New Dynamical Window onto the Landscape for Forced Protein Unfolding

Zu Thur Yew, Tom McLeish, and Emanuele Paci*

Institute of Molecular and Cell Biology, Astbury Centre for Structural Molecular Biology Physics & Astronomy, University of Leeds, Leeds LS2 9JT, United Kingdom

(Received 4 June 2008; published 12 December 2008)

The unfolding of a protein by the application of an external force pulling two atoms of the protein can be detected by atomic force and optical tweezers technologies as have been broadly demonstrated in the past decade. Variation of the applied force results in a modulation of the free-energy barrier to unfolding and thus, the rate of the process, which is often assumed to have single exponential kinetics. It has been recently shown that it is experimentally feasible, through the use of force clamps, to estimate the distribution of unfolding times for a population of proteins initially in the native state. In this Letter we show how the analysis of such distributions under a range of forces can provide unique information about the underlying free-energy surface such as the height of the free-energy barrier, the preexponential factor and the force dependence of the unfolding kinetics without resorting to *ad hoc* kinetic models.

DOI: 10.1103/PhysRevLett.101.248104 PACS numbers: 87.15.Cc, 82.37.Np, 82.37.Rs, 87.15.ap

Single molecule protein unfolding studies are shedding light on the free-energy landscape underlying the macroscopic properties of proteins. In particular, atomic force microscopy is providing a picture of the free-energy landscape at an unprecedented level of detail, revealing the presence of distinct states and parallel unfolding routes [1,2], the barriers and free-energies between states [3] and in some instances, the secondary structure elements that make up the structured core of unfolding intermediates [4,5]. Traditionally, the analysis of force spectroscopy experiments have relied on a phenomenological model due to Bell [6,7], where the mean unfolding time and the maximum unfolding force is predicted to scale exponentially and logarithmically with the force and pulling speed, respectively. It has been recently shown [8–10] that even though Bell's formalism works well for average quantities (e.g., maximum force or mean unfolding time), it fails to reproduce the variance and the distribution of those quantities. Furthermore, Bell's model assumes that (i) the integration over all other degrees of freedom other than the reaction coordinate, x, induces a one-dimensional free-energy profile, $G(x)$, that accurately describes the transition, (ii) the ''distance'' between the native state and the transition state along x is invariant with the force and (iii) the kinetics is single exponential. To this end, Dudko, Hummer, and Szabo [9] proposed an alternative formalism (DHS) that corrects for assumption (ii) and found that their theory is able to account for the variances in the maximum unfolding force at different pulling velocities. Intriguingly, in contrast to Bell, DHS also appears to allow the estimation of the naked barrier height without recourse to arbitrary values of the preexponential factor. However, like Bell, DHS also assume that conditions (i) and (iii) hold.

Below, we show that the distribution of unfolding times not only imposes more severe constraints on a ''model'' (e.g., Bell or DHS), but also contains unprecedented information regarding the free-energy landscape of the protein, amplifying the importance of such single molecule experiments. Remarkably, we observe that at high forces (i.e., low effective barrier), the distribution of unfolding times is not single exponential. An analogous deviation from the expected single exponential kinetics has been recently observed experimentally [11] although the explanation of the deviation from single exponential kinetics was quite different from the one we present below. In addition, by solving the diffusion equation for the probability density, one can fit the distribution of unfolding times and not only extract a *model-free* estimate of the preexponential factor, the activated time for the unfolding and the force dependence of the unfolding kinetics, but also an independent estimate of x_u, which in the mechanical unfolding literature, is commonly taken to be the distance to the transition state. We note that this interpretation of x_u is only valid in the one-dimensional case and that x_u can be understood in more general terms in the context of a Taylor's expansion of the free-energy around the barrier (see below).

To generate the distributions, we used Langevin dynamics simulations with a coarse-grained (only C_α atoms are represented) native-centric protein model [12]. While it is arguable whether such models are realistic enough to describe how a protein collapses and folds [13–15], they are accurate in predicting the mechanical properties of proteins [16,17].

