



Towards Engineering an Orthogonal Protein Translation Initiation System

Byeong Sung Lee¹, Woon Jong Choi¹, Sang Woo Lee¹, Byoung Joon Ko² and Tae Hyeon Yoo^{1,3*}

¹Department of Molecular Science and Technology, Ajou University, Suwon, South Korea, ²School of Biopharmaceutical and Medical Sciences, Sungshin Women's University, Seoul, South Korea, ³Department of Applied Chemistry and Biological Engineering, Ajou University, Suwon, South Korea

In the last two decades, methods to incorporate non-canonical amino acids (ncAAs) into specific positions of a protein have advanced significantly; these methods have become general tools for engineering proteins. However, almost all these methods depend on the translation elongation process, and strategies leveraging the initiation process have rarely been reported. The incorporation of a ncAA specifically at the translation initiation site enables the installation of reactive groups for modification at the N-termini of proteins, which are attractive positions for introducing abiological groups with minimal structural perturbations. In this study, we attempted to engineer an orthogonal protein translation initiation system. Introduction of the identity elements of *Escherichia coli* initiator tRNA converted an engineered *Methanococcus jannaschii* tRNA^{Tyr} into an initiator tRNA. The engineered tRNA enabled the site-specific incorporation of O-propargyl-L-tyrosine (OpgY) into the amber (TAG) codon at the translation initiation position but was inactive toward the elongational TAG codon. Misincorporation of Gln was detected, and the engineered system was demonstrated only with OpgY. We expect further engineering of the initiator tRNA for improved activity and specificity to generate an orthogonal translation initiation system.

Keywords: translation initiation, non-canonical amino acid, initiator tRNA, amber codon, *Methanococcus jannaschii* tRNA

INTRODUCTION

Site-specific modification of the N-terminus of a protein is an attractive strategy to introduce unnatural groups with minimal effects on original protein functions. Protein termini tend to be exposed and flexible but not buried in the core (Christopher and Baldwin 1996; Jacob and Unger 2007). The N-terminal α -amino group is a good nucleophile, and its reactivity has been utilized to modify the N-terminus of proteins specifically. While the pKa of a lysine side-chain is around 10.5, the pKa values of the N-terminal amines have been reported between 6.8 and 9.1, with an average of 7.7 ± 0.5 (Serada and Mant, 1993). The reactivities of N-terminal and lysine side chain amines can be different at a given pH, and several chemical reactions have been reported as methods to modify the N-terminal amine groups specifically (Rosen and Francis 2017). Enzymes have been found to catalyze the peptide-bond formation reaction, and some of them have been used to modify proteins at their N-termini, including sortase (Mao et al., 2004; Proft 2010), intein (Evans and Xu 1999; Xu and Evans 2001), and proteases (Liebscher et al., 2014; Weeks and Wells 2020).

Biological methods for incorporating non-canonical amino acids (ncAAs) into proteins have become invaluable tools for protein engineering (Wang and Schultz 2004). In particular, ncAAs with

OPEN ACCESS

Edited by:

Jiantao Guo,
University of Nebraska-Lincoln,
United States

Reviewed by:

Tapan Kumar Mohanta,
University of Nizwa, Oman
Chenguang Fan,
University of Arkansas, United States

*Correspondence:

Tae Hyeon Yoo
taehyeonyoo@ajou.ac.kr

Specialty section:

This article was submitted to
Chemical Biology,
a section of the journal
Frontiers in Chemistry

Received: 08 September 2021

Accepted: 14 October 2021

Published: 26 October 2021

Citation:

Lee BS, Choi WJ, Lee SW, Ko BJ and
Yoo TH (2021) Towards Engineering
an Orthogonal Protein Translation
Initiation System.
Front. Chem. 9:772648.
doi: 10.3389/fchem.2021.772648

reactivities can provide sites for protein modification (Wang et al., 2009; Wals and Ovaas 2014; Chin 2017). *Escherichia coli* wild-type and variant methionyl-tRNA synthetase (MetRS) can charge the *E. coli* initiator methionyl-tRNA (*Ec*-tRNA^{Met}) with isostructural analogs of methionine (Met), and several analogs with orthogonal reactivities have been successfully installed at the N-termini of proteins (Link et al., 2003; Johnson et al., 2010; Agostini et al., 2017; Tharp et al., 2020; Pagar et al., 2021). However, they were also incorporated into the internal Met positions, and thus, site-specific modification at the N-terminal position was not achieved, besides cases where there was no internal Met residue. Interestingly, *Ec*-tRNA^{Met} charged with ncAA is translationally active for protein synthesis initiation in eukaryotes (Ngo et al., 2013). The system was demonstrated to tag the proteome N-termini but was unsuitable for producing homogeneous recombinant proteins because either ncAA or Met occupied the N-terminal positions. Budisa et al. proposed a clever strategy to incorporate azidohomoalanine at the protein synthesis initiation position by deleting the elongator tRNA^{Met} genes from the *E. coli* genome and introducing an orthogonal pair of MetRS/tRNA^{Met} from *Sulfolobus acidocaldarius* (Simone, Acevedo-Rocha et al., 2016). However, the system was partially successful because of the incomplete orthogonality of the system.

Orthogonal pairs of aminoacyl-tRNA synthetase (aaRS) and tRNA have been developed to introduce diverse ncAAs into proteins in a site-specific manner in various organisms (Wang and Wang 2012; Xiao et al., 2013; Smolskaya and Andreev 2019). These methods were designed based on codons that are not generally used for coding amino acids (such as stop codons (Liu and Schultz 2010) and four-base codons (Hohsaka et al., 2001; Anderson et al., 2004; Wang et al., 2012; Lee et al., 2017)). Strategies have also been reported to incorporate ncAAs into sense codons by modifying orthogonal pairs (Lee et al., 2015; Mukai et al., 2015). However, all orthogonal tRNAs are elongator tRNAs, and thus, cannot be used to install ncAAs at the N-termini of proteins. Recently, Söll et al. reported an engineered *Ec*-tRNA^{Met} that enabled the initiation of protein synthesis with ncAAs (Tharp et al., 2020). The initiator tRNA, named itRNA^{Tyr}2, had an identity element in *Methanococcus jannaschii* tRNA^{Tyr} (*Mj*-tRNA^{Tyr}) by mutations of A72G and the anticodon for the amber codon (CAU→CUA). The initiator tRNA was a substrate for *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*-TyrRS), and the tRNAs charged with ncAAs could be used for *E. coli* protein synthesis initiation. Despite this success, the initiator tRNA had the C1:G72 base pair and could not distinguish between the initiation and elongation amber codons.

