Retraction

CHEMISTRY

The authors note the following: “We wish to retract our article because interpretation of the experimental data was based on incorrect calibration of the apparatus in viscous solutions and there is no basis for the major conclusions of the study on the structure of the transition state for mechanical unfolding.”

www.pnas.org/cgi/doi/10.1073/pnas.1118431108
Solvent molecules bridge the mechanical unfolding transition state of a protein

Lorna Dougan**, Gang Feng†, Hui Lu‡, and Julio M. Fernandez**

**Department of Biological Sciences, Columbia University, New York, NY 10027; and †Department of Bioengineering, University of Illinois, Chicago, IL 60607

We demonstrate a combination of single molecule force spectroscopy and solvent substitution that captures the presence of solvent molecules in the transition state structure. We measure the effect of solvent substitution on the rate of unfolding of the I27 titin module, placed under a constant stretching force. From the force dependency of the unfolding rate, we determine $\Delta x_u$, the distance to the transition state. Unfolding the I27 protein in water gives a $\Delta x_u = 2.5$ Å, a distance that compares well to the size of a water molecule. Although the height of the activation energy barrier to unfolding is greatly increased in both glycerol and deuterium oxide solutions, $\Delta x_u$ depends on the size of the solvent molecules. Upon replacement of water by increasing amounts of the larger glycerol molecules, $\Delta x_u$ increases rapidly and plateaus at its maximum value of 4.4 Å. In contrast, replacement of water by the similarly sized deuterium oxide does not change the value of $\Delta x_u$. From these results we estimate that six to eight water molecules form part of the unfolding transition state structure of the I27 protein, and that the presence of just one glycerol molecule in the transition state is enough to lengthen $\Delta x_u$. Our results show that solvent composition is important for the mechanical function of proteins. Furthermore, given that solvent composition is actively regulated in vivo, it may represent an important modulatory pathway for the regulation of tissue elasticity and other important functions in cellular mechanics.

Water is recognized as an active participant in biological processes, responding structurally and dynamically to the presence of other molecules in subtle and nonintuitive ways (1, 2). In particular, water is thought to play a dynamic role in protein folding and unfolding (3, 4). Water facilitates the necessary changes of the hydrogen bonding network, allowing fast conformational changes (5, 6). During protein folding, water mediates the collapse of the chain and search for native state conformations, contributing to both enthalpic and entropic stabilization (4). An understanding of the mechanisms of protein unfolding and refolding must incorporate the solvating environment which envelopes the molecule. Indeed, the energy landscape of a protein can be affected by changing the solvent environment, suggesting different structures and altering the free energy of the folding transition state. Processes, responding structurally and dynamically to the presence of another molecule in subtle and nonintuitive ways (1, 2). In particular, water is thought to play a dynamic role in protein folding and unfolding (3, 4). Water facilitates the necessary changes of the hydrogen bonding network, allowing fast conformational changes (5, 6). During protein folding, water mediates the collapse of the chain and search for native state conformations, contributing to both enthalpic and entropic stabilization (4). An understanding of the mechanisms of protein unfolding and refolding must incorporate the solvating environment which envelopes the molecule. Indeed, the energy landscape of a protein can be affected by changing the solvent environment, suggesting different structures and altering the free energy of the folding transition state.

Results and Discussion

In our experiments, we construct polyproteins with eight repeats of the human cardiac titin domain I27 (27). Polyproteins are multidomain proteins composed of identical repeats of a single protein (28). The I27 protein is ideal for these experiments given that its mechanical properties have been well characterized both experimentally (21, 28–30) and also in silico using molecular dynamics techniques (31–33). The use of polyproteins is advantageous because it allows for a clear mechanical fingerprint to distinguish them against a background of spurious interactions and also provide us with a larger number of events per recording than otherwise possible with monomers (21).

We measure the properties of the mechanical unfolding transition state of the I27 protein by measuring the force dependency of the unfolding rate of single I27 proteins. When a protein is subjected to an external force its unfolding rate, $k_u$, is well described by an Arrhenius term of the form $k_u(F) = k_u^0 \exp(-\Delta G_u/k_BT)$ (16, 19, 34) where $k_u^0$ is the unfolding rate in the absence of external forces, $F$ is the applied force and $T$ is the temperature. The force dependency of the unfolding rate is typical fit with a straightforward Arrhenius term that measures properties of the unfolding transition state (19). In its simplest representation, the unfolding transition state is characterized by two parameters: the size of its activation energy, $\Delta G_u$, and the elongation of the protein necessary to reach the transition state, $\Delta x_u$. Of particular interest are the force spectroscopy measurements of $\Delta x_u$ which provide a direct measure of the length scales of a transition state, which were hitherto unknown. For example, for protein folding the distance to the folding transition state, $\Delta x_u$ was found to be between 8 Å (17) and 60 Å (23), in rough agreement with the expected role of long range hydrophobic forces (24). For protein unfolding $\Delta x_u$ was found to be much shorter, in the range of 1.7–2.5 Å (16, 19, 21). These values of $\Delta x_u$ are comparable to the size of a water molecule (25), suggesting that water molecules are integral components of the unfolding transition state.

Here we use single molecule force-clamp spectroscopy to test this prediction by measuring the distance to the unfolding transition state, $\Delta x_u$ of the human cardiac titin domain I27 in the presence of glycerol and deuterium oxide. Glycerol is a good hydrogen bonding molecule which is ubiquitous in living systems (26) known to enhance protein stability (9), and is larger in size than water. Deuterium oxide forms stronger hydrogen bonds than water while having a similar size (45). Our experiments directly demonstrate that solvent molecules form part of the structure of the mechanical transition state of a protein.

Author contributions: L.D., H.L., and J.M.F. designed research; L.D. and G.F. performed research; L.D. and H.L. contributed new reagents/analytic tools; L.D. and G.F. analyzed data; and L.D., H.L., and J.M.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence may be addressed. E-mail: ldougan@biology.columbia.edu or jfernandez@illinois.edu

This article contains supporting information online at www.pnas.org/cgi/content/full/0706075105/DC1.

