THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

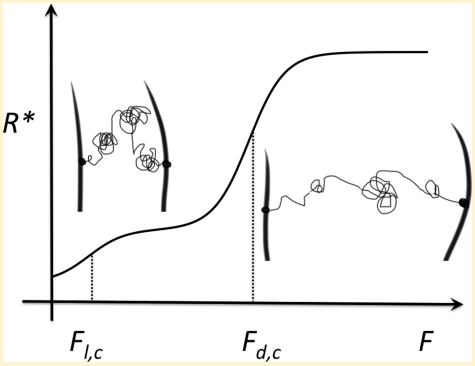
pubs.acs.org/JPCB

Physical Microscopic Model of Proteins Under Force

Nikolay V. Dokholyan*

Department of Biochemistry and Biophysics, University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27599, United States

ABSTRACT: Nature has evolved proteins to counteract forces applied on living cells, and has designed proteins that can sense forces. One can appreciate Nature's ingenuity in evolving these proteins to be highly sensitive to force and to have a high dynamic force range at which they operate. To achieve this level of sensitivity, many of these proteins are composed of multiple domains and linking peptides connecting these domains, each of them having their own force response regimes. Here, using a simple model of a protein, we address the question of how each individual domain responds to force. We also ask how multidomain proteins respond to forces. We find that the end-to-end distance of individual domains under force scales linearly with force. In multidomain proteins, we find that the force response has a rich range: at low force, extension is predominantly governed by "weaker" linking peptides or domain intermediates, while at higher force, the extension is governed by unfolding of individual domains. Overall, the force extension curve comprises multiple sigmoidal transitions governed by unfolding of linking peptides and domains. Our study provides a basic framework for the understanding of protein response to force, and allows for interpretation experiments in which force is used to study the mechanical properties of multidomain proteins.



Nature widely utilizes mechanical force in order to control biological systems from the level of molecules to cells to organs. One particularly fascinating example of mechanical stress used as a tool to control a number of biological processes is found in mechano-sensing proteins. These proteins are responsible for cytoskeletal organization[1] (e.g., actin fibers[2,3]) and remodeling (e.g., filamins[4,5]), cellular transport (e.g., myosin and other motor proteins[6−9]), cell division,[10] contractility (e.g., titin[11,12]), extracellular matrix organization (e.g., tenascin,[13] collagen, elastin[14,15]), and other biological processes. Mechano-sensing proteins, such as filamin, titin, and collagen, are structurally tailored to provide a diverse response to mechanical stimuli or to induce mechanical stress.

Interestingly, the majority of these proteins have *modular* organization: individual cooperatively folding domains[16] are either independent proteins that are self-organized into fibers (e.g., actin fibers), or covalently linked within a larger protein (e.g., filamin and titin). Modular organization of some of these proteins results in a wide dynamic response range to mechanical stimuli. Depending upon the mechanical force acting on them, the response is tailored to serve specific biological functions. For example, filamin, which comprises 24 sequential immunoglobulin (Ig)-like domains connected by 2−6 residue linkers,[17] responds to mechanical stimuli induced by over 70 binding partners in a wide range of force. Diverse structural elements contribute to various force response regimes: linkers balance low external force; midrange force is balanced by structural rearrangements of individual domains, such as the unfurling of a β-strand; and the unfolding of individual domains balances high force. Understanding how external force regulates protein structure is vital to our understanding of a wide variety of biological processes associated with these proteins. Here we ask two specific questions: First, how does an external force affect the end-to-end distance distribution in a single protein domain? Second, what is the total extension of a multidomain protein in response to a force?

Previous approaches treated general polymers under force as freely jointed chains or freely rotating chains,[18,19] or a worm-like chain.[20] Unlike most polymers, proteins have been designed by nature to fold cooperatively in an all-or-none transition. This unique property is vital for protein survival in the cell, as unfolded proteins are usually targeted for degradation. Hence, it is important to consider free energy barriers when considering the folding of multidomain proteins (see, e.g., ref 21). Other approaches have offered elegant models that describe forces acting on protein-like constructs.[22−25]

Here, we present a simple model that addresses the question of protein deformation response to force. We first consider a nonideal self-avoiding model of a protein and determine the equilibrium distribution of the end-to-end distances upon the application of force. We then consider a model of a multidomain protein that consists of independently folded domains joined consecutively by short linkers, and determine the role of linkers on the extension of such multidomain

