

The folding cooperativity of a protein is controlled by its chain topology

Elizabeth A. Shank^{1,2,*†}, Ciro Cecconi^{1,2,*†}, Jesse W. Dill^{2,3,*}, Susan Marqusee^{1,2} & Carlos Bustamante^{1,2,4,5,6}

The three-dimensional structures of proteins often show a modular architecture comprised of discrete structural regions or domains. Cooperative communication between these regions is important for catalysis, regulation and efficient folding; lack of coupling has been implicated in the formation of fibrils and other misfolding pathologies¹. How different structural regions of a protein communicate and contribute to a protein's overall energetics and folding, however, is still poorly understood. Here we use a single-molecule optical tweezers approach to induce the selective unfolding of particular regions of T4 lysozyme and monitor the effect on other regions not directly acted on by force. We investigate how the topological organization of a protein (the order of structural elements along the sequence) affects the coupling and folding cooperativity between its domains. To probe the status of the regions not directly subjected to force, we determine the free energy changes during mechanical unfolding using Crooks' fluctuation theorem. We pull on topological variants (circular permutants) and find that the topological organization of the polypeptide chain critically determines the folding cooperativity between domains and thus what parts of the folding/unfolding landscape are explored. We speculate that proteins may have evolved to select certain topologies that increase coupling between regions to avoid areas of the landscape that lead to kinetic trapping and misfolding.

Most mechanical unfolding studies have focused on small globular domains, using either the atomic force microscope (AFM)^{2–4} or optical tweezers⁵. By grabbing the molecule at specific residues, one can select both the unfolding reaction coordinate and the region of the molecule subjected to mechanical force. AFM studies designed to pull along different molecular axes have demonstrated that the orientation of the applied force can determine the accessibility to different unfolding pathways^{6–8}. As we show here, mechanical manipulation provides a unique method to induce the selective unfolding of a particular region of a protein, making it possible to characterize its effect on the rest of the molecule and to investigate the importance of protein topology on the folding cooperativity and communication between domains.

We chose to study the unfolding of a cysteine-free variant of T4 lysozyme (*T4L), a protein composed of two globular regions, which we refer to as domains (Fig. 1). These domains are discontinuous in sequence: the carboxy-terminal domain (residues 60–164) also contains the amino-terminal helix A (residues 1–12) (Fig. 1c). This re-entrant topology, termed a 'discontinuous subdomain'⁹, is not uncommon among enzymes. Although traditional ensemble studies show that *T4L folds as a cooperative two-state protein¹⁰, a variety of spectroscopic techniques indicate some degree of structural and

energetic independence between these domains^{11–15}. In this study, we investigate the coupling between these two regions by exploiting our ability to unfold each domain selectively through the choice of pulling positions and the use of circular permutants^{15,16}.

We constructed optical tweezer samples by attaching double-stranded DNA molecular 'handles'^{17,18} via thiol chemistry to two cysteines engineered at specific sites where force is to be applied (see Supplementary Information and Supplementary Table 1). A construct with cysteines at residues 16 and 159 (16,159 WT*T4L), was designed to apply force across both protein domains. Force ramp studies show that

