CHAPTER 1.3

Real-time Detection of Thiol Chemistry in Single Proteins

EDWARD C. ECKELS*[a,b], DANIEL J. ECHELMAN[a,b], JAIME ANDRÉS RIVAS-PARDO[a] AND JULIO M. FERNÁNDEZ[a]

[a]Department of Biological Sciences, Columbia University, New York, NY 10027, USA; [b]Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University Medical Center, New York, NY 10032, USA
*E-mail: ece2117@columbia.edu

1.3.1   Introduction

The cell is exquisitely tuned to alter its functions and gene expression in response to changes in its redox environment. Initiation of signaling pathways thus requires proteins that are sensitive to redox conditions, which is usually achieved through strategic placement of cysteine residues within the protein. A major challenge for protein biochemists studying these pathways is to detect changes in the redox status in one or several cysteine residues in a protein in a reliable and time-resolved manner. Many techniques used to detect the redox state of a particular amino acid require treatment of the protein samples with chemical probes that react with side-chains of a certain oxidation state. The probes do not always react with absolute specificity for the side-chains, and there is always concern that the redox state of the side-chains may change during the isolation and treatment process. Single-molecule force spectroscopy (SMFS) provides a means for executing

controlled protein redox experiments, either chemically or enzymatically driven, with high-fidelity real-time readout. Cysteine chemistry is rapid and reversible, and is therefore well suited for studies by SMFS. Here we discuss how to detect Ångstrom-level changes in the conformations of proteins due to the formation or cleavage of intra- and intermolecular (thiol) bonds. These studies allow for direct observation of the scission of covalent bonds under massive forces, a biological phenomenon as old as the first time one animal sunk its teeth into another. To control the rate at which the molecular bonds rupture, SMFS utilizes cryptic bonds that have low solvent accessibility and require unfolding of the substrate in order for the redox reaction to occur. Because force can be used to denature only the substrate, it is possible to observe the effect of oxidoreductases on unfolded substrates without the need for thermal or chemical denaturation.

1.3.1.1 Single-molecule Pulling Instruments

Basic execution of a force spectroscopy experiment requires tethering of a single protein to a microscopic pulling device, commonly referred to as a force probe. Although there are many ways to apply force to proteins, the three most often used techniques are atomic force microscopy, magnetic tweezers and optical tweezers.[1] The atomic force microscope (AFM) is widely used for these experiments because of its high-throughput ability to probe thousands of individual proteins per day and the relative ease of the attachment to the force probe.[2–7] It will therefore be the central focus of the single-molecule assays described in this chapter. In an AFM (Figure 1.3.1A), a silicon nitride cantilever with a probe tip radius of tens of nanometers is used to pick up single proteins attached to a glass or gold-coated surface. The surface is mounted on a piezoelectric (piezo) actuator with nano- to picometer resolution[8] that is used to apply force by downward displacement of the piezo away from the cantilever. If a single protein is suspended between the cantilever and the surface, the cantilever bends under the applied load. A laser beam reflected off the tip of the cantilever allows the measurement of cantilever deflection with high accuracy. Cantilevers with spring constants of ~15 pN nm^{-1} are commonly used as force probes for protein pulling experiments, meaning that if the cantilever is bent downwards by just 1 nm, it will apply 15 pN of force to the protein. A major innovation for single-protein AFM studies was the development of a force-clamp, which rapidly positions the pulling surface so as to maintain constant deflection of the cantilever through electronic feedback.[9–11] The force clamp is to single-thiol redox chemistry what the voltage clamp[12,13] was to single-ion channel recordings – namely that the force clamp is absolutely indispensable for studying the kinetics, and hence the mechanism, of thiol cleavage and re-formation.

The geometry of a magnetic tweezers instrument (Figure 1.3.1B) is similar, in that the protein is suspended between a glass surface and the force probe, but in this case the probe is a superparamagnetic bead. An electromagnet or
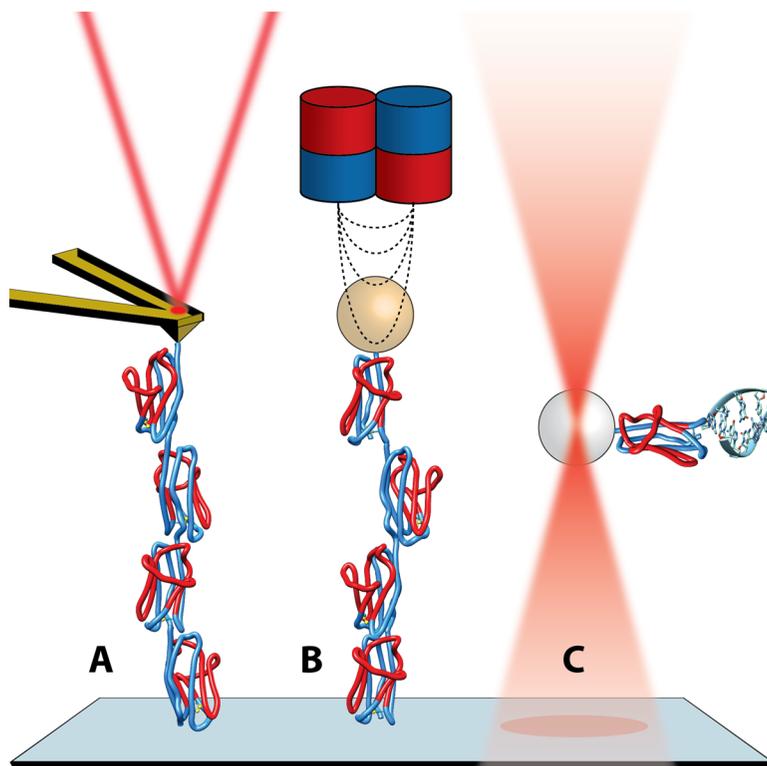
Figure 1.3.1 Single-molecule force spectroscopy probes the redox status of single disulfide bonds. The three most popular modes of single-molecule force spectroscopy (SMFS) utilize the atomic force microscope (A), magnetic tweezers (B) and optical tweezers (C). In the atomic force microscope (AFM), a single polyprotein of identical domains is suspended between a flexible microcantilever and a rigid glass surface. Force is applied to the protein by displacement of the glass surface by a piezoelectric actuator. The length of the protein is tracked by deflection of a tightly focused laser beam off the back of the cantilever. (B) In magnetic tweezers, a single polyprotein construct is tethered to a superparamagnetic bead through biotin–streptavidin linkage and to the glass through covalent HaloTag attachment chemistry. Force on the protein increases as a pair of permanent magnets approaches the fluid chamber. (C) Optical tweezers utilize the trapping of a dielectric microbead in a laser focus to apply force to proteins. Pulling on the protein through the DNA linkage displaces the microbead away from the focus of the laser, generating a restoring force that stretches the protein.

a pair of permanent magnets generates a magnetic force while the position of the bead is tracked with nanometer resolution using high-speed video-microscopy.[14] The main advantage of the magnetic tweezers is that a magnetic field provides a passive force clamp so that experiments operate under constant force without the need for electronic feedback. This permits folding

and unfolding reactions to be tracked at forces down to a few piconewtons and innovations in attachment chemistry have permitted pulling experiments lasting from hours to weeks.[14–16] Both magnetic and optical tweezers instruments are often constructed on top of inverted microscopes so that simultaneous fluorescence imaging during application of force can be readily achieved.[17–19] Magnetic tweezers have been used to track individual folding events of titin immunoglobulin (Ig) domains containing internal disulfide bonds.[20]

Optical tweezers utilize the trapping forces on dielectric microbeads in a tightly focused laser beam to apply forces to proteins (Figure 1.3.1C). The protein is suspended between two microbeads, often using long pieces of double-stranded DNA. In one arrangement, a single microbead is optically trapped while the other is glued to a glass micropipette mounted on a piezoelectric actuator. When the actuator pulls its bead away from the trap, the entire DNA–protein complex is under force. Small displacements of the microbead due to protein folding and unfolding are tracked with a position-sensitive photodiode placed in the path of the laser, after the microbead. In an alternative setup, the second bead is also in an optical trap that can be displaced towards or away from the first trapped bead to apply force. Optical tweezers suffer from the disadvantages of complex instrumentation and difficult to achieve force-clamp conditions.[21,22] Furthermore, the protein–DNA linkages are often achieved through disulfide bonds,[23] which is not suitable for redox experiments. Although some groups have found alternative attachment chemistries,[24] these drawbacks are serious enough that no significant studies of protein thiol chemistry have been achieved with optical tweezers. It is highly recommended to use the AFM for high-throughput studies or magnetic tweezers for low-force observation of folding trajectories in single-molecule thiol chemistry assays.

1.3.1.2 Mechanical Fingerprints Provide Unambiguous Identification of Single-protein Tethers
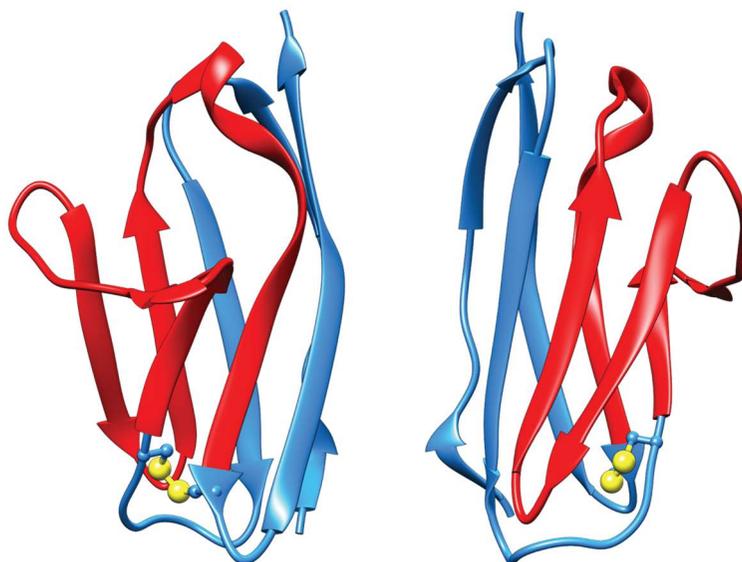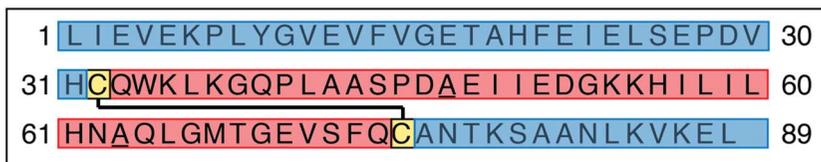
In SMFS, there is a high probability of having something other than a single protein molecule attach to the force probe. It is possible to have two or more molecules connected to the probe in parallel, or even aggregates of proteins. In the case of AFM and magnetic tweezers, there are also non-specific interactions between the force probe and the surface that can be alleviated with surface passivation but are difficult to avoid altogether. To overcome these problems, the concept of a "mechanical fingerprint" was developed. For AFM and magnetic tweezers, a single polypeptide consisting of several identical tandem protein domains is immobilized on the pulling surface (Figure 1.3.1A). Hence, when the protein is unfolded with force, there should be one unfolding event per domain. Typically, experiments are performed with protein constructs consisting of eight identical tandem domains.[25,26] The most commonly used protein domain is a titin Ig fold

