

Site-Specific Incorporation of a Thioester Containing Amino Acid into Proteins

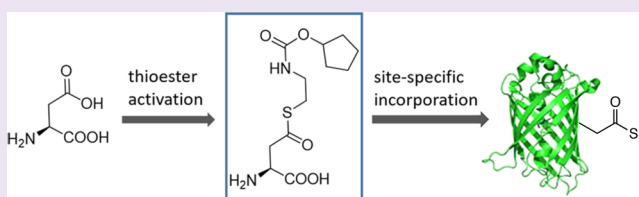
Weimin Xuan,^{†,§} Daniel Collins,^{‡,§} Minseob Koh,[†] Sida Shao,[†] Anzhi Yao,[†] Han Xiao,[†] Philip Garner,^{*,‡} and Peter G. Schultz^{*,†,§}

[†]Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N Torrey Pines Road, La Jolla, California 92037, United States

[‡]Department of Chemistry, Washington State University, Pullman, Washington 99164-4630, United States

S Supporting Information

ABSTRACT: Here, we report the site-specific incorporation of a thioester containing noncanonical amino acid (ncAA) into recombinantly expressed proteins. Specifically, we genetically encoded a thioester-activated aspartic acid (ThioD) in bacteria in good yield and with high fidelity using an orthogonal nonsense suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pair. To demonstrate the utility of ThioD, we used native chemical ligation to label green fluorescent protein with a fluorophore in good yield.



The site-specific modification of proteins with biophysical probes, post-translational modifications, drugs, oligonucleotides, and other moieties is useful for the study of protein structure and function, as well as the generation of therapeutic proteins such as antibody drug conjugates.^{1,2} To this end, a number of methods have been developed including the selective modification of cysteine residues with electrophiles,^{3,4} the fusion of peptide tags to proteins of interest which can be subsequently chemically or enzymatically modified,^{5–8} and the incorporation of noncanonical amino acids (ncAAs) into proteins through semisynthetic^{9–11} or recombinant methods.^{12,13} In nature, the thioester group provides a means for selective protein conjugation; examples include protein ubiquitination,¹⁴ intein splicing,¹⁵ and the attachment of complement factors to pathogens.¹⁶ In protein chemistry, C-terminal thioesters have been very useful for the semisynthesis of protein variants by native chemical ligation.¹¹ While C-terminal protein thioesters may be accessed using recombinant intein-based technology,¹⁵ the selective incorporation of thioesters at internal sites in recombinant proteins is difficult. Robust methods to do so would enable the formation of a variety of small molecule and polypeptide conjugates, including branched and lariat proteins.

One solution to this problem is to genetically encode a thioester containing ncAA in living cells using an orthogonal nonsense suppressor tRNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pair.^{12,13} This approach has been used to encode a large number of ncAAs with structurally diverse side chains including fluorophores, metal chelators, stable analogues of post-translationally modified amino acids, photoaffinity probes, and bio-orthogonal chemically reactive amino acids.¹⁷ ncAAs have been site-specifically incorporated into proteins in both prokaryotic^{18,19} and eukaryotic cells²⁰ and whole organisms²¹ in

excellent yields and with high fidelity. We now report the genetic incorporation of a thioester-containing ncAA into proteins in *Escherichia coli* (*E. coli*).

We chose to encode an aspartate derived thioester into proteins. The design of the candidate ncAA was based on the idea that the aspartic acid derivative should loosely resemble a known substrate of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (PylRS). The use of this system as a starting point for encoding ncAAs has enjoyed particular success because of the substrate promiscuity of PylRS, which extends well beyond the native substrate pyrrolysine (Pyl, 4). Native PylRS and a number of PylRS mutants have been engineered to accommodate a variety of ncAA side chains in its large hydrophobic binding pocket.²² With that in mind, we began by modifying the known PylRS substrate 5 (Figure 1).²³ Incorporation of an aspartyl β -thioester moiety into the Pyl structure led to 6, but the acid sensitivity of the thioaminal group necessitated the insertion of a methylene group to give the more stable analogue 7 (ThioD). Molecular docking of ThioD in wild type PylRS suggested that ThioD is likely to be accommodated in the active site of PylRS (Figure S1).

