ORTHOGONALITY AND CODON PREFERENCE OF THE PYRROLYSYL-tRNA SYNTHETASE-tRNA^{PYL} PAIR IN *Escherichia coli* FOR THE GENETIC CODE EXPANSION

A Thesis

by

KETURAH AMARKIE ODOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2012

Major Subject: Chemistry

Orthogonality and Codon Preference of the Pyrrolysyl-tRNA Synthetase-tRNA^{Pyl} Pair in *Escherichia coli* for the Genetic Code Expansion

ORTHOGONALITY AND CODON PREFERENCE OF THE PYRROLYSYL-tRNA SYNTHETASE-tRNA^{PYL} PAIR IN *Escherichia coli* FOR THE GENETIC CODE EXPANSION

A Thesis

by

KETURAH AMARKIE ODOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Wenshe Liu
Committee Members,	Tadhg Begley
	Frank Raushel
	Coran Watanabe
Head of Department,	David H. Russell

May 2012

Major Subject: Chemistry

ABSTRACT

Orthogonality and Codon Preference of the Pyrrolysyl-tRNA Synthetase-$tRNA^{Pyl}$ Pair in *Escherichia coli* for the Genetic Code Expansion. (May 2012)

Keturah Amarkie Odoi, B.S., Southwestern Oklahoma State University

Chair of Advisory Committee: Dr. Wenshe Liu

Systematic studies of basal nonsense suppression, orthogonality of $tRNA^{Pyl}$ variants, and cross recognition between codons and tRNA anticodons are reported. *E. coli* displays detectable basal amber and opal suppression but shows a negligible ochre suppression. Although detectable, basal amber suppression is fully inhibited when a pyrrolysyl-tRNA synthetase (PylRS)-$tRNA^{Pyl}_{CUA}$ pair is genetically encoded. $tRNA^{Pyl}_{CUA}$ is aminoacylated by an *E. coli* aminoacyl-tRNA synthetase at a low level, however, this misaminoacylation is fully inhibited when both PylRS and its substrate are present. Besides that it is fully orthogonal in *E. coli* and can be coupled with PylRS to genetically incorporate a NAA at an ochre codon, $tRNA^{Pyl}_{UUA}$ is not able to recognize an UAG codon to induce amber suppression. This observation is in direct conflict with the wobble base pair hypothesis and enables using an evolved *M. jannaschii* tyrosyl-tRNA synthetase-$tRNA^{Pyl}_{UUA}$ pair and the wild type or evolved PylRS-$tRNA^{Pyl}_{UUA}$ pair to genetically incorporate two different NAAs at amber and ochre codons. $tRNA^{Pyl}_{UCA}$ is charged by *E. coli* tryptophanyl-tRNA synthetase, thus not orthogonal in *E. coli*. Mutagenic studies of

$tRNA^{Pyl}_{UCA}$ led to the discovery of its G73U form which shows a higher orthogonality. Mutating $tRNA^{Pyl}_{CUA}$ to $tRNA^{Pyl}_{UCCU}$ not only leads to the loss of the relative orthogonality of $tRNA^{Pyl}$ in *E. coli* but also abolishes its aminoacylation by PylRS.

DEDICATION

To my parents who have sacrificed and made it their priority to see me succeed in life, I dedicate this to you. Thank you for all your love and support, and being there for me every step of the way.

ABBREVIATIONS

Pyl	pyrrolysine
aaRS	aminoacyl-tRNA synthetase
NAA	noncanonical amino acid
BocK	N^{ε} –(tert-butyloxycarbonyl)-L-lysine
AzF	para-azido-L-phenylalanine
PylRS	pyrrolysyl-tRNA synthetase
MjTyrRS	*Methanococcus jannaschii* tyrosyl-tRNA synthetase
ESI-MS	electrospray ionization mass spectrometry
PCR	polymerase chain reaction
sfGFP	superfolder green fluorescent protein
sfGFP134TAG	sfGFP gene with an amber mutation at N134 and a 6x His tag at its C-terminus
sfGFP134TGA	sfGFP gene with an opal mutation at N134 and a 6× His tag at its C-terminus
sfGFP134TAA	sfGFP gene with an ochre mutation at N134 and a 6× His tag at its C-terminus
sfGFP134AGGA	sfGFP gene with an quadruple AGGA mutation at N134 and a 6× His tag at its C-terminus
sfGFP2TAG134TAA	sfGFP gene with an amber mutation at S2, an ochre mutation at N134 and a 6× His tag at its C-terminus
sfGFP2TGA	sfGFP gene with an opal mutation at S2 and a 6× His tag at its C-terminus

IPTG	isopropyl β-D-thiogalactopyranoside
RF	release factor
TrpRS	tryptophanyl-tRNA synthetase
ArgRS	arginyl-tRNA synthetase
Trp	tryptophan
Lys	lysine
Glu	glutamate
Gln	glutamine
Phe	phenylalanine
Arg	arginine

TABLE OF CONTENTS

Page

LIST OF FIGURES

1. INTRODUCTION

Pyrrolysine (Pyl, Fig. 1), the 22nd amino acid that was initially discovered in methanogenic methylamine methyltransferase, is genetically encoded by the RNA nucleotide triplet UAG, a stop codon that halts translation of mRNA during a regular protein translation process (*1-2*). The delivery of Pyl to ribosome is mediated by a unique tRNA, pylT, which is specifically acylated by a unique aminoacyl-tRNA synthetase (aaRS), pyrrolysyl-tRNA synthetase (PylRS) (*3-4*). It contains a special nucleotide triplet, CUA, which is an anticodon that recognizes a UAG stop codon in mRNA (Fig. 2). Unlike $tRNA^{Sec}$ that needs a special elongation factor (SelB in E. coli and EFsec in mammalian cells) and a mRNA secondary structure to deliver to the ribosome A site to bind to a UGA stop codon, the $tRNA^{Pyl}$ hijacks the regular translation elongation process to suppress a UAG codon for the incorporation of Pyl (*5-7*).

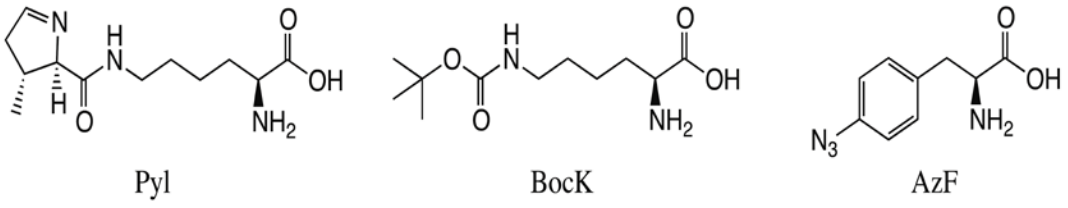


Figure 1. **The structures of the NAAs: Pyl, BocK and AzF.** Pyl is the 22nd amino acid encoded using the amber stop codon. BocK and AzF were also incorporated into protein using the three stop codons (amber, ochre and opal).

This thesis follows the style of *The Journal of Biological Chemistry.*

Previous studies demonstrated that PylRS shows remarkably high substrate promiscuity and is able to charge $tRNA^{Pyl}_{CUA}$ with a variety of noncanonical amino acids (NAAs). For these reasons and the naturally high orthogonality of the PylRS-$tRNA^{Pyl}_{CUA}$ pair in bacteria, yeast, and mammalian cells, this pair has been directly transferred to *E. coli*, *S. cerevisiae*, and human cells for the genetic incorporation of many NAAs such as N^{ε}-(cyclopentyloxycarbonyl)-L-lysine, N^{ε}-(tert-butyloxycarbonyl)-L-lysine (BocK), N^{ε}-D-prolyl-L-lysine, N^{ε}-D-cysteinyl-L-lysine, N^{ε}-(allyloxycarbonyl)-L-lysine, N^{ε}-(propargyloxycarbonyl)-L-lysine, etc into proteins at amber mutation sites (*8-10*).

Evolving PylRS has also allowed the incorporation of N^{ε}-acetyl-L-lysine, N^{ε}-methyl-L-lysine, two photocaged lysines, two photocrosslinking lysine derivatives, a keto-containing lysine analog, and even phenylalanine derivatives into proteins (*11-15*). The genetic incorporation of these NAAs into proteins and their following modifications have enabled a variety of biochemistry studies such as the functional investigation of protein posttranslational modifications, protein folding analysis, biosensor development, tracking signal transduction processes, and probing enzyme mechanisms (*16*).

Recently, two methods were also independently developed in Chin group and our group to genetically incorporate two different NAAs into one protein in *E. coli* by coupling the PylRS-$tRNA^{Pyl}$ pair with another suppressing aaRS-tRNA pair initially derived from the *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)-$tRNA^{Tyr}$ pair (*17-18*). The method developed in Chin group uses one amber codon and one quadruple AGGA codon that were suppressed by the PylRS-$tRNA^{Pyl}_{CUA}$ pair and an evolved

*Mj*TyrRS-$tRNA^{Tyr}_{UCCU}$ pair to code two different NAAs. A special engineered ribosome, Ribo-Q1, has to be used to improve the AGGA suppression level. Our method relies on the suppression of two stop codons, namely one amber UAG codon and one ochre UAA codon, which is achieved by genetically encoding an evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pair for amber suppression and a wild type or evolved PylRS-$tRNA^{Pyl}_{UUA}$ pair for ochre suppression in *E. coli*. One intrinsic merit of our method is its simplicity. No ribosome engineering is necessary. The genetic incorporation of two different NAAs into one protein can be potentially applied to install a FRET pair to a protein for conformation and dynamic studies. We demonstrated in a separate publication how to synthesize proteins with two different posttranslational modifications for their functional investigation, and build phage-displayed peptide libraries with the expanded chemical diversities (18).

Although the PylRS-$tRNA^{Pyl}$ pair has been used extensively to expand the genetic code for the incorporation of different NAAs in the past few years, there are several critical questions related to the pair that have not been seriously addressed.

First, the pair is directly transferred to *E. coli*, yeast, and human cells without a thorough investigation on whether the pair is truly orthogonal in these cell lines. Several publications in which the wild type PylRS-$tRNA^{Pyl}_{UUA}$ pair was used to code NAAs clearly showed a low level of amber suppression even without providing any NAA. Given that PylRS is specific for Pyl, this low level of amber suppression may be due to

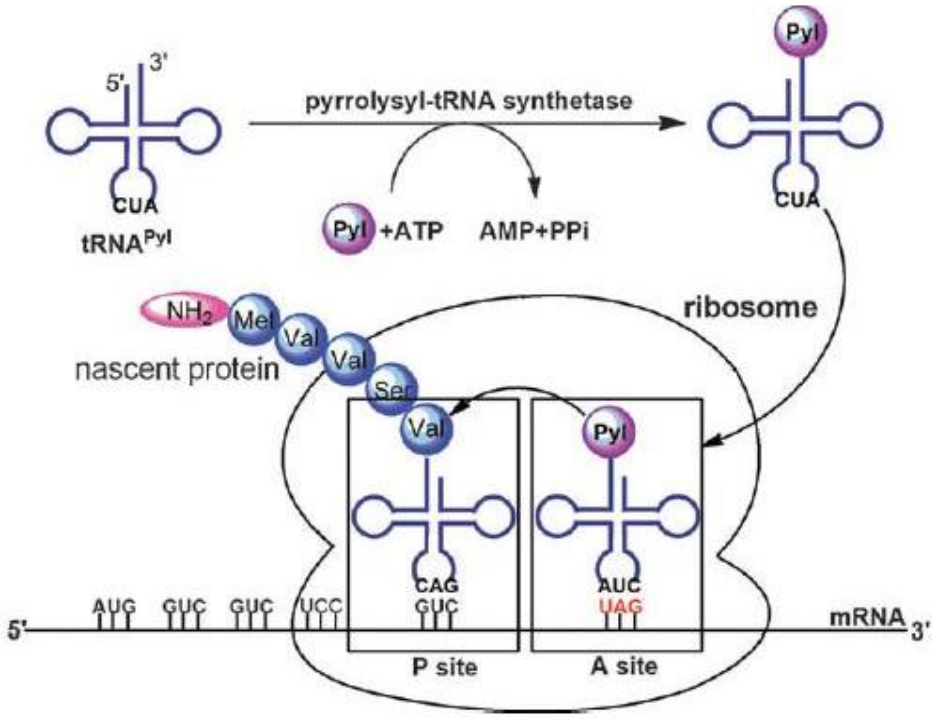


Figure 2. **The Pyl (22nd amino acid) incorporation machinery.** Following the incorporation method of the Pyl, several unnatural amino acids have been site specifically incorporated into proteins using an engineered aaRS-tRNA pair. The engineered aaRS is used to charge the desired unnatural amino acid on to an orthogonal tRNA.

the misaminoacylation of $tRNA^{Pyl}$ by endogenous aaRSs in *E. coli*. Second, PylRS does not specifically recognize the anticodon of $tRNA^{Pyl}_{CUA}$, but whether we can directly mutate the anticodon of $tRNA^{Pyl}_{CUA}$ and use mutant $tRNA^{Pyl}$ forms as orthogonal tRNAs in *E. coli*, yeast, and other cells needs to be further investigated. Last but not the least, a study is necessary to clarify whether an aminoacylated $tRNA^{Pyl}_{UUA}$ can lead to suppression at a UAG codon since an ochre anticodon can recognize an amber codon based on the wobble base pairing hypothesis (*19*). In this study, we attempt to address these issues and also analyze basal suppression of nonsense codons in *E. coli*.

2. MATERIALS AND METHODS

Materials

Phusion high-fidelity DNA polymerase, T4 DNA ligase, T4 polynucleotide kinase (T4 PNK), and restriction enzymes were purchased from New England Biolabs. Oligonucleotide primers were ordered from Integrated DNA Technologies (IDT). Ni-NTA superflow resins were purchased from Qiagen. All polymerase chain reactions (PCRs) were performed using Phusion high-fidelity DNA polymerase. BocK was purchased from Chem Impex. *p*-Azido-L-phenylalanine (AzF) was synthesized according to a revised literature procedure (*20*). DNA sequencing services were provided by Gene Technologies Lab at Texas A&M University. Electrospray ionization mass spectrometry (ESI-MS) analysis of purified proteins was provided by Laboratory for Biological Mass Spectrometry at Texas A&M University.

