Internal protein dynamics shifts the distance to the mechanical transition state

Daniel K. West
School of Physics & Astronomy and
School of Biochemistry & Microbiology, University of Leeds, Leeds LS2 9JT, United Kingdom

Emanuele Paci and Peter D. Olmsted*
School of Physics & Astronomy and
Astbury Centre for Structural Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

Mechanical unfolding of polyproteins by force spectroscopy provides valuable insight into their free energy landscapes. Most phenomenological models of the unfolding process are two-state and/or one dimensional, with the details of the protein and its dynamics often subsumed into a zero-force unfolding rate and a single distance x_u^{1D} to the transition state. We consider the entire phase space of a model protein under a constant force, and show that the distance x_u^{1D} contains a sizeable contribution from exploring the full multidimensional energy landscape. Proteins with more degrees of freedom are expected to have larger values for x_u^{1D}. We show that externally attached flexible linkers also contribute to the measured unfolding characteristics.

Atomic force microscopy or optical tweezers are now routinely used to study the mechanical properties of proteins [1, 2]. An important issues is the unfolding behavior of folded domains, including the strength, and the dependence on fold topology [3] and secondary structure [3, 4]. The simplest description of unfolding treats the unfolding domain as moving in a one dimensional potential $G(x)$, where the reaction coordinate x is assumed to be directly coupled to the applied force. In perhaps the simplest approximation, the rate of unfolding $k_u(F)$ of a two-state (native and denatured) protein under force can then be calculated using Bell's formula based on Kramers' relation for escape from a well [1, 5]:

$$k_u(F) \simeq k_u^0 \exp\left(\frac{F x_u^{1D}}{k_B T}\right), \qquad (1)$$

where k_B is Boltzmann's constant, T is the absolute temperature, x_u^{1D} is the transition state displacement along the force projection, and k_u^0 is the unfolding rate constant at zero force. This result can be used to calculate the distributions of unfolding times (for applied force) or unfolding forces (for applied pulling speed), and has been applied to many different proteins [6–10]. This simple result has been corrected to use the entire shape of $G(x)$ rather than just the barrier height and displacement x_u^{1D} and incorporate the cantilever compliance [11, 12], to relax the diffusive limit to faster speeds [13], and to allow for multiple pathways or states [14, 15]. There has been extensive work on the utility of a the x_u and the simple 1D picture, linear force dependence, etc [16, 17] **[More work to do here!!!]**

However, the assumption of a one dimensional reaction coordinate grossly simplifies physical reality. The unfolding rate depends dramatically on pulling direction [18, 19], and hence on the multidimensional nature of the free energy landscape. Moreover, the one dimensional parameters have no satisfactory physical interpretation: x_u^{1D} is defined along the pulling direction, while the actual unfolding takes place along an unknown reaction coordinate(s), presumably involving a few key amino acids. While $x_u^{1D} \simeq 0.2\,\mathrm{nm}$ is the right order of magnitude for hydrogen bonds, the explicit connection with molecular configurations remains unclear. In this Letter we explore the role of the multidimensional energy landscape in determining the effective 1D unfolding parameters. The key physical ingredient is that an applied force perturbs fluctuations transverse to the forcing direction, because of the highly anharmonic nature of angular bonds. Specifically, force restricts a protein's conformational search among the dihedral states of the polypeptide backbone. We calculate this using molecular dynamics (MD) simulations of a simple off-lattice Gō model [20, 21] for a

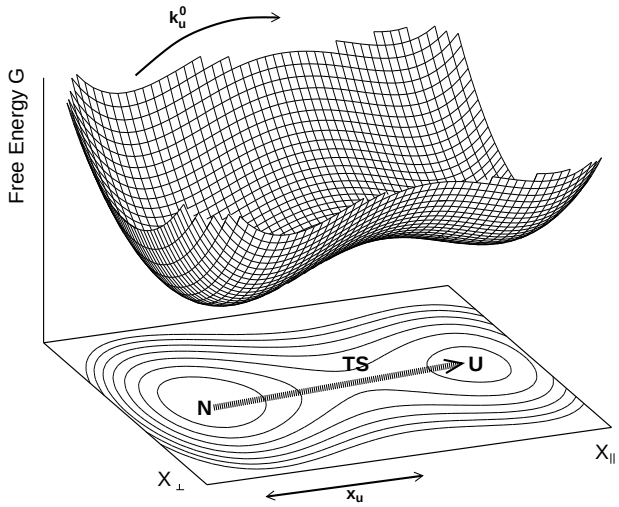

FIG. 1: The free energy surface of a two-state system, with native (N) and unfolded (U) minima. An external force F parallel to $X_\parallel$ lowers the barrier to unfolding by Fx_u.

topologically simple protein, and show that this restriction leads to a sizeable contribution to $x_u^{\rm 1D}$.

