Mapping the Energy Landscape of Biomolecules Using Single Molecule Force Correlation Spectroscopy: Theory and Applications

V. Barsegov,* D. K. Klimov,‡ and D. Thirumalai*†

*Biophysics Program, Institute for Physical Science and Technology, †Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland; and ‡Bioinformatics and Computational Biology Program, School of Computational Sciences, George Mason University, Manassas, Virginia

ABSTRACT We present, to our knowledge, a new theory that takes internal dynamics of proteins into account to describe forced-unfolding and force-quench refolding in single molecule experiments. In the current experimental setup (using either atomic force microscopy or laser optical tweezers) the distribution of unfolding times, $P(t)$, is measured by applying a constant stretching force $\mathbf{f}_S$ from which the apparent $\mathbf{f}_S$-dependent unfolding rate is obtained. To describe the complexity of the underlying energy landscape requires additional probes that can incorporate the dynamics of tension propagation and relaxation of the polypeptide chain upon force quench. We introduce a theory of force correlation spectroscopy to map the parameters of the energy landscape of proteins. In force correlation spectroscopy, the joint distribution $P(T, t)$ of folding and unfolding times is constructed by repeated application of cycles of stretching at constant $\mathbf{f}_S$ separated by release periods T during which the force is quenched to $\mathbf{f}_Q < \mathbf{f}_S$. During the release period, the protein can collapse to a manifold of compact states or refold. We show that $P(T, t)$ at various $\mathbf{f}_S$ and $\mathbf{f}_Q$ values can be used to resolve the kinetics of unfolding as well as formation of native contacts. We also present methods to extract the parameters of the energy landscape using chain extension as the reaction coordinate and $P(T, t)$. The theory and a wormlike chain model for the unfolded states allows us to obtain the persistence length l_p and the $\mathbf{f}_Q$-dependent relaxation time, giving us an estimate of collapse timescale at the single molecular level, in the coil states of the polypeptide chain. Thus, a more complete description of landscape of protein native interactions can be mapped out if unfolding time data are collected at several values of $\mathbf{f}_S$ and $\mathbf{f}_Q$. We illustrate the utility of the proposed formalism by analyzing simulations of unfolding-refolding trajectories of a coarse-grained protein ($S1$) with β-sheet architecture for several values of $\mathbf{f}_S$, T, and $\mathbf{f}_Q = 0$. The simulations of stretch-relax trajectories are used to map many of the parameters that characterize the energy landscape of $S1$.

INTRODUCTION

Several biological functions are triggered by mechanical force. These include stretching and contraction of muscle proteins such as titin (1,2), rolling and tethering of cell adhesion molecules (3–6), V. Barsegov, D. Klimov, and D. Thirumalai, unpublished), translocation of proteins across membranes (7–10), and unfoldase activity of chaperonins and proteasomes. Understanding these diverse functions requires our probing the response of biomolecules to applied external tension. Dynamical responses to mechanical force can be used to characterize in detail the free energy landscape of biomolecules. Advances in manipulating micron-sized beads attached to single biomolecules have made it possible to stretch, twist, unfold, and even unbind proteins using forces on the order of tens of picoNewtons (11–13). Single molecule force spectroscopy on a number of different systems has allowed us to obtain a glimpse of the unbinding energy landscape of biomolecules and protein-protein complexes (14–17). In atomic force microscopy (AFM) experiments, used to unfold proteins by force, one end of a protein is adsorbed on a template and a constant or a time-dependent pulling force is applied to the other terminus (18–24). By measuring the distribution of forces required to completely unfold proteins and the associated unfolding times, the global parameters of the protein energy landscape can be estimated (25–30). These insightful experiments when combined with theoretical studies (31–33) can give an unprecedented picture of forced-unfolding pathways.

Current experiments have been designed primarily to obtain information on forced-unfolding of proteins and do not probe the reverse folding process. Although force-clamp AFM techniques have been used recently to probe (re)folding of single ubiquitin polyprotein (23), the lack of theoretical approaches has made it difficult to interpret these pioneering experiments (34,35). Secondly, the resolution of multiple timescales in protein folding and refolding requires not only novel experimental tools for single molecule experiments but also new theoretical analysis methods. Minimally, unfolding of proteins by a stretching force, $\mathbf{f}_S$, is described by the global unfolding time $\tau_U(\mathbf{f}_S)$, timescales for propagation of the applied tension, and the dynamics describing the intermediates or protein-coil states. Finally, if the external conditions (loading rate or the magnitude of $\mathbf{f}_S$) are such that these processes can occur on similar timescales, then the analysis of the data requires new theoretical ideas.

For forced unfolding, the variable conjugate to $\mathbf{f}_S$, namely, the protein end-to-end distance X, is a natural reaction coordinate. However, X is not appropriate for describing

Submitted October 12, 2005, and accepted for publication January 12, 2006.

Address reprint requests to D. Thirumalai, Tel.: 301-405-4303; E-mail: thirum@glue.umd.edu.

0006-3495/06/06/3827/15 $2.00

doi: 10.1529/biophysj.105.075937

protein refolding which, due to substantial variations in the duration of folding barrier crossing, may range from milliseconds to few minutes. To obtain statistically meaningful distributions of unfolding times, a large number of complete unfolding trajectories must be recorded, requiring repeated application of the pulling force. The inherent heterogeneity in the duration of folding and the lack of correlation between evolution of X and (re)folding progress creates initial state ambiguity when force is repeatedly applied to the same molecule. As a result, the interpretation of unfolding time data is complicated, especially when the conditions are such that the reverse folding process at the quenched force $\mathbf{f}_Q$ can occur on a long timescale, $\tau_F(\mathbf{f}_Q)$.

Motivated by the need to assess the effect of the multiple timescales on the energy landscape of folding and unfolding, we develop a new theoretical formalism to describe correlations between the various dynamical processes. Our theory leads naturally to a new class of single molecule force experiments, namely, the force correlation spectroscopy (FCS), which can be used to study both forced unfolding as well as force-quenched (re)folding. Such studies can lead to more detailed information on both kinetic and dynamic events underlying unfolding and refolding. In the FCS, cycles of stretching ($\mathbf{f}_S$) are separated by periods T of quenched force $\mathbf{f}_Q < \mathbf{f}_S$, during which the stretched protein can relax from its unfolded state X_U to coil state X_C or even (re)fold to the native basin of attraction (NBA) state. The two experimental observables are X and the unfolding time t. The central quantity in the FCS is the distribution of unfolding times $P(T, t)$ separated by recoil or refolding events of duration T. The higher order statistical measure embedded in $P(T, t)$ is readily accessible by constructing a histogram of unfolding times for varying T and does not require additional technical developments. The crucial element in the proposed analysis is that $P(T, t)$ is computed by averaging over final (unfolded) states, rather than initial (folded) states. This procedure removes the potential ambiguity of not precisely knowing the initial distribution of conformations in the NBA. Despite the uniqueness of the native state there are a number of conformations in the NBA that reflect the fluctuations of the folded state. The proposed formalism is a natural extension of unbinding-time data analysis. Indeed, $P(T, t)$ reduces to the standard distribution of unfolding times $P(t)$ when T exceeds protein (re)folding timescale $\tau_F(\mathbf{f}_Q)$.

The complexity of the energy landscape of proteins demands FCS and the theoretical analysis. Current single molecule experiments on poly-Ub or poly-Ig27 (performed in the $T \to \infty$ regime) show that in these systems unfolding occurs abruptly in an apparent all-or-none manner or through a dominant intermediate (31). On the other hand, refolding upon force-quench is complex, and surely occurs though an ensemble of collapsed coiled states (23). A number of timescales characterize the stretch-release experiments. These include besides $\tau_F(\mathbf{f}_Q)$, the $\mathbf{f}_S$-dependent unfolding time, and the relaxation dynamics in the coiled states $\{C\}$ upon force-quench $\tau_d(\mathbf{f}_Q)$. In addition, if we assume that X is an appropriate reaction coordinate, then the location of the NBA, $\{C\}$, the transition state ensembles, and the associated widths are required for a complete characterization of the underlying energy landscape. Most of these parameters can be extracted using the proposed FCS experiments and the theoretical analysis presented here.

In a preliminary study (36), we reported the basics of the theory used to propose a new class of single molecule force spectroscopy methods for deciphering protein-protein interactions. This article is devoted to further developments in the theory, with application to forced-unfolding and force-quench refolding of proteins. In particular, we illustrate the efficacy of the FCS by analyzing single unfolding-refolding trajectories generated for a coarse-grained model (CGM) protein $S1$ with β-sheet architecture (37,38). We showed previously that forced-unraveling of $S1$, in the limit of $T \to \infty$, can be described by an apparent two-state kinetics (38,39). The thermodynamics and kinetics observed in $S1$ is a characteristic of a number of proteins where folding/unfolding fits well two-state behavior (40). Thus, $S1$ serves as a useful model to illustrate the efficacy of the FCS. Here, we show that by varying T and the magnitude of the stretching ($\mathbf{f}_S$ or $\mathbf{f}_Q$), the entire dynamical processes, starting from the NBA to the fully stretched state, can be resolved. In the process we establish that $P(T, t)$, which can be measured using AFM or laser optical tweezer (LOT) experiments, provides a convenient way of characterizing the energy landscape of biomolecules in detail.

MODELS AND METHODS

Theory of force correlation spectroscopy (FCS)

In single-molecule atomic force microscopy (AFM) experiments used to unfold proteins by force, the N-terminus of a protein is anchored at the surface and the C-terminus is attached to the cantilever tip through a polymer linker. The molecule is stretched by displacing the cantilever tip and the resulting force is measured. From a theoretical perspective it is more convenient to envision applying a constant stretching force $\mathbf{f}_S = f_S\mathbf{x}$ in the $\mathbf{x}$-direction (Fig. 1). The free energy in the constant force formulation is related to the experimental setup by a Legendre transformation. More recently, it has become possible to apply a constant force in AFM or laser or optical tweezer (LOT) experiments to the ends of a protein. With this setup the unfolding time for the end-to-end distance X to reach the contour length L can be measured for each molecule. For a fixed $\mathbf{f}_S$, repeated application of the pulling force results in a single trajectory of unfolding times ($t_1, t_2, t_3, \ldots$, Fig. 1) from which the histogram of unfolding times $P(t)$ is obtained. The $\mathbf{f}_S$-dependent unfolding rate K_U is obtained by fitting a Poissonian formula $K_U^{-1} \exp[-K_U t]$ to the kinetics of population of folded states p_F, which is related to $P(t)$ as $p_F(t) = 1 - \int_0^t dsP(s)$.

Because K_U is a convolution of several microscopic processes, it does not describe unfolding in molecular detail. For instance, mechanical unfolding of fibronectin domains Fn^{III} involves the intermediate aligned state (26) with partially disrupted hydrophobic core, which cannot be resolved by knowing only K_U. Even when the transition from the folded state F to the globally extended state U (26) does not involve parallel routes as in Fig. 2, or multistate kinetics, the force-induced unfolding pathway must involve formation of intermediate coiled states $\{C\}$. The subsequent transition from $\{C\}$

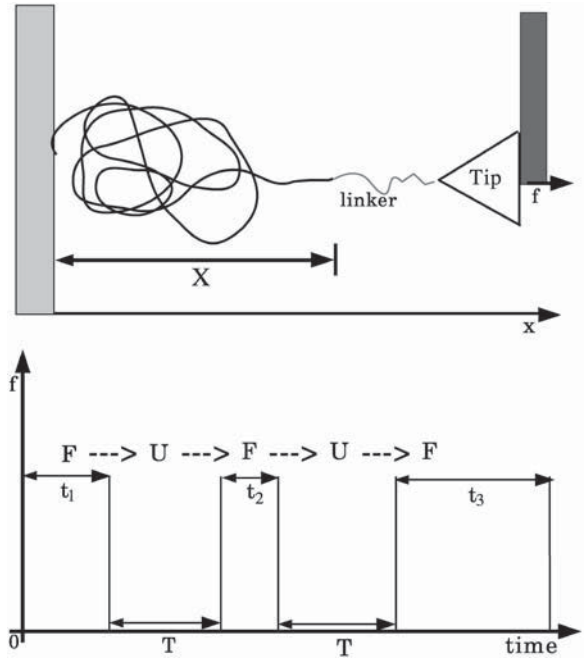

FIGURE 1 (*Top*) A typical AFM setup: constant force $\mathbf{f} = \mathbf{f}_S = f_S\mathbf{x}$ is applied through the cantilever tip linker in the direction $\mathbf{x}$ parallel to the protein end-to-end vector $\mathbf{X}$. Stretching cycles are interrupted by relaxation intervals T during which the force is quenched, $\mathbf{f} = \mathbf{f}_Q = f_Q\mathbf{x}$ ($f_S > f_Q$). (*Bottom*) A single trajectory of forced unfolding times $t_1, t_2, t_3, \ldots$, separated by fixed relaxation time T, during which the unfolded protein can either collapse into the manifold of coiled states $\{C\}$ if T is short or reach the native basin of attraction (NBA) if T is long.

