Traceless and Site-Specific Ubiquitination of Recombinant Proteins

Satpal Virdee, Prashant B. Kapadnis, Thomas Elliott, Kathrin Lang, Julia Madrzak, Duy P. Nguyen, Lutz Riechmann, and Jason W. Chin*

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K.

Supporting Information

ABSTRACT: Protein ubiquitination is a post-translational modification that regulates almost all aspects of eukaryotic biology. Here we discover the first routes for the efficient site-specific incorporation of δ-thiol-L-lysine (7) and δ-hydroxy-L-lysine (8) into recombinant proteins, via evolution of a pyrrolysyl-tRNA synthetase/tRNA_CUA pair. We combine the genetically directed incorporation of native isopeptide bond between substrate proteins and ubiquitin. We exemplify this approach by demonstrating the synthesis of a ubiquitin dimer and the first synthesis of ubiquitinated SUMO.

Post-translational modification of target proteins with the 76 amino acid protein ubiquitin regulates almost all aspects of eukaryotic biology. Ubiquitination is the process by which the ε-amino group of a lysine residue within the substrate protein is linked to the C-terminal carboxylate of ubiquitin via an isopeptide bond. In vivo, ubiquitin is attached to its substrates by a series of enzymes (E1s, E2s, E3s) that direct isopeptide bond formation. Studying the molecular consequences of protein ubiquitination is challenging, since there are >600 E3 ubiquitin ligases when the ligases are known, they may not drive the reaction to completion or at a unique site in vitro.

Several investigators have addressed the creation of ubiquitin conjugates that are connected via non-native linkages, including a disulfide bond, an oxime, triazoles, and isopeptide bonds in which the universally conserved C-terminal glycine of ubiquitin is mutated to ε-cysteine or alanine in a non-traceless native chemical ligation. While some of these non-native linkages have found utility, a clear and important challenge is to address the creation of methods for the ubiquitination of any protein at a user-defined site via a native isopeptide bond.

We recently described a new approach, in protein chemistry termed GOPAL, for creating native isopeptide bonds between ubiquitin and a specific lysine in a target protein. Using the M.arkeri (Mb) pyrrolysyl-tRNA synthetase (PyrRS)/tRNA_CUA pair, which naturally introduces pyrrolysine (I, Figure 1) into proteins in certain methanogens, we site-specifically inserted Nε-(t-butyloxycarbonyl)-L-lysine (2) into ubiquitin and developed a series of selective protection and deprotection steps that allowed us to direct site-selective isopeptide bond formation. Using this approach, we were able to generate important ubiquitin dimers linked through specific isopeptide bonds, solve the crystal structure of Lys6-linked diubiquitin, and reveal new deubiquitinase specificity. Since this method relies on cellular protein synthesis to generate the component proteins, it is in principle scalable for the creation of traceless isopeptide bonds between proteins of any size. However, this method does require multiple protection and deprotection steps and generates highly protected hydrophobic intermediates that can be poorly soluble. Moreover, the method may be challenging to apply to proteins that cannot be refolded.

Received: March 28, 2011
Published: June 28, 2011
1,2-Aminothiolcs can react reversibly with thioesters to form a new thioester-linked intermediate. Intramolecular nucleophilic attack on the thioester intermediate, in a favorable S-to-N acyl shift, leads to formation of an amide bond. This reaction has been used to assemble polypeptides from fragments bearing C-terminal thioesters and fragments bearing N-terminal cysteines and forms a basis for native chemical ligation.11 Recently, four distinct lysine derivatives (3–6), bearing 1,2-aminothiols, which can be incorporated into synthetic peptides via solid-phase peptide synthesis methods have been described.12–18 3 can be ligated with C-terminal thioesters of ubiquitin, and subsequent auxiliary removal allows the generation of a native isopeptide bond.12,15 The deprotection of 4–6 allows their ligation with C-terminal thioesters, and subsequent desulfurization yields native isopeptide bonds.14–17 Unfortunately, while these amino acids can be incorporated into longer peptides via rounds of native chemical ligation with thioesters, these steps require further thioreduction, decrease the yield of protein conjugates, and ultimately limit the length of proteins and/or the positions within proteins to which these approaches might be applied.

