Expanding the Genetic Code of *Escherichia coli*

Lei Wang,[1] Ansgar Brock,[2] Brad Herberich,[1] Peter G. Schultz[1,2]*

A unique transfer RNA (tRNA)/aminoacyl-tRNA synthetase pair has been generated that expands the number of genetically encoded amino acids in *Escherichia coli*. When introduced into *E. coli*, this pair leads to the in vivo incorporation of the synthetic amino acid *O*-methyl-L-tyrosine into protein in response to an amber nonsense codon. The fidelity of translation is greater than 99%, as determined by analysis of dihydrofolate reductase containing the unnatural amino acid. This approach should provide a general method for increasing the genetic repertoire of living cells to include a variety of amino acids with novel structural, chemical, and physical properties not found in the common 20 amino acids.

The genetic code of all organisms encodes the same 20 common amino acids. These amino acids can be modified by posttranslational modification, e.g., phosphorylation or oxidation, and in rarer instances, augmented by selenocysteine (*1*). Nonetheless, it is remarkable that polypeptides synthesized from 20 simple building blocks carry out all of the complex processes of life. Is it possible that the properties of proteins, or possibly an entire organism, could be enhanced by expanding the genetic code to include additional amino acids with novel biological, chemical, or physical properties? To begin to address this question, we developed a strategy that makes it possible to site-specifically incorporate unnatural amino acids directly into proteins in living cells. This methodology should also provide a powerful tool for analyzing protein function both in vitro and in vivo.

Unnatural amino acids can be site-specifically incorporated into proteins in vitro by the addition of chemically aminoacylated suppressor tRNAs to protein synthesis reactions programmed with a gene containing a desired amber nonsense mutation (*2–6*). One can also substitute a number of the common 20 amino acids with close structural homologs using auxotrophic strains (*7–9*). However, the addition of a new amino acid to the genetic repertoire in vivo requires additional components for the biosynthetic machinery. A new tRNA must be constructed that is not recognized by existing *E. coli* aminoacyl-tRNA synthetases, but functions efficiently in translation (an orthogonal tRNA). This tRNA must deliver the novel amino acid in response to a codon that does not encode any of the common 20 amino acids, e.g., nonsense or four base codons. The former have been used together with suppressor tRNAs in conventional protein mutagenesis (*10*); the latter can be suppressed efficiently by tRNAs containing modified anticodon loops (*11*, *12*). A new aminoacyl-tRNA synthetase (an orthogonal synthetase) is also required that aminoacylates the orthogonal tRNA, but does not recognize any of the endogenous *E. coli* tRNAs. This synthetase must aminoacylate the tRNA with only the desired unnatural amino acid and none of the common 20 amino acids. Likewise, the unnatural amino acid cannot be a substrate for the endogenous synthetases. Lastly, the amino acid, when added to the growth medium, must be efficiently transported into the cytoplasm.

