



Cite this: *RSC Adv.*, 2022, 12, 34142

Received 5th August 2022  
Accepted 19th November 2022

DOI: 10.1039/d2ra04898a

rsc.li/rsc-advances

# Genetic encoding of isobutyryl-, isovaleryl-, and $\beta$ -hydroxybutyryl-lysine in *E. coli*<sup>†</sup>

Jayani A. Christopher, Sahan A. Galbada Liyanage, Eve M. Nicholson, William D. Kinney and T. Ashton Cropp \*

Here we report the synthesis and genetic encoding of the lysine post translational modifications,  $\beta$ -hydroxybutyryl-lysine, isobutyryl-lysine and isovaleryl-lysine. The ability to obtain a homogenous protein samples with site-specific incorporation of these acylated lysine residues can serve as a powerful tool to study the biological role of lysine post translational modifications.

Lysine residues undergo several different types of post-translational modifications (PTMs) such as methylation, SUMOylation, ubiquitination, glycation, and acylation.<sup>1</sup> Lysine acylation is the most common type of PTM that has been studied extensively in the past due to its direct impact on cell functions such as gene transcription,<sup>2,3</sup> energy metabolism,<sup>4</sup> nuclear transport,<sup>4</sup> sub-cellular protein localization,<sup>5</sup> protein-DNA interactions,<sup>5</sup> apoptosis,<sup>6</sup> and cell metabolism.<sup>7–10</sup> Acylation occurring in lysine residues at the tails of the histone proteins can have a profound effect on the regulation of gene transcription.<sup>11–13</sup> At physiological pH, lysine residues of histones carry positive charges due to the protonation of epsilon nitrogen, and hence leads to the tight binding with negatively charged DNA backbone in chromatin structure. By undergoing acylation, these amine groups lose the positive charge, thereby minimizing the tightly bound interactions giving fluidity to the DNA strands. This can increase gene expression because the DNA becomes more accessible to be transcription.<sup>13–15</sup>

The lysine PTM,  $\beta$ -hydroxybutyryl-lysine, (HBK) (Fig. 1), was discovered in 2016 as a modification of histone proteins.<sup>9</sup> It was shown that this modification is directly related to the concentration of the corresponding ketone bodies available in the cell, and is thus heavily influenced by metabolism. Interestingly, HBK not only quenches the lysine positive charge, but the presence of a hydroxyl group introduces the potential for additional hydrogen bonding, that could contribute to the alteration of protein-protein or protein DNA interactions.<sup>8</sup> Interestingly a similar structure, isobutyryl-lysine (IBK), was also recently discovered as a histone modification mark.<sup>16</sup> Indeed, many other acylations that are derived from the corresponding acyl-CoAs have been observed, including isovaleryl-lysine (IVK).<sup>17</sup> Given the structural similarity of these amino

acids, and the importance of understanding epigenetic regulation, we attempted to add these three amino acids to the genetic code.

Previously, we reported the successful incorporation of a different lysine PTM, 2-hydroxyisobutyryl lysine PTM into the genetic code of *E. coli* using the pyrrolysyl-tRNA synthetase (PylRS) and the orthogonal tRNA<sup>CUA</sup> originally from *Methanosarcina mazei* species.<sup>18</sup> Xiao and co-workers also successfully incorporated this PTM into recombinant histone proteins in *E. coli* and mammalian cells using an evolved PylRS/tRNA<sup>CUA</sup> pair which was initially obtained from *Methanosarcina barkeri* (Mb) species.<sup>19</sup> Here we show that this pair can also be used for the successful genetic encoding of these related PTMs. These analogues have been identified as potential lysine PTMs due to the availability of the corresponding acyl-CoAs from metabolic

