High Precision AFM-Based SMFS of Mechanically Labile Type III Secretion

System Effectors

By Marc-André LeBlanc

B.S. Lyon College, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

2018

This thesis for the Doctor of Philosophy degree entitled:

High Precision AFM-Based SMFS of Mechanically Labile Type III Secretion System Effectors

written by Marc-André LeBlanc

has been approved for the Department of Chemistry and Biochemistry

_____

Marcelo Sousa

_____

Tom Perkins

_____

Date

The final copy of this thesis has been examined by the signatories, and we

find that both the content and the form meet acceptable presentation standards

of scholarly work in the above mentioned discipline.

Abstract

LeBlanc, Marc-André (Ph.D. Chemistry and Biochemistry)

High Precision AFM-Based SMFS of Mechanically Labile Type III Secretion System Effectors

Thesis directed by Professor Marcelo C. Sousa

Pathogenic bacteria have developed a wide range of tools for circumventing or overcoming the host's defenses. Over time, these tools have become increasingly complex, allowing bacteria to live and thrive within a wide variety of host environments. One such tool is the Type III Secretion System (T3SS), a needle-like complex that allows bacteria to directly inject proteins, known as effectors, from their cytoplasm into host cells. Once inside host cells, effector proteins have a wide range of effects, from shutting down the host immune response to rearranging the host cytoskeleton to accommodate invading bacteria.

Because the T3SS needle presents a narrow channel (< 2 nm), effector proteins must be mechanically unfolded before passing through. Proteins are unfolded by a molecular motor that associates with the base of needle and pulls protein into the channel. While this motor can unfold and secrete many proteins, it is unable to unfold proteins that have high mechanical stability. This indicates a need for effectors to be mechanically labile no matter their function. This may be one of the reasons effectors have very low sequence and structural similarity to other members of their protein super-families. This spurred our investigation into how effectors respond to mechanical force. To investigate effector protein stability, I used atomic force microscopy (AFM) to mechanically unfold the proteins.

Here I show that effector proteins of the T3SS unfold at very low force, despite containing a wide variety folds and functions. This supports our hypothesis that to facilitate efficient secretion, effectors evolved to be mechanically labile. Because effector proteins unfold at such low force, it was critical for me to utilize site-specific attachment to both the AFM tip and surface, increasing both the amount of data I could collect and the quality of collected data. Site-specific attachment resulted in a 70-fold improvement in the yield of high quality data, allowing rapid characterization of mechanically labile α-helical proteins.

Combining site-specific attachment with modified cantilevers allowed the collection of unfolding data for 5 effector proteins, finding they all unfold at low force (<20 pN), making them some of the most mechanically labile proteins studied to date by AFM-SMFS. Comparing the mechanical stability of effector proteins to their *in vivo* secretion rates, showed that unfolding force does not always correlate with *in vivo* secretion rate. However, the distance to the transition state does correlate with *in vivo* secretion rate.

To elucidate how effector proteins have evolved to be efficiently secreted, the mechanical stability of an effector protein, NleC, was compared with a non-secreted homologue, protealysin. While the initial unfolding event of NleC occurs below the detection limit of our AFM platform, the unfolding of an intermediate along the unfolding pathway was measured. When compared to the unfolding of protealysin, the unfolding intermediate of NleC was shown to be significantly less stable, supporting our hypothesis that effector proteins have evolved to unfold at low force to facilitate efficient secretion. We now have an array of tools that allow for the efficient mechanical characterization of diverse proteins with high-precision.

Acknowledgements

I would like to thank my thesis advisor, Marcelo Sousa, for the academic freedom to pursue the projects that interested me the most. His guidance and advice were critical to the work presented here. I would also like to thank my secondary advisor, Tom Perkins, who welcomed me into his lab and was a constant resource as I learned atomic force microscopy. Without the collaboration of the Perkins lab, the work presented here would not have been possible. I want to thank the members of my thesis committee, Jim Goodrich, Joe Falke, Michael Stowell, and Amy Palmer for their support and advice.

I owe a huge debt of gratitude to the other members of the Sousa lab. Dr. Michelle Turco who started the investigation into the Type III Secretion System, Dr. Alex Hopkins for his advice and critical discussion, Dr. Pamela Arden Doerner-Barbour for assisting me in an uncountable number of experiments, and Megan Mitchel for general discussions of biochemistry and biophysics. I also had the privilege of working with excellent undergraduate students Isabel Lippincott and Morgan Fink, who brought fresh eyes to my research projects.

The gracious assistance provided by the Perkins lab was instrumental in my ability to perform AFM experiments. This included my initial training from Dr. Matt Siewny, analytical help from Patrick Heenan, surface and tip functionalizations developed by Dr. Rob Walder, and tip modifications developed by Dr. Devin Edwards. Every member of the Perkins lab challenged me to think more deeply about my project and lent advice wherever they could.

My love of science was kindled by my brother Gabriel, and endlessly encouraged by my parents, Sharon and Gervais. Most importantly could not have made it this far without the

Table of Contents

List of Tables

List of Figures

Chapter I Introduction to the T3SS and Single Molecule Force Spectroscopy

The Type III Secretion System Relies on Mechanically Unfolding Proteins *In Vivo*

The Type III Secretion System Apparatus

The Type III Secretion System (T3SS) is an essential virulence factor in a wide range of gram-negative bacteria, including many common strains such as *Shigella*, *Salmonella*, *Yersinia*, and *E. coli*[1]. At the core of the T3SS is multi-subunit machine responsible for secreting proteins directly from the bacterial cytoplasm into the cytoplasm of host cells known as the injectisome
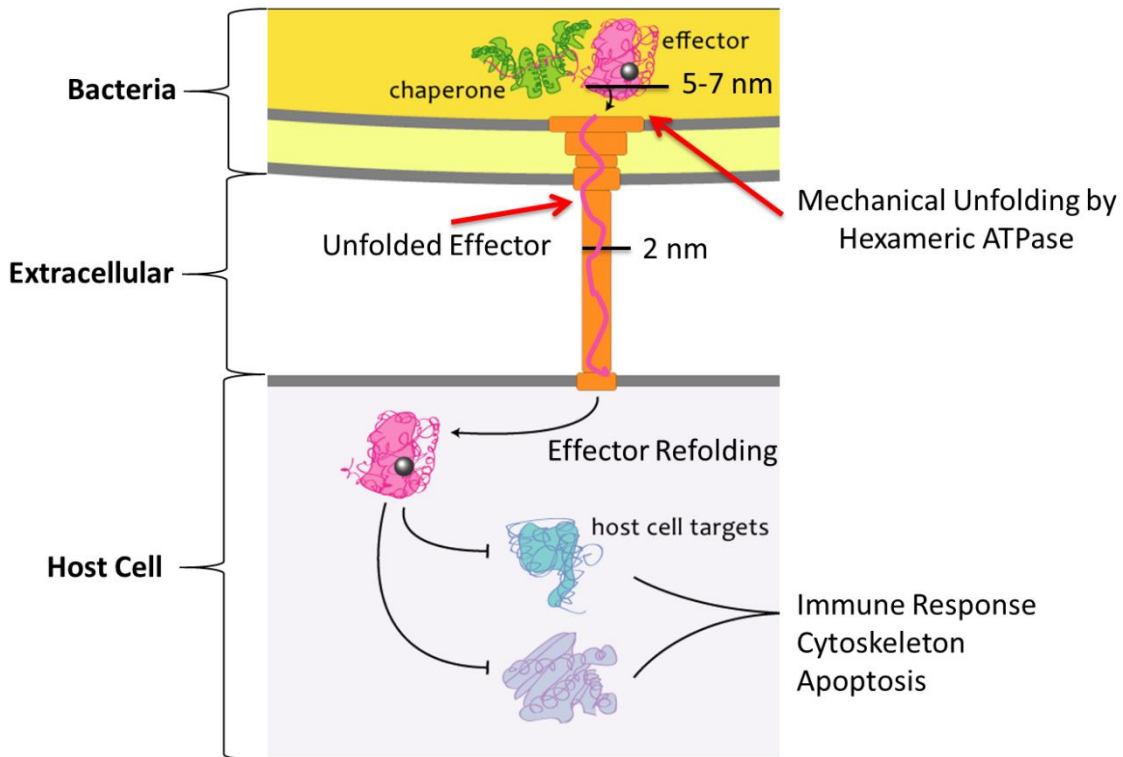
Figure I-1. Schematic of the Type III Secretion System (T3SS). The T3SS consists of a needle complex which spans both the inner and outer membranes of gram-negative bacteria. Substrate proteins known as effectors are expressed and able to fold in the cytoplasm, where they bind chaperones and are brought to the base of the needle complex. To pass through the needle, effectors must be mechanically unfolded by an associated ATPase. Once inside the host cell, effectors can have a wide range of effect, ranging from shutting down the host immune response, rearranging the cytoskeleton, or even inducing apoptosis.

(Figure I-1). Resembling a syringe, the tip of the injectisome combines with translocase subunits open a pore in the host membrane. At the base of the injectisome, cytoplasmic protein complexes are responsible for recognizing substrate proteins, ensuring proper assembly of the needle and preventing premature release of substrate proteins. Secretion is powered by a dedicated ATPase that associates with the base of the injectisome.[1–4]

Chaperones Maintain Effector Proteins with an Extended N-terminal Domain

Substrates of the T3SS, known as effector proteins, are expressed in the bacterial cytoplasm where they are typically bound by chaperones. The chaperones of the T3SS are specialized to bind the N-terminal 50-100 amino acids (the chaperone binding domain) and maintain this region in an extended, unstructured state. While the chaperone binding domain remains unstructured, the C-terminal portions of effector proteins are able to fold into their native conformation.[5] Effector proteins are targeted for secretion by an N-terminal signal sequence but various experiments have demonstrated that chaperones also play a role in effector targeting and sorting.[6,7] Once targeted to the base of the injectisome, the chaperone must be displaced, and the effector is secreted from N- to C-terminus through the T3SS needle.

Secretion Through the T3SS Requires Mechanical Unfolding by a Specialized ATPase

In order to pass through the T3SS needle the effectors must pass through a narrow, needle-like channel with an inner diameter less than 2 nm (Figure I-2A)[2,3,8]. Yet most effector proteins have folded C-terminal domains with a diameter greater than 4.5 nm. Because of this, proteins secreted through the T3SS must be at least partially unfolded to make it through the needle[4,9,10]. This is done by an ATPase associated with the base of the needle (Figure I-2B). The T3SS associated ATPase is proposed to unfold protein through a mechanism similar to AAA+

2
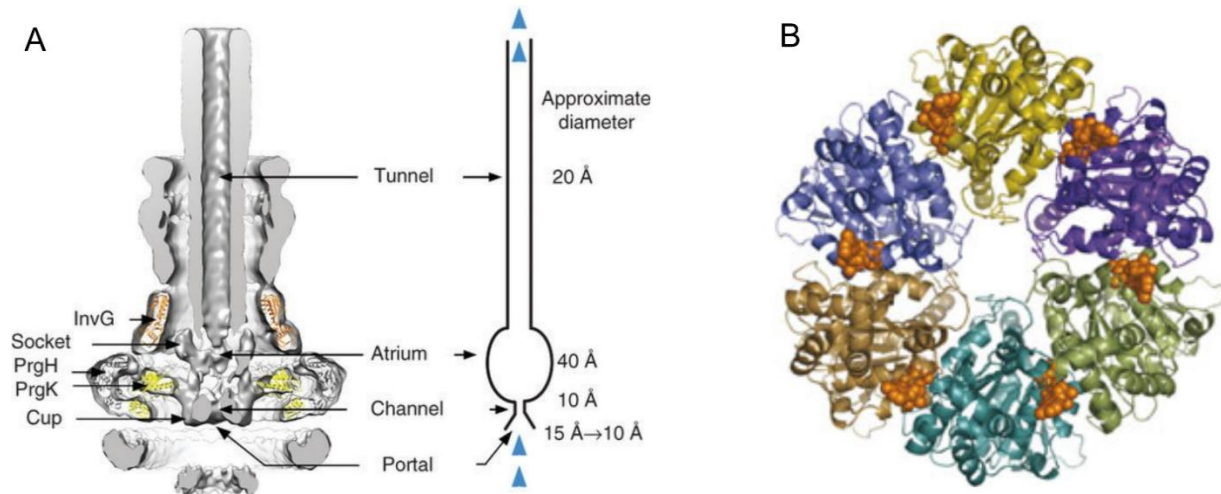
Figure I-2. (**A**) Cryo-EM structure of the T3SS from *Salmonella*, from *Radics et al., Nat Struct Mol Biol* (2014). (**B**) Proposed structure of hexameric ATPase EscN from Enteropathogenic *E. Coli*. From Zarivach, R. et. al, *Nat Struct Mol Biol* (2007).

ATPases, such as ClpX.[2,6] Similarly to AAA+ ATPases, the T3SS ATPases are proposed to form hexameric rings, through which effector proteins are pulled.[2] However, while the mechanism is similar to AAA+ ATPases, the T3SS associated ATPases are more structurally similar to the catalytic β-subunit of the F_1 ATPase, a rotary motor used for ATP synthesis.[2,11] Whereas AAA+ unfoldases are capable of unfolding even very tightly folded substrates such as GFP and ubiquitin, studies of the T3SS have found that it is unable to secrete such stable substrates. Importantly, studies have shown that the associated ATPase can't unfold any protein with a secretion signal, failing to secrete very stable proteins such as GFP[6]. We hypothesize that there is an evolutionary pressure for proteins secreted by the T3SS to unfold at low force, enhancing their ability to be secreted.

T3SS Effector Proteins Have Diverse Functions and Unique Sequences

The proteins secreted through the T3SS, known as effectors, have a wide range of effects necessary for pathogenesis.[12,13] They can cause cytoskeletal rearrangement, shut down the host immune response, and even result in cell death[14]. While the general mechanism of secretion is

3

conserved across a range of pathogens, effector proteins vary between strains.[15–17] Interestingly, most effector proteins share greater sequence and structural similarity to other effector proteins than to members of their protein superfamilies.[18,19] While this may not be particularly surprising given the N-terminal signal sequence and chaperone binding domain, it may also point to the unique constraints placed upon effector proteins. It does not seem unreasonable that efficient secretion through the T3SS would require specialized structural consideration that enable effectors to unfold at low force.

While some studies have investigated the thermodynamic stability of effector proteins,[20,21] there has yet to be a thorough investigation of the mechanical stability of these proteins. It is important to remember that thermodynamic stability does not have to correlate with mechanical stability, and proteins with similar thermodynamic stability can have very different mechanical stability[22]. This discrepancy makes sense when you consider the vastly different pathways a protein experiences when unfolded by different denaturants. Thermal or chemical denaturation result in a globule denatured state where non-native contacts can still stabilize the unfolded protein. However, mechanical unfolding results in an extended denatured state, where no native or non-native contacts are possible.[23–26] Moreover, mechanical unfolding occurs from one end of the protein to the other, potentially allowing the protein to sample different unfolding intermediates than when globally denatured by temperature or chaotropes[23]. Therefore, since effectors are mechanically unfolded *in vivo,* we need to test their mechanical stability using force spectroscopy.

Much of this thesis deals with high-precision, high-throughput methods to measure the amount of force required to unfold a protein. While I used these tools to investigate effector

proteins of the T3SS, it's important to understand that there is general interest in understanding how proteins respond to force. While the folding of proteins is critical to their function, it is often overlooked how important protein unfolding is to the working and upkeep of a healthy cell. Many proteins must be unfolded at some point, whether for transport to various intracellular areas, secretion out of the cell, or degradation within the cell. For a cell to survive, it must have the ability to degrade proteins that are misfolded or are no longer needed. This is most often completed by an ATPase that mechanically unfolds the protein. Proteins that are too mechanically stable to be properly degraded can wreak havoc on the cell, and often result in cell death. Therefore, while I am particularly interested in secretion through the T3SS, keep in mind that there are many applications that benefit from a precise and efficient method to study how proteins respond to force.

Atomic Force Microscopy as a High-Throughput, High-Precision Tool

Force Spectroscopy Relies in Measure Macromolecular Extension as a Function of Force

To investigate how a protein responds to force, a method is needed to apply precise amounts of force while monitoring protein structure. Current methods almost exclusively rely on single-molecule techniques, where it's possible to measure individual proteins unfolding. In these experiments, a protein is anchored to substrate (surface or bead) and "pulled" from one point. As the protein unfolds under an applied force, it goes from a compact three-dimensional structure to an extended polypeptide chain. By monitoring the length of the molecule length (ΔL_c) as force is applied, it's possible to identify the moment the protein unfolds and, in some cases, stable intermediates populated by the protein as it unfolds. Force-spectroscopy is the umbrella term used to describe this type of experiment. Monitoring a single-molecule allows researchers

to identify unique events that may be masked when measuring thousands of molecules at a time, as is done in bulk experiments. Moreover, single-molecule experiments allow researchers to delineate heterogeneity within populations. While a given protein or complex may always be made of the same basic pieces, minute differences in their structure can result in different behavior from molecule to molecule. Measuring these differences allows researchers to more fully understand how a protein folds or misfolds and the effects those differences have on its function.

Necessary Features for High-Throughput, High-Quality Force Spectrocopy

The ideal method for a single-molecule force spectroscopy (SMFS) experiment would provide the following: high-throughput data acquisition, high force stability, high temporal resolution, high force precision, and low force noise. High-throughput data acquisition allows for screening of many proteins and the acquisition of larger data sets. This enables the experimenter to not only test many more individual proteins, but also mutations that may affect how an individual protein folds and unfolds. High force stability, or the ability to maintain a set force over long periods of time, allows for equilibrium experiments. A protein can be stretched under a set force and allow it to reversibly unfold and refold. This is a powerful technique that, with proper analysis, can yield the folding energy landscape of a protein (i.e. the path the molecule takes to go from unfolded polypeptide to folded protein). High temporal resolution means that short lived intermediates in the folding or unfolding pathway can be measured. The more intermediates that can be measured, the more precise the picture of how the protein folds or unfolds. High force precision allows a researcher to know exactly how much force is being applied to the protein at any given time. This not only makes the assignment of the unfolding force much easier and more

accurate, but allows researchers to accurately derive the energy parameters that underlie protein folding. Finally, low force noise prevents small shifts in the applied force or extension from being missed. The lower the noise (the better the signal:noise ratio), the more events that can be measured. To date, there is no method that is excels in all these categories. The strengths and weaknesses of the current methods is discussed below.

Comparison of Three Common Force Spectroscopy Techniques

There are three main methods to carry out single-molecule force spectroscopy experiments: optical tweezers, magnetic tweezers, and atomic force microscopy (AFM). Optical tweezers trap a bead within a laser beam. One end of a target macromolecule (typically DNA, RNA, or protein) is attached to the bead while the other end is anchored to the surface or to another bead. The laser can then be moved away from the anchor point, exerting a force upon the protein stretched between the two points.[27] The position of the bead is measured by a quadrant photodiode, allowing researchers to track the extension (length) of the protein as it unfolds. Optical tweezers are notable for their force stability and temporal resolution, allowing the application of 0.1-100 pN of force with μs temporal resolution.[28] A key advantage of optical tweezers over other SMFS techniques is the low noise, allowing very small changes in force or extension to be measured. However, optical tweezers are notoriously difficult to use, and obtaining high precision measurement requires very careful and time-consuming calibration and tuning of the system. Moreover, because each experiment is initiated by finding, trapping, and positioning the bead, it's difficult to test more than a few proteins in a day. Finally, the amount of force optical tweezers can exert is limited to ~100 pN. Since the goal of my project was to test

a diverse set of proteins that could unfold across a wide range of forces, and to do so quickly, optical tweezers were not the best method for my purposes.

Another common method for SMFS experiments is magnetic tweezers. Magnetic tweezers are similar to optical tweezers in that it relies on the measurement of a trapped bead. However, whereas the optical trap used a laser to trap the bead, in a magnetic tweezer experiment the bead is polystyrene mixed with iron and is controlled by a magnetic field. Because the magnetic field can be applied to many beads at the same time, many individual proteins can be tested at the same time. However, the spatial and temporal resolution of magnetic tweezers is too low to see small, short lived unfolding events[28]. Since I wanted the ability to measure transient unfolding states, I wanted a method with better spatial and temporal resolution. I therefore turned to atomic force microscopy.

Atomic force microscopy is a relatively straightforward technique with a wide range of uses. Developed in the 1980s, it utilizes a cantilever containing a sharp tip to interact with the surface (Figure I-2)[29]. AFM-based SMFS was pioneered in the 1990s, when it was used to investigate the force necessary to rupture complementary strands of DNA[30]. Since then, enormous progress has been made to increase the force precision, force stability, and temporal resolution of AFM-based SMFS. Recent advances have focused on modifications to the AFM cantilever, resulting in drastic improvements in all three areas[31–34]. The deflection of the cantilever is accurately measured by reflecting a laser off the back of the cantilever and onto a quadrant photodetector. If the spring constant of the cantilever is known, the deflection can be interpreted into applied force. This force can either be applied to the surface (pushing) for applications such as AFM imaging, or to a molecule attached to the AFM tip (pulling) as is the
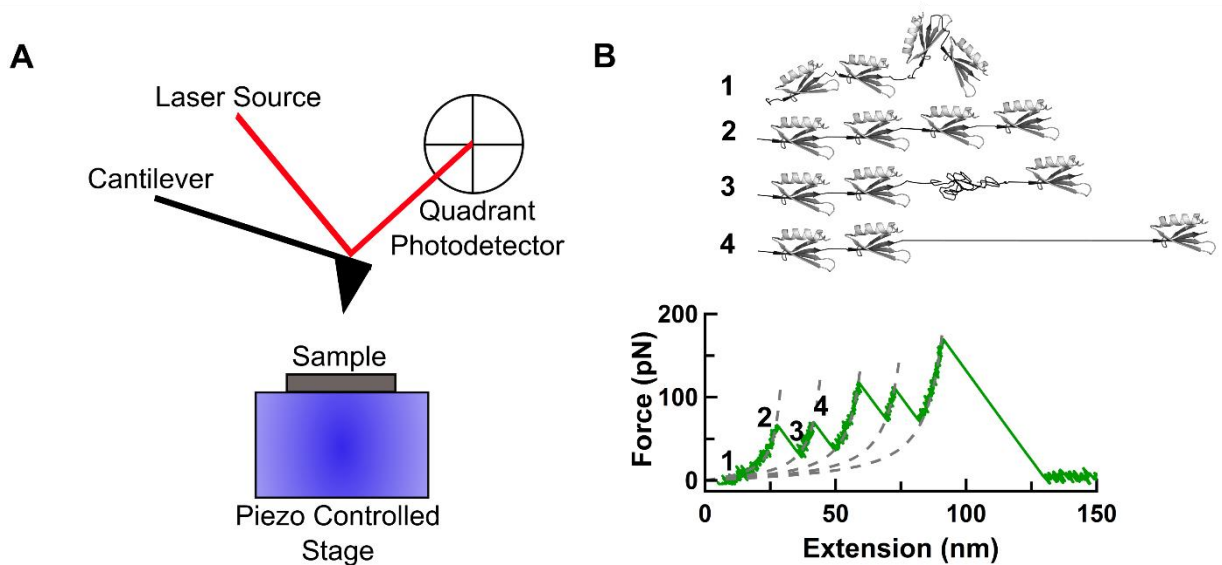
Figure I-3. A) Schematic of an atomic force microscope B) Representative force extension curve (FEC) with a structural representation of protein unfolding. After tip attachment (1) the protein is relaxed and no force is being exerted by the tip. As the tip starts to move away from the surface, the protein is put under tension (2). The force increases until one protein domain unfolds (3) which results in a rapid drop in force. As the tip continues to move away from the surface the protein is again put under tension (4). This process repeats until the entire polyprotein unfolds.

case with AFM-based SMFS. In pushing applications, the cantilever is pushed into the surface, causing the tip to bend away from the surface. As the tip passes over objects on the surface, the tip deflects further. The larger the object the tip encounters, the more the tip deflects. This is read out as an AFM topograph, and can yield sub-nm resolution depending on the sharpness of the tip.

For pulling applications as is the case for SMFS experiments (Figure I-3B), the cantilever is moved away from the surface. If the tip is attached to a surface anchored protein, the cantilever deflects toward the surface. As the cantilever continues to move away from the surface, the protein prevents the tip from moving away, increasing the deflection as force is applied to the protein. At some point, the protein will unfold, allowing the tip to move back into line with the

cantilever, read out as a reduction in the deflection as the tension in the system is released. This is recorded as a force extension curve (FEC), where the force (calculated from the spring constant and deflection) is plotted on the y-axis and the extension (how far away the tip is from the surface) is plotted on the x-axis (Figure I-3B). The peaks in the curve occur when the protein is under maximum tension. When the protein unfolds, the tension is released and the force drops rapidly. If multiple proteins are connected into a polyprotein, this process repeats until all proteins are unfolded or the tip detaches from the protein. The FEC in Figure I-3B is the measurement of a polyprotein of 4 identical proteins (NuG_2). Note the 5 peaks in the curve; the first four peaks correspond to NuG_2 unfolding (as illustrated in the schematic) while the final peak corresponds to the protein rupturing away from the tip. This is an ideal AFM FEC, allowing for easy interpretation of the force necessary to unfold the protein.

Using a Worm-Like Chain to Fit Force-Extension Curves

Interpreting the AFM FEC (also known as an AFM trace) starts by fitting a worm-like chain (WLC) model to the force extension curve. The WLC model (Equation 1), where F is force, k_BT is thermal energy, L_c is contour length of the polymer and x is extension away from the surface, treats protein unfolding as the extension of an elastic polymer, and when fit to the data provides the contour length of the folded and unfolded protein[35]. Because the force, thermal energy, and

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

extension away from the surface are all measured during the experiment (or are known physical constants in the case of thermal energy), it's possible to solve for contour length. The contour length is the length of the protein at its maximum possible extension. To put it another way, the contour length is the longest the protein can get without breaking, and does not depend on the

applied force. In Figure 3B, the grey dashed lines indicate the WLC fits to the data, with each peak

is fit to its own WLC model. That enables me to identify the length of polyprotein before any of

the four NuG_2s unfold, after the first NuG_2 unfolds, after the second unfolding, the third, the

fourth, and finally the length of the completely unfolded polyprotein. I can then extract the

change in length that occurs after every unfolding event, denoted as the change in contour length

(ΔL_c). The change in contour length from the folded to the unfolded state allows us to identify

what protein or protein fragment unfolds at each step.

Applying Force to Protein Shifts the Free Energy Landscape

After determining the contour length of measured unfolding events, the next step in

analyzing an AFM FEC is to identify the force at which each protein unfolds, and by measuring

the slope of the trace prior to protein unfolding, calculate the loading rate, or how quickly the

force was applied to the protein. These values are important because, with further analysis, they

provide information about how stiff a protein is and the unfolding rate at zero force. How these

factors relate protein unfolding under force is best explained by looking at the free energy

landscape of the protein (Figure I-4A).[36] Interpreting the free-energy landscape of a force-

spectroscopy experiment is somewhat simplified compared to other experiments because the

reaction coordinate is the extension away from the surface as the protein is pulled. As force is

applied to the protein through the retraction of the AFM tip, it tilts the energy landscape and

lowers the activation barrier, favoring the unfolded state. The reaction coordinate can be

understood as a real physical value, and the distance from the folded state (f) to the transition

state (‡), x_u, represents the width of the energy barrier. This relates to how far the protein must

be pulled before it passes over the energy barrier and unfolds. The further this distance is, the

softer, more compliant a protein is. Proteins with a short distance to the transition state are thought of as brittle, unfolding all at once when the barrier is reached. Typically, α-helical proteins are more compliant (longer distance to the transition state) than proteins with β-sheets. This can be an important physiological characteristic since a more compliant protein allows the force to be applied in a more stepwise fashion, and may allow ATPases to more easily unfold a compliant protein than a stiff protein, even if the total amount of force needed to unfold the proteins are the same.
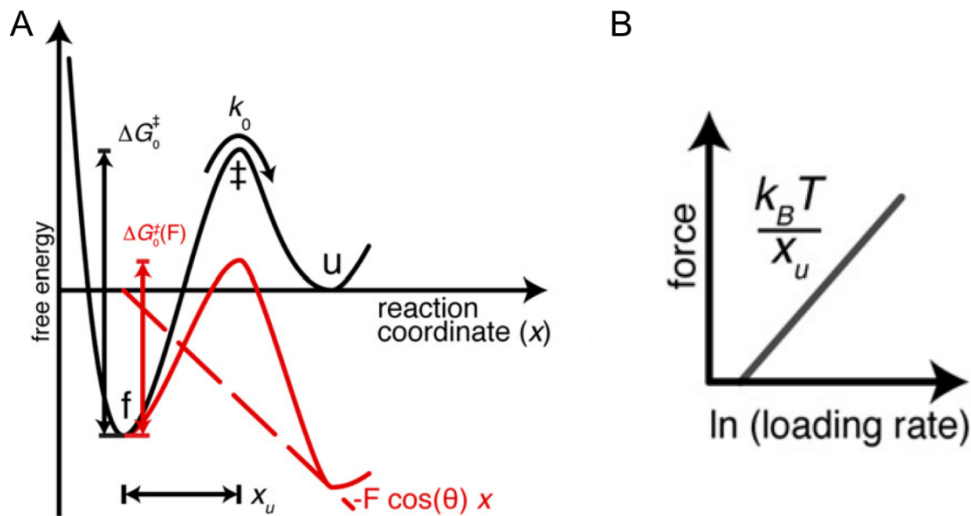
Figure I-4. (A) Free energy landscape shifts due to applied force. In a force-spectroscopy experiment, the reaction coordinate is defined as the extension away from the surface (x). For the protein to go from the folded state (f) to the unfolded state (u), it must pass through a high energy transition state (‡). The height of this barrier is the activation free energy ($\Delta G_0^{\ddagger}$) and the distance to the transition state along the reaction coordinate is denoted by x_u. x_u defines the width of the energy barrier. The rate of crossing the energy barrier is given as k_0. As pulling force is applied to the protein, a mechanical potential ($-F \cos(\theta)x$, where F is force, θ is the angle between the applied force and the extension away from the surface, x is extension away from the surface) is added to the landscape, as shown by the dashed red line. This tilts the energy landscape (red line), lowering the activation barrier and favoring the unfolded state. (B) A model dynamic force spectrum showing the dependence of the unfolding force on the loading rate (how fast the force is applied to the protein). The result is a linear fit where the slope is proportional to $1/x_u$, which is to say the slope is proportional to the inverse distance to the transition state. The longer the distance, the smaller the slope, and the less the loading rate changes the unfolding force. Adapted from Bippes and Muller, *Rep. Prog. Phys.*, 2011.

Protein unfolding is a thermally driven process, with the energy fluctuations constantly probing the unfolding energy barrier. Therefore, the longer a protein can sample the energy barrier, the more likely it is to eventually have enough thermal energy to pass over the barrier. Given enough time, proteins will spontaneously unfold at a given rate, a measure of how stable the protein fold is. As force is applied to the energy landscape, it lowers the activation barrier, making it more likely that the protein can cross the barrier. This is balanced by the rate at which the cantilever is retracted from the surface, because the rate directly influences how long the protein can sample the energy barrier. The faster the retraction, the less likely a thermal fluctuation will add enough energy to cross the barrier, and the higher the force necessary before the protein unfolds. This can be illustrated by plotting the most probable unfolding force against the log of the loading rate in a dynamic force spectrum (Figure I-4B).[36] The relationship is linear, with higher loading rates leading to higher unfolding forces. The slope of the line is proportional to $1/x_u$, meaning that the force needed to unfold compliant proteins changes less as the loading rate changes.

Bell-Evans Model Provides Off-Rate at Zero Force and Distance to the Transition State

One of the common analysis used in force-spectroscopy is fitting the linear dynamic force spectrum with the Bell-Evans model (Equation 2), where *F(r)* is the most probable unfolding force, k_BT is thermal energy, x_β is the location of the energy barrier, *r* is the loading rate, and k_0 is the off-rate constant at zero force (or the unfolding rate at zero force when looking at protein

$$F(r) = \left(\frac{k_B T}{x_\beta} \right) \ln \frac{r \, x_\beta}{k_0 \, k_B T}$$

(2)

unfolding instead of ligand/receptor unbinding). When fit, we can not only extract the distance to the transition state, but also the unfolding-rate at zero force.[37] These are fundamental

characteristics of the protein, and give us an idea of the stability and compliance of the protein. I will look to apply this analysis to effector proteins and their homologues, giving us a method to compare the proteins beyond just the force it takes to unfold the protein.

SMFS is Hindered by Non-Specific Attachment

With a better idea of what information can be obtained from a force-spectroscopy experiment, it's possible to turn our attention to what is necessary to collect a high-quality force extension curve. To unfold the protein, you must attach the protein to both the surface and the tip. Ideally, the attachment sites are at the termini of the protein, allowing the protein to be completely unfolded. Only a single protein or polyprotein can be attached to the tip, otherwise the unfolding of multiple proteins in parallel or in series convolutes the resulting trace, making it impossible to interpret. As the tip is retracted, it is important that the tip not stick to the surface. If this happens, there will be a large force peak when the tip eventually ruptures from the surface, obscuring any protein unfolding that occurs at low force near the surface (where we expect to observe effector proteins unfolding). Once the tip is away from the surface, the connection to the protein must be strong enough to survive the unfolding of the protein of interest yet weak enough to unbind when high force is applied. This allows the tip to remain clean and capable of attaching to another protein as the cycle of approach and retraction is repeated. This is the ideal scenario for a force-spectroscopy experiment that sadly is not how most experiments are performed. Traditionally, AFM experiments are time consuming, tedious affairs hampered by non-specific adhesion, making interpretation of the data difficult if not impossible. Each step in this process had to be addressed for us to obtain high-throughput of high-quality data (Chapter 2). We wanted a way to collect data quickly, while improving the specificity and precision of the

collected data, simplifying our interpretation and making it easier to compare traces between proteins.

Traditional AFM-based SMFS is done with non-specific attachment to both the AFM surface and tip. This means depositing protein onto unfunctionalized glass slides and probing with an unfunctionalized silicon nitride tip. While simple to implement, non-specific attachment results in a very low percentage of attempts where the tip attaches to a protein (< 1%). Furthermore, even if the tip attached to a protein, most of the time it detaches before the protein of interest unfolded. Even if both of those problems can be overcome, the use on non-specific attachment means you can't know where the protein attaches to either the surface or the tip. This makes it difficult to interpret the data, particularly when testing a protein for the first time.

