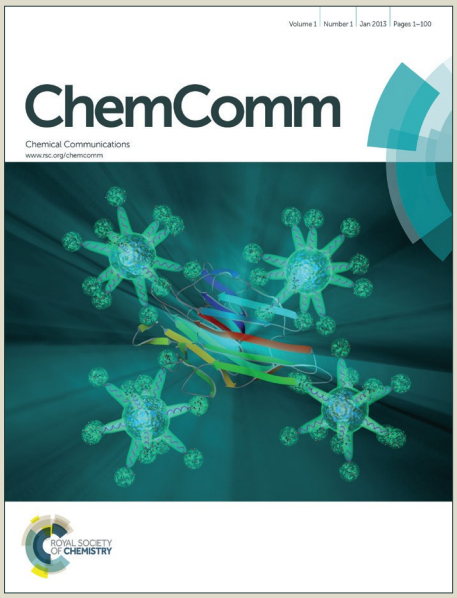
ChemComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Liu, Y. Cao, G. K. T. Nguyen and J. P. Tam, *Chem. Commun.*, 2015, DOI: 10.1039/C5CC07227A.



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

View Article Online
DOI: 10.1039/C5CC07227A

Journal Name

COMMUNICATION

Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yuan Cao, Giang K T Nguyen, James P Tam* and Chuan-Fa Liu*

Using a recently discovered peptide ligase, butelase 1, we developed a novel method to access protein thioesters in good yield. We successfully combined it with native chemical ligaiton and sortase-mediated ligation in tandem for protein C-terminal labeling and dual-terminal labeling to exploit the orthogonality of these three ligation methods.

Over the past three decades, considerable efforts have been devoted to developing chemoselective peptide ligation strategies[1] that operate in aqueous conditions and without the use of a protection group scheme, to enable protein synthesis and site-specific tagging of macromolecules for various biochemical applications.[2] The chemical strategies thus developed often exploit the unique bifunctional moiety of a peptide's N-terminal amino-acid residue that has a nucleophilic side chain such as Cys or Ser/Thr to react with another peptide containing a C-terminal "activated" moiety. These include, in chronological order, pseudo-proline ligation,[3] native chemical ligation,[4] thioacid capture ligation,[5] and serine/threonine ligation,[6] although the last one requires a non-aqueous solvent system. The most successful of these is native chemical ligation (NCL) to which the key is the use of a peptide C-ter thioester of aryl or alkyl nature for mediating the ligation reaction with a cysteinyl moiety.[4, 7] Besides being used in NCL, peptide thioesters are also building blocks in the more classic peptide segment condensation method via silver ion activation.[8] Thus, an efficient and convenient method to prepare a peptide/protein thioester is always desirable and continues to be well sought after. Whereas peptide thioesters can be synthesized directly by Boc chemistry[7b, c, 8] or indirectly by Fmoc chemistry,[9] protein thioesters are obtained mostly through the intein technology.[10] This method requires the fusion expression of the target protein with an engineered intein and the addition of an external thiol to release the protein thioester from the immobilized intein (Fig. 1A).[10d, e] Other methods using subtiligase[11] or SrtA[12] to prepare peptide/protein thioesters have also been developed.[13]

Recently a novel Asn/Asp(Asx)-peptide ligase, butelase 1, was discovered.[14] Butelase 1 performs the cyclization step in the cyclotide biosynthesis in the medicinal plant *Clitoria ternatea*. It recognizes a –N(D)HV motif at the C-terminus and cleaves the N(D)H bond to conjugate with an incoming N-terminal amino group (Xaa) to form an Asx-Xaa at the ligation junction. Butelase 1 has been shown to achieve success in both inter- and intramolecular peptide ligations with a high catalytic efficiency and near "traceless" feature.[14]

Herein we report a butelase-mediated ligation (BML) method to prepare protein thioesters (Fig. 1B) with high efficiency. We further coupled this method with NCL and sortase-mediated ligation (SML) for labelling proteins to show that these three ligation strategies are orthogonal to each other.