Specifically, we study the mechanically resistant protein, ubiquitin, which has a well-defined transition state both in simulations [16,18] and experiments [19]. We first characterized the protein by performing simulations over a broad range of temperatures. At 300 K, the native state is very stable and no unfolding event was recorded during a 1.25 μs simulation. Along this simulation, 5000 phase points (coordinates and velocities) were extracted as initial

conditions for constant force pulling simulations. Applying a force of 250 pN to the two ends of the polypeptide chain, ubiquitin unfolds in about 0.7 ns on average over 5000 independent simulations. (Unfolding is defined as the protein reaching an extension of 100 Å, which is about 40% the maximal length. Unfolding occurs in a highly cooperative manner at an extension of ~42 Å even at the largest forces probed in this work and as such, the time distributions do not depend on the exact definition of an unfolding event). At large enough forces, the cumulative probability of unfolding times [Fig. 1(a)] is clearly not single exponential, but can be well described by a double exponential.

A simple explanation of the nonexponentiality at high forces arises from assumption (iii) used in deriving one-dimensional landscape models of protein unfolding (see above), which is equivalent to taking only the eigenfunction belonging to the lowest eigenvalue of the diffusion equation for the probability density $P(x, t)$ (where $\beta = 1/k_B T$):

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left(D(x)\left(\frac{\partial P}{\partial x} + \beta P \frac{\partial G}{\partial x}\right)\right). \quad (1)$$

This is not exact because there is no guarantee that the form of the first eigenfunction is identical to the equilibrium population of the denatured state, which itself forms the initial condition for the evolution of Eq. (1). The single exponential approximation becomes exact only in the limit of large barriers (i.e., $\gg 1 k_B T$). Away from this limit, the next most important term in the series become non-negligible. The structure of the first two terms of this series is [20,21]

$$P(t) \simeq 1 - A_1 e^{-t/\tau_1} - A_2 e^{-t/\tau_2} + \ldots, \quad (2)$$

where $\tau_1 \sim \tau_0 e^{\beta \Delta \mathcal{G}_u}$ (τ_0 is the usual Kramers prefactor)

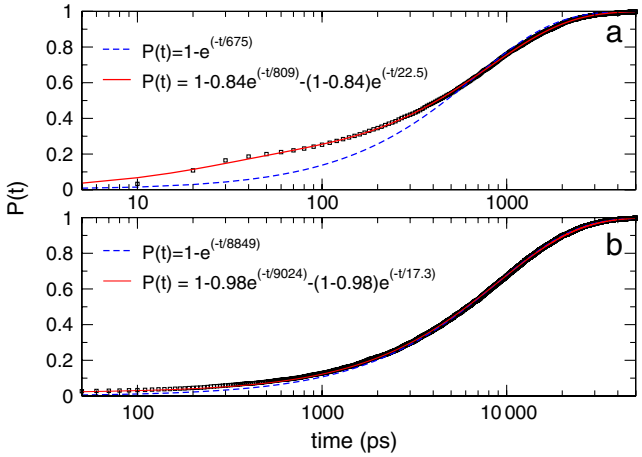

FIG. 1 (color online). Cumulative probability of unfolding times at two different forces and 300 K. (a) A double exponential is required to fit the curve at $F = 250$ pN. (b) At a lower force ($F = 200$ pN) the free-energy barrier is larger and the distribution of unfolding times is closer to single exponential.

and τ_2 and time scales of all higher terms are of the order of τ_0 (i.e., do not suffer the Boltzmann penalty). Higher-order terms, A_n, in Eq. (2) arise from higher eigenfunctions of the Smoluchowski equation [Eq. (1)] of index $n = 2, 3, 4, \ldots$. From the continuity of the density function, $P(x, t)$, these terms have amplitudes $A_n \simeq 1/n^2$ which results in the weakly perturbed region being dominated by A_2.

Crucially, the amplitude of the first term (A_1) is generically dominating [22] by a factor proportional to the exponential of the effective barrier height, i.e.,

$$\frac{A_1}{A_2} = \gamma e^{\beta \Delta \mathcal{G}_u}, \quad (3)$$

where γ is a number of order one that depends on the shape of the free-energy landscape $G(x)$.

