Chapter 8

Site-specific Incorporation of Amino Acid Analogs into Proteins *In Vivo* by an Engineered Yeast Phenylalanyl-tRNA Synthetase

Portions of this chapter are adapted from a paper in preparation by Pin Wang, Inchan Kwon, Soojin Son, Yi Tang and David A. Tirrell

Abstract

Aminoacyl-tRNA synthetases (aaRSs) catalyze aminoacylation and establish the rules of genetic code. Precise manipulation of synthetase activity can alter the aminoacylation specificity to attach non-canonical amino acids to the intended transfer RNA (tRNA). Subsequently by codon-anticodon interaction between messenger RNA (mRNA) and tRNA the amino acid analogs can be assigned to specific sites in the growing polypeptide chain. Thus introduction of non-natural amino acids into proteins *in vivo* relies heavily on manipulation of amino acid specificity of aaRS. In this chapter, we generated and characterized a mutant phenylalanyl-tRNA synthetase (PheRS) from *Saccharomyces cerevisiae* with a point mutation (T415G) in the α-subunit of the enzyme. The promiscuous activity of this mutant was extensively explored by ATP-PP_i exchange assays *in vitro*. A broad activation profile toward many non-natural amino acids was observed. A phenylalanine auxotrophic *E. coli* strain transformed with this mutant synthetase and its cognate suppressor tRNA enable the assignment of an amber nonsense codon to the non-natural amino acids 3-(2-naphthyl)alanine and 3-benzothiophenylalanine. Therefore, this variant synthease and its cognate tRNA could serve as an additional "21st" pair for site-specific incorporation of novel amino acids into proteins *in vivo*.

1. Introduction

Aminoacyl-tRNA synthetases (aaRSs) insure the high fidelity of translation of genetic code sequences into functional proteins through a two-step aminoacylation reaction (Scheme 1) (*1-6*). In first step, the cognate amino acid is activated by aaRS in the presence of ATP to form the amino acid adenylate; subsequently the 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 aaRS (*4, 7, 8*). Manipulation of such reaction could potentially alter the genetic code to allow incorporation of novel amino acids into proteins *in vivo* (*9, 10*).

With depletion of the intracellular pool of one natural amino acid in a bacterial auxotrophic cells, the cellular level of aminoacylation is perturbed in a way that the aaRS is forced to charge the non-natural amino acid supplemented in the cultures as a surrogate into the cognate tRNA (*10*). Through the subsequent translation by the ribosomal machinery *in vivo* the genetic code is re-assigned and the non-natural amino acid is incorporated into proteins. We have successfully implemented this method to incorporate alkenes (*11*), alkynes (*12*), alkyl azides (*13*) and fluorinated side chains (*14*) into proteins. By enhancing the cellular aminoacylation reactivity through over-expression of the wild-type aaRSs in the host, we found that some sluggish amino acid analogs could also be introduced into proteins (*15-17*). The methods described above require the cellular wild-type aaRSs to recognize these novel amino acids. Engineering new synthetase activities is almost inevitable when one wants to incorporate the amino acid that is not recognizable

by wild-type enzyme. To date, we and others have shown that re-design of the synthetic site of synthetases (*18-20*) and attenuation of editing function of synthetases (*21, 22*) could be two complementary strategies to further expand our ability to introduce non-natural amino acids 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 (*10*). By manipulation of both the synthetase and the tRNA in the host, one can accomplish the site-specific introduction of a single copy of a novel amino acid into proteins *in vivo* (*23-31*). This method is derived from an *in vitro* approach to non-natural 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 of aminoacylation *in vitro* suffers from technical difficulty and the intrinsic low yield of protein production, which limits its application (*36*). In 1998, Furter engineered the cellular aminoacylation reaction by importing a "twenty-first pair" comprising a yeast suppressor tRNA and a yeast phenylalanyl-tRNA synthetase (*y*PheRS) (*23*). This approach allowed efficient site-specifically incorporation of *p*-fluorophenylalanine *in vivo* in response to an amber codon. Recently Schultz and coworkers have devised a powerful selection method to find mutant forms of tyrosyl-tRNA synthetase from archaebacterium *Methanococcus jannaschii* (*9, 24*). By introduction of such a mutant and a cognate suppressor, a number of non-natural amino acids have been incorporated into proteins *in vivo* (*24-31*).

In this chapter, we present a rationally designed variant of *y*PheRS with new amino acid specificity, allowing incorporation of aromatic amino acid analogs (**7-8**) into

proteins site-specifically. The variant of *y*PheRS 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 a number of aromatic amino acid (**1-12**). We observed that this variant activates tryptophan (Trp, **6**) 82-fold faster than cognate substrate phenylalanine (Phe, **1**). An phenylalanine auxotrophic bacterial strain transformed with this mutant and the cognate suppressor tRNA can assign analog **7** to the amber codon with rather high fidelity (>95%). Such a transformed strain also allows the incorporation of **8** into proteins at the amber site with competitive misincorporation of Trp.

2. Materials and Methods

2.1. Materials

Amino acids **1** and **5-6** were obtained from Sigma (St. Louis, MO). **2-4** and **9-11** were purchased from Chem-Impex (Wood Dale, IL). **7, 8** and **12** were purchased from RSP Amino Acid Analogues (Shirley, MA). [^{32}P]-labeled sodium pyrophosphate was purchased from NEN Life Science (Boston, MA).

