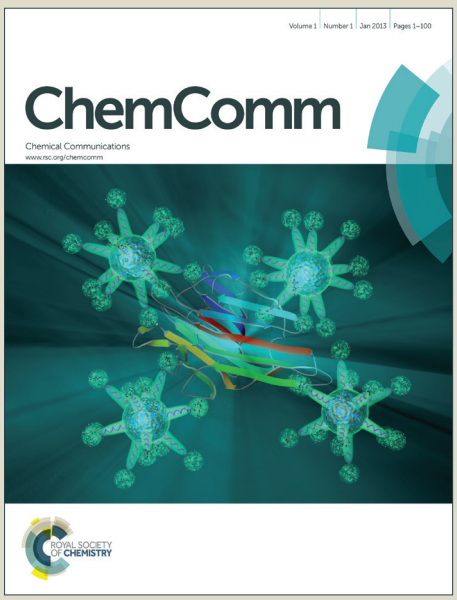
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

ChemComm

COMMUNICATION

Total chemical synthesis of the site-selective azide-labeled [I66A]HIV-1 protease

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yun-Kun Qi,[‡ab] Hao-Nan Chang,[‡ab] Kai-Mai Pan[b] and Ji-Shen Zheng*[ab]

The first total chemical synthesis of the site-selective azide-labeled [I66A]HIV-1 protease is described by native chemical ligation. Chemical synthesis of azide-labeled proteins would provide useful protein tools for biochemical, biophysical or medical studies.

Proteins modified with artificial probes are attractive therapeutics, diagnostic agents and imaging tools for biophysical and biomedine studies.[1] Generally, to install such modifications, three main kinds of reactive functional groups are involved.[2] The first functional groups are the native reactive groups (amine, carboxyl, thiol, *etc*) exposed on protein surfaces.[3] The heterogeneity and randomness of this method may influence the protein architecture and/or biological activity.[4] The second reactive handles involve the recombinant fusion tags to the amino terminus (N-terminus) or carboxyl terminus (C-terminus) of the protein.[5] The protein with fusion tags could then be covalently conjugated to support surface through enzyme such as sortase.[6] However, for some proteins, the N- or C-terminus of the protein is wrapped inside the protein, which may cause low covalent efficiency.[2] The third reactive handles for the site-selective protein modification are the unnatural orthogonal ''click'' groups through orthogonal aminoacyl-tRNA synthetase–tRNA pairs.[7] The most popular ''click'' reaction is copper-catalysed Huisgen [2+3] azide-alkyne cycloaddition reported by Sharpless. The nearly 100% yield and a broad range of solvent tolerance have made the azide-alkyne click chemistry an excellent candidate for the site-specific labeling of proteins.[8] Nonetheless, bioengineering technology often suffers from low yield and incomplete post-translational modifications.[8c]

Total chemical synthesis, which enables introduction of unnatural amino acids with more flexibility and precision[9], is an intriguing alternative approach to install azide/alkyne groups on proteins. Moreover, chemical synthesis can produce proteins (e.g., post-translationally modified proteins[9a-c], mirror-image proteins[10a,b], photon-activatable proteins[10c]) that are difficult to obtain using recombinant technology. Indeed, solid phase peptide synthesis (SPPS) permitted the synthesis of azide/alkyne labeled peptides for peptide immobilizing, polyethylene glycol (PEG)ylation, stapling or ^{18}F-labeling etc.[11] However, SPPS technology can only assemble polypeptide chains containing 40-60 amino acids.[9e,12] To the best of our knowledge, there has been no report on chemical synthesis of azide/alkyne-labeled proteins through chemoselective ligation[13] of multiple peptide segments. It is envisaged that chemical ligation will be an important method for large azide/alkyne-labeled proteins.

Herein, we reported the first total chemical synthesis of site-specifically azide-labeled HIV-1 protease.[14] It was found that the azide and alkyne groups were not compatible with the free-radical-based desulfurization reaction which has been widely applied in protein chemical synthesis.[15] Combining the optimized peptide ligation and azide incorporation new method, the total synthesis of azide-labeled [I66A]HIV-1 protease was achieved. The enzymatic activity of the synthetic azide-labeled HIV-1 [I66A]protease was demonstrated by the autocleavage and substrate hydrolysis assay.

