

Direct Observation of Tug-of-War during the Folding of a Mutually Exclusive Protein

Qing Peng and Hongbin Li*

Department of Chemistry, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada

Received April 29, 2009; E-mail: Hongbin@chem.ubc.ca

Abstract: Although most protein folding studies are carried out on single-domain proteins, over two-thirds of proteins in proteomes are built up from multiple individually folded domains. A significant fraction of these multidomain proteins are domain-insertion proteins, in which one guest domain is inserted into a surface loop of a host protein. Intricate thermodynamic and kinetic coupling between the two domains can have a profound impact on their folding dynamics. Here we use an engineered mutually exclusive protein as a model system to directly illustrate one such complex dynamic process: the “tug-of-war” process during protein folding. By inserting a guest protein I27w34f into a host protein GB1-L5 (GL5), we engineered a novel, mutually exclusive protein, GL5/I27w34f, in which only one domain can remain folded at any given time due to topological constraints imposed by the folded structures. Using stopped-flow techniques, we obtained the first kinetic evidence that the guest and host domains engage in a folding tug-of-war as they attempt to fold, in which the host domain folds rapidly into its three-dimensional structure and is then automatically unfolded, driven by the folding of the guest domain. Our results provided direct evidence that protein folding can generate sufficient mechanical strain to unravel a host protein. Using single-molecule atomic force microscopy, we provide direct evidence for the existence of a conformational equilibrium between the two mutually exclusive conformations. Our results highlight important roles played by the intricate coupling between folding kinetics, thermodynamic stability, and mechanical strain in the folding of complex multidomain proteins, which cannot be addressed in traditional single-domain protein folding studies.

Introduction

Domains are fundamental building blocks, both structurally and functionally, for proteins. Over two-thirds of proteins in the prokaryote and eukaryote proteomes display modular architecture and are built from multiple domains to accomplish widely diversified functions.^{1,2} In most multidomain proteins, individual domains are arranged in tandem, with one domain following the next in a contiguous order. However, a significant fraction of modular proteins are found to be exceptions to this general pattern.^{3,4} In such proteins, one guest domain is inserted into the loop of another host protein, thus disrupting the sequence continuity of the host protein and creating domain insertion proteins. Among these domain insertion proteins, engineered mutually exclusive proteins^{5–7} are of particular interest for the study of protein folding/unfolding dynamics. In addition, such proteins can potentially serve as a new class of cytotoxic proteins that can be activated by cell-specific effector molecules.⁵

Mutually exclusive proteins are a special class of engineered domain insertion proteins.⁵ In mutually exclusive proteins, the distance between the N–C termini of the inserted guest protein is significantly longer than the distance between the two ends of the surface loop of the host protein where the guest domain is inserted. Topological constraints arising from such structural incompatibility will allow only one domain to remain folded at any given time, leading to the “mutual exclusiveness” of folding between the two domains. Pioneering work on mutually exclusive proteins using thermodynamic methods has demonstrated the mutual exclusiveness of folding between two domains on the ensemble level using a barnase-ubiquitin-based mutually exclusive protein.^{5–7} However, the kinetic aspects of this mutual exclusiveness remains unexplored, and the mutual exclusiveness between the folding of two domains needs to be tested at the single-molecule level. In this paper, we have designed a novel, mutually exclusive protein composed of the host protein GB1-L5 (GL5)⁸ and the inserted guest protein I27w34f, which is a Trp34Phe mutant of the 27th Ig domain of the muscle protein titin⁹ (Figure 1). The host protein can fold into its native state much faster than its guest protein. Combining fast kinetics and single-molecule atomic force microscopy (AFM) techniques,^{9,10} we report the first kinetic evidence of the “tug-of-war” occurring between the folding of the two

(1) Ponting, C. P.; Russell, R. R. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 45–71.

(2) Vogel, C.; Berzuini, C.; Bashton, M.; Gough, J.; Teichmann, S. A. *J. Mol. Biol.* **2004**, *336*, 809–823.

(3) Selvam, R. A.; Sasidharan, R. *Nucleic Acids Res.* **2004**, *32* (Database issue), D193–D195.

(4) Aroul-Selvam, R.; Hubbard, T.; Sasidharan, R. *J. Mol. Biol.* **2004**, *338*, 633–641.

(5) Radley, T. L.; Markowska, A. I.; Bettinger, B. T.; Ha, J. H.; Loh, S. N. *J. Mol. Biol.* **2003**, *332*, 529–536.

(6) Cutler, T. A.; Loh, S. N. *J. Mol. Biol.* **2007**, *371*, 308–316.

(7) Cutler, T. A.; Mills, B. M.; Lubin, D. J.; Chong, L. T.; Loh, S. N. *J. Mol. Biol.* **2009**, *386*, 854–868.

(8) Li, H.; Wang, H. C.; Cao, Y.; Sharma, D.; Wang, M. *J. Mol. Biol.* **2008**, *379*, 871–880.

(9) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694–3699.

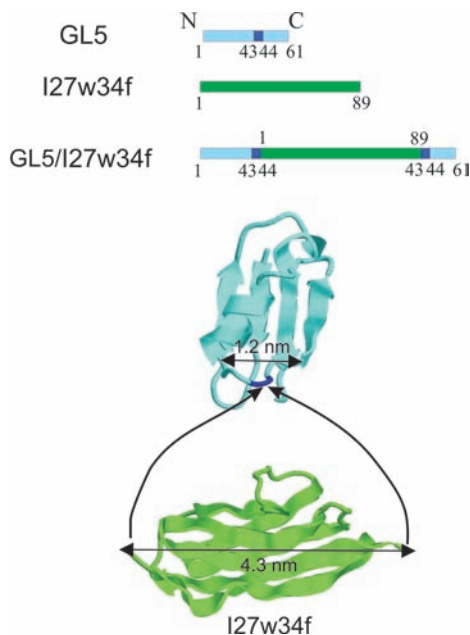


Figure 1. Construction of the designed mutually exclusive protein GL5/I27w34f. The three-dimensional structure of GL5 (cyan) was obtained by structural homology modeling using the Web-based package Swiss-Model by First Approach Mode (<http://swissmodel.expasy.org/>), and the structure of I27w34f was assumed to be the same as that of the wt I27 (green, PDB code 1TIT). In the mutually exclusive protein GL5/I27w34f, the I27w34f domain is inserted into the second loop of the host protein GL5. The duplication of residues 43 and 44 results from the *Ava*I site used in constructing the gene of GL5/I27w34f.

