

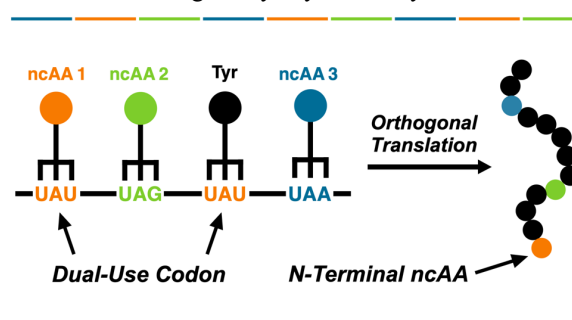
Genetic Encoding of Three Distinct Noncanonical Amino Acids Using Reprogrammed Initiator and Nonsense Codons

Jeffery M. Tharp,[†] Oscar Vargas-Rodriguez,[†] Alanna Schepartz,^{*,§,||} and Dieter Söll^{*,†,‡}

[†]Department of Molecular Biophysics & Biochemistry and [‡]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[§]Department of Chemistry and ^{||}Department of Molecular & Cell Biology, University of California, Berkeley, California 94705, United States

ABSTRACT: We recently described an orthogonal initiator tRNA (*itRNA*^{Ty2}) that can initiate protein synthesis with noncanonical amino acids (ncAAs) in response to the UAG nonsense codon. Here we report that a mutant of *itRNA*^{Ty2} (*itRNA*^{Ty2}_{AUA}) can efficiently initiate translation in response to the UAU tyrosine codon, giving rise to proteins with an ncAA at their N-terminus. We show that, in cells expressing *itRNA*^{Ty2}_{AUA}, UAU can function as a dual-use codon that selectively encodes ncAAs at the initiating position and tyrosine at elongating positions. Using *itRNA*^{Ty2}_{AUA}, in conjunction with its cognate tyrosyl-tRNA synthetase and two mutually orthogonal pyrrolysyl-tRNA synthetases, we demonstrate that UAU can be reassigned along with UAG or UAA to encode two distinct ncAAs in the same protein. Furthermore, by engineering the substrate specificity of one of the pyrrolysyl-tRNA synthetases, we developed a triply orthogonal system that enables simultaneous reassignment of UAU, UAG, and UAA to produce proteins containing three distinct ncAAs at precisely defined sites. To showcase the utility of this system, we produced proteins containing two or three ncAAs, with unique bioorthogonal functional groups, and demonstrate that these proteins can be separately modified with multiple fluorescent probes.



INTRODUCTION

Most lifeforms on Earth utilize a ‘universal’ genetic code in which 64 triplet codons encode the twenty canonical amino acids and three stop signals. In the laboratory, however, artificial genetic code expansion (GCE) has enabled the biosynthesis of proteins containing diverse noncanonical amino acids (ncAAs) at precisely defined sites. This methodology employs orthogonal aminoacyl-tRNA synthetase (o-aaRS) and tRNA (o-tRNA) pairs to co-translationally install ncAAs into proteins, typically in response to a redefined nonsense codon. Laboratory GCE is not limited to the incorporation of a single ncAA. Through the combined action of multiple, mutually orthogonal o-aaRS•o-tRNA pairs, proteins with multiple distinct ncAAs can be produced in living cells.¹⁻⁵ This technology has broad-ranging applications, from introducing reactive moieties into proteins for site-specific bioconjugation at multiple sites^{4, 6}, to producing homogeneously modified proteins featuring genetically encoded posttranslational modifications^{7, 8}. An exciting prospect of GCE is the ability one day to produce completely unnatural polypeptides with new-to-Nature functions. However, further expansion of the genetic code is limited by the lack of ‘blank’ codons to encode additional unnatural moieties.

Most commonly, GCE relies on stop codon suppression, wherein a nonsense codon (i.e., UAG, UAA, or UGA) is reas-

signed to encode an ncAA. Recently, mutually orthogonal o-aaRS•o-tRNA pairs were used to simultaneously reassign all three nonsense codons and direct site-specific installation of three distinct ncAAs for the first time *in vivo*.⁴ However, nonsense suppression can only provide three blank codons for GCE. Moreover, reassigning all three nonsense codons requires modified expression systems that assist in accurately terminating translation.⁴ Four-base codons such as UAGA or AGGA represent a promising alternative with the possibility of providing 256 unique blank codons.⁹ Indeed, four-base codons have been used alone, and in combination with nonsense codons, to encode two and three distinct ncAAs in the same protein gene.^{1, 5, 10} However, translation of four-base codons is very inefficient on wildtype ribosomes.^{3, 11, 12} It is likely for this reason that four-base codons have not been widely adopted. A third strategy to further expand the genetic code is to reassign one of the 61 sense codons that normally encode a canonical amino acid.^{13, 14} This strategy is particularly challenging due to competition with endogenous aminoacyl-tRNAs for suppression of sense codons.¹⁵ Because of this competition, sense codon reassignment most often affords a heterogeneous product containing a mixture of the canonical and noncanonical amino acids.^{14, 16-24}

In a recent study, we reported the development of an orthogonal initiator tRNA (*itRNA*^{Ty2}) that is a substrate for the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase

(*Mj*TyrRS)—an o-aaRS that is commonly used for GCE.²⁵ We demonstrated that *it*RNA^{Ty2} can initiate translation at UAG codons with a variety ncAAs. The unique ability of *it*RNA^{Ty2} to initiate translation is afforded by a conserved set of sequence motifs that are exclusive to initiator tRNAs.^{25, 26} Here, we asked whether *it*RNA^{Ty2} could be engineered to initiate translation at a reassigned sense codon. We hypothesized that endogenous elongator tRNAs, which lack structural motifs required for initiation, would be unable to compete with *it*RNA^{Ty2} for suppression of a reassigned initiator codon, thus abrogating a major hinderance to sense codon reassignment. We demonstrate that an anticodon mutant of *it*RNA^{Ty2} can efficiently initiate translation with ncAAs in response to the UAU tyrosine codon. Initiation at UAU was achieved without detectable tyrosine incorporation giving rise to proteins with an ncAA at the N-terminus. Using this mutant initiator tRNA, alongside two mutually orthogonal pyrrolysyl-tRNA synthetase (PylRS) and tRNA pairs, we demonstrate that UAU can be reassigned along with UAA and UAG to simultaneously encode two and three distinct ncAAs in the same protein. Finally, we demonstrate that proteins containing two and three reactive ncAAs can be separately modified with multiple fluorescent probes.

