Tandem Repeating Modular Proteins Avoid Aggregation in Single Molecule Force Spectroscopy Experiments†

Lorna Dougan* and Julio M. Fernandez

Department of Biological Sciences, Columbia University, New York, New York 10027

Received: July 31, 2007; In Final Form: October 12, 2007

We have used single molecule force spectroscopy to explore the unfolding and refolding behavior of the immunoglobulin-like I27 protein in aqueous 2,2,2-trifluoroethanol (TFE). In bulk solution experiments, a 28% v/v TFE solution has previously been observed to enhance intermolecular attractions and lead to misfolding and aggregation of tandem modular proteins of high sequence identity. In our single molecule experiments, however, we measure successful refolding of the polyprotein $I27_8$ in all TFE solutions up to 35% v/v. Using a single molecule micromanipulation technique, we have shown that refolding of a polyprotein with identical repeats is not hindered by the presence of this cosolvent. These experimental results provide new insight into the properties of tandem repeating proteins and raise interesting questions as to the evolutionary success of such proteins in avoiding misfolding and aggregation.

Introduction

In living systems, molecular structures have the ability to self-assemble with high precision. An example of such self-assembly is the folding of proteins into their three-dimensional structures. Protein folding allows key functional groups to be brought into close proximity and allows the function of biological processes. Failure of a protein to fold correctly or to remain correctly folded is the origin of a wide variety of diseases.[1−6] Some of these diseases, such as cystic fibrosis and some cancers, result from proteins folding incorrectly and not being able to exercise their proper function. In other cases, proteins with a high propensity to misfold escape all protective mechanisms and form aggregates within cells or in extracellular space. For example, Alzheimer's, Huntington's, and Parkinson's disease are directly associated with the deposition of such aggregates in tissues in the human body.[1−3] Each of these neurodegenerative diseases is associated with abnormalities in the folding of a particular protein. However, it is striking that the molecular pathways leading to misfolding and aggregation of proteins, and the mechanisms by which these processes lead to disease, are similar.[1] Therefore, understanding the mechanisms that control misfolding of a protein and the initial formation of aggregates is an important issue in protein folding and for developing a strategy to treat these neurodegenerative illnesses.

Indeed, only correctly folded proteins have long-term stability in a crowded biological environment, allowing them to fulfill their function. In the human body, a large fraction of proteins have structural and mechanical function. A common feature of these mechanical proteins is that they contain multiple individually folded protein domains. Multidomain proteins, containing more than one domain, account for greater than 70% of all eukaryotic proteins.[42] Furthermore, tandem repeats of protein domains exist in up to 20% of the proteins in the genome. These tandem repeating proteins present an interesting case in which misfolding has been avoided in a crowded environment that contains similar domains.

One important example of a tandem modular protein is the immunoglobulin-like domain that is found in a variety of proteins. The giant protein titin is composed of around 300 repeats of immunoglobulin-like domains and fibronectin type III (FnIII) domains (Figure 1) and spans half of the muscle sarcomere.[7] Titin plays a number of important roles in muscle contraction and elasticity, as well as controlling chromosome shape in the cell nucleus.[12] The region of titin located in the sarcomere I-band is believed to be responsible for titin extensibility and passive elasticity. The half I-band is composed of distinct regions, referred to as the proximal immunoglobulin-like region, N2B and PEVK region, and distal immunoglobulin-like region (Figure 1). These regions include a head-to-tail linear array of immunoglobulin-like domains interrupted at intervals by less highly structured linker N2B sequences and PEVK sequences (70% proline, glutamic acid, valine, and lysine residues). The I27 protein, using the old nomenclature of Labeit and Kolmerer,[7] is located in the distal immunoglobulin-like region. This protein's secondary structure consists of eight β-strands that are stabilized primarily by hydrogen bonds between amide and carbonyl groups of the main chain.

In recent years, single molecule force spectroscopy has made it possible to explore the energy landscape of proteins such as I27 by applying a mechanical force.[8] In a mechanical unfolding experiment, the pathway of unfolding of a single protein is biased along the direction of the applied force, which is precisely known. Single molecule force spectroscopy experiments have been highly successful in improving our understanding of protein folding.[8−14] These experiments provide a valuable tool to measure the mechanical properties of proteins. Indeed, force is a natural probe of protein folding and unfolding, unlike the use of harsh denaturants and/or large temperature jumps to perturb a folded protein. Furthermore, observing only population averages, as in bulk experiments, can hide dynamic or mechanistic features of biological molecules that are important for function. Single molecules exist at any given time in particular conformational states with their solvent environment. Single

† Part of the "Giacinto Scoles Festschrift".

* Corresponding authors. E-mail: (L.D.) ldougan@biology.columbia.edu.