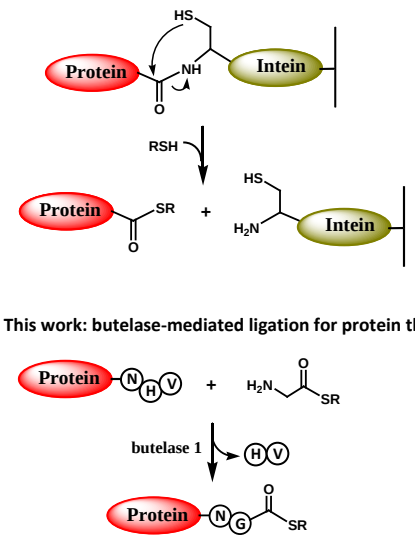
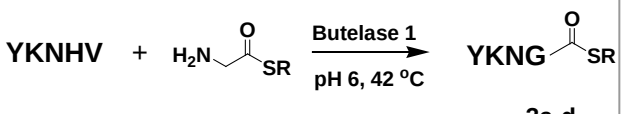


Fig. 1 Protein thioester preparation using (A) intein-mediated technology and (B) butelase-mediated ligation.

[a] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551.E-mail: jptam@ntu.edu.sg (J. P. Tam), cfliu@ntu.edu.sg (C.-F. Liu)

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

Table 1 Ligation between a model peptide and different glycine thioesters



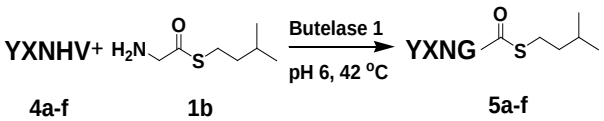
2 **1a-d** **3a-d**

Glycine thioester	R	Yield [%]
1a	R1	>95
1b	R2	>95
1c	R3	>95
1d	R4	>95

All reactions were carried out in the same condition: 50 μM peptide, 2.2 mM glycine thioester, 100 nM butelase 1, pH 6, 20 mM phosphate buffer, 42 °C. Yields were measured after 2 h when reactions reached equilibrium.

Four glycine thioesters Gly-COSR **1a-d** (Table 1) were first used to demonstrate butelase-mediated thioester synthesis. Since these simple glycyl thioesters were easily synthesized from Boc-Gly-OH and commercially available thiols, one could afford to use them in large excess (44 fold to the acyl peptide substrate, Table 1) to outcompete the reverse reaction, essentially overcoming the reversibility problem of BML. As a result, all four glycine thioesters ligated efficiently with the model peptide YKNHV **2** in the presence of butelase 1 to afford the corresponding peptide thioesters YKNG-COSR **3a-d** in near quantitative yields (Table 1, Fig. S1†) as analysed by HPLC. To determine the pH effect of this enzymatic ligation, reaction between **1b** and **2** was performed in different pH conditions: pH 4, 5, 6, 6.5, 7.5 and 8. We found that the reaction proceeded efficiently at pH 5, 6 and 6.5 (>90%), and less efficiently at pH 4, 7.5 and 8. This result indicates that butelase 1 performs ligation optimally under mildly acidic pH conditions, a featu desirable for the stability of thioesters. For convenience, we use the glycine thioester **1b** in our following work. We further prepa a series of peptides **4a-f** with a common sequence YXNHV varyi the residue at P2 position (X = Val, Leu, Phe, Ser, Nle, D-Ala). All six peptides, including two peptides containing unnatural amino acid ligated with excessive glycine thioester **1b** at pH 6 to afford the

Table 2 Ligation between **1b** and different peptides



4a-f **1b** **5a-f**

Peptide	X	Yield [%]
4a	Val	>95
4b	Leu	93
4c	Ser	81
4d	Phe	>95
4e	Nle	>95
4f	D-Ala	91

All reactions were carried out in standard condition: 50 μM peptide, 2 mM glycine thioesters, 100 nM butelase 1, 20 mM phosphate buffer, pH 6, 42 °C. All yields were measured after 95 min when reactions reached equilibrium.

