BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 1667–1677 1667
DOI: 10.1021/bi901947r

Generating Permissive Site-Specific Unnatural Aminoacyl-tRNA Synthetases†

Shigeki J. Miyake-Stoner,‡ Christian A. Refakis,‡ Jared T. Hammill,‡ Hrvoje Lusic,§ Jennifer L. Hazen,‡ Alexander Deiters,§ and Ryan A. Mehl*,‡

‡*Department of Chemistry, Franklin & Marshall College, Lancaster, Pennsylvania 17604-3003, and* §*Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695-8204*

Received November 11, 2009; Revised Manuscript Received December 21, 2009

ABSTRACT: Genetically incorporated unnatural amino acid (UAA) technologies are powerful tools that are greatly enhancing our ability to study and engineer biological systems. Using these techniques, researchers can precisely control the position and number of novel chemical moieties in a protein, via introducing the novel R group of UAAs, that are genetically encoded in the protein's primary structure. The substrate recognition properties of a natural aminoacyl-tRNA synthetase (aaRS) must be modified in order to incorporate UAAs into proteins. Protocols to do so are technically simple but require time and optimization, which has significantly limited the accessibility of this important technology. At present, engineered unnatural aminoacyl-tRNA synthetases (UaaRS) are evaluated on their translational efficiency (the extent to which they allow for incorporation of UAAs into protein) and fidelity (the extent to which they prevent incorporation of natural amino acids). We propose that a third parameter of substrate recognition, permissivity, is equally important. Permissive UaaRSs, whose relaxed substrate recognition properties allow them to incorporate multiple unnatural amino acids (but not natural amino acids), would eliminate the need to generate new UaaRSs for many new UAAs. Here, we outline methods for quickly and easily assessing the permissivity of existing UaaRSs and for generating permissive UaaRSs. In proof of principle experiments, we determined the degree of permissivity of two UaaRSs for a family of structurally related fluorinated UAAs (^{19}F-UAAs). We then increased the permissivity of the initial UaaRSs to allow for incorporation of the family of ^{19}F-UAAs. Finally, we validated the utility of these new ^{19}F-UAAs as probes for fluorine NMR studies of protein structure and dynamics. We expect that results of this work will increase the accessibility of UAA technology and the use of new UAAs in proteins.

The recent engineering of aminoacyl-tRNA synthetase/tRNA pairs (aaRS/tRNA)[1] to enable the production of chemically modified proteins *in vivo* via incorporation of unnatural amino acids (UAAs) has had a dramatic impact on the fields of biochemical research and medicinal proteins (*1*−*7*). While improvements in site-specific incorporation technology have provided access to UAA−protein from prokaryotes and eukaryotes, the field would benefit from access to methods that increase the diversity of UAAs that can be incorporated (*8*−*11*). Currently, the development of site-specific UAA incorporation methods begins with a library of orthogonal aaRS/tRNA pairs, which after selections result in a UaaRS/tRNA pair that can incorporate a UAA in response to a unique codon with high efficiency and fidelity (*1*, *12*). The double-sieve selections (*13*) used to identify UaaRS/tRNA pairs consist of positive selections, which select for the ability to incorporate a specific UAA, and negative selections, which select against natural amino acid structures. The result is a unique UaaRS active site that incorporates the new UAA during translation in response to a unique codon with high efficiency and fidelity. The high efficiency of the UaaRS is key to producing high levels of UAA−protein, while high fidelity ensures that no natural amino acids are incorporated at the programmed site.

The UaaRS selections and characterization methods are technically simple for those well versed in molecular biology but require time and optimization, an effort that investigators interested in generating only a single UaaRS for their particular application may not consider worthwhile. The need for improving the process and increasing accessibility led us to find alternate, less technical methods for generating UaaRSs and characterizing them by effectively monitoring the efficiency and fidelity for the sum of translational processes.

A few recent reports using site-specific methods have utilized what seems to be a relaxed substrate specificity from selected UaaRS/tRNA pairs to incorporate multiple similar UAAs with the same UaaRS *in vivo* (*14*−*17*). We hypothesized that this lack of specificity, a quality we define as permissivity, may be a generally useful feature of UaaRSs. Once a UaaRS/tRNA pair for a specific UAA is selected and shown to have high translational fidelity, additional UAAs with similar structures can be

†This work was supported by F&M Hackman and Eyler funds, NSF-MCB-0448297, Research Corporation (CC6364), and ACS-PRF (42214-GB4). A.D. is a Beckman Young Investigator and a Cottrell Scholar.

*To whom correspondence should be addressed. E-mail: ryan.mehl@fandm.edu. Phone: 717-291-4125. Fax: 717-291-3939.

[1]Abbreviations: aaRS, aminoacyl-tRNA synthetase; Amp, ampicillin; Cm, chloramphenicol; EF-Tu, elongation factor Tu; ESI-Q-Tof, electrospray ionization quadrupole time of flight; ^{19}F-UAA, fluorinated unnatural amino acid; Km, kanamycin; *Mj*, *Methanococcus jannaschii*; NMR, nuclear magnetic resonance; PBS, phosphate-buffered saline; pmdfF, 2,3-difluoro-4-methylphenylalanine; pmF, 4-methylphenylalanine; pmmfF, 3-fluoro-4-methylphenylalanine; pmtfF, 2,3,5,6-tetrafluoro-4-methylphenylalanine; ptfmF, 4-trifluoromethylphenylalanine; ptfmmfF, 3-fluoro-4-trifluoromethylphenylalanine; sfGFP, super-folder GFP; Tet, tetracycline; UAA, unnatural amino acid; UaaRS, unnatural aminoacyl-tRNA synthetase; UAA-sfGFP, unnatural amino acid-containing super-folder GFP.

Published on Web 01/18/2010

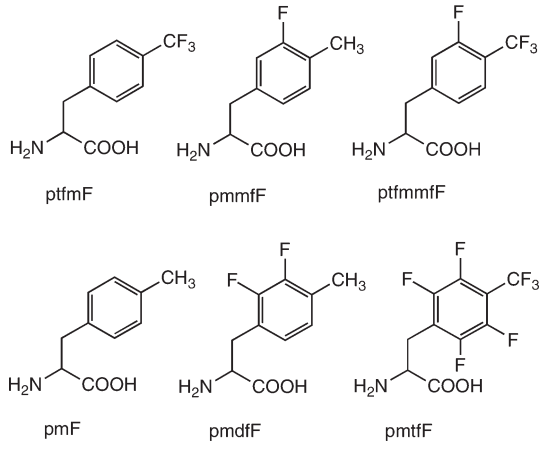


FIGURE 1: Family of unnatural amino acids: 4-trifluoromethylphenylalanine (ptfmF), 3-fluoro-4-methylphenylalanine (pmmfF), 3-fluoro-4-trifluoromethylphenylalanine (ptfmmfF), 4-methylphenylalanine (pmF), 2,3-difluoro-4-methylphenylalanine (pmdfF), and 2,3,5,6-tetrafluoro-4-methylphenylalanine (pmtfF).

incorporated without the need to generate additional UaaRSs. Permissivity may be a feature of many site-specific UaaRSs, and taking advantage of this feature may allow others to avoid performing selections and fidelity characterization. Recently, we showed that the UaaRSs that can site-specifically incorporate *p*-benzoylphenylalanine (*18*) and naphthylalanine (*19*) have varying levels of permissivity for similar structured UAAs (*20*). Since the current positive and negative selections for generating UaaRSs do not impose selective pressure for or against UaaRS permissivity, it is unlikely to be a universal feature for selected UaaRSs. In fact, a few attempts to utilize UaaRS permissivity have also been shown to fail (*17*, *21*). To accurately and quickly determine efficiency, fidelity, and permissivity of UaaRS/tRNA pairs under standard UAA-expression conditions, we developed a fluorescent reporter of UAA−protein production that is functional in autoinduction medium.

Clearly, the current >50 different site-specific UaaRSs will not incorporate all future UAAs (*1*, *14*, *15*, *20*, *22*−*25*). Thus we sought to develop a method to generate generally permissive UaaRS/tRNA pairs and monitor the relationship of efficiency, fidelity, and permissivity of the selection process. We hypothesized that one way to make UaaRSs permissive to a family of structurally related UAAs is to exchange active site residues of UaaRSs that are independently selected for members of the UAA family. To this end, we employed identical standard site-specific selection methods in order to generate UaaRSs for two structurally similar fluorinated UAAs, 4-trifluoromethylphenylalanine (ptfmF) and 3-fluoro-4-methylphenylalanine (pmmfF) (Figure 1). Starting from the same aaRS library of 6×10^7 members, separate double-sieve selections were performed in the presence and absence of ptfmF and pmmfF. Each UaaRS selection resulted in a single unique UaaRS sequence for each UAA that was quantitatively characterized for efficiency and fidelity. We then studied the incorporation efficiency, permissivity, and fidelity of producing UAA−protein using these two unique UaaRSs with a matrix of UAA chemical structures that contain the chemical features of the parent UAAs (Figure 1). We exchanged the unique active site residues in each parent UaaRS with one another to determine whether the hybrid sequences would have high translational permissivity and efficiency, without compromising fidelity. We have shown that construction of generally permissive UaaRSs is possible by assessing hybrids of two different UaaRS sequence solutions and that this is a viable method for developing efficient, high-fidelity UaaRSs. We have also shown that the current UaaRS selection process may not be producing the most efficient UaaRSs. These findings and methods will allow for development of more efficient UaaRSs and effective selection methods, as well as provide access to a family of new ^{19}F-NMR probes for studying protein structure and dynamics (*26*−*31*).

