Chapter 4

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

This work was completed in collaboration with Pin Wang, Kent Kirshenbaum, Yi Tang, Deepshika Datta and Steve Mayo.

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 to 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: cognate amino acids react with ATP to form aminoacyl-adenylates; subsequently these activated forms of the amino acids are attached to their cognate tRNAs by esterification. 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].

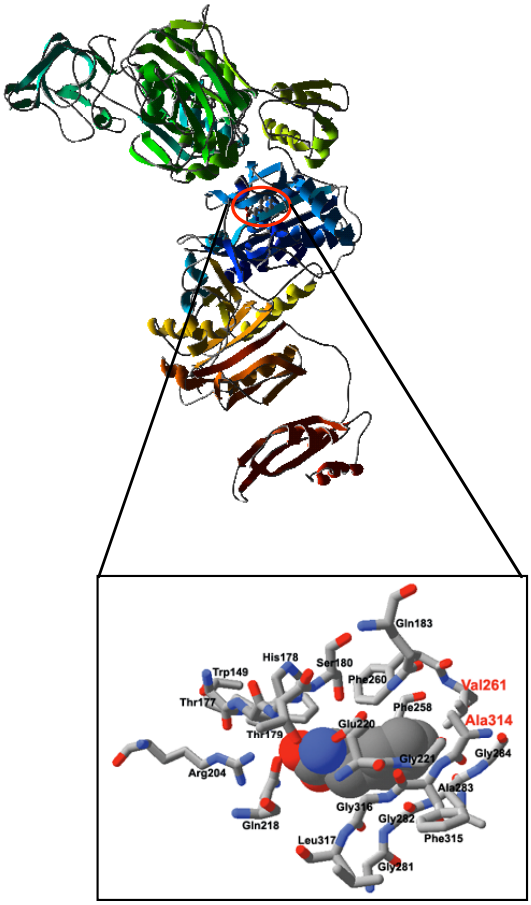
The error rate for recognition of tRNAs by aaRSs is extremely low (10^{-6} or lower) [5] because of the large tRNA/aaRS contact area. Discrimination of substrate amino acids is more challenging because they are much smaller and have less structural variability. On the basis of mutually exclusive sequence 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 an in-line nucleophilic displacement reaction [4].

Among Class II aaRSs, phenylalanyl-tRNA synthetase (PheRS) is unique in that it attaches its cognate amino acid 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* (PheRS) reveals that the α-subunit is the catalytic unit and the major function for the β-subunit is recognition and binding of $tRNA^{Phe}$ (Figure 4.1) [9, 10].

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

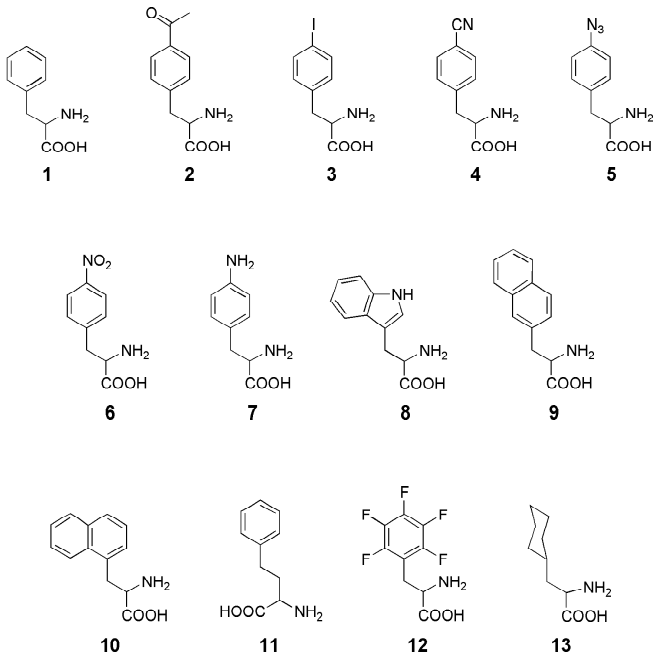
Figure 4.1. Ribbon representation of the portion of catalytic -subunit of PheRS from *T. thermophilus*.. The active site, expanded below, demonstrates bound Phe in space filling model and proximal residues in stick representation. Val251 and A314, highlighted in red, define the distal endof the binding pocket.



