In Vivo Site-Specific Protein Tagging with Diverse Amines Using an Engineered Sortase Variant

Jeff E. Glasgow,[†] Marc L. Salit,[†] and Jennifer R. Cochran*,[‡]

[†]National Institute of Standards and Technology, Stanford, California 94305, United States
[‡]Departments of Bioengineering and Chemical Engineering, Stanford University, Stanford, California 94305, United States

S *Supporting Information*

ABSTRACT: Chemoenzymatic modification of proteins is an attractive option to create highly specific conjugates for therapeutics, diagnostics, or materials under gentle biological conditions. However, these methods often suffer from expensive specialized substrates, bulky fusion tags, low yields, and extra purification steps to achieve the desired conjugate. *Staphylococcus aureus* sortase A and its engineered variants are used to attach oligoglycine derivatives to the C-terminus of proteins expressed with a minimal LPXTG tag. This strategy has been used extensively for bioconjugation in vitro and for protein–protein conjugation in living cells. Here we show that an enzyme variant recently engineered for higher activity on oligoglycine has promiscuous activity that allows proteins to be tagged using a diverse array of small, commercially available amines, including several bioorthogonal functional groups. This technique can also be carried out in living *Escherichia coli*, enabling simple, inexpensive production of chemically functionalized proteins with no additional purification steps.

Site-specific modification of proteins is an essential technique in many scientific fields.[1,2] As an example, the efficacy of antibody–drug conjugates, a therapeutic approach to cancer treatment, is enhanced by the inherent uniformity that stems from site-specific attachment of small-molecule chemotherapeutics.[3] Other in vitro protein conjugates for use in materials,[4] imaging,[5] diagnostics,[6] catalysis,[7] or devices[8] can similarly benefit from the homogeneity of site-specific conjugation. Chemical methods for modifying proteins have historically relied on the different reactivities of specific amino acids, e.g., lysine, cysteine, and tyrosine; however, in recent years significant advances have been made to modify unique sites such as N-terminal residues,[9] C-terminal residues,[10] or glycosylated residues.[11]

In vivo tagging of proteins, though challenging, can be used to illuminate protein localization, function, and intermolecular interactions or to allow modified protein production in fewer steps than other approaches.[12] Amber stop codon suppression using unnatural amino acids (UAAs), one of the most heavily used methods for in vivo protein labeling, can install a multitude of different functional groups in a wide variety of cell types; however, this method can be prone to decreased protein yield, truncation, and misincorporation.[13] Enzyme fusions that employ mechanistic-based protein labeling such as SNAP,

CLIP, TMP, and Halo tags feature exquisite specificity but are limited by their molecular size and expensive substrates.[14] Numerous other chemoenzymatic methods developed to ligate orthogonal functional group adaptors allow for smaller size and greater versatility of tags. A host of natural and engineered enzymes, such as 4′-phosphopantetheinyl transferase (Sfp),[15] glutathione-S-transferase (GST),[16] transglutaminase,[17] tubulin tyrosine ligase,[18] and phosphocholine transferase,[19] are used to attach functional groups in vitro. Alternatively, the enzymes N-myristoyl transferase,[20] biotin ligase,[21] lipoic acid ligase,[22] and formylglycine generating enzyme[23] have been coexpressed with a protein of interest fused to a recognition sequence so that they attach a unique functional group in the cytoplasm.

Another widely used chemoenzymatic bioconjugation approach utilizes the transpeptidase sortase A from *Staphylococcus aureus* to label the N- or C-terminus.[24] This enzymatic approach is popular because of its versatility, only requiring an LPXTG recognition sequence (the "C-peptide")[25] and an oligoglycine nucleophile (the "N-peptide") (see Figure 1A).[26,27] By means of this scheme, proteins have been attached to lipids,[28] nucleic acids,[29] polymers,[30] drugs,[31] inorganic materials,[32] surfaces,[33] thioesters,[34] depsipeptides,[35] and other proteins.[36] Because of the peptidic nature of the substrates, this approach has been largely limited to in vitro or cell-surface labeling, though non-natural protein–protein ligations have
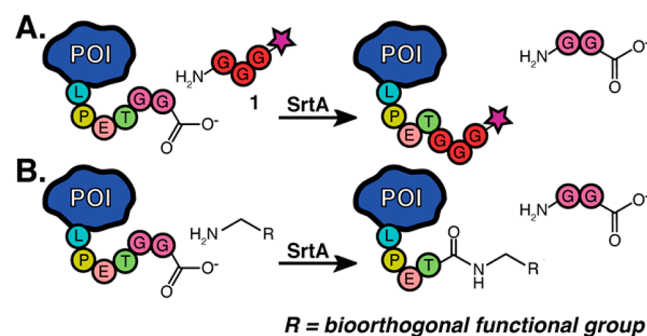
Figure 1. C-terminal transpeptidation of proteins using sortase A. (A) Classic sortase reaction. The enzyme recognizes an LPXTG motif on the protein of interest (POI) and replaces the terminal glycine with tagged oligoglycine **1**. (B) New sortase activity in which oligoglycine is replaced by an inexpensive amine containing a cell-permeable, bioorthogonal chemical handle.

