Possible Use of DNAs in single molecule force spectroscopy to probe single protein unfolding signatures

Author:

Saw Thuan Beng

Supervisors:

A/Prof Wang Zhisong

Prof Lim Chwee Teck

A THESIS SUBMITTED
FOR THE DEGREE OF MASTERS OF SCIENCE
DEPARTMENT OF PHYSICS
NATIONAL UNIVERSITY OF SINGAPORE

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.

I have dully acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Saw Thuan Beng

29 July 2013

I. Preface/Acknowledgements

The project started off with the aim of characterizing the mechanical aspects of the protein, important to elucidate cell-cell adhesion at the molecular level. However, half way through the project, we found that the bulk of the data from AFM, although sufficient for preliminary deductions, is far from satisfactory. This situation was not improved because protein engineering constructs that could improve our study was not immediately available due to difficulty in construction.

At the same time, I was taking the advanced biophysics course in NUS by Prof. Yan Jie. It was in one of his classes that I heard of his idea to use DNAs to innovate and improve AFM single protein unfolding signal recognition. Understanding that the signal recognition problem was the main issue with AFM results and key to improving the quality of my results, I approached Prof. Yan and volunteered to help him develop his idea. After development of this technique, it can potentially be used for the initial project.

There is a list of people I wish to thank: Prof. Yan Jie (Physics department) and Prof. Lim Chwee Teck (BioEngineering department) for supervising my work. Prof. Yan Jie for the conception of the new AFM-DNA idea. Lu Chen (RA in Prof. Liu's lab) for ~ 80% of the Magnetic Tweezers work presented in the thesis which I used for comparison with my AFM results. Prof Rene-Marc Mege for providing the α-Catenin constructs. WuFei and QiuWu (PhD students) for teaching me how to use AFM and Magnetic Tweezers. KongFang, Brenda and LiuMin (Post Doc and RA at Prof. Lim's lab) for helping me with the AFM at GEM4 lab in SMART. Zhang XingHua for providing me with the DNAs and related advice. Prof. Liu RuChuan (Physics department), Prof. Wang ZhiSong (Physics department) and Prof. Benoit Ladoux (Biology Department) for advice and helpful discussions.

Table of Contents

II. Abstract

Using atomic force microscope (AFM) to characterize single protein mechanics has certain limitations in terms of obtaining and recognizing single protein unfolding signals. This led us to develop a new approach through using DNA molecules as markers to probe the unfolding of our proteins. One of the basic protein parameters that can be extracted is the unfolding structure (i.e. change in contour length after unfolding). Our experiments using both AFM and Magnetic Tweezers suggest that the protein of our study, α-Catenin constructs, has $4-5$ unfolding units, with each increasing the contour length by $\sim 10\ nm$. However, AFM gives poorer results in terms of significantly larger histogram distribution. The main issue here is with single-molecule signal recognition. Here, we introduce a new approach using AFM which can potentially overcome this problem and improve upon existing methods for enhancing the quality of the AFM results, e.g. heteromeric polyprotein using $I27$. We aim to couple DNA overstretching and streptavidin-biotin interaction specificity to more unambiguously identify protein signals. We provided a working protocol to immobilize DNA for AFM manipulation. Preliminary experiments were first done only with DNAs and without proteins, and the results showed encouraging features which were important for its efficient use in single molecule force spectroscopy: 1) short tip-surface pause time of $\sim 1-2\ s$ 2) reusable streptavidin-biotin bond after breakage 3) $\sim 10\%$ of all curves having DNA signals, verified by fitting the extensible Worm-Like Chain (WLC) model. Most importantly, there is a stable overstretching force range of $50-70\ pN$ and a clear force plateau extension of $\sim 50\%$ of the fabricated DNA contour length, which can provide the two important marker parameters for protein identification.

III. List of Tables

IV. List of Figures

this additional force, exposing a binding site (purple arrow) for more F-Actin. This way, the junction recruits more forces to stabilize the adhesion (following [16]).

Figure 4: The whole composite bridge (leading to cell nucleuses at two ends) for cell-cell adhesion. Basic Materials: (F-Actin, α-Catenin) for Cell 1 and Cell 2, and other molecular complexes (two blue rods) linking them. Weak points of bridge (usually non-covalent bonding [18]) are pointed by dark arrows.

Figure 5: Simplified picture of protein folding/unfolding. (A) Folded protein held by two bonds in a solution. Due to Brownian motion, the outer bond can break (with certain rate, k_0) and reveals the inner part of protein. (B) When exerted by force at two ends, protein unfolds at different rate, k_F. r denotes the distance between protein ends.

Figure 6: (A) Bond energy landscape/potential as a function of bond length, r (i.e. distance between protein ends (Figure 5)). Potential energy shows two local minimum. Minimums correspond to folded state (unfolded state), at x_f (x_u) and is separated by barrier (height, E_b), at x_b. The landscape at x_f (x_b) is approximated by (inverted) harmonic potentials with stiffness, $\omega_f\sqrt{m}$ ($\omega_b\sqrt{m}$). The protein is in the folded state (blue circle). (B) (Black curve) Initial bond potential. (Red line) Constant external force potential. (Blue curve) Modified potential (i.e. sum of bond and external force potential).

Figure 7: One recent molecular model of α-Catenin monomer. (A) Linear amino acid sequence for α-Catenin monomer, separated into four main domains, $D1$, $D3a, b$, $D4$ and $D5$. Numbers $1-906$ indicate amino acid number. $D1$ and $D5$ bind molecular partners to form complete molecular bridge. $D3a$ contains Vinculin binding site (cyan). $D3b$ and $D4$ form M domain, modulating $D3a$. Our experiments use a recombinant construct of bracketed region, $D3a$ and M. (B) α-Catenin consists of a series of α-helical bundles, color code follows (A). $D5$ position is rather flexible so is omitted to facilitate visualisation. Adapted from [57].

Figure 8: Not drawn to scale. (A) Magnetic Tweezers with Total Internal Reflection fluorescence (TIRF) technique. The immobilized protein attached to paramagnetic bead is pulled by magnetic field which exerts force on the bead. Vertical extension, z, of protein measured by evanescent wave from total internal reflected laser beam. (B) Atomic Force Microscopy (AFM). An immobilized protein is pulled by the flexible AFM cantilever controlled by a motorized piezo. AFM cantilever acts like a spring and exerts force on protein.

Figure 9: (Cantilever, bead and protein not drawn to scale) (A1) Typical force-extension curve for AFM constant-v mode which detected protein. (Light red curve) During cantilever approach to surface, cantilever has no deflection (i.e. F=0). When AFM cantilever touches surface, cantilever is deflected upward and force increases positively. (Dark red curve) When cantilever is retracted from surface, the first straight peak shows non-specific interaction with surface bending the cantilever backward (i.e. F negative). After leaving surface, there are saw-tooth patterned peaks corresponding to protein pulling and unfolding. Unfolding corresponds to the straight part between two saw-tooth patterns (blue arrows). (A2) Zoom into one saw-tooth pattern. {1} Protein (green chain) is pulled and accumulates tension. {2} Protein unfolds and releases tension (i.e. decrease in cantilever deflection). {3} Protein pulling cycle continues. (M1) Typical extension-time curve for Magnetic Tweezers constant-F mode which detected protein. Curve shows increase of bead-protein extension with time. Protein unfolding corresponds to step increase of the extension (red arrows). (M2) Zoom into one plateau-step pattern. {1} Protein is taut, thus extension is constant (average over noise). {2} Protein unfolds, and there is sudden (step) increase in protein extension. {3} Protein pulling cycle continues.

Figure 10: AFM data for unfolding length. (A) Blue circle shows example of data points that we collect i.e. contour length change during unfolding, ΔL, and corresponding unfolding force, F. Other histogram parameters are clearly stated in the example. For each fixed velocity experiment, we analysed ~ 40 – 80 curves. (B) 2D colour graph shows three experiments at different velocity, v $(= 100, 600, 3600\ nm/s)$, plotting ΔL, against F. Colour signifies relative frequency of data, e.g. red means highest frequency. F ranges from $0 - 180\ pN$ for all v. For $v = 100\ nm/s$, there is one red frequency peak at $\sim \Delta L = 10\ nm$. For $v = 600\ nm/s$, there are two red frequency peaks at $\sim \Delta L = 10, 30\ nm$, and one yellow peak at $\sim \Delta L = 40\ nm$.. For $v = 3600\ nm/s$, there are two red frequency peaks at $\sim \Delta L = 20 - 30\ nm$, and two yellow peaks at $\sim \Delta L = 10 - 20, 40 - 50\ nm$. Frequency peaks are shifted to higher ΔL with increasing v. (C) Histogram lumps all ΔL data points of all constant velocity experiments of different v (range $100 - 3600\ nm/s$). There are $500 - 600$ curves and over 1000 data points. Only one single peak at $\sim \Delta L = 10\ nm$. Half width is $\sim 40\ nm$. (D) Histogram shows total unfolding contour length (i.e. sum of all ΔL in one curve) per pulling curve for all experiments of different v. Single peak at $20\ nm$, but half width ranges from $10 - 40\ nm$. (E) Histogram shows number of unfolding per pulling curve for all experiments of different v's. Most curves have two unfoldings while some have maximum of six unfoldings.

Figure 11: Magnetic Tweezers data for change in contour length at unfolding, ΔL. Histogram lumps all ΔL for all experiments at different constant F (range from $0 - 15\ pN$). Single peak at $\sim 10\ nm$. By experience, average number of unfolding per pulling curve is $4 - 6$, e.g. in Figure 12. (work with Lu Chen, a research assistant in Prof. Liu's lab. ~80% of data from Lu Chen)

Figure 12: Three typical constant force pulling curves (different forces) for Magnetic Tweezers, plots z extension of bead against time. Red arrows show unfolding steps. On average, all unfolding events finish within duration $50 - 150\,s$. By experience, most curves follow this trend.

Figure 13: AFM constant velocity pulling data. Each point represents unfolding force average, $\bar{F}$, of all data for one single pulling velocity experiment. Graph plots mean unfolding force, $\bar{F}$ against log of the pulling velocity. Points can be roughly separated into two regimes, one where points hover around a plateau ($Ln(v) < 7$) and another where points steadily increase ($Ln(v) > 7$).

Figure 14: (adapted from [48]) Upper row: sandwich heteromeric polyprotein, with analye (red) and marker (blue). Below: Example of unfolding signal in force-extension curve from the construct. Red line fitted curves are from marker. Black arrow is analyte signal.

Figure 15: Envisioned configuration of experimental setup. DNA on AFM cantilever tip can search for protein on glass slide with correct chemistry.

Figure 16: (A) dsDNA double helix and dimensions (adapted from [50]). (B) Typical force extension curve of dsDNA in a SMFS experiment. Regime 1 ($< \sim 10\,pN$) can be fitted with WLC, with persistence length, $p \sim 50\,nm$. Regime 2 and 4 ($\sim 10 - 65\,pN$; $> 65\,pN$) can be fitted with extensible WLC, with different parameters i.e. p and stretching modulus, k. Regime 3 is overstretching plateau, extension $\sim 70\%$ of contour length, $F \sim 65\,pN$ depending on experimental conditions: temperature, salt concentration, etc.

Figure 17: Expected setup schematic for protocol in 3.3.3) Experiment Design. Length scales are not representative. Functional surfaces: BSA-biotin cantilever and Streptavidinated glass slide. Both DNA ends are biotinylated. Biotin and streptavidin have very specific binding affinity and can bind upon meeting. Some DNAs form loops. Some DNAs are capped with Streptavidin and have free end. The latter is available for pulling and stretching.

Figure 18: Three typical force-extension curves from AFM pulling using setup in Figure 17. Light red curves show extension of AFM cantilever towards surface, while dark red curves show retraction from surface. Vertical deflection (pN) is not always indicative of real force but has to be normalised by the horizontal dotted line, taken as $0\,pN$. Top panel (No DNA signal): represent $\sim 90\%$ of total curves, associated to background force and no DNA being stretched. Bottom panel: Both signals represent $\sim 10\%$ of total curves.

Among them, $\sim 75\%$ is One-DNA signal, $\sim 25\%$ is Two-DNA signal. Green line fits the short "tail part" of the stretching after the plateau using extensible WLC. Fitted parameters are very similar. Contour length (nm/bp) is calculated using contour length (nm) divided by 3 kbp for One-DNA and 6 kbp for Two-DNA.

Figure 19: (A) Pipetting fluid induces shear flow on DNA but it is verified that DNA-bead stays intact after normal pipetting. (B) Example of cantilever pulling DNA which eventually breaks at the SV-biotin bond at surface. DNA transferred to cantilever.

Figure 20: (A) Bond energy landscape/potential as a function of bond length, r (i.e. distance between protein ends). Potential energy shows two local minimum. Minimums correspond to folded state (unfolded state), at x_f (x_u) and is separated by barrier (height, E_b), at x_b. The landscape at x_f (x_b) is approximated by (inverted) harmonic potentials with stiffness, $\omega_f\sqrt{m}$ ($\omega_b\sqrt{m}$). The protein is in the folded state (blue circle). (B) (Black curve) Initial bond potential. (Red line) Constant external force potential. (Blue curve) Modified potential (i.e. sum of bond and external force potential). (C1) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, minimum near x_f. (Blue curve) Modified potential with only one minimum, close to initial folded state position (i.e. x_f). (C2) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, with minimum between x_f and x_u, close to x_b. (Blue curve) Modified potential with two minimums. The minimum on the right represents new unfolded state. (C3) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, minimum near x_u. (Blue curve) Modified potential with one minimum, close to initial unfolded state position (i.e. x_u).

1. Introduction

It has recently become clear that mechanical forces and factors have a direct impact on many of the most important life processes, e.g. cell differentiation [2], cell migration [3] and cell-substrate adhesion [4-6]. This invites physicists to study relevant and important biological questions. At the molecular level, the key functional parts of a cell are proteins and DNAs. Therefore, studying the mechanics and the mechanotransduction mechanisms involving proteins and DNAs constitute an integral part of this emerging field, called Mechanobiology. Some important questions at the molecular level being answered are how DNA biomechanics regulates gene expression [5] and how protein mechanotransduction accounts for cell functions such as cell-substrate adhesion [7].

The inside of a cell is an extremely complex environment (ref. Figure 1). One way to simplify the study of macromolecular mechanics is to isolate the relevant molecules from a cell and study them in vitro. Still, this is a mammoth task due to the sizes involved (DNA: coiled volume $\sim 0.1\,\mu m$, proteins: $\sim 1-10\,nm$). Past work by molecular biologists has allowed specific bio-molecule isolation to be possible. However, another intrinsic difficulty lies in that these macromolecules are soft-matter objects and can have important structural and mechanical changes induced by small changes in forces, temperature, solution pH, etc [8]. This complicates experimental efforts to study them.

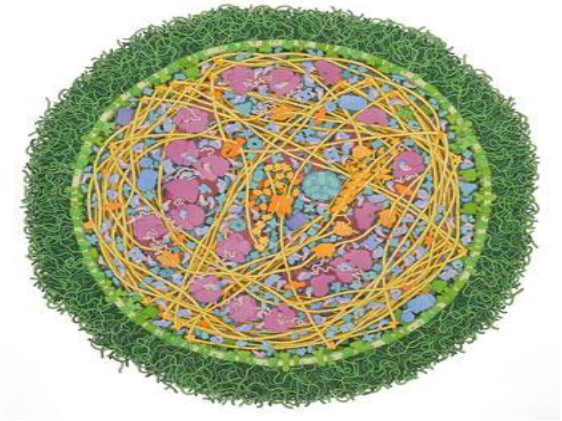


Figure 1 : (drawing by scientist, David Goodsell) *Mycoplasma mycoides* bacteria. Extremely packed cellular condition with DNA shown in orange, cytoplasmic proteins in blue and pink [50].

Historically, in vitro molecular studies are done in bulk where only the average values of many molecules tested were obtained (e.g. electrophoresis to study molecule structural size, circular light dichroism to study protein denaturation, etc). Mechanical information of the molecules had to be inferred indirectly from these bulk measurements, e.g. single DNA elasticity and bendability [9]. However, the last two decades of intense instrumentation research in this field has seen the development and maturation of truly single-molecule experimental tools. They can probe bio-molecules one at a time, e.g. Atomic Force Microscopy (AFM), Optical Tweezers, Magnetic Tweezers (MT), Biomembrane-Force-Probes, etc [10, 11] . Single-molecule tools are in most cases preferred over bulk assays because we do not miss any information from averaging [12]. Moreover, they allow for precise measurement and control ($\sim nm$) and can directly apply physiologically relevant forces ($\sim 10\, pN$) on the bio-molecules. This gives a direct investigation of the role of forces on biological processes. Among the existing tools, AFM is the most developed and commercialised technique. However, after working with it for the bulk of the project, we find that the single-molecule signal recognition for AFM is still a problem i.e. noise from the environment masks the real protein unfolding signals

We thus suggest and work on a new AFM single-molecule signal recognition method that capitalizes on DNA biomechanics research of almost two decades. Hopefully, this can help to more unambiguously identify the protein unfolding signals.

This Master's Thesis has two parts. The first part intends to serve as a primer to single-molecule biomechanics research and its tools, allowing the reader to appreciate the use and subsequent need for AFM improvement. After an elaborate introduction (i.e. scientific background, protein unfolding theory, etc.) to the field with a specific case study on α-Catenin and cell-cell adhesion, AFM and Magnetic Tweezers results on the protein unfolding are shown. We found that the AFM high-throughput-data-collection does not translate to an overall advantage in data quality and efficiency over Magnetic Tweezers i.e. unfolding structure histogram for AFM is much more widely distributed than that of Magnetic Tweezers even though AFM has much more data.

This naturally leads us to the second part of the report where we propose a new method that we hope can improve the quality of our AFM results[a]. We discuss existing methods for aiding the recognition problem and getting better AFM results, but we also observe that these methods have their own intrinsic problems. We hope that our new method, which consists of using DNAs for protein searching, can potentially overcome all the obstacles faced by the preferable current method i.e. use of $I27$ in heteromeric polyproteins. Finally, we present some encouraging results from tests on DNAs alone and discuss necessary follow-up work to consolidate the idea.

[a] The inability to use current methods for improving our AFM results played a big part for us to start working on the new method directly, relegating the protein characterization project for the time being.

