J|A|C|S ARTICLES

Published on Web 09/11/2009

Domain Insertion Effectively Regulates the Mechanical Unfolding Hierarchy of Elastomeric Proteins: Toward Engineering Multifunctional Elastomeric Proteins

Qing Peng and Hongbin Li*

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

Received May 3, 2009; E-mail: Hongbin@chem.ubc.ca

Abstract: The architecture of elastomeric proteins controls fine-tuned nanomechanical properties of this class of proteins. Most elastomeric proteins are tandem modular in structure, consisting of many individually folded domains of varying stability. Upon stretching, these elements unfold sequentially following a strict hierarchical pattern determined by their mechanical stability, where the weakest element unfolds first and the strongest unfolds last. Although such a hierarchical architecture is well-suited for biological functions of elastomeric proteins, it may become incompatible with incorporating proteins of desirable functionality in order to construct multifunctional artificial elastomeric proteins, as many of these desired proteins are not evolved for mechanical purpose. Thus, exposure to a high stretching force will result in unraveling of these proteins and lead to a loss of their functionality. To overcome this challenge, we combine protein engineering with single molecule force spectroscopy to demonstrate that domain insertion is an effective strategy to control the mechanical unfolding hierarchy of multidomain proteins and effectively protect mechanically labile domains. As a proof-of-principle experiment, we spliced a mechanically labile T4 lysozyme (T4L) into a flexible loop of a mechanically stronger host domain GL5 to create a domain insertion protein. Using single molecule force spectroscopy, we showed that the mechanically labile T4L domain unfolds only after the mechanically stronger host domain GL5 has unfolded. Such a reverse mechanical unfolding hierarchy effectively protects the mechanically labile T4L domain from applied stretching force and significantly increased the lifetime of T4L. The approach demonstrated here opens the possibility to incorporate labile proteins into elastomeric proteins for engineering novel multifunctional elastomeric proteins.

Introduction

Mechanical proteins within living biological systems self-assemble into complex mechano-chemical machinery in order to sense, generate, and bear mechanical forces,[1] as well as to accomplish a diverse range of biological functions. For example, elastomeric proteins, a special class of mechanical proteins,[2,3] are subject to mechanical tension under physiological conditions and serve as molecular springs in a wide variety of biological machinery to establish elastic connections and provide mechanical strength, elasticity, and extensibility.[2,4−10] As well as performing important biological functions *in vivo*, elastomeric proteins also play indispensable roles by serving as functional building blocks for constructing nanomechanical devices with well-defined applications.[11−13] For example, molecular motors have been incorporated into nanomechanical devices to provide locomotive forces.[12] It can be envisioned that elastomeric proteins can be incorporated into nanomechanical systems to serve as molecular springs, latches, switches, and sensors.[13−15]

Recent developments in single molecule atomic force microscopy (AFM) have enabled the characterization of the mechanical properties of elastomeric proteins, both natural and artificial, in great detail.[4−6,16,17] For example, the mechanical unfolding of green fluorescent protein (GFP) was investigated

(1) Bao, G.; Suresh, S. *Nat. Mater.* **2003**, *2*, 715.
(2) Tatham, A. S.; Shewry, P. R. *Trends Biochem. Sci.* **2000**, *25*, 567.
(3) Gosline, J.; Lillie, M.; Carrington, E.; Guerette, P.; Ortlepp, C.; Savage, K. *Philos. Trans. R. Soc. Lon. B* **2002**, *357*, 121.
(4) Li, H.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkvliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M. *Nature* **2002**, *418*, 998.
(5) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature* **1998**, *393*, 181.
(6) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109.
(7) Rief, M.; Pascual, J.; Saraste, M.; Gaub, H. E. *J. Mol. Biol.* **1999**, *286*, 553.
(8) Ohashi, T.; Kiehart, D. P.; Erickson, H. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2153.
(9) Miller, M. K.; Granzier, H.; Ehler, E.; Gregorio, C. C. *Trends Cell Biol.* **2004**, *14*, 119.
(10) Carl, P.; Kwok, C. H.; Manderson, G.; Speicher, D. W.; Discher, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1565.
(11) Hess, H.; Clemmens, J.; Brunner, C.; Doot, R.; Luna, S.; Ernst, K. H.; Vogel, V. *Nano Lett.* **2005**, *5*, 629.
(12) Soong, R. K.; Bachand, G. D.; Neves, H. P.; Olkhovets, A. G.; Craighead, H. G.; Montemagno, C. D. *Science* **2000**, *290*, 1555.
(13) Goodsell, D. S. *Bionanotechnology*; Wiley-Liss, Inc: Wilmington, DE, 2004.
(14) Ball, P. *Nature* **2001**, *409*, 413.
(15) Zhang, S. *Nat. Biotechnol.* **2003**, *21*, 1171.
(16) Carrion-Vazquez, M.; Oberhauser, A.; Diez, H.; Hervas, R.; Oroz, J.; Fernandez, J.; Martinez-Martin, A. Protein nanomechanics as studied by AFM single-molecule force spectroscopy. In *Advanced Techniques in Biophysics*; Arrondo, J., Alonso, A., Eds.; Springer: New York, 2006; p 163.
(17) Li, H. *Adv. Funct. Mater.* **2008**, *18*, 2643.

10.1021/ja903589t CCC: $40.75

in the hope that GFP could be used as a force sensor.[18,19] Although some nonmechanical proteins can exhibit significant mechanical stability and thus serve as artificial elastomeric proteins,[20−22] the vast majority of nonmechanical proteins are mechanically labile and cannot function under a stretching force,[16,23−31] making it challenging to employ specific proteins, such as enzymes, in the construction of multifunctional nanomechanical systems.

