PROTEIN ENGINEERING VIA SITE-SPECIFIC INCORPORATION OF NONNATURAL AMINO ACIDS

Thesis by

Inchan Kwon

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended January 19, 2007)

Acknowledgments

I would like to acknowledge my thesis advisor David Tirrell. I still remembered Dave's research presentation in my first year at Caltech. I was captivated by his clear presentation and fascinating research area, which influenced me to join his lab. During my days in his lab, I was deeply impressed by his endless curiosity in science and his systematic approach to solve problems with deep knowledge and strong enthusiasm. Without his timely advice and guidance throughout my PhD program, most of my projects would not have been successful. Dave was very supportive of my future career in all aspects.

I also would like to acknowledge other committee members, Prof. Frances Arnold, Prof. Dennis Dougherty, and Prof. Christina Smolke. They also served as my candidacy committee and provided me with many comments and much help since then. I appreciate their time and efforts to review my thesis as well as their great support for my career development.

I am grateful to all former and current Tirrell laboratory members. I was very lucky to work with such wonderful colleagues. In particular, Pin, Jamie, and Rebecca helped me a lot to settle down when I joined Tirrell laboratory. They knew everything I needed and were always willing to assist me. Three postdoctoral fellows, Kent, Jinsang, and Jin impressed me deeply with their commitment and enthusiasm to pursue science. They were also willing to share their experience and helped me to go through many non-scientific issues. They were invaluable role models as young scientists. I thank Caglar and Kimberly for their efforts and time reviewing my writing many times as well as for their helpful

discussions on diverse topics. I appreciate Julie and Jamie for sharing their experience in career development.

Finally, I wish to thank my wife Junghi and my son Hyungju for their continual support and encouragement throughout my PhD program. My daughter Suah, who was born recently, has added great pleasure to my life.

Abstract

Nonnatural amino acid incorporation has been one of the most important protein engineering techniques. Particularly site-specific incorporation of nonnatural amino acids would allow design of artificial proteins containing a nonnatural amino acid with minimal perturbation of their native properties. Site-specific incorporation of phenylalanine (Phe) analogs and tryptophan (Trp) analogs, such as *p*-bromophenylalanine (pBrF), *p*-iodophenylalanine, *p*-azidophenylalanine, 6-chlorotrytophan, 6-bromotryptophan, 5-bromotrytophan, and benzothienylalanine, into proteins in *Escherichia coli* has been realized by *E. coli* strains outfitted with a mutant yeast phenylalanyl-tRNA suppressor ($ytRNA^{Phe}_{CUA}$) and a mutant yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) with a point mutation in the active site of the enzyme. In order to reduce Trp and lysine (Lys) misincorporation at an amber codon, the $ytRNA^{Phe}_{CUA\_UG}$ containing the optimized sequence and the yPheRS (T415A) showing higher specificity toward pBrF were developed. Combining $ytRNA^{Phe}_{CUA\_UG}$ and yPheRS (T415A) allowed incorporation of pBrF into murine dihydrofolate reductase in response to an amber codon with at least 98% fidelity.

Re-assignment of degenerate sense codons offers the prospect of a substantially expanded genetic code and a correspondingly enriched set of building blocks for natural and artificial proteins. Here we describe the use of a mutant yeast phenylalanine transfer RNA ($ytRNA^{Phe}_{AAA}$) containing a modified anticodon to break the degeneracy of the genetic code in *E. coli*. By using an *E. coli* strain co-transformed with $ytRNA^{Phe}_{AAA}$ and a mutant yPheRS (T415G), we demonstrated efficient replacement of Phe by L-3-(2-naphthyl)alanine (2Nal) at UUU, but not at UUC codons.

However, this method had two limitations. First, the yPheRS (T415G) also activated Trp, which led to Trp misincorporation. Second, 2Nal was misincorporated at UUC Phe codons, due to the relaxed codon recognition of AAA anticodon in the $ytRNA^{Phe}_{AAA}$. High-throughput screening of a yPheRS library led to a more selective yPheRS variant (yPheRS_naph). The rationally designed $ytRNA^{Phe}_{CAA}$, which has the CAA anticodon recognizing only a UUG (Leu) codon, allowed incorporation of 2Nal only at UUG codon. Combined use of yPheRS_naph and $ytRNA^{Phe}_{CAA}$ achieved multiple-site-specific incorporation of 2Nal into proteins. These results illustrate a general method for increasing the number of distinct, genetically-encoded amino acids available for protein engineering.

Table of Contents

List of Tables **Page**

List of Figures **Page**

Chapter 1

Expansion of the Number of the Genetically Encoded Amino Acids in *E. coli*

Abstract

Nonnatural amino acid incorporation has been one of the most important protein engineering platforms. In particular, site-specific incorporation of nonnatural amino acids allows design of proteins containing a nonnatural amino acid with minimal perturbation of native properties. Until now more than 30 nonnatural amino acids have been introduced into proteins in *E. coli* hosts outfitted with orthogonal pairs of cognate tRNAs and aminoacyl-tRNA synthetases (aaRS). Substrate specificity of aaRS has been altered to recognize various nonnatural amino acids through either rational design of the active site of aaRS or high-throughput screening of aaRS library. In this approach, site-specificity has been achieved by assigning a nonnatural amino acid to a stop codon, a degenerate codon, or a frameshift codon for efficient discrimination of the programmed sites from unwanted sites.

Introduction

Genetically encoded nonnatural amino acids have been used to endow recombinant proteins with novel chemical, physical, or biological properties. Several distinct methods have been developed to introduce nonnatural amino acids into recombinant proteins at programmed sites in vivo. Residue-specific incorporation involves the global replacement of a particular natural amino acid with a nonnatural amino acid in a target protein using auxotrophic hosts.[1-14] The strength of this technique lies in efficient protein translation and multi-site incorporation of a nonnatural amino acid, because sense codons are reassigned for a nonnatural amino acid. Residue-specific incorporation of 5',5',5'-trifluoroleucine (TFL) into a leucine-zipper protein enhanced its thermal stability.[15] A reactive

phenylalanine analog, *para*-ethynylphenylalanine, has been used for selective dye-labeling of newly synthesized protein in *E. coli*.[16] Newly synthesized proteins in mammalin cells have been selectively indentified using bioorthogonal noncanonical amino acid tagging.[17] Trp analogs containing selenophene and thienyl functional groups have been used in the X-ray crystallography of proteins.[18,19] Aminotryptophan was used to design protein-based pH sensors[3] and greatly changed the spectral properties of fluorescent proteins.[1] Regardless of many successful applications of residue-specific incorporation, its further application to the design of biologically-active macromolecules could be restricted by two limitations. First, reassignment of sense codons for a nonnatural amino acid usually leads to exclusion of one natural amino acid, which is encoded by the sense codons, in protein translation. Second, native properties of a target protein may be impaired by incorporation of a nonnatural amino acid at non-permissive sites of a target protein,[20,21] though the Tirrell laboratory recently showed that the impaired properties may be rescued through directed evolution of the target protein.[21]

Site-Specific Incorporation

Site-specific incorporation involves introduction of a nonnatural amino acid into a target protein at any position selected. Access to all 20 natural amino acids and site-specificity of the method may allow design of proteins containing a nonnatural amino acid with minimal perturbation of native properties. Site-specific incorporation requires outfitting the cell with an “orthogonal pair” comprising a suppressor tRNA and a cognate aminoacyl-tRNA synthetase (aaRS) that operate independently of the endogenous synthetase-tRNA pairs in *E. coli*. The site-specific incorporation of a nonnatural amino acid into recombinant proteins via this strategy was reported by Furter in 1998[22] and the Schultz

laboratory has developed powerful selection methods to identify heterologous synthetases and tRNAs.[23-26] A general strategy for site-specific incorporation consists of four steps and each step will be discussed in more detail in the following sections.

Choice of a Nonnatural Amino Acid

First of all, it is necessary to choose a nonnatural amino acid suitable for goals. Until now more than 30 nonnatural amino acids have been successfully introduced to proteins in a site-specific manner[27,28] (Figure 1) including reactive nonnatural amino acids, such as *para*-acetylphenylalanine **1**, *meta*-acetylphenylalanine **2**, *para*-(3-oxobutanoyl)-L-phenylalanine **3**, *para*-(2-amino-3-hydroxylenthynyl)phenylalanine **4**, *para*-ehthynylthiocarbonyl-phenylalanine **5**, *para*-propargyloxyphenylalanine **6**, *para*-azidophenylalanine **7**, *para*-ethynylphenylalanine **32**, *para*-iodophenylalanine **20**, and *para*-bromophenylalanine **31**. These reactive nonnatural amino acids can be used to modify proteins through bio-orthogonal chemical transformations. For example, the keto and β-diketo functional groups (**1**-**3**) can selectively react with both hydrazides and alkoxyamines.[29-31] Azide functional groups (**7**) can be ligated to alkynes through copper-catalyzed [3+2] cycloaddition.[16,32,33] Aryl halides (**20** and **31**) can be conjugated to terminal alkynes or alkenes via palladium-catalyzed coupling reactions.[34-39] Nonnatural amino acids containing photoactive side chains including *para*-azidophenylalanine **7**,[40,41] *para*-benzoylphenylalanine **9**,[42,43] *O*-(2-nitrobenzyl)tyrosine **12**,[44] and *S*-(2-nitrobenzyl)cysteine **10**[45] have been inserted into proteins.

Genetically-encoded nonnatural amino acids can be used as probes of various kinds. First, nonnatural amino acids containing heavy atoms (I and Se) have been used for X-ray

crystallographic studies of protein structure.[18,46] Second, fluorinated phenylalanine and ^{15}N-labelled methoxy phenylalanine have been used to label proteins for NMR study.[22,47] Third, *para*-L-cyanophenylalanine **22** containing nitrile group has been introduced to proteins as an infrared probe to investigate changes in local environment.[48] Fourth, introduction of a fluorophore into proteins at defined sites can greatly facilitate study of protein function and structure. For example, the two fluorescent amino acids, 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (dansylalanine) **18** and L-(7-hydroxycoumarin-4-yl)ethylglycine **17,** have been introduced to proteins to monitor unfolding of proteins.[49,50]

Orthogonal Pairs

Activation of an amino acid followed by charging the activated amino acid into a cognate tRNA by a cognate aminoacyl-tRNA synthetase (aaRS) is a key step to ensure high fidelity in a protein translation process. One orthogonal pair of aaRS/tRNA is assigned to each natural amino acid. Therefore, a new orthogonal aaRS/tRNA pair is necessary for a nonnatural amino acid to be utilized in protein translation machinery. Therefore, the development of orthogonal pairs to each living organism is one of the key steps in achieving site-specific incorporation of a nonnatural amino acid into proteins in vivo.

Modification of Endogenous Orthogonal Pairs. Initially Schultz and colleagues tried to modify an *E. coli* endogenous RNA/aaRS pair to obtain a new orthogonal pair. Three sites in $tRNA^{Gln}_2$ sequence, which are critical for binding with *E. coli* glutamyl-tRNA synthetase (eGlnRS), were mutated to reduce affinity to eGlnRS. Then eGlnRS was evolved to

recognize the mutant $tRNA^{Gln}_2$ and retain its orthogonality with the endogenous tRNAs in *E. coli*. However, the approach was unsuccessful due to poor discrimination of the mutant $tRNA^{Gln}_2$ against wild-type $tRNA^{Gln}_2$ by even the best eGlnRS mutant.[51]

Orthogonal Pairs Derived from Yeast. A second approach to generate an orthogonal pair was the use of a heterologous tRNA/aaRS pair obtained from another organism in *E. coli*. The yeast phenylalanyl-tRNA/yeast phenylalanyl-tRNA synthetase pair was known to be orthogonal to the *E. coli* endogenous system. In vitro aminoacylation studies indicated that there was little cross-talk between *E. coli* and yeast phenylalanyl-tRNA/synthetase pairs.[52,53] Aminoacylation activity of yeast $tRNA^{Phe}$ by *E. coli* PheRS is 10-fold lower than that of *E. coli* $tRNA^{Phe}$.[22] Aminoacylation activity of *E. coli* $tRNA^{Phe}$ by yeast PheRS is 24- or 34-fold lower than that of yeast $tRNA^{Phe}$,[54,55] because U20 in *E. coli* $tRNA^{Phe}$ replaces G20, one of the identity elements of yeast $tRNA^{Phe}$ recognized by yPheRS (Figure 2).[54] A yeast amber suppressor derived from $tRNA^{Phe}_{GAA}$ ($ytRNA^{Phe}_{CUA\_wt}$) has been widely used to study suppression of amber codons, though *E. coli* lysyl-tRNA synthetase (eLysRS) shows a weak activity to aminoacylate $ytRNA^{Phe}_{CUA\_wt}$ with Lys.[22] A wild-type tRNA sequence can be engineered to eliminate the cross-reactivity with *E.coli* synthetase but preserve efficient recognition by the cognate aaRS. When $ytRNA^{Phe}_{CUA\_wt}$ was co-expressed with wild-type yPheRS in an *E. coli* host, 60% of the amber codon was occupied by Lys.[22] The mischarging by Lys was likely due to the presence of U36 and A73 in the $ytRNA^{Phe}_{CUA\_wt}$, which are the main identity elements for eLysRS.[56] However, when G37 was replaced by A37 and $ytRNA^{Phe}_{CUA}$ was co-expressed with wild-type yPheRS, only 5% of the amber codon was occupied by Lys. The remaining 5% of Lys misincorporation was

eliminated by rational design of ytRNA sequence. Disruption of a Watson-Crick base pair between nucleotides 30 and 40 in ytRNA$^{Phe}_{CUA}$ reduced mischarging by the eLysRS. As an alternative orthogonal tRNA/synthetase pair, an amber suppressor derived from yeast tyrosyl-tRNA (ytRNA^{Tyr}) and yeast tyrosyl-tRNA synthetase (yTyrRS) pair was examined.[57,58] RajBhandary and colleagues generated yTyrRS variants using error-prone PCR to enhance discrimination of the suppressor tRNA over *E. coli* proline tRNA by factors of 2.2- to 6.8-fold. Nishikawa and colleagues optimized the sequence of yeast amber suppressor tRNA^{Tyr} to minimize Lys charging by eLysRS.[59] This optimized yeast tRNA^{Tyr}, together with yeast TyrRS, could be used as an orthogonal pair for incorporation of nonnatural amino acids into proteins. Schultz and colleagues developed two orthogonal pairs derived from yeast. An amber suppressor derived from yeast glutamyl-tRNAs (ytRNA$^{Gln}_{2_CUA}$) is not a substrate for any *E. coli* endogenous synthetase. In turn, yeast glutamyl-tRNA synthetase (yGlnRS) aminoacylated ytRNA$^{Gln}_{2}$ in *E. coli* but did not charge *E. coli* tRNA^{Gln}.[60] An amber suppressor derived from yeast aspartidyl-tRNA (ytRNA$^{Asp}_{CUA}$) is not recognized by *E. coli* synthetases. Introduction of E188K mutation in yeast aspartidyl-tRNA synthetase (yAspRS) led to reduction in aminoacylation activity with *E. coli* tRNA^{Asp}.[61] Therefore, both ytRNA$^{Gln}_{2_CUA}$/yGlnRS pair and ytRNA$^{Asp}_{CUA}$/yAspRS (E188K) form orthogonal pairs in *E. coli* suitable for the incorporation of nonnatural amino acids into proteins in vivo.

Orthogonal Pairs Derived from Archaebacteria. Schultz and colleagues developed an orthogonal tyrosyl-tRNA/synthetase pair obtained from archaea, since archeal aminoacyl-tRNA synthetases are more similar to eukaryotic than prokaryotic counterparts. The amber

suppressor tyrosyl-tRNA (*Mj*tRNA^{Tyr})/synthetase (*Mj*TyrRS) pair derived from *Methanococcus jannaschii* was selected and evolved to be orthogonal to the *E. coli* endogenous system.[25,62] The amber suppressor *Mj*tRNA$^{Tyr}_{CUA}$ was efficiently aminoacylated by *Mj*TyrRS in *E. coli*. However, *E. coli* endogenous synthetases showed weak activity towards *Mj*tRNA$^{Tyr}_{CUA}$. Therefore, a selection method to obtain orthogonal tRNAs in *E. coli* has been developed to enhance orthogonality with respect to *E. coli* endogenous synthetases and retain recognition by the cognate synthetase and the protein translational machinery (Figure 3).[25] This method includes positive and negative selection of a mutant tRNA library. In the negative selection, amber codons were introduced into a toxic barnase protein at permissive sites that allow replacement of amino acids. Eleven nucleotides of *Mj*tRNA$^{Tyr}_{CUA}$ that do not bind to *Mj*TyrRS were saturated with all four bases to generate a tRNA library. When the tRNA library was expressed in minimal medium without *Mj*TyrRS, cells containing tRNA variants aminoacylated by *E. coli* endogenous synthetases resulted in cell death. The plasmids were isolated from the cells survived, and co-transformed with a plasmid harboring both a *Mj*TyrRS gene and a β-lactamase gene containing an amber codon at a permissive site into *E. coli* hosts. When cells were cultured in liquid medium in the presence of ampicillin, cells harboring tRNA variants aminoacylated by *Mj*TyrRS survived. After the negative and positive screening of tRNA library, the tRNAs were orthogonal to the *E. coli* endogenous synthetases and retained their affinity to the cognate synthetase. Schultz and colleagues developed another orthogonal pair obtained from Archea. An amber suppressor leucyl-tRNA and a leucyl-tRNA synthetase were derived from *Halobacterium sp. NRS-1* and *Methanobacterium thermoautotrophicum*, respectively. Aminoacylation activity of the suppressor leucyl-

tRNA by the leucyl-tRNA synthetase and its orthogonality with *E. coli* synthetases were significantly enhanced by extensive mutations in the anticodon loop and acceptor stem.[23]

Substrate Specificity Change

Another issue to achieve incorporation of a nonnatural amino acid into protein in vivo is alteration of substrate specificity of an aminoacyl-tRNA synthetase toward the nonnatural amino acid. Aminoacyl-tRNA synthetases ensure the high fidelity of transforming genetic code sequences into biologically functional proteins through a two-step aminoacylation reaction.[63-67] In the first step, the cognate amino acid is activated by aaRS in the presence of ATP to form the amino acid adenylate. Subsequently aaRS catalyzes the esterification reaction to join the amino acid to the 2'- or 3'-OH of the terminal ribonucleotide of its cognate tRNA. Manipulation of such reactions could potentially alter the genetic code to allow incorporation of novel amino acid into proteins in vivo.[68] Some nonnatural amino acids including TFL, azidohomoalanine, and *para*-fluorophenylalanine, are recognized by wild-type cognate aminoacyl-tRNA synthetases.[5,15,22] However, the majority of nonnatural amino acids are very poor substrates for natural aminoacyl-tRNA synthetases. Engineering the new synthetase activity is almost inevitable when one wants to incorporate an amino acid that is not recognized by wild-type enzymes.

Rational Design of the Active Sites of Aminoacyl-tRNA Synthetases. Alteration of substrate specificity of aminoacyl-tRNA synthetases has been achieved in several labs by the rational design of the binding pocket to accommodate nonnatural amino acid. Tirrell

and colleagues have been interested in exploring the possibility of incorporating substituted phenylalanines, such as *p*-bromophenylalanine, *p*-iodophenylalanine, and *p*-azidophenylalanine. On the basis of the crystal structure of the PheRS (tPheRS) from *Thermus thermophilus* (Figure 4a and b), Safro and colleagues proposed that Val 261 and Ala 314 in the amino acid binding pocket of α-subunit of the tPheRS are critical in the discrimination of Phe from its amino acid competitors. Sequence alignment indicates that Ala 314 in α-subunit of the tPheRS corresponds to Ala 294 in α-subunit of the ePheRS (Figure 4c), and the Hennecke group showed that the substrate specificity of the ePheRS can be relaxed by a mutation at Ala 294. The A294G mutant was shown to enable incorporation of *para*-chlorophenylalanine into recombinant proteins.[69] A subsequent computational simulation, consistent with the Safro prediction, identified two cavity-forming mutations (T251G and A294G) in ePheRS binding pocket. These two mutations led to relaxed substrate specificity and efficient in vivo replacement of Phe by *para*-acetylphenylalanine (pAcF).[4] Sequence alignment showed that Thr 415 in yPheRS is equivalent to Thr 251 in ePheRS. Therefore, a T415G mutation was introduced into the yPheRS to enlarge the active site. The yPheRS (T415G) variant showed activity for *p*-bromophenylalanine, *p*-iodophenylalanine, and *p*-azidophenylalanine in ATP-PPi exchange assays in vitro (see Chapters 2 and 4). There was an interesting finding that the yPheRS (T415G) efficiently activated Trp and 2-naphthylalanine as well as Phe analogs. Considering the spectrum of nonnatural amino acids recognized by yPheRS (T415G), Tirrell and colleagues explored whether the yPheRS (T415G) can activate even Trp analogs. In vitro activation studies showed that the yPheRS (T415G) efficiently activated four Trp analogs, 6-chlorotryptophan (6ClW), 6-bromotryptophan (6BrW), and 5-

bromotrytophan (5BrW), which were not utilized by the endogenous *E. coli* translational system (see Chapter 3). A phenylalanine auxotrophic *E. coli* strain transformed with this yPheRS (T415G) and cognate suppressor tRNA enabled the assignment of an amber nonsense codon to Phe analogs and Trp analogs in vivo (see Chapters 2-4). The cavity formed due to the T415G mutation in the binding pocket of yPheRS led to the activation of Trp as well as nonnatural amino acids, which resulted in misincorporation of Trp at the programmed sites. Therefore, the binding site of yPheRS was re-designed to enhance specificity for pBrF. Specifically, Tirrell and colleagues used the T415A variant, which exhibits 5-fold higher activity towards pBrF as compared to Trp in ATP-PPi exchange assays. Use of the yPheRS (T415A) eliminated misincorporation of Trp at programmed sites in proteins.[70]

Nishikawa and colleagues have changed substrate specificity of yeast tyrosyl-tRNA synthetase (yTyrRS) to accommodate tyrosine analogs.[71] Based on the crystal structure of TyrRS obtained from *Bacillus stearothermophilus* TyrRS and sequence alignment of two homologous TyrRSs, active site residues in yTyrRS have been mutated to generate a series of yTyrRS variants. Among the yTyrRS variants, one containing the Tyr43Gly mutation (yTyrRS (Y43G)) was found to show activity for several 3-substituted tyrosine analogs in aminoacylation assays in vitro. yTyrRS (Y43G) efficiently charged ytRNA^{Tyr} with 3-iodo-tyrosine, which was not utilized by wild-type yTyrRS. The yTyrRS (Y43G) showed 400-fold lower activity towards tyrosine than wild-type yTyrRS. Similarly, Yokoyama and colleagues[72] screened *E. coli* TyrRS variants recognizing 3-iodotyrosine via in vitro biochemical assays. The *E. coli* TyrRS variant that has two mutations (Y37V and Q195C) in the active site activated 3-iodotyrosine 10-fold more efficiently than tyrosine.[72]

Combined use of *E. coli* TyrRS (V37C195) and a suppressor tRNA derived from *Bacillus stearothermophilus* allowed incorporation of 3-iodotyrosine into proteins in mammalian cells.[73]

Directed Evolution of Aminoacyl-tRNA Synthetase Substrate Specificity. Aminoacyl-tRNA synthetases are known to be readily evolvable. Schultz and colleagues have developed powerful selection methods to change the substrate specificity of tyrosyl-tRNA synthetase (*Mj*TyrRS) derived from *Methanococcus jannaschii* toward nonnatural amino acids.[23,24,26,74] The selection method consists of a series of positive and negative selections.[68] Based on the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus*, five residues in the active site of *Mj*TyrRS were selected and randomized to generate a *Mj*TyrRS mutant library. In the positive selection, an amber codon was inserted at permissive sites in a chloramphenicol acetyltransferase (CAT) gene. The library was cultured in medium containing chloramphenicol in the presence of a nonnatural amino acid. Only the cells containing *Mj*TyrRS variants allowing suppression of an amber codon in CAT gene survived. Surviving cells were subjected to the negative selection. In the negative selection, the cells were regrown in medium containing chloramphenicol in the absence of the nonnatural amino acid. Those cells that could not survive were isolated from the replica plates containing the nonnatural amino acid. Using this method, substrate specificity of *Mj*TyrRS was changed to selectively recognize *O*-methyl-*L*-tyrosine.[68]

Alternative selection methods for evolving substrate specificity of *Mj*TyrRS towards nonnatural amino acids have been developed. One of them involves a positive

selection similar to the previous one, but a different negative selection.[60] The negative selection is based on suppression of amber codons inserted at permissive sites in the barnase gene that is lethal to *E. coli*. When the cells survived from the positive selection were cultured in the absence of nonnatural amino acids, the cells containing *Mj*TyrRS variants recognizing natural amino acids resulted in cell death due to expression of the barnase protein. Another method for evolving *Mj*TyrRS specificity to nonnatural amino acids involves an amplifiable fluorescent reporter, a green fluorescent protein.[24] Amber codons were introduced at various sites in a T7 RNA polymerase gene that controlled expression of a green fluorescent protein. In the positive screening, the *Mj*TyrRS library cells were cultured in the presence of a nonnatural amino acid. The cells containing *Mj*TyrRS variants that recognized the nonnatural amino acid and suppressed the amber codons in T7 RNA polymerase gene became fluorescent and were collected with a fluorescence-activated cell sorter (FACS). In the negative screening, the cells collected from the positive screening were cultured in the absence of the nonnatural amino acid; the cells containing *Mj*TyrRS variants that did not recognize any natural amino acid were not fluorescent and collected with a FACS. Therefore, the cells collected from the positive and negative screenings contained the *Mj*TyrRS variants that can selectively charge the nonnatural amino acid into the cognate tRNA. One advantage of the latter method is that both reporter genes are encoded in a single plasmid, which eliminates the need for DNA isolation after each selection step. However, the method requires a FACS for high-throughput screening.

Most attempts to evolve aminoacyl-tRNA synthetase substrate specificity have been limited to prokaryotic or archaeal aminoacyl-tRNA synthetases. Recently, Tirrell and

colleagues have evolved a eukaryotic aminoacyl-tRNA synthetase, yPheRS, to change its substrate specificity towards a hydrophobic phenylalanine analog, 2-naphthylalanine (see Chapter 6). The yPheRS library was constructed by randomizing four residues (N412, T415, S418, and S437) that contact the Phe substrate inside the binding pocket (Figure 5a). The screening consisted of a series of positive and negative screenings (Figure 5b), and involved expression of a GFP reporting folding status. The active synthetases would allow incorporation of either 2Nal or other natural amino acids at non-permissive sites of a GFP, which resulted in poorly-folded GFP showing weak fluorescence. Therefore, in the positive screening, the library cells were cultured in the presence of 2Nal and then the cells containing active synthetases were collected by high-throughput screening with a FACS. The yPheRS variants with very low activity toward natural amino acids will not misincorporate any natural amino acid at Phe sites in GFP and so GFP would retain full intensity of fluorescence. In the negative screening, the cells obtained from the positive screening were cultured in the absence of 2Nal and then the cells containing selective synthetases were collected with a FACS.

New Codon-Anticodon Interactions

Among the sixty-four codons, sixty-one (sense codons) are assigned to one of the twenty natural amino acids. The other three codons (nonsense codons) are stop codons that terminate protein translation. In order to incorporate a nonnatural amino acid into proteins in vivo, we need to either re-assign an existing codon or generate a new codon for the nonnatural amino acid. Until now three distinct approaches have been explored, re-

assignment of a stop codon, reassignment of a degenerate codon, and assignment of a frameshift codon.

Nonsense Suppression. In 1989 incorporation of a nonnatural amino acid into proteins in vitro was demonstrated by two independent groups, the Chamberlin group[75] and the Schultz group.[76] Nonsense suppression means that one of the stop codons (nonsense codons) can be read by a suppressor charged with an amino acid. Otherwise protein translation will be terminated at the stop codon. Three stop codons, amber (UAG),[46,68,70,74,77] ochre (UAA),[23,78,79] and opal (UGA) stop codons,[23,78] have been reassigned for nonnatural amino acids. The amber codon has been used most widely,[22,40,42,43,48,70,74,76,80-83] because it is the least common stop codon in *E. coli*, and because several naturally occurring suppressor tRNAs recognize it efficiently.[84,85] Use of the amber codon to encode nonnatural amino acids requires outfitting the cell with an "orthogonal pair" comprising a suppressor tRNA and a cognate aminoacyl-tRNA synthetase (aaRS) that operate independently of the endogenous synthetase-tRNA pairs in *E. coli* as described in previous sections. To date, more than thirty nonnatural amino acids have been incorporated into proteins in response to the amber stop codon with high fidelity.[27,28,70]

Breaking the Degeneracy of the Genetic Code. Suppression of the amber stop codon has worked well for single-site insertion of novel amino acids. However, the utility of this approach is limited by moderate suppression efficiency. Competition between elongation and termination produces mixtures of full-length and truncated chains, and protein yields

are reduced accordingly. Efficient multi-site incorporation has been accomplished by replacement of natural amino acids in auxotrophic *E. coli* strains,[1,6,9-11,86,87] and by using aminoacyl-tRNA synthetases with relaxed substrate specificity[4] or attenuated editing activity.[11] Although this method provides efficient incorporation of analogs at multiple sites, it suffers from the limitation that one of the natural amino acids must be excluded from the engineered protein. Efficient multiple-site-specific incorporation can be achieved by re-assignment of a sense codon, if the sense codon can be efficiently distinguished from other sense codons. Tirrell and colleagues investigated re-assignment of a degenerate codon for a nonnatural amino acid. As a test case for establishing the feasibility of breaking the degeneracy of the code, phenylalanine codons were investigated (see Chapter 5).

Although 2Nal incorporaton was biased to UUU codons in the first study, misincorporation of 2Nal at unwanted sites (UUC codons) could not be avoided due to ambiguity in codon recognition by the AAA anticodon of ytRNA$^{Phe}_{AAA}$. Misincorporation of a nonnatural amino acid at unwanted sites might cause perturbation or loss of function of a target protein.[20,21] Based on experimental findings[88-95] and stereo-chemical modeling[96-99] of codon-anticodon interactions, an expanded wobble rule was proposed by Lim and Curran.[90] The expanded wobble rule indicates that A in the first position of the anticodon can recognize all four bases in the third position of the codon, although its affinity to C is lower than that to U. The observed incomplete discrimination of UUU from UUC codons by the AAA anticodon of ytRNA^{Phe} motivated us to explore a different set of degenerate codons. Several considerations recommend the UUG codon. First, Leu is encoded by six codons: UUA, UUG, CUA, CUG, CUU, and CUC. Discrimination of UUG from CUN (N = A/U/G/C) codons might be highly efficient due to discrimination at the first position in

the codon. Second, our existing yeast orthogonal pair should be readily adapted to the incorporation of Phe analogs in response to UUG codons. Third, according to the expanded wobble rules, C in the first position of the anticodon can recognize only G in the third position of the codon (see Chapter 6).

Frameshift Suppression. The utility of stop or degenerate codons might be restricted by limited number of codons. In order to overcome this limitation, extended codons have been investigated. Although three-base codons are universal for most living organisms, frameshift suppression of a four-base codon occurs is a regulatory mechanism in viruses and bacteria, and is mediated either by slipping of normal tRNAs at the ribosome or by natural frameshift suppressor tRNAs with four-base anticodons. Sisido and colleagues have pioneered frameshift suppression for incorporation of nonnatural amino acids into proteins in *E. coli*.[100-103] In this approach, four-base codons were inserted into a target protein at programmed sites. When the four-base codons are read by a modified tRNA that is charged with nonnatural amino acid and contains the corresponding anticodon, the full length target protein can be synthesized. If the four-base codons are read as three-base codons, truncated proteins will be formed due to early termination of protein translation. Four-base codons were usually derived from rarely used codons to minimize reading as three-base codons by the endogenous tRNAs. The translation efficiencies of some four-base codons were higher than those of three-base stop codons. In particular, a GGGU codon showed 86% translation efficiency,[104] while the suppression efficiencies of an amber codon were ca. 20-60%.[75,105] A number of four-base codons; CGGG, GGGU, CGGU, AGGU, CCCU, and CUCU have been generated.[101-103,106-108] The four-base codon strategy has also been applied to

incorporate two different nonnatural amino acids into two programmed sites in a protein, which showed orthogonality of each four-base codon to other four-base codons as well as three-base codons.[103,104,109] The frameshift suppression method has been extended to five-base codon to incorporate nonnatural amino acids.[110] Recently, Sisido and colleagues have demonstrated frameshift suppression of a four-base codon (UAGN) in mammalian cells using tRNA containing NCUA anticodon.[108]

Schultz and colleagues developed a combinatorial approach to select efficient suppressors of four-base codons; AGGA, UAGA, CCCU, and CUAG.[23,111] This approach has been extended to five- and six-base codons with tRNAs containing 6-10 nucleotides in the anticodon loops.[112] They identified suppressors of several five-base codons, such as AGGAU, CUACU, and CUAGU. However, translation efficiency of five-base codons was ca. 10%, which was lower than that of four-base codons. Suppression of six-base codons was too weak to be detected.[112] Recently, Schultz and colleagues have shown that a four-base codon can be used to incorporate a nonnatural amino acid into proteins in *E. coli* outfitted with a heterologous tRNA/synthetase orthogonal pair. In particular, L-homoglutamine was charged into archeal $tRNA^{Lys}$ by the evolved archeal lysyl-tRNA synthetase, and then was inserted into proteins in response to an AGGA codon in *E. coli*.[113]

Conclusions

In 1989 the concept of site-specific incorporation in vitro was independently demonstrated by the Chamberlin group and the Schultz group. Recently, several groups achieved site-specific incorporation of nonnatural amino acids into proteins in vivo using *E. coli* hosts outfitted with a twenty-first orthogonal pair of tRNA/aaRS. Schultz and

colleagues altered the substrate specificity of TyrRS, which was derived from *M. jannaschii*, to various nonnatural amino acids *via* high-throughput screening of a *Mj*TyrRS library. Tirrell and colleagues altered yPheRS substrate specificity toward Phe analogs, which led to site-specific incorporation of Phe analogs in *E. coli*. The moderate suppression efficiency of amber codons has motivated exploration of alternative codons that encode nonnatural amino acids. Breaking the degeneracy of the genetic code or frameshift suppression greatly expands the theoretical number of codons that can be assigned to nonnatural amino acids.

References

1. 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.
2. Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455-459.
3. 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.
4. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.
5. 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.

7. 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. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.

14. Budisa, N. *Angew. Chem. Int. Ed.* **2004**, *43*, 6426-6463.

15. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem.Int. Ed.* **2001**, *40*, 1494-1496.

16. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 14150-14151.

17. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9482-9487.

18. Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2001**, *309*, 925-936.

19. Budisa, N.; Pal, P. P.; Alefelder, S.; Birle, P.; Krywcun, T.; Rubini, M.; Wenger, W.; Bae, J. H.; Steiner, T. *Biol. Chem.* **2004**, *385*, 191-202.

20. Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893-904.

21. Montclare, J. K.; Tirrell, D. A. *Angew. Chem. Int. Ed.* **2006**, *45*, 4518-4521.

22. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

23. Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598-9608.

24. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

25. Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883-890.

26. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1-11.

27. Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.

28. Xie, J. M.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548-554.

29. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 56-61.

30. Zhang, Z. W.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735-6746.

31. Zeng, H. Q.; Me, J. M.; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5356-5359.

32. Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.

33. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602.

34. Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; Yokoyama, S. *ChemBioChem* **2006**, *7*, 134-139.

35. Nguefack, J. F.; Bolitt, V.; Sinou, D. *Tetrahedron Lett.* **1996**, *37*, 5527-5530.

36. Sakakibara, Y.; Ido, Y.; Sasaki, K.; Sakai, M.; Uchino, N. *Bull. Chem. Soc. Jpn.* **1993**, *66*, 2776-2778.

37. Schoenbe, A.; Heck, R. F. *J. Org. Chem.* **1974**, *39*, 3327-3331.

38. Yoneyama, M.; Kakimoto, M.; Imai, Y. *Macromolecules* **1988**, *21*, 1908-1911.

39. Yoneyama, M.; Konishi, T.; Kakimoto, M. A.; Imai, Y. *Makromol. Chem., Rapid Commun.* **1990**, *11*, 381-386.

40. 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.

41. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

42. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.

43. Chin, J. W.; Schultz, P. G. *ChemBioChem* **2002**, *3*, 1135-1137.

44. Deiters, A.; Groff, D.; Ryu, Y. H.; Xie, J. M.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2006**, *45*, 2728-2731.

45. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306-14307.

46. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22*, 1297-1301.

47. Deiters, A.; Geierstanger, B. H.; Schultz, P. G. *ChemBioChem* **2005**, *6*, 55-58.

48. Schultz, K. C.; Supekova, L.; Ryu, Y. H.; Xie, J. M.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984-13985.

49. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9785-9789.

50. Wang, J. Y.; Xie, J. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 8738-8739.

51. Liu, D. R.; Magliery, T. J.; Pasternak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10092-10097.

52. Doctor, B. P.; Mudd, J. A. *J. Biol. Chem.* **1963**, *238*, 3677-3681.

53. Kwok, Y.; Wong, J. T. F. *Can. J. Biochem.* **1980**, *58*, 213-218.

54. Sampson, J. R.; Direnzo, A. B.; Behlen, L. S.; Uhlenbeck, O. C. *Science* **1989**, *243*, 1363-1366.

55. Roe, B.; Sirover, M.; Dudock, B. *Biochemistry* **1973**, *12*, 4146-4154.

56. Tamura, K.; Himeno, H.; Asahara, H.; Hasegawa, T.; Shimizu, M. *Nucleic Acids Res.* **1992**, *20*, 2335-2339.

57. Kowal, A. K.; Kohrer, C.; RajBhandary, U. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2268-2273.

58. Ohno, S.; Yokogawa, T.; Fujii, I.; Asahara, H.; Inokuchi, H.; Nishikawa, K. *J. Biochem. (Tokyo).* **1998**, *124*, 1065-1068.

59. Fukunaga, J.; Yokogawa, T.; Ohno, S.; Nishikawa, K. *J. Biochem. (Tokyo).* **2006**, *139*, 689-696.

60. Liu, D. R.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4780-4785.

61. Pastrnak, M.; Magliery, T. J.; Schultz, P. G. *Helv. Chim. Acta* **2000**, *83*, 2277-2286.

62. Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010-5011.

63. Francklyn, C.; Perona, J. J.; Puetz, J.; Hou, Y. M. *RNA* **2002**, *8*, 1363-1372.

64. de Pouplana, L. R.; Schimmel, P. *Trends Biochem. Sci.* **2001**, *26*, 591-596.

65. Ibba, M.; Soll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617-650.

66. Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881-889.

67. Schimmel, P. *Annu. Rev. Biochem.* **1987**, *56*, 125-158.

68. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

69. Ibba, M.; Kast, P.; Hennecke, H. *Biochemistry* **1994**, *33*, 7107-7112.

70. Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

71. Ohno, S.; Yokogawa, T.; Nishikawa, K. *J. Biochem. (Tokyo).* **2001**, *130*, 417-423.

72. Kiga, D.; Sakamoto, K.; Kodama, K.; Kigawa, T.; Matsuda, T.; Yabuki, T.; Shirouzu, M.; Harada, Y.; Nakayama, H.; Takio, K.; Hasegawa, Y.; Endo, Y.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9715-9720.

73. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res.* **2002**, *30*, 4692-4699.

74. Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 388-389.

75. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.

76. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

77. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

78. Kohrer, C.; Xie, L.; Kellerer, S.; Varshney, U.; RajBhandary, U. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14310-14315.

79. Kohrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. *Chem. Biol.* **2003**, *10*, 1095-1102.

80. Alfonta, L.; Zhang, Z. W.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 14662-14663.

81. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964-967.

82. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 2910-2914.

83. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 935-939.

84. Thorbjarnardottir, S.; Uemura, H.; Dingermann, T.; Rafnar, T.; Thorsteinsdottir, S.; Soll, D.; Eggertsson, G. *J. Bacteriol.* **1985**, *161*, 207-211.

85. Yoshimura, M.; Inokuchi, H.; Ozeki, H. *J. Mol. Biol.* **1984**, *177*, 627-644.

86. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148-2152.

87. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

88. Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *J. Mol. Biol.* **2002**, *317*, 481-492.

89. 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.

90. Lim, V. I.; Curran, J. F. *RNA* **2001**, *7*, 942-957.

91. Inagaki, Y.; Kojima, A.; Bessho, Y.; Hori, H.; Ohama, T.; Osawa, S. *J. Mol. Biol.* **1995**, *251*, 486-492.

92. Curran, J. F. *Nucleic Acids Res.* **1995**, *23*, 683-688.

93. 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.

94. Osawa, S.; Jukes, T. H.; Watanabe, K.; Muto, A. *Microbiol. Rev.* **1992**, *56*, 229-264.

95. Munz, P.; Leupold, U.; Agris, P.; Kohli, J. *Nature* **1981**, *294*, 187-188.

96. Lim, V. I.; Aglyamova, G. V. *Biol. Chem.* **1998**, *379*, 773-781.

97. Lim, V. I. *J. Mol. Biol.* **1997**, *266*, 877-890.

98. Lim, V. I. *J. Mol. Biol.* **1995**, *252*, 277-282.

99. Lim, V. I.; Venclovas, C. *FEBS Lett.* **1992**, *313*, 133-137.

100. Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. *Biomacromolecules* **2000**, *1*, 118-125.

101. Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Sisido, M. *J. Am. Chem. Soc.* **1998**, *120*, 7520-7529.

102. Muranaka, N.; Hohsaka, T.; Sisido, M. *FEBS Lett.* **2002**, *510*, 10-12.

103. Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *J. Am. Chem. Soc.* **2002**, *124*, 14586-14590.

104. Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060-11064.

105. Anderson, R. D.; Zhou, J.; Hecht, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 9674-9675.

106. Sisido, M.; Ninomiya, K.; Ohtsuki, T.; Hohsaka, T. *Methods* **2005**, *36*, 270-278.

107. Taki, M.; Hohsaka, T.; Murakami, H.; Taira, K.; Sisido, M. *FEBS Lett.* **2001**, *507*, 35-38.

108. Taki, M.; Matsushita, J.; Sisido, M. *ChemBioChem* **2006**, *7*, 425-428.

109. Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 34-40.

110. Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Nucleic Acids Res.* **2001**, *29*, 3646-3651.

111. Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *J. Mol. Biol.* **2001**, *307*, 755-769.

112. Anderson, J. C.; Magliery, T. J.; Schultz, P. G. *Chem. Biol.* **2002**, *9*, 237-244.

113. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7566-7571.

114. Dudock, B. S.; Diperi, C.; Michael, M. S. *J. Biol. Chem.* **1970**, *245*, 2465-2468.

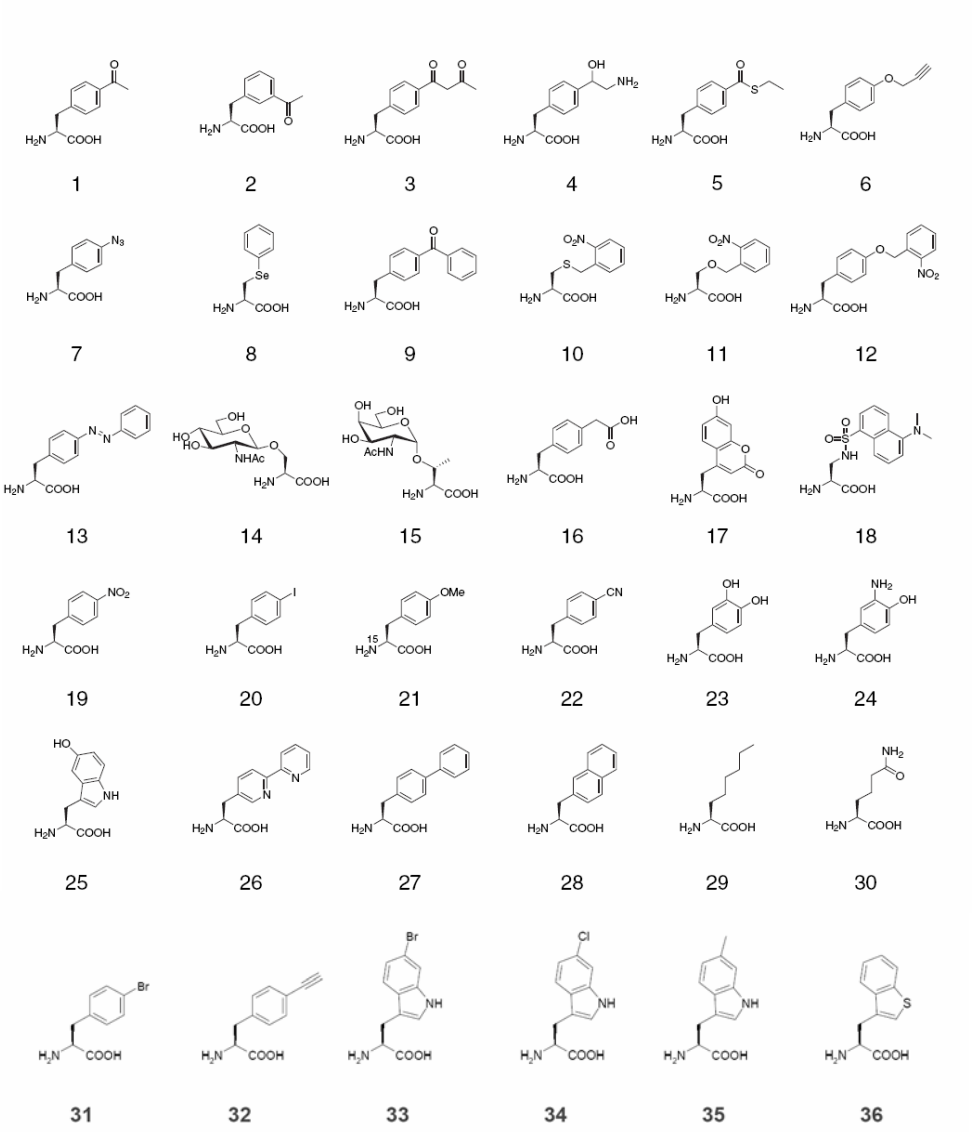


Figure 1: Nonnatural amino acids that have been genetically incorporated into proteins in bacteria, yeast or mammalian cells. This figure is reprinted, with permission, from the *Annual Review of Biophysics and Biomolecular Structure*, Volume **35**, Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code, 225,[27] Copyright (2006) by Annual Reviews www.annualreviews.org.

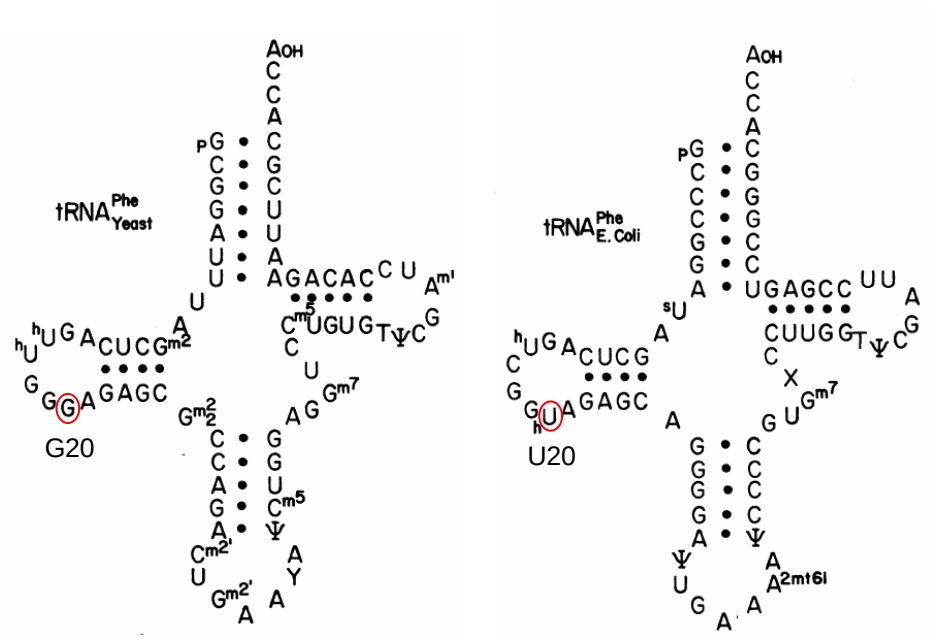
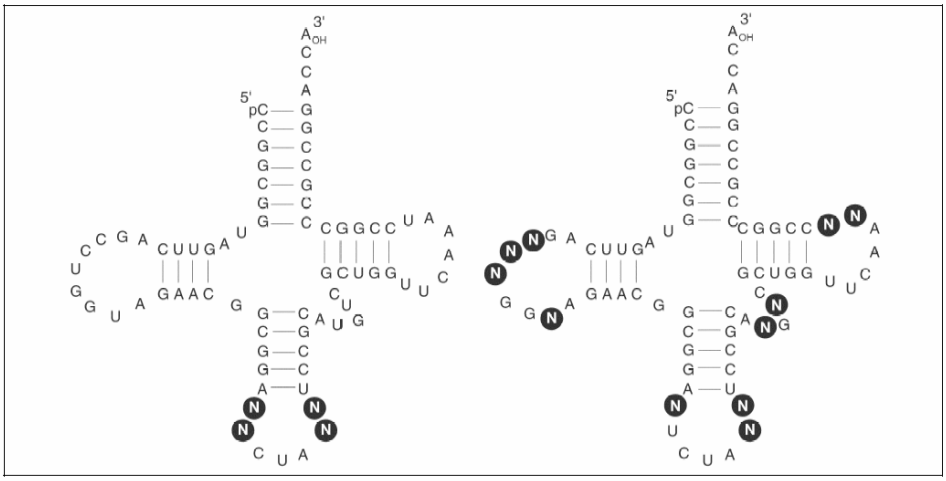


Figure 2: Structures of yeast tRNA^{Phe} (left) and *E. coli* tRNA^{Phe} (right). This figure is reprinted from Dudock, B. S.; Diperi, C.; Michael, M. S. *J. Biol. Chem.* **1970**, *245*, 2465-2468,[114] Copyright 1970 with permission from the American Society for Biochemistry and Molecular Biology.