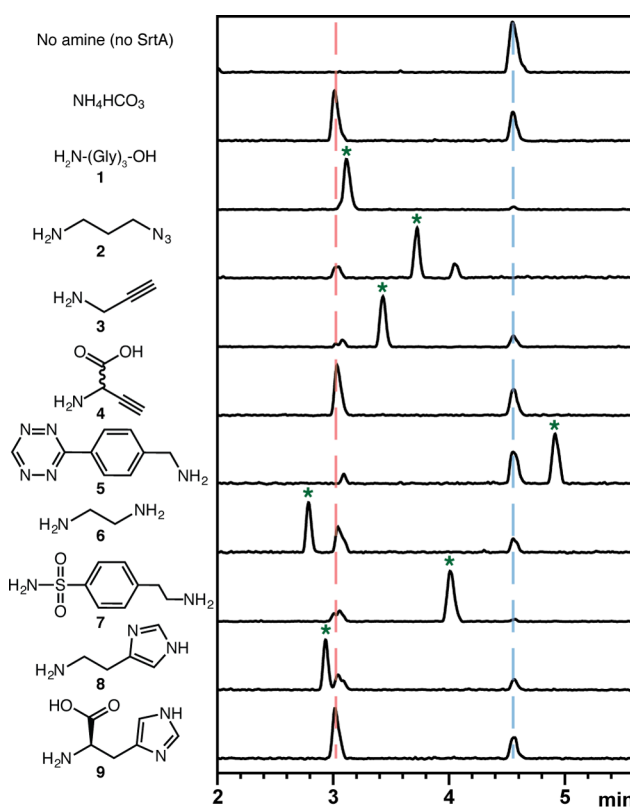been demonstrated in both live mammalian[37] and *E. coli*[38] cells. Central to this latter bacterial example were two sets of mutations engineered into the sortase enzyme (termed 7M). One set, [P94R/D160N/D165A/K190E/K196T],[39] dramatically increases the activity of the enzyme. Another set, [E105K/E108A], confers calcium independence to the enzymatic activity.[40] These previous studies highlight the vast potential of sortase, but the requirement of synthetic peptide substrates limits the majority of applications to lab-scale in vitro labeling in research groups or organizations with the resources to custom-make the desired substrates. An inexpensive, cell-permeable, commercially available bioorthogonal adaptor would greatly extend the potential of sortase for producing large-scale and/or in vivo conjugates.

Here we describe the use of the engineered sortase variant 7M (SrtA7M) to create bioorthogonally tagged proteins directly from *E. coli* culture. By simply coexpressing the sortase variant with the protein of interest and adding commercially available substrate mimics, such as 3-azido-1-propanamine (Azp) or propargylamine (Figure 1B), we can produce large quantities of labeled protein with no extra purification steps. Additionally, we show that this method works in vivo on a variety of protein substrates.

Wild-type sortase enzymes have been used to install non-glycine nucleophiles such as aminohexoses,[41] lysine-containing sequences,[42] some amines,[26,43,44] and hydrazines.[45] We hypothesized that the engineered sortase variant SrtA7M would more efficiently activate LPETG sequences with little specificity for the nucleophile. We incubated purified SrtA7M with the peptide LPETGSW and several potentially useful amines and analyzed the reactions by LC/MS to determine which could act as nucleophiles in the transpeptidation reaction. Six of the eight amines tested, in addition to triglycine, showed significant conversion in just 2 h (Figure 2). The reaction was tolerant of the bioorthogonally reactive groups on Azp **2**, propargylamine (**3**) (but not DL-propargylglycine (**4**)), and tetrazine amine **5** in addition to charged ethylenediamine (**6**) and bulky aminoethylbenzenesulfonamide (**7**). Interestingly, the enzyme was able to act on histamine (**8**) but not histidine (**9**). The two amines that did not participate in the transpeptidation, **4** and **9**, are branched at the α-carbon, suggesting that the 7M variant maintains the wild-type enzyme's preference for the unbranched primary amine of oligoglycine. In the absence of a suitable amine, ammonia from the ammonium bicarbonate buffer was also able to add to the peptide (Figure 2). Curiously, the enzyme was more active in this buffer; however, significant activity was also seen in Tris buffer (Supplementary Figure S1). Minimal enzymatic activity was seen in phosphate buffer. We also tested the pH sensitivity of the reaction with propargylamine in Tris buffer and found it to work significantly better above pH 7.5 (Supplementary Figure S2). Wild-type SrtA was only able to catalyze the reaction when 15-fold more enzyme was added and incubated for 20 h, resulting in modest conversion (Supplementary Figure S3). These data show that the enzymatic activity, but not substrate specificity, has been altered with SrtA7M compared with wild-type sortase.

