Low force unfolding of a single-domain protein by parallel pathways

Pavel I. Zhuravlev,[†] Michael Hinczewski,[‡] and D.Thirumalai[¶]

†*Biophysics Program, Institute for Physical Science and Technology, Department of Chemistry & Biochemistry, University of Maryland, College Park, MD, 20742*

‡*Department of Physics, Case Western Reserve University, Cleveland, OH, 44106*

¶*Department of Chemistry, The University of Texas, Austin, TX, 7712*

E-mail:

Abstract

Deviations from linearity in the dependence of the logarithm of protein unfolding rates, $\log k_u(f)$, as a function of mechanical force, f, measurable in single molecule experiments, can arise for many reasons. In particular, upward curvature in $\log k_u(f)$ as a function of f implies that the underlying energy landscape must be multidimensional with the possibility that unfolding ensues by parallel pathways. Here, simulations using the SOP-SC model of a wild type β-sandwich protein and several mutants, with immunoglobulin folds, show upward curvature in the unfolding kinetics. There are substantial changes in the structures of the transition state ensembles as force is increased, signaling a switch in the unfolding pathways. Our results, when combined with previous theoretical and experimental studies, show that parallel unfolding of structurally unrelated single domain proteins can be determined from the dependence of $\log k_u(f)$ as a function of force (or $\log k_u[C]$ where $[C]$ is the denaturant concentration).

1

Introduction

Since the pioneering studies on forced-unfolding of titin over twenty years ago,[1–3] single molecule pulling experiments have generated a wealth of data that continue to provide unprecedented glimpses into the response of biomolecules and their complexes to mechanical forces. In a typical pulling experiment,[4–7] mechanical force (f) is applied to two points on a protein, and the response is monitored as a function of f. The experiments (optical tweezers or atomic force microscopy (AFM)) could be done at a constant f (or similarly constant extension) or a constant loading rate, r_f. The constant f experiments generate a folding trajectory in which the protein (or RNA) could hop between a number of states accessible at f. If during the observation time, which can be extremely long, the system is ergodic then both the equilibrium and f-dependent rates between all the relevant states can be computed, provided some corrections due to the experimental setup are taken into account[8,9] In the constant loading rate mode, sometimes referred to as dynamic force spectroscopy, the distribution of rupture forces is measured by linearly increasing f at a constant loading rate, $r_f = \frac{df}{dt}$. In many instances one encounters slip bonds in which plots of the most probable rupture force (f^*) as a function of $\log r_f$ ([$\log r_f, f^*$] plot) or the dependence of unfolding rate ($k_u(f)$) on f, (referred to as a [$\log k_u(f), f$] plot) are linear, which is explained using the Evans-Ritchie (for constant r_f) theory[10,11] or the Bell model (constant f).[12] However, there are a number of examples (forced unfolding of monomeric proteins[13,14] and rupture of cell adhesion and other complexes[15]), which show curvature or change in slope in the [$\log r_f, f^*$][16] and [$\log k_u(f), f$] plots. The data in these experiments could arise for the following reasons: (i) The location ($x^{\ddagger}$) of the transition state (TS) in the one dimensional free energy profiles as a function of extension, conjugate to f and accessible in pulling experiments, moves with f or r_f due to an interplay between force and curvature of the energy landscape.[17–19] (ii) The upward curvature is a consequence of unfolding occurring by parallel pathways in a strongly multidimensional landscape,[14] which is surely the case in the SH3 domain.[13] (iii) An extremely interesting and counter intuitive scenario arises when the

upward curvature is so pronounced that it gives rise to catch bonds, observed in a number of protein complexes.[20,21] In this case as well as in (ii) the extension alone cannot serve as a reaction coordinate. (iv) Rupture of protein complexes (biotin-streptavidin[16,22]) involves overcoming two sequential barriers with two TS locations, one at low forces (large value of $x^{\ddagger}$) and the other (small value of $x^{\ddagger}$) at high forces. (v) In forced-rupture of certain cell adhesion complexes, it is possible to construct a collective one dimensional coordinate,[23] which reflects the change in the slopes observed in the $[\log r_f, f^*]$ plots. These studies show that new physics of biological molecules and complexes could be teased out using the data from single molecule experiments complemented by theory and well-designed simulations.[14,24]

Here, we revisit forced-unfolding of a single domain β-sheet immunoglobulin domain (I27) (Fig. 1) in order to determine whether the low force unfolding occurs in a multidimensional landscape. It could be argued that the magnitudes of such high forces used in a majority of previous experimental and computational studies were not biologically relevant. However, the use of high forces was warranted in order to demonstrate how such experiments can probe the unfolding of a single protein under mechanical forces. Due to its mechanical resilience, I27 unfolding is not easily realized in low-force single molecule spectroscopy, either in the laser optical trap (LOT) setup, or *in silico* at room temperature in the absence of denaturants. Because $[\log k_u(f), f]$ plots for I27 were not readily available at low forces it has been difficult to determine whether forced-unfolding proceeds by parallel pathways with a switch from one to another at a critical force.

