Supplementary Information

Verification of Sortase for Protein Conjugation by Single-Molecule Force Spectroscopy and Molecular Dynamics Simulations

Fang Tian, Guoqiang Li, Bin Zheng, Yutong Liu, Shengchao Shi, Yibing Deng, and Peng Zheng*

State Key Laboratory of Coordination Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, Jiangsu, 210023, P. R. China

*Corresponding author: Correspondence and requests for materials should be addressed to P.Z. (Email: pengz@nju.edu.cn)

This Supplementary Information Includes:

Supplementary Methods 3-4

Protein engineering 3

Sortase-mediated I27 conjugation 3

Single-molecule AFM experiment.....3-4

Molecular dynamics simulations.....4

Supplementary Note.....5

Protein sequences 5

Supplementary Figures 5-7

Figure S1. SDS-PAGE results of I27 monomer 5

Figure S2. ΔLc of conjugated I27 unfolding with different linkers.....6

Figure S3. Single-molecule AFM results of marker protein CBM.....6

Figure S4. Representative open and closed conformations of I27 dimer with linker LPETGGGG (4G) from MD simulations 6

Figure S5. The radius of gyration (Rg) for each construct with different linkers as a function of simulations time 7

Figure S6. SMD simulation results from the two closed conformations of the 4G-linked I27. 7

Supplementary Videos 8

Video 1. MD simulations for I27 dimer with linker RS .....

Video 2. MD simulations for I27 dimer with linker LPETG .....

Video 3. MD simulations for I27 dimer with linker LPETGGG.....

Video 4. MD simulations for I27 dimer with linker LPETGGGG. .....

Viedo 5. SMD simulations for the closed conformation of the 4G-linker I27 dimer...

Viedo 6. SMD simulations for open conformation of the 4G-linker I27 dimer .....

Supplementary References 8

Supplementary Methods

Protein engineering. The plasmids encoding protein domain I27, CBM (cellulose-binding module), type III cohesion-dockerin-X module domain complex from *Ruminococcus flavefaciens* were purchased from Genscript. Dr. Peng Cheng from Peking University kindly provides us the plasmid for mgsrtA. It contains nine mutations from the wild-type Sortase A. The protein sequence is provided in the supplementary notes. The genes for the three fused protein monomers, Coh-Tev'-X-I27-LPETGG (X: G, GGG, or GGGG), were obtained by Gibson assembly-based method. Gene for Coh-Tev-$(I27)_4$-LPETGG was obtained using standard molecular biology techniques by a classic three-restriction digestion enzyme system.[1] They were all subcloned into the pQE80L vector and expressed in BL21 (DE3) *E. coli* cells. The cells were cultured in LB medium at 37°C, and protein expression was induced for 6 h by the addition of 1 mM IPTG (final concentration). The cells were collected by centrifugation at 5000 × g for 20 min at 4°C, lysed by high-pressure homogenization. Then, the debris was depleted by further centrifugation at 10000 × g for 40 min. The polyprotein was purified using a Ni^{2+} or Co^{2+} affinity chromatography column. After washing (50 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 7.4), the polyprotein was eluted in the buffer (50 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.4). All three I27 ligation units with different linkers were identified by SDS-PAGE gel electrophoresis (Figure S1).

Sortase-mediated I27 conjugation. The conjugated I27 proteins with different linkers were built using the stepwise ligation and cleavage method.[2-4] A GL-ELP_{20nm} functionalized glass coverslip was used.[2] Here, the ELP functions as a spacer from single-molecule measurement to eliminate the non-specific interaction between AFM tip and surface.

Firstly, the protein ligation unit containing N-terminal LPETGG was incubated with mgSrtA in a 2:1 ratio in Ca^{2+} buffer (50 mM Tris, 150 mM NaCl and 1 mM Ca^{2+}, pH 7.4) on the functionalized glass surface at 25°C for 1 h. The coverslips were washed with Milli-Q water first and then with filtered Ca^{2+} buffer to remove unreacted protein. As a result, the protein monomer was immobilized. Then, we added ~800 nM TEV protease solution (75 mM NaCl, 0.5 mM EDTA, 25 mM Tris, pH 8.0, 10% [v/v] glycerol) to cleave the TEV site. The reaction took 1 h at 25°C and the N-terminal glycine was exposed. Then, TEV protease was washed away with Milli-Q water and buffer. These ligation and cleavage procedures were repeated about six times and polyprotein Coh-Tev'-$(LPETX-I27)_N$-LPETGG on the glass coverslip was obtained. Based on our AFM results, the protein polymerization number ranges from 2 to 4.

