nature chemical biology

Genetically encoding N^{ε}-acetyllysine in recombinant proteins

Heinz Neumann, Sew Y Peak-Chew & Jason W Chin

N^{ε}-acetylation of lysine (1) is a reversible post-translational modification with a regulatory role that rivals that of phosphorylation in eukaryotes. No general methods exist to synthesize proteins containing N^{ε}-acetyllysine (2) at defined sites. Here we demonstrate the site-specific incorporation of N^{ε}-acetyllysine in recombinant proteins produced in *Escherichia coli* via the evolution of an orthogonal N^{ε}-acetyllysyl-tRNA synthetase/$tRNA_{CUA}$ pair. This strategy should find wide applications in defining the cellular role of this modification.

N^{ε}-acetylation of lysine is a dynamic, reversible post-translational modification first described on histones. Acetylation and deacetylation are mediated by histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively. The acetylation of lysine residues on several histones modulates chromatin condensation[1] and may be an epigenetic mark as part of the histone code[2]. In recent years it has emerged that hundreds of eukaryotic proteins (beyond histones) are acetylated[3]. Over 60 transcription factors and coactivators are acetylated, including the tumor suppressor p53, and numerous enzymes including MDM2, acetyl coenzyme A synthase, HATs, HDACs, kinases and phosphatases are acetylated[3,4]. Acetylation regulates diverse cellular processes including DNA replication, DNA repair, DNA recombination, maintenance of genomic stability, cytoskeletal dynamics, metabolism, cytokine-activated signal transduction, protein folding and trafficking, and apoptosis[3,4]. Acetylation is also emerging as a key regulator in mitochondrial processes, with over 20% of mitochondrial proteins subject to acetylation[3]. Overall it is emerging that N^{ε}-acetylation is a modification with a diversity of roles and a functional importance that rivals that of phosphorylation[5,6].

Despite the huge importance of lysine acetylation there is no general method of producing homogeneous recombinant proteins that contain N^{ε}-acetyllysine at defined sites. Semisynthetic methods to install N^{ε}-acetyllysine using native chemical ligation were crucial in demonstrating the role of acetylation of Lys16 in histone H4 on chromatin decompaction[1]. These experiments give a hint of the impact that a general method to produce homogeneously acetylated proteins (that is, one that does not require the synthesis of large quantities of modified peptide thioester and is not limited to terminal residues) would have in defining the molecular and cellular role of lysine acetylation. Researchers have used purified HAT complexes to acetylate recombinant proteins, but this is often an unsatisfactory solution because (i) HATs for particular modifications are unknown, (ii) *tour de force* efforts are often required to prepare active HAT complexes, (iii) HAT-mediated reactions are often difficult to drive to completion, leading to a heterogeneous sample and (iv) HATs may acetylate several sites, making it difficult to interrogate the molecular consequences of acetylation at any one site. To address these deficits we envisioned genetically encoding the incorporation of N^{ε}-acetyllysine into recombinant proteins with high translational fidelity and efficiency in response to the amber codon, via the generation of an orthogonal N^{ε}-acetyllysyl-tRNA synthetase/tRNA pair.

Certain methanogenic bacteria, including *Methanosarcina barkeri* MS, incorporate pyrrolysine (3) in response to the UAG codons present in several methyltransferase genes[7,8]. The incorporation of pyrrolysine in *M. barkeri* MS is directed by a pyrrolysyl-tRNA synthetase (MbPylRS) and its cognate amber suppressor, $MbtRNA_{CUA}$, in response to an amber codon[9]. Previous work demonstrates that the MbPylRS/$MbtRNA_{CUA}$ pair functions in *E. coli* and that $MbtRNA_{CUA}$ is not an efficient substrate for endogenous aminoacyl-tRNA synthetases in *E. coli*[8,10]. The MbPylRS/$MbtRNA_{CUA}$ pair therefore appears to satisfy two of the three criteria for orthogonality with respect to endogenous aminoacyl-tRNA synthetases and tRNAs[11]. These observations, in combination with the realization that acetyllysine is a substructure of pyrrolysine (but not a substrate for PylRS[12]), led us to investigate the evolution of the MbPylRS/$MbtRNA_{CUA}$ pair into an N^{ε}-acetyllysyl-tRNA synthetase/$tRNA_{CUA}$ pair for the genetic incorporation of acetyllysine into proteins expressed in *E. coli*.

