ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Translation System Engineering in *Escherichia coli* Enhances Non-Canonical Amino Acid Incorporation into Proteins

Rui Gan,[1] Jessica G. Perez,[1] Erik D. Carlson,[1] Ioanna Ntai,[2,3] Farren J. Isaacs,[4,5] Neil L. Kelleher,[2,3,6] Michael C. Jewett[1,3,7,8,9]

[1]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3120; telephone: 8474675007; fax: 8474913728; e-mail: m-jewett@northwestern.edu

[2]Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3120

[3]Chemistry of Life Processes Institute, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

[4]Systems Biology Institute, Yale University, West Haven, Connecticut

[5]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut

[6]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois

[7]Interdisciplinary Biological Sciences Program, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-0001

[8]Northwestern Institute on Complex Systems, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

[9]Simpson Querry Institute, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

ABSTRACT: The ability to site-specifically incorporate non-canonical amino acids (ncAAs) into proteins has made possible the study of protein structure and function in fundamentally new ways, as well as the bio synthesis of unnatural polymers. However, the task of site-specifically incorporating multiple ncAAs into proteins with high purity and yield continues to present a challenge. At the heart of this challenge lies the lower efficiency of engineered orthogonal translation system components compared to their natural counterparts (e.g., translation elements that specifically use a ncAA and do not interact with the cell's natural translation apparatus). Here, we show that evolving and tuning expression levels of multiple components of an engineered translation system together as a whole enhances ncAA incorporation efficiency. Specifically, we increase protein yield when incorporating multiple *p*-azido-phenylalanine(pAzF) residues into proteins by (i) evolving the *Methanocaldococcus jannaschii p*-azido-phenylalanyl-tRNA synthetase anti-codon binding domain, (ii) evolving the elongation factor Tu amino acid-binding pocket, and (iii) tuning the expression of evolved translation machinery components in a single vector. Use of the evolved translation machinery in a genomically recoded organism lacking release factor one enabled enhanced multi-site ncAA incorporation into proteins. We anticipate that our approach to orthogonal translation system development will accelerate and expand our ability to site-specifically incorporate multiple ncAAs into proteins and biopolymers, advancing new horizons for synthetic and chemical biotechnology.

Biotechnol. Bioeng. 2017;114: 1074–1086.

KEYWORDS: noncanonical amino acid; orthogonal translation system; aminoacyl-tRNA synthetase; elongation factor Tu; synthetic biology; genomically recoded organism; directed evolution

Conflict of interest: None.

Abbreviations: IPTG, β-D-1-thiogalactopyranoside; ncAA, noncanonical amino acid; CFPS, cell-free protein synthesis; OTS, orthogonal translation system; pAzF, p-azido-phenylalanine; pAzFRS, p-Azido-phenylalanyl-tRNA synthetase; GFP, green fluorescence protein; EF-Tu, elongation factor Tu; o-tRNA, orthogonal tRNA; aaRS, aminoacyl-tRNA synthetase; RF1, release factor 1.

Correspondence to: M.C. Jewett

Contract grant sponsor: Defense Advanced Research Projects Agency
Contract grant number: N66001-12-C-4211
Contract grant sponsor: David and Lucile Packard Foundation
Contract grant sponsor: The Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust
Contract grant sponsor: Arnold and Mabel Beckman Foundation
Contract grant sponsor: National Resource for translational and Developmental Proteomics
Contract grant number: P41 GM108569
Received 28 July 2016; Revision received 28 October 2016; Accepted 13 December 2016
Accepted manuscript online 13 December 2016;
Article first published online 2 February 2017 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.26239/abstract).
DOI 10.1002/bit.26239

Introduction

The site-specific, co-translational incorporation of non-canonical amino acids (ncAAs) into proteins expands the range of genetically encoded chemistry in proteins. This expansion can be a powerful tool for both fundamental science and applied technology (Davis and Chin, 2012; Forster et al., 2003; Liu and Schultz, 2010). For example, photocaged (Wu et al., 2004), fluorescent (Summerer et al., 2006), and bio-orthogonal reactive (Chin et al., 2002a,b; Lang et al., 2012) ncAAs have provided new ways to study protein structure and dynamics (Liu and Schultz, 2010). In addition, ncAAs that mimic natural post-translational modifications have helped to elucidate the role of such modifications in previously unattainable ways (Bröcker et al., 2014; Davis and Chin, 2012; Lee et al., 2013; Neumann et al., 2009; Oza et al., 2015; Park et al., 2011; Tian et al., 2014). Further, ncAA incorporation into proteins has opened the way to novel antibody drug conjugates (Axup et al., 2012; Zimmerman et al., 2014), modified human therapeutics (Cho et al., 2011), and protein biomaterials (Albayrak and Swartz, 2014), among other applications. With the ability to construct bio-based products beyond the limits of nature, expanding the genetic code has emerged as a major opportunity in synthetic and chemical biology (Des Soye et al., 2015; Hoesl and Budisa, 2012; Liu and Schultz, 2010; Quast et al., 2015).

Expansion of the genetic code by orthogonal translation systems (OTSs) involves codon re-assignment and engineered translation machinery. The engineered translation machinery is evolved to recognize a ncAA and ideally can operate alongside the cell's natural translation apparatus in a parallel and independent fashion. Most commonly, the amber UAG stop codon is re-assigned to a ncAA in an approach called amber suppression (Wang et al., 2006). In this approach pioneered by Wang and Schultz (2002), an orthogonal tRNA (o-tRNA) anticodon is mutated to allow for decoding of the UAG stop codon (Wang et al., 2000), and an orthogonal aminoacyl-tRNA synthetase (o-aaRS) is evolved to aminoacylate the ncAA to the o-tRNA. These o-tRNA/o-aaRS pairs are typically derived from phylogenetically distant organisms such as *Methanocaldococcus jannaschii* or *Methanosarcina mazei*, and are not efficiently recognized by the aaRS/tRNA pairs of the host, that is *Escherichia coli* (Ambrogelly et al., 2007; Plass et al., 2011). The o-tRNA/o-aaRS pairs are then evolved so that the o-aaRS charges a ncAA to the o-tRNA.

Once charged, the ncAA-tRNA complex must be properly delivered to the ribosome by elongation factor Tu (EF-Tu). As part of key proofreading steps, the thermodynamic interactions between elongation factor Tu (EF-Tu) and aminoacylated tRNAs are finely balanced to bind the canonical amino acids with their cognate tRNAs (Schrader et al., 2009), and not mismatched amino acyl-tRNA substrates. Consequently, ncAA-tRNA substrates may not bind EF-Tu efficiently because it is a non-native substrate, preventing efficient delivery to the ribosome. While still a subject of debate for ncAA incorporation efforts (Rogerson et al., 2015), EF-Tu may thus require engineering to allow for the efficient incorporation of a given ncAA (Fan et al., 2016; Park et al., 2011). Beyond o-aaRSs, o-tRNAs, and EF-Tu, ribosomes must be able to accommodate the ncAA and in some instances ribosome engineering may be necessary. Towards such a goal, orthogonal ribosome systems could be used, noting the recent advance of a tethered ribosome system in cells, termed Ribo-T, that enables modification of the large and small subunits of the ribosome while avoiding viability constraints (Fried et al., 2015; Orelle et al., 2015; Rackham and Chin, 2005; Wang et al., 2007).

To date, over 150 ncAAs have been co-translationally incorporated into proteins in *E. coli* (Des Soye et al., 2015; Lang and Chin, 2014). While these advances highlight our ever-expanding understanding of the workings of the translational apparatus, they have also revealed areas for improvement. Rigorous analyses of several o-aaRSs reveal that these enzymes, while functional, generally have poor catalytic efficiency as compared to native synthetases (Boniecki et al., 2008; Ling and Soll, 2010; Nehring et al., 2012; Reynolds et al., 2010; Tanrikulu et al., 2009; Umehara et al., 2012; Wiltschi et al., 2008). This has led to limitations in site-specific, multi-site ncAA incorporation into proteins. Moreover, o-aaRSs show polyspecificity, meaning a single OTS can incorporate multiple ncAAs. This is useful for expanding ncAA diversity without the need for further evolution, but hinders incorporation of multiple distinct ncAAs by existing o-aaRSs (Miyake-Stoner et al., 2010; Wan et al., 2014; Wang et al., 2012; Young et al., 2011). Native EF-Tu also shows limited capacity for incorporation of bulky or charged ncAAs (Fan et al., 2016; Park et al., 2011), and could be the target of engineering efforts if thermodynamic interactions limit delivery of aminoacyl-tRNA substrates (Wang et al., 2016). Further, the presence of release factor 1 (RF1) can cause early termination of proteins when using amber suppression technology because it competes for the UAG codon (Hong et al., 2014b; Johnson et al., 2011). Additionally, overexpression of OTS system components can be detrimental to cellular growth (Hong et al., 2015), ultimately affecting protein production and cell viability.

Recent advances have led to major improvements in ncAA incorporation and are beginning to address many of the aforementioned limitations. First, optimized OTS expression systems have been shown to enhance suppression efficiency and allow for two different ncAAs to be incorporated into the same protein (Chatterjee et al., 2013; Young et al., 2010). Second, efforts to suppress or delete RF1 have removed the competition with the ncAA-o-tRNA species at the UAG codon to increase incorporation efficiencies (Johnson et al., 2011; Lajoie et al., 2013; Loscha et al., 2012; Mukai et al., 2010; Ohtake et al., 2012). Third, efforts to further engineer individual OTS components have led to improved ncAA aminoacylation efficiencies and increased overall yields of modified proteins, for example, o-tRNA (Chatterjee et al., 2012; Young et al., 2010), o-aaRS (Amiram et al., 2015; Chatterjee et al., 2012; Liu et al., 1997), EF-Tu (Doi et al., 2007; Fan et al., 2016; Park et al., 2011; Wang et al., 2015), and ribosome (Barrett and Chin, 2010; Neumann et al., 2010; Rackham and Chin, 2005; Wang et al., 2007).

