Enzymatic biosynthesis and immobilization of polyprotein verified at the single-molecule level

Yibing Deng¹, Tao Wu¹, Mengdi Wang¹, Shengchao Shi¹, Guodong Yuan¹, Xi Li¹, Hanchung Chong²,³, Bin Wu²,³ & Peng Zheng 1

The recent development of chemical and bio-conjugation techniques allows for the engineering of various protein polymers. However, most of the polymerization process is difficult to control. To meet this challenge, we develop an enzymatic procedure to build polyprotein using the combination of a strict protein ligase OaAEP1 (Oldenlandia affinis asparaginyl endopeptidases 1) and a protease TEV (tobacco etch virus). We firstly demonstrate the use of OaAEP1-alone to build a sequence-uncontrolled ubiquitin polyprotein and covalently immobilize the coupled protein on the surface. Then, we construct a poly-metalloprotein, rubredoxin, from the purified monomer. Lastly, we show the feasibility of synthesizing protein polymers with rationally-controlled sequences by the synergy of the ligase and protease, which are verified by protein unfolding using atomic force microscopy-based single-molecule force spectroscopy (AFM-SMFS). Thus, this study provides a strategy for polyprotein engineering and immobilization.

¹State Key Laboratory of Coordination Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, Jiangsu 210023, P. R. China. ²School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore. ³NTU Institute of Structural Biology, Nanyang Technological University, EMB 06-01, 59 Nanyang Drive, Singapore 636921, Singapore. Correspondence and requests for materials should be addressed to P.Z. (email: pengz@nju.edu.cn)
Protein conjugation and polymerization is a natural biochemical process and has important applications in biomaterial and biomedicine engineering. Compared with synthetic polymers, one of the unique features of most biopolymers, such as protein, is a uniform structure with a well-controlled sequence of amino acids, while multi-domain protein consists of similar or different protein subdomains. This clustering of the same or different protein domains often results in enhanced biological function and stability. Several biochemical reaction-based methods, especially the cysteine-based coupling, have been developed for building protein polymer, in which protein monomer is designed with additional cysteines or specific residues as the basic unit for ligation. However, unlike natural multi-domain proteins, it is rarely reported that these polymers and biomaterials are of well-controlled subunit sequence like their natural origin, and a bio-synthetic route for this purpose remains a key challenge. Another approach is to build the complete gene into one open reading frame for the artificial protein oligomer, just like the natural way. For example, a so-called polyprotein strategy has been developed to build protein oligomer to mimic natural modular protein. The fused polyprotein comprises identical or different multiple protein domains whose genes are built using recombinant DNA technology. However, the engineering of toxic or large-sized protein polymer is often challenging. Also, many proteins such as metalloprotein and delicate enzyme may need purification as a monomer. Thus, the application of recombinant DNA technology for building poly-protein is also limited.

To address this challenge, we develop an enzymatic, stepwise construction of protein polymer/polyprotein with a relatively-precise controlled sequence using a protein ligase and a protein protease. OaAEP1 is a recently developed, efficient and strict endopeptidase, which links two peptides/proteins covalently as a peptide bond through two termini in less than 30 min. It requires a ligation unit with only two N-terminal GL residues (NH2-Gly-Leu) and three C-terminal NGL residues (Asn-Gly-Leu-COOH) (Fig. 1a). Thus, based on the ligation unit GL-POI-NGL (POI: protein of interest), OaAEP1 can be used to build polyprotein with an uncontrolled sequence in the solution, similar to the bi-cysteine or other ligase-based coupling methods. The synthesized polyprotein is characterized by the SDS-PAGE gel method, at the ensemble level. Moreover, the polymerization indication the formation of a protein polymer mixture with a varied length under different conditions. The reaction is fast (completed in 20 min) with the formation of protein polymer at least up to a pentamer. The polymerization is robust under acidic pH from 4 to 7, as well as in the presence of most biologically-relevant metal ions, except Hg.

Fig. 1 Protein polymerization by OaAEP1. a Principle of OaAEP1-alone for protein polymerization is based on a GL-POI-NGL protein unit in which an N-terminal NGL (colored in red) reacts with a C-terminal GL (colored in blue) and leads an NGL linkage. b–d The PAGE gel results of ubiquitin polymerization indicate the formation of a protein polymer mixture with a varied length under different conditions. The reaction is fast (completed in 20 min) with the formation of protein polymer at least up to a pentamer (b). The polymerization is robust under acidic pH from 4 to 7 (c), as well as in the presence of most biologically-relevant metal ions (d), except Hg.
designed polymers are unambiguously verified by protein domain unfolding using AFM-based SMFS, at the single-molecule level. AFM-based SMFS can mechanically unfold each protein domain and characterize protein mechanics. The unfolding of polyprotein leads to a characteristic sawtooth-like force-extension curve, in which each force peak corresponds to the unfolding of each protein domain. Thus, the polymerization number of the protein polymer can be directly counted to verify our design. A poly-ubiquitin molecule is built and characterized. Compared with the PAGE gel method, the protein is measured under a native condition at room temperature, and only well-folded protein shows expected stability. Also, any linker effect between ligated proteins can be examined. Consequently, SMFS measurement of the protein polymer not only confirms the polymer design but also provides complementary information about protein stability and folding.

