Pulling Direction as a Reaction Coordinate for the Mechanical Unfolding of Single Molecules†

Robert B. Best,[‡,⊥] Emanuele Paci,[§] Gerhard Hummer,[‡] and Olga K. Dudko*[,∥,#]

Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom, and Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892

Received: July 27, 2007; In Final Form: November 8, 2007

The folding and unfolding kinetics of single molecules, such as proteins or nucleic acids, can be explored by mechanical pulling experiments. Determining intrinsic kinetic information, at zero stretching force, usually requires an extrapolation by fitting a theoretical model. Here, we apply a recent theoretical approach describing molecular rupture in the presence of force to unfolding kinetic data obtained from coarse-grained simulations of ubiquitin. Unfolding rates calculated from simulations over a broad range of stretching forces, for different pulling directions, reveal a remarkable "turnover" from a force-independent process at low force to a force-dependent process at high force, akin to the "roll-over" in unfolding rates sometimes seen in studies using chemical denaturant. While such a turnover in rates is unexpected in one dimension, we demonstrate that it can occur for dynamics in just two dimensions. We relate the turnover to the quality of the pulling direction as a reaction coordinate for the intrinsic folding mechanism. A novel pulling direction, designed to be the most relevant to the intrinsic folding pathway, results in the smallest turnover. Our results are in accord with protein engineering experiments and simulations which indicate that the unfolding mechanism at high force can differ from the intrinsic mechanism. The apparent similarity between extrapolated and intrinsic rates in experiments, unexpected for different unfolding barriers, can be explained if the turnover occurs at low forces.

1. Introduction

Single molecule experiments with atomic force microscopes, optical tweezers, and related techniques have now been widely applied to studies of folding and unfolding of proteins and nucleic acids in the presence of a stretching force. In contrast to the situation encountered for small RNA hairpins,[1−3] it has generally not been possible to study the exchange of the folded and unfolded states of proteins or larger RNAs at equilibrium.[4,5] Promising steps in this direction are the observation of equilibrium hopping between the unfolded state and an intermediate for RNase H[6] and zipping and unzipping of helical coiled coils.[7] A reason for this difficulty of obtaining an equilibrium between the folded and the unfolded states of proteins under mechanical tension is that the two states are, with some exceptions,[8] separated by high free energy barriers. Consequently, when using a constant pulling force, unfolding experiments must be conducted at very high forces, and refolding experiments must be conducted at low forces. Intrinsic (un)folding rates, in the absence of a stretching force, are generally obtained by extrapolation from the data collected with a large pulling force.

Several models based on one-dimensional Kramers theory[9] have been developed which relate the force-dependent folding kinetics to the intrinsic unfolding dynamics. A recent formalism[10] unifies several previous approaches[11−14] and allows one to extract three parameters describing a one-dimensional landscape from kinetic data collected under different forces or pulling speeds. These parameters describe the height of the barrier and its distance from the initial state along the pulling coordinate and the intrinsic rate (see Figure 1a). The pulling coordinate is defined here as the molecular extension on which the pulling force acts, for example, the distance between the ends of a protein if it is pulled from the termini. The reduction to a one-dimensional model is a huge simplification, which may be justified in the case of protein folding with a funneled energy landscape.[15] However, finding a good (collective) one-dimensional reaction coordinate for such a high-dimensional system is challenging.[16−20] Thus, as has been previously observed,[21] it is not clear that the pulling coordinate, defined by only a single pair distance, will be relevant to the intrinsic folding pathway and therefore whether the fitted parameters are relevant descriptors of the (un)folding reaction, even for a protein with an ideally funneled landscape; indeed, such a coordinate may not even separate the folded and unfolded states.[22,23]

In the case of unfolding, a large extrapolation is generally required to determine the rate coefficients in the absence of force. The apparent similarity between the extrapolated rates and the rates measured in standard nonmechanical unfolding experiments for some proteins, particularly the extensively studied titin I27,[11,24,25] initially suggested that mechanical unfolding experiments probed the same free energy barriers. More recently, both simulations and sophisticated protein engineering experiments have indicated that the unfolding pathways at high force differ substantially from those in the

† Part of the "Attila Szabo Festschrift".

* Corresponding author. E-mail: dudko@physics.ucsd.edu.
‡ NIDDK, National Institutes of Health.
§ University of Leeds.
∥ Center for Information Technology, National Institutes of Health.
⊥ Present address: Department of Chemistry, University of Cambridge, Lensfield Road Cambridge CB2 1EW, U.K.
Present Address: Department of Physics and CTBP, University of California at San Diego, La Jolla, CA.

10.1021/jp075955j CCC: $40.75
Published on Web 02/06/2008

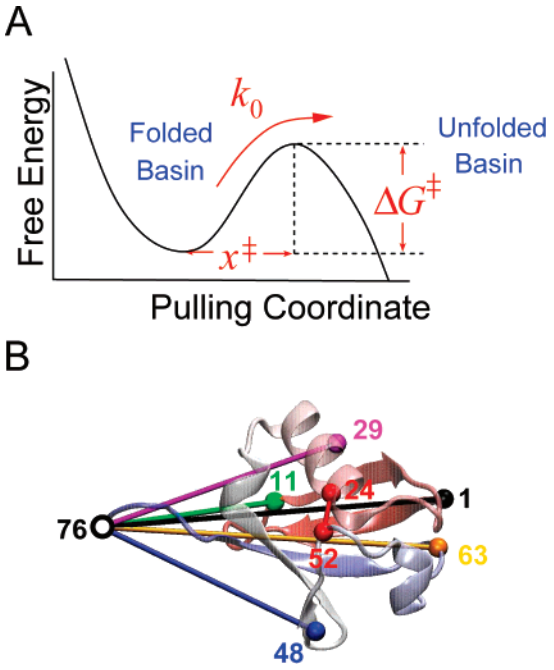

Figure 1. One-dimensional energy landscape for unfolding and the structure of ubiquitin. (a) Parameters of the one-dimensional theory[10] in which the reaction coordinate is given by the pulling coordinate, defined as the vector connecting the two residues to which force is applied. (b) The five naturally occurring pulling coordinates for ubiquitin (1−76: black, 11−76: green, 29−76: magenta, 48−76: blue, 63−76: orange) and a "designed" coordinate (24−52: red) discussed in this work are shown by lines connecting the residues which are pulled.

absence of force.[26−31] Is it just a coincidence that the extrapolated rates are similar to the intrinsic rates? Do mechanical unfolding experiments provide any information which cannot be obtained by more conventional means?