EXPERIMENTAL PROCEDURES

General Methods. Chemical reagents were purchased from Sigma-Aldrich and Fisher Scientific and used without further purification. Unnatural amino acids L-4-trifluoromethylphenylalanine (ptfmF), D-4-methylphenylalanine, and L-4-methylphenylalanine (pmF) were purchased from Bachem. Racemic 4-trifluoromethylphenylalanine (ptfmF), 3-fluoro-4-methylphenylalanine (pmmfF), 3-fluoro-4-trifluoromethylphenylalanine (ptfmmfF), 2,3-difluoro-4-methylphenylalanine (pmdfF), and 2,3,5,6-tetrafluoro-4-methylphenylalanine (pmtfF) were synthesized (Supporting Information). Oligonucleotides, DH10B cells, and *pBadA* were purchased from Invitrogen. All primer sequences are provided in Supporting Information Table 1.

Construction of Plasmids. *pDule-ptfmF and pDule-pmmfF* plasmids were generated by amplifying the aminoacyl-tRNA RS gene (aaRS) from the *pBK* plasmid isolated from the library using primers RS-move-F and RS-move-R. The amplified DNA fragments were cloned into a *pDule* plasmid using the incorporated *Nco*I and *Kpn*I sites (*11*). The sfGFP gene developed by Waldo et al. (*32*), codon-optimized for *Escherichia coli* with a C-terminal 6-His affinity tag, was a gift from Dr. Robert P. Hammer. The optimized sfGFP gene was subcloned into *pBadA* generating *pBad-sfGFP*. The *pBad-sfGFP-134TAG*, *pBad-sfGFP-146TAG*, and *pBad-sfGFP-150TAG* plasmids were generated by site-directed mutagenesis using Stratagene's Quikchange mutagenesis protocol on *pBad-sfGFP*. The mutated *pDule-ptfmF* and *pDule-pmmfF* plasmids (*ptfmF A65V*, *ptfmF A65V S158A*, *ptfmF A65V S158G*, *ptfmF H70L*, *ptfmF S158A*, *ptfmF S158G*, *pmmfF V65A*, *pmmfF L70H*, *pmmfF L70H G158A*, and *pmmfF G158A*) were generated by overlap PCR mutagenesis. See Supporting Information Table 1 for a list of primers and genetic sequences.

Selection of ptfmF- and pmmfF-Specific Aminoacyl-tRNA Synthetases. The selections for each UaaRS were performed identically using the same library plasmid and selection plasmids as described previously (*22*). The library plasmid, *pBK-3D-Lib*, was moved sequentially between cells containing a positive selection plasmid (*pCG*) and cells containing a negative selection plasmid (*pNEG*). The library of aaRSs was encoded on the plasmid *pBK-3D-Lib* and focused on the active site amino acids surrounding the *para* position of the tyrosine substrate. In the aaRS library (3D library) (*33*), Leu65, His70, Gln155, and Ile159 were randomized to all 20 natural amino acids, Tyr32 was randomized to 15 natural amino acids (not Trp, Phe, Tyr, Cys, and Ile), Asp158 was restricted to Gly, Ser, or Val, Leu162 was restricted to Lys, Ser, Leu, His, and Glu, and Phe108 and Gln109 were restricted to the following pairs: Trp-Met, Ala-Asp, Ser-Lys, Arg-Glu, Arg-Pro, Ser-His, and Phe-Gln. After a total of three rounds of positive and negative selection, the remaining *pBK-3D-Lib* members were transformed into positive selection cells (*pCG* electrocompetent DH10B *E. coli*) and grown on LB medium plates in the presence of 1 mM L-ptfmF and 2 mM racemic pmmfF, 60−80 μg/mL Cm, 50 μg/mL Km, 25 μg/mL

Tet, and 0.002% arabinose. Individual colonies (10) were selected from the surviving library and screened in the same medium in the presence and absence of unnatural amino acid (UAA) and varying concentrations of Cm from 0 to 120 μg/mL. Individuals (5) showing the least growth on Cm plates in the absence of UAA were selected and sequenced using the primer 5′-CCTAATCTCTTCTGGAGAGTC-3′.

Comparison of UAA Incorporation Using the sfGFP Reporter. All protein expressions were performed in DH10B cells transformed with either *pBad-sfGFP* or *pBad-sfGFP-150TAG* and a *pDule* vector to produce wt-sfGFP or UAA-sfGFP. The arabinose autoinduction medium and method were used (*11*, *34*). Racemic UAAs were dissolved in sterile water and 1 mol equiv of aqueous NaOH and then added to the appropriate medium to a final concentration of 1 mM. Negative control cultures of the mutant proteins, containing no UAA, were grown simultaneously. Cultures at 37 °C were inoculated with 1:100 dilutions of saturated noninducing cultures. Standard expressions were performed in duplicate with 3 mL of medium in 5 × 17 mm culture tubes with shaking at 37 °C for 40 h. Error bars represent the standard deviation obtained from duplicate expressions.

For the UAA concentration-dependent sfGFP overexpressions, the appropriate UAA was dissolved to 3 mM in arabinose autoinduction medium and then serially diluted with medium to obtain concentrations of 0.25, 0.5, 1.0, 2.0, and 3.0 mM UAA. Equivalent temperature, volumes, medium, and vessels were used for these expressions. At the beginning, during the course, and at the end of culturing cells with UAA, no UAA precipitate was visible.

Fluorescence measurements of the cultures were collected 40 h after inoculation using a HORIBA Jobin Yvon FluoroMax-4. The emission from 500 to 520 nm (1 nm bandwidth) was summed with excitation at 488 nm (1 nm bandwidth). Samples were prepared by diluting suspended cells directly from culture 100-fold with phosphate-buffered saline (PBS).

Expression and Purification of sfGFP Containing Unnatural Amino Acids. DH10B *E. coli* cells cotransformed with one of the three *pBad-sfGFP-TAG* vectors and the machinery plasmid *pDule-UaaRS* were used to inoculate 5 mL of noninducing medium containing 100 μg/mL Amp and 25 μg/mL Tet. The noninducing medium culture was grown to saturation with shaking at 37 °C, and 1.0 mL was used to inoculate 100 mL of autoinduction medium with 100 μg/mL Amp, 25 μg/mL Tet, and 1 mM UAA. After 40 h of shaking at 37 °C, cells from 25 mL of culture were collected by centrifugation.

The protein was purified using BD-TALON cobalt ion-exchange chromatography. The cell pellet was resuspended in wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, pH 7) containing 1 mg/mL chicken egg white lysozyme and sonicated 3 × 1 min while cooled on ice. The lysate was clarified by centrifugation, applied to 0.5 mL bed-volume resin, and bound for 20 min. Bound resin was washed with >50 volumes of wash buffer.

Protein was eluted from the bound resin with 5 mL volumes of elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 150 mM imidazole, pH 7) and checked for concentration with a Bradford protein assay. Purity was assessed by SDS-PAGE. For all mass spectrometry and ^{19}F-NMR analyses, the *pDule-ptfmF-A65V-S158A* plasmid was used to express UAA-sfGFP.

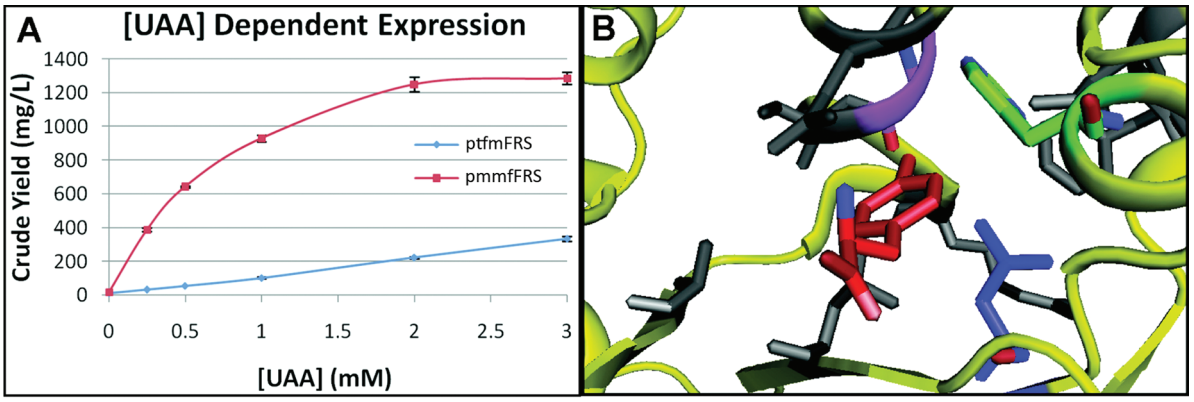
^{19}F-NMR Analysis of UAA-sfGFP. ^{19}F-NMR data were collected using a Varian Unity INOVA 500 MHz spectrometer fitted with a 5 mm broad-band probe. The proton coil was tuned to the fluorine frequency (470.114 MHz). Standard decoupling parameters were used with a 100000 Hz spectral width, 9 μs pulse length (approximately 45°), 2.00 s acquisition time, and 1.00 s relaxation delay. A 10 Hz line broadening was applied, and all spectra were recorded at 25 ± 0.2 °C. All spectra were locked and referenced to a solvent mixture (0.2% solution of 2-fluorophenol in toluene-d_8) in a capillary tube within every ^{19}F-NMR sample. The 2-fluorophenol ^{19}F signal in this internal reference capillary tube was set at −143.687 ppm referenced from trifluorotoluene at −65.000 ppm. The common practice of adding D_2O to the buffer was avoided due to a noticeable deuterium isotope effect on the ^{19}F chemical shift. Purified protein samples were assessed in elution buffer at protein concentrations of 5−15 mg/mL.