In addition to the polyprotein engineering, OaAEP1 is implemented for protein immobilization for SMFS, together with the specific cohesin-dockerin (Coh-XDoc) receptor-ligand pairs. This configuration ensures the complete unfolding of all protein subdomains. And the polymerization number is counted correctly. Traditionally, the polyprotein sample is deposited on a glass coverslip and is picked up by AFM tip randomly through a non-specific interaction. Although this set up is simple, the pick-up ratio of the high-quality single-molecule event is very low (<0.01%). Based on the development of surface (bio)chemistry methods for covalent anchoring of protein such as thiol chemistry, chloroalkane chemistry, isopeptide bond, click chemistry and specific and reversible receptor-ligand pairs for reversible linkage of sample such as streptavidin-biotin, cohesin-dockerin, and other antigen-antibody interactions, site-specific anchoring and probing configuration for AFM experiment has become possible and is becoming a standard method now.

Our method provides another alternative for such a purpose which can be adopted in all these similar AFM systems. Only two short peptide tags are needed for the ligation, which leads to simple protein preparation without further chemical modification.

Moreover, this OaAEP1 ligase-mediate method is cysteine-free and achieved at the monomeric protein module level. It enables the study of the challenging protein system such as metalloprotein, which needs initial purification as a monomer. Because the overexpression of metalloprotein often results in different metal forms as a mixture, it requires additional purification to obtain a pure-metal form monomer for ligation.

Lastly, we take advantage of a removable TEV protease site, which is compatible with OaAEP1 ligation, to achieve the stepwise protein polymerization on the surface. When a TEV site (ENLYFQ/G) plus a leucine (L) is engineered at the N-terminus of the protein unit as ENLYFQ/G-L-POI, the TEV cleavage results in an N-terminal GL residue as GL-POI, and is compatible with further OaAEP1 ligation. Consequently, an enzymatic, stepwise biosynthesis of polyprotein with a relatively well-controlled sequence is achieved. Thus, our enzymatic method provides a method for polyprotein sample preparation, both sequence-uncontrolled and controlled, as well as protein immobilization for single-molecule studies, especially for the complex metalloprotein system.

Results
A sequence-uncontrolled polyprotein built by OaAEP1-only. Protein ubiquitin (Ub), which has been well characterized by single-molecule AFM before, was chosen for demonstration. The poly-ubiquitin is also a natural signal for protein degradation with biological function. A construct GL-Ub-NGL was built for the OaAEP1 polymerization and reacted in the buffer solution.

The Coomassie-stained SDS-PAGE gel results of the product clearly showed the rapid construction of Ub polymer at least up to a pentamer within 20 min (Fig. 1b). Quantitatively, ~25% of the protein unit was ligated to a dimer, 10% to a trimer, while ~60% of the protein remained a monomer (Fig. 1b). As expected, the polyprotein length was uncontrollable, and the yields diminished rapidly as the chain grew. Nevertheless, its yields and dispersity were comparable to most other protein monomer-based polymerization method, such as the bi-cysteine and sortase-based method.

In addition, we found that OaAEP1 is also a robust ligase under harsh conditions, such as under the acidic solutions and in the presence of metal ions. First, the same ubiquitin polymerization reaction was tested under different pH levels. The SDS-PAGE gel results showed the polymerization was well-performed under acidic condition (pH 4 to 7) (Fig. 1c). Moreover, the reaction was performed in the presence of the most biologically-relevant metal ions. Twelve different metal ions including Fe(III), Co(II), Ni(II), Cu(II), Zn(II), Mn(II), Ca(II), Cd(II), Hg(II), and Pb(II) were tested, respectively. The SDS-PAGE gel results showed that the reaction was not affected by most metal ions at the concentrations of 0.2 mM, except 0.002 mM for Hg (Fig. 1d, Supplementary Fig. 1). These concentrations are much higher than the free metal level in most metalloprotein solutions. These experiments indicate that OaAEP1 ligase is a versatile enzyme that is suitable for constructing challenging protein under harsh conditions.

