Locating the Barrier for Folding of Single Molecules under an External Force

Olga K. Dudko,[1,*] Thomas G. W. Graham,[2] and Robert B. Best[2,†]

[1]Department of Physics and Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92093, USA

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 7 July 2011; published 7 November 2011)

Single-molecule pulling experiments on the folding of biomolecules are usually interpreted with one-dimensional models in which the dynamics occurs on the ''pulling coordinate.'' Paradoxically, the free-energy profile along this coordinate may lack a refolding barrier, yet a barrier is known to exist for folding; thus, it has been argued that pulling experiments do not probe folding. Here, we show that transitions monitored in pulling experiments probe the true folding barrier but that the barrier may be hidden in the projection onto the pulling coordinate. However, one-dimensional theory using the pulling coordinate still yields physically meaningful energy landscape parameters.

DOI: 10.1103/PhysRevLett.107.208301 PACS numbers: 82.37.−j, 87.10.Tf, 87.15.ap

Single-molecule force experiments are a powerful tool for revealing details of the folding free-energy landscape of proteins and nucleic acids inaccessible to ensemble methods [1]. Precise measurements at low force have revealed hopping between two or more metastable states with different molecular extensions, attributed to force-dependent folding and unfolding [2–5]. A quantitative description of this dynamics is challenging because it occurs on a complex high-dimensional energy landscape. To enable interpretation of single-molecule pulling experiments, the complexity is usually reduced by invoking one-dimensional (1D) models, in which the original dynamics is projected onto a single coordinate—the molecular extension, or ''pulling coordinate'' [6,7]. Projection of the full energy landscape onto a 1D coordinate yields a free-energy profile, or potential of mean force (PMF) [8]. A remarkable simplification occurs if the projected dynamics can be described as diffusion on this 1D PMF, as folding energy landscape theory suggests is possible for certain collective coordinates [9]. However, although a 1D PMF is formally defined for an arbitrary coordinate, diffusion on this coordinate may not, in general, capture the folding dynamics accurately; therefore, interpretation of the PMF as an energy landscape may be misleading.

Using a 1D model with the pulling coordinate as reaction coordinate, it has been argued that ''hopping'' phenomena seen in experiments on single molecules under stretching force do not reflect true folding events but rather downhill collapse of the unfolded polymer [10]. This conclusion rested on the lack of a barrier in the 1D PMF for the pulling coordinate at zero force, with a barrier only emerging once the energy profile is tilted by the applied force. Thus, it was suggested that observed hopping events are experimental artifacts [10].

Here, we use molecular simulation to investigate the role of force on the hopping phenomena observed in single biomolecules held under a stretching force. We show that, at zero force, the 1D PMF for the pulling coordinate may indeed lack a refolding barrier, in accord with the intriguing proposal by Fernandez and colleagues [10]. However, we find that this is not because the barrier does not exist but because it is hidden by overlap with the folded and unfolded states in the projection onto the pulling coordinate. The barrier is clearly visible in the PMF for a good folding coordinate, and hopping occurs over the same barrier as that probed by pulling experiments. Therefore, the hopping phenomena observed in experiments are true folding events. Remarkably, 1D kinetic models based on the pulling coordinate provide meaningful parameters of the energy landscape even when the pulling coordinate is a poor reaction coordinate.

As prototypical examples, we study the 56-residue protein G and a 44-residue RNA hairpin, similar to systems studied in pulling experiments [3,11]; an alternative hairpin is presented in the Supplemental Material [12]. We use a coarse-grained molecular model, with a Gō-like potential, for each molecule. The model for protein G has been previously described [13], and the nucleic acid model is described in the Supplemental Material [12]. An advantage of such models is that free-energy surfaces can be constructed from this detailed molecular representation without assuming [10] phenomenological energy surfaces.

