conclusion

Although rubredoxin has been well studied with single molecule AFM experiments, it is the first time to be investigated with optical tweezers technique. Several types of optical tweezers experiments were carried out on the recombinant protein cys-*pf*RD-GB1-cys. We started from simple single molecule pulling experiment using constant pulling speed protocol and found that rubredoxin unfolds in two steps. The outer part of the protein unfolds first followed by the unfolding of metal center part. Whereas, only the unfolding of metal center part is detectable under experimental conditions of optical tweezers. Besides, the unfolding/folding process of rubredoxin is reversible which means the metal center can be completely restored. To investigate if force can induce the loss of the metal ion, we recorded the data after the molecule had been exposed to the force sufficiently. It turns out that the unfolding/folding force distribution does not change

significantly, indicating that the mechanical stability of metal center part keeps constant and the structure is fully recovered. In the condition of EDTA (50mM), most of the molecule will lose the metal ion and the holo-RD is converted into apo-RD which shows no characteristic force peak in the FECs. Based on our results, some other valuable questions may remain unknown and requires further investigations. In the next chapter, an outlook for this project will be presented.

Chapter 4: Future outlook

4.1 Prospects on Top7 project

In Chapter **2**, three protein constructs, including wt-Top7, Top7's C-fragment, and cm-Top7, have been discussed in terms of the results from optical tweezers experiments. To further investigate the unfolding/folding pathways, some other truncated fragments of Top7 than C-fragment may be worth studying with the single molecule technique. How different fragments coordinate with each other during the unfolding/folding process is an interesting and sophisticated question in single molecule studies. To understand the relationship between the structure and the unfolding/folding pathway is of vital importance in protein sequence design. In addition to single molecule experiments, performing steered molecular dynamics (SMD) simulations for single truncated fragment may be another way to supply more theoretical evidence.

