

Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins

Madeleine B. Borgia^{1,2}, Alessandro Borgia², Robert B. Best¹, Annette Steward¹, Daniel Nettels², Bengt Wunderlich², Benjamin Schuler² & Jane Clarke¹

A large range of debilitating medical conditions¹ is linked to protein misfolding, which may compete with productive folding particularly in proteins containing multiple domains². Seventy-five per cent of the eukaryotic proteome consists of multidomain proteins, yet it is not understood how interdomain misfolding is avoided. It has been proposed that maintaining low sequence identity between covalently linked domains is a mechanism to avoid misfolding³. Here we use single-molecule Förster resonance energy transfer^{4,5} to detect and quantify rare misfolding events in tandem immunoglobulin domains from the I band of titin under native conditions. About 5.5 per cent of molecules with identical domains misfold during refolding *in vitro* and form an unexpectedly stable state with an unfolding half-time of several days. Tandem arrays of immunoglobulin-like domains in humans show significantly lower sequence identity between neighbouring domains than between non-adjacent domains³. In particular, the sequence identity of neighbouring domains has been found to be preferentially below 40 per cent³. We observe no misfolding for a tandem of naturally neighbouring domains with low sequence identity (24 per cent), whereas misfolding occurs between domains that are 42 per cent identical. Coarse-grained molecular simulations predict the formation of domain-swapped structures that are in excellent agreement with the observed transfer efficiency of the misfolded species. We infer that the interactions underlying misfolding are very specific and result in a sequence-specific domain-swapping mechanism. Diversifying the sequence between neighbouring domains seems to be a successful evolutionary strategy to avoid misfolding in multidomain proteins.

Multidomain proteins comprise covalently linked, frequently similar domains, resulting in high effective local protein concentration. It is therefore probable that these proteins have evolved to avoid interdomain misfolding *in vivo*. Co-translational, 'domain-by-domain' folding is assumed to be important for avoiding misfolding⁶, but many multidomain proteins that are long lived or subject to tensile forces will fold and unfold numerous times during their lifetimes, resulting in neighbouring domains being unfolded at the same time, and therefore may be particularly vulnerable to misfolding. The giant muscle protein titin, for example, undergoes reversible domain unfolding that may have a role in muscle elasticity⁷.

Single-molecule techniques are ideal for detecting rare events⁸ such as misfolding in native conditions. Indeed, the first evidence for the misfolding of adjacent domains in long tandem arrays of the well-characterized twenty-seventh domain from the I band of titin, I27, was obtained using single-molecule atomic force microscopy⁹ (AFM). An alternative approach is single-molecule Förster resonance energy transfer^{4,5} (FRET), whose great sensitivity allows the detection of very small populations. FRET permits the mapping of intramolecular distances by means of the distance-dependent efficiency of excitation energy transfer between donor and acceptor fluorophores attached to specific positions of the protein¹⁰ (for details, see Supplementary Fig. 1).

We proposed that denaturation of tandem constructs of titin domains with guanidinium chloride (GdmCl), followed by rapid

refolding into native conditions, might allow formation of misfolded species, which should be detectable using single-molecule FRET¹¹. We labelled a tandem construct of I27 (I27–I27) in the A strand of domain 1 (E3C) and the G strand of domain 2 (N83C) (Fig. 1a) with a donor (Alexa Fluor 488) and an acceptor (Alexa Fluor 594) fluorophore, attached by means of cysteine residues engineered on the protein surface. For a misfolded domain formed by strands from domains 1 and 2, we predicted that these two strands must be adjacent for the domain to have the mechanical properties observed in the previous AFM experiments⁹. The correctly folded tandem domain would then have low transfer efficiency, whereas a misfolded state would have high transfer efficiency. A monomer of I27 was labelled in the corresponding positions to provide a model for the misfolded state (Fig. 1b). Labelling was found to have little effect on the stability of I27 (Supplementary Fig. 2), and doubly labelled proteins that had not previously been unfolded in GdmCl ('never unfolded') showed single, correctly folded populations with transfer efficiencies of $E = 0.37 \pm 0.01$ and 0.93 ± 0.01 for the tandem I27–I27 and monomer I27, respectively (Fig. 2a, b). (Uncertainties quoted in the text are s.d.)

We conducted refolding experiments by diluting unfolded I27–I27 into refolding buffer. The resulting transfer efficiency histograms then showed two populations (Fig. 2c): one corresponding to the correctly folded native state ($E = 0.37$) and one with precisely the same transfer efficiency as the analogously labelled monomer ($E = 0.93$). This

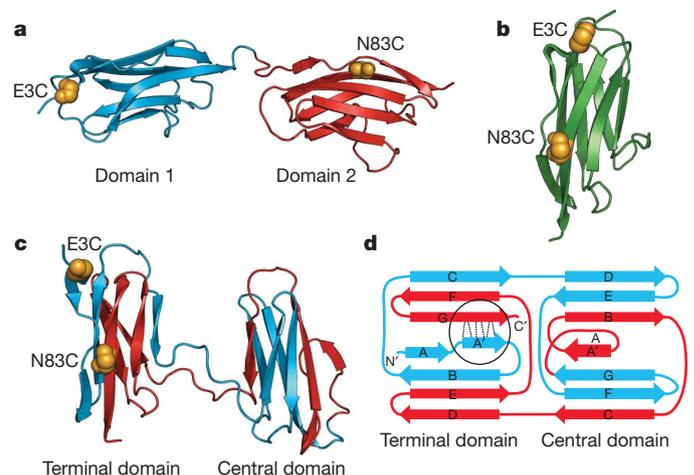


Figure 1 | Structures of native and misfolded I27 constructs. **a**, Natively folded I27–I27 tandem repeat with labelling positions highlighted (golden spheres). **b**, Native I27 crystal structure (Protein Data Bank ID, 1TTT). **c**, One of the domain-swapped, misfolded state structures formed in Gō-model simulations. **d**, Schematic of the misfolded state topology in **c**: hydrogen bonds that are perpendicular to the direction of applied force in AFM mechanical unfolding are shown by dashed lines (circled). Four other misfolded state topologies were populated in the simulations (Supplementary Fig. 5b). We note that we cannot distinguish between such topologies from the results presented here.

¹University of Cambridge Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK. ²Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

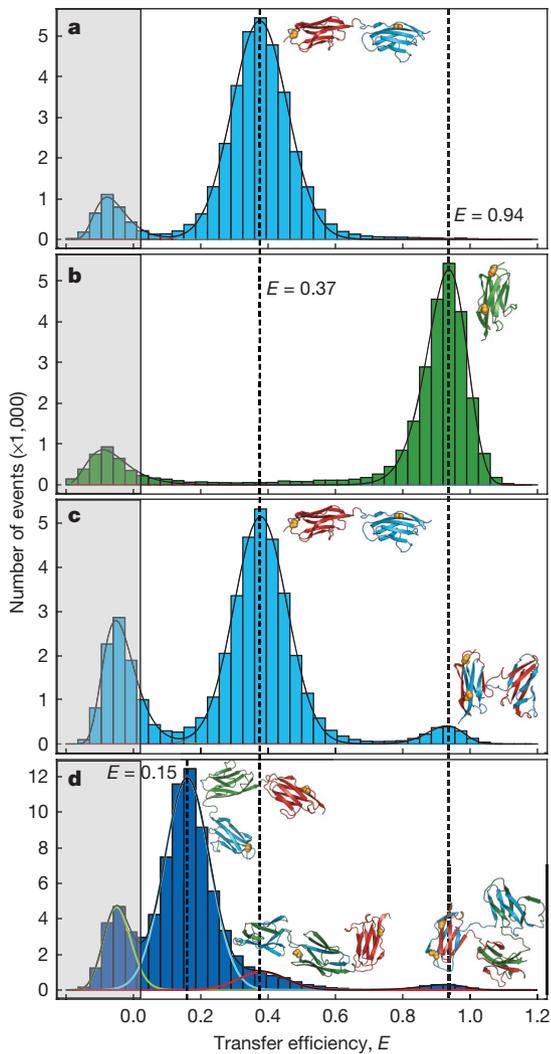


Figure 2 | Transfer efficiency histograms of doubly labelled I27 constructs. **a**, Never-unfolded I27–I27. **b**, Never-unfolded monomeric I27. **c**, Refolded I27–I27. **d**, Refolded I27–I27–I27; fits of individual populations shown as coloured lines for clarity. Histograms are fitted with normal or log-normal distributions. The peak in the grey shaded area consists of events from molecules without an active acceptor fluorophore²⁸. We note that in these experiments a short, four-amino-acid linker (Arg-Ser-Glu-Leu) is included between the domains in the I27–I27 tandem to allow direct comparison with previous AFM and aggregation experiments^{3,9}. Never-unfolded I27–I27–I27 is shown in Supplementary Fig. 6.

observation reveals that the A strand of the first domain and the G strand of the second domain are arranged as in the monomer. A quantitative analysis (Supplementary Table 1) showed that $5.5 (\pm 0.2)\%$ of the molecules are found in the misfolded form.