We generate Langevin dynamics trajectories of the full protein and RNA molecular models in the presence of an added pulling potential $V(x) = -Fx$, where x is the pulling coordinate and F the force. In Fig. 1, we monitor the folding of protein G projected onto two different coordinates: either the fraction of native contacts, Q, effective for monitoring folding in simulations [14], or the pulling coordinate (x_{1-56}, for pulling on residues 1 and 56) itself. Folding is often described as 1D diffusion along such reaction coordinates [9], characterized by a 1D free-energy profile, or PMF, and a diffusion coefficient D (possibly position-dependent [15]). The PMF for Q (Fig. 1), given by $G(Q) = -\ln \int \exp[-V(\mathbf{R})]\delta[Q - Q(\mathbf{R})]d\mathbf{R}$, where V is the potential energy and $\mathbf{R}$ the $3N$ coordinates of the

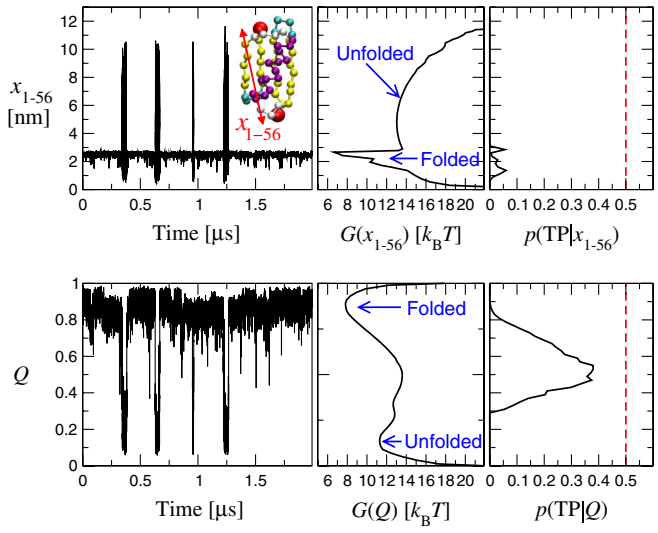

FIG. 1 (color online). Equilibrium folding of protein G at zero force projected onto the x_{1-56} pulling coordinate (top) and onto the folding coordinate Q (bottom). Left: trajectories $x_{1-56}(t)$ and $Q(t)$ both revealing hopping at zero force. Center: the corresponding 1D PMFs $G(x_{1-56})$ and $G(Q)$. Right: the conditional probability of being on a transition path for a given value of the coordinate, $p(\mathrm{TP}|x_{1-56})$ and $p(\mathrm{TP}|Q)$. The vertical dashed red line indicates the theoretical maximum of $p(\mathrm{TP}|Q)$, $p(\mathrm{TP}|x)$.

protein, reveals the expected free-energy minima for the unfolded state (low Q) and folded state (high Q), separated by a barrier. However, the form of the PMF $G(x_{1-56})$ for the pulling coordinate is unusual, with a single folded energy minimum and a very broad unfolded state with a vanishing barrier to refolding. While a barrier may not be needed for two-state folding in higher dimensions [16], lack of a barrier in 1D seems unexpected.

To resolve the paradox of a barrier appearing in $G(Q)$ but not in $G(x)$, we construct the two-dimensional (2D) free-energy landscape $G(Q, x_{1-56})$, shown in Fig. 2(a), where 2D umbrella sampling has been used to probe accurately the high free-energy regions [13]. Two low free-energy regions are seen, a narrow folded minimum at high Q and low x (green arrow, labeled with F) and a broad unfolded minimum at low Q (blue arrow, labeled with U); the unfolded minimum is consistent with the end-end distribution of a wormlike chain with persistence length 0.6 nm for the protein and 2.0 nm for the RNA, similar to experimental values (see Supplemental Material [12]). Note that the pulling coordinate does not reliably separate folded and unfolded states in the absence of force, since there is a significant probability of unfolded and folded chains with similar extensions [17]. Notable is the location of the barrier, at intermediate Q but very small x (red arrow, labeled with ‡). This explains the presence of the barrier in Q [PMF $G(Q)$, shown on the right face of Fig. 2(a)], where it is separated from unfolded and folded states, and the absence of one in x [PMF $G(x)$, shown on the front face of Fig. 2(a)], where it is hidden in the projection by the lower free energy of the folded and unfolded states. Also shown in Fig. 2 are free-energy surfaces for protein G for a different pulling coordinate [x_{10-32}, Fig. 2(b)] and for the hairpin pulled from the termini [Fig. 2(c)]. Although the location and height of the barrier differ, the same general features appear: a barrier is visible on Q but hidden by the folded and unfolded states on x.

