Chapter 6

Breaking the Degeneracy of the Leucine Codons in Escherichia coli

Abstract

Previously we have shown that combination of yPheRS (T415G) and the mutant ytRNA^{Phe}_{AAA} enables incorporation of 2Nal in response to UUU (Phe) codons. However, this method has two limitations. First, the yPheRS (T415G) activates Trp, which leads to Trp misincorporation. Second, 2Nal was incorporated at UUC codons as well as UUU codons, due to the relaxed codon recognition of the AAA anticodon in ytRNA^{Phe}_{AAA}. In order to enhance the selectivity of yPheRS, we performed high-throughput screening of a yPheRS library. The screen revealed a yPheRS variant (yPheRS_naph) with 6-fold higher relative activity toward 2Nal (vs. Trp) in ATP-PPi exchange assays. To enhance codon selectivity, we explored degenerate leucine codons instead of phenylalanine degenerate codons. Combined use of ytRNA^{Phe} containing the CAA anticodon, which recognizes only the UUG (Leu) codon, and yPheRS_naph allowed incorporation of 2Nal into murine dihydrofolate reductase in response to six UUG codons with 50% fidelity, but not to other Leu codon sites.

Introduction

An expanded set of genetically encoded amino acids has allowed design of proteins with novel chemical, physical, or biological properties. Several methods have been developed to introduce nonnatural amino acids into recombinant proteins at programmed sites *in vivo*. First, residue-specific incorporation involves the global replacement of a particular natural amino acid with a nonnatural analog in target proteins.¹⁻¹² The strength of this technique lies in efficient protein translation and multi-site incorporation of a nonnatural amino acid, because sense codons are re-assigned for a

nonnatural amino acid. However, its application could be restricted due to the exclusion of one natural amino acid for protein translation and lack of site-specificity. Second, single-site-specific incorporation involves utility of a heterologous orthogonal aminoacyltRNA synthetase/tRNA pair and suppression of either a stop codon or a frameshift codon.¹³⁻²⁰ Access to all 20 natural amino acids and site-specificity of the technique allows the incorporation of a nonnatural amino acid at a single site in a target protein with minimal perturbation of the native protein structure and function. However, application of the method is limited to the insertion of a nonnatural amino acid at only one or two positions in a target protein, due to moderate suppression efficiency.

Although residue-specific incorporation and single-site-specific incorporation have been used for many different applications,²¹⁻²⁷ one can envision other circumstances in which a nonnatural amino acid needs to be inserted at multiple positions in the presence of all twenty natural amino acids. For example, by introducing a nonnatural amino acid only at permissive sites in a target protein, the protein may be equipped with multiple reaction sites or probes with minimal loss of its native properties. Multiple-site-specific incorporation *in vivo* was realized in our lab by breaking the degeneracy of the Phe codons.²⁸ The combined use of a mutant yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) and mutant yeast phenylalanyl-tRNA containing modified AAA anticodon (ytRNA^{Phe}_{AAA}) that favors UUU Phe codons over UUC Phe codons, achieved efficient incorporation of 2Nal into mDHFR at five Phe sites encoded as UUU. Although the method successfully demonstrated the concept of breaking the degeneracy of the genetic code, its application has been restricted by two limitations.

First, the rationally designed yPheRS (T415G) efficiently activated a natural

148

amino acid, Trp, due to the enlarged binding pocket generated by the T415G mutation.¹³ The relaxed substrate specificity of yPheRS (T415G) showed several drawbacks. Cell growth rate decreased, likely due to misincorporation of Trp at Phe sites in essential proteins. Leaky expression of target protein even under uninduced conditions was prominent, perhaps due to the impaired repressor proteins. Furthermore, misincorporation of Trp as well as 2Nal was observed at Phe residues encoded as UUU codons, which prevented high fidelity incorporation of 2Nal at programmed sites. These drawbacks prompted us to pursue more selective yPheRS variants. Previously we reported that an other rationally designed yPheRS (T415A) showed 10-fold higher activity toward pBrF than Trp. However, vPheRS (T415A) did not exhibit enhanced selectivity toward pBrF against Phe.¹³ Therefore, we need to explore different approaches to obtain highly selective yPheRS variants. Aminoacyl-tRNA synthetases are known to be readily evolvable. Schultz and co-workers have developed powerful screening methods to change the substrate specificity of tyrosyl-tRNA synthetase (mjTyrRS) derived from Methanococcus jannaschii toward nonnatural amino acids.^{14,18,22,29} Recently our lab reported a novel screening method to adapt E. coli methionyl-tRNA synthetase to a reactive methionine analog, azidonorleucine.⁷ However, until now there have been no reports about evolving eukaryotic aminoacyl-tRNA synthetases to change their substrate specificity toward a nonnatural amino acid. In this report, we describe the highthroughput screening of a yPheRS library to obtain yPheRS variants of which substrate specificity is changed to a nonnatural amino acid, 2Nal.

Second, 2Nal was misincorporated at the unwanted sites (UUC codon) as well as programmed sites (UUU codon) due to relaxed codon recognition of the AAA anticodon

of ytRNA^{Phe}_{AAA}. Misincorporation of a nonnatural amino acid at unwanted sites might cause perturbation or loss of native properties of the target protein.^{23,30} According to Crick's wobble rules proposed in 1966,³¹ the base A in the first position of the anticodon recognizes only the base U in the third position of the codon. Therefore, the AAA anticodon was expected to recognize only UUU, not UUC codons. The discrepancy between the experimental result and Crick's wobble rule may be explained by the expanded wobble rule proposed by Lim and Curran in 2001.³² The expanded wobble rule is based on new experimental finding³²⁻³⁹ and stereochemical modeling⁴⁰⁻⁴³ of codon-anticodon interactions. The expanded wobble rule indicates that A in the first position of the anticodon of the anticodon can recognize all four bases in the third position of the codon, though its affinity to C is lower than that to U.

Due to the poor discrimination of UUU codon froms UUC by the AAA anticodon of ytRNA^{Phe}, we explored breaking the degeneracy of leucine (Leu) codons. Several considerations recommend degenerate Leu codons. First, Leu is encoded as six codons: UUA, UUG, CUA, CUG, CUU, and CUC. Discrimination of UUG from CUN (N = A/U/G/C) codons should be highly efficient due to discrimination at the first position in codon. Second, our existing yeast orthogonal pair should be readily adapted to the incorporation of Phe analogs in response to UUG codons. In practical terms, generalization of the concept of breaking the degeneracy of the genetic code is limited by the availability of orthogonal pairs. Third, the modified CAA anticodon would more efficiently discriminate UUG from UUA. According to the expanded wobble rules, C in the first position of the anticodon recognizes only G in the third position of the codon. In this report, we show that an engineered ytRNA^{Phe} containing a modified CAA anticodon (CAA) can completely discriminate UUG from the remaining five Leu codons, and achieve incorporation of 2Nal at multiple programmed sites in recombinant proteins.

Materials and Method

Materials. Unless otherwise noted, reagents were obtained from commercial suppliers and were used without further purification. 2Nal was purchased from Chem-Impex (Wood Dale, IL). Restriction enzymes were obtained from New England Biolabs (Beverly, MA). Quikchange mutagenesis kits were purchased from Stratagene (La Jolla, CA). Nickel-nitrilotriacetic acid affinity columns and plasmid pREP4 were obtained from Qiagen (Valencia, CA). DNA primers were obtained from Integrated DNA Technologies (Coralville, IA) and Operon Technologies (Huntsville, AL). Sequences of the DNA primers used in this research are listed in Appendix C.