It has been shown that the introduction of the main determinants of *Ec*-tRNA^{Met} into elongator tRNAs, such as glutamyl-tRNA and methionyl-tRNA, could convert them into tRNAs, enabling initiation of protein synthesis in *E. coli* (Varshney and RajBhandary 1990; Lee et al., 1991; Varshney et al., 1993). Based on the achievements of these studies, we aimed to engineer *Mj*-tRNA^{Tyr} into an initiator tRNA, which was intended to work only for an amber codon at the translation initiation position, but not for an internal amber codon

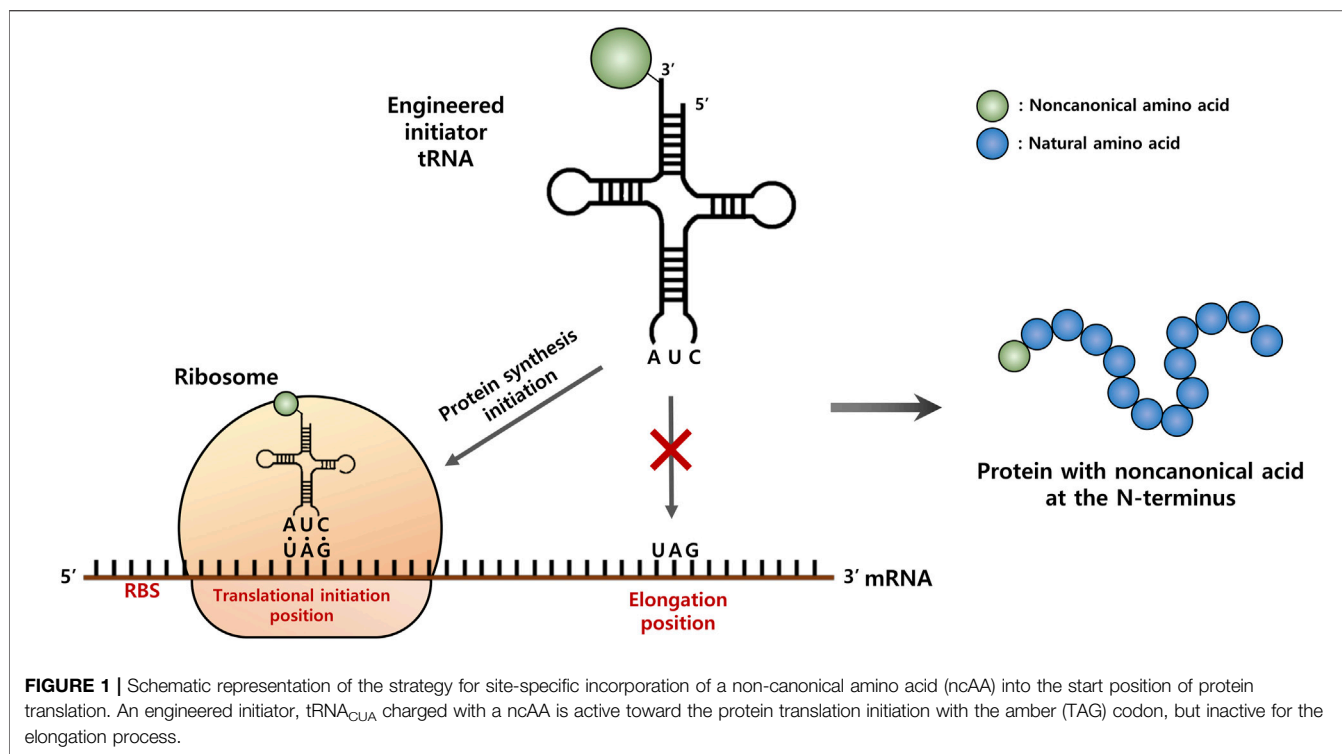
(**Figure 1**). Identity elements for interactions with factors involved in the initiation of protein syntheses, such as methionyl-tRNA transformylase (MTF), the P-site of the 30S ribosomal subunit, and initiation factor-2 (IF-2), were introduced into *Mj*-tRNA^{Tyr}. The engineered *Mj*-tRNA was active toward the initiation amber codon, but did not support the coding of internal amber codons. We believe that the results obtained in this study open a new route for developing an orthogonal translation initiation system to incorporate ncAAs at the N-termini of proteins.

MATERIALS AND METHODS

Plasmid Construction

The plasmids, primers, and proteins used in this study are shown in **Table 1**; **Supplementary Tables S1, S2**, respectively. The synthesized initiator tRNA genes, including the ProK promoter and terminator (*Mj*-itRNA-1 and *Mj*-itRNA-2; the DNA sequences are shown in **Supplementary Table S3**) were cloned into the modified pEVOL (Young et al., 2010) plasmid, in which one copy of the azidophenylalaninyl-tRNA synthetase gene (AzF-RS) is located under the AraBAD promoter (Lee et al., 2015), using the *Apa*LI and *Xho*I sites. The resulting plasmids were named pSEPL773 for *Mj*-itRNA-1 and pSPEL541 for *Mj*-itRNA-2. The methionyl-tRNA transformylase (MTF) gene was amplified (primers 1 and 6) from the chromosome of *E. coli* DH10 β and then cloned into the plasmid containing itRNA-1 (pSPEL773) or itRNA-2 (pSPEL541) using the *Nde*I and *Pst*I sites. The two *Pst*I sites present in the MTF gene were removed by assembly PCR using primer pairs (primers 2 and 3 for the first *Pst*I site; primers 4 and 5 for the second *Pst*I site). The MTF gene was located under the *gln*S promoter. The resulting plasmids were named pSPEL527 (MTF/*Mj*-itRNA-1) and pSPEL528 (MTF/*Mj*-itRNA-2), respectively. The initiator factor-2 (IF-2) gene was amplified (primers 7 and 14) from the chromosome of *E. coli* DH10 β and then cloned into the plasmid containing itRNA-1 (pSPEL773) or itRNA-2 (pSPEL541) using the *Nde*I and *Pst*I sites; the three *Pst*I sites present in the IF-2 gene were removed by assembly PCR using two pairs of primers (primers 8 and 9 for the first *Pst*I site; primer 10 and 11 for the second *Pst*I site; primers 12 and 13 for the third *Pst*I site). The resulting plasmids were named pSPEL780 (IF-2/*Mj*-itRNA-1) and pSPEL562 (IF-2/*Mj*-itRNA-2), respectively. Plasmids expressing both MTF and IF-2 were constructed by cloning the IF-2 gene with a ribosome-binding sequence (Kinneer et al., 2019) into the *Pst*I site of pSPEL527 or pSPEL528, and the orientation of the RBS-IF-2 genes was confirmed by DNA sequencing. The resulting plasmids were named pSPEL781 (MTF/IF-2/*Mj*-itRNA-1) and pSPEL563 (MTF/IF-2/*Mj*-itRNA-2), respectively.

