Check for updates

Received: 18 June 2022 | Revised: 18 August 2022 | Accepted: 24 August 2022

DOI: 10.1002/pro.4443

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Expanding the eukaryotic genetic code with a biosynthesized 21st amino acid

Kuan-Lin Wu[1] | **Joshua A. Moore**[2,3] | **Mitchell D. Miller**[2] | **Yuda Chen**[1] | **Catherine Lee**[1] | **Weijun Xu**[2] | **Zane Peng**[1] | **Qinghui Duan**[1] | **George N. Phillips Jr**[1,2] | **Rosa A. Uribe**[2,3] | **Han Xiao**[1,2,4]

[1]Department of Chemistry, Rice University, Houston, Texas, USA

[2]Department of Biosciences, Rice University, Houston, Texas, USA

[3]Biochemistry and Cell Biology Program, Rice University, Houston, Texas, USA

[4]Department of Bioengineering, Rice University, Houston, Texas, USA

Correspondence

George N. Phillips Jr, Department of Chemistry and Department of Biosciences, Rice University, Houston, TX, USA.
Email: georgep@rice.edu

Rosa A. Uribe, Department of Biosciences and Biochemistry and Cell Biology Program, Rice University, Houston, TX, USA.
Email: ru5@rice.edu

Han Xiao, Department of Chemistry, Department of Biosciences, and Department of Bioengineering, Rice University, Houston, TX, USA.
Email: han.xiao@rice.edu

Funding information

Cancer Prevention and Research Institute of Texas, Grant/Award Numbers: RR170014, RR170062; National Science Foundation, Grant/Award Number: STC1231306; U.S. Department of Defense, Grant/Award Number: W81XWH-21-1-0789; Biological and Environmental Research, Grant/Award Number: KP1605010; Division of Intramural Research, National Institute of Allergy and Infectious Diseases, Grant/Award Number: R01-AI165079; National Cancer Institute, Grant/Award Numbers: ACB-12002, R01-CA217255, R21-CA255894; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Numbers: R01-, DK124804;

Abstract

Genetic code expansion technology allows for the use of noncanonical amino acids (ncAAs) to create semisynthetic organisms for both biochemical and biomedical applications. However, exogenous feeding of chemically synthesized ncAAs at high concentrations is required to compensate for the inefficient cellular uptake and incorporation of these components into proteins, especially in the case of eukaryotic cells and multicellular organisms. To generate organisms capable of autonomously biosynthesizing an ncAA and incorporating it into proteins, we have engineered a metabolic pathway for the synthesis of *O*-methyltyrosine (OMeY). Specifically, we endowed organisms with a marformycins biosynthetic pathway-derived methyltransferase that efficiently converts tyrosine to OMeY in the presence of the co-factor *S*-adenosylmethionine. The resulting cells can produce and site-specifically incorporate OMeY into proteins at much higher levels than cells exogenously fed OMeY. To understand the structural basis for the substrate selectivity of the transferase, we solved the X-ray crystal structures of the ligand-free and tyrosine-bound enzymes. Most importantly, we have extended this OMeY biosynthetic system to both mammalian cells and the zebrafish model to enhance the utility of genetic code expansion. The creation of autonomous eukaryotes using a 21st amino acid will make genetic code expansion technology more applicable to multicellular organisms, providing valuable vertebrate models for biological and biomedical research.

KEYWORDS

biosynthesis, cell engineering, genetic code expansion, metabolic engineering, noncanonical amino acid, site-specific incorporation, unnatural amino acid

© 2022 The Protein Society.

National Institute of General Medical Sciences, Grant/Award Numbers: AGM-12006, P30GM138396, R01-GM115261, R35-GM133706, P30GM133893; National Institute on Minority Health and Health Disparities, Grant/Award Number: F31-HD104474; Welch Foundation, Grant/Award Numbers: C-1970, C-2118; Michigan Technology Tri-Corridor, Grant/Award Number: 085P1000817; Brookhaven National Laboratory, Grant/Award Number: DE-SC0012704; U. S. Department of Energy (DOE), Grant/Award Number: DE-AC02-06CH11357

Review Editor: John Kuriyan

1 | INTRODUCTION

Genetic code expansion technology has been widely used in living cells for site-specific modification of proteins with amino acids having novel chemical, physical, and biological properties.[1–8] This technology is achieved by engineering cells with a bioorthogonal translation system. This consists of an aminoacyl-tRNA synthetase (aaRS) that specifically recognizes a unique tRNA and a noncanonical amino acid (ncAA) that is present at a high intracellular level.[2,9] Since the first report in 2001 of a biorthogonal translation system encoding *O*-methyltyrosine (OMeY), more than 200 ncAAs with diverse chemical, physical, and biological properties have been successfully incorporated into proteins in bacteria using genetic code expansion technology based on evolved archaeal and eukaryotic aaRS/tRNA pairs. To extend genetic code expansion technology to eukaryotes, the pyrrolysyl-tRNA synthetase/tRNA pair from archaea, and the tyrosyl- and leucyl-tRNA synthetase/tRNA pairs from *Escherichia coli* (*E. coli*) have been evolved to genetically incorporate ncAAs into mammalian proteins.[1,2,10] Recently, the scope of this technology has been further expanded from simple cell culture to animals, including worms, fruit flies, zebrafish, and mice.[11–16] These eukaryotic models serve as critical tools for the study of complex biological processes, including tissue development, neurobiological functions, and signaling pathways.

Despite this progress in genetic code expansion technology, the important capability for autonomous biosynthesis of ncAAs has been underdeveloped.[17] Attempts to generate organisms capable of autonomously biosynthesizing ncAAs and incorporating them into proteins have been focused on bacterial cells. In a limited number of examples, the employment of biosynthesized ncAAs has been shown to be beneficial for both the efficiency of ncAA incorporation and the application of genetic code expansion to biological systems. Several synthetic gene clusters for ncAA biosynthesis, including *p*-aminophenylalanine,[18,19] phosphothreonine (pThr),[20] 5-hydroxytryptophan,[21] and 3,4-dihydroxyphenylalanine (DOPA),[22] have been coupled to their corresponding orthogonal mutant aaRS/tRNA pairs to generate completely autonomous *E. coli* cells with an expanded genetic code. In other examples, aminotransferase, tyrosine phenol-lyase, and cysteine β-replacement enzyme have been used to create *E. coli* strains able to biosynthesize and incorporate phenylalanine analogs,[23] DOPA,[24] and cysteine analogs[25] into proteins. However, even in these cases, biosynthesis of DOPA and cysteine analogs requires exogenous addition of 1–10 mM levels of the catechol and aromatic thiols precursors, respectively. In 2017, an autonomous *E. coli* strain harboring a pThr synthetic enzyme from *Salmonella enterica* and an engineered phosphoseryl-tRNA synthetase/tRNA pair from Methanogenic archaea was reported.[20] Compared to exogenous feeding of 1 mM pThr, the biosynthetic system produced a 40-fold higher level of intracellular pThr, resulting in greater incorporation efficiency. Additionally, a completely autonomous bacterial strain utilizing DOPA as a 21st amino acid for protein synthesis has been constructed.[22] The yield of DOPA-containing protein from this autonomous strain is greater than that from cells exogenously fed with 9 mM DOPA. More recently, an autonomous bacterium engineered for autonomous 5-hydroxytryptophan biosynthesis and incorporation has been used to monitor oxidative stress in real-time.[21] This result cannot be reproduced by exogenously supplying 5-hydroxytryptophan. In spite of these benefits, none of the ncAA biosynthetic pathways have been used to generate autonomous eukaryotes with a 21st amino acid.

Currently, the creation of most ncAA-containing proteins in eukaryotic cells has required supplementation with exogenously supplied ncAAs at 1–10 mM levels. This is necessary in order to achieve intracellular ncAA concentrations sufficient for robust incorporation of the ncAAs into proteins. In addition, in many cases, structural modifications of

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

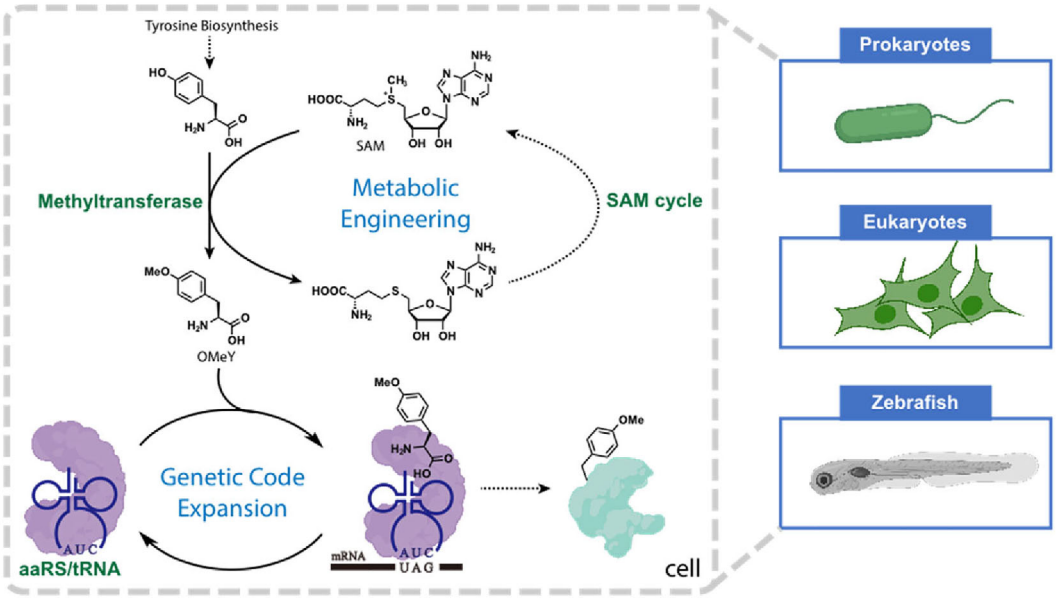

FIGURE 1 Schematic illustration of engineering autonomous organisms with the 21st amino acid, OMeY. In the presence of co-factor SAM, the methyltransferase can efficiently convert tyrosine to OMeY in both prokaryotes and eukaryotes. The biosynthesized OMeY is then site-specifically incorporated into proteins in response to an amber codon