domains in a mutually exclusive protein and the conformational equilibrium between the two possible exclusive conformations. Our results showed that, upon folding, the two domains are engaged in a tug-of-war: the host protein folds into its native state first, with the subsequent slow folding of the guest protein competing with the folded host protein. The folding of the guest domain generates mechanical strain, leading to the mechanical unfolding of the host protein. Our results provide direct experimental evidence that the folding of the guest protein can generate enough mechanical strain to trigger the unfolding of the host protein, and they highlight how intricate coupling between folding kinetics, thermodynamic stability, and mechanical strain determines the outcome of the folding tug-of-war, and thus the final folded state of the mutually exclusive protein. Moreover, our single-molecule AFM results provide direct evidence for the existence of conformational equilibrium between the two mutually exclusive conformations and also reveal that, on rare occasions, a minute fraction of the protein can exist in a conformation with both host and guest domains being folded at the same time. These novel findings clearly demonstrate the great potential of single-molecule AFM in elucidating the folding mechanism of such complex multidomain proteins.

Materials and Methods

Protein Engineering. The gene of the host protein GL5 was constructed as described.⁸ The gene of GL5 contains a 5' *Bam*HI restriction site and 3' *Bg*III and *Kpn*I sites. In addition, the DNA sequence encoding residues 43 and 44 is a non-palindromic *Ava*I site, which allows the insertion of the gene of the guest protein

I27w34f. The gene of the guest protein I27w34f, which carries a single-point mutation w34f, was constructed using the 27th immunoglobulin domain of human cardiac titin (I27) as the template via standard site-directed mutagenesis. I27w34f, carrying a non-palindromic *Ava*I site at its 5' and 3' ends, was then amplified using polymerase chain reaction and subcloned into GL5 between its residues 43 and 44 to obtain the gene of the mutually exclusive protein GL5/I27w34f. GL5/I27w34f was then subcloned into expression vector pQE80L, and its DNA sequence was confirmed by direct DNA sequencing. The final construct GL5/I27w34f, which carries an N-terminal His tag for purification, has the following amino acid sequence:

MRGSHHHHHHGSMDTYKLILNGKTLKGETTTEAVDAATAEK-VFKQYANDNGVGGGLCLIEVEKPLYGVEVFGVGETAHFEIEL-**SEPDVHGQFKLKGQPLAASPDC**EIEDGKKHILILHNCQLGM-TGEVSFQAANTKSAANLKVKELLDGGEWTYDDATKTFVTTERS

where the sequence in italics is from the host domain GL5 and the sequence in bold is from the guest domain I27w34f. The junction between GL5 and I27w34f is Leu-Gly that resulted from the *Ava*I site, the N-terminal Gly-Ser resulted from the *Bam*HI site, and the C-terminal Arg-Ser resulted from the *Bg*III site. The mutually exclusive protein GL5/I27w34f was overexpressed in *Escherichia coli* strain DH5 α and purified by Ni²⁺ affinity chromatography.

The gene of the polyprotein chimera (GB1)₄-GL5/I27w34f-(GB1)₄ was constructed using a previously described step-by-step method,⁹ which is based on the identity of the sticky ends of *Bam*HI and *Bg*III restriction sites after digestion. The polyprotein chimera was overexpressed in *E. coli* strain DH5 α and purified by Ni²⁺ affinity chromatography.

Stopped-Flow Fluorescence and CD Measurements. The thermodynamic stability of isolated GL5 and I27w34f was determined using guanidinium chloride (GdmCl) denaturation by monitoring their circular dichroism (CD) signals at 221 and 230 nm, respectively. The folding kinetics of the mutually exclusive protein GL5/I27w34f was monitored using two different methods: by following its tryptophan fluorescence at 350 nm (excitation wavelength 297 nm) and by monitoring its ellipticity change at 221 nm. The folding kinetics of isolated I27w34f was monitored by its CD signal at 230 nm, and the folding kinetics of isolated GL5 was monitored by following its tryptophan fluorescence at 350 nm. Stopped-flow experiments were carried out on a BioLogic SFM-300 stopped-flow instrument, and CD data were measured using a Jasco-J810 spectropolarimeter flushed with nitrogen gas.

Single-Molecule AFM. Single-molecule AFM experiments were carried out on a custom-built atomic force microscope as described.¹¹ The spring constant of each cantilever (Si₃N₄ cantilevers from Veeco, with a typical spring constant of 40 pN nm⁻¹) was calibrated in solution using the equipartition theorem before and after each experiment. For each experiment, 1 μ L of polyprotein sample (\sim 200 μ g/mL) was loaded into 50 μ L of PBS buffer on a clean glass coverslip. The protein sample was allowed to adsorb onto the coverslip for 5 min before starting the experiment.

Results and Discussion

Design of the Mutually Exclusive Protein GL5/I27w34f. The proposed folding tug-of-war between the two domains in a mutually exclusive protein should depend on the relative thermodynamic stability of the two domains as well as their relative folding kinetics. For example, if the guest protein, whose primary sequence is continuous, folds faster and is also thermodynamically more stable than the host protein, the folding of the guest protein will occur first and result in the thermodynamically stable form of the mutually exclusive protein. In such a scenario, the folding of the more stable guest protein

(10) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109–1112.

(11) Cao, Y.; Li, H. *Nat. Mater.* **2007**, *6*, 109–114.