The earliest demonstration[25] that I27 unfolds by parallel pathways used the chemical denaturant guanidinium chloride (GdmCl) as a perturbation. In the wild type (WT) and several mutants, it was found that $\log k_u$ as a function of GdmCl concentration exhibited upward curvature. Using this data it was surmised that I27 must unfold by parallel pathways as GdmCl concentration is increased. Subsequently, other experiments on different proteins have established the presence of upward curvature in the unfolding kinetics as the denaturant concentration is increased.[26–29]

More recently, it has been shown using optical tweezer experiments that the $[\log k_u(f), f]$ plot for the SH3 domain exhibits upward curvature,[13] which was interpreted in the same way as in GdmCl-induced unfolding of I27. That upward curvature in $[\log k_u(f), f]$ (and by analogy any perturbation that couples linearly to the protein conformations) implies that unfolding could occur by parallel pathways was rigorously demonstrated in general and for SH3 in particular.[14]

In this paper, we report computer simulations of I27 unfolding at low forces $(0 - 50 \text{ pN})$ at elevated temperature, for the wild type and a series of single-point mutants. We show that, at least for some of the mutants of I27 the $[\log k_u(f), f]$ plots show upward curvature, suggesting that the underlying energy landscape is strongly multidimensional, especially at low forces. Elsewhere[14] we showed that if $k_u(f)$ violates either $\frac{d}{df} \log k_f \geq 0$ or $\frac{d^2}{df^2} \log k_f \leq 0$ then the underlying energy landscape is strongly multi-dimensional (SMD). Forced unfolding in such a SMD landscape likely involves a switch in the pathway at a critical force (possibly over a small force range). Below the critical force, I27 unfolds predominantly along one route and at higher forces it follows a different pathway. We analyze and compare transition state ensembles (TSE) at 5 pN and 35 pN, using the P_{fold} analysis,[31] showing the change in the preferential order in which tertiary structure elements rupture. Thus, along with switch in the pathway there is a change in the TSEs. Taken together our study shows that careful theoretical and computational analysis of data from single molecule pulling experiments could be used to determine the dimensionality of the energy landscape of proteins provided $[\log k_u(f), f]$ is known for a broad range of f.

Results

We performed molecular simulations using the coarse-grained (SOP-SC) model[32,33] of titin I27 domain (PDB 1TIT) by applying a constant force to the ends of the protein. The goal of the simulations was to explore the force range, inaccessible to either AFM (which intrinsically

uses high forces) or LOT setups (because unfolding times for I27 at low force are likely too long to be observed in LOT). The problem of very long unfolding times is particularly exacerbated in computer simulations, even using SOP-SC simulations. To alleviate this problem, we performed simulations at 400 K.

Curvature in the $[\log k_u, f]$ **plots:** We generated 100 unfolding trajectories for the WT and each of the six mutants in order to calculate the $[\log k_u(f), f]$ plots. As shown before,[14] the existence of upward curvature in these plots would be an unmistakable indication of strongly multidimensional unfolding dynamics. One consequence is parallel rupture unfolding pathways. Fig. 2 shows the $[\log k_u(f), f]$ plots for the wild-type and five of the mutants, from the set of mutants reported previously.[25] When there is no curvature, the $[\log k_u(f), f]$ is linear, and can be fit with the single-exponential Bell model, $k_u(f) = k_0 \exp\left(\frac{fx}{k_B T}\right)$]. To identify the presence of upward curvature in the plot, we compare the Bell model fit to one with a double exponential $k_u(f) = k_1 \exp\left(\frac{fx_1}{k_B T}\right) + k_2 \exp\left(\frac{fx_2}{k_B T}\right)$ using the Akaike Information Criterion (AIC).[34]

AIC is an estimator of relative quality of statistical models for a given set of data, based on information theory.[34] It provides a means for model selection by estimating how much information (from the data) is lost when using the model to represent the process which produced those data. The AIC value of a model is $AIC = 2k - 2\ln\hat{L}$, where k is the number of parameters in the model and $\hat{L}$ is the maximum value of the likelihood function of the model. The model that has the minimal AIC is most likely to lose the least information from the data, and thus is selected. If the models have values $AIC_1, AIC_2, ...,$ and we define $AIC_{\min}$ as the minimum of those values, then the quantity $\exp\left(AIC_{\min} - AIC_i\right)/2$ can be interpreted as proportional to the probability that the i-th model minimizes the estimated information loss. Then $P_2/P_1 = \frac{\exp\left(AIC_{\min} - AIC_2\right)/2}{\exp\left(AIC_{\min} - AIC_1\right)/2} = \exp\left(AIC_1 - AIC_2\right)/2$, for example, is the ratio of the probability P_2 that model 2 minimizes the information loss to the probability P_1 that model 1 minimizes the information loss. Applied to our case of choosing between the single- and double-exponential fits of $[\log k_u(f), f]$, we calculate the

5

P_2/P_1 ratio of probabilities that a double- or single- exponential model respectively minimizes the information loss.

If $P_2/P_1 > 1$, the double exponential model is more likely to minimize the information loss. If, in addition $x_2 > x_1$, we conclude that the upward curvature fit is superior, implying strong dimensionality of the unfolding process. (We avoid the temptation to over interpret the fitting parameters and use this procedure of fitting to a double-exponential model as a statistical means to establish the presence of the curvature. Double exponential, which comports with the physics of force-induced unfolding, is one of the simplest models that exhibits upward curvature in $[\log k_u(f), f])$.