© 2008 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0706075105

PNAS | March 4, 2008 | vol. 105 | no. 9 | 3185–3190
Δx_u is the distance from the native state to the transition state along the pulling direction. By measuring how the unfolding rate changes with an applied force we can readily obtain estimates for the values of both, $k_u^0$ and $Δx_u$ (16, 19, 34). Given that $k_u^0 = A\exp(-ΔG_u/k_BT)$ and assuming a prefactor, $A = 10^{13}$ s$^{-1}$ (28), we can estimate the size of the activation energy barrier of unfolding $ΔG_u$. The distance to transition state, $Δx_u$, determines the sensitivity of the unfolding rate to the pulling force and measures the elongation of the protein at the transition state of unfolding. Given that both $k_u^0$ and $Δx_u$ reflect properties of the transition state of unfolding, we expected these variables to be strongly influenced by the solvent environment.

Under force-clamp conditions, stretching a polyprotein results in a well defined series of step increases in length, marking the unfolding and extension of the individual modules in the chain (16). The size of the observed steps corresponds to the number of amino acids released by each unfolding event. Stretching a single I27s polyprotein at a constant force of 160 pN, results in a series of step increases in length of 24 nm (Fig. 1A). The time course of these events is a direct measure of the unfolding rate at 160 pN. We measure the unfolding rate by fitting a single exponential to an average of 20 traces similar to the ones shown in Fig. 1A. We define the unfolding rate as $k_u(F) = 1/τ(F)$, where $τ(F)$ is the time constant of the exponential fits to the averaged unfolding traces, shown in Fig. 1B Top. Furthermore, we obtain an estimate of the standard error of $k_u(F)$, using the bootstrapping technique (34, 35). We repeated these measurements over the force range between 120 and 220 pN and obtained the force-dependency of the unfolding rate in standard aqueous solution (Fig. 1C, filled squares). We fit the Arrhenius rate equation to the data (Fig. 1C, solid line over filled squares), and obtained $k_u^0 = 1.25 \times 10^{-4} \pm 10^{-5}$ s$^{-1}$ and $Δx_u = 2.5 \pm 0.1$ Å. From the measured value of $k_u^0$ we can readily estimate the value of $ΔG_u = 23.11$ kcal mol$^{-1}$.

To probe the role of the solvent in setting the structure of the unfolding transition state, we studied the effect of solvent substitution on the force dependency of the unfolding rate. Fig. 1B Middle and Bottom shows averaged unfolding traces and their corresponding exponential fits obtained at different forces in solutions containing 20% glycerol (Fig. 1B Middle) and 30% glycerol (Fig. 1B Bottom). Fig. 1C shows the force dependency of the unfolding rate in 20% glycerol (filled triangles) and 30% glycerol (filled circles). These data showed that replacing water by glycerol has a large effect on the force dependence of unfolding. It is readily apparent that the introduction of glycerol decreases the value of $k_u^0$ (larger $ΔG_u$) while increasing the slope ($Δx_u/k_BT$). Fits of the Arrhenius term to these data measured the increase in the value of $ΔG_u$ (26.16 kcal mol$^{-1}$ and 27.16 kcal mol$^{-1}$ in 20% and 30% glycerol, respectively). Our estimates of the unfolding activation energy, $ΔG_u$, depend on the value of the preexponential factor, $A$. The value of $A$ is not known and, for proteins, ranges from $10^{13}$ s$^{-1}$ to values as low as $10^6$ s$^{-1}$ (36). Here we use $10^9$ s$^{-1}$, which is used in bulk studies of solution protein biochemistry (28). Although the absolute values of $ΔG_u$ depend on the value taken for $A$, the observation that $ΔG_u$ increases with the concentration of glycerol remains unchanged. Although previous bulk studies have measured the stabilization of proteins using glycerol (9, 10, 26), our experiments provide a single-molecule-level demonstration of the stabilization of a protein by an osmolyte. More strikingly, the introduction of glycerol caused a large increase in the distance to transition state from $Δx_u = 2.5$ Å in aqueous and up to $4.0 \pm 0.1$ Å and $4.1 \pm 0.1$ Å in 20% and 30% glycerol, respectively (Fig. 1C). Earlier molecular dynamics simulations of forced unfolding of I27 suggested that when a stretching force is applied between the protein’s termini, resistance to unfolding originated from a set of hydrogen bonds between two parallel β-strands (A’ and G) of the protein structure (31–33). These β-strands must slide past one another for unfolding to occur. Because the hydrogen bonds are perpendicular to the axis of extension, they must rupture simultaneously to allow relative movement of the two termini. Hence, the rupture of these bonds defined the unfolding transition state for the I27 protein. A molecule like glycerol, larger than water (5.6 Å versus 2.5 Å, respectively) (37) but equally competent to form backbone hydrogen bonds would lead to a larger separation between the two key β-strands, enlarging the value of $Δx_u$. If this simple view were correct, the value of $Δx_u$ should increase to a value of 5.6 Å, when glycerol fully replaced water as the solvent.

The large stabilizing effect of glycerol (Fig. 1C) combined with the increased slope of the force-dependency of the unfolding rate made it difficult to do force-clamp experiments at concentrations >30%. Indeed, Fig. 1C (filled diamond) shows that the unfolding rate at 200 pN dropped 12-fold when the glycerol concentration

![Image](image-url)
was increased from 30% to 50%. Our force clamp instrumentation works well measuring rates as slow as ∼0.01 s⁻¹. Outside of this range, measurements become limited by cantilever drift in recordings that last longer than ∼60 s. Hence, a very steep force dependency such as the one we encountered for the unfolding of the I27₈ polyprotein at 50% glycerol, becomes much harder to measure.