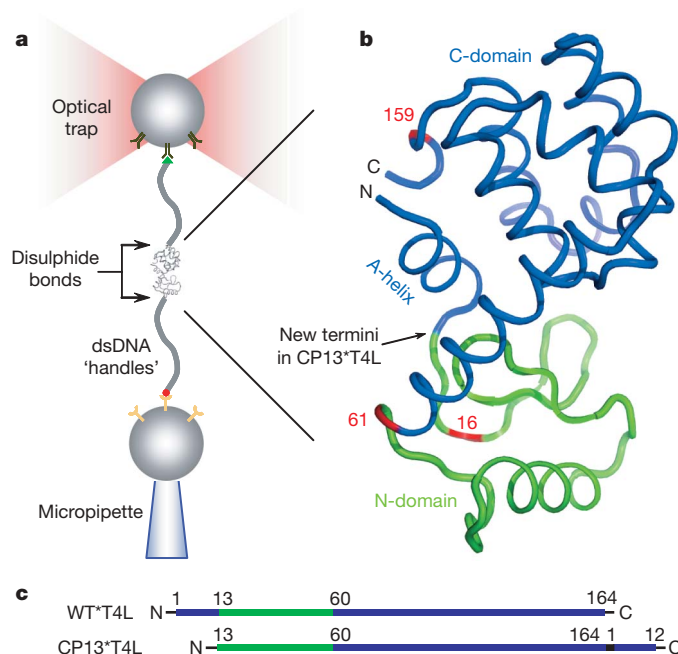


Figure 1 | Experimental set-up and structure of *T4L variants. **a**, Schematic of the optical tweezer experimental set-up. The protein is tethered between two polystyrene beads, one held by suction to a micropipette and the other trapped in an optical trap. The protein contains cysteines that permit covalent attachment to two ~500 base pair dsDNA molecular handles through disulphide linkages. The dsDNA is attached to the micropipette bead via biotin/streptavidin and to the optical trap bead through a digoxigenin/anti-digoxigenin antibody association. By moving the two beads relative to each other, force and extension can be applied to the protein. **b**, Three-dimensional protein structure of T4L coloured blue and green to distinguish the two energetically distinct C- and N-domains, respectively¹⁴. The sites of DNA attachment (residues 16, 61 and 159) are coloured red. **c**, Schematic of WT*T4L and CP13*T4L variants coloured as in **b**.

¹Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA. ²Jason L. Choy Laboratory of Single Molecule Biophysics, Institute for Quantitative Biosciences, University of California, Berkeley, California 94720, USA. ³Biophysics Graduate Group, University of California, Berkeley, California 94720, USA. ⁴Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ⁵Department of Physics, University of California, Berkeley, California 94720, USA. ⁶Department of Chemistry, University of California, Berkeley, California 94720, USA. [†]Present addresses: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA (E.A.S.); Department of Physics, University of Modena and Reggio Emilia, Via Campi 213/A 4100 Modena, Italy (C.C.).

*These authors contributed equally to this work.

unfolding occurs at high forces (30–50 pN), whereas refolding is at approximately 5 pN (Fig. 2a; Supplementary Fig. 1); under these conditions, the unfolding/refolding cycle is not an equilibrium process. The sizes of the unfolding transitions, assuming a worm-like chain model and a persistence length of 0.65 nm (ref. 19), are consistent with complete unfolding between the attachment points (the change in contour length, $\Delta L_{c<meas>} = 47 \pm 2$ nm ($n = 68$); $\Delta L_{c<calc>} = 47.93$ nm: 143 residues \times 0.36 nm per residue – 3.55 nm (the distance between residues 16 and 159 in the folded protein)). These force-extension curves yield an unfolding rate extrapolated to zero force of $k_u^0 = 1.4 \times 10^{-3} \pm 7 \times 10^{-4} \text{ s}^{-1}$ and a distance to the transition state $\Delta x_{N \rightarrow U}^{\ddagger} = 0.64 \pm 0.05$ nm (Supplementary Table 2), similar to those obtained in previous AFM pulling studies across the whole molecule^{20,21}.

Next, we investigated the coupling between the two domains. Unlike most denaturing methods, which act globally, mechanical manipulation can be used to apply force solely to one domain. Thus, we changed the attachment points to apply force exclusively on the N-domain (16,61 WT*T4L)^{14,16} (Fig. 2b). Although a much smaller portion of the protein backbone is forced open ($\Delta L_{c<meas>} = 14 \pm 1$ nm ($n = 99$); $\Delta L_{c<calc>} = 14.75$ nm) compared to 16,159 WT*T4L, the forces at which this construct unfolds are similar and even somewhat higher than those of 16,159 WT*T4L at all pulling speeds (Fig. 3). This behaviour translates into comparable distances to the unfolding transition state ($\Delta x_{N \rightarrow U}^{\ddagger} = 0.49 \pm 0.05$ nm) and extrapolated zero-force unfolding rates ($k_u^0 = 1.3 \times 10^{-3} \pm 8 \times 10^{-4} \text{ s}^{-1}$) (Supplementary Table 2), and indicates that these two constructs might unfold over similar trajectories. Unfortunately, the experimental observable—the change in extension upon unfolding—only yields information about the conformation of the amino acids between the pulling points, and therefore we cannot obtain any direct information about the conformation of the C-domain in this construct.