We next determined the ability of SrtA7M to modify purified proteins. We incubated several proteins containing a C-terminal LPETGG sequence (srt) with 10 μM enzyme and either 100 mM Azp or propargylamine, followed by quenching with formic acid. Figure 3 shows complete conversion of maltose binding protein (MBP-srt), anti-HER2 nanobody 5f7 (5f7-His-

Figure 2. Transpeptidation of peptides using SrtA7M. Peptide LPETGSW (1 mM, dashed blue line) was incubated with SrtA7M (20 μM) and amines **1**−**9** (10 mM) at 37 °C for 2 h in ammonium bicarbonate buffer (pH 7.8), diluted 10-fold with 0.2% formic acid, and analyzed by LC/quadrupole MS. The negative-mode base-peak chromatograms show significant conversion to the desired conjugates (indicated by green asterisks). Ammoniolysis (dashed red line) occurs in the absence of unbranched primary amines.
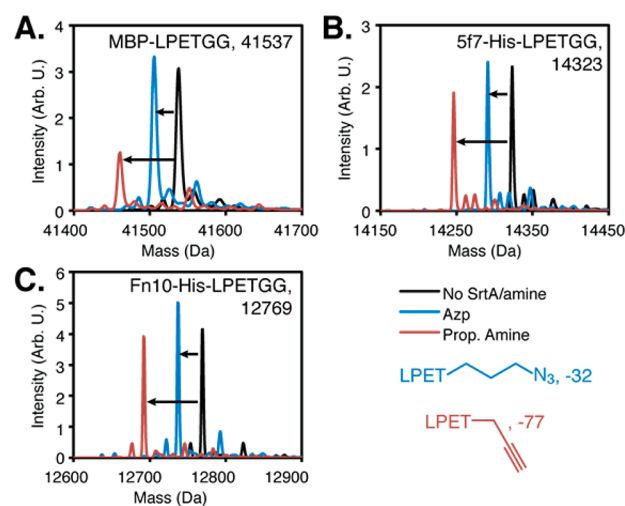
Figure 3. Transpeptidation of purified proteins. Purified proteins expressed with a C-terminal LPETGG sequence were incubated with 10 μM SrtA7M for 8 h at 37 °C along with the desired amine. Proteins were then diluted 10-fold with 0.2% formic acid and analyzed by ESI-TOF LC/MS. (A) Maltose binding protein, (B) 5f7 nanobody, and (C) fibronectin Fn10 all show 80−100% conversion to the desired conjugate.

srt),[46] and the engineered fibronectin domain Fn10 (Fn10-His-srt).[47] Additionally, little to no conversion was observed with

the superfolder GFP (His-sfGFP-srt) under the same conditions until the 6xHis tag was positioned as a spacer, allowing the enzyme access to the LPETGG sequence (sfGFP-His-srt) (Supplementary Figure S6). To define the effective range of reaction conditions, we tested the conjugation of MBP-srt with Azp at different time points and substrate concentrations. These experiments suggested that relatively high concentrations of the amine are needed but that efficient conjugation takes place in less than 1 h (Supplementary Figures S4 and S5).