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 the intracellular pools 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 binding pockets for recognition of novel amino acid substrates.

In this report, we investigate the ability of variants of *E. coli* PheRS with relaxed substrate specificity to allow incorporation of a diverse set of non-natural amino acids (**2-9**)(Figure 4.2). One newly identified mutant (T251G) and two previously studied mutants, A294G (PheRS*, Chapter 2) and T251G/A294G (PheRS**, Chapter 3), of 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**). 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*.

Figure 4.2. Chemical structures of the amino acids involved in this study. 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**).



MATERIALS AND METHODS

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).

Plasmid construction for synthetase expression

The PheRS gene was cloned directly from *E. coli* genomic DNA with flanking primers encoding 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 PheRS. Briefly, a pair of complementary oligos, designated as primer 3 and primer 4, was designed to carry a specific mutation at position 294 or 251 of the a subunit of PheRS. In one reaction primer 1 and primer 4 were used to yield the first DNA fragment of the PheRS gene. In another reaction primer 3 and primer 2 were used to yield the second DNA fragment of the PheRS gene. The two fragments were then mixed for further amplification in the presence of primer 1 and primer 2 to afford the entire 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 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.

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 possibility of chromosomal recombination with the endogenous *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 by 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 ºC. The concentration of the purified enzyme was determined by absorbance at 280 nm under denaturing conditions.

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 [27]. 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 the synthetase 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.

Plasmid construction for in vivo incorporation assays

The *E. coli pheS* gene for the α subunit of *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.

Analog incorporation assays in vivo

A phenylalanine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AF (*HsdS gal* (*λ*cI*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 ºC, 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 %).

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.

RESULTS AND DISCUSSION

Expression and purification of PheRS variants

The wild-type *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) is a gene fragment that encodes the α and β subunits of *PheRS* and a 14 bp of an intercistronic region [28]. Enzymes were expressed in the *E. coli* strain XL-1 blue, which harbors the plasmid of individual pQE30 derivatives containing variants of *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.

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**) are shown in Table 4.1 and 4.2. The K_m value obtained from our measurement for **1** by wild-type *PheRS* is comparable with previously reported value [29], although k_{cat} values are lower; this is presumably due to the different buffer conditions and different methods of measuring

Table 4.1. ATP-PPi exchange kinetics of wild-type and mutant forms of PheRS toward canonical (**1**) and non-canonical (**2-5**) amino acids.

Amino Acid	Enzyme	K_m (□M)	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

Table 4.2. ATP-PPi exchange kinetics of wild-type and mutant forms of PheRS toward canonical (**8**) and non-canonical (**6-8, 9, 10**) amino acids.

Amino Acid	Enzyme	K_m (□M)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)
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

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 *PheRS* (Chapter 2). T251G displayed extremely high reactivity toward **5** (k_{cat}/K_m 3-times that of wild-type toward **1**). *para*-NitroPhe (**6**) was activated by both T251G and T251G/A294G enzymes with similar k_{cat}/K_m values. When we measured the activation rates of bulkier amino acids such as **8**, **9** and **10**, we found that only T251G and T251G/A294G could activate these analogs. Their k_{cat}/K_m values ranging from 9 to 176-times lower than wild-type PheRS toward **1**. Both mutants exhibit striking activities toward the canonical amino acid **8**; only a single mutation can pose significant threats to the fidelity of PheRS. Under our assay conditions, we did not observe above-background activation of **11-13** or tyrosine by any of the synthetases.

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 measured. From the crystal structure, V261, equivalent to T251 in *E. coli*, is located in a loop region between two β-strands [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 4.1); Phe is 16-times poorer a substrate for A294G than for T251G.