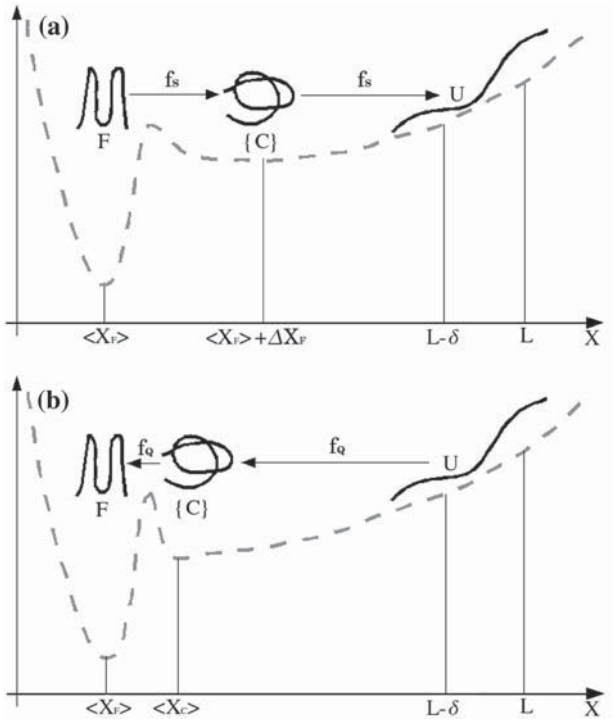

FIGURE 2 Schematic of the free energy profile of a protein (*solid black lines*) upon stretching at constant force $\mathbf{f}_S$ and force-quench $\mathbf{f}_Q$. (*a*) The projections of energy landscape (*dashed lines*) is in the direction of $\mathbf{X}$, which is a suitable reaction coordinate for unfolding induced by force $\mathbf{f}_S$. The average end-to-end distance in the native basin of attraction is $\langle X_F \rangle$. Upon application of $\mathbf{f}_S$, rupture of contacts that stabilize the folded state F results in the formation of an ensemble of high energy extended (by ΔX_F) conformations $\{I\}$. Subsequently, transitions to globally unfolded state U (with $L - \delta \leq X \leq L$) occurs. (*b*) Free energy profile for force-quench refolding, which occurs in the order $U \rightarrow \{C\} \rightarrow F$. Refolding is initiated by quenching the force $\mathbf{f}_S \rightarrow \mathbf{f}_Q < \mathbf{f}_C$, where $\mathbf{f}_C$ is the equilibrium critical force needed to unfold the native protein. The initial event in the process is the formation of an ensemble of compact structures. The mean end-to-end distance of $\{C\}$ is $\langle X_C \rangle$ and the width is ΔX_C, which is a measure of heterogeneity of the refolding pathways. These states may or may not end up in the native basin of attraction (NBA) depending on the duration of T. We have used $\mathbf{X}$ as a reaction coordinate during force-quench for purposes of illustration only.

results in the formation of the globally unfolded state U. The incomplete time resolution prevents current experiments from probing the signature of the collapsed states. To probe the contributions from the underlying $\{C\}$ states to global unfolding requires sophisticated experiments that can resolve contributions from dynamic events underlying forced unfolding. We propose a novel experimental procedure which, when supplemented with unfolding-time data analysis described below, allows us to separately probe the kinetics of native interactions and the dynamics of the protein coil (i.e., the dynamics of end-to-end distance X when the native contacts are disrupted).

Consider an experiment in which stretching cycles (triggered by applying $\mathbf{f}_S$) are interrupted by relaxation intervals T during which force is quenched to $\mathbf{f}_Q < \mathbf{f}_S$. In the time interval T, the polypeptide chain can relax into the manifold $\{C\}$ or even refold to the native state F if T is long enough. If $\mathbf{f}_S > \mathbf{f}_C$ and $\mathbf{f}_Q < \mathbf{f}_C$ where $\mathbf{f}_C$ is the equilibrium critical unfolding force at the specific temperature (see phase diagram for *S*1 in (38)), these transformations can be controlled by T. In the simplest implementation, we set $\mathbf{f}_Q = 0$. The crucial element in the FCS experiment is that the same measurements are repeated for varying T. In the FCS the unfolding times are binned to obtain the joint histogram $P(T, t)$ of unfolding events of duration t generated from the recoil manifold $\{C\}$ or the native basin of attraction (NBA) or both, depending on the duration of the relaxation time T. In the current experiments, $T \rightarrow \infty$. As a result, the dynamics of additional states in the energy landscape that are explored during folding or unfolding are not probed.

The advantages of $P(T, t)$ over the standard distribution of unfolding times $P(t)$ are twofold. First, $P(T, t)$ is computed by averaging over well-characterized fully stretched states. This eliminates the problem of not knowing the distribution of initial protein states encountered in current experiments. Indeed, due to intrinsic heterogeneity of the protein folding pathways, after the first unfolding event the protein may or may not refold into the native conformation, which creates the initial state ambiguity in the next (second, third, etc.) pulling cycle. Therefore, statistical analysis based on averaging over final (stretched) states rather than initial (folded) states allows us to overcome this difficulty. Secondly, statistical analysis of unfolding data performed for different values of T allows us to separately probe the kinetics of native interactions and the dynamics of X. In addition, the entire energy landscape of native interactions can be mapped out when stretch-quench cycles are repeated for several values of $\mathbf{f}_S$, $\mathbf{f}_Q$, and T.

Regime I ($T \ll \tau_F$)

In the simplest unfolding scenario, application of $\mathbf{f}_S$ results in the disruption of the native contacts ($F \rightarrow \{C\}$) followed by stretching of the manifold $\{C\}$ into U (Fig. 2). When stretching cycles are separated by short T compared to the protein folding timescale τ_F at $\mathbf{f}_Q = 0$, $P(T;t)$ is determined by the evolution of the coil state. Then the unfolded state population $p_U(T;t)$ is given by the convolution of protein relaxation (over time T) from the fully stretched state $X_U \approx L$, to an intermediate coiled-state X_1, and stretching

X_1 into final state X_f over time t. Thus, $P(T;t)$ is obtained from $p_U(T;t)$ by taking the derivative with respect to t,

$$P(T \ll \tau_F; t) = \frac{d}{dt} p_U(T \ll \tau_F; t) = \frac{d}{dt} \frac{1}{N(T)} \int_{L-\delta}^{L} dX_f 4\pi X_f^2 \times \int_0^L dX_1 4\pi X_1^2 \int_0^L dX_U 4\pi X_U^2 \times G_S(X_f, t; X_1) G_Q(X_1, T; X_U) P(X_U), \quad (1)$$

where $N(T)$ is the T-dependent normalization constant obtained by taking the last integral in the right-hand side of Eq. 1 from $X_f = 0$ to $X_f = L$, and $P(X_U)$ is the distribution of unfolded states. If X is well controlled, X_U is expected to be centered around a fixed value $\bar{X}_U$ and $P(X_U) \sim \delta(X_U - \bar{X}_U)$. In Eq. 1, $G_Q(X', t;X)$ and $G_S(X', t;X)$ are, respectively, the quenched and the stretching force-dependent conditional probabilities to be in the coiled state X' at time t arriving from state X at time $t = 0$. The integral over X_f is performed in the range $[L - \delta;L]$, with $X = L - \delta$ (Fig. 2) representing unfolding distance at which the total number of native contacts Q is at the unfolding threshold, $Q \approx Q^*$. It follows that $P(T;t)$ (Eq. 1) contains information on the dynamics of X. By assuming a model for X and fitting $P(T;t)$, obtained by differentiating the integral expression appearing in Eq. 1, to the histogram of unfolding times, separated by short $T \ll \tau_F$, we can resolve the dynamics of the polypeptide chain in the coil state, which allows us to evaluate the $\mathbf{f}_Q$-dependent coil dynamical timescale τ_d using single-molecule force spectroscopy. The fit of Eq. 1 could be analytical or numerical depending on the model of X.

Regime II ($T \gg \tau_F$)

When stretching cycles are interrupted by long relaxation periods, $T \gg \tau_F$, the coiled states refold to X_F (Fig. 2). In this regime, the initial conformations in forced-unfolding always reside in the NBA. In this limit, $P(T;t)$ reduces to the standard distribution of unfolding times $P(T, t) \to P(t)$. When $T \gg \tau_F$, $P(T;t)$ is given by the convolution of the kinetics of rupture of native contacts, resulting in protein extension ΔX_F, and dynamics of X from state $X_F + \Delta X_F$ to final state X_f,

$$P(T \gg \tau_F; t) = P(t) = \frac{d}{dt} p_U(T \gg \tau_F; t) = \frac{d}{dt} \frac{1}{N'(T)} \int_{L-\delta}^{L} dX_f 4\pi X_f^2 \int_0^L dX_F 4\pi X_F^2 \int_0^t dt' \times G_S(X_f, t; X_F + \Delta X_F, t') P_F(t', X_F; \mathbf{f}_S), \quad (2)$$

where $N'(T)$ is the normalization constant obtained as in Eq. 1, and $P_F(t, X_F;\mathbf{f}_S)$ is the probability of breaking the contacts over time t that stabilize the native state X_F. By assuming a model for $P_F(t, X_F;\mathbf{f}_S)$ and employing information on the dynamics of X, obtained from the short T-experiment (Eq. 1), we can probe the disruption kinetics of native interactions. By repeating long T-measurements at several values of $\mathbf{f}_S$, we can map out the energy landscape of native interactions projected on the direction of the end-to-end distance vector.

Regime III ($T \sim \tau_F$)

In this limit, some of the molecules reach the NBA, starting from extended states ($X \approx L$), whereas others remain in the basin $\{C\}$. The fraction of folding events ρ_F depends on T, during which X approaches the average extension $\langle X_C \rangle$ facilitating the formation of native contacts. Thus, $P(T \sim \tau_F)$, obtained in the intermediate T-experiment, involves contributions from both $\{C\}$ and F initial conditions and is given by a superposition,

$$P(T \sim \tau_F; t) = \rho_F(T) P(T \gg \tau_F; t) + \rho_C(T) P(T \ll \tau_F; t), \quad (3)$$

where the probability to arrive to F from $\{C\}$ at time T is given by

$$\rho_F(T) = \int_0^L dX_1 4\pi X_1^2 \int_0^L dX_U 4\pi X_U^2 P_C(T, X; \mathbf{f}_Q) \times G_Q(X_1, T; X_U) P(X_U), \quad (4)$$

and the probability to remain in $\{C\}$ is $\rho_C(T) = 1 - \rho_F(T)$. In Eq. 4, $P_C(T, X;\mathbf{f}_Q)$ is the refolding probability determined by the kinetics of formation of native contacts. Because the dynamics of X is weakly correlated with formation of native contacts, X in P_C is expected to be broadly distributed. Therefore, Eqs. 3 and 4 can be used to probe kinetics of formation of native interactions.