Our initial synthetic design was to introduce azide group into HIV-1 protease. HIV-1 protease is a 21.4 kDa retroviral aspartyl protease consisting of two identical 99-residue monomers.[14] Inhibition of HIV-1 protease dramatically disrupts HIV replication and infection, which makes HIV-1 protease an extremely attractive drug target.[16,17] To prepare the azide-labeled HIV-1 protease, the 99-residue protease was divided into three peptide segments: PR[Pro^{1}–Gly^{27}]-NHNH$_2$ **1**, PR[Cys^{28}–Lys(N$_3$)^{41}–Lys^{70}]-NHNH$_2$ **2** and PR[Cys^{71}–Phe^{99}]-NH$_2$ **3**. The three peptide segments were obtained through 9-fluorenylmethyloxycarbonyl (Fmoc) SPPS.[12] Ala^{28} and Ala^{71} were mutated to Cys, which could be converted back to Ala through a ligation-desulfurization strategy (Figure **1A**).[15] Fmoc-Lys(N$_3$)-OH (Figure **1C**) was used to introduce the azide group at Lys^{41} which is far away from the active site and on the edge of the protease according to the HIV-1 protease X-ray crystal structure.[14] The thiols of Cys^{67} and Cys^{95} were protected by acetamidomethyl (Acm) group to enable the selective desulfurization of Cys^{28} and Cys^{71} in the full-length polypeptide. The Acm groups of Cys^{67} and Cys^{95} could then be removed by AgOAc[18] to afford the azide-labeled HIV-1 protease. The three peptide segments were assembled through N-to-C sequential native chemical ligations (NCL) of peptide

[a] *High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230031, China. E-mail: jszheng@hmfl.cas.cn.*
[b] *Department of Chemistry,Tsinghua University, Beijing 100084, China*
† Electronic Supplementary Information (ESI) available: Experimental section and supplementary figures. See DOI: 10.1039/x0xx00000x
‡ These authors contributed equally to this work.

ChemComm

COMMUNICATION

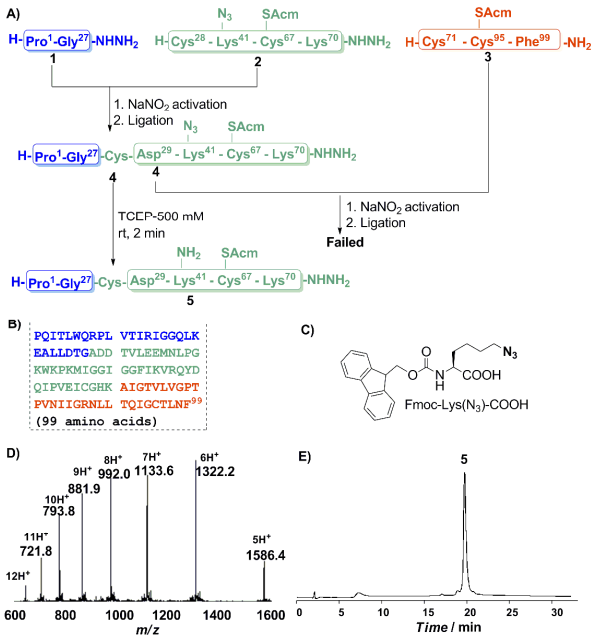


Figure 1. The synthetic scheme for azide-labeled HIV-1 protease. **A)** Synthetic route. **B)** The amino acid sequence of 99-residue HIV-1 protease monomer. **C)** The molecular structure of Fmoc-Lys(N_3)-OH. **D)** The ESI-MS spectrum of purified Peptide **5**. The spectrum gave an observed mass of 7928.3 Da (calcd 7929.2 Da, average isotopes). **E)** Analytic RP-HPLC chromatogram of purified **5**.

