Cooperativity in Thermal and Force-Induced Protein Unfolding: Integration of Crack Propagation and Network Elasticity Models

Amit Srivastava[1] and Rony Granek[1,2]

[1]*The Stella and Avram Goren-Goldstein Department of Biotechnology Engineering, Ben-Gurion University of The Negev, Beer Sheva 84105, Israel*

[2]*The Ilse Katz Institute for Meso and Nanoscale Science and Technology, Ben-Gurion University of The Negev, Beer Sheva 84105, Israel*

(Received 1 October 2012; published 25 March 2013)

We investigate force-induced and temperature-induced unfolding of proteins using the combination of a Gaussian network model and a crack propagation model based on ''bond''-breaking independent events. We assume the existence of threshold values for the mean strain and strain fluctuations that dictate bond rupture. Surprisingly, we find that this stepwise process usually leads to a few cooperative, first-order-like, transitions in which several bonds break simultaneously, reminiscent of the ''avalanches'' seen in disordered networks.

DOI: 10.1103/PhysRevLett.110.138101 PACS numbers: 87.15.A−, 87.10.Pq, 87.14.E−

Protein folding and unfolding under temperature variation and other ''denaturation'' conditions, such as pulling forces or solvent changes, have remained one of the fundamental enigmas of modern molecular biology [1–3]. In the last few years, some progress has been made both theoretically and experimentally [4,5]. Experimental progress has been achieved by single molecule techniques such as fluorescence resonance energy transfer [6], atomic force microscopy [7], and laser tweezers [8]. Fluorescence resonance energy transfer allows us to follow one coordinate in time during the transition, and so far, has been used under temperature variation and denaturating solvent. Single molecule force manipulations are mainly done by atomic force microscopy and are limited to a short period of pulling [9], although recently the use of a laser tweezers setup allowed long time investigation of the fold structure [10]. Theoretically, unfolding dynamics has been studied mostly by molecular dynamics (MD) [5], Monte Carlo simulations [11,12], and by phenomenological models [13]. Some coarse-grained elastic network models [14,15] also have been used to study the unfolding behavior of proteins. Most of these experimental and theoretical studies follow the kinetics of unfolding under a large temperature jump [3–5] or force jump [3–5], and therefore, may detect both nonequilibrium and equilibrium pathways of the folding-unfolding transition.

Yet, there is a lack of understanding of this transition. Biochemists usually refer to a unique unfolding (or denaturation) temperature above which the protein entirely unfolds. The common two state model allows for the two states to coexist at any temperature with fractions determined by the Boltzmann factor. The continuous decrease in enzyme activity above its optimal temperature is usually explained by the increased fraction of unfolded enzymes vs folded ones. However, on the microscopic level, one could expect that unfolding occurs as a stepwise process in which neighboring amino acids (not along the backbone) are disconnecting from each other and attain higher mobility, similar to monomers of a solvated polymer. An intriguing question is therefore: Does a stepwise process lead to a single unfolding temperature, or a single force level, above which the whole protein unfolds?

In this Letter, we commence by investigating thermal unfolding of proteins in a quasiequilibrium situation based on the stepwise rupture of the native contacts between the residues (''bonds''). We assume the existence of a critical value for local (nearest-neighbor) strain fluctuations that dictates the rupture of a bond. To apply this idea, we use the Gaussian network model (GNM) that describes interactions between amino acids on a coarse grained level and is based on a known folded protein structure [16]. Using the GNM, we compute the variance of thermal fluctuations of all bond lengths at a given temperature. When this variance reaches a threshold value, the bond ruptures. The GNM is now applied to the modified network, and any bond having fluctuations above the threshold, at the same temperature, ruptures too. We then raise the temperature by a small amount and repeat the process. The temperature is further raised until all (or most) bonds rupture mimicking an unfolded protein. Surprisingly, we find that this stepwise, single event, unfolding process usually leads to a few cooperative, first-order-like, transitions in which several bonds rupture at the same temperature. In some proteins, only two distinct group of states appear, the folded and the unfolded structures. This *cooperativity* is the result of an instability that develops when a single bond ruptures. As the network becomes more loose after the break, thermal fluctuations are enhanced, which can (although they do not have to) cause the rupture of other bonds having thermal fluctuations above the threshold.