To estimate ΔGa and ΔNₐ at concentrations of glycerol higher than 30%, we used the standard constant-velocity mode of single molecule atomic force microscopy (AFM) (28–30, 38–41, 50). Extending a polyprotein at constant velocity gives a very different perspective on the unfolding events. In these experiments, a polyprotein is extended by retracting the sample holding substrate away from the cantilever tip at a constant velocity of 400 nm s⁻¹. As the protein extends, the pulling force increases rapidly causing the unfolding of one of the I27 modules in the chain. Unfolding then extends the overall length of the protein relaxing the solvent bridging model (Eq. 6 (solid line), and 9 (long dashed line)). The peak of the unfolding force distribution as a function of the concentration of glycerol. Mechanical stability of I27 in glycerol solutions displays a biphasic force dependence for unfolding at a pulling speed of 400 nm s⁻¹ (red squares).

algorithm that computes, at a given extension and force, whether a protein module has unfolded. In these simulations, the force-dependent probability of unfolding is calculated from both ΔGa and ΔNₐ (28, 39, 41). The aim of these simulations is to find a set of unfolding parameters that can account for the distribution of unfolding forces (red traces in the histograms of Fig. 2B) measured from the sawtooth pattern data. Typically, the Monte Carlo simulations have to predict both the distribution of unfolding forces and also simultaneously predict how this distribution shifts with the pulling velocity. This stringent criterion can generate good estimates for ΔGa and ΔNₐ. Indeed, earlier estimates of the value of ΔNₐ for the unfolding of I27₈ in water obtained through the use of such Monte Carlo simulations (28) are identical to those measured here directly, using force-clamp techniques (2.5 Å in both cases; Fig. 1C).

Fig. 2 shows the effect of glycerol on the distribution of unfolding forces obtained from sawtooth pattern unfolding data. We obtained sawtooth pattern unfolding traces over the entire range of glycerol concentrations up to 100%. Sawtooth pattern traces obtained in 0%, 30%, and 100% glycerol (Fig. 2A) showed a surprising trend. Increasing the concentration of glycerol from 0% to 30% caused a decrease in the unfolding forces (Fig. 2A, top and middle traces), which then increased sharply as the concentration of glycerol was raised further up to 100% (Fig. 2A, bottom trace). This trend is better observed in histograms of unfolding forces obtained at different glycerol concentrations (Fig. 2B). A plot of the peak of the unfolding force distribution as a function of the concentration of glycerol is shown in Fig. 2C. As noted before, there is a surprising biphasic dependency of the unfolding forces on glycerol concentration. This results from the nonlinear dependency of ΔNₐ on the glycerol concentration (see SI Text). Monte Carlo simulations of the unfolding force distributions measured the changes of both ΔGa and ΔNₐ as a function of the glycerol concentration, right up to a pure glycerol solution (Fig. 3A and B, squares). The constant velocity sawtooth pattern
data confirms and extends the observations made under force-clamp conditions (Fig. 1C). Indeed, Fig. 3A shows that, whereas \( \Delta G_u \) increases linearly with the addition of glycerol, the value of \( \Delta u \) increases in a sharply nonlinear manner (Fig. 3B). Confidence in the values obtained through the Monte Carlo simulations is gained by comparing the estimates of \( \Delta G_u \) and \( \Delta u \), with those measured directly under force clamp conditions (Fig. 3A and B, circles).

Our measurements show that the size of the activation energy barrier for unfolding, \( \Delta G_u \), increases linearly with the concentration of glycerol (Fig. 3A). The increase can be as large as 11.49 kcal/mol in a pure glycerol solution. This increase is comparable to that measured for other thermally denatured proteins (10). The large increase in the activation energy barrier causes an increase in the average pulling force required to unfolded the I27 protein from 200 pN (0% glycerol) up to 310 pN (100% glycerol; see Fig. 2C). The known function of protective osmolytes in vivo is to stabilize proteins against a wide variety of environmental challenges such as high pressure (42) or temperature (10). Here, we have discovered that a protective osmolyte, glycerol, greatly increases the activation energy barrier to mechanical unfolding, expanding the repertoire of known protective function of these osmolytes.

Fig. 3B now reveals the full picture of the effect of glycerol on the distance to the unfolding transition state, \( \Delta u \). It is clear that the interaction in \( \Delta u \) with glycerol concentration is highly nonlinear, and saturates at a value of \( \Delta u \approx 4.4 \AA \). This increase is smaller than the full 5.6 \AA expected if \( \Delta u \) simply followed the size of the solvent molecule (see above). Nonetheless, the near doubling of \( \Delta u \) in pure glycerol is highly significant because it directly points to an integral structural role of solvent molecules in the unfolding transition state of the I27 protein. The unfolding transition state of a protein under force can be defined as that structure which, taken as a starting point, leads to either full unraveling or to a stable fold, with equal probability (43). Transition state structures are extremely short lived, typically requiring femtosecond laser spectroscopy for their capture (44). However, under a constant stretching force, the effect of mechanical work on the energy landscape of the unfolding protein is felt by the transition state structure, regardless of its lifetime. Thus, our finding that the distance to transition state of protein unfolding is sensitive to the size of the solvent suggests that solvent molecules are part of this short lived structure.

