APS physics

CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Force-Clamp Experiments Reveal the Free-Energy Profile and Diffusion Coefficient of the Collapse of Protein Molecules

H. Lannon, J. S. Haghpanah, J. K. Montclare, E. Vanden-Eijnden, and J. Brujic

Phys. Rev. Lett. **110**, 128301 — Published 22 March 2013

DOI: 10.1103/PhysRevLett.110.128301

Force-clamp experiments reveal the free energy profile and diffusion coefficient of the collapse of proteins

H. Lannon,[1] J.S. Haghpanah,[2] J.K. Montclare,[2] E. Vanden-Eijnden,[3,*] and J. Brujic[1,†]

[1] *Department of Physics and Center for Soft Matter Research, New York University, 4 Washington Place, New York, NY, 10003, USA*

[2] *Department of Chemical and Biological Sciences, Polytechnic Institute of NYU, Brooklyn, New York 11201, USA*

[3] *Courant Institute of Mathematical Sciences, New York University, 251 Mercer St, New York, NY, 10012, USA*

We present force-clamp data on the collapse of ubiquitin polyproteins from a highly extended state to the folded length, in response to a quench in the force from 110pN to 5 or 10pN. Using a recent method for free energy reconstruction from the observed nonequilibrium trajectories, we find that their statistics is captured by simple diffusion along the end-to-end length. The estimated diffusion coefficient of $\sim$100nm^{2}s^{-1} is significantly slower than expected from viscous effects alone, possibly because of the internal degrees of freedom of the protein. The free energy profiles give validity to a physical model in which the multiple protein domains collapse all at once and the role of the force is approximately captured by the Bell model.

By measuring the end-to-end length of proteins and RNA in response to force perturbations, single molecule experiments open a window into the complex dynamics of these molecules on their multi-dimensional energy potentials [1–3]. For example, a protein is unfolded by the application of a constant pulling force, while quenching the force to a low value triggers the collapse of the molecule [1, 4]. This dynamical collapse has been modeled as a one-dimensional diffusion of the measured end-to-end length on a free energy profile in the case of protein monomers [5] and RNA molecules [6]. By contrast, dynamics in degrees of freedom hidden from the experiment were thought to govern the large diversity in the end-to-end length of trajectories visited by collapsing polyproteins [7]. Whether the trajectories can be described by simple diffusion along the measured reaction coordinate or require multiple dimensions remains an open question that requires novel analysis tools.

This question is non-trivial because the collapsing traces are out-of-equilibrium and standard techniques to estimate the free energy profile from the available data based on the Jarzynski equality [8] or Crook's fluctuation theorem [9] are not applicable to the force quench experimental protocol. In force-clamp experiments the work exerted on the system is concentrated in the brief time it takes to quench the force ($\sim$50ms), which would require a prohibitively large pool of data to ensure the statistical accuracy of the free energy estimator based on the Jarzynski equality [10–13]. A second difficulty is that the free energy alone is not sufficient to describe the dynamics of the collapse. If the collapse can be described by an overdamped Langevin equation for the end-to-end length of the protein [5], then a diffusion coefficient must be estimated besides the free energy [14, 15].

Here we introduce an analysis method to reconstruct the free energy profile [16] directly from the collapse trajectories of ubiquitin polyproteins. This procedure tests whether diffusive dynamics along the measured reaction coordinate of length is sufficient to describe the data. By varying the number of protein domains in the chain, we quantify the extent to which these domains are independent of one another [17]. Moreover, we vary the applied quench force to measure its effect on the shape of the free energy landscape. Furthermore, we measure the diffusion coefficient and test its constancy along the measured reaction coordinate. Finally, we interpret the resulting free energy landscapes in terms of polymer physics models [6] to give a microscopic mechanism for the collapse.

We use Atomic Force Microscopy (AFM) in the force-clamp mode to monitor the unfolding and refolding trajectories of ubiquitin polyproteins under a constant stretching force, following the same experimental procedures as those described in [18] and in Sec. 1 of the Supplementary Materials (SM). The experiment consists of exposing a mechanically stable polyprotein to a high pulling force of 110pN, which leads to the stepwise unfolding and extension of each of the protein domains in the polypeptide chain shown in the example in Fig. 1. Subsequently, quenching the force to a lower value of 10pN triggers the collapse of the whole polyprotein from a fully extended state back to a collapsed state with the same end-to-end length as the folded protein. The question is then to understand the mechanism of the collapse dynamics from recordings of these trajectories.