containing a single engineered disulfide bond (Figure 1.3.2). Figure 1.3.3A shows a typical unfolding trace in the AFM using a construct consisting of eight tandem domains of I91 under force-clamp conditions. In the AFM, it is common to pick up proteins not exactly at their N-terminus so that fewer than eight domains are observed, but in Figure 1.3.3A all eight domains are unfolded during the initial force pulse (blue curve, Figure 1.3.3A). The more consecutive domains that are observed, the stronger is the "fingerprint" and the greater the experimental confidence in the identity of the molecule. Under force-clamp conditions, the end-to-end length of a protein follows a stepwise or staircase trajectory under constant force (Figure 1.3.3A). One of the most popular substrates for single-molecule pulling experiments is I91, the 91st immunoglobulin domain of the titin sequence, located in the elastic (I-band) portion of the sarcomere. This domain is also sometimes referred to as I27 for its position within the cardiac isoform of titin. The I91 domain consists of an Ig-like beta-sandwich fold with high mechanical resilience caused by a network of hydrogen bonds between the first (A–A′) and last (G) beta strands in the fold, often referred to as the "mechanical clamp" that withstands the pulling force in SMFS experiments.[27] When the mechanical clamp breaks due to the pulling force, the remainder of the I91 domain immediately loses its structure and extends into the conformation of a random polypeptide coil under force. The step size of unfolding events and the presence of all eight domains ensure that only a single polypeptide chain is being manipulated with the AFM cantilever. The step size of unfolding at any given force can be predicted using simple models of polymer elasticity such as the worm-like chain and the freely jointed chain if the number of amino acids participating in the fold is known.[28,29]

Real-time detection of redox chemistry depends on the ability of the SMFS instrument to measure the step size of unfolding and the force at which the unfolding occurs with high precision. As we shall see, introduction or cleavage of intramolecular bonds between side-chains of a protein domain will alter the unfolding step size. A variety of assays have been developed around this principle, which are detailed in the following sections. The majority of the chapter will focus on the cleavage and formation of disulfide bonds, but the techniques outlined here have also recently been extended to study disulfide isomerization, *S*-glutathionylation reactions and the formation and cleavage of thioester bonds.

1.3.2 Mechanochemical Cleavage of Cryptic Protein Disulfide Bonds

In order to study the force-dependent cleavage of disulfide bonds at a single-molecule level, the thiol residues participating in the bond must be placed strategically within the protein domain. Most single-molecule redox experiments are carried out with an engineered version of I91, but theoretically any protein that satisfies the requirements discussed here could be used

1	L I E V E K P L Y G V E V F V G E T A H F E I E L S E P D V	30
31	H C Q W K L K G Q P L A A S P D A E I I E D G K K H I L I L	60
61	H N A Q L G M T G E V S F Q C A N T K S A A N L K V K E L	89

Figure 1.3.2 The engineered I91$^{(32C/75C)}$ titin domain allows for direct monitoring of the (real-time) oxidation status of a disulfide bond within a protein. The well-characterized I91 (also known as I27 in the cardiac isoform) domain is a useful scaffold for introduction of engineered disulfide bonds. Wild-type I91 unfolds under force by breaking a series of hydrogen bonds between the first (A–A′) and last (G) beta strands in the sequence. The protein domain here is based on the crystal structure of wild-type I91 with several changes: the two native cysteine residues (C47 and C63) are mutated to alanine (underlined in the sequence A). Two new cysteine residues were introduced through the mutations G32C and A75C. The resulting domain is referred to as I91$^{(32C/75C)}$ throughout this chapter. When unfolding this molecule with force, only the residues in blue experience force when the disulfide bond is formed. All of the residues sequestered behind the disulfide bond (red) are protected from force and are stretched only if the disulfide bond is cleaved. I91$^{(32C/75C)}$ is often utilized in reduction and oxidative folding experiments because the disulfide bond is buried in the hydrophobic core of the protein and is solvent inaccessible. Therefore, the disulfide must be primed through mechanical unfolding of the surrounding residues before any reaction can occur. This structure is based on the crystal structure of the wild-type I91 domain (pdb 1WAA) with an energy-minimized disulfide bond introduced using protein structure visualization software.[77]
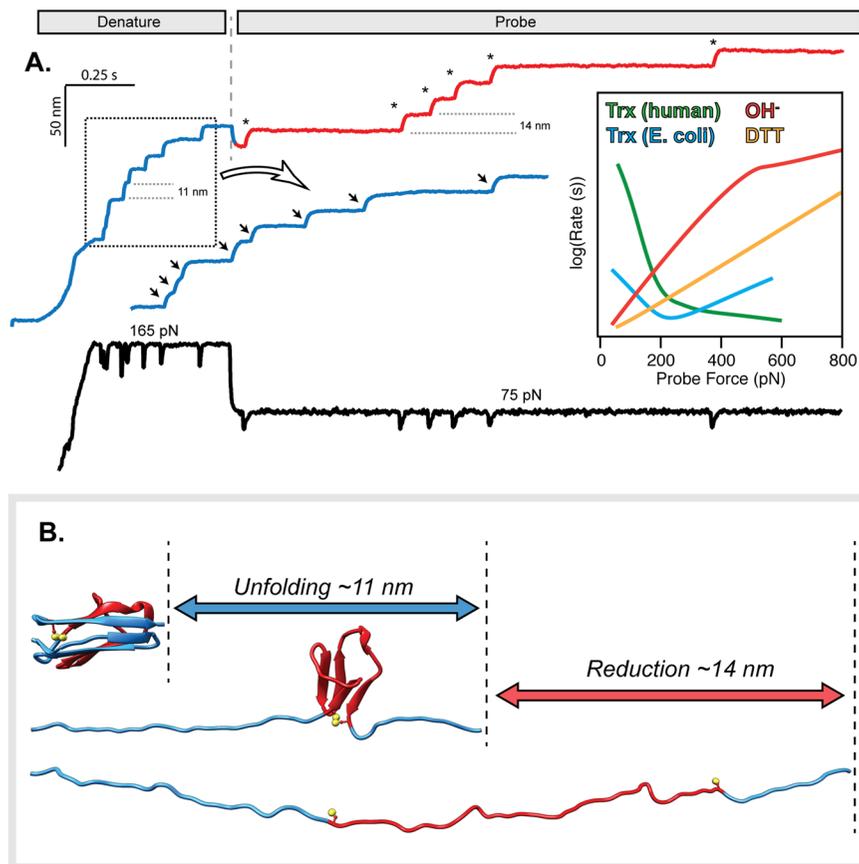
Figure 1.3.3 Detecting single-disulfide reduction events using the AFM in force-clamp mode. Trajectories of a polyprotein consisting of seven tandem domains of I91$^{(32C/75C)}$. (A) The protein is first unfolded during the *denature* pulse at a force of 165 pN. Each of the eight unfolding events (black arrows) results in an upwards 11 nm step caused by loss of structure of the amino acids in the blue segment (B and Figure 1.3.2). Immediately after the denature pulse, the force is changed to a new value for the probe pulse during which reduction of the disulfide will occur, here 75 pN. Six of the eight exposed disulfides are cleaved during this probe pulse (red trace). Each cleavage event (asterisk) results in a 14 nm increase in length as the remaining residues are unfolded. Inset: systematic variation of the probe force across a span of hundreds of piconewtons reveals the mechanism of disulfide cleavage by several nucleophiles. The force dependence of DTT (gold curve) shows a simple Bell-like dependence of the cleavage rate on force, and the hydroxyl ion (red curve) shows evidence of a force-induced switch above ~500 pN. Although the force dependence of both thioredoxins is negative in the low force regime, the cleavage rate of *E. coli* thioredoxin accelerates at higher forces whereas the rate of human thioredoxin does not.

as a substrate in redox experiments. To create a suitable substrate for single-molecule thiol chemistry, the sequence of I91 was altered through several rounds of mutagenesis to achieve a single, cryptic disulfide bond. First, the two cysteine residues native to I91 (C47 and C63) were mutated to alanine because they are too far apart to participate in disulfide bonding (underlined residues, Figure 1.3.2A). Next, several pairs of residues in close proximity in the crystal structure were mutated to cysteine. Four of the mutants containing completely different pairs of cysteines could be expressed and shown to contain internal disulfide bonds.[28] The most used of these mutants contains mutations G32C and A75C (yellow residues, Figure 1.3.2A), which will be referred to as I91$^{(32C/75C)}$. The sequence and structure of I91 and the location of the disulfide bond are shown in Figure 1.3.2B, with amino acid residues sequestered within the disulfide shown in red.