amino acids and engineering of periplasmic amino acid-binding proteins have been carried out to enhance cell membrane permeability to ncAAs.[26–29] Autonomous intracellular biosynthesis of ncAAs provides an alternative strategy for achieving high intracellular ncAA levels independent of their permeability through the cell membrane. This approach should lead to higher yields of ncAA-containing proteins in eukaryotic cells, as well as facilitate the extension of this technology to studying and perturbing complex biological processes in whole organisms. Here, we report our generation of the first autonomous eukaryotic cells and multicellular animals utilizing OMeY as a 21st amino acid for protein synthesis (Figure 1). The resulting autonomous organisms exhibit enhanced efficiency of OMeY incorporation into proteins, compared with that achieved by exogenous feeding of OMeY. Since the exogenous feeding approach has been the gold standard for successful use of genetic code expansion, this comparison underscores the utility of autonomous ncAA biosynthesis for future genetic code expansion in eukaryotes.

2 | RESULTS

2.1 | Identification of *O*-methyltransferases for the biosynthesis of OMeY

OMeY can be biosynthesized from tyrosine by *O*-methyltransferases that use *S*-adenosyl-L-methionine (SAM) as a methyl group donor to carry out the *O*-methylation of phenol-containing compounds. To identify the best transferase for OMeY biosynthesis, we compared the activities of five *O*-methyltransferases (OMeTs): (1) Sky37 within the skyllamycin gene cluster,[30] (2) MfnG for the synthesis of marformycins,[31] (3) catechol OMeT SafC,[32] (4) SfmM3 for the methylation of saframycin,[33] and (5) Catechol *O*-methyltransferase (S-COMT) for the modification of the 3,4-diphenyltyrosine derivatives. These OMeTs have all been reported to have activity with phenol derivatives structurally similar to tyrosine. Genes encoding these OMeTs were codon-optimized based on *E. coli* strain B and subcloned into a P_{BAD} promotor-driven plasmid to yield pBad-OMeT. To evaluate the ability of the different OMeTs to biosynthesize OMeY for genetic incorporation, we employed an amber codon suppression system based on the superfolder green fluorescence protein (sfGFP) signal. This system encodes a mutant *Mj*TyrRS/$tRNA_{Tyr}^{CUA}$ pair recognizing OMeY (encoded on plasmid pUltra-polyRS) and an sfGFP mutant with an amber codon at a permissive site (sfGFP*, encoded on plasmid pLei-sfGFP-D134*).

Using OMeY produced by the respective transferases, suppression of the amber codon allows OMeY incorporation into full-length sfGFP. *E. coli* BL21 (DE3) cells with plasmids pUltra-polyRS and pLei-sfGFP-D134* were independently transformed with each of the pBad-OMeT plasmids. A quantitative fluorescence assay was carried out in 2xYT medium in the presence or absence of L-arabinose to induce P_{BAD} mediated expression of OMeT

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

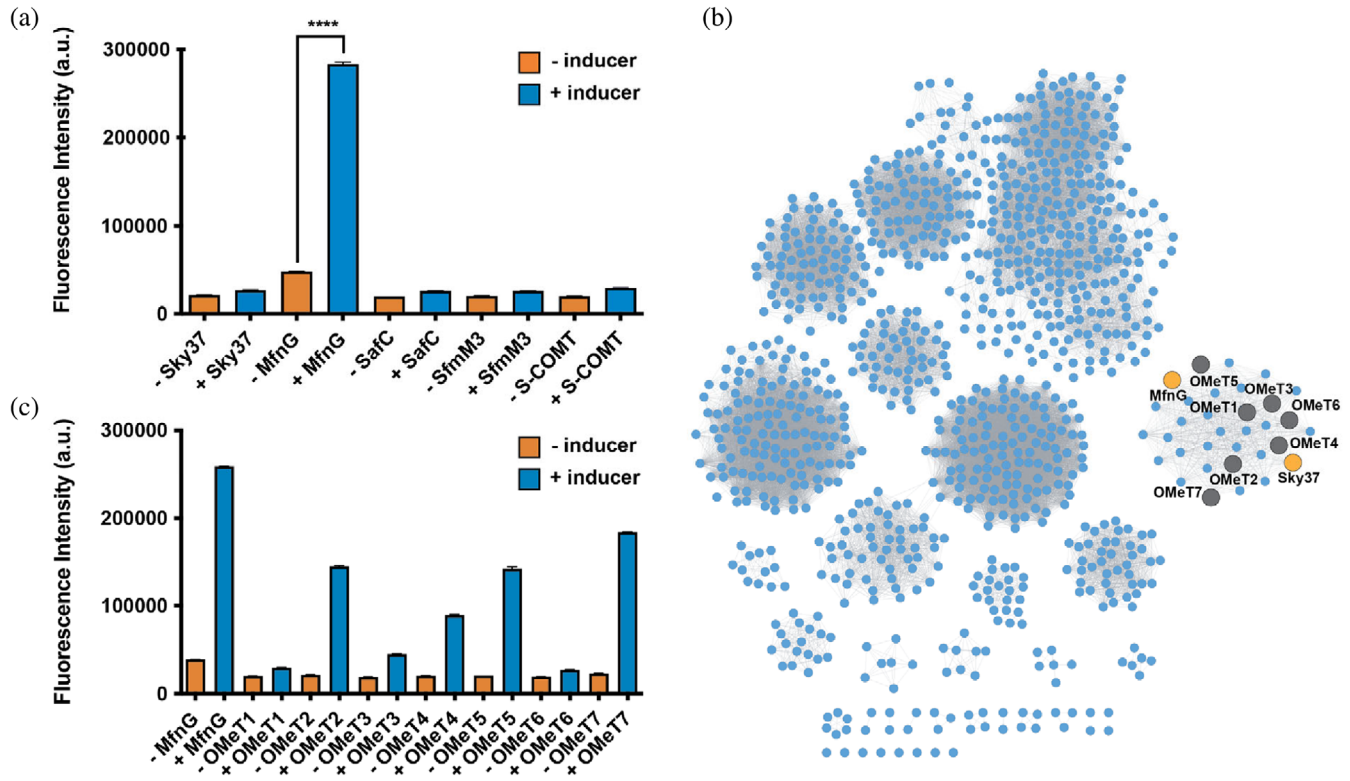

FIGURE 2 Screening of OMeT activity. (a) Fluorescence intensity measurement of BL21 (DE3) cells containing pUltra-polyRS and pLei-sfGFP-D134* in the presence (+) or absence (−) of Sky37, MfnG, SafC, SfmM3, and S-COMT. (b) Sequence similarity network (SSN) based on MfnG sequence. The edge was defined as an alignment score higher than 60. (c) Fluorescence intensity measurement of BL21 (DE3) cells containing pUltra-polyRS and pLei-sfGFP-D134* in the presence (+) or absence (−) of different OMeTs selected from SSN. ****$p < 0.0001$. p value was calculated by two-tailed unpaired student's t test

(Figure 2a). Gratifyingly, compared to the uninduced control, a 10-fold fluorescence increase was observed with MfnG in the presence of the inducer, suggesting the successful biosynthesis and incorporation of OMeY into sfGFP.

Next, we constructed a SSN based on the MfnG amino acid sequence to explore better OMeTs for the biosynthesis of OMeY (Figure 2b).[34] We chose seven enzymes (labeled as OMeT1-7) in this cluster in addition to MfnG and Sky37 for a second round of screening for OMeY production and incorporation. *E. coli* BL21 (DE3) cells containing pUltra-polyRS, pLei-sfGFP-D134* were transformed with pBad-OMeT1-7, and the resulting *E. coli* strains were induced for the production of OMeY-containing sfGFP. In the presence of the inducer, increases in fluorescence were observed for several of the enzymes, including OMeT2, OMeT3, OMeT4, OMeT5, and OMeT7 (Figure 2c). Among these OMeTs, MfnG exhibited the highest activity and was therefore used in subsequent studies.

To quantify levels of biosynthesized OMeY by MfnG, the time course intracellular OMeY concentration was determined using *E. coli* BL21 (DE3) transformed with the pBad-MfnG plasmid. We found that during the early log phase to mid-log phase (1–10 h after induction), a much higher intracellular level of OMeY could be detected in the MfnG transformed cells than in cells fed exogenously with OMeY (Figure S1). These data show that enhanced production of OMeY can be achieved with autonomously biosynthesized OMeY.