2. Single-Molecule Biophysics and Tools

Single-molecule biophysics/mechanics studies life processes at the molecular level. Some of these studies, regardless of the scientific questions, start with the mechanical characterization of the molecule involved. However, for the reader to appreciate this field, we will put the mechanical characterization problem in the context of a specific, open question that we are working on.

The scientific question is to understand the stability of cell-cell adhesion (i.e. how cells stay connected under dynamic conditions), central to basic biological functions such as tissue wound healing, maintenance of skin integrity, cancer metastasis, etc. Cell-cell adhesion is a complex process that is dependent on the ability of cells to sense and react to other cells surrounding it [5]. We would like to see whether minute physiological forces play a role in cell sensing-response and investigate this at the molecular level. The important molecule (i.e. α-Catenin protein) implicated in the process has recently been identified by biologists, so our job is to exert very small forces ($\sim 10 - 100\, pN$) on this macromolecule ($\sim 5\, nm$) and see how it reacts. This is to simulate typical forces experienced by molecules in our cells. To do this, we used two different single-molecule techniques i.e. Atomic Force Microscopy (AFM) and Magnetic Tweezers[b], with an emphasis on AFM.

The background on cell-cell adhesion and cell sensing-response will be given and defined. An interesting molecular model of cell sensing-response and the role of forces are shown. In short, minute physiological forces are hypothesized to be able to weaken the adhesion-protein's bond sufficiently. This will lead to bond breaking and protein unfolding. The unfolding finally reveals a specific functional site to recruit other adhesion stabilizing molecules. Interestingly, this implies that initial bond breaking leads to the cells staying connected. We state clearly the goals of the project, aimed at proving this

[b] Magnetic Tweezers work is shared between group Research Assistant (80%) and me (20%).

mechanistic view of adhesion stability. In the theory section, we introduce the model of a chemical bond as a basis to understand protein unfolding. We also describe force loading of a chemical bond to show the importance of force in this process (i.e. increase protein unfolding rate). Finally, we discuss an overview of the single-molecule experimental techniques used.

2.1) Scientific Background:

2.1.1) Cell adhesion – cell sensing response

Biological cell-cell adhesion means the sticking of two cells which are close together and helps in tissue formation [17]. Interestingly, cell-cell adhesion has two contradicting features. Firstly, the adhesion has to be dynamic enough to allow continuous tissue growth and renewal (i.e. neighbouring cells need to part from each other momentarily to accommodate new cells). However, the "sticking" of cells has to be stable enough such that the tissue stays intact. The worst case scenario of an unstable tissue is when individual component cells become too mobile and invade into other parts of our body (i.e. metastatic cancer cells). Thus, cell-cell adhesion is in stark contrast with simple physics systems where "dynamic" and "stable" are usually mutually exclusive. It is this stability of cell-cell adhesion under dynamic conditions that we are interested in investigating.

The physical structures which form at the interface of two cells adhering to each other are called cell junction. They are complex protein assemblies found at the edges linking two neighbouring cells (Figure 2). In this project, we focus on one of the three junctions, called Adherens junction, which is directly responsible for adhesion stability and skin maintenance. Recently, people have re-discovered a key protein at the Adherens junction, called α-Catenin, which involves actively

in the stabilizing function. The key structural features of α-Catenin (as all proteins) are that it can be in two different functional states (i.e. folded or unfolded state). Unfolded state signifies opening of certain chemical binding site which is initially hidden in the folded state.

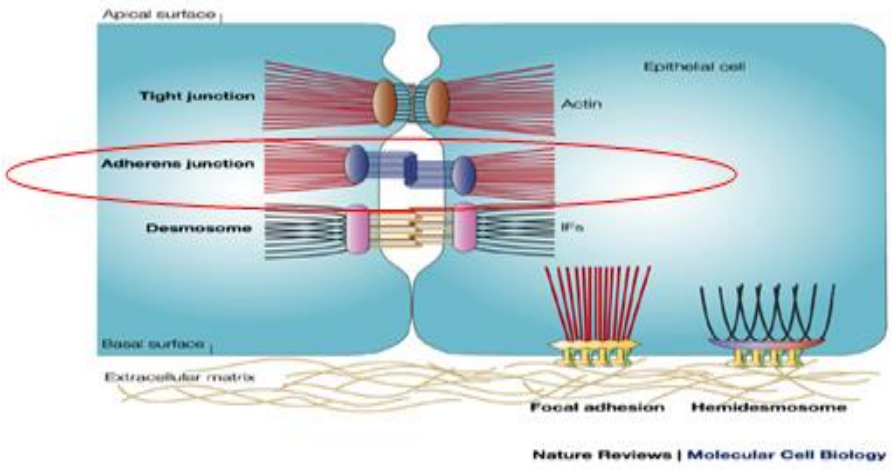


Figure 2: Side-view of two neighbouring cells sitting on substrate (extracellular matrix). At cell-cell interface, three types of junctions (Tight junction, Adherens junction and Desmosome) are formed to connect cells. Adherens junction (circled in red) is linked to cell backbone (F-actin) and is important for cell recognition, skin maintenance and morphogenesis [16]. [56]

In 2010, Yonemura et al. proposed a mechanistic model which describes how α-Catenin help stabilize cell-cell adhesion [16, 19]. Very recently in 2013, Thomas et al. independently did work that supported this model [20]. The model proposes that α-Catenin can sense minute mechanical force change when two adhering cells start to part (i.e. weakening of adhesion), and translates this into chemical signalling in the cell. The α-Catenin can then help the cell respond to enhance the adhesion. How does the protein do this? The model is explained below in some depth. Finally we give a more precise "definition" for cell sensing-response.

In Figure 3(A), two cells are initially in the adhered state. At their junction, α-Catenin is initially folded, and is directly linked to some F-Actins (i.e. "cell

backbone"), forming an initial composite bridge (Figure 4) that are then linked to both the cell nuclei (not shown). The integrity of this composite bridge ensures that the two neighbouring cells are always close together (i.e. cell-cell adhesion). Conversely, if the bridge breaks somewhere, cell-cell adhesion is broken. The most probable breaking points are the weak points where different individual components connect (dark arrows in Figure 4). Weak points are usually non-covalently bonded [18], 10 – 100 times weaker than covalent bonds that make up the individual components forming the composite bridge. The Yonemura model neglects breaking of the adhesion molecular complexes (two blue rods) at the cell-cell interface and only concentrates on the α-Catenin – F-Actin connection.

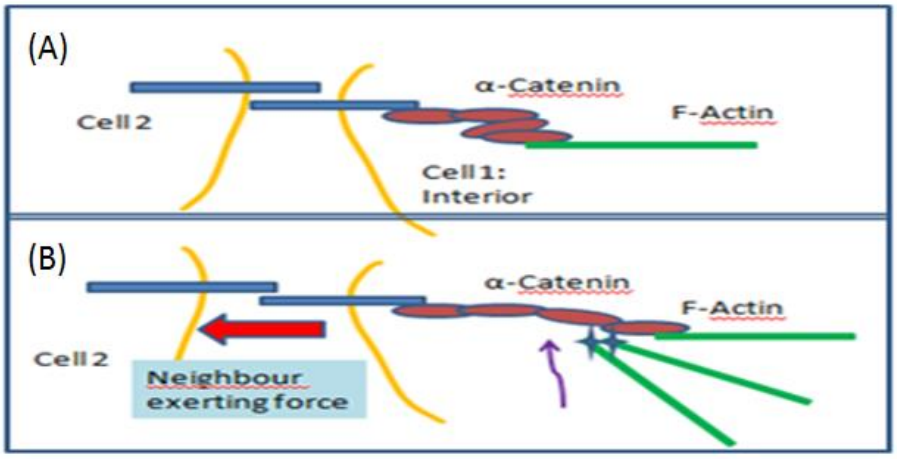


Figure 3: (A) Two cells adhered together. At junction, α-Catenin is folded and linked to F-Actin ("cell backbone") for baseline stability. (B) Yonemura model [16]: Second cell pulling away, thus exerting force on the junction of first cell. α-Catenin is unfolded by this additional force, exposing a binding site (purple arrow) for more F-Actin. This way, the junction recruits more forces to stabilize the adhesion (following [16]).

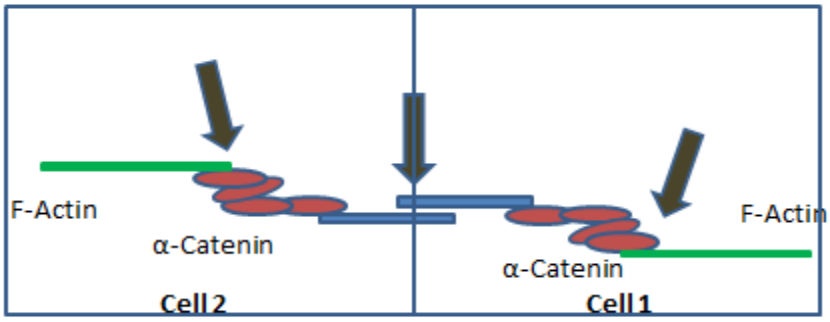


Figure 4: The whole composite bridge (leading to cell nucleuses at two ends) for cell-cell adhesion. Basic Materials: (F-Actin, α-Catenin) for Cell 1 and Cell 2, and other molecular complexes (two blue rods) linking them. Weak points of bridge (usually non-covalent bonding [18]) are pointed by dark arrows.

The Yonemura model proposed the following, as in Figure 3(B). When cell 2 moves too far away from cell 1, it induces additional tension/force in the cell-cell bridge and could potentially cause breakage of the α-Catenin – F-Actin connection. However, α-Catenin can unfold under this additional force, opening up a binding site for more Vinculin - F-Actins to bind. The new F-Actins are transported by a protein called Vinculin and it is the Vinculins that bind to the opening of α-Catenin. With more α-Catenin – F-Actin connections, the bridge is less likely to break totally and thus cell-cell adhesion is stabilized under these dynamic conditions. Actually, as shown in subsection 2.1.3) Theory: Protein Unfolding (Non-Covalent bond breaking), it is the increasing of α-Catenin unfolding rate with increasing force that is crucial for more efficient F-Actin recruitment. This is because all chemical bond breakage is probabilistic in nature. The important idea in this model is that the minute physiological forces ($\sim 10 - 100\ pN$) is sufficient to unfold the proteins (i.e. weaken and break the protein bonds).

Finally, "definition" for cell-cell adhesion, cell sensing and cell response in this thesis is given in Table 1. With this overview of the biological motivation, we can go on to state the aims of the project.

Table 1 - Definitions for key events

(D1) Cell-cell adhesion	Integrity of composite bridges (Figure 4) linking the two cell nuclei.
(D2) Cell sensing	α-Catenin increased unfolding rate with increasing force, before cell-cell adhesion breaks down.
(D3) Cell response	More Vinculin - F-Actin recruited to unfolded α-Catenin in shorter time.

2.1.2) Scientific Aims

The Yonemura model [16, 19, 20] proposes that α-Catenin acts as a force transducer (i.e. sensing forces from the environment and translating it into chemical signalling in cells) by allowing Vinculin binding after it unfolds. The plan is to do a direct mechanical investigation of this model at the single molecular level, which involves:

(I1) Showing that Vinculin only binds to α-Catenin when it is in the unfolded state i.e. to test *(D3) Cell response*.

(I2) Showing that the relevant unfolding rate increases significantly with forces i.e to test *(D2) Cell sensing*.

In this thesis, we report some work done on (I2) for our purpose. More specifically, the direction is 1) characterizing the "relevant" unfolding structures under force for single molecule α-Catenin and 2) determining unfolding rates as a function of force for single molecule α-Catenin. A description of a chemical bond is given in the next subsection to show protein unfolding features, including how the bond dissociates naturally or when a force is applied to the bond.

2.1.3) Theory: Protein Unfolding (Non-Covalent bond breaking)

To understand protein unfolding structure/rates, we need to know how proteins unfold. We also discuss how force can help increase protein unfolding rate, which is key to the *D2) Cell Sensing* feature of the Yonemura model.

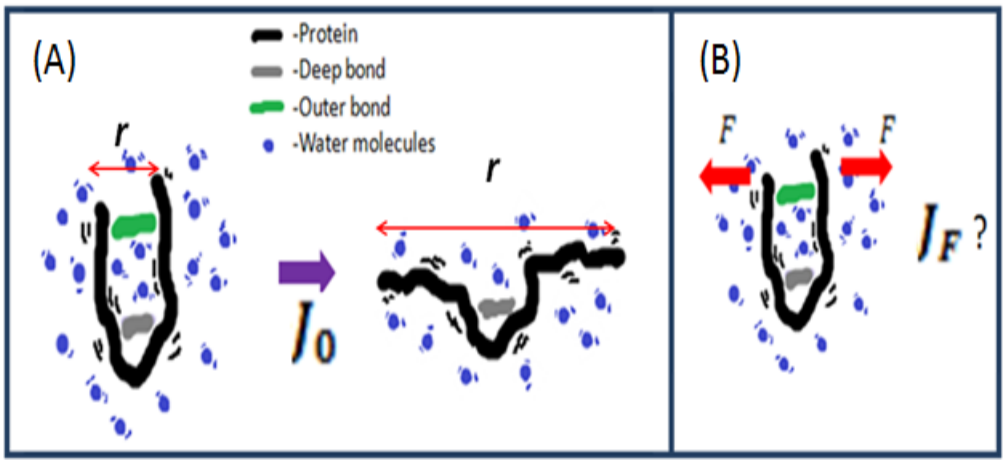


Figure 5: Simplified picture of protein folding/unfolding. (A) Folded protein held by two bonds in a solution. Due to Brownian motion, the outer bond can break (with certain rate, k_0) and reveals the inner part of protein. (B) When exerted by force at two ends, protein unfolds at different rate, k_F. r denotes the distance between protein ends.

Proteins are linear macromolecules, made up of amino acid monomers. As all polymers, proteins are flexible and can be folded into three dimensional structures. The Left Panel in Figure 5 shows a simplified picture. The protein is in constant Brownian motion because it is in solution and is constantly bombarded by water molecules. Thus, the two ends of the folded protein try to move apart to increase entropy, and are only limited by non-covalent bonds and hydrophobic interactions holding them together. However, the bonds have a limited range ($\sim 1 - 10\, nm$). If one end receives a big enough Brownian kick, the outer bond can break, revealing the inner functional structure. This bond breaking happens with a certain rate, J_0. The simplest characterization of protein unfolding structure (project aim 1) is the increase in contour length of the protein after bond breakage. The protein unfolding rate (project aim 2) is more subtle and deals with the kinetic problem of a state represented by the distance between folded protein ends. The state moves in an energy landscape which represents the non-covalent bond [21]. It is shown below.

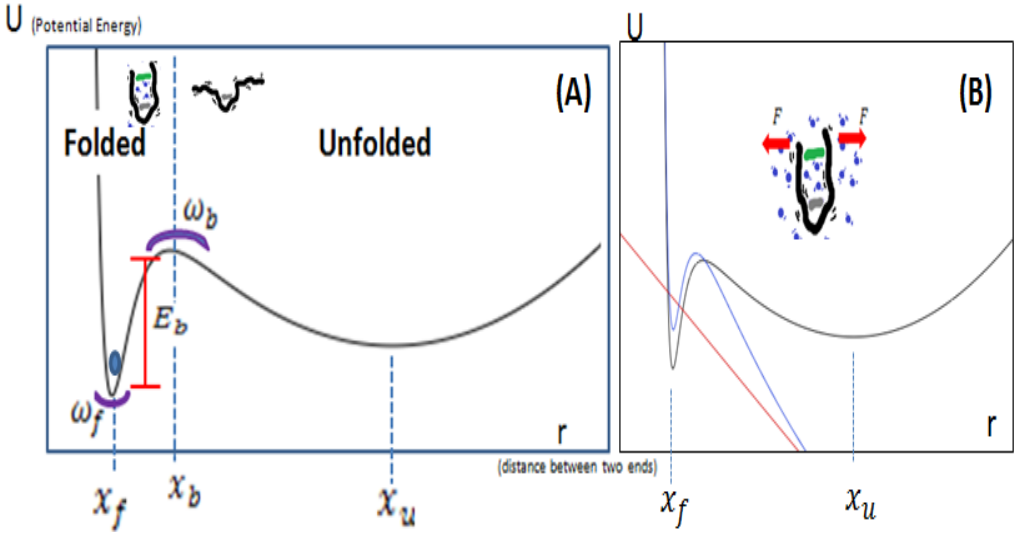


Figure 6:

(A) Bond energy landscape/potential as a function of bond length, r (i.e. distance between protein ends (Figure 5)). Potential energy shows two local minimum. Minimums correspond to folded state (unfolded state), at x_f (x_u) and is separated by barrier (height, E_b), at x_b. The landscape at x_f (x_b) is approximated by (inverted) harmonic potentials with stiffness, $\omega_f\sqrt{m}$ ($\omega_b\sqrt{m}$). The protein is in the folded state (blue circle).

(B) (Black curve) Initial bond potential. (Red line) Constant external force potential. (Blue curve) Modified potential (i.e. sum of bond and external force potential)

The energy landscape of a protein bond, U_0, is shown in Figure 6(A) (similar to [22, 23]). There are two local energy minimums corresponding to the folded (f) and unfolded state (uf) at x_f and x_u. The main feature of the bond is the energy barrier (height, E_b) separating the minimums. Protein unfolding (i.e. bond breakage) corresponds to the $(f) \rightarrow (uf)$ transition, which is probabilistic because proteins are in Brownian motion. Now, take an ensemble of protein with most of them starting in (f), and assume that unfolded proteins cannot fold back. Then Kramer's theory [24] shows that the main factor that influence *the* rate of protein unfolding, J_0, for an average protein depends on the parameters of the bond energy landscape (Figure 6*):* ω_f *(folded state* potential frequency), ω_b (barrier potential frequency), E_b (barrier height). J_0 is driven by diffusion and convection of the two ends of the protein and is given by (details in Appendix A: Kramer's Theory):

$$J_0 = \frac{D\beta m\omega_f \omega_b}{2\pi} e^{-\beta E_b}, \quad (1)$$

where $\sqrt{m}\omega$ is the stiffness parameter of the potential, $\beta = 1/k_B T$, T is absolute temperature, k_B is Boltzman constant. The fact that J_0 is exponentially decreasing with increasing E_b is loosely linked to the equilibrium Boltzman distribution probability, proportional to $e^{-\beta U}$.