Escape from a multidimensional energy landscape— The rate of escape from an $\mathcal{N}$-dimensional energy landscape under an applied force F (Fig. 1) is given by [22]

$$k_u(F) = \Gamma(F) \exp\left(\frac{F x_u}{k_B T}\right), \tag{2}$$

$$\Gamma(F) = \frac{\sqrt{|G_{TS}^{''RC}|}}{2\pi\gamma} \frac{\prod_{k=1}^{\mathcal{N}} \sqrt{G_N^{''k}(F)}}{\prod_{k=1}^{\mathcal{N}-1} \sqrt{G_{TS}^{''k}}} \exp\left(\frac{-\Delta G_{TS-N}}{k_B T}\right), \tag{3}$$

where x_u is the distance to the transition state, γ is a friction coefficient, and ΔG_{TS-N} is the height of the free energy barrier relative to the native basin. $G_{TS}^{''RC}$ is the curvature in the unstable direction at the transition state, $G_{TS}^{''k}$ are the $\mathcal{N}-1$ stable curvatures at the transition state, and $G_N^{''k}(F)$ are the $\mathcal{N}$ positive curvatures about the native basin under an applied force. If the transition state is sharp *i.e.*, $|G_C^{''RC}| \gg F/x_u$, the curvatures $G_{TS}^{''}$ at the transition state, as well as x_u, are approximately independent of force. A more physical representation of the attempt frequency $\Gamma(F)$ follows by relating a curvature $G^{''k}$ to the associated fluctuations by $\langle \delta x_k^2 \rangle = k_B T / G^{''k}$. This yields

$$\Gamma(F) = \frac{k_B T}{2\pi\gamma l_{TS}^{RC}} \frac{V_{TS}}{V_N(F)} \exp\left(\frac{-\Delta G_{TS-N}}{k_B T}\right), \tag{4}$$

where l_{TS}^{RC} is the width of the transition state along the unfolding reaction coordinate, V_{TS} is the volume of phase space available for fluctuations at the transition state and $V_N(F)$ is the corresponding volume in the native basin.

In one dimension the weak force dependence of the prefactor $\Gamma(F)$ can be safely ignored [23]. A weak perturbation of the native basin volume $V_N(F)$ leads to

$$k_u(F) \simeq k_u^0 \exp\left(\frac{F\left[x_u + \lambda k_B T\right]}{k_B T}\right), \tag{5}$$

where $\lambda = -\partial[\ln V(F)]/\partial F|_{F=0}$. Thus, any change in the volume of the native basin with force will shift $x_u^{\rm 1D}$ in the equivalent one dimensional model,

$$x_u^{\rm 1D} = x_u + \lambda k_B T \equiv x_u + \delta x_u, \tag{6}$$

where δx_u is a *dynamical*, or entropic, contribution to the transition state displacement. If the volumes associated with different degrees of freedom randomly increased or decreased with an applied force, there would be little effect. However, we expect the volume of most perturbed degrees of freedom to decrease under an applied force, so that δx_u is proportional to the number of perturbed degrees of freedom. For a force-independent volume we recover Eq. 1, with $\delta x_u = 0$.

Calculation of phase space volumes— Phase space fluctuation volumes were calculated from MD simulation trajectories. MD was performed for protein L (PDB reference: 1HZ6 [24]) using the C_α Gō model of Refs. [20, 25]. The simulation protocol is described in detail in [3]. A real protein does not fluctuate about a single well-defined average structure; because dihedral angles typically access discrete values, the accessible states are fluctuations about many well defined structures, or *nodes* in phase space. Fig. 2 shows the phase space explored by a tetramer with 2 unimodal and 2 bimodal dihedral distributions.

The total unfolding rate $k_u^{\rm tot}(F)$ is the weighted sum of the escape rates $k_u^\beta(F)$ from all nodes $\beta = 1 \ldots M$ (assuming ΔG_{TS-N} and x_u are the same for all nodes);

$$k_u^{\rm tot}(F) = \frac{k_B T}{2\pi\gamma l_{TS}^{RC}} \frac{V_{TS}}{V_N^{\rm eff}(F)} \exp\left[-\frac{(\Delta G_{TS-N} - F x_u)}{k_B T}\right], \tag{7}$$

$$\frac{1}{V_N^{\rm eff}(F)} = \sum_{\beta=1}^{M} \frac{P^\beta(F)}{V_N^\beta(F)}, \tag{8}$$

where $V_N^\beta(F)$ is the volume and $P^\beta(F)$ the occupation probability of node β. The quantity $V_N^{\rm eff}(F)$ is the *effective* phase space volume of the native basin.