The steep dependency of \( \Delta u \) on the glycerol concentration can be understood by developing a simple model of solvent occupancy which does not require any additional information about the system, such as preferential solvation. We assume that there are \( N \) interaction sites that can be occupied by water molecules at the transition state structure, resulting in a value \( \Delta u \). We assume that if a single water molecule is replaced by a glycerol molecule, the transition state elongates to a value, \( \Delta u \). Under these conditions, the observed value of \( \Delta u \) for an ensemble of unfolding I27 proteins will be

\[
\Delta u_{[\text{gly}]}, \text{[gly]} = (P_w)^N \Delta u + (1 - (P_w)^N) \Delta u_{[\text{gly}]}, \tag{1}
\]

where the probability of occupancy by a water molecule \( P_w \) is defined in terms of the glycerol concentration as \( P_w = (1 - [\text{gly}] \), where \([\text{gly}]\) is the volume fraction of glycerol in the solution (Fig. 3). Therefore, \( (P_w)^N \) is the probability that \( N \) sites are occupied by water molecules. \( 1 - (P_w)^N \) corresponds to the probability that not all sites are occupied by water molecules, i.e., at least one water molecule is replaced by a glycerol molecule. If we set \( \Delta u_w = 2.5 \AA \) and \( \Delta u_{\text{gly}} = 4.4 \AA \), the measured values in pure water and pure glycerol, respectively, we can readily reproduce the steep dependency of \( \Delta u \) on glycerol. Fig. 3B shows plots of \( \Delta u_{[\text{gly}]}, \text{[gly]} \) for various values of \( N \) (\( N = 1 \), 3, 6, 9; black lines). Best fits to the measured values of \( \Delta u \) were obtained for \( N = 6 \) [\( \chi^2 = 1.5 \), \( v = 6 \), \( p(\chi^2) = 0.96 \)]. However, \( N = 7, 8 \) also had similar scores [see SI Table 1 containing \( \chi^2 \) and \( p(\chi^2) \) for all \( N \) from 1 to 10]. Remarkably, the optimal values of \( N \) are similar to the known number of hydrogen bonding sites (six) between \( \beta \)-strands \( \alpha' \) and \( G \), which are likely to be part of the unfolding transition state structure of the I27 protein (31–33, 39).

The exact nature of the transition state structure of I27 unfolding under a stretching force is unknown. Steered Molecular Dynamics (SMD) simulations can complement our AFM observations by providing a detailed atomic picture of stretching and unfolding of individual protein domains (29, 31, 32). SMD simulations involve the application of external forces to molecules in molecular dynamics simulations. The simulations are carried out by fixing one terminus of the protein and applying external forces to the other terminus (see SI Text for details). Earlier SMD simulations showed that the simultaneous rupture of six backbone hydrogen bonds between \( \beta \)-strands \( \alpha' \) and \( G \) of the I27 protein (Fig. 4A) was a necessary event in its mechanical unfolding (31–33). Furthermore, these simulations showed that the rupture of these interstrand hydrogen bonds could be followed by bonding to water molecules that formed bridges between the two separating strands (Fig. 4B). One way to interpret our results would be that the transition state structure
is then formed by six to eight water molecules bridging the gap between separating β-strands and taking the place of some of the broken interstrand hydrogen bonds. Our SMD simulations of forced unfolding of the I27 protein in glycerol solutions showed that the resistance to unfolding still originates from the same set of hydrogen bonds between the A’ and G β-strands. In glycerol solutions, the larger size of this cosolvent could lead to a greater gap between the separating β-strands (Fig. 4C). Given that there is a multitude of possible transition state structures formed by water, glycerol, and the protein backbone, there is no straightforward way to link a wider gap between the β-strands A’ and G in the simulations, and the experimentally measured values of Δxu. In Fig. 4 B and C, we define the pulling coordinate for the separating β-strands as the distance between the first amino acid of strand A’ (Y9) and the last amino acid of strand G (K87). This distance, x(Y9)–x(K87), gets longer as the two β-strands separate under a constant force (Fig. 4D), filling the gap with solvent molecules until a transition state is reached (Fig. 4D, arrows). The elongation of the x(Y9)–x(K87) distance up to the transition state is defined as the distance to transition state, ΔxA(G) (Fig. 4D). The crossing of the transition state is marked by an abrupt increase in the separation length (Fig. 4D, arrows) that leads to the complete unraveling of the protein. We repeated SMD simulations of I27 unfolding measured ΔxA(G) in water (Fig. 4E, black bars) and in 30% glycerol solutions (Fig. 4E, red bars). The simulations show that ΔxA(G) increases from 2.9 ± 0.6 Å in water, up to 3.9 ± 2.1 Å in glycerol, in qualitative agreement with our observations and with the simple solvent bridging model of the unfolding transition state. From these simulations, it is not possible to determine which exact structural snapshot corresponds to the transition state of unfolding. However, it is significant that we readily find glycerol molecules between the β-strands A’ and G during the early stages of separation (Fig. 4C). It is also apparent that water molecules are found bridging β-strands A’ and G during the early stages of separation (Fig. 4C). It is interesting to consider whether a pure water transition structure might also exist in glycerol solutions. If this were true, then a glycerol molecule bridging would simply create a more stable transition state structure that is further displaced along the measured coordinate than a pure water containing transition state structure, at lower energy states. Such water molecule containing structures would then become intermediate unfolding states in glycerol solutions. However, such intermediate structures would only be ~2 Å away from the glycerol bridging transition state and thus would be difficult to capture using our current instrumentation. Additionally, the measured steep dependency of Δxu on glycerol concentration (Fig. 3B) suggests that the range over which both a glycerol bridging transition state and a pure water transition state would be expected would be very narrow.

As a further test of the solvent bridging hypothesis, we replaced water with deuterium oxide (D2O), or heavy water. D2O is similar in size to water but forms hydrogen bonds that are stronger by ~0.1–0.2 kcal mol−1 at the same thermodynamic conditions of temperature and number density (45). Substituting water with D2O in our aqueous solutions increases the hydrogen bond strength of the solvent environment and enhances the stability of the protein (46). As before (Fig. 1 B and C), we measured ΔG0 and Δxu in the D2O solution by fitting an Arrhenius term (19) to the force dependency of the unfolding rate (Fig. 5). Our measurements showed that, although ΔG0 increased from 23.11 kcal mol−1 in aqueous solution to 24.67 kcal mol−1 in D2O aqueous solution, Δxu remained unchanged (Fig. 5B). This result lends further support to the solvent bridging model.

