

Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide

Giang K. T. Nguyen, Yuan Cao, Wei Wang, Chuan Fa Liu, and James P. Tam*

Abstract: An efficient ligase with exquisite site-specificity is highly desirable for protein modification. Recently, we discovered the fastest known ligase called butelase 1 from *Clitoria ternatea* for intramolecular cyclization. For intermolecular ligation, butelase 1 requires an excess amount of a substrate to suppress the reverse reaction, a feature similar to other ligases. Herein, we describe the use of thiodepsipeptide substrates with a thiol as a leaving group and an unacceptable nucleophile to render the butelase-mediated ligation reactions irreversible and in high yields. Butelase 1 also accepted depsipeptides as substrates, but unlike a thiodepsipeptide, the depsipeptide ligation was partially reversible as butelase 1 can tolerate an alcohol group as a poor nucleophile. The thiodepsipeptide method was successfully applied in N-terminal labeling of ubiquitin and green fluorescent protein using substrates with or without a biotin group in high yields.

Site-specific protein modifications with tags and probes offer useful tools for studying protein–protein interactions^[1] and structure–function relationships.^[2] A common approach is by chemical means that usually employ N-terminal residue, or functional side chains of cysteine or lysine for derivatization.^[3] Despite being powerful and robust, chemical methods generally require an excess amount of a labelling reagent, and carefully controlled reaction conditions.^[4] Often, site specificity becomes a challenge when multiple copies of a targeted amino acid or functional group are present in a protein substrate.^[5] Recently, enzymatic approaches using peptide ligases provide an attractive alternative with exquisite site-specificity for protein modification.^[6] Currently, sortase A is the most popular ligase,^[7] but it has low catalytic efficiency, requires a long reaction time, and a high molar equivalent of enzyme (typically 0.1 to 1 molar ratio).^[8] Furthermore, sortase A has a stringent substrate requirement, leaving behind a sorting sequence LPXTG in the resulting modified proteins.^[7] A ligase with a broad substrate specificity, efficient kinetics, and traceless ligation would be highly desirable.

Recently, we isolated an Asn/Asp(Asx)-specific ligase called butelase 1 from the medicinal plant *Clitoria ternatea*.^[9]

Butelase 1 is a cysteine ligase mediating the backbone cyclization in the biosynthesis of cyclotides, a family of circular plant defense peptides.^[10] Compared to sortase A, butelase 1 has the following advantages: 1) It has very high catalytic efficiencies, which typically requires 100- to 1000-fold less amount of enzyme in a ligation reaction.^[9a] 2) Butelase 1 is C-terminal-specific for Asx with a sorting sequence of two amino acids or less after Asx. Importantly, no sorting signal is left behind in the ligated protein products. Thus, butelase 1 can be exploited for the synthesis of naturally occurring circular proteins, a feature which eludes sortase A. 3) Butelase 1 displays a very broad specificity for the acceptor nucleophilic amino acids (all natural amino acids except Pro) to form a new Asx-Xaa peptide bond. These features make butelase 1 an attractive new tool for protein engineering. Our previous study showed that butelase-mediated intramolecular cyclization proceeds with high efficiency and is irreversible, whereas the intermolecular peptide ligation is reversible and requires an excess amount of substrate to drive the reaction to completion,^[9a] a condition also found in sortase A.^[8a–d]

Herein, we report the use of thiodepsipeptide as an acceptable sorting signal but a poor competing nucleophile after its release to render the butelase-mediated intermolecular ligation irreversible (Scheme 1). We obtained quantitative ligation yields of >95 % for a model peptide at 0.0005 molar equivalent of butelase 1 and two molar equivalents of the thiodepsipeptide. We also successfully applied the proposed method to label ubiquitin and GFP with high yields. This method is based on our previous work showing that incubation of butelase 1 with a model peptide KALVINHV, with the dipeptide HV as a leaving group, in the presence of various alkyl and aryl thiols did not lead to any detectable amount of peptide thioester.^[9a]

