

# An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon

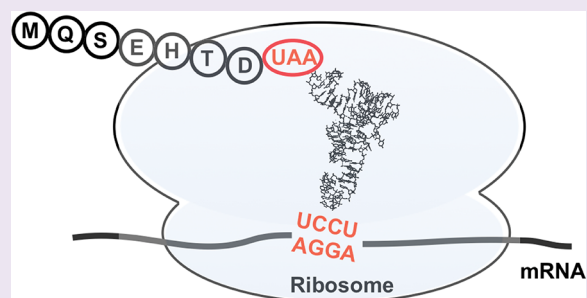
Wei Niu,<sup>†</sup> Peter G. Schultz,<sup>‡</sup> and Jiantao Guo<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, United States

<sup>‡</sup>Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, United States

## S Supporting Information

**ABSTRACT:** We have utilized *in vitro* evolution to identify tRNA variants with significantly enhanced activity for the incorporation of unnatural amino acids into proteins in response to a quadruplet codon in both bacterial and mammalian cells. This approach will facilitate the creation of an optimized and standardized system for the genetic incorporation of unnatural amino acids using quadruplet codons, which will allow the biosynthesis of biopolymers that contain multiple unnatural building blocks.



The genetic code describes how information contained in a nucleic acid sequence, i.e., triplet nucleotide codons, is translated into a polypeptide sequence. Among the 64 DNA codons, 61 encode the 20 common amino acids, and 3 are generally used as stop signals with the rare exceptions of opal (encodes selenocysteine) and amber (encodes pyrrolysine in certain archaeobacteria) nonsense codons. Recently, the amber codon has been used to encode unnatural amino acids (UAAs) in living cells using orthogonal aminoacyl-tRNA synthetase (aaRS)–amber suppressor tRNA ( $tRNA_{CUA}$ ) pairs.<sup>1–6</sup> Engineered archaeal tyrosyl-tRNA synthetase ( $MjTyrRS$ )– $tRNA_{CUA}^{Tyr}$  and pyrrolysyl-tRNA synthetase (PylRS)– $tRNA_{CUA}^{Pyl}$  pairs have enabled the incorporation of more than 60 UAAs into proteins in bacteria and higher organisms.<sup>1–6</sup> To further realize the potential of this expanded genetic code, more “blank” codons (codons that do not encode a natural proteinogenic amino acid) are needed. To this end, quadruplet (frameshift) codons have been utilized to encode UAAs. The incorporation of UAAs in response to quadruplet codons has been achieved in cell-free systems using *in vitro* protein translation methods,<sup>7–11</sup> in *Xenopus* oocytes using micro-injected, chemically aminoacylated tRNAs,<sup>12</sup> and in bacteria.<sup>13,14</sup> However, many of these approaches are hampered by low UAA incorporation efficiency. Here we report a new approach to enhance the efficiency of quadruplet codon suppression based on engineering of the tRNA, rather than other components of the translational machinery. Furthermore, we show for the first time that the evolved tRNA can be readily utilized to incorporate UAAs into proteins in mammalian cells.<sup>15</sup>

## RESULTS AND DISCUSSION

**tRNA Library Construction and Selection.** To determine whether modifications to tRNA structure can lead to improved

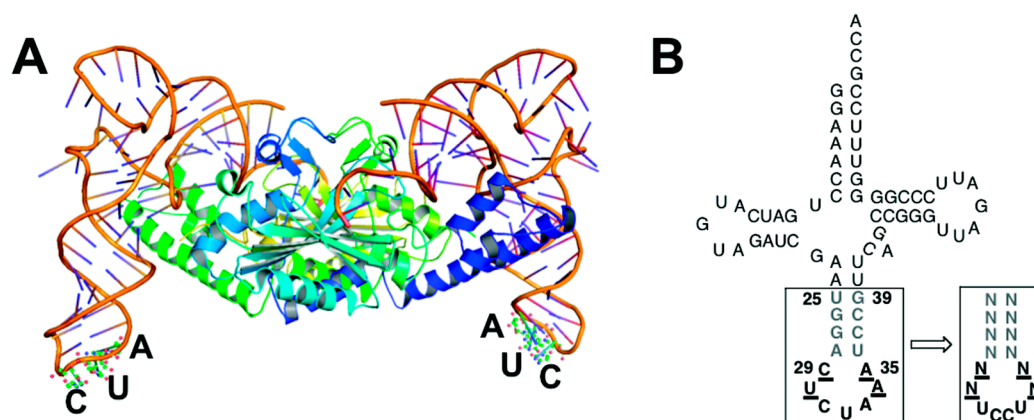
quadruplet codon suppression, we chose to investigate  $tRNA_{CUA}^{Pyl}$ , because its anticodon loop may not act as an important recognition element for its cognate PylRS<sup>16,17</sup> (Figure 1A). In fact, a recent report showed that PylRS could charge  $tRNA_{UUA}^{Pyl}$  and  $tRNA_{UCA}^{Pyl}$  as well.<sup>18</sup> Therefore, mutation of the CUA anticodon to a quadruplet anticodon in  $tRNA_{CUA}^{Pyl}$  may not significantly affect its interaction with PylRS. In addition, previous work has shown that  $tRNA_{CUA}^{Pyl}$  is not recognized by endogenous aaRSs in both *E. coli* and mammalian cells as a result of its unique structural features.<sup>19,20</sup> This property allows us to carry out the *in vitro* evolution experiments in *E. coli* and subsequently shuttle the evolved  $tRNA^{Pyl}$  directly into mammalian cells.