1.3.2.1 Disulfide Cleavage with Low Molecular Weight Thiols

The presence of this disulfide bond in I91$^{(32C/75C)}$ greatly shortens the observed increase in length during the unfolding transition. When the disulfide bond is formed, the protein unfolds from the mechanical clamp up to the disulfide bond. This generates a step of ~11 nm (blue residues, Figure 1.3.2A and B). The remaining residues are protected by the disulfide bond and are not released by the unfolding event. The force required to pull apart a disulfide bond is much greater than the forces used in SMFS experiments on proteins, so cleavage is not expected to occur spontaneously.[30] What is not immediately apparent from the structure depicted in Figure 1.3.2B is that this disulfide bond is *cryptic* – due to the positioning of the two cysteine residue side-chains, the thiols are not susceptible to nucleophilic attack when the I91 domain is folded. Some of the other engineered disulfide bonds have various degrees of solvent accessibility; for example, the disulfide bond in I91$^{(24C/55C)}$ shows complete reduction after ~90 min in 20 mM dithiothreitol (DTT), *without* mechanical or chemical unfolding of the protein.[31] By contrast, I91$^{(32C/75C)}$ requires protein unfolding for reduction of the disulfide bond. The very first single-molecule disulfide reduction experiments were carried out by Wiita *et al.* on an octamer protein of I91$^{(32C/75C)}$ under force-clamp conditions using DTT as a nucleophile.[32] Figure 1.3.3A depicts a typical "double-pulse" protocol employed to study the reduction reaction. First, the protein is stretched at an intermediate force, here at 165 pN, to unfold each of the domains up to the disulfide bond during the denature pulse (blue trace, Figure 1.3.3A). Each unfolding event results in an 11 nm step at this force. After the seven domains have unfolded, the force is quickly clamped to a new value during the "probe" pulse. During the probe pulse, DTT collides with the disulfide bond, reducing it and releasing the remaining contour length of the protein (red trace, Figure 1.3.3A). Each reduction event results in a 14 nm step increase in the length of the protein construct. A schematic of the configurations of a single I91$^{(32C/75C)}$ domain during the denature/probe protocol is shown in Figure 1.3.3B. In I91 domains that are reduced prior

to unfolding, the expected step size during the "denature" pulse would be 25 nm, simply the sum of the two observed steps.

These experiments and several others utilizing different small organic reducing agents are useful for understanding the physical chemistry and mechanism of disulfide bond cleavage. Although the disulfide bond cannot be broken with force in these experiments, it was discovered that the load applied to the disulfide can alter the bond geometry in such a way that facilitates nucleophilic attack of the thiols and cleavage of the disulfide bond. In the presence of low molecular weight thiol compounds, the rate of cleavage of the disulfide bond was found to accelerate exponentially with the applied force, increasing 10-fold over a range of 300 pN. By fitting the force dependence of the reaction with a Bell-like model, it was found that the distance to the transition state ($\Delta x^{\ddagger}$) for the thiol–disulfide exchange is 0.34 Å,[32] meaning that the disulfide bond lengthens by this amount when crossing the thermodynamic barrier during the reaction. Studies with various reducing agents, including glutathione hydrochloride, 2-mercaptoethanol, 1,4-dithioerythritol, L-cysteine, tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and tris(hydroxypropyl)phosphine, showed that the bond elongation during the thiol–disulfide exchange reaction is highly dependent on the nucleophile, with distances to the transition state ranging from 0.23 to 0.46 Å.[33]

1.3.2.2 Two Pathways for Disulfide Cleavage by the Hydroxyl Ion

Force can lower the energy barrier of disulfide cleavage by straining the disulfide bond, but further increasing the force can cause unexpected rearrangements of the polypeptide backbone. Utilizing non thiol-based nucleophiles, Garcia-Manyes *et al.* were able to study the reduction of I91[(32C/75C)] at extremely high forces.[34] Unlike the reducing agents DTT, 2-mercaptoethanol and glutathione, which interfere with the gold–thiol attachment chemistry used to anchor proteins to the pulling surface, the hydroxyl ion (HO⁻) allowed disulfide cleavage to be observed at forces up to 2 nN. These experiments showed that there were two regimes in the rate of cleavage of the disulfide bond as a function of force. At low forces, the rate of nucleophilic attack increased according to a Bell-like model with a $\Delta x^{\ddagger}$ value of 0.5 Å. However, at 500 pN, there is an abrupt switch to a regime where the cleavage rate is less sensitive to force and has a $\Delta x^{\ddagger}$ value of only 0.1 Å. There are two proposed mechanisms in this work: there may be two separate barriers, an inner and an outer barrier, for the total cleavage of the disulfide bond. The first barrier would require a weak complex to form between the hydroxyl and the two thiols, elongating the bond by 0.1 Å. The actual cleavage requires further lengthening of the bond by an additional 0.4 Å. Although this view is consistent with the data, it is very different from the known S_N2 reaction mechanism, which proceeds through a single concerted step. Another explanation is that the peptide backbone undergoes a conformational switch at high forces. Whereas the CSSC dihedral angle

is ~90° for unconstrained disulfides, it is conceivable that application of high forces drives the disulfide into a *trans* conformation that increases the distance between the alpha-carbons participating in the disulfide bond. This reduces the energy of the disulfide bond, making it less susceptible to cleavage than at forces below which the conformational transition occurs. Others have suggested that scission of the C–S bond is responsible for one of the two regimes.[35] Although these forces are likely not relevant on physiological scales, such experiments demonstrate how force spectroscopy can be used to probe reaction pathways. Furthermore, reaction of thiols with oxygen compounds such as the hydroxyl ion can lead to non-enzymatic oxidative folding, and this will be discussed later.

1.3.2.3 Rules for Single-protein Thiol Chemistry Experiments

For all of these experiments and for all the experiments discussed below, it is crucial that the disulfide bond being studied is cryptic. This permits the rate of the reduction reaction to be determined with high accuracy because the reaction, for all practical reasons, starts only when the mechanical clamp breaks and the protein unfolds, exposing the disulfide. If the disulfide was completely solvent accessible, all of the disulfide bonds would become reduced before the cantilever had a chance to pull on them, beginning when the reductant is added to the buffer.[31] This would make it very difficult to study the rate and mechanism of disulfide cleavage. Alegre-Cebollada *et al.* provided a useful summary of the rules[36] that should be followed when planning single-molecule studies for the chemical cleavage of covalent bonds using force spectroscopy, paraphrased as follows. (i) The disulfide bond should be contained within a host protein domain that provides a distinct fingerprint upon unfolding. Ideally, the protein should be expressed as multiple repeats along a single polypeptide chain to provide an unambiguous signature for a single tether. (ii) The disulfide bond should be cryptic or at least have limited ability to react with the nucleophile/enzyme while the host protein domain is folded. (iii) Unfolding of the host protein domain should cause exposure of the disulfide bond to the solvent. (iv) Cleavage of the disulfide bond by the nucleophile or enzyme should cause a conformational change in the host protein, such as a step change in length, that is easily detectable. Although it is possible to study the presence or absence of intermolecular disulfide bonds that are not cryptic using SMFS, detailed analysis of reaction rates to extract sub-ångstrom parameters such as the distance to the transition state requires the bonds to be cryptic.

1.3.3 Enzymatic Cleavage of Protein Disulfide Bonds

1.3.3.1 Single-molecule Reduction by Thioredoxins

Although the cytosol is highly reducing due to millimolar amounts of reduced glutathione, there are a variety of enzymes that participate in the enzymatic cleavage of disulfide bonds and participate in redox control. Thioredoxin (Trx)

is an ancient and well-conserved cytosolic enzyme that acts as an antioxidant by participating in thiol–disulfide exchange with protein substrates.[37] The thioredoxin fold is shared among an entire class of proteins related to maintenance of redox both inside and outside the cell, including human protein disulfide isomerase (PDI), bacterial DsbA, glutaredoxin, glutathione *S*-transferase and glutathione peroxidase.[38] All of these enzymes contain a catalytic site CXXC motif that executes the thiol–disulfide exchange reaction with substrate proteins. Although the overall activity of oxidoreductases can be measured with various bulk assays – spectrophotometrically (*e.g.* Ellman's reagent,[39] NADPH oxidation or tryptophan fluorescence[40]) or using light scattering (*e.g.* insulin turbidity assays[41]) – these experiments lack the sensitivity needed for probing the chemical mechanism of catalysis.

In many ways, cleavage of disulfides by thioredoxin is similar to the reaction with the small nucleophiles described above, but the protein fold of thioredoxin is optimized to position the catalytic residues in such a way that further accelerates the reaction. Reduction of the disulfide bond in I91[(32C/75C)] by *Escherichia coli* Trx was carried out in the same manner as for the small nucleophiles, except that a regeneration system consisting of thioredoxin reductase, NADPH and FAD was also present to convert any oxidized Trx back to its reduced form.[42] As had been demonstrated with bulk techniques and other substrates, it was found that Trx cleaved the disulfide bond in I91[(32C/75C)] nearly 30 000 times faster than DTT. Micromolar amounts of Trx were sufficient to observe complete reduction of all eight I91[(32C/75C)] domains in the polyprotein construct, in contrast with the millimolar concentrations needed for cleavage by DTT. However, the single-molecule assay provided further insight into the mechanism of the thiol–disulfide exchange reaction. To study the force dependence of the reduction reaction, the "double-pulse" force protocol was again employed (Figure 1.3.3A). After unfolding each of the eight I91 domains to expose the disulfide bond at a force of 165 pN, the force clamp was quickly adjusted to a value between 50 and 600 pN. Whereas for small nucleophiles such as DTT the applied force only accelerated the reaction, the Trx reaction rate showed a biphasic nature (light–blue curve, Figure 1.3.3A inset). Below 200 pN, the reduction reaction rate had a negative dependence on the force, meaning that any increase in the pulling force slowed the reaction. Above 200 pN, the reaction was accelerated in the same way that it was for the small nucleophilic molecules. This indicates two competing pathways for the thiol–disulfide exchange reaction. Fitting the reaction with a simple Bell-like model demonstrated that under one cleavage mechanism the disulfide bond must lengthen by 0.2 Å, but under the competing pathway the substrate must shorten by 0.8 Å. How might such a shortening of the substrate occur?

This force-dependent inhibition of the thiol–disulfide exchange reaction can be understood by looking at a solution NMR structure of thioredoxin bound to a short substrate polypeptide through a disulfide bond (pdb 1MDI).[43] In this structure, the angle between the disulfide bond and the

major axis of the peptide binding groove in Trx is ~70°. However, the nucleophilic attack of a thiol on a disulfide bond occurs through an S_N2 reaction that requires alignment of all three sulfur atoms at 180°, due to orbital geometry. Thus, when Trx binds to I91$^{(32C/75C)}$ and encounters the disulfide bond, the bond must undergo rotation, driven by thermal motion, within the binding groove in order for the reaction to occur. Increasing the force on the substrate containing the disulfide bond prevents the bond from "wiggling" into the correct conformation for the reaction to proceed. The negative distance to the transition state of 0.8 Å represents the distance that the peptide must shorten within the binding groove, projected onto the pulling axis. Above 200 pN, it is likely that the peptide is so extended by force that it no longer fits into the peptide groove. Then cleavage of the disulfide occurs when the catalytic thiol of Trx collides with the disulfide bond without actually binding to the peptide substrate. Further evidence that these two competing pathways are mechanistically distinct came from studies of Trx mutants P34H and G47S, which affect the descending limb of the force dependence of the cleavage but not the ascending limb.[37,42] Analysis of the dwell times between exposure of the disulfide bond and cleavage by Trx also demonstrated that the rate was best fitted with a double exponential, suggesting two competing pathways.[44] In contrast, reduction of the disulfide bond with phosphine-based reducing agents (*e.g.* TCEP) was best fitted with only a single exponential.