Theoretically, if we denote by x the end-to-end length, the overdamped Langevin equation reads

$$\dot{x} = -\beta D G'(x) + \sqrt{2D}\,\eta(t) \tag{1}$$

where $\beta = 1/(k_B T)$, $\eta(t)$ is a white-noise term accounting for thermal effects, $G(x)$ is the equilibrium free energy profile and D is the diffusion coefficient which we assume to be constant (this assumption is validated below). Both $G(x)$ and D can be estimated from the collapsing traces using the techniques introduced in [16]. The procedure to calculate the free energy $G(x)$ from the collapsing traces is to cut out pieces of trajectories from the moment

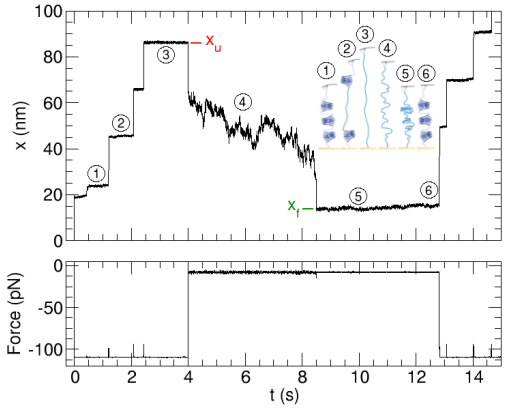

FIG. 1. A typical force-clamp trajectory of the unfolding and refolding of a polyubiquitin chain with $N_d = 3$ domains. The circled numbers indicate the various stages of the pulling experiment, with a schematic representation of the protein conformation in each stage shown in the inset. Here we are interested in the collapse from stage 3 to 5: for details on the other stages, see Sec. 1 of the SM.

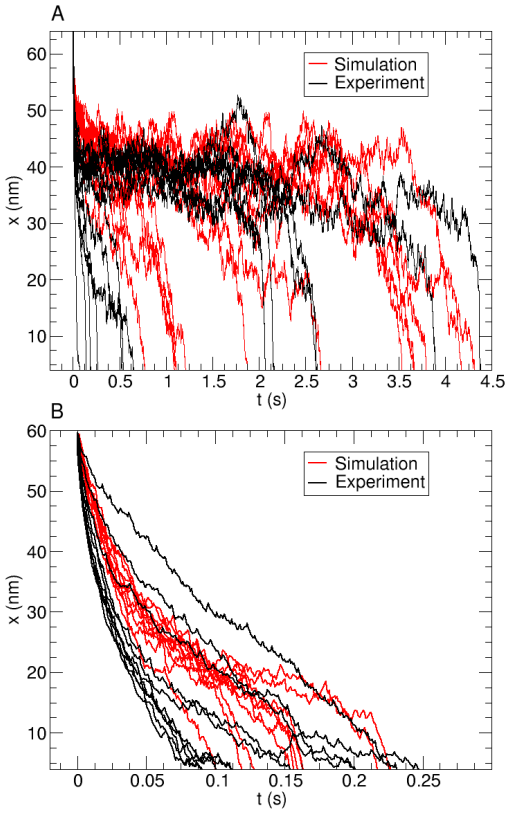

FIG. 2. Collapsing trajectories are grouped by their total length ($N_d = 3$) and recorded from the time of the force quench from 110pN to 10pN in (A) and 5p N in (B). The experimental trajectories are compared with those generated by simulations of diffusive dynamics on the reconstructed free energy profiles.