(a)



(b)

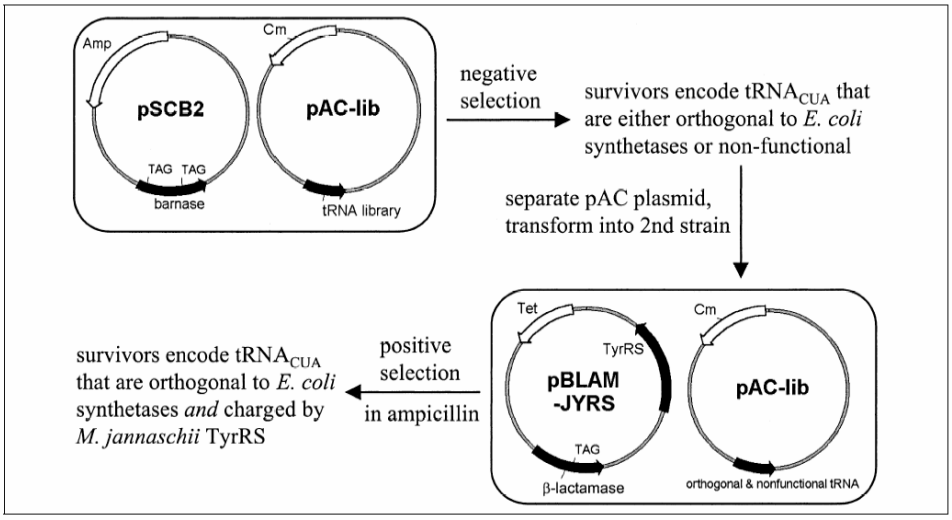
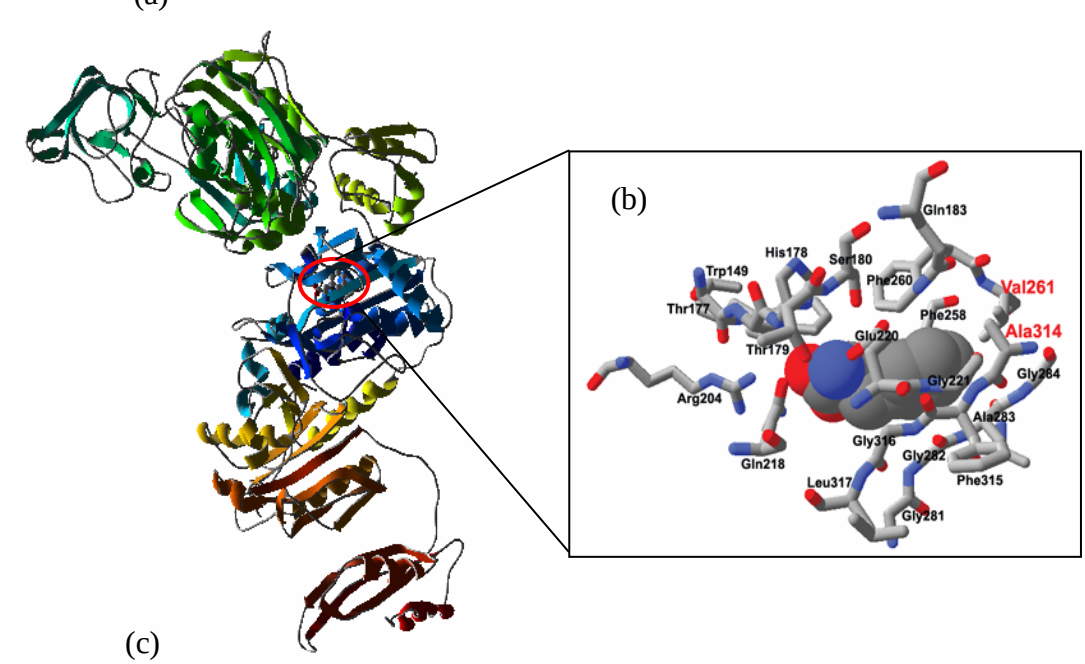


Figure 3: (a) Anticodon-loop tRNA library (left) and all-loop tRNA library (right) derived from *M. jannaschii* tRNA$^{Tyr}_{CUA}$ (N means all four nucleotides (A/U/G/C)). (b) A general selection scheme for suppressor tRNAs that are orthogonal to *the E. coli* endogenous systems and charged efficiently by a cognate synthetase. These figures are reprinted from from *Chem. Biol*., **8**, Wang, L.; Schultz, P.G., A general strategy for orthogonal tRNA, 883,[25] Copyright (2001) with permission from Elsevier.

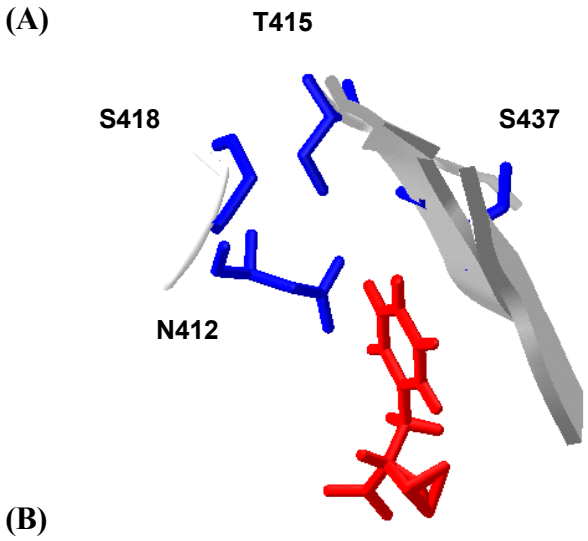
(a)



(c)

T. thermophilus	257YFPFVEP263	·····	313FAFGLGVERLAML325
E. coli	247YFPFTEP253	·····	293FAFGMGMERLTML305
S. cerevisiae	410YNPYTEP416	·····	456LGWGLSLERPTMI468
B. burgdorferi	441YFPFTEP447	·····	484IAWGIGIDRMALM496
B. subtilis	251FFPFTEP257	·····	308FAFGMGVERIAML320
A. aeolicus	247YFPFTEP253	·····	304FAFGMGVERLAML316
H. pylori	240FFPFTEP246	·····	295FAFGMGIERLAML307
H. influenzae	249YFPFTEP255	·····	295FAVGMGVERLTML307
C. trachomatis	252YFPFVEP258	·····	308YALGMGIERLAML320
M. genitalium	249HFPFTEP255	·····	306IAAGIGIERLAML318
H. sapiens	409YNPYTEP415	·····	455IAWGLSLERPTMI467
C. muridarum	251YFPFVEP257	·····	307YALGMGIERLAML319
C. pneumoniae	249YFPFVEP255	·····	305YAVGMGIERLAML317
D. radiodurans	242YYPFVEP248	·····	301FAFGLGLERIAML313
C. jejuni	241FFPFTEP247	·····	296YAFGLGVERFAML308
P. abyssi	420YYPFTEP426	·····	463IAWGIGIDRLAMF475
R. prowazekii	254FFPFTEP260	·····	301FAFGLGVERFAML313
N. meningitidis	249FFPFTEP255	·····	295FAFGIGLDRFAML307
U. urealyticum	243FFPFTEP249	·····	299MAFGVGIDRIAML311
T. pallidum	476YFPFTEP482	·····	519MAWGLGVDRMALL531
T. maritima	246FFPFTEP252	·····	292YAFGMGVERIAML304

Figure 4: (a) Crystal structure of the *T. thermophilus* PheRS (*t*PheRS, pdb 1B70) in ribbon model. Only ($\alpha\beta$) monomeric portion of structure is shown. (b) Active sites of *t*PheRS. Only residues within 6 Å of substrate Phe are shown. Substrate is shown as space-filling model, while residues surrounding the substrate are shown in stick model. (c) Sequence alignment of PheRS from 21 different organisms. Only residues flanking equivalent residues of V261 and A314 in *T. thermophilus* are represented. The shaded residues are the conserved residues equivalent to V261 and A314 from *T. thermophilus*. Sequences adjacent to V261 and A314 are also highly conserved. The sequences are obtained from aminoacyl-tRNA synthetase database (http://rose.man.poznan.pl/aars/). These figures were adapted from Wang's thesis (2003).[41]


(A)
T415
S418
S437
N412
(B)

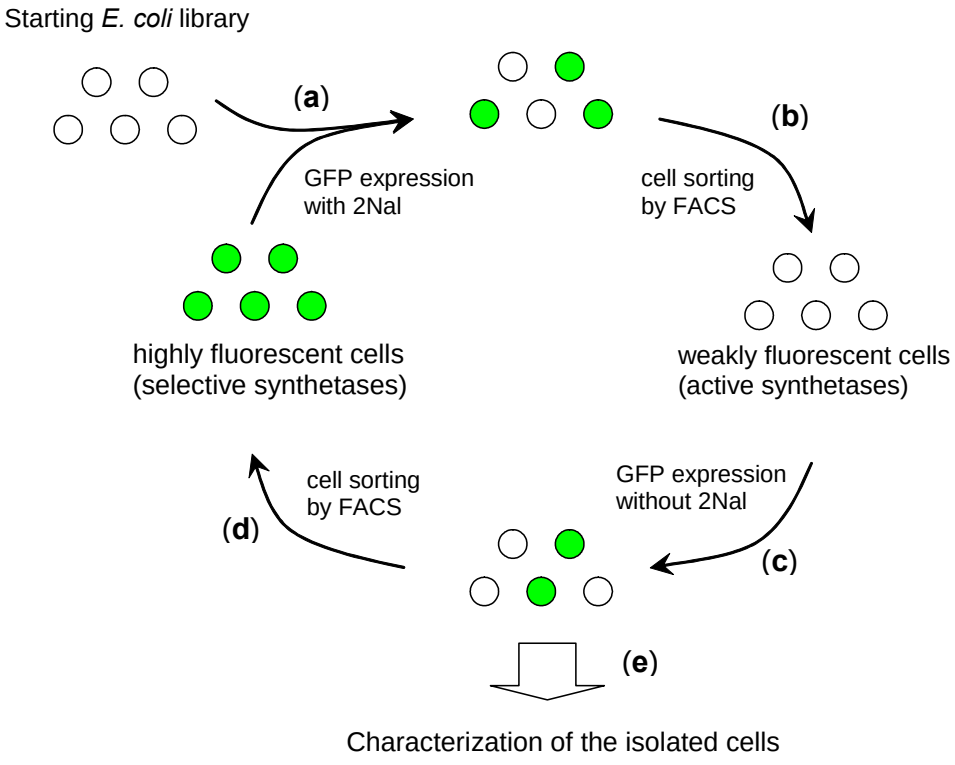

Starting E. coli library
(a)
GFP expression
with 2Nal
(b)
cell sorting
by FACS
highly fluorescent cells
(selective synthetases)
weakly fluorescent cells
(active synthetases)
cell sorting
by FACS
(d)
GFP expression
without 2Nal
(c)
(e)
Characterization of the isolated cells

Figure 5: (**A**) 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. (**B**) 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**).

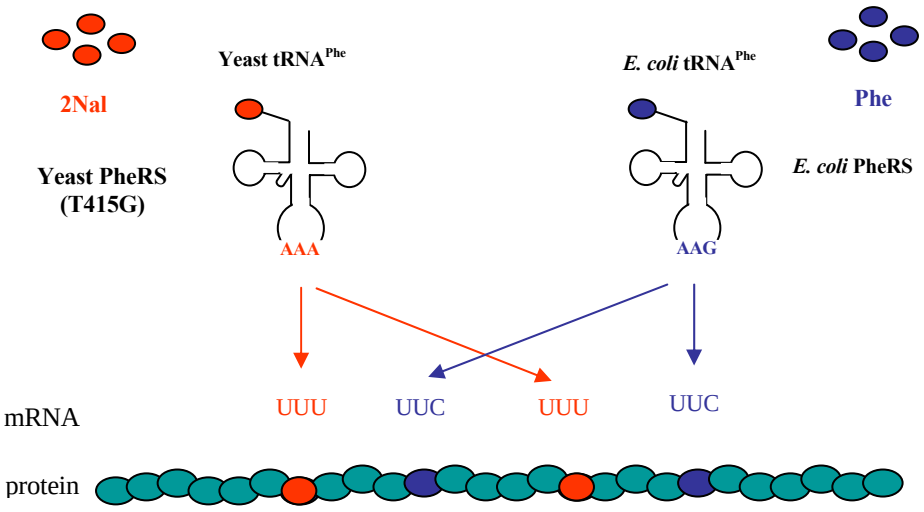


Figure 6: A schematic diagram describing the concept of breaking the degeneracy of the genetic code.

Chapter 2

Site-Specific Incorporation of Phenylalanine Analogs into Proteins in Vivo by an Engineered Yeast Phenylalanyl-tRNA Synthetase

The work in this chapter was performed in collaboration with Pin Wang, Soojin Son, and Yi Tang.

Abstract

Aminoacyl-tRNA synthetases (aaRSs) catalyze the aminoacylation reaction to establish the rules of the genetic code. Precise manipulation of synthetase activity can alter aminoacylation specificity to attach non-canonical amino acids to the cognate transfer RNAs (tRNA). The codon-anticodon interaction between messenger RNA (mRNA) and tRNA determines which amino acid is delivered into a growing polypeptide chain. Thus introduction of nonnatural amino acids into proteins in vivo relies heavily on manipulation of amino acid specificity of the aaRS. In this report, we describe the generation and characterization of a mutant phenylalanyl-tRNA synthetase (PheRS) from *Saccharomyces cerevisiae* with a point mutation (T415G) in the α-subunit of the enzyme. The rationale for the mutation is to allow binding of bulkier amino acid substrates. The promiscuous substrate specificity of this mutant was extensively explored by ATP-PP_i exchange assays in vitro. A broad activation profile toward many nonnatural amino acids was observed. A phenylalanine auxotrophic *E. coli* strain transformed with this mutant synthetase and cognate suppressor tRNA enable the assignment of an amber nonsense codon to the amino acid tryptophan (Trp) or to the nonnatural amino acid 3-(2-naphthyl)alanine. Expression strains engineered as phenylalanine, tryptophan double auxotrophs or phenylalanine, tryptophan and lysine triple auxotrophs outfitted with this pair of mutant synthetase and tRNA makes possible the efficient incorporation of *p*-bromophenyl-, *p*-iodophenyl-, and *p*-azidophenylalanine into recombinant proteins. Therefore, this synthetase and its cognate tRNA could serve as "21st" pair for site-specific incorporation of novel amino acid into proteins in vivo.

Introduction

Aminoacyl-tRNA synthetases (aaRSs) insure the high fidelity of transforming genetic code sequences into biologically functional proteins through a two-step aminoacylation reaction.[1-6] In the first step, the cognate amino acid is activated by aaRS in presence of ATP to form the amino acid adenylate; subsequently aaRS catalyzes the esterification reaction to join the amino acid to the 2'- or 3'-OH of the terminal ribonucleotide of its cognate tRNA. Through the aminoacylation reaction, the rules of the genetic code are settled; the strict correlation between triple nucleotide code and amino acid is established with the assistance of the aaRS.[2,7,8] Manipulation of such a reaction could alter the genetic code to allow incorporation of novel amino acid into proteins in vivo.[9,10]

With depletion of the intracellular pool of one natural amino acid and use of the corresponding bacterial auxotrophic cells, the rates of aminoacylation can be perturbed to enable charging of the nonnatural amino acid supplemented in the culture medium as a surrogate into cognate tRNA.[9] Through subsequent translation by ribosomal machinery, the genetic code is re-assigned and the nonnatural amino acid is incorporated into proteins. We have successfully implemented such a method to incorporate alkenes,[11] alkynes,[12] alkyl azide[13] and fluorinated side chains[14] into proteins. After enhancing the cellular aminoacylation reactivity by over-expression of wild-type aaRS in the host, we found that some sluggish amino acid analogs could also be introduced into proteins.[14-16] The method described above requires the cellular wild-type aaRSs to recognize these novel amino acids. Engineering new synthetase activity is almost inevitable when one wants to incorporate amino acids that are not recognized by the wild-type enzymes. To date, we and others have

shown that the re-design of the synthetic site of synthetase[17-19] and the attenuation of the editing function of synthetase[20,21] could be two complementary strategies to further expand our ability to introduce nonnatural amino acid into proteins.

We have shown that this multi-site incorporation method can utilize altered sets of 20 amino acids to design and engineer proteins and protein-like macromolecules.[9] By manipulation of both synthetase and tRNA in the host, one can also accomplish site-specific introduction of a single copy of a novel amino acid into proteins in vivo.[22-31] This method is derived from an in vitro approach to nonnatural amino acid incorporation through nonsense (stop codon) suppression, in which a stop codon (amber codon) was suppressed by a suppressor tRNA that had been chemically misacylated with the amino acid analog of interest.[32-36] Such a chemical method to alter the aminoacylation in vitro suffers from the technical difficulty and the intrinsic low yield of protein production, which limits its application. In 1998, Furter modified the cellular aminoacylation reaction by imparting a "twenty-first pair" comprising a yeast suppressor tRNA and a yeast phenylalanyl-tRNA synthetase (*y*PheRS).[26] This approach allowed site-specific incorporation of L-*p*-fluorophenylalanine in vivo in response to an amber codon. More recently Schultz and colleagues have devised powerful selection methods to find useful mutant forms of tyrosyl-tRNA synthetase from archaebacterium *Methanococcus jannaschii.*[10,37] By introduction of such a mutant and a cognate suppressor, many nonnatural amino acids have been incorporated into proteins in vivo.[10,22-25,28-31,38-46]

In this report, we present a rationally designed variant of *y*PheRS with new amino acid specificity, allowing incorporation of a variety of phenylalanine analogs (**2-4, 7,** and **8**) into proteins site-specifically. The variant of yPheRS contains a single T415G mutation at

the active site of the synthetase, which was shown to be very effective to open up the binding pocket. We performed a study on the amino acid activation kinetics of this mutant and found that this mutant can activate several interesting and useful aromatic amino acids (**1-8**). We also observed that this variant activates tryptophan (Trp, **6**) 82-fold faster than cognate substrate phenylalanine (Phe, **1**). A phenylalanine auxotrophic bacterial strain transformed with such a mutant and cognate suppressor tRNA can assign analog **7** to the amber codon with high fidelity (>90%). Such a transformed strain also allows the incorporation of **2-4** into proteins at the amber site with misincorporation of Trp. Further refinements of such a system by constructing a Phe/Trp/Lys triple auxotrophic strain enable us to incorporate analog **2-4** into the desired amber site with fidelity of 86%, 82%, and 90%, respectively.

Materials and Methods

Materials. Amino acids **1, 5** and **6** were obtained from Sigma (St. Louis, MO). Amino acids **2-4,** and **7** were purchased from Chem-Impex (Wood Dale, IL). Amino acid **8** was purchased from Biosynth International (Naperville, IL). [^{32}P]-labeled sodium pyrophosphate was purchased from NEN Life Science (Boston, MA).

Plasmid Construction for Synthetase Expression. The yPheRS gene was amplified from template plasmid pUC-ASab2[26,47] encoding the α- and β-subunits of PheRS gene with a 14-base pair (bp) intergenic sequence containing a translational reinitiation site upstream of the ATG start codon of the β-subunit gene. The following primers were used for PCR: 5'-CGA TTT TCA CAC AGG ATC CAG ACC ATG AAT CTA G-3' (primer 1 with

restriction site *Bam*HI) and 5'-GAC GGC CAG TGA ATT CGA GCT CGG TAC-3' (primer 2 with restriction site *Kpn*I). The resulting DNA was introduced into the *BamH*I and *Kpn*I sites of pQE32 to give pQE32-yFRS. The mutant yPheRS was generated by using the 4-primer mutagenesis method. Two complementary oligonucleotides, designated as primer 3 (5'-CTA CCT ACA ATC CTT ACG GCG AGC CAT CAA TGG AAA TC-3') for the forward primer and primer 4 (5'-GAT TTC CAT TGA TGG CTC GCC GTA AGG ATT GTA GGT AG-3') for the reverse primer, were synthesized to carry the specific mutation at position 415 of the α-subunit of *y*PheRS. In the initial two reactions, primer 1/primer 4 and primer 2/primer 3 were added into individual tubes and two DNA fragments were generated from these two PCRs. With the mixture of two reaction products and additional outside primers, a 3400 bp fragment of DNA was obtained and purified. This DNA was subjected to digestion by *Bam*HI and *Kpn*I and inserted into pQE32 to yield pQE32-T415G. The cloned enzymes contain the N-terminal sequence MRGSHHHHHHGIQTMNLE to facilitate protein purification. The entire yPheRS gene was sequenced for each of these constructs. Proofreading polymerase *Pfx* (Invitrogen) was used for all PCR amplication described in this work.

Synthetase Expression and Purification. The plasmids pQE32-yFRS and pQE32-T415G were individually transformed into *E. coli* strain BLR (Novagen) to form expression strains BLR(pQE32-yFRS) and BLR(pQE32-T415G). Synthetase expression was conducted in LB media. At an OD of 0.6, expression of the wild-type and mutant forms of yPheRS was induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4 hours. Cells were harvested and proteins were purified over a nickel-nitrilotriacetic

acid affinity column under native conditions according to the manufacturer's protocol (Qiagen). Imidazole in the elution buffer was removed by desalting column (Bio-rad) and proteins were eluted into a buffer containing 50 mM Tris-HCl (pH=7.5), 1 mM DTT. Aliquots of proteins were stored at -80^{o}C with 50% of glycerol.

Amino Acid Activation Assay. The amino acid-dependent ATP-PP_i exchange reaction was used to evaluate the activation of amino acid analogs by *y*PheRS. This assay was performed in 200 μL of reaction buffer containing 50 mM HEPES (pH=7.6), 20 mM $MgCl_2$, 1 mM DTT, 2 mM ATP and 2 mM [^{32}P]-PP_i with specific activity of 0.2-0.5 TBq/mol. Depending on the rate of activation of the analog by the synthetase, the amino acid concentration varied from 10 μM to 5 mM and the enzyme concentration varied from 10 nM to 100 nM. Aliquots (20 μL) were removed from the reaction solution at various time points and quenched into 500 μL of buffer solution containing 200 mM $NaPP_i$, 7% w/v $HClO_4$ and 3% w/v activated charcoal. The charcoal was spun down and washed twice with 500 μL of 10 mM $NaPP_i$ and 0.5% $HClO_4$ solution. The [^{32}P]-labeled ATP absorbed into the charcoal was quantified via liquid scintillation methods. The specificity constants were calculated by nonlinear regression fit of the data to a Michaelis-Menten model.

Plasmid and Strain Construction for in Vivo Incorporation Assays. pQE16 (Qiagen), chosen as the expression plasmid, encodes the marker protein murine dihydrofolate reducase (mDHFR) with a C-terminal hexa-histidine tag under control of a bacteriophage T5 promoter and t_0 terminator. A Quick-change mutagenesis kit was used to place an

amber codon (UAG) at position 38 of mDHFR (mDHFR (38Am)) (with two complementary oligonucletides (5'-CCG CTC AGG AAC GAG TAG AAGTAC TTC CAA AGA ATG-3'; 5'-CAT TCT TTG GAA GTA CTT CTA CTC GTT CCT GAG CGG-3') to afford pQE16am. The mutant *y*PheRS (T415G) gene was amplified from pQE32-T415G and a constitutive *tac* promoter with an abolished *lac* repressor binding site was added upstream of the start codon of this gene.[26] The entire expression cassette was inserted into the *Pvu*II site of pQE16 to yield pQE16am-T415G. The mutant yeast suppressor tRNA (*mu*tRNA^{Phe}(CUA)) was constitutively expressed under control of an *lpp* promoter. The expression cassette for *mu*tRNA^{Phe}(CUA) was inserted into pREP4 to form pREP4-tRNA as described by Furter.[26] A Phe auxotrophic strain AFK10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi *T2*^{R} *pheA*18), constructed in this laboratory,[26] was used as the expression strain. A Phe/Trp double auxotrophic strain AFW (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi *T2*^{R} *pheA*18, *trpB*114) and a Phe/Trp/Lys triple auxotrophic strain AFWK (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi *T2*^{R} *pheA*18, *trpB*114, *lysA*) were prepared by P1 phage-mediated transduction with *trpB::Tn10* and *lysA::Tn10* transposons.

Analog Incorporation Assay in Vivo. The auxotrophic strains AF, AFW, and AFWK were transformed with the plasmids pQE16am-T415G and pREP4-tRNA to afford expression strains AF [pQE16am-T415G/pREP4-tRNA] and AWF [pQE16am-T415G/pREP4-tRNA] respectively. Expression of mDHFR (38Am) was investigated in 20 mL cultures. The *E. coli* expression strains were grown in M9 minimal medium supplemented with glucose, thiamin, $MgSO_4$, $CaCl_2$, 20 amino acids (at 25 mg/L), and antibiotics (kanamycin and ampicillin). When cells reached an OD_{600} of 1.0, they were

sedimented by centrifugation, washed twice with cold 0.9% NaCl, and shifted to supplemented M9 medium containing 17 amino acids (at 25 mg/L), 3 mM analog of interest, and the indicated concentrations of Phe, Trp and Lys. Protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (1 mM). After 4 hours, cells were pelleted and the protein was purified by passage through a Ni-NTA spin column according to the supplier's instructions (Qiagen).

Compositional Analysis of Mutant mDHFR. Mutant mDHFR (38Am) was purified under denaturing conditions and isolated in elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH = 4.5). For trypsin digestion, 10 μL of the solution was diluted into 90 μL of 75 mM $(NH_4)_2CO_3$ and the pH was adjusted to 8. 2 μL of modified trypsin (Promega, 0.2 μg/μL) was added. The sample was incubated at room temperature overnight. For digestion by endoproteinase Lys-C, 10 μL of the solution was diluted into 90 μL of 25 mM Tris-HCl, pH = 8 and 1 mM EDTA and pH was adjusted to optimal working pH 8 for Lys-C. 2 μL of Lys-C (Calbiochem, 0.2 μg/μL) was added and the reaction was incubated at 37 ^oC for 10 hours. The digestion reaction was stopped by addition of 2 μL trifluoroacetic acid (TFA). The solution was subjected to $ZipTip_{C18}$ (Millipore) purification and the digested peptides were eluted with 3 μL of 50% CH_3CN, 0.1% TFA, of which 1 μL was used for matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis with μ-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid as the matrix. The analysis was performed on a PerSeptive Biosystems (Framingham,

Massachusetts) Voyager DE PRO MALDI-TOF mass spectrometer operating in linear and positive ion modes.

LC-MS/MS analysis of protease-digested peptides was conducted on Finnigan LCQ ion trap mass spectrometer with HPLC pump and ESI probe. 3 μL of digested peptide solution eluted from ZipTip$_{C18}$ was diluted with 20 μL of distilled water and injected into the HPLC pump. Peptides were separated by Magic C18 column (Michrom, 5μL, 200Å, 0.3x150mm) and eluted at a flow rate of 30 μL/min using a linear gradient of 0-60% of acetonitrile in 45 min. The column eluent was transferred to the electrospray source and tandem mass sequencing was carried out by fragmentation of the precursor ion with m/z corresponding to the protease-digested fragment (residues 26-39; NGDLPWPPLRNE***Z***K) including an amber site at position 38 of mutant mDHFR (38Am).

Results and Discussion

Rationales for Engineering New Synthetase Specificity. Engineering the new synthetase specificity toward amino acid analogs is one of the critical steps for the successful implementation of the general strategy to incorporate nonnatural amino acids into proteins in vivo.[29] We have been especially interested in exploring the possibilities of incorporating substituted phenylalanine (Figure 2) such as *p*-bromophenylalanine (2), *p*-iodophenylalanine (3), *p*-azidophenylalanine (4), and 3-(2-naphthyl)alanine (7). These non-proteinogenic amino acids either carry chemical functionality not present in proteins or are characterized by enhanced hydrophobic interactions. The crystal structure of *Thermus thermophilus* PheRS (tPheRS) complexed with phenylalanine is available[48] and the

sequence identity (*vs Saccharomyces cerevisiae*) in the active site region is about 40% (Figure 1). We used this crystal structure as a starting point for the design of a mutant yeast PheRS that charges the nonnatural amino acids of interest. Safro and colleagues identified two residues, V261 and A314, which constitute a back-wall structure to hold substrate phenylalanine and play an important role to hinder binding of amino acids larger than Phe into the active site of *T. thermophilus* PheRS.[49] In fact, an *E. coli* PheRS variant prepared by mutation of the corresponding residues (T251G and A314G, Figure 1), has been shown to exhibit relaxed substrate specificity.[17] In *S. cerevisiae* PheRS, the corresponding residues are T415 and G460. We reasoned that mutation from T415 to G might enlarge the active site to accommodate larger phenylalanine analogs.

Synthetase Expression and Purification. The wild-type yPheRS gene was obtained by PCR amplification of vector pUC-ASab2 encoding the α- and β-subunits of yeast PheRS (our assignment of α- and β-subunits are opposed to what was assigned by Sanni; we followed the tradition that the catalytic subunit of PheRS is designated the α-subunit.[47,48,50]). A 14-bp intergenic sequence obtained from upstream region of the *E. coli* pheST gene was inserted in the middle of the α- and β-subunits of yeast PheRS. Four-primer mutagenesis was employed to create the mutant T415G. The synthetases were over-expressed by *E. coli* BLR strains under control of the bacteriophage T5 promoter. An N-terminal hexa-histidine tag fused to the α-subunit of yPheRS facilitated purification of the enzymes under native conditions. Both *E. coli* and yeast synthetases are $\alpha_2\beta_2$ hetero-tetramers and their molecular weights for each subunit are rather different (α(ePheRS)=37

kDa; α(yPheRS)=57 kDa; β(ePheRS)=87 kDa; β(yPheRS)=67.5). SDS-PAGE analysis of purified yPheRS indicated no observable contamination by *E. coli* PheRS. The intensities of the bands for the α- and β-subunits are not stoichiometric (the α-subunit is stronger than the β-subunit); Sanni and colleagues reported a similar observation and suggested that this result stems from proteolysis of the β-subunit and from one β-fragment co-migrating with the α-subunit.[47] So a fraction of the purified yPheRS used in our experiments must be inactive. Our aim is to investigate the altered substrate specificity from the mutation (T415G for α-subunit) and compare the ability of the mutant enzyme to activate different amino acid analogs (**2**-**8**) in vitro; as long as we paralleled the measurement and used the same batch of the enzyme, it is still possible to deduce the specificity constants for the different analogs and make comparisons.

Amino Acid Specificity of Mutant Yeast PheRS. The activities of the yPheRS variant with Phe analogs (**2**-**8**) were examined via amino acid-dependent ATP-PP_i exchange assay. Kinetic parameters for amino acid **1**-**8** were shown in Table 1. Our measured value of K_m for **1** by the wild-type yeast PheRS matched previously reported values,[51] although the obtained k_{cat} was lower than that reported,[51] and this difference might come from the different buffer conditions and different methods to determine the enzyme concentration. The efficiency k_{cat}/K_m for **1** by the mutant T415G is reduced approximately 40-fold compared to **1** by the wild-type enzyme. The loss of catalytic activity is apparent in both the increasing value of K_m and decreasing value of k_{cat}, which suggests that the T415G mutation significantly impairs both the binding and catalytic events for **1**. However, this mutant exhibited higher activity for analogs **2**-**4** and **6**-**8**, with values of k_{cat}/K_m ranging

from 8- to 58-fold higher than that for Phe (**1**). It appears that the mutation effectively opens up the active site of the synthetase so that T415G favors binding and catalysis of bulkier substrates. Canonical amino acid Trp (**6**) showed especially high activity, as manifested by 82-fold higher specificity constant than that of **1**.

Site-Specific Incorporation of Phe Analogs into mDHFR (38Am) in Vivo Using a Phenylalanine Auxotrophic Strain. To test the utility of the engineered yPheRS to incorporate aromatic amino acid analogs into proteins in vivo at programmed sites, a constitutive expression system for the mutant synthetase was constructed. An amber stop codon was placed at residue 38 of the marker protein mDHFR (38Am), which was equipped with a C-terminal hexa-histidine tag followed by an ochre stop codon (Figure 3). The levels of full length protein production were examined by SDS-PAGE analysis. Without over-expression of suppressor mutRNA^{Phe}(CUA), we did not obtain detectable levels of full length mDHFR (38Am).[52] As a positive control, we co-expressed wild-type yPheRS with mutRNA^{Phe}(CUA), and efficient suppression by phenylalanine was observed.[52] When mutant yPheRS was co-expressed with mutRNA^{Phe}(CUA) in the presence of analogs **2**-**8** (3 mM), full length mDHFR (38Am) was formed.[52] Overall protein yields varied depending on the analogs used and were ca. 2-10 mg/L, estimated by SDS-PAGE. The same strain yield of mDHFR without amber codon in minimal medium was approximately 25 mg/L.

It should be noted that the observed pattern of protein expression by SDS-PAGE does not necessarily mean that the analogs of interest are inserted at the amber codon position. Proteolytic peptide analysis was performed by MALDI-MS and liquid

chromatography tandem mass spectrometry for purified protein samples. Our protocol for digestion by endoproteinase Lys-C yielded two peptides in the mass range 1,600-1,800 Da; these fragments were assigned to residues 26-39 (Peptide A) and 40-53 (Peptide B) of mDHFR (38Am) (Figure 3). Peptide A includes an amber codon at the 38th position; Peptide B contains a single phenylalanine residue at position 41. In a control experiment, in which mDHFR (38Am) was expressed in medium supplemented with **1** (3 mM) (Figure 4a), the mass of Peptide A indicates that **1** was incorporated into the amber site. Our in vitro assay shows that T415G mutant synthetase activates **6** 82-fold faster than **1**; while our in vivo translational experiment reveals that **1** is favored over **6** to enter amber site, despite the fact that Phe auxotrophic strain can still synthesize **6**. In the medium supplemented with **7** (3 mM) and **1** (0.03 mM), we unambiguously confirmed incorporation of **7** into the amber site as shown in Figure 4b; Peptide A was shifted up in mass by 50 Da, consistent with the increased mass of **7** relative to **1**; and Peptide B was unchanged in mass, indicating that the Phe codon site is decoded as **1**. Liquid chromatography tandem mass spectrometry was employed to confirm the sequence of Peptide A. The precursor ion at m/z 867.7 Da, which corresponds to the doubly charged ion of Peptide A, was separated and fragmented with an ion trap mass spectrometer. As shown in Figure 5, the sequence obtained from the spectrum clearly reveals site-specific incorporation of **7** into mDHFR at position 38; we could find no indication of any other amino acid at the amber site after carefully checking all possible peptides with masses that could be associated with Peptide A, suggesting that the fidelity of incorporation of **7** is better than 95%. The kinetics of activation of **7** and **1** by mutant yPheRS showed that the efficiency (k_{cat}/K_m) for **7** is 1,500±100 ($M^{-1}s^{-1}$) (Table 1), 8-fold larger than that for **1**. Therefore, when the ratio of **7** to **1** in the culture medium is

high (100 to 1), $tRNA^{Phe}$(CUA) should be charged predominantly with **7**, consistent with our observation of high percentage incorporation of **7** into proteins at amber sites.

Similar analyses show incorporation of **2**-**4** into mDHFR (38Am). As a representative, incorporation of **2** into mDHFR (38Am) was shown (Figure 4c). We observed competitive misincorporation of Trp (**6**) and lysine (Lys, **8**) at the amber site. Incorporation of Trp is not surprising in view of the in vitro data. The rates of activation of **2**-**4** by T415G are 10-33% of that of **6** (Table 1). Therefore, **6** is able to compete significantly with **2**-**4** for the attachement into $mutRNA^{Phe}$(CUA) and insertion into the amber site. Lys misincorporation at the amber codon is likely due to mischarging of Lys to $mutRNA^{Phe}$(CUA) by *E. coli* lysyl-tRNA synthetase (eLysRS), because $mutRNA^{Phe}$(CUA) has U36 and A73, which are the main identity elements for eLysRS .[53,54] No observation of misincorporation of **6** or **8** with **7** (Figure 4b) could be due to the better recognition of **7** by mutant synthetase than that of **2**-**4** (Table 1) and the better charging of **7** into $mutRNA^{Phe}$(CUA) by mutant synthetase than that of eLysRS.

Site-Specific Incorporation of Phe Analogs into mDHFR in Vivo Using a Phe/Trp Double Auxotrophic Strain. To test our hypothesis of intracellular tryptophan inhibiting the incorporation of analogs (**2**-**4**), we decided to construct a Phe/Trp double auxotrophic strain so that the cellular concentration of Trp could be controlled. This strain, designated AFW, was transformed with mutant yPheRS, mDHFR (38Am) and $mutRNA^{Phe}$(CUA) genes to form a new expression system. In order to confirm our ability to control the cellular pool of Trp in this new strain, we performed a comparison experiment where we set one high concentration of Trp (3 mM) and one low concentration of Trp (0.03 mM) in

M9 minimal medium to examine the effect on site-specific incorporation of **2** (3 mM) in response to an amber stop codon. Proteolytic cleavage of mDHFR (38Am) by trypsin generated two peptides in the mass range 1550-1800 Da, which consistently appeared in the mass spectra and could be assigned to residues 26-39 (Peptide A) and 85-98 (Peptide C) of mDHFR (Figure 4, Figure 5). Peptide C contains a single phenylalanine residue at position 95. MALDI-MS analysis revealed that Trp concentration has a significant effect on incorporation of **2**. As shown in Figure 6a, **2** is incorporated into mDHFR along with Trp; the ratio of MALDI signal intensities suggests that **2** and **6** are incorporated at comparable rates at the amber site, although signal intensity and the peptide concentration are not strictly correlated. In contrast, as we lowered the Trp concentration to 0.03 mM, the signal for Trp incorporation greatly decreased; the amber position was mostly assigned by analog **2** (Figure 6b). Since the biosynthesis of Trp in the auxotrophic strain is impeded, cellular concentration of Trp is dependent on the amino acid transport from medium to cytoplasm. Our previous work showed that the cellular concentration of **2** is 14-17 times higher than that of Phe in the Phe auxotrophic strain cultured in media supplemented with 2 mM of **2** and 0.12 mM of Phe.[19] Compared with Trp, we believe that the 10-fold lower amino acid activation of **2** by mutant yPheRS (Table 1) can be compensated by high concentration of **2** inside the cells; mutRNA^{Phe}(CUA) is charged mostly with **2**. However, we observed that Lys was still misincorporated at amber codon (Figure 6b).

Site-Specific Incorporation of Phe Analogs into mDHFR in Vivo Using a Phe/Trp/Lys Triple Auxotrophic Strain. Based on our hypothesis that Lys was mischarged into mutRNA^{Phe}(CUA) by eLysRS, we decided to control the intracellular concentration of Lys

by constructing a Phe/Trp/Lys triple auxotrophic strain (AFWK). When a high concentration of Lys (0.2 mM) was supplemented into the medium, misincorporation of Lys was detected (Figure 6b). However, when low concentration of Lys (0.01 mM) was supplemented into the medium, the signal for the peptide containing Lys at amber codon is no longer detected (Figure 6c). In conclusion, under the condition that high concentrations of **2** and low concentrations of Phe, Trp and Lys were supplemented into the medium, incorporation of **2** is predominant at the programmed amber stop codon in mDHFR (38Am) with 86% fidelity. Further experiments on analog **3** (Figure 6d) and analog **4** manifest a similar pattern of the site-directed analog incorporation with 82% and 90% fidelity, respectively.[52]

Conclusions

The amino acid binding pocket of yeast PheRS can be engineered very simply to accommodate a variety of aromatic side chain functionalities. Cells outfitted with mutant yeast synthetase (T415G) and suppressor tRNA can site-specifically incorporate several amino acid analogs (**2**-**4**, and **7**) into proteins in response to the UAG codon. Coupled with recent reports from several laboratories,[22-25,27,29-31,38-42,44,45,55-57] these results begin to demonstrate that the import of heterologous synthetase pairs should constitute a general strategy for site-specific incorporation of amino acids into proteins in vivo. Chemical and physical properties of these new side chains can enlarge our ability to manipulate proteins and provide us additional tools to study and design new functions of biomacromolecules.[29,33,41,58-60] For instance, bromine (**2**) and iodine (**3**) once introduced into proteins can facilitate phasing of crystallographic data and therefore be useful in X-ray

diffraction studies of protein structure. Photoactivity of aryl azide (**4**) can be used for intramolecular crosslinking and protein immobilization.[61] Hydrophobic nature of naphthalene (**7**) can be applied to design specific ligand-receptor recognition.[62] Chemical utilities of these side chains equip us with new means to site-specifically modify proteins, most of which are orthogonal to natural amino acid side chains in the proteins. Proteins with aryl-halide side chains can be subjected to palladium-mediated coupling reactions to attach molecules containing the ethynyl group.[63] Azide chains in the proteins can be applied to staudinger ligation to tether molecules associated with the triarylphosphine group.[13] Recently, there is a growing interesting in possible bioconjugation of proteins using copper(I)-catalyzed azide-alkyne cycloaddition reactions.[64-66] Our group is currently doing these investigations using recombinant proteins biosynthesized in vivo in the engineered bacterial hosts described in this report.

Acknowledgments

We would like to thank Dr. Mona Shahgholi for her assistance with tryptic digest mass spectrometry. We are grateful to Dr. Aaron J. Link for helpful discussions on P1 phage-mediated transduction.

References

1. Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881-889.
2. de Pouplana, L. R.; Schimmel, P. *Cell. Mol. Life Sci.* **2000**, *57*, 865-870.
3. Francklyn, C.; Perona, J. J.; Puetz, J.; Hou, Y. M. *RNA* **2002**, *8*, 1363-1372.
4. Ibba, M.; Soll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617-650.

5. Schimmel, P. *Annu. Rev. Biochem.* **1987**, *56*, 125-158.

6. Soll, D.; RajBhandary, U. L., Structure, Biosynthesis, and Function. In *tRNA*, ASM Press: Washington, D. C., 1995.

7. Nangle, L. A.; Lagardt, V. D.; Dodring, V.; Schimmel, P. *J. Biol. Chem.* **2002**, *277*, 45729-45733.

8. Tamura, K.; Schimmel, P. *Nat. Biotechnol.* **2002**, *20*, 669-670.

9. van Hest, J. C. M.; Tirrell, D. A. *Chem. Commun.* **2001**, 1897-1904.

10. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1-11.

11. van Hest, J. C. M.; Tirrell, D. A. *FEBS Lett.* **1998**, *428*, 68-70.

12. van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 1282-1288.

13. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 19-24.

14. Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 11089-11090.

15. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148-2152.

16. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

17. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

18. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235-237.

19. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37-40.

20. Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501-504.

21. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

22. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11020-11024.

23. 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.

24. Deiters, A.; Geierstanger, B. H.; Schultz, P. G. *ChemBioChem* **2005**, *6*, 55-58.

25. Deiters, A.; Groff, D.; Ryu, Y. H.; Xie, J. M.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2006**, *45*, 2728-2731.

26. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

27. Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

28. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 935-939.

29. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

30. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

31. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

32. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.

33. Dougherty, D. A. *Curr. Opin. Chem. Biol.* **2000**, *4*, 645-652.

34. Gilmore, M. A.; Steward, L. E.; Chamberlin, A. R., Incorporation of noncoded amino acids by in vitro protein biosynthesis. In *Implementation and Redesign of Catalytic Function in Biopolymers*, 1999; Vol. 202, pp 77-99.

35. Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468-1473.

36. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

37. Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883-890.

38. Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598-9608.

39. Chin, J. W.; Schultz, P. G. *ChemBioChem* **2002**, *3*, 1135-1137.

40. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782-11783.

41. Schultz, K. C.; Supekova, L.; Ryu, Y. H.; Xie, J. M.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984-13985.

42. Wang, J. Y.; Xie, J. M.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 8738-8739.

43. Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.

44. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 56-61.

45. Xie, J. M.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548-554.

46. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22*, 1297-1301.

47. Sanni, A.; Walter, P.; Ebel, J. P.; Fasiolo, F. *Nucleic Acids Res.* **1990**, *18*, 2087-2092.

48. Reshetnikova, L.; Moor, N.; Lavrik, O.; Vassylyev, D. G. *J. Mol. Biol.* **1999**, *287*, 555-568.

49. Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1534-44.

50. Fayat, G.; Mayaux, J. F.; Sacerdot, C.; Fromant, M.; Springer, M.; Grunbergmanago, M.; Blanquet, S. *J. Mol. Biol.* **1983**, *171*, 239-261.

51. Gabius, H. J.; Vonderhaar, F.; Cramer, F. *Biochemistry* **1983**, *22*, 2331-2339.

52. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

53. McClain, W. H.; Foss, K.; Jenkins, R. A.; Schneider, J. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 9260-9264.

54. Tamura, K.; Himeno, H.; Asahara, H.; Hasegawa, T.; Shimizu, M. *Nucleic Acids Res.* **1992**, *20*, 2335-2339.

55. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964-967.

56. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306-14307.

57. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8882-8887.

58. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 2910-2914.

59. Liu, H. T.; Wang, L.; Brock, A.; Wong, C. H.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 1702-1703.

60. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9785-9789.

61. Fleming, S. A. *Tetrahedron* **1995**, *51*, 12479-12520.

62. Chan, D. C.; Chutkowski, C. T.; Kim, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 15613-15617.

63. Genet, J. P.; Savignac, M. *J. Organomet. Chem.* **1999**, *576*, 305-317.

64. Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021.

65. Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.

66. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602.

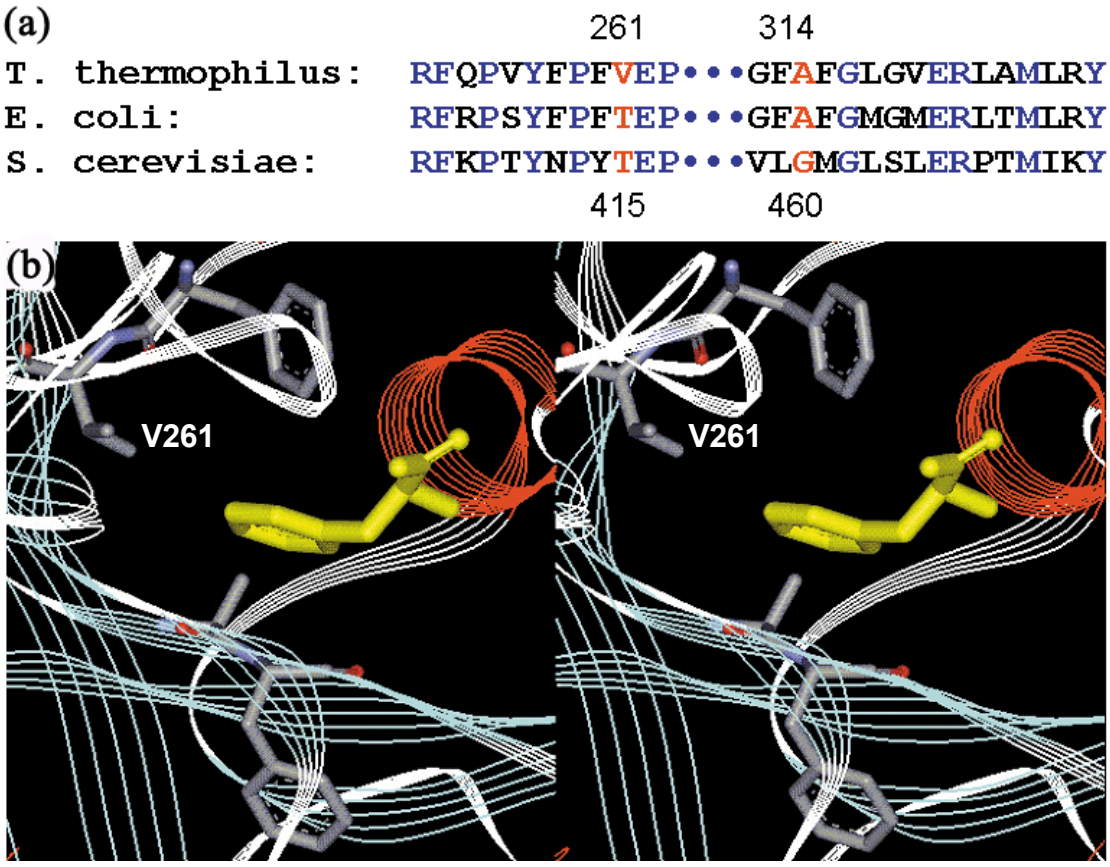


Figure 1: (a) Sequence alignment of PheRS variants from *Thermus thermophilus, Escherichia coli*, *Saccharomyces cerevisiae.* Only portions of sequences flanking the V261 and A314 positions in *T. thermophilus* are shown. The sequences adjacent to V261 and A314 are highly conserved. The equavilent residues in *E. coli* and *S. cerevisiae* that correspond to residues V261 and A314 in *T. thermophilus* are shown in red. (b) Stereoview of active site of PheRS from *T. thermophilus*. Substrate phenylalanine is shown in yellow.

