

DOI: 10.1002/((smt.d.201800039))

Article type: Communication

Mechanical polyprotein assembly using Sfp and sortase-mediated domain oligomerization for single-molecule studies

Haipei Liu[#], Duy Tien Ta[#], Michael A. Nash^{*}

[#] Equally contributing

Dr. H. Liu, Dr. D. T. Ta, Prof. Dr. M. A. Nash
Institute for Physical Chemistry, Department of Chemistry, University of Basel and
Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

E-mail: michael.nash@unibas.ch

Keywords: single-molecule force spectroscopy, protein folding, atomic force microscopy, sortase, cohesion and dockerin, polyprotein.

Abstract. Single-molecule force spectroscopy with the atomic force microscope provides remarkable details on the energy landscapes governing protein (un)folding and intermolecular complex dissociation. In such experiments, multi-domain polyproteins consisting of multiple copies of independently folded domains are used to provide internal controls that can be identified by characteristic contour length increments, unfolding forces, and/or unfolding substeps. Here we present a new approach to mechanical polyprotein synthesis which relies on post-translational enzyme-mediated oligomerization of domain monomers. We engineered mutant variants of the immunoglobulin 27 (I27) domain, as well as a fusion protein consisting of a bacterial cellulose binding module (CBM), an Ig-like X-module (XMod), and a mechanostable receptor called Dockerin (Doc). By utilizing Sfp phosphopantetheinyl transferase and Sortase enzymes, our system allowed for I27-domain oligomerization into polyproteins of varying lengths followed by C-terminal capping with a mechanostable Doc receptor. We characterized the number of oligomerized domains per molecule, the single-domain unfolding forces, and the complex rupture forces of the post-translationally assembled polyproteins using >40 hour automated AFM-SMFS with a Cohesin (Coh) -modified cantilever. Use of the specific Coh-Doc interaction to unfold the polyproteins provided a high

yield of ~3,800 specific single-molecule interaction curves. Our approach is advantageous for assembly of polyproteins from domains that lack proper folding or are insoluble in a polyprotein format.

Main text.

Techniques involving the application of mechanical force to individual molecules, typically DNA, RNA and protein molecules, are collectively referred to under the term ‘single-molecule force spectroscopy’ (SMFS)^[1–8]. SMFS provides a wealth of information on biomolecules, describing in a high level of detail biomolecular behaviors such as spontaneous protein folding and unfolding rates^[9], adoption of short-lived intermediate folded states^[10], response of DNA helices to torque and tension^[11], directionally anisotropic response of proteins to forces^[12–14], and mechanically activated catch bonding^[15,16], to name but a few. The associated techniques used for performing SMFS all have in common the coupling of molecules to nano-to-micro scale force transducers. Magnetic small particles^[17], optically trapped beads^[18], microfabricated silicon nitride cantilevers^[19–21], and beads tethered to a surface under centrifugal forces^[22,23] are all examples of the diversity of methods used for SMFS.

When attempting to measure individual molecules using SMFS, often times it is necessary to have a signal that discriminates specific molecular behaviors (e.g., protein unfolding events, complex unbinding) from the unavoidable signals that arise due to non-specific interactions between the force transducer and the surface. To address this issue, SMFS with the atomic force microscope (AFM) often times relies on internal control elements that are genetically encoded into the sequence of a so-called ‘polyprotein’^[24].

Polyproteins are large multi-domain proteins containing a domain of interest sandwiched between multiple tandem copies of independently foldable marker domains. The marker domains, sometimes referred to as ‘fingerprint’ domains, aid in AFM-SMFS data analysis by providing clear features such as contour length increments, unfolding forces, and/or

the presence of unfolding substeps in the trace that can be searched for in an algorithmic way. Characterization of the molecular response of polyproteins has been instrumental in the development of the field of nanomechanics and mechanobiology^[25-27].

Production of polyproteins has in the past been achieved in a variety of ways. Due to the repetitive nature of the DNA sequences encoding polyproteins, care must be taken during PCR and cloning such that multiple primer annealing sites are avoided. A restriction digest and ligation-based cloning system involving unique restriction sites flanking tandem I27 modules was among the earliest reported techniques^[28] for forming polyprotein gene sequences. More recently, the use of tandem repeats with shuffled codons was also reported^[29]. Gibson assembly cloning for synthesizing repetitive polyprotein genes was also recently reported^[30], as well as plasmid systems specifically developed for force spectroscopy^[31]. Furthermore, previously halo-tags were combined with engineered cysteines and used to achieve continuous long-term measurements of single polyprotein molecules for SMFS studies^[32]. Although these techniques are able to produce polyproteins at reasonable titers using heterologous expression in *E. coli*, they all have in common the limitation that folding of the proteins into their native conformation needs to occur while the domain is tethered to other domains in the polyprotein chain. In some cases, this may limit the application of the technique to relatively simple protein folds. To address this issue, several posttranslational methods for polyprotein assembly have been reported which circumvent the issue. Both bis-maleimide cross-linkers as well as oxidative disulfide bond formation have been used to posttranslationally assemble polyproteins containing 2 engineered cysteine residues^[33,34]. Such approaches have allowed for assembly of polyproteins linked through different anchoring geometries and allowed for measurement of directionality dependent unfolding properties (i.e., mechanical anisotropy) to be studied^[35,36].

Here we present an alternative method to posttranslational polyprotein assembly. We developed an approach wherein the marker domain I27 was engineered to contain distinct N- and C-terminal peptide tags which were ligated in a head-to-tail fashion by the transpeptidase

Srt A from *Staphylococcus aureus*^[37,38]. Sortase A (Srt) is an enzyme that has received considerable attention as a tool for protein engineering and macromolecular assembly^[39,40]. Srt has been used for formation of protein conjugates, heterodimeric enzyme complexes^[41], and linkages between DNA and proteins^[42]. Recently we reported the use of Doc domains as universal pulling handles for highly parallelized SMFS^[43]. We also extended this method through attachment of Doc domains onto full length polyproteins to improve specificity during molecule pickup^[44]. We furthermore have reported the use of Srt-mediated ligation to functionalize elastin-like polypeptides for single-molecule studies^[45]. Here we expand on this work by demonstrating for the first time the post-translational assembly and single-molecule mechanical unfolding of multi-domain polyproteins using a combination of marker domain engineering, enzymatic protein ligation, and specific receptor-ligand interaction.