Plasmid construction

The plasmid maps are presented in the appendix A. The full sequence of the different modification of $tRNA^{Pyl}$ and sfGFP are also provided in the appendix A.

Plasmids pBAD-sfGFP that carries a sequence-optimized superfolder green fluorescent protein (sfGFP) gene with a 6×His tag at its C terminus and pBAD-sfGFP134TAG that carries a sequence-optimized superfolder GFP (sfGFP) gene with an amber mutation at the N134 position and a 6×His tag at its C terminus (the sfGFP134TAG gene) were kind gifts from Dr. Ryan Mehl of Franklin & Marshall College. Plasmid pEVOL-pylT that carries a $tRNA^{Pyl}_{CUA}$ gene under control of a *proK*

promoter and a *proK* terminator was derived from a pEVOL plasmid obtained from Dr. Peter Schultz of Scripps Research Institute. Its construction was shown in a separate publication (*21*).

Plasmid pETtrio-pylT(UUA)-PylRS-MCS was derived from pPylRS-pylT-GFP1TAG149TAA and carries a $tRNA^{Pyl}_{UUA}$ gene (a C34U form of $tRNA^{Pyl}_{CUA}$) under control of the *lpp* promoter and the *rrnC* terminator, the wild type *Methanosarcina mazei* PylRS gene are under the control of the *glnS* promoter and terminator, and multiple cloning sites including *NcoI, NotI, SalI and KpnI are* under the control of the T7 promoter and terminator.

To construct pETtrio-pylT(UUA)-PylRS-MCS, two oligonucleotide primers forward primer NP2 5'-AGCGCGGCCGCGTCGACGGTACCCTCGAGTCTGGTAAAG-3' and reverse primer NP2 5'- ATTGCGGCCGCCCATGGTATATCTCCTTCTTATACTTAAC-3' were used to undergo PCR to amplify pPylRS-pylT-GFP1TAG149TAA. The blunt-end PCR product was phosphorylated at its two 5' ends and directly ligated using T4 DNA ligase to form pETtrio-pylT(UUA)-PylRS-MCS.

Plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAG was constructed by cloning the sfGFP134TAG gene to the *NcoI* and *KpnI* sites of pETtrio-pylT(UUA)-PylRS-MCS. The sfGFP134TAG gene was PCR amplified from pBAD-sfGFP134TAG using two oligonucleotide primers sfGFP134TAG-F 5'-GATATACCATGGTTAGCAAAGGTGAAGAACTG-3' and sfGFP134TAG-R 5'-

CTCGAGGGTACCTCAATGGTGATGATGATGGTG-3'. The PCR for the sfGFP134TAG gene amplification was carried out in the following conditions: 10 µL of 5X Phusion buffer, 1 µl of 10 mM dNTP mix, 0.5 µL of each primer (sfGFP134TAG-F and sfGFP134TAG-R), 0.5 µL of template (pBAD-sfGFPN134TAG), 0.4 µL Phusion polymerase, and 37 µL of water to get 50 µL total volume. The cycling conditions are as follows: (1) 94ºC - 5min, (2) 95ºC - 30sec, (3) 50ºC - 30sec, 4) 72ºC - 60sec, 5) Go to step 2, 29 times, 6) 72ºC - 5min. 1% agarose gel electrophoresis was used to confirm the sfGFP gene. QIAquick gel extraction kit was used to extract and clean the PCR product. The amplified gene was digested by *NcoI* and *KpnI* restriction enzymes and ligated to pETtrio-pylT(UUA)-PylRS-MCS that was predigested by the same enzymes.

Several plasmids listed below are derived from pETtrio-pylT(UUA)-PylRS-sfGFP134TAG. They vary at anticodon of $tRNA^{Pyl}$ and have different nonsense mutations at N134 position of the sfGFP gene. Constructions of these plasmids were carried out using a site-directed mutagenesis protocol based on Phusion DNA polymerase.

In brief, two oligonucleotide primers, one of which covers the mutation site were used to amplify the whole plasmid of pETtrio-pylT(UUA)-PylRS-sfGFP134TAG to give a blunt-end PCR product. This PCR product was phosphorylated by T4 PNK and then ligated to itself using T4 DNA ligase. Primers pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TAG-R 5'-TCTAAATCCGTTCAGCCGGGTTAG-3' were used to construct pETtrio-pylT(CUA)-

PylRS-sfGFP134TAG that carries a $tRNA^{Pyl}_{CUA}$ gene and sfGFP134TAG. Primers sfGFP134-F 5'- GGCAACATTCTGCATAAACTGGA-3' and sfGFP134TAA-R 5'-TTATTCTTTAAAATCAATACCTTTCAGTTCAATGC-3' were used to construct pETtrio-pylT(UUA)-PylRS-sfGFP134TAA that carries a $tRNA^{Pyl}_{UUA}$ gene and a sfGFP gene with an ochre mutation at N134 position and a C-terminal 6×His tag (sfGFP134TAA). Two set of primers: (1) pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TGA-R 5'-TTCAAATCCGTTCAGCCGGGTTAG-3'and (2) sfGFP134-F 5'-GGCAACATTCTGCATAAACTGGA-3' and sfGFP134TGA-R 5'-TCATTCTTTAAAATCAATACCTTTCAGTTCAATGC-3' were used to run two consecutive site-directed mutagenesis reactions to obtain plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134TGA that carries a $tRNA^{Pyl}_{UCA}$ gene and a sfGFP gene with an opal mutation at N134 position and a C-terminal 6×His tag (sfGFP134TGA). The pylT (UCA) gene sequence was confirmed using the primer PylT-SphI-F 5'-GGAATGGTGCATGCTCGAACTTTT-3'.

Two sets of primers (1) pylT-F 5'- GTCCATTCGATCTACATGATCAGGTT-3' and pylT-AGGA-R 5'- TTCCTAATCCGTTCAGCCGGGTTAG-3' and (2) sfGFP134-F 5'- GGCAACATTCTGCATAAACTGGA-3' and sfGFP134AGGA-R 5'-TCCTTTCTTTAAAATCAATACCTTTCA-3' were also used to run two consecutive site-directed mutagenesis reactions on pETtrio-pylT(UUA)-PylRS-sfGFP134TAG to generate plasmid pETtrio-pylT(UCCU)-sfGFP134AGGA that carries a $tRNA^{Pyl}_{UCCU}$ gene

with a quadruple UCCU anticodon and a sfGFP gene with a quadruple AGGA mutation at N134 position and C-terminal 6×His tag (sfGFP134AGGA). Plasmid pETtrio-pylT(CUA)-PylRS-sfGFP134TAA carries a $tRNA^{Pyl}_{CUA}$ gene and sfGFP134TAA. This plasmid was made from plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAA by the same site-directed mutagenesis using two primers pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TAG-R 5'-TCTAAATCCGTTCAGCCGGGTTAG-3. *Duet-F* 5'-TTGTACACGGCCGCATAATC-3') was used to confirm the mutation at the 134 position on sfGFP gene.

Plasmid pETtrio-pylT(UUA)-PylRS-sfGFP2TAG134TAA that carries the sfGFP gene with an amber mutation at S2, an ochre mutation at N134 position, and a 6×His tag at its C-terminus (sfGFP2TAG134TAA) was constructed from pETtrio-pylT(UUA)-PylRS-sfGFP134TAA. Two primers sfGFP2TAG-F 5'-GTTAGCAAAGGTGAAGAACTGTTTACCGG-3' and sfGFP2TAG-R 5'-CTATGCCATGGTATATCTCCTTCTTATACTTAAC-3' were used to amplify pETtrio-pylT(UUA)-sfGFP134TAA. The PCR product was phosphorylated and then ligated to form pETtrio-pylT(UUA)-PylRS-sfGFP2TAG134TAA

Plasmid pETtrio-sfGFP134TGA was derived from pETtrio-pylT(UCA)-PylRS-sfGFP134TGA by digesting it with *SphI* restriction enzyme to remove the $tRNA^{Pyl}_{UCA}$ and parts of the PylRS gene and self-ligate the purified digested plasmid backbone. Plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA contains a $tRNA^{Pyl}_{UCA}$ gene with a UCA

anticodon, a PylRS gene and the sfGFP gene with an opal mutation at S2 and a 6×His tag at its C-terminus (sfGFP2TGA). To construct this plasmid, the sfGFP gene in pBAD-sfGFP was PCR amplified using primers sfGFP2TGA-F 5'-AAAGGTGAAGAACTGTTTACCGGCGT-3' and sfGFP2TGA-R 5'-TCAAACCATGGTATATCTCCTTCTTAT-3'. The PCR product was digested by *NcoI* and *KpnI* restriction enzymes and cloned into the same two sites of pETtrio-pylT(UCA)-PylRS-sfGFP134TGA which was precut by *NcoI* and *KpnI* restriction enzymes to remove the sfGFP134TGA insert. The sfGFP2TGA gene sequence was confirmed using the sfGFP134TAG-R 5'-CTCGAGGGTACCTCAATGGTGATGATGATGGTG-3'.

Three plasmids pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA, and pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA carry mutations that change G73 of $tRNA^{Pyl}_{UCA}$ to C, U, and A, respectively. These plasmids were derived from pETtrio-pylT(UCA)-PylRS-sfGFP134TGA using Phusion DNA polymerase-based site-directed mutagenesis. To make pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, two primers pylT-G73-F 5'-GGAAACCCCGGGAATCTAACCCGG-3' and pylT-G73C-R 5'-CCCACTGCCCATCCTTAGCGAAAGC-3' were employed. Primers pylT-G73-F 5'-GGAAACCCCGGGAATCTAACCCGG-3'and pylT-G73U-R 5'-TCCACTGCCCATCCTTAGCGAAAGC-3' were used to construct pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA. Primers pylT-G73-F 5'-GGAAACCCCGGGAATCTAACCCGG-3'and pylT-G73A-R 5'-

ACCACTGCCCATCCTTAGCGAAAGC -3' were used to construct pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA.

Protein expression of codon-anticodon pairs

Background suppression of an amber mutation at N134 position of sfGFP *E. coli* Top10 cells were transformed with pBAD-sfGFP134TAG and grown in 1 L of 2YT medium that contained 100 μg/mL ampicillin. Cells were let grow to OD_{600}~1.2. Then 0.2% arabinose was added to the medium to induce expression of sfGFP. The induced cells were let grow at 37 °C for 7 hours and then collected by centrifugation (4200 rpm for 20 minutes). The collected cells were resuspended in 35 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication in an ice water bath (temperature for ice bath was 10 °C before each 3 mins run of the sonication). After sonication, the lysed cells were clarified by centrifugation (10000 rpm for 1 hour) and the supernatant was decanted and let bind to 5 mL of Ni-NTA superflow resin at 4 °C for 1 hour. The mixture of the supernatant and resin was then loaded to an empty Qiagen Ni-NTA superflow column and allowed to flow through column by gravity. The resin was washed with 5 times of lysis buffer and sfGFP was then eluted using elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0).

To further purify the expressed sfGFP, the protein was equilibrated against buffer A (20 mM Bis-Tris, pH 6.1) via dialysis and then loaded to a cation exchange column from GE Health Science. The protein was eluted out by running a gradient from buffer A to 100% of buffer B (20 mM Bis-Tris, 1 mM NaCl, pH 6.1) in 60 minutes. The purified protein was then desalted and concentrated to a desired volume using the Amicon ultra 4

centrifugal filter units from Millipore (10 kDa NMWL). The collected protein samples were analyzed by 12 % SDS-PAGE. Protein bands stained with Coomassie blue until visible. Destained gel with 40% methanol, 10% acetic acid, 50% water.

To analyze the purified protein by electrospray ionization mass spectrometry (ESI-MS), the buffer of the purified protein was changed to the phosphate buffer saline. The same experiment was carried out with cells transformed with pBAD-sfGFP.

A separate expression was carried out with cells cotransformed with pBAD-sfGFP134TAG and pEVOL-pylT. Except the addition of 34 μg/mL chloramphenicol to maintain pEVOL-pylT in cells, protein expression conditions, protein purification procedures, and characterization of the finally purified protein were the same as those discussed above. Without further indication, protein purification and characterization in the following experiments were the same as well.

Amber, opal, and ochre suppression analysis of the PylRS-tRNA^{Pyl} pairs

Plasmids pETtrio-pylT(CUA)-PylRS-sfGFP134TAG, pETtrio-pylT(UCA)-PylRS-sfGFP134TGA, and pETtrio-pylT(UUA)-PylRS-sfGFP134TAA were individually used to transform *E. coli* BL21(DE3) cells. For each plasmid, a single colony was selected and let grow in 5 mL of LB medium containing 100 μg/mL ampicillin at 37 °C overnight. The overnight culture was inoculated into 500 mL of 2YT medium supplemented with 100 μg/mL ampicillin and let grow at 37 °C to OD_{600}~1.2. 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and 5 mM BocK were then added to the medium to induce expression of sfGFP. Control experiments in which only 1 mM IPTG

was added to the medium was also carried out. The induced cells were let grow at 37 °C for 7 hours.

Studies on anticodon-codon pairs

Anticodon-codon cross recognition analysis Plasmids pETtrio-pylT(CUA)-PylRS-sfGFP134TAG, pETtrio-pylT(CUA)-PylRS-sfGFP134TAA, pETtrio-pylT(UUA)-PylRS-sfGFP134TAG, and pETtrio-pylT(UUA)-PylRS-sfGFP134TAA were individually used to transform *E. coli* BL21(DE3) cells. A single colony for each plasmid was then selected and let grow in 5 mL of LB medium supplemented with 100 μg/mL ampicillin at 37 °C overnight. This overnight culture was then inoculated into 500 mL of 2YT medium supplemented with 100 μg/mL ampicillin and let grow to OD_{600}~1.2. 1 mM IPTG and 5 mM BocK were then provided and cells continued to grow at 37 °C for 7 h. For pETtrio-pylT(CUA)-PylRS-sfGFP134TAG and pETtrio-pylT(UUA)-PylRS-sfGFP134TAG, expression of sfGFP in the absence of BocK was also tested.