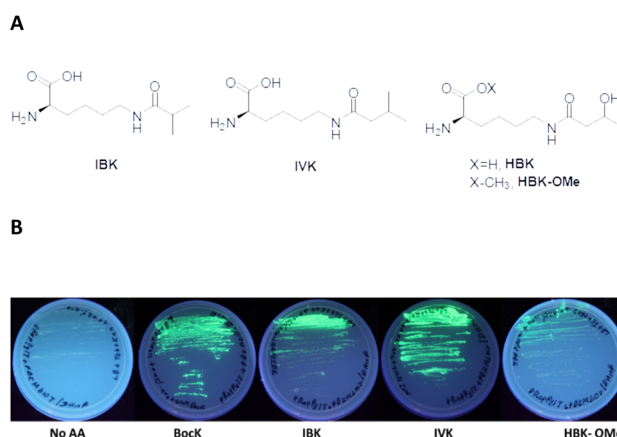


Fig. 1 (A) Chemical structures of isobutyryl-lysine (IBK), isovaleryl-lysine (IVK), and  $\beta$ -hydroxybutyryl-lysine (HBK). (B) Reporter for T7-controlled expression of GFPuv in the presence and absence of 4 mM acylated lysine PTMs. N<sup>ε</sup>-Boc-lysine (Bock) is shown as positive control. Suppression of amber stop codon in T7-RNA polymerase leads to induced expression of GFPuv.

Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284, USA. E-mail: tacropp@vcu.edu

<sup>†</sup> Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2ra04898a>



pathways and similar counter parts have been discovered in different proteins.<sup>17,20</sup>

We first prepared the HBK amino acid in several steps from commercially available benzoyl protected Boc-lysine ( $N^{\epsilon}$ -Cbz-Boc-lysine). The acylating agent, ethyl-3-hydroxybutyrate was protected by a TBDMS group to avoid reactivity of the 3-hydroxyl group, which we found critical for success (Scheme S2†). The sequence of our synthesis allowed us to produce the amino acid as a methyl ester (HBK-OMe), which was expected to be more bioavailable than the free amino acid. Methyl esters of amino acids can be hydrolysed in the cell,<sup>21</sup> and we found that in the case of HBK, esterification was essential for successful protein production. Both IVK and IBK were synthesized as free amino acids using a similar synthetic route starting with Boc-lysine and isovaleryl chloride and isobutyl chloride for acylation, respectively (Scheme S3†). It was not necessary to use these as methyl esters.

As an initial approach, we performed a genetic selection against a *Methanosarcina mazei* PylRS (MbPylRS) active site library using HBK-OMe. Surprisingly, the vast majority of enriched clones from this library consisted of wild-type active PylRS sequences. Importantly, these gene sequences consisted of many variants of degenerate codons that encoded the wild-type residues, suggesting that the wild-type sequence was enriched from the selection. To confirm the substrate specificity of the synthetase, a screening was performed by using a reporter plasmid that expresses GFPuv from a T7 promoter, and a T7-RNA polymerase gene containing an in-frame amber stop codon. This reporter is quite sensitive as a single amber suppression even can lead to multiple copies of GFPuv being expressed. As shown in Fig. 1, fluorescence was observed only in the presence of the acylated lysines, with little to no background expression in the absence of supplemented amino acids. The observed activity was similar to a positive control,  $N^{\epsilon}$ -Boc-Lys, which is often used as a benchmark in unnatural amino acid mutagenesis studies. This indicates that all three amino acids serve as substrates for the wild-type MbPylRS.

To isolate protein containing these amino acids, we performed an expression of superfolder GFP (His6-sfGFP) containing an amber stop codon mutation in place of Y151, and an N-terminal 6X-histidine tag. The expression was performed by growing the cultures in the presence of 2 mM IBK, IVK and HBK-OMe and purified the His tagged proteins with  $Ni^{2+}$  affinity chromatography under native purification conditions. Protein

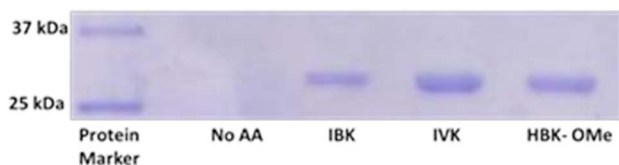


Fig. 2 His6-sfGFP (Y151TAG) expression in the presence and absence of the 2 mM acylated lysines. No expression was seen in the absence of the amino acids. In the presence of the amino acid a band at 28 kDa was observed which corresponds to His6-sfGFP containing a mutation at position 151.