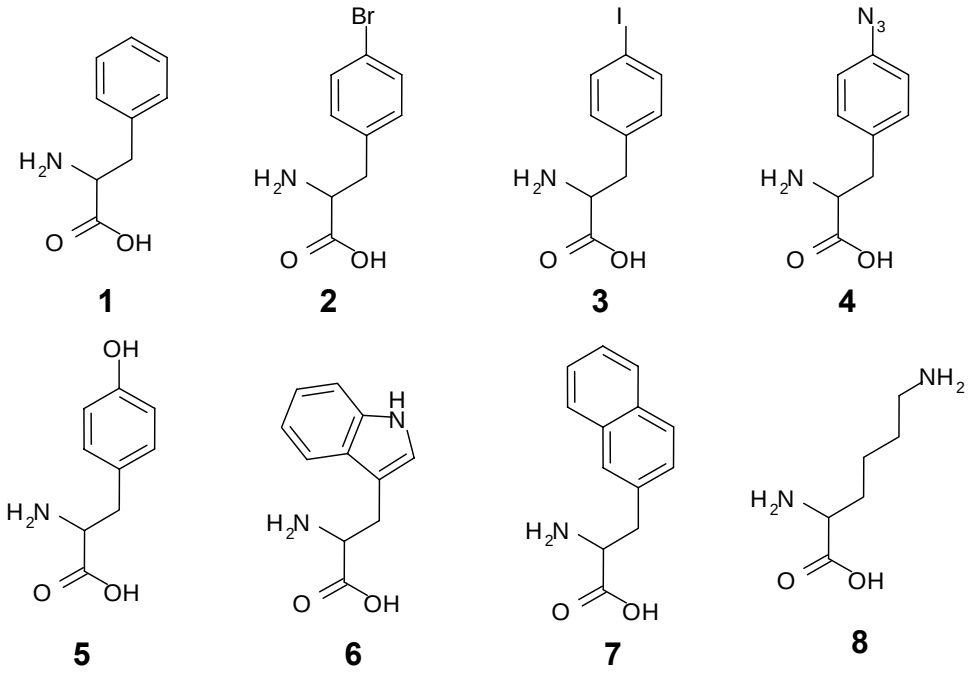


Figure 2: Amino acids involved in this study. (**1**) phenylalanine; (**2**) *p*-bromophenylalanine; (**3**) *p*-iodophenylalanine; (**4**) *p*-azidophenylalanine; (**5**) tyrosine; (**6**) trytophan; (**7**) 3-(2-naphthyl)alanine; and (**8**) L-lysine.

Table 1: Kinetic parameters for ATP-PPi exchange of amino acids (**1**-**7**) by the mutant yeast PheRS

Amino Acid	Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)*
1	T415G	264±42	0.05±0.002	180±30	1
2	T415G	22±3	0.03±0.001	1,500±200	8
3	T415G	12±2	0.05±0.001	4,400±800	24
4	T415G	11±3	0.05±0.002	4,600±1,200	25
5	T415G	757±149	0.04±0.003	48±10	1/4
6	T415G	20±5	0.30±0.006	15,000±4,000	82
7	T415G	27±2	0.04±0.001	1,600±100	8
1	Wild-type	68±20	0.52±0.093	7,600±2,700	41

*relative to k_{cat}/K_m for **1** by T415G

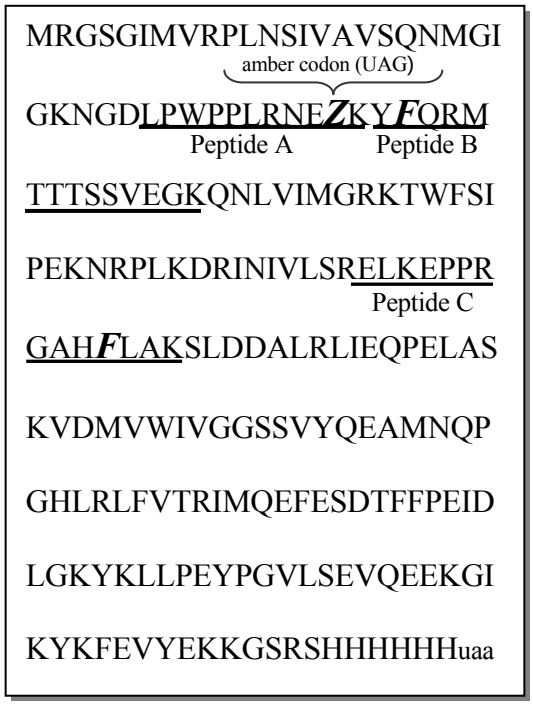


Figure 3: Amino acid sequence of marker protein mDHFR (38Am). Four proteolytic peptide fragments (Peptide A, Peptide B, and Peptide C) used for MALDI-MS and LC-MS/MS analyses are underscored. The positions with amber codon and ochre codon are labeled.

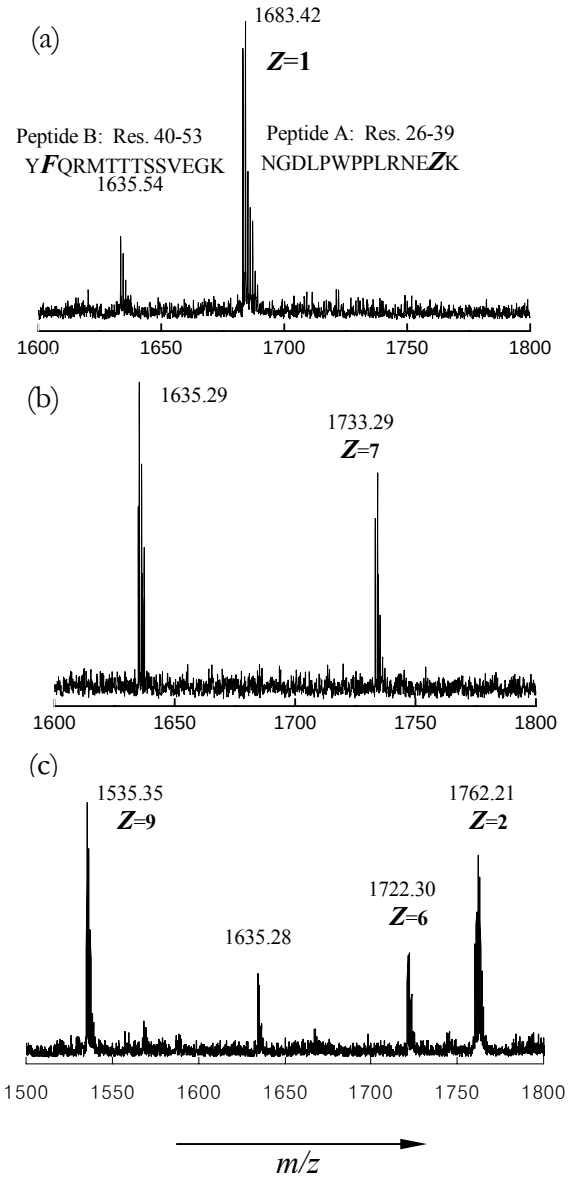

(a)
1683.42
Z=1
Peptide B: Res. 40-53
YFQRMTTTSSVEGK
1635.54
Peptide A: Res. 26-39
NGDLPWPPLRNEZK
1600
1650
1700
1750
1800
(b)
1635.29
1733.29
Z=7
1600
1650
1700
1750
1800
(c)
1535.35
Z=9
1762.21
Z=2
1722.30
Z=6
1635.28
1500
1550
1600
1650
1700
1750
1800
m/z

Figure 4: MALDI-MS of proteolytic peptide fragments derived from mDHFR expressed in media supplemented with (a) **1** (3 mM); (b) **7** (3 mM)/**1** (0.03 mM); (c) **2** (3 mM)/**1** (0.03). mDHFR (38Am) was expressed in phenylalanine auxotrophic strain (AF) transformed with the mutant yeast PheRS and cognate suppressor tRNA. No tryptophan is added, except that 1 mM of Trp is supplemented in (c). Protein was digested by endoproteinase Lys-C. Peptide B, containing one Phe codon, remains the same for all the experimental conditions, indicating normal Phe codons are assigned as Phe. Peptide A contains one amber codon; the residue responding to this codon is designated as ***Z***. In case of Lys misincorporation into Peptide A (c), C-terminal Lys was further cleaved to produce a shorter Peptide A (NGDLPWPPLRNEK).

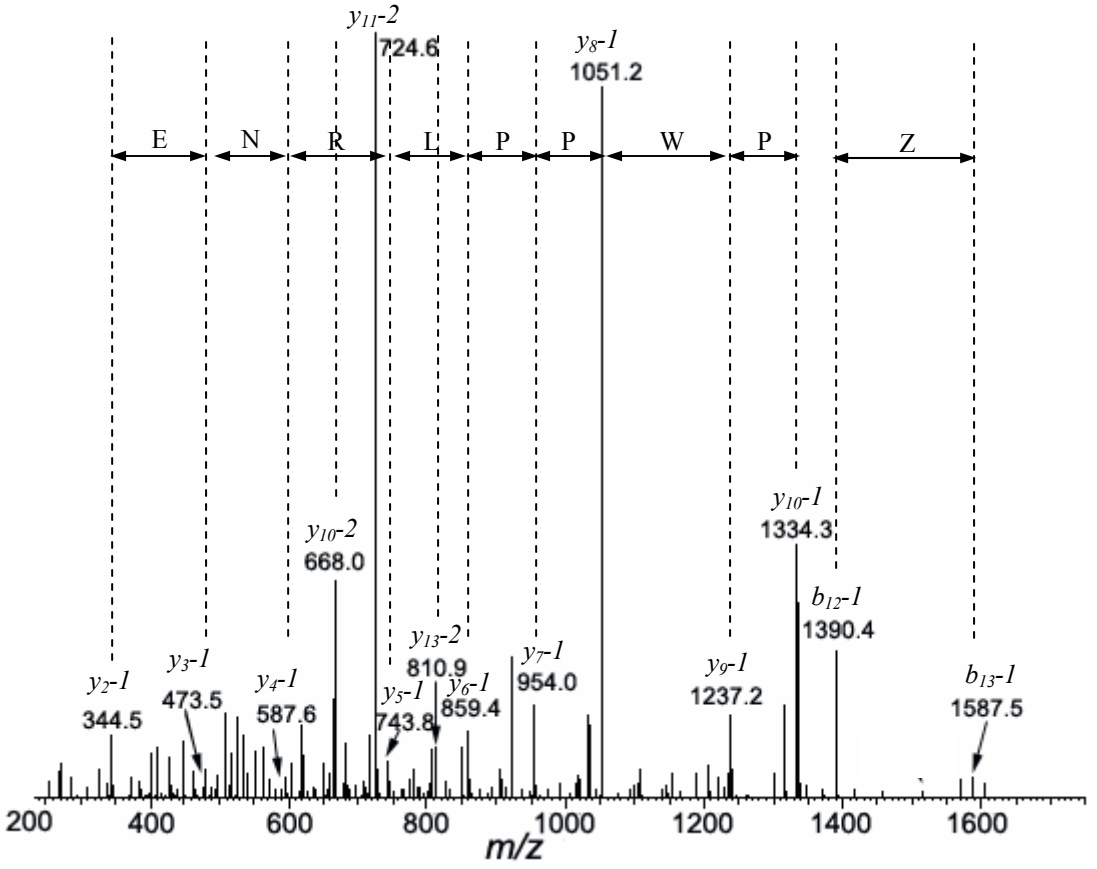


Figure 5: The tandem mass spectrum of Peptide A (NGDLPWPPLRNE***Z***K) derived from mDHFR (38Am) expressed in media supplemented with **7** (3 mM)/**1** (0.03 mM). mDHFR (38Am) was expressed in phenylalanine auxotrophic strain (AF) transformed with the mutant yeast PheRS and cognate suppressor tRNA. Partial sequence of PWPPLRNE and residue Z (corresponds to **7**) of Peptide A can be assigned from the annotated y and b ion series respectively.

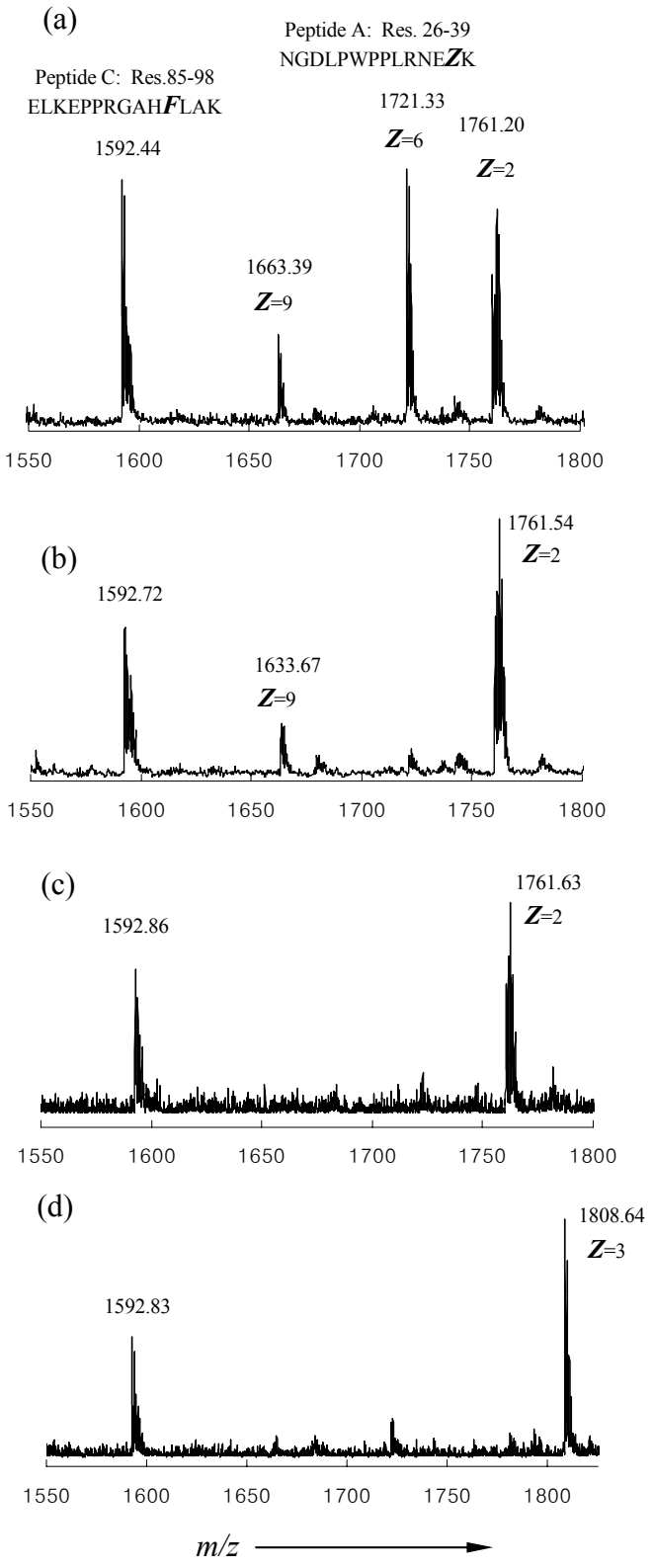

(a)
Peptide A: Res. 26-39
NGDLPWPPLRNEZK
Peptide C: Res.85-98
ELKEPPRGAHFLAK
1592.44
1663.39
Z=9
1721.33
Z=6
1761.20
Z=2
1550
1600
1650
1700
1750
1800
(b)
1592.72
1633.67
Z=9
1761.54
Z=2
(c)
1592.86
1761.63
Z=2
(d)
1592.83
1808.64
Z=3
m/z

Figure 6: MALDI-MS of proteolytic peptide fragments of mDHFR (38Am) expressed in media supplemented with (a) **2** (3 mM)/**6** (3 mM)/**1** (0.03 mM)/**8** (0.2 mM); (b) **2** (3 mM)/**6** (0.01 mM)/**1** (0.03 mM)/**8** (0.2 mM); (c) **2** (3 mM)/**6** (0.01 mM)/**1** (0.03 mM)/**8** (0.01 mM); (d) **3** (3 mM)/**6** (0.01 mM)/**1** (0.03 mM)/**8** (0.01 mM). mDHFR (38Am) was expressed in an phenylalanine, tryptophan and lysine triple auxotrophic strain (AFWK) transformed with the mutant yeast PheRS and cognate suppressor tRNA. The purified mDHFR (38Am) was digested by modified trypsin. Peptide C, containing one Phe codon, remains the same for all the experimental conditions, indicating normal Phe codons are assigned as Phe. Residue ***Z***, the amino acid incorporated in response to the amber codon, is assigned based on the mass units of Peptide A.

Chapter 3

Design of a Bacterial Host for Site-Specific Incorporation of *p*-Bromophenylalanine into Recombinant Proteins

Abstract

Introduction of a yeast suppressor tRNA ($ytRNA^{Phe}_{CUA}$) and a mutant yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) into an *E. coli* expression host allows in vivo incorporation of phenylalanine analogs into recombinant proteins in response to amber stop codons. However, high fidelity incorporation of non-natural amino acids is precluded in this system by mischarging of $ytRNA^{Phe}_{CUA}$ with tryptophan (Trp) and lysine (Lys). Here we show that $ytRNA^{Phe}_{CUA}$ and yPheRS can be re-designed to achieve high fidelity amber codon suppression through delivery of p-bromophenylalanine (pBrF). Two strategies were used to reduce misincorporation of Trp and Lys. First, Lys misincorporation was eliminated by disruption of a Watson-Crick base pair between nucleotides 30 and 40 in $ytRNA^{Phe}_{CUA}$. Loss of this base pair reduces mischarging by the *E. coli* lysyl-tRNA synthetase. Second, the binding site of yPheRS was re-designed to enhance specificity for pBrF. Specifically, we used the T415A variant, which exhibits five-fold higher activity toward pBrF as compared to Trp in ATP-PPi exchange assays. Combining mutant $ytRNA^{Phe}_{CUA}$ and yPheRS (T415A) allowed incorporation of pBrF into murine dihydrofolate reductase in response to an amber codon with at least 98% fidelity.

Introduction

Non-natural amino acids carrying a wide variety of novel functional groups have been incorporated into recombinant proteins in both prokaryotic and eukaryotic cells.[1-24] Although global replacement of natural amino acids with non-natural analogs is useful for many purposes,[1,16,25,26] there are situations in which single-site substitution by a non-natural amino acid is required. In such circumstances, a codon must be assigned uniquely

to the non-natural amino acid. Amber (UAG),[2,4,7-9,27-29] ochre (UAA),[30,31] opal (UGA) stop codons, [30,32] and four base codons[32,33] have been explored for this purpose. The amber codon has been used most widely,[2,8,9,27-29] because it is the least common stop codon in *E. coli* and because several naturally occurring suppressor tRNAs recognize it efficiently.[34,35] Use of the amber codon to encode non-natural amino acids requires outfitting the cell with an "orthogonal pair" comprising a suppressor tRNA and a cognate aminoacyl-tRNA synthetase (aaRS) that operate independently of the endogenous synthetase-tRNA pairs in *E. coli*. The site-specific incorporation of a non-natural amino acid into recombinant proteins via this strategy was reported by Furter in 1998,[9] and the Schultz laboratory has developed powerful selection methods to identify heterologous synthetases and tRNAs.[2,22,32,36,37] The tyrosyl pair derived from the archaebacterium *Methanococcus jannaschii* has been especially useful in this regard.[38]

In a complementary approach, rational modification and virtual screening methods have been used to design the amino acid binding pocket of the *E. coli* PheRS (ePheRS).[5,39] On the basis of the crystal structure of the PheRS (tPheRS) from *Thermus thermophilus*, Safro and colleagues proposed that Val 261 and Ala 314 in the amino acid binding pocket are critical in the discrimination of Phe from its amino acid competitors.[40] Sequence alignment indicates that Ala 314 in tPheRS corresponds to Ala 294 in ePheRS, and the Hennecke group showed that the substrate specificity of ePheRS can be relaxed by a point mutation at Ala 294. The A294G mutant was shown to enable incorporation of *para*-chlorophenylalanine into recombinant proteins.[41] A subsequent computational simulation, consistent with the Safro prediction, identified two cavity-forming mutations (T251G and A294G) in ePheRS binding pocket. These two mutations led to relaxed substrate specificity and efficient in vivo replacement of Phe by *para*-

acetylphenylalanine (pAcF).[5] Sequence alignment shows that Thr 415 in yPheRS is equivalent to Thr 251 in ePheRS. We therefore anticipated that the yPheRS (T415G) variant would activate a variety of Phe analogs.[42] However, when the T415G variant was co-transformed with ytRNA$^{Phe}_{CUA}$ into an *E. coli* host, substantial misincorporation of Trp and Lys was observed. Use of a triple auxotrophic *E. coli* host (Phe, Trp, and Lys) led to good incorporation (85-95%) of a variety of novel aromatic amino acids in response to amber codons,[32] but the limited specificity of this expression system prompted us to consider its refinement.

In this report, we describe the design of a bacterial host capable of high fidelity site-specific incorporation of *para*-bromophenylalanine (pBrF) in response to an amber codon. Introduction of aryl halides, such as pBrF[19,22] or *para*-iodophenylalanine (pIF)[13,22] into recombinant proteins allows site-specific modification via versatile palladium catalyzed cross-coupling reactions with terminal alkene or alkyne reaction partners.[43-48] In order to achieve high fidelity incorporation of pBrF, two different strategies were applied to reduce misincorporation of Trp and Lys, while providing good yields of recombinant protein. First, Lys misincorporation was eliminated by modification of the sequence of ytRNA$^{Phe}_{CUA}$ to reduce mischarging of ytRNA$^{Phe}_{CUA}$ by eLysRS. Second, the binding site of yPheRS was re-designed to enhance specificity for pBrF. Finally, the combined use of these two strategies provided high fidelity amber codon suppression through delivery of pBrF.

Materials and Methods

Materials. Unless otherwise noted, reagents were obtained from commercial suppliers and were used without further purification. Natural amino acids were from Sigma (St.

Louis, MO), and [^{3}H]-amino acids from Amersham Pharmacia Biotech (Piscataway, NJ). pBrF was obtained from Chem-Impex (Wood Dale, IL). Purified total yeast tRNA was purchased from Roche Biochemical (Indianapolis, IN). The nickel-nitrilotriacetic acid affinity column and repressor plasmid pREP4 were from Qiagen (Valencia, CA). *E. coli* strain BLR (RecA$^-$ derivative of BL21) was purchased from Novagen (Madison, WI).

Plasmid Construction for Synthetase Expression. Genes encoding the α- and β-subunits of yPheRS were amplified from template plasmid pUC-ASab2[9,49] and inserted between the *BamH*I and *Kpn*I sites of pQE32 to give pQE32-yPheRS. pQE32-T415G and pQE32-T415A were constructed from pQE32-yPheRS by PCR mutagenesis at position 415 of the α-subunit of yPheRS by use of a QuikChange mutagenesis kit (Stratagene). Two complementary oligonucleotides, 5'-CT ACC TAC AAT CCT TAC GCC GAG CCA TCA ATG GAA ATC-3' for the forward primer and 5'-GAT TTC CAT TGA TGG CTC GGC GTA AGG ATT GTA GGT AG-3' for the reverse primer, were used to introduce the T-to-A mutation at position 415 of the α-subunit of yPheRS. The entire yPheRS gene was verified by DNA sequencing for each of these constructs. Proofreading polymerase *Pfx* was used in all PCR described in this work. The *E. coli lysS* gene was amplified by PCR from template plasmid pXLysKS1,[48] by using two primers, 5'-GCA CTG ACC ATG GCT GAA CAA CAC GCA CAG-3' (which includes an NcoI restriction site) and 5'-GGA CTT CGG ATC CTT TCT GTG GGC GCA TCG C-3' (which carries a *BamH*I restriction site). The resulting DNA was inserted between the *Nco*I and *BamH*I sites of pQE60 to yield pQE60-eLysS. The cloned enzymes contained N-terminal or C-terminal hexa-histidine tags to facilitate protein purification.

Synthetase Expression and Purification. The plasmids pQE32-yPheRS, pQE32-T415G, pQE32-T415A, and pQE60-eLysS were individually co-transformed with a repressor plasmid pREP4 into *E. coli* strain BLR to form expression strains BLR (pQE32-yFRS), BLR (pQE32-T415G), BLR (pQE32-T415A), and BLR (pQE60-eLysS), respectively. Synthetase expression was conducted in 2xYT media with 100 μg/mL of ampicillin and 35 μg/mL of kanamycin. At an OD of 0.6, expression of each of the synthetase variants was induced with 1 mM IPTG. After four hours, cells were harvested and proteins were purified over a nickel-nitrilotriacetic acid affinity column under native conditions according to the manufacturer's protocol (Qiagen). The imidazole in the elution buffer was removed by a desalting column (Amersham Lifescience) and proteins were eluted into a buffer containing 50 mM Tris-HCl (pH=7.5), 1 mM DTT. Aliquots of proteins were stored in -80 ^{o}C in 50% glycerol. Concentrations of the yPheRS variants and eLysRS, the lysS gene product, were determined by UV absorbance at 280 nm using calculated extinction coefficients[50] of 99,060 and 29,420 cm^{-1} M^{-1} for yPheRS and eLysRS, respectively.

Amino Acid Activation Assay. The amino acid-dependent ATP-PP_i exchange reaction was used to determine the kinetics of activation of amino acid analogs by yPheRS. The reaction buffer consisted of 50 mM N-(2-hydroxethyl) piperazine-N'-(2-ethanesulfonic acid) (potassium-HEPES) (pH=7.6), 20 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 2 mM adenosine triphosphate (ATP) and 2 mM [^{32}P]-pyrophosphate (PPi) (NEN Life Science) with specific activity of 10-50 Ci/mol. The amino acid concentration varied from 10 μM to 2.5 mM and the enzyme concentration varied from 10 nM to 100 nM. Aliquots (20

μL) were removed from the reaction solution at various time points and quenched into 500 μL of buffer solution containing 200 mM $NaPP_i$, 7% w/v $HClO_4$ and 3% w/v activated charcoal. The charcoal was sedimented by centrifugation and washed twice with 500 μL of 10 mM $NaPP_i$ and 0.5% $HClO_4$ solution. The [^{32}P]-labeled ATP absorbed on the charcoal was quantified via liquid scintillation methods. The specificity constants were calculated by a nonlinear regression fit of the data to a Michaelis-Menten model.

Plasmid Construction for ytRNA^{Phe} Expression. The mutant yeast amber suppressor tRNA (ytRNA$^{Phe}_{CUA}$) was constitutively expressed under control of an lpp promoter. The expression cassette for ytRNA$^{Phe}_{CUA}$ was inserted into the repressor plasmid pREP4 to form pREP4_ytRNA^{Phe} as described by Furter.[9] A mutant yeast suppressor ytRNA$^{Phe}_{CUA}$_30U40G (ytRNA$^{Phe}_{CUA_UG}$) was constructed from ytRNA$^{Phe}_{CUA}$ by use of a QuikChange mutagenesis kit. Two complementary oligonucleotides, designated as primer UG_f (5'- GAA CAC AGG ACC TCC ACA TTT AGA GTA TGG CGC TCT CCC -3') for the forward primer and primer UG_r (5'- GGG AGA GCG CCA TAC TCT AAA TGT GGA GGT CCT GTG TTC -3') for the reverse primer, were used to introduce the desired mutations at positions 30 and 40 of the yeast suppressor tRNA. The resulting plasmid carrying the gene encoding ytRNA$^{Phe}_{CUA_UG}$ is designated pREP4_ytRNA^{Phe}_UG. In order to construct plasmids for in vitro transcription of ytRNA^{Phe}, the ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ genes were amplified from pREP4_ytRNA^{Phe} and pREP4_ytRNA^{Phe}_UG, respectively. At the end of the tRNA sequence, a *BstN*I site was inserted to allow linearization prior to transcription. A T7 promoter sequence was added for in vitro transcription of ytRNA^{Phe} by T7 RNA polymerase. The following primers were used for PCR: 5'-CTG GGT AAG CTT CGC TAA GGA TCT GCC CTG GTG

CGA ACT CTG-3’ (which includes restriction sites for *Hind*III and *BstN*I) and 5’-GAT TAC GGA TTC CTA ATA CGA CTC ACT ATA GCG GAC TTA GCT C-3’ (which carries an *EcoR*I restriction site and a T7 promoter sequence). The resulting DNA fragments were introduced between the *Hind*III and *EcoR*I sites of pUC18 to yield pUC18_ytRNA^{Phe}_CUA and pUC18_ytRNA^{Phe}_UG, respectively. In order to facilitate DNA manipulation, a *BstN*I site close to the T7 promoter sequence of pUC18_ytRNA^{Phe}_CUA was removed to increase the size of the DNA fragment containing the ytRNA$^{Phe}_{CUA}$ gene from 180 bp to 500 bp after *BstN*I digestion. Two complementary oligonucleotides, 5’-CGG AAG CAG AAA GTG TAA AGA GCG GGG TGC CTA ATG AGT G-3’ for the forward primer and 5’-CAC TCA TTA GGC ACC CCG CTC TTT ACA CTT TAT GCT TCC G-3’ for the reverse primer, were used to introduce this mutation.

In Vitro Transcription. Linearized DNAs were prepared by BstNI digestion of pUC18_ytRNA^{Phe}_CUA and pUC18_ytRNA^{Phe}_UG as described previously.[51] In vitro transcription of linearized DNA templates, and purification of transcripts were performed as described previously with minor alterations.[52] The in vitro transcription of linearized DNA to produce 76-mer tRNA transcripts was performed with the Ambion T7-MEGAshortscript kit. Transcripts were isolated by extraction with 25:24:1 phenol:$CHCl_3$:isoamyl alcohol. The organic layer was re-extracted with water and a 24:1 $CHCl_3$:isoamyl alcohol extraction was performed on the aqueous layers. The water layer was then mixed with an equal volume of isopropanol, precipitated overnight at -20°C, pelleted, dried, and re-dissolved in water. Unreacted nucleotides in the tRNA solution

were eliminated using CHROMA SPIN-30 DEPC-H_2O (BD Bioscience) spin columns. Concentrations of the transcripts were determined by UV absorbance at 260 nm.

Aminoacylation Assay. Aminoacylation of wild-type ytRNA$^{Phe}_{GAA}$ with Phe and Trp by yPheRS variants was performed as described earlier.[51] Aminoacylation reactions were carried out in buffer (100 μL) containing 30 mM HEPES (pH=7.45), 15 mM $MgCl_2$, 4 mM DTT, 25 mM KCl, and 2 mM ATP at 30 °C. Purified yeast total tRNA was used in the assay at a final concentration of 4 mg/mL (ytRNA$^{Phe}_{GAA}$ concentration approximately 2.24 μM). For aminoacylation with Phe, 13.3 μM [^{3}H]-Phe (5.3 Ci/mmol) and 80 nM yPheRS variants were used; for aminoacylation with Trp, 3.3 μM [^{3}H]-Trp (30.0 Ci/mmol) and 160 nM yPheRS variants were used. Aminoacylation of ytRNA^{Phe} transcripts reactions was carried out in buffer (100 μL) containing 100 mM potassium-HEPES (pH=7.4), 10 mM $MgCl_2$, 1 mM DTT, 0.2 mM EDTA, 2 mM ATP, and 4 units/mL yeast inorganic pyrophosphatase (Sigma) at 37°C for eLysRS. For aminoacylation with Lys, 4 μM of ytRNA^{Phe} transcript, 1.1 μM [^{3}H]-Lys (91 Ci/mmol) and 80 nM eLysRS were used. The tRNAs were annealed before use by heating to 85 ^oC for 4 min in annealing buffer (60 mM Tris, pH=7.8, 2 mM $MgCl_2$) followed by slow cooling to room temperature. Reactions were initiated by adding the enzyme and 10 μL aliquots were quenched by spotting on Whatman filter disks soaked with 5% trichloroacetic acid (TCA). The filters were washed for three 10 min periods in ice-cold 5% TCA, washed in ice-cold 95% ethanol, and counted via liquid scintillation methods.

Strain and Plasmid Construction for in Vivo Incorporation of pBrF. A Phe/Trp double auxotrophic strain (AFW) and a Phe/Trp/Lys triple auxotrophic strain (AFWK) were constructed from the Phe auxotrophic strain AF (K10, *Hfr*(Cavalli) *pheS*13rel-1 *tonA*22 *thi* $T2^R$ *pheA*18, *trpB*114)[9] by P1 phage-mediated transposon transduction as described previously.[42] pQE16 (Qiagen) was chosen as the expression plasmid; pQE16 bears a gene encoding the marker protein murine dihydrofolate reducase (mDHFR) outfitted with a C-terminal hexa-histidine tag under control of a bacteriophage T5 promoter and t_0 terminator. A QuikChange mutagenesis kit with two complementary oligonucleotides (5'-CCG CTC AGG AAC GAG TAG AAGTAC TTC CAA AGA ATG-3'; 5'-CAT TCT TTG GAA GTA CTT CTA CTC GTT CCT GAG CGG-3') was used to place an amber codon (TAG) at the 38^{th} position of mDHFR; the resulting plasmid was designated pQE16am. The mutant yPheRS genes T415G and T415A were amplified from pQE32-T415G and pQE32-T415A, respectively, and a constitutive tac promoter with an abolished lac repressor binding site was added upstream of the start codon of this gene.[9] The resulting expression cassettes were inserted into the PvuII site of pQE16 to yield pQE16am-T415G and pQE16am-T415A.

In Vivo Incorporation of pBrF. The auxotrophic strains AF, AFW, and AFWK were transformed with pQE16am and pREP4_ytRNA^{Phe} vectors to enable pBrF incorporation. mDHFR expression was investigated in 20 mL cultures. The *E. coli* expression strains were grown in M9 minimal medium supplemented with glucose, thiamin, $MgSO_4$, $CaCl_2$, 20 amino acids (at 25 μg/mL), and antibiotics (35 μg/mL of kanamycin and 100 μg/mL of ampicillin). When the cultures reached an OD_{600} of 0.8-1.0, cells were sedimented by centrifugation, washed twice with cold 0.9% NaCl, and shifted to expression media

supplemented with 17 amino acids (at 20 μg/mL), 6 mM of pBrF, and the indicated concentrations of Phe, Trp and Lys. Protein expression was induced by addition of 1 mM IPTG. After four hours, cells were pelleted by centrifugation, and the protein was purified on a Ni-NTA spin column under denaturing conditions according to the supplier's instructions (Qiagen). After purification, expression levels of mDHFR were determined by UV absorbance at 280 nm using a calculated extinction coefficient[50] of 24,750 cm^{-1} M^{-1}.

Quantitative Analysis of pBrF Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS). LC-MS and LC/MS/MS analyses of tryptic digests of mDHFR were conducted on a Finnigan LCQ ion trap mass spectrometer equipped with an HPLC pump and ESI probe. Mutant mDHFR was prepared in elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH=4.5). After concentration of the protein by ultracentrifugation (Millipore), 10 μL of the concentrate was diluted into 90 μL of 75 mM NH_4HCO_3 for trypsin digestion. Modified trypsin (1 μL, Promega, 0.2 μg/ μL) was added. The sample was incubated at 37 °C for 2 to 6 h and the reaction was stopped by addition of 12 μL of 5% trifluoroacetic acid (TFA) solution. Digested peptide solution was subjected to desalting on a C_{18} Vydac Microspin column (The Nest group) and eluted with 50 μL of 80% acetonitrile/20% 0.1% w/v formic acid. The peptide solution eluted from the Microspin column was dried, redissolved in 10% acetonitrile/90% 0.1% TFA solution, and injected into the HPLC. Peptides were se*para*ted on a Magic C_{18} column (Michrom, 300 Å, 0.3x150 mm) and eluted at a flow rate of 30 μL/min using a gradient of 10-95% of solvent A (90% acetonitrile/10% 0.1 M aqueous acetic acid solution) and solvent B

(2% acetonitrile/98% 0.1 M aqueous acetic acid solution) for 30 min. The column eluent was introduced to the electrospray source and sequencing was carried out by fragmentation of the precursor ion corresponding to the fragment bearing the residue at position 38 of mutant mDHFR.

Results and Discussion

Aminoacylation of ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ with Lys. When the yeast phenylalanine amber suppressor (ytRNA$^{Phe}_{CUA_WT}$) was co-expressed with wild-type yPheRS in Furter's *E. coli* expression strain, 60% of the amber codon sites were decoded as Lys.[9] Use of the G37A mutant of ytRNA$^{Phe}_{CUA_WT}$ (ytRNA$^{Phe}_{CUA}$) reduced the extent of Lys misincorporation to 5%.[9] Because insertion of Lys in response to the amber codon is likely a consequence of charging of ytRNA$^{Phe}_{CUA}$ by the *E. coli* lysyl-tRNA synthetase (eLysRS), we modified ytRNA$^{Phe}_{CUA}$ to eliminate cross-reactivity with eLysRS. We focused on the base pair between nucleotides 30 and 40, which is thought to enhance recognition of yeast isoleucine amber suppressor (ytRNA$^{Ile}_{CUA}$) by eLysRS.[53] Buttcher and coworkers showed that introduction of a wobble base pair between nucleotides 30 and 40 reduced charging of ytRNA$^{Ile}_{CUA}$ by eLysRS.[53] Similarly, we generated a mutant yeast phenylalanine amber suppressor (ytRNA$^{Phe}_{CUA_UG}$) containing the wobble base pair 30U-40G, and compared the rates of charging of ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ with Lys by eLysRS (Figure 1).[54] As expected, the rate of aminoacylation of ytRNA$^{Phe}_{CUA_UG}$ was about three-fold lower than that of ytRNA$^{Phe}_{CUA}$ (Figure 1); however, there was only about a 40% reduction in the rate of aminoacylation of ytRNAPhe$_{CUA_UG}$ with Phe as compared to that of ytRNA$^{Phe}_{CUA}$. Since ytRNA$^{Phe}_{CUA}$ variants are competitively

aminoacylated by either eLysRS or yPheRS, the three-fold reduction in the rate of aminoacylation by eLysRS was expected to reduce the level of Lys-charged ytRNA^{Phe}.

Occupancy of the Amber Site. Occupancy of the amber site in mDHFR-38Am was analyzed by LC-MS/MS (Figure 2). The results are summarized in Table 1. LC-MS/MS has been used successfully for quantitative analysis of protein modification.[43] Mutant mDHFR containing the amber codon at the 38th position (mDHFR_38Am) was expressed in *E. coli* strains (AF, AFW, and AFWK) co-transformed with pQE16_yPheRS (T415G) and either pREP4_ytRNA^{Phe}_CUA or pREP4_ytRNA^{Phe}_UG. We focused on Peptide 38 (residues 26-39; NGDLPWPPLRNEAmK) which contains the amber site. Peptide 38 variants containing Trp and pBrF are designated Peptides W38 and Z38, respectively. When Lys is incorporated in response to the amber codon, there are two consecutive lysines at the C-terminus of Peptide 38, and the C-terminal Lys can be further cleaved by trypsin. The fully cleaved (NGDLPWPPLRNEK) and partially cleaved (NGDLPWPPLRNEKK) variants of Peptide K38 were designated K38S (short) and K38L (long), respectively. Peptides K38S, K38L, W38 and Z38 were readily separated by liquid chromatography (Figure 2a), and the relative yields of the four variants were determined by comparing the integrated areas of the corresponding chromatographic signals. Furthermore, the fragment ion masses could be unambiguously assigned (data not shown), confirming the identity of the amino acid inserted in response to the amber codon. The amplitude of the most intense signal in the chromatograms in Figure 2 is ca. $1x10^8$ (instrument-dependent arbitrary units) compared to a noise level of ca. $5x10^5$, suggesting that misincorporation of natural amino acids can be analyzed at the 0.5% level.

Reduced Lys Misincorporation by ytRNA$^{Phe}_{CUA_UG}$. When mDHFR(38Am) was expressed in a triple auxotrophic *E. coli* host (AFWK) outfitted with ytRNA$^{Phe}_{CUA}$ and yPheRS (T415G) in medium supplemented with 0.03 mM Phe, 0.01 mM Trp, 1 mM Lys and 6 mM pBrF, Lys misincorporation was observed at a level of 7.6% (Figure 2a). The level of Lys misincorporation could be reduced to 2.7% by restricting the Lys concentration to 0.01 mM (Table 1). However, when ytRNA$^{Phe}_{CUA_UG}$ was co-transformed into *E. coli* strain AFWK with yPheRS (T415G), Lys misincorporation was not detected, even in media supplemented with 1.0 mM Lys (Figure 2b). Considering the detection limit of the ESI MS detector, this establishes an upper limit of 0.5% Lys misincorporation. Similar results were obtained with Phe auxotrophic strain AF outfitted with ytRNA$^{Phe}_{CUA_UG}$ (Figure 2e); the Lys auxotrophic strain is not required to eliminate Lys misincorporation.

Redesign of Phenylalanyl-tRNA Synthetase. Although use of ytRNA$^{Phe}_{CUA_UG}$ eliminated Lys at the amber site, ca. 10% Trp misincorporation was still detected (Figure 2b). Trp misincorporation could result from mischarging of ytRNA$^{Phe}_{CUA}$ with Trp either by yPheRS or by *E. coli* tryptophanyl-tRNA synthetase (eTrpRS). Because misincorporation of Trp was not detected in the absence of yPheRS (T415G), mischarging of ytRNA$^{Phe}_{CUA}$ by eTrpRS is probably negligible. However, the Trp-dependent ATP-PPi exchange rate of yPheRS (T415G) was comparable to that observed for pBrF (Table 2). Therefore, mischarging of ytRNA$^{Phe}_{CUA}$ with Trp by yPheRS (T415G) is the more likely explanation for Trp misincorporation at the amber codon. We speculated that yPheRS (T415A) might exclude Trp.

ATP-PPi Exchange and Aminoacylation of ytRNA^{Phe} Catalyzed by yPheRS Variants. The kinetics of amino acid-dependent ATP-PPi exchange catalyzed by three variants of yPheRS (wild-type, T415G, and T415A) were measured; the kinetic parameters are reported in Table 2. Wild-type yPheRS activates neither Trp nor pBrF, whereas the T415G variant shows ten-fold higher activity toward pBrF and Trp as compared to Phe. As expected, the Trp-dependent ATP-PPi exchange rate of yPheRS (T415A) is ca. thirty-fold lower than that of yPheRS (T415G). In contrast, the effects of the G415A mutation on the rates of activation of Phe and pBrF are much smaller (Table 2). Aminoacylation of ytRNA$^{Phe}_{CUA}$ with Phe was comparable for yPheRS (T415G) and for yPheRS (T415A) (Figure 3a); however, aminoacylation of ytRNA$^{Phe}_{CUA}$ with Trp by yPheRS (T415A) is much slower than that by yPheRS (T415G) (Figure 3b).

The improved selectivity of yPheRS (T415A) with respect to exclusion of Trp may be understood at least in part on the basis of the side-chain volumes of the amino acid substrates and the active-site residues of the synthetase. Replacement of Gly415 by Ala reduces the volume of the substrate-binding site by ca. 16 Å^3.[55] Occlusion of the binding site reduces the binding affinity of both Trp and pBrF, but because Trp is the larger of the two amino acids (by ca. 5 Å^3), it suffers the larger reduction in affinity. The G415A mutation increases K_m by a factor of ca. 17 for Trp as compared to ca. 7 for pBrF (Table 2).

High Fidelity Site-Specific Incorporation of pBrF. Use of expression hosts outfitted with yPheRS (T415A) reduced misincorporation of Trp at the amber site to levels below the limit of detection (Figure 2c-e). Our initial studies used the triple auxotrophic

expression host AFWK. However, construction of the triple auxotroph was time-consuming and the limited cellular pools of several of the natural amino acids reduced protein yield. In order to overcome these problems, a single (Phe) auxotrophic expression host (AF) outfitted with the yPheRS (T415A) and ytRNA$^{Phe}_{CUA_UG}$ was tested for pBrF incorporation. High fidelity incorporation of pBrF in response to the amber codon was observed (Figure 2e and Table 1), indicating that the triple auxotrophic expression host is not necessary. Furthermore, the AF expression host outfitted with yPheRS (T415A) and ytRNA$^{Phe}_{CUA_UG}$ yielded twice as much protein as the triple auxotroph (4.3 mg/L vs. 2.0 mg/L) (Table 1).

Conclusions

We describe the engineering of a bacterial expression host that allows high fidelity incorporation of pBrF into recombinant proteins in response to amber stop codons. Rational modification of the yeast amber suppressor tRNA eliminated cross-reactivity with the *E. coli* LysRS, while the substrate specificity of yPheRS was enhanced by a point mutation in the amino acid binding pocket. When the modified yeast amber suppressor tRNA and the mutant yPheRS were co-expressed in a Phe auxotrophic *E. coli* expression host, at least 98% of the amber sites in full-length recombinant mDHFR were occupied by pBrF. Aryl bromides at programmed positions should enable chemo-selective ligation of proteins with terminal alkene or alkyne reaction partners.[43]

Acknowledgments

We thank Dr. P. Plateau for providing the plasmid pXLysKS1 and Dr. M. Shahgholi for assistance with mass spectrometry. We are grateful to Dr. K. Shin, Dr. J.

Petersson, and Dr. D. Beene for help with in vitro transcription of tRNA. We thank I. C. Tanrikulu for help with analysis of the PheRS crystal structure and for review of this manuscript. We also thank Dr. A. J. Link and K. E. Beatty for discussion and valuable comments on this project. This work was supported by National Institutes of Health grant GM62523.

References

1. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 14150-14151.
2. Bose, M.; Groff, D.; Xie, J. M.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 388-389.
3. Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893-904.
4. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964-967.
5. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.
6. Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. *J. Am. Chem. Soc.* **2003**, *125*, 6616-6617.
7. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782-11783.
8. Deiters, A.; Geierstanger, B. H.; Schultz, P. G. *ChemBioChem* **2005**, *6*, 55-58.
9. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

10. Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* **2004**, *73*, 147-176.

11. Ibba, M.; Stathopoulos, C.; Soll, D. *Curr. Biol.* **2001**, *11*, R563-R565.

12. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

13. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235-237.

14. Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 7512-7513.

15. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.

16. Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164-11165.

17. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602.

18. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res.* **2002**, *30*, 4692-4699.

19. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37-40.

20. Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 11089-11090.

21. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

22. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

23. Xie, J. M.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548-554.

24. 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.

25. Lim, V. I. *J. Mol. Biol.* **1995**, *252*, 277-282.

26. Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem.Int. Ed.* **2001**, *40*, 1494-1496.

27. Cload, S. T.; Liu, D. R.; Froland, W. A.; Schultz, P. G. *Chem. Biol.* **1996**, *3*, 1033-1038.

28. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 2910-2914.

29. Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435-462.

30. Kohrer, C.; Sullivan, E. L.; RajBhandary, U. L. *Nucleic Acids Res.* **2004**, *32*, 6200-6211.

31. Kohrer, C.; Yoo, J. H.; Bennett, M.; Schaack, J.; RajBhandary, U. L. *Chem. Biol.* **2003**, *10*, 1095-1102.

32. Anderson, J. C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 9598-9608.

33. Hohsaka, T.; Muranaka, N.; Komiyama, C.; Matsui, K.; Takaura, S.; Abe, R.; Murakami, H.; Sisido, M. *FEBS Lett.* **2004**, *560*, 173-177.

34. Thorbjarnardottir, S.; Uemura, H.; Dingermann, T.; Rafnar, T.; Thorsteinsdottir, S.; Soll, D.; Eggertsson, G. *J. Bacteriol.* **1985**, *161*, 207-211.

35. Yoshimura, M.; Inokuchi, H.; Ozeki, H. *J. Mol. Biol.* **1984**, *177*, 627-644.

36. Liu, D. R.; Magliery, T. J.; Pasternak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10092-10097.

37. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

38. Cropp, T. A.; Schultz, P. G. *Trends Genet.* **2004**, *20*, 625-630.

39. Wang, P.; Vaidehi, N.; Tirrell, D. A.; Goddard, W. A. *J. Am. Chem. Soc.* **2002**, *124*, 14442-14449.

40. Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1534-44.

41. Ibba, M.; Kast, P.; Hennecke, H. *Biochemistry* **1994**, *33*, 7107-7112.

42. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

43. Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; Yokoyama, S. *ChemBioChem* **2006**, *7*, 134-139.

44. Nguefack, J. F.; Bolitt, V.; Sinou, D. *Tetrahedron Lett.* **1996**, *37*, 5527-5530.

45. Sakakibara, Y.; Ido, Y.; Sasaki, K.; Sakai, M.; Uchino, N. *Bull. Chem. Soc. Jpn.* **1993**, *66*, 2776-2778.

46. Schoenbe.A; Heck, R. F. *J. Org. Chem.* **1974**, *39*, 3327-3331.

47. Yoneyama, M.; Kakimoto, M.; Imai, Y. *Macromolecules* **1988**, *21*, 1908-1911.

48. Yoneyama, M.; Konishi, T.; Kakimoto, M. A.; Imai, Y. *Makromol. Chem., Rapid Commun.* **1990**, *11*, 381-386.

49. Sanni, A.; Walter, P.; Ebel, J. P.; Fasiolo, F. *Nucleic Acids Res.* **1990**, *18*, 2087-2092.

50. Gill, S. C.; Vonhippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

51. Sampson, J. R.; Uhlenbeck, O. C. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1033-1037.

52. Nowak, M. W.; Gallivan, J. P.; Silverman, S. K.; Labarca, C. G.; Dougherty, D. A.; Lester, H. A., In vivo incorporation of unnatural amino acids into ion channels in *Xenopus oocyte* expression system. In *Ion Channels, Pt B*, **1998**, Vol. 293, pp 504-529.

53. Buttcher, V.; Senger, B.; Schumacher, S.; Reinbolt, J.; Fasiolo, F. *Biochem. Biophys. Res. Commun.* **1994**, *200*, 370-377.

54. Nishikawa and coworkers used an alternative strategy to reduce misacylation of yeast amber suppressor tRNATyr by eLysRS. Fukunaga, J.; Yokogawa; T., Ohno; Nishikawa, K. J. Biochem. 2006, 139, 689-696. In.

