Chapter 3

A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient *in vivo* Incorporation of Aryl Ketone Functionality into Proteins

The work presented here is taken from:
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
Reprinted with permission of the publisher.

Introduction

Protein engineering is a powerful tool for modification of the structural, catalytic and binding properties of natural proteins and for the *de novo* design of artificial proteins. Although amino acid replacement is normally limited to the twenty proteinogenic amino acids, it is becoming increasingly clear that incorporation of non-natural amino acids can extend the scope and impact of protein engineering methods [1, 2].

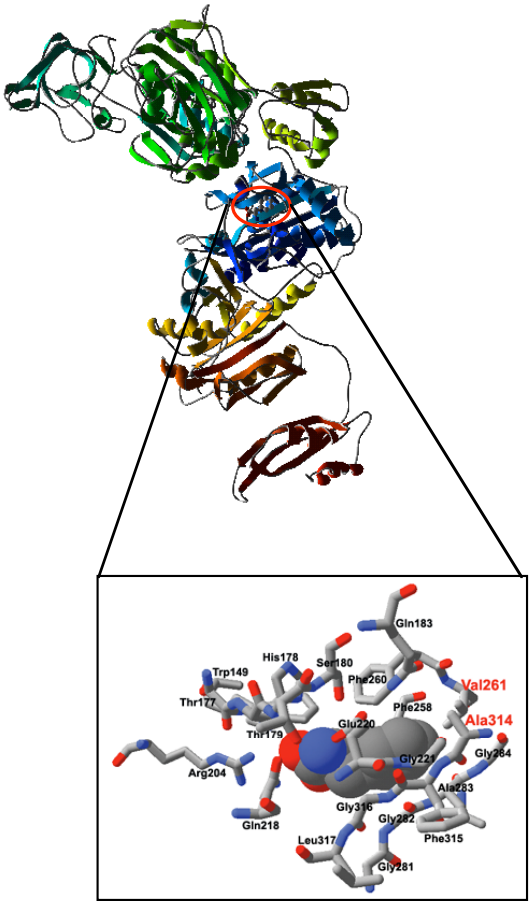
We have previously exploited the ability of auxotrophic *Escherichia coli* strains to effect efficient incorporation of amino acid analogues into proteins in a multi-site fashion. The method is simple and produces high protein yields, and incorporation of the analogue at multiple sites offers significant advantages with respect to control of protein properties such as thermal and chemical stability [3-6].

In this study, we report a computationally designed variant of the *E. coli* phenylalanyl-tRNA synthetase (PheRS), which allows efficient *in vivo* incorporation of aryl ketone functionality into proteins. In 1991, Kast and coworkers [7-9] introduced a variant of the *E. coli* PheRS (termed PheRS*), which bears an Ala294Gly mutation in the β-subunit and which thereby acquires relaxed substrate specificity. We have recently shown that over-expression of PheRS* can be exploited to effect efficient incorporation of *p*-bromo-, *p*-iodo-, *p*-ethynyl-, *p*-cyano- and *p*-azidophenylalanines into recombinant proteins in *E. coli* hosts (Chapter 2) [10]. But similar experiments with *p*-acetylphenylalanine (**2**) failed; even in a host in which PheRS* was over-expressed, phe-depleted cultures supplemented with **2** did not produce substantial yields of protein.

Our interest in **2** arises from the chemical versatility of the side-chain ketone function, which can be chemoselectively ligated with hydrazide, hydroxylamino, and thiosemicarbazide reagents under physiological conditions [11-14]. Cornish and coworkers have accomplished site-specific incorporation of ketone functionality into recombinant proteins via *in vitro* translation [15]; however, at the time of this work there were no previous reports of *in vivo* methods of introducing ketone functionality into recombinant proteins.

We sought PheRS mutants that would allow efficient incorporation of **2** into recombinant proteins *in vivo*. The crystal structure of *Thermus thermophilus* PheRS complexed with **1** is available[16, 17] (Figure 3.1) and while there is 43% overall sequence identity between the *T. thermophilus* PheRS and *E. coli* PheRS; sequence identity in the identified active site region is 80%. We therefore employed a previously described protein design algorithm [18] to identify potentially useful mutants of the *T. thermophilus* PheRS, with the intention to prepare and evaluate the corresponding mutant forms of *E. coli* PheRS.

Figure 3.1. 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 that surrounds residues V261 and A294. Side chains of residues V261 and A314 are highlighted in red.



Materials and Methods

2 was purchased from RSP Analogs. Biotin hydrazide was purchased from Molecular Probes.