At very high forces, we expect a linear $\log k_u(f)$ dependence, with the value of the slope that can be higher, lower or the same as the slopes we observe at lower forces. It is not known where this switch to the linear regime should occur for any of the constructs. In the case when the switch occurs below 50 pN (the upper limit of the curves in Fig. 2), the double exponential model fit only makes sense up to the point of the switching. The data points not contributing to the fits in Fig. 2 are shown without error bars. For L36A, G32A, L60A, C47A and F21L we used the fit up to the point (38-45 pN), visually identified as the change in slope. WT and I23A have a curvature in the $[\log k_u(f), f]$ plots even when the fitting is extended to 50 pN, according to the AIC statistic, which we use for selection of the fitting model (single vs. double exponential). C47A, F21L and G23A have curvature when fitted up to ~ 40 pN, where there is an apparent downward change in the slope. L36A and L60A do not exhibit statistically significant curvature.

Since the changes in the force field for point mutations are small, it is tempting to test for the curvature treating all of the simulations as the simulations of the same molecule, thus increasing the available statistics. The bottom right panel of Fig.2 shows the points for all the mutants on the same plot, and the black symbols represent the combined statistics. Akaike Information Criterion selects the model with curvature with very high probability, confirming the existence of parallel pathways in I27 unfolding.

Secondary and tertiary structure elements: I27 consists of eight β-strands (named A, A', B, C, D, E, F and G along the sequence). They form two β-sheets: $ABED$ and $CFGA'$ (Fig.1). Rupturing of pairs of strands can be therefore characterized by parameters $\chi_{AB}, \chi_{BE}, \chi_{ED}, \chi_{CF}, \chi_{FG}$ and $\chi_{GA'}$, where χ_{XY} is the structural overlap for two parts of the protein X and Y: the fraction of broken native contacts between X and Y,[35]

$$\chi_{XY}(\{r\}) = \frac{1}{M_{XY}} \sum_{\substack{i \in X \\ j \in Y}} \Theta\left(\left||r_i - r_j| - |r_i^0 - r_j^0|\right| - \Delta\right). \tag{1}$$

In the above equation, the summation is over the coarse-grained beads belonging to the parts X and Y, M_{XY} is the number of contacts between X and Y in the native state, $\Theta(x)$ is the Heaviside function, $\Delta = 2\text{Å}$ is the tolerance in the definition of a contact, and $r_{i,j}$ and $r_{i,j}^0$, respectively, are the coordinates of the beads in a given conformation $\{r\}$ and the native state. To compare low-force and high-force unfolding pathways we analyzed the trajectories at 5 pN and 35 pN for the I23A mutant, which does exhibit upward curvature in the $[\log k_u(f), f]$ plot.

Order of rupturing of β-strand pairs: Table 1 shows the percentage of trajectories in which one pair of β-strands (row) fully ruptured ($\chi=1$) before the other (column). The comparisons where there is a significant change between low and high force are marked in bold. Most notably, at low force the CF pair never breaks after FG and almost never after GA'. At high force, CF breaks after FG and after GA' more frequently. Also CF almost always breaks after AB at high force, but less frequently at low force. The fraction of trajectories where AB breaks after ED changes from two thirds at low force to one third at high force. With high probability, BE strand rupture precedes GA' breakage at low forces, but in a third of the trajectories GA' ruptures before BE at high forces. These observations are intuitively consistent with force being applied to the ends of the protein, A and G strands, making unfolding pathways starting with rupturing of pairs involving these two strands more accessible at high forces. In other words, high force promotes rupturing of

7

AB and GA' before other elements.

Transition state ensembles (TSEs) at high and low force: Another way to look at the unfolding pathway is to identify the transition state ensemble (TSE). We define TSE through P_{fold},[31] which is the probability to reach the folded state before the unfolded state, starting from a given conformation. If $P_{fold} = 0.5$ for some conformation, then this conformation belongs to the TSE. We select TSE candidates from the saddle point of the (χ, E) free energy surface, where E is energy and χ is the full overlap parameter (χ_{XY} where both X and Y correspond to the whole protein) We obtain the free energy surface by calculating the histograms of the unfolding trajectories at 5 and 35 pN. Running 10 trajectories for each of the candidates, we select a smaller set of candidates where $0.3 \leq P_{fold} \leq 0.7$, and then run 50 more trajectories for each candidate in the smaller set. We obtain the TSE as the ensemble of candidates with $0.4 < P_{fold} < 0.6$ calculated from the 60 trajectories for each structure. The TSEs for low and high force unfolding (TSE-L and TSE-H) are shown in Fig. 3 (5 pN) and Fig. 4 (35 pN). Visually analyzing the structures, one can see CF to be more disrupted in TSE-H. Also, there are structures in TSE-H with AB fully ruptured, but none in TSE-L. To quantify these observations (along with the same observations about other structural elements) we calculated pairwise χ for each of the TSEs (Tables 1, 2, and Fig. 5). From the data, one can conclude that CF and AB are less structured in TSE-H (and there are structures in the ensemble where one of them is fully ruptured, but not both) than TSE-L. On the other hand, FG and GA' are more structured in TSE-H than in TSE-L. These observations indicate that force induced unfolding of I27 is heterogeneous, rather than having a dominant pathway, and the details of the heterogeneity (e.g. contribution of different pathways) changes with the applied force.

Conclusion

We have shown that at low forces, accessible in LOT experiments, the unfolding of I27 does not occur over a single pathway, but is heterogeneous. The finding that low forced unfolding must occur by parallel pathways is based on the the observation of upward curvature in the $[\log k_u(f), f]$ plot and by establishing that TSE structures, determined without assuming an underlying reaction coordinate, change as f is increased. Our prediction can be confirmed experimentally, by measuring the f-dependence of $\log k_u(f)$ plot, like we did in simulations.