1.3.3.2 Prokaryotic and Eukaryotic Thioredoxins Demonstrate Alternative Mechanisms

The reduction of disulfides by thioredoxin at forces exceeding 300 pN requires that both the thiols of I91$^{(32C/75C)}$ and the oxidoreductase be solvent exposed. Solvent exposure of the catalytic cysteine residues in thioredoxin is dependent on the width and depth of the catalytic groove. If the catalytic thiols are buried deep within the groove, this will preclude reduction of disulfide-bonded substrates under hundreds of piconewtons of load.[42] Perez-Jimenez *et al.* observed a signature of evolution of the thioredoxin groove by measuring the catalytic rates of eight different thioredoxins from four different kingdoms of life using single-molecule force-clamp spectroscopy.[45] Thioredoxins of prokaryotic origin, such as *E. coli* Trx used by Wiita *et al.*,[42] all demonstrated a biphasic reduction rate with respect to the pulling force as described above. Alternatively, thioredoxins of eukaryotic origin such as human Trx1 and *Plasmodium falciparum* Trx1 showed a decreasing reaction rate up to 200 pN, at which the rate plateaued to low values (inset, Figure 1.3.3A). This is speculated to be caused by a deepening of the thioredoxin groove during the evolution of Eukarya, preventing the S_N2 like mechanism of cleavage at higher forces. Interestingly, thioredoxins from eukaryotic organisms but contained in cellular compartments of prokaryotic origin (*e.g.* pea chloroplast Trx and human mitochondrial Trx2)

showed the signature of a shallow prokaryotic groove with reaction rates increasing again at high forces.

1.3.4   Inferring Molecular Pathways of Oxidative Folding

Although cleavage of disulfide bonds by Trx undoubtedly plays an important role in reversing the undesirable effects of oxidative stress in the cytosol, a more interesting question is how disulfide bonds can be introduced into proteins, especially disulfide bonds that are located within the hydrophobic core of the protein. It is generally thought that oxidoreductase enzymes interact with nascent polypeptide chains to catalyze simultaneously the formation of disulfide bonds and folding of the substrate into its native structure.[46] Many of the oxidoreductases interact with unfolded polypeptide substrates as they emerge from translocon pore complexes into, for example, the endoplasmic reticulum (ER) or periplasm.[47] The assays described above can be adapted to study the effects of such enzymes.

1.3.4.1   A Single-molecule Assay for Oxidative Folding by Protein Disulfide Isomerase

A consequence of the single-molecule reduction assays described above is that once the disulfide has been cleaved at high force, the two cysteine residues of I91$^{(32C/75C)}$ are too far apart to be brought into a disulfide bond. A new force pulse assay was developed to lower the force on the peptide substrate to allow for reoxidation of the disulfide bond (Figure 1.3.4A). After full unfolding of the polyprotein to ascertain a single molecule attachment, the tip of the cantilever can be quickly approached back to the surface to which the protein is attached. By slightly pushing the cantilever against the surface, the effective end-to-end distance of the protein is reduced to close to zero and the force on the protein is zero. Although I91 is known to fold back readily to its native structure when left at zero force for several seconds, the disulfide will not re-form without thiol–disulfide exchange, which can be mediated by oxidoreductase enzymes.

Protein disulfide isomerase is an enzyme known to catalyze the efficient formation of disulfide bonds and the rearrangement of incorrectly formed disulfides that may impair the folding process or lead to aggregation. PDI is found in millimolar amounts in the ER, where it interacts with nascent polypeptide chains passing from the ribosome through the translocon pore.[48] These polypeptide chains must be in an extended and unfolded conformation to pass through the pore and PDI assists correct folding and placement of disulfide bonds when the polypeptide emerges into the ER. SMFS provides a perfect means to mirror the environment of a nascent polypeptide chain interacting with the PDI machinery, which must first form a mixed disulfide to engage its substrate. In an experimental scenario, this can be achieved
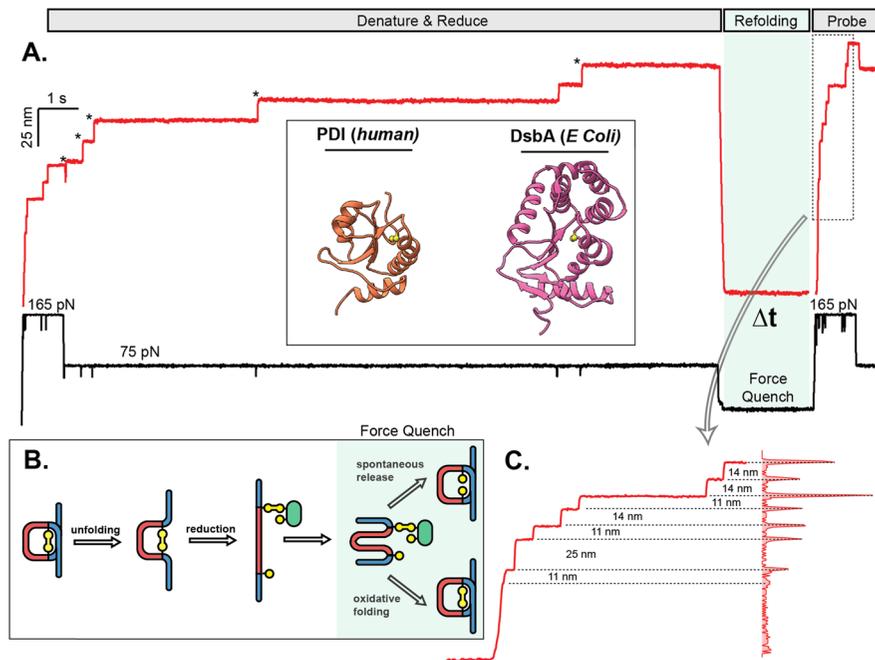
Figure 1.3.4 Reoxidation of single disulfide bonds occurs during protein collapse and folding. (A) A single-molecule force spectroscopy assay for oxidative folding utilizing oxidized I91$^{(32C/75C)}$ and reduced A-domain of human protein disulfide isomerase (PDI, green oval) and the resulting single-molecule trajectory. During the denature pulse at 165 pN, each of the eight I91 domains unfolds (11 nm), exposing a disulfide for attack by PDI. When the reduced PDI attacks the disulfide, the protein domain extends by an additional 14 nm (asterisks), and the PDI remains attached to the I91 substrate through a mixed disulfide (B). After cleavage of all eight disulfide bonds, the force on the substrate peptide is quenched (light-green panel), and the polypeptide chain collapses. In the collapsed state, the I91 disulfide may reoxidize if the free thiol of I91 attacks the mixed disulfide, or the I91 disulfide may remain reduced if the free thiol of PDI attacks the mixed disulfide leading to spontaneous release (A and B). During the probe pulse, the force is increased again to 165 pN, allowing for the counting of reoxidized (11 + 14 nm) steps and reduced (25 nm) steps from spontaneous release. (C) Kinetics of oxidative folding and spontaneous release by PDI are determined by measuring the amplitude of the populations of the oxidized and reduced steps while varying the duration of the denature plus refolding pulse.

in two ways: a unit of oxidized PDI can react with a reduced cysteine side-chain in its substrate, or a unit of reduced PDI can form a mixed disulfide by reacting with an oxidized substrate. The purpose of these experiments is to understand how the disulfide bonds are introduced into the protein during the folding process, so how the mixed disulfide is achieved is irrelevant.

Even though PDI is maintained in its oxidized state within the ER, it is more practical to carry out the SMFS experiments using oxidized I91[(32C/75C)] and reduced PDI. This is the strategy employed by Kosuri *et al.*[48] in the experiments described below. Furthermore, only the oxidoreductase-active A1 domain of PDI was utilized in these experiments, so all references to PDI hereafter refer to only that domain (Figure 1.3.4A).

To measure the ability of PDI to catalyze the re-formation of protein disulfide bonds, Kosuri *et al.* utilized two unfolding pulses with an intervening refolding period ("denature–refolding–probe," Figure 1.3.4A). During the denature pulse, eight domains of I91 containing the oxidized disulfide bond are quickly unfolded at a pulling force of 165 pN (denature pulse) in ~11 nm steps. Subsequently lowering the force to 75 pN allows efficient cleavage of the disulfide bond (asterisks, Figure 1.3.4A) by reduced PDI. In contrast with the thioredoxin assays, the force on the protein is then quickly quenched to zero by bringing the tip into contact with the surface *while PDI is still attached to I91 through the mixed disulfide* (Figure 1.3.4B), and is held there for a short duration Δt (refolding pulse). During this time, the substrate begins the conformational search for its folded native state. After the Δt time has elapsed, the protein is quickly extended at 165 pN again. By counting the relative number of reduced I91 domains (25 nm) and reoxidized I91 domains (11 + 14 nm steps), one can measure the catalytic efficiency of PDI for re-formation of disulfide bonds (Figure 1.3.4C).