We confirmed the activity of the MbPylRS/$MbtRNA_{CUA}$ pair (the sequences of all genes and proteins used are available in the **Supplementary Methods** online) in *E. coli* by using this pair to incorporate the pyrrolysine analog N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc, 4; previously demonstrated to be an efficient substrate of MbPylRS[12]) in response to an amber codon at position 4 of a modified sperm whale myoglobin-His_6 gene (*Myo-his6(4TAG)*). The purified yield of myoglobin-His_6 (2 mg l^{-1} of culture) was comparable to that obtained when we used the *Methanococcus jannaschii* tyrosyl-tRNA synthetase/$tRNA_{CUA}$ ($MjTyrRS/MjtRNA_{CUA}$) pair to insert tyrosine (5) in response to the amber codon in *Myo-his6(4TAG)* (**Fig. 1**), which demonstrates that the MbPylRS/$MbtRNA_{CUA}$ pair directs Cyc incorporation with an efficiency comparable to that of a pair used successfully for genetic code expansion. Electrospray ionization mass spectra of purified myoglobin-His_6 from cells containing Cyc, MbPylRS/$tRNA_{CUA}$ and *Myo-his6(4TAG)* demonstrates the incorporation of a single Cyc residue into myoglobin-His_6 (**Fig. 1**). This demonstrates that Cyc is not measurably incorporated in response to sense codons (that is, MbPylRS is orthogonal with respect to cellular tRNAs) and confirms that natural amino acids are not measurably incorporated in response to the amber codon in the presence of Cyc.

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH England, UK. Correspondence should be addressed to J.W.C. (chin@mrc-lmb.cam.ac.uk).

Received 25 October 2007; accepted 22 January 2008; published online 17 February 2008; doi:10.1038/nchembio.73

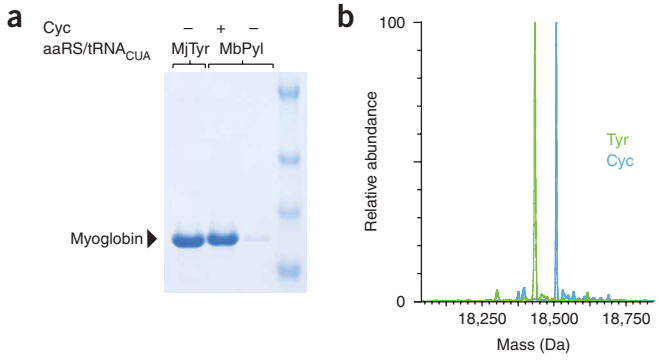

Figure 1 The MbPylRS/MbtRNA$_{CUA}$ pair efficiently and specifically incorporates Cyc in response to an amber stop codon in the gene for myoglobin. (**a**) Production of myoglobin-His$_6$ from *Myo-his6(4TAG)* in Myo4TAGPylT depends on the presence of Cyc in the growth medium. Myoglobin is produced from *Myo-his6(4TAG)* in the presence of MjTyrRS/MjtRNA$_{CUA}$ (lane 1) or in the presence of MbPylRS/MbtRNA$_{CUA}$ with or without 1 mM Cyc (lanes 2 and 3, respectively). Myoglobin was purified by Ni^{2+} affinity chromatography, analyzed by SDS-PAGE and stained with Coomassie. (**b**) ESI-MS analysis of myoglobin produced by MjTyrRS/MjtRNA$_{CUA}$ (Tyr) revealed a mass of 18,433.2 Da (calcd. 18,431.2 Da), whereas the myoglobin produced by MbPylRS/MbtRNA$_{CUA}$ (Cyc) had a mass of 18,510.7 Da. The expected mass difference (m(Cyc) – m(Tyr) = 258.3 Da – 181.2 Da = 77.1 Da) corresponds well to the mass difference observed (77.5 Da).

We detected only a trace of full-length myoglobin-His$_6$ by Coomassie stain when Cyc or MbPylRS was withheld from cells containing *Myo-his6(4TAG)*, which suggests that there is a very low level of aminoacylation of MbtRNA$_{CUA}$ by endogenous aminoacyl-tRNA synthetases and further demonstrates that MbtRNA$_{CUA}$ is orthogonal to endogenous aminoacyl-tRNA synthetases. The specificity of the MbPylRS/MbtRNA$_{CUA}$ pair was further confirmed by phenotypic assays (data not shown). Overall, phenotypic data, protein expression data and mass spectrometry data demonstrate that the MbPylRS/MbtRNA$_{CUA}$ pair is a highly active, specific and orthogonal pair in *E. coli*.