Eq. (1) describes the tendency for a protein to unfold when a large energy barrier is exceedingly low. However, the situation changes when we exert an external force on two ends of the protein (Figure 5(B)). Intuitively, the force weakens the bond and helps the protein to unfold. Quantitatively (Figure 6(B)), consider a constant force which introduces an additional potential ($U_F(r)$: red line) so that it tilts the initial bond potential downwards ($U_0(r) + U_F(r)$: blue curve). It is straightforward to show that the only change caused by this constant force is the lowering of the initial energy barrier by exactly $\Delta E_F = -F(x_b - x_f)$. Thus, the new average protein unfolding rate when a constant force is applied, J_F, (with $x^* = x_b - x_f$: distance between energy minimum and barrier) is:

$$J_F = J_0 e^{\beta F x^*}. \quad (2)$$

Eq. (2) provides a direct way to evaluate the importance of force in cell-cell adhesion (i.e. if we accept the definitions of cell adhesion *(D1)*, *(D2)* and *(D3)*). We substitute in physiological values of ($k_B T, F, x^*$) to calculate $e^{\beta F x^*}$. $k_B T$ ($\sim 4 * 10^{-21}\, J$) is the value at physiological temperature $37°C$, $F(\sim 10\, pN)$ is the typical force felt by protein complexes in the body before it breaks [25], $x^*(\sim 1\, nm)$ is typical of the range of hydrophobic interactions [26] governing protein folding and together give: $e^{\beta F x^*} \sim 10$. This suggests that physiological forces can increase protein unfolding significantly (by ten-fold) and so play an equally important role as chemical factors in dictating cell-cell adhesion.

However, Eq. (1) and (2) just give the average unfolding rate. Since the protein unfolding is intrinsically probabilistic, we can assume that there is a survival probability, $S(t) = \Pr\ (T > t)$, related to the protein, which gives the probability that a protein is still in the folded state if we measure it at time, t. $\Pr(T > t)$ is the probability that the exact unfolding time, T happens after the measurement time, t. $S(t)$ is further assumed to obey a first-order rate equation (taking rate constant, J, as the average protein unfolding rate) :

$$\dot{S}(t) = -JS(t), \tag{3}$$

which gives:

$$S(t) = \exp\left[-\int_0^t J dt'\right]. \tag{4}$$

Having seen the description of protein unfolding (with and without force), we can now introduce the two mechanical single-molecule experimental techniques that we use and discuss how they are used to characterize protein unfolding.

2.1.4) α-Catenin molecular structures and properties

We need to be more precise about the molecular details of α-Catenin to appreciate unfolding results. There are several molecular models for the α-Catenin monomer structures and functions, but we choose to describe the most recent one to our knowledge [57] (Figure 7). The model is derived from crystallographic studies and comparisons with Vinculin, which is also its homolog in addition to being its important binding partner.

α-Catenin has four main domains, $D1, D3, D4$ and $D5$. $D1$ and D5 allow binding to different molecular partners to complete the molecular bridge that links two adjacent cells. $D3$ can be further subdivided into two domains $D3a$ and $D3b$, where $D3a$ is mapped to contain the Vinculin binding site by biochemical assays. Although $D3a$ currently has no structural data [57], it is hypothesized to have two

parallel juxtaposed α-helices forming a bundle (Figure 7(B)). Unfolding of this bundle is thought to be crucial for Vinculin binding. $D3b$ and $D4$ forms a modulation domain, signifying that its presence can block the availability of $D3a$ and need to be displaced, either by force or chemical means, for Vinculin to bind. In this project, we work on a recombinant protein construct consisting of only the Vinculing binding and M domain, which is the minimal structure to study forces involved for the mechano-activation of α-Catenin. Another notable fact is that $D1$ and $D4$ can each homodimerize with the same domain on another α-Catenin molecule [57].

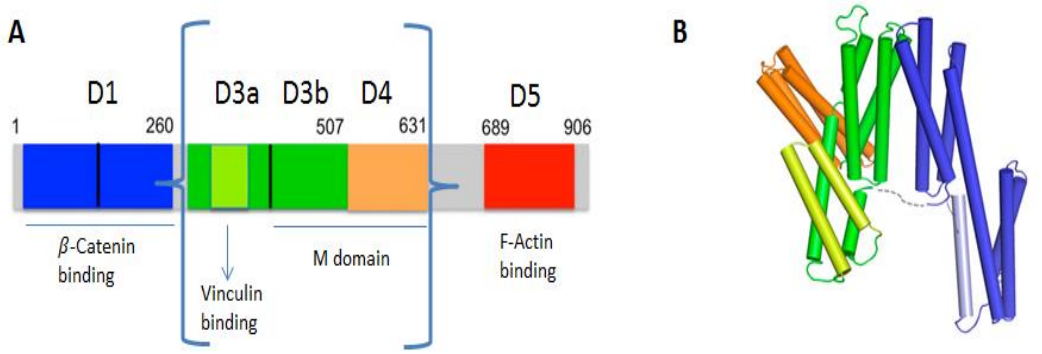


Figure 7: One recent molecular model of α-Catenin monomer. (A) Linear amino acid sequence for α-Catenin monomer, separated into four main domains, $D1$, $D3a, b$, $D4$ and $D5$. Numbers $1-906$ indicate amino acid number. $D1$ and $D5$ bind molecular partners to form complete molecular bridge. $D3a$ contains Vinculin binding site (cyan). $D3b$ and $D4$ form M domain, modulating $D3a$. Our experiments use a recombinant construct of bracketed region, $D3a$ and M. (B) α-Catenin consists of a series of α-helical bundles, color code follows (A). $D5$ position is rather flexible so is omitted to facilitate visualization. Adapted from [57].

2.2) Tool

2.2.1) Atomic Force Microscopy (AFM) and Magnetic Tweezers

Generally, we want to track 1) forces exerted on protein, 2) protein extension as a function of force (i.e. gives unfolding structures and 3) time traces of experiments (i.e. give unfolding rate). Only the basic principles and intrinsic advantages/limitations of both the techniques are described here. Refer to subsection 2.2.2) How to get Unfolding features, for details.

In Atomic Force Microscopy (AFM), proteins in a buffer solution can fix on an open silicon glass slide randomly by strong non-covalent bonds (e.g. biotin-streptavidin) as shown in Figure 8(B). The AFM cantilever is moved by a motorized piezo and its tip approaches the slide to probe for proteins. The cantilever is then retracted from the surface by the piezo to exert force on a possibly attached protein. The approach-retraction cycle is repeated systematically on different points in a given surface where proteins randomly sit. On average, 1 – 10 % of all tip approaches will hit a protein (depending on protein concentration) and the retraction curve gives us information about the protein unfolding.

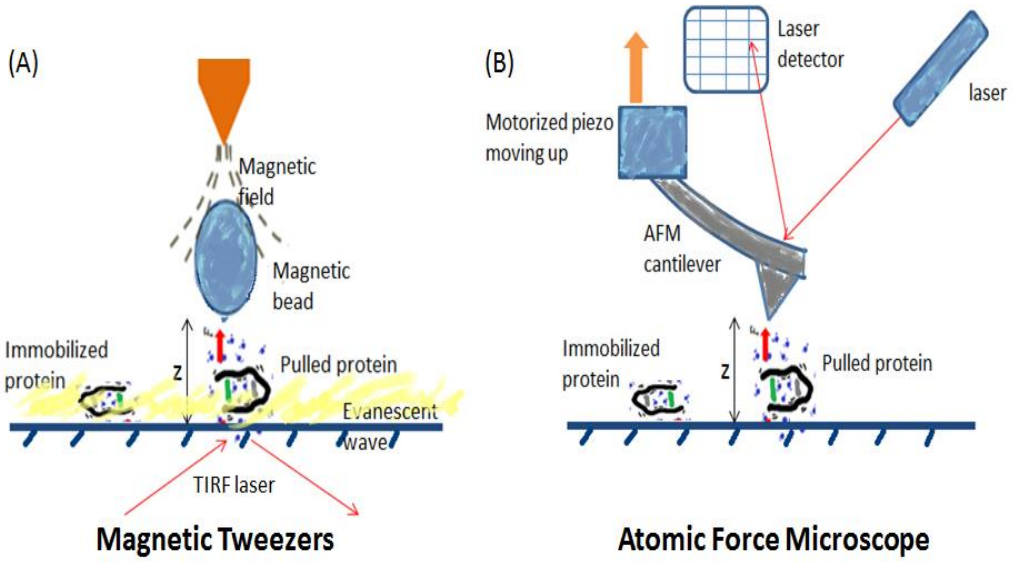


Figure 8: Not drawn to scale. (A) Magnetic Tweezers with Total Internal Reflection fluorescence (TIRF) technique. The immobilized protein attached to paramagnetic bead is pulled by magnetic field which exerts force on the bead. Vertical extension, $\boldsymbol{z}$, of protein measured by evanescent wave from total internal reflected laser beam. (B) Atomic Force Microscopy (AFM). An immobilized protein is pulled by the flexible AFM cantilever controlled by a motorized piezo. AFM cantilever acts like a spring and exerts force on protein. Vertical extension, z, of cantilever is measured by laser deflected from cantilever to a detector. Both techniques have $\boldsymbol{nm}$ resolution and are suited to study protein unfolding steps ~ 10 nm.

The force exerted on the protein is measured by the deflection of the flexible AFM cantilever, which obeys Hooke's Law, $F = kx$, where the spring constant, k is calibrated before the experiment. k ranges from $5 - 30\, pN/nm$. The deflection of the cantilever can be monitored by a laser beam which is reflected from the back of the tip onto a photodiode. The noise of the force

detected is about 10 pN and sets the minimum reliable force detectable. This is consistent with random Brownian deflection of the cantilever given by the equipartition theorem ($k\langle\Delta x^2\rangle = k_B T$ and $F = k\Delta x$). Most importantly, the $\Delta x \sim 1\, nm$ noise in extension measurements sets a resolution good enough for detecting protein unfolding structures. Finally, we want to plot a force versus extension (i.e. z in Figure 8(B)) curve to extract the unfolding structures (explained in 2.2.2) How to get Unfolding features). The extension of the protein is given by the position of the piezo (minus the deflection of the cantilever).

In the Magnetic Tweezers and TIRF technique, the proteins are allowed to fix randomly on a silicon glass slide (ref. Figure 8(A)). Chemically treated paramagnetic beads ($\sim \mu m$) are introduced into a micro-channel by pipetting and can stick to the other ends of the proteins by specific binding. The beads allow us to locate the proteins using a wide-field microscope. Then, we exert a force on the bead (with protein) using an electromagnet. Here, the smallest accurate force that can be exerted is much smaller than that of AFM ($\sim pN$), since it is directly controlled by electric current. Force is measured by observing the variation of the position fluctuations (with a high speed camera, frequency $\sim 500\, Hz$) of the bead in the $x - y$ plane (perpendicular to protein extension) and calculated using the equipartition theorem ($k\langle\Delta x^2\rangle = k_B T$, where $k = F/z$). However, for short molecules like protein, the limitation of the camera frequency coupled with high frequency vibration of the bead (tethered to the short molecule) sets an upper limit for which force on the magnetic bead can be measured with confidence [27]. A camera frame-rate smaller than the bead vibration frequency will underestimate $\langle\Delta x^2\rangle$ and thus overestimates F. For a typical bead size ($\sim 1.5\, \mu m$), the maximum reliable force measured is $F_{max} \sim 15\, pN$.

Also, to calculate force, we need the protein extension, z. This is readily obtained from the exponentially decaying evanescent wave intensity, $\propto e^{-z/l_0}$, of a total internal reflected (TIRF) laser beam which we illuminate the bead from beneath. l_0 is the penetration depth and is typically $\sim 100\, nm$ for a laser wavelength of $\sim 488\, nm$. The intrinsic z resolution is even better than that of

AFM (sub-nanometre) [28] but depends on the thermal fluctuations of the magnetic beads in practice.

In both techniques, the protein has to be pulled upwards to exert a force on it. The pulling can be done in a few ways e.g. constant velocity pulling (i.e. $\dot{z} =$ constant), constant force pulling (i.e. $F =$ constant) and constant loading rate pulling (i.e. $\dot{F} =$ constant). We use constant velocity mode for AFM, and constant force mode for Magnetic Tweezers. Below, we show how to get the unfolding features of the protein from these two pulling modes in practice.

2.2.2) How to get Unfolding features

To extract protein unfolding features from experiments, we show typical protein pulling curves for each technique from our experiments. (Figure 9(A1)) AFM usually operates in the constant velocity pulling mode, and the pulling is best represented by a force-extension curve. As the cantilever is moved up by the piezo (dark red curve), the protein is being stretched and accumulates tension in itself and the cantilever (i.e. increasing cantilever deflection). Each instant where tension is released between two saw-tooth patterns (blue arrows) corresponds to an unfolding event. The physical situation for a saw-tooth pattern is detailed in Figure 9(A2). For Magnetic Tweezers, the most natural way of pulling is the constant force mode, where data is presented in an extension-time curve. When the protein is taut, the z extension of the bead-protein extension stabilises (average over noise) onto a plateau. However, when the protein unfolds, there is first a step increase in extension (red arrows) before it quickly stabilises to another plateau.

Having identified the protein unfolding events on our experimental curves, the next step is to relate them to the unfolding features of a protein. The change in contour length due to unfolding, ΔL can be estimated using the Worm-Like-Chain

(WLC) formula under force [29, 30], widely used to determine force-extension curves of rigid biopolymers i.e. DNA and protein (derivation details in Appendix B: Worm-Like-Chain (WLC) and Extensible WLC Theory).

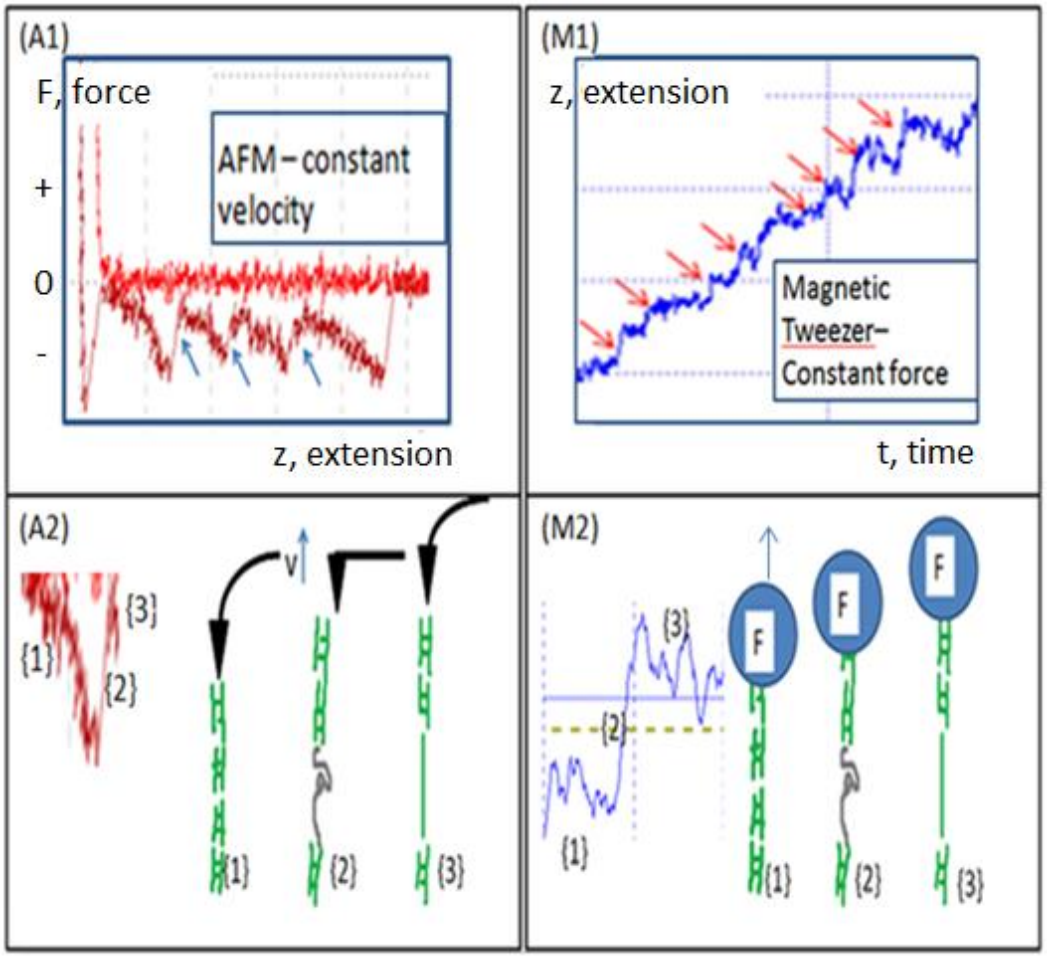


Figure 9: (Cantilever, bead and protein not drawn to scale)

(A1) Typical force-extension curve for AFM constant-v mode which detected protein. (Light red curve) During cantilever approach to surface, cantilever has no deflection (i.e. F=0). When AFM cantilever touches surface, cantilever is deflected upward and force increases positively. (Dark red curve) When cantilever is retracted from surface, the first straight peak shows non-specific interaction with surface bending the cantilever backward (i.e. F negative). After leaving surface, there are saw-tooth patterned peaks corresponding to protein pulling and unfolding. Unfolding corresponds to the straight part between two saw-tooth patterns (blue arrows).

(A2) Zoom into one saw-tooth pattern. {1} Protein (green chain) is pulled and accumulates tension. {2} Protein unfolds and releases tension (i.e. decrease in cantilever deflection). {3} Protein pulling cycle continues.

(M1) Typical extension-time curve for Magnetic Tweezers constant-F mode which detected protein. Curve shows increase of bead-protein extension with time. Protein unfolding corresponds to step increase of the extension (red arrows).

(M2) Zoom into one plateau-step pattern. {1} Protein is taut, thus extension is constant (average over noise). {2} Protein unfolds, and there is sudden (step) increase in protein extension. {3} Protein pulling cycle continues.