Preparation of *E. coli* **hosts**. Preparation of the Phe/Trp double auxotrophic strain (AFW) and the Phe/Trp/Lys triple auxotrophic strain (AFWK) was described previously.¹³ A Phe/Leu double auxotrophic strain, MPC390 (*leuB6(Am)*, *PheA18::Tn10*), was obtained from the *E. coli* Genetic Stock Center (CGSC) at Yale University. A Phe auxotrophic derivative of DH10B (Stratagene) *E. coli* strain was prepared by chemical mutagenesis⁴⁴ and designated DHF.

Plasmid Construction for Reporter Gene Expression. The GFP_{UV} gene was amplified from pGFPuv (Clontech, Palo Alto, CA) using two primers (GFP1 and GFP2) containing a *Pst*I restriction site. The amplified GFPuv gene was inserted into pQE9

(Qiagen) at the PstI site to generate pQE9_GFP_{UV}. Since GFP_{UV} has excitation maxima at 475 and 395 nm,⁴⁵ it was mutated into an EGFP variant (GFP3)⁴⁵ with excitation maximum at 488 nm suitable for FACS analysis. A series of PCR mutagenesis reactions were performed at four residues (F64L, S65T, S99F and T153M) using three pairs of complementary primers (F64LS65T f/F64LS65T r; S99F f/S99F r; T153M_f/T153M_r). The GFP3 gene expression cassette was inserted into pQE16_mDHFR_yPheRS (T415G)²⁸ between the AatII and NheI restriction sites to generate pQE9_GFP3_yPheRS (T415G). GFP3 has twelve Phe residues, of which seven are encoded as UUC. A GFP variant (GFP6) was prepared by replacing all UUC Phe codons and one CUG Leu codon (at position 64) with UUU codons via gene fragment reassembly using sixteen primers (Primer 1-16). The GFP6 expression cassette was removed by restriction digestion with the AatII and NheI, and inserted into pQE9_GFP3_yPheRS (T415G) between the AatII and NheI sites to yield pQE9 GFP6 yPheRS (T415G). In order to introduce an extra copy of the *lacI* gene into pQE9_GFP6_yPheRS (T415G), the lacI gene expression cassette was amplified from pREP4 (Qiagen) with two primers (*lac*I_Af/II_f and *lac*I_SacII_r). Af/II and SacII restriction sites were introduced into pQE9_GFP6_yPheRS (T415G) by PCR reactions with two primers (pQE_AflII_f and pQE_SacII_r). The amplified lacI gene expression cassette was inserted into pQE9_GFP6_yPheRS (T415G) between the AflII and SacII sites to generate pQE9_GFP6_lacI_yPheRS (T415G). A GFP variant (GFP3_WC) containing 12 Phe residues encoded as only UUC codons were generated by the gene fragment reassembly method⁴⁶ using twelve primers (Primer 1, 16, 17-26). The GFP3_WC gene expression cassette was ligated into pQE9_GFP6_lacI_yPheRS (T415)

between the *Aat*II and *Nhe*I sites to yield pQE9_GFP3_WC_lacI_yPheRS (T415G). Information of the marker proteins used in this work is summarized in Appendix B.

Construction of vPheRS Library. Four residues (N412, T415, S418 and S437) in the active site of yPheRS were saturated by two step PCR mutagenesis. First, NNK (N = A/U/G/C; K = G/U) degenerate codons were introduced into the 437th position in the α subunit of yPheRS by PCR mutagenesis with two complementary primers (437 f and 437_r) using pQE9_GFP6_lacI_yPheRS (T415G) as a template. The PCR product was digested by *Dpn*I for 1 hr at 37 °C and transformed into XL-1 blue (Stratagene) competent cells. The plasmids isolated from the culture of transformants were used as a template for the second PCR mutagenesis with one pair of complementary primers (412 415 418 f and 412 415 418 r) to saturate the 412th, 415th and 418th positions in the α -subunit of yPheRS. The PCR product was digested by *Dpn*I for 1 hr at 37 °C and desalted on a spin column. The eluted plasmids were transformed into ElectroTen-Blue electro-competent cells (Stratagene) according to the manufacturer's protocol, and plasmids were isolated from the culture of pooled transformants. The plasmid fragments encoding yPheRS were removed by digestion with NsiI and BglII restriction enzymes and ligated to the large fragments of pQE9_GFP6_lacI_yPheRS (T415G) obtained by digestion with NsiI and BglII. The ligated plasmids were desalted and transformed into ElectroTen-Blue electro-competent cells to generate six million transformants. The yPheRS library plasmids (P_yFS_20) were isolated from the culture of pooled transformants.

Construction of Expression Library. pREP4_ytRNA_AAA, prepared previously,²⁸ was transformed into DHF electro-competent cells to generate DHF_AAA cells. The P_yFS_20 library was transformed into DHF_AAA electro-competent cells to generate the yPheRS expression library (L_01) consisting of ten million transformants. The transformants were transferred to 0.5 L 2xYT media with 200 μ g/mL ampicillin and 35 μ g/mL kanamycin. When OD₆₀₀ reached 1.0, glycerol stocks were prepared and kept in a -80 °C freezer.

Screening of yPheRS Library. A half mL of glycerol stock of the expression library was inoculated into 100 mL of M9 minimal medium supplemented with glucose, thiamin, MgSO₄, CaCl₂, 20 amino acids (at 25 mg/L), antibiotics (35 µg/mL of kanamycin and 200 μ g/mL of ampicillin). When the culture reached an OD₆₀₀ of 0.6–0.8, cells were spun down, washed twice with ice-cold 0.9% NaCl, and shifted to expression medium supplemented with 18 amino acids (at 25 mg/L), and the indicated concentrations of Phe, Trp and 2Nal. Expression of GFP was induced by addition of 1 mM IPTG. After 3 hrs, 1 mL of the culture (based on OD of 1.0) was collected and washed twice with 0.5 mL of PBS (pH 7.4). 300 μ L of cells were diluted with 3 mL of distilled water, and then subjected to cell sorting using a MoFlo cell sorter (DakoCytomation, Ft. Collins, CO). The excitation and emission wavelengths were 488 and 525 nm, respectively. The library was subjected to both positive and negative screening (Figure 3). FACS gates were set based upon FSC/SSC and FL. In order to enrich active yPheRS variants, 5x10⁴ weakly fluorescent cells were collected by positive screening in the first round. The screened cells were incubated in 2xYT medium containing 200 µg/mL of ampicillin and 35 µg/mL of kanamycin. When OD_{600} reached 1.0, glycerol stocks of the cells were prepared. 0.1 mL of the glycerol stocks was inoculated into 20 mL of M9 minimal medium containing 20 amino acids. In order to enrich selective yPheRS variants, the top 1% of highly-fluorescent cells were collected. One more negative screening was applied to enrich the population in bright cells. The positive and negative screening steps were repeated. Finally, one more positive screening was performed, and the collected cells were spread onto a 2xYT agar plate containing 35 µg/mL of kanamycin and 200 µg/mL of ampicillin. After overnight incubation at 37 °C, ten single colonies were isolated and subjected to characterization.