A DNA double helix for a multiple cloning site (*Nco*I-*Bam*HI-*Hind*III-*Not*I-*Xho*I-His₆ Tag) prepared by annealing primers 15 and 16 was phosphorylated using T4 polynucleotide kinase and then ligated into the pBbE6a (Lee et al., 2011) plasmid digested with *Nde*I and *Xho*I. The RBS-GFP gene with the TAG codon at the translation initiation site was generated by PCR amplification

**TABLE 1** | Plasmids used in this study.

Name	Characteristics ^a	Source
pEVOL	<i>Mj</i> -tRNA ^{Tyr} , AzFRS, Cm ^R , p15A ori	Young et al. (2010)
pSPEL773	pEVOL- <i>Mj</i> -itRNA-1, AzFRS, Cm ^R , p15A ori	This study
pSPEL541	pEVOL- <i>Mj</i> -itRNA-2, AzFRS, Cm ^R , p15A ori	This study
pSPEL527	pEVOL- <i>Mj</i> -itRNA-1, AzFRS, MTF, Cm ^R , p15A ori	This study
pSPEL528	pEVOL- <i>Mj</i> -itRNA-2, AzFRS, MTF, Cm ^R , p15A ori	This study
pSPEL780	pEVOL- <i>Mj</i> -itRNA-1, AzFRS, IF-2, Cm ^R , p15A ori	This study
pSPEL562	pEVOL- <i>Mj</i> -itRNA-2, AzFRS, IF-2, Cm ^R , p15A ori	This study
pSPEL781	pEVOL- <i>Mj</i> -itRNA-1, AzFRS, MTF, IF-2, Cm ^R , p15A ori	This study
pSPEL563	pEVOL- <i>Mj</i> -itRNA-2, AzFRS, MTF, IF-2, Cm ^R , p15A ori	This study
pQE-80L	Expression vector, Amp ^R , ColE1 ori	Qiagen
pSPEL542	pQE-80L-TAG-Z domain, Amp ^R , ColE1 ori	This study
pBbE6a	Expression vector, Amp ^R , ColE1 ori	Lee et al. (2011)
pSPEL530	pBbE6a-TAG-11.3.3, ColE1 ori	This study
pGEX-4T-1	Expression vector with GST, Amp ^R , pBR322 ori	Amersham Bioscience
pSPEL236	pGEX-4T-1-GST-TAG-Z domain, Amp ^R , pBR322 ori	This study
pBbS2K-ProRS	pBbS2K-Prolyl tRNA synthetase, Kan ^R , SC101 ori	Lee et al. (2015)

^a*Mj*, *Methanococcus jannaschii*; AzFRS, azidophenylalanyl tRNA synthetase; Cm, chloramphenicol; Amp, ampicillin; Kan, kanamycin; MTF, methionyl-tRNA transformylase; IF-2, initiation factor-2.

using the 11.3.3 GFP gene (Yoo et al., 2007) as a template with primers 17 and 18, and the product was cloned into the modified pBbE6a mentioned above using the *EcoRI* and *HindIII* sites, resulting in pSPEL530. The synthesized gene of the Z domain (the amino acid sequence in **Supplementary Table S2**) with an amber (TAG) codon at the translation initiation site was cloned into pQE-80L (Qiagen) using the *EcoRI* and *HindIII* sites, resulting in pSPEL542. The Z domain gene was amplified using pSPEL542 as a template with primers 19 and 20; the product was cloned into pGEX-4T-1 (Amersham Bioscience)

EcoRI and *XhoI* sites. To introduce an amber (TAG) codon between the glutathione-S-transferase (GST) and Z domain, an annealed DNA double helix of primers 21 and 22 was phosphorylated using T4 polynucleotide kinase and then cloned into the pGEX-4T-1 with the Z domain gene using the *BamHI* and *EcoRI* sites, resulting in pSPEL236.

Non-Canonical Amino Acid Incorporation

The plasmid coding Z domain (pSPEL542) or GFP (pSPEL530) with an amber (TAG) codon at the translation initiation position

was co-transformed with one of the pEVOL derivatives (pSPEL731, 541, 527, 528, 780, or 562) and pBbS2K-ProRS (Lee et al., 2015) encoding *E. coli* prolyl-tRNA synthetase under the tetracycline-inducible promoter into *E. coli* DH10 β . The cells were grown in 2xYT media containing 200 μ g/ml ampicillin, 34 μ g/ml chloramphenicol, and 35 μ g/ml kanamycin at 37°C until the OD₆₀₀ reached 0.5. To induce AzF-RS and prolyl-tRNA synthetase expression, 0.2% L-arabinose and 50 nM anhydrotetracycline were added to the culture. When OD₆₀₀ reached 1.0, 1 mM isopropyl β -D-thiogalatoside and 1 mM nCAA were added to express the model protein. After 3 h, the cells were harvested by centrifugation at 14,000 \times g at 4°C for 20 min, and the cell pellets were stored at -20°C until use.

Copper (I)-Catalyzed Click Reaction and Western Blotting

The cell pellets were resuspended in 1% SDS solution and heated at 95°C for 5 min for lysis. The supernatant obtained after centrifugation at 14,000 \times g for 5 min was subjected to a copper (I)-catalyzed click reaction with biotin-PEG₃-azide (Click Chemistry Tools) following the method described (Hong et al., 2009). The reaction mixtures were analyzed by western blotting using a streptavidin-horseradish peroxidase (HRP) conjugate (Click Chemistry Tools) and an anti-His₅-HRP conjugate (Sigma-Aldrich, St. Louis, MO).

Protein Purification

The Z-domain protein with an N-terminal His₆ tag was purified using a Ni-immobilized resin (Clontech, Mountain View, CA) under native conditions, following the manufacturer's instructions. The cell pellets stored at -20°C were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, pH 7.4) and incubated on ice for 30 min with 50 μ g/ml lysozyme. After sonication, the lysed cells were centrifuged at 9,300 \times g at 4°C for 1 h. The supernatant was incubated at 4°C for 1 h with Ni resin, pre-equilibrated with lysis buffer. The resin was loaded onto a gravity flow column (Thermo Fisher Scientific, Waltham, MA) and washed three times with washing buffer (50 mM sodium phosphate, 300 mM sodium chloride, 40 mM imidazole, pH 7.4). The protein was eluted using elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 300 mM imidazole, pH 7.4). The protein solutions were buffer-exchanged with phosphate buffer saline (PBS) solution (10 mM KH₂PO₄, 150 mM NaCl, pH 7.4) using a centrifugal filter unit (Millipore, 3000 MWCO).

Mass Spectrometry

The intact masses of the proteins were analyzed using a Waters ACQUITY I class UPLC system (Milford, MA) with an ACQUITY UPLC Protein BEH C4 column (2.1 mm \times 100 mm, 1.7 μ m particle size; Waters). The mobile phases were 0.1% formic acid in water (eluent A) and 0.1% formic acid in

acetonitrile (eluent B). The gradient applied was: 0–3 min, 5% eluent B; 3–13 min, linear increase to 50% eluent B at 0.2 ml/min. The eluent was injected into a Thermo Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) and ionized with an electrospray source. MS spectra were acquired in the mass range of 400–2,000 m/z and 120,000 resolution at m/z 200. The deconvoluted mass spectra were generated using Protein Deconvolution 2.0 (Thermo Fisher Scientific, Waltham, MA).