RESULTS AND DISCUSSION

Reassigning sense codons for translation initiation with noncanonical amino acids. To investigate whether anticodon mutants of *it*RNA^{Ty2} can initiate translation at reassigned sense codons, we employed a fluorescence-based assay that we previously used for measuring initiation at UAG.²⁵ This assay relies on a superfolder green fluorescent protein (sfGFP)²⁷ reporter in which the initiating AUG codon is replaced with UAG (sfGFP[1UAG]). The reporter is co-expressed with *it*RNA^{Ty2} and a polyspecific *Mj*TyrRS variant (pCNFRS²⁸ or AzFRS.2.t1²⁹). Under these conditions, sfGFP production relies on the ability of *Mj*TyrRS to charge *it*RNA^{Ty2} with an ncAA (provided in the growth media, **Figure 1**), and on the ability of aminoacyl-*it*RNA^{Ty2} to initiate translation at UAG. To modify this assay for measuring initiation at sense codons, we constructed a series of reporter plasmids in which the initiating codon of sfGFP was replaced with one of eight sense codons (**Figure 2A**). Because *Mj*TyrRS interacts with the anticodon of its cognate tRNA, we limited our search to codons that required minimal base substitutions in *it*RNA^{Ty2}.³⁰ First, to assess whether endogenous *E. coli* tRNAs can compete for initiation at these codons, we measured the expression of each sfGFP reporter in the absence of an *it*RNA^{Ty2} mutant. Without co-expression of a mutant initiator tRNA very little sfGFP expression was detected with each reporter (**Figure 2B**), supporting our assumption that endogenous *E. coli* tRNAs cannot initiate translation at these sense codons.

Next, we measured the expression of each sfGFP mutant in *E. coli* cells that were simultaneously expressing *it*RNA^{Ty2} (with complementary anticodon mutations) and pCNFRS. With four of the initiating sense codons (UAU, UUC, AAC, CAC) we observed significant sfGFP expression when the ncAA *para*-methyl-L-phenylalanine (pMeF) was added to the growth media (**Figures 1** and **2C**). This indicates that the *it*RNA^{Ty2} mutants were successfully charged with pMeF and that the aminoacyl-tRNA can initiate translation at the reassigned sense codon. Of the four codons that showed significant sfGFP expression with pMeF, the tyrosine codon UAU was most efficient, affording a fluorescence/OD₆₀₀ signal

~24% as intense as UAG initiation (**Figure 2C**). We measured sfGFP[1UAG] expression in wildtype *E. coli* DH10B, and in an engineered strain in which redundant copies of the methionine initiator tRNA gene were deleted from the genome (strain DH10BΔ*metZWW*). We have shown that initiation with ncAAs is far more efficient in DH10BΔ*metZWW* than in wildtype cells.²⁵ Consistent with our assumption that the mutant initiator tRNA (*it*RNA^{Ty2}_{AUA}) is initiating translation at UAU, sfGFP[1UAG] expression was nearly five-fold greater in DH10BΔ*metZWW* than in the wildtype strain (**Figure S1**).

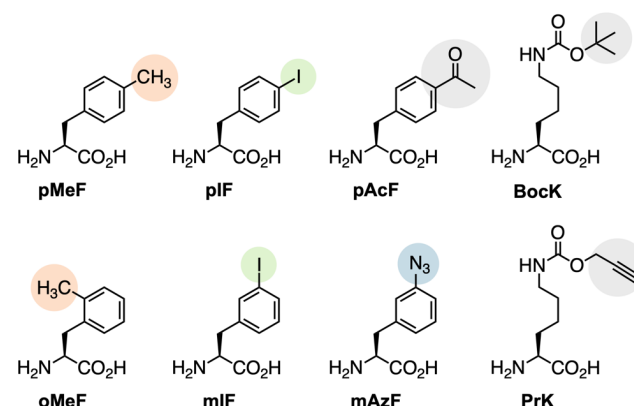


Figure 1. The structures of ncAAs used in this study.

To confirm pMeF incorporation in response to UAU, we expressed and purified sfGFP[1UAG] using pCNFRS and *it*RNA^{Ty2}_{AUA}. Consistent with in-cell fluorescence measurements, robust expression of the reporter only occurred in the presence of pMeF (**Figure 2D** and **S2A**). LC-MS of the purified protein revealed a major peak matching the expected mass of sfGFP with pMeF in place of the initiating methionine (sfGFP-1pMeF, **Figure 2E**). We also observed a peak corresponding to *N*-formyl-sfGFP-1pMeF (**Figure 2E**). We have shown that when ncAAs initiate translation in DH10B-Δ*metZWW* a significant fraction of sfGFP retains an *N*-terminal formyl modification.²⁵ This result further supports the conclusion that pMeF-charged *it*RNA^{Ty2}_{AUA} is initiating translation at UAU. Importantly, no peaks corresponding to tyrosine incorporation at the initiating UAU codon were detected, further demonstrating that the endogenous *E. coli* tyrosine tRNA (*Ect*RNA^{Tyr}) is unable to initiate translation at UAU.

In a previous study we showed that *it*RNA^{Ty2} can both initiate and elongate translation at UAG codons.²⁵ In addition to the initiating UAU, our sfGFP reporter has five elongating UAU codons; however, pMeF was not detected at these positions. This could be due to the low mass difference between pMeF and tyrosine (Δmass = 1.97 Da) which makes it difficult to distinguish between these residues based on mass alone. Therefore, we expressed sfGFP[1UAG] with *para*-iodo-L-phenylalanine (pIF; **Figure 1**) which has a larger mass than tyrosine (Δmass = 109.9 Da). Intact LC-MS of the purified protein revealed major peaks corresponding to sfGFP-1pIF and *N*-formyl-sfGFP-1pIF (**Figure S3**). No peaks corresponding to tyrosine incorporation at the initiating UAU were detected. Likewise, no peaks corresponding to pIF incorporation at elongating UAU codons were detected. We further analyzed this protein by tandem mass spectrometry (MS/MS) following proteolysis. In this analysis we identified peptides with masses and fragmentation patterns consistent with pIF incorporation at the *N*-terminal position (**Figure S4**); we were unable to