the force is quenched at the unfolded length, x_u, until the moment they first reach the folded length at low force, x_f. These pieces are out of equilibrium and sample a stationary probability density $\rho(x)$ which is different from the equilibrium density $\rho_e(x) = C^{-1}e^{-\beta G(x)}$ and biased towards values of x that are closer to x_u, where we begin to record the trajectories directly after the quench of the force. Yet as shown in [16] (see also Sec. 2 of the SM for a derivation), this bias can be removed and the free energy $G(x)$ can be related to the density $\rho(x)$ as

$$G(x) = -k_B T \ln \rho(x) - k_B T \rho'(x_f) \int_x^{x_u} dx'/\rho(x') \quad (2)$$

where $\rho'(x_f)$ denotes the derivative of $\rho(x)$ estimated at x_f. The additional term besides $-k_B T \ln \rho(x)$ in (2) corrects for the nonequilibrium bias in $\rho(x)$. Next we apply (2) to analyze force-clamp trajectories, such as the one shown in Figure 1. Since the length of the polyprotein chain and the polypeptide linker to the surface vary from one experiment to the next, we compare all trajectories in terms of the total length of the collapse $L_{tot} = x_u - x_f$. We find that L_{tot} clusters in increments of a monomer ubiquitin extension of $\sim$ 20nm with a standard deviation of $\sim$ 6nm. We therefore group the clusters of similar collapse lengths and estimate the number of domains in the polyprotein chain as $N_d = L_{tot}/20\text{nm}$. Setting the lowest value of L_{tot} within a group of a given N_d to be x_u at time zero and x_f to 4nm as the folded length of the protein, leads to the alignments of trajectories shown in Figs. 2A and 2B for the 10pN and 5pN force quench, respectively, in the group of $N_d = 3$. The non-equilibrium distribution $\rho(x)$ of the end-to-end length segregated by N_d is shown in Fig. 3. We find that they approximately scale linearly with N_d at both forces, as shown in the insets. At a quench force of 10pN, the extended polypeptides often plateau at $\sim 70\%$ of the contour length before their final collapse. Lowering the quench force to 5pN reveals faster collapse trajectories that visit all end-to-end lengths with a similar probability.

Using the observed distributions we then obtain $G_{N_d}(x)$, the free energy of a polyprotein of N_d domains and rescale these different profiles by

$$G(x + x_f) \equiv G_{N_d=1}(x + x_f) = \frac{1}{N_d} G_{N_d}\left(N_d(x + x_f)\right) \quad (3)$$

This linear rescaling collapses the free energy profiles at different N_d onto a single one, as shown in Fig. 4(A), which implies that the domains in the chain collapse all at once rather than in a stepwise manner. This scenario is inconsistent with previously proposed models for the stochastic refolding of individual domains [19] or the aggregation of the unfolded domains [20]. Instead, our result in (3) suggests a collapse of the entire polypep-

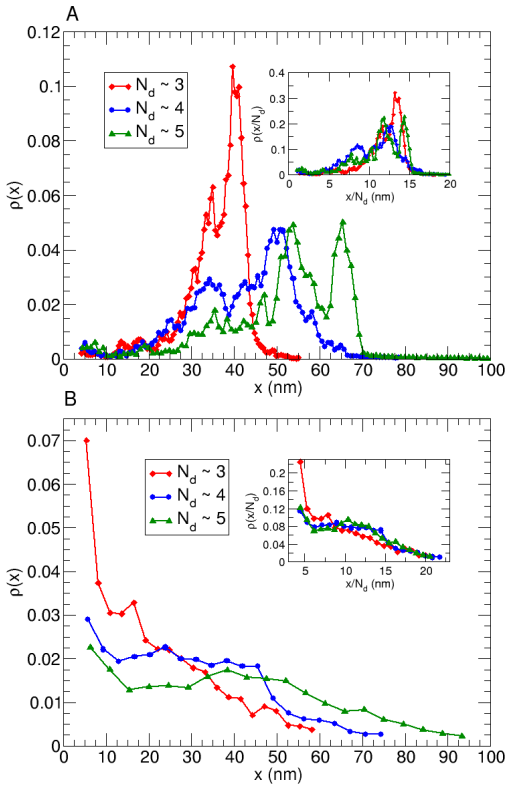

FIG. 3. The nonequilibrium distribution $\rho(x)$ for each N_d collected at a force quench of 10pN in (A) and 5pN in (B). The linear rescaling by N_d is shown in the inset, which indicates a cooperative mechanism for the collapse.