Competitive recognition of the third nucleotide of an amber codon Plasmid pEVOL-AzFRS was a gift from Dr. Peter Schultz at Scripps Research Institute (*21*). It carries one $tRNA^{Tyr}_{CUA}$ gene under control of a *proK* promoter and a *proK* terminator, one *Mj*TyrRS gene under control of a *glnS* promoter and a *glnS* terminator, and one evolved *Mj*TyrRS gene under control of a pBAD promoter. This plasmid together with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG was used to cotransform *E. coli* BL21(DE3) cells. One single colony was selected and let grow in 5 mL of LB medium supplemented with

100 μg/mL ampicillin and 34 μg/mL chloramphenicol at 37 °C overnight. This overnight culture was inoculated into 500 mL of 2YT medium supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol and let grow to OD_{600}~1.2. Expression of sfGFP was then induced. Four induction conditions were tested, including (1) 1 mM IPTG only, (2) 1 mM IPTG and 1 mM AzF, (3) 1 mM IPTG and 5 mM BocK, and (4) 1 mM IPTG, 1 mM AzF, and 5 mM BocK.

Expression of sfGFP incorporated with two different NAAs Plasmids pEVOL-AzFRS and pETtrio-pylT(UUA)-sfGFP2TAG134TAA were used to cotransform *E. coli* BL21(DE3) cells. A single colony was then chosen to express sfGFP at different conditions. The expression procedures including four induction conditions were the same as those shown in the previous competitive recognition analysis.

Mutagenic analysis of opal codon and tRNA$^{Pyl}_{UCA}$

Suppression of an opal mutation at S2 of sfGFP Plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then used to express sfGFP at two induction conditions: (1) 1 mM IPTG and (2) 1 mM IPTG and 5 mM BocK.

Orthogonality test of tRNA$^{Pyl}_{UCA}$ Plasmid pETtrio-sfGFP134TGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then selected to do protein expression that was induced by the addition of 1 mM IPTG.

Mutagenic analysis of tRNA$^{Pyl}_{UCA}$ Plasmids pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA, and pETtrio-

pylT(UCA)G73A-PylRS-sfGFP134TGA were used individually to transform *E. coli* BL21(DE3) cells. A single colony for each plasmid was then selected to do protein expression at two induction conditions: (1) 1 mM IPTG and (2) 1 mM IPTG and 5 mM BocK.

Suppression of a quadruple codon

Plasmids pETtrio-pylT(UCCU)-PylRS-sfGFP134AGGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then selected to expression sfGFP at two induction conditions: (1) 1 mM IPTG and (2) 1 mM IPTG and 5 mM BocK.

3. RESULTS

Background suppression and orthogonality of $tRNA^{Pyl}_{CUA}$

The strict protein translation termination at an amber stop codon and its suppression in *E. coli* were initially studied with T4 phage and later extended to protein models including T4 lysozyme, LacZ, and proteins coded in the trp operon. (*22*) *E. coli* cells with no genetic defects in their translation machineries typically exhibited undetectable or extremely low suppression at an amber mutation site when grown in a minimal medium except that an amber suppressor tRNA was genetically coded. These early studies led to our speculation that *E. coli* cells grown in a rich medium would also display very low to undetectable amber suppression. However, a low expression level of sfGFP was detected in *E. coli* Top10 cells that were transformed with only pBAD-sfGFPN134TAG and grown in 2YT medium, indicating a basal suppression level of the amber mutation at the N134 position of sfGFP (Fig. 3A). The expression yield was 0.156 mg/L which could either be lysine (Lys), glutamate (Glu), or glutamine (Gln) residue incorporated at N134. *E. coli* Top10 cells which were transformed with pBAD-sfGFP gave a sfGFP expression yield of 776 mg/L at the same expression condition. This indicates read through at an amber codon (0.02 %) caused by basal suppression for a sense codon.

Cells cotransformed with both pBAD-sfGFP134TAG and pEVOL-pylT showed an increased basal amber suppression level but not to a significant extent, suggesting that $tRNA^{Pyl}_{CUA}$ is charged by certain endogenous aaRS(s) in *E. coli* (Fig. 3).

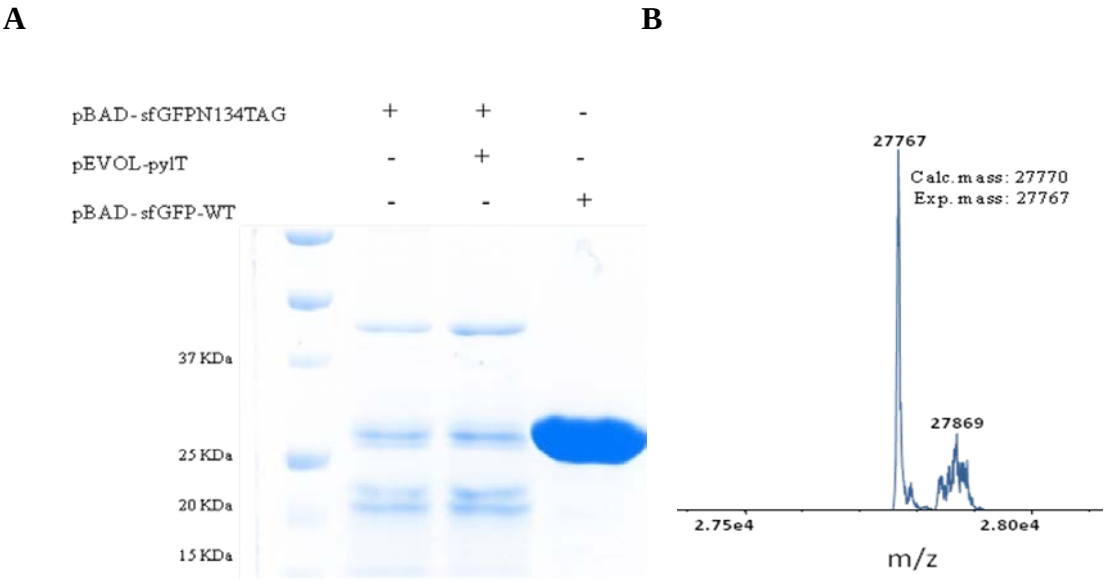


Figure 3. **Background suppression and orthogonality of tRNA$^{Pyl}_{CUA}$.** (A) Background suppression of an amber mutation at N134 of sfGFP. The first lane shows expression of sfGFP in cells that were transformed with pBAD-sfGFP134TAG. The second lane show expression of sfGFP in cells transformed with pBAD-sfGFP134TAG and pEVOL-pylT. (B) ESI-MS of sfGFP expressed in cells transformed with pBAD-sfGFP134TAG and pEVOL-pylT

Amber, opal, and ochre suppression efficiencies of the PylRS-tRNA^{Pyl} pairs

To demonstrate amber suppression efficiency of the PylRS-tRNA$^{Pyl}_{CUA}$ pair, *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG were used to express sfGFP in the absence or presence of 5 mM BocK, a substrate of PylRS. Without BocK in the growth medium, only basal level expression of sfGFP was observed. On the contrary, the addition of BocK promoted sfGFP overexpression (Fig. 4). This demonstrates that the PylRS-tRNA$^{Pyl}_{CUA}$ pair is an efficient amber suppressing

aaRS-tRNA pair and its mediated suppression is dependent on the presence of a NAA. The ESI-MS analysis of the purified sfGFP displayed two major mass peaks at 27,809 Da and 27,940 Da that agree well with the theoretical molecular weights of sfGFP with BocK incorporated at N134 (27,940 Da for the full-length protein; 27809 Da for the full-length protein without the first methionine (M1)) (Fig. 5A). No mass peaks that match a Lys/Glu/Gln residue at N134 of sfGFP could be detected. This indicates that misincorporation resulted from basal amber suppression was inhibited.

It has been previously demonstrated that $tRNA^{Pyl}$ is not hardwired to recognize an amber codon. (7) Mutagenizing the anticodon of $tRNA^{Pyl}$ from CUA to other triplet anticodons does not significantly alter the aminoacylation potential of $tRNA^{Pyl}$ by PylRS. (7) As expected, mutating C34 in $tRNA^{Pyl}_{CUA}$ to U34 in $tRNA^{Pyl}_{UUA}$ did not significantly change its recognition by PylRS. Cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA showed a negligible expression level of sfGFP when BocK was absent in the growth medium, indicating a high orthogonality of $tRNA^{Pyl}_{UUA}$ in *E. coli*. The addition of BocK induced sfGFP overexpression (Fig. 4). The ESI-MS analysis of the purified sfGFP showed two mass peaks (27,939 Da and 27,809 Da) that agree well with the theoretical molecular weights of sfGFP with BocK incorporated at N134 (Fig. 5B). It is obvious that $tRNA^{Pyl}_{UUA}$ is charged with BocK by PylRS and delivers BocK to the ochre mutation site at N134 of sfGFP. In comparison to the amber suppression method, the expression level of sfGFP using the ochre suppression method is lower.

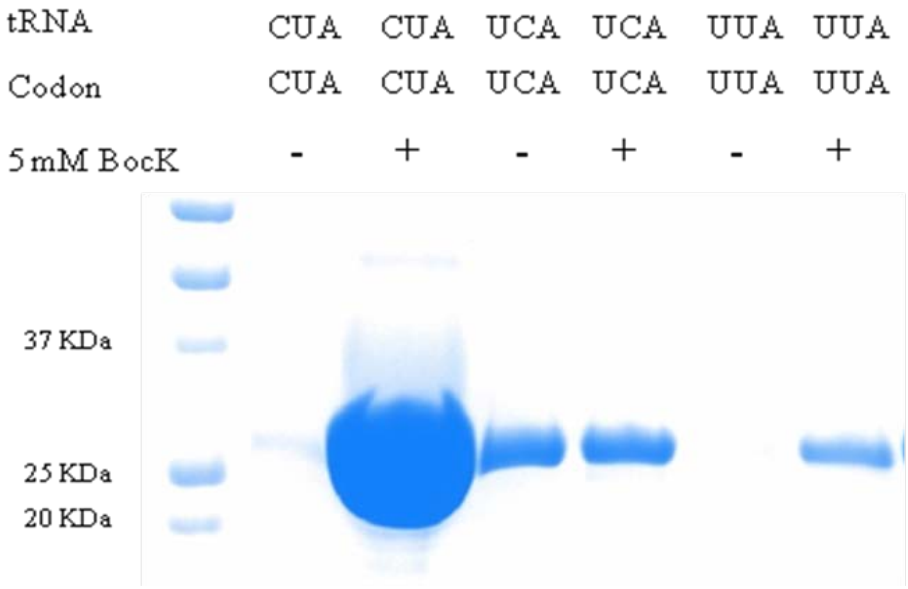


Figure 4. **Expression analysis of the sfGFP variants generated.** Suppression of amber, opal, and ochre mutations at N134 of sfGFP by their corresponding suppressing PylRS-tRNA^{Pyl} pairs in the absence and presence of BocK. Proteins shown in this gel represent their real expression levels. The first two lanes show sfGFP expression levels in cells transformed with pETtrio-PylT(CUA)-PylRS-sfGFP134TAG; the third and fourth lanes show sfGFP expression levels in cells transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA; the last two lanes show sfGFP expression levels in cells transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAA.

On the contrary to $tRNA^{Pyl}_{CUA}$ and $tRNA^{Pyl}_{UUA}$, $tRNA^{Pyl}_{UCA}$ that is a C34U/U35C mutant form of $tRNA^{Pyl}_{CUA}$ is apparently not orthogonal in *E. coli*. Cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA exhibited a very high expression level of sfGFP both in the absence and in the presence of BocK in the media (Fig. 4B).

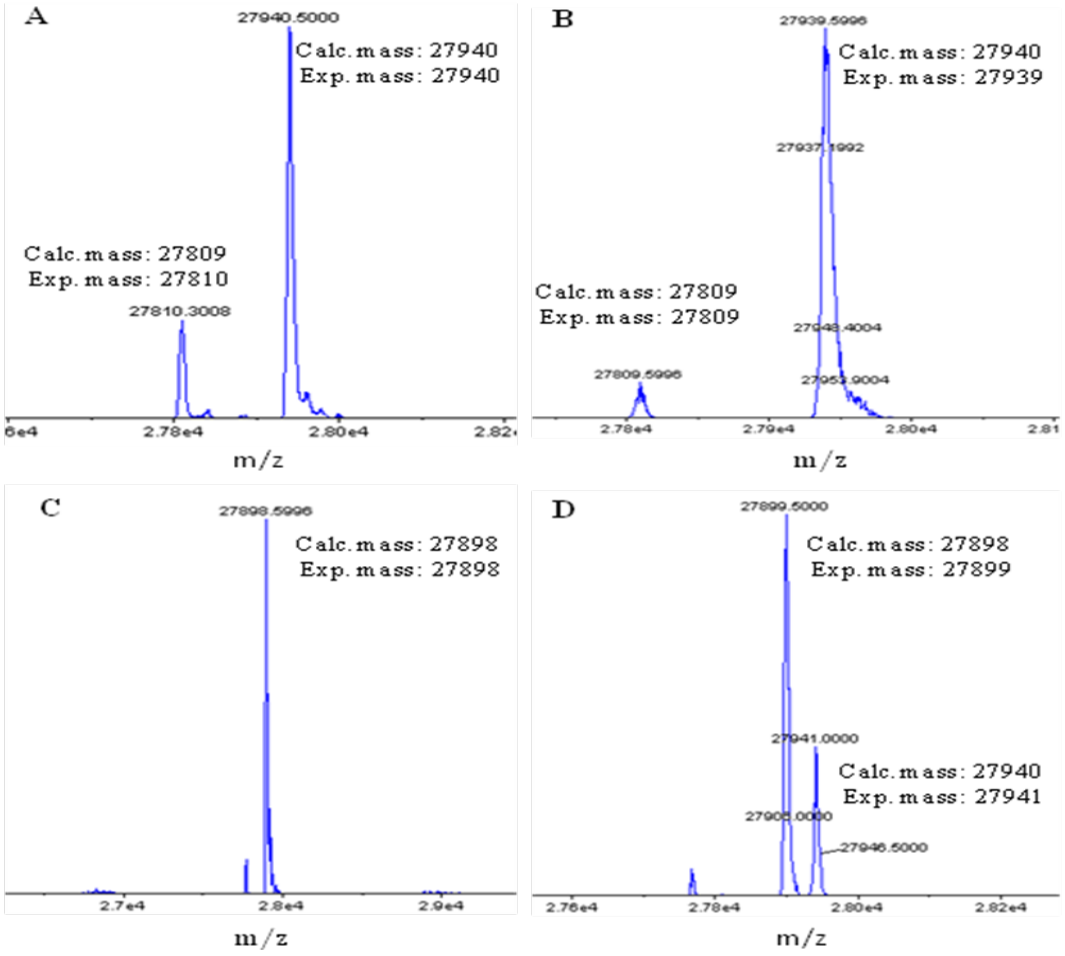


Figure 5. **ESI-MS of sfGFP expressed in cells**. (A) transformed with pETtrio-PylT(CUA)-PylRS-sfGFP134TAG and grown in the presence of 5 mM BocK, (B) transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAA and grown in the presence of 5 mM BocK, (C) transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA and grown in the absence of 5 mM BocK, and (D) transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA and grown in the presence of 5 mM BocK.