Here we demonstrate the site-specific incorporation of δ-thiol-L-lysine (7), which has recently been used in peptide ligation at several sites,17,18 into proteins that are overexpressed using the cell’s translational machinery (Figure 1). Native chemical ligation of proteins containing 7 with a ubiquitin thioester and subsequent desulfurization yields ubiquitin conjugates linked via a native isopeptide bond. This work provides a simple, scalable, and broadly accessible route to ubiquitinated proteins. In the process of this work, we also incorporated another important post-translational modification, δ-hydroxy-L-lysine (8).

Since 7 differs from lysine only by the insertion of a sulfur atom, it may be thermodynamically challenging to create a synthetase that will recognize 7 but exclude lysine by a factor of 103–105, as required to maintain the fidelity of natural protein translation. L-Butyloxy-carbonyl (Boc)-protected lysine (2) is a good substrate for PyrRS,19,20 and we have previously demonstrated that, while the PyrRS/tRNA$_{CUA}$ pair does not directly incorporate the Nε-methyl-L-lysine, it does direct the incorporation of an Nε-methyl derivative of lysine bearing an Nε-Boc group.20 Since the Boc group can be removed after incorporation of the amino acid into the protein, this provides a paradigm for installing modifications on the ε-aminogroup that cannot be installed directly. Therefore, we investigated whether the addition of an Nε-Boc group will also facilitate the incorporation of δ-substituted lysine derivatives (9, 10).

We first synthesized Nε-Boc-protected versions of amino acids bearing δ-substituents (9–11) (Supporting Information (SI) Schemes 1 and 2 and Methods). We demonstrated that none of these amino acids are incorporated into proteins in response to the amber codon using the wild-type PyrRS/tRNA$_{CUA}$ pair. Next, we aimed to discover an evolved PyrRS/tRNA$_{CUA}$ pair for the incorporation of amino acids 9–11. The crystal structure of PyrRS in complex with pyrrolysine reveals two prominent residues (N311 and Y349) that are within 5 Å of the δ-carbon of pyrrolysine (SI Figure 1). N311 in PyrRS binds to the carbonyl group in the bound pyrrolysine, and mutation of this amino acid destroys the ability of the enzyme to discriminate this substrate from natural amino acids (data not shown). Since this carbonyl group is conserved in our designed substrates (9–14), we decided to maintain N311 as a potential positive specificity determinant.

We created a library in which Y349 of MbtPyrRS is mutated to all natural amino acids and selected MbtPyrRS/tRNA$_{CUA}$ variants that confer chloramphenicol resistance on cells bearing a chloramphenicol acetyltransferase gene with an amber codon at a permissive site (D112TAG) in the presence of δ-hydroxy-Nε-(β-butyloxy-carbonyl)-lysine (9). We performed the initial selections in the presence of 9 since it is valence isoelectronic with its thiol analogue (10) but can be prepared in gram quantities in a single step from commercial starting materials (SI Scheme 1), and since we were concerned that a fraction of the δ-thiol compound might undergo oxidation that could potentially lead to the selection of synthetases that recognize oxidized forms of the amino acid. These selections yielded a single mutant, Y349W, in PyrRS. Subsequent selections using the δ-thiol compound 10 directly yielded the same mutation. Selections using a number of more complex libraries did not yield alternative or improved mutants, nor did any of the libraries tested allow the incorporation of the disulfide-protected compound 11 (data not shown).