In vivo evaluation by DHFR 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 (Figure 4.3). 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. Representative tryptic mass spectra for situations where the media contain either no Phe or analog **2, 5, 9** and **10** are shown in Figure 4.4. Table 4.3 lists MALDI-TOF data for fragment 2 derived from mDHFR expressed in media supplemented with **1**, no analog or analogs **2-13**. If 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 4.4e); 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;

Table 4.3. 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 4.4. Amino acid sequence of target protein mDHFR. The protein contain 209 residues, of which 9 are phenylalanines. Two commonly observed 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

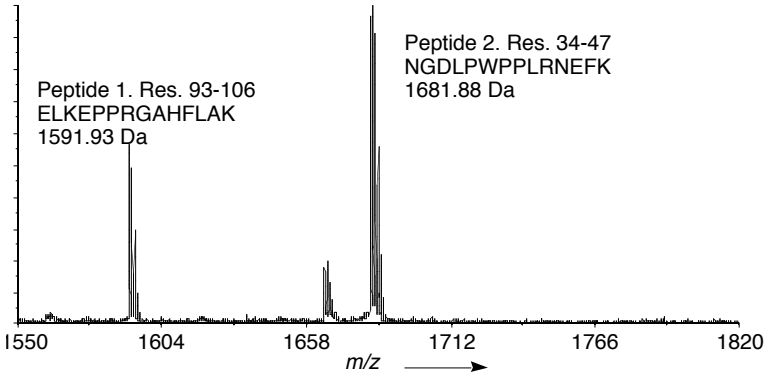
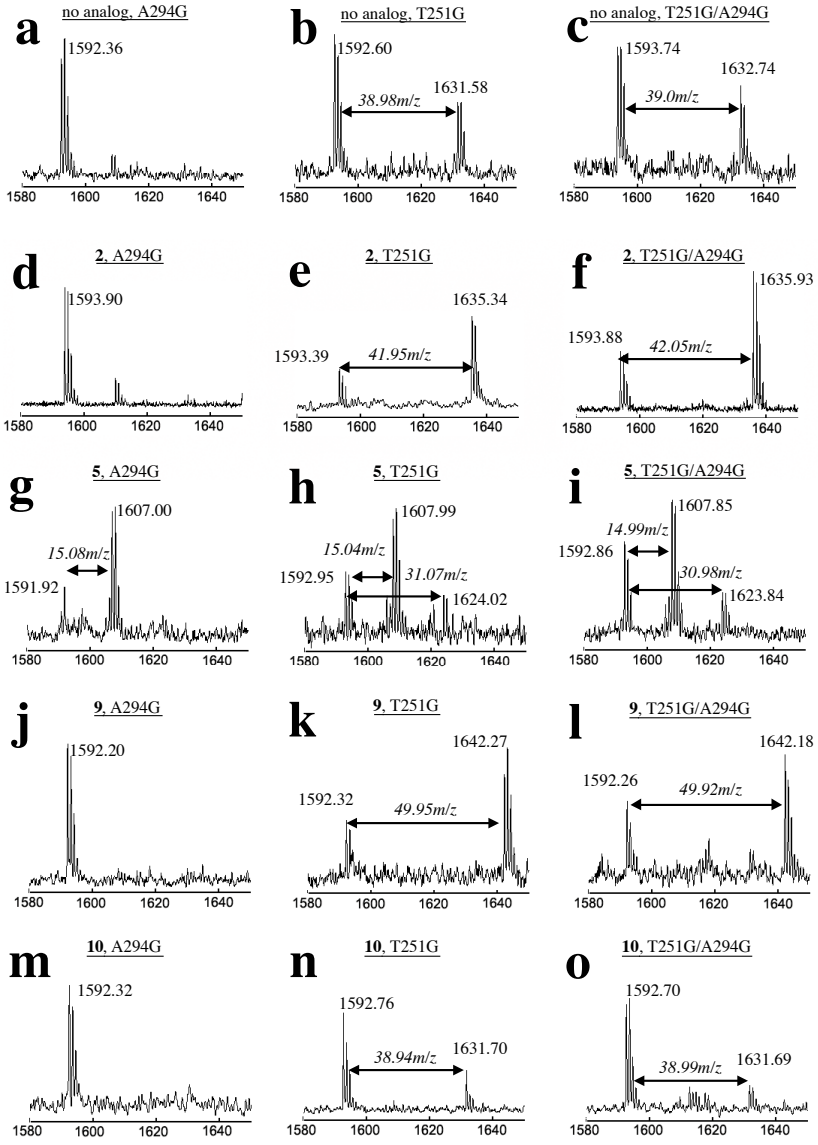