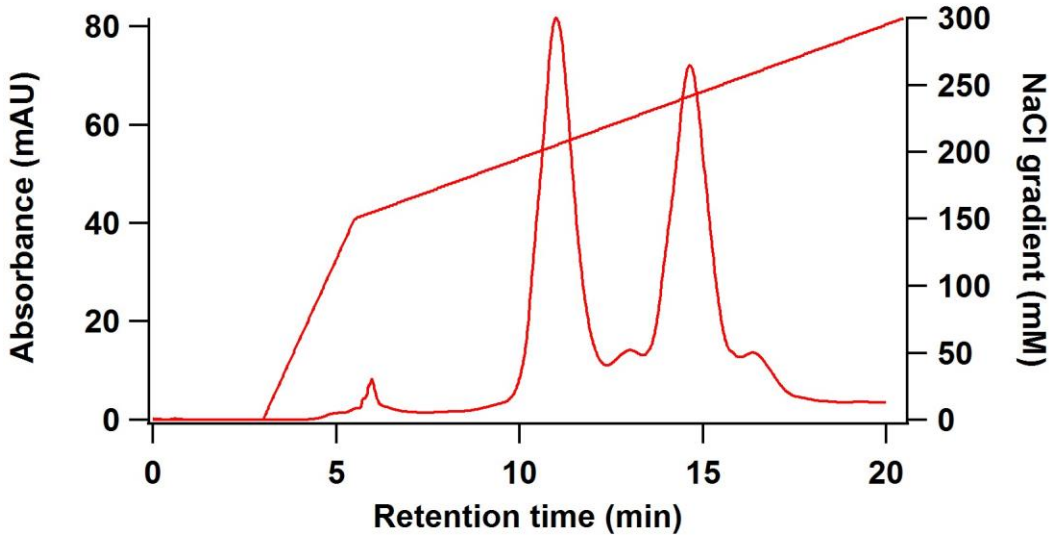
4.2 Prospects on rubredoxin project

In Chapter **3**, mechanical properties of rubredoxin have been characterized and discussed. Based on our results, the holo-rubredoxin unfolds with a two-step manner and can fold back to the native structure without showing any intermediate states regardless whether it is iron-form or zinc-form. To further investigate the unfolding/folding behaviors of those two forms separately, fast protein liquid chromatography (FPLC) is a useful tool to purify two fractions. Corresponding FPLC purification experiments for protein construct of cys-*pf*RD-GB1-cys using anion-exchange chromatography are undergoing. The Mono-Q 5/50GL anion exchange column (GE Healthcare) and AKTA FPLC system (GE Healthcare) are used in the series of FPLC purifications. The Fe-form protein can be eluted first at 200 mM NaCl by using a linear gradient elution (0-300 mM NaCl in 10mM Tris buffer and 1mM TCEP (Tris(2-carboxyethyl)phosphine), pH 8.5). Similarly,

the Zn-form protein can be collected after at 250 mM NaCl (Fig. 4.1A). The ratio of absorbance at 494 nm and absorbance at 280 nm in the UV-Vis spectrum (0.23 for 100% purity) is used to determine the purity of Fe-rubredoxin. The purified Fe-RD protein sample consists of 87% Fe-form protein according to the UV-Vis spectrum (Fig. 4.1B). Two fractions of solution are concentrated through 3k Amicon Ultra – 4 centrifugal filters (Merck Millipore Ltd.). The coupled DNA-protein samples are prepared with the protocol mentioned before (sub-section **2.2.2**). The relevant optical tweezers experiments for two forms of rubredoxin is still being attempted whereas acquired valid data is limited. For the future experiments, the mechanical stability and related kinetic information of two forms of protein can be compared. With those comparisons, the different roles Fe and Zn play can be further discussed.

Optical tweezers experiments on non-purified cys-*pf*RD-GB1-cys in the presence of EDTA (50 mM) have been carried out. The results show that the most of the metal centers have been chelated before the pulling experiments. For future studies, other concentrations of EDTA can be attempted on two forms of protein samples respectively. From that, the difference in the capacity of resisting the chelation between Fe-form and Zn-form proteins can be discussed.

A)



B)

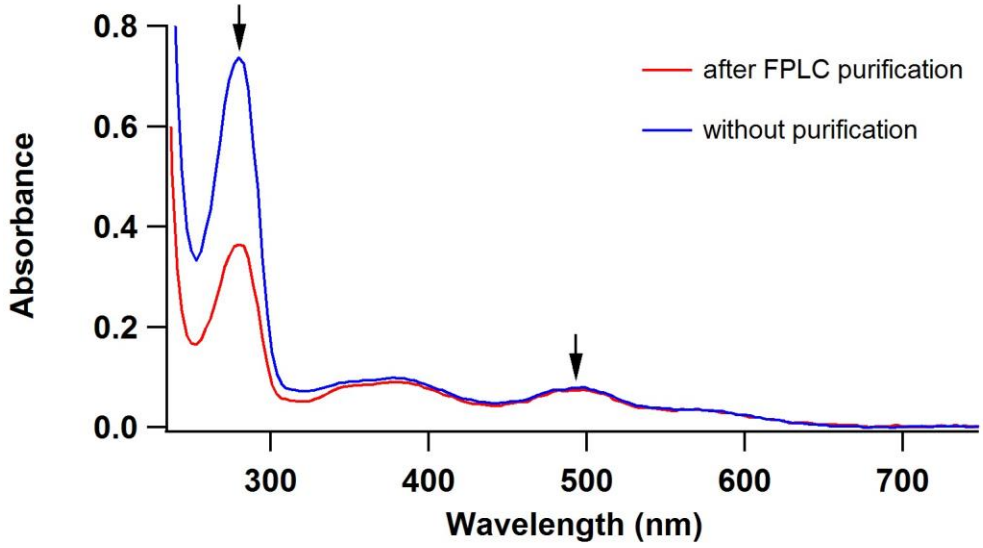


Figure 4.1 FPLC purification using anion exchange chromatography for cys-*pf*RD-GB1-cys. A) The protein solution is subjected to the Mono-Q 5/50GL anion exchange column and eluted with a linear gradient elution (0-300 mM NaCl in 10mM Tris buffer and 1mM TCEP, pH 8.5). Fe-form (first peak) and Zn-form (second peak) of the protein is successfully separated. B) The UV-Vis spectrum of purified Fe-form protein sample is compared with that of non-purified protein sample. A494/A280 is used to calculate the purity of the Fe-form. The purity of Fe-form is significantly improved by using anion exchange chromatography.

