STUDIES ON THE MECHANICAL STABILITY OF A PROTEIN BY SINGLE-MOLECULE ATOMIC FORCE MICROSCOPY

by

HUI-CHUAN WANG

M.Sc., University College London, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(CHEMISTRY)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May, 2009

© Hui-Chuan Wang, 2009
ABSTRACT

Elastomeric proteins are subject to mechanical tensions under biological settings and possess mechanical properties that underlie the elasticity of natural adhesive, cell adhesion and muscle proteins. Single molecule atomic force microscopy has made it possible to directly probe the mechanical properties of elastomeric proteins and provides insights into the molecular design of elastomeric proteins. Combining the single molecule atomic force microscopy and protein engineering techniques allows us study the mechanical stability of proteins and develop methods to tune the mechanical stability. Mechanical tensions are also found in some nonmechanical proteins. Based on the results from single-molecule atomic force microscopy, nonmechanical protein of GB1 shows high mechanical stability that is comparable or superior to those of known elastomeric proteins. Here, we use a small protein, GB1, the B1 IgG binding domain of protein G from Streptococcus, as a model system to directly investigate the mechanical properties of GB1 mutants and loop mutants by using single-molecule AFM. Point mutations in proteins may disrupt the intermolecular interactions and affect the chemical and mechanical stability of the protein. Φ-value analysis together with single-molecule atomic force microscopy is used to probe the mechanical stability of the protein and gives a complete picture on how proteins are structured in the transition state during folding/unfolding event. Results from chapter 2 indicate that GF30L and GT53A mutation decrease the mechanical stability as well as accelerate the unfolding kinetics of GB1. This is due to the disruption of the hydrogen bond networking between the terminal β-strands or unraveled the hydrophobic interactions and side chain interactions, resulting in lower unfolding forces with Φ-values closer to one. Configurational entropy plays crucial roles in defining the thermodynamic stability as well as the folding/unfolding kinetics of proteins. Here, we directly probe the role of configurational entropy in the
mechanical unfolding kinetics and mechanical stability of proteins by using single molecule atomic force microscopy and protein engineering methods. Chapter 3 demonstrates that the mechanical stability of GB1 decreases as the number of inserted amino acid residues into loop 2 of GB1 increases. This result can be explained by the loss of configurational entropy upon closing an unstructured flexible loop using classical polymer theory, highlighting the important role of loop regions in the mechanical unfolding of proteins. The findings from these experiments are of critical importance towards engineering artificial elastomeric proteins with tailored nanomechanical properties.
# TABLE OF CONTENTS

Abstract.................................................................................................................. ii
Table of Contents .................................................................................................. iv
List of Tables ........................................................................................................ vii
List of Figures ....................................................................................................... viii
Acknowledgements .............................................................................................. x
Dedication ................................................................................................................. xi
Co-authorship statement ....................................................................................... xii

Chapter 1: Mechanical unfolding of proteins by single-molecule atomic force microscopy (AFM)................................................................. 1
1.1 Introduction ..................................................................................................... 1
1.2 AFM on single-molecule protein study .......................................................... 3
1.2.1 The mechanical unfolding of proteins by single molecule AFM ............... 4
1.2.2 The mechanical unfolding of a domain can be modelled as a two-state process.... 9
1.3 Protein folds have different mechanical stability ............................................ 11
1.3.1 Titin immunoglobulin domains possess high mechanical stability .............. 13
1.3.2 The I27 module............................................................................................ 15
1.3.3 Length phenotype is sensitive to the precise location of the mutation with respect to the mechanical topology of single proteins .............................. 18
1.3.4 Single protein misfolding events captured by AFM ..................................... 19
1.3.5 Nonmechanical protein can have significant mechanical stability ............ 21
1.4 Thesis overview ............................................................................................... 23
1.5 References ........................................................................................................ 24

Chapter 2: Point mutation alters the mechanical stability of protein GB1........... 28
2.1 Introduction ..................................................................................................... 28
2.2 Methods and materials ................................................................................... 31
2.2.1 Protein engineering: polyprotein construction .......................................... 31
2.2.2 Single-molecule AFM experiment ............................................................. 32
2.2.3 Monte Carlo simulations ........................................................................... 33
2.3 Results .............................................................................................................. 33
2.3.1 Engineering heteropolyproteins and homopolyproteins of GB1 mutants ...... 33
2.3.2 AFM unfolds the proteins and results in a characteristic saw-tooth pattern .... 33
2.3.3 Mutation, F30L and T53A result in the decrease of the mechanical stability of GB1 ........................................................................................................... 36
2.3.4 All GB1 mutants follow the identical unfolding pathway ............................ 38
2.3.5 Refolding depends exponentially on the amount of time that the domain remains
relaxed ............................................. 40
2.4 Discussion .................................. 42
2.4.1 Calculation of the Φ-value...... 42
2.4.2 Mutations F30L and T53A of GB1 destabilize the transition state of unfolding 44
pathway from mechanical studies..... 44
2.4.3 Mechanical and chemical Φ-values compared........ 46
2.5 Conclusions................................ 48
2.6 References................................ 50

Chapter 3: Configurational entropy modulates the mechanical stability of GB1 ........ 53
3.1 Introduction......................... 53
3.2 Methods and materials........... 55
3.2.1 Protein engineering.............. 55
3.2.1.1 Construction of GB1-L2 gene.......... 56
3.2.1.2 Construction of GB1-L5 gene.......... 56
3.2.1.3 Construction of GB1-L24 and GB1-L46 gene........ 57
3.2.1.4 Polyprotein construction........... 57
3.2.2 Single-molecule AFM experiment ......... 58
3.2.3 Monte Carlo simulations..... 58
3.2.4 Circular dichroism measurements....... 59
3.3 Results................................ 59
3.3.1 The second loop has a high tolerance of loop elongation without affecting GB1’s 59
native structure............................... 59
3.3.2 Loop elongation decreases the mechanical stability of GB1 and accelerates the 61
mechanical unfolding kinetics .......... 61
3.4 Discussion............................. 70
3.4.1 Configurational entropy plays important roles in mechanical unfolding kinetics 70
of GB1 ............................................. 70
3.4.2 Loop 2 of GB1 is partially disrupted at the mechanical transition state.......... 73
3.5 References................................ 75

Chapter 4: Summary and conclusions........ 78
4.1 Summary.............................. 78
4.2 Future plan................................ 80
4.3 References................................ 83

Appendices........................................ 84
Appendix A.................................... 84
LIST OF TABLES

Table 2.1
Comparison of mechanical and chemical unfolding data of GB1 mutants .......... 46

Table 3.1
Mechanical experimental data for GB1 loop mutants ..................................... 70
LIST OF FIGURES

Figure 1.1
A schematic diagram of the custom-built AFM ....................................................4

Figure 1.2
Using single-molecule AFM to measure the elasticity of single proteins ............6

Figure 1.3
A schematic diagram of two-state free energy diagram for the mechanical unfolding of proteins.......................... .................. ...........................................................10

Figure 1.4
Three-dimensional structure of representative proteins that have been studied by using single-molecule AFM ................ ...........................................................13

Figure 1.5
Force-extension curves of titin proteins show periodic saw-tooth patterns that are consistent with their modular construction......................................................14

Figure 1.6
Mechanical properties of single human cardiac titin immunoglobulin domains ....16

Figure 1.7
Point mutations in an I27 domain alter its mechanical stability............................17

Figure 1.8
Single-molecule AFM measurements of the contour length of engineered polyproteins.......................... .................. ...........................................................19

Figure 1.9
Misfolding of I27 domains... ...........................................................20

Figure 1.10
The single molecule AFM results reveal that the GB1 protein is mechanically stable...........................................................22

Figure 2.1
Backbone ribbon diagram indicating structure of GB1 and 10 point mutations made on GB1...........................................................31

Figure 2.2
The force-extension relationships of the polyprotein chimeras constructed as (mGB1-I27)$_4$ and (mGB1)$_8$... ...........................................................36

Figure 2.3
Histograms for the mechanical unfolding force of GB1 mutants and wt-GB1......38

Figure 2.4
Stretching a single polyprotein at different pulling speeds.................................39

Figure 2.5
The folding kinetics of GB1 mutants...........................................................41
Figure 2.6
Energy diagram displaying the relationship between the native and unfolded state correlated with transition states

Figure 3.1
Loop insertions in GB1

Figure 3.2
Far-UV CD spectra of the loop insertion mutants indicate that loop 2 is tolerant of loop insertion and GB1 mutants are properly folded

Figure 3.3
Force-extension relationships of GB1 loop mutants

Figure 3.4
(A) Histograms for ΔLc of loop insertion mutants; (B) Histograms for the mechanical unfolding force of loop insertion mutants

Figure 3.5
Mechanical stability shows nonlinear dependence on the number of amino acid residues inserted in loop 2

Figure 3.6
Speed dependence of the mechanical unfolding forces of wt GB1 and GB1 loop mutants

Figure 3.7
The free energy barrier for mechanical unfolding shows a decreased nonlinear dependence with the length of the flexible linker

Figure B1
A schematic diagram to show polyprotein construction procedures

Figure B2
Directional DNA multiple steps cloning strategy is done by ligation of the sticky ends of the palindromic BamHI and BglII restriction site of the fragments

Figure B3
Map of the custom-built pQE80L expression vector

Figure C1
AFM cantilever calibration
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Hongbin Li, who not only took his time to read and comment on the thesis but also provided invaluable assistance on my research projects throughout the time I worked in the lab. Millions of thanks and appreciations!!

I am grateful to the lab members from Li’s group for their comments and suggestions as well as their help and support. Among them, I especially wish to thank Dr. Yi Cao, who taught me all the useful laboratory techniques when I started in the lab. I also really appreciate all the help from the members in Bioservice centre.

It is also a pleasure to thank all the funding to support our research projects and make our ideas come true.

Finally, I would like to thank UBC for providing such a wondering research environment.
DEDICATION

This is dedicated to my family, my father, my mother, and my brother, for their constant supports and encouragements.
CO-AUTHORSHIP STATEMENT

The idea of Chapter 2 was identified and designed by Dr. Hongbin Li, Dr. Yi Cao and myself. I have performed the research and data analysis on wild-type GB1 GL7A, GT11A, GT16A, GT25A, GF30L, GD46 and GT53 mutants. Dr. Yi Cao performed the research and data analysis on GK28A, GK31A and GN35A. Manuscript preparation was done by Hui-Chuan Wang.

The idea of Chapter 3 was designed by Dr. Hongbin Li. Dr. Deepak Sharma and Dr. Meijia Wang performed the research and data analysis on GB1-L2 and GB1-L5. I carried out the research and analyzed the data on GB1-L24 and GB1-L46. Dr. Yi Cao performed the pulling speed experiment for all the GB1 loop mutants. The manuscript preparation was written by Dr. Hongbin Li.
Chapter 1: Mechanical unfolding of proteins by single-molecule atomic force microscopy (AFM)

1.1 Introduction

Many proteins are exposed to mechanical tensions. Elastomeric proteins are an important class of such proteins that are subject to stretching forces. They act as molecular springs and are in charge of multiple biological processes responsible for a wide variety of mechanical functions. Examples of proteins under mechanical tensions include the muscle protein titin, and the extracellular matrix proteins (ECM) fibronectin and tenascin.

All proteins have their unique functions owing to the specific folding of their polypeptide chains. The stability of the folded proteins is typically investigated by thermal or chemical denaturation. These classical biochemical techniques work with ensembles of molecules and provide average parameters out of the Avogadro numbers of molecules. The advent of single molecule techniques, however, brings forth an understanding to the physical properties, and its underlying the mechanical functions of individual proteins at the single-molecule level. These techniques eliminate the ‘population noise’ and enable actual distribution of physical properties of the individual protein. Some of the instruments used to probe the intramolecular and intermolecular bonding are based on either soft springs or stiff springs and include optical tweezers, micro needles, biomembrane force probes and atomic force microscopy (AFM).

Among all, single-molecule AFM is especially suitable to measure the mechanical properties of proteins due to its excellent spatial resolution (~Å), force sensitivity (~10 pN), force range (from ~10 pN to nN), and obviates specific chemical immobilization. The combination of single-molecule AFM together with protein engineering methods has been used extensively to analyze and examine the
mechanical properties of proteins and their individual domains, in terms of the molecular elasticity and binding. The mechanical properties of elastomeric proteins, such as titin\(^7-11\), tenascin\(^16,17\) and fibronectin\(^12-15\), have been studied entirely by using a single-molecule AFM. Protein engineering techniques are employed to construct polyproteins made from identical tandem repeats and assess the mechanical stability of a specific domain from the protein of interest\(^40\). For instance, a direct tandem repeats of Ig module 27 from the I band of human cardiac titin (I27)\(^41\) reveals a mechanical unfolding force of \(~200\) pN and a contour length increment (\(\Delta L_c\)) of \(~28\) nm, indicating a high mechanical stability that will be discussed later. Mechanical proteins are just a small fraction of the total proteins in cells. Many other proteins, called non-mechanical proteins, are not subject to stretching forces under their physiological conditions. Inspired by the naturally evolved elastomeric proteins, artificial non-mechanical proteins are constructed and also show mechanical properties that are comparable to those of naturally evolved elastomeric proteins.

AFM studies convince us that high mechanical stability is a property not only possessed by mechanical proteins but also owned by some non-mechanical proteins. Some examples include, barnase\(^42\) and calmodulin\(^40\), that are mechanically labile and show marginal mechanical stability, unlike green fluorescent protein (GFP)\(^43-44\), which demonstrates a higher mechanical stability (\(~100\) pN). Others like B1 IgG binding domain of protein L\(^63\), B1 IgG binding domain of protein G\(^32-33\) and Top 7\(^45\) also exhibit significant mechanical stability.

In order to understand why these non-mechanical proteins possess high mechanical stability, the structure of mechanical proteins is determined and used for study by Steered Molecular Dynamic simulation (SMD)\(^46-49\). Mechanical stability is determined by the same set of non-covalent interactions-hydrogen bonds, hydrophobic interactions, and disulfide bonds- that determine the overall three-
dimensional structures and thermodynamic stability of proteins. Shear topology of the
two terminal force-bearing β strands also appears to be a common feature of
mechanical stable proteins. The protein topology, therefore, plays a critical role in
determining the mechanical stability of proteins. If non-mechanical proteins are
structurally similar to mechanical proteins, it is likely that non-mechanical proteins
will exhibit significant mechanical stability\textsuperscript{49,50}.

This chapter will outline the investigation on the mechanical stability of
proteins, as well as give insights into the main concepts and ideas of AFM.

1.2 Atomic force microscopy (AFM) on single-molecule protein study

AFM was developed by Gerb Binning, Calvin Quate and Christoph Gerber in
1986\textsuperscript{64} and has since been widely used to scan sample surfaces and examine a variety
of protein interactions with atomic resolution. AFM is a mechanical force transducer,
consisting of a flexible cantilever ended with a sharp tip, photodiodes, a laser diode
and a piezoelectric positioner. AFM contains a Digital Instruments AFM detector head
mounted on the top of the piezoelectric positioner (Figure 1.1). The nanomechanical
properties of materials and the binding forces of receptor-ligand systems and
complementary DNA strands, are used to study the mechanical unfolding forces of
protein domains\textsuperscript{6-11,13,16,28} at the single molecule level. The combination of
single-molecule AFM together with protein engineering methods makes it possible to
analyze and to examine the mechanical properties of proteins in depth. These
gineered polyproteins are then measured by AFM using force-extension mode, in
which the protein is stretched and unravelled by AFM.
**Figure 1.1** A schematic diagram of the custom-built AFM\(^6,4^0\). The AFM instrument consists of a laser diode, a photodiode, a cantilever and a piezoelectric positioner.