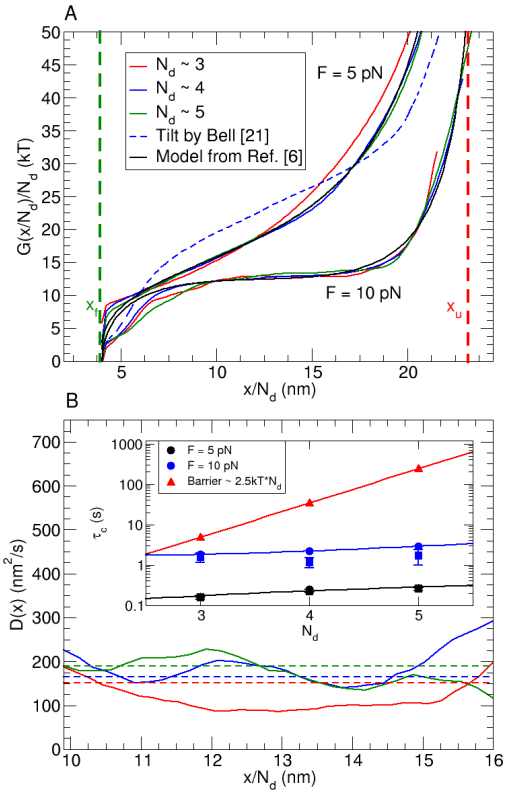

FIG. 4. (A)Experimental free energy profiles as a function of the end-to-end length, rescaled by N_d. (B) Diffusion coefficients $D(x)$ derived from Eq. (5) at 10pN (solid lines) compare well with those estimated from the free energy reconstruction in Eq. (2) (dashed lines). The inset shows τ_c dependence on N_d (squares), consistent with simulated data (circles).

tide chain into a final state that has been shown to be a random compact globule that forms native contacts over time [18]. The shape of the resulting free energy profiles $G(x)$ at quench forces of 5pN and 10pN are shown in Fig. 4A. Interestingly, none of the profiles exhibit a second minimum at an intermediate value of x between x_u and x_f. Rather, the collapse at $F = 10pN$ involves a diffusive slide on a plateau in the free energy that accelerates as the end-to-end length reaches a value $\sim$ 5nm away from x_f. Lowering the force to $5pN$ promotes a downhill collapse that is limited by friction alone. The difference in the profiles at 5pN and 10pN is roughly consistent with the Bell model [21], which predicts that this difference stems from the additional work done on the protein, $\Delta F x$ with $\Delta F = 5pN$, as shown by the dashed line in Fig. 4(A).

A functional form of the free energy landscape of ubiquitin was proposed in [5]. This form, however, does not fit our free energy profiles accurately due to its propensity to form barriers over a wide range of quench forces. A better fit is achieved using the physical model proposed for the collapse of RNA molecules in [6]. At large extensions near x_u, the polypeptide is a purely entropic spring, modeled as a worm like chain [22], that collapses to the intermediate plateau length where intramolecular interactions begin to play a role. These interactions lead to globules of many amino acids that assemble into a cylinder. Finally, the cylinder undergoes a hydrophobic collapse to a sphere at x_f, whose volume is the same as that of the cylinder. This mechanism for the polypeptide collapse has been introduced as the 'expanding sausage' model [23]. The full model is represented by the sum of the entropic term, the work done on the protein and the enthalpic model for the collapse:

$$G(x) = \frac{2k_BT}{\xi^2}\sqrt{\pi\Omega(x-x_f)} - F\,(x-x_f) + k_BT\frac{L_c}{l_P}\int_0^{\frac{x-x_f}{L_c-x_f}}\left(\frac{1}{4(1-y)^2} - \frac{1}{4} + y\right)dy \quad (4)$$

Here F is the applied force, L_c and l_P are the contour and persistence lengths of the extended protein, respectively, Ω is the volume of the cylinder, and ξ is the size of a globule inside the cylinder [23]. Fits to $G(x)$ in Fig. 4 give

$L_c = 26$nm, predicted by the extended length of ubiquitin (76 residues$\times 0.36 = 27.4$ nm), and $l_P = 0.82$nm at 5pN and 1.45nm at 10pN, in agreement with chain stiffening along the backbone due to intramolecular interactions [7]. To obtain the values of Ω and ξ from their ratio $\sqrt{\Omega}/\xi^2$ given by the fit, we assume that the size of the individual monomers in the sausage is l_P.