Most naturally occurring elastomeric proteins are tandem modular proteins that consist of a series of individually folded mechanical elements (domains) of varying stability.[5,6,32] Upon stretching, individual domains in this linear arrangement are subject to the same mechanical stretching force simultaneously and will unfold sequentially according to their mechanical stability; for example, the unfolding of each individual domain follows a strict hierarchical pattern from low to high mechanical stability where the weakest one unfolds first and the strongest one unfolds last.[5,6,32,33] However, such a hierarchical architecture of elastomeric proteins significantly limits the incorporation of proteins of desirable functionality into artificial elastomeric proteins, as many of these desired proteins are mechanically labile and will unravel to lose their functionality when they are subject to a high stretching force.

To overcome this challenge, novel engineering of elastomeric proteins will be required. As a proof-of-principle, here, we combine protein engineering with single molecule AFM to demonstrate that domain insertion[34,35] is an effective method to regulate the mechanical unfolding hierarchy of proteins and effectively protect mechanically labile proteins. By inserting the mechanically weaker T4 lysozyme domain (T4L)[24,25,36] into the mechanically stronger GB1-L5 (GL5)[37] domain, we demonstrate that the mechanically labile T4L domain unfolds only after the stronger host GL5 domain has unfolded. This reverse unfolding hierarchy effectively protects the mechanically labile T4L domain from applied stretching force and significantly increases the lifetime of T4L under a stretching force. The approach demonstrated here opens the avenue to incorporating labile proteins into elastomeric proteins for engineering novel multifunctional elastomeric proteins.

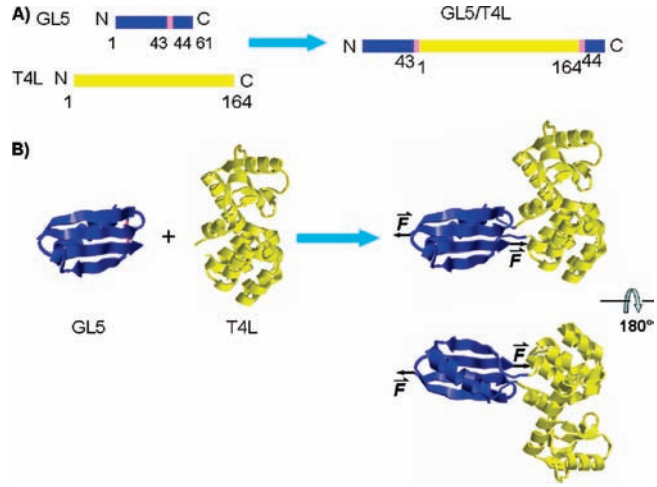

Figure 1. Structure of the designed domain insertion protein GL5/T4L. (A) The construction of GL5/T4L hybrid protein from GL5 and T4L. GL5 contains a nonpalindromic *Ava*I restriction site in its second loop between residues 43 and 44. The domain insertion protein GL5/T4L was constructed by inserting the guest protein T4L into the host protein GL5 between residues 43 and 44 in the second loop. (B) Three-dimensional structures of GL5, T4L (PDB code 1L63), and GL5/T4L. The structure of GL5 was assumed to be the same as that of wild-type GB1 (PDB code 1PGA), and the structure of GL5/T4L was generated by combining the three-dimensional structures of GL5 and T4L.

Materials and Methods

Protein Engineering. The plasmid encoding the pseudowild-type T4 lysozyme protein (T4L) was generously provided by Prof. Brian W. Matthews of the University of Oregon. The DNA sequence for T4L, flanked with nonpalindromic *Ava*I restriction sites (CTCGGG) at both 5′ and 3′ ends, was amplified by polymerase chain reaction. The gene of GL5 was constructed as described previously[37] and contains a nonpalindromic *Ava*I restriction site, which encodes residues 43 and 44, in the second loop (Figure 1A).

To construct the combined hybrid protein GL5/T4L, T4L was first digested with the *Ava*I restriction enzyme and then ligated into pUC19/GL5, which was linearized by digestion with *Ava*I. GL5/T4L was then subcloned into expression vector pQE80L, and its DNA sequence was confirmed by direct DNA sequencing. The final construct GL5/T4L (Figure 1B), which carries an N-terminal His tag for purification, has the following amino acid sequence:

MRGSHHHHHHGS*MDTYKLILNGKTLKGETTTEAVDAATAEKV-FKQYANDNGVGGGLG*MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHL-LTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVR-GILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRM-LQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTW-DAYKNL*LGDGEWTYDDATKTFTVTE*RS, where the sequence in italic is from the host domain GL5 and the sequence in bold is from the guest domain T4L. The junction between GL5 and T4L is Leu-Gly that is a result of the *Ava*I site. The N-terminal Gly-Ser results from the *Bam*HI site, and the C-terminal Arg-Ser results from the *Bgl*II site. The combined hybrid protein GL5/T4L was overexpressed in the *Escherichia coli* strain DH5α and purified by Ni^{2+} affinity chromatography.

