HETEROGENEITY IN PROTEIN FOLDING EXPLORED BY SINGLE MOLECULE FORCE SPECTROSCOPY

By

Dena Izadi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Physics- Doctor of Philosophy
Quantitative Biology- Dual Major

2017

ABSTRACT

HETEROGENEITY IN PROTEIN FOLDING EXPLORED BY SINGLE MOLECULE FORCE SPECTROSCOPY

By

Dena Izadi

The B1 domain of protein G is a well-studied model system of protein folding that has a mixed secondary structure, high stability and relatively slow folding kinetics. However, recent measurements on the sub-millisecond timescale show significant complexity before the final folding step at ~1ms, which depends on how folding is observed. Applying force is one natural method for denaturing proteins in order to observe the process of refolding in detail, but the time resolution of standard instruments typically make investigation of fast folding proteins difficult. In this study, two alkyne-containing unnatural amino acid residues were added to both N and C termini of GB1 and connected via click chemistry to double stranded DNA handles. The protein-DNA chimeras were then suspended between a pair of polystyrene beads held in high-resolution dual optical traps and reversible folding/unfolding events were observed both during force-extension pulling/relaxation experiments as well as under constant force feedback conditions. Presented results for standard, out-of-equilibrium force ramp measurements of GB1 show transitions between two apparently unique folded and unfolded states in agreement with polymer modeling and simple two-state modeling. However, force distributions suggest the presence of intermediate states or multiple folding pathways. Long duration fixed trap position measurements directly reveal equilibrium folding and unfolding reactions and confirm complexity in unfolding and the presence of intermediates. For GB1, while the folding rates are consistent over many molecules and depend strongly on force, the unfolding rates vary widely between molecules.

To my parents, Shahla and Hossein

ACKNOWLEDGMENTS

I would like to thank all the people who contributed in some way to the work described in this thesis. First of all, I thank my academic advisor, Professor Lisa Lapidus, for accepting me into her group. When I joined her lab, I had absolutely no experience in biophysical research. Without her guidance and support, finishing this dissertation would not have been possible. In addition, a thank you to our collaborator, and my co-advisor, Professor Matthew Comstock, without his efforts my job would have undoubtedly been more difficult. He walked me in small steps to become an independent researcher and experimentalist.

I am thankful to my committee members, Professor Phillip Duxbury, Professor Stuart Tessmer and Professor Christina Chan, whose comments have always been helpful. I am especially grateful to Professor Duxbury, who has been helping me from the day I got accepted to Michigan State University for my doctorate studies.

Every result described in this thesis was accomplished with the help and support of fellow lab mates and collaborators from both Lapidus lab and Comstock lab. I am especially thankful to Miles Whitmore, who was a great mentor to me, Kinshuk, Cho-Ying and many others who are all my best friends. I am very grateful for the experimental data shared with me from Yujie Chen and Joseph Slivka, who are further expanding the future work on the single molecule GB1 project.

I owe a debt of gratitude to all the members of Physics and Astronomy Department at Michigan State University, specially Professor Mahanti, Professor Pratt, Professor Tollefson, Kim Crosslan, and Cathy Cords.

I thank all the journals which gave me the permission to include copyrighted pictures as a part of my dissertation. I am also thankful for the funding sources that allowed me to pursue my graduate school studies.

I am thankful to all members of Lurie Nanofabrication Facility at University of Michigan, who all helped and mentored me through the microfabrication process that was done there and is part of my dissertation.

I would like to acknowledge all my friends who supported me during my stay at MSU and made my time a lot more fun.

Finally, I would like to thank my mom, Shahla, my dad, Hossein, and my brother, Alborz, for their constant love and support. Alborz supported me through moments of hopelessness and played the role of my whole family during my stay in East Lansing. I could not finish my studies if I did not have such a great parents, whose skype calls motivated me to stay strong, be grateful for what I have and never give up. I am obligated to both of them for never giving up on me and sacrificing a lot to let me be who I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis has two major parts: the first (chapter two) describes fabrication of ultra-rapid microfluidic mixers being used in our lab using a new, simplified, and low-cost method, and the second involves detailed investigation of the kinetics and behavior of GB1 using high resolution dual optical traps (chapters three, four, and five). In this chapter, we provide detailed background needed for understanding both parts: broad introduction to protein folding and single molecule force spectroscopy (SMFS) and methods used to study them throughout this thesis.

1.1.Protein Folding

Proteins are a family of biological macromolecules, composed of amino acid subunits or monomers. Each protein has a particular sequence of amino acids, which are covalently attached to one another to form long linear chains called polypeptides. These chains are able to fold and generate compact domains with specific three-dimensional structures [1], as shown in Figure 1. These folded domains can serve as building blocks for larger assemblies, provide specific catalytic or binding sites, or regulate the function of DNA. These folded polypeptide chains are functional by themselves in some cases and in other cases they combine with additional polypeptide chains to form the final protein structure. Proteins also perform essential functions throughout the human body. As examples, they are responsible for synthesizing and repairing DNA, transporting materials across the cell, receiving and sending chemical signals, and catalyzing chemical reactions. Three-dimensional structures of proteins can be determined by different experimental methods such as protein crystallography, electron microscopy including electron crystallography

and single particle reconstruction, x-ray crystallography and nuclear magnetic resonance (NMR), small-angle x-ray and neutron scattering (SAXS and SANS). X-ray crystallography and NMR have made the largest contribution to our understanding of protein structures. Each method has its advantages and limitations [2, 3] and all methods have provided useful information from which a set of basic principles of protein structure has emerged. X-ray crystallography can provide very detailed atomic information, showing every atom in a protein or nucleic acid along with atomic details of ligands, inhibitors, ions, and other molecules that are united into the crystal. NMR provides information on proteins in solution, as opposed to those locked in a crystal or bound to a microscope grid, and thus, NMR spectroscopy is the premier method for studying the atomic structures of flexible proteins. On the other hand, electron micrographic studies often combine information from X-ray crystallography or NMR spectroscopy to sort out the atomic details [4]. All these methods help understand how protein structure is generated, see fundamental relationships between different proteins, identify mutual structural themes, and relate structures to functions.

The folding process happens spontaneously and is determined by physical interactions of polypeptide chain. All amino acids have the same backbone structure; hanging off their middle carbon is a different chain. This gives amino acids their unique physical characteristics, which influences the protein information and function. Understanding the physical properties of amino acids, including polarity, solubility, melting point, pKa, optical properties and charge, helps in understanding the structure of proteins. To understand the biological function of proteins in the cell, we would like to be able to predict the three-dimensional structure from amino acid sequence. Therefore, the general folding problem, for a given sequence, is to find the final structure and folding pathway. Despite all the efforts so far over the years, it has not been possible to predict the

detailed folding trajectories theoretically and the folding problem is still unsolved even for a small protein. Experimentally we can determine folding trajectory but only for certain probes and time scales. In addition, many different sequences can fold to similar structures. All that said we are still very far from being able to predict a folding pathway from only the protein sequence.

In the early days of protein experimentation, it was found to be feasible to denature or unfold a protein from its native structure in the lab. In 1929, Mirsky and Anson, discovered that this process was reversible [5]. The connection between a protein's function and its structure was later well established and in 1961 Anfinsen et. al. showed that as the concentration of denaturant decreased in a ribonuclease, the enzymatic activity increased [6]. They were also the first group to propose a thermodynamic explanation stating that the free energy of the protein is at a minimum value when it lies in native folded state [7]. This statement by itself does not explain the process of folding from unfolded to the final folded structure and there are a vast number of theoretical approaches trying to understand and explain the protein folding kinetic process [8] such as folding pathways [9] and energy landscapes [10-15]. First theories to understand the formation of protein structures were developed after Anfinsen showed that proteins are able to fold by themselves [16]. In addition, Levinthal [17] observed that a huge number of conformations are possible for a polypeptide chain and this led Levinthal group to believe that proteins follow a programmed 3D structure formation, called a folding pathway. An energy landscape is a mapping or graph of all those possible conformations and their corresponding energy levels (Gibbs free energy) across the configuration space of the amino acid sequence.

A small fraction of any population of protein molecules occupy each possible higher energy, partially unfolded state, or even the fully unfolded state, as described by the Boltzmann distribution, when in native conditions. The study of these partially unfolded forms (intermediates)

provides useful information about nature of protein structure and can define the unfolding and refolding pathways of a protein even though the intermediates are not visible to many measurement methods.

Protein aggregation is the process by which misfolded proteins adopt a conformational structure, which causes its polymerization into aggregates that clump together and produce organized fibrils. Protein aggregation can occur due to a variety of causes and those causes can be classified to four different groups. Aggregation can cause by mutations, problems in synthesis, aging, or environmental stresses [18-20]. Protein misfolding and aggregation are often correlated with many human diseases such as many neurodegenerative diseases (amyloidosis), Prion disease, Parkinson disease and Alzheimer's disease. To prevent aggregation or to understand its kinetics, we need to understand how a protein folds from its unfolded amino acid structure to its final native structure and this gives us the perspective to sort out many possible pathways in the folding process that might be trapped in the aggregation state. This state also normally has a huge barrier and it is mostly nonreversible. This is one important reason to determine whether a polypeptide chain will fold to its native state or aggregate.

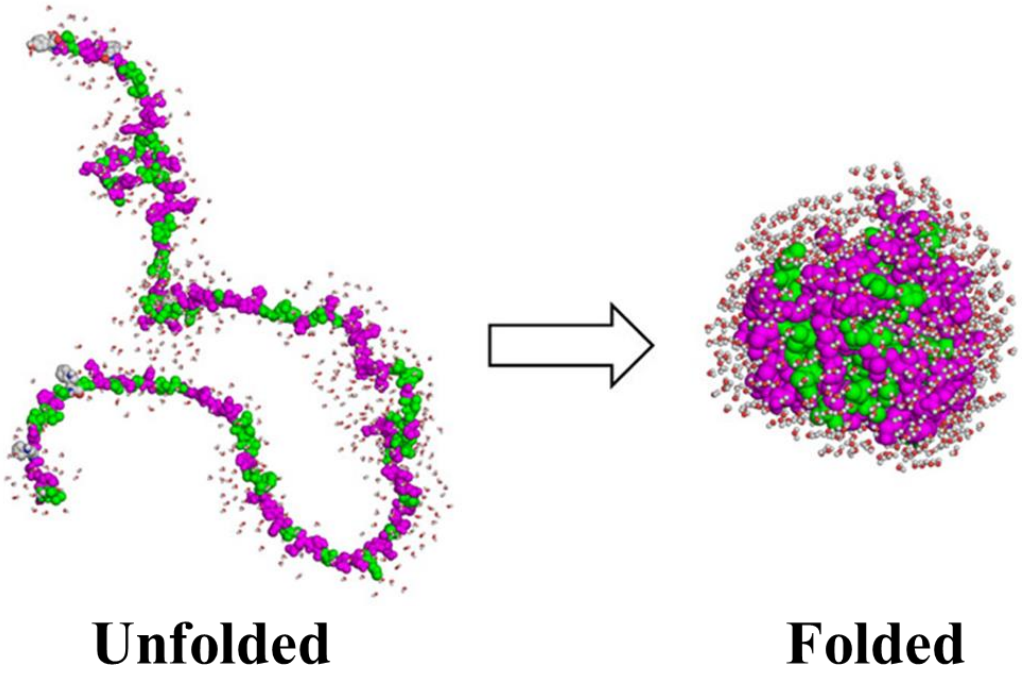

Figure 1: Polypeptide amino acid sequence folding to its native 3D folded structure[1]

1.2.Folding Funnel, Energy Landscape, and Chevron Plots

Figure 2 shows one possible explanation for how proteins fold into their native structures, which is based on the concept of minimizing free energy. This pictorial representation of the energy landscape theory of protein folding was first introduced by Wolynes and Bryngleson as a folding funnel diagram in 1989 and was described more later by Onuchic and Thirumalai [21-23]. Afterwards, the focus of approaches to the protein-folding problem shifted from phenomenological models to a consideration of the general characteristics of the energy surface of a polypeptide chain. In this model, the unfolded protein has both high entropy and high free energy. The high

[1] https://biomedical.closeupengineering.it

entropy corresponds to there being a large number of possible conformational states or in other words, the molecule can take on many different three-dimensional shapes. The high free energy means that the molecule is unstable, and switches easily between the different possible conformational states. As the protein starts to fold, the free energy drops and the number of available conformational states (denoted by the width of the funnel) decreases. At the bottom of the funnel, the free energy is at a minimum and there is only one conformational state available to the protein molecule. This is called its native state, and is the ground state of the system. There are local minima along the way that can trap the protein in a metastable state for some time, slowing its progress towards the free energy minimum. Since the original publication, funnel diagrams have become a fixture in protein folding concepts [24]. The concept introduced by Levinthal [25] is that the appropriate point of reference for protein folding is a random search problem. This can mean that all conformations of the polypeptide chain, except the native state, are equally likely to happen. This theoretically leads to an enormously long folding timescales and experiments show that for small proteins with 100 residues or fewer folding occurs on the order of microseconds to seconds meaning that for any of the models to work, there must be energetic factors that bias the folding process. This bias is presented in three-dimensional folding funnel. Although the field of protein folding has made remarkable progress during the last decades, predicting the folding pathway trajectory from the sequence is still impossible. One of the reasons is that our understanding of unfolded portion of the potential energy landscape is not complete.

The timescales over which folding processes occur is extremely wide and simulation is not possible for this large a range. Many experimental and computational improvements have been made to better understand the unfolded configurations of a protein and be able to better predict the folding pathways. Large-scale computational approaches have recently produced models that

predict many states and many independent trajectory pathways that describes experimental results quantitatively [26-29]. This suggests that at top of the funnel, which has a high energy and low entropy, many details have been missing. See Figure 2 for a schematic 3D view of an energy landscape for protein folding.

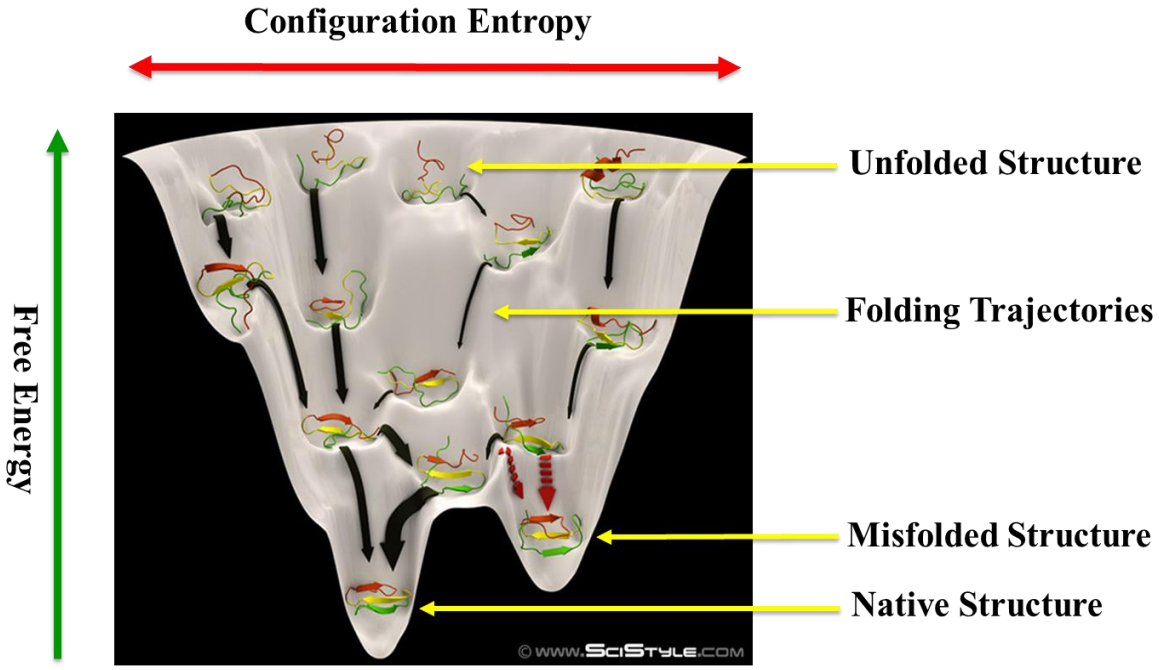

Figure 2: 3D energy landscape for protein folding showing unfolded structures of the peptide on the rim and folding pathways all along the funnel. Native folded structure has the lowest free energy possible.[2]

For showing the dependence of reaction rates to temperature, Svante Arrhenius proposed an equation in 1889 for the temperature dependence of equilibrium constants for both forward and

[2] www.SciStyle.com

reverse reactions. This equation shows the dependence of the rate constant of a chemical reaction to the absolute temperature:

$$k = Ae^{-E_a/(k_B T)}, \quad \text{(Eq. 1)}$$

Where k is the rate constant, T is the absolute temperature (in Kelvin), A is the pre-exponential factor, a constant for each chemical reaction that defines the rate due to frequency of collisions in the correct orientation, E is the activation energy for the reaction (in Joules), and k_B is the Boltzmann constant. Activation energy is actually the minimum energy, which must be available to a system in order to result in a chemical reaction. This energy is also defined as the minimum energy required for starting a chemical reaction and is given in units of kilojoules per mole (KJ/mol) or kilocalories per mole (kcal/mol). This energy can be considered as the energy barrier, which separates two minima of potential energy of the protein. See Figure 3.

It is often convenient to project the landscape onto a one-dimensional energy landscape, since every measurement has only one reaction coordinate. This coordinate could be thermal perturbation, chemical denaturant, force (change in extension) denaturant, etc. The one-dimensional energy landscape along the reaction coordinate is shown as a polynomial curve that has the potential well for native state on one side, a transition barrier in the middle, and an unfolding potential well. The change in the Gibbs free energy, ΔG, of the protein is what defines this one-dimensional profile. The relative heights of the wells change with relative stability of the folded state in that dimension. Folding rates can be described using Kramer's theory [30]. Physical derivation of the rates is as follows (Figure 3):

$$k = k_0 \exp(-\Delta G^{\ddagger}/kBT), \quad \text{(Eq. 2)}$$

where $k_0=\frac{\sqrt{\omega_w \omega_b}}{2\pi k_B T} D$, D is the diffusion constant over the energy barrier, ω_w is the stiffness (curvature) of the potential well, ω_b is the stiffness of the barrier, and k_B is the Boltzmann constant.

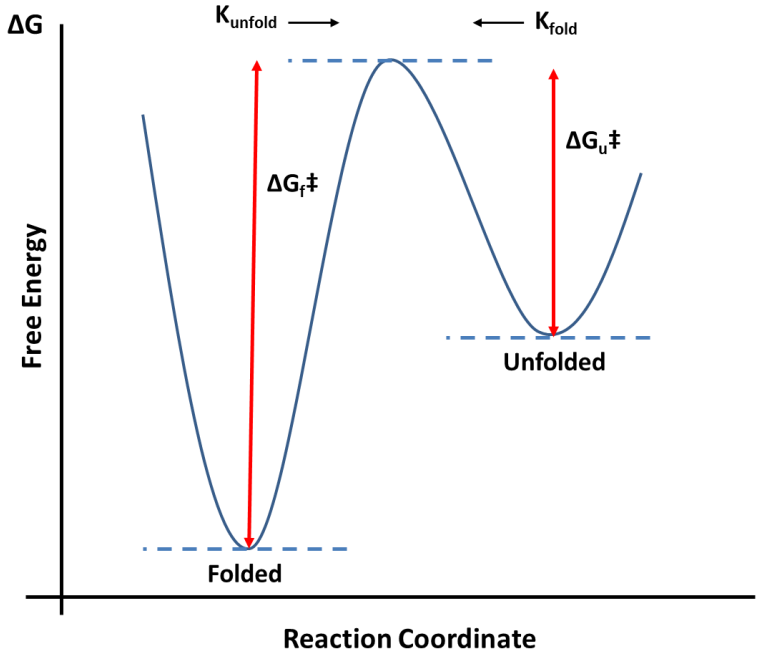

Figure 3: Graphical picture of Kramer's theory reaction rates. The folding and unfolding rates are calculated using$k = k_0 \exp(-\Delta G^{\ddagger}/kBT)$, .

For a simple two-state protein folding model, the folding or unfolding rate is independent of the initial conditions of the experiment and is determined by the free energy barrier after the change in solvent conditions that prompts folding and a pre-exponential term that is constant under all conditions. Basic assumption of two-state model is that the unfolded protein structure will follow one specific pathway and single energy barrier to reach to the lowest energy of the folded state. Many studies on the millisecond time scale showed that the folding reaction of many small proteins follows an exponential time course. For experiments using a change in chemical denaturant, a linear relation between the activation free-energy and the concentration of chemical

denaturant characterizes the two-state behavior, resulting in a V-shaped plot of the log- (relaxation rate) vs denaturant concentration, called a "chevron" plot (Figure 4). The relaxation rate (inverse of relaxation time) is the sum of the folding and unfolding constant rates:

$$k_{obs} = k_f + k_u, \qquad \text{(Eq. 3)}$$

The equilibrium constant is the ratio:

$$K_{eq} = \frac{k_u}{k_f}, \qquad \text{(Eq. 4)}$$

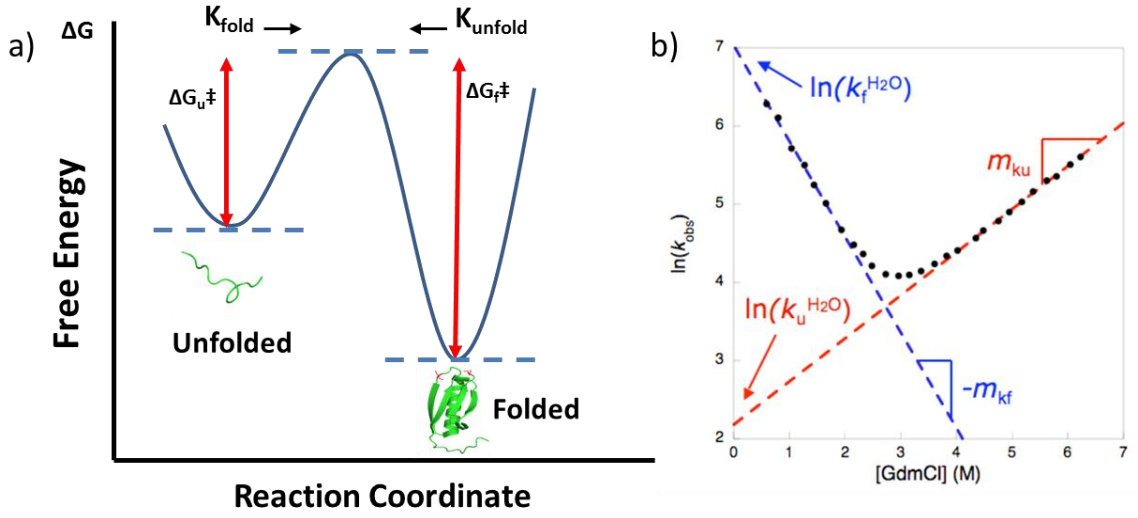

Figure 4: a) 2D view of the energy barrier and b) chevron plot[3] corresponding to the kinetics of the protein.

1.3. Folding Timescales

One reason studying the folding problem is a challenge is the enormous range of relevant folding timescales. The folding timescale ranges from nanoseconds to few seconds depending on the size and sequence of the protein (Figure 5). Chemical kinetic models, generally composed of

[3] http://ftp.aip.org/epaps/hfsp_j/E-HJFOA5-2-00780

two or three states, work quite well particularly for ensemble experiments to describe folding on millisecond and longer time scales. However, evidence shows that large deviations from those simple models can occur when ensemble measurements of folding are pushed below the millisecond time scale [31-35], when folding is followed with different chemical probes [26, 36], or when folding starts from different unfolded ensembles or starting conditions [37-40].

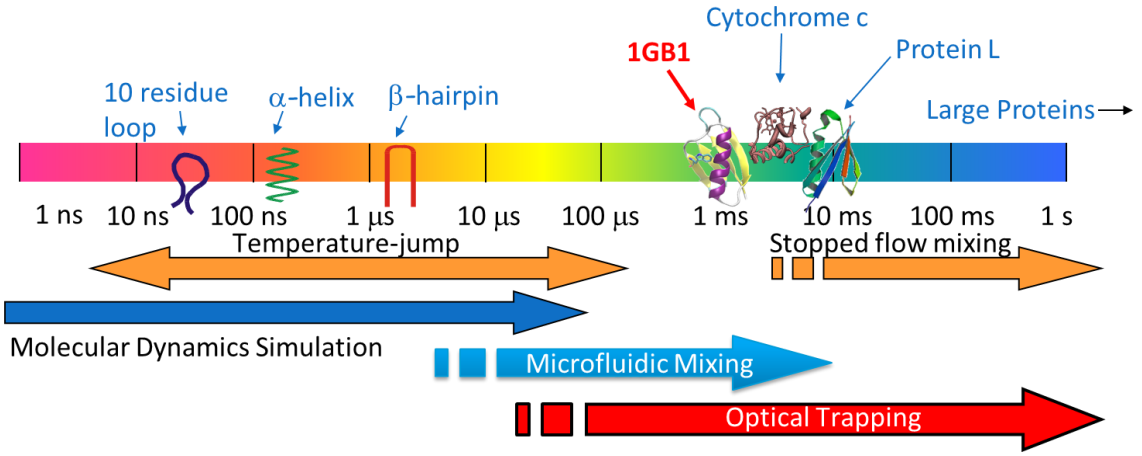

Figure 5: Timescale spectrum of folding events in proteins, showing a number of proteins studied so far. Arrows are showing different approaches used so far to understand one part of this huge spectrum.

Many years of effort has been put into studying folding but mostly at the slow phases, and only over the past fifteen years or so the technology has been available to look at faster timescales. To investigate complexity, we use microfluidic mixing that observe folding events in bulk at timescales as low as 8 µs [41]. The details of how we fabricate microfluidic mixers is discussed in chapter 2.

Most experimental techniques cannot access the whole timescale spectrum for folding events. They are not even capable of observing every amino acid simultaneously and they have to rely on ensemble methods that tend to average out complexity and limit modeling to 2- or 3-states. In other words, each of experimental techniques are different and there is no one technique that can survey the entire landscape. In this thesis, we are examining these two hypotheses: folding can be considered a linear sequence of events, and folding is parallel, multi-pathway processes. We used new experiments for measuring not just averages of structural observables, but also single molecules and we are using two different types of perturbations, chemical denaturant and single molecule force spectroscopy, to probe and re-examine these statements.

We are interested in studying small fast folding proteins because, first, they are computationally accessible and experimentally challenging and second, the study of small fast-folding proteins can provide key information for understanding and fundamental aspects of kinetics and mechanism of much larger proteins. Computational approaches are also available to simulate folding trajectories of small proteins, using molecular dynamic (MD) simulations, in full atomic details but just up to 100-millisecond timescale. We, as experimentalists, are trying to bridge MD simulations to experiment using our high-tech instruments that can measure folding events happening on microsecond timescales.

1.4.B1 domain of protein G (GB1)

Protein G is a large multi-domain cell surface protein of group G Streptococcus and its role in the cell is to help the organism evade the host defenses through its protein binding properties [42]. In this thesis I study single molecules of the B1 domain of protein G [43], using force as the probe. See Figure 6 for secondary structure of this domain. This small protein is one of the best-studied systems in protein folding and has been studied for two decades with bulk measurements.

GB1 has 56 amino acids in its sequence and it is small and therefore accessible to different computational modeling. It also has mixed secondary structure, high stability and slow folding kinetics that make it comparable with much longer proteins. Previous measurements on the milliseconds timescale show that this protein folds in 1ms under standard conditions and it follows a two-state system model [44, 45]. Those measurements were limited to following Trp fluorescence kinetically on the millisecond time scale and NMR analysis in equilibrium [43, 46]. Those experiments revealed basic, out-of-equilibrium force-ramp measurements of unfolding and folding of GB1 versus force. GB1 folding is very well approximated as two-state system folder, on millisecond time scales or faster [44, 45, 47, 48].

On the other hand, there are some measurements that show nonlinearity in their chevron plot and this is a sign that GB1 can have intermediate states [49, 50]. Moreover, Lapidus group and Pande group, reexamined the folding pathway of this domain using microfluidic mixers and Markov State Model (MSM) predictions [26]. Folding work was observed on the microsecond timescale using number of probes including circular dichroism, photochemical oxidation and tryptophan fluorescence. Each probe provided different kinetics. Comparing those observations with MSM predictions confirmed the existence of many folding pathways before the final native state and that those paths did not have large free energy barriers. See Figure 7 for the predicted energy landscape.

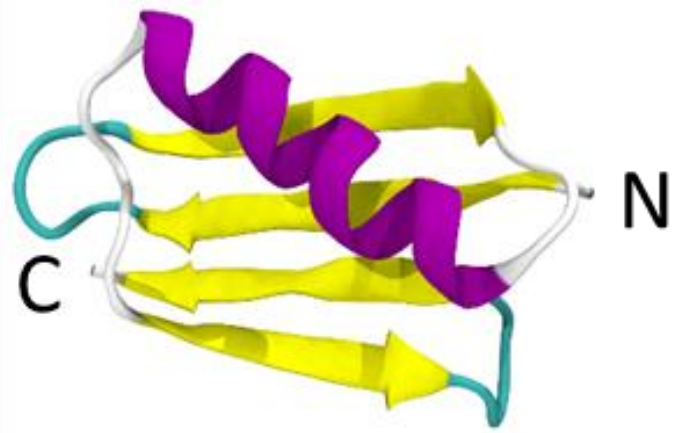

Figure 6: Structure of model system for GB1 with NC termini labels. GB1 has a mixed secondary structure with α-helix in the middle and some β-sheets.

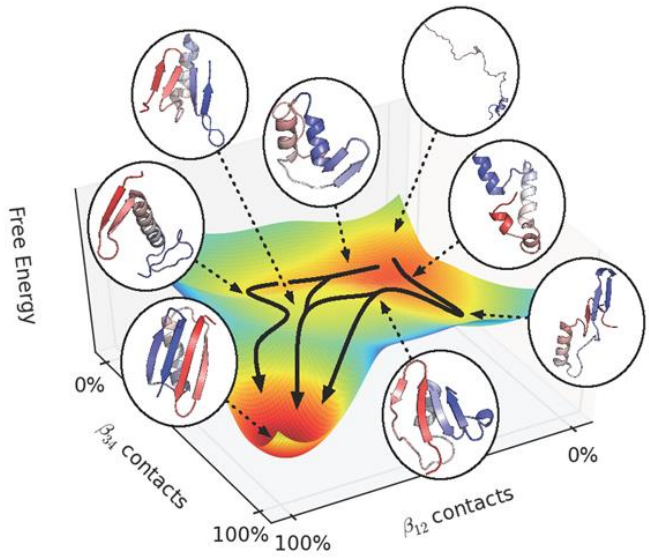

Figure 7: Protein folding potential energy landscape shown for GB1, predicted by Markov State Model, adopted from [26].

1.5.Introduction to Single Molecule Force Spectroscopy

Single molecule force spectroscopy (SMFS) is a powerful technique for studying biological processes such as protein synthesis, protein folding or ribosomal motion [51]. The main categories of SMFS are optical tweezers, atomic force spectroscopy (AFM) and magnetic tweezers [52]. Main goal in folding studies is to probe the free energy landscape of folding/unfolding directly via direct measurement of individual molecules. Figure 8 shows schematic view of dual optical trap, trapping two beads, used in our lab. The application of force can both tune and probe the landscape, which will be discussed in detail in this thesis. Therefore, in some ways it is the same for the chemical denaturant bulk methods. However, since force unfolds the protein in a direction, the unfolded state under force is likely much different than the unfolded state under chemical denaturant. Another advantage is that SMFS directly measures extension, which is a well-defined reaction coordinate for constructing a free energy landscape.

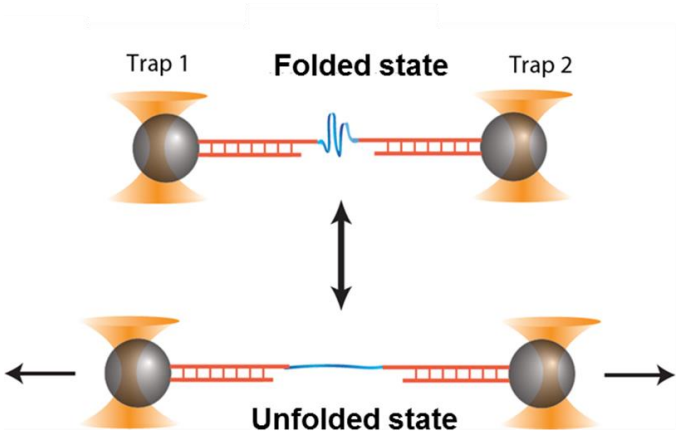

Figure 8: Schematic view of dual optical trap, trapping two beads. The beads are holding a tether, which consists of two double stranded DNA handles and a single protein molecule in the middle. The top scheme shows the trap before the force is applied in which the peptide is in the folded state and the bottom scheme shows the tether with force applied on it and the protein unfolded.