The ESI-MS analysis of purified sfGFP expressed in the absence of BocK showed a mass peak at 27,898 Da that clearly matched a tryptophan (Trp) residue at N134 of sfGFP (calculated mass: 27,898 Da) (Fig. 5C), suggesting that the endogenous tryptophanyl-tRNA synthetase (TrpRS) efficiently aminoacylates $tRNA^{Pyl}_{UCA}$. The ESI-MS analysis of sfGFP expressed in the presence of BocK displayed a very interesting spectrum. Mass peaks for both a Trp residue at N134 of sfGFP (27,899 Da) and a BocK residue at N134 of sfGFP (27,940 Da) were observed and the mass peak for the Trp isoform was much more intensive than the BocK isoform (Fig. 5D). This indicates TrpRS is able to effectively inhibit the PylRS-catalyzed aminoacylation reaction of $tRNA^{Pyl}_{UCA}$.

Anticodon-codon cross recognition

We previously used an evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pair for amber suppression and a wild type or evolved PylRS-$tRNA^{Pyl}_{UUA}$ pair for ochre suppression to genetically incorporate two different NAAs into one protein in *E. coli*. (*18*) One precondition of using this technique is that an aminoacyl-$tRNA^{Pyl}_{UUA}$ does not lead to efficient suppression at a UAG codon, yet a UUA anticodon is supposed to recognize a UAG codon based on the wobble base pairing hypothesis. (19) To prove this, we carried out an anticodon-codon cross recognition analysis. Plasmid pETtrio-pylT(CUA)-PylRS-sfGFP134TAA which carries $tRNA^{Pyl}_{CUA}$ and sfGFP134TAA were transformed into *E. coli* BL21(DE3) cells. In the presence of BocK, the transformed cell did not lead to

detectable expression of sfGFP, proving that a CUA anticodon does not cross recognize a UAA codon and translation termination at a UAA codon is stringent. A similar experiment was tested with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG. Cells transformed with this plasmid did show sfGFP expression. However, in comparison to expression of sfGFP whose nonsense mutation was suppressed by its corresponding suppressor, the expression level was much lower (Fig. 6).

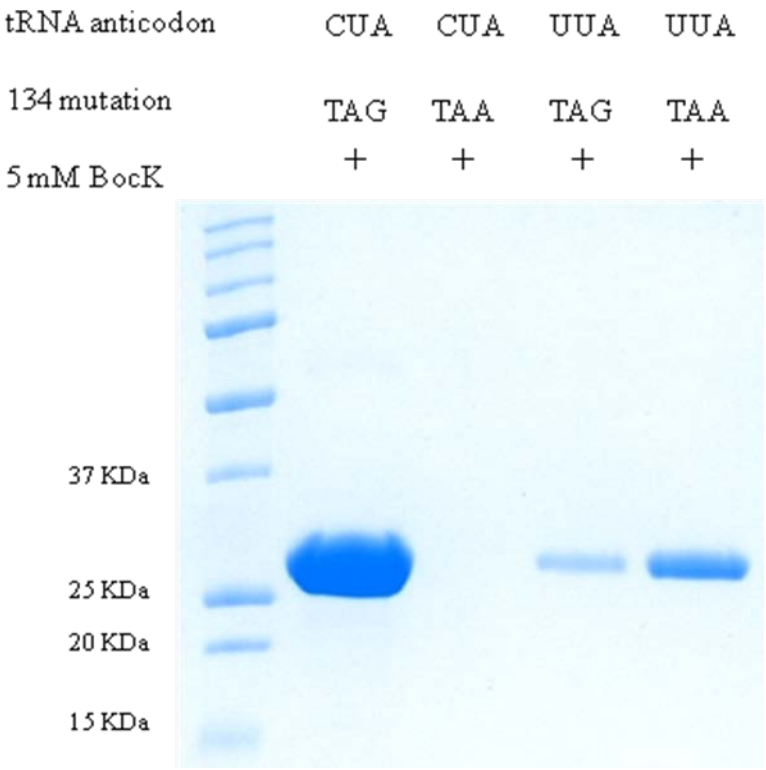


Figure 6. **Cross recognitions between different anticodons of tRNA^{Pyl} and nonsense mutations at N134 of sfGFP**. Cells were transformed with pETtrio-PylT(NNN)-PylRS-sfGFP134N'N'N' and grown in the presence of 5 mM BocK (NNN and N'N'N' denote anticodon and codon specified in the figure).

To see whether BocK promoted suppression of the amber mutation at N134 position of sfGFP by the PylRS-$tRNA^{Pyl}_{UUA}$ pair, cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG was also grown in 2YT medium without BocK supplemented. As shown in Fig. 7A, the sfGFP expression levels both in the absence and in the presence of BocK were very similar. Addition of BocK did not lead to significant increase of amber suppression. We suspected expression of sfGFP in both conditions arose mainly from basal amber suppression in *E. coli*.

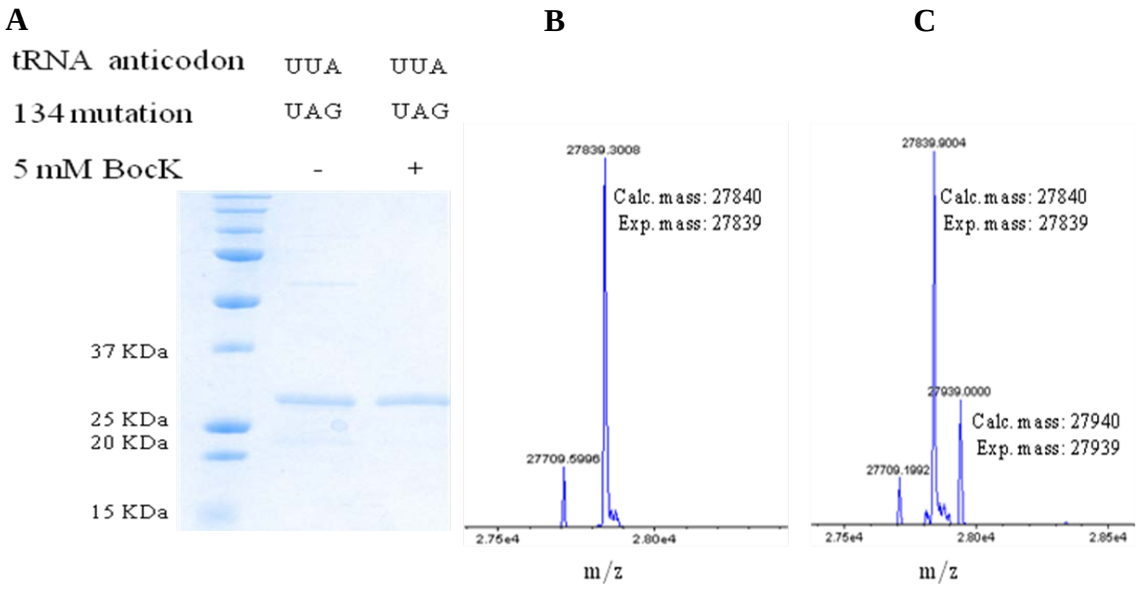


Figure 7. **Recognition efficiency between the UAG codon by a UUA anticodon.** (A) An amber (UAG) codon is not efficiently recognized by a UUA anticodon. Expression of sfGFP in cells transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAG and grown in 2YT medium supplemented without or with BocK. (B) ESI-MS of sfGFP expressed in the absence of BocK. (C) ESI-MS of sfGFP expressed in the presence of BocK.

This was confirmed by the ESI-MS analysis of purified sfGFP. As shown in Figs. 7B and C, sfGFP expressed in both conditions displayed a mass peak (27,838 Da in Fig. 7B and 27,839 Da in Fig. 7C)

An alternative experiment was also carried out to demonstrate that the PylRS-$tRNA^{Pyl}_{UUA}$ pair does not interfere with suppression of an amber mutation mediated by an evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pair. Two plasmids, pETtrio-pylT(UUA)-pylRS-sfGFP134TAG and pEVOL-AzFRS were used to transform cells. pEVOL-AzFRS codes $tRNA^{Tyr}_{CUA}$ and an evolved *Mj*TyrRS, AzFRS that is specific for AzF. As shown in Fig. 8, growing the transformed cells in four conditions led to different expression levels of full-length sfGFP. When no NAA or only BocK was provided in the medium, only a basal level of sfGFP expression was detected, verifying that the PylRS-$tRNA^{Pyl}_{UUA}$ pair does not mediate strong amber suppression. However, addition of AzF to the medium supplemented with or without BocK promoted sfGFP overexpression.

The ESI-MS analysis of purified sfGFP in all four conditions displayed very interesting spectra. Although a mass peak at 27,899 or 27,900 Da for sfGFP expressed in two conditions with the supplement of AzF (Figs. 9B and D) clearly matches the calculated molecular weight (27,901 Da) of sfGFP with AzF incorporated at N134, the mass peak at 27,858/27859 Da for sfGFP expressed in two conditions without the supplement of AzF did not match the molecular weight of sfGFP with either Lys/Glu/Gln or BocK incorporated at N134 (Figs. 9A and B). Instead, this mass peak

agrees well with the molecular weight of sfGFP with phenylalanine (Phe) incorporated at N134 (calculated mass: 27,859 Da).

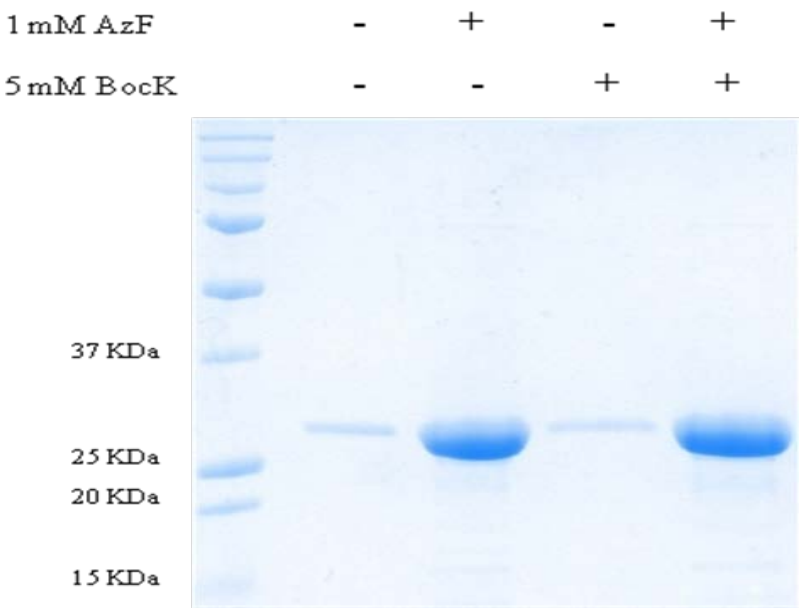


Figure 8. **Expression of sfGFP in cells transformed using two plasmid system.** Cells were transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAG and pEVOL-AzFRS and grown in 2YT medium supplemented with different NAAs.

Genetic incorporation of two different NAAs into one protein in E. coli

To genetically incorporate both AzF and BocK to sfGFP at S2 and N134 respectively, two plasmids pEVOL-AzFRS and pETtrio-pylT(UUA)-PylRS-sfGFP2TAG134TAA were used to transform *E. coli* BL21(DE3) cells. Growing cells transformed with these two plasmids in different conditions displayed different expression levels of sfGFP. Without any provided NAA, there was negligible sfGFP expression. When either AzF or BocK was supplemented, a very low level of full-length sfGFP was detected. However, adding both AzF and BocK into the medium induced

sfGFP overexpression (Fig. 10A). SfGFP expressed in the presence of both AzF and BocK showed a major mass peak at 28,066 Da that match the expected molecular weight of sfGFP with AzF and BocK incorporated at S2 and N134 respectively (calculated mass: 28,067 Da).

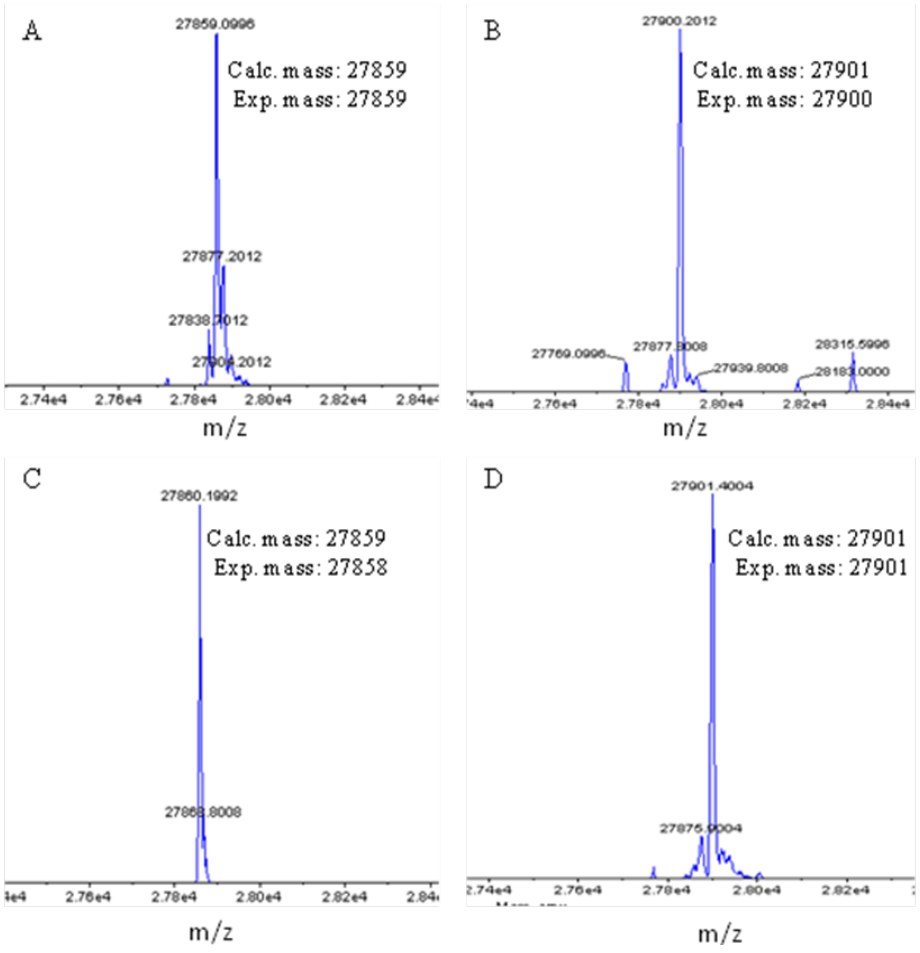


Figure 9. **ESI-MS of sfGFP expressed in cells transformed using two plasmid system.** (A) Cells containing pETtrio-PylT(UUA)-PylRS-sfGFP134TAG and pEVOL-AzFRS were grown in the absence of both AzF and BocK (B) The same cells were grown in the presence of AzF. (C) Cells were grown in the presence of BocK. (D) They were also grown in the presence of both AzF and BocK.

