J|A|C|S COMMUNICATIONS

Published on Web 09/22/2009

Genetically Encoding N^{ϵ}-Methyl-L-lysine in Recombinant Histones

Duy P. Nguyen, Maria M. Garcia Alai, Prashant B. Kapadnis, Heinz Neumann, and Jason W. Chin*

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

Received August 4, 2009; E-mail: chin@mrc-lmb.cam.ac.uk

The N^{ε}-methylation status of specific lysine residues on histone proteins in chromatin controls heterochromatin formation, X-chromosome inactivation, genome imprinting, DNA repair; regulates transcription; and may define epigenetic status.[1−3] The reversible post-translational methylation of lysine residues in histones is mediated by methylases and demethylases, and lysine residues are found in mono-, di-, and trimethylated states. The state and site of modification correlate with the functional outcome in ways that are beginning to be deciphered.[4]

A molecular understanding of the organismal phenomena orchestrated by lysine N^{ε}-methylation is impeded by the challenge of producing site-specifically and quantitatively methylated histones. Researchers have used methyltransferases to methylate histones,[5] but in many cases this is unsatisfactory because it is difficult to control the site, extent, or degree of methylation using these enzymes *in vitro*. And in many cases the specific methyltransferase is simply unknown. Native chemical ligation has been used to construct histones with modified N-terminal tails,[6−8] and this approach has been extended, *via* multiple ligations, to address ubiquitylation outside the tail of a histone.[9] These experiments are often challenging and require synthesis of large quantites of peptide thioesters. Thioether analogues of N^{ε}-methyl-L-lysine in which the γ-methylene unit of lysine is replaced with a sulfur atom can be installed in proteins.[10,11] While these analogues are simple to employ, they are longer than the native amino acids by 0.3 Å,[12] decrease the pK_a of the ammonium protons by 1.1 unit,[13] and have more degrees of freedom, which may lead to altered specificity or affinity in binding interactions.[12] Moreover, one method of creating the linkage may also lead to racemization at the α carbon of the amino acid.[10] Taken together these differences may lead to unpredictable effects on the properties of the analogues. Since these analogues are created for the purpose of discovering unknown properties of the natural system, or explaining known phenomena in molecular detail, differences between the analogues and the natural modification are potentially problematic.

To understand the native system one would ideally install the natural modification via a scalable method that quantitatively introduces the modification at any defined site. We recently demonstrated that another important post-translationally modified amino acid, N^{ε}-acetyl-L-lysine, can be quantitatively and site-specifically genetically encoded in recombinant proteins in response to the amber codon using an evolved pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$ pair that is orthogonal in *E. coli*.[14] This approach is facilitating a molecular understanding of the role of lysine acetylation.[15] In principle, it is possible to use a similar approach to evolve an orthogonal tRNA-synthetase/$tRNA_{CUA}$ pair that specifically recognizes methyl-L-lysine (**3**) and directs its incorporation into recombinant proteins. However, creating a synthetase that will use methyl-L-lysine, but discriminate against L-lysine (**4**, a smaller amino acid which cannot be sterically excluded from the active site, differs by only a single methyl group and is abundant in the cell) by a factor of 10^3 to 10^4 as required for translation, is thermodynamically challenging in the absence of an amino acid editing site.[16,17] Indeed the pyrrolysyl-tRNA synthetase does not accept methyl-lysine as a substrate,[18] and our efforts to evolve a pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$ pair for the direct genetic encoding of methyl-L-lysine, essentially as previously described, did not yield specific enzymes (data not shown).

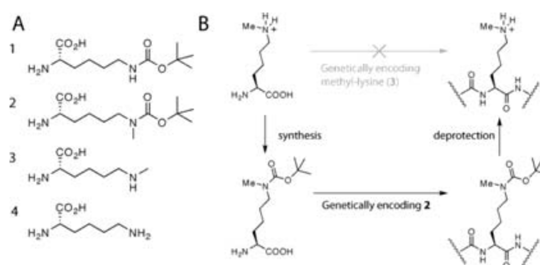

Figure 1. Strategies for encoding lysine methylation. (A) Amino acids used. (B) Schemes for encoding **3** in recombinant proteins.

