Published in final edited form as:

ACS Synth Biol. 2019 May 17; 8(5): 1168–1174. doi:10.1021/acssynbio.9b00051.

Genetic Incorporation of Noncanonical Amino Acids Using Two Mutually Orthogonal Quadruplet Codons

Erme Daniel Hankore[†], Linyi Zhang[†], Yan Chen[†], Kun Liu[†], Wei Niu[‡], and Jiantao Guo^{*†}

[†]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, United States

[‡]Department of Chemical & Biomolecular Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, United States

Abstract

Genetic incorporation of noncanonical amino acids has emerged as a powerful tool for the study of protein structure and function. While the three triplet nonsense codons have been widely explored, quadruplet codons have attracted attention for the potential of creating additional blank codons for noncanonical amino acid mutagenesis. Here we demonstrated for the first time that two orthogonal quadruplet codons could be used to simultaneously encode two different noncanonical amino acids within a single protein in bacterial cells. To achieve this, we fine-tuned the interaction between aminoacyl-tRNA synthetase and tRNA, which afforded up to 21-fold improvement in quadruplet codon decoding efficiency. This work represents a significant step toward the use of multiple quadruplet codons for noncanonical amino acid mutagenesis. Simultaneous incorporation of two or more noncanonical amino acids is of significant importance for biological applications that can benefit from multiple unique functional groups, such as fluorescence resonance energy transfer and nuclear magnetic resonance studies, and ultimately for the synthesis of completely unnatural biopolymers as new biomaterials.

Graphical Abstract

*Corresponding Author jguo4@unl.edu.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.9b00051.

Additional figures, data, plasmid maps, and sequences (PDF)

The authors declare no competing financial interest.

Keywords

quadruplet codon; four-base codon; unnatural amino acids; orthogonal codon; genetic code

Proteins are responsible for all aspects of a cell's performance, ranging from DNA replication to transport of molecules and preservation of homeostasis. They are normally composed of only 20 canonical amino acids the sequence of which defines protein structure and function. While proteins generated from these 20 amino acids perform a wide range of biochemical functions, post-translational modifications are widely used in nature to further increase the functional diversity of proteins. Therefore, it is plausible that the proteome could be expanded to include members with improved or novel activities by increasing the repertoire of amino acid building blocks. In this regard, noncanonical amino acids (ncAAs) have been incorporated into proteins using various techniques.^{1–6} Among them, a widely employed approach for the site-specific incorporation of ncAAs into proteins relies on nonsense codon suppression using engineered tRNA and aminoacyl-tRNA synthetase (aaRS) pairs.^{4–6} Since this method was first demonstrated in 2001,⁷ it has been used to incorporate over 150 ncAAs into proteins in bacterial, yeast, and mammalian cells.⁸

Even though nonsense codon suppression provides an efficient means of ncAA incorporation, the number of available triplet nonsense codons (three in total including amber UAG, ochre UAA, and opal UGA codons) limits the utility of this technique by restricting the number of unique ncAAs that can be simultaneously incorporated into proteins. Additional blank codons (codons that do not encode any canonical amino acids) are required for extensive cellular genetic code expansion. Quadruplet codons provide an intriguing alternative in this regard. While the decoding efficiency is usually lower than that of the nonsense codon, quadruplet codons indeed deliver additional choices for ncAA mutagenesis. In principle, a quadruplet codon table provides a maximum of 256 blank codons, which enable the genetically programmed functionalization of proteins with more ncAAs in living cells. The incorporation of ncAAs in response to quadruplet codons has been achieved in cell-free systems using *in vitro* protein translation methods,^{3,9–12} in *Xenopus* oocytes by microinjecting chemically aminoacylated tRNAs,¹³ and in live cells with engineered tRNA-aaRS pairs.^{13–20} However, many of these approaches are hampered by low ncAA incorporation efficiency. We^{21–23} and others²⁴ recently reported a new approach to enhance the efficiency of quadruplet decoding, which is based on the engineering of tRNA. In the present work, we demonstrate the first example of simultaneous incorporation of two different ncAAs in a single *E. coli* strain using two mutually orthogonal quadruplet codons. Previous examples relied on either two triplet nonsense codons^{25–33} or a combination of an amber nonsense codon and a quadruplet codon.^{16,17,24} Overall, this work will significantly augment recent efforts in cellular genetic code expansion.

RESULTS AND DISCUSSION

Research Design.