We surmise, however, that if the C-domain unfolds during the observed mechanical unfolding of 16,61 WT*T4L, the overall free energy change of the transition should reflect the contribution of both domains and not merely that of the N-domain. It is impossible, however, to extract free energies directly from the work done unfolding the molecule along these non-equilibrium trajectories. We therefore turned to Crooks' fluctuation theorem (CFT)²², a statistical mechanical result shown to be a powerful tool for obtaining unfolding free energies ($\Delta G_{<CFT>}$) from an ensemble of single molecule

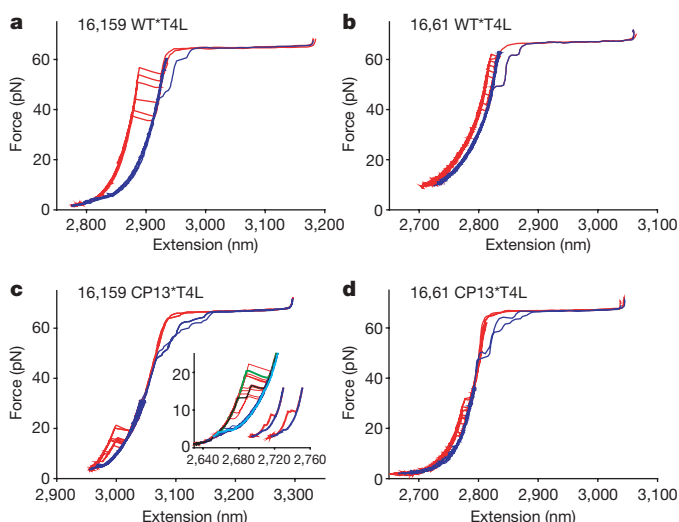


Figure 2 | Unfolding and refolding force-extension curves of WT*T4L and CP13*T4L variants. Representative force-extension curves from **a**, 16,159 WT*T4L; **b**, 16,61 WT*T4L; **c**, 16,159 CP13*T4L and **d**, 16,61 CP13*T4L (red, unfolding; blue, refolding). All data shown were collected using a 50 Hz sampling rate and a pulling speed of 180 nm s^{-1} (except for 16,61 WT*T4L, which were collected using a pulling speed of 60 nm s^{-1}). The insert in **c** shows the curves for 16,159 CP13*T4L in finer detail, illustrating the diverse unfolding and refolding behaviour. These curves were analysed as described previously¹⁷.

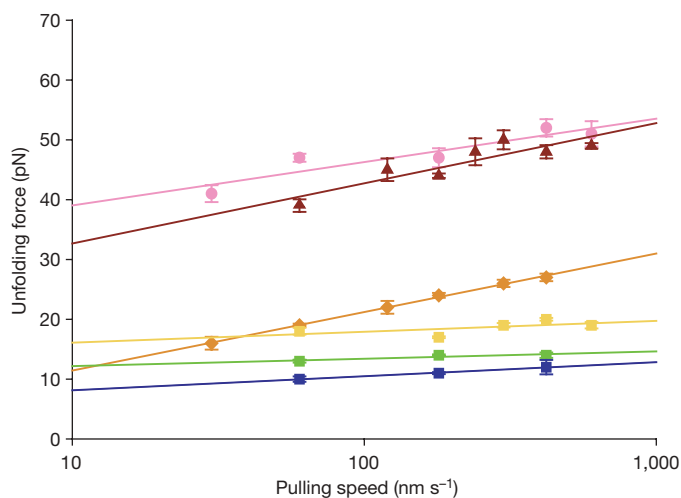


Figure 3 | The dynamic force spectrum of *T4L variants. The protein unfolding force dependence on experimental pulling speed. All points represent the average unfolding force measured at each pulling speed and the error bars are the standard error of the mean of the measured unfolding forces; the regression lines are included to guide the eye. 16,61 WT*T4L (pink circles, 460 total unfolding events from 48 individual proteins); 16,159 WT*T4L (maroon triangles, 1,189 total unfolding events from 47 proteins); 16,61 CP13*T4L (orange diamonds, 877 total unfolding events from 30 proteins); and 16,159 CP13*T4L full protein (yellow squares), C-domain (green squares) and N-domain (blue squares) (1,869 total unfolding events from 38 proteins).

mechanical unfolding trajectories in which the system is far from equilibrium^{23,24}. Application of CFT requires repeated determination of the work needed to unfold and refold the protein by integrating the force-extension curve with bounds that enclose the observed transitions. This integration, however, also includes the work carried out to reversibly stretch the DNA handles and the unfolded protein. Here, we implemented a new method that mathematically eliminates this offset (see Supplementary Information).

