# Contrasting the Individual Reactive Pathways in Protein Unfolding and Disulfide Bond Reduction Observed within a Single Protein

Sergi Garcia-Manyes,*[,†] Tzu-Ling Kuo,[‡] and Julio M. Fernández*[,†]

[†]Department of Biological Sciences and [‡]Department of Physics, Columbia University, New York, New York 10027, United States

S *Supporting Information*

ABSTRACT: Identifying the dynamics of individual molecules along their reactive pathways remains a major goal of modern chemistry. For simple chemical reactions, the transition state position is thought to be highly localized. Conversely, in the case of more complex reactions involving proteins, the potential energy surfaces become rougher, resulting in heterogeneous reaction pathways with multiple transition state structures. Force-clamp spectroscopy experimentally probes the individual reaction pathways sampled by a single protein under the effect of a constant stretching force. Herein, we examine the distribution of conformations that populate the transition state of two different reactions; the unfolding of a single protein and the reduction of a single disulfide bond, both occurring within the same single protein. By applying the recently developed static disorder theory, we quantify the variance of the barrier heights, σ^2, governing each distinct reaction. We demonstrate that the unfolding of the I27 protein follows a nonexponential kinetics, consistent with a high value of $\sigma^2 \sim 18$ $(pN\ nm)^2$. Interestingly,

shortening of the protein upon introduction of a rigid disulfide bond significantly modulates the disorder degree, spanning from $\sigma^2 \sim 8$ to ~ 21 $(pN\ nm)^2$. These results are in sharp contrast with the exponential distribution of times measured for an S_N2 chemical reaction, implying the absence of static disorder $\sigma^2 \sim 0$ $(pN\ nm)^2$. Our results demonstrate the high sensitivity of the force-clamp technique to capture the signatures of disorder in the individual pathways that define two distinct force-induced reactions, occurring within the core of a single protein.
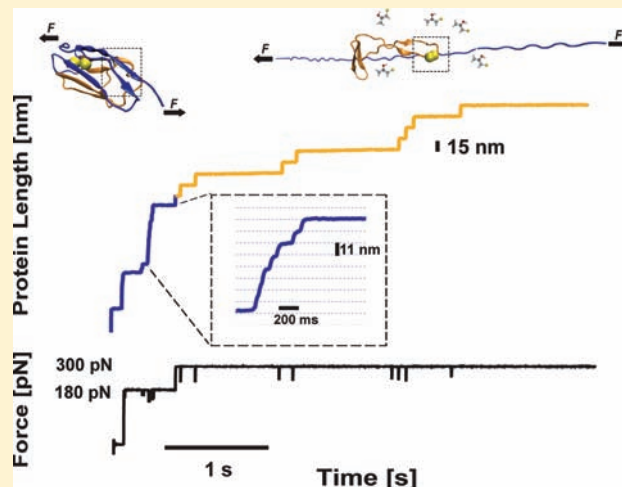
INTRODUCTION

Elucidating the dynamics of reactant molecules along the particular pathway that connects them with the resulting products has become a cornerstone in modern chemistry, allowing detailed reconstruction of the potential energy surface of a chemical reaction. In general, for small systems composed of only a few molecules, the valley connecting reactants with products is narrow, the reaction pathways are well-defined, and the position of the transition state is highly localized.[1,2] For example, the study of elementary chemical reactions such as the simple S_N2 nucleophilic substitution, mainly using crossed molecular beams and ab initio calculations, has provided detailed information regarding the position and height of the free-energy barriers determining the process.[3−7] By contrast, in the case of more complex reactions involving bigger molecular systems such as proteins, the potential energy surfaces become rougher, which results in heterogeneous reaction pathways with multiple transition state structures.[1] The complexity of the resulting energy landscapes has been mainly revealed by the use of computational methods, which sample the individual (un)folding trajectories of a single protein molecule while under particular force-field conditions.[8−14] The high degree of static and dynamic heterogeneity between individual trajectories required statistical mechanical tools to interpret these data.[12,13] In sharp contrast with the in silico observations, the macroscopic nature of traditional protein folding experiments conducted in the bulk averages out the pathways in the ensemble, thus resulting in a smooth free-energy landscape that mainly captures two stable thermodynamic states, namely, the folded and unfolded conformations.[15−17] The conversion between these two states has been traditionally given the treatment of a first order chemical reaction, implying that the concentration of the reactant species—the protein native state—decreases exponentially with time.[15] This observation entails that each individual molecule has a well-defined and unique native state, the unfolding path is the same, and the saddle point defining the transition state is conserved.

The advent of single molecule techniques, primarily fluorescence spectroscopy, has allowed measurement of the fluctuation dynamics of single enzyme molecules over time.[18] These experiments revealed a wide range of catalytic activity within individual enzymes present in the ensemble population ("static disorder") and also large variations in the catalytic activity of an individual enzyme ("dynamic disorder").[18] Such a complex behavior, invisible to bulk experiments, manifests itself in the nonexponential kinetics measured for different enzymes.[19,20] Theoretical works describing these experiments have carefully identified the main features of the potential energy surfaces (often exhibiting multiple saddle points) that govern the large number of reactive paths measured in the experiment.[21−29] The combination of experimental and theoretical work is therefore required to ascribe the complexity in the kinetics of catalysis to the molecular mechanisms involved in the conformational changes of the single enzyme over time.[20,30] The complex kinetics experimentally measured for large biomolecules such as enzymes contrasts with the apparent simplicity of chemical reactions with small molecules, which now can also be investigated one molecule at a time.[31,32]

Thus, it becomes clear that the emergence of single molecule techniques provides a new vista on chemical reactivity. Of particular relevance, single molecule force-clamp spectroscopy excels at probing the conformational dynamics of a single protein under force.[33] By applying a calibrated constant force to a single protein, the individual unfolding pathways of single proteins can be directly monitored.[34] In these experiments, the multidimensional energy surface can be projected onto a single dimension, the end-to-end length of the stretched protein, which defines the reaction coordinate with sub-Ångström resolution. Using a large pool of individual unfolding trajectories, force-clamp measurements revealed that the time course of unfolding of the small protein ubiquitin clearly departed from the single exponential survival probability associated with a simple two-state unfolding scenario.[35−37] Follow-up experiments demonstrated an increased broadening in the distribution of unfolding pathways of ubiquitin as the constant pulling force was increased, scaling with F^2. These latter results were interpreted in the framework of static disorder in the classical Arrhenius description.[38] This theory provides the means to quantify the variance, σ^2, in the amount of disorder in the barrier heights with respect to average value assuming a Gaussian distribution. The implications derived from the unfolding kinetics obtained from our single molecule experiments are deep, indicating that force-clamp spectroscopy can be now used to directly probe the broad conformational diversity defining the native state of proteins using the mechanical stability of each conformation as its structural fingerprint.

Contrary to these observations on the ubiquitin protein, mechanical unfolding of protein G resulted in a simple all-or-none two-state transition, consistent with an exponentially distributed unfolding times at a given constant force.[39] These results suggest that the degree of disorder in the unfolding pathways measured for a single protein under force is likely to be highly protein specific, varying over a wide range. The challenge lies now on discovering the molecular origin of the nonexponential kinetics encountered in mechanical unfolding of proteins and to distinguish it from a distinct molecular process that follows first-order kinetics occurring under the same experimental conditions, within the structure of a single protein.

