Chapter 6

Engineering Relaxed Substrate Specificity in *E. coli* Phenylalanyl-tRNA Synthetase to Incorporate a Diverse Set of Non-natural Amino Acids

Portions of this chapter are adapted from a paper in preparation by Pin Wang, Isaac S. Carrico, Kent Kirshenbaum, Yi Tang, Deepshikha Datta, Stephen L. Mayo and David A. Tirrell

Abstract

Translational fidelity in protein synthesis is maintained by the aminoacyl-tRNA synthetases (aaRSs), which accurately attach the amino acids to the cognate tRNAs. Here we present a computational method to identify critical residues to relax substrate specificity of the *E. coli* phenylalanyl-tRNA synthetase (*e*PheRS). Together with a previously described mutant, we isolated and characterized three *e*PheRS mutants (A294G, T251G and T251G/A294G) with point mutations at positions A294 and T251 in the α-subunit of the enzyme. Their relaxed substrate specificity was extensively explored by use of the ATP-PP_i exchange assays, yielding a broad activation profile toward many aromatic non-natural amino acids. Engineered *E. coli* hosts with over-expression of the mutants enabled the re-assignment of Phe codons with non-natural amino acids including *p*-acetylphenyl-, *p*-iodophenyl-, *p*-cyanophenyl-, *p*-azidophenyl-, *p*-nitrophenyl-, *p*-aminophenyl- and 3-(2-naphthyl)-alanine. These results further demonstrate that engineering the synthetic active sites of aminoacyl-tRNA synthetases can be a powerful route to *in vivo* incorporation of novel amino acid side chains into recombinant proteins.

1. Introduction

Protein biosynthesis is characterized by high fidelity in the translation of nucleic acid sequences into protein sequences (*1*). This requires a class of remarkable enzymes, the aminoacyl-tRNA synthetases (aaRSs), to attach chemically diversified amino acids into their cognate tRNAs, which are subsequently shuttled to the ribosome and site-specifically added to the growing polypeptide chain (*2-4*). aaRSs catalyze the formation of aminoacyl-tRNA by two-step reactions (Scheme1): cognate amino acids react with ATP to form aminoacyl-adenylates (step 1); subsequently these activated forms of the amino acids are attached to their cognate tRNAs by esterification (step 2). These catalytic reactions depend upon the ability of the aaRSs to recognize amino acids, ATP and cognate tRNAs. The substrate specificity of these enzymes is essential to ensure the accurate transformation of genetic information into proteins (*5*).

The error rate for recognition of tRNAs by aaRSs is extremely low (10^{-6} or lower) (*5*) because of the large tRNA/aaRS contact area. Descrimination of substrate amino acids is more challenging because they are much smaller and have less structural variability. On the basis of mutually exclusive sequence motifs at catalytic domains, aaRSs can be divided into two classes (Class I and Class II) (*6*). Class I enzymes have a representative Rossmann-fold catalytic domain, approach their tRNA substrates from the minor groove of the acceptor stem, and aminoacylate the 2'OH of the terminal ribose. In contrast, Class II synthetases have a catalytic domain consisting of an antiparallel β-fold connected by α-helices, access their tRNAs from the major groove of the acceptor stem and attach amino acids to the 3'OH. Extensive structural studies have been conducted on the aaRSs and most of their structures have been determined, providing considerable

insights into their amino acid recognition and activation (*4*). Class I synthetases require the binding energy gained from enzyme-ATP interaction to stabilize a transition state to accommodate and activate the cognate amino acid, while Class II enzymes constitute a rigid template at the active site so that the amino acid and ATP can bind with an optimal orientation to facilitate its in-line nucleophilic displacement reaction (*4*).

Among Class II aaRSs, phenylalanyl-tRNA synthetase (PheRS) is unique in that it attaches its cognate amino acid (Phe, **1**) to the 2'OH of the terminal ribose of the $tRNA^{Phe}$ (*7, 8*); further it is an $\alpha_2\beta_2$ hetero-tetrameric enzyme, rather than an α_2 homo-dimer as most of this class of enzymes (*9, 10*). The crystal structure of PheRS from *Thermus thermophilus* (*t*PheRS) reveals that the α-subunit is the catalytic unit and the major function for β-subunit is recognition and binding of $tRNA^{Phe}$ (Figure 6-1) (*9, 10*). As is characteristic of Class II enzymes, the active site of PheRS is relatively rigid, as revealed by the similar conformations of the ligand-free, Phe-bound and Phe-adenylate analog bound structures of *t*PheRS (*10*). Recognition of Phe by PheRS invovles hydrogen-bonding interactions with the polar ammonium and carboxylate moieties and multiple van der Waals interactions with hydrophobic side chains (Figure 6-1B). The phenyl ring of substrate Phe is oriented between the hydrophobic side chains of F258 and F260 in the α-subunit, with additional back wall constraints constituted by the side-chains of V261 and A314 in the binding pocket (Figure 6-1B) (*10*). Sequence alignment of PheRSs from different organisms shows much conservation on active site residues (Figure 6-1C), in contrast other domains within these enzymes demonstrate considerable diversity.

One of the goals of our laboratory is to enlarge the available amino acid repertoire to expand our capabilities to design of biomacromolecules with programmable chemical

and physical properties (*11-22*). Many chemically distinct amino acids have been introduced into proteins through ribosomal biosynthesis *in vivo*, including alkenes (*12, 13, 15*), alkynes (*12, 13, 16*), cyclobutenes, aryl halides (*11, 20*) and other functional groups (*19, 21*). We have shown that introduction of fluorinated side chains can stabilize coiled-coil proteins to an extent that would be very difficult to achieve by only canonical amino acids. Site-specific modification of proteins can be facilitated by successful incorporation of alkyl azide (*19*) and aromatic ketone functions (*21*). Our method of multi-site replacement of novel amino acid side chains can address protein and material design issues such as stability (*16, 18*) and surface properties (*23*). This method requires the re-assignment of genetic codons to new amino acids, which can be accomplished by using auxotrophic strains and by depletion of intracellular pool of the competing natural amino acids prior to induction of protein expression. Success of these experiments depends on the promiscuity of the aminoacyl-tRNA synthetases. Codon re-assignment can be enhanced by over-expression of aaRS in the host (*13, 18*). When the non-canonical amino acids are not recognized by the wild-type aaRS, re-design of synthetase activity is required. Both computational design (*21, 24*) and combinatorial approaches (*25, 26*) could be powerful tools to design amino acid sequences of binding pockets for recognition of novel substrates.

In this chapter, we describe rationally designed variants of *E. coli* PheRS (*e*PheRS) with relaxed substrate specificity to allow incorporation of a diverse set of non-natural amino acids (**2-9**). One newly identified mutant (T251G) and two previously reported mutants (A294G and T251G/A294G) of *e*PheRS were subjected to extensive *in vitro* and *in vivo* studies. We carried out studies of amino acid activation kinetics of the

wild-type enzyme and all three mutants and found that these mutant synthetases can activate a number of aromatic amino acids (**1-10**). Our *in vitro* aminoacylation experiments showed that the T251G and T251G/A294G enzymes have impaired activity with respect to aminoacylation by Phe but can synthesize tryptophanyl-$tRNA^{Phe}$. When these mutants were over-expressed in an *E. coli* host, many of the analogs that displayed activity *in vitro* were incorporated into recombinant proteins *in vivo*, except analog **10**, which is believed to be due to its incapability of being accepted by A site of ribosome.

2. Materials and Methods

2.1. Materials

Amino acid **1** was purchased from Sigma (St. Louis, MO). **2** was obtained from RSP Amino Acid Analogues (Shirley, MA). **3-13** were purchased from Chem-Impex (Wood Dale, IL). [^{3}H]-labeled amino acids were obtained from Amersham Pharmacia Biotech (Piscataway, NJ). [^{32}P]-labeled sodium pyrophosphate was obtained from NEN Life Science (Boston, MA).