Application of CFT to 170 unfolding and 110 folding events for 16,61 WT*T4L yielded a value for $\Delta G_{<CFT>} = 12.3 \pm 0.6 \text{ kcal mol}^{-1}$ (Fig. 4a and Supplementary Fig. 2). In comparison, the $\Delta G_{<bulk>}$ from ensemble equilibrium denaturation experiments is $14.1 \pm 0.8 \text{ kcal mol}^{-1}$ (ref. 15). The similarity between these values indicates that when WT*T4L is pulled from residues 16 and 61 the entire protein unfolds—even though the attachment points flank only the N-domain—and indicates a high degree of coupling between the domains. Notably, this powerful non-equilibrium approach made it possible to infer the folding status of regions of the protein not bounded by the pulling points.

What structural element mediates this domain coupling? One obvious candidate is the topologically re-entrant helix A. To test this hypothesis we used a circular permutant, CP13*T4L (ref. 15), in which the first twelve residues are attached to the C terminus, creating two continuous domains discretely segregated in the sequence (Fig. 1c). This topological variant selectively alters the physical connectivity of the backbone, leaving all of the native interactions intact. CP13*T4L is active, folds cooperatively, and possesses a global stability and a structure very similar to those of WT*T4L (refs 15, 16). Native state hydrogen exchange studies indicate that the C-domain is energetically less coupled to the rest of the protein¹⁶.

Force application across both domains in 16,159 CP13*T4L resulted in more complex unfolding trajectories than its non-permuted counterpart (compare Fig. 2c to Fig. 2a). Three classes of unfolding transitions were observed, each associated with different structural transitions on the basis of ΔL_c values (Supplementary Fig. 3a). Individual protein molecules exhibited all three of these unfolding transitions in different relaxation and stretching cycles. The first class ($n = 741$) corresponds to the complete unfolding of the protein: a single rip with $\Delta L_c = 46 \pm 3$ nm. The second class ($n = 309$) also corresponds to

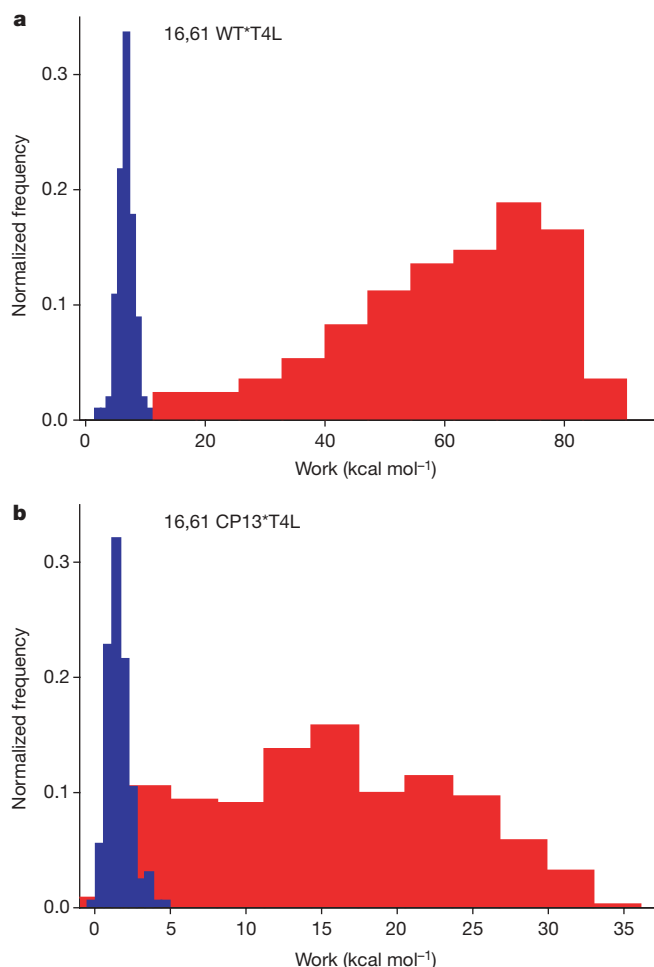


Figure 4 | The normalized probability curves of unfolding and refolding work of 16,61 WT*T4L and CP13*T4L. The normalized probability curves of the work required for the unfolding (in red) and refolding (in blue) measured as described in the main and Supplementary texts. The CFT was used to calculate the free energy from the single molecule experiments. **a**, The calculated free energy for 16,61 WT*T4L ($\Delta G_{\text{CFT}} = 12.3 \pm 0.6 \text{ kcal mol}^{-1}$) agrees well with the free energy measured in bulk solution unfolding experiments ($\Delta G_{\text{bulk}} = 14.1 \pm 0.6 \text{ kcal mol}^{-1}$) (ref. 15). **b**, The calculated free energy of 16,61 CP13*T4L ($\Delta G_{\text{CFT}} = 3.6 \pm 0.2 \text{ kcal mol}^{-1}$) corresponds to the free energy of the N-domain measured in bulk solution using native state hydrogen exchange ($\Delta G_{\text{NSHX}} = 6.1 \pm 1.0 \text{ kcal mol}^{-1}$) (ref. 14).

