Internal protein dynamics shifts the distance to the mechanical transition state

Daniel K. West
School of Physics and Astronomy and School of Biochemistry and Microbiology, University of Leeds, Leeds LS2 9JT, United Kingdom

Emanuele Paci and Peter D. Olmsted*
School of Physics and Astronomy and Astbury Centre for Structural Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

(Received 9 May 2006; published 29 December 2006)

Mechanical unfolding of polyproteins by force spectroscopy provides valuable insight into their free energy landscapes. Most experiments of the unfolding process have been fit to two-state and/or one dimensional models, with the details of the protein and its dynamics often subsumed into a zero-force unfolding rate and a distance x_u^{1D} to the transition state. We consider the entire phase space of a model protein under a constant force, and show that x_u^{1D} contains a sizeable contribution from exploring the full multidimensional energy landscape. This effect is greater for proteins with many degrees of freedom that are affected by force; and surprisingly, we predict that externally attached flexible linkers also contribute to the measured unfolding characteristics.

DOI: 10.1103/PhysRevE.74.061912 PACS number(s): 87.15.Aa, 82.37.Np, 87.15.He, 87.15.La

I. INTRODUCTION

Atomic force microscopy or optical tweezers are now routinely used to study the mechanical properties of proteins [1,2]. An important issue is the unfolding behavior of folded domains, including the strength, and the dependence on fold topology [3] and secondary structure [3,4]. The simplest description of unfolding treats the unfolding domain as moving in a one-dimensional (1D) potential $G(x)$, where the reaction coordinate x is assumed to be directly coupled to the applied force; the rate of unfolding $k_u(F)$ of a two-state (native and denatured) protein under force can then be calculated from $G(x)$ using transition state theory. The earliest approximate solution to this problem is due to Bell, based on Kramers' relation for escape from a well [1,5]:

$$k_u(F) \simeq k_u^0 \exp\left(\frac{Fx_u^{1D}}{k_B T}\right), \tag{1}$$

where k_B is Boltzmann's constant, T is the temperature, and x_u^{1D} is the transition state displacement parallel to the force. This has been used to characterize the distributions of unfolding times (for applied force) or forces (for applied pulling speed) for many proteins [6–10] (see Fig. 1).

Bell's solution has been repeatedly refined in the past decade: to use the entire shape of $G(x)$ instead of the barrier height and displacement x_u^{1D} and incorporate the cantilever compliance [11,12], to relax the diffusive limit to faster speeds [13], and to consider multiple pathways or states [14,15]. Moreover, the simple linear dependence Fx_u^{1D} only holds for a mathematically sharp transition state, and generally a more complex shape applies instead [16,17]. Another limitation is that a 1D potential $G(x)$ grossly simplifies physical reality. The unfolding rate depends dramatically on pulling direction [18,19], and hence on the multidimensional nature of the free energy landscape. Moreover, the 1D parameters have no satisfactory physical interpretation: x_u^{1D} is defined along the pulling direction, while unfolding initiates along an unknown reaction coordinate(s) whose connection with molecular configurations remains unclear.

In this paper we explore a specific consequence of the multidimensional energy landscape for the effective 1D unfolding parameters. Because the linear approximation Fx_u^{1D} in Eq. (1) is an excellent approximation for many experiments [6,10] we will follow this convention. The key physical ingredient is that an applied force alters a protein's conformational search among the dihedral states of the polypeptide backbone. For example, the force could reduce the fluctuations transverse to the forcing direction, leading to fewer pathways over the transition state and a smaller unfolding rate under an applied force. Using molecular dynamics (MD) simulations of a simple protein, we show that, in the main, fluctuations are indeed reduced by an applied force, which leads to a sizeable contribution to x_u^{1D}.