To demonstrate the concept of post-translationally assembled mechanical polyproteins (**Figure 1**), we built the molecular system from the bottom-up by first linking the sulfhydryl group of Coenzyme A (CoA) to an aminosilanized coverglass surface through a 5kDa NHS-PEG-Maleimide spacer, similarly to previously published reports^[46]. PEG served as an elastic spacer and passivation agent at the surface, while CoA allowed for covalent bond formation and tethering to the surface of proteins carrying a ybbR peptide tag (DSLEFIASKLA) using the enzyme Sfp phosphopantetheinyl transferase^[47-49].

We engineered one variant of the I27 domain of the sarcomeric protein Titin, a classical marker domain for AFM-SMFS, to contain an N-terminal ybbR tag and a C-terminal Srt tag (-LPETGG). Incubation of the PEG-CoA surfaces with μM quantities of ybbR-I27-LPETGG in the presence of Sfp resulted in covalent linkage of the I27 marker domain to the surface through its N-terminus while preserving the reactive C-terminal -LPETGG motif for further derivatization with subsequent domains of the polyprotein using Srt. In the second cycle, multi-step oligomerization of the polyprotein was achieved upon addition of Srt together with another variant of I27, this one containing an N-terminal GGG-tag and a C-terminal -LPETGG tag.

At first we produced a recombinant Met-HIS-TEV-GGG-I27-LPETGG containing an N-terminal His₆ tag, followed by a TEV protease cleavage site and a GGG Srt tag. TEV treatment was performed to generate the protein with GGG directly at the N-terminus of the molecule which is required for Srt-mediated ligation. However, we found that the cleavage efficiency by TEV protease was as low as 10% (data not shown). To overcome this issue, we next produced the GGG-HIS-I27-LPETGG with the N-terminal GGG tag located immediately following the Met start site (i.e., MGGG-). We relied on post-translational processing of the protein by the native *E. coli* machinery, which removed the Met amino acid at the N-terminus. Mass spectrometry analysis (now shown) demonstrated that a majority of the protein was natively processed and lacked the N-terminal Met residue resulting in the GGG oligoglycine motif being present directly at the N-terminus of the molecule. This result was in accordance with a prior study^[50] which reported that the N-terminal Met was mostly cleaved if the adjacent residue is glycine.

The I27 domains bearing the corresponding GGG- and -LPETGG sequences were able to self-polymerize in the presence of Srt. The resulting polyproteins contained different numbers of I27 domains ranging from 1-5 (see below). For all reactions, we used an enhanced Srt (Δ 59 Srt) with mutations P94R/D160N/D165A/K190E/K196T to confer significantly improved K_m and K_{cat} compared to the wild type Srt enzyme^[44,51].

Following rinsing, the construct was capped (Figure 1, step 6) by a multi-domain protein (GGG-CBM-XMod-Doc), again by Srt ligation. The capping protein contained an N-terminal GGG-tag for Srt linkage, a cellulosomal CBM domain, and an XMod-Doc tandem domain. XMod-Doc is a receptor domain that forms a mechanically stable linkage (>500 pN rupture forces^[46]) with its complementary Cohesin (Coh) domain. Coh was expressed as a ybbR-CBM-Coh fusion protein for attachment to the cantilever tip. The CBM-fused Cohesin provided an additional CBM marker domain to be attached to the cantilever.

Following heterologous over expression and His-tag affinity purification of the engineered protein domains described above, we performed bulk experiments to test Srt-mediated oligomerization of 50 μ M GGG-I27-LPETGG by incubation with 15 μ M of Srt A at room temperature for 1, 2 or 3 hours. The progression of the reaction was monitored using denaturing SDS polyacrylamide gel electrophoresis (**Figure 2**). The results showed after 1 hour the appearance of bands corresponding primarily to dimers, and trimers, with some tetramers and larger molecular weight polyproteins appearing only very faintly. We also observed the presence of a band at a molecular weight of \sim 30 kDa, which was consistent with an acyl-enzyme intermediate complex between the I27 monomer and Srt^[52]. These results confirmed the functionality of the engineered Srt tags and the ability to assemble I27 polyproteins in one-pot reactions using Srt-mediated oligomerization.

After confirming the basic functionality of the assembly strategy in bulk, we prepared a coverglass surface following the scheme shown in Figure 1, and probed the surface with Coh-CBM covalently attached to a silicon nitride cantilever through a C-terminal ybbR tag. A typical single-molecule force vs. distance trace from such an experiment is shown in **Figure 3**. The data from the single traces were histogrammed in contour length space (Figure 3b) using a worm-like chain transformation with an assumed persistence length of 0.4 nm^[53,54]. By noting that the amino acid sequence lengths of the domains are 89 amino acids for I27 and 164 amino acids for CBM, and subtracting the respective folded lengths of 4.32 nm and 2.2 nm, we calculated a theoretically expected contour length increment of 28 nm for I27 and 58 nm for CBM. This assumes a length of 0.365 nm contour length per amino acid. The distances between the peaks in the contour length histogram were then assigned to the unfolded marker domains, and peak assignments in the force distance trace could be made. This analysis demonstrated that, for the trace shown in Figure 3a, a polyprotein containing 3 I27 domains and one CBM-XMod-Doc was stretched from the N- to C-terminus. The second CBM unfolding length

increment was attributed to the CBM domain attached to the Cohesin on the cantilever, while the XMod in this trace did not unfold.

The surface was probed continuously for >42 hours using automated SMFS and 71,252 force-distance traces in total were acquired. The majority of such curves were not usable because they contained either no interactions or very complex multiple interactions. For the analysis, we only considered curves which contained 2 CBM contour length increments. This ensured that we only analyzed traces representing single polyproteins stretched from end to end. The total number of curves matching this criterion was 3796 out of 71,252, representing a curve yield of 5.33%. Due to the stochastic nature of the enzyme-mediated oligomerization approach, we obtained polyproteins with differing numbers of I27 domains. **Figure 4** shows example unfolding traces of polyproteins bearing anywhere from 1-5 copies of the I27 domain, respectively. In some cases, an unfolding event corresponding to the XMod sub-domain of the capping protein was also observed (e.g., Figure 4, N=3 and N=5).