Application of single molecule techniques for studying protein folding started in 1997 with the AFM work of Gaub and laser tweezers work of Bustamante on individual titin molecules, [53, 54]. Titin is a giant polypeptide, which plays an important role in maintaining the structural integrity in muscle. It is designed to withstand force in the muscle rather than heat or chemical environment change. Since its mechanical properties is important for its biological function, the idea of studying it under force arose. Since then, single molecule techniques have been applied to many other proteins including fast folding proteins. First studies were mostly limited to fluorophore measurements and since fluorophores have nanosecond time scales, the number of excitation-emission cycles was limited [55]. By analyzing data from high–time-resolution SMFS measurements, the average transit time across transition paths, τ_{tp}, was found for two small proteins, α_3D and a WW domain [56, 57] and for the protein GB1 [56, 58] and a DNA hairpin [59]. Eaton group studied transition paths, which is a property that can only be studied by single-molecule methods, for protein GB1 and their analysis shows a transition path time of <200 μs. This path time is >10000 times shorter than the mean waiting time in the unfolded state. Their measurements, along with Szabo's theory for diffusive transition paths [60], predicts the transition path time to be insensitive to the folding rate, with 2-fold difference for rate coefficients that differ by 10^5- fold. They did not have enough resolution at the time to resolve transition path times. The fast folding cold shock protein B (Csps), which was considered to be a simple two-state kinetic model, was FRET labeled and studied by single molecule methods first in 2004 [61]. Due to the instrumental limitations, this group could not resolve the actual transition steps between folded and unfolded states. However, when the single molecule experiments have high time resolutions as they do these days, they can observe the actual transit time instead. Cold shock protein B was studied recently, using single molecule force spectroscopy , under moderate pulling forces using

force-clamp AFM [62]. In this nanoscale study, the unfolding of Csps shows heterogeneity with trajectories that range from single short-lived intermediate state to different combinations of multiple long-lived intermediates. As a result, using a single sequential pathway on the energy landscape of Csps is not capable of explaining such a heterogeneity. Rief group was able to monitor and then analyze the rapid folding transitions of a single Villin headpiece (HP35) for up to 15 minutes and with 300 kHz bandwidth [63]. Their measurements show a difference in folding pathway and cooperativity between wild type and stable variant of headpiece 35. Woodside group reconstructed the full landscape profile for the prion protein PrP from force-extension curves. They show a double-well potential with an extended, partially unfolded transition state. This means that the native folding pathway involves only two states, without evidence for partially folded intermediates that had been proposed to be the cause for misfolding. This group also applied SMFS to observe the misfolding of the prion protein PrP directly and some transitions were observed into off-pathway intermediates. They observed three different misfolding pathways, all starting from the unfolded state and the misfolding rate was higher than the rate for native folding. Therefore, they were able to map out non-native folding pathways for prion protein using SMFS methods [51, 64]. Perkins group reexamined the unfolding of individual bacteriorhodopsin molecules in native lipid bilayers. Their results show newly detected intermediates in the unfolding pathway, indicating complex dynamics, including frequent refolding [65]. Protein src SH3 is another example that unfolds under multiple folding pathways at single molecule level [66-70]. Marqusee group has shown recently for single-domain proteins of src SH3 [70] that unfolding involve a switch in the pathway as force is increased from a low to a high value. This suggests that the underlying energy landscape for those proteins should be strongly multidimensional. Pierse and

Dudko have recently investigated theoretically the types of observations expected for protein unfolding with multiple pathways and/or intermediates [71].

Figure 9 shows a graphical picture of a two-dimensional energy landscape plotted for the force perturbation, which is interpreted as the change in the extension of the protein. The force's effect on the protein's extension affects the unfolded protein mainly, simply because it is a polymer. The folded state likely is not changing extension with force. However, in either case, whenever we move along the extension reaction coordinate (both for folding and unfolding) the trap does work, and force acts as perturbation, which shifts the energy landscape and thus affects the rates. The main effect is the change in the relative transition state/barrier height. When force is applied to a single protein molecule, the force will do work on the protein and as a result the end-to-end extension of the unfolded protein is different from zero force and the energy landscape is slightly tilted on the unfolded potential well side. This also results in some changes to the two-state equations, which is shown in the following equations:

$$k_{fold} \sim Ae^{\frac{-(\Delta G^{\ddagger} + x_F^{\ddagger} \times F)}{k_B T}}, \quad \text{(Eq. 5)}$$

$$k_{unfold} \sim Ae^{\frac{-(\Delta G^{\ddagger} - x_u^{\ddagger} \times F)}{k_B T}}, \quad \text{(Eq. 6)}$$

$$\log k_{fold} \sim \log\left(Ae^{\frac{-\Delta G^{\ddagger}}{k_B T}}\right) - \frac{x_F^{\ddagger}}{k_B T} \times F, \quad \text{(Eq. 7)}$$

$$\log k_{unfold} \sim \log\left(Ae^{\frac{-\Delta G^{\ddagger}}{k_B T}}\right) + \frac{x_u^{\ddagger}}{k_B T} \times F, \quad \text{(Eq. 8)}$$

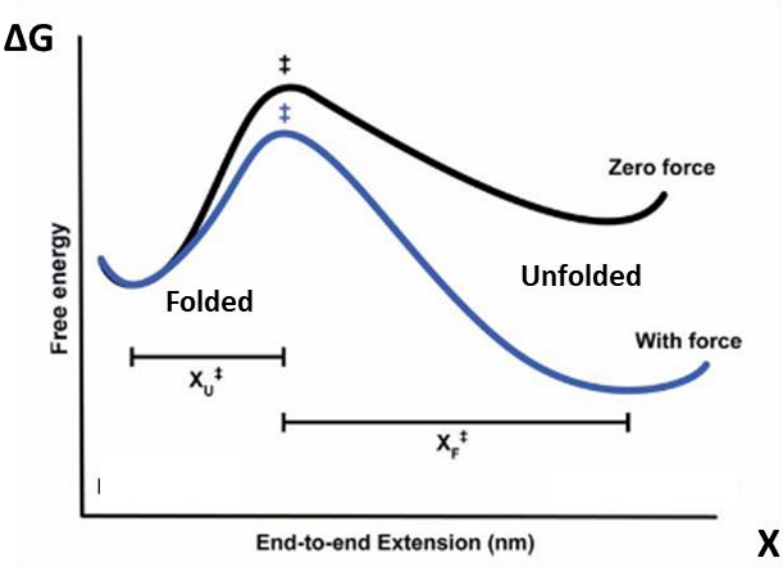

Figure 9: Graphical picture of the two-dimensional free energy landscape of a protein: at zero force (black curve), and when force is applied (blue curve). As shown here, whenever we move along the extension reaction coordinate (both for folding and unfolding) the trap does work, which shifts the energy landscape and thus affects the rates. Plot adopted from [72].

In this thesis, we are using force as our perturbation and probe to study the folding kinetics of GB1 under force. This represents a new initial condition for folding molecules of GB1 and allows us to compare the results with bulk measurements performed with mixing and lead to a better understanding of folding pathways of this domain.

CHAPTER 2

MICROFLUIDIC MIXERS

In this chapter, we present a method to fabricate a fused silica microfluidic device by employing low viscosity KMPR photoresists. The resulting device is a continuous-flow microfluidic mixer based on hydrodynamic focusing. The advantages of this new fabrication method [73] compared to the traditional approach [41] are simplification, and time and cost reduction, while still preserving the quality and the performance of the mixers. This process results in devices in which the focusing channel has an aspect ratio of 10:1. The newly fabricated mixer is successfully used to observe the folding of the Pin1 WW domain at the microsecond time scale. The work in this chapter has been published as:

Izadi, D., T. Nguyen, and L. Lapidus, Complete Procedure for Fabrication of a Fused Silica Ultra-rapid Microfluidic Mixer Used in Biophysical Measurements [73].

All sorts of environmental variables have been used over the years as a folding prompt to study protein folding either in bulk or in single molecule measurements, for examples rapid changes in pH [74], pressure [75], temperature [76] or physical manipulation techniques such as laser or magnetic tweezers, atomic force microscope or optical tweezers. Although these methods cover different time ranges, computer simulations are generally able to only measure up to microsecond timescales with picosecond or greater resolution. Ultrafast laser spectroscopy measures up to microseconds with picosecond resolution [77]. Ultrafast measurements are typically limited by the available experimental prompts such as T-jump and photo-physical triggers. The most generic prompt for most proteins is the dilution of chemical denaturant,

requiring a fluidic mixer. Stopped-flow mixing experiments can measure folding following a dead time of about one millisecond. Ultra-rapid mixers have a mixing time of about 5-10 μs and can measure folding up to the first few milliseconds.

The initial solution contains the protein and a high concentration of denaturant, typically 6M GdnHcl or 8 M Urea. In this environment, almost all protein molecules are fully unfolded. This solution is then rapidly diluted with a buffer solution with no denaturant, which drops the total denaturant concentration below the threshold concentration for unfolding, and proteins start to refold.

Mixers are divided into two main groups based on their design and on how the liquid flows in the device: stopped-flow mixers and continuous-flow mixers. Conventional stopped-flow mixers dispense a defined volume of two liquids into an observation volume, and have "dead times" of 1–5 ms, largely limited by turbulence. In our lab, we use continuous-flow mixers in which solutions flow continuously at a constant flow rate. In these type of mixers, the observations taken at different positions along the channel corresponding to different times passed after mixing.

For laminar flow Stokes equation is derived from Navier-Stokes equation:

$$\rho \frac{D\vec{V}}{Dt} = \rho\vec{B} - \nabla p + \mu\nabla^2\vec{V},\ \nabla\cdot\vec{V} = 0, \qquad \text{(Eq. 9)}$$

In the above relations, μ is the dynamic viscosity and p is the thermodynamic pressure. In order to be in the laminar flow regime and avoid turbulent flow, Reynolds number needs to be small (<100). Reynolds number is actually defined as the product of density times velocity multiplied by the length divided by the viscosity coefficient.

$$R_e = \frac{\rho \upsilon L}{\mu}, \qquad \text{(Eq. 10)}$$

This is proportional to the ratio of inertial forces and viscous forces (forces resistant to change and heavy and gluey forces) in a fluid flow. Flowing fluids normally follow along

streamlines. If a flow is laminar, fluids will move along smooth streamlines and all motion is symmetric in time; consequently, all motion is reversed if the applied pressures are reversed. If the flow is turbulent, these streamlines breakup and the fluid will move in an irregular manner.

The hydrodynamic focusing mixer first introduced by Brody et al. [78] is one of the first designed that uses laminar flow and its first design had a minimum feature size of 1-2 um and was a cross shape device having four channels meeting at the center. Since then, the mixer has been optimized and used intensively in the field of protein folding due to its extremely low mixing time, which is on the order of microseconds [32, 41, 78-86]. In the initial design, the protein denaturant solution enters from the top channel and meets with the buffer coming from two side channels and the two solutions are mixed at the center and exit through the bottom channel. The fluid requirements of this device are also very low compared to turbulent devices, with protein consumption in the order of nL/s. The design described in this work has a mixing time of 8 μs (as defined as an 80% decrease in concentration [87]), one of the fastest reported to date, and has been measured experimentally and confirmed by simulation of the device by finite element analysis [38, 39, 80, 88].

Our mixer design, Figure 10, consists of a sample stream entering the mixing region through a central nozzle of width w_c. The stream is constrained symmetrically from sides by two side channels of width w_s. In this design, flow is approximated with two-dimensional (2D) flow simulations. The non-dimensional Navier-Stokes equation of motion are:

$$\frac{\partial u*}{\partial x*} + (\alpha\beta)\frac{\partial \upsilon*}{\partial y*} = 0, \qquad \text{(Eq. 11)}$$

$$\mathrm{Re}\left(u*\frac{\partial u*}{\partial x*} + \alpha\beta\upsilon*\frac{\partial u*}{\partial y*}\right) = -\frac{\beta}{\alpha}\frac{\partial P*}{\partial x*} + \frac{\beta}{\alpha}\left(\frac{\partial^2 u*}{\partial x*^2} + \alpha^2\frac{\partial^2 u*}{\partial y*^2}\right), \qquad \text{(Eq. 12)}$$

$$\mathrm{Re}\left(u*\frac{\partial u*}{\partial x*} + \alpha\beta\upsilon*\frac{\partial u*}{\partial y*}\right) = -\frac{\partial P*}{\partial y*} + \frac{\beta}{\alpha}\left(\frac{\partial^2 \upsilon*}{\partial x*^2} + \alpha^2\frac{\partial^2 \upsilon*}{\partial y*^2}\right), \qquad \text{(Eq. 13)}$$

where $\alpha = w_c/w_s$, $\gamma = w_e/w_c$, $\beta = Us/Uc$, and Reynolds number is defined as:

$Re_w = U_s w_s / v$, (Eq. 14)

where U is the maximum velocity, w is the sample nozzle width, and v is the kinematic viscosity of the solution. u and v are scaled using Uc and Us, respectively, and pressure is scaled as Ucμ/wc, where μ is the dynamic viscosity. The subscripts s, c, and e refer to the side, center, and exit channels. The full Navier–Stokes formulation is valid for Re_w as high as 15 [88].

Some physical limitations exist that affect optimization constraints. Those limitations include the lithography step in fabrication, which does not allow feature size of smaller than 1-2 micrometers. Another issue would be contaminants left over from the fabrication process, which can be present in the buffers and clog the side channels at the nozzle. Clogging is kept to a minimum with filter post spacings of 1–2 μm. The depth of the channels is also limited to ~10 μm to optimize the fluorescence signal with a confocal system. Moreover, the physical properties of buffers and denaturants limits the maximum aspect ratio (aspect ratio = width/depth) of the regions of interest to less than one. This aspect ratio constraint makes 2D fluid flow approximation possible at the vertical cross section of the mixer.

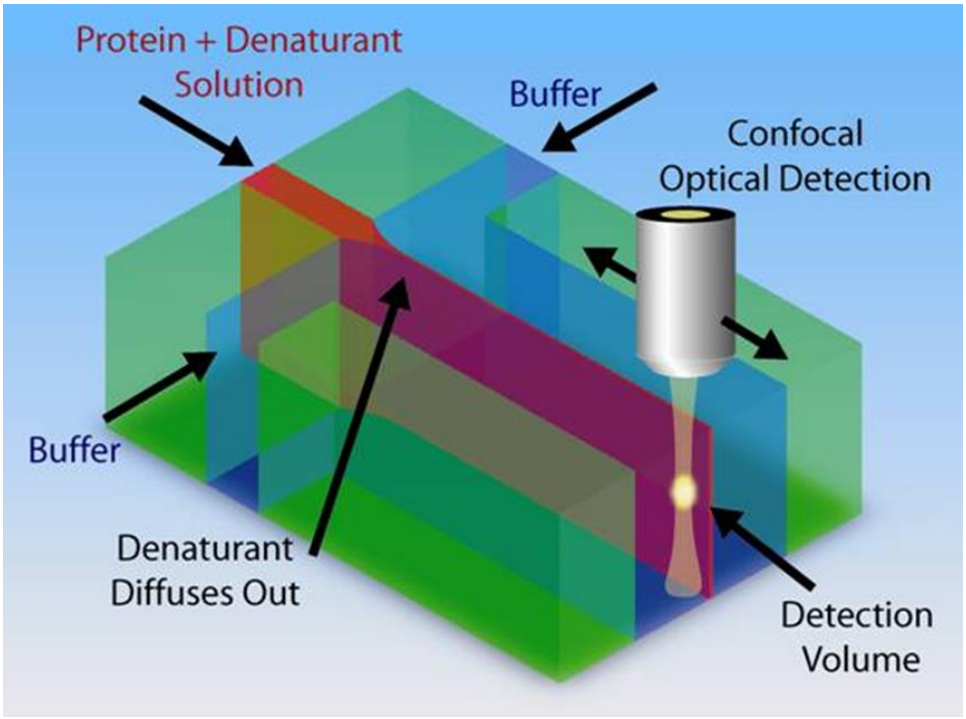

Figure 10: Schematic view of mixing experiment set up with T-mixer. Protein dissolved and placed in the center channel. The geometry constricts the protein flow into a narrow jet ~100nm wide. Molecules of denaturant are light and small and their diffusion is fast. Protein jet flows down at a constant rate and fluorescence during folding can be observed using scanning microscope. Image adopted from [89].

Features on the mixer device are etched using a photolithography process at a Micro-Electro-Mechanical System (MEMS) fabrication facility. Since protein folding in a microfluidic mixer requires optical observation of the sample inside the mixing and downstream regions, these devices have been fabricated from transparent materials, such as borosilicate glass for visible fluorescence and fused silica for UV fluorescence. In our lab, this method was originally based on utilizing a polycrystalline-silicon (poly-Si) mask. A thin layer of poly-Si is deposited on the fused silica surface in a low-pressure chemical vapor deposition (LPCVD) furnace. The mask pattern is imprinted on the poly-Si surface by photolithography. Then the pattern is etched into the poly-Si by deep reactive ion etching (DRIE). The poly-Si then functions as a mask during the fused silica

etching step. Once the fused silica is etched, the poly-Si is removed via a xenon-difluoride etcher. This procedure requires highly specialized cleanroom machines for deposition, etching, and removal of the poly-Si mask and, consequently, is expensive and time consuming. Additionally, it is not feasible to achieve a poly-Si layer of more than about three μm thickness due to stress on the wafer. As a result, the deepest channels achieved were 38 μm [90]. To overcome these disadvantages, I started working on developing a new fabrication method which not only is cost effective and less time consuming, but it is also reproducible and reliable. This fabrication method will be explained in detail in this chapter. Recently, It was reported that high viscosity KMPR photoresist (KMPR 1025) can be used to etch fused silica [91, 92] but no completed devices have been made. Our new fabrication method for a complete microfluidic mixer uses low viscosity KMPR (KMPR 1005 and KMPR 1010), which has not been previously tried for fused silica fabrication. This new protocol allows us to etch deeper than 40 microns in fused si, which was not feasible with the old protocol. The resulting channels in the mixers are of high quality with an aspect ratio as high as 10:1. The unetched surfaces remain atomically smooth, suitable for fusion bonding. These advantages make this fabrication method affordable and simple enough for use by researchers in the field of protein folding, in particular, and in the biotechnology community, in general.

2.1. Microfabrication

The fabrication of these microfluidic mixers was done in Robert H. Lurie Nanofabrication Facility at the University of Michigan and the Keck Microfabrication Facility at Michigan State University

2.1.2. Chemicals and Materials

KMPR 1005 and KMPR 1010 were obtained from MicroChem Corp. (Westborough, MA, USA). Four-inch (0.525 and 0.170 mm thick) fused silica wafers were purchased from Plan Optik Company (Elsoff, Germany). KMPR 1005 for the fabrication of the hydrodynamic focusing mixer (T-mixer) which should eventually have the channel depth of ~10 μm. For the demonstration of deep etching, we use KMPR 1010. Figure 11 shows the SEM image of uniform KMPR 1010 layer after developing. The thickness of this layer is ~17 μm measured using a Dektak 8M Surface Profilometer (Bruker Corporation, Billerica, MA, USA). Figure 12 shows the SEM of the etched fused silica channel after removing KMPR 1010. The depth of this channel is ~45 μm as measured by the Dektak 8M Surface Profilometer (Table 1).

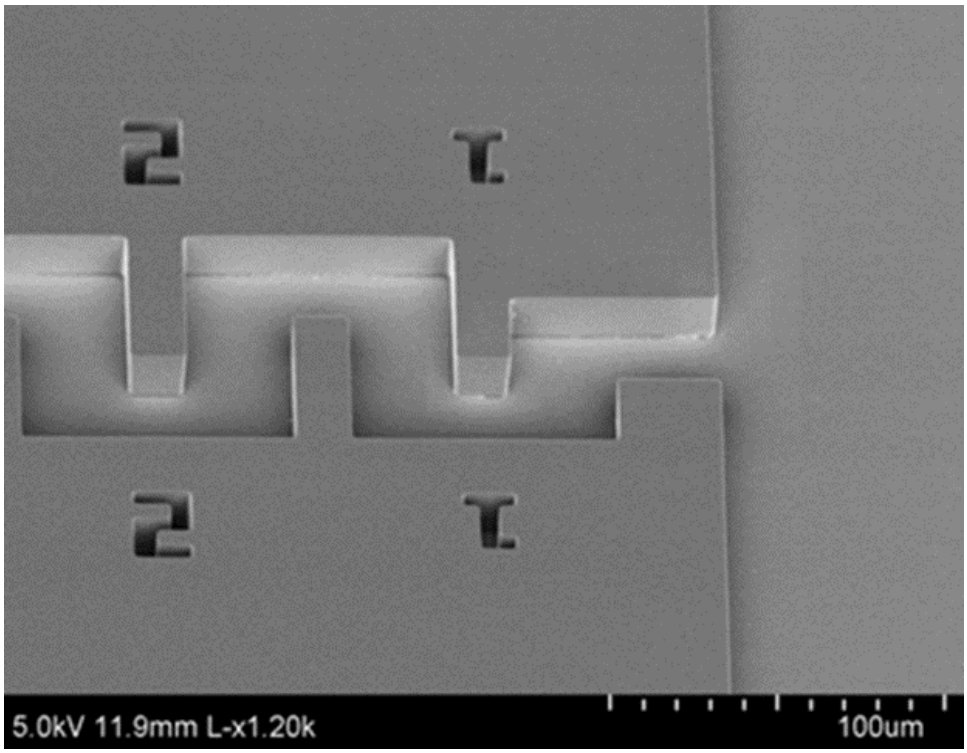

Figure 11: SEM image of KMPR 1010 layer after developing on the fused silica.

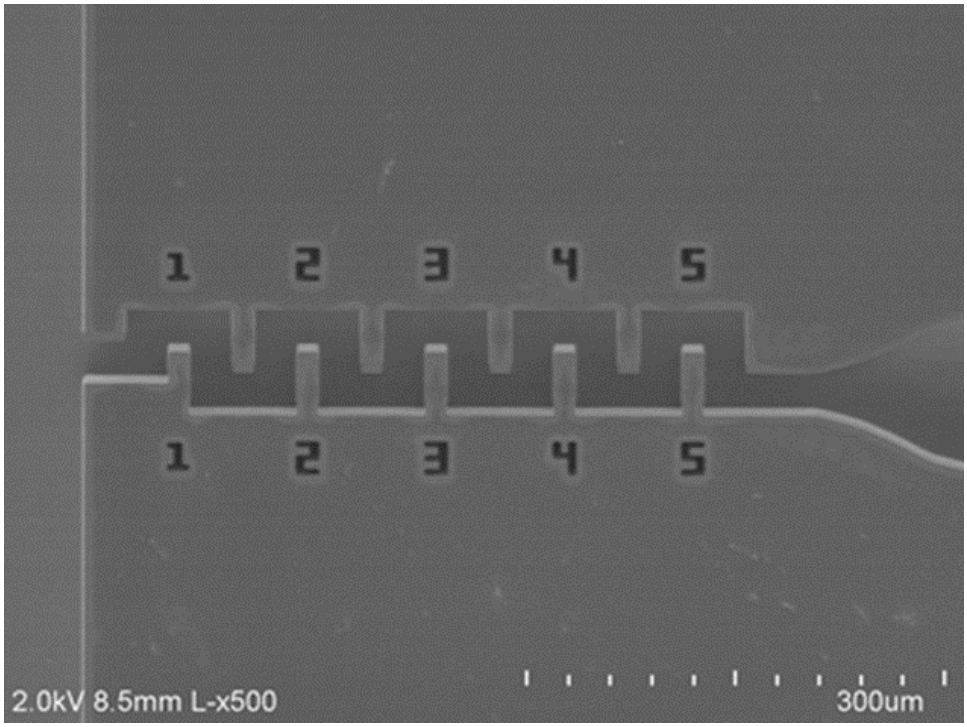

Figure 12: SEM image of etched fused silica channels after removing KMPR1010

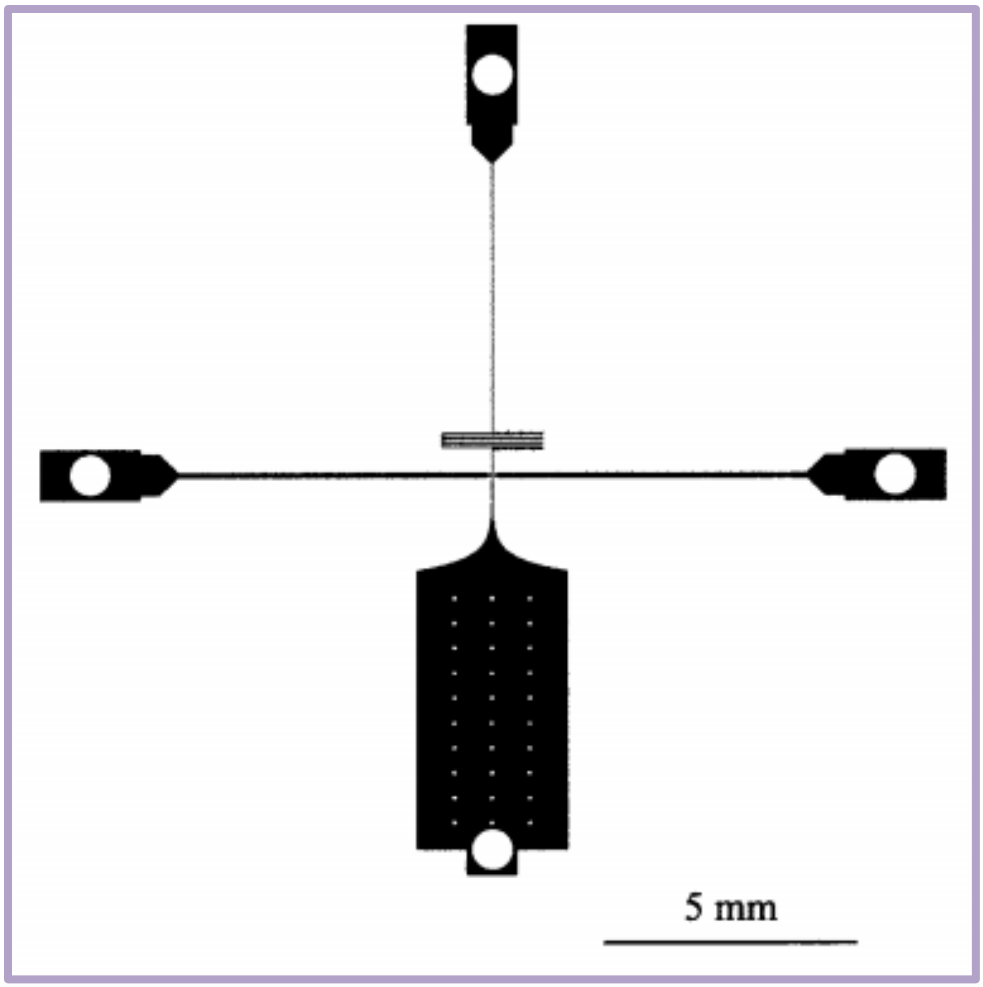

Figure 13: Lay out of individual T-mixer chip

2.1.3. Process Flow

The chip consists of two fused silica parts permanently bonded together. The top wafer (a) has a thickness of 170 micron. The bottom wafer (b) has a thickness of 525 micron. The following process (except step (g)) is for wafer b. See Figure 14.

2.1.4. Cleaning

The fused silica wafers were cleaned with Nano-strip, a stabilized formulation of sulfuric acid and hydrogen peroxide (cyantek.com), for 10 min at 60 °C. Wafers are then rinsed in de-

ionized water over four cycles in a DI Quick Dump Rinse tank for 5 min. The wafers are finally dried with nitrogen gas.

2.1.5. Photoresist Coating

KMPR was spin-coated on the wafer using an ACS200 cluster tool (SUSS MicroTec, Munich, Germany). The viscosity of KMPR 1005 is 95 centistokes (cSt). The ACS200 cluster tool performed a soft-bake at 100 °C for 7 min before dispensing for three s at 4500 rpm.

2.1.6. Photolithography

A negative photolithography mask was fabricated from Compugraphics USA (Fremont, CA, USA). A Karl Suss MA-6/BA-6 (SUSS MicroTec, Garching, Germany) contact aligner was used to expose the mask and substrate to UV light. This aligner uses a wedge error compensation (WEC) head to apply 1.0 bar pressure for high contact between the mask and the wafer. The UV source emits 20 mW/cm^2 at 405 nm. The exposure time is 5.5 s. After exposure, the substrate is baked at 100 °C for 7 min. The image of the mask will be visible in the KMPR, confirming there was sufficient exposure.

2.1.7. Developing

Development of the pattern uses AZ-300 MIF developer (EMD Performance Materials, Somerville, NJ, USA) for 135 s. After development, the sample was spray rinsed for about 60 s with de-ionized water and dried with nitrogen gas.

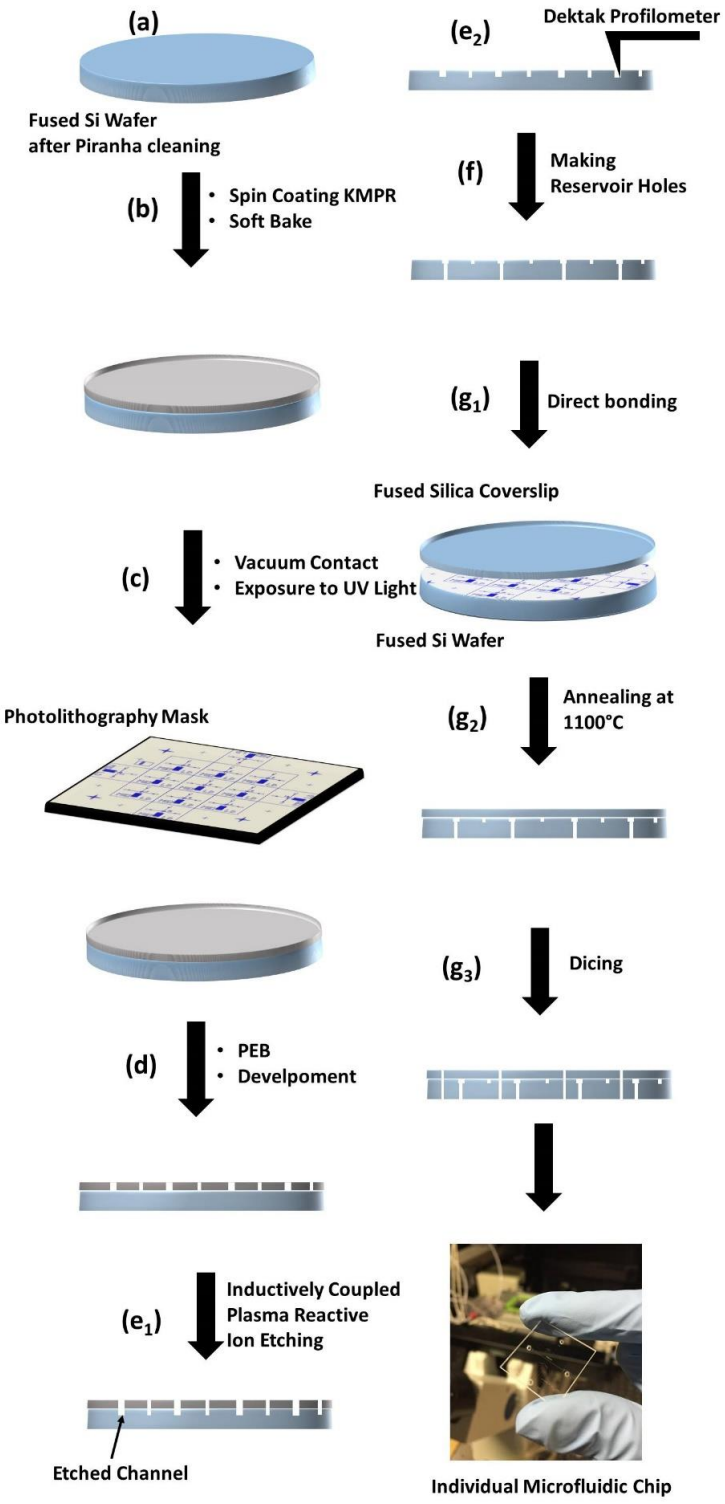

Figure 14: Schematic view of fabrication process for Tmixer design

2.1.8. Etching

The channels were etched using reactive ion etching optimized for oxide substrates. The wafer was mounted on a six-inch silicon wafer. This silicon wafer provides mechanical support to the substrate. The fused si wafer is then etched with the SPTS APS Dielectric Etch tool (STS Glass Etcher, SPTS technologies, Newport, UK). The etch rate was 0.50 μm/min. Therefore, 10 μm deep channels required 20 min of etching. The remaining KMPR was removed with REMOVER PG (MicroChem Corp., Westborough, MA, USA), a proprietary N-methyl-2-pyrrolidone-based solvent. The wafers were immersed for 30–45 minutes at 65–80 °C and then left to soak in the solvent overnight. The wafers were rinsed with isopropyl alcohol and de-ionized water, and the channel depths were confirmed with a Dektak® profilometer (Bruker Corporation, Billerica, MA, USA) measurement.

2.1.9. Sandblasting

A high-precision sandblaster (Crystal Mark Inc., Glendale, CA, USA) was used to create the openings, which allowed fluid flow into the channels from the sample reservoirs sealed by O-rings to the back of the mixer. The aluminum powder was comprised of 27.5-micron diameter particles.

2.1.10. Bonding and Dicing

Prior to bonding, both wafers A and B are cleaned with RCA reverse cleaning for a total of 100 min. Then the two wafers are placed together and a pressure from a single finger will cause the pre-bonding front to expand out to the perimeter. The wafer is then baked at 1100 °C for 2 h and the individual chips diced using a laser engraver (VLS2.30, Universal Laser Systems, Scottsdale, AZ, USA).

Figure 15 shows the SEM of a uniform layer of KMPR 1005 after the developing step. The thickness of this layer is measured by using Dektak 8M Surface Profilometer and is ~seven μm. Figure 15b shows the mixing region after etching. At the end of the center channel, as shown in Fig. 2b, the width of the focusing nozzle J is ~ 1 μm (the blue line in Figure 15b). The depth of the channel, determined using Dektak, is ~10 μm. These measurements indicate that, by using KMPR 1005 photoresist, we obtain in the focusing region an aspect ratio of 10:1, which is comparable to previous work using a poly-Si mask [41], while requiring fewer fabrication steps, hence reducing the experimental time and cost.