hydrazide.[12,19] Peptide **1** (1 equiv., 2 mM) was converted to the thioester by *in situ* $NaNO_2$ activation and thiolysis, and then reacted with peptide **2** (1.5 equiv.) to afford the desired product **4** in 48% isolated yield at pH 6.6-6.8 within 5 h (Supporting Information). However, the next ligation between peptide **4** and **3** failed to give satisfactory yield because of the poor stability of the peptide thioester which was *in situ* converted from peptide **4** and the favorable aggregation of peptide **3** in ligation buffer (Supporting Information).[17] To increase the solubility of peptide **3**, PR[Cys[71]–Arg[115]]-NH_2 **7** was synthesized with a solubilizing Arginine (Arg)-tag at C-terminus. Besides, it was found that although the azide group could survive the Tris(2-carboxyethyl)phosphine (TCEP) during NCL with a low TCEP concentration and a quick operation (< 5min), it could be reduced into amino group under high concentration of TCEP (500 mM) conditions for desulfurization (Figure **1D**). The results indicated that the azide group was not compatible with the free-radical-based desulfurization reaction.

Another route to labeled HIV-1 protease was to incorporate the alkyne group. Fmoc-Pra-OH (Fmoc-Propargylglycine) was used to directly introduce the alkyne functional group at Lys[41] during Fmoc SPPS (Figure **2C**). By using the peptide assemble strategy described in Figure **2A**, the peptide **1** (1 equiv.) was first reacted with **6** (1.5 equiv.) to generate peptide **8** with 52% isolated yield (Supporting Information). Peptide **8** (1 equiv.) was then reacted with peptide **7** (2 equiv.) to afford the product **9** in 43% isolated yield (Supporting Information). Remarkably, with six arginine residues at the monitored by analytical RP-HPLC trace. Unfortunately, MALDI-TOF mass analysis of peptide **10** (Observed mass 13117.1 Da) showed an increase of 250.4 Da compared to the desired desulfurization product (Calcd mass 12866.7 Da, average isotopes). To confirm the modification site and the source of the modification, Fmoc-Pra-OH was treated at the free-radical-based desulfurization conditions.The product was characterized by ^{1}H-NMR, ^{31}P-NMR, and high-resolution mass spectrometry. The results indicated that the vinylphosphonium salt was formed (Figure **2E**, Supporting Information). Three different conditions were tested to prevent the modification. But those attempts failed (Supporting Information).

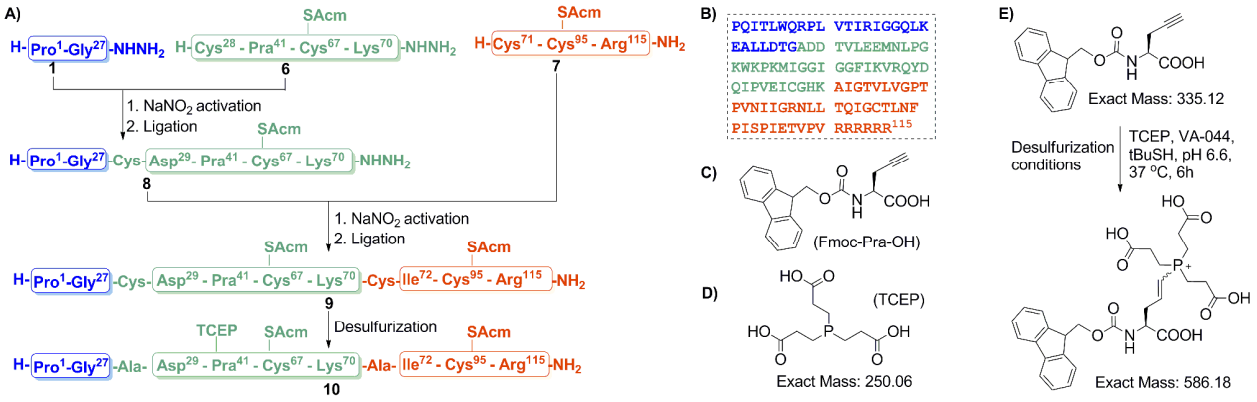


Figure 2. The synthetic scheme for alkyne-labeled HIV-1 protease. **A)** Synthetic route. **B)** The amino acid sequence of HIV-1 protease precursor. **C)** The molecular structure of Fmoc-Pra-OH. **D)** The molecular structure of TCEP. **E)** The reaction between alkyne and TCEP.