Special Issue: Harold A. Scheraga Festschrift

Received: December 28, 2011
Revised: February 25, 2012
Published: March 1, 2012

   dx.doi.org/10.1021/jp212543m | *J. Phys. Chem. B* 2012, 116, 6806−6809

proteins. It should be noted that the assumption of independence does not hold for all proteins.[26]

END-TO-END DISTANCE OF A PROTEIN UNDER FORCE

We model a protein as a nonideal self-avoiding polymer and consider this protein to be at equilibrium after force is applied. The distribution of the end-to-end distances R in a nonideal self-avoiding polymer (of size $N \gg 1$)[27] under external force F can be written as[28]

$$P(R|F) \propto R^2 \exp\left(-\frac{3R^2}{2Nb^2} - \frac{N^2\upsilon_c}{2R^3} - \frac{E(R)}{k_BT} + \frac{F\delta R}{k_BT}\right) \quad (1)$$

where ν_c is the effective volume, b is the bond length between monomers, T is the temperature, k_B is the Boltzmann constant, $E(R)$ is the potential energy of a polymer, and δR is the distance by which one needs to stretch a protein so that the crucial interactions responsible for the folding barrier are disrupted (Figure 1). The sum of the first three terms in the exponent is the free energy of the polymer chain.[28]

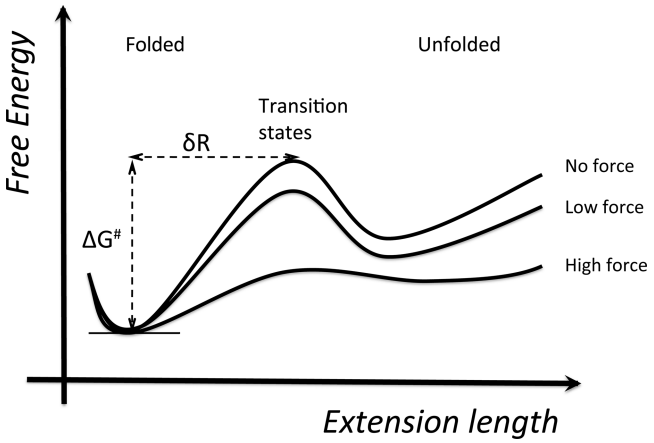


Figure 1. Schematic representation of the free energy of a two-state protein as a function of extension length. At low force, the protein folding is dominated by the transition across a single free energy barrier. As force increases, the transition barrier decreases, allowing proteins under stress to sample more unfolded states.

The most probable end-to-end distance is determined by the max $P(R|F) = P(R^*,F)$, which can be obtained upon differentiating eq 1 as a solution of the following equation:

$$\xi^5 + \omega\xi^4 - \xi^3 = \eta \quad (2)$$

where $\xi = R^*/R_0^*$, $R_0^* = (2Nb^2/3)^{1/2}$, $\omega = \{[E'(R) - F\delta R]/k_BT\}R_0^*$, and $\eta = 3N^2\upsilon_c/4R_0^{*3}$.

Although it is difficult to obtain an exact solution to eq 2, one can estimate the asymptotic behavior of the most probable solution as a function of F and N ($N \gg 1$). For simplicity, we consider a protein model where $E = E_0$ when a protein is folded and $E = 0$, when it is unfolded. For large F, $R^* \propto N$, ${R^*}_0 \propto N^{1/2}$, the equation will be dominated by the first two terms: $\xi^5 + \omega\xi^4 \approx 0$, and the solution is

$$R^* \approx \frac{FR_0^{*2}}{k_BT} \quad (3)$$

The linear scaling of the most probable end-to-end distance with the length N (Figure 2A), as well as the result that the extension is linear with force, is not surprising at high forces.

For small F, $\xi \geq 1$, and eq 2 is dominated by the first two terms on the left-hand-side and the right-hand-side: $\xi^5 + \omega\xi^4 \approx \eta$ can be rewritten as $\xi \approx \eta^{1/5}(1 - (1/5)(\omega/\xi))$. Solving for ξ, we obtain $R^* \approx R_0^*\eta^{1/5}(1 - (1/5)(\omega/\eta^{2/5}))$. In our simple two-state protein model,

$$R^* \propto N^{3/5}\left(1 + \frac{F}{k_BT}\frac{\lambda}{\sqrt{N}}\right) \quad (4)$$

where λ is a constant. As expected, at zero force ($F = 0$), we recover Flory scaling $R^* \propto N^{3/5}$.

Interestingly, this model predicts that protein extension, as a response to force, is linear at both low and high forces (Figure 2B). This response becomes nonlinear at the ranges of forces at which $F\delta R$ becomes comparable with the protein free energy barrier.