The synthesis of ThioD (Scheme S1) began with the conversion of cyclopentanol (8) to a chloroformate, which reacted with ethanolamine to give alcohol 9. This alcohol was converted to tosylate 10, and then to the thiol 11 by displacement with thiourea followed by hydrolysis. Steglich thioesterification of Boc-Asp-O^tBu produced compound 12, which underwent deprotection under acidic conditions to give ThioD as its TFA salt in 26% overall yield. The stability of

Received: November 21, 2017

Accepted: January 23, 2018

Published: January 23, 2018

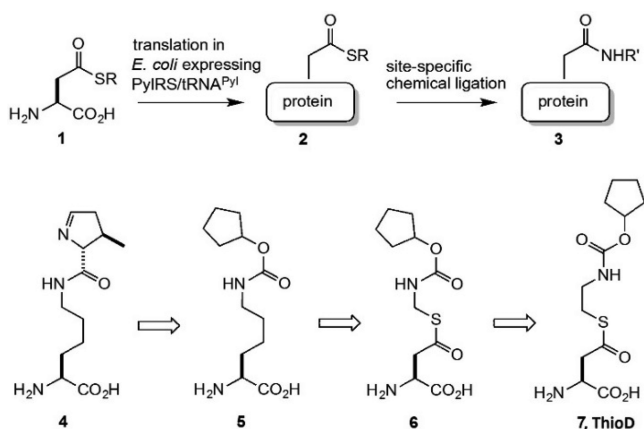


Figure 1. Side chain chemical ligation scheme and ncAA design.

ThioD was evaluated in 60 mM phosphate buffer, pH 7.1, containing 30 mM GSH (glutathione), conditions that simulate GSH levels in *E. coli*.²⁴ After 36 h, 52% of ThioD remained intact; the primary decomposition pathway involved hydrolysis to aspartic acid and free thiol (Supporting Information). This study suggested that ThioD would be stable enough to proceed with the recombinant studies.

Wild type *Methanosarcina barkeri* PylRS failed to incorporate ThioD into proteins in *E. coli*. We therefore generated a *Methanosarcina barkeri* PylRS library with five residues in the pyrrolysine binding site (Ala267, Tyr271, Asn311, Cys313, and Tyr349, Figure 2A) randomized to NNK (N = A, T, C or G; K = T or G) based on the X-ray crystal structure. Over 1×10^8 transformants were obtained to cover the theoretical library size of 3×10^7 with 96% expected completeness. Multiple rounds of positive and negative selection in *E. coli* DH10B were performed in the presence of 2 mM ThioD via the published method.²⁵ One colony was isolated that survived two positive

and one negative selection; DNA sequencing revealed a PylRS variant (ThioDRS, Figure 2B) harboring the mutations Ala267Tyr, Tyr271Ala, Asn311Thr, Cys313Gly; and Tyr349Phe. Analysis of these mutations in ThioDRS reveals a strong rationale for its recognition of ThioD—the Asn311Thr mutation appears to make space for the thioester group while the Cys313Gly and Tyr271Ala mutations make the cavity larger to accommodate the hydrophobic cyclopentane ring. The Ala267Tyr mutation may result in a hydrogen bond to the carbamate group.

To confirm the incorporation of ThioD, we inserted ThioDRS into the pUltra vector (pUltra-ThioDRS), which was designed for efficient ncAA expression, and transformed it into *E. coli* DH10B.²⁶ A C-terminal His-tag sfGFP variant with an amber mutation at a solvent exposed, permissive site (sfGFP-Y151TAG) was coexpressed on plasmid pET22b (containing a T5 promoter) in the presence or absence of 2 mM ThioD. After fermentation at 25 °C for 6 h in LB media, strong fluorescence enhancement was observable only in the presence of 2 mM ThioD (Figure 2C); purification by passing it through a Ni-NTA column afforded 8 mg/L of the sfGFP mutant. The purified sfGFP variant was also analyzed by SDS-PAGE gel (Figure 2D) and electrospray ionization quadrupole time-of-flight (ESI-QTOF) mass spectrometry (Figure 2E). The major (60%) observed mass peak (27720.0 Da) is consistent with the expected mass (27720 Da) for ThioD incorporation. However, a significant amount (35%) of protein was obtained as the GSH adduct (27838.1 Da) due to the rapid thiol-exchange reaction; we also observed a small amount (5%) of sfGFP-Y151Asp (27549.0 Da) due to thioester hydrolysis. These experiments confirmed the successful incorporation of ThioD into recombinant proteins in *E. coli*. One concern is the susceptibility of incorporated ThioD to hydrolysis during manipulation, especially at solvent exposed sites. We therefore examined the stability of the purified sfGFP-Y151ThioD by