Figure 5 presents the rupture force distributions obtained for Coh-Doc with and without XMod unfolded (Figure 5a), as well as the unfolding force distributions for CBM (Figure 5b) and I27 (Figure 5c). CBM had a lower median unfolding force than I27, and so it was usually unfolded prior to unfolding of the I27 domains in the traces shown in Figure 4. Some overlap in the unfolding force distributions of CBM and I27 is observed and occasionally I27 unfolded first (Figure 4, N=3). For polyproteins containing 3 or more I27 domains, a statistical effect also led to I27 unfolding events occurring prior to CBM unfolding in the force-distance traces. Simply based on the fact that there were more I27 domains than CBM domains present in the polyprotein, many times I27 was observed to unfold prior to a CBM domain in the single-molecule unfolding traces, despite that the I27 has a higher median unfolding force. We fit each of the rupture force distributions using the probability distribution function from a Bell-Evans model of force induced molecular transitions (**Equation 1**)^[55,56]:

$$p(F) = \frac{k_0}{r} \exp \left[\frac{F\Delta x}{k_B T} - \frac{k_0 k_B T}{r\Delta x} \left(e^{\frac{F\Delta x}{k_B T}} - 1 \right) \right] \quad (1)$$

Here, k_0 is the natural off rate of the complex dissociation or unfolding transition, r is the loading rate, F is the rupture or unfolding force, Δx is the distance to the transition state along the reaction coordinate, and $k_B T$ is the thermal energy. The results from this fitting are shown in Figure 5 for each of the unfolding/unbinding events of interest, as well as the corresponding means loading rate for each population of unfolding events. The Δx value obtained for I27 in this experiment is lower than previously published studies which reported Δx values from 0.25 - 0.3 nm^[25,30,57]. The observed differences could be attributable to differences in cantilever stiffness, or the fitting method (we used here only a single averaged loading rate). Furthermore, we analyzed here all I27 domain unfolding events from the traces and not only the final one. This tends to broaden the force distributions and leads to lowering of the Δx values and increasing of k_{off} values due to the so called 'N-effect' ^[58]. For a more precise estimate of the unfolding parameters of I27, only the final unfolding events should be considered.

Due to the stochastic nature of the Srt-mediated assembly process, the resulting polyproteins were polydisperse, containing variable numbers of oligomerized I27 domains. In **Figure 6**, we quantified the number of I27 domains per molecule that passed our selection filter of 2 CBM unfolding events during the AFM measurement. The results demonstrate that the majority of I27 domains appeared as monomer and dimer species (49.3%, and 39.7% respectively), while ~10% were assembled into trimers. Less than 1% of total polyproteins which passed our selection filter of 2 CBM unfolding events contained 4 or more I27 domains. We note that in addition to the I27 domains, the polyproteins formed in this manner all contained C-terminal CBM-XMod-Doc. The bulky nature of the globular I27 domains limits the ultimate sizes that can be achieved through assembly of polyproteins through enzymatic ligation. In addition to the reported reaction containing 15 μ M Srt, we also tested the reaction

at 7.5 μ M Srt. Both conditions resulted in the same *in vitro* polymerization and SMFS outcomes which were limited to polyproteins of maximally 5 I27 repeats. For our purposes, and for the purposes of many groups in the field, large polyproteins containing >4 domains are not required, therefore we believe this method is of value even if the ultimate size of the polyproteins assembled in this way is limited to ~3-4 domains.

In summary, we have described a new approach to the formation of multidomain mechanical polyproteins for single-molecule studies. We described an approach that relies on Sfp and Srt enzymes to first covalently link a modular building block to a coverglass surface using Sfp, and then oligomerize marker domains off of this building block using Srt to create polydisperse multi-domain proteins. The oligomerization requires only installation of a GGG-motif at the N-terminus and an LPETGG sequence at the C-terminus of the monomer domains. Finally, XMod-Doc was ligated to the C-terminus of the assembled polyproteins to provide a specific receptor that bound non-covalently to a Coh-CBM protein covalently linked to the cantilever.

Coh-Doc interactions are very reliable in AFM-SMFS studies^[59-62] because Coh attached to the cantilever can withstand reliably >50,000 rupture cycles without loss of binding activity. The C-terminal capping molecule (GGG-CBM-XMod-Doc) has itself two built-in marker domains, providing identifiable unfolding events that confirm the full length polyprotein was stretched from end to end. The high mechanical stability of XMod meant it only unfolded occasionally prior to Coh-Doc complex rupture. In traces where XMod did unfold, the Coh-Doc complex was found to be significantly destabilized and ruptured at much lower forces, consistent with prior studies^[46,59]. The CBM domain with an unfolding force of ~150 pN regularly unfolded prior to Coh-Doc rupture. The presence of 2 CBM unfolding events, one derived from the CBM molecule on the cantilever and one derived from the CBM attached at the C-terminal end of the surface-assembled polyproteins was used as a filter for curve

analysis which ensured only full length constructs were analyzed. This approach allowed us to quantify the distribution of polyprotein lengths resulting from Srt-mediated oligomerization in terms of the number of I27 monomers added per molecule. Previously, AFM-SMFS had been used to quantify the polydispersity of living free radical polymerization reactions performed using reversible addition fragmentation chain transfer polymerization^[63]. Our analysis here extends this concept of polydispersity characterization at the single-molecule level for an enzyme mediated reaction. These tools should provide alternative options for groups wanting to assemble a wide variety of polyproteins. Specifically, this approach will be advantageous for studying the mechanical properties of proteins with potentially complex folding pathways that are not able to correctly fold when expressed *in vivo* as polyproteins.

Experimental Section

Cloning

The plasmids: pET28a-ybbR-HIS-I27-LPETGG (a), pET28a-HIS-TEV-CBM-XMod-Doc (b), pET28a-Coh-CBM-HIS-ybbR (c), and pET28a-Srt A (d) were kind gifts from Hermann Gaub's lab at Ludwig-Maximilians-Universität Munich. The GGG-HIS-I27-LPETGG gene was amplified from plasmid (a) using primers 1 and 2, which added the 3 residues GGG in front of the HIS-I27-LPETGG sequence, resulting in the construct pET28a-GGG-HIS-I27-LPETGG (e). The GGG-CBM-XMod-Doc gene was amplified from plasmid (b) using primers 1 and 3, adding the 3 residues GGG in front of the HIS-TEV-CBM-XMod-Doc. This resulted in the construct pET28a-GGG-CBM-XMod-Doc (f). All constructs were verified by sequencing.