Next, the polyprotein sample obtained by OaAEP1-only ligation was used directly for single-molecule AFM characterization and study. The sample purified by gel-filtration chromography with higher polymerization degrees was used for better performance, in which Ub tetramer presented most (Supplementary Fig. 2). The protein solution was deposited on a clean glass coverslip and was then pressed and captured by AFM tip (Fig. 2a). Stretching the polyproteins resulted in a typical sawtooth-like force-extension curve with multiple peaks, which corresponded to the unfolding of each ubiquitin monomer (Fig. 2b). For example, seven unfolding peaks were observed in curve 2 of Fig. 2b, indicating that seven ubiquitin units were unfolded. Two key experimental results, the contour length increment (∆Lc) and the force (F) upon unfolding were analyzed, and the results are shown in Fig. 2c. The contour length increment is related to how many proteins residues are unfolded and extended upon protein unfolding. Previous single-molecule AFM unfolding experiments of ubiquitin polyprotein built by the recombinant DNA method showed a ∆Lc of ~24 nm, which was from the full extension of 76 amino acids of Ub (76aa×0.36nm=27 nm, 4 nm is the distance between the N and C termini of the folded Ub), and an average unfolding force of 203 pN. Here, our ligase-built ubiquitin polymer showed comparable results, with an average ∆Lc and standard deviation of 23.1 ± 2 nm, and an average unfolding force and standard deviation of 202 ± 44 pN, number (n) = 198. These results validate our method for building polyprotein for single-molecule measurement. Moreover, it indicates no linker effect from the three new connection NGL residues between protein monomers as the same unfolding forces for Ub measured in our construct. Finally, there was a large dispersity, as many different numbers of ubiquitin unfolding peaks from one to seven were recorded for each molecule. Thus, more experiments were performed to obtain statistics. As shown in the Fig. 2d, most curves showed three Ub unfolding peaks (n = 51, 52%) and four Ub unfolding peaks (n = 26, 27%). The curves with only two or one ubiquitin molecules were not analyzed, as their signals were strongly affected by the non-specific interaction for a short molecule.
It is noted that Ub trimer (52%) was detected most instead of tetramer (27%), which was different from the previous polyprotein mixture identification (Supplementary Fig. 2). Here, the polyprotein sample absorbed on the coverslip was picked up at a random location along the long protein polymer, and by a weak non-specific interaction between the protein and AFM tip or coverslip. It is possible that the full-length protein polymer may not have been completely captured and unfolded. For example, the seven unfolding peaks of Ub observed above (Fig. 2b) cannot guarantee that a ubiquitin heptamer was captured. It was perhaps an octamer, but the tip only pressed on the seventh Ub domain, and one domain was not unfolded and counted. Another possibility is that an octamer was fully captured, while the stability of one domain was higher than the non-specific protein detachment force from either the AFM tip or the glass coverslip. Thus, the domain remained folded when the octamer was detached. Both scenarios have led to force-extension curves detected with fewer unfolding peaks and a large dispersity.

**The covalent immobilization of (poly)protein in the AFM system.** To solve this problem, a covalent attachment of the protein sample in the AFM system with defined immobilization geometry and strong attachment force was developed using OaAEP1 as well. We demonstrated this application based on a strong and reversible type III cohesin-dockerin-Xmodule (Coh-XDoc, or Coh-Doc) receptor-ligand pair developed by Dr. Nash and Dr. Gaub for single-molecule studies. Protein polymer, GL-(POI)$_n$-NGL, as obtained above, was directly used here. First, it was covalently linked to the NH$_2$-Gly-Leu functionalized glass coverslip using OaAEP1 through its C-terminal Asn-Gly-Leu residues. Then, it was ligated with the Coh-NGL through its N-terminus as Coh-(POI)$_n$-NGL-Glass. Finally, it was directly probed by a Protein Marker-XDoc functionalized AFM tip forming a complete force loop: AFM Tip-Protein Marker-(XDoc-Coh)-(POI)$_n$-Glass (Fig. 3a). The experimental details are described in Supplementary Methods. (Ub)$_n$ obtained previously was used directly and measured in this configuration. It showed similar unfolding force and $\Delta L_c$ results to before (Fig. 3b). In contrast, from the statistical analysis, it is clear that protein with a larger polymerization number was detected with higher frequency, and the tetramer presents most now (32%, Fig. 3c, d). This indicates the polyprotein was stretched between the two ends in this configuration. Fifty-two curves were randomly selected for force analysis (Fig. 3e). The force peak with a $\Delta L_c$ of ~55 nm was from the unfolding of the protein marker CBM (cellulose-binding module), and sometimes Xmodule in the XDoc unfolded before the Coh-XDoc complex dissociated, with a $\Delta L_c$ of ~34 nm.