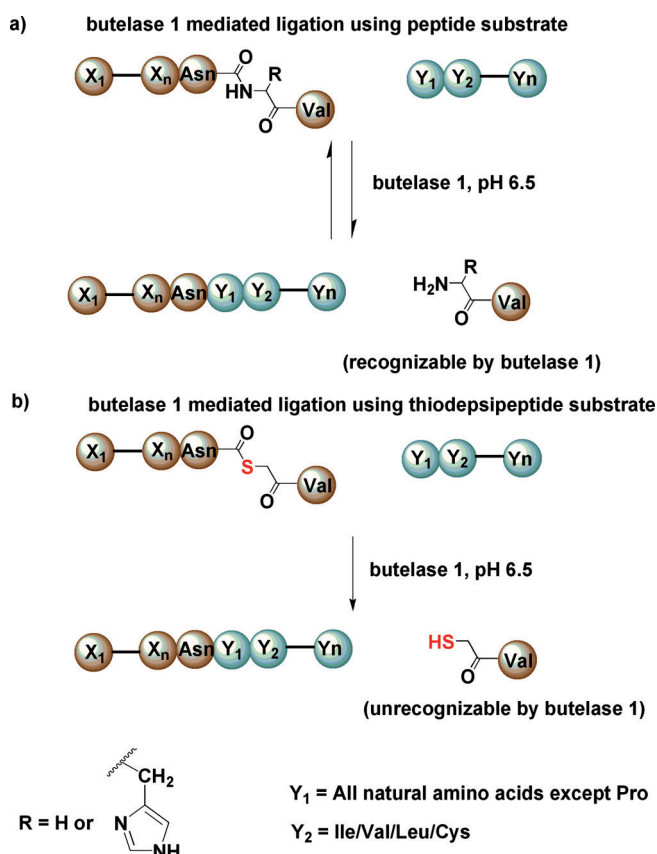
This result suggested that thiol groups are poor competing nucleophiles as compared to the HV dipeptide sorting signal. Thus, we hypothesized that intermolecular ligation reactions would be irreversible if the scissile asparaginyl amide bond is replaced by a thioester linkage. For comparison, we also synthesized a depsipeptide because it has been demonstrated that ligation yields were improved for sortase A using a depsipeptide substrate.^[8c] However, the applications of depsipeptide for N-terminal protein labeling required 0.1–0.2 molar equivalents of sortase A and a time-consuming solution-phase synthesis of depsipeptide.

To support our hypothesis, we prepared four different peptide substrates sharing a common sequence YKNXV 1–4 with a XV as a sorting signal (X = thio- or glycolic acid, His, and Gly, respectively; Table 1). Thiodepsipeptide 1 and depsipeptide 2 contain a S- and O-ester bond as a (S)GV and (O)GV analog, respectively, at the scissile bond compared

[*] Dr. G. K. T. Nguyen,^[+] Y. Cao,^[+] Dr. W. Wang, Prof. Dr. C. F. Liu, Prof. Dr. J. P. Tam
School of Biological Sciences, Nanyang Technological University
60 Nanyang Drive, Singapore 637551 (Singapore)
E-mail: jptam@ntu.edu.sg

[†] These authors contributed equally to this work.

Supporting information for this article, including experimental details, is available on the WWW under <http://dx.doi.org/10.1002/anie.201506810>.



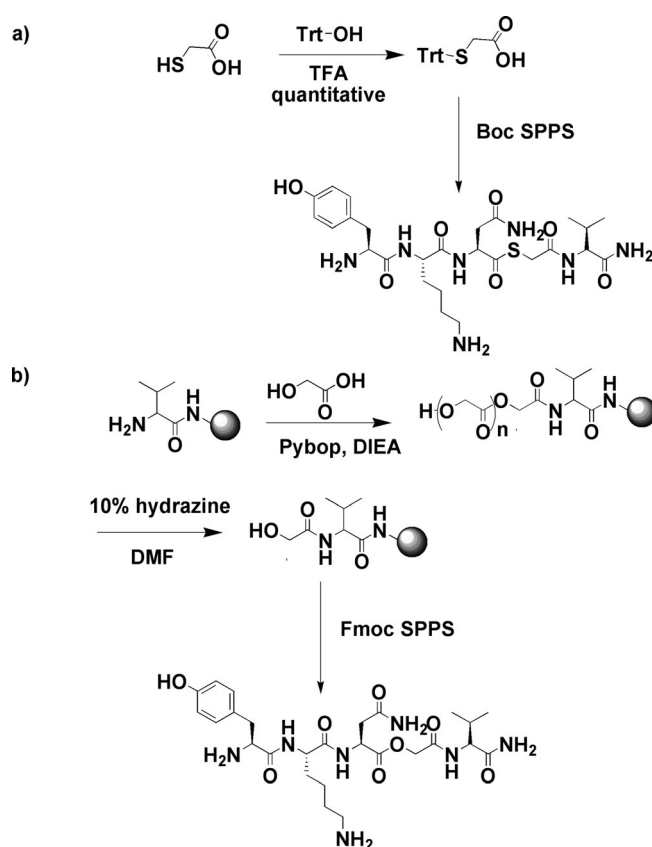
Scheme 1. Butelase 1 ligation using a) natural peptide and b) thiopeptide.