Bibliography

(1) Ashkin, A. *Physical Review Letters* **1970**, *24*, 156.

(2) Neuman, K. C.; Block, S. M. *The Review of scientific instruments* **2004**, *75*, 2787.

(3) Neuman, K. C.; Nagy, A. *Nature Methods* **2008**, *5*, 491.

(4) Smith, S. B.; Cui, Y.; Bustamante, C. In *Methods in Enzymology*; Academic Press: 2003; Vol. Volume 361, p 134.

(5) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science (New York, N.Y.)* **1994**, *265*, 1599.

(6) Dantas, G.; Watters, A. L.; Lunde, B. M.; Eletr, Z. M.; Isern, N. G.; Roseman, T.; Lipfert, J.; Doniach, S.; Tompa, M.; Kuhlman, B.; Stoddard, B. L.; Varani, G.; Baker, D. *Journal of molecular biology* **2006**, *362*, 1004.

(7) Li, H.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkvliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M. *Nature* **2002**, *418*, 998.

(8) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science (New York, N.Y.)* **1997**, *276*, 1109.

(9) Ciechanover, A. *The EMBO Journal* **1998**, *17*, 7151.

(10) Brujić, J.; Hermans, R. I. Z.; Garcia-Manyes, S.; Walther, K. A.; Fernandez, J. M. *Biophysical journal* **2007**, *92*, 2896.

(11) Cao, Y.; Li, H. *Langmuir* **2011**, *27*, 1440.

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

(13) Zhang, X.; Halvorsen, K.; Zhang, C. Z.; Wong, W. P.; Springer, T. A. *Science (New York, N.Y.)* **2009**, *324*, 1330.

(14) Wang, M. D.; Yin, H.; Landick, R.; Gelles, J.; Block, S. M. *Biophysical journal* **1997**, *72*, 1335.

(15) Bell, G. I. *Science (New York, N.Y.)* **1978**, *200*, 618.

(16) Evans, E.; Ritchie, K. *Biophysical journal* **1997**, *72*, 1541.

(17) Hummer, G.; Szabo, A. *Biophysical journal* **2003**, *85*, 5.

(18) Dudko, O. K.; Hummer, G.; Szabo, A. *Proceedings of the National Academy of Sciences* **2008**, *105*, 15755.

(19) King, W. T.; Su, M.; Yang, G. *International journal of biological macromolecules* **2010**, *46*, 159.

(20) Oberbarnscheidt, L.; Janissen, R.; Oesterhelt, F. *Biophysical journal* **2009**, *97*, L19.

(21) Lei, H.; He, C.; Hu, C.; Li, J.; Hu, X.; Hu, X.; Li, H. *Angewandte Chemie International Edition* **2016**, n/a.

(22) He, C.; Hu, C.; Hu, X.; Hu, X.; Xiao, A.; Perkins, T. T.; Li, H. *Angewandte Chemie International Edition* **2015**, *54*, 9921.

(23) Zhang, X.; Ma, L.; Zhang, Y. *The Yale Journal of Biology and Medicine* **2013**, *86*, 367.

(24) Stigler, J.; Ziegler, F.; Gieseke, A.; Gebhardt, J. C. M.; Rief, M. *Science (New York, N.Y.)* **2011**, *334*, 512.

(25) Dudko, O. K.; Hummer, G.; Szabo, A. *Physical Review Letters* **2006**, *96*, 108101.

(26) Zhang, Y.; Dudko, O. K. *Proc Natl Acad Sci U S A* **2013**, *110*, 16432.

(27) Alemany, A.; Mossa, A.; Junier, I.; Ritort, F. *Nat Phys* **2012**, *8*, 688.

(28) Guinn, E. J.; Jagannathan, B.; Marqusee, S. *Nature communications* **2015**, *6*, 6861.