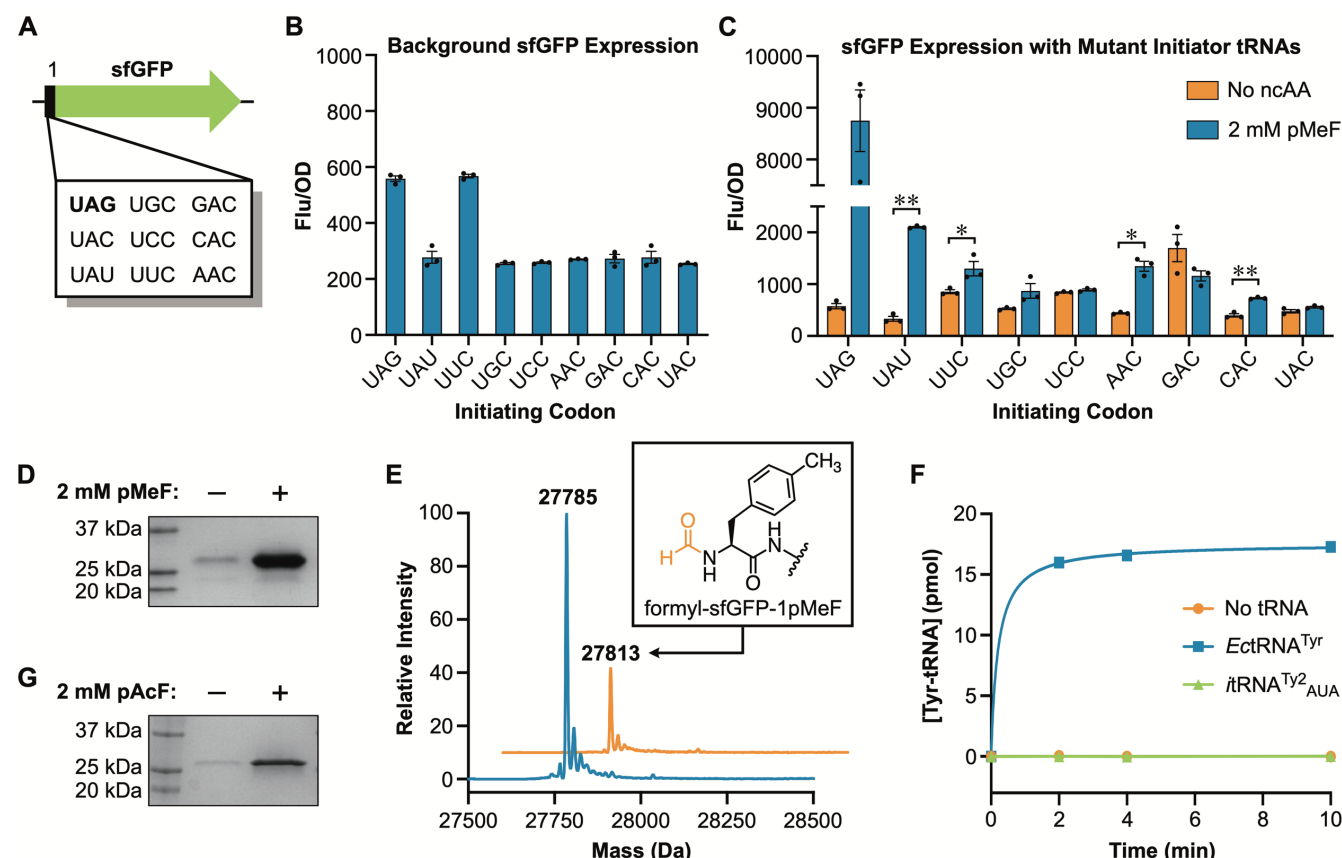


Figure 2. Reassigning sense codons for initiation with ncAAs. (A) A series of sfGFP reporters were constructed in which the initiating methionine codon was replaced with UAG or one of eight sense codons. (B) Endogenous *E. coli* tRNAs cannot initiate at UAG or any of the sense codons. Data are displayed as the mean ± SEM of three biological replicates. (C) Co-expression of a mutant initiator tRNA enables translation initiation with ncAAs at select codons. Data are displayed as the mean ± SEM of three biological replicates (* = $p < 0.05$, ** = $p < 0.005$, paired t-test). (D) Expression of sfGFP[1UAU] is dependent on addition of pMeF to the growth media. (E) LC-MS of D. Peaks corresponding to sfGFP-1pMeF (theoretical mass = 27786 Da) and N-formyl-sfGFP-1pMeF (theoretical mass = 27814 Da) were detected. (F) $itRNA^{Ty2}_{AUA}$ is not aminoacylated by the *E. coli* TyrRS. Timepoints represent the mean of three independent experiments. (G) ncAA-dependent expression of codon-optimized sfGFP in which all elongating UAU codons were replaced with UAC.

detect pIF incorporation in response to any elongating UAU codon. Thus, our data suggest that while $itRNA^{Ty2}_{AUA}$ can initiate translation with ncAAs at UAU, under these conditions, it is unable to compete with endogenous $EctRNA^{Tyr}$ for suppression of elongating UAU codons. Taken together, our results indicate that, in cells expressing $itRNA^{Ty2}_{AUA}$, UAU can function as a dual-use codon: at the initiating position UAU encodes an ncAA, whereas at elongating positions UAU encodes tyrosine. Similar dual-use codons have been reported *in vitro*³¹ and partial dual reassignment of the methionine AUG codon has been reported *in vivo*¹⁸; however, to our knowledge, this is the first time that complete reassignment of a dual-use codon has been achieved in living cells. It should be noted that this observation might be codon context dependent. That is, incorporation of ncAAs at elongating UAU codons cannot be ruled out for every protein at this time. To avoid possible complications from elongating UAU codons in subsequent experiments, we synthesized a codon-optimized sfGFP in which all elongating UAU codons were replaced with the synonymous codon UAC. As with our previous reporter, robust expression of the codon-optimized variant occurred only when an ncAA was provided in the growth media (Figure 2G and S2B). When this optimized gene was expressed with *para*-acetyl-L-phenylalanine (pAcF; Figure 1), sfGFP-1pAcF was obtained with a yield of ~260 mg per liter of culture.

Several of the codons that we tested showed an increase in sfGFP expression irrespective of whether an ncAA was provided in the growth media (Figure 2C). A possible explanation for this observation is that mutating the anticodon of $itRNA^{Ty2}$ converted the tRNA into a substrate for an endogenous *E. coli* aaRS. For example, we observed an increase in sfGFP expression when the anticodon of $itRNA^{Ty2}$ was mutated to GUC (for initiating at GAC). In *E. coli* GAC encodes aspartate and the aspartyl-tRNA synthetase uses anticodon bases to recognize its cognate tRNA.³² Similar mis-aminoacylation was observed when the anticodon of pyrrolysine tRNA ($tRNA^{Pyl}$) was mutated to reassign an arginine codon in *Mycoplasma capricolum*.¹⁶ While no tyrosine incorporation at the initiating UAU codon was detected, it is possible that $itRNA^{Ty2}_{AUA}$ is still being mis-charged with tyrosine by *E. coli* TyrRS ($EctTyrRS$). Proteins that contain an N-terminal tyrosine are extremely unstable in *E. coli*³³, thus, mis-charging of $itRNA^{Ty2}_{AUA}$ with tyrosine might go undetected if the reporter protein is rapidly degraded *in vivo*. To investigate this possibility, we performed *in vitro* aminoacylation assays using purified $EctTyrRS$ and an $itRNA^{Ty2}_{AUA}$ transcript. While $EctTyrRS$ charged its cognate tRNA with radiolabeled tyrosine, no aminoacylation of $itRNA^{Ty2}_{AUA}$ was detected (Figure 2F), indicating that the mutant tRNA is orthogonal to $EctTyrRS$.

