Published in final edited form as:
*Nature*. ; 474(7353): 662–665. doi:10.1038/nature10099.

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

**Madeleine B. Borgia**[*], **Alessandro Borgia**[=], **Robert B. Best**[*], **Annette Steward**[*], **Daniel Nettels**[=], **Bengt Wunderlich**[=], **Benjamin Schuler**[=], and **Jane Clarke**[*]

[*]University of Cambridge Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, UK
[=]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

## Abstract

A large range of debilitating medical conditions[1] are linked to protein misfolding, which may compete with productive folding particularly in proteins containing multiple domains[2]. With 75% of the eukaryotic proteome consisting of multidomain proteins, how is inter-domain misfolding avoided? It has been proposed that maintaining low sequence identity between covalently linked domains is a mechanism to avoid misfolding[3]. Here we use single-molecule Förster Resonance Energy Transfer (FRET) experiments[4,5] to detect and quantify rare misfolding events in tandem Ig domains from the I-band of titin under native conditions. About 5.5% of molecules with identical domains misfold during refolding *in vitro* and form a surprisingly stable state with an unfolding half time of several days. Tandem arrays of immunoglobulin-like (Ig-like) domains in humans exhibit significantly lower sequence identity between neighbouring domains than between non-adjacent domains[3]. In particular, the sequence identity of neighbouring domains has been found to be preferentially below 40%[3]. Interestingly we observe no misfolding for a tandem of naturally neighbouring domains with low sequence identity (24%), whereas misfolding occurs between domains which are 42% identical. Coarse-grained molecular simulations predict the formation of domain-swapped structures, which 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 appears 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 inter-domain misfolding *in vivo*. While co-translational, “domain-by-domain” folding is assumed to be important for avoiding misfolding[6], many proteins that are long-lived or subject to tensile forces will fold and unfold numerous times during their lifetime and may thus be particularly vulnerable to misfolding. The giant muscle protein titin, for example, undergoes reversible domain unfolding which may play a role in muscle elasticity[7].

Correspondence and requests for materials should be addressed to J.C. (jc162@cam.ac.uk) and B.S. (schuler@bioc.uzh.ch)..
**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.** The authors declare no competing interests.



**Supplementary Information:** Supporting information is provided with this document.


Europe PMC Funders Author Manuscripts

Single-molecule techniques are ideal for detecting rare events[8] such as misfolding in native conditions. Indeed, the first evidence for misfolding of adjacent domains in long tandem arrays of the well-characterised 27th domain from the I-band of titin, I27, was obtained using single-molecule atomic force microscopy (AFM)[9]. An alternative approach is single-molecule FRET[4,5], whose great sensitivity allows the detection of very small populations. FRET enables the mapping of intramolecular distances by means of the distance-dependent efficiency of excitation energy transfer between a donor and acceptor fluorophore attached to specific positions of the protein[10] (for details, see Supplementary Fig. 1).

We hypothesised 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 FRET[11]. 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 via cysteine residues engineered on the protein surface. For a misfolded domain to be formed by strands from domains 1 and 2, we predicted that these two strands must be adjacent for this domain to have the mechanical properties observed in the previous AFM experiments[9]. The correctly folded tandem would then have low transfer efficiency, while 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') show single, correctly folded populations with transfer efficiencies ($E$) of 0.37 (±0.01) and 0.93 (±0.01) for the tandem I27-I27 and monomer I27, respectively (Fig. 2 a & b).

We conducted refolding experiments by diluting unfolded I27-I27 into refolding buffer. The resulting transfer efficiency histograms now exhibited 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 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(±0.2)% of the molecules are found in the misfolded form.

Based on the results of the AFM studies[9] 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 ($\tau \approx 34$ minutes[12]). We therefore investigated the unfolding kinetics of the misfolded state[13] (Fig. 3a). At high GdmCl concentrations the decay in the number of high-transfer efficiency events ($E > 0.8$), corresponding to the misfolded state, is fitted well by a single exponential (Fig. 3b) with rate constants slightly higher than the unfolding rate constant for I27wt determined in ensemble measurements[12] (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, converting to the correctly folded form only on a time scale 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, partially folded intermediates[14-16] sometimes termed "misfolded" because they contain some non-native interactions[17].