Single-molecule AFM. Single-molecule AFM experiments were performed using a Nanowizard4 Atomic Force Microscope (JPK). A Si_3N_4 cantilever (MLCT, Bruker) was used and calibrated in the Tris buffer, yielding a typical spring constant of 50 $pN \cdot nm^{-1}$. All AFM experiments were performed under a constant pulling velocity of 400 $nm \cdot s^{-1}$ at 25°C. For accurate comparison between different constructs, the same cantilever was used for at least two

different constructs in one AFM experiment. The 3G and 4G-linked constructs were measured successfully using the same cantilever for three times and seven independent AFM experiments were performed.

During the AFM experiment, the CBM-XDoc functionalized AFM tip was pressed against the glass to specifically pick up a polyprotein via the reversible XDoc-Coh interaction (Fig. 1, Fig. 2A). By moving the AFM tip vertically, conjugated I27 polyprotein and marker protein CBM were stretched, and each protein was unfolded sequentially. Finally, the molecule was detached from the rupture of the XDoc-Coh interaction, and the tip moved up to another location to pick up another molecule for repeated cycles.

Molecular dynamics simulations. We performed molecular dynamics (MD) simulations on I27 dimer with four different linkers (RS, LPETG, LPETGGG, and LPETGGGG), of which three times were performed for the 4G-linker construct and one time for all other constructs. The protein sequences were submitted to the I-TASSER server[5] and the secondary structure of each protein was obtained. The structure was prepared using VMD[6] and the plugin QwikMD.[7] The MD simulations in the present study were performed employing the NAMD molecular dynamics package.[8] The CHARMM36 force field[9] was used in all MD simulations. Before the MD simulations, all the systems were submitted to an energy minimization protocol for 1000 steps. The MD simulations without constraints were performed with explicit solvent using the TIP3[10] water model in the NpT ensemble. The temperature was maintained at 300.00 K using the Langevin dynamics. The pressure was maintained at 1 atm using the Nosé-Hoover Langevin piston.[11] A distance cut-off of 12.0 Å was applied to short-range, non-bonded interactions, and 10.0 Å to the smothering functions. Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method.[12] Each structure underwent a similar equilibration and minimization procedure. Final simulations were run for at least 100 ns.

To understand the unfolding mechanism between the open and closed conformation of I27, we also performed steered molecular dynamics (SMD) for the construct with the 3G linker (one time) and the 4G liner (three times), respectively. The simulations were performed under a constant velocity, employing a pulling speed at 10 $m \cdot s^{-1}$. The spring constant of the pulling spring was set to 7.0 $kcal \cdot mol^{-1} \cdot Å^{-1}$ (1 $kcal \cdot mol^{-1} = 69.479\ pN \cdot Å^{-1}$). In all simulations, SMD was employed by restraining the position of the C-terminal, while pulling on the N-terminus of the I27 dimer. All analyses of MD trajectories were carried out by employing VMD and its plug-ins.

Supplementary Notes

Protein sequences of Coh-***Tev'***-X-I27-***LPETGG*** (X: **G, GGG or GGGG**):

MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEF
KVSGAEGKYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENN
GNGVFVASGADDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSK
GDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIK
AGEP***ENLYFQX***LIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPL
AASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKSAANLKVKELRS***LPE***
TGG

mgSrtA:

QAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAKENQSLDD
QNISIAGHTFIGRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPT
AVGVLDEQKGKDKQLTLITCDDLNRETGVWETRKILVATEVK