Dual incorporation of noncanonical amino acids using reprogrammed initiator and nonsense codons. After demonstrating that *tRNA*^{Ty2}_{AUA} can initiate translation with ncAAs at UAU codons, we next asked whether UAU could be used together with nonsense codons to encode two distinct ncAAs in the same protein. The two most commonly used codons for GCE are UAG and UAA; however, *tRNA*^{Ty2}_{AUA} can base pair with these triplets at all but the wobble position. To encode distinct ncAAs using a combination of UAU and UAG/UAA, *tRNA*^{Ty2}_{AUA} must be able to recognize UAU and discriminate against UAG and UAA. Therefore, to evaluate whether *tRNA*^{Ty2}_{AUA} recognizes these nonsense codons, we constructed two reporters in which the second position of sfGFP was mutated to UAG or UAA (sfGFP[2UAG] and sfGFP[2UAA], respectively; **Figure S5A**). We co-expressed these mutant reporters, along with *tRNA*^{Ty2}_{AUA} and AzFRS.2.t1, in *E. coli* DH10BΔ*metZWW*, and we monitored sfGFP production in the presence of pMeF. As expected, sfGFP[1UAU] afforded a strong fluorescence signal, whereas no significant fluorescence was observed in cells expressing sfGFP[2UAG] or sfGFP[2UAA] (**Figure S5B**), demonstrating that *tRNA*^{Ty2}_{AUA} is orthogonal to these nonsense codons.

After confirming orthogonality of *tRNA*^{Ty2}_{AUA} towards UAG and UAA, we next explored whether these codons could be used, together with a UAU, to encode two different ncAAs. For introducing a second ncAA we conscripted the PylRS•*tRNA*^{Pyl} pair which is orthogonal in bacteria and eukaryotes and has been engineered to recognize numerous ncAA substrates.^{34, 35} The most widely used PylRS•*tRNA*^{Pyl} pairs for GCE are those originating from the *Methanosarcina* species *mazei* (*MmPylRS*•*MmtRNA*^{Pyl}) and *barkeri*. Importantly, the PylRS and *MjTyrRS* systems are mutually orthogonal and have been used to install distinct ncAAs in response to UAA and UAG codons within the same gene.² Here, we asked whether the *MjTyrRS* and *MmPylRS* can be used to introduce two unique ncAAs in response to UAU and UAA, respectively. With this objective we assembled a plasmid system for simultaneous expression of the two o-aaRS•o-tRNA pairs, together with a sfGFP reporter containing a UAU mutation at position 1 and a UAA mutation at position 151 (sfGFP[1UAU-151UAA]; **Figure S6A**). In this system we employed the *MjTyrRS* variant AzFRS.2.t1, together with *tRNA*^{Ty2}_{AUA}, to incorporate pMeF in response to the initiating UAU and wildtype *MmPylRS*, together with a mutant opal-suppressor *MmtRNA*^{Pyl} (*MmtRNA*^{Pyl}_{UUA})³⁶, to introduce *N*^ε-boc-L-lysine (BocK; **Figure 1**) in response to the elongating UAA. We measured sfGFP production in *E. coli* DH10BΔ*metZWW* in the presence of both pMeF and BocK. Under these conditions, robust expression occurred only when both ncAAs were provided in the growth media (**Figure 3A**), suggesting that the UAU and UAA codons were simultaneously suppressed to afford full-length sfGFP with the desired ncAAs.

Next, we evaluated dual ncAA incorporation using a combination of UAU and UAG. To introduce a second ncAA in response to UAG, we employed a recently characterized PylRS•*tRNA*^{Pyl} pair from the methanogenic archaeon “*Candidatus* Methanomethylophilus alvus” (*MaPylRS*•*MatRNA*^{Pyl}).³⁷ Like the *MmPylRS* pair, the *MaPylRS*•*MatRNA*^{Pyl} pair is orthogonal in bacteria and eukaryotes.^{37, 38} Furthermore, we have shown that this pair is mutually orthogonal with the engineered *MjTyrRS*•*tRNA*^{Ty2} pair.²⁵ We employed the same plasmid-based system to co-express these two o-aaRS•o-tRNA

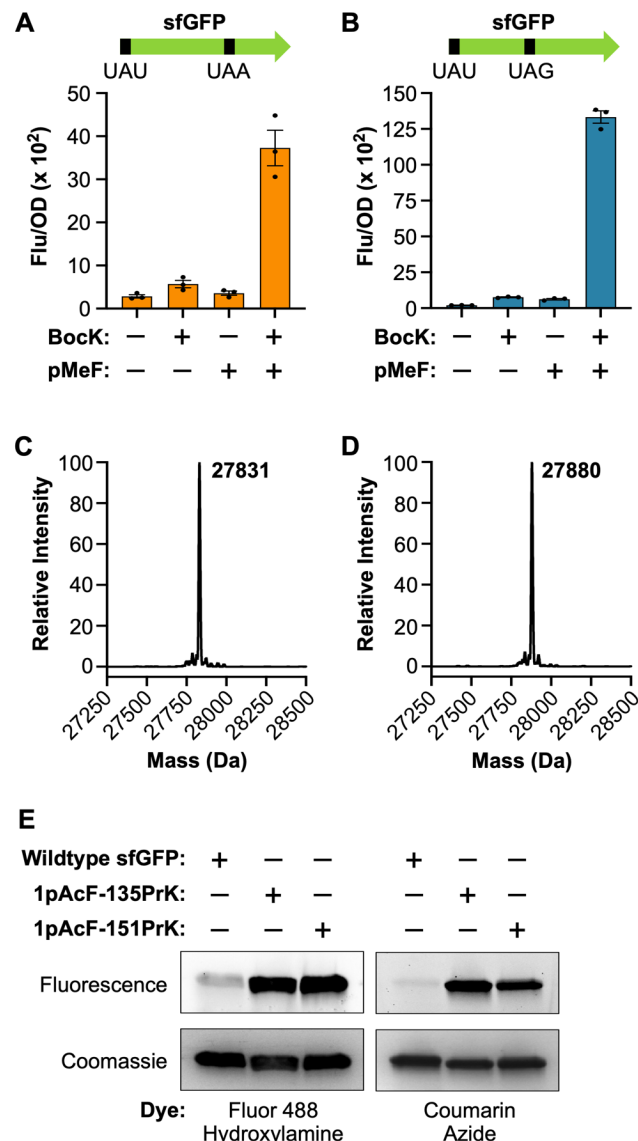


Figure 3. Co-translational installation of two distinct ncAAs encoded by UAU/UAA and UAU/UAG. (A-B) Expression of sfGFP[1UAU-151UAA] (A) and sfGFP[1UAU-135UAG] (B) in the presence of pMeF and BocK. Data are displayed as the mean ± SEM of three biological replicates. (C) LC-MS of sfGFP-1pAcF-135PrK (theoretical mass = 27833 Da). (D) LC-MS of sfGFP-1pAcF-151PrK (theoretical mass = 27882 Da). (E) Labeling of sfGFP-1pAcF-135PrK and sfGFP-1pAcF-151PrK with Fluor 488-hydroxylamine and coumarin azide. Proteins were resolved by SDS-PAGE and visualized by in-gel fluorescence and Coomassie staining.

pairs along with a sfGFP mutant containing a UAU mutation at position 1 and a UAG mutation at position 135 (sfGFP[1UAU-135UAG]; **Figure S6B**). AzFRS.2.t1 and *tRNA*^{Ty2}_{AUA} were used to incorporate pMeF in response to the initiating UAU, while wildtype *MaPylRS* and a variant of *MatRNA*^{Pyl} (*MatRNA*(6)^{Pyl}), which was engineered to be orthogonal to *MmPylRS*, were used to incorporate BocK in response to UAG.³⁷ Again, robust sfGFP expression occurred only when both pMeF and BocK were added to the growth media, demonstrating dual suppression of UAU and UAG and successful incorporation of both ncAAs (**Figure 3B**).