What is happening to the protein substrate during the refolding pulse? When the force on the reduced I91 molecule is lowered to zero, the polypeptide backbone collapses and the cysteine of I91 that was freed during the thiol–disulfide exchange is brought into close proximity to the mixed disulfide with PDI ("force quench," Figure 1.3.4B). If the time for refolding Δt is long enough, then the free cysteine can attack the mixed disulfide and release PDI. This would result in re-formation of the I91[(32C/75C)] mixed disulfide. Another possibility for the course of events is that the other thiol in PDI, which is in a reduced state, can attack the mixed disulfide, leading to spontaneous release of the oxidoreductase from the I91 polypeptide. In the latter scenario, the oxidation occurs at the disulfide of PDI, leaving the I91 cysteines in their reduced form, hence the 25 nm steps. Alternatively, if neither the free thiol of I91 nor the free thiol of PDI attacks, leaving the mixed disulfide intact, the I91 will also remain reduced. On the probe pulse, evidence for both oxidative folding (re-formed 11 nm steps) and spontaneous release of PDI (25 nm steps) can be observed (Figure 1.3.4C).[48]

To distinguish between spontaneous release of the oxidoreductase and slow cleavage of the mixed disulfide, a mutant form (C35S) of the previously studied Trx was used. Whereas the mixed disulfide is slow to resolve in PDI, it was well known that wild-type Trx undergoes very rapid spontaneous release of its substrates through attack of its other catalytic thiol on the

mixed disulfide. Indeed, experiments utilizing the "denature–refolding–probe" pulse with wild-type Trx showed only reduced (25 nm) steps on the probe pulse due to the fast release.[48] However, mutant Trx[(C35S)] is prevented from undergoing spontaneous release from its substrate, hence altering its redox capabilities. The mixed disulfide between Trx[(C35S)] and I91[(32C/75C)] can only be resolved through attack of the free thiol on I91. Experiments using Trx[(C35S)] demonstrated full reoxidation of the I91 disulfide bond due to the presence of 11 nm, but not 25 nm, steps during the probe pulse. This indicates that the 25 nm steps observed in the presence of PDI are not due to failure of the disulfide to reoxidize, but rather from spontaneous release of PDI from its substrate. By varying the duration of the denature and refolding pulse, the kinetics of disulfide reoxidation by PDI and spontaneous release of PDI from its substrate were found to be approximately equal.

These experiments measured for the first time the coordination of folding and introduction of disulfide bonds into proteins. The implications of the mechanism of oxidative folding by PDI are striking and unexpected. Experiments performed in the absence of PDI demonstrated that oxidized I91[(32C/75C)] domains fold 30 times faster than when the disulfide bond is reduced (7.74 *versus* 0.27 s^{-1}). This is expected because the disulfide bond reduces the conformational space that the protein domain must search before it finds its minimum energy native state, an effect that has been observed in many other proteins.[49,50] In the presence of PDI, the I91 domains are found to fold at a rate almost identical with that of reduced I91[(32C/75C)], suggesting that acquisition of the native fold is the rate-limiting step during oxidative folding of I91[(32C/75C)] and must precede formation of the disulfide bond.

It is remarkable that I91[(32C/75C)] can fold even while PDI remains attached through a mixed disulfide. Recall that the two cysteine residues of I91[(32C/75C)] are buried, so it would be expected that attachment of the entire PDI domain to I91 would hinder folding. If anything, PDI marginally increases the speed of folding compared with reduced I91[(32C/75C)]. It appears that PDI is specially adapted to perform oxidative folding in this manner. If PDI could only introduce disulfide bonds while its reduced substrate was in its nascent or denatured state, there would be a high chance of introducing non-native disulfides, which Kosuri *et al.* showed would lead to a misfolded, unfunctional substrate.[48] By permitting folding to occur first, cysteines closest in the native state have the greatest chance of forming a disulfide. All members of the Trx family have a short hydrophobic groove surrounding the catalytic site,[45] supposedly for assisting substrate recognition. This groove is known to have chaperone activity independent of the oxidoreductase activity[51] and may play an important role in guiding the substrate to its native fold despite the presence of the mixed disulfide. PDI and other oxidoreductases are capable of identifying unfolded substrates that contain disulfides through this groove and form mixed disulfides to act as a placeholder to prevent incorrect disulfide formation throughout the folding process.

1.3.4.2   DsbA is a High-efficiency Bacterial Oxidoreductase

A variety of other Trx homologs are found in all the branches of life. Oxidative folding is especially important for the secretory system of bacteria, which must introduce disulfide bonds into proteins when they emerge from the bacterial translocon pore. *E. coli* DsbA is an oxidoreductase that is part of the Dsb system, which is composed of several proteins responsible for proper folding of periplasmic proteins.[52] Although DsbA null mutations are not lethal, they lead to loss of extracellular structures such as adhesive pili and flagella.[53] Like PDI, DsbA is a member of the Trx family of folds, so it shares a similar topology and CXXC motif. The only difference in the global structure compared with Trx is the insertion of a helical domain after the second beta strand in DsbA (Figure 1.3.4A).[38] Interestingly, the CXXC redox-active motif also varies among these oxidoreductases: Trx (CGPC), PDI (CGHC) and DsbA (CPHC). The structure–function relationships are responsible for the large differences in redox potentials for each of these three enzymes, with Trx as the most reducing (−271 mV), PDI at an intermediate potential (−160 mV) and DsbA the most oxidizing (−120 mV).[54] These differences are manifested directly at the single-molecule level during oxidative folding experiments in the AFM.

Using the same denature–refolding–probe pulse protocol employed for PDI, Kahn *et al.* used reduced DsbA to cleave and re-form the intramolecular disulfide bonds of I91$^{(32C/75C)}$.[55] This allowed a direct comparison between the oxidative folding mechanisms of DsbA, PDI and Trx. The major finding from these experiments was that the folding rate of I91 under coordination of DsbA was three times faster than that of PDI, but still slower than the folding of oxidized I91$^{(32C/75C)}$ domains in the absence of DsbA. This suggests again that DsbA introduces the disulfide bond shortly after folding but does so more quickly than PDI. Hence protein folding drives disulfide bond formation during oxidative folding under control of both of these oxidoreductases. The spontaneous release rate of DsbA was also found to be much slower than that of PDI, which itself is much slower than that of Trx. These findings demonstrate that the release rate is inversely proportional to the redox potential of the oxidoreductase and partially explains why DsbA is such an efficient catalyst of oxidative folding. The slower release rate would also increase the fidelity of the disulfide bond formation, especially for slow-folding substrates.

These experiments provide further evidence for the chaperone activity of the thioredoxin fold. Using observation of ~25 nm reduced steps as evidence for spontaneous release, the mechanism of release can be inferred by changing the Δt of the denature plus refolding time. Not only were the kinetics of DsbA release much slower than those of PDI, but also the absolute *amplitude* of the observed fraction of reduced domains was smaller. With PDI, longer wait times Δt lead to greater numbers of reduced domains, approaching 84% at infinite time. With DsbA, however, only 44% of the domains will be reduced during the probe pulse at infinite time. One explanation for this

difference is that even after spontaneous release of DsbA from its mixed disulfide, there are non-covalent interactions that prevent diffusion of DsbA away from the substrate. This allows DsbA to re-form its mixed disulfide with the reduced I91$^{(32C/75C)}$ and gives it another chance to allow the free thiol of I91 to attack the mixed disulfide, lowering the total fraction of reduced domains during long quench durations Δt. This dynamic equilibrium between DsbA and its substrate could only arise if the "off" rate of DsbA from the folded substrate is much slower than the rate of thiol–disulfide exchange. Although the affinity of DsbA for the folded substrate was not directly measured, this observed behavior is characteristic of chaperones, which tend to bind to and stabilize the folded state. From these experiments, it seems that the catalytic groove of these enzymes controls the redox equilibrium of the oxidoreductase both by changing the electronic environment (pK_a) of the catalytic thiols and by adjusting the affinity for the unfolded and folded substrates.[55]

1.3.4.3 Non-enzymatic Oxidative Folding

Although the cellular redox environment is buffered with large amounts of antioxidants such as glutathione, which is maintained in the reduced state through various enzymatic pathways, large transient increases in oxidative stress can generate a variety of reactive oxygen species (ROS) that can modify protein thiols. The thiol of cysteine can undergo several rounds of oxidation by ROS. The first oxidation reaction converts the reduced thiol (–SH) to sulfenic acid (–SOH). The second and third oxidations convert sulfenic acid to sulfinic acid (–SO$_2$H) and sulfonic acid (–SO$_3$H). Of these reactions, only the conversion of the reduced thiol to sulfenic acid is thought to be reversible.[56] Such oxidations can alter the charge and sterics of the cysteine side-chain and may thus influence the folding rate or the stability of the protein domain. However, if the cysteine undergoing oxidation by the ROS can form a disulfide bond with a neighboring residue, there are different consequences for protein folding.

Although the hydroxyl radical, generated by the decomposition of hydrogen peroxide, is often used to initiate oxidation of cysteine side-chains, it is difficult to use it in single-molecule experiments. The radical is promiscuous, leading to oxidation of side-chains other than cysteine, and very reactive, causing cysteine to reach its higher irreversible oxidation states more quickly. Building on the previously discussed studies of disulfide cleavage by the hydroxyl ion (OH$^-$), Beedle *et al.* used this nucleophile to explore non-enzymatic reoxidation of protein disulfide bonds.[57] Instead of using oxidized I91$^{(32C/75C)}$ as the substrate, the engineered I91$^{(24C/55C)}$ was used in these studies. This disulfide bond cannot be accessed by hydroxyl ions when the protein is in its folded state as all disulfides seem to be formed during the initial denature pulse. To accelerate the rate of cleavage of the disulfide, the hydroxyl ion concentration was increased by raising the solution pH to 12.8 which is equivalent to a concentration of 63 mM hydroxyl ion.

These assays again utilize the usual "denature–refolding–probe" force protocol used for the previous oxidative folding experiments (Figure 1.3.4A). The disulfide bonds in this particular engineered I91 domain result first in a ~15 nm step from unfolding up to the disulfide bond and then a ~10 nm step after cleavage of the disulfide by the hydroxyl anion due to the different placement of the engineered cysteines. After the force quench to permit refolding, the hydroxyl-cleaved disulfides demonstrated re-formation of the disulfide bonds through reappearance of the ~15 nm steps in the probe pulse. Similarly to PDI and DsbA, the hydroxyl ion acts as a placeholder for the cleaved disulfide bond by formation of an *S*-sulfenylated (–SOH) intermediate.[57] When the force on the protein is quenched during the refolding pulse, the free thiol of I91 can attack the sulfenic moiety *via* an S_N2 reaction, releasing the hydroxyl ion back into solution and re-forming the disulfide bond. However, the –SOH group is short lived and can progress to the higher thiol oxidation states of sulfinic and sulfonic acids. Experiments performed with long denature pulses to prolong the exposure of the thiols to solution demonstrated a loss of reoxidation of the disulfide bond but also a loss of the ability of the I91 domain to refold at all. This indicates that (a) the higher oxidation states of cysteine are not reversible under attack by the free thiol of I91 and (b) the sterics or electrostatics created by the additional oxygen atoms on the sulfur are unfavorable enough to prevent refolding of the I91 domain. To prove the existence of the *S*-sulfenylated intermediate, Beedle *et al.* performed experiments at pH 12.8 in the presence of 30 mM dimedone, which reacts irreversibly with sulfenic acid, but not any other oxidation state of the cysteine side-chain.[57] The bulky dimedone side-group blocked protein folding of I91$^{(24C/55C)}$ in single-molecule experiments and its presence was also confirmed using mass spectrometry. There are several important differences between non-enzymatic oxidative folding and what is seen with the oxidoreductase enzymes. There is no way for the sulfenic acid to release the hydroxyl ion spontaneously, so there are never any ~25 nm steps observed in the probe pulse. Additionally, even for the shortest denature pulses that allowed for cleavage of the disulfide bond by hydroxyl but minimized exposure to the oxidizing environment, the maximum re-formation of the disulfide bond only approached 55%.