For Eqs. 1 and 2 to be of use, one needs to know the (re)folding timescale τ_F. The simplest way to evaluate τ_F is to construct a series of histograms $P(T_n, t)$ ($n = 1, 2, \ldots, N$) for a fixed $\mathbf{f}_S$ and increasing relaxation time $T_1 < T_2 < \ldots < T_N$, and compare $P(T_n, t)$ values with the distribution $P(T^*, t)$ obtained for sufficiently long $T^* \gg \tau_F$. If $T = T^*$, then all the molecules are guaranteed to reach the NBA. The difference

$$D(T_n) = |P(T_n, t) - P(T*, t)| \quad (5)$$

is expected to be nonzero for $T_n \le \tau_F$ and should vanish if T_n exceeds τ_F. Statistically, as T_n starts to exceed τ_F, increasingly more molecules will reach the NBA by forming native contacts. Then, more unfolding trajectories will start from folded states, and when $T \gg \tau_F$ all unfolding events will originate from the NBA. Therefore, $D(T_n)$ is a sensitive measure for identifying the kinetic signatures for forming native contacts. The utility of $D(T_n)$ is that it is a simple yet accurate estimator of τ_F, which can be utilized in practical applications. Indeed, one can estimate τ_F by identifying it with the shortest T_n at which $P(T_n;t) \approx P(T^*, t)$, i.e., $T_n \approx \tau_F$. We should emphasize that to obtain τ_F from the criterion that $D(\tau_F) \approx 0$ no assumptions about the distribution of refolding times have been made. Having evaluated τ_F one can then use Eqs. 1 and 2 for short and long T-measurements to resolve protein coil dynamics and rupture kinetics of native contacts.

Let us summarize the major steps in the FCS. First, we estimate τ_F by using $D(T)$ (Eq. 5). We next probe protein coil dynamics by analyzing $P(T \ll \tau_F; t)$ obtained from short-T-measurements (Eq. 1). In the third step, we use information on protein coil dynamics to resolve the kinetics of rupture of native interactions contained in $P(T \gg \tau_F; t)$ of long-T-measurements (Eq. 2). Finally, by employing the information on protein coil dynamics and kinetics of rupture of native interactions, we resolve the kinetics of formation of native contacts by analyzing $P(T \sim \tau_F;t)$ from intermediate T-measurements (Eqs. 3 and 4).

The beauty of the proposed framework is that these experiments can be readily performed using available technology. In the current AFM experiments, T can be made as short as a few microseconds. Simple calculations show that the relaxation of a short 50-amino-acid protein from the stretched state, with $L \approx 19$ nm, to the coiled states $\{C\}$, with, say, $X \approx 2$ nm, occurs on the timescale $\tau_d \approx \Delta x^2/D \sim 10$ μs, where $\Delta x = L - X \approx 17$ nm and $D \approx 10^{-7}$ cm^2/s is the diffusion constant. Clearly, the time of formation of native contacts, which drives the transition from $\{C\}$ to the NBA, prolongs τ_F by a few microseconds to a few milliseconds or larger, depending on folding conditions. In the experimental studies of forced unfolding and force-quenched refolding of ubiquitin, τ_F was found to be of the order of 10–100 ms (23). Computer simulation studies of unzipping-rezipping transitions in short 22-nt RNA hairpin P5GA have predicted that τ_F is of the order of a few hundreds of microseconds (34).

Model for the kinetics of native contacts

To interpret the data generated by FCS it is useful to have a model for the time evolution of the native contacts and X. We first present a simple kinetic model for rupture and formation of native contacts represented by probabilities P_F and P_C in Eqs. 2 and 4, respectively, and a model for the dynamics of X given by the propagator $G_{S,Q}(X', t;X)$. To describe the force-dependent evolution of native interactions we adopt the continuous-time-random-walk (CTRW) formalism (41–45). In the CTRW model, a random walker,

representing rupture (formation) of native contacts, pauses in the native (coiled) state for a time t before making a transition to the coiled (native) state. The waiting-time distribution is given by the function $\Psi_\alpha(t)$ ($\alpha = r$ or f, where r and f refer to rupture and formation of native contacts, respectively). We assume that the probabilities $P_F(t, X_F;\mathbf{f}_S)$ and $P_C(t, X_C;\mathbf{f}_Q)$ are separable so that

$$P_F(t, X_F;\mathbf{f}_S) \approx P_{eq}(X_F)P_r(t;\mathbf{f}_S),$$
$$P_C(t, X_C;\mathbf{f}_Q) \approx P_C(X_C)P_f(t;\mathbf{f}_Q), \quad (6)$$

where $P_{eq}(X_F)$ is the equilibrium distribution of native states, $P_C(X_C)$ is the distribution of coiled states, and $P_r(t;\mathbf{f}_S)$ and $P_f(t;\mathbf{f}_Q)$ are the force-dependent probabilities of rupture and formation of native contacts, respectively. Factorization in Eq. 6 implies that application of force does not result in the redistribution of states X_F and X_C in the NBA and in the manifold of coiled states $\{C\}$, but only changes the timescales for NBA $\rightarrow \{C\}$ and $\{C\} \rightarrow$ NBA transitions, and thus, the probabilities P_r and P_f. We expect the approximation in Eq. 6 to be valid provided the rupture of native contacts and refolding events are cooperative.

During stretching cycles, for $\mathbf{f}_S$ well above $\mathbf{f}_C$, we may neglect the reverse folding process. Similarly, global unfolding is negligible during relaxation periods with $\mathbf{f}_Q < \mathbf{f}_C$. Then, the master equations for $P_r(t)$ is

$$\frac{d}{dt}P_r(t) = -\int_0^t d\tau\Phi_r(\tau)P_r(t-\tau), \quad (7)$$

where $\Phi_r(t)$ is the generalized rate for the rupture and formation of native interactions. In the Laplace domain, defined by $\bar{f}(z) = \int_0^\infty dt f(t)\exp[-tz]$, $\Psi_r(t)$ is related to $\Phi_r(t)$ as

$$\bar{\Phi}_r(z) = z\bar{\Psi}_r(z)[1-\bar{\Psi}_r(z)]^{-1}. \quad (8)$$

The structure of the master equation for $P_f(t)$ is identical to Eq. 7, with the relationship between $\Phi_f(t)$ and $\Psi_f(t)$ being similar to Eq. 8. The general solution to Eq. 7 is

$$\bar{P}_r(z) = [z-\Phi_r(z)]^{-1}\bar{P}_r(0), \quad (9)$$

where $\bar{P}_r(0) = 1$ is the initial condition and the solution in the time domain is given by the inverse Laplace transform, $P_r(t) = L^{-1}\{\bar{P}_r(z)\}$. The solution for $\bar{P}_f(z)$ is obtained in a similar fashion (see Eq. 9) with initial condition of $\bar{P}_f(0) = 1$.

Model for the polypeptide chain

In the extended state, when the majority of native interactions that stabilize the folded state are disrupted, the molecule can be treated roughly as a fluctuating coil. Simulations and analysis of native structures (46) suggest that proteins behave as wormlike chains (WLCs). For convenience we use a continuous WLC description for the coil state whose Hamiltonian is

$$H = \frac{3k_BT}{2l_p}\int_{-L/2}^{L/2} ds\left(\frac{\partial\mathbf{r}(s,t)}{\partial s}\right)^2 + \frac{3l_pk_BT}{8}\int_{-L/2}^{L/2} ds\left(\frac{\partial^2\mathbf{r}(s,t)}{\partial s^2}\right)^2 + \frac{3k_bT}{4}\left[\left(\frac{\partial\mathbf{r}(-L/2,t)}{\partial s}\right)^2 + \left(\frac{\partial\mathbf{r}(-L/2,t)}{\partial s}\right)^2\right] + \mathbf{f}_{S,Q}\int_{-L/2}^{L/2} ds\left(\frac{\partial\mathbf{r}(s,t)}{\partial s}\right), \quad (10)$$

where l_p is the protein coil persistence length. A large number of force-extension curves obtained using mechanical unfolding experiments in proteins, DNA, and RNA have been analyzed using a WLC model. In Eq. 10, the three-dimensional Cartesian vector $\mathbf{r}(s, t)$ represents the spatial location of the s^{th} protein monomer at time t. The first two terms describe chain connectivity and bending energy, respectively. The third term represents fluctuations of the chain free ends and the fourth term corresponds to coupling of $\mathbf{r}$ to $\mathbf{f}_{S,Q}$. The end-to-end vector is computed as $\mathbf{X}(t) = \mathbf{r}(L/2, t) - \mathbf{r}(-L/2, t)$.

We need a dynamical model in which X is represented by the propagator $G(X, t;X_0)$. Although bond vectors of a WLC are correlated, the statistics of X can be represented by a large number of independent modes. It is therefore reasonable, at least in the large L limit, to describe $G_{S,Q}(X, t;X_0)$ by a Gaussian,

$$G_{S,Q}(X,t;X_0) = \left(\frac{3}{2\pi\langle X^2\rangle_{S,Q}}\right)^{3/2}\frac{1}{(1-\phi^2_{S,Q}(t))^{3/2}} \times \exp\left[-\frac{3(X-\phi_{S,Q}(t)X_0)^2}{2\langle X^2\rangle_{S,Q}(1-\phi^2_{S,Q}(t))}\right], \quad (11)$$

specified by the second moment $\langle X^2\rangle_{S,Q}$ and the normalized correlation function $\phi(t)_{S,Q} = \langle X(t)X(0)\rangle_{S,Q}/\langle X^2\rangle_{S,Q}$. Calculations of $\langle X^2\rangle_{S,Q}$ and $\phi(t)_{S,Q}$ are given in the Appendix (46,47). In the absence of force, we obtain

$$\langle X(t)X(0)\rangle_0 = 12k_BT\sum_{n=1}^{\infty}\frac{1}{z_n}\psi_n^2(L/2)e^{-z_nt/\gamma}, \quad n = 1, 3, \ldots, 2q+1, \quad (12)$$

where $\psi_n(X)$ and z_n are the eigenfunctions and eigenvalues of the modes of the operator that describe the dynamics of $\mathbf{r}(s, t)$ (see Eq. A1). To construct the propagator $G_{S,Q}(X, t;X_0)$ for $\mathbf{f}_{S,Q}$, Eq. A1 is integrated with $\mathbf{f}_{S,Q}$ added to random force. We obtain $\langle X^2\rangle_{S,Q} = \langle X^2\rangle_0 + \mathbf{f}^2_{S,Q}\sum_{n=1}^{\infty}\psi_n^2(L/2)/z_n^2$, where $n = 1, 3, \ldots, 2q+1$. We analyze the distributions of unfolding times $P(T, t)$ for the model sequence $S1$ (Fig. 3) obtained using simulations, the CTRW model for evolution of native interactions (Eqs. 6–9), and Gaussian statistics of the protein coil (Eq. 11).

Simulations of model β-sheet protein

The usefulness of FCS is illustrated by computing and analyzing the distribution function $P(T;t)$ for a model polypeptide chain with β-sheet architecture. Sequence $S1$, which is a variant of an off-lattice model introduced sometime ago (37), is a coarse-grained model (CGM) of a polypeptide chain, in which each amino acid is substituted with a united atom of appropriate mass and diameter at the position of the C_α-carbons (38,39). The $S1$ sequence is modeled as a chain of 46 connected beads of three types—hydrophobic B, hydrophilic L, and neutral N—with the contour length $L = 46a$, where $a \approx 3.8$ Å is the distance between two consecutive C_α-carbon atoms. The coordinate of the j^{th} residue is given by the vector $\mathbf{x}_j$ with $j = 1, 2, \ldots, N$.