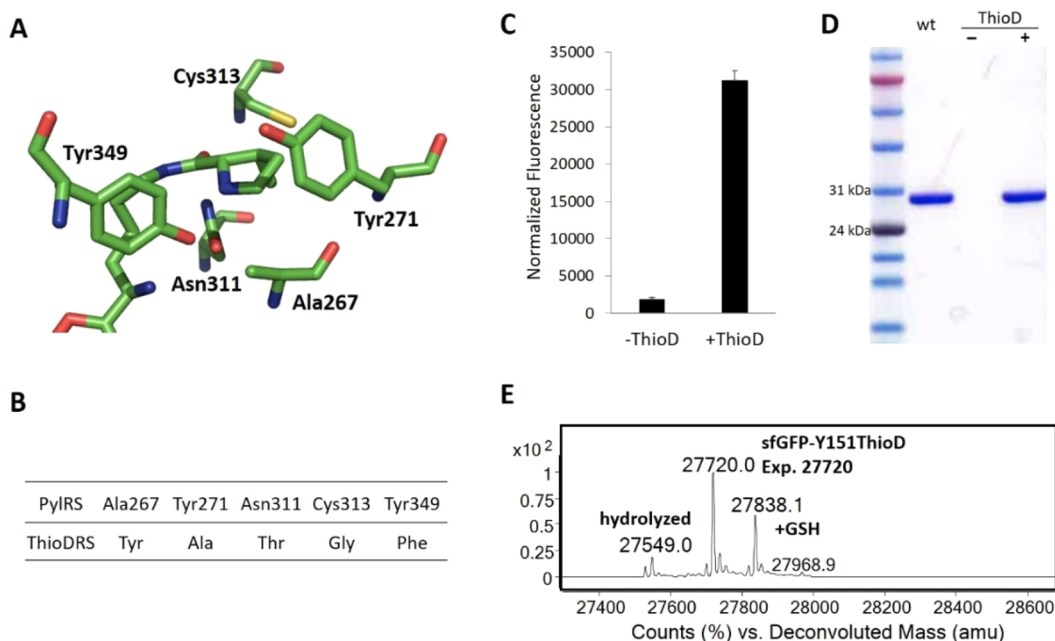


Figure 2. (A) X-ray crystal structure of the *Methanosarcina mazei* PylRS complexed with adenylated pyrrolysine (PDB: 2Q7H). The residue labeling is based on the corresponding residues in *Methanosarcina barkeri* PylRS. (B) The mutations found in the selected clone. (C) Fluorescence analysis of ThioDRS for its ability to incorporate ThioD into proteins in *E. coli*; fluorescence was normalized to OD600. (D) SDS-PAGE analysis of sfGFP-Y151TAG expressed with ThioDRS in the presence or absence of ThioD. (E) Mass spectrum of sfGFP-Y151ThioD purified from *E. coli*.

