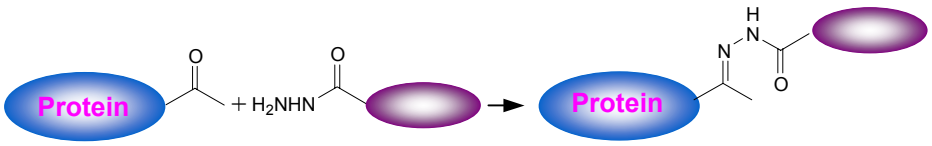
Chapter 5

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

Portions of this chapter are adapted from:
*Deepshikha Datta, Pin Wang, Isaac S. Carrico, Stephen L. Mayo and David A. Tirrell, J. Am. Soc. Chem., (**2002**), 124, 5652-5653*

Abstract

Incorporation of non-natural amino acids into proteins *in vivo* expands the scope of protein synthesis and design. *p*-Acetylphenylalanine was incorporated into recombinant dihydrofolate reductase (DHFR) in *Escherichia coli* via a computationally designed mutant form of the phenylalanyl-tRNA synthetase of the host. DHFR outfitted with ketone functionality can be chemoselectively ligated with hydrazide reagents under mild conditions.



---

1. 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 (*1-3*).

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 (*4, 5*).

In this chapter, we report a computationally designed variant of the *E. coli* phenylalanyl-tRNA synthetase (*e*PheRS), which allows efficient *in vivo* incorporation of aryl ketone functionality into proteins. In 1991, Kast and coworkers (*6-9*) introduced a variant of *e*PheRS (termed *e*PheRS*), which bears an Ala294Gly mutation in the α-subunit and which thereby acquires relaxed substrate specificity. We have recently shown that over-expression of *e*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 (*10, 11*). But similar experiments with *p*-acetylphenylalanine (**2**) failed; even in a host in which *e*PheRS* was over-expressed, phe-depleted cultures supplemented with **2** did not produce substantial yields of protein (Figure 5-2).

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 (*12-18*). Cornish and coworkers have accomplished site-specific incorporation of ketone functionality into recombinant proteins via *in vitro* translation (*19*); however, we are unaware of previous reports of *in vivo* methods of introducing ketone functionality into recombinant proteins.

2. Materials and Methods

2.1. 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'CTAGCAGTTGACAA-TTAATCATCGGCTCGTATAATGGATCGAATTGTGAGCGGAATCGATTTTCACA CAGGAAACAGACC**ATG**GATCTTCGTCGCCATCCTCGGGTCGACGTCTGTTTG CAAGCTTG-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.

2.2. 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* (*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. 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 ^{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 **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.

2.3. 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.

2.4. Tryptic Peptide Analysis

10 μL of purified protein in 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. 10 μL 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.

2.5. Chemical Modification of Protein

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 hours. 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.

3. Results and Discussion

We sought *e*PheRS mutants that would allow efficient incorporation of **2** into recombinant proteins *in vivo*. The crystal structure of *Thermus thermophilus* PheRS (*t*PheRS) complexed with **1** is available (*20, 21*) and there is 43% overall sequence identity between *e*PheRS and *t*PheRS; sequence identity in the identified active site region is 80%. We therefore employed a previously described protein design algorithm (*22*) to identify potentially useful mutants of *t*PheRS, with the intention to prepare and evaluate the corresponding mutant forms of *e*PheRS. We generated a backbone-independent rotamer library for **2**, in which both the χ_1 and χ_2 torsional angles were varied by ±20° (in increments of 5°) from the values of **1** in the *t*PheRS structure. Design calculations were performed by fixing the identity of the substrate (**2**) and by allowing each of 11 non-anchor sites in the amino acid binding pocket of *t*PheRS (determined from the crystal structure) to be occupied by any of the twenty natural amino acids except for proline, methionine and cysteine. The anchor residues (Glu128, Glu130, Trp149, His178, Ser180, Gln183, and Arg204) were held fixed in identity and conformation in all calculations. These residues contribute important electrostatic interactions with the substrate and it is reasonable to assume that such interactions are equally critical for the binding of **2**.