Table 1: List of peptide substrates and their corresponding molecular weight.

Peptide	Sequence	MW (MH ⁺)	
		Calc.	Obs.
1	YKN(S)GV	596.28	596.28
2	YKN(O)GV	580.30	580.19
3	YKNHV	659.36	659.27
4	YKNGV	579.32	579.16
5	GIGGIR	571.36	571.3
6	(O)GVYKV	565.33	565.28
7	ERLYRGRLYRRNHV	1888.05	1888.2
8	biotin-TYKN-thioglc-V	923.4	923.31

with an amide bond in the two control peptides **3** (HV) and **4** (GV). Thiopeptide **1** was prepared by a standard Boc solid-phase synthesis (Scheme 2a). Alternatively, it can also be prepared by Fmoc chemistry using the modified Fmoc deprotection cocktail containing 25 % 1-methylpyrrolidine, 2 % hexamethylenimine, and 2 % HOBT in a NMP/DMSO (1:1) mixture.^[11] For depsipeptide **2**, instead of using solution-phase synthesis,^[8c] we developed a fully solid-phase compatible and straightforward approach (Scheme 2b). The depsipeptides were synthesized on Rink amide MBHA resin with Fmoc amino acids and glycolic acid.

Hydrazinolysis of the glycolic acid-coupled product removed all side products of oligomers, leaving only one glycolic acid attached to the preceding residue as a stable



Scheme 2. Synthetic strategy for a) thiopeptide **1** and b) depsipeptide **2**.

amide bond. Our synthesis route provides a convenient method for preparing the depsipeptide.

We evaluated the ligation efficiency of each peptide substrate to a model peptide GIGGIR **5** to form the ligated product YKNGIGGIR by HPLC. The reactions were performed in the presence of 100 μ M GIGGIR, one or two molar equivalents of each peptide substrate, and 50 nM of butelase 1 (0.0005 molar equivalent). Time-course analysis showed that thiopeptide **1** was the most efficient substrate followed by depsipeptide **2** (Figure 1). Peptide **3** and peptide **4** were the least efficient. All ligation reactions reached equilibrium after 60 min with marginal improvement in yields after 90 min. At two molar equivalents of **1**, > 95 % of peptide **5** was converted into the ligation product within 1 h (Figure 1). In contrast, the ligation yields only reached 31, 59, and 68 %, respectively, for peptide **4**, **3** and **2** under similar conditions. Even at one molar equivalent of thiopeptide **1**, the ligation yield still reached 73 %, confirming thiopeptide is the best of the four substrates. Table 2 shows a kinetic study to quantify the difference among peptides **1–4**. The result was in agreement with the time-course experiments, with thiopeptide **1** having the highest catalytic efficiency (k_{cat}/K_m) and peptide **4** being the lowest.

Next, we sought to determine what shifts the equilibrium to a higher yield for thiopeptide **1** as compared to depsipeptide **2**, because their leaving groups are a thiol or an alcohol, both of which are poor nucleophiles for butelase 1.

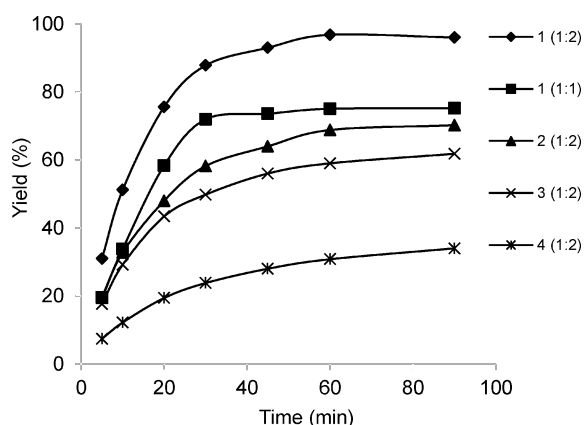


Figure 1. Time-course of butelase-mediated ligation of peptides 1–4 with peptide 5. The reaction was monitored by HPLC. Reaction conditions: 50 nM butelase 1, 100 μ M peptide 5, corresponding ratios of peptide 1–4, 20 mM phosphate buffer, 1 mM EDTA, pH 6.5, 42 °C.