The potential energy U of a chain conformation is

$$U = U_{bond} + U_{bend} + U_{da} + U_{nb}, \quad (13)$$

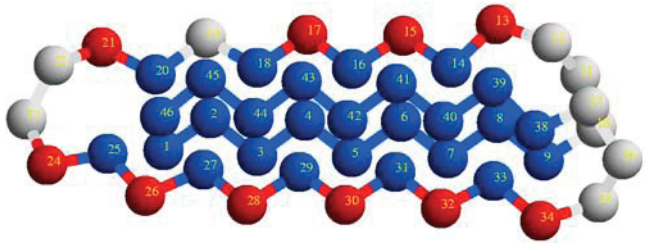

FIGURE 3 Native structure of the model protein $S1$. The model polypeptide chain has a β-sheet architecture of the native state. The β-strands of the model chain are formed by native contacts between hydrophobic residues (given by *blue spheres*). The hydrophilic residues are shown (*red spheres*) and the residues forming the turn regions are given (*gray spheres*).

where U_{bond}, U_{bend}, and U_{da} are the energy terms, which determine local protein structure, and U_{nb} corresponds to nonlocal (nonbonded) interactions. The bond-length potential U_{bond}, which describes the chain connectivity, is given by a harmonic function

$$U_{bond} = \frac{k_b}{2}\sum_{j=1}^{N-1}\left(|\mathbf{X}_j - \mathbf{x}_{j+1}| - a\right)^2, \quad (14)$$

where $k_b = 100\epsilon_h/a^2$ and ϵ_h ($\approx$ 1.25 kcal/mol) is the energy unit roughly equal to the free energy of a hydrophobic contact. The bending potential U_{bend} is

$$U_{bend} = \sum_{j=1}^{N-2}\frac{k_\theta}{2}(\theta_j - \theta_0)^2, \quad (15)$$

where $k_\theta = 20\epsilon_h/rad^2$ and $\theta_0 = 105°$. The dihedral angle potential U_{da}, which is largely responsible for maintaining proteinlike secondary structure, is taken to be

$$V_{da} = \sum_{i=1}^{N-3}[A_i(1+\cos\phi_i) + B_i(1+\cos 3\phi_i)], \quad (16)$$

where the coefficients A_i and B_i are sequence-dependent. Along the β-strands, *trans*-states are preferred and $A = B = 1.2E_h$. In the turn regions (i.e., in the vicinity of a cluster of N residues), $A = 0$, $B = 0.2E_h$. The nonbonded 12-6 Lennard-Jones interaction U_{nb} between hydrophobic residues is the sum of pairwise energies

$$U_{nb} = \sum_{i<j+2} U_{ij}, \quad (17)$$

where U_{ij} depends on the nature of the residues. The double summation in Eq. 17 runs over all possible pairs excluding the nearest-neighbor residues. The potential U_{ij}^{BB} between a pair of hydrophobic residues B is given by $U_{ij}^{BB}(r) = 4\lambda\epsilon_h[(a/r)^{12} - (a/r)^6]$, where λ is a random factor unique for each pair of B residues (39) and $r = |\mathbf{x}_i - \mathbf{x}_j|$. For all other pairs of residues $U_{ij}^{\alpha\beta}$ is repulsive (39).

Although an off-lattice CGM drastically simplifies the polypeptide chain structure, it does retain important characteristics of proteins, such as chain connectivity and the heterogeneity of contact interactions. The local energy terms in S1 provide accurate representation of the protein topology. The native structure of $S1$ is a β-sheet protein that has a topology similar to the much-studied immunoglobulin domains (Fig. 3). When the model sequence is subject to $\mathbf{f}_S$ or $\mathbf{f}_Q$, the total energy is written as $U_{tot} = U - \mathbf{f}_\alpha\mathbf{X}$ ($\alpha = S$ or Q), where $\mathbf{X}$ is the protein end-to-end vector, and $\mathbf{f}_{S,\,Q} = (f_{S,Q}, 0, 0)$ is applied along the $\mathbf{x}$-direction (Fig. 1).

The dynamics of the polypeptide chain is assumed to be given by the overdamped Langevin equation—which, in the absence of $\mathbf{f}_S$ or $\mathbf{f}_Q$, is

$$\eta\frac{d}{dt}\mathbf{x}_j = -\frac{\partial U_{tot}}{\partial \mathbf{x}_j} + \mathbf{g}_j(t), \quad (18)$$

where η is the friction coefficient and $\mathbf{g}_j(t)$ is a Gaussian white noise, with the statistics

$$\langle \mathbf{g}_j(t)\rangle = 0,\ \langle \mathbf{g}_i(t)\mathbf{g}_j(t')\rangle = 6k_BT\eta\delta_{ij}\delta(t-t'). \quad (19)$$

Equation 18 is integrated with a step size $\delta t = 0.02\tau_L$, where $\tau_L = (ma^2/\epsilon_h)^{1/2} = 3$ ps is the unit of time and $m \approx 3\times10^{-22}g$ is a residue mass. In Eq. 18, the value of $\eta = 50\ m/\tau_L$ corresponds roughly to water viscosity.

RESULTS

Simulations of unfolding and refolding of *S*1

For the model sequence $S1$ we have previously shown that the equilibrium critical unfolding force is $f_C \approx 22.6$ pN (38) at the temperature $T_s = 0.692\epsilon_h/k_B$ below the folding transition temperature $T_F = 0.7\epsilon_h/k_B$. At this temperature 70% of native contacts are formed (see the phase diagram in (38)). To simulate the stretch-relax trajectories, the initially folded structures in the NBA were equilibrated for 60 ns at T_s. To probe forced unfolding of S1 at $T = T_s$, constant pulling force $f_S = 40$ pN and 80 pN was applied to both terminals of $S1$. For these values of f_S, $S1$ globally unfolds in $t = 90$ ps and 50 ps, respectively. Cycles of stretching were interrupted by relaxation intervals during which the force is abruptly quenched to $f_Q = 0$ for various durations of T. Unfolding-refolding trajectories of $S1$ have been recorded as a time-series of X and the number of native contacts Q.

In Fig. 4 we present a single unfolding-refolding trajectory of X and Q of $S1$, generated by stretch-relax cycles. Stretching cycles of constant force $f_S = 80$ pN applied for 30 ns are interrupted by periods of quenched force relaxed over 90 ns. A folding event is registered if it results in the formation of 92% of the total number of native contacts $Q_F = 106$, i.e., $Q \geq 0.92\ Q_F$ for the first time. An unfolding time is defined as the time of rupture of 92% of all possible contacts for the first time. With this definition, the unfolded state end-to-end distance is $X \geq X_U \approx 36a$. In Fig. 4, folded (unfolded) states correspond to minimal (maximal) X and maximal (minimal) Q. Inspection of Fig. 4 shows that refolding events are essentially stochastic. Out of 36 relaxation periods only nine attempts resulted in refolding of $S1$. Both X and Q show that refolding of $S1$ occurs though an initial collapse to a coiled state with the end-to-end distance $X_C/a \approx 15$ ($Q \approx 20$), followed by the establishment of additional native contacts ($Q \approx 90$) stabilizing the folded state with $X_F/a \approx$ (1–2).

We generated ~1200 single unfolding-refolding trajectories and monitored the time-dependent behavior of X and Q. In the first set of simulations we set $f_S = 40$ pN and used several values of $T = 24$, 54, 102, 150, and 240 ns. In the second set, $f_S = 80$ pN, and $T = 15$, 48, 86, 120, and 180 ns. Each trajectory involves four stretching cycles separated by three relaxation intervals in which $f_Q = 0$. Typical unfolding-refolding trajectories of X and Q for $f_S = 40$ pN, $f_Q = 0$, and $T = 102$, 150, and 240 ns are displayed in Figs. 5–7, respectively. Due to finite duration of stretching cycles (90 ns), unfolding of $S1$ failed in few cases—which were not included in the subsequent analysis of unfolding times. Only first stretching cycles in each trajectory are guaranteed to start from the NBA, and for $T = 102$ ns (Fig. 5), relatively few relaxation intervals result in refolding (with large Q). This implies that the distribution of unfolding times $P(T, t)$ obtained from these trajectories are dominated by contributions from the coiled states, with the kinetics of formation of the native contacts playing only a minor role. Not unexpectedly, refolding events are more frequent when T is increased to 150 ns and 240 ns. At $T = 150$ ns, Q reaches higher values (≈65–75) and the failure to refold is rare (Fig. 6). This implies that as T starts to exceed the (re)folding time τ_F, the

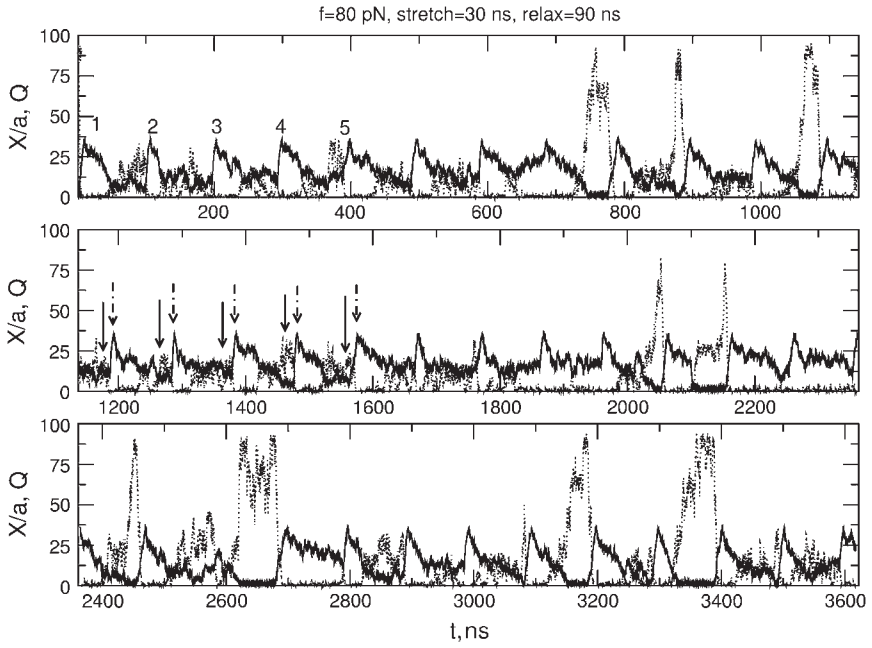

FIGURE 4 A single unfolding-refolding trajectory of the end-to-end distance X/a (*solid lines*) and the total number of native contacts Q (*dotted lines*) as a function of time t for $S1$. The trajectory is obtained by repeated application of stretch-quench cycles with stretching force $f_S = 80$ pN and quenched force $f_Q = 0$. The duration of stretching cycle and relaxation period is 30 ns and 90 ns, respectively. The first five unfolding events corresponding to large X/a and small Q are marked explicitly by numbers 1, 2, 3, 4, and 5. Force-stretch and force-quench for the stretch-quench cycles 13, 14, 15, 16, and 17 (*middle panel*) are denoted by solid and dash-dotted arrows.

distribution of unfolding events, parameterized by $P(T, t)$, is characterized by diminishing contribution from the coiled states $\{C\}$ and is increasingly dominated by the folded conformations in the NBA. Note that failed refolding events are observed even at $T = 240$ ns (Fig. 7), which implies large heterogeneity in the duration of folding barrier crossing events. Figs. 5–7 suggest that the folding time τ_F at the temperature T_U is in the range 100–240 ns. Direct computations of the folding time τ_F from hundreds of folding trajectories starting with the fully stretched states gives $\bar{\tau}_F \approx 176$ ns. The agreement between $\bar{\tau}_F$ and τ_F validates our stretch-release simulations.

Analysis of the distribution of unfolding times of *S*1

The theoretical considerations in our formalism suggest that the T-dependent heterogeneous unfolding processes occur not only from the NBA but also from the intermediate coil $\{C\}$ states. The T-dependent protein dynamics can be

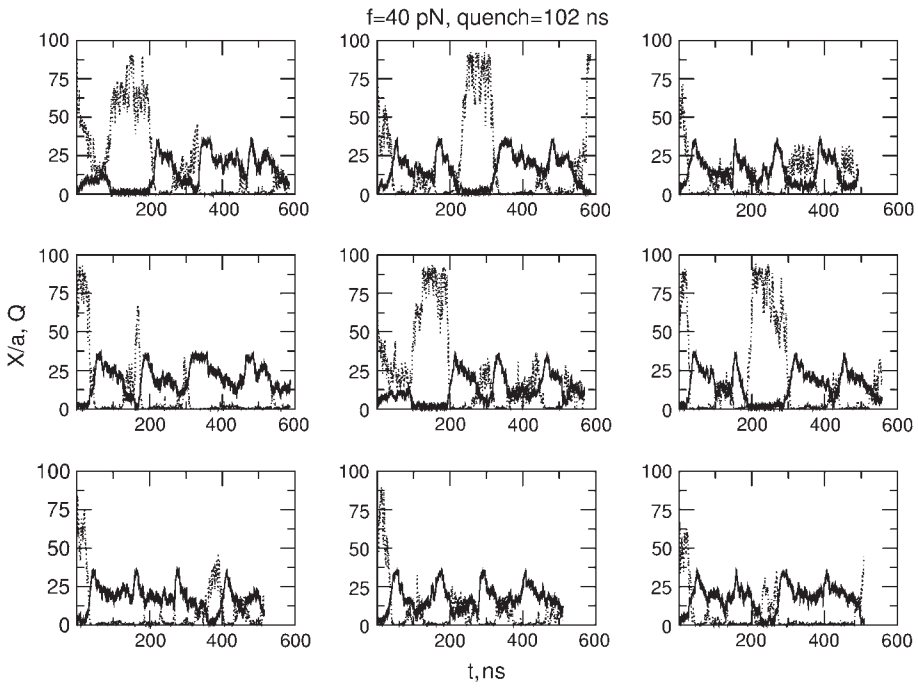

FIGURE 5 Typical unfolding-refolding trajectories of X/a (*solid lines*) and Q (*dotted lines*) for $S1$ as functions of time t, simulated by applying four stretch-quench cycles at the pulling force $f_S = 40$ pN and quenched force $f_Q = 0$. The duration of relaxation time $T = 102$ ns.