To verify all subdomains were unfolded in this covalent attachment setup, a (Ub)$_n$ polyprotein with a known number of subdomains was built by the recombinant DNA method and tested. Coh-(Ub)$_n$ was used for coverslip attachment, and a CBM-XDoc was used for the tip functionalization. As a result, a polyprotein CBM-(XDoc-Coh)-(Ub)$_n$ including one marker protein CBM and six ubiquitin was used. Most AFM measurement showed an expected full-length polyprotein unfolding scenario with six Ub (292 out of 322, 91%), as demonstrated by the overlap of the force-extension curves with CBM unfolding first. (Fig. 3f, $n=52$). Consequently, this OaAEP1-facilitated covalent protein immobilization method enables the complete unfolding and correct counting of folded subdomains in the polyprotein molecule.

**Pure-metal form polyprotein built from the purified monomer.** Our monomer-based ligation method enables the study of more challenging protein system such as metalloprotein, which sometimes shows several different metal-forms and needs purification as a monomer first. A well-characterized iron-sulfur protein rubredoxin (Rd) with a native Fe(Cys)$_4$ metal center was chosen for demonstration. The overexpression of rubredoxin in *E. coli*
results in a mixture of native iron-form and zinc-substituted form, whose polyprotein built by the classic recombinant DNA method lead to a mixed and uncontrolled metal form (Fig. 4a)44. Instead as a monomer, the protein mixture solution can be separated into pure Fe-Rd and pure Zn-Rd firstly using ion-exchange chromatography (Fig. 4b). The purity was confirmed by their characteristic UV-Vis spectra, respectively (Fig. 4c). As a result, the pure-metal form Rd can be polymerized to poly-CBM unfolding method (polyprotein. Source data of Fig. 3e are provided as a Source Data by their characteristic UV-Vis spectra, respectively (Fig.4c). As a result, the pure-metal form Rd can be polymerized to poly-

![Diagram](https://example.com/diagram.png)

**Fig. 3** The OaAEP1-dependent polyprotein immobilization verified by single-molecule AFM. a The schematic shows that OaAEP1 facilitates a covalent anchoring of (poly)protein on a glass surface as well as a specifically probing by an XDoc-functionalized AFM tip. b Representative curves of (Ub)n, in the specific attachment showed a high detachment force from the Coh-XDoc interaction and more unfolding peaks. c, d The statistical analysis of the (Ub)n indicated the number of curves with a specific number of ubiquitin peaks. Most curves showed the number of Ub unfolding peaks between three and six (>90%). e The scatter plot shows the relationship between unfolding/rupture force and ∆Lc for (Ub)n, Xmodule of the XDoc (∆Lc = 35 nm), and the CBM. f The overlay of all curves with CBM unfolding first show the complete unfolding of the polyprotein (Ub)n using the OaAEP1-dependent polyprotein immobilization method (n = 52). The covalent immobilization with a high detachment force enables the complete unfolding of all subdomains in the polyprotein. Source data of Fig. 3e are provided as a Source Data file.

Sequence-controlled polyproteins built on the glass surface. To rationally control the sequence of the polyprotein built by OaAEP1, we first validated in principle that the TEV cleavage site is compatible with our OaAEP1 ligation system for a stepwise protein polymerization. As shown in the schematic of Fig. 5a, when a TEV site (ENLYFQ/G) plus a leucine (L) was added at the N-terminal part of a protein, the TEV protease cleavage results in an N-terminal Gly-Leu. As a result, the cleaved protein was then compatible with further OaAEP1 ligation and ultimately led to a protein polymer. Ubiquitin was used here for demonstration again. The SDS-PAGE gel result proved the validity of such a procedure for the construction of ubiquitin dimer based on Coh-tev-L-Ub (for cleavage) and Coh-tev-L-Ub-NGL (for ligation) (Fig. 5b).

However, the ligation efficiency for OaAEP1 was not sufficiently high enough for efficient polymerization. Previously, the OaAEP1-only ligation resulted in Ub dimer formation with a yield of 25%. Here, for a one-step ligation, the efficiency was of 20%. Although this is satisfactory for the one-step protein labeling or a dimer construction, it is still challenging to build a relatively long protein oligomer/polymer. The yield decreases exponentially after several rounds of reaction. Consequently, it is necessary to increase ligation efficiency.

For a chemical reaction, when one reactant is added in excess, the chemical equilibrium will be pushed toward the product. Thus, we modified the ratio between the two reactants to increase the yield. The one-step ligation between Coh-tev-L-Ub-NGL and GL-Ub was used here for the test. When the two reactants were in an equal molar ratio, the ligation efficiency was 20%.