Here, we use molecular simulations of a coarse-grained model of the 76-residue protein ubiquitin to determine directly the unfolding rates as a function of force and for different pulling coordinates (i.e., different attachment points). Related models have been successfully used to investigate mechanical refolding[32] and unfolding[22,33,34] for ubiquitin and other proteins.[35] Ubiquitin is synthesized as a linked polyubiquitin chain, where the ubiquitin domains are connected at their N and C termini (pulling coordinate 1−76). In addition, it can be conjugated into short polyubiquitin chains in the cell where a lysine side-chain on one domain is linked to the C terminus of the next. Ubiquitin performs a variety of signaling roles, determined by which lysine is used to form the conjugates[36] (Lys11, Lys29, Lys48, Lys63, or pulling coordinates 11−76, 29−76, 48−76, 63−76). We have initially focused on the above "naturally occurring" pulling coordinates since they will presumably be easiest to study experimentally (Figure 1b).[25]

The paper is outlined as follows. First, we study the unfolding kinetics as a function of force under conditions where the protein is stable ($T < T_f$) and compare intrinsic parameters extrapolated from high force for two different pulling directions. Next, we use simulations at a higher temperature where the unfolding barrier is smaller to determine directly the force dependence of unfolding rates over the full range of force. In this case, we find a remarkable transition from a force-independent rate at low force to a force-dependent rate at higher force. On the basis of our interpretation of the turnover, we design a novel pulling coordinate which greatly reduces the extent of turnover. We explore the "turnover" using Brownian simulations on minimalist two-dimensional surfaces. The origin of the turnover is discussed in terms of the quality of the pulling coordinate as a reaction coordinate. In conclusion, we discuss the implications of our findings for the choice of pulling coordinate in experiments.

2. Methods

2.1. Coarse-Grained Model. The protein is described by a coarse model in which residues are represented by a particle centered on the α carbon. For interactions between residues, we use a previously described Go-like potential,[37] that has the crystal structure 1UBQ[38] as its global energy minimum. Briefly, the residues are connected by rigid bonds (constrained using SHAKE[39]) and harmonic pseudo-angles with minima taken from the native structure. A statistical potential derived from the PDB is applied to the torsion angles, and a Go-like pair contact potential is used for nonbonded interactions.

2.2. Langevin Simulations in the Presence of a Pulling Force. Pulling simulations were run with an additional constant force F acting on the vector $\mathbf{r}_{ij}$ connecting the attachment residues i and j, that is, with the potential $V = V_0 - F|\mathbf{r}_{ij}|$, where V_0 is the original force-field. Unfolding kinetics were characterized using mean first passage times, which are an accurate approximation for diffusive dynamics,[40] for simulations starting from the folded state. To account for runs that did not unfold, the maximum likelihood estimate (assuming single-exponential kinetics) of the first passage time was calculated as[35,41]

$$\langle \tau_{\mathrm{FP}} \rangle = \frac{\sum_{i=1}^{N_r} \tau_i + (N - N_r)\tau_{\mathrm{sim}}}{N_r} \qquad (1)$$

where N is the number of trials of length τ_{sim} and N_r runs cross the barrier with first passage times τ_i. The associated error is $\langle \tau_{\mathrm{FP}} \rangle N_r^{-1/2}$. Different trials were distinguished by the choice of random seed for the initial velocities and random forces in the integrator. An initial short equilibration in the absence of force was done for each trial. For unfolding, $N = 5000$ trials were used at each pulling force. Under most conditions, all of the trials resulted in unfolding: the largest fraction of trials in which the protein remained folded was ≈1%. Refolding simulations were done in an analogous way to those for unfolding; initial conditions were obtained by equilibrating the unfolded protein at a higher force (100 pN) as was done in the experimental refolding of ubiquitin.[4] One hundred refolding trials were run at zero force, resulting in a total of 76 refolding events.

Simulations were run with CHARMM,[42] using Langevin dynamics with the BBK integrator,[43] a time step of 10 fs and a friction coefficient of 0.1 ps^{-1}. A low friction coefficient was employed in the Langevin simulations, in order to accelerate the folding and unfolding dynamics. Typical values approximating the friction of water are in the range 50−100 ps^{-1}.[44,45] Using a higher friction coefficient is not expected to alter the mechanism[18] significantly, but as a consequence of the low friction, all of the rates directly obtained from simulation are faster than the experimental rates. Assuming high friction scaling of the rates (rate ∝ 1/friction), all times have therefore been scaled by a factor of 1000, accounting approximately for the friction of water.[46]

2.3. Free Energy Surfaces from Umbrella Sampling. To inform the kinetic analysis, we used thermodynamic information and free energy surfaces from umbrella sampling. Harmonic

umbrella potentials were placed on the fraction of native contacts Q, defined by

$$Q = \frac{1}{N_Q}\sum_{(ij)}\frac{1}{1+e^{\lambda(r_{ij}-\gamma r_{ij,0})}} \quad (2)$$

where r_{ij} is the distance between residues i and j, $r_{ij,0}$ is the distance between the same residues in the native structure, N_Q is the number of pairs (ij) of residues forming contacts in the native structure,[47] and the parameters λ and γ were chosen as 5 Å^{-1} and 1.2, respectively.

Ten umbrella "windows" were used with target values Q_i of 0.1, 0.2, 0.3, 0.38, 0.43, 0.52, 0.61, 0.7, 0.78, and 0.9 and a spring constant k_Q of 500 k_BT. The potential of the ith window is $V_i = (1/2)k_Q(Q - Q_i)^2$. Replica exchange between adjacent Q windows was used to enhance sampling. That is, denoting the ith replica's current coordinates $\mathbf{R}_i$, if $V_i(\mathbf{R}_{i+1}) + V_{i+1}(\mathbf{R}_i) < V_i(\mathbf{R}_i) + V_{i+1}(\mathbf{R}_{i+1})$ the swap is always accepted, otherwise it is accepted with probability $\exp[-\beta(V_i(\mathbf{R}_{i+1}) + V_{i+1}(\mathbf{R}_i) - V_i(\mathbf{R}_i) + V_{i+1}(\mathbf{R}_{i+1}))]$, $\beta = 1/k_BT$ and $\mathbf{R}$ represents the Cartesian coordinates of the whole protein. The weighted histogram analysis method (WHAM) was used to combine the data from different umbrella windows to calculate thermodynamic properties.

2.4 Brownian Dynamics Simulations on 2D Potential

Two-dimensional Brownian dynamics simulations, equivalent to Langevin dynamics in the high friction limit, were run corresponding to the stochastic differential equation $\dot{\mathbf{r}}(t) = -D\nabla V(\mathbf{r}(t)) + \mathbf{F} + \mathbf{R}(t)$; D is an isotropic diffusion coefficient, $\mathbf{F}$ is a constant pulling force in an arbitrary direction, and the $\mathbf{R}(t)$ are random forces with zero mean and variance $\langle R_\alpha(t)R_\alpha(t')\rangle = 2D\delta(t - t')$, $\alpha = x,y$. In terms of finite differences, $\mathbf{r}(t + \Delta t) = \mathbf{r}(t) - D\Delta t\nabla V(\mathbf{r}) + \mathbf{F}\Delta t + g(t)\Delta t$, where the x,y components of g are chosen independently from a Gaussian distribution with zero mean and variance $2D\Delta t$. For these simulations, the units of energy are k_BT, while length and time units are arbitrary; D was set to unity.