55. Volumes were estimated as solvent-excluded volumes (1.4 Å probe) by using Chem3D Pro 5.0 (CambridgeSoft: Cambridge, MA)

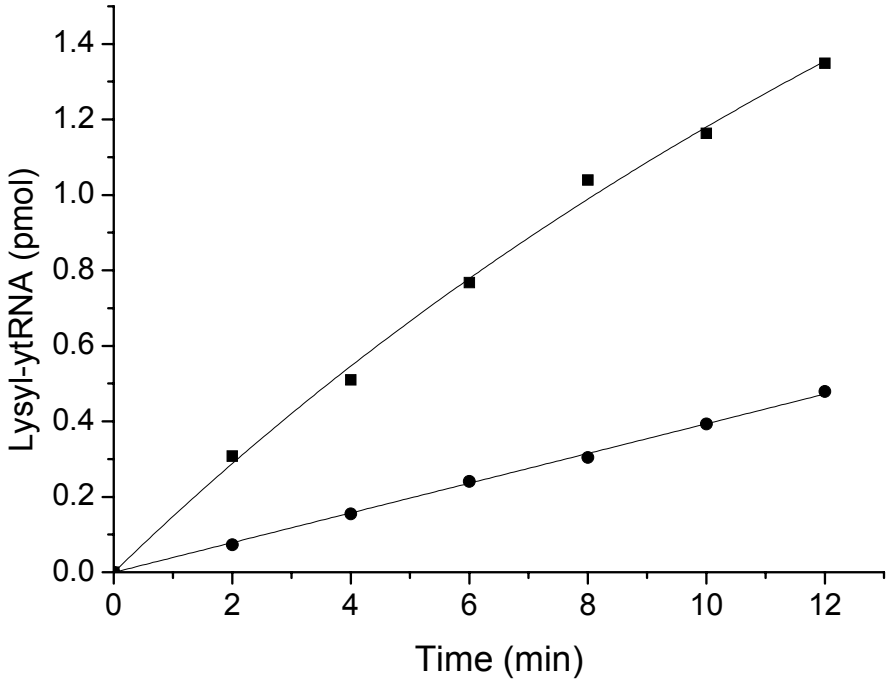


Figure 1: Aminoacylation of ytRNA$^{Phe}_{CUA}$ (■) and ytRNA$^{Phe}_{CUA_UG}$ (●) with Lys by eLysRS. The reaction was carried out at 37°C in mixtures containing 2 mM ATP, 1.1 μM [^{3}H]-Lys, 80 nM eLysRS, and 4 μM of ytRNA$^{Phe}_{CUA}$ or ytRNA$^{Phe}_{CUA_UG}$.

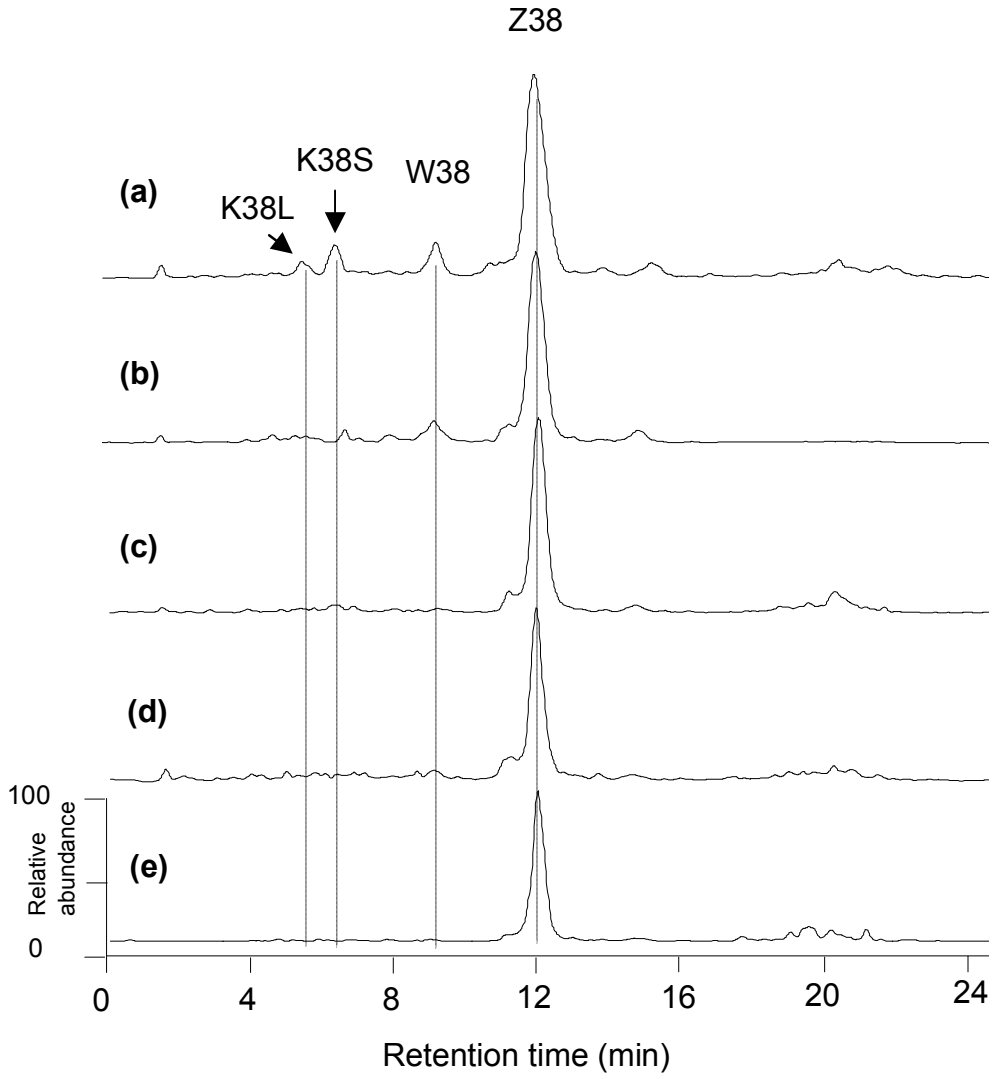
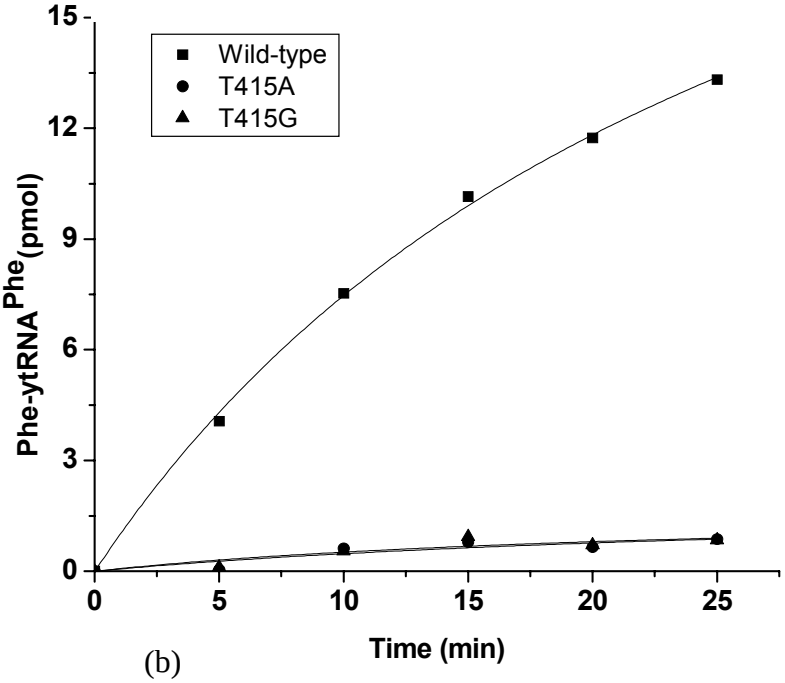

Z38
K38S
W38
K38L
(a)
(b)
(c)
(d)
(e)
100
Relative abundance
0
0
4
8
12
16
20
24
Retention time (min)

Figure 2: LC-MS chromatograms of tryptic digests of mDHFR. Peptide 38 (residues 26-39; NGDLPWPPLRNEAmK; Am indicates an amber codon) contains an amber codon at the 38th position. Peptide 38 variants containing Lys (Peptides K38S and K38L), Trp (Peptide W38), and pBrF (Peptide Z38) were se*para*ted and detected by MS. mDHFRs were synthesized in a triple auxotrophic expression host outfitted with: (a) ytRNA$^{Phe}_{CUA}$ and yPheRS (T415G); (b) ytRNA$^{Phe}_{CUA_UG}$ and yPheRS (T415G); (c) ytRNA$^{Phe}_{CUA}$ and yPheRS (T415A); (d) ytRNA$^{Phe}_{CUA_UG}$ and yPheRS (T415A); or in a single auxotrophic strain with ytRNA$^{Phe}_{CUA_UG}$ and yPheRS (T415A) (e). The expression media were supplemented with 6.0 mM pBrF, 0.01 mM Trp, 1.0 mM Lys, 0.03 mM Phe (a, b and c) or 0.01 mM Phe (d, e and f), and 25 μg/mL of 17 amino acids.

(a)


Wild-type
T415A
T415G
Phe-ytRNA^Phe(pmol)
Time (min)
0
3
6
9
12
15
0
5
10
15
20
25

(b)

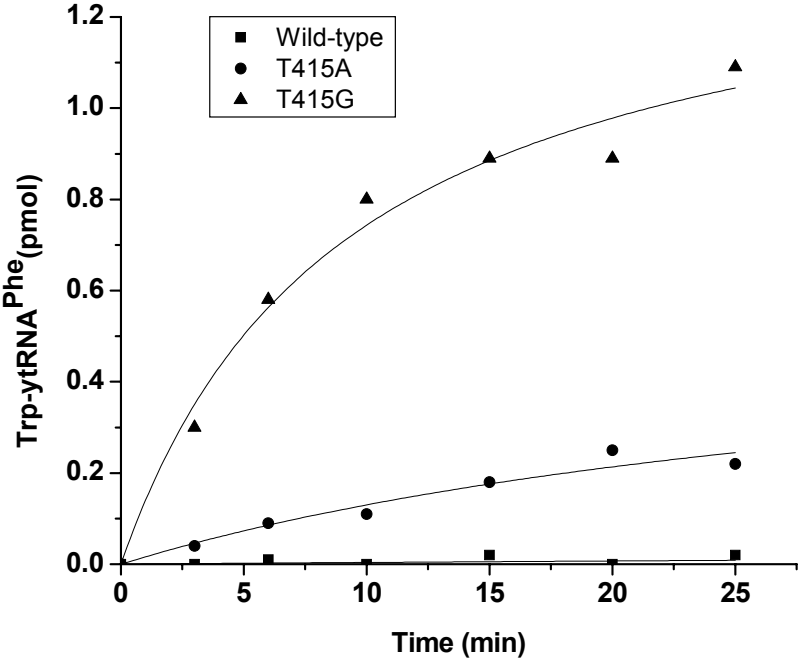

Wild-type
T415A
T415G
Trp-ytRNA^Phe(pmol)
Time (min)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
5
10
15
20
25

Figure 3: Charging of Phe (a) and Trp (b) by wild-type yPheRS (■), yPheRS (T415G) (▲) and yPheRS (T415A) (●). Reactions were carried out at 30°C in mixtures containing 2.24 μM $ytRNA^{Phe}_{GAA}$, 2 mM ATP, and either (a) 13.3 μm [^{3}H]-Phe and 80 nM yPheRS variants or (b) 3.3 μm [^{3}H]-Trp and 160 nM yPheRS variants in aminoacylation buffer.

Table 1: Occupancy of amber sites and expression yields.

Host strain	yPheRS	ytRNA^{Phe}	Occupancy of amber sites (%)			Yield (mg/L)[a]
			Lys	Trp	pBrF	
AFWK[b]	T415G	CUA[c]	7.6	9.5	83	1.4 ± 0.4
AFWK	T415G	CUA	2.7[d]	10.2	87	1.5 ±0.2
AFWK	T415G	CUA_UG[e]	ND[f]	10.4	90	1.4 ± 0.1
AFWK	T415A	CUA	2.7	ND	97	3.1 ± 0.3
AFWK	T415A	CUA_UG	ND	ND	> 98	2.0 ± 0.2
AF[g]	T415A	CUA_UG	ND	ND	> 98	4.3 ± 0.4

[a] Volumetric yields are given as mg of purified mDHFR_38Am per liter of culture volume.

[b] Triple (Phe, Lys, and Trp) auxotrophic *E. coli* strain.

[c,e] The ytRNA$^{Phe}_{CUA}$ and ytRNA$^{Phe}_{CUA_UG}$ are designated as CUA and CUA_UG, respectively.

[d] 0.01 mM of Lys was supplemented into the media rather than 1.0 mM.

[f] Not detected in either LC/MS or MALDI analyses.

[g] Single (Phe) auxotrophic *E. coli* strain.

Table 2: Kinetic parameters for ATP-PPi exchange of amino acids by the wild-type yPheRS, yPheRS (T415G), and yPheRS (T415A) variant.

Amino Acid	Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)[a]
Phe	T415A	185±56	0.207±0.02	1,163±240	2.1
Trp	T415A	913±310	0.129±0.01	152.4±49	0.28
pBrF	T415A	255±73	0.217±0.03	892±220	1.6
Phe	T415G	499±51	0.278±0.06	553.0±79	1.0
Trp	T415G	55.0±24	0.261±0.08	5,001±1,000	9.0
pBrF	T415G	36.3±5.0	0.211±0.09	6,116±3,400	11
Phe	Wild-type	29.2±9.8	0.302±0.13	10,669±3,700	19
pBrF	Wild-type	ND[b]	ND	ND	ND

[a] Relative to k_{cat}/K_m for Phe by T415G.

[b] Not detected.

Chapter 4

Site-Specific Incorporation of Tryptophan Analogs into Recombinant Proteins in Vivo

Abstract

An expansion of the number of genetically encoded tryptophan (Trp) analogs in *Escherichia coli* has been restricted by two limitations. First, Trp analogs inactive in the native *E. coli* translation system could not be incorporated into proteins. Second, most of the previous attempts to incorporate Trp analogs have been limited to global replacement of all Trp residues throughout a target protein. Here we show that a rationally designed yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) efficiently activated four Trp analogs: 6-chlorotrytophan (6ClW), 6-bromotryptophan (6BrW), 5-bromotrytophan (5BrW), and benzothienylalanine (BT), which were not utilized by the endogenous *E. coli* translational system. Introduction of the yPheRS (T415G) and a mutant yeast phenylalanine amber suppressor ($ytRNA^{Phe}_{CUA\_UG}$) into an *E. coli* expression host allowed the site-specific incorporation of three Trp analogs: 6ClW, 6BrW, and BT, into murine dihydrofolate reductase in response to amber stop codons with at least 98% fidelity. Subsequently, all three Trp analogs were introduced at the Trp66 position in the chromophore of a cyan fluorescent protein (CFP) to investigate changes in spectral properties. CFP variants containing Trp analogs showed a blue-shift in the fluorescence emission peak as well as in the absorption maximum. In particular, a CFP variant with BT featured an unusually large Stokes' shift (56 nm). As shown in this report, an expanded set of the genetically encoded Trp analogs would enable the design of new proteins with novel properties.

Introduction

Tryptophan (Trp) is an attractive target for protein engineering for several reasons. First, it is the main source of UV absorbance and fluorescence of proteins. Second, Trp is involved in various interactions in proteins, such as ligand binding[1-3] and DNA-protein interaction.[4,5] Third, numerous Trp analogs are available through diverse indole chemistry. Fourth, it is the rarest amino acid,[6] and so its substitution by other amino acids minimally perturbs structure.[7] Ever since 7-azatryptophan and 2-azatryptophan were first incorporated into proteins in *E. coli* in the 1950s,[8,9] many researchers have tried to replace Trp with Trp analogs. In the last few decades, a residue-specific incorporation method utilizing a Trp auxotrophic strain realized the incorporation of various Trp analogs containing fluoro, hydroxyl, methyl, amino, selenophene, and thienyl functional groups.[10-12] Fluorinated Trp analogs, in particular, have been used in ^{19}F-NMR studies.[13,14] Trp analogs containing selenophene and thienyl functional groups were used in the X-ray crystallography of proteins.[15,16] Aminotryptophan was used to design protein-based pH sensors[17] and greatly changed the spectral properties of fluorescent proteins.[18]

Although diverse Trp analogs have been introduced into proteins, further expansion of the genetic code for Trp has been restricted by two limitations. First, Trp analogs incorporated into proteins in *E. coli* have so far been limited to the substrates for the native *E. coli* translation system. Therefore, a large number of Trp analogs, such as benzothienylalanine (BT), 5-methyltryptophan, 6-methyltryptophan, 6-aminotryptophan, and 5-cyanotryptophan, were found not to be incorporated into proteins.[10] Recently, several research groups have devised various methods to introduce translationally

inactive nonnatural amino acids into recombinant proteins in *E. coli*.[10,12,19] Overexpression of aminoacyl-tRNA synthetase (aaRS),[20,21] utilization of attenuated editing mutant aaRS,[22,23] rational design of the binding pocket of aaRS,[24,25] and screening of an aaRS library[26,27] have been effective in expanding the genetic code for leucine, isoleucine, valine, methionine, phenylalanine, and tyrosine. Recently, Schultz and colleagues modified *Bacillus stearothermophilus* trytophanyl-tRNA synthetase (*Bs*TrpRS) to incorporate 5-hydroxytryptophan into a recombinant protein in mammalian cells.[28] However, until now there have been no reports about modification of substrate specificity of aaRS to expand the number of genetically encoded Trp analogs in *E. coli*. In this report, we have shown that the rational design of the binding pocket of yPheRS can expand the genetic code for Trp in *E. coli*.

Second, most of the previous attempts to incorporate Trp analogs have been limited to the global replacement of Trp throughout a protein. Although Trp is the rarest amino acid, the replacement of Trp with Trp analogs at non-permissive sites could result in the loss of function. For example, the replacement of three Trp residues in barstar with 4-aminotryptophans greatly reduced its stability due to increased polarity of 4-aminotryptophan at position 53 inside the protein core.[11] In order to overcome such a limitation, we have explored the feasibility of site-specific incorporation of Trp analogs into recombinant proteins in *E. coli*, which would allow incorporation of Trp analogs only at permissive sites. In 1998, Furter showed site-specific incorporation of *p*-fluorophenylalanine into murine dihydrofolate reductase (mDHFR) in *E. coli* through the use of an orthogonal yPheRS/ytRNA^{Phe} pair.[29] More recently, Schultz and colleagues have developed powerful screening methods to change the substrate specificity of

tyrosyl-tRNA synthetase (*mj*TyrRS) derived from *Methanococcus jannaschii* toward nonnatural amino acids.[19,26,30] *E. coli* strains equipped with the screened orthogonal *mj*TyrRS/*mj*tRNA^{Tyr} pairs have allowed incorporation of diverse Phe and Tyr analogs into proteins in a site-specific manner. Tirrell and colleagues refined the orthogonality of the yPheRS/ytRNA^{Phe} pair and the specificity of yPheRS to achieve high fidelity incorporation of *p*-bromophenylalanine into proteins *in vivo*.[25]

In this report, we describe the design of a bacterial host to achieve site-specific incorporation of Trp analogs into a recombinant protein in response to an amber stop codon in *E. coli*. Since no orthogonal tryptophanyl-tRNA synthetase/tRNA^{Trp} pairs have been reported for *E. coli* translation systems, we needed an alternate orthogonal pair. Our previous studies showed that a rationally designed yPheRS (T415G) variant efficiently activates Trp as well as various Phe analogs.[25,31] We hypothesized that the relaxed substrate specificity of yPheRS (T415G) may allow it to recognize a variety of Trp analogs. Therefore, a mutant yPheRS (T415G) and a mutant yeast amber suppressor (ytRNA$^{Phe}_{CUA_UG}$)[32] were introduced into an *E. coli* expression strain to insert Trp analogs at a specific site in a protein. Four Trp analogs, BT, 6ClW, 6BrW, and 5BrrW, were examined for incorporation *in vivo*. In our preliminary studies, none of the four Trp analogs tested were incorporated at UGG codons, even in a Trp auxotrophic *E. coli* strain under a Trp-depleted condition (data not shown). BT is isosteric to Trp and so might result in minimal perturbation of a protein structure upon the replacement of Trp. Introduction of aryl halides into recombinant proteins can allow for site-specific modification through a palladium-catalyzed coupling reactions with alkenes and alkynes.[33,34] Bromine and iodine, once introduced into proteins, can facilitate the phasing

of crystallographic data and therefore can be useful in X-ray diffraction studies of protein structures.[35] As expected, introduction of an orthogonal yeast PheRS (T415G) /ytRNA$^{Phe}_{CUA_UG}$ pair into *E. coli* expression strains enabled high-fidelity (ca. 98%) incorporation of BT, 6BrW, and 6ClW into mDHFR in a site-specific manner.

The *E. coli* strain, equipped with an engineered orthogonal pair, has been used to change the spectral properties of fluorescent proteins. Considering that Trp is a main contributor to UV absorbance and the fluorescence of proteins, it is not surprising that the replacement of Trp with other amino acids resulted in changes in the spectral properties of proteins in many different ways. Schultz and his colleagues have characterized the spectral properties of green fluorescent protein (GFP) variants containing various Phe analogs in the chromophore.[36] Fluorescence emission peaks of all GFP variants were blue-shifted to some extent, depending on the electron-donating ability of the Phe analogs. Similarly, Budisa and his colleagues replaced Trp residues in either GFP or the cyan fluorescent protein (CFP) with translationally active Trp analogs.[11] In particular, the fluorescence emission peak of 4-aminotryptophan in the chromophore of a fluorescent protein was red-shifted significantly.[18] In this report, we have explored whether we could design novel spectral classes of fluorescent proteins containing Trp analogs that would be translationally inactive for the endogenous *E. coli* translation system. Specifically, we replaced a Trp residue in the chromophore of a CFP variant (CFP6) with 6ClW, 6BrW, and BT to generate CFP6_6ClW, CFP6_6BrW, and CFP6_BT, respectively. Similar to halogenated Phe analogs, fluorescence emission peaks of CFP6_6ClW and CFP6_6BrW were blue-shifted. However, quantum yields of halogenated Trp analogs were 3- to 4-fold higher than their Phe counterparts. CFP6_BT exhibited a larger Stokes' shift (56 nm) than

that of CFP6 (ca. 40 nm).

Materials and Methods

Materials. Amino acids **1-2** were obtained from Sigma (St. Louis, MO). **3** and **4** were purchased from Biosynth International (Naperville, IL). **5** and **6** were from Aldrich (St. Louis, MO) and Chem-Impex (Wood Dale, IL), respectively. [^{32}P]-labeled sodium pyrophosphate was purchased from NEN Life Science (Boston, MA).

Amino Acid Activation Assay. The kinetics of activation of amino acids by the yPheRS (T415G) variant were determined by the amino acid dependent adenosine triphosphate (ATP)-[^{32}P]-pyrophosphate (PP_i) exchange reaction. The plasmid construction, expression, and purification of the yPheRS (T415G) variant with C-terminal hexa-histidine tag were described previously.[25] The assay buffer contained 50 mM *N*-(2-hydroxethyl) piperazine-*N*'-(2-ethanesulfonic acid) (potassium-HEPES) (pH=7.6), 20 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 2 mM ATP, and 2 mM [^{32}P]- PP_i. The amino acid concentrations varied from 100 nM to 2.5 mM and the enzyme concentration varied from 100 nM to 400 nM. Aliquots (20 μL) of solution were quenched at various time points in 500 μL of buffer solution containing 200 mM $NaPP_i$, 7% w/v $HClO_4$ and 3% w/v activated charcoal. The charcoal was centrifuged and washed three times with 500 μL of 10 mM $NaPP_i$ and 0.5% $HClO_4$ solution. The [^{32}P]-labeled ATP absorbed on the charcoal was counted via liquid scintillation methods. The kinetic parameters were calculated by a nonlinear regression fit of the data to a Michaelis-Menten model.

Strain and Plasmid Construction for in Vivo Incorporation Assays. pQE16am-T415G[25] encodes a murine dihydrofolate reductase (mDHFR) outfitted with a C-terminal hexa-histidine tag and an amber codon at position 38 . The GFP_{UV} gene was cloned from pGFPuv (Clontech) and inserted into the *Pst*I site of pQE9 (Qiagen) to yield a pQE9_GFP_{UV} plasmid. A series of QuikChange mutations (Stratagene) were performed on GFP_{UV} gene to generate a GFP6 gene containing four sequence changes (F64L, S65T, S99F, and T153M) (see Chapter 6). Either an amber codon (UAG) or a tryptophan codon (UGG) was inserted at position 66 of GFP6 by QuikChange mutagenesis using two complementary pairs of oligonucleotides, Am66_f (5'-CTT GTC ACT ACT CTG ACC TAG GGT GTT CAA TGC TTT TCC-3')/Am66_r (5'-GGA AAA GCA TTG AAC ACC CTA GGT CAG AGT AGT GAC AAG-3') and W66_f (CTT GTC ACT ACT CTG ACC TGG GGT GTT CAA TGC TTT TCC-3')/W66_r (5'-GGA AAA GCA TTG AAC ACC CCA GGT CAG AGT AGT GAC AAG-3'), to yield pQE9_CFP6_66Am and pQE9_CFP6, respectively. The entire expression cassette from either CFP6_66Am or CFP6 was inserted between the *Aat*II and *Nhe*I sites of pQE16am-T415G to generate pQE9_CFP6_66Am-T415G or pQE9_CFP6-T415G, respectively. pREP4_ytRNA^{Phe}_UG was used to constitutively express a mutant yeast amber suppressor tRNA (ytRNA$^{Phe}_{CUA_UG}$) shows minimal cross-reactivity with the *E. coli* lysyl-tRNA synthetase.[25] A Phe/Trp double auxotrophic strain (AFW) was derived from the Phe auxotrophic strain AF (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi $T2^R$ *pheA*18, *trpB*114) by P1 phage-mediated transposon transduction as described previously.[25,31] The double auxotrophic strain, AFW, was co-transformed with either pQE16am-T415G or pQE9_CFP6_66Am-T415G, and with pREP4_ytRNA^{Phe}_UG to yield expression strains

AFW[pQE16am-T415G/pREP4_ytRNA^{Phe}_UG] or AFW[pQE9_CFP6_66Am-T415G/pREP4_ytRNA^{Phe}_UG], respectively. DH10B (Invitrogen) *E. coli* strain was co-transformed with pQE9_CFP6-T415G and pREP4_ytRNA^{Phe}_UG to generate DH10B[pQE9_CFP6-T415G/pREP4_ytRNA^{Phe}_UG].

In Vivo Incorporation Assays. *E. coli* expression strains were grown in M9 minimal medium supplemented with 0.4% glucose, 35 mg/L thiamin, 1mM $MgSO_4$, 1mM $CaCl_2$, 20 amino acids (at 25 mg/L), 35 mg/L kanamycin, and 200 mg/L ampicillin. The overnight culture was diluted 20-fold in fresh M9 minimal medium. The cells were grown to an OD_{600} between 0.9 and 1.0, sedimented by centrifugation, and washed twice with cold 0.9% NaCl. The cell pellet was resuspended in fresh M9 minimal medium containing 18 amino acids (at 25 mg/L), 3 mM analog of interest, and the indicated concentrations of Phe and Trp. After 10 min, 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) was added to induce protein expression. After 4 hours, cells were pelleted and kept at -80 °C.

A different procedure was used for expression strain DH10B[pQE9_CFP6-T415G/pREP4_ytRNA^{Phe}_UG]. A single colony was inoculated into 5 mL of 2xYT medium and grown overnight. 5 mL of the overnight culture was transferred to 100 mL of fresh 2xYT medium. When the cultures reached an OD_{600} of 0.6, 1 mM IPTG was added to induce protein expression. After 3 hours, cells were sedimented and stored at -80 °C. Whole cell lysates were analyzed by SDS-PAGE.

Quantitative Analysis of Analog Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS). Quantitative analysis of amber codon occupancy by various amino acids was performed by LC-MS as described previously.[25] In a minor alteration, mutant mDHFR was purified on a Ni-NTA spin column under denaturing conditions according to the supplier's instructions (Qiagen). After purification, mutant mDHFR concentrations were determined by UV absorbance at 280 nm using a calculated extinction coefficient[37] of 24,750 cm^{-1} M^{-1}. After concentration of the protein by ultrafiltration (Millipore), 10 μL of the solution was diluted into 90 μL of 75 mM $(NH_4)_2CO_3$ for trypsin digestion. 1 μL of modified trypsin (Promega, 0.2 μg/μL) was added. The samples were incubated for 2 hrs at 37 ^oC. The reaction was quenched by addition of 13 μL of 5% trifluoroacetic acid (TFA) solution. The solution was then directly subjected to LC-MS/MS analysis. LC-MS/MS analysis of protease-digested peptides was conducted on a QSTAR XL LC-MS/MS system (Applied Biosystems/MDS SCIEX) with HPLC system (NanoLC-2DTM, Eksigent Technologies) and ESI probe. Peptides were separated on a C18 reversed column (3 μm, 300Å, 0.1 x 70mm) and eluted at a flow rate of 0.2 μL/min using a gradient of 0-50% of solvent B (89% acetonitrile/10% H_2O/0.02% TFA/1% formic acid) and solvent A (2% acetonitrile/97% H_2O/0.02% TFA/1% formic acid) in 55 min. The column eluent was transferred to the electrospray source and sequencing was carried out by fragmentation of the precursor ion corresponding to the peptide containing the residue at position 38 of mutant mDHFR.

Fluorescent Protein Isolation. Fluorescent proteins were purified by Ni-NTA affinity chromatography according to the manufacturer's protocol (Qiagen) under native

conditions. Purified proteins were desalted over a Sephadex PD-10 column (Amersham Pharmacia, Piscataway, NJ) using 20 mM Tris-HCl, 1 mM EDTA, pH 8.0. Protein samples other than CFP6 were further purified by ammonium sulfate precipitation.[38] Dry ammonium sulfate was added to the bright fluorescent solution to 40% saturation (about 1.6 M). Then solutions were incubated on ice for 25 min and the precipitated proteins were removed by centrifugation at 14,000 rpm for 15 min at 4 ^{o}C. Dry ammonium sulfate was added to the fluorescent supernatants to 70% saturation (about 2.8 M). The solutions were incubated on ice for 30 min and the fluorescent proteins were collected by centrifugation at 14,000 rpm for 30 min at 4 ^{o}C. The fluorescent proteins were resuspended in 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 and desalted over polyacrylamide P-6 columns (Bio-rad) using 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 buffer. Protein concentrations were determined by the BCA protein assay kit (Pierce Biotechnology, Rockford, IL) using bovine serum albumin as a reference.

Characterization of Spectral Properties. Absorption spectra of fluorescent proteins were measured on a Cary 50 UV-visible spectrophotometer (VARIAN, Palo Alto, CA). Protein samples were diluted to four different concentrations ($0.1 < OD_{\lambda max} < 0.5$). Extinction coefficients were calculated by applying Beer's law to the absorption spectra of diluted protein samples. A molar mass of 30,287 Da was used for all fluorescent proteins. Fluorescence spectra were measured on a PTI QuantaMaster fluorescence spectrofluorometer (Photon Technology International, Birmingham, NJ) at a scan rate of 1 nm/s. For the quantum yield calculation, protein samples in assay buffer and fluorescein (Invitrogen) in borate buffer (pH 9.1) were adjusted to similar OD ($OD_{\lambda ex} < 0.1$). The

fluorescence emission spectra were recorded with excitation at the maxima in the corresponding excitation spectra. The emission spectra were corrected by the correction functions supplied with the instrument. The quantum yields of fluorescent proteins were calculated by comparing the integrated intensities of the corrected emission spectra with that of fluorescein of known quantum yield (0.91).[39]

Results and Discussion

Rational Design of yPheRS and Choice of Trp Analogs. In order to increase the number of genetically encoded Trp analogs, we engineered the substrate specificity of yPheRS. Studies of the crystal structure of PheRS (tPheRS) derived from *Thermus thermophilus* revealed that Val 261 in the amino acid binding pocket is critical for distinguishing Phe from other non-aromatic amino acids.[40] Sequence alignments showed that Val 261 in tPheRS corresponds to Thr 415 in yPheRS. Introduction of a single mutation (Thr to Gly) at position 415 in the binding pocket of yPheRS (yPheRS (T415G)) created extra room to accommodate various Phe analogs.[25,31] There was an interesting finding that yPheRS (T415G) activated Trp and 2-naphthylalanine as well as substituted Phe analogs. Considering the spectrum of nonnatural amino acids recognized by yPheRS (T415G), we hypothesized that yPheRS (T415G) may activate Trp analogs. In order to achieve site-specific incorporation of Trp analogs, we needed to choose Trp analogs that can be recognized by yPheRS (T415G) but by the native *E. coli* translation system. We have chosen four Trp analogs (6ClW, 6BrW, 5BrW, and BT) that can not be utilized by the endogenous *E. coli* translational system. Based on previous data,[10,11] we proposed a model for the adaptability of Trp analogs to the binding pocket of *E. coli*

TrpRS (eTrpRS) (Figure 2). Recognition of Trp analogs by eTrpRS depends on the size and position of substituents on the indole ring. Methyl (ca. 16.8 $Å^3$),[41] amino (ca. 12.3 $Å^3$), and fluorine groups (ca. 3.2 $Å^3$) were the largest functional groups tolerated at the C4, C5 and C6 positions of the indole ring, respectively.[10] Therefore, introduction of bulkier groups at each position would result in translationally inactive Trp analogs. Based on our hypothetical model, we chose three translationally inactive Trp analogs, **3**, **4**, and **5** containing chlorine (ca. 14.3 $Å^3$) or bromine (ca. 20 $Å^3$) at the C6 position, and bromine at the C5 position of the indole ring. Additionally, we chose **6,** known as an translationally inactive Trp analog.[42] Replacement of the nitrogen in the indole ring of Trp with other hetero-atoms, such as selenium, sulfur, or oxygen, results in translationally inactive amino acids.[15,42] As expected, none of the four Trp analogs (**3**-**6**) were incorporated into proteins, even in Trp auxotrophic *E. coli* strain under Trp depleted conditions (data not shown).

Activation of Trp Analogs by yPheRS (T415G). The relative rates of activation of Phe, Trp and Trp analogs (**1**-**6**) by yPheRS (T415G) were examined by ATP-PP_i exchange assays. The kinetic parameters are shown in Table 1. Derivatives of Trp (**3**-**6**) revealed different activities with respect to substitution at different positions of the indole ring. Substitution at position 6 (**3** and **4**) was highly favorable; **3** displayed the highest activity among all the analogs we tested. The increase in side-chain volume upon replacement of **2** with **3** (ca. 39 $Å^3$) matched very well the cavity (ca. 39 $Å^3$) generated by the T415G mutation in the binding pocket of yPheRS. **4** exhibited a k_{cat}/K_m value comparable to that of Trp. Notably, **4** showed 7-fold larger K_m value than that of **3,** indicating less favorable

binding of **4** to the binding pocket of yPheRS. This result is consistent with the fact that the calculated increase in side-chain volume (45 $Å^3$) upon replacement of **2** with **4** is larger than the cavity generated by the T415G mutation. Substitution at position 5 of the indole ring was less favorable than that at position 6. The activation rate of **5** by yPheRS (T415G) was one half of that of **4**. Replacement of an imine in the indole ring with a sulfur atom reduced binding to yPheRS (T415G). The K_m of **6** was 3-fold greater than that of **1**. However, the activation rate of **6** by yPheRS (T415G) was 6.4-fold faster than that of **2**. The spectrum of activity of the yPheRS (T415G) supported our proposal that the yPheRS (T415G) may recognize Trp analogs.

Site-Specific Incorporation of Trp Analogs into mDHFR in Vivo. To investigate the utility of yPheRS (T415G) to incorporate Trp analogs into proteins in vivo, mDHFR was chosen as a test protein. An amber stop codon was placed at the 38th position of mDHFR to generate mDHFR_38Am. The expression of full length mDHFR_38Am with various Trp analogs (**3**-**6**) was examined by SDS-PAGE analysis (Figure 3). All of the Trp analogs supported expression of full length mDHFR_38Am. The band intensities of mDHFR_38Am variants were in the order of **4** > **3** > **5** > **6**. Although expression of full length mDHFR_38Am strongly indicated that the amber stop codon was suppressed, it should be noted that the mDHFR_38Am expression does not necessarily mean that the Trp analogs were incorporated at the amber site. In order to verify incorporation of Trp analogs at the amber site, the occupancy of the site in mDHFR_38Am by various amino acids was determined by LC-MS/MS analysis. Trypsin digestion of purified mDHFR_38Am variants generated diverse peptide fragments. Among them, a peptide

containing the amber site (residues 26-39, NGDLPWPPLRNEAmK) was designated as peptide 38. Peptide 38 variants containing Phe, Trp or Trp analogs were separated on a C_{18} reversed phase column and relative yields were determined by comparing the integrated areas of the corresponding signals in the chromatogram. The results are summarized in Table 2. Furthermore, the identity of each peptide was confirmed by tandem-mass spectrometry. As a representative, the identity of peptide 38 containing 6BrW at the amber site was determined from the spectrum shown in Figure 4. We observed that **3**, **4**, and **6** occupied at least 98% of the amber sites in full length mDHFR_38Am (Table 2). However, **5** occupied only 2% of the amber site, though it was efficiently activated by yPheRS (T415G). Instead, **2** occupied 97% of the amber site. The discrepancy between in vitro activation and in vivo incorporation for **5** could be rationalized by the model proposed by Sisido and colleagues for the adaptability of non-natural amino acids to the *E. coli* ribosome.[43] The model suggests that the *E. coli* ribosome has the ability to define the allowed and excluded regions of the aromatic groups of aryl-*L*-alanines to be bound for further translation. The addition of bromine increases the volume of the indole ring by 20 $Å^3$. We reason that the addition of bromine at the C5 position of the indole ring may result in exclusion of **5** by the ribosome. Another interesting finding from the experiment for analog **5** is that at similarly low concentrations of Phe and Trp in the culture medium, Phe was incorporated into the amber site instead of Trp. According to our in vitro activation assays, Trp was activated 18-fold faster than Phe by yPheRS (T415G). This discrepancy between the results of in vitro activation assays and in vivo translation might be explained by elevated Trp degradation in the presence of Trp analogs. Expression of tryptophanase, a major enzyme

involved in Trp degradation, is regulated by tryptophan-induced transcription antitermination in *E. coli*.[44,45] Even 1-methyltryptophan, a Trp analog, acted as an effective inducer in vitro.[45] Therefore, we reasoned that supplementation of **5** into the culture medium accelerated Trp degradation via formation of tryptophanase and reduced the intracellular concentration of Trp levels lower than that of Phe. Since the ytRNA$^{Phe}_{CUA_UG}$ charged with **5** was rejected in ribosome, the ytRNA$^{Phe}_{CUA_UG}$ charged with **2** was dominantly utilized for translation.

Expression yields of mDHFR_38Am with various Trp analogs were determined by measuring the concentration of purified mDHFR_38Am. The results are summarized in Table 2. One interesting finding was that **4** provided 25% higher expression level of mDHFR_38Am as compared to **3** despite its 3.5-fold lower k_{cat}/K_m value. This discrepancy can be rationalized by higher k_{cat} of **4** than that of **3**. The concentration (3 mM) of Trp analogs supplemented into culture medium was at least 900-fold higher than K_m values of both **3** and **4**. In such a case, the k_{cat} value became more important than k_{cat}/K_m in determining relative activation rates. Therefore, **4** with 60% higher k_{cat} value than that of **3** could be more efficiently used for protein translation.

Spectral Analysis of Fluorescent Proteins. The spectral properties of CFP6 and CFP6_66Am proteins containing Trp analogs are summarized in Table 3. Absorption maxima, emission maxima, and extinction coefficient of CFP6 were consistent with literature values.[46] Compared to CFP6, CFP6_6ClW and CFP6_6BrW showed blue-shifts of 14 nm in the fluorescence emission maxima (Figure 6). Similar trends were found in previous work in which Tyr at position 66 of GFP was replaced by *p*-

bromophenylalanine.[36] In addition, solution studies indicated that the absorption and fluorescence spectra of fluorinated tryptophans were blue-shifted;[16] the electron-withdrawing property of halogen functional groups can account for the blue-shift in the fluorescence emission peak. Compared with CFP6, the quantum yields of the CFP6_6ClW and CFP6_6BrW decreased 2.5 to 3-fold (Table 3), probably due to collisional perturbation of spin-orbital coupling in the π-electron orbitals of the fluorophore by the halogen atom.[47] CFP6_BT showed blue-shifts of 16 and 34 nm in fluorescence emission and absorption maxima, respectively. Similar blue-shift in absorption maxima in CFP variants was observed by replacing a tryptophan with sulfur containing tryptophan surrogates, such as β-(thieno[2,3-*b*]pyrrolyl)-*L*-alanine ([2.3]Tpa) and β-(thieno[3,2-*b*]pyrrolyl)-*L*-alanine ([3.2]Tpa).[16] Extinction coefficient of CFP6_BT (8,100 $M^{-1}cm^{-1}$) was similar to those of CFP variants containing either [2,3]Tpa (6840 $M^{-1}cm^{-1}$) or [3,2]Tpa (6350 $M^{-1}cm^{-1}$). Introduction of either [2,3]Tpa or [3,2]Tpa in the chromophore of CFP resulted in the disappearance of fluorescence;[16] however, CFP6_BT exhibited substantial fluorescence, with quantum yield only 3-fold smaller than that of CFP6. CFP6_BT is the only fluorescent protein containing a sulfur atom in its chromophore. CFP6_BT has several advantages as a fluorescence resonance electron transfer (FRET) partner. It can be efficiently excited by a common violet diode laser (405-415 nm), which can simplify cell screening on the basis of FRET signals. In addition, the Stokes' shift of CFP6_BT (56 nm) is larger than that of CFP6 (37 nm). This large Stokes' shift is advantageous for fluorescence measurements, since it allows emission photons to be detected against a low background and isolated from excitation photons. CFP and yellow fluorescent protein (YFP) have been commonly used as a FRET

pair. Because the fluorescence emission spectrum of CFP_BT (Figure 6 **b**) overlaps the excitation spectrum of GFP, it can be used as a FRET partner with GFP.

Conclusions

In this paper, we have shown that properly engineered *E. coli* strains can allow high-fidelity incorporation of Trp analogs into recombinant proteins in response to an amber stop codon. Rational modification of the binding pocket in yPheRS enabled activation of Trp analogs (**3**-**6**). When the engineered yPheRS and the yeast amber suppressor tRNA were introduced into a Phe/Trp auxotrophic *E. coli* expression host, at least 98% of the amber sites in recombinant mDHFR were occupied by Trp analogs (**3**, **4** and **6**). In a subsequent experiment, three Trp analogs (**3**, **4**, and **6)** were introduced at the Trp66 position of CFP. CFP variants containing Trp analogs (**4, 5**, and **6**) in the chromophore showed blue-shifts in fluorescence emission and absorption maxima. In particular, CFP6BT provided unusually large Stokes' shift (56 nm). Introduction of aryl bromides into recombinant proteins should allow site-specific modification via bio-orthogonal palladium-catalyzed coupling reactions. An expanded set of genetically encoded Trp analogs should enable design of new proteins with novel properties.

Acknowledgments

We thank I. C. Tanrikulu, R. Connor, and T. H. Yoo for discussion and valuable comments. We are grateful to Dr. J. Zhou for help with LC-MS analysis. This work was supported by National Institutes of Health grant.

References

1. Fulton, K. F.; Jackson, S. E.; Buckle, A. M. *Biochemistry* **2003**, *42*, 2364-2372.
2. Uchiyama, T.; Katouno, F.; Nikaidou, N.; Nonaka, T.; Sugiyama, J.; Watanabe, T. *J. Biol. Chem.* **2001**, *276*, 41343-41349.
3. Bray, M. R.; Johnson, P. E.; Gilkes, N. R.; McIntosh, L. P.; Kilburn, D. G.; Warren, R. A. J. *Protein Sci.* **1996**, *5*, 2311-2318.
4. Marie, G.; Serani, L.; Laprevote, O.; Cahuzac, B.; Guittet, E.; Felenbok, B. *Protein Sci.* **2001**, *10*, 99-107.
5. Eason, D. E.; Shepherd, A. T.; Blanck, G. *Biochim. Biophys. Acta - Gene Struct. Express.* **1999**, *1446*, 140-144.
6. Dayoff, M. O., *Atlas of protein sequence and structure* National Biomedical Research Foundation: Washington DC, 1972.
7. Wong, C. Y.; Eftink, M. R. *Protein Sci.* **1997**, *6*, 689-697.
8. Brawerman, G.; Ycas, M. *Arch. Biochem. Biophys.* **1957**, *68*, 112-117.
9. Pardee, A. B.; Shore, V. G.; Prestidge, L. S. *Biochim. Biophys. Acta* **1956**, *21*, 406-407.
10. Budisa, N. *Angew. Chem. Int. Ed.* **2004**, *43*, 6426-6463.
11. Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893-904.
12. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.
13. Pratt, E. A.; Ho, C. *Fed. Proc.* **1974**, *33*, 1463-1463.
14. Seifert, M. H.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 7932-7942.
15. Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2001**, *309*, 925-936.

16. Budisa, N.; Pal, P. P.; Alefelder, S.; Birle, P.; Krywcun, T.; Rubini, M.; Wenger, W.; Bae, J. H.; Steiner, T. *Biol. Chem.* **2004**, *385*, 191-202.

17. 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.

18. 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.

19. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1-11.

20. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

21. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235-237.

22. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

23. Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501-504.

24. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

25. Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

26. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

27. 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.

28. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8882-8887.

29. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

30. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

31. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

32. The mutant yeast amber suppressor was used to reduce lysine misincorporation at the amber sites, because it showed minimal cross-reactivity with *E. coli* lysyl-tRNA synthetase.

33. Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; Yokoyama, S. *ChemBioChem* **2006**, *7*, 134-139.

34. Ojida, A.; Tsutsumi, H.; Kasagi, N.; Hamachi, I. *Tetrahedron Lett.* **2005**, *46*, 3301-3305.

35. Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22*, 1297-1301.

36. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

37. Gill, S. C.; Vonhippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

38. Yakhnin, A. V.; Vinokurov, L. M.; Surin, A. K.; Alakhov, Y. B. *Protein Expr. Purif.* **1998**, *14*, 382-386.

39. Velapoldi, R. A.; Tonnesen, H. H. *J. Fluoresc.* **2004**, *14*, 465-472.

40. Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1534-44.

41. Volumes were estimated as solvent-excluded volumes (1.4 Å probe) by using Chem3D Pro 5.0 (Cambridge, MA)

42. Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. *Protein Sci.* **2001**, *10*, 1281-1292.

43. Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 34-40.

44. Gong, F.; Ito, K.; Nakamura, Y.; Yanofsky, C. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8997-9001.

45. Gong, F.; Yanofsky, C. *J. Biol. Chem.* **2001**, *276*, 1974-1983.

46. Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

47. Kasha, M. *J. Chem. Phys.* **1952**, *20*, 71-74.

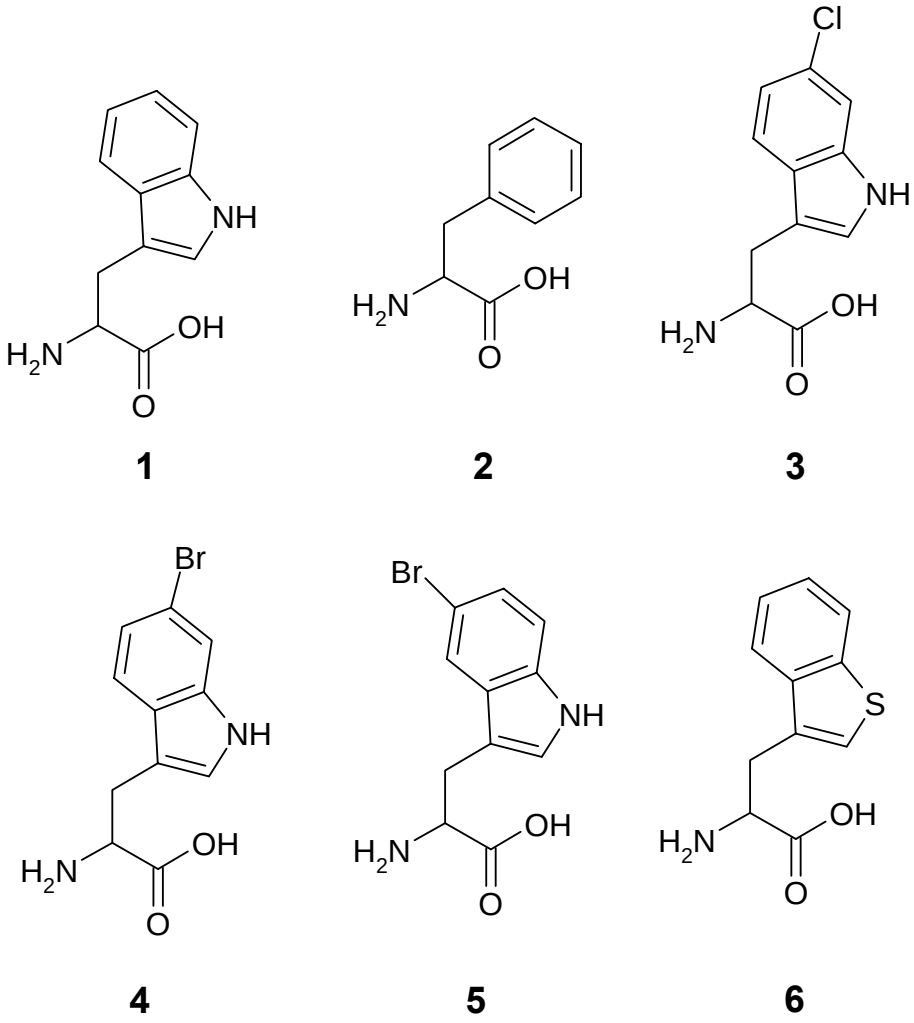


Figure 1: Amino acids involved in this study. (**1**) tryptophan, (**2**) phenylalanine, (**3**) 6-chlorotryptophan, (**4**) 6-bromotryptophan, (**5**) 5-bromotryptophan, (**6**) benzothienylalanine.

