Elastic Bond Network Model for Protein Unfolding Mechanics

Hendrik Dietz[1,*] and Matthias Rief[2,†]

[1]Dana-Farber Cancer Institute and BCMP, Harvard Medical School, Boston, Massachusetts 02115, USA
[2]Physik Department, Technische Universitaet Muenchen, Garching bei Muenchen, D85748 Germany

(Received 4 April 2007; published 4 March 2008)

Recent advances in single molecule mechanics have made it possible to investigate the mechanical anisotropy of protein stability in great detail. A quantitative prediction of protein unfolding forces at experimental time scales has so far been difficult. Here, we present an elastically bonded network model to describe the mechanical unfolding forces of green fluorescent protein in eight different pulling directions. The combination of an elastic network and irreversible bond fracture kinetics offers a new concept to understand the determinants of mechanical protein stability.

DOI: 10.1103/PhysRevLett.100.098101 PACS numbers: 87.15.La, 82.37.Rs, 87.14.E−, 87.15.A−

Mechanical experiments with single protein molecules have provided a host of novel information on the mechanics and unfolding of protein structures. The interpretation of such experiments has greatly profited from molecular dynamics simulations that helped to discover the presence and structure of intermediate states [1–6]. The major drawback of full atom molecular dynamics simulations is the huge discrepancy between experimental (seconds) and simulation time scales (nanoseconds). Even though approaches have been presented to bridge this gap in time scales [7], quantitative comparison of simulated and experimental protein fracture forces has largely remained impossible. This has led researchers to apply and develop more coarse-grained models to tackle such questions [8–12].

In a recent study we combined cysteine engineering with single molecule force spectroscopy [13] to explore the anisotropic deformation response of green fluorescent protein (GFP) [14]. The origin of the observed mechanical anisotropy has so far been unclear. Here, we describe a network model of elastic bonds that allows quantitative description of the fracture of a protein structure loaded with force in various directions. The basis of our model is illustrated in Fig. 1. The protein structure is described as a network of identical bonds which may break irreversibly under force load. Network nodes are positioned at C_α coordinates. Bonds are assigned between nodes l and m if their distance is smaller than a cutoff radius R_C. The rate coefficient for thermally activated fracture of an individual bond is accelerated exponentially by a force F_{lm} acting on the bond [15]: $k_{\mathrm{off}}(F_{lm}) = k_0 \exp(F_{lm}\Delta x_1/k_B T)$, where Δx_1 describes the distance between the bond's bound and transition states and k_0 its zero-force dissociation rate coefficient, k_B is Boltzmann's constant, and T is temperature. The probability for the bond to withstand a force F_{lm} applied to the bond with constant force-loading rate η is $P_{lm}(F_{lm}) = \exp[\frac{k_0}{\eta}\frac{k_B T}{\Delta x_1}(1 - \exp\frac{F_{im}\Delta x_1}{k_B T})]$. We treat fracture of the whole network as a highly cooperative process and assume that fracture of the first bond in the network will induce fracture of the whole network. This assumption is motivated by the generally observed high cooperativity of protein unfolding in mechanical single molecule experiments [16]. Therefore, in our model, a folded protein requires all bonds to be intact at any given time. The probability for all bonds to be intact under an external force F_{ij} applied via network nodes i and j is given by the product of all single bond probabilities:

$$P_N(F_{ij}) = \prod_{l<N}\prod_{m<l} e^{(k_0/\eta)(k_B T/\Delta x_1)(1/\alpha_{lm}|_{ij})(1-e^{\alpha_{lm}|_{ij}(F_{ij}\Delta x_1/k_B T)})}. \quad (1)$$

The probability density $g_N(F_{ij}) = -dP_N/dF_{ij}$ to observe the first fracture event of any single bond in the network at a certain applied force F_{ij} is given by

$$g_N(F_{ij}) = P_N(F_{ij})\left(\sum_{l<N}\sum_{m<l}\frac{k_0}{\eta}e^{\alpha_{lm}|_{ij}(F_{ij}\Delta x_1/k_B T)}\right). \quad (2)$$

$g_N(F_{ij})$ allows for the analytical calculation of fracture force distributions that can be directly compared to experimental data. In both Eqs. (1) and (2), the force F_{lm} acting on each bond has been replaced by the product of a coefficient $\alpha_{lm}|_{ij}$ and the external force F_{ij}. $\alpha_{lm}|_{ij}$ defines