2.2. Computational Design of Mutant Forms of PheRS

We used a protein design algorithm (*27, 28*), ORBIT, to predict the optimal amino acid sequences of the binding pocket of PheRS from *Thermus thermophilus* (*t*PheRS; PDB code: 1B70) (*10*) for binding to the amino acid analogs. Selection of amino acids for optimization is carried out using a very efficient search algorithm that relies on a discrete set of allowed conformations for each side chain and empirical potential energy

functions that are used to calculate pairwise interactions between side chain and backbone and between side chains.

In our design calculations, optimization was performed by varying the torsional angles of the analogs and side chains lining the pocket simultaneously. This required generating rotamer libraries for the analogs, since they are not included in the standard rotamer libraries. For all the natural amino acids, the possible χ_1 and χ_2 angles are derived from database analysis. As this is not feasible in the case of amino acid analogs, the closest approximation for χ_1 and χ_2 angles for Phe analogs were taken to be the same as those for Phe. In addition, this could provide us a better chance to select for conformations of analogs that were as close as possible to the orientation of Phe in the binding pocket, so that they can be activated through reaction with ATP in the pocket. Accordingly we generated a backbone independent rotamer library for analog **2** (Figure 6-2). The torsional angles of substrate Phe complexed with *t*PheRS in the crystal structure (χ_1: -101^o; χ_2 : -104^o) were also included in the new rotamer libraries for both Phe and analog. Charges were assigned only to the heavy atoms of the analog **2** to be consistent with the way that charges for the natural amino acids are represented in ORBIT.

Since the residues in the pocket are buried in the protein structure, we employed force field parameters similar to those utilized in previous protein core design algorithm. The design algorithm uses energy terms based on a force field that includes van der Waals interactions, electrostatic interactions, hydrogen bonding and solvation effects (*29*). Calculations were performed by anchoring the substrate to be **2** and varying 11 residues within 6 Å of the substrate in the binding pocket of *t*PheRS (L137, V184, M187,

L222, F258, F260, V261, V286, V290, V294, A314). Positions 137, 184, 258, 260, 261, 286, 290, 294 and 314 were allowed to be occupied by any of the 20 natural amino acids except proline, methionine and cysteine. Methionine was allowed at position 187 since it is the wild-type residue at this position; only hydrophobic amino acids were allowed at position 222. Most of these positions are buried in the core and a number of them pack against Phe in the crystal structure. The anchor residues (E128, E130, W149, H178, S180, Q183 and R204) were fixed both in identity and conformation in all the calculations. These residues make very important electrostatic interactions with the substrate and we reason that this kind of interaction is probably equally critical for analog **2**. From the crystal structure it seems that the anchor residues hold the Phe zwitterions in a way that the carbonyl group of the zwitterions are close to the ATP binding site. This proximity could be important for reactions to form the aminoadenylate, the first step in aminoacylation. Since this reaction is required for amino acid incorporation into proteins *in vivo*, it seems important to make sure that the zwitterions of **2** are also anchored the same way as the natural substrate.

2.3. Plasmid Construction for Synthetase Expression

The *e*PheRS gene was cloned directly from *E. coli* genomic DNA with flanking primers encoded the restriction sites *Sac*I and *Hind*III (primer 1: 5'-CAC CAC TGA CAC AAT GAG CTC AAC CAT GTC ACA TCT CG-3'; primer 2: 5'-CAT ATG GCT AGC AAG CTT CAT AGG TTC AAT CCC-3'). The resulting 3500 base-pair DNA fragment was gel-purified, digested with *Sac*I and *Hind*III, and ligated into the expression plasmid pQE30 (Qiagen) to yield pQE-pheST, which encodes both the α and β subunits

of wild-type *E. coli* PheRS. Four-primer mutagenesis method was employed to generate desired mutant forms of *e*PheRS. Briefly, a pair of complementary oligos, designated as primer 3 and primer 4, was designed to carry a specific mutantation at position 294 or 251 of the α subunit of *e*PheRS. In one reaction primer 1 and primer 4 was used to yield the first DNA fragment of *e*PheRS gene. In another reaction primer 3 and primer 2 was used to yield the second DNA fragment of *e*PheRS gene. The two fragments were then mixed for further amplification in the presence of primer 1 and primer 2 to afford the entire *e*PheRS gene, which was cloned into pQE30 to yield pQE30-A294G, pQE30-T251G or pQE30-T251G/A294G, encoding the A294G, T251G and T251G/A294G mutant forms of *e*PheRS, respectively. The cloned enzymes contained the N-terminal leader sequence MRGSHHHHHHTDPHASST for purification. Platinum *Pfx* DNA polymerase (Invitrogen) was used for the PCR reactions. The integrity of each cloned gene was confirmed by DNA sequencing.

2.4. Synthetase Expression and Purification

Plasmids pQE30-pheST, pQE30-A294G, pQE30-T251G and pQE30-T251G/A294G were independently transferred to *E. coli* RecA$^-$ strain XL-1 blue (Stratagene) to minimize the possible chromosomal recombination with the endogenous *e*PheRS gene. Protein expression was induced at OD_{600}=0.6 with 1 mM IPTG. After three hours, the cells were harvested. The enzyme was purified using Ni-NTA agarose resin under native conditions according to the manufacturer's instructions (Qiagen). The eluted protein solutions contained 250 mM of imidazole, which was removed on an ion-exchange column eluted with buffer A (50 mM Tris-HCl, 1 mM DTT). Purified enzymes

were stored in buffer A with 50% glycerol at –80 ^{o}C. The concentration of the purified enzyme was determined by absorbance at 280 nm under denaturing conditions.

2.5. Amino Acid Activation Assays

Assays were performed at ambient temperature by measuring the kinetics of the amino acid dependent ATP-pyrophosphate (PP_i) exchange reaction (*30*). The reaction was conducted in 200 µl of reaction buffer (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 upon the activity of synthetases toward the substrate, the enzyme concentration varied from 10 nM to 100 nM with substrate concentrations of 10 µM to 5 mM. Aliquots were taken at various times and quenched into a 500 µL solution containing 200 mM PP_i, 7% w/v $HClO_4$ and 3% w/v activated charcoal. The charcoal was spun down and washed twice with a solution containing 10 mM $NaPP_i$ and 0.5% perchloric acid. The [^{32}P]-labeled ATP absorbed by the charcoal was counted via liquid scintillation methods. Kinetic constants (K_m and k_{cat}) were extracted by a nonlinear regression fit of the data to a Michaelis Menten model.

2.6. Aminoacylation Assays

Aminoacylation assays were performed at 37 ^{o}C in the buffer (100 µL) containing 30 mM HEPES (pH=7.4), 10 mM $MgCl_2$, 1mM DTT and 2 mM ATP (*22, 31-33*). Purified *E. coli* total tRNA (Roche Biochemical) was used in the assay at a final concentration of 12 mg/mL ($tRNA^{Phe}$ concentration approximately 7.5 µM). tRNAs were annealed before use by heating to 85 ^{o}C for 4 min in annealing buffer (60 mM Tris,

pH=7.8, 2 mM $MgCl_2$) followed by slow cooling to room temperature. For aminoacylation of Phe (**1**), 20 μM [^{3}H]-Phe (2000 dpm/pmol) and 10 nM *e*PheRS variants were used; for aminoacylation of Trp (**8**), 200 μM [^{3}H]-Trp (4000 dpm/pmol) and 40 nM enzymes were used. 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 and counted via liquid scintillation methods.