The anticodon of  $tRNA_{CUA}^{Pyl}$  was replaced with UCCU to generate  $tRNA_{UCCU}^{Pyl}$ , which recognizes the quadruplet codon AGGA. On the basis of the notion that mutation of the anticodon loop in  $tRNA_{UCCU}^{Pyl}$  might modulate its binding to the ribosome, we created a tRNA library (library I, theoretical diversity 256, >99% coverage) in which four bases of the anticodon loop (29, 30, 34, 35) were randomized (Figure 1B). The tRNA library was subjected to a positive selection to identify functional  $tRNA_{UCCU}^{Pyl}$  sequences in the presence of BocLysRS, chloramphenicol (34  $\mu\text{g mL}^{-1}$ ), and a chloramphenicol acetyltransferase mutant gene containing the AGGA codon at a permissive site (the codon for Gln98 is replaced with AGGA). The survivors were further screened in the presence and absence of *N*-*tert*-butyloxy-carbonyl-L-lysine (Boc-Lys),<sup>21</sup> and seven individual clones that only survived on chloramphenicol in the presence of Boc-Lys were chosen for DNA sequencing. Five of the sequenced clones converged on a

Received: March 7, 2013

Accepted: May 10, 2013

Published: May 10, 2013



**Figure 1.** Structure of tRNA<sup>Pyl</sup>. (A) Crystal structure of PylRS-tRNA<sup>Pyl</sup> from *Desulfotibacterium hafniense* (PDB ID 2ZNI; the cocrystal structure lacks the N-terminal part of the PylRS).<sup>16</sup> (B) Clover leaf structures of tRNA<sup>Pyl</sup> from *Methanosarcina barkeri*. In bold, anticodon; underlined, nucleotides in the anticodon loop that were randomized; gray in bold, nucleotides in the anticodon stem that were randomized.

unique sequence (referred to as M1) with the mutations C29A and A35C, while the other two converged to a second related sequence (referred to as M2) with the mutation A35U (Table 1).

**Table 1.** Evolved tRNA<sup>Pyl</sup><sub>UCCU</sub> Variants with Improved AGGA Suppression Activity<sup>a</sup>

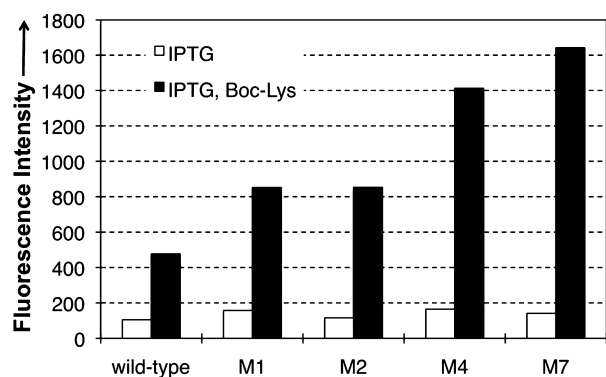
library	tRNA	positions <sup>b</sup>			
		25–28	29, 30	34, 35	36–39
	wt	UGGA	CU	AA	UCCG
Lib I	M1		<u>A</u> <u>U</u>	<u>A</u> <u>C</u>	
	M2		CU	<u>A</u> <u>U</u>	
Lib II	M3	<u>U</u> <u>C</u> <u>G</u> <u>G</u>	<u>A</u> <u>U</u>	<u>A</u> <u>C</u>	<u>C</u> <u>C</u> <u>U</u> <u>A</u>
	M4	<u>U</u> <u>G</u> <u>G</u> <u>G</u>	<u>A</u> <u>U</u>	<u>A</u> <u>C</u>	<u>C</u> <u>C</u> <u>U</u> <u>A</u>
Lib III	M5	<u>A</u> <u>G</u> <u>G</u> <u>G</u>	CU	<u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>G</u> <u>U</u>
	M6	<u>U</u> <u>G</u> <u>G</u> <u>G</u>	CU	<u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>U</u> <u>A</u>
	M7	<u>G</u> <u>G</u> <u>G</u> <u>G</u>	CU	<u>A</u> <u>U</u>	<u>C</u> <u>C</u> <u>G</u> <u>C</u>
consensus		<u>D</u> <u>S</u> <u>G</u> <u>G</u>	<u>M</u> <u>U</u>	<u>A</u> <u>Y</u>	<u>C</u> <u>C</u> <u>K</u> <u>H</u>