Genes encoding polyproteins $(GL5)_4$-T4L-$(GL5)_4$ and $(GL5)_4$-GL5/T4L-$(GL5)_4$ were constructed in a pQE80L vector using a previously described method based on the identity of the sticky

(18) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16192.
(19) Perez-Jimenez, R.; Garcia-Manyes, S.; Ainavarapu, S. R.; Fernandez, J. M. *J. Biol. Chem.* **2006**, *281*, 40010.
(20) Brockwell, D. J.; Beddard, G. S.; Paci, E.; West, D. K.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophys. J.* **2005**, *89*, 506.
(21) Cao, Y.; Lam, C.; Wang, M.; Li, H. *Angew. Chem., Int. Ed. Engl.* **2006**, *45*, 642.
(22) Cao, Y.; Li, H. *Nat. Mater.* **2007**, *6*, 109.
(23) Best, R. B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. *Biophys. J.* **2001**, *81*, 2344.
(24) Yang, G.; Cecconi, C.; Baase, W. A.; Vetter, I. R.; Breyer, W. A.; Haack, J. A.; Matthews, B. W.; Dahlquist, F. W.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 139.
(25) Peng, Q.; Li, H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1885.
(26) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. *Science* **2005**, *309*, 2057.
(27) Ainavarapu, S. R.; Li, L.; Badilla, C. L.; Fernandez, J. M. *Biophys. J.* **2005**, *89*, 3337.
(28) Junker, J. P.; Hell, K.; Schlierf, M.; Neupert, W.; Rief, M. *Biophys. J.* **2005**, *89*, L46.
(29) Wilcox, A. J.; Choy, J.; Bustamante, C.; Matouschek, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15435.
(30) Junker, J. P.; Ziegler, F.; Rief, M. *Science* **2009**, *323*, 633.
(31) Sulkowska, J. I.; Cieplak, M. *J. Phys.: Condens. Matter* **2007**, *19*, 283201.
(32) Oberhauser, A. F.; Badilla-Fernandez, C.; Carrion-Vazquez, M.; Fernandez, J. M. *J. Mol. Biol.* **2002**, *319*, 433.
(33) Li, H.; Oberhauser, A. F.; Fowler, S. B.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6527.
(34) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11241.
(35) Ostermeier, M. *Protein Eng. Des. Sel.* **2005**, *18*, 359.
(36) Matthews, B. W. *FASEB J.* **1996**, *10*, 35.
(37) Li, H.; Wang, H. C.; Cao, Y.; Sharma, D.; Wang, M. *J. Mol. Biol.* **2008**, *379*, 871.

ends generated by *BamH*I and *Bgl*II restriction enzymes.[38] The resultant polyproteins were overexpressed in the DH5α *E. coli* strain and purified from the supernatant using Ni^{2+} affinity chromatography. The polyproteins were kept at 4 °C in PBS buffer at a concentration of ~500 μg/mL.

The molecular weights of the combined hybrid protein GL5/T4L, polyprotein $(GL5)_4$-T4L-$(GL5)_4$, and $(GL5)_4$-GL5/T4L-$(GL5)_4$ were characterized using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The theoretical molecular weight for GL5/T4L, $(GL5)_4$-T4L-$(GL5)_4$, and $(GL5)_4$-GL5/T4L-$(GL5)_4$ is 27k, 75k, and 82k, respectively. As shown in Figure S1 of the Supporting Information, apparent molecular weights for the three proteins are in good agreement with the theoretical ones. From the intensity of protein bands on the PAGE gel, we estimated that the purity of the two polyprotein chimera was >95% and the purity of the combined hybrid protein GL5/T4L was >98%. Both host GL5 and guest T4L domains are folded in the combined hybrid protein GL5/T4L, as probed by far-ultraviolet circular dichroism spectroscopy (far-UV-CD) (Figure S2 of the Supporting Information).

Single Molecule AFM. Single molecule AFM experiments were carried out on a custom built AFM, which was constructed as described.[39] All the force−extension and force−clamp measurements were carried out in PBS buffer. In a typical experiment, the polyprotein sample (1 μL) was deposited onto a clean glass coverslip covered by PBS buffer (50 μL) and allowed to adsorb for approximately 5 min before performing the force−extension experiment. The spring constant of each individual cantilever (Si_3N_4 cantilevers from Vecco, with a typical spring constant of 40 pN/nm) was calibrated in solution using the equipartition theorem before and after each experiment.

In single molecule AFM experiments, it is essential to ensure that force−extension curves are indeed resulted from the unfolding of a single molecule. The use of GL5 in the polyprotein chimera provides a fingerprint that allows us to identify single molecule stretching events without any ambiguity. To ensure that force−extension curves contain the signature of unfolding of the combined hybrid protein GL5/T4L, we only analyzed force−extension curves containing at least five unfolding events of GL5 domains. The unfolding force was taken as the peak value of a given unfolding event. The last force peak in force−extension curves is usually resulted from the detachment of the polyprotein chain from the glass substrate or the cantilever tip and thus not included in the unfolding force analysis. The worm-like-chain model (WLC)[40] of polymer elasticity was used to fit force−extension curves in order to measure the contour length increment (ΔLc) of individual unfolding events, as well as the persistence length of the polyprotein chain. All these data analysis were done using routines custom written in Igor Pro 5 (Wavemetrics, Lake Oswego).

Chemical Unfolding Kinetics. Unfolding kinetics experiments were carried out on a Bio-Logic SFM300 instrument equipped with an MOF-200 spectrometer, and tryptophan fluorescence was used as the probe. Proteins were excited at 297 nm, and a 320 nm lowpass filter was used to collect the fluorescence data. The dead time of the stopped-flow instrument for the unfolding experiment at 4 M guanidine chloride (GdmCl) is ~4.3 ms. For unfolding experiments, a stock solution of ~10 μM native protein in PBS was rapidly mixed with different volumes of 7 M GdmCl stock solution to trigger the unfolding reaction. Kinetic traces were recorded over intervals of 20 μs and 1 ms for GL5 and T4L, respectively. The recorded fluorescence−time traces were then analyzed using the software supplied by Bio-Logic.