To achieve simultaneous and site-specific incorporation of two ncAAs in *E. coli*, we intended to use a *Mj*TyrRS-tRNA^{Tyr} pair derived from *Methanocaldococcus jannaschii* and a

MbPylRS-tRNA^{Pyl} pair from *Methanosarcina barkeri*. These two pairs are not only orthogonal to endogenous aaRSs and tRNAs of the host cell, but also mutually orthogonal to each other.^{17,24,25} In our previous work, we have engineered a *BocLysRS*-tRNA^{UCCU} pair that could efficiently decode an AGGA codon.²¹ *BocLysRS* is a *MbPylRS* derivative that can charge tRNA^{UCCU} (a tRNA^{Pyl} derivative) with *Nε*-(*tert*-butoxy-carbonyl)-L-lysine (*BocLys*).³⁴ An *AcPheRS*-tRNA^{UCUA} pair was also reported to decode a UAGA codon.²⁰ *AcPheRS* is a *MjTyrRS* derivative that can charge tRNA^{UCUA} (a tRNA^{Tyr} derivative) with *p*-acetylphenyl-alanine (*pAcPhe*).³⁵

In this work, we demonstrated that the *BocLysRS*-tRNA^{UCCU} pair and the *AcPheRS*-tRNA^{UCUA} pair do not have any cross reactivity. However, a simultaneous decoding of AGGA and TAGA codons was hindered by the low efficiency of the *AcPheRS*-tRNA^{UCUA} pair. This is mainly due to the fact that the reported tRNA^{UCUA} was generated by simply replacing the CUA anticodon in the amber suppressing tRNA^{CUA} (a derivative of tRNA^{Tyr}; Figure 1A) into the UCUA anticodon. It was reported in a structural study³⁶ that the anticodon region of tRNA was engaged in a specific interaction with *MjTyrRS* (Figure 1B). The switch from a triplet to a quadruplet anticodon would impair the recognition of tRNA by its cognate *MjTyrRS*, which can likely lead to a lower aminoacylation efficiency. Therefore, we sought to fine-tune the interaction between *MjTyrRS*-derived *AcPheRS* and tRNA^{UCUA} by conducting directed evolution of *AcPheRS*. Our previous research showed successful application of this strategy to efficiently improving amber suppression efficiency.³⁷

Directed Evolution of *AcPheRS*.

To generate *AcPheRS* mutants with higher efficiency in charging tRNA^{UCUA} with *AcPhe*, we sought to engineer the interface between *AcPheRS* and tRNA^{UCUA}. As shown in Figure 1B, the anticodon of wild-type *M. jannaschii* tyrosyl tRNA^{Tyr} (with an anticodon that decodes UAC codon with Tyr) is engaged in a specific interaction with the C-terminal domain of *MjTyrRS*. Since the main differences between the anticodon region of tRNA^{UCUA} and that of tRNA^{Tyr} are the replacement of G34 with C34 and the insertion of an extra nucleotide, U33.5 (Figure 1A; Nucleotides of tRNA were numbered by following conventional rule.³⁸), we decided to modify the region of *MjTyrRS* that recognizes the anticodon region, especially the G34 and U35 residues, in order to accommodate alterations in tRNA^{UCUA}. As shown in the crystal structure³⁶ (Figure 1B), the G34 of tRNA^{Tyr} is sandwiched between Phe261 and His283 residues of *MjTyrRS*. G34 also interacts with Asp286 through hydrogen bonding. Furthermore, G34 and U35 of tRNA^{Tyr} are located close to Cys231 and Met285 of *MjTyrRS* as well. Therefore, these amino acid residues (Phe261, His283, Asp286, Met285, and Cys231) of *MjTyrRS* were targeted in our directed evolution efforts to improve the aminoacylation activity and the UAGA codon decoding efficiency.

Two *AcPheRS* libraries (Library-1: Phe261, His283, Met285, and Asp286. Library-2: Cys231, Phe261, His283, and Asp286) were created by randomizing the indicated amino acid residues through an overlapping polymerase chain reaction (PCR) with primers containing NNK (N = A, C, T, or G, K = T or G) codons at the randomization sites. The theoretical diversity of each library was 1.05×10^6 (>99% coverage). DNA sequencing