On the basis of the results of the AFM studies⁹, we had supposed that the misfolded species consisted of a single strand-swapped titin domain with the remaining sequence unstructured and, thus, with an unfolding time similar to that of a native domain ($\tau \approx 34$ min; ref. 12). We therefore investigated the unfolding kinetics of the misfolded state¹³ (Fig. 3a). At high GdmCl concentrations, the decays in the number of high-transfer-efficiency events ($E > 0.8$), corresponding to the misfolded state, are fitted well by single exponentials (Fig. 3b) with rate constants slightly higher than the unfolding rate constants for wild-type I27 determined in ensemble measurements¹² (Fig. 3c). We can also estimate the unfolding rate constant of the correctly folded species, which agrees well with the ensemble data (Fig. 3c) (see Supplementary Fig. 3 and Supplementary Information). In the absence of denaturant, however, the misfolded species was surprisingly long lived,

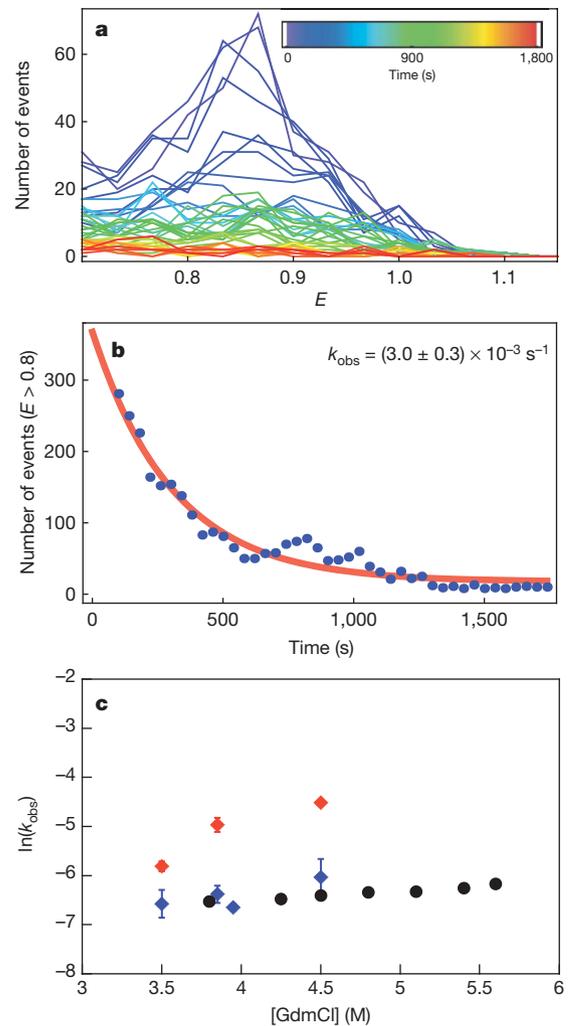


Figure 3 | Unfolding kinetics. **a**, Evolution of transfer efficiency histograms over time ($E \geq 0.7$) from single-molecule double-jump experiments in which refolded/misfolded I27–I27 (doubly labelled) was unfolded in 3.5 M GdmCl. Histograms were constructed for a moving window of 120 s that was shifted by 30 s for each increment (colour scale). **b**, The number of events with $E > 0.8$ for each histogram in **a** was summed and the resulting kinetics fitted with a single exponential decay. The observed rate constants (k_{obs}) are unaffected by different window sizes or the use of non-overlapping windows. **c**, Unfolding rate constants for the wild-type I27 monomer (black; ensemble data from ref. 12) and for the misfolded and natively folded states of I27–I27 from single-molecule measurements (red and blue, respectively), as functions of GdmCl concentration. Each point was obtained by fitting data sets composed by merging at least six repeats, with error bars representing the standard error of the fit (Methods). For some data points, the error bars are smaller than the symbols. I27 domains have the same unfolding rate constants in tandem repeat proteins as in isolated domains.

converting to the correctly folded form only on a timescale of days (Supplementary Fig. 4). The formation of the misfolded structure is thus under kinetic, rather than thermodynamic, control. Its remarkable kinetic stability clearly distinguishes the misfolded species described here from short-lived, partly folded intermediates^{14–16} sometimes termed ‘misfolded’ because they contain some non-native interactions¹⁷.

An explanation for the slow unfolding under native conditions is suggested by folding simulations of I27–I27 with a Gō-like model¹⁸. In these simulations, only native interactions are attractive, and interactions between a given pair of residues are considered equal, independently of whether they are in the same or different domains. Although most trajectories result in two correctly folded domains, misfolded species with two fully folded, strand-swapped domains are occasionally formed. Five different strand-swapped topologies were observed

(Fig. 1c, d and Supplementary Fig. 5). Such an extensively misfolded structure explains its persistence: correct folding cannot occur while either misfolded domain remains folded. Because refolding rate constants are much higher than unfolding rate constants under native conditions, the simultaneous unfolding of both domains is very unlikely and conversion to the native state is extremely slow¹⁹.

We can test the validity of our model further by investigating the refolding of a three-domain tandem of I27 with the FRET labels in domain 1 (E3C) and domain 3 (N83C). If domain-swapped structures were to be formed, we would expect to see two misfolded populations: one with the FRET efficiency of the monomer (misfolding between domains 1 and 3) and another population with the efficiency of the I27–I27 tandem (misfolding between domains 1 and 2 or 2 and 3). This is precisely what we observe (Fig. 2d). The proportion of monomer-like (high-FRET) species in the trimeric tandem is significantly lower than before ($2.8 (\pm 0.6)\%$); this is likely to reflect the lower probability of association between domains that are more distant in sequence. The population of the misfolded species with dimer-like FRET efficiency ($9 (\pm 2)\%$) was instead almost twice as large as in the two-domain tandem ($5.5 (\pm 0.2)\%$); this is probably due to the two alternative possibilities to misfold in an analogous way to the dimeric tandem (domain 1 with 2 and domain 2 with 3). Simulations with the Gō-like model also predict domain-swapped structures with monomer- and dimer-like FRET efficiencies, with relative populations similar to experiment (Supplementary Fig. 6).

Much work has been dedicated to investigating the sequence specificity of protein aggregation^{20,21}, including the hypothesis that there is selective pressure to prevent oligomerization by a domain-swapping mechanism^{22,23}. Misfolding is often considered to precede aggregation, suggesting that sequence-specific behaviour observed in aggregation also applies to misfolding. Our single-molecule FRET experiments allow us to test this hypothesis directly by investigating mixed tandem constructs, I27–I28 and I27–I32 (Fig. 4). Indeed, the construct comprising I27 and I28, natural neighbours in titin with only 24% sequence identity, did not yield any detectable population of high-FRET misfolded species on refolding (Fig. 4c, d). However, misfolding is seen on refolding of I27–I32 (sequence identity, 42%; Fig. 4e, f), to the same extent as for I27–I27 (Supplementary Table 1). This misfolded species also unfolds with the same rate constant as I27–I27 (Supplementary Fig. 7). The I27–I32 misfold is consistent with previous experiments showing chimaeric domains of I27 and I32 to be stable^{24,25}. This result strongly supports the idea that protein misfolding is sequence specific. In proteins where sequence identity between neighbouring domains is high, the topology may prevent formation of stable misfolded species¹⁹.