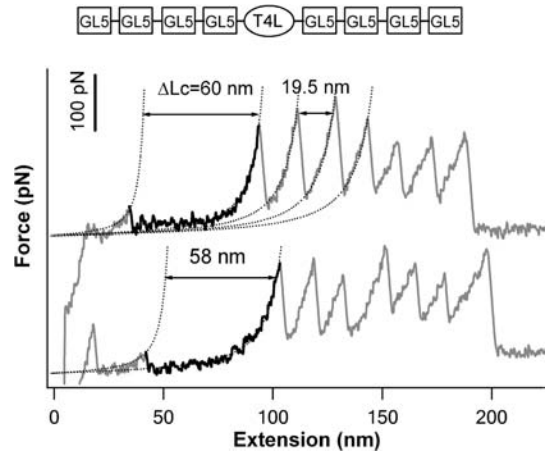

Figure 2. Force−extension curves of $(GL5)_4$-T4L-$(GL5)_4$ reveal the mechanical unfolding hierarchy between T4L and GL5 domains when they are placed in tandem. Stretching the polyprotein $(GL5)_4$-T4L-$(GL5)_4$ results in force−extension curves with a characteristic sawtooth pattern appearance. The mechanical unfolding of the well-characterized GL5 domains (in gray) occurred at ~130 pN with a contour length increment ΔLc of ~20 nm, while the mechanical unfolding of T4L (in black) is characterized by unfolding forces of ~50 pN and a ΔLc of ~60 nm. The mechanical unfolding of T4L (in black) always proceeds the mechanical unfolding of GL5 domains (in gray). Dotted lines correspond to WLC fits to the experimental data. Top panel shows the schematic of the polyprotein $(GL5)_4$-T4L-$(GL5)_4$.

Results

Protein Design. To demonstrate the feasibility of controlling the mechanical unfolding hierarchy of proteins by domain insertion, we used well-characterized T4L (Figure 1, in yellow) and GL5 (Figure 1, in blue) as model systems. The mechanical unfolding of T4L and GL5 has been previously studied in detail using single molecule AFM.[24,25,37] The mechanical unfolding of T4L is characterized by a contour length increment ΔLc of ~60 nm and an unfolding force of ~50 pN at a pulling speed of 400 nm/s.[25] In contrast, GL5 is a loop insertion mutant of the mechanically stable protein GB1, in which five residues (GGGLG) were inserted into the second loop of GB1. The folded structure of GL5 was not significantly different when compared with that of GB1.[37] The mechanical unfolding of GL5 is characterized by a ΔLc of ~20 nm and an unfolding force of ~130 pN at a pulling speed of 400 nm/s.[37] The sharp contrast in the mechanical properties of T4L and GL5 domains offers a convenient mechanical fingerprint that allows us to identify the mechanical unfolding of both domains in the constructed protein chimera without any ambiguity and allows for the determination of a mechanical unfolding hierarchy between the two domains.

Since T4L is mechanically weaker than GL5, T4L always unfolds prior to the mechanical unfolding of GL5 domains when T4L and GL5 are arranged in tandem in the modular protein $(GL5)_4$-T4L-$(GL5)_4$, as demonstrated by the fact that the unfolding event of T4L always precedes those of GL5 domains in a given force−extension curve of $(GL5)_4$-T4L-$(GL5)_4$ (Figure 2). This result clearly reveals the mechanical unfolding hierarchy between GL5 and T4L domains. To reverse the mechanical unfolding hierarchy, we employed the domain insertion approach[34,35] to design a novel combined hybrid protein GL5/T4L, in which the mechanically stable domain GL5 is used as the host protein while the mechanically weaker domain T4L is

(38) 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.
(39) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674.
(40) Marko, J. F.; Siggia, E. D. *Macromolecules* **1995**, *28*, 8759.

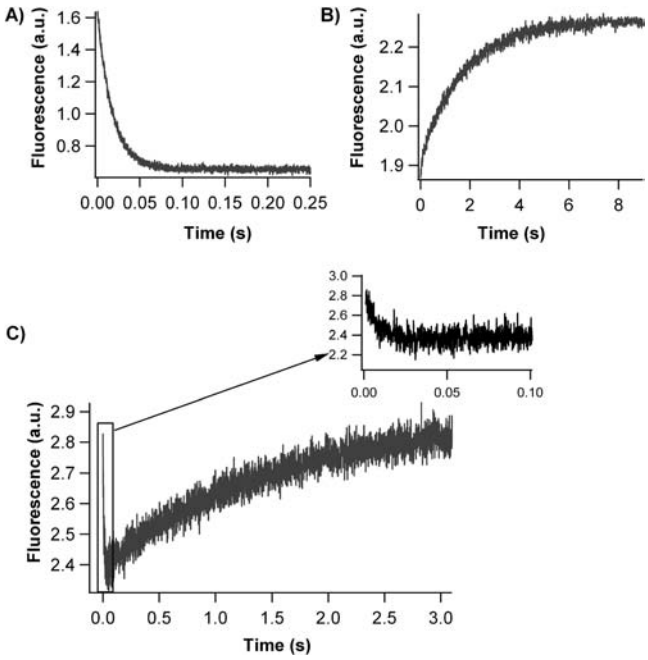

Figure 3. Unfolding kinetics of the domain insertion protein GL5/T4L. (A, B) Unfolding traces of isolated GL5 and T4L domains in the presence of 4 M GdmCl, respectively. (C) A typical unfolding trace of the domain insertion protein GL5/T4L in 4 M GdmCl. The tryptophan fluorescence of the combined protein exhibits a fast decay process followed by a slower increasing phase. Inset is a zoom in of the fast decay process on a shorter time scale.

inserted into the second loop of the host GL5 domain (Figure 1B). Since the N- and C-termini of T4L are within close proximity (0.8 nm apart from each other, PDB code 1L63), GL5 should be able to tolerate the insertion of T4L from a topological perspective. However, it is of note that the insertion of T4L does disrupt the sequence continuity of GL5.