2.7. Plasmid Construction for *In Vivo* Incorporation Assays

The *E. coli pheS* gene for the α subunit of *e*PheRS was amplified by PCR from plasmid pQE30-pheST. Amplified *pheS* was subjected to site-directed mutagenesis to create the coding sequences for the intended A294G, T251G and T251G/A294G mutants. To allow constitutive expression of the α subunit of the synthetase, a linker encoding a *tac* promoter with an abolished *lac* repressor binding site was prepared with terminal *Nhe*I restriction sites and internal *Nco*I and *Hind*III sites. The linker sequence is 5'CTA GC AGT TGA CAA TTA ATC ATC GGC TCG TAT AAT GGA TCG AAT TGT GAG CGG AAT CGA TTT TCA CAC AGG AAA CAG ACC **ATG** GAT CTT CGT CGC CAT CCT CGG GTC GAC GTC TGT TTG CAA GCT TG-3' (the –35 and –10 sequences are underlined and the start codon is in bold). The linker was cloned into the *Nhe*I site of vector pET5a (Novagen) to yield pET5a-tac. PCR amplified fragments encoding the A294G, T251G and T251G/A294G mutants were independently cloned into pET5a-tac at the *Nco*I and *Hind*III sites. Genes for A294G, T251G and T251G/A294G (now equipped with the constitutive *tac* promoter) were cut out at the flanking *Nhe*I sites,

and inserted into expression plasmid pQE15 (Qiagen) to form pQE15-A294G, pQE15-T251G and pQE15-T251G/A294G, respectively. Expression plasmids pQE15-A294G, pQE15-T251G and pQE15-T251G/A294G encode the test protein murine dihydrofolate reductase (mDHFR) under control of a bacteriophage T5 promoter.

2.8. Analog Incorporation Assays *In Vivo*

A phenylalanine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AF (*HsdS gal* (*λc*I*ts*857 *ind* 1 *S*am7 *nin*5 *lac*UV5-T7 gene *1*) *pheA*) and constructed in our laboratory, was used as the expression host (*11, 20*). The AF strain was transformed with repressor plasmid pLysS-IQ and with pQE15-A294G, pQE15-T251G or pQE15-T251G/A294G to afford expression strains AF-IQ[pQE15-A294G], AF-IQ[pQE15-T251G] or AF-IQ[pQE15-T251G/A294G], respectively.

Small scale (10 ml) cultures were used to investigate the *in vivo* incorporation of amino acid analogs **2-13**. M9 minimal medium (50 mL) supplemented with 0.2 % glucose, 1 mg/L thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 19 amino acids (at 20 mg/L), antibiotics (ampicillin 200 mg/L, chloramphenicol 35 mg/L) and phenylalanine (at 20 mg/L) was inoculated with 1 mL of an overnight culture of the expression strain. When the optical density at 600 nm reached 0.8-1.0, a medium shift was performed. Cells were sedimented by centrifugation for 15 min at 3100g at 4 ^{o}C, the supernatant was removed and the cell pellets were washed twice with 0.9% NaCl. Cells were resuspended in supplemented M9 medium containing either: (a) 250 mg/L analog, (b) 20 mg/L Phe (**1**) (positive control), (c) no Phe or analog (negative control). Protein expression was induced 10 min after the medium shift by addition of isopropyl-β-D-thiogalactoside

(IPTG) to a final concentration of 1 mM. Cells were cultured for 4 hours post-induction and protein expression was monitored by SDS polyacrylamide gel electrophoresis (PAGE, 12 %).

2.9. Target Protein Composition Analysis

Target protein mDHFR as expressed in this work contains an N-terminal hexahistidine sequence, which was utilized to purify the protein by nickel affinity chromatography with stepwise pH gradient elution under denaturing conditions according to the recommendations of the supplier (Qiagen). Purified protein in 10 μL of elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH=4.5) was mixed with 90 μl 75 mM NH_4OAc, to which 2 μL of modified trypsin (Promega, 0.2 μg/μL) was added. The solution was allowed to digest overnight at room temperature. The reaction was quenched by addition of trifluoroacetic acid to pH < 4.0. The digest was subjected to sample clean-up by using a $ZipTip_{C18}$, which provided 2 μL of purified sample solution. A 10 μL volume of the MALDI matrix (α-cyano-β-hydroxycinnamic acid, 10 mg/ml in 50% CH_3CN) was added, and 0.5 μL of the resulting solution was spotted directly onto the sample plate. Samples were analyzed in the linear mode on an Applied Biosystems Voyager DE Pro MALDI-TOF mass spectrometer.

3. Results and Discussion

3.1. Computational design

Although the crystal structure of *E. coli* PheRS (*e*PheRS) is not available, the structure of *Thermus thermophilus* PheRS (*t*PheRS) complexed with Phe has been solved

to 2.7 Å resolution (Figure 6-1, PDB code: 1B70). Sequence alignment showed 43% identity between *e*PheRS and *t*PheRS. Accurate description of the binding pocket for Phe is essential for our computational design approach since the calculation is based on the protein backbone from the crystal structure. As shown in Figure 6-1, the aromatic ring of substrate Phe is buried in the hydrophobic pocket while the carboxylate and ammonium groups make extensive electrostatic contacts with the charged and polar resides at the pocket. Our intention is to design the binding pocket for the analogs so that they could bind to *t*PheRS in the same orientation as Phe because this orientation is vital for the adenylation reaction. The analog we chose for this work is *p*-acetylphenylalanine (**2**), an amino acid with the chemically versatile side-chain ketone functionality.

In the first calculation we allowed all the rotamers of **2** that were generated in the rotamer library. The rotamers selected from this calculation were not buried in the binding pocket and most of them were found to be solvent exposed. In the second calculation we only included those **2** rotamers that would be able to pack into the binding pocket. These are the rotamers with all possible combinations of χ_1 of -101^o ($\pm 20^o$) and χ_2 of -104^o ($\pm 20^o$) in increments of 5^o. The structure generated in this calculation has the aromatic ring of **2** buried in the pocket and is almost completely superimposable with Phe in the active site of *t*PheRS. A control calculation was also performed where the substrate is Phe. From Table 6-1, we compared the sequences predicted for **1** and **2** and observed that most significant difference between the two sequences was the prediction of two important cavity-forming mutations for the needed enzyme to accommodate **2**: V261 (T251 in *E. coli*) to G and A314 (A294 in *E. coli*) to G.

Not only are these predications in a good agreement with the report of Reshetnikova and coworkers that A314 and V261 form the back wall of the amino acid binding pocket and prevent amino acids larger than Phe (e.g. tyrosine) to bind (Figure 6-1) (*10*), but also with our *in vivo* experiments, which indicated that over-expression of the A294G mutant form of the α subunit *e*PheRS could incorporate of *p*-bromo-, *p*-iodo- (**3**), *p*-ethynyl-, *p*-cyano- (**4**) and *p*-azidophenylalanines (**5**) (*11, 20*). Efficient incorporation of **2** was realized in *E. coli* armed with the double mutant (V251G/A314G)PheRS. Sequence alignment of PheRS from different organisms showed that the most corresponding residue for V261 from *T. thermophilus* is T251 and residue for A314 is A294, indicating conservation and relevance of these two residues for substrate binding (Figure 6-1). Coupled with results from computational design and previous experimental results, we decided to evaluate three mutant forms of *e*PheRS (A294G, T251G and T251G/A294G) more systematically and thoroughly *in vivo* and *in vitro* with respect to activating a broader pool of amino acid analogs (**2**-**13**).

3.2. Expression and Purification of ePheRS Variants

The wild-type *e*PheRS gene was amplified from *E. coli* genomic DNA and used as a template to generate mutants A294G, T251G and T251G/A294G through four-primer PCR. The PCR product (approximately 3500 bp) of gene fragment encodes the α and β subunits of *e*PheRS and a 14 bp of an intercistronic region (*34*). Enzymes were expressed in *E. coli* strain XL-1 blue, which harbors the plasmid of individual pQE30 derivatives containing variants of *e*PheRS gene in frame with a N-terminal $(His)_6$ tag. Ni-NTA affinity chromatography showed nearly identical levels of expression of the α

and β subunits by SDS-PAGE analysis, indicating the high efficiency of the intercistronic sequence. Significantly slower growth was observed for the strain bearing the vector encoding the T251/A294G mutant. The reduced growth rate might be associated with mischarging of one or more non-cognate amino acids by T251/A294G.

E. coli strain XL-1 blue contains an endogenous copy of the *e*PheRS gene. According to SDS-PAGE analysis, our mutant proteins (A294G, T251G and T251/A294G) were over-expressed at the level approximately 50-fold higher than the chromosomally encoded *e*PheRS. In light of fact that *e*PheRS is a hetero-tetramer ($\alpha_2\beta_2$ type) (*10*), we expected that the mutant enzymes could have a low level of contamination by the endogenous α subunit of *e*PheRS by our His-tag purification scheme with His-tag at the N-terminal of α subunit of the enzyme. One could estimate that less than 2% of hetero-tetramer could have one mutant α subunit and one wild-type α subunit. The contaminant enzyme has wild-type activity toward Phe (**1**) but exhibits no activity toward any of the amino acid analogs we examined, so that the observation of amino acid analog activation can only be attributed to the desired mutant forms of the enzymes.