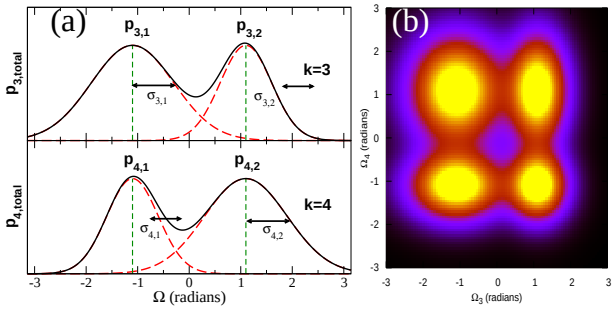

FIG. 2: The phase space of an oligomer with 4 dihedral angles: two of these are unimodal ($k = 1, 2$, not shown) and two are bimodal ($k = 3, 4$). (a) The probability distribution function $p_{k,\rm total}$ for each bi-modal dihedral angle (black solid line) can be resolved into separate distributions $p_{k,n}$ (red dashed lines) about well defined averages (green vertical dotted lines). There are four possible structures corresponding to fluctuations around $\{\bar\Omega_{3,1},\bar\Omega_{4,1}\}$,$\{\bar\Omega_{3,1},\bar\Omega_{4,2}\}$, $\{\bar\Omega_{3,2},\bar\Omega_{4,1}\}$ and $\{\bar\Omega_{3,2},\bar\Omega_{4,2}\}$ (b) The phase space projected onto $\{\Omega_3,\Omega_4\}$.

The occupation probability $P^\beta(F)$ of each node is

$$P^\beta(F) = \left\langle \prod_{k=1}^{N-3} \frac{p_{k,n_\beta}(\Omega_k(t))}{p_{k,\rm total}(\Omega_k(t))} \right\rangle, \tag{9}$$

where $\sum_{\beta=1}^{M} P^\beta(F) = 1$ and $\langle \ldots \rangle$ is the average over the MD trajectory. A node is specified by a particular set of occupancies of each dihedral angle $\mathbf{\Omega} = (\Omega_1, \ldots \Omega_{N-3})$,

where N is the number of atoms. Each term in the product is the normalised probability that, in node β, a given dihedral angle Ω_k participates in its nth dihedral state (peak) (Fig. 2a). For large numbers of nodes M a mean field approach, in which all nodes are assumed to be equally populated, works well when states are sufficiently uncorrelated in time, as in this case [26].

The volume $V_N^\beta(F)$ of fluctuations about each node β is given by $V_N^\beta(F) = \sqrt{\det \mathbf{C}^\beta}$, where $\mathbf{C_{ij}}^\beta = \langle \delta\mathbf{r}_i \delta\mathbf{r}_j \rangle_\beta$ is the covariance matrix for fluctuations $\delta\mathbf{r}_i$ in each C_α position $\mathbf{r}_i$. Here the angle brackets denote an average within node β. We calculate $\mathbf{C}^\beta$ by transforming coordinates to bond lengths, bond angles, and dihedrals angles. We ignore correlations between bond and dihedral angles, which is an excellent approximation here [26]. Hence, the effective volume of phase space is given by

$$\frac{1}{V_N^{\rm eff}(F)} \simeq \frac{1}{V_\theta(F)} \sum_{\beta=1}^{M} \frac{P^\beta(F)}{V_\Omega^\beta(F)}. \qquad (10)$$

where $V_\theta(F)$ and $V_\Omega^\beta(F)$ are the volumes of phase space explored by the bond and dihedral angles respectively.