2.2 | Structural insight into MfnG catalysis

There are nine major classes of SAM-dependent methyltransferases (MeTs) that participate in the methylation of a wide range of secondary metabolites.[35–39] Class I MeTs comprised a Rossmann-like αβα sandwich structure, which most OMeTs belong to this class. Based on the definition of a protein domain and family search website Prosite (https://prosite.expasy.org/), OMeTs can be further divided into 2 classes. Class I OMeTs are relatively small (ca. 22 kDa), with activities that depend on divalent metal ions. S-COMT is an example of this type of OMeT. MfnG is an example of Class II OMeTs. These OMeTs are

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

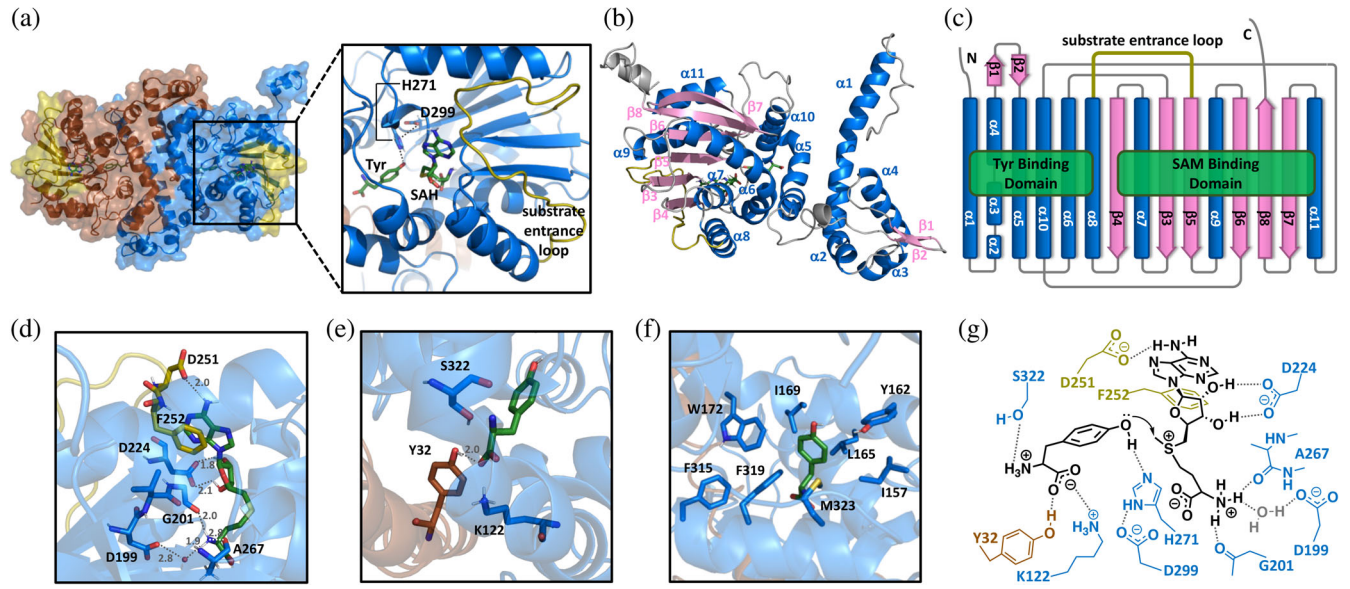

FIGURE 3 Crystal structure of MfnG. (a) Schematic surface representation of MfnG dimer (chain A marine and chain B brown). The enlarged image shows the catalytic pocket with substrates. The flexible loop in olive color controls substrate entry. (b) The Rossmann-fold-like architecture of MfnG. Helices are shown in marine, β-strands in pink, and loops in gray. (c) Schematic representation of the topology and structural motifs of MfnG, indicating the SAM and tyrosine binding domain. (d) Interaction between MfnG and SAH. The water molecule is labeled as a red sphere. (e) Interaction between MfnG and carboxyl and amino group of tyrosine. (f) Hydrophobic interaction between MfnG and the tyrosine aromatic moiety. (g) Schematic representation of the enzyme-substrate interactions in the catalytic pocket

metal independent, with prominent *N*-terminal substrate-binding domains and larger molecular masses (36–43 kDa). To better understand the structure/function parameters of MfnG, we crystallized the enzyme along with SAM and determined the crystal structure at 1.2 Å resolution by molecular replacement (MR, PDB: 7UX7) (the expression, purification, and characterization of MfnG are in Supporting Information Experimental Procedures and Figure S2). Crystalline MfnG exists as a tightly intertwined homodimer, in which the *N*-terminal α-helix of each protomer penetrates the other protomer to generate an engaging interface of 4700 Å^2 primarily from the interaction of helices (Figure 3a). The SAM-binding core of MfnG exhibits a six-stranded β-sheet Rossmann-like fold surrounded by α-helices (Figure 3b, c). We observed the *S*-adenosyl-L-homocysteine (SAH) instead of SAM in the binding pocket since the methyl group can be easily taken away by random environmental nucleophiles. The olive-colored loop extending from Gly240 to Asp262 is characterized by weak electron density (Figure 3a,b), indicative of structural flexibility that could allow substrate access to the catalytic pocket. No metal ions are observed bound to the enzyme, consistent with the metal independent nature of enzyme activity. These characteristics are in line with the PROSITE search result that MfnG (A0A0D4WTP2) is a SAM-dependent class II OMeT. To better understand the structural domains and functions of MfnG, we performed a Dali search (http://ekhidna2.biocenter.helsinki.fi/dali/), which is a webserver specialized in searching similar protein structures by aligning input structure with the implemented PDB library, to pull out known similar proteins for structure and function comparison. The Dali search returned 358 structures from the PDB50 subset, where we examined the top 14 structures. The structural alignment of MfnG with those 14 OMeTs allowed us to identify several conserved residues associated with SAM binding (Figure 3d,g and S3). It appears that the alpha-amino group of SAH participates in two hydrogen-bonding networks. This amino group directly interacts with the backbone carbonyls of Gly201 and Ala267 as well as with the side chain of Asp199 through a water molecule (Figure 3d, S4, and S5). The side chain of Asp224 anchors the SAH ribose moiety through two strong hydrogen bonds. The adenine of SAH is stabilized by Asp251 and Phe252. The side chain of Asp251 hydrogen bonds to adenine, and the aromatic side chain of Phe252 provides π–π stacking stabilization. Both Asp251 and Phe252 are located on the flexible olive-colored loop. We speculate that after the SAM enters the binding pocket, its interactions with Asp251 and Phe252 drive the tightening of the flexible loop and closure of the binding pocket. The MfnG *N*-terminal substrate-binding domain contains almost no residues that are conserved with other OMeTs, consistent with the unique nature of its substrate specificity.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To better understand tyrosine binding in MfnG, we soaked the substrate L-tyrosine into SAH/MfnG crystals in order to determine the holoenzyme crystal structure with both SAH and tyrosine bound (PDB: 7UX8). The tyrosine substrate fits well into the cavity adjacent to the SAH binding site. Interestingly, both protomers of the dimeric MfnG contribute to the interaction with tyrosine. Specifically, Lys122 and Ser322 from one protomer bind to the carboxyl and amino groups of the substrate tyrosine, respectively (Figure 3e). The phenolic oxygen of Tyr32′ from the other protomer stabilizes the tyrosine substrate by forming a strong hydrogen bond with its carboxyl group (Figure 3e). The amino and carboxyl groups of SAH are further stabilized by two water molecules which also hydrogen bond to the side chains of His125, Asp126, Trp134, and Trp39′ (from the other protomer; Figures S4 and S5). MfnG contains a hydrophobic pocket between two α-helices to accommodate the tyrosine aromatic core (Figure 3f). MfnG is likely to tolerate a broad range of benzylic substrates with different substituent groups since there is no obvious electronic interaction between MfnG and the aromatic region of the tyrosine. We observed a niacin-like metabolite from expression in *E. coli* copurifying and bound in the acceptor binding site, which speaks to its binding promiscuity (Figure S6). In terms of a possible catalytic mechanism, His271 and Asp299 may form a hydrogen bond network that is involved in the catalytic cycle (Figure 3a,g), where His271 is conserved among Class II OMeT and often has carboxylic acid hydrogen-bonding the histidine (Figure S3).

2.3 | Engineering *E. coli* for production of OMeY containing proteins

To investigate the efficiency and specificity of incorporating biosynthesized OMeY into proteins in response to the amber codon, *E. coli* BL21 (DE3) cells were transformed with the three plasmids pBad-MfnG, pUltra-polyRS, and pLei-sfGFP-D134* (Figure 4a). OMeY-containing sfGFP protein (sfGFP-D134OMeY) was expressed by induction of these transformed cells, followed by Ni-NTA affinity purification of the GFP product. Cells containing pUltra-polyRS, and pLei-sfGFP-D134* with or without exogenously fed 1 mM OMeY were used as positive and negative controls, respectively, for comparison with cells autonomously biosynthesizing OMeY. SDS-PAGE gel analysis revealed that full-length sfGFP was only observed in the presence of exogenously fed 1 mM OMeY or when the biosynthesis of OMeY was induced (Figure 4b). To our delight, more than 1.6-fold sfGFP-D134OMeY protein was obtained from cells biosynthesizing OMeY than from cells fed with 1 mM OMeY. ESI-MS analysis afforded two observed masses of 27,827 Da (without *N*-terminal methionine) and 27,958 Da, which is in good agreement with the calculated masses as well as the mass of sfGFP-D134OMeY obtained from exogenously fed cells (Figure 4c,d). The site-specific incorporation of OMeY at 134 position was confirmed by MS/MS analysis (Figure S7).