Our simulations provide structural insights into heterogeneities and differences of unfolding pathways at low and high forces. We predict that single molecule spectroscopy of the mechanically resilient titin I27 domain (both wild types and mutants) at low forces and elevated temperature should provide evidence that the energy landscape is strongly multidimensional. In addition, we show that there ought to be a switch in the folding pathway and associated changes in the structures of the transition state ensemble. Because it has not been done, we assume that varying temperature by a significant amount in optical tweezer experiments must be difficult. This could be mitigated by performing pulling experiments in the presence of small amount of a mild denaturant (urea) as has been done for the SH3 domain.[36]

Methods

We used the SOP-SC model,[37,38] which is a coarse-grained model for proteins with two interaction beads per residue to simulate force-induced unfolding of I27. The interactions are based on the native structure, but the amino acid identity of each residue is also taken into account. A mutation is applied to the PDB structure of the wild type, so the interactions are sorted into native and non-native subsequently. Thus, the set of native contacts remains the same, but a few interaction parameters change because of mutations. The interactions between the side chains are given by the statistical potential.[39] The details of the model and

the simulations may be found elsewhere.[38]

We used maximum likelihood fitting to discriminate between a double exponential and single exponential (Bell) models for the dependence of the unfolding rate $k_u(f)$ on force. We compared the model fits by the Akaike Information Criterion[34] in order to find out whether there is upward curvature in the $[\log k_u, f]$ plot, which would establish the multidimensionality of energy landscape leading to parallel unfolding pathways.

References

(1) Rief, M, Gautel, H, Oesterhelt, F, Fernandez, J. M, & Gaub, H. E. (1997) Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM *Science* **276**, 1109–1111.

(2) Kellermeyer, M. Z, Smith, S. B, Granzier, H. L, & Bustamante, C. (1997) Folding-Unfolding Transitions in Single Titin Molecules Characterized by Force-Measuring Laser Tweezers *Science* **276**, 1112–1116.

(3) Tskhovrebova, L, Trinick, J, Sleep, J, & Simmons, R. (1997) Elasticity and unfolding of single molecules of the giant muscle protein titin *Nature* **387**, 308–312.

(4) Gebhardt, J. C. M, Bornschlögl, T, & Rief, M. (2010) Full distance-resolved folding energy landscape of one single protein molecule *Proc. Natl. Acad. Sci. USA* **107**, 2013–2018.

(5) Cecconi, C, Shank, E. A, Bustamante, C, & Marqusee, S. (2005) Direct Observation of Three-State Folding of a Single Protein Molecule *Science* **309**, 2057–2060.

(6) Woodside, M. T, Garcia-Garcia, C, & Block, S. M. (2008) Folding and unfolding single rna molecules under tension *Curr. Opin. Chem. Biol.* **12**, 640–646.

(7) Zhuang, X & Rief, M. (2003) Single-molecule folding *Curr. Opin. Struct. Biol.* **13**, 86–97.

(8) Hinczewski, M, Gebhardt, J. C. M, Rief, M, & Thirumalai, D. (2013) From mechanical folding trajectories to intrinsic energy landscapes of biopolymers *Proc. Natl. Acad. Sci.* **110**, 4500–4505.

(9) Ramm, B, Stigler, J, Hinczewski, M, Thirumalai, D, Herrmann, H, W, G, & Rief, M. (2014) Sequence-resolved free energy profiles of stress-bearing vimentin intermediate filaments *Proc. Natl. Acad. Sci.* **111**, 11359–11364.

(10) Evans, E & Ritchie, K. (1999) Strength of a Weak Bond Connecting Flexible Polymer Chains *Biophys. J.* **76**, 2439–2447.

(11) Evans, E. (2001) Probing the Relation between Force-Lifetime-and Chemistry in Single Molecular Bonds *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105–128.

(12) Bell, G. I. (1978) Models for the specific adhesion of cells to cells *Science* **200**, 618–627.

(13) Jagannathan, B, Elms, P. J, Bustamante, C, & Marqusee, S. (2012) Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein *Proc. Natl. Acad. Sci.* **109**, 17820–17825.

(14) Zhuravlev, P. I, Hinczewski, M, Chakrabarti, S, Marqusee, S, & Thirumalai, D. (2016) Force-dependent switch in protein unfolding pathways and transition-state movements *Proc. Natl. Acad. Sci.* **113**, E715–E724.

(15) Wojcikiewicz, E. P, Abdulreda, M. H, Zhang, X, & Moy, V. T. (2006) Force spectroscopy of LFA-1 and its ligands, ICAM-1 and ICAM-2 *Biomacromolecules* **7**, 3188–3195.

(16) Merkel, R, Nassoy, P, Leung, A, Ritchie, K, & Evans, E. (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy *Nature* **397**, 50–53.

(17) Hyeon, C & Thirumalai, D. (2003) Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl. Acad. Sci. USA* **100**, 10249–10253.

(18) Dudko, O. K, Hummer, G, & Szabo, A. (2006) Intrinsic rates and activation free energies from single-molecule pulling experiments *Phys. Rev. Lett.* **96**, 108101.

(19) Dudko, O. K, Hummer, G, & Szabo, A. (2008) Theory, Analysis, and Interpretation of Single-Molecule Force Spectroscopy Experiments *Proc. Natl. Acad. Sci. USA* **105**, 15755–15760.

(20) Marshall, B. T, Long, M, Piper, J. W, Yago, T, McEver, R. P, & Zhu, C. (2003) Direct observation of catch bonds involving cell-adhesion molecules *Nature* **423**, 190–193.