Figure 4.5. MALDI-MS of tryptic peptide fragment 2 derived from mDHFR expressed from media containing no analog, **2**, **5**, **9** or **10**. For each amino acid, experiment has been tested with over-expression of three individual mutant synthetases. The peptide has the sequence ELKEPPRGAH**F**LAK. 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.



mass spectra indicate that no mischarges occur at Phe sites in A294G (Figure 4.4a, Table 4.3), 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 4.4b to 4.4c, Table 4.3). Similar patterns are observed when we supplemented media with **8** (250 mg/L, Table 4.3). This suggests 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 4.2); Figure 4.4c 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 PheRS, which charges Phe 25-times faster than T251G/A294G charges **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 PheRS (Table 4.3). Since analog **4** is activated relatively slowly (Table 4.1) by T251G and T251G/A294G, mass spectra manifest the partial substitution of **8** in place of Phe in addition to substitution of **4** (Table 4.3). From Table 4.1, 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**. Tryptic fragment 2 containing **6** or **7** both shows mass distributions characteristic of incorporation of the analog at Phe sites in addition to partial replacement with **8**. This is in contrast to the *in vitro* data showing that **6** is

a relatively good substrate for both T251G and T251G/A294G (Table 4.2). Despite the high concentration of this analog in the growth media, we still find the co-substitution of **8**, indicating that either poor transport of this analog into the cytoplasm results in low cellular concentration of **6**, or one of the downstream translational components prohibits this amino acid from becoming incorporated into proteins [5].

Based on the acceptance of **8** and the high *in vitro* activities of **9** and **10**, we anticipated that these analogs would be incorporated into DHFR. As expected, the analog **9** incorporated into fragment 2 can be clearly identified (with an additional mass peak; mass shift of +50 Da; Figure 4.4k to 4.4l) 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). Considering the high concentration of **9** in our experiments, 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 4.1,4.2 and 4.3). 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 the ribosomal A site. Analog **10** has one of these "forbidden" ring structures [30], 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.

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, 31-33]. 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, 31], or attenuation of editing abilities of synthetases [22, 33]. The results described in this chapter show additional evidence that design of new synthetase activities can 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.

References

1. Soll, D. and U. L. RajBhandary, eds. *t-RNA, structure, biosynthesis, and function*. 1995, ASM Press: Washington, D.C.

2. Schimmel, P., Annu. Rev. Biochem., 1987. **56**: pp. 125-158.

3. Carter, C. W., Jr., Cognition, *Mechanism, and Evolutionary Relationships in Aminoacyl-Transfer Rna-Synthetases* Annu. Rev. Biochem., 1993. **62**: p. 715-748.

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

5. Sankaranarayanan, R. and D. Moras, *The fidelity of the translation of the genetic code*. Acta Biochimica Polonica, 2001. **48**(2): p. 323-325.

6. Cusack, S., *Eleven down and nine to go*. Nature Struct. Biol., 1995. **2**: p. 824-831.

7. Sprinzl, M. and F. Cramer, *Site of aminoacylation of trnas from E. coli with respect to the 2'- or 3'- hydroxyl group of the terminal adenosine*. Proc. Natl. Acad. Sci. USA, 1975. **72**: p. 3049-3053.

8. Fraser, T. H. and A. Rich, *Amino acids are not initially attached to the same position on trna molecules*. Proc. Natl. Acad. Sci. USA, 1975. **72**: p. 3044-3048.

9. Mosyak, L., et al., *Structure of phenylalanyl-trna synthetase from Thermus thermophilus*. Nature Struct. Biol., 1995. **2**: p. 537-547.