Opal suppression and orthogonality of tRNA$^{Pyl}_{UCA}$

Our initial study using cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA showed that sfGFP expressed in the presence or the absence of BocK mainly contained Trp at N134. To determine whether the misrecognition of the UGA codon at N134 by the endogenous Trp-tRNA^{Trp} was caused by the nucleotide contents around it, we constructed plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA for analysis. Cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP2TGA expressed sfGFP in the presence or the absence of BocK. Expression levels in both conditions were much higher than those for plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134TGA (Fig. 11).

The ESI-MS analyses of purified sfGFP from cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP2TGA displayed similar patterns as sfGFP from cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA. Mass peaks at 27,795 Da and 27,926/27,927 Da for sfGFP expressed in the absence or the presence of BocK are shown in Figs. 11B and C. In the absence of BocK the molecular weights of sfGFP matches Trp incorporated at S2 (calculated mass: 27,926 Da for the full-length protein, 27,794 Da for the full-length protein without M1). For sfGFP expressed in the presence of BocK, there are two peaks at 27,967 Da and 27,834 Da that agree well with the molecular weights of sfGFP with BocK incorporated at S2 (calculated mass: 27,967 Da for the full-length protein, 27,836 Da for the full-length protein without M1). It is obvious that nucleotide contents around the opal mutation site do not specifically recruit Trp-tRNA^{Trp} but does have effects on the overall suppression level. To further understand opal suppression, plasmid pETtrio-sfGFP134TGA that carries no opal

suppressor tRNA was used to transform *E. coli* BL21(DE3) cells. The transformed cells displayed a detectable level of sfGFP expression, though sfGFP resulted from this basal opal suppression had a much lower expression yield than sfGFP expressed in cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA (Fig. 12).

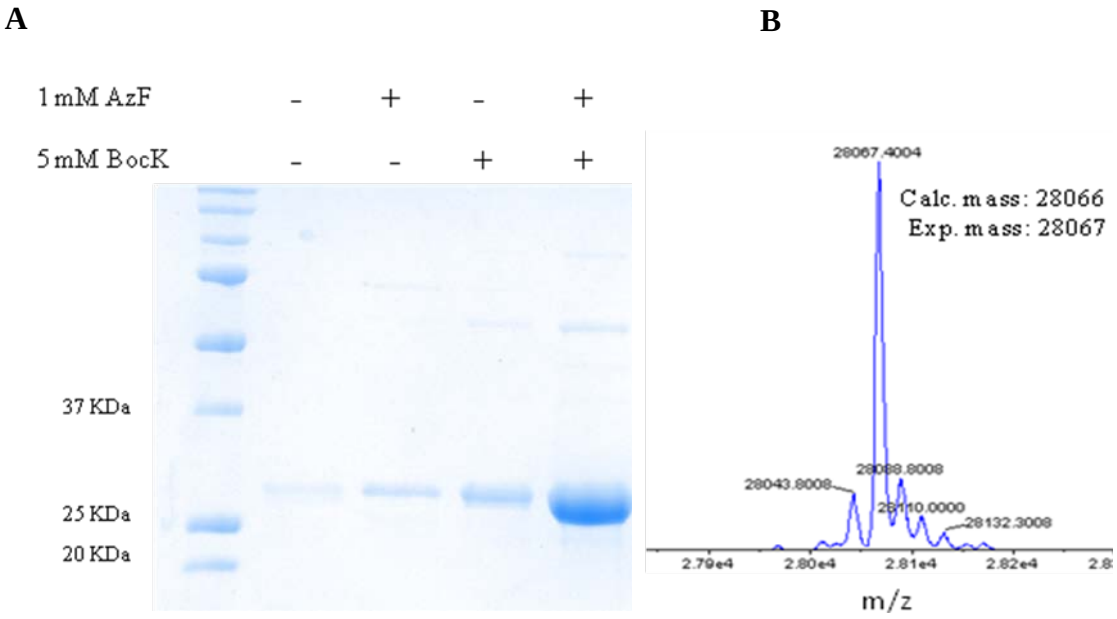


Figure 10. **Expression of sfGFP in cells using two stop codons .** (A) Cells were transformed with pETtrio-PylT(UUA)-PylRS-sfGFP2TAG134TAA and pEVOL-AzFRS and grown in 2YT medium supplemented with different NAAs. (B) ESI-MS of sfGFP expressed in the presence of both AzF and BocK.

The ESI-MS analysis of sfGFP resulted from the basal opal suppression showed a major mass peak at 27,899 Da that matches the molecular weight of sfGFP with Trp incorporated at N134.

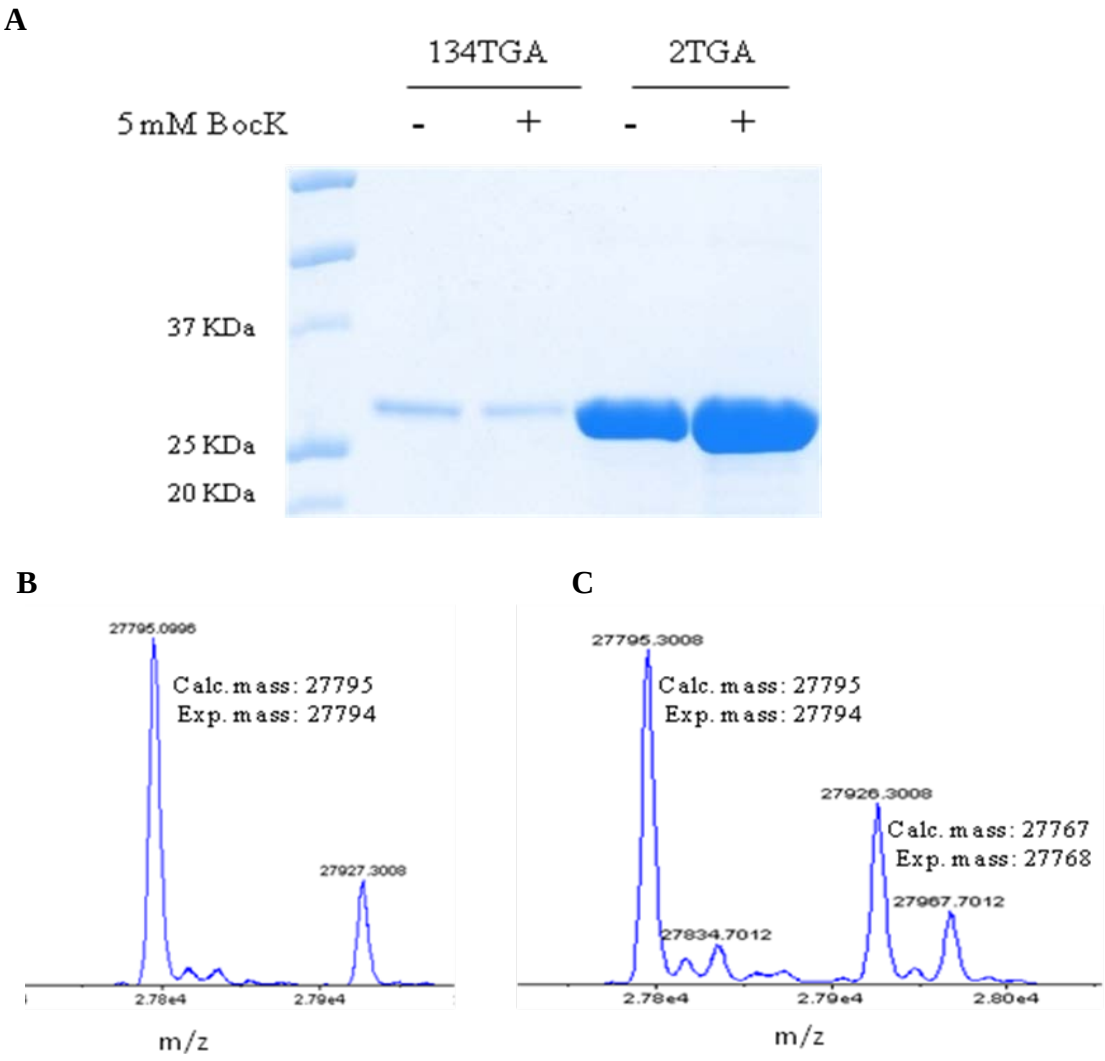


Figure 11. **Opal suppression at 2nd position of sfGFP using the PylRS-tRNA$^{Pyl}_{UCA}$ pair.** (A) Expression of sfGFP with an opal mutation. The first two lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)-sfGFP134TGA and grown in the absence or presence of BocK. The last two lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)-sfGFP2TGA and grown in the absence or presence of BocK. ESI-MS of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)-sfGFP2TGA and grown in the (B)absence or (C) presence of BocK.

In order to increase the orthogonality of $tRNA^{Pyl}_{UCA}$ in *E. coli*, we carried out mutagenic studies of this tRNA. As shown in Fig. 13, *E. coli* $tRNA^{Trp}$ has four strong recognition elements for its binding to TrpRS, three of these recognition elements C35, A36, and G73 exist in $tRNA^{Pyl}_{UCA}$. Since C35 and A36 are part of the anticodon, we focused on the mutagenic studies of G73.

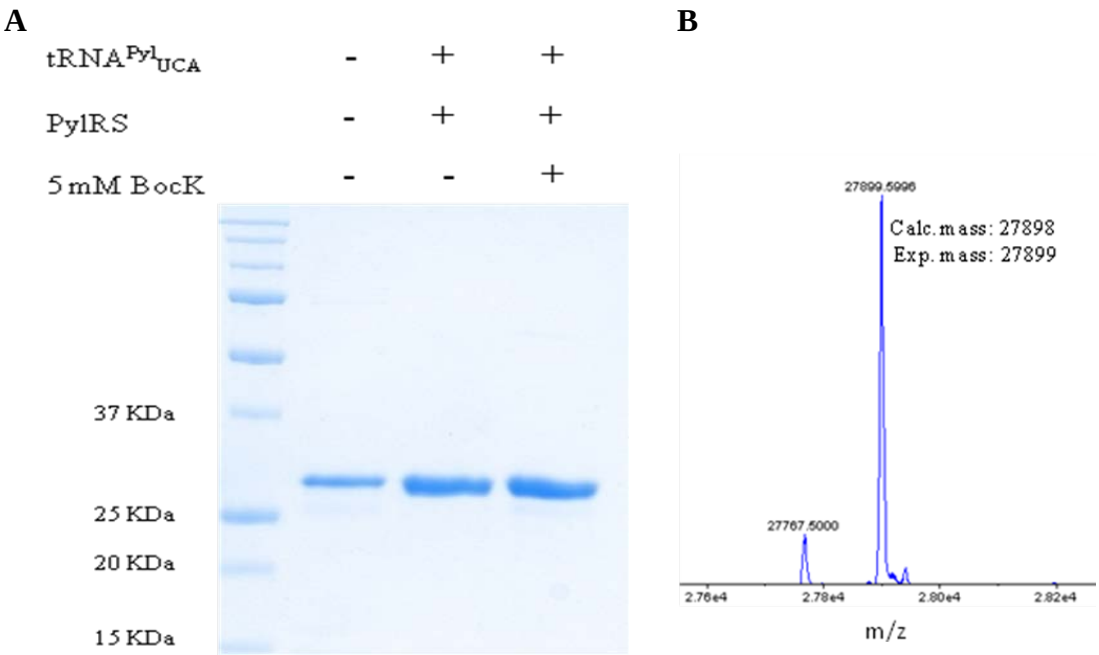
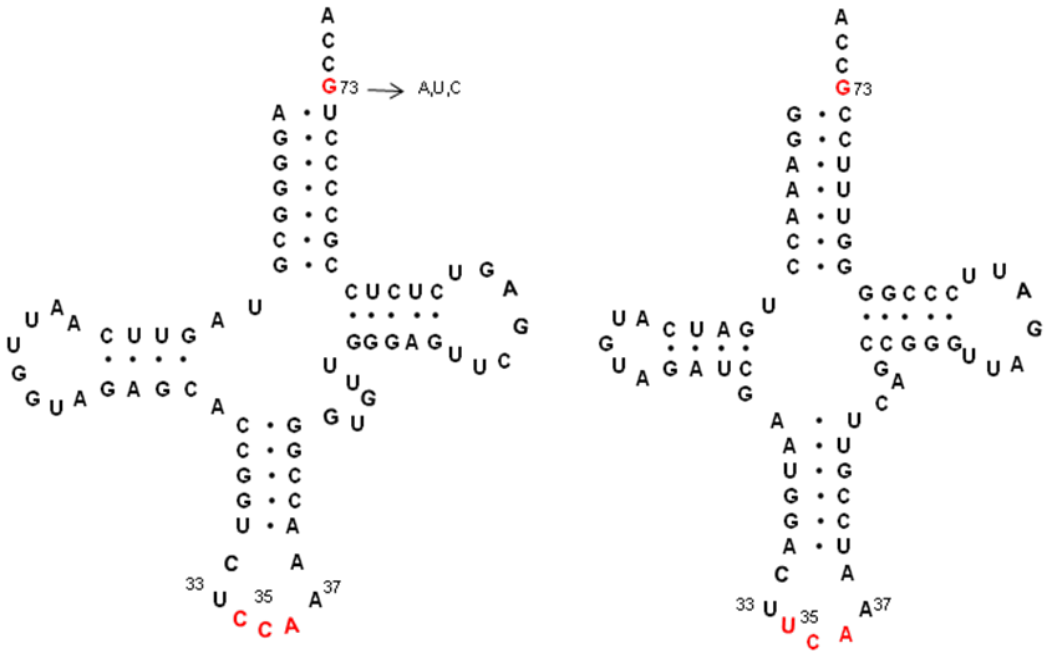


Figure 12. **Suppression of an opal mutation at N134 of sfGFP at different conditions.** (A) The first lane shows expression of sfGFP in cells transformed with pETtrio-sfGFP134TGA. The second and third lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA and grown in the absence or presence of BocK. (B) ESI-MS of sfGFP expressed in cells transformed with pETtrio-sfGFP134TGA.

