Check for updates

View Article Online
View Journal

ChemComm
Chemical Communications

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: F. Tian, G. Li, B. Zheng, Y. Liu, S. Shi, Y. Deng and P. zheng, *Chem. Commun.*, 2020, DOI: 10.1039/D0CC00714E.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

Please do not adjust margins

ChemComm

View Article Online
DOI: 10.1039/D0CC00714E

COMMUNICATION

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

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

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

Sortase is one of the most widely used enzymes for covalent protein conjugation which links protein and protein/small molecules together in a site-specific way. It typically recognizes the "GGG" and "LPXTG" peptide sequences and conjugates them into an "LPXTGGG" linker. As a non-natural linker with several flexible glycine residues, it is unknown whether it affects the properties of the conjugated protein. To verify the use of sortase for protein-protein conjugation, we combined single-molecule force spectroscopy (SMFS) and molecular dynamics (MD) simulations to characterize sortase-conjugated I27 with three different linkers. We found that the polyprotein I27 with classic linkers "LPETGGG" and "LPETG" from sortase ligation were of normal stability. However, a protein with a longer artificial linker "LPETGGGG" showed a 15% lower unfolding force. MD simulations revealed that the 4G linker showed a high probability of a closed conformation, in which the adjacent monomer has transient protein-protein interaction. Thus, we verify the use of sortase for protein conjugation, and a longer linker with higher glycine content should be used with caution.

Protein conjugation is a powerful methodology for both basic and applied research fields, including the study of protein mechanism and function, the construction of protein arrays for drug screening and proteomics, and protein-based drugs.[1-4] Among different approaches, enzyme-mediated protein conjugation is of great interest because it allows for a site-specific and covalent peptide bond linkage between ligation units.[5-9] For example, sortase, a transpeptidase known of for decades, has been widely used for protein ligation, labeling, and immobilization.[8,10] Sortase A recognizes specific residues on the protein terminus and eventually forms a seven amino-acid-length "LPXTGGG" linkage. However, such a non-natural and long linker with several flexible glycines may affect the structure and properties of the individual protein unit, leading to an unwanted "linker effect".[11-14]

Recently, a paper reported the use of sortase for building I27 polyprotein with a GGGG linkage leading to much-reduced protein stability.[15] Interestingly, another work using sortase with a GGG as the linker for building I27 polyprotein showed a similar unfolding force as a normal protein.[16] This raises concerns about the use of sortase for protein conjugation. To verify the use of sortase for protein-protein conjugation, we combined atomic force microscopy (AFM)-based SMFS, MD simulation, and protein engineering to characterize the stability of protein unit in the sortase-mediated conjugated polyprotein I27 with three different linkers, "LPETG" (1G), "LPETGGG" (3G) and "LPETGGGG" (4G). AFM-based SMFS is a powerful way to manipulate protein mechanically and measures protein unfolding force.[17-25] Together with MD simulations and theoretical methods, it characterizes protein stability and mechanics.[26-34] Compared with other classic protein characterization methods, such as SDS-PAGE gel or mass spectroscopy, it measures the native protein in aqueous solution at room temperature. Moreover, it directly measures the unfolding force/stability of the protein, which verifies the correct folding and structure of the protein.[35, 36] Thus, it is an alternative way to characterize ligated protein and provide complementary information for the conjugated protein under a native state.

mgSrtA, a Sortase A variant with nine amino acid mutations, is used for building conjugated I27 with different linkers.[5, 37] It recognizes both a classic N-terminal GGG sequence and a single G sequence with high ligation efficiency. Thus, it was used here for protein conjugation and to test the linker effect. The 27th immunoglobulin (Ig) domain of sarcomeric protein titin (I27) was chosen as the model protein. Titin is the largest naturally occurring protein molecule in human body with hundreds of Ig domains and was the first protein used for AFM-based SMFS protein unfolding experiments.[17,18,38,39] Thus, it has been well characterized and is widely used as a marker protein for single-molecule studies. Previous measurements using the recombinantly expressed polyprotein I27 with an RS linker showed an average force of ≈200 pN with a contour length increment (ΔLc) of ≈28 nm upon unfolding.[18]

State Key Laboratory of Coordination Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, China. E-mail: pengz@nju.edu.cn
Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

Published on 11 March 2020. Downloaded by Columbia University Libraries on 3/11/2020 12:51:48 PM.