3.3. Activation of Analogs by Variant Enzymes *In Vitro*

We examined the aminoacyl adenylate synthesis by monitoring the amino acid dependent ATP-PP_i exchange assay in the presence of either the wild type PheRS or one of the described mutants. Kinetic parameters for activation of canonical amino acids (**1** and **8**) and non-canonical amino acids (**2-7**, **9-10**) were shown in Table 6-2. K_m value obtained from our measurement for **1** by wild-type *e*PheRS is comparable with previously reported value (*35*), although k_{cat} values is lower; this is presumably due to the

different buffer conditions and different methods of measuring concentrations of the enzyme. Compared to wild-type enzyme, A294G, T251G and T251G/A294G exhibit higher K_m and lower k_{cat} toward **1**, and the specificity constant (k_{cat}/K_m) is decreased by factors of 251, 16 and 806, respectively. The T251G mutation has minimal effect on the ability of PheRS to recognize **1**. Analog **2** was activated by both T251G and T251G/A294G and our previous experiments showed that **2** could be introduced into proteins *in vivo* in an *E. coli* host outfitted with the T251G/A294G form of *e*PheRS. All three mutant enzymes recognize *para*-substituted Phe analogs **3**, **4**, **5** and **7**, with different reactivity. T251G displayed extremely high reactivity toward **5** (k_{cat}/K_m 3-times faster than wild-type toward **1**). *para*-Nitro substituted Phe (**6**) was activated by both T251G and T251G/A294G enzymes with similar k_{cat}/K_m values. When we measured the substrate properties of even bulkier amino acids such as **8**, **9** and **10**, we found that only T251G and T251G/A294G could activate these analogs with k_{cat}/K_m ranging from 9 to 176-times poorer than wild-type *e*PheRS toward **1**. Both mutants exhibit striking activities to the canonical amino acid **9**; only a single mutation can pose significant threats to the fidelity of *e*PheRS. Under our assay conditions, we did not observe above-background activation of **11-13** or tyrosine by any of the syntetases.

A general observation from the activation assay is that T251G tends to show higher activity than A294G (higher k_{cat}/K_m) for almost all the analogs we measured. From the crystal structure, V261, equivalent to T251 in *E. coli*, is located in a loop region between two β-strands (Figure 6-3) (*10*), and A314, equivalent to A294 in *E. coli*, is positioned in the middle of one β-strand (*10*). Because of its placement the T251G mutation would be expected to generate more flexibility so that it might be more effective

in opening up space to allow binding of bulkier substrates. In addition, A294 is situated at the third characteristic motif of the class II aaRSs, which provides critical residues for substrate binding; other than offering more room, mutation of this residue might jeopardize the ability of the active site to recognize amino acids. This can be manifested by the k_{cat}/K_m values of Phe (Table 6-2); Phe is 16-times poorer a substrate for A294G than for T251G.

3.4. Aminoacylation by Variant Enzymes *In Vitro*

We investigated the aminoacylation activities of the purified wild-type and the mutant forms of *e*PheRS with canonical amino acids **1** and **8** by standard TCA precipitation assays. As expected, the wild-type enzyme showed the fastest aminoacylation rate of **1**. The kinetics of attaching **1** to $tRNA^{Phe}$ by the mutants are rather interesting; the amino acid activation rate (k_{cat}/K_m) for **1** by T251G was 16-times faster than that by A294G (Table 6-2), but T251G displayed only a slightly higher rate of aminoacylation than A294G (Figure 6-4A). Our assay conditions (10 nM *e*PheRS, 20 μM Phe), no significant aminoacylation reactions were observed by T251G/A294G. For reaction of synthesis of Trp-$tRNA^{Phe}$, we used higher enzyme concentration (40 nM) and higher concentration of **8** (200 μM) to facilitate the detections. As shown in Figure 6-4B, both T251G and T251G/A294G can synthesize Trp-$tRNA^{Phe}$; T251G exhibited slightly higher rate of reaction than T251G/A294G. Since both wild-type and A294G could not activate **8**, it is not surprising to observe no aminoacylation of **8** by these enzymes.

We noted that sometimes mutants performed poorly during amino acid activation, but showed similar abilities to make aminoacylated tRNA. Aminoacylation relies on the

abilities of synthetase to bind the amino acid, ATP, and subsequently the adenylate. Greater activation only indicates that the enzyme is able to catalyze the formation of amino acid adenylate faster, but does not imply that it can hold adenylate better for next step reaction to form aminoacylated tRNA. T251G mutation can synthesize phenylalanyl-adenylate faster, as manifested by k_{cat}/K_m of activation, indicating that it might have better binding of Phe than A294G. However, this mutation might not facilitate holding the adenylate for subsequent aminoacylation reactions. It seems that *e*PheRS has to balance the abilities of amino acid activation and tRNA attachment to yield a total rate of aminoacylation. Similar observations have been reported before (*35-37*).

3.5. *In Vivo* Effects of Mutant Synthetases

Assays for determination of translational activity of amino acid analogs (**2-13**) by mutant synthetases *in vivo* were performed. Our previous work has demonstrated that auxotrophic *E. coli* strain outfitted with the A294G α subunit mutant *e*PheRS could facilitate incorporation of *p*-bromo-, *p*-iodo- (**3**), *p*-ethynyl-, *p*-cyano- (**4**), *p*-azido- (**5**) and *p*-acetyl-phenylalanines (**2**) into protein in the place of **1**. It should be noted that most of these experiments were carried out with an expression plasmid encoding A294G under inducible T7 promoter. In order to make a consistent comparison, we repeated the *in vivo* incorporation studies on **2-5** using our new constructs. The synthetase expression plasmid pQE30-pheST was the template for PCR amplification and site-directed mutagenesis to generate genes for the A294G, T251G and A294G/T251G α subunits. A linker was designed to contain a constitutive *tac* promoter with an abolished *lac* operator

and added upstream of the start codon of these genes. This promoter has been successfully used to constitutively express PheRS from *S. cerevisiae* in an *E. coli* host (*38*). The entire expression cassettes of the A294G, T251G and A294G/T251G α subunits were then individually inserted into expression plasmid pQE15 to yield pQE15-A294G, pQE15-T251 and pQE15-A294/T251 respectively. These three plasmids all carry a gene encoding a 209-residue murine dihydrofolate reductase (mDHFR) protein (Figure 6-5); the protein contains six his-tag residues and 9 Phe residues as the possible sites of substitution by analogs. Transformation of *E. coli* phenylalanine auxotrophic strain AF with the repressor plasmid pLys-IQ and individual expression plasmids pQE15-A294G, pQE15-T251 and pQE15-A294/T251 afforded bacterial host strains AF-IQ[pQE15-A294G], AF-IQ[pQE15-T251G] and AF-IQ[pQE15-T251G/A294G] respectively. AF-IQ contains an endogenous copy of wild-type *e*PheRS; because α subunit of mutant synthetases has multiple copies and expresses strongly with *tac* promoter, the active *e*PheRS in the host would be the predominately mutant forms. Strain AF-IQ[pQE15-T251G/A294G] exhibited a slow growth phenotype.

Our *in vivo* protein expression assays require a medium-shift procedure. Normally before protein induction the bacterial cells were grown in M9 medium supplemented with all 20 natural amino acids; cells should be making mutant synthetase during this growth phase. Upon induction, cells were transferred into a new medium supplemented with only 19 natural amino acids and the analog of interest. After 3-hour protein expression, the levels of the accumulated target protein were examined by SDS-PAGE analysis. The results showed that mDHFR was expressed, at some level, in all the samples (data not shown), even in cultures where Phe was nominally depleted during

protein expression. Presumably even in the auxotrophic stain under conditions of starvation in Phe, Phe is still available in the cell from protein degradation or incomplete depletion of intracellular Phe. Nevertheless, visualization of mDHFR by SDS-PAGE demonstrated variable levels of expression dependent on which analogs (**2-13**) were present in the medium. Confirmation of the possible substitution of analogs required further characterization of target protein.