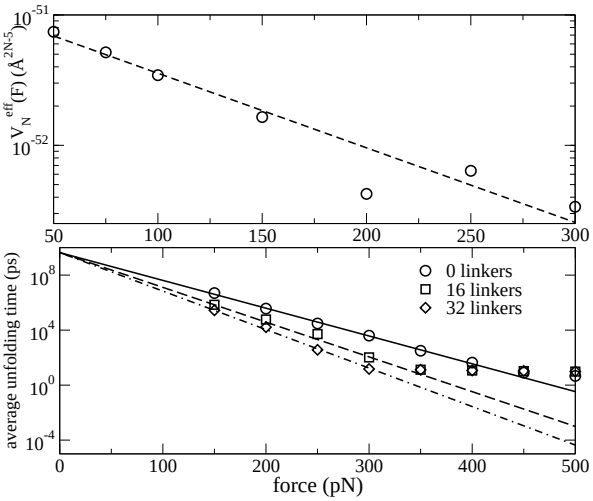

FIG. 3: (a) Effective phase space volume $V_N^{\rm eff}(F)$ for protein L (○). Dashed line is a fit to $\ln V_N^{\rm eff}(F) \simeq \ln V_N^{\rm eff}(0) - \delta x_u F/k_BT$, which yields the dynamic contribution to the transition state placement $\delta x_u = 0.054 \pm 0.008$ nm (Eq. 6). (b) Average unfolding times for protein L using MD at $T = 300$ K, with n_l attached glycine linkers (○: $n_l = 0, x_u^{\rm 1D} = 0.191 \pm 0.004$ nm, □: $n_l = 16, x_u^{\rm 1D} = 0.241 \pm 0.004$ nm, ◇: $n_l = 32, x_u^{\rm 1D} = 0.267 \pm 0.004$ nm). The linear fits yield $\log \tau = A - x_u^{\rm 1D} F/(k_BT)$. Error bars are of order the symbol size.

Results— Fig. 3a shows the phase space volume as a function of force calculated from MD simulations of protein L. The dynamic contribution to the transition state placement is $\delta x_u = 0.054 \pm 0.008$ nm. The reduction of phase space volume comes from (1) the narrowing of the dihedral distributions, and (2) the reduction in the number of multi-modal dihedral peaks. The latter effect dominates, since the loss of a single dihedral peak immediately removes many nodes of phase space. Simulations of the unfolding of the same protein L domain (Fig. 3b) yield an effective 1D transition displacement $x_u^{\rm 1D} = 0.191 \pm 0.004$ nm, from measuring an exponential dependence of the unfolding time τ_u on applied force, $\tau_u \sim e^{-Fx_u^{\rm 1D}/(k_BT)}$, as predicted from a single reaction coordinate description, Eq. (1). Hence we conclude that the bare transition state position was $x_u = 0.137$ nm, and the large shift of $\delta x_u = 0.054 \pm 0.008$ nm is between 34% and 45%.

Linker Effects— For convenience, protein domains are often pulled with long linkers, or unfolded protein strands. The linkers *also* fluctuate about discrete dihedral states when stretched. The "lumpiness" of this phase space is irrelevant for weakly stretched strands, but dominates the response for strongly stretched strands. Since force is coupled to the folded domain through the linkers, the total available phase space is the product of protein and linker phase spaces, and the measured $x_u^{\rm 1D}$ depends on the restriction of the linkers' phase space.

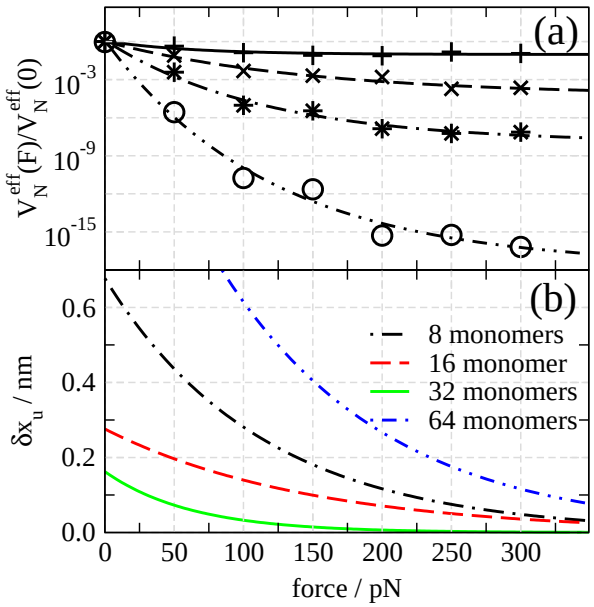

FIG. 4: (a) Volume of phase space calculated using the mean field method and (b) the dynamic contribution to the distance to the transition state $\delta x_u = -k_BT\partial[\ln V_N^{\rm eff}]/\partial F$, for strands of flexible glycine linkers. A constant force was applied for a total time of 1 μs at T=300K.

To test this, homogeneous linkers strands were constructed from a dihedral potential based on glycine. Fig. 4 shows the normalised effective phase space volume $V_N^{\rm eff}(F)/V_N^{\rm eff}(0)$ and the corresponding δx_u as a function of force for different number of atoms n_l per linker. The effect is greater for longer linkers, since more nodes are available to remove. We have ignored any force depen-

dence of the volume of the transition state V_{TS} (Eq. 4). Although we cannot easily characterize the (unstable) transition state, we can compare the shifts δx_u measured directly from the unfolding times (Fig. 3) with predictions from the phase space volumes (Fig. 4). The difference $\delta x_u(n_l = 32) - \delta x_u(n_l = 16) \simeq 0.03 - 0.05$ nm from the calculation of phase space volumes (at forces of order $200 - 300$ pN) agrees with the difference $x_u(n_l = 32) - x_u(n_l = 16) = 0.026$ nm measured from pulling simulations. This gives us confidence that for this model of protein L the transition state is sharp and its volume does not change appreciably under an applied force.