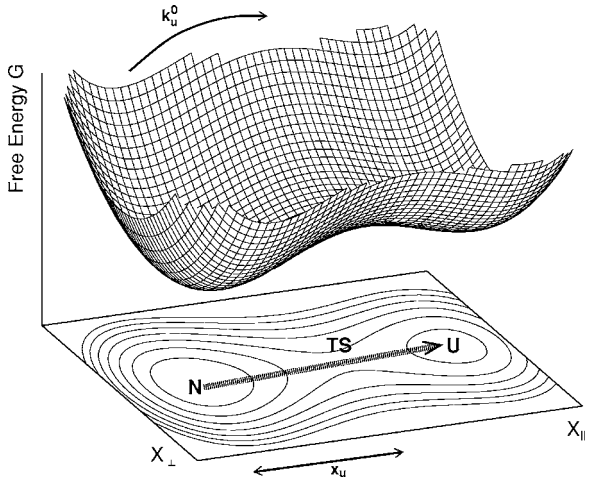

FIG. 1. The free energy surface of a two-state system, with native (N) and unfolded (U) minima. An external force F parallel to $X_\parallel$ lowers the barrier to unfolding by Fx_u.

*Electronic address: p.d.olmsted@leeds.ac.uk

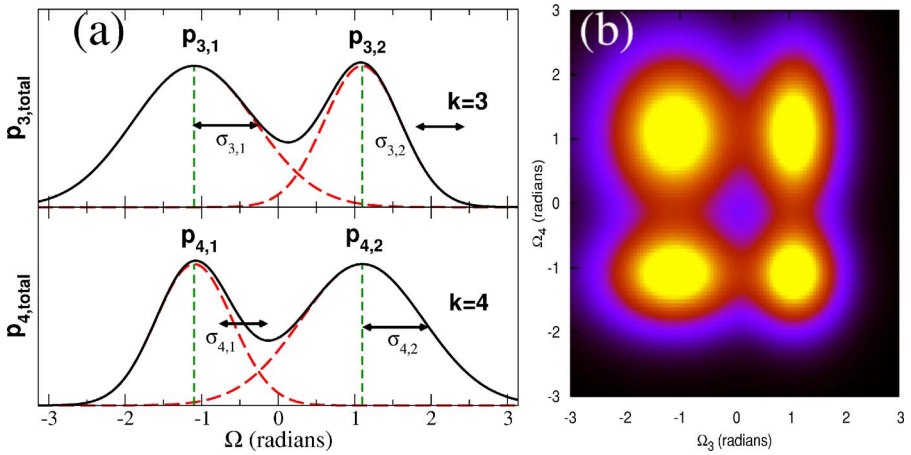

FIG. 2. (Color online) The phase space of an oligomer with four dihedral angles: two of these are unimodal ($k=1,2$, not shown) and two are bimodal ($k=3,4$). (a) The probability distribution function $p_{k,\mathrm{total}}$ for each bimodal dihedral angle (black solid line) can be resolved into separate distributions $p_{k,n}$ (red dashed lines) about well defined averages (green vertical dotted lines). There are four possible structures corresponding to fluctuations around $\{\bar{\Omega}_{3,1},\bar{\Omega}_{4,1}\}$, $\{\bar{\Omega}_{3,1},\bar{\Omega}_{4,2}\}$, $\{\bar{\Omega}_{3,2},\bar{\Omega}_{4,1}\}$, and $\{\bar{\Omega}_{3,2},\bar{\Omega}_{4,2}\}$. (b) The phase space projected onto $\{\Omega_3,\Omega_4\}$.

II. ESCAPE FROM A MULTIDIMENSIONAL ENERGY LANDSCAPE

The rate of escape from an $\mathcal{N}$-dimensional energy landscape under an applied force F is given in the Kramers approximation by $k_u(F)=\Gamma(F)e^{Fx_u/(k_BT)}$, where [5,20]

$$\Gamma(F)=\frac{\sqrt{|G''^{\,RC}_{TS}|}}{2\pi\gamma}\frac{\prod_{k=1}^{\mathcal{N}}\sqrt{G''^{\,k}_{N}(F)}}{\prod_{k=1}^{\mathcal{N}-1}\sqrt{G''^{\,k}_{TS}}}\exp\left(-\frac{\Delta G_{TS\text{-}N}}{k_BT}\right), \tag{2}$$