Besides studying protein (un)folding, the combination of protein engineering and force-clamp spectroscopy has allowed direct examination of the effect of a constant force on the potential energy surface of a chemical reaction, at the single bond level.[40] Our results have demonstrated that the reduction of a single disulfide bond embedded within the core of a protein is activated by the pulling force in the presence of nucleophiles.[41] These experiments have permitted reconstruction of the energy landscape of a chemical reaction under force, allowing capture of the most plausible conformation of the reaction transition state.[41−43] Herein, we compare the effect of a mechanical constant force on the energy landscape of two different reactions: the unfolding of a single protein and the reduction of a single disulfide bond by a small nucleophile, both occurring within the same single protein. Our results demonstrate that, similar to ubiquitin, the transition state of unfolding of the 27th immunoglobulin-like domain of cardiac titin (I27), the mechanical properties of which have been amply studied,[44,45] is not unique but instead is composed of a set of similar structures. Interestingly, shortening of the I27 protein by the presence of disulfide bonds has a strong effect on the measured heterogeneity in the unfolding pathways. Our results are quantitatively discussed within the framework of the static disorder theory. In sharp contrast with these results, the energy landscape of an S_N2 chemical reaction, occurring within a much shorter length scale, is smooth, implying a high degree of homogeneity within the individual reactive pathways connecting the reactants with the reaction products. Our results demonstrate the ability of the force-clamp technique, through the acquisition of a large statistical pool of data, to uncover fine details of the ensemble of transition state structures populating the saddle point of the potential energy surface governing protein (un)folding and chemical reactions, at the single molecule level.

■ MATERIALS AND METHODS

Protein Engineering. PolyI27 ($I27_8$) and its mutants were constructed by consecutive subcloning of the respective monomers, using the *Bam*HI and *Bgl*II restriction sites.[44] The disulfide bond mutants were constructed using a cysteine-free I27 protein in which native Cys[47] and Cys[63], which do not form a disulfide bond, were mutated to alanines. Using the QuickChange site-directed mutagenesis kit (Stratagene), we introduced additional mutations to form disulfide bonds at specific positions within the I27 protein sequence: Gly[32] to Cys, Ala[75] to Cys ($I27_{G32C−A75C}$); Pro[28] to Cys, Lys[54] to Cys ($I27_{P28C−K54C}$); and Glu[24] to Cys, Lys[55] to Cys ($I27_{E24C−K55C}$).[46] As in the case of the wt-($I27_8$) protein, we constructed an eight-domain N−C linked polyprotein for each I27 mutant. The eight-domain polyproteins were cloned into the pQE80L (Qiagen) expression vector and transformed into the BLR DE3 *Escherichia coli* expression strain. Each polyprotein construct was finally purified by histidine metal-affinity chromatography with Talon resin (Clontech) and by gel filtration using Superdex 200 HR column (GE BioSciences).

Force Spectroscopy. Force-clamp AFM experiments were conducted at room temperature using a homemade setup under force-clamp conditions described elsewhere.[34] The sample was prepared by depositing 1−10 mL of protein in PBS solution (at a concentration of 1−10 mg mL^{-1}) onto a freshly evaporated gold coverslide. Each cantilever (Si_3N_4 Veeco MLCT-AUHW) was individually calibrated using the equipartition theorem, which gave a typical spring constant of 15 pN/nm. Single proteins were picked up from the surface by pushing the cantilever onto the surface with a contact force of 500−1000 pN to promote the nonspecific adhesion of the proteins on the cantilever surface. The
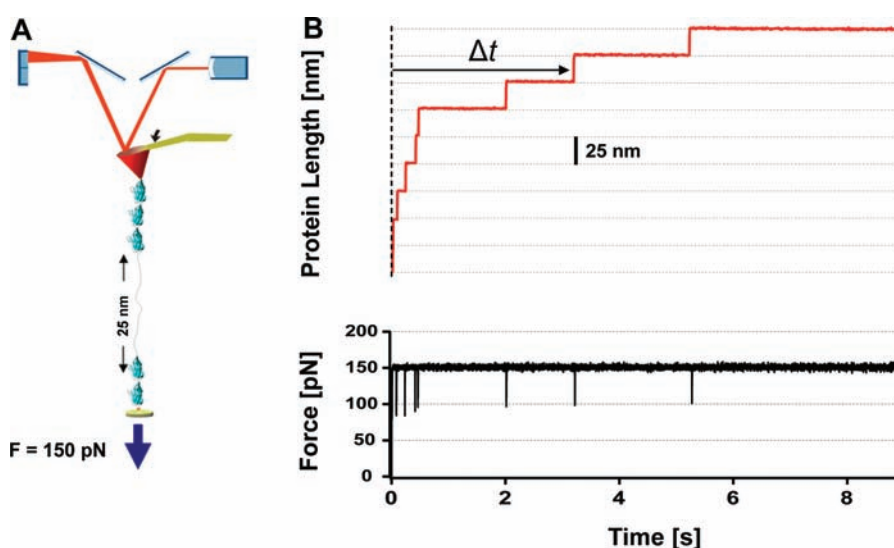
Figure 1. Unfolding the wild-type I27$_8$ polyprotein in a force clamp. (A) Schematics of the experimental setup. (B) Unfolding the I27 polyprotein under a constant force of 150 pN results in a staircase-like elongation, where the unfolding of each monomer in the chain occurs stochastically at a time Δt after the application of force, eliciting steps of $\sim$25 nm in length.

piezoelectric actuator was then retracted to produce a set deflection (force), which was kept constant throughout the experiment using an external, active feedback mechanism, while the extension was recorded. The force feedback was based on a proportional, integral, and differential amplifier, the output of which was fed to the piezoelectric positioner. The feedback response is limited to $\sim$2−5 ms. The high resolution piezoelectric actuator allowed our measurements of protein length a peak-to-peak resolution of $\sim$0.5 nm. Data from the force traces were filtered using a pole Bessel filter at 1 kHz. The reaction rate constants that we measured ($\sim$ 1 s^{-1}) were chosen such that they were not compromised by the feedback bandwidth or cantilever drift occurring over much longer time scales. In the chemical reduction experiments, a PBS solution containing 10 mM 1-mercapto-2-propanol was used, and the pH value was adjusted to 7.4.