We realized that we might be able to encode N^{ε}-methyl-L-lysine (**3**) indirectly by providing the synthetase enzyme with a substrate that was significantly different from both N^{ε}-methyl-L-lysine and L-lysine if we were able to subsequently effect the facile, quantitative, and specific post-translational conversion of this precursor to N^{ε}-methyl-L-lysine on the synthesized protein. Since N^{ε}-*tert*-butyl-oxycarbonyl-L-lysine (**1**) is an efficient substrate for the pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$ pair,[19] we asked whether N^{ε}-methyl-L-lysine (**3**) could be incorporated into proteins in a two-step process in which N^{ε}-*tert*-butyl-oxycarbonyl-N^{ε}-methyl-L-lysine (**2**) is genetically incorporated into proteins and the *tert*-butyl-oxycarbonyl group is removed post-translationally to reveal N^{ε}-methyl-L-lysine (Figure 1).

To investigate whether **2** can be incorporated using the pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$ pair we prepared the amino acid in 95% yield by deprotection of commercially available N^{α}-Fmoc-N^{ε}-*tert*-butyl-oxycarbonyl-N^{ε}-methyl-L-lysine **5**. In addition we directly synthesized **2** from **3** in 77% yield (Supplementary Schemes 1 and 2, Supplementary methods, and Supplementary Figure 1). We transformed *E. coli* with pBKPylS (which encodes the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase, *Mb*PylRS) and p*Myo*4TAGPylT-his_6 (which encodes *Mb*$tRNA_{CUA}$ and a C-terminally hexahistidine tagged sperm whale myoglobin gene with an amber codon at position 4)[14] and induced protein expression with and without the addition of **2** to midlog phase cells. Full-length myoglobin was only produced and purified in good yield in the presence of **2** (Figure 2). To demonstrate that **2** can be incorporated with high fidelity into recombinant proteins and is not subjected to *in vivo* modification,[14] we performed electrospray ionization mass spectrometry (ESI-MS) on the purified protein. The ESI-MS spectra of myoglobin-His_6 demonstrates the quantitative incorporation of **2** (Supplementary Figure 2A). These data demonstrate that **2** can be genetically encoded in proteins in good yield and with high fidelity using the *Mb*PylRS/*Mb*$tRNA_{CUA}$ pair.

10.1021/ja906603s CCC: $40.75 © 2009 American Chemical Society

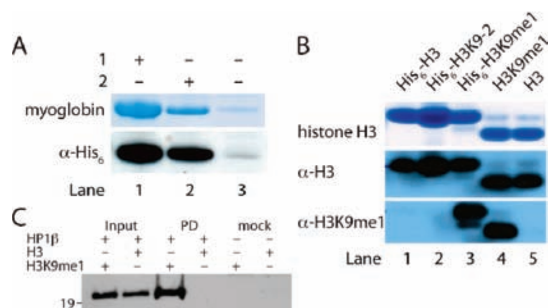

Figure 2. **2** can be site-specifically incorporated into recombinant proteins in response to an amber codon and quantitatively, post-translationally converted to **3**. (A) Myoglobin-His$_6$ is purified from *E. coli* containing p*Myo*4TAGPylT-*his$_6$*, and pBKPylS in the presence of amino acids **1** or **2**. (B) Synthesis of H3K9me1, lane 3, His$_6$ H3 incorporating **2** in place of K9 and deprotected with 2% TFA, lanes 4 and 5 are postcleavage of the N-terminal His$_6$ tag with TEV protease. (C) HP1 specifically recognizes H3K9me1. HP1 was used to immunoprecipitate H3 or H3K9me1. The immunoprecipitation was probed for H3 using an anti H3 antibody. Input: 2% of total Histone H3. PD "pull down". Mock: no HP1 added.

To specifically and efficiently introduce **2** in a histone at a physiologically relevant site, we transformed *E. coli* BL21(DE3) with pBKPylS and pCDF-PylT-H3K9TAG (a vector which encodes *Mb*tRNA$_{CUA}$ and a N-terminally hexahistidine tagged histone H3 gene in which the codon for lysine 9 is replaced with an amber codon).[15] We grew the cells in the presence of 2 mM **2** and expressed and purified the recombinant histone in good yield (2 mg per liter of culture). ESI-MS analysis of the purified histone confirms the incorporation of **2** into histone H3 (Supplementary Figure 2B).

To demonstrate that the *tert*-butyl-oxycarbonyl group can be quantitatively removed from the histone under mild conditions, the purified H3K9-**2** was treated with a solution of 2% trifluoroacetic acid (TFA) for 4 h at 37 °C. Western blots with an anti-H3K9me1 antibody against unmodified H3, H3 bearing **2** at position 9 (H3K9-**2**), and the TFA treated H3K9-**2** confirmed the presence of methyl-L-lysine at position 9 in the deprotected sample (Figure 2B, lane 3). The ESI-MS spectra of the deprotected H3K9-**2** sample (Supplementary Figure 2C) demonstrates that the auxiliary is quantitatively removed under these conditions to reveal N^{ε}-methyl-L-lysine. MS/MS protein sequencing (Supplementary Figure 2D) further confirms that the site of lysine methylation is as genetically encoded. H3K9me1 can be assembled into nucleosomes in vitro with a comparable efficiency to unmodified H3 (Supplementary Figure 3).