2.5. Theory of Force-Induced Unfolding. To relate the kinetics to the underlying free energy surface, we use a one-dimensional theory[10] derived from Kramers' description of irreversible rupture over a single barrier.[9] The advantage of the Kramers-based description over previous approaches based on the phenomenological Bell model[13,14] is that it accounts for the dependence of the transition state location on force, thus explaining the curvature in the force dependence of the kinetics and resulting in a fitted barrier height ($\Delta G^\ddagger$). The dependence of the unfolding rate $k(F) = 1/\tau(F)$ on force F for different forms of one-dimensional landscape may be expressed in a unified way[10] as

$$k_{1D}(F) = k_0\left(1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger}\right)^{1/\nu - 1} e^{\Delta G^\ddagger[1-(1-(\nu F x^\ddagger/\Delta G^\ddagger))^{1/\nu}]} \quad (3)$$

where k_0 is the "intrinsic" rate at zero force, $\Delta G^\ddagger$ is the height of the barrier and $x^\ddagger$ is the distance to the transition state (Figure 1a). For a linear-cubic form of the energy surface,[12] $\nu = 2/3$, while for a cusp-like surface,[11] $\nu = 1/2$; the Bell model is captured by eq 3 with $\nu = 1$. We present only fits with $\nu = 2/3$ in this paper, however alternate fits with $\nu = 1/2$ produce very similar results, indicating that the extracted parameters are robust and not sensitive to the precise shape of the free energy surface as long as the barrier height is accounted for in the fits (i.e., $\nu \neq 1$). Errors of the fitted parameters were estimated by generating 100 "synthetic" data sets. Each synthetic set consisted of the same number of points as the original set. To each mean unfolding time $\langle\tau\rangle$ was added a random deviate from a normal distribution with standard deviation equal to the standard error $\{(\langle\tau^2\rangle - \langle\tau\rangle^2)/N\}^{1/2}$ of unfolding times from the N simulations at that force. The error on each fitted parameter was estimated as the standard deviation of the parameters fitted to the synthetic data sets.

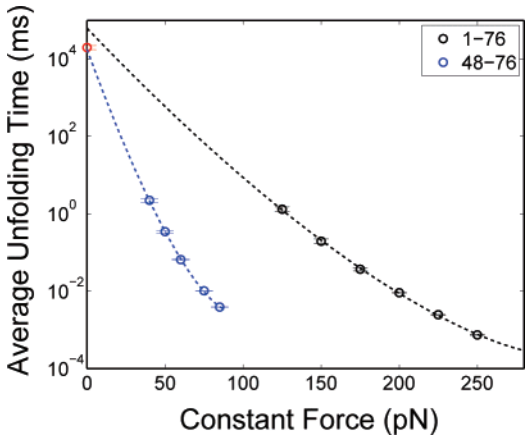

Figure 2. Unfolding times of ubiquitin at 300 K. Fits of eq 3 to the unfolding times in the range of force accessible to simulation were used to extrapolate to low forces for pulling directions 1−76 (black symbols) and 48−76 (blue symbols). The error bars are smaller than the symbol size and are shown inside the symbols. The fit parameters thus obtained are $k_0 = 1.6 \times 10^{-5}$ ms^{-1}, $x^\ddagger = 4.1$ Å, $\Delta G^\ddagger = 21.2\ k_BT$ for 1−76 and $k_0 = 5.6 \times 10^{-5}$ ms^{-1}, $x^\ddagger = 10.7$ Å, $\Delta G^\ddagger = 17.3\ k_BT$ for 48−76. The intrinsic unfolding time at zero force (red point) was obtained under a two-state assumption, from a combination of protein stability determined by umbrella sampling, and refolding times in the absence of force.

3. Results and Discussion

3.1. Extrapolation from High-Force Unfolding Kinetics. In order to match experimental conditions as closely as possible, we first study unfolding of ubiquitin at 300 K, a temperature at which the folded protein is stable ($T_f \approx 322$ K). Just as in the experiment, however, we need to use high forces under these conditions in order to unfold the protein on a practical time scale. We determined the unfolding time as a function of pulling force when pulling on the termini of the protein (pulling coordinate 1−76) and also when pulling on residue 48 and the C-terminus (coordinate 48−76). The dependence of unfolding time on pulling force is shown in Figure 2 for these two pulling directions. The data for each pulling direction can be fitted to eq 3, suggesting that a one-dimensional model is adequate to describe the data in the high-force regime. The curvature at high forces is due to the "Hammond" effect: stabilizing the unfolded state causes the barrier to move toward the folded state.[48]

An independent estimate of the unfolding time in the absence of force was obtained using a two-state approximation, $\beta\Delta G = \ln(k_f/k_0)$. From the folding rate $k_f = 0.33(0.04)$ ms^{-1} estimated from refolding simulations at zero force and the stability $\Delta G = 8.7k_BT$ from umbrella sampling, the intrinsic unfolding rate was estimated to be $5.0(0.6) \times 10^{-5}$ ms^{-1}.

We can assess how well the extrapolations describe the intrinsic unfolding (in the absence of a pulling force) by comparing the parameters obtained from the fits to the two pulling coordinates. If the two pulling directions are probing the same barrier, then the fitted intrinsic unfolding rates $k_0 = \tau_0^{-1}$ and barrier heights $\Delta G^\ddagger$ should be the same for the different coordinates (although $x^\ddagger$ may differ). This is a necessary

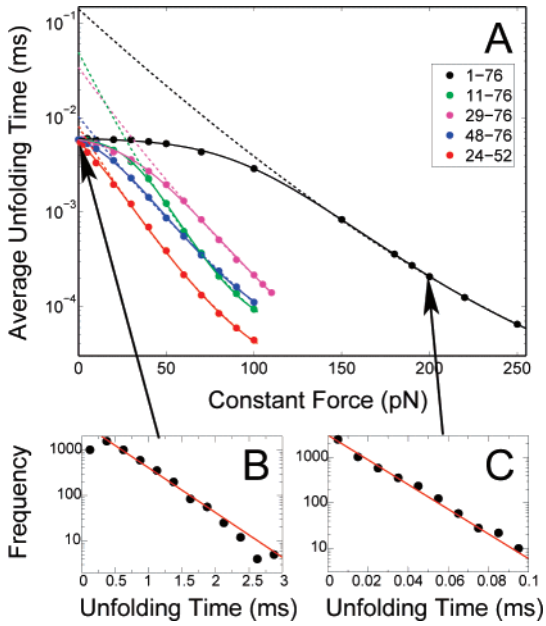

Figure 3. Dependence of unfolding time on pulling force at 380 K. (a) Unfolding times for the different attachment points are distinguished by color with simulation data given as symbols (error bars are smaller than the symbol size). Fits of the one-dimensional theory (eq 3) are shown as broken lines, while fits of the sum of a force-independent and force-dependent process (eq 4) are shown as solid lines. The distributions of unfolding times (b) at zero force and (c) for pulling with a 200 pN force on 1−76 are close to single-exponential.