To begin to evolve the MbPylRS/MbtRNA$_{CUA}$ orthogonal pair for the incorporation of N^{ε}-acetyllysine in response to the amber codon, we created a library of 10^8 MbPylRS mutants in which six residues (Leu266, Leu270, Tyr271, Leu274, Cys313 and Trp383) are randomized (**Fig. 2**). These residues were chosen on the basis of the structure of *Methanosarcina mazei* PylRS in complex with pyrrolysine[13], and they are within 6 Å of the bound pyrroline ring of pyrrolysine. To select mutant MbPylRS/MbtRNA$_{CUA}$ pairs that direct the genetic incorporation of N^{ε}-acetyllysine, we performed three rounds of selection (positive, negative, positive), as described in the **Supplementary Methods**.

Aminoacyl-tRNA synthetase clones that survived the selection were isolated and transformed with pREP-pylT (which contains the *pylT* gene and confers green fluorescent protein (GFP) fluorescence and chloramphenicol resistance on cells bearing functional amber suppressors). 22 clones out of the 96 we screened conferred chloramphenicol resistance on *E. coli* up to 150 μg ml^{-1} and 20–30 μg ml^{-1} chloramphenicol in the presence and absence of 2 mM N^{ε}-acetyllysine, respectively; these clones also showed amino acid–dependent GFP fluorescence. The large difference in chloramphenicol resistance in the presence and absence of N^{ε}-acetyllysine suggests that the selected synthetases have a substantial *in vivo* specificity for the insertion of N^{ε}-acetyllysine over all 20 common amino acids found in the cell, in response to the amber codon. Sequencing revealed that the clones have two distinct sequences, which we designated AcKRS-1 and AcKRS-2. AcKRS-1 has six mutations (L266V, L270I, Y271F, L274A, C313F and the nonprogrammed mutant D76G), whereas AcKRS-2 has four mutations (L270I, Y271L, L274A and C313F) with respect to MbPylRS. It is likely that