Supplementary Figures

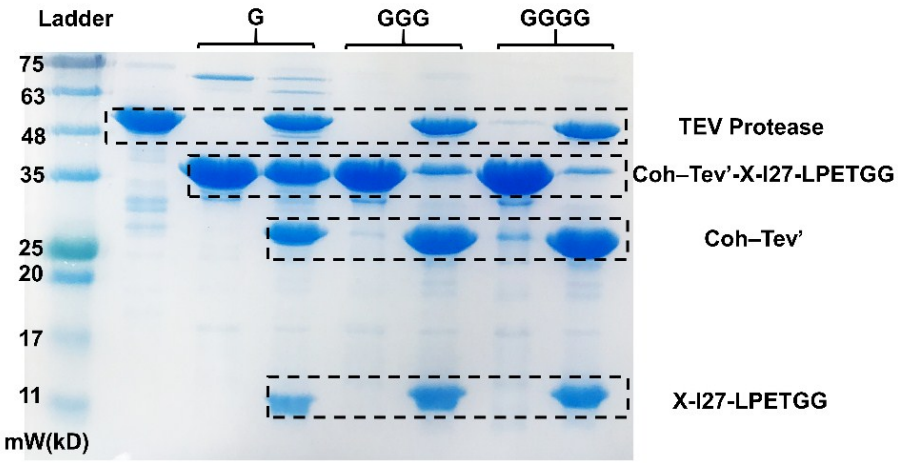

Figure S1. SDS-PAGE gel results for each protein unit before and after TEV cleavage. Lane 1 is the protein ladder. Lane 2 is pure TEV protease (sfGFP-TEV, MW: ~55kDa). Lanes 3, 5, and 7 are Coh-Tev'-G-I27-LPETGG, Coh-Tev'-GGG-I27-LPETGG, and Coh-Tev'-GGGG-I27-LPETGG, respectively (theoretical MW: ~ 43 kDa). Lanes 4, 6, 8 show the protein mixture after TEV cleavage, pure TEV protease (sfGFP-TEV), and cleaved X-I27-LPETGG (MW: ~11 kDa). All ligation units can be cleaved, but with different efficiencies.

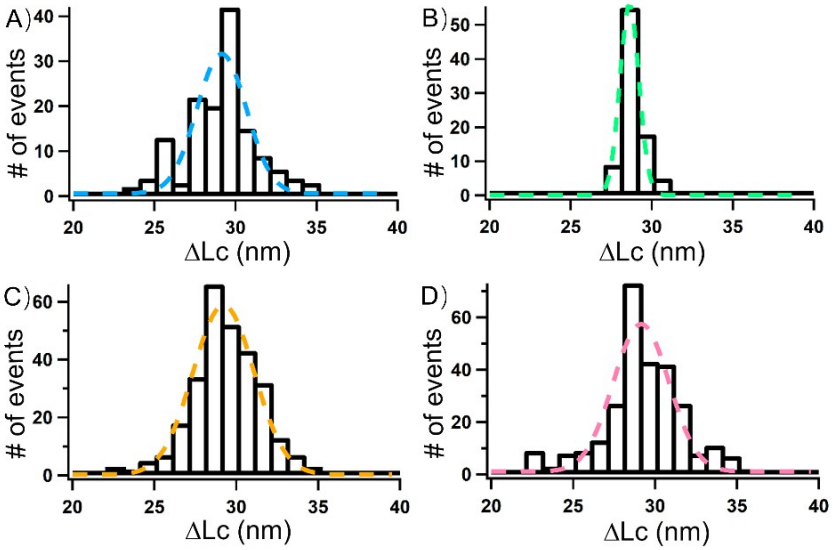

Figure S2. ΔLc results of conjugated I27 unfolding with different linkers. Histograms show the value for I27 polyprotein with linker RS (A), LPETG (B), LPETGGG (C) and LPETGGGG (D) is 28.6 ± 2.3 nm, 28.1 ± 0.7 nm, 28.7 ± 2.1 nm and 28.6 ± 2.5 nm, respectively.

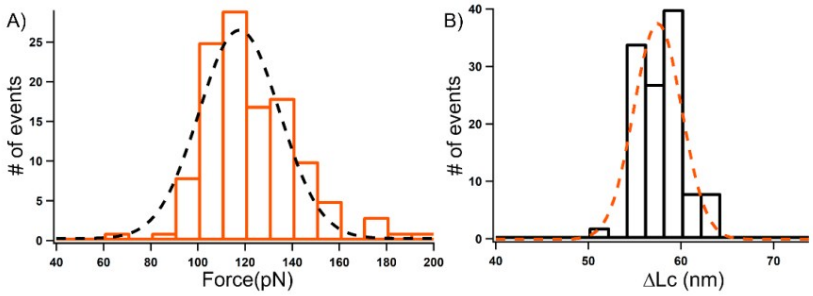

Figure S3. AFM results of the marker protein CBM. (A) The unfolding force of CBM is 112 ± 20.6 pN. (B) The average ΔLc value is 56.4 ± 2.3 nm.