### 1.2.1 The mechanical unfolding of proteins by single molecule AFM

A schematic diagram of a force-measuring mode measurement on a polyprotein is shown \(^6,4^0\) in Figure 1.2. The two ends of the proteins are pulled apart at a constant velocity by moving the piezoelectric positioner away. All of the experimental measurements for polyprotein are carried out in PBS (phosphate buffered saline) buffer. A small amount of polyprotein sample is deposited onto a clean glass cover slip covered by \(~50\) µL PBS buffer and is allowed to adsorb for approximately 10 min. The spring constant of each individual cantilever is calibrated before each experiment. The single-molecule AFM pulling experiments are then performed. The AFM tip and substrate are brought together and pressed against each other, and then withdrawn. Some of the molecules are adsorbed at the tip. As the distance of the two surfaces increases, non-specific interactions between the two layers are broken leaving only a single molecule protein between the tip and the substrate. The first resistance to extension of the protein is entropic forces. Further increase in the distance between
the two surfaces extends the protein and generates an entropic restoring force that bends the cantilever giving a cantilever deflection $\Delta z_c$. This is detected and measured by the photodiode (See Figure 1.2(2)). The extension of the protein can be calculated from the piezoelectric positioner movement in the $z$-axis ($\Delta z_p$). Figure 1.2 indicates that the protein extension ($L_1-L_0$) pulled by the piezoelectric positioner can be measured by subtracting cantilever deflection ($\Delta z_c$) from the movement of the piezoelectric positioner in the $z$-axis ($\Delta z_p$). Thus, the cantilever deflection is obtained: $L_1-L_0=\Delta z_p - \Delta z_c$. The photodiode output is related to the cantilever deflection $\Delta z_c$, and therefore the applied forces are calculated by applying to the equation: $F=-k \cdot \Delta z_c$ (where the spring constant $k$ is determined from the calibration). The contour length of the protein increases when a domain of the protein unfolds, and eventually the force acting on the cantilever is relaxed (Figure 1.2(3)). When stress is applied to a folded domain, the protein will unravel. The extension of another domain of the same protein begins. The force-measuring mode of the AFM results in force-extension curves with a characteristic saw-tooth pattern appearance.
**Figure 1.2** Using single-molecule AFM to measure the elasticity of single proteins:

(Upper) The unfolding of a protein by an external force. The protein extension \((L_1-L_0)\) pulled by the piezoelectric positioner can be calculated by subtracting cantilever deflection \((\Delta z_c)\) from the movement of the piezoelectric positioner in the z-axis \((\Delta z_p)\). Thus, the cantilever deflection can be measured from the equation: \(L_1-L_0 = \Delta z_p - \Delta z_c\).

The applied forces are then acquired by applying to the equation: \(F=-k \cdot \Delta z_c\) (where the spring constant \(k\) is determined from the AFM calibration).

(Lower) A force-extension curve of a polyprotein is well-fitted with the worm-like chain (WLC) model (blue solid lines), which describes a polyprotein as a continuous string of a given total length (contour length, \(L_c\)). Each individual peak of the saw-tooth pattern force-extension curve represents the unravelling of individual domains of a protein. \(\Delta L_c\) represents the contour length increment, which gives an indication of the mechanical structure of the domain.
Proteins tend to coil up in order to maximize the entropy of the polymer chain. Extension of the protein generates an opposing force due to the decrease of the entropy of the polymer chain. Such behaviour is known as the entropic elasticity of polymers. The entropic elasticity of proteins can be described by the worm-like chain (WLC) model of polymer elasticity, which predicts the entropic restoring force \( F \) generated upon extension of a protein \( (x) \\).

\[
F(x) = \frac{kT}{p} \left[ \frac{1}{4} \left( 1 - \frac{x}{L_c} \right)^2 - \frac{1}{4} + \frac{x}{L_c} \right]
\]

where \( k \) and \( T \) represent the Boltzmann constant and temperature in Kelvin, respectively. \( p \) is the persistence length, which reflects the polymer flexibility. \( x \) and \( L_c \) are the extension of the protein and its contour length, respectively. Two fitting parameters of the WLC model are the persistent length and its contour length. Each force peak of the saw-tooth pattern corresponds to the unravelling of each individual domains of a protein. The last peak of the force-extension curve represents the final extension and the subsequent detachment of the completely unfolded polyprotein chain from either the cantilever tip or the glass substrate. The consecutive unfolding force peaks of the polyprotein are well fitted by the worm-like-chain model (WLC).

The contour length increment \( (\Delta L_c) \) gives an indication of the mechanical structure of the domain. The contour length increment, which is defined as the length of the ‘force hidden’ or ‘trapped’ amino acid region of the protein between the bonds forming the major mechanical resistance to mechanical unfolding, can simply calculated by subtracting the length of the folded domain (= distance of the N- and C- termini) from the length of the unfolded and fully extended domain (=number of amino acid within a domain × the length of each amino acid, ~0.36nm).
Protein unfolding is a stochastic event. The unfolding force of a protein is an indication for mechanical stability. The mechanical stability of a protein can be explained as the resistance to unfolding in response to an external force, this unfolding force to a protein specifically means ‘the most probable unfolding force’ at a given pulling speed. Once enough data points are collected for a certain pulling speed, and the histogram is built, the mechanical unfolding force of a protein is taken to measure the most probable unfolding force at any given pulling speed from the histogram. From the histograms, the unfolding distance ($\Delta x_u$) and the unfolding rate constant at zero force ($\alpha_0$) can also be analyzed by using Monte Carlo simulation. However, the histogram of mechanical unfolding force might broaden due to experimental errors, resulting in incorrectly predicted values of unfolding distance and the unfolding rate constant at zero force. Therefore, pulling speed dependence experiments are performed by using the AFM to further confirm the unfolding distance and the unfolding rate constant at zero force. AFMs have a broad pulling speed range. Since an unfolding event is a non-equilibrium reaction, the slower the pulling speed is, the lower the force required to unfold a domain. The average unfolding force is dependent on the pulling rate. At very low pulling speeds, the limiting factor for collecting reasonable data points is the thermal drift, whereas viscous drag and cantilever response time may introduce errors at very high pulling speeds. The errors caused by pulling at very low speeds and very high speeds can be reduced by allowing samples and the AFM to equilibrate and using a short, low noise cantilever, respectively.
1.2.2 The mechanical unfolding of a domain can be modelled as a two-state process

The purpose of data analysis is to determine an unfolding rate constant at zero force \((k_u^0)\) and the unfolding distance \((\Delta x_u)\). The unfolding of a domain under an applied force is a probabilistic event that can be modeled as a two-state Markovian process with force-dependent rate constants. In the two-state process, the protein adopts only a native or a denatured conformation separated by a high energy transition state. Figure 1.3 shows how when proteins unfold under an applied force, the barrier to unfolding is lowered, relative to the native state. Such an effect was first addressed by Bell in 1978\(^5^3\), with Evans and Richie providing a more detailed explanation in 1997\(^5^4\). Along the reaction coordinate of an unfolding/folding reaction, the distance between the native to the mechanical transition state is called the unfolding distance of a protein \((\Delta x_u)\). The force-dependent unfolding rate constant \(k_u(F)\) is shown in the following equation and is determined by the activation energy of unfolding, \(\Delta G_{TS-N}\) and unfolding distance over which the force must apply to reach the transition state.

\[
k_u(F) = \nu \kappa \exp\left[-\left(\Delta G_{TS-N} - F\Delta x_u\right)/k_B T\right]
\]

where \(\nu\) is the vibrational frequency at the transition state, and \(\kappa\) is the transmission coefficient. \(k_B\) and \(T\) are the Boltzmann constant and the temperature in Kelvin, respectively. \(F\) represents the applied force. When the force applies to the system, it lowers the activation energy barrier to unfolding during the pulling experiment. Therefore, the external force on the unfolding rate constants is calculated as
\[ k_u(F) = k_u^0 \exp\left(\frac{F \Delta x_u}{k_u T}\right) \]  

where \( k_u^0 \) is the unfolding rate constant under zero force. According to equation (3), for a given \( k_u^0 \), if the unfolding distance is small, the unfolding force will be high, and vice versa.

**Figure 1.3** A schematic diagram of a two-state process (from folded state: N to unfolded/denatured state: U with a high energy barrier transition state: TS) free energy for the mechanical unfolding of proteins. The blue curve represents landscape at zero force, while the red curve indicates the unfolding under applied force F. The energy barrier between N and TS (\( \Delta G_{TS,N} \)) is lowered by \( F x_u \) where \( x_u \) represents the unfolding distance between N and TS. \( \Delta G_{U,N} \) indicates the thermodynamic stability of protein (the free energy difference between the unfolded and folded states of proteins)\(^6\).
Stretching a polyprotein results in the force-extension relationship that shows a complex sawtooth pattern appearance. To simulate the complex behaviour\textsuperscript{6,7}, Monte Carlo techniques are used to model and reproduce the probability of unfolding versus the applied force in order to estimate the unfolding distance $\Delta x_u$ and the unfolding rate constant at zero force $k_u^0$. The Monte Carlo simulation is used to fit the pulling kinetic experiments and the speed dependence of the unfolding forces. The best fits for each unfolding force are obtained. The values of $k_u^0$ and $\Delta x_u$ in the simulations are adjusted until the simulation data match the experimental ones.

\subsection*{1.3 Protein folds have different mechanical stability}

Each type of protein has its particular three-dimensional structure\textsuperscript{1}. Protein folding is usually a spontaneous reaction. A protein can be unfolded or denatured by treatment with certain solvents, thermal melt, an externally applied force, or change of the pH to disrupt the non-covalent interactions holding the folded chain together. Eventually, proteins lose their natural structure. When the factors involved in unfolding the protein are removed, the protein often refolds spontaneously into its original conformation.

When proteins fold improperly, they may form aggregates that can damage cells and even whole tissues, leading to a number of neurodegenerative disorders, including Alzheimer’s disease, Cystic fibrosis, and Huntington’s disease\textsuperscript{1}. Molecular chaperones\textsuperscript{1} are a special class of proteins that assist in protein folding in a living cell. Once folding is complete, the chaperone will leave its current protein and like a catalyst, will move on to support the folding of another protein. Molecular chaperones bind to partly folded protein domains and help them to fold along the most energetically favourable pathway to make the folding process more efficient and reliable.
As an external force is applied to the N- and C- termini of a protein by single-molecule AFM, the protein starts to unfold. Single-molecule AFM has been used to study proteins with different folds: some of them consider to be an ‘all beta’ structure (β-sandwich topology), some of them to be an α/β structure, and others have an ‘all alpha’ structure. Examples of ‘all beta’ structure from typical mechanical proteins are immunoglobulin (Ig) domain and I27 module. Proteins with α/β structure include protein G B1 domain. A typical example ‘all alpha’ proteins is calmodulin, which has a fold with low mechanical stability. Since the studies on titin published in 1997, many other mechanical proteins and non-mechanical proteins have been examined by single-molecule AFM to study their mechanical properties in a similar manner. What we have learned from these contributions is that many mechanical proteins and non-mechanical proteins are able to withstand repeated cycles of force stretching and relaxation. Figure 1.4 indicates the three-dimensional structures of representative proteins that have been investigated by using single-molecule AFM. This section will outline several studies on the combination of AFM and molecular biology to study the mechanical stability of proteins. These proteins include titin, I27, and protein G B1 domain (GB1).
Figure 1.4 Three dimensional-structures of representative proteins: mechanically labile proteins (upper), natural elastomeric proteins (middle) and non-mechanical proteins have been examined by using single-molecule AFM.

1.3.1 Titin immunoglobulin domains possess high mechanical stability

AFM studies on titin\textsuperscript{7-11} provided insights of how the passive elasticity of muscle is finely regulated by the collective mechanical properties of the constituting folded immunoglobulin Ig-like domains as well as a small fraction of non-structured sequences, such as PEVK sequence and N2B sequence.\textsuperscript{11} The mechanical properties of titin are essential for its biological function. Titin is a giant elastic protein in the sarcomeres of cardiac and skeletal muscles. The mechanically active region of titin in the sarcomeric I band is assembled from immunoglobulin (Ig)-like domains arranged in tandem. An Ig domain comprises ~90-100 amino acids and folds into $\beta$-sandwich structure, which is composed of seven $\beta$-strands folding into a $\beta$-barrel. The PEVK domain contains coiled conformations that are elongated when the muscle is stretched.
Single-molecule AFM stretch experiments have been performed with native titin. Stretching titin results in force-extension curves with saw-tooth pattern appearance, which display a hierarchical ordering of mechanical stability varied between 150 and 300 pN. The adjacent peaks of the force-extension curves are spaced between 25 to 28 nm, shown in Figure 1.5\(^7\). Titin is a modular protein that is constituted by heterogeneous populations of domains differing in size, structure, sequence and stability. It is impossible to assign each force peak to the unfolding of a specific Ig domain. The construction of polyproteins (tandem repeats of a single domain) allows the study of the mechanical properties of specific domains. Polyproteins have a perfect repetitive structure that results in a periodical saw-tooth pattern in force extension curves. This ‘fingerprint’ clearly indicates that domains from a protein are being unfolded and provide the measurement of the unfolding force for these domains as well as their size. The 27\(^{th}\) immunoglobulin (I27) domain of the human cardiac muscle protein titin was selected and studied by single-molecule AFM to show the advantages of constructing polyproteins.

**Figure 1.5** Force-extension curves of titin proteins show periodic saw-tooth patterns that are consistent with their modular construction. Middle and lower traces show the
stretching of polyproteins preceded by a spacer region of variable length. Each force peak represents the stretching and unfolding of titin domains.  

1.3.2 The I27 module

The 27th immunoglobulin (I27) domain of the human cardiac muscle protein titin was constructed into direct tandem repeats domain by protein engineering method. I27 was chosen because the tertiary structure and its stability have been determined. The unfolding and stretching of I27 had also been simulated using SMD. I27 is composed of 89 amino acids and has a typical Ig I topology composed of seven β-strands (strands A-G), which fold into two face-to-face β-sheets through backbone hydrogen bonds and hydrophobic core interactions. Most adjacent β-strands are in anti-parallel orientation, except A’G strands are arranged in parallel. The mechanical properties of I27 have been studied by using single-molecule AFM. I278 and I2712 polyprotein were constructed. The resulting force-extension curves with saw-tooth patterns show each unfolding event of contour length increment of ~28 nm, shown in Figure 1.6. A histogram of force peaks displays an asymmetric distribution maximum at ~200 pN, this mechanical unfolding force is a parameter that indicates the mechanical stability of a particular folded domain under the specific experimental conditions (ie. pulling speeds).
**Figure 1.6** Mechanical properties of single human cardiac titin immunoglobulin domains. (A) Construction of I27 polyproteins: Agarose gel stained with ethidium bromide showing the sizes of the I27 multimers (left). Coomassie blue staining the purified (I27)$_8$ polyprotein (~90 kDa) separated by using SDS-PAGE (right)$^{57}$. (B) Force-extension curves of (I27)$_{12}$ (upper) and (I27)$_8$ (lower) polyproteins show periodic saw-tooth patterns with equally spaced force peaks. Fits of the WLC model to the consecutive peaks give a contour length increment of ~28 nm, represented by blue lines$^{57}$. (C) The mechanical topology of I27. The structure shows the shear topology, where the hydrogen bonds break simultaneously as the forces apply to N- and C- terminus of the strands, indicated by black arrows. The dashed lines indicate the hydrogen bonds which are responsible for the mechanical resistance under an applied force$^{48}$.

According to SMD$^{48}$, a cluster of backbone hydrogen bonds connecting A’ strand to the G strand provides the major mechanical resistance to unfolding. Therefore, the A’G region, considered to be the mechano-active site, is believed to be the key to the mechanical stability of I27. These hydrogen bonds are perpendicular to
the direction of the applied force at N- and C- termini of β-strands (a shear topology), unfolding is done by rupturing the hydrogen bonds simultaneously. Changing the mechanical stability can be achieved by tuning the mechano-active site of the protein. A’G region is the mechano-active site of I27. Using site-directed mutagenesis, the key backbone hydrogen bonds between A’ and G strands were disrupted by proline point mutation. As anticipated and shown in Figure 1.7, proline point mutation at V11P, V13P and V15P significantly reduces the mechanical stability of I27, whereas point mutation at Y9P increases the mechanical unfolding. Despite the unexpected result from Y9P point mutation, this study indicates the possibility that the mechanical stability of systems can be tune by tuning the interactions in the mechano-active site region.