<sup>a</sup>Sequences of each tRNA<sup>Pyl</sup><sub>UCCU</sub> variant at randomized positions are listed. <sup>b</sup>The positions with mutations are underlined and in bold. S = G, C; M = A, C; K = G, U; Y = C, U; H = A, U, C; D = U, A, G.

We next sought to further improve the AGGA suppression efficiency by exploring mutation of additional structural elements in tRNA<sup>Pyl</sup><sub>UCCU</sub>. To this end, we constructed two additional tRNA libraries (library II is based on M1, theoretical diversity  $6.6 \times 10^4$ , >99% coverage; library III is based on M2, theoretical diversity  $6.6 \times 10^4$ , >99% coverage) in which the sequences at positions 25–28 and 36–39 (Figure 1B) of the two hits from library I were randomized. We hypothesized that sequence changes in the anticodon stem might have an effect on the interaction between the anticodon loop and ribosome and therefore affect the efficiency of quadruplet codon suppression. Libraries II and III were subjected to the same selection/screening procedure that was used for library I, but with an increased concentration of chloramphenicol ( $50 \mu\text{g mL}^{-1}$ ). Out of the more than  $10^3$  clones that survived on the selection plate, 96 individual clones were replicated on selection plates with a range of chloramphenicol concentrations ( $34, 50, 75, 100, 150 \mu\text{g mL}^{-1}$ ). This experiment provided a qualitative comparison of the efficiency of quadruplet codon suppression by tRNA<sup>Pyl</sup><sub>UCCU</sub> mutants. Eight clones that grew with 100 or  $150 \mu\text{g mL}^{-1}$  chloramphenicol in the presence of Boc-Lys but not

with  $34 \mu\text{g mL}^{-1}$  of chloramphenicol in the absence of Boc-Lys were chosen for sequencing. Three out of eight clones were derived from library II; two of them converged on a unique sequence (referred to as M4; Table 1). The other five clones were derived from library III; two of them converged on a unique sequence (referred to as M6) and another two of them converged on a second unique sequence (referred to as M7; Table 1). All eight clones contain the same A28G and U36C mutations (Table 1), which leads to the replacement of the A28-U36 with a G28-C36 base pair. No mutations were observed at positions 27 and 37, which leaves the G27-C37 base pair intact. The two G-C pairs apparently strengthen the part of the anticodon stem that is directly adjacent to the anticodon loop. Interestingly, all clones contain a mismatched Watson–Crick base pairing at position 26–38 (Table 1). Based on molecular simulation with energy minimization using a CHARMM27 forcefield as implemented in Discovery Studio 2.5, the above mutations result in structural changes (Supplementary Figure 1) in the anticodon (especially the last two bases C and U), anticodon loop, and anticodon stem (at base pairs G26-C38, G27-C37, and A28-U36). Such structural changes in tRNA<sup>Pyl</sup><sub>UCCU</sub> may allow a favorable conformational adjustment of tRNA upon its binding to the A site and subsequent translocation to the P and E sites of the ribosome.<sup>22,23</sup>