The calculations identified two important cavity-forming mutations: Val261 (Thr251 in *E. coli*) to Gly, and Ala314 (Ala294 in *E. coli*) to Gly (Figure 5-1). These predictions are consistent with the results of Reshetnikova and coworkers (*20, 21*), who pointed out that Ala314 and Val261 hinder the binding of amino acids larger than phe (e.g., tyrosine) into the active site of *t*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 (*10, 11*). We were thus encouraged to test whether the additional Thr251Gly mutation would relax the specificity of *e*PheRS* sufficiently to allow incorporation of **2** into proteins *in vivo*.

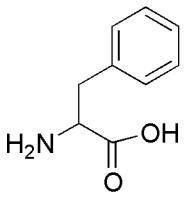
*E. coli pheS** (which encodes *e*PheRS*) was amplified by the polymerase chain reaction (PCR) from vector pQE-FS. Amplified *pheS** was subjected to PCR mutagenesis to create the coding sequence for the desired Thr251Gly mutant, which we designate *pheS****. To allow constitutive expression of the synthetase, a *tac* promoter with an abolished *lac* repressor binding site was inserted upstream of the start codon of *pheS**** (*23*). The expression cassette was then cloned into pQE15 (Qiagen), which encodes the marker protein mouse dihydrofolate reductase (mDHFR). The resulting plasmid was designated pQE-FS**. As a control, plasmid pQE-FS* (which contained *pheS** under control of a constitutive *tac* promoter) was constructed similarly. AF-IQ, a phenylalanine auxotrophic *E. coli* strain carrying the repressor plasmid pLysS-IQ (*10*), was transformed with pQE15, pQE-FS*, or pQE-FS** to generate expression systems AF-IQ[pQE15], AF-IQ[pQE-FS*], and AF-IQ[pQE-FS**], respectively. 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 5-2), neither AF-IQ[pQE15] nor AF-IQ[pQE-FS*] exhibits protein expression above background (-phe) in media supplemented with **2** (*24*). 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 yield of about 20 mg/L, approximately 60% of that obtained from cultures supplemented with **1**. MALDI-TOF mass spectrometry showed that the mass of mDHFR-**2** was increased by 304 Da, which corresponds to approximately 80% replacement of **1** by **2** (mDHFR contains 9 phe residues) (Figure 5-3). Incorporation of **2** was confirmed by tryptic digestion of mDHFR-**2** (Figure 5-4). For mDHFR, two peptides in the mass range 1550-1750 Da were assigned to residues 34-47 and 93-106, respectively (Figure 5-4a). Each fragment contains a single phe residue. The corresponding fragments of mDHFR-**2** (Figure 5-4b) were shifted up in mass by 42 Da, consistent with the increased mass of **2** relative to **1** (*25*).

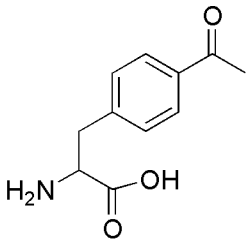
We have completed preliminary studies of the reactivity of mDHFR-**2** toward hydrazide reagents. Purified mDHFR and mDHFR-**2** were dissolved in PBS buffer (pH 6.0) and treated either with 5 mM biotin hydrazide (BH) or with PBS buffer as a negative control. The reaction products were analyzed by western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

4. Conclusion

In 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.



1



2

Figure 5-1

Active sites in *t*PheRS/Phe. The proteins are shown as a stick model and the substrate Phe is shown as a ball-and-stick model. For the proteins, yellow colored residues are involved in hydrophobic interaction with substrate while blue colored residues are involved in hydrophilic recognition toward to substrate. (a) Wild type; (b) A314G mutant; (c) T261G/A314G mutant.