N-Terminal Sequencing

The purified Z-domain protein was transferred to a PVDF membrane (0.45 μ m pore size; Pall), and the membrane was washed with ultrapure water and dried in air. The membrane was subjected to an N-terminal sequencing analysis using Procise[®] LC492 Protein Sequencing System (Applied Biosystems, Waltham, MA). The retention times of phenylthiohydantoin (PTH)-amino acids were compared with those of the standards (Tokyo Chemical Industry, Tokyo). PHT-OpgY was synthesized following a previously published method (Steiman et al., 1985).

RESULTS

Introduction of the identity elements of Ec-tRNA^{Met} converted Mj-tRNA^{Tyr} into an initiator tRNA

The two main identity elements of the *E. coli* initiator tRNA (Ec-tRNA^{Met}) are 1) the absence of a Watson-Crick base pair between positions 1 and 72 in the acceptor stem (pink in **Figure 2A**) and 2) three consecutive G:C base pairs in the anticodon stem (green in **Figure 2A**). The C1:A72 mismatch plays a critical role in the interaction with MTF, and the 3G:C pairs are important for targeting the initiator tRNA to the P-site of the 30S ribosomal subunit (Kozak 1999; Laursen et al., 2005; Louise et al., 2009). It was previously demonstrated that an *E. coli* glutamyl-tRNA (Ec-tRNA^{Gln}) could be converted into an initiator tRNA by introducing these two determinants. The engineered tRNA with the CUA anticodon for the amber (TAG) nonsense codon was activated by the endogenous glutamyl-tRNA synthetase (GlnRS), and the initiator tRNA charged with Gln enabled the synthesis of recombinant proteins with a TAG codon at their initial position (Varshney et al., 1993). Based on these results, we introduced the two elements into an engineered Mj-tRNA^{Tyr} (Wang and Schultz 2001) (**Figure 2B**) by mutations of G72→A72 for the C1:A72 mismatch and A31:U39→G31:C39 for the three consecutive G:C pairs in the anticodon stem. In addition to the C1:A72 mismatch, several other elements in the acceptor and the D stem (blue in **Figure 2A**) have been reported to play a role in the formylation by MTF (RajBhandary 1994, Louise, L. et al., 2009). Based on a previously published report in which Mj-tRNA^{Tyr} was engineered for its orthogonality to the 20 endogenous aminoacyl-tRNA synthetases of *E. coli* (Guo et al., 2009), several additional mutations were introduced into the acceptor stem. Since any changes in the A11:U24 pair significantly inhibited charging the tRNA in our previous experiments (unpublished results), we decided not to introduce the element in the D stem into Mj-tRNA^{Tyr}, and the

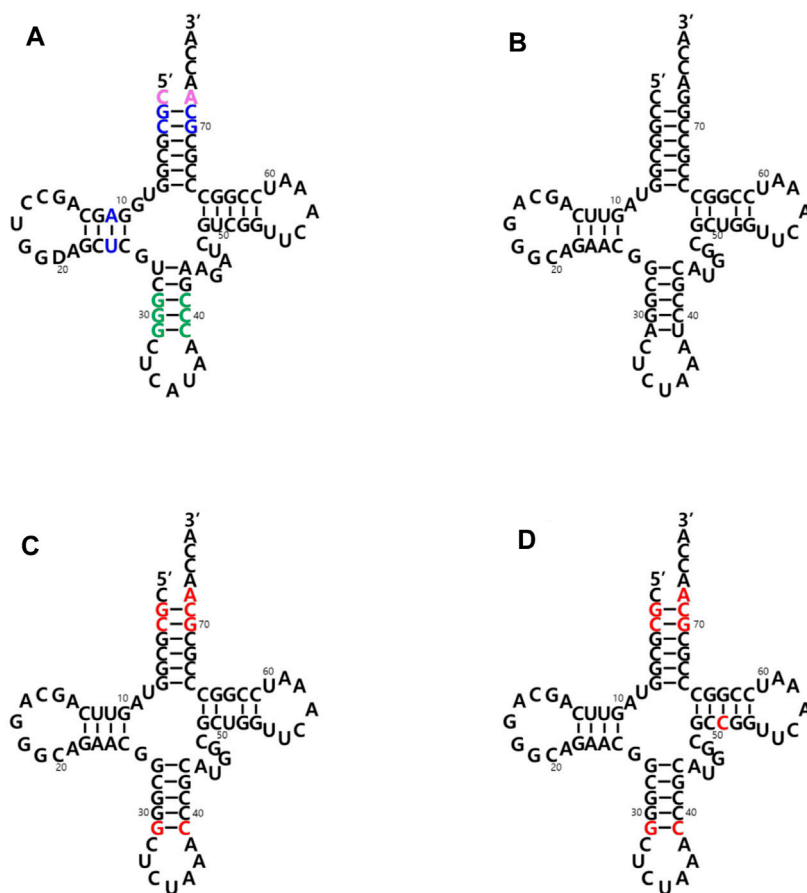


FIGURE 2 | Cloverleaf depiction of tRNAs used in this study. **(A)** *E. coli* initiator tRNA ($Ec\text{-}tRNA^{\text{Met}}$). The two main identity elements of $Ec\text{-}tRNA^{\text{Met}}$ are the absence of a Watson-Crick base pair between positions C1 and A72 in the acceptor stem (in pink) and three consecutive G:C base pairs in the anticodon stem (in green). Some additional elements playing a role in the formylation reaction are shown in blue. **(B)** Engineered *M. jannaschii* tyrosyl- $tRNA_{\text{CUA}}$ ($Mj\text{-}tRNA^{\text{Tyr}}$). **(C)** *Mj* initiator $tRNA_{\text{CUA}}\text{-1}$ ($Mj\text{-}itRNA\text{-1}$). **(D)** *Mj* initiator $tRNA_{\text{CUA}}\text{-2}$ ($Mj\text{-}itRNA\text{-2}$). The changed nucleotides in $Mj\text{-}itRNA\text{-1}$ and $Mj\text{-}itRNA\text{-2}$ compared with $Mj\text{-}tRNA^{\text{Tyr}}$ are marked by red letters.