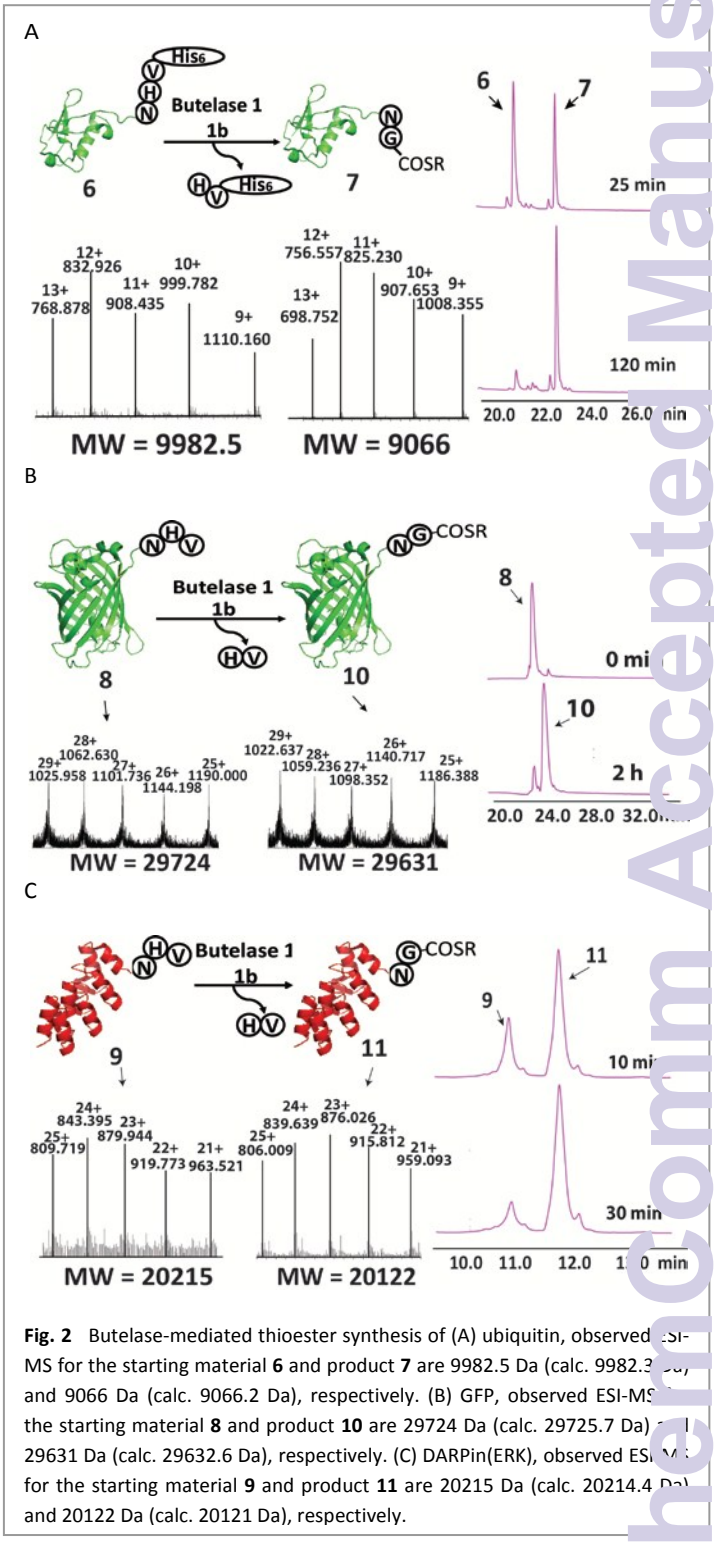


Fig. 2 Butelase-mediated thioester synthesis of (A) ubiquitin, observed ESI-MS for the starting material **6** and product **7** are 9982.5 Da (calc. 9982.3 Da) and 9066 Da (calc. 9066.2 Da), respectively. (B) GFP, observed ESI-MS the starting material **8** and product **10** are 29724 Da (calc. 29725.7 Da) 29631 Da (calc. 29632.6 Da), respectively. (C) DARPin(ERK), observed ESI-M for the starting material **9** and product **11** are 20215 Da (calc. 20214.4 Da) and 20122 Da (calc. 20121 Da), respectively.