An explanation for the slow unfolding under native conditions is suggested by folding simulations of I27-I27 with a Gō-like model[18]. 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. While 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. Since 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 slow[19].

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. 2 d). The proportion of monomer-like (high-FRET) species in the trimeric tandem is significantly lower than before (2.8(±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(±2)%) was instead almost twice as high as in the two-domain tandem (5.5(±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 oligimerisation 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 now allow us to test this hypothesis directly by investigating mixed tandem constructs, I27-I28 and I27-I32 (Fig. 4b). Indeed I27-I28, natural neighbours in titin with only 24% sequence identity, did not yield any detectable population of high-FRET misfolded species upon refolding (Fig. 4c,d). However, misfolding is seen upon refolding of I27-I32 (sequence identity 42%) (Fig. 4e,f), to the same extent as I27-I27 (Supplementary Table 1). This misfolded species also unfolds with the same rate constant as that of I27-I27 (Supplementary Fig. 7). The I27-I32 misfold is consistent with previous experiments showing chimeric 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 species[19].

Misfolding in our experiments is more frequent than had been observed in AFM experiments[9], 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 twice the chain length is released upon unfolding. Our results are entirely compatible with this finding. While the misfolded species has two folded domains, and is thus stable in folding conditions, only the terminal domain would experience shearing of the H-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 I27[26], resulting in the same mechanical stability. Since 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).

In summary, 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 of the puzzle in understanding the problems encountered during the crucial evolutionary transition from single to multi-domain proteins.

## Methods Summary

For details of protein production, ensemble equilibrium measurements and labelling see **Methods**. Single-molecule experiments, instrumentation, data reduction and analysis are detailed in **Methods**. The resulting relative populations from all experiments and analysis techniques are summarised in Supplementary Table 1. Folding simulations using a Gō-like model were run using the CHARMM code[27] as described in **Methods**. For details of mechanical unfolding simulations see **Methods**.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

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 for 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. 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 helpful discussions and valuable contributions to data analysis.

## Methods

## Protein expression and labelling

Cysteine residues were introduced by site-directed mutagenesis. E3C in domain 1 (always I27) and N83C in domain 2 if I27 or I32, and K83C for I28 (with the numbering relative to a single domain), for the two domain constructs and E3C in domain 1 and N83C in domain 3 for the three domain construct of I27. 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. Both dyes were mixed simultaneously with reduced protein in equimolar ratios and incubated at 4°C for ~10 hours. Un-reacted 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, Uppsala, Sweden). 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., CA, USA) monitoring intrinsic tryptophan fluorescence as described previously[29], 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, Milpitas, CA, USA) and an Olympus UplanApo 60x/1.20W objective. After a dichroic mirror that separates excitation and emission light (500DCXR, Chroma Technology, Rockingham, VT), fluorescence emission passed through a 100 μm pinhole and was split by a second dichroic mirror (585DCXR, Chroma Technology) into donor and acceptor fluorescence. Donor fluorescence then passed a filter (ET525/50M, Chroma Technology) before being focused onto a single-photon avalanche diode (MPD 100ct, Micro Photon Devices, Bolzano, Italy) while acceptor fluorescence passed a filter (QT 650/100) before being focused onto a single-photon avalanche diode (SPCM-AQR-13, PerkinElmer Optoelectronics, Vaudreuil, QC, Canada). The arrival time of every photon was recorded with a two-channel time-correlated single-photon counting module (PicoHarp300, PicoQuant, Berlin, Germany). 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 proteins[31], while the photoprotective agents β-mercaptoethanol (Sigma) and cysteamine hydrochloride (Sigma) were employed to minimise chromophore damage[32]. 'Never-unfolded' experiments 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 4.4M GdmCl (PBS, 0.01% Tween, 10 mM β-mercaptoethanol) 1:100 with single-molecule buffer, to a final GdmCl concentration of 0.04 M. 8 to 10 hour measurements were made for all constructs with one or more repeats. The absence of aggregates was ensured in all experimental conditions as previously described[33]. Fluorescence bursts were identified by combining successive photons separated by 150 μs or less, and events comprised of 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 $E$ 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 summed, and the relative populations of misfolded species computed. Only the latter method was amenable for 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 summarised 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 refolded experiments above, but with higher protein concentration) with GdmCl in single-molecule buffer. A moving-window analysis[13] was applied to the time-resolved photon trajectory with a window size ($\Delta t$) of 120s. 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. 4a). 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. 4b). 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 experiments is detailed in Supplementary Information and Supplementary Fig. 5.