The far-UV-CD spectrum of the combined hybrid protein GL5/T4L suggested that the host GL5 and the guest T4L domains are well-folded and assume secondary structures that are similar to that of isolated GL5 and T4L domains (Figure S2 of the Supporting Information). To further verify that both domains fold properly in the engineered combined hybrid protein GL5/T4L, we also used stopped-flow techniques to monitor their chemical unfolding dynamics by following the change of tryptophan fluorescence of both domains. The unfolding of isolated GL5 and T4L domains exhibits characteristic features that are easily distinguishable from one another: the unfolding of isolated GL5 results in a single exponential decay of the tryptophan fluorescence (Figure 3A), while the unfolding of isolated T4L results in an exponential increase in the tryptophan fluorescence (Figure 3B). Using these spectroscopic features, we can directly confirm the folding of both GL5 and T4L domains in the engineered domain insertion protein GL5/T4L. As shown in Figure 3C, the tryptophan fluorescence of GL5/T4L follows two distinct phases upon unfolding in the presence of 4 M GdmCl: the fluorescence showed a fast single exponential decay process followed by a slower increasing phase. The fast decay phase is the signature of the unfolding of the folded host protein GL5, and the slower increasing phase is the unfolding signature of the folded guest protein T4L in the combined hybrid protein. This result confirmed that both the host protein GL5 and the guest protein T4L are well-folded in the domain insertion protein GL5/T4L, and the result is consistent with previously reported examples of domain insertion proteins.[34,35,41]

Mechanical Unfolding Hierarchy of GL5 and T4L is Reversed in the Combined Hybrid Protein GL5/T4L. To facilitate characterizing the mechanical unfolding of the combined hybrid protein GL5/T4L, we constructed the polyprotein $(GL5)_4$-GL5/T4L-$(GL5)_4$ (Figure 4A), where $(GL5)_4$ serves as "mechanical handles" and helps to identify mechanical unfolding events of hybrid protein GL5/T4L. Stretching the polyprotein $(GL5)_4$-GL5/T4L-$(GL5)_4$ results in characteristic sawtoothlike force−extension curves such as those shown in Figure 4B. The mechanical unfolding events of GL5 domains (colored in red) can be readily identified from these force−extension curves as those occurring with a ΔLc of $\sim$20 nm.[37] In addition to the unfolding events of GL5 domains in force−extension curves containing five or more GL5 unfolding events, we also always observed the appearance of a pair of unfolding force peaks (colored in green): a high unfolding force peak ($\sim$130 pN) followed immediately by a smaller amplitude unfolding force peak ($\sim$50 pN). These pairs of unfolding force peaks always begin with a higher unfolding force peak followed by a lower unfolding force peak. Fitting the WLC model of polymer elasticity to these pairs of unfolding force peaks reveals that the higher unfolding force peak has a contour length increment ΔLc of $\sim$20 nm, while the lower unfolding force peak has a ΔLc of $\sim$60 nm (Figure S3 of the Supporting Information). The sum of ΔLc for the two unfolding events measures an average $\Delta Lc_{(total)}$ of 80.6 $\pm$ 2.7 nm (avg $\pm$ SD, $n = 140$) (Figure 4C). The unfolded and fully extended hybrid protein GL5/T4L is $\sim$81.7 nm in length ((61 aa + 164 aa + 2 aa) $\times$ 0.36 nm/aa), and the distance between the N- and C-termini of GL5 is $\sim$2.4 nm. Therefore, the complete mechanical unfolding of the combined hybrid protein GL5/T4L will result in a theoretical ΔLc of 79.3 nm (81.7 nm − 2.4 nm = 79.3 nm), in good agreement with the experimentally determined $\Delta Lc_{(total)}$. These results strongly suggest that these pairs of high and low unfolding force peaks correspond to the mechanical unfolding of hybrid GL5/T4L domains, in which the high unfolding force peak corresponds to the unfolding of the host GL5 domain and the lower unfolding force peak corresponds to the subsequent unfolding of the inserted guest protein T4L.

It is also of note that, if the unfolding of the hybrid GL5/T4L occurs toward the end of the force−extension curves, the unfolding of the hybrid GL5/T4L frequently manifests itself as a single unfolding event with a $\Delta Lc_{(total)}$ of $\sim$80 nm, as those shown in curves d and e in Figure 4B. These apparent "single" unfolding events are due to the fact that, after the unfolding of the host protein GL5, the relaxed force is still sufficiently high such that it will trigger the immediate unfolding of T4L. Therefore, the unfolding of T4L will occur concurrently with the unfolding of the host protein GL5. In comparison, if the unfolding of GL5/T4L occurs at the beginning of the force−extension curves, the unfolding of GL5 will result in the relaxation of the stretching force to a value that is well below 50 pN (see curves a−c in Figure 4B). Thus, T4L will temporarily remain folded after GL5 unfolds, allowing us to observe the unfolding of T4L separately.