**Figure 1.7** Point mutations in an I27 domain alter its mechanical stability. (A) A cartoon diagram shows the structure of I27 and the amino acids that are substituted by Pro residues. Black bars indicate the hydrogen bonds connecting the two strands which are the major source for the responsibility of the mechanical stability. (B) Force-extension curves for each mutants. V13P, V11P and V15P show reduced mechanical stability when compared to wt-I27. In contrast, V9P exhibits an increased
mechanical unfolding force ~50pN in comparison to wt-I27\textsuperscript{56}.

1.3.3 Length phenotype is sensitive to the precise location of the mutation with respect to the mechanical topology of single proteins

Molecular dynamic simulations predict that hydrogen bond networks between A’ and G β-strands of I27 module exhibit a high mechanical resistance that must be ruptured to trigger unfolding\textsuperscript{48}. These I27 domains contain a mechanically isolated region (force-hidden region) that is placed behind the point of high mechanical resistance in the A’G overlap. In order to demonstrate the validity of the prediction, AFM experiments combined with protein engineering techniques were carried out. Glycine residues were inserted in the locations either inside the mechanically isolated region or outside the region near N- or C- termini of I27. Polyproteins were constructed based on wild-type and mutant forms of the I27 module with a five-Gly insert in the FG hairpin loop at position 75 (within the predicted force-hidden region, I27::75Gly\textsubscript{5}), in the N terminus region at position 7 (I27::7Gly\textsubscript{5}), and after the C-terminus at position 89 (I27::89Gly\textsubscript{5})\textsuperscript{57}. The length of the protein mutants before and after an applied force were determined and measured, which gives an explicit mechanical topology of the I27 module. The contour length increment of I27::75Gly\textsubscript{5} mutants, in comparison to the wt-I27, has lengthened, on average, by 2.00 nm per domain, which corresponds to the length of the five inserted amino acids. These hidden amino acids within the mechanically isolated region become exposed after the unfolding therefore an increased in the contour length increment is observed only after an applied force. By contrast, the contour length increments of (I27::7Gly\textsubscript{5}) and (I27::89Gly\textsubscript{5}) mutants do not show any changes as compared to wt-I27 because these regions are already pre-stretched before unfolding takes place. This concept gives us an idea of making GB1 loop mutants and studies their mechanical stability.
Figure 1.8 Single-molecule AFM measurements of the contour length of engineered polyproteins. Structure of the I27 Ig-like module and locations of the five Gly insertion sites (center). A comparison of the force-extension curves of wt-I27 and mutant polyproteins: I27::7Gly₅ (left), I27::7Gly₅ (middle), and I27::89Gly₅ (right). After 10 consecutive unfolding events, I27::75Gly₅ polyprotein (red curve, left) shows 19.1 nm longer than the wild-type polyprotein (black curve, left). The force-extension curves of 27::7Gly₅ and I27::89Gly₅ (red traces, middle and right) superimpose on the saw-tooth pattern obtained from the wt-I27 (black traces, middle and right)⁵⁷.

1.3.4 Single protein misfolding events captured by atomic force microscopy

Mechanical unfolding has allowed the identification of misfolding events in polyproteins. Missing peaks are observed during AFM refolding experiments. Typically, a single polyprotein remains attached to an AFM tip allowing for repeated stretching and relaxation cycles (double-pulse protocol). During this repeated stretching/relaxation cycles, the sawtooth pattern of the force-extension curves shows some missing peaks. Such misfolding events are observed in the refolding experiment of I27⁵⁸. Figure 1.9 indicates an example of misfolding events in (I27)₈ polyprotein. A first extension (upper, Fig. 1.9) shows five unfolding peaks. Mechanically unfolded I27 polyprotein readily refolds when relaxed to their resting length and then extended,
and shows only four peaks (middle, Fig. 1.9). Five peaks are revealed again in the third pull (lower, Fig. 1.9). The misfolding of I27 is fully reversible and changes the mechanical topology of the domain while it is about as stable as the original structure. Such misfolding events are likely to occur due to the interaction between A’ and G patches of neighbouring I27 domains, where the force bearing amino acids linking these modules are hidden by becoming part of the ‘skip’ fold\(^{58,59}\).

**Figure 1.9** Misfolding of I27 domains. (A) Consecutive force-extension curve of \((I27)_8\) at 10 second intervals. The middle force-extension curve appears to have a ‘skip’ force peak\(^{58}\). (B) A schematic diagram shows how misfolding events are likely to occur\(^{59}\).

### 1.3.5 Non-mechanical protein can have significant mechanical stability

I27 of titin displays a high mechanical stability due to its shear topology\(^{48}\). The key structural feature of I27 is its two terminus β strands that are arranged in parallel and connected by a series of backbone hydrogen bonds. Based on this topological consideration, non-mechanical protein GB1 has been identified from the pool of small
model proteins and possibly possesses high mechanical stability. The B1 immunoglobulin G (IgG) binding domain of *streptococcal* protein G is one of the smallest candidates\(^{32}\). The B1 IgG binding domain of protein G (GB1) is a small \(\alpha/\beta\) protein consisted of 56 amino acid residues and is composed of a four-strand \(\beta\)-sheet packed against a long \(\alpha\) helix. According to its structure, the two central \(\beta\)-sheet strands with N- and C- terminus are parallel to each other. The extensive hydrogen bonds are networking among the two central strands, which potentially exhibit mechanical resistance to the applied force. Based on this shear topology, GB1 will be mechanically stable. Polyprotein of (GB1)\(_8\) was constructed and studied by single-molecule AFM\(^{32}\). The results reveal that GB1 protein has high mechanical stability at pulling speed of 400 nm/s. These force peaks from the force-extension curves are equally spaced and fitted by WLC model, displaying contour length increment of \(~18\) nm. The amplitudes of the force peaks vary around 180 pN. The unfolding force histogram for GB1 has its maximum force at 178 pN, which displays a high mechanical stability. The unfolding distance, \(\Delta x_u\) of 0.17 nm and the unfolding rate constant at zero force, \(\alpha_0\) of 0.039 s\(^{-1}\) are obtained from the unfolding force histogram by using Monte Carlo simulation. The pulling speed experiment of GB1 was carried out and compared that with I27. Clearly, the data show that the unfolding force for GB1 has a much stronger dependence on the pulling speed than that of the I27\(^{32}\). Since GB1 has been extensively studied and possess high mechanical stability, we use GB1 as a model system to directly investigate the effect of mutations on mechanical properties of GB1 by using single-molecule AFM.
Figure 1.10 The single molecule AFM results reveal that the GB1 protein is mechanical stable. (A) Force-extension measurements on (GB1)$_8$ polyproteins show a periodic saw-tooth pattern with equally spaced force peaks. WLC model can be fitted to the individual unfolding force peaks with a contour length increment of ~17.7 nm, represented by gray lines. (B) A histogram of unfolding forces for GB1 protein has the maximum at ~178 pN. The solid black line is the Monte Carlo simulation of the mechanical unfolding of (GB1)$_8$ polyprotein, with unfolding rate constant of 0.039 s$^{-1}$ and an unfolding distance of 0.17 nm at pulling speed of 400 nm/s. (C) Unfolding kinetics of the GB1 protein (black square). The measured unfolding forces at high pulling speed are corrected for the hydrodynamic drag force that acts on the AFM cantilever. The unfolding kinetics of GB1 and I27 are fitted by Monte Carlo simulations, respectively (black, solid line and grey line). GB1 has a smaller unfolding distance than I27$^{32}$. 
1.4 Thesis overview

This chapter covers several studies on the combination of single-molecule AFM with protein engineering techniques, as well as explains the principle methodology of AFM. In this thesis, we use a small protein, GB1, the B1 IgG binding domain of protein G from *Streptococcus*, as a model system to directly investigate the mechanical stability of GB1 by using single-molecule AFM in combination with protein engineering techniques. In nature, some domains of proteins have been found to contain mutations. Many of these mutations change the functions of proteins, such as the ligand binding properties. It is also believed that many of these mutations change the mechanical stability of the proteins. Recent studies have demonstrated that mutations affect the resistance to mechanical unfolding of GB1. However, how mutations affect the mechanical stability of non-mechanical proteins have not been investigated. In chapter 2, we demonstrate the study on the effect of point mutations on the mechanical stability of the non-mechanical protein, GB1. Such effects are studied by using Φ-value analysis. Loop region plays a critical role in protein structures and functions. Chapter 3 reveals that the effect of loop insertion in the second loop of GB1 changes the mechanical stability. The mechanical stability of GB1 is modulated by the configuration entropy. In chapter 3, we conclude that as the number of inserted amino acid residues into loop 2 of GB1 increases, it reduces the mechanical stability of GB1 and accelerates the mechanical unfolding rate constant. Both chapters provide insights that by tuning the mechano-active site of its domain, the mechanical stability of a protein is altered. Chapter 4 summarizes the findings of the thesis and outlines the future directions on these topics.
1.5 References


Chapter 2: Point mutation alters the mechanical stability of protein GB1

2.1 Introduction

Characterization of the structure and energetic of the transition state is a significant piece of our understanding towards protein folding. Unfortunately, the transition state is the least well-populated, transient and partially unstructured state. Therefore, its structure is extremely difficult to determine and characterize by using the direct traditional methods such as NMR or X-ray crystallography. Instead, indirect approaches include the change in the environmental conditions, such as pH, temperature or denaturant concentration, and consequences of mutations, were allowed to study the properties of the transition state. Application of protein engineering techniques has currently taken place into the understanding of the structure of transition state of protein folding/unfolding. Using protein engineering method to probe the folding pathways of proteins was initiated by the Fersht’s laboratory. The protein engineering method becomes the most common used method to evaluate the importance of the interactions mediated by specific amino acid side chains that change the stability of the folding/unfolding transition state structure of a protein. Mutants contain amino acid substitutions located throughout a protein and are probed for their folding kinetics and equilibrium thermodynamics. Results are analyzed and allowed the calculation of the extent of formation of the structural surrounding a particular mutation site in the folding transition state, which is defined as the $\Phi$-value.

In $\Phi$-value analysis, the folding/unfolding kinetics and folding/unfolding free

---

1 A version of this chapter will be submitted for publication. Wang, H-C. and Li, H. Point mutation alters the mechanical stability of protein GB1.
energies of a protein with one or more point mutations are compared with those of the wild-type protein, giving the range from 0 to $1^{7-10}$. Such comparison indicates the contribution of the degree of mutation on a protein towards the folding/unfolding pathways. Φ-value analysis has been used widely to study the transition state structure of the folding/unfolding pathway for many proteins, using chemical ensemble methods, which include barnase$^{5,6,11,12}$, CI2$^{13}$, SH3 domains$^{14,15}$, immunoglobulin domain$^{16}$ and individual domains of protein G$^{17,18}$. Despite the detailed studies of Φ-values using the ensemble chemical methods, there is very little study on the mechanical Φ-value analysis of the mechanical unfolding of proteins$^{21}$.

Many proteins are exposed to mechanical tensions$^{22-29}$. Single-molecule force spectroscopy studies combining with protein engineering techniques is an efficient way to reveal the mechanical property of a protein. Atomic force microscopy (AFM) has been largely associated with protein engineering techniques to investigate the folding/unfolding kinetics as well as the mechanical stability of proteins, at the single-molecule level$^{30-34,47}$. Here, we use a small protein, GB1, the B1 IgG binding domain of protein G from Streptococcus, as a model system to directly investigate the effect of single point mutation on the mechanical stability of GB1 by using single-molecule AFM. GB1 contains 56 amino acid residues and is an α/β protein composed of a four strand β-sheet packed against a α-helix$^{35}$. GB1 has high thermo stability and contains no disulfide bonds or prosthetic groups. Its mechanical folding and unfolding dynamics have been investigated in detail using single-molecule AFM$^{22,36,37}$. Recent study indicates that GB1 has excellent mechanical properties and is easy to manipulate by AFM$^{22,37}$, which makes it a good candidate for single-molecule force spectroscopy studies. Using protein engineering techniques, we constructed 10 site-directed mutants, L7A, T11A, T16A, T25A, K28A, F30L, K31A, N35A, D46A and T53A, which distribute at different regions of GB1 (Fig.2.1), and
AFM is employed to investigate the effect of each point mutation on the mechanical stability and mechanical folding/unfolding kinetics of GB1. Substitutions should be as non-disruptive as possible. We mainly used Ala (A) substitution because Ala is less likely to form new non-native interactions when substituted into a protein than any other amino acids\textsuperscript{10,17,18}. We found that point mutations F30L and T53A in GB1 reduce the mechanical stability of GB1. F30L and T53A mutations that have relatively higher $\Phi$-values can be considered to be as unstructured in the transition state as they are in the unfolded state. It indicates that the transition state structures of F30L and T53A are destabilized by the mutation. In addition, we compared $\Phi$-values of GB1 mutants from AFM with those obtained on isolated, un-tethered modules by using standard chemical folding/unfolding methods\textsuperscript{17} and found that different point mutations might affect mechanical and chemical stability of GB1 with different degrees. Our results not only demonstrate that point mutation of GB1 may alter its mechanical stability and change its transition state due to the disruption of any specific interaction, but also indicate that the unfolding/folding rate ratio and $\Phi$-values obtained from the mechanical and chemical studies may be different.
Figure 2.1 Backbone ribbon diagram indicating structure of GB1 and 10 point mutations made on GB1. GB1 is a small $\alpha/\beta$ protein with 56 amino acid residues and is composed of a four-strand $\beta$ sheet packed against an $\alpha$-helix. All the point mutations are shown in different colors.

2.2 Methods and materials

2.2.1 Protein engineering: polyprotein construction

The multiple-step cloning technique was used. The plasmid encoding the GB1 was generously provided by David Baker of the University of Washington, Seattle. Mutants L7A, T11A, T16A, T25A, K28A, F30L, K31A, N35A, D46A and T53A were constructed using PCR-based site-directed mutagenesis techniques and confirmed by direct DNA sequencing. This led to the products of desired GB1 mutants (mGB1)-pUC19. The I27 insert containing BamHI and KpnI was ligated into the (mGB1)-pUC19 (mGB1= L7A, T11A, T16A, T25A, F30L, K31A, N35A, D46A and T53A) at restriction site KpnI and BglII. Insert (mGB1-I27) flanked by a 5’ BamHI site and a 3’ BglII site followed by KpnI site was obtained. The (mGB1-I27)$_4$ (see Appendix D)
genes were constructed by iterative cloning monomer into monomer, dimer into
dimer, tetramer into tetramer based on the identity of the sticky ends generated by the
*BamHI* and *BgII* restriction enzymes. Insert (GK28A), (GK31A) or (GN35A), also
containing *BamHI* and *KpnI* was ligated into (mGB1)-pUC19 (mGB1= K28A, K31A,
and N35A). The (mGB1)$_8$ genes were constructed using procedures similar to the
construction of heteropolyprotein, which are based on the identification of the sticky
ends of *BamHI* and *BgII* sites. The 8mers (polyprotein- (mGB1)$_8$ and
heteropolyprotein (mGB1-I27)$_4$) were subcloned into pQE80L expression vector,
which contains an N-terminal (His)$_6$ purification tag to facilitate the purification of
expressed proteins. The products were then transformed into DH5α competent cells
for expression to generate proteins. The polyproteins and heteropolyproteins were
purified by using Ni-NTA affinity chromatography and then analyzed by SDS-PAGE
to confirm the size and the expression level (see Appendix E). The final constructs
contained eight repeat tandem domains of (mGB1)$_8$ or (mGB1-I27)$_4$.