results confirmed randomization and revealed excellent coverage of the library. The resulting libraries were subjected to two rounds of positive selection with one round of intermediate negative selection by following reported protocols.³⁷ A chloramphenicol acetyltransferase gene containing a TAGA codon at a permissive site (Gln98) was used as the positive selection marker. Functional AcPheRS mutants would lead to cell growth in the presence of chloramphenicol. A barnase gene containing two TAGA codons at permissive sites (Gln2TAGA and Asp44TAGA) was used for the negative selection. AcPheRS mutants that can recognize any of the 20 canonical amino acids would result in cell death due to the production of toxic barnase protein. Both positive and negative selections were carried out in C321. A.exp cells that do not express release factor 1 (RF1; recognizes the UAG nonsense codon as the termination signal of protein translation) and do not contain endogenous amber UAG nonsense codon (replaced with ochre UAA nonsense codon).³⁹ This genomically recoded *E. coli* strain represents an excellent host for the directed evolution of AcPheRS for UAGA codon decoding by eliminating any in-frame UAGA codons in the host genome. Surviving library members were replicated on plates with or without chloramphenicol (50–100mg L⁻¹) to ensure selectivity toward AcPhe. Colonies with the fastest rate of growth in the presence of chloramphenicol were further screened and characterized.

Characterization of AcPheRS Mutants.

To examine the efficiency and fidelity of UAGA codon decoding by tRNA^{UCUA} in the presence of evolved AcPheRS mutants, a fluorescence-based screening was conducted. A reporter system was established through the introduction of a UAGA codon into GFP_{UV} (a variant of green fluorescent protein with an excitation maximum of 395 nm) by replacing the Asn149 codon to yield GFP-Asn149UAGA (encoded in plasmid pLei-GFP1Q-tRNA^{UCUA}). Higher expression of GFP-Asn149UAGA in the presence of AcPhe would indicate a more efficient decoding of UAGA codon by the AcPheRS mutant. Low or no expression of GFP-Asn149UAGA in the absence of AcPhe would suggest good fidelity of the mutant. In general, it was observed that all the selected hits gave fluorescence values higher than that of the wild-type AcPheRS with minimal background. Among the characterized mutants, AcPheRS-2 and AcPheRS-4 showed significant overall increase in fluorescence (Figure 2). The best one (AcPheRS-2) displayed a 21-fold increase in fluorescence relative to that of the parental wild-type AcPheRS (AcPheRS-wt). As another comparison, the expression level of GFP-Asn149UAGA was about 55% of that of the wild-type GFP (GFP-wt) when AcPheRS-2 was used (Figure 2). The background fluorescence signals in the absence of AcPhe are comparable between the evolved AcPheRS mutants and the parental AcPheRS-wt (Figure 2). A “Fidelity Index”, which was based on the ratio of fluorescence intensities of cells that were cultivated in the presence and in the absence of AcPhe, was established to achieve quantitative evaluation of the fidelity of AcPhe incorporation in response to UAGA codon. As shown in Figure S1, all evolved AcPheRS variants displayed better Fidelity Index than that of the parental AcPheRS-wt. Among them, AcPheRS-8 had the best Fidelity Index, which represented a more than 8-fold improvement over AcPheRS-wt. Interestingly, while the two hits (AcPheRS-2 and AcPheRS-4) from library-1 showed the highest activity, hits from library-2 generally possessed a better Fidelity Index.

All 10 evolved AcPheRS variants were sequenced. No obvious consensus of mutations was observed. This indicates that a large number of solutions exist to accommodate an expanded anticodon loop of tRNA_{UCUA}. As a side note, Cys190 was not a randomized residue by design, but Cys190Phe mutation was commonly observed among hits from both AcPheRS library-1 and library-2. The mutation at this site was likely to be introduced during the construction of libraries.

Encoding ncAAs with Two Orthogonal Quadruplet Codons.

Simultaneous decoding of AGGA and UAGA codons with two different ncAAs was examined using a GFP mutant. The AGGA codon was placed at the Tyr39 position and was decoded by the previously established BocLysRS-tRNA_{UCCU} pair with BocLys.²¹ The UAGA codon was placed at the Asn149 position of GFP and was decoded by the (AcPheRS-3)-tRNA_{UCUA} pair with AcPhe. The decoding efficiency of the (AcPheRS-3)-tRNA_{UCUA} pair matched that of the BocLysRS-tRNA_{UCCU} pair very well. A matching decoding efficiency of the two pairs can likely benefit the double decoding experiments. In addition, AcPheRS-3 was chosen for both its high efficiency in decoding and its over 3-fold better Fidelity Index than the parental enzyme.