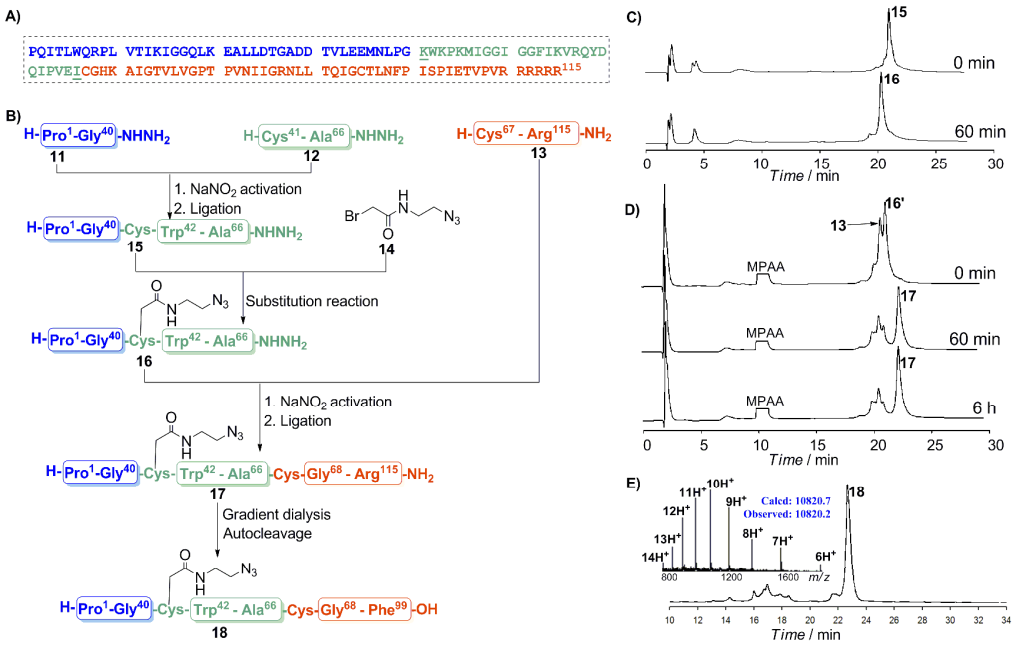


Figure 3. Total chemical synthesis of azide-labeled [I66A]HIV-1 protease. **A)** The amino acid sequence of [I66A]HIV-1 protease precursor. **B)** Synthetic route for azide-labeled [I66A]HIV-1 protease. **C)** The substitution reaction between peptide **15** and compound **14**. 0 and 60 min after the addition of **14**. **D)** The RP-HPLC traces of ligation of peptide **16** and **13**. **E)** Analytic RP-HPLC chromatogram (λ = 214 nm) of protein folding (after the overnight dialysis, crude folding solution). Inner: the ESI-MS spectrum of the objective protease **18**.

The above results showed that neither azide nor alkyne group was compatible with the desulfurization conditions. To overcome the problem, we optimized a new peptide ligation and azide/alkyne incorporation strategy. In the new method, the 115-residue [I66A]HIV-1 protease precursor was divided into three peptide segments, i.e., PR[Pro^{1}–Gly^{40}]-NHNH$_2$ **11**, PR[Cys^{41}–Ala^{66}]-NHNH$_2$ **12** and PR[Cys^{67}–Arg^{115}]-NH$_2$ **13**. It was worth mentioning that Lys^{41} was replaced by Cys to provide an active site for the peptide ligation and the following bioconjugation reaction with functional groups (Figure **3A** and **3B**). The sterically hindered Ile^{66} was mutated to Ala to increase the efficiency of peptide ligation. Note that the azide and alkyne would both be added to polypeptide through a highly efficient substitution reaction between the thiol group of cysteine and brominated derivatives.[21]