The WLC formula is given by:

$$F = \frac{k_B T}{P}\left[\frac{1}{4}\left(1 - \frac{z}{L}\right)^{-2} - \frac{1}{4} + \frac{z}{L}\right], \tag{5}$$

where F is the protein stretching force, P is the persistence length of a polymer, z is protein extension and L is protein contour length. P characterizes the local bending stiffness of a flexible polymer and is typically $\sim 0.1 - 1\ nm$ for proteins. For AFM, we fit eq. (5) to two successive saw-tooth patterns to get their respective contour lengths, L_1 and L_2. Then we calculate the change in contour length with $\Delta L = L_2 - L_1$. To note, WLC does not directly take into account hydrophobic interactions between unfolded sub-domains of a protein, but the effect should be minimal for large unfolding forces [58,59], as we have in AFM studies. Further, we can expect an indirect effect of the interactions to appear in an effective persistence length, P_{eff}, which has also been fitted. The details will not be considered as it is not within the scope of this Masters thesis.

For our Magnetic Tweezers, since F is constant, this gives $(z + \Delta z)/(L + \Delta L) = z/L$. So we can replace (z/L) by $(\Delta z/\Delta L)$ in eq. (5) and obtain ΔL from:

$$F = \frac{k_B T}{P}\left[\frac{1}{4}\left(1 - \frac{\Delta z}{\Delta L}\right)^{-2} - \frac{1}{4} + \frac{\Delta z}{\Delta L}\right]. \tag{6}$$

The unfolding rate determination is dependent on the pulling mode. In constant force mode (Magnetic Tweezers), we fix a force and take the ensemble average of numerous extension-time curves similar to that in Figure 9(M1) for all time points. The averaged curve will smooth out the steps in the individual curves

and can be fitted with eq. (4): $S(t) = \exp(-J_F t)$, since force is fixed. J_F is the inverse of the time constant in the average curve. Repeat the above procedure for different forces and plot $J_F(F)$, which we fit with eq. (2): $J_F = J_0 e^{\beta F x^*}$ to get protein unfolding rate at zero force, J_0. Double check that x^*, distance between energy minimum and barrier is $\sim 1\, nm$. This procedure assumes that protein refolding rate is negligible to the unfolding rate at all times.

In constant velocity mode (AFM), we can use the formula:

$$\beta \bar{F} x^* = \ln\left(\frac{r x^* e^{-\gamma} \beta}{J_0}\right), \tag{7}$$

to get J_0. $\bar{F}$ is the average of all the breaking forces (i.e. force at the tip of the saw-tooth patterns just before protein unfolding) for curves (similar to Figure 9(A1)) pulled at the same piezo velocity, v. x^* is the distance between energy minimum and barrier, $\gamma = 0.5872$ ... is the Euler-Mascheroni constant [31], and $r = kv$ is the loading rate [c], where k can be taken as AFM cantilever stiffness. This procedure assumes v to be large enough (details in Appendix C: AFM Constant Velocity Experiment Design).

2.3) Methods and materials

α-Catenin construct

The recombinant α-Catenin monomer construct is obtained from collaborators in France. The construct consists of $D3a$ and M domains (ref. Figure 7) and is made by PCR and purification from cells. The specific chemistry at both ends are biotin ($D3a$ end) and 6xHis-tag ($D4$ end).

[c] Strictly speaking, the loading rate of protein, r_p should be proportional to the velocity at the protein end and not the velocity of the piezo. But after checking our data, their difference is negligible (i.e. two orders smaller than the values)

AFM Protocol

Prepare slides for AFM:

1. Clean normal glass slides (sonicate slides with DI water, then Acetone/Ethanol, and $1M\ NaOH$, each for 30 min).
2. Incubate the slides in amino-silane (2 %) in Acetone ($\sim 30\ min$).
3. Incubate slides in glutaraldehyde 1 % in DI water ($\sim 2\ hr$).
4. Incubate slides in NTA-amino ($1\ \mu g/ml$) in DI water ($\sim 2\ hr$).
5. Incubate slides in $NiS0_4.6H_2O$ ($100\ mM$) in DI water (whole day).
6. Wash slides with DI water before use.

Prepare AFM setup:

1. Incubate α-Catenin ($\sim 10\ \mu g/m$) (buffer: HEPES $10\ mM$, $NaCl$ $100\ mM$) on treated glass slides. Let it sit for $\sim 30\ min$.
2. Fix AFM cantilever (brand: Nanosensors) on cantilever holder (need careful handling!)
3. Choose force spectroscopy mode on AFM (brand: JPK).
4. Calibrate the sensitivity and spring constant.
5. Start experiment.

Magnetic Tweezers Protocol

Prepare channels on slides for Magnetic Tweezers:

1. Clean normal glass slides (sonicate slides with DI, Acetone, $1M\ NaOH$, DI, each for $30\ min$. Before each new step, use DI to rinse slide.)
2. Heat slides in DI water at 100 ℃ $\sim 30\ min$ before quickly incubating it in (ethanol + silane (5 %) +DI) for $5\ min$.
3. Rinse slides with ethanol and blow dry with Nitrogen gas. (Check for hydrophobicity of treated slides by seeing whether water forms round

droplets or stays flat, more pronounced hydrophobicity after treatment means more successful treatment)

4. Make channels on slide using double sided tape.
5. Do PEGylation (PEG in HEPES and $NaOH$, $1 : 10^6/10^7$) of surface.
6. Do further surface blocking with BSA.
7. Wash channel (with Hepes + NaCl) and inject Neutravidin with α-Catenin, wait for $1\ hr$.
8. Wash channel (with Hepes + NaCl), and inject biotinylated magnetic beads .
9. Wash channel thoroughly (with Hepes + NaCl).

Prepare Magnetic Tweezers setup:

1. Mount treated channels with proteins on the microscope and fix it with tape to reduce drift. Start experiment.

2.4) Results

The bulk of the experiments are done with AFM. However, some Magnetic Tweezers results are presented for comparison so that we can critically evaluate the performance of AFM. The important thing to note is that control experiments are done systematically. For AFM, the negative control is done by performing tests on a bare slide before putting proteins on the same slide to pull again. Without proteins, saw-tooth pattern frequencies were $1 - 2$ orders lower than that with proteins ($\sim 3 - 10$ % of all pulling curves). So we are confident that the data are signals from α-Catenin among others even though we use non-specific interaction between tip and protein to do the pulling.

For Magnetic Tweezers, the idea is to associate the presence of magnetic beads to the presence of proteins. This is done by using specific binding to

selectively bind the beads only to the protein. The control then compares the presence of beads in two channels, one with protein, and the other none, after washing the channels sufficiently to get rid of the protein in the bulk solution. The results are divided into change in contour length during unfolding, ΔL, and unfolding rate. We will then discuss some implications of the results.

2.4.1) Unfolding contour length, ΔL

For AFM, constant velocity experiments with velocities ranging from $100 - 3600\, nm/s$ were performed and each experiment produced $\sim 40 - 80$ curves for analysis. From the colour plots in Figure 10(B), we see that the average forces at the point of unfolding is typically around $60\, pN$. We also see that typical change in contour lengths during unfolding are $\Delta L \sim 10,\ 20,\ 30,\ 40\, nm$. With higher v, the weightage for bigger ΔL increases. However, when $v \to 100\, nm/s$ (tends to the smallest experimented velocity), ΔL coverges to a single value at $\sim 10\, nm$. This suggests that there are multiple bonds binding α-Catenin and that at higher v, multiple bonds tend to break simultaneously and thus ΔL is bigger[d]. However, when v is small, bonds tend to break one by one and since ΔL converges to a single value, this suggests all unfolding structures have the same length, $\sim 10\, nm$. All the arguments are consistent with the requirement that v should not change the intrinsic ΔL of unfolding structures.

[d] This bigger ΔL could also be explained by the possibility that some inner bonds break before the outer bonds but are not detectable because the protein ends are still held by the outer bonds. In this case, when outer bond is finally broken, this information will be added on the already broken inner bond and thus we see a bigger ΔL.

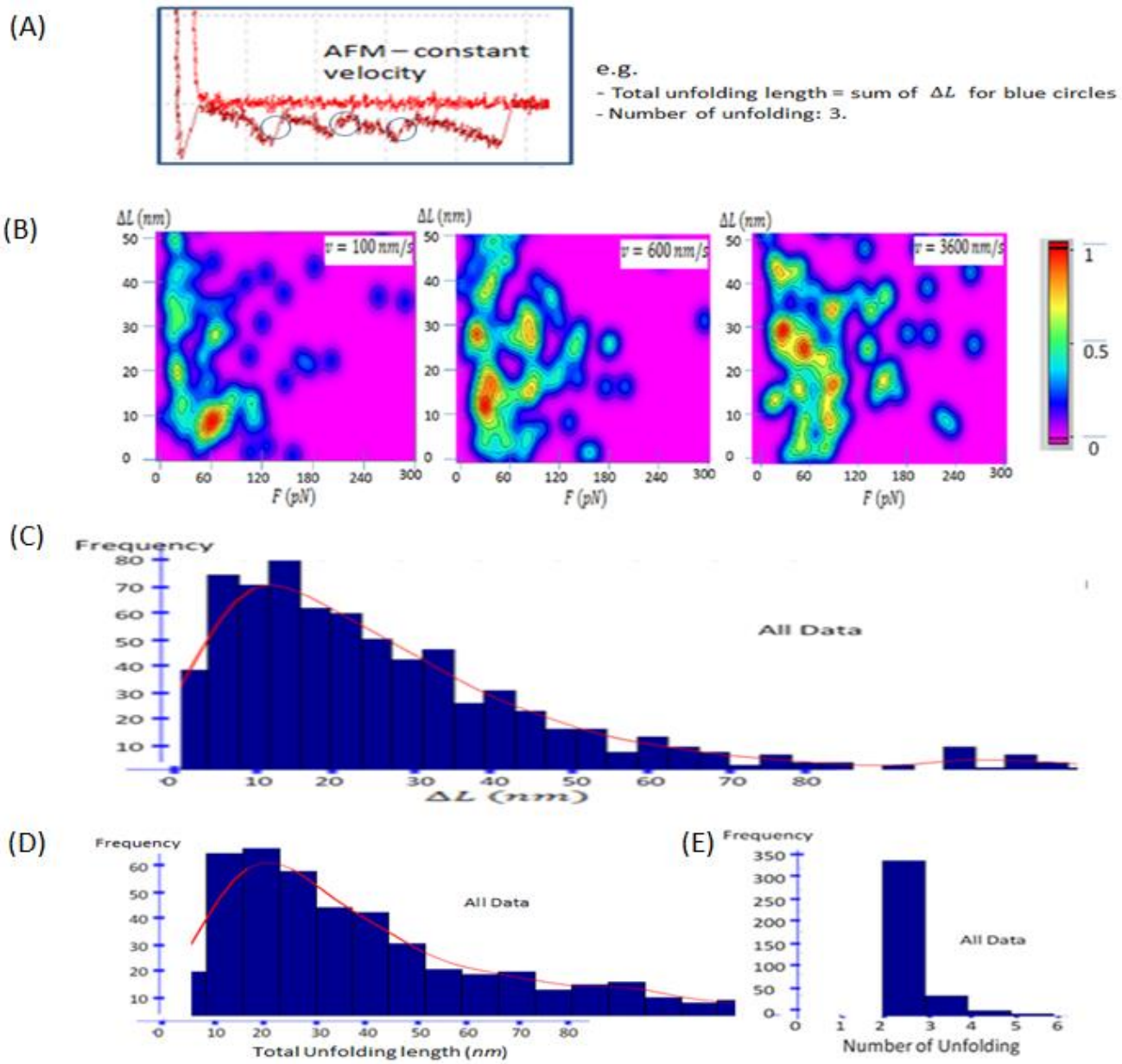


Figure 10: AFM data for unfolding length. (A) Blue circle shows example of data points that we collect i.e. contour length change during unfolding, ΔL, and corresponding unfolding force, F. Other histogram parameters are clearly stated in the example. For each fixed velocity experiment, we analysed ~ 40 – 80 curves. (B) 2D colour graph shows three experiments at different velocity, v $(= 100, 600, 3600\ nm/s)$, plotting ΔL, against F. Colour signifies relative frequency of data, e.g. red means highest frequency. F ranges from $0 - 180\ pN$ for all v. For $v = 100\ nm/s$, there is one red frequency peak at ~ $\Delta L = 10\ nm$. For $v = 600\ nm/s$, there are two red frequency peaks at $\sim \Delta L = 10, 30\ nm$, and one yellow peak at $\sim \Delta L = 40\ nm$.. For $v = 3600\ nm/s$, there are two red frequency peaks at $\sim \Delta L = 20 - 30\ nm$, and two yellow peaks at $\sim \Delta L = 10 - 20, 40 - 50\ nm$. Frequency peaks are shifted to higher ΔL with increasing v. (C) Histogram lumps all ΔL data points of all constant velocity experiments of different v (range $100 - 3600\ nm/s$). There are $500 - 600$ curves and over 1000 data points. Only one single peak at $\sim \Delta L = 10\ nm$. Half width is ~ $40\ nm$. (D) Histogram shows total unfolding contour length (i.e. sum of all ΔL in one curve) per pulling curve for all experiments of different v. Single peak at $20\ nm$, but half width ranges from $10 - 40\ nm$. (E) Histogram shows number of unfolding per pulling curve for all experiments of different v's. Most curves have two unfoldings while some have maximum of six unfoldings.

The argument that all unfolding structures have similar values $\sim 10\,nm$ is further verified when we lump all ΔL values for all experiments into one histogram (Figure 10(C)). The histogram shows one single peak at $\Delta L \sim 10\,nm$. We also plot a histogram of total unfolding length for each experimental curve (i.e. each protein pulling) (Figure 10(D)). We find a single peak at $\sim 20\,nm$ but the distribution is quite flat and the half width reaches $\sim 40 - 50\,nm$. The last histogram (Figure 10(E)) plotting the total number of unfolding per experimental curve just confirms with the colour plots that there are indeed many large ΔL unfolding. This is because we see the histogram peaks sharply at (number of unfolding $= 2$) although the distribution of ΔL and total unfolding length are relatively flat.

Rough inference of the unfolding structures involved in the AFM pulling: Possible multiple unfolding $(4 - 5)$, each with similar length $\sim 10\,nm$. The latter conclusion comes from the fact that smaller v causes ΔL to converge to a single value at $\sim 10\,nm$. The former conclusion is not absurd even though the histogram of total unfolding length peaks at $\sim 20\,nm$. This is because the distribution of this histogram is rather flat and its half width reaches $\sim 40 - 50\,nm$. Also, we suspect some initial unfolding are hidden in the relatively large non-specific interaction between the AFM cantilever tip and the slide surface, shown by a sharp V-shaped peak at the start of pulling (Figure 10(A)) which looks very different from the saw-tooth patterns.

For Magnetic Tweezers, experiments are operated at forces, $F \sim (0 - 15\,pN)$. We do not have ample magnetic Tweezers results as compared to AFM so we just plot a histogram that lumps all ΔL from all experiments at different F (Figure 11). We find a single peak at $\Delta L \sim 10\,nm$. The half width is only $\sim 20\,nm$, only half of that of the corresponding AFM histogram (Figure 10(C)). This better quality of results compensates for the lesser data collected. By experience (e.g. Figure 12), the number of unfolding for each pulling curve is $4 - 8$. These results are similar to the AFM conclusion.

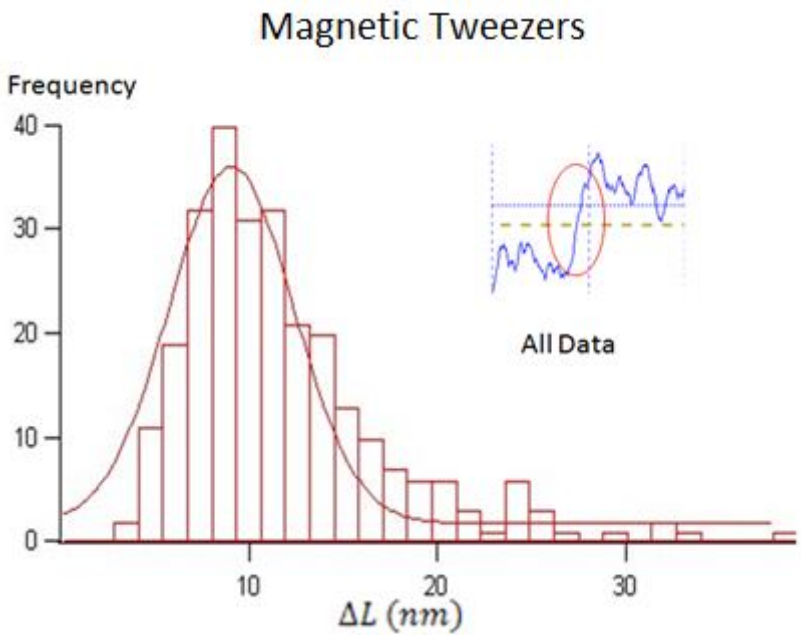


Figure 11: Magnetic Tweezers data for change in contour length at unfolding, ΔL. Histogram lumps all ΔL for all experiments at different constant F (range from $0-15\,pN$). Single peak at $\sim 10\,nm$. By experience, average number of unfolding per pulling curve is $4-6$, e.g. in Figure 12. (work with Lu Chen, a research assistant in Prof. Liu's lab. ~80% of data from Lu Chen)

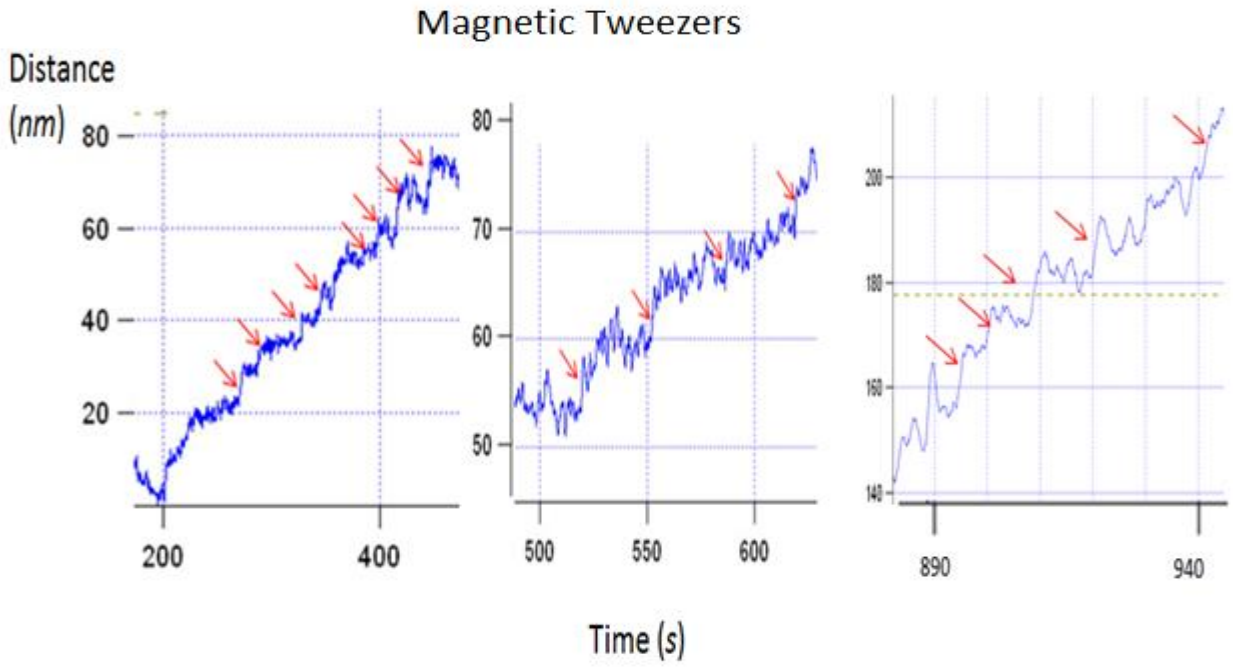


Figure 12: Three typical constant force pulling curves (different forces) for Magnetic Tweezers, plots z extension of bead against time. Red arrows show unfolding steps. On average, all unfolding events finish within duration $50-150\,s$. By experience, most curves follow this trend.