The overlap of the folded, unfolded, and transition states on the pulling coordinate renders it a poor reaction coordinate at zero force. We quantify how well a coordinate ζ

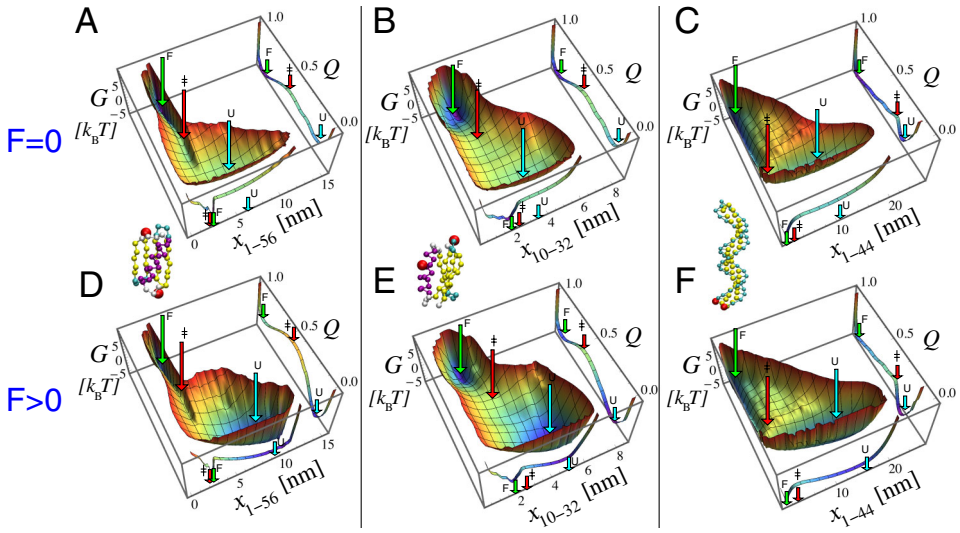

FIG. 2 (color online). 2D free-energy landscapes $G(Q, x)$. Folding free-energy surfaces at zero force for (a) protein G, $G(Q, x_{1-56})$; (b) protein G, $G(Q, x_{10-32})$; (c) RNA hairpin, $G(Q, x_{1-44})$; and for corresponding systems at forces of (d) 1.75 k_BT/nm, (e) 2.5 k_BT/nm, and (f) 0.5 k_BT/nm. 1D PMFs for the pulling coordinate x and the fraction of native contacts Q are shown on the front and right faces, respectively. The approximate positions of the *true* folded, unfolded, and transition states are indicated by green, blue, and red arrows (and corresponding labels F, U, and ‡) on the 2D PMF $G(Q, x)$ and 1D PMFs $G(Q)$, $G(x)$. The folded structure and pulling coordinate are illustrated next to the matching energy surfaces.

separates the transition state from stable states using a Bayesian criterion [18] $p(\mathrm{TP}|\zeta)$, the probability of being on a transition path rather than in the folded or unfolded state, for a given value of ζ. For a good coordinate, transition states will be collected at a single value of $\zeta^{\ddagger}$, where the value of $p(\mathrm{TP}|\zeta)$ will be near the theoretical maximum of 0.5. By this measure, Q is a good coordinate (Fig. 1, bottom right), while x_{1-56} is clearly a poor coordinate at zero force (Fig. 1, top right).