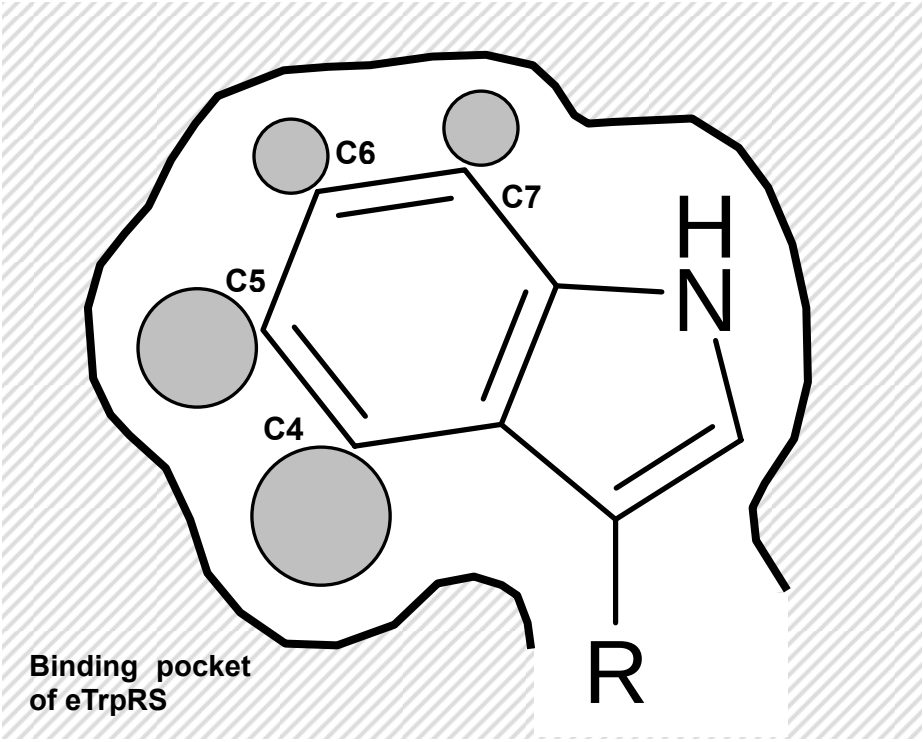


Figure 2: A hypothetical model for the adaptability of Trp analogs to the binding pocket of eTrpRS. Position and size of substituents on the indole ring that can be tolerated in the binding pocket of eTrpRS are represented as circles. Indole rings containing larger substituents can not fit into the binding pocket of eTrpRS.

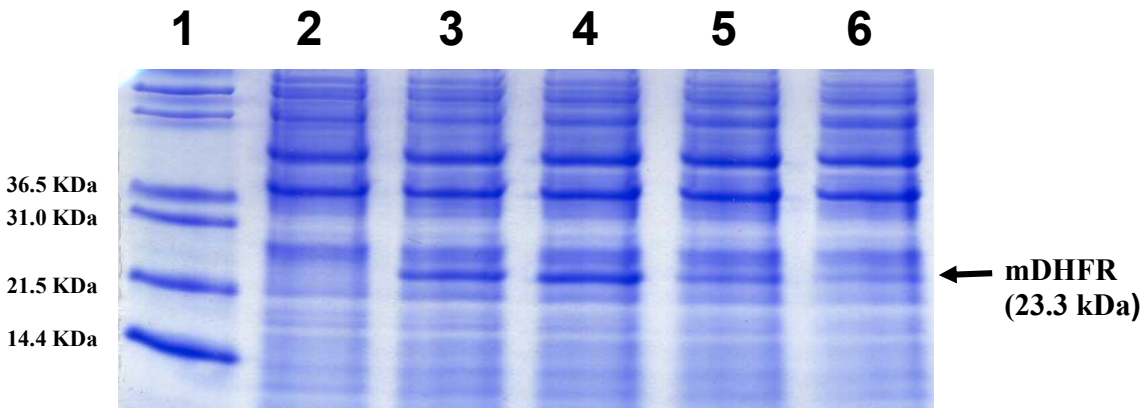


Figure 3: SDS-PAGE analysis of mDHFR_38Am expression. Cells were equipped with yPheRS (T415G) and ytRNA$^{Phe}_{CUA_UG}$. Cultures were supplemented with 18 amino acids (at 25 mg/L), one of Trp analogs (at 3 mM) and the indicated concentrations of Phe and Trp. Lane 1: molecular weight standards, lane 2: before induction, lane 3: 6ClW, 0.03 mM Phe and 0.01 mM Trp, Lane 4: 6BrW, 0.015 mM Phe and 0.05 mM Trp, lane 5: 5BrW, 0.03 mM Phe and 0.01 mM Trp, lane 6: BT, 0.01 mM Phe and 0.0025 mM Trp.

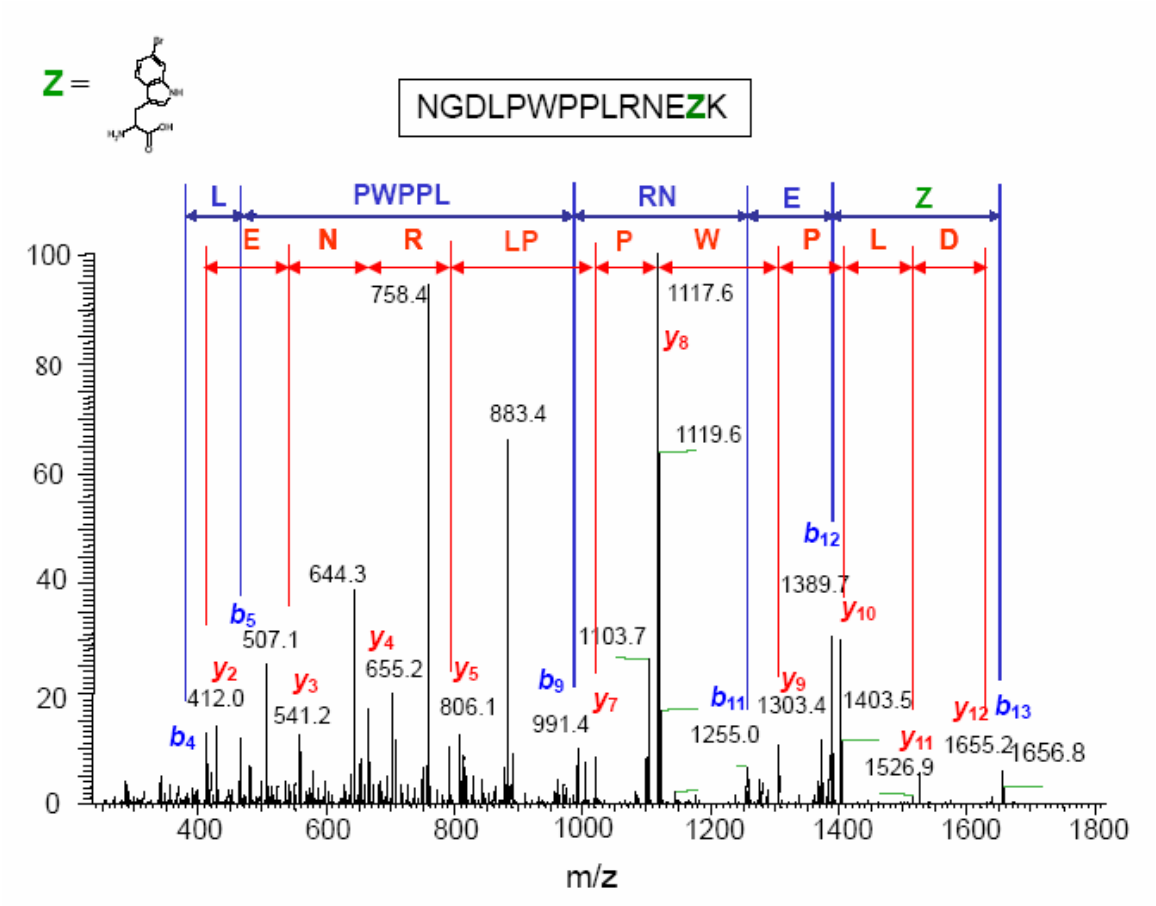


Figure 4: Tandem mass spectrum of the peptide (NGDLPWPPLRNEZK). The partial sequence DLPWPPLRNEZ containing 6BrW (**Z**) can be confirmed by assigning masses of series of fragmented ions (*b* and *y*).

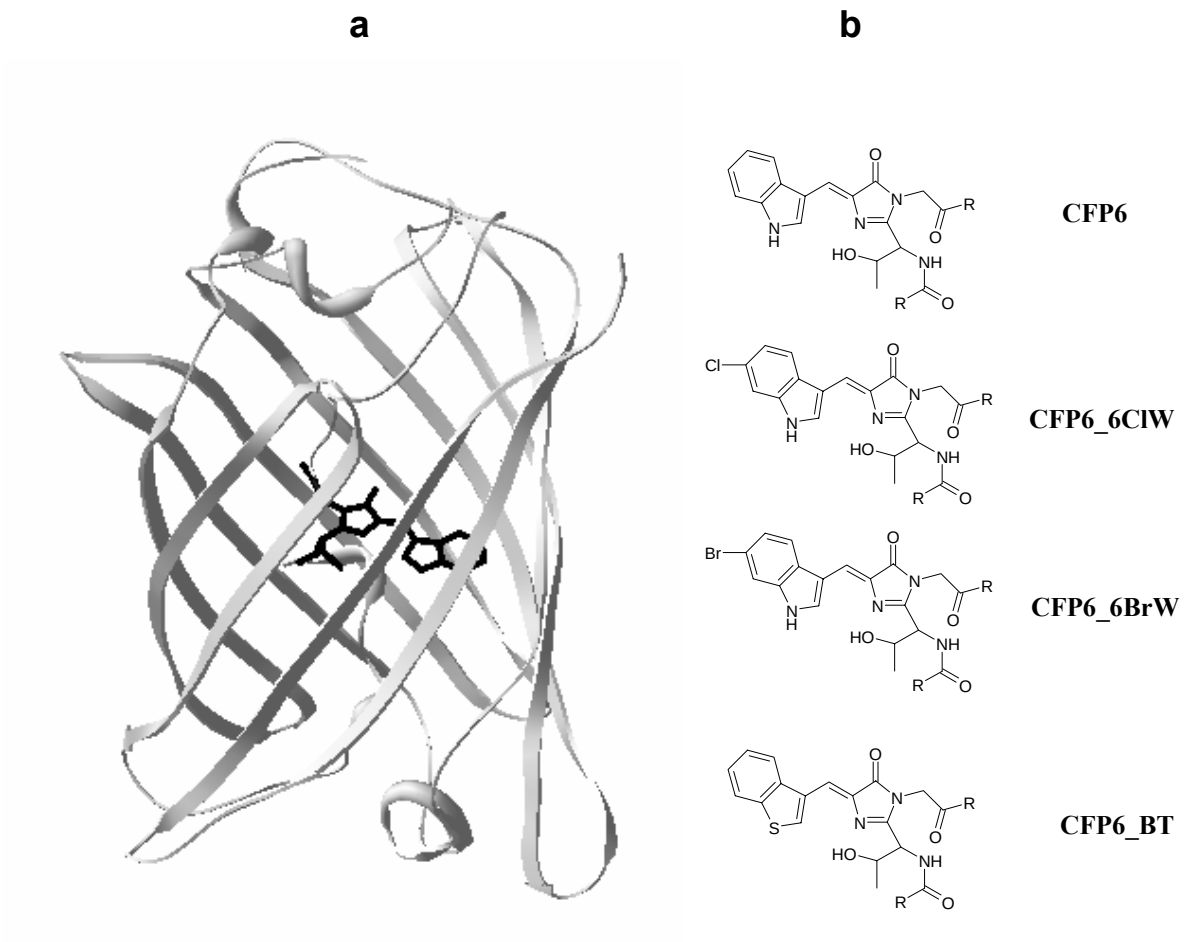


Figure 5: Crystal structure of enhanced cyan fluorescent protein (PDB code 1OXD)[18] (**a**). The chromophore is shown in black. Chromophores of CFP6 variants containing **1**, **3**, **4**, and **6** (**b**).

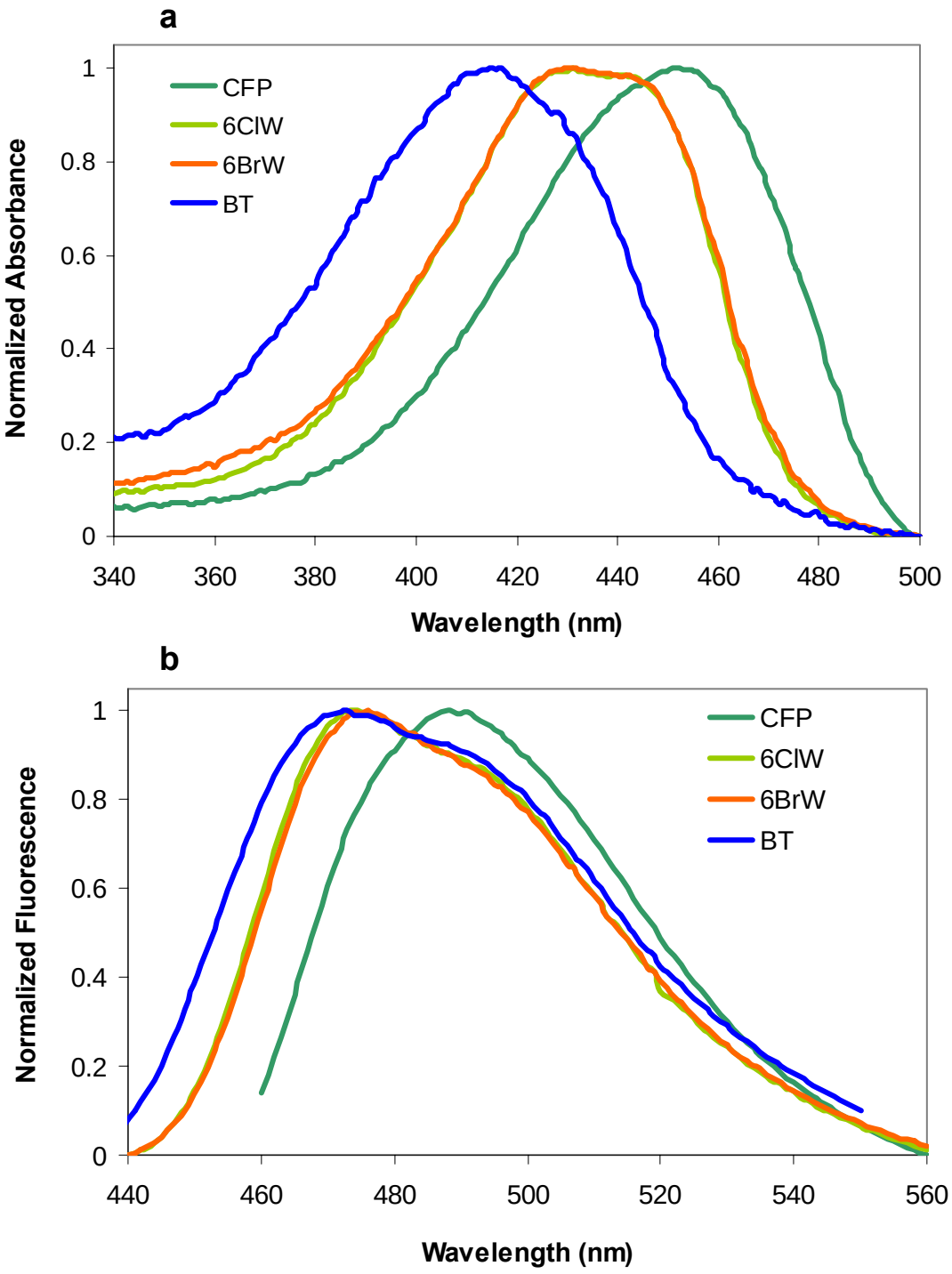


Figure 6: Absorption (**a**) and fluorescence emission spectra (**b**) for CFP6 variants containing different Trp analogs at the Trp66 position.

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

Amino Acid	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)[a]
Phe (**2**)	55±14	0.20±0.11	3,500±1,100	1
Trp (**1**)	2.8±1.6	0.15±0.003	63,200±34,600	18
6ClW (**3**)	0.51±0.21	0.11±0.04	224,100±15,300	64
6BrW (**4**)	3.4±1.3	0.18±0.11	61,800±53,900	18
5BrW (**5**)	0.62±0.02	0.021±0.002	33,900±3,600	9.7
BT (**6**)	8.1±3.7	0.15±0.06	22,500±18,000	6.4

[a] relative to k_{cat}/K_m for Phe by yPheRS (T415G)

Table 2: Occupancy of amber sites and expression yields

Trp analogs	Occupancy of amber sites (%)				Yield (mg/L)[a]
	Phe	Trp	Lys	Naa	
Trp (**1**)	0.4 ± 0.5	ND[b]	1.2 ± 1.3	98.4 ± 1.7	5.4 ± 0.3
6ClW (**3**)	0.4 ± 0.7	ND	ND	99.6 ± 0.7	6.8 ± 1.3
6BrW (**4**)	97.0 ± 1.4	0.3 ± 0.6	0.7 ± 0.5	1.8 ± 0.7	3.6 ± 0.6
BT (**6**)	0.6 ± 0.6	0.7 ± 0.6	0.9 ± 1.5	97.9 ± 2.4	2.8 ± 0.6

[a] Volumetric yields are given as mg of purified mDHFR_38Am per liter of culture.
[b] Not detected in LC-MS analysis.

Table 3: Spectral properties of CFP6 variants

amino acid at position 66	absorption maximum (nm)	extinction coefficient (M^{-1} cm^{-1})	emission maximum (nm)	quantum yield
Trp (**1**)	449	25900 ± 1000	488	0.14 ± 0.008
6ClW (**3**)	430	19800 ± 6400	474	0.050 ± 0.001
6BrW (**4**)	430	20600 ± 3800	474	0.058 ± 0.001
BT (**6**)	415	8100 ± 1100	472	0.048 ± 0.008

Chapter 5

Breaking the Degeneracy of the Genetic Code

Abstract

A mutant yeast phenylalanine transfer RNA (ytRNA$^{Phe}_{AAA}$) containing a modified (AAA) anticodon was generated to explore the feasibility of breaking the degeneracy of the genetic code in *Escherichia coli*. By using an *E.coli* strain co-transformed with ytRNA$^{Phe}_{AAA}$ and a mutant yeast phenylalanyl-tRNA synthetase, we demonstrate efficient codon-biased replacement of phenylalanine (Phe) by L-3-(2-naphthyl)alanine (2Nal), a non-proteinogenic analog. Site-specific incorporation of 2Nal in response to UUU codons was confirmed by mass spectrometric analysis of recombinant murine dihydrofolate reductase. These results illustrate a general method for increasing the number of distinct, genetically-encoded amino acids available for protein engineering and for exploration of the chemistry and physics of protein-like macromolecules.

Introduction

Organisms use a canonical set of 20 amino acids to generate the proteins that sustain the life of the cell. In recent years, several laboratories have pursued an expansion in the number of genetically-encoded amino acids, by using either a nonsense suppressor or a frameshift suppressor tRNA to incorporate non-canonical amino acids into proteins in response to amber or four-base codons, respectively.[1-10] Such methods have worked well for single-site insertion of novel amino acids; however, their utility in multi-site incorporation is limited by modest (20-60%) suppression efficiencies.[1, 5, 11]

Efficient multi-site incorporation has been accomplished by replacement of natural amino acids in auxotrophic *E. coli* strains,[12-15] and by using aminoacyl-tRNA synthetases with relaxed substrate specificity or attenuated editing activity.[14, 16] Although this method

provides efficient incorporation of analogs at multiple sites, it suffers from the limitation that the novel amino acid must "share" codons with one of the natural amino acids. We present here a potential solution to this coding problem.

Material and Methods

Materials. All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. L-3-(2-naphthyl)alanine and other natural amino acids were obtained from Sigma (St. Louis, MO).

Strains and Plasmids. An *E. coli* strain XL1-Blue (Stratagene) was used for plasmid propagation and isolation. A Phe auxotrophic strain K10-F6Δ (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi $T2^R$ *pheA*18)[4] was a gift from Rolf Furter (University of Massachusetts). Plasmids carrying modified yeast $tRNA^{Phe}$ variants were derived from pRO117[4] in which yeast $tRNA^{Phe}$ expression cassette was inserted at the *Sna*I site of the pREP4 (Qiagen). Plasmids for expression of mDHFR and overexpression of mutant yeast phenylalanyl-tRNA synthetases (yPheRS) were derived from pRO148[4] in which yPheRS genes under constitutive *tac* promoter control was inserted at the *PvuII* site of pQE16 (Qiagen).

Construction of Plasmid Carrying Mutant Yeast $tRNA^{Phe}$. pREP4_ytRNA^{Phe}_AAA containing a mutant yeast $tRNA^{Phe}$ with a modified anticodon (AAA) ($ytRNA^{Phe}_{AAA}$) was generated by PCR mutagenesis kit (Stratagene) using pRO117[4] as a template.

Construction of Plasmid Carrying mDHFR and Mutant Yeast Phenylalanyl-tRNA Synthetase. An intact mDHFR expression cassette was obtained by endonuclease restriction of pQE16 at the *Aat*II and *Nhe*I sites. This expression cassette was inserted between the *Aat*II and *Nhe*I sites of pRO148 to generate pQE16_mDHFR_yPheRS. A mutation of threonine to glycine at the 415th position of α-subunit of yPheRS was performed by PCR mutagenesis kit (Stratagene) using pQE16_mDHFR_yPheRS as a template to generate pQE16_mDHFR_yPheRS (T415G).

Protein Expression and Purification. Cultures of K10-F6Δ outfitted with pREP4_ytRNA_AAA and pQE16_mDHFR_yPheRS (T415G) were grown in M9 minimal medium supplemented with 0.4% (w/v) of glucose, 0.1 mM of $CaCl_2$, 1.0 mM of $MgSO_4$, 35 μg/mL of thiamine, 20 amino acids (25 mg/L), 100 μg/mL of ampicillin, and 35 μg/mL of kanamycin. When an optical density of the culture reached 0.8 to 1.0 at 600nm (OD_{600}), the cultures were centrifuged for 7 min (6,000 rpm) at 4 °C. The cell pellets were washed twice with 0.9% (w/v) of NaCl solution. The cells were resuspended in M9 minimal medium supplemented with 18 amino acids (25 mg/L) and indicated amounts of 2Nal and Phe. After 10 min incubation, 1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to induce mDHFR protein expression. The OD_{600} of the culture was measured 4 hrs after induction. Then the cells were harvested by centrifugation for 10 min (6,000 rpm) at 4°C and stored at -70°C. After thawing, mDHFR proteins were purified under denaturing conditions according to manufacturer's

protocol (Qiagen). Protein expression was evaluated by SDS-PAGE with coomassie blue staining. Loading of the gel was normalized for cell densities as determined by OD_{600}.

Amino Acid Analysis. The purified mDHFR solutions were concentrated 10-fold by ultrafiltration (Millipore) followed by a buffer exchange against 0.1% trifluoroacetic acid (TFA) solution. Samples were submitted to the Molecular Structure Facility at the University of California, Davis, for amino acid analysis on a Beckman 6300 instrument (Fullerton, CA).

Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI–MS) Analysis. The purified mDHFR solutions were desalted by $ZipTip_{C18}$ (Millipore) and eluted with 3 μL of 50% CH_3CN/0.1% TFA. 1 μL was used for matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis with sinapinic acid (30% (v/v) of acetonitrile and 70% (v/v) of 0.1% TFA solution) as the matrix. The analysis was performed on a PerSeptive Biosystems (Framingham, Massachusetts) Voyager DE PRO MALDI-TOF mass spectrometer operating in linear and positive ion modes. For MALDI-MASS analysis of tryptic digests of purified mDHFR, 10 μL of the concentrated protein solution was added to 90 μL of 75 mM ammonium bicarbonate solution. 2 μL of 0.2 g/L of modified trypsin (Promega) was added, and the solution was incubated at 37°C for 2.0 hrs. 12 μl of 5% TFA solution was added to quench the reaction. Chromatography on $Ziptip_{C18}$ columns (Millipore) provided purified peptide samples (1 μL), which were added to 2,5-dihydroxybenzoic acid (DHB) MALDI matrix solution for MALDI-MASS analysis.

Results and Discussion

The genetic code is degenerate, in that the protein biosynthetic machinery utilizes 61 mRNA sense codons to direct the templated polymerization of a set of 20 amino acid monomers.[17] Just two amino acids, methionine and tryptophan, are encoded by unique mRNA triplets. Re-assignment of degenerate sense codons therefore offers the prospect of a substantially expanded genetic code and a correspondingly enriched set of building blocks for natural and artificial proteins.

As a test case for establishing the feasibility of breaking the degeneracy of the code, we chose the biosynthetic machinery responsible for incorporation of phenylalanine (Phe) into the proteins of *E. coli*. Phe is encoded by two codons, UUC and UUU. Both codons are read by a single ytRNA, which is equipped with the anticodon sequence GAA. The UUC codon is therefore recognized through standard Watson-Crick base-pairing between codon and anticodon; UUU is read through a G-U wobble base pair at the first position of the anticodon.[18] Thermal denaturation of RNA duplexes has yielded estimates of the Gibbs free energies of melting of G-U, G-C, A-U, and A-C base pairs as 4.1, 6.5, 6.3, and 2.6 kcal/mol, respectively, at 37^{o}C.[19] Thus the wobble base pair, G-U, is less stable than the Watson-Crick base pair, A-U. On this basis, we proposed that a mutant ytRNA^{Phe} outfitted with the AAA anticodon (ytRNA$^{Phe}_{AAA}$) might be engineered to read UUU codons faster than wild-type ytRNA^{Phe}. If ytRNA$^{Phe}_{AAA}$ can then be charged selectively with an amino acid analog, one should be able to accomplish codon-biased incorporation of the analog at multiple sites in recombinant proteins. With respect to reading of UUC, an unmodified A in the first position of the anticodon is known to read codons ending with C, as well as U, in the absence of ytRNAs containing G in the

first anticodon position.[20-22] However, the binding of *E. coli* $tRNA^{Phe}_{GAA}$ should dominate the binding of $ytRNA^{Phe}_{AAA}$ owing to differences in the stability of A-C and G-C base pairs (see above).

The approach used here is a modification of the method introduced by Furter for site-specific insertion of amino acid analogs *in vivo*.[4] The method involves introduction into *E. coli* of a heterologous aminoacyl-tRNA synthetase and its cognate tRNA. If cross-charging between the heterologous pair and the translational apparatus of the host is slow or absent, and if the analog is charged only by the heterologous synthetase, insertion of the analog can be restricted (or at least biased) to sites characterized by productive base-pairing between the heterologous tRNA and the messenger RNA of interest.

In order to test these ideas, we prepared a yeast $tRNA^{Phe}$ ($ytRNA^{Phe}_{AAA}$) with an altered anticodon loop. The first base (G^{34}) of the $ytRNA^{Phe}$ anticodon (GAA) was replaced with A to provide specific Watson-Crick base-pairing to the UUU codon. Furthermore, G^{37} in the extended anticodon site was replaced with A in order to increase translational efficiency.[10, 23] We believe that charging of yeast $tRNA^{Phe}_{AAA}$ by *E. coli* PheRS can be ignored, because the aminoacylation rate of yeast $tRNA^{Phe}_{AAA}$ by *E. coli* PheRS is known to be 0.1% of that of *E.coli* $tRNA^{Phe}_{GAA}$.[24]

Since wild-type yeast PheRS does not activate amino acids significantly larger than phenylalanine, a mutant form of the synthetase with relaxed substrate specificity was prepared in order to accommodate L-3-(2-naphthyl)alanine (2Nal).[10] On the basis of prior work from this laboratory,[16] the mutant yeast PheRS (yPheRS (T415G)) was prepared by introduction of a Thr415Gly mutation in the α-subunit of the synthetase. The kinetics of activation of 2Nal and Phe by yPheRS (T415G) was analyzed *in vitro* via the

pyrophosphate exchange assay. The specificity constant (k_{cat}/K_M) for activation of 2Nal by yPheRS (T415G) was found to be $1.55x10^{-3}$ ($S^{-1}\mu M^{-1}$), 8-fold larger than that for Phe.[10] Therefore, when the ratio of 2Nal to Phe in the culture medium is high, $ytRNA^{Phe}_{AAA}$ should be charged predominantly with 2Nal. Recently, we have shown that 2Nal can be incorporated with better than 95% efficiency via amber suppression in an *E. coli* strain co-transformed with yPheRS (T415G) and $ytRNA^{Phe}_{CUA}$.

Murine dihydrofolate reductase (mDHFR), which contains nine Phe residues, was chosen as the test protein. The expression plasmid pQE16 encodes mDHFR under control of a bacteriophage T5 promoter; the protein is outfitted with a C-terminal hexahistidine tag to facilitate purification via immobilized metal affinity chromatography. In this construct, four of the Phe residues of mDHFR are encoded by UUC codons, five by UUU. A full-length copy of the yPheRS (T415G) gene, under control of a constitutive tac promoter, was inserted into pQE16. The gene encoding $ytRNA^{Phe}_{AAA}$ was inserted into the repressor plasmid pREP4 (Qiagen) under control of the constitutive promoter *lpp*. *E. coli* transformants harboring these two plasmids were incubated in Phe-depleted minimal medium supplemented with 3 mM 2Nal, and then treated with 1 mM IPTG to induce expression of mDHFR. Although the *E. coli* strain (K10-F6Δ) used in this study is a Phe auxotroph,[4] a detectable level of mDHFR was expressed even under conditions of nominal depletion of Phe (Figure 2), probably due to release of Phe through turnover of cellular proteins. In negative control experiments, mDHFR was expressed in the absence of either $ytRNA^{Phe}_{AAA}$ or yPheRS (T415G). mDFHR expression levels in these experiments were similar, indicating that neither $ytRNA^{Phe}_{AAA}$ nor yPheRS (T415G) significantly reduces the protein synthesis rate (Figure 2). However, MALDI-MS spectra

and amino acid analyses of purified mDHFRs showed differences among samples prepared under these conditions (Table 1). The molar mass of mDHFR prepared in the absence of 2Nal, ytRNA$^{Phe}_{AAA}$, or yPheRS (T415G) was 23,287 daltons, precisely that calculated for His-tagged mDHFR. However, when ytRNA$^{Phe}_{AAA}$ and yPheRS (T415G) were introduced into the expression strain and 2Nal was added to the culture medium, the observed mass of mDHFR was 23,537 daltons. Because each substitution of 2Nal for Phe leads to a mass increment of 50 daltons, this result is consistent with replacement of five Phe residues by 2Nal. No detectable mass shift was found in the absence of either ytRNA$^{Phe}_{AAA}$ or yPheRS (T415G), confirming that the intact heterologous pair is required for incorporation of 2Nal. For mDHFR isolated from the strain harboring the heterologous pair, amino acid analysis indicated replacement of 4.4 of the 9 Phe residues by 2Nal (Table 1). Without ytRNA$^{Phe}_{AAA}$ or yPheRS (T415G), no incorporation of 2Nal into mDHFR was detected by amino acid analysis.

Because neither MALDI nor amino acid analysis of intact mDHFR shows which Phe residues have been replaced by 2Nal, tryptic digests were analyzed to determine the occupancy of individual Phe sites. Tryptic digestion of mDHFR yields peptide fragments that are readily analyzed by MALDI mass spectrometry as shown in Figure 3. Peptide 1_{UUU} (residues 184-191, YKFEVYEK) contains a Phe residue encoded as UUU, whereas Peptide 2_{UUC} (residues 62-70, KTWFSIPEK) and Peptide 3_{UUC} (residues 26-39, NGDLPWPPLRNEFK) each contain a Phe residue encoded by UUC. In the absence of 2Nal, Peptide 1_{UUU} was detected with a monoisotopic mass of 1105.55 daltons, in accord with its theoretical mass (Figure 3A). However, when 2Nal was added, a strong signal at a mass of 1155.61 daltons was detected, and the 1105.55 signal was greatly reduced in

intensity (Figure 3B). As described earlier, each substitution of 2Nal for Phe leads to a mass increase of 50.06 daltons; the observed shift in the experimental mass is thus consistent with replacement of Phe by 2Nal in response to the UUU codon. Liquid chromatography-tandem mass spectrometry of Peptide 1_{UUU} (Nal) was also carried out in order to determine more directly the origin of the observed increase in mass. The fragment ion masses could be unambiguously assigned as shown in Figure 4, confirming replacement of Phe by 2Nal. The ratio of MALDI signal intensities, though not rigorously related to relative peptide concentrations, suggests that 2Nal incorporation is dominant at the UUU codon.

Similar analyses were conducted for Peptide 2_{UUC} and Peptide 3_{UUC}. In the absence of added 2Nal, the observed masses of Peptides 2_{UUC} and 3_{UUC} are 1135.61 (Figure 3A) and 1682.89 daltons (Figure 3D), respectively. These observed masses match well the corresponding theoretical masses (1135.61 daltons for Peptide 2_{UUC}, and 1682.86 daltons for Peptide 3_{UUC}). Upon addition of 2Nal to the expression medium, the signals at these masses (Figure 3B and 3E) were not substantially reduced, and only very weak signals were observed at masses of 1185.60 and at 1733.03, which would be expected for peptides containing 2Nal in place of Phe. 2Nal incorporation thus appears to be rare at UUC codons under the conditions used here for protein expression. Other peptides containing encoded Phe show similar codon-biased selective incorporation of 2Nal and Phe.

There is at least a formal possibility that the observed codon-biased incorporation of 2Nal might be dependent on codon context rather than - or in addition to - codon identity. In order to test this possibility, mutant mDHFR genes were prepared by

mutating the UUU codon in Peptide 1_{UUU} to UUC, and the UUC codon in Peptide 3_{UUC} to UUU. The resulting peptides were designated Peptide 1_{UUC} and Peptide 3_{UUU}, respectively. In Peptide 1_{UUC}, 2Nal incorporation was greatly reduced (Figure 3C), whereas for Peptide $3_{UUU,}$ 2Nal is readily detected (Figure 3F). 2Nal incorporation is unambiguously codon-biased to UUU.

Conclusions

The results described here show conclusively that a heterologous pair comprising a genetically engineered tRNA and cognate aminoacyl-tRNA synthetase can be used to break the degeneracy of the genetic code in *E. coli*. This method should provide a general strategy for multi-site incorporation of non-canonical amino acids, without the requirement that one of the natural amino acids be excluded. Introduction of more than one set of orthogonal pairs should allow several types of non-canonical amino acids to be incorporated in site-specific fashion. Ongoing experiments address the quantitative selectivity and the generality of the approach demonstrated here.

Acknowledgments

This work was supported by NIH GM62523 and by the NSF MRSEC program. We thank Dr. Mona Shahgholi, Dr. Gary M. Hathaway, and Dr. Jie Zhou for mass spectrometry studies. We thank Dr. Pin Wang and Dr. Soojin Son for helpful discussions.

References

1. Anderson, R. D.; Zhou, J.; Hecht, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 9674-9675.

2. Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537-539.

3. 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.

4. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

5. Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060-11064.

6. Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *J. Mol. Biol.* **2001**, *307*, 755-769.

7. Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.

8. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

9. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

10. Wang, P., Ph.D. thesis, California Institute of Technology, Pasadena, 2003.

11. Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.

12. Kast, P.; Hennecke, H. *J. Mol. Biol.* **1991**, *222*, 99-124.

13. Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37-40.

14. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

15. Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205-215.

16. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

17. Crick, F. H.; Brenner, S.; Watstobi, R. J.; Barnett, L. *Nature* **1961**, *192*, 1227-1232.

18. Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548-555.

19. Meroueh, M.; Chow, C. S. *Nucleic Acids Res.* **1999**, *27*, 1118-1125.

20. 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.

21. Chen, P.; Qian, Q.; Zhang, S. P.; Isaksson, L. A.; Bjork, G. R. *J. Mol. Biol.* **2002**, *317*, 481-492.

22. Inagaki, Y.; Kojima, A.; Bessho, Y.; Hori, H.; Ohama, T.; Osawa, S. *J. Mol. Biol.* **1995**, *251*, 486-492.

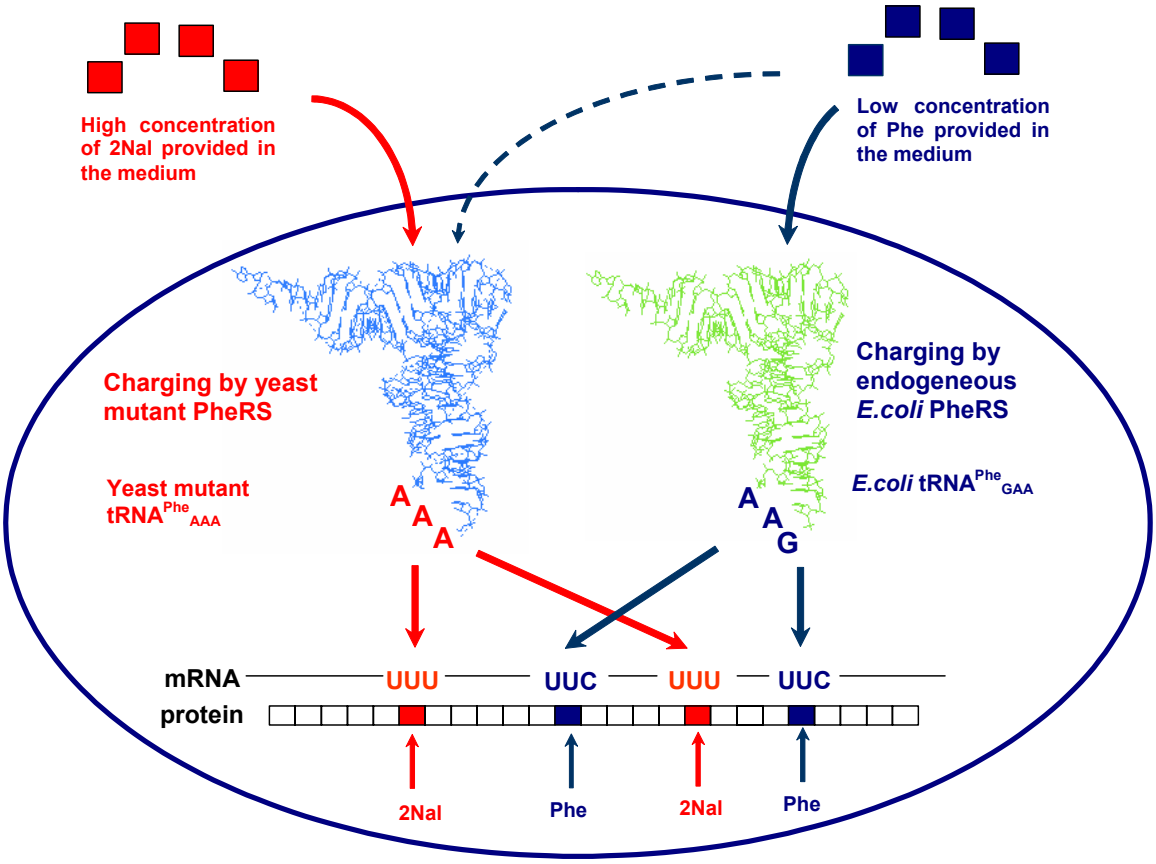
23. Yarus, M. *Science* **1982**, *218*, 646-652.

24. Peterson, E. T.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 10380-10389.

Table 1: Molar masses and numbers of 2Nal residues observed for mDHFR samples prepared under various conditions

$ytRNA^{Phe}_{AAA}$	+	+	-	+
yPheRS (T415G)	+	-	+	+
2Nal (3mM)	-	+	+	+
Phe (3mM)	+	-	-	-
Mass of intact mDHFR	23287	23287	23287	23537
Number of 2Nal residues	ND[a]	ND	ND	4.4

a. Not detected



Phenylalanine auxotrophic *E. coli* strain

Figure 1: A strategy for multi-site incorporation of 2Nal into recombinant proteins by breaking the degeneracy of the phenylalanine codons. ytRNA images were obtained from the Protein Data Bank (www.rcsb.org/pdb/).

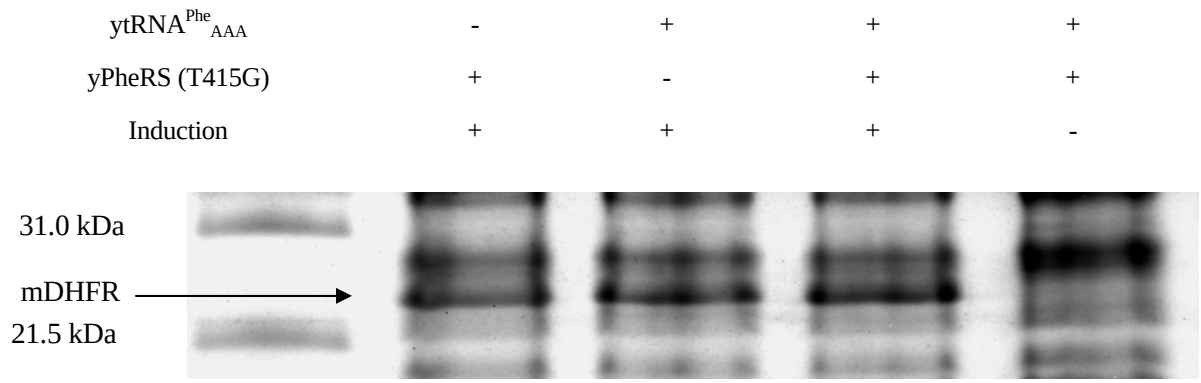


Figure 2: SDS-PAGE analysis of mDHFR prepared in minimal media supplemented with 3 mM 2Nal and free of exogenous Phe. Conditions are noted at the top of each lane. Lane 1 shows molecular weight standards.

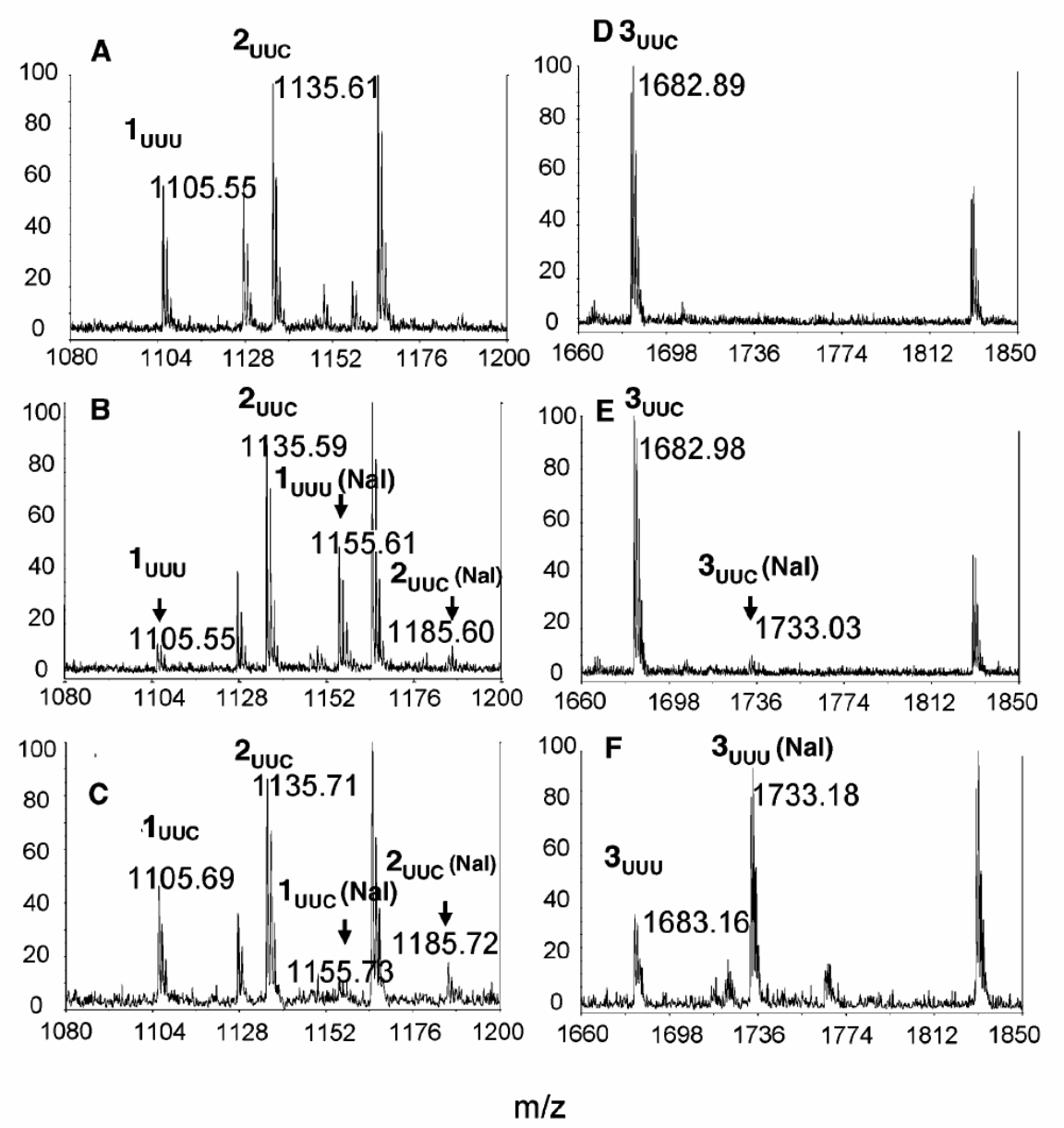

A
2UUC
1135.61
1UUU
1105.55
D 3UUC
1682.89
B
2UUC
1135.59
1UUU (NaI)
1155.61
1UUU
2UUC (NaI)
1105.55
1185.60
E 3UUC
1682.98
3UUC (NaI)
1733.03
C
2UUC
1135.71
1UUC
1105.69
2UUC (NaI)
1UUC (NaI)
1185.72
1155.73
F
3UUU (NaI)
1733.18
3UUU
1683.16
100
80
60
40
20
0
1080
1104
1128
1152
1176
1200
1660
1698
1736
1774
1812
1850
m/z

Figure 3: Replacement of Phe by 2Nal can be detected in MALDI-MS spectra of tryptic fragments of mDHFR. Peptide 1_{UUU} (residues 184-191, YKFEVYEK) contains a Phe (F) residue encoded by UUU, whereas in Peptide 1_{UUC} this codon has been mutated to UUC. Peptides 2 (residues 62-70, KTWFSIPEK) and 3 (residues 26-39, NGDLPWPPLRNEFK) are designated similarly. Peptide 1_{UUU} (Nal) refers to the form of the peptide containing 2Nal in place of Phe. In the absence of 2Nal, Peptide1_{UUU} was detected at a mass of 1105.55 (A). Upon addition of 2Nal, a strong signal corresponding to Peptide 1_{UUU}(Nal) was detected at m/z = 1155.61, and the signal for Peptide 1_{UUU} was greatly reduced (B). For Peptide 1_{UUC}, 2Nal incorporation was much less efficient (C). Signals corresponding to Peptides 2_{UUC} (B) and 3_{UUC} (E) were not substantially reduced upon addition of 2Nal, and only very weak signals for Peptides 2_{UUC} (Nal) and 3_{UUC} (Nal) were detected (B, E). When the UUC codon in Peptide 3_{UUC} is mutated to UUU, a strong signal for Peptide 3_{UUU} (Nal) is detected (F). These data confirm that incorporation of 2Nal is strongly biased to UUU codons.

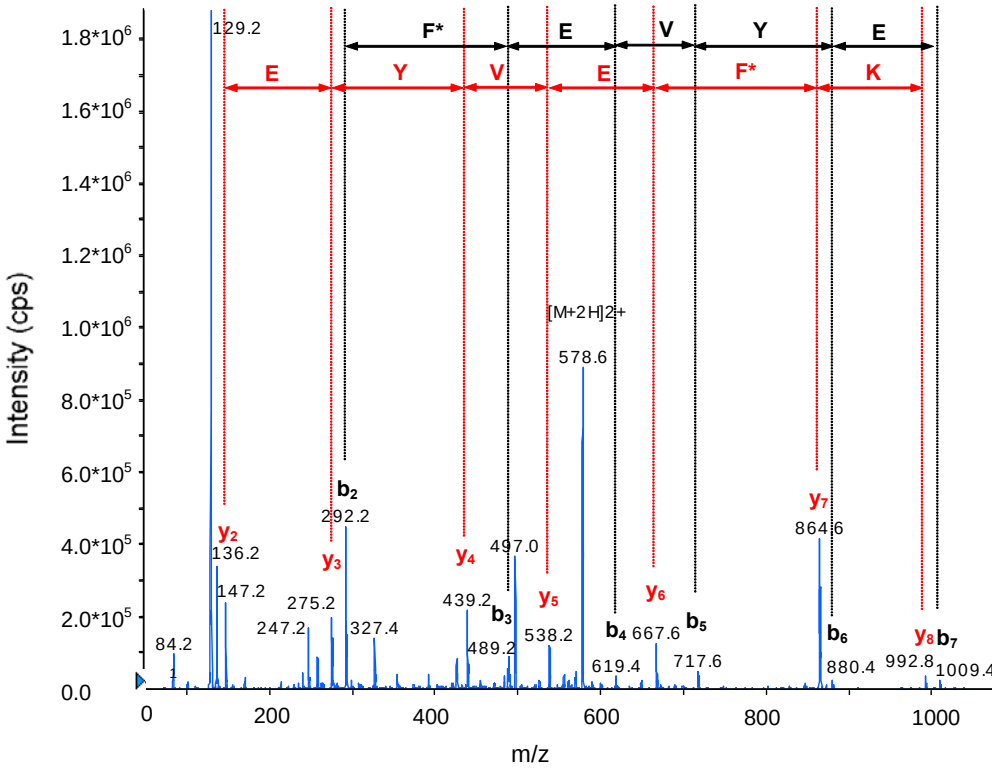


Figure 4: Tandem mass spectrum of Peptide 1_{UUU} (Nal) YKF*EVYEK. The doubly-charged ion at 578.6 daltons was selected and fragmented. The sequence of the peptide containing 2Nal (F*) can be read from the annotated b (black) or y (red) ion series.