To determine if biosynthesized OMeY can be incorporated at multiple sites in the same protein, we generated pLei plasmids encoding sfGFP proteins harboring one (Asp134TAG), two (Asp134TAG/Val2TAG), or three (Asp134TAG/Val2TAG/Tyr152TAG) amber codons at permissive sites. *E. coli* BL21 (DE3) cells transformed with a pLei plasmid, pUltra-polyRS, and pBad-MfnG were grown in 2xYT medium with 0.2% L-arabinose for 16 h at 30°C to allow protein expression. For comparison with the previously reported suppression system, cells with and without feeding of 1 mM OMeY were used as controls.[9,40] The autonomously biosynthetic cells exhibited a 2–2.5 fold higher fluorescence signal than the cells fed exogenously with OMeY (Figure 4e), which is consistent with the SDS-PAGE gel analysis. The yields of single, double, and triple OMeY-containing sfGFP mutants obtained from OMeY biosynthetic cells were 16, 12, and 10 mg L^{-1}, compared to feeding control 11, 7, and 4 mg L^{-1}, respectively (Figure 4f). The significant increase of protein expression level can be explained by the much higher intracellular OMeY concentration at the log phase for cells with biosynthetic OMeY than exogenously fed OMeY (Figure S1). ESI-MS analysis confirmed the presence of the expected number of OMeY residues in each case (Figure S8). Thus, the autonomous OMeY biosynthetic *E. coli* cells exhibited improved performance compared with the existing amber suppression system utilizing exogenously fed OMeY, permitting this ncAA to be incorporated with high efficiency at both single and multiple sites.

2.4 | Generation of autonomous mammalian cells using OMeY as a 21st amino acid

Using genetic code expansion to engineer eukaryotic proteins containing ncAAs is of great utility for both biochemical and biomedical research. Despite the progress of this technology in eukaryotes, codon suppression systems in mammalian cells have exclusively relied on exogenous feeding of high concentrations of chemically synthesized ncAAs for successful uptake and utilization of the ncAAs by eukaryotic cells. To explore the value and feasibility of using mammalian cells with the

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

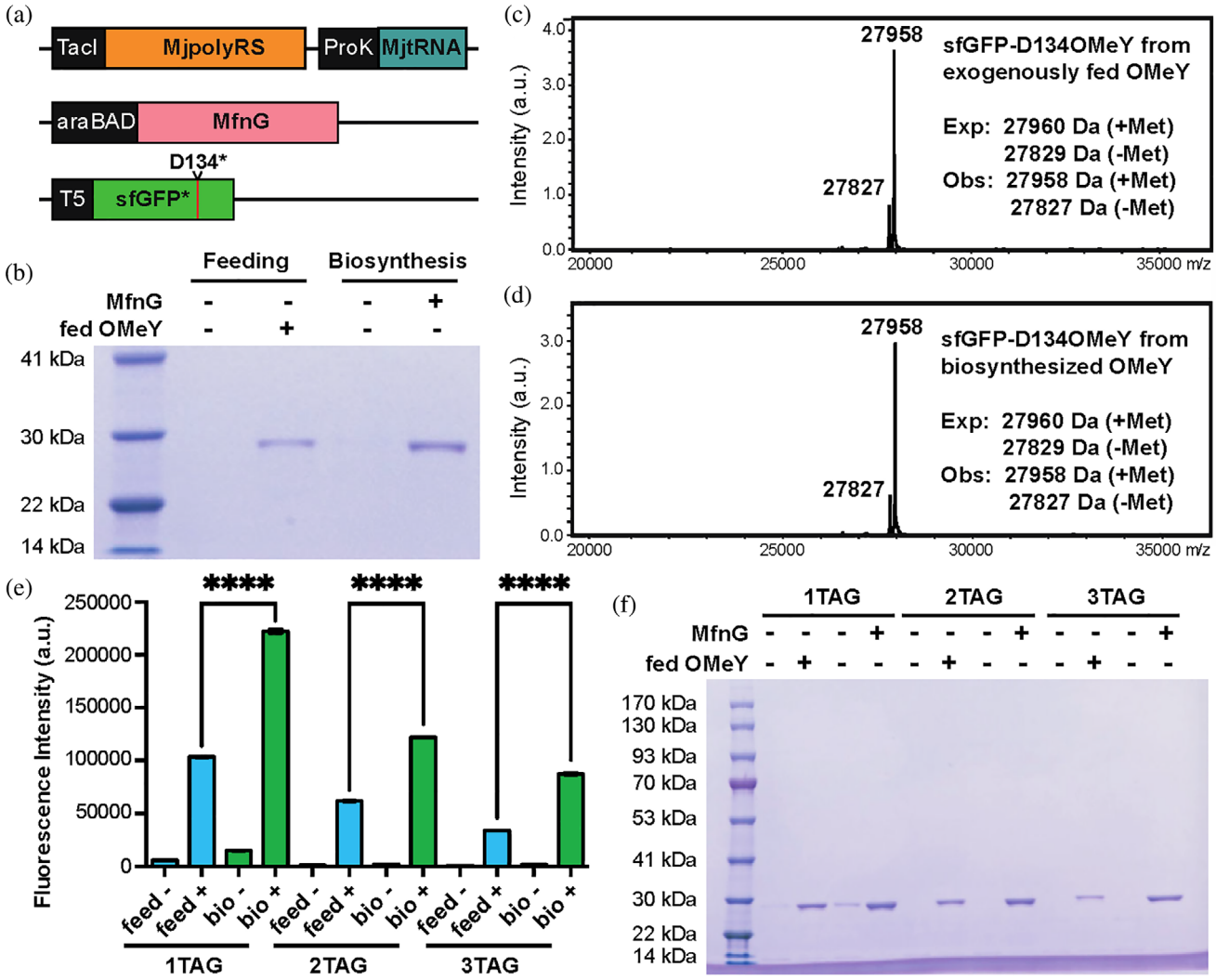

FIGURE 4 Biosynthesis and incorporation of OMeY into protein in *E. coli*. (a) The plasmid constructs used for the genetic incorporation of OMeY in *E. coli*. pUltra-polyRS encodes a TacI promoter-driven polyRS (polyRS is derived from a *Methanococcus jannaschii* TyrRS) and a Prok promoter-driven MjtRNA$_{Tyr}^{CUA}$. pBad-MfnG encodes an araBad promoter-driven MfnG. pLei-sfGFP-D134* encodes a T5 promoter-driven sfGFP with an amber codon at Asp134 position. (b) SDS-PAGE analysis of the mutant sfGFP-D134OMeY proteins from cells in the presence (+) or absence (−) of exogenously fed OMeY or biosynthesized OMeY. (c,d) Mass spectra of sfGFP-D134OMeY expressed from *E. coli* with exogenously fed OMeY or biosynthesized OMeY. (e) Fluorescence intensity measurement of BL21 (DE3) cells containing pUltra-polyRS and pLei-sfGFP* containing different amounts of TAG mutations in the presence (+) or absence (−) of OMeY or MfnG. (f) SDS-PAGE analysis of the purified sfGFP mutants from (e). ****$p < 0.0001$. p values were calculated by two-tailed unpaired student's t tests

autonomous ability to biosynthesize the ncAA and genetically incorporate it into proteins, we used a recently reported mammalian suppression system for OMeY.[40] This system encodes an enhanced green fluorescent protein (EGFP) with a Tyr39TAG mutation and a *C*-terminal hexahistidine affinity tag, two copies each of the *E. coli* and *Bacillus stearothermophilus* tRNA$_{Tyr}^{CUA}$, and a mutant *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS), evolved to charge the tRNA with OMeY. The resulting plasmid (pAcBac2.tR4-OMeYRS/EGFP*) was cotransfected into HEK293T cells along with pLenti-MfnG harboring a CMV-driven MfnG expression cassette (Figure 5a). As a control, the empty pLenti vector was also co-transfected into HEK293T cells with pAcBac2.tR4-OMeYRS/EGFP*. EGFP expression in the presence or absence of 1 mM OMeY was monitored by fluorescence microscopy. After culturing for 48 h, we observed strong green fluorescence in cells harboring MfnG as well as in cells fed exogenously with 1 mM OMeY, indicating that the transfected MfnG can efficiently *O*-methylate tyrosine in mammalian cells (Figure S9A). Furthermore, SDS-PAGE gel analysis confirmed that biosynthesized