Characterization of the Isolated Clones. The ten single colonies were transferred to 0.5 L 2xYT media with 200 µg/mL ampicillin and 35 µg/mL kanamycin. When OD₆₀₀ reached 1.0, glycerol stocks were prepared and kept in a -80 °C freezer. 200 µL of each glycerol stock was inoculated into minimal medium and incubated until OD₆₀₀ reached 0.6. The cultures were washed twice with 0.9% NaCl solution, and resuspended with 20 mL of minimal medium supplemented with 18 amino acids, 50 µM Trp, and 5 µM Phe. The resuspended cells were divided into two fractions, and transferred into two flasks. 3 mM 2Nal was supplemented into one of them. GFP6 was expressed by addition of 1 mM IPTG. After 3 hrs, 1 mL of the culture was collected, and washed twice with 0.5 mL of PBS (pH 7.4). 100 µL of cells were diluted with 3 mL of distilled water. Fluorescence intensities of the cells were analyzed by a MoFlo cell sorter. At least 20,000 events were collected in each measurement. Data were analyzed with Summit software (DakoCytomation). One positive clone showing weakly fluorescence with 2Nal and

highly fluorescence without 2Nal was subjected to further characterization. The plasmid isolated from the positive clone was subjected to DNA sequencing, which revealed that a yPheRS variant contained four mutations (N412G, T415G, S418C, and S437F). The yPheRS variant and the plasmid encoding it were designated yPheRS_naph and pQE9_GFP6_lacI_yPheRS_naph, respectively.

Amino Acid Activation Assay. pQE32-yPheRS (T415G) was prepared previously.¹³ The plasmid fragment encoding yPheRS removed from was pQE9_GFP6_lacI_yPheRS_naph by digestion with NsiI and BglII, and then ligated with a large fragment of pQE32-yPheRS (T415G) generated by digestion with NsiI and BglII to yield pQE32-yPheRS_naph. Expression and purification of yPheRS variants were described previously.¹³ The kinetics of activation of amino acids by vPheRS (T415G) and yPheRS_naph were determined by the amino acid-dependent adenosine triphosphate $(ATP)-[^{32}P]$ -pyrophosphate (PP_i) exchange assay. The assay buffer included 50 mM N-(2hydroxethyl) piperazine-N'-(2-ethanesulfonic acid) (potassium-HEPES) (pH=7.6), 20 mM MgCl₂, 1 mM dithiothreitol (DTT), 2 mM ATP, and 2 mM [³²P]- PP_i. The amino acid concentration varied from 100 nM to 10 mM and the enzyme concentration varied from 100 nM to 800 nM. Detail procedures of the assay were described previously.¹³

Codons. The gene fragment containing four mutations in the binding pocket of yPheRS_naph was excised by *Nsi*I and *Bgl*II digestion of pQE9_GFP6_lacI_yPheRS_naph and inserted into pQE16_mDHFR_yPheRS (T415G)

and pQE9_GFP3_WC_lacI_yPheRS (T415G) between the NsiI and BglII sites to generate pQE16_mDHFR_yPheRS_naph and pQE9_GFP3_WC_lacI_yPheRS_naph, respectively. Both pQE16_mDHFR_yPheRS (T415G) and pQE16_mDHFR_yPheRS naph were co-transformed with pREP4_ytRNA^{Phe}_AAA into AFW competent cells to generate AFW [pQE16 mDHFR yPheRS (T415G)/pREP4 ytRNA^{Phe} AAA] and AFW [pQE16_mDHFR_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA], respectively. In order to express intact mDHFR, pQE16 (Qiagen) and pREP4 plasmids were co-transformed into AFW competent cells to generate AFW [pQE16/pREP4]. Both pQE9_GFP6_lacI_yPheRS_naph and pQE9_GFP3_WC_lacI_yPheRS_naph were transformed DHF AAA electrocompetent into cells to construct DHF [pQE9_GFP6_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] DHF and [pQE9_GFP3_WC_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA], respectively. In order to determine Phe codon occupancy by various amino acids, the AGA (Arg) codon in the second position of GFP6 was mutated to either a UUU or a UUC (Phe) codon by PCR mutagenesis. The PCR reactions were conducted with two pairs of complementary $(2F_UUU_f/2F_UUU_r;$ $2F_UUC_f/2F_UUC_r$) primers using pQE9_GFP6_lacI_yPheRS_naph as a template to generate pQE9_GFP6 (2FUUC)_lacI_yPheRS_naph, (2FUUU)_lacI_yPheRS_naph pQE9_GFP6 and respectively. Removal of a lac promoter and a lac operator, and replacement of a weak ribosome binding site with a stronger one in the upstream of α -subunit of yPheRS gene were performed by two QuickChange mutagenesis reactions using two pairs of complementary primers (R lacP f/R lacP r and Ins_SD_A_Del_lacO_f/Ins_SD_A_Del_lacO_r) using (2FUUU)_lacI_yPheRS_naph and

pQE9_GFP6 (2FUUC)_lacI_yPheRS_naph as templates to generate pQE9_GFP6 (2FUUU)_lacI_SD_yPheRS_naph and pQE9_GFP6 (2FUUC)_lacI_SD_yPheRS_naph and pQE9_GFP6 (2FUUC)_lacI_SD_yPheRS_naph were transformed into DHF_AAA electrocompetent cells to construct DHF [pQE9_GFP6 (2FUUU)_lacI_SD_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] and DHF [pQE9_GFP6 (2FUUC)_lacI_SD_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA].

Residue- and Site-Specific Incorporation of 2Nal into Recombinant Proteins. The AAA anticodon of ytRNA^{Phe}AAA was mutated into GAA by PCR mutagenesis with one pair of primers (ytRNA-GAA_f/ytRNA-GAA_r) using pREP4_ytRNA_AAA as a template. The resulting pREP4_ytRNA_GAA plasmid contained the ytRNA^{Phe}_{GAA} gene. Both pQE9_GFP6 (2FUUU)_lacI_SD_yPheRS_naph and pREP4_ytRNA^{Phe}_GAA were co-transformed into DHF competent cells to generate DHF [pQE9 GFP6 (2FUUU)_lacI_SD_yPheRS_naph/pREP4_ytRNA^{Phe}_GAA]. pREP4_ytRNA_UG was prepared previously.¹³ The yPheRS_naph gene fragment excised by NsiI and BglII digestion of pQE9_GFP6_lacI_SD_yPheRS_naph was inserted into pQE16_mDHFR (38Am)_yPheRS (T415G) between the NsiI and BglII sites to generate pQE16_mDHFR (38Am)_SD_yPheRS_naph. An AAG (Lys) codon was changed to a UAG codon at position 158 of GFP6 sequence in pQE9_GFP6_lacI_yPheRS_naph by PCR mutagenesis with one pair of complementary primers (K158_UAG_f/K158_UAG_r) to yield pQE9_GFP6 (158Am)_lacI_SD_yPheRS_naph. Both pQE16_mDHFR (38Am)_yPheRS_naph and pQE9_GFP6 (158Am)_lacI_SD_yPheRS_naph were cotransformed with pREP4_ytRNA^{Phe}_UG into AFWK and DHF competent cells to construct AFKW [pQE16_mDHFR (38Am)_yPheRS_naph/pREP4_ytRNA^{Phe}_UG] and DHF [pQE9_GFP6 (158Am)_lacI_SD_yPheRS_naph/pREP4_ytRNA^{Phe}_UG], respectively.