As sortase has only rarely been used in living systems, we next determined whether SrtA7M is able to install useful functional groups to proteins as they were expressed in *E. coli*. SrtA7M, under a rhamnose-inducible promoter, was coexpressed in BL21(DE3) cells with the protein GST-His-srt, thioredoxin-fused nanobody (Trx-5f7-His-srt), or sfGFP-His-srt under T7-inducible promoters in the presence of Azp (Figure 4). The azide-tagged proteins were then labeled with Cy3-

Figure 4. Modification of proteins in living *E. coli* cells. SrtA7M and each protein of interest were coexpressed, and amine (25 mM) was added to the culture medium. Clarified lysate was then incubated with Cy3-DBCO. (A) Cy3 fluorescence and (B) Coomassie-stained SDS-PAGE gels show high levels of modification of sfGFP, GST, Trx-5f7 nanobody (before and after Ni^{2+}/NTA purification), and MBP. (C) Mass spectra of the superfolder GFP expressed alone or with SrtA7M and Azp or propargylamine and purified by Ni^{2+}/NTA chromatography, showing complete conversion to the desired conjugates. (D) Flow cytometry with Trx-5f7-Cy3. Trx-5f7 was conjugated to Azp in vivo, labeled with Cy3-DBCO on the Ni^{2+}/NTA column, and bound to HER2 molecules on SK-BR-3 cells. Blocked cells were incubated with 100-fold excess unlabeled Trx-5f7 to discern nonspecific Cy3 binding to the cell surface.

dibenzocyclooctyne (Cy3-DBCO) in cell lysate and assayed by SDS-PAGE with fluorescence and Coomassie staining (Figure 4A,B). The LPETGG-tagged proteins were specifically conjugated only when SrtA7M was coexpressed. Trx-5f7-His-srt was only modestly expressed in *E. coli*, but subsequent purification showed that the protein was also effectively labeled with Cy3-DBCO (Figure 4A,B). The RefSeq protein database includes 24 proteins containing an LPXTG sequence in the *E.*

coli BL21(DE3) proteome;[48] however, incubating lysate of SrtA7M-expressing cells with Cy3-DBCO (Figure 4A) or with biotin alkyne (Supplementary Figure S7) revealed minimal off-target protein conjugation with azide above background. This indicates high specificity of the sortase reaction to the protein substrate. Increased expression of sortase did not result in higher levels of protein conjugation (Supplementary Figure S8).

Additionally, we coexpressed MBP-srt with SrtA7M under rhamnose- and IPTG-inducible promoters in DH5a cells (Figure 4A,B). Protein labeling with Cy3-DBCO was also successful under these conditions, indicating that the sortase reaction is not dependent on cell or plasmid type.

To quantify the reaction in more detail, sfGFP-His-srt was conjugated in vivo with several different substrates and purified. At the time of induction we added 25 mM Gly_3, Azp, or propargylamine to the cultures and incubated the mixtures for 24 h at 30 °C. After expression and conjugation, the proteins were purified with Ni^{2+}/NTA resin and analyzed by LC/MS. Incubation with Gly_3 resulted in partial conjugation of the LPETGG tag, while incubation with Azp or propargylamine resulted in complete conjugation of the protein (Figure 4C and Supplementary Figure S9). In the absence of amine, several other uncharacterized peaks were present in the purified protein, suggesting that SrtA7M is able to conjugate intracellular *E. coli* metabolites or medium components in addition to hydrolysis at threonine (Supplementary Figure S9).

Finally, we demonstrated the utility of our simple labeling method by making a Cy3-tagged HER2-binding imaging agent in a single expression and purification step. After expression and Azp conjugation, Trx-5f7 was bound to Ni^{2+}/NTA resin and labeled on-column with Cy3-DBCO. This probe was effective in detecting HER2 expression on SK-BR-3 breast cancer cells (Figure 4D).

Site-specific modification with alternate nucleophiles has been demonstrated in other sortase-mediated protein labeling experiments, but only in specialized cases. Similar thioester-trapping techniques have been developed using intein domains[49] and butelase,[50] but not in living cells. Here we have shown for the first time a general, high-yielding protein modification strategy that uses inexpensive bioorthogonal reagents. In addition, we have shown that this strategy is effective in living *E. coli* cells, paving the way for further engineering of specific highly active enzymes for in vivo protein experiments. Future engineering of sortase to repress proteolytic activity and activity on other cellular amines will improve the performance and specificity for amine nucleophiles.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b03836.