where x_u is the distance to the transition state, γ is a friction coefficient, and $\Delta G_{TS\text{-}N}$ is the height of the free energy barrier relative to the native basin. $G''^{\,RC}_{TS}$ is the curvature in the unstable direction at the transition state, $G''^{\,k}_{TS}$ are the $\mathcal{N}-1$ stable curvatures at the transition state, and $G''^{\,k}_{N}(F)$ are the $\mathcal{N}$ positive curvatures about the forced native basin. As in the 1D case we assume a very sharp transition state, i.e., $|G''^{\,RC}_{C}|\gg F/x_u$, so that x_u and the curvatures G''_{TS} at the transition state are approximately independent of force.

A physical representation of the attempt frequency $\Gamma(F)$ follows by relating G''^{k} to the associated fluctuations by $\langle\delta x_k^2\rangle=k_BT/G''^{k}$. This yields

$$\Gamma(F)=\frac{k_BT}{2\pi\gamma l^{RC}_{TS}}\frac{V_{TS}}{V_N(F)}\exp\left(\frac{-\Delta G_{TS\text{-}N}}{k_BT}\right), \tag{3}$$

where $l^{RC}_{TS}=\sqrt{k_BT/G''^{\,RC}_{TS}}$ is the width of the transition state along the unfolding reaction coordinate, and V_{TS} and $V_N(F)$ are the volumes of phase space available for fluctuations about the transition and force-perturbed native states, given by $V=\sqrt{\det\mathbf{C}}$, where $\mathbf{C}_{ij}=\langle\delta\mathbf{r}_i\delta\mathbf{r}_j\rangle$ is the covariance matrix for position fluctuations $\delta\mathbf{r}_i$.

In one dimension the weak force dependence of the prefactor $\Gamma(F)$ can be safely ignored [21]. A weak perturbation of the native basin volume $V_N(F)$ leads to

$$k_u(F)\simeq k_u^0\exp\left(\frac{F[x_u+\lambda k_BT]}{k_BT}\right), \tag{4}$$

where $\lambda=-\partial[\ln V(F)]/\partial F|_{F=0}$. This corresponds to a shift of x_u^{1D} in the equivalent 1D model,

$$x_u^{1D}=x_u+\lambda k_BT\equiv x_u+\delta x_u, \tag{5}$$

where δx_u is an *entropic*, or dynamic, contribution to the transition state displacement. If the volumes of perturbed degrees of freedom randomly increased or decreased with an applied force, there would be little effect. However, the dominant contribution to λ is actually a decrease due to freezing out unfavorable dihedral states under an applied force, so that $\delta x_u>0$ is extensive in the number of perturbed degrees of freedom.

III. CALCULATION OF PHASE SPACE VOLUMES

Phase space fluctuation volumes were calculated from MD simulation trajectories for protein L (PDB reference: 1HZ6 [22]) using the C_α Gō model of Ref. [23], which replaces each entire amino acid by a coarse-grained "atom" and uses potentials based on the native folded structure (the simulation protocol is described in Refs. [3,24]). A real protein does not fluctuate about a single well-defined average structure because dihedral angles typically access discrete values. Hence, the accessible phase space comprises *nodes* of fluctuations about many well defined structures. Figure 2 shows the phase space explored by a tetramer with two unimodal and two bimodal dihedral distributions.

The total unfolding rate $k_u^{\mathrm{tot}}(F)=\Gamma_{\mathrm{tot}}(F)e^{Fx_u/(k_BT)}$ is the weighted sum of the escape rates $k_u^{\beta}(F)$ from all nodes $\beta=1\cdots M$ (assuming $\Delta G_{TS\text{-}N}$ and x_u are the same for all nodes), leading to

$$\Gamma_{\mathrm{tot}}(F)=\frac{k_BT}{2\pi\gamma l^{RC}_{TS}}\frac{V_{TS}}{V_N^{\mathrm{eff}}(F)}\exp\left(-\frac{\Delta G_{TS\text{-}N}}{k_BT}\right), \tag{6}$$

$$\frac{1}{V_N^{\mathrm{eff}}(F)} = \sum_{\beta=1}^{M} \frac{P^\beta(F)}{V_N^\beta(F)}, \tag{7}$$

where $V_N^\beta(F)$ is the volume and $P^\beta(F)$ the occupation probability of node β. The quantity $V_N^{\mathrm{eff}}(F)$ is the effective phase space volume of the native basin.

