CrossMark
click for updates

PNAS PLUS

BIOPHYSICS AND
COMPUTATIONAL BIOLOGY

CHEMISTRY

Force-dependent switch in protein unfolding pathways and transition-state movements

Pavel I. Zhuravlev[a,1], Michael Hinczewski[b,1], Shaon Chakrabarti[a], Susan Marqusee[c], and D. Thirumalai[a]

[a]Biophysics Program, Institute for Physical Science and Technology, Department of Chemistry & Biochemistry, University of Maryland, College Park, MD 20742; [b]Department of Physics, Case Western Reserve University, Cleveland, OH 44106; and [c]California Institute for Quantitative Biosciences, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Although it is known that single-domain proteins fold and unfold by parallel pathways, demonstration of this expectation has been difficult to establish in experiments. Unfolding rate, $k_u(f)$, as a function of force f, obtained in single-molecule pulling experiments on src SH3 domain, exhibits upward curvature on a $\log k_u(f)$ plot. Similar observations were reported for other proteins for the unfolding rate $k_u([C])$. These findings imply unfolding in these single-domain proteins involves a switch in the pathway as f or [C] is increased from a low to a high value. We provide a unified theory demonstrating that if $\log k_u$ as a function of a perturbation (f or [C]) exhibits upward curvature then the underlying energy landscape must be strongly multidimensional. Using molecular simulations we provide a structural basis for the switch in the pathways and dramatic shifts in the transition-state ensemble (TSE) in src SH3 domain as f is increased. We show that a single-point mutation shifts the upward curvature in $\log k_u(f)$ to a lower force, thus establishing the malleability of the underlying folding landscape. Our theory, applicable to any perturbation that affects the free energy of the protein linearly, readily explains movement in the TSE in a β-sandwich (I27) protein and single-chain monellin as the denaturant concentration is varied. We predict that in the force range accessible in laser optical tweezer experiments there should be a switch in the unfolding pathways in I27 or its mutants.

protein folding | parallel pathways | single-molecule force spectroscopy

That single-domain proteins should fold by parallel or multiple pathways emerges naturally from theoretical considerations (1–3) and computational studies (4–7). The generality of the conclusions reached in the theoretical studies is sufficiently compelling, which makes it surprising that they are not routinely demonstrated in typical ensemble folding experiments. The reasons for the difficulties in directly observing parallel folding or unfolding pathways in monomeric proteins can be appreciated based on the following arguments. Consider a protein that reaches the folded state by two different pathways. The ratio of flux through these pathways is proportional to $\exp\left[\frac{\Delta G_L^{\ddagger} - \Delta G_H^{\ddagger}}{k_B T}\right]$, where $\Delta G_L^{\ddagger}$ and $\Delta G_H^{\ddagger}$ are, respectively, the free energy barriers separating the folded and unfolded states along the two pathways, k_B is the Boltzmann constant, and T is the temperature. If $\Delta\Delta G^{\ddagger} = \Delta G_H^{\ddagger} - \Delta G_L^{\ddagger} > 0$ is large compared with $k_B T$ then the experimental detection of flux through the high free energy barrier pathway (H) is unlikely. External perturbations (such as mechanical force f or denaturants [C]) could reduce $\Delta\Delta G^{\ddagger}$. However, the values of f (or [C]) should fall in an experimentally accessible range for detecting a potential switch between pathways. Despite these inherent limitations, Clarke and coworkers showed in a most insightful study that unfolding of Ig domain (I27) induced by denaturants occurs by parallel routes (8). Subsequently, additional experiments on single-chain monellin (9), using denaturants and spectroscopic probes, have firmly shown the existence of multiple paths. Thus, it seems that multiple folding routes can be detected in standard folding experiments (10, 11) provided the flux through the higher free energy barriers is not so small that it escapes detection. In addition, parallel folding

pathways have been observed in repeat proteins, where inherent symmetry in the connectivity of the individual domains (12) results in parallel assembly.

Single-molecule pulling experiments in which f is applied to specific locations on the protein have demonstrated that unfolding of many proteins follows complex multiple routes. Mechanical force, unlike denaturants, does not alter the effective microscopic interactions between the residues, thus allowing for a cleaner interpretation. More importantly, by following the fate of many individual proteins the underlying heterogeneity in the routes explored by the protein can be revealed. Indeed, pulling experiments and simulations on a variety of single-domain proteins (13–15) show clear signatures of many routes for f-induced unfolding. It could be argued that in many of these studies the network of states connecting the folded and unfolded states is a consequence of complex topology, although they are all single-domain proteins. However, the src SH3 domain is a small protein with no apparent symmetry in the arrangement of secondary structure elements that folds in an apparent two-state manner. Thus, the discovery that it unfolds using parallel pathways (16, 17) is unexpected and requires a firm theoretical explanation and structural interpretation.

In single-molecule pulling experiments, performed at constant force or constant loading rate, only a 1D coordinate, the molecular extension (x), is readily measurable. When performed at constant f, it is possible to generate folding trajectories (x as a function of time), from which equilibrium 1D free energy profiles, $F(x)$, can be extracted using rigorous theory (18). The utility of $F(x)$ hinges on the crucial assumption that all other degrees of

Significance

Single-domain proteins with symmetrical arrangement of secondary structural elements in the native state are expected to fold by diverse pathways. However, understanding the origins of pathway diversity, and the experimental signatures for identifying it, are major challenges, especially for small proteins with no obvious symmetry in the folded states. We show rigorously that upward curvature in the logarithm of unfolding rates as a function of force (or denaturants) implies that the folding occurs by diverse pathways. The theoretical concepts are illustrated using simulations of src SH3 domain, which explain the emergence of parallel pathways in single-molecule pulling experiments and provide structural description of the routes to the unfolded state. We make testable predictions illustrating the generality of the theory.
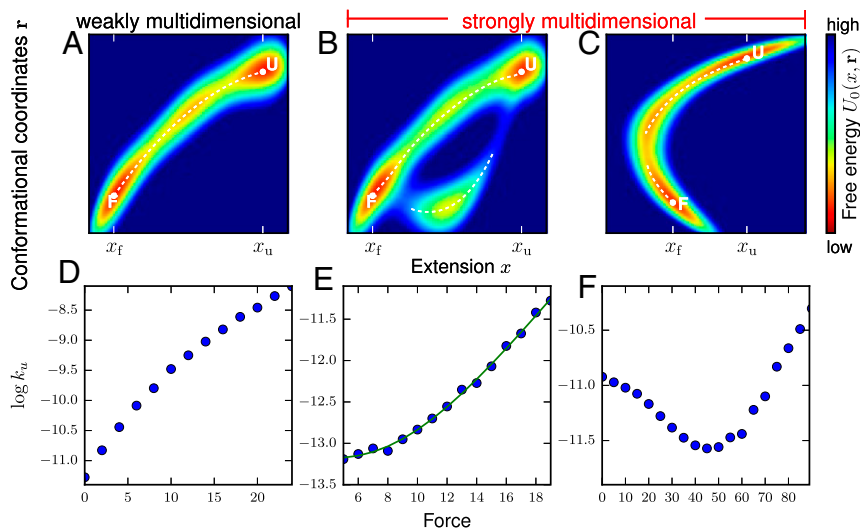
Fig. 1. (*A–C*) Schematic illustration of three multidimensional free energy landscapes $U_0(x, \mathbf{r})$ at zero force, plotted in terms of end-to-end extension x versus the other conformational degrees of freedom $\mathbf{r}$. N and U could represent the native and unfolded basins. The white dotted lines are local minima of $U_0(x, \mathbf{r})$ with respect to $\mathbf{r}$ at a given x. The landscape in *A* is weakly multidimensional with respect to x, according to the conditions described in the main text, whereas *B* and *C* are strongly multidimensional. (*D–F*) The [$\log k_\mathrm{u}(f), f$] plots for the energy landscapes given in *A–C* obtained from Brownian dynamics simulations. The WMD landscape (left) produces downward curvature. The landscape suggesting parallel pathways (*B*) produces upward curvature (*E*), where the solid line is the least squares fit to a double exponential model. The SMD in *C* produces even stronger upward curvature, giving rise to catch bond behavior with $k_\mathrm{u}(f)$ decreasing at low forces and subsequently increasing at larger f values.

freedom in the system including the solvent come to equilibrium on time scales much faster than in x, so that x itself may be considered to be an accurate reaction coordinate.

A straightforward way to assess whether a 1D picture is adequate is to analyze the f dependence of the unfolding rate $k_\mathrm{u}(f)$, which can be experimentally obtained at constant f or computed from unfolding rates measured at different loading rates (18, 19). The observed upward curvature in the [$\log k_\mathrm{u}(f), f$] plot in src SH3 (16) was shown to be a consequence of unfolding by two pathways, one dominant at low forces and the other at high forces. It was succinctly argued that the measured [$\log k_\mathrm{u}(f), f$] data cannot be explained by multiple barriers in a 1D $F(x)$ or a 1D profile with a single barrier in which the unfolding rate is usually fit using the Bell model $k_\mathrm{u}(f) = k^0 \exp\left(\frac{fx^\ddagger}{kT}\right)$, where k^0 is the unfolding rate at $f = 0$, and $x^\ddagger$ is the location of the barrier in $F(x)$ at zero force with respect to the folded state. (Throughout this paper, by location of the barrier, or the transition state, we mean the location with respect to the folded state.) The upward curvatures in the monotonic [$\log k_\mathrm{u}(f), f$] as well as [$\log k_\mathrm{u}([C]), [C]$] plots, observed experimentally, necessarily imply that parallel routes are involved in the unfolding process. (A nonmonotonic [$\log k_\mathrm{u}(f), f$] plot suggests catch bond behavior.)

To provide a general framework for a quantitative explanation of a broad class of experiments, we first present a rigorous theoretical proof that upward curvature in [$\log k_\mathrm{u}(f), f$] (or [$\log k_\mathrm{u}([C]), [C]$]) implies that the folding landscape is strongly multidimensional (SMD). Hence, such SMD landscapes cannot be reduced to 1D or superposition of physically meaningful 1D landscapes, which can rationalize the observed convex [$\log k_\mathrm{u}(f), f$] plot. We note *en passant* that the shape of the measured [$\log k_\mathrm{u}(f), f$] plot cannot be justified using $F(x)$ even if $x^\ddagger$ were allowed to depend on f, moving toward the folded state as f increases. The theory only hinges on the assumption that the perturbation (f or [C]) is linearly coupled to the effective energy function of the protein. To illustrate the structural origin of the upward curvature in the [$\log k_\mathrm{u}(f), f$] plot we also performed simulations of f-induced unfolding of src SH3 domain, to identify the structural details of the unfolding pathways including the movement of transition

states as the force is increased. The results of the simulations semiquantitatively reproduce the experimental [$\log k_\mathrm{u}(f), f$] plots for both the WT and the V61A mutant. More importantly, we also provide the structural basis of the switch in the unfolding pathways as f is varied, which cannot be obtained using pulling experiments. We obtain structures of the transition-state ensembles (TSEs), demonstrating the change in the TSE structures as f is increased from a low to a high value.