In the presence of a force pulling the two ends of the protein apart, the height of the free-energy barrier becomes a decreasing function of the force.

The broadly used "Bell" expression

$$\Delta \mathcal{G}_u = \Delta G_u - F x_u \quad (4)$$

constitutes the leading term only; here x_u is the distance along the one-dimensional reaction coordinate to the free-energy barrier and $\Delta G_u = \Delta G(x_u)$. Even in the case where Eq. (4) holds (i.e., for $\Delta G_u \to \infty$, $F \to 0$), when a large number of degrees of freedom are involved, the coefficient x_u contains entropic contributions as well [23]; furthermore, high-dimensional folding landscapes strongly compromise the conclusions of one-dimensional analyses of folding times [22,24], making interpretations of x_u as a physical distance perilous. However, the validity of Eq. (3) does not depend on the validity of Eq. (4). The latter can still be assumed in a small range of forces where $\Delta \mathcal{G}_u$ can be approximated as a linear function of F, which leads to unambiguous values of x_u.

It should be noted that γ in Eq. (3) is computable if the form of the potential is known. For example, in the flat hypersphere case [22] treated by Bicout and Szabo, $\gamma = 9\pi^2/24 \simeq 3.7$. If γ is independent of the force, the value of the coefficient, x_u, extracted from Eq. (3) is independent of the exact value of γ. This is in strong contrast with models that attempt to extract x_u from the force dependence of the mean unfolding time alone. We will show below that the assumption that γ is force independent is consistent with our simulations.

By performing the simulations in a broad range of forces, we evaluated the distribution of unfolding times at each force and extracted the 2 times, τ_1 and τ_2, and the amplitudes, A_1 and A_2.

From the double exponential fits of the distribution of unfolding times of our ubiquitin model, τ_2 is approximately constant with an average of 18.9 ps [Fig. 2(a)]. This value is lower than the ~10 ns rate of looping in peptides [25] or the ~1 μs folding rate of the fastest proteins [26] that are often taken as rough estimations of

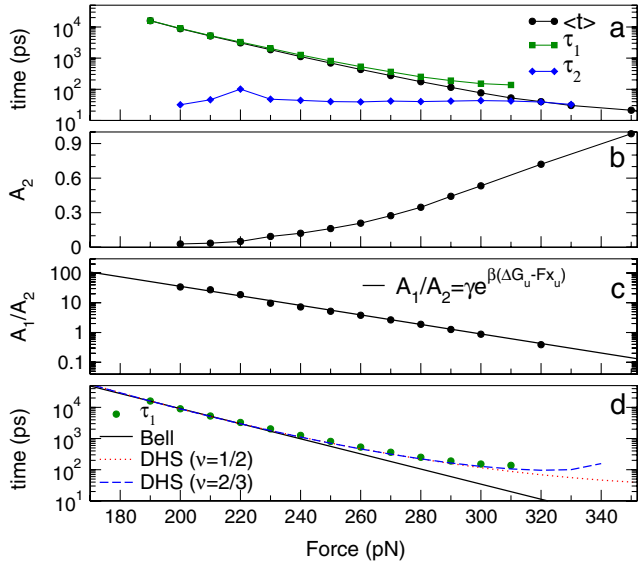

FIG. 2 (color online). (a) Average unfolding time $\langle t \rangle$ and fitting parameters from the double-exponential fit of the time distribution τ_1 and τ_2 ($\tau_1 + \tau_2 \simeq \langle t \rangle$). (b) Amplitude A_2 of the leading term of the expansion in Eq. (2); A_2 tends to 0 (i.e., $P(t)$ can be approximated with a single exponential) for $F \leq 200$ pN. (c) Fit of A_1/A_2 to Eq. (3). (d) Activated time τ_1 fitted using Bell's relation (black line), DHS with $\nu = 1/2$ (red line) and $\nu = 2/3$ (blue line); data points for $F > 300$ pN have been disregarded in the fits.

the preexponential factor. There are two reasons for this: one is the low external friction at which the simulations have been performed (0.1 $ps^{-1} \sim 1000$ times lower than water) and the other is that the "internal friction" of the model is lower than that of real proteins due to the smoother free-energy surface of a structure-based model designed to be "minimally frustrated."