Construction of Plasmids and Expression Hosts for Incorporation of 2Nal at Leu Codons. The AAA anticodon of ytRNA^{Phe}AAA was mutated to CAA by PCR mutagenesis with primers ytRNA-CAA_f and ytRNA-CAA_r using pREP4_ytRNA^{Phe}_AAA as a template to yield pREP4_ytRNA^{Phe}_CAA. The expression cassette of mDHFR was excised from pQE16 (Qiagen) by digestion with AatII and NheI and inserted into pQE9_GFP6_lacI_yPheRS_naph between the AatII and NheI sites to generate pQE16_mDHFR_lacI_yPheRS_naph. In order to increase the number of Leu residues encoded as UUG, UUC and UUU (Phe) codons in position 38 and 95 of mDHFR were changed to UUG by PCR mutagenesis reactions with two complementary pairs of primers using pQE16_mDHFR_lacI_yPheRS_naph as a template to generate pQE16_mDHFR2_lacI_yPheRS_naph. PCR mutagenesis reaction was performed to mutate UUG to UUA at position 100 of mDHFR2 with two complementary primers (L100_UUA_f/L100_UUA_r) to yield pQE16_mDHFR2 (100UUA)_lacI_yPheRS_naph. Either pQE16_mDHFR2_lacI_yPheRS_naph or pQE16 mDHFR2 (100UUA) lacI yPheRS naph was co-transformed with ytRNA^{Phe}_{CAA} into MPC390 competent cells MP to yield [pQE16_mDHFR2_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_CAA] or [pQE16_mDHFR2 (100UUA) lacI yPheRS naph/pREP4 ytRNA^{Phe} CAA], respectively. In order to

express intact mDHFR, pQE16 (Qiagen) and pREP4 were co-transformed into MPC390 competent cells to generate MP [pQE16/pREP4]. In order to generate GFP with Cterminal hexa-histidine tag (GFP3C), a PstI restriction site was added to pQE70 (Qiagen) using two complementary primers (PstI 70 f/PstI 70 r) to yield pQE70*. The coding sequence for GFP3 was excised by *PstI* digestion from pQE9 GFP3 yPheRS (T415G) and inserted into pQE70* at the PstI site to generate pQE70*_GFP3C. Expression cassette of GFP3C isolated from pQE70*_GFP3C inserted was into pQE9_GFP3_lacI_yPheRS_naph to yield pQE70*_GFP3C_lacI_yPheRS_naph. Lys 158 in GFP3C was mutated to Leu (encoded as UUG) by PCR mutagenesis with two complementary (K158 UUG f/K158 UUG r) primers using pQE70*_GFP3C_lacI_yPheRS_naph as a template to yield pQE70*_GFP3C (158UUG)_lacI_yPheRS_naph. Both pQE70*_GFP3C_lacI_yPheRS_naph and (158UUG)_lacI_yPheRS_naph pQE70*_GFP3C co-transformed with were pREP4 ytRNA^{Phe} CAA into **MPC390** competent cells to construct MP [pQE70* GFP3C lacI yPheRS naph/pREP4 ytRNA^{Phe} CAA] MP and [pQE70*_GFP3C (158UUG)_lacI_yPheRS_naph/ pREP4_ytRNA^{Phe} CAA], respectively.

Expression of mDHFR Variants and GFP Variants in Vivo. Both AFW and AFWK expression strains co-transformed with pQE plasmid variants and pREP4 plasmid variants were grown in M9 minimal medium supplemented with 0.4 wt % glucose, 35 mg/L thiamin, 1mM MgSO₄, 1mM CaCl₂, 20 amino acids (at 25 mg/L), 35 mg/L kanamycin, and 200 mg/L ampicillin. The overnight cultures of expression strains were diluted 20-fold in fresh M9 minimal medium and incubated at 37 °C. When cells reached

OD of 0.8 - 1.0, cells were spun down and washed twice with cold 0.9% NaCl. The cultures were resuspended in fresh M9 minimal medium supplemented with 18 amino acids (25 μg/mL), and the indicated concentrations of Phe, Trp, and 2Nal. After 10 min incubation, 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) was added to induce expression of GFP at 30 °C. After 4 hours, cells were harvested and either kept at -80 °C or subjected to fluorescence measurement according to the procedures described earlier (see Characterization of the Isolated Clones). Whole cell lysates were analyzed by SDS-PAGE. Due to slow growth of DHF and MPC390 expression hosts co-transformed with pQE plasmid variants and pREP4 plasmid variants, transformants were grown in 2xYT medium to prepare glycerol stocks first. Then glycerol stocks were inoculated into minimal medium supplemented with 20 amino acids (at 25 mg/L) and incubated overnight at 37 °C. The remaining steps were similar to those for AF and AFWK expression hosts.

Quantitative Analysis of Codon Occupancy. Quantitative analysis of codon occupancy was performed by either N-terminal protein sequencing or LC-MS analysis of tryptic digests. The GFP6 (2UUU) and GFP6 (UUC) variants were expressed in minimal medium and purified by Ni-NTA affinity chromatography according to the manufacturer's protocol (Qiagen) under denaturing conditions. The purified GFP variants were subjected to N-terminal protein sequencing using a 492 cLC Procise protein microsequencer (Applied Biosystems, Foster City, CA). Occupancy of Phe codons in mDHFR and Leu codons in GFP was determined by LC-MS analysis. mDHFR expressed in minimal medium were subjected to purification via Ni-NTA affinity chromatography

according to the manufacturer's protocol (Qiagen) under denaturing conditions. After purification, expression levels of GFP and mDHFR were determined by UV absorbance at 280 nm using a calculated extinction coefficient of 20,010 cm⁻¹ M⁻¹ and 24, 750 cm⁻¹ M^{-1} , respectively. The purified proteins were concentrated by ultrafiltration (Millipore). 10 µL of the concentrate was diluted into 90 µL of 75 mM (NH₄)₂CO₃ solution and then 1 µL of modified trypsin (Promega, 0.2 µg/µL) was added. Reaction was carried out for 2-4 hrs at 37 °C and quenched by addition of 13 µL of 5% trifluoroacetic acid (TFA) solution. The solution was then directly subjected to LC-MS analysis conducted on a LCT Premier XE MICROMASS MS system (MS Technologies, Montgomery Village, MD) with Acquity UPLCTM system (Waters, Milford, MA). Tryptic digests were separated by Acquity BEH300 C18 column (1.7 µm, 300 Å, 2.1 x 50mm) using a gradient of 5-95% of solvent B (90% of acetonitrile/10% of 0.1% formic acid solution) and solvent A (2% of acetonitrile/98% of 0.1% formic acid solution) in 10 min. The column eluent was transferred to the electrospray source and mass spectra were recorded.

Results and Discussion

Misincorporation of Trp into Recombinant Proteins in Response to UUU Codons. mDHFR was expressed in AFW [pQE16_mDHFR_yPheRS (T415G)/pREP4_ytRNA^{Phe}_AAA] expression hosts in minimal medium supplemented with 18 amino acids (MM18_FW), 2.5 μ M Phe, 3 mM Trp. Occupancy of the UUU codon sites was determined by LC-MS analysis of tryptic digests of mDHFR expressed with or without 3 mM 2Nal. We focused on Peptide **1** (residues 140-144; LF_{UUU}VTR), which contains a Phe residue at position 141 as UUU. LC-MS analysis of Peptide **1** variants indicated that 10% of position 141 was occupied by Trp (Figure 1b). Trp misincorporation at UUU site can be explained by the relaxed substrate specificity of yPheRS (T415G). yPheRS (T415G) showed 2-fold higher k_{cat}/K_m for Trp than for 2Nal in ATP-PPi exchange assays (Table 1).