**Characterization and Validation of Quadruplet Codon Suppression in *E. coli*.** To further determine the efficiency and fidelity of AGGA suppression by the wild-type and the mutant tRNA<sup>Pyl</sup><sub>UCCU</sub> variants in *E. coli*, a frameshift mutation (AGGA) was introduced into GFP<sub>UV</sub> to replace the Asn149 codon (GFP<sub>UV</sub>-N149AGGA). We constructed plasmid pLei-GFP<sub>UV</sub>-N149AGGA, which contains GFP<sub>UV</sub>-149AGGA (driven by a T5 promoter) and a copy of a tRNA<sup>Pyl</sup><sub>UCCU</sub> variant (wild-type or a mutant; expressed under the control of a *lpp* promoter). Plasmid pLei-GFP<sub>UV</sub>-N149AGGA was cotransformed into *E. coli* Genehog strain with plasmid pBK-BocLysRS, which carries a PylRS variant (Y384F mutant; from *Methanosarcina mazei*) that specifically aminoacylates tRNA<sup>Pyl</sup><sub>CUA</sub> with Boc-Lys.<sup>21</sup> Protein expression was carried out in LB medium supplemented with and without 5 mM Boc-Lys. Fluorescence analysis of *E. coli* cultures showed that full-length GFP<sub>UV</sub> protein was produced only in the presence of Boc-Lys for all tRNA<sup>Pyl</sup><sub>UCCU</sub> variants that were examined (Figure 2). This result indicates that both the wild-type and evolved tRNA<sup>Pyl</sup><sub>UCCU</sub>



**Figure 2.** GFP<sub>UV</sub> fluorescence assays of cells expressing tRNA<sup>Pyl</sup><sub>UCCU</sub> variants. Fluorescence readings of *E. coli* Genehog cells expressing wt tRNA<sup>Pyl</sup><sub>UCCU</sub> (leftmost bar) or the evolved tRNA<sup>Pyl</sup><sub>UCCU</sub> mutants, each coexpressed with BocLysRS. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplet measurements with standard deviation.

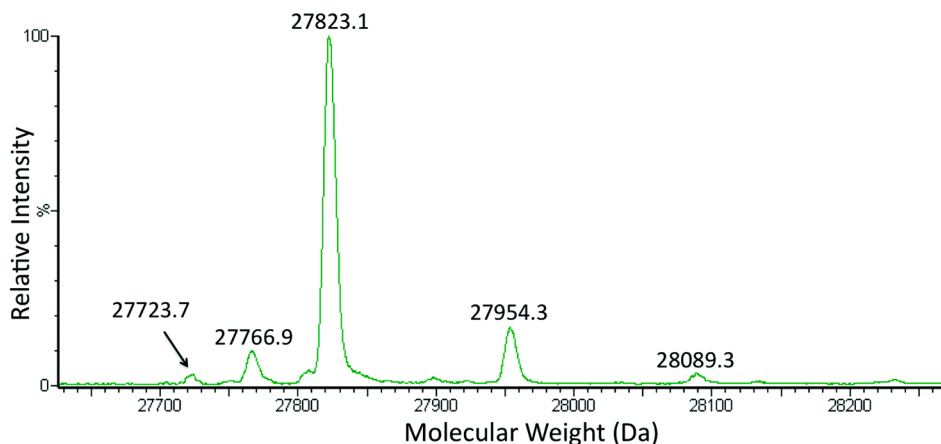
variants are not cross-active with any endogenous aaRSs in *E. coli*. All mutants (M1 and M2 from library I; M4 and M7 from libraries II and III) exhibited enhanced AGGA suppression in the presence of Boc-Lys relative to that of the wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub>. Two representative mutants (M4 and M7) from libraries II and III displayed the largest improvement based on the fluorescence assays (Figure 2).

We further confirmed incorporation of Boc-Lys into GFP<sub>UV</sub>-149AGGA by mass spectrometry (Figure 3). ESI-MS analysis of the partially purified GFP<sub>UV</sub> mutant showed two major signals (27823.1 and 27954.3 Da) corresponding to GFP<sub>UV</sub> protein containing Boc-Lys with and without the N-terminal Met (Figure 3). The yield of the mutant GFP<sub>UV</sub> protein was 1 mg/L after partial purification when M7 was used. For comparison, the yield of GFP<sub>UV</sub>-149TAG mutant in the presence of BocLysRS and tRNA<sup>Pyl</sup><sub>CUA</sub> under the same conditions was 4 mg/L. It was recently reported that the *E. coli* tRNA<sup>Arg</sup><sub>CCU</sub> could efficiently suppress an AGGA codon at position 134 of a sequence-optimized sfGFP variant.<sup>24</sup> However, we did not detect the signal (27752.1 Da) corresponding to the Arg incorporation in GFP<sub>UV</sub>-149AGGA (Supplementary Figure 3 and 4). Since we also did not detect obvious natural amino acid