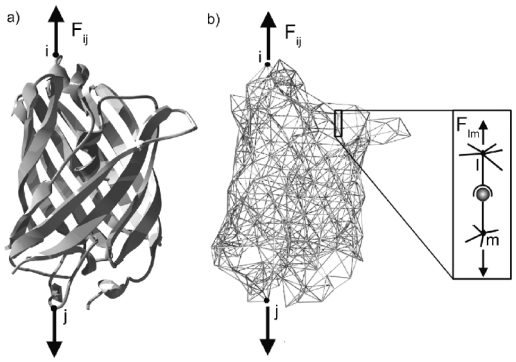

FIG. 1. (a) Applying force to a protein structure (here GFP) via two amino acids i and j, using cysteine engineering combined with single molecule force spectroscopy. (b) GFP modeled as an elastic bond network. (Inset) Network connections are springs with identical spring constant that may irreversibly break.

a force-action matrix that describes the fraction of the total applied load F_{ij} that each bond between nodes l and m experiences in the network when force is applied via two nodes i and j. It is important to note that $\alpha_{lm}|_{ij}$ enters the model as a parameter in order to predict the fracture kinetics of the bond network under force. To calculate the force-action matrix $\alpha_{lm}|_{ij}$ we modeled the protein structure as an elastic network [17,18] and assumed the bonds in the system as elastic springs. Identical spring constants of 10 N/m were assigned to each network connection. We solved numerically the equations of motion of the elastically coupled mass point system [19] while moving nodes i and j apart until a desired maximum deformation $d_{\max}$ was achieved. The total restoring force F_{ij} arising on nodes i and j was monitored, while internal forces F_{lm} between nodes l and m were calculated from the deformation of each spring. The force-action matrix was then calculated via $\alpha_{lm}|_{ij} = F_{lm}/F_{ij}$. Experimental evidence for GFP [14,20] strongly suggested that the protein shows negligible distortions up to force loads as high as 700 pN. A maximum deformation of $d_{\max} = 0.5$ Å was thus chosen in order minimize perturbation of network topology. $\alpha_{lm}|_{ij}$ was set to zero if no bond had been assigned between nodes l and m by the cutoff criterion R_C.

A graphic representation of the force-action matrix $\alpha_{lm}|_{ij}$ for three different pulling directions in GFP [$(i, j) = (132, 212)$, $(3, 132)$ and $(6, 221)$] is given in Fig. 2. Obviously, the force propagation through the network depends strongly on the pulling direction. The leftmost structure (132, 212) shows high forces acting predominantly at the springs located between residues 132 and 212. In contrast, in the rightmost structure, the forces are distributed among parallel springs with each spring bearing much lower load.

We applied this elastic bond network model to experimental data measured in force spectroscopy experiments with GFP. Data from five different pulling geometries have been published previously [14]. In addition, we measured

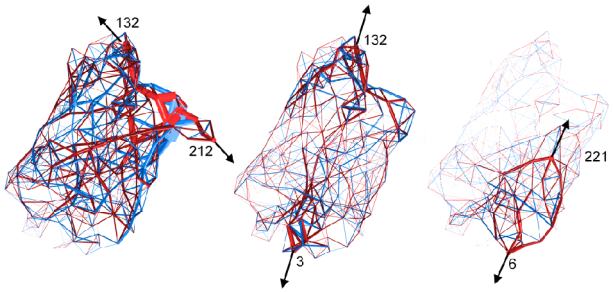

FIG. 2 (color online). Illustration of the force-action matrix $\alpha_{lm}|_{ij}$ when force is applied to the GFP network via nodes (132, 212), (3, 132), and (6, 221). The width of the network connections codes for the values of $\alpha_{lm}|_{ij}$, light (red) colors denote extension, dark (blue) colors denote compression. Cutoff radius $R_C = 6.75$ Å.

three further pulling directions to provide a large data set to test the validity of our model. The experimental fracture force distributions together with predicted $g_N(F_{ij})$ using Eq. (2) (solid black lines) are shown in Fig. 3. With only one exception, both the average unfolding force in the different pulling directions and the width of the force distributions are well reproduced. The discrepancy between model and data for the direction (26, 198) may be attributed to an overassignment of bonds in the region around residue 198 due to the simple cutoff criterion applied. The good quantitative agreement with the rest of the data set argues for a surprising simplicity underlying the determinants of mechanical stability of proteins. Our model only relies on structural information and entirely neglects the potential influence of sequence variations as