10.1021/jp076095+ CCC: $37.00
Published on Web 11/20/2007

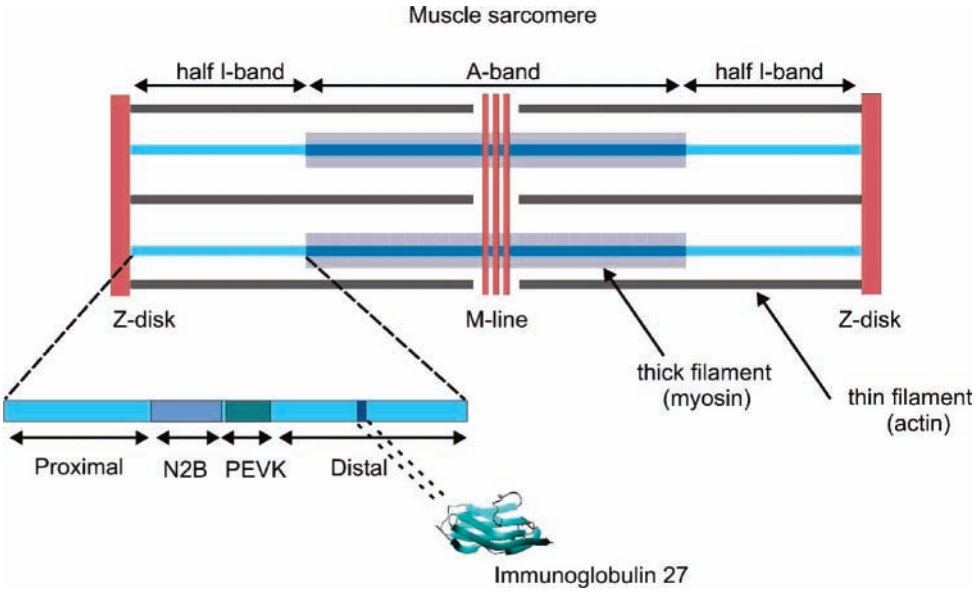

Figure 1. Schematic diagram showing the position of titin in the sarcomere. Titin spans the distance between the M-line and the Z-disk. It is made up of immunoglobulin-like (Ig) and fibronectin type III (fnIII) domains with additional less structured regions such as the N2B and PEVK region. The half I-band connects the Z-disk and the A-band and is composed of distinct regions, referred to as the proximal immunoglobulin-like region, N2B and PEVK region, and distal immunoglobulin-like region. The immunoglobulin-like protein I27 is located in the distal region of the half I-band.

molecule techniques allow the measurement of the distribution of activities of individual proteins.

Polyproteins are multidomain proteins composed of identical repeats of a single protein.[15] When folded, tandem modular polyproteins do not misfold or aggregate. Indeed, they are very stable, can be produced in large quantities, and remain usable after many months of storage. Under unfolding conditions, there may be competition between protein refolding and aggregation. Single molecule force spectroscopy experiments have shown that folding is favored with less than 2% refolding events showing misfolding.[16] Furthermore, polyproteins unfold completely reversibly in equilibrium denaturation experiments.[17] This suggests that rapid and efficient refolding is a mechanism that can help prevent aggregation of adjacent unfolded domains. Interestingly, the domains of the giant protein titin that are most likely to unfold on application of force are those that are most able to recover quickly. Of the distal I-band domains of titin that have been examined, the I27 protein is mechanically the weakest,[18] and yet chemical denaturation experiments have shown that it refolds more rapidly than any of the other distal I-band domains that have been studied.[19] Furthermore, light scattering experiments have shown that the use of denaturants in bulk solution experiments amplifies the probability of misfolding. Indeed, a recent study using such methods revealed that coaggregation between different domains only occurs when the domains have a high sequence identity.[20] In these bulk solution experiments, the polyprotein $I27_8$ was observed to misfold and aggregate in a 28% v/v 2,2,2-trifluoroethanol (TFE) solution, a substance known to enhance intermolecular interactions.[21] Immunoglobulin-like domains with more than 70% sequence identity were found to be highly prone to coaggregation, while those with less than 30−40% sequence identity did not detectably interact. The conclusion of this work was that sequence diversity is important in avoiding aggregation in the evolution of proteins.[20]

Following from this idea, we have carried out a different kind of study to explore misfolding and aggregation of proteins. Instead of studying aggregation by modifying the density of proteins in a bulk solution, we probed the unfolding and folding behavior of individual tandem repeating protein domains in TFE solutions. The very nature of a polyprotein, where repeats of identical proteins are in very close proximity to each other, results in an intrinsic high density of proteins. We have completed single molecule force spectroscopy experiments on polyprotein $I27_8$ in a number of TFE solutions. We have utilized a single molecule micromanipulation technique that has proved to be successful for studying protein folding.[15,16,22,23] These experiments probe the specific folding behavior of tandem repeats of proteins unfolded in an environment known to promote intermolecular interactions in bulk solutions.[21] These single molecule force spectroscopy experiments provide a new perspective on the folding and aggregation properties of tandem repeating proteins.

Experimental Procedures

Protein Engineering and Purification. For efficient single molecule experiments, we first constructed polyproteins using protein engineering. The details of the polyprotein engineering and purification have been reported previously.[15] Briefly, we constructed an eight domain N−C linked polyprotein of I27, the 27th immunoglobulin-like domain of cardiac titin, through successive cloning in modified pT7Blue vectors and then expressed the gene using vector pQE30 in *Escherichia coli* strain BLR(DE3). Pelleted cells were lysed by sonification, and the His-6-tagged soluble protein was purified first by immobilized metal ion affinity chromatography and then by gel filtration. The protein was stored at 4 °C in 50 mM sodium phosphate/150 mM sodium chloride buffer (pH 7.2). The polyproteins have two Cys residues in their carboxyl terminus to facilitate the attachment of the molecules to the surface.

Solvent Environment. Samples of TFE and water were obtained from Sigma-Aldrich and used without additional purification. Solvent mixtures were prepared to obtain the desired volume fraction (v/v) of TFE. The solvent mixtures were prepared in PBS buffer (pH 7.4) at 25 °C.