Data Analysis. All data were recorded and analyzed using custom software written in Igor Pro 6.0 (Wavemetrics). The fingerprint of a single polyprotein in our unfolding experiments was considered to be at least six well-resolved steps of the corresponding height in each construct exhibiting long detachment times in order not to bias the unfolding probability.[37] In the case of the reduction experiments, no traces that included unfolding events during the second force pulse were included in the analysis. Only data with more than five reduction events in the second pulse were included in the analysis. The survival probability is defined as $\langle S(t)\rangle = 1 - \int_0^t p(t)\,dt$, where $p(t)\,dt$ corresponds to the probability of unfolding during the period of time between t and $t + dt$. The survival probability at each force is obtained from the associated probability density histogram of unfolding times, $p(t)$. The integral in the fitting function (eq 2) is evaluated by taking a sum of the integrand over the interval $[-6\sigma, 6\sigma]$, with the interval width of $dr = 0.001$ pN nm. All of the fitting procedures are performed by the Levenberg−Marquardt least-squares algorithm, implemented in the Igor Pro 6 software package. The errors of k and σ^2 are estimated by the bootstrap method.[47]

■ RESULTS

In our force-clamp essay, we make use of polyproteins composed of eight identical domains of the I27 protein (Figure 1A). Under a constant force of 150 pN, each individual protein module in the polyprotein chain unfolds, yielding a step increase in length of $\sim$25 nm (Figure 1B and Figure 2A, red trace).[37] The resulting unfolding trajectories resemble a staircase,

where the unfolding of each particular domain occurs stochastically at a time Δt from the moment that the force is applied to the protein (vertical dotted line in Figure 1B). By capturing 319 individual unfolding trajectories such as those shown in Figure 1B and Figure 2A, we collected 2296 independent unfolding events, the dwell time distribution of which is shown in Figure 2B. The average unfolding rate calculated as $k = 1/\langle t\rangle$, where $\langle t\rangle$ corresponds to the average time for unfolding and is $k = 0.83$ s^{-1}. However, a single exponential distribution with a rate constant of 0.83 s^{-1} (black line in Figure 2B) fails to capture the full time distribution of events. Similar conclusions can be drawn from the cumulative distribution of unfolding times shown in Figure SI1A in the Supporting Information; the single exponential fit to the experimental data (dotted black line) does not capture the distribution of unfolding times. Notably, the experimental distribution of unfolding times shows a high population of events that occur faster than that predicted by a single exponential distribution as shown in the normalized residuals plot in Figure SI1A in the Supporting Information (inset) and also in Figure 2B.

Unlike classical unfolding experiments using chemicals or large temperature jumps, the mechanical unfolding of proteins is a highly localized process mainly involving the rupture of a few key hydrogen bonds within the structure of the protein native state, which constitute the crucial structural motif that provides the protein with mechanical stability.[48,49] In the case of the I27 protein, such a mechanical clamp is placed between the A′G strands, encompassing amino acids 9−15 and 83−87.[49] While mutations in the amino acids constituting the protein transition state are known to significantly alter the mechanical stability of the protein,[48] the nature of the amino acids composing the peripheral regions beyond such a mechanical clamp and the total length of the protein are generally believed not to considerably affect its mechanical resistance. However, it is plausible that changes in the protein length involve little rearrangements in the conformation of the protein's native state, which would remain invisible to ensemble experiments, especially when the 3D crystal structure is not available. To test this hypothesis, we engineered a double mutation (Gly32Cys and Ala75Cys) into the core of the
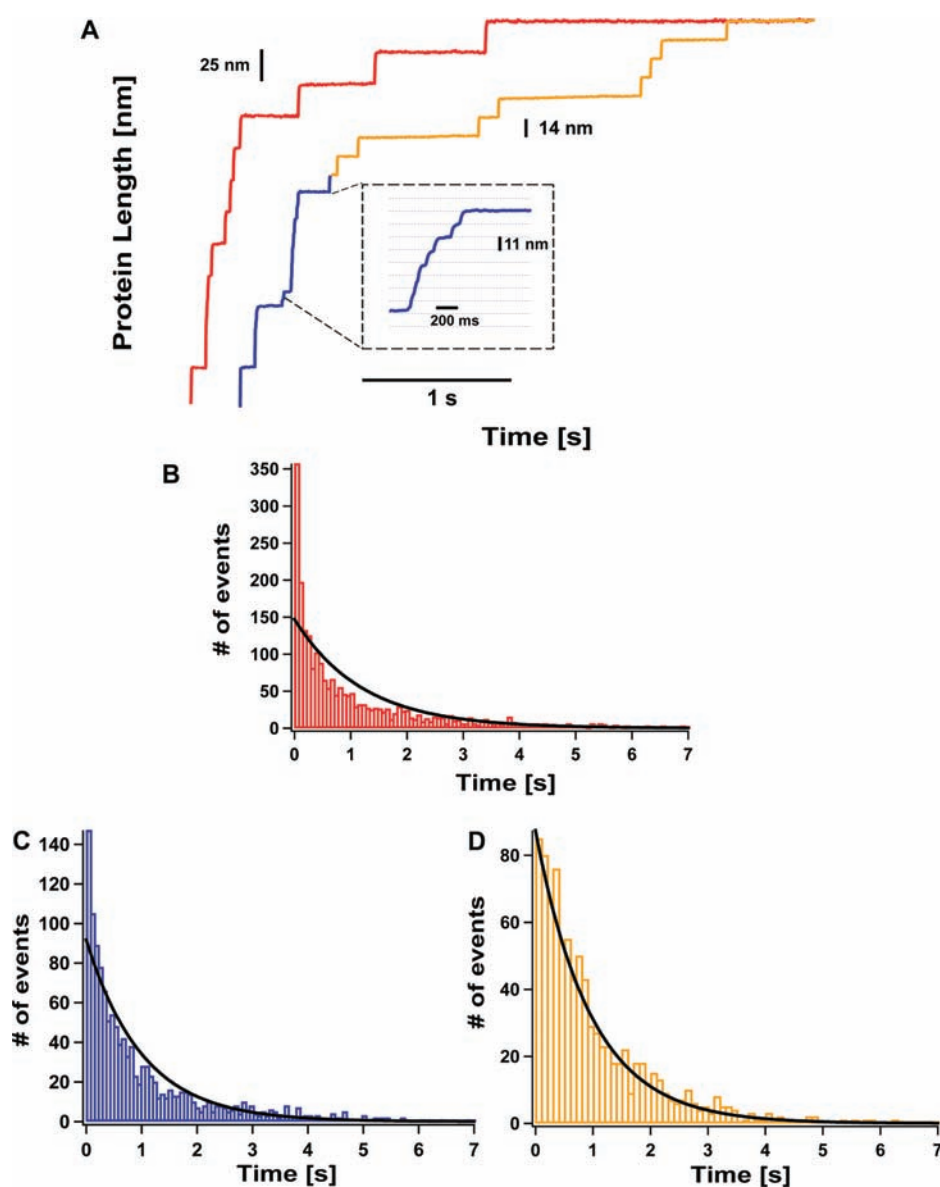
Figure 2. Probing the kinetics of protein unfolding and disulfide reduction within the same single molecule. (A) Unfolding the $(I27_{G32C-A75C})_8$ polyprotein under a constant force of 180 pN (blue trace) results in a stepwise elongation of 11 nm (inset), corresponding to the extension of the protein up to the position of the mechanically rigid disulfide bond. Exposing the disulfide bond to a solution containing 10 mM 1-mercapto-2-propanol at 300 pN triggers the reduction of the disulfide bond, which is marked by a ~14 nm stepwise elongation (orange trace), corresponding to the extension of the amino acids trapped behind the disulfide bond. The total protein elongation of the $(I27_{G32C-A75C})_8$ construct after the unfolding and reduction reactions matches the length of the unfolded wild-type I27 protein (red trace). (B) A histogram corresponding to 2296 unfolding events of the wt-I27 protein measures the probability of unfolding at a constant pulling force of 150 pN. At short dwell times, the distribution deviates significantly from a single exponential (black line). (C) A histogram of 1269 unfolding events of the $(I27_{G32C-A75C})_8$ protein at a constant pulling force of 130 pN cannot be reproduced by a single exponential distribution with an associated characteristic time $k = 1/\langle t \rangle$ (black line). (D) Time−course distribution corresponding to 897 individual disulfide reduction events occurring under a constant force of 300 pN. The distribution can be precisely captured by a single exponential fit (black line).