**2.2.2 Single-molecule AFM experiment**

Single-molecule AFM experiments were carried out in a custom-built atomic
force microscope, which was constructed as described. All of the experimental
measurements for GB1 mutant polyproteins were carried out in PBS buffer. 1uL of
~0.6mg/mL (mGB1-I27)$_8$ heteropolyprotein sample or (mGB1)$_8$ polyprotein sample
was deposited onto a clean glass cover slip covered by ~50 µL PBS buffer and was
allowed to adsorb for approximately 10 min. The spring constant of each individual
cantilever (Si$_3$N$_4$ cantilevers from Veeco, with a typical spring constant of 40 pN nm$^{-1}$)
was calibrated in solution using the equipartition theorem before each experiment.
Single-molecule AFM pulling experiments were performed. The pulling speed for all
the unfolding experiments was 400 nm s$^{-1}$, except when reported otherwise.
2.2.3 Monte Carlo simulations

Monte Carlo simulations of the stretching on the GB1 polyprotein mutants were carried out using standard procedures as described previously (44, Appendix H). The mechanical unfolding of the wt- and mutant GB1 polyproteins was described as a two-state Markovian process with a force-dependent rate constant. Both unfolding rate constant at zero force $k_0^u$ and the distance of the native state to the transition state $\Delta x_u$ along the reaction coordinate of the mechanical unfolding reaction were estimated.

2.3 Results

2.3.1 Engineering heteropolyproteins and homopolyproteins of GB1 mutants

We used site-directed mutagenesis\textsuperscript{33,38-40} to construct point mutations at 10 different positions of the GB1 domain. Nine point mutations were substituted by Ala residues, while one point mutation was substituted by Leu residue. Substitutions were made as non-disruptive as possible in order to avoid any introduction of new interactions. Substitutions were located in different regions of the GB1. Point mutations at L7A, T11A, T16A, T25A, F30L, D46A and T53A of GB1 were engineered with Ig modules 27 (I27) and made into heteropolyproteins (polyprotein chimeras), while point mutations at K28A, K31A and N35A were constructed into eight repeated domain polyproteins (homopolyprotein).

2.3.2 AFM unfolds the proteins and results in a characteristic saw-tooth pattern

To investigate the mechanical stability of point mutation GB1, we constructed seven polyprotein chimeras, which contain 4 repeats of Ig modules 27 (I27) and GB1 mutant dimer (L7A-I27)\textsubscript{4}, (T11A-I27)\textsubscript{4}, (T16A-I27)\textsubscript{4}, (T25A-I27)\textsubscript{4}, (F30L-I27)\textsubscript{4}, (D46A-I27)\textsubscript{4} and (T53A-I27)\textsubscript{4}. Stretching single molecular polyprotein chimeras resulted in force-extension curves with a characteristic saw-tooth pattern appearance
which showed as many as 8 peaks of two different levels of unfolding forces with two
distinct spacing. The polyprotein chimera was picked up randomly by the AFM tip,
the majority of force-extension curves corresponds to the stretching and unfolding of
part of the polyprotein. The individual force peaks correspond to the mechanical
unfolding of the individual GB1 or I27 domains in the polyprotein chain. The last
peak of the force-extension curve represents the final extension and the subsequent
detachment of the completely unfolded polyprotein chain from either the cantilever
tip or the glass substrate. The consecutive unfolding force peaks of the polyprotein are
fitted by the worm-like-chain model (WLC), which predicts the entropic restoring
force generated on the extension of a polymer. In the polyprotein chimera, the well
characterized I27 domains serve as fingerprints for identifying single-molecule
stretching events and discerning the signatures of the mechanical unfolding of GB1
mutants. The mechanical unfolding of I27 is characterized by the contour-length
increment of $\Delta L_c$ of ~28.4 nm and unfolding force of ~ 200 pN in
phosphate-buffered saline (PBS) at a pulling speed of 400 nm s$^{-1}$. Its structure has
been determined by using NMR and has known its chemical and mechanical
stability. The unfolding force of I27 modules in the heteropolyproteins is similar
to that of the (I27)$_8$ homopolyprotein. The mechanical property of the wild type
GB1 (wt-GB1) is well understood from the previous studies. The mechanical
unfolding force of polyprotein wild type GB1 is ~ 180 pN with the contour-length
increment $\Delta L_c$ of ~ 18.0 nm, which is in good agreement with the calculated contour
length increment. Each GB1 domain has 56 amino acid residues, and its proper
folding is 2.6 nm between the N- and C- termini of wt-GB1. Thus, it gives the $\Delta L_c$
of ~18.0 nm (56*0.36-2.6) $^{22}$. Point mutation of GB1 would not alter the contour length
increment since the mutation does not affect the proper folding of GB1.
According to the experimental data, the unfolding force of wt-GB1 modules in heteropolyproteins is similar to that of (wt-GB1)$_8$ homopolyproteins$^{22}$. Figure 2.2 represents examples of the stretching and unfolding of polyprotein chimeras, consisting of up to eight unfolding events occurring at ~200 pN with $\Delta L_c$ of ~ 28.4 nm, which can be easily identified as the mechanical unfolding of the I27 domains, and with $\Delta L_c$ of ~ 18.0 nm of GB1 mutants with various unfolding forces for different mutants. In addition to these polyprotein chimeras, we further engineered three GB1 mutant homopolyproteins, (K28A)$_8$, (K31A)$_8$ and (N35A)$_8$, which are composed of eight identical tandem repeats of GB1 mutants. Force-extension curves showed as many as eight equally spaced unfolding peaks for the three GB1 mutant homopolyproteins. Fits of the WLC to the consecutive peaks resulting from the unfolding events of the GB1 mutant homopolyproteins give a contour length increment of $\Delta L_c$ ~ 18.0 nm.
Figure 2.2 The representative force-extension curves of the polyprotein chimeras constructed as (mGB1-I27)$_4$ and (mGB1)$_8$. Stretching a single polyprotein results in a force-extension curve with a saw-tooth pattern appearance, which is well-described by the WLC equation. Each individual unfolding force peak corresponds to the mechanical unraveling of each individual domain in the polyprotein chain. Fits of the WLC to the consecutive unfolding force peaks of (GL7A-I27)$_4$ (red lines) indicate that the contour length increment $\Delta L_c$ of GB1 mutants and I27 are $\sim$18.0nm and $\sim$27.0nm, respectively. The color green indicates the contour length increment of GB1 mutants, $\Delta L_c$ $\sim$18.0nm. Peaks for I27 are indicated in black, contour length increment of I27, $\Delta L_c$ $\sim$28.0nm.

2.3.3 Mutations, F30L and T53A result in the decrease of the mechanical stability of GB1

Unfolding force histograms from each of the GB1 mutants and wt-GB1 are shown in Fig. 2.3. The mechanical unfolding forces of F30L and T53A are $\sim$154pN and $\sim$151pN, respectively. Histograms for other GB1 mutants L7A, T11A, T16A,
T25A and D46A, of the polyproteins show similar mechanical unfolding forces compared to the wt-GB1. The mechanical unfolding forces from the histograms for the GB1 mutants of L7A, T11A, T16A, T25A, and D46A of the polyprotein have the maximum at 190pN, 188pN, 171pN, 183pN, and 198pN, respectively. These results clearly reveal that the point mutations F30L and T53A reduce the mechanical stability of GB1, while point mutations at L7A, T11A, T16A, T25A and D46A do not alter the mechanical stability of GB1. Unfolding force histograms for the GB1 mutant homopolyproteins, (K28A)$_8$, (K31A)$_8$ and (N35A)$_8$, illustrate the similar mechanical stability as wt-GB1 as their average mechanical unfolding force is 178pN, 175pN and 179pN, respectively. Therefore, the alanine mutagenesis at K28, K31 and N35 residues does not alter the mechanical stability of GB1.

Monte Carlo techniques are used to model and reproduce the unfolding force distribution versus the applied force in order to estimate the unfolding distance from the native state to the transition state $\Delta x_u$ and the unfolding rate constant at zero force $k_u^0$. Using the previous described method $^{41,43-45}$, the values of $k_u^0$ and $\Delta x_u$ in the simulations were adjusted until the simulation data matched the experimental ones. The best fits for the histograms of unfolding forces were obtained using an unfolding distance, $\Delta x_u = 0.17$ nm for all GB1 mutants and wt-GB1, and unfolding rates at zero force, $k_u^0$ of 0.0314 s$^{-1}$ for L7A, 0.0340 s$^{-1}$ for T11A, 0.0610 s$^{-1}$ for T16A, 0.0440 s$^{-1}$ for T25A, 0.050s$^{-1}$ for K28A, 0.1130 s$^{-1}$ for F30L, 0.0540 s$^{-1}$ for K31A, 0.0480 s$^{-1}$ for N35A, 0.0260 s$^{-1}$ for D46A, 0.1190 s$^{-1}$ for T53A, and 0.0460 s$^{-1}$ for wild type GB1.
Figure 2.3 Histograms for the mechanical unfolding force of GB1 mutants and wt-GB1. The red solid lines correspond to the Monte Carlo simulations of the unfolding forces of GB1 mutant domains of the polyprotein when performing at the pulling speed of 400 nm s$^{-1}$. The average unfolding forces, $\Delta x_u$ and $k_u^0$ are indicated in Table 1. It is of note that the detection limit measured for the unfolding force in our AFM is ~20pN. The average unfolding force is $190 \pm 40$pN (n=447) for GL7A, $188 \pm 45$pN (n=691) for GT11A, $171 \pm 40$pN (n=627) for GT16A, $183 \pm 42$pN (n=608) for GT25A, $178 \pm 38$pN (n=2106) for GK28A, $154 \pm 34$pN (n=1111) for GF30L, $175 \pm 43$pN (n=2002) for GK31A, $179 \pm 40$pN (n=1461) for GN35A, $198 \pm 44$pN (n=489) for GD46A, $151 \pm 40$pN (n=1479) for GT53A, and $180 \pm 36$pN (n=643) for wt-GB1.

2.3.4 All GB1 mutants follow the identical unfolding pathway

To further investigate how the point mutations affect the mechanical unfolding rate constant at zero force and unfolding distance from the native state to the transition state of GB1, we carried out force-extension measurements of GB1 at different pulling speeds ranging from 50-4000 nm s$^{-1}$. Since the mechanical unfolding
is carried out under non-equilibrium condition, the slower the pulling speed is, the lower the force required to unfold GB1. The average unfolding force is thus dependent on the pulling rate (Figure 2.4). The speed dependence of the unfolding force of all GB1 mutants has slopes similar to that for wt-GB1, which means that all the mutants have similar $\Delta x_u$ as wt-GB1. The unfolding events for each of the heteropolyproteins obtained from the Monte Carlo simulations can adequately fit into their corresponding histograms using the following parameters: unfolding rate constant at zero force, $k_u^0$, 0.0314 s$^{-1}$ for L7A, 0.0340 s$^{-1}$ for T11A, 0.0610 s$^{-1}$ for T16A, 0.0440 s$^{-1}$ for T25A, 0.1130 s$^{-1}$ for F30L, 0.0260 s$^{-1}$ for D46A, 0.1190 s$^{-1}$ for T53A, and 0.0460 s$^{-1}$ for wt-GB1; all mutants have the same unfolding distance $\Delta x_u = 0.17$ nm as wt-GB1$^{[30]}$.

![Figure 2.4](image)

**Figure 2.4** Stretching a single polyprotein at different pulling speeds. The average unfolding forces of wt-GB1 and all GB1 mutant heteropolyproteins increase as the pulling speeds increase. The measured unfolding forces at high pulling speed (above 2000 nms$^{-1}$) are corrected for the hydrodynamic drag force that applies directly on the AFM cantilever. Pulling speed experiment shows that there is no change in slope of all GB1 mutants and wt-GB1. To quantify these data, we used Monte Carlo simulations to estimate the $\Delta x_u$ and $k_u^0$. The black solid line is the fit of the Monte Carlo simulations.
Carlo simulation to the wt-GB1 experimental data, which is indicated by black hollow circle.

The point mutations of GB1, F30L and T53A, reduce the force required to unfold the domains and accelerate the unfolding rate constants. Mutants F30L on the helix and T53A on the fourth $\beta$-strand unfold around 3 times faster than the wt-GB1 and rest of the GB1 mutants. Such destabilization effects may be caused by the removal of any specific interactions such as hydrogen bonding, hydrophobic or side-chain interactions. The unfolding rate constants of these two mutants measured by the Monte Carlo simulations also indicate that the free energy barrier of unfolding is changed by the Leu or Ala substitution from Phe and Thr residues, respectively.

2.3.5 Refolding probability depends exponentially on the amount of time that the domain remains relaxed

By using a double-pulse protocol, we measured the folding kinetics of GB1 mutants (Figure 2.5(left)). A first extension of the protein allows counting of the available folded domains in the polyprotein chain, $N_{\text{total}}$. The total extension of the protein is limited to prevent the detachment of the polyprotein from the AFM tip or the glass slide. The protein quickly relaxes back to its initial length. After a relaxation time, $\Delta t$, the protein is stretched again, and several force peaks are observed, $N_{\text{refolded}}$. By plotting the refolded fraction, $N_{\text{refolded}}/N_{\text{total}}$, versus the relaxation time, we found that the $N_{\text{refolded}}/N_{\text{total}}$ depends exponentially on the amount of the time that the polyprotein remains relaxed, and the data can be described as a first-order reaction:

$$ P_f(t) = 1 - e^{-\lambda \Delta t} $$

(a)
where \( P_f(t) \) is the folding probability, \( \Delta t \) is the delayed time interval, \( k_f^0 \) is the folding rate constant. Figure 5 shows one of the GB1 mutant examples, GD46A. Fitting equation (a) to the experimental data measures a folding rate constant of GD46A \( k_f^0 = 1.9 \text{ s}^{-1} \) (Figure 2.5). D46A mutant refolds ~50 times slower than the wt-GB1.

