

Supporting Information

Instantaneous splice and excise of inteins to synthesize polyproteins on a substrate with tunable linkers

Veerpal Kaur¹, Surbhi Garg¹, Sabyasachi Rakshit^{1,2*}

¹Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India, 140306

²Centre for Protein Science Design and Engineering, Indian Institute of Science Education and Research Mohali, Punjab, India, 140306

Table of Contents

Supplementary Figures 1 to 4

Index

Material and Methods

Figure S1. Fluorescence intensity profile showing the sequential polyprotein growth.

Figure S2: Verification of intein-mediated polyprotein synthesis using AFM unfolding experiments.

Figure S3: Bar plot showing the frequency of events.

Figure S4: Comparison of unfolding forces reported for (I27)₈ synthesized using SE, DC, and ES.

Material and Methods

Plasmid Construction

pSABAD14-111(Plasmid#61824), pSARSF53-259(Plasmid#61757), pTWIN1-His6-Ssp-Httex1-43Q-A2C-A60C (Plasmid #114623) plasmids having genes encoding SspDnaB intein and NpuDnaB inteins, respectively, are ordered from Addgene and used throughout the study. Overlap PCR (Polymerase Chain Reaction) protocol is used to assemble split inteins, I27, MBP, and TEV protease tag site in four different constructs as per construct design (Figure 1). Modified pMal-c2X vector (having restriction site XhoI) double digested with EcoRI-HF and XhoI restriction enzyme is used as plasmid backbone. The main motive of using the pMal-c2X vector is to attach the MBP tag to proteins on N-terminals.

After amplifying the SspDnaB^C encoding gene from pTWIN1-His6-Ssp-Httex1-43Q-A2C-A60C plasmid, we did its overlap PCR with I27 having LPETGGG on its C-terminal, the resulting gene sequence is used as insert in Gibson assembly and modified pMAL-c2X (having XhoI site) as a vector to get MBP-TEV- SspDnaB^C-I27-LPETGGG (construct a). For the construction of MBP-TEV-NpuDnaB^C-I27-SspDnaB^N (construct b) genes NpuDnaB^C, SspDnaB^N are PCR amplified from pSARSF53-259 and pTWIN1-His6-Ssp-Httex1-43Q-A2C-A60C plasmids, respectively. I27 template is PCR amplified from the plasmid in the lab. Then overlap PCR of all three segments is done. The resulting amplified PCR product is double digested with EcoRI-HF, and XhoI enzymes then ligated into modified pMAL-c2X (having inserted XhoI site and native EcoRI site) vector using Gibson assembly. TEV cut site is added using primer and MBP tag in the aid of vector backbone. For the construction of MBP-TEV-SspDnaB^C-I27-NpuDnaB^N (construct c), we use our constructed plasmid pMAL-c2X. SspDnaB^C-I27-LPETGGG (construct a), double digest it with BamHI-HF and XhoI to remove I27-LPETGGG sequence and ligate it with insert I27-NpuDnaB^N.

SspDnaB^N mini-intein encoding gene sequence is PCR amplified from pTWIN1-His6-Ssp-Httex1-43Q-A2C-A60C plasmid and ligated into modified pMAL-c2X vector after doing its overlap PCR with I27 to get MBP-TEV-I27-SspDnaB^N (construct d).

All the plasmids are confirmed by sanger sequencing.