In this work, we sought to explore the effects of evolving both the o-aaRS and EF-Tu for improved *p*-azido-phenylalanine (pAzF) incorporation into proteins in *E. coli*. The foundational principle was that a systematic engineering approach of evolving more than one component of the orthogonal translation machinery could provide beneficial opportunities to enhance ncAA incorporation into proteins. To achieve this goal, our study involved three steps. First, we evolved the aminoacyl-tRNA synthetase of *M. jannaschii* to

improve pAzF incorporation into proteins. Previous works have already identified *M. jannaschii* o-tRNA/o-aaRS pairs that can be used to site-specifically install pAzF into proteins (Amiram et al., 2015; Chin et al., 2002b). However, based on our recent work and that of others (Amiram et al., 2015; Wang et al., 2015), we hypothesized that the anticodon recognition domain of the o-aaRS could be further engineered. We explored a total of 10 amino acid residues residing in the anticodon recognition domain of the *M. jannaschii* o-aaRS and demonstrated the ability to isolate variants with increased pAzF incorporation efficiency. Second, we evolved the *E. coli* EF-Tu to improve pAzF incorporation into proteins. Given the important role of EF-Tu in the system of translation, we hypothesized that EF-Tu mutants might yield an enhanced OTS. Indeed, our results showed that we could further enhance ncAA incorporation into proteins by combining the beneficial mutants of both the o-aaRS and EF-Tu. Third, we assessed the ability of our engineered OTS to synthesize proteins containing multiple site-specifically introduced pAzF residues. Though our engineered system does not benefit the already high suppression efficiency of a single UAG amber codon, multi-site incorporation of several pAzF residues is significantly improved, an observation that is amplified in a genomically recoded strain lacking RF1 (Lajoie et al., 2013) and consistent with another recent report (Zheng et al., 2016). In sum, our work demonstrates that a comprehensive engineering approach is advantageous for genetic expansion efforts that require highly efficient OTSs. This work has implications for biotechnology, protein engineering, and synthetic biology projects.

Materials and Methods

Reagents and Buffers

Chemicals and media were purchased from Sigma–Aldrich unless designated otherwise. DNA polymerases, T4 polynucleotide kinase, T4 DNA ligase, and restriction endonucleases were purchased from New England Biolabs (NEB). Plasmids were extracted using Omega E.Z.N.A. DNA/RNA Isolation Kits (Omega Bio-Tek). All DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. (IDT). pAzF was purchased from P212121, LLC.

Strains and Plasmids

DH10-beta competent cells were purchased from NEB. All oligonucleotides used for cloning are shown in Supplementary Table SI, and all vectors are shown in Supplementary Figure S1. The genes of the orthogonal translation system (EF-Tu, pAzFRS, and o-tRNA variants) were cloned into expression vectors using standard methods. The initial source of these genes are as follows: *E. coli* native EF-Tu-coding sequence was amplified from the genome of *E. coli* KC6 strain (Jewett et al., 2008); pAzF aminoacyl tRNA synthetase (pAzFRS) was amplified from the plasmid pEVOL-pAzF (Young et al., 2010); the amber codon suppressor tRNA, *M. jannaschii* $tRNA^{tyr}_{CUA}$, was amplified from the plasmid pEVOL-pAzF (Young et al., 2010). Both EF-Tu and pAzFRS were inserted into plasmid vector pDLppTT1 (kanamycin) flanked by the lpp promoter and the T1T2 terminator. The resulting plasmids were named pDLppEAzRSTT1 and pDLppEFTuTT1, respectively (Supplementary Fig. S1). Plasmids bearing super folder green fluorescent protein (sfGFP) based reporter genes were constructed by insertion of the sfGFP gene from the plasmid pY71sfGFP (Bundy and Swartz, 2010) with 1 amber codon at position D190 into the plasmid vector pDT7TT2 (carbenicillin resistance). The resulting plasmid, which harbors a T7 promoter and T1T2 terminator flanking the sfGFP gene, was termed pDT7sfGFP1TAGTT2. From this construct, we used standard cloning procedures to generate the 3-amber (D36, K101, and D190) and 5-amber (D36, K101, E132, D190, and E213) sfGFP reporter plasmids (Hong et al., 2014b), which were named pDT7sfGFP3TAGTT2 and pDT7sfGFP5TAGTT2, respectively. We also generated a series of constructs with the T7 RNA polymerase promoter replaced by the native promoter PtacI (de Boer et al., 1983) for use in *E. coli* C321.ΔA (Lajoie et al., 2013), which lacks a genomic copy of T7 RNA polymerase. These plasmids were named pDPtacIsfGFPXTAGTT2 (X: 1, 3, or 5) (Supplementary Fig. S1). To create the library of orthogonal translation constructs with different promoter strengths (termed JGPX (X: 1–27), the three OTS component genes were inserted into the JGPX plasmids, with either the T7, PtacI, or lpp promoters in all combinations. The pAzFRS, o-tRNA, and EF-Tu were inserted using NheI and EcoRI, BamHI and EcoRI, BamHI and NheI restriction sites at the 5′ and 3′ ends, respectively. All plasmid constructs were sequence verified. For the construction of JGP26BM, we replaced the parental pAzFRS and wildtype EF-Tu with the evolved mutants, respectively in the plasmid JGP26 (Supplementary Fig. S1). Of note, the sfGFP reporter system described here was used because it has been previously optimized for the top-down mass spectrometry experiments in our previous work (Hong et al., 2014b, 2015).

p-Azido-phenylalanine-tRNA Synthetase Mutagenesis and Selection

Based on the 3-dimensional structure of the *M. jannaschii*-based pAzFRS/tRNA pair (PDB ID: 1J1U), a total of 10 amino acids in the vicinity of the o-tRNA anti-codon loop were mutated to assess the ability to enhance pAzF incorporation efficiency. This was achieved through the use of several libraries as described in the text following a positive-negative selection procedure reported previously in the literature (Wang and Schultz, 2001; Wang et al., 2003). For selection, we used LB medium instead of 2xYT, because we found that LB caused less false positive colonies during the positive-negative selection strategy as compared to 2xYT. The selection scheme is shown in Supplementary Figure S2. In the positive selection, functional aaRS constructs are tested in vivo for their ability to suppress an in frame amber codon in a chloramphenicol acetyl transferase (*cat)* reporter that confers resistance to chloramphenicol. In the negative selection, synthetases that loaded a natural amino acid were selected against using the toxin protein barnase.

In library 1 (RsLib1), A233, P258, F261, H283, M285, and D286 were mutated as NNK (N: A, T, G, or C; K: G or T) combinatorially. After construction, the plasmid library was transformed by electroporation into *E. coli* DH10-beta containing the pRepCMD112 plasmid (Supplementary Fig. S1) derived from pRepCM3 (Melancon and Schultz, 2009) to produce $\sim 10^8$ transformants.

Winning transformants were identified by their ability to grow on LB solid media supplemented with 25 μg/mL kanamycin and 10 μg/mL tetracycline in the presence of 0.5 mM IPTG, 1 mM pAzF, and 10 μg/mL chloramphenicol (first-round selection), or 20 μg/mL chloramphenicol (second-round selection), at 37°C for 36 h. Plasmids from surviving colonies were extracted and transformed into DH10-beta *E. coli* with pLWJ17B3 (which encodes the barnase gene) for negative selection (Wang et al., 2003). The negative selection was carried out as above but in the absence of pAzF. Following negative selection, plasmids from surviving colonies were transformed into DH10-beta *E. coli* harboring pDPtacIsfGFP1TAGTT2 plasmid for the final round of screening based on sfGFP expression in a 96-well plate assay. This well plate growth and expression assay was carried out in 100 μL of 2xYT media with 25 μg/mL kanamycin and 10 μg/mL tetracycline in the presence of 1 mM IPTG, 1 mM pAzF, or 0 mM pAzF for negative control samples (Supplementary Fig. S1).

In library 2 (RsLib2), another four amino acids (C231, P232, P284, and K288), which are close to the U35 and A36 of the o-tRNA anticodon loop, were mutated to NNK individually (only one mutation site per DNA molecule). Rather than the antibiotic selection described above, we focused here on the analysis of sfGFP expression by intact cell fluorescence measurements, where a sfGFP containing an amber codon will be expressed at a higher level, if an orthogonal aaRS/tRNA cognate pair has enhanced activity due to a beneficial mutation. After library construction, the plasmid library was transformed into DH10-beta cell with pDULE-tRNA and pDPtacIsfGFP1TAGTT2 (the sfGFP reporter plasmid), and grown on LB solid media (25 μg/mL kanamycin, 10 μg/mL tetracycline; and 50 μg/mL carbenicillin) at 37°C for 24 h. Six hundred single colonies were then picked and grown in 96-well plates in 2xYT liquid media with 25 μg/mL kanamycin, 10 μg/mL tetracycline, and 50 μg/mL Carbenicillin. At an optical density (OD_{600}) = 0.2, 1 mM IPTG, and 1 mM pAzF were added to induce protein expression allowing for pAzF incorporation into sfGFP containing an amber codon. The fluorescence (excitation/emission = 488/525 nm) of each well was read in Synergy2 (BioTek) with shaking and normalized by optical density with the wavelength of 600 nanometers (Fluorescence per OD_{600}). Each aaRS variant was treated as a separate experiment, and sfGFP levels were compared with the progenitor from which it was evolved.

In library 3 (RsLib3), H283, P284, M285, and R286 were mutated to NNK (N: A, T, G, or C; K: G, or T) combinatorially. Selection was performed as described for RsLib1.