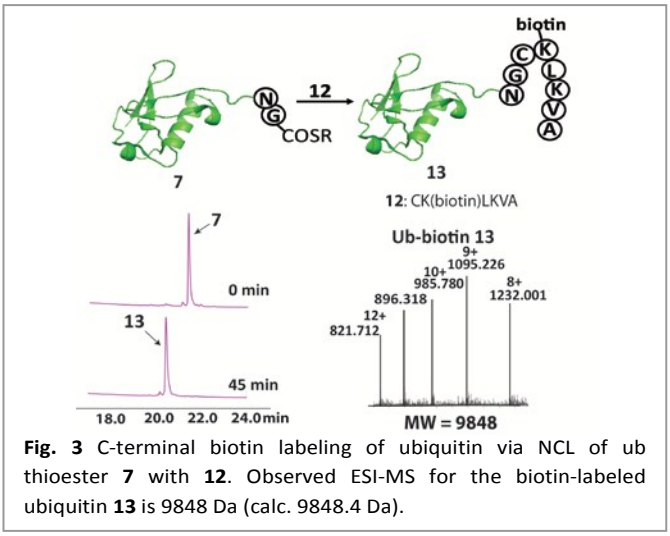


Fig. 3 C-terminal biotin labeling of ubiquitin via NCL of ub thioester **7** with **12**. Observed ESI-MS for the biotin-labeled ubiquitin **13** is 9848 Da (calc. 9848.4 Da).

respective products YXNG-COSR **5a-f** in good yields after 95 min (Table 2). This result clearly showed the generality of our method and that butelase 1 tolerates a variety of amino acids N-terminal to the NHV tripeptide recognition motif. Clearly, an advantage of this approach is that a large excess of the Gly-COSR can be used to force the reaction to completion in a short duration.

We also tested the thioesters of four other amino acids (Ala, Leu, Ile, Phe) in butelase-mediated ligation with model peptide **2**. Surprisingly, only alanine thioester gave some modest ligation product, while the other amino thioesters did not work in the ligation reaction (data not shown). So, a limitation of this method is that, among the amino-acid thioesters, only glycinyl thioesters are efficient substrates of butelase 1.

To show the compatibility of BML with NCL, the peptide thioester YKNG-COSR **3b** was ligated with a cysteinyl peptide CGYKNHV to afford YKNGCGYKNHV in quantitative yield under standard NCL conditions (Fig. S2†).

To show the utility of this approach for proteins, we prepared an engineered ubiquitin **6** with the NHV recognition motif and a His_6-tag for affinity purification at its C-terminus. Reaction of **6** (100 μM) with glycine thioester **1b** (2 mM, 20 eq.) in the presence of 200 nM butelase 1 (0.002 eq.) afforded the protein thioester product **7** in >90% yield after 2 h (Fig. 2A). We also prepared two other proteins: GFP **8** and an ERK-ankyrin repeat protein (DARPin(ERK))[15] **9**, both engineered to have a C-ter NHV motif. Using the same reaction, we obtained the thioester product **10** and **11** in *ca.* 85% yield in short time (Fig. 2B and C). These results show again the high catalytic efficiency of butelase 1. Since the –NHV motif is very small, any protein tagged with it can be easily expressed as shown in the three protein examples here. These features make our approach a very robust and convenient method to prepare protein thioesters.