Proteins Expression and Purification

a) Inteins

The expression plasmids containing desired four protein-encoding sequences are transformed in Lemo21(DE3) competent cells. Single colonies of transformed cells are picked up for primary inoculation, then the secondary culture is grown overnight in LB media (having 1% glucose, 750 μM Rhamanose, 100 μg/mL Ampicillin, 34 μg/mL Chloramphenicol) with continuing Shaking at 37°C till OD₆₀₀ reaches 0.5, then 0.4 mM IPTG is used to induce production of recombinant proteins, and after induction cell culture is allowed to grow for next 4 hours at 37°C continuously. After centrifugation, cell pellets are resuspended and lysed by sonication in lysis buffer (50mM HEPES, 150mM NaCl, 100mM KCl, 2mM CaCl₂, pH 7.5). The supernatant containing soluble fraction after centrifugation passed through the Amylose resin column as all our proteins have MBP tag and eluted from the resin with buffer +20mM maltose and dialyzed. The purity of proteins is checked using 12% SDS-PAGE.

b) TEV

Primary inoculation of TEV Protease Rosetta cells is done in LB media with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol antibiotics. The secondary culture is grown in LB media (having 100 μg/mL ampicillin, 34 μg/mL chloramphenicol) at 37°C till OD₆₀₀ reaches 0.4. Overexpression of protein is induced with 0.3mM IPTG for 20 hours at 16°C. Cells are

pellet down at 4°C at 10000g for 30 minutes. Resuspend the pellet in lysis buffer (25mM HEPES, 20 mM KCl, 150 mM NaCl, 20mM MgCl₂, pH 7.4). Pellet is lysed with lysozyme (2mg/g of the pellet) and PMSF (1mM) while shaking at 4°C for 1h. Cells are sonicated, followed by centrifugation at 4°C at 10000g for 30mins. The supernatant is passed through the pre-equilibrated Ni-NTA column. Column is washed with washing buffer (25mM HEPES, 20 mM KCl, 650 mM NaCl, 20mM MgCl₂, 20mM Imidazole, pH 7.4). TEV Protease is eluted using elution buffer (25mM HEPES, 20mM KCl, 650 mM NaCl, 20mM MgCl₂, 300mM Imidazole, pH 7.4). Protein is dialyzed and run the 12% SDS-PAGE to check the purity of the protein.

Polyglycine exposed surface preparation

After doing plasma cleaning of coverslips, coverslips are kept in piranha solution (3 parts conc. H₂SO₄: 1part 30% H₂O₂) for 2 hours followed by 3 times proper washing with MQ water by sonication for 5 min. Then silanization of coverslips is done with 2% APTES (3-Aminopropyltriethoxy silane) in 95% acetone for 30 minutes followed by washing with acetone and water, then dried for 1h at 110 °C in a vacuum oven. PEGylation of the surfaces is done with PEG buffer (100mM NaHCO₃ and 600mM K₂S₀₄), then PEGylated surfaces are washed and incubated with GGGGC peptide in Ca²⁺ buffer (pH 7.4) for 7h. After washing with MQ surfaces are ready to use.

Polyprotein synthesis on the surface

Polyglycine exposed surface is incubated with protein of interest (POI), I27 here, having N-terminal MBP tag, TEV protease cut site, Int^C(Ssp), and LPETGGG at its C-terminal (1μM) and Sortase (1.25μM) in 4:5 molar ratio to facilitate the formation of a thioester intermediate in 2mM Ca²⁺, 50mM HEPES, 150mM NaCl, 100mM KCl buffer at pH 7.4 in room temperature (RT) for 1 hour, then the surface is washed properly using a buffer to remove unreacted sortase and POI. Then surface having attached I27 and exposed MBP tag is incubated with TEV protease (10μM) for 1 hour at RT for chopping out the MBP tag followed by washing with buffer. In the second step of layering of I27 on the surface, we incubated the surface with another protein POI i.e I27 having MBP tag, TEV protease site, Int^C(Npu) at its N-terminal, and Int^N(Ssp) on C-terminal (1μM) for 5 minutes for trans-splicing reaction of Int^C(Ssp) and Int^N(Ssp) followed by washing with buffer. After proteolytic cleavage of MBP tag from I27 dimer attached to surface, we incubated the surface with our next protein having I27 with MBP tag, TEV protease site, Int^C(Ssp) at its N-terminal and Int^N(Npu) on C-terminal (1μM) for 5 minutes for trans-splicing reaction. These steps are repeated to form polyproteins on a solid surface.