Table 1. Primer sequences used for cloning.

Primer	Sequence (5' to 3')
1	GTTGTTTCATATGGGGGGCGGTCACCACCACCACCACCACGGAGAA
2	AGCCGGATCTTACTCGAGTTA
3	ATCTTACTCGAGTTATTCTTCTTC

Protein expression and purification

Plasmids (a), (c), (d), (e) and (f) were transformed to *E. coli* BL21(DE3) cells and positive cells were selected on LB-agar plates supplemented with 50 µg/mL kanamycin (LB^{kan}) at 37 °C overnight. Single colonies were picked and inoculated into 3 mL LB^{kan}. This preculture was grown at 37 °C overnight with shaking. On the following day, 1 mL of the preculture was inoculated into 100 mL LB supplemented with 50 µg/mL kanamycin. The cultures were grown at 37 °C with shaking until the OD₆₀₀ reached 0.5. After that, protein expression was induced with 0.5 mM IPTG. The expression was performed at 25 °C for 6 hrs with shaking. The cells were then harvested by centrifugation at 4000x g for 10 min and re-suspended in 10 mL lysis buffer (50 mM Tris, 50 mM NaCl, pH 8.0 containing 1 mg lysozyme and 20 Units of *DNaseI*) and incubated on ice for 15 min before sonication (35% amplitude, 2 second pulse on, 2 second pulse off) for 10 min. The cell lysate was then centrifuged at 18,000x g at 4 °C for 30 min. All His₆-tagged proteins were purified on a His-Trap FF column using a GE-AKTA chromatography system. The purified proteins were analysed on 12% SDS-PAGE gel for purity and ESI-Mass spectrometry for correct masses. The protein concentration was determined using the BCA assay.

In vitro Srt-mediated polymerization

For *in vitro* Srt-mediated domain oligomerization, a 500 µl mixture containing 50 µM GGG-I27-LPETGG, 15 µM Srt A, 1 mM CaCl₂ in TBS buffer pH 7.5 was incubated at room

temperature. Aliquots of 20 μl were sampled after 1, 2, and 3 hrs and were analyzed on 12% SDS-PAGE gel.

AFM Sample Preparation

AFM cantilevers and coverglasses were first cleaned by UV-ozone treatment followed by soaking in piranha etching solution and rinsing in water. Next, 3-Aminopropyl (diethoxy) methylsilane (APDMES, ABCR GmbH, Karlsruhe, Germany) was used to silanize the surface of levers and coverglasses with amine groups. The amine groups were subsequently conjugated to a heterobifunctional NHS-PEG-Mal linker (5 kDa; Rapp Polymere, Tübingen, Germany) in 50 mM HEPES buffer pH 7.5 for 30 min. Both PEGylated cantilevers and coverglasses were incubated with Coenzyme A (CoA, 200 μM) in coupling buffer (50mM sodium phosphate, 50mM NaCl, 10mM EDTA, pH 7.2) for 2 h at room temperature. Sfp-catalyzed coupling of the proteins Coh-CBM-ybbR and ybbR-I27-LPETGG to the CoA coated levers or surfaces was done in Ca-TBS measurement buffer (25 mM Tris, 72 mM NaCl, 1mM CaCl_2 , pH7.2) supplemented with 10 mM MgCl_2 at room temperature for 2 h. The cantilevers were washed extensively with measurement buffer and stored in it before measurement.

To build the I27 polyprotein structures, Srt-mediated polymerization of GGG-I27-LPETGG was done by incubating the ybbR-I27-LPETGG-modified coverglass with 80 μM GGG-I27-LPETGG and 15 μM Srt in Srt reaction buffer (50 mM Tris, 150 mM NaCl, 1mM CaCl_2 , pH7.2) at room temperature for 1 hour. Next, the coverglass was rinsed and again incubated with 60 μM GGG-CBM-XMod-Doc and 15 μM Srt in the same buffer at room temperature for 1 hour. Both cantilevers and coverglasses were rinsed extensively with measurement buffer before measurement.

AFM SMFS measurements and Data Analysis

SMFS measurements using the high-affinity Coh-Doc complex as a specific pulling handle were carried out using automated AFM-based SMFS (Force Robot 300, JPK Instruments). SMFS data were recorded in calcium supplemented TBS buffer (25 mM Tris, 72 mM NaCl, 1 mM CaCl₂, pH 7.2), at room temperature and at a constant pulling speed of 500 nm s⁻¹. Force distance curves were analysed using custom software scripts written in python. The data traces were filtered by searching for contour length increments that matched the lengths of the specific protein fingerprint domains: I27(~28 nm) and CBM(~58 nm). The histograms of unfolding forces were fitted with Bell-Evans model to determine values of the most probable unfolding force ($\langle F \rangle$), effective distance to the transition state (Δx) and the natural offrate (k_{off}).

Acknowledgments

This work was supported by the Branco Weiss Fellowship (ETH Zurich), by the Swiss National Center for Competence in Research Molecular Systems Engineering (NCCR-MSE), and by the European Research Council (Starting Grant 715207). The authors thank Lukas Milles (LMU Munich) for providing the python software for analysis of SMFS data, and thank Wolfgang Ott, Markus Jobst, and Ellis Durner (LMU Munich) for plasmids.