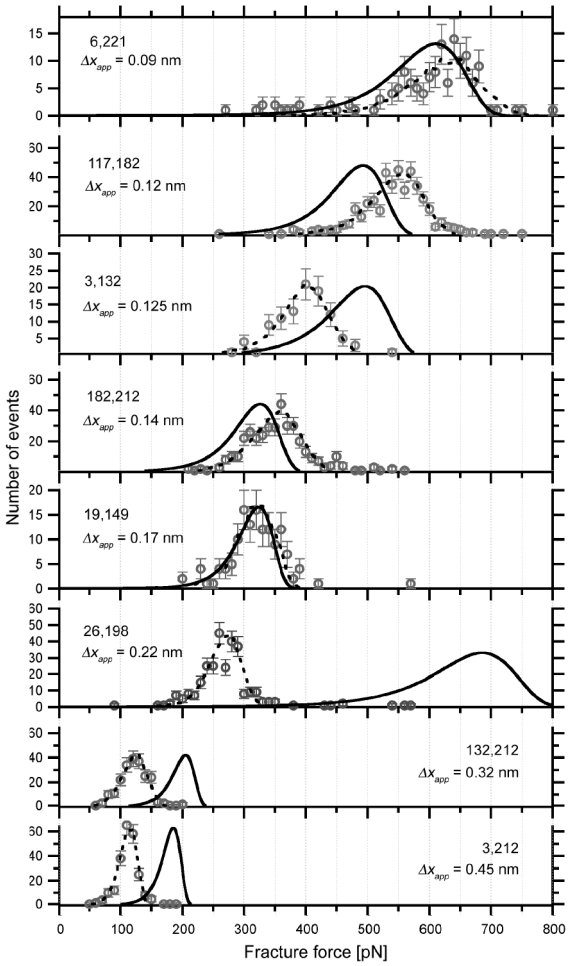

FIG. 3. Comparison of experimental fracture force distributions (open circles) for GFP with predicted fracture force distributions from the elastic bond network model (solid lines). Parameters used: cutoff radius $R_C = 6.75$ Å, transition state position $\Delta x_1 = 0.28$ nm, zero-force dissociation rate $k_0 = 1 \times 10^{-3}$ s^{-1}. Calculated $g_N(F_{ij})$ have been normalized to match observed number of events. Experimental conditions were as in [14]. Broken lines: two-state analysis as described in [14]. Apparent transition state positions Δx_{app} are given in the figure.

long as they conserve the structure. Consistent with this conclusion, many structure conserving mutations within the immunoglobulin domain Ig27 from titin only showed minor effects ($\leq$30 pN) on the unfolding force [21,22] while β-sheet breaking proline mutations did alter the stability of Ig27 significantly [23].

To assess the significance of our results, it is important to note that our model contains four free parameters: two parameters ($\Delta x_1 = 0.28$ nm and $k_0 = 1 \times 10^{-3}$ s^{-1}) describing single bond kinetics and two more parameters (cutoff radius $R_C = 6.75$ Å and maximum deformation $d_{\max} = 0.5$ Å) in order to calculate the force-action matrix for the elastic network. We found that choosing a cutoff radius R_C of 6.75 Å, consistent with the literature values reported for elastic network modeling of protein B factors [17,18], yields best agreement with the experimental data, where smaller R_C eliminate long-range connectivity while larger R_C erase anisotropic mechanics. Larger $d_{\max}$ values result in deviation from the experimentally observed stability order for direction (3, 212).

One interesting trend in the data in Fig. 3 is the strong correlation between the width of the force distributions and the average fracture force: distributions with low average force are generally narrow, while high force distributions are wide. Such a trend is well conserved for fracture force distributions of various proteins in the literature [14,24–26]. Our model now offers a simple explanation for this general trend. If fracture of a protein structure is determined by the first fracture event of any bond in the network, both fracture forces and distribution widths will be simply determined by the topology of interconnection of the bonds in the network. To illustrate this, we calculated the fracture forces for two limiting cases of bond networks: a serial and a parallel combination of N springs (see Fig. 4).

In the case of the serial combination, fracture forces of the network lie close to those of an individual spring with only a weak, logarithmic dependence on N. For a parallel

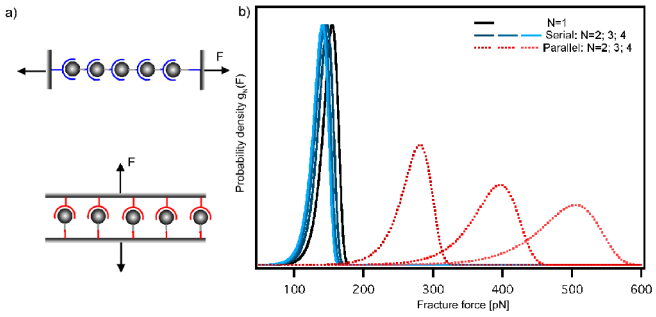

FIG. 4 (color online). (a) Scheme of a serial and a parallel combination of N identical bonds. The amount of the applied force transduced to individual bonds (i.e., $\alpha_{lm}|_{ij}$) is $\alpha = 1$ in the serial configuration, while $\alpha = 1/N$ in the parallel configuration. (b) Fracture probability densities $g_N(F)$, calculated with Eq. (2) for different number N of bonds loaded.

combination, both the distribution widths and the average fracture forces grow proportionally in N. Hence, the correlations found experimentally between average forces and distribution widths may be explained by a variable number of otherwise almost identical parallel bonds in a protein bond network.