2.2. Plasmid Construction for Synthetase Expression

The *y*PheRS gene was amplified from a template plasmid pUC-ASab2 (*23, 37*) encoding the α- and β-subunits of PheRS with a 14 base pair intergenic sequence containing a translational reinitiation site upstream of ATG start code of β-subunit gene. The following primers were used for the 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 intended mutant *y*PheRS was generated by using a 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. For 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 the additional outside primers, a 3400-bp fragment of DNA was obtained and purified. This DNA was subjected to digestion by *BamH*I and *Kpn*I and inserted into pQE32 to yield pQE32-T415G. The cloned enzymes contain an N-terminal leader sequence MRGSHHHHHHGIQTMNLE to facilitate protein purification. The entire *y*PheRS gene was DNA-sequenced for each of these constructs to verify its integrity. Proofreading polymerase *Pfx* (Invitrogen) was used for all PCR described in this work.

2.3. 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]. Over-expression was conducted in LB media; at an OD of 0.6, over-expression of wild-type and mutant forms of *y*PheRS was induced with 1 mM

isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4 hours. Then 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 on an ion-exchange column 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 in 50% glycerol. The concentration of the protein was determined by measuring the absorbance at 280 nm under denaturing conditions.

2.4. Amino Acid Activation Assays

The amino acid-dependent ATP-PP_i exchange reaction was used to evaluate the activation of amino acid analogs by *y*PheRS. The 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 activity of the analogs, the amino acid concentration varied from 10 µM to 5 mM and 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 charcoal was quantified via liquid scintillation methods. The specificity constants were calculated by nonlinear regression fit of the data to a Michaelis Menten model.

2.5. Plasmid Construction for *In Vivo* Incorporation Assays

pQE16 (Qiagen) was chosen as our expression plasmid, and encodes the marker protein murine dihydrofolate reducase (mDHFR) with C-terminal $(His)_6$-tag gene under control of a bacteriophage T5 promoter and t_0 terminator. A quick-change mutagenesis kit was used to place an amber codon (TAG) at the 38th position of mDHFR 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 gene T415G 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 (*23*). The entire expression cassette of T415G was inserted into the *Pvu*II site of pQE16 to yield pQE16am-T415G. The expression cassette of *mut*RNA^{Phe}(CUA) was inserted into repressor plasmid pREP4 to form pREP4-tRNA as described by Furter (*23*).

2.6. Analog Incorporation Assays *In Vivo*

The phenylalanine auxotrophic strain AF (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi $T2^R$ *pheA*18) (*23*) was transformed with plasmid pQE16am-T415G and pREP4-tRNA to afford expression strain AF[pQE16am-T415G/pREP4-tRNA]. The *E. coli* expression strains were grown in M9 minimal medium supplemented with glucose, thiamin, $MgSO_4$, $CaCl_2$, 20 natural amino acids (at 20 mg/L), 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 3 mM analog of interest and the indicated concentrations of Phe and Trp. In a control expression, cells were shifted to medium supplemented with 0.2 mM Phe and Trp.

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 virtue of C-terminal $(His)_6$ tag through a Ni-NTA spin column according to the supplier's instructions (Qiagen). The aliquots of purified protein were used for SDS-PAGE analysis.

2.7. Composition Analysis of the Mutant mDHFR

The mutant mDHFR purified under denaturing conditions was 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 $(NH4)_2CO_3$ 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 then subjected to $ZipTip_{C18}$ (Millipore) purification and the digested peptides were eluted with 3 μL of 50% CH_3CN and 0.1% TFA, of which 1 μL was used for matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis with α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid as matrices. 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 a 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 HPLC pump. Peptides were separated by a Magic C18 HPLC column (5 micron, 200 Å, 0.3mm ID x 150 mm length, Michron Bioresources) and eluted at a flow rate of 10 μL/min using a linear gradient of 0-60% of solvent B (acetonitrile) over 45 min. The column eluent flow to the electrospray source and the tandem mass sequencing was carried out by fragmentation of the precursor ion with m/z corresponding to the protease-digested fragment that contains the residue at positon 38 of mutant mDHFR.

3. Results and Discussion

3.1 Rationale for Engineering the New Synthetase Specificity

Engineering the new synthetase specificity toward amino acid analog is the one of critical steps for the successful implementation of the general strategy to incorporate non-natural amino acids into proteins *in vivo* (*9*). We have been especially interested in exploring the possibilities of incorporating substituted phenylalanines (Figure 8-2) such as *p*-bromophenylalanine (**2**), *p*-iodophenylalanine (**3**), *p*-azidophenylalanine (**4**), 3-(2-naphthyl)alanine (**7**), etc. 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 (*t*PheRS) complexed with phenylalanine is available (*38*) and the sequence identity (vs *Saccharomyces cerevisiae*) in the active site region is about 40% (Figure 8-1). We used this crystal structure as a starting point for design of a yeast mutant PheRS that charges the non-

natural amino acids of interest. In the previous work, we identified two residues, V261 and A314, which constitute a back-wall structure to hold substrate phenylalanine and hinder binding of amino acids larger than Phe into the active site of *T. thermophilus* PheRS (*20*). In fact, an *E. coli* PheRS variant prepared by mutation of the corresponding residues (T261G and A314G, Figure 8-1), has been shown to exhibit relaxed substrate specificity (*20*). 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.