Computational redesign of the Thermus thermophilus PheRS active site

We used a protein design algorithm [18, 19], ORBIT, to predict the optimal amino acid sequences of the binding pocket of PheRS from *Thermus thermophilus* (PheRS; PDB code: 1B70) [20] for binding to **2**. 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 **2**, the closest approximations for χ_1 and χ_2 angles for **2** 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. Accordingly we generated a backbone independent rotamer library for analog **2**. The torsional angles of substrate Phe complexed with *T. thermophilus* PheRS in the crystal structure (χ_1: -101°; χ_2 : -104°) were also included in the new rotamer

libraries for both Phe and **2**. 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 [21]. Calculations were performed by anchoring the substrate, **2,** and varying 11 residues within 6 Å of the substrate in the binding pocket of PheRS (L137, V184, M187, L222, F258, F260, V261, V286, V290, V294, A314). At positions 137, 184, 258, 260, 261, 286, 290, 294 and 314 any of the 20 natural amino acids were allowed 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.

Plasmid Construction

*E. coli pheS** was amplified by the polymerase chain reaction (PCR) from vector pQE-FS. Amplified *pheS** was subjected to PCR mutagenesis to create the coding sequence for the desired Thr251Gly mutant, which we designate *pheS***. To allow constitutive expression 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'CTAGCAGTTGACAATTAATCATCGGCTCGTATAATGGATCGAATT
GTGAGCGGAATCGATTTTCACACAGGAAACAGACC**ATG**GATCTTCGTCGC
CATCCTCGGGTCGACGTCTGTTTGCAAGCTTG-3'

(the –35 and –10 sequences are underlined and start codon is in bold). This linker was cloned into the *Nhe*I site of vector pET5a (Novagen) to form pET5a-tac. PCR amplified fragments containing *pheS** and *pheS*** were cloned into pET5a-tac at the *Nco*I and *Hind*III sites. *pheS** and *pheS*** outfitted with the *tac* promoter were cut out as *Nhe*I fragments and inserted into expression plasmid pQE15 (Qiagen) to yield pQE-FS* and pQE-FS** respectively. Expression plasmids pQE15, pQE-FS* and pQE-FS** encode murine dihydrofolate reductase (mDHFR) under control of a bacteriophage T5 promoter.

Determination of translational activity

Buffer and media were prepared according to standard protocols. A phenylalanine auxotrophic derivative of *E. coli* strain BL21(DE3), designated AF (*HsdS gal* (*cI*ts857 *ind* 1 *S*am7 *nin*5 *lac*UV5-T7 gene *1*) *pheA*) and constructed in our laboratory, was used as the expression host. The AF strain was transformed with repressor plasmid pLysS-IQ and with pQE15, pQE-FS* or pQE-FS** to afford expression strains AF-IQ[pQE15], AF-IQ[pQE-FS*] or AF-IQ[pQE-FS**] respectively.

Small-scale (10 ml) cultures were used to investigate the *in vivo* translational activity of **2**. M9 minimal medium (50 ml) supplemented with 0.2 % glucose, 1mg/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 **2**, (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 %), using a normalized OD_{600} of 0.2 per sample.

Protein purification

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). The eluted protein was buffer-exchanged (Millipore, MWCO=5 kDa) against distilled water three times and the purified protein was subjected to matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis.

Tryptic peptide analysis

10 ml of purified protein in elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH=4.5) was mixed with 90 ml 75 mM NH_4OAc, to which 2 mL of modified trypsin (Promega, 0.2 mg/mL) 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 ml of purified sample solution. 10 ml of the MALDI matrix (a-cyano-b-hydroxycinnamic acid, 10 mg/ml in 50% CH_3CN) was added, and 0.5 ml 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.

*Chemical modification of DHFR-**2** with biotin hydrazide*

Purified proteins (mDHFR-wt and mDHFR-2) were dissolved in 200 μl of PBS buffer (pH=6.0) and added to 20 μl of 5mM biotin hydrazide (BH, dissolved in

PBS). Protein/BH mixtures were incubated at room temperature for 1 to 1.5 h. Reaction solutions were then washed twice with distilled water using a buffer-exchange column (Millipore, MWCO=5 kDa). Standard western blotting procedures were used to identify proteins modified with BH as well as those bearing an N-terminal hexahistidine tag.