The occupation probability $P^\beta(F)$ of each node is

$$P^\beta(F) = \left\langle \prod_{k=1}^{N-3} \frac{p_{k,n_\beta}(\Omega_k(t))}{p_{k,\mathrm{total}}(\Omega_k(t))} \right\rangle, \tag{8}$$

where $\Sigma_{\beta=1}^{M} P^\beta(F)=1$ and $\langle\cdots\rangle$ is the average over the MD trajectory. A node is specified by a particular set of occupancies of each dihedral angle $\mathbf{\Omega}=(\Omega_1,\ldots,\Omega_{N-3})$, where N is the number of atoms. Each term in the product is the normalized probability that, in node β, a given dihedral angle Ω_k participates in its nth dihedral state (peak) [Fig. 2(a)]. For many nodes $M \gg 1$, a mean field approach in which all nodes are assumed to be equally populated works well when states are sufficiently uncorrelated in time, as in this case [25].

We calculate the volume $V_N^\beta(F)=\sqrt{\det \mathbf{C}^\beta}$ of fluctuations about each node β by transforming coordinates to bond lengths and angles, and dihedral angles. We ignore correlations between bond and dihedral angles, which is an excellent approximation here [25]. Hence, the effective volume of phase space is given by

$$\frac{1}{V_N^{\mathrm{eff}}(F)} \simeq \frac{1}{V_\theta(F)} \sum_{\beta=1}^{M} \frac{P^\beta(F)}{V_\Omega^\beta(F)}. \tag{9}$$

where $V_\theta(F)$ and $V_\Omega^\beta(F)$ are the volumes of phase space explored by the dihedral and (unimodal) bond angles, respectively.

IV. RESULTS

Figure 3(a) shows the phase space volume as a function of force calculated from MD simulations of protein L. The entropic contribution to the transition state placement is $\delta x_u = 0.054 \pm 0.008$ nm. The reduction of phase space volume comes from the narrowing of the dihedral distributions and the reduction in the number of multimodal dihedral peaks. The latter effect dominates, since the loss of a single dihedral peak immediately removes many nodes of phase space. Simulations of the same protein L domain [Fig. 3(b)] yield an effective 1D transition displacement $x_u^{1D}=0.191\pm0.004$ nm, from measuring an exponential dependence of the unfolding time τ_u on applied force, $\tau_u \sim e^{-Fx_u^{1D}/(k_BT)}$, as predicted by Eq. (1). Hence we conclude that the bare transition state position was $x_u \simeq 0.137$ nm, and the large shift of $\delta x_u = 0.054 \pm 0.008$ nm is between 34 and 45 %.

V. LINKER EFFECTS

For convenience, protein domains are often pulled with long linkers, or unfolded protein strands. The linkers *also* fluctuate about discrete dihedral states when stretched. The