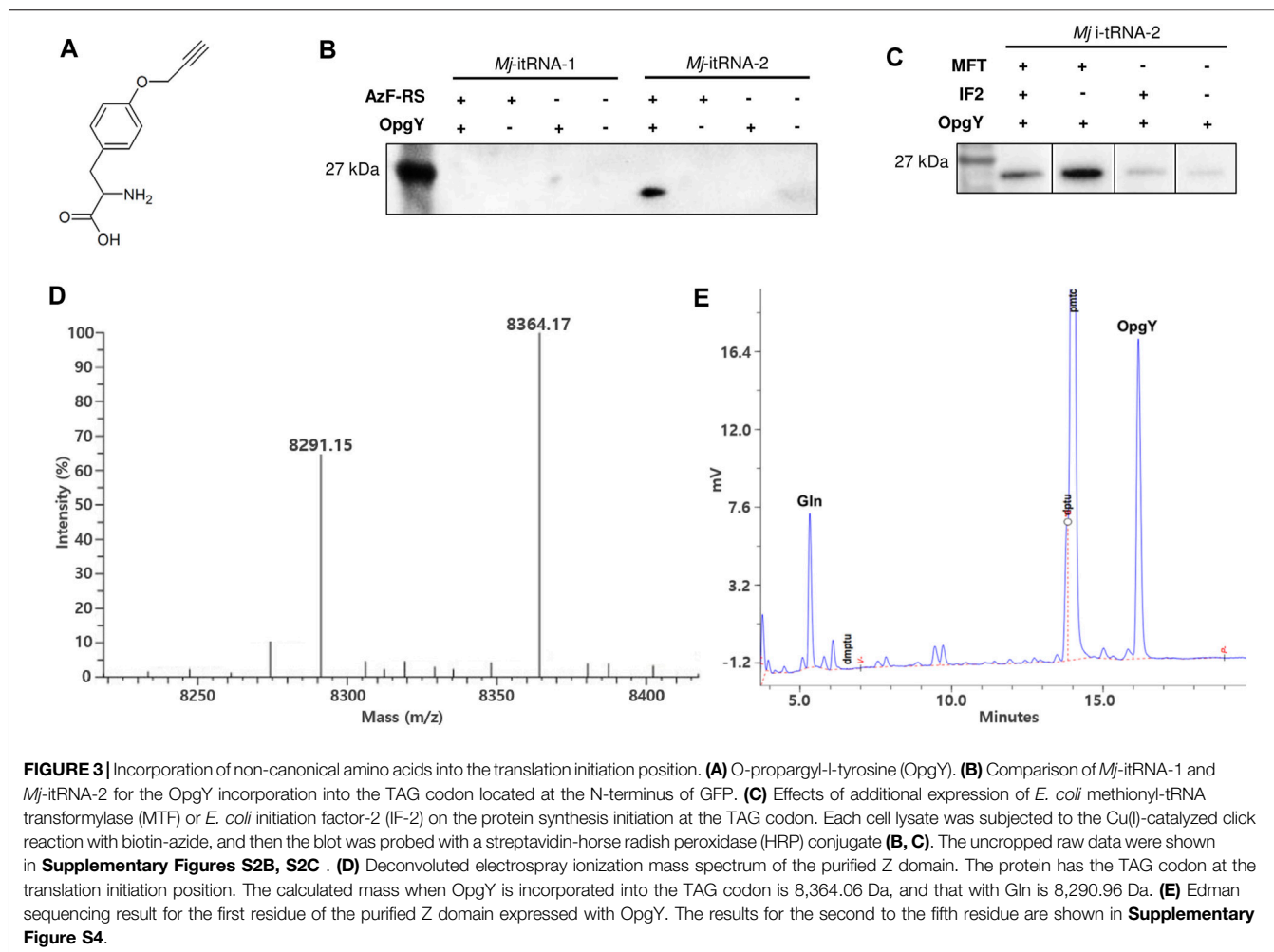
resulting molecule was named *Mj*-itRNA-1 (**Figure 2C**). The C51G52 motif was reported to interact with IF-2 (Louise, L. et al., 2009), although it is weak, and the element was additionally introduced into *Mj*-itRNA-1 by a mutation of U51→C51, resulting in *Mj*-itRNA-2 (**Figure 2D**).

The start codon of a GFP gene (Yoo et al., 2007) was changed to the TAG amber codon, and the gene was cloned into pBbE6a (Lee et al., 2011). The tRNA gene in the pEVOL-AzF plasmid (Young et al., 2010) was replaced with *Mj*-itRNA-1 or *Mj*-itRNA-2. pEVOL-AzF encodes an *Mj*-TyrRS variant (AzF-RS) that is active toward both AzF and O-propargyl-L-tyrosine (OpgY) (Lee et al., 2015). The resulting plasmids were transformed into *E. coli* DH10 β with pBbs2k-ProRS, which overexpresses *E. coli* prolyl-tRNA synthetase ($Ec\text{-}ProRS$) and suppresses the mischarging of *E. coli* prolyl-tRNA with nCAAs (Lee et al., 2015). Expression of AzF-RS and $Ec\text{-}ProRS$ was induced by L-arabinose and anhydrotetracycline, respectively, and the GFP protein with the TAG codon at its N-terminus was expressed by isopropyl β -D-thiogalactopyranoside in the presence of OpgY (**Figure 3A**). The cell lysate reacted with biotin-azide via the Cu(I)-catalyzed

click reaction, and the products were analyzed by western blotting using a HRP conjugate (**Figure 3B**). The signal indicates that the protein has an alkyne group of OpgY. Only when all *Mj*-itRNA-2, AzF-RS, and OpgY were present, the western signal was detected; which suggests that OpgY was incorporated into the TAG codon located at the initial position of GFP. That is, *Mj*-itRNA-2 was charged with OpgY by AzF-RS, and the α -amine group of OpgY in the aminoacylated tRNA was probably formylated by MTF, and then the resulting tRNA, formyl-OpgY-*Mj*-itRNA-2, supported the translation initiation with the amber codon.

Overexpression of MTF Increased the Efficiency of the Protein Synthesis Initiation with nCAA

Even though *Mj*-itRNA-2 enabled the initiation of protein synthesis at the amber codon, its interactions with the factors involved in forming the 30S initiation complex might be relatively weak compared with those of $Ec\text{-}tRNA^{\text{Met}}$. An approach to address this would be to increase the concentrations of these



factors instead of engineering the tRNA further. The three interactions of MTF, 30S ribosome, and IF-2 with *Mj*-itRNA-2 aminoacylated with OpgY could be considered to increase translation initiation with OpgY. We decided to explore the overexpression of MTF or IF-2; increasing the concentration of the 30S ribosome could affect various aspects of physiology and thus was not tested in this study. The genes for MTF or IF-2 were cloned under the constitutive Gln promoter of pEVOL-AzF; a bicistronic gene was used to express MTF and IF-2 simultaneously. Additional expression of MTF significantly improved the incorporation of OpgY into the amber codon at the initiation position of GFP, while the effect of overexpressing IF-2 was marginal (**Figure 3C**). This observation suggested that the formylation of the α -amine group of OpgY-*Mj*-itRNA-2 was the rate-determining step in translation initiation with the TAG codon. Expressing IF-2 in addition to MTF via the bicistronic construct decreased the signal compared with the case of expressing only MTF. This might be because the expression of IF-2 decreased the level of MTF due to the limited resources for transcription and translation in expressing multiple genes. The aminoacyl-tRNA synthetase of AzF-RS was originally engineered to activate AzF (Chin et al., 2002) and incorporate AzF into the

initiation TAG codon was attempted using the same *E. coli* strain. However, unlike OpgY, *Mj*-itRNA-2 did not support protein translation initiation with AzF (data not shown). In the case of *Mj*-itRNA-1, additional expression of MTF or IF-2 did not result in detectable incorporation of OpgY into the initiation TAG codon (data not shown).