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

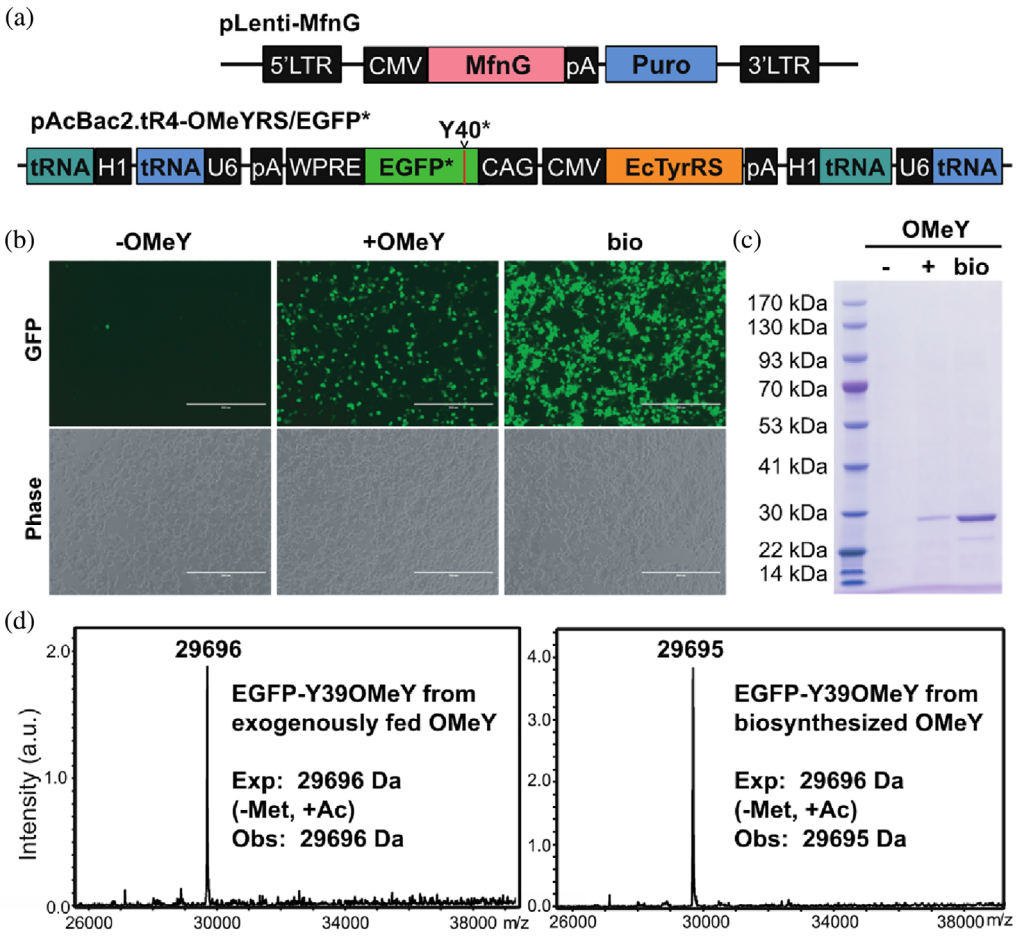

FIGURE 5 Biosynthesis and incorporation of OMeY in HEK293T protein. (a) pLenti-MfnG plasmid encodes a CMV promoter-driven MfnG. pAcBac2.tR4-OMeYRS/EGFP* encodes a CAG-promoter-driven EGFP mutant, CMV-promoter-driven *Ec*TyrRS, and two copies of *E. coli* $tRNA_{Tyr}^{CUA}$ and *B. stearothermophilus* $tRNA_{Tyr}^{CUA}$. (b) Fluorescence imaging of EGFP signal of HEK293T cells with fed OMeY and HEK293T-MfnG transfected with pAcBac2.tR4-OMeYRS/EGFP*. (c) SDS-PAGE gel analysis of purified EGFP mutants with OMeY from cells with exogenously fed OMeY (+OMeY) or biosynthesized OMeY (bio). (d) Mass spectra of EGFP mutants purified from HEK293T (left) or HEK293T-MfnG (right)

OMeY was incorporated into purified EGFP (Figure S9B). To avoid the heterogeneous expression of MfnG due to transient transfection, we subsequently used lentivirus to stably integrate MfnG into the HEK293T genome, creating a stable, homogeneous cell line for the biosynthesis of OMeY. For this purpose, the MfnG-containing lentivirus was incubated with HEK293T for 24 h, followed by puromycin selection to identify MfnG-containing HEK293T cells (HEK293T-MfnG). To examine the efficiency of incorporating biosynthesized OMeY into protein in this stable cell line, HEK293T-MfnG cells were transfected with pAcBac2.tR4-OMeYRS/EGFP*. HEK293T cells capable of autonomous OMeY biosynthesis exhibited a significant enhancement in EGFP expression compared to HEK293T cells fed exogenously with 1 mM OMeY (Figure 5b). SDS-PAGE gel and ESI-MS analysis confirmed that biosynthesized OMeY was incorporated into purified EGFP with a much higher yield compared to feeding control (Figure 5c,d). Flow cytometry analysis indicated that more than 60% of HEK293T-MfnG cells (with biosynthesized OMeY) were transfected with pAcBac2.tR4-OMeYRS/EGFP* expressed EGFP, whereas fewer than 36% of HEK293T cells were fluorescent in the presence of exogenously fed 1 mM OMeY (Figure S10). The average EGFP intensity for HEK293T-MfnG was 2.4 times higher compared to HEK293T (Figure S10). The EGFP proteins with biosynthesized or exogenously supplied OMeY were isolated by affinity chromatography using Ni-NTA 48 h post-transfection. The yield of purified EGFP from HEK293T-MfnG cells was significantly higher (ca. 3.6 μg/10^7 cells) than that obtained from the previously reported system using exogenously fed OMeY (ca. 1.1 μg/10^7 cells). ESI-mass spectrometric analysis of both proteins yielded observed masses in agreement with the calculated mass. Furthermore, we did not observe any toxicity due to MfnG overexpression (Figure S11).

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

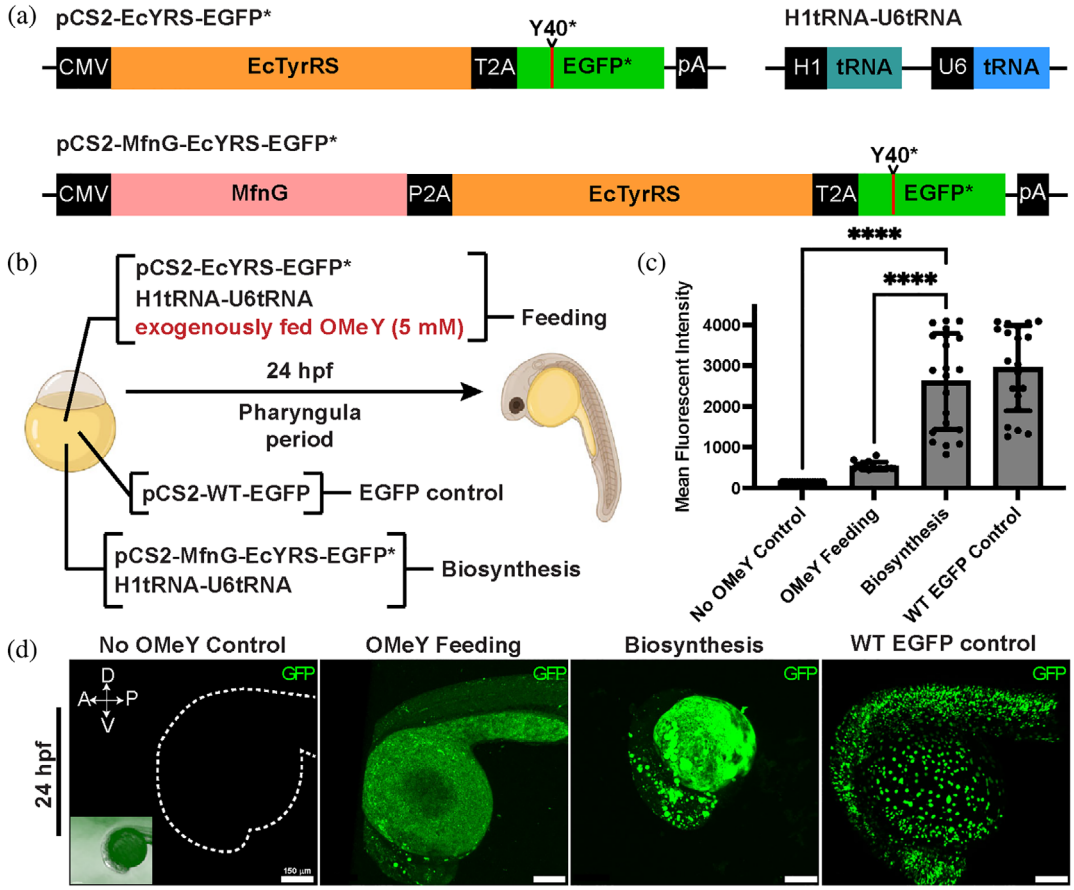

FIGURE 6 Biosynthesis and incorporation of OMeY in zebrafish protein. (a) Plasmid constructs used in zebrafish microinjections. pCS2-EcYRS-EGFP* encodes a CMV-promoter-driven EcTyrRS and mutant EGFP expression cassette. pCS2-MfnG-EcYRS-EGFP* encodes a CMV-promoter-driven MfnG, EcTyrRS, and mutant EGFP expression cassette. H1tRNA-U6tRNA encodes an *E. coli* $tRNA_{Tyr}^{CUA}$ and *B. stearothermophilus* $tRNA_{Tyr}^{CUA}$ driven by U6 and H1 promoter, respectively. (b) Schematic illustration of the workflow to incorporate OMeY in zebrafish protein. (c) Quantitative analysis of EGFP fluorescence signal in the presence or absence of exogenously fed OMeY or biosynthesized OMeY in zebrafish embryo at 24 hpf. (d) Fluorescence images of zebrafish embryos in the presence or absence of exogenously fed OMeY or biosynthesized OMeY at 24 hpf. ****$p < 0.0001$. p values were caulculated by two-tailed unpaired student's t tests. Scale bar = 150 µm