I27 protein, which spontaneously forms a disulfide bond that is buried and inaccessible to the external solvent environment.[46] We constructed and expressed a polyprotein that consisted of eight identical repeats of this mutated domain, $(I27_{G32C-A75C})_8$. Pulling this construct under a constant force of 130 pN results in a staircase in which the length of each step measures ~11 nm (Figure SI3B in the Supporting Information). Such protein extension corresponds to the unfolding of the protein up to the position of the disulfide bond, which acts as an internal covalent mechanical clamp that prevents the protein from

completely unraveling once the hydrogen bond motif has been disrupted upon the application of force. We gathered 180 individual full unfolding trajectories containing 1269 unfolding events, which yielded an average unfolding rate of $k = 0.99 \text{ s}^{-1}$. As before, a single exponential distribution with a rate constant of 0.99 s^{-1} (black line in Figure 2C) does not fully capture the time distribution of unfolding events (Figure SI1B in the Supporting Information).

The results presented here demonstrate that the unfolding of both the wt-I27 protein (Figure 2B) and the shorter disulfide

bond truncated form (Figure 2C) exhibit unfolding pathways that are heterogeneous, evocative of a native state composed of an ensemble of similar conformations with slightly different energies.[35] Therefore, the unfolding process exhibits a higher degree of complexity than expected for a simple first-order chemical reaction with a single reactant species exhibiting a smooth energy surface. Besides allowing investigation of the individual unfolding pathways (Figure 2C), the $(I27_{G32C-A75C})_8$ polyprotein provides the perfect platform to experimentally investigate, at the single bond level, the individual reaction pathways of the simple S_N2 chemical reaction that involves the reduction of the disulfide bond that shortcuts the I27 protein in the presence of a nucleophile. Following the experimental approach of our previous studies,[40,41] we used a double pulse protocol to study the force-dependent kinetics of disulfide bond reduction (Figure 2A). The first force pulse (180 pN) rapidly unfolded the modules in the polyprotein, resulting in the $\sim$11 nm steps (blue trace). Unfolding the protein exposes the buried disulfide bond to the solution, which allows the chemical attack provided that a nucleophile agent is present in the solution. A second force pulse at a higher force (300 pN) is applied to the protein shortly after all of the modules unfolded. In the presence of 10 mM 1-mercapto-2-propanol in the solution, the reduction of each disulfide bond is marked by a $\sim$14 nm elongation step (Figure 2A, orange trace), which corresponds to the total length of the residues trapped behind the disulfide bond. The total length of the unfolded protein after complete disulfide bond reduction exactly matches the length of the I27 wild-type protein (red trace). Figure 2D shows the time distribution corresponding to 897 individual disulfide reduction events stemming from 144 individual reactive trajectories. In sharp contrast with the time distribution corresponding to the unfolding process (Figure 2C), the reduction of a disulfide bond follows a perfect exponential time–course distribution, as revealed by the single exponential fit (black line) to the data, yielding a reduction rate, $k = 1.03$ s^{-1}, which perfectly corresponds to the reaction rate calculated as $k = 1/\langle t \rangle = 1.01$ s^{-1}. This result suggests that the reduction of a disulfide bond by an S-based nucleophile in solution follows a rather precise pathway with a conserved transition state structure. We then examined whether the specific location of the disulfide bond within the I27 structure had a role in the disulfide bond reactivity. We repeated the experiments described in Figure 2A,C,D with the $(I27_{E24C-K55C})_8$ construct, where the disulfide bond shortcuts the protein between positions 24 and 55. In this case, unfolding the protein up to the disulfide bond elicits steps of $\sim$14.5 nm. The reduction of the disulfide bond, studied during the second force pulse at 300 pN, is identified by a step increase in length of $\sim$10.4 nm, which corresponds again to the length of the amino acids trapped behind the disulfide bond (Figure SI2 in the Supporting Information). As before, the time–course distribution of the unfolding events in this case follows a perfect single exponential, demonstrating that the apparent simplicity observed for the disulfide bond reduction is independent of the chemical environment of the reactive site.

In our experiments, we chose the particular conditions of force and nucleophile concentration such that the three different processes studied here—unfolding of the wild-type protein, unfolding of the truncated protein, and reduction of the disulfide bond—occur at the same rate ($\sim$1 s^{-1}, see Figure 2B–2D and Figure SI1 in the Supporting Information). However, the time–course distribution of the individual events in the case of protein unfolding or disulfide reduction is significantly different. The

subtleties emerge only when the individual pathways are investigated under constant force conditions at the single molecule level. To quantify the diversity in the number of pathways and transition state structures visited during the reaction, we recently developed a generalized Arrhenius equation,[50] based on Zwanzig's static disorder theory,[51,52] to account for the nonexponential survival probability that we observe in our experiments[38] (Figure 2B,C). From the molecular perspective, proteins are known to exhibit conformational diversity in the native state ensemble.[53] In this scenario, static disorder entails that the interconversion rates between the different structural conformers is slow as compared to the reaction rate. Such structural complexity is likely to result in a diverse set of unfolding pathways, which would lead to fluctuations in the height of the free-energy barriers and also in the conformation of the reaction transition state.

Adapting the Arrhenius/Bell formalism[50,54] to include the disorder in the barrier height, $\sigma_{\Delta G}$, and in the transition state conformation, $\sigma_{\Delta x}$, the mean rate of a reaction is given by:

$$\langle k(F) \rangle = \int_{-\infty}^{\infty} k(F, r)\, f(r)\, \mathrm{d}r$$

$$= A \exp\left[-\left(\frac{\Delta G_{avg} - F\Delta x_{avg}}{k_B T}\right)\right] \exp\left[\frac{\sigma_{\Delta G}^2 + F^2 \sigma_{\Delta x}^2}{2(k_B T)^2}\right]$$

$$(1)$$

where r is the amount of disorder in the barrier heights with respect to the average value, $\Delta G_{avg} - F\Delta x_{avg}$. In this equation, A stands for the preexponential factor, k_B is the Boltzmann constant, T is the absolute temperature, ΔG is the height of the free-energy barrier of reaction in the absence of force, and Δx represents the actual distance from the native, initial conformation to the transition state conformation along the reaction coordinate. In our force-clamp experiments, we measure the ensemble-averaged survival probability (Figure 2B–D). In the presence of static disorder, the ensemble-averaged survival probability $\langle S(t, F) \rangle$ is given by the superposition of the survival probability for each reaction pathway, weighted by the probability of the corresponding pathway, giving rise to:

$$\langle S(t, F) \rangle = \int_{-\infty}^{\infty} S(t, F, r)\, f(r)\, \mathrm{d}r$$

$$= \int_{-\infty}^{\infty} \exp\left[-k \exp\left(-\frac{r}{k_B T}\right) \times t\right] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{r^2}{2\sigma^2}\right) \mathrm{d}r$$