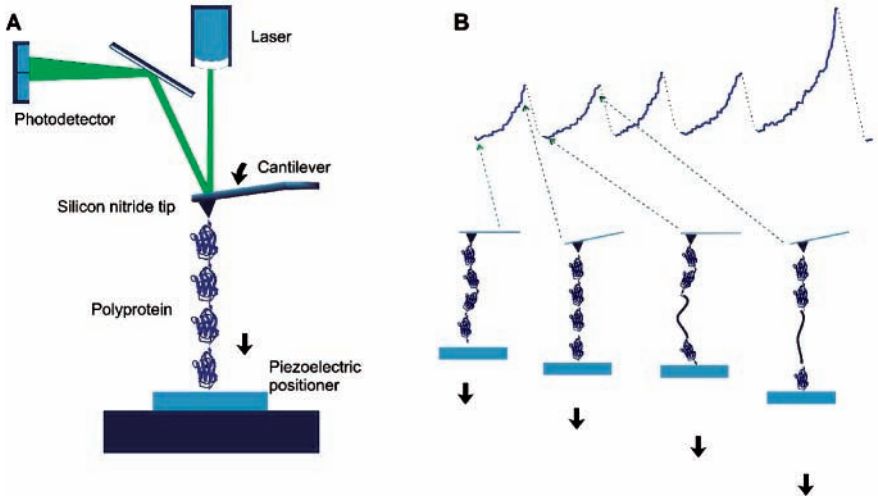

Figure 2. Force−extension mode of the atomic force microscope. When pressed against the layer of protein attached to a substrate, the cantilever tip can adsorb a single protein molecule. Extension of a molecule by retraction of the piezoelectric positioner results in deflection of the cantilever.

Single Molecule Force−Extension Spectroscopy. We used a custom-built atomic force microscope (AFM) equipped with a PicoCube P363.3-CD piezoelectric translator (Physik Instrumente, Karlsruhe, Germany) controlled by an analogue PID feedback system that has been described previously.[24] Silicon-nitride cantilevers (Veeco, Santa Barbara, CA) were calibrated for their spring constant using the equipartition theorem as reported.[25] The average spring constant was ∼65 pN/nm. All data were obtained and analyzed using custom software written for use in Igor 5.0 (Wavemetrics, Oswego, OR).

In an atomic force microscopy experiment, a single molecule is stretched between the microscopic silicon nitride tip of a flexible cantilever and a flat substrate that is mounted on a highly accurate piezoelectric positioner (Figure 2A). The proteins were suspended in buffer at a concentration of ∼10 μg/mL and adsorbed onto freshly evaporated gold coverslips. Single protein molecules were stretched by first pressing the cantilever on the gold-coated coverslide for 3 s at 500−700 pN, then retracting at a constant velocity. As the distance between tip and substrate increases, extension of the molecule generates a restoring force that causes the cantilever to bend. This causes deflection of a laser beam directed toward the upper surface of the cantilever, measured using four highly sensitive photodetectors. The output of the photodetectors can be related to the angle of the cantilever and therefore to the applied force, if the elastic properties of the cantilever are known. This system then allows spatial manipulation of less than a nanometer and can measure forces of a few piconewtons.

In the AFM experiments, an O-ring was used to minimize the rate of evaporation of the solvent buffer. The O-ring fits into the fluid cell and allows a seal to be formed for the protein in solution between the fluid cell and the coverslip. Additionally, experiments were carried out over a period of 2−3 h to minimize the effects of any evaporation, ensuring that the concentration of TFE in the solution remained fixed.

Data Analysis. The force−extension curves are well-described by the worm-like-chain (WLC) model of polymer elasticity[11]

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{x}{L_C}\right)^{-2} - \frac{1}{4} + \frac{x}{L_C}\right]$$

which predicts the entropic restoring force (F) generated on the extension (x) of a polymer, where L_C is the contour length, x is the end-to-end distance of the chain, p is the persistence length, and k_B and T are Boltzmann's constant and temperature, respectively. Consecutive peaks in the sawtooth pattern (described next) were fitted by WLC with the same persistence length and contour length increment, ΔL_C.

Results

Fingerprint of Protein Unfolding. We construct polyproteins with eight repeats of the immunoglobulin-like human cardiac titin domain I27. The I27 protein is an 89 residue β-sandwich protein with well-characterized properties.[15,26] The use of polyproteins is advantageous in that they provide a clear mechanical fingerprint to distinguish them against a background of spurious interactions.

In the force−extension experiment, single polyproteins were fully extended by retracting the sample holding substrate away from the cantilever tip at a constant velocity of 400 nm s^{-1}, generating a periodic sawtooth pattern (Figure 2B). This sawtooth pattern is then the fingerprint for single molecule studies of protein unfolding and results from the sequential extension and unfolding of the protein modules. The peak force reached before an unfolding event measures the mechanical stability of the protein module, while the spacing between peaks is a measure of the increased contour length, $\Delta L_C = \sim 28.4$ nm, of the protein as it unfolds. The last peak in the sawtooth pattern represents the final extension of the unfolded protein prior to detachment from the atomic force microscope tip.