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 aminoacyl-tRNA 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

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, yPheRS (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 high-throughput 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 finding[32-39] and stereochemical modeling[40-43] of codon-anticodon interactions. The expanded wobble rule indicates that A in the first position 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 mutagenesis[44] 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 *Pst*I 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 *Aat*II and *Nhe*I 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 re-assembly using sixteen primers (Primer 1–16). The GFP6 expression cassette was removed by restriction digestion with the *Aat*II and *Nhe*I, and inserted into pQE9_GFP3_yPheRS (T415G) between the *Aat*II and *Nhe*I sites to yield pQE9_GFP6_yPheRS (T415G). In order to introduce an extra copy of the *lac*I gene into pQE9_GFP6_yPheRS (T415G), the *lac*I gene expression cassette was amplified from pREP4 (Qiagen) with two primers (*lac*I_*Afl*II_f and *lac*I_*Sac*II_r). *Afl*II and *Sac*II restriction sites were introduced into pQE9_GFP6_yPheRS (T415G) by PCR reactions with two primers (pQE_*Afl*II_f and pQE_*Sac*II_r). The amplified *lac*I gene expression cassette was inserted into pQE9_GFP6_yPheRS (T415G) between the *Afl*II and *Sac*II 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 method[46] 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 yPheRS 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 *Nsi*I and *Bgl*II restriction enzymes and ligated to the large fragments of pQE9_GFP6_lacI_yPheRS (T415G) obtained by digestion with *Nsi*I and *Bgl*II. 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 μg/mL ampicillin and 35 μg/mL kanamycin. When OD_{600} 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, $MgSO_4$, $CaCl_2$, 20 amino acids (at 25 mg/L), antibiotics (35 μg/mL of kanamycin and 200 μg/mL of ampicillin). When the culture reached an OD_{600} 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 μL 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, $5x10^4$ 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 OD_{600} 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 OD_{600} 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 was removed from pQE9_GFP6_lacI_yPheRS_naph by digestion with *Nsi*I and *Bgl*II, and then ligated with a large fragment of pQE32-yPheRS (T415G) generated by digestion with *Nsi*I and *Bgl*II to yield pQE32-yPheRS_naph. Expression and purification of yPheRS variants were described previously.[13] The kinetics of activation of amino acids by yPheRS (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*-(2-hydroxethyl) piperazine-*N*'-(2-ethanesulfonic acid) (potassium-HEPES) (pH=7.6), 20 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 2 mM ATP, and 2 mM [^{32}P]- 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]

Construction of Plasmids and Expression Hosts for Incorporation of 2Nal at Phe 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 *Nsi*I and *Bgl*II 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 into DHF_AAA electrocompetent cells to construct DHF [pQE9_GFP6_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_AAA] and DHF [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 primers (2F_UUU_f/2F_UUU_r; 2F_UUC_f/2F_UUC_r) using pQE9_GFP6_lacI_yPheRS_naph as a template to generate pQE9_GFP6 (2FUUU)_lacI_yPheRS_naph and pQE9_GFP6 (2FUUC)_lacI_yPheRS_naph, 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. Both pQE9_GFP6 (2FUUU)_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 *Nsi*I and *Bgl*II digestion of pQE9_GFP6_lacI_SD_yPheRS_naph was inserted into pQE16_mDHFR (38Am)_yPheRS (T415G) between the *Nsi*I and *Bgl*II 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 co-

transformed 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 *Aat*II and *Nhe*I and inserted into pQE9_GFP6_lacI_yPheRS_naph between the *Aat*II and *Nhe*I 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 to yield MP [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 C-terminal hexa-histidine tag (GFP3C), a *Pst*I restriction site was added to pQE70 (Qiagen) using two complementary primers (*Pst*I_70_f/*Pst*I_70_r) to yield pQE70*. The coding sequence for GFP3 was excised by *Pst*I digestion from pQE9_GFP3_yPheRS (T415G) and inserted into pQE70* at the *Pst*I site to generate pQE70*_GFP3C. Expression cassette of GFP3C isolated from pQE70*_GFP3C was inserted 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 primers (K158_UUG_f/K158_UUG_r) using pQE70*_GFP3C_lacI_yPheRS_naph as a template to yield pQE70*_GFP3C (158UUG)_lacI_yPheRS_naph. Both pQE70*_GFP3C_lacI_yPheRS_naph and pQE70*_GFP3C (158UUG)_lacI_yPheRS_naph were co-transformed with pREP4_ytRNA^{Phe}_CAA into MPC390 competent cells to construct MP [pQE70*_GFP3C_lacI_yPheRS_naph/pREP4_ytRNA^{Phe}_CAA] and MP [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 $MgSO_4$, 1mM $CaCl_2$, 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 micro-sequencer (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 $(NH_4)_2CO_3$ 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 μM 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 $6x10^6$ yPheRS transformants (see Materials and Methods), a population somewhat larger than the theoretical library size (10^6). By transforming yPheRS library plasmids into DHF expression hosts, $1x10^7$ 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 $5x10^4$ 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 $5x10^4$ 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 yPheRS 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-site-specific 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 was achieved by AFWK [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 μM Phe, 50 μM Trp, 50 μM 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. **A**ccording 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 $\mathbf{6}_{UUA}$ was tested. Peptide $\mathbf{6}_{UUA}$ is the same as Peptide **6** except both Leu residues are encoded as UUA codons. Since the Peptide $\mathbf{6}_{UUA}$ variant containing 2Nal was not detected, we conclude that 2Nal incorporation is highly specific to the UUG codon.

UUG Codon-Specific Incorporation of 2Nal. Previously we 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 expressed in *E. coli* strain MP [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 μg/mL), 50 μM Phe, 50 μM Trp, 1.25 μM 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. 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

1. 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.

2. Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455-459.

3. 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.

4. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

5. 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.

7. 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.

20. 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.

29. 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.

33. 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.

35. 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.

37. 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.

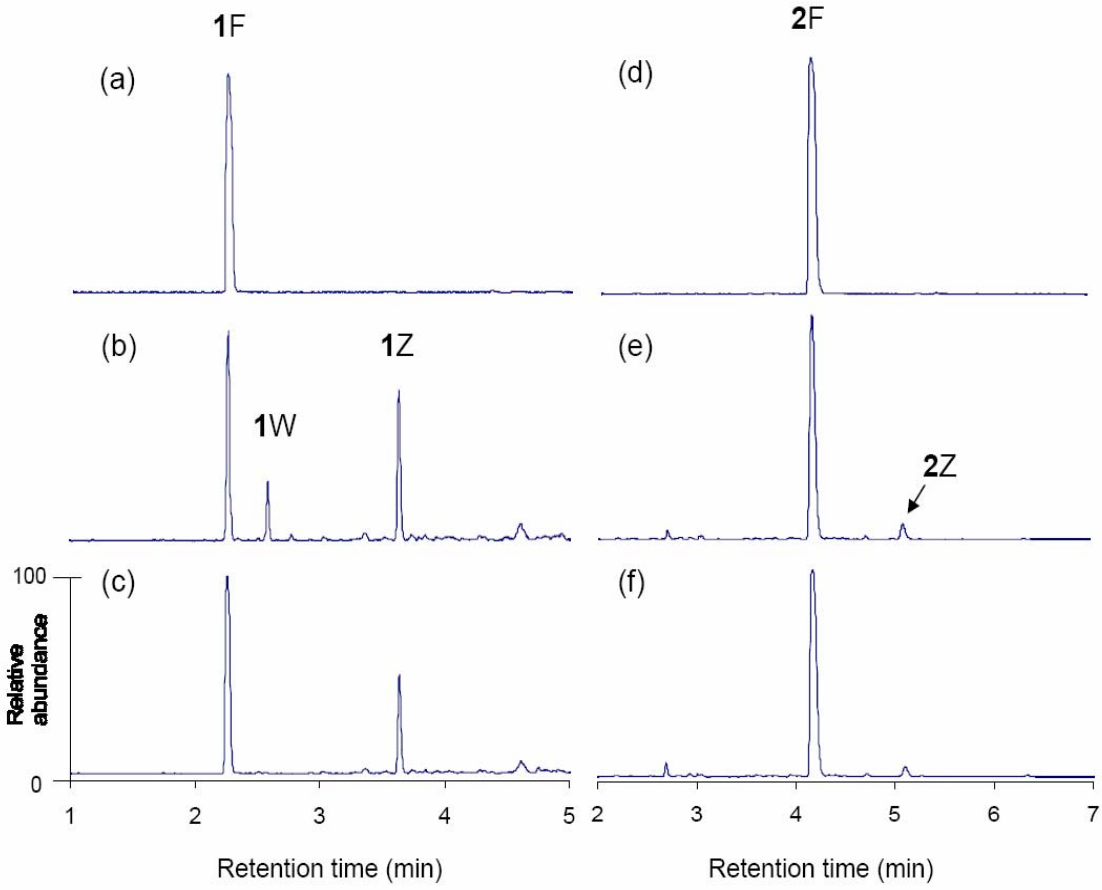

1F
(a)
2F
(d)
(b)
1Z
1W
(e)
2Z
100
(c)
(f)
Relative abundance
0
1
2
3
4
5
2
3
4
5
6
7
Retention time (min)
Retention time (min)

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.

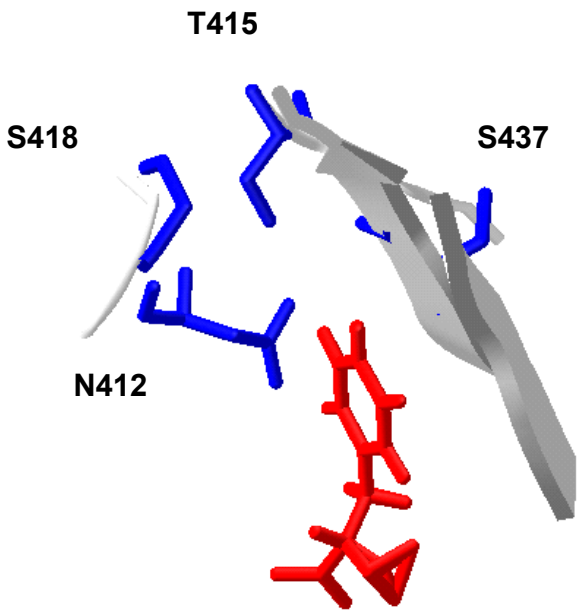


Figure 2: Phe substrate (red) and four residues (blue) within 7 Å of the *para*-position of the phenyl ring of the substrate inside the binding pocket of a homology model of yPheRS. These four residues were subjected to mutagenesis to generate a yPheRS library. The homology model of yPheRS was prepared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

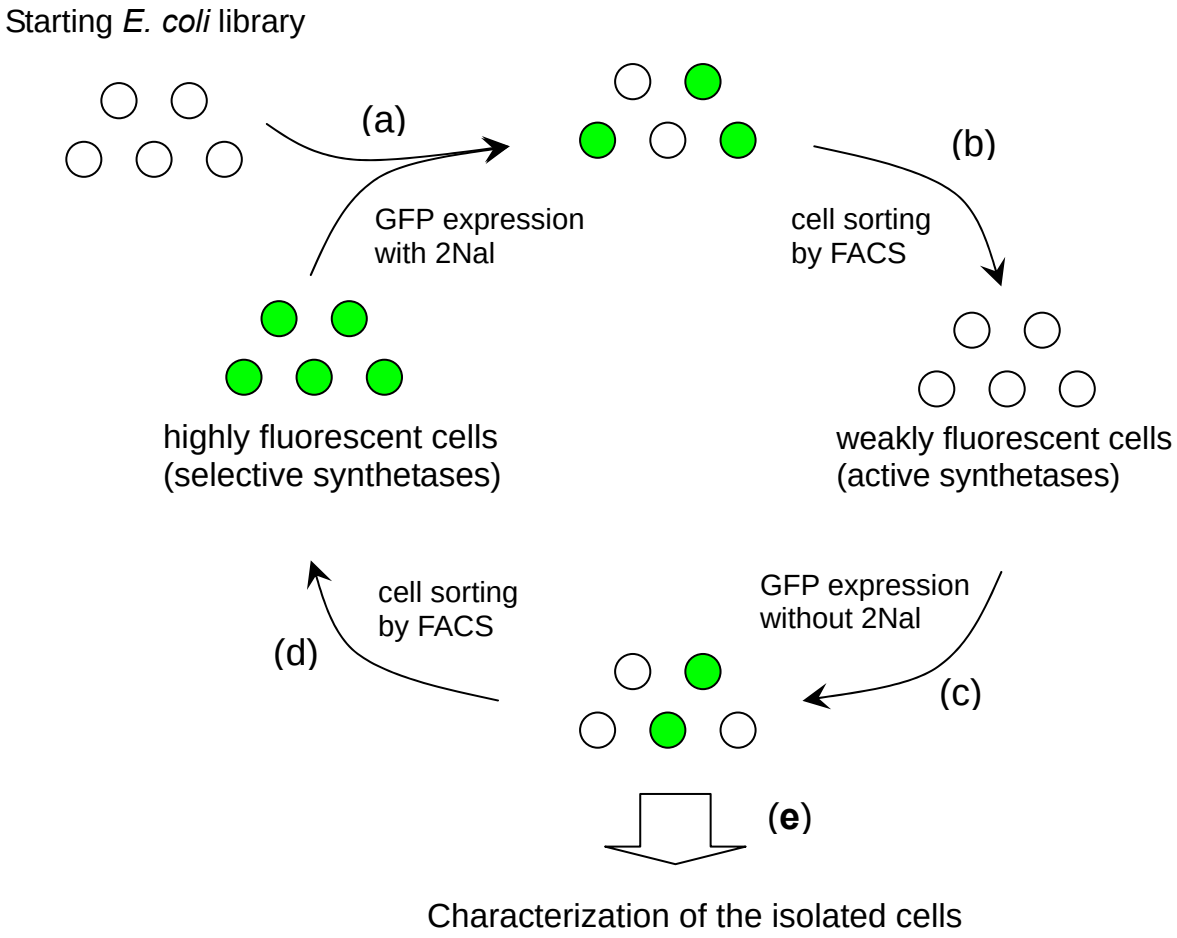


Figure 3: A screening scheme for yPheRS library. GFP6 in yPheRS expression library *E. coli* cells outfitted with ytRNA$^{Phe}_{AAA}$ and yPheRS library was expressed in the presence of 2Nal (a). Weakly fluorescent cells that contain active yPheRS variants were enriched by FACS (b). GFP6 in the collected 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).

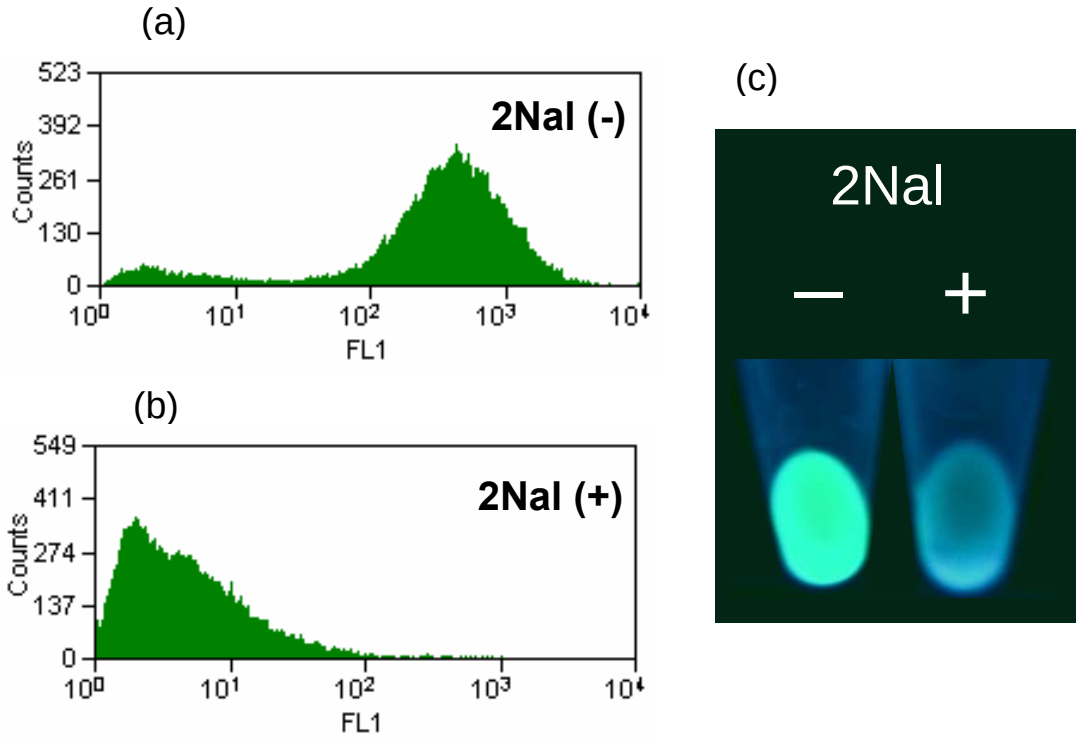


Figure 4: Fluorescence intensities of cells containing GFP6 expressed in minimal medium without 2NaI (a); with 2NaI (b). Visual comparison of fluorescence of cells containing GFP6 expressed in the absence of 2NaI (c left); in the presence of 2NaI (c right).

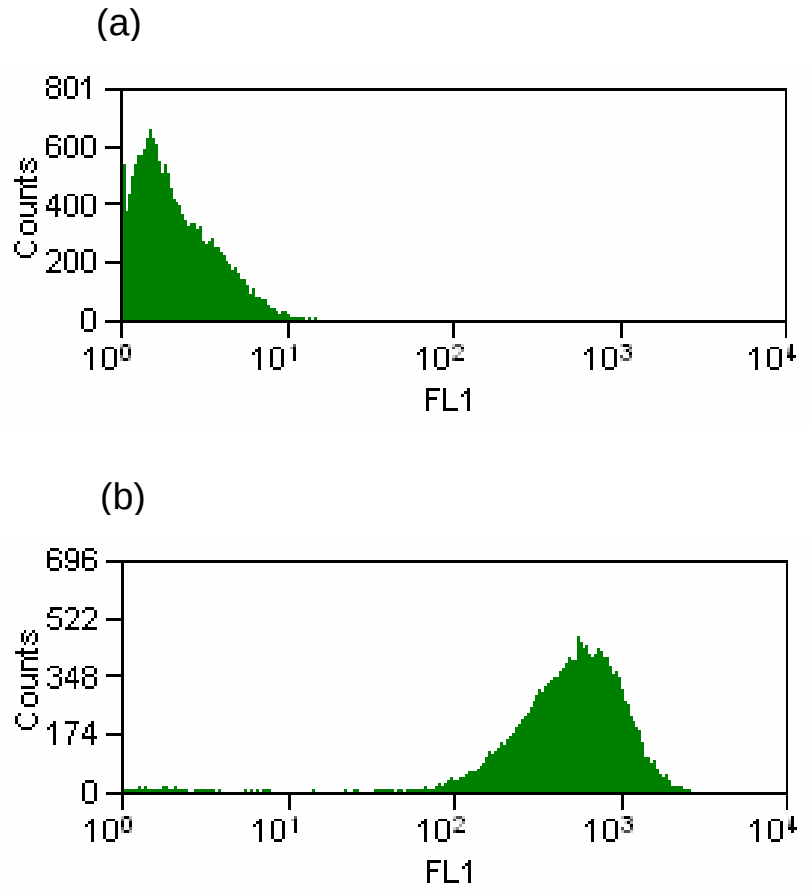


Figure 5: Fluorescence intensities of cells under uninduced conditions (a); cells expressing full length GFP6 (158Am) in the presence of 2Nal (b). The DHF *E. coli* strains were outfitted with ytRNA$^{Phe}_{CUA_UG}$ and yPheRS_naph. The expression media were supplemented with 18 amino acids (25 μg/mL),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

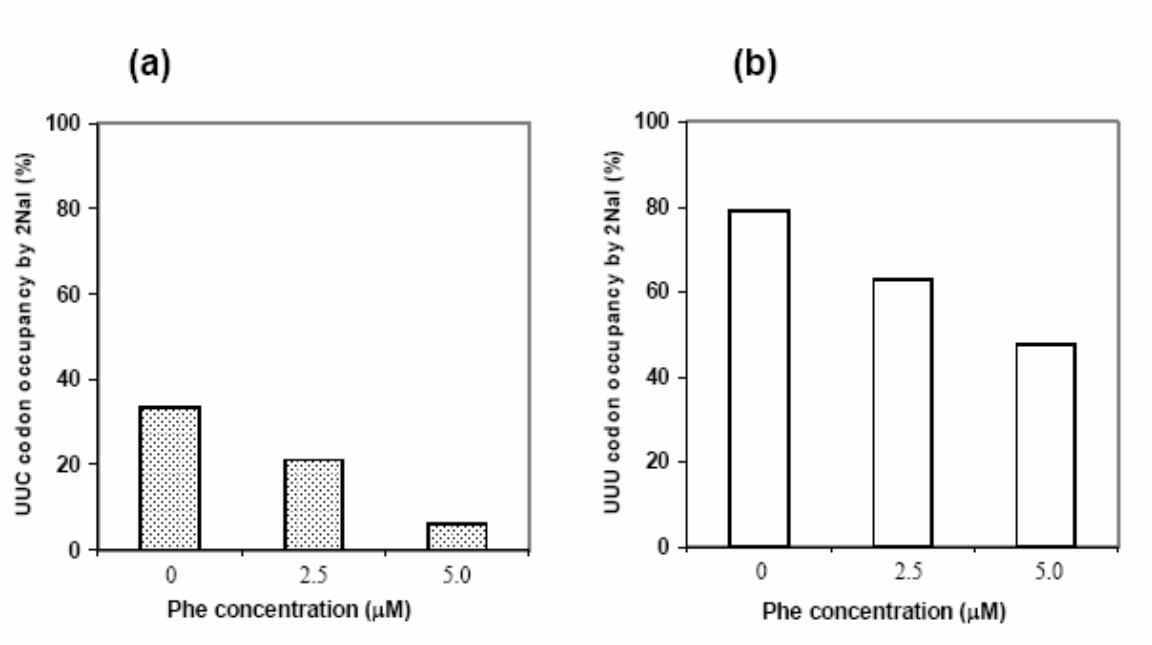


Figure 6: UUC and UUU codon occupancy by Phe and 2Nal. Both GFP6 (2UUC) and GFP6 (2UUU) were expressed in DHF expression hosts outfitted with yPheRS_naph and ytRNA$^{Phe}_{AAA}$ in minimal medium supplemented with 18 amino acids (25 μg/mL), 50 μM Trp, 3 mM 2Nal, and 0; 2.5 μM; 5.0 μM Phe. The UUC (a)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

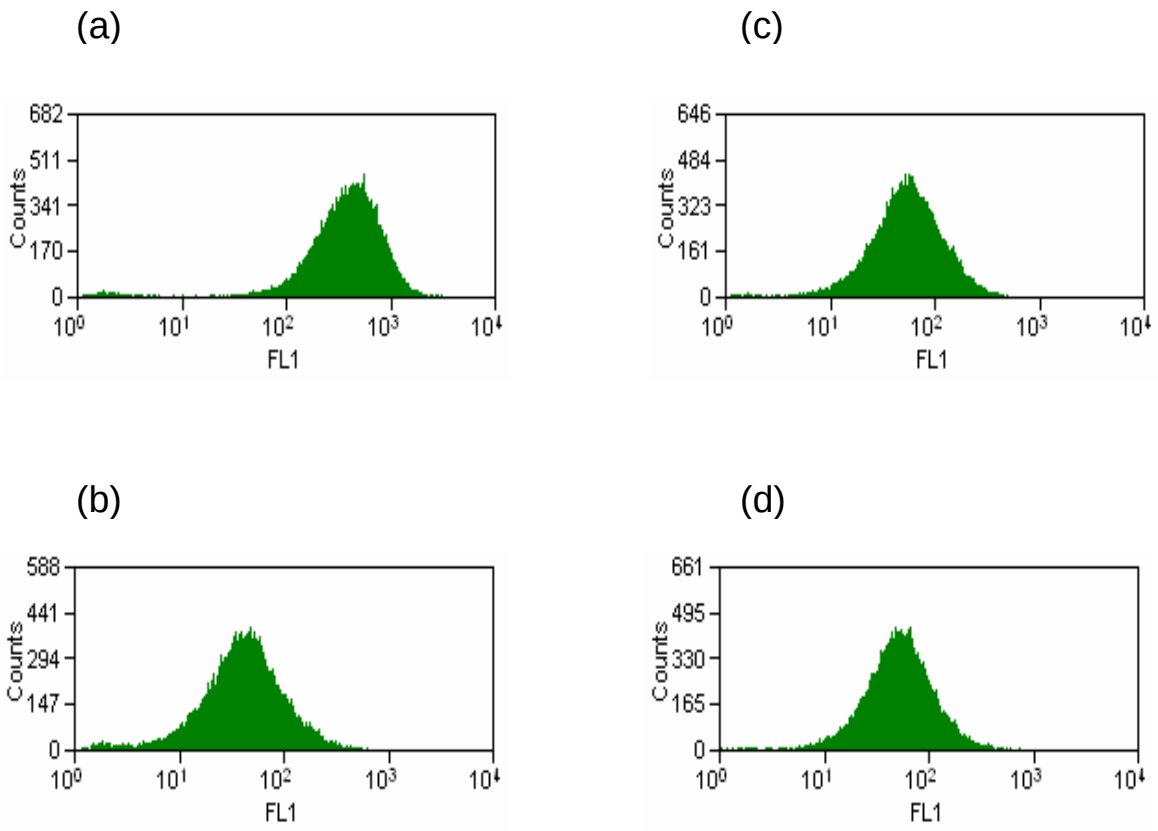


Figure 7: Fluorescence intensities of cells expressing GFP variants. GFP3_WC was expressed in DHF expression hosts outfitted with yPheRS_naph and ytRNA$^{Phe}_{AAA}$ in minimal medium supplemented with 18 amino acids, 5.0 μM Phe, 50 μM Trp, and no 2Nal (a); 3 mM 2Nal (b). GFP3 was expressed in MPC390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

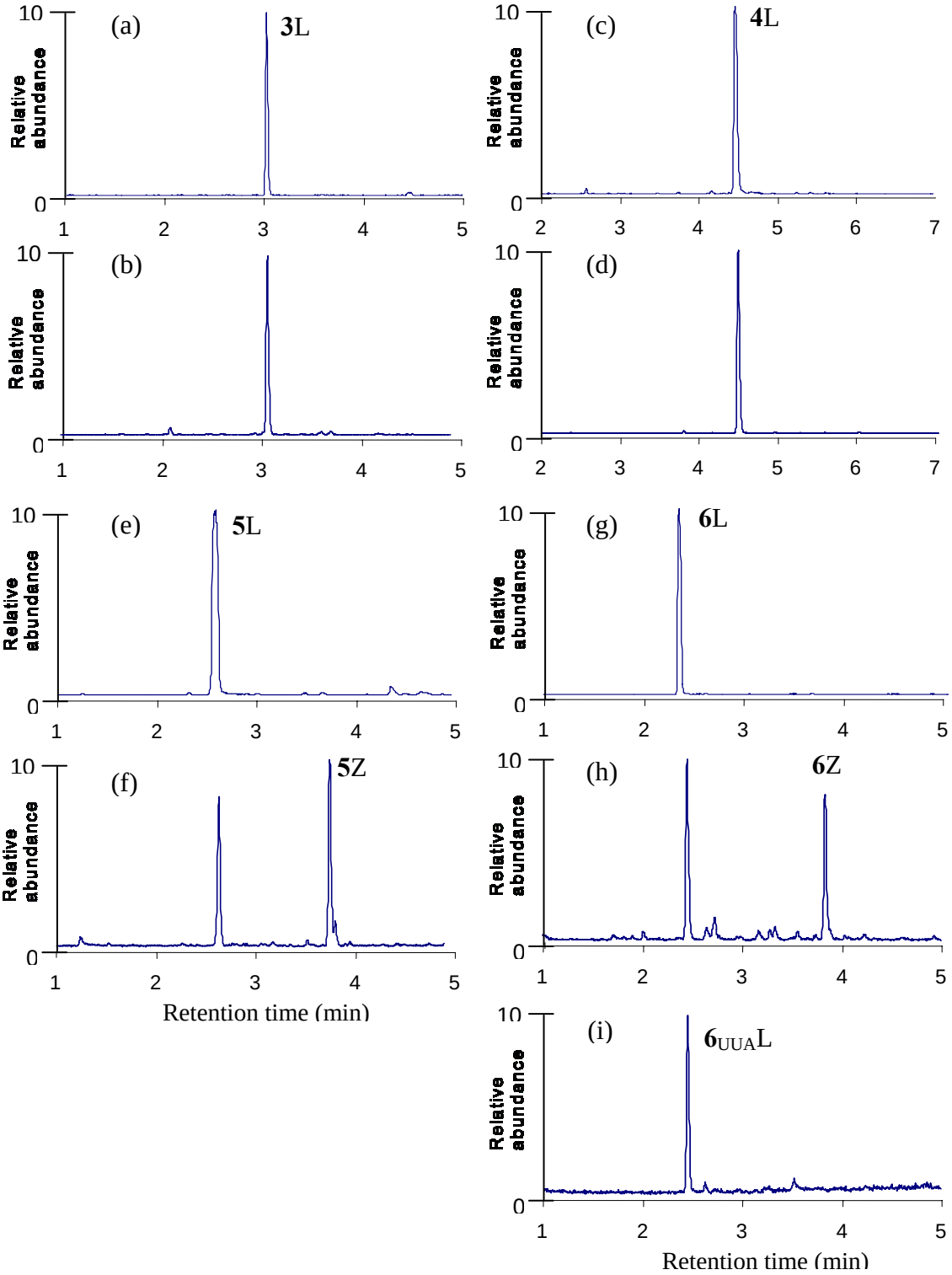
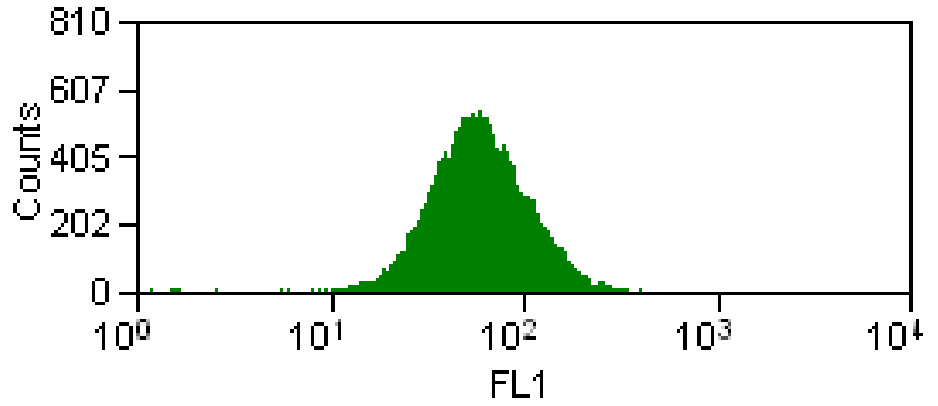

(a) 3L
(b)
(c) 4L
(d)
(e) 5L
(f) 5Z
(g) 6L
(h) 6Z
(i) 6UUAL
Relative abundance
Retention time (min)

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 $\mathbf{6}_{UUA}$ is the same as Peptide **6** except both Leu residues are encoded as UUA codon. Peptide **3; 4; 5; 6;** $\mathbf{6}_{UUA}$ variants containing Leu and 2Nal were designated **3**L and **3**Z; **4**L and **4**Z; **5**L and **5**Z; **6**L and **6**Z; $\mathbf{6}_{UUA}$L and $\mathbf{6}_{UUA}$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 **3**Z, **4**Z, or $\mathbf{6}_{UUA}$Z was detected by LC-MS analysis.

(a)



(b)

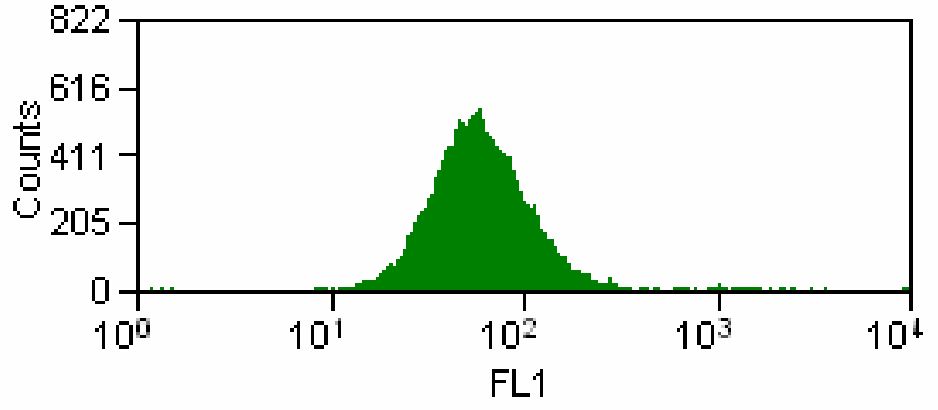


Figure 9: Fluorescence intensities of cells containing GFP3 (158UUG) expressed in minimal medium supplemented with 17 amino acids (25 μg/mL), 1.25 μM Leu, 5.0 μM Phe, and 50 μM Trp without 2Nal (a); with 3 mM 2Nal (b). MPC390 expression hosts were outfitted with yPheRS_naph and ytRNA$^{Phe}_{CAA}$.

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

Amino Acid	Enzyme	K_m (μM)	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 ± 1,100	**1**[a]
Trp	T415G	2.83 ± 1.6	0.153 ± 0.003	63,200 ± 34,600	**18**[a]
2Nal	T415G	7.03 ± 0.14	0.208 ± 0.04	29,500 ± 5,800	**8.4**[a]
Phe	naph	11,000 ± 2,700	0.0095 ± 0.0021	0.855 ± 0.007	**1**[b]
Trp	naph	1,400 ± 600	0.0035 ± 0.0009	2.52 ± 0.44	**2.9**[b]
2Nal	naph	2,000 ± 700	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.

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

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

[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.

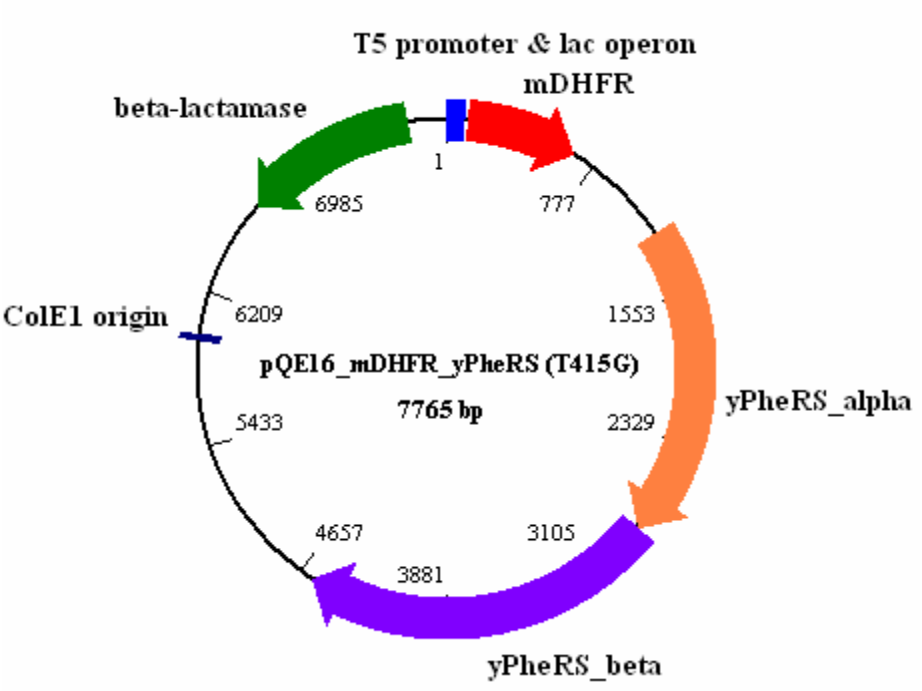
[d] The 38^{th} position in the amino acid sequence of mDHFR_38Am.

Appendix A

Plasmid Information

Following are sequences and information on plasmids described in the thesis.

1. pQE16_mDHFR_yPheRS (T415G) (7765 bp)



T5 promoter and lac operon	7 - 87 bp
mDHFR_His	115 - 660 bp
yPheRS_alpha subunit	1226 - 2787 bp
yPheRS_beta subunit	2802 - 4589 bp
ColE1 origin	5972 - 5997 bp
beta-lactamase	7560 - 6700 bp

Full sequence

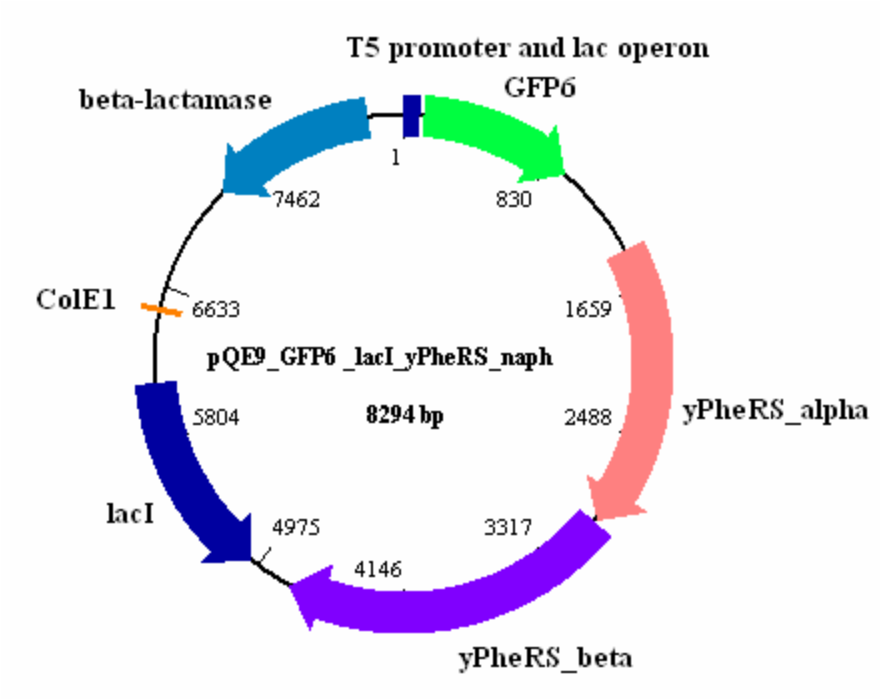
```
CTC GAG AAA TCA TAA AAA ATT TAT TTG CTT TGT GAG CGG ATA ACA ATT ATA ATA
GAT TCA ATT GTG AGC GGA TAA CAA TTT CAC ACA GAA TTC ATT AAA GAG GAG AAA
TTA ACT ATG AGA GGA TCC GGC ATC ATG GTT CGA CCA TTG AAC TCG ATC GTC GCC
GTG TCC CAA AAT ATG GGG ATT GGC AAG AAC GGA GAC CTA CCC TGG CCT CCG CTC
AGG AAC GAG TTC AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT TCA GTG GAA GGT
AAA CAG AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG AAG
AAT CGA CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA
```

CCA CCA CGA GGA GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT
ATT GAA CAA CCG GAA TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC
AGT TCT GTT TAC CAG GAA GCC ATG AAT CAA CCA GGC CAC CTT AGA CTC TTT GTG
ACA AGG ATC ATG CAG GAA TTT GAA AGT GAC ACG TTT TTC CCA GAA ATT GAT TTG
GGG AAA TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT GAG GTC CAG GAG
GAA AAA GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA GGT TCC AGA TCT
CAT CAC CAT CAC CAT CAC TAA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA
GAT CCA GTA ATG ACC TCA GAA CTC CAT CTG GAT TTG TTC AGA ACG CTC GGT TGC
CGC CGG GCG TTT TTT ATT GGT GAG AAT CCA AGC TAG CTT GGC GAG ATT TTC AGG
AGC TAA GGA AGC TAA AAT GGA GAA AAA AAT CAC TGG ATA TAC CAC CGT TGA TAT
ATC CCA ATG GCA TCG TAA AGA ACA TTT TGA GGC ATT TCA GTC AGT TGC TCA ATG
TAC CTA TAA CCA GAC CGT TCA GCT GGC ACG ACA GGT TTC CCG ACT GGA AAG CGG
GCA GTG AGC GCA ACG CAA TTA ATG TGA GTT AGC TCA CTC ATT AGG CAC CCC AGG
CTT TAC ACT TTA TGC TTC CGG CTC GTA TGT TGT GTG GAA TTG TGA GCG GAT AAC
AAT TTC ACA CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG CAT GCC TGC
AGT TGA CAA TTA ATC ATC GGC TCG TAT AAT GGA TCC AAT TGT GAG CGG AAT CGA
TTT TCA CAC AGG AAA CAG ACC ATG AAT CTA GAG ATG TCT GAC TTC CAA TTA GAA
ATT CTA AAG AAA CTA GAT GAA TTG GAT GAG ATC AAG TCC ACA CTG GCA ACT TTC
CCT CAG CAC GGC TCT CAA GAT GTT CTT TCC GCT TTG AAC TCT TTG AAA GCC CAC
AAC AAG TTA GAG TTT TCC AAG GTC GAC ACG GTT ACG TAT GAC TTG ACC AAA GAA
GGT GCT CAA ATT TTG AAT GAA GGT TCG TAC GAA ATT AAA CTA GTC AAG CTC ATC
CAA GAG TTG GGT CAA CTT CAA ATC AAA GAT GTG ATG TCC AAA CTA GGC CCT CAA
GTT GGT AAG GTC GGT CAG GCT AGA GCT TTC AAG AAC GGC TGG ATC GCC AAA AAC
GCC TCA AAC GAG CTT GAA CTC TCC GCA AAA TTG CAA AAT ACC GAT TTA AAT GAG
CTT ACT GAT GAA ACG CAA TCT ATT CTA GCG CAA ATC AAG AAC AAC TCG CAT CTG
GAT AGC ATT GAC GCC AAG ATT TTG AAC GAC TTG AAG AAA AGA AAG TTA ATT GCT
CAA GGT AAA ATC ACA GAT TTC AGT GTC ACC AAA GGG CCA GAG TTC TCG ACC GAC
CTC ACC AAA TTG GAA ACC GAT CTT ACC TCC GAC ATG GTC TCC ACC AAT GCA TAC
AAG GAC TTG AAG TTC AAG CCT TAC AAT TTC AAT TCT CAA GGT GTG CAA ATA TCT
TCA GGT GCT CTT CAC CCC TTA AAC AAA GTC AGA GAG GAA TTT AGA CAA ATT TTC
TTT TCC ATG GGA TTC ACA GAG ATG CCC TCG AAC CAA TAC GTC GAG ACA GGT TTC
TGG AAC TTC GAT GCC CTT TAC GTC CCA CAA CAG CAT CCT GCT CGT GAC CTG CAA
GAC ACT TTC TAC ATC AAG GAC CCA CTA ACC GCT GAG TTG CCC GAT GAC AAG ACA
TAC ATG GAC AAT ATC AAA GCC GTT CAC GAA CAG GGG AGA TTC GGG TCC ATC GGT
TAT CGT TAC AAC TGG AAG CCA GAA GAA TGT CAA AAA TTG GTC TTG AGA ACT CAC
TCC ACA GCC ATC TCT GCC AGA ATG CTG CAC GAT TTG GCC AAA GAT CCA AAG CCC
ACC AGA TTG TTT TCT ATC GAC CGT GTT TTC CGT AAC GAA GCA GTT GAC GCC ACC
CAT TTG GCC GAA TTC CAC CAG GTG GAA GGT GTT CTT GCC GAC TAC AAC ATT ACT
CTG GGT GAC CTG ATC AAG TTC ATG GAA GAG TTT TTC GAA AGA ATG GGT GTC ACC
GGT TTG AGA TTC AAG CCT ACC TAC AAT CCT TAC GGC GAG CCA TCA ATG GAA ATC
TTT TCT TGG CAC GAA GGT TTG CAA AAA TGG GTC GAA ATC GGT AAC TCT GGT ATG
TTC AGA CCA GAA ATG CTC GAG TCC ATG GGT CTA CCA AAG GAT CTA AGA GTC CTT
GGT TGG GGG TTA TCC TTG GAA AGA CCT ACC ATG ATC AAA TAT AAG GTT CAA AAC
ATC AGA GAA CTG TTA GGT CAT AAA GTC TCT TTG GAC TTT ATC GAA ACC AAT CCT
GCT GCT AGA TTG GAC GAA GAC TTG TAC GAA TAA GGC AGG AAT AGA TTA TGC CTA
CCG TCT CCG TGA ACA AGC AGC AAT TAT TTG ATC TTC TAG GCA AAG ACT ACA CTT
CCC AAG AGT TCG ACG AAT TAT GTT TTG AAT TCG GTA TGG AAA TGG ACG AAG ACA
CCA CAG AAG AGG CCT TGA AAA CCG GGG AGG AGC CGG AAT TGA AGC TTG ATA TCA
GTG CCA ATC GTT ACG ATT TGC TTT GTA TCG AAG GTA TTT CAC AGT CGC TGA ACG
AAT ACT TGG AAC GTA AAG AAA GAC CTG ACT ATA AAT TAA GCA AGC CAA CCA CTA
AGT TGA TCA TCG ACA AAT CAA CGG AGC AAA TTA GAC CTT TTG CTA CCG CTG CTG
TAT TGA GAA ATA TCA AGC TTA ACG AAA AAT CTT ACG CTT CTT TTA TTG CCT TGC
AAG ATA AAT TAC ATG CCA ATC TAT GTA GAA ACA GAA GCT TGG TTG CCA TGG GTA
CTC ACG ATT TAG ATT CAA TTG AAG GTC CAT TCC ATT ACA GAG CTC TAC CAC CAA
AGG ACA TCA AGT TCG TAC CAT TGA ATC AAA CCC AAG AGT TTA CTG GTG ACA AAT
TGA TCG AGT TTT ATA AAT CTC CAG AAC AGA AAA ACA ACA TAG GGA GAT ACG TTC
ACA TTA TTG AGG ATT CTC CAG TCT TCC CAG TTA TTA TGG ACA GCA AAG ATC GTG