The selected synthetase (δSHKRS)/tRNA$_{CUA}$ pair conferred chloramphenicol resistance on cells containing a chloramphenicol acetyltransferase gene with an amber codon at position 112 of 200 μg mL$^{-1}$ in the presence of 9 and <50 μg mL$^{-1}$ in the absence of 10. We produced C-terminally His-tagged ubiquitin with an amber codon at position 6 from UbTAG6-His$_6$, in the presence of δSHKRS/tRNA$_{CUA}$ and 9 or 10, in reasonable yield (0.5 mg L$^{-1}$ (SI Figure 2)). No protein was produced in the absence of the unnatural amino acid. The incorporation of each amino acid was conclusively demonstrated by ESI-MS analysis (SI Figure 3), and ubiquitin bearing the δ-thiol-L-lysine (7) at position 6 was prepared by the quantitative removal of the Boc group from ubiquitin bearing δ-thiol-Nε-(β-butyloxy-carbonyl)-L-lysine at position by the addition of 60% TFA for 1 h at 22 °C and characterized by mass spectrometry (SI Figure 3).

Taken together, the phenotypic experiments, protein expression experiments, and mass spectrometry data conclusively demonstrate that δSHKRS/tRNA$_{CUA}$, in the presence of 9 or 10, directly incorporates δ-hydroxy-Nε-(β-butyloxy-carbonyl)-L-lysine or δ-thiol-Nε-(β-butyloxy-carbonyl)-L-lysine into recombinant proteins in response to the amber codon and allows the preparation of proteins containing a site-specifically incorporated δ-thiol-L-lysine (7). However, we were interested in improving two aspects of this approach. First, the yield of recombinant protein produced when using this synthetase was ~10 times lower than that obtained with 2 and the PyrRS/tRNA$_{CUA}$ Pair, a combination that we and others have shown is very efficient.10,19,20 Second, the deprotection conditions are denaturing, making this approach to installing 7 incompatible with proteins that cannot be reversibly refolded. To improve the method, we combined our progress up to this point with some observations we had made while investigating the evolvability of the PyrRS/tRNA$_{CUA}$ pair. This allowed us to very efficiently install δ-substituted derivatives of lysine (7 and 8) into proteins under native conditions.

In the process of investigating the scope of amino acids that can be incorporated using the PyrRS/tRNA$_{CUA}$ pair, we discovered a variant synthetase (nitroCbzKRS) that incorporates Nε-(p-nitrocobenzoyloxy)-L-lysine (12). This synthetase was selected by five rounds of positive and negative selection, in the presence and absence of 12, on a 10^5-member synthetase library, which contains all combinations of mutations at residues M241, A267, Y271, L274, and C313. The evolved synthetase contains the mutations Y271M, L274G, and C313A. The selected nitroCbzKRS/tRNA$_{CUA}$ pair conferred chloramphenicol resistance on cells containing a chloramphenicol acetyltransferase
gene with an amber codon at position 112 of >300 \( \mu \text{g mL}^{-1} \) in the presence of 12 and <50 \( \mu \text{g mL}^{-1} \) in the absence of 12. Cells containing the nitroCbzKRS/tRNA\(_{CUA} \) pair directed expression of UbTAG6-His\(_6 \) in the presence 12 to produce good yields of ubiquitin (10 mg L\(^{-1} \)). Ubiquitin expression was clearly amino acid dependent (SI Figure 4).

The amino acid dependence observed in the phenotypic and protein expression experiments demonstrated that 12 is incorporated during protein translation and that there is little translational incorporation of natural amino acids in response to the amber codon. However, mass spectrometry of the purified protein revealed that the protein contained lysine in place of 14. While we do not yet know the exact mechanism by which the nitro-substituted Cbz group is removed, a likely mechanism is incorporated during protein translation and that there is little translational incorporation of natural amino acids in response to the amber codon. However, mass spectrometry of the purified protein revealed that the protein contained lysine in place of 14. While we do not yet know the exact mechanism by which the nitro-substituted Cbz group is removed, a likely mechanism

We next synthesized \( \delta \)-substituted derivatives of Ne-(p-nitrocarbobenzyloxy)lysine (13 and 14; SI Scheme 3 and Methods) and aimed to site-specifically incorporate these into proteins. The incorporation of these amino acids and subsequent removal of p-nitrocarbobenzyloxy group, as described above, should lead to a clear mass shift in the protein, corresponding to the mass added by the \( \delta \)-substituent, and provide a direct route to the incorporation of \( \delta \)-substituted lysine derivatives.