EF-Tu Mutagenesis and Screening

Based on the known 3-dimensional structure of EF-Tu (PDB ID: 1OB2), we created a mutant library in the EF-Tu substrate recognition pocket to assess its ability to enhance pAzF incorporation. This library included a total of 10 amino acids: S65, H66, E215, D216, V217, F218, T228, F261, N273, and V274. As this library was too large to fully mutate to all other amino acids, we pursued an alanine scanning approach whereby we mutated each amino acids residue to alanine in all combinations using the method depicted in Supplementary Figure S3. The wildtype EF-Tu-coding sequence was amplified as five fragments. Each fragment carried the desired mutations introduced by DNA oligos. Then the five fragments, each with overlapping regions, were assembled to full-length EF-Tu sequence by a thermal cycling procedure: 95°C, 1 min; 20 cycles of 98°C, 20 s; 50°C, 20 s; 72°C, 2 min; and a final extension of 72°C, 5 min. The full-length product was cloned into pDLppTT1 plasmids using flanking NdeI and XhoI restriction sites. Twenty samples were sequenced to generate snapshot of the library's diversity (Supplementary Fig. S4). Next, the EF-Tu library-coding fragment was digested by the restriction endonucleases NheI and BamHI and inserted into pDLppEAzRSTT1 via the same cut sites, resulting in pDLppEA-zEFTT1 (Supplementary Fig. S1). Lastly, the plasmid library was transformed into DH10-beta strain harboring the plasmid pDPtacIsfGFP1TAGTT2 and pDULE-tRNA (Supplementary Fig. S1). Single colonies were picked from 2xYT solid media plate and inoculated into fresh 2xYT liquid media (25 μg/mL kanamycin, 10 μg/mL tetracycline, and 50 μg/mL carbenicillin) in a 96-well plate (100 μL media/well). The cells were cultivated to early exponential phase ($OD_{600} \sim = 0.2$), at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) and pAzF were supplemented with the final concentration of 1 mM each. The cultures were further grown for 8 h at 30°C with shaking to saturation. Then, the bulk fluorescence (525 nm) of each well was read in Synergy2 (BioTek) and normalized by the cell number measured by the optical density at 600 nm.

Analysis of GFP Expression by Intact Cell Fluorescent Measurements

Liquid cell cultures of strains harboring plasmid based OTSs and GFP reporter plasmids were inoculated from frozen stocks and grown to confluence overnight in 3 mL of 2xYT media at 37 °C. Cultures were then inoculated at 1:50 dilution in 3 mL of 2xYT media with necessary antibiotics, and the cells were allowed to grow at 37°C to an OD_{600} of 0.5–0.7 in a shaking plate incubator at 250 r.p.m. (∼1.5 h). IPTG and pAzF were added to a final concentration of 1 mM to induce protein expression for 8 h. About 5.5×10^7 of cells of each sample were collected and washed twice with 200 μL of 1 × PBS buffer. The cells were resuspended in 100 μL of 1 × PBS buffer in 96-well plate. The fluorescence of sfGFP was measured on a BioTek Synergy 2 spectrophotometric plate reader using excitation and emission wavelengths of 488 and 525 nm, respectively. Fluorescence signals were normalized by the OD_{600} reading.

For in vivo fluorescence measurements in the promoter library assays, total 27 plasmids, plasmid JGP1 to 27, were transformed into BL21(DE3) individually. Three millilitre of LB media was inoculated with a single colony in a test tube and grown overnight at 37°C. Next, saturated culture was diluted 50-fold into 100 μL of LB media in 96-well plates in triplicate. Cultures and inducers were added individually to each well. Cultures were grown in Synergy (BioTek) plate at 37°C until early mid-exponential phase at which point the cultures were induced with 1 mM IPTG, 0.02% arabinose, and 0 or 10 mM pAzF. Fluorescence and OD_{600} readings were taken at 7.5 h after induction for maximum fluorescence/OD_{600} signal.

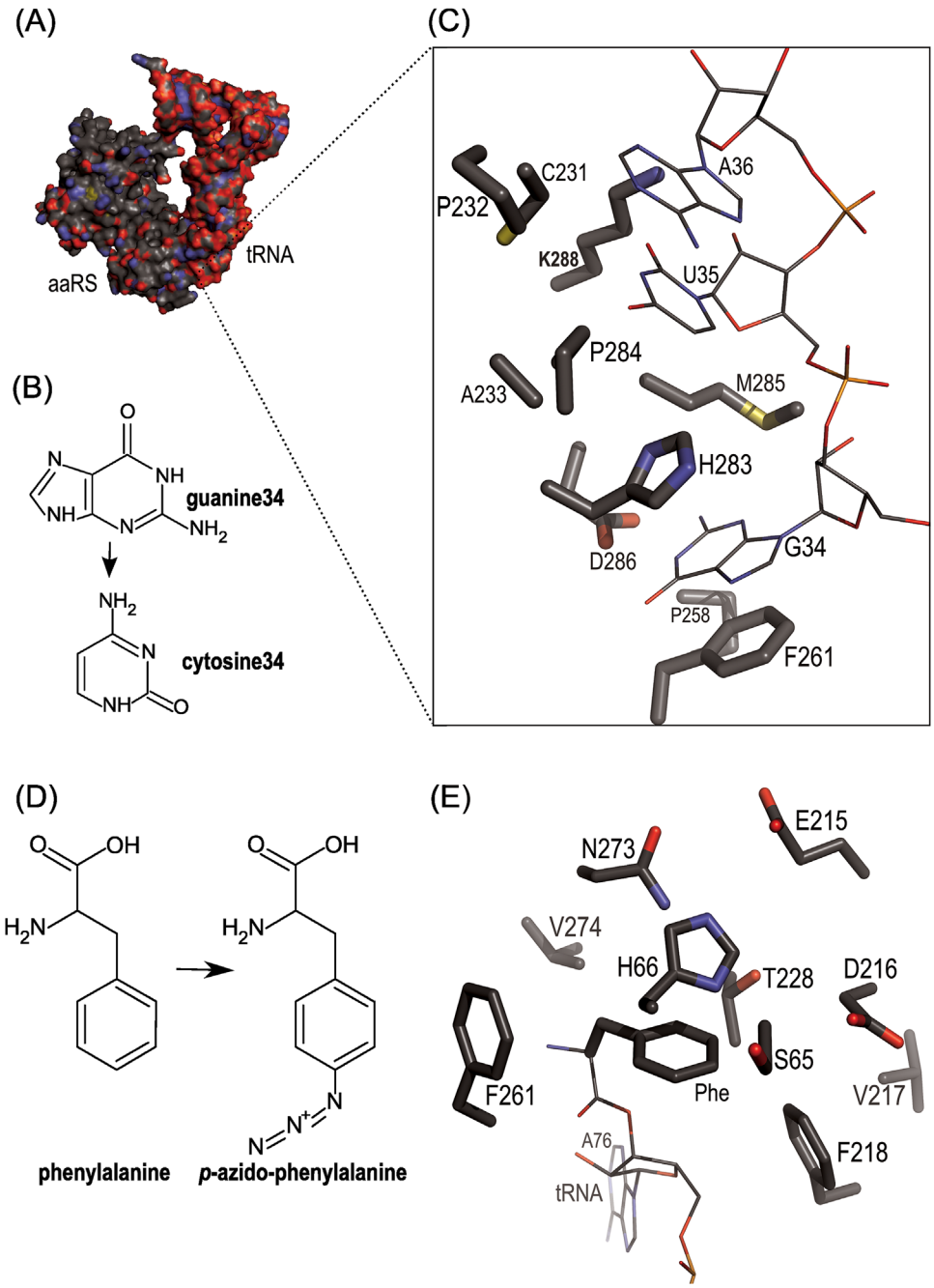

Figure 1. Three-dimensional structure information and library design. (**A**) Crystal structure of the *M. jannaschii* tyrosyl-tRNA synthetase (left: aaRS) and tyrosyl-tRNA^{Tyr} (right: tRNA) complex (PDB: 1J1U); (**B**) Molecular structure of guanine and cytosine; (**C**) tRNA-recognition domain of the synthetase and anticodon loop of tRNA. The residues C231, P232, A233, P258, F261, H283, P284, M285, D286, and K288 in tRNA-recognition domain underwent mutagenesis. The nucleotide G34 in tRNA was changed to cytosine for amber codon suppression. (**D**) Molecular structure of phenylalanine and p-azido-phenylalanine (pAzF); (**E**) The amino acid-binding pocket of elongation factor Tu and phenylalanyl-tRNA (PDB: 1OB2). Ten amino acids: S65, H66, E215, D216, V217, F219, T228, F261, N273, and V274 were mutated to alanine in this study. Phe, phenylalanine charged in tRNA molecule.

For the evaluation of solubility of sfGFP harboring pAzF, 5 $OD_{600} \times$ mL of cells of each sample was collected and washed twice with 500 μL of PBS buffer. Then, the cells were resuspended in 500 μL of 1 × PBS buffer. The cells were further sonicated with 250–300 joules to extract total protein. The soluble fraction was separated by spinning the lysate at 21,000 × g for 10 min. The insoluble fraction was prepared by suspending the pellet in 500 μL of PBS buffer. The total protein, soluble fraction, and insoluble fraction were resolved in NuPAGE™ Novex™ 4–12% Bis-Tris Protein Gels (Invitrogen, CA).