Decreasing the applied force from 250 to 200 pN results in the average unfolding time increasing to 8.5 ns and the distribution of unfolding times becoming single exponential [Fig. 1(b)]: the effective barrier becomes high enough that the kinetics are indistinguishable from single exponential. In this case, the amplitude of the second term A_2 [Fig. 2(b)] decreases to zero at low forces as expected.

The logarithm of A_1/A_2 [Fig. 2(c)] is perfectly linear, compatible with the assumption that γ and x_u are independent of the force. Fitting the ratio of the two amplitudes to Eq. (3), the slope of $\ln(A_1/A_2)$ uniquely determines $x_u = 1.5$ Å, which is independent on the shape of the underlying free-energy profile. On the other hand, ΔG_u and γ are related by $\beta \Delta G_u + \ln(\gamma) = 10.9$: if γ is that of a Bicout-Szabo barrier ($\gamma \simeq 3.7$), we obtain $\Delta G_u = 5.8$ kcal/mol; if γ varies by 1 order of magnitude, the estimation of ΔG_u only varies by 1.4 kcal/mol.

To compare the values of ΔG_u, x_u and τ_2 obtained with Eq. (2) and (3), we independently determined the parameters by fitting τ_1 to the DHS equation [9].

$$\tau(F) = \tau(0)\left(1 - \frac{\nu F x_u}{\Delta G_u}\right)^{(\nu-1)/\nu} e^{-\beta \Delta G_u \{1-[1-(\nu F x_u)/\Delta G_u]^{1/\nu}\}}, \quad (5)$$

where $\nu = 1/2$ corresponds to the cusp case, $\nu = 2/3$ to the linear-cubic case and the Bell form is recovered for either $\nu = 1$ or $\Delta G_u \to \infty$. The fits of the τ_1 using the Eq. (5) with both $\nu = 1/2$ and $2/3$ are shown in Fig. 2(d). From the fit of the the activated time, τ_1, we obtained $x_u = 2.3$ Å ($\nu = 1$), $x_u = 3.9$ Å ($\nu = 2/3$), $x_u = 4.6$ Å ($\nu = 1/2$), and $\Delta G_u = 14.8$ kcal/mol ($\nu = 2/3$) and $\Delta G_u = 12.8$ kcal/mol ($\nu = 1/2$). These values of x_u and ΔG_u however, are those for zero force and cannot be directly compared with those obtained from Eqs. (3) and (4). Given a linear-cubic free-energy profile ($\nu = 2/3$) with $\Delta G_u = 14.8$ kcal/mol and $x_u = 3.9$ Å as estimated above, ΔG_u and x_u drop to 10–13 kcal/mol and 1.9–2.7 Å, respectively, when the applied force is 200–300 pN. Such values are slightly larger than those obtained using our "model-free" approach in the same range of forces. However, the values estimated from DHS depend on the assumed one dimensionality and shape of the free-energy profile (which is a central result of Ref. [9]), while those from the analysis of the distributions of unfolding times do not. Indeed, we have checked by performing Brownian dynamics on a one-dimensional linear-cubic free-energy profile that in the range of forces where our approach [i.e., Eqs. (3)] is applicable (i.e., the kinetics are double exponential), it gives estimates of ΔG_u and x_u identical to directly fitting ΔG_u of the linear-cubic profile to Eq. (4). Therefore, both the atomistic model of ubiquitin and 1D Brownian dynamics simulations corroborate the validity of Eq. (3). The similarity of the estimates of ΔG_u and x_u to the "true" values at zero force however, depends on the form of ΔG_u, which for real proteins, is not only likely to show significant deviations from the simple shapes analyzed here (i.e., linear cubic), but is also likely to be multidimensional.