An orthogonal tRNA/synthetase pair in *E. coli* can be generated by importing a pair from a different organism if cross-species aminoacylation is inefficient and the anticodon loop is not a key determinant of synthetase recognition. One such candidate pair is the tyrosyl tRNA/synthetase pair of *Methanococcus jannaschii*, an archaebacterium whose $tRNA^{Tyr}$ identity elements differ from those of *E. coli* $tRNA^{Tyr}$ (in particular, the first base pair of the acceptor stem is GC in *E. coli* and CG in *M. jannaschii*), and whose tyrosyl-tRNA synthetase (TyrRS) has only a minimalist anticodon loop binding domain (*13*). In addition, the *M. jannaschii* TyrRS does not have an editing mechanism (*14*) and, therefore, should not proofread an unnatural amino acid ligated to the tRNA. We have shown that the *M. jannaschii* TyrRS efficiently aminoacylates an amber suppressor tRNA derived from its cognate $tRNA^{Tyr}$ (*15*), but does not aminoacylate *E. coli* tRNAs (*13*). Moreover, the *M. jannaschii* $tRNA^{Tyr}_{CUA}$ is a poor substrate for the *E. coli* synthetases but functions efficiently in protein translation in *E. coli* (*15*). To further reduce recognition of the *M. jannaschii* $tRNA^{Tyr}_{CUA}$ by *E. coli* synthetases, 11 nucleotides of the tRNA that do not interact directly with the *M. jannaschii* TyrRS (C16, C17, U17a, U20, C32, G37, A38, U45, U47, A59, and U60) were randomly mutated to generate a suppressor tRNA library. This tRNA library was passed through a negative selection (suppression of amber mutations in the barnase gene), which removes tRNAs that are aminoacylated by *E. coli* synthetases, and then a positive selection for tRNAs that are efficiently aminoacylated by *M. jannaschii* TyrRS (suppression of amber mutations in the β-lactamase gene) (*16*). The orthogonal nature of the resulting suppressor tRNAs was tested by an in vivo complementation assay, which is based on suppression of an amber stop codon at a nonessential position (Ala^{184}) of the TEM-1 β-lactamase gene carried on plasmid pBLAM. Aminoacylation of a transformed suppressor tRNA by any endogenous *E. coli* synthetase results in cell growth in the presence of ampicillin. *E. coli* transformed with the *M. jannaschii* $tRNA^{Tyr}_{CUA}$ and pBLAM survive at 55 μg/ml ampicillin. When the best mutant suppressor tRNA ($mutRNA^{Tyr}_{CUA}$) selected from the library was expressed (*17*), cells survive at only 12 μg/ml ampicillin; similar values are obtained in the absence of any suppressor tRNA. When the *M. jannaschii* TyrRS is coexpressed with this $mutRNA^{Tyr}_{CUA}$, cells survive at 440 μg/ml ampicillin. Thus, the $mutRNA^{Tyr}_{CUA}$ is a poorer substrate for the endogenous synthetases than the *M. jannaschii* $tRNA^{Tyr}_{CUA}$ but is still aminoacylated efficiently by the *M. jannaschii* TyrRS.

To alter the amino acid specificity of the orthogonal TyrRS so that it charges the $mutRNA^{Tyr}_{CUA}$ with a desired unnatural amino acid, a library of TyrRS mutants was generated and screened. On the basis of the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus* (*18*), five residues (Tyr^{32}, Glu^{107}, Asp^{158}, Ile^{159}, and Leu^{162}) in the active site of *M. jannaschii* TyrRS that are within 6.5 Å of the *para* position of the aryl ring of bound tyrosine were mutated (Fig. 1) (*19*). These residues were all initially mutated to alanine, and the resulting inactive Ala_5 TyrRS was used as a template for polymerase chain reaction (PCR) random mutagenesis with doped oligonucleotides. A positive selection was then applied that is based on suppression of an amber stop codon at a nonessential position (Asp112) in the chloramphenicol acetyltransferase (CAT) gene (*20*). Cells transformed with the mutant TyrRS library and $mutRNA^{Tyr}_{CUA}$ gene were grown in media containing the unnatural amino acid and were selected for their survival in the presence of various concentrations of chloramphenicol. If a mutant TyrRS charges the orthogonal $mutRNA^{Tyr}_{CUA}$ with any amino acid, either natural or unnatural, the cell produces CAT and survives. The surviving cells were then grown

[1]Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, CA 92121, USA.

*To whom correspondence should be addressed. E-mail: schultz@scripps.edu

in the presence of chloramphenicol and the absence of the unnatural amino acid. Those cells that did not survive, i.e., those that encode mutant TyrRS's which charge the orthogonal $mutRNA^{Tyr}_{CUA}$ with an unnatural amino acid, were isolated from a replica plate supplemented with the unnatural amino acid. The mutant TyrRS genes were isolated from these cells, recombined in vitro by DNA shuffling (*21*), and transformed back into *E. coli* for further rounds of selection with increasing concentrations of chloramphenicol.

A tyrosine analog with the *para* hydroxyl group substituted with a methoxy group was used in the selection. Two rounds of selection and DNA shuffling were carried out and a clone evolved whose survival in chloramphenicol was dependent on the addition of 1 mM *O*-methyl-L-tyrosine to the growth media (*22*). In the absence of *O*-methyl-L-tyrosine, cells harboring the mutant TyrRS were not viable on minimal media plates containing 1% glycerol, 0.3 mM leucine (GMML), and 15 μg/ml of chloramphenicol. Cells were able to grow on GMML plates with 125 μg/ml chloramphenicol in the presence of 1 mM *O*-methyl-L-tyrosine. Similar results were obtained in liquid GMML. As a control, cells with the $mutRNA^{Tyr}_{CUA}$ and the inactive Ala_5 TyrRS did not survive at the lowest concentration of chloramphenicol used, either in the presence or absence of 1 mM *O*-methyl-L-tyrosine. Addition of 1 mM *O*-methyl-L-tyrosine itself does not significantly affect the growth rate of *E. coli*.