3.6. Protein Tryptic Peptide Analysis

All the cell-lysate samples were subjected to nickel-affinity purification and purified mDHFRs were tryptically digested. As a result of our work-up protocols two peptide fragments, each containing a single Phe site, consistently appeared in our MALDI mass spectra as shown in Figure 6-5. Analysis of these two peptides gave consistent results, but for clarity we chose to focus on fragment 2 (N-ELKEPPRGAH**F**LAK-C); the mass for the unaltered peptide is 1592.89 Da. The representative tryptic mass spectra for situations where the media contain either no Phe or analog **2, 5, 9** and **10** are shown in Figure 6-6; they include all three mutant synthetases. Table 6-3 lists MALDI-TOF data for fragment 2 derived from mDHFR expressed in media supplemented with **1**, no analog or analog **2-13**. If the Phe is partially substituted by an analog, an additional peak with mass difference corresponding to the difference between Phe and the analog appears in the spectrum. For example, after over-expression of T251G in the host, substitution of Phe with **2** results an additional peak with mass increment of 42 mass units (Figure 6-6e); this is consistent with our *in vitro* activation assays, which show that T251G can activate **2** with k_{cat}/K_m 178-times poorer than wild-type PheRS toward **1**. In the media without

Phe and any analogs, we observed expression of mDHFR from SDS-PAGE analysis; mass spectra indicate that no mischarges occur at Phe sites in A294G (Figure 6-6a, Table 6-3), but in the presence of mutants T251G and A294G/T251G an increment of 39 mass units implies that Phe sites are partially substituted by amino acid **8** (Figure 6-6b to 6-6c, Table 6-3). Similar patterns are observed when we supplemented media with **8** (250 mg/L, Table 6-3). This result supports our hypothesis that the slow growth phenotype appearing in strains bearing the T251G/A294G mutant results from toxicity imposed by mischarging non-cognate amino acid **8** into Phe codons in many cellular proteins. According to our *in vitro* measurements, **8** is activated 32-times more rapidly than Phe by T251/A294G (Table 6-2); Figure 6-6c indicates that even with deprivation of Phe, **8** still cannot completely replace Phe at all Phe codons; we attribute this largely to the endogenous wild-type copy of *e*PheRS, which charges Phe 25-time faster than T251G/A294G toward **8**; additionally there are other factors along the translational pathway which might favor cognate amino acid Phe.

Consistent with our previous *in vivo* results, analogs **3-5** are not only activated by A294G *in vitro*, they are able to infiltrate into Phe codons *in vivo* in the host carrying the A294G mutant of *e*PheRS (Figure 6-6g, Table 6-3). Since analog **4** is activated relatively slowly (Table 6-2) by T251G and T251G/A294G, mass spectra manifest the partial substitution of **8** in place of Phe in addition to substitution of **4** (Table 6-3). From Table 6-2, we observed that both T251G and T251G/A294G could activate **3** and **5** relatively rapidly; other than fractional replacement of **3** and **5** at Phe sites, we do not observe co-substitution of **8** (Figure 6-6h to 6-6i, Table 6-3). Tryptic fragment 2 containing **6** or **7** both shows mass distributions characteristic of incorporation of the analog at Phe sites

with co-replacement of **8**, as discussed above for fragment containing **4**; despite the *in vitro* data showing that **6** is a relatively good substrate for both T251G and T251G/A294G (Table 6-2) and the high concentration of this analog in the growth media, we still find the co-substitution of **8**, indicating that either poor transportation of this analog to cytoplasm results in low cellular concentration of **6**, or translational apparatus somehow does not favor this analog to support protein biosynthesis.

In view of activation, aminoacylation and *in vivo* translational activities of **8** by mutant synthetases T251G and T251G/A294G, we decided to test properties of analogs (**9-10**) with naphthyl side chains to further explore the capacity of our newly designed synthetase. Compared to the indole ring, the naphthyl group is more hydrophobic and bulkier (side-chain volume: V(**8**)=152.42 $Å^3$; V(**9**)=171.71 $Å^3$; V(**10**)=171.13 $Å^3$; Figure 6-2). T251G displayed higher activity toward **9** than **10**; T251G and T251G/A294G showed similar activity toward **9**. As expected, the analog **9** incorporated into fragment 2 can be clearly identified (with an additional mass peak; mass shift of +50 Da; Figure 6-6k to 6-6l; Table 6-3) in response to both mutants; consistent with relative *in vitro* activities of **8** and **9** (**9** is a slightly better substrate for both T251G and T251G/A294G) and considering the high concentration of **9** in our experiment conditions, it is not surprising to observe no co-incorporation of **8** at Phe sites. After careful examination of all possible peptide fragments, we were not able to detect the incorporation of **10** into mDHFR with any of the mutants. The specificity constant k_{cat}/K_m for **10** by T251G is reduced by 124-fold compared to **1** by wild-type; this reduced activity is not sufficient to explain the lack of translational activity since we have observed incorporation of **4** and **7** despite their even lower specificity constants (Table 6-2 and 6-3). Sisido and co-worker investigated

the adaptability of aromatic non-natural amino acids to the *E. coli* ribosome and found that certain ring structures are not allowed to occupy ribosome A site. Analog **10** has one of these "forbidden" ring structures (*39*), implying that our mutant synthetases might be able to attach **10** into $tRNA^{Phe}$, however this aminoacylated tRNA could not be adapted into *E. coli* ribosome for further translation reactions.

Although we did not observe *in vitro* activities of analogs **11-13** by any of the variant synthetases, we still examined their abilities to support protein biosynthesis *in vivo* and found that none of them showed a detectable level of incorporation, confirming that amino acid activation is pivotal for the *in vivo* translation system to utilize analogs for protein synthesis.

4. Conclusion

One of our objectives for this work is to expand the set of amino acid building blocks for protein engineering and biomaterial engineering. We and others have developed several *in vivo* methods to accomplish this goal (*13, 18, 21, 22, 25, 38, 40, 41*). Generally, alteration of cellular aminoacylation reactions could enable us to introduce many chemically and biophysically interesting side chains into recombinant proteins. The methods include over-expression of wild-type aminoacyl-tRNA synthetases (*13, 18*), introduction of newly designed synthetase activities (*11, 20, 21*), import of a novel tRNA/synthetase pair(*25, 38*), or attenuation of editing abilities of synthetases (*22, 41*). The results described in this chapter show additional evidence that design of new synthetase activities could be a powerful tool to introduce non-canonical amino acids into proteins in a multi-site fashion. With the mutant synthetases (T251G and

T251G/A294G), amino acids such as **6-7** and **9** are suitable substrates for protein synthesis *in vivo*. Along with previous success of incorporation of **2-5**, we can introduce many aromatic side chains with distinct photo- or electro-physical properties into biomacromolecules.

Scheme 1

$$aaRS + aa + ATP \leftrightarrow aaRS \circ (aa - AMP) + PP_i \quad \textbf{(1)}$$

$$aaRS \circ (aa - AMP) + tRNA^{aa} \leftrightarrow aa - tRNA^{aa} + aaRS + AMP \quad \textbf{(2)}$$

Table 6-1: ORBIT Calculation for *p*-Acetyl-phenylalanine (2) Binding into *t*PheRS.

