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 (*mut*RNA^{Phe}(CUA)) was constitutively expressed under control of an *lpp* promoter. The expression cassette for *mut*RNA^{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.

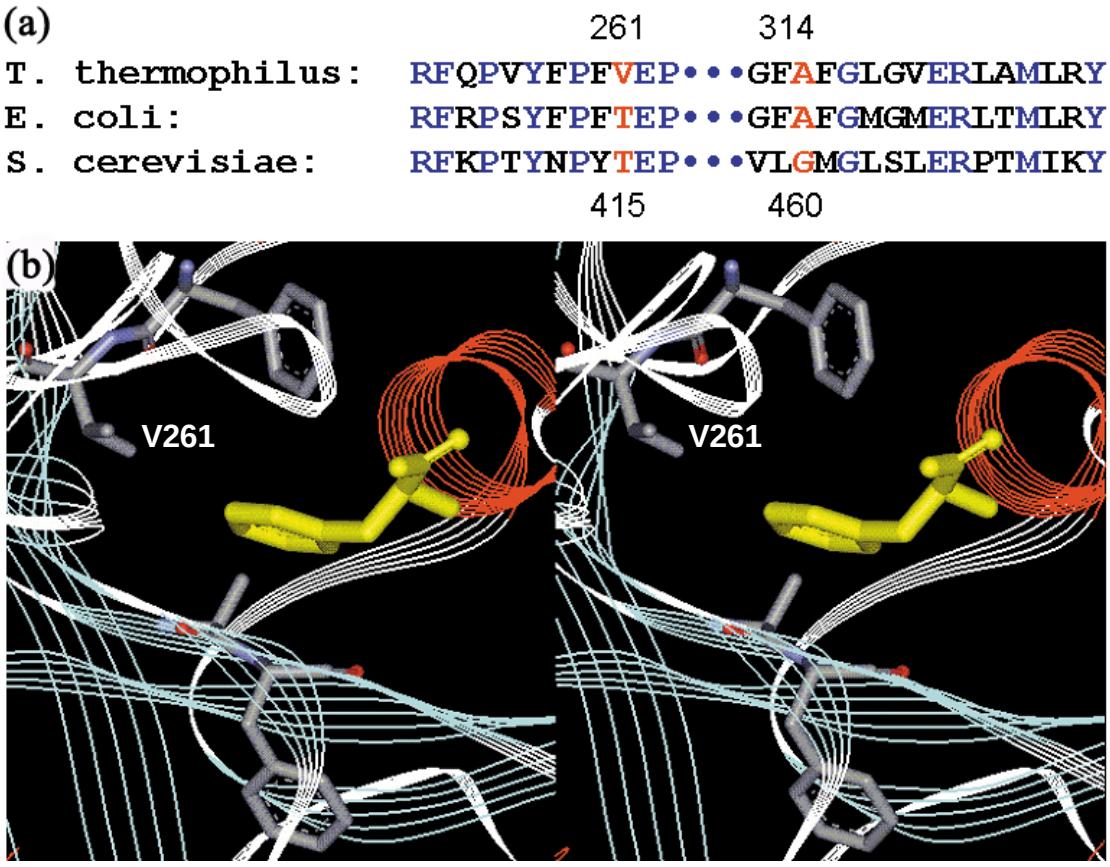

Figure 1: (a) Sequence alignment of PheRS variants from *Thermus thermophilus*, *Escherichia coli*, *Saccharomyces cerevisiae.* Only portions of sequences flanking the V261 and A314 positions in *T. thermophilus* are shown. The sequences adjacent to V261 and A314 are highly conserved. The equavilent 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.

