Published in final edited form as:

Biochemistry. 2018 January 30; 57(4): 441–445. doi:10.1021/acs.biochem.7b00952.

Expanding the scope of single and dual noncanonical amino acid mutagenesis in mammalian cells using orthogonal polyspecific leucyl-tRNA synthetases

Yunan Zheng[†], Raja Mukherjee[†], Melissa A. Chin, Peter Igo, Martin J. Gilgenast, and Abhishek Chatterjee^{*}

Department of Chemistry, Boston College, 2609 Beacon Street, Chestnut Hill, MA 02467, USA.

Abstract

Engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that enable site-specific incorporation of noncanonical amino acids (ncAAs) into proteins in living cells have emerged as powerful tools in chemical biology. The *E. coli* derived leucyl-tRNA synthetase (EcLeuRS)/tRNA pair is a promising candidate for ncAA mutagenesis in mammalian cells, but it has been engineered to charge only a limited set of ncAAs so far. Here we show that two highly polyspecific EcLeuRS mutants can efficiently charge a large array of useful ncAAs into proteins expressed in mammalian cells, while discriminating against the 20 canonical amino acids. When combined with an opal-suppressing pyrrolysyl pair, these EcLeuRS variants further enabled site-specific incorporation of different combinations of two distinct ncAAs into proteins expressed in mammalian cells.

Noncanonical amino acid (ncAA) mutagenesis provides enabling new ways to probe and manipulate protein function in living mammalian cells.^{1–8} This technology uses engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that do not cross-react with their endogenous counterparts (i.e., orthogonal) to incorporate the ncAA of interest in response to a nonsense codon. Four such orthogonal aaRS/tRNA pairs have been developed to date for ncAA mutagenesis in mammalian cells: tyrosyl,^{6, 9} leucyl (EcLeu)^{7, 10} and tryptophanyl¹¹ pairs from *E. coli*, and the unique pyrrolysyl pair (Pyl)^{1–4, 12, 13} from archaea. However, so far, the overwhelming majority of ncAAs have been genetically encoded in mammalian cells using the Pyl pair.^{1–3, 11, 12} The natural substrate-promiscuity of the pyrrolysyl-tRNA synthetase (PylRS), and the ability to further engineer its substrate-specificity using a facile *E. coli* based selection scheme, have contributed to its remarkable success. However, the overwhelming dependence on a single platform limits the structural diversity of the genetically encoded ncAAs in mammalian cells. Additionally, it has been recently shown that two different aaRS/tRNA pairs can be used together to site-specifically incorporate two

^{*}Corresponding Author abhishek.chatterjee@bc.edu.

[†]These authors contributed equally.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Experimental methods, supplementary figures, and supplementary tables (PDF)

Notes

The authors declare no competing financial interests.

distinct ncAAs into proteins expressed in mammalian cells.^{14, 15} However, the scope of this promising technology is significantly restricted by the limited selection of ncAAs that can be currently incorporated using the non-Pyl pairs.^{2, 3, 11, 15}

The *E. coli* derived leucyl-tRNA synthetase (EcLeuRS)/tRNA^{EcLeu} pair has been previously used to incorporate a small number of ncAAs into proteins expressed in mammalian cells.^{7, 15–17} Despite the limited number, the significant structural diversity^{2, 4, 10, 18–22} of the ncAAs that the EcLeuRS/tRNA^{EcLeu} pair has been engineered to charge suggests that it is a promising candidate for further expanding the mammalian genetic code. We were particularly intrigued by a class of EcLeuRS mutants which were previously found to charge a series of long and hydrophobic ncAAs, such as homologs of alanine, cysteine and methionine.^{18, 19} Such “polyspecific” aaRS variants, which charge a series of structurally related ncAAs but none of the 20 canonical amino acids, are useful as they enable rapid expansion of the catalog of genetically encoded ncAAs without labor-intensive selection of unique aaRS mutants for each individual member.^{2, 23–25} We wondered if these EcLeuRS variants can charge other ncAAs of similar structure, but with additional enabling chemical functionalities (e.g., bioconjugation handles). The access to such polyspecific EcLeuRS mutants for genetic code expansion in mammalian cells will be highly valuable.