Mass Spectrometry Protein Sample Preparation. All protein samples processed for MS analysis were the same protein samples used for ^{19}F-NMR studies. The samples in elution buffer were exchanged into 20 mM ammonium acetate buffer, pH 7, using PD10 gel filtration columns. Proteins in 20 mM ammonium acetate buffer were lyophilized overnight under vacuum. Protein samples for full protein mass spectrometry were resuspended in 1:1 water:acetonitrile with 0.2% formic acid. The samples were analyzed at the Mass Spectrometry Facility at University of Illinois Urbana−Champaign under the direction of Dr. Furong Sun using their ESI-Q-Tof Ultima.

RESULTS AND DISCUSSION

Selection of ptfmF- and pmmfF-Specific Aminoacyl-tRNA Synthetase. Two parallel selections were performed to obtain UaaRSs: one to incorporate ptfmF and the other to incorporate pmmfF. We began with a modified *Methanococcus jannaschii* (*Mj*) tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair, which directs the incorporation of tyrosine in response to the amber (TAG) codon (*1*, *12*). This pair (*Mj*aaRS/tRNA) is orthogonal to the endogenous aminoacyl-tRNA synthetases and tRNAs in *E. coli*. The library of the gene for *Mj*aaRS has randomized codons corresponding to nine amino acid residues (Tyr32, Leu65, His70, Phe108, Gln109, Gln155, Asp158, Ile159, Leu162) that are within 7 Å of the bound tyrosine in the active site (*33*). The same *Mj*aaRS library was used for each UAA selection. The aaRS library was subjected to three rounds of alternating positive and negative selections in the presence and absence of ptfmF or pmmfF (*1*, *22*).

Of the resulting aaRS clones for each UAA, 40−50 were isolated and transformed into cells bearing *pREP2-JYCUA*, which contains *Mj* $tRNA_{CUA}$ and confers amber suppressor-dependent expression of chloramphenicol acetyltransferase. All RS clones showed amino acid-dependent expression of both reporters, surviving at C_m concentrations of 120 μg/mL in the presence of the respective UAA, but only at concentrations of less than 15 μg/mL in the absence of the UAA. Sequencing of 10 distinct clones for each of the selections revealed that each selection fully converged and resulted in only one unique UaaRS solution for each UAA. The selection in the presence of ptfmF resulted in the ptfmFRS (Val32, Ala65, His70, Trp108, Met109, Gln155, Ser158, Met159, Leu162), while the selection in the presence of pmmfF resulted in the pmmfFRS (Leu32, Val65, Leu70, Trp108, Met109, Asn155, Gly158, Phe159, Pro162) (Figure 2C). For further characterization of the incorporation of ptfmF and pmmfF into proteins in response to the amber codon, we cloned each UaaRS into a *pDule* vector that contains one copy of *Mj* $tRNA_{CUA}$ to create *pDule-ptfmF* and *pDule-pmmfF*, respectively.



C

Mj Y parent	Tyr32	Leu65	His70	Phe108	Gln109	Gln155	Asp158	Ile159	Leu162
Parent RS-tfmF	Val	Ala	wt-His	Trp	Met	wt-Gln	Ser	Met	wt-Leu
Parent RS-pmmfF	Leu	Val	Leu	Trp	Met	Asn	Gly	Phe	Pro

FIGURE 2: Characterization of ptfmFRS and pmmfFRS. (A) Cells containing *pDule-UaaRS* and *pBad-sfGFP-150TAG* were grown in autoinduction medium in the presence of increasing concentrations of the respective UAA. (B) Crystal structure of the UaaRS active site. The UAA substrate is in red. Active site residues in the *meta* and *para* positions of the substrate are Leu65 in blue, His70 in green, and Asp158 in purple. The crystal structure image was generated from PDB 1ZH6. (C) Sequence of wild-type *Mj*UaaRS compared to the selected ptfmFRS and selected pmmfFRS.

To test the efficiency and verify the fidelity of the newly selected UaaRSs (ptfmFRS and pmmfFRS) with their respective UAAs, we performed the standard assessment of protein yields in the presence and absence of UAA as well as ESI-Q-Tof mass analysis. We cotransformed each *pDule-UaaRS* with *pBad-sfGFP-150TAG* (super-folder GFP gene containing an amber codon at site 150) and expressed UAA-sfGFP in the presence and absence of the selected UaaRS's UAA in autoinduction medium. In the presence of 1 mM racemic ptfmF, full-length ptfmF-sfGFP could be cobalt affinity purified with a yield of 110 mg/L, while no sfGFP was purified in the absence of UAA (Figure 3A). In the presence of 1 mM racemic pmmfF, full-length pmmfF-sfGFP could be purified with a yield of 330 mg/L (Figure 3A). Native sfGFP expressed and purified under identical conditions yielded 700 mg/L. Lack of sfGFP in the absence of the UAA also confirms the high fidelity of the new UaaRSs. For further confirmation of the incorporation of the UAAs with their respective UaaRSs, the mass of sfGFP was compared to that of the full-length ptfmF-sfGFP and pmmfF-sfGFP by ESI-Q-Tof mass analysis (Figure 4). The native sfGFP has the expected mass of 27829.0 ± 1 Da. Full-length UAA protein, ptfmF-sfGFP and pmmfF-sfGFP, showed masses that would correspond to the incorporation of single ptfmF and pmmfF amino acid at 27894.0 ± 1 Da and 27930.0 ± 1 Da, respectively. Overall, the results of protein expression with affinity purification and SDS−PAGE and MS analysis demonstrate the high fidelity and efficient incorporation of ptfmF and pmmF at genetically programmed sites in sfGFP using their respective *pDule-UaaRS*.

Construction of a Fluorescent Reporter of UAA−Protein Expression. A facile assessment of UaaRS selectivity and promiscuity under UAA−protein expression conditions was important to quickly increase the utility of site-specific incorporation methods. The need to quantify a large, dynamic range of UAA−protein production, to monitor small differences in incorporation ability, and to quantify low-yielding protein expression *in vivo* led us to develop a fluorescent reporter. A reporter was previously developed by Santoro et al., but this used an unoptimized GFP protein and was designed to work under genetic selection conditions, not UAA−protein overexpression

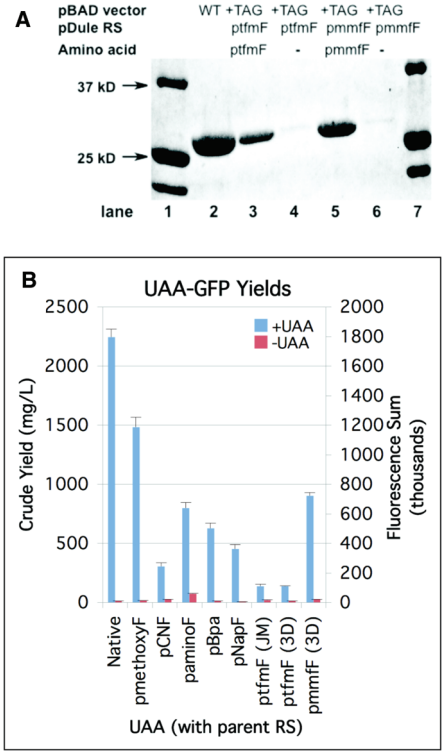


FIGURE 3: Validation of selected synthetases, ptfmFRS and pmmfFRS and UAA-sfGFP reporter system. (A) SDS−PAGE analysis of ptfmF and pmmfF incorporated into sfGFP with their respective UaaRS/tRNA pair. All proteins were purified by metal-immobilized affinity chromatography. Lanes 1 and 7 contain molecular mass markers, while protein production conditions are indicated at the top for lanes 2−6. (B) The first blue bar is the native sfGFP crude yield in autoinduction medium. In comparison, the first red bar represents a TAG codon introduced at site 150 in the sfGFP sequence. Subsequent bars are the crude UAA-sfGFP produced using *E. coli* cells containing *pDule-UaaRS*/*pBad-sfGFP-150TAG*. Blue bars are crude yields employing the indicated UaaRS and its UAA in autoinduction medium, while the red bars for each are expressions in the absence of UAA. The error for each sample was calculated from separate expressions performed in parallel. The synthetase ptfmF(JM)-RS was selected from the JM-RS library (*37*), whereas the ptfmF(3D)-RS was selected from the 3D-RS library (*33*).