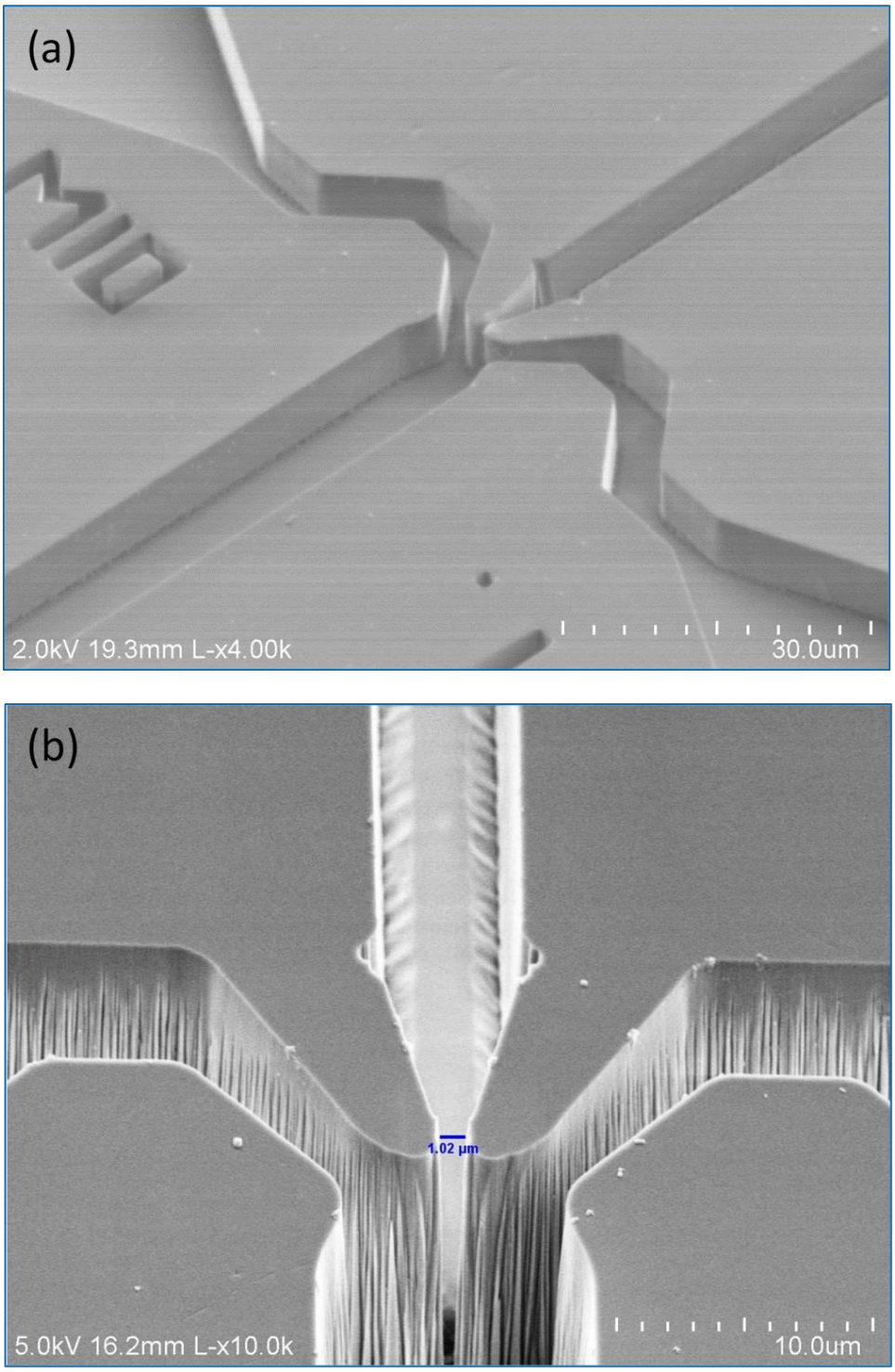

Figure 15: (a) SEM of a uniform KMPR 1005 layer after developing step and (b) SEM of the etched fused silica.

The mixing time is determined, in part, by the width of the nozzles in the mixing region. This feature is limited to ~1 μm by the resolution of our lithography. However, a narrow nozzle also makes the center channel prone to clogging by particle contamination. The chip design features a filter region made of a series of fused silica micron-sized posts to catch particulates within the flow streams (see Figure 16), and cleaning with Piranha solution can dissolve most organic particulates. However, this system does not catch one common type of particulate on the small, irregular shards of fused silica created during the drilling of inlet holes (Figure 17) where the fluid enters the channels. These shards tend to stick strongly to the inside of the hole and are difficult to remove by washing or air pressure prior to bonding. However, continuous flow during use of the chip will eventually dislodge such shards that may be caught in the mixing region, frequently disabling the device. We, therefore, have improved another step of the fabrication method, using a sandblaster to create the openings instead of a diamond-tipped drill. Sandblasting with 27.5 μm aluminum oxide powder does not leave shards of fused silica in the resulting holes (Figure 19) and, therefore, improves the reproducibility of fabricating these chips.

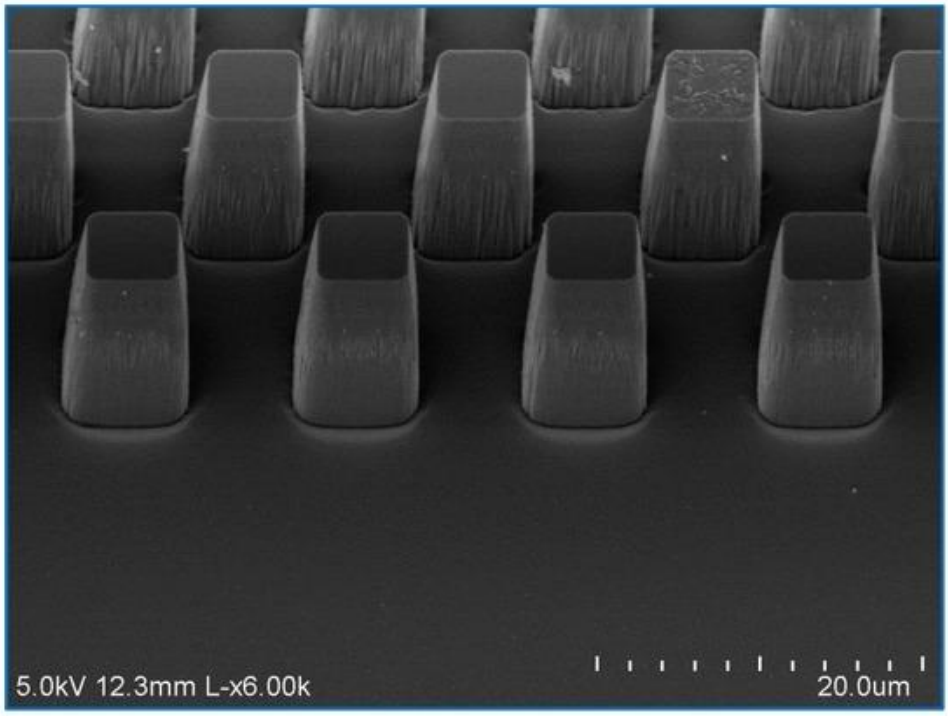

Figure 16: SEM of a filter region made of a series of fused silica micron-sized posts.

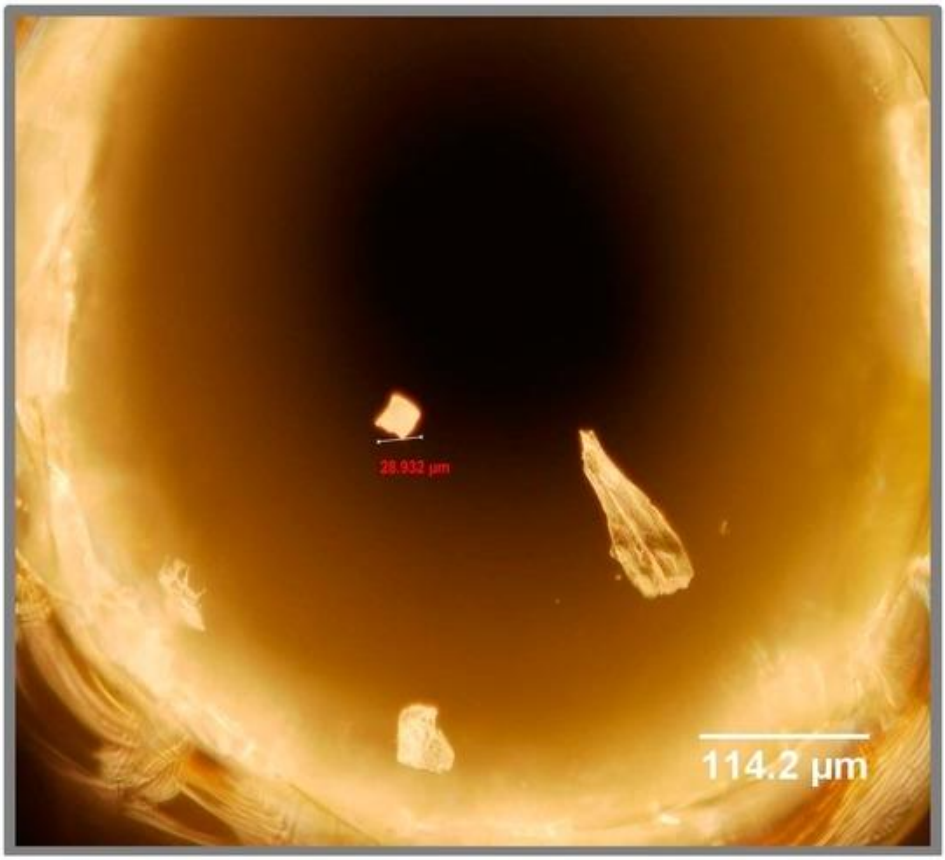

Figure 17: Microscope picture of the inlet hole made from a drilling machine with resulting shards falling off the fused silica surface.

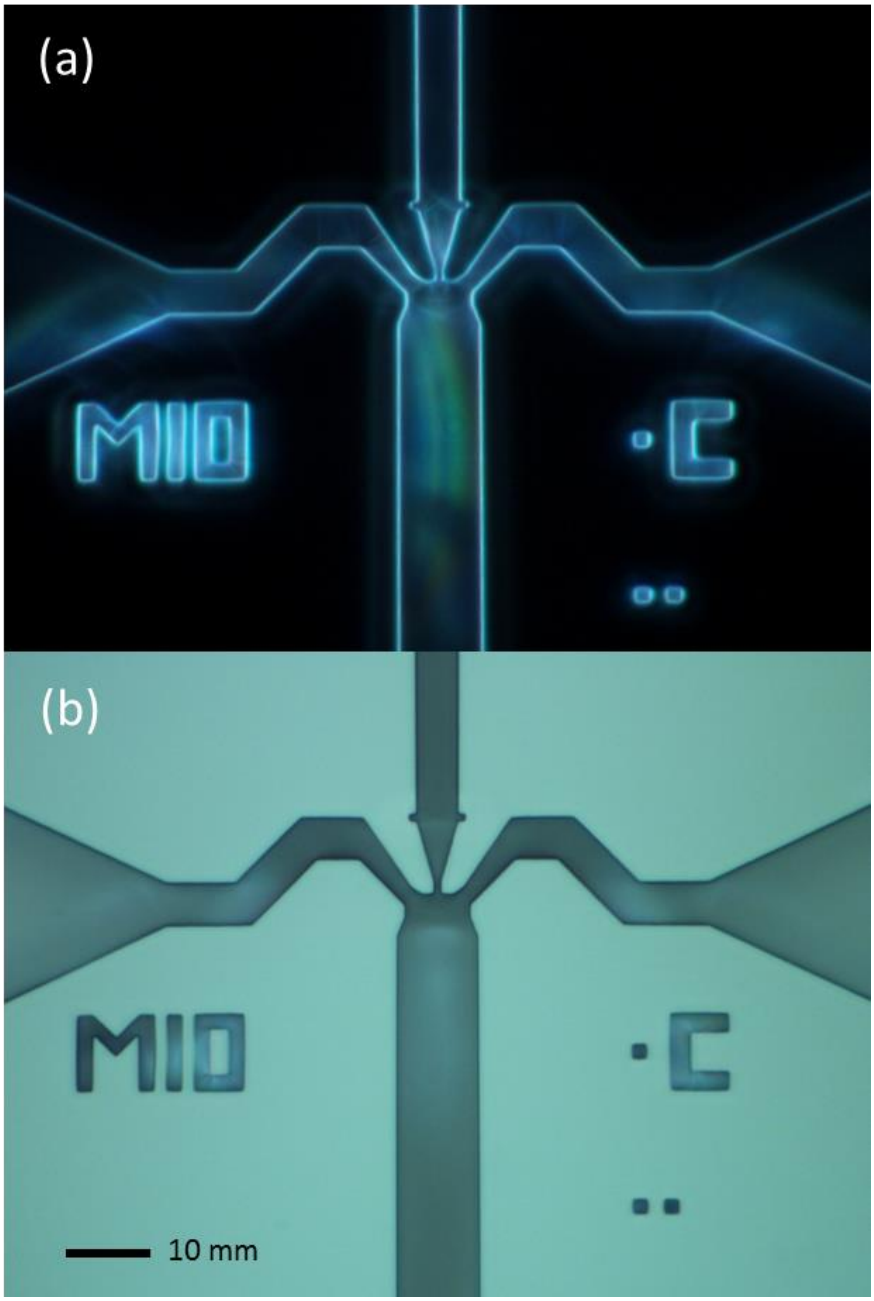

Figure 18: Optical microscope image of a Tmixer taken at KMF

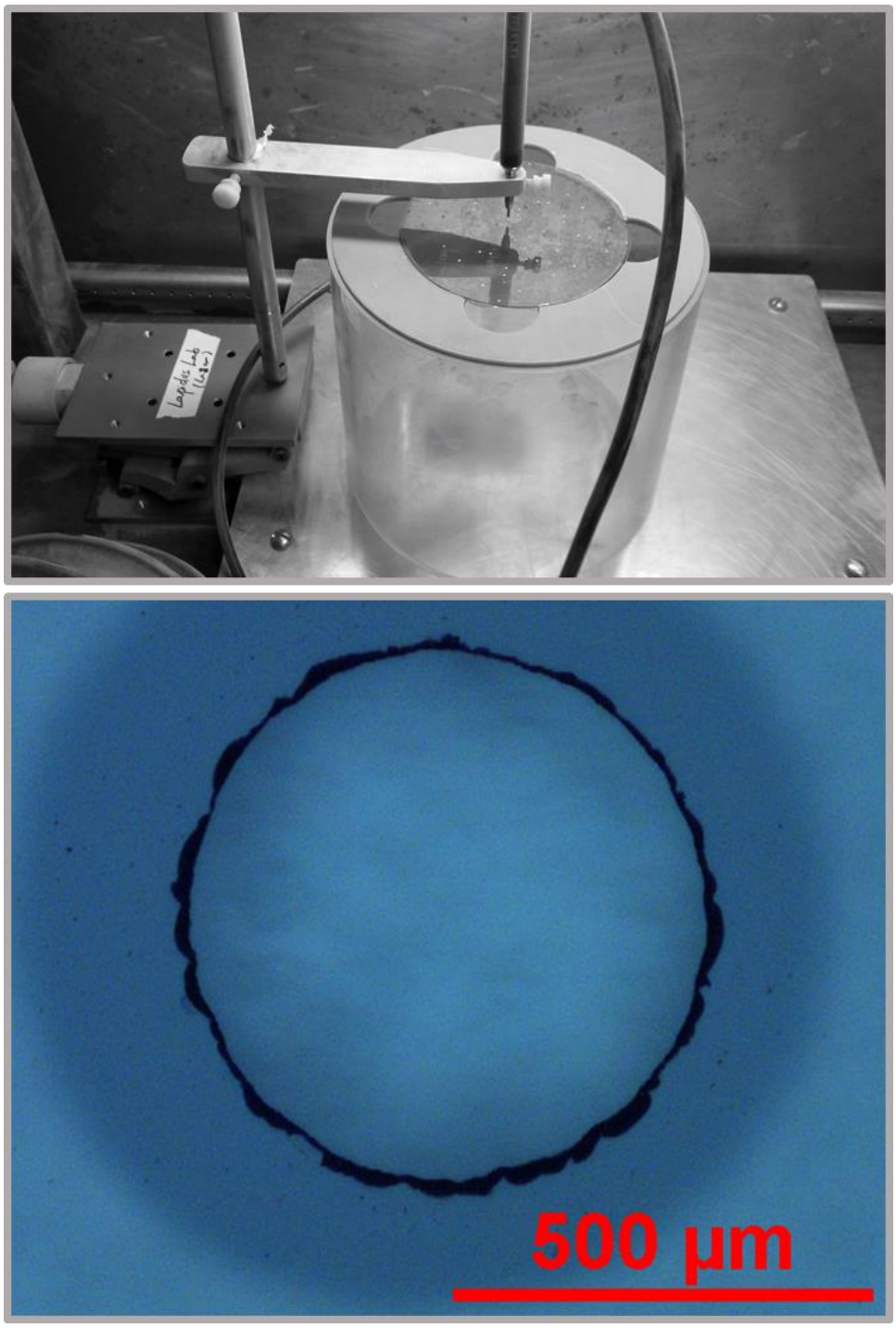

Figure 19: (Top) sandblaster set-up. (Bottom) microscope picture of the inlet hole made from the sandblaster.

2.2. Protein Refolding Experiments

We used this hydrodynamic focusing mixer and the scanning confocal setup to capture the fluorescence intensity of the folding of the Pin1 and hYAP WW domain. The experimental setup is shown in [41].

Briefly, the fluorescence within the mixer is observed with a custom-built confocal microscope. The protein is first unfolded by dissolving lyophilized protein (300 μM) in 6 M guanidine hydrochloride (GdnHCl). Folding was initiated by mixing them with 100 mM potassium phosphate buffer (PPB, pH 7.0) in the microfluidic mixer. Mixing is achieved by flowing the buffer at the side channels at ~100 times the flow rate of the unfolded protein in the center channel, constricting the protein stream to a jet ~100 nm wide. After mixing, the flow proceeds down the exit channel at a constant rate. During the folding experiment, the Trp fluorescence decay along the jet was measured at different locations with a photon counter.

The Trp fluorescence is excited by an argon-ion laser frequency doubled to emit 258 nm light. The light is focused to ~1 μm spot by a fused silica microscope objective and emitted fluorescence is captured by the same objective and detected by a photon-counting module. The chip is scanned across the objective to measure the fluorescence at different positions in the supply and exit channel. The channel position is converted to time by the known flow rate. The fluorescence intensity during folding (I) was normalized by a baseline measurement (I_0) where the protein is mixed with the same denaturant solution it was already dissolved in. This normalization removes optical artifacts due to surface inconsistencies from etching and the large drop in fluorescence during formation of the jet. Figure 20 shows the relative fluorescence vs. time. Data collected in the center channel above the mixing region is assigned a negative time. Any change in fluorescence during the mixing time would suggest extra phases that cannot be resolved by this

device. For this protein, as can be seen in Figure 20, there does not appear to be a substantial "burst phase" (hydrophobic collapse). The rise in fluorescence is fit to a single exponential with a rise time of 133 μs (red curve in Figure 20), in agreement with the folding rate measured by laser T-jump [93].

2.3. Conclusions

This chapter shows a new procedure developed for making an ultrafast fused silica microfluidic mixer by using low-viscosity KMPR photoresist (KMPR 1005 and KMPR 1010). This new procedure is simpler, less time consuming and less expensive without compromising quality or performance of the microfluidic device compared to the previous method, using poly Si as the mask. We successfully demonstrated the utilization of the newly fabricated mixer to investigate the folding of the Pin1 WW domain in microsecond time scale. The new procedure developed in this work can also be employed to other applications such as lab on a chip or biophysical measurements, which require the microfabrication of fused silica devices.

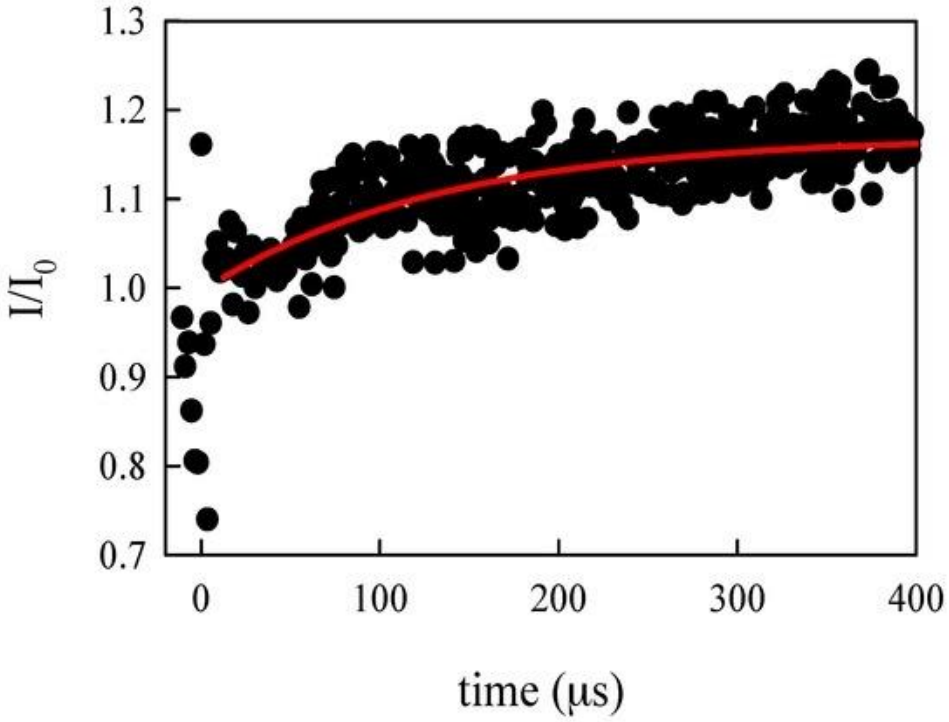

Figure 20: Relative fluorescence of the refolding experiment versus time.

Table 1: Fabrication process for KMPR1010

Process Parameter	Process Tool	Serpentine Mixer Design
Photoresist	ACS 200 Cluster	KMPR® 1010
Photoresist spin speed	ACS 200 Cluster	1.5 RPM
Average resist Thickness	Dektak 8M Surface Profilometer	17.0 μm
UV exposure time	Karl Suss MA6	18.0 seconds
Development time	AZ 300 MIF	285 seconds
Total etch time	STS glass etcher	90 minutes
Etched channel depth	Dektak 8M	45.0 μm

CHAPTER 3

BIO-CONJUGATION WET LABORATORY CHEMISTRY

To be able to apply force to an individual protein using our dual optical trap, we need to somehow connect a single protein to a pair of beads. The most common and facile approach is to use double stranded DNA "handles" that function as spacers between the trapped beads and the protein. One reason for using DNA molecules as handles is that they are relatively easy to synthesize and modify in the laboratory. Choosing the DNA handle length is based on an optimization between convenience (longer) and measurement resolution (shorter). ~3 kb dsDNA has been previously shown in many experiments to be a good compromise. In this part, we describe how we generate DNA-protein chimeras for GB1. The DNA construct in this study contains a ssDNA linker bound to two long DNA handles. These pieces were made by fusing three segments together (ligation) to make the pairing permanent: left handle ("LH", ~1.1 kb), insert ("GB1-Insert"), and right handle ("RH", 1.8 kb). See Figure 21.

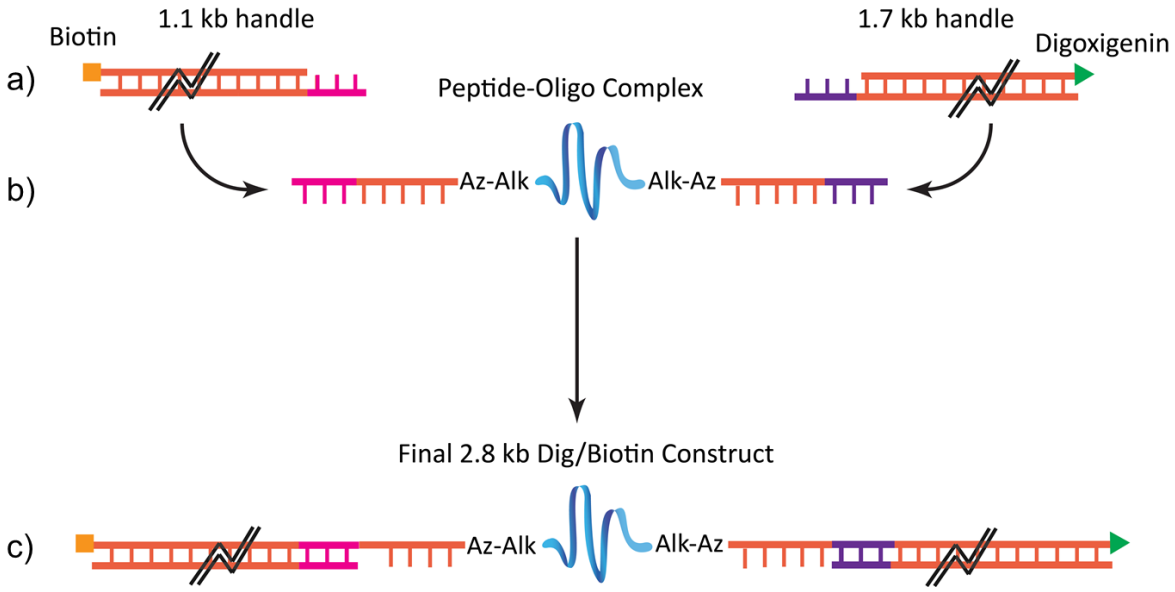

Figure 21: Schematic of protein-DNA chimera preparation, using click chemistry reaction. a) DNA preparation steps before attaching Insert, b) Insert made using click chemistry reaction to attach alkine groups on both ends of protein to azide groups on the ends of ssDNA, c) After ligation and cleaning, final 2.8 kb construct is ready to be held in the trap.

It is an uncommon challenge to construct a protein with two specific DNA linker oligos attached. We explored numerous methods including different disulfide bond approach. For the disulfide bond approach, while one end of each DNA molecule was covalently attached to a cysteine residue, the other end is bound to a bead through either streptavidin-biotin or digoxigenin-antibody interaction. The disulfide method worked, but had the drawback of often creating multimers of protein in the construct. We finally settled on a ‘click chemistry’ method. This method has the advantage of preventing multimers and is straightforward and reproducible. In this chapter, we explain how this method works. Check Appendix A for full, detailed protocols.

3.1. GB1-insert preparation

Two approaches were used to attach single-stranded DNA oligos to peptide: disulfide thiol linked protein insert, which uses disulfide bonds to attach single stranded oligos to peptide [94]

and Copper catalyzed azide-alkyne cycloaddition (CuAAC, ‘click chemistry’), which uses azide-alkyne attachment instead [95, 96].

CuAAC is the most popular protocol of click chemistry and was introduced by the Sharpless group in 2001 [97]. This method forms a covalent bond between single azide and alkyne ligands. The reaction is very slow unless catalyzed by Cu (I) atoms. Azides and alkynes only react with each other with very high specificity. This is an advantage of this method compared to the disulfide bonding, because it does not let peptides bond to each other. The reaction yield is often ‘quantitative’ (very high). Reaction conditions are biomolecule compatible. In any ligation methodology, notable challenges exist to produce sufficient sample yield while considering the fragile nature of the biomolecules. In our laboratory, we first had to explore various click chemistry method options (e.g., the presence of DMSO, reaction temperate and timing, etc.) to get the reaction to work. The reason we did that was we were not initially able to get a high yield of our desired construct. Therefore, we tried exploring various methods to be able to achieve a final high yield of desired construct made for GB1 single molecule experiments.

The main challenge is the unique structure of our particular peptide, GB1. The reaction requires the accessibility of the ligands on the peptide NC termini. If the azide or alkyne group on the biomolecule is sterically hindered or somehow inaccessible to the catalyst and the coupling partner, increasing the reaction temperature or adding solubilizing agents such as DMSO can have a beneficial effect. We tried both adding DMSO and increasing temperature. The reason for this to be true is that increases in the ability of the medium or temperature to solvate hydrophobic domains can make hindered sites expose to potent catalyst. Figure 22 shows PAGE gel result for all methods we used. The top three bands in the magnified area on Figure 22a are 'two left oligos and the peptide', 'one left oligo, peptide and the right oligo' (desired insert), and 'two right oligos

and the peptide', from bottom to top respectively. Lane 2,3 and 7, showing the highest yields, correspond to either 10% DMSO reaction or high temperature reaction. Figure 22b shows another set of reactions with a control lane (no peptide added), different ratios of peptide to oligos at high temperature, reaction with 10% DMSO, and reaction with 2.6 M GdnHcl. The results are consistent with previous gel, showing that adding DMSO and increasing temperature both help increase the reaction yield. Figure 23 shows the size exclusion results for some of the reactions. SE results are consistent with the PAGE results, showing better yield with DMSO and increased temperature. The first peak in each sample is collected as the desired GB1-insert.

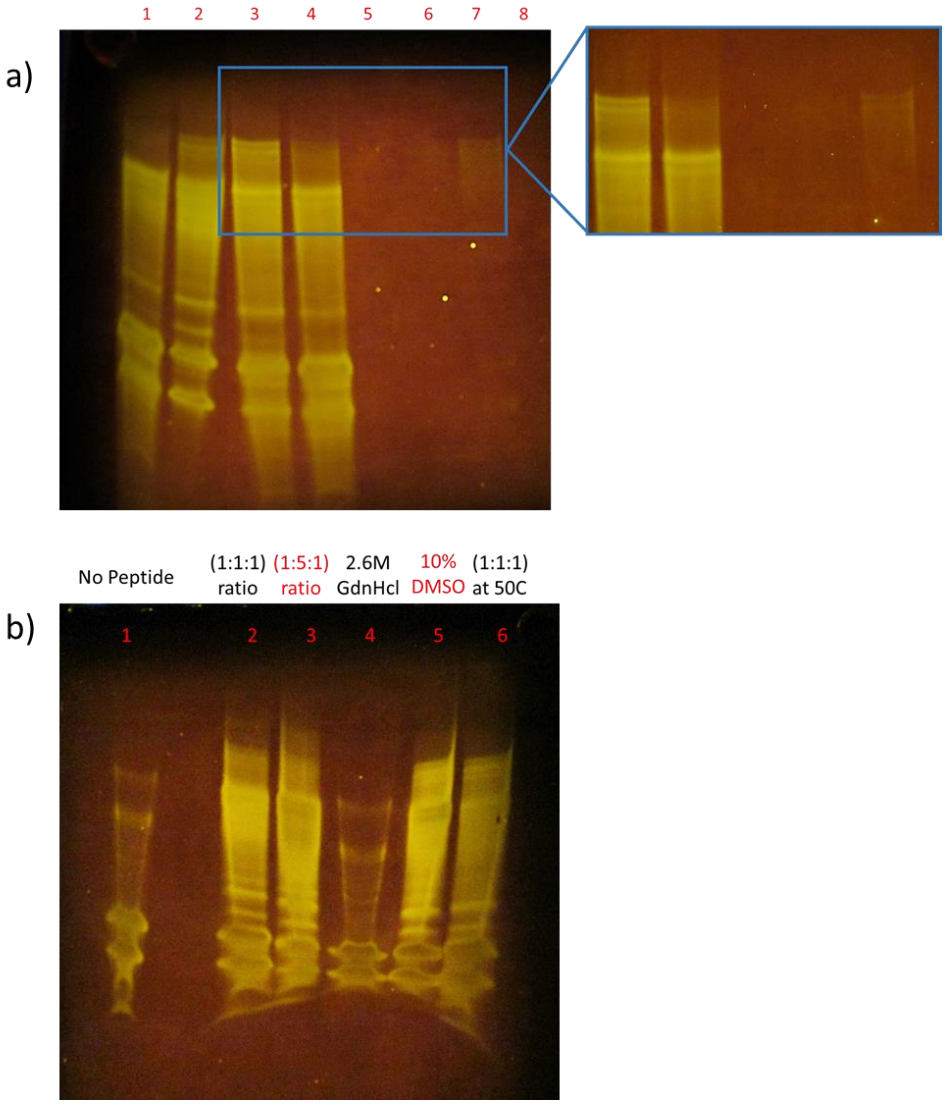

Figure 22: PAGE gel result of 'GB1-insert' for the following reactions: a) Lane 1: Reaction with (1:1:1) ratio- 2 hrs. incubation at RT, Lane 2: Reaction with 10% DMSO- 2 hrs. incubation at RT, Lane 3: Reaction with (1:1:1) ratio- 2 hrs. incubation at 50 C, Lane 4: Reaction with (1:1:1) ratio- 6 hrs. incubation at RT, Lane 5: Reaction with (1:1:1) ratio- 2 hrs. incubation at RT, Lane 6) Reaction with 10% DMSO- 2 hrs. incubation at RT, Lane 7- Reaction with (1:1:1) ratio- 2 hrs. incubation at 50 C, Lane 8- Reaction with (1:1:1) ratio- 6 hrs. incubation at RT. Results show better yield with DMSO and increased temperature. b) Lane 1: Reaction with no peptide- 90 minutes incubation at RT, Lane 2: Reaction with (1:1:1) ratio of L-oligo, R-oligo and peptide, Lane 3: Reaction with 5X peptide- 90 minutes at RT, Lane 4: Reaction with 2.6 M GdnHcl- 90 minutes at RT, Lane 5: Reaction with 10% DMSO- 90 minutes at RT, Lane 6: Reaction with (1:1:1) ratio, 90 minutes reaction at 50 C

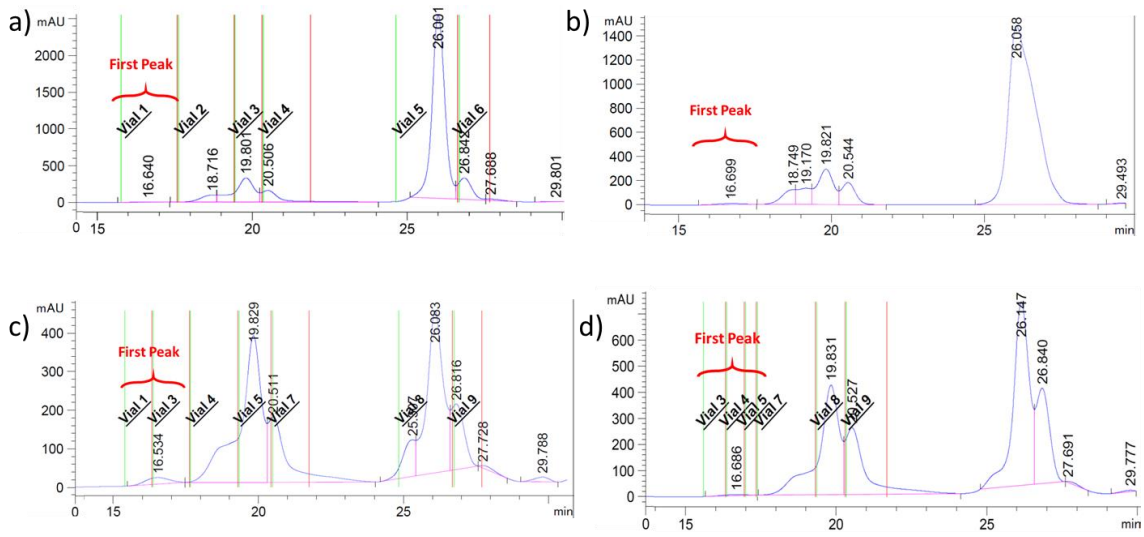

Figure 23: Size exclusion 260 nm absorption results for some of the reactions of the previous figure. The first peak shown in red in all plots correspond to the desired construct. a) Reaction with (1:1:1) ratio- 2 hrs. incubation at RT, b) Reaction with 10% DMSO- 2 hrs. incubation at RT, c) Reaction with (1:1:1) ratio- 2 hrs. incubation at 50 C, d) Reaction with (1:1:1) ratio- 6 hours incubation at RT

3.2. DNA preparation

Our protein-DNA chimera samples, shown in Figure 24, need to bind to two functionally distinct polystyrene beads: one bead coated with anti-digoxigenin antibodies and the other, with streptavidin. In order to do that, we attach two distinctly modified dsDNA handles to the protein molecules. All oligonucleotides used in this study were purchased from Integrated DNA Technology (IDT). All used oligonucleotides are listed in

Table 2 in Appendix A. We used two different DNA handle lengths on left side and right side of the peptide so that our desired construct has a single unique length in the final agarose gel

purification after ligation. The most challenging part in constructing peptide-DNA chimera is attachment of peptide to DNA handles.