Results and Discussion

Mutational predictions based on ORBIT calculations

The calculations identified two important cavity-forming mutations: Val261 (Thr251 in *E. coli*) to Gly, and Ala314 (Ala294 in *E. coli*) to Gly (Table 2.1)(Figure 3.2). These predictions are consistent with the results of Reshetnikova and coworkers [16], who pointed out that Ala314 and Val261 hinder the binding of amino acids larger than phe (e.g., tyrosine) into the active site of PheRS. Further confidence in the prediction was engendered by the fact that the Ala294Gly mutant allows incorporation of an interesting set of *para*-substituted phenylalanines, as described earlier (Chapter 2). We were thus encouraged to test whether the additional Thr251Gly mutation would relax the specificity of PheRS* sufficiently to allow incorporation of **2** into proteins *in vivo*.

*PAGE analysis of mutant synthetase effects on incorporation of **2** into mDHFR*

The capacity of **2** to support protein synthesis in each expression system was determined by induction of mDHFR expression in phenylalanine-free minimal media supplemented with **2**. As shown in SDS-PAGE analysis of whole cell lysates (Figure 3.3), neither AF-IQ[pQE15] nor AF-IQ[pQE-FS*] exhibits protein expression above background (-phe) in media supplemented with **2**. In contrast, similarly supplemented cultures of AF-IQ[pQE-FS**] yield high levels of mDHFR expression. The histidine-tagged protein from the latter culture (mDHFR-**2**) was purified in a

Table 3.1. ORBIT Calculation for *para*-acetylphenylalanine (**2**) binding into *Thermus thermophilus* 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

Figure 3.2. Active site of *Thermus thermophilus* PheRS/Phe. Protein residues are shown as stick models and the substrate Phe is shown as a ball-and-stick model. Yellow protein residues are involved in hydrophobic interactions with the substrate. (a) wild type; (b) A314G mutant; (c) L261G/A314G mutant.

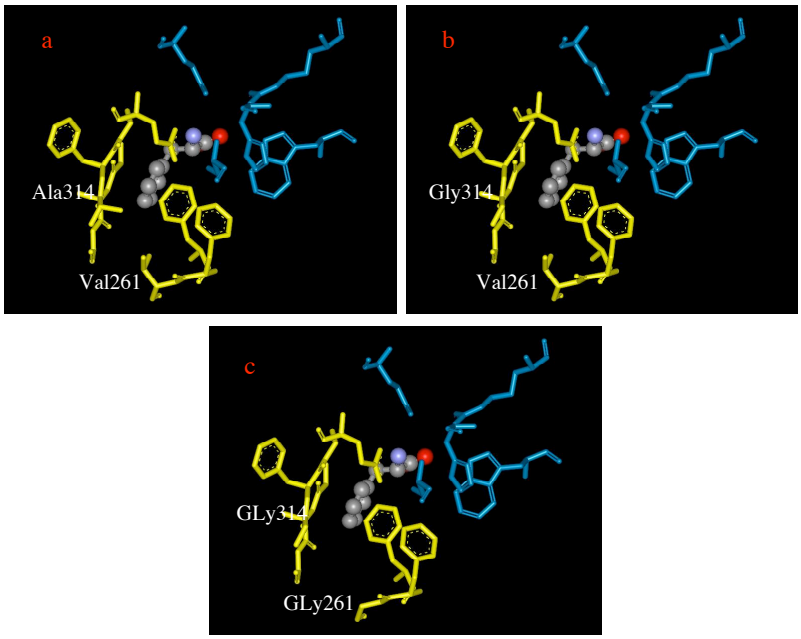
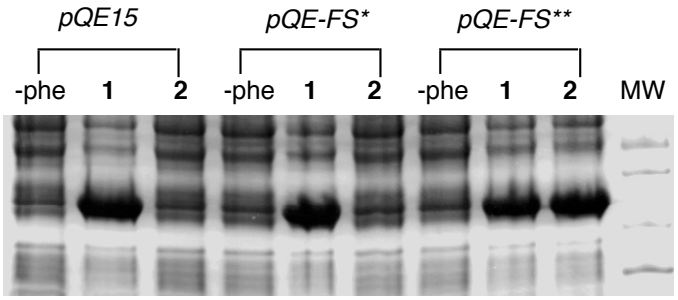


FIgure 3.3. SDS-PAGE of cell lysates of 4 hr post-induction with 1 mM IPTG. Expression plasmids and amino acid supplements are indicated. Concentration of **1**=20mg/L; **2**=250mg/L. Lane MW: molecular weight marker (36.5, 31, 21.5, 14.4 kDa).



yield of about 20 mg/L, approximately 60% of that obtained from cultures supplemented with phenylalanine.