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

References

- [1] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, M. A. Nash, *J. Struct. Biol.* **2016**, *197*, 3.
- [2] T. Hoffmann, L. Dougan, *Chem. Soc. Rev.* **2012**, *41*, 4781.
- [3] K. C. Neuman, A. Nagy, *Nat. Methods* **2008**, *5*, 491.
- [4] T. E. Fisher, P. E. Marszalek, J. M. Fernandez, *Nat. Struct. Biol.* **2000**, *7*, 719.
- [5] E. M. Puchner, H. E. Gaub, *Curr. Opin. Struct. Biol.* **2009**, *19*, 605.
- [6] C.-K. Lee, Y.-M. Wang, L.-S. Huang, S. Lin, *Micron* **2007**, *38*, 446.
- [7] D. Alsteens, H. E. Gaub, R. Newton, M. Pfreundschuh, C. Gerber, D. J. Müller, *Nature Reviews Materials* **2017**, *2*, 17008.
- [8] D. Leckband, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 1.
- [9] M. Rief, J. M. Fernandez, H. E. Gaub, *Phys. Rev. Lett.* **1998**, *81*, 4764.
- [10] H. Yu, M. G. W. Siewny, D. T. Edwards, A. W. Sanders, T. T. Perkins, *Science* **2017**, *355*, 945.
- [11] J. Lipfert, J. W. J. Kerssemakers, T. Jager, N. H. Dekker, *Nat. Methods* **2010**, *7*, 977.
- [12] B. Jagannathan, P. J. Elms, C. Bustamante, S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820.
- [13] P. Zheng, C.-C. Chou, Y. Guo, Y. Wang, H. Li, *J. Am. Chem. Soc.* **2013**, *135*, 17783.
- [14] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, S. E. Radford, *Nat. Struct. Biol.* **2003**, *10*, 731.
- [15] N. Strohmeyer, M. Bharadwaj, M. Costell, R. Fässler, D. J. Müller, *Nat. Mater.* **2017**, *16*, 1262.
- [16] B. T. Marshall, M. Long, J. W. Piper, T. Yago, R. P. McEver, C. Zhu, *Nature* **2003**, *423*, 190.
- [17] C. Gosse, V. Croquette, *Biophys. J.* **2002**, *82*, 3314.
- [18] M. T. Woodside, S. M. Block, *Annu. Rev. Biophys.* **2014**, *43*, 19.
- [19] H.-J. Butt, B. Cappella, M. Kappl, *Surf. Sci. Rep.* **2005**, *59*, 1.
- [20] D. T. Edwards, T. T. Perkins, *J. Struct. Biol.* **2017**, *197*, 13.
- [21] M. S. Bull, R. M. A. Sullan, H. Li, T. T. Perkins, *ACS Nano* **2014**, *8*, 4984.
- [22] D. Yang, A. Ward, K. Halvorsen, W. P. Wong, *Nat. Commun.* **2016**, *7*, 11026.
- [23] K. Halvorsen, W. P. Wong, *Biophys. J.* **2010**, *98*, L53.
- [24] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3694.
- [25] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, *Science* **1997**, *276*, 1109.
- [26] M. Schlierf, H. Li, J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7299.
- [27] J. Brujic, R. I. Z. Hermans, S. Garcia-Manyes, K. A. Walther, J. M. Fernandez, *Biophys. J.* **2007**, *92*, 2896.
- [28] A. Steward, J. L. Toca-Herrera, J. Clarke, *Protein Sci.* **2002**, *11*, 2179.
- [29] Z. N. Scholl, E. A. Josephs, P. E. Marszalek, *Biomacromolecules* **2016**, *17*, 2502.
- [30] T. Hoffmann, K. M. Tych, T. Crosskey, B. Schiffrin, D. J. Brockwell, L. Dougan, *ACS Nano* **2015**, *9*, 8811.
- [31] J. Oroz, R. Hervás, M. Carrión-Vázquez, *Biophys. J.* **2012**, *102*, 682.
- [32] J. Valle-Orero, J. A. Rivas-Pardo, I. Popa, *Nanotechnology* **2017**, *28*, 174003.
- [33] P. Zheng, Y. Cao, H. Li, *Langmuir* **2011**, *27*, 5713.
- [34] G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 139.
- [35] H. Dietz, M. Rief, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1244.
- [36] H. Dietz, F. Berkemeier, M. Bertz, M. Rief, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12724.
- [37] S. Tsukiji, T. Nagamune, *ChemBiochem* **2009**, *10*, 787.

- [38] T. Proft, *Biotechnol. Lett.* **2009**, *32*, 1.
- [39] E. Steen Redeker, D. T. Ta, D. Cortens, B. Billen, W. Guedens, P. Adriaenssens, *Bioconjug. Chem.* **2013**, *24*, 1761.
- [40] L. Schmohl, D. Schwarzer, *Curr. Opin. Chem. Biol.* **2014**, *22*, 122.
- [41] K. Li, R. Zhang, Y. Xu, Z. Wu, J. Li, X. Zhou, J. Jiang, H. Liu, R. Xiao, *Sci. Rep.* **2017**, *7*, 3081.
- [42] M. A. Koussa, M. Sotomayor, W. P. Wong, *Methods* **2014**, *67*, 134.
- [43] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, H. E. Gaub, *Nat. Methods* **2014**, *11*, 1127.
- [44] E. Durner, W. Ott, M. A. Nash, H. E. Gaub, *ACS Omega* **2017**, *2*, 3064.
- [45] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, H. E. Gaub, *ACS Nano* **2017**, *11*, 6346.
- [46] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, M. A. Nash, *Nat. Commun.* **2014**, *5*, 1.
- [47] W. Ott, T. Nicolaus, H. E. Gaub, M. A. Nash, *Biomacromolecules* **2016**, *17*, 1330.
- [48] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15815.
- [49] J. Yin, A. J. Lin, D. E. Golan, C. T. Walsh, *Nat. Protoc.* **2006**, *1*, 280.
- [50] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, S. Blanquet, *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 8247.
- [51] I. Chen, B. M. Dorr, D. R. Liu, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11399.
- [52] S. Massa, N. Vikani, C. Betti, S. Ballet, S. Vanderhaegen, J. Steyaert, B. Descamps, C. Vanhove, A. Bunschoten, F. W. B. van Leeuwen, S. Hernot, V. Caveliers, T. Lahoutte, S. Muyldermans, C. Xavier, N. Devoogdt, *Contrast Media Mol. Imaging* **2016**, *11*, 328.
- [53] M. A. Jobst, C. Schoeler, K. Malinowska, M. A. Nash, *J. Vis. Exp.* **2013**, e50950.
- [54] E. M. Puchner, G. Franzen, M. Gautel, H. E. Gaub, *Biophys. J.* **2008**, *95*, 426.
- [55] E. Evans, K. Ritchie, *Biophys. J.* **1997**, *72*, 1541.
- [56] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten, *Biophys. J.* **1997**, *72*, 1568.
- [57] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 468.
- [58] R. C. Zinober, D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, D. A. Smith, *Protein Sci.* **2009**, *11*, 2759.
- [59] C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash, H. E. Gaub, *Nano Lett.* **2015**, *15*, 7370.
- [60] J. Wu, P. Li, C. Dong, H. Jiang, Bin Xue, X. Gao, M. Qin, W. Wang, Bin Chen, Y. Cao, *Nat. Commun.* **2018**, *9*, 620.
- [61] H. Takahashi, F. Rico, C. Chipot, S. Scheuring, *ACS Nano* **2018**, DOI 10.1021/acsnano.7b08973.
- [62] A. M. Vera, M. Carrión-Vázquez, *Angew. Chem. Int. Ed Engl.* **2016**, *55*, 13970.
- [63] T. Tischer, R. Gralla-Koser, V. Trouillet, L. Barner, C. Barner-Kowollik, C. Lee-Thedieck, *ACS Macro Lett.* **2016**, *5*, 498.