Polyproteins Facilitate Accurate Assignment of Unfolding Data

Because using non-specific attachment results in messy, difficult to interpret FECs, virtually all protein studies using AFM-based SMFS experiments are done with polyproteins. Early experiments fused multiple copies of one protein together, giving researchers an increased chance to see the protein of interest unfold and increasing the number of unfolding events upon rare attachments[38–40]. While this was an essential development that enabled many early AFM SMFS studies, it still had drawbacks. Despite being able to get multiple unfolding events for every attachment, the throughput is still very low. Moreover, as each protein unfolds, it adds compliance to the system which can change the rupture force of subsequent unfolding events. This broadens the average rupture force at any given pulling speed, decreasing the accuracy of the results[41].

To obtain more precise measurements, researchers started only adding one copy of their protein of interest to their polyproteins (usually in the middle), with other proteins being added to act as markers. As the polyprotein unfolds, the marker proteins, which have known rupture forces and result in known changes in contour length, act as a mechanical fingerprint. Using marker domains allows researchers to know the polyprotein has fully unfolded, a very important check when using non-specific attachment since it's possible to only unfold part of a protein. Marker proteins also allow researchers to ensure only one polyprotein is being unfolded at a time. If multiple proteins attach to the tip, the resulting force extension curve will either contain too many peaks (if the proteins unfold serially) or the peaks will have a significantly higher unfolding force (if the proteins unfold in parallel). By ensuring the correct number of marker domains and the correct unfolding force, any additional peaks can be attributed to the unfolding of the protein of interest. If the trace doesn't include enough marker domain peaks, the tip either ruptured before the full polyprotein could unfold, or the tip attached to an internal site within the polyprotein. While these fingerprint domains can add a great deal of confidence to results of an AFM experiment, requiring complete polyprotein unfolding means even fewer traces can be used.

Collecting High-Quality, High-Throughput SMFS Data Requires Site-Specific Attachment

It became apparent that to achieve high-throughput collection of high-quality data, site-specific attachment was necessary. By not relying on non-specific adhesion to the tip, a higher percentage of attempts will yield protein unfolding. Placing unique chemical or protein handles at the termini of the protein means that when the tip attaches to a protein most of traces should contain the full polyprotein unfolding. This wasn't a new idea, but most of the schemes relied on

cysteine-maleimide linkages which wouldn't work for our proteins of interest, all of which contain solvent accessible cysteines. Moreover, the majority of labs using specific attachment only attached protein to the AFM surface. Finally, the procedures for surface and tip functionalization were finicky, time consuming, and typically had to be performed immediately prior to any experiment. The second chapter of this thesis describes the development of a polyprotein construct to facilitate AFM-based SMFS measurements and the collaboration with the Perkins lab to implement their novel surface functionalization protocols and cantilever improvements to yield high-throughput acquisition of high-quality force extension curves. The final product is a highly efficient process for surface and tip functionalization compatible with a wide range of protein functionalizations that I developed and optimized.

Chapter II A Heterobifunctional Polyprotein for Efficient AFM Analysis

As discussed in Chapter 1, I developed and optimized a bifunctional polyprotein construct, enabling site-specific attachment to both the AFM surface and tip. By bifunctionally labeling proteins with dibenzocyclooctyne (DBCO) and biotin, I could site-specifically attach the protein to both the AFM surface and tip. In doing so, I achieved a 70-fold improvement in acquiring high-quality force extension curves. In this chapter I will present the development and optimization of two polyprotein constructs which utilize the aldehyde tag[42] to facilitate both homo- and hetero-bifunctional labeling, as well as their application to not only test proteins of interest, but also test newly developed AFM cantilevers.

While traditional AFM-based SMFS continues to rely on non-specific attachment to both the AFM surface and tip, there has been growing interest in developing efficient, easy to implement methods to site-specifically attach proteins to both the AFM surface and tip. The first attempts at site-specific attachment covalently linked proteins to AFM surfaces but continued to probe with unfunctionalized tips[43–47]. This significantly improved the number of attachments, but because the tip attachment was non-specific, it remained difficult get high-quality data at high rates. Therefore, researchers started exploring methods to site-specifically attach proteins to both the AFM-surface and tip. Initially, covalent linkages were used on both sides.[48,49] The problem with these methods is that every bond in the chain is equally strong, meaning that when the force is high enough to break the covalent bonds (>2 nN), the protein is equally likely to break internally as at the site of tip attachment. The result that the tip quickly became clogged with ruptured proteins, limiting the lifetime of the tip and impeding throughput. To increase throughput, researchers used well-characterized receptor-ligand pairs that formed strong but

non-covalent linkages. The most commonly used pair is streptavidin-biotin, which ruptures at 100-200 pN[50–52]. This is enough force to unfold most proteins-of-interest, but is weak enough to ensure that the streptavidin-biotin linkage will break before the internal covalent bonds. This leaves a clean tip, meaning that thousands of attachments can be achieved with an individual tip[53]. Here I present the developments in tip and surface functionalization, as well as the development of a bifunctional polyprotein capable of specific attachment to both the AFM surface and tip.

Results

A Simplified Method for Functionalizing AFM Surfaces and Tips

The idea of using site-specific attachment to both the AFM-surface and tip was certainly not a new concept when I started in the lab. However, it was hampered by the complexity and time-consuming nature of most of the surface and tip functionalization protocols. For that
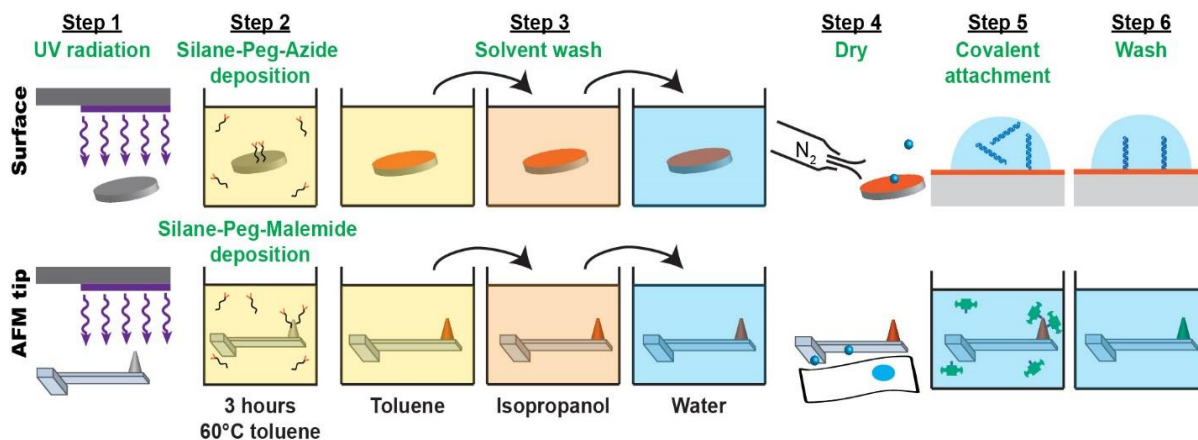
Figure II-1. Efficient six-step protocol for functionalization of glass coverslips (a) and silicon-nitride tips (b) using hetero-bifunctional PEG. Tips and coverslips were activated using UV radiation and then silane-PEG functionalized in toluene at elevated temperature resulting in azide-derivatized coverslips and maleimide-derivatized AFM tips. After a series of solvent washes, the surfaces were dried. Biomolecules were covalently bound to coverslips using copper-free click chemistry, while thiol-modified streptavidin was covalently attached to maleimide-coated AFM tips. Finally, coverslips and tips were washed and stored in buffer at 4 °C until use.

reason, Dr. Walder in the Perkins lab developed a new protocol that utilizes functionalized peg-silanes to functionalize glass slides or AFM cantilevers.[53] The protocol relies on a simple 3 h incubation in toluene dissolved silane-PEGs, saving ~5 h compared to previously published methods (Figure II-1). When combined with a wide variety of commercially available PEG-silanes, this method allows us to quickly and easily functionalize surfaces and tips.

Azide Functionalized Surfaces are Stable for a Month

While maleimide functionalized surfaces did work to improve protein deposition, the inability to store them meant preparing surfaces was a daily task that took ~5 hours. Ideally, surfaces could be prepared in large batches that could be stored until needed, drastically reducing the amount of time spent functionalizing surfaces. To tackle this issue, Rob looked for other methods of covalently coupling proteins. One attractive option was copper-free click chemistry which allows for efficient, covalent coupling between an azide and a strained alkyne.[54,55] Conveniently, there are commercially available silane-PEG-azides (Nanocs Inc.) which could be used in the same protocol developed for silane-PEG-maleimide (Figure II-1). The major advantage of functionalizing surfaces with azide instead of maleimide is that the surfaces could be stored for over a month without a significant drop in activity.[53] This meant large batches of surfaces could be prepared and stored until needed, not only greatly reducing the amount of time spent functionalizing surfaces, but also increasing reproducibility between surfaces.

A Heterobifunctional Polyprotein with Bioorthogonal Labels

With the significant improvement in surface and tip functionalization, I needed a way to functionalize proteins. To attach proteins to azide functionalized surfaces and probe them with streptavidin functionalized tips, it was necessary to bifunctionally label proteins with biotin and a strained alkyne (such as dibenzocyclooctyne (DBCO)). While many methods have been developed to specifically label proteins, I wanted a method that enabled bio-orthogonal labeling with high yield while using a small tag. The natural place to start was cysteine-mediated maleimide labeling.[56] However, since all the effector proteins we were interested in studying contained solvent accessible cysteines, simply relying on cysteine-mediated labeling was not a viable option. While I designed constructs containing the the HaloTag[57] as a method of attaching our proteins to the surface or tip, it introduces a large protein domain and would have required a new functionalization scheme to introduce a chloroalkane to the surface or tip. For similar reasons, I also decided against the SNAP-Tag[57] and CLIP-Tag[58,59]. Instead, I turned to the aldehyde
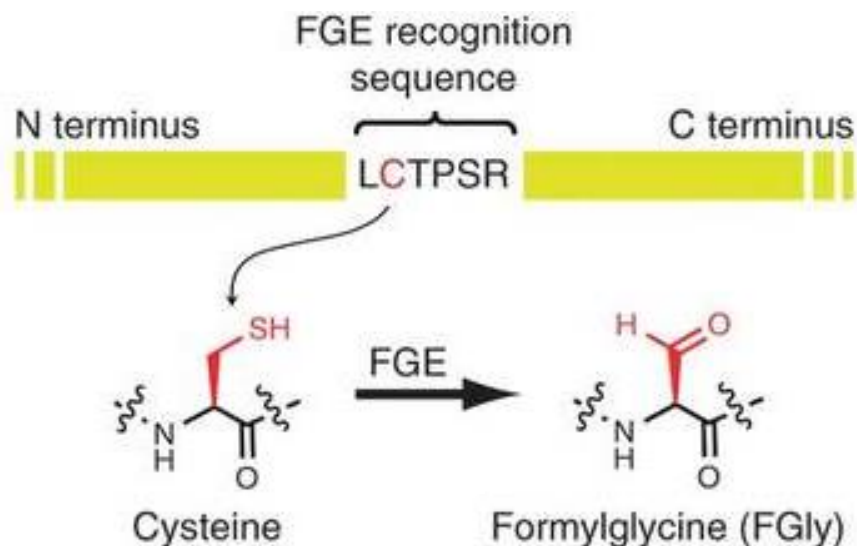
Figure II-2. Aldehyde tag conversion by FGE. Upon reaction with formylglycine generating enzyme (FGE), the cysteine within the aldehyde tag (LCTPSR) is converted to formylglycine, an aldehyde containing amino acid. Adapted from Rabuka *et al, Nature Protocols* (2012).

tag, a small 6-amino-acid tag (LCTPSR) which is recognized by formylglycine-generating enzyme (FGE)[60]. Upon reaction with FGE, the internal cysteine is converted to an aldehyde containing formylglycine (Figure II-2). The aldehyde provides a bio-orthogonal tag compatible with a wide range of commercially available hydrazide and oxyamine reagents such as biotin-PEG$_3$-oxyamine (Quanta Biodesign) and DBCO-PEG3-oxyamine (ClickChemTools).

After deciding on the aldehyde tag, I designed a two polyprotein constructs that contained aldehyde tags, a marker protein (NuG2) to provide a mechanical fingerprint, and a cloning site for the insertion of proteins of interest (Figure II-3). NuG2 has been well characterized in various SMFS experiments (rupture force ~40 pN, ΔLc ~18 nm)[26,61] and provides a convenient, fast folding marker within the construct. The four copies of NuG2 allow verification that the polyprotein completely unfolded, increasing confidence that the protein of interest completely unfolded before the protein detached from the tip.

The first construct I designed contained an N-terminal cysteine, compatible with maleimide functionalization, and a C-terminal aldehyde tag (Figure II-3A). Known as the
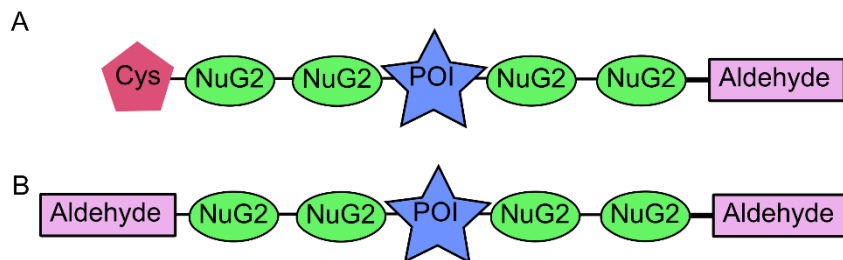
Figure II-3. Aldehyde tag conversion by FGE. Upon reaction with formylglycine generating enzyme (FGE), the cysteine within the aldehyde tag (LCTPSR) is converted to formylglycine, an aldehyde containing amino acid. Adapted from Rabuka *et al, Nature Protocols* (2012).

directional construct, it provides the highest labeling efficiency by specifically labeling each end

of the protein, but with the drawback of relying on cysteine-mediated labeling.

I also wanted polyprotein that could be bifunctionally labeled without the need for

cysteines which could be used for proteins of interest that contained solvent accessible cysteines.

To do so, I created a construct with aldehyde tags on both ends (Figure II-3B). This allowed for

completely biorthogonal labeling, without reliance on cysteine mediated labeling. This was a key

distinction that allowed me to test proteins that contained solvent accessible cysteines with an

efficiency equal or greater than methods that rely on maleimide reagents for site-specific

attachment.

In Vivo Aldehyde Conversion Unsuccessful with Co-Expressed FGE

After verifying I could express and purify the polyprotein constructs, I needed to ensure

that the cysteine in the aldehyde tag was being converted to a formylglycine by FGE. As noted in

the methods, the plasmids containing the polyprotein construct also contain a PET-Duet cassette,

enabling co-expression of FGE with our protein of interest. While other labs have shown

successful conversion by co-expressing FGE,[62] the previous work used two plasmids to achieve
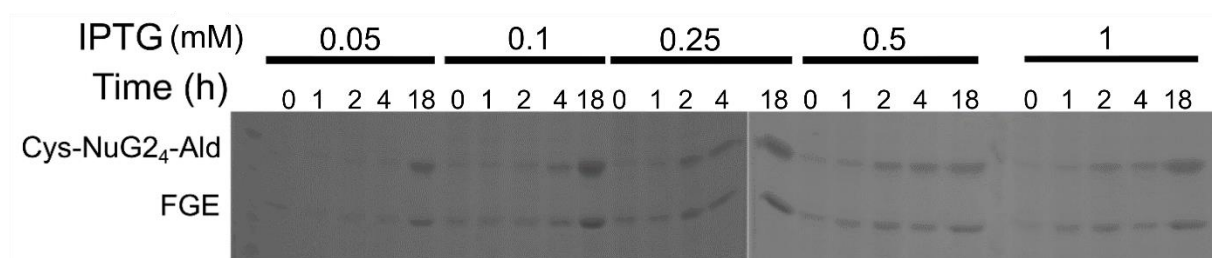
Figure II-4. Expression test of pMS 1088 (directional construct [cysteine-NuG2$_4$-Aldehyde] [*M. Tb* FGE]). Cultures of *E. Coli* BL-21 (DE3) transformed with pMS 1088 were grown to OD$_{600}$ = 0.55 at 37°C before induction with the noted amount of ITPG. Cultures continued shaking at 37°C for 4 h before being moved to an 18°C incubator overnight. 100 uL samples were taken, spun down, and resuspended in 50 uL TBS before being prepared for SDS-PAGE gel.

co-expression instead of a PET-Duet cassette. Therefore, it was unknown if using a PET-Duet cassette would achieve similar results. As shown in Figure II-4, FGE expressed at roughly equal levels as our polyprotein construct at all tested induction conditions. This gave me hope that the *in vivo* conversion from cysteine to formylglycine would be as robust as previously reported (>95%)[63].

To test for aldehyde conversion in a fast, qualitative way, I labeled the purified protein with Alexa Fluor 350 Hydrazide. As noted in the methods, I used this label to compare two labeling methods that had previously been used for *in vitro* labeling of aldehyde tagged protein. Because I was also interested in whether the position of the aldehyde tag effected conversion efficiency, I tested polyproteins with a single aldehyde tag at either the N-terminus (1091) or C-terminus (1088). As seen in Figure II-5, the low pH Bertozzi buffer resulted in protein aggregation and most
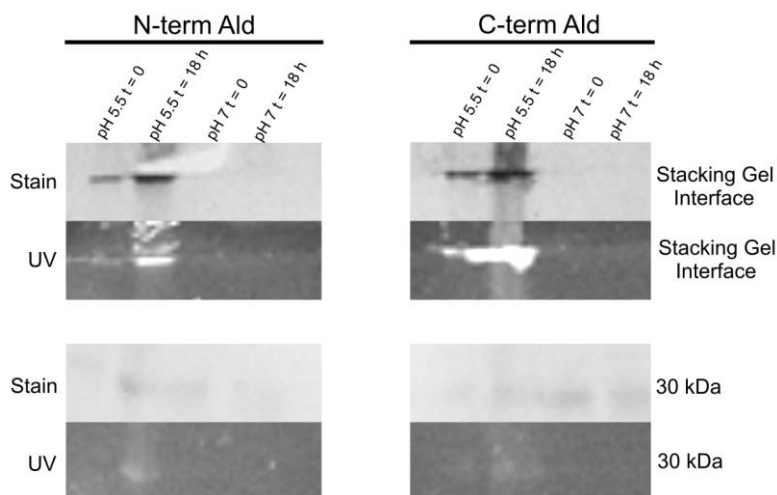
Figure II-5. Alexa Fluor 350 Hydrazide labeling test of N- and C-terminal aldehyde tagged proteins in two different buffers. Protein was exchanged into either the Ha labeling buffer (250 mM KH_2PO_4, pH 7.0, 500 mM KCl, 5 mM DTT) or Bertozzi labeling buffer (100 mM MES, pH 5.5) before the addition of 100 mM (Ha) or 15 mM (Bertozzi) Alexa Fluor 350 Hydrazide. Samples were incubated for 24 h at 4°C (Ha) or 37°C (Bertozzi) overnight. Gels were imaged with 365 nm light before being stained with Coomassie R-250. Upper images show the stacking gel interface, where protein aggregates are retained.

of the protein was unable to make it into the gel. While it does appear that labeling may be better

in the Bertozzi buffer, neither 1088 or 1091 in either buffer seemed well labeled. This was

confirmed by LC-MS analysis of the purified protein, which showed <5% conversion of the

cysteine to formylglycine *in vivo.* However, I was worried that the problem was the detection of

the dye was adding further complication. I wanted a second method to quickly test for aldehyde

conversion.

Because I wanted to biotinylate the proteins for my AFM experiments, I decided to also

use biotin as a qualitative method to test aldehyde conversion. Specifically, I labeled the proteins

with biotin-PEG$_2$-oxyamine using the same two protocols as above, ran the proteins on an SDS-

PAGE gel which was western blotted onto PVDF membrane. After probing with streptavidin-HRP

the membrane could be exposed with ECL Pro chemiluminescent reagent. This method provided

me a with an easy way to qualitatively measure aldehyde conversion and labeling, which was

sufficient to test which labeling buffer was best. As shown in Figure II-6, the Bertozzi labeling
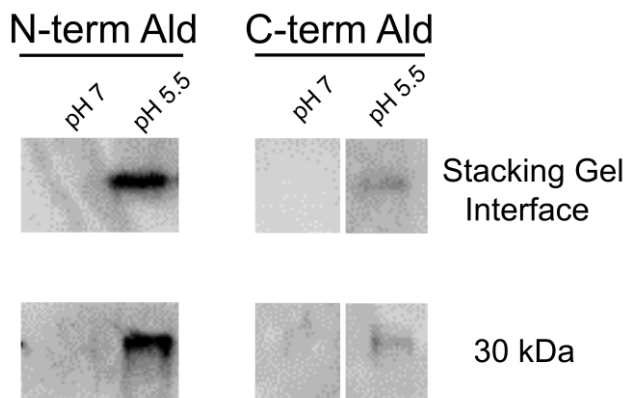
Figure II-6. Biotin-PEG$_2$-Oxyamine labeling test of pMS 1088 in two different buffers. Protein was exchanged into either the Ha labeling buffer (250 mM KH$_2$PO$_4$, pH 7.0, 500 mM KCl, 5 mM DTT) and Bertozzi labeling buffer (100 mM MES, pH 5.5) before the addition of 100 mM (Ha) or 15 mM (Bertozzi) dye. Samples were run on an SDS-PAGE gel and western blotted onto PVDF membrane, probed with streptavidin-HRP and visualized with ECL Pro reagent.

buffer resulted in much higher levels of aldehyde labeling than the neutral pH Ha labeling buffer, which resulted in no measurable labeling. However, it is important to note that the lower pH used in the Berozzi method does result in protein aggregation, which can be seen by the upper band on the western in Figure II-6. Since I needed to biotinylate the polyproteins for our site-specific attachment scheme (biotin-streptavidin linkage to the tip), western blotting became my primary test for aldehyde conversion and labeling.

LC-MS was used to quantitatively assess the aldehyde conversion *in vivo.* Trypsin digested samples were tested for the mass change associated with converting a cysteine to a formylglycine (-17.9882). The first sample we tested was the same used for the biotin labeling experiment shown in Figure II-5. To our surprise, despite a strong band in the western blot, the *in vivo* aldehyde conversion was significantly worse than previously reported results. I was never able to get over 35% conversion of the cysteine to formylglycine *in vivo*, and many preps yielded <5% conversion to formylglycine. This helped explain why our labeling tests were yielding such poor results and indicated a need for a different approach to convert the cysteine in the aldehyde tag to formylglycine.

In Vitro FGE Conversion is >95% Effective

To test whether the problem with aldehyde conversion was caused by expressing FGE using the PET-Duet cassette, I transformed a separate FGE expression plasmid- pBAD/myc–his A Rv0712 (FGE)[42] (Addgene plasmid #16132)- along with our polyprotein plasmid. Despite the pBAD plasmid conferring resistance, I was never able to induce FGE expression with arabinose in BL-21 (DE3) cells. While I could show arabinose induction of FGE expression in Top-10 cells, I never successfully had a system that coexpressed the polyprotein and FGE from two plasmids. I believe

that with more time, this would be a workable solution, given the success other labs have had with this method.

At the time, we were looking for an easy, efficient solution that would work for both polyproteins, regardless of the protein of interest inserted within them. We reached out to David Rabuka, one of the students from Carolynn Bertozzi's lab who had done the initial work on FGE conversion. He suggested trying *in vitro* aldehyde conversion using purified FGE, and graciously sent us samples of purified FGE, both the human *M. tuberculosis* variants. When reacted with our purified polyprotein at a 1:10 molar ratio in FGE conversion buffer (50 mM triethanolamine (TEA), pH 9.0, 10 mM NaCl, 1 mM TCEP) overnight at 18°C, we a saw a dramatic increase in aldehyde conversion. The increase in aldehyde conversion was evident in my qualitative tests (both fluorophore and biotin labeling) and subsequently verified by LC-MS analysis of trypsin digested samples which showed routine conversion of >80% to formylglycine. There was no significant difference in activity between the human and *M. tuberculosis* variants.

While the gift of purified FGE was an excellent way to demonstrate that *in vitro* FGE conversion was an effective solution, I needed a way to purify my own stock of FGE. I attempted to simply purify *M. tuberculosis* FGE from the pBad plasmid, but despite strong expression I was unable to purify the protein because it lacked a 6xHis tag. I therefore moved the gene for *M. tuberculosis* FGE into our 6xHis-SUMO tag construct (pMS 984). The resulting fusion protein expressed well and allowed me to purify high concentrations of active FGE. While my initial FGE purifications were completed in the same way as the polyproteins, David Rabuka reached out in 2015 to provide an optimized method. His group at Catelent Biologics published a paper which included not only an optimized purification method, but demonstrated that adding copper (II)

27

sulfate could increase the activity of purified FGE[64]. However, adding copper sulfate to my purified protein resulted in almost complete precipitation of the protein. Despite the papers assertion that the precipitation could be reversed by adding EDTA, I was never able to recover the bulk of the protein. Therefore, the optimized method that continues to be used in the lab does not include the copper sulfate activation but continues to yield highly active FGE capable of converting >95% of cysteines to formylglycine *in vitro*.

Oxyamine Reagents are not Stable Enough for Prolonged AFM Experiments

Once I had a robust method for incorporating aldehydes into our polyprotein, I turned my attention to what labels would best facilitate site-specific attachment to our AFM surfaces and tips. My first attempt used the Bertozzi labeling scheme to biotinylate the hetero-bifunctional construct with biotin-PEG$_2$-oxyamine. The result was a polyprotein with a cysteine at the N-terminus and a biotin at the C-terminus. This was deposited on a maleimide functionalized surface and probed with a streptavidin functionalized tip. While this was successful and yielded a 15-fold improvement in attachment rate compared to non-specific attachment (Figure II-13F), I quickly realized that the attachment rate was dropping over the course of a 3 h experiment (Figure II-7, red line). While I was still able to get many attachments using oxyamine labeled protein, it prevented me from using a tip to probe multiple proteins. This makes it much more difficult to compare the results between proteins, which was a major goal for the project. The drop in attachment rate indicated that our tip was getting clogged, either by protein that detached from the surface or biotin that detached from the protein. Since the attachment to the surface was covalent, I was confident that the problem was the oxime linkage that attached the
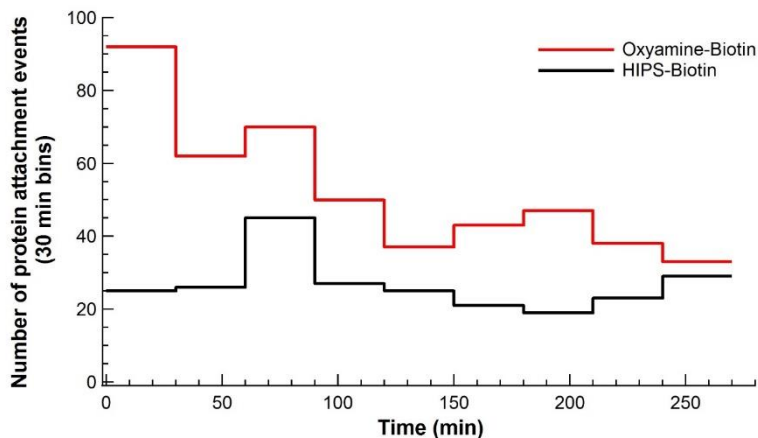
Figure II-7. Comparison of attachment efficiency of oxyamine and HIPS biotin labelled protein over time. The directional protein construct (aldehyde-NuG2$_4$-cysteine) was labelled with either oxyamine-PEG$_2$-biotin or HIPS-Biotin and attached to a maleimide functionalized surface at a deposition concentration of 0.3 mg/mL. While the protein labeled with HIPS-Biotin showed consistent attachment over 4.5 h, the oxyamine-biotin labeled protein showed a rapid decline in protein attachment. After 2.5 h, the number of protein attachments dropped by 60%.

biotin to the protein. This was a surprising idea because oxime linkages are typically regarded very stable.[65]

To test whether the oxime linkage was simply dissociating in our AFM buffer, I labeled a batch of protein with biotin-PEG$_2$-oxyamine and ran a time course to look for loss of biotin labeling. Samples were stored at 4°C, 25°C, and °C with aliquots taken at 30 m, 1 h, 2 h, 4 h, 1 d, 3 d, 4 d, and 7 d. As seen in Figure II-8, oxime degradation seems to be relatively minor over the timespan of a typical AFM experiment (4 hours). However, at longer times, particularly at warmer temperatures, the oxime bond does seem to degrade. While the degradation may seem minor over the timespan of an AFM experiment, it's important to remember that only a small amount of free biotin is necessary to occupy the available streptavidin on the tip. While degradation of the oxime linkage may simply be occurring in solution, it's also possible that the oxime linkage degrades under force. If this occurs, the biotin would remain in the streptavidin binding site,
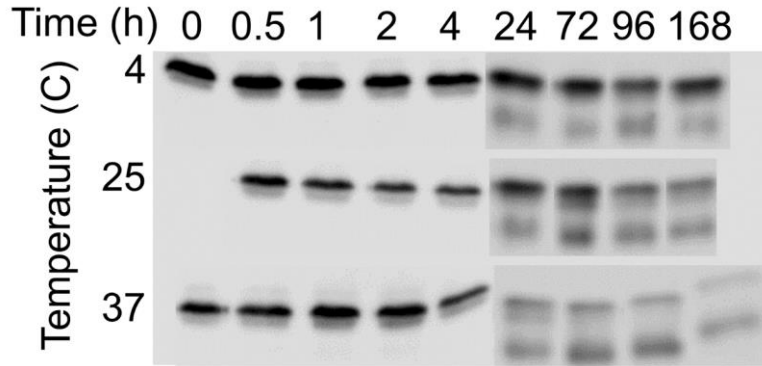
Figure II-8. Oxime degradation time course. Biotin-PEG$_2$-Oxyamine labeled 1088 (directional construct) samples were incubated for 7 days (168 hours) at 4, 25, or 37°C. Aliquots were taken at the indicated time points and placed in 4x SDS dye and frozen to halt the reaction. While degradation is slow, there is apparent loss of biotin after 2 h at 25°C, typical conditions for an AFM experiment. Interestingly, after a day of incubation, a lower molecular weight product appears on the western blot which may indicate the protein itself is susceptible to degradation over time.

preventing it from binding to a new protein. It is difficult to say with certainty which mechanism results in the decreased attachment efficiency. It's important to note that the biotin-oxyamine still results in a much greater data acquisition than using non-specific attachment. However, because the tip becomes occupied by free biotin, it cannot easily be used between samples, introducing additional error when trying to compare the unfolding force between proteins. Therefore, we concluded that the oxime linkage, while robust enough for most biochemical experiments, is insufficiently stable for the AFM experiments we were most interested in performing.

Conveniently, David Rabuka's group also wanted a more stable aldehyde reactive label and developed the Hydrazion-Pictet-Spengler (HIPS) ligation. Based off of a hydrazide reaction, HIPS reagents cyclize upon reaction with an aldehyde as shown in Figure II-9, forming an incredibly stable linkage that doesn't degrade, even if left for weeks in harsh conditions.[66] Another advantage of HIPS is that labeling can be done at a higher pH than traditional hydrazide or oxyamine reactions, with maximum efficiency at pH 6 instead of pH 5.5. The only real drawback is that HIPS has slower labeling kinetics than either hydrazide or oxyamine, requiring a 3-day incubation at 37°C for optimal labeling. I was willing to accept the tradeoff if it allowed me to use an individual tip for many samples without a significant drop in activity. Dr. Rabuka graciously sent some HIPS-reagent that could be derivatized to create HIPS-biotin and HIPS-DBCO. Our collaborator Dr. Jake Greenberg synthesized the reagents, details of which can be found in Appendix A. After we had been using HIPS reagents for ~2 years, it was discovered by chemists at ClickChem Tools that HIPS degrades over time, even when stored at -20°C. This likely explains why I saw variation in HIPS labeling based on the age of the reagents. To avoid this problem
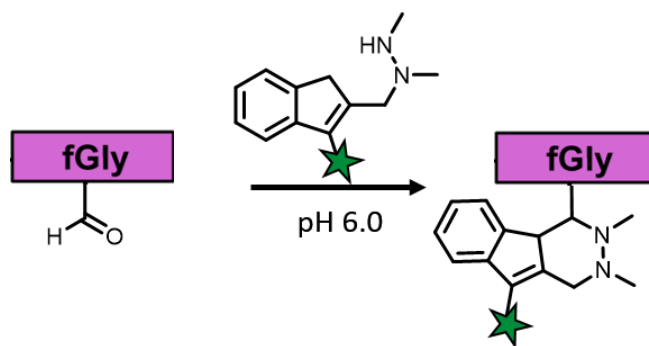
Figure II-9. Hydrazino-Pictet-Spengler (HIPS) Reaction. The active group of HIPS is very similar to a hydrazide, but cyclizes upon reaction with an aldehyde, resulting in a much more stable linkage. Note that the reaction proceeds at pH 6.0, a higher pH than other aldehyde reactive groups.

moving forward, the HIPS reagents are protected by Fluorenylmethyloxycarbonyl (FMOC) during storage, and deprotected immediately prior to use. This has been an effective solution in my hands, and the current stocks of HIPS reagents were made in this way (Appendix A).

To quantify the labeling efficiency of the new reagents, I again used LC-MS to analyze trypsin digested samples. This time, samples were taken before and after *in vitro* aldehyde conversion as well as after labeling with HIPS reagents. Labeling with HIPS-Biotin was significantly more efficient than labeling with HIPS-DBCO (>80% vs <35%) because HIPS-biotin is much more soluble in our labeling buffer than HIPS-DBCO, likely due to the hydrophobicity of DBCO. HIPS-DBCO had very limited solubility in our labeling buffer, and would precipitate when added at 10x molar concentration, even when the protein concentration was <0.1 mg/mL. As shown by other labs, aldehyde labeling is highly dependent on the concentration of label, with higher concentrations resulting in significantly improved labeling kinetics and yields.[62] Therefore, it wasn't particularly surprising that I never saw over 35% labeling with HIPS-DBCO and that it had slower labeling kinetics than HIPS-Biotin. When labeling the dual-aldehyde construct, 10 molar equivalents of HIPS-DBCO were first added to the buffer exchanged protein and the mixture was incubated overnight at 37 °C before adding 10 molar equivalents of HIPS-Biotin and continuing the incubation for 3 days at 37 °C. When both reagents were added at the same time, less than 20% of the protein would be labeled with HIPS-DBCO. By delaying the addition of HIPS-Biotin, the percentage of protein labeled with HIPS-DBCO could be increased to 35%.