By combining single-molecule force spectroscopy techniques with solvent substitution, we have shown that solvent molecules form part of the mechanical transition state structure of a protein. Although we have demonstrated the crucial roles played by solvent molecules in titin, the giant elastic protein of muscle, it is likely that osmolytes also control the mechanical transition state structure of other proteins. Indeed, it will be interesting to elucidate the role of osmolytes in the mechanical transition state structure of proteins that have a distinct topology from the I27 protein. Interestingly, solvent composition is actively regulated in vivo. For example, a member of the aquaporin family of membrane channel proteins, GilpF, is highly selective for permeation of glycerol (47, 48), a naturally occurring osmolyte (26).

Thus, regulation of the cellular solvent composition may be an important mechanism under conditions of mechanical stress and/or mechanical injury (34) where sustained mechanical forces applied to tissues may trigger widespread protein unfolding. A rapid compensatory increase of the cellular osmolyte concentration may therefore be “mechanically” protective.

**Materials and Methods**

**Protein Engineering and Purification.** To allow for efficient single-molecule experiments, we first constructed polyproteins using protein engineering. The details of the polyprotein engineering and purification were reported (28). Briefly, we constructed an eight-domain N–C-linked polyprotein of I27, the 27th Ig-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).

**Solvent Environment.** Samples of glycerol (~99%), deuterium oxide, and water were obtained from Sigma-Aldrich and used without additional purification. Solvent mixtures were prepared to obtain the desired volume fraction, vol/vol, ratio of the cosolvent and viscosity. Viscosities were confirmed by using a falling ball viscometer.

**Single-Molecule Force Spectroscopy.** We used a custom-built atomic force microscope equipped with a PicoCube P363.3-CD piezoelectric translator (Physik Instrumente, Karlsruhe, Germany) controlled by an analog PID feedback system described in refs. 16 and 28. Silicon nitride cantilevers (Veeco, Santa Barbara, CA) were calibrated for their spring constant using the equipartition theorem as reported (49). The average spring constant was ~60 pN/nm for force-extension experiments and ~15 pN/nm for force-clamp experiments. All data were obtained and analyzed using custom software written for use in Igor 5.0 (Wavemetrics, Oswego, OR). There was ~0.5 nm of peak-to-peak noise and a feedback response time of ~5 ms in all experiments. To estimate the error on our experimentally obtained rate constant, we carried out the nonparametric bootstrap method (35).

**ACKNOWLEDGMENTS.** We thank Arun Wila, Sergi Garcia-Manyes for careful reading of the manuscript and Bruce Berne, Devarajan Thirumalai, and members of the Fernandez lab for helpful discussions. This work was supported by National Institutes of Health Grants HL66030 and HL61228 (to J.M.F.) and P01 AI 060915 (to H.L.).
Supporting Text

Biphasic Mechanical Stability
We extend the I278 polyprotein using force extension spectroscopy, retracting the sample holding substrate away from the cantilever tip at a constant velocity of 400 nm s^{-1} and 50 nms^{-1}. The saw-tooth pattern of peaks that is observed corresponds to sequential unraveling of individual domains of a modular protein (Fig. 2A). A plot of the peak of the unfolding force distribution as a function of the concentration of glycerol is shown in SI Fig. 6, at 400 nms^{-1} (red squares) and 50 nms^{-1} (black circles). The mechanical stability of the I27 protein in aqueous glycerol solutions displays a biphasic force dependence for unfolding at both pulling velocities.

To extract kinetic information from the force-extension experiments, Monte Carlo simulations (described below) were completed to compile the force required to unfold a domain during simulated extensions (1). The Monte Carlo simulations have to predict both the distribution of unfolding forces (Fig. 2B) and also simultaneously predict how this distribution shifts with the pulling velocity (1-3), generating good estimates for the size of the activation energy barrier, \( \Delta G_U \), and the distance from the native state to the transition state along the pulling direction \( \Delta x_u \).

We can use these estimates for \( \Delta G_U \) and \( \Delta x_u \) to calculate the most likely force of unfolding, \( F_{\text{max}} \),

\[
F_{\text{max}} = \frac{k_B T}{\Delta x} \ln \left[ \frac{\Delta x_u k_B \nu}{k_u^0 k_B T} \right] \tag{1}
\]

where \( k_c \) is the spring constant of the cantilever, \( \nu \) is the pulling velocity and \( k_u^0 \) is the unfolding rate at zero force, where \( k_u^0 = A \exp(-\Delta G_U/k_B T) \). In SI Fig. 6 the dashed line shows the expected force dependence, \( F_{\text{max}} \), at a pulling speed of 400 nms^{-1} (red squares) and 50 nms^{-1} (black circles). Clearly the values of \( F_{\text{max}} \) obtained using equation 1 do not exactly agree with the measured values of the peak of the unfolding force distribution. This is unsurprising given that the equation 1 assumes a constant velocity whereas in
force-extension experiments the force varies dynamically throughout the experiment, making it difficult to accurately measure force-dependent parameters. Nevertheless, the biphasic mechanical stability of the protein I27 in aqueous glycerol is evident.

Monte Carlo Simulations

The values of both $\Delta G_U$ and $\Delta x_u$ can be estimated indirectly through the use of Monte Carlo techniques that simulate the unfolding of a polyprotein pulled at constant velocity. The aim of the Monte Carlo simulations is to find a set of unfolding parameters, $k_U^0$ and $\Delta x_u$ that can account for the distribution of unfolding forces (red traces in the histograms of Fig. 2B) measured from the sawtooth pattern data (1, 4, 5).