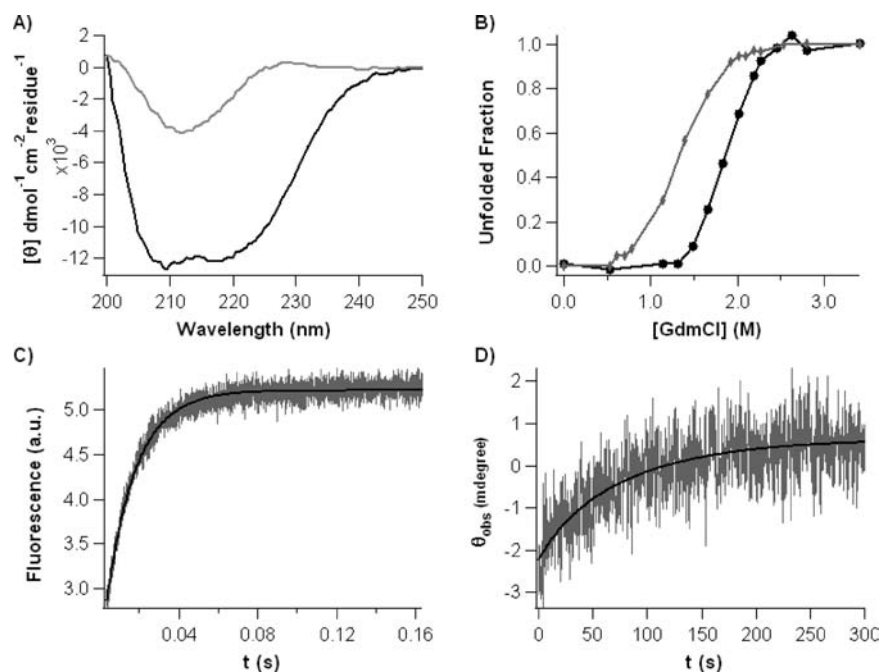


Figure 2. Thermodynamic and kinetic properties of isolated GL5 and I27w34f domains used to construct the mutually exclusive protein GL5/I27w34f. (A) Far-UV CD spectra of isolated GL5 (in black) and I27w34f (in gray) in PBS buffer at pH 7.4. (B) Chemical denaturation curves of GL5 (in gray) and I27w34f (in black) in PBS buffer monitored by CD. The chemical denaturation of GL5 and I27w34f was monitored at 221 and 230 nm, respectively. (C) Folding kinetics of isolated GL5 in 0.3 M GdmCl followed by the tryptophan fluorescence at 350 nm. The black line is a single-exponential fit to the data with a folding rate constant of 67 s^{-1} . (D) Folding kinetics of isolated I27w34f in 0.3 M GdmCl followed by monitoring the CD signal at 230 nm. The black line is a double-exponential fit to the data with folding rate constants of 0.039 and 0.011 s^{-1} in 0.3 M GdmCl.

prevents the folding of the host protein, and no evidence for the tug-of-war of folding can be observed. In contrast, if the host protein, whose primary sequence continuity is disrupted by the insertion of the guest protein, folds much faster than the guest protein while the thermodynamic stability of the guest protein is higher than that of the host protein, the host protein will be likely to fold first and the mutually exclusive protein forms a metastable state. In this scenario, the thermodynamically more stable guest protein will also attempt to fold, leading to a folding tug-of-war. If the folding of the guest protein can generate enough mechanical strain, the folded host protein will be unraveled. Therefore, the key to directly observing the folding tug-of-war is to design a mutually exclusive protein in which the thermodynamically weaker domain folds faster than the thermodynamically more stable domain.

On the basis of this design principle, we endeavored to engineer a mutually exclusive protein in which the host protein is thermodynamically weaker and folds much faster than the guest protein. For this purpose, we used mutants of two well-characterized small proteins, GB1⁸ and I27,¹² to engineer a mutually exclusive protein, GL5/I27w34f, in which GL5 serves as the host domain and I27w34f serves as the guest domain. GL5 is a loop insertion mutant of a small protein, GB1,^{8,13} and its three-dimensional structure obtained by homology modeling is shown in Figure 1. The second loop of GL5 contains 10 residues, and the distance between the two termini of the loop is $\sim 1.2 \text{ nm}$.^{14,15} In contrast, I27w34f is a tryptophan mutant of

I27, in which tryptophan 34 was mutated to phenylalanine. We assumed that the three-dimensional structure of I27w34f is similar to that of wild type I27.¹² The distance between the N- and C-termini of I27w34f is $\sim 4.3 \text{ nm}$, significantly larger than the size of the second loop of GL5. Therefore, the insertion of the folded I27w34f into GL5 will lead to the unraveling of GL5, while the folding of GL5 will prevent the folding of I27w34f, the very nature of mutual exclusiveness.

To determine the thermodynamic stability of isolated GL5 and I27w34f domains, we used far-UV CD spectroscopy to determine their chemical denaturation curves. As shown in Figure 2A, GL5 exhibits a far-UV CD spectrum that is typical of α/β proteins and characterized by strong negative bands at 208 and 221 nm. In contrast, I27w34f shows a typical all- β protein CD spectrum with a negative band at $\sim 215 \text{ nm}$. In addition, there is a weak maximum at $\sim 230 \text{ nm}$, which is characteristic of asymmetric aromatic residues. These features are indistinguishable from those of wild type I27¹⁶ and have been used to monitor the folding/unfolding transitions of I27 and other proteins.^{17–19} Here we used ellipticity at 221 nm to monitor the unfolding/folding transition of GL5 and ellipticity at 230 nm to monitor the unfolding/folding of I27w34f. Figure 2B shows the chemical denaturation curves of both domains. As characterized by the midpoint, $[\text{GdmCl}]_{0.5}$, of the chemical denaturation curve, the thermodynamic stability of isolated I27w34f ($[\text{GdmCl}]_{0.5} = 1.87 \text{ M}$) is higher than that of GL5

(12) Improta, S.; Politou, A. S.; Pastore, A. *Structure* **1996**, *4*, 323–337.
 (13) McCallister, E. L.; Alm, E.; Baker, D. *Nat. Struct. Biol.* **2000**, *7*, 669–673.
 (14) Gronenborn, A. M.; Filipula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657–661.
 (15) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721–4729.

(16) Politou, A. S.; Thomas, D. J.; Pastore, A. *Biophys. J.* **1995**, *69*, 2601–2610.
 (17) Vuilleumier, S.; Sancho, J.; Loewenthal, R.; Fersht, A. R. *Biochemistry* **1993**, *32*, 10303–10313.
 (18) Gur, E.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16113–16118.
 (19) Clarke, J.; Hamill, S. J.; Johnson, C. M. *J. Mol. Biol.* **1997**, *270*, 771–778.