Construction of yPheRS Library. In order to enhance the specificity of yPheRS with respect to 2Nal, we explored high-throughput screening of a yPheRS library. On the basis of the crystal structure of the homologous PheRS (tPheRS) from *Thermus thermophilus*⁴⁷ and sequence alignment between tPheRS and yPheRS, four residues (N412, S418, T415, and S437) within 7 Å of the *para* position of the phenyl ring of Phe bound to yPheRS were subjected to saturation mutagenesis (Figure 2). PCR mutagenesis generated 6x10⁶ yPheRS transformants (see Materials and Methods), a population somewhat larger than the theoretical library size (10⁶). By transforming yPheRS library plasmids into DHF expression hosts, 1x10⁷ transformants were obtained and pooled to construct the yPheRS expression library.

High-throughput Screening of yPheRS Library. Screening of the yPheRS library entailed both positive and negative screenings to obtain active and selective yPheRS variants, respectively. Active synthetases would allow incorporation of either 2Nal or natural amino acids at non-permissive sites of GFP6, where amino acids other than Phe would lead to reduction in the fluorescence of cells. Our preliminary study showed that misincorporation of 2Nal at 12 Phe sites of GFP6 resulted in more than 20-fold reduction in fluorescence of cells, likely due to the perturbed folded structure of GFP6 (data not

shown). For positive screening, the yPheRS library was induced with 1 mM IPTG to express GFP6 in minimal medium supplemented with 3 mM 2Nal. The fluorescence histogram of yPheRS library cells expressing GFP6 was obtained by flow cytometry. The gate in the fluorescence channel was set to collect 1% of cells for which fluorescence was slightly above background. The 5×10^4 cells collected were regrown in 2xYT medium and subjected to negative screening to enrich cells containing selective yPheRS variants. The negative screening was based on the fact that synthetases selective for 2Nal should not misincorporate any natural amino acid at Phe sites in GFP6 but E. coli endogenous Phe orthogonal pair will incorporate Phe at the Phe sites. The resulting GFP6 would retain full intensity of fluorescence. For negative screening, GFP6 was expressed in minimal medium in the absence of 2Nal. The top 1% of cells in the fluorescence channel was collected to obtain 5×10^4 cells. The collected cells were regrown in 2xYT medium and entered a next round of screening. After two rounds of positive and negative screening, one more positive screening was carried out to enrich active vPheRS variants toward 2Nal. A portion of the collected cells were spread on agar plates containing suitable antibiotics and ten colonies were subjected to further characterization.

Characterization of the Selective yPheRS Variant. One out of ten clones showed the fluorescence characteristics expected of a selective yPheRS variant. The other clones were false positives that have two peaks coincidently overlapped with the gates for sorting in the histogram of fluorescence. In the absence of 2Nal, cells expressing GFP6 retained full intensity of fluorescence (Figure 4a). However, in the presence of 2Nal, cells expressing GFP6 showed substantially reduced intensity of fluorescence (Figure 4b).

Similar trends were observed in the visual comparison of fluorescence of cell pellets (Figure 4c). The plasmid DNA coding the selective yPheRS variant was isolated from the culture of the clone and subjected to further analysis. DNA sequencing analysis of the plasmid showed that there were mutations of all four positions (N412G, T415G, S418C, and S437F) in the binding pocket of yPheRS. The isolated yPheRS variant was designated yPheRS_naph.

Amino Acid Activation Analysis by yPheRS_naph. Activation of Phe, Trp and 2Nal by both yPheRS (T415G) and yPheRS_naph were examined *in vitro*. The kinetic parameters are listed in Table 1. yPheRS (T415G) showed 2-fold lower activity toward 2Nal than Trp. However, the yPheRS_naph showed 6-fold higher activity toward 2Nal than Trp, which translates into 12-fold enhanced selectivity toward 2Nal (vs. Trp) compared to yPheRS (T415G). The yPheRS_naph also showed 17-fold higher activity toward 2Nal than Phe, while yPheRS (T415G) showed 8.4-fold higher activity. Previously we reported that rationally designed yPheRS (T415A) showed 10-fold higher activity toward pBrF than Trp. However, the activation of Phe by yPheRS (T415A) was comparable to that of pBrF. Therefore, yPheRS_naph is the first yPheRS variant that efficiently discriminates a nonnatural amino acid from all twenty natural amino acids.

The poor binding of Phe and Trp by yPheRS_naph may be understood by loss of favorable interactions between the aromatic rings of the substrate and active-site residues of the synthetase. In a crystal structure of tPheRS,⁴⁷ the side chain of Phe 258 (which corresponds to Asn 412 in yPheRS) makes direct contact with the phenyl ring of the substrate. Since the interaction makes the substrate recognition highly specific and very

favorable energetically, we can readily assume that N412G mutation in yPheRS would lead to poor binding of aromatic amino acid by the PheRS. However, the S418C mutation may not be critical to substrate specificity for 2Nal, because the yPheRS variant containing only three mutations (N412G, T415G and S437F) showed activity toward 2Nal similar to that of yPheRS_naph in ATP-PPi exchange assays. Analysis of fluorescence changes upon incorporation of 2Nal revealed that omission of the S437F mutation leads to a loss of activity toward 2Nal (data not shown). Therefore, S437F mutation was thought to play a key role in discriminating 2Nal from both Phe and Trp.

Elimination of Trp Misincorporation at UUU Codons by yPheRS_naph in Vivo. AFW [pQE16_mDHFR_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] cells were induced to express mDHFR in MM18_FW medium supplemented with 2.5 μM Phe, 50 μM Trp and 3 mM 2Nal. Peptide 1 (residues 140-144; LF_{UUU}VTR), one of the tryptic fragments of purified mDHFR, contains a Phe residue encoded as UUU. Occupancy of the UUU site in Peptide 1 was investigated by LC-MS analysis. The results revealed that use of yPheRS_naph completely eliminated misincorporation of Trp at the UUU site (Figure 1c), while yPheRS (T415G) allowed misincorporation of Trp at the UUU site (Figure 1b). Similar results were obtained by N-terminal sequencing of purified intact GFP6 (2UUU). DHF [pQE9_GFP6_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] cells were induced to express GFP6 (2UUU) in MM18_FW medium supplemented with 50 μM Trp and 3 mM 2Nal. 80% and 20% of the UUU codon at the 2nd position of GFP6 (UUU) were decoded as 2Nal and Phe, respectively; but Trp was not detected at this position (Table 2). Residue- and Single-Site-Specific Incorporation of 2Nal in Vivo. With an appropriate tRNA, the selective yPheRS_naph variant can be used for residue- and single-sitespecific incorporation of 2Nal into proteins. In order to realize residue-specific incorporation of 2Nal, DHF [pQE9 GFP6 (2UUU) lacI yPheRS naph/pREP4 ytRNA^{Phe} GAA] expression hosts were induced to express GFP6 (2UUU) in MM18_FW medium supplemented with 50 µM Trp and 3 mM 2Nal. N-terminal sequencing of the purified GFP6 (2UUU) showed that 92% of position 2 was occupied by 2Nal (Table 2), slightly higher than the 80% occupancy achieved by multiple-site-specific incorporation. The enhanced 2Nal incorporation may be a consequence of the known 12-fold higher aminoacylation rate for ytRNA^{Phe}_{GAA} by yPheRS as compared to ytRNA^{Phe}AAA.⁴⁸ Single-site-specific incorporation of 2Nal into mDHFR_38Am achieved AFWK was by [pQE16_mDHFR_yPheRS_naph/pREP4_ytRNA^{Phe}_UG] in miminal medium supplemented with 25 µg/mL 17 amino acids (MM17 FWK), 50 µM Phe, 50 µM Trp, 50 µM Lys, and 3 mM 2Nal. MALDI-MS analysis of tryptic digests of mDHFR_38Am revealed that 2Nal was dominant at the amber site. Neither Trp nor Phe was detected, confirming the high selectivity of yPheRS_naph toward 2Nal.