complete unfolding of the molecule, but with two rips of unequal size in rapid succession ($\Delta L_c = 17 \pm 2 \text{ nm}$ and $\Delta L_c = 28 \pm 3 \text{ nm}$); these double-rip transitions are consistent with unfolding of the N-domain followed by unfolding of the C-domain (Supplementary Figs 3a, 4, 5). The third, minor class of unfolding transitions ($n = 181$), consisted of a single rip with $\Delta L_c = 28 \pm 3 \text{ nm}$, consistent with unfolding of just the C-domain, indicating that during the previous stretch/relaxation cycle only the C-domain refolded (Supplementary Fig. 6).

Because a significant fraction of the unfolding events in the circular permutant involved an apparent three-state sequential unfolding, we wondered if some, or even all, of the apparently single cooperative unfolding rips also involved two successive transitions with a short-lived intermediate that was masked by our low sampling rate (50 Hz). We therefore repeated these experiments using a newer instrument with a 1 kHz sampling rate. Indeed, at this higher sampling rate, >90% of the unfolding events involve two transitions (see Supplementary Information). These higher resolution data indicate that, within this force regime, the dominant unfolding pathway for the circular permutant when pulled across both domains (16,159 CP13*T4L) is three-state. Importantly, similar studies on the wild-type topology (16,159 WT*T4L) did not uncover this three-state unfolding pathway (0 out of

136 showed two transitions at 1 kHz). Taken together, these results indicate that transferring the A-helix to the C terminus decouples the two domains and reduces their cooperativity.

The existence of an intermediate comprised of just a folded C-domain is consistent with both protein fragmentation and native state hydrogen exchange studies^{15,16}. In addition, an isolated C-domain fragment is mechanically stable, showing cooperative unfolding in the 10 pN range (data not shown). Conversely, an isolated N-domain fragment is not cooperatively folded^{15,16} and we saw no evidence for the isolated mechanical unfolding of the N-domain. These data support the idea that in the circular permutant, the C-domain can refold independently of the rest of the protein and is mechanically stable even in the absence of an organized N-domain.

The refolding trajectories of 16,159 CP13*T4L also showed reduced interdomain coupling compared to 16,159 WT*T4L, displaying refolding trajectories with either one or two clear compactions (Fig. 2c and Supplementary Fig. 3b, c). Thus, the refolding of 16,159 CP13*T4L involves at least one kinetically resolvable intermediate (see Supplementary Information).

The forces at which 16,159 CP13*T4L unfolded were all significantly lower than those observed for 16,159 WT*T4L (Fig. 3). This permutant also has a much shallower response of unfolding force to experimental pulling speed compared to both WT*T4L constructs (Fig. 3), and a larger calculated distance to the transition state (Supplementary Table 2), indicating that 16,159 CP13*T4L is mechanically more compliant than wild type and explores a qualitatively different unfolding trajectory. This result is consistent with the A-helix no longer acting as a bridge between the domains, mechanically weakening the interdomain region and giving rise to more compliant behaviour.

