J|A|C|S
COMMUNICATIONS

Published on Web 04/26/2002

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

Deepshikha Datta, Pin Wang, Isaac S. Carrico, Stephen L. Mayo, and David A. Tirrell*

Division of Chemistry and Chemical Engineering, Howard Hughes Medical Institute, Division of Biology, Biochemistry Option, California Institute of Technology, Pasadena, California 91125

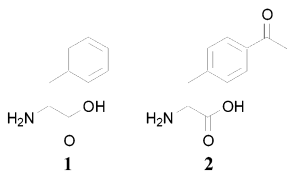
Received December 7, 2001

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 20 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]

We have previously exploited the ability of auxotrophic *Escherichia coli* strains to effect efficient incorporation of amino acid analogues into proteins in a multisite 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.[2]

In this study, 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 co-workers[3] 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 overexpression 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.[4,5] But similar experiments with *p*-acetylphenylalanine (**2**) failed; even in a host in which *e*PheRS* was overexpressed, phe-depleted cultures supplemented with **2** did not produce substantial yields of protein (Figure 1).

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.[6,7] Cornish and co-workers have accomplished site-specific incorporation of ketone functionality into recombinant proteins via in vitro translation;[8] however, we are unaware of previous reports of in vivo methods of introducing ketone functionality into recombinant proteins.



1 **2**

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[9] 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[10] to identify potentially useful mutants of

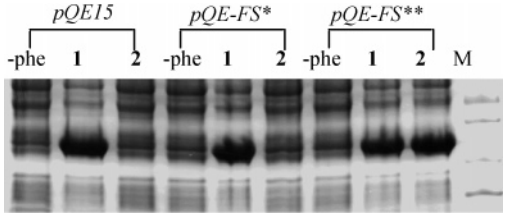


Figure 1. SDS-PAGE of cell lysates of 4 h post-induction with 1 mM IPTG. Expression plasmids and amino acid supplements are indicated. The concentration of **1** is 20 mg/L and that of **2** is 250 mg/L. Lane M: molecular weight marker (36.5, 31, 21.5, 14.4 kDa).

*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 20 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. These predictions are consistent with the results of Reshetnikova and co-workers,[9] 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.[4,5] 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.[4] 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***.[11] 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

* Corresponding author. E-mail: tirrell@caltech.edu.

10.1021/ja0177096 CCC: $22.00

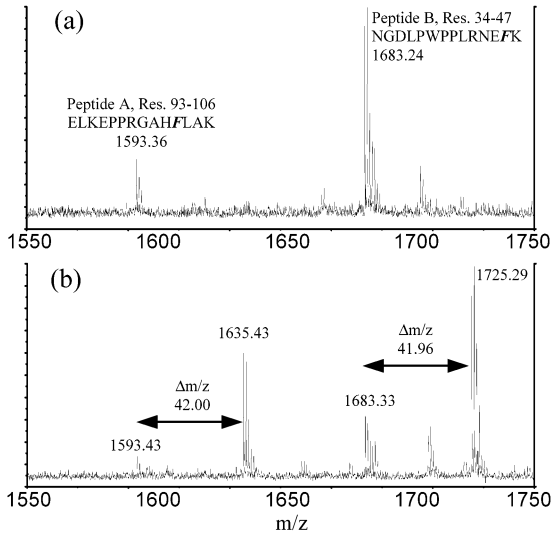


Figure 2. MALDI TOF mass spectra of tryptic peptides digested from mDHFR expressed in media supplemented with **1** (a) or **2** (b).

auxotrophic *E. coli* strain carrying the repressor plasmid pLysS-IQ,[4] 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 1), neither AF-IQ[pQE15] nor AF-IQ[pQE-FS*] exhibits protein expression above background (-phe) in media supplemented with **2**.[12] 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). Incorporation of **2** was confirmed by tryptic digestion of mDHFR-**2** (Figure 2). For mDHFR, two peptides in the mass range 1550−1750 Da were assigned to residues 34−47 and 93−106, respectively (Figure 2a). Each fragment contains a single phe residue. The corresponding fragments of mDHFR-**2** (Figure 2b) were shifted up in mass by 42 Da, consistent with the increased mass of **2** relative to **1**.[13]

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). The products were also examined for the presence of the 6×His 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 3b).

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.