requirement for the extrapolations to refer to the intrinsic unfolding pathway. The fitted rates are indeed very similar, 1.6 $\times$ 10^{-5} ms^{-1} for 1−76 and 5.6 $\times$ 10^{-5} ms^{-1} for 48−76. The close agreement between the two extrapolated rates, and the estimated intrinsic rate (see previous paragraph) suggests that the extrapolations may be used to obtain information on the intrinsic pathway. However, the fitted barrier heights for the two pulling directions differ by 4 k_BT, suggesting that these pulling directions are probing different barriers (we note that the magnitude of the barrier heights is qualitatively consistent with the thermodynamic calculations of Li et al.,[33] who found a higher barrier for 1−76 than 48−76). Similarly, a good quality fit could not be obtained when the intrinsic rate k_0 and barrier height $\Delta G^{\ddagger}$ were globally optimized for the two pulling directions. Since the fits to the two pulling directions are inconsistent with a common barrier, it is unlikely that they both probe the intrinsic unfolding barrier.

We further note that the separation of the folded state and transition state, $x^{\ddagger}$ = 4.1 Å, fitted to high force data for the pulling direction 1−76, is within the thermal fluctuations in this coordinate in the folded state at zero force. Since an essential requirement of a reaction coordinate is that it should separate the reactive states from those most populated at equilibrium,[17] the parameters fitted to the high force data are probably not relevant to the intrinsic folding pathway, at least for this coordinate.

3.2. Kinetic "Turnover" in Unfolding Times. To explore the origin of the apparent differences in unfolding barriers probed by different pulling directions, we chose to study the unfolding under relatively destabilizing conditions (380 K). At this temperature, the unfolding rate is faster, and consequently it is computationally feasible to calculate unfolding times over the full range of forces, and even in the absence of force, from simulation. Figure 3 shows the force dependence of unfolding times for each of the naturally occurring pulling coordinates at 380 K. We note that the 63−76 direction has been omitted as the application of force to this coordinate has almost no effect on the unfolding rate. This is easily explained by observing that these two residues lie at opposite ends of an extended β strand in the folded protein, so that the coordinate is not extensible. At high forces, the unfolding times for all coordinates are force dependent: as shown by the broken lines in Figure 3, the data can be fitted by the one-dimensional theory[10] (eq 3) in the intermediate to high force regime (fit parameters are summarized in Table 1). The fits are all well-defined as shown by the relatively small errors in the fitted parameters. The estimates of the intrinsic unfolding times, k_0^{-1}, for different pulling directions vary by up to an order of magnitude (note, however, that the differences in Figure 3 would appear much smaller in the axis range of Figure 2). This, together with the wide variation in fitted barrier heights (see Table 1 and Figure 3), indicates that the unfolding barriers probed by the different pulling directions are not the same. The extrapolated unfolding times are all longer than the direct estimate of unfolding time obtained from simulations at zero force. While the extrapolated zero-force unfolding times are close to the direct estimate for some pulling directions, they are remote for others.

The explanation for these discrepancies can be found in the low force regime of Figure 3: for all of the pulling coordinates the curves show "turnover" to an approximately force-independent process at low forces, meeting at zero force, which is well outside the errors of the calculated unfolding times (5000 trials were used for each force); note that the assumption of a single-exponential distribution used in the maximum likelihood estimate of unfolding times (eq 1) is justified by the actual distributions, Figure 3b,c. The turnover can be explained by a switch from the intrinsic barrier probed by pulling at very low forces, to a different barrier at high force. Force apparently has little effect on the intrinsic barrier for the naturally occurring pulling coordinates and only accelerates the rate in the high-force region. In fact, the unfolding times over the complete range of forces can be described by a sum of two processes:

$$k(F) = k_{const} + k_{1D}(F) \qquad (4)$$

a force-independent unfolding time describing the intrinsic unfolding at low forces, k_{const}, and a force-dependent process describing the "high-force" pathway, $k_{1D}(F)$, given by eq 3.

The parameters from the fit of eq 4 (solid lines in Figure 3) which describe the high force pathway match those from the fit of eq 3 to the high force data only (broken lines), confirming the above interpretation. At sufficiently low forces, unfolding proceeds via the intrinsic pathway, which appears insensitive to force applied to any of the pulling coordinates. Such a turnover in mechanism was in fact proposed on the basis of a revised interpretation of the experimental unfolding kinetics of titin I27.[29]

What distinguishes the responses to pulling along the different coordinates? The 1−76 coordinate, relating to ubiquitin with N−C linkages (the most commonly used pulling coordinate for proteins), shows the most dramatic evidence of turnover, with force dependence of the unfolding times emerging only above $\approx$60 pN. The relatively high barrier inferred from the fit (10.9 k_BT) indicates that the "high force" pathway is initially highly disfavored, and the short distance to the transition state for unfolding (1.9 Å) means that force accelerates unfolding via this pathway only weakly (the potential for the folded basin is effectively "stiff"), hence it only becomes relevant at high forces. Although less dramatic, a similar turnover is found for the other pulling coordinates: in these examples, the larger distance to

TABLE 1: Parameters Fitted to the Unfolding Times in Figure 3 (Errors in Brackets)[a]

pulling direction	fit parameters				
$i-j$	k_{const} [ms^{-1}]	k_0 [ms^{-1}]	$x^{\ddagger}$ [Å]	$\Delta G^{\ddagger}$ [k_BT]	$k_{const} + k_0$ [ms^{-1}]
1−76	164 (1)	3.0 (0.3)	1.9 (0.04)	10.9 (0.1)	167 (1)
		7.0 (1.1)	1.7 (0.1)	10.5 (0.2)	
11−76	163 (1)	7.2 (0.2)	4.4 (0.04)	8.9 (0.02)	170 (1)
		20.2 (0.6)	3.6 (0.05)	8.1 (0.04)	
29−76	150 (2)	20.9 (1.4)	2.6 (0.1)	10.5 (0.7)	171 (2)
		29.3 (2.5)	2.5 (0.1)	9.4 (0.4)	
48−76	127 (2)	43.3 (1.5)	3.0 (0.05)	7.1 (0.03)	170 (2)
		97.3 (1.8)	2.3 (0.03)	8.2 (0.2)	
24−52	79.4 (2.6)	98.7 (1.6)	3.2 (0.02)	7.2 (0.01)	178 (2)
		123 (1.2)	2.9 (0.02)	7.2 (0.02)	
Zero force					171 (2)

[a] The upper rows give the fit of all the data to the sum of force-independent and force-dependent processes, eq 4. The lower rows give the fit of the intermediate to high force data to a force-dependent process only, eq 3. Note that, for the fits to eq 4, the net rate at zero force is given by the sum of k_{const} and k_0 (last column); for 48−76 and 24−52 these two rates are of comparable magnitude, whereas for the other pulling directions, k_{const} dominates. The intrinsic unfolding rate obtained from simulations at zero force is listed for reference.

the high force transition state (i.e., a "softer" folded basin) is principally responsible for the earlier turnover.