TTT GCT CCC TGC CAC CAT TAA TCA ATA GTG AAC ATT CGA AGA TCT CTG TGA ACA
CCC GTA ACA TTT TGA TTG ATA TAA CCG CCA CCG ATA AGA CCA AAG CCG AGA TCG
TTT TGA ACA TAT TAA CTA CAA TGT TCT CAC GTT ATT GTG ACG AAC CAT TCA CGG
TTG AGC CTG TAG AAA TTG TCT CTG AAC ACA ATG GCC AAT CCC GTT TGG CGC CAA
ACT TCA ACG ATA GAA TTA TGG ATG TCT CCA TTA AGT ACA TCA ACT CCT GTC TTG
GCC TAG ATC AAT CCG CTG ATG AAA TTG CTC ATT GTC TAA AAA AGA TGT CGT TGC
ATG CCG TTC AAT CAA AGG AAG ACA AGG ACA TCT TGC ACG TTG ACA TTC CGG TAA
CTA GAC CTG ATA TTT TGC ACG CTT GTG ATA TAA TGG AAG ATG CCG CTG TCG GTT
ATG GTT TCA ATA ATT TGC CAA AGG GTG AGA AAT TAT CCA ATG CCA ACT TCA TTG
CCA AAC CAT TAC CAA TCA ACA AGG TTT CTG ATA TTT TCA GAG TTG CAT CCT CTC
AAG CCA CGT GGG TTG AGG TTT TAC CAT TGA CCT TAT GTT CGC ACG ATG AAA ACT
TTA AAT TTC TAA GAC AAT CCG ACA ATG ATG ATT TAG CTG TCA AAT TGG CCA ACC
CAA AGA CTT TGG AAT ACC AAG TTG TTA GAA CCA CTT TAT TGC CTG GTA TCT TAA
AGA CAG TCA AGG AAA ACA GAA AAC ATT CCT TAC CAA TCA AAG TCT TTG AAA CCG
GTG ACG TTG TAT TTA AAG ACG ACA AAC TAG AAA GGA AGG CGT ACA ATG AAC GTC
ACT GGG CTG CCA TCT ACG TGG GTA AGA ATT CTG GGT TTG AAA TCA TTC AAG GGT
TAT TGG GTA AAA TCA TGC AAA CTT TTA GAA CAG AGT GGA TTG CAG ACT ACG GTG
CTG CTG CTT CTG GCA GAG GTT ACT GGA TTG AAG AAG ACG ATT CTG TGA AAA CCT
ATT TCC CAG GTA GAG GTG CCA AGG TCA TGT TCA GAT CCA AAG AAG GCG CTG AGC
CAA AGC AAA TCG GCC ACT TGG GTG TCT TGC ATC CTG AAG TCA TGA TGA ATT TCG
ACG TTC CAT TCG CTG CAT CCT TTG TAG AGG TTA ATG CCG AAG TCT TCC TAT AAT
GTA ATG TTC TAA CAA AAA TTT TTA CTG ATT TAT AAA ACT TAT ATA GAT AGA TAG
ACA TAT ATA TAT GGG TAC CGA GCT CGA ATT CAC TGG CCG TCG TTT TAC AAC GTC
GTG ACT GGG AAA ACC CTG GCG TTA CCC AAC TTA ATC GCC TTG CAG CAC ATC CCC
CTT TCG CCA GCT GGA TAT TAC GGC CTT TTT AAA GAC CGT AAA GAA AAA TAA GCA
CAA GTT TTA TCC GGC CTT TAT TCA CAT TCT TGC CCG CCT GAT GAA TGC TCA TCC
GGA ATT TCG TAT GGC AAT GAA AGA CGG TGA GCT GGT GAT ATG GGA TAG TGT TCA
CCC TTG TTA CAC CGT TTT CCA TGA GCA AAC TGA AAC GTT TTC ATC GCT CTG GAG
TGA ATA CCA CGA CGA TTT CCG GCA GTT TCT ACA CAT ATA TTC GCA AGA TGT GGC
GTG TTA CGG TGA AAA CCT GGC CTA TTT CCC TAA AGG GTT TAT TGA GAA TAT GTT
TTT CGT CTC AGC CAA TCC CTG GGT GAG TTT CAC CAG TTT TGA TTT AAA CGT GGC
CAA TAT GGA CAA CTT CTT CGC CCC CGT TTT CAC CAT GGG CAA ATA TTA TAC GCA
AGG CGA CAA GGT GCT GAT GCC GCT GGC GAT TCA GGT TCA TCA TGC CGT CTG TGA
TGG CTT CCA TGT CGG CAG AAT GCT TAA TGA ATT ACA ACA GTA CTG CGA TGA GTG
GCA GGG CGG GGC GTA ATT TTT TTA AGG CAG TTA TTG GTG CCC TTA AAC GCC TGG
GGT AAT GAC TCT CTA GCT TGA GGC ATC AAA TAA AAC GAA AGG CTC AGT CGA AAG
ACT GGG CCT TTC GTT TTA TCT GTT GTT TGT CGG TGA ACG CTC TCC TGA GTA GGA
CAA ATC CGC CGC TCT AGA GCT GCC TCG CGC GTT TCG GTG ATG ACG GTG AAA ACC
TCT GAC ACA TGC AGC TCC CGG AGA CGG TCA CAG CTT GTC TGT AAG CGG ATG CCG
GGA GCA GAC AAG CCC GTC AGG GCG CGT CAG CGG GTG TTG GCG GGT GTC GGG GCG
CAG CCA TGA CCC AGT CAC GTA GCG ATA GCG GAG TGT ATA CTG GCT TAA CTA TGC
GGC ATC AGA GCA GAT TGT ACT GAG AGT GCA CCA TAT GCG GTG TGA AAT ACC GCA
CAG ATG CGT AAG GAG AAA ATA CCG CAT CAG GCG CTC TTC CGC TTC CTC GCT CAC
TGA CTC GCT GCG CTC GGT CTG TCG GCT GCG GCG AGC GGT ATC AGC TCA CTC AAA
GGC GGT AAT ACG GTT ATC CAC AGA ATC AGG GGA TAA CGC AGG AAA GAA CAT GTG
AGC AAA AGG CCA GCA AAA GGC CAG GAA CCG TAA AAA GGC CGC GTT GCT GGC GTT
TTT CCA TAG GCT CCG CCC CCC TGA CGA GCA TCA CAA AAA TCG ACG CTC AAG TCA
GAG GTG GCG AAA CCC GAC AGG ACT ATA AAG ATA CCA GGC GTT TCC CCC TGG AAG
CTC CCT CGT GCG CTC TCC TGT TCC GAC CCT GCC GCT TAC CGG ATA CCT GTC CGC
CTT TCT CCC TTC GGG AAG CGT GGC GCT TTC TCA ATG CTC ACG CTG TAG GTA TCT
CAG TTC GGT GTA GGT CGT TCG CTC CAA GCT GGG CTG TGT GCA CGA ACC CCC CGT
TCA GCC CGA CCG CTG CGC CTT ATC CGG TAA CTA TCG TCT TGA GTC CAA CCC GGT
AAG ACA CGA CTT ATC GCC ACT GGC AGC AGC CAC TGG TAA CAG GAT TAG CAG AGC
GAG GTA TGT AGG CGG TGC TAC AGA GTT CTT GAA GTG GTG GCC TAA CTA CGG CTA
CAC TAG AAG GAC AGT ATT TGG TAT CTG CGC TCT GCT GAA GCC AGT TAC CTT CGG
AAA AAG AGT TGG TAG CTC TTG ATC CGG CAA ACA AAC CAC CGC TGG TAG CGG TGG
TTT TTT TGT TTG CAA GCA GCA GAT TAC GCG CAG AAA AAA AGG ATC TCA AGA AGA

TCC TTT GAT CTT TTC TAC GGG GTC TGA CGC TCA GTG GAA CGA AAA CTC ACG TTA
AGG GAT TTT GGT CAT GAG ATT ATC AAA AAG GAT CTT CAC CTA GAT CCT TTT AAA
TTA AAA ATG AAG TTT TAA ATC AAT CTA AAG TAT ATA TGA GTA AAC TTG GTC TGA
CAG TTA CCA ATG CTT AAT CAG TGA GGC ACC TAT CTC AGC GAT CTG TCT ATT TCG
TTC ATC CAT AGC TGC CTG ACT CCC CGT CGT GTA GAT AAC TAC GAT ACG GGA GGG
CTT ACC ATC TGG CCC CAG TGC TGC AAT GAT ACC GCG AGA CCC ACG CTC ACC GGC
TCC AGA TTT ATC AGC AAT AAA CCA GCC AGC CGG AAG GGC CGA GCG CAG AAG TGG
TCC TGC AAC TTT ATC CGC CTC CAT CCA GTC TAT TAA TTG TTG CCG GGA AGC TAG
AGT AAG TAG TTC GCC AGT TAA TAG TTT GCG CAA CGT TGT TGC CAT TGC TAC AGG
CAT CGT GGT GTC ACG CTC GTC GTT TGG TAT GGC TTC ATT CAG CTC CGG TTC CCA
ACG ATC AAG GCG AGT TAC ATG ATC CCC CAT GTT GTG CAA AAA AGC GGT TAG CTC
CTT CGG TCC TCC GAT CGT TGT CAG AAG TAA GTT GGC CGC AGT GTT ATC ACT CAT
GGT TAT GGC AGC ACT GCA TAA TTC TCT TAC TGT CAT GCC ATC CGT AAG ATG CTT
TTC TGT GAC TGG TGA GTA CTC AAC CAA GTC ATT CTG AGA ATA GTG TAT GCG GCG
ACC GAG TTG CTC TTG CCC GGC GTC AAT ACG GGA TAA TAC CGC GCC ACA TAG CAG
AAC TTT AAA AGT GCT CAT CAT TGG AAA ACG TTC TTC GGG GCG AAA ACT CTC AAG
GAT CTT ACC GCT GTT GAG ATC CAG TTC GAT GTA ACC CAC TCG TGC ACC CAA CTG
ATC TTC AGC ATC TTT TAC TTT CAC CAG CGT TTC TGG GTG AGC AAA AAC AGG AAG
GCA AAA TGC CGC AAA AAA GGG AAT AAG GGC GAC ACG GAA ATG TTG AAT ACT CAT
ACT CTT CCT TTT TCA ATA TTA TTG AAG CAT TTA TCA GGG TTA TTG TCT CAT GAG
CGG ATA CAT ATT TGA ATG TAT TTA GAA AAA TAA ACA AAT AGG GGT TCC GCG CAC
ATT TCC CCG AAA AGT GCC ACC TGA CGT CTA AGA AAC CAT TAT TAT CAT GAC ATT
AAC CTA TAA AAA TAG GCG TAT CAC GAG GCC CTT TCG TCT TCA C

2. pQE9_GFP6_lacI_yPheRS_naph (8,294 bp)



T5 promoter and lac operon	7 - 87 bp
His_GFP6	115 - 921 bp
yPheRS_alpha subunit	1449 - 2972 bp
yPheRS_beta subunit	2987 - 4774 bp
ColE1 origin	6501 - 6526 bp
beta-lactamase	8089 - 7229 bp

Full sequence

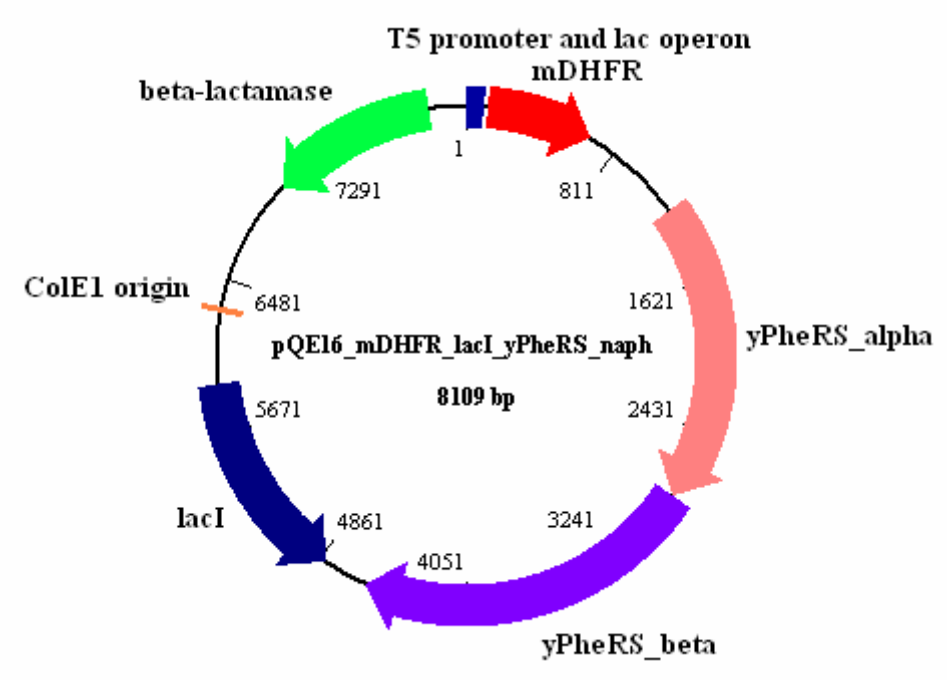
```
CTC GAG AAA TCA TAA AAA ATT TAT TTG CTT TGT GAG CGG ATA ACA ATT ATA ATA
GAT TCA ATT GTG AGC GGA TAA CAA TTT CAC ACA GAA TTC ATT AAA GAG GAG AAA
TTA ACT ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GTC GAC CTG CAG
CCC CGG GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTT ACT GGA GTT
GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC
AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT
TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTT ACC
TAT GGT GTT CAA TGC TTT TCC CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTT
TTT AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TTT TTT AAA
GAT GAC GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT
GTT AAT CGT ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTC
GGA CAC AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC
```

AAA CAA AAG AAT GGA ATC AAA GCT AAC TTT AAA ATT CGC CAC AAC ATT GAA GAT
GGA TCC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC
CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCG ACA CAA TCT GCC CTT TCG AAA
GAT CCC AAC GAA AAG CGT GAC CAC ATG GTC CTT CTT GAG TTT GTA ACT GCT GCT
GGG ATT ACA CAT GGC ATG GAT GAG CTC TAC AAA CTG CAG CCA AGC TTA ATT AGC
TGA AGC TTG GAC TCC TGT TGA TAG ATC CAG TAA TGA CCT CAG AAC TCC ATC TGG
ATT TGT TCA GAA CGC TCG GTT GCC GCC GGG CGT TTT TTA TTG GTG AGA ATC CAA
GCT AGC TTG GCG AGA TTT TCA GGA GCT AAG GAA GCT AAA ATG GAG AAA AAA ATC
ACT GGA TAT ACC ACC GTT GAT ATA TCC CAA TGG CAT CGT AAA GAA CAT TTT GAG
GCA TTT CAG TCA GTT GCT CAA TGT ACC TAT AAC CAG ACC GTT CAG CTG GCA CGA
CAG GTT TCC CGA CTG GAA AGC GGG CAG TGA GCG CAA CGC AAT TAA TGT GAG TTA
GCT CAC TCA TTA GGC ACC CCA GGC TTT ACA CTT TAT GCT TCC GGC TCG TAT GTT
GTG TGG AAT TGT GAG CGG ATA ACA ATT TCA CAC AGG AAA CAG CTA TGA CCA TGA
TTA CGC CAA GCT TGC ATG CCT GCA GTT GAC AAT TAA TCA TCG GCT CGT ATA ATG
GAT CCA ATT GTG AGC GGA ATC GAT TTT CAC ACA GGA AAC AGA CCA TGA ATC TAG
AGA TGT CTG ACT TCC AAT TAG AAA TTC TAA AGA AAC TAG ATG AAT TGG ATG AGA
TCA AGT CCA CAC TGG CAA CTT TCC CTC AGC ACG GCT CTC AAG ATG TTC TTT CCG
CTT TGA ACT CTT TGA AAG CCC ACA ACA AGT TAG AGT TTT CCA AGG TCG ACA CGG
TTA CGT ATG ACT TGA CCA AAG AAG GTG CTC AAA TTT TGA ATG AAG GTT CGT ACG
AAA TTA AAC TAG TCA AGC TCA TCC AAG AGT TGG GTC AAC TTC AAA TCA AAG ATG
TGA TGT CCA AAC TAG GCC CTC AAG TTG GTA AGG TCG GTC AGG CTA GAG CTT TCA
AGA ACG GCT GGA TCG CCA AAA ACG CCT CAA ACG AGC TTG AAC TCT CCG CAA AAT
TGC AAA ATA CCG ATT TAA ATG AGC TTA CTG ATG AAA CGC AAT CTA TTC TAG CGC
AAA TCA AGA ACA ACT CGC ATC TGG ATA GCA TTG ACG CCA AGA TTT TGA ACG ACT
TGA AGA AAA GAA AGT TAA TTG CTC AAG GTA AAA TCA CAG ATT TCA GTG TCA CCA
AAG GGC CAG AGT TCT CGA CCG ACC TCA CCA AAT TGG AAA CCG ATC TTA CCT CCG
ACA TGG TCT CCA CCA ATG CAT ACA AGG ACT TGA AGT TCA AGC CTT ACA ATT TCA
ATT CTC AAG GTG TGC AAA TAT CTT CAG GTG CTC TTC ACC CCT TAA ACA AAG TCA
GAG AGG AAT TTA GAC AAA TTT TCT TTT CCA TGG GAT TCA CAG AGA TGC CCT CGA
ACC AAT ACG TCG AGA CAG GTT TCT GGA ACT TCG ATG CCC TTT ACG TCC CAC AAC
AGC ATC CTG CTC GTG ACC TGC AAG ACA CTT TCT ACA TCA AGG ACC CAC TAA CCG
CTG AGT TGC CCG ATG ACA AGA CAT ACA TGG ACA ATA TCA AAG CCG TTC ACG AAC
AGG GGA GAT TCG GGT CCA TCG GTT ATC GTT ACA ACT GGA AGC CAG AAG AAT GTC
AAA AAT TGG TCT TGA GAA CTC ACT CCA CAG CCA TCT CTG CCA GAA TGC TGC ACG
ATT TGG CCA AAG ATC CAA AGC CCA CCA GAT TGT TTT CTA TCG ACC GTG TTT TCC
GTA ACG AAG CAG TTG ACG CCA CCC ATT TGG CCG AAT TCC ACC AGG TGG AAG GTG
TTC TTG CCG ACT ACA ACA TTA CTC TGG GTG ACC TGA TCA AGT TCA TGG AAG AGT
TTT TCG AAA GAA TGG GTG TCA CCG GTT TGA GAT TCA AGC CTA CCT ACG GGC CTT
ACG GGG AGC CAT CTA TGG AAA TCT TTT CTT GGC ACG AAG GTT TGC AAA AAT GGG
TCG AAA TCG GTA ACT TTG GTA TGT TCA GAC CAG AAA TGC TCG AGT CCA TGG GTC
TAC CAA AGG ATC TAA GAG TCC TTG GTT GGG GGT TAT CCT TGG AAA GAC CTA CCA
TGA TCA AAT ATA AGG TTC AAA ACA TCA GAG AAC TGT TAG GTC ATA AAG TCT CTT
TGG ACT TTA TCG AAA CCA ATC CTG CTG CTA GAT TGG ACG AAG ACT TGT ACG AAT
AAG GCA GGA ATA GAT TAT GCC TAC CGT CTC CGT GAA CAA GCA GCA ATT ATT TGA
TCT TCT AGG CAA AGA CTA CAC TTC CCA AGA GTT CGA CGA ATT ATG TTT TGA ATT
CGG TAT GGA AAT GGA CGA AGA CAC CAC AGA AGA GGC CTT GAA AAC CGG GGA GGA
GCC GGA ATT GAA GCT TGA TAT CAG TGC CAA TCG TTA CGA TTT GCT TTG TAT CGA
AGG TAT TTC ACA GTC GCT GAA CGA ATA CTT GGA ACG TAA AGA AAG ACC TGA CTA
TAA ATT AAG CAA GCC AAC CAC TAA GTT GAT CAT CGA CAA ATC AAC GGA GCA AAT
TAG ACC TTT TGC TAC CGC TGC TGT ATT GAG AAA TAT CAA GCT TAA CGA AAA ATC
TTA CGC TTC TTT TAT TGC CTT GCA AGA TAA ATT ACA TGC CAA TCT ATG TAG AAA
CAG AAG CTT GGT TGC CAT GGG TAC TCA CGA TTT AGA TTC AAT TGA AGG TCC ATT
CCA TTA CAG AGC TCT ACC ACC AAA GGA CAT CAA GTT CGT ACC ATT GAA TCA AAC
CCA AGA GTT TAC TGG TGA CAA ATT GAT CGA GTT TTA TAA ATC TCC AGA ACA GAA
AAA CAA CAT AGG GAG ATA CGT TCA CAT TAT TGA GGA TTC TCC AGT CTT CCC AGT
TAT TAT GGA CAG CAA AGA TCG TGT TTG CTC CCT GCC ACC ATT AAT CAA TAG TGA
ACA TTC GAA GAT CTC TGT GAA CAC CCG TAA CAT TTT GAT TGA TAT AAC CGC CAC

CGA TAA GAC CAA AGC CGA GAT CGT TTT GAA CAT ATT AAC TAC AAT GTT CTC ACG
TTA TTG TGA CGA ACC ATT CAC GGT TGA GCC TGT AGA AAT TGT CTC TGA ACA CAA
TGG CCA ATC CCG TTT GGC GCC AAA CTT CAA CGA TAG AAT TAT GGA TGT CTC CAT
TAA GTA CAT CAA CTC CTG TCT TGG CCT AGA TCA ATC CGC TGA TGA AAT TGC TCA
TTG TCT AAA AAA GAT GTC GTT GCA TGC CGT TCA ATC AAA GGA AGA CAA GGA CAT
CTT GCA CGT TGA CAT TCC GGT AAC TAG ACC TGA TAT TTT GCA CGC TTG TGA TAT
AAT GGA AGA TGC CGC TGT CGG TTA TGG TTT CAA TAA TTT GCC AAA GGG TGA GAA
ATT ATC CAA TGC CAA CTT CAT TGC CAA ACC ATT ACC AAT CAA CAA GGT TTC TGA
TAT TTT CAG AGT TGC ATC CTC TCA AGC CAC GTG GGT TGA GGT TTT ACC ATT GAC
CTT ATG TTC GCA CGA TGA AAA CTT TAA ATT TCT AAG ACA ATC CGA CAA TGA TGA
TTT AGC TGT CAA ATT GGC CAA CCC AAA GAC TTT GGA ATA CCA AGT TGT TAG AAC
CAC TTT ATT GCC TGG TAT CTT AAA GAC AGT CAA GGA AAA CAG AAA ACA TTC CTT
ACC AAT CAA AGT CTT TGA AAC CGG TGA CGT TGT ATT TAA AGA CGA CAA ACT AGA
AAG GAA GGC GTA CAA TGA ACG TCA CTG GGC TGC CAT CTA CGT GGG TAA GAA TTC
TGG GTT TGA AAT CAT TCA AGG GTT ATT GGG TAA AAT CAT GCA AAC TTT TAG AAC
AGA GTG GAT TGC AGA CTA CGG TGC TGC TGC TTC TGG CAG AGG TTA CTG GAT TGA
AGA AGA CGA TTC TGT GAA AAC CTA TTT CCC AGG TAG AGG TGC CAA GGT CAT GTT
CAG ATC CAA AGA AGG CGC TGA GCC AAA GCA AAT CGG CCA CTT GGG TGT CTT GCA
TCC TGA AGT CAT GAT GAA TTT CGA CGT TCC ATT CGC TGC ATC CTT TGT AGA GGT
TAA TGC CGA AGT CTT CCT ATA ATG TAA TGT TCT AAC AAA AAT TTT TAC TGA TTT
ATA AAA CTT ATA TAG ATA GAT AGA CAT ATA TAT ATA TCT ATA TAT AGA AAC ACA
ACT AAA GTT TAC CAT GTT TTA TAT AGG GTA CCG AGC TCG AAT TCA CTG GCC GTC
GTT TTA CAA CGT CGT GAC TGG GAA AAC CGC GGT TAC GTG CAG TCG ATG ATA AGC
TGT CAA ACA TGA GAA TTG TGC CTA ATG AGT GAG CTA ACT TAC ATT AAT TGC GTT
GCG CTC ACT GCC CGC TTT CCA GTC GGG AAA CCT GTC GTG CCA GCT GCA TTA ATG
AAT CGG CCA ACG CGC GGG GAG AGG CGG TTT GCG TAT TGG GCG CCA GGG TGG TTT
TTC TTT TCA CCA GTG AGA CGG GCA ACA GCT GAT TGC CCT TCA CCG CCT GGC CCT
GAG AGA GTT GCA GCA AGC GGT CCA CGC TGG TTT GCC CCA GCA GGC GAA AAT CCT
GTT TGA TGG TGG TTA ACG GCG GGA TAT AAC ATG AGC TGT CTT CGG TAT CGT CGT
ATC CCA CTA CCG AGA TAT CCG CAC CAA CGC GCA GCC CGG ACT CGG TAA TGG CGC
GCA TTG CGC CCA GCG CCA TCT GAT CGT TGG CAA CCA GCA TCG CAG TGG GAA CGA
TGC CCT CAT TCA GCA TTT GCA TGG TTT GTT GAA AAC CGG ACA TGG CAC TCC AGT
CGC CTT CCC GTT CCG CTA TCG GCT GAA TTT GAT TGC GAG TGA GAT ATT TAT GCC
AGC CAG CCA GAC GCA GAC GCG CCG AGA CAG AAC TTA ATG GGC CCG CTA ACA GCG
CGA TTT GCT GGT GAC CCA ATG CGA CCA GAT GCT CCA CGC CCA GTC GCG TAC CGT
CTT CAT GGG AGA AAA TAA TAC TGT TGA TGG GTG TCT GGT CAG AGA CAT CAA GAA
ATA ACG CCG GAA CAT TAG TGC AGG CAG CTT CCA CAG CAA TGG CAT CCT GGT CAT
CCA GCG GAT AGT TAA TGA TCA GCC CAC TGA CGC GTT GCG CGA GAA GAT TGT GCA
CCG CCG CTT TAC AGG CTT CGA CGC CGC TTC GTT CTA CCA TCG ACA CCA CCA CGC
TGG CAC CCA GTT GAT CGG CGC GAG ATT TAA TCG CCG CGA CAA TTT GCG ACG GCG
CGT GCA GGG CCA GAC TGG AGG TGG CAA CGC CAA TCA GCA ACG ACT GTT TGC CCG
CCA GTT GTT GTG CCA CGC GGT TGG GAA TGT AAT TCA GCT CCG CCA TCG CCG CTT
CCA CTT TTT CCC GCG TTT TCG CAG AAA CGT GGC TGG CCT GGT TCA CCA CGC GGG
AAA CGG TCT GAT AAG AGA CAC CGG CAT ACT CTG CGA CAT CGT ATA ACG TTA CTG
GTT TCA CAT TCA CCA CCC TGA ATT GAC TCT CTT CCG GGC GCT ATC ATG CCA TAC
CGC GAA AGG TTT TGC ACC ATT CGA TGG TGT CGG AAT TTC GGG CAG CGT TGG GTC
CTG GCC ACG GGT GCG CAT GAC TTA AGC GGT GTG AAA TAC CGC ACA GAT GCG TAA
GGA GAA AAT ACC GCA TCA GGC GCT CTT CCG CTT CCT CGC TCA CTG ACT CGC TGC
GCT CGG TCT GTC GGC TGC GGC GAG CGG TAT CAG CTC ACT CAA AGG CGG TAA TAC
GGT TAT CCA CAG AAT CAG GGG ATA ACG CAG GAA AGA ACA TGT GAG CAA AAG GCC
AGC AAA AGG CCA GGA ACC GTA AAA AGG CCG CGT TGC TGG CGT TTT TCC ATA GGC
TCC GCC CCC CTG ACG AGC ATC ACA AAA ATC GAC GCT CAA GTC AGA GGT GGC GAA
ACC CGA CAG GAC TAT AAA GAT ACC AGG CGT TTC CCC CTG GAA GCT CCC TCG TGC
GCT CTC CTG TTC CGA CCC TGC CGC TTA CCG GAT ACC TGT CCG CCT TTC TCC CTT
CGG GAA GCG TGG CGC TTT CTC AAT GCT CAC GCT GTA GGT ATC TCA GTT CGG TGT
AGG TCG TTC GCT CCA AGC TGG GCT GTG TGC ACG AAC CCC CCG TTC AGC CCG ACC
GCT GCG CCT TAT CCG GTA ACT ATC GTC TTG AGT CCA ACC CGG TAA GAC ACG ACT

TAT CGC CAC TGG CAG CAG CCA CTG GTA ACA GGA TTA GCA GAG CGA GGT ATG TAG
GCG GTG CTA CAG AGT TCT TGA AGT GGT GGC CTA ACT ACG GCT ACA CTA GAA GGA
CAG TAT TTG GTA TCT GCG CTC TGC TGA AGC CAG TTA CCT TCG GAA AAA GAG TTG
GTA GCT CTT GAT CCG GCA AAC AAA CCA CCG CTG GTA GCG GTG GTT TTT TTG TTT
GCA AGC AGC AGA TTA CGC GCA GAA AAA AAG GAT CTC AAG AAG ATC CTT TGA TCT
TTT CTA CGG GGT CTG ACG CTC AGT GGA ACG AAA ACT CAC GTT AAG GGA TTT TGG
TCA TGA GAT TAT CAA AAA GGA TCT TCA CCT AGA TCC TTT TAA ATT AAA AAT GAA
GTT TTA AAT CAA TCT AAA GTA TAT ATG AGT AAA CTT GGT CTG ACA GTT ACC AAT
GCT TAA TCA GTG AGG CAC CTA TCT CAG CGA TCT GTC TAT TTC GTT CAT CCA TAG
CTG CCT GAC TCC CCG TCG TGT AGA TAA CTA CGA TAC GGG AGG GCT TAC CAT CTG
GCC CCA GTG CTG CAA TGA TAC CGC GAG ACC CAC GCT CAC CGG CTC CAG ATT TAT
CAG CAA TAA ACC AGC CAG CCG GAA GGG CCG AGC GCA GAA GTG GTC CTG CAA CTT
TAT CCG CCT CCA TCC AGT CTA TTA ATT GTT GCC GGG AAG CTA GAG TAA GTA GTT
CGC CAG TTA ATA GTT TGC GCA ACG TTG TTG CCA TTG CTA CAG GCA TCG TGG TGT
CAC GCT CGT CGT TTG GTA TGG CTT CAT TCA GCT CCG GTT CCC AAC GAT CAA GGC
GAG TTA CAT GAT CCC CCA TGT TGT GCA AAA AAG CGG TTA GCT CCT TCG GTC CTC
CGA TCG TTG TCA GAA GTA AGT TGG CCG CAG TGT TAT CAC TCA TGG TTA TGG CAG
CAC TGC ATA ATT CTC TTA CTG TCA TGC CAT CCG TAA GAT GCT TTT CTG TGA CTG
GTG AGT ACT CAA CCA AGT CAT TCT GAG AAT AGT GTA TGC GGC GAC CGA GTT GCT
CTT GCC CGG CGT CAA TAC GGG ATA ATA CCG CGC CAC ATA GCA GAA CTT TAA AAG
TGC TCA TCA TTG GAA AAC GTT CTT CGG GGC GAA AAC TCT CAA GGA TCT TAC CGC
TGT TGA GAT CCA GTT CGA TGT AAC CCA CTC GTG CAC CCA ACT GAT CTT CAG CAT
CTT TTA CTT TCA CCA GCG TTT CTG GGT GAG CAA AAA CAG GAA GGC AAA ATG CCG
CAA AAA AGG GAA TAA GGG CGA CAC GGA AAT GTT GAA TAC TCA TAC TCT TCC TTT
TTC AAT ATT ATT GAA GCA TTT ATC AGG GTT ATT GTC TCA TGA GCG GAT ACA TAT
TTG AAT GTA TTT AGA AAA ATA AAC AAA TAG GGG TTC CGC GCA CAT TTC CCC GAA
AAG TGC CAC CTG ACG TCT AAG AAA CCA TTA TTA TCA TGA CAT TAA CCT ATA AAA
ATA GGC GTA TCA CGA GGC CCT TTC GTC TTC AC

3. pQE16_mDHFR_lacI_yPheRS_naph (8109 bp)



T5 promoter and lac operon	7 - 87 bp
mDHFR_His	115 - 660 bp
yPheRS_alpha subunit	1264 - 2787 bp
yPheRS_beta subunit	2802 - 4589 bp
$lacI^q$	5926 - 4842 bp
ColE1 origin	6501 - 6526 bp
beta-lactamase	7904 - 7044 bp

Full sequence

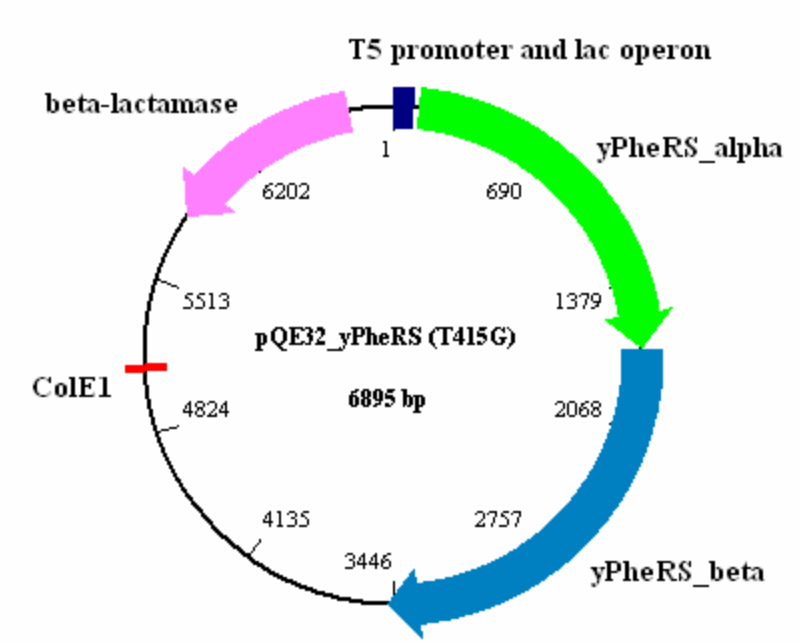
```
CTC GAG AAA TCA TAA AAA ATT TAT TTG CTT TGT GAG CGG ATA ACA ATT ATA ATA
GAT TCA ATT GTG AGC GGA TAA CAA TTT CAC ACA GAA TTC ATT AAA GAG GAG AAA
TTA ACT ATG AGA GGA TCC GGC ATC ATG GTT CGA CCA TTG AAC TCG ATC GTC GCC
GTG TCC CAA AAT ATG GGG ATT GGC AAG AAC GGA GAC CTA CCC TGG CCT CCG CTC
AGG AAC GAG TTC AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT TCA GTG GAA GGT
AAA CAG AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG AAG
AAT CGA CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA
```

CCA CCA CGA GGA GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT
ATT GAA CAA CCG GAA TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC
AGT TCT GTT TAC CAG GAA GCC ATG AAT CAA CCA GGC CAC CTT AGA CTC TTT GTG
ACA AGG ATC ATG CAG GAA TTT GAA AGT GAC ACG TTT TTC CCA GAA ATT GAT TTG
GGG AAA TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT GAG GTC CAG GAG
GAA AAA GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA GGT TCC AGA TCT
CAT CAC CAT CAC CAT CAC TAA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA
GAT CCA GTA ATG ACC TCA GAA CTC CAT CTG GAT TTG TTC AGA ACG CTC GGT TGC
CGC CGG GCG TTT TTT ATT GGT GAG AAT CCA AGC TAG CTT GGC GAG ATT TTC AGG
AGC TAA GGA AGC TAA AAT GGA GAA AAA AAT CAC TGG ATA TAC CAC CGT TGA TAT
ATC CCA ATG GCA TCG TAA AGA ACA TTT TGA GGC ATT TCA GTC AGT TGC TCA ATG
TAC CTA TAA CCA GAC CGT TCA GCT GGC ACG ACA GGT TTC CCG ACT GGA AAG CGG
GCA GTG AGC GCA ACG CAA TTA ATG TGA GTT AGC TCA CTC ATT AGG CAC CCC AGG
CTT TAC ACT TTA TGC TTC CGG CTC GTA TGT TGT GTG GAA TTG TGA GCG GAT AAC
AAT TTC ACA CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG CAT GCC TGC
AGT TGA CAA TTA ATC ATC GGC TCG TAT AAT GGA TCC AAT TGT GAG CGG AAT CGA
TTT TCA CAC AGG AAA CAG ACC ATG AAT CTA GAG ATG TCT GAC TTC CAA TTA GAA
ATT CTA AAG AAA CTA GAT GAA TTG GAT GAG ATC AAG TCC ACA CTG GCA ACT TTC
CCT CAG CAC GGC TCT CAA GAT GTT CTT TCC GCT TTG AAC TCT TTG AAA GCC CAC
AAC AAG TTA GAG TTT TCC AAG GTC GAC ACG GTT ACG TAT GAC TTG ACC AAA GAA
GGT GCT CAA ATT TTG AAT GAA GGT TCG TAC GAA ATT AAA CTA GTC AAG CTC ATC
CAA GAG TTG GGT CAA CTT CAA ATC AAA GAT GTG ATG TCC AAA CTA GGC CCT CAA
GTT GGT AAG GTC GGT CAG GCT AGA GCT TTC AAG AAC GGC TGG ATC GCC AAA AAC
GCC TCA AAC GAG CTT GAA CTC TCC GCA AAA TTG CAA AAT ACC GAT TTA AAT GAG
CTT ACT GAT GAA ACG CAA TCT ATT CTA GCG CAA ATC AAG AAC AAC TCG CAT CTG
GAT AGC ATT GAC GCC AAG ATT TTG AAC GAC TTG AAG AAA AGA AAG TTA ATT GCT
CAA GGT AAA ATC ACA GAT TTC AGT GTC ACC AAA GGG CCA GAG TTC TCG ACC GAC
CTC ACC AAA TTG GAA ACC GAT CTT ACC TCC GAC ATG GTC TCC ACC AAT GCA TAC
AAG GAC TTG AAG TTC AAG CCT TAC AAT TTC AAT TCT CAA GGT GTG CAA ATA TCT
TCA GGT GCT CTT CAC CCC TTA AAC AAA GTC AGA GAG GAA TTT AGA CAA ATT TTC
TTT TCC ATG GGA TTC ACA GAG ATG CCC TCG AAC CAA TAC GTC GAG ACA GGT TTC
TGG AAC TTC GAT GCC CTT TAC GTC CCA CAA CAG CAT CCT GCT CGT GAC CTG CAA
GAC ACT TTC TAC ATC AAG GAC CCA CTA ACC GCT GAG TTG CCC GAT GAC AAG ACA
TAC ATG GAC AAT ATC AAA GCC GTT CAC GAA CAG GGG AGA TTC GGG TCC ATC GGT
TAT CGT TAC AAC TGG AAG CCA GAA GAA TGT CAA AAA TTG GTC TTG AGA ACT CAC
TCC ACA GCC ATC TCT GCC AGA ATG CTG CAC GAT TTG GCC AAA GAT CCA AAG CCC
ACC AGA TTG TTT TCT ATC GAC CGT GTT TTC CGT AAC GAA GCA GTT GAC GCC ACC
CAT TTG GCC GAA TTC CAC CAG GTG GAA GGT GTT CTT GCC GAC TAC AAC ATT ACT
CTG GGT GAC CTG ATC AAG TTC ATG GAA GAG TTT TTC GAA AGA ATG GGT GTC ACC
GGT TTG AGA TTC AAG CCT ACC TAC GGG CCT TAC GGG GAG CCA TCT ATG GAA ATC
TTT TCT TGG CAC GAA GGT TTG CAA AAA TGG GTC GAA ATC GGT AAC TTT GGT ATG
TTC AGA CCA GAA ATG CTC GAG TCC ATG GGT CTA CCA AAG GAT CTA AGA GTC CTT
GGT TGG GGG TTA TCC TTG GAA AGA CCT ACC ATG ATC AAA TAT AAG GTT CAA AAC
ATC AGA GAA CTG TTA GGT CAT AAA GTC TCT TTG GAC TTT ATC GAA ACC AAT CCT
GCT GCT AGA TTG GAC GAA GAC TTG TAC GAA TAA GGC AGG AAT AGA TTA TGC CTA
CCG TCT CCG TGA ACA AGC AGC AAT TAT TTG ATC TTC TAG GCA AAG ACT ACA CTT
CCC AAG AGT TCG ACG AAT TAT GTT TTG AAT TCG GTA TGG AAA TGG ACG AAG ACA
CCA CAG AAG AGG CCT TGA AAA CCG GGG AGG AGC CGG AAT TGA AGC TTG ATA TCA
GTG CCA ATC GTT ACG ATT TGC TTT GTA TCG AAG GTA TTT CAC AGT CGC TGA ACG
AAT ACT TGG AAC GTA AAG AAA GAC CTG ACT ATA AAT TAA GCA AGC CAA CCA CTA
AGT TGA TCA TCG ACA AAT CAA CGG AGC AAA TTA GAC CTT TTG CTA CCG CTG CTG
TAT TGA GAA ATA TCA AGC TTA ACG AAA AAT CTT ACG CTT CTT TTA TTG CCT TGC
AAG ATA AAT TAC ATG CCA ATC TAT GTA GAA ACA GAA GCT TGG TTG CCA TGG GTA
CTC ACG ATT TAG ATT CAA TTG AAG GTC CAT TCC ATT ACA GAG CTC TAC CAC CAA
AGG ACA TCA AGT TCG TAC CAT TGA ATC AAA CCC AAG AGT TTA CTG GTG ACA AAT
TGA TCG AGT TTT ATA AAT CTC CAG AAC AGA AAA ACA ACA TAG GGA GAT ACG TTC
ACA TTA TTG AGG ATT CTC CAG TCT TCC CAG TTA TTA TGG ACA GCA AAG ATC GTG

TTT GCT CCC TGC CAC CAT TAA TCA ATA GTG AAC ATT CGA AGA TCT CTG TGA ACA
CCC GTA ACA TTT TGA TTG ATA TAA CCG CCA CCG ATA AGA CCA AAG CCG AGA TCG
TTT TGA ACA TAT TAA CTA CAA TGT TCT CAC GTT ATT GTG ACG AAC CAT TCA CGG
TTG AGC CTG TAG AAA TTG TCT CTG AAC ACA ATG GCC AAT CCC GTT TGG CGC CAA
ACT TCA ACG ATA GAA TTA TGG ATG TCT CCA TTA AGT ACA TCA ACT CCT GTC TTG
GCC TAG ATC AAT CCG CTG ATG AAA TTG CTC ATT GTC TAA AAA AGA TGT CGT TGC
ATG CCG TTC AAT CAA AGG AAG ACA AGG ACA TCT TGC ACG TTG ACA TTC CGG TAA
CTA GAC CTG ATA TTT TGC ACG CTT GTG ATA TAA TGG AAG ATG CCG CTG TCG GTT
ATG GTT TCA ATA ATT TGC CAA AGG GTG AGA AAT TAT CCA ATG CCA ACT TCA TTG
CCA AAC CAT TAC CAA TCA ACA AGG TTT CTG ATA TTT TCA GAG TTG CAT CCT CTC
AAG CCA CGT GGG TTG AGG TTT TAC CAT TGA CCT TAT GTT CGC ACG ATG AAA ACT
TTA AAT TTC TAA GAC AAT CCG ACA ATG ATG ATT TAG CTG TCA AAT TGG CCA ACC
CAA AGA CTT TGG AAT ACC AAG TTG TTA GAA CCA CTT TAT TGC CTG GTA TCT TAA
AGA CAG TCA AGG AAA ACA GAA AAC ATT CCT TAC CAA TCA AAG TCT TTG AAA CCG
GTG ACG TTG TAT TTA AAG ACG ACA AAC TAG AAA GGA AGG CGT ACA ATG AAC GTC
ACT GGG CTG CCA TCT ACG TGG GTA AGA ATT CTG GGT TTG AAA TCA TTC AAG GGT
TAT TGG GTA AAA TCA TGC AAA CTT TTA GAA CAG AGT GGA TTG CAG ACT ACG GTG
CTG CTG CTT CTG GCA GAG GTT ACT GGA TTG AAG AAG ACG ATT CTG TGA AAA CCT
ATT TCC CAG GTA GAG GTG CCA AGG TCA TGT TCA GAT CCA AAG AAG GCG CTG AGC
CAA AGC AAA TCG GCC ACT TGG GTG TCT TGC ATC CTG AAG TCA TGA TGA ATT TCG
ACG TTC CAT TCG CTG CAT CCT TTG TAG AGG TTA ATG CCG AAG TCT TCC TAT AAT
GTA ATG TTC TAA CAA AAA TTT TTA CTG ATT TAT AAA ACT TAT ATA GAT AGA TAG
ACA TAT ATA TAT ATC TAT ATA TAG AAA CAC AAC TAA AGT TTA CCA TGT TTT ATA
TAG GGT ACC GAG CTC GAA TTC ACT GGC CGT CGT TTT ACA ACG TCG TGA CTG GGA
AAA CCG CGG TTA CGT GCA GTC GAT GAT AAG CTG TCA AAC ATG AGA ATT GTG CCT
AAT GAG TGA GCT AAC TTA CAT TAA TTG CGT TGC GCT CAC TGC CCG CTT TCC AGT
CGG GAA ACC TGT CGT GCC AGC TGC ATT AAT GAA TCG GCC AAC GCG CGG GGA GAG
GCG GTT TGC GTA TTG GGC GCC AGG GTG GTT TTT CTT TTC ACC AGT GAG ACG GGC
AAC AGC TGA TTG CCC TTC ACC GCC TGG CCC TGA GAG AGT TGC AGC AAG CGG TCC
ACG CTG GTT TGC CCC AGC AGG CGA AAA TCC TGT TTG ATG GTG GTT AAC GGC GGG
ATA TAA CAT GAG CTG TCT TCG GTA TCG TCG TAT CCC ACT ACC GAG ATA TCC GCA
CCA ACG CGC AGC CCG GAC TCG GTA ATG GCG CGC ATT GCG CCC AGC GCC ATC TGA
TCG TTG GCA ACC AGC ATC GCA GTG GGA ACG ATG CCC TCA TTC AGC ATT TGC ATG
GTT TGT TGA AAA CCG GAC ATG GCA CTC CAG TCG CCT TCC CGT TCC GCT ATC GGC
TGA ATT TGA TTG CGA GTG AGA TAT TTA TGC CAG CCA GCC AGA CGC AGA CGC GCC
GAG ACA GAA CTT AAT GGG CCC GCT AAC AGC GCG ATT TGC TGG TGA CCC AAT GCG
ACC AGA TGC TCC ACG CCC AGT CGC GTA CCG TCT TCA TGG GAG AAA ATA ATA CTG
TTG ATG GGT GTC TGG TCA GAG ACA TCA AGA AAT AAC GCC GGA ACA TTA GTG CAG
GCA GCT TCC ACA GCA ATG GCA TCC TGG TCA TCC AGC GGA TAG TTA ATG ATC AGC
CCA CTG ACG CGT TGC GCG AGA AGA TTG TGC ACC GCC GCT TTA CAG GCT TCG ACG
CCG CTT CGT TCT ACC ATC GAC ACC ACC ACG CTG GCA CCC AGT TGA TCG GCG CGA
GAT TTA ATC GCC GCG ACA ATT TGC GAC GGC GCG TGC AGG GCC AGA CTG GAG GTG
GCA ACG CCA ATC AGC AAC GAC TGT TTG CCC GCC AGT TGT TGT GCC ACG CGG TTG
GGA ATG TAA TTC AGC TCC GCC ATC GCC GCT TCC ACT TTT TCC CGC GTT TTC GCA
GAA ACG TGG CTG GCC TGG TTC ACC ACG CGG GAA ACG GTC TGA TAA GAG ACA CCG
GCA TAC TCT GCG ACA TCG TAT AAC GTT ACT GGT TTC ACA TTC ACC ACC CTG AAT
TGA CTC TCT TCC GGG CGC TAT CAT GCC ATA CCG CGA AAG GTT TTG CAC CAT TCG
ATG GTG TCG GAA TTT CGG GCA GCG TTG GGT CCT GGC CAC GGG TGC GCA TGA CTT
AAG CGG TGT GAA ATA CCG CAC AGA TGC GTA AGG AGA AAA TAC CGC ATC AGG CGC
TCT TCC GCT TCC TCG CTC ACT GAC TCG CTG CGC TCG GTC TGT CGG CTG CGG CGA
GCG GTA TCA GCT CAC TCA AAG GCG GTA ATA CGG TTA TCC ACA GAA TCA GGG GAT
AAC GCA GGA AAG AAC ATG TGA GCA AAA GGC CAG CAA AAG GCC AGG AAC CGT AAA
AAG GCC GCG TTG CTG GCG TTT TTC CAT AGG CTC CGC CCC CCT GAC GAG CAT CAC
AAA AAT CGA CGC TCA AGT CAG AGG TGG CGA AAC CCG ACA GGA CTA TAA AGA TAC
CAG GCG TTT CCC CCT GGA AGC TCC CTC GTG CGC TCT CCT GTT CCG ACC CTG CCG
CTT ACC GGA TAC CTG TCC GCC TTT CTC CCT TCG GGA AGC GTG GCG CTT TCT CAA
TGC TCA CGC TGT AGG TAT CTC AGT TCG GTG TAG GTC GTT CGC TCC AAG CTG GGC

TGT GTG CAC GAA CCC CCC GTT CAG CCC GAC CGC TGC GCC TTA TCC GGT AAC TAT
CGT CTT GAG TCC AAC CCG GTA AGA CAC GAC TTA TCG CCA CTG GCA GCA GCC ACT
GGT AAC AGG ATT AGC AGA GCG AGG TAT GTA GGC GGT GCT ACA GAG TTC TTG AAG
TGG TGG CCT AAC TAC GGC TAC ACT AGA AGG ACA GTA TTT GGT ATC TGC GCT CTG
CTG AAG CCA GTT ACC TTC GGA AAA AGA GTT GGT AGC TCT TGA TCC GGC AAA CAA
ACC ACC GCT GGT AGC GGT GGT TTT TTT GTT TGC AAG CAG CAG ATT ACG CGC AGA
AAA AAA GGA TCT CAA GAA GAT CCT TTG ATC TTT TCT ACG GGG TCT GAC GCT CAG
TGG AAC GAA AAC TCA CGT TAA GGG ATT TTG GTC ATG AGA TTA TCA AAA AGG ATC
TTC ACC TAG ATC CTT TTA AAT TAA AAA TGA AGT TTT AAA TCA ATC TAA AGT ATA
TAT GAG TAA ACT TGG TCT GAC AGT TAC CAA TGC TTA ATC AGT GAG GCA CCT ATC
TCA GCG ATC TGT CTA TTT CGT TCA TCC ATA GCT GCC TGA CTC CCC GTC GTG TAG
ATA ACT ACG ATA CGG GAG GGC TTA CCA TCT GGC CCC AGT GCT GCA ATG ATA CCG
CGA GAC CCA CGC TCA CCG GCT CCA GAT TTA TCA GCA ATA AAC CAG CCA GCC GGA
AGG GCC GAG CGC AGA AGT GGT CCT GCA ACT TTA TCC GCC TCC ATC CAG TCT ATT
AAT TGT TGC CGG GAA GCT AGA GTA AGT AGT TCG CCA GTT AAT AGT TTG CGC AAC
GTT GTT GCC ATT GCT ACA GGC ATC GTG GTG TCA CGC TCG TCG TTT GGT ATG GCT
TCA TTC AGC TCC GGT TCC CAA CGA TCA AGG CGA GTT ACA TGA TCC CCC ATG TTG
TGC AAA AAA GCG GTT AGC TCC TTC GGT CCT CCG ATC GTT GTC AGA AGT AAG TTG
GCC GCA GTG TTA TCA CTC ATG GTT ATG GCA GCA CTG CAT AAT TCT CTT ACT GTC
ATG CCA TCC GTA AGA TGC TTT TCT GTG ACT GGT GAG TAC TCA ACC AAG TCA TTC
TGA GAA TAG TGT ATG CGG CGA CCG AGT TGC TCT TGC CCG GCG TCA ATA CGG GAT
AAT ACC GCG CCA CAT AGC AGA ACT TTA AAA GTG CTC ATC ATT GGA AAA CGT TCT
TCG GGG CGA AAA CTC TCA AGG ATC TTA CCG CTG TTG AGA TCC AGT TCG ATG TAA
CCC ACT CGT GCA CCC AAC TGA TCT TCA GCA TCT TTT ACT TTC ACC AGC GTT TCT
GGG TGA GCA AAA ACA GGA AGG CAA AAT GCC GCA AAA AAG GGA ATA AGG GCG ACA
CGG AAA TGT TGA ATA CTC ATA CTC TTC CTT TTT CAA TAT TAT TGA AGC ATT TAT
CAG GGT TAT TGT CTC ATG AGC GGA TAC ATA TTT GAA TGT ATT TAG AAA AAT AAA
CAA ATA GGG GTT CCG CGC ACA TTT CCC CGA AAA GTG CCA CCT GAC GTC TAA GAA
ACC ATT ATT ATC ATG ACA TTA ACC TAT AAA AAT AGG CGT ATC ACG AGG CCC TTT
CGT CTT CAC