**Figure 2.5** The folding kinetics of GB1 mutants: (Left) The folding kinetics of (GD46A)_8 polyprotein is probed by AFM using a double-pulse protocol. The polyprotein is first stretched to count the number of domains in the chain that unfold \( (N_{\text{total}}) \), and then the unfolded polyprotein is quickly relaxed back to its initial length. After a period of relaxation time, \( \Delta t \), the protein is stretched again to measure the number of domains that have refolded during relaxation, \( N_{\text{refolded}} \). (Right) Plot of the fraction of refolded domains, \( N_{\text{refolded}}/N_{\text{total}} \) versus \( \Delta t \). GD46A refolds at a much slower rate in comparison to wt-GB1. The solid lines are the fits of the data to the single exponential function: equation (a), where \( k_f^0 = 1.9 \text{ s}^{-1} \). GD46A mutant and wt-GB1 proteins are said to be a simple two-state model for the folding reactions.
2.4 Discussion

2.4.1 Calculation of the $\Phi$- value

$\Phi$- value analysis is a protein engineering method used to study the structure of the unfolding or folding transition state in protein domains that unfold in a two-state mechanism. The objective of this analysis is to evaluate the importance of the mutation sites on the effects of kinetic and equilibria of protein folding/unfolding as well as to map out which portions of the protein contribute to the stabilization of the transition state. This analysis assumes a close relationship between structure and energy. A point mutation in the protein is engineered and eventually will cause a difference in stability between mutants and wild-types. It corresponds to the changes in the height of the energy barrier between the folded state and the transition state along the reaction coordinate. Unfolding rate constant at zero force $k_u^0$ is obtained from unfolding kinetic and speed dependence experiment fitted with Monte Carlo simulations. By comparing the $k_u^0$ of GB1 mutant heteropolyprotein with that of the wt-GB1, the change in the free energy barrier of mechanical unfolding upon mutation between the native state and the transition state, $\Delta \Delta G_{N-TS}$ can be calculated as the following:

$$\Delta \Delta G_{N-TS} = -RT \ln \left( \frac{k_u^{0\text{mut}}}{k_u^{0\text{wt}}} \right)$$

In the same manner, the change in the free energy barrier of mechanical folding upon mutation between the unfolded state and the transition state, $\Delta \Delta G_{TS-U}$ can be calculated by comparing the values of $k_f^0$ between GB1 mutants and wt-GB1, of the heteropolyprotein.

$$\Delta \Delta G_{TS-U} = -RT \ln \left( \frac{k_f^{0\text{mut}}}{k_f^{0\text{wt}}} \right)$$

The total change in the free energy caused by the mutation is $\Delta \Delta G_{N-U}$. When $\Phi$- value
analysis combines with single molecule AFM measurements, it is used to probe the structure of the transition state and the mechanical stability during the unfolding process. The $\Phi$-value analysis for unfolding is defined as the ratio of the energetic destabilization introduced by the mutation to the transition state versus that introduced to the native folded state:

\[
\Phi(u) = \frac{\Delta \Delta G_{N-TS}}{\Delta \Delta G_{N-U}} \quad (5)
\]

When $\Phi(u) = 1$, it indicates that the residue being mutated is as unstructured in the transition state as in the unfolded state, whereas when $\Phi(u) = 0$, it means that the degree to which the transition state is destabilized by the mutation is equal to the degree to which the folded state is destabilized, indicating that the residue being mutated on the protein has no effect on the unfolding rate but slows down the folding rate (Figure 2.6). It implies that the residue is as structured in the transition state as in the folded state. Fractional $\Phi$-values indicate the native-state interactions of a point mutation are only partially formed in the transition state or show a multiple folding/unfolding pathways. A $\Phi$-value closer to 1 indicates that the region surrounding the mutation is relatively unfolded or unstructured in the transition state, while a $\Phi$-value closer to 0 suggests that the local structure around the mutation site in the transition state closely resembles the structure in the native state.
2.4.2 Mutations F30L and T53A of GB1 destabilize the transition state of unfolding pathway from mechanical studies

According to the AFM unfolding kinetic data, mutants F30L and T53A of GB1 show the changes in unfolding free energy barrier to these in the wt-GB1, while unfolding free energy barriers of other GB1 mutants remain unchanged when compared to that of the wt-GB1. The position of the transition state of all GB1 mutants along the reaction coordinate remains unchanged according to the pulling rate dependence experiment. Φ-values are calculated by using equation (5), shown in table 2.1. The single point mutations at L7A, T11A, T16A, T25A, K28A, K31A, N35A and D46A of GB1 do not affect the mechanical unfolding force and unfolding free energy barrier significantly, resulting in Φ-values of closer to 0, which suggests that each of these point mutations is assumed to be fully structured in the transition state. Each of these point mutations is located at different regions of GB1: L7A- first β-strand;
T11A-first β-turn; T16A-second β-strand; T25A-helix; K28A-centre of the helix; K31A and N35A- C-terminal portion of the helix; D46A-second β-turn. Φ-values of near 0 from these point mutations reflect that these regions appear to be largely formed in the unfolding transition state. The amino acid substitution at F30L and T53A of GB1 reduces the mechanical unfolding force and accelerates the unfolding rate, resulting in a relatively higher Φ-value, which indicates that each of the amino acid substitutions in a region is considerably unstructured in the unfolding transition state. Amino acid substitution from Phe to Leu at residue 30 of GB1 has relatively high Φ-value, which can be explained by the removing of buried interactions between the helix and the first β-hairpin. Since Phe is pointing towards the first and second β-hairpin of GB1, substitution with Leu may remove some buried interactions and have huge impact on the mechanical stability. Amino acid substitution from Thr to Ala at residue 53 of GB1 is situated on the fourth β-hairpin also has a relatively higher Φ-value. Since Thr is a polar residue, substitution with Ala may eliminate the hydrogen bonds interacting with solvents or other part of the protein and cause destabilization. Moreover, the destabilizing behaviour of mutant T53A of GB1 may also be attributed to the removal of the side chain interactions in the hairpin closest to the turn. Hydrogen bonding between the terminal β-strands is not the only interaction that is significant to mechanical stability. Buried interactions and side chain interactions play an important role as well. Mechanical unfolding force, Δxu, k′u, and chemical and mechanical Φ(u)-values are tabulated in Table 2.1.
### Table 2.1 Comparison of mechanical and chemical unfolding data of GB1 mutant

<table>
<thead>
<tr>
<th></th>
<th>Mechanical unfolding force (pN)</th>
<th>Histogram Band Width (pN)</th>
<th>Mech. $k_u$ (s⁻¹)</th>
<th>Mech. $k_f$ (s⁻¹)</th>
<th>Mech. $\Delta \Delta G_{N-TS}$ (kcal/mol)</th>
<th>Chem. $\Delta \Delta G_{N-U}$ (kcal/mol)</th>
<th>$\Delta x_u$ (nm)</th>
<th>Mech. $\Phi(u)$</th>
<th>Chem. $\Phi(u)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L7A</td>
<td>190</td>
<td>40</td>
<td>0.0314</td>
<td>N/A</td>
<td>-0.23</td>
<td>1.85</td>
<td>0.17</td>
<td>-0.12</td>
<td>0.68</td>
</tr>
<tr>
<td>T11A</td>
<td>188</td>
<td>45</td>
<td>0.0400</td>
<td>N/A</td>
<td>-0.08</td>
<td>0.60</td>
<td>0.17</td>
<td>-0.13</td>
<td>0.98</td>
</tr>
<tr>
<td>T16A</td>
<td>171</td>
<td>40</td>
<td>0.0610</td>
<td>N/A</td>
<td>0.17</td>
<td>0.38</td>
<td>0.17</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>T25A</td>
<td>183</td>
<td>42</td>
<td>0.0440</td>
<td>N/A</td>
<td>-0.03</td>
<td>-0.22</td>
<td>0.17</td>
<td>0.14</td>
<td>1.81</td>
</tr>
<tr>
<td>K28A</td>
<td>178</td>
<td>38</td>
<td>0.0500</td>
<td>N/A</td>
<td>-0.05</td>
<td>N/A</td>
<td>0.17</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>F30L</td>
<td>154</td>
<td>34</td>
<td>0.1130</td>
<td>N/A</td>
<td>0.53</td>
<td>1.42</td>
<td>0.17</td>
<td>0.37</td>
<td>0.95</td>
</tr>
<tr>
<td>K31A</td>
<td>175</td>
<td>43</td>
<td>0.0540</td>
<td>N/A</td>
<td>-0.09</td>
<td>N/A</td>
<td>0.17</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>N35A</td>
<td>179</td>
<td>40</td>
<td>0.0480</td>
<td>N/A</td>
<td>-0.03</td>
<td>N/A</td>
<td>0.17</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>D46A</td>
<td>198</td>
<td>44</td>
<td>0.0260</td>
<td>1.9</td>
<td>-0.34</td>
<td>1.74</td>
<td>0.17</td>
<td>-0.20</td>
<td>0.04</td>
</tr>
<tr>
<td>T53A</td>
<td>151</td>
<td>40</td>
<td>0.1190</td>
<td>N/A</td>
<td>0.56</td>
<td>1.91</td>
<td>0.17</td>
<td>0.29</td>
<td>0.73</td>
</tr>
<tr>
<td>WT</td>
<td>180</td>
<td>36</td>
<td>0.0460</td>
<td>720</td>
<td>N/A</td>
<td>N/A</td>
<td>0.17</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### 2.4.3 Mechanical and chemical $\Phi$-values compared

We compared the $\Phi(u)$-values distributions of GB1 from both mechanical and chemical experimental results. T11A, T16A and F30L of GB1 mutants have high (near 1) $\Phi$-values from chemical folding kinetics. These three mutation sites affect the structure of the unfolding transition state of GB1 significantly, and its high $\Phi$-values suggest that the first $\beta$-turn (T11A), second $\beta$-strand (T16A) and the centre of helix (F30L) are highly unstructured at the unfolding transition state. Results show that L7A and T53A of GB1 mutants have intermediate $\Phi$-values from the chemical folding.
kinetics. L7A is situated in the first $\beta$-strand and pointed towards the core of GB1. Substitution Leu to Ala at residue 7 may have removed interactions with the second $\beta$-hairpin. While T53A is located at the centre of the fourth $\beta$-hairpin, this point mutation may remove interactions between the adjacent hairpins. Amino acid substitution from Asp to Ala at residue 46 removes a hydrogen bond between one of the side chain oxygen atoms and the amino acid residue of 48, it slows down the refolding rate and has no effect on the unfolding rate. D46A of GB1 mutant gives $\Phi$-values of near 0 from chemical folding kinetics. Since GD46A is located at the second $\beta$-turn, it illustrates that the second $\beta$-turn is highly structured in the unfolding transition state. T25A located on the helix may have resulted in removing buried interaction between helix and the first $\beta$-hairpin has $\Phi$-value of near 2$^{17}$. From chemical studies, the second $\beta$-turn is assumed to be highly structured in the unfolding transition state. Point mutations on the first and fourth $\beta$-strands have intermediate $\Phi$-values, indicating partly formed sites of these strands in the unfolding transition state, while point mutation on the first $\beta$-turn, second $\beta$-strand and the centre of the helix may not affect the unfolding transition state. From the mechanical studies, the first $\beta$-strand, first $\beta$-turn, beginning of the helix and second $\beta$-turn appear to be intact in the unfolding transition state, while the centre of the helix (F30L) and the fourth $\beta$-hairpin (T53A) appears to be relatively disordered. T16A located on the second $\beta$-strand appears to be relatively unstructured at the unfolding transition state, according to the $\Phi$-value analysis. $\Delta\Delta G_{N,U}$ must be sufficiently large enough so the $\Phi$-values can be determined with confidence. $\Delta\Delta G_{N,U}$ for T16A and T25A are too small$^{17}$. Thus, the $\Phi$-values for both chemical and mechanical experiments might not be accurate enough.
2.5 Conclusion

Point mutations in proteins may affect their chemical and mechanical stability. The mechanical stability of proteins can be measured by single-molecule AFM. The saw-tooth patterns of all the GB1 mutants are well described by WLC model. All GB1 mutants show identical unfolding pathways as the wild type GB1, indicating by the identical unfolding distance $\Delta x_u$. The unfolding process of these mutants has a two-state manner, from the folded state to the unfolded state. Combining $\Phi$-value analysis and AFM measurements together with protein engineering techniques gives a complete picture on how proteins are structured in the transition state during folding or unfolding event. According to our AFM results, F30L and T53A of GB1 mutants resulted in a lower mechanical unfolding force than wt-GB1 and have near one $\Phi$-value, which demonstrate that the transition states of the two mutants are unstructured during unfolding process. On the other hand, the rest of the GB1 mutants show near zero $\Phi$-value, which indicates that the transition states of these mutants are highly structured during folding process. Thus, these GB1 mutants show similar unfolding forces as wild type GB1 and have no effect on the unfolding rate. However, experimental errors can be high in measuring protein stability and folding/unfolding rates for the GB1 mutants and wt-GB1 during AFM measurements, which may lead to some errors associated with $\Phi$-value analysis. To minimize the errors on the experimental data and $\Phi$-value, more numbers of data points were collected during the experiments. By using the $\Phi$-value analysis, it clearly indicates that to which the mutation site has the strongest effect on the mechanical stability of GB1. Based on our $\Phi$-value analysis of GB1 mutants, it suggests that hydrogen bonding between the terminal $\beta$-strands may not be the main and only interaction that is important to mechanical stability. Some buried interactions and side chain interactions can play significant roles on the mechanical stability of the protein. From our experimental
data, none of these point mutations enhance the mechanical stability. However, GD46A may be an exception. By combining the protein engineering techniques with single-molecule AFM, it is possible to find out to which single point mutation could possibly increase the mechanical stability of the proteins.
2.6 References


(21) Ng, S. P.; Rounsevell, R. W.; Steward, A.; Geierhaas, C. D.; Williams, P. M.;


(38) Brockwell, D. J.; Beddard, G. S.; Clarkson, J.; Zinober, R. C.; Blake, A. W.;


Chapter 3: Configurational entropy modulates the mechanical stability of GB1

3.1 Introduction

Loop regions of proteins have been described as a ‘random coil’ even if they adopt to form well-defined conformations found in most globular proteins. Loop regions in proteins play important roles in connecting secondary structural elements, \( \alpha \) helices and \( \beta \) strands, as well as forming binding sites and enzyme active sites. Length of the flexible loop significantly influences the thermodynamic stability and the folding/unfolding kinetics of proteins. Such influence can be explained by the entropic cost associated with loop closure in proteins using classical polymer theory. The effect of configurational entropy on the thermodynamic stability as well as chemical folding/unfolding kinetics of proteins with different length of flexible loops has been widely studied by using ensemble methods. For example, the transition state and energy barrier of \( \alpha \)-spectrin SH3 mutants had been investigated by using a stopped-flow instrument; insertion of 10 Gly or Gln residues into the long reactive loop of chymotrypsin inhibitor-2 (CI2) was used to study the effect of the sequence on protein stability and on the folding pathways by the model system. It is of note that the long insertion into the loop of a folded protein could have destabilizing effects on thermodynamic stability and often slows down folding reactions while accelerating the unfolding reactions. However, the effect of loop length on protein mechanical folding and unfolding dynamics still remains unknown.

Application of protein engineering methods in combination with single-molecule atomic force microscopy (AFM) technique is an efficient way to provide information on studying protein mechanical folding and unfolding dynamics.

along a well-defined reaction coordinate defined by the stretching force and has been extensively used to investigate the molecular determinants of the mechanical stability of proteins.\textsuperscript{8-12, 13} Large number of studies on loop insertion in single-molecule protein mechanics techniques has been mainly focused on locating the structural motif that provides mechanical resistance or mapping the location of an unfolding intermediate state.\textsuperscript{14, 15} Here, we use a small protein, GB1, the B1 IgG binding domain of protein G from \textit{Streptococcus}, as a model system to directly investigate the effect of loop length on the mechanical stability and mechanical unfolding kinetics of GB1 by using single-molecule AFM. GB1 contains 56 amino acid residues and is a $\alpha/\beta$ protein composed of four strand $\beta$-sheet packed against a $\alpha$-helix.\textsuperscript{16} GB1 has high thermostability and contains no disulfide bonds or prosthetic groups. Its mechanical folding and unfolding dynamics have been studied in detailed using single-molecule AFM\textsuperscript{17-19}. Recent study\textsuperscript{16, 17-19} indicates that GB1 has excellent mechanical properties and is easy to manipulate by AFM, which makes it a good candidate for single-molecule force spectroscopy studies. The structure of GB1 shows two loops connecting the $\alpha$ helix to $\beta$ strands 2 and 3, respectively (Fig. 3.1). Using protein engineering techniques, we constructed four GB1 loop mutants, GB1-L2, GB1-L5, GB1-L24 and GB1-L46. Each mutant has various numbers of inserted amino acid residues (2, 5, 24 and 46 residues) into loop 2 of GB1, between residue 39 and 40, to elongate the flexible loop. The single-molecule AFM is used to investigate the effect of loop length on the mechanical stability and mechanical unfolding kinetics of GB1. L2, L5, L24 and L46 indicate the number of inserted amino acid residues on loop 2 of GB1. It was found that the mechanical stability of GB1 decreases and the mechanical unfolding kinetics of GB1 accelerate as the elongation of the second loop increases. The results reflect that the mechanical stability and unfolding kinetics of GB1 loop
mutants are directly associated with the loss of configurational entropy upon closing an unstructured loop region in proteins, telling us the loop region of protein plays critical roles in mechanical unfolding. It is also important to understand whether the loop formation is complete or not in the transition state compared with that in the native state, the effect of configurational entropy can be different. This can be determined from the free energy barriers of the native state to the transition state, which explains the effect of the loop length on the mechanical unfolding kinetics. We found that in our study, the loop is partially formed in the transition state, which means the loop length will affect the free energy barrier for mechanical unfolding by accelerating the unfolding kinetics. Our results not only demonstrate that the second loop extension of GB1 plays important roles on mechanical stability as well as mechanical unfolding kinetics but also indicate that the mechanical stability of proteins can be modulated by using configurational entropy.