Fluorescence-based imaging experiments using Total internal reflection fluorescence microscope

After tethering the first layer of the protein (1μM) to the surface using sortagging, we incubated the surface with anti-His antibody (1:1500 dilution) (rabbit) for 1 h to probe the 6XHis-tag on the N-terminus of MBP tag. Coverslip is gently washed thrice with buffer after 1h incubation of primary antibody to remove unreacted antibodies. Then the surface is incubated with fluorescently labeled secondary antibody-Alexa488 labeled (1:3000 dilution) for 45min. Again the coverslip is gently washed thrice with buffer after 45min incubation of secondary antibody to remove unreacted antibodies and imaged using TIRF microscope to monitor fluorescence signal. As fusion protein contains 6xHis at the N-terminus so there is a bright fluorescence signal. Then we incubated the surface with TEV protease (10μM) for 1h to chop

the MBP tag and expose the split intein required for the next intein splicing reaction. The coverslip is gently washed thrice with buffer after 1h incubation of TEV protease to remove unreacted protease and imaged using TIRF microscope to monitor fluorescence signal. This TEV protease reaction chop off MBP tag having 6xHis, thus we observed loss of fluorescence signals as expected from successful TEV cut. After completion of one round of intein splice and excise(SE) followed by TEV cut, the process is repeated for the next successive SE and TEV cut until the desired length of polyprotein is achieved. We monitor the extent of TEV cut and intein splicing after every step.

To quantify the growth of intein-mediated polyprotein, we calculated the absolute fluorescence intensity after background subtraction for each intein splicing and TEV cut reaction.

Single-molecule force-spectroscopy experiments

For SMFS using AFM (Nano Wizard 3, JPK Instruments, Berlin, Germany), the protein precursor is immobilized on the glass coverslip and Si₃N₄ cantilever (Olympus, OMCL-TR400PSA-1), using a high-specificity immobilization protocol with 10% bifunctional PEG (NHS-PEG-maleimide, MW 5000 Da) in mono-functional PEG (NHS-PEG, MW 5000 Da). For unfolding experiments, the surface is modified with the construct (a), and then subsequent layers of polyproteins are formed through intein-mediated polyprotein synthesis. The polyprotein is non-specifically approached by the cantilever at 2000nm/s velocity. The interaction between cantilever and polyprotein from the surface is allowed for 0.5s, and then the cantilever is retracted away from the surface with 2000nm/s velocity. Different areas on the surface are scanned. At each area, we have recorded 5625 force-distance curves. The spring constant of the cantilever is 35.5pN/nm measured from the power spectra of thermal noise using thermal fluctuation methods for quantitative estimation of force from each force-distance curve. The analysis is done with home-written MATLAB programs, and force curves are fitted to the Worm-like Chain (WLC) extension model. A histogram of unfolding force and change in contour length for polyprotein are plotted from a non-parametric kernel density estimate where the bin-width is estimated from the Freedman–Diaconis rule using home written MATLAB program for the given equation(1).

$$\text{Bin width} = \frac{2 * IQR}{\sqrt[3]{n}} \quad \text{Equation(1)}$$

Where IQR is the inter-quartile range and n is no. of measurements.