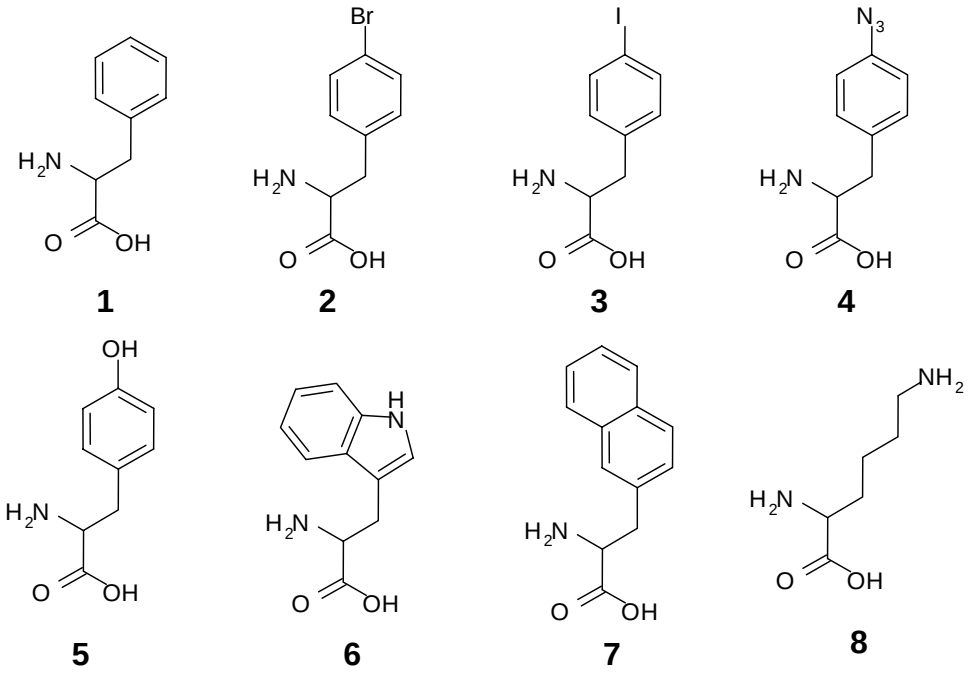

Figure 2: Amino acids involved in this study. (**1**) phenylalanine; (**2**) *p*-bromophenylalanine; (**3**) *p*-iodophenylalanine; (**4**) *p*-azidophenylalanine; (**5**) tyrosine; (**6**) trytophan; (**7**) 3-(2-naphthyl)alanine; and (**8**) L-lysine.

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

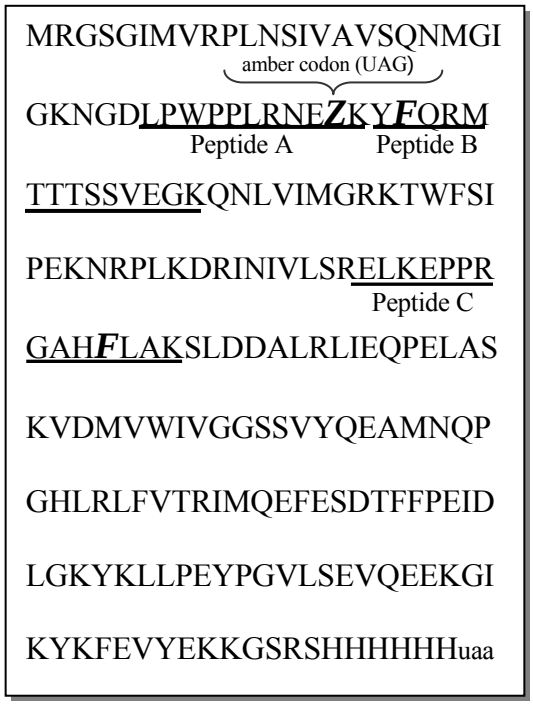

Figure 3: Amino acid sequence of marker protein mDHFR (38Am). Four proteolytic peptide fragments (Peptide A, Peptide B, and Peptide C) used for MALDI-MS and LC-MS/MS analyses are underscored. The positions with amber codon and ochre codon are labeled.