The TAG Start Codon was Encoded with Either ncAA or Gln by *Mj*-itRNA-2

A small protein (Z domain (Nilsson et al., 1987)) was used to evaluate OpgY incorporation using a mass spectrometry. The start codon of Z domain was changed to TAG codon, and the resulting gene was cloned into pQE-80 L. The Z domain was expressed using *E. coli* cells expressing *Mj*-itRNA-2, AzF-RS, MTF, and ProRS in the presence of OpgY. The purified protein was analyzed by liquid chromatography-mass spectrometry. The calculated mass of the Z domain with OpgY at its N-terminus was 8,364.06 Da, and a mass of 8,364.17 was detected (**Figure 3D**). This result, along with the western blot results (**Figures 3B,C**), indicated that OpgY was incorporated into the TAG codon of the translation initiation position. However, another mass peak

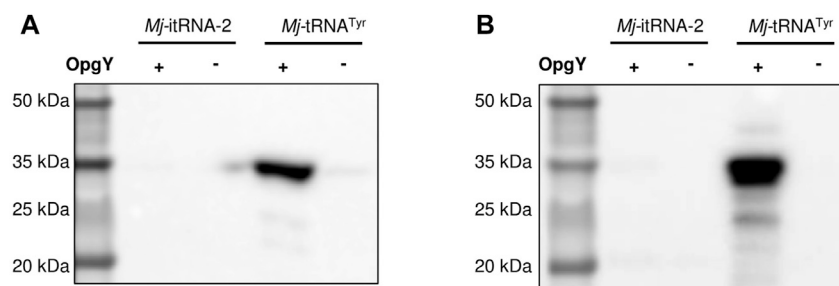


FIGURE 4 | Suppression of an elongational TAG codon with *Mj*-itRNA-2. There is a TAG codon between GST and the Z domain-His₆, and the fusion protein was expressed with *Mj*-itRNA-2 or *Mj*-tRNA^{Tyr}. Each cell lysate was subjected to the Cu(I)-catalyzed click reaction with biotin-azide, and then the blot was probed with an anti-His₆-HRP conjugate (**A**) or a streptavidin-HRP conjugate (**B**). The uncropped raw data were shown in **Supplementary Figures S5B, S5C**.

(8,291.15 Da) was observed in the Z-domain sample. Since the mass of the wild-type Z domain (Met at its N-terminus) is 8,294.03 Da (**Supplementary Figure S3**: the observed mass was 8,294.12), it was suspected that a canonical amino acid besides Met was incorporated into the TAG codon. The N-terminal residue of the Z domain was determined by the Edman degradation analysis, and two residues, OpgY and Gln, were detected (**Figure 3E**). The peaks for Gln and OpgY were integrated (597 for Gln and 1477 for OpgY), and the ratio of OpgY to Gln incorporated into the TAG position was calculated as 2.55. The mass of 8,291.15 Da was consistent with the calculated mass (8,290.96 Da) when Gln was incorporated into the TAG codon.

Mj-itRNA-2 Did Not Support the Incorporation of ncAA into an Internal Amber Codon

Next, we examined whether *Mj*-itRNA-2 could function as an elongator tRNA. The two main elements of *Ec*-tRNA^{Met}, the C1: A72 mismatch, and the three consecutive G:C pairs in the anticodon stem, are known to distinguish the initiator tRNA from elongator tRNAs and to prevent it from being involved in the translation elongation process (Laursen et al., 2005). In addition to the two elements, *Mj*-itRNA-2 had mutations in the acceptor and the D stem; the T stem has also been reported to interact with *E. coli* elongation factor Tu (Guo et al., 2009). A plasmid was constructed by introducing the TAG codon between GST and the Z domain, and the protein had a C-terminal His₆-tag. The GST-TAG-Z protein was expressed with either *Mj*-tRNA^{Tyr} (**Figure 2B**) or *Mj*-itRNA-2, and the cell lysates were analyzed by western blotting after the Cu(I)-catalyzed click reaction with biotin-azide. While OpgY was efficiently incorporated into the TAG codon with the elongator *Mj*-tRNA^{Tyr}, *Mj*-itRNA-2 could not suppress the internal nonsense codon (**Figure 4**). These results indicate that *Mj*-itRNA-2 1) is charged with ncAA, 2) initiates protein translation at the TAG codon, and 3) is not active in the elongation process of protein synthesis. *Mj*-itRNA-2 is the first reported tRNA enabling the orthogonal incorporation of ncAA into the amber codon only at the translation initiation position.

DISCUSSION

Installation of ncAAs into specific positions of the protein has provided invaluable tools for not only studying proteins but also engineering them. Orthogonal pairs of tRNA/aaRS have been engineered for these purposes, and novel proteins with various unnatural functional groups have been created. These systems have been developed to target the elongation process of protein synthesis. However, a method to incorporate ncAAs at the translation initiation position in a site-specific manner has rarely been reported. This may be attributed to the difficulty in engineering functional initiator tRNAs compared with elongation tRNAs. The protein synthesis initiation process of *E. coli* is distinct from the elongation process, and the initiator tRNA (*Ec*-tRNA^{Met}) plays an important role in this process. *Ec*-tRNA^{Met} interacts with MTF, 30S ribosome, and IF-2 (RajBhandary 1994; Laursen et al., 2005); the identity elements for these interactions also prevent association with the elongation process. In this study, we introduced the identity elements of *Ec*-tRNA^{Met} into an elongator *Mj*-tRNA^{Tyr} variant. One of the designed initiator tRNAs, *Mj*-itRNA-2, incorporated OpgY into the TAG codon at the translation initiation position, but was not active for an internal TAG codon.