ChemComm Accepted Manuscript

Please do not adjust margins

View Article Online
DOI: 10.1039/D0CC01715E

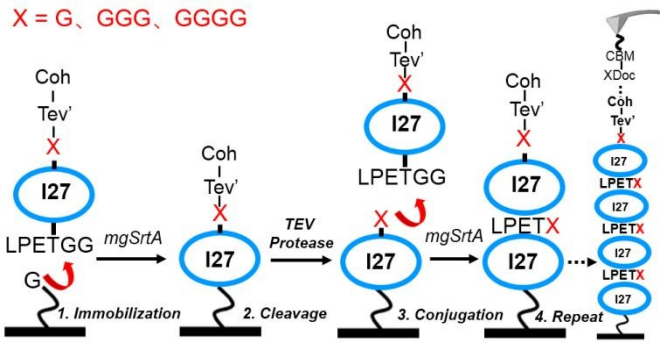


Figure 1. Schematic for building conjugated protein $(I27)_4$ for AFM measurement. A protein ligation unit, Coh-Tev'-X-I27-LPETGG, was used. X stands for three different sequences. Firstly, the protein unit was enzymatically immobilized on a glycine-functionalized surface using mgSrtA. Then, the unit was cleaved by TEV protease, exposing the X. Thirdly, an additional unit was conjugated to the exposed immobilized protein. These cycles can be repeated four times with the construction of $(I27)_4$. Finally, a CBM-XDoc functionalized AFM tip probed the polyprotein for measurement.

To construct sortase-mediated polyprotein $(I27)_N$ with different linkers for AFM measurement, a stepwise ligation and cleavage method using both mgSrtA and TEV protease was used.[6,15] Accordingly, the protein Coh-Tev'-X-I27-LPETGG was constructed as the basic ligation unit. Tev' stands for peptide sequence ENLYFQ, which, with an additional glycine, is the complete TEV cleavage site. X stands for G, GGG, and GGGG. The different linkages between I27 in the polyprotein were achieved by adding different residue(s) X before the I27, leading to an "LPETX" linkage (Fig. 1). The AFM system measured the protein sample in a well-defined site-specific way.[6,40] First, the protein unit is immobilized on a glycine-functionalized glass coverslip through its C terminal LPETGG by mgSrtA (Fig. 1, Step 1). Here, the Tev' site served as temporary N-terminal protection. Then, TEV protease cleaved the protein and exposed the X at the N-terminus (Step 2, Fig. S1). As a result, mgSrtA was able to ligate another protein unit leading to a protein dimer with the linkage "LPETX". This cycle can be repeated N times to allow for the construction of the polyprotein $(I27)_N$. It is noted that the real polymerization degree is less than the number of cycles because the sortase hydrolyzes the linker itself. Typically, we repeated the cycles six times, and I27 dimer to tetramer were all obtained. For example, $(I27)_4$ can be built as Coh-$(I27\text{-}LPETX)_4$-Glass. Here, the Coh (Cohesin) was used as the pulling handle for SMFS, because it forms a reversible Cohesin/Dockerin-Xmodule (Coh/XDoc) protein-protein interaction. Thus, a CBM-XDoc functionalized AFM tip can probe the polyprotein on the surface site-specifically, and the CBM with ΔLc of ≈55 nm was used as the single-molecule marker (Fig. 2A, Fig. S3). To compare the different constructs accurately, at least two constructs were measured using the same cantilever at the same experiment. The details of AFM measurement can be found in SI.

First, the shortest linkage "LPETG" was tested for AFM measurement. Accordingly, the protein unit Coh-Tev'-G-I27-LPETGG was constructed. The stepwise protein conjugation protocol mentioned above was used, leading to a polyprotein Coh-$(I27\text{-}LPETG)_N$-Glass. The force-extension curve of stretching the polyprotein by single-molecule AFM showed the characteristic saw-tooth like peaks from the stepwise protein unfolding (Fig. 2, Curve 1). By fitting the elasticity of the curve using the worm-like chain model, unfolding events with ΔLc of ≈28 nm were detected, which agreed well with previous results from I27 unfolding. This result indicated the successful conjugation of I27 using sortase, and at least four I27s were linked. More importantly, the unfolding force for sortase-mediated polymerized I27 was 200±39 pN (average+s.t., n=364), which is also comparable with previous results. This result indicated no linker effect from the sortase-derived LPETG (1G) linker between protein monomers.