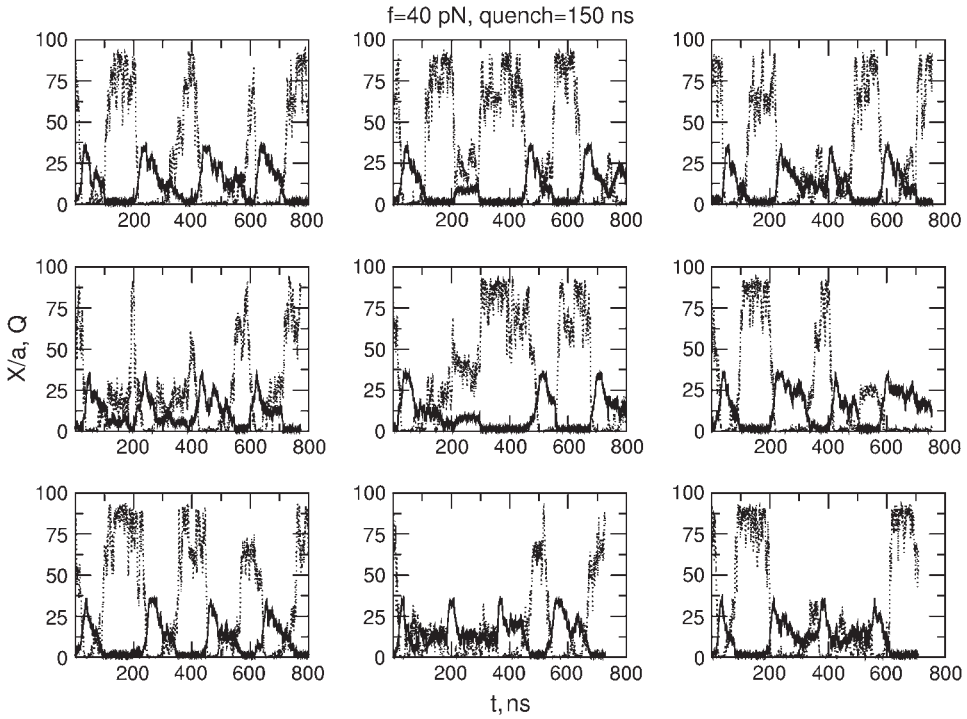

FIGURE 6 Examples of unfolding-refolding trajectories of X/a (*solid lines*) and Q (*dotted lines*) for $S1$ as a function of time t. The pulling force is $f_S = 40$ pN and the quenched force is $f_Q = 0$. The duration of relaxation time $T = 150$ ns.

utilized to separately probe the coil dynamics of the polypeptide chain and the kinetics of formation/rupture of native contacts (Q). We now utilize unfolding-refolding trajectories of $S1$, simulated for short, intermediate and long T, to build the histograms of unfolding times $P(T, t)$. Using $P(T, t)$ we provide quantitative description of the polypeptide chain dynamics in the coil state and the kinetics of rupture and formation of native interactions by employing CTRW model for Q and Gaussian statistics for X.

We computed $P(T, t)$ using the distribution of unfolding times obtained for $f_S = 80$ pN, $T = 15$, 48, and 86 ns (Fig. 8), and $f_S = 40$ pN, $T = 24$, 54, and 102 ns (Fig. 9). In both cases $f_Q = 0$. We excluded unfolding times corresponding to the first stretch-quench cycle of each trajectory, which were used

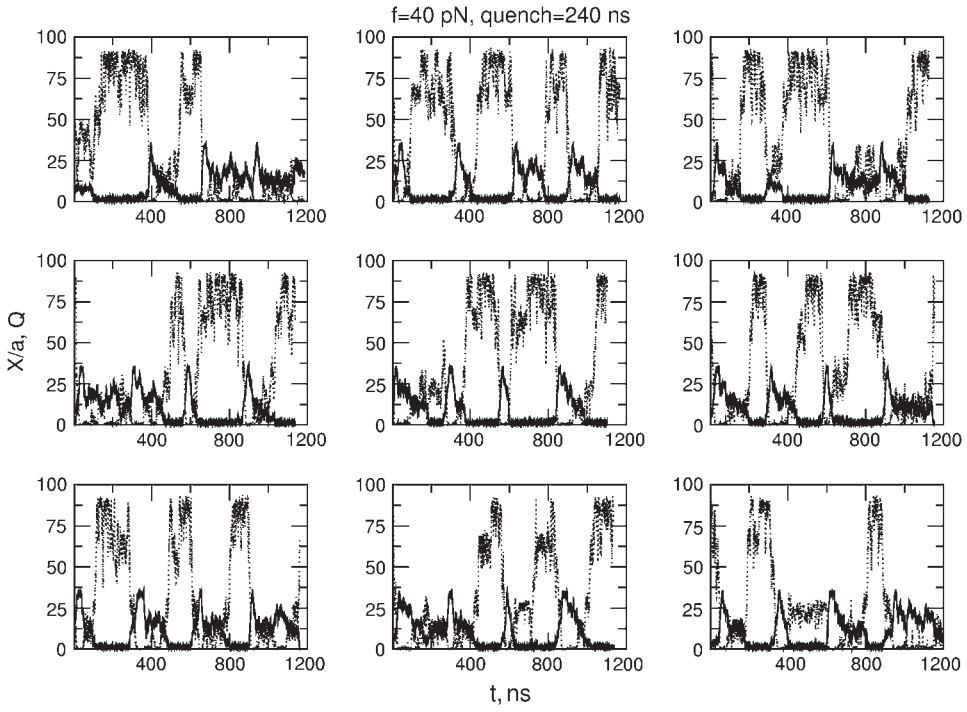

FIGURE 7 Same as Fig. 6, except $T = 240$ ns.

to construct $P(t)$ for the purposes of comparing $P(t)$ with $P(T, t)$ for long T. Single peaked $P(T, t)$ obtained for $T = 15$ ns (Fig. 8) and $T = 24$ ns (Fig. 9), represent contributions to $S1$ unfolding from coil manifold $\{C\}$ alone. When T is increased to 48 ns (Fig. 8) and 54 ns (Fig. 9), position of the peak shifts to longer times, i.e., from $t \approx 2.5$ ns to $t \approx 5$ ns (Fig. 8) and from $t \approx 6$ ns to $t \approx 10$ ns (Fig. 9). Furthermore, $P(T, t)$ develops a shoulder at $t \approx 10$ ns and $t \approx 25$ ns, observed for $T = 86$ ns (Fig. 8) and $T = 102$ ns (Fig. 9), which indicates a growing (with T) contribution to unfolding from relaxation trajectories that reach the NBA. At longer $T = 150$ ns, when most relaxation periods result in refolding of $S1$, contribution from coiled states diminishes and at $T = 240$ ns, $P(T, t)$ is identical to the standard distribution $P(t)$ constructed from unfolding times of the first stretch-quench cycle of each trajectory. This implies that for $f_Q = 0$, $\tau_F \approx 240$ ns and that $P(T, t) \rightarrow P(t)$ for $T > 240$ ns. The distribution $P(T, t) = P(t)$ constructed from unfolding times separated by $T = 300$ ns is presented in Figs. 8 and 9 (*top left panel*).

We use the CTRW formalism to analyze the histograms of unfolding times $P(T, t)$ from which the parameters that characterize the energy landscape of $S1$ can be mapped. We describe the kinetics of rupture and formation of native contacts by the waiting-time distributions Ψ_r, Ψ_f,

$$\Psi_r(t) = N_r t^{\nu_r - 1} e^{-k_r t}, \quad \Psi_f(t) = N_f t^{\nu_f - 1} e^{-k_f t}, \qquad (20)$$

where k_r (dependent on f_S) and k_f (dependent on f_Q) are the rates of rupture and formation of native interactions, respectively, $N_{r,f} = k_{r,f}/\Gamma(\nu_{r,f})$ are normalization constants ($\Gamma(x)$ is γ-function), and $\nu_{r,f} \geq 1$ are phenomenological parameters quantifying the deviations of the kinetics from a Poissonian process. For instance, $\nu_{r,f} = 1$ implies Poissonian process and corresponds to standard chemical kinetics with constant rate $k_{r,f}$. We assume that both the folded and the unfolded states are sharply distributed around the mean native and unfolded end-to-end distance $\langle X_F \rangle$ and $\langle X_U \rangle$, respectively (Fig. 2),

$$P_{eq}(X_F) = \delta(X - \langle X_F \rangle), \quad \text{and} \quad P(X_U) = \delta(X - \langle X_U \rangle), \qquad (21)$$

where $\langle X_U \rangle / a = 36$ residues corresponds to the definition of unfolded state. For $S1$ the contour length $L/a = 46$. Thus, $S1$ is unfolded if X/a exceeds $\langle X_U \rangle$, which implies $\delta/a = 10$ residues (see Fig. 2 and the lower limit of integration in Eq. 1). We describe the distribution of states $\{C\}$ before the transition to the NBA by a Gaussian,

$$P_C(X) = e^{-(X - \langle X_C \rangle)^2 / 2\Delta X_C^2}, \qquad (22)$$

with the width, ΔX_C, centered around the average distance, $\langle X_C \rangle$.

We performed numerical fits of the histograms presented in Figs. 8 and 9 using Eqs. 1–4. By fitting the theoretical curves to $P(T, t)$ constructed from short $T = 15$ ns and $T = 48$ ns simulations (Fig. 8) and $T = 24$ ns and $T = 54$ ns (Fig. 9), we first studied the dynamics of X to estimate the dynamical timescale τ_d, i.e., the longest relaxation time corresponding to the smallest eigenvalue z_n (Eq. 12), and persistence length l_p of $S1$ in the coil states $\{C\}$. By using the values of τ_d and l_p, we used our theory to describe $P(T, t)$ constructed from long $T = 300$ ns simulations. This analysis allows us to estimate the parameters characterizing the rupture of native contacts k_r, ν_r, $\langle X_F \rangle$, and ΔX_F. Finally, the parameters k_f, ν_f, $\langle X_C \rangle$, and ΔX_C, characterizing formation of native contacts, were estimated using τ_d, l_p, k_r, ν_r, $\langle X_F \rangle$, and ΔX_F, and fitting Eqs. 3 and 4 to $P(T, t)$ for intermediate $T = 86$ ns (Fig. 8) and $T = 102$ ns (Fig. 9).