Interestingly, HIPS-DBCO reacts with higher efficiency to the C-terminal aldehyde tag than the N-terminal aldehyde tag (35% versus 8%). This increases the amount of protein that is labeled with both HIPS-biotin and HIPS-DBCO, as most of the protein labeled with DBCO at the C-terminus

will be labeled with biotin at the N-terminus. A schematic of protein labeling is shown in Figure II-10.

To combat the low solubility of DBCO, commercial maleimide-DBCO reagents have sulfonated derivatives which increase the aqueous solubility. To achieve similar results, David Rabuka's group recently synthesized a HIPS-Sulfo-DBCO for us. This variant has much better
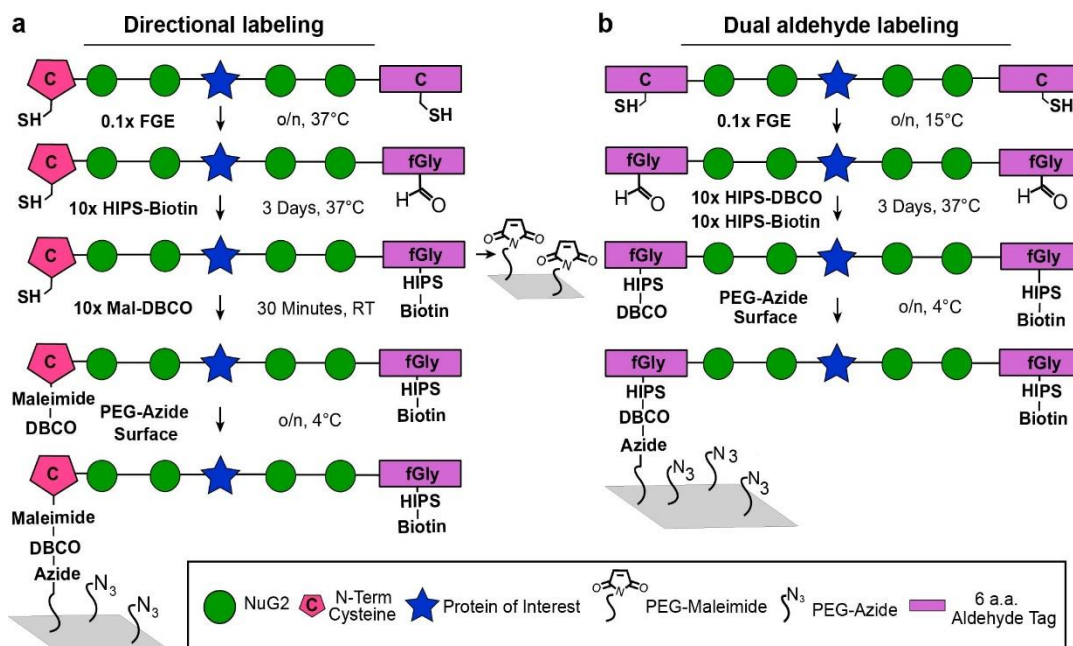
Figure II-10. Overview of two schemes for polyprotein functionalization and surface attachment based upon (A) a directional and (B) a dual-aldehyde construct. The polyprotein consisted of two repeats of two NuG2 (green circles) flanking the protein of interest (blue star). To generate an aldehyde, a 6-amino-acid tag (LCTPSR, purple rectangle) was introduced into the DNA coding for the polyprotein (not shown). After expression and purification, the polyprotein was reacted overnight (o/n) with formylglycine-generating enzyme (FGE), which converts the cysteine only in the short tag into a formylglycine (fGly), an aldehyde-containing amino acid. Next, the aldehyde at one or both ends of the polyprotein was functionalized with a HIPS-based reagent. In both constructs, one end was labelled with a biotin and, in the right column, the other end with a DBCO, a copperless click chemistry reagent that reacts with azide (N$_3$) moieties. If desired, the directional construct could be directly linked to a maleimide functionalized surface using a terminal cysteine or subsequently DBCO-labelled for linkage to an azide-functionalized surface. Note that while the dual-aldehyde construct only yields a fraction of polyproteins with both DBCO and biotin labels, only such heterobifunctionally labelled proteins are efficiently stretched between an azide-functionalized surface and a streptavidin-coated tip.

solubility in our HIPS labeling buffer than HIPS-DBCO. While we have yet to do a full characterization of the compounds ability to label aldehyde modified proteins, it is likely that it will significantly improve the labeling. A simple reaction diagram of the synthesis of FMOC protected HIPS-Sulfo-DBCO is shown in Figure II-11. Detailed synthesis conditions can be found in Appendix A.
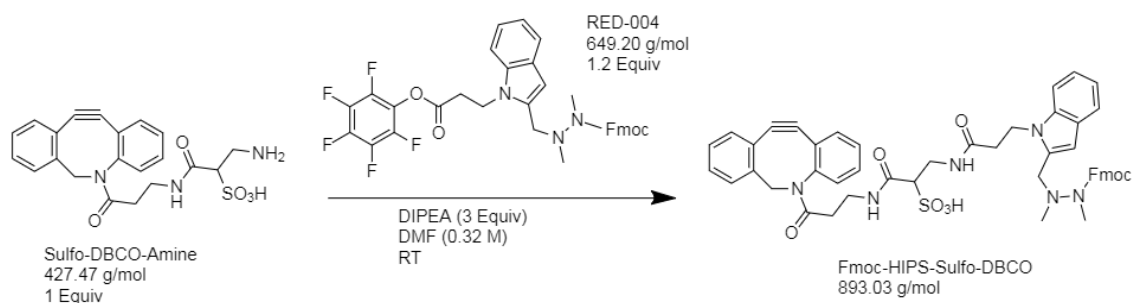
Figure II-11. Synthetic scheme of HIPS-Sulfo-DBCO. Synthesis performed by Catalent Biologics. Note that the final product contains both a sulfate group to increase aqueous solubility and that the HIPS is FMOC protected, preventing degradation during storage.

HIPS Reagents are More Stable than Oxyamine Reagents for AFM Experiments

To test whether HIPS reagents provided a more stable linkage than the oxyamine reagents I originally used, the directional protein construct (aldehyde-NuG2$_4$-cysteine) was labelled with either oxyamine-PEG$_2$-biotin or HIPS-Biotin and attached to a maleimide functionalized surface at a deposition concentration of 0.3 mg/mL. The surface was probed with a streptavidin functionalized cantilever by raster scanning in a grid pattern with spots separated by four μm and each spot was probed three times. Data were collected for 4.5 h, and the number of traces showing protein attachment were binned in 30 min intervals (Figure II-7). While the protein labeled with HIPS-Biotin showed consistent attachment over 4.5 h, the oxyamine-biotin labeled protein showed a rapid decline in protein attachment. After 2.5 h, the number of protein attachments dropped by 60%. As noted above, the reduction in attachment events is likely due

to degradation of the oxime linkage either in solution or when under force. It cannot be determined whether the degradation is force induced or simply the release of free biotin from surface bound proteins in solution. While still a major improvement over non-specific attachment, the use of oxyamine-biotin prevents the reuse of the streptavidin functionalized cantilevers over multiple experiments.

Maleimide Labeling for Dual Cysteine and Directional Constructs

Because the aldehyde tag contains a cysteine (LCTPSR), the polyproteins can be labeled with maleimide reagents if the *in vitro* aldehyde conversion is skipped. Commercially available Maleimide-DBCO and Maleimide-PEG$_2$-Biotin (ThermoFisher Scientific) can be used to bifunctionally label the polyproteins if the protein of interest does not contain solvent accessible cysteines. While this results in stochastic labeling as with the dual aldehyde construct, only those that are doubly labeled will be efficiently measured on the AFM. The directional construct (cysteine-NuG2(4x)-aldehyde), previously used for deposition onto maleimide functionalized surfaces, can also be functionalized with maleimide-DBCO and HIPS-Biotin, providing a fully heterobifunctional protein. Maleimide-DBCO is used instead of HIPS DBCO because while maleimide-DBCO also has low solubility, the maleimide reaction is much more efficient than the aldehyde reaction, yielding better labeling in less time than with HIPS-DBCO. Recently, I've switched to maleimide-sulfo-DBCO, which has much better aqueous solubility and results in better labeling.

Homobifunctional Labeling is Sufficient for AFM-Based SMFS

Because the Dual-Aldehyde polyprotein was homobifunctional, fewer proteins are labeled with both DBCO and biotin, but only proteins that are will be measured during the AFM assay. Our assay was designed to efficiently stretch only proteins heterobifunctionally labelled with DBCO and biotin. Yet, at best, 50% of individual homobifunctional polyproteins will be terminally labelled with DBCO and biotin while 25% will be doubly labelled with DBCO and 25% doubly labelled with biotin. Doubly labelled DBCO protein attaches efficiently to the surface but is unlikely to make a connection to the streptavidin-coated tip for three reasons: (*i*) it lacks biotin for site-specific attachment (Figure II-12), (*ii*) both labelled ends are likely to be anchored to the azide surface via the DBCO moiety, and (*iii*) non-specific adhesion is suppressed by the low
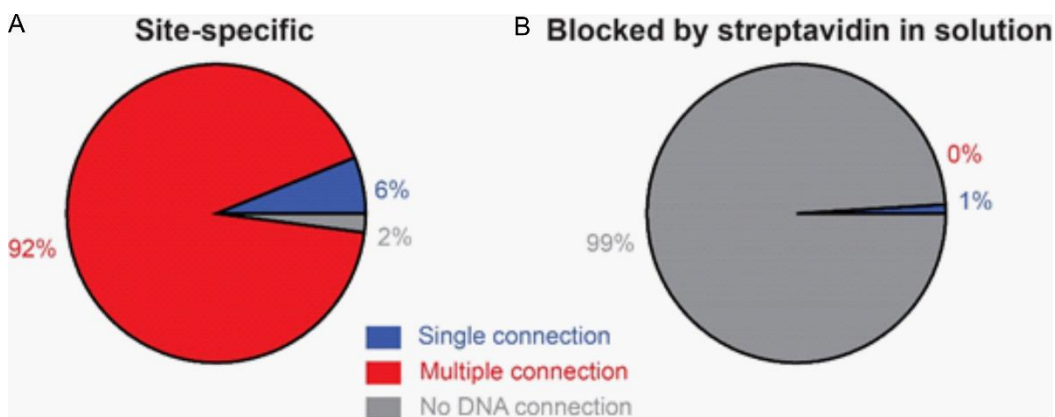
Figure II-12. Control experiment demonstrating site-specific connection between the sample and the tip via a streptavidin-biotin bond. (A) A pie chart quantifies the results when probing a surface densely coated with biotinylated DNA using a streptavidin-coated AFM tip. Most measurements (92%) yielded multiple connections (*red*) when retracting the tip from the surface ($N_{attempts}$ = 200), with a much smaller percentage of force-extension curves showing a single (*blue*) or no connection (*grey*). (B) A pie chart quantifies the same surface after incubating the sample with streptavidin (100 μL of 200 μg/mL for 1 h) to block site-specific attachment to the biotinylated DNA. Essentially all measurements (99%) yielded no connections (*grey*) with 1% of the records showing a single connection (*blue*) ($N_{attempts}$ = 200). Hence, the attachment between the tip and DNA arises from a site-specific streptavidin-biotin bond. Data collected and prepared by Rob Walder.

contact force (100 pN) along with the tip's PEG coating. For protein doubly labelled with biotin, the protein lacks the functionalization to site specifically bind to the azide-PEG-coated surfaces and should be efficiently removed during subsequent washing of the relatively non-stick surfaces. Therefore, while we only expected to label 50% of proteins with both biotin and DBCO based on stochastic labeling, we could make up for the lower labeling efficiency by simply depositing a higher concentration of protein.

Site-Specific Attachment to the AFM Surface and Tip Yields 70-Fold Increase in High-Quality Data

I demonstrated improvements in the rate of acquiring high-quality protein records by incorporating our site-specific bioconjugation scheme with the polyprotein construct (Figure II-13). Specifically, I deposited the empty polyprotein, consisting of four repeats of NuG2 [*i.e.*,
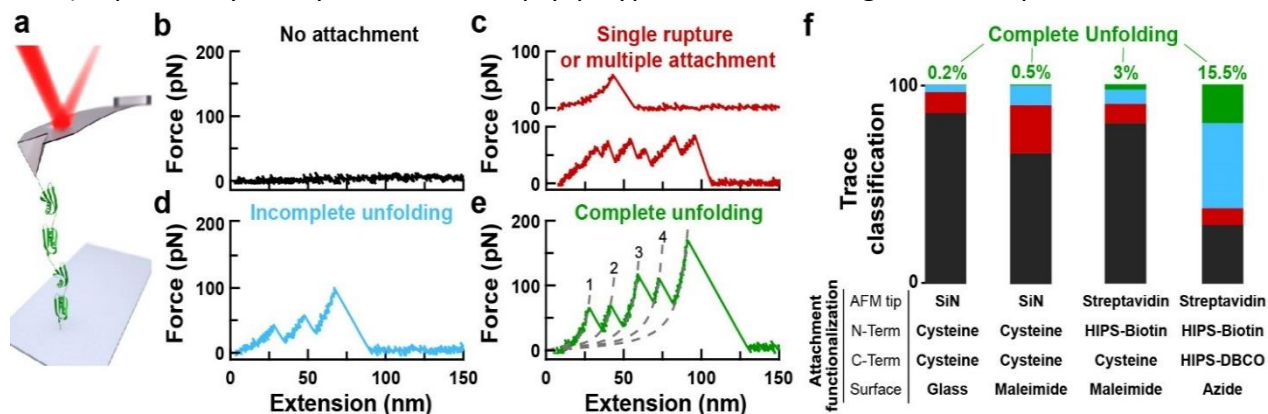
Figure II-13. Improved single-molecule force spectroscopy. (A) A schematic of the experiment shows a polyprotein with four NuG2 domains being stretched between the surface and an AFM tip. (B-E) Force-extension curves (FECs) typically show one of four classes of mechanical fingerprints ranging from no attachment to full unfolding of the polyprotein. In panel e, the segments of the FEC between domain ruptures are well described by a worm-like-chain model (dashed lines). FEC data smoothed to 1 kHz. (F) Bar graphs characterizing the fraction of records with no unfolding (black), traces showing a single unfolding event or pulling on multiple proteins in parallel (red), incomplete unfolding of a single polypeptide (blue), and complete unfolding (green). Site-specific attachment to the functionalized, PEG-coated surface used either a maleimide-thiol reaction or a reaction between dibenzocyclooctyne (DBCO), a copperless-click-chemistry reagent, and an azide-derivatized surface. Hydrazino-Pictet-Spengler (HIPS) reagents labelled genetically encoded aldehydes with biotin and DBCO. Going from left to right, the bar graphs were based on 595, 595, 1237, and 898 individual stretching attempts. SiN, silicon nitride.

(NuG2)$_4$], with variable chemical functionalizations at a fixed concentration (0.3 mg/ml) (Figure II-13A). The tip was gently (100 pN) pressed into the surface for 1 s and then retracted at a constant velocity (400 nm/s), changing locations every third attempt. When using nonspecific adhesion, the vast majority (85.5%) of attempts yielded no attachment (Figure II-13B). Less often, we observed a trace showing a single rupture or pulling on multiple proteins in parallel (Figure II-13C) or a trace showing incomplete unfolding of a single polyprotein (Figure II-13D), at 10.8% and 3.5% respectively. The desired high-quality data, the full unfolding of the polyprotein (Figure II-13E), was observed least frequently (0.2%). As summarized in Figure II-13F, the sequential introduction of site-specific surface coupling and tip attachment led to an increased yield of high-quality data. In particular, we achieved a ~75-fold increase in the rate of acquiring high-quality data when using the dual-aldehyde construct in conjunction with an azide-functionalized surface and a streptavidin-coated tip. Optimization of protein deposition concentration is likely to further increase this improvement.

Site-Specific Attachment Enables Repeated Unfolding of Individual Polyproteins

While collecting data from many molecules is an important part of AFM SMFS experiments, it can mask variations in protein unfolding. Heterogeneity in the folding and unfolding of individual RNA molecules has been observed by single-molecule fluorescence.[67,68] Such heterogeneity is masked in traditional AFM studies where dynamic force spectra are derived from a large number of rupture forces from a multitude of different molecules. To demonstrate the capability of performing true single molecule experiments, Rob collected a full dynamic spectrum from a single individual polyprotein (Figure II-14 C,D). As illustrated in Figure II-14C, he was able to collect 357 unfolding events in 2 h across multiple pulling speeds. This allowed him

to construct the first reported dynamic force spectrum (Figure II-14C) of a single individual polyprotein collected by AFM.

Importantly, analysis of the force-extension curves (Figure II-14B) yielded a change in contour length (ΔL = 17.4 nm) that matched the literature values for NuG2 ($\Delta L \approx$ 17.6 nm).[61,69] Moreover, when we compared our data with existing results using a Bell-Evans model,[37,69] our results agree, indicating that our attachment scheme does not interfere with protein unfolding. Analysis of a single individual polyprotein yielded k_o = 0.03 ± 0.02 s^{-1} and $\Delta x^{\ddagger}$ = 4.2 ± 0.4 Å, where k_o is the off-rate at zero force, and $\Delta x^{\ddagger}$ is the distance to the unfolding-transition state. These parameters are in quantitative agreement with the recently reported values (0.04 s^{-1}, and 4.2 Å).[69] We therefore conclude that the rupture force of the streptavidin-biotin linkage did not alter
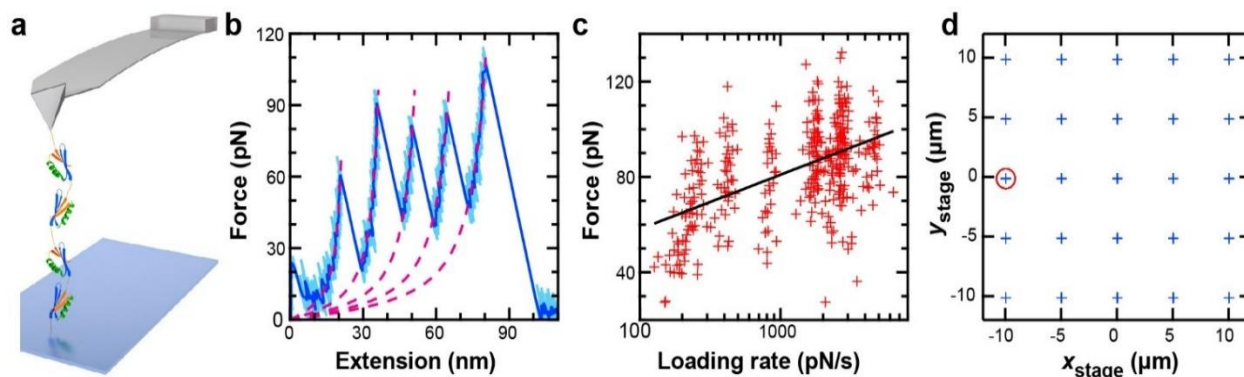
Figure II-14. Dynamic force spectroscopy of a single individual polyprotein. (A) Schematic of experiment showing a polyprotein of four repeats of NuG2 [*i.e.*, (NuG2)$_4$] covalently bound to a PEG-coated surface and reversibly bound to a streptavidin-coated cantilever. (B) A force-extension curve shows the full unfolding of (NuG2)$_4$ and minimal adhesion at small extensions (<10 nm). High-bandwidth data (50 kHz, light blue) were filtered to 200 Hz (dark blue). Magenta dashed lines represent worm-like chain fits. (C) A dynamic force spectrum for NuG2 generated from one individual polyprotein where each data point (red; N = 357) represents the rupture of an individual NuG2 domain. Bell-Evans analysis of this data (black line) yielded values for the zero-force off-rate (k_o = 0.03 s^{-1}) and distance to the transition ($\Delta x^{\ddagger}$ = 4.2 Å). (D) Surface locations (x_{stage} and y_{stage}, blue crosses) probed during polyprotein experiment. All data for (C) were measured from one location (red circle) with no indication of multiple attachments.

the results for NuG2, which is considered a mechanically robust protein.[61] The strong but reversible attachment of the polyprotein to the tip in conjunction with its covalent anchoring to an otherwise nonstick surface significantly accelerated the acquisition of high-quality AFM-based SMFS data and enables probing for kinetic differences between individual proteins and, more generally, rapidly acquiring large, high-quality data sets.

Dual Aldehyde Polyprotein is Suitable to Test Wide Variety of Target Proteins

To demonstrate the versatility of our technology, I tested three different proteins: calmodulin, rubredoxin, an iron-binding protein containing four internal cysteines, and α_3D, a computationally designed three helix bundle[70] not previously characterized by SMFS. These proteins were chosen for three distinct purposes. Calmodulin was well characterized by various SMFS experiments[71–74] and set a nice benchmark for comparison. Rubredoxin, with its four internal cysteines, was an excellent example of the type of proteins that benefit from our bio-orthogonal labels. It also unfolds at relatively high force (higher than NuG2), providing a good test of our streptavidin-biotin linkage. α3D, as an entirely α-helical protein, allowed us to demonstrate the ability to characterize mechanically weak proteins.
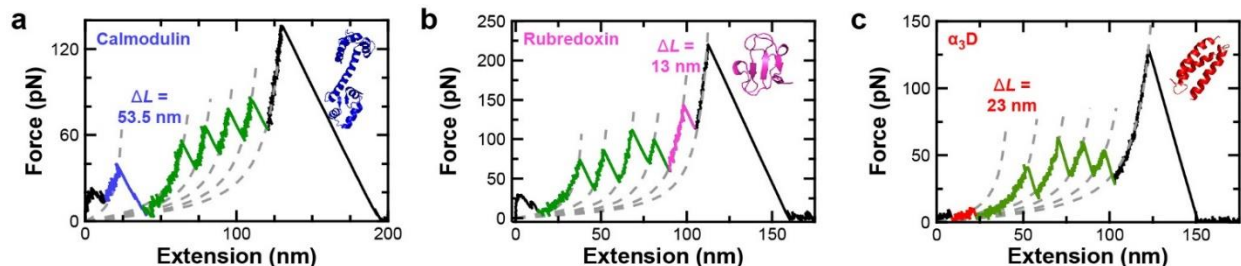
Figure II-15. Efficient SMFS studies of diverse proteins. (A-C) FECs show the complete unfolding of a polyprotein containing calmodulin, rubredoxin and α_3D, respectively. Segments of the FEC between domain ruptures are well described by a worm-like chain model (grey dashed lines). The contour length (ΔL) increases associated with each protein agree with the number of unfolded amino acids and the known structure. FEC data smoothed to 1 kHz.

The plasmid underlying the polyprotein construct enables efficient insertion of new protein sequences through a multiple cloning site amenable to both typical restriction digest cloning and sticky end cloning,[75] a method that does not rely on digestion of the DNA insert. All three proteins expressed well within the dual-aldehyde construct and only minimal amounts of protein were lost during the labeling process. All three proteins yielded high-quality SMFS data with a mechanical fingerprint matching the size of the inserted protein (Figure II-15 A-C). As expected, the rupture forces for calmodulin and rubredoxin agreed with previous studies.[71,76]

Because unfolding of proteins with high α-helical content generally occurs at low forces and low extensions, it is often masked by surface adhesion and the typical AFM force noise of 5–10 pN.[28] Thus, it is noteworthy that minimal surface adhesion was observed in the force-extension curves at low extensions, even when α_3D unfolded at less than 12 pN (Figure II-16A).This is due in part to the combination of site-specific anchoring to PEG-coated surfaces and improved force stability achieved by removing the metallic coating from the cantilevers.[33] Additionally, the traces were exceptionally consistent, requiring minimal lateral shifts. For example, we generated a heat map, a standard AFM analysis, by overlaying 11 individual force-
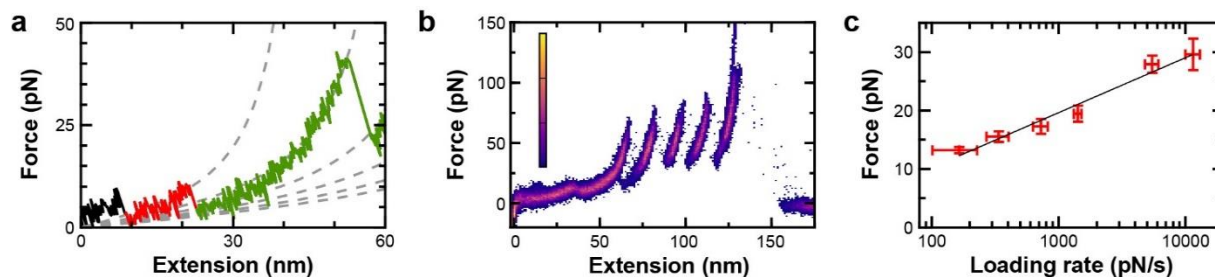
Figure II-16. Dynamic force spectrum for α₃D. (A) A force-extension curve showing an expanded view of α3D unfolding. (B) A heat map of 11 traces of α3D generated with no lateral offset, a commonly applied correction in AFM-based SMFS. (C) A dynamic force spectrum for α₃D shows the mean rupture force as a function of loading rate. Errors represent the standard error of the mean. Analysis of this data with a Bell-Evans model (black line) yielded the distance to the transition ($\Delta x_u^{\ddagger}$ = 1.0 ± 0.1 nm) and the zero-force off-rate (k_0 = 2.0 ± 0.7 s-1). FEC data smoothed to 1 kHz.

extension curves without any lateral offset (Figure II-16B). Importantly, insertion of mechanically labile target proteins (*e.g.*, α3D and calmodulin) into the polyprotein resulted in similarly high efficiencies of the complete unfolding of the full construct (Figure II-17). I was able to demonstrate acquisition of high quality data of diverse mechanically labile proteins while maintaining excellent data throughput.

Site-Specific Attachment Enables Mechanical Characterization of a Mechanically Labile Protein

I leveraged the increased efficiency and high sensitivity at low force to mechanically characterize α₃D, obtaining its dynamic force spectrum (Figure II-16C). To do so, the rupture force was measured as a function of loading rate, a process that can often take days to weeks. In contrast, I acquired a spectrum of this mechanically fragile protein in just ~24 h of instrument time, despite only analyzing force-extension curves that exhibited the full unfolding of the polyprotein (i.e., α₃D plus four NuG2 ruptures followed by tip detachment). Data acquisition could be further accelerated if we optimized the data protocol to acquire hundreds of records from individual molecules and averaged over multiple molecules. However, since this was the

first SFMS study of α_3D, I chose a more traditional approach that incorporated raster scanning of the tip across the protein-coated surface.
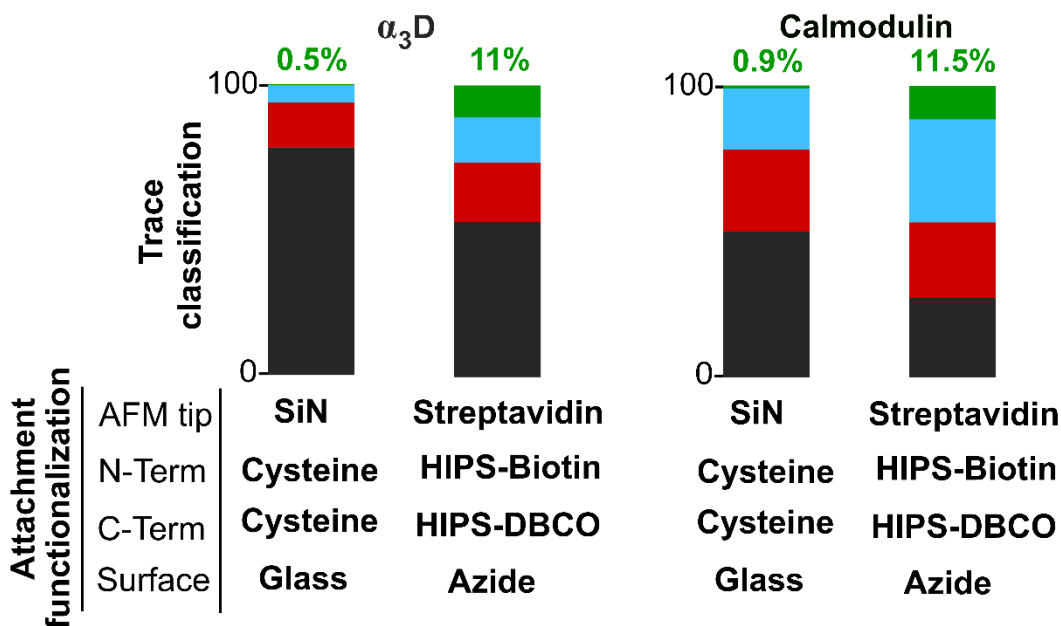
Figure II-17. Comparison of the yield of high-quality curves with and without site specific attachment for polyproteins containing α_3D and calmodulin. Bar graphs characterizing the fraction of records with no unfolding (black), traces showing a single unfolding event or pulling on multiple proteins in parallel (red), incomplete unfolding of a single polypeptide (blue), and complete unfolding (green). Examples of these different types of traces for the polyprotein prior to target protein insertion is shown in Figure II-13. Site-specific attachment to the functionalized, PEG-coated surface used a reaction between dibenzocyclooctyne (DBCO), a copperless-click-chemistry reagent, and an azide-derivatized surface. Hydrazino-Pictet-Spengler (HIPS) reagents labelled genetically encoded aldehydes with biotin and DBCO. α_3D samples were deposited at 3 mg/mL, while calmodulin samples were deposited at 1 mg/mL. Going from left to right, the bar graphs were based on 614, 682, 653, and 611 individual stretching attempts. SiN, silicon nitride.

I analyzed the α_3D dynamic force spectrum using the Bell-Evans model,[43] which is the standard for AFM-based SMFS. For α_3D, this analysis yielded $\Delta x^{\ddagger} = 1.0 \pm 0.1$ nm, and $k_0 = 2.0 \pm 0.7$ s^{-1}. As expected for an all-α-helical protein, α_3D unfolds at low force and is mechanically compliant (large $\Delta x^{\ddagger}$) relative to proteins that contain β-sheet structures,[77] including NuG2 ($\Delta x^{\ddagger}$

= 0.42 nm).[61,69] The mean rupture force of α_3D at 165 pN/s (v = 50 nm/s) is less than that of calmodulin,[71] making it arguably the mechanically weakest protein studied to date by AFM.

Site-Specific Attachment Facilitates Testing of FIB Modified Cantilevers

Modifying cantilevers has recently been shown to be a powerful tool to improving the force stability, precision, and noise in AFM experiments. One of the first major modifications to the AFM cantilever was to remove the gold coating. Traditionally, the gold coating was added to increase the amount of laser light reflected onto the detector, which theoretically improved the signal-to-noise ratio. However, the Perkins lab found that the gold coating was responsible for the significant force drift that plagued AFM experiments for decades.[33] By simply removing the gold and underlying chromium coatings through a brief etching, they drastically reduced the force drift, improving the force precision by an order of magnitude. Moreover, despite the lower amount of reflected light (~10x less), they measured sub-pN force precision that was stabile for over 100 s. This increased precision and stability opened the door to measuring low force events while pulling at slower speeds, a realm that simply wasn't previously accessible in AFM experiments.

The next step in cantilever modification was to modify the shape of commercial cantilevers with focused ion beam (FIB) milling. The ideal AFM cantilever has good force stability over long periods of time and a fast response time. Removing the gold coating from long, soft cantilevers achieved one of those goals.[33] However, temporal resolution was limited due to the size of the cantilever. According to the fluctuation-dissipation theorem $\Delta F = (4k_BT\Delta f\beta)^{1/2}$, where ΔF is the force precision, k_BT is the thermal energy, Δf is the bandwidth of the measurement, and β is the hydrodynamic drag of the cantilever, the smaller the hydrodymic drag, the better the

force precision. The larger a cantilever is, the more hydrodynamic drag it experiences, limiting the temporal resolution. Therefore, many labs moved to shorter, stiffer cantilevers to improve their temporal resolution.[78] While force precision isn't effected by stiffer cantilevers, the Perkins lab found that long-term stability, dominated by instrumental drift, became worse as the cantilever became stiffer[33]. Therefore, researchers had to choose between high temporal resolution with low force noise (stiff cantilever) or long-term force stability with lower temporal resolution and higher force noise. The most recent modifications to AFM cantilevers address this issue,[32,34] providing both long-term stability and high temporal resolution. Testing and characterization of modified cantilevers is significantly easier when using site-specific chemistry.

Through FIB modification, the shape of the cantilever can be changed, making softer (lower spring constant) short cantilevers. The modifications resulted in short, soft cantilevers with even less hydrodynamic drag, that maintained fast response times and long-term stability that rivaled the long, soft cantilevers.[31,34] The most recent advance has resulted in what is known as a "Warhammer" cantilever geometry, consisting of a 40-µm-long cantilever with a 4 x 4 µm^2 gold coated patch at the end of a 2-µm-wide shaft (Figure II-18A). Made by focused ion beam (FIB) milling, these cantilevers combine 9 µs temporal resolution with sub-pN force stability over 100 s and are capable of 0.1 pN force precision (Figure II-18 B-C)[32]. The increased resolution and precision is important for my studies because it makes it much easier to identify protein unfolding at low force where the deflection of the cantilever is small.
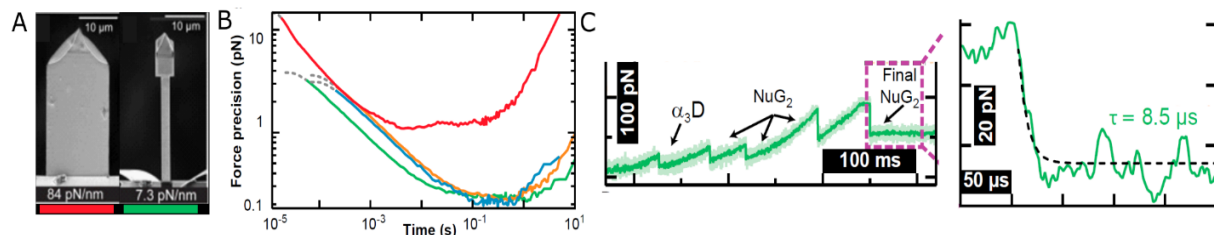
Figure II-18. Improvements in force precision and temporal resolution by FIB modification. (A) Images of an unmodified Biolever Mini (red) and a FIB modified "Warhammer" (green) (B) Allan deviation of the unmodified cantilever (red) and the Warhammer (green). The Warhammer achieves 0.1 pN force precision, on order of magnitude better than the Mini. (C) Measurement of the temporal resolution of the Warhammer. A polyprotein was unfolded until only one domain remained folded and held at 80 pN until the final NuG2 unfolded. An exponential curve was fit to the unfolding event to extract the temporal resolution. Adapted from Edwards *et. al.* (2017).