To explore this possibility, we first screened several previously reported^{18, 19} EcLeuRS mutants for charging 2-aminocaprylic acid (**1**, Figure 1; known substrate of these mutants) into proteins in mammalian cells. These mutants were cloned into the pAcBac3-EcLeu_{TAG}-EGFP* plasmid that also encodes expression cassettes for its cognate tRNA_{CUA}^{EcLeu} as well as the EGFP-39-TAG reporter. We introduced mutation in the editing domain of these EcLeuRS variants (T252) to improve their ncAA-charging efficiency. The resulting plasmids were transfected into HEK293T cells and the expression of full-length EGFP-39-TAG reporter was monitored in the presence and absence of **1** by measuring its characteristic fluorescence in cell-free extract. Two mutants, named PLRS1 and PLRS2 (Figure 2A), enabled efficient expression of EGFP selectively in the presence of **1** (Figure 2B). Isolation of the full-length reporter protein by immobilized metal ion chromatography (IMAC) using a C-terminal poly-histidine tag followed by its mass-spectrometry (MS) analysis confirmed successful incorporation of **1** (Figure S1).

Next, to assess the tolerance of these EcLeuRS variants for varying side-chain sizes, we synthesized a series of ncAAs by modifying cysteine with linear alkyl halides of increasing length (**2a–2f**, Figure 1). We evaluated how efficiently the two polyspecific EcLeuRS variants charge them using the aforementioned EGFP-39-TAG expression assay (Figure 2C). Both variants were found to charge nearly all of these ncAAs, although the efficiency dropped for the shortest and the longest members of the series. This experiment highlights the remarkable plasticity of these active sites, and provides an estimate for the optimal side-chain length for its substrate ncAAs. It also indicates the basis for their discrimination against the canonical hydrophobic amino acids, whose side chains are significantly smaller.

To take further advantage of their substrate polyspecificity, we evaluated if these EcLeuRS variants can charge additional linear hydrophobic ncAAs of similar side-chain length with a variety of other non-natural chemical functionalities. Guided by the results described in

Figure 2C, a series of ncAAs were synthesized by S-alkylating cysteine with various alkyl halides containing different functional groups, such as alkyne, azide, ketone, alkene, halogen, etc (**3a–c**, **4a–c**, **5**, **6**, **7**; Figure 1). All of these ncAAs were found to be substrates for PLRS1 and PLRS2 (Figure 2D; Figure S2). As both mutants showed comparable efficiencies for charging these ncAAs, here we present the data associated with only PLRS1. Incorporation of each of the ncAAs was confirmed by purifying the corresponding full-length reporter protein by IMAC, followed by MS analysis (Figure S1, Figure S2, Table S1). The efficiency of ncAA incorporation was comparable with the well-established pyrrolysyl pair charging one of its most efficient substrates (*e*-Boc-lysine). The ability to site-specifically incorporate ncAAs harboring ketone, azide, or alkyne will allow chemoselective modification of the resulting protein via well-established bio-orthogonal conjugation reactions^{26–28} (e.g., oxime condensation,²⁹ and Cu(I)-dependent or strain-promoted azide-alkyne cycloaddition reactions^{30, 31}). It is highly likely that additional structurally similar ncAAs with other non-natural functional groups will also be substrates for PLRS1, making it a powerful platform for introducing new chemistries into the mammalian genetic code.