![Diagram](https://example.com/diagram.png)
By increasing their ratio from 1 to 3, the yield improved to 40%, and 50% at a ratio of 5. Finally, when a ratio of 10 to 1 was used, the efficiency increased to 75% (Fig. 5c). Here, the ligation efficiency is calculated based on the formation of the product, Coh-tev-L-(Ub)2 dimer, whose band color in the gel intensifies clearly as the reactant ratio increases. To resolve between the reactant monomer Coh-tev-L-Ub-NGL and the product dimer, the sample migrates for a long time. As a result, the band of the other smaller reactant monomer GL-Ub is unclear upon the long migration. By increasing the concentration ten times for all the samples, the decrease of this monomer can also be detected and used for the efficiency calculation with similar results (Supplementary Fig. 3). Consequently, the yield (75%) for OaAEP1-ligation is obtained and is enough to build a polyprotein.

We then built the protein polymer on a functionalized glass surface based on the stepwise OaAEP1 ligation and TEV cleavage reaction (Fig. 6a). First, the C-terminus of the protein unit was linked to the glass surface which naturally protects the protein unit from the sequence-uncontrolled polymerization. In addition, it allows the removal of excessive protein monomer and enzymes by simple buffer washing. Moreover, the resultant protein copolymer (Ub-Rd)3 was also built by adding the protein units Ub and Rd one by one. Qualitatively, protein polymer cohesin incorporated here was for single-molecule AFM measurement using an XDoc-functionalized tip. The AFM results of (Ub)1 showed the corresponding unfolding event with only one Ub peak (Number, N = 54). Then, (Ub)1 was cleaved by protease TEV as GL-(Ub)1-NGL-Glass to expose the GL residues for the second-round ligation. By adding the protein unit Coh-tev-L-Ub-NGL in excess with OaAEP1, Coh-tev-L-(Ub)2-NGL-Glass as a ubiquitin dimer was obtained. (Ub)2 was also characterized by single-molecule AFM, showing the corresponding dimer formation (N = 155, 88%). By repeating this stepwise cleavage and ligation procedure, we obtained ubiquitin polymer up to (Ub)5. The AFM experiments showed that 31% (N = 83) of curves had five Ub unfolding peaks (Fig. 6c). Through the whole process, the same polyprotein sample was characterized at each polymerization stage using the same AFM tip, and the maximum unfolding peaks of polymer picked up corresponded to the desired length (Fig. 6b). Nevertheless, there was still a fraction of molecules that were not reacted, as the ligation efficiency was not 100%. Indeed, when the polymer grew to hexamer, the yield became even lower (Supplementary Fig. 4).

Similarly, metalloprotein Rd was constructed as a sequence-controlled polyprotein on the glass surface and characterized. It demonstrated the feasibility of this stepwise method for constructing metalloprotein polymer (Fig. 6b). Finally, the protein copolymer (Ub-Rd)2 was also built by adding the protein units Ub and Rd one by one. Qualitatively, protein polymer
is both thermally and mechanically stable, proved by the joint SDS-PAGE gel and single-molecule force spectroscopy experiments.

By adding one reactant in excess, the ligation efficiency increases from 20 to 75%. This allows for the construction of a relatively long protein oligomer. Statistically, pentameric protein, such as (Ub)$_5$, can be obtained on the glass surface with a small polydispersity and percentage of ~30%. And a long protein decamer (Ub-RD)$_5$ can be obtained but only with several molecules under current ligation efficiency. Further improvement of the ligation efficiency or a new method to purify the resultant protein is necessary to achieve such a long protein polymer. Nevertheless, these results are already an advance compared with other ligase-based protein ligation method. Previous protein ligase-dependent polymerization seldom reported the construction of a protein pentamer, and the yield for a protein tetramer was ~1% at the single-molecule level. Thus, our method provides a tool for biotechnology and protein engineering, which is suitable for both protein coupling and immobilization.

There are other enzymatic methods which are well-suited for linking protein as a dimer or for protein labeling, such as sortase, split intein, bultease, and SpyCather-SpyTag. Sortase is another similar peptidase which is promising for building long protein polymer. Several groups have used it for single-molecule AFM studies, both for protein covalent attachment and polypeptide construction, with important discoveries. Unfortunately, the wild-type sortase is not a strict ligase and can hydrolyze the ligated linker itself. The engineering for a better sortase is under intense study with substantial improvements being developed. The SpoonTag/SpoonCatcher system based on the SpyCatcher-SpyTag system is another powerful ligation method with a high ligation efficiency of >95% by forming an intermolecular isopeptide formation. Thus, a long protein polymer like a decamer can be obtained in a similar, stepwise fashion with a high yield. By comparison, our OaAEP1 requires two and three additional amino acids at the two ends, and a short three amino-acid length linkage is present after ligation. Thus, either of these two methods may be applied based on different requirements.