(21) Chakrabarti, S, Hinczewski, M, & Thirumalai, D. (2017) Phenomenological and microscopic theories for catch bonds *J. Struct. Biol.* **197**, 50–56.

(22) Hyeon, C & Thirumalai, D. (2012) Multiple barriers in forced rupture of protein complexes *J. Chem. Phys.* **137**, 055103.

(23) Toan, N & Thirumalai, D. (2018) Forced-rupture of cell-adhesion complexes reveals abrupt switch between two brittle states *J. Chem. Phys.* **148**, 123332.

(24) Hinczewski, M, Hyeon, C, & Thirumalai, D. (2016) Directly measuring single-molecule heterogeneity using force spectroscopy *Proc. Natl. Acad. Sci.* **113**, E3852–E3861.

(25) Wright, C. F, Lindorff-Larsen, K, Randles, L. G, & Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. & Mol. Biol.* **10**, 658–662.

(26) Aksel, T & Barrick, D. (2014) Direct Observation of Parallel Folding Pathways Revealed Using a Symmetric Repeat Protein System *Biophys. J.* **107**, 220–232.

(27) Jha, S. K & Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein *Proc. Natl. Acad. Sci.* **106**, 12289–12294.

(28) Aghera, N & Udgaonkar, J. B. (2013) The Utilization of Competing Unfolding Pathways of Monellin Is Dictated by Enthalpic Barriers *Biochemistry* **52**, 5770–5779.

(29) Zaidi, F, Nath, U, & Udgaonkar, J. (1997) Multiple intermediates and transition states during protein unfolding *Nat. Struct. & Mol. Biol.* **4**, 1016–1024.

(30) Lalng, C & Schlick, T. (2010) Computational approaches to 3D modeling of RNA *J. Phys.: Condens. Matter* **22**, 283101.

(31) Du, R, Pande, V. S, Grosberg, A. Y, Tanaka, T, & Shakhnovich, E. I. (1998) On the transition coordinate for protein folding *J. Chem. Phys.* **108**, 334.

(32) Hyeon, C, Dima, R. I, & Thirumalai, D. (2006) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins *Structure* **14**, 1633–1645.

(33) Liu, Z & Thirumalai, D. (2018) Denaturants Alter the Flux through Multiple Pathways in the Folding of PDZ Domain *J. Phys. Chem. B* **122**, 1408–1416.

(34) Burnam, K & Anderson, D. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* (Springer, New York).

(35) Guo, Z & Thirumalai, D. (1995) Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways *Biopolymers* **36**, 83–102.

(36) Guinn, E. J, Jagannathan, B, & Marqsee, S. (2015) Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein *Nat. Comm.* **6**, 6861.

(37) Liu, Z, Reddy, G, O'Brien, E. P, & Thirumalai, D. (2011) Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc Natl Acad Sci USA* **108**, 7787–7792.

(38) Zhuravlev, P. I, Reddy, G, Straub, J. E, & Thirumalai, D. (2014) Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins. *J Mol Biol* **426**, 2653–2666.

(39) Betancourt, M & Thirumalai, D. (1999) Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes *Prot. Sci.* **8**, 361–369.

Table 1: Percentage of trajectories where the row contacts fully break after column contacts, at 5pN; at 35 pN

	χ_{AB}	χ_{BE}	χ_{ED}	χ_{CF}	χ_{FG}	$\chi_{GA'}$
χ_{AB}		14; 4	**66; 34**	**24; 4**	0; 0	2; 2
χ_{BE}	86; 96		98; 94	62; 56	2; 8	**4; 30**
χ_{ED}	34; 66	2; 6		12; 14	0; 2	2; 8
χ_{CF}	76; 96	38; 44	88; 86		**0; 8**	**6; 22**
χ_{FG}	100; 100	97; 91	100; 97	100; 91		71; 79
$\chi_{GA'}$	97; 98	95; 68	97; 92	93; 78	28; 20	

Table 2: Percentage of trajectories where the row contacts break 95% after column contacts, at 5pN; at 35 pN