Two plasmids were constructed for the double decoding experiments (Figure 3B). A GFP mutant gene, GFP-Tyr39AGGA-Asn149TAGA, was cloned on the plasmid pLei-GFP2Q-AcPheRS-tRNA_{UCUA} (*Cm^R*) together with the (AcPheRS-3)-tRNA_{UCUA} pair. The other plasmid, pBK-BocKRS-tRNA_{UCCU} (*Kan^R*), harbored the BocLysRS-tRNA_{UCCU} pair. We conducted double decoding experiments at room temperature by inducing protein expression at OD₆₀₀ of 1. As shown in Figure 3C, significant GFP fluorescence was only observed when both AcPhe (1 mM) and BocLys (5 mM) were added to the cell culture media. The mutant GFP displayed identical fluorescence properties to that of the parental wild-type GFP (Figure S2). As controls, very low fluorescence was detected in the absence of either AcPhe, or BocLys, or both. In addition to the BocLys-AcPhe combination, we also conducted the double decoding experiments with the AlkyneLys-AcPhe combination (Figure S3). AlkyneLys (*N*ε-(pent-4-yn-1-yloxy-carbonyl)-L-lysine; Figure S3A) contains an alkyne functional group and is useful for protein modification through click chemistry. Since the BocLysRS is active on AlkyneLys, the same two plasmids for the BocLys-AcPhe pair incorporation (Figure 3B) were used in this experiment. As shown in Figure S3B, significant GFP fluorescence was only observed when both AcPhe (1 mM) and AlkyneLys (1 mM) were added to the cell culture media. As controls, very low fluorescence was detected in the absence of either AcPhe, or AlkyneLys, or both. These observations indicated that the two quadruplet codons are indeed orthogonal to each other and can be used to simultaneously encode two different ncAAs.

Next, we conducted a larger scale (10 mL cell culture) expression experiment of the GFP-Tyr39AGGA-Asn149TAGA mutant. A yield of 9.6 mg/L of protein was obtained in the presence of both AcPhe and BocLys after a purification using affinity chromatography. As a comparison, the yield of GFP-wt (no quadruplet codon) was 65.3 mg/L by following the same protocol for protein expression and purification. On the other hand, only a trace amount of protein was detected by SDS-PAGE in the absence of either AcPhe, or BocLys, or

both (Figure 3D). This observation was consistent with our fluorescence-based assay (Figure 3C). We further analyzed the purified protein by mass spectrometry after SDS-PAGE separation and trypsin digest. While the expected mass for the peptide fragment containing AcPhe was observed (Figure S4B), a lysine residue was detected at position Tyr39 instead of BocLys (Figure S4A). This is mainly due to the cleavage of carbamate in BocLys under mass spectrometry conditions, which was also observed previously with the electron spray ionization process.^{22,40,41} In addition, we have shown previously that BocLysRS could not charge tRNA with Lys.²² Overall, mass spectrometry analyses (Figure S4) of the purified protein confirmed that AcPhe and BocLys were site specifically incorporated at predefined positions.

In conclusion, we have identified AcPheRS mutants that can be paired with tRNA_{UCUA} to decode UAGA codon with improved efficiency. For the first time, we achieved the simultaneous genetic incorporation of two ncAAs in response to two orthogonal quadruplet codons. As an alternative to nonsense codon suppression strategy, quadruplet codon decoding provides additional blank codons for ncAA mutagenesis, which will likely augment recent efforts in cellular genetic code expansion. An improvement in decoding efficiency would further promote the application of quadruplet codons in ncAA mutagenesis. The simultaneous incorporation of multiple ncAAs with different physical and chemical properties at specific sites within proteins in live cells will be of significant importance for many biological investigations, such as FRET and NMR, and for synthetic biology applications, such as rewiring biological systems and synthesizing new functions.

METHODS

Reagents.

All primers were purchased from Sigma-Aldrich. AcPhe was purchased from Santa Cruz Biotechnology and BocLys was purchased from Bachem. AlkyneLys was synthesized by following a literature report.⁴² KOD hot start DNA polymerase was purchased from EMD Millipore. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs and Thermo Scientific. Standard molecular biology techniques were employed in all cloning experiments. The *E. coli* GeneHogs and C321. A.exp strains were used for cloning and selection, respectively. All solutions were prepared with water that was purified by the Barnstead Nanopure ultrapure water purification system. Antibiotics were added where appropriate to the following final concentrations: ampicillin, 100 mg L⁻¹; kanamycin, 50 mg L⁻¹; tetracycline, 12.5 mg L⁻¹; chloramphenicol, 34–100 mg L⁻¹.

Construction of Plasmids.