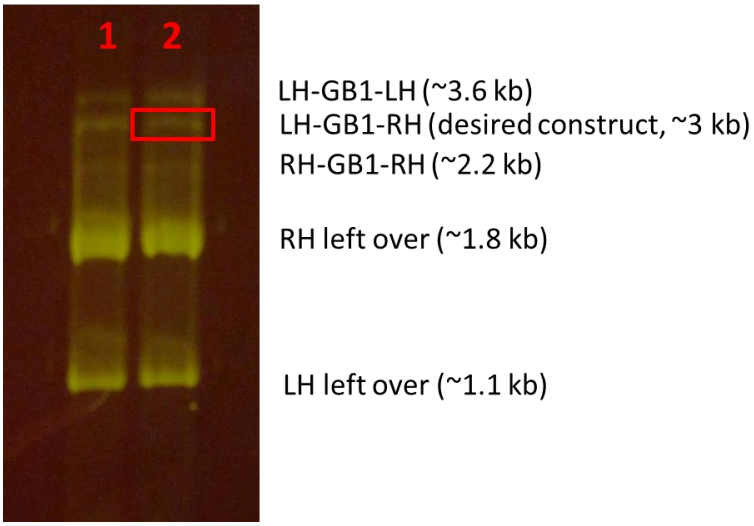

Figure 24: Agarose gel showing two identical samples run for protein-DNA purification. The band in the red box is our desired construct.

3.3. Preparing DNA handles for experiment

First step in making DNA-protein chimera is preparing DNA handles for bead attachment. To ensure the maximum yield of attaching two DNA handles to each protein, we first synthesize and purify both handles. For more information on primer name, sequence, and detailed protocol on how to purify handles, check Appendix A. Restriction digestion is the process of cutting a dsDNA molecule into two pieces at a specific site using an enzyme. This step creates a ssDNA overhang at one end of the dsDNA handle molecules for ssDNA attachments from protein-insert. As the final step in DNA preparation, agarose gel purification is used to purify the desired portion of the cut DNA handles from the undesired portion. See Appendix A for details of the protocol.

3.4. Ligation of LH and RH to Protein-Insert

We ligate handles to the protein-insert ssDNA linker oligos to form the final construct product. To achieve a good yield of the final construct, in theory, it is important that the three components (LH, protein-insert and RH) are combined stoichiometrically 1:1:1. The reason for picking this ratio is that too much insert and both handles will have an insert ligated to themselves, so then they cannot attach to each other. Too little insert will reduce the yield for having our desired construct since there will be too many left over handles. However, experimentally we found that higher final yields are achieved when using 2X to 5X more protein-insert compared to handles (i.e., 1:2:1 and 1:5:1). This might be because of inaccuracies in measuring the peptide concentration, or the fact that some fraction of our peptide insert is damaged and inactive.

CHAPTER 4

MATERIALS AND METHODS

Single molecule force spectroscopy (SMFS) is a very powerful technique that has been used for studying biomolecular systems [98, 99] and to study protein folding in particular. The primary technique used in this thesis is optical tweezers, which consist of one or more optical traps. In this chapter, we discuss the basics of how optical tweezers work and the single-molecule force spectroscopy methods that we use for our experiments. Dr. Matthew Comstock's group at Michigan State University built this high resolution dual optical trap and fluorescence tool to study single molecules. Our group collaborated with the Comstock group to study protein folding and unfolding events under force.

4.1. Optical Trapping

Optical trapping is a powerful single-molecule technique to probe and, to study biomolecules. Optical trapping, due to the piconewtons and nanometer ranges of force and distance, has been used to study motion of individual molecules, such as proteins and DNAs. Properties of these molecules can be probed through the application of forces to the bead, to which the object is attached. As mentioned in Chapter one, using optical tweezers for studying the kinetics of single molecule proteins started back in 1997, with Bustamante work on the giant protein titin [54]. Since then, optical tweezers have been used to study single molecules under applied tension and force, as the probe and perturbation. In this section we discuss how optical traps trap beads and how the measurements are calibrated.

The optical trap consists of a tightly focused, high power laser beam. The trap light can exert forces on small, micron-scale objects within the laser focus. In our work, micron-diameter polystyrene beads are held within the laser focus by the trap. For small bead displacements from the trap center, the trap acts on the bead as a simple Hookean spring on the bead:

$$F = -\kappa x, \qquad \text{(Eq. 15)}$$

where F is the restoring force, and κ is the spring constant which is proportional to the trap laser intensity. The spring constant depends both on the trap properties (e.g., how well focused it is), and the bead properties (e.g., the radius of the bead and its dielectric properties). The restoring force brings the bead towards the laser focus, where the light intensity is the highest.

This is caused by the transfer of momentum from scattering of incident photons. The laser light changes direction in the bead, thus changing its momentum. The direction of this momentum changes when the light interacts with an object by either reflection or refraction. This change in momentum of the light must be accompanied by an equal and opposite change in momentum of the bead. The force is the time rate of change of the momentum and therefore is proportional to the rate of photons being refracted, which of course is proportional to the intensity of the laser light.

Modeling how a trap works depends on the size of the bead compared to the size of the trap laser focus. Small beads use the field gradient model, larger beads use Ray-optics (refraction) model [48]. Our beads are ~1 μm in size and our trap size is about the same size, and therefore fit imperfectly in both regimes. The force applied on the bead by incident light can be divided into two components: a scattering force and a restoring force. See Figure 25 for the schematic view of how force components move the trapped objects from balanced position. The scattering force comes from specular reflection and absorption at the surface of the bead, while restoring force

comes from refraction as light passes through the object [100]. The scattering component of the force pushes the bead in the direction of the light propagation (Figure 25a), while the restoring force causes a momentum transfer, which results in the object being pushed toward the center of the intensity gradient (Figure 25b). When incident light reaches the particle from one direction, it scatters in different directions and some of the incident light might get absorbed. Stable trapping needs a very steep gradient of the light and is achieved by focusing the beam of light to a diffraction-limited spot using a high-numerical aperture objective lens (Figure 25c). In other words, in order to trap an object stably in three dimensions, the axial gradient component of the force (restoring force), which pulls the particle towards the focal point, must be bigger than the scattering component of the force, which pushes the particle away from that region. What matters most in having a better trap is the external rays since they have bigger refraction angles and consequently higher restoring forces.

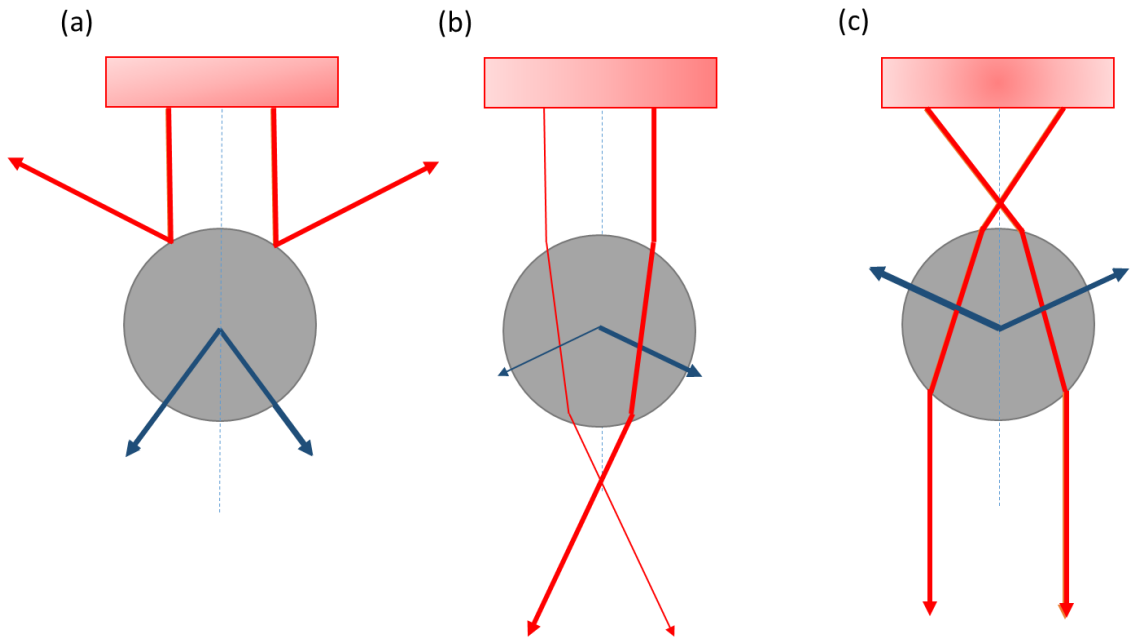

Figure 25: Ray-optics approximation of optical trapping. Beads are shown as the grey spheres. The red lines (of different thicknesses) show two rays of light (of different intensities) from the beam. Blue arrows show the forces. (a) The scattering force from the reflection off the bead surface results in a force that pushes the bead further downstream. (b) The beam of light illuminates the bead. Beam's intensity gradient is increasing from left to right. (c) In a stable trap, the light is focused using a high numerical-aperture lens. The bead refracts the two rays, but this leads to a net force towards the focus.

4.2. Position Measurement

By measuring the displacement of the trapped beads, we can measure the changing extension of any tether between the beads. The changing extension of the tether comes from the changing extension of the inserted protein as it unfolds and refolds. For measuring the bead displacement, there are different approaches such as video-based detection [101], direct object imaging [102], and laser-based position detection methods such as back-focal plane interferometry [103, 104]. The last method has the highest sensitivity and is the method we use.

Laser based position detection can use a single laser for both trapping and position detection.

Back focal-plane interferometry relies on the interference pattern produced by the forward scattered light from the object and the non-scattered light from the trap. See Figure 26 for detection of bead displacement using this technique. When the object is displaced from its equilibrium position, it will create a change in the interference pattern and this pattern is imaged on a position sensitive photodetector, such as a quadrant photodiode (QPD). After calibration, the QPD signal gives the bead displacement from its equilibrium position, and the force on the bead [48]. The QPD measurement is an x and y voltage output that is proportional to the bead displacement from the center of the trap. This measurement becomes non-linear with increasing bead displacement and requires precise centering of the QPD detector with the trap laser.

Axial motion within the optical trap can be determined from the total laser intensity in the back focal plane. The axial position signal is detected from the interference between the un-scattered laser beam and the beam scattered by the trapped object. Axial position is inversely proportional to the NA of the detector [48] and recording the total intensity on the position detector, provides the axial position of the trapped object relative to the laser focus (Figure 26c). However, in the experiments presented here, the beads and tether remained within the trap laser focal plane, and axial motion was not recorded and can be ignored.

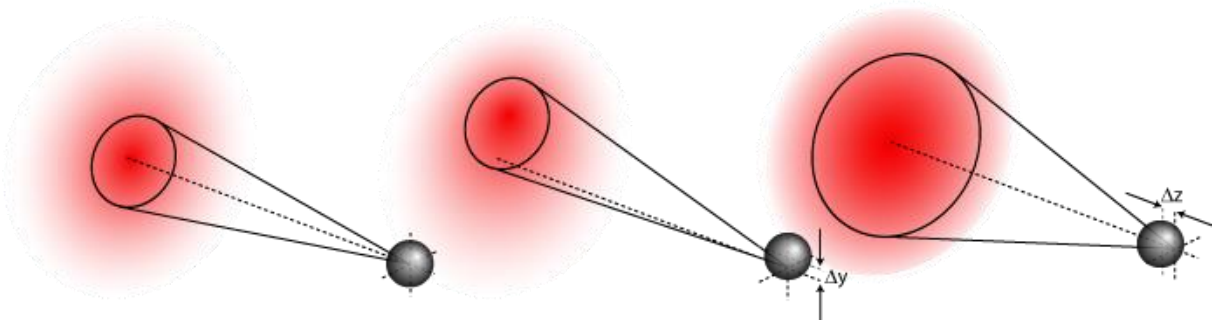

Figure 26: Back-focal plane interferometry, detection of bead displacement by back-plane interferometry. Interference between un-scattered & forward scattered light by bead (grey sphere) is highly dependent on bead position. The interference pattern is shown on a position sensitive detector (red area) for (a) bead in equilibrium position, (b) displaced laterally by Δy, (c) displaced axially by Δz. Figure adopted from [105].

For the most common, traditional optical tweezers setups, the molecule under study is tether between a single trapped bead and a sample chamber surface. Random drift of the sample stage relative to the trap laser focus location is one of the main sources of drift in traditional optical tweezers instrument and the main resolution limiting factors. By instead tethering the molecule between a pair of trapped bead in a dual-trap set up, and making a differential tether extension measurement between the two trapped bead, we disconnect the measurement from the drifting stage. Our instrument achieves a dual-trap design using the 'time-sharing' approach [106]. In such a setup, a single trap laser is rapidly switched between two trap locations using and acousto optic device (AOD) to rapidly deflect the beam between two positions. See Figure 27 for schematic view of our instrument set up and the interlacing cycle from time-sharing. AOD produces a diffraction grating inside a crystal in response to an input radio frequency sound wave. AODs control the trap position through deflection and the trap stiffness through light level. In this setup, since the two traps are formed from a single laser source, they follow an almost identical optical path. This will remove almost all the noise coming from fluctuations in the beam path.

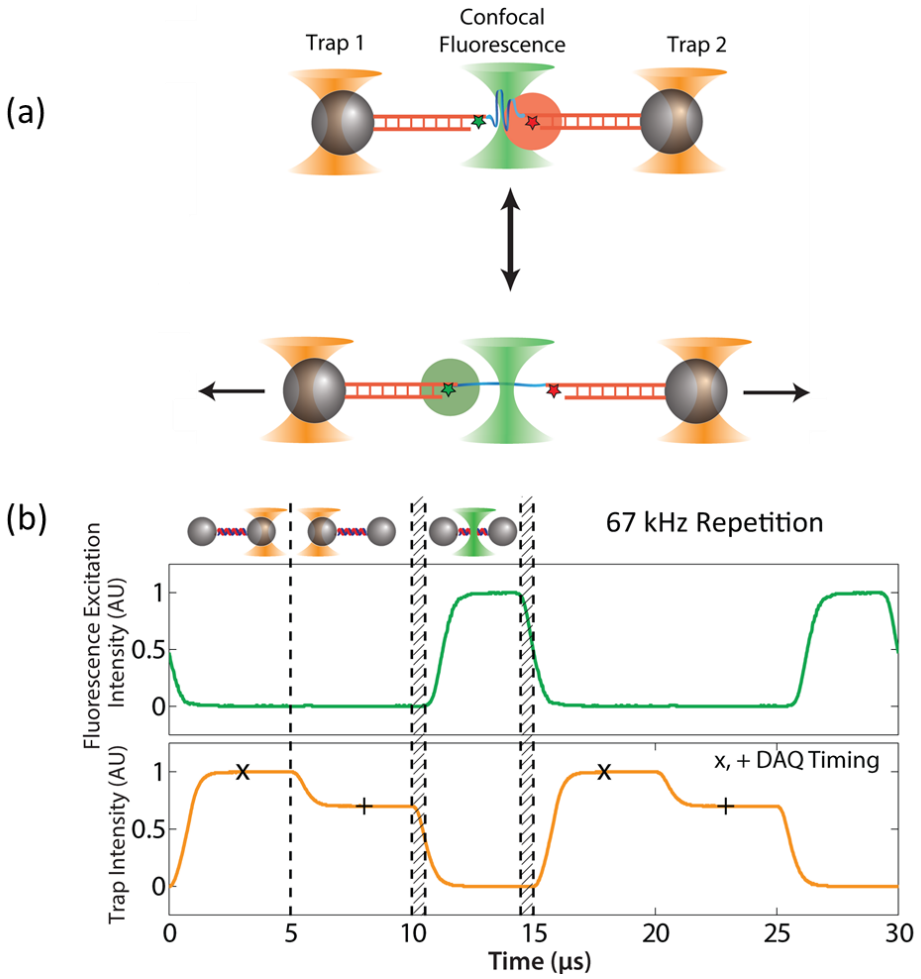

Figure 27: Schematic view of high time resolution fleezers. (a) Schematic view of an experiment setup using high-resolution fleezers setup. Two beads (grey spheres) are held in optical traps (orange cones) tethered together by a protein-DNA chimera molecule (red and blue construct). Both ends of the protein are labeled with cy3 and cy5 for FRET measurements (red and green stars). A fluorescence excitation laser (green cone) excites the fluorophore. (b) The interlacing cycle. The two traps (orange cones) and the fluorescence excitation laser (green cone) are switched on and off in five microseconds intervals (black dotted lines). To decrease the probability of two-photon photo bleaching, a short time delay of 625 nanoseconds is included between the traps and the excitation laser (grey shaded regions). The two traps are set to two different intensities for clarity.

4.3. Position Measurement Calibration

For this measurement we use the method that records the Brownian motion of a trapped object to achieve the conversion factors between bead position and detector output and the trap

stiffness [48, 107]. Figure 28 shows the calibration for bead deflections with known displacement on the glass surface, with known force caused by the flow or Brownian motion.

Precise calibration is limited to spherical objects though. Beads can be used either alone or in attachment to DNA handles for calibration process. To compute the applied force to each trapped object, precise determination of the displacement of the object from its equilibrium position is required. Position determination for the trap depends on motion calibration of the trap itself in the movement surface against beam deflection using acousto-optic deflection (AOD). This is accomplished by video tracking a trapped bead with a fully calibrated piezo-electric stage. When this step is done, the detector is calibrated in both lateral dimensions.

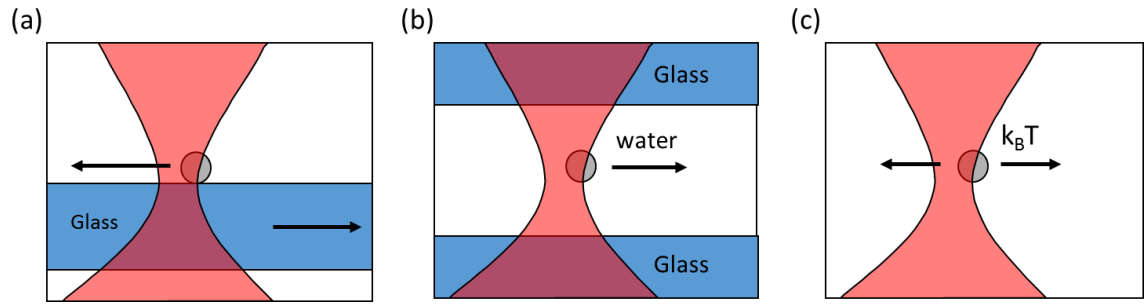

Figure 28: Calibrating bead deflections with (a) a known displacement or with (b) a known force based on Stokes law ($F_{\text{flow}} = \beta v$, $\beta = 6\pi\eta a$) or (c) a known force based on Brownian motion.

We perform position detector calibration based on a measurement of the Brownian motion of the trapped but untethered beads [48]. The equation for the one-sided power spectrum for a trapped bead is:

$$S_{xx}(f) = \frac{k_B T}{\pi^2 \beta (f_0{}^2 + f^2)}, \text{(displacement}^2\text{/Hz)} \qquad \text{(Eq. 16)}$$

where K_B is the Boltzmann's constant, T is the absolute temperature, β is the hydrodynamic drag coefficient of the object and this coefficient for a sphere of radius a in a medium with viscosity η is $\beta = 6\pi\eta a$ (Stokes law). f_0 for a trap stiffness of α comes from:

$f_0 = \alpha(2\pi\beta)^{-1}$, (Eq. 17)

The detector, in fact, measures the uncalibrated power spectrum $S_{\upsilon\upsilon}(f)$, which is related to the true power spectrum by:

$S_{\upsilon\upsilon}(f) = \rho^2 . S_{xx}(f)$, (Eq. 18)

where ρ is the linear sensitivity of the detector in volts/unit distance. For $f >> f_0$, this gives us:

$\rho = [S_{\upsilon\upsilon}(f)\pi^2\beta/k_B]^{1/2}$, (Eq. 19)

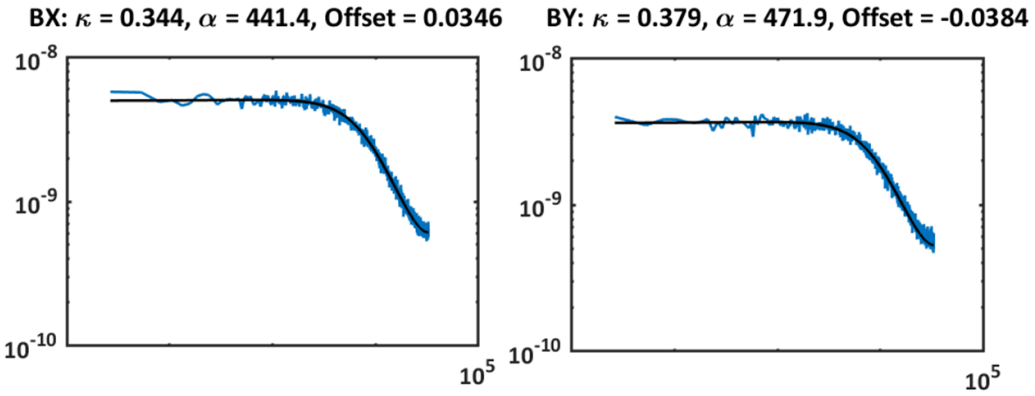

Figure 29: Power spectrum of a trapped bead. Power spectrum recorded with a position sensitive detector (blue curve) for both X and Y positioning. The raw power spectrum fit (black line) to a Lorentzian ($S_{xx}(f) = \frac{k_B T}{\pi^2\beta(f_0{}^2+f^2)}$, (displacement2/Hz)).

The power spectrum from $S_{xx}(f) = \frac{k_B T}{\pi^2\beta(f_0{}^2+f^2)}$, (displacement2/Hz) can be fit to find the overall scaling factor and roll off frequency, f_0, from which the trap stiffness α is calculated from $f_0 = \alpha\,(2\pi\beta) - 1$, . See Figure 29 for details of calibration method.

4.4. Reducing tether molecule damage and oxygen scavenger systems

It is critical to remove oxygen from any solutions where trapping experiment is performed. Reactive oxygen species, such as singlet oxygen, can break tethers during trapping. We used enzyme cocktails such as “gloxy” (GOx; glucose oxide, catalase, and glucose) or “poxy” (POx;

pyranose oxidase, catalase, and glucose) to remove oxygen from trapping buffers [52]. In our experiments, we used both GOx and POx, although we found that the Pox has some advantages over GOx. GOx can cause a pH drop in the buffer over time and, as a result, requires higher concentration of buffering agents, whereas POx does not [108]. In addition, POx can be stored at 4ºC for several months without noticeable change in activity while GOx aggregates over the course of one to two weeks when stored at 4ºC.

4.5. Flow chamber

Sample chambers we use are made of two No. 1 microscope coverslips that contains a layer of sealing film (Parafilm) sandwiched between them. The channels are cut into the sealing film using our laser engraver. Our flow chamber design is shown in Figure 31. The top and bottom channels contain anti-digoxigenin and streptavidin beads, respectively. Two thin glass capillaries connect top and bottom channel to central channel for bead entrance (Figure 30 and Figure 31). The sample flow chamber is mounted on a custom-built U-shaped aluminum bracket [52]. See Appendix A for more details.

The central fluid stream contains an oxygen scavenging system to increase the lifetime of the tethers or fluorophores, if any [109, 110]. All measurements were carried out at room temperature. During an experiment, we first form a tether between an Adig bead and a DNA bead. Streptavidin-coated beads are already covered with DNA handles during sample preparation. By moving trapped beads back and forth, we fish for tethers so that Antidigoxigenin-coated beads attach to our DNA handle through the non-covalent interaction of digoxigenin (dig) modified DNA with an anti-digoxigenin antibody on the bead. We then move the incipient tether to the opposite side of the capillaries entrances, away from sample flow, in the fluid stream to start measurements

and record activity of the tether. This reduces the chance of junk and extra bead drops in the trap during measurements. Check Figure 31 for a schematic view of the explained set up.

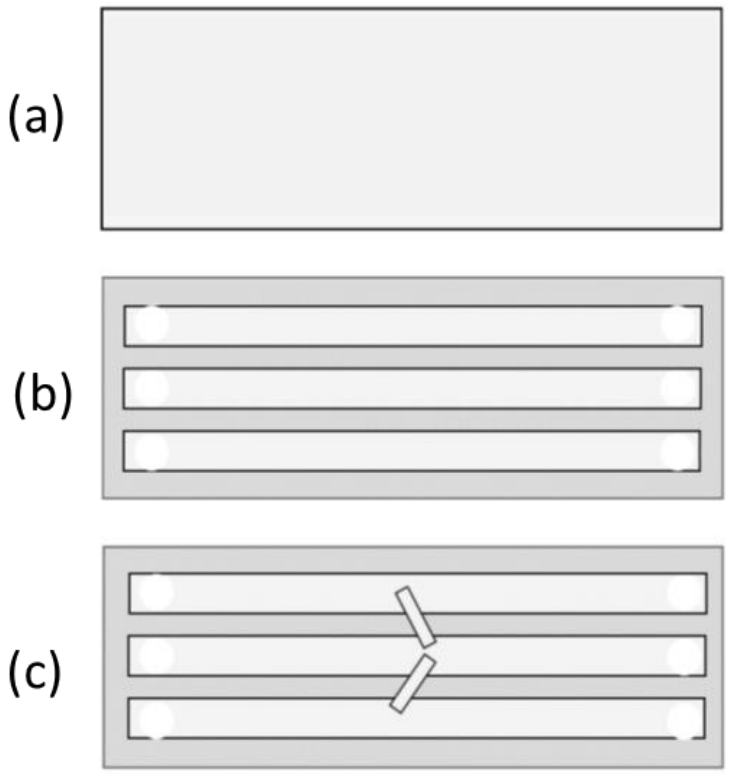

Figure 30: Schematic view of the flow chamber assembly showing (a) glass coverslip with no holes, which sits on top, (b) cut Parafilm aligned in the middle, and (c) bottom coverslip with holes on it and two capillaries placed in the appropriate position.

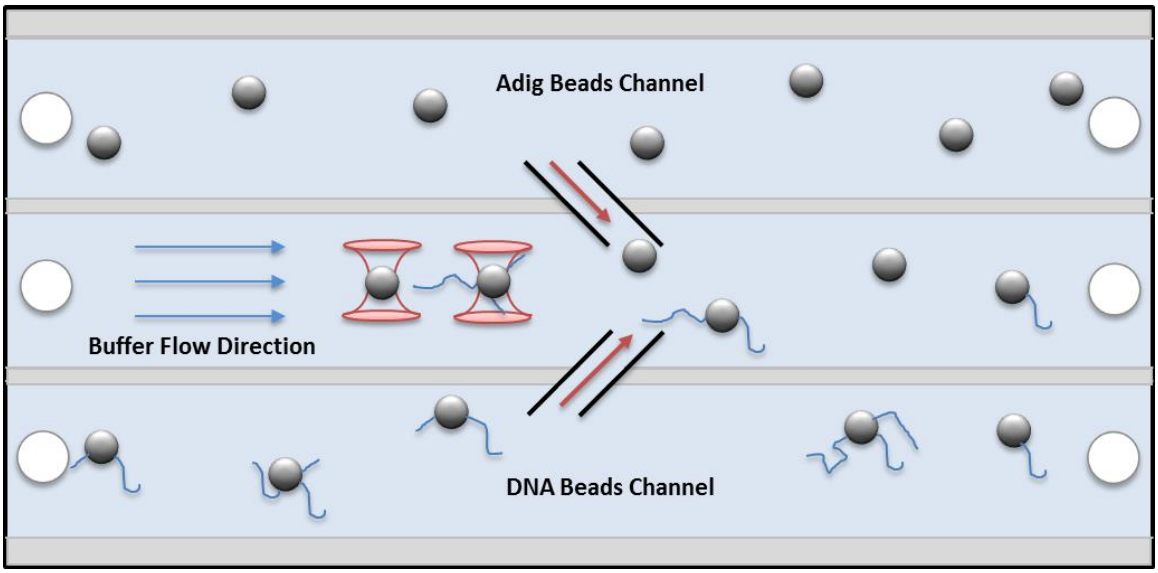

Figure 31: Schematic of flow chamber used for GB1 experiments. Adig beads flow in the top channel and DNA beads flow in the bottom channel and capillaries will let beads into the central channel for measurements. Central channel is showing the fluid stream direction and two beads fishing for trapping. Measurements are done on the opposite side of the capillary entrance, away from sample flow, to reduce the chance of junk and extra bead drops in the trap during measurements.

4.6. Trapping Experimental Conditions and Setup

In a trapping experiment, we attach a protein-DNA chimera (our PG construct), which consists of two relatively long DNA handles on the side and a single PG protein in the middle, to two beads. The two beads are attached to the ends of both DNA handles and are trapped by the laser beam. We pull the trapped construct in various ways and study the kinetics of GB1. More details of our data analysis method are explained in data collection methods.

4.7. Data Modeling Method (Worm-Like Chain Model)

Many polymeric biological molecules such as DNA, RNA and peptides have an intrinsically elastic structure that makes them rigid at small length scales and flexible over longer lengths. Over the past 25 years, the elastic properties of long nucleic-acid duplexes have been

revealed in detail by manipulating individual strands using single-molecule force techniques. Earlier studies of long strands of dsDNA using magnetic tweezers discovered that these polymers are well described by the WLC model [111-113]. Worm-like chain (WLC) model, first introduced by Kratky and Porod in 1949 [114], is one of the simplest theories that has been able to explain many biological molecules characteristic behavior and describing the elastic properties of a variety of bio-molecules, such as ss-DNA, ds-DNA, and polypeptide chains. This model assumes that the polymer is inextensible (i.e., the bonds connecting subunits do not stretch), has linear elastic bending energy and it can have thermal fluctuations and is defined as an infinitely thin rod with bending rigidity κ. The WLC model describes the chain as a deformable continuum semi-flexible rod characterized by its "persistence length". The stiffness κ is conventionally characterized by the persistence length l_p

$$lp = \left(\frac{\kappa}{kBT}\right), \qquad \text{(Eq. 20)}$$

Persistence length is a measure of stiffness that corresponds to the length at which a polymer behaves as a rigid rod, which for dsDNA was found to be ~50 nm (~150 bp) [111]. For contour lengths, lc, that are much shorter than lp, the wormlike chain becomes a rigid rod, whereas for lc >> lp, the end-to-end distribution comes from a random walk of the polymer. We use a slightly modified version of WLC to model our dsDNA handles: the XWLC. This is the extensible WLC. This adds to the model the ability of the segment bonds to stretch. We need to include this when we pull the DNA tether nearly to their full contour length, as we do when F > a few pN, which is basically all the time.

Unfolded proteins are polypeptide chains and can also be modeled well using XWLC model [54, 115]. The persistence length for the protein case is considered 0.38 nm and the subunit length for the protein, the amino acid, is 0.8 nm. The model for the entire tether extension vs force

is simply the sum of the models for each portion. For the relevant tether, our peptide insert tether, there are three parts to sum: dsDNA handles, unfolded peptide, AND ssDNA linkers. For ssDNA we do not use WLC, but instead we use the XFJC.

For a wormlike chain, the bending energy to form a contact can be estimated by [115]

$$\frac{E_l}{K_B T} = \frac{k_l^2 l_p l_c}{2} \approx \frac{2\pi^2 l_p l_c}{(l_c+a)^2}, \quad \text{(Eq. 21)}$$

where k_l is the curvature of the loop. If the loop is approximated as a circle of circumference (lc + a) and k_l is approximated by the reciprocal of the loop radius. The bending energy becomes significant as l_c reached l_p, since the bending energy varies approximately as n^{-1}.

The WLC model fit the stretch data at low to moderate forces and the release data at moderate to high forces. For a WLC, end-to-end extension is related to the external force by [116]:

$$f = \frac{k_B T}{\varepsilon}\left(\frac{1}{4\left(1-\frac{Z}{L}\right)^2} - \frac{1}{4} + \frac{Z}{L}\right), \quad \text{(Eq. 22)}$$

where Z is the end-to-end extension, L is total contour length of the chain, $\mathcal{E}$ is the persistence length, k_B is Boltzmann constant, T is temperature and f is the applied force.

This model is not adequate over the larger range of stretching force, though, which exceeds beyond dsDNA's contour length [112, 117, 118]. Figure 32 shows the experimental force vs extension curve which fit with the WLC model curve.