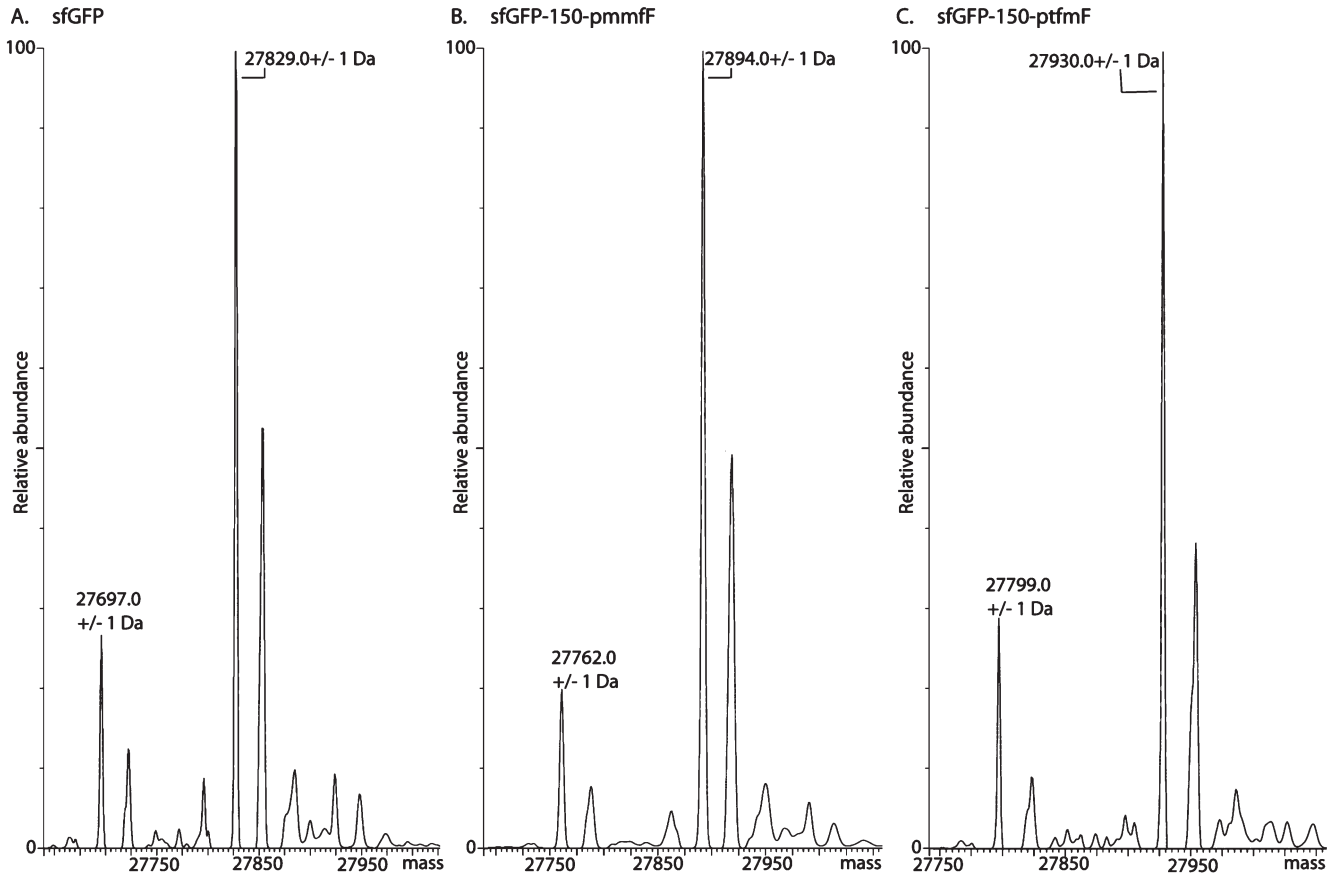


FIGURE 4: ESI-Q-Tof MS of sfGFP and UAA-sfGFP samples demonstrates the efficient high-fidelity incorporation of a single UAA (pmmfF or ptfmF) in response to an amber stop codon, TAG, using the parent pmmfFRS and parent ptfmFRS, respectively. (A) ESI-Q-Tof MS analysis of sfGFP shows a single major peak at 27829.0 ± 1 Da. (B) ESI-Q-Tof MS analysis of sfGFP-150-pmmfF shows a single major peak at 27894.0 ± 1 Da. (C) ESI-Q-Tof MS analysis of sfGFP-150-ptfmF shows a single major peak at 27930.0 ± 1 Da. Each sample did show a +22 sodium adduct. The samples also show a small −131 ± 1 Da peak, indicating minor amounts of peptidase-based removal of N-terminal methionines. No other peaks were observed that would correlate with background incorporation of a natural amino acid.

conditions (*35*). The sfGFP developed by Waldo et al. (*32*) was adapted to report on the activity of a UaaRS by incorporating an amber (TAG) codon at site 150. The position of the codon allows for incorporation of a UAA that would be structurally unimposing on the surface of the β-barrel in the middle of the sequence. Lack of suppression yields a truncated protein of 149 amino acids (resulting in lack of fluorescence), while successful suppression produces full-length sfGFP, a very stable protein even at high concentrations (*32*). Placing the sfGFP gene on the *pBad* expression plasmid allowed for UAA-sfGFP protein overexpression when combined with the plasmid containing the unnatural tRNA/RS pair, *pDule-UaaRS*, in arabinose autoinduction medium (*11*, *34*). The addition of a hexahistidine affinity tag to the C-terminus of sfGFP gene allowed for subsequent purification of expressed sfGFP and UAA-sfGFP for the determination of protein yields. Since the autoinduction medium enables the cells to begin expressing UAA-sfGFP when the cells reach midlog growth without external monitoring or intervention, small volume expressions can be used to reproducibly monitor UaaRS efficacy via fluorescent intensity of cell cultures after 40 h of growth (*11*). As seen in Figure 3B, when native sfGFP is overexpressed in autoinduction medium, the crude protein yield reproducibly reaches 2.2 g of sfGFP/L of medium as based on fluorescence intensity. However, if the amber codon disrupted sfGFP gene is expressed, only truncated protein is produced, resulting in only low background autofluorescence of the cells and buffer (Figure 3). Since *in vivo* and pure sfGFP fluorescence intensity has been shown to be unaffected by cells and medium, we have correlated pure sfGFP fluorescence intensity (Figure 3B, right *y*-axis) with the protein yield and report our values in milligrams of crude protein per liter of medium (left *y*-axis) (*32*). To confirm that this UAA-dependent sfGFP expression would report effectively on UAA incorporation, we cotransformed it with six different literature-characterized *pDule-UaaRS* plasmids (*pDule-methoxyF*, *pDule-pCNF*, *pDule-pAF*, *pDule-pBpa*, *pDule-NapF*, *pDule-ptfmF*(JM)) (*11*, *17*, *22*, *36*). Each of these different *pDule-UaaRS*/*pBad-sfGFP-150TAG* cells were grown in the presence and absence of the appropriate UAA at 1 mM concentration (Figure 3B). Provided autoinduction medium is used and the cultures are allowed to grow for 40 h, normalization of growth rate is not necessary. A range of UAA-sfGFP protein quantities was produced (100−1500 mg/L) in the presence of the UAA, while very low background fluorescence was detected in the absence of the UAA. The UAA-sfGFP yields correlate with expected literature yields; furthermore, the fluorescent intensity of purified UAA-sfGFP was also measured to confirm that the UAA does not affect fluorescence. The high fidelity of all of these *pDule-UaaRS* systems has previously been characterized in the presence of their corresponding UAA (1 mM) under identical or similar expression conditions to the expressions reported here (*11*, *17*, *20*, *22*, *36*, *37*). While there is slight variation in background fluorescence for these *pDule-UaaRS* systems, this UAA-sfGFP reporter correlates well with the published efficiency and fidelity measurements. The very high precision of the

sfGFP/autoinduction assay allows for confident distinctions between UAA incorporation efficiency and small changes in background incorporation. An additional benefit to this reporter method is that the conditions match that of UAA−protein expression methods; therefore, conditional changes that may alter fluorescent intensity correlate in changes with expected UAA−protein yield.

Characterizing Parent Synthetase Selectivity and Efficiency. The sfGFP reporter was used to assess the efficiency of the new *pDule-ptfmF* and *pDule-pmmfF* plasmids. Crude cell culture fluorescence showed that the UaaRS selected to incorporate pmmfF was able to produce seven times more pmmfF-sfGFP than the ptfmFRS was to produce ptfmF-sfGFP. Both showed very low background fluorescence in the absence of the UAA (Figure 3B). The protein yields and low background fluorescence are consistent with the quantitative UaaRS efficiency and fidelity characterization. However, overloading the metal affinity resin prevents the purified yields from matching the *in vivo* yields. This was confirmed by the high amount of fluorescent protein lost during purification of native sfGFP and pmmfF-sfGFP. The selectivity of ptfmFRS and pmmfFRS for their respective UAAs is confirmed by their high yield of UAA-sfGFP in the presence of UAA, and the lack of protein produced in the absence of UAA, as observed by fluorescence and purified protein. An additional goal of these selections was to produce a more efficient UaaRS for the incorporation of ptfmF into proteins for ^{19}F-NMR studies than the previously selected ptfmFRS(JM) from a different library (*37*). Interestingly, Figure 3B shows that selections from both libraries produced high-fidelity UaaRSs with lower than average efficiencies for incorporating ptfmF.