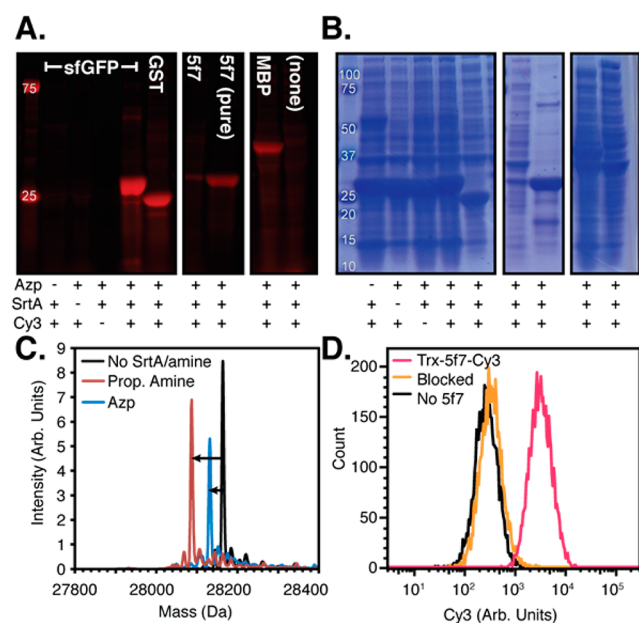Materials, methods, and supplementary figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*jennifer.cochran@stanford.edu

Notes
The authors declare the following competing financial interest(s): The authors are listed as inventors on a pending patent application related to technology described in this work.

■ REFERENCES

(1) Schumacher, D.; Hackenberger, C. *Curr. Opin. Chem. Biol.* **2014**, *22*, 62.

(2) Spicer, C.; Davis, B. *Nat. Commun.* **2014**, *5*, 4740.

(3) Strop, P.; Delaria, K.; Foletti, D.; Witt, J. M.; Hasa-Moreno, A.; Poulsen, K.; Casas, M. G.; Dorywalska, M.; Farias, S.; Pios, A.; Lui, V.; Dushin, R.; Zhou, D.; Navaratnam, T.; Tran, T. T.; Sutton, J.; Lindquist, K. C.; Han, B.; Liu, S. H.; Shelton, D. L.; Pons, J.; Rajpal, A. *Nat. Biotechnol.* **2015**, *33* (7), 694.

(4) Chen, Y.; Triola, G.; Waldmann, H. *Acc. Chem. Res.* **2011**, *44* (9), 762.

(5) James, M. L.; Gambhir, S. S. *Physiol. Rev.* **2012**, *92* (2), 897.

(6) Luo, X.; Davis, J. *Chem. Soc. Rev.* **2013**, *42* (13), 5944.

(7) Hernandez, K.; Fernandez-Lafuente, R. *Enzyme Microb. Technol.* **2011**, *48* (2), 107.

(8) Sapsford, K. E.; Algar, W. R.; Berti, L.; Gemmill, K.; Casey, B.; Oh, E.; Stewart, M.; Medintz, I. L. *Chem. Rev.* **2013**, *113* (3), 1904.

(9) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136* (27), 9572.

(10) Marshall, C. J.; Agarwal, N.; Kalia, J.; Grosskopf, V. A.; McGrath, N. A.; Abbott, N. L.; Raines, R. T.; Shusta, E. V. *Bioconjugate Chem.* **2013**, *24* (9), 1634.

(11) Agarwal, P.; Bertozzi, C. R. *Bioconjugate Chem.* **2015**, *26* (2), 176.

(12) Takaoka, Y.; Ojida, A.; Hamachi, I. *Angew. Chem., Int. Ed.* **2013**, *52* (15), 4088.

(13) Lang, K.; Chin, J. W. *Chem. Rev.* **2014**, *114* (9), 4764.

(14) Chen, Z.; Cornish, V. W.; Min, W. *Curr. Opin. Chem. Biol.* **2013**, *17* (4), 637.

(15) Zhou, Z.; Koglin, A.; Wang, Y.; McMahon, A. P.; Walsh, C. T. *J. Am. Chem. Soc.* **2008**, *130* (30), 9925.

(16) Zhang, C.; Spokoyny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L. *Angew. Chem., Int. Ed.* **2013**, *52* (52), 14001.

(17) Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W. *Chem. Commun.* **2014**, *50* (50), 6604.

(18) Schumacher, D.; Helma, J.; Mann, F. A.; Pichler, G.; Natale, F.; Krause, E.; Cardoso, M. C.; Hackenberger, C. P. R.; Leonhardt, H. *Angew. Chem., Int. Ed.* **2015**, *54* (46), 13787.

(19) Heller, K.; Ochtrop, P.; Albers, M. F.; Zauner, F. B.; Itzen, A.; Hedberg, C. *Angew. Chem., Int. Ed.* **2015**, *54* (35), 10327.

(20) Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6* (13), 2308.

(21) Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. *Nat. Methods* **2005**, *2* (2), 99.