3.3. Pulling Coordinate as Reaction Coordinate. The ineffectiveness of the stretching force in accelerating unfolding via the intrinsic pathway means that none of the "naturally occurring" pulling coordinates are particularly good as reaction coordinates at low force. This implies that the projections of the folded state and transition state onto the pulling coordinate are almost identical, so the distance to the transition state, $x^{\ddagger}$, is approximately zero. This can be directly observed by calculating a two-dimensional free energy surface as a function of the pulling coordinate and the fraction of native contacts in the absence of force. Figure 4a shows such a surface for the pulling coordinate 1−76. Because of the destabilizing conditions, the unfolded state appears at a lower free energy than the folded state, and the barrier to unfolding is small. An expansion of the region around the native state and transition state (Figure 4b) clearly shows that they lie at the same position along the pulling coordinate. Analogous projections for other naturally occurring pulling directions reveal a similar overlap of the native and transition states on the pulling coordinate. However, at high force, the good fit to the one-dimensional theory[10] shows that distance along the pulling coordinate becomes a good reaction coordinate. Clearly any alternative mechanism which is accelerated by force would supplant a force-independent intrinsic mechanism above some critical force. Consequently, almost any pulling coordinate will become a good reaction coordinate at high forces.

An ideal pulling coordinate would be one which is also a good reaction coordinate at zero or low force (i.e., for the intrinsic pathway). Such a perfect coordinate would give rise to a simple acceleration of the intrinsic rate, described by eq 3. In the absence of such an ideal coordinate, one can try to find a coordinate in which the high force fit of eq 3 gives a rate and barrier height that are as close as possible to the intrinsic ones. Of the naturally occurring pulling coordinates, the high force fit to 48−76 comes the closest to predicting the correct intrinsic rate.

3.4. Design of a Pulling Coordinate Relevant to Intrinsic Unfolding. Can we design a better pulling coordinate than the naturally occurring ones? We adopted the following very simple procedure to attempt this. First, we approximate the transition state as consisting of structures for which $Q \approx 0.73$ (see Figure 4b). We calculated average distance matrices for all possible pair distances from structures close to the transition state and close to the folded state ($Q \approx 0.9$) picked from the umbrella sampling simulations at zero force, with the correct weights. By taking the difference of these distance matrices, we obtain for each pair of residues the difference in pair distance between the folded state and the transition state: these are predictions of the $x^{\ddagger}$ for the pulling coordinates defined by each pair of residues. We chose a "designed" coordinate, 24−52, by selecting a pair of residues having a particularly large estimated $x^{\ddagger}$ and also a large sequence separation: this should correspond to a pulling coordinate which will be sensitive to force. Note that this is possible in practice, since artificial pulling coordinates have been introduced using lipoyl groups for the protein E2lip3[49] and with disulfide linkages in early experiments on lysozyme[50] and recent experiments on green fluorescent protein.[51,52]

The folding free energy surface as a function of Q and the new pulling coordinate 24−52 suggests that it should be more sensitive to force than 1−76 (Figure 4c). Expansion of the region close to the native state (Figure 4d) shows that the apparent transition state (the saddle leading out of the native state) is separated from the native state along the pulling coordinate. Thus, the intrinsic pathway should be sensitive to force.

The unfolding kinetics for 24−52 are also included in Figure 3 (red curve). It is not possible to fit the data by the one-dimensional model given by eq 3 alone, so the coordinate is not ideal. However, like the other coordinates, it can be fitted to the sum of force-independent and force-dependent processes given by eq 4 (note that it can also be fitted by a sum of two force-dependent processes, but the fitting parameters are not well determined). According to our stated criteria, the new coordinate is the best of those considered, since the rates extrapolated from high force data are very close to the true rates. Moreover, the rates fitted for "high force" and "low force" processes (Table 1) compete at zero force (i.e., k_0 is close to k_{const}).

Exploring the unfolding kinetics for pulling along the designed good coordinate 24−52 reveals an interesting feature of the unfolding mechanism: a transient unfolding intermediate becomes visible in the time series of the end-to-end distances, $r_{ij}(t)$. The fit of the one-dimensional theory[10] to the unfolding times remains justified provided that the rate-limiting barrier is the first one encountered along the pulling coordinate. This assumption is supported by the free energy surface for 24−52, Figure 4c, where the intermediate appears to be on the unfolded side of the major transition state. Closer examination indicates that an intermediate is also visited when pulling along the other coordinates; while not visible in the pulling coordinate itself, it is seen in Q. Inspection of the folding free energy surface for 1−76 in Figure 4d shows how the intermediate is separated from the native state along Q but not along the pulling