Discussion

Maleimide Functionalization, Ideal for Tips, is Insufficient for Surfaces

Initially, the newly developed salinization protocol for surfaces and tips was used to create PEG-maleimide functionalized surfaces and tips because maleimide functionalization was widely used in the field.[44] The advantage of maleimide functionalization is it allows for rapid deposition (<30 m) of proteins containing solvent accessible cysteines at near neutral pH (pH 6.5-7.5). However, maleimides are quickly hydrolyzed in aqueous solution,[79] meaning that maleimide functionalized surfaces or tips had to be made fresh before every experiment.

Because the surfaces would be used to test a variety of proteins, the inability to store maleimide functionalized surfaces was a major drawback. However, the tips were only functionalized with a cysteine tagged variant of streptavidin. Therefore, tips could be prepared in batches and immediately functionalized with streptavidin, thereby avoiding maleimide

hydrolysis. Once the tips were functionalized with streptavidin, they could be stored for up to two weeks at 4°C without a significant drop in activity.

HIPS Labeling Results in Acceptable Protein Loss

It is important to note that not all proteins of interest were stable at pH 6.0. Indeed, most of the tested proteins had some precipitation in our HIPS labeling buffer. While it is possible to do the HIPS-labeling at pH 7.0, the kinetics are significantly slower[66]. I decided that losing a portion of the protein was an acceptable tradeoff to avoid exposing the proteins to a longer incubation. Because AFM SMFS requires only a very small amount of protein I was consistently able to get sufficient amounts of labeled protein to run experiments, despite losing some protein during incubation. It is this step that presents the main challenge when preparing a new protein of interest for mechanical characterization. If the aldehyde labels (oxyamine, hydrazide, or HIPS) are soluble, increasing the concentration should allow aldehyde labeling to be carried out at a higher, more favorable pH without a reduction in labeling efficiency.[62]

Conclusion

To accelerate the acquisition of high-quality SMFS data by AFM, I co-developed a new polyprotein construct featuring a versatile and mechanically robust HIPS chemistry in parallel with a conjugation protocol to site-specifically anchor biomolecules to otherwise non-stick surfaces in an accessible and significantly simplified manner. To do so required the optimization of FGE conversion, moving from an *in vivo* to an *in vitro* scheme, optimization of aldehyde labeling, and switching labels to include a more stable chemical handle.

Using protein labelled at one end with DBCO, a copper-free click chemistry reagent, and biotin at the other end, I demonstrated the increased acquisition rate. The DBCO label enabled us to leverage the efficiency and bio-orthogonality of click chemistry while avoiding the detrimental effects of copper ions on many biomolecules.[54] The biotin label provided a simple and accessible means to strongly but reversibly attach the DNA or polyprotein to a streptavidin coated tip, avoiding an undesired macromolecular coating at the apex of a tip.

This reversible linkage allowed hundreds to thousands of individual molecules to be probed and re-probed by a single cantilever over days to weeks, improving throughput and precision. Overall, we achieved a 75-fold increase in the yield of high-quality protein-unfolding data in comparison to traditional AFM-based assays that rely upon non-specific attachment. Our advances in site-specific coupling are not limited to AFM-based SMFS, but should benefit other SMFS modalities. Indeed, Rob demonstrated such broader utility by overstretching DNA, a relatively high-force transition[80,81] using both an AFM and an optical trap. Additionally, our labelling scheme using a cysteine and a 6-amino-acid tag can be extended to internal sites within a protein to control the directionality of the applied force, similar to prior work using a dual cysteine system,[82] but with the added benefit of orthogonal coupling chemistry. More broadly, the increased rate of acquiring high-quality data demonstrated here should facilitate rapid collection of large data sets, enabling more complete investigation of macromolecular folding over a range of physiochemical conditions (pH, temperature, denaturant).

With the advances in surface and tip functionalization, coupled with my experience in protein engineering and labeling, I was in a position to start investigating effector proteins of the Type III secretion system. While modifications were made to the methods for attaching protein

to the AFM surface and tip, this work was fundamental to our ability to robustly test protein unfolding by AFM. The procedures described in this chapter are more than sufficient for studying a wide variety of target proteins and have been used in several papers to help validate and demonstrate improved technical aspects of AFM based SMFS.[32,34,83] However, as with any method, it doesn't work for every target protein and every experimental design. The streptavidin-biotin linkage ruptures at ~150 pN, precluding its use to study proteins that rupture at higher force. Moreover, proteins may be unstable and precipitate or become inactive during the long incubation times necessary for HIPS labeling. While we are interested in further refining this methodology, there is no reason to avoid the use of other site-specific schemes that have been developed by other lab, as will be discussed in the following chapter.

Methods

Preparing Streptavidin Functionalized Tips

An overview of tip functionalization is shown in Figure 4. Before functionalization, the gold and underlying chromium layer was chemically stripped from long, soft cantilevers [L = 100 µm; k = 6 pN/nm BioLever (Olympus)] to improve force stability.[33] To do so, the cantilevers were first washed by sequential immersion in toluene, isopropanol, and ultrapure water in individual 100-mL beakers and dried by gently pressing the chip against a Kimwipe. Next, individual cantilevers were immersed in gold etchant (Type TFA, Transene) in a 10-mL beaker for 30 s, ultrapure water in a 500-mL beaker for 40 s, chromium etchant (Cr Etchant, Transene) in a 10-mL beaker for 30 s, and then a second 500-mL beaker of ultrapure water for 40 s. Finally, the cantilevers were dried by gently pressing the chip—not the cantilever—against a Kimwipe to wick away excess water. Before silanization, we irradiated them with UV light for 30 min (PSD-UV8,

Novascan Technologies, Inc.). The cantilevers were immediately immersed in a 0.15 mg/mL solution of silane-PEG-maleimide (MW_{PEG} = 600; PG2-MLSL-600, Nanocs Inc.) in toluene at 60 °C in a glass-capped 50-mL beaker for 3 h. To increase the consistency of the functionalized tips, the toluene solution was stirred at ~600 rpm. A custom Teflon container was made to fit inside the beaker to prevent the stir bar from destroying the cantilevers. After the 3 h incubation, cantilevers were sequentially rinsed in toluene, isopropanol, and ultrapure water before drying by pressing the chip against a Kimwipe.

A commercially available variant of streptavidin that was thiol-modified (SAVT, Protein Mods LLC) was then deposited onto the maleimide functionalized tips. Immediately after drying, we immersed the cantilever tips in 50 μL of 200 μg/mL modified streptavidin dissolved in PBS (pH 6.75) supplemented with 1 mM TCEP [Tris(2-carboxyethyl)phosphine] for 3 h at room temperature. The lower pH was used to reduce maleimide hydrolysis while maintaining in the optimal pH range (6.5–7.5) for maleimide-thiol reactions.[84] After protein deposition the coated cantilevers were then washed by rinsing the cantilevers in 3 separate 10-ml beakers for 30 s each in PBS (pH 7.4) to remove unreacted streptavidin. Streptavidin-coated cantilevers can be stored at 4 °C in PBS using a sealed wafer holder within humidity chambers (pipette racks with water in the base sealed with parafilm) for up to 2 weeks. Importantly this protocol can be used with any of the cantilevers commonly used for SMFS, including focused-ion-beam modified (FIB) BioLever Mini[31] and FIB-modified BioLever Fasts.[34]

Preparing Azide Functionalized Surfaces

An overview of the surface functionalization is shown is Figure 4. Circular glass coverslips are first cleaned with acetone, ethanol, and 3 M ethanolic KOH and then rinsed twice in ultrapure

water. We dried the slides with a gentle stream of dry nitrogen (N_2). The coverslips are then irradiated with UV light for 30 min (PSD-UV8, Novascan Technologies, Inc.) to increase the efficiency of the silane coupling. While initially the surfaces were simply placed on the bottom of a glass-capped beaker containing a 0.15 mg/mL solution of silane-PEG-azide (MW_{PEG} = 600, PG2-AZSL-600, Nanocs Inc.) dissolved in toluene, we later found that using a custom Teflon holder which holds the surfaces upright yielded more consistent surfaces. Importantly, the Teflon holder allows the solution to be stirred during incubation which further enhances surface consistency. In either case, the surfaces were incubated for 3 h at 60 °C. The functionalized surfaces were washed for ~30 s sequentially in toluene, isopropanol, and ultrapure water held in individual 200-mL beakers and then gently dried them using a stream of N_2. It is critical to rapidly transfer the surfaces from the functionalization solution (silane-PEG-azide) into the toluene wash solution to prevent drying and PEG agglomeration. Azide-functionalized surfaces stored dry at 4 °C in sealable, 1" wafer holders (*e.g.*, H22-101-0615, Entegris) remained functional for up to 1 month, with optimal results in the first 2 weeks.

Design of Aldehyde Tagged Polyproteins

My aldehyde tag plasmids were based on the pET-32 Xa/LIC vector. All primers used in this work are listed in Appendix B. The initial vector, provided by Ting-Fang Wang,[85] included both the cleavable SUMO tag and a 6x-His tag to increase protein solubility and allow for Ni^{2+} affinity purification (pMS 927). A small linker region was inserted and the antibiotic resistance was switched to kanamycin by Sandra in our lab (pMS 984). Because it was previously reported that aldehyde conversion (reaction with FGE) could be done *in vivo* by co-expressing FGE, we inserted a pET-Duet cassette containing our polyprotein construct and a copy of *M. tuberculosis*

formylglycine generating enzyme (Mt. FGE) (pMS 1087), provided by Xinghua Shi. This allowed for co-expression of Mt. FGE and the polyprotein construct.

The core of the polyprotein construct was synthesized by Genscript and contained four copies of NuG2,[86] with a multiple cloning site between the second and third NuG2. The NuG2 repeats were first inserted into pMS 984 SfoI/SacI (primers U1063, L946, L947) while adding an N-terminal cysteine tag (pMS 1059), before being moved into pMS 1087 NdeI/SacI using U2007/L2007, which added a C-terminal aldehyde tag[42] (LCTPSR) (pMS 1088) [pET-32/Duet (Cys-NuG2$_4$-Ald) (Mt. FGE)]. To use EcoRV for the insertion of proteins of interest, I had to remove the EcoRV site from the DNA sequence coding for FGE using the QuikChange site-directed mutagenesis kit (primers U2086/L2086). This construct was labeled the "directional construct" because of the ability to heterobifunctionally label the protein using cysteine-mediated and aldehyde-mediated labeling.

For the dual-aldehyde, or homobifunctional polyprotein construct, I first created a complementary construct with the cysteine tag on the C-terminus and an aldehyde tag on the N-terminus (pMS 1091T). To this end, I amplified the NuG2$_4$ construct out of pMS 1088 using U2041, L2041, and L2041.1 in a three primer system. U2041 adds the aldehyde tag to the N-terminus, while L2041 and L2041.1 add the cysteine tag to the C-terminus. The construct (Ald-NuG2$_4$-Cys) was inserted into pMS 984 digested SfoI/SacI (pMS 1091T). I then digested pMS 1088 (C-terminal aldehyde tag) and pMS 1091T (N-terminal aldehyde tag) NheI/NcoI. The insert from pMS 1091 was ligated into the digested pMS 1088, yielding a construct with aldehyde tags at both termini (pMS 1251) [pET-32/Duet (Ald-NuG2$_4$-Ald) (Mt. FGE)].

Polyprotein Expression and Purification

All polyproteins were overexpressed in *E. Coli* strain BL-21 (DE3) transformed with the appropriate plasmid. After transformation, 1 L cultures were grown to an $OD_{600} \approx 0.55$ and then cold-shocked by transfer to an ice/water bath for 20 min. Protein expression was induced with 0.2 mM IPTG (concentration determined by expression test shown in Figure 5) and the cultures incubated overnight at 18 ºC. Cells were harvested using a Beckman JLA 8.1000 rotor (3,500 RPM, 30 min, 4 ºC), the cell pellet was resuspended in 10 mL lysis buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 1 mM PMSF) and lysed using an EmulsiFlex-C3 homogenizer. Cell lysate was clarified by centrifugation on a Beckman JA 20 rotor (15,500 RPM, 15 min, 4 ºC), and the clarified lysate was added to Ni-NTA (Qiagen) beads (1.5 mL beads per liter of culture) equilibrated with TBS (25 mM Tris, pH 8.0, 150 mM NaCl). The mixture was incubated for 1 h at 4 ºC. The beads were washed with 5 column volumes (CV) of wash buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 20 mM imidazole) and eluted with 3 CV elution buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 250 mM imidazole). While my early purifications used 5 mM BME as the reducing agent, I noticed some reduction of the nickel despite the relatively low concentration of BME. Switching to 2 mM TCEP alleviated the issue and slightly improved the yield of purified product.

All of my polyproteins contain a 6xHis-SUMO tag[85] for increased solubility and ease of purification. However, if left on the protein after purification, it can cause protein dimerization and aggregation. I therefore cleaved the SUMO tag during dialysis using ULP1. The eluted protein was diluted with an equal volume of TBS to reduce the concentration of imidazole before adding ULP1 (1:1000 w/w) (a plasmid for expression of ULP1 was a kind gift from Dr. Christopher Lima

(Memorial Sloan Kettering Cancer Center)). The sample was dialyzed into TBS + 5 mM BME, and further purified using a Superdex 200 16/600 size exclusion column (SEC) equilibrated with TBS + 5 mM BME. Peak fractions were pooled and concentrated to ~5 mg/mL using a 10,000 Da molecular weight cutoff (MWCO) centrifugal concentrator (Amicon). Typical yield was ~20 mg of protein per liter of culture.

Labeling Aldehyde Modified Protein with Alexa-Fluor 350 Hydrazide

There were two protocols in the literature for labeling aldehyde modified proteins with hydrazide or oxyamine reagents. The first was developed by T.J. Ha's lab and relied on a high concentration of dye in a neutral pH labeling buffer (250 mM KH_2PO_4, pH 7.0, 500 mM KCl, 5 mM DTT) incubated at 4°C.[62] The second, developed in the Bertozzi lab, used a more traditional low pH aldehyde reaction buffer (100 mM MES, pH 5.5) at 37°C to facilitate a hydrazide or oxyamine reaction with the aldehyde[42]. While the Bertozzi protocol uses a harsher environment, it yields efficient labeling and is the more commonly used method for labeling aldehyde modified proteins. Notably, the Bertozzi protocol requires less dye to achieve a similar level of labeling. Both methods were reported to efficiently label aldehyde modified proteins during an overnight incubation.

Protein was exchanged into the appropriate labeling buffer using centrifugal concentrators (Amicon 0.5 mL, 10 kDa). The buffer exchanged protein was added to dried aliquots of Alexa Fluor 350 Hydrazide, to yield final dye concentrations of 100 mM in the Ha buffer and 15 mM in the Bertozzi buffer. The reactions were incubated overnight at either 4°C or 37°C respectively. After incubation, I used a centrifugal concentrator to remove some of the excess

dye before running the sample on an SDS-PAGE gel. Before staining, I visualized the dye-labeled protein using a 365 nm UV light.

Labeling with Biotin-PEG₂-Oxyamine and Visualization by Western Blotting

Proteins were labeled in the same buffers used for hydrazide labeling. Biotin-PEG$_2$-Oxyamine in dimethyl sulfoxide (DMSO) (100 mM) was added in 10x molar excess and incubated overnight at either 4°C (Ha Buffer) or 37°C (Bertozzi Buffer). After labeling, the protein was run on an SDS-PAGE gel and blotted onto PVDF membrane. After blocking with 5% milk solution, the membrane was probed with Pierce High Sensitivity Streptavidin-HRP and visualized by chemiluminescent detection of the western blot when exposed to ECL reagent (PerkinElmer).

LC-MS of Trypsin Digested Samples Testing for Aldehyde Conversion and Labeling

To test for aldehyde conversion quantitatively, I performed LC-MS on trypsin digested proteins before and after labeling. Protein was first treated with iodoacetamide to label any unmodified cysteines. A mixture of 25 µg protein in PBS (25 mM sodium phosphate, pH 7.2, 150 mM NaCl, 2 mM TCEP), ammonium bicarbonate (100 mM), and iodoacetamide (15 mM) was incubated for 45 min at room temperature in the dark. The mixture was then exchanged into buffer containing 7 M urea, 50 mM Tris, pH 7.75, using a 10,000 Da MWCO centrifugal concentrator (Amicon) before boiling for 5 min. After boiling, the protein was exchanged into 50 mM Tris, pH 7.75 (final volume ~100 µL) using the same centrifugal concentrator. 1 µg of trypsin (Promega) was added and the reaction was incubated overnight at 37 °C. The reaction was quenched with formic acid (1.75%).

Ultra-high performance liquid chromatography (UPLC) coupled to a high-resolution mass spectrometer (SYNAPT G2 HDMS, Waters Corp.) was used to mass analyze trypsin-generated peptides. Online peptide desalting was achieved following sample injection into a Waters HDX Manager nanoAcquity instrument using a VanGuard 2.1 x 5 mm C18 UPLC column (Waters) over 3 min at 100 µL/min 0.1% formic acid in water. High-resolution separations were then achieved using an Acquity 1.0 x 100 mm C18 UPLC column (Waters) and a 12 min gradient with 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B) at a flow rate of 40 µL/min (Table 1). Data were analyzed manually by generating extracted ion chromatograms with 30 ppm mass tolerances for the respective monoisotopic masses (Table 2). Following Savitzky-Golay peak smoothing, the integrated peak areas were used to quantify the populations of cysteine, carboxyamidomethylcysteine, or formylglycine-containing peptides. If the protein was labeled through the formylglycine, the amount labeled with HIPS-DBCO and HIPS-Biotin was also quantified.

Time (m)	%A	%B
0	92	8
6	60	40
7.5	15	85
9	92	8
12	92	8

Table 1. Chromatography gradient used for LC-MS. Solvent A is 0.1% formic acid in water and solvent B is 0.1% formic acid in acetonitrile.

N-terminal Peptide- MAPLCTPSR				
Peptide Modification	Mass Change	+3	+2	+1
Cysteine		325.8299	488.2412	975.475
Carboxyamidomethylcysteine	57.0215	344.837	516.7519	1032.4965
Formylglycine	-17.9882	319.8338	479.2471	957.4868
w/HIPS DBCO	+483.2646	486.918	729.8734	1458.7396
w/HIPS-Biotin	+493.2847	490.2581	734.8835	1468.7597

C-terminal Peptide- SSGLCTPSR				
Peptide Modification	Mass Change	+3	+2	+1
Cysteine		303.1482	454.2187	907.4302
Carboxyamidomethylcysteine	57.0215	322.1554	482.7295	964.4517
Formylglycine	-17.9882	297.1522	445.2246	889.442
w/HIPS DBCO	+483.2646	464.2364	695.851	1390.6948
w/HIPS-Biotin	+493.2847	467.5765	700.8611	1400.7149

Table T2. Monoisotopic peptide masses for trypsin digested proteins.

Cloning of FGE Expression Plasmid

The DNA for M. tuberculosis formylglycine generating enzyme was amplified from plasmid pBAD/myc–his A Rv0712 (FGE)[4] (Addgene plasmid #16132) using primers U2139 and L2139. This fragment was digested with Xho and ligated into pMS 984 digested with Sfo and Xho restriction sites to yield a fusion with a cleavable 6xHis-SUMO tag for purification (pMS 1243).

FGE Expression and Purification

The current purification protocol is essentially the same as described in Holder *et. al.*,[64] with minor modifications. Briefly, *E. Coli* strain BL-21 (DE3) cells transformed with pMS 1243 are grown in 1 L culture to an $OD_{600} \approx 0.55$. The incubation temperature is reduced to 18 ºC and protein expression was induced with 0.2 mM IPTG. After an overnight incubation, cells were harvested and lysed as described in the polyprotein purification. Before the cell lysate was clarified, 1% (w/v) streptomycin sulfate was added to remove DNA. After a 15-minute incubation at 4 °C, the lysate was clarified as previously described, mixed with Ni-NTA agarose beads and rotated for ~1 h at 4 °C before washing away unbound proteins with 10 CV wash buffer. FGE was eluted with 3 CV elution buffer and the SUMO tag was cleaved with 1:1000 molar ratio of Ulp1. The protein was further purified by loading onto an Superdex 200 16/600 size exclusion column equilibrated with storage buffer (25 mM TEA, pH 7.4, 8% (v/v) glycerol). Peak fractions were pooled and concentrated using a 10,000 Da MWCO centrifugal concentrator (Amicon) to a final concentration of approximately 10 mg/mL and stored at -70 °C. Typical yield was 20–30 mg per liter of culture.

In Vitro Aldehyde Conversion

While the polyprotein purification remained the same, minor additions were made to facilitate *in vitro* aldehyde conversion. After elution, the polyprotein is diluted 2x in TBS and 1:1000 molar ratio of Ulp1 is added to cleave the SUMO tag. The protein is dialyzed directly into FGE conversion buffer (25 mM triethanolamine (TEA), pH 9.0, 50 mM NaCl, 1 mM BME). While doing SEC at this stage does increase the final purity of the protein, it isn't necessary to achieve a good sample for AFM. This is because the protein will go through two more buffer exchanges

(aldehyde reaction buffer and storage buffer/AFM buffer) and the SEC runs used for those exchanges are typically sufficient to achieve a high purity final product. Importantly, FGE conversion buffer does not impact the ability of ULP1 to cleave the SUMO tag. After dialysis, the purified FGE was added in a 1:10 FGE:polyprotein molar ratio. The mixture was incubated overnight at 18 °C with shaking. This consistently resulted in 95% conversion of the cysteine to formylglycine, as quantified by LC-MS.

Labeling Aldehyde Tagged Protein with HIPS-Biotin and HIPS-DBCO

The current HIPS labeling protocol is based on the protocol from the Rabuka group[66]. While initially my labeling buffer included 0.1% Triton-X100 in an attempt to increase protein stability, I found it didn't make a significant difference to protein stability and hampered the ability to properly quantify protein concentration. I removed the Triton X-100 from the protocol without any noticeable change in labeling efficiency. To label aldehyde modified proteins, the protein was exchanged into HIPS labeling buffer (50 mM sodium phosphate, pH 6.0, 150 mM NaCl, 1 mM TCEP (Tris(2-carboxyethyl)phosphine)), using a size exclusion column (Superdex 75 10/300 GL). The peak fractions were pooled and the protein concentrated to 0.5–2 mg/mL in a 10,000 Da MWCO centrifugal concentrator (Amicon).

For the directional construct, 10 molar equivalents of HIPS-Biotin were added and the mixture was incubated for 3 days at 37 °C. This resulted in more than 90% of protein being labeled with HIPS-Biotin as determined by LC-MS. The reaction can be performed with only an overnight incubation, resulting in 77% of the protein being labeled. Excess label was removed using a size exclusion column (Superdex 75 10/300 GL) equilibrated with PBS (25 mM phosphate, pH 7.1, 150

mM NaCl, 1 mM TCEP). Peak fractions were pooled and concentrated to 0.5–2 mg/mL in a 10,000 Da MWCO centrifugal concentrator (Amicon).

Maleimide Labeling of Dual Cysteine and Directional Constructs

To label either the directional or homobifunctional construct (without aldehyde conversion), the protein was exchanged into PBS (pH 7.4) using a size exclusion column (Superdex 75 10/300 GL). Peak fractions were pooled and the protein was concentrated to 0.5–2 mg/mL in a 10,000 Da MWCO centrifugal concentrator (Amicon). For labeling the directional construct, 10 molar equivalents of Maleimide-DBCO or Maleimide-Sulfo-DBCO (Click Chemistry Tools) was added, and the mixture was incubated overnight at room temperature in the dark. If labeling the homobifunctional construct, 10 molar equivalents of Maleimide-DBCO or Maleimide-Sulfo-DBCO and Maleimide-PEG$_2$-Biotin (ThermoFisher Scientific) were added and the mixture was incubated overnight at room temperature in the dark. After incubation, excess label was removed using a size exclusion column (Superdex 75 10/300 GL) equilibrated with PBS (pH 7.4). Peak fractions were pooled and concentrated using a 10,000 Da MWCO centrifugal concentrator (Amicon), to a final protein concentration of 0.5–2 mg/mL. Protein was snap-frozen in liquid nitrogen and stored at -70°C.

Protein Deposition and Surface Preparation for AFM

Glass coverslips functionalized with PEG-maleimide or PEG-azide were mounted to metal pucks with epoxy, allowing the surfaces to be mounted to the magnetic sample holder in the Asylum Cypher AFM. 20 µL of the polyprotein construct (50–300 ng/µL) was deposited onto the functionalized coverslip. The coverslip was sealed in a wafer holder contained in a simple humidity chamber for 1 h (maleimide surface) or 4–24 h (azide surfaces) at 4 °C. To remove

unbound molecules after this incubation, samples were rinsed 10 times with 1 mL of PBS (150 mM NaCl, 25 mM phosphate buffer pH 7.4). Rinsing was done by flowing the buffer over an angled surface, taking care to avoid de-wetting the surface. Recent work by Patrick in the Perkins lab indicates that this washing procedure may subject deposited proteins to significant force from the flow of the buffer. An alternative approach is to flood the wafer holder with buffer and vacuum off the excess 10 times. It has not currently known whether this is necessary for all samples, or only those that unfold at low force and are unable to refold.

AFM Assay and Analysis

AFM experiments were performed on a commercial AFM (Cypher ES, Asylum Research). Cantilevers were calibrated using standard protocols.[87] Specifically, the stiffness was calibrated using the thermal method[87] far from the surface and sensitivity measured by pressing the cantilever into hard contact with the surface. On average, the cantilevers used in this chapter (Biolever Long) had a spring constant $k \approx 8$ pN/nm. This is relatively soft for AFM cantilevers, due to the long cantilever length. As is typical in AFM-based SMFS, force-extension curve acquisitions were initiated by bringing the cantilever into hard contact with the surface (v = 400 nm/s). However, to reduce non-specific attachment, the tip was pushed into the surface for a brief period (~1–2 s) at 100 pN, approximately 10-fold less force than is generally used to promote non-specific attachment.[43] The tip was then retracted at 50–3,800 nm/s while digitizing at 50 kHz. For presentation purposes, these high-bandwidth records were boxcar averaged to the indicated bandwidth.

The sample surface was typically probed with a raster scan. Specifically, the AFM tip was moved in a grid pattern, with locations separated by 4 µm, and each location was probed 5 times.

This standard scheme was further optimized by repeatedly probing the same location when a molecule was detected. The repeated probing of the same individual molecule was discontinued after ~10 consecutive attempts failed to yield a connection.

Force was determined by cantilever deflection after scaling for the sensitivity and stiffness of each cantilever. Extension was calculated by subtracting the deflection of the cantilever from the motion of the piezo controlled stage. For each rupture observed in a force-extension curve the loading rate (pN/s) was determined by fitting a line to the force-versus-time curve in the immediate vicinity of a rupture.

Chapter III Comparing the Mechanical Stability of Effector Proteins

For proteins to be secreted through the T3SS, they must first be unfolded by an ATPase associated with the base of the needle. While analogous to AAA(+) unfoldases which pull substrate proteins through a narrow central pore, the ATPase associated with the T3SS is a weak unfoldase structurally related to F1 ATP synthase rotary motor.[2,6] Early studies of the T3SS showed that fusion proteins containing very stably folded proteins such as GFP or ubiquitin could not be secreted through the T3SS, even if they contained the necessary and sufficient signal sequence.[6,88] The inability for these fusions to be secreted was not related to protein size, but rather dictated by the stability of the protein. This was demonstrated by showing that conformationally destabilized mutants of ubiquitin could be secreted.[89] Because the overwhelming majority of protein stability studies measure thermodynamic stability with thermal or chemical denaturants, the consensus model presumes that thermodynamic stability dictates whether a protein can be secreted.[6]

What is often forgotten when considering the factors that contribute to a proteins ability to be secreted is that proteins must undergo mechanical unfolding before secretion. With this requirement in mind, my central hypothesis is that mechanical stability can serve as a predictor for secretion competence. This differs from the current consensus model which uses thermodynamic stability as a predictor for secretion competency, despite a variety of research showing that melting temperature does not predict unfolding force and thermodynamic stability does not correlate with mechanical stability.[90–94] When this is considered, the protein fusion studies show that mechanical stability can the rate limiting step in protein secretion. If the mechanical stability is always the rate limiting step, there will be a correlation between

mechanical stability and rate at which proteins can be secreted through the T3SS. In support of this idea, a recent paper studying AAA+ proteases showed that local mechanical stability has a strong effect on the rate of unfolding.[95] This may be even more important for the T3SS ATPase since it is a weak unfoldase.

Two possible interpretations arise from the idea that mechanical stability can be the rate limiting step in protein secretion. The first and most simplistic model is that protein unfolding is always the rate limiting step. If this is the case, protein secretion rate will correlate with mechanical stability. In the second model, proteins must be able to unfold below a defined threshold force that the ATPase can produce, but once that threshold is met, other factors become rate limiting. If this is the case, any protein too mechanically stable to be unfolded will not be secreted, but secretion rate does not need to correlate with mechanical stability. To identify which model is correct, it is necessary to test both the *in vivo* secretion rate of effectors and their mechanical stability.

Measuring the *in* vivo secretion kinetics of effector proteins is not trivial. While several studies have shown that effector proteins can be secreted in a defined order,[8,96] only one study has been published that measured *in vivo* secretion kinetics.[97] This study, focused on SopE2 and SptP, found that *in vivo* secretion rates differ between effector. SopE2 and SptP were chosen for study because they have antagonistic activity. SopE2 is a guanyl-nucleotide-exchange factor that activates Cdc42, resulting in changes to the host cell cytoskeleton that facilitate invasion of *Salmonella*.[98] However, after the bacteria is internalized, the host cell cytoskeleton returns to normal. This is caused by SptP, a GTPase-activating protein (GAP), which inhibits Cdc42.[99] Because of this antagonistic activity, the order in which these effectors are secreted is essential

for bacterial pathogenesis and the mechanisms that underlie their secretion order are of particular interest. Previous work characterized the *in vivo* secretion rate of SopE2 and SptP. Using a novel fluorescence approach, they found that SopE2 is secreted ~2x faster than SptP, showing for the first time that effector proteins can have different secretion kinetics[97].

SopE2   SptP
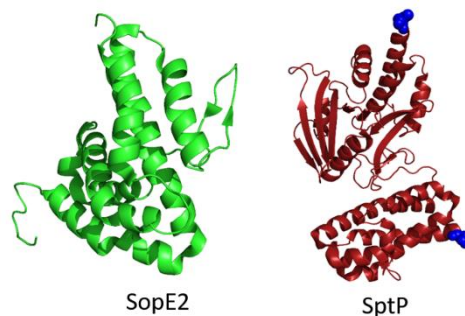
Figure III-1. Crystal structures of SopE2 and SptP. Note that SopE2 is entirely α-helical, while SptP contains β-sheets. β-strand shearing requires more force than unfolding α-helices and may help explain why SptP is secreted ~2x slower than SopE2. PDB accession numbers: 1R9K (SopE2), 1G4U (SptP)

With data available for the *in vivo* secretion kinetics of SopE2 and SptP, I turned my attention to testing their mechanical stability. Interestingly, SopE2 is entirely α-helical, while SptP has extended β-sheets as seen in Figure III-1. Because α-helical proteins typically unfold at lower force,[100] I suspected that SopE2 was less mechanically stable than SptP. This would be consistent with the idea that the amount of force necessary to unfold a protein correlates with its secretion kinetics. I used these proteins to investigate whether mechanical stability correlates with *in vivo* secretion kinetics. With the ability to measure protein unfolding with exquisite force and temporal resolution by AFM, we had a suitable platform to carry out these experiments.

SopE2 and SptP, both *Salmonella* effector proteins, provide a convenient system for testing the correlation between mechanical stability and *in vivo* secretion kinetics. Not only are they the only effector pair with measured *in vivo* secretion rates, but both have known

structures.[101,102] This gives us a roadmap to help identify unfolding intermediates we may find in the AFM traces. It also gave us confidence that the proteins could be overexpressed in *E. Coli* without too much difficulty. Having the structures also allows us to run steered molecular dynamics simulations of the proteins unfolding, a technique which has been critical to elucidating unfolding and rupture pathways of other proteins.[41,103–106]

Results

Implementation of Genetically Encoded Bifunctional Polyproteins for Simplified AFM Experiments

In Chapter 2, I described the development of two novel polyprotein constructs (directional and dual-aldehyde) capable of site-specific attachment to AFM surfaces and tips using the aldehyde tag. These constructs provided high-efficiency acquisition of high quality FEC that enabled the collection of a dynamic force spectrum of the most mechanically labile protein studied to date. The constructs also facilitated advancements in surface and tip functionalization, and greatly simplified the process of testing modified cantilevers. However, the method required reaction with FGE after purification followed by derivatization with custom made reagents in conditions that required optimization for different proteins of interest. Therefore, I continued searching for complimentary methods that would allow for rapidly screening new target proteins with simpler experimental overhead. The ideal method would include genetically encoded tags that did not require post-purification modification, allowing purified proteins to be immediately tested.