Plasmid pRep-TAGA-tRNA_{UCUA} was constructed as follows: The tRNA_{UCUA} was amplified from tRNA^{Tyr} (*M. jannaschii*) by overlapping PCR with primers pRep-FP1, pRep-RP1, pRep-FP2, and pRep-RP2 (Table S1). The resulting PCR product was digested with *Xba*I and ligated into plasmid pRepCM12b-UAGA²² that was pretreated with the same restriction enzyme to afford pRep-TAGA-tRNA_{UCUA}.

Plasmid pNeg-TAGA-tRNA_{UCUA} was constructed as follows: The gene cassette that contains a barnase gene (with two TAGA codons at positions Gln2 and Asp44) and tRNA_{UCUA} was amplified from plasmid pNEG³⁷ by overlapping PCR with primers pNeg-FP1, pNeg-RP1, pNeg-FP2, pRep-RP2, pRep-RP1, and pNeg-RP3 (Table S1). The resulting PCR product was then digested with *EcoR*I and *Nde*I and ligated into plasmid pNEG that was pretreated with the same set of restriction enzymes to afford pNeg-TAGA-tRNA_{UCUA}.

Plasmid pBK-BocKRS-tRNA_{UCCU} was constructed as follows: The vector pBK-BocKRS²¹ was used as template for PCR amplification using primers pBk-FP1 and pBK-RP1 (Table S1). The tRNA_{UCCU} was PCR amplified from pBK-tRNA²¹ using primers pBk-FP3 and pBk-FP4 (Table S1). Sequence ligation independent cloning (SLIC) was used to ligate the above two PCR products to afford plasmid pBK-BocKRS-tRNA_{UCCU}.

Plasmid pLei-GFP2Q-AcPheRS-tRNA_{UCUA} was constructed as follows: The DNA fragment that contains tRNA_{UCUA} was obtained by digestion of plasmid pRep-TAGA-tRNA_{UCUA} with *Spe*I and *Pst*I. It was subsequently ligated into plasmid pGFP_{UV}-UCUA-wt²² that was pretreated with the same set of restriction enzymes to afford pLei-GFP1Q-tRNA_{UCUA}. The quadruplet codon in pLei-GFP1Q-tRNA_{UCUA} was at position Asn149 of GFP. The second quadruplet codon (at position Tyr39) was introduced by overlapping PCR using primers 2Q-FP1, 2Q-RP1, 2Q-FP2, and 2Q-RP2. AcPheRS-3 mutant (Table 1) was PCR amplified and the resulting PCR product was digested with *Sph*I and ligated to pLei-GFP2Q-tRNA_{UCUA} pretreated with the same restriction enzyme to afford pLei-GFP2Q-AcPheRS-tRNA_{UCUA}.

Library construction.

The two AcPheRS mutant libraries were generated by overlapping PCR. Library-1 was constructed by using primers Lib1-F1, Lib1-R1, Lib1-F2, and Lib1-R2 (Table S2). Library-2 was constructed by using primers Lib2-F1, Lib2-R1, Lib2-F2, Lib2-R2, Lib2-F3, and Lib2-R3 (Table S2). The overlapping PCR product was digested with *Bam*HI and *Pst*I and ligated into the pBK²¹ vector that was pretreated with the same set of restriction enzymes.

Positive Selection.

An AcPheRS mutant library of interest was transformed into C321. A.exp cells containing plasmid pRep-TAGA-tRNA_{UCUA}. For the first round positive selection, cells ($\sim 1 \times 10^7$) were grown on agar plates containing kanamycin, tetracycline, chloramphenicol (34 or 50 mg L⁻¹), and 1 mM AcPhe. As a control, the same number of cells was also plated on agar plates without AcPhe. For the second round of positive selection, cells ($\sim 1 \times 10^4$) were plated on 50 or 100 mg L⁻¹ chloramphenicol and 1 mM AcPhe.

Negative Selection.

AcPheRS variants that survived the first round of positive selection were isolated and subjected to the negative selection. C321. A.exp cells ($\sim 1 \times 10^7$) harboring pNeg-TAGA-tRNA_{UCUA} and AcPheRS library were plated on agar plates contacting kanamycin, ampicillin, and 0.2% L-arabinose. Cells were grown at 37 °C for as much time as is necessary for colonies to be visible (24–48 h). Cells from the 0.2% L-arabinose plate were collected and AcPheRS variants were isolated for a second round of positive selection.

Hit Validation.