Acknowledgment. The authors thank William A. Goddard III, Nagarajan Vaidehi, Kent Kirshenbaum, and Yi Tang for helpful discussions. This work was supported by NIH Grants R01-GM62523 and T32-GM08501, the NSF Center for the Science and Engineering of Materials at Caltech, the Howard Hughes Medical Institute, the Ralph M. Parsons Foundation, and an IBM shared University Research Grant.

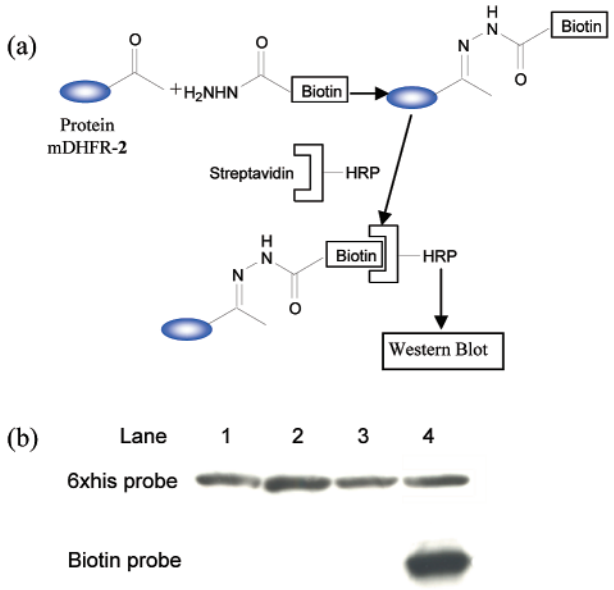


Figure 3. Western blot showing chemoselective modification of ketone functionality in mDHFR. (a) Modified protein was treated with biotin hydrazide (BH), stained with HRP conjugated streptavidin, and analyzed by western blot. (b) Western blot analysis of the products. Lane 1: mDHFR-wt + buffer. Lane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

Supporting Information Available: MALDI-MS for proteins and details of experimental procedures (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Dougherty, D. A. *Curr. Opin. Chem. Biol.* **2000**, *4*, 645−652. (b) Minks, C.; Alefelder, S.; Moroder, L.; Huber, R.; Budisa, N. *Tetrahedron* **2000**, *56*, 9431−9442. (c) Behrens, C.; Nielsen, J. N.; Fan, X. J.; Doisy, X.; Kim, K. H.; Praetorius-Ibba, M.; Nielsen, P. E.; Ibba, M. *Tetrahedron* **2000**, *56*, 9443−9449.
(2) (a) Tang, Y.; Giovanna, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 1494−1496. (b) We recognize that multisite incorporation of amino acid analogues may in some cases result in protein misfolding and loss of function.
(3) (a) Ibba, M.; Hennecke, H. *FEBS Lett.* **1995**, *364*, 272−275. (b) Ibba, M.; Kast, P.; Hennecke, H. *Biochemistry* **1994**, *33*, 7107−7112. (c) Kast, P.; Hennecke, H. *J. Mol. Biol.* **1991**, *222*, 99−124.
(4) Sharma, N.; Furter, R.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37−40.
(5) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *Chem. BioChem.* **2001**, in press.
(6) (a) Rose, K. *J. Am. Chem. Soc.* **1994**, *116*, 30−33. (b) Canne, L. E.; Ferre-DAmare, S. K.; Burley, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 2998−3007. (c) Rideout, D.; Calogeropoulou, T. *Biopolymers* **1990**, *29*, 247−262. (d) Rideout, D. *Cancer Invest.* **1994**, *12*, 189−202. (e) Shao, J.; Tam, J. P. *J. Am. Chem. Soc.* **1995**, *117*, 3893−3899.
(7) (a) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125−1128. (b) Yarema, K. J.; Mahal, L. K.; Bruehl, R. E.; Rodriguez, E. C.; Bertozzi, C. R. *J. Biol. Chem.* **1998**, *273*, 31168−31179.
(8) Cornish, V. W.; Hahn, K. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150−8151.
(9) (a) Reshetnikova, L.; Moor, N.; Lavrik, O.; Vassylyev, D. G. *J. Mol. Biol.* **1999**, *287*, 555−568. (b) Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr.* **2001**, *D57*, 1534−1544.
(10) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.
(11) Furter, R. *Protein Sci.* **1998**, *7*, 419−426.
(12) Phe starvation does not completely eliminate background expression, presumably because of incomplete depletion of the cellular pool of phe.
(13) 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.

JA0177096