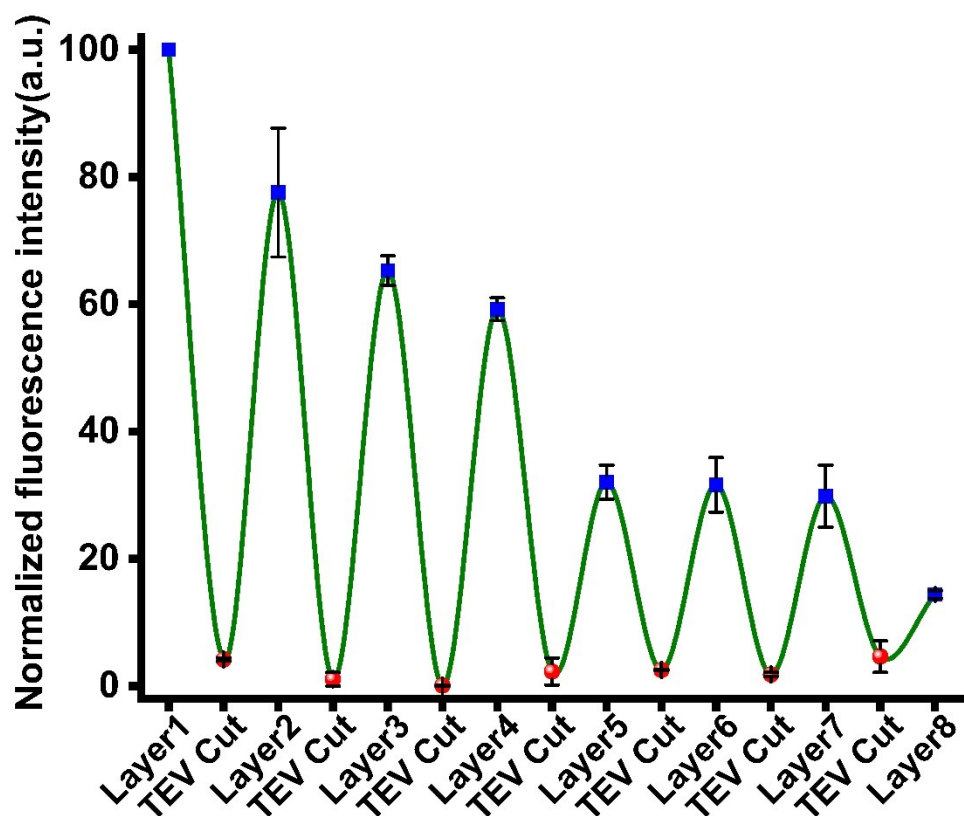


Figure S1. Fluorescence intensity profile showing the sequential polyprotein growth. The change in fluorescence intensity during sequential intein-mediated polyprotein synthesis (blue) followed by TEV cut (red) is shown. The alternative On-Off in the fluorescence signal corresponds to the Intein-splicing and TEV proteolysis, respectively. We also noticed a gradual loss in overall fluorescence intensity in the 'On-state'. The gradual loss in fluorescent intensity with the reaction steps refers to the extent of the reaction synthesizing the octamers of I27.