## Simulations

A coarse-grained Gō-like model was generated based on the structure of I27 (1tit.pdb[36]) using a standard procedure[37]. Briefly, all bond lengths are fixed by constraints, harmonic terms are used for the angles, a knowledge based potential 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 3 $k_B$T, by using the free energy barrier projected onto the fraction of native contacts as a lower bound. 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 cut-off of 0.15 nm.

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 the end-end distance. The CHARMM molecular simulation package was used for all calculations[27].

## References

1. Gregersen N, Bross P, Vang S, Christensen JH. Protein misfolding and human disease. Annu Rev Genomics Hum Genet. 2006; 7:103–124. [PubMed: 16722804]
2. Jaenicke R, Seckler R. Protein misassembly in vitro. Adv. Protein Chem. 1997; 50:1–59. [PubMed: 9338078]
3. Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. Nature. 2005; 438:878–881. [PubMed: 16341018]
4. Joo C, Balci H, Ishitsuka Y, Buranachai C, Ha T. Advances in single-molecule fluorescence methods for molecular biology. Annu. Rev. Biochem. 2008; 77:51–76. [PubMed: 18412538]
5. Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Curr. Opin. Struct. Biol. 2008; 18:16–26. [PubMed: 18221865]
6. Fedorov AN, Baldwin TO. Cotranslational protein folding. J. Biol. Chem. 1997; 272:32715–32718. [PubMed: 9407040]

7. Li H, et al. Reverse engineering of the giant muscle protein titin. Nature. 2002; 418:998–1002. [PubMed: 12198551]
8. Borgia A, Williams PM, Clarke J. Single-molecule studies of protein folding. Annu. Rev. Biochem. 2008; 77:101–125. [PubMed: 18412537]
9. Oberhauser AF, Marszalek PE, Carrion-Vasquez M, Fernandez JM. Single protein misfolding events captured by atomic force microscopy. Nat. Struct. Biol. 1999; 6:1025–1028. [PubMed: 10542093]
10. Stryer L. Fluorescence energy transfer as a spectroscopic ruler. Annu. Rev. Biochem. 1978; 47:819–846. [PubMed: 354506]
11. Gambin Y, et al. Direct single-molecule observation of a protein living in two opposed native structures. Proc. Natl. Acad. Sci. USA. 2009; 106:10153–10158. [PubMed: 19506258]
12. Fowler SB, Clarke J. Mapping the folding pathway of an immunoglobulin domain: structural detail from phi value analysis and movement of the transition state. Structure. 2001; 9:355–366. [PubMed: 11377196]
13. Hofmann H, et al. Single-molecule spectroscopy of protein folding in a chaperonin cage. Proc. Natl. Acad. Sci. USA. 2010; 107:11793–11798. [PubMed: 20547872]
14. Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Gianni S. Folding and misfolding in a naturally occurring circularly permuted PDZ domain. J Biol Chem. 2008; 283:8954–8960. [PubMed: 18263589]
15. Gianni S, et al. Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat Struct Mol Biol. 2010
16. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. A transient and low-populated protein-folding intermediate at atomic resolution. Science. 2010; 329:1312–1316. [PubMed: 20829478]
17. Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol. 2002; 9:209–216. [PubMed: 11875516]
18. Yang S, et al. Domain swapping is a consequence of minimal frustration. Proc. Natl. Acad. Sci. USA. 2004; 101:13786–13791. [PubMed: 15361578]
19. Arora P, Hammes GG, Oas TG. Folding mechanism of a multiple independently-folding domain protein: double B domain of protein A. Biochemistry. 2006; 45:12312–12324. [PubMed: 17014084]
20. Jaenicke R. Folding and association of proteins. Prog Biophys Mol Biol. 1987; 49:117–237. [PubMed: 3327098]
21. Straub JE, Thirumalai D. Principles governing oligomer formation in amyloidogenic peptides. Curr Opin Struct Biol. 2010; 20:187–195. [PubMed: 20106655]
22. Bennett MJ, Sawaya MR, Eisenberg D. Deposition diseases and 3D domain swapping. Structure. 2006; 14:811–824. [PubMed: 16698543]
23. Mitraki A. Protein aggregation from inclusion bodies to amyloid and biomaterials. Adv Protein Chem Struct Biol. 2010; 79:89–125. [PubMed: 20621282]
24. Borgia A, Steward A, Clarke J. An effective strategy for the design of proteins with enhanced mechanical stability. Angew. Chem. Int. Ed. Engl. 2008; 47:6900–6903. [PubMed: 18666188]
25. Balamurali MM, et al. Recombination of protein fragments: a promising approach toward engineering proteins with novel nanomechanical properties. Protein Sci. 2008; 17:1815–1826. [PubMed: 18628239]
26. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys. J. 1998; 75:662–671. [PubMed: 9675168]
27. Brooks BR, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 1983; 4:187–217.
28. Dahan M, et al. Ratiometric measurement and identification of single diffusing molecules. Chemical Physics. 1999; 247:85–106.