This implies that the number N of these monomers must be $N = L_c/l_P$. Following de Gennes' argument, we then set $\xi = l_P\sqrt{g}$ and $\Omega = L_c\pi\xi^2 = L_c\pi l_P^2 g$, where g is the number of monomers inside each globule and becomes the fit parameter that replaces the ratio. Fits to $G(x)$ in Fig. 4 thus yield $\xi = 2.6$nm at 5pN and 2.7nm at 10pN, in rough agreement with the value $\xi = 2$nm estimated for the hydrophobic collapse [24], and $\Omega = 204\text{nm}^3$ at 5pN and 374nm^3 at 10pN. The volumes of the cylinder correspond to the size of the collapsed protein, assumed to be spherical at the end of the trajectory, whose diameter is 7.3nm and 8.9nm under a force of 5 and 10pN, respectively. This makes sense because the folded length of the protein in solution is 4nm. This result indicates that the packing of globules inside the initial sausage is different for the two quench forces. Note also that the functional form of this free energy is consistent with the scaling with N_d in Eq. (3) since the volume of a polyprotein with N_d domains is $N_d\Omega$ and its contour length $N_d L_c$ while all the other parameters in Eq. (4) are unaffected by N_d. The collapsing traces can also be used to calculate the diffusion coefficient $D(x)$ on the reconstructed landscape and thereby verify our assumption that it is constant, $D(x) \approx D$. The idea is to cut the traces from x_u till the first moment they reach x and recalculate their non-equilibrium probability density ρ. The probability flux of these traces through the end-point x can be expressed in two ways: it is given by $D(x)\rho'(x)$, and it can also be estimated directly as $1/\tau_c(x)$, where $\tau_c(x)$ is the average time it takes them to collapse from x_u to x. Equating these two expressions and solving for $D(x)$ gives

$$D(x) = 1/(\tau_c(x)\rho'(x)) \tag{5}$$

This estimator for $D(x)$ is new and has the advantage over the standard one using quadratic variation of the trajectory [14] that it is insensitive to the time resolution of the instrument. Because of the small number of traces per N_d per force (~ 15), the estimate for $D(x)$ is accurate over the plateau regime in the end-to-end length in the data set at 10pN and not in the drift dominated parts of the landscape. The results obtained for polyproteins with different N_d in Fig. 4B are in good agreement with each other, within the experimental error, and show that the diffusion coefficient is roughly constant as a function of x, consistent with the assumption made in Eq. (1). The average value of $D = 180\text{nm}^2/\text{s}$ is consistent with the fact that the end-to-end length of the protein needs to diffuse by 15nm in about 1s on a barrierless free energy profile. Yet it is an order of magnitude smaller than that found in the highly extended regime of the entropic recoil of the polypeptide [25]. Even this larger value of the diffusion coefficient is vastly smaller than $D = 10^8\text{nm}^2/\text{s}$, estimated from molecular dynamics simulations of the polypeptide recoil. The surprisingly slow diffusion found in [25] was attributed to the viscous drag of the cantilever tether on the protein. The diffusion coefficient we find corresponds to moving a milimeter sized bead through water, which is larger than any component in the experimental system. Here we argue that the diffusion is further slowed down by the internal degrees of freedom that come into play when the molecule collapses beyond the initial entropic recoil. This hypothesis is consistent with recent single molecule experiments that show an independent friction with the end-to-end length of a folding protein [26, 27]. It is likely that many local barriers in other degrees of freedom can be mimicked by an effective diffusion constant.

To verify our results, we generate artificial traces using Eq. (1) with the estimated $G(x)$ and the constant $D = 180\text{nm}^2/\text{s}$, and show that they are in good agreement with experimental traces in Fig. 2. In addition, the fact that traces generated using D derived at 10pN reproduce the spread of times to collapse and the noise fluctuations in the experimental traces at 5pN suggests that D does not change with the quench force. We estimate that $\sim 70\%$ of experimental trajectories are consistent with the 1D diffusive model, while the outliers do not agree with the synthetic distribution of collapse times. Such trajectories have been observed previously [7]. Nevertheless, the simulated and the experimental average times to collapse τ_c agree well at both quench forces and for all N_d, as shown in the inset in Fig. 4B. By contrast, a linear scaling with N_d of a barrier-limited $G(x)$ [5] would lead to a much steeper dependence of τ_c with N_d, which is inconsistent with our and published polyprotein data [1].