AzFRS.2.t1 and the wildtype PylRSs are polyspecific syn-

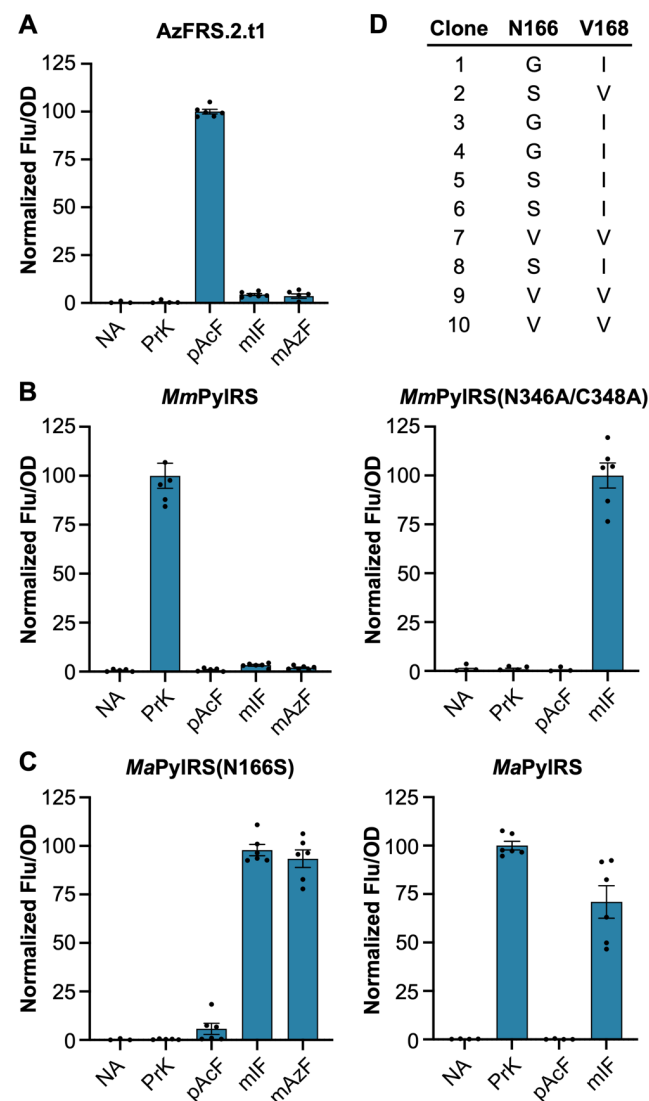


Figure 4. Substrate specificities of three mutually orthogonal aminoacyl-tRNA synthetases. (A-C) Expression of sfGFP[2UAG] in the presence of PrK, pAcF, mIF, or mAzF facilitated by (A) the *Mj*TyrRS variant AzFRS.2.t1, (B) wildtype *Mm*PylRS and variant *Mm*PylRS(N346A/C348A), and (C) wildtype *Ma*PylRS and variant *Ma*PylRS(N166S). Data are displayed as the mean \pm SEM of six biological replicates. NA = None Added. (D) *Ma*PylRS mutants identified from an N166/V168 library selected on media containing mIF (clones 1–5) or oMeF (clones 6–10).

thetases that recognize a number of substrates.^{29, 34} Among these are ncAAs with bioorthogonal functional groups that enable site-specific protein bioconjugation. For example, in addition to pMeF, AzFRS.2.t1 recognizes the ketone-containing ncAA pAcF which can undergo oxime ligation with hydroxylamine probes.³⁹ Furthermore, in addition to BocK, wildtype PylRS recognizes *N*^ε-propargyl-L-lysine (PrK; **Figure 1**) which can undergo Cu(I)-catalyzed cycloaddition with azides.⁴⁰ We tested whether these o-aaRS•o-tRNA pairs could be used to produce proteins containing ketone and alkyne reactive moieties at precisely defined sites by expressing sfGFP[1UAU-135UAG] and sfGFP[1UAU-151UAA] in the presence of pAcF and PrK. Again, robust expression of these reporter proteins occurred only when both ncAAs were provided in the growth media (**Figure S6C–D**). Next, we purified

these proteins and confirmed pAcF and PrK incorporation by LC-MS; in both cases we observed mass peaks consistent with sfGFP containing the expected ncAA substitutions (**Figure 3C–D**). We also confirmed the site-specificity of pAcF and PrK incorporation by MS/MS analysis (**Figure S7–8**). Finally, to demonstrate the utility of this system, we showed that these proteins, which contain ketone and alkyne bioconjugation handles, can be readily labeled with both hydroxylamine- and azide-based fluorescent dyes (**Figure 3E and S9**).

Co-translational installation of three distinct noncanonical amino acids. The above data demonstrate that *Mj*TyrRS•tRNA^{Ty2}_{AUA} can be used, in conjunction with two different PylRS•tRNA^{Pyl} pairs, to incorporate distinct ncAAs in response to UAU and UAG/UAA. Next, we asked whether these mutually orthogonal o-aaRS•o-tRNA pairs can be combined to direct site-specific installation of three unique ncAAs. As a prerequisite, we first sought to identify three o-aaRS variants that recognize distinct ncAAs substrates. The *Mj*TyrRS variant AzFRS.2.t1 primarily recognizes *para*-substituted phenylalanine derivatives, whereas wildtype *Mm*PylRS and *Ma*PylRS both recognize *N*^ε-substituted lysine derivatives and have overlapping substrate specificities.⁴¹ A major contributor to the substrate specificity of PylRS is the so-called *gatekeeper residue*, a conserved asparagine located in the enzyme's substrate binding pocket. This residue is involved in several interactions with the substrate amino acid, including a hydrogen bond with the side chain amide oxygen of pyrrolysine, and other *N*^ε-substituted lysine derivatives (**Figure S10A**).^{42, 43} Mutating this asparagine in *Mm*PylRS to alanine (N346A) obliterates recognition of *N*^ε-substituted lysine derivatives.⁴⁴ Introducing a second alanine mutation (C348A) affords an *Mm*PylRS variant (*Mm*PylRS(N346A/C348A)) that recognizes more than thirty *ortho*-, *para*-, and *meta*-substituted phenylalanine derivatives.^{44–46} However, this variant poorly recognizes derivatives with small *para* substitutions, such as those recognized by AzFRS.2.t1.⁴⁴ Therefore, we hypothesized that *Mm*PylRS(N346A/C348A) and AzFRS.2.t1 could be used together in the same cell to selectively install *meta*- and *para*-substituted phenylalanine derivatives, respectively. In addition, we hypothesized that wildtype *Ma*PylRS could serve as a third o-aaRS to install *N*^ε-substituted lysine derivatives.