These considerations can be generalized to an arbitrary network of bonds. First fracture in a discrete bond network will be dominated by the bond that bears most of the externally applied force. A consequence of Eq. (2) is that the average force at which first fracture will be observed is scaled by the largest value of the force-action matrix:

$$\langle \tilde{F} \rangle \simeq \langle F_1 \rangle / \max \alpha_{lm}|_{ij}, \tag{3}$$

where $\langle F_1 \rangle$ denotes the average fracture force of the generic bond type in the system when loaded individually. An example is provided in Fig. 4: In a parallel combination of bonds, force is equally distributed among many bonds, $\max \alpha_{lm}$ will be small and hence the fracture force high.

In the literature, cooperative unfolding of proteins under load is generally discussed as a two-state process. Such a treatment (broken lines in Fig. 3) generally yields a transition state position Δx_{app} and a zero-force unfolding rate k_0. In this treatment, the width of fracture force distributions is reciprocal to Δx_{app} while the position of the distribution on the force axis shifts logarithmically with k_0 [27]. How can the fracture of a network of multiple bonds be reconciled with a two-state treatment?

If fracture of a multiple bond system is treated as a two-state system, the analysis of fracture force distribution widths yields an *apparent* transition state position:

$$\Delta x_{\text{app}} \simeq \max \alpha_{lm}|_{ij} \Delta x_1. \tag{4}$$

Again, this effect is illustrated by the parallel bonded network of Fig. 4 where the width of the force distributions grows proportionally with N, indicating a decreasing Δx_{app}. The interpretation of the fracture forces in the framework of a bonded network hence suggests varying degrees of parallel bonding as the source for the different transition state positions Δx_{app} found in mechanical protein unfolding. Therefore, care must be taken when interpreting Δx_{app} in terms of a real length. A direct prediction of Eqs. (3) and (4) is that the product of the average unfolding force and apparent transition state position should be constant:

$$\langle F \rangle \Delta x_{\text{app}} \simeq \langle F_1 \rangle \Delta x_1 = \text{const.} \tag{5}$$

Obviously, the connectivity of the network and the direction of force load represented in the $\alpha_{lm}|_{ij}$ matrix is canceled out in the product and the product reflects only the properties of the generic bond type in the system. Supporting this interpretation, the product of average fracture force and apparent transition state position observed experimentally at the same loading rate for the eight directions in GFP (Fig. 3) is well conserved (54 $\pm$ 7 pN nm).

This product is of the same magnitude for many proteins which have been mechanically investigated so far (e.g., Ig27: 50 pN nm [28]; tenascin: 54 pN nm [29]; ubiquitin: N-C: 51 pN nm, 48-C: 54 pN nm [24]; T4 lysozyme: 52 pN nm [30]; protein L: 46 pN nm [31]; spectrin: 50 pN nm [32], etc.), providing further support for interpretation of fracture forces in the framework of a bond network.

The simplified model presented here and its choice of parameters is certainly not unique to all proteins and may have to be adapted for other proteins. With our choice of parameters, forces significantly smaller than 100 pN cannot be explained. Therefore, proteins with predominantly α-helical structure exhibiting significantly lower unfolding forces, like spectrin [32], T4 lysozyme [30], or leucine zippers [33], require a different choice of parameters. Comprehensive data sets of various pulling directions also for other proteins are necessary for further refinement of the model. Another limitation of our model concerns the strict irreversibility of the fracture of the individual bonds. This assumption will certainly not hold true if the mechanical unfolding process approaches equilibrium as has been observed for RNAse H [34], myosin tail [35], and leucine zippers [33]. Kinetic modeling for multiple reversible bonds will become important in this case [36]. Moreover, the model presented here is not capable of predicting the structure of potential intermediate states occurring along the mechanical unfolding pathway. For these purposes, application of more realistic models [4,8,11] is essential.

In summary, the description of mechanical unfolding of proteins as fracture of an elastic bond network yielded quantitative agreement with experimental data, thus explaining the large anisotropy of protein mechanics.