Results

Nature of the Energy Landscape from [$\log k_\mathrm{u}(f), f$] Plots. Let us assume that the unfolding rate $k_\mathrm{u}(\alpha)$, as a function of a controllable external perturbation α, can be measured. We assume that α decreases the stability of the folded state linearly, as is the case in the pulling experiment with $\alpha = f$, the force applied at two points of the protein. However, the discussion below is quite general and applies to any external parameter with a linear, additive contribution to the effective protein energy function. For a protein under force, the total free energy has the general form $U(x, \mathbf{r}, f) = U_0(x, \mathbf{r}) - fx$, with a force contribution fx, where x is the end-to-end extension of the protein. Here, U_0 is the free energy in the absence of applied tension and the vector $\mathbf{r}$ represents all of the additional conformational degrees of freedom besides x.

In the derivation below, we model the dynamics of the protein as diffusion of a single particle on the multidimensional landscape $U(x, \mathbf{r}, f)$. The unfolding of the protein would correspond to the particle starting in the reference protein conformation ($x_\mathrm{f}, \mathbf{r}_\mathrm{n}$) in the folded state energy basin F and diffusing to any other conformation, with a given extension $x_\mathrm{u} > x_\mathrm{f}$, representing the unfolded basin U (Fig. 1). The unfolding time for a particular trajectory is the time when the particle reaches the target conformation for the first time (known as first passage time). Averaging this time over all trajectories yields the mean first passage time (MFPT) from the unfolded to folded state, which we denote as $t_\mathrm{u}(x_\mathrm{f}, \mathbf{r}_\mathrm{n}, f)$, or the average unfolding time. The unfolding rate is the inverse of the unfolding time, $k_\mathrm{u}(f) \equiv 1/t_\mathrm{u}(x_\mathrm{f}, \mathbf{r}_\mathrm{n}, f)$.

We are interested in finding the curvature of $\log k_\mathrm{u}(f)$ as a function of f, and in particular the sign of $\frac{d^2}{df^2} \log k_\mathrm{u}$. Starting from the diffusion equation, we find expressions for the MFPT from any

Zhuravlev et al.

conformation with extension x_f, $t_u(x_f, \mathbf{r}, f)$, and then for $\log k_u(f)$ and its first two derivatives. It turns out that if we use the assumptions of a single unfolding pathway, the second derivative is negative and the curvature of $[\log k_u(f), f]$ has to be downward.

The summary of the subsequent derivation is as follows: (i) We start from the equation for $t_u(x_f, \mathbf{r}, f)$ which can be obtained from the diffusion equation (20), (ii) integrate it over the $\mathbf{r}$ degrees of freedom, (iii) use two assumptions for evaluating the integral with $t_u(x_f, \mathbf{r}, f)$ inside, (iv) solve the ordinary differential equation for the unfolding time, and (v) establish that the solution implies certain constraints on the shape of the $[\log k_u(f), f]$ plot. Following this derivation in detail is not necessary for understanding the other parts of the paper.

The equation for the MFPT $t_u(x, \mathbf{r}, f)$ can be obtained from the diffusion equation (in Fokker–Planck form) by integration over x, $\mathbf{r}$, and t, followed by some rearrangements (20). The result is called the backward Kolmogorov equation:

$$D(x, \mathbf{r})e^{\beta U(x, \mathbf{r}, f)}\left\{\frac{\partial}{\partial x}\left[e^{-\beta U(x, \mathbf{r}, f)}\frac{\partial}{\partial x}t_u(x, \mathbf{r}, f)\right]\right\}$$
$$+ D(x, \mathbf{r})e^{\beta U(x, \mathbf{r}, f)}\left\{\nabla_{\mathbf{r}}\cdot\left[e^{-\beta U(x, \mathbf{r}, f)}\nabla_{\mathbf{r}}t_u(x, \mathbf{r}, f)\right]\right\} = -1, \quad [1]$$

with the boundary condition $t_u(x_u, \mathbf{r}, f) = 0$, with $\beta = 1/k_B T$ and $D(x, \mathbf{r})$ being the diffusion constant, which for generality is allowed to depend on the conformation. By dividing both sides of Eq. **1** by $De^{\beta U}$ and integrating over the conformational coordinates $\mathbf{r}$, we obtain

$$\frac{\partial}{\partial x}\left[\int d\mathbf{r}\, e^{-\beta U(x, \mathbf{r}, f)}\frac{\partial}{\partial x}t_u(x, \mathbf{r}, f)\right] = -\int d\mathbf{r}\, D^{-1}(x, \mathbf{r})e^{-\beta U(x, \mathbf{r})}. \quad [2]$$

To get the result in Eq. **2**, we have assumed that $U(x, \mathbf{r}) \to \infty$ at the integration limits of the coordinate space of $\mathbf{r}$ (i.e., the diffusion process is bounded). We rewrite Eq. **2** as

$$\frac{\partial}{\partial x}\left[e^{\beta fx}\int d\mathbf{r}\, e^{-\beta U_0(x, \mathbf{r})}\frac{\partial}{\partial x}t_u(x, \mathbf{r}, f)\right] = -e^{\beta fx}G(x), \quad [3]$$

where $G(x) \equiv \int d\mathbf{r}\, D^{-1}(x, \mathbf{r})\exp(-\beta U_0(x, \mathbf{r}))$.

Further simplification of the MFPT expression depends on the nature of the multidimensional free energy $U_0(x, \mathbf{r})$. In particular, we define a class of free energies that satisfy the following two conditions:

i) $U_0(x, \mathbf{r})$ has a single minimum with respect to $\mathbf{r}$ at each point x in the range x_f to x_u. We denote the location of this minimum as $\mathbf{r}_m(x)$.
ii) The Boltzmann factor $\exp(-\beta U_0(x, \mathbf{r}))$ for $\mathbf{r}$ near $\mathbf{r}_m(x)$ is sharply peaked, so the thermodynamic contribution from conformations with coordinates far from $\mathbf{r}_m(x)$ is negligible. In other words, we assume fast equilibration along the $\mathbf{r}$ coordinates at each x, compared with the timescale of first passage between N and U.

A schematic illustration of a $U_0(x, \mathbf{r})$ satisfying these requirements is shown in Fig. 1A. Diffusion is essentially confined to a single, narrow reaction pathway in the multidimensional space. We will call any $U_0(x, \mathbf{r})$ in this category weakly multidimensional (WMD) with respect to x, because the diffusion process is quasi-1D in terms of the reaction coordinate x. In contrast, any $U_0(x, \mathbf{r})$ that violates either one of the above conditions will be called SMD, because it has characteristics that qualitatively distinguish it from any 1D diffusion process. Note that this categorization makes no other assumptions about the shape of $U_0(x, \mathbf{r})$ except for those specified above: For example, there could be one or many free energies barriers separating N and U, or none at all. Fig. 1 B and C show two examples of $U_0(x, \mathbf{r})$ that are SMD. In

both cases, condition i is violated, because in the range $x_f < x < x_u$ there is no unique minimum in $U_0(x, \mathbf{r})$ along $\mathbf{r}$. For Fig. 1B, there are two possible reaction pathways between N and U, whereas for Fig. 1C there is a single pathway, but it is nonmonotonic in x.

For an energy landscape that is WMD, there are rigorous bounds on the first and second derivatives of $\log k_u(f)$ with respect to f. To derive these bounds, note that the WMD assumptions allow us to make a saddle-point approximation to the integral over $\mathbf{r}$ on the left-hand side of Eq. **3**, setting the value of $\mathbf{r}$ in $\partial t_u(x, \mathbf{r}, f)/\partial x$ to $\mathbf{r}_m(x)$. Because this will be the dominant contribution, we approximate Eq. **3**:

$$\frac{\partial}{\partial x}\left[e^{\beta fx}\frac{\partial}{\partial x}t_u(x, \mathbf{r}_m(x), f)\int d\mathbf{r}\, e^{-\beta U_0(x, \mathbf{r})}\right] \approx -e^{\beta fx}G(x). \quad [4]$$

By simplifying the notation by defining $\tau_u(x, f) \equiv t_u(x, r_m(x), f)$ and $I(x) \equiv \int d\mathbf{r}\exp(-\beta U_0(x, \mathbf{r}))$, Eq. **4** becomes

$$\frac{\partial}{\partial x}\left[e^{\beta fx}\frac{\partial}{\partial x}\tau_u(x, f)I(x)\right] \approx -e^{\beta fx}G(x). \quad [5]$$

The solution for $\tau_u(x, f)$ from Eq. **5**, with boundary condition $\tau_u(x_u, f) = 0$, can be written as a Laplace transform of a function $H(x, s)$,

$$\tau_u(x, f) = \int_0^\infty ds\, e^{-\beta fs}H(x, s),$$
$$H(x, s) \equiv \int_x^{x_u} dx' I^{-1}(x')G(x' - s). \quad [6]$$

Both $I(x)$ and $G(x)$ are nonnegative functions [because $D(x, \mathbf{r}) > 0$ and $\exp(-\beta U(x, \mathbf{r})) \geq 0$ for all x and $\mathbf{r}$], so the function $H(x, s)$ is likewise nonnegative, $H(x, s) \geq 0$ for $x \leq x_u$. From this property, it follows that $\tau_u(x, f) \geq 0$, and

$$\frac{\partial}{\partial f}\tau_u(x, f) = -\int_0^\infty ds\, \beta s\, e^{-\beta fs}H(x, s) \leq 0,$$
$$\frac{\partial^2}{\partial f^2}\tau_u(x, f) = \int_0^\infty ds\, (\beta s)^2 e^{-\beta fs}H(x, s) \geq 0, \quad [7]$$

for $x \leq x_u$. Because the experimental data are typically plotted in terms of $\log k_u(f) = -\log \tau_u(x_f, f)$ with respect to f, we are specifically interested in the corresponding derivatives of $\log k_u(f)$,

$$\frac{d}{df}\log k_u = -\frac{1}{\tau_u}\frac{\partial \tau_u}{\partial f}, \qquad \frac{d^2}{df^2}\log k_u = \frac{1}{\tau_u^2}\left[\left(\frac{\partial \tau_u}{\partial f}\right)^2 - \tau_u\frac{\partial^2 \tau_u}{\partial f^2}\right]. \quad [8]$$

From Eq. **7** we see that $d\log k_u/df \geq 0$. The sign of $d^2\log k_u/df^2$ requires establishing the sign of the term in the square brackets in Eq. **8**, which can be done by using the Cauchy–Schwarz inequality. Let us define two functions, $g_1(x, s) \equiv \Theta(s)\sqrt{e^{-\beta fs}H(x, s)}$ and $g_2(x, s) \equiv \Theta(s)\beta s\sqrt{e^{-\beta fs}H(x, s)}$, where $\Theta(s) = 1$ for $s \geq 0$ and 0 for $s < 0$. Then, from Eqs. **6** and **7** we obtain $\tau_u = \int_{-\infty}^\infty ds|g_1(x, s)|^2$, $\frac{\partial \tau_u}{\partial f} = -\int_{-\infty}^\infty ds\, g_1^*(x, s)g_2(x, s)$, $\frac{\partial^2 \tau_u}{\partial f^2} = \int_{-\infty}^\infty ds|g_2(x, s)|^2$. Using the Cauchy–Schwarz inequality

$$\left|\int_{-\infty}^\infty ds\, g_1^*(x, s)g_2(x, s)\right|^2 \leq \int_{-\infty}^\infty ds|g_1(x, s)|^2 \int_{-\infty}^\infty ds|g_2(x, s)|^2 \quad [9]$$

we find that $(\partial \tau_u/\partial f)^2 \leq \tau_u \partial^2 \tau_u/\partial f^2$. Hence, from Eq. **8** we see that $d^2\log k_u/df^2 \leq 0$. In summary, we can now state the full criterion for the validity of WMD for describing force-induced unfolding.