4. pQE32_yPheRS (T415G) (6,895 bp)



T5 promoter and lac operon	7 - 87 bp
yPheRS_alpha subunit	115 - 1680 bp
yPheRS_beta subunit	1695 - 3482 bp
ColE1 origin	5103 - 5127 bp
beta-lactamase	6690 - 5830 bp

Full sequence

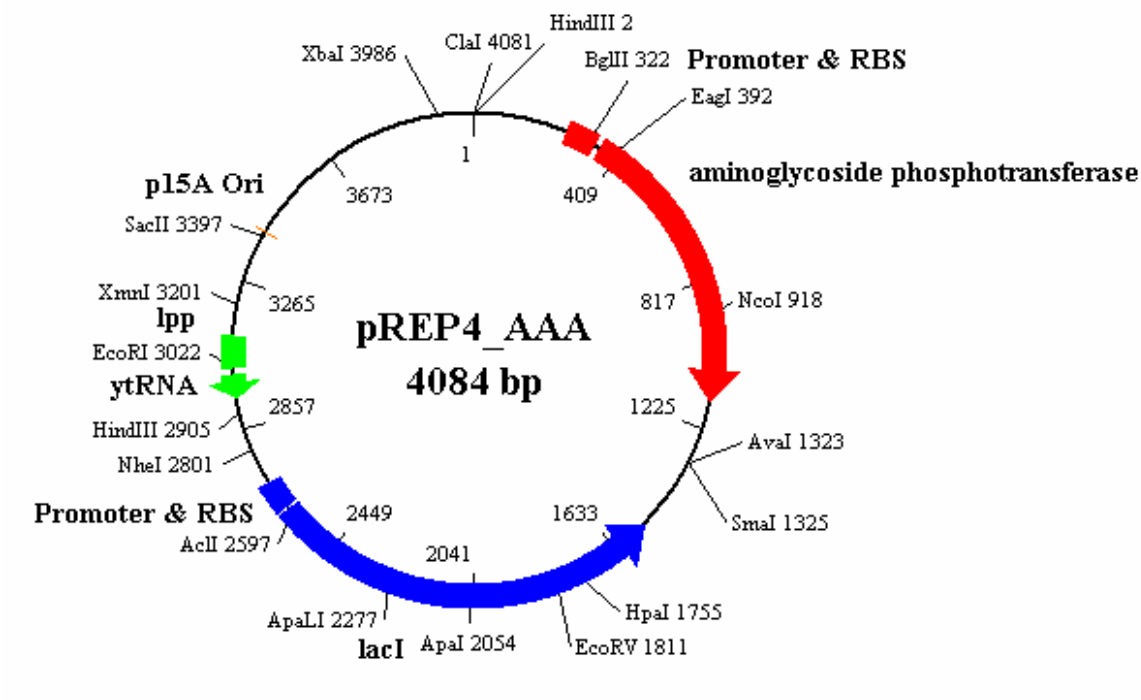
```
CTC GAG AAA TCA TAA AAA ATT TAT TTG CTT TGT GAG CGG ATA ACA ATT ATA ATA
GAT TCA ATT GTG AGC GGA TAA CAA TTT CAC ACA GAA TTC ATT AAA GAG GAG AAA
TTA ACT ATG AGA GGA TCT CAC CAT CAC CAT CAC CAT GGG ATC CAG ACC ATG AAT
CTA GAG ATG TCT GAC TTC CAA TTA GAA ATT CTA AAG AAA CTA GAT GAA TTG GAT
GAG ATC AAG TCC ACA CTG GCA ACT TTC CCT CAG CAC GGC TCT CAA GAT GTT CTT
TCC GCT TTG AAC TCT TTG AAA GCC CAC AAC AAG TTA GAG TTT TCC AAG GTC GAC
ACG GTT ACG TAT GAC TTG ACC AAA GAA GGT GCT CAA ATT TTG AAT GAA GGT TCG
TAC GAA ATT AAA CTA GTC AAG CTC ATC CAA GAG TTG GGT CAA CTT CAA ATC AAA
GAT GTG ATG TCC AAA CTA GGC CCT CAA GTT GGT AAG GTC GGT CAG GCT AGA GCT
TTC AAG AAC GGC TGG ATC GCC AAA AAC GCC TCA AAC GAG CTT GAA CTC TCC GCA
AAA TTG CAA AAT ACC GAT TTA AAT GAG CTT ACT GAT GAA ACG CAA TCT ATT CTA
GCG CAA ATC AAG AAC AAC TCG CAT CTG GAT AGC ATT GAC GCC AAG ATT TTG AAC
GAC TTG AAG AAA AGA AAG TTA ATT GCT CAA GGT AAA ATC ACA GAT TTC AGT GTC
```

ACC AAA GGG CCA GAG TTC TCG ACC GAC CTC ACC AAA TTG GAA ACC GAT CTT ACC
TCC GAC ATG GTC TCC ACC AAT GCA TAC AAG GAC TTG AAG TTC AAG CCT TAC AAT
TTC AAT TCT CAA GGT GTG CAA ATA TCT TCA GGT GCT CTT CAC CCC TTA AAC AAA
GTC AGA GAG GAA TTT AGA CAA ATT TTC TTT TCC ATG GGA TTC ACA GAG ATG CCC
TCG AAC CAA TAC GTC GAG ACA GGT TTC TGG AAC TTC GAT GCC CTT TAC GTC CCA
CAA CAG CAT CCT GCT CGT GAC CTG CAA GAC ACT TTC TAC ATC AAG GAC CCA CTA
ACC GCT GAG TTG CCC GAT GAC AAG ACA TAC ATG GAC AAT ATC AAA GCC GTT CAC
GAA CAG GGG AGA TTC GGG TCC ATC GGT TAT CGT TAC AAC TGG AAG CCA GAA GAA
TGT CAA AAA TTG GTC TTG AGA ACT CAC TCC ACA GCC ATC TCT GCC AGA ATG CTG
CAC GAT TTG GCC AAA GAT CCA AAG CCC ACC AGA TTG TTT TCT ATC GAC CGT GTT
TTC CGT AAC GAA GCA GTT GAC GCC ACC CAT TTG GCC GAA TTC CAC CAG GTG GAA
GGT GTT CTT GCC GAC TAC AAC ATT ACT CTG GGT GAC CTG ATC AAG TTC ATG GAA
GAG TTT TTC GAA AGA ATG GGT GTC ACC GGT TTG AGA TTC AAG CCT ACC TAC AAT
CCT TAC GGC GAG CCA TCA ATG GAA ATC TTT TCT TGG CAC GAA GGT TTG CAA AAA
TGG GTC GAA ATC GGT AAC TCT GGT ATG TTC AGA CCA GAA ATG CTC GAG TCC ATG
GGT CTA CCA AAG GAT CTA AGA GTC CTT GGT TGG GGG TTA TCC TTG GAA AGA CCT
ACC ATG ATC AAA TAT AAG GTT CAA AAC ATC AGA GAA CTG TTA GGT CAT AAA GTC
TCT TTG GAC TTT ATC GAA ACC AAT CCT GCT GCT AGA TTG GAC GAA GAC TTG TAC
GAA TAA GGC AGG AAT AGA TTA TGC CTA CCG TCT CCG TGA ACA AGC AGC AAT TAT
TTG ATC TTC TAG GCA AAG ACT ACA CTT CCC AAG AGT TCG ACG AAT TAT GTT TTG
AAT TCG GTA TGG AAA TGG ACG AAG ACA CCA CAG AAG AGG CCT TGA AAA CCG GGG
AGG AGC CGG AAT TGA AGC TTG ATA TCA GTG CCA ATC GTT ACG ATT TGC TTT GTA
TCG AAG GTA TTT CAC AGT CGC TGA ACG AAT ACT TGG AAC GTA AAG AAA GAC CTG
ACT ATA AAT TAA GCA AGC CAA CCA CTA AGT TGA TCA TCG ACA AAT CAA CGG AGC
AAA TTA GAC CTT TTG CTA CCG CTG CTG TAT TGA GAA ATA TCA AGC TTA ACG AAA
AAT CTT ACG CTT CTT TTA TTG CCT TGC AAG ATA AAT TAC ATG CCA ATC TAT GTA
GAA ACA GAA GCT TGG TTG CCA TGG GTA CTC ACG ATT TAG ATT CAA TTG AAG GTC
CAT TCC ATT ACA GAG CTC TAC CAC CAA AGG ACA TCA AGT TCG TAC CAT TGA ATC
AAA CCC AAG AGT TTA CTG GTG ACA AAT TGA TCG AGT TTT ATA AAT CTC CAG AAC
AGA AAA ACA ACA TAG GGA GAT ACG TTC ACA TTA TTG AGG ATT CTC CAG TCT TCC
CAG TTA TTA TGG ACA GCA AAG ATC GTG TTT GCT CCC TGC CAC CAT TAA TCA ATA
GTG AAC ATT CGA AGA TCT CTG TGA ACA CCC GTA ACA TTT TGA TTG ATA TAA CCG
CCA CCG ATA AGA CCA AAG CCG AGA TCG TTT TGA ACA TAT TAA CTA CAA TGT TCT
CAC GTT ATT GTG ACG AAC CAT TCA CGG TTG AGC CTG TAG AAA TTG TCT CTG AAC
ACA ATG GCC AAT CCC GTT TGG CGC CAA ACT TCA ACG ATA GAA TTA TGG ATG TCT
CCA TTA AGT ACA TCA ACT CCT GTC TTG GCC TAG ATC AAT CCG CTG ATG AAA TTG
CTC ATT GTC TAA AAA AGA TGT CGT TGC ATG CCG TTC AAT CAA AGG AAG ACA AGG
ACA TCT TGC ACG TTG ACA TTC CGG TAA CTA GAC CTG ATA TTT TGC ACG CTT GTG
ATA TAA TGG AAG ATG CCG CTG TCG GTT ATG GTT TCA ATA ATT TGC CAA AGG GTG
AGA AAT TAT CCA ATG CCA ACT TCA TTG CCA AAC CAT TAC CAA TCA ACA AGG TTT
CTG ATA TTT TCA GAG TTG CAT CCT CTC AAG CCA CGT GGG TTG AGG TTT TAC CAT
TGA CCT TAT GTT CGC ACG ATG AAA ACT TTA AAT TTC TAA GAC AAT CCG ACA ATG
ATG ATT TAG CTG TCA AAT TGG CCA ACC CAA AGA CTT TGG AAT ACC AAG TTG TTA
GAA CCA CTT TAT TGC CTG GTA TCT TAA AGA CAG TCA AGG AAA ACA GAA AAC ATT
CCT TAC CAA TCA AAG TCT TTG AAA CCG GTG ACG TTG TAT TTA AAG ACG ACA AAC
TAG AAA GGA AGG CGT ACA ATG AAC GTC ACT GGG CTG CCA TCT ACG TGG GTA AGA
ATT CTG GGT TTG AAA TCA TTC AAG GGT TAT TGG GTA AAA TCA TGC AAA CTT TTA
GAA CAG AGT GGA TTG CAG ACT ACG GTG CTG CTG CTT CTG GCA GAG GTT ACT GGA
TTG AAG AAG ACG ATT CTG TGA AAA CCT ATT TCC CAG GTA GAG GTG CCA AGG TCA
TGT TCA GAT CCA AAG AAG GCG CTG AGC CAA AGC AAA TCG GCC ACT TGG GTG TCT
TGC ATC CTG AAG TCA TGA TGA ATT TCG ACG TTC CAT TCG CTG CAT CCT TTG TAG
AGG TTA ATG CCG AAG TCT TCC TAT AAT GTA ATG TTC TAA CAA AAA TTT TTA CTG
ATT TAT AAA ACT TAT ATA GAT AGA TAG ACA CAT AAA TGT ATC TAT ATA TAG AAA
CAC AAC TAA AGT TTA CCA TCT TAA AGG TAA TGG GTA CCC CGG GTC GAC CTG CAG
CCA AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT GAT AGA TCC AGT AAT GAC CTC
AGA ACT CCA TCT GGA TTT GTT CAG AAC GCT CGG TTG CCG CCG GGC GTT TTT TAT
TGG TGA GAA TCC AAG CTA GCT TGG CGA GAT TTT CAG GAG CTA AGG AAG CTA AAA

TGG AGA AAA AAA TCA CTG GAT ATA CCA CCG TTG ATA TAT CCC AAT GGC ATC GTA
AAG AAC ATT TTG AGG CAT TTC AGT CAG TTG CTC AAT GTA CCT ATA ACC AGA CCG
TTC AGC TGG ATA TTA CGG CCT TTT TAA AGA CCG TAA AGA AAA ATA AGC ACA AGT
TTT ATC CGG CCT TTA TTC ACA TTC TTG CCC GCC TGA TGA ATG CTC ATC CGG AAT
TTC GTA TGG CAA TGA AAG ACG GTG AGC TGG TGA TAT GGG ATA GTG TTC ACC CTT
GTT ACA CCG TTT TCC ATG AGC AAA CTG AAA CGT TTT CAT CGC TCT GGA GTG AAT
ACC ACG ACG ATT TCC GGC AGT TTC TAC ACA TAT ATT CGC AAG ATG TGG CGT GTT
ACG GTG AAA ACC TGG CCT ATT TCC CTA AAG GGT TTA TTG AGA ATA TGT TTT TCG
TCT CAG CCA ATC CCT GGG TGA GTT TCA CCA GTT TTG ATT TAA ACG TGG CCA ATA
TGG ACA ACT TCT TCG CCC CCG TTT TCA CCA TGG GCA AAT ATT ATA CGC AAG GCG
ACA AGG TGC TGA TGC CGC TGG CGA TTC AGG TTC ATC ATG CCG TTT GTG ATG GCT
TCC ATG TCG GCA GAA TGC TTA ATG AAT TAC AAC AGT ACT GCG ATG AGT GGC AGG
GCG GGG CGT AAT TTT TTT AAG GCA GTT ATT GGT GCC CTT AAA CGC CTG GGG TAA
TGA CTC TCT AGC TTG AGG CAT CAA ATA AAA CGA AAG GCT CAG TCG AAA GAC TGG
GCC TTT CGT TTT ATC TGT TGT TTG TCG GTG AAC GCT CTC CTG AGT AGG ACA AAT
CCG CCC TCT AGA GCT GCC TCG CGC GTT TCG GTG ATG ACG GTG AAA ACC TCT GAC
ACA TGC AGC TCC CGG AGA CGG TCA CAG CTT GTC TGT AAG CGG ATG CCG GGA GCA
GAC AAG CCC GTC AGG GCG CGT CAG CGG GTG TTG GCG GGT GTC GGG GCG CAG CCA
TGA CCC AGT CAC GTA GCG ATA GCG GAG TGT ATA CTG GCT TAA CTA TGC GGC ATC
AGA GCA GAT TGT ACT GAG AGT GCA CCA TAT GCG GTG TGA AAT ACC GCA CAG ATG
CGT AAG GAG AAA ATA CCG CAT CAG GCG CTC TTC CGC TTC CTC GCT CAC TGA CTC
GCT GCG CTC GGT CGT TCG GCT GCG GCG AGC GGT ATC AGC TCA CTC AAA GGC GGT
AAT ACG GTT ATC CAC AGA ATC AGG GGA TAA CGC AGG AAA GAA CAT GTG AGC AAA
AGG CCA GCA AAA GGC CAG GAA CCG TAA AAA GGC CGC GTT GCT GGC GTT TTT CCA
TAG GCT CCG CCC CCC TGA CGA GCA TCA CAA AAA TCG ACG CTC AAG TCA GAG GTG
GCG AAA CCC GAC AGG ACT ATA AAG ATA CCA GGC GTT TCC CCC TGG AAG CTC CCT
CGT GCG CTC TCC TGT TCC GAC CCT GCC GCT TAC CGG ATA CCT GTC CGC CTT TCT
CCC TTC GGG AAG CGT GGC GCT TTC TCA TAG CTC ACG CTG TAG GTA TCT CAG TTC
GGT GTA GGT CGT TCG CTC CAA GCT GGG CTG TGT GCA CGA ACC CCC CGT TCA GCC
CGA CCG CTG CGC CTT ATC CGG TAA CTA TCG TCT TGA GTC CAA CCC GGT AAG ACA
CGA CTT ATC GCC ACT GGC AGC AGC CAC TGG TAA CAG GAT TAG CAG AGC GAG GTA
TGT AGG CGG TGC TAC AGA GTT CTT GAA GTG GTG GCC TAA CTA CGG CTA CAC TAG
AAG GAC AGT ATT TGG TAT CTG CGC TCT GCT GAA GCC AGT TAC CTT CGG AAA AAG
AGT TGG TAG CTC TTG ATC CGG CAA ACA AAC CAC CGC TGG TAG CGG TGG TTT TTT
TGT TTG CAA GCA GCA GAT TAC GCG CAG AAA AAA AGG ATC TCA AGA AGA TCC TTT
GAT CTT TTC TAC GGG GTC TGA CGC TCA GTG GAA CGA AAA CTC ACG TTA AGG GAT
TTT GGT CAT GAG ATT ATC AAA AAG GAT CTT CAC CTA GAT CCT TTT AAA TTA AAA
ATG AAG TTT TAA ATC AAT CTA AAG TAT ATA TGA GTA AAC TTG GTC TGA CAG TTA
CCA ATG CTT AAT CAG TGA GGC ACC TAT CTC AGC GAT CTG TCT ATT TCG TTC ATC
CAT AGT TGC CTG ACT CCC CGT CGT GTA GAT AAC TAC GAT ACG GGA GGG CTT ACC
ATC TGG CCC CAG TGC TGC AAT GAT ACC GCG AGA CCC ACG CTC ACC GGC TCC AGA
TTT ATC AGC AAT AAA CCA GCC AGC CGG AAG GGC CGA GCG CAG AAG TGG TCC TGC
AAC TTT ATC CGC CTC CAT CCA GTC TAT TAA TTG TTG CCG GGA AGC TAG AGT AAG
TAG TTC GCC AGT TAA TAG TTT GCG CAA CGT TGT TGC CAT TGC TAC AGG CAT CGT
GGT GTC ACG CTC GTC GTT TGG TAT GGC TTC ATT CAG CTC CGG TTC CCA ACG ATC
AAG GCG AGT TAC ATG ATC CCC CAT GTT GTG CAA AAA AGC GGT TAG CTC CTT CGG
TCC TCC GAT CGT TGT CAG AAG TAA GTT GGC CGC AGT GTT ATC ACT CAT GGT TAT
GGC AGC ACT GCA TAA TTC TCT TAC TGT CAT GCC ATC CGT AAG ATG CTT TTC TGT
GAC TGG TGA GTA CTC AAC CAA GTC ATT CTG AGA ATA GTG TAT GCG GCG ACC GAG
TTG CTC TTG CCC GGC GTC AAT ACG GGA TAA TAC CGC GCC ACA TAG CAG AAC TTT
AAA AGT GCT CAT CAT TGG AAA ACG TTC TTC GGG GCG AAA ACT CTC AAG GAT CTT
ACC GCT GTT GAG ATC CAG TTC GAT GTA ACC CAC TCG TGC ACC CAA CTG ATC TTC
AGC ATC TTT TAC TTT CAC CAG CGT TTC TGG GTG AGC AAA AAC AGG AAG GCA AAA
TGC CGC AAA AAA GGG AAT AAG GGC GAC ACG GAA ATG TTG AAT ACT CAT ACT CTT
CCT TTT TCA ATA TTA TTG AAG CAT TTA TCA GGG TTA TTG TCT CAT GAG CGG ATA
CAT ATT TGA ATG TAT TTA GAA AAA TAA ACA AAT AGG GGT TCC GCG CAC ATT TCC

CCG AAA AGT GCC ACC TGA CGT CTA AGA AAC CAT TAT TAT CAT GAC ATT AAC CTA
TAA AAA TAG GCG TAT CAC GAG GCC CTT TCG TCT TCA C

5. pREP4_ytRNA^{Phe}_AAA (4084 bp)



Vector size	4084 bp
agpt promoter	259-260 bp
RBS	336-344 bp
Aminoglycosie phosphotransferase	357-1151 bp
lacI promoter	2638-2639 bp
lacI repressor	1529-2611 bp
lpp promoter	3029-3083 bp
ytRNA	2945-3020 bp
transcriptional terminator(rrnC)	2938-2909 bp
p15A origin of replication	3408-3411 bp

Full sequence

```
AAG CTT CAC GCT GCC GCA AGC ACT CAG GGC GCA AGG GCT GCT AAA GGA AGC GGA
ACA CGT AGA AAG CCA GTC CGC AGA AAC GGT GCT GAC CCC GGA TGA ATG TCA GCT
```

ACT GGG CTA TCT GGA CAA GGG AAA ACG CAA GCG CAA AGA GAA AGC AGG TAG CTT
GCA GTG GGC TTA CAT GGC GAT AGC TAG ACT GGG CGG TTT TAT GGA CAG CAA GCG
AAC CGG AAT TGC CAG CTG GGG CGC CCT CTG GTA AGG TTG GGA AGC CCT GCA AAG
TAA ACT GGA TGG CTT TCT TGC CGC CAA GGA TCT GAT GGC GCA GGG GAT CAA GAT
CTG ATC AAG AGA CAG GAT GAC GGT CGT TTC GCA TGC TTG AAC AAG ATG GAT TGC
ACG CAG GTT CTC CGG CCG CTT GGG TGG AGA GGC TAT TCG GCT ATG ACT GGG CAC
AAC AGA CAA TCG GCT GCT CTG ATG CCG CCG TGT TCC GGC TGT CAG CGC AGG GGC
GCC CGG TTC TTT TTG TCA AGA CCG ACC TGT CCG GTG CCC TGA ATG AAC TGC AGG
ACG AGG CAG CGC GGC TAT CGT GGC TGG CCA CGA CGG GCG TTC CTT GCG CAG CTG
TGC TCG ACG TTG TCA CTG AAG CGG GAA GGG ACT GGC TGC TAT TGG GCG AAG TGC
CGG GGC AGG ATC TCC TGT CAT CTC ACC TTG CTC CTG CCG AGA AAG TAT CCA TCA
TGG CTG ATG CAA TGC GGC GGC TGC ATA CGC TTG ATC CGG CTA CCT GCC CAT TCG
ACC ACC AAG CGA AAC ATC GCA TCG AGC GAG CAC GTA CTC GGA TGG AAG CCG GTC
TTG TCG ATC AGG ATG ATC TGG ACG AAG AGC ATC AGG GGC TCG CGC CAG CCG AAC
TGT TCG CCA GGC TCA AGG CGC GCA TGC CCG ACG GCG AGG ATC TCG TCG TGA CCC
ATG GCG ATG CCT GCT TGC CGA ATA TCA TGG TGG AAA ATG GCC GCT TTT CTG GAT
TCA TCG ACT GTG GCC GGC TGG GTG TGG CGG ACC GCT ATC AGG ACA TAG CGT TGG
CTA CCC GTG ATA TTG CTG AAG AGC TTG GCG GCG AAT GGG CTG ACC GCT TCC TCG
TGC TTT ACG GTA TCG CCG CTC CCG ATT CGC AGC GCA TCG CCT TCT ATC GCC TTC
TTG ACG AGT TCT TCT GAG CGG GAC TCT GGG GTT CGA AAT GAC CGA CCA AGC GAC
GCC CAA CCT GCC ATC ACG AGA TTT CGA TTC CAC CGC CGC CTT CTA TGA AAG GTT
GGG CTT CGG AAT CGT TTT CCG GGA CGC CGG CTG GAT GAT CCT CCA GCG CGG GGA
TCT CAT GCT GGA GTT CTT CGC CCA CCC CGG GCT CGA TCC CCT CGC GAG TTG GTT
CAG CTG CTG CCT GAG GCT GGA CGA CCT CGC GGA GTT CTA CCG GCA GTG CAA ATC
CGT CGG CAT CCA GGA AAC CAG CAG CGG CTA TCC GCG CAT CCA TGC CCC CGA ACT
GCA GGA GTG GGG AGG CAC GAT GGC CGC TTT GGT CGA CAA TTC GCG CTA ACT TAC
ATT AAT TGC GTT GCG CTC ACT GCC CGC TTT CCA GTC GGG AAA CCT GTC GTG CCA
GCT GCA TTA ATG AAT CGG CCA ACG CGC GGG GAG AGG CGG TTT GCG TAT TGG GCG
CCA GGG TGG TTT TTC TTT TCA CCA GTG AGA CGG GCA ACA GCT GAT TGC CCT TCA
CCG CCT GGC CCT GAG AGA GTT GCA GCA AGC GGT CCA CGC TGG TTT GCC CCA GCA
GGC GAA AAT CCT GTT TGA TGG TGG TTA ACG GCG GGA TAT AAC ATG AGC TGT CTT
CGG TAT CGT CGT ATC CCA CTA CCG AGA TAT CCG CAC CAA CGC GCA GCC CGG ACT
CGG TAA TGG CGC GCA TTG CGC CCA GCG CCA TCT GAT CGT TGG CAA CCA GCA TCG
CAG TGG GAA CGA TGC CCT CAT TCA GCA TTT GCA TGG TTT GTT GAA AAC CGG ACA
TGG CAC TCC AGT CGC CTT CCC GTT CCG CTA TCG GCT GAA TTT GAT TGC GAG TGA
GAT ATT TAT GCC AGC CAG CCA GAC GCA GAC GCG CCG AGA CAG AAC TTA ATG GGC
CCG CTA ACA GCG CGA TTT GCT GGT GAC CCA ATG CGA CCA GAT GCT CCA CGC CCA
GTC GCG TAC CGT CTT CAT GGG AGA AAA TAA TAC TGT TGA TGG GTG TCT GGT CAG
AGA CAT CAA GAA ATA ACG CCG GAA CAT TAG TGC AGG CAG CTT CCA CAG CAA TGG
CAT CCT GGT CAT CCA GCG GAT AGT TAA TGA TCA GCC CAC TGA CGC GTT GCG CGA
GAA GAT TGT GCA CCG CCG CTT TAC AGG CTT CGA CGC CGC TTC GTT CTA CCA TCG
ACA CCA CCA CGC TGG CAC CCA GTT GAT CGG CGC GAG ATT TAA TCG CCG CGA CAA
TTT GCG ACG GCG CGT GCA GGG CCA GAC TGG AGG TGG CAA CGC CAA TCA GCA ACG
ACT GTT TGC CCG CCA GTT GTT GTG CCA CGC GGT TGG GAA TGT AAT TCA GCT CCG
CCA TCG CCG CTT CCA CTT TTT CCC GCG TTT TCG CAG AAA CGT GGC TGG CCT GGT
TCA CCA CGC GGG AAA CGG TCT GAT AAG AGA CAC CGG CAT ACT CTG CGA CAT CGT
ATA ACG TTA CTG GTT TCA CAT TCA CCA CCC TGA ATT GAC TCT CTT CCG GGC GCT
ATC ATG CCA TAC CGC GAA AGG TTT TGC GCC ATT CGA TGG TGT CAA CGT AAA TGC
ATG CCG CTT CGC CTT CGC GCG CGA ATT GTC GAC CCT GTC CCT CCT GTT CAG CTA
CTG ACG GGG TGG TGC GTA ACG GCA AAA GCA CCG CCG GAC ATC AGC GCT AGC GGA
GTG TAC TGG CGA AAG GGG GAG GTG CTG CAA GGC GAT TAA GTT GGG TAA CGC CAG
GGT TTT CCC AGT CAC GAC GTT GTA AAA CGA CGG CCA GTG CCA AGC TTA AAA AAA
ATC CTT AGC TTT CGC TAA GGA TCT GCA GTG GTG CGA ACT CTG TGG ATC GAA CAC
AGG ACC TCC AGA TTT TTA GTC TGG CGC TCT CCC AAC TGA GCT AAG TCC GCG AAT
TCA GCG TTA CAA GTA TTA CAC AAA GTT TTT TAT GTT GAG AAT ATT TTT TTG ATG
GGG CGC CAC TTA TTT TTG ATC GTT CGC TCA AAG AAG CGG CGC CAG GGT GGT TTT
TCT TTT CAC CAG TGA GAC GGG CAA CAG CTG GCT TAC TAT GTT GGC ACT GAT GAG
GGT GTC AGT GAA GTG CTT CAT GTG GCA GGA GAA AAA AGG CTG CAC CGG TGC GTC
AGC AGA ATA TGT GAT ACA GGA TAT ATT CCG CTT CCT CGC TCA CTG ACT CGC TAC
GCT CGG TCG TTC GAC TGC GGC GAG CGG AAA TGG CTT ACG AAC GGG GCG GAG ATT
TCC TGG AAG ATG CCA GGA AGA TAC TTA ACA GGG AAG TGA GAG GGC CGC GGC AAA
GCC GTT TTT CCA TAG GCT CCG CCC CCC TGA CAA GCA TCA CGA AAT CTG ACG CTC
AAA TCA GTG GTG GCG AAA CCC GAC AGG ACT ATA AAG ATA CCA GGC GTT TCC CCT
GGC GGC TCC CTC GTG CGC TCT CCT GTT CCT GCC TTT CGG TTT ACC GGT GTC ATT

CCG CTG TTA TGG CCG CGT TTG TCT CAT TCC ACG CCT GAC ACT CAG TTC CGG GTA
GGC AGT TCG CTC CAA GCT GGA CTG TAT GCA CGA ACC CCC CGT TCA GTC CGA CCG
CTG CGC CTT ATC CGG TAA CTA TCG TCT TGA GTC CAA CCC GGA AAG ACA TGC AAA
AGC ACC ACT GGC AGC AGC CAC TGG TAA TTG ATT TAG AGG AGT TAG TCT TGA AGT
CAT GCG CCG GTT AAG GCT AAA CTG AAA GGA CAA GTT TTG GTG ACT GCG CTC CTC
CAA GCC AGT TAC CTC GGT TCA AAG AGT TGG TAG CTC AGA GAA CCT TCG AAA AAC
CGC CCT GCA AGG CGG TTT TTT CGT TTT CAG AGC AAG AGA TTA CGC GCA GAC CAA
AAC GAT CTC AAG AAG ATC ATC TTA TTA ATC AGA TAA AAT ATT TCT AGA TTT CAG
TGC AAT TTA TCT CTT CAA ATG TAG CAC CTG AAG TCA GCC CCA TAC GAT ATA AGT
TGT TAA TTC TCA TGT TTG ACA GCT TAT CAT CGA T

Appendix B

Marker Protein Information

1. GFP_{UV} (Clontech) (without a N-terminal hexa-histidine tag)

DNA sequence

```
ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA
GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT
GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT
GTT CCA TGG CCA ACA CTT GTC ACT ACT TTC TCT TAT GGT GTT CAA TGC TTT TCC
CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA
GGT TAT GTA CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC AAG ACG
CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT CGT ATC GAG TTA AAA
GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTC GGA CAC AAA CTC GAG TAC AAC
TAT AAC TCA CAC AAT GTA TAC ATC ACG GCA GAC AAA CAA AAG AAT GGA ATC AAA
GCT AAC TTC AAA ATT CGC CAC AAC ATT GAA GAT GGA TCC GTT CAA CTA GCA GAC
CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC
CAT TAC CTG TCG ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG CGT GAC
CAC ATG GTC CTT CTT GAG TTT GTA ACT GCT GCT GGG ATT ACA CAT GGC ATG GAT
GAG CTC TAC AAA CTG CAG CCA AGC TTA ATT AGC TGA
```

Amino acid sequence

```
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCF
SRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE
YNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE
KRDHMVLLEFVTAAGITHGMDELYKLQPSLIS
```

2. GFP_{UV}

DNA sequence

```
ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GTC GAC CTG CAG CCC CGG
GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA
ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA
GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT
ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTC TCT TAT GGT
GTT CAA TGC TTT TCC CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTT TTC AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC
GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT
CGT ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTC GGA CAC
AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ACG GCA GAC AAA CAA
AAG AAT GGA ATC AAA GCT AAC TTC AAA ATT ATG AGA GGA TCG CAT CAC CAT CAC
CAT CAC GGA TCC GTC GAC CTG CAG CCC CGG GTA CCG GTA GAA AAA ATG AGT AAA
GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT
GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC
GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG
CCA ACA CTT GTC ACT ACT TTC TCT TAT GGT GTT CAA TGC TTT TCC CGT TAT CCG
GAT CAT ATG AAA CGG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA
CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC AAG ACG CGT GCT GAA
GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT CGT ATC GAG TTA AAA GGT ATT GAT
TTT AAA GAA GAT GGA AAC ATT CTC GGA CAC AAA CTC GAG TAC AAC TAT AAC TCA
CAC AAT GTA TAC ATC ACG GCA GAC AAA CAA AAG AAT GGA ATC AAA GCT AAC TTC
AAA ATT CGC CAC AAC ATT GAA GAT GGA TCC GTT CAA CTA GCA GAC CAT TAT CAA
CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG
```

TCG ACA CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG CGT GAC CAC ATG GTC
CTT CTT GAG TTT GTA ACT GCT GCT GGG ATT ACA CAT GGC ATG GAT GAG CTC TAC
AAA CTG CAG CCA AGC TTA ATT AGC TGA

Amino acid sequence

MRGSHHHHHHGSVDLQPRVPVEKMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLQPSLIS

3. GFP3

F64L, S65T, S99F, and T153M mutations in GFP_{UV}

DNA sequence

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GTC GAC CTG CAG CCC CGG
GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA
ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA
GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT
ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTG ACC TAT GGT
GTT CAA TGC TTC TCC CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTT TTC AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TTC TTC AAA GAT GAC
GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT
CGT ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTC GGA CAC
AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA
AAG AAT GGA ATC AAA GCT AAC TTC AAA ATT CGC CAC AAC ATT GAA GAT GGA TCC
GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC
CTT TTA CCA GAC AAC CAT TAC CTG TCG ACA CAA TCT GCC CTT TCG AAA GAT CCC
AAC GAA AAG CGT GAC CAC ATG GTC CTT CTT GAG TTT GTA ACT GCT GCT GGG ATT
ACA CAT GGC ATG GAT GAG CTC TAC AAA CTG CAG CCA AGC TTA ATT AGC TGA

Amino acid sequence

MRGSHHHHHHGSVDLQPRVPVEKMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLQPSLIS

4. GFP6

L64F mutation and all Phe codons are changed to UUU in GFP3

DNA sequence

ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GTC GAC CTG CAG CCC CGG
GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTT ACT GGA GTT GTC CCA
ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA
GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC ACT

```
ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTT ACC TAT GGT
GTT CAA TGC TTT TCC CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTT TTT AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TTT TTT AAA GAT GAC
GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT
CGT ATC GAG TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTC GGA CAC
AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA
AAG AAT GGA ATC AAA GCT AAC TTT AAA ATT CGC CAC AAC ATT GAA GAT GGA TCC
GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC
CTT TTA CCA GAC AAC CAT TAC CTG TCG ACA CAA TCT GCC CTT TCG AAA GAT CCC
AAC GAA AAG CGT GAC CAC ATG GTC CTT CTT GAG TTT GTA ACT GCT GCT GGG ATT
ACA CAT GGC ATG GAT GAG CTC TAC AAA CTG CAG CCA AGC TTA ATT AGC TGA
```

Amino acid sequence

```
MRGSHHHHHHGSVDLQPRVPVEKMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTFTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLQPSLIS
```

5. GFP_WC

All Phe codons are changed to UUC in GFP3

DNA sequence

```
ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC GTC GAC CTG CAG CCC CGG
GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA
ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA
GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTC ATT TGC ACT
ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT CTG ACC TAT GGT
GTT CAA TGC TTC TCC CGT TAT CCG GAT CAT ATG AAA CGG CAT GAC TTC TTC AAG
AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA CGC ACT ATA TTC TTC AAA GAT GAC
GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG TTC GAA GGT GAT ACC CTT GTT AAT
CGT ATC GAG TTA AAA GGT ATT GAT TTC AAA GAA GAT GGA AAC ATT CTC GGA CAC
AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA
AAG AAT GGA ATC AAA GCT AAC TTC AAA ATT CGC CAC AAC ATT GAA GAT GGA TCC
GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT GTC
CTT TTA CCA GAC AAC CAT TAC CTG TCG ACA CAA TCT GCC CTT TCG AAA GAT CCC
AAC GAA AAG CGT GAC CAC ATG GTC CTT CTT GAG TTC GTA ACT GCT GCT GGG ATT
ACA CAT GGC ATG GAT GAG CTC TAC AAA CTG CAG CCA AGC TTA ATT AGC TGA
```

Amino acid sequence

```
MRGSHHHHHHGSVDLQPRVPVEKMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLQPSLIS
```

6. GFP6 (2UUU)

AGA (Arg) to UUU (Phe) mutation in position 2 of GFP6

7. GFP6 (2UUC)

AGA (Arg) to UUU (Phe) mutation in position 2 of GFP6

8. GFP6 (158Am)

AAA (Lys) to UAG (an amber codon) mutation in position 158 of GFP6

9. GFP3C

GFP3 with a C-terminal hexa-histidine tag

DNA sequence

```
ATG CGA GGA TCC CTG CAG CCC CGG GTA CCG GTA GAA AAA ATG AGT AAA GGA GAA
GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT
GGG CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA
CTT ACC CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA
CTT GTC ACT ACT CTG ACC TAT GGT GTT CAA TGC TTC TCC CGT TAT CCG GAT CAT
ATG AAA CGG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG GAA
CGC ACT ATA TTC TTC AAA GAT GAC GGG AAC TAC AAG ACG CGT GCT GAA GTC AAG
TTT GAA GGT GAT ACC CTT GTT AAT CGT ATC GAG TTA AAA GGT ATT GAT TTT AAA
GAA GAT GGA AAC ATT CTC GGA CAC AAA CTC GAG TAC AAC TAT AAC TCA CAC AAT
GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA ATC AAA GCT AAC TTC AAA ATT
CGC CAC AAC ATT GAA GAT GGA TCC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT
ACT CCA ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCG ACA
CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG CGT GAC CAC ATG GTC CTT CTT
GAG TTT GTA ACT GCT GCT GGG ATT ACA CAT GGC ATG GAT GAG CTC TAC AAA CTG
CAG CAT CAC CAT CAC CAT CAC TAA
```

Amino acid sequence

```
MRGSLQPRVPVEKMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWP
TLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID
FKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY
LSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKLQHHHHHH
```

10. GFP3C (158UUG)

AAA (Lys) to UUG (Leu) mutation in position 158 of GFP3C

11. mDHFR (Qiagen)

C-terminal hexa-histidine tag

DNA sequence

```
ATG AGA GGA TCC GGC ATC ATG GTT CGA CCA TTG AAC TCG ATC GTC GCC GTG TCC
CAA AAT ATG GGG ATT GGC AAG AAC GGA GAC CTA CCC TGG CCT CCG CTC AGG AAC
GAG TTC AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT TCA GTG GAA GGT AAA CAG
AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG AAG AAT CGA
CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA CCA CCA
CGA GGA GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT GAA
CAA CCG GAA TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC AGT TCT
GTT TAC CAG GAA GCC ATG AAT CAA CCA GGC CAC CTT AGA CTC TTT GTG ACA AGG
ATC ATG CAG GAA TTT GAA AGT GAC ACG TTT TTC CCA GAA ATT GAT TTG GGG AAA
TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT GAG GTC CAG GAG GAA AAA
GGC ATC
```

Amino acid sequence

```
MRGSGIMVRPLNSIVAVSQNMGIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSIPEKN
RPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSSVYQEAMNQPGHLRLFV
TRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGI
```

12. mDHFR2

UUC (Phe) in position 38 and UUU (Phe) in position 95 to UUG (Leu) mutations in mDHFR

DNA sequence

```
ATG AGA GGA TCC GGC ATC ATG GTT CGA CCA TTG AAC TCG ATC GTC GCC GTG TCC
CAA AAT ATG GGG ATT GGC AAG AAC GGA GAC CTA CCC TGG CCT CCG CTC AGG AAC
GAG TTG AAG TAC TTC CAA AGA ATG ACC ACA ACC TCT TCA GTG GAA GGT AAA CAG
AAT CTG GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG AAG AAT CGA
CCT TTA AAG GAC AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA CCA CCA
CGA GGA GCT CAT TTG CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT GAA
CAA CCG GAA TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC AGT TCT
GTT TAC CAG GAA GCC ATG AAT CAA CCA GGC CAC CTC AGA CTC TTT GTG ACA AGG
ATC ATG CAG GAA TTT GAA AGT GAC ACG TTT TTC CCA GAA ATT GAT TTG GGG AAA
TAT AAA CTT CTC CCA GAA TAC CCA GGC GTC CTC TCT GAG GTC CAG GAG GAA AAA
GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA GGT TCC AGA TCT CAT CAC
CAT CAC CAT CAC
```

Amino acid sequence

```
MRGSGIMVRPLNSIVAVSQNMGIGKNGDLPWPPLRNELKYFQRMTTTSSVEGKQNLVIMGRKTWFSIPEKN
RPLKDRINIVLSRELKEPPRGAHLLAKSLDDALRLIEQPELASKVDMVWIVGGSSVYQEAMNQPGHLRLFV
TRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFEVYEKKGSRSHHHHHH
```

13. mDHFR2 (100UUA)

UUG to UUA mutation in position 100 of mDHFR2

Appendix C

Names and Sequences of PCR Primers

Primer name	Primer sequence
GFP1	5’- AC CTGCAG CCC CGG GTA CCG GTA GAA - 3’
GFP2	5’- AC CTGCAG TTT GTA GAG CTC ATC CAT - 3’
F64LS65T_f	5’- CTT GTC ACT ACT CTG ACC TAT GGT GTT CAA TGC TTC TCC CGT - 3’
F64LS65T_r	5’- ACG GGA GAA GCA TTG AAC ACC ATA GGT CAG AGT AGT GAC AAG - 3’
S99F_f	5’- GTA CAG GAA CGC ACT ATA TTC TTC AAA GAT GAC GGG AAC - 3’
S99F_r	5’- GTT CCC GTC ATC TTT GAA GAA TAT AGT GCG TTC CTG TAC - 3’
T153M_f	5’- CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA - 3’
T153M_r	5’- TCC ATT CTT TTG TTT GTC TGC CAT GAT GTA TAC ATT GTG - 3’
Primer 1	5’- GTG CCA CCT GAC GTC TAA GAA ACC ATT ATT ATC ATG ACA TTA ACC - 3’
Primer 2	5' - GAG TAA AGG AGA AGA ACT TTT TAC TGG AGT TGT CCC AAT TC - 3'
Primer 3	5’- GAA TTG GGA CAA CTC CAG TAA AAA GTT CTT CTC CTT TAC TC - 3’
Primer 4	5’- GGC CAA CAC TTG TCA CTA CTT TTA CCT ATG GTG TTC AAT GCT T - 3’
Primer 5	5’- AAG CAT TGA ACA CCA TAG GTA AAA GTA GTG ACA AGT GTT GGC C - 3’
Primer 6	5’- CAT ATG AAA CGG CAT GAC TTT TTT AAG AGT GCC ATG CCC GAA G - 3’
Primer 7	5’- CTT CGG GCA TGG CAC TCT TAA AAA AGT CAT GCC GTT TCA TAT G - 3’
Primer 8	5’- GTT ATG TAC AGG AAC GCA CTA TAT TTT TCA AAG ATG ACG GGA ACT ACA A - 3’
Primer 9	5’- TTG TAG TTC CCG TCA TCT TTG AAA AAT ATA GTG CGT TCC TGT ACA TAA C - 3’
Primer 10	5’- ACA AAA GAA TGG AAT CAA AGC TAA CTT TAA AAT TCG CCA CAA CAT TGA AGA TG - 3’
Primer 11	5’- CAT CTT CAA TGT TGT GGC GAA TTT TAA AGT TAG CTT TGA TTC CAT TCT TTT GT - 3’
Primer 12	5’- CGC CAA GCT AGC TTG GAT TCT CAC CAA TAA AAA ACG CCC - 3’
Primer 13	5’- TAC CTA TGG TGT TCA ATG CTT TTC CCG TTA TCC GGA TCA TAT G - 3’
Primer 14	5’- CAT ATG ATC CGG ATA ACG GGA AAA GCA TTG AAC ACC ATA GGT A - 3’

Primer 15	5’- GTT ATG TAC AGG AAC GCA CTA TAT TTT TTA AAG ATG ACG GGA ACT ACA AG - 3’
Primer 16	5’- CTT GTA GTT CCC GTC ATC TTT AAA AAA TAT AGT GCG TTC CTG TAC ATA AC - 3’
Primer 17	5’- GT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG GGT- 3’
Primer 18	5’- ACC CTC TCC ACT GAC AGA GAA TTT GTG CCC ATT AAC ATC AC- 3’
Primer 19	5’- CA TAC GGA AAA CTT ACC CTT AAA TTC ATT TGC ACT ACT GGA AAA CTA CC- 3’
Primer 20	5’- GGT AGT TTT CCA GTA GTG CAA ATG AAT TTA AGG GTA AGT TTT CCG TAT G- 3’
Primer 21	5’- AT CAT ATG AAA CGG CAT GAC TTC TTC AAG AGT GCC ATG CCC- 3’
Primer 22	5’- GGG CAT GGC ACT CTT GAA GAA GTC ATG CCG TTT CAT ATG AT- 3’
Primer 23	5’- CG CGT GCT GAA GTC AAG TTC GAA GGT GAT ACC CTT GT- 3’
Primer 24	5’- AC AAG GGT ATC ACC TTC GAA CTT GAC TTC AGC ACG CG- 3’
Primer 25	5’- CGT ATC GAG TTA AAA GGT ATT GAT TTC AAA GAA GAT GGA AAC ATT CTC GGA- 3’
Primer 26	5’- TCC GAG AAT GTT TCC ATC TTC TTT GAA ATC AAT ACC TTT TAA CTC GAT ACG - 3’
*lac*I_*Afl*II_f	5’- CAT ATT CTT AAG TCA TGC GCA CCC GTG GCC AGG ACC C - 3’
*lac*I_*Sac*II_r	5’- CAA CTC CCG CGG TTA CGT GCA GAT GAT AAG CTG TCA AAC - 3’
pQE_*Afl*II_f	5’-CAA GTA CTT AAG CGG TGT GAA ATA CCG CAC AGA TGC GTA AGG -3’
pQE_*Sac*II_r	5’- GAT CAT CCG CGG TTT TCC CAG TCA CGA CGT TGT AAA ACG - 3’
437_f	5’ - GTC GAA ATC GGT AAC NNK GGT ATG TTC AGA CCA GAA ATG CTC G -3’
437_r	5’ - CGA GCA TTT CTG GTC TGA ACA TAC CMN NGT TAC CGA TTT CGA C - 3’
412_415_418_f	5’ - C AAG CCT ACC TAC NNK CCT TAC NNK GAG CCA NNK ATG GAA ATC TTT T - 3’
412_415_418_r	5’- A AAA GAT TTC CAT MNN TGG CTC MNN GTA AGG MN N GTA GG T AGG CTT G - 3’
2F_UUU_f	5’ - C ATT AAA GAG GAG AAA TTA ACT ATG TTT GGA TCG CAT CAC - 3’
2F_UUU_r	5’ - GTG ATG CGA TCC AAA CAT AGT TAA TTT CTC CTC TTT AAT G - 3’
2F_UUC_f	5’ - C ATT AAA GAG GAG AAA TTA ACT ATG TTC GGA TCG CAT CAC - 3’

2F_UUC_r	5’ - GTG ATG CGA TCC GAA CAT AGT TAA TTT CTC CTC TTT AAT G - 3’
R_lacP_f	GCA CCC CAG TCT TTA CAC TAT GCC TAC CGT CTC CGT GAA C
R_lacP_r	GTT CAC GGA GAC GGT AGG CAT AGT GTA AAG ACT GGG GTG C
Ins_SD_A_Del_lacO_f	CGT ATA ATG GAT CC C GAT TTT CAC ATA AGG AGC AGA CC ATG
Ins_SD_A_Del_lacO_r	CAT GGT CTG CTC CTT ATG TGA AAA TCG GGA TCC ATT ATA CG
ytRNA-GAA_f	5’ - CAC AGG ACC TCC AGA TTT TCA GTC TGG CGC TCT CC - 3’
ytRNA-GAA_r	5’ - GGA GAG CGC CAG ACT GAA AAT CTG GAG GTC CTG TG - 3’
ytRNA-CAA_f	5’ - CAC AGG ACC TCC AGA TTT TGA GTC TGG CGC TCT CC - 3’
ytRNA-CAA_r	5’ - GGA GAG CGC CAG ACT CAA AAT CTG GAG GTC CTG TG - 3’
L100_UUA_f	5’ - CTT GCC AAA AGT TTA GAT GAT GCC TTA AGA C - 3’
L100_UUA_r	5’ - GTC TTA AGG CAT CAT CTA AAC TTT TGG CAA G - 3’
PstI_70_f	5’ - GCA TGC GAG GAT CCC TGC AGC ATC ACC ATC ACC - 3’
PstI_70_r	5’- GGT GAT GGT GAT GCT GCA GGG ATC CTC GCA TGC - 3’
K158_UUG_f	5’ - ATG GCA GAC AAA CAA TTG AAC GGA ATC AAA GCT AAC TTC -3’
K158_UUG_r	5’ - GAA GTT AGC TTT GAT TCC GTT CAA TTG TTT GTC TGC CAT - 3’
K158_UAG_f	5’ - CAT CAT GGC AGA CAA ACA ATA GAA TGG AAT CAA AGC TAA C - 3’
K158_UAG_r	5’ - GTT AGC TTT GAT TCC ATT CTA TTG TTT GTC TGC CAT GAT G- 3’