1.3.5   Conformational Changes and Allostery Through Cysteine Redox and Disulfide Isomerization

Formation and cleavage of intramolecular disulfide bonds are only one mechanism by which a protein might change form or function in response to altered cellular redox conditions. There are a variety of other intra- and intermolecular bonds that might form with protein thiols that have profound effects. Altered protein stabilities and conformations through disulfide bond isomerization, *S*-glutathionylation and thioesterification are discussed here.

1.3.5.1 Disulfide Isomerization in Superoxide Dismutase 1 (SOD1)

SMFS experiments are especially apt for distinguishing rearrangements of disulfides within proteins. Intramolecular disulfide isomerization occurs when a single-protein domain contains three or more cysteine residues in close proximity. Depending on the geometry and local electronic environment of the cysteine side-chains, various numbers of different disulfide bonds could possibly form. This was first realized by Alegre-Cebollada *et al.* using an engineered I91 domain containing four cysteine residues:[58] I91$^{(24C/32C/55C/75C)}$.[58] After unfolding this protein in the presence of L-cysteine as a nucleophile, four distinct reduction step sizes were observed. This indicates that disulfide bonds other than those present in I91$^{(32C/75C)}$ and I91$^{(24C/55C)}$ had formed during folding. The authors further demonstrated that when a single reduced thiol was sequestered in the loop behind a formed disulfide exposed under force, it preferred to attack one thiol of the disulfide over the other. These experiments demonstrate that disulfide isomerization could occur transiently and spontaneously within proteins, leading to changes in protein topology that might affect function. The observed regioselectivity for attacking one thiol over the other is speculated to be related to the pK_a of the leaving thiol in the exchange reaction, assuming that thiolates with lower pK_a values are better leaving groups.[58]

In the above experiments, force is used as a switch to "uncage" a single internal cysteine residue so that it can participate in thiol–disulfide exchange. However, small changes in protein structure or alterations in cysteine pK_a due to mutations can also induce disulfide isomerization. SOD1 is an enzyme important for the degradation of superoxide anion ($O_2^{\cdot-}$) into less toxic ROS, and changes in the redox state of SOD1 are associated with the onset of the neurodegenerative disease amyotrophic lateral sclerosis (ALS). SOD1 contains four cysteines: conserved Cys57 and Cys146 form a disulfide bond required for proper function, whereas Cys6 and Cys111 are involved in the formation of non-native disulfide bonds that may predispose SOD1 to aggregation, a pathological hallmark of ALS. Solsona *et al.* found that two disease-causing mutations (G39A and A4V) alter the reactivity and disulfide bonding patterns of SOD1 using single-molecule force-clamp assays.[59] First, it was clear from single-disulfide cleavage experiments by TCEP on wild-type SOD1 that the Cys57–Cys146 disulfide bond is buried, as cleavage occurred only after unfolding, generating a characteristic increase in protein length upon reduction. After unfolding, single isomerization events due to attack of Cys111 on the disulfide were observed. The G39A mutation greatly increased the regiospecificity of the isomerization reaction, making attack at Cys146 and formation of the Cys111–Cys146 disulfide 2.5 times more likely than in the wild-type SOD1.[59] Experiments on the A4V mutant did not show reduction or isomerization steps. This suggests that the A4V mutation prevents any disulfide from forming, or alternatively induces isomerization of the disulfide to a solvent-accessible state, rendering it cleavable by TCEP even before

unfolding. These experiments highlight the sensitivity of thiol chemistry to the surrounding amino acid environment.

1.3.5.2   *S*-Glutathionylation of Cryptic Thiols Regulates Muscle Elasticity

Post-translational modification (PTM) of thiols is a well-known regulator of protein function. Because the intracellular pool of reduced glutathione is in the millimolar range, overwhelming the cellular redox machinery with oxidative stress can generate a large pool of oxidized glutathione. This can happen especially during events such as ischemia/reperfusion injury in the aftermath of myocardial infarction. In one study of a mouse model of myocardial infarction, an ultra-high molecular weight protein bearing glutathione was seen in proteins isolated from the infarcted cardiac tissue.[60] Although no confirmatory experiment was performed, a protein of such a size was assumed to be titin, which can reach over 3 MDa in size in some muscle isoforms. Indeed, the elastic I-band region of titin contains a disproportionately high number of cysteine residues and is thought to be redox sensitive. Using the wild-type form of I91, which contains cysteines at positions 47 and 63, Alegre-Cebollada *et al.* demonstrated that these residues could undergo *S*-glutathionylation only after unfolding of the I91 domain.[25] Furthermore, using single-cysteine mutagenesis, it was found that *S*-glutathionylation at position Cys63 lowered the mechanical stability of the I91 fold while modification of Cys47 blocked folding (Figure 1.3.5). Blockage of folding was found to be reversible in the presence of glutaredoxin, which removes the glutathione from the cysteine residue. Interestingly, these experiments could also be performed at the tissue level in single stretched human cardiomyocytes.[25] Stretching cardiomyocytes transmits piconewton-level forces to titin, unfolding many domains and exposing cryptic thiols. When the cardiomyocytes were soaked in oxidized glutathione for 30 min while being stretched, they lost their elastic recoil, presumably owing to the inability of individual titin Ig domains to refold. This effect could be reversed by exchanging the buffer for reduced glutathione or DTT, a signature of thiol–disulfide exchange chemistry. Hence the mechanical properties of cardiac and skeletal muscle are directly regulated by the redox status of titin Ig domains.

### 1.3.5.3   A Protein Thioester Responsible for Bacterial Adhesion

Cysteines are also commonly found in a second type of *inter*- and *intra*-molecular covalent bond as a component of thioester bonds. Here, the cysteine thiol conjugates with a carbonyl to yield a high-energy bond of the form R–S–CO–R′ (Figure 1.3.6A). Often the cysteine originates in the enzyme and the carbonyl originates in the substrate. The carbonyl is contributed
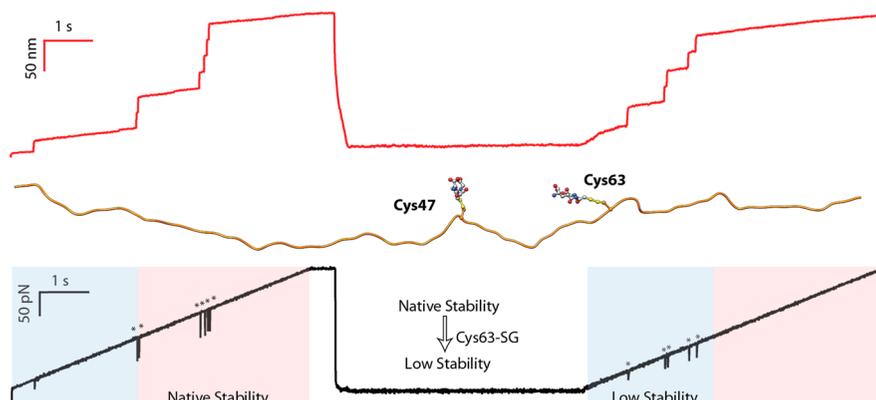
Figure 1.3.5 *S*-Glutathionylation of cryptic thiols alters the stability of the native I91 fold. Top: force-ramp unfolding of a mutant $(I91^{C47A})_8$ polyprotein in the presence of soluble oxidized glutathione (GSSG). Each protein domain has only a single cysteine at position 63 that can be oxidatively modified by glutathione. After each unfolding step of ~25 nm, Cys63 is exposed to the solution and can react with GSSG. Bottom: *S*-glutathionylation of the thiol is detected during the second force ramp: each domain containing a modified thiol unfolds at a lower force, indicating destabilization of the native fold by the presence of a covalently bound glutathione molecule. During the second ramp, five out of six domains refolded, and all five have low stability, suggesting post-translational modification by glutathione. Inset (center): *S*-glutathionylation at Cys47 and Cys63 together blocks folding, and utilizing a mutant $(I91^{C63A})_8$ it was shown that *S*-glutathionylation of Cys47 alone is sufficient to block folding, whereas *S*-glutathionylation of Cys63 alone only weakens the mechanical stability of the fold.

by the polypeptide backbone in E1/E2/E3-mediated ubiquitinylation[61] and sortase-mediated pilus assembly,[62] where the thioester bond links the enzyme–polyprotein complex in these examples of anabolic metabolism. Alternatively, the active-site thiol of a cysteine protease catabolizes proteins *via* a thioester intermediate that is then hydrolyzed by water. Significantly, whereas thioesters react readily within the enzyme active site, hydrolysis is otherwise slow in solution, with a half-life reaching tens to hundreds of days.[63]