Misfolding in our experiments is more frequent than had been observed in AFM experiments⁹, suggesting that the tethering in those experiments reduces misfolding; this might be advantageous for titin domains *in vivo*. Unfolding of the misfolded species observed with AFM showed them to have the same mechanical resistance as correctly folded I27, but with twice the chain length being released on unfolding. Our results are entirely compatible with this finding. Although the misfolded species has two folded domains, and is therefore stable in folding conditions, only the terminal domain would experience shearing of the hydrogen bonds between the parallel A' and G strands (Fig. 1d, circled) perpendicular to the direction of the applied force, in the same way as the correctly folded I27²⁶, resulting in the same mechanical stability. Because force is not applied to the A and G strands in the central domain, this domain is likely to unfold at low force, together with the terminal domain. This hypothesis is supported by simulations (Supplementary Fig. 8).

Our results suggest that diversifying the sequence composition between neighbouring domains is an effective evolutionary strategy to ensure efficient folding in multidomain proteins and avoid the formation of stable misfolded species. This adds a significant piece to the puzzle of understanding the problems encountered during the crucial evolutionary transition from single to multidomain proteins.

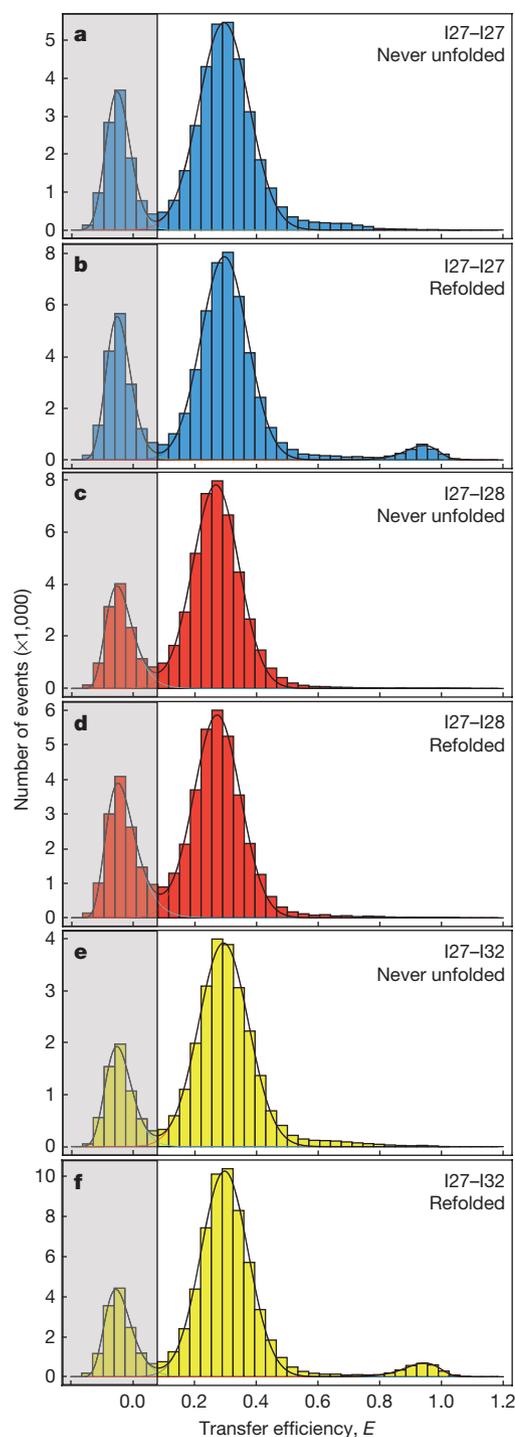


Figure 4 | Transfer efficiency histograms of tandem constructs with identical and non-identical domains. **a, b**, I27–I27 never-unfolded control (**a**) and refolded (**b**) constructs. **c, d**, I27–I28 never-unfolded control (**c**) and refolded (**d**) constructs. **e, f**, I27–I32 never-unfolded (**e**) control and refolded (**f**) constructs. To mimic the natural protein, no linker was added between the domains in these experiments. We note that the frequency of misfolding was the same for I27–I27 with and without the linker, $5.5 (\pm 0.2)\%$ and $5.7 (\pm 0.5)\%$, respectively (compare Fig. 2c with Fig. 4b and Supplementary Table 1). Addition of the four-amino-acid linker also made no difference to the results for I27–I28 (Supplementary Fig. 9).

METHODS SUMMARY

For details of protein production, ensemble equilibrium measurements and labeling, see Methods. Single-molecule experiments, instrumentation, data reduction and analysis are also detailed in Methods. The resulting relative populations from all experiments and analysis techniques are summarized in Supplementary Table 1.

Folding simulations using a Gō-like model were run using the CHARMM code²⁷ as described in Methods. For details of mechanical unfolding simulations, see Methods.

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

Received 6 January; accepted 1 April 2011.

Published online 29 May 2011.

- Gregersen, N., Bross, P., Vang, S. & Christensen, J. H. Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* **7**, 103–124 (2006).
- Jaenicke, R. & Seckler, R. Protein misassembly *in vitro*. *Adv. Protein Chem.* **50**, 1–59 (1997).
- Wright, C. F., Teichmann, S. A., Clarke, J. & Dobson, C. M. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881 (2005).
- Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77**, 51–76 (2008).
- Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
- Fedorov, A. N. & Baldwin, T. O. Cotranslational protein folding. *J. Biol. Chem.* **272**, 32715–32718 (1997).
- Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
- Borgia, A., Williams, P. M. & Clarke, J. Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77**, 101–125 (2008).
- Oberhauser, A. F., Marszalek, P. E., Carrion-Vasquez, M. & Fernandez, J. M. Single protein misfolding events captured by atomic force microscopy. *Nature Struct. Biol.* **6**, 1025–1028 (1999).
- Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).
- Gambin, Y. *et al.* Direct single-molecule observation of a protein living in two opposed native structures. *Proc. Natl Acad. Sci. USA* **106**, 10153–10158 (2009) CrossRef.
- Fowler, S. B. & Clarke, J. Mapping the folding pathway of an immunoglobulin domain: structural detail from phi value analysis and movement of the transition state. *Structure* **9**, 355–366 (2001).
- Hofmann, H. *et al.* Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl Acad. Sci. USA* **107**, 11793–11798 (2010).
- Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M. & Gianni, S. Folding and misfolding in a naturally occurring circularly permuted PDZ domain. *J. Biol. Chem.* **283**, 8954–8960 (2008).
- Gianni, S. *et al.* Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nature Struct. Mol. Biol.* **17**, 1431–1437 (2010).
- Korzhev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
- Capaldi, A. P., Kleanthous, C. & Radford, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216 (2002).
- Yang, S. *et al.* Domain swapping is a consequence of minimal frustration. *Proc. Natl Acad. Sci. USA* **101**, 13786–13791 (2004).
- Arora, P., Hammes, G. G. & Oas, T. G. Folding mechanism of a multiple independently-folding domain protein: double B domain of protein A. *Biochemistry* **45**, 12312–12324 (2006).
- Jaenicke, R. Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237 (1987).
- Straub, J. E. & Thirumalai, D. Principles governing oligomer formation in amyloidogenic peptides. *Curr. Opin. Struct. Biol.* **20**, 187–195 (2010).
- Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **14**, 811–824 (2006).
- Mitraki, A. Protein aggregation from inclusion bodies to amyloid and biomaterials. *Adv. Protein Chem. Struct. Biol.* **79**, 89–125 (2010).
- Borgia, A., Steward, A. & Clarke, J. An effective strategy for the design of proteins with enhanced mechanical stability. *Angew. Chem. Int. Ed.* **47**, 6900–6903 (2008).
- Balamurali, M. M. *et al.* Recombination of protein fragments: a promising approach toward engineering proteins with novel nanomechanical properties. *Protein Sci.* **17**, 1815–1826 (2008).
- Lu, H., Israilewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671 (1998).
- Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
- Dahan, M. *et al.* Ratiometric measurement and identification of single diffusing molecules. *Chem. Phys.* **247**, 85–106 (1999).

