CHEMBIOCHEM

DOI: 10.1002/cbic.200900690

A Genetically Encoded ε-*N*-Methyl Lysine in Mammalian Cells

Dan Groff,[a] Peng R. Chen,[b] Francis B. Peters,[a] and Peter G. Schultz*[a]

The post-translational methylation of lysine modulates the activity, stability, localization and biomolecular interactions of many eukaryotic proteins. For example, monomethylation of Lys372 in the mammalian tumor suppressor p53 has been shown to affect protein stability and localization.[1] Protein methylation plays a particularly important role in gene expression due to its involvement in the histone code, in which specific modifications to histone proteins modulate the transcriptional status of specific genes. Methylation of distinct histone lysine residues has been correlated with both transcriptional activation and repression depending on the lysine modified.[2] To better understand the functional consequences of lysine methylation, methods are needed to generate proteins with defined methylation status both in vitro and in living cells.

One straightforward in vitro method for lysine methylation makes use of methyltransferase enzymes,[3] which transfer the *S*-methyl group of the cofactor *S*-adenosyl methionine to the sidechains of lysine and arginine. These enzymes have very specific sequence requirements, and their activity can lead to heterogeneous mixtures of methylation products.[4] To circumvent these limitations, one can use solid-phase peptide synthesis to directly incorporate methylated lysine residues into full-length proteins by using native chemical ligation.[5] In addition, in vitro translation has been adapted for the production of histone tails with multiple modified lysine residues.[6] However, these methods can suffer from low yields, restrictions on the site of modification, and are not easily adapted to cellular studies. Recently, an in vitro chemical modification strategy was developed that takes advantage of the unique reactivity of cysteine. The reaction between cysteine and *N*-methyl aminoethylhalides generates a thioether adduct structurally similar to methyl lysine. Mono-, di-, and tri-methylated lysine analogues have been generated in this way (Scheme 1).[7] By using a similar strategy, unnatural amino acid mutagenesis was used to selectively incorporate phenylselenocysteine into proteins. It can be subsequently oxidized to dehydroalanine and reacted with aminoethylthiols to again produce methyl lysine analogues. However, an initial oxidation step is required, so this strategy is not compatible with proteins containing redox active cysteine or methionine residues.[8]

[a] *Dr. D. Groff,[+] F. B. Peters,[+] Prof. Dr. P. G. Schultz*
Department of Chemistry, The Scripps Research Institute
10550 N. Torrey Pines, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu

[b] *Dr. P. R. Chen[+]*
Department of Chemical Biology, Peking University
Beijing 100871 (China)

[+] *These authors contributed equally to this work.*

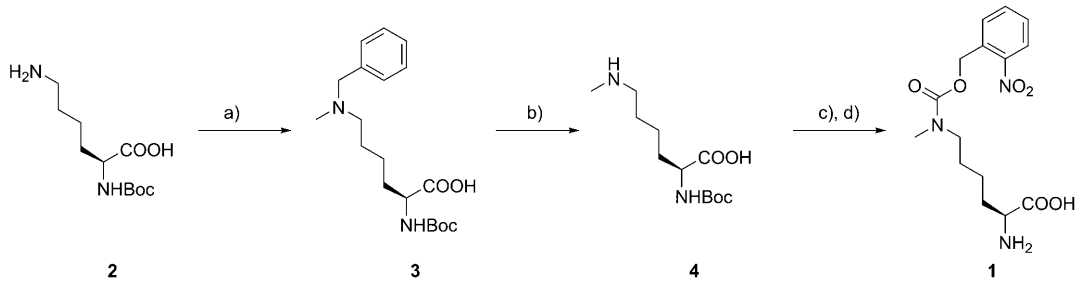
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200900690.