2.4.2) Unfolding rate

For unfolding rate, the data collected for AFM and Magnetic Tweezers is not sufficient for any serious deduction, but we give some comments.

Estimating unfolding rate from the Magnetic Tweezers results is straightforward. According to the typical pulling curves (Figure 12), most curves have all unfolding events finishing within $50 - 150\,s$. This translates to an overall unfolding rate under force: $J_F \sim (\,7 * 10^{-3} - \,2 * 10^{-2})\,s^{-1}$ for force range, $F \sim (0 - 15\,pN)$.

For the AFM, eq. (7) gives a formula to estimate protein unfolding rate at zero force, J_0, from average unfolding forces, $\bar{F}$, of a single velocity pulling experiment. However, we can only fit this equation to the part of the data characterized by large enough velocity [32]. Practically, a large enough velocity regime (different regime for different proteins) can be identified with the regime where the points in the $\bar{F}\ vs\ Ln(v)$ graph start to ascend after a plateau regime, indicated in Figure 13[e]. However, only three data points in Figure 13 fit this criterion. We therefore do not try to fit this data to get an estimate of J_0.

[e] A detailed discussion in Appendix C: AFM Constant Velocity Experiment Design can explain why it is eligible to draw the experimental points into these two regimes, and why the "plateau" regime has significant oscillations of points.

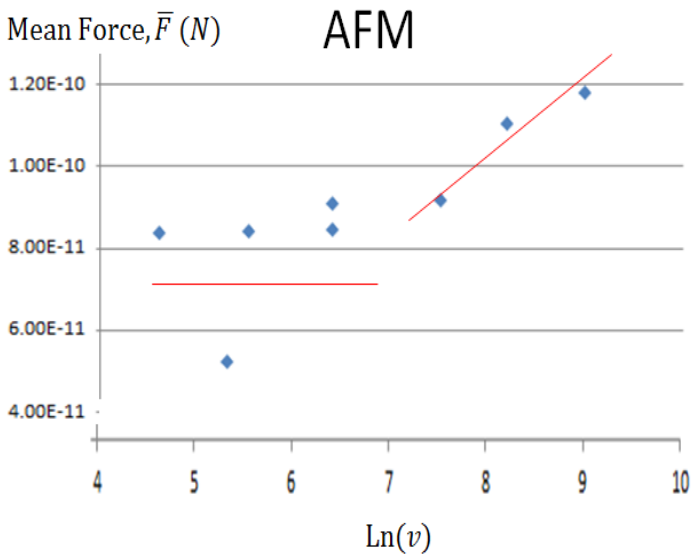


Figure 13: AFM constant velocity pulling data. Each point represents unfolding force average, $\bar{F}$, of all data for one single pulling velocity experiment. Graph plots mean unfolding force, $\bar{F}$ against log of the pulling velocity. Points can be roughly separated into two regimes, one where points hover around a plateau ($Ln(v) < 7$) and another where points steadily increase ($Ln(v) > 7$).

2.5) Discussions

2.5.1) Result Implications and Possible Errors

In all experiments, we try to pull on single recombinant α-Catenin constructs, consisting of the $D3a$ and M domain. Hence, the sudden length increase or force release during the pulling of the protein construct in experiments will theoretically be due to the sudden structural changes in the different domains in this construct monomer. For AFM, non-specific interaction between cantilever tip and protein is used to pick up the molecule. Hence, it is possible for some cases that the molecule is pulled from the middle and we may not get all the unfolding events. This problem is circumvented in the Magnetic Tweezers experiment, since the biotin end of the molecule is bound to the biotinylated

magnetic beads through a neutravidin molecule. The specific interaction between biotin-neutravidin increases the possibility that the molecule is pulled end-to-end, giving all the unfolding events. The structural changes for the molecule could be due to the unfolding of the helical bundle of $D3a$ to reveal the Vinculin binding domain [57] and the displacements of the M domain to release $D3a$. However, to delve into these details requires at least working with several other mutant constructs (with deletion of different domains), which is not within the scope of this Masters thesis.

Here we recapitulate and compare unfolding structure result inference for both techniques:

Table 2 - Summary of α-Catenin unfolding structures

	Number of Unfolding units (per curve)	Unfolding length
AFM	$2-5$	Similar for all unfolding: $\sim 10\ nm$
Magnetic Tweezers	$4-8$	Similar for all unfolding: $\sim 10\ nm$

The unfolding structure inferred from both AFM and Magnetic Tweezers are similar and consistent with each other. Both shows that the unfolding can go through multiple steps, where the basic unfolding units have similar unfolding length, $\sim 10\ nm$. Furthermore, we find that the total unfolding length result is consistent with the α-Catenin construct length scale, $\sim 100\ nm$ ($\sim 0.3\ nm/aa$, Figure 7).

For the unfolding rate, only Magnetic Tweezers can give us a simple estimation from the limited data. J_F ($\max \sim 0.02\ s^{-1}$) is estimated from the Magnetic Tweezers (at physiological forces, $\sim 10\ pN$). The time scale involved

in this forced unfolding rate is well within usual cell reaction time scales (~ minutes), suggesting that α-Catenin is a possible candidate for mechano-sensing at cell-cell adhesion junctions. However, more data and analysis should be made (from AFM and Magnetic Tweezers) to deduce J_0. Only by comparing J_F and J_0 we can better understand the role of physiological forces for increasing α-Catenin unfolding rate and enhancing cellular response.

We address some possible experimental problems that may influence our results. For AFM, two pitfalls are identified. First, the usage of non-specific binding to pull the protein may lead to pulling of random objects in the solution. This is addressed by negative control experiments which assure us that our results are from α-Catenin and not other elements on the slide. Also, our $\bar{F}$ vs $\ln(v)$ data trend follows the theoretical predictions for a chemical bond being pulled by a spring at constant velocity [22, 31, 32, 35]. This suggests that we are indeed consistently pulling on the same chemical bond (i.e. that of α-Catenin). Moreover, the smallest unfolding length is $\sim 10\ nm$ and is always present and independent of pulling speed, v, of the experiment. This is consistent with common sense that the length of a structure should not depend on pulling speed.

Second, the question is whether our results come from dimer or monomer[f] α-Catenin since the protein concentration used in AFM is high ($\sim 10\ \mu g/ml$). In this respect, we check that α-Catenin dimer percentage is only $\sim 30\%$ at our AFM working concentration if we use physiologically determined dimerization and off-rates of the α-Catenin monomers and dimers [36]. We are further convinced that our results come mainly from monomers when we consider results from Magnetic Tweezers. This is because Magnetic Tweezers gives similar unfolding structures as AFM even though its protein working concentration is a few orders lower than that in AFM.

[f] α-Catenin, like many other molecules, is found to form dimers (i.e. two molecules sticking together and could lead to structural changes) at a certain rate in solution. The dimers can also dissociate to reform two monomers at some other rate. The percentage of dimers will be high if the initial concentration of proteins is high and vice-versa. Knowledge of dimer association and dissociation rate allows one to calculate the dimer concentration at equilibrium.

Finally, technical problems from the machine itself can affect our results. Possible errors for both techniques come from force measurement. For AFM, the measurement of force as a function of protein extension hinges on the pre-calibration of the unknown cantilever stiffness. For the commercialised JPK AFM that we use in our experiments, the cantilever stiffness is calibrated by looking at the cantilever deflection (i.e. thermal oscillation method), assuming the deflection is only due to solution molecules. The equipartition theorem ($k\langle x^2\rangle = k_B T$) then links the cantilever stiffness to the deflection. In actual cases, the cantilever deflection can easily be influenced by other noise as well and causes one to underestimate the cantilever stiffness (typical error estimates are at ~ 10 % [37]). This possible loophole can cause a systematic error in force measurements. For Magnetic Tweezers, the force calibration is actually more subtle than discussed previously and could lead to more serious overestimation of the force for short molecules like proteins [27, 38]. However, this does not change the preliminary conclusion that α -Catenin can be a good force transducer because an overestimated F means that J_F can only be bigger, and will continue to be substantially bigger than J_0.

2.5.2) AFM vs. Magnetic Tweezers

The two different techniques used in the project allowed direct comparison of results (especially for unfolding structure). It also allowed us to better understand the strengths and weaknesses of both the techniques in the context of this project. Below is a summary of their complimentary features (from our own experiences) to help in future experimental designs that deal with single molecule unfolding.

Table 3 - Summary of complimentary aspects of AFM (constant velocity mode) and Magnetic Tweezers (constant force mode) from our working experience. Each row compares a complimentary aspect of the two techniques. Orange highlight means disadvantage while blue highlight means advantage.

AFM (constant velocity mode)	**Magnetic Tweezers (constant force mode)**
- Large force	- Small force (closer to physiology conditions)
- Direct unfolding of a whole structure held by multiple bonds	- Structure held by multiple bonds unfold bond by bond
- More dimers (high concentration)	- More monomers (low concentration)
- Theory to extract unfolding features is more complicated	- More straightforward to extract unfolding features
- High throughput	- Low throughput

As shown in Table 3, AFM usually operates under larger forces (at high speed) than Magnetic Tweezers and could be less relevant for single molecule studies. At high speed, AFM shows more simultaneous unfolding of multiple units and could interfere with result interpretation. In contrast, Magnetic Tweezers' small force allows the multiple structures to unfold one by one. Moreover, the protein working concentration in AFM is much higher than in Magnetic Tweezers and leads to more dimer formation, which is undesirable for our study. Lastly, the theory to extract unfolding features from AFM constant velocity mode is more complicated than that of the constant force mode Magnetic Tweezers.

In short, the AFM constant velocity mode has the advantage that it is a high throughput method. However, Magnetic Tweezers is more simple and straightforward in acquiring protein unfolding parameters at physiological forces. Finally, we note that AFM can also be operated in the constant force mode [39], but is more prone to drift and machine feedback problems [10] as compared to Magnetic Tweezers.

2.5.3) Quality of Results

Comparing the AFM and Magnetic Tweezers results, we see that the unfolding structure histogram for AFM has a much larger distribution half width ($\sim 3-4$ times that of Magnetic Tweezers), suggesting lower AFM result quality. This overall inefficiency of the AFM in obtaining reliable data is all the more glaring because AFM has ~ 6 times more data than Magnetic Tweezers, and should in principle have better statistics. From experience, this large uncertainty comes from the difficulty to recognize protein unfolding signals from the force-extension curves. Non-specific interactions between cantilever tip-surface at small distances where protein usually unfolds is the main culprit. We have discussed above that AFM is a high throughput method. However, this advantage can only be presided over other single molecule tools (e.g. Magnetic Tweezers) if the AFM signal recognition problem is reduced effectively.

There are indeed existing methods devised to overcome this problem as discussed in the next chapter. However, they have their own problems and are far from perfect. Most importantly, they are not available for our experiments within this time-frame. In light of this, the development of a better and simple-to-use AFM method for signal recognition will be much appreciated.

3. New AFM-DNA method

We describe the protein signal recognition problem and introduce the current experimental methods which are not perfect. After stating some of the problems with the current methods, we present the new idea which tries to make use of the DNA as a better marker. The idea is described with its proposed advantages over the currently preferred method, i.e. heteromeric polyprotein using I27, discussed in ref. 3.2) Current methods and problems. Some background knowledge is presented for the DNA mechanics before we explain how we will use this in the new method. Finally, we present some preliminary but encouraging results and end with discussions.

3.1) Protein signal recognition problem

The idea of AFM single-molecule force spectroscopy (SFMS) is very simple. A mechanical cantilever touch-and-pull systematically on a surface randomly scattered with the unknown bio-molecule (in our case, protein) that we want to study. The protein unfolding signal is embedded in the force-extension curves produced by the AFM. The intrinsic flaw is that the experimenter chooses the part of the curve which is the protein, before further analysis is done. This human bias can lead to two important consequences: 1) miss out important unfolding information and 2) associate unfolding to the irrelevant part of the curve. Curve fitting with the Worm-like-Chain (WLC) curve can in principle reduce the second bias. However, since only one of the two parameters (i.e. persistence length) in the WLC is subjected to a phenomenological constraint ($\sim 0.3\, nm$ [29]) during fitting, this is not very reliable for recognising protein signals.

Situations that could lead to bad/partial signal recognition are: 1) AFM cantilever tip-surface interactions [42] superposing on the protein unfolding signals when the tip-surface distance is very small ($\sim 1 - 10\ nm$), 2) cantilever pulling on multiple proteins, 3) proteins not picked up strictly at the end and only unfolds partially, etc. In practice, the experimenter needs to resort to heavy statistical analysis by collecting a lot of data. Hopefully, the real protein signal finally stands out among other noise signals in a histogram. In short, there is a lot of uncertainty in result interpretation and also inefficiency in data collection if no further improvements are done to the experimental setup.

3.2) Current methods and problems

The protein signal recognition problem will welcome more effective methods. This is evident from the many recent articles on complex statistical algorithms developed to recognize single-molecule unfolding [43-45] from noise data. In this section (and report), we will concentrate on the current experimental methods trying to solve the problem. There are two typical experimental methods employed currently to reduce human bias from two different aspects: 1) reduce amplitude of unwanted signals (e.g. electrostatic, Van der Waals, etc.) from tip-surface interactions and 2) enhance recognisability of the protein signal.

Reducing unwanted signals is done by allowing molecular blocking agents to adsorb onto the probe and slide surfaces. Examples of effective and commonly used blocking agents are proteins: Bovine Serum Albumin (BSA), polysorbate surfactants[g]: TWEEN 20 and short single stranded DNA (ssDNA) oligomers[h]. These molecules have variable stability under mechanical pulling but are all considerably stable (i.e. will normally not incur more ambiguous signal to the force-extension curves). They introduce additional surface repulsive forces that

[g] Polysorbate surfactants – a class of emulsifiers to stabilize colloids which tends to stick to the surface of a solid.
[h] Oligomers – short macromolecule with a few monomer units.

can cancel out the common initial non-specific attractions between cantilever tip and surface. The origins of the repulsive forces are not clear cut but could be due to entropic, electrostatic, steric and hydration forces [42, 46].

The second method aims to enhance recognisability of protein signals by using protein engineering techniques. In these techniques [47-49], large molecules are constructed by linking several proteins one by one (e.g. using short peptide linkers). These constructs can be classified into two groups: 1) homomeric polyproteins– chain of same proteins and 2) heteromeric polyproteins – chain of different proteins. Homomeric polyproteins serves to multiply the signal from our protein of study, analyte, while HETE aims to attach a well characterized protein (marker) signal to that of the analyte.

More specifically, a homomeric polyprotein is made of N analytes, $(-A-)_N$ and works as a positive control, i.e. if there are m unfolding events for one analyte, there should ideally be $(N * m)$ unfolding events for the homomeric polyprotein. For the heteromeric polyprotein (Figure 14), two different proteins are usually hooked in a sandwich pattern, $(-B-)_N - A - (-B-)_N$, where the marker proteins flank the two ends of the single analyte protein in the middle. Usually, we choose the marker to be a much more stable protein than the analyte, with well characterized mechanical properties (i.e. I27 module of the titin protein[i]). Signals coming from the marker will serve as the fingerprint to identify our construct. Further, the sandwich configuration significantly enhances the probability that the analyte is unfolded if we see $(N + 1)$ unfolding events from the marker, because the analye will most probably be hanging in the solution, 'free' for pulling.

[i] Titin – large proteins that gives elasticity to muscle cells. I27 module – one structural domain in titin.

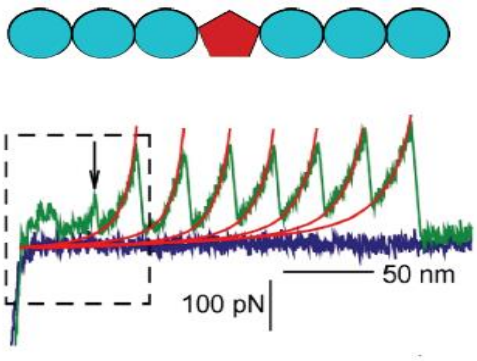


Figure 14: (adapted from [48]) Upper row: sandwich heteromeric polyprotein, with analye (red) and marker (blue). Below: Example of unfolding signal in force-extension curve from the construct. Red line fitted curves are from marker. Black arrow is analyte signal.