2.5 | Genetic code expansion in zebrafish using biosynthesized OMeY

Genetic code expansion technology has recently been expanded for use in multicellular organisms, including *Caenorhbditis elegans*, *Drosophila melanogaster*, *Danio rerio*, and *Mus musculus*. To enhance the cellular availability of ncAAs for protein production in intact organisms, direct injections of ncAAs at high concentrations are required. We propose that the generation of multicellular organisms with the ability to autonomously biosynthesize ncAAs, and use them for protein synthesis, will greatly facilitate the application of genetic code expansion technology to whole organisms. To illustrate this principle, we have chosen zebrafish as a model animal. Zebrafish are a popular vertebrate model organism for *in vivo* imaging due to their embryonic transparency and *ex utero* development. In addition, many molecular and cellular mechanisms are conserved between zebrafish and mammals. Importantly, the microinjection of the desired mRNA and/or plasmid DNA into the 1- or 2-cell stage of the embryo is a standard approach for introducing extrinsic genes for transcript and protein expression.[41] To generate zebrafish embryos with the ability to synthesize and genetically incorporate an ncAA into protein, we constructed a plasmid (pCS2-MfnG-EcYRS-EGFP*) containing CMV promoter-driven MfnG, *Ec*TyrRS, as well as an EGFP gene with an amber codon at a permissive site (Tyr40TAG) (Figure 6a). As for feeding control, we constructed another plasmid (pCS2-EcYRS-EGFP*) without the MfnG (Figure 6a). Compared to negative controls lacking OMeY, animals injected with MfnG exhibited a significant increase in EGFP fluorescence 24 h post-fertilization (hpf), suggesting the successful incorporation of OMeY into protein in

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

response to the amber codon (Figure 6b). Furthermore, zebrafish that biosynthesized OMeY exhibited a much higher level of fluorescence than the control fish fed with synthetic OMeY in E3 embryo media (Figure 6c,d).

3 | DISCUSSION

To date, more than 200 ncAAs have been genetically incorporated into proteins using genetic code expansion technology. However, the current technology requires efficient cellular uptake of chemically synthesized ncAAs exogenously fed at high concentrations, greatly limiting the efficiency and utility of the approach. The disadvantage of exogenous ncAA feeding is even more significant when this technology is applied to complex eukaryotic organisms. For example, the genetic code has recently been expanded to include ncAAs in the model organisms *C. elegans* and *D. rerio*. To achieve optimal ncAA incorporation efficiency, supplementation of the growth solution with 2.5–5 mM ncAA is required.[12,16] The site-specific incorporation of ncAAs into proteins in *D. melanogaster* has been demonstrated at different developmental stages and in subsets of cells within tissues.[13] However, this requires the presence of 10 mM ncAA in the food.[13] In a study of zebrafish embryo signaling pathways, ncAAs were delivered by microinjection in order to reach a whole embryo concentration of 1 mM.[14] For genetic code expansion in the widely used *M. musculus* model, a high ncAA dose of 50 mg/mouse was administered daily via intraperitoneal injection.[42] These examples demonstrate that significantly large quantities of exogenous ncAAs are required for genetic code expansion in eukaryotic organisms, suggesting that the poor pharmacokinetics and bioavailability of ncAAs seriously limit the efficient incorporation of ncAAs into proteins in multicellular organisms. However, a high dose of ncAA might cause systematic toxicity to the animals, with examples showing that high extracellular ncAA concentration could reduce the animal viability and affect the encoding yield.[16,43] Thus, the creation of cells and organisms with the endogenous ability to biosynthesize and use ncAAs for protein synthesis will not only enhance ncAA incorporation efficiency but also facilitate the application of genetic code expansion to whole animals.

In this work, we have generated multiple "completely autonomous" species that possess the biosynthetic and translational machinery for making proteins that contain the 21st amino acid OMeY. This is the first systematic approach to creating functional and completely autonomous eukaryotic organisms that make proteins containing an ncAA. Excitingly, the endogenous biosynthesis of ncAA generates higher intracellular ncAA levels than attained through exogenous ncAA feeding, leading to greater efficiency of the genetic code expansion technology. Importantly, the limited bioavailability of exogenously fed ncAAs in multicellular systems can also be overcome by endogenous biosynthesis of the ncAAs, as illustrated by the significant increase in yield of ncAA-containing proteins in autonomous zebrafish. These advantages of the autonomous system not only allow for enhanced efficiency of ncAA incorporation but also open new opportunities in multicellular systems for encoding ncAAs with poor bioavailability. For example, several types of biologically important post-translational amino acid modifications, including phosphorylation, sulfation, and glycosylation, have never been studied at the organismal level. The poor membrane permeability of these charged and hydrophilic ncAAs worsen their bioavailability in exogenous feeding models but is overcome in our autonomous ncAA models. We are now in the process of creating cells with these types of 21st amino acid modifications.

4 | METHODS

4.1 | Expression and purification of sfGFPs in *E. coli*

For feeding control, plasmid pUltra-polyRS, pBad-empty control, and pLei-sfGFP-D134* (or pLei-sfGFP-D134*V2*, pLei-sfGFP-D134*V2*Y152*) was cotransformed into BL21(DE3) cells; for experimental group, plasmid pUltra-polyRS, pBad-MfnG, and pLei-sfGFP-D134* (or pLei-sfGFP-D134*V2* and pLei-sfGFP-D134*V2*Y152*) was co-transformed into BL21(DE3) cells. The transformed BL21(DE3) cells were cultured in 2xYT medium at 37°C, 210 rpm, and only the feeding control was supplemented with 1 mM OMeY. When the OD600 of the culture reached 0.6, protein expression was induced by the addition of IPTG and L-(+)-arabinose to a final concentration of 1 mM and 0.2% w/v, respectively, and cells were grown for an additional 16 h at 30°C, 210 rpm. Cells were harvested by centrifugation at 4750 × g for 10 min and lysed with BugBuster® Protein Extraction Reagent (with 10 units/ml Pierce Universal Nuclease) following the manufacturer's instructions. Protein purification was carried out with HisPur™ Ni-NTA Resin following the manufacturer's instructions. The purified protein was then analyzed by SDS-PAGE and ESI-MS.

4.2 | Expression and purification of EGFP in HEK293T cells

HEK293T or HEK293T-MfnG cells were cultured in DMEM medium supplemented with 10% FBS and 1%

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

antibiotic–antimycotic at 37°C in a humidified chamber with 5% CO_2. After seeding a 6-well plate with 4×10^5 cells/well overnight, cells were transfected with pAcBac2.tR4-OMeYRS/EGFP* using Lipofectamine 2000 (Thermo Fisher Scientific) in the presence or absence of OMeY. The images were taken at 48 h post-transfection using AMG EVOS FL Fluorescence Imaging Microscope. The medium was removed, and the cells were washed gently with PBS for three times. The cells were lysed by pipetting 1 ml of Mammalian Cell PE LB™ (G Bioscience) following the manufacturer's instructions. Protein purification was carried out with HisPur™ Ni-NTA Resin, and the products were analyzed by SDS-PAGE and ESI-MS.

4.3 | Crystallization of MfnG

Purified MfnG protein (~11 mg/ml in 10 mM HEPES, pH 7.4) was pre-incubated with 2 mM SAM prior to crystal tray assembly. Crystal screening trays were set up with a mosquito LCP pipetting robot (SPT Labtech, Boston, MA, USA) and commercially available screens including Wizard Classic 1 and 2 and Wizard Classic 3 and 4 (Rigaku Reagents, Inc., Bainbridge Island, WA, USA), PEGs suite and PEGs suite II (NeXtal Biotechnologies, Holland, OH, USA), Morpheus and MIDAS (Molecular Dimensions, Holland, OH, USA), and PEGRx-HT and IndexHT (Hampton Research, Alisa Viejo, CA, USA). Protein crystals were grown via the sitting drop method of vapor diffusion at a 1:1 (v:v) ratio of protein/precipitant (200 nL/200 nL), at 20°C. Crystals grew from several conditions including with precipitant solutions containing 0.1 M Tris, pH 8.0, 30% (w/v) polyethylene glycol (PEG) monomethylether 2000 (PEGRx condition B12) or containing 0.2 M ammonium formate, 20% (w/v) PEG 3350 (PEGs suite I condition F11), or containing 0.1 M Tris pH 8.0, 28% (w/v) PEG 4000 (PegRx condition C09).