To demonstrate the biochemical activity of the methylated histone generated by our approach, we performed immunoprecipitations with heterochromatin protein 1 (HP1) (Figure 2C), a chromodomain protein[20] that does not bind to unmethylated H3 but is known to specifically bind to short peptides based on a histone H3 tail bearing mono-, di-, or trimethylated K9 (with a preference for di- and trimethylated H3 K9).[21] HP1 immunoprecipitation of full-length H3K9me1, synthesized by our approach, and full length H3 allows us to demonstrate that HP1 binds specifically to full-length H3K9me1 over unmethylated H3.

In conclusion, we have created a general method for the quantitative, site-specific incorporation of N^{ε}-methyl-L-lysine in recombinant proteins. The method has two steps: first, an amino acid containing an auxiliary group is used to differentiate N^{ε}-methyl-L-lysine from L-lysine and to provide a good substrate for the pyrrolysyl synthetase; second, the auxiliary group is removed to reveal N^{ε}-methyl-L-lysine. We have demonstrated the utility of the method by site-specifically installing N^{ε}-methyl-L-lysine into full-length histone H3 and demonstrated that the modified H3 specifically recruits HP1.[21] We are currently extending our approach to installing other modifications implicated in the histone code and epigenetic inheritance to understand how combinations of post-translational modifications program cellular outcomes.

Acknowledgment. We are grateful to Daniela Rhodes, Andrew Routh, and Lynda Chapman for reagents and discussions and to Sew Peak-Chew for assistance acquiring MS/MS spectra. This work was supported by The Medical Research Council, UK, and grants from the European Research Council & the Human Frontiers of Science Program to J.W.C.

Supporting Information Available: Experimental protocols and supplementary figures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kohler, C.; Villar, C. B. *Trends Cell Biol.* **2008**, *18*, 236–43.
(2) Spivakov, M.; Fisher, A. G. *Nat. Rev. Genet.* **2007**, *8*, 263–71.
(3) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–49.
(4) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41–5.
(5) Martino, F.; Kueng, S.; Robinson, P.; Tsai-Pflugfelder, M.; van Leeuwen, F.; Ziegler, M.; Cubizolles, F.; Cockell, M. M.; Rhodes, D.; Gasser, S. M. *Mol. Cell* **2009**, *33*, 323–34.
(6) Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. *Science* **2006**, *311*, 844–7.
(7) He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F.; Kelleher, N.; McCafferty, D. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12033–8.
(8) Shogren-Knaak, M. A.; Fry, C. J.; Peterson, C. L. *J. Biol. Chem.* **2003**, *278*, 15744–15748.
(9) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. *Nature* **2008**, *453*, 812–6.
(10) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2008**, *47*, 6399–401.
(11) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128*, 1003–12.
(12) Gellman, S. H. *Biochemistry* **1991**, *30*, 6633–6.
(13) Gloss, L. M.; Kirsch, J. F. *Biochemistry* **1995**, *34*, 3990–8.
(14) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–4.
(15) Neumann, H.; Hancock, S. M.; Buning, R.; Routh, A.; Chapman, L. Somers, J.; Owen-Hughes, T.; van Noort, J.; Rhodes, D.; Chin, J. W. *Mol. Cell.*, in press.
(16) Fersht, A. R. *Biochemistry* **1977**, *16*, 1025–30.
(17) Nureki, O.; Vassylyev, D. G.; Tateno, M.; Shimada, A.; Nakama, T.; Fukai, S.; Konno, M.; Hendrickson, T. L.; Schimmel, P.; Yokoyama, S. *Science* **1998**, *280*, 578–82.
(18) Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695–700.
(19) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–22.
(20) Kim, J.; Daniel, J.; Espejo, A.; Lake, A.; Krishna, M.; Xia, L.; Zhang, Y.; Bedford, M. T. *EMBO Rep.* **2006**, *7*, 397–403.
(21) Fischle, W.; Wang, Y.; Jacobs, S. A.; Kim, Y.; Allis, C. D.; Khorasanizadeh, S. *Genes Dev.* **2003**, *17*, 1870–81.

JA906603S