To assemble the 115-residue azide-labeled HIV-1 protease precursor, we carried out an N-to-C sequential ligation of three peptide segments. First, peptide **11** (1 equiv.) was reacted with **12** (2 equiv.) to afford the ligation product PR[Pro^{1}–Ala^{66}]-NHNH$_2$ **15** within 5 h, with an isolated yield of 43% (Supporting Information). Subsequently, a synthetic azide-containing compound **14** (15 equiv., Supporting Information) was selectively installed to the Cys^{41} residue of purified peptide **15** (1 equiv.) at room temperature within 1 h (Figure **3C**). The site-selective azide-labeled peptide PR[Pro^{1}–Cys(N$_3$)^{41}–Ala^{66}]-NHNH$_2$ **16** was obtained with 55% isolated yield. The final ligation between peptide **16** (1 equiv.) and **13** (2 equiv.) proceeded smoothly to generate the desired [I66A]HIV-1 protease precursor PR[Pro^{1}–Cys(N$_3$)^{41}–Arg^{115}]-NH$_2$ **17** within 6 h, with an isolated yield of 41% (Figure **3D**). Using the optimized peptide ligation and azide incorporation strategy, the azide-labeled HIV-1 protease precursor was synthesized with high-efficiency.

To fold the azide-labeled HIV-1 protease, we conducted the gradient dialysis experiment.[17] After folding, the synthesized HIV-1 protease precursor [I66A]PR[Pro^{1}–Cys(N$_3$)^{41}–Arg^{115}]-NH$_2$ **17** can automatically cleave the C-terminal 16 amino-acid residues and then become the mature active [I66A]HIV-1 protease, just as the HIV-1 protease did naturally (Supporting Information). After 12 h dialysis at 4 °C, ESI-MS analysis of the folding solution showed that reverse transcriptase sequence with six arginines was cleaved successfully between residues Phe^{99} and Pro^{100}. PR[Pro^{1}–Cys(N$_3$)^{41}–Phe^{99}]-OH **18** with auto-cleavage activity was obtained (Figure **3B**). The ESI-MS spectrum and HPLC trace of azide-labeled [I66A]HIV-1 protease were shown in Figure **3E**.

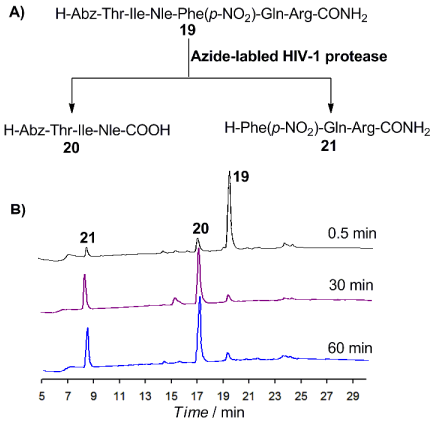


Figure 4. A) Substrate peptide **19** hydrolysis by synthetic azide-labeled HIV-1 protease. Abz = 2-aminobenzoyl, Nle = norleucine, Phe(p-NO$_2$) = p-nitrophenylalanine. **B)** Hydrolysis process monitored by RP-HPLC after the incubation of 0.5 min, 30 min and 60 min.

To examine the enzymatic activity of the synthetic azide-labeled HIV-1 protease, the substrate hydrolysis assay was performed. A peptide substrate **19** of HIV-1 protease was synthesized (Figure **4**).[22] Analytic RP-HPLC and ESI-MS were used to monitor the hydrolysis procedure of synthetic azide-labeled and commercial recombinant HIV-1 protease. As shown in Figure **4B**, hydrolysis of substrate **19** occurred immediately after the addition of the synthetic HIV-1 protease **18**. Peptide **19** was almost hydrolyzed within 30 min. According to the RP-HPLC analysis and mass spectrometry data, the azide-labeled HIV-1 protease showed good enzymatic hydrolysis activity of peptide **19**. According to the enzymatic activity assay, synthetic azide-labeled and commercial recombinant HIV-1 protease showed similar hydrolysis acitvity (Supporting Information). The data indicated that the bioactive azide-labeled [I66A]HIV-1 protease was produced by the new synthetic strategy.