3.2 Synthetase Expression and Purification

The wild-type *y*PheRS gene was obtained from PCR amplification of original vector pUC-ASab2 encoding the α- and β-subunits of yeast PheRS and a 14 base pair intergenic sequence from operon of *E. coli pheST* gene in the middle (our assignment of α- and β-subunits are opposed to what was assigned by Sanni et al.; we followed the tradition that the catalytic subunit of PheRS is designated as the α-subunit (*38*)) (*37, 39*). Four-primer mutagenesis was employed to create the mutant T415G. The synthetases were over-expressed by *E. coli* BLR strains under the bacteriophage *T5* promoter. The N-terminal $(His)_6$ tag fused to α-subunit of *y*PheRS facilitated our purification of the enzymes under native conditions. Both *E. coli* and yeast synthetases are $\alpha_2\beta_2$ hetero-tetramers and their molecular weighs for each subunit are rather different (α(*e*PheRS)=37 kDa; α(*y*PheRS)=57 kDa; β(*e*PheRS)=87 kDa; β(*y*PheRS)=67.5). The SDS-PAGE analysis of purified *y*PheRS indicated no observable contamination of *E. coli* PheRS. The intensity of bands for α- and β-subunits are not stoichiometric (α-subunit is stronger

than β-subunit); Sanni and co-workers reported the similar observation and suggested that this stems from proteolysis of the β-subunit resulting in one fragment from this proteolysis co-migrating with the α-subunit (*37*). So some portions of purified *y*PheRS were 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 (**2-12**) *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; we recognized that the absolute value of k_{cat} may be off.

3.3 Amino Acid Specificity of the Mutant Yeast PheRS

The activities of *y*PheRS variant with Phe and analogs (**2-12**) were examined via the amino acid-dependent ATP-PP_i exchange assay. As the first step reaction (Scheme 1), *y*PheRS recognizes amino acid substrate and ATP and catalyzes the formation of amino acid adenylate with the release of pyrophosphates. With incubation of amino acid, ATP, [^{32}P]-labeled pyrophosphate and enzyme, the reverse reaction consumes labeled PP_i to generate labeled ATP, which is quantified via liquid scintillation method. The kinetic parameters for amino acids **1-12** are shown in Figure 8-3 and Table 8-1. Our measured value of K_m for **1** by the wild-type yeast PheRS matched previously reported value (*40*); the obtained k_{cat} was lower than that reported (*40*) and this difference might come from the different buffer conditions and different methods to determine enzyme concentration. The specificity constant of k_{cat}/K_m for **1** by the mutant T415G is reduced approximately 40-fold compared to **1** by wild-type enzyme. The loss of catalytic activity is apparent in

both increasing value of K_m and decreasing value of k_{cat}, which suggests that T415G mutation significantly impairs both binding and catalysis. However, this mutant exhibited higher activity for analog **2-4** and **6-12** with values of k_{cat}/K_m ranging from 3 to 385-fold higher than that for Phe (**1**). It appears that mutation efficiently opens up the active site of the synthetase so that the T415G favors binding and catalysis of bulkier substrates. The canonical amino acid Trp (**6**) showed relatively high activity as manifested by 82-fold higher specificity constant than that of **1**. The derivatives of Trp (**8-11**) revealed different activities with respect to substitution at different position of indole ring. Substitution at the 6th position (**9-10**) was highly favorable and **10** displayed the highest activity among all the analogs we tested, suggesting that mutation tends to be more effective in response to this position and substitution could complement nicely to the cavity created by this mutation. The spectrum of activity of the mutant *y*PheRS supports our proposal that the T415G mutation would create extra room to allow binding of amino acids larger than **1**.

3.4. Site-specific Incorporation of Analogs into mDHFR *In Vivo* Using a Phenylalanine Auxotrophic Strain.

To test the utility of the engineered *y*PheRS to incorporate aromatic amino acid analogs into proteins *in vivo* at a specific site, a constitutive expression system for the mutant synthetase was constructed. An amber stop codon was placed at the desired site (Phe38) of marker protein mDHFR with a fused C-terminal $(His)_6$ tag followed by an ochre stop codon as a programmed stop codon (Figure 8-4). The mutant yeast suppressor tRNA (*mut*RNA^{Phe}(CUA)) was constitutively expressed under control of an *lpp* promoter

(23). Phenylalanine auxotrophic cells transformed with mutant *y*PheRS, mDHFR and *mut*RNA^{Phe}(CUA) genes were grown in M9 minimal media supplemented with 20 natural amino acids. When OD_{600}=1, cells were switched to fresh M9 media supplemented with 18 natural amino acids (20 mg/L), analog of interest (3 mM), and Phe and Trp at controlled concentrations. mDHFR was induced for 3 hours, after which the marker protein was isolated by Ni-NTA affinity chromatography. The levels of full-length protein production were examined by SDS-PAGE analysis. Without over-expression of suppressor *mut*RNA^{Phe}(CUA), we did not obtain detectable levels of full-length mDHFR (41). As a positive control, we co-expressed wild-type *y*PheRS with *mut*RNA^{Phe}(CUA) and efficient suppression by phenylalanine was observed (41). When mutant *y*PheRS was co-expressed with *mut*RNA^{Phe}(CUA) in the presence of analogs **2-12** (3 mM), full length mDHFR was formed (41). Overall protein yields varied depending on the analog (ca. 2-10 mg/L, estimated by SDS-PAGE). As a comparison, protein yield in the same strain for production of mDHFR (without the amber codon) in minimal media 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-TOF and liquid chromatography tandem mass spectrometry for purified protein samples. Following work-up of proteolytic protein digestion by endoproteinase Lys-C, two peptides in the mass range 1600-1800 Da consistently appeared in the mass spectra and could be assigned to residues 26-39 (peptide A) and 40-53 (peptide B) of mDHFR, respectively (Figure 8-4, Figure 8-5). Peptide A has the amber codon at the 38th position; peptide B