An analogous scheme is used to investigate the pulling force induced unfolding of proteins at constant force

conditions and at "zero temperature," which is considered mostly for illustrative purposes. Here, too, unfolding is visualized as a series of "bond"-rupture independent events. Force is applied to a pair of amino acids, not necessarily those at the sequence ends, disregarding possible experimental difficulties. Starting with computing the GNM elastic response to the applied force, we locate the most strained (nonbackbone) bond in the network and determine its rupture based on a threshold value. Upon rupture, a new (slightly loose) network is formed, whose deformation at the same applied force is recalculated, and a check is performed for bonds whose strain is already above the threshold strain value. If such bonds are found, they are ruptured too. The force is then increased by a small increment and the above process is repeated until all bonds are ruptured. Similar to the thermal unfolding, here, too, we find a few cooperative, first-order-like, unfolding events, that lead to plateaus in force-extension curves. It is interesting to note that similar models have been considered for studying crack propagation in disorder networks [17], even though, in our case, the network is not random, but dictated by the native fold structure and the GNM cutoff distance R_c defining the interaction range. Interestingly, "avalanches" of bond-breaking events are also seen in disordered networks [17]. Closely related is the self-organized criticality, observed in random resistor networks in which resistors cut ("burn") when the current passing through them overcomes a threshold [18].

The GNM is a coarse grained model of a bead-spring network which is topology based and is independent of the sequence specificity [19]. In this model, each amino acid is represented by its α carbon (C_α), and is interacting with its first-shell neighbors residing below a cutoff distance R_c via harmonic springs. Despite its simplicity, GNM has been successful in predicting both local fluctuations and large scale conformational motion [20–22]. In the conventional GNM, all spring constants are identical; however, for better accuracy, we have distinguished between the (nearest-neighbor) backbone interactions, characterized by a spring constant $c\gamma$, and the nonbackbone inter-residue interactions, described by a spring constant γ [15]. The Hamiltonian of the system can be written as

$$H = \frac{1}{2}\gamma \sum_{\langle ij \rangle} \Gamma_{ij}(\vec{u}_i - \vec{u}_j)^2, \tag{1}$$

where the sum runs over all pairs $\langle ij \rangle$, $\{\vec{u}_i\}$ are the displacement vectors in 3D space, Γ_{ij} is a connectivity (spring constant) matrix distinguishing between backbone and nonbackbone springs,

$$\Gamma_{ij} = \begin{cases} c & \text{if } |i-j| = 1 \\ 1 & \text{if } |i-j| > 1 \quad \text{and} \quad R_{ij} \le R_c, \\ 0 & \text{if } |i-j| > 1 \quad \text{and} \quad R_{ij} > R_c \end{cases} \tag{2}$$

R_{ij} is the distance between ith and jth C_αs, and R_c is the cutoff distance (7 Å in this work). Transforming to a standard quadratic form, we get

$$H = \frac{1}{2}\gamma \sum_{m=x,y,z} \mathbf{u}_m^{\mathrm{T}} \cdot \mathbf{\Lambda} \cdot \mathbf{u}_m, \tag{3}$$

where the elements of the matrix $\mathbf{\Lambda}$ are

$$\Lambda_{ij} = \begin{cases} -\Gamma_{ij} & \text{if } i \neq j \\ \sum_{k \neq i} \Gamma_{ik} & \text{if } i = j \end{cases}, \tag{4}$$

and $\mathbf{u}_m$ is a vector in residue space of length N, the number of residues. To determine the value of γ and c for a particular protein, we have fitted the theoretically calculated B factors ($\frac{8\pi^2}{3}\langle \vec{u}_i^2 \rangle$) [16,23] to their x-ray values.