Next, we showed the prepared protein thioester was useful for NCL. So, a cysteinyl peptide CK(biotin)LKVA **12** was ligated with the ubiquitin thioester **7** under standard NCL conditions to afford, in near quantitative yield as observed by HPLC analysis (Fig. 3), the product ubiquitin **13** containing now a biotin probe at its C-ter end.

To show that SML can also be used in conjunction with the above two reactions for a three-step tandem ligation, another recombinant ubiquitin **14** was prepared. It contained a –NHV motif followed by a His_6-tag at the C-terminus. At N-terminus, it displayed

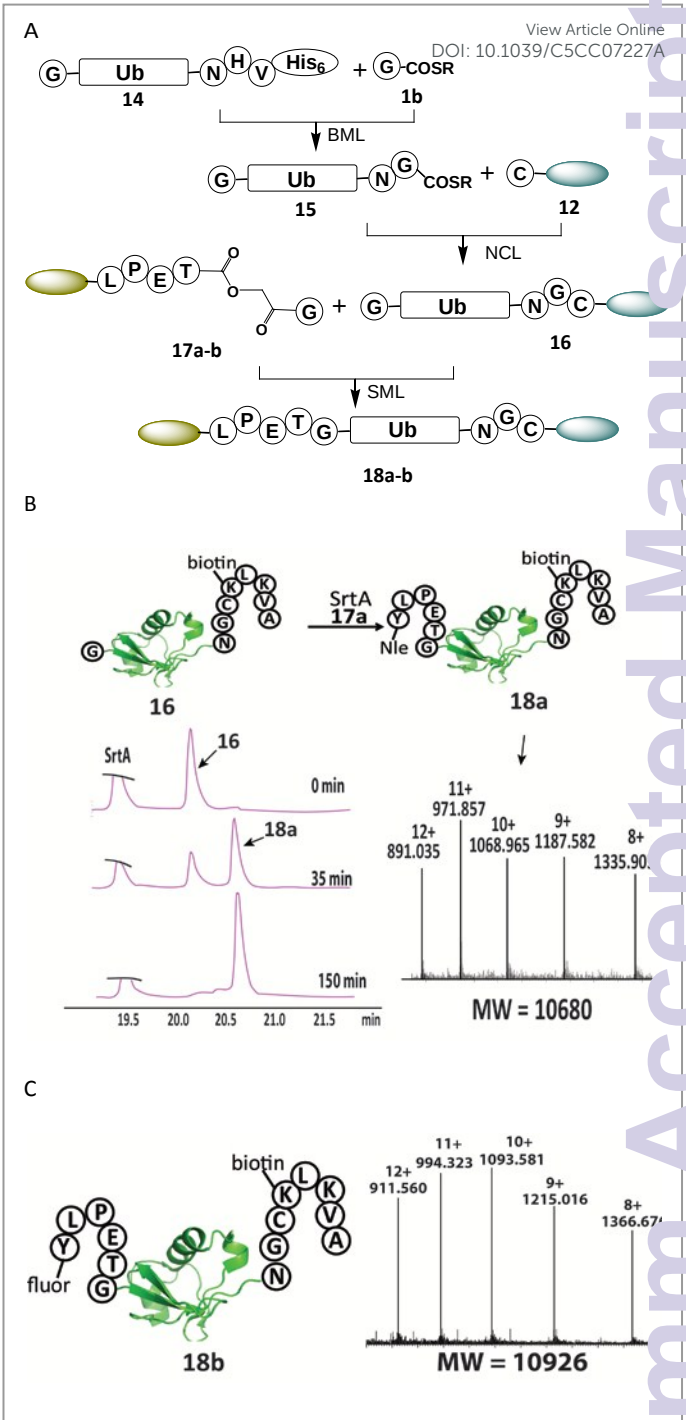


Fig. 4 (A) Schematic illustration for the dual-termini labelling of ubiquitin (B) Sortase-mediated N-terminal labeling of the Ub-biotin **16** with Nle YLPET-glc-G **17a**. Observed ESI-MS for the final dual-terminal labeled product **18a** is 10680 Da (calc. 10680.2 Da). (C) Observed ESI-MS of the dual-labeled fluor-Ub-biotin **18b** is 10926 Da (calc. 10925.7 Da) obtained via SML of **16** with fluor-YLPET-glc-G **17b** .