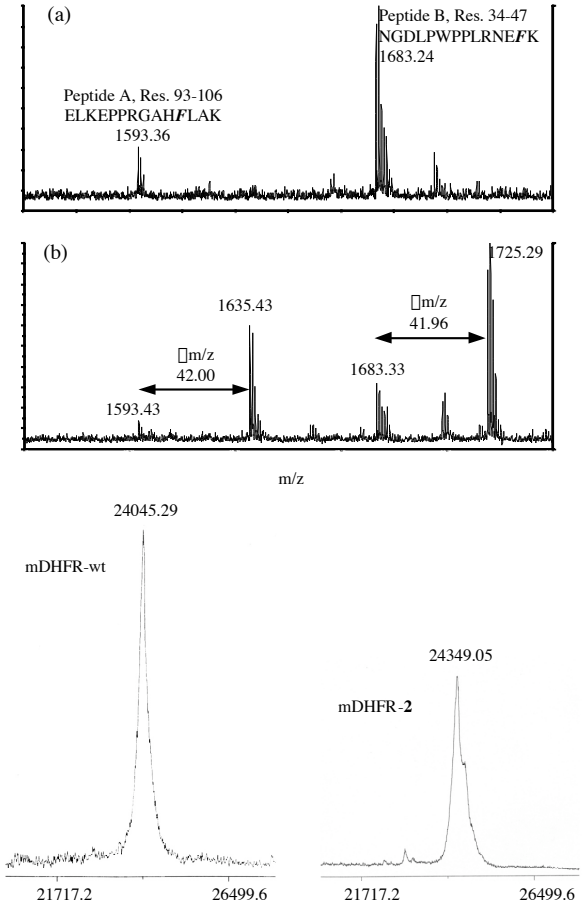
*MALDI-TOF analysis of incorporation of **2***

MALDI-TOF mass spectrometry showed that the mass of mDHFR-**2** was increased by 304 Da, which corresponds to approximately 80% replacement of phe (**1**) by **2** (mDHFR contains 9 phe residues). Incorporation of **2** was confirmed by tryptic digestion of mDHFR-**2** (Figure 3.4a and b). For mDHFR, two peptides in the mass range 1550-1750 Da were assigned to residues 34-47 and 93-106, respectively (Figure 3.4a). Each fragment contains a single phe residue. The corresponding fragments of mDHFR-**2** (Figure 3.4b) were shifted up in mass by 42 Da, consistent with the increased mass of **2** relative to **1**. In addition whole protein analysis of DHFR expressed in the presence of **2** demonstrated a mass increase of 300, with respect to DHFR expressed with **1**. This corresponds to approximately 7 analogs incorporated or an 87% replacement of phe by **2**.

*Selective modification of DHFR-**2** with biotin hydrazide*

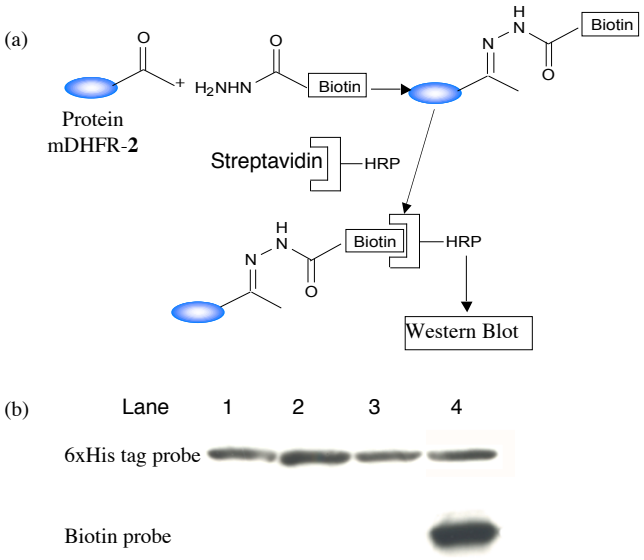
We have completed preliminary studies of the reactivity of mDHFR-**2** toward hydrazide reagents. Purified mDHFR and mDHFR-**2** were dissolved in PBS buffer (pH 6.0) and treated either with 5 mM biotin hydrazide (BH) or with PBS buffer as a negative control. The reaction products were analyzed by western blotting and visualized by treatment with a biotin-specific streptavidin-HRP conjugate (Figure 3.5). The products were also examined for the presence of the 6xHis tag of mDHFR

Figure 3.4. MALDI-TOF mass spectra of tryptic peptides derived from DHFR expressed in media supplemented with phe (a) or analog **4** (b). Two prominent mass peaks in (a) correspond to peptides 34-47 and 93-106, each containing one phe residue. Two new mass peaks are observed in (b) with a $\Delta m/z$ of 41.96 and 42.00, consistent with the increased mass of **2** relative to phe. Undigested MALDI-TOF mass spectra of DHFR expressed in media supplemented with phe (c) or analog **2** (d).



to ensure the identity of the protein band and to probe the possibility of chain cleavage under the ligation conditions. The results are consistent with chemoselective ligation without chain cleavage (Figure 3.5b).