$$(2)$$

where

$$\sigma^2 = \sigma_{\Delta G}^2 + F^2 \sigma_{\Delta x}^2 \qquad (3)$$

In the absence of disorder $[\sigma \to 0;\, f(r) \to \delta(r)]$, the survival probability becomes a single exponential, in good agreement with the classical Arrhenius conception. Conversely, in the presence of disorder ($\sigma > 0$), the survival probability becomes nonexponential. Using this refined statistical approach,[38] we tested this theory on our experimental data on protein unfolding and chemical reaction. The histograms of measured unfolding dwell times in each case (Figure 2B,C) correspond to the probability of unfolding, $p(t)$. From these histograms, we simply calculate the ensemble-averaged survival probability as $\langle S(t) \rangle = 1 - \int_0^t p(t')\, \mathrm{d}t'$.
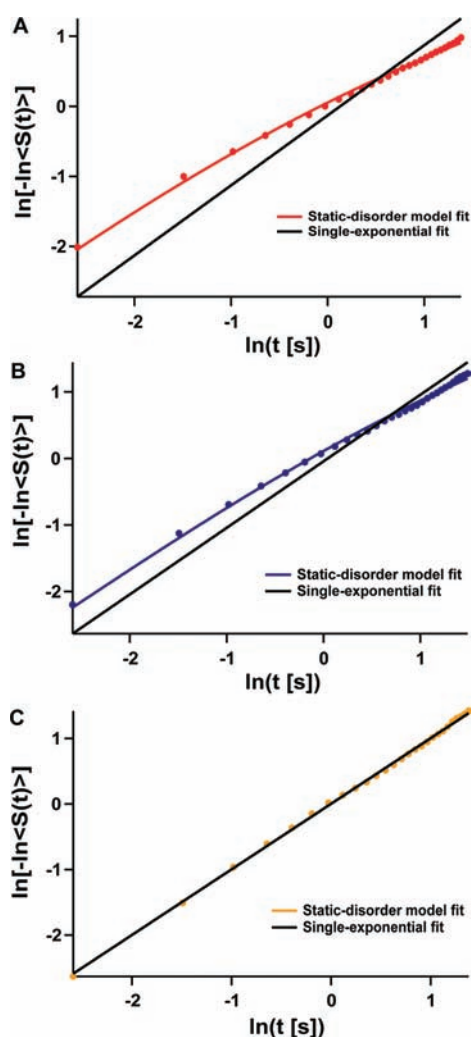
Figure 3. Surviving probability for the unfolding and reduction reactions is well described by static disorder theory. Plot of $\ln[-\ln\langle S(t)\rangle]$ vs $\ln t$ corresponding to the unfolding of the wt-I27 protein at 150 pN (A), the unfolding of the truncated $(I27_{G32C-A75C})_8$ protein at 130 pN (B), and the reduction of the disulfide bond in the $(I27_{G32C-A75C})_8$ protein at 300 pN (C). The solid color lines represent the fits of the static disorder survival probability (eq 2) to the data in each case with the reaction rate k and the variance of the barrier heights σ^2 as fit parameters. The obtained values in each case are listed in Table 1. For comparison, the fits to a single exponential are included in the figures (black traces). While the single exponential fit captures well the time−course distribution for the disulfide bond reduction (C), it clearly fails to reproduce the data corresponding to the unfolding reaction (A and B).

Figure 3A,B shows the plot of the $\ln[-\ln\langle S(t)\rangle]$ versus $\ln t$ obtained for the wild-type and the $(I27_{G32C-A75C})_8$ polyproteins, respectively. The fit of the static disorder model (eq 2) reproduces quite well the experimental data in both cases, yielding values for k and σ^2 that are listed in Table 1. The errors in the fit parameters were estimated using the bootstrap method.[47] For comparison, the single exponential fit is included in the plots as a continuous black line. The obtained value for σ^2 combines the contribution to the disorder arising from the fluctuations in both the height of the energy barrier and the distance to the transition state. While in the experiments described here we cannot decouple both effects, σ^2 is a quantitative reporter of the degree of heterogeneity in the transition state sampled in each particular force-activated

Table 1. Kinetic Parameters for Protein Unfolding and Disulfide Bond Reaction from the Static Disorder Model Fit[a]

	k (s^{-1})	σ^2 [(pN nm)^2]
I27$_{wt}$ unfolding	1.1 ± 0.1	17.5 ± 1.7
I27$_{G32C-A75C}$ unfolding	1.2 ± 0.1	7.7 ± 1.3
I27$_{G32C-A75C}$ reduction	1.0 ± 0.1	$3.4 \times 10^{-4} \pm 0.2$
I27$_{E24C-K55C}$ unfolding	1.9 ± 0.1	21.3 ± 2.4
I27$_{P28C-K54C}$ unfolding	3.5 ± 0.2	15.7 ± 2.1
I27$_{E24C-K55C}$ reduction	1.0 ± 0.1	$9.8 \times 10^{-4} \pm 0.7$

[a] The reaction rate (unfolding or chemical reduction) k, and the variance of the barrier heights, σ^2, were obtained from the static disorder model fit to the data. The errors were estimated using the bootstrap method.

reaction. The obtained high values of σ^2 for the wt-I27 protein (17.5 ± 1.7) (pN nm)^2 and for the truncated I27$_{G32C-A75C}$ form (7.7 ± 1.3) (pN nm)^2 quantitatively confirm the high degree of heterogeneity measured in the protein unfolding pathways under force. In apparent contrast with these results, fitting the static disorder model (eq 2) to the experimental time distribution corresponding to the disulfide bond reduction under force (Figure 2C) results in a value of $(\sigma^2 = 3.41 \times 10^{-5} \pm 0.2)$ (pN nm)^2 (Table 1), thus implying the absence of disorder and the precise agreement with a single exponential time distribution consistent with simple first order kinetics (Figure 3C).