In conclusion, we demonstrate that the distribution of unfolding times in a range of forces provides crucial information not obtainable from average times: an independent estimate of both the preexponential factor (i.e., τ_2) and x_u. Remarkably, both the preexponential factor $\tau_0 \simeq \tau_2$ and x_u obtained from the analysis of the time distributions do not depend on a particular model (i.e., a form for the underlying free-energy profile) unlike all previously proposed models. It should be noted that it is not possible to obtain an estimate of the preexponential factor or ΔG with the broadly used Bell model. Conversely, the determination of ΔG_u will depend on the value of γ, which is model dependent; however, the dependence on the model is much weaker than that observed within the DHS framework. Lastly, we show that a deviation from single exponential kinetics can be simply explained by a low barrier, which is a well-known result (see, e.g., Ref. [27]). Thus, as fast-improving atomic force instruments make the deter-

mination of the distribution of unfolding times [11,28] at the single molecule level more routine, the approach proposed here can provide model-free estimates of the relevant parameters that characterize the response of proteins to a force. It also allows for the verification of hypotheses and models that are often used by default without critical assessment of the meaningfulness of the parameters provided.

We acknowledge stimulating discussions with Attila Szabo and Olga Dudko. Z. T. Y. acknowledges the Wellcome Trust and the University of Leeds for financial support.

*e.paci@leeds.ac.uk

[1] M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief, Proc. Natl. Acad. Sci. U.S.A. **104**, 20 268 (2007).
[2] D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu, and H. Li, Proc. Natl. Acad. Sci. U.S.A. **104**, 9278 (2007).
[3] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee, Science **309**, 2057 (2005).
[4] I. Schwaiger, M. Schleicher, A. A. Noegel, and M. Rief, EMBO Rep. **6**, 46 (2005).
[5] M. Bertz and M. Rief, J. Mol. Biol. **378**, 447 (2008).
[6] G. I. Bell, Science **200**, 618 (1978).
[7] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[8] O. K. Dudko, A. E. Filippov, J. Klafter, and M. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11 378 (2003).
[9] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[10] M. Schlierf and M. Rief, Biophys. J. **90**, L33 (2005).
[11] J. Brujic, R. I. Hermans, K. A. Walther, and J. M. Fernandez, Nature Phys. **2**, 282 (2006).
[12] J. Karanicolas and C. L. I. Brooks, Protein Sci. **11**, 2351 (2002).
[13] E. Paci, M. Vendruscolo, and M. Karplus, Proteins: Struct. Funct. Genet. **47**, 379 (2002).
[14] E. Paci, M. Vendruscolo, and M. Karplus, Biophys. J. **83**, 3032 (2002).
[15] A. Cavalli, M. Vendruscolo, and E. Paci, Biophys. J. **88**, 3158 (2005).
[16] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2005).
[17] J. I. Sulkowska and M. Cieplak, J. Phys. Condens. Matter **19**, 283201, (2007).
[18] M. Cieplak and P. E. Marszalek, J. Chem. Phys. **123**, 194903 (2005).
[19] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, Nat. Struct. Biol. **10**, 738 (2003).
[20] T. C. B. McLeish, J. Phys. Condens. Matter **18**, 1861 (2006).
[21] H. L. Frisch, V. Privman, C. Nicolis, and G. Nicolis, J. Phys. A **23**, L1147 (1990).
[22] D. J. Bicout and A. Szabo, Protein Sci. **9**, 452 (2000).
[23] D. K. West, E. Paci, and P. D. Olmsted, Phys. Rev. E **74**, 061912 (2006).
[24] T. C. McLeish, Biophys. J. **88**, 172 (2004).
[25] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, Proc. Natl. Acad. Sci. U.S.A. **96**, 9597 (1999).
[26] J. Kubelka, J. Hofrichter, and W. A. Eaton, Curr. Opin. Struct. Biol. **14**, 76 (2004).
[27] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1981).
[28] Y. Cao, R. Kuske, and H. Li, Biophys. J. **95**, 782 (2008).