Extracting the energy landscape parameters of S1

There are a number of parameters that characterize the energy landscape and the dynamics of the major components in the NBA $\rightarrow$ U transition. The numerical values of the model parameters are summarized in Table 1. The values of $\nu_r = 6.9$

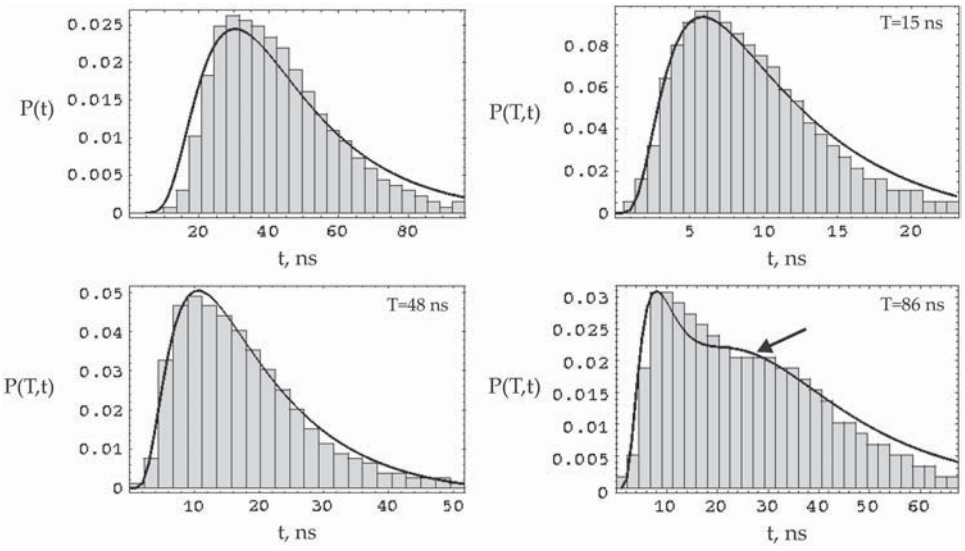

FIGURE 8 Histograms of forced unfolding times $P(t)$ and the joint distributions of unfolding times separated by relaxation periods of the quenched force $P(T, t)$. The distribution functions are constructed from single unfolding-refolding trajectories of $S1$ simulated in stretch-quench cycles of $f_S = 80$ pN and $f_Q = 0$ for $T = 15$ ns, 48 ns, and 86 ns. Simulated distributions are shown by shaded bars with the contribution to global unfolding events from coiled conformations $\{C\}$ indicated by an arrow for $T = 86$ ns. The results of the numerical fits obtained by using Eqs. 1–4 are represented by solid lines. The energy landscape parameters of $S1$ are summarized in Table 1.

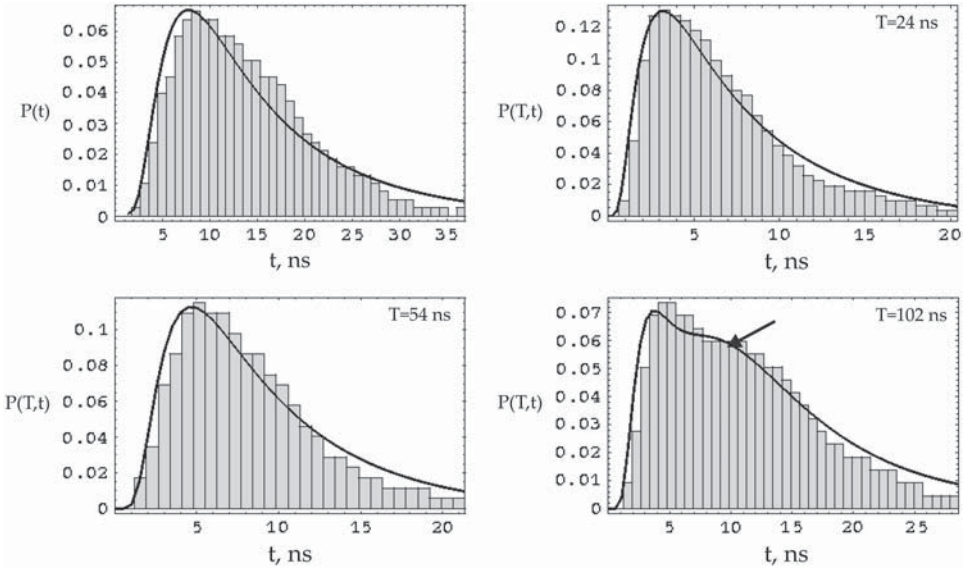

FIGURE 9 Histograms of forced unfolding times $P(t)$ and $P(T, t)$ constructed from single unfolding-refolding trajectories for $S1$. The stretch-quench cycles were simulated with $f_S = 40$ pN and $f_Q = 0$ for $T = 24$ ns, 54 ns, and 102 ns. Simulated distributions are shown by shaded bars with the contribution to global unfolding events from coiled conformations $\{C\}$ indicated by an arrow for $T = 102$ ns. The results of numerical fit obtained by using Eqs. 1–4 are represented by solid lines. The values of the parameters are given in Table 1.

for $f_S = 40$ pN and $\nu_r = 5.1$ for $f_S = 80$ pN indicate that rupture of native contacts is highly cooperative, especially at the lower $f_S = 40$ pN. This agrees with the previous findings on kinetics of forced unfolding of $S1$ (38), which were based solely on unfolding $S1$ by applying a constant force. In contrast, the formation of native contacts is characterized by $\nu_f \approx 1$, implying an almost-Poissonian distribution for the kinetics of formation of native contacts. The structural characteristics of the coil states are obtained using the relaxation of the polypeptide chain upon force-quench from stretched states. The value of the persistence length l_p, which should be independent of $\mathbf{f}_Q$ provided $\mathbf{f}_Q/\mathbf{f}_C \ll 1$, is found to be ~4.8 Å (Table 1). This value is in accord with the results of the recent experimental measurements based on kinetics of loop formation in denatured states of proteins (48).

Upon rupture of native contacts, the chain extends by $\Delta X_F/a = 6.4$ (for $f_S = 40$ pN) and $\Delta X_F/a = 6.7$ (for $f_S = 80$ pN). This distance separates the basins of folded states with $\langle X_F\rangle/a = 4.5$ at $f_S = 40$ pN and $\langle X_F\rangle/a = 4.6$ at $f_S = 80$ pN from high free-energy states when the polypeptide chain is stretched in the direction of $\mathbf{f}_S$ (Fig. 2 *a*). Because these high free-energy states are never populated, we expect that forced-unfolding of $S1$ must occur in an apparent two-step manner when $T \to \infty$. Explicit simulations of $S1$ unfolding at constant $\mathbf{f}_S$ (≈69 pN) shows that mechanical unfolding occurs in a single step (see Fig. 2 in (38)).

From the refolding free-energy profile upon force-quench (see Fig. 2 *b*) we infer that the initial stretched conformation must collapse to an ensemble of compact structures $\{C\}$. From the analysis of $P(T;t)$ using the CTRW formalism we find that the average end-to-end distance $\langle X_C\rangle$ for the manifold $\{C\}$ is close to $\langle X_F\rangle$ (see Table 1), which suggests that the ensemble of the $\{C\} \to$ NBA transition states is close to the native state. There is a broad distribution of coiled states $\{C\}$, which is manifested in the large width $\Delta X_C/a = 2.2$. Due to the broad conformational distribution, there is substantial heterogeneity in the refolding pathways. This feature is reflected in the long tails in $P(T, t)$ (see Figs. 8 and 9). As a result, we expect the kinetic transition to be sharp. The estimated timescale ($\sim 1/k_f$) for forming native contacts for

TABLE 1 Energy landscape parameters for *S1* extracted from FCS

f_S, pN*	$l_p/a^\dagger$	τ_d, ns‡	k_r, 1/ns§	$\nu_r^\P$	$\langle X_F\rangle/a^\parallel$	$\Delta X_F/a$**	k_f, 1/ns††	$\nu_f^{\ddagger\ddagger}$	$\langle X_C\rangle/a^{\S\S}$	$\Delta X_C/a^{\P\P}$
40	1.2	19.6	0.02	6.9	4.5	6.4	0.26	1.1	4.8	2.2
80	1.1	15.2	0.11	5.1	4.6	6.7	0.25	1.1	4.7	2.2

*f_S is the magnitude of the stretching force.
†l_p is the persistence length of $S1$ in the coiled state (Eq. 10) measured in units of a (≈Å).
‡τ_d is the f_Q-dependent longest relaxation time in the coil state (Eq. 12).
§k_r (k_f) is the rate of rupture (formation) of native interactions (Eq. 20) and is a function of f_S (f_Q).
¶ν_r (ν_f) quantifies deviations of the native contacts rupture (formation) kinetics from the Poisson process.
‖$\langle X_F\rangle$ ($\langle X_C\rangle$) is the average end-to-end distance of $S1$ in the NBA (manifold $\{C\}$) (Fig. 2 *b*, Eqs. 21 and 22).
**ΔX_F is the extension of the chain before rupture of all native contacts (Fig. 2 *a* and Eq. 2).
††k_r (k_f) is the rate of rupture (formation) of native interactions (Eq. 20) and is a function of f_S (f_S).
‡‡ν_r (ν_f) quantifies deviations of the native contacts rupture (formation) kinetics from the Poisson process.
§§$\langle X_F\rangle$ ($\langle X_C\rangle$) is the average end-to-end distance of $S1$ in the NBA (manifold $\{C\}$) (Fig. 2 *b*, Eqs. 21 and 22).
¶¶ΔX_C is the width of the distribution of coiled states of $S1$ (Eq. 23), a measure of the refolding heterogeneity.

$S1$ is shorter than the coil dynamical timescale τ_d (for the values of f_S used in the simulations). This indicates that the dynamical collapse of $S1$ from the stretched state $X_U \approx L$ and equilibration in the coiled manifold $\{C\}$ constitutes a significant fraction of the total folding time ($\approx \tau_d + k_f^{-1}$). From the analysis of folding of $S1$ ($P(T;t)$ at intermediate T) we also infer that the transition state ensemble for $\{C\} \rightarrow N$ must be narrow.

From the rates of rupture of native contacts k_r at the two f_S values and assuming the Bell model for the dependence of k_r on f_S,

$$k_r(f_S) = k_r^0 e^{\sigma f_S / k_B T}, \tag{23}$$

we estimated the force-free rupture rate k_r^0 and the critical extension σ, at which folded states of $S1$ become unstable. We found that $k_r^0 = 8\times10^{-4}\,\mathrm{ns}^{-1}$ is negligible compared to the rate of formation of native contacts, $k_f = 0.25\,\mathrm{ns}^{-1}$. The location of the transition state of unfolding $X = \langle X_F \rangle + \sigma$ is characterized by $\sigma = 1.5\,a \approx 0.03\,L$. The value of σ is short compared to ΔX_C, which is a measure of the width of the $\{C\}$ manifold. Small σ implies that the major barrier to unfolding is close to the native conformation. A similar values of σ was obtained in the previous study of $S1$ by using an entirely different approach (38). These findings are consistent with AFM experiments (49) and computer simulations (50), which show that native structures of proteins appear to be brittle upon application of mechanical force.

The parameter τ_d is an approximate estimate of the collapse time, τ_c, from the stretched to the coiled state. Using direct simulations of the decay of the radius of gyration, R_g, starting from a rodlike conformation, we obtained $\tau_c \approx 80$ ns (see the Supplementary Information in (51)). The value of τ_d (≈ 20 ns) is in reasonable agreement with the estimate of τ_c. This exercise shows that reliable estimates of timescales of conformational dynamics, which are difficult to obtain, can be made using FCS. To ascertain the extent to which the estimate of K_U agrees with independent calculations, we obtained the K_U by applying a constant force to unfold $S1$. The value of K_U, obtained by averaging over 200 trajectories, is ~90 ns at $f_S = 40$ pN, which is in rough accord with $K_U \approx \tau_d + k_r^{-1} \approx 70$ ns. This further validates the efficacy of FCS in obtaining the energy landscape of proteins. We also estimated K_U^0 from the value of K_U obtained by direct simulation and the Bell model. The $\mathbf{f}_S$-dependent unfolding rate $K_U \approx \tau_d + k_r^{-1}$ increases with $\mathbf{f}_S$ in accord with Eq. 23. The prefactor (K_U^0) is ~10-fold smaller than k_r^0. The difference may be either due to the failure of the assumption that $k_r^0 = K_U^0$, or to the breakdown of the Bell model (52).

DISCUSSION

In this section we summarize the main steps for practical implementation of the proposed force correlation spectroscopy (FCS) to probe the energy landscape of proteins using forced unfolding of proteins.