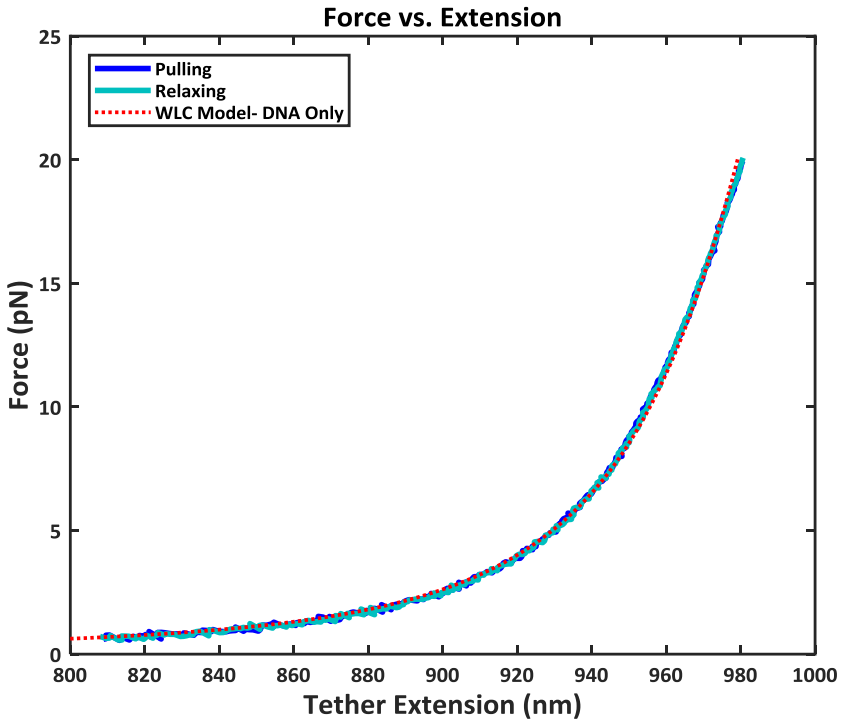

Figure 32: Worm-Like Chain (WLC) Model, force extension curve for DNA fit with the model curve. The insert shows the optical trap setup used to perform this measurement.

Figure 32 shows a force-extension curve obtained by manipulating a single tether. When no protein is involved and the measurements are done just on double stranded DNA handles, we expect to see a XWLC model increase in force and extension. For DNA handles, no transition from the WLC curve is expected. However, when another molecule, such as a protein, is present in the tether, this behavior changes depending on how the protein behaves. In other words, we expect to see different WLC curves for different tether lengths. When a protein unfolds, there is a switch from one length of the tether to the other and consequently, we expect an abrupt increase in the extension and a drop in force. By analyzing the size of the transitions in the recorded traces, it is possible to calculate the number of amino acids involved in every single unfolding/folding process.

4.8. Data Collection Methods

We considered two separate approaches for our data collection. The first measurement approach is under non-equilibrium conditions in which the optical trap separation continuously changes with a constant loading rate. This will pull and relax the force on the protein. This method which is also called force ramps in most single-molecule force measurements, applies high forces to the molecule and then relax it at the same rate. The second measurement approach is under equilibrium conditions, holding trap positions constant while taking data. In the second approach, bead positions are fixed in respect to each other and the only change in the extension (and force) comes from the tethered molecule, which is the polypeptide chain in our measurements. Both approaches provide useful information regarding the kinetics or behavior of the protein molecule under force.

4.8.1. Approach 1: Data Collection Under Non-Equilibrium Condition

Depending on the data collection approach, we can either pull to a specific high force, which makes the protein unfold, and then relax back to low force to let the peptide refold, or pull until we reach to a specific force or extension to let the protein fold and unfold by itself under equilibrium conditions. Each of these data collection approaches occurred in several ways, and details for both methods is discussed in the next two sections.

Figure 8 in Chapter one, shows schematic view of the dual optical trap experiment set up. During the experiment, the protein is stretched and relaxed by moving the beads apart. By moving the beads apart, we can apply force to the protein.

To generate sufficient unfolding/folding events with less error contributed by instrumentation drift, the force is swept up and down rapidly with unfolding and folding occurring

far above/below the equivalence force, i.e. the force where the folded/unfolded states have equal free energy and are equally likely. Thus folding/unfolding reactions happen far from equilibrium, at extreme forces, and various data analysis routines are used to infer equilibrium properties [60, 119-121]. These analysis methods work best for simple cases of 2-state folding or cases where one or two intermediate states are well understood but validation becomes challenging [121]. In addition, it is likely that high forces can exceedingly perturb the folding free-energy landscape such that alternate states and pathways are induced and observed at high forces versus low forces.

4.8.1.1. Data Analysis Methods for Non-Equilibrium Conditions

Extensions of DNA tether versus tether tension can be predicted using standard polymer models which was also explained in detail in last section as WLC model. This modeling for our measurements consists of numerical calculations performed using MATLAB [106]. Total extension of each tether consists of the sum of the extensions of segments of the tether that form it. The model tethers are decomposed into ds- and ssDNA and amino acids segments and compute the extension of all segments separately. We use worm-like chain (WLC) model to compute the extensions of the ds-DNA segments and depending on the number of amino acids in the tether (56 for 1GB1) we add that length to our model to get final model for full extended molecule. So consequently, the DNA-only model is considered a fully folded model with no amino acids involved in the chain and depending on the number of the amino acids being stretched, a fit model is assigned for each length. For all calculations used in our studies, we use the same polymer parameters. For WLC calculations: persistence length of the ds-DNA = 53 nm, persistence length of the peptide = 0.4 nm, stretch modulus for ds-DNA = 1200 pN, and the contour length per single ds-DNA base pair = 0.338 nm/bp and the counter length of each amino acid in the peptide sequence = 0.38 nm.

Another important factor in our equilibrium measurements is tether extension, which is calculated from the known trap separation and is calibrated as is explained earlier in this chapter.

It is important to know that this change in extension is independent of the absolute trap separation and bead sizes and any offsets needed to fit the construct force-extension behavior. The predicted change in tether extension upon probe strand binding at a given force F, $\Delta x_h(F)$, is thus calculated from the difference in the extension of ssDNA and dsDNA at that force. Most of our measurements are performed with the traps held at constant positions, however, not at constant force, and because of that, the relative extension of the entire tether compared to the two traps must be considered. For constant trap position measurements, we have a new method of putting the direct polymer models onto the F vs time traces. Therefore, we plot time traces as F vs time, not Ext vs time.

4.8.2. Approach 2: Data Collection under Equilibrium Condition

When taking data under equilibrium conditions, we can directly measure folding and unfolding rate constants, which can be compared with other in vitro measurements and other denaturant methods [44].

In order to keep the protein molecule under equilibrium condition, one method is to keep the trap positions constant. This will let the peptide to fold/unfold under equilibrium conditions. In this set of experiments, we hold the protein (tether) at a constant trap position and this will let the protein fold and unfold by itself and study its kinetics by measuring the change in extension (delta extension). Different trap positions correspond to different forces. Therefore, in this method we can watch either the change in delta trap position or, equivalently, the change in force over time. By measuring the length changes, we are able to count the number of amino acids in the

protein sequence that are opening and figure out if the protein is folded, unfolded or partially folded when at equilibrium.

Optical tweezers instrument with high spatial resolution can also be operated in high time resolution. While doing measurements at highest time resolution, we down sample our data points appropriately during data analysis. This is actually helping with balancing out the noise and being able to catch events easily. Increased time averaging of protein extension measurement decreases spatial noise while reducing the time sampling rate and vice versa. As a rule of thumb, high-resolution optical tweezers instruments such as ours can typically achieve 0.1 nm spatial resolution at 100 ms time resolution with 10 pN of applied force. Previously these instruments were aimed at resolving base pair, 0.34 nm, stepping of nucleic acid processing helicases and polymerases. However, the change in extension upon folding from the extended polymer state to the native folded state of even the small model proteins we are investigating is far larger, approximately 15 nm, and we can easily achieve 1 ms full folding/unfolding time resolution. We perform folding measurements with 15 μs sampling. This time resolution is the highest that our instrument can achieve and considering that our instrument has the highest time-resolution among all optical traps, it gives us the ability to catch possible events that were not observable with slower probe.

4.8.2.1. Data Analysis Methods for Equilibrium Conditions

Direct dwell time measurement is one of the quantities that can be measured in single molecule experiments [122, 123]. The dwell time of a state or configuration of the protein (e.g., the folded or unfolded state) is defined as the time the protein spends in that state prior to transitioning to another state. This is marked by a step change in force or extension in the trajectories. Measuring the dwell times for a large number of individual folding or unfolding events in a large number of proteins gives us statistical information about the protein's kinetics. The

inverse of the mean dwell time is the associated reaction rate constant. In bulk measurements, and in any measurement that enough data point is available, the dwell time measurement is fitted with a single exponential curve. A single exponential decay of a state simply implies that we are starting with a single homogeneous distribution of initial states, going to a single homogeneous set of final states. In our measurements, since we do not have enough data points to be able to fit the data to exponential curve for most forces, we average over the time spent at the unfolded or folded state to find the mean dwell time.

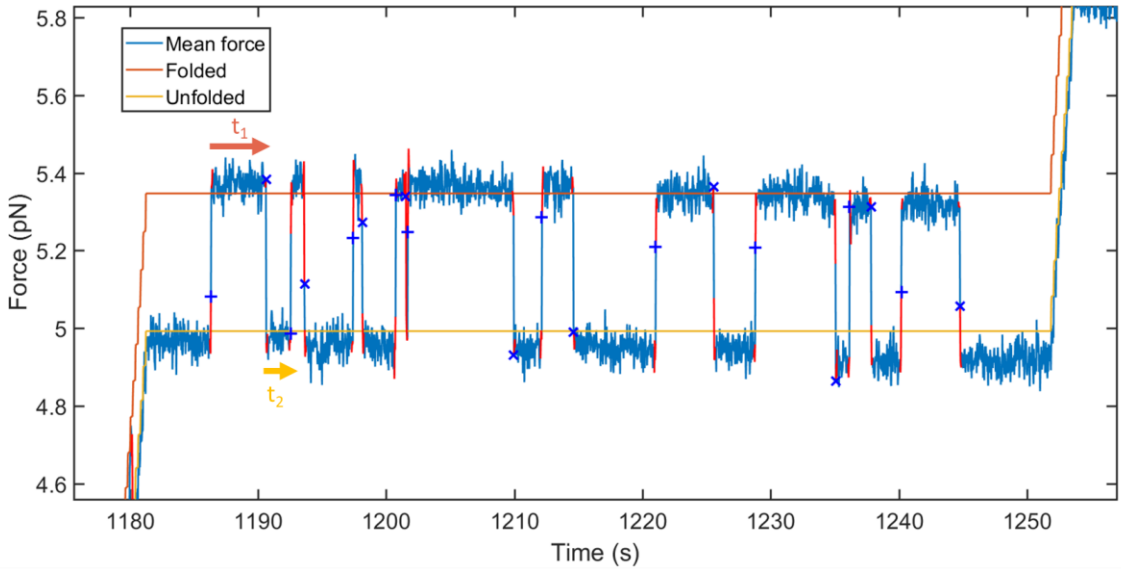

Figure 33: Example of a time trace with transitions between folded and unfolded states. 'x' and '+' marks show transition to unfolded and folded states respectively, while t_1 and t_2 show the time spent at each state before the next transition.

CHAPTER 5

GB1 EXPERIMENT RESULTS AND CONCLUSIONS

Protein G is a large multi-domain cell surface protein of group G Streptococcus. B1 domain of protein G, called GB1, is one of the best-studied systems in protein folding over the last two decades. GB1 shows slow folding kinetics and previous studies on the millisecond timescale, show a folding time of ~1ms for this protein. In these studies, GB1 follows a simple two-state system model [44, 45, 47]. There are other measurements that show some nonlinearity in their chevron plot [49, 50], and some ultrarapid fast mixer measurements in agreement with Markov State Model predictions, which proves that GB1 does not behave like a simple two-state folder [26].

In this thesis I study single molecules of GB1, using force as the denaturant and the probe. The GB1 folding system, which is compatible with high-resolution measurement, along with a variety of powerful biochemical protocols for making suitable protein-DNA hybrid construct is explained in detail in the previous two chapters. The protein-DNA tether, used in this work, consists of two double-stranded DNA "handles", one 1.1 kb long and the other 1.8 kb long, joined by two short stretches of single-stranded DNA overhangs covalently linked to the N- and C-termini of GB1, as shown in Figure 21. In this chapter, results from the GB1 single molecule experiments and their interpretation are presented.

As part of our data collection methods, we completed the characterization of folding/unfolding vs force, including the force dependence, time scales and presence of indeterminate states in the reactions. Here in this chapter, results from the experiments and conclusions is discussed.

5.1. Results of High Spatial Resolution Measurements of Folding Complexity Out of Equilibrium

In Figure 34, results from a typical force ramp experiment on a single GB1 tether molecule is shown. By pulling the beads apart in the trap, force is applied to the GB1 tether and as a result, extension increases. Aside from the protein, the force extension curve for this construct is well modeled as a random polymer of known contour and persistence length (black dotted lines in Figure 34). Ideally, when GB1 unfolds, it also acts like a random polymer with an added known contour and persistence length. The pulling curve on the force extension plot (blue curve) follows the polymer WLC model (black dotted line). At ~7pN unfolding of the protein causes a sudden decrease in force and an abrupt increase in extension. When GB1 completely unfolds, the pulling trajectory instantaneously jumps from one polymer model to the other. From that point, the tether follows the polymer model with a longer length, which corresponds to the length of the unfolded peptide added to the previous length of the DNA handles. We pull on this tether to a force of ~25 pN and then we relax it. This is shown as the red curve on the plot. By moving the beads closer, the force applied to the GB1 tether decreases and the peptide refolds at ~6 pN. This molecule is showing a simple two state behavior at which the protein is either at the folded or unfolded state.

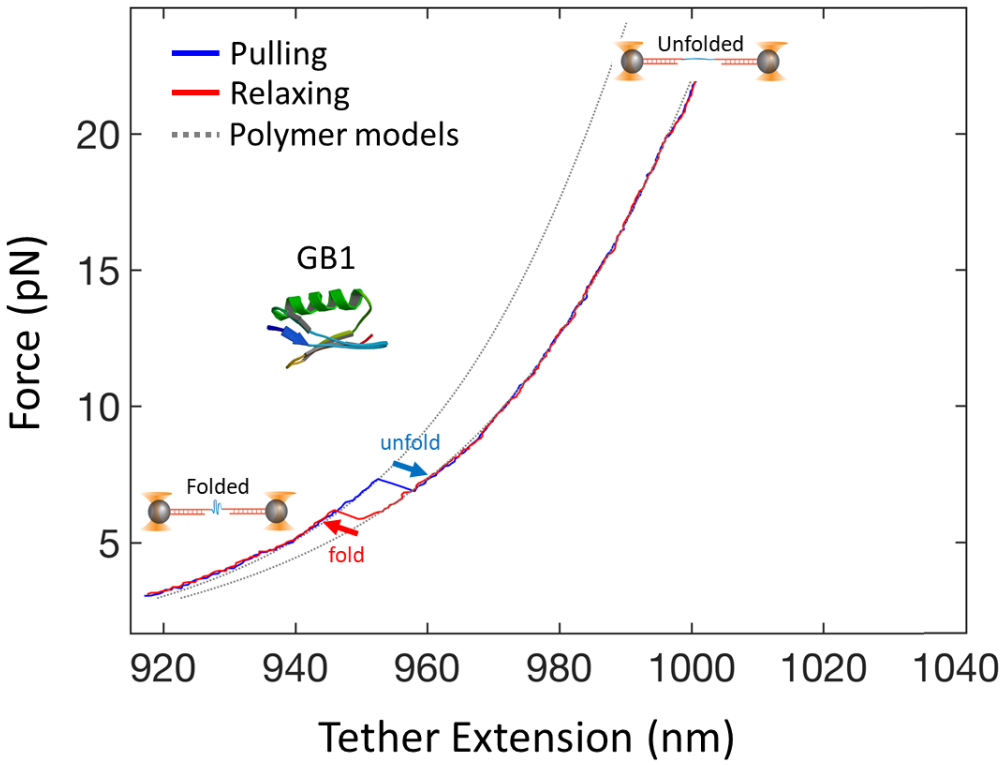

Figure 34: An example of force extension curve for a single GB1 molecule tether showing pulling curve (blue curve) and relaxing curve (red curve) with WLC polymer model (black dotted lines). When beads are pulled apart, the protein unfolds at ~7 pN with a sudden drop in force and abrupt increase in extension. When beads are moved close again, on the relaxing curve, protein refolds at ~6 pN. This plot shows an example of the basic two-state behavior for GB1 molecule.

Figure 35a shows the same GB1 molecule tether that is shown in Figure 34, with multiple force scans applied to the molecule. Individual force extension curves are offset in this plot for better visualization. As shown in the first three curves on the right, when pulling, the GB1 molecule is unfolding at different forces at each pull and refolding at different forces on the relaxing curves. The second curve from left is showing some repeated openings and closings that are occurring during the relaxing curve at low force. The last curve on the right on this plot shows cumulative force scans for the same molecule. We observed reversible two-state complete unfolding/folding results for GB1 and all curves agree that the single GB1 molecule is fully unfolding and refolding

and there is no partial opening for this molecule. In addition, this figure shows that GB1 is unfolding at a large range of forces, F~7-25 pN, but refolds at a much smaller range of forces, F~4-6 pN.

We compiled distributions of the forces of individual unfolding and folding events for each scan speed (1nm/sec-10nm/sec) for multiple molecules. Figure 35b is showing this distribution of folding and unfolding events as a function of force for different scan speeds. The folding distributions are simple, showing a single narrow peak at ~5 pN regardless of speed. This is consistent with a simple 2-state folding model. For a two-state folder, we would expect to observe the dependence of the unfolding forces on loading rate, when the system is out of equilibrium [124]. However, unfolding appears to show two distributions: a low force, ~5 pN distribution likely corresponding to equilibrium unfolding, and an even larger broad high force distribution. This figure shows that not only unfolding occurs over a wide range of forces, but also is somewhat speed-dependent. The speed-dependence of unfolding distribution is well aligned with the characteristics of a two-state system, but then we do not expect to see a low force peak at ~5pN. This suggests that there are some complexities in the folding pathway of GB1 when studied under force at single molecule level. Therefore, what we observe here is not what we expect for the simplest unfolding model. The low force peak is not moving with pulling speed, so perhaps it is also close to equilibrium. The high force spread, though, is moving with speed and it is likely farther out of equilibrium.

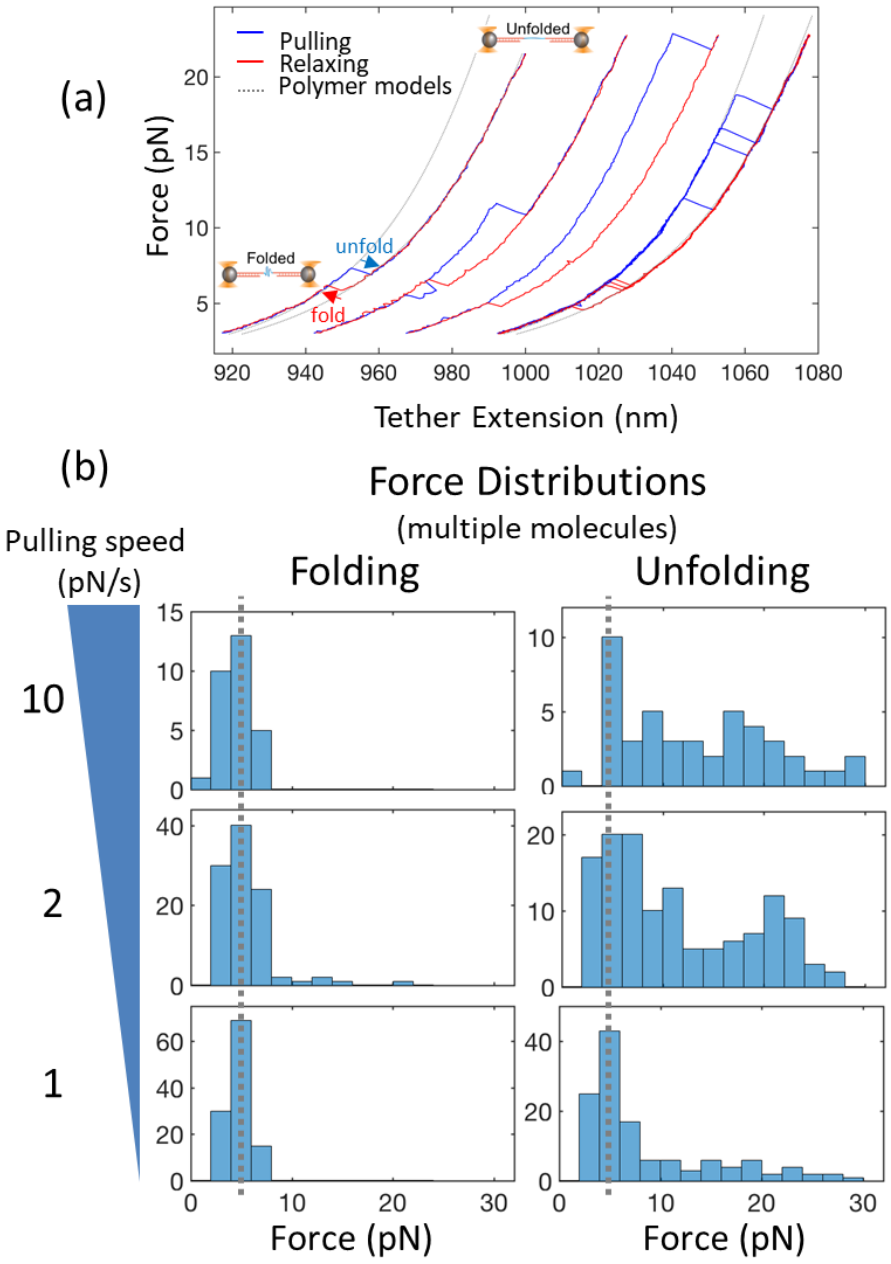

Figure 35: a) Example of force-extension curves of one single GB1 molecule showing multiple force scans at the same force range of F~ 2-25 pN. Repeated pulling (blue) and relaxing (red) at a speed of 10 nm/s. Individual force extension curves are offset in this plot for better visualization. Note the large range of forces for unfolding events and low force range of forces for refolding events. b) Histogram of folding and unfolding events vs. force for multiple molecules at various pulling speeds. The folding distributions are showing a single narrow peak at ~5 pN regardless of speed, the unfolding rates are somehow speed dependent, but still show a peak at the same force of ~5pN (vertical dashed line on folding and unfolding distributions).

5.2. Results of High Spatial Resolution Measurements of Folding Complexity in Equilibrium

The 'messy' nature of the unfolding histograms makes determining the unfolding rates from the histograms impossible for our measurements. Instead, we used another approach for near equilibrium measurements and that is 'constant trap position' condition. This method allows us to find folding and unfolding rates by directly measuring the dwell time distribution under constant conditions: trap position or force.

Figure 36 shows a typical time trace at the constant trap position corresponding to ~5 pN force. The traps, and the bead positions, are fixed during this measurement and the protein state is monitored over time. When the position is fixed, the beads will remain at a specific distance from each other and the tether only changes its length when the protein unfolds or refolds. The unfolding and refolding of the protein causes a drop or increase in the extension and consequently in the force. Therefore, when the protein is folded we expect a lower extension (higher force) in the time trace and when the protein unfolds, extension increases and force drops. This type of measurement is similar, but not equivalent, to force feedback measurements. In force feedback measurements, the force is held constant with a high accuracy. The reason we pick constant trap position over force feedback is that constant trap position measurements are done at the highest resolution of the trap instrument, but force feedback has some limitations in terms of the highest bandwidth. As shown in the magnified view of this figure, GB1 molecule jumps back and forth between folded and unfolded states for ~1100 seconds. A histogram of the raw trap position for this trace is shown in Figure 37a, for three different bandwidths: 1.5 ms, 15 ms (10Xaveraged), and 150 ms (100X averaged). Similar histogram of the folded and unfolded state locations, the distributions of the mean positions of each folded/unfolded event, is shown in Figure 37b. Both histograms are

showing two main peaks corresponding to the folded and unfolded states of the protein. The data do not show any evidence of multiple folded or unfolded states.

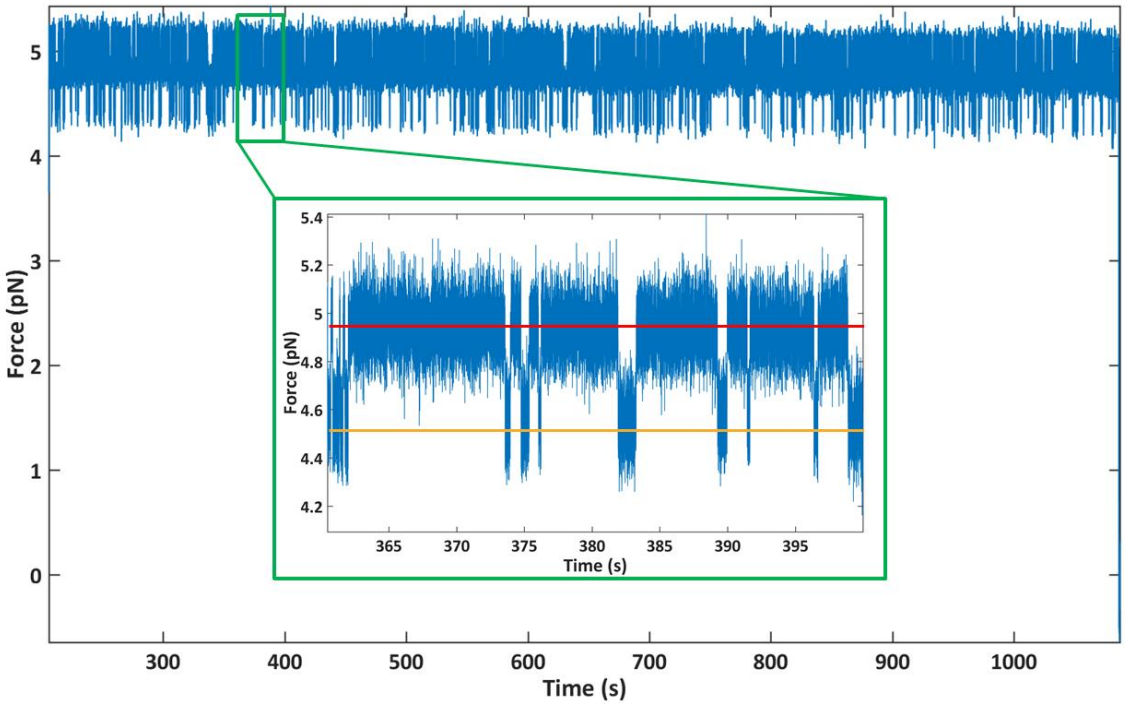

Figure 36: A typical time trace at the trap position corresponding to 5 pN force collected at 1 ms bandwidth. This molecule is showing a simple two-state behavior as shown in the magnified view of the plot. Fully folded/unfolded polymer models in red/yellow.

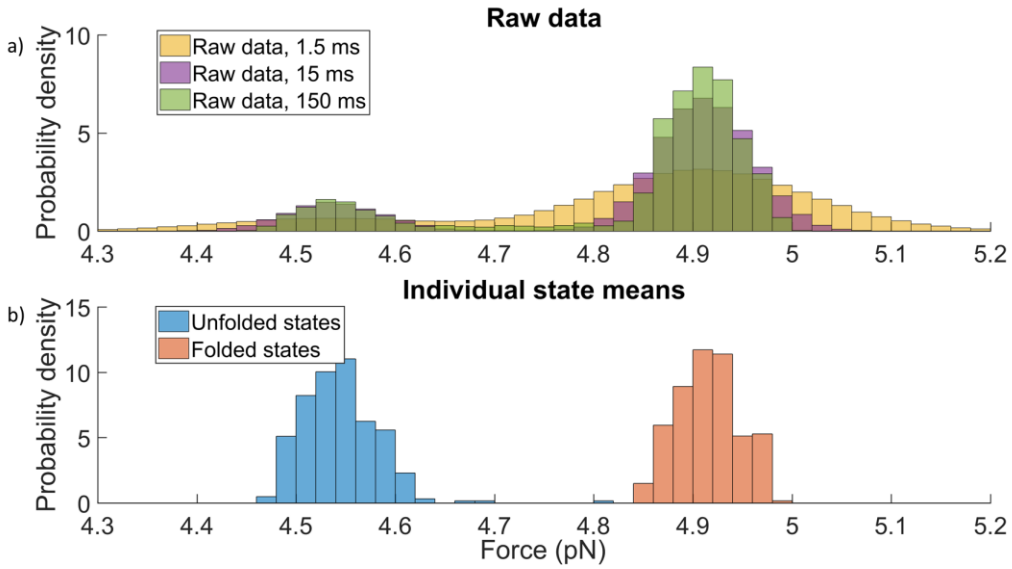

Figure 37: a) Distribution of folded and unfolded GB1 forces for the raw trap position for the trace shown in Figure 36 for three different time resolutions: 1.5 ms, 15 ms, and 150 ms. b) Distribution of folded and unfolded GB1 of averaged data for each individual state.

Figure 38a and c show the dwell time distribution of folded states and unfolded states, respectively, for the long time trace shown in Figure 36. Since this trace has considerable number of folding and unfolding events and is also showing an exponential behavior, we fit it with a single exponential curve as shown in Figure 38b and d. As discussed in chapter four, the folding and unfolding rates are typically calculated from the mean of the dwell times for folding and unfolding because of limited events in most observations. Figure 38a and c show the average dwell times of 2.28 seconds for folding and 0.55 for unfolding. The decay constants are 2.24 seconds and 0.64 seconds, which are comparable with the rate constants calculated from the mean.

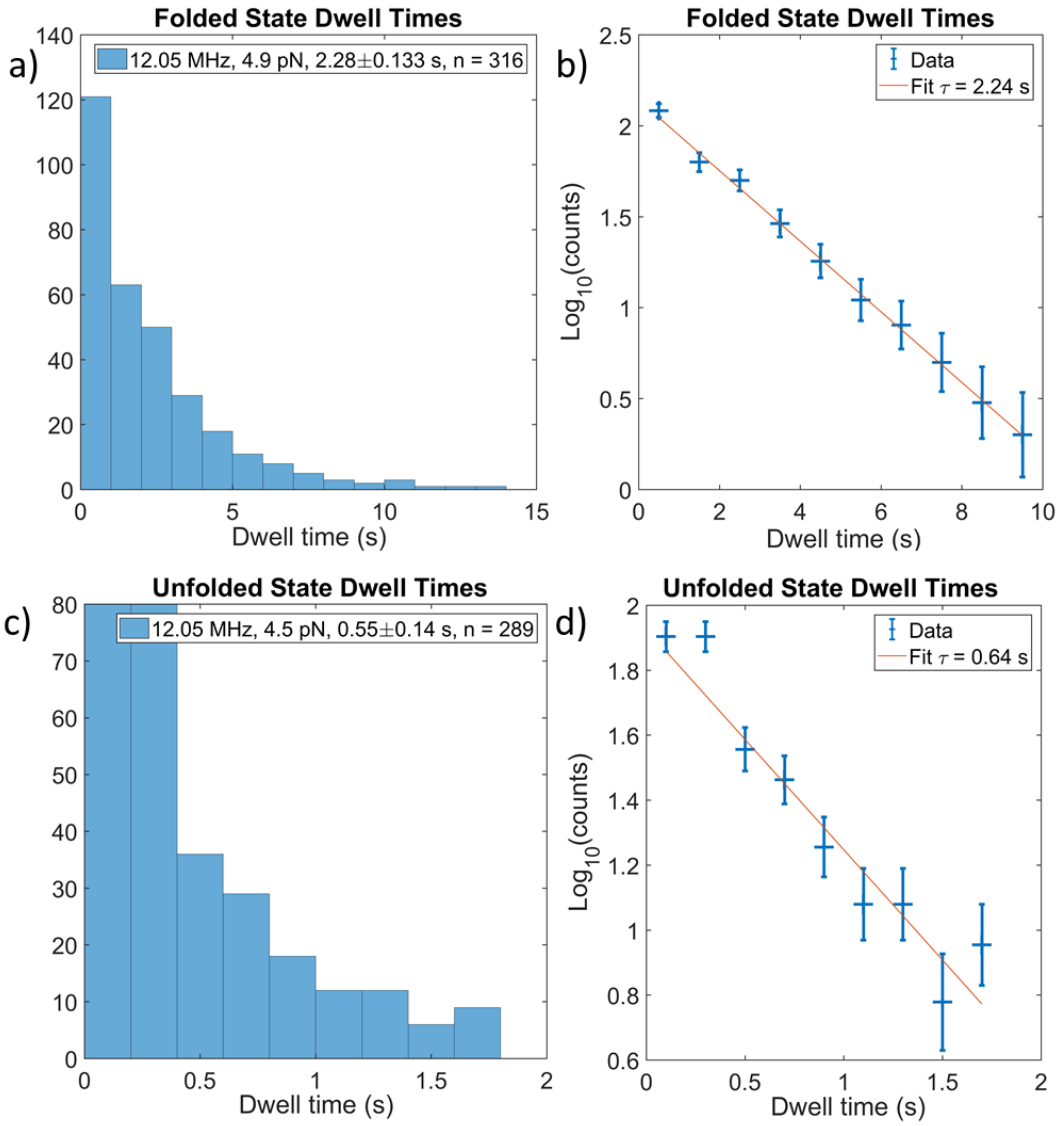

Figure 38: Dwell time measurement distribution for a single GB1 molecule at 5 pN force (see Figure 36) for (a) folding and (c) unfolding. b, d) shows the folded state dwell time fit, with 2.24 sec dwell time for folding, and 0.64 sec dwell time for unfolding state.

To determine the dependence of folding and unfolding rate constants on the applied force, another set of experiments was carried out. In these measurements, the trap is jumped into different fixed positions, which pulls the tether to different forces of F<10 pN, during the time trace measurement. Figure 39 shows a typical time trace for a single GB1 molecule at different trap

positions. As it is expected for a two-state model and is noticeable in the figure, when the average force increases, the protein spends less time in the folded state and tends to stay unfolded for most of the time. Figure 40 shows the folding and unfolding rate constants at different forces for this molecule. When the force increases, the constant rate for folding is increasing, but the unfolding constant rate does not. The unfolding rate is ~0.4 s^{-1}, regardless of force.

We repeated the same measurement for 15 unique molecules under a variety of equilibrium forces, which are plotted in Figure 41. While the folding constant rates (determined from the unfolded dwell times) are consistent over all of these molecules and depend strongly on force, the unfolding constant rates (determined from folded dwell times) vary widely between molecules.