contains a single phenylalanine residue at position 41. In the control experiment, peptides A and B derived from mDHFR expressed in media supplemented with **1** (3 mM) and without **6** is shown in Figure 8-5a; the mass of peptide A indicates that **1** is incorporated into the amber position. Our *in vitro* assay shows that T415G mutant synthetase activates **6** 82-fold faster than **1**; *in vivo* translational experiment reveals that **1** is favored over **6** to enter amber site, despite the fact that the Phe auxotrophic strain can still synthesize **6**. When in the media supplemented with **7** (3 mM) and **1** (0.03 mM), we unambiguously confirmed the incorporation of L-3-(2-naphthyl)alanine (**7)** into the amber site as shown in Figure 8-5b; peptide A was shifted up in mass by 50 Da, consistent with the increased mass of **7** relative to **1**; peptide B was unchanged in mass, indicating that the normal Phe codon site remains to be occupied by **1**. Liquid chromatography tandem mass spectrometry of the digested peptides was employed to further confirm the sequence of the peptide A. The precursor ion at m/z 867.7 Da, which corresponds to the doubly charged ion of the peptide A is separated and fragmented with an ion trap mass spectrometer. As shown in Figure 8-6, the sequence information obtained from the spectrum clearly reveals the site-specific incorporation of **7** into mDHFR at intended site; we could find no indication of any other amino acid at the amber site after carefully checking all possible peptides with the mass 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 *y*PheRS showed that the specificity constant (k_{cat}/K_m) for **7** is 1,550±125 ($M^{-1}s^{-1}$) (Table 8-1), 8-fold larger than that for **1**. Therefore, when the ratio of **7** to **1** in the culture media is high (100 to 1), $tRNA^{Phe}$(CUA)

should be charged predominantly with **7**, which could rationalize our observation of high percentage incorporation of **7** into proteins at the amber site.

Similar analyses show incorporation of **8** into mDHFR, but we observed competitive misincorporation of tryptophan (**6**) (Figure 8-5c) at the amber site, despite the fact that we did not supplement **6** into the culture media. Incorporation of Trp is not surprising in view of the *in vitro* data; activation of **8** by T415G is 23-fold slower than that of **6** (Table 8-1). Although high concentration (millimolar) of analog was added into growth media, the intracellular concentration usually remained to be in micromolar range (*18*). So the difference of intracellular concentrations of **8** and **6** is much less than that in culture media. Intracellular **6** is able to compete significantly with **8** for attachment to *mu*tRNA^{Phe}(CUA) and insertion into the amber site. The fact that we did not observe misincorporation of **6** with **7** (Figure 8-5b) could be due to a higher intracellular concentration of **7** and the better recognition of **7** by mutant synthetase than that of **8** (Table 8-1).

For the other analogs (**2-4**), neither MALDI analysis nor LC-MS/MS produced any evidence of analog incorporation, although *in vitro* amino acid activation assays showed substantial activity; instead we always observed the incorporation of **6** at the amber site (we only showed the example of **2** in Figure 8-5d). We reasoned that intracellular pool of **6** is enough to compete with analogs (**2-4**) so **6** is predominantly charged to suppressor tRNA and incorporated into the amber position.

Using a phenylalanine and tryptophan double auxotrophic strain, Kwan in our laboratory has recently incorporated several other analogs (**2-4** and **9-10**) into recombinant proteins in response to the amber codon. The results clearly show that the

high concentration of cellular Trp interferes with analog incorporation; with depletion of Trp in media, the mutant yeast PheRS is able to attach analogs to suppressor tRNA and incorporate them at the intended amber site. *In vivo* results also indicated no incorporation of analog **11** and **12** into proteins at the desired amber site. The discrepancy between the results of *in vivo* translation and *in vitro* activation assays for **11** could be rationalized by the hypothetical model proposed by Sisido and co-workers for the adaptability of non-natural amino acids to the *E. coli* ribosome (*42*). The model suggests that *E. coli* ribosome has the ability to define the allowed and excluded regions of the aromatic groups of L-arylalanine to be bound for further translations. We reason that substitution at the 5th position of indole ring could result in intolerance of **11** by ribosome, while ribosome is bearable for tryptophan analogs substituted at the 6th position (**9-11**). Similar argument can be applied to analogs **7** and **12**; **12** is active *in vitro* by mutant synthetase (Table 8-1) but translationally inactive *in vivo*; as reported before, ribosome favors aromatic configurations of analog **7**, not **12** (*42*).

4. Conclusion

The amino acid binding pocket of yeast PheRS appears to be highly designable and the synthetase can be engineered to accommodate a variety of aromatic side chains. Cells outfitted with mutant synthetase and suppressor tRNA can site-specifically incorporate many of amino acid analogs (**2-4, 7-10**) into proteins in response to UAG codon. Coupled with recent reports from the Schultz laboratory (*24-31*), 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 (*28, 31, 36, 43-45*). 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 (*46*). Hydrophobic nature of naphthalene (**7**) can be applied to design specific ligand-receptor recognition (*47*). Chemical utilities of these side chains equip us the new means to site-specifically modify proteins, most of which are orthogonal to natural amino acid side chains in the proteins. Proteins with halo-aryl side chains can be subjected to palladium-mediated coupling reactions to attach molecules containing ethynyl group (*48*). Azide chains in the proteins can be applied to staudinger ligation to tether molecules associated with triarylphosphine group (*13*). Recently there is a growing interest in possible bioconjugation of proteins using copper(I)-catalyzed azide-alkyne cycloaddition reactions (*49-52*). Our group is currently undergoing these investigations using recombinant proteins biosynthesized *in vivo* in the engineered bacterial hosts described in this report.