To further demonstrate that the observed phenotype is due to the site-specific incorporation of *O*-methyl-L-tyrosine by the orthogonal $mutRNA^{Tyr}_{CUA}$/mutant TyrRS pair in response to an amber stop codon, an *O*-methyl-L-tyrosine mutant of dihydrofolate reductase (DHFR) was generated and characterized. The third codon of the *E. coli* DHFR gene (a permissive site) was mutated to TAG, and a COOH-terminal His6 tag was added to separate the mutant protein from endogenous *E. coli* DHFR. As a control, the $mutRNA^{Tyr}_{CUA}$ was coexpressed with the wild-type *M. jannaschii* TyrRS, resulting in efficient suppression of the nonsense codon in DHFR with tyrosine (Fig. 2). When the mutant TyrRS was expressed in the presence of $mutRNA^{Tyr}_{CUA}$ and 1 mM *O*-methyl-L-tyrosine in liquid GMML growth media, full-length DHFR was also produced and could be purified by Ni affinity chromatography with an isolated yield of 2 mg/liter (*23*). In the absence of either *O*-methyl-L-tyrosine, $mutRNA^{Tyr}_{CUA}$ or mutant TyrRS, no DHFR (<0.1% by densitometry) was observed by analysis with SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and silver staining (Fig. 2). Western analysis further demonstrated that no trace amount of DHFR was produced in the absence of either $mutRNA^{Tyr}_{CUA}$, mutant TyrRS, or *O*-methyl-L-tyrosine (Fig. 2).

The identity of the amino acid inserted in response to the TAG codon was confirmed by mass analysis of both the intact protein and tryptic fragments. The average mass of the intact protein was determined by electrospray ionization Fourier transform–ion cyclotron resonance mass spectrometry (FT-ICR MS). The observed value for the monoisotopic mass from the cluster next to the internal calibrant was 18,096.002 daltons, which is within 5 parts per million (ppm) of the theoretical mass of 18,095.908 daltons and clearly demonstrates the incorporation of *O*-methyl-L-tyrosine (*24*). This result also indicates that other endogenous *E. coli* synthetases do not use *O*-methyl-L-tyrosine as a substrate. Liquid chromatography tandem mass spectrometry of tryptic digests was carried out to confirm the sequence of the NH_2-terminal peptide. An example of a tandem MS spectrum is shown in Fig. 3. The doubly charged precursor ion at 691.5 daltons, corresponding to the NH_2-terminal tryptic peptide MIY*MIAALAVDR (*25*), was selected and fragmented in an ion trap mass spectrometer (ITMS). The fragment ion masses could be unambiguously assigned as shown in Fig. 3, confirming the site-specific incorporation of *O*-methyl-L-tyrosine. Neither the protein mass spectra nor the tryptic peptide maps gave any indications of the incorporation of tyrosine or other amino acids in place of *O*-methyl-L-tyrosine; a minimum 95% incorporation purity for *O*-methyl-L-tyrosine was obtained from the signal-to-noise ratio of the protein mass spectra. Taken together, the cell growth, protein expression, and mass spectrometry experiments demonstrate that the $mutRNA^{Tyr}_{CUA}$/mutant TyrRS orthogonal pair is capable of selectively inserting *O*-methyl-L-tyrosine into proteins in response to the amber codon with a fidelity rivaling that of the natural amino acids. The fact that this amino acid, which is structurally similar to phenylalanine and tyrosine, is incorporated with high fidelity suggests that it will also be possible to incorporate other less homologous amino acids with high fidelity.