The re-entrant A-helix seems to be responsible for the high degree of cooperativity observed in WT*T4L, coupling the two domains and giving rise to an all-or-none, two-state unfolding behaviour. These results indicate that the topological organization of the polypeptide chain can dictate the degree of a protein's inter-domain coupling, and, consequently, its mechanical unfolding pathway. To test this idea, we investigated the mechanical unfolding of 16,61 CP13*T4L. If our interpretation is correct, pulling this construct—where the force acts directly only on the N-domain—should lead to unfolding exclusively that domain. This construct unfolds at low forces compared to its corresponding wild-type analogue (Fig. 3) in a single transition with ΔL_c corresponding to the N-domain ($\Delta L_{c,\text{meas}} = 15 \pm 1 \text{ nm}$ ($n = 84$); $\Delta L_{c,\text{calc}} = 14.75 \text{ nm}$). To uncover the status of the C-domain, we used CFT to determine the free energy associated with this transition. In contrast to the ΔG_{CFT} for 16,61 WT*T4L, the ΔG_{CFT} for 16,61 CP13*T4L is $\sim 3.6 \pm 0.2 \text{ kcal mol}^{-1}$ (Fig. 4b), comparable to the energy required to unfold only the N-domain in CP13*T4L (as determined from native state hydrogen exchange, $\Delta G_{\text{NSHX}} = 6.1 \pm 1.0 \text{ kcal mol}^{-1}$; ref. 16). Thus, CFT analysis shows that the unfolding of the N-domain in CP13*T4L is no longer coupled to the rest of the protein, confirming our inference that the discontinuous-domain topology of the wild-type polypeptide is responsible for mechanically coupling the two domains. Circular permutation removes this coupling and transforms a mechanically cooperative system into a non-cooperative one that goes through a long-lived structural intermediate where only the C-domain is folded.

T4 lysozyme has been the subject of two previous mechanical unfolding studies using the AFM^{20,21}. The first involved pulling across the whole protein from residues 21 and 124 and revealed a cooperative, single unfolding transition²⁰. Residues 21 and 124 proved refractory to our DNA modification, but we observe a similar extrapolated unfolding rate and distance to the transition state for 16,159 WT*T4L. Recently, other researchers used the AFM to unfold T4 lysozyme and a circular permutant similar to the one we studied here²¹. In these AFM studies, the samples are polypeptides and hence each construct is pulled from its respective N and C terminus. Therefore, it is impossible to apply force to a single region of the protein in this experiment. For each construct, the authors observed

multiple parallel unfolding pathways, and they were unable to directly correlate the heterogeneous changes in contour lengths with a particular structural transition. In our study, we find no evidence for such heterogeneous kinetic partitioning. Instead, our results are most consistent with the formation of a well-defined unfolding intermediate, one that we can associate by its change in contour length with the unfolding of the N-domain. The differences seen in the two studies may arise from the difference in loading rates afforded by the two methods, the nature of the sample (a tethered polypeptide versus single molecule tether), or the placement of the applied force.

Cooperativity is a hallmark of natural proteins. Mutations that lower cooperativity and increase the population of partially unfolded forms promote misfolding. Classic examples are variants of human lysozyme that fold stably^{25,26}, but in which selective destabilization of one region lowers the coupling between the domains and allows for misfolding and amyloid formation^{25,26}. Our results indicate that a discontinuous topology, although seemingly complex, may confer an advantage by promoting coupling between regions. In fact, circular permutation of enzymes is usually tolerated for both folding and function²⁷, suggesting that a discontinuous topology may fine-tune the folding landscape such that these molecules avoid regions that may lead to kinetic trapping and frustration. The mechanical unfolding trajectories depend markedly on this topological feature.

Spontaneous folding and self-assembly are fundamental processes of bio-morphogenesis. Although driven thermodynamically, these processes can be frustrated by the formation of intermediates, which are often kinetically trapped and unproductive. Our results indicate that productive cooperative interactions among protein domains, and which regions of the folding landscape are explored, depend not only on the local details of the protein's structure, but also on the topological organization of its polypeptide chain.

METHODS SUMMARY

Variants of *T4L with cysteines at specific sites were constructed using standard genetic engineering techniques. All cysteines were placed in solvent-exposed sites. Ensemble folding and equilibrium denaturation studies showed that all of the cysteine variants used here display minimal perturbation in either the energetics or folding kinetics of the protein (see Supplementary Table 1). These cysteines were then used to attach the protein to two DNA handles. The DNA handles were generated by PCR, using pGEMEX 1 plasmid DNA from Promega as template. One handle was synthesized using the primers 5'-thiol-GCTACCGTAATTGAGACCAC-3' and 5'-biotin-CAAAAAACCCCTCAAGACCC-3'; the other was synthesized using the same 5'-thiol primer together with 5'-digoxigenin-version of the biotin primer. The chemistry and protocol for handle attachment was as described in ref. 18. The optical tweezer experiments were performed in 250 mM NaCl, 10 mM Tris pH 7.0 and 1 mM EDTA at ambient room temperature.