Residue	184	222	258	261	286	290	294	314	Ligand Energy (Kcal/mol)	Total Energy (Kcal/mol)
*t*PheRS-1 (wild-type)	V	L	F	V	V	V	V	A	-16.91	-240.71
*t*PheRS-1 (calculated)	I	A	Y	V	L	I	I	A	-16.87	-242.14
*t*PheRS-2 (calculated)	I	L	Y	G	L	I	V	G	-21.40	-225.13

Table 6-2: ATP-PPi Exchange Kinetics of Wild-type and Mutant Forms of ePheRS toward Canonical (1 and 8) and Non-canonical (2-7, 9-10) Amino Acids

Amino Acid	Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)
1	Wild-type	28±10	1.4±0.12	49,593±18,074	1
1	A294G	455±281	0.09±0.018	197±128	1/251
1	T251G	45±17	0.14±0.013	3,076±1,171	1/16
1	T251G/A294G	976±208	0.06±0.005	62±14	1/806
2	Wild-type	–	–	–	–
2	A294G	–	–	–	–
2	T251G	502±48	0.14±0.005	279±28	1/178
2	T251G/A294G	36±4	0.07±0.002	1,918±236	1/26
3	Wild-type	–	–	–	–
3	A294G	3936±1942	0.02±0.006	5±2	1/9761
3	T251G	40±5	0.14±0.003	3,441±447	1/14
3	T251G/A294G	34±6	0.04±0.001	1,158±195	1/43
4	Wild-type	–	–	–	–
4	A294G	4526±2142	0.02±0.001	4±2	1/11223
4	T251G	568±60	0.12±0.004	211±23	1/235
4	T251G/A294G	2056±216	0.03±0.001	15±2	1/3340
5	Wild-type	–	–	–	–
5	A294G	124±43	0.03±0.003	242±86	1/205
5	T251G	4.6±0.98	0.61±0.041	131,466±29,165	3/1
5	T251G/A294G	324±63	5.1±1.63	15,648±5,897	1/3
6	Wild-type	–	–	–	–
6	A294G	–	–	–	–
6	T251G	55±11	0.05±0.002	909±186	1/54
6	T251G/A294G	13±2	0.02±0.001	1575±316	1/31
7	Wild-type	–	–	–	–
7	A294G	4555±2400	0.12±0.043	26±17	1/1882
7	T251G	294±66	0.05±0.003	170±40	1/292
7	T251G/A294G	323±97	0.01±0.001	31±10	1/1601
8	Wild-type	–	–	–	–
8	A294G	–	–	–	–
8	T251G	248±84	0.07±0.007	282±100	1/176
8	T251G/A294G	5.0±1.5	0.01±0.001	2012±624	1/25
9	Wild-type	–	–	–	–
9	A294G	–	–	–	–
9	T251G	26±10	0.14±0.011	5,291±2,045	1/9
9	T251G/A294G	17±5	0.07±0.005	4,028±1,220	1/12
10	Wild-type	–	–	–	–
10	A294G	–	–	–	–
10	T251G	50±20	0.02±0.002	400±164	1/124
10	T251G/A294G	ND	ND	ND	ND

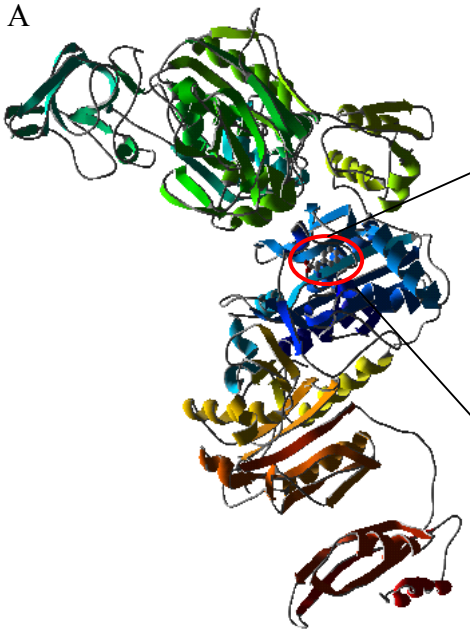
Table 6-3: Mass Data for Peptide Fragment 1 Derived from mDHFR Expressed in Media Supplemented with 1, No Analog or Analog 2-13.

Analog supplemented in media	Theoretical Δm/z	A294G		T251G		T251G/A294G	
		Observed Δm/z	Observed Analog Incorporation	Observed Δm/z	Observed Analog Incorporation	Observed Δm/z	Observed Analog Incorporation
1	0	0	–	0	–	0	–
No analog	0	0	–	38.993	**8**	38.997	**8**
2	42.040	0	–	41.950	**2**	42.050	**2**
3	125.897	125.804	**3**	125.871	**3**	125.829	**3**
4	24.995	24.992	**4**	24.946	**4**	24.973	**4**
		24.970		38.974	**8**	38.984	**8**
5	41.001	15.016	**7**	14.980	**7**	14.995	**7**
				30.968	*solvolysis*	31.007	*solvolysis*
6	44.985	0	–	28.971	*photolysis*	28.999	*photolysis*
				38.949	**8**	38.996	**8**
				44.978	**6**	44.972	**6**
7	15.011	15.032	**7**	14.974	**7**	14.995	**7**
		58.000	*	38.975	**8**	38.993	**8**
				57.969	*	58.006	*
8	39.011	0	–	39.024	**8**	39.004	**8**
9	50.016	0	–	49.968	**9**	49.964	**9**
10	50.016	0	–	38.988	**8**	38.982	**8**
11	14.016	0	–	39.013	**8**	39.004	**8**
12	89.953	0	–	38.995	**8**	38.980	**8**
13	6.047	0	–	38.986	**8**	38.998	**8**

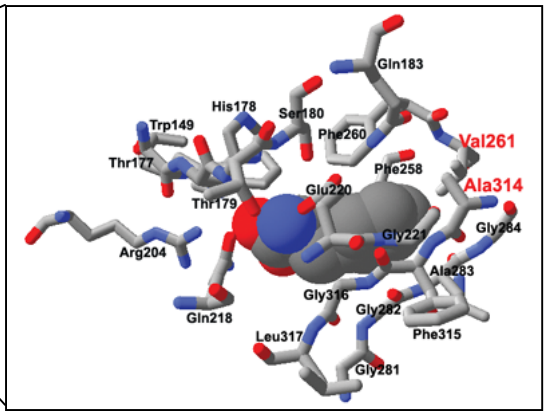
Figure 6-1:

(A) Cystal structure of the *T. thermophilus* PheRS (*t*PheRS, pdb 1B70) in ribbon model. Only an ($\alpha\beta$) portion of the 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 residues of V261 and A314 from *T. thermophilus*. Sequences adjacent to V261 and A314 are also highly conserved. The sequences are obtained from the aminoacyl-tRNA synthetase database (http://rose.man.poznan.pl/aars/).

A



B

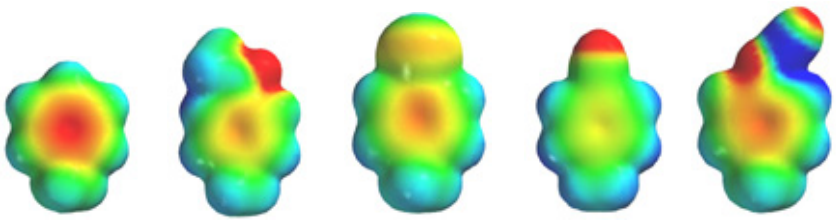


C

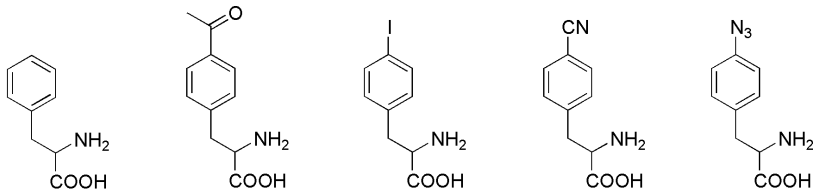
T. thermophilus	257YFPFVEP263	·····	313FAFGLGVERLAML325
E. coli	247YFPFTEP253	·····	293FAFGMGMERLTML305
S. cerevisiae	410YNPYTEP416	·····	456LGWGLSLERPTMI468
B. burgdorferi	441YFPFTEP447	·····	484IAWGIGIDRMALM496
B. subtilis	251FFPFTEP257	·····	308FAFGMGVERIAML320
A. aeolicus	247YFPFTEP253	·····	304FAFGMGVERLAML316
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 6-2:

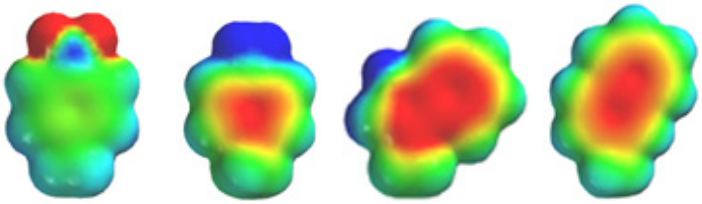
Chemical structures of the amino acids involved in this study. For each amino acid, from top to bottom, electron density map of side chain; side chain volume; chemdraw structure. The electron density map is generated from geometry optimization calculation with AM force field performed by using the program MacSpartan (Wavefunction, Inc., Irvine, CA). The side chain volume is derived from the same calculation. The amino acids are: phenylalanine (**1**), *p*-acetylphenylalanine (**2**), *p*-iodo-phenylalnine (**3**), *p*-cyano-phenylalanine (**4**), *p*-azido-phenylalanine (**5**), *p*-nitro-phenylalanine (**6**), *p*-amino-phenylalanine (**7**), tryptophan (**8**), 3-(2-naphthyl)alanine (**9**), 3-(1-naphthyl)alanine (**10**), homo-phenylalanine (**11**), *penta*-fluorophenylalanine (**12**) and cyclohexaalanine (**13**).



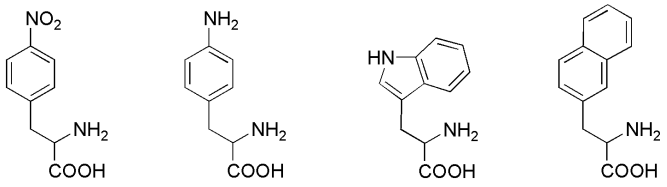
Side Chain Volume (Å^3) 121.11 157.49 154.65 135.02 139.20



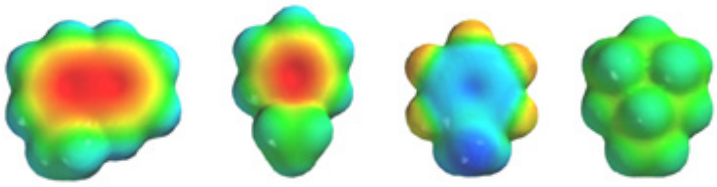
1 **2** **3** **4** **5**



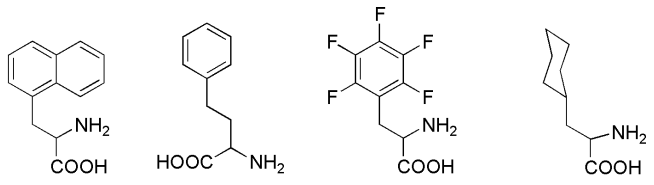
Side Chain Volume (Å^3) 133.94 131.91 152.42 171.71



6 **7** **8** **9**



Side Chain Volume (Å^3) 171.13 141.92 123.23 147.27



10 **11** **12** **13**

Figure 6-3:

Ribbon representation of the portion of catalytic α-subunit of PheRS from *T. thermophilus* that surrounds residues V251 and A294. Side chains of residues V251 and A294 are shown in stick model.

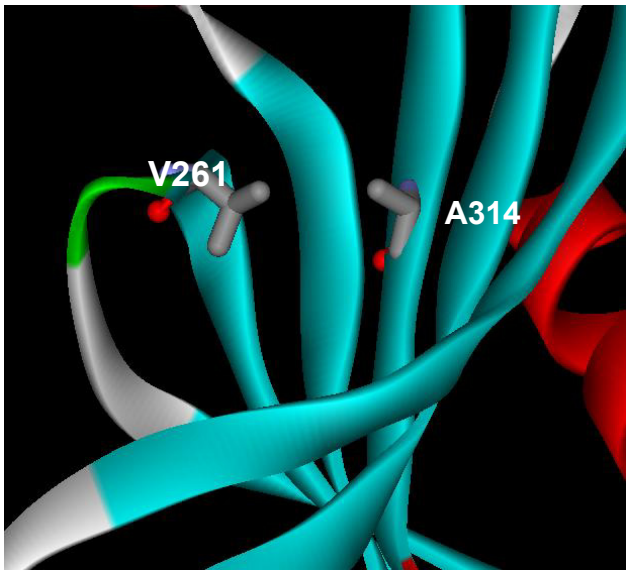
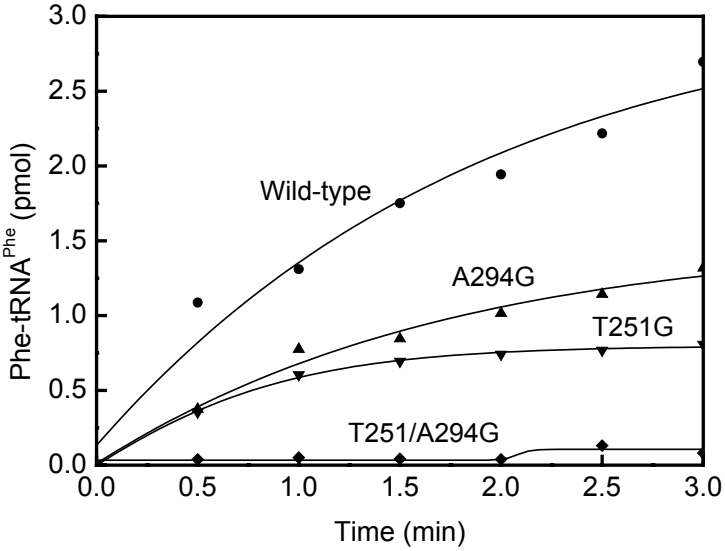

V261
A314

Figure 6-4:

Aminoacylation of $tRNA^{Phe}$ by wild-type and mutant *e*PheRS enzymes at 37°C. Each enzyme (wild-type, A294G, T251G and T251/A294G) was evaluated for its ability to aminoacylate Phe (A) and Trp (B): (●) wild-type enzyme; (▲) A294G; (▼) T251G and (♦)T251G/A294G.


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Phe-tRNA^{Phe} (pmol)
Wild-type
A294G
T251G
T251/A294G
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time (min)

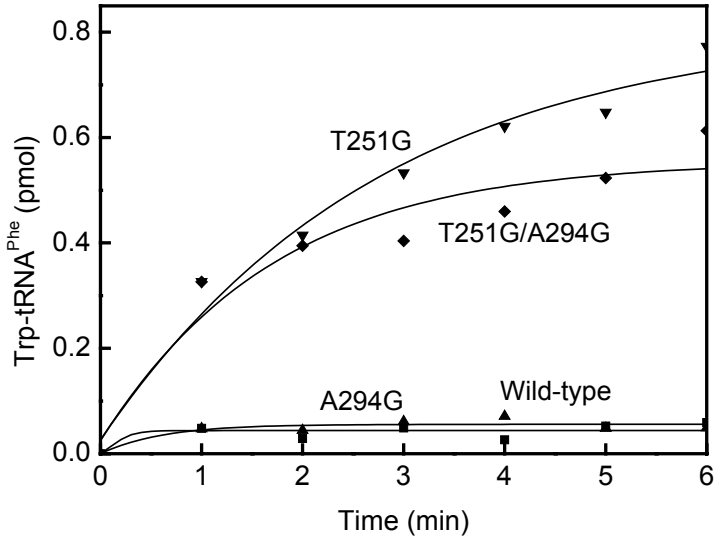

0.8
0.6
0.4
0.2
0.0
Trp-tRNA^{Phe} (pmol)
T251G
T251G/A294G
A294G
Wild-type
0
1
2
3
4
5
6
Time (min)

Figure 6-5:

Amino acid sequence of target protein mDHFR. The protein contains 209 residues, of which 9 are phenylalanines. Two common tryptic fragments are underlined as shown. Fragment 1 has one Phe and its expected mass is 1682.86; fragment 2 has one Phe and its expected mass is 1592.89.