Autoinduction medium and cells containing *pDule-UaaRS* with *pBad-sfGFP-150TAG* were also used to determine the concentration of UAA in the medium that is necessary to most effectively produce UAA−protein with these new UaaRSs (Figure 2A). As expected, UAA-sfGFP yield increased with an increasing UAA concentration, indicating little UAA toxicity. The two different UaaRSs showed surprising differences in their effectiveness for producing UAA proteins. The titration yield of pmmfF-sfGFP with pmmfFRS rose quickly with increasing concentration of pmmfF and leveled off at a 2 mM concentration, whereas the titration yield of the ptfmFRS and ptfmF showed a linear dependence up to 3 mM. While the sfGFP reporter assesses the sum of the processes involved for translating unnatural protein, the close structural similarity of pmmfF and ptfmF makes it unlikely that the difference in UAA dependence is due to differences in solubility, bioavailability, elongation factor Tu (EF-Tu) transport, or structural effects on sfGFP. Although both UaaRSs were selected in the presence of 1 mM L-UAA, their inherent difference in ability with regard to concentration effects future experimental design. Since many of the UAAs currently in use are expensive or challenging to synthesize, understanding and optimizing the UAA concentration to afford maximum yield per mole of UAA is useful. The reproducibility and precision of these expression measurements were high, with standard deviations of $< 5\%$ for each expression yield. To better visualize the error bars, a bar graph analysis is included (Supporting Information Figure 1).

While both UAAs used in these selections had very similar chemical structures, the resulting UaaRSs (ptfmFRS and pmmfFRS) had significantly different sequences and significantly different ability to produce UAA-sfGFP (crude *in vivo* yields of

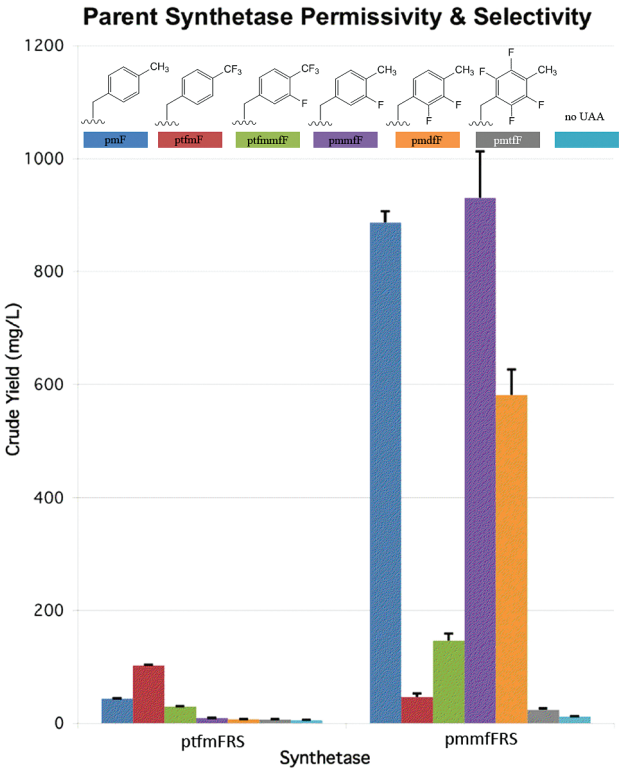


FIGURE 5: Incorporation selectivity and fidelity of parent UaaRSs, ptfmFRS, and pmmfFRS using the different UAAs. Cells containing *pDule-UaaRS* and *pBad-sfGFP-150TAG* were grown in autoinduction medium in the presence of 1 mM racemic UAA.

100 mg/L vs 930 mg/L, respectively) in the presence of 1 mM UAA. Both had high fidelity, not permitting incorporation of natural amino acids, but the pmmfFRS was more efficient at producing UAA-sfGFP at both high and low UAA concentrations. While the UAAs have similar structures, the selected UaaRSs contained seven different active site residues out of a maximum of nine residues (Figure 2C). One explanation for the pmmfFRS's superior efficiency is that it works better with this general chemical shape (*para*-substituted phenylalanine) than does the ptfmFRS.

In order to determine each new UaaRS's level of specificity for its selected substrate, a UAA family with different fluorine substitution patterns was synthesized and tested for incorporation into sfGFP-150TAG. Protein was expressed in cells containing *pBad-sfGFP-150TAG*/*pDule-ptfmF* or *pBad-sfGFP-150TAG*/*pDule-pmmfF* in the presence of 1 mM racemic 4-methylphenylalanine (pmF), 4-trifluoromethylphenylalanine (ptfmF), 3-fluoro-4-trifluoromethylphenylalanine (ptfmmfF), 3-fluoro-4-methylphenylalanine (pmmfF), 2,3-difluoro-4-methylphenylalanine (pmdfF), and 2,3,5,6-tetrafluoro-4-methylphenylalanine (pmtfF), as well as no UAA (Figure 5). While both of the new UaaRSs were able to incorporate most of the family of UAA structures at useful levels, each revealed a distinct preference for its parent structural scaffold. The size of the *para* substituent seems to be a key recognition feature for both UaaRSs. The pmmfFRS was able to incorporate pmF with the same efficiency as pmmfF, but if the larger *p*-trifluoromethyl substituent was used, ptfmF and ptfmmfF incorporation dropped significantly. The ptfmFRS had problems accommodating the smaller *p*-methyl substituent and larger *m*-fluoro substituent. The UAA-sfGFP yield with ptfmFRS was halved with the smaller pmF amino acid and reduced to

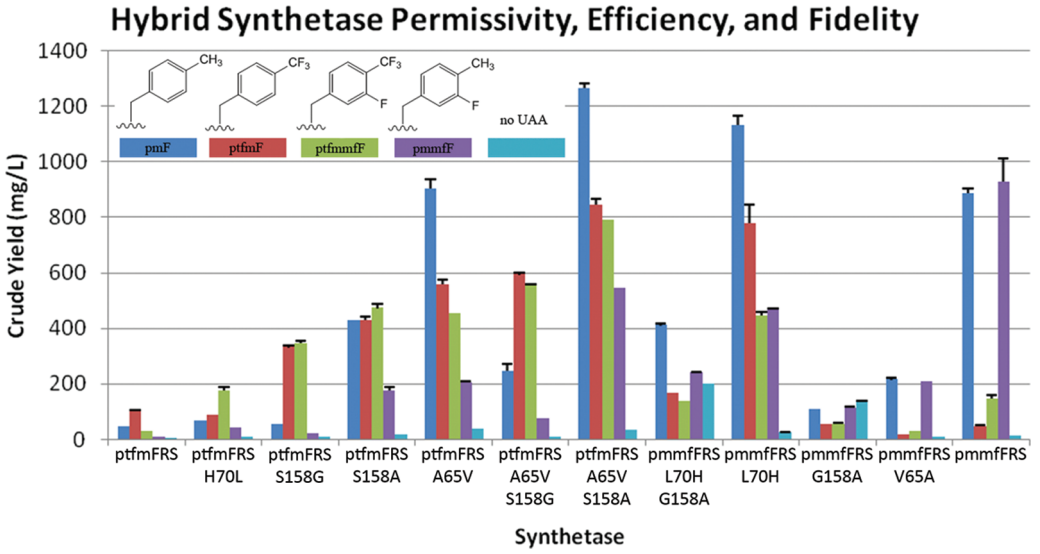


FIGURE 6: Incorporation efficiency, permissivity, and fidelity of hybrid UaaRSs compared to the parent UaaRSs using the different UAAs. Cells containing *pDule-UaaRS* and *pBad-sfGFP-150TAG* were grown in autoinduction medium in the presence of 1 mM racemic UAA.

background with pmmfF. Both UaaRSs were unable to effectively utilize the bulky pmtfF UAA.

Even though each UaaRS shows some low levels of permissivity toward similar chemical structures, the individual sequence solutions for each parent UaaRS (pmmfFRS and ptfmFRS) have shown selectivity for its UAA's general chemical shape. In order to verify that UaaRS permissivity is not restricted to the parent RSs selected in this report, other previously selected UaaRSs were assessed. The UaaRSs that can site-specifically incorporate *p*-benzoylphenylalanine (*18*) and naphthylalanine (*19*) were shown to have high efficiency and permissivity for similar structured UAAs and that the permissivity could be altered by active site mutagenesis (*20*). Based on the selectivity of pmmfFRS and ptfmFRS, a hybrid UaaRS containing elements of both UaaRS active sites should have reduced selectivity and increased permissivity.