Mass Spectrometry

The purified protein was analyzed by nanocapillary LC-MS using a 100 mm × 75 μm ID PLRP-S column in-line with an Orbitrap Elite

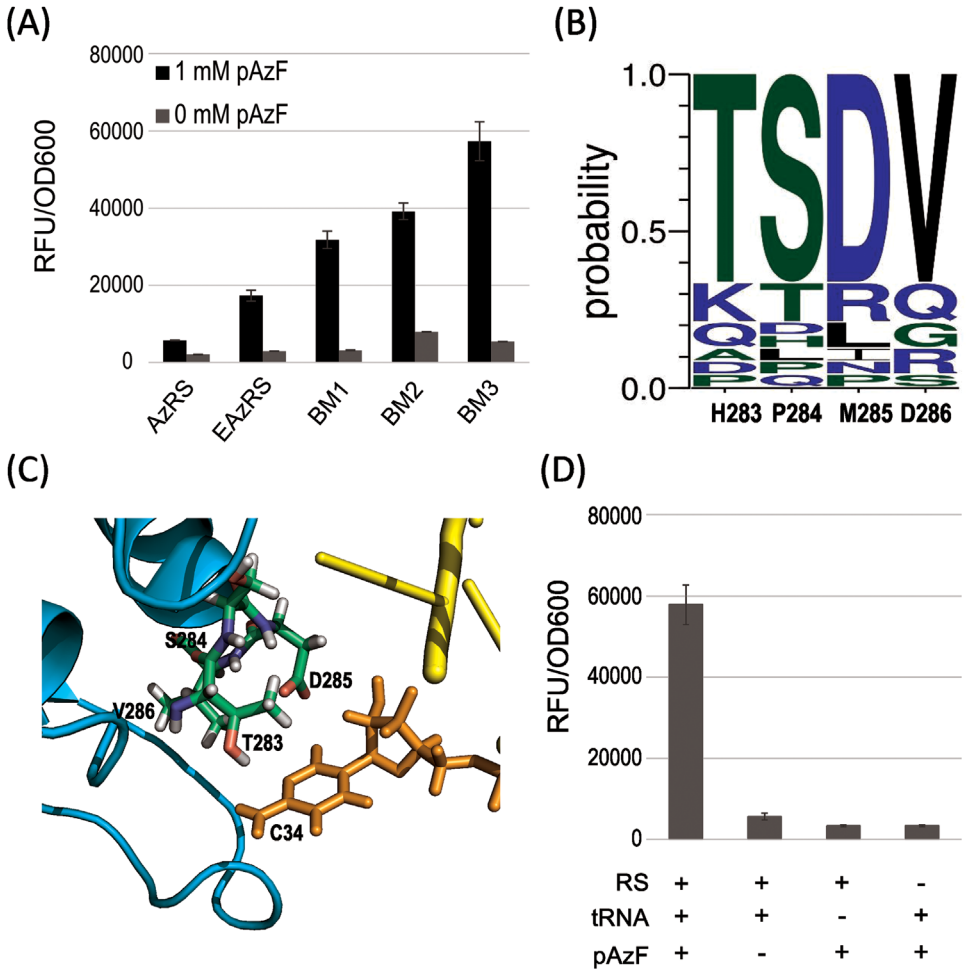

Figure 2. The evolution and identification of beneficial variants of *p*-azido-phenylalanine tRNA synthetase (pAzFRS). (**A**) Representative variants identified during evolution. AzRS: the progenitor pAzFRS first reported by Schultz's group (Chin et al., 2002b). EAzRS: the pAzFRS harboring a D286R mutation reported by Yokoyama's group (Kobayashi et al., 2003). BM1: the top variant obtained from the selection of RsLib1 with the mutations: P258A, F261P, H283L, M285F, and D286Y. BM2: A beneficial variant identified from RsLib2 with a single-site mutation of P284V in addition to the known D286R mutation. BM3: a beneficial variant from the selection of RsLib3 with the mutations: H283T, P284S, M285D, and D286V. BM3 mutant showed approximate eightfold improvement over the parental pAzFRS and threefold over the reported D286R mutant. (**B**) Sequence convergence after the selection of RsLib3. (**C**) Highlighted mutated sites of BM3 in 3-D model of pAzFRS. (**D**) Negative control experiments for BM3 lacking one component for ncAA incorporation. RS, BM3 synthetase; tRNA, o-tRNA; pAzF, p-azido-phenylalanine. Standard deviation accounts for the day-to-day variation in three independent experiments and three colonies were picked for each samples.

(Thermo-Fisher, Waltham, MA). All MS methods included the following events: (i) FT scan, m/z 400–2000, 120,000 resolving power and (ii) data-dependent MS/MS on the top two peaks in each spectrum from scan event one using higher-energy collisional dissociation (HCD) with normalized collision energy of 25, isolation width 15 m/z, and detection of ions with resolving power of 60,000. All data were analyzed using QualBrowser, part of the Xcalibur software packaged with the ThermoFisher Orbitrap Elite.

Results

Evolution of p-azido-phenylalanine–tRNA Synthetase

We hypothesized that mutations at or near the anti-codon recognition domain of the pAzFRS might yield variant aaRSs that more efficiently incorporate pAzF at the amber codon. This hypothesis was based on recent works that show that impaired binding of the pAzFRS to its cognate o-tRNA$_{CUA}$ might reduce efficiency of the system (Amiram et al., 2015), likely because the native *M. jannaschii* TyrRS (*Mj*TyrRS) normally recognizes the GUA anticodon rather than the CUA anticodon to decipher the amber stop codon (Kobayashi et al., 2003). The following general approach was used to improve the interaction between the o-tRNA anti-codon loop and the pAzFRS. First, we used crystal structure-guided information to generate a combinatorial mutation library. Simultaneously, we carried out site-directed mutagenesis to interrogate several individual residues. At last, we sequenced and characterized improved OTS variants obtained via selections.

Crystal structure analysis revealed at least 10 residues in the tRNA-recognition domain of the pAzFRS as targets for modification (Fig. 1). Since a total random combinatorial mutation library of all residues was not feasible (20 ^ 10), we created three distinct

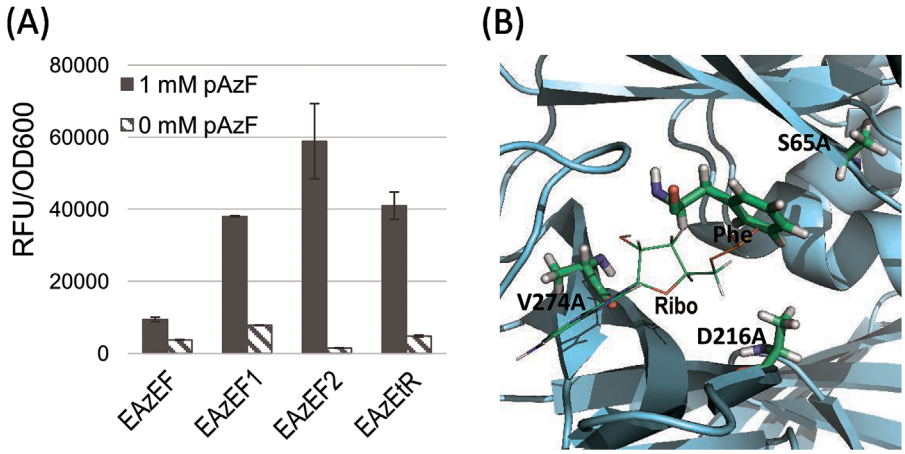

Figure 3. The evolution and identification of beneficial EF-Tu mutants for *p*-azido-phenylalanine (pAzF) incorporation. (**A**) pAzF incorporation efficiency accomplished by wildtype and evolved synthetase and EF-Tu molecules, respectively. EAzEF, plasmid-born wildtype EF-Tu and EAzRS (D286R); EAzEF1, plasmid-born beneficial EF-Tu variant EFBM1 and EAzRS (D286R); EAzEF2, plasmid-born evolved EF-Tu EFBM1 and evolved synthetase BM3; EAzEtR, plasmid-born evolved EF-Tu EFBM1 and evolved synthetase BM3 working with a previously optimized orthogonal tRNA (Young et al., 2010). All experiments were performed in DH10-beta cells with the expression of native EF-Tu from genome. (**B**) Three dimensional model of EFBM1 with highlighted mutated sites: S65A, D216A, and V274A. Phe, the phenylalanine residue of the aminoacyl-tRNA substrate; Ribo, the ribose ring of the A76 of the aminoacyl-tRNA substrate. Standard deviation accounts for the day-to-day variation in three independent experiments and three colonies were picked for each samples.

libraries from the residues identified and sought to evolve the pAzFRS with a plasmid-based orthogonal translation system using a positive and negative selection system previously described (Chin et al., 2002b). The first library (RsLib1) diversified six residues (A233, P258, F261, H283, M285, and D286) in close proximity to the o-tRNA C34 nucleotide (wildtype G34). After two rounds of positive-negative selection, 300 surviving pAzFRS variants were further evaluated quantitatively for their ability to produce superfolder green fluorescent protein (sfGFP) with one amber codon (sfGFP(1UAG)). Direct comparisons to the progenitor enzyme were carried out in a 96-well plate fluorescence-based assay (Supplementary Fig. S2). Sequencing revealed that 25% of the isolated clones converged to a single variant (Supplementary Table S2). The top variant, termed BM1 (P258A, F261P, H283L, M285F, and D286Y) showed approximately a fivefold increase in normalized fluorescence as compared to the progenitor aaRS and a twofold increase as compared to a reported variant harboring D286R single-point mutation, which we term EAzRS in our study (Kobayashi et al., 2003) (Fig. 2A).