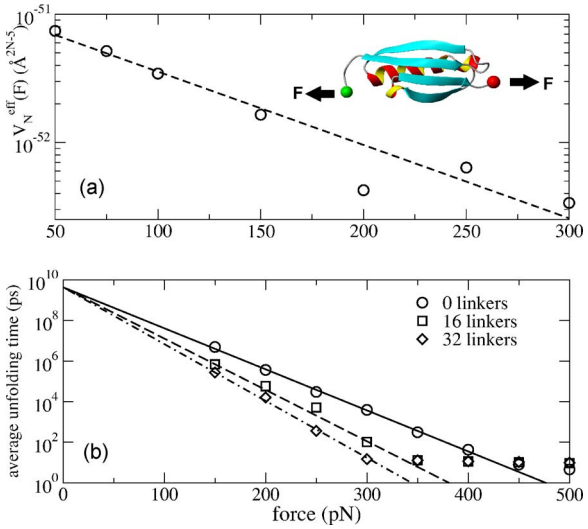

FIG. 3. (Color online) (a) Effective phase space volume $V_N^{\mathrm{eff}}(F)$ for protein L(○). The dashed line is a fit to $\ln V_N^{\mathrm{eff}}(F) \simeq \ln V_N^{\mathrm{eff}}(0) - \delta x_u F/k_BT$, which yields the entropic contribution to the transition state placement $\delta x_u = 0.054 \pm 0.008$ nm [Eq. (5)]. (b) Average unfolding times for protein L using MD at $T=300$ K, with n_l attached glycine linkers (○: $n_l=0$, $x_u^{1D}=0.191\pm0.004$ nm, □: $n_l=16$, $x_u^{1D}=0.241\pm0.004$ nm, ◇: $n_l=32$, $x_u^{1D}=0.267\pm0.004$ nm). The linear fits yield $\ln \tau = A - x_u^{1D}F/(k_BT)$. Error bars are of order the symbol size.

"lumpiness" of this phase space is irrelevant for weakly stretched strands, but dominates the response for strongly stretched strands. Since force is coupled to the folded domain through the linkers, the total available phase space is the product of protein and linker phase spaces, and the measured x_u^{1D} depends on the restriction of the linkers' phase space. To test this, linkers were constructed from a dihedral potential based on glycine. Figure 4 shows the normalized effective phase space volume $V_N^{\mathrm{eff}}(F)/V_N^{\mathrm{eff}}(0)$ and the corresponding δx_u as a function of force for different number of

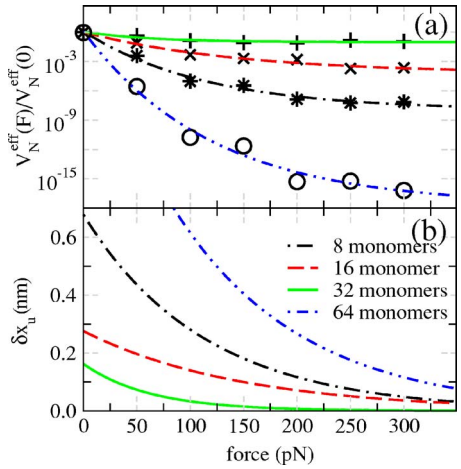

FIG. 4. (Color online) (a) Volume of phase space calculated using the mean field method and (b) the entropic contribution to the distance to the transition state $\delta x_u = -k_BT\partial[\ln V_N^{\mathrm{eff}}]/\partial F$, for strands of flexible glycine linkers. A constant force was applied for a total time of 1 μs at $T=300$ K.

atoms n_l per linker. The effect is greater for longer linkers, since more nodes are available to remove.

VI. APPROXIMATIONS

We have ignored the force dependence of the volume of the transition state V_{TS} [Eqs. (2), (3)], which is acceptable for very sharp or small (comprising few nodes) transition states. In principle, the native and transition states should be analyzed in exactly the same way. However, the linker calculations provide an opportunity to evaluate this approximation. We can compare the shifts δx_u measured directly from the MD unfolding times of proteins with different linker lengths (Fig. 3) with predictions from the phase space volumes (Fig. 4). The difference $\delta x_u(n_l=32)-\delta x_u(n_l=16)\simeq 0.03-0.05$ nm from the calculation of phase space volumes (at forces of order 200–300 pN) agrees surprisingly well with the difference $x_u(n_l=32)-x_u(n_l=16)=0.026$ nm measured from pulling simulations, suggesting that in this case the transition state is sharp and its volume does not change appreciably under an applied force. We have also assumed that all nodes are separated from the transition state by the same distance x_u and energy $\Delta G_{TS\text{-}N}$. This should be relaxed in a more detailed calculation, and unfolding rates averaged over this more complicated landscape; this could lead to more complex kinetics [26]. Finally, our calculation rests on the Kramers solution in the diffusive limit; obviously the entire energy landscape is important [13], but this does not change the primary message that the degeneracy of the native state contributes significantly to the unfolding rate.