The 2D free-energy surfaces show that the lack of a barrier in the 1D PMF for x at zero force is due to x being a poor reaction coordinate at zero force and not because of the nonexistence of a barrier. This deficiency of the pulling coordinate raises a question over the interpretation of single-molecule experiments, where 1D models are widely employed; 2D models have only recently been proposed [7,19,20] and require more parameters. We discuss two possible scenarios: in the first, 1D models are justified because it only takes a small force to shift the "softer" unfolded state and transition state to larger extensions and thus reveal all three states in the PMF along x. This scenario is illustrated by protein G pulled along the x_{10-32} coordinate: the barrier, being clearly visible on the 2D landscape but lost in the projection on x [Fig. 2(b)], moves to larger extensions at low force such that its location on the 1D PMF along x soon coincides with the true location on the 2D landscape [Fig. 2(e)]. At these forces, x becomes a better coordinate [21]; further increasing force reveals "Hammond behavior" (as also expected for 1D models), where the transition state moves toward the folded state and the two eventually merge as the barrier to unfolding vanishes. A similar scenario is observed for the hairpin [Fig. 2(c) and 2(f)]. In the second scenario, force has little effect on the displacement of the "stiffer" transition state from the folded state. This scenario is illustrated by protein G pulled along the x_{1-56} coordinate, where shearing of antiparallel β strands results in a much stiffer transition state. Because, in this scenario, x remains a poor reaction coordinate up until higher force, the justification of 1D models appears less straightforward.

To investigate the implications for interpreting kinetic data typically available in experiments, we have studied force-dependent folding and unfolding kinetics using both molecular simulations of the full Gō models and Brownian dynamics simulations on the molecular free-energy surfaces in Fig. 2. Good agreement between the kinetics obtained by the two methods (Fig. 3) confirms that the 2D landscape captures essential features of the unprojected energy landscape. The (un)folding kinetics from the Brownian dynamics simulations fit well in all cases to the 1D theory of force-dependent kinetics [6] (Fig. 3). For the first scenario above, a good fit of the 1D theory might be expected, since force quickly makes x a good coordinate and the effect of higher dimensions is weak. Indeed, if x becomes a good coordinate upon application of force, then the fitted parameters (Table I) do literally correspond to the 1D PMF on x, since force simply tilts this PMF. However, the meaning of the 1D parameters for

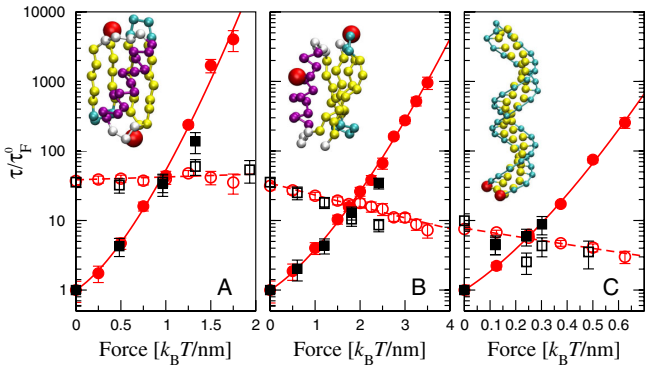

FIG. 3 (color online). Folding (filled symbols) and unfolding (empty symbols) lifetimes for protein G with (a) x_{1-56} and (b) x_{10-32} pulling coordinates and (c) for the hairpin. Black squares represent equilibrium Gō model simulations, red circles represent 2D Brownian dynamics simulations, and lines represent fits to a one-dimensional theory [6].

the second scenario is less clear: indeed, the parameters for x_{1-56} are clearly inconsistent with the formal 1D $G(x_{1-56})$ [Fig. 2(a), front face], as the fitted transition state displacement is negative (-0.1 nm; see Table I), indicating that the barrier is shorter than the folded state—obviously impossible in a 1D PMF.