We applied our models to two well characterized proteins: (i) chymotrypsin inhibitor [CI2, Protein Data Bank (PDB) code 2CI2] that is an 83 residue protein consisting of an α-helix and three β-sheet strands, and (ii) barnase (PDB code 1A2P), that is a 110-residue ribonuclease that consists of four α-helix and five β-sheet strands (see the Supplemental Material [24]). We commence with a description of thermal unfolding. We denote the temperature at which the first bond ruptures as T_0. Figure 1 shows results for CI2. In Fig. 1(A), the number of ruptured bonds is plotted against the reduced temperature T/T_0. Figure 1(B) shows the corresponding connectivity maps at different reduced temperatures. We can observe four large jumps, where in each jump several bonds rupture simultaneously, due to an avalanchelike effect. Each jump corresponds to the opening of a different domain: from state (a) to state (b), bonds connecting the N term and β_3 domains break; from (c) to (d), the connecting bonds α-loop2, β_2-β_3, and β_1-β_2 break simultaneously; from (d) to (e), bonds associated with different β-sheet strands and loops rupture; and from (e) to (f), the α-helix bonds open. Thus, each jump in the plot signifies a first-order-like transition between different partially unfolded states. Apparently, the first two transitions correspond mainly to tertiary bond ruptures (i.e., bonds forming the ternary structure), and the last two transitions are secondary bond ruptures.

These results are consistent with MD [25,26] and MC [12] simulations. We note however that these simulations follow the kinetics of unfolding upon a large temperature jump, rather than predicting the equilibrium state at each temperature. Moreover, although unfolding is seen to take place by a few major cooperative transitions between intermediate states [26], these transitions are much clearer in the present study. We also note that in a similar, GNM based, study of unfolding, no cooperative behavior was revealed [15]. It should be emphasized that these cooperative transitions are not a preassumption, and it is gratifying that this cooperativity is predicted by such a simple model, even if some details may be inaccurately

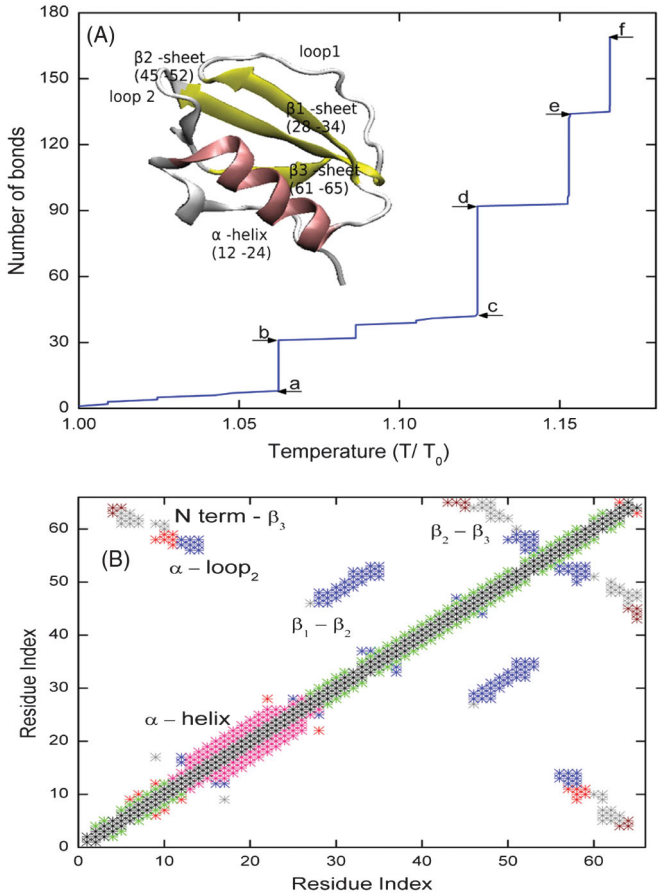

FIG. 1 (color). (A) Phase diagram for CI2: Number of ruptured bonds vs temperature. Different protein states are denoted by (a)–(f). The inset shows a cartoon representation of CI2, generated by using visual molecular dynamics [30]. For ease of comparison with previous protein studies, residues are renumbered to begin with 1 instead 20, since information regarding the first 19 residues is missing from the PDB. (B) Contact map of CI2 at different states. Each native contact is shown by the symbol *. The different colors correspond to the contacts being removed between the consecutive states marked in Fig. 1(A): (i) wine color—up to state (a), total of 7 native contacts get opened, (ii) gray—from (a) to (b), 24 contacts, (iii) red—from (b) to (c), 12 contacts, (iv) blue—from (c) to (d), 50 contacts, (v) green—from (d) to (e), 40 contacts, and (vi) pink—from (e) to (f), total of 37 native contacts get opened.