Although the use of heteromeric polyprotein is becoming a standard for today's SMFS experiments, it is far from perfect. The possible problems are highlighted in Table 4. Firstly, the proteins in the polyprotein could interact with each other leading to altered protein properties compared to the single monomeric protein [48]. Some polyproteins can even aggregate and not be successfully expressed by the cell. Now, assuming the polyproteins have non-interacting monomers, there are still other outstanding issues. First of all, the marker being a protein itself, gives signals which are probabilistic in nature (e.g. breaking force at each unfolding event is not a fixed value even if we fix the pulling velocity). At the same time, the marker signals also come from unfolding events, so will resemble analyte signals to certain degrees (i.e. can be fitted with the WLC model with similar parameter values). Last but not least, the uncertainty remains about pulling a single or multiple polyproteins. Also the proteins could unfold partially because the proteins could have more than "one point" stuck to the cantilever tip due to small tip-surface separation.

Table 4 - Problems with using heteromeric polyproteins for protein signal recognition at the two different stages (A), (B) in the method.

(A) Expressing polyproteins	1) Possible interactions between proteins in the chain.
(B) AFM Pulling	1) Marker also gives probabilistic, unfolding signals. 2) More than one protein being pulled or individual proteins unfold only partially.

3.3) New AFM-DNA method

3.3.1) Methodology and Advantages

We propose a novel and simple method for the protein SFMS signal recognition problem that can not only circumvent the obstacles in the use of heteromeric polyprotein, but also provide some other advantages. We propose using DNA with its very distinct overstretching transition (ref. next section) as the marker for our protein analyte. The idea is to stick DNAs on the cantilever tip and fish for proteins adsorbed on a glass surface. Both macromolecules are easily functionalized on both the cantilever and glass slide surfaces and can search for each other with the correct chemistry. Figure 15 shows the envisioned simplified scenario.

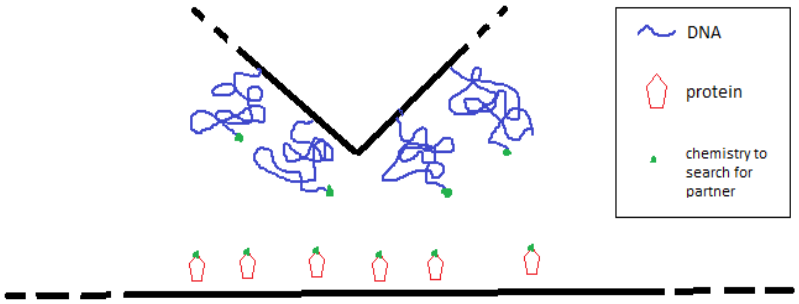


Figure 15: Envisioned configuration of experimental setup. DNA on AFM cantilever tip can search for protein on glass slide with correct chemistry.

The table below shows the problems (mentioned in previous section) in using heteromeric polymers and how the new AFM-DNA method could possibly solve them. Another advantage of using this new method would be the ability to double check the calibration of the cantilever stiffness. The usual calibration is done by the thermal fluctuation method, and can have significant deviations (up to three times difference from personal experience) depending on the real time local fluctuations. Moreover, there will be no problem with the construction of DNAs as it is a very mature field [8].

Problems with Heteromeric polyproteins	**Possible solution with AFM-DNA method**
1. Probabilistic signal and look-alike to analyte	1. overstretching signal is deterministic and very different from analyte i.e.happens as a constant force plateau ($\sim 65\ pN$).
2. Protein stretching at small tip-surface separation ($\sim nm$): (a) Non-specific interactions and (b) Possible partial protein unfolding	2. Protein unfolding likely at much larger tip-surface separation, about the length of DNA ($\sim \mu m$), linked in series with protein.

3. More than one protein pulled each time	3. Cantilever force for overstretching will be roughly quantized ($\sim N * \ 65\ pN$), proportional to number of proteins, N, pulled in parallel.
-	4. Bonus: Overstretching force can help calibrate cantilever stiffness.

Although the new idea can be advantageous in several key aspects of single protein signal recognition, we acknowledge that there are other limitations to it. The most notable would be that it is highly unlikely to use this idea for proteins which can interact strongly with DNA. Below, we discuss the relevant DNA micromechanics and how we can use it to identify protein unfolding signals. We also describe the strategy for immobilizing DNAs for experiments.

3.3.2) DNA micromechanics and How to recognize protein unfolding

DNA micromechanics: The DNA that we are using is the double stranded DNA (usually in B-DNA form). It is made up of two strands of nucleotide bases, denoted as (A, C, G, T). Each strand is complimentary to one another, with the bases paired up i.e. AT and CG. The two strands stick together by hydrogen (H) bonds within the base pairs and inter base pair interactions called base-pair stacking. CG rich DNA sequences are mechanically more stable than AT rich sequences [14]. A picture of dsDNA is in Figure 16(A).

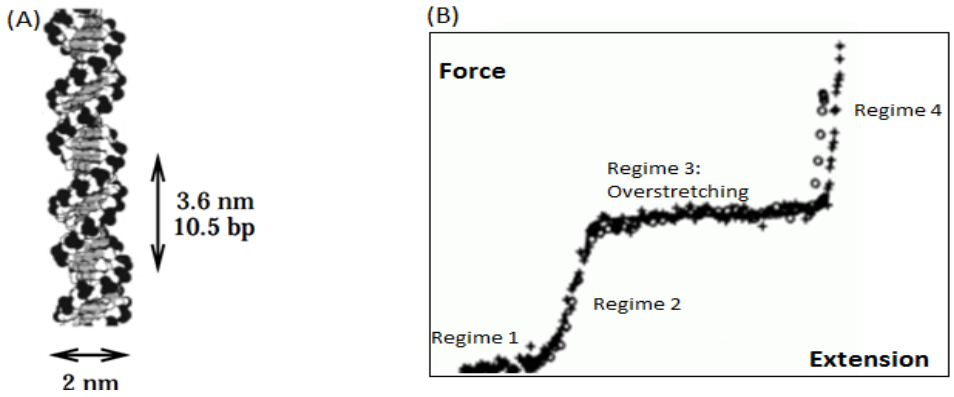


Figure 16: (A) dsDNA double helix and dimensions (adapted from [50]). (B) Typical force extension curve of dsDNA in a SMFS experiment. Regime 1 ($< \sim 10\ pN$) can be fitted with WLC, with persistence length, $p \sim 50\ nm$. Regime 2 and 4 ($\sim 10 - 65\ pN$; $> 65\ pN$) can be fitted with extensible WLC, with different parameters i.e. p and stretching modulus, k. Regime 3 is overstretching plateau, extension ~ 70% of contour length, $F \sim 65\ pN$ depending on experimental conditions: temperature, salt concentration, etc.

Single dsDNA molecules studied in SMFS pulling conditions are quite rigid in terms of bendability and exhibits non-linear responses. More specifically, the extension of the DNA under force can be roughly divided into four regimes [51]: At low forces ($< \sim 10\ pN$), the DNA contour length is almost non-stretched and its elasticity is purely entropic. In this case, the DNA can be reasonably fitted with a WLC model, with parameters persistence length (usually $\sim 50\ nm$) and contour length. At higher forces ($\sim 10 - 65\ pN$), the inter-base-pair distances are stretched, leading to enthalpic elasticity. The DNA extends very close to its initial contour length and can be fitted with an extensible WLC model which has an additional parameter (ref. Appendix B: Worm-Like-Chain (WLC) and Extensible WLC Theory): stretching modulus. Around $\sim 65 \pm 10\ pN$, the DNA extension increases sharply at almost constant force. This overstretching transition allows extension of DNA to $1.5 - 2$ times its initial contour length. For even higher forces ($> \sim 65\ pN$), the DNA continues to extend and can be fitted with another extensible WLC with different parameters.

The overstretching characteristics (e.g. overstretching force and length) depend sensitively on temperature, salt concentration, base-pair content of DNA [8]. All these factors affect the base-pair stability of the DNA, e.g. lower

temperature, higher salt concentration and higher CG content lead to more stable DNA and higher overstretching force. Another obvious effect is the appearance of hysteresis in the force-extension curve during the relaxation of a DNA for less stable DNA. However, most importantly, the variations in overstretching force and length is small and hopefully will not complicate its use as a marker.

How to recognize protein unfolding: When we have a pulling configuration of a protein coupled to the end of a DNA, we have a system of two non-linear springs in series. When the cantilever is pulled, the tension will be shared equally between the protein and DNA, thus stretching both molecules together. When the protein suddenly unfolds at some point, the tension in the DNA will drop simultaneously, corresponding to the DNA's length decrease during this process, (in Figure 16(B)). Protein unfolding is thus tagged with a DNA tension drop, which has to happen in a regime where DNA tension is dependent on length i.e. not in the overstretching regime of Figure 16(B). Despite this, the DNA overstretching plateau is absolutely crucial for easy identification of a DNA-protein pulling event. Only when we are certain that a protein is pulled by a DNA then can we try to search for where the unfolding has occurred in the force-extension curve.

To ensure that protein unfolding happens at the correct regime, we can vary the AFM cantilever pulling speed as the average protein unfolding force is dependent on its loading rate (as Eq. (7)). The ideal regime for protein unfolding to be recognized will be in the stiffer Regime 2 and 4 in Figure 16(B), as a larger DNA tension change happens for the same decrease in length. Finally, the forces involved for unfolding the protein here is $\sim 60\, pN$, which is not far from the conventional AFM forces obtained from our study of α-Catenin shown in Figure 10(B). We remind the reader that these large forces are obtained (compared to the physiological molecular forces of $\sim 10\, pN$ in the cell) because pulling speed is high.

Immobilization for pulling: To pull the DNA, we need to immobilize it on a surface. Strong, specific interactions are chosen for better control. We can

functionalize the DNA at its two ends with specific functional groups to match the surface functionalization e.g. thiol-SMCC (covalent, $\sim > 1000\, pN$) and streptavidin-biotin (one of the strongest known non-covalent protein-protein specific interaction [52-54], $\sim 200\, pN$ depending on loading rate). Also, to get better pulling curves, the use of stable DNAs is crucial. As mentioned above, we can use DNAs with high CG content and/or close the DNA ends by ligation[j] [8]. For stable DNAs going through a overstretching, the mechanism is usually a phase transition called B to S transition, where the B-DNA changes form into a S-DNA.

Below we present the working materials and protocol for the first step to realizing the AFM-DNA method - immobilizing the DNA and pulling it in the AFM setup. Comments are made regarding rationale of choosing the materials and protocol.

3.3.3) Experiment Design

Materials	Equipment
1. 1X PBS (as buffer) ($\sim 137 mM\ Na^{+}$) 2. dsDNA - high CG content (60%), open ended, 3kbp length and both ends functionalised with biotin 3. Glutaraldehyde 1% in DI water 4. Streptavidin (SV) in buffer ($0.1\ mg/ml$) 5. BSA-biotin in buffer ($0.1 mg/ml$) 6. BSA ($>\ 1\ mg/ml$) 7. APTES (1% in Acetone)	1. Plasma cleaner (Harrick Plasma) 2. AFM (JPK) 3. tipless cantilever (Bruker, stiffness: $0.01\ N/m$)

[j] DNA ligase: An enzyme which facilitates the joining of DNA strands.

Protocol

1) **Glass slide preparation:** Cleaning of glass slide (DI water $30\,min \rightarrow$ Acetone $30\,min \rightarrow NaOH\ 30\,min$). Incubation for covalent anchoring (APTES $30min \rightarrow$ Glutaraldehyde $1\,hr \rightarrow$ blow dry using N_2 gas $\rightarrow$ sonicate $20\,min$ in buffer $\rightarrow$ SV $30\,min$, wash with buffer $\rightarrow$ BSA $20\,min$, wash with buffer $\rightarrow$ DNA $5 - 10\,min$, wash with buffer $\rightarrow$ SV $20\,min$, wash with buffer $\rightarrow$ pull in BSA ($0.1\,mg/ml$) - buffer).
   Attention: A) SV to be washed away really clean from bulk solution before DNA addition to avoid DNA aggregation. B) DNA to be washed gently to avoid de-adsorption from surface.

2) **Cantilever preparation:** Plasma $5\,min \rightarrow$ BSA-biotin overnight, wash with buffer $\rightarrow$ BSA 20 min, wash with buffer.

The Biotin-SV bond was chosen as our DNA linkage to surface due to ease of use, high specificity and strength of bond [55]. We cannot have SV at one end and biotin at the other end else the DNA can instantaneously form closed-end loops or aggregate. Hence, double biotin ends are chosen. Glass slides uses glutaraldehyde to bind SV strongly (covalent) at the surface for DNA binding. Electrostatic interaction (by plasma treatment) for binding of BSA-biotin on cantilever is less well controlled but was chosen to facilitate cantilever manipulation. DNA with both ends biotinylated can easily form loops on the SV surface and not be available for pulling. Therefore, SV incubation was done shortly ($5 - 10\,min$) after DNA incubation to block the other biotin end from interacting with the SV surface, and reduce loop formation. BSA incubation is to block non-specific binding at surfaces.

3.4) Results

As the first step to setting up the AFM-DNA method, DNAs were incubated on a chemically treated glass slide as described in the previous section. The incubated slide was stretched and their force-extension signals were analyzed. The expected setup schematic is shown in Figure 17. From the many force-extension curves, three typical curves can be extracted and are shown in Figure 18. Supposedly DNA signals can be seen superposed on background signals (Figure 18).

Pulling statistics: ~ 800 curves are pulled at different locations on the glass slides. Clearly, not all areas have DNA below. Of those which have signals superposed on the background signals, ~ 250 curves were analyzed, ~ 10% have one of the clear plateau signals in the bottom panel in Figure 18, indicating the presence of DNAs. The signal with short plateau is labelled as One-DNA signal, while the other is called Two-DNA signal. The reason for these

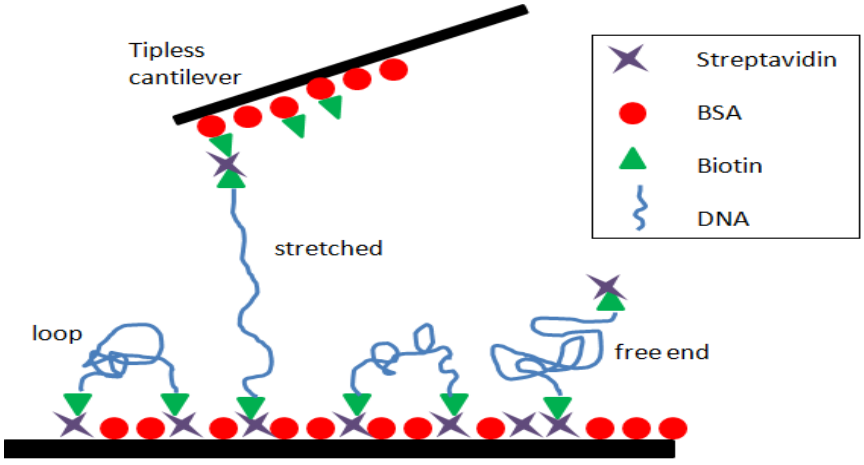


Figure 17: Expected setup schematic for protocol in 3.3.3) Experiment Design. Length scales are not representative. Functional surfaces: BSA-biotin cantilever and Streptavidinated glass slide. Both DNA ends are biotinylated. Biotin and streptavidin have very specific binding affinity and can bind upon meeting. Some DNAs form loops. Some DNAs are capped with Streptavidin and have free end. The latter is available for pulling and stretching.

names will be made clear later. Among the curves with DNA signals, $\sim 75\%$ are One-DNA signals and $\sim 25\%$ are Two-DNA signals.

Analysis: We would like to associate the plateaus in Figure 18 bottom panel and their two stretching tails (before and after the plateau) with the well identified DNA stretching Regimes 2,3 and 4 in Figure 16. To do this, we fitted the second tail after the plateaus with the extensible WLC because forces are large ($> 10\ pN$). Although this model has three parameters, we fixed the contour length (taken as the length just after the plateau ends) and allow only the persistence length and stretching modulus to vary for fitting. This is because the sampling rate of the usually short tail part is not high enough and can give absurd results if we allow all three parameters to vary for fitting. To decrease bias of fixing contour length, several contour lengths were fitted and the averages were taken. Also, we do not try to fit the first tail and the plateau itself because as can be seen in Figure 18 top panel, the background signal is an exponential decay signal for large distances (up to $\sim 1-2\ \mu m$) and interferes with these two parts of the supposedly DNA signals. All fitting are done with JPK DataProcessing and results (averages $\pm$ standard deviations) are shown in

Table 5.