Table 2: Kinetic parameters of butelase 1 for peptides 1–4.

Peptide	Sequence	k_{cat} [s^{-1}]	K_{m} [mM]	$k_{\text{cat}}/K_{\text{m}}$ [$\text{M}^{-1} \text{s}^{-1}$]
1	YKN-thioglc-V	56.5 ± 7.5	2.76 ± 0.3	20480
2	YKN-glc-V	12.4 ± 0.32	0.74 ± 0.01	16840
3	YKNHV	4.1 ± 0.65	0.3 ± 0.01	13490
4	YKNGV	7.9 ± 0.72	0.88 ± 0.08	8900

We found that our model peptide accepts glycerol as a nucleophile to give about 20% of the ligated product (Supporting Information, Figure S1). This observation suggests that butelase 1 can recognize an alcohol as an acceptor nucleophile. To demonstrate that butelase 1 indeed accepts an alcohol group, we synthesized peptide 6, (O)GVYKV, with glycolic acid as the N-terminal residue. Ligation of peptide 6 with peptide 7 (ERLYRGRLYRRNHV) led to a ligated product ERLYRGRLYRRN(O)GVYKV with about 25% yield (Figure S2). No detectable ligation product was formed when 7 was mixed with a (S)GV peptide under the same condition. These results indicate that the butelase-mediated ligation is reversible for depsipeptide and irreversible for thiodepsipeptide, which could explain the high yields of thiodepsipeptide 1.

To apply butelase 1 for N-terminal labelling of proteins, we used ubiquitin and green fluorescent protein (GFP) as examples (Figure 2). Since butelase 1 prefers a hydrophobic amino acid (Ile/Leu/Val) or Cys at the P2'' position of an acceptor nucleophile, we prepared a recombinant ubiquitin protein with Gly-Ile at the N-terminus and His-tag at the C-terminus. We also examined whether our method would work if the His-tag was located at the N-terminus. We thus prepared a recombinant GFP with an Ile inserted between the start codon and the N-terminal His-tag. Both proteins were labelled with the model thiodepsipeptide 1 in the presence of 0.001 molar equivalent of butelase 1. The labelling of proteins was slower than the labelling of peptide 5, probably because the N-terminal amino groups of proteins may have a lower accessibility to butelase 1 than short peptides. Furthermore, we found that the half-life of the thiodepsipep-

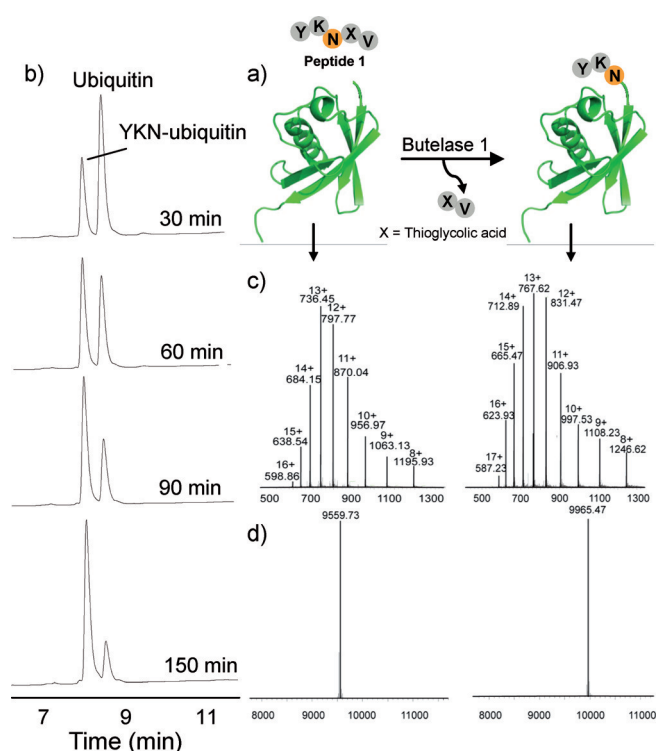


Figure 2. a) Illustration of the butelase-mediated N-terminal modification of ubiquitin by the use of thiodepsipeptide 1. b) Time-course analysis of the ubiquitin labelling reaction. c) ESI spectra of unmodified and modified ubiquitin. d) Deconvoluted ESI spectra of unmodified and modified ubiquitin.