Biosynthetic approaches have also been used to incorporate methyl lysine into proteins. For example, the ε-methyl lysine precursor N^{ε}-*tert*-butyl-oxycarbonyl N^{ε}-methyl-L-lysine (Boc methyl lysine) has been genetically inserted into proteins in *E. coli*. This strategy produces good yields of homogeneously monomethylated proteins largely independent of sequence context.[4] However, Boc deprotection in aqueous TFA is necessary to generate monomethyl lysine and can cause denaturation and loss of biological activity with many proteins.[9]

All of the above methods for introducing methyl lysine require in vitro manipulation to produce the final, methylated protein. This restriction generally prevents studies of methylated proteins in their native cellular context. To overcome this limitation, we have genetically encoded the photocaged *N*-methyl lysine, N^{ε}-*o*-nitrobenzyl-oxycarbonyl-N^{ε}-methyl-L-lysine **1**, in both bacteria and mammalian cells. Photocaged second messengers and proteins are widely used tools for the spatial and temporal control of a variety of cellular processes because light allows noninvasive generation of the active photoproducts in the cell. Furthermore, it has been previously demonstrated that *o*-nitrobenzyl-*O*-tyrosine and dimethoxy-*o*-nitrobenzyl-*O*-serine can be efficiently deprotected with light in *E. coli*,[10] *Xenopus* oocytes,[11] and yeast.[12]

The synthesis of photocaged methyl lysine **1** involves two sequential reductive aminations of N^{α}-*tert*-butyl-oxycarbonyl-L-lysine with benzaldehyde and formaldehyde using STABH to afford N^{α}-*tert*-butyl-oxycarbonyl-N^{ε}-benzyl-N^{ε}-methyl-L-lysine.[13] Reductive debenzylation quantitatively yielded N^{α}-*tert*-butyl-oxycarbonyl-N^{ε}-methyl-L-lysine **4** which was then coupled to *o*-nitrobenzyl chloroformate to produce **5**. Boc deprotection proceeded quantitatively with HCl in dioxane to generate photocaged methyl lysine **1**.

To genetically encode **1**, we used a pyrrolysyl-tRNA synthetase from *M. barkeri* (MbPylRS) and a pyrrolysyl tRNA ($tRNA^{Pyl}_{CUA}$) from *M. mazei*,[14] which previously were adapted for the site-specific incorporation of unnatural amino acids into proteins in response to the amber nonsense codon TAG.[15–17] The orthogonality of this pair to endogenous tRNAs and aminoacyl-tRNA synthetases (aaRSs) has been demonstrated in both *E. coli* and mammalian cells, so that an aaRS evolved to incorporate the unnatural amino acid N^{ε}-*o*-nitrobenzyloxycarbonyl-L-lysine in *E. coli* can also be used in mammalian cells. An aaRS library was created in *E. coli* in which the codons for residues L270, Y271, L274 and C313 of the pyrrolysyl-tRNA synthetase from *M. barkeri* were all randomized as NNK. Directed evolution with two positive rounds and one negative round of selection as previously described[18] resulted in the identification of the aaRS G12 (experimental details in Supporting Information), which is capable of incorporating **1** into proteins. aaRS G12 has the following mutations: Y271I, L274M, and C313A.



Scheme 1. Synthesis of photocaged methyl lysine **1**. a) Benzaldehyde and STABH followed by formaldehyde and STABH; b) H_2 with Pd/C; c) *o*-nitrobenzyl chloroformate; d) 4 N HCl/dioxanes

To determine the fidelity and efficiency with which **1** can be incorporated into proteins in *E. coli*, a myoglobin gene with an amber mutation at codon 99 and a C-terminal His_6 tag was expressed in the presence of the G12 aaRS/$tRNA^{Pyl}_{CUA}$ pair and **1**. In the absence of unnatural amino acid, no full length myoglobin was evident by SDS-PAGE analysis. In Terrific Broth media supplemented with 1 mM **1**, the yield of mutant protein was 4.2 mg protein L^{-1} (Figure 1 A), compared to yields of around 1 mg L^{-1} for protein expressed with the WT pyrrolysine synthetase and the pyrrolysine analogue N^{ε}-cyclopentyloxycarbonyl-L-lysine[15] and 2 mg L^{-1} for Myo4TAG suppressed with $tRNA^{Tyr}_{CUA}$ and the WT tyrosyl aminoacyl-tRNA synthetase from *M. jannaschii*.[16] To confirm incorporation of **1** at position 99, the mutant myoglobin was purified with nickel-affinity chromatography and then subjected to electrospray mass spectral analysis. For protein expressed in the presence of **1**, mass peaks of 18540 Da corresponding to the photocaged methyl lysine mutant, 18408 Da corresponding to the methyl lysine mutant with an additional acetylation, and 18351 Da corresponding to the lysine mutant were present (Figure 1 A). The methyl lysine mutant is likely due to enzymatic degradation of the nitrobenzyl carbamate moiety; this has been observed previously.[15] The lysine mutant might result from demethylation of methyl lysine; it was not observed in the mass spectrum in cases in which protein was expressed in the absence of **1**; this demonstrates that this peak is not due to misincorporation of lysine at this position.