Scheme 1

$$aaRS + aa + ATP \leftrightarrow aaRS \bullet (aa - AMP) + PP_i \quad (1)$$

$$aaRS \bullet (aa - AMP) + tRNA^{aa} \rightarrow aa - tRNA^{aa} + aaRS + AMP \quad (2)$$

Table 8-1: The kinetic Parameters for the ATP-PPi exchange of Amino Acids (1-12) by the Yeast Mutant 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	184±30	1
2	T415G	22±3	0.03±0.001	1,538±228	8
3	T415G	12±2	0.05±0.001	4,365±797	24
4	T415G	11±3	0.05±0.002	4,558±1,186	25
5	T415G	757±149	0.04±0.003	48±10	1/4
6	T415G	20±5	0.30±0.006	15,000±4,063	82
7	T415G	27±2	0.04±0.001	1,550±125	8
8	T415G	52±10	0.03±0.001	531±102	3
9	T415G	20±8	0.20±0.018	10,256±4,562	56
10	T415G	8±4	0.55±0.097	70,876±34,843	385
11	T415G	31±18	0.06±0.005	1,939±1,149	10
12	T415G	94±50	0.05±0.006	533±293	3
1	Wild-type	68±20	0.52±0.093	7,627±2,664	41

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

Figure 8-1

(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.

(a)

```
                          261       314
T. thermophilus:   RFQPVYFPFVEP•••GFAFGLGVERLAMLRY
E. coli:           RFRPSYFPFTEP•••GFAFGMGMERLTMLRY
S. cerevisiae:     RFKPTYNPYTEP•••VLGMGLSLERPTMIKY
                          415       460
```

(b)

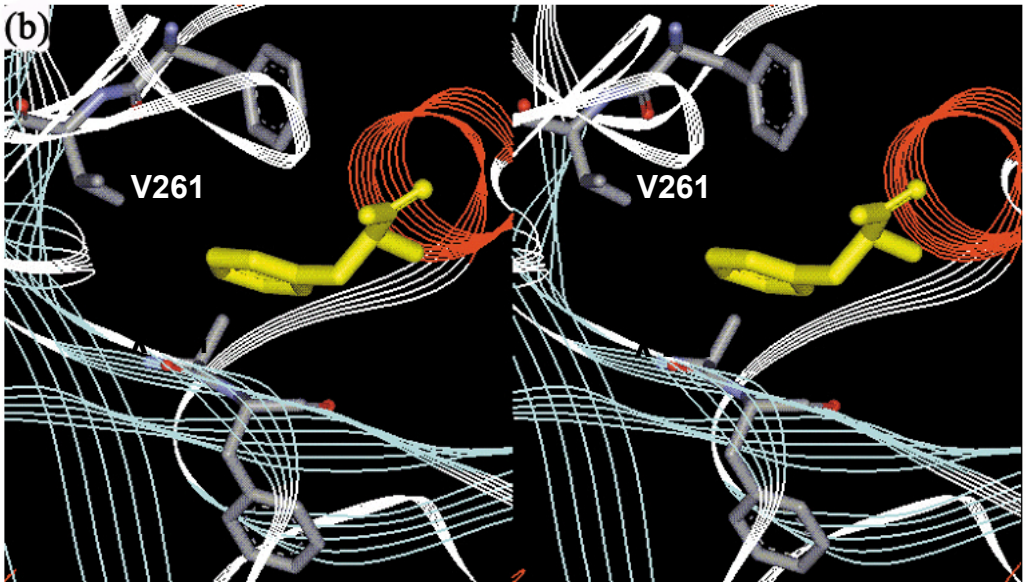
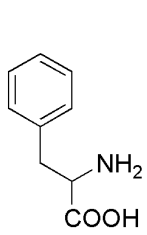
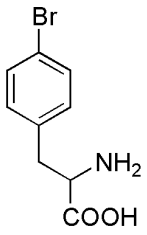


Figure 8-2

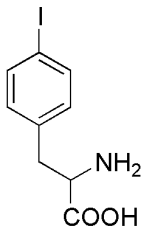
Amino acids involved in this study. (**1**) phenylalanine; (2) *p*-bromophenylalanine; (**3**) *p*-idiophenylalanine; (**4**) *p*-azidophenylalanine; (**5**) tyrosine; (**6**) trytophan; (**7**) 3-(2-naphthyl)alanine; (**8**) 3-benzothiophenylalaline; (**9**) 3-(6-chloroindolyl)alanine; (**10**) 3-(6-bromoindolyl)alanine; (**11**) 3-(5-bromoindolyl)alanine; (**12**) 3-(1-naphthyl)alanine.



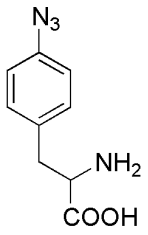
1



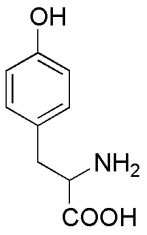
2



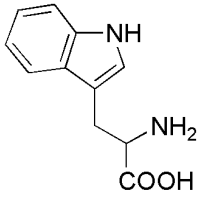
3



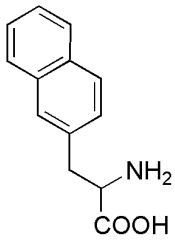
4



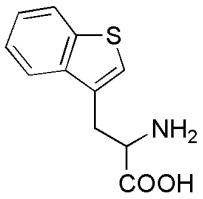
5



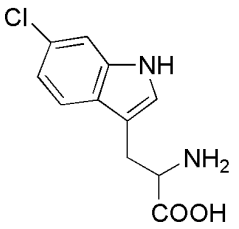
6



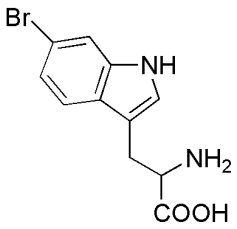
7



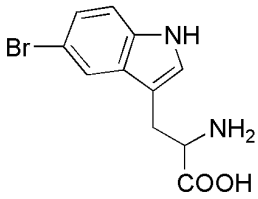
8



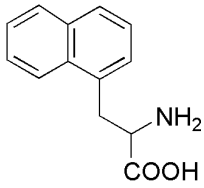
9



10



11



12

Figure 8-3

Activation of amino acids (**1-12**) by yeast mutant PheRS. The relative specificity constant $k_{cat}/K_m(rel)$ is the specificity constant relative to k_{cat}/K_m for **1** by yeast mutant PheRS.