(29) Jahn, M.; Buchner, J.; Hugel, T.; Rief, M. *Proceedings of the National Academy of Sciences* **2016**, *113*, 1232.

(30) Yu, H.; Liu, X.; Neupane, K.; Gupta, A. N.; Brigley, A. M.; Solanki, A.; Sosova, I.; Woodside, M. T. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 5283.

(31) Ziegler, F.; Lim, N. C.; Mandal, S. S.; Pelz, B.; Ng, W. P.; Schlierf, M.; Jackson, S. E.; Rief, M. *Proc Natl Acad Sci U S A* **2016**, *113*, 7533.

(32) Chistol, G.; Liu, S.; Hetherington, C. L.; Moffitt, J. R.; Grimes, S.; Jardine, P. J.; Bustamante, C. *Cell* **2012**, *151*, 1017.

(33) Heidarsson, P. O.; Naqvi, M. M.; Otazo, M. R.; Mossa, A.; Kragelund, B. B.; Cecconi, C. *Proceedings of the National Academy of Sciences* **2014**, *111*, 13069.

(34) Woodside, M. T.; Block, S. M. *Annual Review of Biophysics* **2014**, *43*, 19.

(35) Yu, H.; Dee, D. R.; Liu, X.; Brigley, A. M.; Sosova, I.; Woodside, M. T. *Proc Natl Acad Sci U S A* **2015**, *112*, 8308.

(36) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science (New York, N.Y.)* **2003**, *302*, 1364.

(37) Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H.; Fernandez, J. M. *Progress in Biophysics and Molecular Biology* **2000**, *74*, 63.

(38) Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. *Cell* **2007**, *128*, 613.

(39) Zhang, Z.; Chan, H. S. *Biophysical journal* **2009**, *96*, L25.

(40) Scalley-Kim, M.; Baker, D. *Journal of molecular biology* **2004**, *338*, 573.

(41) Yadahalli, S.; Gosavi, S. *Proteins* **2014**, *82*, 364.

(42) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of molecular graphics* **1996**, *14*, 33.

(43) Sharma, D.; Perisic, O.; Peng, Q.; Cao, Y.; Lam, C.; Lu, H.; Li, H. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 9278.

(44) Goldman, D. H.; Kaiser, C. M.; Milin, A.; Righini, M.; Tinoco, I.; Bustamante, C. *Science (New York, N.Y.)* **2015**, *348*, 457.

(45) Mohanty, S.; Meinke, J. H.; Zimmermann, O. *Proteins* **2013**, *81*, 1446.

(46) Gaye, M. L.; Hardwick, C.; Kouza, M.; Hansmann, U. H. E. *EPL (Europhysics Letters)* **2012**, *97*, 68003.

(47) Thomson, A. J.; Gray, H. B. *Current Opinion in Chemical Biology* **1998**, *2*, 155.

(48) Schenkman, J. B.; Greim, H. *Cytochrome P450*; Springer-Verlag: New York;Berlin;, 1993; Vol. 105;105.;.

(49) Co, M.; Merck; Co *Vitamin B12*; Merck, Chemical Division: Rahway,N.J, 1958.

(50) Jenney Jr, F. E.; Adams, M. W. W. In *Methods in Enzymology*; Academic Press: 2001; Vol. Volume 334, p 45.

(51) Day, M. W.; Hsu, B. T.; Joshua-Tor, L.; Park, J. B.; Zhou, Z. H.; Adams, M. W.; Rees, D. C. *Protein Science : A Publication of the Protein Society* **1992**, *1*, 1494.

(52) Blake, P. R.; Park, J. B.; Bryant, F. O.; Aono, S.; Magnuson, J. K.; Eccleston, E.; Howard, J. B.; Summers, M. F.; Adams, M. W. W. *Biochemistry* **1991**, *30*, 10885.

(53) Blake, P. R.; Park, J. B.; Zhou, Z. H.; Hare, D. R.; Adams, M. W.; Summers, M. F. *Protein Science : A Publication of the Protein Society* **1992**, *1*, 1508.

(54) Bachmayer, H.; Piette, L. H.; Yasunobu, K. T.; Whiteley, H. R. *Proceedings of the National Academy of Sciences of the United States of America* **1967**, *57*, 122.