At approximately the same time we were developing our polyprotein, Professor Herman Gaub's lab in Munich Germany was also developing systems for site-specific attachment.[41,106–110] Many of their methods covalently attached proteins to Coenzyme A (CoA) functionalized surfaces through the YBBR tag. The YBBR tag is an 11-amino acid tag (D**S**LEFIASKLA) which can be covalently attached to CoA through the hydroxyl group of the first serine, in a reaction catalyzed by phosphopantetheinyl transferase Sfp as shown in Figure III-2.[111,112] This reaction occurs rapidly (30 minutes) at room temperature in neutral pH buffer. Because the tag can be fused to the protein at either N- or C-terminus, it allows the covalent attachment to the surface in either direction. Conveniently, CoA contains a free sulfhydryl, allowing us to use the same protocol detailed in Chapter 2 to prepare PEG-Maleimide surfaces or tips which could then be immediately reacted with CoA. I designed three polyprotein constructs which utilize YBBR for efficient attachment to AFM surfaces. Three constructs were designed to allow different tip attachment schemes, including the aldehyde tag, Dockerin III, and Fgβ, which are discussed in more detail
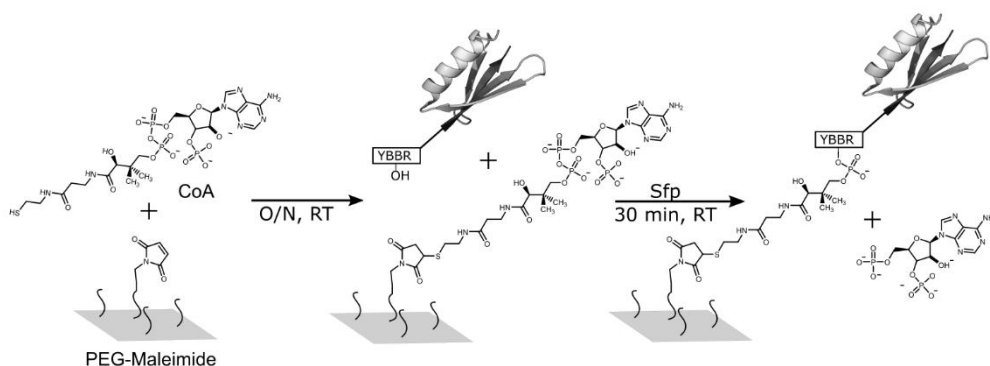
Figure III-2. Attaching YBBR tagged protein to CoA functionalized surfaces. PEG-Maleimide surfaces are incubated with CoA overnight at 25°C, resulting in a CoA functionalized surface. The surfaces can be stored for over a month at 4°C before use. YBBR tagged protein is co-incubated with Sfp for 30 m at RT. Non-bound protein is washed off the surface before testing by AFM.

below.

Cohesin:Dockerin III as a More Force Stable Tip Attachment

While streptavidin binds biotin with very high affinity (K_d ~10^{-14} mol/L)[113] and is widely used as an anchoring system in a wide variety of experiments, the streptavidin-biotin linkage ruptures at 100-200 pN. While this is enough for the characterization of mechanically labile proteins, it can hamper the study of more mechanically stable proteins because the streptavidin-biotin linkage can rupture before the unfolding of the protein of interest has occurred. Therefore, there was a need in the field for another non-covalent method to attach proteins for force probes. While researching cellulosomes, protein networks that degrade biomass, the Gaub lab discovered that Cohesin:Dockerin pairs, which bind with high affinity,[114,115] form a very
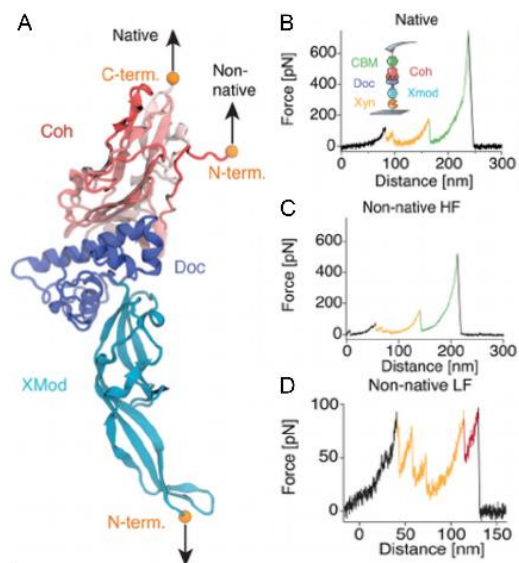
Figure III-3. Cohesin:Dockerin III provides mechanically stable tip attachment. (A) Structure of Cohesin:Dockerin-Xmod (PDB 4IU3) with attachment points marked in orange. In the native geometry, Cohesin is anchored from the C-terminus and Dockerin-Xmod is anchored from the N-terminus. (B) Pulling in the native geometry results in a rupture force >600 pN. (C-D) Pulling in the non-native geometry (Cohesin anchored from N-terminus) results in two unfolding paths, one rupturing at high force (C) and the other at low force (D). Adapted from Schoeler *et al*. NanoLetters (2015).

mechanically stable linkage. Specifically, when stabilized by an X-module (Xmod)- a cellulosome protein of unknown function- the Cohesin:Dockerin III interaction ruptures at >700 pN (Figure III-3 A-B).[116] This makes it one the most mechanically stable linkages currently known, and because it binds with high affinity and specificity, it is an excellent option for force spectroscopy experiments. Interestingly, if Cohesin is anchored from the N-terminus instead of the native C-terminus, the rupture occurs through two pathways (Figure III-3 C-D). Deemed the high-force and low-force pathways, one ruptures at >600 pN while the other ruptures at <100 pN.[106] Therefore, for all our experiments, Cohesin was C-terminally anchored to the tip. While the additional strength was not necessary to study effector proteins, the capability provided additional flexibility, particularly when comparing effector protein to non-secreted homologues (Chapter 4) that may unfold at higher forces than those allowed by the streptavidin:biotin linkage.

Cohesin:Dockerin III Allows for Rapid Testing of α_3D

A plasmid obtained from the Gaub lab allowed for the expression of a polyprotein containing a C-terminal Dockerin III (Figure III-4A). A linker region between ddFLN4 (a marker protein) and Xmod-Dockerin III facilitated the insertion of proteins of interest with Gibson Assembly.[117] This allows proteins to be inserted quickly, without the need for restriction enzymes
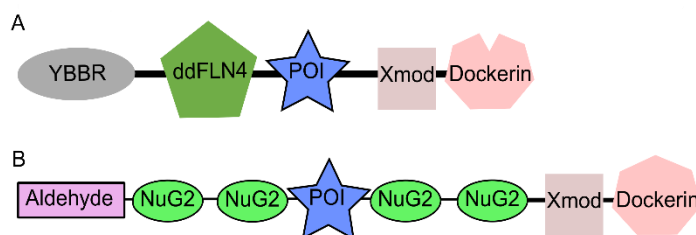
Figure III-4. Dockerin III polyprotein schematics. (A) A polyprotein construct provided by the Gaub lab [6xHis-YBBR-ddFLN4-Xmod-DockerinIII]. (B) A polyprotein construct I designed to utilize the aldehyde tag for surface attachment and Dockerin III for tip attachment.

and without leaving any cloning scars in the final product. As a test of the system, I inserted α_3D into the construct using Gibson Assembly. Expression, and purification of this new α_3D construct was equally simple as the process when using the dual-aldehyde construct and produced purified protein three days after starting the process. Because it contained the YBBR genetically encoded tags for specific-attachment to both the AFM surface and tip, after purification the protein simply needed to be dialyzed into the appropriate buffer and it was ready to test.

In experiments carried out by our collaborator, Dr. Devin Edwards (Perkins Lab), when deposited at a similar concentration, the Dockerin III construct probed with a Cohesin functionalized tip resulted in a much higher rate of multiple attachments than our biotin labeled sample probed with a streptavidin functionalized tip. This is the result of denser protein deposition when using CoA-YBBR to covalently attach proteins to the surface and tip compared to using azide-DBCO or maleimide-cysteine attachment.

Cohesin III Functionalized Tips are Highly Reusable

One of the advantages of using Cohesin:Dockerin is the high stability of Cohesin. Cohesin tips could be used for over a month after being functionalized. Moreover, if the tips were fouled during an experiment (denoted by a decreasing attachment rate), 7 M guanidinium hydrochloride could be used to to denature Cohesin, followed by refolding by immersion in HEPES buffer. This process resulted in complete re-activation of the tip, meaning an individual tip could be used for many protein samples, improving the ability to compare the unfolding forces of different proteins.

However, for all the perks of the Cohesin:Dockerin system, experiments carried out by Dr. Edwards also reported more surface adhesion with the Cohesin functionalized tips than with streptavidin functionalized tips. Therefore, while I created constructs with SopE2 and SptP that used Cohesin:Dockerin for attachment to the AFM tip (Figure III-4 A,B), I was concerned that adhesion may mask the unfolding of effector proteins that were all proposed to unfold at low force. I thus continued to test alternative attachment strategies that could afford high force stability while minimizing surface adhesion.

SdrG:Fgβ Provide Near Covalent Strength Tip Attachment and Low Non-Specific Surface Adhesion

Gram-positive bacteria have developed a sophisticated system of virulence factors to target and adhere to host cells. Known as microbial surface components recognizing adhesive matrix molecules (MSCRAMMs), these systems allow the bacteria to anchor themselves onto host-cells, facilitating pathogenesis.[118] One of the best studied adhesins is the SD-repeat protein G (SdrG) from *Staphylococcus epidermidis*. As shown in Figure III-5, SdrG uses a dock, lock, and latch (DLL) mechanism to bind the N-terminus of the β-chain of human fibrinogen (Fgβ).[119,120] This means
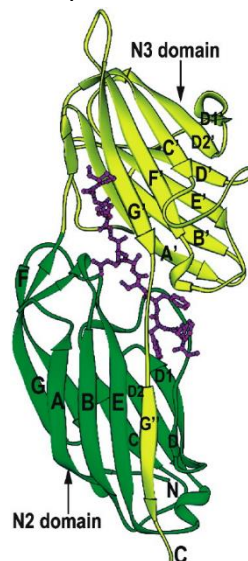
Figure III-5. Structure of SdrG:Fgβ complex, adapted from Ponnuraj *et al.* Cell (2003).

that the Fgβ peptide (purple sticks in Figure III-5) is buried between the N2 and N3 domains of SdrG (green and yellow in Figure III-5), and latched in place by a betta strand from N3 that β-augments into the β-sheet of N2 (Figure III-5). This forms an incredibly stable complex that allow bacteria to remain attached to the N-terminus of fibrinogen against the force of blood flow. A recent study found that this complex required over 2 nN of force before rupture, making it the most mechanically stable complex studied to date.[121]

The rupture force of the SdrG:Fgβ complex rivals the amount of force needed to break covalent bonds (~2.5 nN). This was potentially problematic if we wanted to use this complex as a general method for SMFS. As noted in Chapter 1, the ideal tip attachment for AFM based SMFS is strong but not covalent, allowing the tip to unbind from the protein of interest before breaking internal covalent bonds. This leaves the tip clean and able to repeatedly pick up protein from the surface. If internal covalent bonds break, the tip will be coated protein fragments, preventing attachment to new proteins. However, the Gaub lab showed that SdrG tips could be used to pick up thousands of proteins from a surface before loss of activity, indicating that the SdrG:Fgβ complex tends to rupture before breaking the internal bonds of the protein of interest.[122] Importantly for the characterization of effector proteins, SdrG functionalized tips showed very low adhesion to the surface (Figure III-6C). Therefore, I generated the construct shown in Figure III-6 to mechanically characterize proteins using SdrG:Fgβ for attachment to the AFM tip.

To measure the unfolding of SopE2 and SptP under force, the proteins were incorporated into a construct containing an N-terminal Fgβ tag and a C-terminal YBBR tag (6xHis-SUMO-Fgβ-ddFLN4-MCS-YBBR, illustrated in Figure III-6A). Proteins were deposited on CoA surfaces and initially probed with Biolever Long cantilevers SdrG as illustrated in Figure III-6B. A representative FEC is shown in Figure III-6C. The distinct curvature in the high force event (the rupture of the Fgβ:SdrG interaction) is due to an instrumental artifact caused by the severe cantilever deflection. Because of this, WLC fits were only applied to data <200 pN. When the gold is removed from Biolever Long cantilevers these long, soft cantilevers are stable for hundreds of seconds and are routinely used for AFM-based SMFS experiments.[33,53] Indeed, all the data I presented in Chapter 2 was collected with a similar cantilever. As shown in Chapter 2 Figures 17 and 18, this cantilever was sufficient to measure the unfolding of α_3D, calmodulin, and rubredoxin. However,
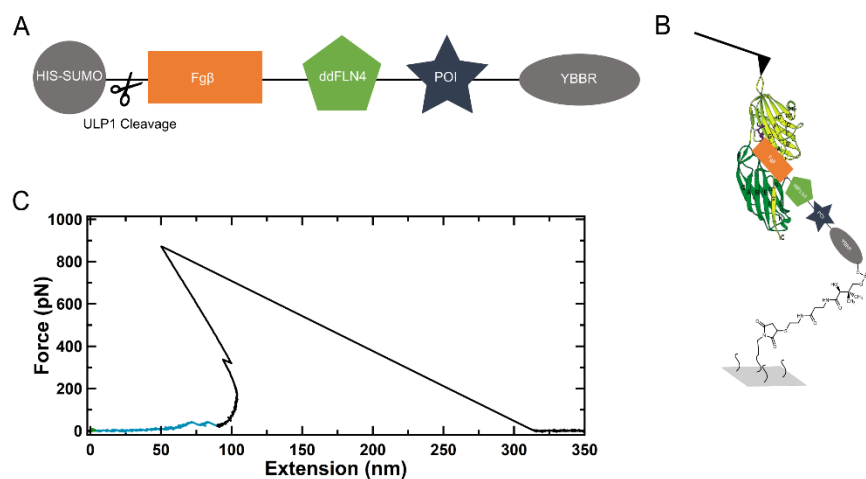
Figure III-6. (A) Schematic of Fgβ construct, 6xHis-SUMO-Fgβ-ddFLN4-MCS-YBBR (B) Schematic of Fgβ polyprotein deposited on a PEG-maleimide-CoA surface, probed with an SdrG functionalized tip. Not to scale. (C) Representative FEC of the Fgβ polyprotein collected with a Biolever Long. The high force event curves back toward the left due to an instrumental artifact. Data collected with 400 nm/s retraction, digitized at 50 kHz and smoothed to 1 kHz.

when I started testing SopE2 and SptP, I found that they both unfold at very low force (<25 pN) (green sections in Figure III-7 A-B) and it was difficult to consistently resolve the effector protein unfolding. While a small number of traces contained discernable peaks corresponding to the effector proteins unfolding, more often the force noise was too high to reliably assign the peaks. This is due to the large hydrodynamic drag experienced by the relatively large Biolever Long cantilever. To increase the data quality, it was necessary to use a modified cantilever. Specifically, I used a Biolever Mini, modified with focused ion beam milling into a "warhammer" geometry.[32]

Specifically, the a Warhammer mini achieves 9-µs temporal resolution with sub-pN force stability over 100 s.[32] This results in a drastic increase in data quality, as seen in Figure III-7 C-D. It's important to note when comparing Figure III-7A to Figure III-7C (SopE2 unfolding) and Figure
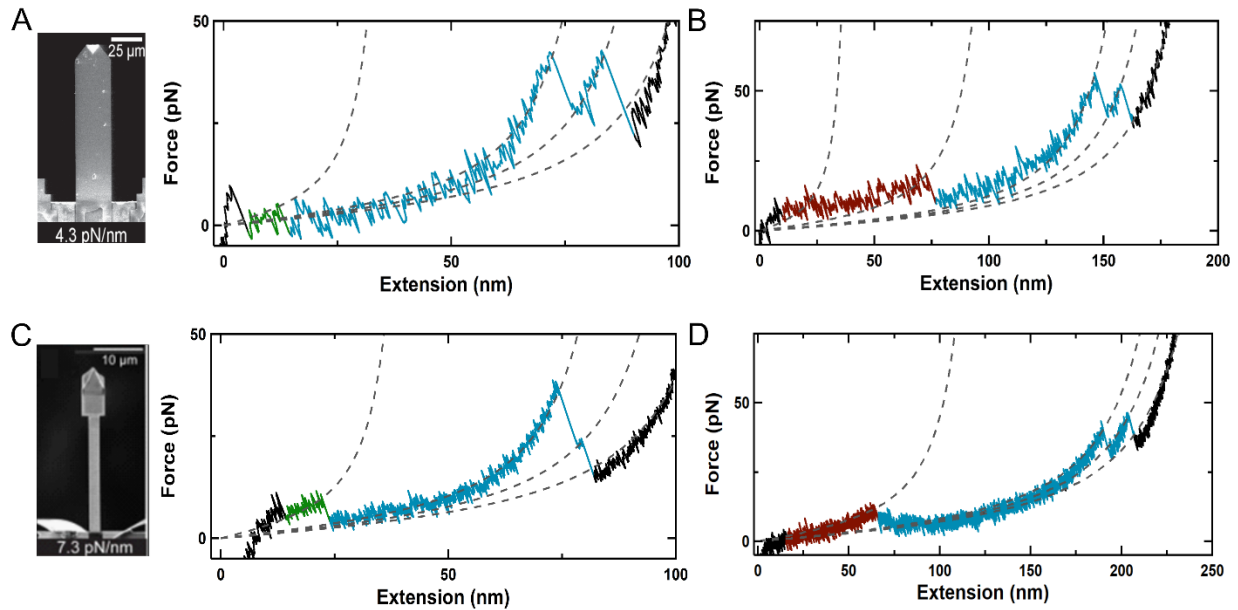
Figure III-7. Comparison of data quality between unmodified and modified cantilevers. (A) SopE2 rupture (green) measured with a Biolever Long. (B) SptP rupture (red) measured with a Biolever Long. (C) SopE2 rupture (green) measured with a "Warhammer" modified cantilever. (D) SptP rupture (red) measured with a "Warhammer" modified cantilever. Grey dashed lines represent the WLC fit. All data collected at 50 kHz and smoothed to 5 kHz for presentation. All measurements collected at 400 nm/s retraction velocity

III-7B to Figure III-7D (SptP unfolding), that the traces were collected at the same retraction velocity (400 nm/s) with the same sampling rate (50 kHz). Therefore, the increased data quality is due purely to the increased performance of the cantilever. The Warhammer mini allowed me to reliably assign the peaks corresponding to SopE2 and SptP unfolding. Importantly, the change in contour length (ΔL_C) measured for both SopE2 and SptP is consistent with the values predicted by subtracting the end-to-end length from the crystal structures from the length of the unfolded polypeptide (ΔL_C SopE2 = 57.4 nm, ΔL_C SptP = 133 nm). Throughout the rest of this chapter, all data was collected with Warhammer modified Biolever Minis that were SdrG functionalized.

Detection of an Unfolding Intermediate in SopE2

The use of modified cantilevers not only made it easier to reliably assign the effector protein unfolding peaks, it also revealed more detail of the unfolding pathway. The improved temporal resolution and reduced force noise provided by the Warhammer allowed me to identify a common unfolding intermediate in the unfolding pathway of SopE2. Figure III-8 shows a representative trace where the intermediate was captured. Passing through this intermediate seems to be the preferred pathway, with ~65% of recorded traces (n = 140) showing this two-step unfolding pathway. The average steps had contour length changes (ΔL_c) of 11.7 ± 3.8 nm for the first step, and 45.3 ± 4.5 nm for the second step. Added together (57 ± 8.3 nm), this represents almost an exact match to the estimated change in contour length(ΔL_c) for the full SopE2 fragment (57.4 nm)- calculated as the total number of amino acids multiplied by the length of an amino acid (0.37 nm) minus the end to end distance in the crystal structure.

In preliminary steered molecular dynamics simulations performed by our collaborators in Dr. Tajkhorshid's lab (University of Illinois), there is a peak in the FEC centered around 12 nm. By analyzing the trajectories and angle between the domains, they suggest that the intermediate may be the result of the two three-helix bundles of SopE2 separating. A simple schematic of this is shown as an inset in Figure III-8.
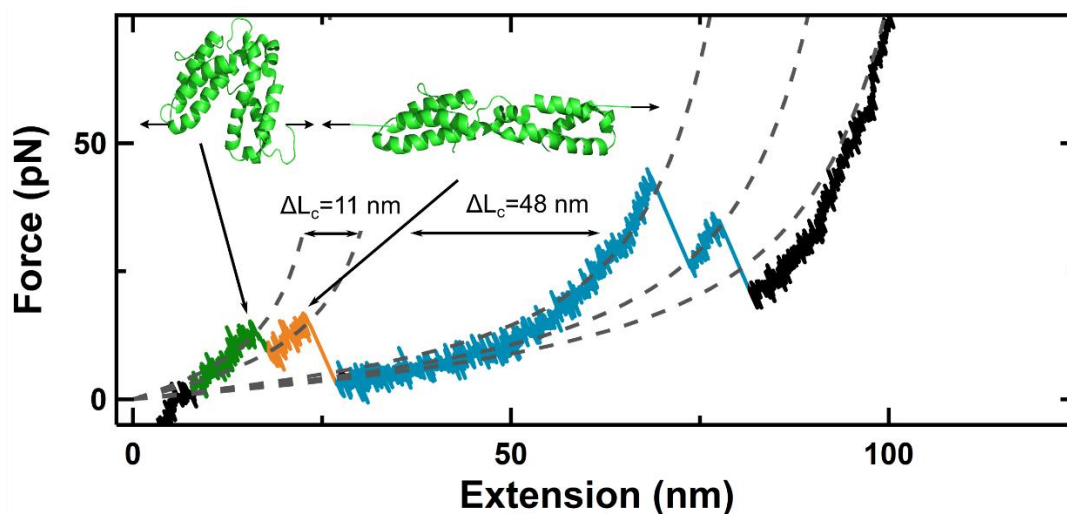
Figure III-8. Representative FEC showing an unfolding intermediate in the SopE2 unfolding pathway. The unfolding intermediate, marked in orange, may correspond to the separation of the three-helix bundles that make up SopE2, as shown in the inset schematic. The change in contour length for the two steps, 11 nm and 48 nm, add together to match the predicted change in contour length for the full protein (57 nm).

SptP Unfolds at Lower Force than SopE2

With the ability to reliably identify the unfolding of SopE2 and SptP, I was ready to compare the amount of force necessary to unfold SopE2 and SptP. As a reminder, SopE2 is secreted ~2x faster *in vivo* than SptP, leading me to expect SopE2 to unfold at a lower force than SptP. To test this, I deposited both proteins on CoA functionalized surfaces at 3 μM and probed

with a streptavidin functionalized tip. Because neither protein had previously been studied by SMFS techniques, I wanted to test many individual proteins to gather the average unfolding force. Therefore, I didn't try probing an individual polyprotein repeatedly, instead relying on raster scanning of the surface to collect my data set.
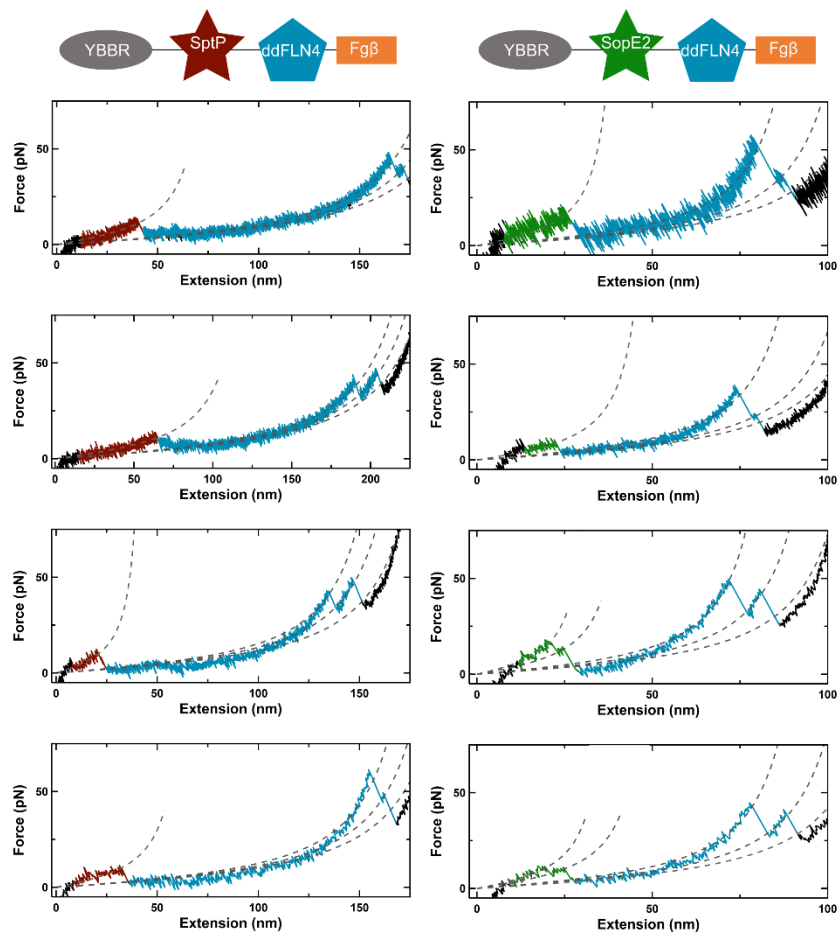
Figure III-9. Representative FECs of SptP and SopE2 at four retraction velocities. Plots show representative traces at the four unfolding velocities (200, 400, 800, and 1600 nm/s from top to bottom respectively). All traces were obtained at 50kHz and smoothed to 5 kHz for presentation. Grey dashed lines are the WLC fits to the data.

Because the amount of force necessary to unfold a protein depends on how quickly the force is applied (discussed in more detail in Chapter 1), I tested SptP and SopE2 unfolding at four different retraction velocities (Figure III-9). For each protein, I collected at least 12 unfolding

events at each speed and averaged the force at which SopE2 or SptP unfolded. If the protein

unfolded in more than one step, I only used the force of the first unfolding event since the first

unfolding event is mandatory. The number of unfolding events is relatively low because I only

used data where the ΔL_C was within 10% of the predicted value. What became immediately

apparent upon analyzing the data was that SptP unfolds at a lower force than SopE2 (32% less on

average) at all tested retraction velocities (Figure III-10). This was a surprising result, not only

because SopE2 is secreted more quickly than SptP, but because SptP has significant β-sheets.
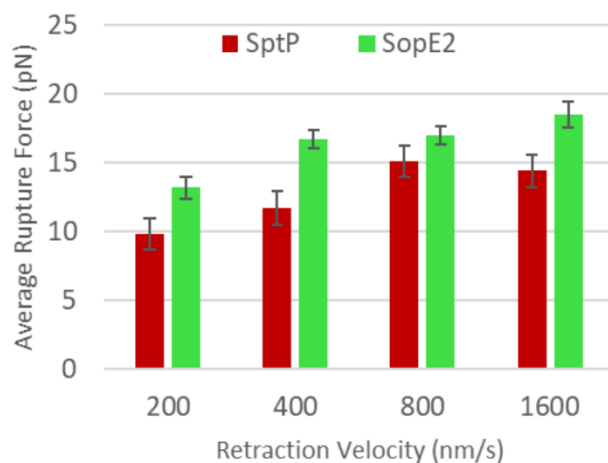
Figure III-10. Comparison of the amount of force necessary to unfold SptP and SopE2 at four retraction velocities. (B) Average rupture force at each retraction velocity. Error bars represent the standard error of the mean rupture force. Average rupture force includes data from at least 20 unfolding events except for SptP at 1600 nm/s (15 events) and SopE2 at 200 nm/s (12 events). Regardless of the retraction velocity, SptP unfolds at lower force than SopE2.

Preliminary Dynamic Force Spectra of SopE2 and SptP

While comparing the average unfolding force of SopE2 and SptP does give us an idea of

their mechanical stability, further analysis of the data can provide a better picture of how force

affects the proteins. I constructed a preliminary dynamic force spectrum with the same data used

to compare the average unfolding force (using at least 12 FECs at each of four pulling speeds). As

shown in Figure III-11, the linear fit to the data is rather poor, particularly for SptP (dashed red

line). This results in large errors in the unfolding rate at zero force and the distance to the transition state as noted in Figure III-11. Nevertheless, while preliminary, the spectra indicate that both effector proteins have large distances to the transition state ($\Delta x^{\ddagger}_{SopE2}$ = 2.1 ± 0.6 nm, $\Delta x^{\ddagger}_{SptP}$ = 1.8 ± 0.5 nm). As a point of comparison, the entirely α-helical α3D, while much smaller, has a $\Delta x^{\ddagger}$ of 1.0 ± 0.1 nm.
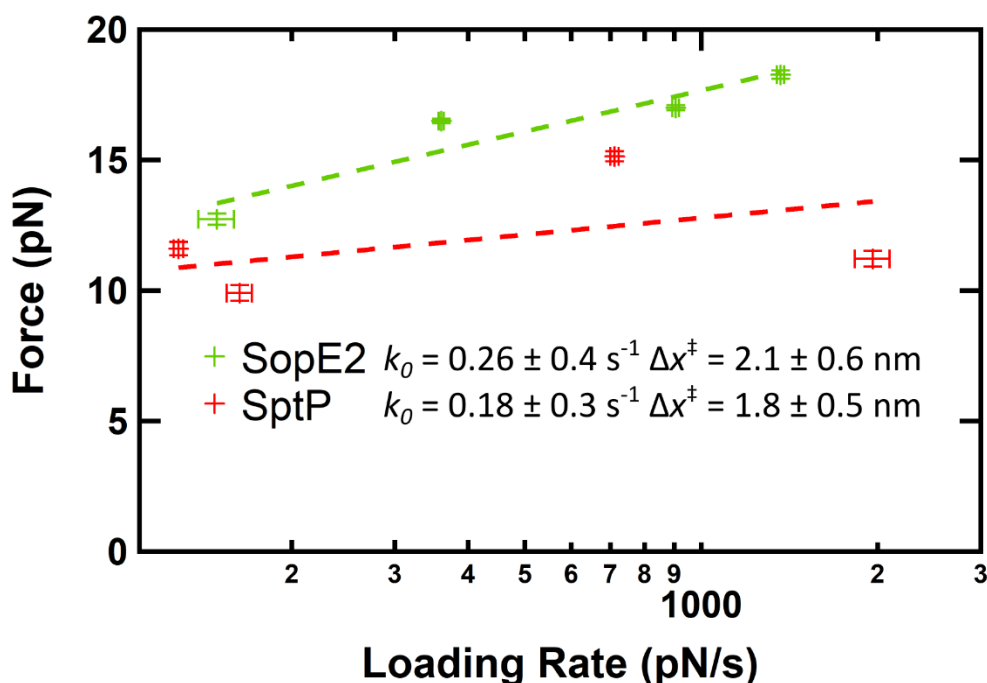
Figure III-11. Preliminary dynamic force spectra of SptP and SopE2. Average rupture force includes data from at least 20 unfolding events except for SptP at 1600 nm/s (15 events) and SopE2 at 200 nm/s (12 events). Error bars represent the standard variance.

Discussion

Genetically Encoded Tags Increase Protein Deposition Efficiency

Because the YBBR tag, Dockerin III, and Fgβ are genetically encoded and require no functionalization before deposition, significantly less protein is needed for deposition than when using DBCO labeled proteins. As discussed in Chapter 2, the low solubility of DBCO reagents

limited labeling to <35%. Moreover, for the homobifunctional dual aldehyde tagged polyprotein, an even lower percentage was labeled with both HIPS-Biotin and HIPS-DBCO. This meant that getting enough protein attached to the surface for efficient AFM experiments often required deposition of 50-100 µM protein. While not always a problem, if a protein precipitated in the HIPS labeling buffer, it could be a challenge to label enough protein. Using the genetically encoded tags, typical deposition concentrations are 0.1-5 µM, an improvement due to every protein containing the tag. Reducing the deposition concentration to 0.1 µM was actually essential to minimize multiple attachments while maintaining a high rate of protein attachment when using the YBBR-Dockerin III polyprotein. Because FGE conversion and derivatization were not necessary, significantly smaller cultures could be used to express the protein. This simplifies the process of purifying multiple proteins for AFM testing.

Using PEG-Methoxy to Reduce Multiple Attachments

Before high-quality SMFS data can be obtained using site-specific attachment to the AFM surface and tip, the protein deposition concentration must be optimized to prevent the tip from attaching to more than one polyprotein at a time (multiple attachment) while maintaining a high attachment efficiency. This process became more problematic with the increased deposition efficiency provided by genetically encoded tags, with small variation in protein concentration often resulting in too many or too few connections. To avoid this optimization process, I am investigating the effect of diluting the silane-PEG-maleimide used for surface deposition. By diluting silane-PEG-maleimide into silane-PEG-methoxy, the number of sites able to covalently bind protein to the surface are decreased and ideally spread out across the surface. This should enable any concentration of protein to be deposited without risking multiple attachments and

would greatly simplify the process of optimizing protein deposition conditions. Initial tests have seen a reduction in multiple attachments, but an increase in non-specific adhesion to the surface. Variation of the PEG-methoxy size is likely needed to create an optimal non-stick surface with sparse functionalization.

Stringent Filtering Resulted in Low Number of Data Points

Because SopE2 and SptP have never been studied by any SMFS technique, I wanted to ensure the data I analyzed was the result of SopE2 and SptP unfolding. Therefore, I only included data where ΔL_C was within 10% of the predicted value. This is a very stringent requirement that resulted in most of my collected data being ignored for this analysis. A major issue when analyzing unfolding events that occur at low force near the surface is that it provides a very small number of data points with which to fit the WLC. Because the WLC is an exponential fit (specifically $\frac{1}{\left(1-\frac{x}{L}\right)^2}$), small variations at the beginning of the fit can result in large changes in the resulting contour length. While the data collection process is now mostly automated, achieving significantly larger data sets may require broadening the range of accepted ΔL_C or optimizing the process for fitting WLCs to low force unfolding events.

SptP Unfolds at Low Force Despite β-Sheets

My results showed that at every tested retraction velocity, SptP unfolded at lower force than SopE2. This was surprising because of the β-sheets within SptP, which are traditionally associated with higher unfolding force. The most likely reason for why the β-sheets don't impart more mechanical stability to SptP is that the force is applied to the protein in a way that does not require the β-strands to shear. Other studies have shown β-sheets are only mechanically robust

when forced to shear across each other. "Peeling" one strand away from another requires significantly less force.[122–125] Changing the β-sheets so that applied force results in peeling instead of shearing may have been an evolutionary adaptation to facilitate secretion of effector proteins. This idea will be explored further in Chapter 4, where I discuss the entirely unique Ψ-loop motif of NleC which contains an offset inner strand that may allow the β-sheet to peel apart.