Overexpression of MTF significantly improved the efficiency of translation initiation with OpgY, and the result implied that OpgY-*Mj*-itRNA-2 was not an efficient substrate for MTF. The same system failed to initiate protein translation with AzF, even though AzF-RS is also active toward ncAA. The identity of amino acids linked to an initiator tRNA affects the activity of MTF (Schmitt et al., 1998; Laursen et al., 2005). It is plausible that AzF-*Mj*-tRNA-2 was not formylated, considering the structural difference between OpgY and AzF. The elements present in the acceptor stem were introduced for interaction with MTF, but some other elements such as A11: U24 in the D stem (Lee et al., 1991; Ramesh et al., 1997) were not tested in this study because of the detrimental effects of mutations in this region in our previous experiments. In addition to OpgY, Gln was incorporated into the start TAG codon, which suggests that *Mj*-itRNA-2 was charged with Gln,

probably by the endogenous glutamyl-tRNA synthetase (*Ec*-GlnRS). These results suggest that further engineering of the initiator tRNA, particularly for enhancing the interaction with MTF and preventing the aminoacylation with Gln by *Ec*-GlnRS, could improve translation initiation efficiency with ncAAs. The newly engineered initiator tRNAs would enable developing an orthogonal translation initiation system active only for the TAG start codon. Furthermore, with many *Mj*-TyrRS variants engineered toward various ncAAs, a repertoire of functional groups would be available for site-specific incorporation at the N-termini of proteins.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

REFERENCES

- Agostini, F., Völler, J. S., Kokschi, B., Acevedo-Rocha, C. G., Kubyskhin, V., and Budisa, N. (2017). Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* 56 (33), 9680–9703. doi:10.1002/anie.201610129
- Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004). An Expanded Genetic Code with a Functional Quadruplet Codon. *Proc. Natl. Acad. Sci.* 101 (20), 7566–7571. doi:10.1073/pnas.0401517101
- Chin, J. W. (2017). Expanding and Reprogramming the Genetic Code. *Nature* 550 (7674), 53–60. doi:10.1038/nature24031
- Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002). Addition of P-Azido-L-Phenylalanine to the Genetic Code of *Escherichia coli*. *J. Am. Chem. Soc.* 124 (31), 9026–9027. doi:10.1021/ja027007w
- Christopher, J. A., and Baldwin, T. O. (1996). Implications of N and C-Terminal Proximity for Protein Folding. *J. Mol. Biol.* 257 (1), 175–187. doi:10.1006/jmbi.1996.0154
- Evans, T. C., Jr., and Xu, M.-Q. (1999). Intein-mediated Protein Ligation: Harnessing Nature's Escape Artists. *Biopolymers* 51 (5), 333–342. doi:10.1002/(sici)1097-0282(1999)51:5<333:aid-bip3>3.0.co;2-#
- Guo, J., Melançon, C. E., 3rd, Lee, H. S., Groff, D., and Schultz, P. G. (2009). Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem. Int. Edition* 48 (48), 9148–9151. doi:10.1002/anie.200904035
- Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001). Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an *Escherichia coli* *In Vitro* Translation System. *Biochemistry* 40 (37), 11060–11064. doi:10.1021/bi0108204
- Hong, V., Presolski, S. I., Ma, C., and Finn, M. A. G. (2009). Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.* 48 (52), 9879–9883. doi:10.1002/anie.200905087
- Jacob, E., and Unger, R. (2007). A Tale of Two Tails: Why Are Terminal Residues of Proteins Exposed. *Bioinformatics* 23 (2), e225–e230. doi:10.1093/bioinformatics/btl318
- Johnson, J. A., Lu, Y. Y., Van Deventer, J. A., and Tirrell, D. A. (2010). Residue-specific Incorporation of Non-canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* 14 (6), 774–780. doi:10.1016/j.cbpa.2010.09.013
- Kinney, K., Flynn, M., Thomas, S. B., Meekin, J., Varkey, R., Xiao, X., et al. (2019). Preclinical Assessment of an Antibody-PBD Conjugate that Targets BCMA on Multiple Myeloma and Myeloma Progenitor Cells. *Leukemia* 33 (3), 766–771. doi:10.1038/s41375-018-0278-7
- Kozak, M. (1999). Initiation of Translation in Prokaryotes and Eukaryotes. *Gene* 234 (2), 187–208. doi:10.1016/s0378-1119(99)00210-3
- Laursen, B. S., Sørensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. (2005). Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.* 69 (1), 101–123. doi:10.1128/mmr.69.1.101-123.2005
- Lee, B. S., Kim, S., Ko, B. J., and Yoo, T. H. (2017). An Efficient System for Incorporation of Unnatural Amino Acids in Response to the Four-Base Codon AGGA in *Escherichia coli*. *Biochim. Biophys. Acta Gen. Subj.* 1861 (11 Pt B), 3016–3023. doi:10.1016/j.bbagen.2017.02.017
- Lee, B. S., Shin, S., Jeon, J. Y., Jang, K.-S., Lee, B. Y., Choi, S., et al. (2015). Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS Chem. Biol.* 10 (7), 1648–1653. doi:10.1021/acscchembio.5b00230
- Lee, C. P., Seong, B. L., and RajBhandary, U. L. (1991). Structural and Sequence Elements Important for Recognition of *Escherichia coli* Formylmethionine tRNA by Methionyl-tRNA Transformylase Are Clustered in the Acceptor Stem. *J. Biol. Chem.* 266 (27), 18012–18017. doi:10.1016/s0021-9258(18)55230-3
- Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., et al. (2011). BglBrick Vectors and Datasheets: A Synthetic Biology Platform for Gene Expression. *J. Biol. Eng.* 5, 12. doi:10.1186/1754-1611-5-12
- Liebscher, S., Schöpfel, M., Aumüller, T., Sharkhuukhen, A., Pech, A., Höss, E., et al. (2014). N-terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chem. Int. Ed.* 53 (11), 3024–3028. doi:10.1002/anie.201307736
- Link, A. J., Mock, M. L., and Tirrell, D. A. (2003). Non-canonical Amino Acids in Protein Engineering. *Curr. Opin. Biotechnol.* 14 (6), 603–609. doi:10.1016/j.copbio.2003.10.011
- Liu, C. C., and Schultz, P. G. (2010). Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* 79, 413–444. doi:10.1146/annurev.biochem.052308.105824
- Louise, C. V. R., Brain, S. L., and Hansu, U. S. (2009). *Initiator tRNAs in Bacteria and Eukaryotes. Encyclopedia of Life Science.* (New York: John Wiley and Sons).
- Mao, H., Hart, S. A., and Pollack, B. A. (2004). Sortase-Mediated Protein Ligation - A New Method for Protein Engineering. *J. Am. Chem. Soc.* 126, 2670–2671. doi:10.1021/ja039915e
- Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015). Reassignment of a Rare Sense Codon to a Non-canonical Amino Acid in *Escherichia coli*. *Nucleic Acids Res.* 43 (16), 8111–8122. doi:10.1093/nar/gkv787
- Ngo, J. T., Schuman, E. M., and Tirrell, D. A. (2013). Mutant Methionyl-tRNA Synthetase from Bacteria Enables Site-Selective N-Terminal Labeling of Proteins Expressed in Mammalian Cells. *Proc. Natl. Acad. Sci.* 110 (13), 4992–4997. doi:10.1073/pnas.1216375110

AUTHOR CONTRIBUTIONS

BL, WC, and TY designed research; BL, WC, and SL performed the experiments and analyzed the data; BK performed mass analyses; BL, WC, and TY wrote the article. All authors approved the final version of the article.

