Bioorganic & Medicinal Chemistry Letters xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

# Enhancing the incorporation of lysine derivatives into proteins with methylester forms of unnatural amino acids

Han Zhou[a], Jenny W. Cheung[b], Tomaya Carpenter[a], Stacey K. Jones Jr.[a], Nhu H. Luong[a], Nhi C. Tran[a], Savannah E. Jacobs[a], Sahan A. Galbada Liyanage[b], T. Ashton Cropp[b,c,*], Jun Yin[a,*]

[a] Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, United States
[b] Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284, United States
[c] Institute for Structural Biology, Drug Discovery and Development, Virginia Commonwealth University, Richmond, VA 23219, United States

ARTICLE INFO

*Keywords:*
Unnatural amino acid
Methyl ester
Ubiquitin
Aminoacyl-tRNA synthetase
Protein expression

ABSTRACT

We have improved the incorporation of L- and D-forms of unnatural amino acid (UAA) $N^{\varepsilon}$-thiaprolyl-L-lysine (ThzK) into ubiquitin (UB) and green fluorescent protein (GFP) by 2–6 folds with the use of the methylester forms of the UAAs in *E coli* cell culture. We also improved the yields of UAA-incorporated UB and GFP with the methylester forms of $N^{\varepsilon}$-Boc-L-Lysine (BocK) and $N^{\varepsilon}$-propargyl-L-Lysine (PrK) by 2–5 folds compared to their free acid forms. Our work demonstrated that using methylester-capped UAAs for protein expression is a useful strategy to enhance the yields of UAA-incorporated proteins.