Unfolding Force Does Not Correlate with the Secretion Kinetics of SopE2 and SptP

While the amount of data remains relatively low, it is striking that at all tested velocities SopE2 unfolds at a higher force than SptP, averaging 32% higher unfolding force across all retraction velocities. While these results may not agree with our hypothesis that the mechanical stability of an effector correlates with its *in vivo* secretion rate, it's important to remember that the *in vivo* secretion rates of SopE2 and SptP only differ by ~2 fold. It is possible that in this instance, the mechanical stability is close enough that other factors account for the difference in secretion rate. Indeed, SptP is ~2x longer than SopE2, which may account for the slower secretion rate. However, it is striking that both SopE2 and SptP unfold at forces <20 pN at all tested retraction velocities, reinforcing the idea that effector proteins have evolved to unfold at very low force. These proteins represent two of the most mechanically labile proteins studied to date by AFM, despite their drastically different structures and functions. This lends support to the idea of mechanical stability as a threshold, with effector proteins needing to unfold below a set force. Once that low force threshold is met, other factors may become rate limiting. This may have allowed bacteria to evolve more advanced systems to modulate the hierarchy of effector protein secretion, a process that is essential for effective pathogenesis.

Distance to the Transition State May Correlate with Secretion Kinetics

As discussed in Chapter 1, plotting the most likely unfolding force against the log of the loading rate (a dynamic force spectrum) results in a linear relationship. When fit with the Bell-Evans model,[37] we can extract the distance to the transition state ($\Delta x^{\ddagger}$) and the unfolding rate at zero force (k_o). $\Delta x^{\ddagger}$ is a measure of protein compliance, with a larger $\Delta x^{\ddagger}$ corresponding to a softer, more flexible protein. This means the force can be applied gradually to the protein, and may make it easier for weak unfoldases, such as the ATPase associated with the T3SS[2] to unfold a protein. My results indicate that SptP has a shorter distance to the transition state than SopE2 (1.8 nm and 2.1 nm respectively).

This suggests that softer, more compliant proteins can be secreted more quickly, even if the amount of force needed to unfold them is higher. This idea is consistent with a model proposed by Bell[126] that postulates that the rate of unfolding under force ($k(F)$) varies exponentially with force as described in the equation:

$$k(F) = k_0 e^{-\frac{\Delta G^{\ddagger} - F\Delta x^{\ddagger}}{k_B T}}$$

where $\Delta G_0^{\ddagger}$ and the k_0 are the barrier height and unfolding rate at zero force respectively (k_B and T are the Boltzmann constant and absolute temperature respectively). As $\Delta G_0^{\ddagger}$ increases, higher forces are necessary to maintain a constant rate. But if proteins have similar $\Delta G_0^{\ddagger}$ and k_0, the one with the larger $\Delta x^{\ddagger}$ will require less force to unfold at the same rate. This is illustrated in Figure III-12.
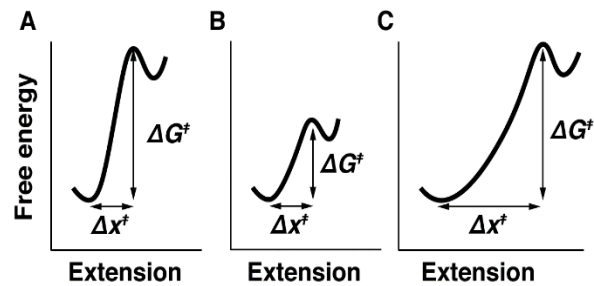
Figure III-12. 1D-energy landscapes projected on the extension coordinate. Compared to A, B and C would both require less force to unfold at the same rate. For B, this is due to a lower $\Delta G_0^{\ddagger}$. For C, the difference is due to the larger distance to the transition state.

This model is of particular interest when considering that the ATPase associated with the T3SS is a weak unfoldase. If the ATPase can only apply a low force, but can sustain that force over time, then more compliant effectors (larger $\Delta x^{\ddagger}$) will be secreted more quickly. This assumes that effector proteins have similar $\Delta G_0^{\ddagger}$ and k_0. It seems unlikely that this would be the only factor that determines the order of effector protein secretion, but it may play a key role in the process. Further studies on both the mechanical stability and *in vivo* secretion kinetics will be necessary to elucidate if $\Delta x^{\ddagger}$ can serve as a predictor for secretion kinetics.

Methods

Preparation of CoA Functionalized Surfaces and Tips

Specifically, after rinsing PEG-Maleimide surfaces in toluene, isopropanol, and ultrapure water, surfaces were dried and immediately covered with 1 mM CoA in CoA reaction buffer (25 mM phosphate, pH 7.0, 150 mM NaCl, 4 mM TCEP). While tips were then stored in a wafer holder in a humidity chamber, surfaces were stored wet in a petri dish contained within a simple humidity chamber for up to a month at 4°C. When surfaces were needed, the CoA solution was rinsed off by immersion in 2 100 mL beakers of ultrapure water, dried, and mounted to a metal puck. 30 μL of 1-5 μM YBBR tagged protein was added to surface in the presence of 25 mM $MgCl_2$ and 3 μM

Sfp. Whereas maleimide functionalized surfaces were unappealing because they had to be used immediately to avoid hydrolysis, CoA is a stable, allowing CoA functionalized surfaces to be stored for up to a month at 4°C without significant reduction in activity. This means that large batches of surfaces can be prepared at once and stored until needed, making them comparable to the azide functionalized surfaces used in Chapter 2. The Gaub lab has reported that PEG-CoA surfaces can be dried and stored under argon for up to a month, but that hasn't been tested in our lab.

Cohesin III and Dockerin III Plasmids

The Gaub lab provided us with plasmids for both 6xHis-Cohesin III-YBBR (pMS 1481) and 6xHis-YBBR-ddFLN4-Dockerin III-Xmod (pMS 1480). The Dockerin construct included the immunoglobulin-like domain 4 (ddFLN4) from *Dictyostelium discoideum* as a marker protein. Because ddFLN4 unfolds in two steps (ΔL_c = 15.1 ± 0.1 nm and 16.9 ± 0.1 nm),[127] it serves as a good marker protein without needing multiple copies.

Cloning, Expression, and Purification of α₃D in Dockerin III Construct

α_3D was amplified with primers U/L 2415 and pMS 1480 was opened with primers U/L 2414, resulting in 15 bp overlaps on each end. The PCR products were digested with DpnI before undergoing Gibson Assembly. The resulting construct (pMS 1561) expressed well and was purified using the same protocol described in Chapter 2 for the aldehyde constructs. The only change was that after elution the protein was simply dialyzed into HEPES buffer (25 mM HEPES, pH 7.2, 150 mM NaCL) overnight at 25°C and concentrated in an Amicon 10 kDa centrifugal concentrator before being delivered to the Perkins lab for testing. 30 µL of 2 µM protein in HEPES buffer was deposited onto a CoA functionalized surface in the presence of 25 mM $MgCl_2$ and 3

µM Sfp for 30 minutes, rinsed 10x with 1 mL HEPES buffer, and probed with a Cohesin functionalized tip. Tip functionalization was performed the same way as the surface, simply using YBBR-Cohesin in place of the Dockerin protein and immersing a CoA functionalized tip in the protein solution.

Incorporating a Multiple Cloning Site into Fgβ Constructs

The Gaub lab provided us with plasmids for SdrG(domains N2 and N3)-6xHis-YBBR (pMS 1505) and Fgβ-ddFLN4-6xHis-YBBR (pMS 1504). It's important to remember that the Fgβ peptide must be at the N-terminus of the polyprotein to achieve high force ruptures. If inverted, the rupture force is <60 pN.[122] While this plasmid was designed use Gibson Assembly for the insertion of a protein of interest, I wanted the option of using standard restriction enzyme cloning. Therefore, I introduced a multiple cloning site (MCS) behind ddFLN4. At the same time, I moved the system into pMS 984, adding the SUMO tag to the polyprotein. This was done by amplifying Fgβ-ddFLN4 out of pMS 1504 with primers U/L 2379 which added a glycine at the N-terminus (necessary for SUMO cleavage by Ulp1) and a MCS to the C-terminus, ending in an NcoI site. The YBBR tag was added back in after the MCS using annealed primers (U/L 2380) that formed compatible sticky ends for NcoI and XhoI. This annealed fragment was ligated to the Fgβ-ddFLN4 PCR product, and inserted into pMS 984 opened with SfoI and XhoI. The resulting construct was 6xHis-SUMO-Fgβ-ddFLN4-MCS-YBBR (pMS 1554), with an identical MCS as our aldehyde constructs. This allowed me to quickly move a protein of interest between the polyprotein constructs without needing to prepare a new insert.

Cloning, Expression and Purification SopE2 and SptP into Fgβ Construct

Effector proteins of the T3SS contain a signal sequence at N-terminus, typically followed by an unstructured region that can bind chaperones, although not all effector proteins require chaperones to be efficiently secreted.[6] In structural studies of effectors, the N-terminal sequence can rarely be studied unless it is in complex with a chaperone. Therefore, the available structures typically only include the active portions of the protein.[101,102,128] Because I was attempting to characterize the unfolding of SopE2 and SptP for the first time, I wanted the best chance of being able to map unfolding events back to the protein structure. Therefore, I focused my initial efforts on only the portions of SopE2 and SptP that with known structures. For SopE2 that includes residues 69-240 (PDB 1R9K)[101] while SptP includes residues 161-543 (PDB 1G4U)[102] as shown in Figure III-2.

The selected fragments of SopE2 and SptP were amplified out of the tetracysteine tagged constructs originally used by the Palmer lab for measurement of *in vivo* secretion rates[97] using primers U2385/L2386 and U2387/L2388 respectively. In both cases, the primers introduce a C-terminal NcoI site so that the fragments could be inserted into pMS 1554 opened EcoRV/NcoI. The resulting plasmids, pMS 1569 (SptP) and 1570 (SopE2) were expressed in *E. Coli* BL-21 (DE3) cells induced with 0.4 mM IPTG and incubated overnight at 18°C after induction. Protein purification was done exactly as described in Chapter 2 for the aldehyde tagged construct. The SopE2 containing plasmid (pMS 1570) expressed at much higher levels than the SptP containing plasmid (pMS 1569) as shown in Figure III-6. However, after purification, there was still more than enough SptP for AFM experiments, aided by the fact that after Ulp1 digestion no further functionalization was necessary. For both polyproteins, the elution from the Ni-NTA column was

diluted in half in TBS + TCEP (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP) and 100 uL of 0.5 mg/mL Ulp1 was added to cleave the SUMO Tag. The solution was dialyzed overnight into HEPES Buffer (25 mM HEPES, pH 7.2, 150 mM NaCl, 5 mM BME) at 25°C, concentrated in an Amicon centrifugal concentrator (10 kDa), and either immediately used for AFM experiments or flash frozen and stored at -80°C.

Cloning and Expression of SptP and SopE2 for Thermal Denaturation Studies

The gene fragments of SptP and SopE2 (the portions with known structures), were amplified using primers U2419/L2419 (SopE2) and U2420/L2420 (SptP) which incorporated an N-terminal glycine (necessary for SUMO cleavage by ULP1) and a C-terminal stop codon. These were inserted into pMS 984 (6xHis-SUMO) opened SfoI/XhoI. Protein expression was carried in 1L autoinduction media (LB + 0.6% v/v glycerol, 0.05% w/v glucose, 0.2% w/v lactose),[129] inoculated with 1 colony forming unit (CFU) of *E. Coli BL*-21 (DE3) transformed with pMS 1597 (SptP) or pMS 1599 (SopE2) and grown overnight at 37°C. Purification was done identically to the procedure for the polyprotein constructs. SptP expresses at a much lower level than SopE2, yielding a lower amount of purified protein. Eluted protein was dialyzed into TBS (25 mM Tris, pH 8.0, 150 mM NaCl) in the presence of ULP1 to cleave the SUMO tag. After dialysis, the protein was run over a 1 mL Ni-NTA column to remove the SUMO tag and ULP1 and purified to homogeneity with an S200 (SptP) or S75 analytical column equilibrated with 25 mM Tris pH 8.0, 1 mM DTT. Salt was omitted to allow for collection of CD down to 190 nm.

Chapter IV Comparing Effector Proteins to Non-Secreted Homologues

While comparing effector proteins to each other sheds light on the parameters that effect secretion rates, comparing effector proteins to their non-secreted homologues can help elucidate what evolutionary modifications were necessary to allow proteins with such a broad array of structures and functions to pass through the T3SS. Early research into effector proteins of the T3SS found that these proteins show very low sequence similarity to other members of their protein superfamilies.[15] This divergence often makes it impossible to identify functional homologues of effector proteins by sequence alignment, often requiring structure determination to identify homologues. [15] This may not be particularly surprising given the need for a signal sequence for secretion and that many also contain an unstructured chaperone binding domain at the N-terminus.[4,18] Yet, even when examining the domains of effector proteins responsible for their pathogenic activity, there continues to be low sequence similarity. Because of the low sequence similarity, it is often necessary to solve the structure of effector proteins to identify their homologues.[128,130,131] The sequence divergence between effector proteins and their non-secreted homologues suggests that efficient secretion through the T3SS imposes unique constraints upon effectors that has driven its sequence divergence.

Comparing the structures of effector proteins and non-secreted homologs often reveals additional signs of divergence. This is because effector proteins also have low structural similarity with members of their superfamilies, with only the most essential domains being conserved.[128,131] An excellent example of this is NleC, a Zincin which has low structural similarity with its protein superfamily. The closes structural relatives are aminopeptidase tricorn (RMSD = 3.3Å) and botulinum toxin (RMSD = 3.3Å), as determined using the Dali server.[132] Interestingly,

NleC does not contain the five-stranded β-sheet present in canonical Zincins (Figure IV-1A).[133] Moreover, the three stranded β-sheet present in NleC is made of a unique modified Ψ-loop motif[128]. Whereas the standard Ψ-loop β-sheet is fully stacked, with the outer two β-strands only interacting with the middle strand, in NleC the Ψ-loop zippers the outer two strands together while the middle strand exits the sheet half way through (Figure IV-1B).[128] Since, in general, shearing β-strands requires high force,[77] this modification may allow NleC to unfold at lower force, increasing the secretion rate *in vivo*. We hypothesize that the unique structural aspects of NleC evolved to facilitate its secretion through the T3SS by reducing the amount of force necessary to unfold. The need for secretion and the ability for the protein to be unfolded by the weak unfoldase associated with the T3SS was likely the driving force behind NleC's divergent sequence. To elucidate how structural modifications may have changed effector proteins response to force, it is necessary to compare the unfolding force spectra of effector proteins to their non-secreted homologues.

Botulinum neurotoxin   NleC
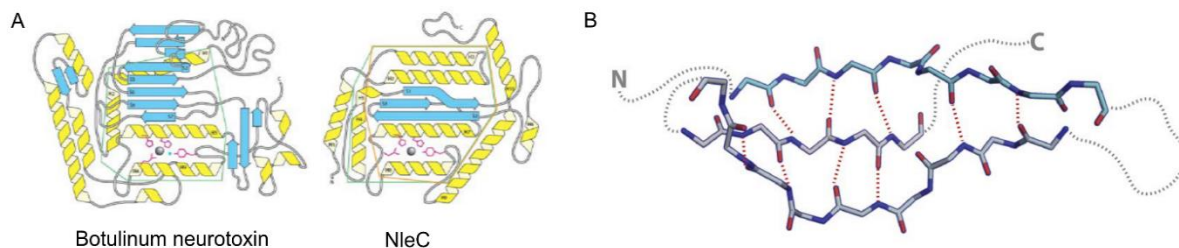
Figure IV-1. NleC contains a unique Ψ-loop motif. (A) Topography diagrams of Botulinum neurotoxin and NleC. Botulinum neurotoxin is one of the closes structural homologues to NleC, with a Z-score of 5.7. (B) Ball and stick model of the modified Ψ-loop motif of NleC highlighting the hydrogen bonds. Note the interaction of the outer two strands which may facilitate unfolding at low force. Adapted from Turco and Sousa, *Biochemistry* (2014).

When considering the reasons for the divergence of effector proteins from members of their protein superfamilies, it is important to remember the steps leading to their secretion. Effector proteins are expressed in the bacterial cytoplasm, where most are bound by chaperones

that maintain the N-terminus in an unfolded state. Interestingly, chaperone binding does not prevent the catalytic domains of effector proteins from folding. For effector proteins to pass through the ~2 nm needle of the T3SS, they must be unfolded by the associated ATPase.[4,15,134] Once secreted into host cells, effector proteins must refold to carry out their various activities, contributing to bacterial pathogenesis.[135] Protein secretion through the T3SS occurs quickly, with a full complement of effector proteins being secreted within minutes of host cell contact and individual proteins being secreted in less than a second.[97,136] Because the T3SS is essential for pathogenicity in many bacteria, there was likely a strong evolutionary advantage to mutations that facilitated rapid unfolding under force. As discussed in Chapter 3, the rate a protein unfolds under force directly rates to the free energy of unfolding, the off rate at zero force, and the distance to the transition state. It is possible that evolutionary adaptation of effector proteins resulted in modification to all three of these characteristics, or simply optimized one of them. The goal of this project is to elucidate how the evolutionary divergence of effector protein sequence and structure has affected their mechanical stability. This may not only provide a better understanding of the T3SS, but also provide general insight into how protein modifications affect their stability under force.

We chose three sets of effector-homologue pairs to characterize with AFM-based SMFS: NleC/Protealysin, SpvB/ART2.2, and YopH/PTP1B. Comparing the crystal structures of the effector proteins to their non-secreted homologues (Figure IV-2) showed that in each case, the fold was retained despite divergent sequences. This facilitates the ability to directly compare the mechanical stability of the effector proteins to their homologues. The three sets of protein carry out diverse functions and provide an array of different folds and sizes. SpvB and ART2.2 are ADP-

ribosyltransferases (ATR) and are the smallest pair of proteins in our chosen set (199 and 226 amino acids respectively).[137,138] The SpvB structure[139] contains a canonical α/β fold and NAD$^+$ binding site, despite sharing only 15-20% sequence identity with ATRs not associated with the T3SS. YopH and PTP1B are protein phosphatases, containing canonical phosphatase folds.[140–142]
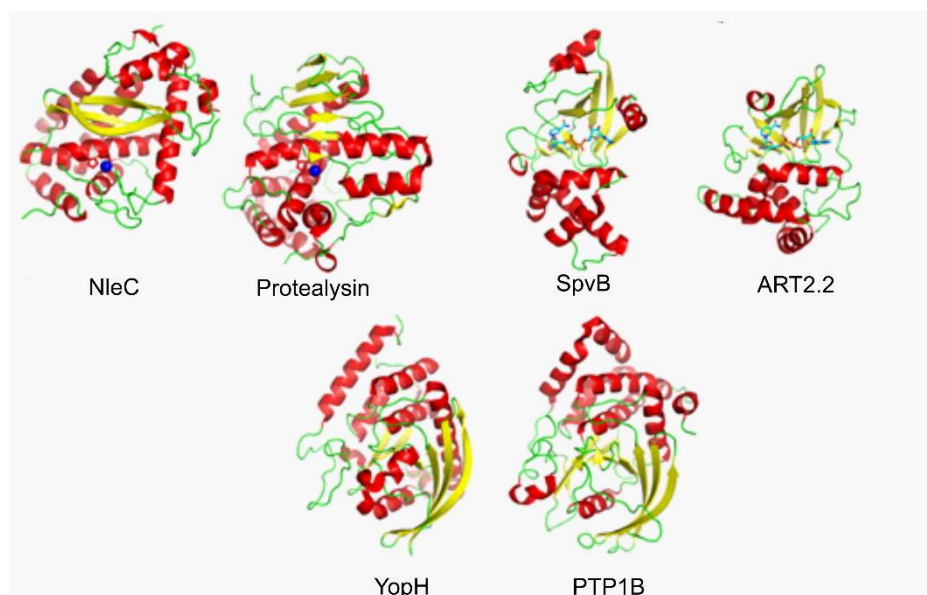
Figure IV-2. Structures of TTSS effectors and non-secreted homologs. Zincins NleC and Protealysin, blue spheres are Zn2+ ions. ADP-ribosyltransferases SpvB and ART2.2, NAD ligand shown in cyan. Protein phosphatases YopH and PTP1B.

This is despite YopH only sharing ~15% sequence identity with non-effector phosphatases.

NleC and Protealysin are both Zincin proteases, a protein superfamily typically associated with high thermodynamic stability.[128,130,143–146] NleC is the most divergent effector being tested, showing no significant sequence similarity to any known protein except for the zinc-binding motif (HExxH). It was not until the structure of NleC was determined that it could be assigned as a Zincin. Because both proteins chelate Zn^{2+} ions, they provide an interesting test case for investigating how ion retention effects mechanical stability. Due to the traditionally high thermodynamic stability of Zincins, characterization of this pair was prioritized. If our hypothesis

is correct, and the divergence of effector proteins enables them to unfold at lower force, this pair is the most likely to have significant differences. For this reason, the data presented in this chapter were collected with NleC and Protealysin. However, I have collected preliminary data for SpvB/ART2.2 and YopH/PTP1B, presented in Appendix C. The data presented here utilizes both the Dual-Aldehyde construct and the YBBR-Fgβ construct.

Results

NleC Remains Active After HIPS Labeling

One of the main concerns when using the Dual-Aldehyde construct was how HIPS labeling would affect the protein of interest. Because a previous member of the Sousa Lab (Dr. Michelle Turco) had developed an activity assay for NleC, it served as an excellent platform to test whether protein could remain active within the Dual-Aldehyde construct after HIPS labeling. To qualitatively test the activity of NleC after HIPS conversion I labeled a Dual-Aldehyde construct containing the catalytic fragment of NleC with HIPS-Biotin and HIPS-DBCO. After labeling, the polyprotein was incubated with freshly purified RelA and RelB, components of NF-κB cleaved by NleC *in vivo*, at room temperature for 30 minutes. As seen in Figure IV-3 both RelA and RelB were efficiently cleaved by the NleC containing polyprotein after HIPS-labeling. This gave me confidence that the HIPS-labeling protocol described in Chapter 2 could be used without deactivating the protein of interest inserted into the construct. While this is not a quantitative evaluation of the effect the polyprotein and HIPS labeling have on NleC, it quantitatively shows that NleC retains activity. This indicates that its ability to fold into a native conformation is not dramatically affected by insertion into the polyprotein and subsequent functionalization.
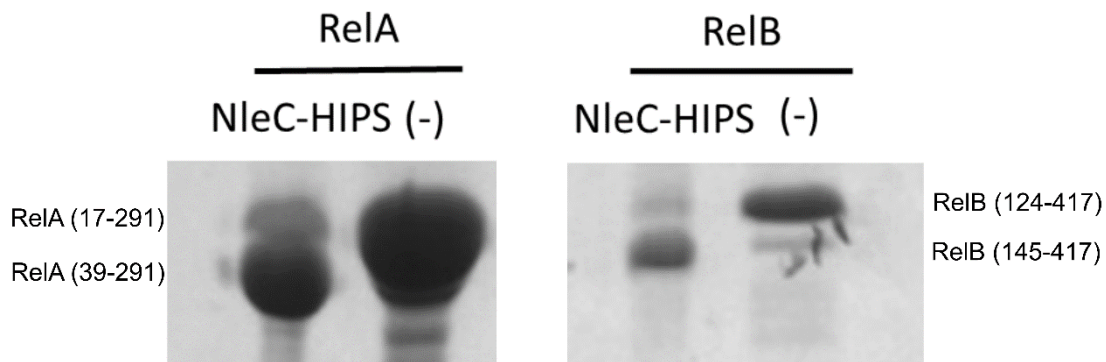
Figure IV-3. Structures of TTSS effectors and non-secreted homologs. Zincins NleC and Protealysin, blue spheres are Zn2+ ions. ADP-ribosyltransferases SpvB and ART2.2, NAD ligand shown in cyan. Protein phosphatases YopH and PTP1B.

NleC Requires Modified Cantilevers to Reliably Assign Low Force Unfolding

With the goal of mechanically characterizing NleC, I used protocols developed in Chapter 2 to incorporate the catalytic core of NleC (fragment 22-279) into the Dual-Aldehyde construct. The resulting polyprotein was labeled with HIPS-Biotin and HIPS-DBCO (Figure IV-4A) and deposited on an azide functionalized surface and probed with a biotin functionalized tip. Initial experiments were performed with a Biolever Long (Olympus) cantilever with the gold removed. However, as was the case with SopE2 and SptP (Chapter 3), NleC unfolds at very low force (<15 pN) (Figure IV-5B). This made it difficult to assign the unfolding event associated with NleC because it was often lost in the force noise. Therefore, it was necessary to once again use a Warhammer Biolever Mini to improve the data quality (Figure IV-5C). The drastic improvement in data quality is made more apparent when looking just at the unfolding of NleC (Figure IV-5 D-E). The improved quality made it possible to reliably assign unfolding peaks to NleC.
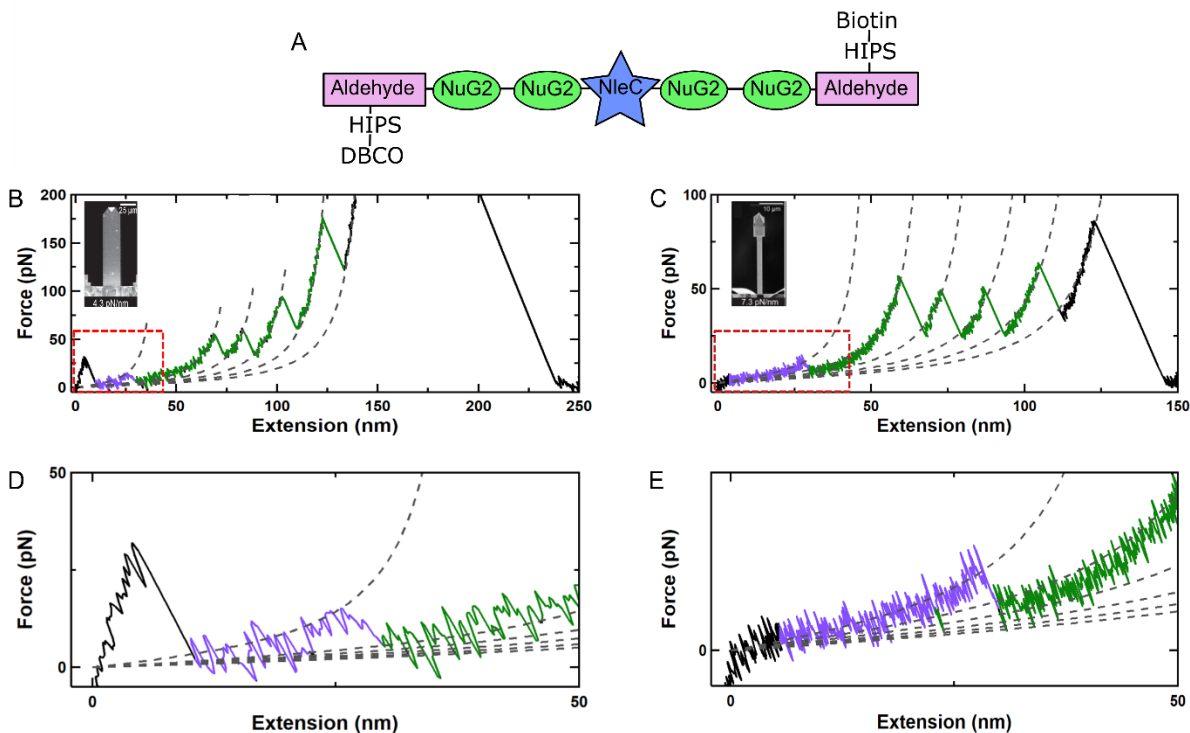
Figure IV-4. Comparison of force extension curves of NleC using Biolever Long and Warhammer cantilevers. (A) Schematic of NleC in the Dual-Aldehyde construct, labeled with HIPS-Biotin and HIPS-DBCO. (B) Representative FEC of NleC in Dual-Aldehyde construct collected with a Biolever Long (inset). Grey dashed lines are the WLC fit. (C) Representative FEC of NleC in Dual-Aldehyde construct collected with a Warhammer Biolever Mini (inset). (D) Expansion of the area highlighted by a red box in B. (E) Expansion of the area highlighted by a red box in C. All FEC collected at a retraction velocity of 400 nm/s, digitized at 50 kHz and smoothed to 1 kHz for presentation.

The Initial NleC Unfolding Event is Below the AFM Detection Limit

In 98 FECs recorded for NleC, the majority (78%) display a single unfolding peak corresponding to a change in contour length 24 ± 3 nm (Figure IV-4 B,C). This is much shorter than the expected ΔL_c of 87 nm for the full unfolding of NleC, suggesting that the protein unfolds in at least 2 steps with the first unfolding step occurring below the detection limit of the AFM instrumentation. Of those FECs containing unfolding events that resulted in ΔL_c greater than 25 nm, no traces contained ΔL_c within 20% of the expected 86 nm, suggesting that there may be more than 2 unfolding steps, but regardless of the number of steps, the initial unfolding step

occurs below the detection limit of the AFM instrumentation. Importantly, all data was collected in a buffer containing 20 µM Zn^{2+} to ensure the NleC contains a chelated zinc ion.

To rule out the possibility that the NleC was partially unfolded during transportation or storage, I tested that the protein could cleave RelA and RelB directly before deposition onto the surface. The results of this test were in qualitative agreement with the tests I performed on freshly purified protein, indicating that the deposited protein did not contain truncated NleC and that the protein was not completely inactivated. To ensure that the 24 nm unfolding event was not caused by the AFM tip picking up multiple proteins at once or an artifact of the polyprotein construct, I analyzed the characteristic fingerprint of the 4 NuG2 marker proteins unfolding. In every FEC with a 24 nm unfolding event, all four marker proteins were observed, with both the ΔL_c (18 nm) and unfolding force (~50 pN) agreeing with literature values.[26,61] If the tip had attached to multiple polyproteins, the resulting FEC would either contain more NuG2 unfolding events (serial unfolding), or the unfolding force would be higher (parallel unfolding). Finally, analysis of the FECs before the unfolding of the first NuG2 domain reveals a contour length of approximately 75 nm, consistent with complete unfolding of NleC. Taken together, these results suggest that the unfolding peak with a contour length change of approximately 24 nm corresponds to an NleC unfolding intermediate, while the initial unfolding event is below the detection limit of the AFM.

Preliminary Dynamic Force Spectrum of NleC Indicates Effector Compliance

With the acquisition of an NleC unfolding intermediate at various retraction velocities, I constructed a preliminary dynamic force spectrum. As with the preliminary dynamic force spectrum of SopE2 and SptP (Chapter 3) it should be noted that the standard in the field is to have the most likely rupture force at five different retraction velocities. I plotted the average the rupture force at three different retraction velocities (200, 400, and 800 nm/s). With that caveat in mind, the preliminary dynamic force spectrum for the NleC intermediate (Figure IV-5) indicates that it has a similarly long distance to the transition state as SopE2 and SptP ($\Delta x^{\ddagger}_{NleC}$ = 1.3 ± 0.17 nm, $\Delta x^{\ddagger}_{SopE2}$ = 2.1 ± 0.6 nm, $\Delta x^{\ddagger}_{SptP}$ = 1.8 ± 0.5 nm). It is important to remember that this is likely the unfolding intermediate, and therefore is likely less stable that the initial unfolding event. However, since this is the only force peak detectable on our platform, this intermediate likely corresponds to the most force stable step in the unfolding pathway. Therefore, the long distant to the transition state of the NLeC intermediate is further evidence that effector proteins have evolved to have long distances to the transition state to facilitate faster secretion. Interestingly, the unfolding rate at zero force (k_0) is 2.0 ± 0.8 s^{-1}. This indicates that the most force stable
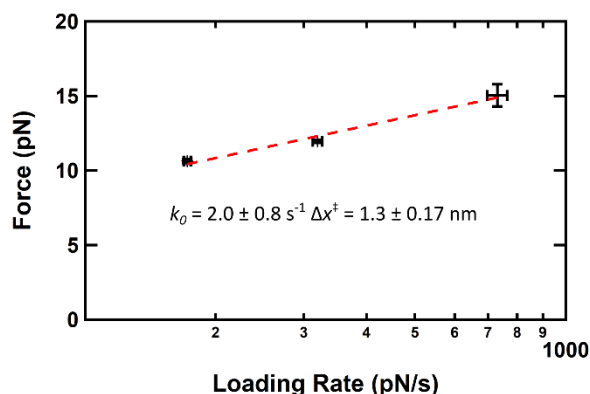
Figure IV-5. Preliminary dynamic force spectrum of NleC. Error bars represent the standard variation. Data was collected at 200, 400, and 800 nm/s retraction velocity, with N = 47, 23, and 8 respectively.

intermediate in the unfolding pathway of NleC is still quite unstable, even when compared to other effector proteins ($k_{0\ SopE2}$ = 0.26 ± 0.4 s^{-1}, $k_{0\ SptP}$ = 0.18 ± 0.3 s^{-1}). While not surprising considering that this unfolding event is likely an intermediate, it does support the idea that NleC has evolved to unfold at low force, facilitating its secretion through the T3SS. More data will help refine the distance to the transition state and unfolding rate at zero force, but is unlikely to result in significant changes to the values presented here.

Protealysin Contains At least One Unfolding Intermediate and Unfolds at Moderate Force

To compare the amount of force necessary to unfold NleC to a non-secreted homologue, I inserted an inactivated mutant of protealysin into the Fgβ-YBBR construct (Figure IV-6). An inactivated protealysin was used to avoid proteolysis during protein purification. Importantly, the mutation did not affect the ability of the protein to chelate a Zn^{2+} ion. Using the Fgβ-YBBR construct allowed me to quickly test the protein without the need for *in vitro* aldehyde conversion and subsequent derivatization. It also provided the added benefit of a much stronger connection to the tip, since the Fgβ:SdrG interaction ruptures at >800 pN. This was advantageous because I anticipated that protealysin would unfold at much higher force than NleC.