MULTIDOMAIN PROTEINS

Many structural proteins, such as filamin, titin, and fibronectin, consist of multiple domains connected by linkers of varying lengths. These proteins respond to mechanical stress at various magnitudes of force acting on these proteins. The multidomain organization of these proteins allows for response to wide range of stress: at low stress, the linkers unfold, while at higher stress, the domains unfold. This combinatorial unfolding offers a rich

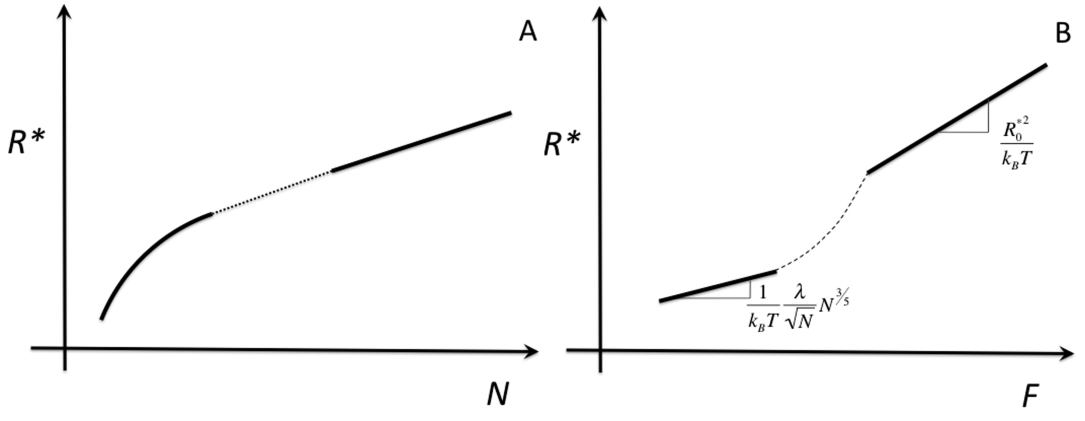


Figure 2. Schematic representation of the most likely extension of a single domain protein as a function of (a) length, and (b) force.

dx.doi.org/10.1021/jp212543m | J. Phys. Chem. B 2012, 116, 6806−6809

range of response to cellular or extracellular mechanical stimuli, which we quantify next.

We assume that each domain is folded independently from each other. The change in protein stability ΔG under force is governed by the changes in stabilities of each individual domain $\Delta G_{d,i}$ as well as of the linkers $\Delta G_{l,(i,i+1)}$:

$$\Delta G = \Delta G_{d,i} + \Delta G_{l,(i,i+1)} - F\delta\tilde{R} - T\Delta S \quad (5)$$

where $\delta\tilde{R}$ is the extension of a protein under force F that contributes to the work done in order to extend the protein. Assume that this force disrupted our protein, which contains n domains and m linkers, and N is the total number of domains (we assume $N \gg 1$). Then the resulting extension of the protein is $\delta R = nx_0 + mx_1$, where x_0 and x_1 are the average extensions of domains and linkers correspondingly (assuming that all domains extend roughly by the same length, as well as all linkers by the same length). The extension of a protein contributing to the work done to disrupt n domain and m linkers is

$$\delta\tilde{R} = nx_0\gamma_0 + mx_1\gamma_1 \quad (6)$$

where γ_0 and γ_1 are factors ($\gamma_{0,1} \leq 1$).

The entropic term in eq 5 accounts for the combinatorial number of unfolding events of n domains and m linkers: $\Delta S = Nk_B(S_N(n) + S_N(m))$, where $S_N(x) \equiv (x/N)\ln(x/N) + ((N-x)/N)\ln((N-x)/N)$.

We minimize eq 5 subject to eq 6:

$$\begin{cases} \left.\dfrac{\partial G}{\partial n}\right|_{n=n^*} = 0 \\ \left.\dfrac{\partial G}{\partial m}\right|_{m=m^*} = 0 \end{cases}$$

in order to obtain the typical number of unfolded n^* domains and m^* linkers

$$\begin{pmatrix} n^* \\ m^* \end{pmatrix} = \frac{N}{1 + \exp\left\{\beta\begin{pmatrix} \Delta G_{d,i} - Fx_0 \\ \Delta G_{l,(i,i+1)} - Fx_1 \end{pmatrix}\right\}} \quad (7)$$

and thus,

$$\frac{\delta\tilde{R}}{N} = \frac{x_0\gamma_0}{1 + \exp\{\beta(\Delta G_{d,i} - Fx_0)\}} + \frac{x_1\gamma_1}{1 + \exp\{\beta(\Delta G_{l,(i,i+1)} - Fx_1)\}} \quad (8)$$