3.2 Methods and materials

3.2.1 Protein engineering

Figure 3.1 Loop insertions in GB1. GB1 has a typical α/β structure with a four-strand β sheet packed against an α helix. There are two loops connecting the helix to strands 2 and 3, respectively. The second loop of GB1 (residue 37-41), connecting the α helix and β strand 3, considers as a long flexible loop. The flexible linker sequences are inserted between residues 39 and 40. The sequences of the flexible linkers are also
shown.

### 3.2.1.1 Construction of GB1-L2 gene

A nonpalindromic *Ava* I restriction site (CTC GGG) was inserted between the 39\textsuperscript{th} and 40\textsuperscript{th} amino acid residues of wt-GB1 by using the standard protocols of site-directed mutagenesis, which gives GB1(*Ava* I)-pUC19. The PCR products were purified and confirmed by restriction digestion followed by DNA sequencing. The insertion of the *Ava* I restriction site in the mutants results in two additional amino acid residues ‘LG’ on loop 2 of wt-GB1.

### 3.2.1.2 Construction of GB1-L5 gene

GB1-L5 genes were constructed by using Mega Primer PCR method. Template encoding GB1 residues was PCR amplified with a 5’ overhang primer of restriction site *Bam* HI (called G.up) and 3’ overhang primer of DNA encoding three Gly residues followed by restriction site *Ava* I. The amplified PCR product was gel purified to acquire the desire PCR mega primer. The template encoding GB1 residues was PCR amplified with the mega primer. The latest amplified PCR product was gel purified and digested with restriction enzymes *Bam* HI and *Ava* I and subcloned into pUC19-GB1(*Ava* I) digested with *Bam* HI and *Ava* I restriction enzymes to produce pUC19-GB1-L5, which was confirmed by DNA sequencing. The PCR mega primer method results in the insertion of five residues, GGGLG, between 39\textsuperscript{th} and 40\textsuperscript{th} amino acid residues of wt-GB1.
3.2.1.3 Construction of GB1-L24 and GB1-L46 gene

In order to construct mutants GB1-L24 and GB1-L46, the flexible linker sequence (GSA)$_6$GS flanked with AvaI restriction site at both ends was used and synthesized by Integrated DNA technologies. The resulting products were digested with restriction enzyme AvaI and ligated into pUC19-GB1(AvaI) also digested with restriction enzyme AvaI. During the ligation, the flexible linkers can ligate into the digested vector, as well as self-ligate to produce ring structure concatemers. The ligation products were screened for plasmids containing one or two repeats of (GSA)$_6$GS by digesting purified plasmids with restriction enzymes BamHI and KpnI, which results in pUC19-GB1-L24, containing one repeat of the flexible linker and pUC19-GB1-L46, containing two repeats of the flexible linkers. DNA sequencing results confirm the insertion of 24 residues, LG(GSA)$_6$GSLG or of 46 residues, LG(GSA)$_6$GSLG(GSA)$_6$GSLG between the 39$^{th}$ and 40$^{th}$ amino acid residues of wt-GB1, respectively. See Appendix F.

3.2.1.4 Polyprotein construction

The gene encoding the GB1 loop mutants from PCR is in the plasmids pUC19. GB1 is flanked by a 5’ BamHI site and a 3’ BglII site followed by KpnI site. Eight-domains polyprotein, (GB1-L2)$_8$, (GB1-L5)$_8$, (GB1-L24)$_8$, and (GB1-L46)$_8$ were constructed by iterative cloning of monomer into monomer, dimer into dimer, and tetramer into tetramer based on the identity of the sticky ends generated by the BamHI and BglII restriction sites. 8mer is subcloned into expression vector pQE80L, which contains an N-terminal (His)$_6$ purification tag to facilitate the purification of expressed proteins and then transformed into a recombinase- defective E. coli cloning strain, DH5$\alpha$. The protein is then purified by Ni-NTA affinity chromatography, which
binds the (His)$_6$ tag polyproteins. The sizes of GB1 mutant polyproteins were confirmed by SDS-PAGE, see Appendix G.

### 3.2.2 Single-molecule AFM experiment

Single-molecule AFM experiments were carried out in a custom-built atomic force microscope, which was constructed as described $^{17-19}$. All of the experiment measurements for GB1 loop mutant polyproteins were carried out in phosphate-buffered saline (PBS) solution. ~1µL protein sample was deposited onto a freshly clean glass cover slip covered by ~50 µL PBS solution and was allowed to adsorb for approximately 10 min. The spring constant of each individual cantilever ($\text{Si}_3\text{N}_4$ cantilevers from Veeco, with a typical spring constant of 40 pN nm$^{-1}$) was calibrated in PBS solution using the equipartition theorem before each experiment with an average error of 10%. Single-molecule AFM pulling experiments were then performed. The pulling speed for all the unfolding experiments was 400 nm s$^{-1}$, except when reported otherwise.

### 3.2.3 Monte Carlo simulations

Monte Carlo simulations of the stretching of the GB1 loop mutants were carried out using standard procedures as described previously $^{27, 39, \text{appendix H}}$. The mechanical unfolding of the polyproteins was described as a two-state Markovian process with force-dependent rate constant. Both unfolding rate constant at zero force $\alpha_0$ and the distance of the native state to the transition state $\Delta x_u$ along the reaction coordinate of the mechanical unfolding reaction were estimated. The minimal detectable unfolding force in Monte Carlo simulation was set to 30 pN in order to reflect the noise level of the unfolding force measurements in the single-molecule AFM experiments.
3.2.4 Circular dichroism measurements

CD was performed on a JASCO-J810 spectropolarimeter flushed with nitrogen gas. The protein samples were measured in 0.1 M PBS at pH 7.4. An average of three scans was carried out for each protein sample. The CD spectra were recorded in a 0.2-cm path-length cuvette at a scan rate of 50 nm/min.

3.3 Results

3.3.1 The second loop has a high tolerance of loop elongation without affecting GB1’s native structure

There are two loops in GB1 connecting the α helix and β strands 2 and 3. First loop and second loop of GB1 consist of 3 and 5 residues, respectively. First loop of GB1 (residue 20-22), connecting β strand 2 to the α helix, considers as a short, rigid and less flexible loop, while the second loop of GB1 (residue 37-41), connecting the α helix and β strand 3, considers as a long flexible loop. Here, we used the second loop of GB1 to investigate the effect of loop length on the mechanical stability and mechanical unfolding kinetics of GB1. Various numbers of flexible amino acid residues were inserted into the loop 2 of GB1 between residue 39 and 40 to obtain GB1 loop mutants GB1-L2, GB1-L5, GB1-L24 and GB1-L46, which elongate loop 2 of GB1 by insertion of 2, 5, 24 and 46 amino acid residues, respectively. GB1-L2 has two residues insertion, LG in loop 2; GB1-L5 has a 5- residues flexible linker insertion, GGGLG, in loop 2; GB1-L24 and GB1-L46 have flexible linkers of Gly-Ser-Ala (GSA) repeats insertion in loop 2²⁰. Glycine and alanine residues within the flexible linkers consider to be the smaller residues, which provide the flexibility to avoid the formation of residue structure in the loop region, while serine residues
increase the solubility and reduce the aggregation. Far-ultraviolet circular dichroism (far-UV CD) spectroscopy is an excellent tool to investigate the structural integrity of the loop insertion mutants. CD spectroscopy characterizes and gives rise to distinct CD spectra for the secondary structures of protein, α helix, β sheet and random coil. The CD spectrum of α helix is characterized by two negative bands at ~222 and 208 nm, while the CD spectrum of β sheets gives rise to one negative band at ~215 nm. The CD spectrum of random coils shows a negative band at ~195 nm. The CD spectrum of wild-type GB1 (wt-GB1) and GB1 loop mutants is shown in Figure 3.2. Wt-GB1 shows two broad negative bands with minima at 208 and 222 nm, which indicate the compact structure of GB1 with the existence of the secondary structures.

As the number of flexible residue insertion in loop 2 increases, the two minima are still clearly visible in the CD spectra for the four GB1 loop mutants, indicating that the compact structures of GB1 with the second structures retain well in the mutants. The minima at 222 nm in the CD spectrum remain unchanged upon the insertion of various numbers of residues into loop 2, although their relative amplitude decrease compared with that of wt-GB1, suggesting that the content of the secondary structures within the compact structure of GB1 has decreased as the number of flexible residue insertion increases. On the other hand, the minima at 208 nm have gradually shifted towards the lower wavelength with increasing number of residues being inserted into loop 2, indicating the gradual increase content of random coil structures in loop insertion mutants since the CD spectrum of random coil is characterized by a negative band at ~195 nm. These results clearly indicate that the second loop is tolerant of loop elongation without affecting GB1’s native structure. Therefore, these GB1 loop mutants provide good model systems to investigate the effect of loop length on the mechanical stability of proteins.
3.2 Loop elongation decreases the mechanical stability of GB1 and accelerates the mechanical unfolding kinetics

Four polyproteins, (GB1-L2)$_8$, (GB1-L5)$_8$, (GB1-L24)$_8$ and (GB1-L46)$_8$ were constructed to study the mechanical stability of GB1 loop mutants. All these mutants are composed of eight identical tandem repeats of the GB1 loop mutants. Stretching the polyproteins gives rise to the sequential unravelling of the individual GB1 loop mutant domains, which resulted in force-extension curves with characteristic sawtooth pattern appearances, shown in Figure 3.3. Each individual force peak corresponds to
the mechanical unfolding of the individual GB1 loop domains in the polyprotein chain. The last peak indicates the stretching and the subsequent detachment of the completely unfolded polyprotein chain from the cantilever tip or the coverslip substrate. Worm-like chain (WLC) model of polymer elasticity\textsuperscript{23} fits to the consecutive unfolding force events reveals the contour length increment ($\Delta L_c$) of the loop insertion mutants increases as the inserted number of amino acid residues increases. According to the force-extension curves of wt-GB1 and GB1 loop mutants, it is clear to see that the spacing between the consecutive unfolding events of the loop insertion mutants increases (the amino acid residues inserted in loop 2 increases), while the unfolding force of each loop mutant decreases. The force-extension curve of wt-GB1 shown in Figure 3.3 is fitted by the WLC model of polymer elasticity to give the contour length increment ($\Delta L_c$) of $\sim 18$ nm and the unfolding force of $\sim 180$ pN\textsuperscript{17}, whereas the insertion of 46 residues into loop 2 results in a $\Delta L_c$ of $\sim 35$ nm, and the unfolding force is decreased to $\sim 63$ pN. $\Delta L_c \sim 35$ nm of mutant GB1-L46 is measured from subtraction of the folded contour length of GB1 from the total unfolded and fully extended length of all the residues. Folded contour length corresponds to the distance between the N- and C- termini of the domain. Since the insertion of 46 amino acid residues into loop 2 of GB1 does not affect the proper folding of GB1, the distance between the N- and C- termini of the mutant (GB1-L46) is $2.6$ nm\textsuperscript{16}, which is identical to wt-GB1. GB1-L46 is 102 residues long, thus the length of the unfolded and fully extended of all the residues is $(102\times0.36)$ 36.72 nm, assuming the length of a single amino acid residue is 0.36 nm\textsuperscript{24}. The complete unfolding of GB1-L46 results in $\Delta L_c$ of $\sim 34.12$ nm (36.72-2.6), which is in a good agreement with the experimentally observed $\Delta L_c$ for GB1-L46. This result also confirms the finding of the contour length increment in the force-extension curves of other GB1 loop mutants, shown in Figure 3 and 4A.
**Figure 3.3** Force-extension relationships of GB1 loop mutants. Stretching GB1 loop mutants gives a characteristic sawtooth-like force-extension curve. Each individual force peaks represents the mechanical unfolding of the individual mutant domains. The mechanical unfolding forces of GB1 loop mutants decrease as the number of the amino acid residues inserted in the loop 2 increases. WLC fits (red lines) to the consecutive unfolding force peaks reveal the contour length increment of loop insertion mutants increases as a function of the inserted number of amino acid residues.

Mechanical clamp of GB1 is composed of the force-bearing β strands 1 and 4 according to the molecular dynamics simulation. Mechanical clamp provides the mechanical resistance to extending and unravelling of GB1. By disrupting the mechanical clamp, the domain will unfold, and the hidden amino acids in the core will expose to the mechanical forces and give rise to the contour length increment ($\Delta L_c$). If additional amino acids are inserted after the mechanical clamp, mechanical unfolding
will result in the increase in $\Delta L_c$. Since loop 2 is located after the mechanical clamp of GB1, an increase in $\Delta L_c$ for the GB1 loop mutants would be observed. The number of amino acids inserted in loop 2 increases as the $\Delta L_c$ of the GB1 loop mutant increases, shown in Figure 3.4A.

**Figure 3.4** (A) Histograms for $\Delta L_c$ of loop insertion mutants. The average of the contour length increment is 18.3±0.5 nm ($n=470$) for wt-GB1, 18.5±0.7 nm ($n=415$) for GB1-L2, 19.8±0.7 nm ($n=307$) for GB1-L5, 26.9±1.1 nm ($n=796$) for GB1-L24 and 35.0±1.0 nm ($n=595$) for GB1-L46. (B) Histograms for the mechanical unfolding force of loop insertion mutants. The average of the unfolding force is 177±40 pN
(n=2593) for wt-GB1, 121±28 pN (n=1483) for GB1-L2, 134±36 pN (n=1209) for GB1-L5, 80±35 pN (n=659) for GB1-L24 and 63±31 pN (n=595) for GB1-L46. Red lines are Monte Carlo fits to the experimental data using the following unfolding distance Δx_u and unfolding rate constant α_0 values: 0.17 nm and 0.039 s\(^{-1}\) for wt-GB1; 0.22 nm and 0.075 s\(^{-1}\) for GB1-L2; 0.19 nm and 0.090 s\(^{-1}\) for GB1-L5; 0.19 nm and 0.40 s\(^{-1}\) for GB1-L24; and 0.19 nm and 0.62 s\(^{-1}\) for GB1-L46. The detection limit measured for the unfolding force in the AFM is ~20pN.

According to Figure 3.4B, the mechanical stability of GB1 loop mutants decreases progressively as the function of the loop length. The unfolding force histograms for the GB1 loop mutants show significant decrease from 180 pN for the wt-GB1 to 63 pN for the (GB1-L46) mutant. A plot of the unfolding force versus the linker length is shown in Figure 3.5, which describes the mechanical stability decreases in a nonlinear fashion as a function of the inserted loop length. As the loop length increases further, the decreasing trend tends to weaken. It is worth noting that (GB1-L2) is an exception, as its unfolding force is clearly lower than that of (GB1-L5) mutants. These results prove that the loop insertion has a significant influence on the mechanical stability of proteins.
Figure 3.5 Mechanical stability shows nonlinear dependence on the number of amino acid residues inserted in loop 2. The effect of mechanical destabilization and acceleration of unfolding kinetics tends to attenuate with further increase of the number of amino acid residues in loop 2. Red line is an exponential fit to the experimental data to illustrate the trend how the unfolding force changes as a function of linker length.