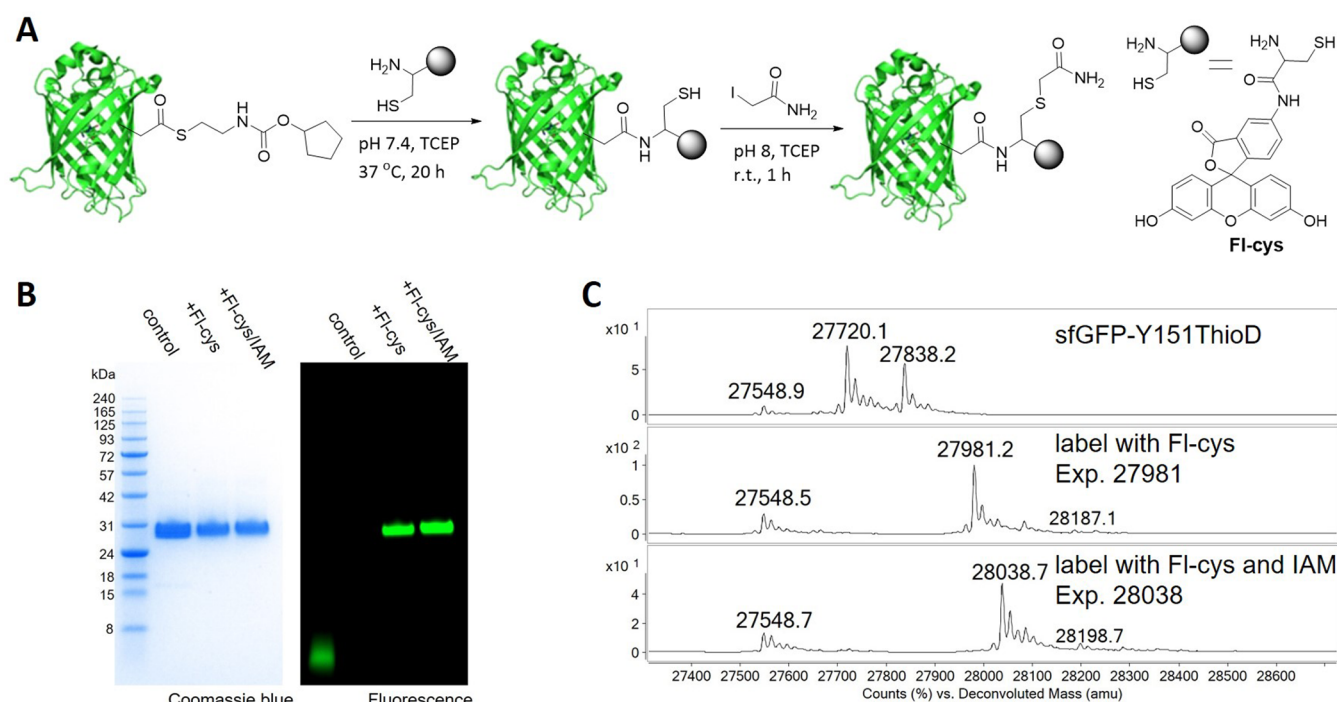


Figure 3. (A) Labeling of sfGFP-Y151ThioD with a fluorescein derivative (Fl-cys) by native chemical ligation. The following alkylation reaction was performed in a pH 8 phosphate buffer containing 0.1 mg mL⁻¹ of the above labeling product, 15 mM iodoacetamide, and 10 mM TCEP at RT. (B) SDS-PAGE gel was visualized by Coomassie blue staining (left) and fluorescence emission (right). (C) Mass spectra analysis of the protein variants in this study.

incubating it in pH 7.4 DPBS at 37 °C for 20 h. Subsequent mass spectral analysis revealed a half-life of over 20 h for ThioD at this site (Figure S5). Importantly, the GSH adduct is also susceptible to hydrolysis, suggesting it can also react with nucleophilic moieties.

Native chemical ligation is a widely used method for covalently linking an N-terminal cysteine residue of one peptide to a C-terminal thioester of another.²⁷ To expand its utility, β -aminothiol-containing ncAAs have been incorporated into proteins to enable native chemical ligation at any defined site,^{28,29} including modification at multiple sites.³⁰ We reasoned that in a similar fashion, β -aminothiol-containing biophysical probes and other moieties can be used to modify protein variants that contain ThioD. To test this notion, we used a fluorescein derivative containing a β -aminothiol (Figure 3A, Fl-cys). The conjugation reaction was carried out by incubating sfGFP-Y151ThioD (3.7 μ M) and Fl-Cys (100 equiv) in 50 mM phosphate, pH 7.4, buffer containing 25 mM TECP [tris(2-carboxyethyl)phosphine] at 37 °C for 20 h. The labeling product was subjected to SDS-PAGE gel analysis, stained with Coomassie blue or directly imaged by fluorescence emission (Figure 3B). ESI-QTOF mass spectrometry was also used to analyze the labeling products. As shown in Figure 3C, the desired product was observed, with complete consumption of the ThioD-containing sfGFP and its GSH adduct. This suggests that the GSH adduct is largely in the thioester form and active in a second thiol-exchange reaction (ring size constraints may limit S to N acyl migration in the GSH adduct). Notably, the hydrolyzed byproduct was only found in small amounts (<20%). To confirm the presence of the amide-linked fluorescein and a free thiol, we further incubated the labeling product with iodoacetamide (IAM) and observed complete disappearance of the labeling product and the generation of a

new peak consistent with the alkylation of the thiol group (Figure 3B and C). Importantly, the thiol handle generated by the acyl transfer reaction can be modified with additional electrophilic agents.

In summary, we have genetically encoded the noncanonical amino acid ThioD in *E. coli* and demonstrate that proteins containing ThioD can be efficiently modified with β -aminothiol derivatives in excellent yields. This work provides another useful tool for carrying out selective chemistry on proteins with exquisite selectivity. We are currently extending this work to the formation of branched and lariat protein structures.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.7b00998.

Materials and Methods, additional text, and NMR spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: ppg@wsu.edu.

*E-mail: schultz@scripps.edu.

ORCID

Peter G. Schultz: 0000-0003-3188-1202

Author Contributions

§These authors contributed equally to this work

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant CHE-1665274 (P.G.) and National Institutes of Health grant R01 GM062159 (P.G.S.). We thank Zhihao Yu (Clowers group, WSU) for assistance with HRMS measurements and Kristen Williams for her assistance in manuscript preparation.