a glycine residue for SrtA recognition. One of the practical applications of this convergent methodology could be dual-terminus modification of proteins (Fig. 4A). This attempt was previously done by using two sortases with different activities.[16] In our case, the C-terminal labeling of **14** was first conducted using the above two-step process: BML with glycine thioester **1b** followed by NCL with biotinyl peptide **12**. 85% and quantitative HPLC

conversions were obtained for the BML and NCL steps, respectively, to give **15** and **16** (Fig. S3 and 4†). After HPLC purification, the C-terminus-labeled ubiquitin **16** was further labeled on the N-terminus using SrtA and a depsipeptide containing a motif LPET-glc-G. The use of the ester derivative of the SrtA motif LPXTG was based on a previous report that it could significantly improve the SML yield due to irreversibility of the reaction.[17] We placed a short sequence –GSGSGS- between the N-terminal Gly residue and the main body of ubiquitin for better accessibility of its N-terminus by SML. A depsipeptide Nle-YLPET-glc-G **17a** was first used for demonstration. The SML reaction gave a near quantitative yield of **18a** after 2.5 h with two molar equivalents of the depsipeptide and 0.1 molar equivalent of SrtA (Fig. 4B). SML requires a high molar ratio of SrtA, which is a significant drawback as compared to BML which requires a very small amount of the enzyme. Nevertheless, a simple treatment with Ni-NTA removed the His-tagged SrtA and subsequent dialysis removed the small peptide from the desired product. Using SML, **16** was also successfully ligated with a fluorescein-peptide fluor-YLPET-glc-G **17b** to give **18b** which is dually labelled with a fluorescence probe and biotin group on the N- and C-ter ends, respectively (Fig. 4C).

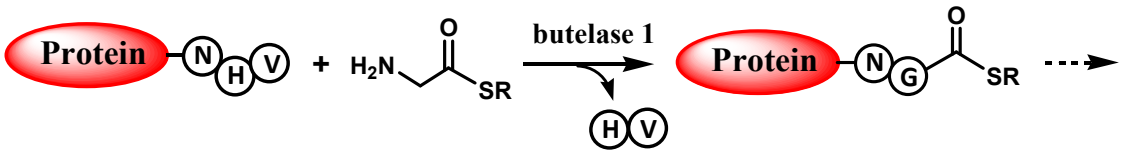
Currently, intein-based technology is the method of choice to access protein thioesters for ligation reactions. We have shown that BML provides a complementary method to prepare protein thioesters conveniently and efficiently as seen with several protein substrates of different sizes. This method requires the substrate protein to have only a small tripeptide motif NHV at the C-terminus and renders little change on the protein sequence after ligation as it leaves behind only a dipeptide trace -NG. We have further shown that this method can be combined successfully in tandem with NCL and SML for sequential ligation to achieve bi-directional and orthogonal labelling of ubiquitin in a model system. These results demonstrate butelase 1 as a versatile tool for protein manipulation. We foresee that, like other immerging methods,[18] BML will offer numerous future opportunities in biotechnology.

We thank Dr. T. Cornvik at Nanyang Technological University for providing the plasmids of some of the model proteins. This work was supported by Singapore National Research Foundation-Competitive Research Programme Grant (NRF-CRP8-2011-05).