While the difference in the measured values for disorder, σ^2, between protein unfolding and chemical reduction reactions is evident, differences within the disorder values obtained for the unfolding process of both I27 protein forms are also noticeable. Because the σ^2 value obtained for the short I27$_{G32C-A75C}$ form is smaller than that corresponding to the wt-I27, it is conceivable that the fluctuations in the protein unfolding pathways directly correlate with the protein length. To test this hypothesis, we investigated the distribution of unfolding pathways of a set of different I27 protein mutants where the position of the disulfide bond is varied within the structure, namely, in positions 32−75, 24−55, and 28−54. Thus, the number of amino acids being extended upon unfolding of the protein (the unsequestered amino acids) is different in each case, namely, 46 amino acids for the $(I27_{G32C-A75C})$ construct, 58 amino acids for the $(I27_{E24C-K55C})$ protein, and 63 amino acids in the case of the $(I27_{P28C-K54C})$ protein. The wild-type I27 protein is composed of 89 residues. Examples of experimental trajectories obtained for each I27 protein mutant are shown in Figure SI3 in the Supporting Information. The distribution of unfolding times corresponding to the unfolding of the $(I27_{G32C-A75C})_8$, $(I27_{E24C-K55C})_8$, and $(I27_{P28C-K54C})_8$ polyproteins at 130 pN are shown in Figure 4A−C. In none of the cases, a single exponential time distribution with an associated time constant $k = 1/\langle t \rangle$ (black line) fully reproduces the experimental distribution of unfolding times. Figure 4D−F shows the fit of the static disorder theory for each protein mutant. The values of the resulting fitting parameters are included in Table 1. While the shortest protein, $(I27_{G32C-A75C})_8$, exhibits the lower value of dispersion ($\sigma^2 = 7.7 \pm 1.3$) (pN nm)^2, the highest departure from the single exponential time distribution, ($\sigma^2 = 21.3 \pm 2.4$) (pN nm)^2, (Figure 4D−F), does not correspond to the longest wild-type form but instead to the $(I27_{E24C-K55C})_8$ protein, which extends only 58 amino acids. Hence, we do not observe correlation between the heterogeneity in the unfolding pathways and the protein length, suggesting that, rather than the protein length,
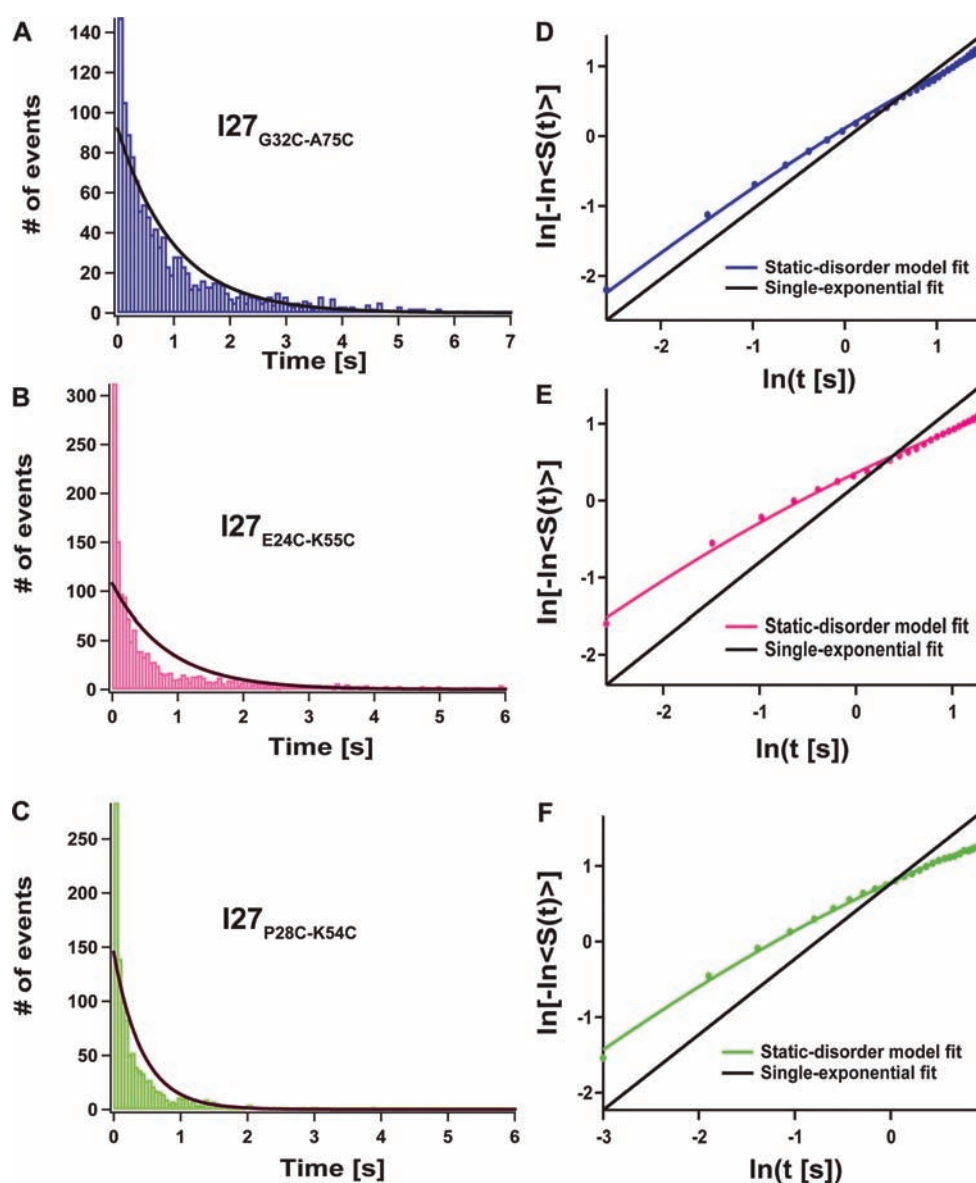
Figure 4. Measured disorder in the unfolding pathways of I27 does not correlate with the protein length. Time−course distribution corresponding to the unfolding of the (A) $(I27_{G32C-A75C})_8$ construct ($n = 1269$), (B) the $(I27_{E24C-K55C})_8$ construct ($n = 1528$), and (C) the $(I27_{P28C-K54C})_8$ polyprotein ($n = 1099$). In none of the cases, a single exponential time distribution with an associated unfolding rate $k = 1/\langle t\rangle$ fully reproduces the experimental distribution of unfolding times. (D and E) Associated plots of $\ln[-\ln\langle S(t)\rangle]$ vs $\ln t$ corresponding to the unfolding of the $(I27_{G32C-A75C})_8$, $(I27_{E24C-K55C})_8$, and $(I27_{E24C-K55C})_8$ constructs, respectively. The fits of the static disorder survival probability (colored lines) and of the single exponential (black line) clearly show the adequacy of the static disorder theory to fit the experimental results. The obtained values for the fitting parameters k and σ^2 are listed in Table 1. Notably, the values of σ^2 do not correlate with the number of amino acids being extended after mechanical unfolding of the protein.

the heterogeneity of unfolding pathways seems to be mainly dictated by the precise conformation of the protein native state, which is subtly modified by the presence of the shortcutting disulfide bonds present in the protein mutants.

■ DISCUSSION

Here, we investigated the effect of applying a constant stretching force on two complete different reactions: the unfolding of a single protein and the reduction of a disulfide bond, both occurring within the structure of the same single molecule. Notably, both reactions give rise to distinct results, highlighting in each case the distribution of sampled transition state structures. It

has been long recognized, mostly using NMR relaxation techniques,[53,55] that proteins are dynamic systems[56] and that the fluctuations of both the backbone and side chains and their interplay with the water molecules that solvate the protein shell give rise to an ensemble of nearly isoenergetic conformations that define the protein's native state. However, direct structural characterization defining each distinct protein conformation, together with their time scale of interconversion, has been mainly restricted to the information provided by all-atom molecular dynamics simulations.[57] Because the "liquidlike" side chain fluctuations are measured to occur within a 10 ns time scale,[57] it is extremely unlikely that they can be captured as "frozen" conformations in our experiments. By contrast, it is plausible that
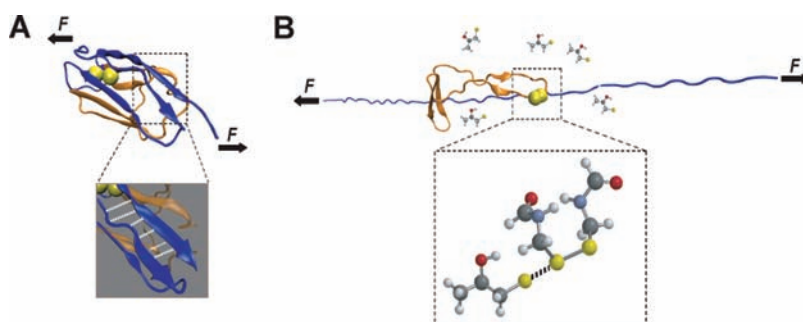
Figure 5. Schematics of the transition state structures of the two distinct reactions probed in our experiments. (A) Unfolding of the I27 protein involves the breakage of a set of key hydrogen bonds (white, inset). Upon disruption, the protein structure (blue region) is extended up to the disulfide bond, which acts as a mechanical covalent clamp. (B) The reduction of the exposed disulfide bond encompasses the back-side attack of the 1-mercapto-2-propanol nucleophile and the sub-Ångström elongation of the S−S bond under force. Upon reduction, the trapped amino acids (orange region) are elongated under the presence of the pulling force. The diversity of conformations defining the transition state for mechanical protein unfolding contrasts with the structural simplicity measured for the S_N2 chemical reaction under force.