Characterizing Permissivity of Sequence Space between Parent Synthetases. The resulting parent UaaRSs for pmmfF and ptfmF have seven active site members that are different from each other. The sequence space in between the two parent UaaRSs is all of the individual amino acid changes that are required to stepwise convert one parent UaaRS active site electrostatic environment into the other. In order to explore the sequence space of the two parent UaaRSs, we choose to make hybrid UaaRSs by exchanging residues in their active sites. We were curious how far into the hybridization process or sequence space from each parent UaaRS was needed to produce a UaaRS that would efficiently incorporate the family of UAAs. On the basis of crystal structures of UaaRSs (*38*, *39*), we limited our seven options for variation to three sites, focusing our mutagenesis on sites in the library that would be most proximal to the *meta* and *para* position on the bound UAA (Figure 2B). Simple exchanges were made between the parent ptfmFRS and pmmfFRS at sites 65, 70, and 158 (Figure 2C). Each of the new offspring UaaRSs was screened for its promiscuity for natural amino acids and permissivity to the family of UAA chemical structures to determine the selectivity and fidelity of these hybrid sequence solutions (Figure 6). An important note is that each of the hybrid UaaRSs' sequences was present in the original selection library; therefore, they were outcompeted by the parent UaaRS during the selection process.

Figure 6 shows each parent UaaRS on opposite ends and the UaaRS with the exchanged residues in the center. General efficiency is visualized by the height of the bar for each UaaRS, while fidelity or promiscuity for natural amino acids is monitored by the height of the last, aqua bar for each UaaRS, where no UAA is provided in the expression. General specificity changes can be derived from the extent to which the parent amino acid is incorporated compared to others for a given UaaRS, while in contrast, general permissivity can be monitored by the heights of all the bars for a given UaaRS. What is immediately apparent is that the sequence space between the parent UaaRSs contains members that show significant increases in efficiency and permissivity, while also maintaining a high level of fidelity.

All changes to the parent ptfmFRS showed an increase in efficiency and permissivity. The most improved single mutant was ptfmFRS-A65V, which is permissive for all members in the UAA family tested showing a 19.5×, 4.5×, 14.5×, and 20.0× increase in yield for pmF, ptfmF, ptfmmfF, and pmmfF, respectively. The background fluorescence in the absence of UAA also increased by 5.5×. This increase in background fluorescence was shown by mass spectral analysis in subsequent double mutant UaaRSs to not compromise fidelity (Figures 4 and 7). To explore the sequence space between serine and glycine at library site 158, alanine was incorporated into both parent UaaRSs. The ptfmFRS-S158A mutant is a more permissive and efficient ptfmFRS for all UAAs tested, indicating that a middle ground between the parent solutions might give rise to a more effective UaaRS. The mutations of the pmmfFRS, however, produced very different results. The pmmfFRS-V65A mutant showed a 4-fold decrease in efficiency for all UAA family members, while maintaining specificity. The pmmfFRS-L70H mutation generated a broadly permissive UaaRS, substantially increasing the yield of *p*-triflouromethyl-containing family members, but reduced the parent amino acid (pmmfF-sfGFP) yield by 2-fold. The pmmfFRS-G158A mutation resulted in a low-fidelity UaaRS and removed any ability to incorporate these UAAs.

The sequence space between the parent UaaRSs was further explored by making double exchanges. Combining a low-fidelity sequence, pmmfFRS-G158A, with a broadly permissive, high-fidelity sequence, pmmfFRS-V65A, resulted in a more efficient,

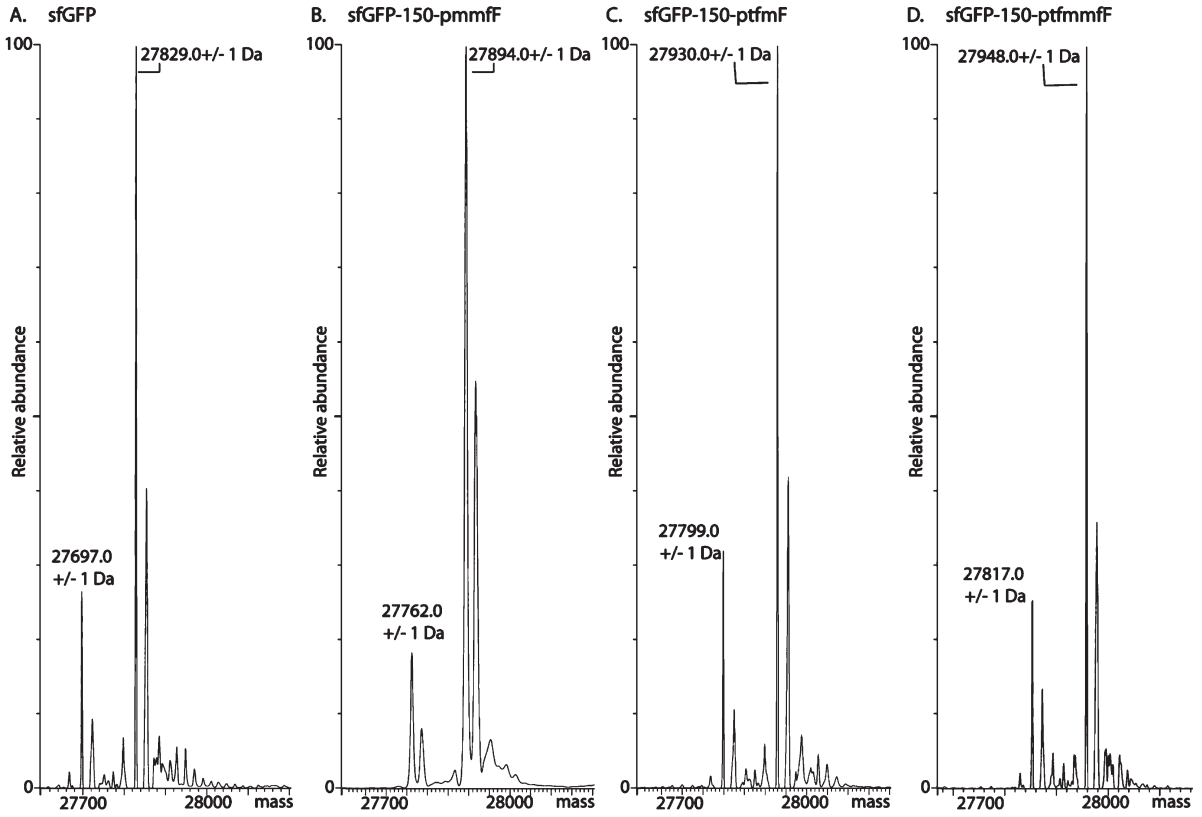


FIGURE 7: ESI-Q-Tof MS of sfGFP and UAA-sfGFP samples demonstrates the efficient high-fidelity incorporation of a single UAA (ptfmF, pmmfF, or ptfmmfF) in response to an amber stop codon, TAG, using the permissive ptfmFRS-A65V-S158A. (A) ESI-Q-Tof MS analysis of sfGFP shows a single major peak at 27829.0 ± 1 Da. (B) ESI-Q-Tof MS analysis of sfGFP-150-pmmfF shows a single major peak at 27894.0 ± 1 Da. (C) ESI-Q-Tof MS analysis of sfGFP-150-ptfmF shows a single major peak at 27930.0 ± 1 Da. (D) ESI-Q-Tof MS analysis of sfGFP-150-ptfmmfF shows a single major peak at 27948.0 ± 1 Da. Each sample did show a +22 sodium adduct. The samples also show a small −131 ± 1 Da peak, indicating minor amounts of peptidase-based removal of N-terminal methionines. No other peaks were observed that would correlate with background incorporation of a natural amino acid.

low-fidelity UaaRS. Combining the broadly permissive ptfmFRS-A65V mutation with ptfmFRS-S158G and ptfmFRS-S158A resulted in ptfmFRS-A65V-S158G and ptfmFRS-A65V-S158A. Compared to the single mutants, the double mutant UaaRSs maintained their broad permissivity, increased their UAA incorporation efficiency by at least 40% for all of the UAA family members, and also improved their fidelity by 10−20%.