Notes and references

1. (a) P. E. Dawson and S. B. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923-960; (b) C. P. R. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030-10074; (c) R. J. Payne and C. H. Wong, *Chem. Commun.*, 2010, **46**, 21-43; (d) J. P. Tam, Q. Yu and Z. Miao, *Biopolymers*, 1999, **51**, 311-332; (e) V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471-479.
2. (a) C. P. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030-10074; (b) J. P. Pellois and T. W. Muir, *Curr. Opin. Chem. Biol.*, 2006, **10**, 487-491; (c) D. Macmillan, *Angew. Chem. Int. Ed.*, 2006, **45**, 7668-7672.
3. C.-F. Liu and J. P. Tam, *J. Am. Chem. Soc., 1994, 116, 4149-4153*.
4. P. E. Dawson, T. W. Muir, I. Clarklewis and S. B. H. Ke, *Science*, 1994, **266**, 776-779.
5. C.-F. Liu, C. Rao and J. P. Tam, *Tetrahedron Lett.*, 1996, **37**, 933-936.
6. X. Li, H. Y. Lam, Y. Zhang and C. K. Chan, *Org. Lett.*, 2010, **12**, 1724-1727.
7. (a) T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang and H. Lau, *Liebigs Ann. Chem.*, 1953, **583**, 129-149; (b) J. P. Tam, Y. A. Lu, C.-F. Liu and J. Shao, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 12485-12489; (c) T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. USA*, 1999, 96, 10068-10073.
8. S. Aimoto, *Biopolymers*, 1999, **51**, 247-265.
9. (a) F. Mende and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 1232-1240; (b) J. Kang and D. Macmillan, *Org. Biomol. Chem.*, 2010, **8**, 1993-2002.
10. (a) K. V. Mills, M. A. Johnson and F. B. Perler, *J. Biol. Chem.*, 2014, **289**, 14498-14505; (b) N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446-461; (c) M. Q. Xu and T. C. Evans, Jr., *Methods*, 2001, **24**, 257-277; (d) T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. USA,* 1998, 95, 6705-6710; (e) T. C. Evans, Jr., J. Benner and M. Q. Xu, *Protein Sci.*, 1998, **7**, 2256-2264.
11. T. K. Chang, D. Y. Jackson, J. P. Burnier and J. A. Wells, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 12544-12548.
12. (a) Luciano A. Marraffini, Andrea C. DeDent and Olaf Schneewind., *Microbiol. Mol. Biol. Rev*. 2006, **70**, 192-221; (b) H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull and O. Schneewind, *Proc. Natl. Acad. Sci. USA*, 1999, 96, 12424-12429.
13. (a) J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao and B. L. Pentelute, *J. Am. Chem. Soc.*, 2012, **134**, 10749-10752; (b) Y. M. Li, Y. T. Li, M. Pan, X. Q. Kong, Y. C. Huang, Z. Y. Hong and L. L. Liu, *Angew. Chem., Int. Ed.*, 2014, **53**, 2198-2202; (c) X. H. Tan, A. Wirjo and C.-F. Liu, *ChemBioChem*, 2007, **8**, 1512-1515.
14. G. K. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam, *Nat. Chem. Biol.*, 2014, **10**, 732-738.
15. H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz and A. Pluckthun, *J. Mol. Biol.*, 2003, **332**, 489-503.
16. (a) J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp and H. L. Ploegh, *J. Am. Chem. Soc.*, 2009, **131**, 10800-10801; (b) P. R. Race, M. L. Bentley, J. A. Melvin, A. Crow, R. K. Hughes, W. D. Smith, R. B. Sessions, M. A. Kehoe, D. G. McCafferty and M. J. Banfield, *J. Biol. Chem.*, 2009, **284**, 6924-6933.
17. D. J. Williamson, M. A. Fascione, M. E. Webb and W. B. Turnbull, *Angew. Chem., Int. Ed.*, 2012, **51**, 9377-9380.
18. G. Veggiani, B. Zakeri and M. Howarth, *Trends Biotechnol.* 2014, **32**, 506-512.

View Article Online
DOI: 10.1039/C5CC07227A



Recombinant proteins with a C-terminal thioester are readily prepared through the catalysis of butelase 1 – a powerful peptide ligase.