Three mutations, G73A, G73C, and G73U of $tRNA^{Pyl}_{UCA}$ were tested. All the three transformed cells, still expressed sfGFP without the addition of BocK (Fig. 14B). The ESI-MS analysis of sfGFP expressed in the absence of BocK showed a major mass peak at 27,896 Da that matches sfGFP with Trp incorporated at N134. Cells transformed with pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA and cells transformed with pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA both displayed high expression levels of sfGFP in both the presence and the absence of BocK. Apparently G73A and G73C mutations did not increase the orthogonality of $tRNA^{Pyl}_{UCA}$ in *E. coli* (Fig. 14).



E. coli $tRNA^{Trp}$ *M. mazei* $tRNA^{Pyl}_{UCA}$

Figure 13. **Secondary structures of *E. coli* $tRNA^{Trp}$ and *M. mazei* $tRNA^{Pyl}_{UCA}$.** Recognition elements of $tRNA^{Trp}$ for *E. coli* TrpRS are highlighted in red. Mutation of G73 to A, U, and C in $tRNA^{Pyl}_{UCA}$ is denoted.

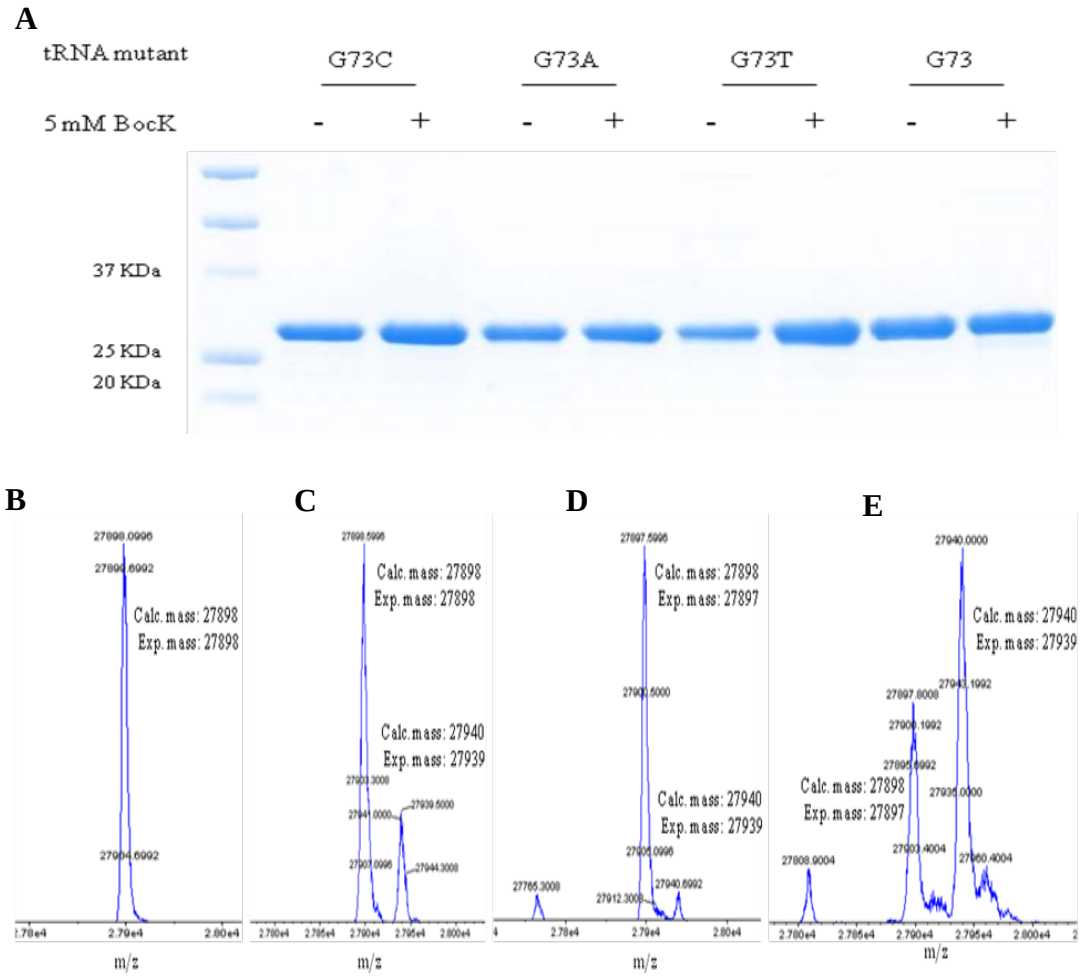


Figure 14. **Effects of the G73 mutation of $tRNA^{Pyl}_{UCA}$ on *E. coli* TrpRS in the absence and presence of BocK.** (A) Suppression of an opal mutation at N134 of sfGFP by different $tRNA^{Pyl}_{UCA}$ variants. The first 2 lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)G73C-sfGFP134TGA. The 3rd and 4th lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)G73A-sfGFP134TGA. The 5th and 6th lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)G73U-sfGFP134TGA. The last 2 lanes show expression of sfGFP in cells transformed with pETtrio-pylT(UCA)-sfGFP134TGA. (B) The ESI-MS analysis of sfGFP expressed in cells transformed with any of the mutants and grown in the absence of BocK. (C) The ESI-MS analysis of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)G73C-sfGFP134TGA and grown in the presence of BocK. (D) The ESI-MS analysis of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)G73A-sfGFP134TGA and grown in the presence of BocK. (E) The ESI-MS analysis of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)G73U-sfGFP134TGA and grown in the presence of BocK.

However, cells transformed with pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA showed significantly different expression levels of sfGFP when growing cells in the presence of BocK.. The addition of BocK promoted the sfGFP expression level to increase. The ESI-MS analysis of the purified sfGFP confirms that the BocK incorporated at N134 became dominant (Fig. 14D). The intensity of the mass peak at 27,939 Da that matches sfGFP with BocK incorporated at N134 is roughly twice of that of the mass peak at 27,897 Da that matches sfGFP with Trp incorporated at N134.

Quadruple suppression efficiency

Given that PylRS does not specifically recognize $tRNA^{Pyl}$ as its anticodon, we suspected that the pair may be engineered to recognize an AGGA quadruple codon. Cells transformed with pETtrio-pylT(UCCU)-PylRS-sfGFP134AGGA showed high expression levels of sfGFP both in the presence and in the absence of BocK, proving that $tRNA^{Pyl}_{UCCU}$ is not orthogonal in *E. coli* (Fig. 15). The ESI-MS analysis of purified sfGFP expressed in both conditions showed major mass peaks that match sfGFP with arginine (Arg) incorporated at N134. With Arg at N134, the theoretic molecular weights of full-length sfGFP and full-length sfGFP without M1 are 27,868 Da and 27,737 Da, respectively. Corresponding peaks can be found in spectra of sfGFP expressed in the absence of BocK (27,735 and 27,867 Da) and sfGFP expressed in the presence of BocK (27,735 and 27,867 Da). Addition of BocK to the medium did not change the ESI-MS patterns of the expressed protein. No sfGFP with BocK at N134 was observed. These data not only demonstrate that $tRNA^{Pyl}_{UCCU}$ is aminoacylated by arginyl-tRNA synthetase

(ArgRS), thus not orthogonal in *E. coli* but it also imply that PylRS is not able to charge $tRNA^{Pyl}_{UCCU}$ with BocK.

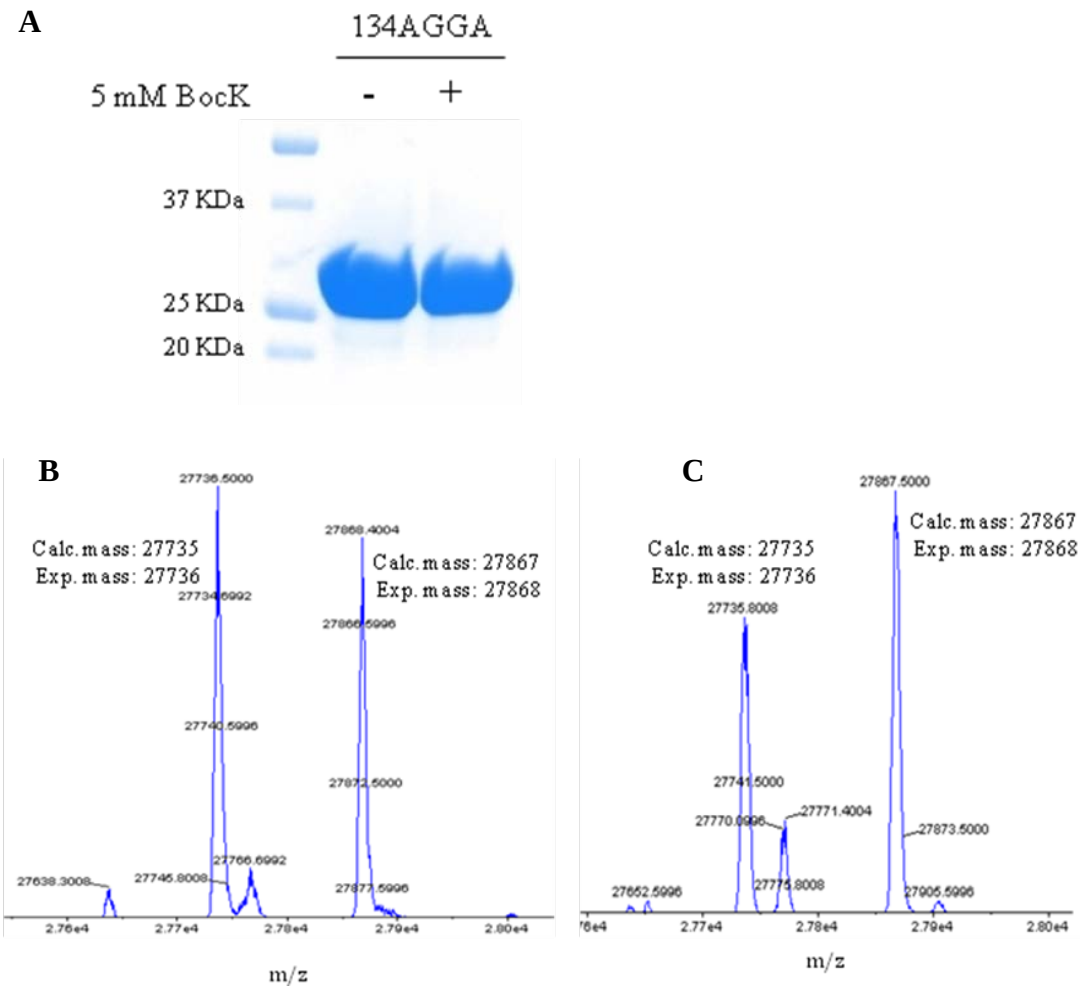


Figure 15. **Orthogonality of $tRNA^{Pyl}_{UCCU}$.** (A) Expression of sfGFP in cells transformed with pETtrio-pylT(UCCU)-sfGFP134AGGA and grown in the presence or absence of BocK. **(**B) ESI-MS of sfGFP expressed in the absence of BocK. (C) ESI-MS of sfGFP expressed in the presence of BocK.

4. DISCUSSION AND SUMMARY

Translation termination at stop codons

The termination of protein synthesis occurs when one of three stop codons enters the ribosome A site and is recognized by a specialized protein factor called release factor (RF). In *E. coli*, there are two RFs; RF-1 recognizes UAG and UAA codons, whereas RF-2 recognizes UGA and UAA codons. Given that precise termination at a stop codon is critical to maintain the proper structure and function of a synthesized protein, the faithful translation termination has been considered an indispensible property of the ribosome and essential for cells. It is also a general practice that proteins recombinantly expressed in *E. coli* have correct amino acid sequences because of the stringent translation termination at stop codons.

Impaired termination at stop codons was previously observed in *E. coli* that had defects in their ribosomes or RFs, coded suppressor tRNAs, or were treated with antibiotics such as streptomycin. (23) However, high basal levels of nonsense suppression in commonly used *E. coli* K12 and B strains have not been reported as far as we are aware. We initially suspected that *E. coli* Top10 cells transformed with pBAD-sfGFP134TAG would show a strict translation termination at the amber mutation at N134 of sfGFP. However, the transformed cells expressed sfGFP whose fluorescence was visible after the cells were harvested. This basal amber suppression level is possibly due to wobble base pairing interactions between the UAG codon and an endogenous tRNA. The ESI-MS analysis of the purified protein indicates a Lys/Glu/Gln residue at

the amber mutation site. Among $tRNA^{Lys}$, $tRNA^{Glu}$, and $tRNA^{Gln}$ that are all near-cognate tRNAs of UAG in *E. coli*, $tRNA^{Gln}$ has an CUG anticodon that can possibly form two Watson-Crick base pairs with the middle and 3' nucleotides of UAG and one wobble UG base pair with 5' nucleotide of UAG (*24-26*). Its interactions with UAG are theoretically stronger than the other two tRNAs. For this reason, we believe the misincorporated amino acid at N134 of sfGFP is Gln. Since *E. coli* cells grown in minimal media usually display undetectable amber suppression (*27*), the following reasons may explain the observed relatively high basal amber suppression level in cells grown in 2YT medium. In comparison to *E. coli* grown in minimal media, *E. coli* cells grown in 2YT medium have a much faster protein translation rate partly due to high concentrations of aminoacyl-tRNAs. High concentration of aminoacyl-tRNAs and a relatively low free RF-1 concentration due to the more active engagement of RF1 in the protein translation process may contribute to the detected high basal amber suppression level.