While the site-specific incorporation of a single ncAA into proteins in mammalian cells is now well-established, the technology for incorporating two distinct ncAAs still remains underdeveloped.^{14, 15} We have recently shown that the EcLeuRS/tRNA_{CUA}^{EcLeu} pair can be used together with the TGA-suppressing PylRS/tRNA_{UCA}^{Pyl} pair to site-specifically incorporate two distinct ncAAs into proteins expressed in mammalian cells with high fidelity.¹⁵ However, so far, the scope of this platform has been severely limited owing to the lack of useful ncAAs genetically encoded using the leucyl pair. An enabling application of the dual-ncAA mutagenesis technology involves site-specific incorporation of two different bioconjugation handles, which can be independently functionalized through mutually compatible chemistries, allowing precise attachment of two distinct entities (e.g., two optical probes) onto one protein.^{15, 32–37} With the expanded set of ncAAs that can be charged using the PLRS1/tRNA_{CUA}^{EcLeu} pair, it should be now possible to achieve this using the EcLeu/Pyl dual nonsense-suppression system. We were particularly attracted to the combination of two ncAAs harboring an azide (e.g., C5Az **4c** charged by PLRS1) and a strained alkene (e.g., CpK **9**,^{35, 38} or NbK **10**,³⁹ charged by PylRS), which can be functionalized under ambient conditions using strain-promoted azide–alkyne cycloaddition and tetrazine–strained alkene click chemistries, respectively, that are mutually compatible.¹⁵

When using two different aaRS/tRNA pairs for incorporating two distinct ncAAs, it is important to ensure that these do not cross-react with each other at the levels of aaRS-ncAA, aaRS-tRNA, or codon-anticodon recognition.¹⁵ We have previously ruled out cross-reactivity between the EcLeuRS/tRNA_{CUA}^{EcLeu} and PylRS/tRNA_{UCA}^{Pyl} pairs at the levels of aaRS-tRNA and codon-anticodon recognition.¹⁵ However, given the structural similarity between the substrates charged by PLRS1 and PylRS, we needed to confirm that the intended substrate for one aaRS is not charged by the other (Figure 3A). Using the aforementioned EGFP-39-TAG expression assay, the *M. barkeri* derived PylRS (MbPylRS) and PLRS1 were assessed for charging C5Az (**4c**), AzK (**8**), CpK (**9**) and NbK (**10**; Figure 1). We found that MbPylRS does not charge C5Az (intended PLRS1 substrate), but PLRS1 has the ability to charge some of the MbPylRS substrates with varying efficiencies: AzK was

an efficient substrate, CpK was a weak substrate, but the larger NbK was a poor substrate (Figure 3B, Figure S3). Thus, it should be possible to selectively charge C5Az and NbK by PLRS1 and MbPylRS, respectively, allowing site-specific incorporation of an azide and a norbornene group into proteins. The suitability of using CpK as the MbPylRS substrate in an analogous dual-ncAA incorporation experiment was less clear, given it is also charged weakly by PLRS1 (Figure 3B, Figure S3). However, since C5Az is a significantly better substrate for PLRS1 than CpK, it is possible that in the presence of both these ncAAs, the former will be charged with acceptable selectivity. Indeed, MS analysis of the EGFP-39-TAG reporter expressed from the aforementioned pAcBac3-EcLeu_{TAG}-EGFP* plasmid in the presence of both C5Az and CpK revealed selective incorporation of the former (Figure S4). Consequently, MbPylRS and PLRS1 should selectively charge C5Az and CpK, respectively, when both amino acids are supplemented in the culture medium, enabling site-specific incorporation of an azide and a cyclopropene into target proteins expressed in mammalian cells.