Mechanical Stability of the I27 Protein in Different Solvent Environments. Stretching the I27 protein in aqueous buffer resulted in a distribution of forces required to unfold the protein (Figure 3A). A histogram of force peaks (unfolding force) measured from 200 unfolding events in aqueous solution reveals a distribution of events that has an average unfolding force of 200 ± 14.5 pN (Figure 3B) in agreement with previous literature.[15,28] Force−extension experiments on the I27 protein were carried out in a range of TFE solutions: 10, 20, 28, and 35% v/v TFE. Figure 3C shows typical sawtooth pattern data for mechanical unfolding of the I27 protein in 28% v/v TFE. A histogram of force peaks (unfolding force) measured from 200 unfolding events in 28% v/v TFE reveals a distribution of events that has an average unfolding force of 185 ± 17.0 pN (Figure 3D). A plot of the peak of the unfolding force distribution as a function of the concentration of TFE is shown in Figure 3E. The dashed line in Figure 3E shows a linear fit to the data with

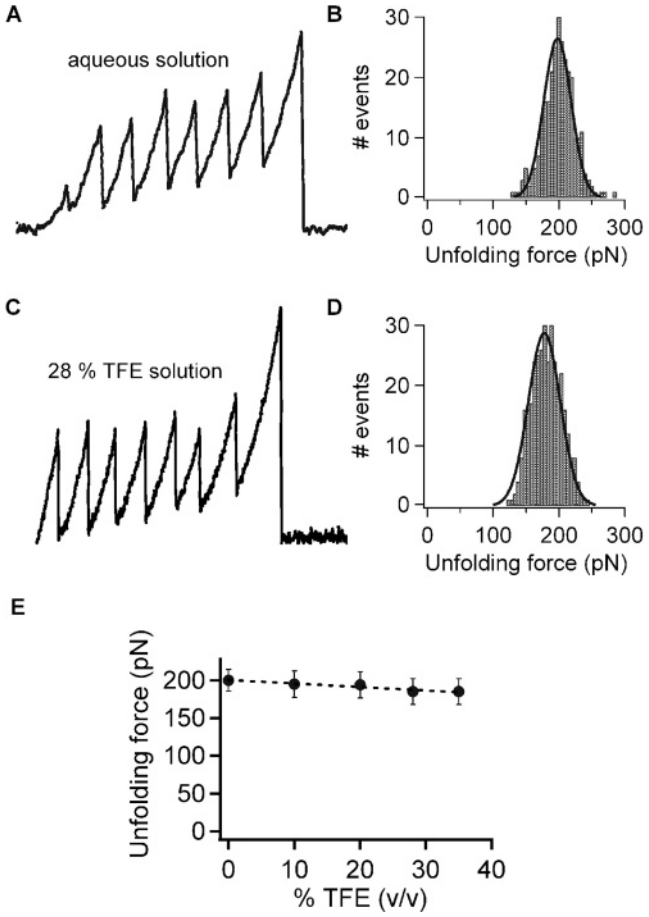

Figure 3. Force−extension experiments reveal unfolding behavior of the polyprotein $I27_8$ in 28% v/v aqueous TFE. Force−extension relationships for the polyprotein $I27_8$ measured using AFM. (A) Stretching of a single I27 polyprotein in aqueous solution results in a force−extension curve with a sawtooth pattern having equally spaced force peaks. (B) Unfolding force histogram for $I27_8$ in aqueous solution showing an average peak unfolding force of 200 ± 14.5 pN. (C) Sawtooth pattern obtained from unfolding of the I27 protein in 28% v/v TFE solution. (D) Unfolding force histogram for $I27_8$ in 28% v/v TFE solution showing an average peak unfolding force of 185 ± 17.0pN. (E) Average peak unfolding force as a function of concentration of TFE v/v. The dashed line shows a linear fit to the data with a slope of 0.45 pN/% TFE (v/v).

a slope of 0.45 pN/% TFE (v/v). While previous bulk studies have measured the destabilization of proteins using TFE,[20,21] our experiments provide the first single molecule level demonstration of the destabilization of a protein by TFE. Furthermore, TFE is known to destabilize proteins under a range of environmental conditions such as temperature and chemical denaturants. Here, we have found that in addition, TFE destabilizes a mechanically unfolded protein. It is interesting to note that this destabilization, although evident, is mild with only a 10% reduction in F_{UN} with the substitution of 35% TFE. Interestingly, there is no universal molecular theory that can explain the mechanism by which solvents such as TFE affect the stability of a protein. Several studies have shown that TFE increases the structure of the TFE/water solution, thereby increasing the energetic cost associated with solvation of the polypeptide backbone.[29] This change in the hydrogen bond capacity of water destabilizes the unfolded state. In some proteins, TFE can affect the three-dimensional structure, stabilizing and inducing helix structures.[30]

Refolding of I27 Protein in TFE Solutions. Previous work has shown that incubation of the small protein, acylphosphatase, in a 28% v/v TFE solution leads to the formation of insoluble amyloid fibrils.[21] Furthermore, this concentration of TFE marks the saturation point for rapid formation of structured aggregates that show characteristics of amyloid fibrils.[20] Therefore, a 28% v/v TFE solution provided an excellent starting point to study the refolding behavior of an individual tandem repeating polyprotein $I27_8$. In the experiment, a single polyprotein remained attached to the atomic force microscope tip allowing for repeated extension and relaxation cycles following a double pulse protocol[22] (Figure 4). A first extension allows counting of the available folded domains. After reaching the extended state, the protein is relaxed to its initial length. After a variable time period, Δt, the protein is again stretched.