Remarkably, we find instead that the parameters from a fit to 1D theory (Table I) correspond to the location and height of the *true* barrier, i.e., that captured by the good reaction coordinate. This can be understood if the dynamics of x is fast relative to that of Q [13] (note that x can be fast even though dynamics of the pulling apparatus, to which the molecule is attached by a linker, may be slow [22]). In this case, the folding dynamics is faithfully captured by the force-dependent 1D PMF for Q, $G(Q;F)$, defined by $\exp[-G(Q;F)] = \int \exp[-G(Q,x) + Fx]dx$, and (un)folding times can be calculated from 1D Kramers theory applied to $G(Q;F)$. Next, note that the Kramers rate expression in terms of Q can be transformed to one in terms of an effective extension $\bar{x}$, given by

$$\exp[F\bar{x}(Q)] = \langle \exp[Fx] \rangle_Q, \tag{1}$$

where $\langle \exp[Fx] \rangle_Q$ denotes the average of $\exp[Fx]$ for a fixed Q over the distribution at zero force [23]. The Kramers expression for $\tau(F)/\tau(0)$ using $\bar{x}$ becomes

TABLE I. Fits of folding (F) and unfolding (U) kinetics in Fig. 3 to 1D theory [6] for the dynamics on the x coordinate, using $\nu = 2/3$. Numbers in italics were fixed to values based on the energy landscape. $\Delta x^{\ddagger}$ is the difference in x between the transition state and reactant (folded or unfolded) state.

Molecule, coordinate	Process	$\Delta x^{\ddagger}$ [nm]	$\Delta G^{\ddagger}$ [k_BT]	τ_0 [μs]
Protein G, x_{1-56}	F	-3.12 ± 0.19	0.86 ± 0.28	0.06 ± 0.01
Protein G, x_{1-56}	U	-0.10 ± 0.06	*6.0*	2.11 ± 0.36
Protein G, x_{10-32}	F	-1.39 ± 0.04	1.05 ± 0.19	0.06 ± 0.01
Protein G, x_{10-32}	U	0.42 ± 0.03	*6.0*	1.90 ± 0.32
Hairpin, x_{1-44}	F	-7.01 ± 0.21	1.43 ± 0.36	0.17 ± 0.04
Hairpin, x_{1-44}	U	1.45 ± 0.10	*6.0*	1.29 ± 0.31

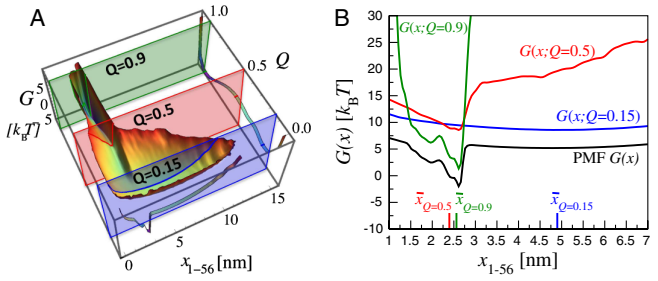

FIG. 4 (color online). Effective extension $\bar{x}_{1-56}$. (a) Slices through the 2D free-energy surface are shown in (b) as labeled, along with the 1D PMF on x. Values of $\bar{x}$ at the folded ($\bar{x}_{Q=0.9}$), unfolded ($\bar{x}_{Q=0.15}$), and transition ($\bar{x}_{Q=0.5}$) states are given by vertical bars at the bottom of the plot. The difference $\bar{x}_{Q=0.5} - \bar{x}_{Q=0.9} \approx -0.15$ nm is consistent with the $\Delta x^{\ddagger}$ from the fit to unfolding kinetics (see Table I).

isomorphic with that for dynamics on x [6] (see Supplemental Material [12]):

$$\frac{\tau(F)}{\tau(0)} = \frac{\int_{\cup} e^{-G(\bar{x})+F\bar{x}}d\bar{x} \int_{\cap} e^{G(\bar{x})-F\bar{x}}d\bar{x}}{\int_{\cup} e^{-G(\bar{x})}d\bar{x} \int_{\cap} e^{G(\bar{x})}d\bar{x}}. \quad (2)$$

At low force, $\bar{x} \to \langle x \rangle_Q$; therefore, the location of the barrier on $\bar{x}$ and the barrier height which captures the kinetics are those of the true barrier (i.e., that seen on the 2D surface) rather than the barrier on the formal 1D PMF on x. As illustrated in Fig. 4 for protein G pulled on x_{1-56}, the $\bar{x}$ computed from the slices through the 2D landscape at fixed Q corresponding to the unfolded, folded, and transition states are in very good accord with the $\Delta x^{\ddagger}$ from the fits of the rates to 1D theory. Landscape parameters on $\bar{x}$ agree well with those from the fit to 1D theory for the other systems; see Supplemental Material [12].