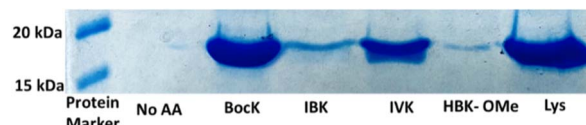


Fig. 3 Histone H3 (K9TAG) expression in the presence and absence of the 4 mM acylated lysines. Each lane corresponds to 20  $\mu$ L of eluted protein loaded, and is indicative of the overall expression levels. No expression was seen in the absence of the exogenous amino acids. In the presence of the amino acid, a band at 17 kDa was observed which corresponds to full-length H3 protein. A control expression of wild-type H3 containing a lysine codon at position 9 (Lys) is included as a control.

expression was only observed in the presence of the amino acids (Fig. 2), again confirming that endogenous amino acids are not substrates for the synthetase. The ESI-MS mass spectra of the intact proteins confirmed incorporation of the correct amino acid (Fig. S1†). The mass spectra did indicate the presence of some free lysine residues, perhaps the result of hydrolysis reactions *in vivo*, or during protein purification (Fig. S1†).

Finally, to demonstrate expression of the proteins most likely to be studied using these amino acids, we expressed human histone H3 protein containing the amber stop codon in place of the codon for lysine 9. This position has been observed to contain lysine modifications, including HBK, in cellular studies. The protein expression was performed with 4 mM BocK, IBK, IVK and HBK-OMe cultures alongside a wild-type histone H3 containing a native K9 residue. H3-proteins were purified by  $Ni^{2+}$  affinity chromatography under denaturing conditions, which has been used in the past to successfully re-fold functional protein.<sup>3</sup> The protein expression was observed only in the presence of the acylated and wildtype lysines. The expression with IVK showed the highest yield when, whereas HBK-OMe expression was observed to be the lowest (Fig. 3). Nevertheless, as seen before, no expression was observed in the absence of exogenous amino acids indicating efficient site-specific incorporation. While HBK expression was low in this example, in other protein expressions, yields are comparable.

In summary, we were able to successfully synthesize and add new lysine PTMs to the genetic code of *E. coli* using the Mb PylRS/PylRS tRNA<sup>CUA</sup> orthogonal pair. This leverages this system and enables the synthesis of natural human histone proteins containing newly discovered PTMs at site-specific locations. These protein products could serve as unique substrates to be used in deciphering the substrate specificity of enzymes that act on PTMs such as histone deacetylases. Moreover, this approach should enable the in-depth study of the consequences of these modifications with residue-level precision.

## Autor contributions

J. Christopher, S. Liyanage, and T. A. Cropp conceived and designed the experiments. J. Christopher, S. Liyanage, E. Nicholson, and W. Kinney performed experiments and analysed data. J. Christopher and T. A. Cropp prepared the manuscript.

## Conflicts of interest

The authors indicate there are no conflicts to declare.

## Acknowledgements

The authors acknowledge NSF CHE-2109051 for support of this work.