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

Acknowledgements This work was supported by the Wellcome Trust (grant number, 064417), the Swiss National Science Foundation (to B.S.) and the Swiss National Center of Competence in Research in Structural Biology (to B.S.). M.B.B. was supported by a UK Medical Research Council studentship. A.B. is supported by a Marie Curie Intra-European Fellowship. R.B.B. is supported by a Royal Society University Research Fellowship. J.C. is a Wellcome Trust Senior Research Fellow. We thank H. Hofmann, A. Soranno and A. Hoffmann for discussions and contributions to data analysis.

Author Contributions M.B.B., A.B., B.S. and J.C. designed the investigation. M.B.B. and A.B. performed the experiments. R.B.B. performed the simulations. D.N. and B.W. built the single-molecule instrumentation. D.N. provided data analysis software. A.S. cloned the gene of the trimeric tandem construct. M.B.B. performed the analysis. M.B.B., J.C. and B.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.C. (jc162@cam.ac.uk) or B.S. (schuler@bioc.uzh.ch).

METHODS

Protein expression and labelling. Cysteine residues were introduced by site-directed mutagenesis: for the two-domain constructs, E3C in domain 1 (always I27), N83C in domain 2 if I27 or I32, and K83C in domain 2 if I28 (with the numbering relative to a single domain); and for the three-domain construct of I27, E3C in domain 1 and N83C in domain 3. DNA sequencing confirmed the mutagenesis.

I27 monomer and the I27–I27, I27–I27–I27, I27–I28 and I27–I32 tandems, with the engineered surface cysteines, were expressed as described previously^{29,30}. Labelling was carried out using Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor) maleimides (Invitrogen) according to the manufacturer's procedures. The dyes were mixed simultaneously with reduced protein in equimolar ratios and incubated at 4 °C for ~10 h. Unreacted dye was removed by gel filtration and the differently labelled variants were separated by ion-exchange chromatography (MonoQ 5/50 GL; GE Healthcare Biosciences AB). I27 has two intrinsic cysteines that were not removed as they are buried in the native state and all labelling was carried out on folded protein in native conditions.

Ensemble measurements. Equilibrium measurements were performed for the doubly labelled I27 monomer to check the effect of labelling (Supplementary Fig. 2). Experiments were performed in GdmCl on a Cary Eclipse fluorimeter (Varian Inc.) monitoring intrinsic tryptophan fluorescence as described previously²⁹, but with lower protein concentrations (0.05–0.5 µM) and the addition of 0.001% Tween 20.

Single-molecule instrumentation. Observations of single-molecule fluorescence were made using a custom-built confocal microscope equipped with a continuous-wave, 488-nm solid-state laser (FCD488-010, JDSU) and an Olympus UplanApo ×60/1.20W objective. After passage through a dichroic mirror that separates excitation and emission light (500DCXR, Chroma Technology), fluorescence emission passed through a 100-µm pinhole and was split into donor and acceptor fluorescence by a second dichroic mirror (585DCXR, Chroma Technology). Donor fluorescence then passed a filter (ET525/50M, Chroma Technology) before being focused onto a single-photon avalanche diode (MPD 100ct, Micro Photon Devices). Similarly, acceptor fluorescence passed a filter (QT 650/100) before being focused onto a single-photon avalanche diode (SPCM-AQR-13, PerkinElmer Optoelectronics). The arrival time of every photon was recorded with a two-channel, time-correlated, single-photon counting module (PicoHarp300, PicoQuant). All measurements were performed with a laser power of 100 µW, measured at the back aperture of the objective (beam waist, 8 mm).

Single-molecule equilibrium measurements. All experiments were performed at protein concentrations between 0.5 and 25 pM, in the same final solution conditions: PBS; 0.001% Tween 20; 140 mM β-mercaptoethanol; 20 mM cysteamine hydrochloride (single-molecule buffer). Tween 20 (Pierce) was used to prevent surface adhesion of the proteins³¹, whereas the photoprotective agents β-mercaptoethanol (Sigma) and cysteamine hydrochloride (Sigma) were used to minimize chromophore damage³². Experiments on never-unfolded proteins were conducted by mixing protein in PBS (0.01% Tween, 10 mM β-mercaptoethanol) 1:99 with 0.04 M GdmCl in single-molecule buffer (to mimic the final conditions in the refolding experiments). Refolding experiments were performed by mixing protein unfolded in 4.4 M GdmCl (PBS, 0.01% Tween, 10 mM β-mercaptoethanol) 1:100 with single-molecule buffer, to a final GdmCl concentration of 0.04 M. Measurements over 8 to 10 h were made for all constructs with one or more repeats. The absence of aggregates was ensured in all experimental conditions as previously described²³. Fluorescence bursts were identified by combining successive photons separated by 150 µs or less, and events comprising 35 or more photons were kept for analysis. Transfer efficiencies were corrected for quantum yields, cross-talk and direct excitation as described previously^{34,35}.

Populations of correctly folded and misfolded molecules in the transfer efficiency histograms were quantified using two methods. Where possible, peaks were fitted using a Gaussian distribution (for populations where $0.1 \leq E \leq 0.8$) or a log-normal distribution (for populations where $E \geq 0.8$), and the resulting fits integrated. Transfer efficiencies quoted in the main text denote the average E value obtained from these fits, with standard deviations calculated from multiple experiments. The populations of misfolded species were determined relative to the sum of natively folded and misfolded populations. Alternatively, ranges of transfer efficiencies were chosen for each population, the corresponding number of bursts were summed and the relative populations of misfolded species were computed.

Only the latter method could be used in the 'never-unfolded' control measurements and all experiments involving I27–I28 tandems, as no misfolded population was observed. The resulting relative populations from all experiments and analysis techniques are summarized in Supplementary Table 1.

Single-molecule kinetic measurements. Unfolding of the misfolded species was achieved by mixing a previously refolded protein sample (prepared as for refolding experiments above, but with higher protein concentration) with GdmCl in single-molecule buffer. To obtain sufficient statistics, at least six repeat measurements were made for each GdmCl concentration and the resulting photon trajectories were merged to give one data set. A moving-window analysis¹³ was applied to the merged, time-resolved photon trajectory with a window size (Δt) of 120 s. Transfer efficiency histograms were calculated from the bursts in that time window, and the window was shifted by $\Delta t/3$ to form each successive time point (Fig. 3a). A time $t = t_s + \Delta t/2$, where t_s is the start time of the window, was assigned to each histogram, and the number of events with $E \geq 0.8$ in each histogram as a function of time was fitted to a single exponential (Fig. 3b). The resulting rate constants were robust for different window sizes and non-overlapping windows. Standard errors were taken from the covariance matrix of the fit (weighted by the average inverse variance of the residuals of the data points with respect to an unweighted fit).

The analysis of the unfolding of the native state from the same type of experiment is detailed in Supplementary Information and Supplementary Fig. 5.

Simulations. A coarse-grained Gō-like model was generated on the basis of the structure of I27 (Protein Data Bank ID, 1TIT³⁶) using a standard procedure³⁷. Briefly, all bond lengths are fixed by constraints, harmonic terms are used for the angles, a knowledge-based potential is used for the torsion angles and non-bonded interactions are treated with a Gō-like potential in which only interactions formed between residues in the folded protein are attractive (with relative strengths given by the Miyazawa–Jernigan matrix), all others being repulsive. Two or three identical I27 sequences were linked by four-residue repulsive linkers to treat the two and three domain tandems, respectively. Interactions between residues i and j in different domains were treated exactly like the interactions between those residues in the same domain, and interactions with the linker were repulsive. A simulation temperature was chosen such that the folding barrier was at least $3k_B T$, by using as a lower bound the free-energy barrier projected onto the fraction of native contacts. Folding simulations were run, starting from fully extended configurations, using Langevin dynamics with a friction of 0.1 ps^{-1} and a time step of 10 fs. The final structures were clustered using a simple leader–follower algorithm with a cutoff of 0.15 on the r.m.s. distance between contact maps.