Table 1. AFM results for conjugated I27 with different linkers

Linker	ΔLc (nm)	Force (pN)	Number
LPETG (1G)	28.1 ± 0.7	200±39	364
LPETGGG (3G)	28.7 ± 2.1	189±35	286
LPETGGGG (4G)	28.6 ± 2.5	174±31	299
RS control	28.6 ± 2.3	204±30	144

Then, we tested the classic srtA-mediated linkage "LPETGGG" (3G), including three glycines between conjugated proteins. The protein unit, Coh-Tev'-GGG-I27-LPETGG, was built. And polyprotein Coh-$(I27\text{-}LPETGGG)_N$ was constructed on the surface and measured by AFM. The same force-extension curve as before was observed (Fig. 2B, curve 2). Moreover, a slightly lower (15 pN, 7%) force of 189±35 pN (n=286) was obtained. These results prove that the classic LPETGGG linkage has a minimal effect on the conjugated protein.

Glycine is a flexible amino acid, and previous studies have suggested a higher frequency of these flexible residues in the linker may cause a non-natural protein-protein interaction.[11-14] It is also widely used to facilitate protein-protein interaction as the linker $(GGS)_n$. To verify such a hypothesis, we tested an artificial linker, LPETGGGG, with four GGGG (4G). A protein unit, Coh-Tev'-GGGG-I27-LPETGG, was constructed accordingly. Similarly, AFM measurement showed the protein was linked and I27 unfolding peaks were observed (Fig. 2B, curve 3). In contrast, the unfolding force was of 174±31 pN (n=299). Compared with the unfolding force (204 pN) of the recombinantly expressed polyprotein[16], it was much lower (30 pN, 15%). To exclude other effects, such as the different protein immobilization methods, polyprotein length, and cantilever, we built polyprotein Coh-$(I27\text{-}RS)_4$-LPETGG, recombinantly, and used the same method for AFM measurement. The unfolding force was 204±30 pN which was the same as previous results. Taken together, we believe the use of sortase for protein conjugation with an LPETG or LPETGGG linkage is suitable and longer linkers, especially with higher glycine content, should be used with caution.

Published on 11 March 2020. Downloaded by Columbia University Libraries on 3/11/2020 12:51:48 PM.