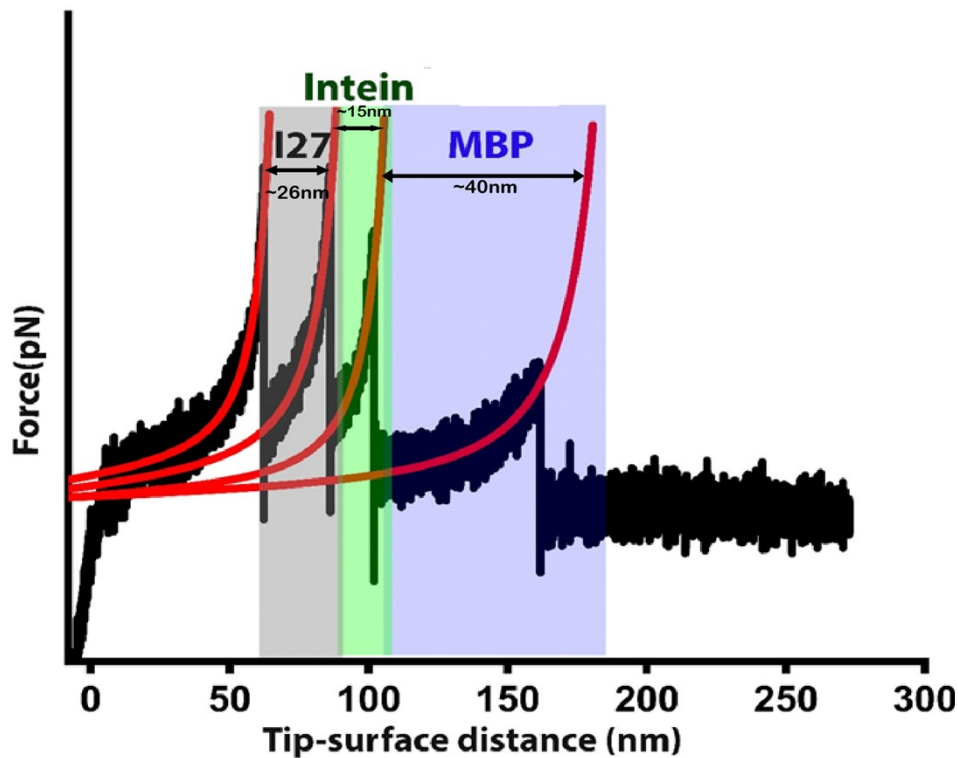


Figure S2. Characteristic force-extension curves of incomplete reaction-products of intein-mediated polyprotein synthesis. The inteins used in polyprotein synthesis possess though high nearly 85% splicing efficiency. This indicates that the step-by-step synthesis of polyproteins using inteins would leave almost 15% unreactive products after every step. Further, the TEV proteolysis is almost 90% efficient, and thus, the number of reactants also drops progressively with increasing steps. Overall, it is expected that step-by-step synthesis of polyproteins will have incomplete by-products containing MBP and inteins in the layers along with I27. As proof, we obtained a fraction of force-extension curves featuring unfolding of MBP and inteins along with I27. We have fit the force-extension peaks to the WLC model and marked the gain in contour lengths for the unfolding of MBP (purple-shadow) and inteins (green-shadow) along with I27 (grey-shadow). The contour-length gains for MBP and I27 are 40.0 ± 0.5 nm and 25.9 ± 0.7 nm, respectively whereas, for inteins is 15.0 ± 0.2 nm.