Mechanical unfolding simulations were performed in which a force of 150 pN was applied to the ends of the I27–I27 tandem, starting from structures belonging to either the folded cluster or one of the misfolded clusters, and monitoring unfolding by measuring the end–end distance. The CHARMM molecular simulation package was used for all calculations²⁷.

29. Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**, 819–829 (2002).
30. Steward, A., Toca-Herrera, J. L. & Clarke, J. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.* **11**, 2179–2183 (2002).
31. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002).
32. Nettels, D. *et al.* Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl Acad. Sci. USA* **106**, 20740–20745 (2009).
33. Hillger, F., Nettels, D., Dorsch, S. & Schuler, B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J. Fluoresc.* **17**, 759–765 (2007).
34. Schuler, B. Application of single molecule Förster resonance energy transfer to protein folding. *Protein Folding Protocols* (eds Bai, Y. & Nussinov, R.) 115–138 (Humana Press, 2007).
35. Hoffmann, A. *et al.* Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA* **104**, 105–110 (2007).
36. Improta, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4**, 323–337 (1996).
37. Karanicolas, J. & Brooks, C. L. III. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* **334**, 309–325 (2003).

Supplementary Information:

S1: Supplementary Methods

Singular value decomposition (SVD) of native state unfolding.

To analyse the unfolding of the native state from single-molecule double-jump experiments performed on I27-I27 (as detailed in **Methods**), singular value decomposition (SVD) was employed due to the overlap of the native and denatured populations.

A moving-window analysis (see **Methods**) was applied to the transfer efficiency range -0.2 to 0.6 (20 bins). The resulting E histograms were combined into an $m \times n$ matrix, \mathbf{A} , for $m=20$ bins and n time points. SVD analysis decomposes the matrix \mathbf{A} into three matrices³⁸: \mathbf{U} , \mathbf{S} and \mathbf{V} :

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$$

\mathbf{U} is an $m \times m$ matrix whose columns form the “output” basis vectors (eigenvectors of $\mathbf{A}\mathbf{A}^T$). \mathbf{S} is an $m \times n$ diagonal matrix whose elements are the singular values, or weighting factors for every basis vector (the square roots of eigenvalues of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$). \mathbf{V}^T is a transposed $n \times n$ matrix whose columns form the “output” amplitude vectors (eigenvectors of $\mathbf{A}^T\mathbf{A}$) that correspond to the kinetics of the process. The output of the SVD allows a reduced representation of the data matrix to be produced, which describes the data in terms of basis vectors and their associated time-dependent amplitude vectors, and the number of non-zero singular values required to reconstruct the data matrix are an estimate of the number of species involved in the reaction. The amplitude vectors containing significant kinetic information were fitted with a single exponential process giving the rate constants for the unfolding of the native state. See Fig. 3c and Supplementary Fig. 3.

S2: Supplementary Table

Supplementary Table 1: Relative populations of the misfolded state^a for all tandem constructs refolded and ‘never-unfolded’ calculated as described in Methods.

		Percentage of Misfolded Species	
	Tandem	Refolded	‘Never-unfolded’
Linker-free ^b	I27(C3)-I27(C83)	5.7±0.5	0.7±0.6
	I27(C3)-I32(C83)	4.6±0.6	0.8±0.3
	I27(C3)-I28(C83) ^c	0.6±0.1	0.5±0.2
Linker	I27(C3)-I27(C38)	5.5±0.2	0.7±0.1
	I27(C3)-I28(C83) ^{c,d}	0.4	0.4
	I27(C3)-I27-I27(C83):		
	<i>Dimer-like misfolded state^e</i>	9±2	2.1±0.1 ^g
	<i>Monomer-like misfolded state^f</i>	2.8±0.6	0.6±0.3

^a The misfolded state is defined as bursts with $E \geq 0.8$ or the integrated fit of the histogram peak for the high FRET population.

^b In linker-free constructs there are no added residues between domains, while linker constructs have four added amino acids (RSEL) between domains.

^c Only analysed by summing bursts assigned to each ‘population’ (as described in Methods)

^d Values quoted for one measurement made for each experiment.

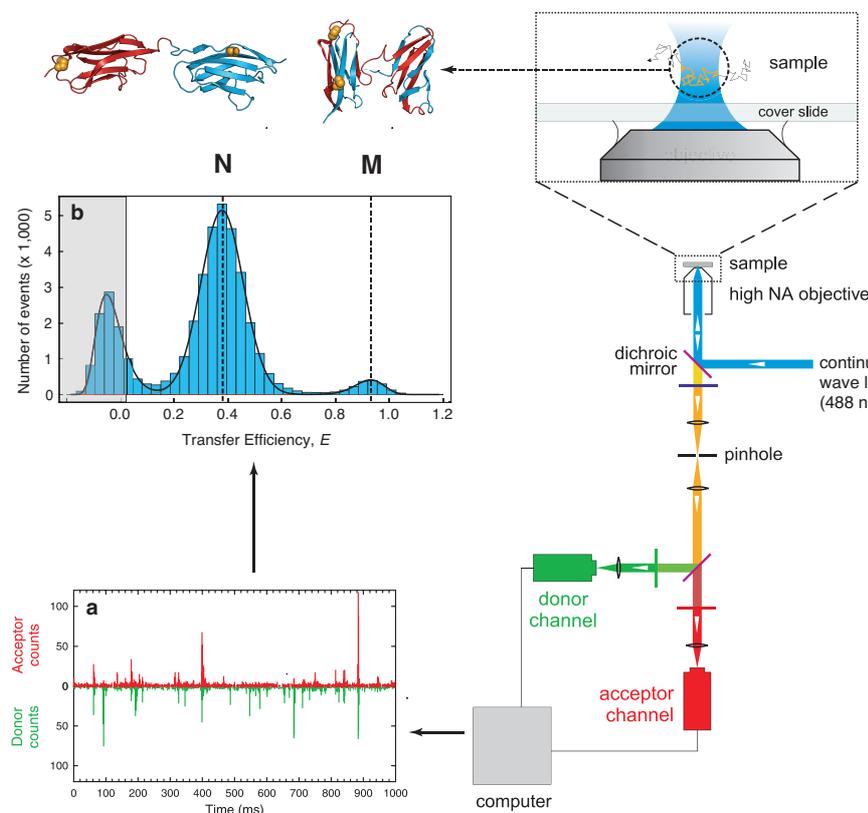
^e Misfolded state where the dyes have the transfer efficiency of the natively folded two-domain tandem.

^f Misfolded state where the dyes have the transfer efficiency of the natively folded monomer.

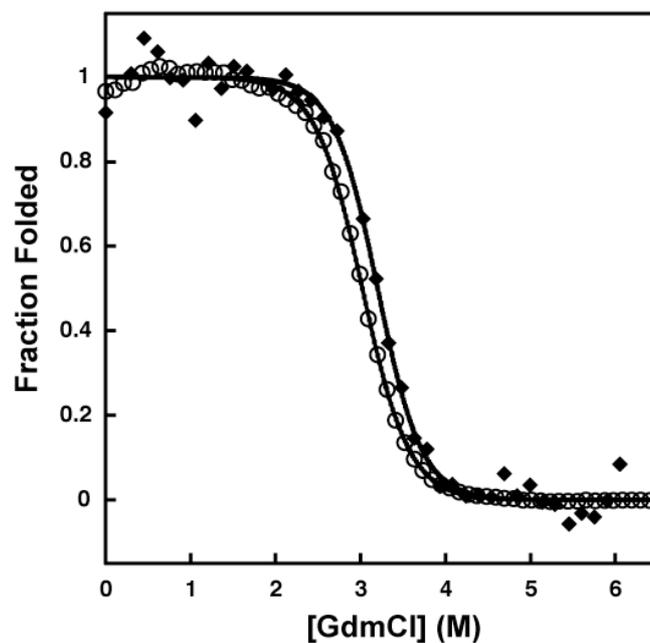
^g Due to the overlap between the E ranges of the correctly folded and dimer-like misfolded state in the refolded tandem histogram, this value was estimated by quantifying the number of bursts within the E range corresponding to the dimer-like species that fall outside the fit of the correctly folded peak in the never-unfolded tandem histogram.

All errors quoted are the unbiased sample standard deviation calculated from two 8-10 hr measurements.