We carried out refolding experiments on our identical tandem repeating polyprotein $I27_8$ in 28% v/v TFE. Stretching the polyprotein $I27_8$ in 28% v/v TFE resulted in force−extension curves with peaks that varied randomly in amplitude with an average value of ∼185 pN. In the example shown (Figure 4A), the protein is extended, and five domains unfold. Since the polyprotein is picked up at random, and the total extension of the protein is limited to prevent detachment, the number of extended domains is typically less than the maximum (eight domains). The protein is then relaxed for a delay time $\Delta t = 1$ s. Upon a second extension, five protein domains unfold. In a second example (Figure 4A), the protein is extended, and four domains unfold. The protein is relaxed for a delay time $\Delta t =$ 0.1 s and upon a second extension, two protein domains unfold. We interpret these results as an indication that some of the domains that unfolded in the first extension of the protein had spontaneously refolded upon relaxation.[15,16,22,23] Such single molecule refolding experiments revealed many events (>200) where the I27 protein had unfolded in TFE solution and refolded without any aggregation or misfolding occurring. During these repeated unfolding and refolding cycles, the sawtooth pattern of the force−extension curve occasionally shows skips in the force peaks, indicating misfolding of a protein.[16] In the example shown (Figure 4B), a first extension shows four unfolding peaks. A second pull, after the protein was relaxed to its resting length and allowed to refold for a time $\Delta t = 1$ s, shows only two force peaks. The missing unfolding events observed during the refolding cycles occurred 4 times in 241 events. This frequency of occurrence of skip events is consistent with previous experiments on the I27 protein in the absence of TFE.[16]

By varying the waiting time Δt from 0.1 to 5 s and measuring the ratio of folded modules, $N_{refolded}/N_{total}$, we deduced the rate constants of refolding for the polyprotein $I27_8$ in 28% v/v TFE. The plot in Figure 4C shows the time evolution of the fraction of refolded modules or the probability of refolding for the I27 protein in aqueous solution (black squares) and in 28% v/v TFE (red circles). The refolding rate constant k_f is deduced from the single exponential fits of a two state model to the data.[22] The solid lines are fits of the data to the function $P(t) = 1 - \exp(-t/k_f)$. The folding rate constant of the I27 protein in aqueous solution is 1.35 ± 0.51 s^{-1}, while in the TFE solution, $k_f = 1.7 \pm 0.15$ s^{-1}. Indeed, it is evident from Figure 4C that the addition of TFE does not alter k_f greatly.

To further explore the refolding behavior of the I27 protein, we carried out force−extension refolding experiments in 35% v/v TFE. In Figure 5, one such example of successive refolding in a 35% v/v TFE solution is shown. The protein is first stretched, and five domains unfold (trace 1). The protein is relaxed to its initial length for a delay time $\Delta t = 1$ s. Upon a