Having identified improvements from modifying the six residues closest to the o-tRNA C34 nucleotide, we next sought to explore the impact of another four amino acid residues in the anticodon binding domain (C231, P232, P284, and K288). These residues were selected because they are close to the o-tRNA U35 and A36 nucleotides of the anticodon, and we hoped they would additionally contribute towards aaRS-$tRNA_{CUA}$ binding optimization. We chose here to use a NNK library to diversify the four positions individually, which still covers all 20 amino acids at each single site. Given the importance of the D286R mutation, shown in our initial selection and in previous works, we chose to incorporate this into our starting pAzFRS gene template. Due to the targeted nature of this library (RsLib2), the four target residues were mutated to NNK for a total of 80 variants ($20 + 20 + 20 + 20$). We picked 600 clones to ensure complete library coverage and screened for improved production of sfGFP(1UAG) using an in vivo fluorescence assay in the presence and in absence of pAzF. This screening identified that the mutant BM2 with a P284V mutation resulted in a 2.2-fold improvement in sfGFP(1UAG) synthesis as compared to the EAzRS (Fig. 2A) (Supplementary Table S3).

Based on the identification of mutants P284V and D286R in a single secondary structure turn in pAzFRS, we created a third library (RsLib3) focusing only on this turn element directly (H283, P284, M285, and D286). Specifically, we generated a combinatorial mutation library at all four positions and performed two rounds of positive-negative selection using the *cat*/barnase system described above (Supplementary Fig. S2). Individual colony analysis of twenty-four mutants revealed that 66% or 16 of 24 converged to a single sequence of H283T, P284S, M285D, and D286V (Fig. 2B and C; Supplementary Table S4). This top variant, termed BM3, with the T283-S284-D285-V286 mutations, was capable of producing ~8-fold more sfGFP(1UAG) than the progenitor enzyme (Fig. 2A). As a control, we verified that expression of sfGFP(1UAG) with the BM3-encoding plasmid was dependent on both the o-tRNA-bearing plasmid and pAzF (Fig. 2D).

Engineering EF-Tu by a Combinatorial Alanine-Scanning Method to Improve pAzF Incorporation

Once o-tRNAs are charged with an ncAA of interest by an o-aaRS, the charged tRNAs must be shuttled to the ribosome by the EF-Tu. EF-Tu recognizes the ncAA-tRNA substrate at two regions: (i) the amino acid-binding pocket and tRNA acceptor stem with residues from the β-barrel domain 1 and GTPase domains; and (ii) the T-stem region of the tRNA with residues from the β-barrel domain 2 and the GTPase domain (Nissen et al., 1995). With the goal of engineering the interaction between pAzF-$tRNA_{CUA}$ and EF-Tu, we sought to identify variants in the amino acid-binding pocket of EF-Tu that might favor pAzF-o-tRNA delivery to the ribosome.

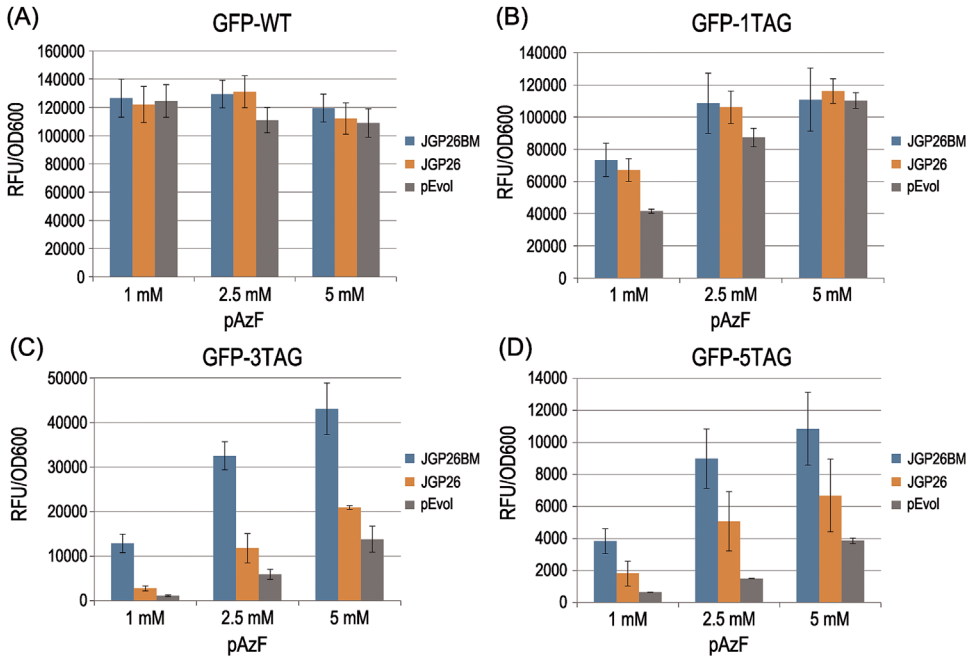

Figure 4. Characterization and optimization of pAzF incorporation and protein expression in BL21(DE3) using an evolved synthetase and EF-Tu. The pAzF incorporation efficiency and protein yield of the evolved synthetase and EF-Tu were demonstrated by suppression 1, 3, and 5 amber codons at 1, 2.5, and 5 mM of pAzF, respectively. Wildtype GFP expression was used as standard of protein yield. GFP-WT, wildtype GFP (panel **A**); GFP-1TAG, GFP with 1 amber codon (panel **B**); GFP-3TAG, GFP with 3 amber codons (panel **C**); GFP-5TAG, GFP with 5 amber codons (panel **D**). JGP26: a 3-part plasmid containing the synthetase, EF-Tu, and tRNA with the optimized promoter usage: lpp promoter for synthetase and EF-Tu, and PtacI promoter for tRNA; JGP26BM: a 3-part plasmid containing the evolved synthetase (BM3), and EF-Tu (EFBM1) and tRNA with the optimized promoter usage: lpp promoter for BM3 and EFBM1, and PtacI promoter for tRNA. Evol: pEvol-pAzF plasmid reported in previous publications (Young et al., 2010). Standard deviation accounts for the day-to-day variation in three independent experiments and three colonies were picked for each samples.

Based on the 3-dimensional structure of EF-Tu (PDB ID: 1OB2), we identified a total of 10 amino acids around the amino acid-binding pocket of EF-Tu (S65, H66, E215, D216, V217, F218, T228, F261, N273, and V274) to modify (Supplementary Fig. S3). As with the pAzFRS evolution, there are too many residues to completely diversify and perform a complete screening of the library. Thus, we chose to carry out a more constrained design by focusing on the construction of an alanine scanning library. By mutating EF-Tu amino acid-binding pocket residues to the smaller amino acid alanine, the pocket would essentially be enlarged for the ncAA. This library of 1024 members was subjected to screening by intact cell fluorescent measurements and its ability to produce sfGFP (1UAG). In this case, the EAzRS synthease (the pAzFRS with single D286R mutation) was used along with its cognate o-tRNA partner, because the work here was done in parallel to the synthetase evolution experiments described above. Beneficial EF-Tu mutants were identified by screening for the ability to improve synthesis of sfGFP(1UAG) as compared to wild-type EF-Tu in a 96-well plate fluorescence assay (Supplementary Fig. S2B). The best mutant, EFBM1, showed approximately a four-fold improvement as compared to the plasmid-bearing the native EF-Tu (Fig. 3A, the bar "EAzEF1"). Individual colony sequencing revealed that the EFBM1 variant carried three alanine mutations at S65A, D216A, and V274A. The S65A and D216A mutations are located near the phenylalanine moiety, so the increased size of the EF-Tu amino acid-binding pocket likely allows for improved pAzF accommodation. V274A is located close to the acceptor stem of the tRNA moiety. This mutation most likely alleviates the steric hindrance of the aminoacylated acceptor stem, perhaps allowing improved acceptance of the non-native aminoacyl-tRNA, and suggests that the tRNA acceptor arm is key engineering target for future efforts. Finally, we combined the top variants isolated from the aaRS evolution (BM3) and the EF-Tu evolution (EFBM1). The combined system with improved o-$tRNA_{CUA}$ binding and pAzF-tRNA binding resulted in a sixfold increase in sfGFP (1UAG) production compared to the combination of EAzRS and wildtype EF-Tu (Fig. 3A, the bar "EAzEF2").

Enhanced ncAA Incorporation into Proteins by Evolved aaRS and EF-Tu

Even though we observed a sixfold increase in sfGFP(1UAG) production with the combination of evolved pAzFRS and EF-Tu compared to their parental parts, we also noticed that the co-expression pAzFRS and EF-Tu (Fig. 3A sample "EAzEF") resulted in a lowered sfGFP(1UAG) production compared to the cell expressing only pAzFRS (Fig. 2A sample "EAzRS"). Apparently, co-expression multiple components negatively influenced the cell growth and protein production ability. To better evaluate the evolved translation system components, we next chose to find a vector to optimize the expression of multiple components in an able host. The pEVOL and pUltra vectors are