Step 1: Evaluating the (re)folding timescale τ_F

In the first phase of the FCS experiments, one needs to collect a series of histograms $P(T_n, t)$, $n = 1, 2, \ldots, N$ of unfolding times for increasing relaxation time $T_1 < T_2 < \ldots < T_N$ by repeated stretch-release experiments. This can be done by discarding the first unfolding time t_1 in the sequence of recorded unfolding times $\{t_1, t_2, \ldots, t_M\}$ for each T_n to guarantee that all the unfolding events are generated from the stretched states with the distribution $P(X_U)$ (see Eq. 1). This is a crucial element of the FCS methodology since it enables us to perform the averaging over the final (stretched) states. It is easier to resolve experimentally the end-to-end distance $X \approx L$, rather than the initial (folded) states in which a number of conformations belong to the NBA. The histograms are compared with $P(T^*, t)$ obtained for sufficiently long $T^* \gg \tau_F$. To ensure that T^* exceeds τ_F, T^* can be as long as a few tens of minutes. The time at which $D(T_n)$, given by Eq. 5, is equal to zero can then be used to estimate τ_F. Notice that our estimate of τ_F does not hinge on whether $P(T \rightarrow \infty;t) \equiv P(t)$ is Poissonian or not! Clearly, the choice of T^* depends on the protein under the study, and prior knowledge or bulk measurements of unfolding times observed under the influence of temperature jump or denaturing agents can serve as a guide to estimate the order of magnitude of T^*.

Step 2: Resolving the dynamics of the polypeptide chain

To this end we have determined the ensemble average (re)folding time, τ_F. In the second phase of the FCS, we perform statistical analysis of the distribution of unfolding times collected at $T \ll \tau_F$, i.e., $P(T \ll \tau_F;t)$ (see Regime I, above). This allows us to probe the dynamic properties of the polypeptide chain, such as the protein persistence length, l_p, and the protein dynamical timescale, τ_d (see Table 1). Indeed, by assuming a reasonable model for the conditional probability, $G(X', t;X)$, of the protein end-to-end distance and the distribution of the stretched states, $P(X_U)$, the values l_p and τ_d can be determined from the fit (either analytically or numerically) of the unfolding time distribution, $P(T \ll \tau_F;t)$, given by Eq. 1, to the histogram of unfolding times collected for $T \ll \tau_F$. To illustrate the utility of the FCS, in this work we assumed a Gaussian profile for $G_{S,Q}(X', t;X)$ (see Eq. 11) and the wormlike chain model for the polypeptide chain. The general formulae (1) allow for the use of more sophisticated models of X, should it become necessary. Recent single molecule FRET experiments on proteins (53,54), dsDNA, ssDNA, and RNA (55) show, surprisingly, that the characteristics of unfolded states obey wormlike chain models. Moreover, all the data in forced unfolding of proteins have been analyzed using WLC models. Thus, the analysis of FCS data using WLC dynamics for unfolded polypeptide chains to a large extent is justified. The values $G_S(X', t;X)$ and $G_Q(X', t;X)$ can be measured in the current AFM and LOT experiments

by computing the frequency of occurrence of the event X after the forced-stretch ($\mathbf{f} = \mathbf{f}_S$) or force-quench ($\mathbf{f} = \mathbf{f}_Q$) from the well-controlled partially stretched state X or the fully stretched state $X \approx L$ of the chain, respectively, over time t ($\ll \tau_F$).

Step 3: Probing the kinetics of rupture of the protein native contacts

Having resolved the dynamics of the protein in extension-time regime, where the number of native interactions that stabilize the native state is small, we can resolve the kinetics of forced rupture of native interactions stabilizing the NBA (see Regime II, above). In the third part of the FCS we analyze the distribution of unfolding times for $T \gg \tau_F$, given by Eq. 2. We use the knowledge about the propagator $G_S(X', t';X, t)$, appearing in the right-hand side of Eq. 2, obtained in Step 2 to perform analytical or numerical fit of the distribution $P(T \gg \tau_F;t)$ to the histogram of unfolding times collected for $T \gg \tau_F$. The new information, gathered in Step 3, sheds the light on the kinetics of native interactions stabilizing the NBA, which is contained in the probability $P_F(t;\mathbf{f}_S, X_F)$ (see Eq. 2). For convenience, we used the continuous time random walk (CTRW) model for $P_F(t;\mathbf{f}_S, X_F)$, which is summarized in Eqs. 7–9, and the assumption of separability, given by Eq. 6. CTRW reduces to the Poissonian kinetics with the rate constants when the waiting-time distribution function for the rupture of native contacts, $\Psi_r(t)$, is an exponential function of t. The CTRW probes the possible deviations of the kinetics of $P_F(t;\mathbf{f}_S, X_F)$ from the Poisson process and allows us to test different functional forms for $\Psi_r(t)$. In the simplest implementation of CTRW utilized in this work, $\Psi_r(t)$ is assumed to be an algebraic function of t, given by Eq. 20, which allows us to estimate the rate of rupture of native interactions, k_r, and parameter ν_r quantifying the deviations of the rupture kinetics from a Poissonian process. Furthermore, by repeating Step 3 for different values of the stretching force, f_S, and assuming the Bell model for $k_r(f_S)$, given by Eq. 23, we can also estimate the force-free rupture rate, k_r^0, and the critical extension σ, which quantifies the distance from the NBA to the transition state along the direction of f_S. We also obtain the average end-to-end distance in the folded state, $\langle X_F \rangle$ from the distribution of the native states, $P_{eq}(X_F)$.

Step 4: Resolving the kinetics of formation of native interactions

In the final step the distributions $P(T \ll \tau_F;t)$ and $P(T \gg \tau_F;t)$, analyzed in Steps 2 and 3, respectively, are used to form a linear superposition (see Eq. 3 in Regime III, above). The T-dependent weights are given by the probabilities $\rho_C(T)$ and $\rho_F(T) = 1 - \rho_C(T)$, respectively. This superposition is used to fit the histogram of unfolding times, $P(T \sim \tau_F;t)$, collected for $T \sim \tau_F$. The estimated probability $\rho_F(T)$ should then be matched with the probability obtained by performing double integration in Eq. 4. This allows us to probe the kinetics of formation of native contacts, $P_C(T;X, \mathbf{f}_Q)$, for the known propagator $G_Q(X', T;X)$ analyzed in Step 2. As in the case of $P_F(t;\mathbf{f}_S, X_F)$, we assumed separability condition for $P_C(t;\mathbf{f}_Q, X_C)$ (Eq. 6) and CTRW for the kinetics of formation of native contacts contained in $P_f(t;\mathbf{f}_Q)$ (see Eqs. 7–9). A simple algebraic form for the waiting-time distribution function, $\Psi_f(t)$, given by Eq. 20, allows us to estimate the force-free rate of formation of native interactions, $k_f(f_Q = 0) = k_f^0$. Moreover, the heterogeneity of the protein-folding pathways can be assessed by analyzing the width, ΔX_C, of the distribution of coiled protein states, $P_C(X)$, centered around the average end-to-end distance, $\langle X_C \rangle$ (see Eq. 22). Similar to the analysis of rupture kinetics, Step 4 could be repeated for the two values of the quenched force, f_Q, to yield the force-free rate of formation of native contacts, stabilizing the native fold, and the distance between $\langle X_C \rangle$ and the transition state for the formation of native contacts. For the purposes of illustration, in this work we used $f_Q = 0$.

At the minimum FCS can be used to obtain model-independent estimate of τ_F. By assuming a WLC description for coiled states, which is justified in light of a number of FRET and forced unfolding experiments, estimates of collapse times and their distribution as well as persistence length can be obtained. If CTRW model is assumed, then estimates of timescale for rupture and formation of native contacts can be made. The utility of FCS for $S1$ illustrates the efficacy of the theory. The potential of obtaining hitherto unavailable information makes FCS extremely useful.

CONCLUSIONS

In this article, we have developed a theory to describe the role of internal relaxation of polypeptide chains in the dynamics of single-molecule force-induced unfolding and force-quench refolding. To probe the effect of dynamics of the chain in the compact manifold of states that are populated in the pathways to the NBA starting from the stretched conformations, we propose using a series of stretch-release cycles. In this new class of single-molecule experiments, referred to as force correlation spectroscopy (FCS), the duration of release times (T) is varied. FCS is equivalent to conventional mechanical unfolding experiments in the limit $T \to \infty$. By applying our theory to a model β-sheet protein we have shown that the parameters that characterize the energy landscape of proteins can be obtained using the joint distribution function of unfolding times $P(T;t)$.

The experimentally controllable parameters are $\mathbf{f}_S$, $\mathbf{f}_Q$, and T. In our illustrative example, we used values of $\mathbf{f}_S$ that are ~2–4 times greater than the equilibrium unfolding force. We set $\mathbf{f}_Q = 0$, which is difficult to realize in experiments. From the schematic energy landscape in Fig. 1 it is clear that the profiles corresponding to the positions of the manifold $\{C\}$, the dynamics of $\{C\}$, and the transition state location and barrier height depend on $\mathbf{f}_Q$. The simple application, used

here for proof-of-principle purposes only, already illustrates the power of FCS. To obtain the energy landscape of $S1$ by using FCS that covers a broader range of $\mathbf{f}_S$ and $\mathbf{f}_Q$, a complete characterization of the landscape can be made. The experiments that we propose based on the new theoretical development can be readily performed using presently available technology. Indeed, the pioneering experimental setup used by Fernandez and Li (23), who utilized force to initiate refolding, can be readily adopted to perform single molecule FCS.

It is known that even for proteins that fold in an apparent two-state manner the energy landscape is rough (21). The scale of roughness ΔE can be measured in conventional AFM experiments by varying temperature. The extent to which the internal dynamics of proteins is affected by ΔE, whose value is between 2 and 5 k_BT (56,57), on the force-quenched refolding is hard to predict. These subtle effects of the energy landscape can be resolved (in principle) using FCS in which temperature is also varied.

APPENDIX: CALCULATION OF $\langle X(T)X(0)\rangle$

In this Appendix we outline the calculation of $\langle X(t)X(0)\rangle$ and $\langle X^2\rangle$ for the force-free propagator $G_0(X, t;X_0)$. By using Eq. 10 (without the last term) and applying the least-action principle to WLC Lagrangian equation, $L = m/2\int_{-L/2}^{L/2} ds(\partial\mathbf{r}/\partial t)^2 - H$, we obtain $m(\partial^2/\partial t^2)\mathbf{r}(s,t)+\epsilon(\partial^4/\partial s^4)\mathbf{r}(s,t) - 2\nu(\partial^2/\partial s^2)\mathbf{r}(s,t) = 0$, where m is the protein segment mass and $\epsilon = 3l_pk_BT/4$, $\nu = 3k_BT/2l_p$. Dynamics of the media is taken into account by including a stochastic force $f(s, t)$ with the white noise statistics, $\langle f_\alpha(s, t)\rangle = 0$, $\langle f_\alpha(s, t)f_\beta(s', t')\rangle = 2\gamma k_BT\delta_{\alpha\beta}\delta(s-s')\delta(t-t')$, where $\alpha = x, y, z$, and γ is the friction coefficient per unit coil length. In the overdamped limit, the equation of motion for $\mathbf{r}(s, t)$ is (46,47)

$$\gamma\frac{\partial}{\partial t}\mathbf{r}(s,t) + \varepsilon\frac{\partial^4}{\partial s^4}\mathbf{r}(s,t) - 2\nu\frac{\partial^2}{\partial s^2}\mathbf{r}(s,t) = \mathbf{f}(s,t), \tag{A1}$$

with the boundary conditions

$$\left[2\nu\frac{\partial}{\partial s}\mathbf{r}(s,t) - \varepsilon\frac{\partial^3}{\partial s^3}\mathbf{r}(s,t)\right]_{\pm L/2} = 0,$$
$$\left[2\nu_0\frac{\partial}{\partial s}\mathbf{r}(s,t) + \varepsilon\frac{\partial^2}{\partial s^2}\mathbf{r}(s,t)\right]_{\pm L/2} = 0, \tag{A2}$$

where $\nu_0 = 3\ k_BT/4$. We solve Eq. A1 by expanding $\mathbf{r}(s, t)$ and $\mathbf{f}(s, t)$ in a complete set of orthonormal eigenfunctions $\{\psi_n(s)\}$, i.e.,

$$\mathbf{r}(s,t) = \sum_{n=0}^{\infty}\boldsymbol{\xi}_n(t)\psi_n(s) \quad \text{and} \quad \mathbf{f}(s,t) = \sum_{n=0}^{\infty}\mathbf{f}_n(t)\psi_n(s). \tag{A3}$$