The unfolding force histogram of the hybrid GL5/T4L domains reveals a bimodal distribution (Figure 4D) with the first peak centered at 50 $\pm$ 14 pN ($n = 64$) and the second at

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

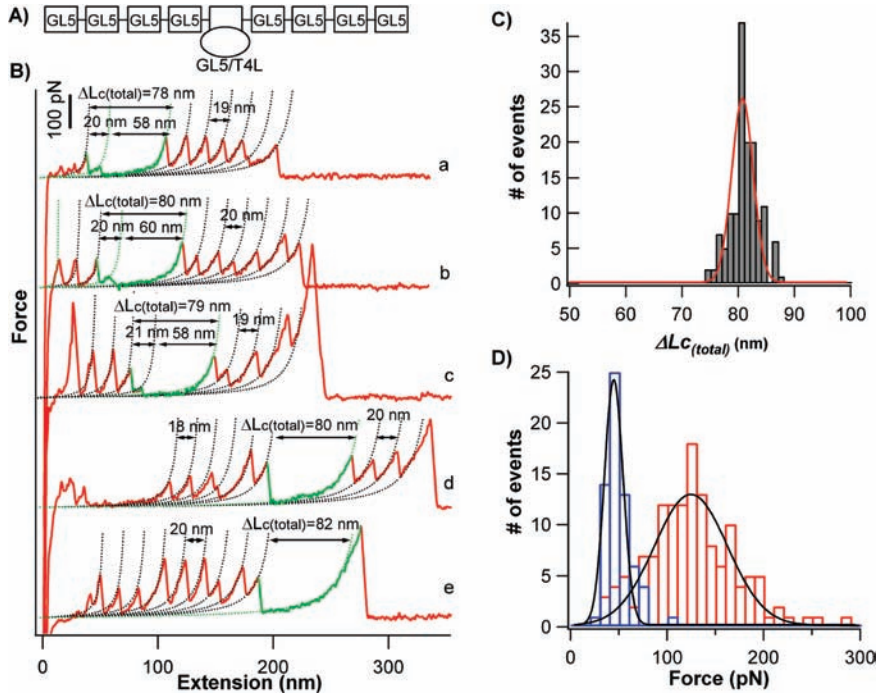

Figure 4. Mechanical unfolding of GL5/T4L displays a reversed mechanical unfolding hierarchy between GL5 and T4L domains. (A) A schematic of the polyprotein chimera $(GL5)_4$-GL5/T4L-$(GL5)_4$. (B) Typical force−extension curves of the polyprotein chimera $(GL5)_4$-GL5/T4L-$(GL5)_4$. In force−extension curves that contain five or more unfolding events of GL5 domains (colored in red) which are characterized by an unfolding force of ∼130 pN and a ΔLc of ∼20 nm, we always observe the unfolding of the combined protein GL5/T4L (colored in green) with a $\Delta Lc_{(total)}$ of ∼80 nm. The mechanical unfolding of GL5/T4L can occur in two steps (in curves a, b, and c) or in an apparent single step (in curves d and e). In both situations, when the unfolding of GL5/T4L occurs at the beginning of the force−extension curves, GL5/T4L tends to occur in two steps, giving rise to a pair of unfolding force peaks with the first one demonstrating a higher unfolding force and the second one being lower in force. WLC fits to the data showed that the higher unfolding force peaks had a ΔLc of ∼20 nm and the lower unfolding force peaks had a ΔLc of ∼60 nm. The sum of the ΔLc for the two events is ∼80 nm. This result suggests that the higher unfolding force event is due to the unfolding of the host protein GL5 while the lower unfolding force event corresponds to the unfolding of the guest protein T4L. Evidently, the mechanically more stable domain GL5 unfolds prior to the mechanical unfolding of the weaker guest protein T4L, a reverse mechanical unfolding hierarchy demonstrated between the two domains in the combined hybrid protein. If the unfolding event of GL5/T4L occurs toward the end of the force−extension curves, the mechanical unfolding of GL5/T4L occurs in an apparently single step with a ΔLc of ∼80 nm, suggesting that T4L unfolds immediately following the unfolding of the host protein GL5. (C) Histogram of total contour length increment $\Delta Lc_{(total)}$ for the mechanical unfolding of GL5/T4L. The Gaussian fit (red line) to the experimental data measures an average $\Delta Lc_{(total)}$ of 80.6 ± 2.7 nm (avg ± SD, n = 140). (D) Unfolding force histogram for the domain insertion protein GL5/T4L. Red bars indicate the unfolding force histogram of the host protein GL5 in the hybrid protein, while blue bars correspond to the unfolding force histogram of the guest T4L. Gaussian fits measured the average unfolding force of 126 ± 46 pN (n = 140) for the host GL5 domain and 50 ± 14 pN (n = 64) for the inserted guest protein T4L, respectively. The unfolding forces were measured at a pulling speed of 400 nm/s.

126 ± 46 pN (n = 140). The amplitude of the two unfolding force peaks is very similar to that measured previously for isolated T4L[24,25] and GL5 domains,[37] suggesting that the domain insertion does not change the mechanical resistance of both domains significantly. It is clear that the mechanical unfolding of the hybrid GL5/T4L does not follow the conventional mechanical unfolding hierarchy. Instead, domain insertion results in a reversed mechanical unfolding hierarchy where the mechanically stable GL5 unfolds first and the mechanically labile T4L unfolds last. Thus, the folded structure of the host protein GL5 effectively prevents T4L from "seeing" the stretching force before GL5 unfolds, providing the mechanically labile T4L with an effective protection against an applied stretching force.