To sum up, we use a non-equilibrium method to analyze trajectories and estimate free energy profiles and the diffusion coefficient of polypeptide collapse directly from the data. Our results show that the chain collapses simultaneously, independent of the number of protein domains that are in the chain. Moreover, the diffusion coefficient along the reaction coordinate is surprisingly small, indicating that internal friction might be at play. Our approach is general and applies to other nonequilibrium processes, such as the relaxation of DNA and RNA molecules, ion channels in voltage-clamp experiments, etc., which have a different structure to proteins.

We thank Alexander Grosberg and Dmitrii Makarov for insightful discussions. This work was supported in part by the New York University MRSEC Award DMR-0820341 and NSF Career Award 0955621. The research of E.V.-E. was supported in part by NSF grant DMS07-08140 and ONR grant N00014-11-1-0345.

* Email: eve2@cims.nyu.edu
† Email: jb2929@nyu.edu

[1] J. Fernandez and H. Li, Science **303**, 1674 (2004).
[2] C. Bustamante, J. Marko, E. Siggia, and S. Smith, Science **265**, 1599 (1994).
[3] M. Woodside, P. Anthony, W. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. Block, Science **314**, 1001 (2006).
[4] S. Garcia-Manyes, J. Brujić, C. Badilla, and J. Fernandez, Biophys. J. **93**, 2436 (2007).
[5] R. Berkovich, S. Garcia-Manyes, M. Urbakh, J. Klafter, and J. Fernandez, Biophys. J. **98**, 2692 (2010).
[6] C. Hyeon, G. Morrison, D. Pincus, and D. Thirumalai, Proc. Natl. Acad. Sci. USA **106**, 20288 (2009).
[7] K. Walther, F. Gräiter, L. Dougan, C. Badilla, B. Berne, and J. Fernandez, Proc. Natl. Acad. Sci. USA **104**, 7916 (2007).
[8] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[9] G. Crooks, J. of Stat. Phys. **90**, 1481 (1998).
[10] G. Hummer, J. of Chem. Phys. **114**, 7330 (2001).
[11] H. Oberhofer, C. Dellago, and P. Geissler, J. of Phys. Chem. B **109**, 6902 (2005).
[12] H. Oberhofer and C. Dellago, J. of Comp. Chem. **30**, 1726 (2009).
[13] D. Minh and J. Chodera, J. of Chem. Phys. **131**, 134110 (2009).
[14] R. Best and G. Hummer, Proc. Natl. Acad. Sci. USA **107**, 1088 (2010).
[15] X. Schlagberger, J. Bayer, J. Rädler, and R. Netz, EPL **76**, 346 (2006).
[16] Q. Zhang, J. Brujić, and E. Vanden-Eijnden, J. Stat. Phys. **144**, 344 (2011).
[17] J. Fernandez, H. Li, and J. Brujić, Science **306**, 411 (2004).
[18] S. Garcia-Maynes, L. Dougan, C. Badilla, J. Brujić, and J. Fernandez, Proc. Natl. Acad. Sci. USA **106**, 10534 (2009).
[19] R. Best and G. Hummer, Science **308**, 498 (2005).
[20] C. Wright, S. Teichmann, J. Clarke, and C. Dobson, Nature **438**, 878 (2005).
[21] G. Bell, Science **200**, 618 (1978).
[22] J. Marko and E. Siggia, Macromolecules **28**, 8759 (1995).
[23] P. de Gennes, J. Physique Lett. **46**, 639 (1985).
[24] D. Thirumalai, D. Klimov, and S. Woodson, Theor. Chem. Acc. **96**, 14 (1997).
[25] R. Berkovich, R. I. Hermans, I. Popa, G. Stirnemann, S. Garcia-Manyes, B. Berne, and J. Fernandez, Proc. Nat. Acad. USA **109**, 14416 (2012).
[26] A. Soranno, B. Buchli, D. Nettels, R. Cheng, S. Müller-Späth, S. Pfeil, A. Hoffman, E. Lipman, D. Makarov, and B. Schuler, Proc. Natl. Acad. Sci. USA (2012).
[27] T. Cellmer, E. Henry, J. Hofrichter, and W. Eaton, Proc. Natl. Acad. Sci. USA **105**, 18320 (2008).