Substituting Eq. A3 into Eq. A1 and separating variables, we obtain

$$\varepsilon\frac{d^4}{ds^4}\psi_n(s) - 2\nu\frac{d^2}{ds^2}\psi_n(s) = z_n\psi_n(s) \quad \text{and}$$
$$\gamma\frac{d}{dt}\boldsymbol{\xi}_n(t) + z_n\boldsymbol{\xi}_n(t) = \mathbf{f}_n(t), \tag{A4}$$

where z_n is the n^{th} eigenvalue. The second expression in Eq. A4 for $\boldsymbol{\xi}(t)$ is solved by

$$\boldsymbol{\xi}_n(t) = \frac{1}{\gamma}\int_{-\infty}^{t} dt'\mathbf{f}_n(t')\exp\left[-\frac{(t-t')z_n}{\gamma}\right], \tag{A5}$$

and the eigenfunctions $\psi_n(s)$ are

$$\psi_0 = \sqrt{1/L}$$
$$\psi_n(s) = \sqrt{c_n/L}\left(\frac{\alpha_n}{\cos[\alpha_nL/2]}\sin[\alpha_ns] + \frac{\beta_n}{\cosh[\beta_nL/2]}\sinh[\beta_ns]\right), \quad n = 1, 3, \ldots, 2q+1$$
$$\psi_n(s) = \sqrt{c_n/L}\left(-\frac{\alpha_n}{\sin[\alpha_nL/2]}\cos[\alpha_ns] + \frac{\beta_n}{\sinh[\beta_nL/2]}\cosh[\beta_ns]\right), n = 2, 4, \ldots, 2q, \tag{A6}$$

where c_n values are the normalization constants and α_n and β_n are determined from Eq. A2,

$$\alpha_n\sin[\alpha_nL/2]\cosh[\beta_nL/2] - \beta_n^3\cos[\alpha_nL/2]\sinh[\beta_nL/2] - \frac{1}{l_p}(\alpha_n^2+\beta_n^2)\cos[\alpha_nL/2]\cosh[\beta_nL/2] = 0,$$
$$n = 1, 3, \ldots, 2q+1$$
$$\alpha_n\cos[\alpha_nL/2]\sinh[\beta_nL/2] + \beta_n^3\sin[\alpha_nL/2]\cosh[\beta_nL/2] + \frac{1}{l_p}(\alpha_n^2+\beta_n^2)\sin[\alpha_nL/2]\sinh[\beta_nL/2] = 0,$$
$$n = 2, 4, \ldots, 2q, \tag{A7}$$

and the parameters α_n and β_n are related as $\beta_n^2 - \alpha_n^2 = (1/l_p^2)$. The eigenvalues z_n are given by $z_n = \epsilon\alpha_n^4 + 2\nu\alpha^2$. Using Eqs. A3 and A5, we obtain $\langle\mathbf{r}(s,t)\mathbf{r}(s',t)\rangle = 3k_BT\sum_{n=0}^{\infty}(1/z_n)\psi_n(s)\psi_n(s')e^{-z_nt/\gamma}$. Then,

$\langle X(t)X(0)\rangle = \langle \mathbf{r}((\mathrm{L}/2),t)\mathbf{r}((\mathrm{L}/2),0)\rangle + \langle \mathbf{r}(-(\mathrm{L}/2),t)\mathbf{r}(-(\mathrm{L}/2),0)\rangle - \langle \mathbf{r}((\mathrm{L}/2),t)\,\mathbf{r}(-(\mathrm{L}/2),0)\rangle - \langle \mathbf{r}(-(\mathrm{L}/2),t)\mathbf{r}((\mathrm{L}/2),0)\rangle$, which yields Eq. 12.

This work was supported in part by a grant from the National Science Foundation through grant No. NSF CHE-05-01456.

REFERENCES

1. Labeit, S., and B. Kolmerer. 1995. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science*. 270:293–296.
2. Minaeva, A., M. Kulke, J. M. Fernandez, and W. A. Linke. 2001. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophys. J*. 80:1442–1451.
3. Ohashi, T., D. P. Kiehart, and H. P. Erickson. 1999. Dynamics and elasticity of the fibronectin matrix in living cell culture visualized by fibronectin-green fluorescent protein. *Proc. Natl. Acad. Sci. USA*. 96: 2153–2158.
4. Marshall, B. T., M. Long, J. W. Piper, T. Yago, R. P. McEver, and C. Zhu. 2003. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*. 423:190–193.
5. Evans, E., A. Leung, D. Hammer, and S. Simon. 2001. Chemically distinct transition states govern rapid dissociation of single L-selectin bonds under force. *Proc. Natl. Acad. Sci. USA*. 98:3784–3789.
6. Barsegov, V., and D. Thirumalai. 2005. Dynamics of unbinding of cell adhesion molecules: transition from catch to slip bonds. *Proc. Natl. Acad. Sci. USA*. 102:1835–1839.
7. Chicurel, M. E., C. S. Chen, and D. E. Ingber. 1998. Cellular control lies in the balance of forces. *Curr. Opin. Cell Biol*. 10:232–239.
8. Henrickson, S. E., M. Misakian, B. Robertson, and J. J. Kasianowicz. 2000. Driven DNA transport into an asymmetric nanometer-scale pore. *Phys. Rev. Lett*. 85:3057–3060.
9. Sung, W., and P. J. Park. 1996. Polymer translocation through a pore in a membrane. *Phys. Rev. Lett*. 77:783–786.
10. Muthukumar, M. 2003. Polymer escape through a nanopore. *J. Chem. Phys*. 118:5174–5184.
11. Liphardt, G., D. Smith, and C. Bustamante. 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol*. 10:279–285.
12. Allemand, J.-F., D. Bensimon, and V. Croquette. 2003. Stretching DNA and RNA to probe their interactions with proteins. *Curr. Opin. Struct. Biol*. 13:266–274.
13. Zhuang, X., L. E. Bartley, H. P. Babcock, R. Russel, T. Ha, D. Herschlag, and S. Chu. 2000. A single molecule study of RNA catalysis and folding. *Science*. 288:2048–2051.
14. Rief, M., H. Clausen-Shaunman, and H. E. Gaub. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol*. 6: 346–349.
15. Chang, K. C., D. F. Tees, and D. A. Hammer. 2000. The state diagram for cell adhesion under flow: leukocyte rolling and firm adhesion. *Proc. Natl. Acad. Sci. USA*. 97:11262–11267.
16. Weisel, J. W., H. Shuman, and R. I. Litvinov. 2003. Protein-protein unbinding induced by force: single molecule studies. *Curr. Opin. Struct. Biol*. 13:227–235.
17. Liphardt, J., S. Dumont, S. B. Smith, Jr., I. Tinoko, and C. Bustamante. 2002. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*. 296:1832–1835.
18. Bartolo, D., I. Derenyi, and A. Ajdari. 2002. Dynamic response of adhesion complexes: beyond the single path picture. *Phys. Rev. E*. 65: 051910–051913.
19. Hummer, G., and A. Szabo. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA*. 98:3658–3661.
20. Lapidus, L. J., W. A. Eaton, and J. Hofrichter. 2000. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA*. 97:7220–7225.
21. Hyeon, C., and D. Thirumalai. 2003. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl. Acad. Sci. USA*. 100:10249–10253.
22. Oberhauser, A. F., H. P. Erickson, and J. M. Fernandez. 1998. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*. 393:181–185.
23. Fernandez, J. M., and H. Li. 2004. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*. 303:1674–1678.
24. Rief, M., J. Pascual, M. Saraste, and H. E. Gaub. 1999. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol*. 286:553–561.
25. Yang, G., C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante. 2000. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl. Acad. Sci. USA*. 97:139–144.
26. Craig, D., A. Krammer, K. Schulten, and V. Vogel. 2001. Comparison of the early stages of forced unfolding for fibronectin type III modules. *Proc. Natl. Acad. Sci. USA*. 98:5590–5595.
27. Carrion-Vazquez, M., A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. USA*. 96:3694–3699.
28. Scott, K. A., A. Steward, S. B. Fowler, and J. Clarke. 2002. Titin: a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol*. 315:819–829.
29. Litvinovich, S. V., and K. C. Ingham. 1995. Interactions between type III domains in the 110 kDa cell-binding fragment of fibronectin. *J. Mol. Biol*. 248:611–626.
30. Krammer, A., H. Lu, B. Isralewitz, K. Schulten, and V. Vogel. 1999. Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proc. Natl. Acad. Sci. USA*. 96:1351–1356.
31. Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol*. 11:224–230.
32. Cieplak, M., T. X. Hoang, and M. O. Robbins. 2004. Thermal effects in stretching of Gō-like models of titin and secondary structures. *Proteins*. 56:285–297.
33. Makarov, D. E., Z. Wang, J. B. Thompson, and H. G. Hansma. 2002. On the interpretation of force extension curves of single protein molecules. *J. Chem. Phys*. 116:7760–7765.
34. Hyeon, C., and D. Thirumalai. 2005. Chemical theory and computation special feature: mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. USA*. 102:6789–6794.
35. Best, R. B., and G. Hummer. 2005. Comment on ''Force-clamp spectroscopy monitors the folding trajectory of a single protein''. *Science*. 308:498.
36. Barsegov, V., and D. Thirumalai. 2005. Probing protein-protein interactions by dynamic force correlation spectroscopy. *Phys. Rev. Lett*. 95:168301–168305.
37. Honeycutt, J. D., and D. Thirumalai. 1990. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. USA*. 87:3526–3529.
38. Klimov, D. K., and D. Thirumalai. 2000. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. USA*. 97:7254–7259.
39. Veitshans, T., D. K. Klimov, and D. Thirumalai. 1997. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des*. 2:1–22.
40. Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des*. 3:R81–R91.
41. Montroll, E. W., and H. Sher. 1975. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B*. 12:2455–2477.
42. Sher, H., and M. Lax. 1973. Stochastic transport in a disordered solid. I. Theory. *Phys. Rev. B*. 7:4491–4502.
43. Barsegov, V., Y. Shapir, and S. Mukamel. 2003. One-dimensional transport with dynamic disorder. *Phys. Rev. E*. 68:011101–011114.

44. Barsegov, V., and S. Mukamel. 2002. Probing single molecule kinetics by photon arrival trajectories. *J. Chem. Phys.* 116:9802–9810.

45. Barsegov, V., and S. Mukamel. 2004. Multipoint fluorescence quenching-time statistics for single molecules with anomalous diffusion. *J. Phys. Chem. A.* 108:15–24.

46. Harnau, L., R. G. Winkler, and P. Reineker. 1995. Dynamic properties of molecular chains with variable stiffness. *J. Chem. Phys.* 102:7750–7757.

47. Dua, A., and B. J. Cherayil. 2002. The dynamics of chain closure in semiflexible polymers. *J. Chem. Phys.* 116:399–409.

48. Lapidus, L. J., P. S. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter. 2002. Effects of chain stiffness on the dynamics of loop formation in polypeptides. (Appendix.). *J. Phys. Chem. B.* 106: 11628–11640.

49. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science.* 276:1109–1112.

50. Klimov, D. K., and D. Thirumalai. 1999. Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding. *Proc. Natl. Acad. Sci. USA.* 96:6166–6170.

51. Li, M. S., C. K. Hu, D. Klimov, and D. Thirumalai. 2006. Multiple stepwise refolding of immunoglobulin domain I27 upon force quench depends on initial conditions. *Proc. Natl. Acad. Sci. USA.* 103:93–98.

52. Hummer, G., and A. Szabo. 2003. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.* 85:5–15.

53. Schuler, B., E. A. Lipman, P. J. Steinbach, M. Kumke, and W. A. Eaton. 2005. Polyproline and the ''spectroscopic ruler'' revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 102: 2754–2759.

54. Laurence, T. A., X. Kong, M. Jaeger, and S. Weiss. 2005. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. USA.* 102:17248–17353.

55. Caliskan, G., C. Hyeon, U. Perez-Salas, R. M. Briber, S. A. Woodson, and D. Thirumalai. 2005. Persistence length changes dramatically as RNA folds. *Phys. Rev. Lett.* 95:268303.

56. Thirumalai, D., and S. A. Woodson. 1996. Kinetics of folding of proteins and RNA. *Acc. Chem. Res.* 29:433–439.

57. Nevo, R., V. Brumfeld, R. Kapon, P. Hinterdorfer, and Z. Reich. 2005. Direct measurement of protein energy landscape roughness. *EMBO Rep.* 6:482–486.