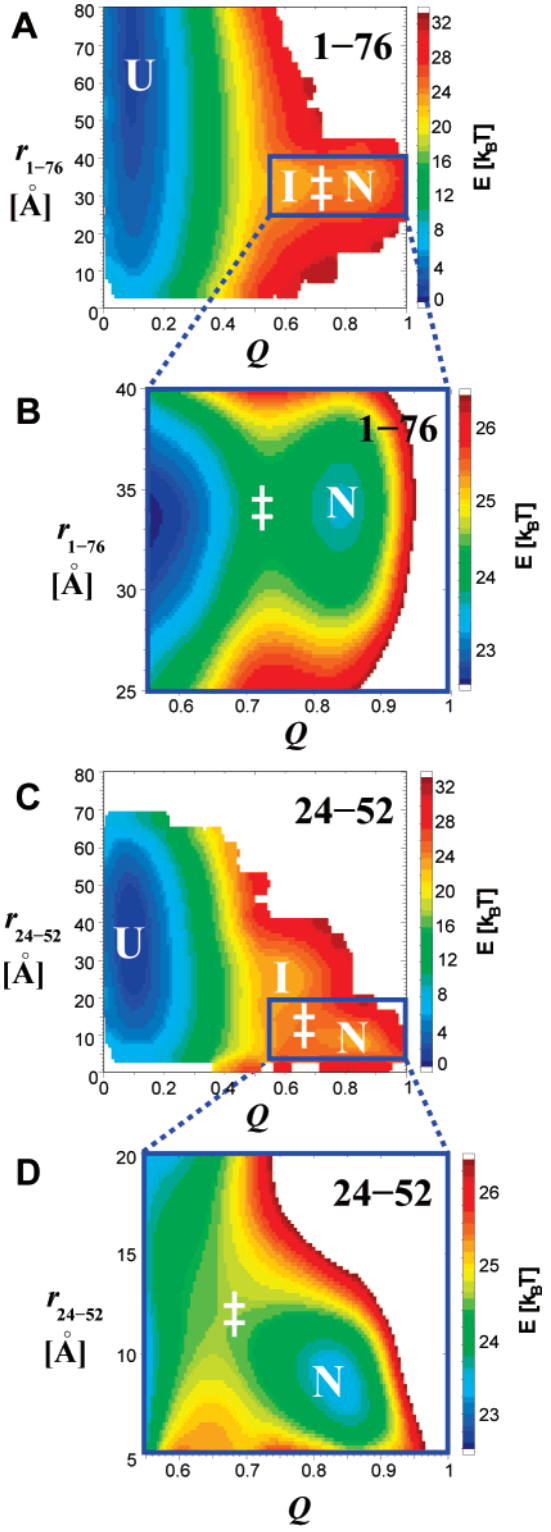

Figure 4. Folding free energy landscapes at 380 K and zero force. (A) Free energy $W(r_{1-76}, Q)$ as a function of the pulling coordinate r_{1-76} and the fraction of native contacts Q. (B) Expansion of (A) showing native state and transition state region. (C) Free energy surface corresponding to the designed pulling coordinate r_{24-52}. (D) Expansion of (C) showing native and transition state regions. The positions of the native state, the major transition state, an unfolding intermediate, and the unfolded state are indicated on the plots by the symbols "N", "‡"', "I", and "U", respectively.

coordinate. In connection with the intermediate, it is noteworthy that a recent experimental study in which GFP was pulled in several directions found one direction for which an unfolding intermediate could be observed in the pulling coordinate.[52]

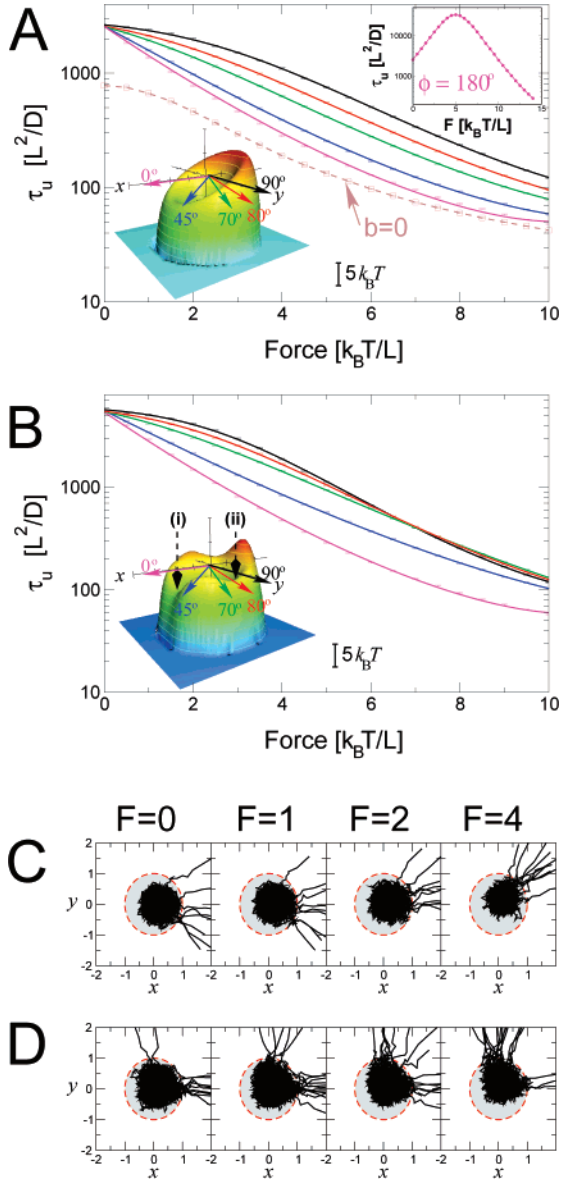

Figure 5. Kinetics on model two-dimensional surfaces, for different pulling directions. (a) Escape kinetics from a basin with a single saddle favoring escape along x (lower left inset), described by eq 5 ($a = 10$, $b = 5$). The force dependence of the escape rate is fitted with the sum of a force-dependent and a force-independent process, eq 4. (b) Escape kinetics from a basin with two alternative saddles (indicated by (i) and (ii) in inset), given by eq 6 ($a = 10$, $b = 5$, $c = 2$, $d = 4$). The data are fitted with the same model as in (a). The pulling directions are color coded, and indicated by arrows in the insets: $\phi = 0°$ (magenta; along x), $\phi = 45°$ (blue), $\phi = 70°$ (green), $\phi = 80°$ (red), and $\phi = 90°$ (black; along y). Data obtained for a system with no preferred escape direction at zero force (eq 5 with $a = 5$ and $b = 0$) are given by brown squares in (a). An example of an initial increase in unfolding time caused by pulling along the $-x$ direction ($\phi = 180°$) is shown in the top right inset. The dependence of escape pathways on force is illustrated using some typical trajectories in (c) and (d), corresponding to the energy surfaces in (a) and (b) respectively: the approximate location of the basin is indicated by the shaded disc.

3.5. Origins of Turnover: Simple 2D Models. The turnover seen in the unfolding kinetics, from a force-independent intrinsic pathway to a force-dependent pathway, indicates a change in the nature of the unfolding barrier. A natural explanation for this would be a discrete switch between two "saddles" in the free energy landscape, one which is sampled at low force and one at high force; this is implied by our heuristic fit of the ubiquitin unfolding data to the sum of a force-independent