The unfolding force of a protein is determined by two parameters: the height of the free energy barrier $\Delta G_{TS-N}$ for mechanical unfolding and the unfolding distance $\Delta x_u$ between the folded state (N) and the unfolding transition state. Both $\Delta G_{TS-N}$ and $\Delta x_u$ can be determined from the unfolding force histogram indirectly and directly, respectively. The spontaneous unfolding rate constant at zero force $\alpha_0$, which is directly related to $\Delta G_{TS-N}$, can be measured from the unfolding force histogram. The width of the unfolding force histogram is directly related to $\Delta x_u^{26}$. It is clear to see that the width of the unfolding force histogram for GB1-L2 is much narrower than those for wt-GB1, whereas the width for the other three mutants GB1-L5, GB1-L24 and GB1-L46 remain unchanged when compared with wt-GB1. In order to further confirm and estimate $\alpha_0$ and $\Delta x_u$ accurately, pulling speed experiment was carried out.
on these GB1 loop mutant proteins, shown in Figure 3.6. The higher the pulling speed is, the higher the force required to unfold the GB1 loop mutant proteins. The speed dependence of the unfolding force of (GB1-L5), (GB1-L24) and (GB1-L46) has a slope similar to that for wt-GB1, while the slope of the speed dependence of the unfolding force of (GB1-L2) is clearly different from that of wt-GB1, as well as other GB1 loop mutants. It is evident that there is a change in $\Delta x_u$ for GB1-L2, which is consistent with the conclusion drawn from the unfolding force histogram. Using standard Monte Carlo simulation procedures$^{27,28,39}$, we reproduced force-extension curves of GB1 loop mutants. By fitting the unfolding force histograms and the pulling speed experiments simultaneously, we have estimated $\alpha_0$ and $\Delta x_u$ for individual mutant proteins. $\Delta x_u$ is estimated to be 0.19nm for GB1-L5, GB1-L24 and GB1-L46, which is slightly bigger than that of wt-GB1, estimated to be 0.17nm$^{17}$, whereas $\Delta x_u$ for GB1-L2 is estimated to be 0.22nm, which is significantly bigger than that of wt-GB1. The small values of $\Delta x_u$ for GB1-L5, GB1-L24, GB1-L46 and wt-GB1 suggest that the mechanical unfolding transition state is highly native-like.
Figure 3.6 Speed dependence of the mechanical unfolding forces of wt-GB1 and GB1 loop mutants. The average unfolding forces of GB1 loop mutants and wt-GB1 increase with increase of the pulling speed. Error bars indicate the standard deviation of unfolding force. Solid lines correspond to Monte Carlo fits to the experimental data using following unfolding distance, $\Delta x_u$ and unfolding rate constant at zero force, $\alpha_0$: 0.17 nm and 0.039 s$^{-1}$ for wt-GB1; 0.22 nm and 0.075 s$^{-1}$ for GB1-L2; 0.19nm and 0.090 s$^{-1}$ for GB1-L5; 0.19 nm and 0.40 s$^{-1}$ for GB1-L24 and 0.19 nm and 0.62 s$^{-1}$ for GB1-L46.

The difference of the unfolding kinetic energy barrier upon loop elongation, $\Delta \Delta G_{TS-N}$, can be obtained by the following equation:

$$\Delta \Delta G_{TS-N} = \Delta G_{TS-N(\text{mut})} - \Delta G_{TS-N(\text{wt})} = RT \ln \left( \frac{\alpha_0(\text{wt})}{\alpha_0(\text{mut})} \right)$$

where $\alpha_0 = \nu \kappa \exp(-\Delta G_{TS-N} / RT)$, in which $\nu$ is the vibrational frequency at the transition state. $\kappa$ is the transmission coefficient. R is the gas constant and T is the temperature. $\Delta G_{TS-N}$ indicates the free energy barrier of GB1 between the native state
to transition state. Since $\alpha_0$ of all GB1 loop mutants and wt-GB1 have been estimated from Monte Carlo simulation procedures, $\Delta \Delta G_{TS-N}$ can be acquired by substituting $\alpha_0$ into equation (1). Plotting $\Delta \Delta G_{TS-N}$ versus loop length allows us to investigate how the mechanical unfolding kinetic barrier changes with the increasing of loop length.

Figure 3.7 shows a nonlinear dependency relationship of $\Delta \Delta G_{TS-N}$ on the loop length and demonstrates the effect of loop length tends to attenuate with increasing number of amino acids in loop2. Based on these results, it is suggested that the effect of loop length on the mechanical stability of GB1 is associated significantly by reducing the free energy barrier for unfolding, rather than changing the mechanical unfolding distance or shifting the mechanical unfolding pathways. However, GB1-L2 seems to be an exception, in that, the effect of loop length on the mechanical stability involves in both the reducing of the free energy barrier and of shifting the mechanical unfolding pathway.

![Figure 3.7](image.png)

**Figure 3.7** The free energy barrier for mechanical unfolding shows a decreased nonlinear dependence with the length of the flexible linker.
Table 3.1 Mechanical experimental data for GB1 loop mutants

<table>
<thead>
<tr>
<th>Mutant</th>
<th>ΔLc (nm)</th>
<th>Unfolding force (pN)</th>
<th>Histogram Band Width</th>
<th>Δxu (nm)</th>
<th>α₀ (s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt-GB1</td>
<td>18.3 ± 0.5</td>
<td>177</td>
<td>40</td>
<td>0.17</td>
<td>0.039</td>
</tr>
<tr>
<td>GB1-L2</td>
<td>18.5 ± 0.7</td>
<td>121</td>
<td>28</td>
<td>0.22</td>
<td>0.075</td>
</tr>
<tr>
<td>GB1-L5</td>
<td>19.8 ± 0.7</td>
<td>134</td>
<td>36</td>
<td>0.19</td>
<td>0.090</td>
</tr>
<tr>
<td>GB1-L24</td>
<td>26.9 ± 1.1</td>
<td>80</td>
<td>35</td>
<td>0.19</td>
<td>0.40</td>
</tr>
<tr>
<td>GB1-L46</td>
<td>35.0 ± 1.0</td>
<td>63</td>
<td>31</td>
<td>0.19</td>
<td>0.62</td>
</tr>
</tbody>
</table>

3.4 Discussion

3.4.1 Configurational entropy plays important roles in mechanical unfolding kinetics of GB1

The unfolding force of a protein is determined by two characteristic parameters of the free energy diagram for the mechanical unfolding reaction: the free energy barrier for unfolding (ΔG_{TS-N}) and the unfolding distance Δxu between the native state and mechanical unfolding transition state. GB1-L5, L24 and L46 only shows small change of Δxu compared with wt-GB1, while Δxu of GB1-L2 is significantly larger than that for other GB1 loop mutants and wt-GB1. The AFM experimental results clearly show that GB1-L2 does not follow the nonlinear decrease trend as other GB1 loop mutants do. The effect of configurational entropy on the mechanical stability of GB1-L2 is thus the combined effect of increased unfolding rate constant at zero force
and unfolding distance. The reason behind the abnormality still remains unclear. One speculation would be a result from the use of the restriction site of AvaI, which leads to two amino acid residues insertion: leucine and glycine. Since leucine is relatively bulky and hydrophobic, it is possible that the inserted leucine in the loop 2 of GB1 interacts with other parts of the protein, leading to a change in the mechanical unfolding distance. Such an effect would probably be solved by obtaining a high-resolution structure of the particular mutant and steered molecular dynamics simulation of the mechanical unfolding process. The unfolding distance for other GB1 loop mutants does not change. This explains that the effect of configurational entropy on the mechanical stability is determined solely by the mechanical unfolding free energy barrier (the mechanical unfolding kinetics).

The insertion of the flexible linker into the second loop of GB1 produces destabilization effects in mechanical properties. Single-molecule AFM study in combination with protein engineering methods demonstrates that the loop insertion decreases the mechanical stability and accelerates the mechanical unfolding kinetics of GB1. Formation of loop regions in proteins is directly associated with the loss of configurational entropy. The change in free energy with loop length can be explained using classical polymer theory. The simply polymer model shows the change in configurational entropy of closing an unstructured loop of $L$ residues relative to the loop of $L_0$ residues can be estimated as followed:

$$
\Delta \Delta S_{N-U} = \Delta S_{N-U}(L) - \Delta S_{N-U}(L_0) = -C^* R \ln \left( \frac{L}{L_0} \right)
$$

(2)

where $L_0$ is the length of the reference loop, which is 5 for GB1 (there are five residues on the second loop of GB1), $L$ is the length of the loop after loop insertion, $C^*$ is a coefficient that is dependent upon the polymer used, and $R$ is the gas constant. Thus, the change in the thermodynamic stability for the two proteins after
loop insertion is:

$$\Delta \Delta G_{U-N} = \Delta G_{U-N}(L) - \Delta G_{U-N}(L_0) = -C * RT \ln \left( \frac{L}{L_0} \right)$$  \hspace{1cm} (3)$$

where $\Delta G_{U-N}(L_0)$ is the thermodynamic stability of the protein with a loop of $L_0$
residues, $\Delta G_{U-N}(L)$ is the thermodynamic stability of the protein after the loop is
elongated to $L$. $T$ is the temperature in Kelvin. Thus, loop insertion will considerably
increase the entropic cost of closing an unstructured loop, resulting in decreased
thermodynamic stability of the protein. A loop insertion approach potentially enables
the tuning of the mechanical stability of a given protein continuously in a
semi-predictable fashion according to the nonlinear dependence of the unfolding
forces on the number of amino acid residues inserted. In the future, if the
experimental $C^*$ is measured and obtained, it is possible to predict the mechanical
stability of the GB1 loop mutants (with any number of amino acid residues being
inserted into loop 2, eg. 15) by substituting values into equation (3). Once $C^*$ is
obtained, the effective loop length of the transition state before and after loop
insertion is also acquired by substituting values into equation (4).

As mentioned previously, the energy barrier for mechanical unfolding is one of
the two parameters that determine the unfolding force of a protein. The change of the
free energy barrier for mechanical unfolding between the folded and transition state of
the protein upon loop insertion depends on the change in the effective length of the
elongated loop in the unfolding reaction. The relationship can be written as:

$$\Delta \Delta G_{TS-N} = \Delta \Delta G_{TS-U} - \Delta \Delta G_{N-U} = C * RT \left[ \ln \left( \frac{L}{L_0} \right) - \ln \left( \frac{L}{L_0'} \right) \right]$$  \hspace{1cm} (4)$$

where $L_0$ and $L'$ correspond to the effective loop length of the transition state before
and after loop insertion, respectively, and $\Delta \Delta G_{TS-N}$ represents the change in the free
energy barrier for unfolding, which can be obtained solely from the experimental
mechanical unfolding rate constant by substituting these values into equation (1).

### 3.4.2 Loop 2 of GB1 is partially disrupted at the mechanical transition state.

The formation of the loop in the transition state is highly related to the free energy barrier for mechanical unfolding, which can be measured from the unfolding rate constant of proteins. Therefore, whether the loop formation is complete or not in the transition state in comparison with that in the native state would have a different influence on the effect of the configurational entropy. If the loop is fully formed in the mechanical unfolding transition state as in the native state, loop insertion will destabilize the transition state as the same amount as it does to the native state, leading to zero for $\Delta \Delta G_{TS,N}$. Thus, the free energy barrier of the folded (native) state and the transition state will not be affected by the loop insertion. On the other hand, if the loop is deformed partially in the mechanical unfolding transition state as compared to that in the native state, loop insertion will cause destabilization effect on the unfolding transition state than it does on the native state, leading to a negative $\Delta \Delta G_{TS,N}$. In this case, loop insertion will reduce the unfolding free energy barrier and accelerate the unfolding kinetics. According to our single-molecule AFM results, we found that the unfolding kinetics accelerates upon the loop elongation, which suggests that the formation of loop 2 in the mechanical unfolding transition state is partially deformed by the stretching forces. Therefore, loop 2 after the insertion in the mechanical unfolding transition state is not as formed as in the native state of GB1.

The study on the amino acid residue insertion into the second loop of GB1 indicates several important conclusions:

1. No conformational changes are observed by the flexible linker insertions. This has been confirmed by CD spectroscopy. The CD spectra still shows that the compact structures with both $\alpha$ and $\beta$ features are well retained in the loop mutants. Therefore, the second loop of GB1 is tolerant of loop elongation.
2. The loop elongation of GB1 accelerates the unfolding rates and decreases the protein mechanical stability as the number of inserted amino acid residues increases. There is a loss of the free energy barrier for mechanical unfolding, which explains the loss of the protein stability with loop elongation of GB1. The free energy for chemical unfolding also decreases as the loop elongation increases.

3. Configurational entropy plays important roles in the mechanical unfolding kinetics of GB1. Loop elongation does not change much on the unfolding distance, the effect of configurational entropy on the mechanical stability can be determined solely on the free energy barrier for mechanical unfolding.

This study provides the general knowledge to use configurational entropy as an effective means to tune the mechanical stability of proteins, especially for the loops in proteins. Based on our experimental data, it is now possible to predict the mechanical stability of GB1 loop mutant with different length of amino acid residues being inserted into loop 2.
3.5 References


Chapter 4: Summary and conclusion

4.1 Summary

The development of the single molecule mechanics and engineering has made significant progress since the first single protein study was revealed many years ago. Using single molecule force spectroscopy techniques, the mechanical properties of individual protein domains can now be examined and studied in great detail. The combination of single molecule AFM with protein engineering techniques has given us a more complete picture in investigating the molecular determinants of mechanical stability of proteins, as well as brought us a step closer towards tuning the mechanical properties of proteins in a rational fashion. These artificial elastomeric proteins, such as GB1, can be used as building blocks for the construction of biomaterials for applications ranging from material sciences to biomedical engineering.

Chapter 1 discusses the factors that affect the mechanical stability of proteins, as well as study and explain the main concepts and principles of AFM. In addition to these, the two-state unfolding process and worm-like chain models are discussed. Numbers of examples on the significant findings of single-molecule AFM of protein are also discussed.

Chapter 2 discusses that point mutation in proteins disrupts the intermolecular interactions and affects the chemical and mechanical stability of the protein with different degrees. \( \Phi \)-value analysis together with single-molecule atomic force microscopy is used to probe the mechanical stability of the protein and gives a complete picture on how proteins are structured in the transition state during folding and unfolding event. GB1 mutants were constructed. Single molecule AFM results reveal that the mutants F30L and T53A show lower unfolding forces in comparison with wild-type GB1 and accelerate their unfolding rate constants. Both mutants (F30L and T53A) have the relatively high \( \Phi \)-values, which demonstrate that the transition
states of the two mutants are unstructured during unfolding process, whereas other mutants show no significant changes on the unfolding forces and unfolding rate constants at zero force to the wt-GB1 and indicate relatively low Φ-values (around zero). All mutants show identical unfolding pathways as the wild type GB1 which reveals that the unfolding process of GB1 mutants follows the two-state manner. Point mutations on F30L and T53A of GB1 may have disrupted the hydrogen bonding between the terminal β-strands or unravelled the hydrophobic interactions and side chain interactions to give lower unfolding forces with relatively higher Φ-values.

Chapter 3 reveals that loop elongation into a folded protein is expected to destabilize the mechanical stability and accelerate the mechanical unfolding kinetics of the protein. This expectation can be explained by the loss of configurational entropy with closure of an unstructured flexible loop using classical polymer theory. Loop insertion of GB1 results in a decrease in thermodynamic stability of the protein. Four mutant polyproteins of loop elongations with two, five, twenty-four, and forty-six amino acid residues into the second loop of GB1 were constructed, respectively. These GB1 mutants were used to study the mechanical stability as well as the mechanical unfolding kinetics of the proteins by using single-molecule atomic force microscopy. Single molecule AFM results show that the mechanical stability of GB1 decreases as the number of inserted amino acids increases and demonstrate that the significant role of the second loop region of GB1 in the mechanical unfolding stability.
4.2 Future plan

Although the topics of single protein mechanics have been around for many years, this is still a burgeoning field full of challenges, and new discoveries in this field are yet to come. Further studies can be carried out by using single-molecule AFM:

1. Tuning the mechano-active site enables the tuning of the mechanical stability of a protein. Based on that, by carrying out protein engineering techniques\(^4\) (eg, site-directed mutagenesis\(^6\), loop insertion\(^5\)), it is possible to tune the mechano-active site of a protein. To carry out further study, by combining loop insertion and engineered disulfide cross-linking, or building a disulfide cross-linking within a native protein using point mutation, it will become possible to tune the mechanical stability of protein by changing the redox potential of the aqueous solution. Further works on the mechanical stability of proteins to ligand binding, protein-protein interactions, applying different voltage, and chemical denaturation will lead to development of engineering artificial elastomeric materials of tailored nanomechanical properties for nanomechanical and biomedical applications. Some proteins that are involved in neurodegenerative diseases or cancer\(^5\) have been studied by single-molecule AFM\(^8\) and give insights on the conformational structure and mechanical properties of disease proteins that might correlate with conditions known to promote the formation of aggregates. Further studies can be made available to directly test the influence of mutations on the conformational structure and mechanical properties of these disease proteins, as well as to find solutions to prevent the disease.