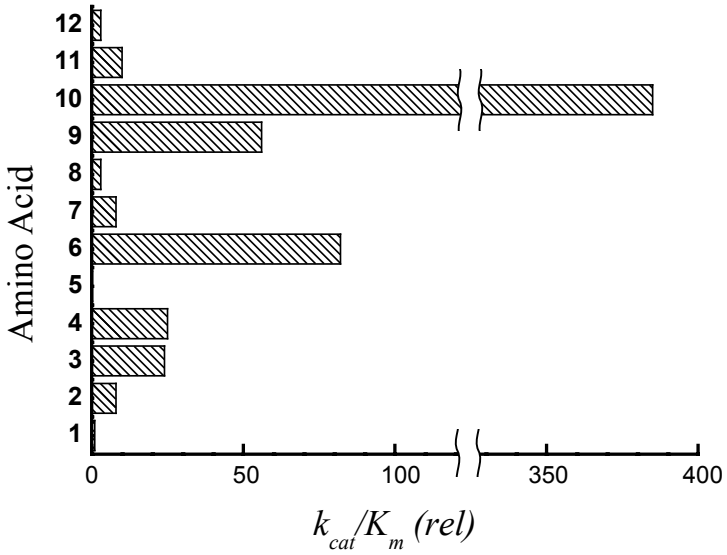

Amino Acid
12
11
10
9
8
7
6
5
4
3
2
1
0
50
100
350
400
k_{cat}/K_m (rel)

Figure 8-4

Amino acid sequence of marker protein mDHFR. Two proteolytic peptide fragments (Peptide A and Peptide B) used for MALDI and LC-MS/MS analyses are underscored. The positions with amber codon and ochre codon are labeled.

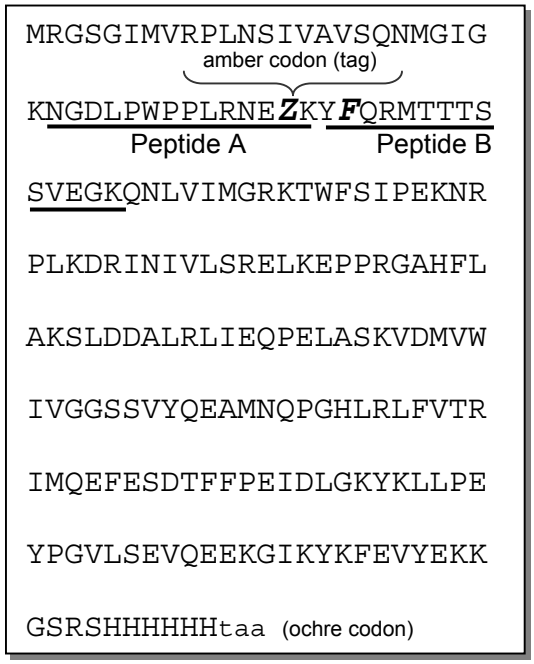

MRGSGIMVRPLNSIVAVSQNMGIG
amber codon (tag)
KNGDLPWPPLRNEZKYFQRMTTTS
Peptide A
Peptide B
SVEGKQNLVIMGRKTWFSIPEKNR
PLKDRINIVLSRELKEPPRGAHFL
AKSLDDALRLIEQPELASKVDMVW
IVGGSSVYQEAMNQPGHLRLFVTR
IMQEFESDTFFPEIDLGKYKLLPE
YPGVLSEVQEEKGIKYKFEVYEKK
GSRSHHHHHHtaa (ochre codon)

Figure 8-5

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) **8** (3 mM)/**1** (0.03); (d) **2** (3 mM)/**1** (0.03 mM). mDHFR was expressed in an phenylalanine auxotrophic strain (AF) transformed with the mutant yeast PheRS and the cognate suppressor tRNA. No tryptophan is supplemented during induction. Protein was digested by endoproteinase Lys-C. Peptide B, containing one Phe codon, remains the same for all the experimental conditions, indicating the normal Phe codons are assigned with Phe. Peptide A contains one amber codon and the residue responding to this codon is designated as Z. Amino acid for Z is assigned based on the mass units for Peptide A at different expression conditions.