TABLE 2: Fits to First Passage Times on Two-Dimensional Landscapes[a]

pulling direction	fit parameters				
ϕ	k_{const}	k_0	$x^{\ddagger}$	$\Delta G^{\ddagger}$ [k_BT]	$k_{const} + k_0$
		Single Saddle:			
90°	3.26×10^{-4}	5.14×10^{-5}	0.699	6.81	3.77×10^{-4}
80°	2.88×10^{-4}	1.00×10^{-4}	0.645	6.35	3.88×10^{-4}
70°	2.33×10^{-4}	1.59×10^{-4}	0.623	5.92	3.92×10^{-4}
45°	5.59×10^{-5}	3.34×10^{-4}	0.595	5.27	3.90×10^{-4}
0°		3.91×10^{-4}	0.643	5.17	3.91×10^{-4}
zero force					3.91×10^{-4}
		Double Saddle:			
90°	1.58×10^{-4}	1.72×10^{-5}	0.882	7.83	1.79×10^{-4}
80°	1.49×10^{-4}	3.24×10^{-5}	0.749	7.45	1.81×10^{-4}
70°	1.16×10^{-4}	6.72×10^{-5}	0.608	7.23	1.83×10^{-4}
45°	1.09×10^{-5}	1.68×10^{-4}	0.559	5.83	1.79×10^{-4}
0°		1.85×10^{-4}	0.712	5.81	1.85×10^{-4}
zero force					1.79×10^{-4}

[a] At $\phi = 0°$ (pulling along x) a single force-dependent process given by eq 3 is fitted, otherwise the sum of a force-independent process with escape rate k_{const} and a force-dependent process (eq 4) is used. Units of length and time are arbitrary. Pulling directions are given as the angle from the x axis. Thus, x corresponds to 0°, and y corresponds to 90°. The last column gives the net rate at zero force for each pulling direction and the escape rate determined in the absence of force.

process and a force-dependent process. However, there is a slightly different possibility, in which force distorts the low-force unfolding barrier continuously in such a way that unfolding becomes increasingly force-sensitive (i.e., the folded and transition states become separated along the pulling coordinate).

We investigate the alternative scenarios for turnover using simplified two-dimensional energy landscapes. The first example, depicted in Figure 5a, represents a scenario in which force causes a continuous distortion of the barrier, such that at high force, the saddle, has shifted from being along x to being close to the pulling direction. The dynamics occurs on a combined energy surface (including the effect of the force F) given by

$$V(x,y) = -a(x^2 + y^2 - 1)^2 - bx - xF\cos\phi - yF\sin\phi \quad (5)$$

which describes a circular barrier centered approximately at (0,0) with a saddle favoring escape along the $+x$ direction: x is the ideal coordinate for this system. The last two terms describe an additional potential corresponding to a constant force of magnitude F at an angle ϕ (the pulling direction) to the x axis. Escape rates were calculated from mean first passage times from Brownian dynamics simulations on this potential. The force dependence of escape times resulting from pulling along the x direction fits eq 3, confirming that it is a good coordinate (Table 2). Pulling along other directions gives rise to "turnover" in the kinetics which can be fitted by the sum of force-independent and dependent processes as for the real unfolding data, using eq 4. The force-dependent process describes escape along the pulling direction at high force (Table 2). This explains why the fitted barrier heights from the force-dependent part of the kinetics increase as the pulling direction gets closer to the y axis. Figure 5c shows the continuous switch in escape path as the pulling force along y is increased. The naturally occurring pulling directions would be best represented by pulling along the y coordinate, the black curve in Figure 5a. The designed coordinate 24−52 might correspond to pulling somewhere between x and y, for example, the green curve in Figure 5a, since the intrinsic pathway is neither orthogonal nor parallel to the pulling direction (Figure 4d). Note that like 24−52 it is possible to fit the green curve in Figure 5a with eq 4, although the intrinsic pathway is expected to have some force dependence.

It is also possible that applying force to the pulling coordinate acts to oppose the intrinsic unfolding mechanism. The most extreme example for the potential illustrated in Figure 5a would be pulling along the $-x$ direction ($\phi = 180°$). Any direction with a component along $-x$ would have a similar, though weaker, effect. Here, the force both lowers the barrier on the $-x$ side of the well while raising it on the $+x$ side. Thus, the escape is initially slowed by the increase in the intrinsic barrier (along $+x$) until the alternative barrier (along $-x$) becomes more favorable and the rate increases, a scenario illustrated in the upper right inset of Figure 5a. While we did not observe this phenomenon for ubiquitin, some evidence for an initial slowing of unfolding rate has been seen in constant force simulations of the I27 domain of titin.[53]

An interesting situation arises when there is no favored direction at zero force: when $b = 0$ in eq 5, the potential is radially symmetric. Without force, no single direction is a perfect coordinate, though $r = (x^2 + y^2)^{1/2}$ is a good parametric coordinate. Applying force causes selective acceleration along the pulling direction, resulting in turnover in the kinetics, included in Figure 5a. This case is not, however, described by eq 4. Thus, the vectorial nature of the force will make the escape process one-dimensional at high force, even if there is no single direction which is a good reaction coordinate at zero force.

In the alternative scenario, a two saddle surface is studied, given by (again including the potential giving rise to the pulling force):

$$V(x,y) = -a(x^2 + y^2 - 1)^2 - bxe^{-dy^2} - cye^{-dx^2} - xF\cos\phi - yF\sin\phi \quad (6)$$

The main saddle in this case is also along $+x$, but there is a smaller subsidiary saddle along $+y$. Figure 5b shows the surface and corresponding escape times. In the absence of force, almost all escape would be along x and applying force along this direction again gives a very good fit to eq 3 (Figure 5b). Pulling along y gives a force-independent escape (along x) at low forces, switching to a force-dependent escape (along y) at high forces. This interpretation is confirmed by the close match of the force-independent escape rate for pulling along y (6.1×10^3) with the extrapolated rate from the fit of eq 3 to pulling along x (5.7×10^3). The discrete switch from escape along x to escape along y can be clearly seen in Figure 5d. Pulling along directions intermediate between x and y produces varying degrees of turnover.

We can conclude that there are at least two "toy" models for a shift of barrier that produce force-dependent lifetimes consistent with the unfolding simulations of ubiquitin. Therefore, the good fit to the sum of a force-independent and force-dependent model does not necessarily imply a discrete switch between two competing pathways but could also be explained by a continuous distortion of the barrier with force. We can, however, make the general deduction that the effect of force is to shift the barrier from a "low force" scenario where the unfolding is almost completely force independent, to a force-dependent scenario at "high force". The effect of force is to shift the barrier in such a way that the pulling coordinate becomes a better reaction coordinate for unfolding.