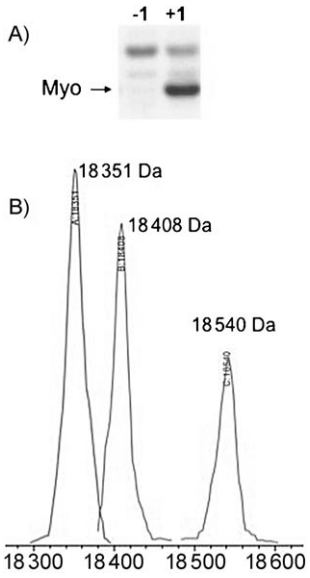


Figure 1. Photocaged methyl lysine can be incorporated into proteins with high fidelity in *E. coli* using the G12 aaRS/$tRNA^{Pyl}_{CUA}$ pair. A) *E. coli* amber suppression of Myo99TAG in Terrific Broth. Left lane: protein expressed without **1**. Right lane: protein expressed with 1 mM **1**. B) Electrospray mass spectrum of TAG99→**1** Myo mutant.

Next we determined if the G12 aaRS/$tRNA^{Pyl}_{CUA}$ pair could be used to selectively incorporate photocaged methyl lysine into proteins in mammalian cells. The genes encoding G12 and $tRNA^{Pyl}_{CUA}$ were introduced into the mammalian vector pCMV, hereafter designated pCMV-G12, with the G12 gene under control of a constitutive CMV promoter and a single $tRNA^{Pyl}_{CUA}$ under control of a human U6 promoter.[15] Chinese hamster ovary (CHO) cells were transiently transfected with vector pCMV-G12 and a reporter plasmid containing the eGFP gene with codon 40 mutated to TAG and a C-terminal His_6 tag. CHO cells grown in media lacking **1** exhibited no visible fluorescence, whereas cells grown in media with 1 mM **1** produced a bright GFP signal (Figure 2 A). This was further verified with an anti-His_6 western blot (Figure 2 B), which indicated that full length protein was produced only in the presence of photocaged methyl lysine **1**. Finally, to verify that the protein contained only photocaged methyl lysine, 21 μg of protein was isolated by nickel affinity chromatography from 2.2×10^7 CHO cells. Electrospray mass spectral analysis showed the presence of only one peak at 29840 *m/z* (Figure 2 C) corresponding to the eGFP40TAG→**1** mutant (calcd mass = 29842 Da). In addition, the lack of peaks corresponding to *N*-methyl lysine, lysine, or additional acetylation further support the hypothesis that these modifications in *E. coli* result from enzymes not present in mammalian cells. To verify that this peak represents photocaged methyl lysine, protein was subjected to photolysis for 20 min with $\geq$ 365 nm light, an irradiation regime demonstrated to be safe for use with mammalian cells.[19] Subsequent mass spectral analysis showed quantitative conversion to a protein with a mass of 29663 Da, 177 mass units lighter, corresponding to the loss of one *o*-nitrobenzyloxycarbonyl group. To demonstrate that methyl lysine can be liberated in living cells, eGFP40→**1** was photolyzed in CHO cells. The previous experiment was repeated with irradiation before cell lysis and protein purification. Two peaks are evident in the mass spectrum, a small peak at 29845 *m/z* corresponding to eGFP con-