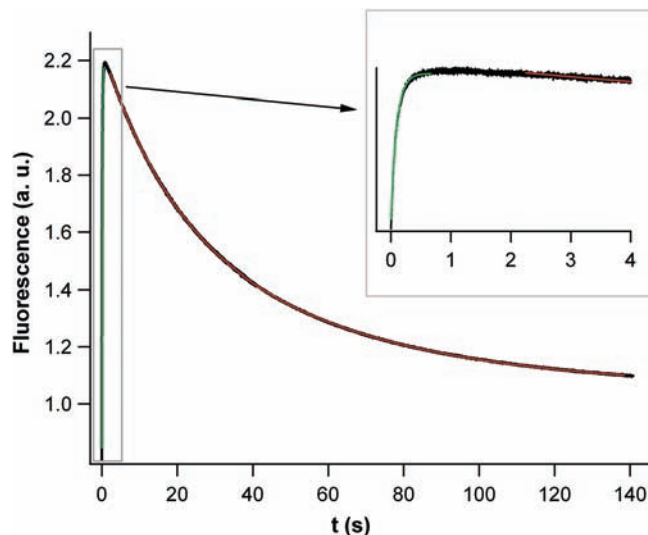


Figure 3. Typical tryptophan fluorescence trace for the folding of GL5/I27w34f in 0.3 M GdmCl. The tryptophan fluorescence of the mutually exclusive protein exhibits a fast rising process, followed by a much slower decay phase. The inset is a close-up view of the fast increasing phase of the tryptophan fluorescence on a shorter time scale. The green line is a single-exponential fit to the folding phase of GL5 with a rate constant of 10 s^{-1} in 0.3 M GdmCl. The red line corresponds to the double-exponential fit to the unfolding phase of GL5 with rate constants of 0.056 s^{-1} (fast phase) and 0.019 s^{-1} (slow phase) in 0.3 M GdmCl.

($[\text{GdmCl}]_{0.5} = 1.34 \text{ M}$) (Figure 2B). The insertion of the guest protein I27w34f into GL5 will be likely to further destabilize the host protein GL5 due to the increased loop size. In addition, GL5 folds much faster than I27w34f ($\sim 67 \text{ s}^{-1}$ for GL5 versus $\sim 0.039 \text{ s}^{-1}$ for I27w34f in 0.3 M GdmCl) (Figure 2C,D). Thus, the utilization of I27w34f as a guest protein and GL5 as a host protein conforms to the ideal specifications necessary for observing tug-of-war folding behavior occurring within mutually exclusive proteins.

The Two Domains in GL5/I27w34f Are Engaged in a Tug-of-War during Folding. To investigate the folding process of the mutually exclusive protein GL5/I27w34f, we carried out stopped-flow spectrofluorimetry studies at different concentrations of chemical denaturant GdmCl, and the tryptophan fluorescence was used as a probe. Since I27w34f does not contain any tryptophan residue, the tryptophan fluorescence measured from GL5/I27w34f solely reflects the folding and unfolding dynamics of the host protein GL5. Figure 3 shows the time evolution of the tryptophan fluorescence of GL5/I27w34f during the folding reaction in 0.3 M GdmCl. The tryptophan fluorescence of GL5/I27w34f showed a characteristic rise and then decay pattern: in 0.5 s, the tryptophan fluorescence of GL5/I27w34f increased rapidly and exponentially to reach a plateau, followed by a slow exponential decay process. It is of note that the exponential decay does not lead to the level of the fluorescence intensity at time zero. The rising phase can be well-described by a single exponential with a rate constant of 10 s^{-1} in 0.3 M GdmCl. However, the decay phase is best described by a double-exponential fit with rate constants of 0.056 and 0.019 s^{-1} in 0.3 M GdmCl, respectively (see Figure 1S in the Supporting Information for a comparison between single- and double-exponential fits). Since the fluorescence signal measured from GL5/I27w34f only reports the folding/unfolding of the host protein GL5, the rising phase thus *presumably* corresponds to the folding of GL5, and the relaxation phase *presumably* corresponds to the unfolding process of GL5. It is of note that

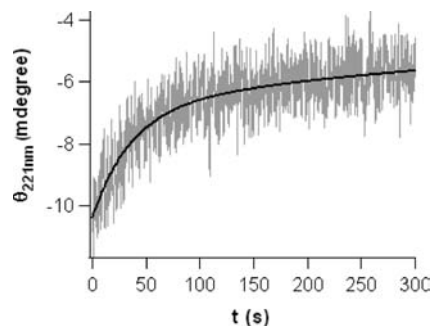


Figure 4. Kinetic information on secondary structural changes of the host domain GL5 during the folding of GL5/I27w34f in 0.3 M GdmCl. Ellipticity at 221 nm was used as the probe in the CD measurements. The black line is a double-exponential fit to the data with rate constants of 0.028 s^{-1} (fast phase) and 0.0022 s^{-1} (slow phase).

the fluorescence data are also consistent with the scenario in which the GL5 domain quickly adopts one conformation and then slowly changes to another conformation that is not the unfolded state. Additional experimental evidence is thus needed to discern such possibilities.

For this purpose, we also monitored the change of secondary structures of the host GL5 domain during the folding of the mutually exclusive protein GL5/I27w34f on a longer time scale. We used changes in the ellipticity at 221 nm to monitor the folding/unfolding transition of the GL5 domain on a longer time scale when the folding reaction is manually initiated. As shown in Figure 4, upon initiating the folding of GL5/I27w34f from its stock solution in 4 M GdmCl, GL5 clearly underwent an unfolding transition on the time scale of 10–150 s, implying that the folding of GL5 was a fast reaction. The unfolding rate constant of GL5 measured in CD experiments is comparable to that measured in stopped-flow experiments via tryptophan fluorescence (0.028 versus 0.056 s^{-1}). This result, in conjunction with stopped-flow data, strongly indicates that the conformational change of the host domain GL5 can only be explained by processes involving both secondary structure change and the exposure of a hydrophobic core, which are signatures of the folding–unfolding transition of GL5. These results demonstrate that, upon initiation of the folding reaction of GL5/I27w34f, the host protein GL5 follows a complex folding and unfolding process: the host protein GL5 folds rapidly from its unfolded state to acquire three-dimensional structure and then unfolds slowly to lose its ordered structure under conditions that strongly favor folding. What are the driving forces for such conformational changes?