tide 1 was relatively short, mainly owing to hydrolysis and aspartimide formation associated with the nature of the asparagine residue. The half-life of thiodepsipeptide 1 in our reaction buffer is about 75 min in the absence of butelase 1 and 40 min in the presence of 0.0005 molar equivalent of butelase 1 (Figure S3). For depsipeptide 2, the half-life is 45 min and 37 min in the absence and presence of butelase 1, respectively. This feature poses a problem when the ligation proceeds slowly, and a moderate yield ($\approx 60\%$) was obtained after five equivalents of thiodepsipeptide 1 being completely consumed in the reaction. Owing to the instability of the thiodepsipeptide, one molar equivalent of the labelling reagent was added to the reaction every thirty minutes instead of adding them all at once. This strategy improved the ubiquitin labeling to 82% yield with five molar peptide equivalents after 2.5 h (Figure 2).

Comparatively, $<10\%$ ligation product was observed when ligating ubiquitin with LPETGG peptide when using sortase A. This result suggests that the N-terminal of ubiquitin may be buried and less accessible for ligation than small peptides, a finding in agreement with previous findings that sortase A is unable to modify myoglobin and fly pumilio RNA binding domain.^[8c] Using the same strategy, 70% of GFP was labelled with four molar peptide equivalents (Figure S3). This is also first example to explore the application of butelase 1 for protein labeling.

To determine if we could improve the ligation yield by introducing a linker to the N-terminus of proteins, we

expressed another recombinant ubiquitin with a short linker peptide GISGSGS. We obtained quantitative labeling for ubiquitin ($\approx 95\%$) with four equivalents of the thiopeptide **1** in the presence of 0.001 molar equivalents of butelase 1 in 100 min (Figure 3a). The ligation yield also reached $> 90\%$ yield for sortase A after introducing the linker peptide (GGSGSGS), a result consistent with previous reports.^[8c,f] However, sortase A required 0.1 molar equivalent of enzyme to catalyze the ligation, compared to 0.001 molar equivalent of butelase 1.