10. Reshetnikova, L., et al., *Crystal structure of phenylalanyl-trna synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue*. J. Mol. Biol., 1999. **287**(2): p. 555-568.

11. Sharma, N., et al., *Efficient introduction of aryl bromide functionality into proteins in vivo*. FEBS Lett., 2000. **467**(1): p. 37-40.

12. van Hest, J. C. M., K. L. Kiick, and D. A. Tirrell, *Efficient incorporation of unsaturated methionine analogues into proteins in vivo*. J. Am. Chem. Soc., 2000. **122**(7): p. 1282-1288.

13. Kiick, K. L., J. C. M. van Hest, and D. A. Tirrell, *Expanding the scope of protein biosynthesis by altering the methionyl-trna synthetase activity of a bacterial expression host*. Angew. Chem., Int. Ed., 2000. **39**(12): p. 2148-2152.

14. Tang, Y., et al., *Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine*. Biochemistry, 2001. **40**(9): p. 2790-2796.

15. Kiick, K. L., R. Weberskirch, and D. A. Tirrell, *Identification of an expanded set of translationally active methionine analogues in escherichia coli (vol 502, pg 25, 2001)*. FEBS Lett., 2001. **505**(3): p. 465-465.

16. Tang, Y., et al., *Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability*. Angew. Chem., Int. Ed., 2001. **40**(8): p. 1494-1496.

17. van Hest, J. C. M. and D. A. Tirrell, *Protein-based materials, toward a new level of structural control*. Chem. Commun., 2001(19): p. 1897-1904.

18. Tang, Y. and D. A. Tirrell, *Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host*. J. Am. Chem. Soc., 2001. **123**(44): p. 11089-11090.

19. Kiick, K. L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the staudinger ligation*. Proc. Natl. Acad. Sci. USA, 2002. **99**(1): p. 19-24.

20. Kirshenbaum, K., I. S. Carrico, and D. A. Tirrell, *Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues*. ChemBioChem, 2002. **3**(2-3): p. 235-237.

21. Datta, D., et al., *A designed phenylalanyl-trna synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins*. J. Am. Chem. Soc., 2002. **124**(20): p. 5652-5653.

22. Tang, Y. and D. A. Tirrell, *Attenuation of the editing activity of the escherichia coli leucyl-trna synthetase allows incorporation of novel amino acids into proteins in vivo*. Biochemistry, 2002. **41**(34): p. 10635-10645.

23. Kothakota, S., *Ph. D. Thesis*, in *Department of Polymer Science and Engineering*. 1995, University of Massachusetts: Amherst. p. 65.

24. Wang, P., et al., *Virtual screening for binding of phenylalanine analogues to phenylalanyl-trna synthetase*. J. Am. Chem. Soc., 2002. **124**(48): p. 14442-14449.

25. Wang, L., et al., *Expanding the genetic code of escherichia coli*. Science, 2001. **292**(5516): p. 498-500.

26. Santoro, S. W., et al., *An efficient system for the evolution of aminoacyl-trna synthetase specificity*. Nature Biotech., 2002. **20**(10): p. 1044-1048.

27. Calendar, R. and P. Berg, Biochemistry, 1966. **5**: p. 1681-1690.

28. Fayat, G., et al., *Escherichia coli phenylalanyl-trna synthetase operon region*. J. Mol. Biol., 1983. **171**: p. 239-261.

29. Ibba, M., P. Kast, and H. Hennecke, *Substrate specificity is determined by amino acid binding pocket site in escherichia coli phenylalanyl-trna synthetase*. Biochemistry, 1994. **33**: p. 7107-7112.

30. Hohsaka, T., et al., *Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems*. J. Am. Chem. Soc., 1999. **121**: p. 34-40.

31. Furter, R., Protein Sci., 1998. **7**: p. 419-426.

32. Wang, L. and P. G. Schultz, *Expanding the genetic code*. Chem. Commun., 2002(1): p. 1-11.

33. Doring, V., et al., Science, 2001. **292**: p. 501-504.