Since the condition under which the protein is present strongly favors folding, the seemingly “spontaneous” unfolding of GL5 must be driven by other thermodynamically or kinetically favorable processes. The guest domain I27w34f, which is inserted in the middle of the sequence of GL5, should also attempt to fold under conditions that favor folding. It is important to note that the unfolding rate constant of the host domain GL5 measured in Figure 3 is similar to the folding rate constant of isolated I27w34f under similar conditions (0.056 versus 0.039 s^{-1} in 0.3 M GdmCl) (Figure 2D), suggesting that the unfolding of the host domain GL5 may be driven by the folding of the guest domain I27w34f. The nature of this driving force is the mechanical strain generated by the folding of I27w34f, which involves “extending” the two termini of I27w34f to a distance of $\sim 4.3 \text{ nm}$ and subsequently imposes mechanical strain on GL5. This result provides direct experimental evidence

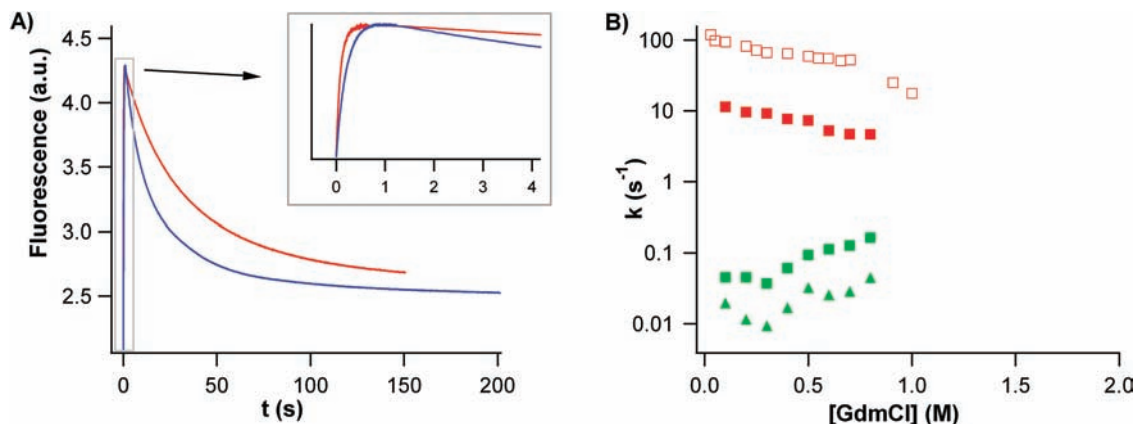


Figure 5. Chemical denaturants slow the folding of the host GL5 but accelerate the folding of the guest protein I27w34f in the mutually exclusive protein GL5/I27w34f. (A) Folding traces of the mutually exclusive protein GL5/I27w34f in the presence of 0.1 M (red) and 0.7 M GdmCl (blue). (B) Semilogarithmic plot of k versus GdmCl concentration. Solid symbols correspond to the data measured from the mutually exclusive protein GL5/I27w34f (red solid squares, folding rate for the host protein GL5; green squares, folding rate constant in the fast phase for the guest protein I27w34f; and green triangles, folding rate constant in the slow phase for the guest protein I27w34f). For comparison, the semilogarithmic plot of k versus [GdmCl] for isolated GL5 domains is also shown (red open squares). It is evident that the folding of GL5 in GL5/I27w34f is slowed by the insertion of the guest protein I27w34f.

that the folding of the guest protein I27w34f can generate enough mechanical strain to trigger the unfolding of the folded host protein GL5.

It is important to note that the short end-to-end distance of the loop in the rapidly folded host protein GL5 imposes topological constraints that prevent the folding of the guest domain I27w34f. Consequently, the guest I27w34f will have to fight against such constraints in order to fold. In return, the folding of I27w34f will impose mechanical strains that attempt to mechanically unravel the host domain. Therefore, our observation that the host protein folds and then unfolds provides the first direct kinetic evidence of the tug-of-war between the competitive folding of the two domains in a mutually exclusive protein GL5/I27w34f. Since I27w34f is thermodynamically more stable than the host GL5, the guest domain I27w34f will tend to win the tug-of-war and fold into a β -sandwich structure with its N- and C-termini pointing in opposing directions. As a consequence, most GL5 domains will “lose” the folding tug-of-war and thus unfold. This conclusion is also supported by the CD result on GL5/I27w34f in PBS (Figure 2S, Supporting Information).

It is of note that the tryptophan fluorescence did not reach the initial intensity when the folding reaction was initiated, indicating that a small portion of GL5/I27w34f (about ~25% estimated from the residual Trp fluorescence) existed with GL5 remaining folded and I27w34f remaining unfolded after the tug-of-war reached equilibrium. Therefore, a conformational equilibrium exists between the two folded forms of the mutually exclusive protein GL5/I27w34f.²⁰

The Folding and Unfolding Dynamics of the Host and Guest Proteins Are Intricately Coupled. By varying the GdmCl concentration, we measured the chemical denaturant dependence of the folding kinetics of the host GL5 and the guest I27w34f domains in GL5/I27w34f (Figure 5A). Since the folding of I27w34f is tightly coupled to the unfolding of GL5 in the mutually exclusive protein, the folding kinetics of I27w34f can be inferred from the unfolding kinetics of GL5 during the fluorescence decay phase (Figure 5). As expected, the logarithm

of the folding rate constant of the host protein GL5 decreases linearly as a function of the GdmCl concentration (Figure 5B, red solid squares). Surprisingly, the logarithm of the folding rate constant of the guest I27w34f increases linearly as a function of the GdmCl concentration (Figure 5B, green symbols). Such dependence is the complete opposite of that observed for most proteins during chemical folding processes. Such dependence is not surprising if we consider this process as the unfolding process of the host protein GL5, despite the fact that the conditions strongly favor folding. A possible explanation in this seemingly contradictory result lies in the fact that the mechanical resistance/stability of proteins will decrease with increasing GdmCl concentration, as demonstrated previously by using single-molecule AFM.²¹ Hence, increasing GdmCl concentration will facilitate the unraveling of the host protein GL5, which is driven by the folding of the guest protein I27w34f. This effect is equivalent to facilitating the folding of the guest domain I27w34f.