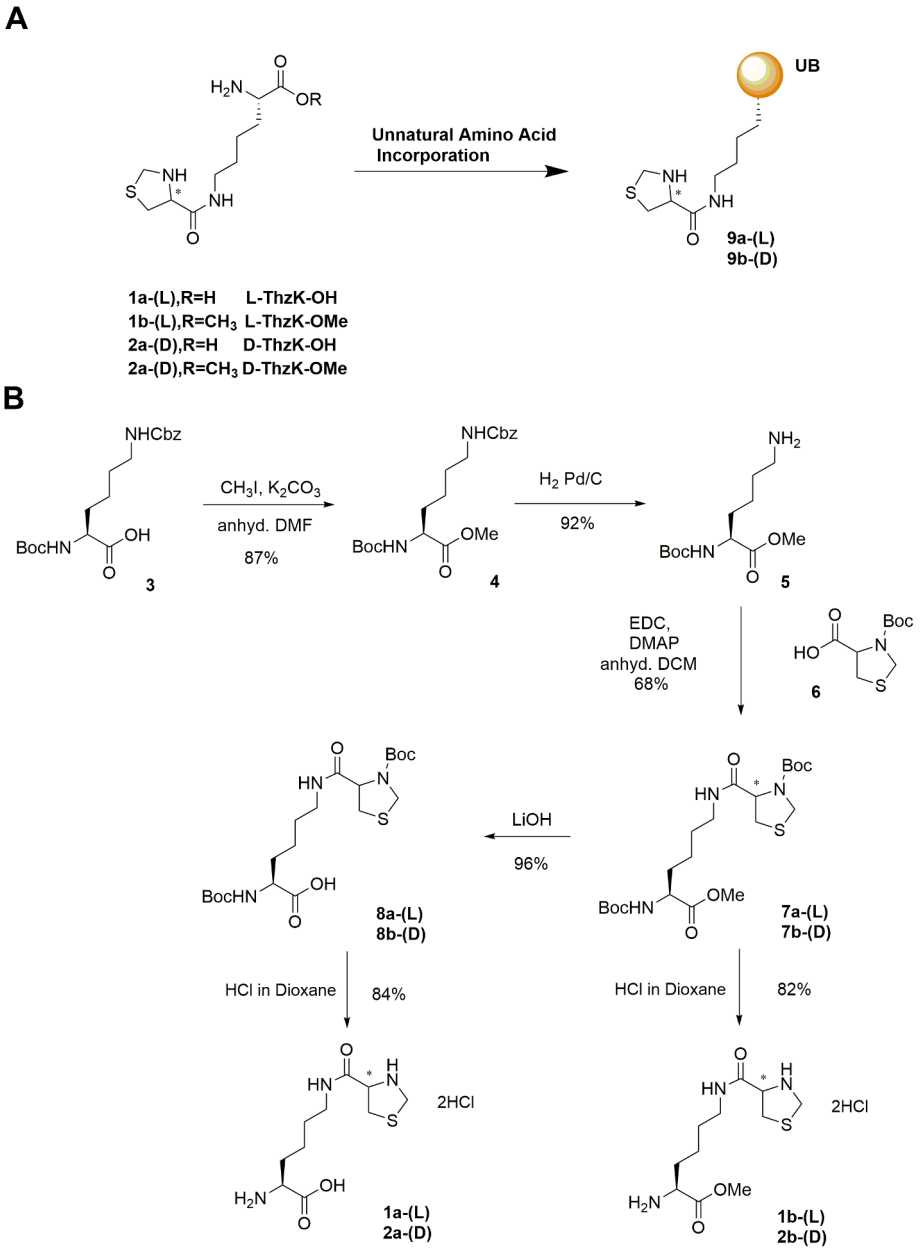
Unnatural amino acid (UAA) incorporation into proteins *via* genetic code expansion has greatly expanded the structural diversity and chemical reactivity of the functionalities on a protein scaffold.[1–5] UAA-incorporated proteins have become powerful tools for biological studies – they can generate precise posttranslational modification patterns,[6] or act as chemical probes to enable photocrosslinking,[7] biorthogonal labeling,[8] photocaging[9] or enzymatic profiling capacities[10] on the proteins. Currently, more than 150 UAAs have been incorporated into proteins with engineered tRNA-tRNA synthetase (RS) pairs.[5,11–13] Enhancing the yield of UAA-incorporated proteins would save efforts on UAA synthesis for protein expression and expand the use of UAA incorporation as a powerful tool of chemical biology. Here we took the incorporation of $N^{\varepsilon}$-L-thiaprolyl-L-lysine (L-ThzK, **1a**), and $N^{\varepsilon}$-D-thiaprolyl-L-lysine (D-ThzK, **2a**) into ubiquitin (UB) and superfolder green fluorescent protein (sfGFP) as a case study (Scheme 1A) and found supplying the *E. coli* cell cultures with the methylester forms of the UAAs would significantly enhance the yields of UAA-incorporated proteins. We also showed that the methylester forms of UAAs are effective in enhancing the incorporation of $N^{\varepsilon}$-Boc-L-Lysine (BocK) and $N^{\varepsilon}$-propargyl-L-Lysine (PrK) into UB and sfGFP. Our work demonstrates that optimizing the cellular uptake of UAAs by methylester capping could be a useful approach to enhance the yields of UAA-incorporated proteins and make them more accessible for the study of biological problems.

Chin and co-workers recently engineered pyrrolysyl (Pyl) tRNA synthetase (RS) from *Methanosarcina barkeri* (*Mb*) for the incorporation of L-ThzK UAA.[14] After the incorporation, the thiaprolyl ring in the ThzK side chain can be opened by treatment with methoxyamine to generate a Cys-conjugated Lys side chain. The 1,2-aminiothiol functionality on the Cys residue can be a reactive handle for conjugation with cyano-derivatized probes or with C-terminal thioester of proteins through expressed protein ligation.[14,15] The same study showed that D-ThzK can be incorporated into proteins with wild type (wt) *Mb* PylRS. Subsequently, variants of PylRS were developed that would incorporate δ-thiol-L-Lys into proteins to enable the synthesis of protein conjugates with isopeptide linkages at specific Lys residues.[16] We rationalized that enhancing the incorporation of L/D-ThzK UAA into proteins would enable the synthesis of protein conjugates to study protein posttranslational modification by UB or UB-like proteins. [17–18]

Instead of preparing ThzK from $N^{\varepsilon}$-Boc-L-Lys-OMe **5**,[14] we chose *N*-Boc-Lys(Cbz)-OH **3,** a cheaper starting material to begin our synthesis of the ThzK UAA (Scheme 1B). We first methylated **3** to generate protected Lys methylester **4** and then deprotected Cbz group on **4** by hydrogenation to afford **5**. We then coupled **5** with L- or D-form of Boc-thiazolidine-4-carboxylic acid **6** to afford Boc and methylester protected ThzK (**7a** and **7b**). We either removed all the protecting groups on **7a** and **7b** to generate the free acid forms of L/D-ThzK-OH (**1a** and **2a**) or only removed the Boc group to generate the methylester forms of ThzK

* Corresponding authors at: Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284, United States (T.A. Cropp), and Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, United States (J. Yin).
*E-mail addresses:* tacropp@vcu.edu (T.A. Cropp), junyin@gsu.edu (J. Yin).

https://doi.org/10.1016/j.bmcl.2019.126876
Received 16 July 2019; Received in revised form 26 November 2019; Accepted 27 November 2019
0960-894X/ © 2019 Elsevier Ltd. All rights reserved.




**Scheme 1.** (A) Incorporating L/D-ThzK in the free acid forms (**1a**/**2a**) or in the methylester forms (**1b**/**2b**) into ubiquitin (UB). (B) Synthetic scheme for L/D-ThzK-OH **1a**/**2a** and L/D-ThzK-OMe **1b**/**2b**.

(ThzK-OMe, **1b** and **2b**). Through this procedure, we achieved an overall yield of 25% for the synthesis of ThzK-OMe. All ThzK UAAs were purified by HPLC and their identities confirmed by ESI and NMR (Supplementary Figs. S1–S4). Free acid and methylester forms of the UAA were stable after HPLC purification and lyophilization. They could be stored in the 4 °C cold box for weeks with no apparent decomposition.

To test the incorporation efficiency, we first used the engineered *Mb* L-ThzKRS[14] to express UB with L-ThzK incorporated at the *K*11 position (Scheme 1A). We supplied the cell culture with 1 mM L-ThzK-OH (**1a**) and found 1 L of cell culture would yield ~ 6 mg of UB with L-ThzK incorporated after purifying the 6 × His-tagged UB by a Ni-NTA affinity column (Fig. 1A). We determined protein concentration in the elution by Bio-Rad protein assay.[19] We also attempted the incorporation of D-ThzK-OH (**2a**) into UB with wt *Mb* PylRS and found a similar yield of UB expression (Fig. 1B). The expression and purification of UB with the *K*11 amber codon in the absence of UAA is shown in Supplementary Fig. S5.

We suspected the L- and D-ThzK-OH in their free carboxylic acid forms could have poor membrane permeability to the *E coli* cell and so they would not be readily available for protein expression inside the cells. We thus tried methylester forms of L- and D-ThzK-OMe (**1b** and **2b**) for ThzK incorporation into K11 of UB. We supplied 1 mM L-ThzK-OMe to the cell culture to compare the efficiency of incorporation with the same concentration of L-ThzK-OH with the coexpression of *Mb* L-ThzKRS. We found we could express 33 mg ( ± 3.0 mg) of L-ThzK-incorporated UB from 1 L of culture among three trials (Fig. 1E). In contrast, the yield of L-ThzK incorporation into UB with L-ThzK-OH was only 6.0 mg ( ± 1.5 mg) among three trails (Fig. 1A). We found similar improvements when we compared the yields of D-ThzK incorporation into UB with D-ThzK-OMe and D-ThzK-OH in the presence of *Mb* PylRS. The methyl ester form of UAA yielded 37 mg ( ± 3.0 mg) of UB from 1 L of culture, while the free acid form yielded 6.3 mg ( ± 1.5 mg) of UB (Fig. 1B and 1F). The correct incorporation of ThzK into UB using L- or

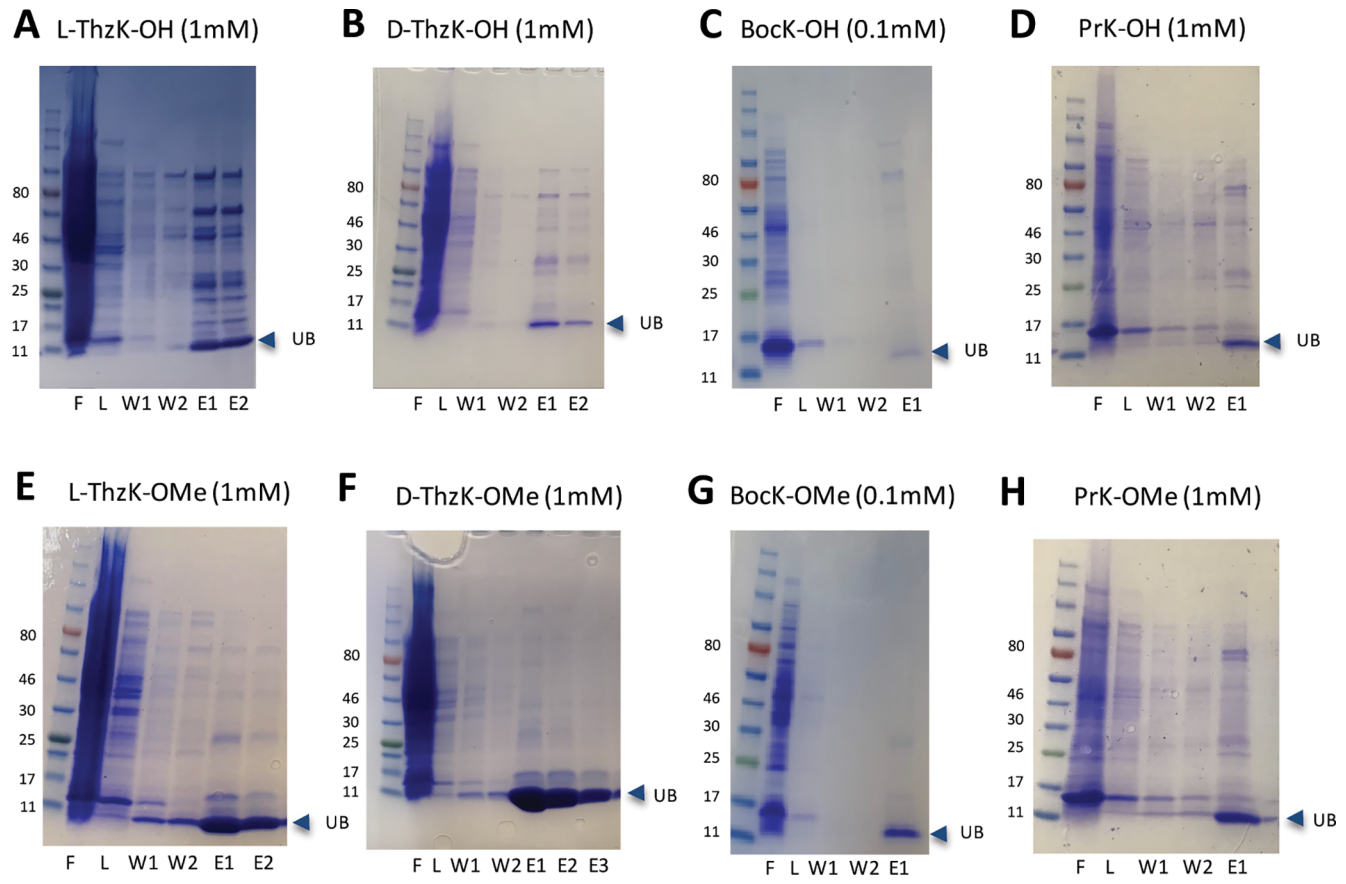



**Fig. 1.** Incorporating L-ThzK, D-ThzK, Bock, Prk into UB at the K11 position with the frees acid or methylester forms of the UAA. (A) UB expression with *Mb* L-ThzKRS and L-ThzK-OH. The PAGE gel was stained with Coomassie blue to analyse samples from the flow through of the Ni-NTA column (F), once wash by lysis buffer (L), twice by wash buffer (W1 and W2), and twicw by elution buffer (E1 and E2). (B) UB expression with wt *Mb* PylRS and D-ThzK-OH. (C) UB expression with wt *Mb* PylRS and BocK-OH. (D) UB expression with wt *Mb* PylRS and PrK-OH. (E) UB expression with *Mb* L-ThzKRS and L-ThzK-OMe. (F) UB expression with wt *Mb* PylRS and D-ThzK-OMe. (G) UB expression with wt *Mb* PylRS and BocK-OMe. (G) UB expression with wt *Mb* PylRS and PrK-OMe. 1.0 mM of free acid or methyester forms of L/D-ThzK and Prk were used for UB expression in (A), (B), (D), (E), (F), and (H), and 0.1 mM of free acid or methylester forms of BocK UAA was used for UB expression in (C) and (G).

D-ThzK-OMe was confirmed by MALDI mass spectrometry comparing the difference of the molecular weights (MW) of UB with either ThzK or Lys incorporated (Supplementary Figs. S6–S8).

We then tested if the methylester forms of other UAAs such as BocK and PrK would enhance their incorporation into UB with wt *Mb* PylRS. We found the yield of BocK incorporation into UB had a high yield of more than 50 mg per liter with 1 mM of either BocK-OH or BocK-OMe. However, when we lowered the concentration of the UAA, the methylester form of BocK showed a clear advantage in enhancing the yield of BocK-incorporated proteins. At 0.1 mM, BocK-OMe yielded 14 mg of UB ( ± 1.5 mg) from 1 L of culture, while BocK-OH yielded 4.0 mg of UB ( ± 1.0 mg) (Fig. 1C and 1G). When the UAA concentrations were lowered to 0.01 mM, BocK-OMe yielded 10 mg of UB ( ± 1.5 mg) from 1 L of culture, while BocK-OH yielded 1.0 mg of UB ( ± 0.5 mg) (Supplementary Fig. S9). The incorporation of BocK into to UB with the methylester form of the UAA was confirmed by MALDI mass spectrometry (Supplementary Fig. S10).

We also prepared PrK-OH and PrK-OMe to compare their efficiency of incorporation into UB with wt *Mb* PylRS. We found 1 mM PrK-OMe yielded 31 mg UB ( ± 1.5 mg) from 1 L of culture while same concertation of PrK-OH yielded 15 mg UB ( ± 1.5 mg) from 1 L of culture (Fig. 1D and 1H). The incorporation of PrK into UB with the methylester form of the UAA was confirmed by MALDI mass spectrometry (Supplementary Fig. S11). Thus, the methylester form of PrK also enhanced the yield of its incorporation into proteins.

Besides UB, we used sfGFP as a model system to compare the yields of protein expression with free acid or methylester forms of the UAAs. After the induction of sfGFP expression, we analyzed the yields of protein expression by polyacrylamide gel electrophoresis (PAGE) and Coomassie blue staining of the gels (Fig. 2). The incorporation efficiency of either L- or D-ThzK-OH into position 151 of sfGFP at 2 mM of the UAA was low when wt PylRS of the *Methanosarcina mazei* (*Mm*) or *Mb* origin were used for protein expression (Fig 2A and 2B). When we used the engineered L-ThzKRS of *Mb* origin for incorporation, the yield of L/D-ThzK-OH incorporation into sfGFP was still low as judged by the weak bands of sfGFP on the PAGE gel (Fig. 2C). When 2 mM BocK-OH was used, there is a decent yield of sfGFP with the coexpression of *Mm* or *Mb* PylRS (Fig. 2A and 2B). BocK UAA is known for its high efficiency of incorporation into various proteins by wt PylRS, so such a result was expected.[20] Correspondingly, the sfGFP fluorescence from the cell culture showed high intensity with BocK incorporation into sfGFP by wt *Mm* or *Mb* PylRS. In contrast, the fluorescence intensities associated with L- or D-ThzK-OH incorporation were low with either wt PylRS or engineered L-ThzKRS (Fig. 2D).

When we performed ThzK incorporation into sfGFP with 2 mM L- or D-ThzK-OMe, we found higher sfGFP expression with the presence of L-ThzKRS compared to the cell cultures with the free acid forms of L/D-ThzK as suggested by the Coomassie blue staining of the PAGE gel of the crude cell lysates (Fig. 2C). sfGFP fluorescence from the cell culture with L- or D-ThzK-OMe was also significantly higher than the corresponding culture using the free acids of UAA (Fig. 2D). Out of three trials, L-ThzK-OMe gave 2-fold stronger fluorescence signal than the free

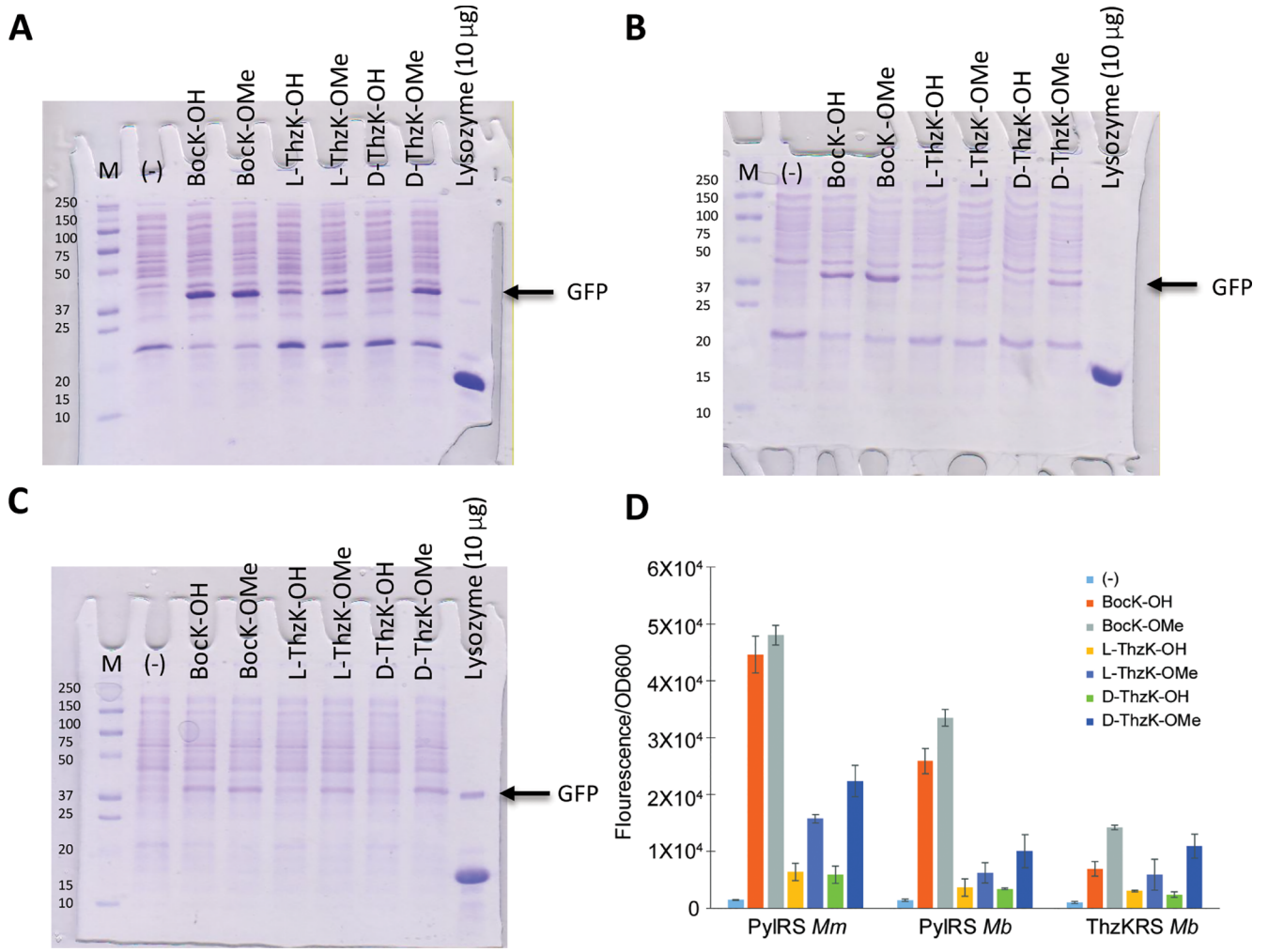



**Fig. 2.** Comparing the yield of BocK and L/D-ThzK incorporation into sfGFP with wt *Mm* and *Mb* PylRS and L-ThzKRS. (A) Using wt *Mm* PylRS to incorporate various forms of UAA into sfGFP. Efficiency of incorporation was evaluated based on the intensity of the sfGFP band on the PAGE gel of the crude cell lysates. (B) Using wt *Mb* PylRS to incorporate various forms of UAA in sfGFP. (C) Using *Mb* L-ThzKRS to incorporate various forms of UAA into sfGFP. (D) The intensities of fluorescences signals from cells expressing sfGFP with various UAA and tRNA synthetase pairs. The results are the average of three trials. In (A) to (D), 2 mM of UAAs were used for protein expression. (−) sign designates the control expression in absence of the UAAs, and M is the protein marker.

acid with either wt *Mb* or *Mm* PylRS or engineered L-ThzKRS for incorporation, and D-ThzK-OMe gave 3-5-fold stronger fluorescence signal than the corresponding free acid with wt or engineered PylRS.

BocK-OH and BocK-OMe did not show as large a difference in the yields of sfGFP expression when they were supplied at high concentrations (2 mM) in the cell culture. We thus varied the concentrations of the two forms of BocK UAA in the lower range of concentrations for sfGFP expression in the presence of wt *Mb* PylRS. We found the yield of sfGFP expression with BocK-OMe was at least 4-fold higher than with Boc-OH when the UAA concentrations were varied between 0.01 and 0.075 mM (Supplementary Fig. S12A). When the BocK UAA concentrations were greater than 0.1 mM, the difference in sfGFP