In the simulation, force-extension curves were generated by assuming the polyprotein to be well described by the Worm Like Chain (WLC) model of polymer elasticity (6), and the cantilever to be a linear spring (1). The unfolding of a domain was modeled as a two state Markovian process. To determine if a still folded protein molecule will unfold at a particular force, a probability is calculated according to the theory developed by Bell (7) and elaborated by Evans and Ritchie (8) for two-state unfolding:

$$P = P(F)\Delta t = A \exp\left(-\frac{\Delta G_U - F\Delta x_u}{k_BT}\right)\Delta t = k_U^0 \exp\left(\frac{F\Delta x_u}{k_BT}\right)\Delta t$$  (2)

where $\Delta G_U$ is the activation energy barrier for unfolding (at zero force), $F$ is the pulling force experienced by the protein, $\Delta x_U$ is the distance between the folded state and the transition state along the pulling direction, $A$ is an attempt frequency, $k_U^0$ is the unfolding rate when no external force is present and $\Delta t$ is the time interval over which force is acting on the protein. At each force level, each folded molecule in the polyprotein is polled for unfolding by comparing the unfolding probability with a randomly generated number, before the chain is pulled further. The Monte Carlo simulations have to predict both the distribution of unfolding forces (Fig. 2B) and also simultaneously predict how this distribution shifts with the pulling velocity (SI Fig 6) (1-3). The values of $k_U^0$ and $\Delta x_U$ for the I27 protein were obtained as adjusting parameters that best fit the simulation results to both sets of experimental data (pulling velocity of 400 nms$^{-1}$ and 50 nms$^{-1}$). This stringent criteria can generate good estimates for the size of the activation energy barrier, $\Delta G_U$, and the distance from the native state to the transition state along the pulling direction $\Delta x_u$.

Effect of viscous drag on the AFM cantilever

A solution with a large quantity of glycerol becomes very viscous. The viscosity in glycerol solutions increases exponentially at high glycerol content. For a 30 % v/v glycerol-water solution the viscosity has increased from 0.89 cP to 3.09 cP at 20 °C, while in a 100 % v/v glycerol solution the viscosity has increased to 934 cP. The measured forces are underestimated due to the viscous drag on the AFM cantilever, and
this underestimation is more severe at high glycerol concentration and higher pulling speeds.

Previously a scaled spherical model has been developed to assess the effect of the hydrodynamic drag on AFM cantilevers for microrheological measurements in liquid at low 16 Reynolds numbers (9). This model has then been used to measure the viscous drag on the unfolding forces of a multi-domain protein in fast pulling experiments with pulling speeds in the range of 5 - 140 µms\(^{-1}\) (10). These authors showed that the unfolding forces could be underestimated by more than 20 % due to viscous drag at a pulling speed of 30 µms\(^{-1}\). In contrast to the experiments discussed above, our experiments were carried out at lower pulling speeds of 50 nms\(^{-1}\) and 400 nms\(^{-1}\), and the unfolding force underestimation due to viscous drag on the cantilever was found not to be significant in the force extension experiments.

**Solvent Bridging Model**

The dependency of \(\Delta x_u\) on the glycerol concentration \([\text{gly}]\) can be understood by developing a simple model of solvent occupancy. The observed value of \(\Delta x_u[\text{gly}]\) for an ensemble of unfolding I27 proteins will be:

\[
\Delta x_u[\text{gly}] = (P_w)^N \Delta x_w + \left(1 - (P_w)^N\right) \Delta x_G
\]

Where the probability of occupancy by a water molecule \(P_w\) is defined in terms of the glycerol concentration as \(P_w = (1-[\text{gly}])\), where \([\text{gly}]\) is the volume fraction of glycerol in the solution (Fig. 3B) and \(N\) is the number of interaction sites. \((P_w)^N\) is the probability that \(N\) sites are occupied by water molecules, while \(1-(P_w)^N\) corresponds to the probability that not all sites are occupied by water molecules. \(\Delta x_w\) and \(\Delta x_G\) are the measured values of the distance to the unfolding transition state in pure water and pure glycerol respectively. Table S1 shows \(\chi^2\) and \(p(\chi^2)\) where \(v = 6\) for the solvent bridging model for interaction sites \(N = 1-10\).

**Steered molecular dynamics simulations**

The I27 protein was subject to a simulated equilibration, constant velocity SMD, and constant force SMD. The aqueous environment was modeled using explicit water with periodic boundary conditions. The water box was large enough for equilibration and for the first 50 Å of stretching (length 135 Å, width 68 Å, height 68 Å). The whole protein-water system contains ~59,300 atoms. The 30% glycerol box has the same size as the pure water box. Since the CHARMM force field parameters are not available for glycerol molecules, they are built from parameters used in lipid head which are provided in CHARMM (11). This procedure is similar to a published work in simulating aquaglyceroporin GlpF (12). The ratio of the numbers of glycerol and water molecules is set to be 1:9, corresponding to 30% volume concentration. Starting from the water box, a python script is written to randomly assign 1227 glycerol molecules in that box to replace waters. The corresponding molecular structure file (.psf) was generated by psfgen in VMD based on the structure of the I27 protein and glycerol molecules. The
total system of protein-water-glycerol contains 60165 atoms. The velocities used in constant velocity SMD simulations were in the range between 0.1m/s to 1m/s, 5-6 orders of magnitude larger than pulling velocities used in AFM experiments. The I27 protein was also stretched by the clamped forces at 600, 800, 900 and 1000 pN, separately, in the constant force SMD simulations. The model preparation and data analysis were done with VMD (13) and MD simulation with NAMD (14). During the 1 nanosecond equilibration the protein is reasonably stable and did not deviate from the initial PDB structure 1TIT, with the RMSD below 1.6 Å. That final structure from the equilibration was the starting structure in the constant velocity and constant force SMD.