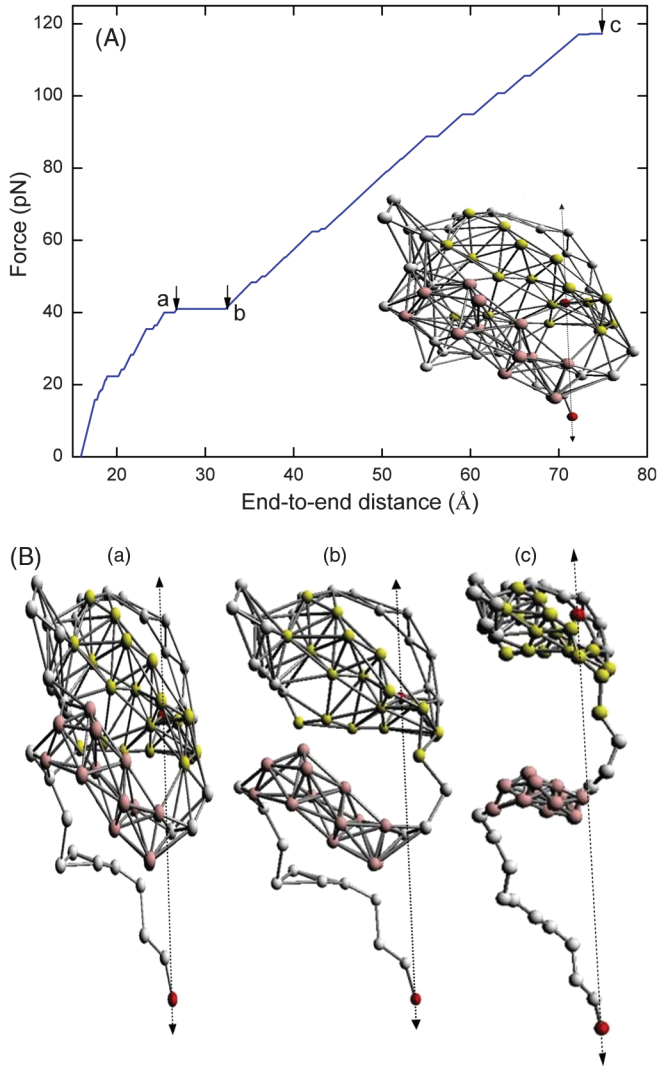

FIG. 2 (color). (A) Phase diagram for CI2: Force vs extension. A pulling force is applied at the end residues, colored in red. The inset shows the CI2 GNM-based network. Nodes' colors correspond to the secondary structure of which they are a part, shown in the inset of Fig. 1(A). (B) Snapshots of CI2 connectivity network at states (a)–(c) marked in Fig. 2(A).

predicted. If we assign T_0 to be the room temperature ($\simeq 300$ K), then complete unfolding is predicted to occur at $T \simeq 350$ K, close to the reported melting temperature (348 K). Results for barnase are presented in the Supplemental Material (see Fig. S1 in the Supplemental Material [24]). Here, too, unfolding occurs via first-order, avalanchelike, transitions. As for CI2, the sequence of unfolding is similar to the sequences observed experimentally by kinetic NMR studies [27], and those predicted by MD [28] and MC studies [29], keeping in mind that these are all kinetic studies that cannot be directly compared to the one presented here.

We now turn to explore unfolding under a pulling force. An obvious question is whether force induces similar unfolding pathways as thermal unfolding. While the temperature is a scalar, the force is a vector. The force can be exerted at different residue pairs, such that the resulting tension is transmitted to the different bonds unevenly, unlike temperature. Each elementary elastic stage of unfolding is calculated using the GNM [cf. Eq. (17) in the Supplemental Material [24]]. Bond rupture occurs, as described in the introductory paragraphs, when the nonbackbone bond stretching increases above a threshold value.

We commence with CI2 with the force applied at the end residues. Figure 2(A) shows the resulting force-extension relation, i.e., the force vs end-to-end distance. We observe

a few first-order transition plateaus in the plot, corresponding to several bonds rupturing simultaneously. Snapshots of the intermediate protein states (a), (b), and (c) are shown in Fig. 2(B). In Fig. 2, up to state (a), the connecting bonds of the N term with β_3 and of the α helix with loop2 rupture, from (a) to (b) the bonds in between the loops rupture, and from (b) to (c) some of the connecting bonds β_1-β_2 and β_2-β_3 rupture too. At state (c), there remain 47% of native bonds, and further increase of the force does not lead to further unfolding. It is interesting to note that the initial unfolding pathway that appears here is similar, even if not identical, to the thermal unfolding.

However, this is not universally true for other residue pairs at which the force is exerted. Figure 3(A) shows the force-extension relation for CI2, when the force is applied at the residue pair (15, 40). We note again a few first-order-like plateaus, demonstrating multistage cooperative unfolding; however, the sequence of unfolding is quite different. Moreover, 81% of the protein get unfolded (and further increasing of the force does not lead to further unfolding). In Fig. 3, up to state (a) a few bonds connecting the α helix and loop2 and connecting different loops get ruptured, from state (a) to state (b) almost all bonds connecting the N term to β_3 and α to loop2, and the majority of bonds connecting β_1 to β_2 and within the α helix, get ruptured, and from (b) to (c) bonds connecting the loops get ruptured. At state (c), the contacts β_2-β_3 remain intact. Thus, this sequence of unfolding is very different from the sequence obtained for a force applied at the end residues.

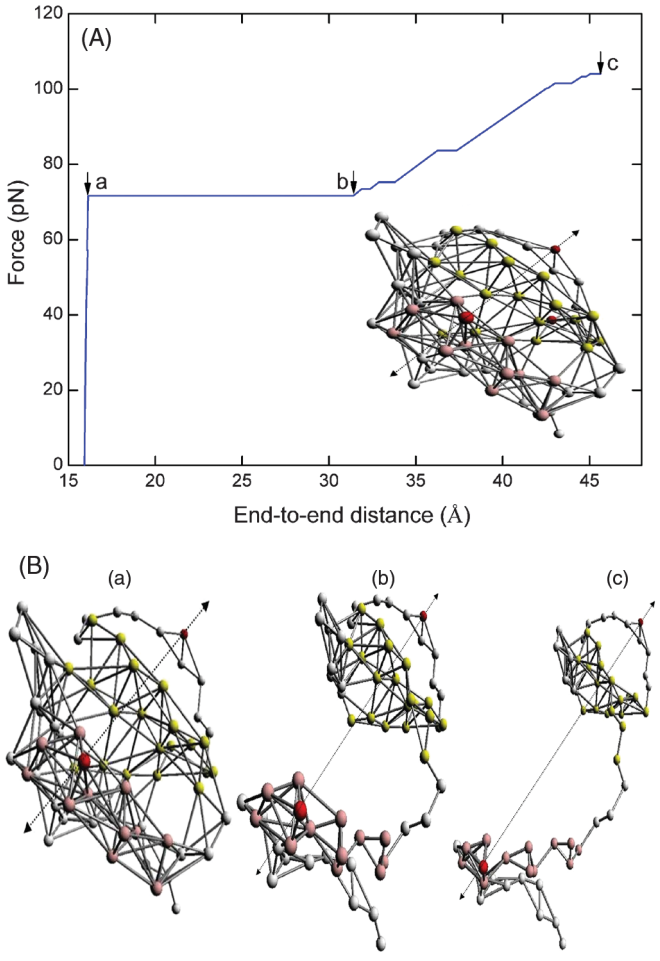

FIG. 3 (color). Same as Fig. 2, but with the force applied at residue pair (15, 40), colored in red.