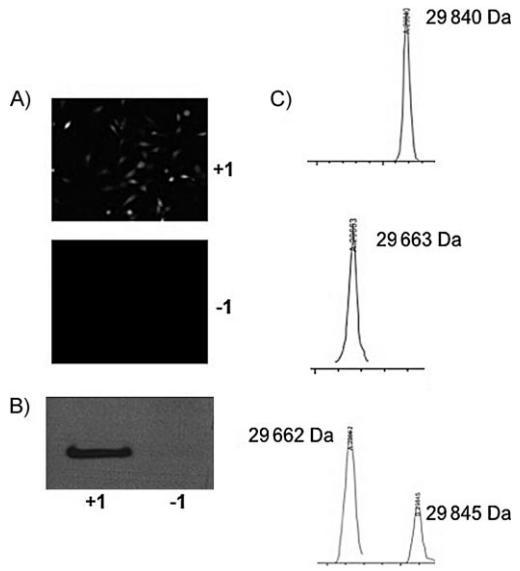


Figure 2. Photocaged methyl lysine **1** can be incorporated into proteins with high fidelity in mammalian cells. A) CHO expression of eGFP73TAG transfected with pCMV-G12 imaged with fluorescence microscopy. Top image shows cells supplemented with 1 mM **1**; bottom image shows cells with no unnatural amino acid. B) Western blot analysis of eGFP40TAG→**1** mutant expression probed with anti-His_6 antibody. C) Electrospray mass spectrum of eGFP40TAG→**1** mutant protein from CHO cells purified by using nickel affinity chromotography. Top: before irradiation, middle: after irradiation $\geq$365 nm for 20 min, bottom: irradiation in CHO cells $\geq$365 nm for 60 min.

taining **1**, and a major peak at 29662 *m/z* corresponding to photolyzed eGFP containing methyl lysine at position 40.

In conclusion, we have shown that N-methyl lysine can be efficiently incorporated into proteins in both *E. coli* and mammalian cells with high fidelity. Although additional effort is necessary to produce homogenously methylated proteins in *E. coli*, this system is capable of introducing photocaged methyl lysine site specifically into proteins in mammalian cells. In addition, we demonstrated that the caging group can be removed in living cells to produce homogeneously methylated proteins. This method will be a useful tool in understanding the role of monomethylation in a variety of biological processes.

Acknowledgements

This work was supported by NIH grants R01M62159 and PN22EY01 (P.G.S.).

Keywords: amino acids · methyl lysine · photochemistry · post-translational modifications · pyrrolysine · unnatural amino acids

[1] S. Chuikov, J. K. Kurash, J. R. Wilson, B. Xiao, N. Justin, G. S. Ivanov, K. McKinney, P. Tempst, C. Prives, S. J. Gamblin, N. A. Barlev, D. Reinberg, *Nature* **2004**, *432*, 353.
[2] P. Hublitz, M. Albert, A. H. Peters, *Int. J. Dev. Biol.* **2009**, *53*, 335.
[3] F. Martino, S. Kueng, P. Robinson, M. Tsai-Pflugfelder, F. van Leeuwen, M. Ziegler, F. Cubizolles, M. M. Cockell, D. Rhodes, S. M. Gasser, *Mol. Cell* **2009**, *33*, 323.
[4] D. P. Nguyen, M. M. Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 14194.
[5] S. He, D. Bauman, J. S. Davis, A. Loyola, K. Nishioka, J. L. Gronlund, D. Reinberg, F. Meng, N. Kelleher, D. G. McCafferty, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12033.
[6] T. J. Kang, S. Yuzawa, H. Suga, *Chem. Biol.* **2008**, *15*, 1166.
[7] M. D. Simon, F. Chu, L. R. Racki, C. C. de La Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, *128*, 1003.
[8] J. Guo, J. Wang, J. S. Lee, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 6399.
[9] H. Kanazawa, M. Nishikawa, A. Mizutani, C. Sakamoto, Y. Morita-Murase, Y. Nagata, A. Kikuchi, T. Okano, *J. Chromatogr. A* **2008**, *1191*, 157.
[10] A. Deiters, D. Groff, Y. Ryu, J. Xie, P. G. Schultz, *Angew. Chem.* **2006**, *118*, 2794; *Angew. Chem. Int. Ed.* **2006**, *45*, 2728.
[11] K. D. Philipson, J. P. Gallivan, G. S. Brandt, D. A. Dougherty, H. A. Lester, *Am. J. Physiol. Cell Physiol.* **2001**, *281*, C195.
[12] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nat. Chem. Biol.* **2007**, *3*, 769.
[13] Z.-P. Huang, X.-Y. Su, J.-T. Du, Y.-F. Zhao, Y.-M. Li, *Tetrahedron Lett.* **2006**, *47*, 5997.
[14] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459.
[15] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052.
[16] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232.
[17] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.
[18] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[19] W. F. Edwards, D. D. Young, A. Deiters, *ACS Chem. Biol.* **2009**, *4*, 441.

Received: November 12, 2009
Published online on April 26, 2010