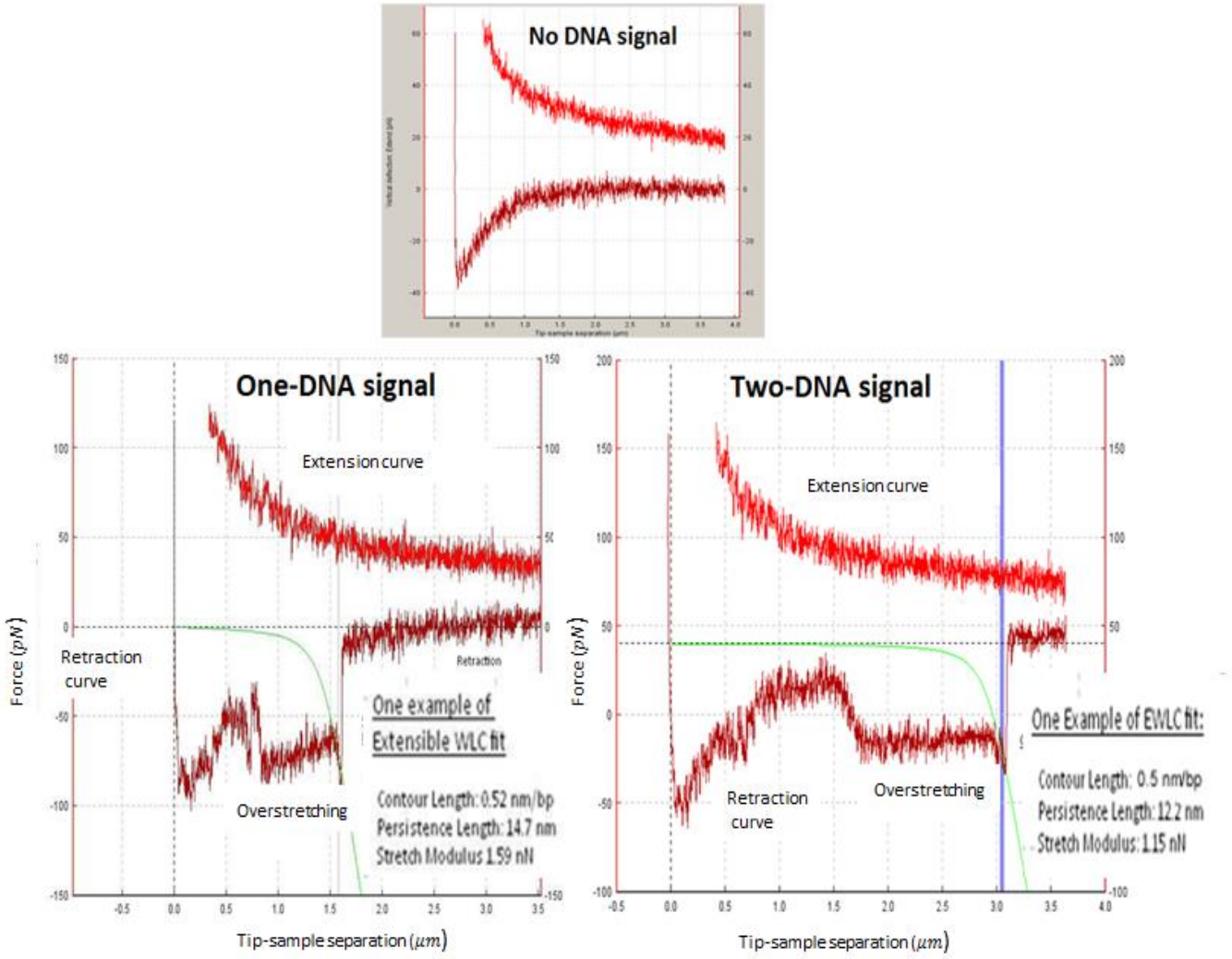


Figure 18: Three typical force-extension curves from AFM pulling using setup in Figure 17. Light red curves show extension of AFM cantilever towards surface, while dark red curves show retraction from surface. Vertical deflection (pN) is not always indicative of real force but has to be normalised by the horizontal dotted line, taken as $0\ pN$. Top panel (No DNA signal): represent ~ 90% of total curves, associated to background force and no DNA being stretched. Bottom panel: Both signals represent ~ 10% of total curves. Among them, ~75% is One-DNA signal, ~25% is Two-DNA signal. Green line fits the short "tail part" of the stretching after the plateau using extensible WLC. Fitted parameters are very similar. Contour length (nm/bp) is calculated using contour length (nm) divided by 3 kbp for One-DNA and 6 kbp for Two-DNA.

Table 5 - Result analysis of tail-like signals after the force plateaus using extensible WLC model.

	One-DNA	Two-DNA
Contour Length (nm)	1.30 ± 0.19	3.09 ± 0.12
Persistence Length (nm)	12.4 ± 5.9	10.2 ± 6.5
Stretch Modulus (nN)	1.62 ± 0.41	1.81 ± 0.60
Contour Length (nm/bp)	0.52 ± 0.03	0.46 ± 0.04

Other important characterizations of our system are:

- Overstretching forces: $50 - 70\ pN$
- Breaking force (force at the end of the second tail after the plateau): $83 \pm 17\ pN$
- Thermal noise: $\pm 10\ pN$

3.5) Discussion

The success of the AFM-DNA idea will depend on several factors: 1) ease of setting up the tool i.e. immobilizing the DNA, 2) ease of use compared to previous methods once it is set up and 3) fulfilling the advantages that were claimed previously.

Firstly, we need to be sure the DNAs are well immobilized on the surface with usual setup procedures (i.e. human pipetting, incubation, etc). The main issues here are whether the DNAs form loops and whether they are still sticking to

the surface after pipetting. Loops can be avoided if we use SV to block one of the ends as in Figure 17. However, the latter concern (i.e. pipetting) remains problematic because the DNAs are long molecules ($\sim \mu m$). The necessary open slide configuration for AFM (unlike possible use of small channels in Magnetic Tweezers setups) leads to harder control of pipetting. This could then generate large shear forces on these long structures and reap them from the surface. To make sure the DNAs are still immobilized on the open surfaces after normal pipetting (rate $\sim 200\,\mu l/5\,s$), micron sized magnetic beads with the correct chemistry (streptavidin beads for DNA with biotin ends) was flown in to label the immobilized DNAs (not shown in results).

Exerting forces on beads by Magnetic Tweezers showed that DNAs remain well attached to the surfaces after normal pipetting (depicted in Figure 19(A)). This is confirmed by smaller fluctuations of the beads at larger forces and a maximum ascend of the bead similar to the length of the DNA used.

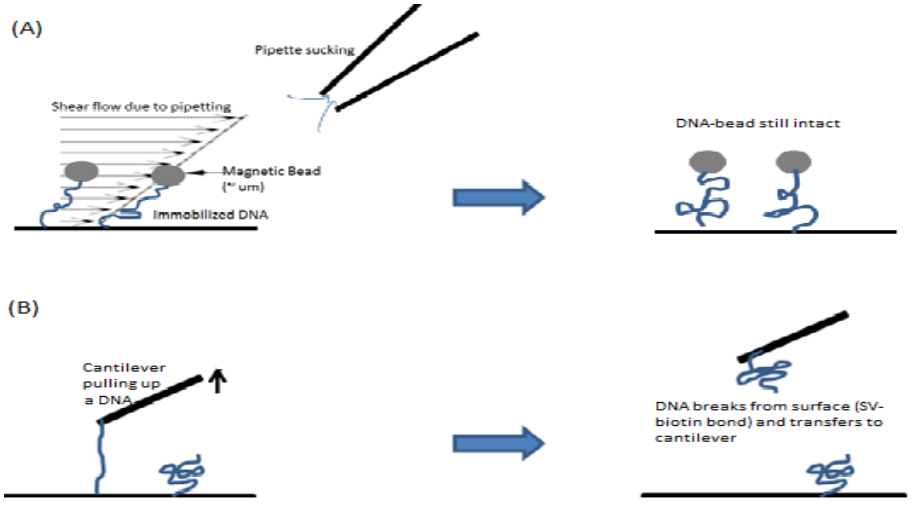


Figure 19: (A) Pipetting fluid induces shear flow on DNA but it is verified that DNA-bead stays intact after normal pipetting. (B) Example of cantilever pulling DNA which eventually breaks at the SV-biotin bond at surface. DNA transferred to cantilever.

A further confirmation of the working DNAs came from the result analysis of the tail-like signals after the force plateaus. There are two types of force plateaus with different lengths which we associate to the DNAs. We hypothesize that the shorter plateau is our single DNA while the longer plateau corresponds to a DNA left in the bulk that has attached in series with a surface immobilized DNA. This is possible for a SV-capped free DNA end. It is tempting to say this because the short plateau length is the one expected for pulling one single immobilized DNA while the longer one is almost twice its length. This is confirmed when we fit the extensible WLC model to the two signals, where they show very similar overlapping persistence length and stretch modulus values, two material parameters which should indeed be independent of the overall structure length. Moreover, if we divide the shorter contour length with 3kbp and the longer contour length with 6kbp, we get again similar contour lengths. More importantly, these fitted values are comparable to that obtained with Magnetic Tweezers at similar experimental materials and conditions in [8] i.e. at $[Na^+]\,150\,mM$: persistence length $13\,nm$, stretching modulus $2700\,pN$ and contour length $0.576\,nm/bp$. Finally, the overstretching forces are around the expected value: $50-70\,pN$.

The next step is to assess the usability of the newly set up AFM-DNA tool compared to the usual methods. A functional and efficient AFM setup depends on high speed pick up and high pick up rate. Both criteria are satisfied because we have $\sim 10\%$ pick up rate and for each pick up, the cantilever just need to touch the surface for $1-2\,s$ at a small force of $40-50\,pN$. Any pick up rate smaller than 10% would be problematic because when we add the proteins that we want to study, the pick-up rate will further decrease because the density of these proteins need to be controlled. A $1-2\,s$ pause at the surface for pick up is also the usual norm for efficient data collecting, contributed by the fast recognition between SV and biotin. Further, we want the DNAs pulled to be immediately reusable after breaking off the surface. This is because we want to incubate the DNAs on a cantilever tip with limited usable surface rather than the glass slide surface in the final ideal product. This is verified when the used cantilever pulls

on a brand new glass slide with only SV (and BSA for blocking) adsorbed on the surface (depicted in Figure 19(B)). We consistently see force plateaus (not analysed in results) indicative of the presence of DNAs on the initially clean cantilever. This shows that these DNA-biotin ends are reusable after SV-biotin breakage.

Finally, the advantages promised by the AFM-DNA method over the use of heteromeric polyprotein with $I27$ have survived the very first experiments. With an extremely well defined and large contour length (standard deviation is $\sim$10% of average) feature, this is an excellent marker candidate that can potentially solve $I27$ problems such as probabilistic and look-alike signals with the analyte, and protein unfolding at small tip-surface separation. Further, with the rather well defined overstretching force range $50 - 70\,pN$ (most probably limited by cantilever thermal noise of $\pm 10\,pN$), this can be used to calibrate cantilever stiffness and also determine how many proteins are pulled at once. However, we acknowledge that we need to involve well-studied proteins with the DNAs in the next phase of experiments to further prove that this idea is workable.

On a side note, there are several interesting observations which we cannot yet explain. Firstly, there is a very short-lived tail section after the force plateau, leading to large standard deviations for the persistence length and stretch modulus with the used data sampling rate i.e. $\sim 1/4 \; - 1/3$ of the average values. This is surprising because studies show that SV-biotin bonds break at only $\sim 200\,pN$ for our typical loading rate in this system i.e. $\sim 1\,\mu m/s$ [52-54]. However, our DNA-cantilever usually breaks off at $< 100\,pN$, giving the tail only an additional length of $\sim 20\,pN$ after the plateau. We are quite sure that most of the break off is at the SV-biotin links and not the surface-DNA links because the used cantilever shows DNA signals when used to pull a new DNA-bare SV slide. Also, the background signal from the tipless cantilever pulling on the surface is very peculiar. Not only is the interaction extremely long range ($\sim 1 - 2\,\mu m$), but there also is a total reversal of force direction i.e. interaction is repulsive during extension of the cantilever towards surface but attractive during retraction from

surface. However, since these interactions are slowly changing (exponential), it does not interfere with the identification of the DNA signals as we have done.

4. Conclusions

Molecular mechanics studies in vitro, although simplistic, have contributed much to our understanding of important life processes at the molecular level. Among the many single molecule force spectroscopy (SMFS) tools, AFM is the most mature and commercialised technique. Yet, from our own working experience (on α-Catenin) and literature review on single protein unfolding, we believe that the AFM SMFS community will gladly invite new, complimentary methods that can improve single molecule signal recognition efficiency. In light of this, we propose a new method, using DNA molecules with its unique overstretching transition to more unambiguously identify protein unfolding signals. Current results on the DNAs alone showed promising trends of using the idea in the usual AFM setting. Having said this, it is crucial for follow-up experiments to verify that DNAs can indeed be successfully coupled to well-studied proteins to obtain expected single protein unfolding characteristics. Finally, apart from the proposed new idea, this thesis can also serve as a short introduction to single molecule biomechanics and two of its working tools (AFM and Magnetic Tweezers).

References

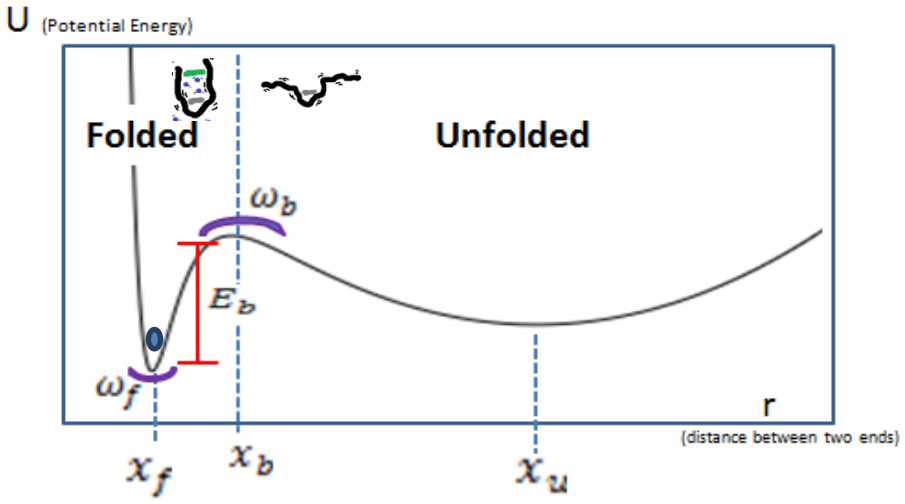
1. Davies, P., *Rethinking cancer.* Physics World, 2010. **23**(6): p. 28-33.
2. Engler, A.J., et al., *Matrix elasticity directs stem cell lineage specification.* Cell, 2006. **126**(4): p. 677-689.
3. Buguin, A., et al., *An array of microfabricated pillars to study cell migration.* M S-Medecine Sciences, 2005. **21**(8-9): p. 765-767.
4. Lim, C.T., A. Bershadsky, and M.P. Sheetz, *Mechanobiology.* Journal of the Royal Society Interface, 2010. **7**: p. S291-S293.
5. Ladoux, B. and A. Nicolas, *Physically based principles of cell adhesion mechanosensitivity in tissues.* Reports on Progress in Physics, 2012. **75**(11): p. 116601.
6. Galbraith, C.G. and M.P. Sheetz, *Forces on adhesive contacts affect cell function.* Current Opinion in Cell Biology, 1998. **10**(5): p. 566-571.
7. Milstein, J.N. and J.C. Meiners, *On the role of DNA biomechanics in the regulation of gene expression.* Journal of the Royal Society Interface, 2011. **8**(65): p. 1673-1681.
8. Zhang, X.H., et al., *Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching by single-molecule calorimetry.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(10): p. 3865-3870.
9. Cloutier, T.E. and J. Widom, *Spontaneous sharp bending of double-stranded DNA.* Molecular Cell, 2004. **14**(3): p. 355-362.
10. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, Magnetic Tweezers and atomic force microscopy.* Nature Methods, 2008. **5**(6): p. 491-505.
11. Gourier, C., et al., *A Nanospring Named Erythrocyte. The Biomembrane Force Probe.* Cellular and Molecular Bioengineering, 2008. **1**(4): p. 263-275.
12. Zhuang, X.W. and M. Rief, *Single-molecule folding.* Current Opinion in Structural Biology, 2003. **13**(1): p. 88-97.
13. Ingber, D.E., *Mechanobiology and diseases of mechanotransduction.* Annals of Medicine, 2003. **35**(8): p. 564-577.
14. Marko, J.F. and S. Cocco, *The micromechanics of DNA.* Physics World, 2003. **16**(3): p. 37-41.
15. Mammoto, T. and D.E. Ingber, *Mechanical control of tissue and organ development.* Development, 2010. **137**(9): p. 1407-1420.
16. Yonemura, S., et al., *α-Catenin as a tension transducer that induces adherens junction development.* Nature Cell Biology, 2010. **12**(6): p. 533-542.
17. Alberts, B., J.H. Wilson, and T. Hunt, *Molecular biology of the cell*. 5th ed. 2008, New York: Garland Science. xxxiii, 1601, 90 p.
18. Hilton, G.R. and J.L.P. Benesch, *Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry.* Journal of the Royal Society Interface, 2012. **9**(70): p. 801-816.
19. Yonemura, S., *A mechanism of mechanotransduction at the cell-cell interface Emergence of alpha-catenin as the center of a force-balancing mechanism for morphogenesis in multicellular organisms.* Bioessays, 2011. **33**(10): p. 732-736.
20. Thomas, W.A., et al., *alpha-Catenin and Vinculin Cooperate to Promote High E-cadherin-based Adhesion Strength.* Journal of Biological Chemistry, 2013. **288**(7): p. 4957-4969.

21. Frauenfelder, H., *Energy landscape and dynamics of biomolecules Extended abstract.* Journal of Biological Physics, 2005. **31**(3-4): p. 413-416.
22. Friddle, R.W., et al., *Near-equilibrium chemical force microscopy.* Journal of Physical Chemistry C, 2008. **112**(13): p. 4986-4990.
23. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The Energy Landscapes and Motions of Proteins.* Science, 1991. **254**(5038): p. 1598-1603.
24. Kramers, H.A., *Brownian motion in a field of force and the diffusion model of chemical reactions.* Physica, 1940. **7**: p. 284-304.
25. Jiang, G.Y., et al., *Two-piconewton slip bond between fibronectin and the cytoskeleton depends on talin.* Nature, 2003. **424**(6946): p. 334-337.
26. Israelachvili, J. and R. Pashley, *The Hydrophobic Interaction Is Long-Range, Decaying Exponentially with Distance.* Nature, 1982. **300**(5890): p. 341-342.
27. Chen, H., et al., *Improved High-Force Magnetic Tweezers for Stretching and Refolding of Proteins and Short DNA.* Biophysical Journal, 2011. **100**(2): p. 517-523.
28. Sarkar, A., R.B. Robertson, and J.M. Fernandez, *Simultaneous atomic force microscope and fluorescence measurements of protein unfolding using a calibrated evanescent wave.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(35): p. 12882-12886.
29. Stirnemann, G., et al., *Elasticity, structure, and relaxation of extended proteins under force.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. 3847-52.
30. Marko, J.F. and E.D. Siggia, *Stretching DNA.* Macromolecules, 1995. **28**(26): p. 8759-8770.
31. Hummer, G. and A. Szabo, *Kinetics from nonequilibrium single-molecule pulling experiments.* Biophysical Journal, 2003. **85**(1): p. 5-15.
32. Noy, A., *Force spectroscopy 101: how to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment.* Current Opinion in Chemical Biology, 2011. **15**(5): p. 710-718.
33. Yonemura, S., *Cadherin-actin interactions at adherens junctions.* Current Opinion in Cell Biology, 2011. **23**(5): p. 515-522.
34. Williams, P.M., et al., *Hidden complexity in the mechanical properties of titin.* Nature, 2003. **422**(6930): p. 446-449.
35. Evans, E. and K. Ritchie, *Dynamic strength of molecular adhesion bonds.* Biophysical Journal, 1997. **72**(4): p. 1541-1555.
36. Dawes, A.T., *A mathematical model of alpha-catenin dimerization at adherens junctions in polarized epithelial cells.* Journal of Theoretical Biology, 2009. **257**(3): p. 480-488.
37. Burnham, N.A., et al., *Comparison of calibration methods for atomic-force microscopy cantilevers.* Nanotechnology, 2003. **14**(1): p. 1-6.
38. te Velthuis, A.J.W., et al., *Quantitative Guidelines for Force Calibration Through Spectral Analysis of Magnetic Tweezers Data.* Biophysical Journal, 2011. **100**(3): p. 481-481.
39. Fernandez, J.M., S. Garcia-Manyes, and L. Dougan, *Force-Clamp Spectroscopy of Single Proteins.* Single Molecule Spectroscopy in Chemistry, Physics and Biology, 2010. **96**: p. 317-335.
40. Shapiro, L. and W.I. Weis, *Structure and Biochemistry of Cadherins and Catenins.* Cold Spring Harbor Perspectives in Biology, 2009. **1**(3).
41. del Rio, A., et al., *Stretching single talin rod molecules activates vinculin binding.* Science, 2009. **323**(5914): p. 638-41.
42. Cappella, B. and G. Dietler, *Force-distance curves by atomic force microscopy.* Surface Science Reports, 1999. **34**(1-3): p. 1-+.