the mechanically distinct substates that we capture stem from the interplay of the slower motions of the "solid" backbone with the surrounding water molecules, which altogether define each independent protein conformation.[58] Indeed, in the case of mechanical unfolding of the I27protein, the transition state is composed of a set of key hydrogen bonds that form the mechanical clamp (Figure 5A). Crucially, the solvent molecules are intricately related to the transition state conformation,[59] and it is likely that the heterogeneity of pathways that we measure have their origin in an ensemble of transition state structures with similar energy that differ in the number and dynamics of water molecules governing the forced unfolding reaction.[38]

Our results demonstrate that shortening the I27 protein by the presence of a disulfide bond, although placed in a region away from the key A′G patch, greatly affects the distribution of unfolding pathways. Interestingly, the breadth of the unfolding time distribution does not correlate with the protein length. These results suggest that slight changes in the protein tertiary conformation have major implications on the distribution of the transition state conformations. We speculate that the change in the protein backbone imposed by the rigid disulfide bond restricts the orientation and dynamics of the water molecules defining the unfolding transition state. It is therefore tempting to hypothesize that the structure and diversity in the transition state structures sampled in each protein might correspond to a natural mechanism that regulates protein fragility from a functional viewpoint, which would be of particular importance in cardiac I27 protein. In the case of protein unfolding, the heterogeneity in the unfolding pathways is interpreted here using the static disorder theory in terms of the multiple conformations that populate the transition state structures.[38] An alternative explanation for the measured distribution of unfolding pathways would emerge from the conformational diversity of the native state of the protein, which would lead to an ensemble of structures with slightly different energy barriers.[35] It is plausible that the interplay of both scenarios, that is, the diversity in both the transition state structures and also in the native state conformation, account for the diversity of unfolding pathways that we measure in our experiments. The heterogeneity measured for protein unfolding contrasts with the simplicity and univocal conformation of the transition state corresponding to the reduction of the disulfide bond that shortcuts the I27 protein. In this case, the transition state geometry of the S_N2 chemical reaction involving the 1-mercapto-2-propanol back-side attack (Figure 5B), and occurring

on the Ångström length scale, is highly conserved. Indeed, the high degree of directionality imposed by the back-side attack that describes the S_N2 chemical reaction restricts the conformational space available for the successful collision to take place. Such stereospecificity is likely to account for the homogeneity in the transition state structures that we measure in our experiments. Alternatively, it is also possible that the reaction transition state involves a set of similar structures rather than a unique one. However, because of the fleeting nature of the reaction transition state conformation (occurring on the femtosecond time scale[60]), the rate of interconversion between the different structures would be much faster than the rates measured in our experiments. In such a case, the dynamic disorder in the sampled transition state structures would explain the exponential distribution of times that we measure. The apparent simplicity of the chemical reaction under force using a small nucleophile such as 1-mercapto-2-propanol contrasts with the complex kinetics observed for the same disulfide bond reduction when an enzyme such as thioredoxin is used.[43,61,62] In any case, our experiments probing chemical reactivity under force using small nucleophiles confirm the adequacy of the simple first-order kinetic law to fully capture the kinetics of the statistically broad set of reactive trajectories, at the single bond level.

The apparent simplicity of the chemical reactions under force contrasts with the complexity in the unfolding mechanism, both reactions being measured within the same molecule and under identical experimental conditions. This experiment serves as a control that rules out the possibility that the origin of the measured nonexponential kinetics for protein unfolding results from the limitations of our experimental setup. These potential sources of experimental error[63,64] include the instrumental fluctuations in the constant force applied to the proteins, the finite time response of the force feedback, the averaging of data obtained during independent experimental days using different cantilever probes, or the thermal fluctuations of the polyproteins.

The heterogeneity in the distribution of unfolding pathways of the native I27 protein closely resembles that found for ubiquitin ($\sigma^2 = 18.5 \pm 1.5$) (pN nm)^2, at the same pulling force.[38] The challenge is now to identify the interplay between structural motifs within the protein (such as the number and position of hydrogen bonds) and the medium conditions (such as solvent or temperature), which could altogether affect the distribution of unfolding pathways that can be captured at the single molecule level. Most importantly, according to our results, it is apparent

that protein unfolding is a much complex reaction, for which its kinetic treatment as a simple first order chemical reaction is an oversimplification that does not capture the essential physical properties underlying the process. Regarding the chemical reactions, it would be interesting to identify signatures of complexity at the single bond length-scale using small nucleophiles by studying reactions that could proceed along two concomitant and parallel reaction mechanisms, exhibiting distinct transition state conformations, as it would be the case of a reducing agent that could either proceed through a nucleophilic attack or following an addition/elimination mechanism.[65,66]

Our results demonstrate the capability of force-clamp spectroscopy to capture a large set of individual reactive trajectories that allow direct identification of the degree of heterogeneity in the transition state structures sampled in each particular reaction. The high statistical throughput that can be now achieved with the technique is therefore perfectly suited to indentify fine details in the potential energy landscape, which can only be unveiled at the single molecule level. Crucially, here, we demonstrated that the variance of the barrier heights, σ^2, is a very sensitive parameter that can vary over a wide range within the structure of a single protein. Indeed, the heterogeneity in the unfolding transition state ensemble is very sensitive to mutations that introduce disulfide bonds in different locations throughout the protein. By contrast, disulfide bond reduction by a small nucleophile such as 1-mercapto-2-propanol shows no disorder, implying a single well-defined transition state structure that is insensitive to its location within the same protein. The long-term goal is to build a comprehensive formalism to reconstruct the full energy landscape of a folding protein and to correlate it with the protein's architecture. The diversity of conformations that we found in the most prominent energy minimum of the landscape, corresponding to the native state, might have important implications on the functionality of proteins with mechanical function, of common occurrence in nature.

■ ASSOCIATED CONTENT

Ⓢ **Supporting Information.** Additional figures illustrating the different behavior of force on the energy landscape of protein unfolding and chemical reactions. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
sergi@biology.columbia.edu; jfernandez@columbia.edu

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 868–893.

(2) Levine, R. D. *Molecular Reaction Dynamics*; Cambridge University Press: Cambridge, United Kingdom, 2005.

(3) Brauman, J. I. *Science* **2008**, *319*, 168.