Europe PMC Funders Author Manuscripts

## References


29. Scott KA, Steward A, Fowler SB, Clarke J. Titin; a multidomain protein that behaves as the sum of its parts. J. Mol. Biol. 2002; 315:819–829. [PubMed: 11812150]
30. Steward A, Toca-Herrera JL, Clarke J. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. Protein Sci. 2002; 11:2179–2183. [PubMed: 12192073]
31. Schuler B, Lipman EA, Eaton WA. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature. 2002; 419:743–747. [PubMed: 12384704]
32. Nettels D, et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc. Natl. Acad. Sci. USA. 2009
33. Hillger F, et al. Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. Angew. Chem. Int. Ed. Engl. 2008; 47:6184–6188. [PubMed: 18618555]
34. Schuler B. Single-molecule fluorescence spectroscopy of protein folding. Chemphyschem. 2005; 6:1206–1220. [PubMed: 15991265]
35. Hoffmann A, et al. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. Proc. Natl. Acad. Sci. USA. 2007; 104:105–110. [PubMed: 17185422]
36. Improta S, Politou AS, Pastore A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure. 1996; 4:323–337. [PubMed: 8805538]
37. Karanicolas J, Brooks CL 3rd. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J. Mol. Biol. 2003; 334:309–325. [PubMed: 14607121]

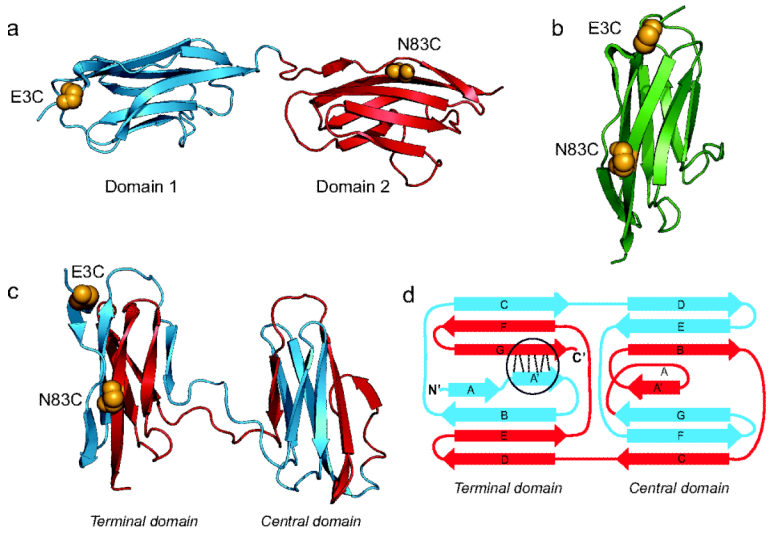


**Figure 1.**

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 (1tit.pdb) with labelling positions corresponding to those expected for the misfolded state of I27-I27. (**c**) One of the domain-swapped misfolded state structures formed in Gō-model simulations. (**d**) Schematic of this misfolded state topology: 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.




Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

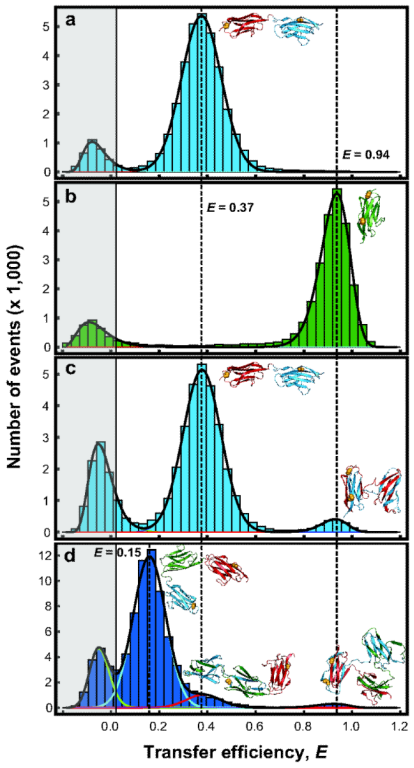


**Figure 2.**

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 fluorophore[28]. 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.

Europe PMC Funders Author Manuscripts

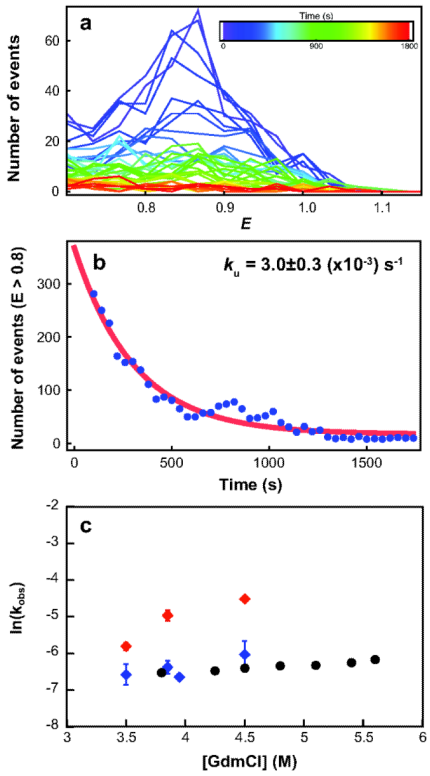


**Figure 3.**

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 120s that was shifted by 30s for each increment (inset, colour key). (**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 rate constants are unaffected by different window sizes or the use of non-overlapping windows. (**c**) Unfolding rate constants for I27wt monomer (black) (ensemble data from[12])*, and for the misfolded and natively folded states of I27-I27 from single-molecule measurements (red and blue, respectively). The error bars represent the standard error of the fit (see **Methods**). Note that 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.

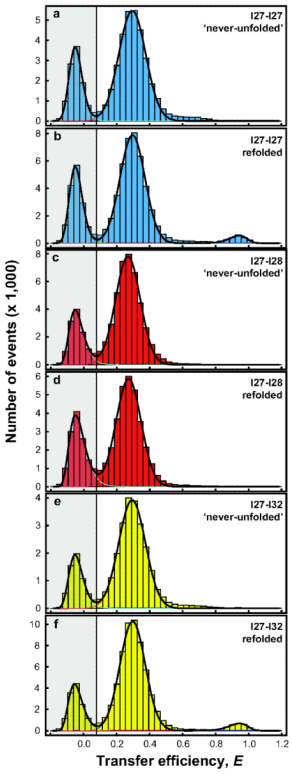


**Figure 4.**

Transfer Efficiency Histograms of Tandem Constructs with Identical and Non-identical Domains. (**a**) and (**b**) I27-I27 'never-unfolded' control and refolded, respectively. (**c**) and (**d**) I27-I28 'never-unfolded' control and refolded, respectively. (**e**) and (**f**) I27-I32 'never-unfolded' control and refolded, respectively. In order to mimic the natural protein, there was no linker added between the domains in these experiments. Note that the frequency of misfolding was the same for I27-I27 with and without the linker, 5.5(±0.2)% and 5.7(±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).




Europe PMC Funders Author Manuscripts