4.4 | X-ray diffraction, data collection, and processing

Crystals were screened for diffraction at the Advanced Photon Source (APS, Argonne National Lab) beamlines LS-CAT 21-ID-D and 21-ID-F and GM/CA-CAT beamline 23-ID-B and at the National Synchrotron Light Source (NSLS-II, Brookhaven National Lab) FMX beamline. Full datasets from the best crystals were collected at LS-CAT 21-ID-D with an Eiger X 9 M detector (Dectris). A native data set was collected from a crystal that diffracted beyond 1.35 Å resolution and was indexed in the $P2_12_12_1$ space group. A second crystal diffracted beyond 1.2 Å and also indexed in the $P2_12_12_1$ space group with different cell lengths. The data were integrated with *XDS*[44] and scaled using *aimless*[45] and *Staraniso*[46] within the *AutoPROC*[47] data processing pipeline (Table S2).

4.5 | Structure phasing and refinement

The structure was phased using MR. A model was generated using the *FFAS* model server[48,49] from the putative OMeTs (NPUN_R0239) from *Nostoc punctiforme* PCC 73102 (PDB: 2R3S; Joint Center for Structural Genomics, unpublished) with side chains built using *SCRWL 4*.[50] MR trials using the full-length homology model were unsuccessful. However, after the model was trimmed to the C-terminal Rossmann-fold domain (residues 164–365), *Phaser*[51] found a solution. Taking account of the identified translational NCS (Patterson peak 49% of origin at 0.135, 0.5, and 0.0), *Phaser* successfully placed one chain and its tNCS-related chain giving a TFZ of 11.1, an LLG of 133, and an R-factor of 0.56 at 2 Å. This partial model was refined through iterative density modification and poly-alanine auto-tracing using *SHELXE*[52] resulting in 660 residues traced with a final CC of 43%. *ARP/wARP*[53] was used to rebuild the model with side chains. *Coot*[54] and *Phenix.refine*[55] were used for rebuilding and structure refinement. Electron density revealed density for two ligands. One was clearly SAH, suggesting the added SAM had lost its reactive methyl group during the weeks of crystal growth. The second appears to be a compound that co-purified with the MfnG from the expression in *E. coli*. The size, shape, and binding environment are consistent with niacin, nicotinamide, benzoate, nitrobenzene, or some similar compound (Figure S6A). This compound was modeled as an unknown ligand since we could not distinguish between these options from the electron density and binding environment.

The second crystal form of MfnG diffracted to higher resolution and was phased by MR with *phaser* using a dimer search model from the first crystal form of MfnG. This second form also contains a dimer in the asymmetric unit in space group $P2_12_12_1$, but the unit cell lengths are different and the packing arrangement does not show the tNCS observed in form I. The maps showed clear density for the SAH as well as the same copurified metabolite observed in the form I crystals (Figure S6B). The final model was refined with restrained anisotropic B-factors. The final refinement statistics are shown in (Table S3).

4.6 | Microinjection of zebrafish embryos

For all microinjection experiments, constructs were injected during the one-cell stage and embryos were

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cultured in E3 embryo media. 30 picograms (pg) of the suppressor tRNA (H1tRNA-U6tRNA) and 30 pg each of the following constructs were injected: pCS2-MfnG-EcYRS-EGFP*, pCS2-EcYRS-EGFP*, and pCS2-WT-EGFP. For the feeding control (pCS2-EcYRS-EGFP*), 5 mM OMeY was added to the E3 embryo media. Embryos were imaged at 24 hpf.

AUTHOR CONTRIBUTIONS

Kuan-Lin Wu: Data curation (equal); formal analysis (equal); investigation (equal); writing – original draft (equal); writing – review and editing (equal). **Joshua A. Moore:** Data curation (equal); formal analysis (equal); writing – original draft (equal). **Mitchell D. Miller:** Data curation (equal); formal analysis (equal); investigation (equal); writing – original draft (equal); writing – review and editing (equal). **Yuda Chen:** Data curation (equal); formal analysis (equal); investigation (equal). **Catherine Lee:** Data curation (equal); formal analysis (equal); investigation (equal). **Weijun Xu:** Data curation (equal); formal analysis (equal); investigation (equal). **Zane Peng:** Data curation (equal); formal analysis (equal). **Qinghui Duan:** Data curation (equal); formal analysis (equal); investigation (equal). **George N. Phillips:** Funding acquisition (equal); project administration (equal); supervision (equal); writing – original draft (equal). **Rosa A. Uribe:** Formal analysis (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Han Xiao:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We thank Dr. Xiao Laboratory members for their insightful comments. This work was supported by the Cancer Prevention & Research Institute of Texas (CPRIT RR170014 to H.X. and CPRIT RR170062 to R.A.U.), National Institutes of Health (R35-GM133706, R21-CA255894, and R01-AI165079 to H.X., F31-HD104474 to J.A.M., R01-GM115261 and R01-CA217255 to G.N.P., and R01-DK124804 to R.A.U.), the Robert A. Welch Foundation (C-1970 to H.X.), the National Science Foundation (STC 1231306 to G.N.P.), and U.S. Department of Defense (W81XWH-21-1-0789 to H.X.). H.X. and R.A.U. are Cancer Prevention & Research Institute of Texas (CPRIT) Scholars in Cancer research. We thank Christopher Pennington and Rice University Mass Spectrometer Facility for conducting MS/MS experiments. This research used resources of the Advanced Photon Source (APS), a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357 and of the National Synchrotron Light Source II, a DOE Office of Science User Facility operated by Brookhaven National Laboratory (BNL) under Contract No. DE-SC0012704. Use of the APS LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant No. 085P1000817). The use of GM/CA. at the APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (Grant No. AGM-12006 and P30GM138396). The FMX beamline use at BNL Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 (Grant No. P30GM133893), and the DOE Office of Biological and Environmental Research (KP1605010). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

Atomic coordinates and structure factors were deposited into the Protein Data Bank (PDB) under theaccession code 7UX6, 7UX7, and 7UX8.

ORCID

Kuan-Lin Wu https://orcid.org/0000-0003-4691-7706
Joshua A. Moore https://orcid.org/0000-0001-7879-0135
Mitchell D. Miller https://orcid.org/0000-0003-1626-4943
Yuda Chen https://orcid.org/0000-0002-5399-1720
Weijun Xu https://orcid.org/0000-0002-0553-9908
Qinghui Duan https://orcid.org/0000-0001-6457-9777
George N. Phillips Jr https://orcid.org/0000-0002-4171-4603
Rosa A. Uribe https://orcid.org/0000-0002-0427-4493
Han Xiao https://orcid.org/0000-0002-4311-971X