Colonies displaying rapid growth in the presence of 100 mg L⁻¹ chloramphenicol on the second round positive selection were replicated on plates containing kanamycin, tetracycline, 50 or 100 mg L⁻¹ chloramphenicol, and with or without 1 mM AcPhe. Hits displaying rapid growth on plates with AcPhe but no growth in the absence of AcPhe were chosen for further analysis.

Fluorescence Assay.

C321. A.exp cells containing pLei-GFP1Q-tRNA_{UCUA} and an AcPheRS variant of interest were cultured in LB media with 0.2 mM IPTG, and with or without 1 mM AcPhe. After 15 h, cells were washed once with PBS buffer. Cell density was measured at OD₆₀₀ and fluorescence was quantified using λ_{Ex} = 390 nm and λ_{Em} = 510 nm. All reported fluorescence intensity values were normalized to cell growth. Each reported value was the average of triplet measurements and presented with standard deviation.

Protein Expression and Purification.

C321. A.exp cells containing pBK-BocKRS-tRNA_{UCCU} and pLei-GFP2Q-Ac-PheRS-tRNA_{UCUA} were grown in LB media. At OD₆₀₀ = 1, expression of GFP was induced with the addition of 0.5 mM IPTG and ncAAs at indicated concentrations (no ncAA, 1 mM AcPhe only, 5 mM BocLys only, or both). GFP fluorescence of cells was quantified as described previously. For protein purification, cells were harvested by centrifuging at 5000g for 10 min at 4 °C. The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate, 0.5 mM NaCl, and 20 mM imidazole) and lysed by sonication. The insoluble fraction was removed by centrifugation for 30 min at 21 000g. The soluble fraction was applied to Ni Sepharose resin (GE healthcare). Protein purification followed the manufacture's protocol. The eluted protein was desalted and buffer exchanged into PBS buffer using Econo-Pac 10-DG desalting column (Bio-Rad). Protein concentration was measured using the Bradford assay (Bio-Rad). Protein samples were analyzed by 18% SDS-PAGE and stained with coomassie blue.

Mass Spectrometry.

Purified protein was analyzed using tandem MS/MS following a trypsin digest. Briefly, the protein sample was reduced with DTT (5 mM), alkylated with iodoacetamide (15 mM), and digested overnight with trypsin at 37 °C. The digestion mixture was then separated by nanoLC-MS/MS using a 1 h gradient on an Acquity UPLC M-Class Peptide CSH C18 column (130A, 0.075 mm × 250 mm; Waters Corp, Milford, MA) feeding into a Q-Exactive HF mass spectrometer (Thermo Fisher). Peptide fragments were identified using the Mascot database search.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation (Grant 1553041 to J.G.) and National Institute of Health (Grant 1R01AI111862 to J.G. and W.N.). The authors thank Drs Sophie Alvarez and Mike Naldrett (Proteomics and Metabolomics Facility) for mass spectrometry analysis. C321. A.exp was a gift from Dr. George Church (Addgene # 49018).

REFERENCES

- (1). Dougherty DA (2000) Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol* 4, 645–652. [PubMed: 11102869]
- (2). Budisa N, Minks C, Alefelder S, Wenger W, Dong F, Moroder L, and Huber R (1999) Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. *FASEB J.* 13, 41–51. [PubMed: 9872928]
- (3). Hohsaka T, and Sisido M (2002) Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol* 6, 809–815. [PubMed: 12470735]
- (4). Wang L, Xie J, and Schultz PG (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct* 35, 225–249. [PubMed: 16689635]
- (5). Wu X, and Schultz PG (2009) Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc* 131, 12497–12515. [PubMed: 19689159]
- (6). Chin JW (2012) Reprogramming the genetic code. *Science* 336, 428–429. [PubMed: 22539711]
- (7). Wang L, Brock A, Herberich B, and Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500. [PubMed: 11313494]
- (8). Dumas A, Lercher L, Spicer CD, and Davis BG (2015) Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci* 6, 50–69. [PubMed: 28553457]
- (9). Hohsaka T, Ashizuka Y, Murakami H, and Sisido M (1996) Incorporation of nonnatural amino acids into streptavidin through *In vitro* frame-shift suppression. *J. Am. Chem. Soc* 118, 9778–9779.
- (10). Ohtsuki T, Manabe T, and Sisido M (2005) Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures. *FEBS Lett.* 579, 6769–6774. [PubMed: 16310775]
- (11). Taira H, Fukushima M, Hohsaka T, and Sisido M (2005) Four-base codon-mediated incorporation of nonnatural amino acids into proteins in a eukaryotic cell-free translation system. *J. Biosci. Bioeng* 99, 473–476. [PubMed: 16233819]
- (12). Taki M, Tokuda Y, Ohtsuki T, and Sisido M (2006) Design of carrier tRNAs and selection of four-base codons for efficient incorporation of various nonnatural amino acids into proteins in *Spodoptera frugiperda* 21 (Sf21) insect cell-free translation system. *J. Biosci. Bioeng* 102, 511–517. [PubMed: 17270715]
- (13). Rodriguez EA, Lester HA, and Dougherty DA (2006) *In vivo* incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. U. S. A* 103, 8650–8655. [PubMed: 16728509]
- (14). Magliery TJ, Anderson JC, and Schultz PG (2001) Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of "Shifty" four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol* 307, 755–769. [PubMed: 11273699]
- (15). Anderson JC, Magliery TJ, and Schultz PG (2002) Exploring the limits of codon and anticodon size. *Chem. Biol* 9, 237–244. [PubMed: 11880038]
- (16). Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, and Schultz PG (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A* 101, 7566–7571. [PubMed: 15138302]
- (17). Neumann H, Wang K, Davis L, Garcia-Alai M, and Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444. [PubMed: 20154731]