## Notes and references

- 1 C. Kontaxi, P. Piccardo and A. C. Gill, *Front. Mol. Biosci.*, 2017, **4**, 56.
- 2 X. J. Yang and E. Seto, *Mol. Cell*, 2008, **31**, 449–461.
- 3 B. J. Wilkins, L. E. Hahn, S. Heitmuller, H. Frauendorf, O. Valerius, G. H. Braus and H. Neumann, *ACS Chem. Biol.*, 2015, **10**, 939–944.
- 4 S. Lee, *Toxicol. Res.*, 2013, **29**, 81–86.
- 5 S. C. Kim, R. Sprung, Y. Chen, Y. Xu, H. Ball, J. Pei, T. Cheng, Y. Kho, H. Xiao, L. Xiao, N. v. Grishin, M. White, X. J. Yang and Y. Zhao, *Mol. Cell*, 2006, **23**, 607–618.
- 6 T. Yanagisawa, T. Umehara, K. Sakamoto and S. Yokoyama, *ChemBioChem*, 2014, **15**, 2181–2187.
- 7 D. Zhang, Z. Tang, H. Huang, G. Zhou, C. Cui, Y. Weng, W. Liu, S. Kim, S. Lee, M. Perez-Neut, J. Ding, D. Czyz, R. Hu, Z. Ye, M. He, Y. G. Zheng, H. A. Shuman, L. Dai, B. Ren, R. G. Roeder, L. Becker and Y. Zhao, *Nature*, 2019, **574**, 575–580.
- 8 B. R. Sabari, D. Zhang, C. D. Allis and Y. Zhao, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 90–101.
- 9 Z. Xie, D. Zhang, D. Chung, Z. Tang, H. Huang, L. Dai, S. Qi, J. Li, G. Colak, Y. Chen, C. Xia, C. Peng, H. Ruan, M. Kirkey, D. Wang, L. M. Jensen, O. K. Kwon, S. Lee, S. D. Pletcher, M. Tan, D. B. Lombard, K. P. White, H. Zhao, J. Li, R. G. Roeder, X. Yang and Y. Zhao, *Mol. Cell*, 2016, **62**, 194–206.
- 10 M. D. Hirschev and Y. Zhao, *Mol. Cell. Proteomics*, 2015, **14**, 2308–2315.
- 11 Z. A. Wang, Y. Zeng, Y. Kurra, X. Wang, J. M. Tharp, E. C. Vatansever, W. W. Hsu, S. Dai, X. Fang and W. R. Liu, *Angew. Chem., Int. Ed.*, 2017, **56**, 212–216.
- 12 L. Dai, C. Peng, E. Montellier, Z. Lu, Y. Chen, H. Ishii, A. Debernardi, T. Buchou, S. Rousseaux, F. Jin, B. R. Sabari, Z. Deng, C. D. Allis, B. Ren, S. Khochbin and Y. Zhao, *Nat. Chem. Biol.*, 2014, **10**, 365–370.
- 13 B. R. Sabari, Z. Tang, H. Huang, V. Yong-Gonzalez, H. Molina, H. E. Kong, L. Dai, M. Shimada, J. R. Cross, Y. Zhao, R. G. Roeder and C. D. Allis, *Mol. Cell*, 2015, **58**, 203–215.
- 14 T. Narita, B. T. Weinert and C. Choudhary, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 156–174.
- 15 M. J. Gattner, M. Vrabel and T. Carell, *Chem. Commun.*, 2013, **49**, 379–381.
- 16 Z. Zhu, Z. Han, L. Halabelian, X. Yang, J. Ding, N. Zhang, L. Ngo, J. Song, H. Zeng, M. He, Y. Zhao, C. H. Arrowsmith, M. Luo, M. G. Bartlett and Y. G. Zheng, *Nucleic Acids Res.*, 2021, **49**, 177–189.
- 17 T. Baldensperger, S. di Sanzo, A. Ori and M. A. Glomb, *Anal. Chem.*, 2019, **91**, 12336–12343.
- 18 W. A. Knight and T. A. Cropp, *Org. Biomol. Chem.*, 2015, **13**, 6479–6481.
- 19 H. Xiao, W. Xuan, S. Shao, T. Liu and P. G. Schultz, *ACS Chem. Biol.*, 2015, **10**, 1599–1603.
- 20 J. Baeza, M. J. Smallegan and J. M. Denu, *Trends Biochem. Sci.*, 2016, **41**, 231–234.
- 21 H. Zhou, J. W. Cheung, T. Carpenter, S. K. Jones, N. H. Luong, N. C. Tran, S. E. Jacobs, S. A. Galbada Liyanage, T. A. Cropp and J. Yin, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 126876.