However, the unfolding constant rates can be roughly divided into two categories. One set of molecules seem to have unfolding constant rates that are insensitive to force, $k_u \sim 0.3\ s^{-1}$. The other set have a stronger dependence on force, though weaker than the folding constant rates. Extrapolation of the folded constant rate to F= 0 pN, yields k_f ~500 s^{-1} consistent with the rate determined in zero denaturant by mixing experiments [26]. Extrapolation of the unfolded constant rate to F= 0 pN of the force-sensitive set of molecules yields k_u ~0.02 s^{-1}, also consistent with the folding experiments.

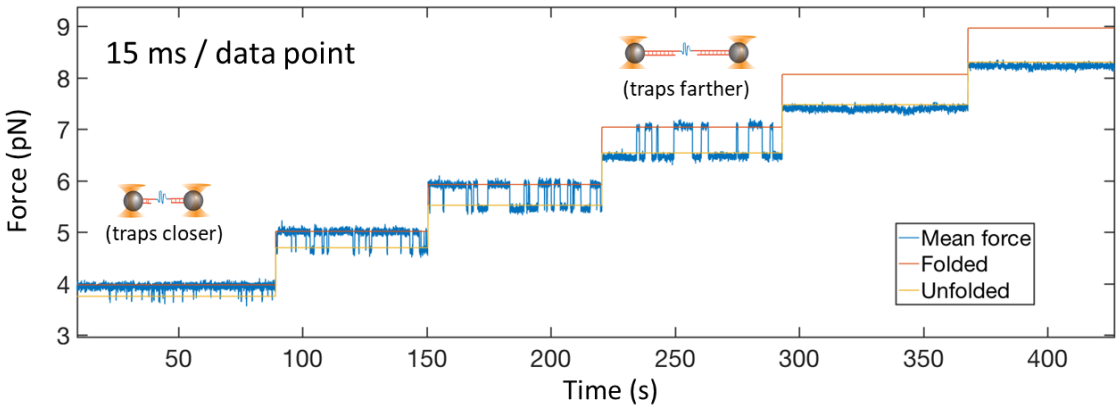

Figure 39: Observation of folding and unfolding at fixed trap positions for a single GB1 molecule. The position of the folded (red) and unfolded (unfolded) forces were determined by the WLC models of folded and unfolded protein G plus DNA handles. At low average force, the protein spends most of its time folded with few, rapid excursions to the unfolded state. At higher forces, the protein spends most of its time in the unfolded state.

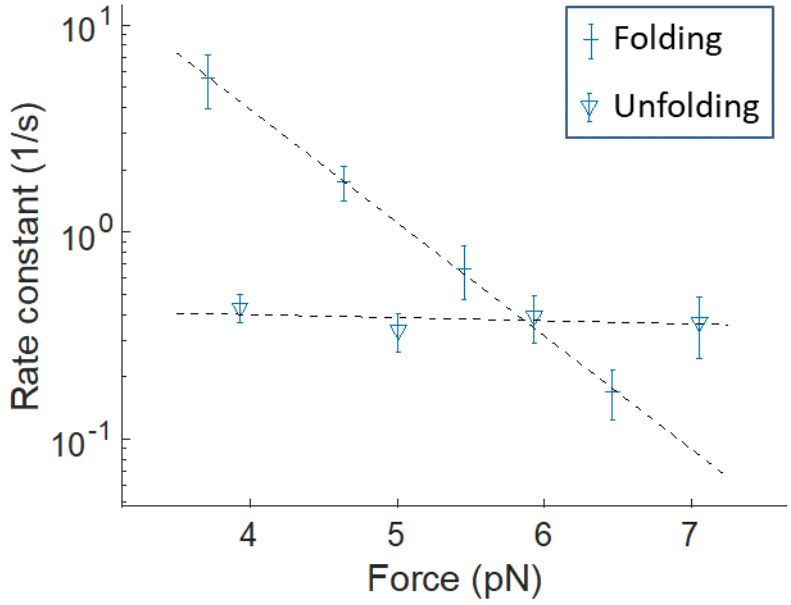

Figure 40: The folding and unfolding rates at different forces for one molecule showing folding rates that an unfolding rate, k_u ~ 0.4 s^{-1}, that is not force dependent. Extrapolate at zero force: ~1.5 ms folding and ~2.5 sec unfolding

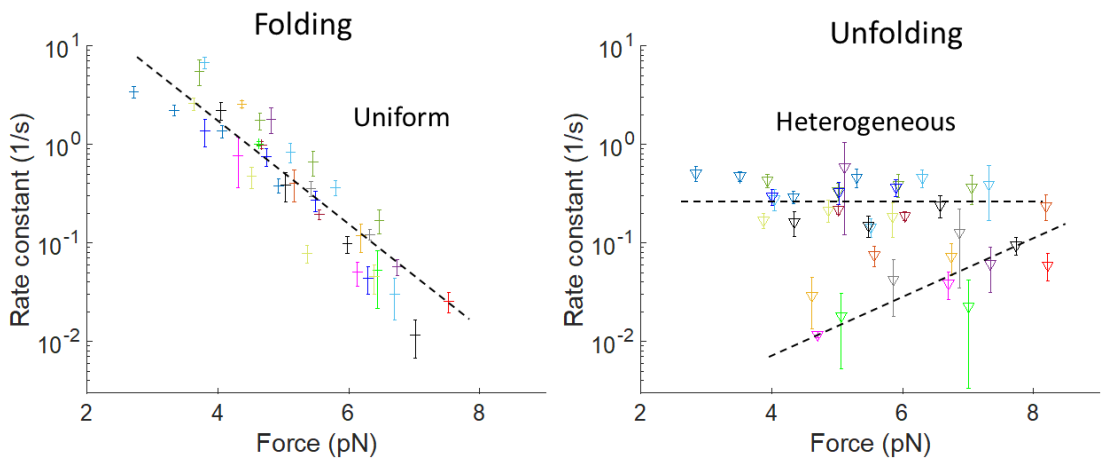

Figure 41: Folding and unfolding rates vs. force determined by dwell times at constant extension for 15 unique molecules. Each color represents a different molecule. The extrapolation of the folding rates to F= 0 pN, gives a rate k_u ~0.02 s^{-1}, in reasonable agreement of the folding rate, k_{H20} = extrapolated from bulk mixing experiments. The ratio of the slopes of these lines is also in agreement with the ratio of m-values from bulk mixing experiments.

One question that needs to be answered here is whether individual molecules can demonstrate both unfolding types. In other words, do these molecules slowly switch between unfolding types? The behavior of the long time trace shown in Figure 36 and the dwell time distribution for folding and unfolding events for the same molecule, Figure 38, does not indicate that the molecule is switching between unfolding types. Otherwise, any deflection from the simple exponential decay is expected. On the other hand, if we go back to the force ramp shown in Figure 35a, we seem to have some data showing individual molecules showing the 'bimodal' unfolding versus force. The histogram shown in Figure 35b supports the idea that individual molecules can switch between unfolding modes. This would suggest that it might be something about the pulling that is helping this, either because it is brought to higher forces, or something about the out of equilibrium speed of pulling.

There is another group of data in our measurements that showed smaller openings than full unfolding in either force extension curves, time traces or both. One example of that type is shown in Figure 42. The force extension curve and time trace in this figure are for the same molecule. In this example, there is an opening of about half of the full unfolding extension in both the force ramp and equilibrium measurements. The offset model for the small opening in the force ramp is found in the same fashion that the fully unfolded fit model is found, which is based on the number of amino acids added to the length of the DNA model only. In the time trace, the fit lines are showing fully folded, intermediate and fully unfolded states and agree with the fit models in force extension curve, meaning that the size of the intermediate state in both equilibrium and non-equilibrium measurements are the same.

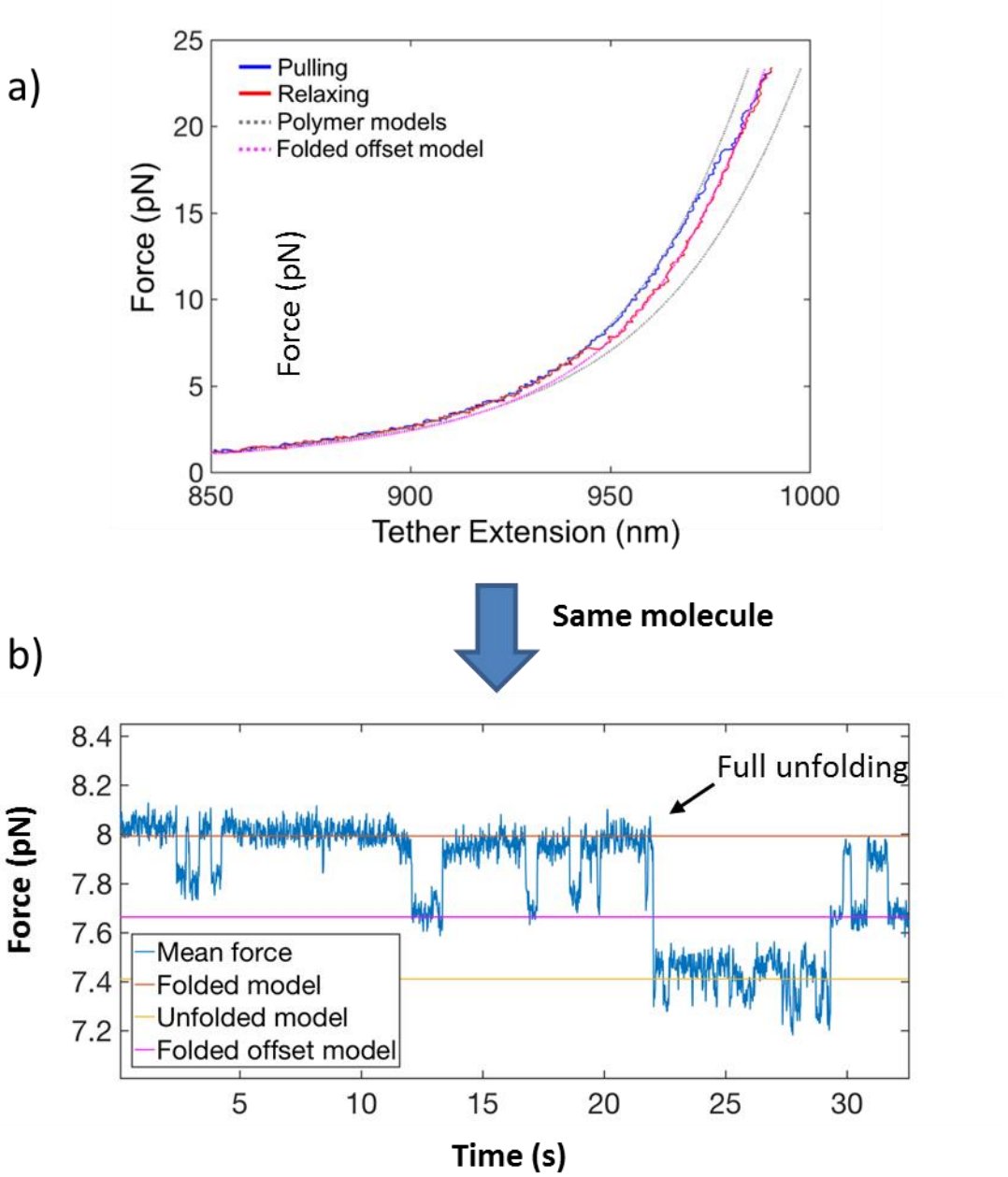

Figure 42: a) force ramp showing a molecule with an intermediate opening, b) equilibrium measurement for the same molecule showing both intermediate state, of the same opening size as in the force ramp, and fully unfolded state. Pink dotted line and pink line, which are the fit models for the small opening, are both showing the same opening size and in agreement with each other.

CHAPTER 6

DISCUSSION AND FUTURE WORK

Our force ramp results for individual GB1 molecules show that in some cases this protein only show switches between two states (folded and fully unfolded) under force and in other cases there is some evidence of intermediate states. The first behavior can be interpreted as a two-state folder, but force distributions for different loading rates indicate that although folding events are all happening at a low force of ~5pN regardless of loading rate, unfolding events are happening over a wide range of forces, including a peak at ~5pN. For a simple two-state folder with a single folding pathway, when loading force is increased, we expect to see a shift to the force distribution for unfolding events. Although we see this shift in the unfolding distribution, there is still a peak at low force for all loading rates that cannot be explained with a two-state system behavior. Our force ramp measurements for some molecules did not show any evidence of folding/unfolding events. This could be a sign that our measurements tolerance limits the ability to see slower unfolding processes that are missed by even the slowest loading rates.

Our results for the fixed trap position measurements for folding events for the cases that show only two states indicate that we are observing the same folding behavior for all the molecules. Folding rates decrease exponentially with increasing force. On the other hand, the same results for unfolding events show two different regimes. One set of molecules unfold at a constant rate regardless of force. The other group have a stronger dependence on force, though weaker than the folding constant rates. One explanation for two pathways can be found in the structure. Figure 43 shows the secondary structure of GB1 and the position of N and C termini in the structure, where

the DNA handles are attached to the protein. In this figure, the direction of the force applied by the trap is also shown. It would not be too surprising to have a weak force dependence since the pulling direction would appear to be applied to shearing adjacent beta strands rather than popping them open one bond at a time. This would suggest that there might be two or more unfolding pathways that the protein can pass to unfold, one of which is slower than the other at low forces. And the fact that most molecules seem to unfold in only one of these pathways suggests pathway switching is slow on the time scale of our observations, ~100-1000 seconds.

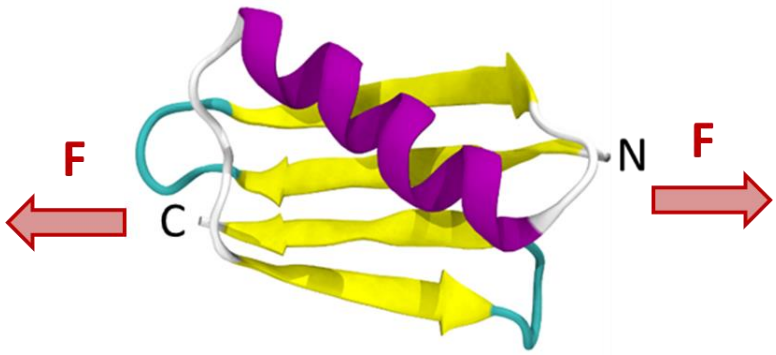

Figure 43: GB1 secondary structure with N and C termini, where the force is applied.

Molecules that show more than two states in their force ramp and/or time traces still need further investigation. Although a considerable number of molecules show at least one intermediate state involved in their folding event, the 'messy' nature of those behaviors has not let us conclude on that fraction of molecules yet. More studies on this behavior is still ongoing and better understanding of this behavior needs further investigation. Currently work in the Comstock group is focused on finding all the intermediate states and any possible switches between different unfolding regimes of GB1.

Pierse and Dudko have recently investigated the types of observations expected for protein unfolding with multiple pathways and/or intermediates [71]. They predict such insensitivity of the unfolding rate of force is indicative of two pathways of equal probability, one of which has an intermediate that acts as at trap. Figure 44 shows some of the possible scenarios that can explain our results. Figure 44a shows a case that the protein only has one native and one unfolded state, but it has two pathways that it can take to either fold or unfold. If this is the case in our results, then we should have expected the molecules to be able to switch between the two pathways and each time it folds or unfolds picks one or the other. Our time trace data shows no evidence of individual molecules switching between possible folding pathways. The other scenario is shown in Figure 44b, where the molecule has two native states and one unfolded state. In this case, the molecule can be in one of two folded states, and take one of two paths to unfold. In this case, for any fixed trap position we should expect a double exponential decay in our force distributions corresponding to two different states. However, this is not what we observed in our long time traces, as shown in Figure 36 and Figure 38. Therefore, this leads us to the other scenarios, shown in Figure 44c and d. Our results show that the GB1 molecule certainly has more than one folding pathway. These pathways could be between different native/unfolded states, Figure 36c, or they could be pathways with some on-pathway intermediate states, Figure 44d. These two examples are just some possible scenarios among all and more clarification needs further investigation. Our time traces show two different unfolding regimes, indicating that there are more than one pathways on the energy landscape of GB1 molecule. As mentioned before, individual molecules show no evidence of pathway switch in fixed trap position measurements. This could be explained by having more than one native and/or unfolded states. On the other hand, since dwell time analysis for fixed trap position data shows a single exponential decay. So, the only way GB1 could show

the unfolding force distribution peaks we observed in force ramps, is when applied force somehow help the molecule pass some barriers that will let it take an ‘off’-pathway and fold/unfold through a new folding pathway (Figure 44c).

We know that GB1 molecules are not behaving like a simple two-state folder, but it is not clear how many pathways we should consider that could explain this behavior completely. Based on Pierse and Dudko's scenario [71], the behavior we observe in our force ramps, could suggest the existence of an intermediate state, Figure 44d. The heterogeneity of the unfolding rates in our data is also suggestive of an intermediate on one pathway that has a high barrier to unfolding but both pathways are used at all forces. Two pathways between two states, folded and unfolded, that dominate at different force ranges, does not produce the broad range unfolding forces that has low dependence on pulling speed. Instead, one pathway must also include an intermediate.

In conclusion, the mechanical unfolding of GB1 is highly heterogeneous when investigated at the single-molecule level using our dual optical trap. The heterogeneity is recognizable as behaviors that occur either in equilibrium or non-equilibrium measurements of GB1. Those behaviors cannot be explained with a simple sequential unfolding pathway. This can be an indication of multipath or multistate unfolding of GB1, which is beyond a simple two-state model. One important point is that, although two different unfolding regimes are observed for fixed trap position and some non-two-state behaviors in force ramps, but we only see molecules switching between unfolding paths during force ramp measurements. We do not see this switch at fixed trap positions. This result indicates that GB1 does not act as a two-state folder in the force ramp experiments. However, individual molecules in time traces all act as two-state folders, with the caveat that they appear to be able to switch some property if you scan to the higher forces in the force ramp.

Overall, our single molecule measurements of GB1, along with other single-molecule studies on more proteins [62], suggests that mechanical unfolding through multiple pathways could be a general occurrence observed for "two-state" folders. It is also important to mention that if we only took the ensemble data, not the individual single molecule data, we would not know the existence of multiple pathways or intermediates that form within the dead time of the measuring instruments, are too brief or too unresolvable to be observed and we would instead start thinking about simpler models, like a simple two-state system model.

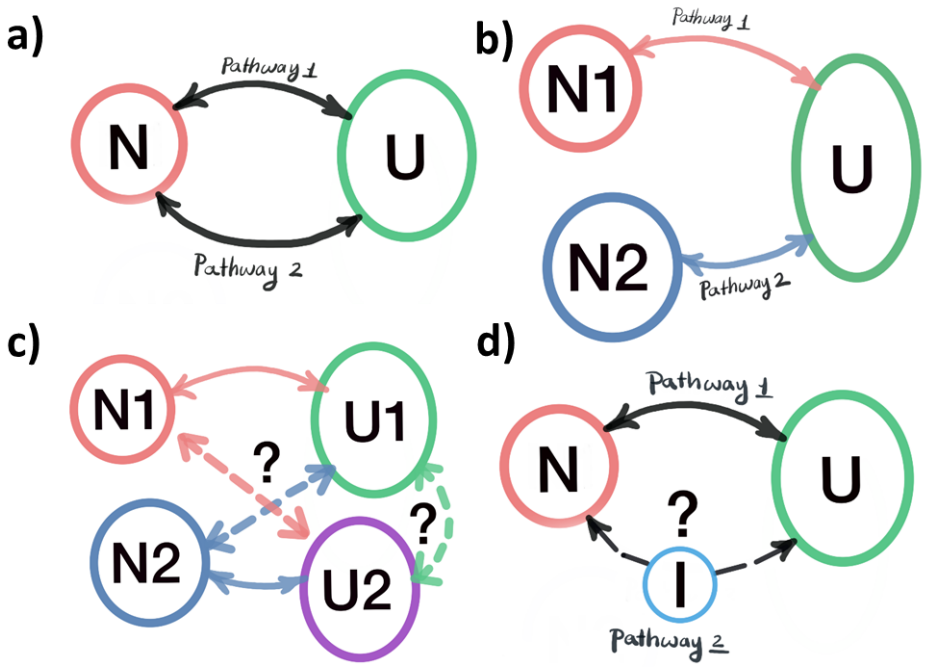

Figure 44: Different scenarios that could explain GB1 molecule behaviour. a) The protein molecule only has one native and one unfolded state, but it has two pathways that it can take to either fold or unfold. b) The protein molecule has two native states and one unfolded state. In this case, the molecule can be in one of two folded states, and take one of two paths to unfold. c, d) The protein molecule certainly has more than one folding pathway. These pathways could be between different native/unfolded states, or they could be pathways with some on-pathway intermediate states. These two are just possible scenarios among all and more clarification needs further investigation.

APPENDIX

APPENDIX A: Protocols for Performing Experiments

A.1. Model system: B1 domain of protein G (GB1)

B1 domain of protein G ($C_{283}H_{433}N_{69}O_{97}S_1$, MW: 6385.98) was ordered with 92.47% purity from NeobioLab (Woburn, MA), with alkyne functional groups, (propargylglycine, Pra), which is an unnatural amino acid added to both N and C termini of the sequence and will let the click chemistry reaction to happen between peptide and oligomers. The sequence for GB1 with 'Pra' is as follows:

Pra-

TYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

-Pra

Protein G was received from NeobioLab in lyophilized form (Figure 45). To prevent or minimize degradation, the peptide is stored in lyophilized form at −20°C, or −80°C freezer. If the peptide is in solution, freeze-thaw cycles should be avoided by freezing individual small amount aliquots in stock. We dissolve small amounts of peptide in sodium phosphate buffer (pH: 7.0, 100 mM) into desired concentration and we store them in −20°C (-80°C preferred) freezers. We use the Nanodrop spectrometer to double check the concentration of the dissolved peptide. The extinction coefficient of the bare GB1 (no 'Pra' attached) is ξ_{280}= 9530 M^{-1} cm^{-1}.

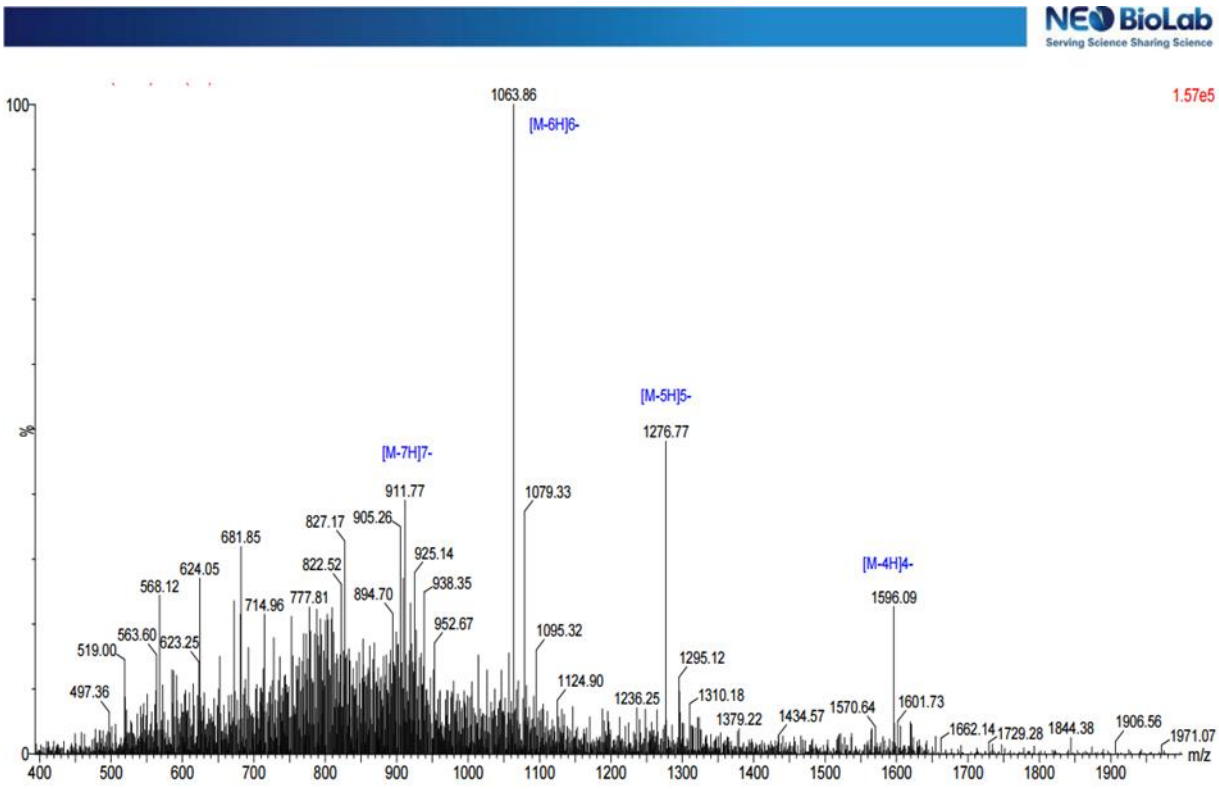

Figure 45: Mass spec peaks from NeobioLab for GB1 purification

A.2. Preparation of ~2.8 kb GB1-Pra insert (click) construct, with degenerate PspGI cut handles

A.2.1. Construct details:

This protocol makes ~2.8 kb construct with GB1 peptide insert with click-based GB1-Pra insert. Full construct has biotin and dig labels. GB1 insert is 59 aa long. LH PCR is 1704 bp, cuts to 1080 (keep) + 624 bp (discard), and RH PCR is 2397 bp, cuts to 1773 (keep) + 624 bp (discard). Final construct dsDNA composition is 2853 bp

A.2.2. Necessary consumable supplies for making construct:

All necessary supplies for PCR reaction is as follows:

- Phusion high fidelity polymerase 2x mastermix (NEB M0531S, public enzymes box, BMB order)
- DMSO (public enzymes box, comes with Phusion)
- nuclease free water (IDT, get your own tube from fridge, store on bench)
- TE, Tris pH 8.0 buffer with EDTA (IDT, get your own tube from fridge, store on bench)
- pBR322 DNA template (NEB N3033S, fridge, public tray in front, BMB order)
- DNA oligo primers (IDT, fridge, public oligo box)
- Qiagen PCR cleanup kit (BMB stock)

Note: Public enzymes box is located in the top of the -20C freezer.

All necessary supplies for restriction digest is as follows:

- PspGI (NEB R0611S, BMB order)

- Cutsmart buffer (comes with enzyme)
- Dephosphorylation reaction: CIP, calf intestinal phosphatase (NEB, BMB stock)

For gels:

- Low melting point agarose (Lifetech)
- GelGreen stain
- 1 kb DNA ladder (NEB BMB stock)
- Qiagen gel cleanup kit (BMB stock)
- Optional: Pre-cast test gel, 1% agarose, TBE, GelGreen stain (Embitec GG4700, PA order)

A.2.3. Construct Protocol

1. **PCR** to produce the two handles' initial material (left and right, 15 min prep, 1 hour wait, prep gel after starting PCR program)**.**

2. Primers should be 100 uM stock. Primers are in blue public oligo box in fridge and should be 100 uM – check label. Make sure you are using the correct version. Add **1 uL** of each forward LH and RH primers to **9 uL IDT TE** pH 8.0 (final concentration 10 uM). Add **1.5 uL** of LH_rev primer to **13.5 uL IDT TE** pH 8.0 (final concentration 10 uM, need 2x volume for this shared primer).

3. Dilute 1 uL of each template DNA 100x in IDT TE pH 8.0. Template DNA is in purple public tray in front of fridge racks.

4. Make **two** sets of the following (50 μL volume each tube, optimal for PCR):

PCR notes: NEB points out that Phusion should be added to the reaction mixture last since it has exonuclease activity and will eat at primers. Assemble the mixtures and begin PCR as soon as possible – do not let stand at RT.

Table 2: Oligonucleotides for synthesizing GB1 construct. Sequences are listed in IDT format.

Primer Name	**Sequence (IDT format, 5' to 3')**
LH forward primer	/5Biosg/ CCA GTG CTG CAA TGA TAC CG
LH reverse primer	TTG CATGAT AAA GAA GAC AGT CAT
RH forward primer	/5DigN/CGC TTT CTC ATA GCT CAC GC
RH reverse primer	TTG CATGAT AAA GAA GAC AGT CAT

Table 3: Construct necessary supplies and the volume needed.

	Left Handle (μL) (1704 bp before cut)	Right Handle (μL) (2397 bp before cut)
2x Phusion mix	**25**	**25**
FWD primer	**2.5** (LHWW_for_bio) /5Biosg/ CCAGTGCTGCAATGATACCG	**2.5** (RHww2 – for – dig) /5DigN/ CGCTTTCTCATAGCTCACGC
REV primer	**2.5** (LH – rev) TTG CATGAT AAA GAA GAC AGT CAT	**2.5** (LH – rev) same as for LH
template DNA	**1** PBR322 DNA (10 ng)	**1** PBR322 DNA (10 ng)
DMSO	**1.5**	**1.5**
dnase-free H2O	**17.5**	**17.5**

5. Place tubes in thermocycler, tighten lid, and run PCR program for 48 minutes.

Note: PCR program consists of the following steps (we suggest using a PCR thermocycler with a temperature gradient of 65-75 ºC:

Run PCR on both reaction mixes, using the following program: (1) 98 ºC for 30 seconds, (2) 98 ºC for 10 seconds, (3) 59 ºC for 10 seconds, (4) 72 ºC for 33 seconds, (5) repeat steps 2-4 thirty times, (6) 72 ºC for 5 minutes, (7) 4 ºC forever. After the PCR, we do one PCR clean up (purify).

6. Optional: Now is a good time to begin prepping the next purification gel so that it is properly set in time to run the handle purification.

7. **Qiagen** PCR clean up kit (10 min)

8. Combine the two left and right handle products respectively into one volume each (final volume 100 uL each).

9. Follow QIAquick PCR Purification Protocol. Don't add isopropanol or indicator dye. **Final volume 30 uL – let EB buffer sit in column for 5 min** before spinning to maximize collection.

10. Measure concentration of PCR products using the Nanodrop. ~100 ng/uL concentrations for each handle is typical. Peak is measured at 260 nm.

11. **Optional test gel** (for newbies, usually the PCR is very reliable): Check PCR with either a pre-cast test gel (Bio-Rad, 1% agarose, EtBr) or Epitec 1% agarose gel with GelGreen and TBE. Run gel for ~30 min at 75 V.

Table 4: Lanes to run in the gel

Handles (2 lanes)	Ladder
0.5 µL DNA	0.5 µL NEB 1
2 µL blue loading dye	2 µL dye
7.5 µL H_2O	7.5 µL H2O

Use pre-cast Epitec 1% agarose gel with GelGreen and TBE (can use lanes from a used gel < 1 week old). 10 µL per lane. Use 0.5X TBE buffer (add gel to well, cover with buffer).

DNA goes to +!

12. **Digestion** to cut the handles (1 hour, for optimal efficiency, prep gels now). Incubate in the PCR thermocycler. Enzymes are in public enzyme -20 C freezer box. Always keep enzymes on ice when not in freezer! If planning on running a test gel, consider saving 1 uL of undigested handle to run in lane adjacent to cut handles. These digestions are performed to produce 5' overhangs for both LH and RH, which will eventually be used to base-pair to Insert.

13. Left handle - cuts to 1080 bp (keep) + 624 bp (discard):

Add to DNA: **3.5 µL Cutsmart** buffer (NEB) + **2 µL PspGI** (NEB R0611S, 10 kU/mL). Incubate for 1 hour at 75 C.

14. Right handle - cuts to 1773 bp (keep) + 624 bp (discard):

Add to DNA: **3.5 µL Cutsmart** buffer (same) + **2 µL PspGI** (same). Incubate 1 hour at 75 C.

15. **Optional test gel**: Run: ladder, left, left cut, right, right cut (as above).

16. **Dephosphorylation** of left and right handle overhangs and deactivating Alkaline Phosphatase Calf Intestinal (CIP).

Note: The PspGI overhang for these handles tend to ligate to other PspGI overhangs (the other handles, not quite clear why). Dephosphorylation of the handle 5' overhang prevents this unwanted handle-handle ligation, maximizing yield of the desired product but also removing the presence of potentially confusing gel bands. Note: there is no need to 'kill' the CIP reaction since we proceed directly to the gel purification (or Qiagen cleanup). If you needed to stop the CIP reaction, you could do so by adding EDTA. Make sure the EDTA will not interfere with downstream reactions (see NEB)

17. Add **1 µL of CIP** enzyme to each handle. No additional buffer needs to be added in this case, as the Cutsmart added previously serves as a sufficient buffer. If the handle sample is not still in the Cutsmart buffer, then the CIP buffer should be added (see NEB).

18. Incubate for **30 min at 37 C**.

19. Optional: Run **Qiagen PCR clean up** kit as above (30 uL final EB volume with 5 min incubation). Nanodrop to verify presence of DNA and check concentrations. Should be ~50 ng/ µL (lose ~1/2).

20. **Gel purify handles** (1 hour, 2 hour gel prep, start gel prep earlier)

21. **Make gel**. Add 50 ml 0.5x TBE buffer, 0.5 g agarose (1%) LMP (low melting point) to a ~100 µL Erlenmeyer flask (add the buffer first, then the agarose to dissolve most easily). Microwave until solution just starts to boil and then swirl to mix. Microwave a second time and mix to make sure all gel is completely dissolved and there are no slivers left in the solution. Cool for ~ 20 or 30 minutes, until solution is still warm but not hot, then add 2 µL GelGreen. While cooling, level a gel box. Pour cooled gel into box. Remove any bubbles by dabbing at them with the corner of a Kimwipe. Place comb into middle spot. Wait 60 to 90 minutes at RT to solidify

(cover with foil). Cover the set up with foil to prevent any contamination. Use the gel the same day, and store it in fridge if not used right away (don't keep it overnight).

22. **Add to PCR products** (~30 µL each):

3 (5 without loading dye) µL 100% glycerol (needed to keep in well, ~10% total)

Optional: 2 µL loading dye (unnecessary with practice). Though it seems the loading dye may improve the sharpness of bands. Run **one lane for each handle** (~35 µL handles per lane).