Domain Insertion Significantly Prolongs the Lifetime of T4L under a Stretching Force. Reversed mechanical unfolding hierarchy delays the unfolding of T4L dramatically, which is equivalent to prolonging the lifetime of T4L under a stretching force. To directly illustrate this point, we have carried out force−clamp spectroscopy[39] measurements on $(GL5)_4$-T4L-$(GL5)_4$ and $(GL5)_4$-GL5/T4L-$(GL5)_4$ to directly quantify the effect of domain insertion on the lifetime of T4L under a stretching force.

When subject to a constant stretching force of 70 pN, the unfolding of $(GL5)_4$-GL5/T4L-$(GL5)_4$ results in characteristic stepwise elongation of the polyprotein (Figure 5A). The unfolding of GL5 domains gave rise to unfolding steps of ∼17 nm amplitude, which agrees well with the extension of a 20 nm long polypeptide chain at a stretching force of 70 pN. In addition to the unfolding steps of GL5 domains, we also observed the unfolding steps of ∼70 nm amplitude (colored in green), which corresponds to the complete unfolding of GL5/T4L hybrid proteins (Figure 5A). Since the stretching force is constant, T4L will be immediately exposed to a high stretching force of 70 pN, which is high enough to trigger the immediate unfolding of T4L after the unfolding of the host protein GL5. Therefore, the unfolding of GL5/T4L manifests as a single step instead of two. The lifetime of T4L in the hybrid protein is thus dictated by the lifetime Δt of the host protein GL5. It is evident that the lifetime of T4L follows a single exponential decay. An exponential fit to the experimental data measures an average lifetime of 123 ms under a stretching force of 70 pN (Figure 5D). This lifetime of T4L in GL5/T4L is essentially determined by the lifetime of the host protein GL5, which is almost identical to the lifetime of stand-alone GL5 domains (Figure S4 of the Supporting Information).

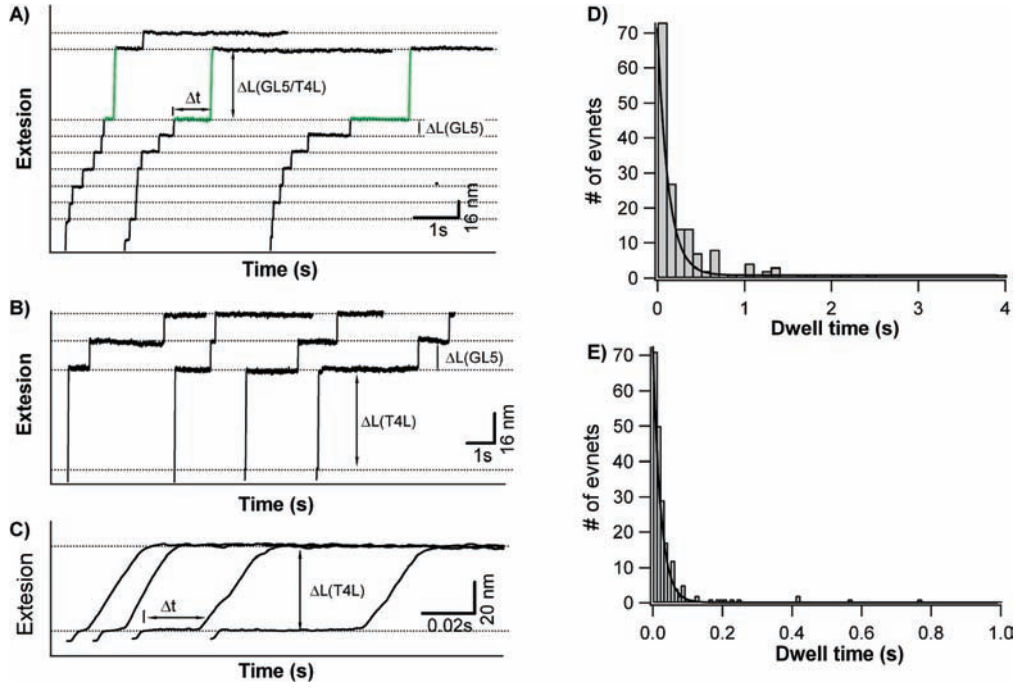

Figure 5. Lifetime of T4L under a stretching force is significantly prolonged in the domain insertion protein GL5/T4L. (A) Typical extension−time curves caused by stretching the polyprotein chimera $(GL5)_4$-GL5/T4L-$(GL5)_4$ at a constant force of 70 pN. The staircaselike traces show the sequential unfolding events of individual domains in the polyprotein. The unfolding of isolated GL5 results in a stepwise elongation of ∼17 nm in the end-to-end distance of the polyprotein chain, which is labeled as ΔL(GL5). The unfolding of the domain insertion protein GL5/T4L results in a much larger elongation of ∼69 nm, which is labeled as ΔL(GL5/T4L). The lifetime of T4L in the domain insertion protein GL5/T4L is labeled as Δt. (B) Typical extension−time curves caused by stretching the polyprotein chimera $(GL5)_4$-T4L-$(GL5)_4$ at a constant force of 40 pN. The unfolding of T4L results in the elongation of the polyprotein by ∼50 nm, which is labeled as ΔL(T4L). Upon exposure to the stretching force of 40 pN, T4L unfolds immediately and precedes the unfolding of all GL5 domains. (C) A close-up view of unfolding events of T4L. It is of note that the lifetime of T4L under a stretching force of 40 pN, labeled as Δt, is in the range of milliseconds. (D) Histogram of the lifetime distribution of T4L in the domain insertion protein GL5/T4L at 70 pN. A single exponential fit (solid line) to the experimental data measured an average lifetime of 123 ms. (E) Histogram of the lifetime distribution of T4L under a stretching force of 40 pN. A single exponential fit (solid line) to the experimental data measured an average lifetime of 22 ms.