MRGSHHHHHHGSGIMVRPLNSIVAVSQNMG

IGKNGDLPWPPLRNE**F**KY**F**QRMTTTSSVEG
Fragment 1: 1682.86
KQNLVIMGRKTW**F**SIPEKNRPLKDRINIVL

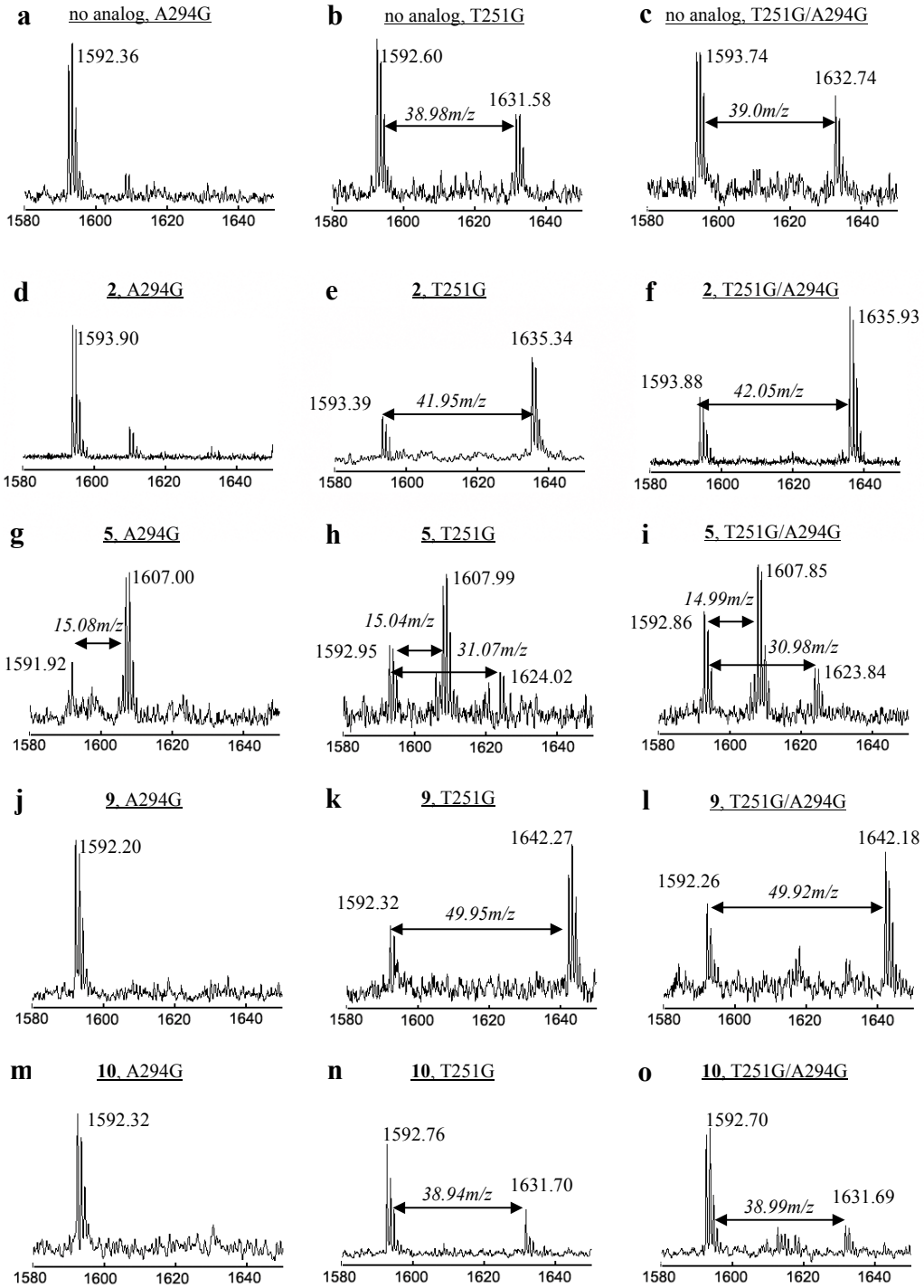
SRELKEPPRGAH**F**LAKSLDDALRLIEQPEL
Fragment 2: 1592.89
ASKVDMVWIVGGSSVYQEAMNQPGHLRL**F**V

TRIMQE**F**ESDT**FF**PEIDLGKYKLLPEYPGV

LSEVQEEKGIKYK**F**EVYEKKGWKILSLIS

Figure 6-6:

MALDI-MS of tryptic peptide fragment 2 derived from mDHFR expressed from media containing no analog, **2**, **5**, **9** or **10**. For each amino acid, experiment has been tested with over-expression of three individual mutant synthetases. The peptide has the sequence ELKEPPRGAH**F**LAK. The expected mass for this fragment is 1592.89. The possible substitution would occur at one Phe site. Mass for each peak and its shift have been labeled as shown.


a
no analog, A294G
1592.36
1580
1600
1620
1640
b
no analog, T251G
1592.60
38.98m/z
1631.58
c
no analog, T251G/A294G
1593.74
1632.74
39.0m/z
d
2, A294G
1593.90
e
2, T251G
1635.34
1593.39
41.95m/z
f
2, T251G/A294G
1635.93
1593.88
42.05m/z
g
5, A294G
1607.00
15.08m/z
1591.92
h
5, T251G
1607.99
15.04m/z
1592.95
31.07m/z
1624.02
i
5, T251G/A294G
1607.85
14.99m/z
1592.86
30.98m/z
1623.84
j
9, A294G
1592.20
k
9, T251G
1642.27
1592.32
49.95m/z
l
9, T251G/A294G
1642.18
1592.26
49.92m/z
m
10, A294G
1592.32
n
10, T251G
1592.76
38.94m/z
1631.70
o
10, T251G/A294G
1592.70
38.99m/z
1631.69

5. References

1. Soll, D., and RajBhandary, U. L. (1995), ASM Press, Washington, D.C.

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

3. Carter, C. W., Jr. (1993) *Annu. Rev. Biochem. 62*, 715-748.

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

5. Sankaranarayanan, R., and Moras, D. (2001) *Acta Biochimica Polonica 48*, 323-325.

6. Cusack, S. (1995) *Nature Struct. Biol. 2*, 824-831.

7. Sprinzl, M., and Cramer, F. (1975) *Proc. Natl. Acad. Sci. USA 72*, 3049-3053.

8. Fraser, T. H., and Rich, A. (1975) *Proc. Natl. Acad. Sci. USA 72*, 3044-3048.

9. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) *Nature Struct. Biol. 2*, 537-547.

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

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

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

14. Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, W. A., DeGrado, W. F., and Tirrell, D. A. (2001) *Biochemistry 40*, 2790-2796.

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

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

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

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

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

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

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

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

23. Kothakota, S. (1995) in *Department of Polymer Science and Engineering* pp 65, University of Massachusetts, Amherst.

24. Wang, P., Vaidehi, N., Tirrell, D. A., and Goddard, W. A. (2002) *J. Am. Chem. Soc. 124*, 14442-14449.

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

26. Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) *Nature Biotech. 20*, 1044-1048.

27. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) *J. Mol. Biol. 273*, 789-796.

28. Dahiyat, B. I., and Mayo, S. L. (1997) *Science 278*, 82-87.

29. Gordon, D. B., and Marshall, S. A. (1999) *Curr. Opin. Struct. Biol. 9*, 509-513.

30. Calendar, R., and Berg, P. (1966) *Biochemistry 5*, 1681-1690.

31. Tocchini-Valentini, G., Saks, M. E., and Abelson, J. (2000) *J. Mol. Biol. 298*, 779-793.

32. Hendrickson, T. L., Nomanbhoy, T. K., and Schimmel, P. (2000) *Biochemistry 39*, 8180-8186.

33. Sampson, J. R., and Uhlenbeck, O. C. (1988) *Proc. Natl. Acad. Sci. USA 85*, 1033-1037.

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

35. Ibba, M., Kast, P., and Hennecke, H. (1994) *Biochemistry 33*, 7107-7112.

36. Avis, J. M., Day, A. G., Garcia, G. A., and Fersht, A. R. (1993) *Biochemistry 32*, 5312-5320.

37. Avis, J. M., and Fersht, A. R. (1993) *Biochemistry 32*, 5321-5326.

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

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

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

41. 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.