It is clear that the major extensions of proteins occur when domains unfold, near their unfolding transition $\Delta G_{d,i} \approx Fx_0$: $\delta\tilde{R}/N \approx (x_0\gamma_0/2)[1 - \beta(\Delta G_{d,i} - Fx_0)] + [x_1\gamma_1/(1 + \exp\{\beta(\Delta G_{l,(i,i+1)} - Fx_1)\}]$, with the slope of the transition curve $(x_0^2\gamma_0/2)\beta$ (Figure 3).

While deriving eq 8, we assume that all domains and all linkers correspondingly behave similarly. If, for example, some linkers are "stronger" than others, one can generalize eq 8 to account for these stronger linkers. One would need to add an extra term on the right-hand side of eq 8, which would look similar to the last term, but account for different values of $\Delta G_{l,(i,i+1)}$, x_1, and γ_1.

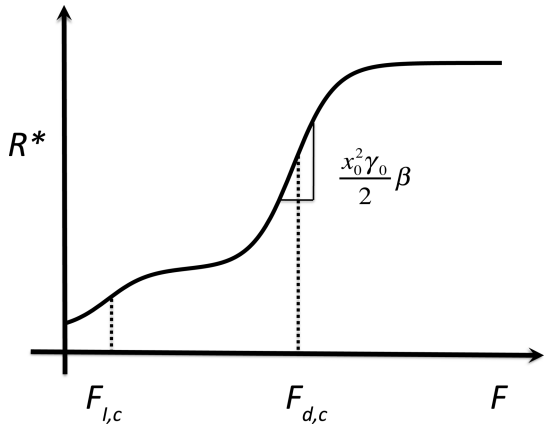


Figure 3. Schematic representation of the most likely extension of a multidomain protein as a function of force. Midpoints of the sigmoidal transitions at lower and higher forces correspond to critical forces $F_{l,c}$ and $F_{d,c}$ corresponding to linker peptide and domain unfolding correspondingly.

Typically, linkers form a smaller number of interactions than do protein domains, making their interactions more likely to be disrupted all at once than are those of proteins. Hence, the dominant contribution to multidomain protein unfolding in eq 8 at low force arises from the linkers (due the exponential nature of the contributions of corresponding terms). As the force increases, domains are more likely to unfold by disruption of the critical nucleus, which allows for a protein domain to overcome the folding free energy barrier. However, because of the multiple linkers in a multidomain protein, combinatorial unfolding of these linkers shields individual domain unfolding. At larger force, when the majority of linkers are extended, the least stable domain ("weakest link") unfolds.

Such unfolding of multidomain proteins has been extensively utilized by nature in order to respond to stress and control the activity of these proteins. Multistep unfolding provides a rich dynamic range for stress response, which is especially important for structural proteins that are responsible for the stress response in cytoskeletal remodeling. For example, filamins are thought to modulate connections between actin fibers; stress transmission to filamins allows for cells to deform without damage. Titins respond to muscle contraction at various forces.[29]

On the other hand, nature utilizes linkers to allow signal transmission only when an applied force exceeds certain limits. These signals may be transmitted from various phosphorylation, protease recognition, and binding sites that are normally protected by the more compact structures formed by linkers, but upon stress become exposed to the relevant cellular machinery. In some cases, these "encrypted" sites can appear in domains as well as linkers. For example, filamin's N-terminal β-strand of domain 20 (and possibly domain 18[30]) is in autoinhibited state when ordered, preventing interactions of other proteins with filamin A. In forced-induced unfolding, Lad et al.[31] suggested that interactions within domains are mediated by the exposed N-terminal β-strand. In filamin A domain 9, there are several phosphorylation sites on the β-strands, such as S1081 and S1084 in the loop between β-strands A and B.[32,33] When the applied force exceeds a threshold of ~35 pN,[30] these sites become exposed and can be phosphorylated. In filamins (proteins that mediate interactions between actin fibers), such

dx.doi.org/10.1021/jp212543m | J. Phys. Chem. B 2012, 116, 6806–6809

stress-dependent conformational changes allow precise control of cytoskeletal dynamics.

Shedding light on the modular organization of proteins that sense force has important implications for protein design and engineering. The rational combination of linkers and repetitive domains may allow for the design of functional proteins that will sense and report on the forces acting inside living cells.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dokh@unc.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Elizabeth A. Proctor for help with reading the manuscript. The work is supported by NIH grant R01GM080742.