Criteria for WMD Unfolding Landscape. The unfolding rate $k_u(f)$ on a WMD free energy landscape under applied force f must satisfy

$$\frac{d}{df}\log k_u \geq 0, \qquad \frac{d^2}{df^2}\log k_u \leq 0. \qquad [10]$$

If $k_u(f)$ fails to satisfy either of the conditions in Eq. **10**, the underlying free energy landscape must be SMD, and analyses of the measured data using the end-to-end distance, x, as a reaction coordinate are incorrect.

Scenarios for [log $k_u(f)$, f] Plots in WMD and SMD. A range of behaviors in [log $k_u(f)$, f] plots can be obtained depending on the nature of the energy landscape. Stochastic simulations in a WMD (Fig. 1 A and D) show that [log $k_u(f)$, f] has a minor downward curvature, which is readily explained by a generalized Bell model in which the transition-state location is allowed to move toward the folded state in accord with the Hammond effect (21). In contrast, in the SMD (Fig. 1 B, C, E, and F) the [log $k_u(f)$, f] plot shows upward curvature. The upward curvature in Fig. 1E indicates loss of flux from the folded state through two channels in Fig. 1B, similar to parallel pathways in protein unfolding experiments. Interestingly, the upward curvature in Fig. 1F from the SMD landscape in Fig. 1C does not come from parallel pathways. Instead, the lifetime of the folded state first increases followed by the usual decrease as f increases. Such a counterintuitive "catch bond" behavior is well documented in a number of protein complexes (22–24). The results

in Fig. 1 E and F show that violations of Eq. **10** implies that the underlying energy landscape must be SMD.

Naive Analyses of the f Dependence of $k_u(f)$. Using the data generated by molecular dynamics simulations of force unfolding the src SH3 domain, with force applied to residues 9 and 59, for a set of forces f_i, we calculated $k_u(f_i)$ for each force f_i as the inverse of the MFPT from the folded state to the unfolded state, by averaging the set of first passage times to unfolding t_{ij} over the trajectory index j (*Materials and Methods*). The [log $k_u(f)$, f] plot for f-induced unfolding is nonlinear with upward curvature implying that the free energy landscape is SMD (Fig. 2B). We note parenthetically that the inadequacy of the Bell model cannot be fixed using movement of the transition state with f or using a 1D free energy profile with two (or more) barriers. Remarkably, the slope change in the simulations qualitatively coincides with measurements on the same protein (16, 17), where constant force was applied to the residues 7 and 59. Thus, both simulations and experiment show that the condition in Eq. **10** is violated, implying that the free energy landscape for SH3 is SMD.

The observed $k_u(f)$ dependence can be fit using a sum of two exponential functions (16),

$$k_u(f) = k_L^0 \exp\frac{fx_L}{kT} + k_H^0 \exp\frac{fx_H}{kT}. \qquad [11]$$

The parameters k_L^0 and k_H^0 (unfolding rates at $f = 0$) and x_L and x_H (putative locations of the transition states) can be precisely
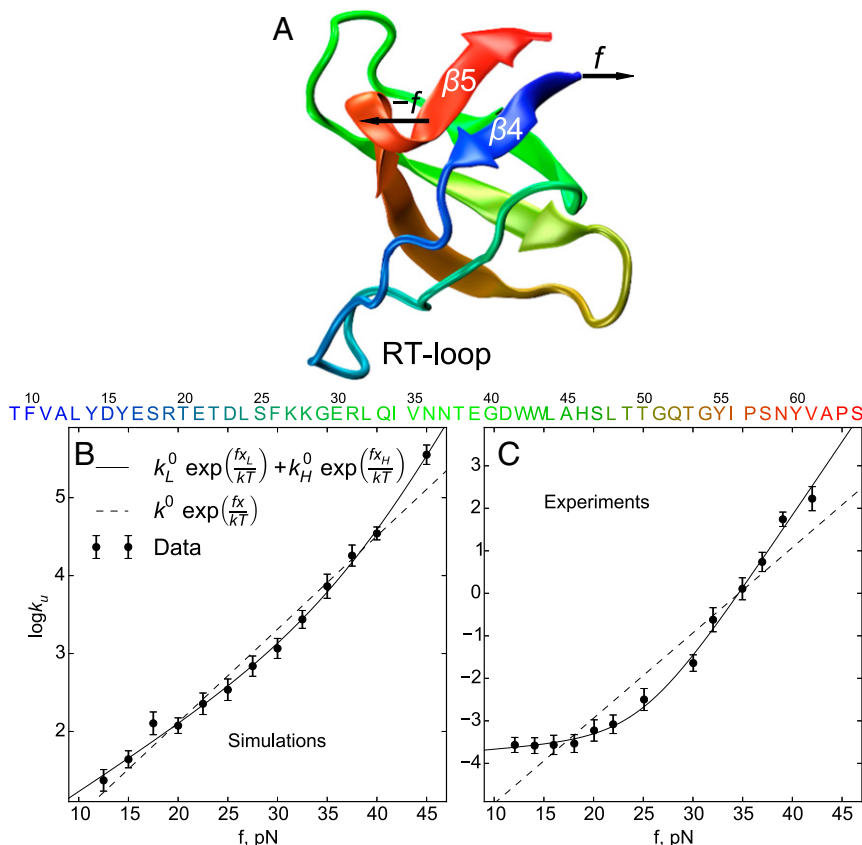
Fig. 2. (*A*) The native structure of *G. gallus* src SH3 domain from Tyr kinase, whose sequence is given below. The N-terminal and C-terminal β-strands are denoted as β4 and β5. In the simulations, the force is applied to residues 9 and 59. RT loop is the longest loop of the domain, which is positioned in sequence right after the N-terminal β-strand. (*B*) The dependence of unfolding rate $k_u(f)$ obtained from the simulations using the SOP-SC model as a function of f given in terms of [log $k_u(f)$, f] plot. The solid line is a two-exponential fit (Eq. **11**) whereas the dashed line is a fit using Bell model. (*C*) Same as *B* except the data are obtained from single-molecule pulling experiments (16).

Table 1. Comparison of fitting parameter sets for two different models for the rates obtained in our simulations and rates from the SMFS experiment (16)

Name	Simulations			Experiment		
	Value	Error	Units	Value	Error	Units
k_L^0	1.5	0.4	Seconds^{-1}	2.0	1.1	10^{-2} s^{-1}
k_H^0	3.9	10.0	10^{-3} s^{-1}	6.1	6.4	10^{-6} s^{-1}
x_L	0.40	0.1	Nanometers	0.08	0.1	Nanometers
x_H	1.16	0.3	Nanometers	1.42	0.12	Nanometers
		$P_2/P_1 \sim 10^3$			$P_2/P_1 \sim 10^{17}$	

The definition of P_2/P_1 is given in *Materials and Methods*. The errors come out of log-likelihood covariance matrix with respect to parameters.

extracted using maximum likelihood estimation (MLE) (*Materials and Methods*). According to the Akaike information criterion (25), the double-exponential model is significantly more probable than the single-exponential model, for both simulations (relative likelihood of the models $P_2/P_1 \sim 10^3$) and experiments (17) ($P_2/P_1 \sim 10^{17}$). The extracted values of x_L and x_H, shown in Table 1, are $x_L = 0.40$ nm, $x_H = 1.16$ nm for the simulations data and $x_L = 0.1$ nm, $x_H = 1.44$ nm for the experimental data from ref. 17 with the MLE procedure, which differ somewhat from the values reported in ref. 17 ($x_L = 0.18$ nm, $x_H = 1.52$ nm). Given that the error in x_L estimated for experimental data using MLE is large we surmise that the simulations and experiments are in good agreement. The switch in the forced unfolding behavior

[estimated as the point where the third derivative of $\log k_u(f)$ in Eq. **11** with parameters given by MLE changes sign] occurs around 25 pN for the experimental data and around 35 pN for the simulation data. These comparisons show that the simulations based on the self-organized polymer with side-chains (SOP-SC) model reproduce quantitatively the shape of the [$\log k_u(f)$] plot. Because simulations are done by coarse-graining the degrees of freedom, involving both solvent and proteins, the $k_u(f)$ from simulations are expected to be larger than the measured values, with the discrepancy being greater at higher forces. Our previous work (26) showed that the unfolding rate in denaturants is larger by a factor of ≈ 150, which is similar to the difference between experiment and simulations in Fig. 2. However, because the inference about parallel pathways relies solely on the shape of $\log k_u(f)$ the inability to quantitatively reproduce the precise value of $k_u(f)$ is irrelevant.