Perhaps the most notable thioester in biology is between an acetyl group and coenzyme A (CoA, also abbreviated as CoASH to emphasize its terminal thiol group), which is integral to multiple pathways in central metabolism, including the citric acid cycle and fatty acid metabolism.[64] The thioester bond is highly exergonic, with a free energy of hydrolysis comparable to that of the ATP pyrophosphate bond and is therefore useful in coupling to less energetically favorable reactions.[65,66] These two features – the high energy of hydrolysis and the centrality of acetyl-CoA in metabolism have led to a "thioester world" hypothesis for prebiotic life, wherein thioester bonds would function as a proto-ATP.[64,66,67]
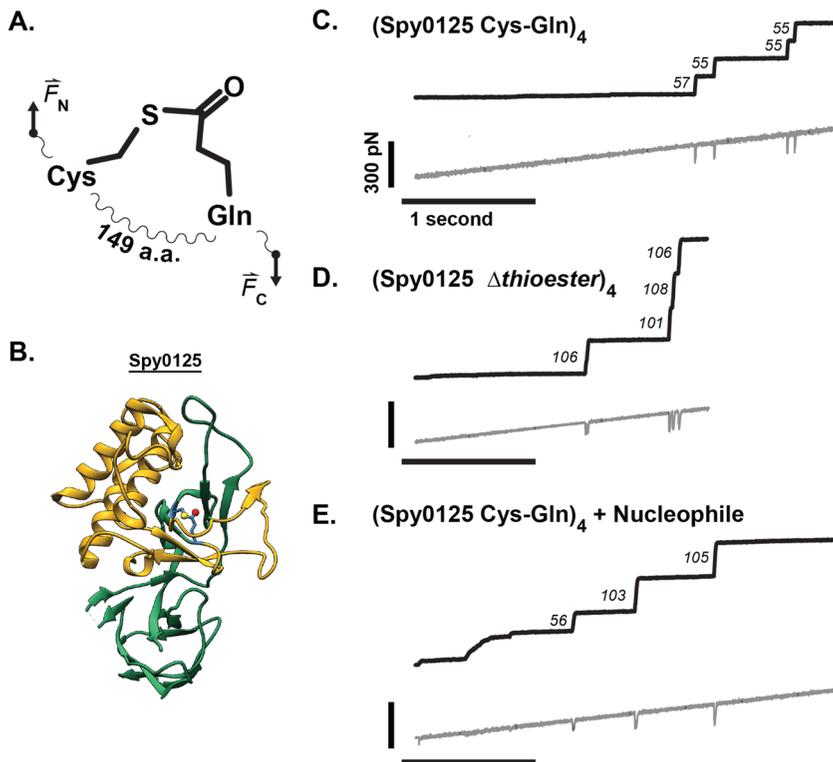
Figure 1.3.6   Resolving the Cys–Gln thioester bond with force spectroscopy. (A) Chemical structure of a cysteine–glutamine thioester bond. In the bacterial adhesin Spy0125, 149 amino acids separate the participating cysteine and glutamine residues. (B) Protein structure of Spy0125 as used by Echelman *et al.*[76] in their thioester force spectroscopy assay. Residues protected by the thioester bond from mechanical unfolding are colored gold and the remaining residues are colored green. The side-chains of the cysteine and glutamine participating in the bond are highlighted in the structure. (C) Force-ramp unfolding of the Spy0125 polyprotein showing repeated stepwise extensions of ~55 nm, consistent with the presence of formed thioester bonds. (D) With a Cys to Ala mutation to remove the thioester bond, the Spy0125 tetramer unfolds with ~107 nm steps. (E) Force-ramp unfolding of the same construct as in (C), but with the addition of a nucleophile (*e.g.* methylamine), allows the study of chemical cleavage at the level of a single thioester bond.

Meanwhile, *intra*molecular thioester bonds are known to occur in three biological systems: in Gram-positive bacterial adhesins, in α_2-macroglobulin (α2M) antiproteases, and in the immune complement proteins C3 and C4.[68–71] In each instance, the thioester bond forms between a cysteine thiol and the side-chain carbonyl of either a glutamine or a glutamate residue. The result is a high-energy bond that can react with nucleophilic ligands to form an *inter*molecular covalent bond.

Each system directs its thioester chemistry towards a specific end. Gram-positive bacterial adhesins place the thioester bond at the distal tip of their adhesive structures to react with and form stable covalent linkages with host ligands, such as fibrinogen of the extracellular matrix.[70] In an example of convergent evolution, the C3 and C4 complement proteins of the innate immune system use the thioester to covalently crosslink exposed sugars or amines on bacterial surfaces.[68] The thioester bond of α2M reacts with the lysine side-chain of a protease, causing inactivation.[69,71]

In the α2M and C3/C4 complement examples, specificity of the thioester bond towards nucleophilic ligands is determined by a proteolytic gating mechanism.[71,72] The thioester bond is guarded within a hydrophobic pocket until a site-specific proteolytic cleavage exposes the bond to solution and to reactions with neighboring nucleophiles. In C3, after cleavage, the thioester reaction proceeds with a half-life on the order of microseconds.[73] Independent of cleavage, there is a low baseline level of C3 thioester hydrolysis or cleavage by amines that exist in equilibrium with thioester bond re-formation.[74] In C3, this equilibrium can be measured by size-exclusion chromatography, as C3 undergoes large-scale conformational rearrangements upon thioester cleavage.[74,75]

A similar equilibrium was recently determined in the Gram-positive bacterial adhesion Spy0125 (Figure 1.3.6B) using AFM-based single-molecule force spectroscopy.[76] In Spy0125, the thioester-forming cysteine and glutamine side-chains are located 149 residues apart, giving a ~59 nm difference in contour length (L_C) between when the bond is there and when it is not (assuming 0.4 nm per residue[28]). This difference in contour lengths produces two distinct fingerprints with force-ramp spectroscopy. Force-ramp unfolding of a Spy0125 polyprotein with its internal Cys–Gln thioester bond intact yields regular ~55 nm steps (Figure 1.3.6C). Removal of the thioester bond by a cysteine to alanine mutation shifts the step size to ~107 nm (Figure 1.3.6D). With these two fingerprints predefined, the thioester bond can be assayed for reactivity to different nucleophiles and under various conditions of pH and concentration.[76] For example, Figure 1.3.6E shows force ramp of the Spy0125 polyprotein in the presence of a small molecule nucleophile, methylamine. The first module unfolds as a 56 nm step, indicating a formed thioester bond, whereas the two following modules unfold with 103 and 106 nm steps, indicating thioester bonds that have been cleaved by the attacking nucleophile before mechanical unfolding.

This single-molecule assay elucidated two critical aspects of the bacterial adhesin's thioester bond. First, Echelman *et al.* observed that mechanical unfolding of the Spy0125 adhesin inhibits thioester reactivity to nucleophiles.[76] In a repeat of the force-clamp reduction assays for the disulfide bond, Spy0125 was unfolded on an initial force pulse and then held at constant forces of 50–350 pN. However, subsequent steps from the nucleophilic cleavage of the exposed thioester bond were not observed, whereas cleavage was readily observed when the protein was folded and at zero force. This differential reactivity between 0 and >50 pN should invite future research into

thioester chemistry in the low-force range, which may reveal a negative force dependence of the reaction rates.

In addition, the thioester bond is seen to re-form spontaneously using the single-molecule force spectroscopy assay. In the presence of a nucleophile, the fraction of ~107 nm steps rises to a plateau due to thioester cleavage. When the nucleophile is subsequently removed, the population returns to a baseline of ~55 nm steps as the thioester bond re-forms.[76] Notably, this re-formation proceeds with an estimated half-life of <7 min, in contrast to the complement protein C3, which has a half-life for thioester re-formation of 1.5 h.[74] The spontaneous thioester bond formation observed in complement and in bacterial adhesins is paradoxical, given that the formation of thioester bonds in solution is energetically unfavorable, on the order of +7 kcal mol^{-1}.[66] Future work based on single-molecule assays may unravel the energetics of thioester bond formation within a folding protein.

1.3.6 Conclusion

Rending apart tissue with a sharp, pointed object concentrates enormous forces upon a small surface area that may contain only a few chemical bonds. The spines of the sea urchin or porcupine, the sting of a bee, the claws of a mantis, the teeth of a grizzly bear, or the arrow of a hunter's bow all penetrate tissue tearing apart molecular bonds. The bonds that are not broken by the force, however, may be strained and primed for chemical reaction as described here. Because nanonewton-level forces are needed to tear apart covalent bonds, it is likely that many peptide chains at the site of tearing sustain a sizeable amount of force. These environments are recreated in single-molecule force spectroscopy experiments. Following the rules laid out for the study of cryptic thiol bonds allows an enormous variety of nucleophiles, from enzymes to organic and even inorganic electron donors, to interact with disulfide bonds under force. The dynamics of the reduction reaction can be inferred by measuring the force dependence of the reaction rate. These studies show that except in the case of the simplest low molecular weight thiol compounds, a simple Bell-like model is inappropriate for describing the reaction rate. Force alters not only the energy of the bond being studied, but also the sterics of the reaction or interaction with enzyme catalysts. Although I91$^{(32C/75C)}$ and its relatives are used in these single-molecule studies, protein substrates containing native cryptic disulfide bonds are abundant and may demonstrate different dynamics under force compared with the studies described here. Going forward, there is much work to be done. Oxidoreductases, transferases, hydrolases, lyases, isomerases and protein ligases all have the possibility of being studied with single-molecule resolution. Many more post-translational modifications of protein thiols and other amino acid side-chains also exist and may have the ability to alter protein dynamics. Cysteine remains the most interesting of all the amino acids for the broad range of adducts that it can form with common biological molecules. The mechanism

by which nitric oxide, arguably one of the most important signaling molecules in biology, exerts its effects on protein structure and function are still poorly understood and widely debated, but perhaps could be studied using these techniques. With carefully designed experiments, single-molecule force spectroscopy will continue to unravel the delicate interplay between protein form and function.

References

1. K. C. Neuman and A. Nagy, *Nat. Methods*, 2008, **5**, 491–505.
2. M. Carrion-Vazquez, P. E. Marszalek, A. F. Oberhauser and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 11288–11292.
3. M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3694–3699.
4. H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6527–6531.
5. H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek and J. M. Fernandez, *Nat. Struct. Mol. Biol.*, 2000, **7**, 1117–1120.
6. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109–1112.
7. H. Li, W. A. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkvliet, H. Lu, P. E. Marszalek and J. M. Fernandez, *Nature*, 2002, **418**, 998–1002.
8. I. Popa, P. Kosuri, J. Alegre-Cebollada, S. Garcia-Manyes and J. M. Fernandez, *Nat. Protoc.*, 2013, **8**, 1261–1276.
9. A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 468–472.
10. J. M. Fernandez and H. Li, *Science*, 2004, **303**, 1674–1678.
11. M. Schlierf, H. Li and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7299–7304.
12. E. Neher and B. Sakmann, *Nature*, 1976, **260**, 799–802.
13. B. Sakmann and E. Neher, *Annu. Rev. Physiol.*, 1984, **46**, 455–472.
14. I. Popa, J. A. Rivas-Pardo, E. C. Eckels, D. J. Echelman, C. L. Badilla, J. Valle-Orero and J. M. Fernández, *J. Am. Chem. Soc.*, 2016, **138**, 10546–10553.
15. H. Chen, G. Yuan, R. S. Winardhi, M. Yao, I. Popa, J. M. Fernandez and J. Yan, *J. Am. Chem. Soc.*, 2015, **137**, 3540–3546.
16. G. Yuan, S. Le, M. Yao, H. Qian, X. Zhou, J. Yan and H. Chen, *Angew. Chem., Int. Ed.*, 2017, **56**, 5490–5493.
17. R. Liu, S. Garcia-Manyes, A. Sarkar, C. L. Badilla and J. M. Fernández, *Biophys. J.*, 2009, **96**, 3810–3821.
18. A. Sarkar, R. B. Robertson and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12882–12886.
19. F. E. Kemmerich, M. Swoboda, D. J. Kauert, M. S. Grieb, S. Hahn, F. W. Schwarz, R. Seidel and M. Schlierf, *Nano Lett.*, 2016, **16**, 381–386.
20. J. A. R. Pardo, E. Eckels, J. Valle-Orero and J. M. Fernadez, *Biophys. J.*, 2017, **112**, 456a.