To confirm the substrate specificity of these o-aaRSs, we measured sfGFP[2UAG] production in cells expressing either AzFRS.2.t1, *Mm*PylRS(N346A/C348A), or wildtype *Ma*PylRS (as well as their cognate amber-suppressor tRNAs) in the presence of three substrates: pAcF, PrK, and *meta*-iodo-L-phenylalanine (mIF; **Figure 1**). Consistent with our hypothesis, we found that AzFRS.2.t1 selectively recognizes the *para*-substituted ncAA pAcF, while *Mm*PylRS(N346A/C348A) selectively recognizes the *meta*-substituted ncAA mIF (**Figure 4A–B**). However, surprisingly, we found that mIF is also recognized by wildtype *Ma*PylRS (**Figure 4C**). Indeed, *Ma*PylRS enabled robust sfGFP expression in the presence of both mIF and PrK. In contrast, wildtype *Mm*PylRS only afforded robust reporter expression in the presence of PrK (**Figure 4B**).

In light of this observation, we decided instead to use *Ma*PylRS to install *meta*-substituted ncAAs and wildtype *Mm*PylRS to install *N*^ε-substituted lysine derivatives. However, this required that we engineer an *Ma*PylRS variant that accepts mIF and rejects PrK. Towards this end, we constructed a two-site *Ma*PylRS library randomizing residues N166 and V168 (corresponding to N346 and C348 in *Mm*PylRS, **Figure**

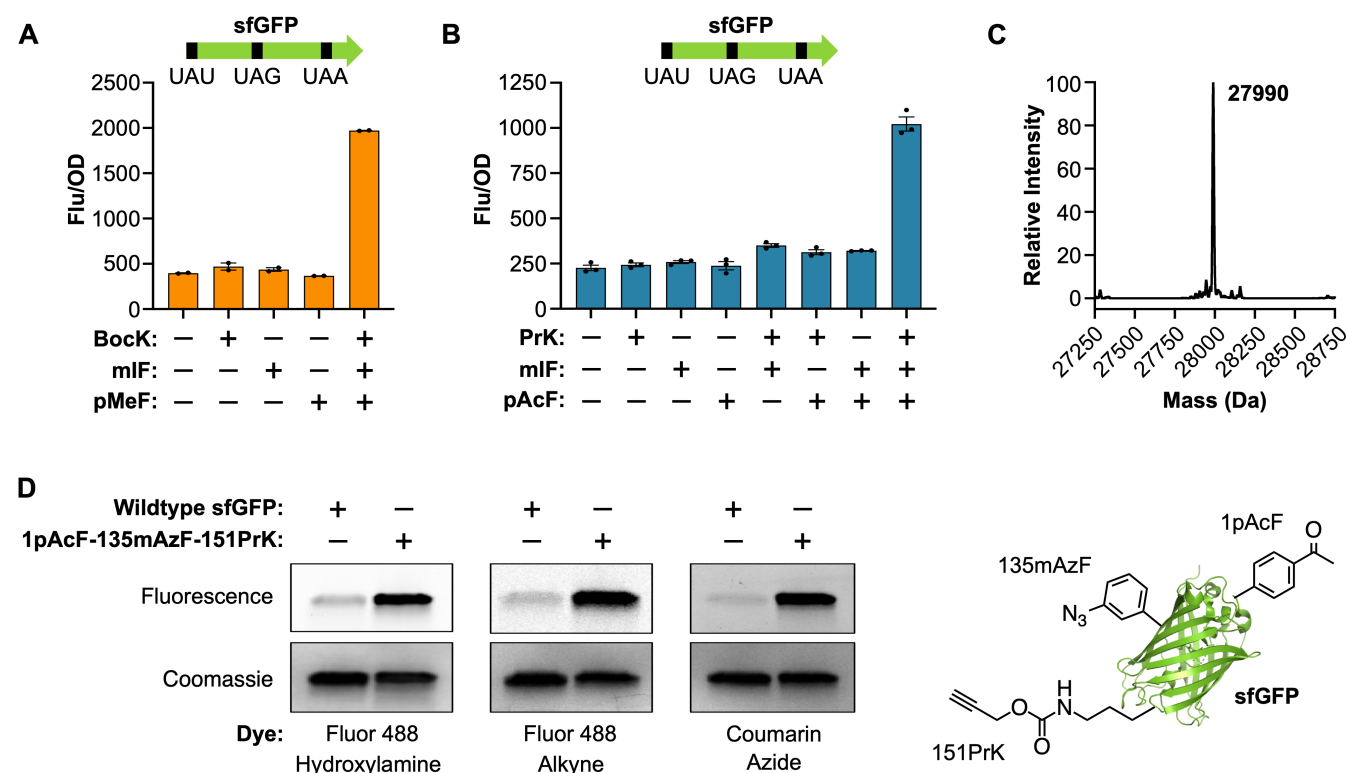


Figure 5. Co-translational installation of three distinct noncanonical amino acids encoded by UAU, UAG, and UAA. (A) Expression of sfGFP[1UAU-135UAG-151UAA] in the presence of BocK, mIF, and pMeF. Data are displayed as the mean \pm SEM of two biological replicates. (B) Expression of sfGFP[1UAU-135UAG-151UAA] in the presence of PrK, mIF, and pAcF. Data are displayed as the mean \pm SEM of three biological replicates. (C) LC-MS of sfGFP-1pAcF-135mIF-151PrK (theoretical mass = 27992 Da). (D) Labeling of sfGFP-1pAcF-135mAzF-151PrK with Fluor 488-hydroxylamine, Fluor 488-alkyne, and coumarin azide. Proteins were resolved by SDS-PAGE and visualized by in-gel fluorescence and Coomassie staining. A cartoon representation of sfGFP-1pAcF-135mAzF-151PrK is also shown (PDB: 2B3P).

S10B). We subjected this library to positive selection using chloramphenicol acetyltransferase with a UAG codon at position 112 (cat[112UAG]).⁴⁷ Cells expressing the *MaPylRS* N166/V168 library, *MatRNA*(6)^{Pyl}, and cat[112UAG] were challenged to grow on media containing 1 mM mIF and 50 μ g/mL chloramphenicol. We also performed a parallel selection using *ortho*-methyl-L-phenylalanine (oMeF; **Figure 1**). Following selection, surviving clones were screened for ncAA-dependent growth (**Figure S11**) and five clones from each selection were sequenced to reveal mutations in *MaPylRS* that altered its specificity. In total, four unique mutants with similar amino acid substitutions were identified (**Figure 4D**). We screened these mutants for selective incorporation of mIF using sfGFP[2UAG]. Gratifyingly, we found that robust expression occurred only in the presence of mIF, indicating that these newly identified *MaPylRS* mutants are selective for the *meta*-substituted ncAA (**Figure 4C** and **S12**). In subsequent experiments we employed the N166S mutant, *MaPylRS*(N166S). In addition to mIF and oMeF, we screened *MaPylRS*(N166S) for the ability to recognize various substituted phenylalanine derivatives. Similar to *MmPylRS*-(N346A/C348A), *MaPylRS*(N166S) recognizes a number of substrates (**Figure S13**) including those with bioorthogonal handles such as *meta*-azido-L-phenylalanine (mAzF; **Figure 1**). Like mIF, we found that mAzF is rejected by AzFRS.2.t1 and wildtype *MmPylRS* (**Figure 4A–C**). To test if this *ortho/para/meta* substrate selectivity is a general feature of *MaPylRS*(N166S) and AzFRS.2.t1 we compared sfGFP[2UAG] expression with these enzymes using a panel of

phenylalanine derivatives containing the same substituent at different positions of the aromatic ring. In all cases we found that AzFRS.2.t1 selectively recognizes *para*-substituted derivatives, while *MaPylRS*(N166S) selectively recognizes derivatives with *ortho* or *meta* substitutions (**Figure S14**).