Interestingly, while our bond rupture procedure, for both thermal and force-induced unfolding, is a "crack propagation" based model, it is possible to show, within first order perturbation theory, that it is very close to a thermodynamic criterion based on free-energy minimization (for details, see the Supplemental Material [24]). The Helmholtz free energy of a Gaussian network under force $\vec{f}$ acting on residues (i, j), after n bonds have ruptured, is (omitting irrelevant constants)

$$F = n\epsilon + \frac{3}{2}k_B T\sum_{\alpha}\ln(\lambda_\alpha) - \frac{f^2}{2\gamma}\sum_{\alpha}\frac{(U_\alpha^i - U_\alpha^j)^2}{\lambda_\alpha}, \quad (5)$$

where $\{\lambda_\alpha\}$ are the eigenvalues of the matrix $\mathbf{\Lambda}$, U_α^i stands for the ith (residue) entry of the eigenvector $\mathbf{U_\alpha}$, and $-\epsilon$ is the bond binding energy ($\epsilon > 0$). If δF is the free-energy change in a single bond rupture event, the bond to be chosen for rupture should be the one that will minimize δF and the event should take place when $\delta F = 0$. When this δF is estimated within first order perturbation theory, the result is an identical criterion for bond rupture in the case of thermal unfolding, and a very similar (even if not identical) criterion in the case of force-induced unfolding at $T = 0$. This is intuitively expected, as a highly fluctuating bond is likely to allow for extra internal protein motion when ruptured, and thus significantly increase its configurational entropy, and a highly tensed bond is likely to allow for a large increase in extension when ruptured, and thus maximize the work. This implies that "crack propagation" unfolding pathways are likely to be similar to thermodynamic equilibrium pathways.

To conclude, single bond rupture events, that are either based on a strain fluctuations threshold (thermal unfolding) or a mean strain threshold (force-induced unfolding), lead to cooperative, avalanchelike events, at which several bonds rupture simultaneously. While some details of our predictions could be inaccurate due to the major simplification of the GNM and our associated algorithm, a few general conclusions can be drawn. Unfolding always occurs through a few intermediate states. These states are not simply kinetic transition states and are stable in a certain range of temperatures or forces. In thermal unfolding, usually the tertiary structure unfolds first. Then, the secondary structure unfolds, where usually β sheets come first and α helix's last. Force-induced unfolding pathways are sensitive to the location of the applied force, and are, in principle, different from thermal unfolding pathways. Pulling at the end residues can sometimes induce unfolding pathways similar to thermal unfolding, while the pathway

generated from pulling at residues far away—along the sequence—from the end residues, is likely to diverge from that of thermal unfolding. Our results contradict the two state, dogmatic, model (folded and unfolded), on which interpretation of kinetic (experimental) studies is based [2]. A few intermediate states, though, have been observed in a recent (constant) pulling force experiment, in accord with our predictions [10].

We thank Yossi Klafter, Shlomi Reuveni, Matthias Rief, and Michael Urbakh for useful discussions. This research was supported by the Deutsch-Israelische Projektkooperation (DIP).

[1] A. R. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).

[2] A. Borgia, P. M. Williams, and J. Clarke, Annu. Rev. Biochem. **77**, 101 (2008); C. M. Dobson, A. Sali, and M. Karplus, Angew. Chem., Int. Ed. Engl. **37**, 868 (1998).

[3] J. E. Shea and C. L. Brooks, Annu. Rev. Phys. Chem. **52**, 499 (2001); S. Kumar and M. S. Li, Phys. Rep. **486**, 1 (2010).

[4] E. M. Puchner and H. E. Gaub, Curr. Opin. Struct. Biol. **19**, 605 (2009); W. J. Greenleaf, M. T. Woodside, and S. M. Block, Annu. Rev. Biophys. Biomol. Struct. **36**, 171 (2007).

[5] J. Sulkowska and M. Cieplak, J. Phys. Condens. Matter **19**, 283201 (2007); R. D. Toofanny and V. Daggett, WIREs Comput. Mol. Sci. **2**, 405 (2012); K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011); A. M. R. de Graff, G. Shannon, D. W. Farrell, P. M. Williams, and M. F. Thorpe, Biophys. J. **101**, 736 (2011).

[6] E. Hass, ChemPhysChem **6**, 858 (2005); B. Schuler, ChemPhysChem **6**, 1206 (2005).