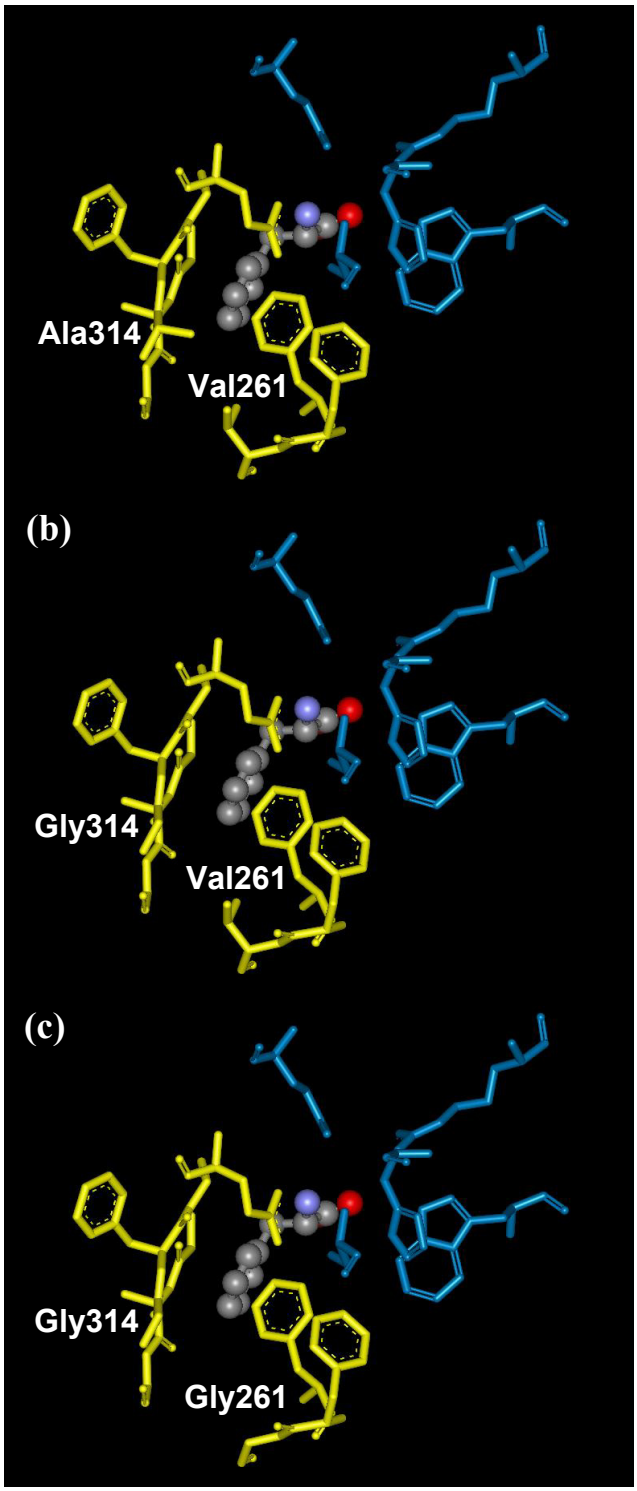

Ala314
Val261
(b)
Gly314
Val261
(c)
Gly314
Gly261

Figure 5-2

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 M: molecular weight marker (36.5, 31, 21.5, 14.4 kDa).