**Force-clamp unfolding of the protein I27 in deuterium oxide**

Under force-clamp conditions, stretching a polyprotein in deuterium oxide solution results in a well defined series of step increases in length, marking the unfolding and extension of the individual modules in the chain (15). We measure the unfolding rate by fitting a single exponential to an average of 20 traces similar to the ones shown in Fig. 1A and obtain an estimate of the standard error of \( r_0(F) \), using the bootstrapping technique (16, 17). We repeated these measurements over the force range between 140 pN and 200 pN to obtain the force-dependency of the unfolding rate in deuterium oxide solution (SI Fig. 7).
Supporting Information for

**Solvent molecules bridge the mechanical unfolding transition state of a protein**

**Lorna Dougan**¹, **Gang Feng**², **Hui Lu**² and **Julio M Fernandez**¹

¹Biological Sciences, Columbia University, New York, NY, 10027, ²Department of Bioengineering, University of Illinois, Chicago, Illinois, 60607

**Biphasic Mechanical Stability**
We extend the I27₈ polyprotein using force extension spectroscopy, retracting the sample holding substrate away from the cantilever tip at a constant velocity of 400 nm s⁻¹ and 50 nms⁻¹. The saw-tooth pattern of peaks that is observed corresponds to sequential unraveling of individual domains of a modular protein (Fig. 2A). A plot of the peak of the unfolding force distribution as a function of the concentration of glycerol is shown in Fig. S1, at 400 nms⁻¹ (red squares) and 50 nms⁻¹ (black circles). The mechanical stability of the I27 protein in aqueous glycerol solutions displays a biphasic force dependence for unfolding at both pulling velocities.

To extract kinetic information from the force-extension experiments, Monte Carlo simulations (described below) were completed to compile the force required to unfold a domain during simulated extensions (1). The Monte Carlo simulations have to predict both the distribution of unfolding forces (Fig. 2B) and also simultaneously predict how this distribution shifts with the pulling velocity (1-3), generating good estimates for the size of the activation energy barrier, ΔGᵤ, and the distance from the native state to the transition state along the pulling direction Δxᵤ.

We can use these estimates for ΔGᵤ and Δxᵤ to calculate the most likely force of unfolding, Fₘₐₓ,

\[ F_{max} = k_c T \ln \left( \frac{\Delta x_u k_c \nu}{k_u^0 k_B T} \right) \]  

(1)

where \( k_c \) is the spring constant of the cantilever, \( \nu \) is the pulling velocity and \( k_u^0 \) is the unfolding rate at zero force, where \( k_u^0 = A \exp(-\Delta G_u/k_B T) \). In Fig. S1 the dashed line shows the expected force dependence, \( F_{max} \), at a pulling speed of 400 nms⁻¹ (red squares) and 50 nms⁻¹ (black circles). Clearly the values of \( F_{max} \) obtained using equation 1 do not exactly agree with the measured values of the peak of the unfolding force distribution. This is unsurprising given that the equation 1 assumes a constant velocity whereas in
force-extension experiments the force varies dynamically throughout the experiment, making it difficult to accurately measure force-dependent parameters. Nevertheless, the biphasic mechanical stability of the protein I27 in aqueous glycerol is evident.

Monte Carlo Simulations

The values of both $\Delta G_U$ and $\Delta x_u$ can be estimated indirectly through the use of Monte Carlo techniques that simulate the unfolding of a polyprotein pulled at constant velocity. The aim of the Monte Carlo simulations is to find a set of unfolding parameters, $k_U^0$ and $\Delta x_u$ that can account for the distribution of unfolding forces (red traces in the histograms of Fig. 2B) measured from the sawtooth pattern data (1, 4, 5).

In the simulation, force-extension curves were generated by assuming the polyprotein to be well described by the Worm Like Chain (WLC) model of polymer elasticity (6), and the cantilever to be a linear spring (1). The unfolding of a domain was modeled as a two state Markovian process. To determine if a still folded protein molecule will unfold at a particular force, a probability is calculated according to the theory developed by Bell (7) and elaborated by Evans and Ritchie (8) for two-state unfolding:

$$P = P(F)\Delta t = A \exp \left( -\frac{\Delta G_U - F\Delta x_u}{k_B T} \right) \Delta t = k_U^0 \exp \left( \frac{F\Delta x_u}{k_B T} \right) \Delta t$$ (2)

where $\Delta G_U$ is the activation energy barrier for unfolding (at zero force), $F$ is the pulling force experienced by the protein, $\Delta x_U$ is the distance between the folded state and the transition state along the pulling direction, $A$ is an attempt frequency, $k_U^0$ is the unfolding rate when no external force is present and $\Delta t$ is the time interval over which force is acting on the protein. At each force level, each folded molecule in the polyprotein is polled for unfolding by comparing the unfolding probability with a randomly generated number, before the chain is pulled further. The Monte Carlo simulations have to predict both the distribution of unfolding forces (Fig. 2B) and also simultaneously predict how this distribution shifts with the pulling velocity (Fig S1) (1-3). The values of $k_U^0$ and $\Delta x_u$ for the I27 protein were obtained as adjusting parameters that best fit the simulation results to both sets of experimental data (pulling velocity of 400 nm/s and 50 nm/s). This stringent criteria can generate good estimates for the size of the activation energy barrier, $\Delta G_U$, and the distance from the native state to the transition state along the pulling direction $\Delta x_u$.

Effect of viscous drag on the AFM cantilever

A solution with a large quantity of glycerol becomes very viscous. The viscosity in glycerol solutions increases exponentially at high glycerol content. For a 30 % v/v glycerol-water solution the viscosity has increased from 0.89 cP to 3.09 cP at 20℃, while in a 100 % v/v glycerol solution the viscosity has increased to 934 cP. The measured forces are underestimated due to the viscous drag on the AFM cantilever, and
this underestimation is more severe at high glycerol concentration and higher pulling speeds.

Previously a scaled spherical model has been developed to assess the effect of the hydrodynamic drag on AFM cantilevers for microrheological measurements in liquid at low 16 Reynolds numbers (9). This model has then been used to measure the viscous drag on the unfolding forces of a multi-domain protein in fast pulling experiments with pulling speeds in the range of 5 - 140 µms\(^{-1}\) (10). These authors showed that the unfolding forces could be underestimated by more than 20 % due to viscous drag at a pulling speed of 30 µms\(^{-1}\). In contrast to the experiments discussed above, our experiments were carried out at lower pulling speeds of 50 nm\(^{-1}\) and 400 nm\(^{-1}\), and the unfolding force underestimation due to viscous drag on the cantilever was found not to be significant in the force extension experiments.