As expected, the nitroCbzKRS/tRNA\(_{CUA} \) pair did not direct the efficient incorporation of the \( \delta \)-substituted amino acids (Figure 2). To discover a synthetase/tRNA pair that uses 13 and 14 to yield 8 and 7 in recombinant proteins, we combined the mutation in \( \delta \)SHKR and the incorporation of \( \delta \)-hydroxy-Ne-(1-butylxycarbonyl)-l-lysine & \( \delta \)-thiol-Ne-(1-butylxycarbonyl)-l-lysine with those in nitroCbzKRS for introducing Ne-(p-nitrocarbobenzyloxy)-l-lysine to create a new synthetase (nitroCbzKRS\(^{\delta} \)).

Mass spectrometry of ubiquitin purified from cells containing the nitroCbzKRS\(^{\delta} \)/tRNA\(_{CUA} \) pair, 13 and a ubiquitin gene bearing in an amber codon at position 6 demonstrates the incorporation of 8 (Supplementary Figure 5). This is consistent with the translational incorporation of \( \delta \)-hydroxy-Ne-(1-butylxycarbonyl)-l-lysine into ubiquitin and the subsequent removal of the p-nitrocarbobenzyloxy group from the protein.

Finally, we incorporated \( \delta \)-thiol-Ne-(p-nitrocarbobenzyloxy)-l-lysine into ubiquitin at position 6 using the nitroCbzKRS/tRNA\(_{CUA} \) pair and 14. Again, protein expression was amino acid-dependent (Figure 2). The yield of recombinant protein (10 mg L\(^{-1} \)) was comparable to that for the incorporation of 2 with the PyRS/tRNA\(_{CUA} \) pair.\(^{10} \) Mass spectrometry revealed a mass of 9490 Da, corresponding to removal of the p-nitrocarbobenzyloxy group from the \( \epsilon \)-amine of lysine and the formation of a thiazolidine adduct between the resulting 1,2-aminothiol and pyruvate\(^{23} \) (Figure 2, SI Figure 6). A second minor peak corresponds to the decarboxylation of the thiazolidine adduct. These adducts are well predecurred for free 1,2-aminothiols resulting from N-terminal cysteines. Treatment of the protein with 200 mM methoxyamine\(^{23} \) in phosphate-buffered saline (PBS) led to quantitative removal of the pyruvate adducts to reveal \( \delta \)-thiol-l-lysine (7) at the genetically directed site in the protein, as characterized by mass spectrometry. We have repeated this experiment at several different, biologically relevant sites in ubiquitin and in SUMO to demonstrate the generality of this approach (SI Figures 7 and 11); it nonetheless remains a possibility that some other proteins will behave differently.

While we do not yet know the exact mechanism by which the nitro-substituted Cbz group is removed, a likely mechanism would include the reduction of the aromatic nitro group, pre or post cell lysis, and the subsequent fragmentation to reveal a free \( \epsilon \)-amino group (SI Figure 6).