Despite the good fits to Eq. **11** neither x_L nor x_H can be associated with transition-state location as is traditionally assumed. We show below that such projections onto a 1D coordinate cease to have physical meaning when the underlying folding landscape is SMD. The apparent barriers to unfolding at $f = 0$ along the pathways can be estimated using $k_L^0 \approx k_{u0} \exp(-\Delta G_L^\ddagger/k_BT)$ and $k_H^0 \approx k_{u0} \exp(-\Delta G_H^\ddagger/k_BT)$. Using the accepted estimates for the prefactor ($k_{u0} \approx 10^6 - 10^7 \text{s}^{-1}$) (27–29), and the values of k_L^0 and k_H^0 from the fits of experimental data (Table 1), we obtain $\Delta G_L^\ddagger \approx (18 - 20)k_BT$, depending on the value of k_{u0} and $\Delta G_H^\ddagger \approx (26 - 28)k_BT$. If these values are reasonable then the ratio of fluxes through the two pathways at $f = 0$ is $\sim \exp(-(\Delta G_H^\ddagger - \Delta G_L^\ddagger)/k_BT) \sim e^{-8} = 3 \cdot 10^{-4}$, which is much smaller than those obtained in the simulations by
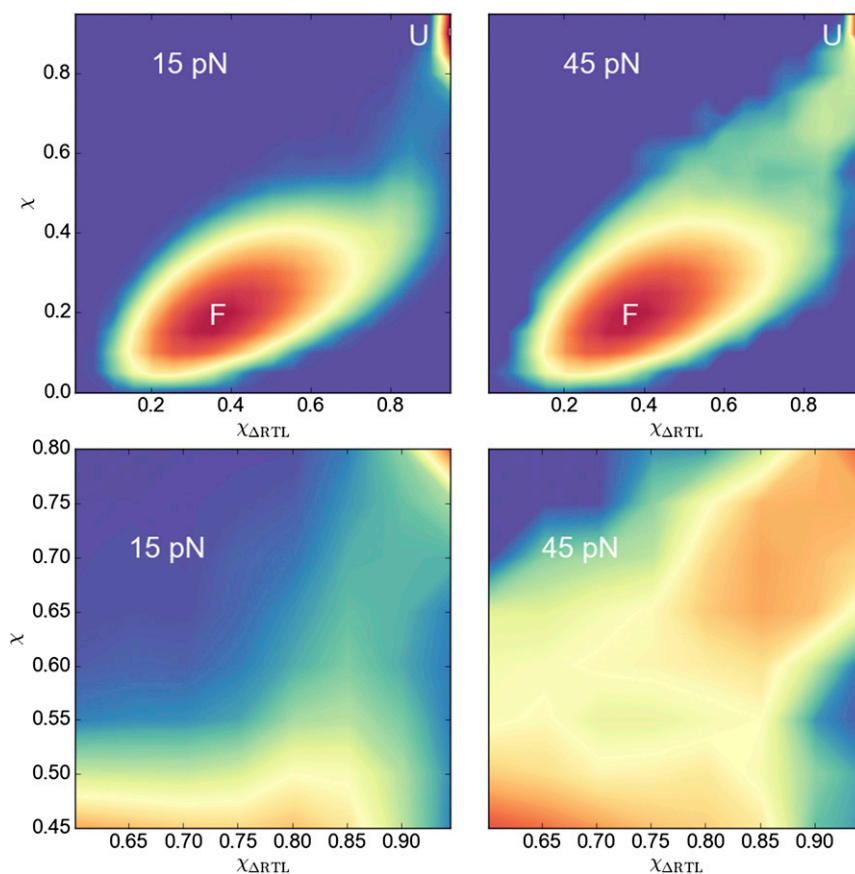
Fig. 3. Histograms of unfolding trajectories at 15 and 45 pN show the difference between unfolding pathways. The overlap parameter χ is the order parameter describing the global unfolding, and χ_{RTL} shows ruptured interactions between the RT loop and the core. The lower panels are blowups of the route to the unfolded (U) state, whereas the upper panels show the profiles connecting F and U states (color scheme is also adapted to higher resolution). The forces are explicitly marked. The rupture process of the RT loop is dramatically different at 15 pN and 45 pN, demonstrating a switch in the unfolding pathway.
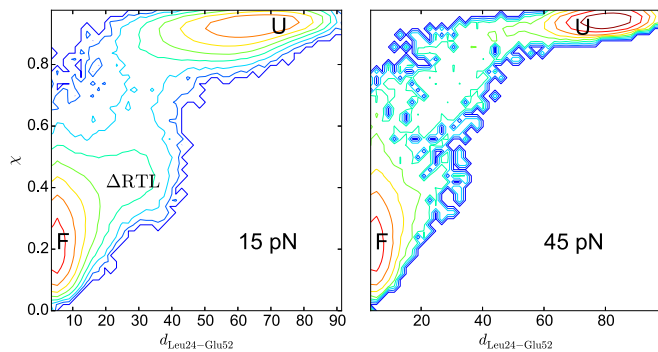
Fig. 4. Histograms of unfolding trajectories at 15 and 45 pN using a 2D map with the global order parameter (χ) and the distance between Leu24 and Glu52 side chains, which shows whether the RT loop interacts with the protein core (Fig. 2A). The profile offers an alternative view of the one in terms of χ and χ_{RTL} (Fig. 3).

direct calculation of the flux through the two pathways. In addition, the finding that $x_H > x_L$ also makes no physical sense, because we expect the molecule under higher tension to be more brittle (19). These are the first indications that the fits using Eq. **11** do not provide meaningful parameters.

Structural Basis of f-Dependent Switch in Pathways. To provide a structural interpretation of the SMD nature of f-induced unfolding of src SH3, we followed the changes in several variables describing the conformations of SH3 as force applied to residues 9 and 59 is varied in the SOP-SC simulations. Most of these are derived from measures assessing the extent to which structures of various parts of the protein overlap with the conformation in the native state. The structural overlap χ_{AB} for two parts of the protein A and B is the fraction of broken native contacts between A and B (30),

$$\chi_{AB}(\{\boldsymbol{r}\}) = \frac{1}{M_{AB}} \sum_{\substack{i \in A \\ j \in B}} \Theta\left(\left| |\boldsymbol{r}_i - \boldsymbol{r}_j| - |\boldsymbol{r}_i^0 - \boldsymbol{r}_j^0| \right| - \Delta\right), \qquad [12]$$

where the summation is over the coarse-grained beads belonging to the parts A and B, M_{AB} is the number of contacts between A and B in the native state, $\Theta(x)$ is the Heaviside function, $\Delta = 2\text{Å}$ is the tolerance in the definition of a contact, and $\boldsymbol{r}_{i,j}$ and $\boldsymbol{r}_{i,j}^0$, respectively, are the coordinates of the beads in a given conformation $\{\boldsymbol{r}\}$ and the native state. Two of the most relevant sets of contacts in the forced rupture of SH3 are the ones between the N-terminal ($\beta4$) and C-terminal ($\beta5$) β-strands (Fig. 2A), computed using the structural overlap, $\chi_{\beta4\beta5}$, and contacts between the RT loop (residues 15–31) and the protein core (strands $\beta1$ and $\beta2$, residues 42–57) quantified by χ_{RTL}. When these structural elements unravel the structural overlap values become close to unity, signaling the global unfolding of the SH3 domain.

Depending on f, in some trajectories the RT loop ruptures from the protein first (χ_{RTL} sharply approaches 1), followed by the break between $\beta4$ and $\beta5$ strands ($\chi_{\beta4\beta5} \approx 1$). In other trajectories, the order is opposite, with $\beta4\beta5$ sheet melting first, without the RT-loop rupture (Figs. 3 and 4 and Fig. S1). The calculated the fraction, $P_{\Delta RTL}$, of trajectories that unravel through rupture of the RT-loop pathway depends strongly on force, suggesting that these are the two major pathways responsible for the change in the slope of the [$\log k_u(f), f$] plot (Fig. 5C). At low forces (15 pN) $P_{\Delta RTL} \approx 0.8$, implying that $\approx 80\%$ of the trajectories unfold through the RT-loop pathway, and this fraction decreases monotonically to $P_{\Delta RTL} \approx 15\%$ at 45 pN. Movies S1 and S2 illustrate the two unfolding scenarios.

Effect of Cysteine Cross-Linking. To further illustrate that the slope change in Fig. 2 is due to the switching of the unfolding routes between the particular pathways discussed above, we created an in silico mutant by adding a disulfide bond between the RT loop and $\beta2$ (mimicking a potential experiment with L24C/G54C mutant). In the cross-link mutant, the enhanced stability of the RT loop to the protein core blocks the ΔRTL unfolding pathway. We generated six 1,500-ms unfolding trajectories at 15 pN and did not observe unfolding in any of them, thus obtaining an estimate for the upper bound of unfolding rate of ≈ 0.6 s^{-1} for this mutant. Comparing this unfolding rate to the rate at 15 pN for the WT (without the disulfide bridge) of 5.2 s^{-1} shows that blocking of the pathway decreases the average unfolding rate at 15 pN. The mutant simulations with the disulfide bridge suggest that the RTL pathway plays a major role at low forces, and the unfolding through the $\beta4\beta5$ pathways is much slower at low force. Furthermore, these simulations also show that rupture of the protein through the $\beta4\beta5$ pathway occurs at a very slow rate at low forces even when the unfolding flux along the RTL pathway is muted. Taken together these simulations explain the structural basis of rupture in the two major unfolding pathways.