Next, we devised a plasmid-based system for simultaneous incorporation of three distinct ncAAs. This three-plasmid system consisted of: (1) an accessory plasmids encoding AzFRS.2.t1, *MaPylRS*(N166S), and *MatRNA*(6)^{Pyl}, (2) a second accessory plasmid encoding *MmPylRS* and *MmtRNA*^{Pyl}_{UUA}, and (3) a reporter plasmid encoding *tRNA*^{Ty2}_{AUA} and a triple-mutant sfGFP with a UAU mutation at position 1, a UAG mutation at position 135, and a UAA mutation at position 151 (**Figure S15A**). We measured sfGFP production *E. coli* DH10B Δ metZWW in the presence of pMeF, BocK, and mIF. With this initial system, we observed significant sfGFP production only in the presence of all three ncAAs, however, the signal was relatively low (**Figure S15B**). We hypothesized that this was due to poor suppression of the UAA codon by *MmtRNA*^{Pyl}_{UUA} since, *MmPylRS* is the least active o-aaRS in this system. Therefore, an additional copy of *MmtRNA*^{Pyl}_{UUA} was added to the reporter plasmid resulting in a \sim 2-fold increase in the fluorescence signal (**Figure S15B**). Next, we purified sfGFP expressed with pAcF, PrK, and mIF and we analyzed the protein by LC-MS. This analysis revealed a major peak consistent with sfGFP containing three ncAAs at the desired positions, however, a second peak was also present (**Figure S16**). This second peak was consistent with misincor-

poration of PrK at the UAG codon. Further analysis by MS/MS confirmed a mixture of mIF and PrK at position 135 (**Figure S17–18**). We surmised that mis-incorporation of PrK likely results from an excess of *MmtRNA*^{Pyl}_{UUA}, since ochre-suppressor tRNAs can suppress both UAA and UAG.⁴⁸ Therefore, we constructed a third reporter plasmid in which the extra copy of *MmtRNA*^{Pyl}_{UUA} was replaced with *MatRNA*(6)^{Pyl} (**Figure S15A**). Using this new plasmid, we again found that sfGFP production occurred only when the growth media was supplemented with a *para*-substituted phenylalanine, a *meta*-substituted phenylalanine, and an *N*^ε-substituted lysine derivative, i.e., pMeF, mIF, and BocK or pAcF, mIF, and PrK. (**Figure 5A–B**). LC-MS analysis of sfGFP expressed with pAcF, mIF, and PrK revealed a major peak consistent with the incorporation of all three amino acids; a peak corresponding to mis-incorporation of PrK was not detected (**Figure 5C**). Furthermore, MS/MS analysis confirmed the incorporation of each ncAA at the desired positions (**Figure S19**). We used this optimized expression system to site-specifically install three reactive nAAs (pAcF, mAzF, and PrK) into sfGFP affording a protein with three unique bioorthogonal functional groups (ketone, azide, and alkyne; **Figure 5D**). Finally, we demonstrated that this single protein could be modified with multiple fluorescent probes bearing hydroxylamine, alkyne, or azide reactive handles (**Figure 5D and S20**).

In summary, we have engineered an orthogonal initiator tRNA that enables efficient reassignment of the UAU codon to initiate translation with nAAs. We demonstrated that this system enables double and triple incorporation of distinct nAAs encoded by UAU, UAG, and UAA codons. Several of the nAAs that we tested for simultaneous incorporation (e.g., pAcF, mIF, mAzF, and PrK) contain unique bioorthogonal reaction handles that enable double and triple labeling of proteins produced from this system. In addition to the demonstrated reactivity of pAcF, mAzF, and PrK, mIF can also be used as a reaction handle for protein conjugation with boronic acids *via* palladium-catalyzed Suzuki-Miyaura cross-coupling.⁴⁹

Safety statement. No unexpected, new, or significant hazards or risks were encountered during the course of this work.

ASSOCIATED CONTENT

Supporting Information

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

Materials and Methods, Supplementary Figures and Tables, DNA Sequences (PDF)

AUTHOR INFORMATION

Corresponding Authors

* E-mail: schepartz@berkeley.edu

* E-mail: dieter.soll@yale.edu

ORCID

Jeffery M. Tharp: 0000-0002-2362-3249
Oscar Vargas-Rodriguez: 0000-0002-2301-2800
Alanna Schepartz: 0000-0003-2127-3932
Dieter Söll: 0000-0002-3077-8986

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Kyle Hoffman and Sebastian Santiago for help with mass spectrometry data collection and analysis, and Erol Vatansever, Jonathan Fischer, Christina Chung, and Natalie Krahn for providing critical feedback on the manuscript. J.M.T. was supported by the Center for Genetically Encoded Materials, an NSF Center for Chemical Innovation (NSF CHE-2021739 to A.S.), O.V.-R. was supported by the National Institute of General Medical Sciences (R35GM122560 to D.S.). The genetic experiments were supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the Department of Energy (DE-FG0298ER20311 to D.S.).

REFERENCES

- (1) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566–7571.
- (2) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew. Chem. Int. Ed.* **2010**, *49*, 3211–3214.
- (3) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441–444.
- (4) Italia, J. S.; Addy, P. S.; Erickson, S. B.; Peeler, J. C.; Weerapana, E.; Chatterjee, A. Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc.* **2019**, *141*, 6204–6212.
- (5) Dunkelmann, D. L.; Willis, J. C. W.; Beattie, A. T.; Chin, J. W. Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **2020**, *12*, 535–544.
- (6) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. Catalyst-free and site-specific one-pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids. *ChemBioChem* **2012**, *13*, 1405–1408.
- (7) Venkat, S.; Sturges, J.; Stahman, A.; Gregory, C.; Gan, Q.; Fan, C. Genetically incorporating two distinct post-translational modifications into one protein simultaneously. *ACS Synth. Biol.* **2018**, *7*, 689–695.
- (8) Zheng, Y.; Gilgenast, M. J.; Hauc, S.; Chatterjee, A. Capturing post-translational modification-triggered protein-protein interactions using dual noncanonical amino acid mutagenesis. *ACS Chem. Biol.* **2018**, *13*, 1137–1141.
- (9) Wang, K.; Schmied, W. H.; Chin, J. W. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed.* **2012**, *51*, 2288–2297.
- (10) Hankore, E. D.; Zhang, L. Y.; Chen, Y.; Liu, K.; Niu, W.; Guo, J. T. Genetic incorporation of noncanonical amino acids using two mutually orthogonal quadruplet codons. *ACS Synth. Biol.* **2019**, *8*, 1168–1174.
- (11) Wang, N.; Shang, X.; Cerny, R.; Niu, W.; Guo, J. Systematic evolution and study of UAGN decoding tRNAs in a genomically recoded bacteria. *Sci. Rep.* **2016**, *6*, 21898.
- (12) Chatterjee, A.; Lajoie, M. J.; Xiao, H.; Church, G. M.; Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem* **2014**, *15*, 1782–1786.
- (13) Link, A. J.; Mock, M. L.; Tirrell, D. A. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.
- (14) Bohlke, N.; Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.* **2014**, *351*, 133–144.