To demonstrate simultaneous incorporation of these ncAA combinations using EcLeu+Pyl pairs (Figure 4A), we used our recently developed two-plasmid system (Figure S5)^{15, 36}: i) pAcBac3-EcLeu_{TAG}-EGFP** encodes PLRS1, multiple copies of tRNA_{CUA}^{EcLeu}, and an EGFP reporter with a TAG and TGA codon at positions 39 and 151, respectively; ii) pAcBac1-Pyl_{TGA} encodes the MbPylRS and multiple copies of tRNA_{UCA}^{Pyl}. These two plasmids were co-transfected into HEK293T cells and the expression of the full-length EGFP reporter was monitored in the presence or absence of C5Az+NbK or C5Az+CpK in the medium using fluorescence microscopy, as well as by measuring the EGFP fluorescence in cell-free extract. Selective appearance of EGFP fluorescence in the presence of ncAAs indicated successful dual nonsense suppression (Figure 4B, Figure S6). MS analysis of the full-length reporter protein, purified in each case using a C-terminal poly-histidine tag, further confirmed the incorporation of the desired ncAAs (Figure S7). Even though there was significant full-length reporter expression in the presence of CpK alone (Figure 4B, Figure S6), consistent with the fact that it is charged weakly by PLRS1, the MS analysis of the reporter protein expressed in the presence of both CpK and C5Az confirmed incorporation of these ncAAs at intended sites. Incorporation efficiency of C5Az+CpK (2 µg/10 cm dish; 1.5% of wild-type protein) was found to be higher than C5Az+NbK (0.6 µg/10 cm dish; 0.5% of wild-type EGFP reporter), likely because CpK is a better substrate for MbPylRS than NbK (Figure 3B).

Site-specific incorporation of an azide and a strained alkene (cyclopropene/norbornene) into a protein should allow its chemoselective modification with two distinct entities through mutually compatible bio-orthogonal chemistries. Indeed, incubating the EGFP-39-C5Az-151-NbK or EGFP-39-C5Az-151-CpK with DBCO-TAMRA or tetrazine-fluorescein (Figure S8) led to covalent attachment of the respective fluorophores onto the protein, as revealed by SDS-PAGE followed by fluorescence imaging (Figure 4C), as well as MS analysis. The NbK-tetrazine conjugation was found to be significantly slower and did not reach completion even after prolonged incubation; in contrast, the CpK-tetrazine conjugation enabled complete protein modification within 30 min (Figure 4C–D). We further demonstrated dual-labeling of EGFP-39-C5Az-151-CpK with two distinct fluorophores simply by sequentially incubating it with DBCO-TAMRA and tetrazine-fluorescein under

ambient conditions (Figure 4D). The same expression system developed here should further enable site-specific incorporation many additional combination of ncAAs, given the large repertoire of substrates that can be charged by each of the two aaRSs used here.

In summary, we have significantly expanded the utility of the EcLeu pair for ncAA mutagenesis in mammalian cells by taking advantage of EcLeuRS mutants with remarkable substrate polyspecificity. Simply by maintaining a linear and hydrophobic side-chain architecture of optimal size, a variety of chemical functionalities were effortlessly genetically encoded in mammalian cells using PLRS1. In addition to those reported in this manuscript, it should be possible to charge additional ncAAs with other chemical functionalities using these mutants. The expanded set of ncAAs that can now be charged using the EcLeu pair also significantly broadens the scope of dual ncAA incorporation using the EcLeu and the Pyl pair. This was highlighted by demonstrating concurrent site-specific incorporation of an azide and a cyclopropene functionality into a reporter protein, which enabled precise attachment of two distinct fluorophores using mutually compatible conjugation chemistries. Taken together, these traits establish the polyspecific EcLeuRS variants described here as powerful new tools for mammalian genetic code expansion.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Boston College Chemistry Department and NIGMS for financial support.

Funding Sources

This work was supported by NIH (R01GM124319 to AC)

ABBREVIATIONS

aaRS	aminoacyl-tRNA synthetase
ncAA	noncanonical amino acid
LeuRS	leucyl-tRNA synthetase
Pyl	pyrrolysyl
IMAC	immobilized metal ion chromatography
MS	mass spectrometry

REFERENCES

1. Chin JW (2014) Annual review of biochemistry 83, 379–408.
2. Dumas A, Lercher L, Spicer CD, and Davis BG (2015) Chemical Science 6, 50–69. [PubMed: 28553457]
3. Italia JS, Zheng Y, Kelemen RE, Erickson SB, Addy PS, and Chatterjee A (2017) Biochemical Society transactions 45, 555–562. [PubMed: 28408495]