We thank Emanuele Paci for stimulating discussions. This work was supported by DFG Project No. RI 990/3-1 and the cluster of excellence NIM.

*dietz@ph.tum.de
†mrief@ph.tum.de

[1] R. Best, S. Fowler, J. Toca-Herrera, and J. Clarke, J. Mol. Biol. **288**, 441 (1999).
[2] P. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. Oberhauser, K. Schulten, and J. Fernandez, Nature (London) **402**, 100 (1999).
[3] M. Gao, D. Craig, O. Lequin, I. Campbell, V. Vogel, and K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **100**, 14 784 (2003).
[4] F. Grater, J. Shen, H. Jiang, M. Gautel, and H. Grubmuller, Biophys. J. **88**, 790 (2004).
[5] V. Ortiz, S. Nielsen, M. Klein, and D. Discher, J. Mol. Biol. **349**, 638 (2005).
[6] E. Paci and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **97**, 6521 (2000).
[7] B. Heymann and H. Grubmuller, Chem. Phys. Lett. **303**, 1 (1999).
[8] D. West, D. Brockwell, P. Olmsted, S. Radford, and E. Paci, Biophys. J. **90**, 287 (2005).
[9] D. West, P. Olmsted, and E. Paci, J. Chem. Phys. **124**, 154909 (2006).
[10] D. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 7254 (2000).
[11] C. Hyeon, R. Dima, and D. Thirumalai, Structure **14**, 1633 (2006).
[12] R. Rohs, C. Etchebest, and R. Lavery, Biophys. J. **76**, 2760 (1999).
[13] H. Dietz, M. Bertz, M. Schlierf, F. Berkemeier, T. Bornschlogl, J. Junker, and M. Rief, Nature Protocols **1**, 80 (2006).
[14] H. Dietz, F. Berkemeier, M. Bertz, and M. Rief, Proc. Natl. Acad. Sci. U.S.A. **103**, 12 724 (2006).
[15] G. Bell, Science **200**, 618 (1978).
[16] M. Carrion-Vazquez, A. Oberhauser, T. Fisher, P. Marszalek, H. Li, and J. Fernandez, Prog. Biophys. Molec. Biol. **74**, 63 (2000).
[17] M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[18] I. Bahar, A. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[19] L. Verlet, Phys. Rev. **159**, 98 (1967).
[20] H. Dietz and M. Rief, Proc. Natl. Acad. Sci. U.S.A. **103**, 1244 (2006).
[21] D. Brockwell, G. Beddard, J. Clarkson, R. Zinober, A. Blake, J. Trinick, P. Olmsted, D. Smith, and S. Radford, Biophys. J. **83**, 458 (2002).
[22] R. Best, S. Fowler, J. Toca-Herrera, and J. Clarke, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 143 (2002).
[23] H. Li, M. Carrion-Vazquez, A. Oberhauser, P. Marszalek, and J. Fernandez, Nat. Struct. Biol. **7**, 1117 (2000).
[24] M. Carrion-Vazquez, H. Li, H. Lu, P. Marszalek, A. Oberhauser, and J. Fernandez, Nat. Struct. Biol. **10**, 738 (2003).
[25] Y. Cao, C. Lam, M. Wang, and H. Li, Angew. Chem., Int. Ed. Engl. **45**, 642 (2006).
[26] Y. Cao and H. Li, J. Mol. Biol. **361**, 372 (2006).
[27] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[28] M. Carrion-Vazquez, A. Oberhauser, S. Fowler, P. Marszalek, S. Broedel, J. Clarke, and J. Fernandez, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999).
[29] A. Oberhauser, P. Marszalek, H. Erickson, and J. Fernandez, Nature (London) **393**, 181 (1998).
[30] G. Yang, C. Cecconi, W. Baase, I. Vetter, W. Breyer, J. Haack, B. Matthews, F. Dahlquist, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **97**, 139 (2000).
[31] D. Brockwell, G. Beddard, E. Paci, D. West, P. Olmsted, A. Smith, and S. Radford, Biophys. J. **89**, 506 (2005).
[32] M. Rief, J. Pascual, M. Saraste, and H. Gaub, J. Mol. Biol. **286**, 553 (1999).
[33] T. Bornschlogl and M. Rief, Phys. Rev. Lett. **96**, 118102 (2006).
[34] C. Cecconi, E. Shank, C. Bustamante, and S. Marqusee, Science **309**, 2057 (2005).
[35] I. Schwaiger, C. Sattler, D. Hostetter, and M. Rief, Nat. Mater. **1**, 232 (2002).
[36] U. Seifert, Phys. Rev. Lett. **84**, 2750 (2000).