FUNDING

This work was supported by the National Research Foundation of Korea (2019R1A6A1A11051471 and 2021R1A2C2003453) and the Commercializations Promotion Agency for R&D Outcomes (2021N100).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.772648/full#supplementary-material>

- Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmlblad, A., Holmgren, E., et al. (1987). A Synthetic IgG-Binding Domain Based on Staphylococcal Protein A. *Protein Eng. Des. Sel* 1 (2), 107–113. doi:10.1093/protein/1.2.107
- Pagar, A. D., Patil, M. D., Flood, D. T., Yoo, T. H., Dawson, P. E., and Yun, H. (2021). Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* 121 (10), 6173–6245. doi:10.1021/acs.chemrev.0c01201
- Proft, T. (2010). Sortase-mediated Protein Ligation: an Emerging Biotechnology Tool for Protein Modification and Immobilisation. *Biotechnol. Lett.* 32 (1), 1–10. doi:10.1007/s10529-009-0116-0
- RajBhandary, U. L. (1994). Initiator Transfer RNAs. *J. Bacteriol.* 176 (3), 547–552. doi:10.1128/jb.176.3.547-552.1994
- Ramesh, V., Gite, S., Li, Y., and RajBhandary, U. L. (1997). Suppressor Mutations in *Escherichia coli* Methionyl-tRNA Formyltransferase: Role of a 16-amino Acid Insertion Module in Initiator tRNA Recognition. *Proc. Natl. Acad. Sci.* 94 (25), 13524–13529. doi:10.1073/pnas.94.25.13524
- Rosen, C. B., and Francis, M. B. (2017). Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* 13 (7), 697–705. doi:10.1038/nchembio.2416
- Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998). Crystal Structure of Methionyl-tRNA^{Met} Transformylase Complexed with the Initiator Formyl-Methionyl-tRNA^{Met}. *EMBO J.* 17 (23), 6819–6826. doi:10.1093/emboj/17.23.6819
- Simone, A. D., Acevedo-Rocha, C. G., Hoessl, M. G., and Budisa, N. (2016). Towards Reassignment of the Methionine Codon AUG to Two Different Noncanonical Amino Acids in Bacterial Translation. *Croat. Chem. Acta* 89 (2), 243–253. doi:10.5562/cca2915
- Smolskaya, S., and Andreev, Y. A. (2019). Site-Specific Incorporation of Unnatural Amino Acids into *Escherichia coli* Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules* 9 (7). doi:10.3390/biom9070255
- Steiman, D. M., Ridge, R. J., and Matsueda, G. R. (1985). Synthesis of Side Chain-Protected Amino Acid Phenylthiohydantoins and Their Use in Quantitative Solid-phase Edman Degradation. *Anal. Biochem.* 145 (1), 91–95. doi:10.1016/0003-2697(85)90331-8
- Serada, T. J., and Mant, C. (1993). Effect of the α -amino Group on Peptide Retention Behaviour in Reversed-phase Chromatography Determination of the pKa Values of the α -amino Group of 19 Different N-Terminal Amino Acid Residues. *J. Chromatogr.* 646, 17–30.
- Tharp, J. M., Ad, O., Amikura, K., Ward, F. R., Garcia, E. M., Cate, J. H. D., et al. (2020). Initiation of Protein Synthesis with Non-Canonical Amino Acids *In Vivo*. *Angew. Chem. Int. Ed.* 59 (8), 3122–3126. doi:10.1002/anie.201914671
- Tharp, J. M., Krahn, N., Varshney, U., and Söll, D. (2020). Hijacking Translation Initiation for Synthetic Biology. *Chembiochem* 21 (10), 1387–1396. doi:10.1002/cbic.202000017
- Varshney, U., Lee, C. P., and RajBhandary, U. L. (1993). From Elongator tRNA to Initiator tRNA. *Proc. Natl. Acad. Sci.* 90 (6), 2305–2309. doi:10.1073/pnas.90.6.2305
- Varshney, U., and RajBhandary, U. L. (1990). Initiation of Protein Synthesis from a Termination Codon. *Proc. Natl. Acad. Sci.* 87 (4), 1586–1590. doi:10.1073/pnas.87.4.1586
- Wals, K., and Ova, H. (2014). Unnatural Amino Acid Incorporation in *E. coli*: Current and Future Applications in the Design of Therapeutic Proteins. *Front. Chem.* 2, 15. doi:10.3389/fchem.2014.00015
- Wang, K., Schmied, W. H., and Chin, J. W. (2012). Reprogramming the Genetic Code: from Triplet to Quadruplet Codes. *Angew. Chem. Int. Ed.* 51 (10), 2288–2297. doi:10.1002/anie.201105016
- Wang, L., and Schultz, P. G. (2004). Expanding the Genetic Code. *Angew. Chem. Int. Ed. Engl.* 44 (1), 34–66. doi:10.1002/anie.200460627
- Wang, L., and Schultz, P. G. (2001). A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* 8 (9), 883–890. doi:10.1016/s1074-5521(01)00063-1
- Wang, Q., Parrish, A. R., and Wang, L. (2009). Expanding the Genetic Code for Biological Studies. *Chem. Biol.* 16 (3), 323–336. doi:10.1016/j.chembiol.2009.03.001
- Wang, Q., and Wang, L. (2012). Genetic Incorporation of Unnatural Amino Acids into Proteins in Yeast. *Methods Mol. Biol.* 794, 199–213. doi:10.1007/978-1-61779-331-8_12
- Weeks, A. M., and Wells, J. A. (2020). Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* 120 (6), 3127–3160. doi:10.1021/acs.chemrev.9b00372
- Xiao, H., Chatterjee, A., Choi, S.-h., Bajjuri, K. M., Sinha, S. C., and Schultz, P. G. (2013). Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chem. Int. Ed.* 52 (52), 14080–14083. doi:10.1002/anie.201308137
- Xu, M.-Q., and Evans, T. C., Jr. (2001). Intein-mediated Ligation and Cyclization of Expressed Proteins. *Methods* 24 (3), 257–277. doi:10.1006/meth.2001.1187
- Yoo, T. H., Link, A. J., and Tirrell, D. A. (2007). Evolution of a Fluorinated green Fluorescent Protein. *Proc. Natl. Acad. Sci.* 104 (35), 13887–13890. doi:10.1073/pnas.0701904104
- Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010). An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli*. *J. Mol. Biol.* 395 (2), 361–374. doi:10.1016/j.jmb.2009.10.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lee, Choi, Lee, Ko and Yoo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.