21. W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri and S. M. Block, *Phys. Rev. Lett.*, 2005, **95**, 208102.

22. M. J. Lang, C. L. Asbury, J. W. Shaevitz and S. M. Block, *Biophys. J.*, 2002, **83**, 491–501.

23. J. C. M. Gebhardt, T. Bornschlogl and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2013–2018.

24. M.-E. Aubin-Tam, A. O. Olivares, R. T. Sauer, T. A. Baker and M. J. Lang, *Cell*, 2011, **145**, 257–267.

25. J. Alegre-Cebollada, P. Kosuri, D. Giganti, E. Eckels, J. A. Rivas-Pardo, N. Hamdani, C. M. Warren, R. J. Solaro, W. A. Linke and J. M. Fernández, *Cell*, 2014, **156**, 1235–1246.

26. J. A. Rivas-Pardo, E. C. Eckels, I. Popa, P. Kosuri, W. A. Linke and J. M. Fernández, *Cell Rep.*, 2016, **14**, 1339–1347.

27. M. Sotomayor and K. Schulten, *Science*, 2007, **316**, 1144–1148.

28. S. R. K. Ainavarapu, J. Brujić, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li and J. M. Fernandez, *Biophys. J.*, 2007, **92**, 225–233.

29. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, 1953.

30. M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann and H. E. Gaub, *Science*, 1999, **283**, 1727–1730.

31. S. R. K. Ainavarapu, A. P. Wiita, H. H. Huang and J. M. Fernandez, *J. Am. Chem. Soc.*, 2008, **130**, 436–437.

32. A. P. Wiita, S. R. K. Ainavarapu, H. H. Huang and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 7222–7227.

33. S. R. Koti Ainavarapu, A. P. Wiita, L. Dougan, E. Uggerud and J. M. Fernandez, *J. Am. Chem. Soc.*, 2008, **130**, 6479–6487.

34. S. Garcia-Manyes, J. Liang, R. Szoszkiewicz, T.-L. Kuo and J. M. Fernández, *Nat. Chem.*, 2009, **1**, 236–242.

35. P. Dopieralski, J. Ribas-Arino, P. Anjukandi, M. Krupicka, J. Kiss and D. Marx, *Nat. Chem.*, 2013, **5**, 685–691.

36. J. Alegre-Cebollada, R. Perez-Jimenez, P. Kosuri and J. M. Fernandez, *J. Biol. Chem.*, 2010, **285**, 18961–18966.

37. R. Perez-Jimenez, A. P. Wiita, D. Rodriguez-Larrea, P. Kosuri, J. A. Gavira, J. M. Sanchez-Ruiz and J. M. Fernandez, *J. Biol. Chem.*, 2008, **283**, 27121–27129.

38. J. L. Martin, *Structure*, 1995, **3**, 245–250.

39. T. Tamura and T. C. Stadtman, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 1006–1011.

40. A. Holmgren, *J. Biol. Chem.*, 1972, **247**, 1992–1998.

41. A. Holmgren, *J. Biol. Chem.*, 1979, **254**, 9627–9632.

42. A. P. Wiita, R. Perez-Jimenez, K. A. Walther, F. Gräter, B. J. Berne, A. Holmgren, J. M. Sanchez-Ruiz and J. M. Fernandez, *Nature*, 2007, **450**, 124–127.

43. J. Qin, G. M. Clore, W. P. Kennedy, J. R. Huth and A. M. Gronenborn, *Structure*, 1995, **3**, 289–297.

44. R. Szoszkiewicz, S. R. K. Ainavarapu, A. P. Wiita, R. Perez-Jimenez, J. M. Sanchez-Ruiz and J. M. Fernandez, *Langmuir*, 2008, **24**, 1356–1364.

45. R. Perez-Jimenez, J. Li, P. Kosuri, I. Sanchez-Romero, A. P. Wiita, D. Rodriguez-Larrea, A. Chueca, A. Holmgren, A. Miranda-Vizuete, K. Becker, S.-H. Cho, J. Beckwith, E. Gelhaye, J. P. Jacquot, E. Gaucher, J. M. Sanchez-Ruiz, B. J. Berne and J. M. Fernandez, *Nat. Struct. Mol. Biol.*, 2009, **16**, 890–896.

46. B. Wilkinson and H. F. Gilbert, *Biochim. Biophys. Acta, Proteins Proteomics*, 2004, **1699**, 35–44.

47. M. Molinari and A. Helenius, *Nature*, 1999, **402**, 90–93.

48. P. Kosuri, J. Alegre-Ceballada, J. Feng, A. Kaplan, A. Inglés-Prieto, C. L. Badilla, B. R. Stockwell, J. M. Sanchez-Ruiz, A. Holmgren and J. M. Fernández, *Cell*, 2012, **151**, 794–806.

49. K. Chakraborty, M. Chatila, J. Sinha, Q. Shi, B. C. Poschner, M. Sikor, G. Jiang, D. C. Lamb, F. U. Hartl and M. Hayer-Hartl, *Cell*, 2010, **142**, 112–122.

50. C. B. Anfinsen and others, *Science*, 1973, **181**, 223–230.

51. R. Kern, A. Malki, A. Holmgren and G. Richarme, *Biochem. J.*, 2003, **371**, 965–972.

52. H. Kadokura, H. Tian, T. Zander, J. C. A. Bardwell and J. Beckwith, *Science*, 2004, **303**, 534–537.

53. F. E. Dailey and H. C. Berg, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 1043–1047.

54. G. Ren, D. Stephan, Z. Xu, Y. Zheng, D. Tang, R. S. Harrison, M. Kurz, R. Jarrott, S. R. Shouldice, A. Hiniker, J. L. Martin, B. Heras and J. C. A. Bardwell, *J. Biol. Chem.*, 2009, **284**, 10150–10159.

55. T. B. Kahn, J. M. Fernández and R. Perez-Jimenez, *J. Biol. Chem.*, 2015, **290**, 14518–14527.

56. C. E. Paulsen and K. S. Carroll, *Chem. Rev.*, 2013, **113**, 4633–4679.

57. A. E. M. Beedle, A. Lezamiz, G. Stirnemann and S. Garcia-Manyes, *Nat. Commun.*, 2015, **6**, 7894.

58. J. Alegre-Cebollada, P. Kosuri, J. A. Rivas-Pardo and J. M. Fernández, *Nat. Chem.*, 2011, **3**, 882–887.

59. C. Solsona, T. B. Kahn, C. L. Badilla, C. Álvarez-Zaldiernas, J. Blasi, J. M. Fernandez and J. Alegre-Cebollada, *J. Biol. Chem.*, 2014, **289**, 26722–26732.

60. B. S. Avner, K. M. Shioura, S. B. Scruggs, M. Grachoff, D. L. Geenen, D. L. Helseth, M. Farjah, P. H. Goldspink and R. John Solaro, *Mol. Cell. Biochem.*, 2012, **363**, 203–215.

61. M. Hochstrasser, *Nature*, 2009, **458**, 422–429.

62. H. Ton-That, L. A. Marraffini and O. Schneewind, *Biochim. Biophys. Acta*, 2004, **1694**, 269–278.

63. P. J. Bracher, P. W. Snyder, B. R. Bohall and G. M. Whitesides, *Origins Life Evol. Biospheres*, 2011, **41**, 399–412.

64. J. E. Goldford, H. Hartman, T. F. Smith and D. Segre, *Cell*, 2017, **168**, 1126–1134 e9.

65. R. K. Thauer, K. Jungermann and K. Decker, *Bacteriol. Rev.*, 1977, **41**, 100–180.

66. C. D. de Duve, *Blueprint for a Cell: The Nature and Origin of Life*, Pergamon Press, 1991.

67. S. N. Semenov, L. J. Kraft, A. Ainla, M. Zhao, M. Baghbanzadeh, V. E. Campbell, K. Kang, J. M. Fox and G. M. Whitesides, *Nature*, 2016, **537**, 656–660.

68. S. K. Law and A. W. Dodds, *Protein Sci.*, 1997, **6**, 263–274.

69. L. Sottrup-Jensen, H. F. Hansen, H. S. Pedersen and L. Kristensen, *J. Biol. Chem.*, 1990, **265**, 17727–17737.

70. M. Walden, J. M. Edwards, A. M. Dziewulska, R. Bergmann, G. Saalbach, S. Y. Kan, O. K. Miller, M. Weckener, R. J. Jackson, S. L. Shirran, C. H. Botting, G. J. Florence, M. Rohde, M. J. Banfield and U. Schwarz-Linek, *eLife*, **4**, e06638.

71. S. G. Wong and A. Dessen, *Nat. Commun.*, 2014, **5**, 4917.

72. H. J. Muller-Eberhard, *Annu. Rev. Biochem.*, 1988, **57**, 321–347.

73. R. B. Sim, T. M. Twose, D. S. Paterson and E. Sim, *Biochem. J.*, 1981, **193**, 115–127.

74. M. K. Pangburn, *J. Biol. Chem.*, 1992, **267**, 8584–8590.

75. N. Nishida, T. Walz and T. A. Springer, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19737–19742.

76. D. J. Echelman, A. Q. Lee and J. M. Fernández, *J. Biol. Chem.*, 2017, **292**, 8988–8997.

77. UCSF Chimera—A visualization system for exploratory research and analysis - Pettersen - 2004-Journal of Computational Chemistry - Wiley Online Library, http://onlinelibrary.wiley.com/doi/10.1002/jcc.20084/abstract, accessed 20 June 2017.