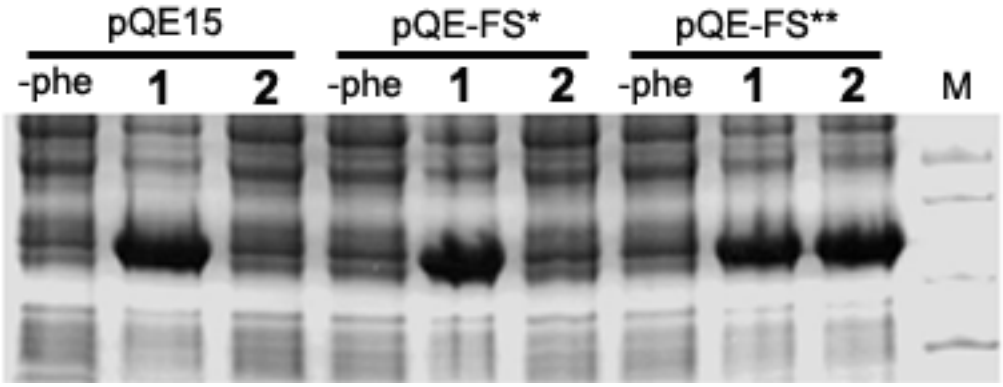

pQE15
pQE-FS*
pQE-FS**
-phe
1
2
-phe
1
2
-phe
1
2
M

Figure 5-3

Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of purified proteins containing **1** (left panel) or **2** (right panel).

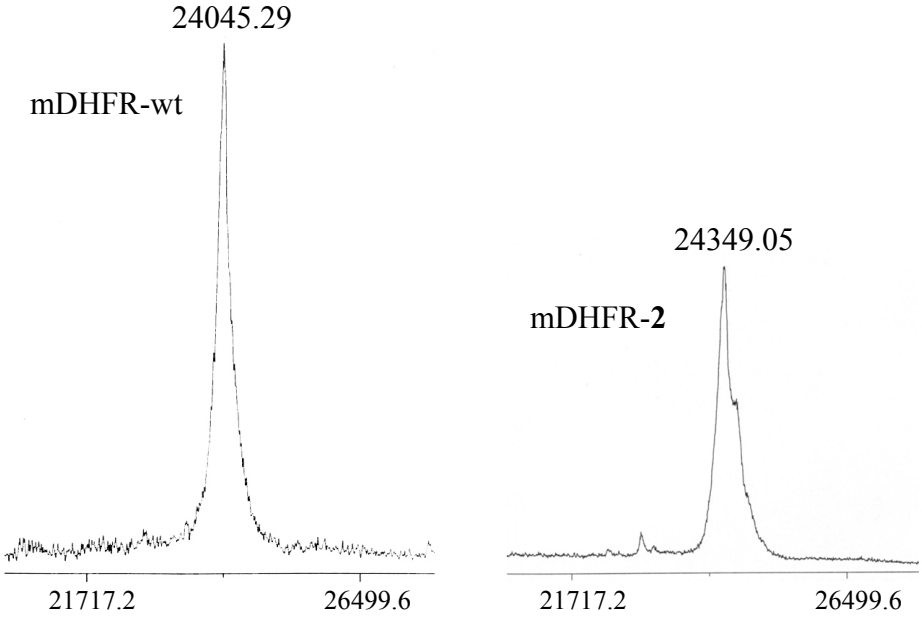

24045.29
mDHFR-wt
21717.2
26499.6
24349.05
mDHFR-2
21717.2
26499.6

Figure 5-4

MALDI TOF mass spectra of tryptic peptides digested from mDHFR expressed in media supplemented with **1** (a) or **2** (b).