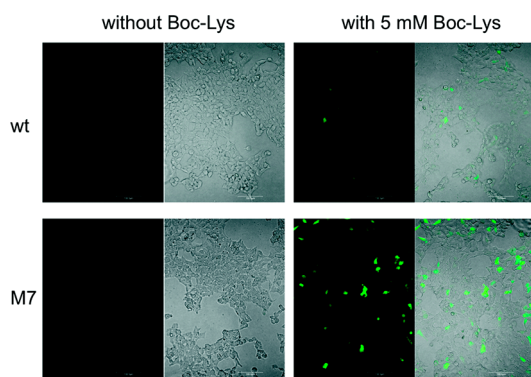
incorporation in response to AGGA codons in chloramphenicol acetyltransferase (position 98, for positive selection) and in barnase (positions 2 and 44, for negative assays; Supplementary Figure 5) for both wt and mutant tRNA<sup>Pyl</sup><sub>UCCU</sub> in our experiments, we hypothesized that the previous reported suppression of AGGA codon by *E. coli* tRNA<sup>Arg</sup><sub>CCU</sub> might be specific to the codon (position 134) context in sfGFP.<sup>24</sup> In addition to Arg, no signals corresponding to the incorporation of other common 20 amino acids at position 149 were observed (Supplementary Figure 3, Supplementary Table 1) either. All of the above observations indicate insufficient levels of natural amino acid incorporation into GFP<sub>UV</sub>-149AGGA and confirmed the good orthogonality of the evolved tRNA<sup>Pyl</sup><sub>UCCU</sub> mutants.

#### Expanded Genetic Code in Mammalian Cells with a Quadruplet Codon.

Next, the activity of the evolved M7 mutant of tRNA<sup>Pyl</sup><sub>UCCU</sub> was examined in mammalian cells. AGGA suppression was monitored using an enhanced GFP (EGFP) with an AGGA mutation at the permissive residue 150 (EGFP-150AGGA). 293T cells were cotransfected with two plasmids, which contained a total of two copies of wild-type or M7 mutant tRNA<sup>Pyl</sup><sub>UCCU</sub> under the control of a human U6 promoter and BocLysRS and EGFP-150AGGA genes under the control of a nonregulated CMV promoter. Following transfection, cells were cultured in DMEM media (containing 10% FBS and 2 mM L-glutamine) with or without 5 mM Boc-Lys for 12 h before visualization under a fluorescence microscope (Figure 4). Full-length EGFP was detected only in cells supplemented with Boc-Lys, while no EGFP was observed otherwise. Significantly enhanced fluorescence was observed in cell cultures transfected with the M7 mutant (Figure 4, panel M7) relative to those transfected with wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> (Figure 4, panel wt). No obvious detrimental effects on cell viability were observed for either the wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> or the M7 mutant. These results correlate well with those in *E. coli*, which indicates that the evolved tRNA<sup>Pyl</sup><sub>UCCU</sub> mutant is also more efficient in decoding AGGA in mammalian cells. Although prokaryotes have 70S ribosomes (30S small subunit, 40S large subunit) whereas eukaryotes have 80S ribosomes (40S small subunit, 50S large subunit), the interaction between ribosome and tRNA may not be that different.



**Figure 3.** Deconvoluted ESI-MS spectra of the GFP<sub>UV</sub>-N149-Boc-Lys mutant. Expected masses: 27824.1 Da (without the N-terminal Met) and 27955.2 (with the N-terminal Met). Observed masses: 27823.1 and 27954.3 Da. The 27723.7 Da signal may correspond to the protein that lost a Boc group under mass spectrometry conditions. The 27766.9 Da signal does not correspond to GFP<sub>UV</sub> mutant that contains any of the 20 proteogenic amino acids at position 149. Supplementary Figure 3 shows the ESI-MS spectra before deconvolution.



**Figure 4.** Suppression of AGGA in EGFP150AGGA protein in 293T cells. The left picture in each panel represents fluorescence image, and the right picture in each panel represents overlapped bright field and fluorescence images.

**Conclusions.** In summary, we have utilized an *in vitro* evolution approach to identify tRNA<sup>Pyl</sup><sub>UCCU</sub> variants with significantly enhanced activity for the incorporation of unnatural amino acids into proteins in response to a quadruplet codon. We also demonstrated that the evolved tRNA<sup>Pyl</sup><sub>UCCU</sub> variants from bacteria can be readily shuttled into a mammalian system. A similar approach is likely applicable to quadruplet codons other than AGGA, which will facilitate the creation of an optimized and standardized system for the genetic incorporation of unnatural amino acids using quadruplet codons. Generation of additional blank codons will allow the biosynthesis of biopolymers that contain more unnatural elements, which will further facilitate the study and manipulation of biological processes.