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

View Article Online
DOI: 10.1039/D0CC00714E

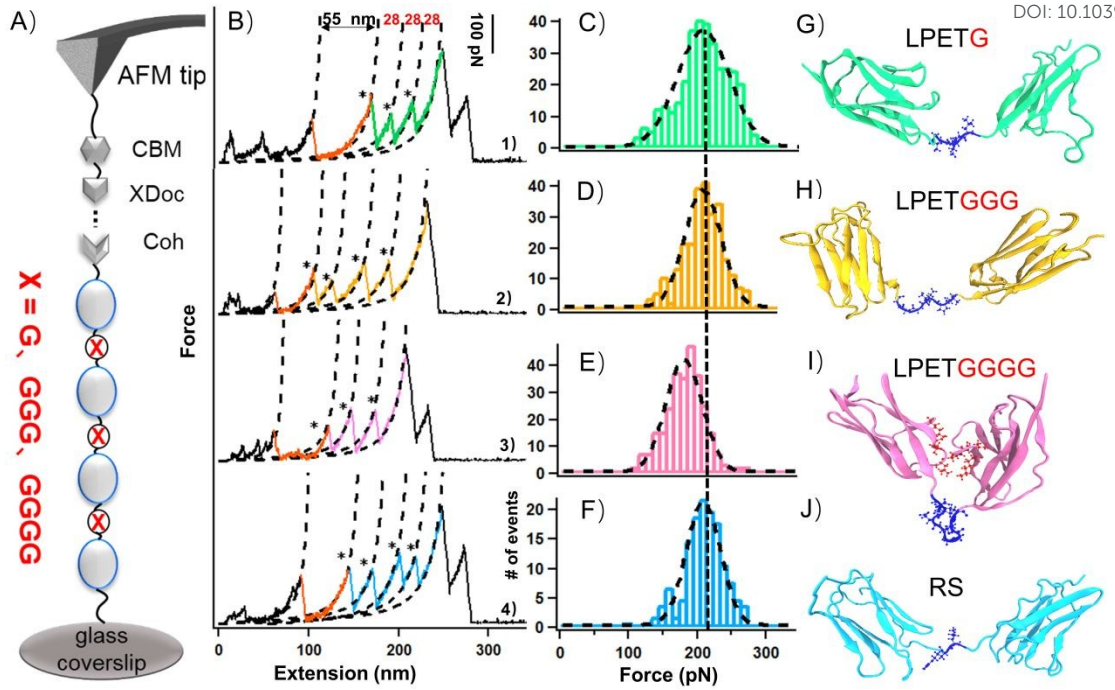


Figure 2. A) Setup of the AFM unfolding experiment on sortase-mediated conjugated polyprotein with different linkages. The protein was covalently immobilized and probed specifically through a reversible Coh/XDoc interaction. B) Representative force-extension curves from $(I27)_N$ with the linker "LPETG", curve 1; "LPETGGG" curve 2; "LPETGGGG", curve 3 and "RS", curve 4. Histograms of their corresponding unfolding force (C-F) and dimer conformation under MD simulations (G-J) are shown. A closed conformation with inter-domain interaction from hydrophobic residues is observed with a high probability for the 4G-construct.

.

To further understand why the longer linker with more glycine residues leads to a different unfolding force of I27 in the conjugated protein, we performed MD simulations.[41-44] The I27 dimer structures with four different linkers were constructed based on the crystal structure (PDB code: 1TIT). The details of the MD simulation are provided in SI.

After 100 ns of simulations, we compared the snapshots of four simulations (Fig. 2G-J). I27 dimer with the linkers of LPETG, LPETGGG, and RS achieved an appropriate separation of connected proteins and reduced their interference with each other like an open state (SI, Videos 1-3). Interestingly, I27 dimers with the 4G linker appeared to twine around each other as a closed state. Indeed, during the simulations, individual I27s were close to each other step-by-step (SI, Video 4). The closed state is from the transient protein-protein interaction between several interdomain hydrophobic residues. Three different MD simulations were performed. Interestingly, Lys 37, Leu 65 and Thr 68 in the first I27 domain were always presented for the interaction, while the residues involved in the other domain were not fixed (Fig. S4). The probability of open and closed conformations in all 100-ns MD simulations for the four linkers supports this difference (Table 2). The 4G-linked I27 gave rise to more closed conformation (39.7±7%, n=3). And the analysis of radius of gyration (Rg) also supports a more closed or compact conformation for it (Fig. S5). Thus, the observation of different probabilities of this closed conformation is the most significant difference between the four linkers.

To further understand these observations, we performed steered molecular dynamics (SMD) simulations on the closed and open conformation of I27 dimer linked by 3G and 4G, respectively. Each representative snapshot of the closed conformation from three MD simulations was selected as the starting state for SMD simulations. The results of the 4G-linked I27 showed the I27 dimer unfolded in the same way as typical protein unfolding. The linker was straightened first, and then the protein was unfolded. (Fig. 3, SI Videos 5&6). It showed the highest force peak when the first β-strand was completely unfolded. Moreover, the simulated unfolding force of the closed conformation is lower than that of the open one (Fig. 3, 1650 pN Vs. 2000 pN). Thus, the SMD simulations result also revealed that the closed conformation showed lower mechanical stability. Nevertheless, the molecular/atomic mechanism for such an effect remains further exploration.

Table 2. The probability of two conformational states of I27 dimer with different linkers in MD simulation.

Linker	Closed (%)
LPETG (1G)	6.1
LPETGGG (3G)	12.4
LPETGGGG (4G)	39.7
RS control	1.3

In conclusion, AFM measurement and SMD simulation verified the use of sortase for conjugating protein with classic LPETGGG linker and LPETG linker. However, a longer linker LPETGGGG may induce a transient protein-protein interaction,

Published on 11 March 2020. Downloaded by Columbia University Libraries on 3/11/2020 12:51:48 PM.

ChemComm Accepted Manuscript

View Article Online
DOI: 10.1039/D0CC01414E

and it should be used with caution. This work was supported by the National Natural Science Foundation of China (Grant No. 21771103, 21977047). The genes were ordered from genscript Inc. The numerical calculations in this paper have been done on the computing facilities in the High Performance Computing Center (HPCC) of Nanjing University.