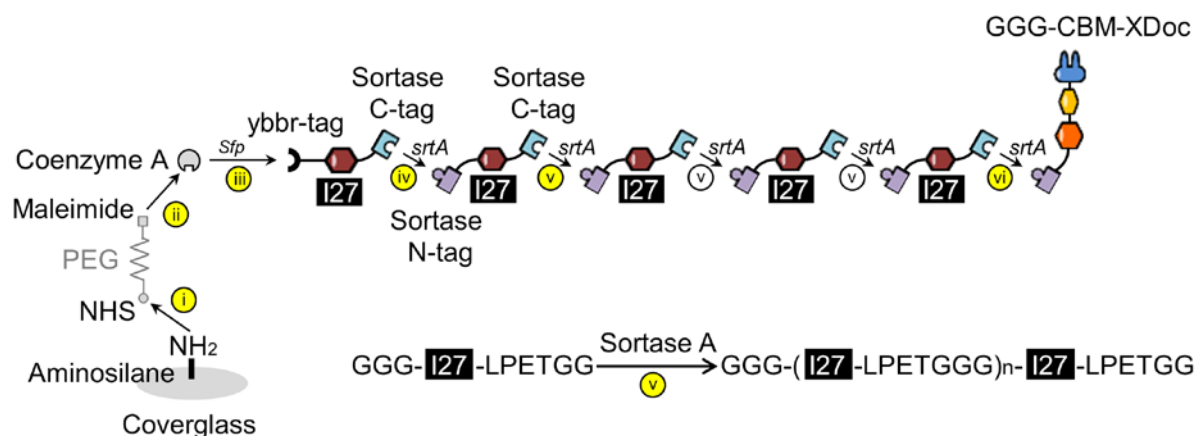


Figure 1. Schematic diagram showing the surface chemistry strategy. (i) Aminosilanized glass was functionalized with NHS-PEG-Maleimide, and (ii) reacted with the free thiol group of Coenzyme A. (iii) An I27 variant bearing an N-terminal ybbR tag was covalently linked through a phosphodiester bond at a conserved serine residue to the CoA surface. (iv, v) Srt mediated oligomerization of an I27 variant bearing both N- and C-terminal Srt tags. (vi) C-terminal capping of the polyprotein using Srt-mediated ligation of GGG-CBM-XDoc.

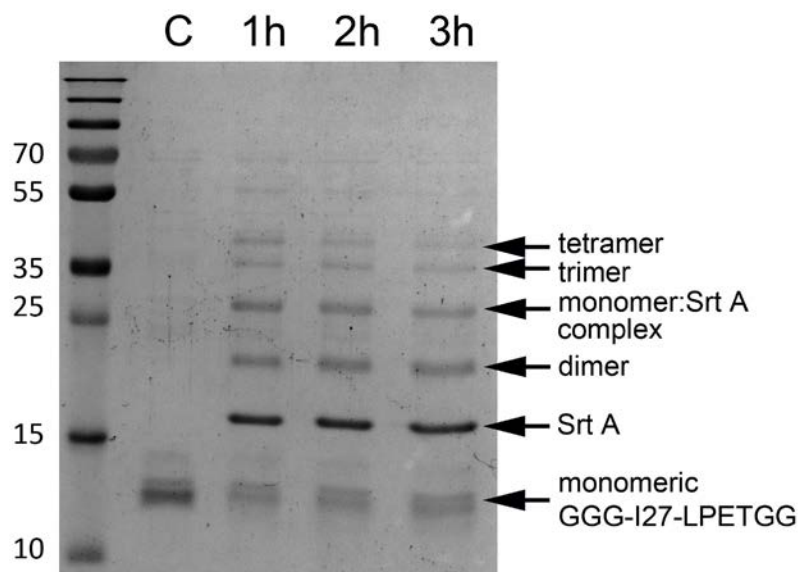


Figure 2. In vitro Srt A-mediated polymerization of GGG-I27-LPETGG showing the formation of dimers and trimers as the most abundant products. Lane C: monomeric GGG-I27-LPETGG loaded as negative control; lane 1h, 2h and 3h: GGG-I27-LPETGG mixed with Srt A after 1 hr, 2 hr and 3 hr reaction time.