Compared with the folding of isolated GL5 (Figure 5B, red open squares), the folding of the host GL5 domain in the mutually exclusive protein GL5/I27w34f is slowed by an order of magnitude. This effect is a direct reflection of the influence of configurational entropy on the folding of GL5. Since the folding of I27w34f is much slower than that of GL5, I27w34f remains largely unstructured during the folding of the host protein GL5. The net effect is equivalent to increasing the size of the second loop of GL5 by 89 residues. It is well known that the configurational entropy penalty associated with increasing the length of an unstructured loop may slow protein folding.^{22–26} Therefore, the impeding effect on the folding of the host protein GL5 in the mutually exclusive protein GL5/I27w34f can be readily explained by the increased cost in configurational entropy. This result highlights the intricate coupling between host and guest domains in a mutually

(21) Cao, Y.; Li, H. *J. Mol. Biol.* **2008**, *375*, 316–324.

(22) Viguera, A. R.; Serrano, L. *Nat. Struct. Biol.* **1997**, *4*, 939–946.

(23) Grantcharova, V. P.; Riddle, D. S.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7084–7089.

(24) Scalley-Kim, M.; Minard, P.; Baker, D. *Protein Sci.* **2003**, *12*, 197–206.

(25) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1525–1529.

(26) Nagi, A. D.; Anderson, K. S.; Regan, L. *J. Mol. Biol.* **1999**, *286*, 257–265.

(20) Although unlikely, the reason that the fluorescence does not return to the initial value could be that the GL5 domain (in the vicinity of the tryptophan) remains partially folded. Our current data cannot completely rule out this possibility.

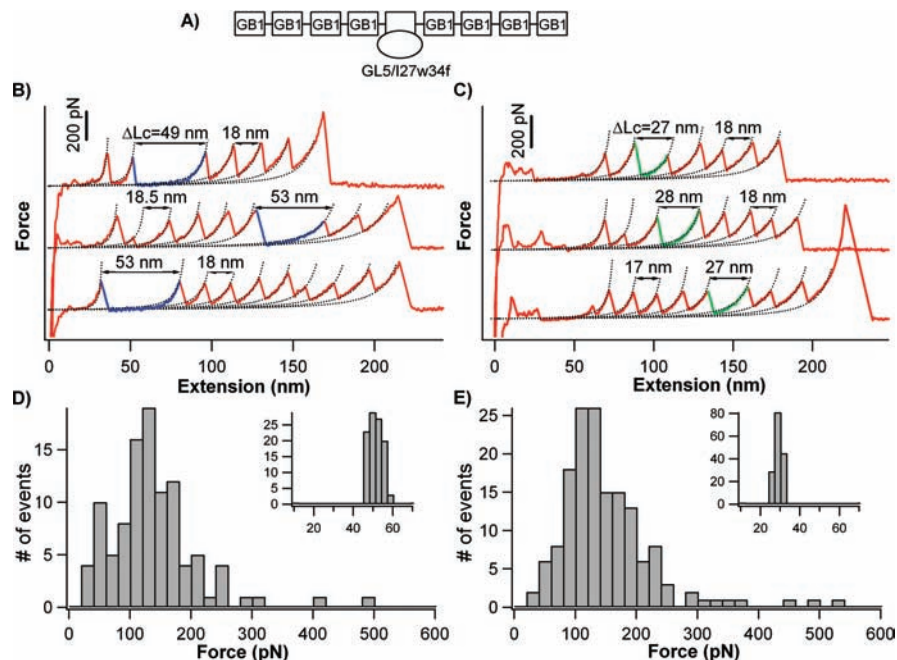


Figure 6. AFM experiments provide direct evidence at the single-molecule level for the coexistence of two populations: GL5_{folded}/I27w34f_{unfolded} and GL5_{unfolded}/I27w34f_{folded} for the mutually exclusive protein GL5/I27w34f. (A) Schematic of the polyprotein chimera (GB1)₄-GL5/I27w34f-(GB1)₄ used in single-molecule AFM studies. (B) Typical force–extension curves show mechanical signatures for the unfolding of GL5_{folded}/I27w34f_{unfolded} (colored in blue). The unfolding of GL5_{folded}/I27w34f_{unfolded} results in unfolding events with a ΔL_c of ~ 52 nm, as determined by Worm-Like Chain model (WLC) fits (thin lines) to the experimental data. (C) Typical force–extension curves show mechanical signatures of the unfolding of GL5_{unfolded}/I27w34f_{folded} (colored in green). (D) Unfolding force histogram of GL5_{folded}/I27w34f_{unfolded}. The average unfolding force is 128 ± 55 pN ($n = 102$). The inset is the histogram of ΔL_c resulted from the unfolding of GL5_{folded}/I27w34f_{unfolded}. The average ΔL_c is 50.8 ± 3.3 nm ($n = 102$). (E) Histogram of unfolding force of GL5_{unfolded}/I27w34f_{folded}. The average unfolding force is 142 ± 58 pN ($n = 155$). The inset is the histogram of ΔL_c of GL5_{unfolded}/I27w34f_{folded}. The average ΔL_c is 28.3 ± 1.8 nm ($n = 155$). The unfolding forces were measured at a pulling speed of 400 nm/s.

exclusive protein. Therefore, the host and guest domains can no longer be treated or assumed as isolated proteins in understanding the folding–unfolding dynamics of domain insertion proteins, including mutually exclusive proteins.

Single-Molecule AFM Experiments Provide Direct Evidence for the Existence of Distinct Forms of GL5/I27w34f. As indicated above, the tryptophan fluorescence of GL5/I27w34f did not relax back to the initial value when the folding reaction was initiated. This result suggests that, in a fraction of the mutually exclusive protein GL5/I27w34f ($\sim 25\%$ in the presence of 0.3 M GdmCl), the host protein GL5 remained folded and the guest protein I27w34f remained unfolded. Since the two populations of GL5/I27w34f are expected to exhibit distinctive mechanical signatures during mechanical unfolding, we carried out single-molecule AFM studies to directly confirm the existence of the two populations of GL5/I27w34f and test the mutual exclusiveness of the folding of two domains at the single-molecule level. In the form of GL5_{folded}/I27w34f_{unfolded}, the fully unfolded and extended GL5/I27w34f is ~ 54.7 nm ($(63 + 89)\text{aa} \times 0.36 \text{ nm/aa}$), and the distance between the N- and C-termini should be the same as that of wild type GL5 (~ 2.4 nm). Thus, the unfolding of GL5_{folded}/I27w34f_{unfolded} should lead to unfolding events with a contour length increment ΔL_c of ~ 52 nm ($54.7 \text{ nm} - 2.4 \text{ nm} = 52.3 \text{ nm}$). In contrast, the unfolding of GL5_{unfolded}/I27w34f_{folded} should lead to unfolding events that are characteristic of I27w34f, which displays a ΔL_c of ~ 28 nm (the same as that of wild type I27⁹), since GL5 is unfolded and will not result in an unfolding force peak. To facilitate the identification of GL5/I27w34f unfolding events, we engineered a heteropolyprotein (GB1)₄-GL5/I27w34f-(GB1)₄ (Figure 6A), in

which well-characterized GB1 domains serve as a fingerprint^{11,27,28} for discerning the mechanical unfolding event of GL5/I27w34f.