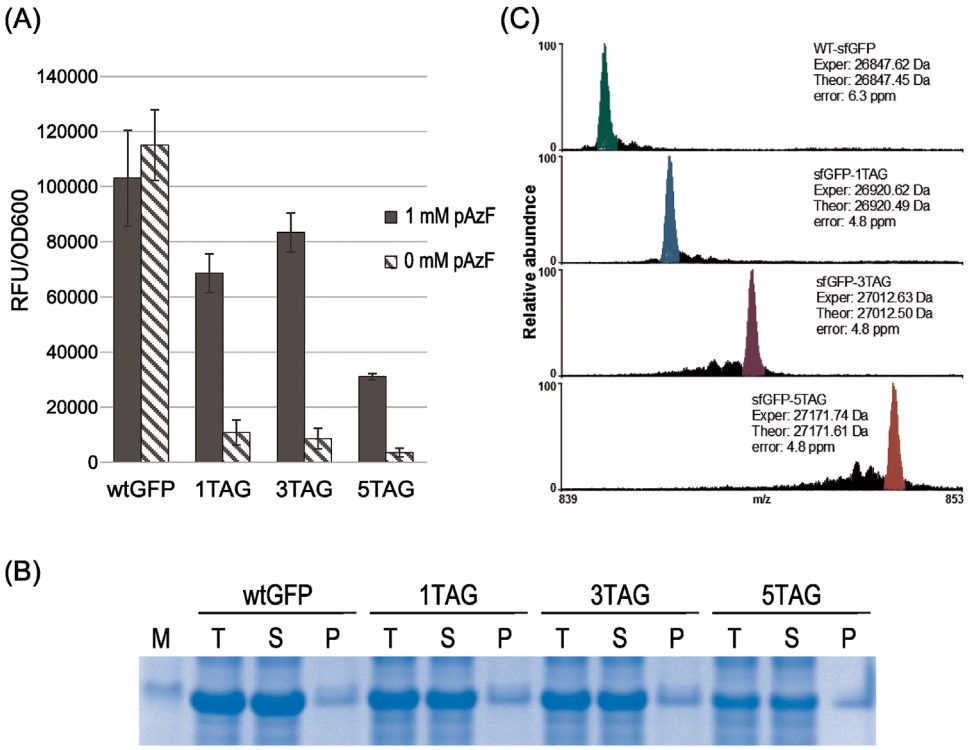

Figure 5. Fidelity and efficiency of pAzF incorporation at multiple amber codon at the absence of release factor 1. (**A**) The active protein yield by suppressing 1, 3, and 5 UAGs in reading frame of GFP by the evolved synthetase and EF-Tu in the plasmid JGP26BM. The pAzF was supplied at 1 mM or 0 mM. The host is a release factor 1-deficient *E. coli* strain (*C321.ΔA*) with all amber codons in genome replaced with ochre codon (Isaacs et al., 2011; Lajoie et al., 2013). (**B**) The solubility of wildtype GFP, GFP with 1 pAzF, 3 pAzFs and 5 pAzFs are high and consistent. T, total protein; S, soluble fraction; P, precipitated fraction. wtGFP, wildtype GFP; 1TAG, GFP with one pAzF; 3TAG, GFP with 3 pAzFs; 5TAG, GFP with 5 pAzFs; M, Molecular weight standard. (**C**) Spectrum of wildtype and various modified sfGFP samples, obtained by top-down mass spectrometry and illustrating site-specific incorporation of pAzF at single and multiple sites. Major peaks (color) in each spectrum agree with the theoretical peaks, respectively. "Exper" indicates experimentally obtained protein mass and "Theor" indicates theoretically calculated protein mass. Standard deviation accounts for the day-to-day variation in three independent experiments and two colonies for each samples.

routinely used in the field as efficient plasmid architectures (Chatterjee et al., 2013; Young et al., 2010). However, they are both designed for orthogonal tRNA/aaRS pairs. The inclusion of three translation system components (aaRS, EF-Tu, tRNA) as compared to two that are usually used (aaRS, tRNA), necessitated us to explore different plasmid architectures.

To achieve our goals, we first sought to place all OTS components on a single plasmid to avoid unnecessary plasmid/antibiotic burden on the cell. Balancing expression levels of orthogonal translation components is essential for cellular fitness and functional activity (Hong et al., 2014a). Thus, the key idea was to identify expression conditions that maximized pAzF incorporation into proteins while alleviating physiological burden to cell growth. We created a library of plasmids carrying promoters of varying strengths for each component: T7 promoter (high activity) (Ikeda et al., 1992), the PtacI promoter (medium activity) (de Boer et al., 1983), and the lpp promoter (low activity) (Inouye and Inouye, 1985). Promoter activity was validated by placing each promoter in front of a sfGFP gene and assessing normalized fluorescence activity (Fluorescence per OD_{600}). The relative ratios of sfGFP expression were 8:3:1 for the T7, PtacI, and lpp promoters, respectively, which verified different promoter activity levels (Supplementary Fig. S5).

We next built 27 individual plasmids, JGP1-27, comprising all possible promoter combinations of the pAzFRS, o-tRNA derived from the pEVOL plasmid (Chin et al., 2002b) and *E. coli*'s endogenous EF-Tu, and screened the library for improved production of sfGFP(1UAG) using an in vivo fluorescence assay in the presence pAzF. We observed that the highest protein yield and fastest cellular growth were obtained, when using the plasmid JGP26, which has the lpp promoter regulating the o-aaRS and EF-Tu, and a PtacI promoter regulating the o-tRNA (Supplementary Fig. S1 and S6). Using this plasmid architecture, we then constructed JGP26BM containing our evolved translation system components BM3 and EFBM1.

Of note, we were somewhat surprised by the fact that the expression of the aaRS from PtacI led to nearly universal drop in performance (Supplementary Fig. S6). Previously, use of this promoter for aaRS expression was quite successful for ncAA incorporation (Chatterjee et al., 2013). We believe our observation is likely a result of two factors. First, our plasmid contains a pBR322 origin, rather than the p15A origin. The pBR322 origin maintains more plasmid copies than the p15A origin (~20 vs. ~10). Second, we co-expressed three components of the translation system (aaRS, EF-Tu, and tRNA) in cell. We suspect that the burden caused by expressing these three components may lead to a different result

than when expressing only two components (aaRS, tRNA). All of this suggests that promoter selection can be affected by many contextual factors, such as, plasmid architecture, orthogonal components, and host strain, among others.

We next measured enhancements in amber codon suppression and protein yield resulting from our evolved translation system by comparing JGP26, JGP26BM, and pEVOL (Supplementary Fig. S1) (Young et al., 2010). In our initial test, we co-transformed BL21 (DE3) with a plasmid carrying the sfGFP gene with 1, 3, or 5 UAGs, or wild type equivalents and the orthogonal translation system plasmids (Fig. 4A). We used fluorescence assay to quantify sfGFP expression. Our results indicated a 2-, 10-, and 5-fold increase in incorporation of pAzF into sfGFP(1UAG) (Fig. 4B), sfGFP(3UAG) (Fig. 4C), and sfGFP(5UAG) (Fig. 4D). While the pEVOL system data is simply to provide a reference for the community, the key comparison is between JGP26BM (evolved components) and JGP26 (parental components). JGP26BM and JGP26 showed the same expression levels of sfGFP(1UAG) under these conditions (Fig. 4B). Our data indicate that the original pAzFRS system, which has been optimized in many reports over the past decade, is fairly efficient for its ability to introduce a single non-canonical amino acid at a single amber codon without our evolved components. However, with an increasing number of amber codons, where the orthogonal translation system is pushed more, the difference between parental and evolved parts increased. We also adjusted the concentration of pAzF in the media from 1 mM to 2.5 mM to 5 mM (Fig. 4). Increasing pAzF concentrations increase overall protein production levels and reduced slightly the beneficial effect of our evolved system. Curiously, single-UAG GFP always showed a lower expression level than that of wildtype GFP (∼70%) in our experiments, when we used only 1 mM of pAzF. However, the evolved components function much more efficiently than progenitor systems and, in all cases, we observed weak expression of proteins with five amber codons.

Evolved and Tuned OTS System Enables Accurate ncAA Insertion at Multiple UAG Sites

It is known that UAG-suppression competes with termination mediated by RF1. While deletion of RF1 is lethal in normal cells, there has been growing interest to alleviate this limitation by inactivating RF1 (Heinemann et al., 2012; Johnson et al., 2011, 2012; Mukai et al., 2010; Ohtake et al., 2012). More recently, efforts to completely recode the genome of *E. coli* to replace all 321 occurrences of the UAG codon with the synonymous UAA codon, permitted the deletion of RF1 and complete reassignment of the amber codon translation function (Isaacs et al., 2011; Lajoie et al., 2013). Here, we set out to use the genomically recoded *C321.ΔA* strain and our engineered translation system components to investigate the ability to enable efficient multi-site ncAA incorporation. For this study, we co-transformed the plasmid JGP26BM harboring BM3 and EFBM1 with the sfGFP reporter gene constructs containing 1, 3, or 5 UAGs, or the wild type equivalents. We found expression levels with sfGFP constructs with a single UAG amber codon were comparable in wild-type sfGFP, and the absence of RF1 to be increasingly beneficial for multi-site ncAA incorporation (Fig. 5A). Specifically, our analysis revealed that the evolved orthogonal translation system could suppress 3-UAG codons at production levels of approximately 80% of wild type sfGFP in the *C321.ΔA* strain. Additionally, we observed significantly enhanced production of sfGFP(5UAG) compared the yield we observed in BL21 (DE3) cell. The solubility of wildtype sfGFP and its derivatives with 1 pAzF, 3 pAzFs, and 5 pAzFs was also assessed. We observed that nearly all of the sfGFP produced was soluble (Fig. 5B).

We next examined the fidelity of multi-site pAzF incorporation into the sfGFP reporter constructs. Specifically, we carried out top-down mass spectrometry (i.e., MS analysis of whole intact proteins) to detect and provide semi-quantitative information for the incorporation of pAzF into sfGFP. Figure 5C shows the 32+ charge state of sfGFP and clearly illustrates mass shifts corresponding to the incorporation of one, three, and five pAzF residues. Site-specific incorporation of pAzF, as detected by MS, was greater than 95% in all samples (Fig. 5C), with less than 3 ppm difference between experimental and theoretical protein masses. In other words, we achieved efficient, high yielding, and pure site-specific pAzF incorporation into sfGFP using our evolved translation system and the *C321ΔA* strain.