From the perspective of single-molecule study, our method also provides a way to construct the polypeptide sample with a better-controlled length using monomeric protein unit. Recombinant DNA technology was employed to build polypeptide for single-molecule AFM study, which composes of identical or different multiple protein domains. Thus, it results in characteristic sawtooth like force-extension curves from the stepwise unfolding of each domain under force. This strategy significantly increases the data reliability and collection efficiency and has become the gold standard for AFM-based SMFS. However, it relies on repetitive cloning cycles and is time-consuming. Recently, a much more efficient Gibson assembly-based method was developed for building long polypeptide genes. Nevertheless, the engineering of toxic or large-sized protein polymer is often challenging, as the misfolding or the inability to express a large-size protein often occurs, and limits its application.

Another simple approach is to build the polymer at the monomeric protein level, by expressing individual protein monomers first and then conjugating them as a polypeptide. A bicysteine based protein coupling method was first developed and is widely used. By engineering two cysteines on a single protein, the formation of the intermolecular disulfide bond enables the construction of polypeptide and has led to many important discoveries. This approach is efficient and enables the construction of larger protein like poly-GFP. A modified method was later developed consisting of a reducing-resist thiol-ether bond formation using maleimide-thiol chemistry. Recently, similar
ligase-based polymerization methods using sortase were also developed\(^\text{12,55}\). All these methods enable the pre-purification of protein monomer before polymerization. Thus, they are suitable for building complex protein such as metalloprotein. However, these polymerization methods all suffer from a sequence-uncontrolled protein polymer with a large dispersity. The protein unit is also linked in a mixed geometry of head-to-tail and head-to-head. For the sortase-based method, a long protein unit is also linked in a mixed geometry of head-to-tail and head-to-head. For the sortase-based method, a long protein polymer was difficult to obtain due to the low ligation efficiency\(^\text{12,55}\). Here, our method utilizes a cysteine-free ligase OaAEP1 for protein ligation, polymerization, and immobilization with relatively high yield.

In conclusion, we develop an enzymatic method to synthesize both sequence-controlled and non-controlled polymerized protein. The robustness and high efficiency of the ligase enable the engineering and study of a wide range of protein, both delicate and complex, as well as providing an efficient way for both protein sample coupling and immobilization for single-molecule studies.

**Methods**

**Protein engineering.** The gene coding for protein of interest: ubiquitin from human (Ub), rubredoxins (Rd, Rd represents Zn-formed rubredoxin from *Clostridium pasteurianum*, if not specified, Fe-form rubredoxin from *Pyrococcus furiosus*), the B1 domain of immunoglobulin G (GB1), the cellulose-binding module (CBM), type III cohesin-dockerin-X module domain complex from *Ruminococcus flavefaciens* (Coh-Xmodule-Doc, or Coh-XDoc), tobacco etch virus (TEV) protease (fused with superfold GFP as GFP-TEV for use), elastin-like polypeptides (ELP) were ordered from Genscript Inc, respectively. Regular PCR procedure was used for the further addition of N-terminal GL and C-terminal NGL to the protein if needed. Typically, a three-restriction digestion enzyme system BamHI-BglII-KpnI was used for connecting the gene of different protein fragments\(^\text{2}\). The same overhang after BamHI and BglII digestion allows the stepwise ligation between their fragments. All genes were finally confirmed by DNA sequencing from GenScript Inc. Typical protein overexpression and purification procedure, the general method for OaAEP1-only polymerization, and corresponding amino acid sequences are all provided in Supplementary Methods and Notes.

First, the pQE80L-POI or pET28a-POI expression plasmids were transformed into *E. coli* BL21(DE3) cells. Single colonies were picked into LB medium containing 100 μg mL\(^{-1}\) ampicillin sodium salt or 50 μg mL\(^{-1}\) kanamycin (continuous shaking, 37 °C, and 16–20 h). After grown to saturation, overnight cultures were diluted 1:50 into fresh LB media containing ampicillin sodium salt or kanamycin and shifted to 20 °C. The culture was grown under the continuous shaking. After the OD\(^\text{600}\) reached 0.5, 1 mM IPTG was added to the media. After 20 h, the culture was harvested and stored at -80 °C for protein extraction.