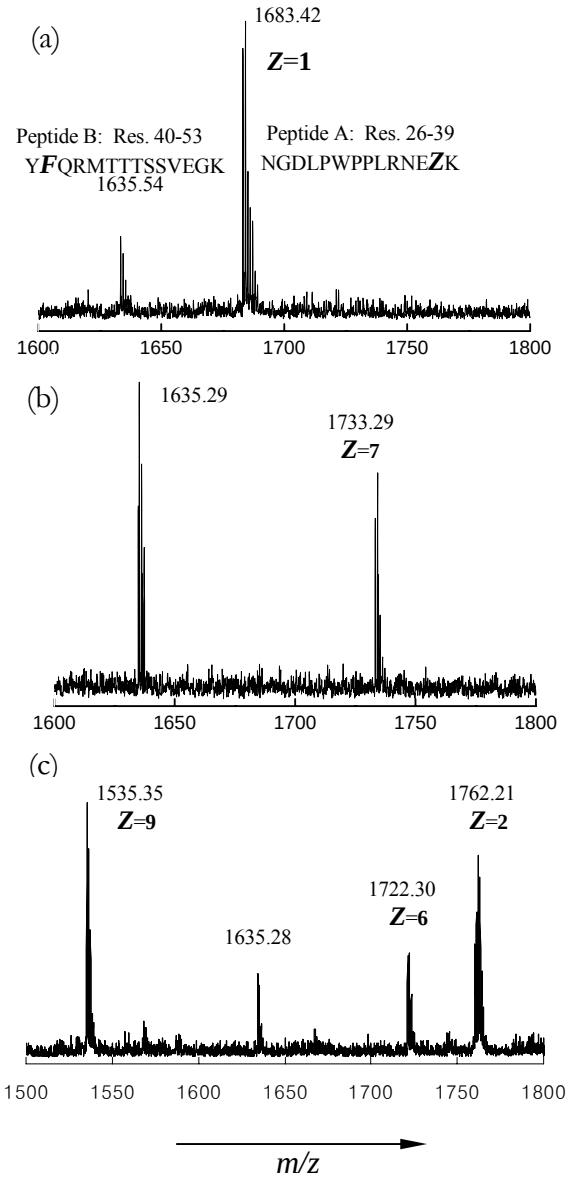
(a)
1683.42
Z=1
Peptide B: Res. 40-53
YFQRMTTTSSVEGK
1635.54
Peptide A: Res. 26-39
NGDLPWPPLRNEZK
1600
1650
1700
1750
1800
(b)
1635.29
1733.29
Z=7
1600
1650
1700
1750
1800
(c)
1535.35
Z=9
1762.21
Z=2
1722.30
Z=6
1635.28
1500
1550
1600
1650
1700
1750
1800
m/z

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).

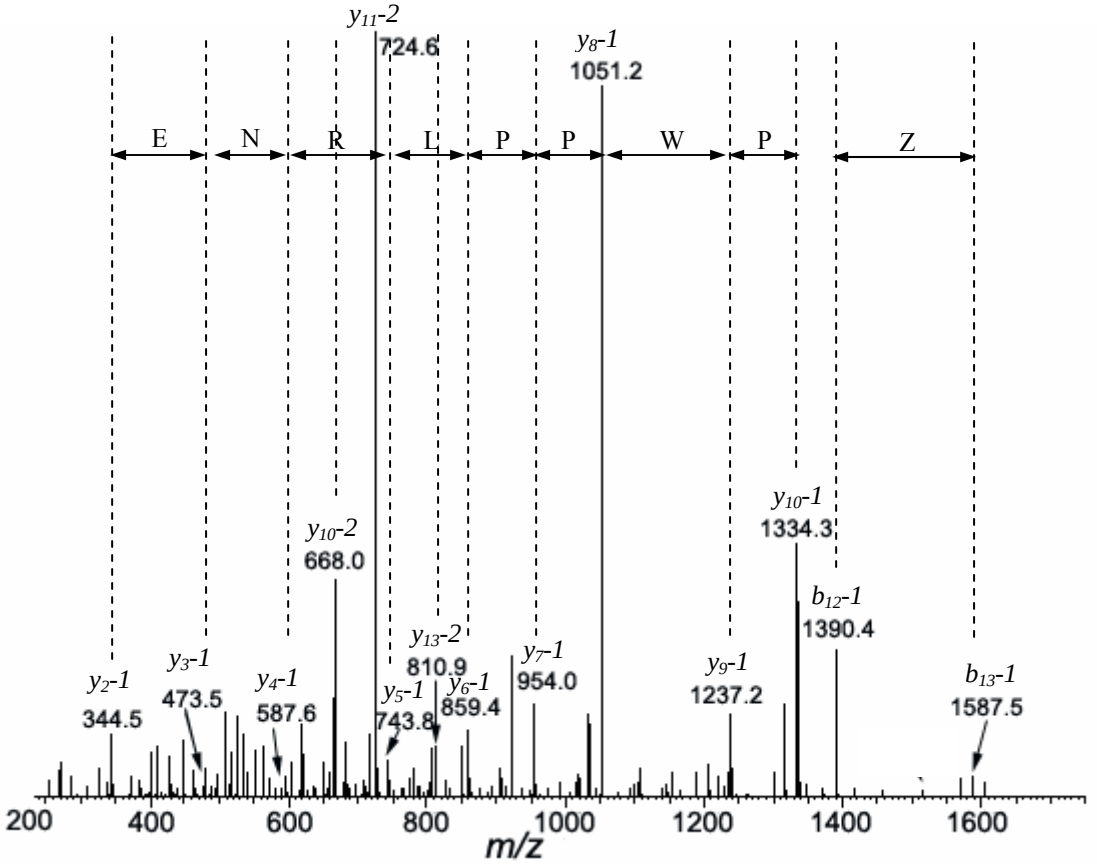

Figure 5: The tandem mass spectrum of Peptide A (NGDLPWPPLRNE**Z**K) derived from mDHFR (38Am) expressed in media supplemented with **7** (3 mM)/**1** (0.03 mM). mDHFR (38Am) was expressed in phenylalanine auxotrophic strain (AF) transformed with the mutant yeast PheRS and cognate suppressor tRNA. Partial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