VII. DISCUSSION

We have studied the phenomenological parameter x_u^{1D}, which is easily extracted experimentally according to its interpretation as a transition state displacement with dimensions of length. This interpretation is misleading, and we have shown that an applied force restricts a fluctuating protein's accessible phase space, which increases x_u^{1D} in the equivalent 1D two-state model. Larger proteins have a potentially larger x_u^{1D}, and proteins whose fold geometries couple to many degrees of freedom to a particular applied force also have a potentially larger x_u^{1D}. Most importantly, x_u^{1D} should be greater for proteins unfolded through longer attached linker strands. This may have biological significance, e.g., the long unfolded PEVK regions in titin [27] may play help modify the unfolding characteristics of titin. Finally, we note that many experiments have unfolded concatamers of multiple domains, for convenience of attachment and to generate larger statistics [9,10,15,18,19,28–31]. We surmise that in all of these cases the entropic contribution to x_u^{1D} typically included a contribution from already unfolded domains, which acted as "linkers" for the remaining domains to unfold.

ACKNOWLEDGMENTS

D.K.W. acknowledges the Wellcome Trust for financial support. We thank D.J. Brockwell, J. Clarke, T. McLeish, and S.E. Radford for helpful discussions.

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).
[2] M. Rief, J. M. Fernandez, and H. E. Gaub, Phys. Rev. Lett. **81**, 4764 (1998).
[3] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2006).
[4] V. Ortiz, S. O. Nielsen, M. L. Klein, and D. E. Discher, J. Mol. Biol. **349**, 638 (2005).
[5] G. I. Bell, Science **200**, 618 (1978).
[6] J. M. Fernandez and H. Li, Science **303**, 1674 (2004).
[7] M. Schlierf, H. Li, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **101**, 7299 (2004).
[8] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **98**, 468 (2001).
[9] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford, Biophys. J. **89**, 506 (2005).
[10] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, and S. E. Radford, Biophys. J. **83**, 458 (2002).
[11] M. Schlierf and M. Rief, Biophys. J. **90**, L33 (2006).
[12] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).
[13] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[14] D. Bartolo, I. Derényi, and A. Ajdari, Phys. Rev. E **65**, 051910 (2002).
[15] P. M. Williams, S. B. Fowler, R. B. Best, J. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke, Nature (London) **422**, 446 (2003).
[16] P. C. Li and D. E. Makarov, J. Chem. Phys. **121**, 4826 (2004).
[17] S. Kirmizialtin, L. Huang, and D. E. Makarov, J. Chem. Phys. **122**, 234915 (2005).
[18] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, Nat. Struct. Biol. **10**, 738 (2003).
[19] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford, Nat. Struct. Biol. **10**, 731 (2003).
[20] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[21] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[22] J. W. O'Neill, D. E. Kim, D. Baker, and K. Y. Zhang, Acta Crystallogr., Sect. D: Biol. Crystallogr. **57**, 480 (2001).
[23] J. Karanicolas and C. L. Brooks III, Protein Sci. **11**, 2351 (2002); J. Mol. Biol. **334**, 309 (2003).
[24] D. K. West, P. D. Olmsted, and E. Paci, J. Chem. Phys. **124**, 154909 (2006).
[25] D. K. West, E. Paci, and P. D. Olmsted (unpublished).
[26] J. Brujíc, R. I. Hermans, K. A. Walther, and J. M. Fernandez, Nat. Phys. **2**, 282 (2006).
[27] W. A. Linke, M. Ivemeyer, P. Mundel, M. R. Stockmeier, and

B. Kolmerer, Proc. Natl. Acad. Sci. U.S.A. **95**, 8052 (1998).
[28] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub, J. Mol. Biol. **286**, 553 (1999).
[29] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke, Biophys. J. **81**, 2344 (2001).
[30] M. Carrion-Vasquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).
[31] R. B. Best, S. Fowler, J. L. Toca-Herrera, A. Steward, E. Paci, and J. Clarke, J. Mol. Biol. **330**, 867 (2003).