**Representative curves with maximum unfolding peaks from the unfolding experiments of homo-polyprotein (Ub)\(_5\), top graph, (Rd)\(_5\), middle graph, and (Ub-Rd)\(_3\), bottom graph, are shown, which were detected during each polymerization step from one to five/six for the same sample.**

**The statistical analysis of the number of the curve with specific unfolding peaks, corresponding to the left curves.**
kanamycin (continuous shaking, 37 °C, t - 3 h, except Fe(III)-Rd, and Zn-Rd which were overexpressed in M9 medium (M9 media supplemented with 0.4% glucose, 0.1 mM CaCl2, 2 mM MgSO4) with continuous shaking 6 h), and induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) based on each protein when OD600 is ~0.6. (For rubredoxin, 1 mM FeCl3 or ZnCl2 is added). The bacterial cultures were allowed to incubate for an additional 4-6 h (37 °C). Finally, 400 mL bacterial culture was pelleted by centrifugation (13,260 × g, 25 min, 4 °C) and stored at ~ 80 °C before purification.

The cells were then resuspended in 25 mL lysis buffer (50 mM Tris, pH 7.4) and lysed on ice using a Biosafer sonicator (15% amplitude for 30 min). The lysate was centrifuged (19,632 × g, 40 min, 4 °C) to pellet cell fragments and the supernatant fluids were applied to a Co-NTA or Ni-NTA affinity column (TALON) and washed with buffer containing 20 mM Tris, 400 mM NaCl, 2 mM imidazole, pH 7.4. The bound protein was eluted with elution buffer (20 mM Tris, 400 mM NaCl, 250 mM imidazole, pH 7.4). For rubredoxin we used an anion exchange chromatography (Mono Q 5/30 GL GE Healthcare) using a continuous salt gradient of 0-30% of buffer B (50 mM Tris, 1 M NaCl, pH 8.5) and then a size-exclusion chromatography (Superdex 200 10/300 GL GE Healthcare) that had been pre-equilibrated in 50 mM Tris, 100 mM NaCl, pH 7.4 buffer in an AKTA FPLC system (GE Healthcare) for further purification to ensure the purity >95%.

Stepwise polypeptide preparation with controlled sequences. Here we used the sample preparation for ubiquitin homo-polymer, Coh-tev-L-(Ub)n-NGL, as an example. Ligation unit Coh-tev-L-Ub-NGL was first linked to the GL-ELP50nm-C functionalized coverslip by OaAEP1 and the C-terminus was blocked. Then, the TEV protease solution was loaded on the TEV sites to cleave the TEV site and the protein unit (0.4 mg mL −1 TEV protease 100 mL, 75 mM NaCl, 0.5 mM EDTA, 25 mM Tris-HCl, pH 8.0, 10% [v/v] glycerol). Typically, it was reacted for ~1 hour at 25 °C to produce GL-UB-NGL-glass with exposed N-terminal GL residues. TEV protease was then washed away. Then, 3 times the amount of the ligation unit, containing the whole information of the protein construct and a high rupture force from the cohesin-dockerin dissociation were selected with a ratio of ~5%.

Data availability
Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available from the corresponding author upon request. The source data underlying Figs. 2c, 3e, 4c, e, f are provided as a Source Data File.

Received: 26 November 2018 Accepted: 23 May 2019
Published online: 24 June 2019

References
1. Yang, Y. J., Holmberg, A. L. & Olsen, B. D. Artificially engineered protein polymers. Annu. Rev. Chem. Biomol. Eng. 8, 549–575 (2017).
2. Pelegri-O’Day, E. M. & Maynard, H. D. Controlled radical polymerization as an enabling approach for the next generation of protein–polymer conjugates. Acc. Chem. Res. 49, 1777–1785 (2016).
3. Yang, J. et al. Polypeptide strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology. Proc. Natl Acad. Sci. USA 115, E8509–E8517 (2018).
4. Dietz, H. et al. Cysteine engineering of polypeptides for single-molecule force microscopy. Nat. Protoc. 1, 80–84 (2006).
5. Albury, C. & Swartz, J. R. Direct polymerization of proteins. Science 333, 353–362 (2014).
6. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276, 1109–1112 (1997).
7. Carrion-Vazquez, M. et al. Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl Acad. Sci. USA 96, 3694–3699 (1999).
8. Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polypeptides. Chem. Soc. Rev. 41, 4781–4796 (2012).
9. Yang, R. et al. Engineering a catalytically efficient recombinant protein ligase. J. Am. Chem. Soc. 139, 5351–5358 (2017).
10. Harris, K. S. et al. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat. Commun. 6, 10199 (2015).
11. Jackman, T. et al. Molecular control over the properties of polypeptides by plant asparaginyl endopeptidases. Nat. Commun. 9, 2411 (2018).
12. Liu, H. P., Ta, D. T. & Nash, M. A. Mechanical polypeptide assembly using sfp and sortase-mediated domain oligomerization for single-molecule studies. Small Methods 2, 39–45 (2018).
13. Zhang, X., Rice, F., Xu, A. & Moy, V. In Handbook of Single-Molecule Biophysics (eds Hinterdorfer, P. & Oijen, A.). (Springer US, 2009).
14. Beedle, A. E. M. et al. Forcing the reversibility of a mechanochemical reaction. Nat. Commun. 9, 3155 (2018).
15. Alonso-Caballero, A. et al. Mechanical architecture and folding of E. coli type 1 pilus domains. Nat. Commun. 9, 2758 (2018).