[7] G. Binnig, C. F. Quate, and Ch. Gerber, Phys. Rev. Lett. **56**, 930 (1986); M. Rief, M. Gautel, F. Osterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997); R. Rounsevell, J. R. Forman, and J. Clarke, Methods **34**, 100 (2004).

[8] M. S. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, Science **276**, 1112 (1997).

[9] M. C. Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Farnandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).

[10] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gabhardt, and M. Rief, Science **334**, 512 (2011).

[11] D. E. Makarov, P. K. Hansma, and H. Metiu, J. Chem. Phys. **114**, 9663 (2001); A. Imparato and A. Pelizzola, Phys. Rev. Lett. **100**, 158104 (2008); A. Kleiner and E. Shakhnovich, Biophys. J. **92**, 2054 (2007).

[12] S. B. Ozkan, G. S. Dalgyn, and T. Haliloglu, Polymer **45**, 581 (2004).

[13] A. Imparato, A. Pelizzola, and M. Zamparo, Phys. Rev. Lett. **98**, 148102 (2007); G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003); M. Cieplak, A. Pastore, and T. X. Hoang, J. Chem. Phys. **122**, 054906 (2005); D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2006).

[14] H. Dietz and M. Rief, Phys. Rev. Lett. **100**, 098101 (2008); E. Eyal and I. Baher, Biophys. J. **94**, 3424 (2008).

[15] J. G. Su, C. H. Li, R. Hao, W. Z. Chen, and C. X. Wang, Biophys. J. **94**, 4586 (2008).

[16] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).

[17] M. J. Alava, P. K. V. V. Nukala, and S. Zapperi, Adv. Phys. **55**, 349 (2006).

[18] G. Paladin and A. Vulpiani, Phys. Rep. **156**, 147 (1987).

[19] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).

[20] S. Kundu, J. S. Melton, D. C. Sorensen, and G. N. Phillips, Jr., Biophys. J. **83**, 723 (2002).

[21] I. Bahar, A. Wallqvist, D. G. Covell, and R. L. Jernigan, Biochemistry **37**, 1067 (1998).

[22] N. A. Tamiz, E. Meirovitch, and I. Baher, Proteins **57**, 468 (2004).

[23] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.110.138101 for details about the model and algorithm for protein unfolding; the elastic free-energy and first-order perturbation theory; and the force and thermal unfolding result of the Barnase protein.

[25] D. E. Otzen, L. S. Itzhaki, N. F. elMasry, S. E. Jackson, and A. R. Fresht, Proc. Natl. Acad. Sci. U.S.A. **91**, 10422 (1994); A. Li and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **91**, 10430 (1994); V. Daggett, A. Li, L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, J. Mol. Biol. **257**, 430 (1996); A. Li and V. Daggett, J. Mol. Biol. **257**, 412 (1996); T. Lazaridis and M. Karplus, Science **278**, 1928 (1997); S. L. Kazmirski, A. Li, and V. Daggett, J. Mol. Biol. **290**, 283 (1999); S. L. Kazmirski, K. Wong, S. M. V. Freund, Y. Tan, A. R. Fersht, and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **98**, 4349 (2001).

[26] R. Day, B. J. Bennion, S. Ham, and V. Daggett, J. Mol. Biol. **322**, 189 (2002).

[27] A. R. Fersht, FEBS Lett. **325**, 5 (1993); J. M. Matthews and A. R. Fersht, Biochemistry **34**, 6805 (1995).

[28] A. Li and V. Daggett, J. Mol. Biol. **275**, 677 (1998); C. J. Bond, K. B. Wong, J. Clarke, A. R. Fersht, and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **94**, 13409 (1997); J. Tirado-Rives, M. Orozco, and W. L. Jorgensen, Biochemistry **36**, 7313 (1997).

[29] X. Salvatella, C. M. Dobson, A. R. Fersht, and M. Vendruscolo, Proc. Natl. Acad. Sci. U.S.A. **102**, 12389 (2005); S. Kmiecik and A. Kolinski, Proc. Natl. Acad. Sci. U.S.A. **104**, 12330 (2007).

[30] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).