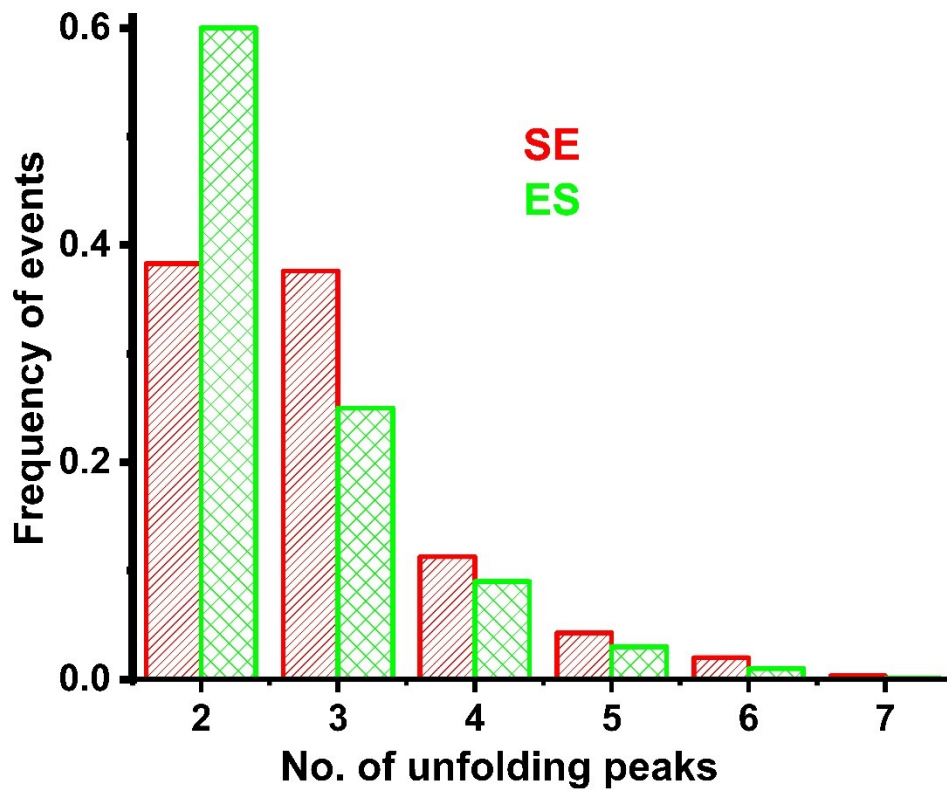


Figure S3. Bar plot showing the frequency of events. Frequency of occurrence of 2 to 7 unfolding peaks of polyprotein obtained from the *in vitro* enzyme-mediated conjugations i.e. by enzymatic staple and cut (ES) (green) and using splice and excise method (SE) (red).

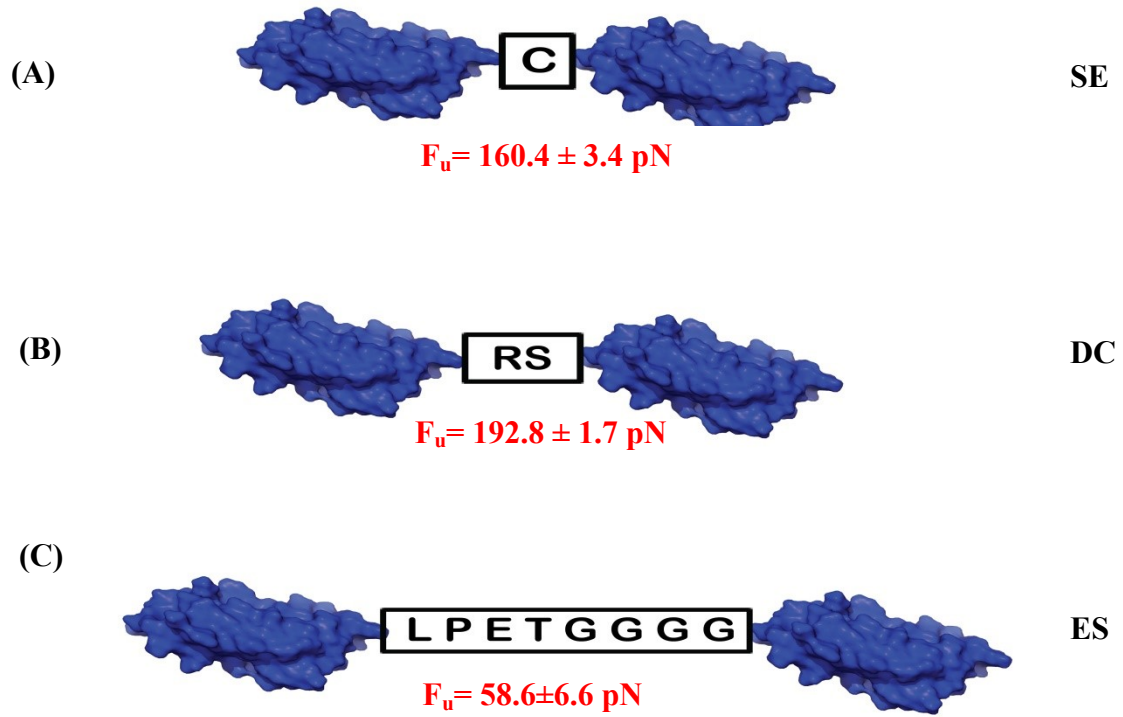


Figure S4. Schematic comparisons of unfolding forces reported for (I27)₈ synthesized using (A) splice and excise method (SE), (B) DNA concatemerization method (DC), (C) enzymatic method i.e., enzymatic staple and cut method.