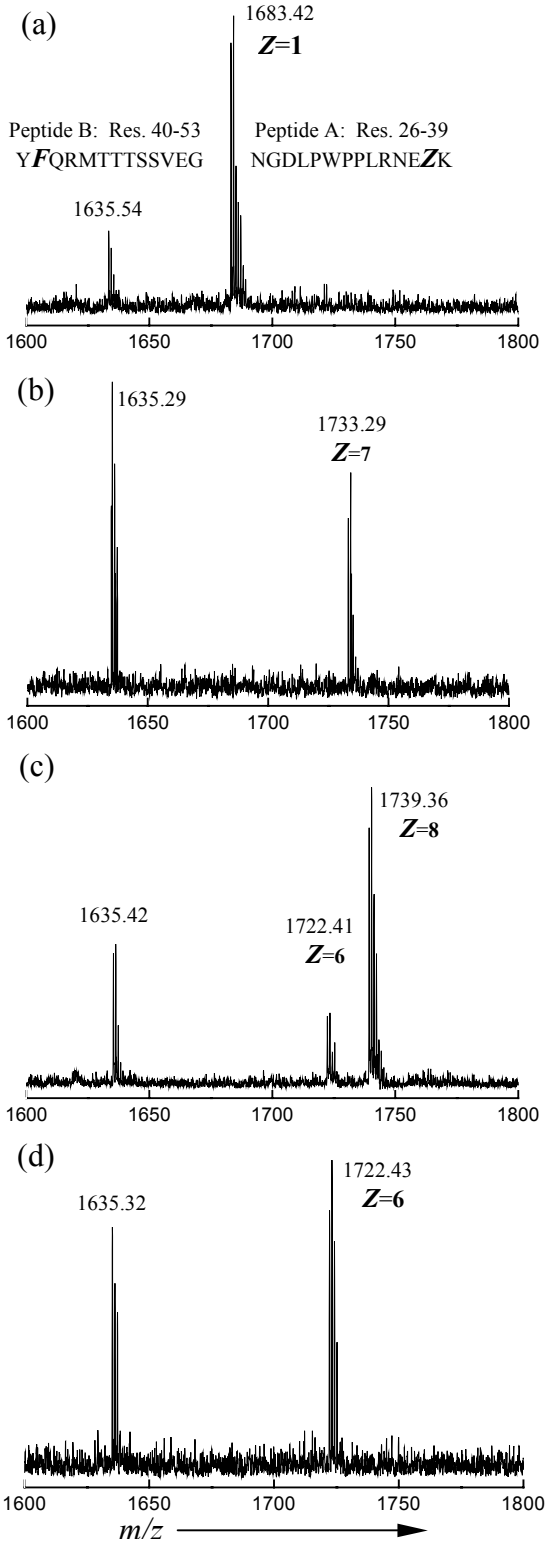
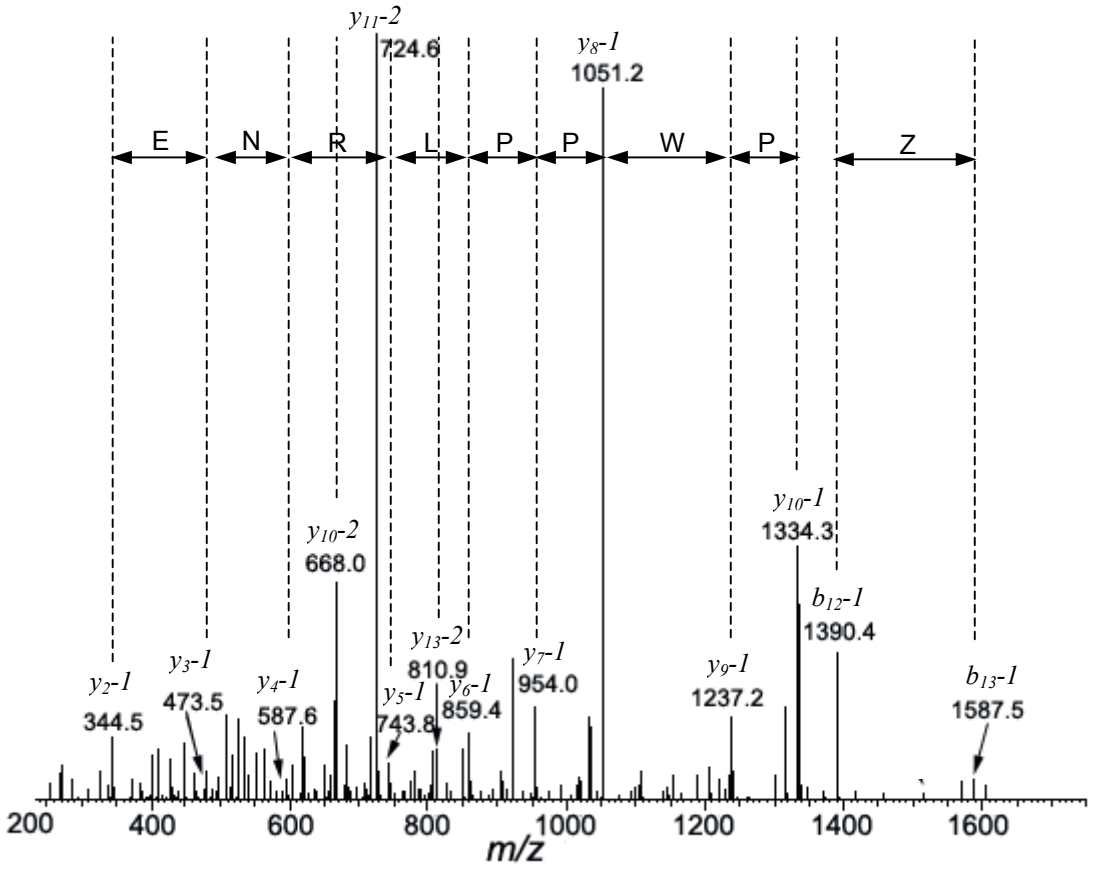

(a)
1683.42
Z=1
Peptide B: Res. 40-53
YFQRMTTTSSVEG
Peptide A: Res. 26-39
NGDLPWPPLRNEZK
1635.54
1600
1650
1700
1750
1800
(b)
1635.29
1733.29
Z=7
1600
1650
1700
1750
1800
(c)
1739.36
Z=8
1635.42
1722.41
Z=6
1600
1650
1700
1750
1800
(d)
1722.43
Z=6
1635.32
1600
1650
1700
1750
1800
m/z

Figure 8-6

The tandem mass spectrum of Peptide A (NGDLPWPPLRNE***Z***K) derived from mDHFR expressed in media supplemented with **7** (3 mM)/**1** (0.03 mM). mDHFR was expressed in an phenylalanine auxotrophic strain (AF) transformed with the mutant yeast PheRS and cognate suppressor tRNA. The 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.


y11-2
724.6
y8-1
1051.2
E
N
R
L
P
P
W
P
Z
y10-1
1334.3
y10-2
668.0
b12-1
1390.4
y13-2
810.9
y7-1
954.0
y3-1
473.5
y2-1
344.5
y4-1
587.6
y5-1
743.8
y6-1
859.4
y9-1
1237.2
b13-1
1587.5
200
400
600
800
1000
1200
1400
1600
m/z

5. References

1. Schimmel, P. (1987) *Annu. Rev. Biochem. 56*, 125-158.

2. Soll, D., and RajBhandary, U. L. (1995) in *tRNA, Structure, Biosynthesis, and Function*, ASM Press, Washington, D.C.

3. Cusack, S. (1997) *Curr. Opin. Struc. Biol. 7*, 881-889.

4. de Pouplana, L. R., and Schimmel, P. (2000) *Cell. Mol. Life Sci. 57*, 865-870.