In summary, our study indicated that the azide and alkyne functional groups were not compatible with the free-radical-based desulfurization reaction which often was employed in protein chemical synthesis. With the optimized peptide ligation and the site-selective azide/alkyne incorporation strategy, the first total chemical synthesis of azide-labeled [I66A]HIV-1 protease with enzymatic activity was achieved. We anticipate that the new method provides a reliable access to azide/alkyne-labeled proteins by total chemical synthesis. The chemical synthesis of azide/alkyne-labeled proteins would provide useful protein molecules for protein stapling,[23] the covalent immobilization[1] (biosensors, microarrays), therapeutics[24] (PEGylation), or diagnostic tools[25] (^{18}F-labeled protein probes).

This work was supported by the National Natural Science Foundation of China (grant number 21402206), Natural Science Foundation of Anhui Province (1508085QB30) and Beijing National Laboratory for Molecular Sciences (BNLMS).

Notes and references

1 (a) P. Jonkheijm, D. Weinrich, H. Schroder, C. M. Niemeyer and H. Waldmann, *Angew. Chem. Int. Ed.*, 2008, **47**, 9618; (b) D. Guan, Y, Kurra, W. Liu and Z. Chen, *Chem. Commun.,* 2015, **51**, 2522.

2 (a) L. S. Wong, F. Khan and J. Micklefield, *Chem. Rev.*, 2009, **109**, 4025; (b) C. D. Spicer and B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740.

3 (a) E. Steen Redeker, D. T. Ta, D. Cortens, B. Billen, W. Guedens and P. Adriaensens, *Bioconjugate Chem.*, 2013, **24**, 1761; (b) R. Kundu and Z. T. Ball, *Chem. Commun.* 2012, **2**, 4166.

4 D. Brady and J. Jordaan, *Biotechnol. Lett.*, 2009, **31**, 1639.

5 L. S. Wong, J. Thirlway and J. Micklefield, *J. Am. Chem. Soc.,* 2008, **130**, 12456.

6 (a) T. Lee, J. Min, H. Hirakawa, T. Nagamune and J. W. Choi, Colloids Surf., B 2014, **120**, 215; (b) B. Kuropka, N. Royla, C. Freund and E. Krause, *Proteomics*, 2015, **15**, 1230; (c) F. Clow, J. D. Fraser and T. Proft, *Biotechnol. Lett.*, 2008, **30**, 1603.

7 (a) C. H. Kim, J. Y. Axup and P. G. Schultz, *Curr. Opin. Chem. Biol.*, 2013, **17**, 412; (b) C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413.

8 (a) H. Neumann, K. H. Wang, L. Davis, M. Garcia-Alai and J. W. Chin, *Nature*, 2010, **464**, 441; (b) B. K. Raliski, C. A. Howard and D. D. Young, *Bioconjugate Chem.*, 2014, **25**, 1916. (c) C. Liu, P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413.

9 (a) B. Fauvet, S. M. Butterfield, J. Fuks, A. Brik and H. A. Lashuel, *Chem. Commun.*, 2013, **49**, 9254; b) P. Wang, S. Dong, J. H. Shieh, E. Peguero, R. Hendrickson, M. A. Moore and S. J. Danishefsky, *Science*, 2013, **342**, 1357; (c) M. M. Muller and T. W. Muir, *Chem. Rev.*, 2015, **115**, 2296; (d) V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596; (e) Y. X. Chen, G. Triola and H. Waldmann, *Acc. Chem. Res.*, 2011, **44**, 762.

10 (a) M. Avital-Shmilovici, K. Mandal, Z. P. Gates, N. B. Phillips, M. A. Weiss and Kent, S. B. H. *J. Am. Chem. Soc.* **2013**, *135*, 3173; (b) M. T. Weinstock, M. T. Jacobsen and M. S. Kay, *Proc. Natl. Acad. Sci. U S A.*, 2014, **111**, 11679; c) X. Chen, S. Tang, J. S. Zheng, R. Zhao, Z. P. Wang, W. Shao, H. N. Chang, J. Y. Cheng, H. Zhao, L. Liu and H. Qi, *Nat. Commun.*, 2015, **6**, 7220.