23. Add a **ladder lane** (same as test gel above, NEB 1 kb ladder, ~10 µL).

24. **Run gel for 60 min @ 85 V**. This should be sufficient time to more than adequately resolve the fragments.

25. Turn on **heating block to 50 C** (to dissolve gel pieces later).

26. **Weigh empty 1.5 mL tube** for each *lane* and write weight and label on tubes.

27. Visualize bands with **blue LED lamp** and **orange filter** plate or glasses. Turning off lab lights helps.

Note: The desired cut left handle should be 1080 bp long (discard 620 bp band). The desired cut right handle should be 1773 bp long (discard 620 bp band). (They should be the second to last band in each lane, with the cut ends appearing below them. Nearly 100% digestion yield is typical. The uncut DNA band may not be visible.)

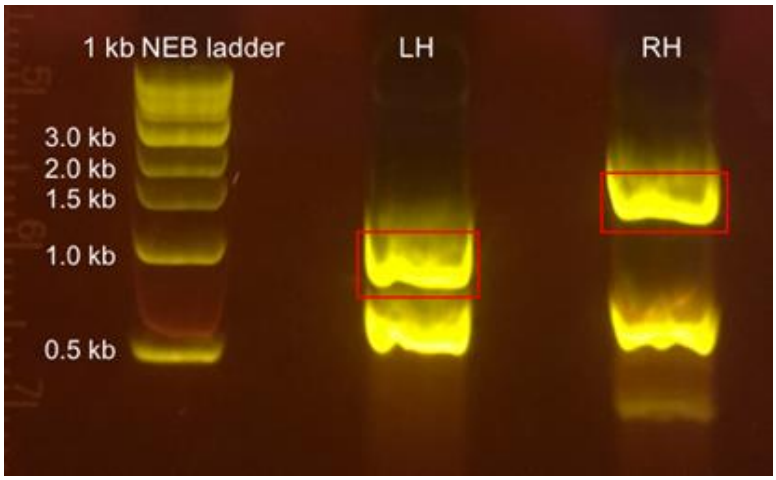

Figure 46: Gel purification of handles visualized with blue LED lamp. The lane on the left is the ladder, the lane in the middle is the left handle and the right lane is the right handle.

28. Using a fresh razor blade, **cut out the desired handle** from each lane and put each into separate pre-weighed tubes. You can cut directly on the gel tray. Little force is needed, cut gently to not needlessly scratch the tray.

29. **Weigh tubes** again and subtract tube weight to get gel weight. Gel slice must be ≤ 400 mg for proper gel extraction cleanup.

30. Do Qiagen **QIAquick gel extraction kit** protocol. (Like PCR purification kit after the initial gel removal step). Depending on gel piece size, may need to load tube more than once (load all of one handle into a single tube in multiple steps). **NO** step 4 (no isopropanol). **Do** add perform optional **extra 500 µL QG gel rinse**. **Do** let **incubate PE rinse for 5 min** before spinning to remove all QG salt. If not, subsequent Nanodrop DNA spectra will be obscured by large low nm peak. Choose more concentrated final volume: 30 µL EB, 5 min incubation.

31. Measure the concentration of your handles using the Nanodrop (be sure to blank with EB).

Note: Handles are now done. At this step, we are ready to move on to ligation step.

32. **Ligate handles** to prepared peptide-oligo complex to form the final construct product. (Should pour the final 1.5% gel before setting up and starting reactions, if planning on running gel right after)

Note: To achieve a good yield of the final construct, combine the three components LH:insert:RH as 1:5:1. We use a **ligation volume of ~80 μL**, **8 μL 10x T4 ligase buffer** and **4 μL T4 ligase** (NEB M0202S, 400 kU/mL) (should usually be a good amount).

Choose handle and linker (peptide insert) volumes such that they will be added in 1:5:1 stoichiometry while having a final volume together of ≤ 68 μL. If < 68 μL total, bring the volume up to 80 μL with nuclease free water. For molar conversions:

Left Handle: ng/ μL x 1.43 gives nM

Right Handle: ng/ μL x 0.87 gives nM

a. Add your handles and linker (peptide insert) mix (your calculated volumes) and the ligase + ligase buffer in a PCR tube and place into thermocycler. Run at **23 C for 30 minutes** (ligation step) **then 65 C for 15 minutes** (kill the ligase) - "Ligate" program.

Note 1: It is likely that the peptide insert concentration is not really 5x the DNA. For most recently prepped inserts, samples consistently measure A280 ~ 0.10, giving ~400 nM sample. But the spectra are poor with large negative regions (bad blank). The used blank is 50 mM NaPi buffer, pH: 7.18, which is the same buffer used in size exclusion column. Assuming the sample is 400 nM and putting in 5x will be consistent with used recipe and should give a usable if not very high yield of final construct.

Note 2: Linker (peptide insert) preparation is explained in detail in the next section.

b. **Optional: Qiagen** PCR cleanup of ligation reaction. If you are not going to run the sample on a gel immediately (e.g., saving for the next day), then you should definitely run the

Qiagen cleanup kit on the sample to put the DNA into a simple and safe storage buffer (EB – Tris). Further, even if running a gel immediately, the gel will likely run better (cleaner, sharper bands, much less streaking) if you do the cleanup before running the gel. It is unclear what is optimal. There will be some sample loss with the Qiagen kit (and it takes a little extra time), but there will also be some sample loss with a bad gel run (streaking) and it could be harder to distinguish bands. It is unclear if performing the Qiagen cleanup step is better or worse overall.

c. **Gel purify final construct (post-stain):** We use a higher % agarose gel and run longer to better separate out the long, 2-4 kb, ligation products. The voltage is lower to not overheat the gel. We get cleaner bands if we *do not include GelGreen* stain in the gel casting, but rather do post staining after the construct is finished running on the gel.

d. **Make** a **1.5%** agarose gel ***without any stain***. Similar as above, but add **0.75 g agarose** to **50 ml 0.5x TBE** (DON'T add GelGreen). Put the comb in the farthest edge location (not the middle). Let stand on the bench for 2 hours before using (cover with Al foil to keep it protected). Very carefully pull the gel away from the mold on the comb end, using a razor blade to assist. It is very easy to mess up and tear the wells. If the wells on one end are torn, undamaged wells can still be used.

e. **Run gel** for **210 min at 80 V**. Use two lanes, ~40 μL each, and a ladder lane (~10 uL). It is also helpful to run excess unused handle as a reference. Use the same amount of handle as used in one lane.

f. **Stain the gel**. Standard stain: mix **135 mL MQ** water, **15 mL 1 M NaCl**, and **45 uL GelGreen** dye in flask. Place gel in clean box (e.g., use leftover tip box cover, rinse out) within larger secondary tray (to catch spills). Cover gel with stain. Cover small box with foil. Gently agitate box (swish fluid) every ~5 min. 30 min of staining will penetrate ~1/4 of gel from top and

bottom (e.g., middle of gel will not be stained). This should be enough to visualize the bands. Visualize gel on the blue light box. The gel can be stained longer if desired (no problems seen after 1 hour).

g. **Cut band.** There will be a substantial amount leftover handles present in the lanes (LH 1080 and RH 1773 bp). Above the handles, there should be three product bands, likely substantially dimmer. The desired final construct band should be the middle of the three products at 2853 bp.

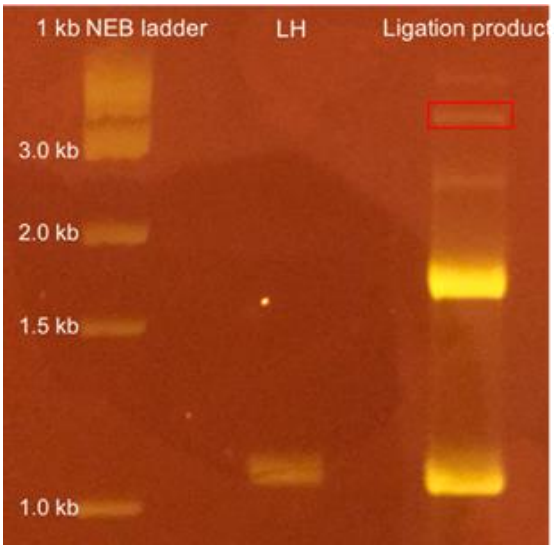

Figure 47: Ligation gel showing the ladder lane, left handle, and ligation product.

Note: Improper binding of handles to each other would also contribute to the three product bands, but should not occur assuming the dephosphorylation reaction was successful.

h. **Qiagen purify** the cut gel bands (as above) with two columns (one for each lane). Do the **extra 500 uL QG wash**. Let the **PE wash stand 5 min** before spinning. Use **30 µL EB** for final elutions and let **stand 5 min** before spinning.

i. **Measure concentration** of construct with Nanodrop. ~4 ng/uL is typical. It is often difficult to get completely clean DNA spectra. Residual QG salt not washed out contributes to absorption < 260 nm and can distort the DNA measurement.

j. **Store construct in the fridge.**

k. Try adding 5 µL construct to 1 µL streptavidin beads for the first trapping session.

Molar Conversions

L-link oligo: ε_{280}= 113,586 L/M

R-link oligo: ε_{280}= 144,138 L/M

Peptide: ε_{280}= 9,530 L/M

Full insert ε_{280}= 267,254 L/M

Left Handle: 1 ng/ µL = 1.43 nM

Right Handle: 1 ng/ µL = 0.87 nM

A.2.4. Preparation of the GB1 linker (peptide insert with GB1-Pra) using click chemistry reaction

As part of construct preparation, we need to make the linker, which attaches to the LH and RH in the ligation step. This linker consists of the GB1 protein in the middle, with alkyne groups (Pra) attached to both N and C termini, and two identical single stranded oligos with 20nt length that have azide groups at one end to connect to the peptide. These oligos attach to the overhangs of the double stranded digested oligos during the ligation process. Figure 48 shows a schematic view of the construct.

Figure 48: Schematic view of DNA-protein chimeras in click chemistry reaction

The reaction is catalyzed by Cu(I) in complex with THPTA. The CuSO4 provides Cu(II) and the sodium ascorbate provides the reduction to Cu(I). Dissolved O2 oxidizes the Cu, so a sufficient excess supply of sodium ascorbate is necessary throughout the reaction. THPTA is a Cu chelator: It binds to the Cu and facilitates catalysis while reducing the quantity of reactive oxygen species generated. It seems that THPTA is sacrificial, so at least five equivalents of THPTA should be used relative to Cu. The Cu concentration should be between 50 and 100 uM. The lower limit is necessary to achieve a sufficient concentration of the proper catalytic complex, which includes more than one metal center. Amino guanidine is included to prevent byproducts of ascorbate oxidation that can covalently modify or crosslink proteins.

A.2.5. Supplies for making insert:

- L link-azide 20” oligo (5'- /5Phos/CCT TTT TTT TTT TTT TTT TT/3AzideN/-3', IDT): the click labeled ss-oligo thar attaches to both NC ends of the GB1 peptide.
- Sodium phosphate (pH 7.0, prepared in the lab): is used as the reaction buffer
- Sodium Ascorbate (Sigma-Aldrich): Sodium ascorbate goes bad on the shelf if reacts with O2 gas. Purge the bottle with N2 gas before closing.
- THPTA (Tris(3-hydroxypropyltriazolylmethyl)amine, 95% Pure, Sigma-Aldrich
- Aminoguanidine hydrochloride (≥98% pure, Sigma_Aldrich)
- CuSO4: prepared as a 1M stock in the lab and stored at RT.
- dH2O: MQ is used for prepping all solutions
- SE column: first prep the hplc to be ready for the end of the reaction.
- HPLC buffers:

1000 mL of 50 mM sodium phosphate buffer pH:7.0-7.2

500 mL of 20% volume ethanol

1000 mL of MQ water

Note 1: It is very important to immediately purify the protein-insert reaction product from the reaction buffer via the size-exclusion HPLC protocol upon completion of the reaction. The Cu click reaction can potentially damage the peptide and DNA molecules, especially any remaining Cu (e.g., buffer not fully exchanged). Cu(II) can apparently catalyze alkyne-alkyne conjugation which would create peptide multimers. Cu(II) can become prevalent at the end of the reaction if the buffer is exposed to air and sodium ascorbate runs out. Thus it is important purify the insert samples from the reaction buffer directly after the completion of the reaction.

Note 2: We pre-mix CuSO4 with the THPTA ligand before adding it to the click reaction solution.

Note 3: We initiate the click reaction finally by the addition of sodium ascorbate to the reaction solution.

Note 4: TCEP (Tris (carboxyethyl) phosphine) is not recommended as a reducing agent instead of sodium ascorbate, because it is an inhibitor of the reaction and can react independently with azides.

Note 5: Tris buffer cannot be used for any component since the tris (hydroxymethyl) aminomethane molecule is a competitive and inhibitory ligand for Cu.

Note 6: Our linker oligos need to be rehydrated in IDT DNAse free water (NOT MQ, NOT any Tris based buffer). But they won't stay good as long in water. The DNA itself can be damaged (bases can be damaged, hydrolysis of backbone) and the NHS-ester coupled azide or alkyne label can be cleaved off. It is best to hydrate in water, then aliquot and freeze samples in the -20. IDT suggests this sample should last only ~6 months in fridge in water, but > 1 year if frozen in water. Don't freeze/thaw the oligos: use a single aliquot for each reaction.

A.2.6. Click chemistry reaction protocol:

Solutions to prepare fresh each time:

1. 100 mM sodium ascorbate in MQ. Add 20 mg sodium ascorbate to 1 mL of water. Back fill sodium ascorbate bottle with dry N2 gas.

2. 50 mM THPTA in MQ. Add 10 mg THPTA to 460 µL water.

Stock solutions needed:

1. 20 mM CuSO4 in MQ. 10 mL volume prepared. Store at room temperature. CuSO4 (MW = 159).

2. 10 mM aminoguanidine hydrochloride. Add 1.1 mg aminoguanidine hydrochloride (MW = 110.55) to 1 mL of water.

3. 1 M sodium phosphate buffer, pH 7.0.

4. GB1 peptide. Stock is aliquoted in -80 C freezer. This concentration ideally should be the same as when you made the stock, but due to evaporation of the buffer, or protein sticking to the side of the stock tubes over time, it is recommended to measure the concentration of the peptide after being thawed using Nanodrop. GB1 MW = 6358.98 Da, GB1 extinction coefficient = 9530 M-1cm-1.

Click Reaction:

1) **Turn on the refrigerated centrifuge** and set to 4 C (for the later step of concentrating the size exclusion purified samples – no need to spin).

2) Prepare the following **fresh click reaction solutions** each time:

a. 100 mM sodium ascorbate: Add **20 mg sodium ascorbate** to **1 mL of MQ** water (use small cut sheet of weighing paper). **Vortex** to dissolve. **Back fill** sodium ascorbate bottle with dry N2 gas (be sure to purge the line first). Wrap in parafilm.

b. 50 mM THPTA: Add **10 mg THPTA** to **460 µL MQ** water. THPTA is a yellow powder. **Vortex** to dissolve. *Be careful not to confuse it with TBTA, which is white and will not dissolve in water.*

c. Premix 5µL of 20 mM CuSO4 stock and 10 µL of the above THPTA solution.

d. If necessary: 100 µM azide oligo. May have stock, if not, dissolve azide linker oligo into nuclease free water to 100 µM. DO NOT use Tris (interferes with click reaction) or MQ (bad

pH). Store in fridge. May last ~6 months, perhaps longer, in the fridge. Do not freeze thaw repeatedly.

3) Prepare click reaction volume. Use 200-µL low-binding tubes. Make two sets of the following (final reaction volume is 100 µL):

1) Calculate how much peptide volume you will use, and add nuclease-free dH2O so that the total peptide volume + water volume = 38.5 µL.

2) Add 10 µL of 1.0 M Na phosphate buffer pH 7.0 (final concentration: 100 mM)

3) Add 5.0 µL of amino guanidine HCl (final concentration: 5 mM)

4) Add 40 µL of 100 µM azide-oligo (final concentration: 40 µM)

5) Add peptide to achieve 20 µM final concentration (volume depends on the stock concentration)

6) Add 1.5 µL of CuSO4/THPTA mixture to the tube, then without delay:

7) **Finally:** Add 5.0 µL of 100 mM Na ascorbate. **This starts the reaction.**

8) Place the tubes in the thermocycler and run at 50 C for two hours. (Program: No. 34)

9) Be ready to immediately purify the sample using the size exclusion HPLC method as soon as the reaction finishes.

3) While waiting for the click reaction incubation, set up the PAGE gel (see below).

After the click reaction completes:

4) Set aside 10 µL of sample from each tube for running in the PAGE gel.

a) Run the PAGE gel as soon as possible while not delaying in performing the HPLC purification.

5) Purify the sample using the size exclusion HPLC method (see below).

The sample should be purified without delay. The sample decays and possibly oligomerizes when it continues to sit in the click reaction solution. Samples have shown slow changes and decay previously when sitting for only ~25 minutes. The entire reaction should be injected into the HPLC and purified in a single shot if possible. Use a large enough insertion loop. We have a 100 uL loop. Excite and measure at 260 nm (optimized to measure the DNA best).

Figure 49 shows all the expected peaks. The desired sample which includes the peptide + two oligos should be the first peak to come off, and the peak should occur between 16 and 18 minutes (peak tat 17.1 minutes).

6) Concentrate the collected sample using a GE Healthcare Vivaspin 500 centrifuge concentrator column (3 KD MW cutoff and max volume of 500 µL). Fill multiple (normally four) centrifuge columns with 500 µL of the desired sample. Spin at 12000 rcf for15 minutes in the cold centrifuge (at 4C). Check your sample to see how much of the buffer is gone. Repeat this process for longer time if needed.

7) Divide the concentrated sample into 20 µL aliquots (use the aliquotting pipette) and store in the -80 freezer.

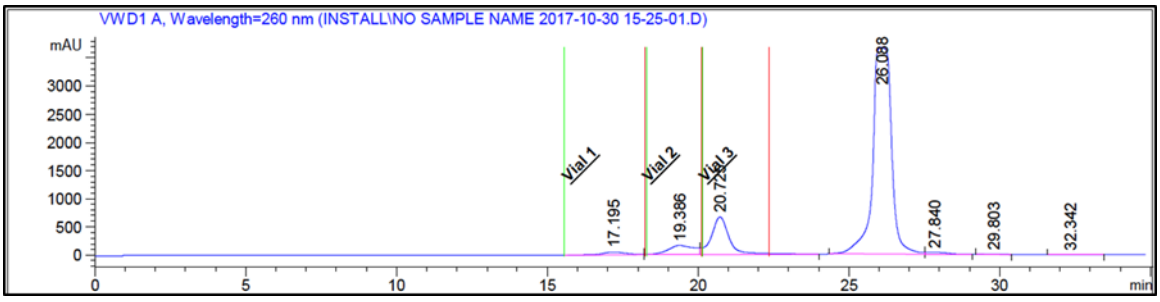

Figure 49: Size exclusion purification of oligo-protein Insert. Vial 1 is showing the desired final Insert. All vials are collected manually.

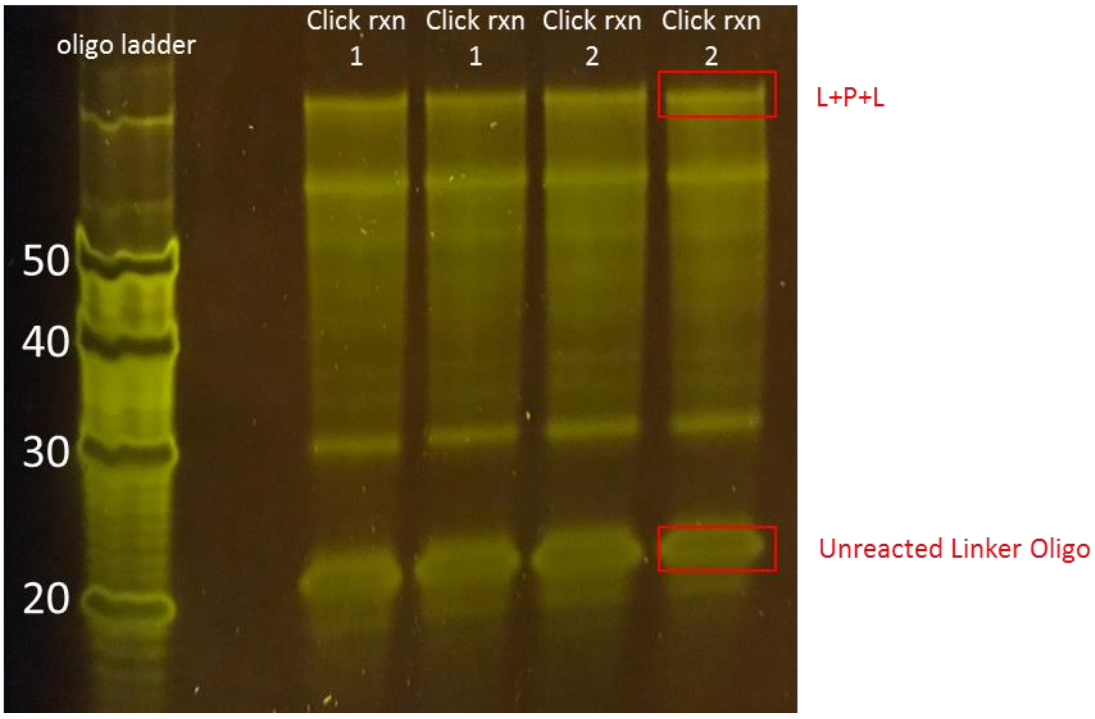

Figure 50: PAGE gel showing ladder of five different oligo length (10 nt, 20 nt, 30 nt, 40 nt, 50 nt) and two lanes of the samples. Red box shows the GB1-insert lane.

A.2.7. Purification Steps

1) SE column preparation:

Supplies:

i. Superdex 200 Increase, 10/300 size exclusion column (GE)

ii. 50 mM sodium phosphate buffer pH:7.0-7.2 (prep 1 L at a time)

iii. 20% by volume ethanol (prep 500 mL at a time)

iv. MQ water (prep 1 L at a time)

Switch to ‘method and run control’

Choose setup method, choose which bottle (C in our case for water). Can also set the flow rate here. The max pressure should already be set to 50 bar.

Prepare the SE column:

b. Starting up the HPLC

a. Verify that the 100 uL sample loop is attached.

b. Startup the ‘online’ version of the software (desktop icon)

c. Open the method ‘Miles_superdex’

d. Switch to the ‘method and run control’ tab

e. Choose ‘setup method’.

i. From here you can select which bottle to flow. Check the bottle labels on the tubing connected to the bottles.

ii. You can also set the flow rate here.

iii. The max pressure should be 50 bar.

f. The following buffers are required (do not need to be fresh). These buffers need to be 0.2 um filtered before using. It is not necessary to degas them since our HPLC machine includes a degassing segment. They do not all need to be attached at once, depending on the usual HPLC setup. E.g., sometimes either the buffer or ethanol are swapped to the same channel.

i. 50 mM sodium phosphate buffer pH:7.0-7.2

ii. 20% by volume ethanol

iii. MQ water

g. Purge air from the solution tubing lines at the pump. When the HLPC sits or bottles/tubing are unused, air bubbles collect in the tubing. It is very important to purge any bubbles from those lines. For each tubing line you plan to use (e.g., C, D, etc.):

i. Switch the black knob at the pump (beneath the bottle rack) so that the pump output is set to the purge line

ii. Begin pumping

iii. Watch the output tubing for bubbles. Continue pumping until all the bubbles appear to have been flushed through

iv. Repeat for each bottle you plan to use (switch bottles in the 'setup method' window)

v. Turn the pump knob back to send the output into the HPLC

c. Preparing the size exclusion column

a. Wash and equilibrate the column in the desired purification buffer: before running the sample, the SE column needs to be equilibrated first with two-column volumes of MQ water and then with at least one-column volume of the desired buffer. The column volume is 25 mL. Use a flow rate of 0.75 mL/min,

i. Run water for 66 minutes

ii. Run the elution buffer for 33 minutes

b. Clean the injection syringe

i. Withdraw and inject a few volumes of ethanol

ii. Withdraw and inject a few volumes of MQ water

c. Clean the sample loop

i. Switch the loop to inject

ii. Using the injection syringe, inject multiple volumes of ethanol, then buffer.

d. Check cleanliness of the setup with a blank buffer test run

e. After and during the click reaction: Keep a continuous, low (0.1 mL/min), flow of buffer through the column while waiting for the click reaction to complete.

d. Run the click sample

a. Rotate the sample injection switch to 'inject'.

b. Insert

c. by injecting it in to the loop. A new run is automatically recorded when the sample injector switch is flipped from inject to load.

e. Collect all peaks in separate glass tubes incase extra test is needed on any of the collected samples. You can either collect automatically (when you know exactly at what time each sample is coming through) or collect manually by hitting the collect button at the beginning and ending of each peak.

Column maintenance and storage:

If the column will not be used for greater than two days, fill the column with 20% ethanol to store it. Use a 0.75 mL/min flow rate and flow for at least 2 hours. Be careful not to exceed the maximum pressure of 50 bar during equilibration. A high pressure may indicate that ethanol gas

bubbles have formed. Decrease the flow rate. The column volume is 25 mL. Store the column in the HPLC drawer underneath the HPLC bench.

2) Urea-PAGE Diagnostic Protocol (20% polyacrylamide):

In addition to the size-exclusion HPLC method, we can also visualize the click reaction products by running them on a denaturing PAGE (poly acrylamide) gel containing urea. This gel is optimized for sorting products according primarily by their ssDNA oligo content. Urea (as opposed to SDS) is optimal for denaturing nucleic acids. With this gel, we can get nucleotide and functional group level resolution on individual oligos, and distinguish between particular double-oligo labeled peptides. Together with the size exclusion method, it is a powerful method of identifying reaction products.

We run a 20% polyacrylamide gel to best separate our samples.

To get the cleanest bands:

- You want as close as possible the same concentration of TBE buffer in the gel and in the gel apparatus wells. Previous protocol had gel with a final TBE concentration of 1.4x for some unknown reason, while the running buffer is 0.5x! In the future: using the same 10x TBE stock, each time prepare an amount of 1x TBE buffer as running buffer, and adjust the TBE in the gel to make a 1x TBE gel. Do the best the match the TBE concentration in the gel and running buffer.
- Fill unused wells with loading buffer.
- Run the samples as soon as possible after the pre-run so that the gel is still warm, while still thoroughly rinsing the wells prior.

Supplies needed:

- 30% acrylamide:bis-acrylamide (29:1) (stored in the fridge door, Bio-Rad #161-0156)
- 10X TBE buffer (Fisher, BP1333-1)
- Urea (Sigma-Aldrich, U5128-5KG, big bucket)
- Ammoniumpersulfate (APS), (from Chem storeroom, >98%, Jade)
- TEMED (Bio-Rad, 161-0800, 5 mL)
- 0.5 M EDTA
- 1 M TrisHCl pH 7.5
- ssDNA for ladder, simple poly-dT oligos (100 uM, 'Ladder 20mer', 'Ladder 30mer', 'Ladder 40mer' and 'Ladder 50mer', IDT, standard desalted, in Dena's fridge box)
- SYBR Gold stain (Invitrogen, S11494, stored in -20 freezer)

The PAGE apparatus includes (all sorted in PAGE drawer beneath PAGE bench area):

1. Two glass plates: a rectangle one and one with comb space
2. Comb (10-well)
3. Yellow rubber band for sealing the glass
4. Vertical black spacers (for spacing the glass plates)
5. Tray for buffer overflow for pouring gel
6. Cell buffer dam (fill with buffer above/below gel)
7. Blue clips for clamping plates together and white clips for holding the glass plates vertically while the gel is running
8. High voltage power supply

Making the PAGE gel:

1. Assemble the glass PAGE gel forms:

a. Rinse glass in DI water.

b. Tight the yellow rubber edge sealing band around the uncut glass (square corners).

c. Add the two vertical black spacers right next to the long edges of the yellow band.

d. Add the cut plate to the top.

e. Clamp with the blue clips, two per side.

f. Verify that the comb fits into the top slot.

g. Set the glass plates horizontally on a rack (Styrofoam pieces) inside a tray to catch the mess later.

2. Prepare the initial PAGE-gel solution:

In a 200 ml beaker (use the usual one labeled PAGE, the spout makes it easier to pour), add the following ingredients:

a. 20 mL of 30% acrylamide:bis-acrylamide (we use a ratio of 29:1 acrylamide:bis-acrylamide) solution

b. 5 mL of 10X TBE buffer (Fisher, BP1333-1)

c. 5 mL of MQ water

d. Add 24 grams of urea (top shelf glass cabinet, big bucket).

e. Mix all the above ingredients and microwave for ~20 seconds. Continue to swirl until all the urea is dissolved.

3. Make a 10% solution by weight of ammonium persulfate (APS). The APS solution can be stored in fridge for a week only. Be careful with APS – it is a skin/eye irritant.

a. Add 1 g APS to 1 mL MQ.

4. Polymerize the gel:

Add to the previous PAGE gel solution to begin polymerization and setting of the gel:

a. 140 µL of APS

b. 40 µL of TEMED (in cabinet). IMPORTANT: gel will quickly begin to polymerize after adding TEMED. Do not delay in adding the gel solution to the forms.

c. Swirl the solution gently but thoroughly to mix.

5. Pour the PAGE gel solution into the gap between the glass plates from one corner, tilting to slowly and evenly fill from one side/corner to the next. Try to keep any bubbles to the opposite side. Sometimes small bubbles will make the last well not usable, but the rest of the wells should be fine.

6. Insert the well comb, being careful to not trap any air bubbles beneath the comb teeth.

7. Set the gel by incubating for 60 minutes with the glass plates placed horizontally on top of Styrofoam or tube racks inside the tray.

Run the gel:

8. Setup the gel:

a. Remove the rubber band from the side and clean the surface of the glass plates using a razor blade and possibly some DI water.

b. Take the glass plates and set them vertically in the cell buffer dam with the aluminum heat sink on the back (positioned near the top) and clip using two of the white clips.

c. Pour 0.5x TBE (*same batch as used to make the gel*) into the upper and lower chambers until the buffer surface reaches the required levels.

9. Pre-run the gel: Pre-run the gel to *remove free urea* and to *warm the gel* prior to running your sample. Turn on the power supply, click on manual (push 3rd button from left) and

change power to 20 W (constant power mode). Hit RUN and run for 30 minutes. Typical output is ~950 V and ~25 mA.

10. While waiting, if necessary: Make the loading buffer: The loading buffer can be stored indefinitely.

a. 24 grams of urea

b. 2 mL of EDTA 0.5 M

c. 0.1 mL Tris 1M, pH:7.5

d. Bring the solution to 50 mL with MQ.

e. Microwave for 20 seconds and vortex to dissolve urea.

11. While waiting: Prepare samples with loading buffer. The wells can safely hold 40 uL. Generally, we dilute samples 50% with loading buffer.

a. For each click sample, combine 10 uL of sample with 10 uL loading buffer

b. Ladder: the ladder is made by combining some simple poly-dT oligos of different length. Oligo stock should be 100 uM. Add 2.5 ul each of desired oligo e.g., 20, 30, 40 and 50 nt long. Add 10 uL loading buffer.

12. After pre-run: quickly clean wells, load samples, and run gel. You want the gel to still be warm when you start the sample run.

a. Clean up wells: After the initial run, some urea-slurry will be seen in the bottom of the wells (~1/2 the well volume). Turn off the power supply. Pipette out slurry using a disposable syringe and ‘blue’ needle. May need to suck out multiple times until clean. Be sure there are no bubbles in the wells. Pipette additional 0.5x TBE in if necessary.

b. Load samples into wells. Be sure to check wells for damage. It can be helpful to use a marker to number each well on the glass that you are going to use. Use a usual pipette with the tip. Add blank loading buffer to ‘empty’ wells (unused, no samples) to improve run quality.

c. Run gel: as above, 20 W for 1 hour.

15. Visualize the gel

a. Prepare the staining buffer

a. 200 mL of 0.5X TBE

b. 20 µL of SYBR Gold stain

b. Fill staining tray (same tray as before) with stain.

c. Turn gel power supply off.

d. Remove the plates from the setup.

e. Removing the gel from the plates: It is very easy to damage the gel – be very careful! Use the special ‘gel removal’ metal sheet (hanging on the drying rack). Gently work top plat off. Put gel + bottom plate into stain and slide the gel off the remaining glass plate.

f. Incubate gel for 45 minutes with gentle swishing every 10-15 min.

g. Illuminate with blue light from Dark Reader. Use the big spatula to transfer the gel to the illuminator.

A.3. Poxy preparation:

Materials:

Pyranose oxidase: Sigma, -20 freezer.

Catalase solution: catalase from bovine liver, 5 mu, EMD-Millipore (formerly Calbiochem), 219001-5MU. Reconstitute into 11 mL T50 and refrigerate.

T50 buffer: standard, 50 mM NaCl and 20 mM Tris-HCl pH 8.0.

Protocol:

1. In a 0.6 mL centrifuge tube, add 20 μL catalase solution (fridge) to 80 μL pH 8.0 T50 buffer.

2. Weigh out 5.8 mg (or recently 2.5 mg seemed sufficient, mjc, 170616) pyranose oxidase (-20 freezer) onto creased paper and add to tube.

3. Carefully dissolve by flicking the tube by hand. We do not want to create foamy bubbles. This is a possible sign of damage to protein. A quick spin in the centrifuge can help to bring down power from the cap area.

4. Centrifuge once at 11 krpm for 5 m to remove any possible junk particles (possibly from catalase), keep the supernatant.

5. Centrifuge filter once at 11 krpm for 1 m (keep the flow through).

6. This tube of poxy should be stored in the fridge (don't freeze/thaw) and used within two weeks. When removing poxy from the tube for use, DO NOT mix the tube! Over time, protein will aggregate and precipitate out of the poxy solution, collecting in the bottom of the tube so much so that you may eventually see a dark brown precipitate in the bottom of the tube.

Poxy taken from the top of the tube may still work, but at this point it is best to prepare fresh poxy.

A.4. Making T50 buffer

For making 50 mL of T50 we use the following protocol:

1) Add 2.5 mL of NaCl 1M (50 mM concentration) to a 50-mL centrifuge tube.

2) Add 1 mL of Tris-Hcl 1M (20 mM concentration); pH: 8.0 to the tube.