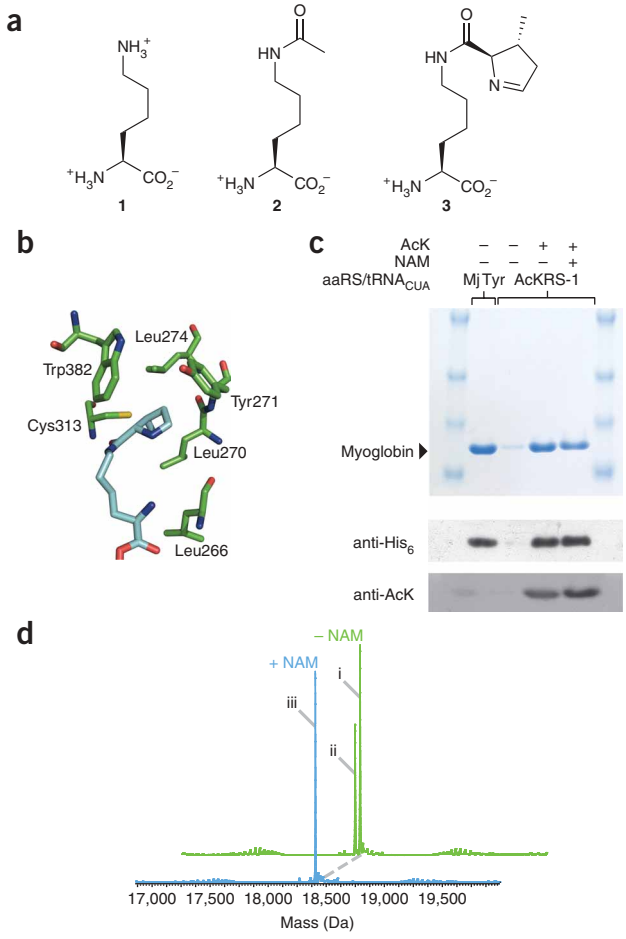

Figure 2 Design and evolution of an MbPylRS/tRNA$_{CUA}$ pair for the genetic incorporation of N^{ε}-acetyllysine. (**a**) Structure of lysine (**1**), N^{ε}-acetyllysine (**2**) and pyrrolysine (**3**). (**b**) Structure of the active site of *M. mazei* PylRS bound to pyrrolysine. The active site residues shown are conserved between *M. mazei* PylRS and *M. barkeri* PylRS. These residues form the hydrophobic binding pocket of pyrrolysine and are mutated in the library to each of the common 20 amino acids. The image was created using PyMol v0.99 (http://pymol.sourceforge.net/) and Protein Data Bank accession code 2Q7H. (**c**) Myoglobin-His$_6$ produced in the presence of MjTyrRS/MjtRNA$_{CUA}$ (lane 1) or in the presence of AcKRS-1 without or with 1 mM N^{ε}-acetyllysine (AcK, lanes 2 and 3, respectively), or in the presence of 1 mM acetyllysine and 50 mM NAM (lane 4). Proteins were purified by Ni^{2+} affinity chromatography, separated by SDS-PAGE and either stained with Coomassie or transferred to nitrocellulose and detected with antibodies to the His$_6$ tag or acetyllysine. (**d**) ESI-MS analysis of the purified acetylated myoglobin-His$_6$. Myoglobin-His$_6$ expressed in the absence of NAM (green) produced two peaks of masses 18,397.6 Da (ii) and 18,439.2 Da (i), which correspond to deacetylated and acetylated myoglobin-His$_6$ (predicted masses 18,396.2 Da and 18,438.2 Da, respectively). When myoglobin-His$_6$ was expressed in the presence of 50 mM NAM (blue), only the peak for the acetylated protein was observed (iii).

the hydrophobic cavity that binds the pyrroline ring in MbPylRS is rearranged to bind the acetyl group, and that the difference in volume between the pyrrolysine and N^{ε}-acetyllysine is compensated for by the larger volume of the mutant amino acids in the evolved synthetases.

To explicitly examine the fidelity and yield of acetyllysine incorporation in response to the amber codon, we used cells containing *Myo-his6(4TAG)*, AcKRS-1 and 1 mM N^{ε}-acetyllysine to produce full-length myoglobin-His_6. The protein was purified with a yield of 1.5 mg l^{-1} of culture (**Fig. 2**), which is comparable to yields reported for the incorporation of unnatural amino acids using the most active variants of the MjTyrRS/MjtRNA$_{CUA}$ pair[11]. Only trace amounts of myoglobin-His_6 were detected by Coomassie stain or western blot against the His_6 tag when N^{ε}-acetyllysine was withheld from cells. Western blots against N^{ε}-acetyllysine further confirmed the incorporation of the amino acid into myoglobin-His_6 and demonstrated that the evolved aminoacyl-tRNA synthetases are very selective for N^{ε}-acetyllysine.

Electrospray ionization mass spectra of myoglobin-His_6 purified from cells incorporating N^{ε}-acetyllysine show two peaks (**Fig. 2**): one peak corresponds to the incorporation of N^{ε}-acetyllysine, and the second peak has a mass of 42 Da less. We assigned the second peak to myoglobin-His_6 bearing lysine derived from post-translational deacetylation in *E. coli*. *E. coli* has a single characterized deacetylase—CobB, which is a sirtuin family nicotinamide adenine dinucleotide–dependent enzyme[14]. Since the sirtuin family of enzymes are known to be potently inhibited by nicotinamide (NAM, **6**), we performed protein expression in the presence of this inhibitor. Electrospray ionization spectra of myoglobin produced from cells containing NAM (**Fig. 2**) gave a single peak corresponding to the acetylated protein, with no peak observed for deacetylated protein. We conclude that NAM completely inhibits the post-translational deacetylation of genetically incorporated N^{ε}-acetyllysine in *E. coli*.