Fernandez and co-workers have highlighted a fundamental issue regarding 1D models which is not often considered: the 1D PMF for the pulling coordinate may have no refolding barrier in the absence of force [10]. However, the 1D PMF for the pulling coordinate lacks a folding barrier not because the barrier does not exist, but rather because it is hidden by the low-energy states when the potential is projected onto the pulling coordinate. As we have shown, the folding barrier probed under force is the same one present at zero force. Thus, the hopping observed in experiments represents true folding-unfolding and not merely barrierless polymer collapse. Note that we do not consider here any switch to an alternative barrier which may occur at a higher force [13,24].

What are the implications for interpreting experiments? If the transition state is easily stretched along a given coordinate (e.g., protein G, x_{10-32}), a small force will shift the barrier to coincide with that in the 1D PMF on x and a 1D model will be a good representation over a wide range of forces. In contrast, stiff transition states (e.g., protein G, x_{1-56}) will be hidden in the 1D PMF at low forces; nonetheless, fits of 1D theory to the kinetics at these forces yield the displacement and height of the true barrier—that captured by a good folding coordinate.

O. K. D. was supported by NSF CAREER Grant No. MCB-0845099 and NSF CTBP Grant No. PHY-0822283. T. W. G. W. was supported by the Winston Churchill Foundation and R. B. by the Royal Society.

*dudko@physics.ucsd.edu
†rbb24@cam.ac.uk

[1] A. Borgia, P. M. Williams, and J. Clarke, Annu. Rev. Biochem. **77**, 101 (2008).
[2] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **292**, 733 (2001).
[3] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, Science **314**, 1001 (2006).
[4] J. P. Junker, F. Ziegler, and M. Rief, Science **323**, 633 (2009).
[5] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante, Nature (London) **465**, 637 (2010).
[6] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[7] O. K. Dudko, G. Hummer, and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **105**, 15755 (2008).
[8] D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, CA, 2000), 1st ed.
[9] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[10] R. Berkovich, S. Garcia-Manyes, J. Klafter, M. Urbakh, and J. M. Fernandez, Biochem. Biophys. Res. Commun. **403**, 133 (2010).
[11] Y. Cao, R. Kuske, and H. Li, Biophys. J. **95**, 782 (2008).
[12] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.208301 for more details of the nucleic acid coarse-grained model, a fit of the unfolded state dimensions to the wormlike chain model, details of the Brownian dynamics simulations, detailed derivation of equations, and description of an alternative hairpin design
[13] T. G. W. Graham and R. B. Best, J. Phys. Chem. B **115**, 1546 (2011).
[14] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **96**, 2031 (1999).
[15] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2009).
[16] D. J. Bicout and A. Szabo, Protein Sci. **9**, 452 (2000).
[17] S. Kirmizialtin, L. Huang, and D. E. Makarov, J. Chem. Phys. **122**, 234915 (2005).
[18] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **102**, 6732 (2005).
[19] Y. Suzuki and O. K. Dudko, Phys. Rev. Lett. **104**, 048101 (2010).
[20] Y. Suzuki and O. K. Dudko, J. Chem. Phys. **134**, 065102 (2011).
[21] G. Morrison, C. Hyeon, M. Hinczewski, and D. Thirumalai, Phys. Rev. Lett. **106**, 138102 (2011).
[22] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **107**, 21441 (2010).
[23] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[24] R. B. Best, E. Paci, G. Hummer, and O. K. Dudko, J. Phys. Chem. B **112**, 5968 (2008).