## METHODS

**Materials and General Methods.** Boc-Lys was purchased from Bachem. Restriction enzymes, Antarctic phosphatase (AP), and T4 DNA ligase were purchased from New England Biolabs. KOD hot start DNA polymerase was purchased from EMD Millipore. Primers were ordered from Eurofins MWG Operon. Protein mass spectrometric data was collected on a Waters Synapt G2 mass spectrometry. Data was processed using Masslynx software (Waters). Standard molecular biology techniques<sup>25</sup> were used throughout. Site-directed mutagenesis was carried out using the QuikChange II site-directed mutagenesis kit (Agilent Technologies) by following the manufacturer's protocol. *E. coli* DH5 $\alpha$  and Genehog were used for routine cloning and DNA propagation. All solutions were prepared in deionized water further treated by Barnstead Nanopure ultrapure water purification system (Thermo Fisher Scientific Inc.). Antibiotics were added where appropriate to the following final concentrations: ampicillin, 100 mg L<sup>-1</sup>; kanamycin, 50 mg L<sup>-1</sup>; tetracycline, 12.5 mg L<sup>-1</sup>.

**Plasmid Construction.** Plasmid pLei-GFP<sub>UV</sub>-N149AGGA was obtained by first introducing the AGGA mutation into plasmid pLei-GFP<sub>UV</sub>-N149TAG<sup>26</sup> by site-directed mutagenesis using following primers, 5'-GAGTACAACACTATAACTCACACAGGAGTATACATCACGGCAGACAA-3' and 5'-TTGTCTGCCGTGATGTATAC-TCCTGTGTGAGTTATAGTTGTACTC-3'. The original tRNA on pLei-GFP<sub>UV</sub>-N149TAG was subsequently replaced with wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub>. Variants of pLei-GFP<sub>UV</sub>-N149AGGA were constructed by replacing the wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> with evolved tRNA mutants.

Plasmid pBK-BocLysRS was constructed by inserting BocLysRS-encoding gene behind the constitutive *glnS* promoter (*P*<sub>*glnS*</sub>) on pBK vector.<sup>27</sup>

Plasmid pREP-BocLysRS-AGGA (for positive selection) was constructed by modifying plasmid pRepCM12b.<sup>26</sup> Specifically, the TAG codon (position 98) in the chloramphenicol acetyltransferase-

encoding gene on pRepCM12b was changed to AGGA by site-directed mutagenesis using following primers, 5'-CGCTCTGGAGTGAATACCACAGGAGATTTCCGGCAGTTTCTACA-3' and 5'-GTAGAAACTGCCGGAATCTCCTGTGGTATTCACTCCAGAGCG-3'. The resulting plasmid, pRepCM12b-AGGA, was digested with *Xba*I and ligated to a DNA fragment containing *P*<sub>*glnS*</sub>-BocLysRS cassette, which was amplified from template pBK-BocLysRS, to yield plasmid pREP-BocLysRS-AGGA.

Plasmid pNEG-AGGA (for negative selection) was constructed by modifying plasmid pNEG(-)tRNA.<sup>26</sup> Codons of Gln2 and Asp44 of the barnase-encoding gene were changed to AGGA by overlapping PCR. The PCR product was digested with *Nde*I and *Eco*RI, then ligated into pNEG(-)tRNA that was treated with the same restriction enzymes to afford plasmid pNEG-AGGA. The resulting plasmid contains the p15A replication origin, the ampicillin resistance marker, the AraC-encoding gene, and the mutant barnase gene that was under the control of the arabinose promoter (pBAD). The overlapping PCR was carried out using the following primers: 5'-GGAATTCATATGGCAAGGAGTTATCAACACGTTTGACGGG-3', 5'-TGCAAGTTCCCTTTTGTATGCCAC-3', 5'-GCATCAAAAAGGGAACCTTGCAAGGAGTCGCTCCGGGGAAAAGCATC-3', and 5'-AGCTTGGCACTGGCCGTCGTT-3'.

Plasmid pCMV-BocLysRS was constructed as follows: The wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> gene lacking 3' CCA termini but with a 3'-TTTTTCT sequence was amplified and fused behind a human U6 promoter by overlapping PCR. The resulting gene cassette was inserted into the *Mfe*I site of pcDNA3.1/hygro(+, Life Technologies) to yield plasmid pcDNA3.1-tRNA<sup>Pyl</sup><sub>UCCU</sub>. The BocLysRS-encoding gene was PCR amplified and inserted behind a nonregulated CMV promoter on pcDNA3.1-tRNA<sup>Pyl</sup><sub>UCCU</sub> to afford plasmid pCMV-BocLysRS. Variants of the pCMV-BocLysRS plasmid were constructed by replacing the wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> with evolved tRNA mutants.