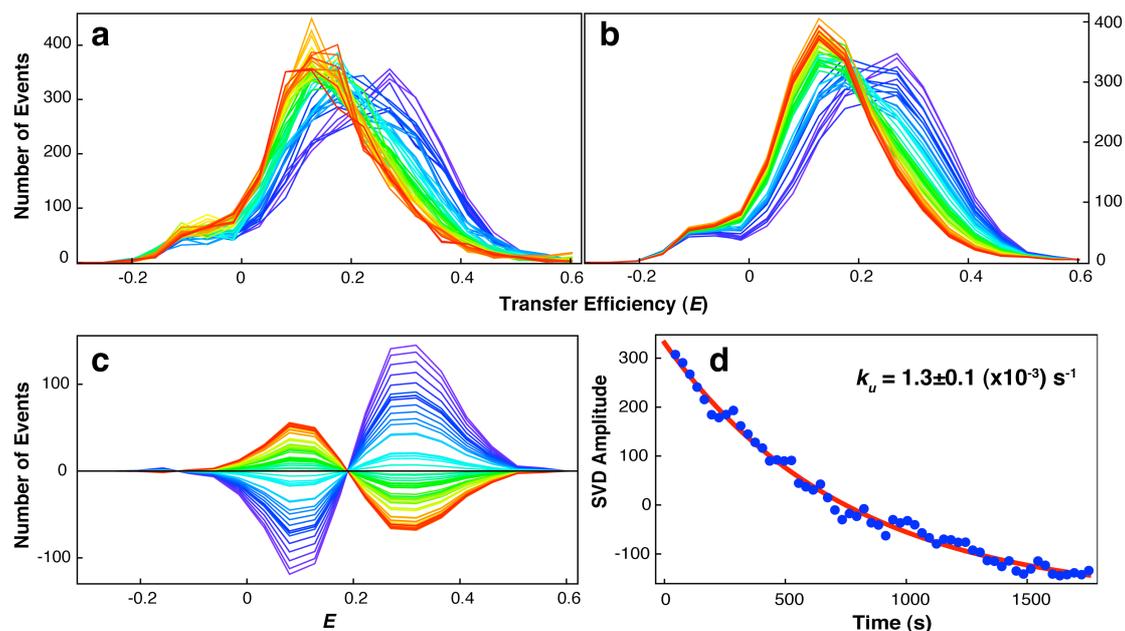
S3: Supplementary Figures



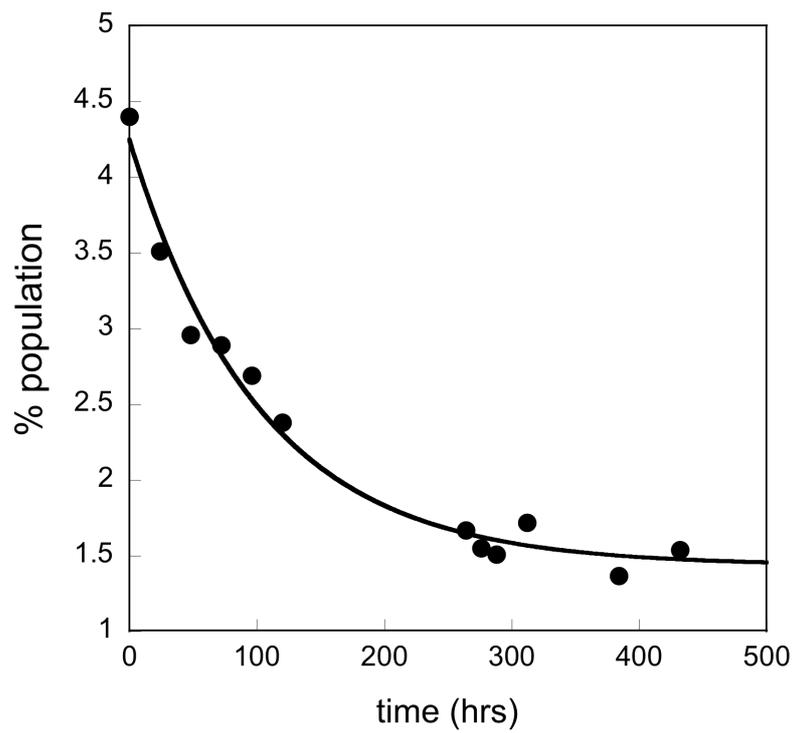
Supplementary Figure 1: Overview of a Single-Molecule FRET Experiment. The diagram on the right shows the main components of the instrument used. For these experiments, the proteins are labelled in specific positions with a suitable donor and acceptor fluorophore. If a protein molecule diffuses through the focused laser beam (circled; upper right), excitation of the donor may result in emission of a fluorescence photon or the transfer of the excitation energy to the acceptor, which can then emit a photon. (a) A typical signal trace of fluorescence photons detected in the donor and acceptor channels; each burst corresponds to a freely diffusing molecule passing through the confocal volume. The efficiency of the energy transfer, E , depends on the separation R of donor and acceptor via $E(R) = (1+(R/R_0)^6)^{-1}$, where R_0 is the Förster radius characteristic of the dye pair^{4,10}. FRET can thus be used to map intramolecular distances in the range of 3 to 8 nm for the R_0 of 5.4 nm³⁹ used in this work. Experimentally, E can be calculated for each burst via $E = (n_A/n_A+n_D)$ where n_A and n_D are the corrected number of acceptor and donor photons detected, respectively (see **Methods**). (b) A two-dimensional histogram of a refolded sample of I27-I27, which shows subpopulations that can be attributed to the correctly folded (N) and misfolded (M) states, and molecules without an active acceptor²⁸ (shaded area).



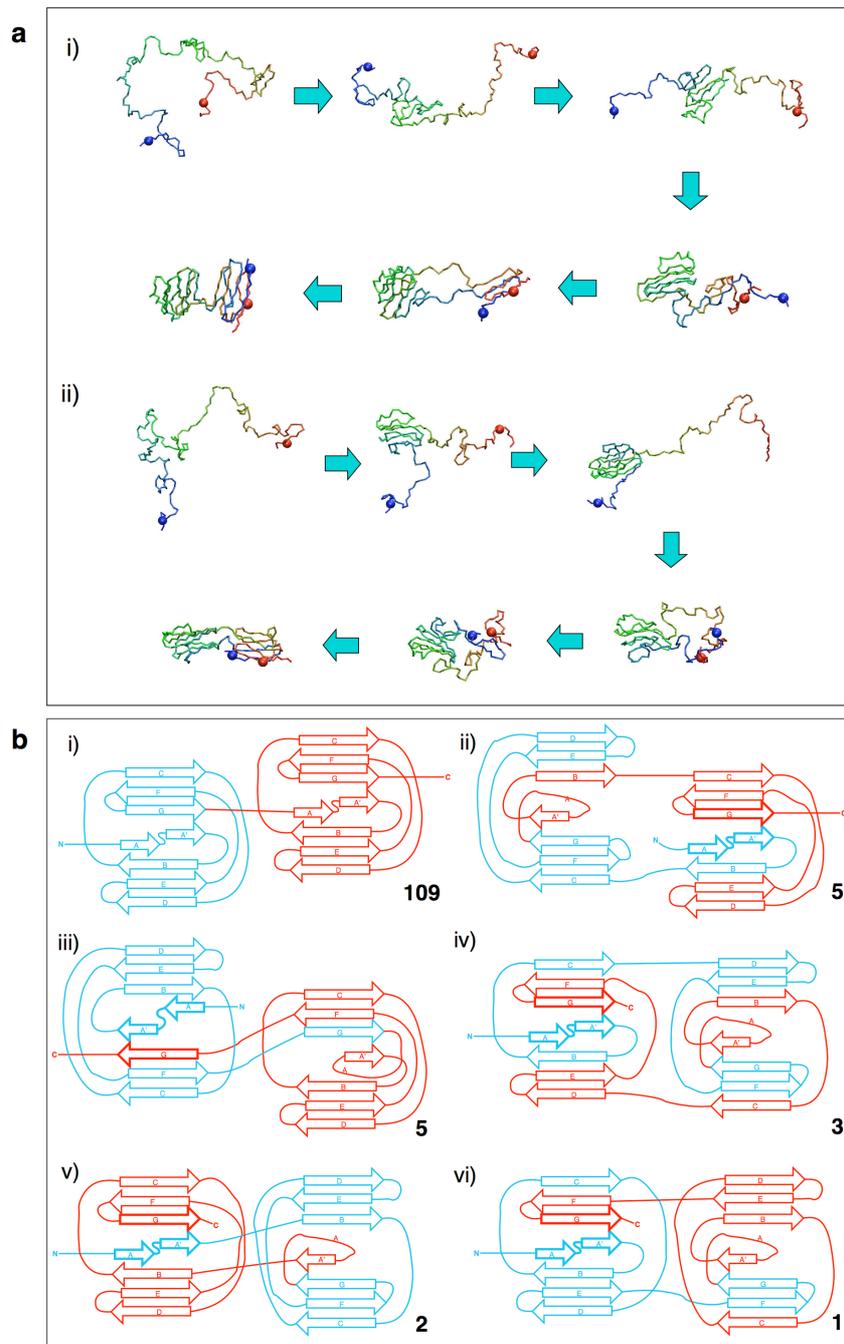
Supplementary Figure 2: Comparison of Ensemble Equilibrium Denaturation for I27wt and Doubly-Labelled I27. Fraction of protein folded as a function of denaturant concentration for I27wt (open circles) (data from reference 12) and doubly-labelled I27 (\blacklozenge) with $\Delta G_{D-N}^{H_2O}$ of 7.6 ± 0.1 and 8.0 ± 0.1 kcal mol⁻¹, respectively (calculated using the average m -value, $\langle m \rangle$,¹²). Mutation and labelling do not significantly alter the stability of I27.