Single-Site-Specific Incorporation of 2Nal into GFP in Vivo. Single-site-specific incorporation of 2Nal into GFP6 was investigated to determine whether a correctly folded fluorescent protein can be obtained upon the addition of a nonnatural amino acid. An AAG (Lys) codon was changed to an amber codon in position 158 in GFP6, which is known to be permissive to replacement of Lys with other natural amino acids or to

circular permutation.⁴⁹ DHF [pQE9_GFP6 (158Am)_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_ UG] cells were used to express GFP6 (158Am) in MM17_FWK medium supplemented with 50 μ M Phe, 50 μ M Trp, 50 μ M Lys, and 3 mM 2Nal. The fluorescence of cells expressing full length of GFP6 (158Am) was 280-fold higher than that in uninduced conditions, which clearly indicated that 2Nal can be inserted into GFP6 without substantial reduction of fluorescence.

Misincorporation of 2Nal at Unwanted Sites (UUC Codons) in Vivo. Site-specific incorporation involves introduction of a nonnatural amino acid into a target protein only at programmed sites. Introduction of a nonnatural amino acid at unwanted site will diminish its advantages over the residue-specific incorporation method. Although AFW [pQE16 mDHFR yPheRS (T415G)/pREP4 ytRNA^{Phe} AAA] realized incorporation of 2Nal at programmed UUU codons in mDHFR (Figure 1c), misincorporation of 2Nal at UUC codons was also observed (Figure 1f). Similar results were observed in two GFP variants, GFP6 (2UUU) and GFP6 (2UUC), expressed in DHF [pQE9_GFP6 (2UUC)_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] and [pQE9_GFP6 (2UUU)_lacI_yPheRS/pREP4_ytRNA^{Phe}_AAA], respectively. Supplementation of 2.5 μM; 5 μM Phe into expression medium lowered 2Nal misincorporation at the unwanted sites (UUC codons) from 34% to 20%; 6%. However, the incorporation level of 2Nal at the programmed sites (UUU codons) also decreased from 80% to 62%; 47% (Figure 6). Incorporation of 2Nal at UUC sites in GFP6 (ca. 20%) with supplementation of 2.5 µM Phe is higher than that in mDHFR (ca. 10%) (Figure 1f), which might be explained by more stable characteristics of GFP6 containing 2Nal as compared to mDHFR. Even 6%

of misincorporation of 2Nal at unwanted sites in a target protein could be unfavorable for the target protein to retain its native properties. As a test case, 2Nal misincorporation into GFP6_WC, which contains 12 unwanted sites, led to 10-fold reduction in the fluorescence of cells (Figure 8).

We reasoned that misincorporation of 2Nal at UUC codon resulted from recognition of UUC codons by the AAA anticodon of ytRNA^{Phe}_{AAA}. According to Crick's wobble rule proposed in 1966,³¹ A in the first position of the anticodon can recognize only U in the third position of the codon. Therefore, UUC codon should not be recognized by the AAA anticodon. The discrepancy between the experimental results and Crick's wobble rule may be explained by the expanded wobble rule proposed by Lim and Curran in 2001.³² According to the expanded wobble rule, A in the first position of the anticodon can recognize all four bases in the third position in codon. The base A in the first position of the anticodon favors bases in the order U > C > G > A, consistent with the codon-biased incorporation of 2Nal observed in this work.

Breaking the Degeneracy of the Leucine Codons. According to the expanded wobble rules, C in the first position of the anticodon will recognize only G in the third position of the codon. Therefore, we hypothesized that $ytRNA^{Phe}_{CAA}$ (containing the modified CAA anticodon) would selectively recognize UUG codons. In order to test this hypothesis, mDHFR was expressed in MP [pQE16_mDHFR2_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_CAA] cells. mDHFR contains twenty Leu codons, of which are six UUG, two UUA and twelve CUN (N = A/T/G/C). The expression level of mDHFR was 2.9 mg/L. Occupancy of each Leu codon by various

amino acids was determined by LC-MS analysis of tryptic digests of mDHFR expressed with and without 2Nal. We focused on four peptides. Peptide **3** (residues 165-180; $L_{CUU}L_{CUC}PEYPGVL_{CUC}SEVQEEK$) contains three Leu residues, encoded as CUU and CUC codons. Peptide **4** (residues 54-61; QNL_{CUG}VIMGR) contains a Leu residue, encoded as a CUG codon. Peptide **5** (residues 62-70; $L_{CUU}IEQPEL_{UUG}ASK$) contains two Leu residues, encoded as CUU and UUG codons. Peptide **6** (residues 99-105; $SL_{UUG}DDAL_{UUA}R$) contains two Leu residues, encoded as UUG and UUA codons. 2Nal was not detected at any CUN codon in Peptide **3** and **4** (Figure 8a-d). However, 50% of UUG codons in Peptide **5** and **6** were occupied by 2Nal (Figure 8e-h). In order to determine UUA codon occupancy by 2Nal, Peptide **6**_{UUA} was tested. Peptide **6**_{UUA} is the same as Peptide **6** except both Leu residues are encoded as UUA codons. Since the Peptide **6**_{UUA} variant containing 2Nal was not detected, we conclude that 2Nal incorporation is highly specific to the UUG codon.

Codon-Specific Incorporation of 2Nal. Previously we UUG showed that misincorporation of 2Nal at unwanted UUC codons in GFP6_WC led to 10-fold reduction in fluorescence of cells, even though there are no UUU codons in GFP6_WC. Similar to GFP6_WC, we prepared GFP3 containing twenty three Leu codons, of which are no UUG, four UUA and nineteen CUN (N = A/T/G/C). In order to investigate effect of misincorporation of 2Nal at unwanted sites (Leu codons other than UUG), GFP3 was Е. MP expressed in coli strain [pQE9_GFP3_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_CAA]. The fluorescence intensities of cells expressing GFP3 without 2Nal or with 2Nal were compared (Figure 7c and d).

There was no detectable difference in the fluorescence of cells prepared under these two conditions, confirming the absence of misincorporation of 2Nal at Leu codons other than UUG.

Minimal Perturbation of Native Properties of GFP upon Incorporation of 2Nal. Ideally introduction of a nonnatural amino acid should impart new properties to a target protein with minimal or no perturbation of native properties. As a test case, we chose fluorescence change of GFP variant upon 2Nal incorporation to evaluate perturbation of GFP structure. A UUG codon, an incorporation site for 2Nal, was introduced at the 158th position of GFP3 to generate GFP3 (158UUG) variant. GFP3 (158UUG) variant was expressed in minimal medium supplemented with 17 amino acids (25 μ g/mL), 50 μ M Phe, 50 μ M Trp, 1.25 μ M Leu, and indicated concentration of 2Nal. There was no significant differences in fluorescence histograms of cells expressing GFP3 (158UUG) in the absence and presence of 2Nal, which meant that GFP3 (158UUG) retained its folded structure even upon 2Nal incorporation at the 158th position.