(4) Chabinyc, M. L.; Craig, S. L.; Regan, C. K.; Brauman, J. I. *Science* **1998**, *279*, 1882–1886.

(5) Sun, L.; Song, K.; Hase, W. L. *Science* **2002**, *296*, 875–878.

(6) Bach, R. D.; Dmitrenko, O.; Thorpe, C. *J. Org. Chem.* **2008**, *73*, 12–21.

(7) Hase, W. L. *Science* **1994**, *266*, 998–1002.

(8) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10–19.

(9) Thirumalai, D. *J. Phys. I* **1995**, *5*, 1457–1467.

(10) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619–1620.

(11) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1277–1281.

(12) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177–188.

(13) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524–7528.

(14) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.

(15) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman: New York, 1999.

(16) Mayne, L.; Englander, S. W. *Protein Sci.* **2000**, *9*, 1873–1877.

(17) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 148–150.

(18) Roeffaers, M. B.; De Cremer, G.; Uji-i, H.; Muls, B.; Sels, B. F.; Jacobs, P. A.; De Schryver, F. C.; De Vos, D. E.; Hofkens, J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12603–12609.

(19) Smiley, R. D.; Hammes, G. G. *Chem. Rev.* **2006**, *106*, 3080–3094.

(20) Ferrer, S.; Tunon, I.; Marti, S.; Moliner, V.; Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. *J. Am. Chem. Soc.* **2006**, *128*, 16851–16863.

(21) Basner, J. E.; Schwartz, S. D. *J. Am. Chem. Soc.* **2005**, *127*, 13822–13831.

(22) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3140–3169.

(23) Lu, H. P.; Xun, L.; Xie, X. S. *Science* **1998**, *282*, 1877–1882.

(24) Xue, Q.; Yeung, E. S. *Nature* **1995**, *373*, 681–683.

(25) Yang, H.; Luo, G.; Karnchanaphanurach, P.; Louie, T. M.; Rech, I.; Cova, S.; Xun, L.; Xie, X. S. *Science* **2003**, *302*, 262–266.

(26) van Oijen, A. M.; Blainey, P. C.; Crampton, D. J.; Richardson, C. C.; Ellenberger, T.; Xie, X. S. *Science* **2003**, *301*, 1235–1238.

(27) English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. *Nat. Chem. Biol.* **2006**, *2*, 87–94.

(28) Velonia, K.; Flomenbom, O.; Loos, D.; Masuo, S.; Cotlet, M.; Engelborghs, Y.; Hofkens, J.; Rowan, A. E.; Klafter, J.; Nolte, R. J.; de Schryver, F. C. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 560–564.

(29) Flomenbom, O.; Velonia, K.; Loos, D.; Masuo, S.; Cotlet, M.; Engelborghs, Y.; Hofkens, J.; Rowan, A. E.; Nolte, R. J.; Van der Auweraer, M.; de Schryver, F. C.; Klafter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2368–2372.

(30) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186–195.

(31) Lu, S.; Li, W. W.; Rotem, D.; Mikhailova, E.; Bayley, H. *Nature Chem.* **2010**, *2*, 921–928.

(32) Garcia-Manyes, S. *Nature Chem.* **2010**, *2*, 905–906.

(33) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674–1678.

(34) Schlierf, M.; Li, H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7299–7304.

(35) Brujic, J.; Hermans, R. I.; Walther, K. A.; Fernandez, J. M. *Nature Phys.* **2006**, *2*, 282–286.

(36) Brujic, J.; Hermans, R. I.; Garcia-Manyes, S.; Walther, K. A.; Fernandez, J. M. *Biophys. J.* **2007**, *92*, 2896–2903.

(37) Garcia-Manyes, S.; Brujic, J.; Badilla, C. L.; Fernandez, J. M. *Biophys. J.* **2007**, *93*, 2436–2446.

(38) Kuo, T. L.; Garcia-Manyes, S.; Li, J.; Barel, I.; Lu, H.; Berne, B. J.; Urbakh, M.; Klafter, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11336–11340.

(39) Cao, Y.; Kuske, R.; Li, H. *Biophys. J.* **2008**, *95*, 782–788.

(40) Wiita, A. P.; Ainavarapu, S. R.; Huang, H. H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7222–7227.

(41) Ainavarapu, S. R. K.; Wiita, A. P.; Dougan, L.; Uggerud, E.; Fernandez, J. M. *J. Am. Chem. Soc.* **2008**, *130*, 6479–6487.

3112

dx.doi.org/10.1021/ja109865z |*J. Am. Chem. Soc.* 2011, 133, 3104–3113

(42) Garcia-Manyes, S.; Liang, J.; Szoszkiewicz, R.; Kuo, T. L.; Fernandez, J. M. *Nature Chem.* **2009**, *1*, 236–242.

(43) Szoszkiewicz, R.; Ainavarapu, S. R.; Wiita, A. P.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Fernandez, J. M. *Langmuir* **2008**, *24*, 1356–1364.

(44) 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*, 3694–3699.

(45) Li, H.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkvliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M. *Nature* **2002**, *418*, 998–1002.

(46) Ainavarapu, S. R.; Brujic, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Li, L.; Walther, K. A.; Carrion-Vazquez, M.; Li, H.; Fernandez, J. M. *Biophys. J.* **2007**, *92*, 225–233.

(47) Efron, B. *The Jacknife, the Bootstrap, and Other Resampling Plans*; Society for Industrial Mathematics: Philedelphia, 1982.

(48) Li, H.; Carrion-Vazquez, M.; Oberhauser, A. F.; Marszalek, P. E.; Fernandez, J. M. *Nat. Struct. Biol.* **2000**, *7*, 1117–1120.

(49) Lu, H.; Schulten, K. *Biophys. J.* **2000**, *79*, 51–65.

(50) Arrhenius, S. *Z. Phys. Chem.* **1889**, *4*, 226–248.

(51) Zwanzig, R. *Abstr. Pap. Am. Chem. Soc.* 1989, *198*, 154.

(52) Zwanzig, R. *J. Chem. Phys.* **1992**, *97*, 3587–3589.

(53) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.

(54) Bell, G. I. *Science* **1978**, *200*, 618–627.

(55) Wuthrich, K.; Wagner, G. *FEBS Lett.* **1975**, *50*, 265–268.

(56) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.

(57) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.

(58) Zhou, Y.; Vitkup, D.; Karplus, M. *J. Mol. Biol.* **1999**, *285*, 1371–1375.

(59) Dougan, L.; Feng, G.; Lu, H.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3185–3190.

(60) Zewail, A. H. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 2586–2631.

(61) Wiita, A. P.; Perez-Jimenez, R.; Walther, K. A.; Grater, F.; Berne, B. J.; Holmgren, A.; Sanchez-Ruiz, J. M.; Fernandez, J. M. *Nature* **2007**, *450*, 124–127.

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

(63) Borgia, A.; Williams, P. M.; Clarke, J. *Annu. Rev. Biochem.* **2008**, *77*, 101–125.

(64) Huang, Z.; Boulatov, R. *Pure Appl. Chem.* **2010**, *82*, 931–951.

(65) Bickelhaupt, F. M.; Baerends, E. J.; Nibbering, N. M. M. *Chem. —Eur. J.* **1996**, *2*, 196–207.

(66) Ensing, B.; Klein, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6755–6759.