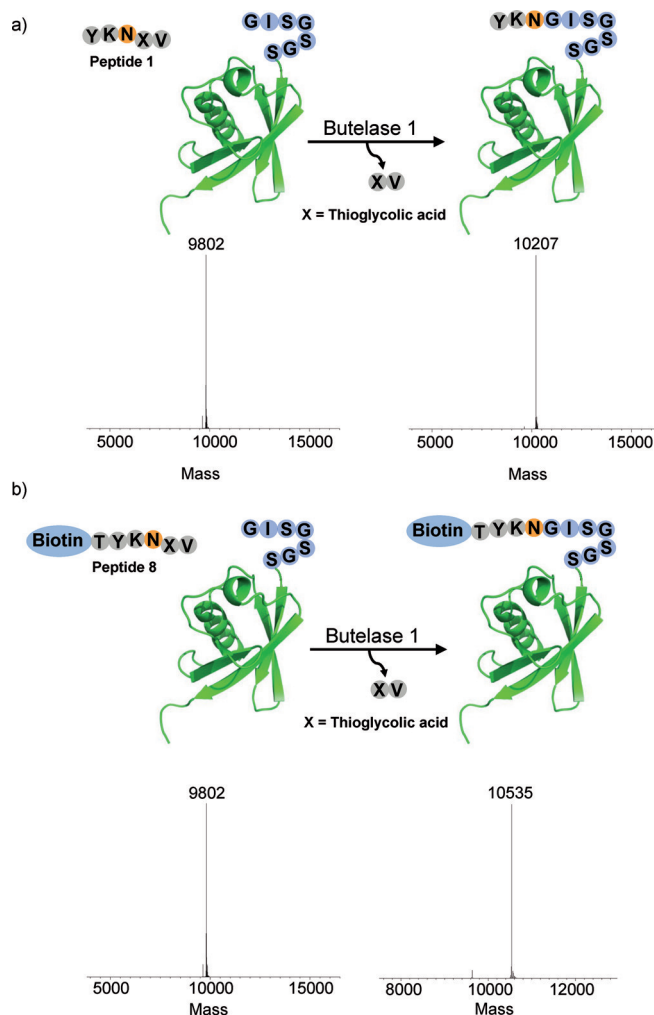


Figure 3. a) Illustration of the butelase-mediated N-terminal ligation of ubiquitin modified with a short linker peptide by the use of thiopeptide **1**. A quantitative ligation yield was obtained, as shown by the deconvoluted ESI spectra. b) Illustration of the butelase-mediated N-terminal modification of ubiquitin by the use of thiopeptide **8** which carries a biotin tag. A quantitative ligation yield was obtained with four equivalent of peptide **8** and 0.001 molar equivalent of butelase 1.

With the success on the model thiopeptide, conjugation with biological functional groups was also investigated. Thiopeptide **8** carrying a biotin group at its N-terminus was synthesized. Successful labelling was achieved on both ubiquitin and GFP with the biotinylated thiopep-

sipeptide **8** (Figure 3b; Supporting Information, Figure S5), demonstrating the applicability of our method to introduce functional tags into peptides and proteins. The labelling yields are comparable between thiopeptides **1** and **8**, which indicates that the biotin group does not affect the labelling efficiency.

In conclusion, we have developed a method for butelase-mediated ligation using thiopeptides as substrates. The ligation yield of $> 95\%$ could be achieved for the model peptide and ubiquitin with a small excess of substrate. The developed method is also applicable for introducing a functional group such as biotin into proteins. The ligation efficiency was greatly improved as the (S)GV byproduct is a poor recognizing substrate in the reverse direction. Furthermore, the preparation of the labelling reagent and the use of this method are simple and straight forward. We anticipate a broad application of this method for N-terminal modification of peptides and proteins.

Acknowledgements

This work was supported by the Singapore National Research Foundation NRF-CRP8-2011-05.

Keywords: butelase 1 · labelling · ligation · thiopeptide

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 15694–15698
Angew. Chem. **2015**, *127*, 15920–15924

- [1] a) T. Matsumoto, S. Sawamoto, T. Sakamoto, T. Tanaka, H. Fukuda, A. Kondo, *J. Biotechnol.* **2011**, *152*, 37; b) T. Sakamoto, S. Sawamoto, T. Tanaka, H. Fukuda, A. Kondo, *Bioconjugate Chem.* **2010**, *21*, 2227; c) J. M. Antos, G. M. Miller, G. M. Grotenbreg, H. L. Ploegh, *J. Am. Chem. Soc.* **2008**, *130*, 16338.
- [2] Y. Kobashigawa, H. Kumeta, K. Ogura, F. Inagaki, *J. Biomol. NMR* **2009**, *43*, 145.
- [3] a) J. M. Chalker, B. G. Davis, *Curr. Opin. Chem. Biol.* **2010**, *14*, 781; b) D. P. Gamblin, P. Garnier, S. van Kasteren, N. J. Oldham, A. J. Fairbanks, B. G. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 828–833; *Angew. Chem.* **2004**, *116*, 846–851; c) F. Li, A. Allahverdi, R. Yang, G. B. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiöld, C. F. Liu, *Angew. Chem. Int. Ed.* **2011**, *50*, 9611–9614; *Angew. Chem.* **2011**, *123*, 9785–9788; d) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776; e) K. K. Pasunooti, R. Yang, S. Vedachalam, B. K. Gorityala, C. F. Liu, X. W. Liu, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6268; f) R. Yang, K. K. Pasunooti, F. Li, X. W. Liu, C. F. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 13592; g) R. Yang, K. K. Pasunooti, F. Li, X. W. Liu, C. F. Liu, *Chem. Commun.* **2010**, *46*, 7199.
- [4] D. P. Baker, E. Y. Lin, K. Lin, M. Pellegrini, R. C. Petter, L. L. Chen, R. M. Arduini, M. Brickelmaier, D. Wen, D. M. Hess, L. Chen, D. Grant, A. Whitty, A. Gill, D. J. Lindner, R. B. Pepinsky, *Bioconjugate Chem.* **2006**, *17*, 179.
- [5] a) J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, *Angew. Chem. Int. Ed.* **2006**, *45*, 5307–5311; *Angew. Chem.* **2006**, *118*, 5433–5437; b) N. Jain, S. W. Smith, S. Ghone, B. Tomczuk, *Pharm. Res.* **2015**, *32*, 3526.
- [6] a) H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670; b) M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, *Nat. Chem. Biol.* **2007**, *3*, 707; c) M. W. L. Popp, H. L. Ploegh, *Angew. Chem. Int. Ed.* **2011**, *50*, 5024–5032; *Angew. Chem.* **2011**, *123*, 5128–5137; d) S. Pritz,