Conclusions

In this chapter, we have shown that the substrate specificity of yPheRS was altered to selectively recognize a nonnatural amino acid, 2Nal. As far as we know, this is the first report of high-throughput screening of a eukaryotic aminoacyl-tRNA synthetase library to obtain synthetase variants selective toward nonnatural amino acids. Use of the selective yPheRS_naph variant resulted in elimination of misincorporation of Trp at UUU codons, which was detected previously in experiments with yPheRS (T415G). Combined use of yPheRS_naph and ytRNA^{Phe}_{CUA_UG} realized high fidelity (ca. 98%) incorporation of 2Nal into proteins in response to an amber codon.

Misincorporation of 2Nal at unwanted sites resulting from the relaxed codon recognition of the AAA anticodon of ytRNA^{Phe} has been overcome by use of the more codon selective ytRNA^{Phe}_{CAA}. The CAA anticodon of ytRNA^{Phe}_{CAA} completely discriminates UUU codons from other five Leu codons. When both yPheRS_naph and ytRNA^{Phe}_{CAA} were overexpressed in *E. coli* expression hosts, 50% of UUG codon sites were occupied by 2Nal, but no other Leu codon sites were occupied by 2Nal, but no other Leu codon sites were occupied by 2Nal. Combined use of yPheRS_naph and the codon selective ytRNA^{Phe}_{CAA} has realized multiple-site-specific incorporation of 2Nal into proteins. We are working to improve the level of incorporation of 2Nal at programmed UUG sites

Acknowledgments

We thank Dr. M. Shahgholi and Dr. J. Zhou for mass spectrometry studies. We also thank Dr. F. Rusnak for help with N-terminal sequencing. We are grateful Dr. A. J. Link for help with the cell sorter. We appreciate I. C. Tanrikulu and P. M. Kekenes-Huskey for help with analysis of the PheRS crystal structure. This work was supported by National Institutes of Health grant.

References

Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* 2003, *328*, 1071-1081.

- Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. Proc. Natl. Acad. Sci. U. S. A. 1998, 95, 455-459.
- Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. Angew. Chem. Int. Ed. 2002, 41, 4066-4069.
- Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. J. Am. Chem. Soc. 2002, 124, 5652-5653.
- Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Proc. Natl. Acad. Sci. U. S. A.
 2002, 99, 19-24.
- 6. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. ChemBioChem 2002, 3, 235-237.
- Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. Proc. Natl. Acad. Sci. U. S. A. 2006, 103, 10180-10185.
- 8. Pratt, E. A.; Ho, C. Fed. Proc. 1974, 33, 1463-1463.
- 9. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. FEBS Lett. 2000, 467, 37-40.
- 10. Tang, Y.; Tirrell, D. A. J. Am. Chem. Soc. 2001, 123, 11089-11090.
- 11. Tang, Y.; Tirrell, D. A. Biochemistry 2002, 41, 10635-10645.
- 12. Wang, P.; Tang, Y.; Tirrell, D. A. J. Am. Chem. Soc. 2003, 125, 6900-6906.
- 13. Kwon, I.; Wang, P.; Tirrell, D. A. J. Am. Chem. Soc. 2006, 128, 11778-11783.
- 14. Anderson, J. C.; Schultz, P. G. Biochemistry 2003, 42, 9598-9608.
- 15.Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* 2001, 40, 11060-11064.
- 16.Kowal, A. K.; Kohrer, C.; RajBhandary, U. L. Proc. Natl. Acad. Sci. U. S. A. 2001, 98, 2268-2273.
- 17. Wang, J. Y.; Xie, J. M.; Schultz, P. G. J. Am. Chem. Soc. 2006, 128, 8738-8739.

- 18. Wang, L.; Schultz, P. G. Chem. Commun. 2002, 1-11.
- 19. Zeng, H. Q.; Me, J. M.; Schultz, P. G. Bioorg. Med. Chem. Lett. 2006, 16, 5356-5359.
- Zhang, Z. W.; Gildersleeve, J.; Yang, Y. Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C. H.;
 Schultz, P. G. *Science* 2004, *303*, 371-373.
- 21. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. J. Am. Chem. Soc. 2005, 127, 14150-14151.
- 22. Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G. J. Am. Chem. Soc. 2006, 128, 388-389.
- 23. Budisa, N.; Pal, P. P. Biol. Chem. 2004, 385, 893-904.
- 24. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. J. Am. Chem. Soc. 2002, 124, 9026-9027.
- 25. Link, A. J.; Tirrell, D. A. J. Am. Chem. Soc. 2003, 125, 11164-11165.
- 26. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. J. Org. Chem. 2003, 68, 174-176.
- 27. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. Nat. Biotechnol.
 2004, 22, 1297-1301.
- 28. Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. J. Am. Chem. Soc. 2003, 125, 7512-7513.
- Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. Nat. Biotechnol.
 2002, 20, 1044-1048.
- 30. Montclare, J. K.; Tirrell, D. A. Angew. Chem. Int. Ed. 2006, 45, 4518-4521.
- 31. Crick, F. H. C. J. Mol. Biol. 1966, 19, 548-555.
- 32. Lim, V. I.; Curran, J. F. RNA 2001, 7, 942-957.
- Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. J. Mol. Biol. 2002, 317, 481-492.

- 34. Watanabe, Y.; Tsurui, H.; Ueda, T.; Furushima, R.; Takamiya, S.; Kita, K.; Nishikawa, K.; Watanabe, K. *Biochim. Biophys. Acta-Gene Struct. Express.* 1997, *1350*, 119-122.
- Inagaki, Y.; Kojima, A.; Bessho, Y.; Hori, H.; Ohama, T.; Osawa, S. J. Mol. Biol.
 1995, 251, 486-492.
- 36. Curran, J. F. Nucleic Acids Res. 1995, 23, 683-688.
- Boren, T.; Elias, P.; Samuelsson, T.; Claesson, C.; Barciszewska, M.; Gehrke, C. W.;
 Kuo, K. C.; Lustig, F. J. Mol. Biol. 1993, 230, 739-749.
- 38. Osawa, S.; Jukes, T. H.; Watanabe, K.; Muto, A. Microbiol. Rev. 1992, 56, 229-264.
- 39. Munz, P.; Leupold, U.; Agris, P.; Kohli, J. Nature 1981, 294, 187-188.
- 40. Lim, V. I.; Aglyamova, G. V. Biol. Chem. 1998, 379, 773-781.
- 41. Lim, V. I. J. Mol. Biol. 1997, 266, 877-890.
- 42. Lim, V. I. J. Mol. Biol. 1995, 252, 277-282.
- 43. Lim, V. I.; Venclovas, C. FEBS Lett. 1992, 313, 133-137.
- 44. Ishii, Y.; Kondo, S. Mutat. Res. 1975, 27, 27-44.
- 45. Tsien, R. Y. Annu. Rev. Biochem. 1998, 67, 509-544.
- 46. Stemmer, W. P. C.; Crameri, A.; Ha, K. D.; Brennan, T. M.; Heyneker, H. L. *Gene* **1995**, *164*, 49-53.
- 47. Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. Acta Crystallogr. D. Biol. Crystallogr.
 2001, 57, 1534-44.
- 48. Sampson, J. R.; Behlen, L. S.; Direnzo, A. B.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 4161-4167.
- 49. Ghosh, I.; Hamilton, A. D.; Regan, L. J. Am. Chem. Soc. 2000, 122, 5658-5659.