4. Conclusion

It is frequently said that mechanical unfolding experiments have a "well-defined" reaction coordinate,[7,54] since the pulling

coordinate is clearly that along which the force acts. But is it also a good reaction coordinate? Our results demonstrate that, at least for the ubiquitin model used in this study, most pulling coordinates are clearly not good reaction coordinates for unfolding at zero force because they produce extrapolated rates which differ from the "true" rates. The reason for this discrepancy is that at very low force (that is, for the intrinsic mechanism) (i) they do not resolve folded and unfolded states and (ii) they are not effective at accelerating the unfolding rate. The latter finding is a consequence of the negligible separation of the folded and transition states on these coordinates at low force (with $x^{\ddagger}$ of a few Å being within thermal fluctuations of the corresponding molecular extension in the folded state). Nevertheless, at higher forces, the unfolding kinetics do become force-dependent and can be fitted by one-dimensional theory,[10] indicating that the application of force turns the pulling coordinate into a relevant reaction coordinate (for the high force mechanism). Force will selectively accelerate any unfolding routes on which it acts (i.e., those with a separation of folded and transition states on the pulling coordinate), such that these will dominate at sufficiently high forces. We have shown that two-dimensional models are sufficient to reproduce this effect. The observation of kinetic turnover in other proteins supports the generality of our findings.[53] Although parameters obtained from one-dimensional fits to high-force data refer only to the high-force mechanism, such fits are nonetheless valuable: a description of the high-force pathway is important since it is that which is directly probed by the experiments. Furthermore, quantitative information on the unfolding mechanism at high force can be used in comparative studies, for example Φ-value analysis of mutants.[30,31,55]

The apparent similarity between zero force unfolding rates extrapolated from high force unfolding data and rates measured by ensemble experiments in some cases seems to contradict the difference in unfolding mechanism at low and high force found in simulations and protein engineering experiments.[26−31] The turnover we observe in the unfolding times from our simulations can be used to reconcile these findings. First, the turnover indicates a shift of the unfolding barrier from one which is not sensitive to the application of force, to a force-sensitive barrier, with increasing force, indicating a change of unfolding mechanism. Second, the turnover generally occurs at low forces. Thus, extrapolation of the high force rates to zero force will not deviate much from the intrinsic rates: indeed in our example, all of the extrapolated rates were within ≈1.5 orders of magnitude of the true rate, typical of the error associated with extrapolations in experiments. We can also use the turnover to suggest an experimental criterion to determine the pulling coordinate which best describes the intrinsic unfolding (for which extrapolated rates will agree best with the true rates at zero force). Since the extrapolated unfolding rates will always be slower than the "true" rates, the best pulling coordinate would be the one for which the fit of eq 3 gives the largest extrapolated rate. Such a coordinate represents the "Achilles' heel" of the protein, that is to say, pulling on it would be the most effective at accelerating the unfolding of the protein and yield parameters most relevant to the intrinsic folding barrier.

Acknowledgment. We thank Attila Szabo for fostering this collaboration and for many helpful discussions on the work. R.B.B. and G.H. were supported by the intramural research program of the NIDDK, NIH, and O.D. was supported by the intramural research program of the CIT, NIH. E.P. acknowledges INTAS for financial support.

References and Notes

(1) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832−1835.
(2) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. *Science* **2006**, *314*, 1001−1004.
(3) Woodside, M. T.; Behnke-Parks, W. M.; Larizadeh, K.; Travers, K.; Herschlag, D.; Block, S. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6190−6195.
(4) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674−1678.
(5) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Nature* **2005**, *437*, 231−234.
(6) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309*, 2057−2060.
(7) Bornschlölgl, T.; Rief, M. *Phys. Rev. Lett.* **2006**, *96*, 118102.
(8) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(9) Kramers, H. A. *Physica* **1940**, *7*, 284−303.
(10) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(11) Hummer, G.; Szabo, A. *Biophys. J.* **2003**, *85*, 5−15.
(12) Dudko, O. K.; Filippov, A. E.; Klafter, J.; Urbakh, M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11378−11381.
(13) Bell, G. I. *Science* **1978**, *200*, 618−627.
(14) Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72*, 1541−1555.
(15) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.
(16) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334−350.
(17) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(18) Best, R. B.; Hummer, G. *Phys. Rev. Lett.* **2006**, *96*, 228104.
(19) Cho, S. S.; Levy, Y.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *103*, 586−591.
(20) Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 12689−12698.
(21) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2031−2035.
(22) Kirmizialtin, S.; Huang, L.; Makarov, D. E. *J. Chem. Phys.* **2005**, *122*, 234915.
(23) West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *125*, 204910.
(24) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694−3699.
(25) Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738−743.
(26) Paci, E.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6521−6526.
(27) Lu, H.; Schulten, K. *Biophys. J.* **2000**, *79*, 51−65.
(28) Li, H.; Carrion-Vazquez, M.; Oberhauser, A. F.; Marszalek, P. E.; Fernandez, J. M. *Nat. Struct. Biol.* **2000**, *7*, 1117−1120.
(29) Williams, P. M.; Fowler, S. B.; Best, R. B.; Toca-Herrera, J. L.; Scott, K.; Steward, A.; Clarke, J. *Nature* **2003**, *422*, 446−449.
(30) Best, R. B.; Fowler, S. B.; Toca-Herrera, J. L.; Steward, A.; Paci, E.; Clarke, J. *J. Mol. Biol.* **2003**, *330*, 867−877.
(31) Ng, S. P.; Rounsevell, R. W.; Steward, A.; Geierhaas, C. D.; Williams, P. M.; Paci, E.; Clarke, J. *J. Mol. Biol.* **2005**, *350*, 776−789.
(32) Best, R. B.; Hummer, G. *Science* **2005**, *308*, 498b.
(33) Li, P.-C.; Makarov, D. E. *J. Chem. Phys.* **2004**, *121*, 4826−4832.
(34) Kleiner, A.; Shakhnovich, E. I. *Biophys. J.* **2007**, *92*, 2054−2061.
(35) West, D. K.; Brockwell, D. J.; Olmsted, P. D.; Radford, S. E.; Paci, E. *Biophys. J.* **2006**, *90*, 287−297.
(36) Weissman, A. M. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 169−178.
(37) Karanicolas, J.; Brooks, C. L., III. *Protein Sci.* **2002**, *11*, 2351−2361.
(38) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(39) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(40) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980**, *72*, 4350−4357.
(41) Yeh, I.-C.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12177−12182.
(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(43) Brooks, C. L., III; Brünger, A.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495−500.
(44) Pastor, R. W.; Karplus, M. *J. Chem. Phys.* **1989**, *91*, 211−218.
(45) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.
(46) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755−1763.

(47) Karanicolas, J.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3954−3959.

(48) Hyeon, C.; Thirumalai, D. *J. Phys.: Condens. Matter* **2007**, *19*, 113101.

(49) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731−737.

(50) Yang, G.; Cecconi, C.; Baase, W. A.; Vetter, I. R.; Breyer, W. A.; Haack, J. A.; Matthews, B. W.; Dahlquist, F. W.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 139−144.

(51) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1244−1247.

(52) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12724−12728.

(53) West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *124*, 154909.

(54) Best, R. B.; Clarke, J. *Chem. Commun.* **2002**, 183−192.

(55) Best, R. B.; Fowler, S. B.; Toca-Herrera, J. L.; Clarke, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12143−12148.