expression became less, but still, the methylester form of BocK gave a higher yield of sfGFP than the free acid form up to 0.5 mM of the UAAs (Supplementary Fig. S12B). The difference in the yields of sfGFP with BocK-OH and BocK-OMe was confirmed by PAGE analysis of the cell lysates (Supplementary Fig. S12C). These results suggest that for a UAA such as BocK, which can be incorporated into proteins with high efficiency in a free acid form, the same yield of protein can be achieved with a lower concentration of the UAA in the methylester form. Thus, UAA material can be saved for large scale protein expressions by being supplied as a methylester. This could be an advantage for using the methylester forms of the UAAs for incorporation since many UAAs are not readily available and require multistep synthesis to prepare.

Our work showed that the yield of L/D-ThzK incorporation into UB could be increased by as much as 5-fold with the use of the methylester forms of the UAA. We also observed a similar enhancement in yields of UAA-incorporated proteins with the methylester forms of BocK and PrK comparing to their free acid forms. With the use of the same tRNA synthetases and *E coli* strains for incorporation, the enhancement of the yield is likely due to the improved membrane permeability of the methylester forms of the UAA to *E coli* cells comparing to the free acid forms. Previously Wang and co-workers demonstrated the beneficial effect of the acetoxymethyl ester form of a Tyr UAA analog in enhancing its incorporation into GFP in mammalian HEK293 cells.[21] Liu's group reported the use of methylester form of L-ThzK to achieve good yield of UB expression in *E. coli* cells.[22,23] Here we carried out a detailed comparison of the yields of L- and D-ThzK incorporation between the methylester and free acid forms of the UAA and found ThzK methylester could enhance UAA incorporation by ~6 fold. We also found the incorporation of BocK and Prk UAA were benefitted from using their methylester forms. We thus suggest UAA methylesters could enhance the yield of UAA-incorporated proteins and expand their use as chemical probes with designed functionalities.