Pathway Switch Occurs at a Lower Force in V61A Mutant. To examine the effect of point mutations, we calculated $k_u(f)$ as a function of f for the V61A mutant. In the laser optical trapping (LOT) experiments, V61A mutant does not show upward curvature in the same force range, and the [$\log k_u(f), f$] plot in that range is linear. However, the curvature can be seen at lower forces. In simulations, we observe the same qualitative change with respect to the WT upon mutation (Fig. 6 and Table S1). If only data for forces above 15 pN are taken into account, the single-exponential model becomes slightly more likely than the double-exponential model, but inclusion of the lower forces data shows double exponential, with pathway switching coming at a lower force than for WT. The fraction of trajectories going through the RT-loop pathway
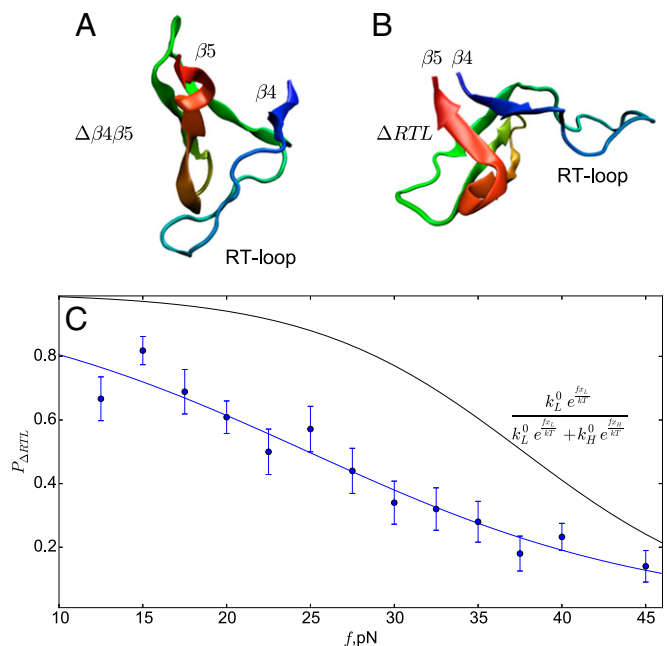
Fig. 5. (A) A snapshot of SH3 with the intact contacts between the RT loop and the core but with ruptured $\beta4$ and $\beta5$ strands (state $\Delta\beta4\beta5$). (B) A snapshot with the RT loop peeled off, but $\beta4\beta5$ intact (state ΔRTL). (C) The flux of molecules through the ΔRTL state as a function of force. The blue symbols are obtained from simulations. The black line, which is the prediction using the double-exponential fit of the [$\log k_u(f), f$] plot with Eq. **11**, differs from direct calculations based on simulations.
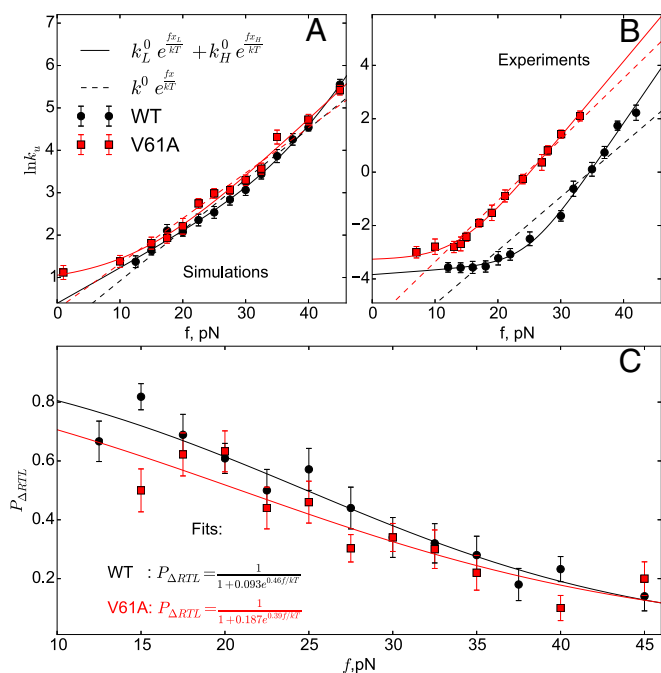
Fig. 6. (*A*) Unfolding rate for the V61A mutant obtained using simulations shows the single-exponential model to be more probable than the double-exponential fit. (*B*) Same as *A* except the data are from experiments. (*C*) Fraction of trajectories that unfold through the low-force pathway as a function of *f* obtained directly from simulations. Lines are the fits to the function $(1 + A \exp(B/kT))^{-1}$ with parameter values as indicated.

decreases compared with the WT [i.e., $P_{\Delta RTL}^{V61A}(f) < P_{\Delta RTL}^{WT}(f)$ for all *f*] (Fig. 6*C*). The loss of upward curvature in the force range above 15 pN can be explained by the more prominent role of the $\beta 4 \beta 5$ pathway at low forces, leading to lesser degree of switching between the pathways. The V61A mutation is in the $\beta 5$ strand, making interactions between $\beta 4$ and $\beta 5$ weaker, thus enabling the sheet to rupture more readily. Parenthetically we note that this is a remarkable result, considering that change in the SOP-SC force field is only minimal, which further illustrates that our model also captures the effect of point mutations.

Free Energy Profiles and Transition States. Let us assume that the free energy landscape projected onto extension as the reaction coordinate accurately captures the *f*-dependent unfolding kinetics. In this case, we expect the Bell model or its variation would hold, and *x* (assumed to be *f*-independent) obtained from the fitting to that model would be the distance to the transition state with respect to the folded state $x^{\ddagger}$. If the underlying free energy landscape were SMD it is still possible to formally construct a 1D free energy profile using experimental (18) or simulation data. It is tempting to associate the distances in the projected 1D profiles with transition-state locations with respect to the folded state, as is customarily done in analyzing single-molecule force spectroscopy (SMFS) data. Such an interpretation suggests that x_L and x_H should correspond to the distances to the two transition states in the two pathways, with $x^{\ddagger}$ increasing with force in an apparent anti-Hammond behavior. To assess whether this is realized, we constructed 1D free energy profiles (of the WT protein) at forces 15, 30, and 45 pN to determine $x^{\ddagger}$. It turns out that $x^{\ddagger}$ decreases rather than increases with force, demonstrating the normally expected Hammond behavior (Fig. 7), as force destabilizes the native state (21, 31) (*Discussion*).

We now demonstrate that x_L and x_H cannot be identified with transition-state locations by calculating the committor probability, P_{fold} (32), the fraction of trajectories that reach the folded state

before the unfolded state starting from x_L or x_H. If x_L and x_H truly correspond to distances to transition states then $P_{fold} \approx 0.5$ (32), that is, the TSE should correspond to structures that have equal probability of reaching folded or unfolded state, starting from x_L or x_H. In sharp contrast to this expectation, the states with x_L are visited hundreds of time before unfolding (Fig. S2), which means $P_{fold}(x_L) \approx 1$. Thus, the usual interpretation of x_L or x_H ceases to have physical meaning, which is a consequence of the strong multidimensionality of the unfolding landscape of SH3.

Force-Dependent Movement of the TSE. The results in Fig. S2 show that the extracted values of x_L and x_H cannot represent the TSE. Because the underlying reaction coordinates for the inherently SMD nature of folding landscapes are difficult to guess, the TSE can only be ascertained with a method that does not use a predetermined form of the reaction coordinate. We use the P_{fold}, based on the theory that the TSE should correspond to structures that have equal probability ($P_{fold} \approx 0.5$) of reaching the folded or unfolded state. To determine the TSEs in our simulations, we picked the putative transition-state structures from the saddle point of the 2D (χ, E) histogram of the unfolding trajectories [$0.66 < \chi < 0.73$; $92 < E < 105$ (kcal/mol) for $f = 15$ pN and $0.59 < \chi < 0.73$; $82 < E < 106$ (kcal/mol) for $f = 45$ pN], where E is the total energy of the protein. We ran multiple trajectories from each of the candidate TS structures, noting when the trajectory reaches the folded or the unfolded state first, to determine the P_{fold}. The set of structures with $0.4 < P_{fold} < 0.6$ is identified with the TSE. The P_{fold} value for the whole ensemble is the total number of trajectories (starting from all of the candidate structures) that reach the folded state first, divided by the total number of trajectories (or, the average of the individual P_{fold} values).

The TSEs for 15 pN and 45 pN are given in Fig. 8. For both sets the $P_{fold} \approx 0.5 \pm 0.07$. The low-force TSE shows that the RT loop is disconnected from the core (ΔRTL state) and the 45 pN TSE has structures where the loop interacts with the core, but the contacts between N- and C-terminal β-strands are broken. The explicit TSE calculations confirm that the TSEs are similar to those found in unfolding trajectories with χ_{RTL} and $\chi_{\beta 4 \beta 5}$.

The experimental analysis of transition states of SH3 using mechanical Φ-values (17) suggests that in the high-force pathway
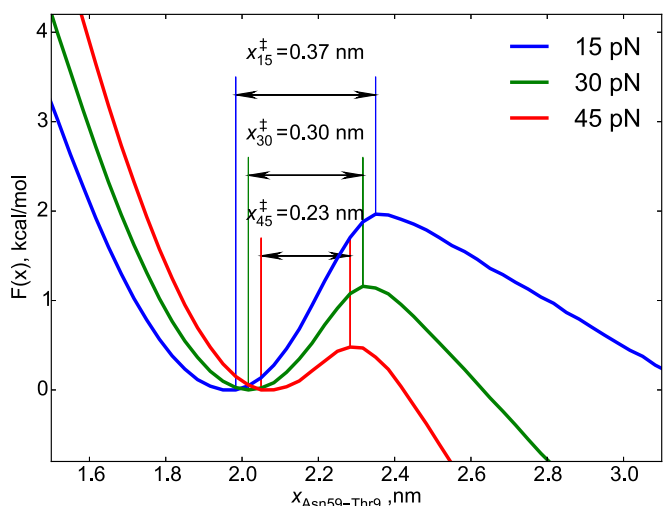
Fig. 7. One-dimensional free energy profiles, $F(x)$, as a function of molecular extension, at three forces, calculated from simulations. The $F(x)$ profiles show that the "distance to the transition state" $x^{\ddagger}$ (the location of the top of the barrier with respect to the folded state) decreases as *f* increases. Fits of the unfolding rate to the double-exponential function produce the exact opposite behavior ($x_H > x_L$), which is an indication that the x_L and x_H extracted from fitting the data to Eq. **11** do not coincide with $x^{\ddagger}$.
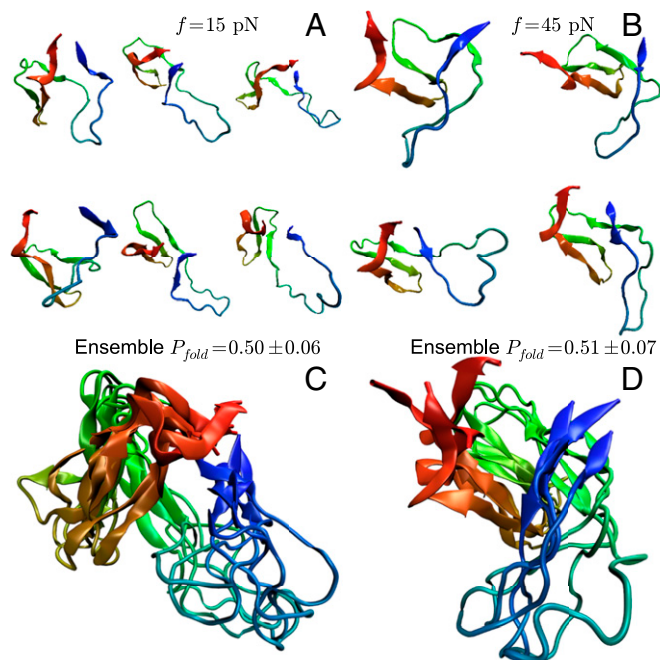
Fig. 8. TSE structures at $f = 15$ pN (*A*) and $f = 45$ pN (*B*). (*C*) Overlay of the TSE at 15 pN and 45 pN (*D*). The TSEs were obtained using P_{fold}, which is independent of the reaction coordinate.

the important residues are Phe-10 and Val-61 (which are in the $\beta4$ and $\beta5$), along with a core residue Leu-44. For the bulk (low/zero force) pathway, Phe-10, Ile-56, and Val-61 are also apparently important in TSE, as is the RT-loop residue Leu-24, which interacts with the protein core. Our simulation results, which provide a complete structural description of the TSEs, support the experimental interpretation, namely, loss of interaction between the RT loop and the core at low forces and rupture of the $\beta4\beta5$ sheet at high forces.