43. Dietz, H. and M. Rief, *Detecting molecular fingerprints in single molecule force spectroscopy using pattern recognition.* Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers, 2007. **46**(8B): p. 5540-5542.
44. Andreopoulos, B. and D. Labudde, *Efficient unfolding pattern recognition in single molecule force spectroscopy data.* Algorithms for Molecular Biology, 2011. **6**.
45. Bosshart, P.D., P.L.T.M. Frederix, and A. Engel, *Reference-Free Alignment and Sorting of Single-Molecule Force Spectroscopy Data.* Biophysical Journal, 2012. **102**(9): p. 2202-2211.
46. Valle-Delgado, J.J., et al., *Interaction forces between BSA layers adsorbed on silica surfaces measured with an atomic force microscope.* Journal of Physical Chemistry B, 2004. **108**(17): p. 5365-5371.
47. Hoffmann, T. and L. Dougan, *Single molecule force spectroscopy using polyproteins.* Chemical Society Reviews, 2012. **41**(14): p. 4781-4796.
48. Uversky, V.N. and S. Longhi, *Instrumental analysis of intrinsically disordered proteins : assessing structure and conformation*. Wiley series on protein and peptide science. 2010, Hoboken, N.J.: Wiley. xxix, 744 p., 16 p. of plates.
49. Carrion-Vazquez, M., et al., *Mechanical design of proteins-studied by single-molecule force spectroscopy and protein engineering.* Progress in Biophysics & Molecular Biology, 2000. **74**(1-2): p. 63-91.
50. Goodsell, D.S., *Illustrating the machinery of life: Viruses.* Biochemistry and Molecular Biology Education, 2012. **40**(5): p. 291-296.
51. Bustamante, C., et al., *Single-molecule studies of DNA mechanics.* Current Opinion in Structural Biology, 2000. **10**(3): p. 279-285.
52. Chivers, C.E., et al., *How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer.* Biochemical Journal, 2011. **435**: p. 55-63.
53. Lo, Y.S., J. Simons, and T.P. Beebe, *Temperature dependence of the biotin-avidin bond-rupture force studied by atomic force microscopy*. Journal of Physical Chemistry B, 2002. **106**(38): p. 9847-9852.
54. Piramowicz, M.D., et al., *Dynamic force measurements of avidin-biotin and streptavdin-biotin interactions using AFM.* Acta Biochimica Polonica, 2006. **53**(1): p. 93-100.
55. Doerr, A., *Protein biochemistry - Streptavidin surprises.* Nature Methods, 2005. **2**(8): p. 573-573.
56. Julius J.Jefferson, Conrad L. Leung, and Ronald K.H.Liem, *Plakins: Goliaths that link cell junctions and the cytoskeleton.* Nature Reviews Molecular Cell Biology 2004. 5, p. 542-553.
57. H.J. Choi, S. Pokutta, G.W. Cadwell, A.A. Bobkov, L.A. Bankston, R.C. Liddington, and W.I. Weis. α *E-Catenin is an autoinhibited molecule that coactivates vinculin*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(22):p. 8576-8581.
58. R.Berkovich, S. Garcia-Manyes, M. Urbakh, J.Klafter and J.M.Fernandez. *Collapse dynamics of single proteins extended by force*. Biophysical Journal, 2010. **98**(11):p. 2692-2701.
59. G.Stirnemann, D.Giganti, J.M.Fernandez and B.J.Berne. *Elasticity, structure, and relaxation of extended proteins under force.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **110**(10): p. 3847-3852.

Appendices

Appendix A: Kramer's Theory



Energy landscape of each protein is a function of its bond length. The potential energy shows two local minimum. Minimums correspond to folded state (unfolded state), at x_f (x_u) and is separated by barrier (height, E_b), at x_b. The landscape at x_f (x_b) is approximated by (inverted) harmonic potentials with stiffness, $\omega_f\sqrt{m}$ ($\omega_b\sqrt{m}$) , m is some mass parameter. The protein moves on this landscape in random motion, so its position is probabilistic.

It is easier to think of an ensemble of proteins where we have different number of proteins, with different bond lengths, $n(r,t)$. The average flux of the number of proteins going from one bond length to the other, j_n depends on diffusion ($= -D\partial n/\partial r$) and convection ($= vn$) of the proteins' two ends, where

D is diffusion constant, v is the convective velocity[k]. Since velocity is related to the drag force, $f_d = \eta v$ (where $\eta = k_B T/D$ is the Einstein relation), we have :

$$j_n = -D\frac{\partial n}{\partial r} - \left(\frac{D}{k_B T}\right) n \frac{dU}{dr}, \tag{A1}$$

where dU/dr represent the force driving the motion (f_d points in opposite direction).

We can rewrite eq. (A1) as:

$$j_n / e^{-U/k_B T} = -D\frac{\partial}{\partial x}\left(\frac{n}{e^{-U/k_B T}}\right). \tag{A2}$$

This just says that the relative diffusion (i.e. relative to equilibrium Boltzman probability) drives the relative flux. Now assume there is a source of protein starting at x_f (i.e. $n(x_f)/n_{total} \propto e^{-\beta U}$ is in equilibrium and obeys Boltzman Distribution), a sink at x_u annihilating the states (i.e. $P(x_u) = 0$) and steady flux, j_n, we can integrate both sides of eq. (A2) by approximating the minimum at x_f (x_b) as (inverted) harmonic potentials with stiffness, $\omega_f\sqrt{m}$ ($\omega_b\sqrt{m}$). The right-hand-side gives just $-D\sqrt{\frac{\beta m\omega_r^2}{2\pi}}$, if we normalize $P(x_r)$ with integration (from $-\infty \to \infty$). Left-hand-side is mostly contributed by the peak around x_b and if we integrate an inverted harmonic potential, we get $e^{\beta E_b}\sqrt{\left(\frac{2\pi}{\beta m\omega_b^2}\right)}$. Rearranging, we finally get:

$$j_n = n_{total}\frac{D\beta m\omega_f\omega_b}{2\pi}e^{-\beta E_b}. \tag{A3}$$

[k] Another name for the diffusion-convection equation is the Smoluchowski Equation.

j_n is larger when ω_f is larger because this allows concentration of states in folded region, increasing probability gradient. j_n is larger when ω_b is larger, because there is larger force at the unfolded region driving the states to the sink. Finally, j_n is smaller when E_b is bigger and is loosely linked to the fact that equilibrium Boltzman distribution probability is proportional to $e^{-\beta U}$.

Appendix B: Worm-Like-Chain (WLC) and Extensible WLC Theory

Polymers are made up of monomer repeats. These monomers move in random fashion subjected to certain physical constraints. The Worm-Like-Chain (WLC) theory assumes the polymer to be an ideal chain (i.e. neglecting all interactions between non-neighboring monomers) with the only energy contribution coming from the bending of neighbouring monomers. If the polymer has a large enough number of monomers, it can be seen as a continuous curve. Thus the energy for a WLC under force, pulling it at its two ends can be written as:

$$\frac{E}{k_B T} = \frac{1}{2}\int_0^L PK^2 ds - Fz, \tag{A4}$$

where L is total contour length, P is persistence length (measures local bending stiffness), K is radius of curvature, F is force pulling the polymer and z is the extension of the polymer in the direction of the pulling force. Now, we want to find the average z extension of a randomly moving polymer under this given force, F. ([30]:Section A) Marko and Siggia derives an approximate formula correct at asymptotically large forces and small forces :

$$F = \frac{k_B T}{P}\left[\frac{1}{4}\left(1 - \frac{z}{L}\right)^{-2} - \frac{1}{4} + \frac{z}{L}\right], \tag{A5}$$

This formula is widely used to analyse DNA at low forces ($\sim < 10\ pN$) and protein force-extension curves.

For other polymers which can have important stretching of their intrinsic inter-monomer distances, there is a stretching modulus, K involved. In this case, the WLC formula is modified to become:

$$\frac{FP}{k_B T} = \frac{1}{4}\left(1 - \frac{x}{L}\right)^{-2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K}. \tag{A6}$$

Or at the high force regime (large extension of polymer), simplifies to become:

$$x = L\left(1 - \frac{1}{2}\left(\frac{k_B T}{FP}\right)^{\frac{1}{2}} + \frac{F}{K}\right). \tag{A7}$$

Eq. (A7) is used to fit DNA stretching with large forces ($> 10\ pN$).

Appendix C: AFM Constant Velocity Experiment Design

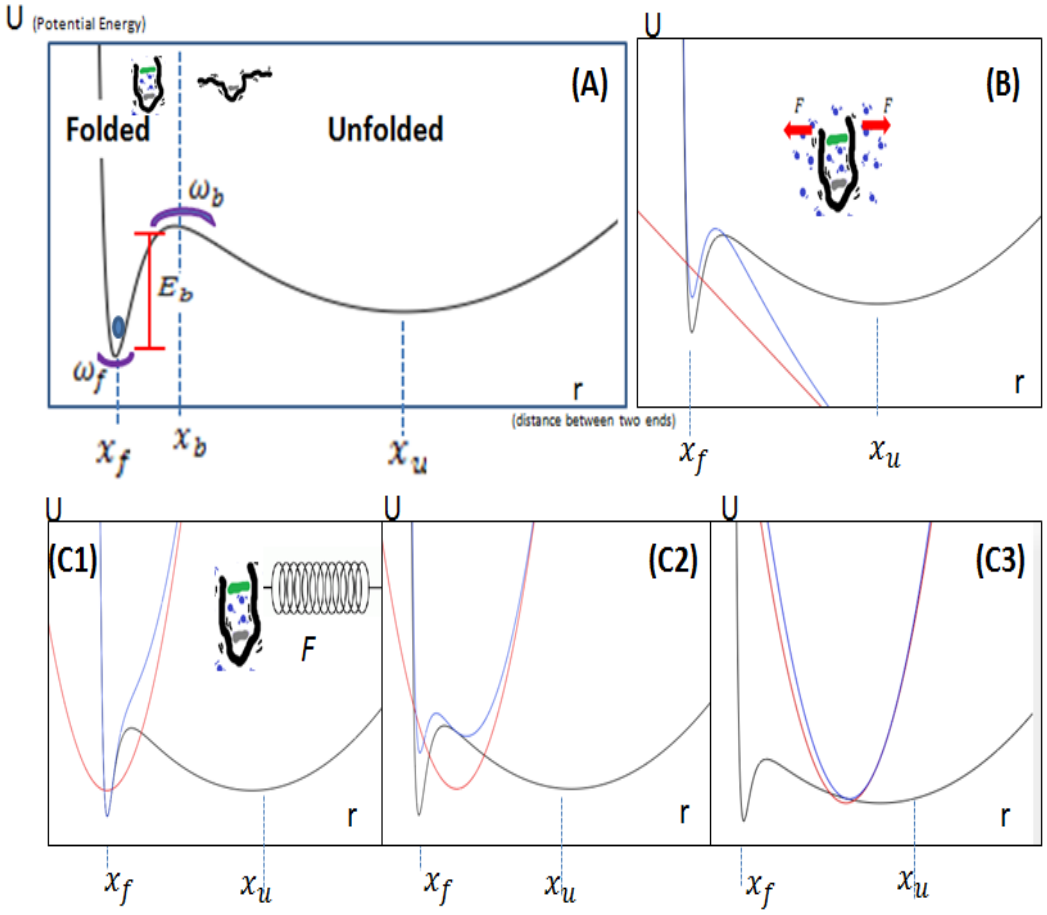


Figure 20:

(A) Bond energy landscape/potential as a function of bond length, $\boldsymbol{r}$ (i.e. distance between protein ends). Potential energy shows two local minimum. Minimums correspond to folded state (unfolded state), at $\boldsymbol{x_f}$ ($\boldsymbol{x_u}$) and is separated by barrier (height, $\boldsymbol{E_b}$), at $\boldsymbol{x_b}$. The landscape at $\boldsymbol{x_f}$ ($\boldsymbol{x_b}$) is approximated by (inverted) harmonic potentials with stiffness, $\boldsymbol{\omega_f\sqrt{m}}$ ($\boldsymbol{\omega_b\sqrt{m}}$). The protein is in the folded state (blue circle).

(B) (Black curve) Initial bond potential. (Red line) Constant external force potential. (Blue curve) Modified potential (i.e. sum of bond and external force potential)

(C1) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, minimum near $\boldsymbol{x_f}$. (Blue curve) Modified potential with only one minimum, close to initial folded state position (i.e. $\boldsymbol{x_f}$).

(C2) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, with minimum between $\boldsymbol{x_f}$ and $\boldsymbol{x_u}$, close to $\boldsymbol{x_b}$. (Blue curve) Modified potential with two minimums. The minimum on the right represents new unfolded state.

(C3) (Black curve) Initial bond potential. (Red line) External spring force, harmonic potential, minimum near $\boldsymbol{x_u}$. (Blue curve) Modified potential with one minimum, close to initial unfolded state position (i.e. $\boldsymbol{x_u}$).

In these constant velocity experiments, we are interested in determining energy barrier height for unfolding, E_b (Figure 20A), and the natural protein unfolding rate, J_0 (eq. (1)). We first present the theory before giving the experimental design guidelines.

In an AFM constant velocity mode, a cantilever (equivalent to spring) is used to pull a chemical bond at constant velocity. This constant v process is always very long compared to the thermal impulses ($< 10^{-12}\ s$) and hence will be very different from the constant force pulling (Figure 20B). In the framework of Kramer's Theory, we can see the spring as a harmonic potential, modifying the initial bond energy as it is pulled away from the surface. Three snapshots of this pulling process are shown in Figure 20(C1), (C2) and (C3). C1 is when the spring is still very close to the folded position, and the modified potential extends infinitely, trapping the protein to be always in its folded state. After moving for some time, the spring reaches the C2 condition and is somewhere near the energy barrier position. The modified potential starts to have two minimum, corresponding to the new folded and unfolded state. In C3, the spring is being pulled further and reaches near x_u. In this situation, the modified potential follows closely the spring potential, and the protein is forced to be in the unfolded state. For practical purposes, most unfolding experimental cases will only be dealing with the C2 situation. This is because there can never be unfolding in C1 , while for most pulling velocities, the protein would have already unfolded before reaching C3.

Focusing on C2, we can separate the speed of a moving harmonic spring potential into two regimes (i.e. slow and fast). In the slow velocity regime, the system is in equilibrium and the modified potential allows similar folding and unfolding rate. Both these rates can be approximated by Kramer's equation (eq. (1)) , but with different corresponding barrier heights. A detailed illustration is given in [22]. By equating these rates, we can get the average protein unfolding force:

$$\bar{F} = \sqrt{2k\left[E_b - k_B T ln\left(\frac{\omega'_u}{\omega'_f}\right)\right]}, \tag{A8}$$

where k describes the stiffness of the spring harmonic potential, and ω'_f (ω'_u) describes the stiffness of the modified potential near the minimum of the folded (unfolded) state. Note that the average force depends on the cantilever stiffness but not the velocity of pulling!

In the high velocity regime, we assume unfolding rate much larger than refolding rate, because the barrier for unfolding is sufficiently small now. In a quasi-static framework, we can just take the survival probability defined in eq. (4), with a time dependent unfolding rate, $J(t)$, due to the moving potential. Since we can approximate $\beta F(t) = kvt$, where k is again spring stiffness, this gives $J(t) = J_0 e^{\beta kvtx^*}$, according to eq. (2). We thus have the survival probability:

$$S(t) = \exp\left[-J_0 \int_0^t exp(\beta kvx^* t')\, dt'\right]. \tag{A9}$$

The mean force at unfolding (or breaking) in this high velocity regime is then:

$$\beta\bar{F} = kv\bar{t} = kv \int_0^\infty S(t)\, dt. \tag{A10}$$

After some manipulation, we retrieve eq. (7): $\beta\bar{F}x^* = \ln\left(\frac{rx^*e^{-\gamma}}{J_0}\right)$, with $r = kv$. Notice that the average breaking force is linear with $\ln(v)$.

Putting the equations of the two regimes together on a $\bar{F}$ vs $\ln(v)$ graph, we should have a regime where $\bar{F}$ is almost constant (corresponds to low velocity), and a regime where $\bar{F}$ grows linearly (corresponds to high velocity). This explains why we can group our AFM results into the two regimes (Figure 13). The oscillation of the points in the slow velocity regime can also be explained if we remember that $\bar{F}$ depends on cantilever stiffness, k at this regime, eq. (A8). Typical commercial cantilevers have a variation of k ($5 - 30\ pN/nm$) and this variation is enough to change the $\bar{F}$ by a factor of 2. [32] gives a good graphical presentation of these discussions.

Based on this theoretical framework, we can state the general constant velocity experimental design to determine bond breakage for any chemical bond (including protein unfolding).

1. Use appropriate chemical specific binding on the tip to search specifically for the protein.
2. Do the pulling over a large speed range ($\sim$ 3 orders of magnitude) to locate the two regimes.
3. Focus on the slow velocity regime for finding the energy barrier, E_b.
4. Focus on the high velocity regime for the unfolding rate, J_0 and distance between folded minimum and energy barrier, x^*.