2. Loop insertion approach enables the tuning of the mechanical stability of a
protein continuously in a semi-predictable fashion according to the nonlinear
dependence of the unfolding forces on the number of amino acid residues
inserted. Further studies will need to be carried out to obtain $C^*$. Performing
the chemical or thermal denaturation of a protein will give us $\Delta\Delta G_{U,N}$. Plotting
$\Delta\Delta G_{U,N}$ versus the number of inserted amino acid residues and fitting with
equation: $\Delta\Delta G_{U,N} = -C^*RT\ln(L/L_0)$ will result in $C^*$. If the experimental $C^*$ is
acquired, it is possible to predict the mechanical stability of the GB1 loop
mutants (with any number of amino acid residues being inserted into loop 2,
eg. 15). Once $C^*$ is obtained, the effective loop length of the transition state
before and after loop insertion is also acquired. Eventually, the transition state
before and after loop insertion for mechanical and chemical experiments can
be compared.

3. Streptokinase (S-K) is a plasminogen activator produced by hemolytic strains
of streptococci\textsuperscript{1,10}. S-K is a protein that converts human plasminogen to
plasmin and dissolves blood clots. Therefore, S-K is employed as an effective
and inexpensive clot-dissolving medication in some cases of acute myocardial
infarction following coronary thrombosis\textsuperscript{11}. Human plasminogen contains
seven structural domains. Recent studies suggest that S-K $\beta$ domain may be
involved in the recognition of substrate plasminogen\textsuperscript{12,13}. This illustrates that
the structure of S-K $\beta$ is important for fully understanding the mechanism of
plasminogen activation by S-K. In addition, the mechanical stretching at
constant speed for S-K $\beta$ domain has been examined by using a coarse-grained molecular dynamic model\textsuperscript{14}. Results reveal that S-K $\beta$ (PDB:1c4p) is a
mechanically strong protein. We can use single-molecule AFM as a tool to
study the mechanical stability of S-K $\beta$ domains (construct into
homopolyprotein), and prove whether the experimental result agrees with the
simulation one. This future study will provide detailed insights into the mechanical properties of individual S-K β domains at the single-molecule level.
4.3 References


Appendices:

Appendix A

A restriction enzyme (restriction endonuclease) is an enzyme that selectively cleaves the phosphodiester bond within double-stranded or single-stranded DNA at specific recognition nucleotide sequences known as restriction sites. Many of these restriction enzymes are palindromic, which means that nitrogenous base sequences read the same backwards and forwards, whereas some are non-palindromic. Some restriction enzymes produce 'sticky' ends, whereas others produce 'blunt' ends. The restriction enzymes we used are listed below:

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source</th>
<th>Recognition sequence</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td><em>Bacillus</em> amyloliquefaciens</td>
<td>5’GGATCC 3’CCTAGG</td>
<td>5’GATCC 3’CCTAG G5’</td>
</tr>
<tr>
<td>KpnI</td>
<td><em>Klebsiella</em> pneumoniae</td>
<td>5’GGTACC 3’CCATGG</td>
<td>5’GGTAC 3’CATGG C3’</td>
</tr>
<tr>
<td>BglII</td>
<td><em>Bacillus globigii</em></td>
<td>5’AGATCT 3’TCTAGA</td>
<td>5’AATCT 3’TCTAG A5’</td>
</tr>
</tbody>
</table>
Appendix B

Construction of octomer:

The method we used routinely to make polyproteins in our lab is based on the multiple-step cloning technique\textsuperscript{55}, shown in Figure B1. The gene encoding the desired DNA sequence from PCR is in the plasmids pUC19 (See Figure B1(1)). DNA monomer is flanked by a 5’ BamHI site and a 3’ BglII site followed by KpnI site. The PCR products are cloned into pUC19 linearized with BamHI and KpnI restriction sites. The PCR products are digested with restriction enzymes BamHI and KpnI to produce desired inserts; and digested with KpnI and BglII to acquire desired vectors, shown in Figure B1(2) and B2. Resulting inserts and vectors are ligated by using T4 ligase on the identity of the sticky ends generated by the BamHI and BglII restriction enzymes, generating fragments that can be ligated into a head-to-tail direction, shown in Figure B1(3) and B2. The ligation products are transformed into XL1/Blue competence cells, which contains a gene that is ampicillin- resistance (this step is called ‘transformation’) in order for plasmids (ampicillin- sensitive) to grow on agar plate. The DNA plasmids are isolated from the bacteria by mini-prep and analyzed directly by using DNA agarose gel staining with ethidium bromide to confirm the size/mass of the concatemers. Agarose gel is achieved by moving negative charge nucleic acid through the gel with an electric field. Shorter molecules move faster than longer ones. 8mer is constructed by iterative cloning of monomer into monomer, dimer into dimer, and tetramer into tetramer based on the identity of the sticky ends generated by the BamHI and BglII restriction enzymes (see Figure B1(4)). 8mer is sub-cloned into expression vector pQE80L, which contains an N-terminal (His)\textsubscript{6} purification tag to facilitate the purification of expressed proteins and then transformed into a recombinase- defective \textit{E. coli} cloning strain, DH5\textalpha. The 5’ primer contains a BamHI restriction site that permitted in-frame cloning of the concatemer into the expression vector. The 3’
primer contains a \textit{Bgl}II restriction site, where there are two Cys residues are located and in-frame with the monomer, also has two in-frame stop codons.

Polyprotein construction:

Clones carrying the appropriate DNA sequence are expressed in the 2.5\% LB culture medium, and then 1M of IPTG acted as an inducer is added to the culture mediums in order to induce the transcription. Pellets are collected by centrifuging the culture medium and stored at -80 degree freezer. The next morning, pellets are lysed with protease inhibitor, lysozyme, DNase, and RNase in ice and then centrifuged at high speed for about an hour. Supernants contain the most of the recombinant proteins are collected. Protein is then purified by Ni- NTA affinity chromatography, which binds the (His)$_6$ tag polyproteins. Eluted proteins are then analyzed by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) to confirm the size and the expression level.

Figure B1 A schematic diagram to show polyprotein construction procedures: (1) PCR (polymerase chain reaction); (2) Digestion with restriction enzymes to produce desired inserts and vectors; (3) Ligation with T4 ligase to make dimer; (4) dimer into
dimer, tetramer into tetramer (repeating step 2, 3); (5) protein expression and protein purification to make polyprotein (each circle represents an individual domain).

**Figure B2** Directional DNA multiple steps cloning strategy is done by ligation of the sticky ends of the palindromic *Bam*HI and *Bgl*II restriction site of the fragments.

Sequence of restriction enzymes, *Bam*HI and *Bgl*II, are complementary to each other. Ligation is done by using T4 ligase enzyme. (1) Plasmids cut with *Bgl*II and *Kpn*I restriction enzymes to generate a vector; (2) Plasmids cut with *Bam*HI and *Bgl*II restriction enzymes to produce an insert; (3) Ligation by using T4 ligase to join the palindromic (symmetric) *Bam*HI and *Bgl*II restriction enzymes in a head-to-tail direction to generate desired products. The restriction enzyme cleaves the phosphodiester bond within a polynucleotide chain (red lines).

pQE80L expression vectors (addgene website) are considered as one of the most powerful system for cloning and expression of recombinant proteins in *E. coli*. This high-level expression vector contains *(His)_6* tag on N-terminus for the affinity purification of the protein. pQE80L has 4751 base pairs and is ampicillin resistance.
Figure B3 indicates a map of the custom-made expression vector. The final products have eight domains (either repeated domains or chimeras) with a short His-tag on the N-terminus. Previous study suggests that His-tag does not affect the properties of the flanking regions.

Figure B3 Map of the custom-built pQE80L expression vector. (Picture acquired from Addgene website).
Appendix C

AFM cantilevers are commercially available, with wide range of sizes and spring constants. Spring constants of AFM cantilevers for most protein experiments are ranged from 10 to 100 pN/nm. In our typical protein pulling experiments, we use cantilever with spring constant ~60 pN/nm. AFM cantilever tip picks out single-molecules non-specifically from a dense layer of proteins. Cantilever tips are made of silicon nitride (Si$_3$N$_4$). The use of silicon nitride as AFM tips is suitable for protein adsorption as well as to adapt specific attachment of the protein. The size of the unsharpened cantilever tips (radius of curvature of ~50nm) is far bigger than that of a typical protein (~5nm), thus the successful rate of picking up single molecule protein with an unsharpened cantilever tip is higher. On the other hand, to collect data with the sharpened tips (radius of curvature ~20nm) may reduce the chance of the tip attaching a single-molecule protein. Gold coated surfaces, located at the back of the cantilever allow the laser to be reflected effectively.

Cantilevers always require to be calibrated prior each single-molecule force spectroscopy (SMFS) experiment. Two constants are to be determined: the ratio of photodiode output voltage to cantilever deflection ($\Delta V/\Delta z_c$) and the spring constant of the cantilever ($k$). Once these values are known, the AFM cantilever acts as the Hooke’s spring: $F=-k \cdot \Delta z_c$ (where $k$ is the spring constant; $\Delta z_c$ is the tip deflection along the $z$-axis), describing a linear elastic spring, and the mechanical resistance force probed with this sensor is determined.$^{50}$

The ratio of the photodiode output voltage to cantilever deflection ($\Delta V/\Delta z_c$) is used to convert the photodiode output into cantilever deflection (nm). The difference of the voltage output can be determined when a laser is reflected on the back of the cantilever and onto the split photodiode detector, which converts the incident light on each half of the split photodiode into voltage and outputs the voltage difference.
$\Delta V/\Delta z_c$ can be determined by bringing the tip into contact with a hard surface, undergoing an approach-retraction cycle, shown in Figure C1. The cantilever deflection should be equal to the piezoelectric positioner movement, which is known from the instrument calibration.

**Figure C1** AFM cantilever calibration. Cantilever tip is away from the substrate (1) and starts to approach the substrate surface (2). When the cantilever tip is in contact with the surface, the cantilever bends (3). The cantilever bends further as the tip and the surface are adhered to each other (4, 5). As the contact between the tip and surface cannot be reached any further, the cantilever starts to pull off and moves away from the surface. Such approach-retraction cycle needs to repeat many times until the cantilever deflection is equal to the movement of the piezoelectric positioner.
Appendix D

The size of multimers, (GL7A-127)ₙ-pUC19 (n=1, 2, or 4) was confirmed by DNA agarose gel. Agarose gel electrophoresis is a method to separate DNA by sizes. The gel is made of agarose, TAE buffer and tiny amount of ethidium bromide (EtBr fluoresces under UV light). Ethidium bromide stained DNA does not make the DNA visible in the gel. DNA plasmids need to mix with negatively charged loading buffers before adding to the gel. Loading buffers are visible in natural light and move at the same speed as DNA of a certain length. The gel is then illuminated under UV light, and DNA plasmids are visible. DNA ladder is added next to the sample as a reference to estimate the size of unknown DNA molecules.
Lane 1 (A, B, C and D): 2-Log DNA ladder

(A) Lane 2: no insert; Lane 3: (GL7A-I27)-pUC19 (GL7A, 56*3=168 base pairs [b.p.]; I27, 89*3=267 b.p.; thus, 168+267=435 b.p.), 1.2% agarose gel.

(B) Lane 2: (GL7A-I27)\textsuperscript{2}-pUC19 (435*2=870 b.p.), 1.0% agarose gel.

(C) Lane 2: (GL7A-I27)\textsuperscript{2}-pQE80L, 1.0% agarose gel.

(D) Lane 2: (GL7A-I27)\textsuperscript{4}-pQE80L (435*4=1740 b.p.), 0.8% agarose gel.

(E) 2-log DNA ladder visualized by ethidium bromide staining on a 1.0% TBE agarose gel. (pictures downloaded from www.neb.com)
Appendix E

The size and expression level of (GL7A)$_8$ polyprotein were confirmed by Tris-Glycine SDS-PAGE. SDS-PAGE was made of 12% resolving gel (on the bottom) and 5% stacking gel (on the top). 12% resolving gel was made of H$_2$O, 30% acrylamide mix, 1.5 M Tris (pH 8.8), 10% SDS, 10% ammonium persulfite and TEMED. 5% stacking gel was made of H$_2$O, 30% acrylamide mix, 1.0 M Tris (pH 6.8), 10% SDS, 10% ammonium persulfate and TEMED.

Lane 1: pre-stained protein marker (~10 uL)

Lane 2: supernatant fraction; Lane 3: washing fraction; Lane 4: Elution 1 fraction; Lane 5: Elution 2 fraction.

(Proteins were eluted by elution buffer to obtain elution 1 fraction and elution 2 fraction. Elution 1 fraction has higher concentration than elution 2 fraction because elution 1 fraction elutes first.)
Appendix F

The size of multimers for GB1 loop mutants was confirmed by DNA agarose gel.

(A) Lane 2: (GB1-L24)\textsubscript{2}-pUC19 (56+24=80, 80*3=240, 240*2=480 b.p.);
Lane 3: (GB1-L46)\textsubscript{2}-pUC19 (56+46=102, 102*3=306, 306*2=612 b.p.).

(B) Lane 2: (GB1-L24)\textsubscript{4}-pQE80L; Lane 3: (GB1-L46)\textsubscript{4}-pQE80L

(C) Lane 1: (GB1-L24)\textsubscript{8}-pQE80L; Lane 2: (GB1-L46)\textsubscript{8}-pQE80L
Appendix G

The size and expression level of GB1 loop mutants were confirmed by Tris-Glycine SDS-PAGE. SDS-PAGE was made of 12% resolving gel (on the bottom) and 5% stacking gel (on the top). 12% resolving gel was made of H$_2$O, 30% acrylamide mix, 1.5 M Tris (pH 8.8), 10% SDS, 10% ammonium persulfate and TEMED. 5% stacking gel was made of H$_2$O, 30% acrylamide mix, 1.0 M Tris (pH 6.8), 10% SDS, 10% ammonium persulfate and TEMED.

Lane 1: prestained protein marker; Lane 2: Elution fraction 1 (GB1-L24)$_8$ polyprotein; Lane 3: Elution fraction 2 (GB1-L24)$_8$ polyprotein; Lane 4: Elution fraction 1 (GB1-L46)$_8$ polyprotein; Lane 5: Elution fraction 2 (GB1-L46)$_8$ polyprotein.
Appendix H

To simulate the complex behaviour of proteins, Monte Carlo techniques are used to model and reproduce the probability of unfolding versus the applied force in order to estimate the unfolding distance $\Delta x_u$ and the unfolding rate constant at zero force $k_u^0$. Here, we assume that the persistence length for all proteins is 0.5nm. AFM extends the polyprotein with a speed $v_c$ starting from $x=0$. The additional extension $\Delta x$ at each time interval $\Delta t$ is: $\Delta x = v_c \times \Delta t$. After each time step, the force generated by stretching the protein chain is first calculated with the WLC model. When an external force is applied to the folded protein causing to unfold, the unfolding rate constant is calculated by using $k(F)=k_u^0 \exp(F\Delta x_u/k_B T)$. A protein gives the probability of observing the unfolding of any domain as $P_u = N_f k(F) \Delta t$, where $N_f$ is the number of folded domains, $k(F)$ is the unfolding rate constant and $\Delta t$ is the polling interval. If the probability of unfolding is higher than a certain number, the protein unfolds to produce force extension curves. If the probability of unfolding is lower than a certain number, the protein does not unfold. We need to try different $\Delta t$ and run the simulation again to see whether $P_u$ is higher or lower than the certain number. The Monte Carlo simulation is used to fit the pulling kinetic experiment and the speed dependence of the unfolding forces. The best fits for each unfolding force are obtained. The values of $k_u^0$ and $\Delta x_u$ in the simulations are adjusted until the simulation data match the experimental ones.