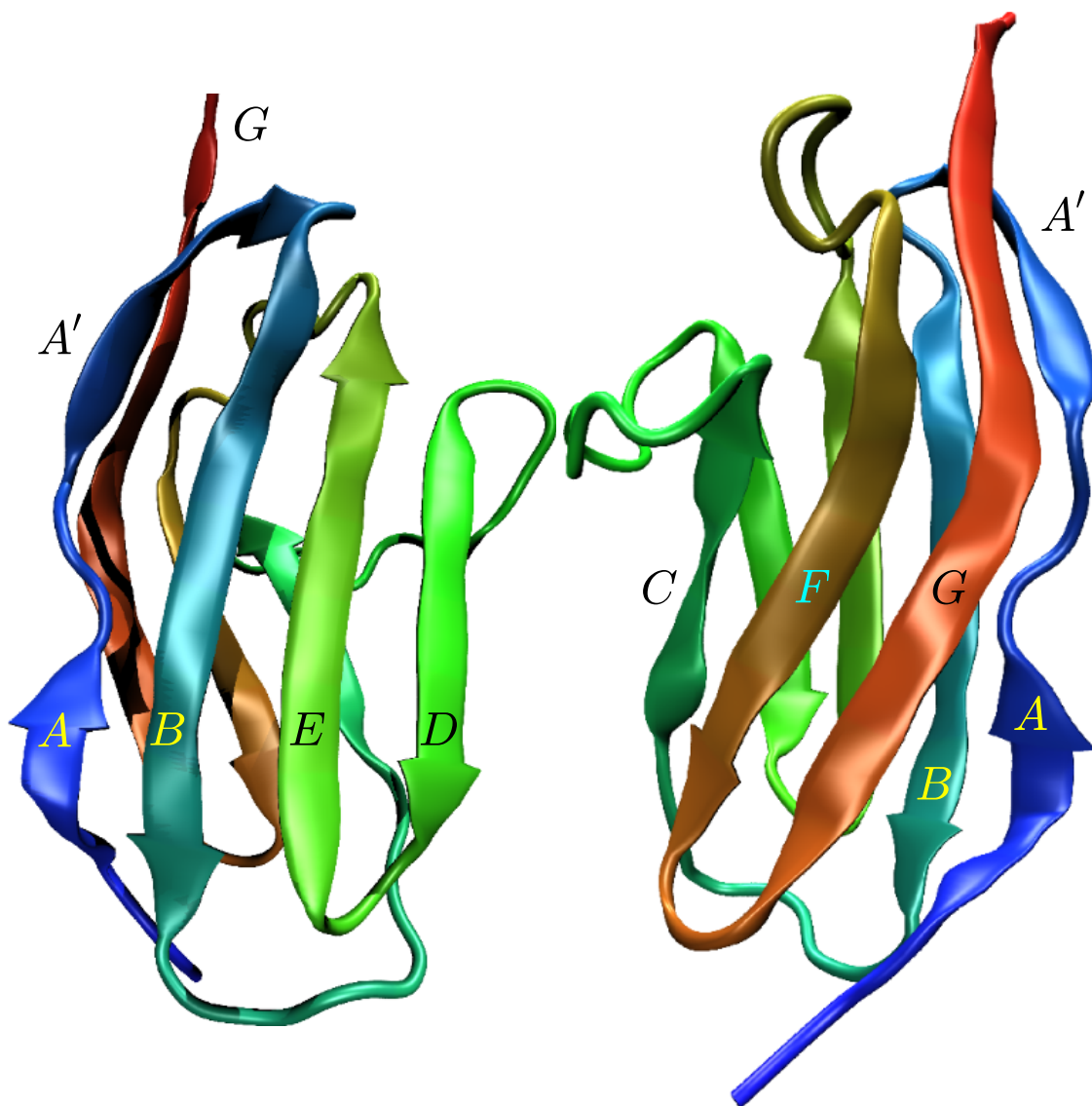|	χ_{AB}	χ_{BE}	χ_{ED}	χ_{CF}	χ_{FG}	$\chi_{GA'}$
χ_{AB}		16; 4	80; 42	20; 6	0; 0	4; 4
χ_{BE}	84; 96		100; 96	44; 52	0; 16	6; 28
χ_{ED}	20; 57	0; 4		2; 10	0; 0	4; 4
χ_{CF}	80; 94	56; 48	98; 90		0; 8	4; 24
χ_{FG}	100; 100	100; 83	100; 100	100; 91		56; 73
$\chi_{GA'}$	96; 96	94; 72	96; 96	96; 74	44; 26	

Figure 1: I27 is a 89 residue single domain protein consisting of two β-sheets $ABED$ and $CFGA'$, made up of eight β-strands A, A', B, C, D, E, F and G. Two orientations are shown for clarity.
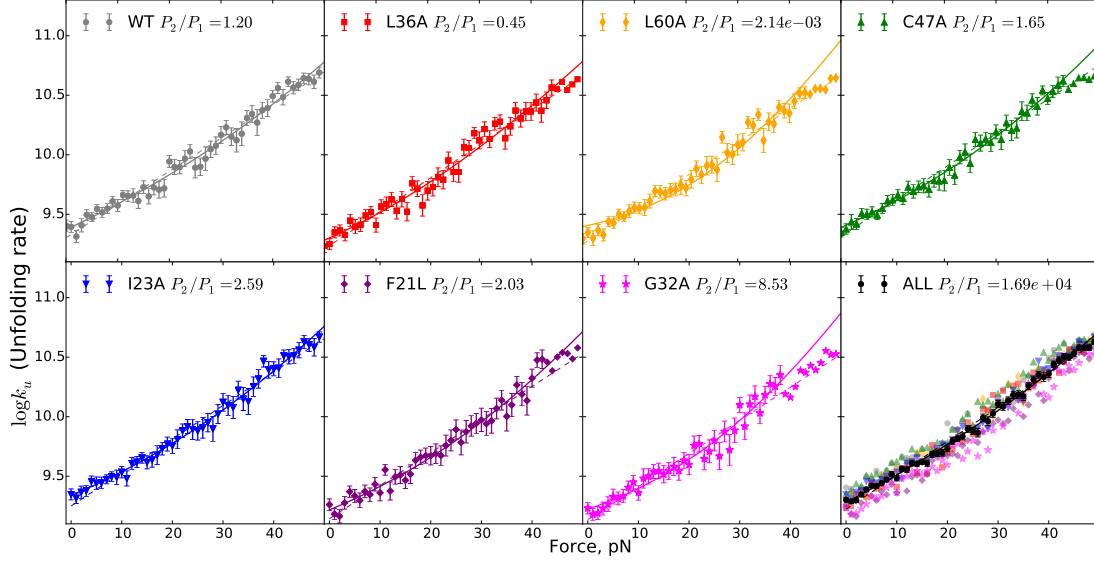
Figure 2: The unfolding rates of titin I27 domain and its mutants. Each point is calculated from the average unfolding time over 50 trajectories. For each mutant, the $[\log k_u(f), f]$ dependence is fitted with a single- (dashed lines) and a double-exponential (solid lines) model. The selection of fitting model is advised by the Akaike Information Criterion. The ratio of probabilities that particular model captures the data better than the other one (P_2/P_1 with P_2 (P_1) being the double (single) exponential fit) are given in the legends. Only L36A and L60A mutants data is better explained by single exponential fit. (Note: For L36A,G32A,L60A,C47A,F21L there is another change in slope around 35-40 pN, hence only the part of the dependence up to that point is fitted. The points that do not contribute to the shown fit and Akaike probability ratio calculations are shown without the error bars. However, using all the data (i.e. up to 50 pN) to fit these five mutants, results in single exponential model being selected (i.e. no curvature), leaving only WT and I23A with curvature.) The last panel uses all the data points, as if they were all for the same molecule, as a proxy (with the argument that changes to the force field for simulating different mutants are tiny) to assess the presence of upward curvature following from tertiary structure of titin I27 domain. With seven times more statistics, the Akaike model selection is strengthened and much more significant statistically, rather than disappeared, as evidenced by a four orders of magnitude higher P_2/P_1.