Plasmid pEGFP-150AGGA was constructed similarly as for pCMV-BocLysRS. The EGFP-150AGGA gene was generated using overlapping PCR to replace the codon of Asn150 with AGGA. The resulting gene was inserted behind the nonregulated CMV promoter on pcDNA3.1-tRNA<sup>Pyl</sup><sub>UCCU</sub> to afford plasmid pEGFP-150AGGA. Variants of the pEGFP-150AGGA plasmid were constructed by replacing the wild-type tRNA<sup>Pyl</sup><sub>UCCU</sub> with evolved tRNA mutants.

**tRNA Library Construction.** Procedures for library construction were adapted from Noren and Noren.<sup>28</sup> Mutants of tRNA were obtained by overlapping PCR using pBK-mmPylT as the template.<sup>29</sup> Resulting PCR products contains DNA cassettes of *lpp* promoter-tRNA mutant-rnC terminator. Digestion of PCR products with *Nco*I and *Xho*I followed by ligation between *Nco*I and *Xho*I sites of pBK vector resulted in the tRNA library. The primers that were used for library construction can be found in Supporting Information.

**Positive Selection.** Library DNAs were transformed into DH10B electrocompetent cells containing plasmid pREP-BocLysRS-AGGA. Transformants were cultivated in LB media containing kanamycin and tetracycline. After 12 h of cultivation, cells were harvested. Based on calculation, a certain number of cells (>4.6  $\times$  the size of the library) were plated on LB agar containing kanamycin, tetracycline, Boc-Lys (5 mM), and chloramphenicol (34 mg L<sup>-1</sup> for library I, 50 mg L<sup>-1</sup> for libraries II and III). The selection plates were incubated at 37 °C for 48–72 h. Selected numbers of single colonies were further screened by replication onto plates with certain concentrations of chloramphenicol in the presence and absence of Boc-Lys. Only the ones that grew in the presence of Boc-Lys and did not grow in the absence of Boc-Lys were selected for further evaluation.

**Negative Selection Assay.** *E. coli* Genehog was cotransformed with plasmids pNEG-AGGA and pBK-tRNA (a plasmid contains tRNA<sup>Pyl</sup><sub>UCCU</sub> variant of interest, see library construction) and plated on solid LB media containing ampicillin (to maintain the pNEG-AGGA plasmid), kanamycin (to maintain the pBK-tRNA plasmid), and 0% or 0.2% L-arabinose (*w/v*; to activate the transcription of mutant barnase gene). The cell growth was checked after 12 h of incubation at 37 °C. If the tRNA<sup>Pyl</sup><sub>UCCU</sub> variant of interest is charged by endogenous aaRSs or if the AGGA codon is suppressed by endogenously charged tRNAs, cells should die due to the expression of full-length toxic barnase

protein. Normal cell growth should be observed if the tRNA<sup>Pyl</sup><sub>UCCU</sub> variant of interest is not charged by endogenous aaRSs and if the AGGA codon is not suppressed by any endogenously charged tRNAs (see Supplementary Figure 5 for results).

**Fluorescence Analysis of Bacterial Culture.** *E. coli* Genehog strain harboring plasmid pBK-BocLysRS and a pLei-GFP<sub>UV</sub>-N149AGGA variant was cultured in 5 mL LB media containing kanamycin and chloramphenicol at 37 °C. The protein expression was induced at the OD<sub>600 nm</sub> of 0.6 by additions of IPTG (0.5 mM) and Boc-Lys (5 mM). Following cultivation at 37 °C for an additional 16 h, 1 mL of cell culture was collected, washed, and then resuspended in 1 mL of potassium phosphate buffer (50 mM, pH 7.4). The processed cells were directly used for fluorescence and cell density measurements using a Synergy H1 Hybrid plate reader (BioTek Instruments). The fluorescence of GFP<sub>UV</sub> was monitored at  $\lambda_{\text{Ex}} = 390$  nm and  $\lambda_{\text{Em}} = 510$  nm. The cell density was estimated by measuring the sample absorbance at 600 nm. Values of fluorescence intensity were normalized to cell growth. Reported data are the average of three measurements with standard deviations.