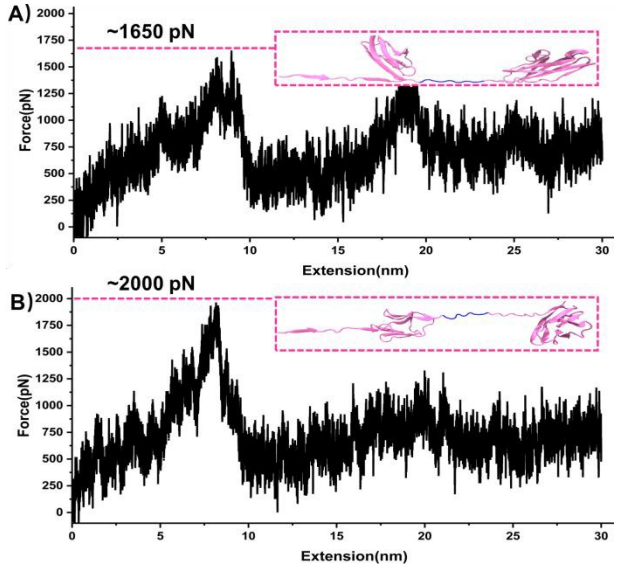


Figure 3. Force-extension curves from SMD simulations of the closed (A) and open (B) conformation of the 4G-linked dimer. The simulated unfolding force of the closed conformation is lower than that of the open conformation.

Notes and References

1. R. E. Thompson and T. W. Muir, *Chem. Rev.*, 2019, DOI: 10.1021/acs.chemrev.9b00450.
2. J. M. Antos, M. C. Truttmann and H. L. Ploegh, *Curr. Opin. Struct. Biol.*, 2016, **38**, 111-118.
3. K. Tsuchikama and Z. An, *Protein Cell*, 2018, **9**, 33-46.
4. V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429.
5. Y. Ge, L. Chen, S. Liu, J. Zhao, H. Zhang and P. R. Chen, *J. Am. Chem. Soc.*, 2019, **141**, 1833-1837.
6. Y. Deng, T. Wu, M. Wang, S. Shi, G. Yuan, X. Li, H. Chong, B. Wu and P. Zheng, *Nat. Commun.*, 2019, **10**, 2775.
7. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao and B. L. Pentelute, *J. Am. Chem. Soc.*, 2012, **134**, 10749-10752.
8. H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670-2671.
9. R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar and B. Wu, *J. Am. Chem. Soc.*, 2017, **139**, 5351-5358.
10. B. M. Dorr, H. O. Ham, C. H. An, E. L. Chaikof and D. R. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 13343-13348.
11. W. Wriggers, S. Chakravarty and P. A. Jennings, *Pept. Sci.*, 2005, **80**, 736-746.
12. X. Chen, J. L. Zaro and W. C. Shen, *Adv. Drug Deliv. Rev.*, 2013, **65**, 1357-1369.
13. M. Frei, S. V. Aradhya, M. S. Hybertsen and L. Venkataraman, *J. Am. Chem. Soc.*, 2012, **134**, 4003-4006.
14. V. P. Reddy Chichili, V. Kumar and J. Sivaraman, *Protein Sci.*, 2013, **22**, 153-167.
15. S. Garg, G. S. Singaraju, S. Yengkhom and S. Rakshit, *Bioconj. Chem.*, 2018, **29**, 1714-1719.
16. H. P. Liu and M. A. Nash, *Small Methods*, 2018, **2**,2366-9608.
17. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109-1112.
18. M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3694-3699.
19. T. Hoffmann, K. M. Tych, T. Crosskey, B. Schiffrin, D. J. Brockwell and L. Dougan, *ACS Nano*, 2015, **9**, 8811-8821.
20. D. Giganti, K. Yan, C. L. Badilla, J. M. Fernandez and J. Alegre-Cebollada, *Nat. Commun.*, 2018, **9**, 185.
21. Y. Song, Z. Ma, P. Yang, X. Zhang, X. Lyu, K. Jiang and W. Zhang, *Macromolecules*, 2019, **52**, 1327-1333.
22. M. Muddassir, B. Manna, P. Singh, S. Singh, R. Kumar, A. Ghosh and D. Sharma, *Chem. Commun.*, 2018, **54**, 9635-9638.
23. H. Yu, P. R. Heenan, D. T. Edwards, L. Uyetake and T. T. Perkins, *Angew. Chem. Int. Ed.* , 2019, **58**, 1710-1713.
24. H. Li and P. Zheng, *Curr. Opin. Chem. Bio.*, 2018, **43**, 58-67.
25. J. Perales-Calvo, D. Giganti, G. Stirnemann and S. Garcia-Manyes, *Sci. Adv.*, 2018, **4**, eaaq0243.
26. F. Sumbul and F. Rico, in *Atomic Force Microscopy: Methods and Protocols*, eds. N. C. Santos and F. A. Carvalho, Springer New York, New York, NY, 2019, pp. 163-189.
27. F. Franz, C. Daday and F. Gräter, *Curr. Opin. Struc. Bio.*, 2020, **61**, 132-138.
28. R. C. Bernardi, E. Durner, C. Schoeler, K. H. Malinowska, B. G. Carvalho, E. A. Bayer, Z. Luthey-Schulten, H. E. Gaub and M. A. Nash, *J. Am. Chem. Soc.*, 2019, **141**, 14752-14763.
29. M. T. Woodside and S. M. Block, *Annu. Rev. Biophys.*, 2014, **43**, 19-39.
30. O. K. Dudko, G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15755-15760.
31. C. A. Plata, Z. N. Scholl, P. E. Marszalek and A. Prados, *J. Chem. Theory Comput.*, 2018, **14**, 2910-2918.
32. P. Zheng, G. M. Arantes, M. J. Field and H. Li, *Nat. Commun.*, 2015, **6**, 7569.
33. T. Dudev, L. M. Frutos and O. Castaño, *Metallomics*, 2020, DOI: 10.1039/C9MT00283A.
34. A. Yadav, S. Paul, R. Venkatramani and S. R. K. Ainavarapu, *Sci. Rep.*, 2018, **8**, 1989.
35. Y. Chen, S. E. Radford and D. J. Brockwell, *Curr. Opin. Struct. Biol.*, 2015, **30**, 89-99.
36. H. Dietz and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16192-16197.
37. L. Chen, J. Cohen, X. Song, A. Zhao, Z. Ye, C. J. Feulner, P. Doonan, W. Somers, L. Lin and P. R. Chen, *Sci. Rep.*, 2016, **6**, 31899.
38. G. Yuan, S. Le, M. Yao, H. Qian, X. Zhou, J. Yan and H. Chen, *Angew. Chem. Int. Ed.*, 2017, **56**, 5490-5493.
39. D. J. Brockwell, G. S. Beddard, J. Clarkson, R. C. Zinober, A. W. Blake, J. Trinick, P. D. Olmsted, D. A. Smith and S. E. Radford, *Biophys. J.*, 2002, **83**, 458-472.
40. S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer and H. E. Gaub, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20431-20436.
41. 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.
42. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph. Model.*, 1996, **14**, 33-38.
43. 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.
44. T. S. Hofer and M. J. Wiedemair, *Phys. Chem. Chem. Phys.*, 2018, **20**, 28523-28534.