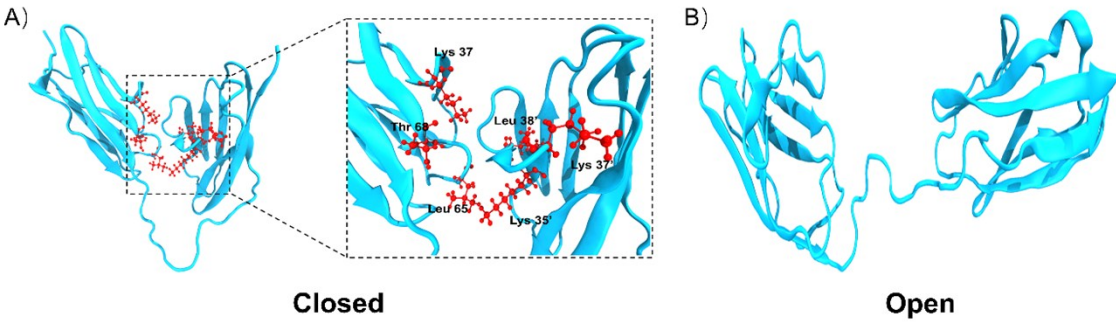

Figure S4. The closed (A) and open (B) conformations of the I27 dimer with the 4G linker from MD simulations used for SMD simulation in the main text are shown. Lys 37, Leu 65, Thr 68, and Lys 35', Lys 37', Leu 38' involved in the inter-domain interaction are highlighted. After 60 ns, the spacing between Leu 65 and Lys 35' decreased substantially, with a distance shorter than 7 Å, whereas Lys 37, Thr 68 and Leu 38' occasionally approached within 7 Å, forming a transient interaction.

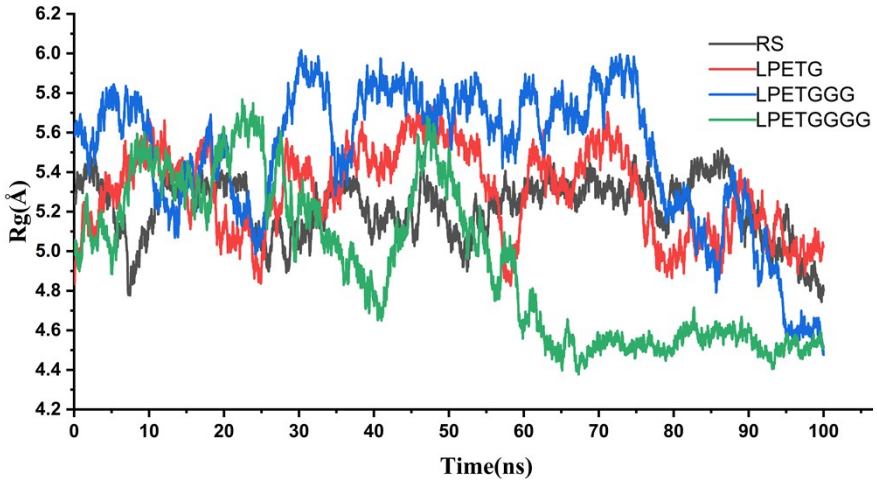

Figure S5. The radius of gyration (Rg) as a function of simulations time for each construct. In general, compared to other structures, the Rg of the trajectory of the 4G-construct became stable and showed a smaller Rg value than those after 60 ns. This result indicates that the I27 dimers with 4G linker appears to twine around each other and formed a protein-protein interaction.