4. Liu CC, and Schultz PG (2010) Annual review of biochemistry 79, 413–444.
5. Liu W, Brock A, Chen S, Chen S, and Schultz PG (2007) Nature methods 4, 239–244. [PubMed: 17322890]
6. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, Kobayashi T, Kitabatake M, Takio K, Saito K, Shirouzu M, Hirao I, and Yokoyama S (2002) Nucleic acids research 30, 4692–4699. [PubMed: 12409460]
7. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, Lee KF, Slesinger PA, and Wang L (2007) Nature neuroscience 10, 1063–1072. [PubMed: 17603477]
8. Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, and Yokoyama S (2005) Nature methods 2, 201–206. [PubMed: 15782189]
9. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, and Schultz PG (2003) Science 301, 964–967. [PubMed: 12920298]
10. Wu N, Deiters A, Cropp TA, King D, and Schultz PG (2004) Journal of the American Chemical Society 126, 14306–14307. [PubMed: 15521721]
11. Italia JS, Addy PS, Wrobel CJ, Crawford LA, Lajoie MJ, Zheng Y, and Chatterjee A (2017) Nature chemical biology 13, 446–450. [PubMed: 28192410]
12. Wan W, Tharp JM, and Liu WR (2014) Biochimica et biophysica acta 1844, 1059–1070. [PubMed: 24631543]
13. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, and Yokoyama S (2008) Biochemical and biophysical research communications 371, 818–822. [PubMed: 18471995]
14. Xiao H, Chatterjee A, Choi SH, Bajjuri KM, Sinha SC, and Schultz PG (2013) Angewandte Chemie (International ed. in English) 52, 14080–14083. [PubMed: 24353230]
15. Zheng Y, Addy PS, Mukherjee R, and Chatterjee A (2017) Chemical Science DOI: 10.1039/c7sc02560b
16. Chatterjee A, Guo J, Lee HS, and Schultz PG (2013) Journal of the American Chemical Society 135, 12540–12543. [PubMed: 23924161]
17. Mitchell AL, Addy PS, Chin MA, and Chatterjee A (2017) Chembiochem 18, 511–514. [PubMed: 28093840]
18. Ai HW, Shen W, Brustad E, and Schultz PG (2010) Angewandte Chemie (International ed. in English) 49, 935–937. [PubMed: 20025009]
19. Brustad E, Bushey ML, Brock A, Chittuluru J, and Schultz PG (2008) Bioorganic & medicinal chemistry letters 18, 6004–6006. [PubMed: 18845434]
20. Lee HS, Guo J, Lemke EA, Dimla RD, and Schultz PG (2009) Journal of the American Chemical Society 131, 12921–12923. [PubMed: 19702307]
21. Lemke EA, Summerer D, Geierstanger BH, Brittain SM, and Schultz PG (2007) Nature chemical biology 3, 769–772. [PubMed: 17965709]
22. Summerer D, Chen S, Wu N, Deiters A, Chin JW, and Schultz PG (2006) Proceedings of the National Academy of Sciences of the United States of America 103, 9785–9789. [PubMed: 16785423]
23. Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, and Schultz PG (2011) Biochemistry 50, 1894–1900. [PubMed: 21280675]
24. Cooley RB, Karplus PA, and Mehl RA (2014) Chembiochem 15, 1810–1819. [PubMed: 25044993]
25. Chatterjee A, Xiao H, Bollong M, Ai HW, and Schultz PG (2013) Proceedings of the National Academy of Sciences of the United States of America 110, 11803–11808. [PubMed: 23818609]
26. Hao Z, Hong S, Chen X, and Chen PR (2011) Accounts of chemical research 44, 742–751. [PubMed: 21634380]
27. Lang K, and Chin JW (2014) Chemical reviews 114, 4764–4806. [PubMed: 24655057]
28. Lang K, and Chin JW (2014) ACS chemical biology 9, 16–20. [PubMed: 24432752]
29. Dirksen A, and Dawson PE (2008) Bioconjugate chemistry 19, 2543–2548. [PubMed: 19053314]
30. Meldal M, and Tornøe CW (2008) Chemical reviews 108, 2952–3015. [PubMed: 18698735]