Analysis of the sequence of the mutant TyrRS that charges the $mutRNA^{Tyr}_{CUA}$ with *O*-methyl-L-tyrosine revealed the following mutations: $Tyr^{32} \rightarrow Gln^{32}$, $Asp^{158} \rightarrow Ala^{158}$, $Glu^{107} \rightarrow Thr^{107}$, and $Leu^{162} \rightarrow Pro^{162}$ (Fig. 1). Based on the x-ray crystal structure of the homologous *B. stearothermophilus* TyrRS, we speculate that loss of the hydrogen-bonding network between Tyr^{32}, Asp^{158}, and substrate tyrosine should disfavor binding of tyrosine to the mutant TyrRS. Indeed, mutation of Asp^{176} (which corresponds to Asp^{158} in *M. jannaschii*) of *B. stearothermophilus* TyrRS yields inactive enzyme (*26*). At the same time, the $Asp^{158} \rightarrow Ala^{158}$ and $Leu^{162} \rightarrow Pro^{162}$

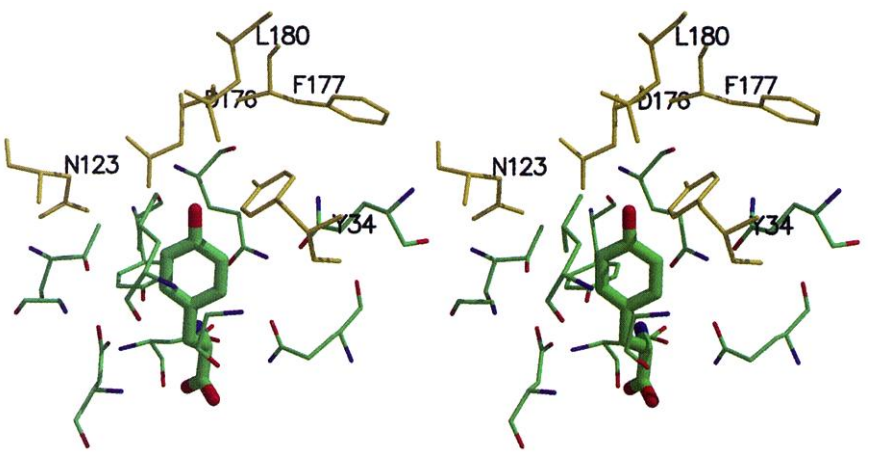

Fig. 1. Stereo view of the active site of TyrRS (*18*). Residues from *B. stearothermophilus* TyrRS are shown in the figure. Corresponding residues from *M. jannaschii* TyrRS are Tyr^{32} (Tyr^{34}), Glu^{107} (Asn^{123}), Asp^{158} (Asp^{176}), Ile^{159} (Phe^{177}), and Leu^{162} (Leu^{180}) with residues from *B. stearothermophilus* TyrRS in parenthesis; mutated residues are in yellow.

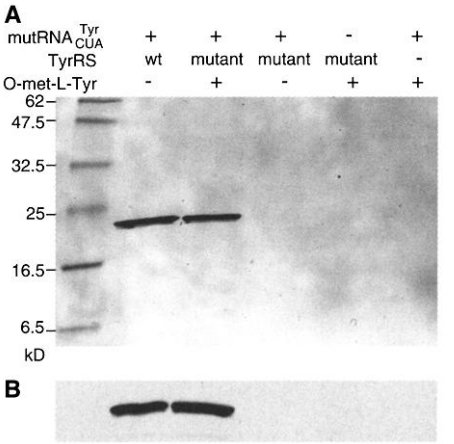

Fig. 2. Accumulation of *E. coli* DHFR protein under different conditions. (**A**) Silver-stained SDS-PAGE gel of purified DHFR. A six-histidine tag was added to the COOH terminus of *E. coli* DHFR, and protein was purified by immobilized metal affinity chromatography. Expression conditions are notated at the top of each lane. The left lane is a molecular weight marker. (**B**) Western blot of gel in (A). A penta-His antibody was used to detect the six-histidine tag at the COOH terminus of DHFR.