11 (a) M. Dobkowski, A. Szychowska, M. Pieszko, A. Miszka, M. Wojciechowska, M. Alenowicz, J. Ruczynski, P. Rekowski, L. Celewicz, J. Barciszewski and P. Mucha, *J. Pept. Sci.*, 2014, **20**, 696; (b) A. I. Fernandez-Llamazares, J. Garcia, J. Adan, D. Meunier, F. Mitjans, J. Spengler and F. Albericio, *Org. Lett.*, 2013, **15**, 4572; (c) Y. H. Lau, Y. Wu, P. de Andrade, W. R. Galloway and D. R. Spring, *Nat. Proto.*, 2015, **10**, 585; (d) L. S. Campbell-Verduyn, L. Mirfeizi, A. K. Schoonen, R. A. Dierchx, P. H. Elsinga and B. L. Feringa, *Angew. Chem., Int. Ed.*, 2011, **50**, 11117.

12 (a) G. M. Fang, Y. M. Li, F. Shen, Y. C. Huang, J. B. Li, Y. M. Lin, H. K. Cui, and L. Liu, *Angew. Chem. Int. Ed.* 2011, **50**, 7645; (b) J. S. Zheng, S. Tang, Y. K. Qi, Z. P. Wang and L. Liu, *Nat. Protoc.*, 2013, **8**, 2483.

13 (a) P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776; (b) V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471; (c) J. C. Monbaliu and A. R. Katritzky, *Chem. Commun.*, 2012, **48**, 11601. (d) L. Raibaut, N. Ollivier and O.Melnyk, *Chem. Soc. Rev.*, 2012, **41**, 7001.

14 M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent and A. Wlodawer, *Science*, 1989, **246**, 1149-1152.

15 Q. Wan and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2007, **46**, 9248.

16 A. K. Ghosh, G. E. Schiltz, L. N. Rusere, H. L. Osswald, D. E. Walters, M. Amano and H. Mitsuya, *Org. Biomol. Chem.*, 2014, **12**, 6842.

17 E. C. B. Johnson, E. Malito, Y. Q. Shen, D. Rich, W. J. Tang and S. B. H. Kent, *J. Am. Chem. Soc.*, 2007, **129**, 11480.

18 B. L. Pentelute and S. B. H. Kent, *Org. Lett.*, 2007, **9**, 687.

19 G. M. Fang, J. X. Wang and L. Liu, *Angew. Chem. Int. Ed.*, 2012, **51**, 10347.

20 J. S. Zheng, M. Yu, Y. K. Qi, S. Tang, F. Shen, Z. P. Wang, L. Xiao, L. Zhang, C. L. Tian and L. Liu, *J. Am. Chem. Soc.*, 2014, **136**, 3695.

21 V. Y. Torbeev and S. B. Kent, *Angew. Chem.*, *Int. Ed.*, 2007, **46**, 1667.

22 Z. Szeltner and L. Polgar, *J. Biol. Chem.*, 1996, **271**, 32180.

23 (a) D. M. Abdeljabbar, F. J. Piscotta, S. Zhang and Link A. James, *Chem. Commun.*, 2014, **50**, 14900; (b) X.-H. Chen, Z. Xiang, Y. S. Hu, V. K. Lacey, H. Cang and L. Wang, *ACS Chem. Biol.*, 2014, **9**, 1956.

24 (a) P. Milla, F. Dosio, and L. Cattel, *Curr. Drug Metab.*, 2011, **13**, 105; (b) N. Nischan and C. P. Hackenberger, *J. Org. Chem.*, 2014, **79**, 10727.

25 (a) F. Wuest, M. Berndt, R. Bergmann, J. van den Hoff, and J. Pietzsch, *Bioconjugate Chem.*, 2008, **19**, 1202; (b) J. D. Way, C. Bergman and F. Wuest, *Chem. Commun.*, 2015, **51**, 3838.