- Mini-Rev. Org. Chem.* **2008**, *5*, 47; e) S. Pritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyermann, *J. Org. Chem.* **2007**, *72*, 3909; f) T. Proft, *Biotechnol. Lett.* **2010**, *32*, 1; g) S. Samantaray, U. Marathe, S. Dasgupta, V. K. Nandicoori, R. P. Roy, *J. Am. Chem. Soc.* **2008**, *130*, 2132; h) T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, *ChemBioChem* **2008**, *9*, 802; i) S. Tsukiji, T. Nagamune, *ChemBioChem* **2009**, *10*, 787; j) S. Liebscher, M. Schopf, T. Aumüller, A. Sharkhuukhen, A. Pech, E. Hoss, C. Parthier, G. Jahreis, M. T. Stubbs, F. Bordusa, *Angew. Chem. Int. Ed.* **2014**, *53*, 3024–3028; *Angew. Chem.* **2014**, *126*, 3068–3072.
- [7] a) P. Cossart, R. Jonquieres, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5013; b) S. Dramsi, S. Magnet, S. Davison, M. Arthur, *FEMS Microbiol. Rev.* **2008**, *32*, 307; c) L. A. Marraffini, A. C. Dedent, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 192; d) S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, *Science* **1999**, *285*, 760; e) S. K. Mazmanian, H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2001**, *40*, 1049; f) H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12424.
- [8] a) J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, G. M. Grotenbreg, M. W. Popp, H. L. Ploegh, *J. Am. Chem. Soc.* **2009**, *131*, 10800; b) F. Liu, E. Y. Luo, D. B. Flora, A. R. Mezo, *J. Org. Chem.* **2014**, *79*, 487; c) D. J. Williamson, M. A. Fascione, M. E. Webb, W. B. Turnbull, *Angew. Chem. Int. Ed.* **2012**, *51*, 9377–9380; *Angew. Chem.* **2012**, *124*, 9511–9514; d) Y. Yamamura, H. Hirakawa, S. Yamaguchi, T. Nagamune, *Chem. Commun.* **2011**, *47*, 4742; e) C. P. Hackenberger, D. Schwarzer, *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074; *Angew. Chem.* **2008**, *120*, 10182–10228; f) C. S. Theile, M. D. Witte, A. E. M. Blom, L. Kundrat, H. L. Ploegh, C. P. Guimaraes, *Nat. Protoc.* **2013**, *8*, 1800.
- [9] a) G. K. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian, J. P. Tam, *Nat. Chem. Biol.* **2014**, *10*, 732; b) Y. Cao, G. K. Nguyen, J. P. Tam, C. F. Liu, *Chem. Commun.* **2015**, DOI 10.1039/C5CC07227A; c) G. K. Nguyen, S. Zhang, N. T. Nguyen, P. Q. Nguyen, M. S. Chiu, A. Hardjojo, J. P. Tam, *J. Biol. Chem.* **2011**, *286*, 24275.
- [10] a) D. J. Craik, N. L. Daly, T. Bond, C. Waine, *J. Mol. Biol.* **1999**, *294*, 1327; b) G. K. Nguyen, Y. Lian, E. W. Pang, P. Q. Nguyen, T. D. Tran, J. P. Tam, *J. Biol. Chem.* **2013**, *288*, 3370; c) G. K. Nguyen, W. H. Lim, P. Q. Nguyen, J. P. Tam, *J. Biol. Chem.* **2012**, *287*, 17598; d) I. Saska, A. D. Gillon, N. Hatsugai, R. G. Dietzgen, I. Hara-Nishimura, M. A. Anderson, D. J. Craik, *J. Biol. Chem.* **2007**, *282*, 29721.
- [11] X. Q. Li, T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **1998**, *39*, 8669.

Received: July 23, 2015

Revised: October 21, 2015

Published online: November 13, 2015