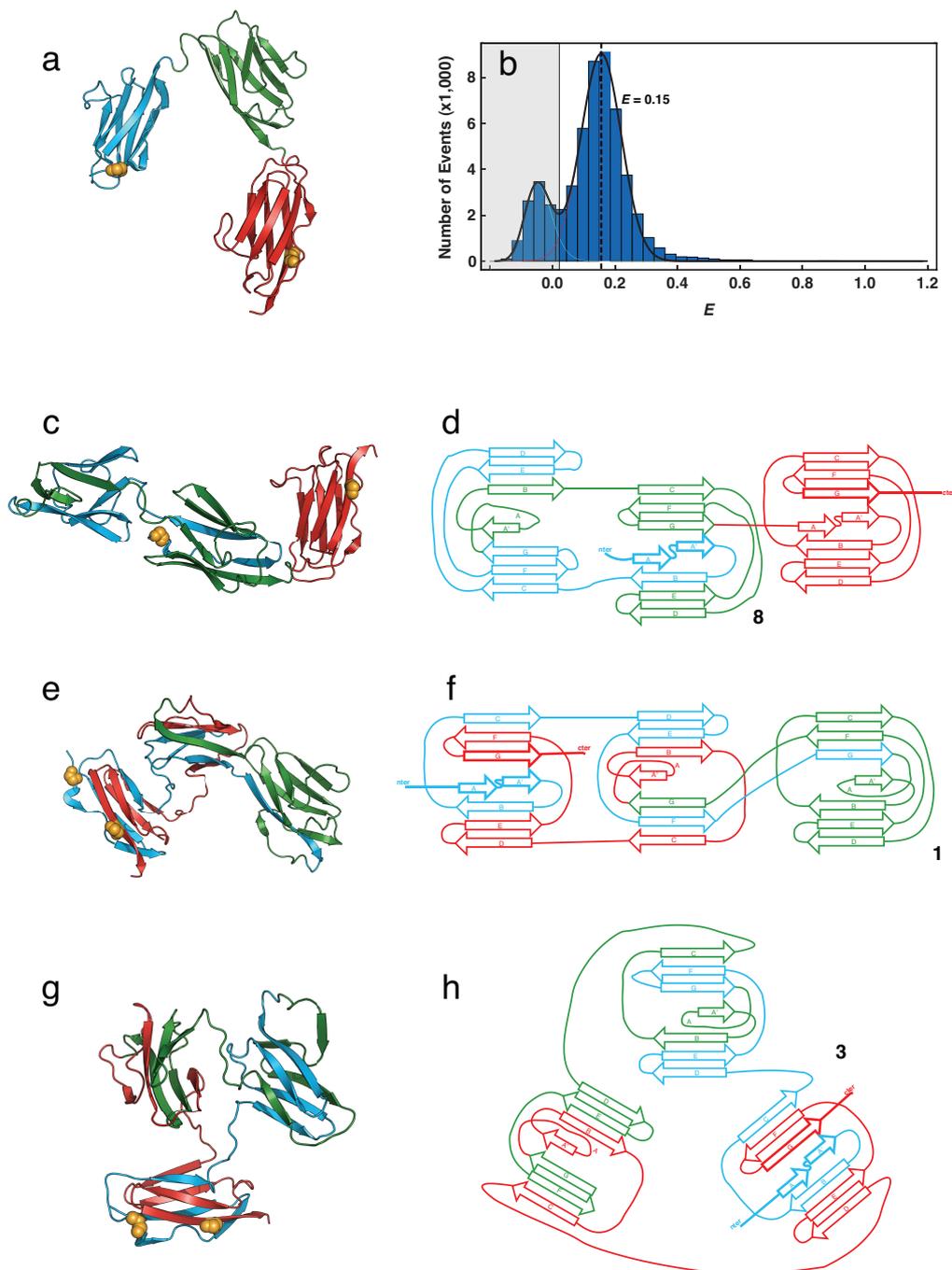
Supplementary Figure 3. SVD analysis of the Native State Unfolding in Single-Molecule Measurements. (a) Transfer efficiency histograms as a function of time (progressing from blue to red) of refolded doubly-labelled I27-I27 unfolding in 3.95 M (as described in **Methods**), for bursts where $-0.2 \leq E \leq 0.6$ (constructed as described in Supplementary Information). (b) Reconstructed transfer efficiency histograms from an SVD of the data in (a) using two components (which describes the process within experimental error). (c) Time evolution of the second basis vector from the SVD of the data in (a) (progressing from blue to red). (d) Kinetics from the second amplitude vector of the SVD of the data in (a) fitted to a single exponential process. It is noted that the first amplitude vector did not contain any kinetic information.



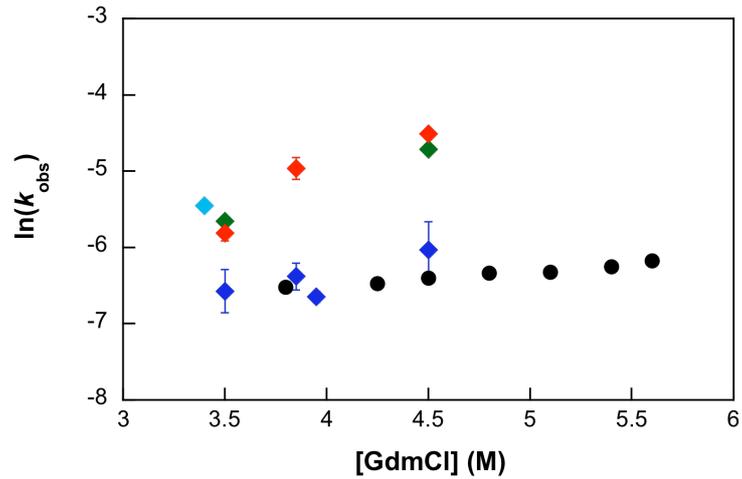
Supplementary Figure 4: Depletion of the Misfolded Species of I27-I27 Over Time in Absence of Denaturant. The points represent the fractional population of the misfolded species. Data are fitted with a single exponential function yielding a rate constant of $2.7 \times 10^{-6} \text{ s}^{-1}$.