ARTICLE IN PRESS



## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by grants from National Science Foundation (NSF) (1420193 to J.Y. and 1710460 to J.Y. and T.A.C.) and National Institutes of Health (NIH) (R01GM104498 to J.Y.) of the United States. We thank Jason Chin of Medical Research Council Laboratory of Molecular Biology, Cambridge, UK, for kindly sharing the *Mb* L-ThzKRS plasmid and helpful discussions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmcl.2019.126876.

## References

1. Lang K, Chin JW. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem Rev.* 2014;114(9):4764–4806.
2. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem.* 2010;79:413–444.
3. Wang L, Schultz PG. Expanding the genetic code. *Angew Chem Int Engl.* 2004;44:34–66.
4. Nguyen TA, Cigler M, Lang K. Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew Chem Int Ed Engl.* 2018;57(44):14350–14361.
5. Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment – expanding the chemistry in biology. *Chem Sci.* 2015;6(1):50–69.
6. Chen H, Venkat S, McGuire P, Gan Q, Fan C. Recent development of genetic code expansion for posttranslational modification studies. *Molecules (Basel, Switzerland).* 2018;23.
7. Pham ND, Parker RB, Kohler JJ. Photocrosslinking approaches to interactome mapping. *Curr Opin Chem Biol.* 2013;17:90–101.
8. Lee KJ, Kang D, Park HS. Site-specific labeling of proteins using unnatural amino acids. *Mol Cells.* 2019;42(5):386–396.
9. Wu N, Deiters A, Cropp TA, King D, Schultz PG. A genetically encoded photocaged amino acid. *J Am Chem Soc.* 2004;126(44):14306–14307.
10. Huguenin-Dezot N, Alonzo DA, Heberlig GW, et al. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature.* 2019;565:112–117.
11. Chin JW. Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem.* 2014;83:379–408.
12. Wang L. Genetically encoding new bioreactivity. *New Biotechnol.* 2017;38(Pt A):16–25.
13. Wang L. Engineering the genetic code in cells and animals: biological considerations and impacts. *Acc Chem Res.* 2017;50(11):2767–2775.
14. Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J Am Chem Soc.* 2011;133(30):11418–11421.
15. Li X, Fekner T, Ottesen JJ, Chan MK. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew Chem Int Ed Engl.* 2009;48(48):9184–9187.
16. Virdee S, Kapadnis PB, Elliott T, et al. Traceless and site-specific ubiquitination of recombinant proteins. *J Am Chem Soc.* 2011;133:10708–10711.
17. Braxton CN, Quartner E, Pawloski W, Fushman D, Cropp TA. Ubiquitin Chains Bearing Genetically Encoded Photo-Cross-Linkers Enable Efficient Covalent Capture of (Poly)ubiquitin-Binding Domains. *Biochemistry.* 2019;58:883–886.
18. Castaneda C, Liu J, Chaturvedi A, Nowicka U, Cropp TA, Fushman D. Nonenzymatic assembly of natural polyubiquitin chains of any linkage composition and isotopic labeling scheme. *J Am Chem Soc.* 2011;133:17855–17868.
19. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976;72:248–254.
20. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem Biophys Res Commun.* 2008;371:818–822.
21. Takimoto JK, Xiang Z, Kang JY, Wang L. Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *ChemBioChem.* 2010;11:2268–2272.
22. Bi X, Pasunooti KK, Lescar J, Liu CF. Thiazolidine-masked alpha-oxo aldehyde functionality for peptide and protein modification. *Bioconjug Chem.* 2017;28:325–329.
23. Bi X, Pasunooti KK, Tareq AH, Takyi-Williams J, Liu CF. Genetic incorporation of 1,2-aminothiol functionality for site-specific protein modification via thiazolidine formation. *Org Biomol Chem.* 2016;14:5282–5285.