A similar test with *E. coli* BL21(DE3) cells transformed with pETtrio-sfGFP134TGA clearly showed that translation termination at an opal UGA codon is not strict either. SfGFP expressed in this condition has a higher expression yield than that resulted from basal amber suppression discussed above. The ESI-MS analysis of the purified protein clearly indicates a Trp residue at N134 of sfGFP. *E. coli* $tRNA^{Trp}$ with a CCA anticodon is a near-cognate tRNA of UGA. It forms two Watson-Crick base pairs with the 5' and middle nucleotides of UGA and a wobble base pair with the 3' nucleotide of UGA, explaining its recruitment to recognize UGA. We think this

relatively high basal level of opal suppression is related to the highly regulated RF-2 expression. The expression of RF-2 is modulated by a mistranslation-regulated feedback loop. When there is little RF-2, a programmed frameshifting event occurs on a UGA codon in the prfB gene transcript to allow production of RF-2. In the presence of sufficient RF-2, translation termination occurs at the premature UGA codon (*28-29*). Thus, RF-2 is maintained at a level to keep the balance of frameshift and translation termination events. We think this regulated level of RF-2 also leaves UGA susceptible for misrecognition by $tRNA^{Trp}$. The observed relatively high readthrough at UGA may also explain why *E. coli* chooses this codon to code selenocysteine (*30*).

Although we did not test ochre suppression specifically, the sfGFP expression analysis with cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAA (Fig. 5) showed an undetectable basal ochre suppression level. This can be explained from several aspects. UAG and UGA are recognized by RF-1 and RF-2, respectively, whereas UAA is recognized by both RF-1 and RF-2. Its recognition by both RF proteins in theory makes the translation termination at UAA more stringent than the other two stop codons. Another reason lies at the nucleotide contents of UAA. Not like UAG and UGA that could involve a GC base pair interaction, UAA could only form AU pairs or wobble pairs. Its interactions with tRNAs are relatively weak, making its misrecognition less possible than UAG and UGA.

Although basal amber suppression and basal opal suppression were both observed in *E. coli* grown in 2YT medium, the readthrough is still low. On the contrary to the high expression yield of wild type sfGFP (776 mg/L), sfGFP resulted from both basal amber

suppression and basal opal suppression did not exceed 1 mg/L. Readthrough of these codons is below 0.02 % of a sense codon. As demonstrated in the results section, this basal readthrough is effectively suppressed when a suppressor tRNA is genetically encoded and does not impair using a nonsense codon to code a NAA. Although this small readthrough at an amber or opal codon may not significantly affect using recombinantly expressed proteins for biochemistry studies, we still recommend using UAA as a stop codon to terminate protein translation for regular recombinant protein expression in *E. coli* in the future.

Orthogonality of tRNA^{Pyl} variants

Although the PylRS-tRNA^{Pyl} pair has been used extensively for the genetic incorporation of NAAs in *E. coli*, its orthogonality has not been carefully investigated. *E. coli* Top10 cells transformed with both pEVOL-pylT and pBAD-sfGFP134TAG led to a sfGFP expression level higher than cells transformed with only pBAD-sfGFP134TAG. It suggests that $tRNA_{CUA}^{Pyl}$ is charged by certain endogenous aaRS(s). The ESI-MS analysis of the purified sfGFP showed a Lys/Glu/Gln residue incorporated at the amber mutation site. Given that several critical identity elements of tRNA^{Gln} such as G73, C34, U35, and A37 also exist in $tRNA_{CUA}^{Pyl}$, $tRNA_{CUA}^{Pyl}$ is probably aminoacylated by *E. coli* glutaminyl-tRNA synthetase (*31-32*), leading to an amber suppression level higher than basal amber suppression. One thing that needs to be pointed out is that this aminoacylation level of $tRNA_{CUA}^{Pyl}$ catalyzed by an endogenous aaRS is relatively low. In addition, pEVOL-pylT has a very strong promoter for $tRNA_{CUA}^{Pyl}$ (*18*). Cells harboring

pEVOL-pylT has highly transcribed $tRNA^{Pyl}_{CUA}$ that also contributes to the increased amber suppression level.

Since $tRNA^{Pyl}_{CUA}$ is not fully orthogonal in *E. coli*, its recognition by an endogenous aaRS is relatively weak and only induces a suppression level that is a little higher than basal amber suppression. The existence of PylRS and its substrate BocK not only significantly improves the amber suppression level but also effectively prevent the recognition of $tRNA^{Pyl}_{CUA}$ by an endogenous aaRS. As shown in Fig. 5A, sfGFP expressed in cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG in the presence of BocK only displayed mass peaks for the protein isoforms with BocK incorporated at N134. Thus, $tRNA^{Pyl}_{CUA}$ can still be applied for the genetic incorporation of NAAs despite its low recognition by an endogenous aaRS.

Given that cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA did not show a detectable expression level of sfGFP in the absence of BocK, we can conclude that $tRNA^{Pyl}_{UUA}$ is fully orthogonal in *E. coli*. In comparison to sfGFP expressed in cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG and grown in the presence of BocK, the sfGFP expression level in cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA and grown in the presence of BocK is lower (Fig. 4). We suspect this is due to the relative weak interactions between the UUA anticodon of $tRNA^{Pyl}_{UUA}$ and a UAA stop codon, which form three two-hydrogen bonding base pairs, and the availability of both RF-1 and RF-2 to stop the translation at a UAA stop codon.

In any case, the ochre suppression level achieved by the PylRS-$tRNA^{Pyl}_{UUA}$ pair is sufficient to promote overexpression of a protein with an ochre mutation.

The efficient recognition of $tRNA^{Pyl}_{UCA}$ by TrpRS may be due to the following reasons. $tRNA^{Pyl}_{UCA}$ and *E. coli* $tRNA^{Trp}$ are structurally similar. Both have a short variable arm and very similar anticodon loops made up of heptanucleotide CUU/CCAAA. Three of four strong recognition elements, G73, C35, and A36 that are critical for $tRNA^{Trp}$ binding to TrpRS exist in $tRNA^{Pyl}_{UCA}$. Since $tRNA^{Pyl}_{UCA}$ is effectively $tRNA^{Trp}_{UCA}$ in *E. coli*, it cannot be applied for the genetic incorporation of NAAs at opal mutation sites. Attempts to alter the recognition of $tRNA^{Pyl}_{UCA}$ by TrpRS led to the discovery of $tRNA^{Pyl}_{UCA}$(G73U) that shows a higher preference to be charged by PylRS than TrpRS. However, $tRNA^{Pyl}_{UCA}$(G73U) is still not orthogonal enough. To increase its orthogonality, further modifications to $tRNA^{Pyl}_{UCA}$(G73U) to diminish its binding to TrpRS is necessary.

Wobble base pair hypothesis

Wobble base pair hypothesis was first introduced by Francis Crick in 1966 to explain the observation that a single tRNA is able to efficiently recognize multiple codons (*19*),(*33*). Based on this hypothesis, an ochre suppressor $tRNA_{UUA}$ is also able to recognize an amber UAG codon. This is a serious concern when both UAG and UAA codons are used to code two different NAAs. However, attempts to use the PylRS-$tRNA^{Pyl}_{UUA}$ pair to deliver BocK at an UAG codon was not successful. Amber suppression

induced by the PylRS- $tRNA^{Pyl}_{UUA}$ pair is even lower than basal amber suppression level in *E. coli* (Fig. 5D). This indicates very weak recognition of UAG by $tRNA^{Pyl}_{UUA}$ and it is in direct conflict with the wobble base pair hypothesis. Weak base pair interactions involved with the UUA anticodon may contribute to the weak recognition of UAG. However, this is definitely not the determining factor since other tRNAs such as $tRNA^{Lys}_{UUU}$ also involve weak base pair interactions to recognize multiple codons (*34-35*).

One explanation for this weak recognition of UAG by $tRNA^{Pyl}_{UUA}$ lies at the tRNA modifications. It has been found that all cognate tRNAs exhibit similar affinities for the ribosome A site when they bind to corresponding codons (36). This uniform binding is unexpected as certain codon-anticodon interactions are expected to be more stable than others due to factors such as the GC base pair content. It has been proposed that the specific sequence and post-transcriptional modification status of the tRNA in the region near the anticodon is tuned to ensure nearly indistinguishable binding of tRNAs to the ribosome A site. This has been the case for tRNAs such as $tRNA^{Lys}_{UUU}$ in which both nucleotides at 34 and 37 are post-transcriptionally modified to achieve similar recognitions of AAA and AAG codons (*33*),(*37*). Unlike endogenous tRNAs that have corresponding modification enzymes, $tRNA^{Pyl}_{UUA}$ is exogenous and may not be targeted by modification enzymes in *E. coli*. $tRNA^{Pyl}_{UUA}$ without modifications at its anticodon loop likely has a very weak binding affinity to the ribosome A site to associate UAG. Our finding points out that wobble base pairing at the 3' nucleotide of a codon is not

sufficient for recruiting a tRNA to the ribosome A-site. Additional interactions are required. This aspect needs to be further investigated. Since $\mathbf{tRNA}^{\mathbf{Pyl}}_{\mathbf{UUA}}$ does not recognize UAG, it is feasible to use an evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pair and a wild type or evolved PylRS-$tRNA^{Pyl}_{UUA}$ pair to code two different NAAs at amber and ochre mutation sites, respectively, in *E. coli*. As we demonstrated, sfGFP incorporated with AzF and BocK that were encoded by UAG and UAA codons could be overexpressed and displayed an expected molecular weight when analyzed by ESI-MS (Fig. 10).

Substrate specificity of an evolved NAA-specific aaRS

During analysis to see whether $\mathbf{tRNA}^{\mathbf{Pyl}}_{\mathbf{UUA}}$ can compete against $\mathbf{tRNA}^{\mathbf{Tyr}}_{\mathbf{CUA}}$ to associate UAG in the ribosome A site, we noticed that Phe was incorporated at N134 of sfGFP which was expressed in cells transformed with pEVOL-AzFRS and pETtrio-pylT(UUA)-PylRS-sfGFP134TAG and grown in either the absence or the presence of BocK. We believe AzFRS can recognize Phe, leading to the misincorporation of Phe. Since AzF is structurally very similar to Phe, one would expect that AzFRS that was originally evolved from *Mj*TyrRS will inevitably recognize Phe at a relatively low level and charge $tRNA^{Tyr}_{CUA}$ with Phe when AzF is absent in the medium. Although not clearly addressed in the existing literature, most evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pairs did show background amber suppression even in minimal media (*38*). Since most evolved *Mj*TyrRS variants are for Phe derivatives, we strongly believe background amber suppression caused by these evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pairs was due to their

recognition of either Phe or tyrosine or both. In this study, it is obvious that this background amber suppression induced by the AzFRS- $tRNA^{Tyr}_{CUA}$ pair inhibited both the basal amber suppression level and amber suppression induced by the PylRS- $tRNA^{Pyl}_{UUA}$ pair. This test provided evidence that amber suppression mediated by the PylRS- $tRNA^{Pyl}_{UUA}$ pair is too low to be of concern. It also points out that substrate specificities of evolved NAA-specific aaRSs need to be further characterized.

The PylRS-tRNA^{Pyl} interaction

PylRS from archaeal origins is composed of two domains, the C-terminal catalytic domain and the less soluble N-terminal domain. Although the PylRS N-terminal domain is not necessary for the aminoacylation of $tRNA^{Pyl}_{CUA}$ *in vitro*, it is indispensible *in vivo*. Söll et al. has showed that deletion of N-terminal residues beyond E30 from *M. barkeri* PylRS severely impaired it *in vivo* activity to charge $tRNA^{Pyl}_{CUA}$ for suppressing an amber mutation (*39*). Because of the insoluble nature of the N-terminal domain, all the current available PylRS structures are for the catalytic domain (40). Based on a recently determined crystal structure of the PylRS-$tRNA^{Pyl}$ complex from *Desulfitobacterium hafniense*,the major recognition sites at $tRNA^{Pyl}$ for the PylRS catalytic domain are located at the acceptor and D arms (*40*)-(*41*). The anticodon of $tRNA^{Pyl}$ does not involve direct interactions with the PylRS catalytic domain. Mutating the anticodon of $tRNA^{Pyl}$ from CUA to UCA and UUA in the current study does not affect its recognition by PylRS. Based on this observation and two previous studies that

showed PylRS was able to charge $tRNA^{Pyl}$ with mutated anticodons (*7*),(*32*), we can conclude that the anticodon of $tRNA^{Pyl}$ is not a recognition element. However, mutating the anticodon of $tRNA^{Pyl}$ to the UCCU quadruple anticodon totally disrupted its recognition by PylRS. A reasonable explanation to this observation is that $tRNA^{Pyl}_{UCCU}$ has an anticodon loop whose conformation is different from that in $tRNA^{Pyl}_{CUA}$, leading to abolishing its aminoacylation by PylRS. Since the anticodon loop does not directly interact with the PylRS catalytic domain, it is reasonable to believe it is engaged in the interaction with the PylRS N-terminal domain that is apparently critical for the PylRS-$tRNA^{Pyl}_{CUA}$ pair to maintain its *in vivo* activity. Because a quadruple anticodon disrupts the interaction between PylRS and $tRNA^{Pyl}$, engineering the PylRS N-terminal domain is necessary to generate a functional PylRS-$tRNA^{Pyl}_{UCCU}$ pair. In order to use $tRNA^{Pyl}_{UCCU}$ in *E. coli*, mutagenesis to eliminate its recognition and aminoacylation by ArgRS also needs to be carried out.