REFERENCES

1. Wan W, Tharp JM, Liu WR. Pyrrolysyl-TRNA Synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta Proteins Proteom. 2014;1844(6): 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.
2. Xiao H, Schultz PG. At the Interface of chemical and biological synthesis: An expanded genetic code. Cold Spring Harb Perspect Biol. 2016;8(9):a023945. https://doi.org/10.1101/cshperspect.a023945.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. Niu W, Guo J. Expanding the chemistry of fluorescent protein biosensors through genetic incorporation of unnatural amino acids. Mol Biosyst. 2013;9(12):2961–2970. https://doi.org/10.1039/c3mb70204a.
4. Young DD, Schultz PG. Playing with the molecules of life. ACS Chem Biol. 2018;13(4):854–870. https://doi.org/10.1021/acschembio.7b00974.
5. Agostini F, Völler J-S, Koksch B, Acevedo-Rocha CG, Kubyshkin V, Budisa N. Biocatalysis with unnatural amino acids: Enzymology meets Xenobiology. Angew Chem Int Ed. 2017;56(33):9680–9703. https://doi.org/10.1002/anie.201610129.
6. Porter JJ, Mehl RA. Genetic code expansion: A powerful tool for understanding the physiological consequences of oxidative stress protein modifications. Oxid Med Cell Longev. 2018;2018:e7607463. https://doi.org/10.1155/2018/7607463 1–14.
7. Huang Y, Liu T. Therapeutic applications of genetic code expansion. Synth Syst Biotechnol. 2018;3(3):150–158. https://doi.org/10.1016/j.synbio.2018.09.003.
8. Chin JW. Expanding and reprogramming the genetic code. Nature. 2017;550(7674):53–60. https://doi.org/10.1038/nature24031.
9. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia Coli. Science. 2001;292(5516):498–500. https://doi.org/10.1126/science.1060077.
10. Chin JW. Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem. 2014;83(1):379–408. https://doi.org/10.1146/annurev-biochem-060713-035737.
11. Brown W, Liu J, Deiters A. Genetic code expansion in animals. ACS Chem Biol. 2018;13(9):2375–2386. https://doi.org/10.1021/acschembio.8b00520.
12. Greiss S, Chin JW. Expanding the genetic code of an animal. J Am Chem Soc. 2011;133(36):14196–14199. https://doi.org/10.1021/ja2054034.
13. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of drosophila melanogaster. Nat Chem Biol. 2012;8(9):748–750. https://doi.org/10.1038/nchembio.1043.
14. Liu J, Hemphill J, Samanta S, Tsang M, Deiters A. Genetic code expansion in zebrafish embryos and its application to optical control of cell signaling. J Am Chem Soc. 2017;139(27):9100–9103. https://doi.org/10.1021/jacs.7b02145.
15. Ernst RJ, Krogager TP, Maywood ES, et al. Genetic code expansion in the mouse brain. Nat Chem Biol. 2016;12(10):776–778. https://doi.org/10.1038/nchembio.2160.
16. Chen Y, Ma J, Lu W, et al. Heritable expansion of the genetic code in mouse and zebrafish. Cell Res. 2017;27(2):294–297. https://doi.org/10.1038/cr.2016.145.
17. Völler J-S, Budisa N. Coupling genetic code expansion and metabolic engineering for synthetic cells. Curr Opin Biotechnol. 2017;48:1–7. https://doi.org/10.1016/j.copbio.2017.02.002.
18. Mehl RA, Anderson JC, Santoro SW, et al. Generation of a bacterium with a 21 amino acid genetic code. J Am Chem Soc. 2003;125(4):935–939. https://doi.org/10.1021/ja0284153.
19. Chen Y, Loredo A, Gordon A, et al. A noncanonical amino acid-based relay system for site-specific protein labeling. Chem Commun. 2018;54(52):7187–7190. https://doi.org/10.1039/C8CC03819H.
20. Zhang MS, Brunner SF, Huguenin-Dezot N, et al. Biosynthesis and genetic encoding of Phosphothreonine through parallel selection and deep sequencing. Nat Methods. 2017;14(7):729–736. https://doi.org/10.1038/nmeth.4302.
21. Chen Y, Tang J, Wang L, et al. Creation of bacterial cells with 5-Hydroxytryptophan as a 21st amino acid building block. Chem. 2020;6(10):2717–2727. https://doi.org/10.1016/j.chempr.2020.07.013.
22. Chen Y, Loredo A, Chung A, Zhang M, Liu R, Xiao H. Biosynthesis and genetic incorporation of 3,4-Dihydroxy-L-phenylalanine into proteins in Escherichia Coli. J Mol Biol. 2021;434(8):167412. https://doi.org/10.1016/j.jmb.2021.167412.
23. Jung J-E, Lee SY, Park H, et al. Genetic incorporation of unnatural amino acids biosynthesized from α-Keto acids by an aminotransferase. Chem Sci. 2014;5(5):1881. https://doi.org/10.1039/c3sc51617b.
24. Kim S, Sung BH, Kim SC, Lee HS. Genetic incorporation of L-Dihydroxyphenylalanine (DOPA) biosynthesized by a tyrosine phenol-Lyase. Chem Commun. 2018;54(24):3002–3005. https://doi.org/10.1039/C8CC00281A.
25. Wang Y, Chen X, Cai W, et al. Expanding the structural diversity of protein building blocks with noncanonical amino acids biosynthesized from aromatic thiols. Angew Chem Int Ed. 2021;60(18):10040–10048. https://doi.org/10.1002/anie.202014540.
26. Takimoto JK, Xiang Z, Kang J-Y, Wang L. Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. Chembiochem. 2010;11(16):2268–2272. https://doi.org/10.1002/cbic.201000436.
27. Bi X, Pasunooti KK, Tareq AH, Takyi-Williams J, Liu C-F. Genetic incorporation of 1,2-Aminothiol functionality for site-specific protein modification via Thiazolidine formation. Org Biomol Chem. 2016;14(23):5282–5285. https://doi.org/10.1039/C6OB00854B.
28. Chen Y, Wu K-L, Tang J, et al. Addition of Isocyanide-containing amino acids to the genetic code for protein labeling and activation. ACS Chem Biol. 2019;14(12):2793–2799. https://doi.org/10.1021/acschembio.9b00678.
29. Ko W, Kumar R, Kim S, Lee HS. Construction of bacterial cells with an active transport system for unnatural amino acids. ACS Synth Biol. 2019;8(5):1195–1203. https://doi.org/10.1021/acssynbio.9b00076.
30. Pohle S, Appelt C, Roux M, Fiedler H-P, Süssmuth RD. Biosynthetic gene cluster of the non-Ribosomally synthesized Cyclodepsipeptide Skyllamycin: Deciphering unprecedented ways of unusual hydroxylation reactions. J Am Chem Soc. 2011;133(16):6194–6205. https://doi.org/10.1021/ja108971p.
31. Liu J, Wang B, Li H, et al. Biosynthesis of the anti-infective Marformycins featuring pre-NRPS assembly line *N* -formylation and *O* -methylation and post-assembly line *C* -hydroxylation chemistries. Org Lett. 2015;17(6):1509–1512. https://doi.org/10.1021/acs.orglett.5b00389.
32. Nelson JT, Lee J, Sims JW, Schmidt EW. Characterization of SafC, a catechol 4-O-Methyltransferase involved in Saframycin biosynthesis. Appl Environ Microbiol. 2007;73(11):3575–3580. https://doi.org/10.1128/AEM.00011-07.
33. Fu C-Y. Biosynthesis of 3-Hydroxy-5-methyl-O-Methyltyrosine in the Saframycin/Safracin biosynthetic pathway. J Microbiol Biotechnol. 2009;19(5):439–446. https://doi.org/10.4014/jmb.0808.484.
34. Zallot R, Oberg N, Gerlt JA. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

databases to discover novel enzymes and metabolic pathways. Biochemistry. 2019;58(41):4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.
35. Schubert HL, Blumenthal RM, Cheng X. Many paths to Methyltransfer: A Chronicle of Convergence. Trends Biochem Sci. 2003;28(6):329–335. https://doi.org/10.1016/S0968-0004(03)00090-2.
36. Wlodarski T, Kutner J, Towpik J, et al. Comprehensive structural and substrate specificity classification of the saccharomyces cerevisiae Methyltransferome. PLoS One. 2011;6(8):e23168. https://doi.org/10.1371/journal.pone.0023168.
37. Liscombe DK, Louie GV, Noel JP. Architectures, mechanisms and molecular evolution of natural product Methyltransferases. Nat Prod Rep. 2012;29(10):1238–1250. https://doi.org/10.1039/c2np20029e.
38. Lee S, Kang J, Kim J. Structural and biochemical characterization of Rv0187, an O-Methyltransferase from mycobacterium tuberculosis. Sci Rep. 2019;9(1):8059. https://doi.org/10.1038/s41598-019-44592-7.
39. Park HL, Lee JC, Lee K, et al. Biochemical characterization of a flavonoid O-Methyltransferase from Perilla leaves and its application in 7-Methoxyflavonoid production. Molecules. 2020;25(19):4455. https://doi.org/10.3390/molecules25194455.
40. Liu W, Brock A, Chen S, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Methods. 2007;4(3):239–244. https://doi.org/10.1038/nmeth1016.
41. Rosen JN, Sweeney MF, Mably JD. Microinjection of zebrafish embryos to analyze gene function. J Vis Exp. 2009;25:e1115. https://doi.org/10.3791/1115.
42. Han S, Yang A, Lee S, Lee H-W, Park CB, Park H-S. Expanding the genetic code of Mus Musculus. Nat Commun. 2017;8(1):14568. https://doi.org/10.1038/ncomms14568.
43. Parrish AR, She X, Xiang Z, et al. Expanding the genetic code of Caenorhabditis Elegans using bacterial aminoacyl-TRNA Synthetase/TRNA pairs. ACS Chem Biol. 2012;7(7):1292–1302. https://doi.org/10.1021/cb200542j.
44. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(2):125–132. https://doi.org/10.1107/S0907444909047337.
45. Evans PR. An Introduction to data reduction: Space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr. 2011;67(4):282–292. https://doi.org/10.1107/S090744491003982X.
46. STARANISO anisotropy & Bayesian estimation server. 2022 [cited May 28, 2022]. Available from: https://staraniso.globalphasing.org/cgi-bin/staraniso.cgi.
47. Vonrhein C, Flensburg C, Keller P, et al. Data processing and analysis with the AutoPROC toolbox. Acta Crystallogr D Biol Crystallogr. 2011;67(4):293–302. https://doi.org/10.1107/S0907444911007773.
48. Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L. The importance of alignment accuracy for molecular replacement. Acta Crystallogr D Biol Crystallogr. 2004;60(7):1229–1236. https://doi.org/10.1107/S0907444904010145.
49. Jaroszewski L, Li Z, Cai X-H, Weber C, Godzik A. FFAS server: Novel features and applications. Nucleic Acids Res. 2011;39:W38–W44. https://doi.org/10.1093/nar/gkr441.
50. Krivov GG, Shapovalov MV, Dunbrack RL. Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct Funct Bioinfor. 2009;77(4):778–795. https://doi.org/10.1002/prot.22488.
51. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007;40(4):658–674. https://doi.org/10.1107/S0021889807021206.
52. Thorn A, Sheldrick GM. Extending molecular-replacement solutions with SHELXE. Acta Crystallogr D Biol Crystallogr. 2013;69(11):2251–2256. https://doi.org/10.1107/S0907444913027534.
53. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/WARP version 7. Nat Protoc. 2008;3(7):1171–1179. https://doi.org/10.1038/nprot.2008.91.
54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of coot. Acta Crystallogr D Biol Crystallogr. 2010;66(4):486–501. https://doi.org/10.1107/S0907444910007493.
55. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. Acta Crystallogr D Struct Biol. 2019;75(10):861–877. https://doi.org/10.1107/S2059798319011471.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wu K-L, Moore JA, Miller MD, Chen Y, Lee C, Xu W, et al. Expanding the eukaryotic genetic code with a biosynthesized 21st amino acid. Protein Science. 2022;31(10):e4443. https://doi.org/10.1002/pro.4443

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4443 by University At Buffalo (Suny), Wiley Online Library on [02/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License