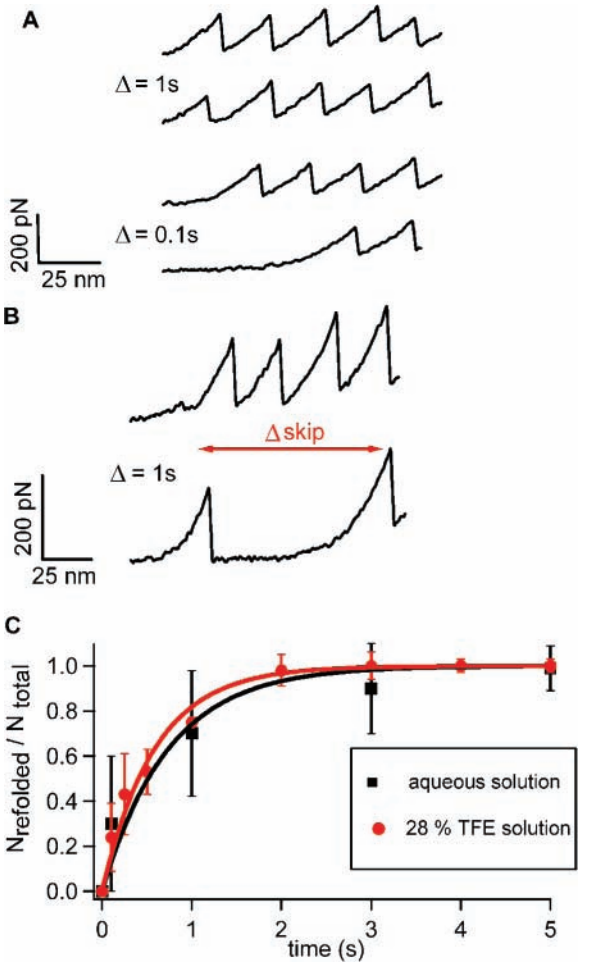

Figure 4. Unfolding and refolding cycles of the polyprotein $I27_8$ in 28% v/v TFE. (A) The protein is first stretched to count the number of domains that unfold and then relaxed to its initial length. A second extension after a delay time, Δt, measures the number of refolded modules. In the first example, the protein is extended, and five domains unfold. The protein is relaxed for a delay time $\Delta t = 1$ s. Upon a second extension, five protein domains unfold. In the second example, the protein is extended, and four domains unfold. The protein is relaxed for a delay time $\Delta t = 0.1$ s. Upon a second extension, two protein domains unfold. (B) Reversible misfolding events in the polyprotein $I27_8$ captured by single molecule AFM. The protein is extended, and four domains unfold. The protein is relaxed for a delay time $\Delta t = 1$ s. Upon a second extension, only two protein domains unfold. The sawtooth pattern shows two skips in the force peaks. (C) Refolding rate of the I27 protein in TFE solution. Plot of the refolded fraction $N_{refolded}/N_{total}$ vs Δt for the I27 protein in aqueous solution (black squares) and the I27 protein in 28% v/v TFE (red circles). For TFE, each symbol (red circle) is the average of 30, 39, 41, 27, 25, 24, 33, and 22 data points obtained from three separate experiments. The solid line is a fit of the data to the function $P_f(t) = 1 - \exp(-t/k_f)$, where $k_f = 1.35 \pm 0.51$ s^{-1} for the I27 protein in aqueous solution and $k_f = 1.70 \pm 0.15$ s^{-1} for the I27 protein in 28% v/v TFE.

second extension, four domains unfold (trace 2 in Figure 5). This cycle is repeated many times, with the four subsequent cycles shown (traces 3−8 in Figure 5). Successive unfolding and refolding is observed. Strikingly, this concentration of TFE (35%) lies well above the saturation point of TFE observed to induce the unfolding of globular proteins and promote intermolecular interactions (28%).[21]

Discussion

It has recently been proposed that low sequence identity may play an important role in safeguarding proteins against misfolding and aggregation.[20] The implication is that tandem modular proteins of high sequence identity may be prone to aggregation. To test this hypothesis and provide new insight into protein misfolding, we have completed single molecule force spectroscopy experiments on the tandem repeating polyprotein $I27_8$. We have completed refolding experiments in solutions containing TFE, a solvent known to enhance intermolecular interactions and promote aggregation of proteins. Successful refolding of the polyprotein $I27_8$ was measured in solutions with up to 35% v/v TFE. These measurements show that tandem repeats of identical domains can successfully refold in a denaturing environment (35% TFE). These new single molecule experiments then raise the question as to why aggregation of polyproteins is observed in light scattering experiments[20] but not observed in single molecule experiments.

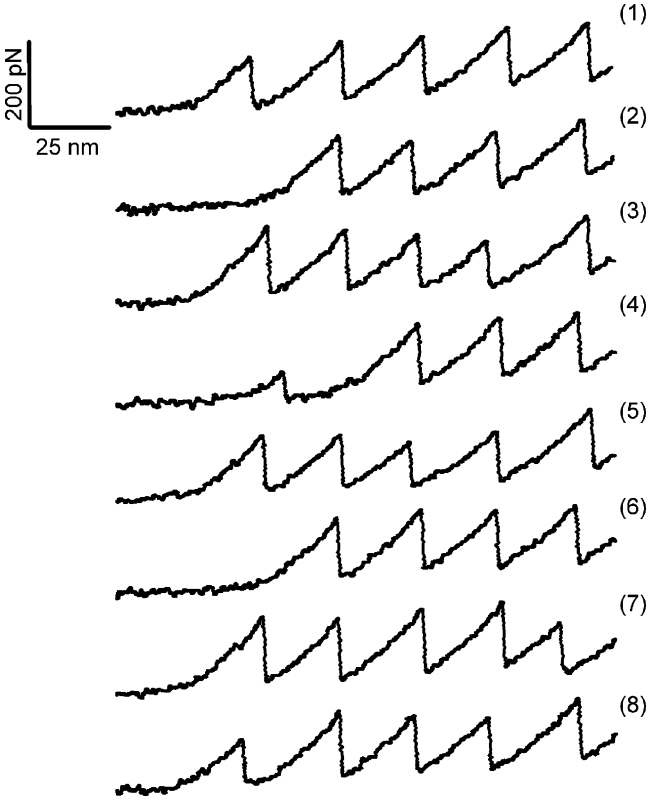

Figure 5. Multiple unfolding and refolding cycles of the polyprotein $I27_8$ in 35% v/v TFE solution. The protein is first stretched, and five domains unfold (1). The protein is relaxed to its initial length for a delay time $\Delta t = 1$ s. Upon a second extension, four domains unfold (2). This cycle is repeated many times, with the four subsequent cycles shown (3−8). Successive unfolding and refolding is observed.

In a force spectroscopy experiment, a single polyprotein, containing identical neighboring repeats, is mechanically unfolded along a well-defined reaction coordinate. Upon mechanical unfolding, each protein in the polyprotein chain is unfolded to a highly extended conformation (86% of the contour length[31]) and is situated next to an identical neighboring unfolded protein. Therefore, each unfolded protein has a local effective protein concentration that is very high and is fully exposed to the solvent environment. Conversely, in a bulk solution experiment, molecules undergoing chemical denaturation explore a wide range of unfolded, structures without a well-defined reaction coordinate. These heterogeneous denatured conformations are close in size (radius of gyration) to the native state of the protein.[32,33] In these experiments, the protein concentration is high to

increase the likelihood that single proteins will be in close enough proximity to interact.

The aggregation rate of the I27 protein in bulk solution has been studied using light scattering techniques and fluorescence techniques.[20] Increasing the concentration of single proteins in the solution led to an increase in the aggregation rates. Furthermore, in this work,[20] the aggregation kinetics of the monomer protein were compared to the aggregation kinetics of polyprotein constructs $I27_2$, $I27_3$, and $I27_8$. The aggregation rates of the polyproteins were 10 times faster than the aggregation rates of the monomer. Interestingly, very little difference in the aggregation rates was measured between polyproteins of different lengths.[20] This study proposed that the initial interaction of a pair of domains is the key step in the aggregation of the system. The authors concluded that the length of the polyprotein chain was not the limiting factor but rather the presence of two neighboring domains. In light of our results, it appears that single molecule techniques and bulk solution techniques are distinct. In single molecule experiments, the local effective concentration is always high, owing to the very nature of the polyprotein construct. In bulk solution experiments, a population of neighbors is achieved through the use of high concentrations of protein in the solution.

In force spectroscopy experiments, aggregation of proteins in solution may occur. However, these aggregates have no accessible mechanical fingerprint and therefore cannot be probed directly with this technique. Conversely, in traditional light scattering experiments, aggregation of proteins can be probed. However, there is no direct measure of the number of correctly folded proteins that remain in the solution, and it is possible that a large number of correctly folded proteins still exist. Instead, only the overwhelming scattering from the aggregated proteins is observed. Therefore, two experimental techniques probe different features of this system. Single molecule refolding experiments provide a bridge between these two experimental techniques. In these experiments, the success of a single protein in unfolding and refolding can be monitored. These measurements show that a single polyprotein can be unfolded, relaxed, and subsequently unfolded in 35% v/v TFE—confirming that the protein has successfully refolded in solution. Interestingly, this concentration of TFE lies above the concentration at which this cosolvent is known to promote intermolecular interactions[21] and result in the formation of structured aggregates.[20]

The unfolding of neighboring modular proteins is likely to be a common theme in the function of modular proteins such as titin, tenascin, spectrin, and many others[31] allowing them to unfold sequentially when subjected to stretching forces applied to their N-C termini.[34] On removal of force, unfolded proteins can refold back to their original folded structure to recover their mechanical stability, maintaining their mechanical properties during repeated stretching and relaxation cycles. From this perspective, the mechanical unfolding of polyproteins captured using force spectroscopy closely reflects the unfolding of modular proteins in vivo, as compared with the study of isolated monomers. Indeed, many proteins have been expressed successfully in tandem array and studied using force spectroscopy, including ubiquitin[9,24] protein G,[35] protein L,[36] green fluorescent protein,[37] and many immunoglobulin-like (Ig-like) domains.[15,28,38] It is possible that tethering a protein, as in force spectroscopy experiments, may favor folding over aggregation and could account for the differences seen when comparing to bulk experiments. However, a recent molecular dynamics simulation study[39] has shown that neither the folding mechanism nor the transition state of a model protein were typically altered significantly when the protein was tethered. Thus, it is likely that experiments done using single molecule techniques and tandem repeat proteins closely resemble conditions that occur in physiological systems. In addition, recent bioinformatic studies have suggested that a significant proportion of proteins in vivo consist of more than one domain.[40,41] Long repeats, containing several domains in tandem, have been observed to be particularly common in multicellular species,[41,42] suggesting that repeats have arisen late in evolution. Therefore, it seems possible that evolutionary mechanisms have developed to avoid the misfolding of these proteins. Interestingly, there is some evidence which suggests that proteins have conserved residues that are important for preventing misfolding or aggregation.[44,45] These have been termed gatekeeper residues; that is, they close the door to the aggregation pathway and allow folding to compete more effectively. These gatekeeper residues are thought to include charged residues and conserved proline or glycine residues. Additionally, it has been suggested that there are a number of specific sequence-based signals in multidomain proteins, and in particular in tandem arrays of identical domains, that prevent interdomain aggregation. However, this hypothesis remains to be tested both by analysis of sequences or by experiment. Indeed, single molecule force spectroscopy is an excellent candidate for exploring the importance and role of gatekeeper residues in correct protein folding.

Under physiological conditions, protein misfolding and aggregation may be triggered by a number of genetic and environmental factors. The mechanisms by which environmental factors might catalyze protein misfolding include changes in pH or oxidative stress, pathological chaperone proteins, macromolecular crowding, and increases in the concentration of misfolded proteins.[1] Recent studies of the molecular mechanism of brain degeneration strongly support the hypothesis that diverse neurodegenerative diseases are caused by the misfolding, aggregation, and accumulation in the brain of an underlying protein.[1] The trigger for the misfolding and aggregation process may depend on hydrophobic interactions and hydrogen bonding between the protein molecules. It has been proposed that protein misfolding and aggregation follows a seeding−nucleation mechanism modulated by several environmental factors and involving the formation of at least two intermediates: oligomers and protofibrils. Kinetic studies have tried to monitor the formation of oligomers, identifying the transition from monomer proteins to dimers and up to decamers.[1] However, structural and biochemical characteristics of these interactions have proved to be challenging because they are transient and unstable. We have shown that single molecule force spectroscopy is a novel technique that provides a new perspective on the question of protein folding and aggregation. Further experimentation, using this new tool, may provide a model for the initial formation of aggregates and intermediates in solution. This model could provide new insight into the diseases caused by protein aggregation while maintaining a clear experimental fingerprint for the protein of interest.[15,28,38]

Acknowledgment. We thank Arun Wiita and Sergi Garcia-Manyes for careful reading of the manuscript and members of the Fernandez lab for helpful discussions. This work was supported by NIH grants to J.M.F. (HL66030 and HL61228).

References and Notes

(1) Soto, C. *Nat. Rev. Neurosci.* **2003**, *4*, 49.

(2) Dobson, C. M. *Philos. Trans. R. Soc. London, Ser. B* **2001**, *356*, 133.

(3) Dobson, C. M. *Nature (London, U.K.)* **2003**, *426*, 884.

(4) Horwich, A. *J. Clin. Invest.* **2002**, *110*, 1221.
(5) Radford, S. E.; Dobson, C. M. *Cell* **1999**, *97*, 291.
(6) Thomas, P. J.; Qu, B. H.; Pedersen, P. L. *Trends Biochem. Sci.* **1995**, *20*, 456.
(7) Labeit, S.; Kolmerer, B. *Science (Washington, DC, U.S.)* **1995**, *270*, 293.
(8) Fisher, T. E.; Oberhauser, A. F.; Carrion-Vazquez, M.; Marszalek, P. E.; Fernandez, J. M. *Trends Biochem. Sci.* **1999**, *24*, 379.
(9) Fernandez, J. M.; Li, H. *Science (Washington, DC, U.S.)* **2004**, *303*, 1674.
(10) Li, H. B.; Fernandez, J. M. *J. Mol. Biol.* **2003**, *334*, 75.
(11) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature (London, U.K.)* **1998**, *393*, 181.
(12) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science (Washington, DC, U.S.)* **1997**, *276*, 1109.
(13) Wiita, A. P.; Ainavarapu, S. R. K.; Huang, H. H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7222.
(14) Walther, K.; Grater, F.; Dougan, L.; Badilla, C. L.; Berne, B. J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7916.
(15) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694.
(16) Oberhauser, A. F.; Marszalek, P. E.; Carrion-Vazquez, M.; Fernandez, J. M. *Nat. Struct. Biol.* **1999**, *6*, 1025.
(17) Rounsevell, R. W. S.; Steward, A.; Clarke, J. *Biophys. J.* **2005**, *88*, 2022.
(18) Li, H. B.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkviliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M. *Nature (London, U.K.)* **2002**, *418*, 998.
(19) Scott, K. A.; Steward, A.; Fowler, S. B.; Clarke, J. *J. Mol. Biol.* **2002**, *315*, 819.
(20) Wright, C. F.; Teichmann, S. A.; Clarke, J.; Dobson, C. M. *Nature (London, U.K.)* **2005**, *438*, 878.
(21) Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3590.
(22) Koti, A. S. R. a.; Brujic, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Walther K. A.; Carrion-Vazquez, M.; Li, H.; Fernandez, J. M. *Biophys. J.* **2007**, *92*, 225.
(23) Li, H. B.; Carrion-Vazquez, M.; Oberhauser, A. F.; Marszalek, P. E.; Fernandez, J. M. *Nat. Struct. Biol.* **2000**, *7*, 1117.
(24) Schlierf, M.; Li, H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7299.
(25) Oberhauser, A. F.; Hansma, P. K.; Carrion-Vazquez, M.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 468.
(26) Marszalek, P. E.; Lu, H.; Li, H. B.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. M. *Nature (London, U.K.)* **1999**, *402*, 100.
(27) Garcia-Manyes, S.; Brujic, J.; Fernandez, J. M. *Biophys. J.* **2007**, *93*, 2436.
(28) Carrion-Vazquez, M.; Li, H. B.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738.
(29) Kentsis, A.; Sosnick, T. R. *Biochemistry* **1998**, *37*, 14613.
(30) Luo, P. Z.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413.
(31) Fernandez, J. M.; Li, H. B.; Brujic, J. *Science (Washington, DC, U.S.)* **2004**, *306*, 411.
(32) Plaxco, K. W.; Gross, M. *Nat. Struct. Biol.* **2001**, *8*, 659.
(33) Shortle, D.; Ackerman, M. S. *Science (Washington, DC, U.S.)* **2001**, *293*, 487.
(34) Tatham, A. S.; Shewry, P. R. *Trends Biochem. Sci.* **2000**, *25*, 567.
(35) Sharma, D.; Perisic, O.; Peng, Q.; Cao, Y.; Lam, C.; Lu, H.; Li, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9278.
(36) Brockwell, D. J.; Beddard, G. S.; Paci, E.; West, D. K.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophys. J.* **2005**, *89*, 506.
(37) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16192.
(38) Oberhauser, A. F.; Badilla-Fernandez, C.; Carrion-Vazquez, M.; Fernandez, J. M. *J. Mol. Biol.* **2002**, *319*, 433.
(39) Friedel, M.; Baumketner, A.; Shea, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8396.
(40) Bjorklund, A. K.; Ekman, D.; Elofsson, A. *PLOS Comput. Biol.* **2006**, *2*, 959.
(41) Ekman, D.; Bjorklund, A. K.; Frey-Skott, J.; Elofsson, A. *J. Mol. Biol.* **2005**, *348*, 231.
(42) Apic, G.; Gough, J.; Teichmann, S. A. *J. Mol. Biol.* **2001**, *310*, 311.
(43) Marcotte, E. M.; Pellegrini, M.; Yeates, T. O.; Eisenberg, D. *J. Mol. Biol.* **1999**, *293*, 151.
(44) Parrini, C.; Taddei, N.; Ramazzotti, M.; Degl'lnnocenti, D.; Ramponi, G.; Dobson, C. M.; Chiti, F. *Structure (Cambridge, MA, U.S.)* **2005**, *13*, 1143.
(45) Steward, A.; Adhya, S.; Clarke, J. *J. Mol. Biol.* **2002**, *318*, 935.