To demonstrate the utility of this system for genetically directing chemoselective protein ubiquitination, we synthesized K6-linked diubiquitin—an important ubiquitin linkage that may be involved in DNA repair-related signaling processes in mammalian cells.\(^{24,25} \) Ubiquitin bearing 7 at position 6 was ligated with an equivalent of ubiquitin thioester, prepared by intein fusion thiolysis.\(^{10} \) After 4 h, the yield of the ubiquitin conjugate linked via an amide bond between \( \delta \)-thiol lysine (7) at position 6 in one ubiquitin and the C-terminus of a second ubiquitin (DiUb\(_6\)SHK6-His\(_6\)) was \( \sim 70\% \), as judged by SDS-PAGE (Figure 3). Ubiquitin species were folded by dialysis against folding buffer (PBS + 1 mM DTT), and the K6-linked diubiquitin conjugate was purified from residual monoubiquitin by ion exchange chromatography (SI Figure 8). The purified ubiquitin chain (DiUb\(_6\)SHK6-His\(_6\)) was a single band by SDS-PAGE, a single peak by HPLC, and the expected mass, confirming the formation of the amide bond (SI Figures 8 and 9). To reveal the entirely native isopeptide bond, DiUb\(_6\)SHK6-His\(_6\) was
desulfurized via a free radical method.\textsuperscript{26} Desulfurization was complete after 2 h, as determined by LC-MS, yielding the first genetically directed native isopeptide linkage to ubiquitin (DiUbK6-His\textsubscript{6}; Figure 3, SI Figure 9). Desulfurization reagents were removed, and concomitant folding was achieved by dialysis of the reaction mixture against 10 mM Tris, pH 7.6 buffer, conditions that are known to yield folded ubiquitin chains.\textsuperscript{27} We confirmed that the K6-linked diubiquitin we have synthesized is folded by circular dichroism spectroscopy, using a folded K6-linked diubiquitin of which we have previously solved the X-ray structure\textsuperscript{10} as an authentic standard (SI Figure 10). To further confirm the biological integrity of the ubiquitin dimers synthesized by our method, we demonstrated that they are efficient substrates for members of the ubiquitin-specific protease (USP) family of deubiquitinases (USP2 and USP5) that we have previously shown are able to readily hydrolyze K6-linked diubiquitin \textsuperscript{10} (SI Figure 10). To further demonstrate the generality of our approach, we incorporated \textdelta\text{-thiol-Ne-(p-nitrocobenzoyl)-L-lysine} into position 11 of recombinant SUMO and prepared ubiquitinated SUMO, a modified protein implicated in promyelocytic leukemia,\textsuperscript{28} by procedures analogous to those used to prepare K6-linked ubiquitin (SI Figure 11).

In conclusion, we have discovered the first route for the efficient site-specific incorporation of \textdelta\text{-thiol-L-lysine} (7) and \textdelta\text{-hydroxy-L-lysine} (8) into recombinant proteins. We have combined the genetically directed incorporation of 7 with native chemical ligation and desulfurization to yield an entirely native isopeptide bond between substrate proteins and ubiquitin. Moreover, we have discovered that the p-nitrocobenzyloxy group can be used to site-specifically install lysine and its close analogues into proteins, and this may facilitate the preparation of proteins bearing site-specific isotopic labels for protein NMR applications.

In the process of developing our approach, we have created synthetases for five new amino acids, including the first for \textdelta\text{-substituted lysines. We have also demonstrated that independently selected synthetase mutations that allow the incorporation of \epsilon\text{-substituted lysine} and that allow the incorporation of \textdelta\text{-substituted lysines can be combined to incorporate lysine derivatives bearing both \textdelta\text- and \epsilon\text-substituents in a single molecule.}

We anticipate that the efficient genetic encoding of unnatural amino acids for native isopeptide bond formation will provide a scalable and broadly accessible route to protein ubiquitination, SUMOylation, and Neddylation and accelerate research into the roles of these important post-translational modifications.

\section*{ASSOCIATED CONTENT}
\section*{Supporting Information.} Supplementary figures and experimental procedures. This material is available free of charge via the Internet at http://pubs.acs.org.

\section*{AUTHOR INFORMATION}
Corresponding Author
chin@mrc-lmb.cam.ac.uk

\section*{ACKNOWLEDGMENT}
We are grateful to the ERC and MRC for funding.