REFERENCES

(1) Fletcher, D. A.; Mullins, R. D. *Nature* **2010**, *463*, 485−92.
(2) Carlsson, A. E. *Annu. Rev. Biophys.* **2010**, *39*, 91−110.
(3) Galletta, B. J.; Mooren, O. L.; Cooper, J. A. *Curr Opin Biotechnol.* **2010**, *21*, 604−10.
(4) Stossel, T. P.; Condeelis, J.; Cooley, L.; Hartwig, J. H.; Noegel, A.; Schleicher, M.; Shapiro, S. S. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 138−45.
(5) van der Flier, A.; Sonnenberg, A. *Biochim. Biophys. Acta* **2001**, *1538*, 99−117.
(6) Sweeney, H. L.; Houdusse, A. *Annu. Rev. Biophys.* **2010**, *39*, 539−57.
(7) Serohijos, A. W.; Chen, Y.; Ding, F.; Elston, T. C.; Dokholyan, N. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18540−5.
(8) Tsygankov, D.; Serohijos, A. W.; Dokholyan, N. V.; Elston, T. C. *J. Chem. Phys.* **2009**, *130*, 025101.
(9) Serohijos, A. W.; Tsygankov, D.; Liu, S.; Elston, T. C.; Dokholyan, N. V. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4840−50.
(10) Bloom, K.; Sharma, S.; Dokholyan, N. V. *Curr. Biol.* **2006**, *16*, R276−8.
(11) LeWinter, M. M.; Granzier, H. *Circulation* **2010**, *121*, 2137−45.
(12) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−8.
(13) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature* **1998**, *393*, 181−5.
(14) Arribas, S. M.; Hinek, A.; Gonzalez, M. C. *Pharmacol. Ther.* **2006**, *111*, 771−91.
(15) Tamburro, A. M. *Nanomedicine (London, U.K.)* **2009**, *4*, 469−87.
(16) Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. *Arch. Biochem. Biophys.* **2008**, *469*, 4−19.
(17) Kesner, B. A.; Milgram, S. L.; Temple, B. R.; Dokholyan, N. V. *Mol. Biol. Evol.* **2010**, *27*, 283−95.
(18) Hanke, F.; Kreuzer, H. J. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2005**, *72*, 031805.
(19) Staple, D. B.; Payne, S. H.; Reddin, A. L. C.; Kreuzer, H. J. *Phys. Biol.* **2009**, *6*, 25005.
(20) Marko, J. F.; Siggia, E. D. *Macromolecules* **1995**, *28*, 8759−8770.
(21) Li, H.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkvliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M. *Nature* **2002**, *418*, 998−1002.
(22) Shen, T.; Canino, L. S.; McCammon, J. A. *Phys. Rev. Lett.* **2002**, *89*, 068103.
(23) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6166−70.
(24) Klimov, D. K.; Thirumalai, D. *J. Phys. Chem. B* **2001**, *105*, 6648−6654.
(25) Morrison, G.; Hyeon, C.; Toan, N. M.; Ha, B.-Y.; Thirumalai, D. *Macromolecules* **2007**, *40*, 7343−7353.
(26) Oberhauser, A. F.; Marszalek, P. E.; Carrion-Vazquez, M.; Fernandez, J. M. *Nat. Struct. Biol.* **1999**, *6*, 1025−8.
(27) Lifshitz, I. M.; Grosberg, A. Yu.; Khokhlov, A. R. *Rev. Mod. Phys.* **1978**, *50*, 683−713.
(28) Doi, M. *Introduction to Polymer Physics*; Oxford University Press: New York, 1995.
(29) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−12.
(30) Kesner, B. A.; Ding, F.; Temple, B. R.; Dokholyan, N. V. *Proteins* **2010**, *78*, 12−24.
(31) Lad, Y.; Kiema, T.; Jiang, P.; Pentikainen, O. T.; Coles, C. H.; Campbell, I. D.; Calderwood, D. A.; Ylanne, J. *EMBO J.* **2007**, *26*, 3993−4004.
(32) Daub, H.; Olsen, J. V.; Bairlein, M.; Gnad, F.; Oppermann, F. S.; Korner, R.; Greff, Z.; Keri, G.; Stemmann, O.; Mann, M. *Mol. Cell* **2008**, *31*, 438−48.
(33) Dephoure, N.; Zhou, C.; Villen, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10762−7.