- (18). Hohsaka T, Ashizuka Y, Taira H, Murakami H, and Sisido M (2001) Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli in vitro* translation system. *Biochemistry* 40, 11060–11064. [PubMed: 11551202]
- (19). Taki M, Matsushita J, and Sisido M (2006) Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* 7, 425–428. [PubMed: 16440374]
- (20). Chatterjee A, Lajoie MJ, Xiao H, Church GM, and Schultz PG (2014) A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem* 15, 1782–1786. [PubMed: 24867343]
- (21). Niu W, Schultz PG, and Guo JT (2013) An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol* 8, 1640–1645. [PubMed: 23662731]
- (22). Wang N, Shang X, Cerny R, Niu W, and Guo J (2016) Systematic evolution and study of UAGN decoding tRNAs in a genomically recoded bacteria. *Sci. Rep* 6, 21898. [PubMed: 26906548]
- (23). Chen Y, Wan Y, Wang N, Yuan Z, Niu W, Li Q, and Guo J (2018) Controlling the replication of a genomically recoded HIV-1 with a functional quadruplet codon in mammalian cells. *ACS Synth. Biol* 7, 1612–1617. [PubMed: 29787233]
- (24). Wang K, Sachdeva A, Cox DJ, Wilf NM, Lang K, Wallace S, Mehl RA, and Chin JW (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labeling and FRET. *Nat. Chem* 6, 393–403. [PubMed: 24755590]
- (25). Wan W, Huang Y, Wang Z, Russell WK, Pai P-J, Russell DH, and Liu WR (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew. Chem., Int. Ed* 49, 3211–3214.
- (26). Xiao H, Chatterjee A, Choi S-H, Bajjuri KM, Sinha SC, and Schultz PG (2013) Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew. Chem., Int. Ed* 52, 14080–14083.
- (27). Zheng Y, Mukherjee R, Chin MA, Igo P, Gilgenast MJ, and Chatterjee A (2018) Expanding the scope of single- and double-noncanonical amino acid mutagenesis in mammalian cells using orthogonal polyspecific leucyl-tRNA synthetases. *Biochemistry* 57, 441–445. [PubMed: 29106828]
- (28). Zheng Y, Gilgenast MJ, Hauc S, and Chatterjee A (2018) Capturing post-translational modification-triggered protein-protein interactions using dual noncanonical amino acid mutagenesis. *ACS Chem. Biol* 13, 1137–1141. [PubMed: 29544052]
- (29). Zheng Y, Addy PS, Mukherjee R, and Chatterjee A (2017) Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. *Chem. Sci* 8, 7211–7217. [PubMed: 29081953]
- (30). Chatterjee A, Xiao H, and Schultz PG (2012) Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci U. S. A* 109, 14841–14846. [PubMed: 22927411]
- (31). Beranek V, Willis JCW, and Chin JW (2019) An evolved *Methanomethylophilus alvus* pyrrolysyl-tRNA synthetase/tRNA pair is highly active and orthogonal in mammalian cells. *Biochemistry* 58, 387–390. [PubMed: 30260626]
- (32). Meineke B, Heimgaertner J, Lafranchi L, and Elsaesser SJ (2018) *Methanomethylophilus alvus* Mx1201 provides basis for mutual orthogonal pyrrolysyl tRNA/aminoacyl-tRNA synthetase pairs in mammalian cells. *ACS Chem. Biol* 13, 3087–3096. [PubMed: 30260624]
- (33). Willis JCW, and Chin JW (2018) Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem* 10, 831–837. [PubMed: 29807989]
- (34). Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, and Yokoyama S (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N ϵ -(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol* 15, 1187–1197. [PubMed: 19022179]
- (35). Wang L, Zhang Z, Brock A, and Schultz PG (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A* 100, 56–61. [PubMed: 12518054]