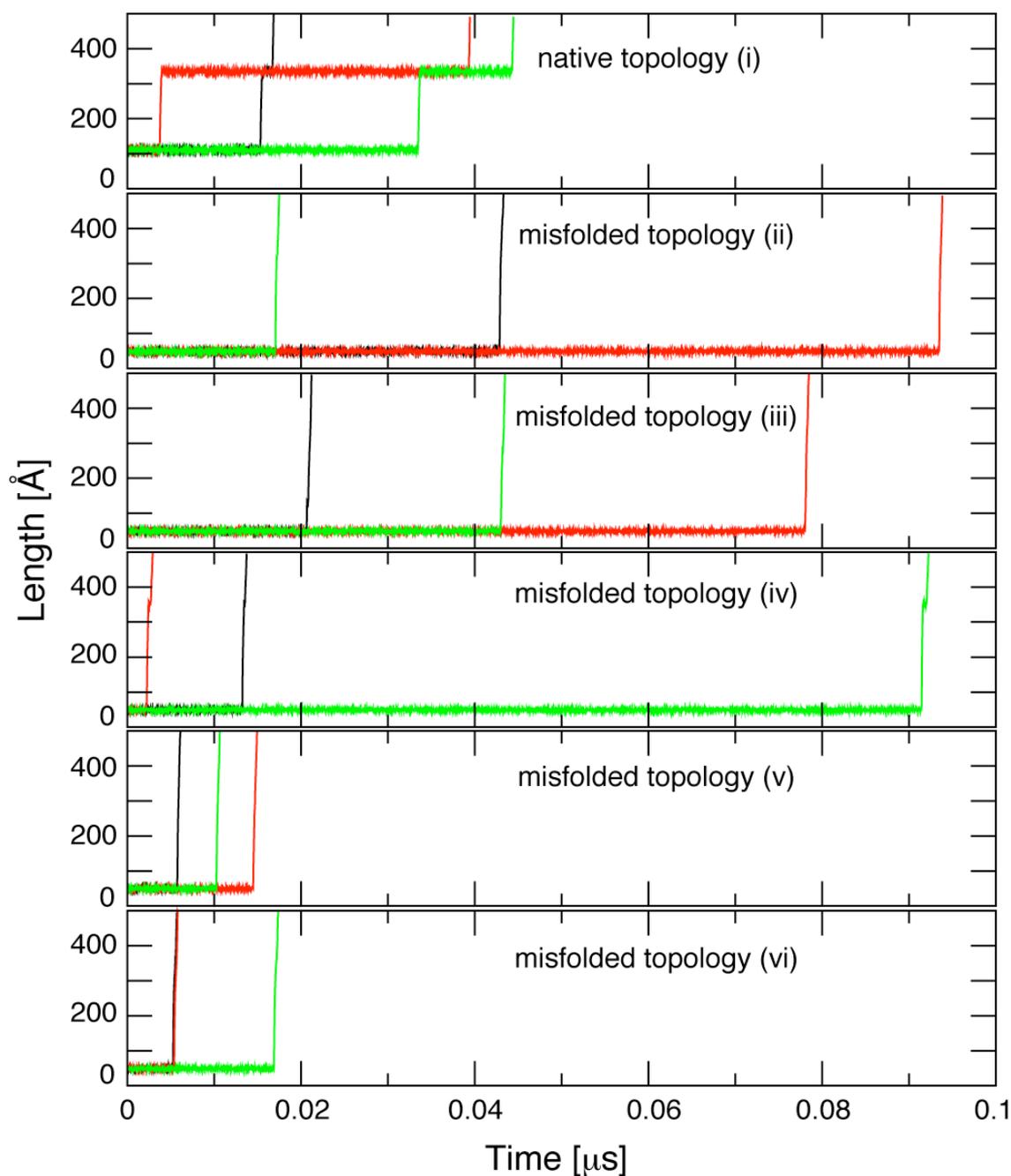
Supplementary Figure 5: I27-I27 Gō-like Model Folding Simulations: Trajectories and Topologies. **(a)** Two observed folding trajectories resulting in misfolded, strand-swapped topologies (i) and (ii) correspond to the topologies (vi) and (iv) in panel **(b)**, respectively.) Attachment sites of the dyes are indicated by spheres (C3 in domain 1 and C83 in domain 2, blue and red, respectively). In all the misfolded topologies the central misfolded domain forms first. **(b)** All topologies seen in 125 completed trajectories. The number of trajectories which resulted in each topology annotated in bold: (i) the correctly folded, native tandem, (ii)-(vi) five strand-swapped topologies observed, classified as misfolded topologies. The N- and C- terminal strands, where the dyes are located, are shown in bold in the misfolded topologies.



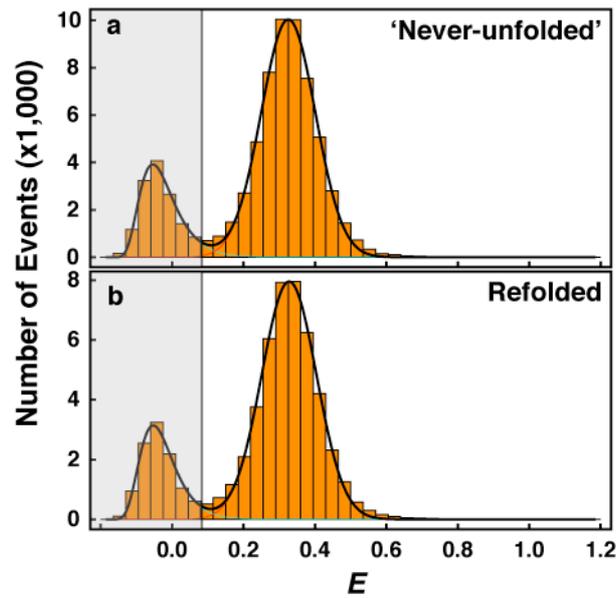
Supplementary Figure 6: I27-I27-I27 Gō-like Model Folding Simulations: Structures, Histogram and Topologies. (a) Natively folded I27-I27-I27 and (b) its “never-unfolded” histogram. 64 folding trajectories were generated: the numbers in black denote the number of trajectories which resulted in each topology shown. In total 18 trajectories resulted in “dimer-like” misfolded structures: the structure (c) and topology (d) are shown for one species formed, but all topologies are like those formed for I27-I27 (Supplementary Fig. 4) with a third domain correctly folded. 4 trajectories resulted in “monomer-like” misfolded structures, both of which are shown: structures (e & g) and topologies (f & h). In all structures the dye attachment sites are shown with golden spheres. In the misfolded topology diagrams the N- and C-terminal strands, where the dyes are located, are shown in bold.



Supplementary Figure 7: Unfolding Rate Constants: I27wt from ensemble measurements (black circles¹²); native state of I27-I27 (with an RSEL linker) measured in single-molecule experiments (blue diamonds); misfolded states (all measured in single-molecule experiments) of I27-I32 (linker-free), I27-I27 (linker-free) and I27-I27 (with an RSEL linker), green, light blue and red diamonds, respectively. Note that error bars are plotted for all data points.



Supplementary Figure 8: Unfolding Trajectories in the Presence of a Pulling Force. Trajectories from simulations, run with a constant pulling force of 150 pN acting between the ends of the chain, for randomly chosen structures for each of the six I27-I27 topologies generated in the G δ -like Model Folding Simulations (Supplementary Figure 5). Three unfolding trajectories, initiated with a different random seed (denoted by different colours), were generated for each topology, monitoring the distance between the chain termini (simulations were terminated once the length exceeded 500 Å). (i)-(vi) numbering is as in Supplementary Figure 5. Only the native topology (i) shows a stable intermediate corresponding to having one domain unfolded with the other remaining folded; all misfolded topologies essentially unfold in one step as observed for the misfolded species in the AFM experiments⁹.



Supplementary Figure 9: Addition of a Linker does not Result in Misfolding in I27-I28. Transfer efficiency histograms for doubly-labelled I27-I28, with an RSEL linker between the domains, (a) ‘never-unfolded’ and (b) refolded. No misfolded species are formed. Histograms are fitted with Gaussian or log-normal distributions.

S4: Supplementary References:

- 38 Henry, E. R. & Hofrichter, J. Singular Value Decomposition - Application to Analysis of Experimental-Data. *Methods in Enzymology* 210, 129-192, (1992).
- 39 Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419, 743-747, (2002).