Characterization of the Most Permissive Synthetase. Exploring the sequence space between these two parent UaaRS solutions by hybridizing parent UaaRS sequences has shown that hybrid UaaRS active sites can have substantially higher permissivity and efficiency. The utility of broadly permissive UaaRSs for a particular family of UAA probes requires that they maintain high fidelity while retaining the ability to produce large quantities of UAA−protein for study. Some of the hybrid sequences between parents show slight increase in background fluorescence in the absence of UAA, which may indicate a decrease in fidelity in the presence of a given UAA. This potential decrease in fidelity could be the reason that these sequence solutions were not obtained during selections, since they would be removed from the pool in the negative selection step. The most permissive and efficient UaaRS, ptfmFRS-A65V-S158A, displays 3−5 times higher background activity compared to parent UaaRSs. The large dynamic range of fluorescence allows detection of very low levels of sfGFP *in vivo*. The high sensitivity of the sfGFP reporter could be detecting translation of sfGFP in the absence of UAA. This would result in a small increase in background

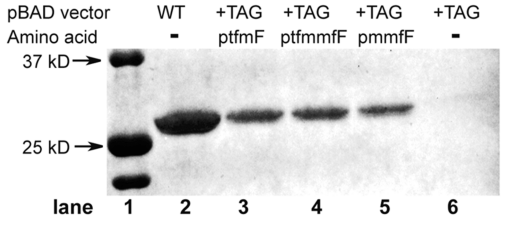


FIGURE 8: Expressed sfGFP, ptfmF-GFP, pmmfF-GFP, and ptfmmfF-GFP using *pDule-ptfmF-A65V-S158A* and *pBad-sfGFP-150* were purified via cobalt affinity chromatography and assessed for purity via SDS−PAGE. Lane 1 contains a molecular mass marker, while protein production conditions are indicated at the top for lanes 2−6. The absence of protein in lane 6 when UAA is withheld from the medium confirms the fidelity of the synthetase.

fluorescence due to low levels of incorporation of natural amino acid when UAA is withheld from the media. When UAA is provided to the medium, it could still be the preferred substrate resulting in translation of entirely UAA-containing protein. To ensure that fidelity was maintained, ptfmFRS-A65V-S158A was used to express UAA-sfGFP in the presence and absence of 1 mM racemic ptfmF, pmmfF, and ptfmmfF. Full-length ptfmF-sfGFP, pmmfF-sfGFP, and ptfmmfF-sfGFP were purified via cobalt affinity chromatography, and their masses were compared using ESI-Q-Tof mass spectroscopy (Figure 7). As expected, the differences in mass between native sfGFP and ptfmF-sfGFP, pmmfF-sfGFP, and ptfmmfF-sfGFP are 65 ± 1, 101 ± 1, and 119 ± 1 Da, respectively. Each UAA-sfGFP spectrum matches the fidelity represented in the sfGFP spectra (Figure 4A), in

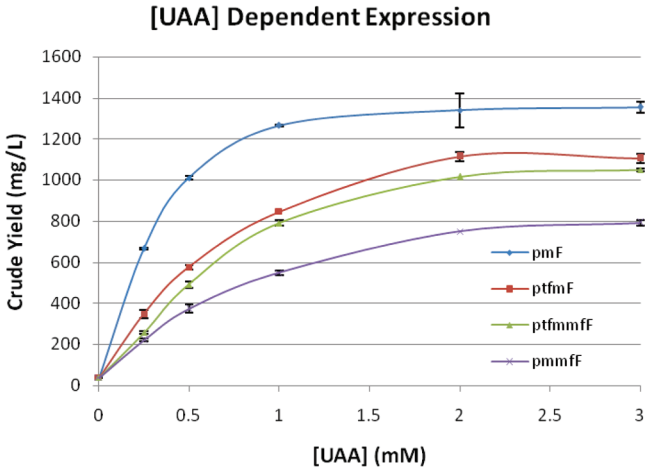


FIGURE 9: Characterization of UAA incorporation by ptfmFRS-A65V-S158A. Cells containing *pDule-ptfmF-A65V-S158A* and *pBad-sfGFP-150TAG* were grown in autoinduction medium in the presence of increasing concentrations of the respective UAA.

agreement with a single addition of the appropriate UAA. No other peaks were observed that could be correlated with background incorporation of a natural amino acid, and the peaks match the characterization of high-fidelity UaaRSs (*22*). Each sample did show small amounts of a 131 ± 1 Da loss from the parent peak, indicating minor amounts of terminal methionine removal. Overall, the results of protein expression and affinity purification followed by SDS−PAGE (Figure 8) and MS analysis demonstrate the high fidelity of the more permissive ptfmFRS-A65V-S158A.

To verify the general utility and scalability of UAA−protein expression with the permissive UaaRS, ^{19}F-labeled sfGFP for ^{19}F-NMR analysis was produced using ptfmFRS-A65V-S158A. The UAAs, ptfmF, pmmfF, and ptfmmfF, were each incorporated at three different locations in sfGFP: exterior site 150 on the surface of the β-barrel, exterior site 134 on a flexible loop at the edge of the barrel, and interior site 146 near the fluorophore (Supporting Information Figure 2A). The completely solvent-exposed 134 site showed sharp single peaks for all *p*-triflouromethyl and *m*-fluoro substituents (Supporting Information Figure 2B). The external edge site 150, which has some interaction with other amino acids on the barrel surface, showed slightly broadened peaks for the *p*-triflouromethyl and *m*-fluoro substituents, indicating multiple closely related environments (Supporting Information Figure 2C) (*30*, *37*, *40*, *41*). The internal site 146 showed greater diversity of signal effects using the different UAAs (Supporting Information Figure 2D). The pmmfF at site 146 showed a sharp single peak indicating a single defined environment; however, ptfmF at that same location showed a very broad and undefined peak, indicating that the *p*-trifluoromethyl substituent is slowly exchanging between multiple environments on the time scale of the measurements. The ptfmmfF showed broadened but defined signals for both the *p*- and *m*-fluorine substituents, indicating that the amino acid is sampling slightly different environments. It is noteworthy that a family of structural probes can be useful in providing different structural information about a particular site, as evidenced by the difference in spectral response of the ^{19}F family of probes between surface and interior incorporation with sfGFP. Thus, permissive UaaRSs that incorporate families of UAA probes have the potential to expedite the structural characterization of proteins.

To further characterize the efficient and broadly permissive ptfmFRS-A65V-S158A, UAA concentration-dependent expressions were performed with the family of UAAs (Figure 9). UAA titration curves of UAA-sfGFP yields clearly show that a more efficient incorporation of an UAA with ptfmFRSA65V-S158A leads to a sharper titration curve and a better production of UAA-sfGFP at low concentrations of UAA in the medium. In general, amino acid solubility and uptake does vary widely with amino acid structure. It should be cautioned that the *in vivo* sfGFP-reporter system here is a measure of all the steps in translating UAA−protein including any differences there might be in UAA uptake, etc. In order to confirm that the translational system containing the more efficient UaaRSs can efficiently produce protein at low concentrations of UAA in solution, titration curves were measured on a single UAA, ptfmF, for each of the hybrid UaaRSs (Supporting Information Figure 3).

While sequence information from multiple parent UaaRSs can be used to obtain UaaRSs that show high permissivity, high efficiency, and good fidelity, the best UaaRSs engineered here are still less than half as efficient at producing UAA-sfGFP *in vivo* compared to wild-type sfGFP production. This inefficiency may result from a need to improve or balance a variety of factors in UAA−protein translation. The need to optimize the UaaRS copy number, along with optimizing the interactions of unnatural tRNAs, aaRSs, release factors, and EF-Tu with the ribosome, could explain some of the reduced efficiency of the UAA-sfGFP production (*42*−*44*). This work shows that the UaaRSs obtained from standard selections may not have the appropriate balance of efficiency and fidelity for maximum protein production. Therefore, it is evident that selection methods have yet to be optimized for obtaining UaaRSs with ideal ratios of translational efficiency and fidelity for maximum protein expression or that other diversity-reducing processes are operating. The standard selection method commonly used to screen large libraries employs a stringent life/death (positive/negative) double-sieve approach. It appears from our titration data that the positive selection could benefit from lowering the UAA concentration from the standard 1 mM during positive selections. The negative selection is likely too stringent, as the fidelity mandated by the negative selection is not necessary for the expression conditions employed here. By tempering the negative selection and reducing the concentration of UAA used in positive selections, a more diverse population of effective UaaRSs should be obtained.

CONCLUSIONS

Traditional selections were initially employed to produce efficient, high-fidelity UaaRSs for the site-specific incorporation of the fluorinated phenylalanine derivatives pmmfF and ptfmF. Using our new sfGFP-reporter system, we assessed the permissivity of the parent UaaRSs for a family of ^{19}F-UAAs (Figure 1). We demonstrated that the parent UaaRSs showed modest permissivity, since they were relatively selective for their parent UAA structure. In an attempt to make UaaRSs with increased permissivity, the sequence space between these two selective and efficient UaaRSs was explored by swapping amino acids from each of their active sites with one another, making hybrid UaaRSs. We showed that many of the hybrid UaaRSs were more efficient and more permissive, while still maintaining fidelity against natural amino acids. There seems to be a range of effective sequence solutions between parent UaaRSs, making it possible to hybridize two or more different UaaRS sequences to

develop permissive UaaRSs. While this method of generating UaaRSs for novel UAAs will certainly not be applicable to all new UAAs, it does show that the *Mj*RS is unusually malleable to active site mutations, and conservative mutations of selected UaaRSs can increase their permissivity. We hope that UaaRS hybridization, when relevant, is more accessible than traditional methods and will encourage new studies with UAA technology.

In the course of our studies on permissivity, we observed that high-quality UaaRSs are eliminated in the standard selection conditions used to generate UaaRSs. Monitoring the UAA concentration-dependent expression of protein and UaaRS fidelity with the UAA-sfGFP reporter indicates that the UaaRSs resulting from selections do not have the efficiency to fidelity ratio that is optimal for maximum protein production. We suspect that more diverse populations of more effective UaaRSs would result if UaaRS selections employed positive selections with lower UAA concentrations and more moderate negative selections. Since the traditional selection process does not impose selective pressure for or against UaaRS permissivity, a more diverse resulting pool of effective UaaRSs would likely translate into a larger number of permissive UaaRSs. By using this sfGFP reporter in autoinduction medium, a large pool of UaaRSs could be screened for efficiency, permissivity, and fidelity with high precision.