To determine the effect of domain insertion on the lifetime of T4L, we also measured the lifetime of T4L in polyprotein $(GL5)_4$-T4L-$(GL5)_4$. Stretching $(GL5)_4$-T4L-$(GL5)_4$ at a constant force of 70 pN results in the instantaneous unfolding of T4L, which is too fast to measure its lifetime. Therefore, we carried out force−clamp experiments at a constant force of 40 pN. Stretching $(GL5)_4$-T4L-$(GL5)_4$ at 40 pN results in staircaselike extension−time relationships, which are characterized by a group of unfolding steps of a step size of ∼16 nm preceded by a single unfolding step of a step size of ∼52 nm (Figure 5B). The unfolding events that occur with a step size of 16 nm are due to the unfolding of GL5 domains in the polyprotein chain, while the unfolding events of a step size of 52 nm correspond to the unfolding of T4L. The measured step sizes for GL5 and T4L agree well with the extension of the unfolded GL5 and T4L at a stretching force of 40 pN, respectively.[25,37] The lifetime of T4L is indicated by Δt (Figure 5C). It is evident that the unfolding of T4L under a stretching force of 40 pN occurs rapidly prior to the unfolding of GL5 domains. An exponential fit to the lifetime distribution of T4L measures an average lifetime of ∼20 ms under a stretching force of 40 pN (Figure 5E).

It is obvious that the lifetime of T4L at 70 pN in the hybrid protein GL5/T4L is longer than the intrinsic lifetime of T4L at 40 pN. To make a more direct comparison, we used the unfolding distance Δx_u of 0.75 nm for T4L[25] to calculate the equivalent lifetime of T4L under a stretching force of 70 pN. This theoretical lifetime is ∼0.08 ms, which is 1500 times shorter than the lifetime of T4L at 70 pN, vividly demonstrating the effects of domain insertion on prolonging the lifetime of T4L.

Discussion

Domain insertion is a powerful new method to engineer proteins of novel functions and has been previously used in applications such as the engineering of allosteric protein switches.[35] For example, a calcium sensor was engineered by inserting a calmodulin domain into GFP.[34] Here, we endeavor to extend the use of the domain insertion approach to the construction of multifunctional artificial elastomeric proteins. Our results demonstrated the feasibility of engineering a novel elastomeric protein containing a reverse mechanical unfolding hierarchy via a domain insertion approach. Although the sequence continuity of the GL5 host protein is disrupted by the insertion of the guest protein T4L, both the host and guest proteins fold correctly and preserve their original mechanical stability, suggesting that there is no mechanical cross-talking between the host and guest proteins. The placement of a mechanically labile enzyme T4L in a mechanically stronger domain GL5 results in a significant increase in the effective lifetime of T4L thus opening up an avenue toward incorporating mechanically labile proteins into nanomechanical devices. Nonmechanical proteins, such as most enzymes, are not designed to withstand mechanical stretching forces under physiological conditions and are typically unstable when exposed to stretching forces. When they are fused to a mechanical protein in a tandem fashion, nonmechanical proteins

will be easily unfolded by an applied stretching force and experience a subsequent loss in activity, such as enzyme functionality. A domain insertion approach allows one to incorporate the enzyme into a mechanically stable host in order to prevent the enzyme from unraveling under a mechanical stretching force. Since the host and guest proteins behave independently of each other in the constructed hybrid protein, the enzyme will preserve its original activity and perform its designated functions under a stretching force.

Hybrid mechanical proteins engineered in this way also represent a prototype of mechanically controlled enzyme switches. Since the host protein GL5 is mechanically stable, the forced unfolding reaction of the host protein will serve as the switch to turn off the enzymatic activities of the guest enzyme in a force−clamp experiment. If the hybrid protein is operated at a lower force, GL5 will be folded, and the guest enzyme will not be subject to the stretching force and will thus remain enzymatically active. If the enzyme needs to be turned off, increasing the stretching force will trigger the mechanical unfolding of GL5 and subsequently cause the unfolding of the guest enzyme. Such concepts will make it possible to design novel multifunctional nanomechanical elements for novel applications in nanomechanics and nanobiotechnology.

Acknowledgment. We thank Ms. Dingyue Khor for her valuable contributions in protein engineering work and Ms. Ashlee Jollymore for critical reading of the manuscript. We are indebted to Prof. Julio M. Fernandez for his support and encouragement of this project and for his intellectual insights into some of the ideas reported in this paper. This work is supported by Natural Science and Engineering Research Council of Canada (NSERC), Canada Research Chairs program, and Michael Smith Foundation for Health Research. H.L. is a Michael Smith Foundation for Health Research Career Investigator. Q.P. is supported by Pacific Century Graduate Scholarship from the Province of British Columbia.

Supporting Information Available: SDS-PAGE gel for combined hybrid protein GL5/T4L and polyproteins $(GL5)_4$-T4L-$(GL5)_4$ and $(GL5)_4$-GL5/T4L-$(GL5)_4$, far-UV-CD spectra of isolated GL5, isolated T4L, and hybrid GL5/T4L proteins, histograms of ΔLc for host GL5 domain and inserted guest domain T4L in hybrid GL5/T4L, and histogram of the dwell time of stand-alone GL5 domains at 70 pN. This material is available free of charge via the Internet at http://pubs.acs.org.

JA903589T