\section*{REFERENCES}
(1) Chen, Z. J.; Sun, L. J. Mol. Cell 2009, 33, 275.
(2) Ikeda, F.; Crosetto, N.; Dikic, I. Cell 2010, 143, 677.
(3) Deshaies, R. J.; Joazeiro, C. A. P. Annu. Rev. Biochem. 2009, 78, 399.
(4) Chen, J.; Ai, Y.; Wang, J.; Haraccka, L.; Zhuang, Z. Nat. Chem. Biol. 2010, 6, 270.
(5) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Nat. Chem. Biol. 2010, 6, 267.
(6) Shanmugham, A.; Fish, A.; Luna-Vargas, M. P. A.; Faessen, A. C.; El Ouailid, F.; Sixma, T. K.; Ovaa, H. J. Am. Chem. Soc. 2010, 132, 8834.
(7) Eger, S.; Scheffner, M.; Marx, A.; Rubini, M. J. Am. Chem. Soc. 2010, 132, 16337.
(8) Li, X.; Fekner, T.; Ottesen, J. J.; Chan, M. K. Angew. Chem., Int. Ed. 2009, 48, 9184.
(9) McGinty, R. K.; Köhn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. ACS Chem. Biol. 2009, 4, 958.
(10) Virdee, S.; Ye, Y.; Nguyen, D. P.; Komander, D.; Chin, J. W. Nat. Chem. Biol. 2010, 6, 750.
(11) Dawson, P. E.; Kent, S. B. Annu. Rev. Biochem. 2000, 69, 923.
(12) Chatterjee, C.; McGinty, R. K.; Pellois, J.-P.; Muir, T. W. Angew. Chem., Int. Ed. 2007, 46, 2814.
(13) McGinty, R.; Kim, J.; Chatterjee, C.; Roeder, R.; Muir, T. W. Nature 2008, 453, 812.
(14) Yang, R.; Pasunooti, K.; Li, F.; Liu, X.; Liu, C. J. Am. Chem. Soc. 2009, 131, 13592.
(15) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Chem. Commun. 2010, 46, 7199.
(16) Ajish Kumar, K.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H.; Brik, A. Angew. Chem., Int. Ed. 2009, 48, 8090.
(17) Kumar, K. S. A.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. Angew. Chem., Int. Ed. 2010, 49, 9126.
(18) El Ouailid, F.; Merks, R.; Ekekebus, R.; Hameed, D. S.; Smit, J. J.; de Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. Angew. Chem., Int. Ed. 2010, 49, 10149.
(19) Yanagisawa, T.; Iishi, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Chem. Biol. 2008, 15, 1187.
(20) Nguyen, D.; García Alai, M.; Kapadnis, P.; Neumann, H.; Chin, J. W. J. Am. Chem. Soc. 2009, 131, 14194.
(21) Kavran, J. M.; Gundlapalli, S.; O’Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. Proc. Natl. Acad. Sci. U.S.A. 2007, 104, 11268.
(22) Neumann, H.; Peak-Chew, S.; Chin, J. Nat. Chem. Biol. 2008, 4, 232.
(23) Ottesen, J. J.; Bar-Dagan, M.; Giovani, B.; Muir, T. W. Biopolymers 2008, 90, 406.
(24) Nishikawa, H.; Ooka, S.; Sato, K.; Arima, K.; Okamoto, J.; Klevit, R.; Fukuda, M.; Ohta, T. J. Biol. Chem. 2004, 279, 3916.
(25) Wu-Baer, F.; Lagrazon, K.; Yuan, W.; Baer, R. J. Biol. Chem. 2003, 278, 34743.
(26) Wan, Q.; Danishefsky, S. J. Angew. Chem., Int. Ed. 2007, 46, 9248.
(27) Pickart, C. M.; Raasi, S. Methods Enzymol. 2005, 399, 21.
(28) Tatham, M. H.; Geoffroy, M. C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E. G.; Palvimo, J. J.; Hay, R. T. Nat. Cell Biol. 2008, 10, 538.