Figure 1: LC-MS chromatogram of tryptic digests of mDHFR. Peptide **1** (residues 140-144; $LF_{UUU}VTR$) contains a Phe residue encoded as UUU. Peptide **2** (residues 63-70; TWF_{UUC}SIPEK) contains a Phe residue encoded as UUC codon. Peptide **1** variants containing Phe, Trp, and 2Nal were designated **1**F, **1**W, and **1**Z, respectively. Peptide **2** variants containing Phe and 2Nal were designated **2**F and **2**Z, respectively. These peptides were separated by LC and detected by MS. Unmodified mDHFR was synthesized in a Phe/Trp auxotrophic expression host (a and d) in 2xYT media. Modified mDHFRs were synthesized in a Phe/Trp auxotrophic expression host outfitted with ytRNA^{Phe}_{AAA} and yPheRS (T415G) (b and e) or yPheRS_naph (c and f). The minimal expression media were supplemented with 18 amino acids (25 µg/mL), 2.5 µM Phe, 3 mM Trp, and 3 mM 2Nal.



Figure 2: Phe substrate (red) and four residues (blue) within 7 Å of the *para*-position of the phenyl ring of the substrate inside the binding pocket of a homology model of yPheRS. These four residues were subjected to mutagenesis to generate a yPheRS library. The homology model of yPheRS was prepared by P. M. Kekenes-Huskey at Caltech.



Figure 3: A screening scheme for yPheRS library. GFP6 in yPheRS expression library *E. coli* cells outfitted with ytRNA^{Phe}_{AAA} and yPheRS library was expressed in the presence of 2Nal (a). Weakly fluorescent cells that contain active yPheRS variants were enriched by FACS (b). GFP6 in the collected cells was expressed in the absence of 2Nal (c). Highly fluorescent cells that contain selective yPheRS variants were enriched by FACS (d). After two rounds of screening, ten colonies were isolated from the enriched cells and characterized (e).



Figure 4: Fluorescence intensities of cells containing GFP6 expressed in minimal medium without 2Nal (a); with 2Nal (b). Visual comparison of fluorescence of cells containing GFP6 expressed in the absence of 2Nal (c left); in the presence of 2Nal (c right).



Figure 5: Fluorescence intensities of cells under uninduced conditions (a); cells expressing full length GFP6 (158Am) in the presence of 2Nal (b). The DHF *E. coli* strains were outfitted with $ytRNA^{Phe}_{CUA_UG}$ and $yPheRS_naph$. The expression media were supplemented with 18 amino acids (25 µg/mL), 50 µM Phe, 50 µM Trp, and 3 mM 2Nal.



Figure 6: UUC and UUU codon occupancy by Phe and 2Nal. Both GFP6 (2UUC) and GFP6 (2UUU) were expressed in DHF expression hosts outfitted with yPheRS_naph and ytRNA^{Phe}_{AAA} in minimal medium supplemented with 18 amino acids (25 μ g/mL), 50 μ M Trp, 3 mM 2Nal, and 0; 2.5 μ M; 5.0 μ M Phe. The UUC (a) and UUU (b) codon occupancy by Phe and 2Nal were determined by N-terminal sequencing.



Figure 7: Fluorescence intensities of cells expressing GFP variants. GFP3_WC was expressed in DHF expression hosts outfitted with yPheRS_naph and ytRNA^{Phe}_{AAA} in minimal medium supplemented with 18 amino acids, 5.0 μ M Phe, 50 μ M Trp, and no 2Nal (a); 3 mM 2Nal (b). GFP3 was expressed in MPC390 expression hosts outfitted with yPheRS_naph and ytRNA^{Phe}_{CAA} in minimal medium supplemented with 17 amino acids, 1.25 μ M Leu, 5.0 μ M Phe, 50 μ M Trp, and no 2Nal (c); 3 mM 2Nal (d).



184

Retention time (min)

Figure 8: LC-MS chromatogram of tryptic digests of mDHFR. Peptide 3 (residues 165-180; L_{CUU}L_{CUC}PEYPGVL_{CUC}SEVQEEK) contains three Leu residues encoded as CUU and CUC codons. Peptide 4 (residues 54-61; QNL_{CUG}VIMGR) contains a Leu residue encoded as CUG codon. Peptide 5 (residues 62-70; L_{CUU}IEQPEL_{UUG}ASK) contains two Leu residues encoded as CUU and UUG codons. Peptide 6 (residues 99-105; SL_{UUG}DDAL_{UUA}R) contains two Leu residues encoded as UUG and UUA codons. Peptide 6_{UUA} is the same as Peptide 6 except both Leu residues are encoded as UUA codon. Peptide 3; 4; 5; 6; 6_{UUA} variants containing Leu and 2Nal were designated 3L and 3Z; 4L and 4Z; 5L and 5Z; 6L and 6Z; $6_{UUA}L$ and $6_{UUA}Z$, respectively. These peptides were separated by LC and detected by MS. Unmodified mDHFR was synthesized in the absence of 2Nal in a Phe/Leu auxotrophic expression host (a, c, e, and g) in 2xYT media. Modified mDHFRs were synthesized in a Phe/Leu auxotrophic expression host outfitted with ytRNA^{Phe}_{CAA} and yPheRS_naph (b, d, f, h and i). The expression minimal media were supplemented with 17 amino acids (25 µg/mL), 1.25 µM Leu, 50 µM Phe, 50 µM Trp, and 3 mM 2Nal. No 3Z, 4Z, or $6_{UUA}Z$ was detected by LC-MS analysis.



Figure 9: Fluorescence intensities of cells containing GFP3 (158UUG) expressed in minimal medium supplemented with 17 amino acids (25 μ g/mL), 1.25 μ M Leu, 5.0 μ M Phe, and 50 μ M Trp without 2Nal (a); with 3 mM 2Nal (b). MPC390 expression hosts were outfitted with yPheRS_naph and ytRNA^{Phe}_{CAA}.

 Table 1:
 Kinetic parameters for ATP-PPi exchange by yPheRS (T415G) and yPheRS_naph

Amino Acid	Enzyme	$K_{m}\left(\mu M\right)$	k_{cat} (s ⁻¹)	$k_{cat}/K_m (M^{-1}s^{-1})$	k_{cat}/K_m (rel)
Phe	T415G	55 ± 14	0.202 ± 0.11	$3,500 \pm 1,100$	1 ^a
Trp	T415G	2.83 ± 1.6	0.153 ± 0.003	$63,200 \pm 34,600$	18 ^a
2Nal	T415G	7.03 ± 0.14	0.208 ± 0.04	$29{,}500\pm5{,}800$	8.4 ^a
Phe	naph	$11,000 \pm 2,700$	0.0095 ± 0.0021	0.855 ± 0.007	1 ^b
Trp	naph	$1,\!400\pm600$	0.0035 ± 0.0009	2.52 ± 0.44	2.9 ^b
2Nal	naph	$2{,}000\pm700$	0.030 ± 0.018	14.54 ± 4.22	17 ^b

^a Relative to k_{cat}/K_m for Phe by yPheRS (T415G).

^b Relative to k_{cat}/K_m for Phe by yPheRS_naph.

Codon	ytRNA	Occupancy of codon (%) ^b				
		2Nal	Phe	Lys		
UUU ^a	ytRNA ^{Phe} AAA	80	20	ND ^c		
UUU	ytRNA ^{Phe} GAA	92	8	ND		
UAG ^d	ytRNA ^{Phe} CUA_UG	98	ND	2		

Table 2: Occupancy of UUU and UAG codons by various amino acids

^a The second position in the amino acid sequence of GFP6.

^b UUU and UAG codon occupancy was determined by N-terminal protein sequencing and

LC-MS analysis, respectively.

^c Not detected.

^d The 38th position in the amino acid sequence of mDHFR_38Am.