5. Ibba, M., and Soll, D. (2000) *Annu. Rev. Biochem. 69*, 617-650.

6. Francklyn, C., Perona, J. J., Puetz, J., and Hou, Y.-M. (2002) *RNA 8*, 1363-1372.

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

8. Tamura, K., and Schimmel, P. (2002) *Nat. Biotechnol. 20*, 669-670.

9. Wang, L., and Schultz, P. G. (2002) *Chem. Commun.*, 1-11.

10. van Hest, J. C. M., and Tirrell, D. A. (2001) *Chem. Commun.*, 1897-1904.

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

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

13. Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002) *Proc. Natl. Acad. Sci. USA 99*, 19-24.

14. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew. Chem., Int. Edit. 40*, 1494-1496.

15. Kiick, K. L., Weberskirch, R., and Tirrell, D. A. (2001) *FEBS Lett. 505*, 465-465.

16. Kiick, K. L., van Hest, J. C. M., and Tirrell, D. A. (2000) *Angew. Chem., Int. Ed. 39*, 2148-2152.

17. Tang, Y., and Tirrell, D. A. (2001) *J. Am. Chem. Soc. 123*, 11089-11090.

18. Sharma, N., Furter, R., Kast, P., and Tirrell, D. A. (2000) *FEBS Lett. 467*, 37-40.

19. Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002) *ChemBioChem 3*, 235-237.

20. Datta, D., Wang, P., Carrico, I. S., Mayo, S. L., and Tirrell, D. A. (2002) *J. Am. Chem. Soc. 124*, 5652-5653.

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

22. Tang, Y., and Tirrell, D. A. (2002) *Biochemistry 41*, 10635-10645.

23. Furter, R. (1998) *Protein Sci. 7*, 419-426.

24. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) *Science 292*, 498-500.

25. Wang, L., Brock, A., and Schultz, P. G. (2002) *J. Am. Chem. Soc. 124*, 1836-1837.

26. Chin, J. W., and Schultz, P. G. (2002) *ChemBioChem 3*, 1135-1137.

27. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *Proc. Natl. Acad. Sci. USA 99*, 11020-11024.

28. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) *J. Am. Chem. Soc. 124*, 9026-9027.

29. Zhang, Z. W., Wang, L., Brock, A., and Schultz, P. G. (2002) *Angew. Chem., Int. Edit. 41*, 2840-+.

30. Wang, L., Zhang, Z. W., Brock, A., and Schultz, P. G. (2003) *Proc. Natl. Acad. Sci. USA 100*, 56-61.

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

32. Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S., and Hecht, S. M. (1984) *Biochemistry 23*, 1468-1473.

33. Bain, J. D., Glabe, C. G., Dix, T. A., and Charmberlin, A. R. (1989) *J. Am. Chem. Soc. 111*, 8013-8014.

34. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) *Science 244*, 182-188.

35. Gillmore, M. A., Steward, L. E., and Charmberlin, A. R. (1999) *Topics Curr. Chem. 202*, 77-99.

36. Dougherty, D. (2000) *Curr. Opin. Chem. Biol. 4*, 645-652.

37. Sanni, A., Walter, P., Ebel, J.-P., and Fasiolo, F. (1990) *Nucleic Acids Res. 18*, 2087-2092.

38. Reshetnikova, L., Moor, N., Lavrik, O., and Vassylyev, D. G. (1999) *J. Mol. Biol. 287*, 555-568.

39. Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., and Blanquet, S. (1983) *J. Mol. Biol. 171*, 239-261.

40. Cabius, H.-J., von der Haar, F., and Cramer, F. (1983) *Biochemistry 22*, 2331-2339.

41. Wang, P., Kwon, I., and Tirrell, D. A. (2003) *unpublished results.*

42. Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H., and Sisido, M. (1999) *J. Am. Chem. Soc. 121*, 34-40.

43. Cornish, V. W., and Schultz, P. G. (1994) *Curr. Opin. Chem. Biol. 4*, 601-607.

44. Wang, L., and Schultz, P. G. (2001) *Chem. Biol. 8*, 883-890.

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

46. Fleming, S. A. (1995) *Tetrahedron 51*, 12479-12520.

47. Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998) *Proc. Nat. Acad. Sci. USA 95*, 15613-15617.

48. Genet, J. P., and Savignac, M. J. (1999) *J. Organomet. Chem. 576*, 305-317.

49. Kolb, H. C., Finn, M. G., and Sharpless, K. B. (2001) *Angew. Chem., Int. Ed. 40*, 2004-2021.

50. Lewis, W. G., Green, L. G., Grynszpan, F., Radic, Z., Carlier, P. R., Taylor, P., Finn, M. G., and Sharpless, K. B. (2002) *Angew. Chem., Int. Ed. 41*, 1053-1057.

51. Wang, Q., Chan, T. R., Hilgraf, R., Fokin, V. V., Sharpless, K. B., and Finn, M. G. (2003) *J. Am. Chem. Soc. 125*, 3192-3193.

52. Speers, A. E., Adam, G. C., and Cravatt, B. F. (2003) *J. Am. Chem. Soc. 125*, 4686-4687.