It is interesting to compute the mean extensions of the two major TSEs. The average distance between force application points for these structures is $x_{15}^{TSE} = 2.25 \pm 0.2$ nm for 15 pN and $x_{45}^{TSE} = 2.7 \pm 0.3$ nm for 45 pN, which (given the distances in the folded state of $x_{15}^{F} = 1.96$ and $x_{45}^{F} = 2.05$ nm) translates to the transition states of $x_{15}^{\ddagger} = 0.29 \pm 0.18$ and $x_{45}^{\ddagger} = 0.65 \pm 0.33$ nm, respectively. These values have no relation to x_L and x_H, further underscoring the inadequacy of using Eq. **11** to interpret $[\log k_u(f), f]$ plots in SMD.

Discussion

Hammond Behavior. Protein folding could be viewed using a chemical reaction framework. Just like in a chemical reaction, transitions occur from a minimum on a free energy landscape (corresponding to reactant or unfolded state) to another minimum (corresponding to a product representing the folded state, or an intermediate) by crossing a free energy barrier. The top of the free energy barrier corresponds to a transition state.

Besides determining the structures of the unfolded and folded states, one of the main goals in protein folding is to identify the TSE and characterize the extent of its heterogeneity. When viewed within the chemical reaction framework, the Hammond postulate provides a qualitative description of the structure of the transition state if it is unique. The Hammond postulate states that "if two states, as, for example, a transition state and an unstable intermediate, occur consecutively during a reaction process and have nearly the same energy content, their interconversion will involve only a small reorganization of the molecular structure" (33). A corollary of the Hammond postulate is that the TS

structure likely resembles the least stable species in the folding reaction.

To apply the Hammond postulate to a protein free energy landscape, perturbed by f, let us assume that at $f = f_m$ the states F and U, with equal free energy, are separated by a transition state. Increasing f will generally destabilize F, and lower the free energy of U. According to Hammond's postulate, the transition state should be more similar to F than U as f increases. If $f < f_m$, then the free energy of F will be lower than U, and consequently the transition state will be more U-like. As a consequence of this argument, in unfolding induced by force, the transition state should move toward the state that is destabilized by f (31), in accord with Hammond behavior. If the opposite were to happen it could be an indication of anti-Hammond behavior.

In a 1D energy landscape, the distance between a minimum and a barrier is reflected in the slope of the $[\log k(f), f]$ plot (m value for $[\log k([C]), [C]]$), which follows from the Arrhenius law and linear coupling in the free energy. Hammond behavior for the unfolding rate would mean movement of the transition state toward the folded state resulting in the decreasing of the slope of the $[\log k_u(f), f]$ plot with f. Hence, the temptation to refer to the opposite change of slope (i.e., increasing with f) as anti-Hammond behavior is natural. However, because the increase of the slope of the $[\log k_u(f), f]$ plot necessarily means that the energy landscape is SMD, referring to movement of the transition state along a single reaction coordinate is not meaningful. Hence, the term "anti-Hammond" behavior in this case does not reflect the opposite of the Hammond postulate in either the original formulation or movement of the transition state. Moreover, even if the energy landscape is formally projected onto the reaction coordinate to which the parameter (f or $[C]$) is coupled (which is possible even in the SMD case albeit without much physical sense), the movement of the transition state on this formal 1D landscape will still obey the Hammond postulate. Such a conclusion follows from a Taylor expansion of the first derivative of the perturbed (by f or $[C]$) free-energy profile $F_f(x) = F_0(x) - fx$ around the barrier top ($x^{\ddagger}$),

$$
\begin{aligned}
F_f'(x)\big|_{x_f^{\ddagger}} &= (F_0(x) - fx)'\big|_{x_f^{\ddagger}} = F_0'\left(x_0^{\ddagger} + \left(x_f^{\ddagger} - x_0^{\ddagger}\right)\right) - f = \\
&= F_0'\left(x_0^{\ddagger}\right) + F_0''\left(x_0^{\ddagger}\right)\left(x_f^{\ddagger} - x_0^{\ddagger}\right) - f = 0,
\end{aligned}
\tag{13}
$$

where $F_0(x)$ and $x_0^{\ddagger}$ are the free energy profile and transition-state position at $f = 0$. Because $F_0'(x_0^{\ddagger}) = 0$ and $F_0''(x_0^{\ddagger}) < 0$, we find $x_f^{\ddagger} - x_0^{\ddagger} = f/F''(x_0^{\ddagger}) < 0$, or $x_f^{\ddagger} < x_0^{\ddagger}$, establishing that the transition state moves toward the native state, in accord with the Hammond behavior. Our conclusion holds for any perturbation f that is linearly coupled to the energy function, and that monotonically destabilizes the folded state. Thus, we surmise that upward curvature in $[\log k_u(f), f]$ or $[\log k_u([C]), [C]]$ plots are not equivalent to anti-Hammond behavior. We note here, though, that the linear coupling of f to the protein Hamiltonian is exact and the perturbation by denaturant is approximate, although the leading order in $[C]$ is linear.

A similar conclusion, that is, a connection between upward curvature and multidimensionality, has been drawn analytically before, in the context of mechanochemistry of small molecules, based on the Taylor expansion of the Bell's model, similar to Eq. **13** (34, 35). In our work, we started from the most general description rather than from the solution of the Kramer's problem. The WMD conditions are similar to 1D assumptions when obtaining the Bell's model, but we do not make any assumptions about the barriers. We also solved directly for the quantity we are interested in, that is, sign of $\frac{d^2}{df^2} \log k_u(f)$, rather than movements of the transition state. Connecting the latter to the curvature of the rate requires some additional steps, which might require more assumptions.

Zhuravlev et al.

SMD in Denaturant-Induced Unfolding. The criterion in Eq. **10** to assess whether experiments can be analyzed using a 1D free energy profile applies to any external perturbation with a linear, additive contribution to the free energy. If we consider the unfolding rate $k_u([C])$ as a function of denaturant concentration $[C]$, a criterion analogous to Eq. **10** would hold if we assume that the energetic contribution due to $[C]$ is linear, proportional to a reaction coordinate related to the solvent-exposed surface area:

$$U(x, \mathbf{r}, [C]) = U_0(x, \mathbf{r}) + S(x)[C], \qquad [14]$$

where $S(x)$ is the solvent-accessible surface area (SASA)-related monotone function of reaction coordinate x. Thus, for any perturbation (f or $[C]$) coupling to the Hamiltonian, the theory and applications also hold when upward curvature in the $[\log k_u([C]), [C]]$ plot is observed.

Typically, the observed nonlinearities in the $[\log k_u([C]), [C]]$ plots are analyzed using a double exponential fit, $k_u([C]) = k_L^0 \exp(m_L[C]) + k_H^0 \exp(m_H[C])$ (8), just like is done to analyze $[\log k_u(f), f]$ plots. Here, m_L and m_H are the analogs of x_L and x_H representing the unfolding m values. It has been shown that for a protein with Ig fold (8) (see Fig. S3 for the fits for several mutants of I27 using the double-exponential model) and for monellin (9) that there is upward curvature in the $[\log k_u([C]), [C]]$ plots, which violates Eq. **10**, implying that the underlying landscape in SMD. If $[\log k_u([C]), [C]]$ plots were linear then the unfolding m value is likely to be proportional to the solvent accessible surface area in the transition state (even if the latter is heterogeneous), and the ensemble of conformations corresponding to the m value may be associated with the TSE ($P_{fold} \approx 0.5$). However, for the $[\log k_u([C]), [C]]$ plots with upward curvature m_L and m_H may not correspond to the SASAs of the respective transition states of the pathways, just as we have shown that the extracted x_L and x_H should not be interpreted as TSE locations at low and high f, respectively. In addition, although the $k_L^{H_2O}$ for the WT is consistent with the expected value for β-sheet proteins with the I27 size, the $k_H^{H_2O} \approx 10^{-4} k_L^0$ seems unphysical. This observation combined with fairly high ratios of m_H/m_L from a double-exponential fit (8) (Table S2 and Fig. S3) suggests that although the double-exponential model above fits the data, inferring the nature of the TSE requires an entirely new set of experiments along the lines reported by Clarke and co-workers (8).

Pathway Switch and Propensity to Aggregate. In our previous work (36) we showed that an excited state N^* in the spectrum of monomeric src SH3 domain has a propensity to aggregate. The structure of the N^*, which is remarkably close to the very lowly populated structure for *Fyn* SH3 domain determined using using NMR (37), has a ruptured interaction between $\beta4$ and $\beta5$. In other words, the value of $\chi_{\beta4\beta5}$ is large. Interestingly, in our simulations unfolding of src SH3 domain occurs by weakening of these interactions at high forces (Fig. 3 and Fig. S1). Thus, the N^* structures are dominant at high forces. Because the probability of populating of N^* is low at low forces [N^* has 1–2% probability of forming at $f = 0$ (36, 37)] it follows that SH3 aggregation is unlikely at low forces but can be promoted at high forces. Thus, SH3 domains have evolved to be aggregation-resistant, and only under unusual external conditions they can form fibrils.

Prediction for a Switch in the Force Unfolding of I27. Based on our theory and simulations we can make a testable prediction for forced unfolding of I27. Because there is upward curvature in the denaturant-induced unfolding of I27 (8), we predict that a similar behavior should be observed for force-induced unfolding as well. In other words, there should be a switch in the pathway as the force used to unfold I27 is changed from a low to a high value. It is likely that this prediction has not been investigated because mechanical unfolding of I27 has so far been probed using only atom force microscopy (38), where high forces are used. It would be most welcome to study the unfolding behavior of I27 using LOT experiments to test our prediction.

Conclusions

We have proven that upward curvature in the unfolding rates as a function of a perturbation, which is linearly coupled to the energy function describing a protein in a solvent, implies that the underlying energy landscape is strongly multidimensional. The observation of upward curvature in the $[\log k_u(f), f]$ plots also implies that unfolding occurs by multiple pathways. In the case of f-induced unfolding of SH3 domain this implies that there is a continuous decrease in the flux of molecules that reach the unfolded state through the low-force pathway as f increases. The numerical results using model 2D free energy profiles allow us to conjecture that if a protein folds by parallel routes then the unfolding rate as a function of the linear perturbation must exhibit upward curvature. Only downward curvature in the $[\log k_u(f), f]$ plots can be interpreted using a single-barrier 1D free energy profile with a moving transition state or one with two sequential barriers (39, 40).

Our study leads to experimental proposals. For example, Förster resonance energy transfer (FRET) experiments especially when combined with force would be most welcome to measure the flux through the two paths identified for src SH3 domain. Our simulations suggest that the FRET labels between the RT loop and the protein core should capture the pathway switch, provided there is sufficient temporal resolution to observe the state with the RT loop unfolded. A more direct way is to block the RT-loop pathway with a disulfide bridge between the RT loop and the core, as we demonstrated using simulations, and assess if the unfolding rate decreases dramatically. Our work shows that the richness of data obtained in pulling experiments can only be fully explained by integrating theory and computations done under conditions that are used in these experiments.