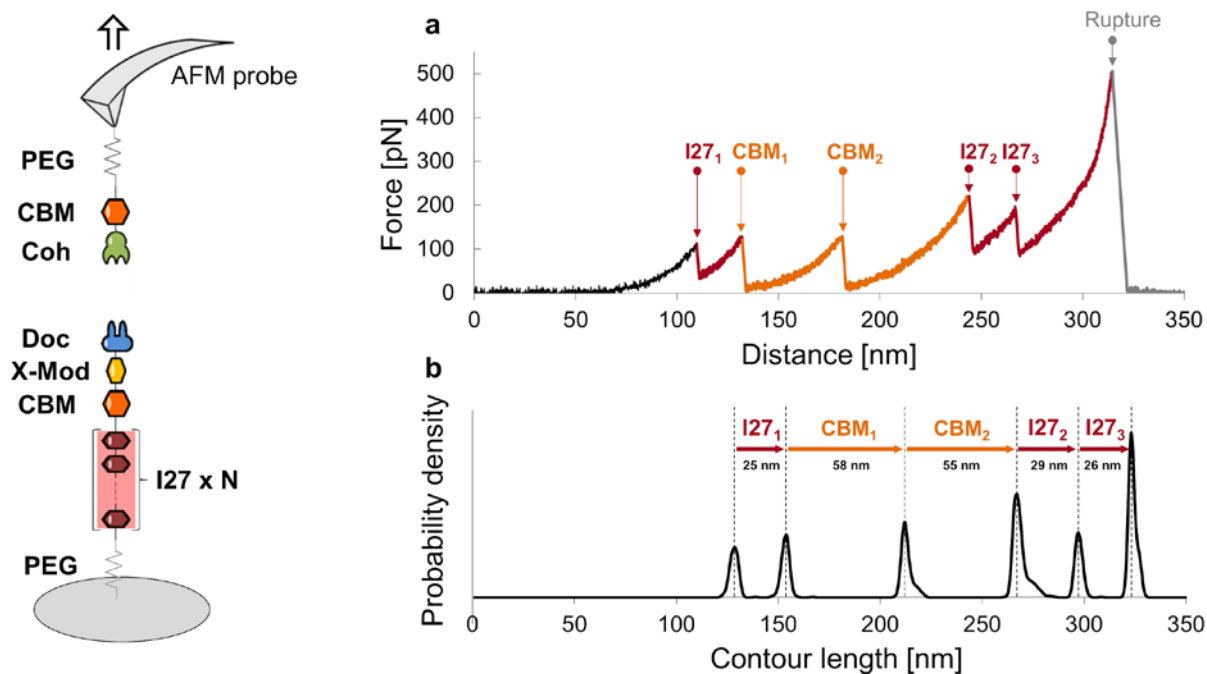


Figure 3. AFM-SMFS unfolding trace of an I27 polyprotein capped at the C-terminal end with CBM-XMod-Doc and probed with a Coh-CBM modified cantilever. (a) Typical force vs. distance trace for a polyprotein containing three copies of I27 and two CBMs. The final peak corresponds to the rupture of the Coh-Doc complex. (b) The data were transformed and histogrammed in contour length space in order to make domain assignments to the unfolding events.

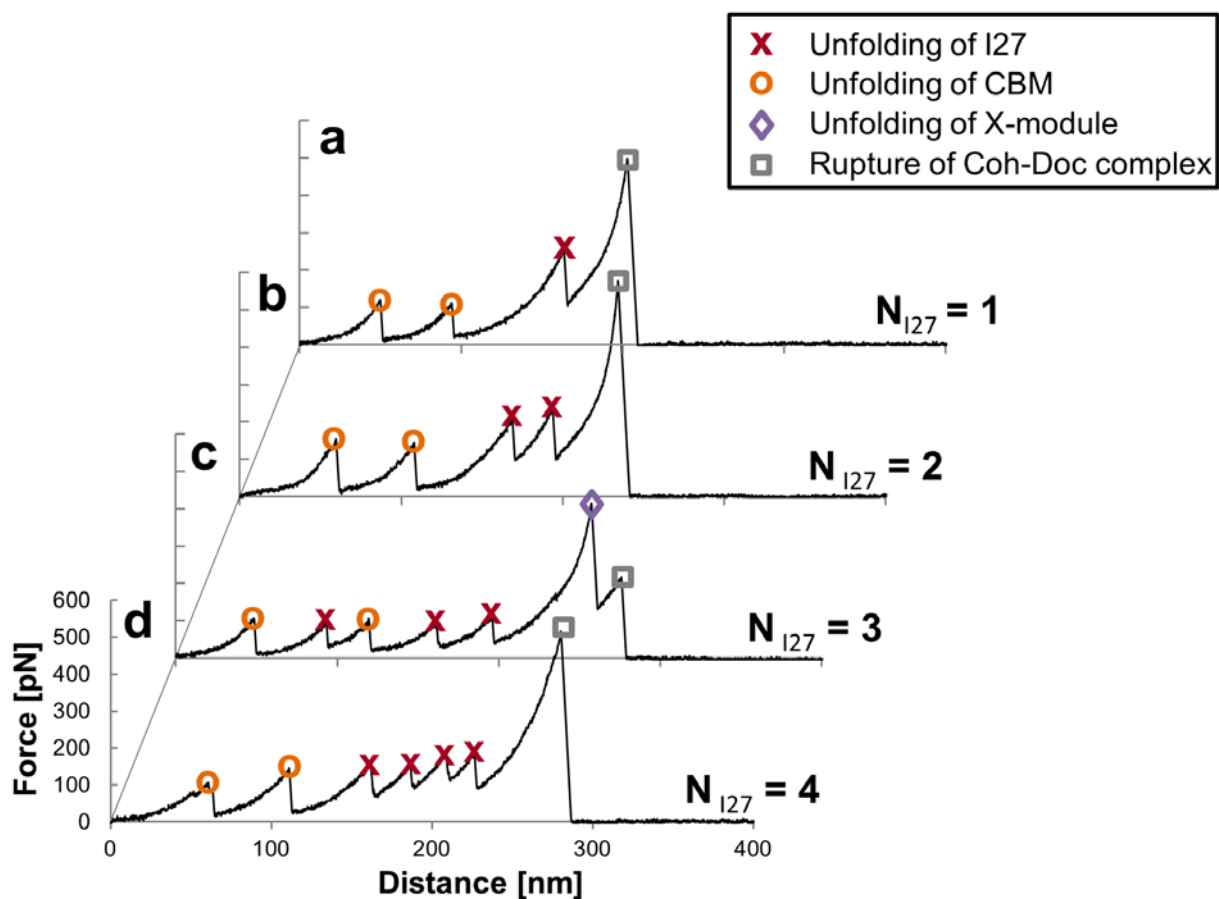


Figure 4. AFM SMFS unfolding traces with numbers of I27 module from 1 to 4.