Discussion

Our work joins an ever-growing collection of reports highlighting the ability to repurpose the translation machinery for genetic code expansion. In this study, we demonstrated the importance of pursuing a systems engineering approach when evolving multiple components of OTSs (e.g., aaRS and EF-Tu) and using RF1 deficient strains to enable high-level and accurate multi-site incorporation of ncAAs into a protein. One of our goals in this manuscript was to explore the possibility that EF-Tu might be a more general target for translation system engineering, especially given that interactions between EF-Tu and aminoacyl-tRNAs are tuned to perfectly match natural amino acids with cognate tRNAs (Schrader et al., 2009). Previous reports suggest that EF-Tu engineering can be beneficial for bulky amino acids (Doi et al., 2007) and charged amino acids as shown for phosphoserine (Park et al., 2011) and phosphotyrosine (Fan et al., 2016). Our results suggest that, it might be a more general target, although additional studies are needed to understand the generality of evolving EF-Tu alongside aaRSs to facilitate integration of a wide range of ncAA targets.

In terms of the pAzFRS, we used a step-wise library selection and screening procedure to improve aaRS activity by ∼8-fold compared to the parental enzyme. The optimization of residues H283-D286 are mainly responsible for the enhancement. In terms of EF-Tu, we used an alanine scanning approach to screen a 1024 member library and observed several mutants having up to fourfold improvement, each with mutations in different residues. By combining our best mutants and tuning expression levels, we observed a 2-, 10-, and 5-fold increase in incorporation of pAzF into sfGFP(1UAG), sfGFP(3UAG), sfGFP(5UAG), respectively, as compared to the progenitor system. We then used top-down mass spectrometry to confirm multiple instances of ncAA incorporation with near perfect fidelity. Taken together, the engineered translation system reported here demonstrates improved performance for protein yield and accurate protein production as compared with the progenitor enzymes.

Of note, we did not observe an additive effect when combining the best aaRS and EF-Tu mutants. One could imagine that alone, an engineered EF-Tu could compensate for poorer synthetase activity. However, if the synthetase is evolved and does a better job at providing the aminoacyl-tRNA substrates then the beneficial effects of the EF-Tu could be muted, or not additive. Thus, our observation may point to the need of future efforts to coordinately tune all system components at the same time.

It is important to note that context plays a key role in assessing the beneficial impact of different evolved orthogonal translation components. For instance, when suppressing only one amber codon with a high concentration of pAzF (5 mM) in BL21(DE3) strain, the evolved mutants do not show an advantage over their parental ones, while when suppressing three amber codons with only 1 mM of pAzF, the evolved mutants showed 10-fold higher efficiency over the parental parts even though the GFP yield is comparatively low. Based on our work and that of others, it is becoming clear that a number of factors influence the performance of OTSs, such as mutant orthogonal translation components, ncAA concentration, amber codon number, strain, and presence or absence of RF1. These factors should be carefully taken into consideration when characterizing beneficial mutants, measuring the ratio of performance/cost in protein production, and designing evolution strategies. For example, we initially planned to use the *C321.ΔA* as our host for the entire project. However, when initially applying this strain to the chloramphenicol acetyl-transferase (CAT) and barnase-based selection strategy, we found that *C321.ΔA* strain did not allow us to select for variants with significantly improved properties (data not shown). We went on to test BL21 (DE3), MG1655, DH5α, and DH10-beta, and finally we found that only DH10-beta provided us the best conditions with the context of our plasmids for a high-throughput CAT/barnase-based selection method. Even though we found benefits for carrying out selections in an RF1 positive strain, we believe that the protein expression should be carried out in an RF1 negative strain, such as *C321.ΔA*.

In sum, our engineering effort demonstrates new details regarding interactions between proteins involved in translation that might be taken into consideration in future designs for constructing similar engineered translation systems. For example, we suspect that promoter tuning of the evolved mutants could lead to further enhancements beyond those observed here. Indeed, we believe that only by treating the translation apparatus as a complex system, whereby all biological parts involved in protein biosynthesis are coordinately optimized (e.g., codons, tRNA, aaRS, EF-Tu, and the ribosome together), will we enable more diverse genetic codes and advanced capabilities. Such advances will be important for harnessing a dramatically expanded genetic code for manufacturing novel therapeutics (Hubbell, 2010; Wals and Ovaa, 2014; Zimmerman et al., 2014), synthesizing genetically-encoded materials (Lee et al., 2013; Romano et al., 2011; Zhong et al., 2012), advancing medicine (Cohen, 2002; Hubbell, 2010; Krishnamoorthy et al., 2013; Pray, 2008; Ramos, 2005; Voronkov et al., 2011; Wals and Ovaa, 2014; Zimmerman et al., 2014), and elucidating fundamental biological insights (e.g., the histone code (Banerjee and Chakravarti, 2011; Lee et al., 2013; Neumann et al., 2009).

This work was supported by the Defense Advanced Research Projects Agency (N66001-12-C-4211), the David and Lucile Packard Foundation, the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust, and the Arnold and Mabel Beckman Foundation. This research was carried out in collaboration with the National Resource for translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences. We thank Prof. Peter G. Schultz for providing the pEVOL plasmid.

References

Albayrak C, Swartz JR. 2014. Direct polymerization of proteins. ACS Synth Biol 3(6):353–362.

Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C, Söll D. 2007. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci USA 104(9):3141–3146.

Amiram M, Haimovich AD, Fan C, Wang YS, Aerni HR, Ntai I, Moonan DW, Ma NJ, Rovner AJ, Hong SH, Kelleher NL, Goodman AL, Jewett MC, Soll D, Rinehart J, Isaacs FJ. 2015. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat Biotechnol 33(12):1272–1279.

Axup JY, Bajjuri KM, Ritland M, Hutchins BM, Kim CH, Kazane SA, Halder R, Forsyth JS, Santidrian AF, Stafin K, Lu Y, Tran H, Seller AJ, Biroc SL, Szydlik A, Pinkstaff JK, Tian F, Sinha SC, Felding-Habermann B, Smider VV, Schultz PG. 2012. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. Proc Natl Acad Sci USA 109(40):16101–16106.

Banerjee T, Chakravarti D. 2011. A peek into the complex realm of histone phosphorylation. Mol Cell Biol 31(24):4858–4873.

Barrett OP, Chin JW. 2010. Evolved orthogonal ribosome purification for in vitro characterization. Nucleic Acids Res 38(8):2682–2691.

Boniecki MT, Vu MT, Betha AK, Martinis SA. 2008. CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. Proc Natl Acad Sci USA 105(49):19223–19228.

Bröcker MJ, Ho JML, Church GM, Söll D, O'Donoghue P. 2014. Recoding the genetic code with selenocysteine. Angew Chem Int Ed Engl 53(1):319–323.

Bundy BC, Swartz JR. 2010. Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. Bioconjugate chemistry 21(2):255–263.

Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG. 2013. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. Biochemistry 52(10):1828–1837.

Chatterjee A, Xiao H, Schultz PG. 2012. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. Proc Natl Acad Sci USA 109(37):14841–14846.

Chin JW, Martin AB, King DS, Wang L, Schultz PG. 2002a. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. Proc Natl Acad Sci USA 99(17):11020–11024.

Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. 2002b. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. J Am Chem Soc 124(31):9026–9027.

Cho H, Daniel T, Buechler YJ, Litzinger DC, Maio Z, A-MH Putnam, Kraynov VS, Sim B-C, Bussell S, Javahishvili T, Kaphle S, Viramontes G, Ong M, Chu S, Becky GC, Lieu R, Knudsen N, Castiglioni P, Norman TC, Axelrod DW, Hoffman AR, Schultz PG, DiMarchi RD, Kimmel BE. 2011. Optimized clinical performance of growth hormone with an expanded genetic code. Proc Natl Acad Sci USA 108(22):9060–9065.

Cohen P. 2002. Protein kinases-the major drug targets of the twenty-first century? Nat Rev Drug Discov 1(4):309–315.

Davis L, Chin JW. 2012. Designer proteins: Applications of genetic code expansion in cell biology. Nat Rev Mol Cell Biol 13(3):168–182.

de Boer HA, Comstock LJ, Vasser M. 1983. The tac promoter: A functional hybrid derived from the trp and lac promoters. Proc Natl Acad Sci USA 80(1):21–25.

Des Soye BJ, Patel JR, Isaacs FJ, Jewett MC. 2015. Repurposing the translation apparatus for synthetic biology. Curr Opin Chem Biol 28:83–90.

Doi Y, Ohtsuki T, Shimizu Y, Ueda T, Sisido M. 2007. Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. J Am Chem Soc 129(46):14458–14462.

Fan C, Ip K, Soll D. 2016. Expanding the genetic code of *Escherichia coli* with phosphotyrosine. FEBS Lett 590(17):3040–3047.

Forster AC, Tan Z, Nalam MN, Lin H, Qu H, Cornish VW, Blacklow SC. 2003. Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proc Natl Acad Sci USA 100(11):6353–6357.

Fried SD, Schmied WH, Uttamapinant C, Chin JW. 2015. Ribosome subunit stapling for orthogonal translation in *E. coli*. Angew Chem Int Ed Engl 54(43):12791–12794.

Heinemann IU, Rovner AJ, Aerni HR, Rogulina S, Cheng L, Olds W, Fischer JT, Soll D, Isaacs FJ, Rinehart J. 2012. Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. FEBS Lett 586(20):3716–3722.

Hoesl MG, Budisa N. 2012. Recent advances in genetic code engineering in *Escherichia coli*. Curr Opin Biotechnol 23(5):751–757.

Hong SH, Kwon YC, Jewett MC. 2014a. Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. Front Chem 2:34.

Hong SH, Ntai I, Haimovich AD, Kelleher NL, Isaacs FJ, Jewett MC. 2014b. Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth Biol 3(6):398–409.

Hong SH, Kwon YC, Martin RW, Soye BJ, de Paz AM, Swonger KN, Ntai I, Kelleher NL, Jewett MC. 2015. Improving cell-Free protein synthesis through genome engineering of *Escherichia coli* lacking release factor 1. Chembiochem 16(5):844–853.