31. Sletten EM, and Bertozzi CR (2011) *Accounts of chemical research* 44, 666–676. [PubMed: 21838330]
32. Chatterjee A, Sun SB, Furman JL, Xiao H, and Schultz PG (2013) *Biochemistry* 52, 1828–1837. [PubMed: 23379331]
33. Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, Russell DH, and Liu WR (2010) *Angewandte Chemie (International ed. in English)* 49, 3211–3214. [PubMed: 20340150]
34. Wang K, Sachdeva A, Cox DJ, Wilf NM, Lang K, Wallace S, Mehl RA, and Chin JW (2014) *Nature chemistry* 6, 393–403.
35. Sachdeva A, Wang K, Elliott T, and Chin JW (2014) *Journal of the American Chemical Society* 136, 7785–7788. [PubMed: 24857040]
36. Zheng Y, Lewis TL, Jr., Igo P, Polleux F, and Chatterjee A (2017) *ACS synthetic biology* 6, 13–18. [PubMed: 27482719]
37. Patterson DM, and Prescher JA (2015) *Current opinion in chemical biology* 28, 141–149. [PubMed: 26276062]
38. Kamber DN, Nazarova LA, Liang Y, Lopez SA, Patterson DM, Shih HW, Houk KN, and Prescher JA (2013) *Journal of the American Chemical Society* 135, 13680–13683. [PubMed: 24000889]
39. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, and Chin JW (2012) *Nature chemistry* 4, 298–304.

Figure 1:
Structures of ncAAs used in this study.

Figure 2. Identification of polyspecific EcLeuRS mutants for ncAA incorporation in mammalian cells. A) Structure of the EcLeuRS (wild type) active site bound to its substrate leucine (purple). Active site residues that are mutated in PLRS1 and PLRS2 (sequence shown below) are highlighted. B) PLRS1 and PLRS2 facilitates efficient incorporation of ncAA **1** into EGFP-39-TAG reporter expressed in mammalian cells, observed by fluorescence microscopy. C) Efficiency of incorporating ncAAs **2a–f** into the EGFP-39-TAG reporter using PLRS1 and PLRS2, measured as EGFP fluorescence in cell-free extract, shows the tolerance of these active sites for different side-chain lengths. D) PLRS1 can be used charge linear hydrophobic amino acids with a variety of functional groups (measured using the aforementioned EGFP-39-TAG expression assay).

Figure 3:

A) For site-specific incorporation of two distinct ncAAs by the Pyl and EcLeu pairs, it is important to rule out cross-reactivity at the level of aaRS-ncAA recognition. B) Evaluation of the relative charging efficiency of indicated ncAAs by PLRS1 and MbPylRS, using the aforementioned EGFP-39-TAG expression assay, reveals that C5Az is not a substrate for MbPylRS, whereas AzK and CpK are charged by PLRS1 with varying efficiencies.

Figure 4:

Site-specific incorporation of two different, mutually compatible bioconjugation handles into a protein expressed in mammalian cells using Pyl and EcLeu pairs. A) Scheme of the dual nonsense suppression. B) Expression of the full-length EGFP-39-TAG-151-TGA reporter (measured as fluorescence in cell-free extract), upon transfecting HEK293T cells with pAcBac3-EcLeu_{TGA}-EGFP** and pAcBac1-Pyl_{TAG}, in the presence of the indicated ncAAs. C) Treatment of EGFP-39C5Az-151CpK or EGFP-39C5Az-151NbK with DBCO-TAMRA or tetrazine-fluorescein results in expected fluorescence labeling, as revealed by fluorescence imaging of these samples following SDS-PAGE. Wild-type EGFP does not get labeled under identical conditions. Labeling efficiency of NbK is significantly lower than CpK. D) Successful single or dual labeling of EGFP-39C5Az-151CpK was further confirmed by ESI-MS analysis.