$f = 5$ pN

Ensemble
$P_{fold} = 0.51 \pm 0.06$

$P_{fold} = 0.48$

$P_{fold} = 0.52$

$P_{fold} = 0.50$

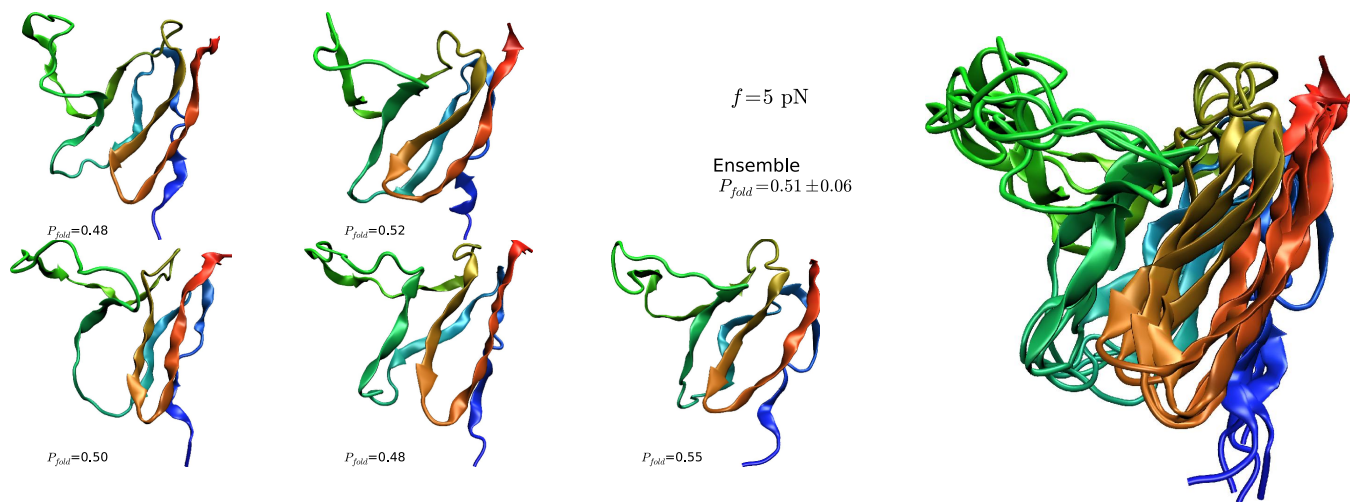$P_{fold} = 0.48$

$P_{fold} = 0.55$

Figure 3: Representative TS structures and the calculated P_{fold} (below each structure) at 5 pN. The calculated value for the TSE is $P_{\text{fold}} = 0.51 \pm 0.06$. $P_{\text{fold}} = 0.51 \pm 0.05$. Superposition of structures in the transition state ensemble.
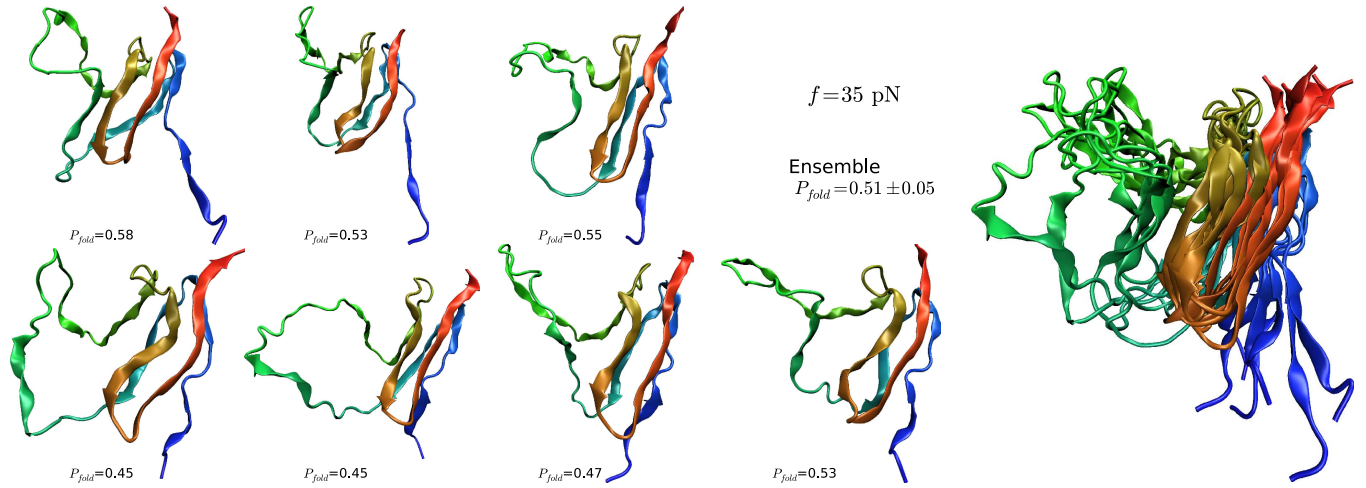
$f = 35$ pN