**Protein Expression in Bacteria.** Similar cell cultivation procedure for fluorescence analysis of bacterial culture was applied to *E. coli* strains that were used for protein purification. The only modification is that 100 mL of culture was grown. Cells were collected by centrifugation at 5000g and 4 °C for 15 min. Harvested cells were resuspended in lysis buffer containing potassium phosphate (20 mM, pH 7.4), NaCl (150 mM), and imidazole (10 mM). Cells were subsequently disrupted by sonication. Cellular debris was removed by centrifugation (21,000g, 30 min, 4 °C). The cell-free lysate was applied to Ni Sepharose 6 Fast Flow resin (GE Healthcare). Protein purification followed manufacturer's instructions. Protein concentrations were determined by Bradford assay (Bio-Rad). Purified protein was desalted prior to MS analysis.

**Protein Expression in 293T Cells.** 293T cells were grown in media containing DMEM, 10% FBS (v/v), and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub> (v/v). When cells reached 60–70% confluency, cell culture media was exchanged to fresh DMEM media (containing 10% FBS and 2 mM L-glutamine) with or without 5 mM Boc-Lys and the cells were grown for an additional 30 min before transfection. Cells were transfected with plasmids pCMV-BocLysRS and pEGFP-150AGGA variants using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol (8  $\mu$ L Lipofectamine + 2  $\mu$ g of pCMV-BocLysRS + 2  $\mu$ g of pEGFP-150AGGA for 2 mL cell culture in Costar 6-well cell-culture clusters). Cells were grown for an additional 12 h before being washed with DMEM base medium, fixed with 4% paraformaldehyde (w/v), washed three times with DPBS, and visualized by an Inverted (Olympus IX 81) confocal microscope.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

Additional figures and data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [jguo4@unl.edu](mailto:jguo4@unl.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work is supported by the New Faculty Startup Fund to J.G. from the Chemistry Department of University of Nebraska-Lincoln. The authors thank R. Cerny for help in protein mass spectrometry and Y. Zhou for help in fluorescence microscope analysis.

## ■ REFERENCES

- (1) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444.
- (2) Wu, X., and Schultz, P. G. (2009) Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc.* 131, 12497–12515.
- (3) Chin, J. W. (2012) Reprogramming the genetic code. *Science* 336, 428–429.
- (4) Ai, H.-W. (2012) Biochemical analysis with the expanded genetic lexicon. *Anal. Bioanal. Chem.* 403, 2089–2102.
- (5) Liu, W.-S. R., Wang, Y.-S., and Wan, W. (2011) Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Mol. Biosyst.* 7, 38–47.
- (6) Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 371, 818–822.
- (7) 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.
- (8) Hohsaka, T., and Sisido, M. (2002) Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol.* 6, 809–815.
- (9) 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.
- (10) 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.
- (11) 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.
- (12) Rodriguez, E. A., Lester, H. A., and Dougherty, D. A. (2006) *In vivo* incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8650–8655.
- (13) Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7566–7571.
- (14) Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444.
- (15) 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.
- (16) Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C., and Söll, D. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3141–3146.
- (17) Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 378, 634–652.
- (18) Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P.-J., Russell, D. H., and Liu, W. R. (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.
- (19) Polycarpo, C., Ambrogelly, A., Berube, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L., and Söll, D. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12450–12454.
- (20) Nozawa, K., O'Donoghue, P., Gundllapalli, S., Arais, Y., Ishitani, R., Umehara, T., Söll, D., and Nureki, O. (2009) Pyrrolysyl-tRNA synthetase-tRNA<sup>Pyl</sup> structure reveals the molecular basis of orthogonality. *Nature* 457, 1163–1167.

(21) Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode *N* $\epsilon$ -(*o*-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* 15, 1187–1197.

(22) Phelps, S. S., Gaudin, C., Yoshizawa, S., Benitez, C., Fourmy, D., and Joseph, S. (2006) Translocation of a tRNA with an extended anticodon through the ribosome. *J. Mol. Biol.* 360, 610–622.

(23) Dunham, C. M., Selmer, M., Phelps, S. S., Kelley, A. C., Suzuki, T., Joseph, S., and Ramakrishnan, V. (2007) Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit. *RNA* 13, 817–823.

(24) O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., and Söll, D. (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA<sup>Pyl</sup> for genetic code expansion. *FEBS Lett.* 586, 3931–3937.

(25) Sambrook, J. F., and Russell, D. W. (2000) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(26) Guo, J., Melancon, C. E., III, Lee, H. S., Groff, D., and Schultz, P. G. (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem., Int. Ed.* 48, 9148–9151.

(27) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500.

(28) Noren, K. A., and Noren, C. J. (2001) Construction of high-complexity combinatorial phage display peptide libraries. *Methods* 23, 169–178.

(29) Chen, P. R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B. H., and Schultz, P. G. (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem., Int. Ed.* 48, 4052–4055.