Discussion— We have shown that an externally applied force restricts a fluctuating protein's accessible phase space volume, which increases the transition state displacement x_u^{1D} in the equivalent one dimensional two-state model. This contribution can be appreciable because of the many degrees of freedom in a protein. Larger proteins have a potentially larger x_u^{1D}, depending on which degrees of freedom couple to the applied force; this will depend critically on the topology of the fold and the direction in which it is pulled. Most importantly, we predict that x_u^{1D} should be greater for proteins unfolded through longer attached linker strands. This may have biological significance; *e.g.* the long unfolded PEVK regions in titin [27] may play help modify the unfolding characteristics of titin. Finally, we note that many experiments have unfolded concatamers of multiple domains, for convenience of attachment and to generate larger statistics [9, 10, 15, 18, 19, 28–31]. We surmise that in all of these cases the dynamic contribution to x_u^{1D} was significant, and also included a contribution from already unfolded domains, which act as "linkers" for the last few domains to unfold in a given pull.

Acknowledgements— DKW acknowledges the Wellcome Trust for a PhD studentship. We thank D. J. Brockwell, J. Clarke, T. McLeish, and S. Radford for helpful discussions.

* Electronic address: p.d.olmsted@leeds.ac.uk

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).
[2] M. Rief, J. M. Fernandez, and H. E. Gaub, Phys. Rev. Lett. **81**, 4764 (1998).
[3] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci, Biophys. J. **90**, 287 (2006).
[4] V. Ortiz, S. O. Nielsen, M. L. Klein, and D. E. Discher, J. Mol. Biol. **349**, 638 (2005).
[5] G. I. Bell, Science **200**, 618 (1978).
[6] J. M. Fernandez and H. Li, Science **303**, 1674 (2004).
[7] M. Schlierf, H. Li, and J. M. Fernandez, Proc. Natl. Acad. Sci. USA **101**, 7299 (2004).
[8] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, Proc. Natl. Acad. Sci. USA **98**, 468 (2001).
[9] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford, Biophys. J. **89**, 506 (2005).
[10] D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith, and S. E. Radford, Biophys. J. **83**, 458 (2002).
[11] M. Schlierf and M. Rief, Biophys. J. **90**, L33 (2006).
[12] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).
[13] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[14] D. Bartolo, I. Derényi, and A. Ajdari, Phys. Rev. E **65**, 051910 (2002).
[15] P. M. Williams, S. B. Fowler, R. B. Best, J. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke, Nature **422**, 446 (2003).
[16] P. C. Li and D. E. Makarov, J. Chem. Phys. **121**, 4826 (2004).
[17] S. Kirmizialtin, L. Huang, and D. E. Makarov, J Chem Phys **122**, 234915 (2005).
[18] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, Nature Struct. Biol. **10**, 738 (2003).
[19] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford, Nature Struct. Biol. **10**, 731 (2003).
[20] J. Karanicolas and C. L. Brooks III, Prot. Sci. **11**, 2351 (2002).
[21] D. K. West, P. D. Olmsted, and E. Paci, J. Chem. Phys. **124**, 154909 (2006).
[22] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[23] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[24] J. W. O'Neill, D. E. Kim, D. Baker, and K. Y. Zhang, Acta Crystallogr. D Biol. Crystallogr. **57**, 480 (2001).
[25] J. Karanicolas and C. L. Brooks III, J. Mol. Biol. **334**, 309 (2003).
[26] D. K. West, E. Paci, and P. D. Olmsted, Phys. Rev. E. **in preparation**, (2006).
[27] W. A. Linke, M. Ivemeyer, P. Mundel, M. R. Stockmeier, and B. Kolmerer, Proc. Natl. Acad. Sci. USA **95**, 8052 (1998).
[28] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub, J. Mol. Biol. **286**, 553 (1999).
[29] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke, Biophys. J. **81**, 2344 (2001).
[30] M. Carrion-Vasquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, Proc. Natl. Acad. Sci. USA **96**, 3694 (1999).
[31] R. B. Best, S. Fowler, J. L. Toca-Herrera, A. Steward, E. Paci, and J. Clarke, J. Mol. Biol. **330**, 867 (2003).