REFERENCES

- (1) Krall, N., da Cruz, F. P., Boutureira, O., and Bernardes, G. J. L. (2016) Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* 8, 103–113.
- (2) Chen, X., and Wu, Y.-W. (2016) Selective chemical labeling of proteins. *Org. Biomol. Chem.* 14, 5417–5439.
- (3) Lin, Y. A., Boutureira, O., Lercher, L., Bhushan, B., Paton, R. S., and Davis, B. G. (2013) Rapid Cross-Metathesis for Reversible Protein Modifications via Chemical Access to Se-Allyl-selenocysteine in Proteins. *J. Am. Chem. Soc.* 135, 12156–12159.
- (4) Zhang, C., Welborn, M., Zhu, T., Yang, N. J., Santos, M. S., Van Voorhis, T., and Pentelute, B. L. (2016) π -Clamp-mediated cysteine conjugation. *Nat. Chem.* 8, 120–128.
- (5) Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* 21, 86–89.
- (6) Carrico, I. S., Carlson, B. L., and Bertozzi, C. R. (2007) Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* 3, 321–322.
- (7) Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klauert, D. H., Bulleit, R. F., and Wood, K. V. (2008) HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* 3, 373–382.
- (8) Miller, L. W., Sable, J., Goelet, P., Sheetz, M. P., and Cornish, V. W. (2004) Methotrexate Conjugates: A Molecular In Vivo Protein Tag. *Angew. Chem., Int. Ed.* 43, 1672–1675.
- (9) Vila-Perelló, M., and Muir, T. W. (2010) Biological Applications of Protein Splicing. *Cell* 143, 191–200.
- (10) Muir, T. W. (2003) Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* 72, 249–289.
- (11) Bondalapati, S., Jbara, M., and Brik, A. (2016) Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. *Nat. Chem.* 8, 407–418.
- (12) Xiao, H., and Schultz, P. G. (2016) At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harbor Perspect. Biol.* 8, a023945.
- (13) Liu, C. C., and Schultz, P. G. (2010) Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* 79, 413–444.
- (14) Cappadocia, L., and Lima, C. D. (2017) Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.*, DOI: 10.1021/acs.chemrev.6b00737.
- (15) Shah, N. H., and Muir, T. W. (2014) Inteins: nature's gift to protein chemists. *Chem. Sci.* 5, 446–461.
- (16) Law, S. K. A., and Dodds, A. W. (1997) The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* 6, 263–274.
- (17) Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* 6, 50–69.
- (18) Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., Reed, S. A., Yang, P.-Y., Wang, R. E., Javahishvili, T., Schultz, P. G., and Young, T. S. (2016) Recombinant thiopeptides containing non-canonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3615–3620.
- (19) He, J., Van Treeck, B., Nguyen, H. B., and Melançon, C. E. (2016) Development of an Unnatural Amino Acid Incorporation System in the Actinobacterial Natural Product Producer *Streptomyces venezuelae* ATCC 15439. *ACS Synth. Biol.* 5, 125–132.
- (20) Wang, L. (2017) Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res.* 50, 2767.
- (21) Chen, Y., Ma, J., Lu, W., Tian, M., Thauvin, M., Yuan, C., Volovitch, M., Wang, Q., Holst, J., Liu, M., Vriza, S., Ye, S., Wang, L., and Li, D. (2017) Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res.* 27, 294–297.
- (22) Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta, Proteins Proteomics* 1844, 1059–1070.
- (23) Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D., and Ambrogelly, A. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* 580, 6695–6700.
- (24) Fahey, R. C., Brown, W. C., Adams, W. B., and Worsham, M. B. (1978) Occurrence of Glutathione in Bacteria. *J. Bacteriol.* 133, 1126–1129.
- (25) Xuan, W., Shao, S., and Schultz, P. G. (2017) Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew. Chem., Int. Ed.* 56, 5096–5100.
- (26) Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013) A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Biochemistry* 52, 1828–1837.
- (27) Malins, L. R., and Payne, R. J. (2014) Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* 22, 70–78.
- (28) Virdee, S., Kapadnis, P. B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D. P., Riechmann, L., and Chin, J. W. (2011) Traceless and Site-Specific Ubiquitination of Recombinant Proteins. *J. Am. Chem. Soc.* 133, 10708–10711.
- (29) Li, X., Fekner, T., Ottesen, J. J., and Chan, M. K. (2009) A Pyrrolysine Analogue for Site-Specific Protein Ubiquitination. *Angew. Chem., Int. Ed.* 48, 9184–9187.
- (30) Wissner, R. F., Batjargal, S., Fadzen, C. M., and Petersson, E. J. (2013) Labeling Energy Transfer Pairs by Combining Unnatural Amino Acid Mutagenesis and Native Chemical Ligation. *J. Am. Chem. Soc.* 135, 6529–6540.