Stretching (GB1)₄-GL5/I27w34f-(GB1)₄ results in force–extension curves of the characteristic sawtooth pattern appearance, in which individual sawtooth peaks correspond to the mechanical unfolding of individual domains in the polyprotein (Figure 6B,C). Since GL5/I27w34f is sandwiched between two (GB1)₄, we can ensure that the force–extension curve contains the mechanical unfolding signature of the GL5/I27w34f protein if five or more GB1 unfolding events are observed in a given force–extension curve. The unfolding events colored in red can be easily identified as the unfolding of GB1 domains, as they are characterized by a ΔL_c of ~ 18 nm and an unfolding force of ~ 180 pN at a pulling speed of 400 nm/s.^{11,27,28} In addition to GB1 unfolding events, we also observed unfolding events of a ΔL_c of ~ 52 nm (Figure 6B, blue) and a ΔL_c of ~ 28 nm (Figure 6C, green). The unfolding events of a ΔL_c of ~ 52 nm are consistent with the unfolding of GL5_{folded}/I27w34f_{unfolded}. In contrast, a ΔL_c of ~ 28 nm is the mechanical unfolding signature of I27w34f domains, suggesting that unfolding events of a ΔL_c of ~ 28 nm in Figure 6C are due to the unfolding of GL5_{unfolded}/I27w34f_{folded}. Therefore, our single-molecule AFM results provide direct evidence for the coexistence of the two forms of GL5/I27w34f under the experimental conditions.

Unfolding force histograms and ΔL_c histograms for the unfolding of GL5_{folded}/I27w34f_{unfolded} and GL5_{unfolded}/I27w34f_{folded} are shown in Figure 6D and 6E, respectively. The average unfolding force is 128 ± 55 pN ($n = 102$) for the host domain

(27) Peng, Q.; Li, H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1885–1890.
(28) Sharma, D.; Perisic, O.; Peng, Q.; Cao, Y.; Lam, C.; Lu, H.; Li, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9278–9283.

GL5 in $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and 142 ± 58 pN ($n = 155$) for the guest domain $I27w34f$ in $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$, respectively. It is interesting to note that the unfolding force for the host GL5 domain in $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ is similar to that of isolated GL5,⁸ despite the fact that $I27w34f$ is largely unfolded in the population of $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$. This result suggests that the mechanical unfolding process of $GL5/I27w34f$ does not affect the loop closure and its associated configurational entropy.

Since single-molecule AFM experiments are initiated from a sample that is presumably at equilibrium, the frequencies of observing $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$ can thus provide valuable information about conformational equilibrium between these two distinctive populations. In our experiments, the relative frequency of observing $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$ is 102:155. In fluorescence decay experiments, we measured the ratio between the two populations $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$ in the presence of different concentrations of GdmCl. Although the ratio between two populations measured in AFM experiments cannot be directly compared with those measured in fluorescence decay experiments, we have plotted the ratio measured in AFM together with those measured in fluorescence decay experiments at different concentrations of GdmCl. It is evident that the ratio measured in AFM is in reasonable agreement with that determined from tryptophan fluorescence decay experiments (Figure 3S, Supporting Information).

Can a Folded GL5 and a Folded I27w34f Coexist in a Mutually Exclusive Protein GL5/I27w34f? By design, $GL5/I27w34f$ is a mutually exclusive protein. In theory, the folding of GL5 and $I27w34f$ should be mutually exclusive, and certainly this mutual exclusiveness is true on the ensemble level. Is it possible that a tiny fraction of $GL5/I27w34f$ can exist with both GL5 and $I27w34f$ folded at the same time? Single-molecule AFM provides an ideal method to probe such a question, as the unfolding of such a population will yield characteristic mechanical unfolding signatures. Out of 261 force–extension curves, we observed four curves showing putative evidence of the existence of $GL5_{\text{folded}}/I27w34f_{\text{folded}}$ (Figure 7). In these four force–extension curves, the unfolding of $GL5/I27w34f$ occurred in two steps: the first step is of a ΔL_c of ~ 24 nm, and the second step is of a ΔL_c of ~ 28 nm. The sum of the two ΔL_c is ~ 52 nm, in good agreement with the complete unfolding of $GL5/I27w34f$. Furthermore, the unfolding event with a ΔL_c of 24 nm always occurred prior to the unfolding event with a ΔL_c of 28 nm. Such unfolding signatures are in good agreement with those expected from the unfolding of $GL5_{\text{folded}}/I27w34f_{\text{folded}}$: the first unfolding step corresponds to the unfolding of the folded host protein GL5, which gives rise to a ΔL_c of ~ 24.4 nm (20.1 nm + 4.3 nm, where 20.1 nm is the ΔL_c of isolated GL5⁸ and 4.3 nm is the N–C distance of $I27w34f$), and the second unfolding step corresponds to the unfolding of the folded guest protein $I27w34f$. These results provide putative evidence that, on rare occasions ($\sim 1.5\%$), it is possible that both the host GL5 and guest $I27w34f$ can remain folded simultaneously in one $GL5/I27w34f$ protein. From the force–extension curves, we can estimate that such conformations can remain folded at least on the time scale of ~ 50 ms (which is the time required to unfold the host domain GL5 at a pulling speed of 400 nm/s). Since the folding of either domain will generate strain on the other one, both domains must be subject to significant mechanical strains in such a coexisting form. Thus, such conformations