(22) Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y. *J. Am. Chem. Soc.* **2012**, *134* (2), 792.

(23) Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L. X.; Bertozzi, C. R. *Bioconjugate Chem.* **2014**, *25* (4), 788.

(24) Schmohl, L.; Schwarzer, D. *Curr. Opin. Chem. Biol.* **2014**, *22*, 122.

(25) Biswas, T.; Pawale, V. S.; Choudhury, D.; Roy, R. P. *Biochemistry* **2014**, *53* (15), 2515.

(26) Warden-Rothman, R.; Caturegli, I.; Popik, V.; Tsourkas, A. *Anal. Chem.* **2013**, *85* (22), 11090.

(27) Parthasarathy, R.; Subramanian, S.; Boder, E. T. *Bioconjugate Chem.* **2007**, *18* (2), 469.

(28) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J. Am. Chem. Soc.* **2008**, *130* (48), 16338.

(29) Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. *J. Org. Chem.* **2007**, *72* (10), 3909.

(30) Popp, M. W.; Dougan, S. K.; Chuang, T. Y.; Spooner, E.; Ploegh, H. L. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (8), 3169.

(31) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One* **2015**, *10* (7), e0131177.

(32) Avvakumova, S.; Colombo, M.; Tortora, P.; Prosperi, D. *Trends Biotechnol.* **2014**, *32* (1), 11.

(33) Raeeszadeh-Sarmazdeh, M.; Parthasarathy, R.; Boder, E. T. *Colloids Surf., B* **2015**, *128*, 457.

(34) Stanger, K.; Maurer, T.; Kaluarachchi, H.; Coons, M.; Franke, Y.; Hannoush, R. N. *FEBS Lett.* **2014**, *588* (23), 4487.

(35) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew. Chem., Int. Ed.* **2012**, *51* (37), 9377−9380.

(36) Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. *J. Am. Chem. Soc.* **2012**, *134* (26), 10749−10752.

(37) Strijbis, K.; Spooner, E.; Ploegh, H. L. *Traffic* **2012**, *13* (6), 780.

(38) Hirakawa, H.; Ishikawa, S.; Nagamune, T. *Biotechnol. J.* **2015**, *10* (9), 1487.

(39) Chen, I.; Dorr, B. M.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11399.

(40) Hirakawa, H.; Ishikawa, S.; Nagamune, T. *Biotechnol. Bioeng.* **2012**, *109* (12), 2955.

(41) Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *J. Am. Chem. Soc.* **2008**, *130* (7), 2132.

(42) Bellucci, J. J.; Bhattacharyya, J.; Chilkoti, A. *Angew. Chem., Int. Ed.* **2015**, *54* (2), 441.

(43) Matsumoto, T.; Takase, R.; Tanaka, T.; Fukuda, H.; Kondo, A. *Biotechnol. J.* **2012**, *7* (5), 642.

(44) Baer, S.; Nigro, J.; Madej, M. P.; Nisbet, R. M.; Suryadinata, R.; Coia, G.; Hong, L. P. T.; Adams, T. E.; Williams, C. C.; Nuttall, S. D. *Org. Biomol. Chem.* **2014**, *12* (17), 2675.

(45) Li, Y.; Li, Y.; Pan, M.; Kong, X.; Huang, Y.; Hong, Z.; Liu, L. *Angew. Chem., Int. Ed.* **2014**, *53* (8), 2198.

(46) Pruszynski, M.; Koumarianou, E.; Vaidyanathan, G.; Revets, H.; Devoogdt, N.; Lahoutte, T.; Zalutsky, M. R. *Nucl. Med. Biol.* **2013**, *40* (1), 52−59.

(47) Richards, J.; Miller, M.; Abend, J.; Koide, A.; Koide, S.; Dewhurst, S. *J. Mol. Biol.* **2003**, *326* (5), 1475−1488.

(48) Pruitt, K.; Brown, G.; Tatusova, T.; Maglott, D. The Reference Sequence (RefSeq) Database. In *The NCBI Handbook* [Internet]; National Center for Biotechnology Information: Bethesda, MD, 2002; Chapter 18; available at http://www.ncbi.nlm.nih.gov/books/NBK21091/.

(49) Borodovsky, A.; Ovaa, H.; Kolli, N.; Gan-Erdene, T.; Wilkinson, K. D.; Ploegh, H. L.; Kessler, B. M. *Chem. Biol.* **2002**, *9* (10), 1149−1159.

(50) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. *Nat. Chem. Biol.* **2014**, *10* (9), 732−738.