16. Walder, R. et al. High-precision single-molecule characterization of the folding of an HIV RNA hairpin by atomic force microscopy. Nano Lett. 18, 6318–6325 (2018).

17. Takahashi, H., Rico, F., Chipot, C. & Scheuring, S. alpha-helix unwinding as force buffer in spectrins. ACS Nano 12, 2719–2727 (2018).

18. Thirumalai, D., Sapra, K. T. & Müller, D. J. Single-molecule force spectroscopy of transmembrane β-barrel proteins. Annu. Rev. Anal. Chem. 11, 375–395 (2018).

19. Zhu, R. et al. Nanopharmacological force sensing to reveal allosteric coupling in transporter binding sites. Angew. Chem. Int. Ed. 55, 1719–1722 (2016).

20. Liu, C. L., Shi, W. Q., Cui, S. X., Wang, Z. Q. & Zhang, X. Force spectroscopy of polymers: beyond single chain mechanics. Curr. Opin. Solid State Mat. Sci. 9, 140–148 (2005).

21. Valbuena, A. et al. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc. Natl Acad. Sci. USA 106, 13791–13796 (2009).

22. Mudassir, M. et al. Single-molecule force-unfolding of titin I27 reveals a correlation between the size of the surrounding anions and its mechanical stability. Chem. Commun. 54, 9635–9638 (2018).

23. Aioanei, D. et al. Single-molecule-level evidence for the osmophobic effect. Nat. Struct. Biol. 10, 738 (2003).

24. Garcia-Manyes, S., Brujic, J., Padilla, C. L. & Fernandez, J. M. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. Biophys. J. 93, 2436–2446 (2007).

25. Durner, E., Ott, W., Nash, M. A. & Gaub, H. E. Post-translational sortase-mediated attachment of high-strength force spectroscopy handles. ACS Omega 2, 3064–3069 (2017).

26. Zhang, P. & Li, H. Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. J. Am. Chem. Soc. 133, 6791–6798 (2011).

27. Garg, S., Singaraju, G. S., Yengkhom, S. & Rakshit, S. Tailored polyproteins using sequential staple and cut. Bioconjug. Chem. 29, 1714–1719 (2018).

28. Carrion-Vazquez, M. et al. The mechanical stability of ubiquitin is linkage dependent. Nat. Struct. Mol. Biol. 20, 1003 (2013).

29. Garcia-Manyes, S., Brujic, J., Padilla, C. L. & Fernandez, J. M. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. Biophys. J. 93, 2436–2446 (2007).

30. Dorn, B. M., Ham, H. O., An, C. H., Chaikoff, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. Proc. Natl Acad. Sci. USA 111, 13343–13348 (2014).

31. Veggiani, G. et al. Programmable polyproteins built using twin peptide locking and enables solid-phase conjugate isolation. J. Am. Chem. Soc. 140, 3008–3018 (2018).

32. Hoffmann, T. et al. Rapid and robust polyprotein production facilitates single-molecule mechanical characterization of beta-barrel assembly machinery polyproteptide transport associated domains. ACS Nano 9, 8811–8821 (2015).

33. Crampton, N. & Brockwell, D. J. Unravelling the design principles for single-protein mechanical strength. Curr. Opin. Struct. Biol. 20, 508–517 (2010).

34. Dietz, H., Berkemeier, F., Bertz, M. & Rief, M. Anisotropic deformation response of single protein molecules. Proc. Natl Acad. Sci. USA 103, 12724–12728 (2006).

35. Zheng, P., Cao, Y. & Li, H. Facile method of constructing polyproteins for single-molecule force spectroscopy studies. Langmuir 27, 5713–5718 (2011).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 21777103), Natural Science Foundation of Jiangsu Province (Grant No. BK20160639), Fundamental Research Funds for the Central Universities (Grant No. 14380171) and Shuangchuang Program of Jiangsu Province for P.Z., B.W. was supported by Minister of Singapore, Tier 1 Grant (2017-T1-001-168).

Author contributions

P.Z. designed the research, Y.B.D. and T.W. performed the experiments, Y.B.D., T.W., M.D.W., G.D.Y., S.C.S., and X.L. analyzed the data and prepared the figures and P.Z., Y.B.D., T.W., M.D.W., G.D.Y., H.C.C., and B.W. wrote the manuscript.