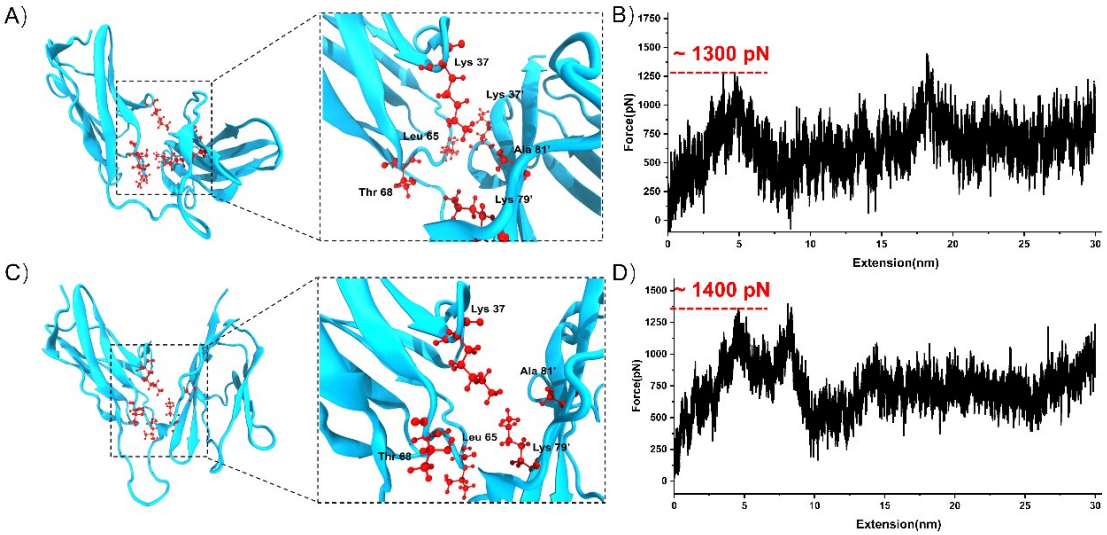

Figure S6. SMD simulation results from the two closed conformations of the 4G-linked I27. A) Residues Lys 37, Leu 65, Thr 68, Lys 79', Ala 81' and Lys 37' involved with the inter-domain hydrophobic interaction are highlighted. Leu 65 and Lys 37' were always present in the interaction. B) Force-extension curve from SMD simulations by stretching the closed conformation from (A). A peak force of ~1300 pN was obtained for the I27 unfolding. C) Residues Lys 37, Leu 65, Thr 68, Lys 79' and Ala 81' involved with the inter-domain hydrophobic are highlighted. Leu 65 and Lys 79' were always present in the interaction. D) Force-extension curve from SMD simulations by stretching the closed conformation from (C). A peak force of ~1300 pN was obtained for the I27 unfolding.

Supplementary Videos

Video 1. 100-ns MD simulations for I27 dimer with linker RS. I27 is colored in cyan, linker RS is colored in red, respectively.

Video 2. 100-ns MD simulations for I27 dimer with linker LPETG. I27 is colored in green, the linker is colored in red, respectively.

Video 3. 100-ns MD simulations for I27 dimer with linker LPETGGG. I27 is colored in yellow, the linker is colored in red, respectively.

Video 4. 100-ns MD simulations for I27 dimer with linker LPETGGGG. I27 is colored in mauve, the linker is colored in red, respectively.

Viedo 5. SMD simulations for closed conformation of I27 dimer with the 4G linker. I27 is colored in cyan, linker LPETGGGG is colored in red, and hydrophobic residues, Lys 37, Leu 65, Thr 68, Lys 35', Lys 37' and Leu 38' are highlighted.

Viedo 6. SMD simulations for open conformation of I27 dimer with the 4G linker. I27 is colored in cyan, linker LPETGGGG is colored in red, respectively.

Supplementary References

1. M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *P. Natl. Acad. Sci. USA*, 1999, **96**, 3694-3699.
2. Y. Deng, T. Wu, M. Wang, S. Shi, G. Yuan, X. Li, H. Chong, B. Wu and P. Zheng, *Nat. Commun.*, 2019, **10**, 2775.
3. S. Garg, G. S. Singaraju, S. Yengkhom and S. Rakshit, *Bioconjug. Chem.*, 2018, **29**, 1714-1719.
4. H. P. Liu, D. T. Ta and M. A. Nash, *Small Methods*, 2018, **2**.
5. A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, 725-738.
6. J. Hsin, A. Arkhipov, Y. Yin, J. E. Stone and K. Schulten, *Curr. Protoc. Bioinformatics*, 2008, **Chapter 5**, Unit 5 7.
7. J. V. Ribeiro, R. C. Bernardi, T. Rudack, J. E. Stone, J. C. Phillips, P. L. Freddolino and K. Schulten, *Sci. Rep.*, 2016, **6**, 26536.
8. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781-1802.
9. R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. Mackerell, Jr., *J. Chem. Theory Comput.*, 2012, **8**, 3257-3273.
10. T. S. Hofer and M. J. Wiedemair, *Phys. Chem. Chem. Phys.*, 2018, **20**, 28523-28534.
11. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613-4621.
12. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.