The protealysin polyprotein was deposited on a CoA functionalized surface and probed with an SdrG functionalized cantilever. The resulting FECs contained unfolding events that correspond to the complete unfolding of protealysin. Strikingly, protealysin unfolds in at least two steps and both can be detected using our AFM platform (Figure IV-6). When the ΔL_c of the two events are added together the sum (107 nm) is within 10% of the predicted value for the complete unfolding of protealysin (118 nm). 68% of collected FECs (N = 44) show multiple steps in the unfolding of protealysin. This is particularly interesting given that NleC also seems to pass

through an intermediate along its unfolding path. Analysis of the rupture force of protealysin shows that while it does unfold at higher force than NleC, the difference is not as drastic as we had predicted. At 800 nm/s the average unfolding force for protealysin is 22 ± 9 pN (N = 8), compared to 15 ± 6 pN (N = 8) for NleC. Much more data is needed to identify whether this initial difference is statistically significant.
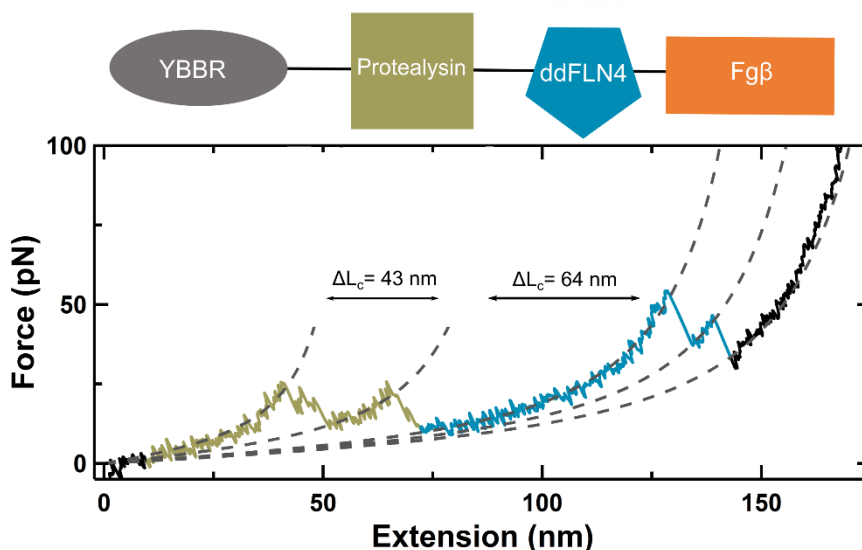
Figure IV-6. Representative FEC of protealysin in the YBBR-Fgβ construct (schematic). Protealysin unfolds in two steps, with ΔL_c = 43 nm for the first event and 64 nm for the second event. When added together (107 nm) they agree with the predicted ΔL_c for the complete unfolding of protealysin (118 nm). Data collected with 1600 nm/s retraction velocity at 50 kHz and smoothed to 5 kHz for presentation.

Discussion

NleC May be Partially Unfolded Before Force is Applied

There are two likely possibilities for why I was unable to record the complete unfolding of NleC. The simplest explanation is that the initial step of NleC unfolding occurs at a force below the detection limit of our AFM platform (<5 pN). This idea is supported by the contour length of the polyprotein before the first NuG2 unfolds, which averaged 73 ± 6 nm (N = 76). While this is still lower than the expected value of 87 nm, it is within 20% of the expected value. The difference between the experimental contour length before NuG2 unfolding and the expected contour length of NleC may be accounted for if the polyprotein binds streptavidin molecules further up the tip. If this occurred, the total contour length of the polyprotein would be reduced by the distance the polyprotein bound up the tip (i.e. the distance between the polyprotein binding site and the apex of the tip). If this is the correct interpretation of our results, it is likely that while I was unable to record the full unfolding of NleC, the portion recorded in my FECs is the most important when considering the *in vivo* unfolding of NleC by the T3SS ATPase. If the unmeasured portion of NleC unfolds below the limit of detection in my experiments, it seems unlikely that it would present any significant hindrance to even a weak unfoldase.

The other possible interpretation for why I was unable to record the complete unfolding of NleC is that the protein was either misfolded or partially unfolded. While my qualitative test of NleC activity showed that it was active before surface deposition, the deposition process and subsequent washing of the surface may have unfolded NleC. If NleC remained in a partially unfolded state, the polyprotein would be more likely to bind higher up the tip, resulting in the observed contour length before NuG2 unfolding. This could also be the case if a portion of NleC

naturally contains a very unstable fold that easily fluctuates between folding and unfolding. This is unlikely because the entire fragment of NleC being tested (22-279) was able to crystalize, typically indicative of a stable fold. If NleC was unfolded during deposition and then misfolded into a structure that differed from the measured crystal structure, the predicted change in contour length would no longer be correct.

Finally, it is possible, although unlikely, that a portion of NleC requires more force to unfold than the rupture force of streptavidin:biotin (~100-150 pN). This can be investigated by moving NleC into the Fgβ-YBBR construct and probing with an SdrG functionalized tip. The high rupture force of Fgβ:SdrG (>800 pN), should provide enough strength to unfold even very stably folded protein.

Near Equilibrium Fluctuations of NleC Will Enable Model-Free Energy Landscape Reconstruction

One of the key advantages of the Warhammer tip geometry is the average temporal resolution of <10 μs combined with sub-pN stability over hundreds of seconds. These characteristics allowed me to collect FECs of NleC at very low pulling rates and see short lived unfolding events. As I decreased the retraction velocity, it was possible to identify near equilibrium fluctuations between the folded and unfolded state of NleC. At a retraction velocity of 20 nm/s, I consistently recorded near equilibrium fluctuations, as shown in Figure IV-7A. If this data is plotted as force vs time (Figure IV-7B), it is possible to assign the lifetimes in the folded and unfolded state.
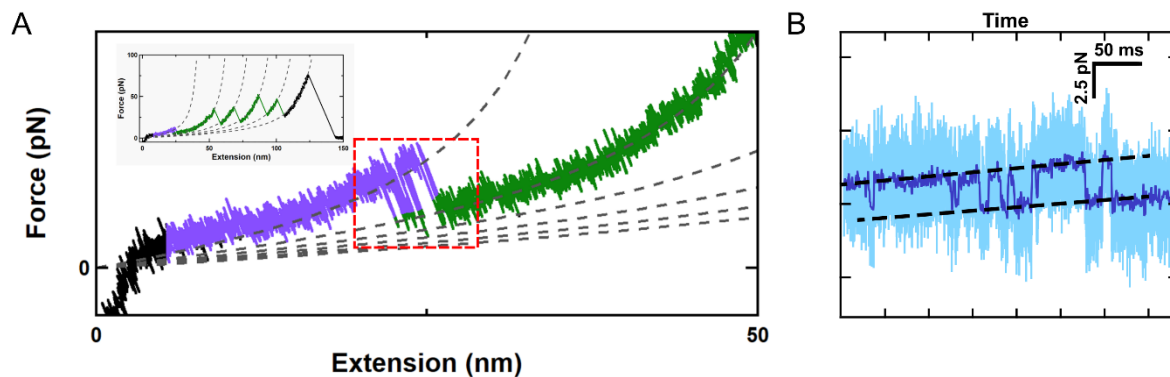
Figure IV-7. Near-equilibrium fluctuations of NleC captured with modified cantilever. (A) A representative FEC taken at 20 nm/s which shows near-equilibrium fluctuations of Nlec between the folded and unfolded state. (B) A force vs time plot of the data within the red box in A. Data was collected at 50 kHz (light blue) and smoothed to 1 kHz (dark blue).

While beyond the scope of this thesis, the ability to capture near-equilibrium fluctuations between the folded and unfolded state of NleC makes it possible to reconstruct the free energy landscape using fluctuation theorems such as Jarzynski's equality or the Dudko model. These methods allow for the recovery of equilibrium free energies from non-equilibrium measurements.[147–149] The Perkins lab is developing tools to analyze the acquired low speed FECs of NleC using the inverse Weierstrass transform (IWT) to recover the energy landscape. IWT is not only a simple and efficient method to analyze the date, it avoids convolving the effect of the cantilever in the calculated energy profile.[150] The ability to extract the free energy landscape of NleC will provide a much better understanding of how NleC reacts to force. If near equilibrium fluctuations can also be obtained for Protealysin, it will be possible to compare their energy landscapes. This may be critical in elucidating how the sequence divergence of NleC has modulated its response to force in comparison to non-secreted homologues.

The Initial Unfolding of Protealysin May Require an Order of Magnitude Higher force than NleC

An interesting aspect of the observed unfolding of protealysin is that like NleC, it unfolds in at least two steps. The key difference between the two is that the initial unfolding step of protealysin can be measured with our AFM platform, whereas the initial unfolding step of NleC is masked. Providing that NleC is not simply misfolded due to its presence in the polyprotein, these results would indicate that the initial unfolding of NleC occurs at a force <5 pN while the initial unfolding of protealysin occurs at 22 pN. This could mean that protealysin requires an order of magnitude more force to initiate unfolding than NleC, a factor that may have driven the sequence divergence. If future studies testing the *in vivo* secretion rate of protealysin and NleC show a significant difference in rates, this is a likely candidate for the underlying cause.

Protealysin May Not Be the Ideal Non-Secreted Homologue for Comparison to NleC

As noted in the introduction, NleC does not share any significant sequence similarity to other Zincins. Therefore, there was little information to use when picking a homologue to mechanically characterize. While protealysin is a well characterized Zincin, it is possible that characterizing other members of the superfamily will provide larger, more easily measure differences to NleC. Future studies will mechanically characterize more Zincins, including Zincin Thermolysin from *Bacillus thermoproteolyticus,* in the hope of finding a homologue that unfolds at high force. This would help elucidate how the sequence divergence of NleC, and the unique aspects of its fold, facilitate its secretion through the T3SS. Moreover, characterizing various members of the Zincin superfamily may more generally provide a better understanding of how protein structure outside of conserved catalytic domains affects mechanical stability. The site-specific attachment schemes developed in this thesis provide a method to rapidly screen a variety

of proteins, opening the door to wide ranging studies of effector proteins and their non-secreted homologues.

Methods

Insertion of NleC into Dual-Aldehyde Construct

To simplify the interpretation of our data, only the catalytic core of NleC (22-279) which has a known structure was amplified and inserted into the Dual-Aldehyde construct. Cloning was done using the Sticky-end PCR,[71] in which the gene fragment was amplified with two primer pairs (U2062/L2062 and U2062/L2062.1), one of which was slightly longer than the other. When the resulting products were mixed, melted, and annealed, a fraction of DNA contained a blunt end and a sticky end compatible with NcoI. This fragment was inserted into pMS 1251 opened EcoRV/NcoI.

Insertion of Inactive Protealysin Into Fgβ-YBBR Construct

To avoid potential proteolysis during expression and purification, the gene for an inactive mutant of protealysin was synthesized by Genewiz. The mutant had been previously characterized and was shown to maintain the chelated Zn^{2+} ion, a factor we believed would be important for its mechanical stability. The gene was amplified with U/L2392, and inserted into pMS 1554 (6xHis-SUMO-Fgβ-ddFLN4-YBBR) EcoRV/NcoI.

Expression, Purification, of Polyproteins

Expression and purification for both NleC in the Dual-Aldehyde construct and protealysin in the Fgβ-YBBR construct was carried out exactly at described in Chapter 2. For both polyproteins, the elution from the Ni-NTA column was diluted in half in TBS + TCEP (25 mM Tris,

pH 8.0, 150 mM NaCl, 2 mM TCEP) and 100 uL of 0.5 mg/mL Ulp1 was added to cleave the SUMO

Tag. NleC in the Dual-Aldehyde construct was dialyzed into Aldehyde Conversion Buffer (25 mM

TEA, pH 9.0, 10 mM NaCl, 5 mM BME). Protealysin in the Fgβ-YBBR construct was dialyzed into

HEPES Buffer (25 mM HEPES, pH 7.2, 150 mM NaCl, 5 mM BME) at 25°C, concentrated in an

Amicon centrifugal concentrator (10 kDa), and either immediately used for AFM experiments or

flash frozen and stored at -80°C.

Aldehyde Conversion and Labeling of NleC in Dual-Aldehyde Construct

After dialysis into aldehyde conversion buffer, the NleC polyprotein was reacted with FGE

in a 1:10 FGE:polyprotein ratio at 18°C overnight. After conversion, the protein was buffer

exchanged into HIPS-conversion buffer using an S200 analytical SEC column. 10x molar excess

HIPS-DBCO was added (a precipitate formed upon addition of HIPS-DBCO) and the mixture was

incubated overnight at 37°C. 10x molar excess HIPS-Biotin was added and the mixture was

incubated for 3 days at 37°C. Excess label was removed by SEC (S75 analytical) equilibrated with

PBS + TCEP (25 mM sodium phosphate, pH 7.2, 150 mM NaCl, 2 mM TCEP).

AFM Surface and Tip Preparation

Azide surfaces and streptavidin tips were made exactly as described in Chapter 2. CoA

surfaces and tips were made exactly as described in Chapter 3. CoA tips were functionalized with

YBBR-SdrG provided by the Gaub lab.

Protein Deposition

HIPS labeled protein was deposited on azide functionalized surfaces as described in

Chapter 2 (20 µL overnight at 4°C). Typically, 0.3-0.5 mg/mL protein was deposited, although as

the stock of HIPS reagents became less active, deposition concentration was increased to 1 mg/mL. Surfaces were rinsed 10x with 1 mL PBS + TCEP, with 20 µM Zn^{2+}.

Protealysin in Fgβ-YBBR was deposited on CoA surfaces at 1-5 µM in the presence of 25 mM $MgCl_2$ and 3 µM Sfp, for 30 minutes at 25°C. Surfaces were rinsed 10x with 1 mL of HEPES buffer + 20 µM Zn^{2+}.

Expression and Purification of RelA and RelB

SUMO-RelA (pMS 982) and SUMO-RelB (pMS 1114) were transformed into *E. Coli* Bl-21 (DE3), and expression and purification was carried out exactly as described in Chapter 2 for the polyprotein construct, including ULP1 cleavage. The proteins were dialyzed into TBS (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP) and used immediately to test NleC activity since Dr. Turco had found that after 1 week, RelA could no longer be cleaved by NleC.

NleC Activity Assay – Degradation of RelA and RelB

10 µL of HIPS labeled NleC polyprotein (~1 mg/mL) was added to 50 uL aliquots of freshly purified RelA (~3 mg/mL) and RelB (~1 mg/mL) and incubated at 25°C for 30 m. Samples were run on a 12% SDS-PAGE gel, stained with Coomassie R-250. NleC cleaves RelA between residues 38 and 39, leaving fragments 1-38 (~4 kDa) and 39-291 (~28 kDa). NleC cleaves RelB between residues 144 and 145, leaving fragments 124-144 (~3 kDa) and 144-417 (~32 kDa). Because of the small N-terminal fragment sizes, I only observed the C-terminal fragment of both proteins after cleavage.

References

1. Troisfontaines, P. & Cornelis, G. R. Type III secretion: more systems than you think. *Physiol. Bethesda Md* **20,** 326–339 (2005).

2. Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B. & Strynadka, N. C. J. Structural analysis of a prototypical ATPase from the type III secretion system. *Nat. Struct. Mol. Biol.* **14,** 131–137 (2007).

3. Loquet, A. *et al.* Atomic model of the type III secretion system needle. *Nature* **486,** 276–279 (2012).

4. Notti, R. Q. & Stebbins, C. E. The Structure and Function of Type III Secretion Systems. *Microbiol. Spectr.* **4,** (2016).

5. Galán, J. E. Common Themes in the Design and Function of Bacterial Effectors. *Cell Host Microbe* **5,** 571–579 (2009).

6. Akeda, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437,** 911–915 (2005).

7. Lara-Tejero, M., Kato, J., Wagner, S., Liu, X. & Galán, J. E. A Sorting Platform Determines the Order of Protein Secretion in Bacterial Type III Systems. *Science* **331,** 1188–1191 (2011).

8. Winnen, B. *et al.* Hierarchical Effector Protein Transport by the Salmonella Typhimurium SPI-1 Type III Secretion System. *PLoS ONE* **3,** e2178 (2008).

9. Akeda, Y. & Galán, J. E. Genetic analysis of the Salmonella enterica type III secretion-associated ATPase InvC defines discrete functional domains. *J. Bacteriol.* **186,** 2402–2412 (2004).

10. Woestyn, S., Allaoui, A., Wattiau, P. & Cornelis, G. R. YscN, the putative energizer of the Yersinia Yop secretion machinery. *J. Bacteriol.* **176,** 1561–1569 (1994).

11. Pozidis, C. *et al.* Type III protein translocase: HrcN is a peripheral ATPase that is activated by oligomerization. *J. Biol. Chem.* **278,** 25816–25824 (2003).

12. Galán, J. E., Lara-Tejero, M., Marlovits, T. C. & Wagner, S. Bacterial Type III Secretion Systems: Specialized Nanomachines for Protein Delivery into Target Cells. *Annu. Rev. Microbiol.* **68,** 415–438 (2014).

13. Cornelis, G. R. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4,** 811–825 (2006).

14. Coburn, B., Sekirov, I. & Finlay, B. B. Type III Secretion Systems and Disease. *Clin. Microbiol. Rev.* **20,** 535–549 (2007).

15. Hueck, C. J. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.* **62,** 379–433 (1998).

16. Figueira, R. & Holden, D. W. Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiol. Read. Engl.* **158,** 1147–1161 (2012).

17. McCann, H. C. & Guttman, D. S. Evolution of the type III secretion system and its effectors in plant-microbe interactions. *New Phytol.* **177,** 33–47 (2008).

18. Arnold, R. *et al.* Sequence-Based Prediction of Type III Secreted Proteins. *PLOS Pathog.* **5,** e1000376 (2009).

19. Goldberg, T., Rost, B. & Bromberg, Y. Computational prediction shines light on type III secretion origins. *Sci. Rep.* **6,** 34516 (2016).

20. Dawson, J. E. & Nicholson, L. K. Folding kinetics and thermodynamics of Pseudomonas syringae effector protein AvrPto provide insight into translocation via the type III secretion system. *Protein Sci. Publ. Protein Soc.* **17,** 1109–1119 (2008).

21. Dawson, J. E. *et al.* Elucidation of a pH-folding switch in the Pseudomonas syringae effector protein AvrPto. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8543–8548 (2009).

22. Rief, M., Gautel, M. & Gaub, H. E. Unfolding Forces of Titin and Fibronectin Domains Directly Measured by AFM. in *Elastic Filaments of the Cell* (eds. Granzier, H. L. & Pollack, G. H.) **481,** 129–141 (Springer US, 2000).

23. Sato, T., Esaki, M., Fernandez, J. M. & Endo, T. Comparison of the protein-unfolding pathways between mitochondrial protein import and atomic-force microscopy measurements. *Proc. Natl. Acad. Sci.* **102,** 17999–18004 (2005).

24. Ng, S. P. *et al.* Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.* **350,** 776–789 (2005).

25. Yagawa Keisuke *et al.* Structural basis for unfolding pathway-dependent stability of proteins: Vectorial unfolding versus global unfolding. *Protein Sci.* **19,** 693–702 (2010).

26. Cao, Y., Kuske, R. & Li, H. Direct Observation of Markovian Behavior of the Mechanical Unfolding of Individual Proteins. *Biophys. J.* **95,** 782–788 (2008).

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

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

29. Binnig, null, Quate, null & Gerber, null. Atomic force microscope. *Phys. Rev. Lett.* **56,** 930–933 (1986).

30. Lee, G. U., Chrisey, L. A. & Colton, R. J. Direct measurement of the forces between complementary strands of DNA. *Science* **266,** 771–773 (1994).

31. Bull, M. S., Sullan, R. M. A., Li, H. & Perkins, T. T. Improved Single Molecule Force Spectroscopy Using Micromachined Cantilevers. *ACS Nano* **8,** 4984–4995 (2014).

32. Edwards, D. T., Faulk, J. K., LeBlanc, M.-A. & Perkins, T. T. Force Spectroscopy with 9-µs Resolution and Sub-pN Stability by Tailoring AFM Cantilever Geometry. *Biophys. J.* **113,** 2595–2600 (2017).

33. Churnside, A. B. *et al.* Routine and Timely Sub-picoNewton Force Stability and Precision for Biological Applications of Atomic Force Microscopy. *Nano Lett.* **12,** 3557–3561 (2012).

34. Edwards, D. T. *et al.* Optimizing 1-µs-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope. *Nano Lett.* **15,** 7091–7098 (2015).

35. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265,** 1599–1600 (1994).

36. Bippes, C. A. & Muller, D. J. High-resolution atomic force microscopy and spectroscopy of native membrane proteins. *Rep. Prog. Phys.* **74,** 086601 (2011).

37. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72,** 1541–1555 (1997).

38. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci.* **96,** 3694–3699 (1999).

39. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393,** 181–185 (1998).

40. Oesterhelt, F. *et al.* Unfolding Pathways of Individual Bacteriorhodopsins. *Science* **288,** 143–146 (2000).

41. Ott, W., Jobst, M. A., Schoeler, C., Gaub, H. E. & Nash, M. A. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. *J. Struct. Biol.* doi:10.1016/j.jsb.2016.02.011

42. Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3,** 321–322 (2007).

43. Bornschlögl, T. & Rief, M. Single-molecule protein unfolding and refolding using atomic force microscopy. *Methods Mol. Biol. Clifton NJ* **783,** 233–250 (2011).

44. Zimmermann, J. L., Nicolaus, T., Neuert, G. & Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat. Protoc.* **5,** 975–985 (2010).

45. Dietz, H. & Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16192–16197 (2004).

46. Kufer, S. K. *et al.* Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. *Eur. Biophys. J.* **35,** 72–78 (2005).

47. Taniguchi, Y. & Kawakami, M. Application of HaloTag Protein to Covalent Immobilization of Recombinant Proteins for Single Molecule Force Spectroscopy. *Langmuir* **26,** 10433–10436 (2010).

48. Popa, I. *et al.* Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.* **135,** 12762–12771 (2013).

49. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109,** E690–E697 (2012).

50. Lee, C.-K., Wang, Y.-M., Huang, L.-S. & Lin, S. Atomic force microscopy: Determination of unbinding force, off rate and energy barrier for protein–ligand interaction. *Micron* **38,** 446–461 (2007).

51. Lee, G. U., Kidwell, D. A. & Colton, R. J. Sensing Discrete Streptavidin-Biotin Interactions with Atomic Force Microscopy. *Langmuir* **10,** 354–357 (1994).

52. Rico, F. & Moy, V. T. Energy landscape roughness of the streptavidin–biotin interaction. *J. Mol. Recognit.* **20,** 495–501 (2007).

53. Walder, R. *et al.* Rapid Characterization of a Mechanically Labile α-helical Protein Enabled by Efficient Site-Specific Bioconjugation. *Manuscr. Prep.*

54. Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* **104,** 16793–16797 (2007).

55. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126,** 15046–15047 (2004).

56. Kim, Y. *et al.* Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjug. Chem.* **19,** 786–791 (2008).

57. Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3,** 373–382 (2008).

58. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21,** 86–89 (2003).

59. Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15,** 128–136 (2008).

60. Carlson, B. L. *et al.* Function and Structure of a Prokaryotic Formylglycine-generating Enzyme. *J. Biol. Chem.* **283,** 20117–20125 (2008).

61. Cao, Y., Balamurali, M. M., Sharma, D. & Li, H. A functional single-molecule binding assay via force spectroscopy. *Proc. Natl. Acad. Sci.* **104,** 15677–15681 (2007).

62. Shi, X. *et al.* Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat. Methods* **9,** 499–503 (2012).

63. Rabuka, D., Rush, J. S., deHart, G. W., Wu, P. & Bertozzi, C. R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat. Protoc.* **7,** 1052–1067 (2012).

64. Holder, P. G. *et al.* Reconstitution of Formylglycine-generating Enzyme with Copper(II) for Aldehyde Tag Conversion. *J. Biol. Chem.* **290,** 15730–15745 (2015).

65. Kalia, J. & Raines, R. T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed Engl.* **47,** 7523–7526 (2008).

66. Agarwal, P. *et al.* Hydrazino-Pictet-Spengler Ligation as a Biocompatible Method for the Generation of Stable Protein Conjugates. *Bioconjug. Chem.* **24,** 846–851 (2013).

67. Tan, E. *et al.* A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9308–9313 (2003).

68. Okumus, B., Wilson, T. J., Lilley, D. M. J. & Ha, T. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.* **87,** 2798–2806 (2004).

69. He, C. *et al.* Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angew. Chem. Int. Ed.* **54,** 9921–9925 (2015).

70. Zhu, Y. *et al.* Ultrafast folding of alpha3D: a de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15486–15491 (2003).

71. Junker, J. P., Ziegler, F. & Rief, M. Ligand-Dependent Equilibrium Fluctuations of Single Calmodulin Molecules. *Science* **323,** 633–637 (2009).

72. Hertadi, R. & Ikai, A. Unfolding mechanics of holo- and apocalmodulin studied by the atomic force microscope. *Protein Sci. Publ. Protein Soc.* **11,** 1532–1538 (2002).

73. Stigler, J. & Rief, M. Calcium-dependent folding of single calmodulin molecules. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17814–17819 (2012).

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

75. Zeng, G. Sticky-end PCR: new method for subcloning. *BioTechniques* **25,** 206–208 (1998).

76. Zheng, P., Takayama, S. J., Mauk, A. G. & Li, H. Single Molecule Force Spectroscopy Reveals That Iron Is Released from the Active Site of Rubredoxin by a Stochastic Mechanism. *J. Am. Chem. Soc.* **135,** 7992–8000 (2013).

77. Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **41,** 4781–4796 (2012).

78. Viani, M. B. *et al.* Small cantilevers for force spectroscopy of single molecules. *J. Appl. Phys.* **86,** 2258–2262 (1999).

79. Khan, M. N. Kinetics and mechanism of the alkaline hydrolysis of maleimide. *J. Pharm. Sci.* **73,** 1767–1771 (1984).

80. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271,** 795–799 (1996).

81. Cluzel, P. *et al.* DNA: an extensible molecule. *Science* **271,** 792–794 (1996).

82. Dietz, H. *et al.* Cysteine engineering of polyproteins for single-molecule force spectroscopy : Article : Nature Protocols. *Nat Protoc.* **1,** 80–84 (2006).

83. Walder, R., Van Patten, W. J., Adhikari, A. & Perkins, T. T. Going Vertical To Improve the Accuracy of Atomic Force Microscopy Based Single-Molecule Force Spectroscopy. *ACS Nano* (2017). doi:10.1021/acsnano.7b05721

84. Hermanson, G. T. 1 - Functional Targets. in *Bioconjugate Techniques* 3–136 (Academic Press, 1996). doi:10.1016/B978-012342335-1/50002-6

85. Lee, C.-D. *et al.* An improved SUMO fusion protein system for effective production of native proteins. *Protein Sci. Publ. Protein Soc.* **17,** 1241–1248 (2008).

86. Nauli, S., Kuhlman, B. & Baker, D. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8,** 602–605 (2001).

87. Sader, J. E., Chon, J. W. M. & Mulvaney, P. Calibration of rectangular atomic force microscope cantilevers. *Rev. Sci. Instrum.* **70,** 3967–3969 (1999).

88. Lee, V. T. & Schneewind, O. Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J. Bacteriol.* **184,** 3740–3745 (2002).

89. Radics, J., Königsmaier, L. & Marlovits, T. C. Structure of a pathogenic type 3 secretion system in action. *Nat. Struct. Mol. Biol.* **21,** 82–87 (2014).

90. Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D. Mechanical Processes in Biochemistry. *Annu. Rev. Biochem.* **73,** 705–748 (2004).

91. Carrion-Vazquez, M. *et al.* The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Mol. Biol.* **10,** 738–743 (2003).

92. Oberhauser, A. F., Badilla-Fernandez, C., Carrion-Vazquez, M. & Fernandez, J. M. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol. Biol.* **319,** 433–447 (2002).

93.   Rief, M., Pascual, J., Saraste, M. & Gaub, H. E. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles11Edited by W. Baumeister. *J. Mol. Biol.* **286,** 553–561 (1999).

94.   Brockwell, D. J. *et al.* Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nat. Struct. Mol. Biol.* **10,** 731–737 (2003).

95.   Olivares, A. O., Kotamarthi, H. C., Stein, B. J., Sauer, R. T. & Baker, T. A. Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. *Proc. Natl. Acad. Sci.* 201707794 (2017). doi:10.1073/pnas.1707794114

96.   Portaliou, A. G. *et al.* Hierarchical protein targeting and secretion is controlled by an affinity switch in the type III secretion system of enteropathogenic Escherichia coli. *EMBO J.* **36,** 3517–3531 (2017).

97.   VanEngelenburg, S. B. & Palmer, A. E. Quantification of Real-Time Salmonella Effector Type-III Secretion Kinetics Reveals Differential Secretion Rates for SopE2 and SptP. *Chem. Biol.* **15,** 619–628 (2008).

98.   Stender, S. *et al.* Identification of SopE2 from Salmonella typhimurium, a conserved guanine nucleotide exchange factor for Cdc42 of the host cell. *Mol. Microbiol.* **36,** 1206–1221 (2000).

99.   Fu, Y. & Galán, J. E. A salmonella protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401,** 293–297 (1999).

100.   Crampton, N. & Brockwell, D. J. Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.* **20,** 508–517 (2010).

101.   Williams, C., Galyov, E. E. & Bagby, S. solution structure, backbone dynamics, and interaction with Cdc42 of Salmonella guanine nucleotide exchange factor SopE2. *Biochemistry (Mosc.)* **43,** 11998–12008 (2004).

102. Stebbins, C. E. & Galán, J. E. Modulation of host signaling by a bacterial mimic: structure of the Salmonella effector SptP bound to Rac1. *Mol. Cell* **6,** 1449–1460 (2000).

103. Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. Can Non-Mechanical Proteins Withstand Force? Stretching Barnase by Atomic Force Microscopy and Molecular Dynamics Simulation. *Biophys. J.* **81,** 2344–2356 (2001).

104. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72,** 1568–1581 (1997).

105. Zhang, Y., Ha, T. & Marqusee, S. Editorial Overview: Single-Molecule Approaches up to Difficult Challenges in Folding and Dynamics. *J. Mol. Biol.* **430,** 405–408 (2018).

106. Schoeler, C. *et al.* Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Lett.* **15,** 7370–7376 (2015).

107. Bauer, M. S., Milles, L. F., Sedlak, S. M. & Gaub, H. E. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. *bioRxiv* 276444 (2018). doi:10.1101/276444

108. Baumann, F. *et al.* Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.* **11,** 89–94 (2016).

109. Durner, E., Ott, W., Nash, M. A. & Gaub, H. E. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega* **2,** 3064–3069 (2017).

110. Erlich, K. R., Baumann, F., Pippig, D. A. & Gaub, H. E. Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods* **1,** n/a-n/a (2017).

111. Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15815–15820 (2005).

112. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase : Article : Nature Protocols. *Nat Protoc.* **1,** 280–285 (2006).

113. Green, N. M. Avidin. in *Advances in Protein Chemistry* (eds. Anfinsen, C. B., Edsall, J. T. & Richards, F. M.) **29,** 85–133 (Academic Press, 1975).

114. Carvalho, A. L. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 13809–13814 (2003).

115. Smith, S. P. & Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* **23,** 686–694 (2013).

116. Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **5,** ncomms6635 (2014).

117. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).

118. Foster, T. J. & Höök, M. Surface protein adhesins of Staphylococcus aureus. *Trends Microbiol.* **6,** 484–488 (1998).

119. Ponnuraj, K. *et al.* A "dock, lock, and latch" Structural Model for a Staphylococcal Adhesin Binding to Fibrinogen. *Cell* **115,** 217–228 (2003).

120. Bowden, M. G. *et al.* Evidence for the "Dock, Lock, and Latch" Ligand Binding Mechanism of the Staphylococcal Microbial Surface Component Recognizing Adhesive Matrix Molecules (MSCRAMM) SdrG. *J. Biol. Chem.* **283,** 638–647 (2008).

121. Herman, P. *et al.* The binding force of the staphylococcal adhesin SdrG is remarkably strong. *Mol. Microbiol.* **93,** 356–368 (2014).

122. Milles, L. F., Schulten, K., Gaub, H. E. & Bernardi, R. C. Molecular mechanism of extreme mechanostability in a pathogen adhesin. 7 (2018).

123. Allison, D. P., Mortensen, N. P., Sullivan, C. J. & Doktycz, M. J. Atomic force microscopy of biological samples. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2,** 618–634 (2010).

124. Single Molecule Studies of Protein Folding Using Atomic Force Microscopy - Springer. in (eds. Bai, Y. & Nussinov, R.) (Humana Press, 2006).

125. Rico, F., Rigato, A., Picas, L. & Scheuring, S. Mechanics of proteins with a focus on atomic force microscopy. *J. Nanobiotechnology* **11,** S3 (2013).

126. Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200,** 618–627 (1978).

127. Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A. & Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* **11,** 81–85 (2004).

128. Turco, M. M. & Sousa, M. C. The structure and specificity of the type III secretion system effector NleC suggest a DNA mimicry mechanism of substrate recognition. *Biochemistry (Mosc.)* **53,** 5131–5139 (2014).

129. Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol. Biol. Clifton NJ* **1091,** 17–32 (2014).

130. Yen, H. *et al.* NleC, a Type III Secretion Protease, Compromises NF-κB Activation by Targeting p65/RelA. *PLoS Pathog.* **6,** (2010).

131. Halavaty, A. S. *et al.* Structure of the Type III Secretion Effector Protein ExoU in Complex with Its Chaperone SpcU. *PLOS ONE* **7,** e49388 (2012).

132. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545-549 (2010).

133. Kurisu, G. *et al.* Structure of the zinc endoprotease from Streptomyces caespitosus. *J. Biochem. (Tokyo)* **121,** 304–308 (1997).

134. Puhar, A. & Sansonetti, P. J. Type III secretion system. *Curr. Biol.* **24,** R784–R791 (2014).

135. Tsai, C.-L., Burkinshaw, B. J., Strynadka, N. C. J. & Tainer, J. A. The Salmonella Type III Secretion System Virulence Effector Forms a New Hexameric Chaperone Assembly for Export of Effector/Chaperone Complexes: FIG 1. *J. Bacteriol.* **197,** 672–675 (2015).

136. Diepold, A. & Armitage, J. P. Type III secretion systems: the bacterial flagellum and the injectisome. *Phil Trans R Soc B* **370,** 20150020 (2015).

137. Tezcan-Merdol, D., Engstrand, L. & Rhen, M. Salmonella enterica SpvB-mediated ADP-ribosylation as an activator for host cell actin degradation. *Int. J. Med. Microbiol. IJMM* **295,** 201–212 (2005).