To begin to demonstrate the utility of our strategy for the synthesis of a protein that is naturally acetylated, we produced recombinant acetylated manganese superoxide dismutase (MnSOD). MnSOD is a tumor suppressor gene and is the principle antioxidant and superoxide scavenger in the mitochondrial matrix. A recent mass spectrometry–based proteomics study demonstrated that Lys44 of MnSOD is subject to acetylation in the mitochondria of HeLa cells[3]. We produced and purified rat mitochondrial His_6-MnSOD, containing N^{ε}-acetyllysine at position 44, from *E. coli* (**Supplementary Fig. 1** online). Electrospray ionization mass spectrometry demonstrates that the protein is homogeneously acetylated (**Supplementary Fig. 2** online), and MS/MS fragmentation of the relevant tryptic peptide confirms that the site of incorporation is as genetically encoded (**Supplementary Fig. 3** online). We find that recombinant purified His_6-MnSOD and His_6-MnSOD(K44AcK) have comparable activities (**Supplementary Fig. 4** online), which suggests that the role of acetylation may be to regulate another aspect of SOD function.

In conclusion, we have confirmed the orthogonality and efficiency of the MbPylRS/MbtRNA$_{CUA}$ pair in *E. coli*. We have evolved this pair to direct the incorporation of N^{ε}-acetyllysine, with high translational fidelity and efficiency, into proteins expressed in *E. coli*, and we have developed an inhibitor-based strategy to eradicate the initially observed post-translational deacetylation of cotranslationally incorporated N^{ε}-acetyllysine in *E. coli*. Finally, we have used our method to produce recombinant rat His_6-MnSOD bearing N^{ε}-acetyllysine at a site defined by mass spectrometry.

We are currently using our strategy for producing other site-specifically acetylated recombinant proteins to decipher the role of this modification in biology. We are addressing the synthesis of nucleosomes and chromatin bearing N^{ε}-acetyllysine at defined sites on particular histones, with the goal of determining the effect of defined modifications on nucleosome and chromatin structure and function[1]. It should be possible to further evolve the MbPylRS/MbtRNA$_{CUA}$ pair for the genetic incorporation of other modified lysines with roles in regulating chromatin structure, function and epigenetic inheritance[6]. Moreover, it should be possible to extend our method to genetically incorporate lysine residues derivatized with diverse functional groups and biophysical probes into proteins. Because MbPylRS does not recognize the anticodon of MbtRNA$_{CUA}$ (ref. 10), it may be possible to combine evolved MbPylRS/MbtRNA pairs with other evolved orthogonal aminoacyl-tRNA synthetase/tRNA$_{CUA}$ pairs and with orthogonal ribosomes possessing evolved decoding properties[15] to direct the efficient incorporation of multiple distinct useful unnatural amino acids in a single protein.

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We are grateful to K. Wang (Medical Research Council Laboratory of Molecular Biology, Cambridge) for assisting with early stages of the project, and to P.G. Schultz (Scripps Institute) for materials. J.W.C. is a European Molecular Biology Organization Young Investigator. This work was funded by the Medical Research Council, UK.

Published online at http://www.nature.com/naturechemicalbiology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Shogren-Knaak, M. *et al. Science* **311**, 844–847 (2006).
2. Jenuwein, T. & Allis, C.D. *Science* **293**, 1074–1080 (2001).
3. Kim, S.C. *et al. Mol. Cell* **23**, 607–618 (2006).
4. Yang, X.J. & Gregoire, S. *EMBO Rep.* **8**, 556–562 (2007).
5. Kouzarides, T. *Cell* **128**, 693–705 (2007).
6. Kouzarides, T. *EMBO J.* **19**, 1176–1179 (2000).
7. Hao, B. *et al. Science* **296**, 1462–1466 (2002).
8. Srinivasan, G., James, C.M. & Krzycki, J.A. *Science* **296**, 1459–1462 (2002).
9. Blight, S.K. *et al. Nature* **431**, 333–335 (2004).
10. Ambrogelly, A. *et al. Proc. Natl. Acad. Sci. USA* **104**, 3141–3146 (2007).
11. Chin, J.W., Martin, A.B., King, D.S., Wang, L. & Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **99**, 11020–11024 (2002).
12. Polycarpo, C.R. *et al. FEBS Lett.* **580**, 6695–6700 (2006).
13. Kavran, J.M. *et al. Proc. Natl. Acad. Sci. USA* **104**, 11268–11273 (2007).
14. Schwer, B., Bunkenborg, J., Verdin, R.O., Andersen, J.S. & Verdin, E. *Proc. Natl. Acad. Sci. USA* **103**, 10224–10229 (2006).
15. Wang, K., Neumann, H., Peak-Chew, S.Y. & Chin, J.W. *Nat. Biotechnol.* **25**, 770–777 (2007).