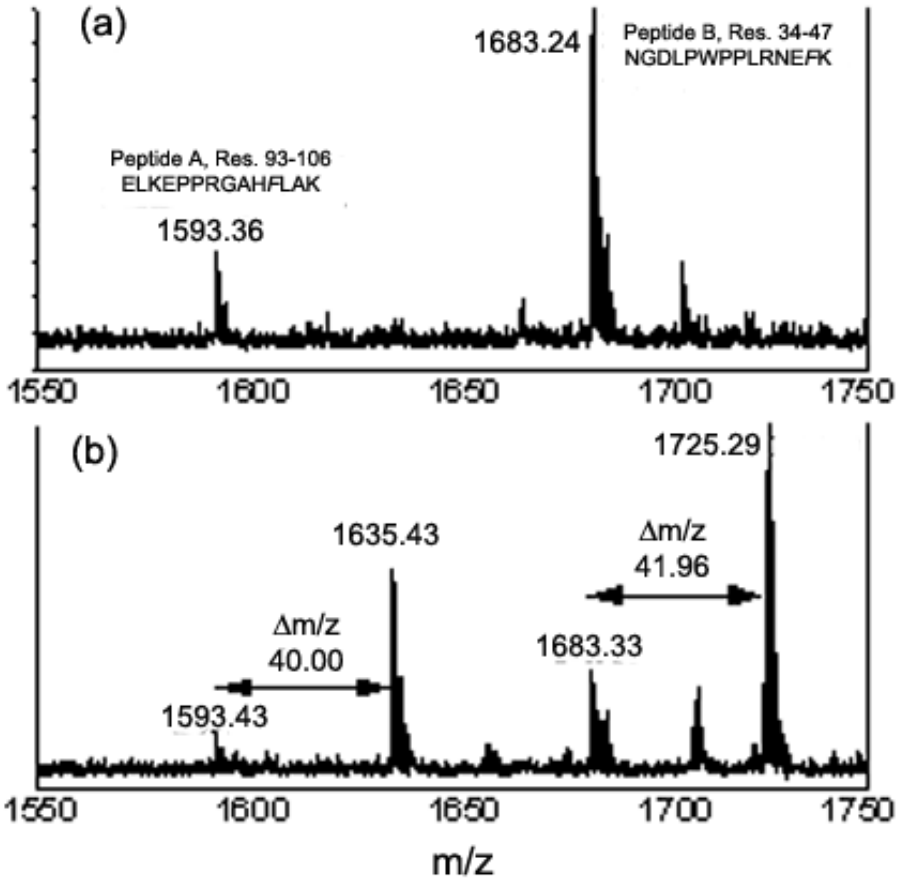
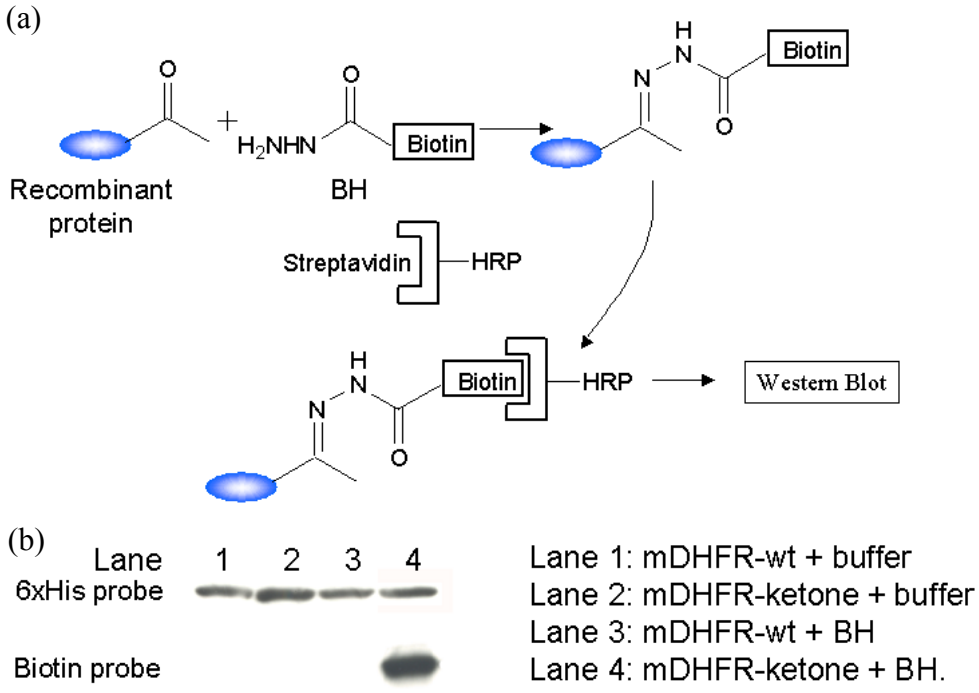