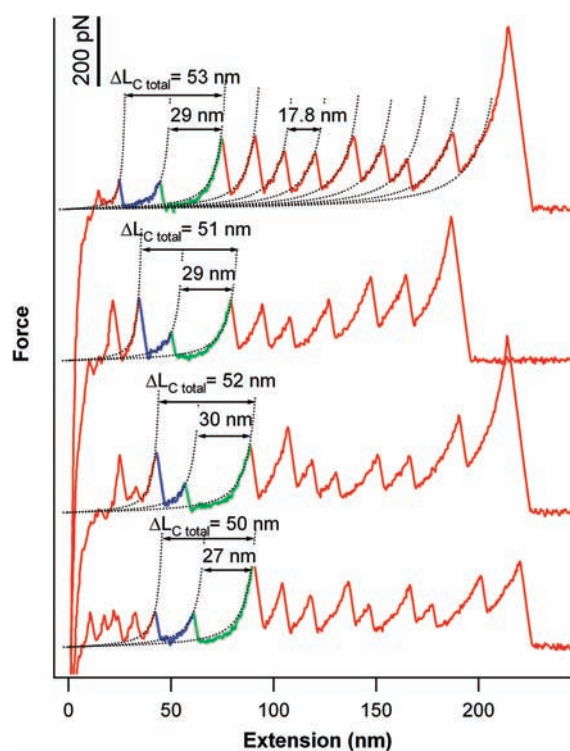


Figure 7. Force–extension curves show putative signatures for the coexistence of both folded GL5 and folded $I27w34f$ domains in $GL5/I27w34f$. Such unfolding events are characterized by the two-step unfolding process: the first unfolding step is of a $\Delta L_c \approx 24$ nm (blue), and the second unfolding step is of a $\Delta L_c \approx 28$ nm (green). The sum of both ΔL_c ($\Delta L_{c\text{ total}}$) is ~ 52 nm. Thin lines are WLC fits to the experimental data.

($GL5_{\text{folded}}/I27w34f_{\text{folded}}$) must be metastable. It will be interesting to investigate whether such a metastable species is an intermediate during the conversion from $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ to $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$, or vice versa.

Thermodynamic, Kinetic, and Mechanical Factors Play Critical Roles in the Folding Dynamics of Mutually Exclusive Proteins. Combining protein engineering, stopped-flow spectrofluorimetry, and single-molecule AFM techniques, we have obtained the first direct kinetic evidence for the tug-of-war that occurs during the folding of two domains in a mutually exclusive protein. Our results show that, during the folding tug-of-war between the two domains, folding kinetics and thermodynamics play critical roles in determining which domain attains a folded structure in a mutually exclusive protein.

In the design of the mutually exclusive protein $GL5/I27w34f$, the host GL5 domain folds faster than $I27w34f$, despite the fact that the insertion of $I27w34f$ slows the intrinsic folding of GL5. Thus, GL5 wins the first round of the folding tug-of-war between GL5 and $I27w34f$. However, since $I27w34f$ is thermodynamically more stable than the host GL5, whose sequence continuity is disrupted by the insertion of the guest domain, $I27w34f$ managed to fold despite topological constraints imposed by the folded GL5. The folding of $I27w34f$ generated enough mechanical strain to unravel the folded host domain GL5. Thus, $I27w34f$ won the tug-of-war of folding in most cases. The intricate coupling that occurs between the kinetics and thermodynamics of the two domains determines the overall folding dynamics and equilibrium conformation of the mutually exclusive protein. For the two dominant populations ($GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$), it remains unclear whether they are in a dynamic equilibrium and interconvertible.

Single-molecule AFM refolding experiments, in particular force-clamp spectroscopy experiments,²⁹ will be ideal to probe such questions. However, instrument drift and relatively slow folding kinetics of I27w34f (on the time scale of a few minutes) make such experiments technically very challenging. Future endeavors will be required to investigate this interesting question in detail.

The mechanical strain generated by the folding of the guest domain I27w34f is another important factor that affects this tug-of-war. The guest domain has to be powerful enough to generate sufficient mechanical strain to trigger the unfolding of the folded host protein. If the mechanical stability of the folded host protein is too high or the mechanical strain generated by the guest protein is too small, the host protein can be “mechanically trapped” in the folded state, despite its lower thermodynamic stability. Therefore, the outcome of the tug-of-war of folding for the two domains in a mutually exclusive protein can also be determined by mechanical factors, such as the competition between the mechanical stability of the host protein and the mechanical force generated by the folding of the guest protein. Therefore, rationally designed mutually exclusive proteins will provide an ideal platform for investigating the importance of these thermodynamic, kinetic, and mechanical determinants during protein folding and unfolding processes.

It is important to point out that, during the folding of mutually exclusive proteins, the two domains do not necessarily have to be effectively engaged in a tug-of-war. If the thermodynamically more stable domain is also the faster folder, the thermodynamically weaker and slower folder will not have any chance to fold during the folding process. In such cases, the two domains effectively are not engaged in a tug-of-war, as the winner will

dominate the tug-of-war from the very beginning. Such scenarios highlight the importance of thermodynamic and kinetic properties of the two domains to the folding of mutually exclusive proteins.

Furthermore, our results also provide direct evidence that the folding of a protein can generate enough mechanical strain to unravel a second protein. The ability of a protein to generate mechanical strain and do mechanical work is a property similar to those of molecular motors, an exciting connection between protein folding and nanotechnology. Such connections raise questions concerning how much force a protein is capable of generating and what factors determine the possible “power output” for different proteins. Single-molecule AFM techniques could prove to be well adapted to answer such questions.

Acknowledgment. We thank Ashlee Jollymore for critical reading of the manuscript. This work is supported by the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs program, and Canada Foundation for Innovation. H.L. is a Michael Smith Foundation for Health Research Career Investigator. Q.P. is supported by a Pacific Century Graduate Scholarship from the Province of British Columbia.

Supporting Information Available: Single- and double-exponential fits to the fluorescence decay phase of GL5/I27w34f in 0.3 M GdmCl; far-UV CD spectra of isolated GL5, isolated I27w34f, and combined GL5/I27w34f in PBS buffer at pH 7.4; and plot of the dependence of the conformational equilibrium between $GL5_{\text{folded}}/I27w34f_{\text{unfolded}}$ and $GL5_{\text{unfolded}}/I27w34f_{\text{folded}}$ on the concentration of GdmCl. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA903480J

(29) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674–1678.