Summary

The orthogonalities of four $tRNA^{Pyl}$ variants in *E. coli* and the cross recognition of UAG codon by $tRNA^{Pyl}_{UUA}$ have been investigated. Among four $tRNA^{Pyl}$ variants, $tRNA^{Pyl}_{CUA}$ can be charged by an endogenous aaRS at a relatively low level, $tRNA^{Pyl}_{UUA}$ is orthogonal, and both $tRNA^{Pyl}_{UCA}$ and $tRNA^{Pyl}_{UCCU}$ are not orthogonal in *E. coli*. Although $tRNA^{Pyl}_{CUA}$ is not fully orthogonal in *E. coli*, its aminoacylation by an endogenous aaRS is at a very low level and can be effectively inhibited when both PylRS and its substrate are

present. Therefore, it is still reliable to use the PylRS-$tRNA^{Pyl}_{CUA}$ pair to genetically incorporate a NAA at an amber mutation site. Besides, its orthogonal nature in *E. coli*, $tRNA^{Pyl}_{UUA}$ does not induce suppression at an amber codon. This is in conflict with the wobble base pair hypothesis and makes it feasible to use an amber suppressing aaRS-tRNA pair and the PylRS-$tRNA^{Pyl}_{UUA}$ pair to code two different NAAs at amber and ochre codons respectively in *E. coli*. Since $tRNA^{Pyl}_{UCA}$ is efficiently charged by *E. coli* TrpRS, it cannot be used to deliver a NAA at an opal mutation site. Attempts to improve its orthogonality led to the discovery of $tRNA^{Pyl}_{UCA}$(G73U) that shows preferential aminoacylation by PylRS. However, further work is still necessary to totally eliminate its aminoacylation by TrpRS. Mutating the anticodon of $tRNA^{Pyl}_{CUA}$ to $tRNA^{Pyl}_{UCCU}$ not only led to the loss of the relative orthogonality of $tRNA^{Pyl}$ in *E. coli* but also abolished its aminoacylation by PylRS. In order to generate a functional PylRS-$tRNA^{Pyl}_{UCCU}$ pair for quadruple codon suppression, further engineering on both PylRS and $tRNA^{Pyl}_{UCCU}$ is required.

REFERENCES

1. Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002) A new UAG-encoded residue in the structure of a Methanogen Methyltransferase. *Science* **296**, 1462-1466

2. Atkins, J. F., and Gesteland, R. (2002) The 22nd amino acid. *Science* **296**, 1409-1410

3. Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459-1462

4. Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004) Direct charging of tRNACUA with pyrrolysine *in vitro* and *in vivo*. *Nature* **431**, 333-335

5. Baron, C., and Böck, A. (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *J. Biol. Chem.* **266**, 20375-20379

6. Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., Hatfield, D. L., and Berry, M. J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Reports* **1**, 158-163

7. 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.* **104**, 3141-3146

8. Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D., and Ambrogelly, A. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695-6700

9. 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

10. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode Nε-(o-Azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187-1197

11. Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J. W. (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Molec. Cell* **36**, 153-163

12. Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N[epsi]-acetyllysine in recombinant proteins. *Natl. Chem. Biol.* **4**, 232-234

13. Wang, Y.-S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W. K., Pai, P.-J., Moe, Y. N., Russell, D. H., and Liu, W. R. (2010) A genetically encoded photocaged N[varepsilon]-methyl-l-lysine. *Molec. BioSyst.* **6**, 1557-1560

14. Huang, Y., Wan, W., Russell, W. K., Pai, P.-J., Wang, Z., Russell, D. H., and Liu, W. (2010) Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg. Med. Chem. Lett.* **20**, 878-880

15. Wang, Y.-S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P.-J., Russell, D. H., and Liu, W. R. (2011) The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of l-phenylalanine and its derivatives, *Molec. BioSyst.* **7**, 714-717

16. Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. (2005) Protein posttranslational modifications: the chemistry of proteome ivdersifications, *Angew. Chem. Intl. Edition* **44**, 7342-7372

17. 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

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. Intl. Edition* **49**, 3211-3214

19. Crick, F. H. C. (1966) Codon—anticodon pairing: The wobble hypothesis, *J. Molec. Biol.* **19**, 548-555

20. Schwyzer, R., and Caviezel, M. (1971) p-Azido-L-phenylalanine: A photo-affinity 'probe' related to tyrosine, *Helv. Chim. Acta* **54**, 1395-1400

21. Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*, *J. Molec. Biol.* **395**, 361-374

22. McPheeters, D. S., Christensen, A., Young, E. T., Stormo, G., and Gold, L. (1986) Translationsl regulation of expression of the bacteriophage T4 lysozyme gene, *Nucleic Acids Res.* **14**, 5813-5826

23. Kisselev, L., Ehrenberg, M., and Frolova, L. (2003) Termination of translation: interplay of mRNA, rRNAs and release factors?, *EMBO J* **22**, 175-182

24. Swanson, R., Hoben, P., Sumner-Smith, M., Uemura, H., Watson, L., and Soll, D. (1988) Accuracy of in vivo aminoacylation requires proper balance of tRNA and aminoacyl-tRNA synthetase, *Science* **242**, 1548-1551

25. Giegé, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res.* **26**, 5017-5035

26. Fukunaga, J.-i., Ohno, S., Nishikawa, K., and Yokogawa, T. A base pair at the bottom of the anticodon stem is reciprocally preferred for discrimination of cognate tRNAs by *Escherichia coli* lysyl- and glutaminyl-tRNA synthetases, *Nucleic Acids Res.* **34**, 3181-3188

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

28. Caskey, C. T., Forrester, W. C., Tate, W., and Ward, C. D. (1984) Cloning of the *Escherichia coli* release factor 2 gene, *J. Bacteriology* **158**, 365-368

29. Bertram, G., Innes, S., Minella, O., Richardson, J. P., and Stansfield, I. (2001) Endless possibilities: translation termination and stop codon recognition, *Microbiol.* **147**, 255-269

30. Bock, A., and Stadtman, T. C. (1988) Selenocysteine, a highly specific component of certain enzymes, is incorporated by a UGA-directed co-translational mechanism, *BioFactors (Oxford, England)* **1**, 245-250

31. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase, *Nature* **352**, 213-218

32. Théobald-Dietrich, A., Frugier, M., Giegé, R., and Rudinger-Thirion, J. (2004) Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA, *Nucleic Acids Res.* **32**, 1091-1096

33. Johansson, M. J. O., Esberg, A., Huang, B., Björk, G. R., and Byström, A. S. (2008) Eukaryotic wobble uridine modifications promote a functionally redundant decoding system, *Molec. Cell. Biol.* **28**, 3301-3312

34. Fukunaga, J.-i., Yokogawa, T., Ohno, S., and Nishikawa, K. (2006) Misacylation of yeast amber suppressor tRNATyr by *E. coli* lysyl-tRNA synthetase and its effective repression by genetic engineering of the tRNA sequence, *J. Biochem.* **139**, 689-696

35. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992) In vitro study of E.coli tRNAArg and tRNALys identity elements, *Nucleic Acids Res.* **20**, 2335-2339

36. van der Gulik, P., and Hoff, W. (2011) Unassigned codons, nonsense suppression, and anticodon modifications in the evolution of the genetic code *J. Molec. Evolution* **73**, 59-69.

37. McClain, W. H., Foss, K., Jenkins, R. A., and Schneider, J. (1990) Nucleotides that determine *Escherichia coli* tRNA(Arg) and tRNA(Lys) acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs, *Proc. Natl. Acad. Sci.* **87**, 9260-9264

38. Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs, *Chem. Biol.* **8**, 883-890

39. Herring, S., Ambrogelly, A., Gundllapalli, S., O'Donoghue, P., Polycarpo, C. R., and Söll, D. (2007) The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable *in vitro* but required for *in vivo* activity, *FEBS Lett.* **581**, 3197-3203

40. 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. Molec. Biol.* **378,** 634-652

41. Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation, *Proc. Natl. Acad. Sci.* **104**, 11268-11273.

APPENDIX A

DNA sequence

pylT$_{CUA}$
ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

pylT$_{UCA}$
ggaaacctgatcatgtagatcgaatggacttcaaatccgttcagccgggttagattcccggggtttccgcca

pylT$_{UUA}$
ggaaacctgatcatgtagatcgaatggacttttaaatccgttcagccgggttagattcccggggtttccgcca

pylT$_{UCCU}$
ggaaacctgatcatgtagatcgaatggacttcctaatccgttcagccgggttagattcccggggtttccgcca

sfGFP134TAG
atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag
cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc
gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta
aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt
aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaatagggcaacattctgggtcataaact
ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca
acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat
aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgc
cgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

sfGFP134TGA
atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag
cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc
gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta
aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt
aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaatgaggcaacattctgggtcataaact
ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca
acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat
aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgc
cgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

sfGFP134TAA

atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag
cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc
gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta
aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt
aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaataaggcaacattctgggtcataaact
ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca
acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat
aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgc
cgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

sfGFP2TGA

atggtttgaaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaatttag
cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttcc
gtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttta
aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagtt
aaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaagatggcaacattctgggtcataaact
ggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcaca
acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccggat
aatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccgc
cgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

sfGFP2TAG134TAA

atggcataggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccata
aatttagcgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgc
cggttccgtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatga
tttctttaaaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcg
gaagttaaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaataaggcaacattctgggtca
taaactggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatcc
gtcacaacgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctg
ccggataatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgt
taccgccgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

Protein sequence

sfGFP134X (denotes the incorporation of the non canonical amino acid)*

mvskgeelftgvvpilveldgdvnghkfsvrgegegdatngkltlkficttgklpvpwptlvttltygvqcfsrypdhmkrh
dffksampegyvqertisfkddgtyktraevkfegdtlvnrielkgidfke*gnilghkleynfnshnvyitadkqkngika
nfkirhnvedgsvqladhyqqntpigdgpvllpdnhylstqsvlskdpnekrdhmvllefvtaagithgmdelykgshhh
hhh

sfGFP2TGA (denotes the incorporation of the non canonical amino acid)*

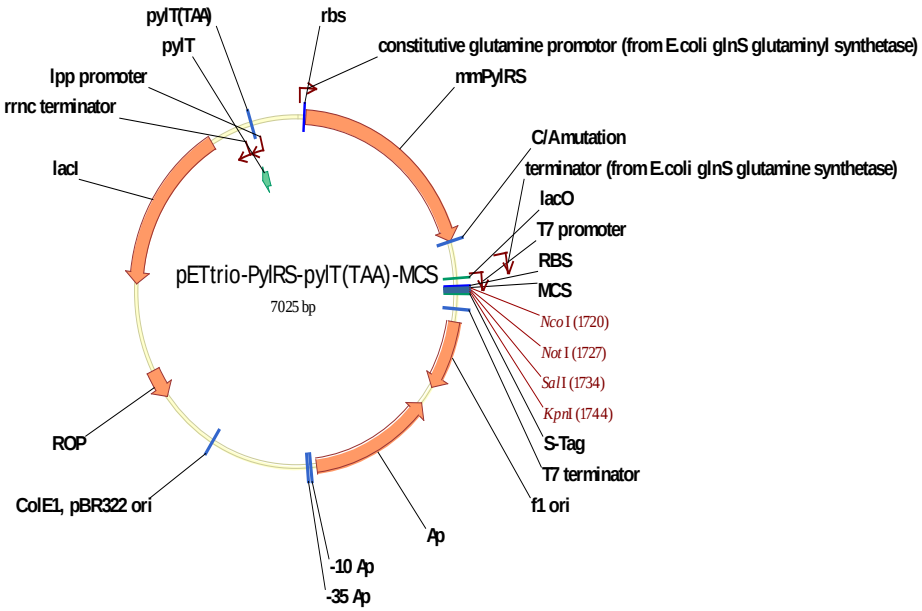
mv*kgeelftgvvpilveldgdvnghkfsvrgegegdatngkltlkficttgklpvpwptlvttltygvqcfsrypdhmkr
hdffksampegyvqertisfkddgtyktraevkfegdtlvnrielkgidfkedgnilghkleynfnshnvyitadkqkngik
anfkirhnvedgsvqladhyqqntpigdgpvllpdnhylstqsvlskdpnekrdhmvllefvtaagithgmdelykgshh
hhhh

msfGFP2TAG134TAA (, + denotes the incorporation of the non-canonical amino acids)*

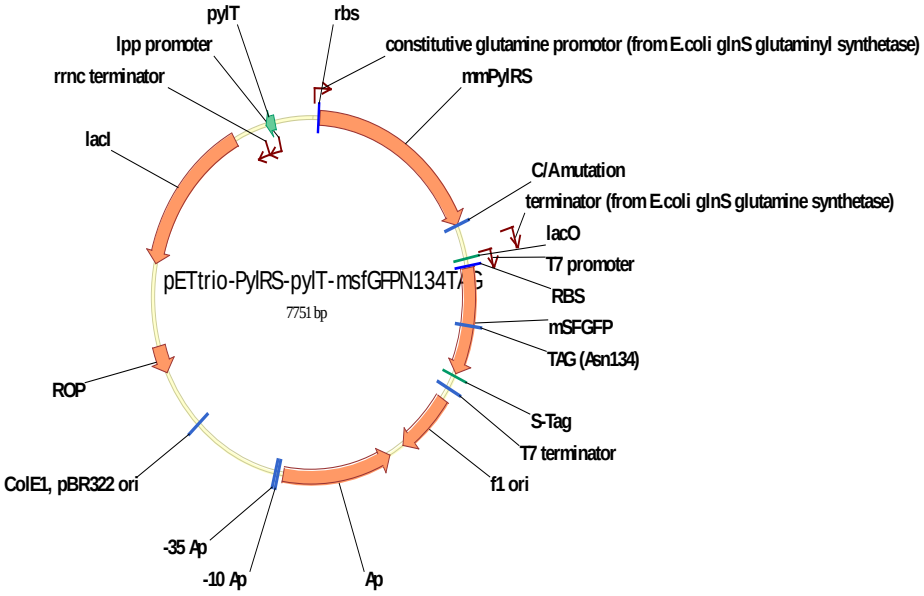
ma*vskgeelftgvvpilveldgdvnghkfsvrgegegdatngkltlkficttgklpvpwptlvttltygvqcfsrypdhm
krhdffksampegyvqertisfkddgtyktraevkfegdtlvnrielkgidfke+

Plasmid maps

pPylRS-pylT-MCS



pPylRS-pylT-sfGFP134TAG



VITA

Keturah Amarkie Odoi received her Bachelor of Science degree in chemistry from Southwestern Oklahoma State University in 2009. She entered the Chemistry program at Texas A&M University in September 2009 and received her Master of Science degree in May 2012.

Ms. Odoi may be reached at Chemistry Department, TAMU 3255, College Station, TX 77843. Her email is keturah.odoi@gmail.com.