Figure 3.5. Western blot showing chemoselective modification of ketone functionality in mDHFR. (a) Modified protein was treated with biotin hydrazide (BH), stained with HRP conjugated streptavidin and analyzed by western blot. (b) Western blot analysis of the products. Lane 1: mDHFR-wt + buffer; Lane 2: mDHFR-**2** + buffer; Lane 3: mDHFR-wt + BH; Lane 4: mDHFR-**2** + BH.



Conclusion

We describe here a new mutant form of the *E. coli* phenylalanyl-tRNA synthetase, which allows efficient *in vivo* incorporation of reactive aryl ketone functionality into recombinant proteins. This study also demonstrates the power of computational protein design in the development of aminoacyl-tRNA synthetases for activation and charging of non-natural amino acids.

References

1. Dougherty, D. A., *Unnatural amino acids as probes of protein structure and function*. Current Opinion in Chemical Biology, 2000. **4**(6): p. 645-652.

2. Behrens, C., et al., *Development of strategies for the site-specific in vivo incorporation of photoreactive amino acids: P- azidophenylalanine, p-acetylphenylalanine and benzofuranylalanine*. Tetrahedron, 2000. **56**(48): p. 9443-9449.

3. Tang, Y., et al., *Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability*. Angewandte Chemie-International Edition, 2001. **40**(8): p. 1494-+.

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

5. 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.

6. 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.

7. Ibba, M. and H. Hennecke, *Relaxing the substrate-specificity of an aminoacyl-transfer-rna synthetase allows in-vitro and in-vivo synthesis of proteins containing unnatural amino-acids*. Febs Letters, 1995. **364**(3): p. 272-275.

8. Ibba, M., P. Kast, and H. Hennecke, *Substrate-specificity is determined by amino-acid binding pocket size in Escherichia coli phenylalanyl-transfer-rna synthetase*. Biochemistry, 1994. **33**(23): p. 7107-7112.

9. Kast, P. and H. Hennecke, *Amino-acid substrate-specificity of escherichia-coli phenylalanyl-transfer rna-synthetase altered by distinct mutations*. Journal of Molecular Biology, 1991. **222**(1): p. 99-124.

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

11. Rose, K., *Facile synthesis of homogeneous artificial proteins*. Journal of the American Chemical Society, 1994. **116**(1): p. 30-33.

12. Canne, L. E., et al., *Total chemical synthesis of a unique transcription factor-related protein - cmyc-max*. Journal of the American Chemical Society, 1995. **117**(11): p. 2998-3007.

13. Rideout, D., *Self-assembling drugs - a new approach to biochemical modulation in cancer-chemotherapy*. Cancer Investigation, 1994. **12**(2): p. 189-202.

14. Rideout, D., et al., *Synergism through direct covalent bonding between agents - a strategy for rational design of chemotherapeutic combinations*. Biopolymers, 1990. **29**(1): p. 247-262.

15. Cornish, V. W., K. M. Hahn, and P. G. Schultz, *Site-specific protein modification using a ketone handle*. Journal of the American Chemical Society, 1996. **118**(34): p. 8150-8151.

16. Reshetnikova, L., et al., *Crystal structures of phenylalanyl-trna synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue*. Journal of Molecular Biology, 1999. **287**(3): p. 555-568.

17. Fishman, R., et al., *Structure at 2.6 angstrom resolution of phenylalanyl-trna synthetase complexed with phenylalanyl-adenylate in the presence of manganese*. Acta Crystallographica Section D-Biological Crystallography, 2001. **57**: p. 1534-1544.

18. Dahiyat, B. I. and S. L. Mayo, *De novo protein design: Fully automated sequence selection*. Science, 1997. **278**(5335): p. 82-87.

19. Dahiyat, B. I., C. A. Sarisky, and S. L. Mayo, *De novo protein design: Towards fully automated sequence selection*. J. Mol. Biol., 1997. **273**(4): p. 789-796.

20. 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.

21. Gordon, D. B. and S. A. Marshall, *Energy functions for protein design*. Curr. Opin. Struct. Biol., 1999. **9**(4): p. 509-513.