- (36). Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, and Yokoyama S (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Mol. Biol* 10, 425–432.
- (37). Wang N, Ju T, Niu W, and Guo J (2015) Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids. *ACS Synth. Biol* 4, 207–212. [PubMed: 24847685]
- (38). Gauss DH, and Sprinzl M (1981) Compilation of tRNA sequences. *Nucleic Acids Res.* 9, r1–r23. [PubMed: 7208351]
- (39). Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, and Isaacs FJ (2013) Genomically recoded organisms expand biological functions. *Science* 342, 357–360. [PubMed: 24136966]
- (40). Nguyen DP, Mahesh M, Elsasser SJ, Hancock SM, Uttamapinant C, and Chin JW (2014) Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc* 136, 2240–2243. [PubMed: 24479649]
- (41). Schmied WH, Elsasser SJ, Uttamapinant C, and Chin JW (2014) Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized Pyrrolysyl tRNA Synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc* 136, 15577–15583. [PubMed: 25350841]
- (42). Li Y, Pan M, Li Y, Huang Y, and Guo Q (2013) Thiol-yne radical reaction mediated site-specific protein labeling via genetic incorporation of an alkynyl-L-lysine analogue. *Org. Biomol. Chem* 11, 2624–2629. [PubMed: 23450369]

Figure 1. Specific interaction between the anticodon of tRNA and *MjTyrRS*. (A) A comparison between tRNA^{UCUA} (with an extended and modified anticodon stem loop for UAGA decoding) and tRNA^{Tyr} (with an anticodon that decodes UAC codon with Tyr). The anticodons are shown in red color. The nucleotides of tRNA were numbered by following the conventional rule.³⁸ (B) The crystal structure of *MjTyrRS*-tRNA^{Tyr} complex (PDB: 1J1U). The anticodon of the tRNA includes G34, U35, and A36.

Figure 2.

GFP fluorescence assays of cells expressing AcPheRS variants. Fluorescence readings of *E. coli* C321. A.exp cells expressing AcPheRS-wt or the evolved AcPheRS mutants together with tRNA_{U^{CUA}} and GFP_{UV}-Asn149UAGA. The expressions were conducted either in the presence or in the absence of 1 mM AcPhe. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplicate measurements with standard deviation. GFP-wt, wild-type GFP without a quadruplet codon.

Figure 3.

Encoding AcPhe and BocLys with two orthogonal quadruplet codons. (A) Chemical structures of AcPhe and BocLys; (B) plasmids for the decoding of two orthogonal quadruplet codons. The GFP mutant, GFP-Tyr39AGGA-Asn149TAGA, contains an AGGA codon at position Tyr39 and a UAGA codon at position Asn149; (C) measurements of fluorescence intensity of cells expressing GFP-Tyr39AGGA-Asn149TAGA under different cultivation conditions. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplet measurements with standard deviation; (D) SDS-PAGE analysis of the expression of GFP-Tyr39AGGA-Asn149-TAGA mutant under different cultivation conditions.

Table 1.

Sequencing Data of the Top Five Hits from Each Library

	Cys231	Phe261	His283	Met285	Asp286	Cys190
Library-1						
AcPheRS-1		Ser	Pro	Pro	Leu	Phe
AcPheRS-2		Phe	Val	Gln	Glu	Phe
AcPheRS-3		Glu	Arg	Ser	Arg	Phe
AcPheRS-4		Ser	His	Met	Glu	Cys
AcPheRS-5		Trp	Gln	Val	Trp	Cys
Library-2						
AcPheRS-6	Phe	Phe	Gln		Pro	Phe
AcPheRS-7	Asn	Pro	His		Gly	Phe
AcPheRS-8	Tyr	Leu	Ala		Gly	Phe
AcPheRS-9	Leu	Ile	His		Asp	Phe
AcPheRS-10	Lys	Tyr	Pro		Val	Phe