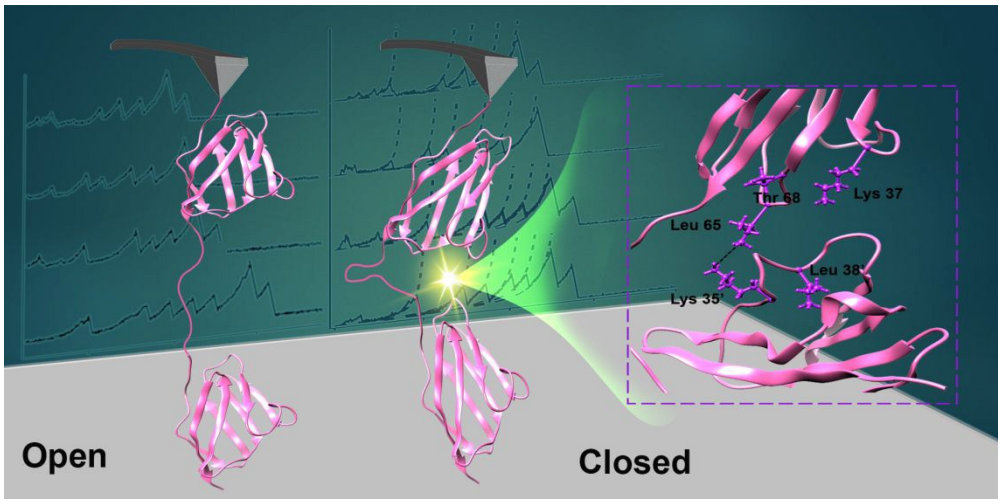
Published on 11 March 2020. Downloaded by Columbia University Libraries on 3/11/2020 12:51:48 PM.

ChemComm Accepted Manuscript

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

View Article Online
DOI: 10.1039/D0CC00714E



SMFS and MD simulations reveal a closed conformation and a decreased stability of sortase-mediated I27 polyprotein when a linker with a high content of glycine is used

Published on 11 March 2020. Downloaded by Columbia University Libraries on 3/11/2020 12:51:48 PM.

ChemComm Accepted Manuscript