Materials and Methods

Force-Dependent Rates for SH3 Domain Using Molecular Simulations. The 56 residue *Gallus gallus* src SH3 domain from Tyr kinase consists of five β strands (PDB ID code 1SRL), which form β-sheets comprising the tertiary structure of the protein (Fig. 2A). Residues are numbered from 9 to 64. The details of the SOP-SC model are described elsewhere (36). A constant force is applied to the N-terminal end (residue 9) and residue 59 (Fig. 2A). We used Langevin dynamics, in the limit of high friction, to compute the f-dependent unfolding rates. We covered a range of forces from 12.5 pN to 45 pN generating between 50–100 trajectories at each force. From these unfolding trajectories, we calculated the first time the protein unfolded ($x_{ee} > 5$ nm), thus obtaining a set of times t_{ij}, for trajectory j at force f_i. We used umbrella sampling with weighted histogram analysis method (41) and low-friction Langevin dynamics (42) to calculate free energy profiles.

Maximum Likelihood Estimation. For the set of M constant forces f_i, with N_i measurements of the unfolding time at each force, assuming exponential distribution of unfolding times $P(t) = ke^{kt}$, where the rate k depends on the force, the log-likelihood function is

$$L = \log \prod_i^M \prod_j^{N_i} k(f_i) e^{-k(f_i)t_{ij}} = \sum_i^M \sum_j^{N_i} \log k(f_i) - k(f_i)t_{ij}. \qquad [15]$$

In the above equation t_{ij} is the unfolding time measured in the j-th trajectory at force f_i. The exponential distribution allows us to take the sum over j and use the average unfolding time $\tau_i = \sum_j t_{ij}/N_i$ for each force,

$$L = \sum_i N_i(\log k(f_i) - k(f_i)\tau_i). \qquad [16]$$

For each of the models (single- and double-exponential) the log-likelihood function L was numerically maximized with the set of data $\{f_i, N_i; \tau_i\}$ (from simulations or experiment). The two maximal values of L (for each model) were plugged into the Akaike information criterion (25) to calculate relative likelihood of the models, that is, the ratio of probabilities that the data are

described by each of the models. The parameters that maximize L are used for fitting the [$\log k_u(f), f$] plots.

Akaike Information Criterion. The lower value of $AIC = 2n - 2L$ (where L is the log-likelihood function and n is the number of parameters in the model) indicates a more probable model, with the relative likelihood of models with AIC_1 and AIC_2 given by $P_2/P_1 = \exp((AIC_1 - AIC_2)/2)$. Thus, for the comparison of Bell's model ($L = L_1; n = 2$) and double-exponential model ($L = L_2; n = 4$), the double exponential is more probable by a factor of $P_2/P_1 = \exp(L_2 - L_1 - 2)$, where L_1 and L_2 are found by maximizing L in Eq. **16**.

1. Harrison SC, Durbin R (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA* 82(12):4028–4030.
2. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267(5204):1619–1620.
3. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21(3):167–195.
4. Guo Z, Thirumalai D, Honeycutt JD (1992) Folding kinetics of proteins: A model study. *J Chem Phys* 97:525–535.
5. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA* 89(18):8721–8725.
6. Klimov DK, Thirumalai D (2005) Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways. *J Mol Biol* 353(5):1171–1186.
7. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA* 106(45):19011–19016.
8. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat Struct Biol* 10(8):658–662.
9. Aghera N, Udgaonkar JB (2012) Kinetic studies of the folding of heterodimeric monellin: Evidence for switching between alternative parallel pathways. *J Mol Biol* 420(3):235–250.
10. Sosnick TR, Barrick D (2011) The folding of single domain proteins–Have we reached a consensus? *Curr Opin Struct Biol* 21(1):12–24.
11. Lindberg MO, Oliveberg M (2007) Malleability of protein folding pathways: A simple reason for complex behaviour. *Curr Opin Struct Biol* 17(1):21–29.
12. Aksel T, Barrick D (2014) Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophys J* 107(1):220–232.
13. Mickler M, et al. (2007) Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc Natl Acad Sci USA* 104(51):20268–20273.
14. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512–516.
15. Kotamarthi HC, Sharma R, Narayan S, Ray S, Ainavarapu SRK (2013) Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS). *J Am Chem Soc* 135(39):14768–14774.
16. Jagannathan B, Elms PJ, Bustamante C, Marqusee S (2012) Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc Natl Acad Sci USA* 109(44):17820–17825.
17. Guinn EJ, Jagannathan B, Marqusee S (2015) Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat Commun* 6:6861.
18. Hinczewski M, Gebhardt JCM, Rief M, Thirumalai D (2013) From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc Natl Acad Sci USA* 110(12):4500–4505.
19. Hyeon C, Thirumalai D (2007) Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *J Phys Condens Matter* 19:113101.
20. van Kampen N (1992) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
21. Hyeon C, Thirumalai D (2006) Forced-unfolding and force-quench refolding of RNA hairpins. *Biophys J* 90(10):3410–3427.
22. Marshall BT, et al. (2003) Direct observation of catch bonds involving cell-adhesion molecules. *Nature* 423(6936):190–193.
23. Buckley CD, et al. (2014) Cell adhesion. The minimal cadherin-catenin complex binds to actin filaments under force. *Science* 346(6209):1254211.
24. Chakrabarti S, Hinczewski M, Thirumalai D (2014) Plasticity of hydrogen bond networks regulates mechanochemistry of cell adhesion complexes. *Proc Natl Acad Sci USA* 111(25):9048–9053.
25. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
26. Liu Z, Reddy G, O'Brien EP, Thirumalai D (2011) Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc Natl Acad Sci USA* 108(19):7787–7792.
27. Li MS, Klimov DK, Thirumalai D (2004) Thermal denaturation and folding rates of single domain proteins: Size matters. *Polymer (Guildf)* 45:573–579.
28. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423(6936):193–197.
29. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. *Curr Opin Struct Biol* 14(1):76–88.
30. Guo Z, Thirumalai D (1995) Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers* 36(1):83–102.
31. Klimov DK, Thirumalai D (1999) Stretching single-domain proteins: Phase diagram and kinetics of force-induced unfolding. *Proc Natl Acad Sci USA* 96(11):6166–6170.
32. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich EI (1998) On the transition coordinate for protein folding. *J Chem Phys* 108:334.
33. Hammond GS (1955) A correlation of reaction rates. *J Am Chem Soc* 77(2):334–338.
34. Konda SSM, Brantley JN, Bielawski CW, Makarov DE (2011) Chemical reactions modulated by mechanical stress: Extended Bell theory. *J Chem Phys* 135(16):164103.
35. Konda SSM, et al. (2013) Molecular catch bonds and the anti-Hammond effect in polymer mechanochemistry. *J Am Chem Soc* 135(34):12722–12729.
36. Zhuravlev PI, Reddy G, Straub JE, Thirumalai D (2014) Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *J Mol Biol* 426(14):2653–2666.
37. Neudecker P, et al. (2012) Structure of an intermediate state in protein folding and aggregation. *Science* 336(6079):362–366.
38. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276(5315):1109–1112.
39. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397(6714):50–53.
40. Hyeon C, Thirumalai D (2012) Multiple barriers in forced rupture of protein complexes. *J Chem Phys* 137(5):055103.
41. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021.
42. Honeycutt JD, Thirumalai D (1992) The nature of folded states of globular proteins. *Biopolymers* 32(6):695–709.

Supporting Information

Zhuravlev et al. 10.1073/pnas.1515730113

### Fits for the Titin I27 Bulk Experiment in Denaturant

We performed maximum likelihood fitting for the data in ref. 8 with single- and double-exponential models and compared the models using Akaike information criterion. The results are given in Table 1. Note the dramatic difference in the prefactors ($k_L^{H_2O}$ and $k_H^{H_2O}$), obtained using a double-exponential fit, which is hard to explain. The difference in m values, if they correspond to the SASA in the transition state, does not seem to be meaningful. These observations suggest, just as in the case for force-induced unfolding, a de facto 1D fit does not yield physically meaningful results.