Single molecule protein tethers were identified by their DNA overstretching transition and those samples were repeatedly stretched and relaxed at a constant speed. Force-extension curves were analysed to obtain unfolding forces and changes in contour length¹⁷. The work done in folding or unfolding the protein was calculated by integrating the area under individual force-extension curves, using equal-force bounds that directly surrounded the refolding or unfolding transition (results described in detail in Supplementary Information).

Analysis of force-extension curves. After tethering a protein sample between the two beads as diagrammed in Fig. 1, the beads were moved away from one another at a constant speed until a predetermined maximum force (typically ~70 pN) was attained. Individual protein tethers that displayed an overstretching transition for DNA at ~67 pN were then repeatedly stretched and relaxed between 0 and ~50 pN of force. These force extension curves were analysed as described previously¹⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 June 2009; accepted 18 March 2010.
Published online 23 May 2010.

- Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).

- Ng, S. P., Randles, L. G. & Clarke, J. Single molecule studies of protein folding using atomic force microscopy. *Methods Mol. Biol.* **350**, 139–167 (2007).
- Borgia, A., Williams, P. M. & Clarke, J. Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77**, 101–125 (2008).
- Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. Stretching single molecules into novel conformations using the atomic force microscope. *Nature Struct. Biol.* **7**, 719–724 (2000).
- Bustamante, C., Chemla, Y. R., Forde, N. R. & Izahy, D. Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
- Brockwell, D. J. *et al.* Pulling geometry defines the mechanical resistance of a β -sheet protein. *Nature Struct. Mol. Biol.* **10**, 731–737 (2003).
- Carrion-Vazquez, M. *et al.* The mechanical stability of ubiquitin is linkage dependent. *Nature Struct. Mol. Biol.* **10**, 738–743 (2003).
- Dietz, H. & Rief, M. Protein structure by mechanical triangulation. *Proc. Natl Acad. Sci. USA* **103**, 1244–1247 (2006).
- Wetlaufer, D. B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA* **70**, 697–701 (1973).
- Elwell, M. & Schellman, J. Phage T4 lysozyme. Physical properties and reversible unfolding. *Biochim. Biophys. Acta* **386**, 309–323 (1975).
- Mchaourab, H. S., Oh, K. J., Fang, C. J. & Hubbell, W. L. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry* **36**, 307–316 (1997).
- Anderson, D. E., Lu, J., McIntosh, L. & Dahlquist, F. W. in *NMR of Proteins* (eds Clore, G. M. & Gronenborn, A. M.) ch. 9 (CRC Press, 1993).
- Zhang, X. J., Wozniak, J. A. & Matthews, B. W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J. Mol. Biol.* **250**, 527–552 (1995).
- Llinás, M., Gillespie, B., Dahlquist, F. W. & Marqusee, S. The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nature Struct. Biol.* **6**, 1072–1078 (1999).
- Llinás, M. & Marqusee, S. Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci.* **7**, 96–104 (1998).
- Cellitti, J. *et al.* Exploring subdomain cooperativity in T4 lysozyme I: Structural and energetic studies of a circular permutant and protein fragment. *Protein Sci.* **16**, 842–851 (2007).
- Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057–2060 (2005).
- Cecconi, C., Shank, E. A., Dahlquist, F. W., Marqusee, S. & Bustamante, C. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur. Biophys. J.* **37**, 729–738 (2008).
- Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).
- Yang, G. *et al.* Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl Acad. Sci. USA* **97**, 139–144 (2000).
- Peng, Q. & Li, H. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: Evidence for a kinetic partitioning mechanism. *Proc. Natl Acad. Sci. USA* **105**, 1885–1890 (2008).
- Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **60**, 2721–2726 (1999).
- Collin, D. *et al.* Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **437**, 231–234 (2005).
- Bornschlög, T. & Rief, M. Single-molecule dynamics of mechanical coiled-coil unzipping. *Langmuir* **24**, 1338–1342 (2008).
- Booth, D. R. *et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793 (1997).
- Dumoulin, M. *et al.* A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature* **424**, 783–788 (2003).
- Lo, W. C., Lee, C. C., Lee, C. Y. & Lyu, P. C. CPDB: a database of circular permutation in proteins. *Nucleic Acids Res.* **37**, D328–D332 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank G. Crooks for assistance in use of Crooks fluctuation analysis, R. Dahlquist and B. Matthews for help in initiating this study, and the entire Bustamante and Marqusee labs for advice and technical help. We would particularly like to thank E. Kwon for her assistance in reagent preparation and data collection. This work was supported in part by NIH grants GM 32543 (C.B.), GM 50945 (S.M.) and a grant from the NSF (S.M.).