Hubbell JA. 2010. Drug development: Longer-lived proteins. Nature 467(7319):1051–1052.

Ikeda RA, Ligman CM, Warshamana S. 1992. T7 promoter contacts essential for promoter activity in vivo. Nucleic Acids Res 20(10):2517–2524.

Inouye S, Inouye M. 1985. Up-promoter mutations in the lpp gene of *Escherichia coli*. Nucleic Acids Res 13(9):3101–3110.

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM. 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333(6040):348–353.

Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR. 2008. An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4:220.

Johnson DBF, Wang C, Xu J, Schultz MD, Schmitz RJ, Ecker JR, Wang L. 2012. Release factor one is nonessential in *Escherichia coli*. ACS Chem Biol 7(8):1337–1344.

Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L. 2011. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7(11):779–786.

Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. 2003. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat Struct Biol 10(6):425–432.

Krishnamoorthy S, Liu Z, Hong A, Zhu R, Chen H, Li T, Zhou X, Gao X. 2013. A novel phosphopeptide microarray based interactome map in Breast cancer cells reveals phosphoprotein-GRB2 cell signaling networks. PLoS ONE 8(6):e67634.

Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ. 2013. Genomically recoded organisms expand biological functions. Science 342(6156):357–360.

Lang K, Chin JW. 2014. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem Rev 114(9):4764–4806.

Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW. 2012. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nat Chem 4(4):298–304.

Lee S, Oh S, Yang A, Kim J, Soll D, Lee D, Park HS. 2013. A facile strategy for selective incorporation of phosphoserine into histones. Angew Chem Int Ed Engl 52(22):5771–5775.

Ling J, Soll D. 2010. Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. Proc Natl Acad Sci USA 107(9):4028–4033.

Liu CC, Schultz PG. 2010. Adding new chemistries to the genetic code. Annu Rev Biochem 79(1):413–444.

Liu DR, Magliery TJ, Pastrnak M, Schultz PG. 1997. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc Natl Acad Sci USA 94(19):10092–10097.

Loscha KV, Herlt AJ, Qi R, Huber T, Ozawa K, Otting G. 2012. Multiple-site labeling of proteins with unnatural amino acids. Angew Chem Int Ed Engl 51(9):2243–2246.

Melancon CE, 3rd, Schultz PG. 2009. One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. Bioorg Med Chem Lett 19(14):3845–3847.

Miyake-Stoner SJ, Refakis CA, Hammill JT, Lusic H, Hazen JL, Deiters A, Mehl RA. 2010. Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. Biochemistry 49(8):1667–1677.

Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. 2010. Codon reassignment in the *Escherichia coli* genetic code. Nucleic Acids Res 38(22):8188–8195.

Nehring S, Budisa N, Wiltschi B. 2012. Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code. PLoS ONE 7(4):e31992.

Neumann H, Hancock SM, Buning R, Routh A, Chapman L, Somers J, Owen-Hughes T, van Noort J, Rhodes D, Chin JW. 2009. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell 36(1):153–163.

Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. 2010. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464(7287):441–444.

Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J. 1995. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. Science 270(5241):1464–1472.

Ohtake K, Sato A, Mukai T, Hino N, Yokoyama S, Sakamoto K. 2012. Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a *prfA*-deficient strain of *Escherichia coli*. J Bacteriol 194(10):2606–2613.

Orelle C, Carlson ED, Szal T, Florin T, Jewett MC, Mankin AS. 2015. Protein synthesis by ribosomes with tethered subunits. Nature 524(7563):119–124.

Oza JP, Aerni HR, Pirman NL, Barber KW, Ter Haar CM, Rogulina S, Amrofell MB, Isaacs FJ, Rinehart J, Jewett MC. 2015. Robust production of recombinant phosphoproteins using cell-free protein synthesis. Nat Commun 6:8168.

Park HS, Hohn MJ, Umehara T, Guo LT, Osborne EM, Benner J, Noren CJ, Rinehart J, Soll D. 2011. Expanding the genetic code of *Escherichia coli* with phosphoserine. Science 333(6046):1151–1154.

Plass T, Milles S, Koehler C, Schultz C, Lemke EA. 2011. Genetically encoded copper-free click chemistry. Angew Chem Int Ed Engl 50(17):3878–3881.

Pray LA. 2008. Gleevec: The breakthrough in cancer treatment. Nat Edu 1(1):37.

Quast RB, Mrusek D, Hoffmeister C, Sonnabend A, Kubick S. 2015. Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. FEBS Lett 589(15):1703–1712.

Rackham O, Chin JW. 2005. A network of orthogonal ribosome x mRNA pairs. Nat Chem Biol 1(3):159–166.

Ramos JW. 2005. Cancer research center hotline: PEA-15 phosphoprotein: A potential cancer drug target. Hawaii Med J 64(3):77–80.

Reynolds NM, Ling J, Roy H, Banerjee R, Repasky SE, Hamel P, Ibba M. 2010. Cell-specific differences in the requirements for translation quality control. Proc Natl Acad Sci USA 107(9):4063–4068.

Rogerson DT, Sachdeva A, Wang K, Haq T, Kazlauskaite A, Hancock SM, Huguenin-Dezot N, Muqit MM, Fry AM, Bayliss R, Chin JW. 2015. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol 11(7):496–503.

Romano NH, Sengupta D, Chung C, Heilshorn SC. 2011. Protein-engineered biomaterials: Nanoscale mimics of the extracellular matrix. Biochim Biophys Acta 1810(3):339–349.

Schrader JM, Chapman SJ, Uhlenbeck OC. 2009. Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis. J Mol Biol 386(5):1255–1264.

Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG. 2006. A genetically encoded fluorescent amino acid. Proc Natl Acad Sci USA 103(26):9785–9789.

Tanrikulu IC, Schmitt E, Mechulam Y, Goddard WA, Tirrell DA. 2009. Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. Proc Natl Acad Sci USA 106(36):15285–15290.

Tian F, Lu Y, Manibusan A, Sellers A, Tran H, Sun Y, Phuong T, Barnett R, Hehli B, Song F, DeGuzman MJ, Ensari S, Pinkstaff JK, Sullivan LM, Biroc SL, Cho H, Schultz PG, DiJoseph J, Dougher M, Ma D, Dushin R, Leal M, Tchistiakova L, Feyfant E, Gerber HP, Sapra P. 2014. A general approach to site-specific antibody drug conjugates. Proc Natl Acad Sci USA 111(5):1766–1771.

Umehara T, Kim J, Lee S, Guo L-T, Söll D, Park H-S. 2012. *N*-Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess *N*-acetyl lysine specificity in vitro and in vivo. FEBS Lett 586(6):729–733.

Voronkov M, Braithwaite SP, Stock JB. 2011. Phosphoprotein phosphatase 2A: A novel druggable target for Alzheimer's disease. Future Med Chem 3(7):821–833.

Wals K, Ovaa H. 2014. Unnatural amino acid incorporation in *E. coli*: Current and future applications in the design of therapeutic proteins. Front Chem 2:15.

Wan W, Tharp JM, Liu WR. 2014. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta 1844(6):1059–1070.

Wang J, Kwiatkowski M, Forster AC. 2016. Kinetics of tRNA(Pyl)-mediated amber suppression in *Escherichia coli* translation reveals unexpected limiting steps and competing reactions. Biotechnol Bioeng 113(7):1552–1559.

Wang K, Neumann H, Peak-Chew SY, Chin JW. 2007. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 25(7):770–777.

Wang L, Magliery TJ, Liu DR, Schultz PG. 2000. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J Am Chem Soc 122(20):5010–5011.

Wang L, Schultz PG. 2001. A general approach for the generation of orthogonal tRNAs. Chem Biol 8(9):883–890.

Wang L, Schultz PG. 2002. Expanding the genetic code. Chem Commun (Camb) 7(1):1–11.

Wang L, Xie J, Schultz PG. 2006. Expanding the genetic code. Annu Rev Biophys Biomol Struct 35:225–249.

Wang L, Zhang Z, Brock A, Schultz PG. 2003. Addition of the keto functional group to the genetic code of *Escherichia coli*. Proc Natl Acad Sci USA 100(1):56–61.

Wang N, Ju T, Niu W, Guo J. 2015. Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids. ACS Synth Biol 4(3):207–212.

Wang Y-S, Fang X, Chen H-Y, Wu B, Wang ZU, Hilty C, Liu WR. 2012. Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. ACS Chem Biol 8(2):405–415.

Wiltschi B, Wenger W, Nehring S, Budisa N. 2008. Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues. Yeast 25(11):775–786.

Wu N, Deiters A, Cropp TA, King D, Schultz PG. 2004. A genetically encoded photocaged amino acid. J Am Chem Soc 126(44):14306–14307.

Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG. 2011. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50(11):1894–1900.

Young TS, Ahmad I, Yin JA, Schultz PG. 2010. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. J Mol Biol 395(2):361–374.

Zheng Y, Lajoie MJ, Italia JS, Chin MA, Church GM, Chatterjee A. 2016. Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. Mol Biosyst 12(6):1746–1749.

Zhong Z, Yang H, Zhang C, Lewis JC. 2012. Synthesis and catalytic activity of amino acids and metallopeptides with catalytically active metallocyclic side chains. Organometallics 31(21):7328–7331.

Zimmerman ES, Heibeck TH, Gill A, Li X, Murray CJ, Madlansacay MR, Tran C, Uter NT, Yin G, Rivers PJ, Yam AY, Wang WD, Steiner AR, Bajad SU, Penta K, Yang W, Hallam TJ, Thanos CD, Sato AK. 2014. Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. Bioconjug Chem 25(2):351–361.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.