(a)
1683.24
Peptide B, Res. 34-47
NGDLPWPPLRNEFK
Peptide A, Res. 93-106
ELKEPPRGAHFLAK
1593.36
1550
1600
1650
1700
1750
(b)
1725.29
1635.43
Δm/z
41.96
Δm/z
40.00
1683.33
1593.43
1550
1600
1650
1700
1750
m/z

Figure 5-5

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.


(a)
Recombinant protein
BH
Biotin
Streptavidin
HRP
Western Blot
(b)
Lane 1 2 3 4
6xHis probe
Biotin probe
Lane 1: mDHFR-wt + buffer
Lane 2: mDHFR-ketone + buffer
Lane 3: mDHFR-wt + BH
Lane 4: mDHFR-ketone + BH.

5. References and Notes

1. Dougherty, D. A. (2000) *Current Opinion in Chemical Biology 4*, 645-652.

2. Minks, C., Alefelder, S., Moroder, L., Huber, R., and Budisa, N. (2000) *Tetrahedron 56*, 9431-9442.

3. Behrens, C., Nielsen, J. N., Fan, X. J., Doisy, X., Kim, K. H., Praetorius-Ibba, M., Nielsen, P. E., and Ibba, M. (2000) *Tetrahedron 56*, 9443-9449.

4. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angewandte Chemie-International Edition 40*, 1494-+.

5. We recognize that multi-site incorporation of amino acid analogs may in some cases result in protein misfolding and loss of function.

6. Kast, P., and Hennecke, H. (1991) *Journal of Molecular Biology 222*, 99-124.

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

8. Ibba, M., and Hennecke, H. (1995) *Febs Letters 364*, 272-275.

9. Ibba, M., Johnson, C. M., Hennecke, H., and Fersht, A. R. (1995) *Febs Letters 358*, 293-296.

10. Sharma, N., Furter, R., Kast, P., and Tirrell, D. A. (2000) *Febs Letters 467*, 37-40.

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

12. Rose, K. (1994) *J. Am. Chem. Soc. 116*, 30-33.

13. Canne, L. E., Ferre-DAmare, S. K., and Burley, S. B. H. (1995) *J. Am. Chem. Soc. 117*, 2998-3007.

14. Rideout, D., and Calogeropoulou, T. (1990) *Biopolymers 29*, 247-262.

15. Rideout, D. (1994) *Cancer Invest. 12*, 189-202.

16. Shao, J., and Tam, J. P. (1995) *J. Am. Chem. Soc. 117*, 3893-3899.

17. Mahal, L. K., Yarema, K. J., and Bertozzi, C. R. (1997) *Science 276*, 1125-1128.

18. Yarema, K. J., Mahal, L. K., Bruehl, R. E., Rodriguez, E. C., and Bertozzi, C. R. (1998) *J. Biol. Chem. 273*, 31168-31179.

19. Cornish, V. W., Hahn, K. M., and Schultz, P. G. (1996) *J. Am. Chem. Soc. 118*, 8150-8151.

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

21. Fishman, R., Ankilova, V., Moor, N., and Safro, M. (2001) *Acta Cryst. D57*, 1534-1544.

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

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

24. Phe starvation does not completely eliminate background expression, presumably because of incomplete depletion of the cellular pool of phe.

25. The enhanced promiscuity of *e*PheRS** allows misincorporation of tryptophan under phe starvation conditions in the absence of **2**. In media supplemented with **2** there is no detectable misincorporation of any of the canonical amino acids. We have not observed misincorporation of tyrosine under any conditions.