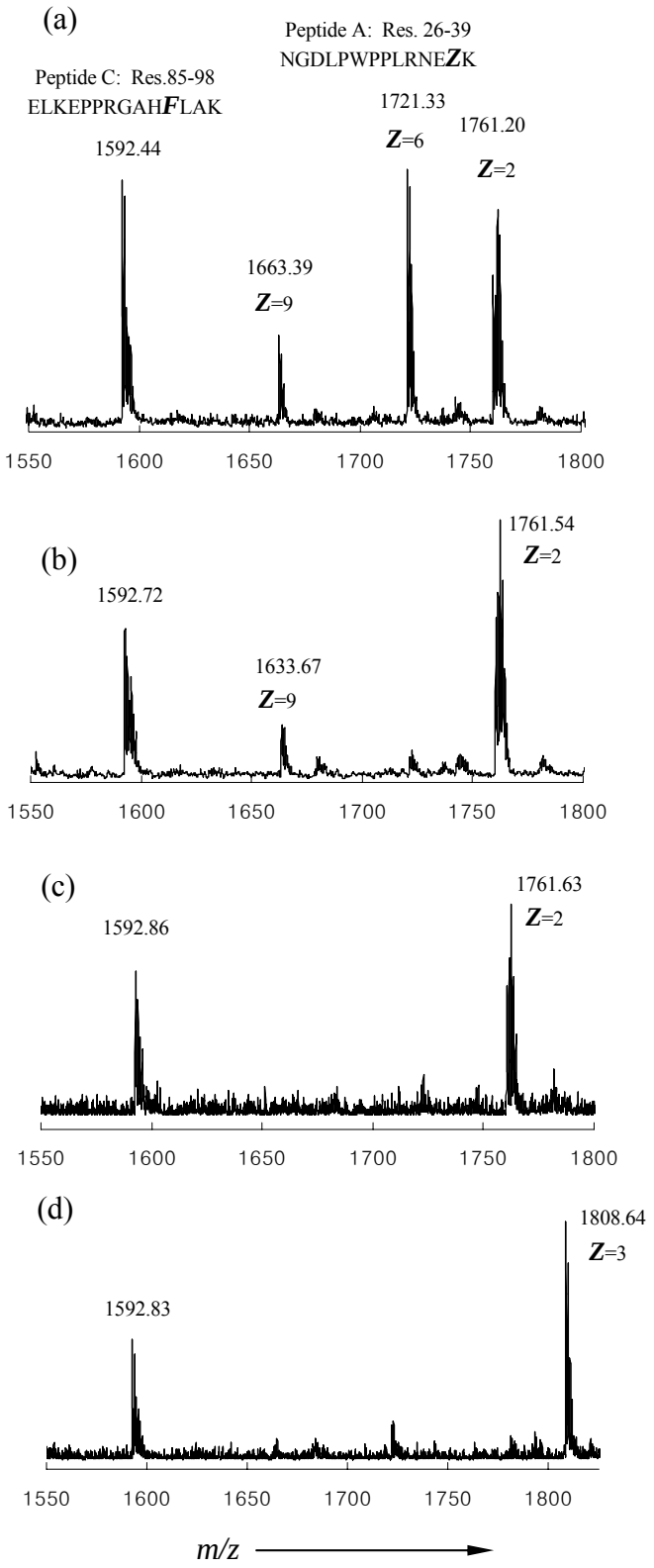
(a)
Peptide A: Res. 26-39
NGDLPWPPLRNEZK
Peptide C: Res.85-98
ELKEPPRGAHFLAK
1721.33
Z=6
1761.20
Z=2
1592.44
1663.39
Z=9
1550
1600
1650
1700
1750
1800
(b)
1761.54
Z=2
1592.72
1633.67
Z=9
1550
1600
1650
1700
1750
1800
(c)
1761.63
Z=2
1592.86
1550
1600
1650
1700
1750
1800
(d)
1808.64
Z=3
1592.83
1550
1600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