**Solvent Bridging Model**

The dependency of \(\Delta x_u\) on the glycerol concentration \([\text{gly}]\) can be understood by developing a simple model of solvent occupancy. The observed value of \(\Delta x_u[\text{gly}]\) for an ensemble of unfolding I27 proteins will be:

\[
\Delta x_u[\text{gly}] = (P_w)^N \Delta x_w + \left(1 - (P_w)^N\right) \Delta x_G
\]

Where the probability of occupancy by a water molecule \(P_w\) is defined in terms of the glycerol concentration as \(P_w = (1-[\text{gly}])\), where \([\text{gly}]\) is the volume fraction of glycerol in the solution (Fig. 3B) and \(N\) is the number of interaction sites. \((P_w)^N\) is the probability that \(N\) sites are occupied by water molecules, while \(1-(P_w)^N\) corresponds to the probability that not all sites are occupied by water molecules. \(\Delta x_w\) and \(\Delta x_G\) are the measured values of the distance to the unfolding transition state in pure water and pure glycerol respectively. Table S1 shows \(\chi^2\) and \(p(\chi^2)\) where \(v = 6\) for the solvent bridging model for interaction sites \(N = 1\)-10.

**Steered molecular dynamics simulations**

The I27 protein was subject to a simulated equilibration, constant velocity SMD, and constant force SMD. The aqueous environment was modeled using explicit water with periodic boundary conditions. The water box was large enough for equilibration and for the first 50 Å of stretching (length 135 Å, width 68 Å, height 68 Å). The whole protein-water system contains ~59,300 atoms. The 30% glycerol box has the same size as the pure water box. Since the CHARMM force field parameters are not available for glycerol molecules, they are built from parameters used in lipid head which are provided in CHARMM (11). This procedure is similar to a published work in simulating aquaglyceroporin GlpF (12). The ratio of the numbers of glycerol and water molecules is set to be 1:9, corresponding to 30% volume concentration. Starting from the water box, a python script is written to randomly assign 1227 glycerol molecules in that box to replaces waters. The corresponding molecular structure file (.psf) was generated by psfgen in VMD based on the structure of the I27 protein and glycerol molecules. The
total system of protein-water-glycerol contains 60165 atoms. The velocities used in constant velocity SMD simulations were in the range between 0.1 m/s to 1 m/s, 5-6 orders of magnitude larger than pulling velocities used in AFM experiments. The I27 protein was also stretched by the clamped forces at 600, 800, 900 and 1000 pN, separately, in the constant force SMD simulations. The model preparation and data analysis were done with VMD (13) and MD simulation with NAMD (14). During the 1 nanosecond equilibration the protein is reasonably stable and did not deviate from the initial PDB structure 1TIT, with the RMSD below 1.6 Å. That final structure from the equilibration was the starting structure in the constant velocity and constant force SMD.

**Force-clamp unfolding of the protein I27 in deuterium oxide**

Under force-clamp conditions, stretching a polyprotein in deuterium oxide solution results in a well defined series of step increases in length, marking the unfolding and extension of the individual modules in the chain (15). We measure the unfolding rate by fitting a single exponential to an average of 20 traces similar to the ones shown in Fig. 1A and obtain an estimate of the standard error of \( r_u(F) \), using the bootstrapping technique (16, 17). We repeated these measurements over the force range between 140 pN and 200 pN to obtain the force-dependency of the unfolding rate in deuterium oxide solution (Fig. S2).
<table>
<thead>
<tr>
<th>Interaction Sites N</th>
<th>$\chi^2$</th>
<th>P($\chi^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67.56</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>29.96</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>14.80</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>7.57</td>
<td>0.0059</td>
</tr>
<tr>
<td>5</td>
<td>3.96</td>
<td>0.0465</td>
</tr>
<tr>
<td>6</td>
<td>1.50</td>
<td>0.22</td>
</tr>
<tr>
<td>7</td>
<td>1.58</td>
<td>0.21</td>
</tr>
<tr>
<td>8</td>
<td>1.77</td>
<td>0.18</td>
</tr>
<tr>
<td>9</td>
<td>2.24</td>
<td>0.13</td>
</tr>
<tr>
<td>10</td>
<td>2.25</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table 1: Solvent bridging model with N interaction sites. Values are shown for all $\chi^2$ and P($\chi^2$) where $\nu = 6$
Figure 1: Biphasic force dependence for unfolding of the I27 protein in glycerol solutions at different pulling velocities: The peak of the unfolding force distribution of the I27 protein as a function of concentration of glycerol, v/v, measured from force-extension unfolding at different pulling velocities. The mechanical stability of the I27 protein in aqueous glycerol solutions displays a biphasic force dependence for unfolding at a pulling speed of 400 nms$^{-1}$ (red squares) and 50 nms$^{-1}$ (black circles). The dashed line shows the expected force dependence at a pulling speed of 400 nms$^{-1}$ (red squares) and 50 nms$^{-1}$ (black circles) using a simple Arrhenius term ($\lambda$) = 0.44 nm.
Figure 2: Force-clamp protein unfolding in aqueous solution and deuterium oxide aqueous solution (A) Multiple trace averages (n > 20 in each trace) of unfolding events measured using force-clamp spectroscopy for the I27 protein in: aqueous solution (light blue circles) for constant force measurements at 200 pN, 180 pN, 160 pN, 140 pN and 120 pN, and deuterium oxide (D$_2$O) aqueous solution (dark blue squares) at 200 pN, 180 pN, 160 pN and 140 pN.
Figure 2: Force-clamp protein unfolding in aqueous solution and deuterium oxide aqueous solution (A) Multiple trace averages (n > 20 in each trace) of unfolding events measured using force-clamp spectroscopy for the I27 protein in: aqueous solution (light blue circles) for constant force measurements at 200 pN, 180 pN, 160 pN, 140 pN and 120 pN, and deuterium oxide (D_2O) aqueous solution (dark blue squares) at 200 pN, 180 pN, 160 pN and 140 pN