3) Add MQ water to reach 50 mL total volume. Filter solution with milipore filters into a fresh tube. Store T50 buffer at room temperature.

A.5. Sample chamber making

Sample flow chambers are used for our experiments to provide an appropriate environment for flowing buffer, DNA and beads. As explained in Chapter 3, sample chambers we use are made of two No. 1 microscope coverslips that contains a layer of sealing film (Parafilm) sandwiched between them. The following describes how we do that:

1) Cut a 4" × 4" square of Parafilm. Place the film, while pulled tight, over a metal frame and secure the edges with paper tape.

2) We use CorelDraw to draw the patterns. The patterns can be reused generally, you do not need to 'make a blueprint' each time. Make a blueprint of the flow channels if you are making a new chamber design.

3) Set the following settings for the laser engraver: Power 8%, Speed 10%, PPI 500, Z-Axis 4".

4) Turn on the air compressor connected to the laser engraver.

5) Run the engraver program and print one copy.

6) When done with engraving, we use tweezers to pull off carefully any loose segments of Parafilm that may be still attached to the channel structure.

The following steps describe how to assemble a flow chamber:

7) Use tweezers to carefully lay down the cut Parafilm on the coverslip with holes, aligning the Parafilm channels with the holes. Moving the Parafilm around tends to leave residue on the coverslips.

8) Use a fresh, clean razor blade to cut two glass capillaries to a length such that they will span the Parafilm between the outer channels and the inner channel.

9) Use tweezers to place the two capillaries at the appropriate positions on the Parafilm with about 45 degrees to the channels. See Figure 30c.

10) Lay the uncut coverslip on top of the film to form a “Parafilm sandwich” between the two coverslips.

11) Preheat a hot plate and a ~0.5 kg weight block before use to ~120-130 ºC.

12) Melt the Parafilm by laying the assembled chamber on the hot plate between two Kimwipes with the block on top of it. This step is to prevent melted Parafilm from sticking to the hot plate. Wait ~3-4 minutes for the Parafilm to melt. It is important that the block applies a uniform pressure over the chamber and this is done by centering it perfectly on top of the Parafilm sandwich.

13) Take the chamber off the hot plate and allow it to cool for ~2 minutes.

14) Mount the chamber onto its bracket, lining up the chamber holes with the bracket holes (see Figure 51). Gently screw the mounts into place to secure the chamber position.

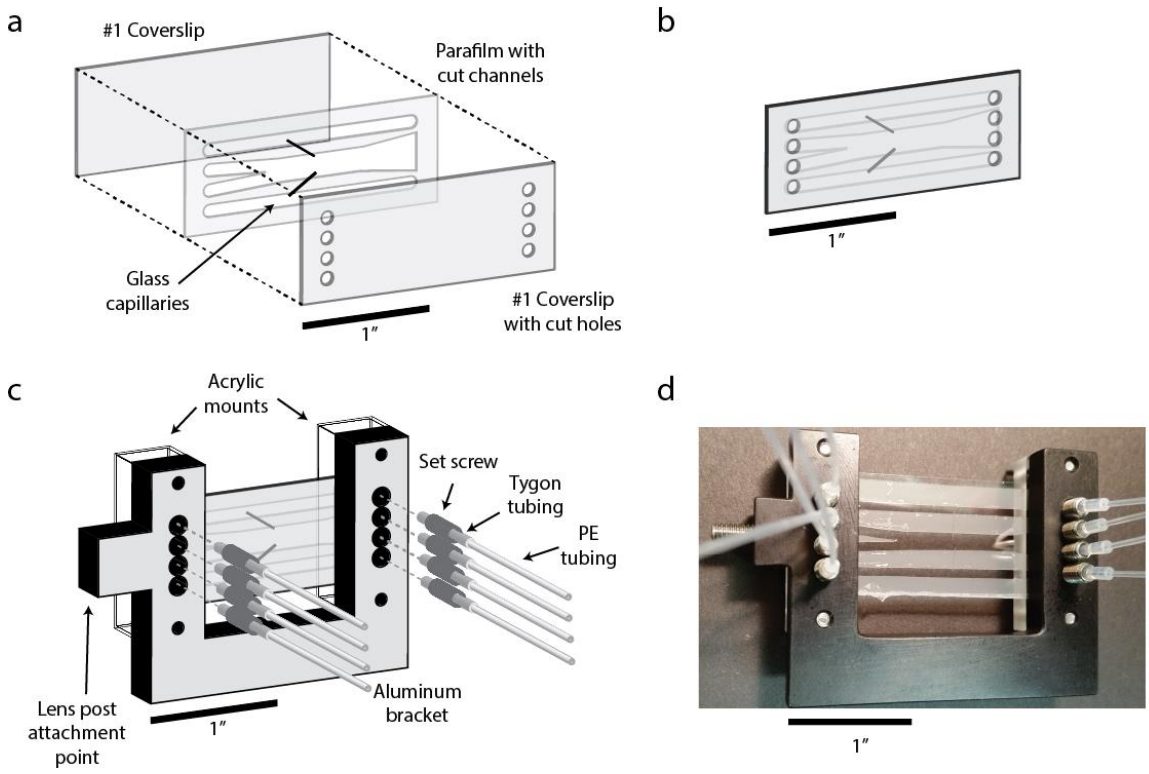

Figure 51: Sample flow chamber assembly as shown in [89] with four-hole chamber mounted.

A.6. Image buffer protocol

1) Add 7.9 mL of MQ water to a 25-mL centrifuge tube.

2) Add 1000 µL of Tris 1M; pH: 8.0 (100 mM concentration) to the tube

3) Add 500 µL of NaCl, 1M

4) Add 500 µL of Glucose (20%) to the tube and vortex the tube for 10 seconds.

5) filter the solution into a fresh tube and store the buffer in 4ºC for up to months.

A.7. Data Collection Protocol

Data collection follows a series of steps as follows:

A.7.1. Chamber setup:

1) Chamber is mounted tightly against both objectives. Mounting is done so that x-motor has a good movement range. Few drops of water cover between each objective and the cover glass.

Note: For running each stream in the chamber, we need a separate glass syringe of prepared sample for each sample to be connected to the entrance of each tube. Those needles connect to the chamber holes. Each syringe is rinsed and dried completely before loading. This prevents any contamination during the experiment.

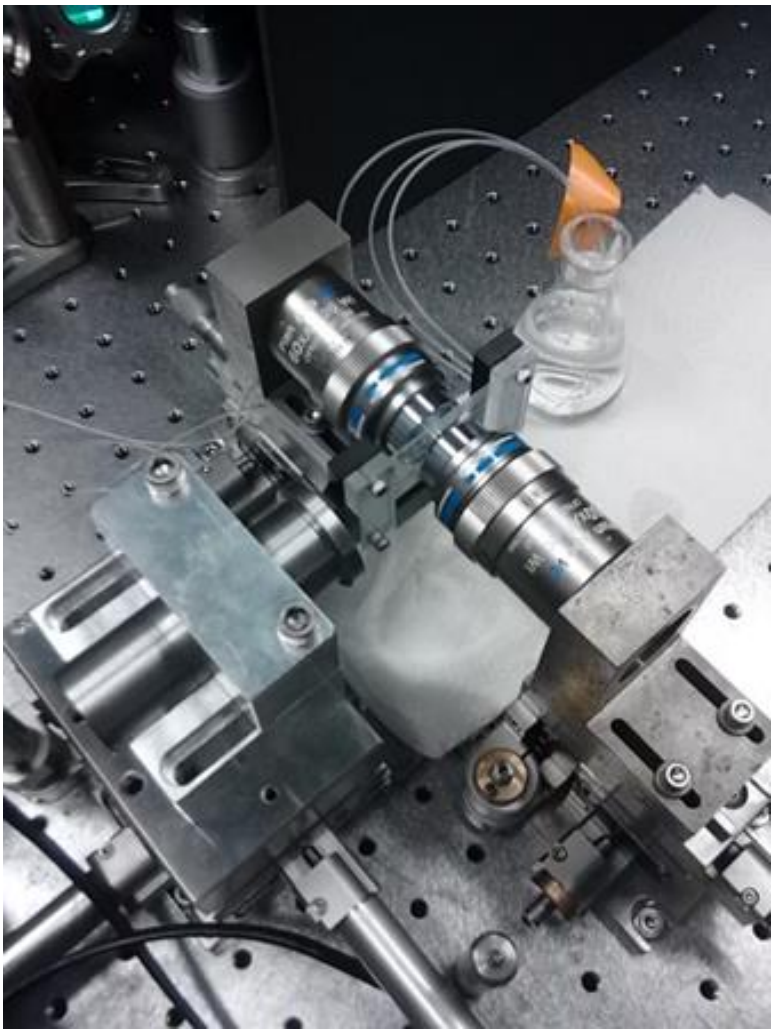

Figure 52: Chamber mounted tightly against both objectives.

2) Beam is collimated using an IR card. To make sure the beam is well collimated, go to "Display all" tab on LabView and make sure that "Trap Sum" is greater than 4.5 Volts.

3) Flip filters to visualize quality of focal spot on CCD camera. This aligns chamber angle in y-direction.

4) Check capillaries for bubbles and flush bubbles out if needed.

5) Motor positions set: using motor control, reset lower capillary to (x,y,z)=(0,0,0). Then use “Sample Stage” tab on LabView to set upper capillary and lower flow upset.

Note: After attaching new syringe of beads or image buffer, always flush 100 µL of it into the chamber before trapping (using “Syringe Pump” tab).

6) Watch the capillaries on the LabView screen to monitor when beads start coming out.

7) Check calibration with Adig beads by hitting “Save Calibration” on main tab. The calibration curve should fit model and have stiffness of k>0.2 pN/nm.

8) Readjust y-angle of the chamber slightly if needed by looking at the interference pattern of the beam on the monitor inside the trap room.

A.7.2. LabView setting for trapping:

1) Set “Averaging Factor” to one for taking data at full bandwidth of 66.66 KHz. Normally people use averaging factor of 125 in the lab, which corresponds to 533.33 KHz

2) In “Main Tab” set trap no. 2 to 90.0 MHz

3) In “Fishing Tab” turn off “Fish again after tether break” button since it will cause problems when tether force goes too low and LabView thinks tether is broken.

4) Set “Closest Delta” to 10 MHz. This value depends on the how low of the force we would like to reach when relaxing the tether.

5) Set “Delta 1” to ~12.4 MHz. Set “Delta 2” to the same.

6) MATLAB: open “plot_trap_raster.m” and have construct set to PG (protein G) for correct model.

A.7.3. Tethering Protocol

1) Hit “Move Cap2” button on “Actions” Tab

a. Script automatically:

i. Moves to upper capillary

ii. Turns top bead pump on

iii. Turns Trap 2 on

b. Trap ADig Bead

2) Hit “Got ADig” button on “Actions” Tab

a. Script automatically:

i. Turns top bead pump off

ii. Moves to lower capillary,

iii. Turns bottom bead pump on

iv. Turns Trap 1 on

b. Trap DNA Bead – Make sure traps are separated by ~17 MHz or more before trapping 2nd bead, or it may fall into the wrong trap

c. Note bottom bead pump must be turned off manually on LabVIEW

3) Hit “Cal/Data “ button on “Actions” Tab

a. Script automatically:

i. Moves to defined lower flow spot

ii. Saves calibration (write down file number)

Note: calibration figure should pop up, if not see earlier explanation of how to fix this

iii. Takes offset scan

(write down file number, number should be = Cal File + 1)

iv. Reads the calibration file into LabVIEW

v. Starts fishing script (start with a closest delta of ~10 MHz)

4) Fishing for DNA tether – fishing script can be turned on and off manually at any time via “Fish for Tether” button on Main Tab

a. Start with a closest delta of ~10 MHz, decrease every 10 - 15 cycles or so until tether is detected

b. When tether is detected, fishing script automatically:

i. Moves to delta 1

ii. Takes Raster Scan Force Extension (write down file number)

iii. Moves to delta 2

5) Taking Raster Scan/Force Extensions (write down file number)

a. Automated Raster Scan runs after fishing script detects tether (above) – starts scan from set “delta 1” and moves a fixed distance.

b. Manual Raster Scans can be taken anytime via “Trap Raster” button on Main Tab – starts scan from current trap delta and moves a fixed distance.

c. Can change Raster Scan settings on Trap Control Tab > AOD Raster Scan Tab. Select correct Raster Scan type (Should be #1 hairpin tether check). Hit “load” after changing any settings.

Note: For single molecule trapping experiments, either it is approach one or two, when fishing is done and an incipient tether is trapped, the first step, as part of the calibration process, is to get a force extension curve. To do that, we pull on the tether to a high force (typically 25-30 pN) and relax it to low force (typically 1-3 pN) and make sure the tethered construct follows the DNA model (WLC model). If this is not true, we know that we should fish for a new tether until our tether follows the model completely. This way we know that our tether does not have any abnormal activity due to a junk or bad DNA.

A.7.4. Taking Time Traces of Tethers at Constant Force

1) Increase Trap Delta on Main Tab until desired mean force is achieved (mean force is displayed on Main Tab left of force plot)

2) On Trap Control Tab > Feedback on Trap Position Tab, hit the following buttons (in order)

a. Set Point

b. Force Feedback On.

c. Remove Action Offset (action gauge should be roughly centered at 0, if not may need to hit button multiple times until gauge is centered)

3) On Main Tab, hit "Start Saving" button (write down file number)

4) Switch back to Feedback Tab to visualize Position vs. Time plot in real time.

• Can hit "Remove Offset" button left of plot to re-center trace in plot window if needed

• Can change y-axis limits by clicking on upper and lower limits on plot axis if needed

5) Hit "Stop Saving" button after tether breaks or done taking data.

6) Make sure Force Feedback is turned off before getting new tether or taking new raster scan.

Note: "Action Scaling" on Trap Control Tab > Feedback on Trap Position Tab should be set to 23, but due to a bug in LabVIEW, its value must be increased to 24 and then decreased back to 23 before running Force Feedback for the first time after LabVIEW is started. Feedback may not work right otherwise.

REFERENCES

REFERENCES

1. Brändén, C.I. and J. Tooze, *Introduction to Protein Structure*. 1999: Garland Pub.

2. Yee, A.A., et al., *NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins.* J Am Chem Soc, 2005. **127**(47): p. 16512-7.

3. Snyder, D.A., et al., *Comparisons of NMR Spectral Quality and Success in Crystallization Demonstrate that NMR and X-ray Crystallography Are Complementary Methods for Small Protein Structure Determination.* Journal of the American Chemical Society, 2005. **127**(47): p. 16505-16511.

4. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

5. Mirsky, A.E. and M.L. Anson, *Protein Coagulation and Its Reversal : The Reversal of the Coagulation of Hemoglobin.* J Gen Physiol, 1929. **13**(2): p. 133-43.

6. Anfinsen, C.B., et al., *The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain.* Proc Natl Acad Sci U S A, 1961. **47**: p. 1309-14.

7. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(4096): p. 223-30.

8. Dill, K.A., et al., *The protein folding problem: when will it be solved?* Curr Opin Struct Biol, 2007. **17**(3): p. 342-6.

9. Bai, Y., et al., *Protein folding intermediates: native-state hydrogen exchange.* Science, 1995. **269**(5221): p. 192-7.

10. Onuchic, J.N., Z. Luthey-Schulten, and P.G. Wolynes, *Theory of protein folding: the energy landscape perspective.* Annu Rev Phys Chem, 1997. **48**: p. 545-600.

11. Dill, K.A., *Polymer principles and protein folding.* Protein Sci, 1999. **8**(6): p. 1166-80.

12. Bicout, D.J. and A. Szabo, *Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model.* Protein Sci, 2000. **9**(3): p. 452-65.

13. Dinner, A.R., et al., *Understanding protein folding via free-energy surfaces from theory and experiment.* Trends Biochem Sci, 2000. **25**(7): p. 331-9.

14. Onuchic, J.N. and P.G. Wolynes, *Theory of protein folding.* Curr Opin Struct Biol, 2004. **14**(1): p. 70-5.

15. Oliveberg, M. and P.G. Wolynes, *The experimental survey of protein-folding energy landscapes.* Q Rev Biophys, 2005. **38**(3): p. 245-88.

16. Anfinsen, C.B., *The formation of the tertiary structure of proteins.* Harvey Lect, 1967. **61**: p. 95-116.

17. Levinthal, C., *Are There Pathways for Protein Folding.* Journal De Chimie Physique Et De Physico-Chimie Biologique, 1968. **65**(1): p. 44-+.

18. Weaver, R.F., *Molecular Biology*. 5th edition ed. 2012, United States: McGraw-Hill Education - Europe.

19. Tyedmers, J., A. Mogk, and B. Bukau, *Cellular strategies for controlling protein aggregation.* Nat Rev Mol Cell Biol, 2010. **11**(11): p. 777-788.

20. Stadtman, E.R. and R.L. Levine, *Free radical-mediated oxidation of free amino acids and amino acid residues in proteins.* Amino Acids, 2003. **25**(3): p. 207-218.

21. Wolynes, P.G., J.N. Onuchic, and D. Thirumalai, *Navigating the folding routes.* Science, 1995. **267**(5204): p. 1619-20.

22. Bryngelson, J.D., et al., *Funnels, pathways, and the energy landscape of protein folding: A synthesis.* Proteins: Structure, Function, and Bioinformatics, 1995. **21**(3): p. 167-195.

23. Bryngelson, J.D. and P.G. Wolynes, *Intermediates and barrier crossing in a random energy model (with applications to protein folding).* The Journal of Physical Chemistry, 1989. **93**(19): p. 6902-6915.

24. Karplus, M., *Behind the folding funnel diagram.* Nat Chem Biol, 2011. **7**(7): p. 401-404.

25. Levinthal, C., *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.*, in *University of Illinois Press*. 1969. p. 22-24.

26. Lapidus, L.J., et al., *Complex Pathways in Folding of Protein G Explored by Simulation and Experiment.* Biophysical Journal, 2014. **107**(4): p. 947-955.

27. Baiz, C.R., et al., *A molecular interpretation of 2D IR protein folding experiments with Markov state models.* Biophys J, 2014. **106**(6): p. 1359-70.

28. Voelz, V.A., et al., *Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment.* J Am Chem Soc, 2012. **134**(30): p. 12565-77.

29. Voelz, V.A., et al., *Unfolded-state dynamics and structure of protein L characterized by simulation and experiment.* J Am Chem Soc, 2010. **132**(13): p. 4702-9.

30. Hänggi, P., P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after Kramers.* Reviews of Modern Physics, 1990. **62**(2): p. 251-341.

31. Sadqi, M., L.J. Lapidus, and V. Munoz, *How fast is protein hydrophobic collapse?* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(21): p. 12117-12122.

32. Lapidus, L.J., et al., *Protein Hydrophobic Collapse and Early Folding Steps Observed in a Microfluidic Mixer.* Biophysical Journal, 2007. **93**(1): p. 218-224.

33. Waldauer, S.A., et al., *Ruggedness in the folding landscape of protein L.* Hfsp Journal, 2008. **2**(6): p. 388-395.

34. Wu, L. and L.J. Lapidus, *Combining Ultrarapid Mixing with Photochemical Oxidation to Probe Protein Folding.* Analytical Chemistry, 2013. **85**(10): p. 4920-4924.

35. Zhu, L., et al., *Sub-millisecond Chain Collapse of the Escherichia coli Globin ApoHmpH.* Journal of Physical Chemistry B, 2013. **117**(26): p. 7868-7877.

36. Prigozhin, M.B., et al., *Mapping fast protein folding with multiple-site fluorescent probes.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(26): p. 7966-7971.

37. Wirth, A.J., et al., *Comparing Fast Pressure Jump and Temperature Jump Protein Folding Experiments and Simulations.* J Am Chem Soc, 2015. **137**(22): p. 7152-9.

38. DeCamp, S.J., et al., *Direct Observation of Downhill Folding of lambda-Repressor in a Microfluidic Mixer.* Biophysical Journal, 2009. **97**(6): p. 1772-1777.

39. Zhu, L., et al., *Evidence of multiple folding pathways for the villin headpiece subdomain.* J Phys Chem B, 2011. **115**(43): p. 12632-7.

40. Narayanan, R., et al., *Exploring the energy landscape of nucleic acid hairpins using laser temperature-jump and microfluidic mixing.* J Am Chem Soc, 2012. **134**(46): p. 18952-63.

41. Waldauer, S.A., et al., *Microfluidic Mixers for Studying Protein Folding.* J Vis Exp, 2012(62): p. e3976.

42. Björck, L. and G. Kronvall, *Purification and some properties of streptococcal protein G, a novel IgG-binding reagent.* The Journal of Immunology, 1984. **133**(2): p. 969-974.

43. Gronenborn, A., et al., *A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G.* Science, 1991. **253**(5020): p. 657-661.

44. Krantz, B.A., et al., *Fast and Slow Intermediate Accumulation and the Initial Barrier Mechanism in Protein Folding.* Journal of Molecular Biology, 2002. **324**(2): p. 359-371.

45. Maxwell, K.L., et al., *Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins.* Protein Sci, 2005. **14**(3): p. 602-16.

46. Selenko, P., et al., *Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes.* Proceedings of the National Academy of Sciences, 2006. **103**(32): p. 11904-11909.

47. McCallister, E.L., E. Alm, and D. Baker, *Critical role of beta-hairpin formation in protein G folding.* Nat Struct Biol, 2000. **7**(8): p. 669-73.

48. Neuman, K.C. and S.M. Block, *Optical trapping.* The Review of scientific instruments, 2004. **75**(9): p. 2787-2809.

49. Park, S.-H., K.T. O'Neil, and H. Roder, *An Early Intermediate in the Folding Reaction of the B1 Domain of Protein G Contains a Native-like Core.* Biochemistry, 1997. **36**(47): p. 14277-14283.

50. Morrone, A., et al., *GB1 is not a two-state folder: identification and characterization of an on-pathway intermediate.* Biophys J, 2011. **101**(8): p. 2053-60.

51. Neupane, K., et al., *Direct observation of transition paths during the folding of proteins and nucleic acids.* Science, 2016. **352**(6282): p. 239-242.

52. Whitley, K.D., M.J. Comstock, and Y.R. Chemla, *High-Resolution "Fleezers": Dual-Trap Optical Tweezers Combined with Single-Molecule Fluorescence Detection*, in *Optical Tweezers: Methods and Protocols*, A. Gennerich, Editor. 2017, Springer New York: New York, NY. p. 183-256.

53. Rief, M., et al., *Reversible unfolding of individual titin immunoglobulin domains by AFM.* Science, 1997. **276**(5315): p. 1109-12.

54. Kellermayer, M.S., et al., *Folding-unfolding transitions in single titin molecules characterized with laser tweezers.* Science, 1997. **276**: p. 1112+.

55. Gelman, H. and M. Gruebele, *Fast protein folding kinetics.* Q Rev Biophys, 2014. **47**(2): p. 95-142.

56. Chung, H.S., et al., *Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times.* Science, 2012. **335**(6071): p. 981-984.

57. Chung, H.S. and W.A. Eaton, *Single-molecule fluorescence probes dynamics of barrier crossing.* Nature, 2013. **502**(7473): p. 685-688.

58. Chung, H.S., J.M. Louis, and W.A. Eaton, *Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories.* Proceedings of the National Academy of Sciences, 2009. **106**(29): p. 11837-11844.

59. Truex, K., et al., *Testing Landscape Theory for Biomolecular Processes with Single Molecule Fluorescence Spectroscopy.* Physical Review Letters, 2015. **115**(1): p. 018101.

60. Dudko, O.K., G. Hummer, and A. Szabo, *Theory, analysis, and interpretation of single-molecule force spectroscopy experiments.* Proc Natl Acad Sci U S A, 2008. **105**(41): p. 15755-60.

61. Rhoades, E., et al., *Two-state folding observed in individual protein molecules.* J Am Chem Soc, 2004. **126**(45): p. 14686-7.

62. Schonfelder, J., R. Perez-Jimenez, and V. Munoz, *A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution.* Nat Commun, 2016. **7**: p. 11777.

63. Zoldak, G., et al., *Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy.* Proc Natl Acad Sci U S A, 2013. **110**(45): p. 18156-61.

64. Yu, H., et al., *Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates.* Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14452-7.

65. Yu, H., et al., *Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins.* Science, 2017. **355**(6328): p. 945-950.

66. Graham, T.G. and R.B. Best, *Force-induced change in protein unfolding mechanism: discrete or continuous switch?* J Phys Chem B, 2011. **115**(6): p. 1546-61.

67. Guinn, E.J., B. Jagannathan, and S. Marqusee, *Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein.* Nat Commun, 2015. **6**: p. 6861.

68. Best, R.B., et al., *Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules.* J Phys Chem B, 2008. **112**(19): p. 5968-76.

69. Li, P.C. and D.E. Makarov, *Simulation of the mechanical unfolding of ubiquitin: probing different unfolding reaction coordinates by changing the pulling geometry.* J Chem Phys, 2004. **121**(10): p. 4826-32.

70. Zhuravlev, P.I., et al., *Force-dependent switch in protein unfolding pathways and transition-state movements.* Proc Natl Acad Sci U S A, 2016. **113**(6): p. E715-24.

71. Pierse, C.A. and O.K. Dudko, *Distinguishing Signatures of Multipathway Conformational Transitions.* Phys Rev Lett, 2017. **118**(8): p. 088101.

72. Jagannathan, B. and S. Marqusee, *Protein Folding and Unfolding Under Force.* Biopolymers, 2013. **99**(11): p. 860-869.

73. Izadi, D., T. Nguyen, and L. Lapidus, *Complete Procedure for Fabrication of a Fused Silica Ultrarapid Microfluidic Mixer Used in Biophysical Measurements.* Micromachines, 2017. **8**(1): p. 16.

74. Rami, B.R. and J.B. Udgaonkar, *pH-jump-induced folding and unfolding studies of barstar: evidence for multiple folding and unfolding pathways.* Biochemistry, 2001. **40**(50): p. 15267-79.

75. Dumont, C., T. Emilsson, and M. Gruebele, *Reaching the protein folding speed limit with large, sub-microsecond pressure jumps.* Nat Methods, 2009. **6**(7): p. 515-9.

76. Liu, F., et al., *A one-dimensional free energy surface does not account for two-probe folding kinetics of protein alpha(3)D.* J Chem Phys, 2009. **130**(6): p. 061101.

77. Ihalainen, J.A., et al., *Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5383-8.

78. Brody, J.P., et al., *Biotechnology at low Reynolds numbers.* Biophys J, 1996. **71**(6): p. 3430-41.

79. Knight, J.B., et al., *Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds.* Physical Review Letters, 1998. **80**(17): p. 3863-3866.

80. Hertzog, D.E., et al., *Optimization of a microfluidic mixer for studying protein folding kinetics.* Analytical Chemistry, 2006. **78**(13): p. 4299-4306.

81. Hertzog, D.E., et al., *Femtomole mixer for microsecond kinetic studies of protein folding.* Analytical Chemistry, 2004. **76**(24): p. 7169-7178.

82. Yao, S. and O. Bakajin, *Improvements in mixing time and mixing uniformity in devices designed for studies of protein folding kinetics.* Anal Chem, 2007. **79**(15): p. 5753-9.

83. Ivorra, B., et al., *Two- and three-dimensional modeling and optimization applied to the design of a fast hydrodynamic focusing microfluidic mixer for protein folding.* Physics of Fluids, 2013. **25**(3): p. 032001-17.

84. Wu, L. and L.J. Lapidus, *Combining ultrarapid mixing with photochemical oxidation to probe protein folding.* Anal Chem, 2013. **85**(10): p. 4920-4.

85. Jiang, L., et al., *Microsecond protein folding events revealed by time-resolved fluorescence resonance energy transfer in a microfluidic mixer.* Anal Chem, 2015. **87**(11): p. 5589-95.

86. Ivorra, B., et al., *Semi-deterministic and genetic algorithms for global optimization of microfluidic protein-folding devices.* International Journal for Numerical Methods in Engineering, 2006. **66**(2): p. 319-333.

87. Yao, S. and O. Bakajin, *Improvements in Mixing Time and Mixing Uniformity in Devices Designed for Studies of Protein Folding Kinetics.* Analytical Chemistry, 2007. **79**: p. 5753-5759.

88. Hertzog, D.E., et al., *Femtomole mixer for microsecond kinetic studies of protein folding.* Anal Chem, 2004. **76**(24): p. 7169-78.

89. Waldauer, S.A., *Early events in protein folding investigated through ultrarapid microfluidic mixing.* 2010.

90. Waldauer, S.A., O. Bakajin, and L.J. Lapidus, *Extremely slow intramolecular diffusion in unfolded protein L.* Proc Natl Acad Sci U S A, 2010. **107**(31): p. 13713-7.

91. Tathagata, R., Z. Haixin, and R.M. Deirdre, *Deep reactive ion etching of fused silica using a single-coated soft mask layer for bio-analytical applications.* Journal of Micromechanics and Microengineering, 2010. **20**(9): p. 097002.

92. Ahamed, M.J., et al., *Study of High Aspect Ratio NLD Plasma Etching and Postprocessing of Fused Silica and Borosilicate Glass.* Journal of Microelectromechanical Systems, 2015. **24**(4): p. 790-800.

93. Jager, M., et al., *The folding mechanism of a beta-sheet: The WW domain.* Journal of Molecular Biology, 2001. **311**(2): p. 373-393.

94. Cecconi, C., et al., *Direct observation of the three-state folding of a single protein molecule.* Science, 2005. **309**(5743): p. 2057-60.

95. Hong, V., et al., *Analysis and Optimization of Copper-Catalyzed Azide–Alkyne Cycloaddition for Bioconjugation.* Angewandte Chemie International Edition, 2009. **48**(52): p. 9879-9883.

96. Presolski, S.I., V.P. Hong, and M.G. Finn, *Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation.* Curr Protoc Chem Biol, 2011. **3**(4): p. 153-162.

97. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angewandte Chemie International Edition, 2001. **40**(11): p. 2004-2021.

98. Greenleaf, W.J., M.T. Woodside, and S.M. Block, *High-resolution, single-molecule measurements of biomolecular motion.* Annu Rev Biophys Biomol Struct, 2007. **36**: p. 171-90.

99. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy.* Nat Methods, 2008. **5**(6): p. 491-505.

100. Whitley, K.D., M.J. Comstock, and Y.R. Chemla, *High-Resolution Optical Tweezers Combined With Single-Molecule Confocal Microscopy.* Methods Enzymol, 2017. **582**: p. 137-169.

101. Gosse, C. and V. Croquette, *Magnetic tweezers: micromanipulation and force measurement at the molecular level.* Biophys J, 2002. **82**(6): p. 3314-29.

102. Molloy, J.E., et al., *Single-molecule mechanics of heavy meromyosin and S1 interacting with rabbit or Drosophila actins using optical tweezers.* Biophys J, 1995. **68**(4 Suppl): p. 298S-303S; 303S-305S.

103. Gittes, F. and C.F. Schmidt, *Interference model for back-focal-plane displacement detection in optical tweezers.* Opt Lett, 1998. **23**(1): p. 7-9.

104. Pralle, A., et al., *Three-dimensional high-resolution particle tracking for optical tweezers by forward scattered light.* Microsc Res Tech, 1999. **44**(5): p. 378-86.

105. Whitley, K., *Single-molecule studies of mechanical and helicase-catalyzed disruption of nucleic acid duplexes.* 2017.

106. Comstock, M.J., T. Ha, and Y.R. Chemla, *Ultrahigh-resolution optical trap with single-fluorophore sensitivity.* Nat Meth, 2011. **8**(4): p. 335-340.

107. Fischer, M., et al., *Active-passive calibration of optical tweezers in viscoelastic media.* Rev Sci Instrum, 2010. **81**(1): p. 015103.

108. Swoboda, M., et al., *Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments.* ACS Nano, 2012. **6**(7): p. 6364-9.

109. Landry, M.P., et al., *Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments.* Biophys J, 2009. **97**(8): p. 2128-36.

110. Ha, T., *Single-molecule fluorescence resonance energy transfer.* Methods, 2001. **25**(1): p. 78-86.

111. Bustamante, C., et al., *Entropic elasticity of lambda-phage DNA.* Science, 1994. **265**(5178): p. 1599-600.

112. Smith, S.B., Y. Cui, and C. Bustamante, *Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules.* Science, 1996. **271**(5250): p. 795-9.

113. Cluzel, P., et al., *DNA: an extensible molecule.* Science, 1996. **271**(5250): p. 792-4.

114. Kratky, O. and G. Porod, *Diffuse small-angle scattering of X-rays in colloid systems.* J Colloid Sci, 1949. **4**(1): p. 35-70.

115. Lapidus, L.J., et al., *Effects of Chain Stiffness on the Dynamics of Loop Formation in Polypeptides. Appendix: Testing a 1-Dimensional Diffusion Model for Peptide Dynamics.* The Journal of Physical Chemistry B, 2002. **106**(44): p. 11628-11640.

116. Marko, J.F. and E.D. Siggia, *Statistical mechanics of supercoiled DNA.* Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 1995. **52**(3): p. 2912-2938.

117. Wang, M.D., et al., *Stretching DNA with optical tweezers.* Biophys J, 1997. **72**(3): p. 1335-46.

118. Smith, S.B., L. Finzi, and C. Bustamante, *Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads.* Science, 1992. **258**(5085): p. 1122-6.

119. Dudko, O.K., T.G. Graham, and R.B. Best, *Locating the barrier for folding of single molecules under an external force.* Phys Rev Lett, 2011. **107**(20): p. 208301.

120. Suzuki, Y. and O.K. Dudko, *Single-molecule rupture dynamics on multidimensional landscapes.* Phys Rev Lett, 2010. **104**(4): p. 048101.

121. Yu, H., et al., *Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape.* Proc Natl Acad Sci U S A, 2015. **112**(27): p. 8308-13.

122. Brujić, J., et al., *Dwell-Time Distribution Analysis of Polyprotein Unfolding Using Force-Clamp Spectroscopy.* Biophys J, 2007. **92**(8): p. 2896-903.

123. Fernandez, J.M. and H. Li, *Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein.* Science, 2004. **303**(5664): p. 1674-1678.

124. Jarzynski, C., *Nonequilibrium Equality for Free Energy Differences.* Physical Review Letters, 1997. **78**(14): p. 2690-2693.