(55) Solomon, E. I.; Gorelsky, S. I.; Dey, A. *Journal of Computational Chemistry* **2006**, *27*, 1415.

(56) Rose, K.; Shadle, S. E.; Eidsness, M. K.; Kurtz, D. M.; Scott, R. A.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *Journal of the American Chemical Society* **1998**, *120*, 10743.

(57) Cavagnero, S.; Debe, D. A.; Zhou, Z. H.; Adams, M. W. W.; Chan, S. I. *Biochemistry* **1998**, *37*, 3369.

(58) Cavagnero, S.; Zhou, Z. H.; Adams, M. W. W.; Chan, S. I. *Biochemistry* **1998**, *37*, 3377.

(59) Zheng, P.; Li, H. *Journal of the American Chemical Society* **2011**, *133*, 6791.

(60) Zheng, P.; Takayama, S.-i. J.; Mauk, A. G.; Li, H. *Journal of the American Chemical Society* **2013**, *135*, 7992.

(61) Zheng, P.; Wang, Y.; Li, H. *Angewandte Chemie* **2014**, *126*, 14284.

(62) Lei, H.; Guo, Y.; Hu, X.; Hu, C.; Hu, X.; Li, H. *Journal of the American Chemical Society* **2017**.

(63) Wei, W.; Sun, Y.; Zhu, M.; Liu, X.; Sun, P.; Wang, F.; Gui, Q.; Meng, W.; Cao, Y.; Zhao, J. *Journal of the American Chemical Society* **2015**, *137*, 15358.

(64) Eidsness, M. K.; O'Dell, S. E.; Kurtz, D. M., Jr.; Robson, R. L.; Scott, R. A. *Protein engineering* **1992**, *5*, 367.

(65) Zheng, P.; Li, H. *Biophysical journal* **2011**, *101*, 1467.

(66) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. *Journal of Proteome Research* **2006**, *5*, 196.

(67) Sénèque, O.; Latour, J.-M. *Journal of the American Chemical Society* **2010**, *132*, 17760.

Appendices

Appendix A DNA sequence

A.1 Forward primer for 802 bp dsDNA handles

5’-Biotin-CAA-AAA-ACC-CCT-CAA-GAC-CC

A.2 Reverse primer for 802 bp dsDNA handles

5’-NH_2-CGA-CGA-TAA-ACG-TAA-GGA-CAT

A.3 Forward primer for 558 bp dsDNA handles

5’-Digoxigenin-CAA-AAA-ACC-CCT-CAA-GAC-CC

A.4 Reverse primer for 558 bp dsDNA handles

5’- NH_2-GCT-ACC-GTA-ATT-GAG-ACC-AC

Appendix B Protein sequence

B.1 GB1

MDTYKLILNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTE

B.2 Wild-type Top7

MGDIQVQVNI DDNGKNFDYT YTVTTESELQ KVLNELKDYI KKQGAKRVRI SITARTKKEA EKFAAILIKV FAELGYNDIN VTWDGDTVTV EGQL

B.3 GB1-Top7's C-fragment-GB1

MDTYKLILNG KTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTERSA KRVRISITA RTKKEAEKFA AILIKVFAEL GYNDINVTWD GDTVTVEGQL EGGSFERSMD TYKLILNGKT LKGETTTEAV DAATAEKVFK QYANDNGVDG EWTYDDATKT FTVTE

B.4 Top7's cooperative mutant (mutation sites are colored in red)

MGDIQIQVNI DDNGKNFDYT YTVTTESEFQ KVLNELKDYF KKQGAKRVRI SITARTKKEA EKFAAILIKV VAELGYNDVN VTVDGDTVTV EGQL

B.5 Cys-Rubredoxin-GB1-Cys

CAKWVCKICG YIYDEDAGDP DNGISPGTKF EELPDDWVCP ICGAPKSEFE KLEDRSMDTY KLILNGKTLK GETTTEAVDA ATAEKVFKQY ANDNGVDGEW TYDDATKTFT VTEC