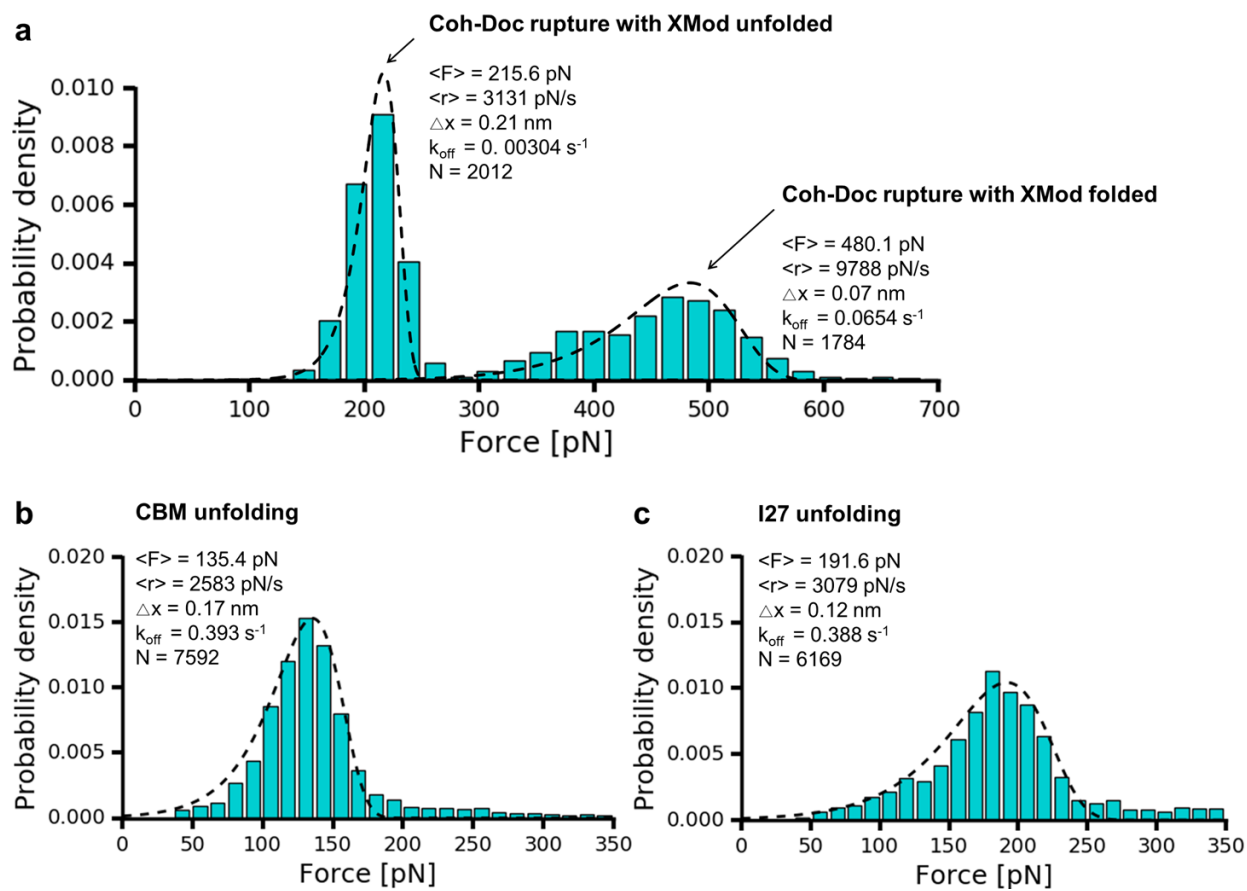


Figure 5. Analysis of rupture forces and unfolding forces. (a) Histogram of rupture force for the final Coh-Doc complex obtained from 3,796 force-extension traces. (b,c) Histogram of unfolding force for (b). CBM and (c) I27 domains.

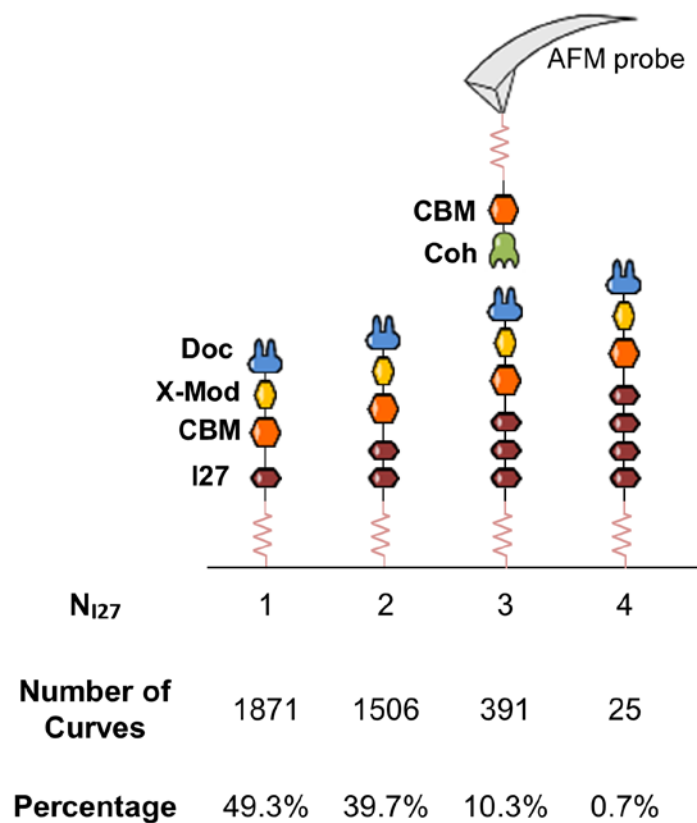


Figure 6. Polydispersity characterization of Sfp/Srt-mediated polyprotein assembly. The number of I27 domain unfolding events in each curve that passed the analysis criterion were counted. In order to be considered for analysis, the curve needed to contain 2 CBM unfolding contour length increments. The majority of curves contained 1-2 I27 unfolding events and 2 CBM events.

Oligomeric polyproteins are used for single-molecule biomechanical experiments to characterize protein folding energy landscapes. Here a new method is presented for posttranslational assembly of polyproteins which relies on two enzymes to achieve site specific surface attachment, multi-domain oligomerization and installation of a mechano-stable receptor for specific protein pickup and stretching under force, resulting in high quality single molecule unfolding datasets.

Keyword: single-molecule force spectroscopy, protein folding, atomic force microscopy, sortase, cohesion and dockerin, polyprotein.

Dr. H. Liu, Dr. D. T. Ta, Prof. Dr. M. A. Nash

H. Liu, D. T. Ta, M. A. Nash*

Mechanical polyprotein assembly using Sfp and sortase-mediated domain oligomerization for single-molecule studies

TOC figure