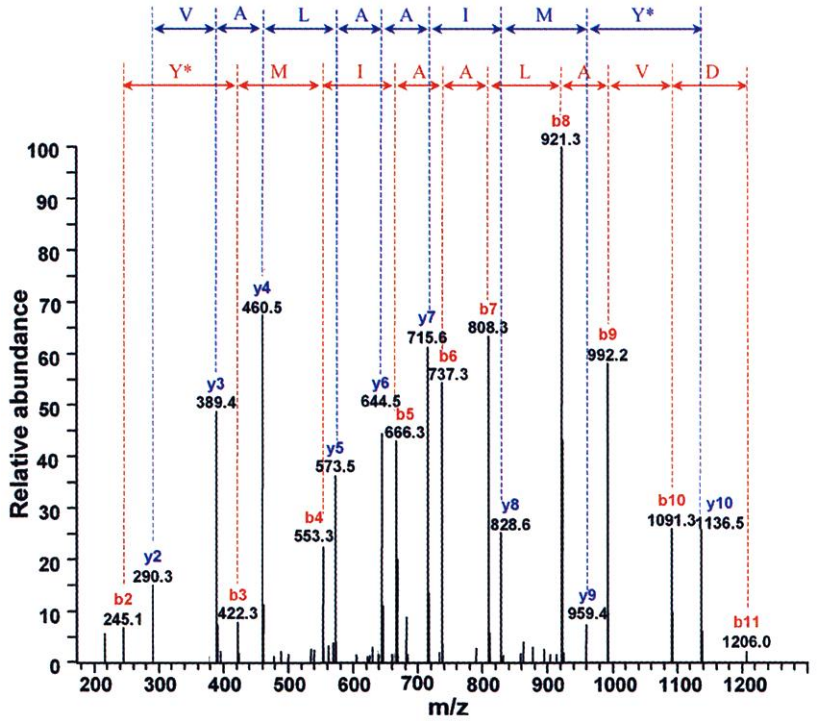

Fig. 3. Tandem mass spectrum of the NH_2-terminal peptide MIY*MIAALAVDR. The partial sequence Y*MIAALAVDR of the peptide containing the *O*-methyl-L-tyrosine residue (Y*) can be read from the annotated b (red) or y (blue) ion series.

mutations create a hydrophobic pocket that allows the methyl group of *O*-methyl-L-tyrosine to extend further into the substrate-binding cavity. Other important catalytic residues in the active site, which bind to the ribose or the phosphate group of the adenylate, were unchanged after two rounds of DNA shuffling.

Kinetics of adenylate formation of *O*-methyl-L-tyrosine and tyrosine with adenosine triphosphate (ATP) catalyzed by the mutant TyrRS were analyzed in vitro using a pyrophosphate-exchange assay at 37°C (*27*). The Michaelis constant (K_m) for tyrosine (5833 ± 902 μM) is approximately 13-fold higher than that for *O*-methyl-L-tyrosine (443 ± 93 μM), and the catalytic rate constant (k_{cat}) for tyrosine (1.8 ± 0.2 × 10^{-3} s^{-1}) is eightfold less than that for *O*-methyl-L-tyrosine (14 ± 1 × 10^{-3} s^{-1}). Thus, the value of k_{cat}/K_m of the mutant TyrRS for *O*-methyl-L-tyrosine is about 100-fold higher than that of tyrosine. The physiological concentration of tyrosine in *E. coli* is about 80 μM, which is far below K_m value (5833 μM) of the mutant TyrRS for tyrosine. Presumably, the concentration of *O*-methyl-L-tyrosine in treated cells is comparable or greater than the K_m (443 μM).

In conclusion, we have shown that it is possible to augment the protein biosynthetic machinery of *E. coli* to accommodate additional genetically encoded amino acids. The ability to introduce novel amino acids into proteins directly in living cells will provide new tools for studies of protein and cellular function and may lead to generation of proteins with enhanced properties. The methodology described here should be generalizable to other amino acids with novel spectroscopic, chemical, or structural properties. Indeed, we have shown that the *E. coli* ribosome is able to incorporate amino acids with a wide array of side chains into proteins using in vitro protein synthesis (*2*). Additional orthogonal tRNA/synthetase pairs (*28*, *29*), as well as four base codons (*11*, *12*), may further expand the number and scope of amino acids that can be incorporated. It should also be possible to generate orthogonal pairs for eukaryotic cells by a similar strategy. Lastly, this study underscores the potential created by the synergistic use of chemistry and biology in the synthesis of new molecular function.