Ensemble
$P_{fold} = 0.51 \pm 0.05$

$P_{fold} = 0.58$ $P_{fold} = 0.53$ $P_{fold} = 0.55$

$P_{fold} = 0.45$ $P_{fold} = 0.45$ $P_{fold} = 0.47$ $P_{fold} = 0.53$

Figure 4: Same as Fig.3 except the value of the force is 35 pN. The average of $P_{\text{fold}} = 0.51 \pm 0.05$
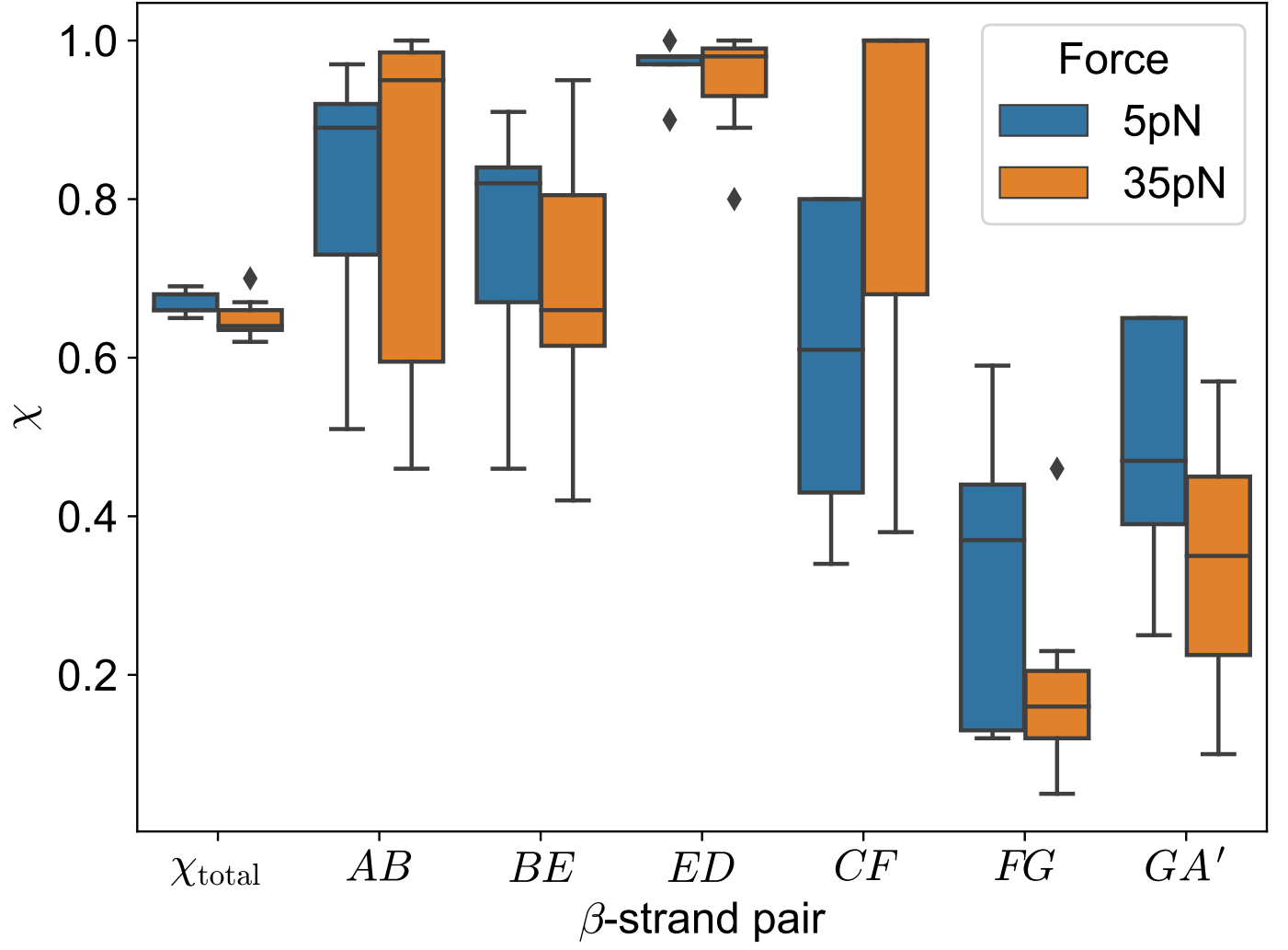
Figure 5: The comparison of ruptured tertiary structure in TSE at 5 and 35 pN (box plot)