### [$\log k_u(f), f$] Plots for 2D Landscapes

To better illustrate the connection between the curvature of the [$\log k, f$] plot and existence of parallel pathways, we performed Brownian dynamics simulations of force-dependent rate of escape of a particle from the bound state for the landscapes given in Fig. 1 of the main text. The resulting curves are given in Fig. 1 *D–F*. For each data point, we generated 8,192 trajectories. The Fig. 1*A* landscape is weakly multidimensional, so the [$\log k, f$] plot does not exhibit upward curvature. For the landscape in Fig. 1*B*, two parallel pathways exist, and flux through the states depends on f as in experiments. The resulting curve has upward curvature. A double exponential fit is shown in Fig. 1*D*. The landscape in Fig. 1*C* gives rise to a more complex behavior.
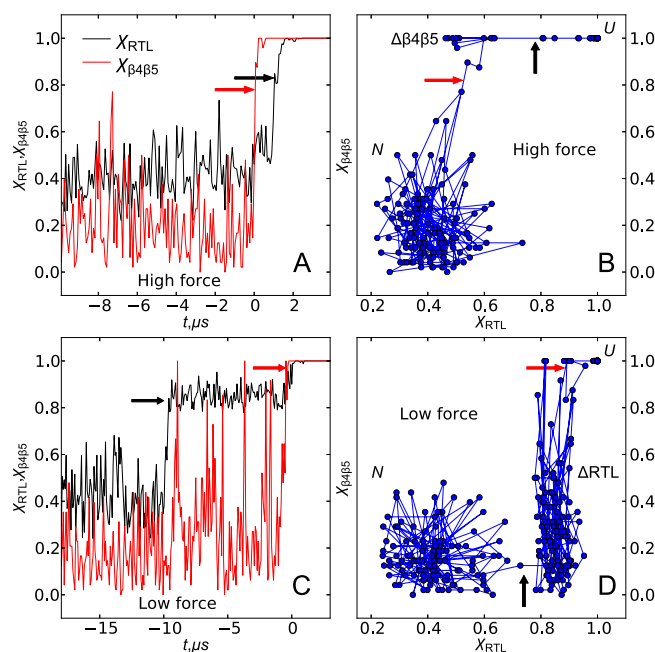
Fig. S1. Typical unfolding trajectories at high (*A* and *B*) and low (*C* and *D*) forces in terms of fractions of ruptured native contacts (χ) calculated for specific parts of the SH3 domain. $\chi_{\beta4\beta5}$ is for contacts between $\beta4$ and $\beta5$, and χ_{RTL} is for contacts between the RT loop and the protein core. At high forces, the $\beta4\beta5$ contacts break first (red arrow in *A* and *B*), followed by the rest of the protein, including RT loop (black arrow in *A* and *B*). There is a transient state $\Delta\beta4\beta5$ where the contacts between N-terminal and C-terminal β strands are broken but the rest of the structure is intact. At low forces, the RT loop is peeled from the core first (black arrow in *C* and *D*), whereas $\beta4$-$\beta5$ contacts are intact. The protein pauses in the ΔRTL state for some length of time before the $\beta4\beta5$ sheet melts (red arrow in *C* and *D*), followed by global unfolding.
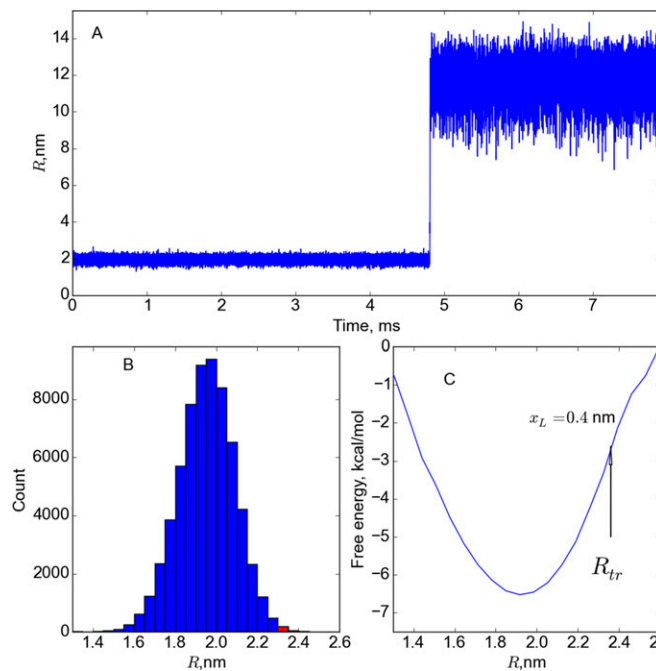
Fig. S2. (*A*) An example of a low-force (15 pN) unfolding trajectory. (*B*) Histogram of the folded state calculated form the trajectory in *A*. (*C*) Free energy profile $F(x) = -k_B T \log P(x)$. If x_L is the distance to the transition state, then the transition state should be located at ≈ 2.34 nm. The states corresponding to 2.34 nm are visited multiple times before the single unfolding event but are always followed by the visit to the folded state. Therefore, P_{fold} calculated from the ensemble of trajectories starting at x_L is near 1. From the histogram in *B*, we find that x_L is visited 231 times before unfolding, which yields $P_{fold} \geq 0.996$.
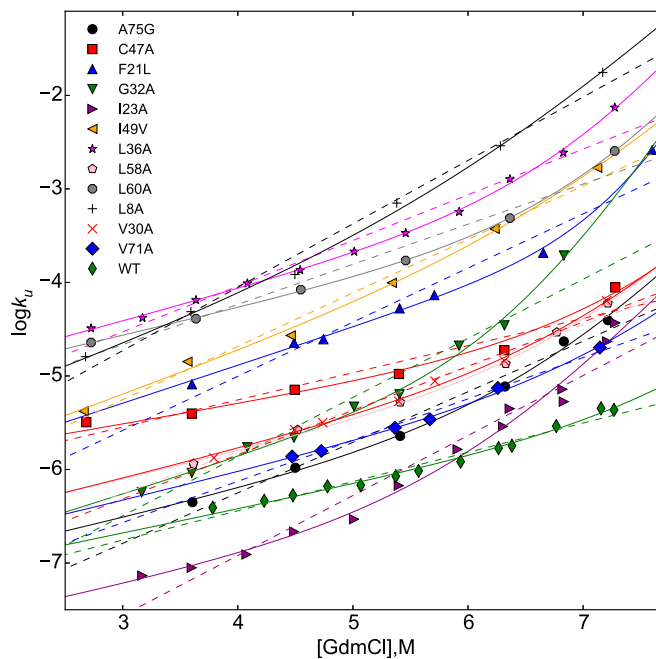
Fig. S3. Unfolding rates of titin I27 domain as a function of guanidinium chloride concentration for the WT and various mutants (8). Dashed lines show single-exponential fits $k_u = k^{H_2O} \exp(m^{\ddagger}[C])$, and solid lines show double exponential fits $k_u = k_L^{H_2O} \exp(m_L[C]) + k_H^{H_2O} \exp(m_H[C])$. The extracted values of the parameters are in Table S2.

Table S1. Fitting parameters for unfolding rates of SH3 domain as a function of force for WT and V61A mutant from simulations and experiment

Name	k_L^0	k_H^0	x_L	x_H	k^0	x	P_2/P_1
Simulation WT	1.47	$3.88 \cdot 10^{-3}$	0.40	1.16	0.76	0.58	10^3
Simulation V61A	2.45	$4.20 \cdot 10^{-1}$	0.02	0.67	1.29	0.52	10^5
Experiment WT	$2.15 \cdot 10^{-2}$	$6.31 \cdot 10^{-6}$	0.07	1.42	$9.59 \cdot 10^{-4}$	0.82	10^{17}
Experiment V61A	$3.83 \cdot 10^{-2}$	$8.46 \cdot 10^{-4}$	0.0	1.15	$3.50 \cdot 10^{-3}$	0.94	10^5
Units	Seconds^{-1}	Seconds^{-1}	Nanometers	Nanometers	Seconds^{-1}	Nanometers	1

Two types of fits were made: exponential fits $k_u = k^0 \exp(fx/k_B T)$ and double-exponential fits $k_u = k_L^0 \exp(fx_L/k_B T) + k_H^0 \exp(fx_H/k_B T)$. P_2/P_1 shows the relative likelihood of the double-exponential model with respect to the single-exponential model, as assessed by Akaike information criterion.

Table S2. Fitting parameters for unfolding rates of titin I27 domain as a function of guanidinium chloride concentration for a set of single-residue mutants and WT (8)

Name	$k_L^{H_2O}$	$k_H^{H_2O}$	m_L	m_H	k^{H_2O}	$m^{\ddagger}$	P_2/P_1
A75G	6.3	4.7	0.28	1.35	2.2	0.54	$\sim 10^{12}$
C47A	21.4	2.9	0.21	1.38	16.3	0.29	$\sim 10^{19}$
F21L	14.8	0.005	0.41	2.40	6.6	0.58	$\sim 10^{40}$
G32A	6.0	0.3	0.38	1.91	2.3	0.63	$\sim 10^{42}$
I23A	3.2	4.3	0.26	1.36	0.8	0.64	$\sim 10^{67}$
I49V	14.4	26.1	0.44	1.31	9.4	0.57	$\sim 10^8$
L36A	44.4	6.0	0.33	1.60	25.0	0.49	$\sim 10^{39}$
L58A	9.6	4.3	0.28	1.35	4.9	0.45	$\sim 10^8$
L60A	43.8	4.1	0.29	1.58	26.0	0.43	$\sim 10^{23}$
L8A	23.2	235.2	0.44	1.19	11.6	0.68	$\sim 10^{14}$
V30A	9.0	4.3	0.30	1.35	4.4	0.47	$\sim 10^6$
V71A	7.3	2.5	0.29	1.33	3.7	0.44	$\sim 10^{0.7}$
WT	5.8	0.6	0.25	1.36	4.6	0.31	$\sim 10^6$
Units	10^{-4}s^{-1}	10^{-7}s^{-1}	mol/L^{-1}	mol/L^{-1}	10^{-4}s^{-1}	mol/L^{-1}	

Two types of fits were made: exponential fits $k_u = k^{H_2O} \exp(m^{\ddagger}[C])$ and double exponential fits $k_u = k_L^{H_2O} \exp(m_L[C]) + k_H^{H_2O} \exp(m_H[C])$. P_2/P_1 shows relative likelihood of the double-exponential model with respect to the single-exponential model, as assessed by Akaike information criterion.
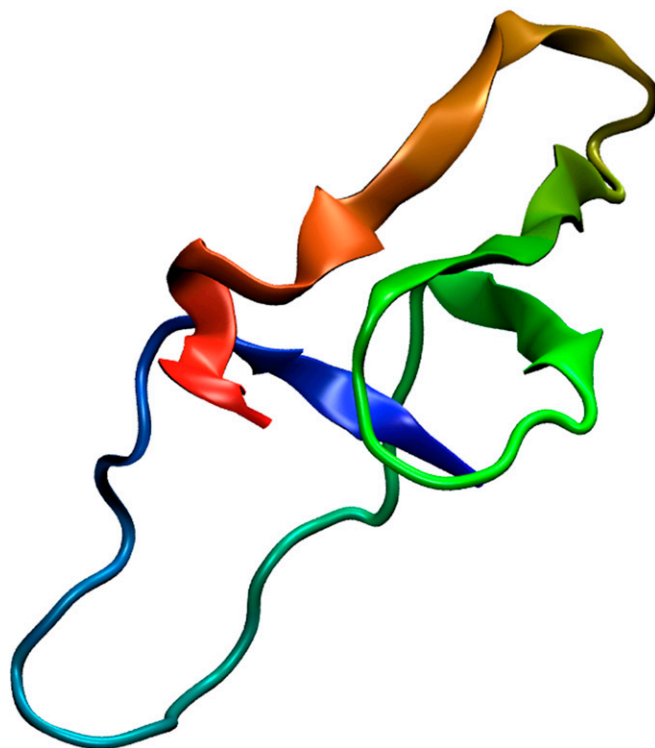
Movie S1.   An example of trajectory unfolding via the RT-loop pathway.
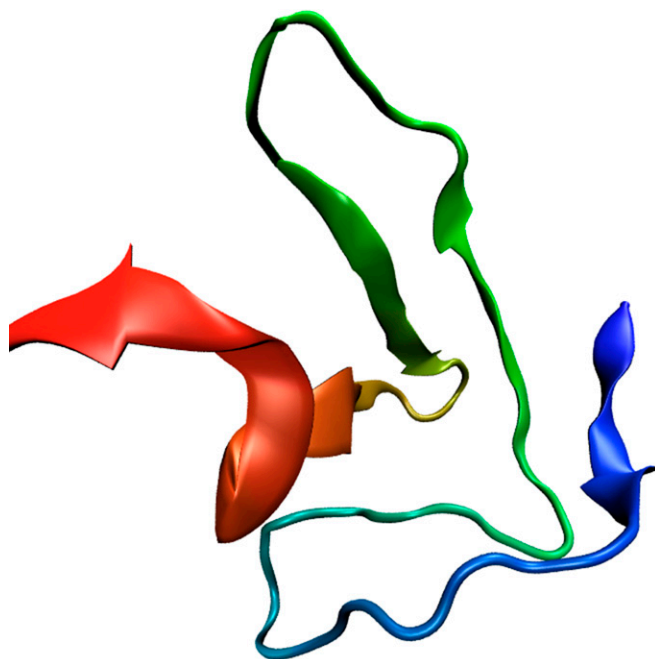
Movie S2.   An example of trajectory unfolding via the $\beta4\beta5$ pathway.