References and Notes

1. A. Böck *et al.*, *Mol. Microbiol.* **5**, 515 (1991).
2. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **244**, 182 (1989).
3. D. Mendel, V. W. Cornish, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 435 (1995).
4. M. W. Nowak *et al.*, *Science* **268**, 439 (1995).
5. J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Diala, *J. Am. Chem. Soc.* **111**, 8013 (1989).
6. J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **356**, 537 (1992).
7. J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **122**, 1282 (2000).
8. F. Hamano-Takaku *et al.*, *J. Biol. Chem.* **275**, 40324 (2000).
9. N. Budisa *et al.*, *FASEB J.* **13**, 41 (1999).
10. M. L. Michaels, C. W. Kim, D. A. Matthews, J. H. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3957 (1990).
11. T. J. Magliery, J. C. Anderson, P. G. Schultz, *J. Mol. Biol.* **307**, 755 (2001).
12. B. Moore, B. C. Persson, C. C. Nelson, R. F. Gesteland, J. F. Atkins, *J. Mol. Biol.* **298**, 195 (2000).
13. B. A. Steer, P. Schimmel, *J. Biol. Chem.* **274**, 35601 (1999).
14. H. Jakubowski, E. Goldman, *Microbiol. Rev.* **56**, 412 (1992).
15. L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **122**, 5010 (2000).
16. L. Wang, P. G. Schultz, unpublished results.
17. The mutant suppressor tRNA contained the following nucleotide substitutions: C17A, U17aG, U20C, G37A, and U47G.
18. P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* **208**, 83 (1988).
19. The TyrRS gene was expressed under the control of *E. coli* GlnRS promoter and terminator in plasmid pBK-JYRS, a pBR322 derived plasmid with kanamycin resistance. Residues Tyr^{32}, Glu^{107}, Asp^{158}, Ile^{159}, and Leu^{162} were substituted with Ala by site-directed mutagenesis to afford plasmid pBK-JYA5. Eight oligonucleotides with NNK (N = A + T + G + C and K = G + T) at the mutation sites were used for PCR amplification of the Ala_5 TyrRS mutant (pBK-JYA5) and ligated back into the Nde I–Pst I digested pBK-JYA5 to generate the TyrRS library. The ligated vectors were transformed into *E. coli* DH10B competent cells to yield a library of 1.6 × 10^9 colony forming unit (cfu). The library was amplified by maxiprep, and supercoiled DNA was used to transform the selection strain containing pYC-J17.
20. M. Pastrnak, T. J. Magliery, P. G. Schultz, *Helv. Chim. Acta.* **83**, 2277 (2000).
21. W. P. C. Stemmer, *Nature* **370**, 389 (1994).
22. The gene encoding $mutRNA^{Tyr}_{CUA}$ was expressed in *E. coli* DH10B cells under the control of the *lpp* promoter and the *rrnC* terminator in plasmid pYC-J17, a pACYC184 derivative that also encodes the $Asp^{112}TAG$ CAT mutant. For the experimental details of the selection with *O*-methyl-L-tyrosine, see supplementary material on *Science* Online at www.sciencemag.org/cgi/content/full/292/5516/498/DC1.
23. When the $mutRNA^{Tyr}_{CUA}$ and wild-type *M. jannaschii* TyrRS are coexpressed, the yield of DHFR is 67 mg/l in 2×YT rich medium and 2.6 mg/l in liquid GMML minimal medium.
24. For this experiment a DHFR mutant lacking the COOH-terminal His tag was used and purified by methotrexate affinity chromatography. In the mutant protein, the third codon was changed to TAG, and the fourth codon was changed from CTG to ATG to improve the amber suppression efficiency, resulting in a $Leu^4 \rightarrow Met^4$ mutation.
25. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The *O*-methyl-L-tyrosine residue is denoted as Y*.
26. G. D. P. Gray, H. W. Duckworth, A. R. Fersht, *FEBS Lett.* **318**, 167 (1993).
27. The mutant TyrRS gene with six histidines at its COOH terminus was cloned into plasmid pQE-60 (Qiagen, Valencia, CA) to generate plasmid pQE-mJYRS. Protein was purified by immobilized metal affinity chromatography according to manufacturer's protocol (Qiagen). Pyrophosphate (PPi) exchange was carried out at 37°C as described (*30*). Values of k_{cat} and K_m were calculated by direct fitting of the Michaelis-Menten equation using nonlinear regression analysis.
28. D. R. Liu, P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4780 (1999).
29. A. K. Kowal, C. Köhrer, U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2268 (2001).
30. T. N. C. Wells, J. W. Knill-Jones, T. E. Gray, A. R. Fersht, *Biochemistry* **30**, 5151 (1991).

31. The authors acknowledge support of this work by the Office of Naval Research (contract 6502573) and the Skaggs Research Institute, and thank D. Liu, T. Magliery, and M. Pastrnak for their helpful suggestions. L. Wang is a graduate student in the Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720, USA.

21 February 2001; accepted 16 March 2001