Unnatural aminoacyl-tRNA synthetases that show permissivity for families of UAA probes have tremendous utility since they allow researchers to express a greater diversity of chemically modified proteins using a single expression system and, therefore, will continue to aid in protein characterization and modification (*14*−*16*, *20*, *22*, *45*). The hybrid UaaRSs we created in this work significantly improved permissivity for a family of ^{19}F-UAAs, and the benefits of using multiple probes in ^{19}F-NMR experiments have been demonstrated. As a proof of principle we studied the incorporation of three different ^{19}F-UAAs into three different sites on sfGFP by ^{19}F-NMR and observed distinct spectral differences depending on the site and the ^{19}F-UAA. If UaaRSs are characterized as permissive or generated to be permissive, the development of additional UAA probes will be greatly accelerated. Further studies are underway to modify the selection process to facilitate the identification of larger pools of efficient and permissive UaaRSs. The hybrids of isolated UaaRSs are also being screened to aid in the development of families of permissive UaaRSs for larger UAAs and toxic UAAs.

ACKNOWLEDGMENT

We thank Beth Buckwalter and Carol Strausser for technical assistance. We gratefully acknowledge the help of Lisa Mertzman.

SUPPORTING INFORMATION AVAILABLE

A description of the synthesis of UAAs, ^{19}F-NMR of proteins, and primers used. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Xie, J., and Schultz, P. G. (2006) A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol. 7*, 775–782.
2. Beatty, K. E. T., D. A. (2009) Noncanonical amino acids in protein science and engineering. *Nucleic Acids Mol. Biol. 22*, 127–153.
3. Cujec, T. P., Roberto, M., Hays Putnam, A.-M. A., Keefe, W. M., Knudsen, N., Ho, L., Pinkstaff, J., and Kraynov, V. (2008) in PCT International Application, p 318, Ambrx, Inc., USA.
4. Seyedsayamdost, M. R., Xie, J., Chan, C. T. Y., Schultz, P. G., and Stubbe, J. (2007) Site-specific insertion of 3-aminotyrosine into subunit alpha 2 of *E. coli* ribonucleotide reductase: Direct evidence for involvement of Y-730 and Y-731 in radical propagation. *J. Am. Chem. Soc. 129*, 15060–15071.
5. Kimmel, B. E., Sim, B., and Daniel, T. O. (2007) in PCT International Application, p 277, Ambrx, Inc., USA.
6. Chin, J. W. (2006) Modular approaches to expanding the functions of living matter. *Nat. Chem. Biol. 2*, 304–311.
7. Carrico, Z. M., Romanini, D. W., Mehl, R. A., and Francis, M. B. (2008) Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.,* 1205–1207.
8. Chen, S., Schultz, P. G., and Brock, A. (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J. Mol. Biol. 371*, 112–122.
9. Chen, P. R., Groff, D., Guo, J. T., Ou, W. J., Cellitti, S., Geierstanger, B. H., and Schultz, P. G. (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem., Int. Ed. 48*, 4052–4055.
10. Melancon, C. E., and Schultz, P. G. (2009) One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorg. Med. Chem. Lett. 19*, 3845–3847.
11. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., and Mehl, R. A. (2007) Preparation of site-specifically labeled fluorinated proteins for ^{19}F-NMR structural characterization. *Nat. Protoc. 2*, 2601–2607.
12. Xie, J., and Schultz, P. G. (2005) An expanding genetic code. *Methods 36*, 227–238.
13. Fersht, A. R. (1998) Research: Protein structure—Sieves in sequence. *Science 280*, 541–541.
14. Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc. 131*, 8720–8722.
15. Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N-epsilon-acetyllysine in recombinant proteins. *Nat. Chem. Biol. 4*, 232–234.
16. Brustad, E., Bushey, M. L., Brock, A., Chittuluru, J., and Schultz, P. G. (2008) A promiscuous aminoacyl-tRNA synthetase that incorporates cysteine, methionine, and alanine homologs into proteins. *Bioorg. Med. Chem. Lett. 18*, 6004–6006.
17. Jackson, J. C., Duffy, S. P., Hess, K. R., and Mehl, R. A. (2006) Improving nature's enzyme acitive sites with genetically encoded amino acids. *J. Am. Chem. Soc. 128*, 11124–11127.
18. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11020–11024.
19. Wang, L., Brock, A., and Schultz, P. G. (2002) Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc. 124*, 1836–1838.
20. Stokes, A. L., Miyake-Stoner, S. J., Peeler, J. C., Nguyen, D. P., Hammer, R. P., and Mehl, R. A. (2009) Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst. 5*, 1032–1038.
21. Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X. S., Zhang, Q., Hu, H. Y., Brittain, S. M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G., and Geierstanger, B. H. (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc. 130*, 9268–9281.
22. Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry 48*, 5953–5962.
23. Liu, C. C., Mack, A. V., Brustad, E. M., Mills, J. H., Groff, D., Smider, V. V., and Schultz, P. G. (2009) Evolution of proteins with genetically encoded "chemical warheads". *J. Am. Chem. Soc. 131*, 9616–9617.
24. Lee, H. S., Spraggon, G., Schultz, P. G., and Wang, F. (2009) Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. *J. Am. Chem. Soc. 131*, 2481–2483.
25. Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A., and Chin, J. W. (2008) Genetically encoding protein oxidative damage. *J. Am. Chem. Soc. 130*, 4028–4033.

26. Li, C. G., Lutz, E. A., Slade, K. M., Ruf, R. A. S., Wang, G. F., and Pielak, G. J. (2009) F-19 NMR studies of alpha-synuclein conformation and fibrillation. *Biochemistry 48*, 8578–8584.
27. Ulrich, A. S. (2003) Solid state ^{19}F NMR methods for studying biomembranes. *Prog. Nucl. Magn. Reson. Spectrosc. 46*, 1–21.
28. Seyedsayamdost, M. R., Yee, C. S., Reece, S. Y., Nocera, D. G., and Stubbe, J. (2006) pH rate profiles of FnY356-R2s (n = 2, 3, 4) in *Escherichia coli* ribonucleotide reductase: Evidence that Y-356 is a redox-active amino acid along the radical propagation pathway. *J. Am. Chem. Soc. 128*, 1562–1568.
29. Leone, M., Rodriguez-Mias, R. A., and Pellecchia, M. (2003) Selective incorporation of ^{19}F-labeled Trp side chains for NMR-spectroscopy-based ligand-protein interaction studies. *ChemBioChem 4*, 649–650.
30. Bann, J. G., and Frieden, C. (2004) Folding and domain-domain interactions of the chaperone PapD measured by ^{19}F NMR. *Biochemistry 43*, 13775–13786.
31. Li, C., Wang, G., Wang, Y., Creager-Allen, R., Lutz, E. A., Scronce, H., Slade, K. M., Ruf, R. A. S., Mehl, R. A., and Pielak, G. J. (2009) Protein ^{19}F NMR in *Escherichia coli*. *J. Am. Chem. Soc.* (ASAP article web Dec 8, 2009).
32. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24*, 79–88.
33. Xie, J. M., Liu, W. S., and Schultz, P. G. (2007) A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem., Int. Ed. 46*, 9239–9242.
34. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif. 41*, 207–234.
35. Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol. 20*, 1044–1048.
36. Farrell, I. S., Toroney, R., Hazen, J. L., Mehl, R. A., and Chin, J. W. (2005) Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat. Methods 2*, 377–384.
37. Jackson, J. C., Hammill, J. T., and Mehl, R. A. (2007) Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc. 129*, 1160–1166.
38. Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2005) Structural characterization of a *p*-acetylphenylalanyl aminoacyl-tRNA synthetase. *J. Am. Chem. Soc. 127*, 14976–14977.
39. Turner, J. M., Graziano, J., Spraggon, G., and Schultz, P. G. (2006) Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc. Natl. Acad. Sci. U.S.A. 103*, 6483–6488.
40. Li, H., and Frieden, C. (2005) NMR studies of 4-^{19}F-phenylalanine-labeled intestinal fatty acid binding protein: Evidence for conformational heterogeneity in the native state. *Biochemistry 44*, 2369–2377.
41. Glaser, R. W., Sachse, C., Durr, U. H., Wadhwani, P., and Ulrich, A. S. (2004) Orientation of the antimicrobial peptide PGLa in lipid membranes determined from ^{19}F-NMR dipolar couplings of 4-CF_3-phenylglycine labels. *J. Magn. Reson. 168*, 153–163.
42. Wang, K. H., Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol. 25*, 770–777.
43. Ryu, Y., and Schultz, P. G. (2006) Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods 3*, 263–275.
44. An, W. L., and Chin, J. W. (2009) Synthesis of orthogonal transcription-translation networks. *Proc. Natl. Acad. Sci. U.S.A. 106*, 8477–8482.
45. Hartman, M. C. T., Josephson, K., and Szostak, J. W. (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. U.S.A. 103*, 4356–4361.