138. Ritter, H., Koch-Nolte, F., Marquez, V. E. & Schulz, G. E. Substrate binding and catalysis of ecto-ADP-ribosyltransferase 2.2 from rat. *Biochemistry (Mosc.)* **42,** 10155–10162 (2003).

139. Margarit, S. M., Davidson, W., Frego, L. & Stebbins, C. E. A steric antagonism of actin polymerization by a salmonella virulence protein. *Struct. Lond. Engl. 1993* **14,** 1219–1229 (2006).

140. Alonso, A. *et al.* Lck Dephosphorylation at Tyr-394 and Inhibition of T Cell Antigen Receptor Signaling by Yersinia Phosphatase YopH. *J. Biol. Chem.* **279,** 4922–4928 (2004).

141. Stuckey, J. A. *et al.* Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 A and the complex with tungstate. *Nature* **370,** 571–575 (1994).

142. Selner, N. G. *et al.* Diverse Levels of Sequence Selectivity and Catalytic Efficiency of Protein-Tyrosine Phosphatases. *Biochemistry (Mosc.)* **53,** 397–412 (2014).

143. Mühlen, S., Ruchaud-Sparagano, M.-H. & Kenny, B. Proteasome-independent Degradation of Canonical NFκB Complex Components by the NleC Protein of Pathogenic Escherichia coli. *J. Biol. Chem.* **286,** 5100–5107 (2011).

144. Pearson, J. S., Riedmaier, P., Marchès, O., Frankel, G. & Hartland, E. L. A type III effector protease NleC from enteropathogenic Escherichia coli targets NF-κB for degradation. *Mol. Microbiol.* **80,** 219–230 (2011).

145. Baruch, K. *et al.* Metalloprotease type III effectors that specifically cleave JNK and NF-κB. *EMBO J.* **30,** 221–231 (2011).

146. Demidyuk, I. V. *et al.* Crystal Structure of the Protealysin Precursor. *J. Biol. Chem.* **285,** 2003–2013 (2010).

147. Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci.* **98,** 3658–3661 (2001).

148. Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. & Bustamante, C. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science* **296,** 1832–1835 (2002).

149. Dudko, O. K., Hummer, G. & Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci.* **105,** 15755–15760 (2008).

150. Hummer, G. & Szabo, A. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci.* **107,** 21441–21446 (2010).

151. Tsai, H.-S., Wang, Y.-Z., Lin, J.-J. & Lien, W.-F. Preparation and properties of sulfopropyl chitosan derivatives with various sulfonation degree. *J. Appl. Polym. Sci.* NA-NA (2009). doi:10.1002/app.31689

Appendix A Synthesis of HIPS Reagents

While in Chapter 2 I demonstrated the ability to efficiently label aldehyde modified proteins with Hydrazino-Pictet Spengler (HIPS) reagents, over time I noticed I needed to deposit more and more protein to achieve similar surface densities of protein. While initially I believed the problem was with inconsistencies in the surface preparation, it was discovered by production chemists at ClickChemTools that HIPS degrades quickly even if stored at -20°C ($t_{1/2}$ = 1 m at -20°C) due to oxidation of the hydrazine moeity. This degradation can be avoided by protecting the hydrazine with fluorenylmethyloxycarbonyl (FMOC). The FMOC can be easily removed directly prior to use, allowing for long term storage of HIPS reagents. Because my collaborator, Dr. Jake Greenberg, had left, Catalent Biologics graciously offered to synthesize new HIPS reagents using FMOC protected HIPS (synthesis and deprotection shown below).

While HIPS-Biotin was made with the same Biotin-PEG$_3$-Amine (ClickChemTools) Dr. Greenberg had used as a precursor, to increase the solubility of HIPS-DBCO, I ordered Sulfo-DBCO-Amine (ClickChemTools) to use a precursor instead of DBCO-Amine. Sulfonation has been shown to dramatically increase the aqueous solubility of various small molecules.[151] Indeed, I found significantly that HIPS-Sulfo-DBCO resulted in significantly less precipitation when added at 10x molar concentration to aldehyde modified protein. I have not yet quantified how much this improves overall labeling efficiency, but expect there to be a significant increase in labeling efficiency with HIPS-Sulfo-DBCO compared to HIPS-DBCO.

HIPS-Biotin and HIPS-DBCO Synthesis and Deprotection from Catalent Biologic
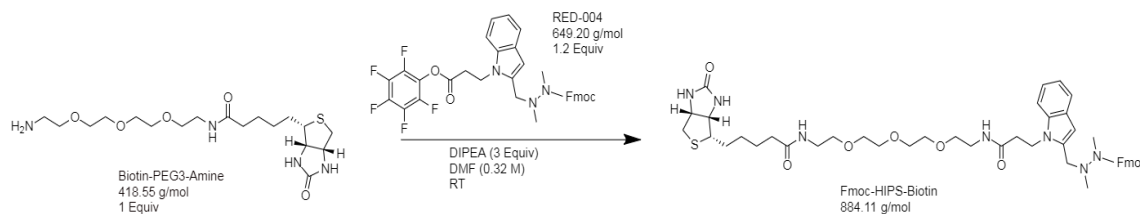
Figure A-1. Synthesis of FMOC Protected HIPS-Biotin

Synthesis of FMOC Protected HIPS-Biotin

Biotin-PEG3-Amine (1 equiv) was added to a 25 mL round bottom flask under nitrogen. Anhydrous DMF (0.32 M) and N,N-diisopropylethylamine (3 equiv) was added and allowed to stir for 5 minutes giving a clear solution. (9*H*-fluoren-9-yl)methyl 1,2-dimethyl-2-((1-(3-oxo-3-(perfluorophenoxy)propyl)-1*H*-indol-2-yl)methyl)hydrazine-1-carboxylate (RED-004,1.2 equiv) was added and the mixture turned slightly yellow. It was monitored by LCMS and allowed to stir for 4 hours. The crude material was taken directly to purification by flash chromatography on C18 using a 5-100% CH_3CN-H_2O (0.05% Formic Acid) gradient as eluent. The purified product was concentrated via rotovap and lyophilized to yield a white solid of Fmoc-HIPS-Biotin. Synthetic scheme shown in Figure A-1

The procedure was performed two times on a small (0.025 g, Biotin-PEG3-Amine) and large (0.468 g, Biotin-PEG3-Amine) scale. Overall yield and purity of combined batches shown below:

Fmoc-HIPS-Biotin (White Solid):

- Yield: 644 mg

- Overall Yield: 62%

- Purity: >98%

- LCMS Retention time: 3.385 min
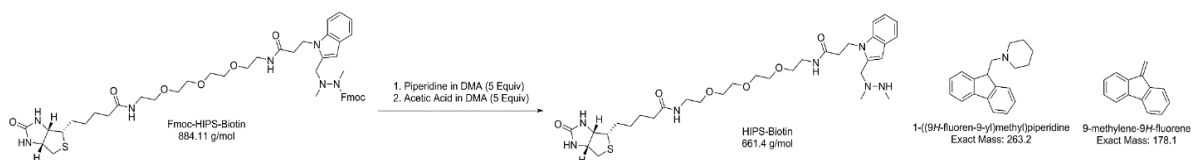
Figure A-2. Deprotection of FMOC Protected HIPS-Biotin

Deprotection of FMOC Protected HIPS-Biotin

Fmoc-HIPS-Biotin was dissolved in dry dimethylacetamide (DMA) to give a 50 mM concentration. 500 mM stock solutions were made of Piperidine and Acetic acid in dimethylacetamide. To the 50 mM Fmoc-HIPS-Biotin solution, 5 equivalents of 500 mM Piperidine stock (overall final concentration: 125 mM) was added and allowed to incubate at room temperature for 15 minutes. The reaction was quenched with 5 equivalents of 500 mM acetic acid stock solution (overall final concentration: 125 mM). This gives a final concentration of 25 mM for HIPS-Biotin solution which can then be stored at -20°C or lower temperatures to ensure long term use. A scheme of this procedure is shown in Figure A-2 and HPLC analysis of the deprotection is shown in Figure A-3. Avoid letting samples sit at room temperature for long periods. Immediately after use, return stock solution to freezer.
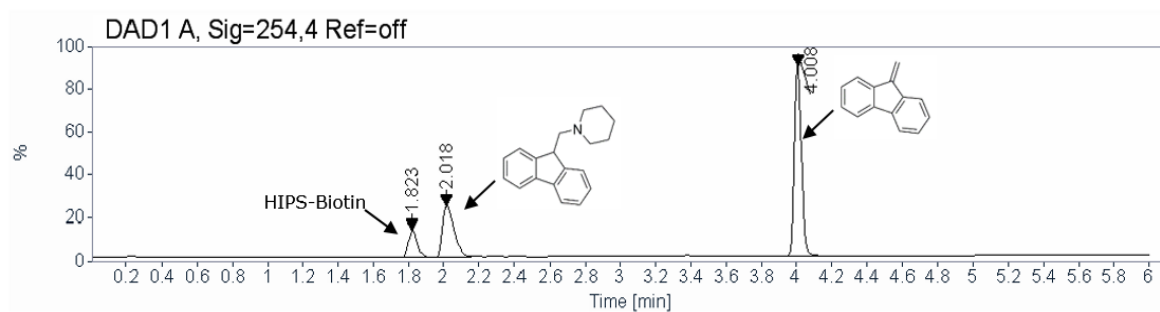
Figure A-3. HPLC of HIPS-Biotin Deprotection
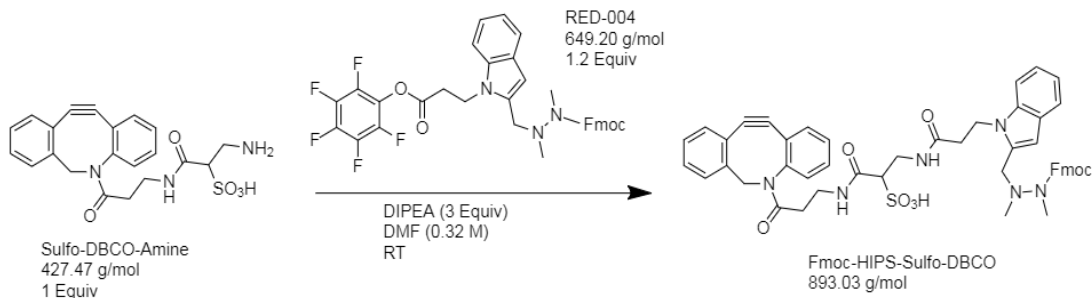
Figure A-4. Synthesis of FMOC protected HIPS-Sulfo-DBCO

Synthesis of FMOC Proctected HIPS-Sulfo-DBCO

Sulfo-DBCO-Amine (1 equiv) was added to a 25 mL round bottom flask under nitrogen. Anhydrous DMF (0.32 M) and N,N-diisopropylethylamine (3 equiv) was added and allowed to stir for 5 minutes giving a clear solution. (9*H*-fluoren-9-yl)methyl 1,2-dimethyl-2-((1-(3-oxo-3-(perfluorophenoxy)propyl)-1*H*-indol-2-yl)methyl)hydrazine-1-carboxylate (RED-004,1.2 equiv) was added and the mixture turned slightly yellow. It was monitored by LCMS and allowed to stir for 4 hours. The reaction was quenched with saturated ammonium chloride and extracted 3X with ethyl acetate. The combined organics were washed 3X using Brine, dried over magnesium sulfate, filtered, and concentrated via rotovap. The crude material was purified by silica gel chromatography (9:1 DCM:MeOH). The purified product was concentrated via rotovap and lyophilized to yield a white solid of Fmoc-HIPS-Sulfo-DBCO. The synthetic scheme is shown in Figure A-4.

LCMS with Formic acid shows broadening of product due to the sulfate group. LCMS using TFA showed sharper peak, however our system does not have mass spec attached to this LCMS. Both spectrums are attached in the packet.

The procedure was performed two times on a small (0.025 g, Sulfo-DBCO-Amine) and large (0.160 g,Sulfo-DBCO-Amine) scale. Overall yield and purity of combined batches shown below:

**Fmoc-HIPS-Sulfo-DBCO (White Solid):**

- Yield: 260 mg

- Overall Yield: 67%

- Purity: >96%

- LCMS Retention time on gradient using 0.1% Formic Acid: 3.861 min

- LCMS Retention time on gradient using 0.05% Trifluoroacetic acid: 3.848 min
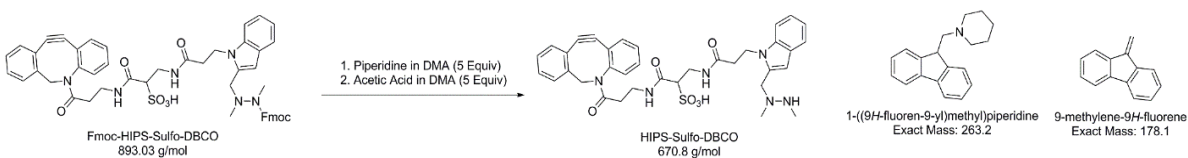
Figure A-5. Deprotection of FMOC protected HIPS-Sulfo-DBCO

Deprotection of FMOC Protected HIPS-Sulfo-DBCO

Fmoc-HIPS-Sulfo-DBCO was dissolved in dry dimethylacetamide (DMA) to give a 50 mM concentration. 500 mM stock solutions were made of Piperidine and Acetic acid in dimethylacetamide. To the 50 mM Fmoc-HIPS-Sulfo-DBCO solution, 5 equivalents of 500 mM Piperidine stock (overall final concentration: 125 mM) was added and allowed to incubate at room temperature for 15 minutes. The reaction was quenched with 5 equivalents of 500 mM

acetic acid stock solution (overall final concentration: 125 mM). This gives a final concentration

of 25 mM for HIPS-Sulfo-Biotin solution which can then be stored at -20 °C or lower temperatures

to ensure long term use. Example is provided below:

Ex. 12.5 µL of 500 mM Piperidine Stock was added to 25 µL of 50 mM Fmoc-HIPS-Biotin

and incubated at room temp for 15 minutes. The reaction was quenched with 12.5 µL of 500 mM

acetic acid stock to give 25 mM HIPS-Sulfo-DBCO solution in DMA.
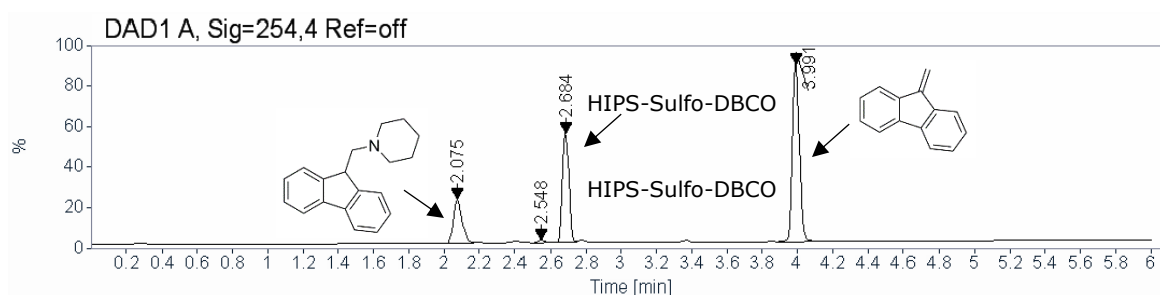
Figure A-6. HPLC Analysis of Deprotected HIPS-Sulfo-DBCO

Appendix B Table of Primers Used for Cloning

Primer	Sequence (5'-3')	Purpose
U850	GCCGGTACCCCATGGGAGCT	Linker insert into pMS 927 to make pMS 952
L850	CCCATGGGGTACCGGC	Linker insert into pMS 927 to make pMS 952
U1063	GGCTGCTCTGGAATGGATAC CTATAAGCTGGTGATTG	Sticky end PCR cloning NuG2 into pMS 984-adds N-terminal Cys Tag
L946	AbGCTCTCACGAGCGTTCGGTCACGG	Sticky end PCR cloning NuG2 into pMS 984-adds C-terminal aldehyde Tag
L947	CTCACGAGCGTTCGGTCACGG	Sticky end PCR cloning NuG2 into pMS 984-adda C-terminal aldehyde Tag
U2007	GAGATATACATATGGGCAG CAGTCATCACCATCATCACC	Cloning pET-Duet containing M. tuberculosis FGE into pMS 984
L2007	TGGTGGTGCTCGAGCTACC CGGACACCGG	Cloning pET-Duet containing M. tuberculosis FGE into pMS 984
U2008	GAGATATACATATGGGCAG CAGTCATCATCATC	Cloning NuG2$_4$ from pMS 1059 into pMS 1087 to make pMS 1088

L2008	AAAAAAAGAGCTCTCAAGA CCGCGATGGTGTGCACAGT CCAGACGAGCGTTCGGTCACGG	Cloning NuG2$_4$ from pMS 1059 into pMS 1087 to make pMS 1088
U2041	GGCATGGCACCACTGTGCAC ACCATCGCGGTCTGGAATGG ATACCTATAAGCTGGTGATTG	Sticky end PCR cloning NuG2$_4$ from pMS 1088 into pMS 984 to make pMS 1091T
L2041	AGCTCTCAGCAAGATCCCGA GCGTTCGGTCACGG	Sticky end PCR cloning NuG$_4$ from pMS 1088 into pMS 984 to make pMS 1091T
L2041. 1	CTCAGCAAGATCCCGAGCGT TCGGTCACGG	Sticky end PCR cloning NuG2$_4$ from pMS 1088 into pMS 984 to make pMS 1091T
U2086	GGACAGCGATATAGCCGACC GAGCCGG	Mutation of EcoRV site in *M. tuberculosis* FGE
U2139	GGCATGCTGACCGAGTTGG	Cloning Mt. FGE into pMS 984
L2139	AGATCCTCGAGCTACCCG	Cloning Mt. FGE into pMS 984
L2193	GACATGTTATCTTCCAGCTT TTCAAAT	Sticky end PCR cloning of rubredoxin into directional protein construct
L2193. 1	CATGGACATGTTATCTTCCA GCTTTTCAAAT	Sticky end PCR cloning of rubredoxin into directional protein construct
U2194	ATGGGCTCCTGGGCAGAAT TCAAACA	Sticky end PCR cloning of α3D into directional

		protein construct
L2194	GACGGGTTGTGACGGTATGCCT	Sticky end PCR cloning of α3D into directional protein construct
L2194.1	CATGGACGGGTTGTGACGGTATGCCT	Sticky end PCR cloning of α3D into directional protein construct
U2144	ATGGCTGACCAACTCACCGAG	Cloning calmodulin into dual aldehyde protein construct
L2144	AAAAACCATGGACTTAGCCGT CATCATTTGCACAAA	Cloning calmodulin into dual aldehyde protein construct
L2323	AAA ATG GTA CCC CCG AGC GTT CGG TC	FOR MOVING 1091 INTO DOCKERIN III CONSTRUCT (C-TERM DOCKERIN)
U2346	GGC AAT GAA GAG GGC TTT TTC AGC GCA CGT GGT CAT CGT CCG CTG GAT TGC TAA GAG CT	FGBETA TAG UPPER_ MAKES A SACI COMPATIBLE STICKY END
L2346	CTT AGC AAT CCA GCG GAC GAT GAC CAC GTG CGC TGA AAA AGC CCT CTT CAT TGC C	FGBETA TAG LOWER _ MAKES A SACI COMPATIBLE STICKY END
U2350	cacgtgGGTACCGCTAGCGGATCCATGGAC TGGTACCaaacactgttacatcagctg	Insert MCS into Dockerin Construct using CPEC
L2350	GTCCATGGATCCGCTAGCGGTACCcacgtg AGAGCCGGAACCAGAACCG	Insert MCS into Dockerin Construct using CPEC

U2351	GTGGTACCaaacactgttacatcagc	Linerarize Dockerin (1480) to insert MCS
L2351	AGAGCCGGAACCAGAACCG	Linerarize Dockerin (1480) to insert MCS
U2372	aaaaCCATGGacCCaaacactgttacatcagctg	MCS into 1480 - Contains NcoI Site
L2372	CTTACTCGAGTTAttcttcttcagcat	MCS into 1480 - Contains XhoI site
U2378	GGAaacgaagaaggcttctttagcg	Move FgBeta/YBBR Construct into 984 SfoI/SacI
U2379	GGTACCAATGAAGAGGGCT	Add MCS to Fg_Beta YBBR construct and moves into 984- SfoI/NcoI
L2379	ACGTCCATGGATCCGCTAGCGGTACCgatatcCGGTGCCGGTTTAACG	Add MCS to Fg_Beta YBBR construct and moves into 984- SfoI/NcoI
U2380	aaaaCCATGGacGACTCTCTGGAATTCATCGCTTCTAAACTGGCTTAACTCGAGTAAG	Add MCS to Fg_Beta YBBR construct and move into 984- NcoI/XhoI
L2380	CTTACTCGAGTTAAGCCAGTTTAGAAGCGATGAATTCCAGAGAGTCgtCCATGGtttt	Add MCS to Fg_Beta YBBR construct and move into 984- NcoI/XhoI
U2385	gagggtagggcggtatt	Clone Crystal Structure SopE2 into AFM Constructs
U2386	actaacataacactatccacccag	Clone Full Length SopE2

		into AFM constructs
L2386	aaaaCCATGGAggaggcattctgaagatacttattcg	Clone SopE2 into AFM constructs - NcoI
U2387	aacgatgttggagcagaaag	Clone Crystal Structure SptP into AFM Constructs
U2388	atgctaaagtatgaggagagaaaatt	Clone Full Length SptP into AFM constructs
L2388	aaaaCCATGGAgcttgccgtcgtcataag	Clone SptP into AFM constructs - NcoI
L2392	aaaaCCATGGAcgccacccctacctg	Protealysin Lower - NcoI
U2392	cacggcagcctgctg	active region Protealysin upper
U2399	AGCGCAGGTACAGGTAGC	1504/1554 Gibson Opening Primer
L2399	ACCTGAACCGCTACCGCTAC	1504/1554 Gibson Opening Primer
U2400	gtagcggtagcggttcaggtAACGATGTTGGAGCAGAAAG	SptP Gibson insertion into 1504/1554
L2400	ccgctacctgtacctgcgctGCTTGCCGTCGTCATAAG	SptP Gibson insertion into 1504/1554
U2401	gtagcggtagcggttcaggtGAGGGTAGGGCGGTATTAAC	SopE2 Gibson insertion into 1504/1554
L2401	ccgctacctgtacctgcgctGGAGGCATTCTGAAGATACTTATTC	SopeE2 Gibson insertion into 1504/1554
U2402	gtagcggtagcggttcaggtCAGCGCGACTGCGCATTAC	Inactive protealysin gibson

		insertion into 1504/1554
L2402	ccgctacctgtacctgcgctCGCCACCCCTACCTGATG	Inactive protealysin gibson insertion into 1504/1554
U2403	gtagcggtagcggttcaggtGAGGCATCAAAGCAGATTC	SpvB Gibson insertion into 1504/1554
L2403	ccgctacctgtacctgcgctTGAGTTGAGTACCCTCATG	SpvB Gibson insertion into 1504/1554
U2404	gtagcggtagcggttcaggtCGTGAACGACCACACACTTC	YopH Gibson insertion into 1504/1554
L2404	ccgctacctgtacctgcgctGCTATTTAATAATGGTCGCCCTTG	YopH Gibson insertion into 1504/1554
U2405	gtagcggtagcggttcaggtCTTGCAGTGCCTTTCATG	ART2.2 Gibson insertion into 1504/1554
L2405	ccgctacctgtacctgcgctACTGAAATCAATGTTGACTG	ART2.2 Gibson insertion into 1504/1554
U2406	gtagcggtagcggttcaggtATGGAGATGGAAAAGGAGTTC	PTP1B Gibson insertion into 1504/1554
L2406	ccgctacctgtacctgcgctGTCCTCGTGGGAAAGCTC	PTP1B Gibson insertion into 1504/1554
U2407	gtagcggtagcggttcaggtATTGCTCCTAATCGTGCTG	NleC Active Gibson insertion into 1504/1554
L2407	ccgctacctgtacctgcgctATACTCTGTGAAATCAGGAC	NleC Active Gibson insertion into 1504/1554
U2408	gtagcggtagcggttcaggtATGAAAATTCCCTCATTACAG	FL_NleC Gibson insertion into 1504/1554
L2408	ccgctacctgtacctgcgctTTGCTGATTGTGTTTGTC	FL_NleC Gibson insertion into 1504/1554

L2414	accagaaccggagcccgg	Opens 1480 for Gibson
U2414	tccggctctgtggtaccaaac	Opens 1480 for Gibson
U2415	ccgggctccggttctggtatgggctcctgggcagaattc	Insert Alpha3D into 1480 Gibson
L2415	tggtaccacagagccggagttgtgacggtatgcctgcag	Insert Alpha3D into 1480 Gibson
U2416	ctccgggctccggttctggtAACGATGTTGGAGCAGAAAG	sptp into 1480 gibson
L2416	tttggtaccacagagccggaGCTTGCCGTCGTCATAAG	sptp into 1480 gibson
U2417	ctccgggctccggttctggtGAGGGTAGGGCGGTATTAAC	sope2 into 1480 gibson
L2417	tttggtaccacagagccggaGGAGGCATTCTGAAGATACTTATTC	sope2 into 1480 gibson
U2418	ctccgggctccggttctggtATTGCTCCTAATCGTGCTG	nlec into 1480 gibson
L2418	tttggtaccacagagccggaATACTCTGTGAAATCAGGAC	nlec into 1480 gibson

Table B-1. Primers used for cloning

Appendix C Preliminary FEC of SpvB and YopH

*Additional Effector Homologue Pairs*

As discussed in Chapter 4, comparing the mechanical stability of effector proteins to their non-secreted homologues provides a unique avenue to investigate what evolutionary modifications effectors went through to facilitate their secretion through the T3SS. While my initial efforts focused on NleC and protealysin, to gain a better understanding of whether or not effector proteins evolved to unfold at lower force, I set out to test two other effector/homologue pairs. SpvB (*Salmonella)* and ART2.2 (*Rattus norvegicus)* are ADP-ribosyltransferases (ATR) and are the smallest pair of proteins in our chosen set (199 and 226 amino acids respectively).[137,138] The SpvB structure[139] contains a canonical α/β fold and NAD$^+$ binding site, despite sharing only 15-20% sequence identity with ATRs not associated with the T3SS. YopH (*Yersinia)* and PTP1B (human) are protein phosphatases, containing canonical phosphatase folds.[140–142] This is despite YopH only sharing ~15% sequence identity with non-effector phosphatases.

Preliminary FEC of SpvB and YopH

Here I show preliminary FEC data for SpvB and YopH. Due to cloning problems and weak expression, I was unable to get preliminary data for ART2.2 and PTP1B. SpvB and YopH continue the trend of effector proteins unfolding at low force, with both unfolding at less than 30 pN. Of the two, YopH unfolds at higher force, averaging ~25 pN at 800 nm/s (N = 10) while SpvB unfolds at 19 pN, although only two high-quality traces have been acquired for SpvB so far.

These proteins served as a test case for two different attachment schemes. SpvB was inserted into the Ald-NuG2$_{(4x)}$-Xmod-Dockerin III construct (Figure C-1 A), while YopH was

inserted into the Dual-Aldehyde construct (Figure C-1 B). The measured ΔL_c is indicative of the complete unfolding for both SpvB (87 nm measured, predicted is 69 nm) and YopH (109 nm measured, 109 nm predicted), demonstrating that I should be able to characterize the complete unfolding of both effector proteins. This may make them better candidates to study than NleC, since the initial unfolding of NleC cannot be detected with our AFM platform.
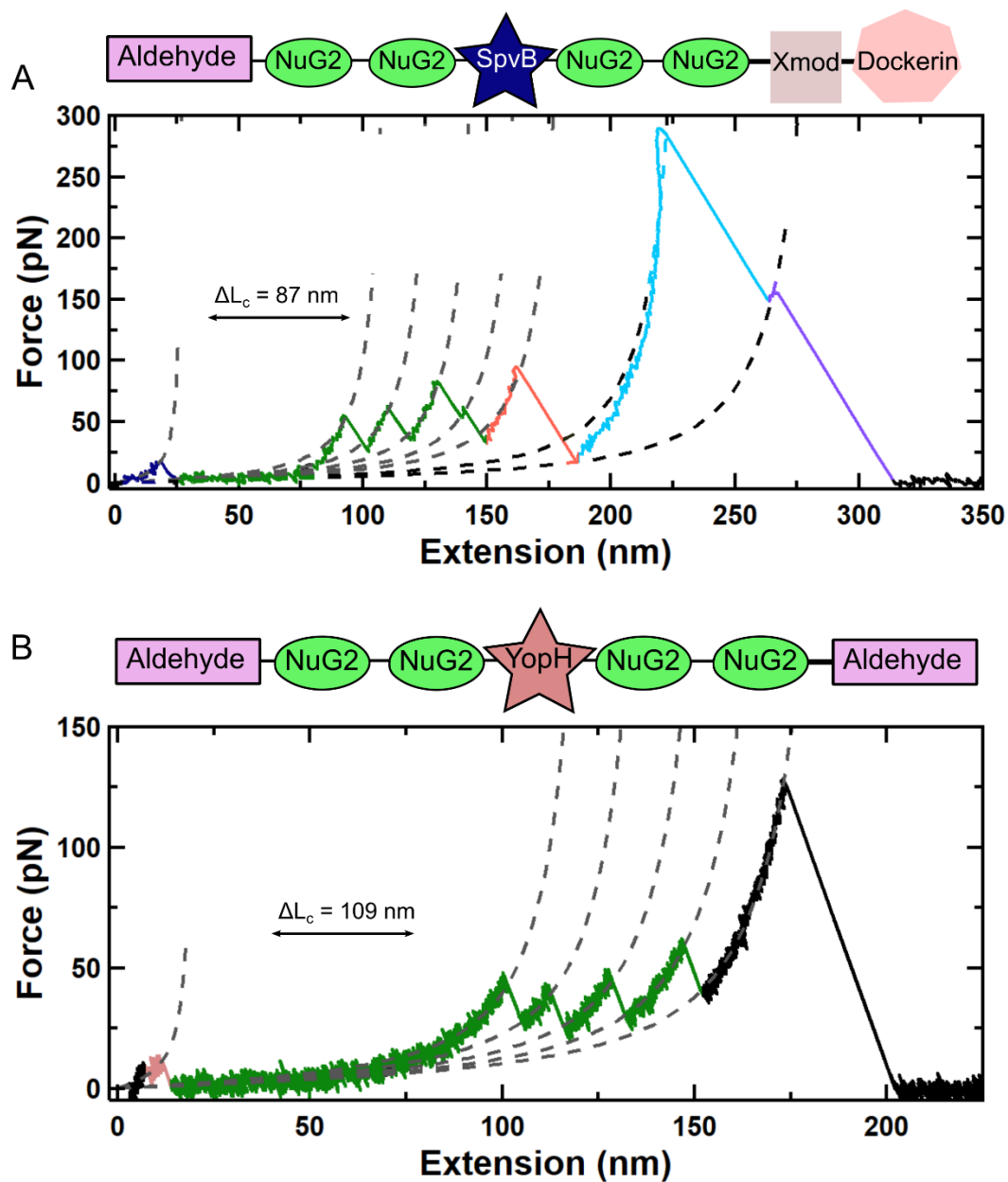
Figure C-1. Representative FEC of SpvB and YopH. (A) FEC of SpvB in the Ald-NUG2(4x)-Xmod-Dockerin III construct. The measured change in contour length is within 20% of the predicted value (87 nm measure, 69 nm predicted). (B) FEC of YopH in the Dual-Aldehyde construct. The measure change in contour length exactly matches the predicted value (109 nm for both). All traces collected at 50 kHz and smoothed to 1 kHz for presentation. (A) was measured at 800 nm/s retraction, (B) was measured at 50 nm/s retraction.

Methods

Cloning of SpvB and YopH

Only the catalytic core of SpvB and YopH were used to simplify data interpretation (SpvB 392-591, YopH 164-468). The genes were acquired through plasmid databases Addgene (SpvB – Plasmid #89455) and DNASU (YopH – Clone ID FLH149883.01X). SpvB 392-591 was amplified using primers U/L 2319 which introduced a C-terminal NcoI site for insertion into pMS 1479 [6xHis-SUMO-Ald-NuG2(4x)-Xmod-Dockerin III] opened EcoRV/NcoI, resulting in pMS 1511. YopH 164-468 was amplified with primers U/L 2321 which added a C-terminal NcoI site for insertion into pMS 1251 [6xHis-SUMO-Ald-NuG2(4x)-Ald] opened EcoRV/NcoI, resulting in pMS 1477.

Polyprotein Expression and Purification

Both polyproteins were expressed and purified exactly as detailed for polyprotein expression and purification in Chapter 2. Briefly, *E. Coli* BL-21 (DE3) cells transformed with the plasmids were grown in 1L culture to OD_{600}= 0.55 and induced with 0.2 mM IPTG before shaking overnight at 18°C. Cultures were spun down, resuspended in 10 mL/L Lysis Buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 1 mM PMSF) and lysed with a C3 Homogenizer. Lysate was clarified and loaded onto 3 mL Ni-NTA columns. Columns were washed with 5 CV Wash Buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 20 mM Imidazole) and eluted with 3 CV Elution Buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP, 250 mM Imidazole). Elutions were diluted in half with TBS + TCEP (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM TCEP) and dialyzed overnight into FGE Conversion Buffer (25 mM TEA, 10 mM NaCl, 5 mM BME) in the presence of ULP1.

In Vitro FGE Conversion

Purified *M. Tb* FGE was added to the purified protein at a 1:10 FGE:polyprotein ratio and incubated overnight at 18°C with shaking.

Oxyamine Labeling

Because these experiments were done before I received FMOC protected HIPS reagents, I used oxyamine-DBCO to label the SpvB polyprotein (single aldehyde), and both oxyamine-Biotin and oxyamine-DBCO to label the YopH polyprotein (Dual-Aldehyde). Therefore, I buffer exchanged the proteins into Oxyamine Labeling Buffer (25 mM sodium phosphate, pH 5.5, 150 mM Nacl, 2 mM TCEP) and added 10x molar concentration of the labels. For the Dual-Aldehyde labeling, I added both HIPS-Biotin and HIPS-DBCO at the same time. Reactions were incubated overnight at 37°C. Excess reagent was removed by running the proteins over an S75 analytical column equilibrated with AFM Buffer (25 mM sodium phosphate, pH 7.2, 150 mM Nacl, 2 mM TCEP).