Author Contributions E.A.S., C.C., J.W.D. conducted the experiments and analysed the data. J.W.D. carried out the Crooks fluctuation analysis. E.A.S., J.W.D., S.M. and C.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M. (Marqusee@berkeley.edu) or C.B. (Carlos@alice.berkeley.edu).

METHODS

Identifying single molecule tethers. Single molecule tethers were identified by examining the DNA handle overstretching transition. Only those fibres that exhibited an overstretching transition of the expected length (230 nm) at the correct force (67 pN) for a single fibre of 1,116 bp DNA were selected for analysis²⁸.

Alternate bounds of integration for the Crooks fluctuation theorem. We calculate the work done in unfolding or refolding a protein by integrating the area under a force-extension curve around the transition. This can be done by overlaying many force-extension curves and setting universal bounds of integration that enclose all of the unfolding and refolding events. The work values must then be corrected for the energy put into stretching the DNA handles and the unfolded polypeptide chain.

The unfolded protein's stretching energy is calculated first by modelling the protein as a wormlike chain and then integrating this theoretical force-extension curve for the unfolded polypeptide chain, with lower and upper bounds defined by the distance between cysteines in the folded state, and the maximum force of the pulling protocol, respectively. The stretching energy of the DNA handles is calculated similarly. Here, we introduce a novel approach that mathematically eliminates the need to calculate the stretching energy of DNA. This removes a key source of uncertainty and simplifies the calculations.

Our approach relies on the fact that, for an equilibrium transition like stretching and relaxing dsDNA, the work done in any trajectory that starts and ends at the same force is zero. We take advantage of this by integrating each force-extension curve individually, and setting the bounds of integration at two points at equal force, on either side of the unfolding or refolding event. As the DNA handles have not undergone any net conformational change, the only necessary correction is to subtract the energy put into stretching the unfolded protein between the native structure and the force at which the transition occurs.

This method depends on the ability to set bounds that enclose the structural transitions, but are also at the same force. Because of this, it is applicable only to two-state transitions. The WT and CP13 mutants of 16,61 *T4L both meet this condition, as the force-bearing region between the cysteines folds and unfolds in an all-or-nothing manner in both constructs.

Pulling using a higher time resolution (1 kHz). Using a new optical tweezers instrument with a 1 kHz sampling rate, we applied a force-ramp protocol with a similar loading rate to the original data ($\sim 16 \text{ pN s}^{-1}$): a 210 nm s^{-1} pulling rate with an optical-trap spring constant of 0.075 pN nm^{-1} . Several (74) force-extension curves were recorded. When averaged down to a sampling rate of 50 Hz, the unfolding kinetics of the 16,159 CP13 *T4L mutant replicated the original observed behaviour: some unfolding events were a single event with an extension change corresponding to the complete unfolding of the protein; some were split into two rips, with the N-terminal domain unfolding before the C-terminal domain; and some showed unfolding of just the C-terminal domain, due to incomplete refolding of the protein in the previous curve.

At 1 kHz, 61 (82%) of the curves showed a discernible pause at the unfolding intermediate. This is a marked increase from the 1,231 force-ramp data recorded at 50 Hz, where only $\sim 25\%$ of the transitions showed this intermediate. Nine force-extension curves (12%) showed unfolding of just the C-terminal subdomain: a similar fraction to the 15% originally recorded. The remaining four curves (5%) taken at 1 kHz showed just one unfolding event, corresponding to the full unfolding of the protein.

These results indicate that the original estimates were limited by the low sampling rate of the instrument. Of the 61 ramps showing a clear unfolding intermediate at 1 kHz, 34 show a pause at the intermediate state lasting for $<40 \text{ ms}$: this is the minimum duration (defined as two sampling intervals) that one could expect to observe an intermediate state at 50 Hz. Thus, as stated above, averaging down the data to 50 Hz again would reduce the fraction of force extension curves with clear unfolding intermediates to 36%—correlating well with the original 25% estimate.

Similar experiments were carried out at this higher sampling rate on 16,159 WT *T4L. Unlike the permuted variant, this protein unfolded in a single transition. Out of 136 pulls, no double-rip transitions were observed, and the contour lengths of the observed transition were consistent with unfolding of the entire region between residues 16 and 159.

28. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).