- (15) Krishnakumar, R.; Ling, J. Experimental challenges of sense codon reassignment: an innovative approach to genetic code expansion. *FEBS Lett.* **2014**, *588*, 383–388.
- (16) Krishnakumar, R.; Prat, L.; Aerni, H. R.; Ling, J. Q.; Merryman, C.; Glass, J. I.; Rinehart, J.; Söll, D. Transfer RNA misidentification scrambles sense codon recoding. *ChemBioChem* **2013**, *14*, 1967–1972.
- (17) Zeng, Y.; Wang, W.; Liu, W. R. Towards reassigning the rare AGG codon in *Escherichia coli*. *ChemBioChem* **2014**, *15*, 1750–1754.
- (18) De Simone, A.; Acevedo-Rocha, C. G.; Hoesl, M. G.; Budisa, N. Towards reassignment of the methionine codon AUG to two different noncanonical amino acids in bacterial translation. *Croat. Chem. Acta* **2016**, *89*, 243–253.
- (19) Mukai, T.; Yamaguchi, A.; Ohtake, K.; Takahashi, M.; Hayashi, A.; Iraha, F.; Kira, S.; Yanagisawa, T.; Yokoyama, S.; Hoshi, H.; Kobayashi, T.; Sakamoto, K. Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* **2015**, *43*, 8111–8122.
- (20) Biddle, W.; Schmitt, M. A.; Fisk, J. D. Evaluating sense codon reassignment with a simple fluorescence screen. *Biochemistry* **2015**, *54*, 7355–7364.
- (21) Lee, B. S.; Shin, S.; Jeon, J. Y.; Jang, K. S.; Lee, B. Y.; Choi, S.; Yoo, T. H. Incorporation of unnatural amino acids in response to the AGG codon. *ACS Chem. Biol.* **2015**, *10*, 1648–1653.
- (22) Wang, Y.; Tsao, M. L. Reassigning sense codon AGA to encode noncanonical amino acids in *Escherichia coli*. *ChemBioChem* **2016**, *17*, 2234–2239.
- (23) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.* **2003**, *125*, 7512–7513.
- (24) Ho, J. M.; Reynolds, N. M.; Rivera, K.; Connolly, M.; Guo, L. T.; Ling, J.; Pappin, D. J.; Church, G. M.; Söll, D. Efficient reassignment of a frequent serine codon in wild-type *Escherichia coli*. *ACS Synth. Biol.* **2016**, *5*, 163–171.
- (25) Tharp, J. M.; Ad, O.; Amikura, K.; Ward, F. R.; Garcia, E. M.; Cate, J. H. D.; Schepartz, A.; Söll, D. Initiation of protein synthesis with non-canonical amino acids in vivo. *Angew. Chem. Int. Ed.* **2020**, *59*, 3122–3126.
- (26) Tharp, J. M.; Krahn, N.; Varshney, U.; Söll, D. Hijacking translation initiation for synthetic biology. *ChemBioChem* **2020**, *21*, 1387–1396.
- (27) Pédélecq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
- (28) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **2011**, *50*, 1894–1900.
- (29) Amiram, M.; Haimovich, A. D.; Fan, C.; Wang, Y. S.; Aerni, H. R.; Ntai, I.; Moonan, D. W.; Ma, N. J.; Rovner, A. J.; Hong, S. H.; Kelleher, N. L.; Goodman, A. L.; Jewett, M. C.; Söll, D.; Rinehart, J.; Isaacs, F. J. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* **2015**, *33*, 1272–1279.
- (30) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* **2003**, *10*, 425–432.
- (31) Goto, Y.; Iseki, M.; Hitomi, A.; Murakami, H.; Suga, H. Nonstandard peptide expression under the genetic code consisting of reprogrammed dual sense codons. *ACS Chem. Biol.* **2013**, *8*, 2630–2634.
- (32) Giegé, R.; Sissler, M.; Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **1998**, *26*, 5017–5035.
- (33) Tobias, J. W.; Shrader, T. E.; Rocap, G.; Varshavsky, A. The N-end rule in bacteria. *Science* **1991**, *254*, 1374–1377.
- (34) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **2014**, *1844*, 1059–1070.
- (35) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing logical codon reassignment—Expanding the chemistry in biology. *Chem. Sci.* **2015**, *6*, 50–69.
- (36) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, *52*, 1828–1837.
- (37) Willis, J. C. W.; Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **2018**, *10*, 831–837.
- (38) Meineke, B.; Heimgärtner, J.; Lafranchi, L.; Elsässer, S. J. *Methanomethylophilus alvus* Mx1201 provides basis for mutual orthogonal pyrrolysyl tRNA/aminoacyl-tRNA synthetase pairs in mammalian cells. *ACS Chem. Biol.* **2018**, *13*, 3087–3096.
- (39) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56–61.
- (40) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
- (41) Yamaguchi, A.; Iraha, F.; Ohtake, K.; Sakamoto, K. Pyrrolysyl-tRNA synthetase with a unique architecture enhances the availability of lysine derivatives in synthetic genetic codes. *Molecules* **2018**, *23*, 2460.
- (42) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268–11273.
- (43) Flügel, V.; Vrabl, M.; Schneider, S. Structural basis for the site-specific incorporation of lysine derivatives into proteins. *PLoS One* **2014**, *9*, e96198.
- (44) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **2012**, *134*, 2950–2953.
- (45) Wang, Y. S.; Fang, X.; Chen, H. Y.; Wu, B.; Wang, Z. U.; Hilty, C.; Liu, W. R. Genetic incorporation of twelve *meta*-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **2013**, *8*, 405–415.
- (46) Tharp, J. M.; Wang, Y. S.; Lee, Y. J.; Yang, Y.; Liu, W. R. Genetic incorporation of seven *ortho*-substituted phenylalanine derivatives. *ACS Chem. Biol.* **2014**, *9*, 884–890.
- (47) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **2001**, *292*, 498–500.
- (48) Eggertsson, G.; Söll, D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol. Rev.* **1988**, *52*, 354–374.
- (49) Chalker, J. M.; Wood, C. S.; Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347.

TOC Image:

