J|A|C|S
ARTICLES
Published on Web 01/04/2003

Generation of a Bacterium with a 21 Amino Acid Genetic Code

Ryan A. Mehl, J. Christopher Anderson, Stephen W. Santoro, Lei Wang, Andrew B. Martin, David S. King,[†] David M. Horn,[‡] and Peter G. Schultz*[,‡]

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received September 4, 2002; E-mail: schultz@scripps.edu

Abstract: We have generated a completely autonomous bacterium with a 21 amino acid genetic code. This bacterium can biosynthesize a nonstandard amino acid from basic carbon sources and incorporate this amino acid into proteins in response to the amber nonsense codon. The biosynthetic pathway for the amino acid *p*-aminophenylalanine (*p*AF) as well as a unique *p*AF synthetase and cognate tRNA were added to *Escherichia coli*. Denaturing gel electrophoresis and mass spectrometric analysis show that *p*AF is incorporated into myoglobin with fidelity and efficiency rivaling those of the common 20 amino acids. This and other such organisms may provide an opportunity to examine the evolutionary consequences of adding new amino acids to the genetic repertoire, as well as generate proteins with new or enhanced biological functions.

Introduction

The genetic code, consisting of four nucleotides specifying 64 unique triplet codons, encodes the 20 amino acid building blocks from which all proteins are synthesized.[1] Is this the ideal number of basic building blocks or would additional amino acids lead to proteins or even organisms with enhanced function? Although various arguments have been put forth to explain the nature and number of amino acids in the code,[2−4] it is clear that proteins require additional factors to carry out many of their natural functions.[5] These groups are provided through post-translational modification and cofactors,[6,7] and in rare instances, organisms have evolved novel translational machinery to incorporate either selenocysteine or pyrrolysine in response to nonsense codons.[8−10] The need for additional chemistries in so many protein-mediated processes raises the question of why the genetic code has not evolved further.

The generation of bacteria with expanded genetic codes containing additional amino acid building blocks would provide an experimental system to address this central question of evolution. To date there have been a number of examples in which one of the common 20 amino acids has been substituted to varying extents with close analogues, including fluorine, nitrogen, and selenium substitution.[11−15] In addition, *in vitro* biosynthetic methods using chemically aminoacylated tRNAs and peptide synthesis have allowed additional amino acids to be incorporated into proteins to study protein structure and function. However, none of these methods have made it possible to genetically encode *additional* amino acids beyond the common 20 in living organisms.

Recently, we showed that one could add new components to the translational machinery of *Escherichia coli* to incorporate additional amino acids added exogenously to the growth media into proteins with high selectivity and fidelity.[16−18] We now report the construction of a completely autonomous bacterium that genetically encodes the additional amino acid *p*-aminophenylalanine (*p*AF). This bacterium has (i) the ability to synthesize *p*AF from simple carbon sources, (ii) an aminoacyl-tRNA synthetase that uniquely utilizes *p*AF and no other amino acid, and (iii) a tRNA that is acylated by this synthetase and no other, and that delivers *p*AF efficiently into proteins in response to the amber codon, TAG. The amino acid is incorporated with fidelity rivaling the common 20 amino acids.

[†] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-1460.
[‡] Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, La Jolla, CA 92037.

(1) Osawa, S. *Evolution of the Genetic Code*; Oxford University Press: Oxford, 1995.
(2) Wong, J. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1909−1912.
(3) Ronneberg, T.; Landweber, L.; Freeland, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13690−13695.
(4) Giulio, M.; Medugno, M. *J. Mol. Evol.* **1999**, *49*, 1−10.
(5) Bock, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. *Mol. Microbiol.* **1991**, *5*, 515.
(6) Lesley, S. *Drugs Pharm. Sci.* **2000**, *101*, 191−205.
(7) Krishna, R.; Wold, F.; Med, S. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1993**, *67*, 265−98.
(8) Chambers, I.; Frampton, J.; Goldfarb, P.; Affara, N.; McBain, W.; Harrison, P. *EMBO J.* **1986**, *5*, 1221−7.
(9) Zinoni, F.; Birkmann, A.; Stadtman, T.; Boeck, A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4650−4.
(10) Srinivasan, G.; James, C.; Krzycki, J. *Science* **2002**, *296*, 1459−61.
(11) Doring, V.; Mootz, H.; Nangle, L.; Hendrickson, T.; Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501−504.
(12) Ibba, M.; Hennecke, H. *FEBS Lett.* **1995**, *364*, 272−275.
(13) Lemeignan, B.; Sonigo, P.; Marliere, P. *J. Mol. Biol.* **1993**, *231*, 161−166.
(14) Kirshenbaum, K.; Carrico, I.; Terrell, D. *ChemBioChem* **2002**, 235.
(15) Bacher, J.; Ellington, A. *J. Bacteriol.* **2001**, *183*, 5414−5425.
(16) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500.
(17) Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1−10.
(18) Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836.

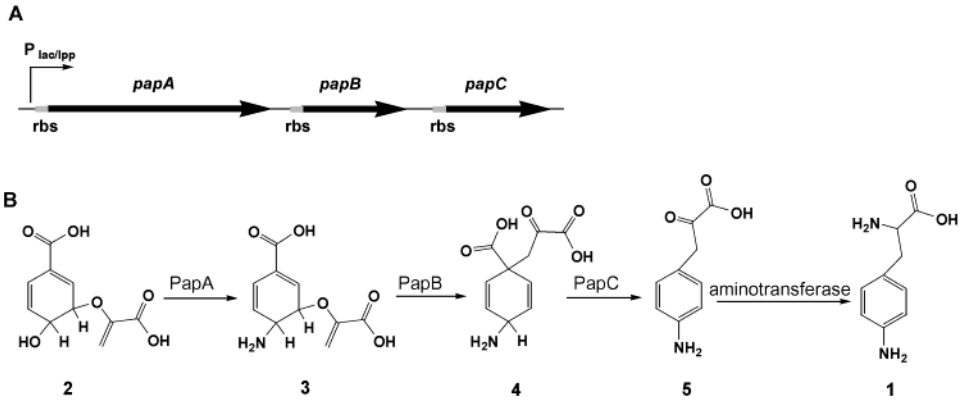

Figure 1. *S. venezuelae* genes *papA*, *papB*, and *papC* cloned into pSC101-derived pLASC plasmid under control of the lac or lpp promotor (A). Proteins PapA, PapB, and PapC convert chorismate **2** to *p*-aminophenylpyruvic acid (**5**), while *E. coli* aromatic aminotransferase completes the biosynthesis to afford *p*-AF (**1**) (B).

Results and Discussion

Selection of a *p*AF Synthetase. For the purposes of these experiments, we chose the nonproteinogenic amino acid *p*-aminophenylalanine (**1**, *p*AF), on the basis of its interesting physical properties (e.g., π-donating effects, hydrogen-bonding properties, and weak basicity), stability, and lack of toxicity to *E. coli*, and the fact that it is a known secondary metabolite. Moreover, the genes that lead to the production of *p*AF as a metabolic intermediate in the production of chloramphenicol and pristinamycin have been identified in *Streptomyces venezuelae* and *Streptomyces pristinaespiralis*, respectively.[19,20]

To generate an orthogonal tRNA−synthetase pair for *p*AF, the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS) and mutant tyrosine amber suppressor tRNA ($mutRNA_{CUA}^{Tyr}$) pair was used as a starting point.[21,22] Natural and synthetic amber codon suppressors have been extensively used to mutate proteins with little negative effect on cell growth.[23,24] The *p*AF-specific synthetase (*p*AFRS) was generated by modifying the amino acid specificity of the *M. jannaschii* TyrRS to accept *p*AF and not any of the common 20 amino acids. A combination of positive and negative selections[25] was used to identify the *p*AFRS enzyme from a library of TyrRS variants containing random amino acids at five positions (Tyr^{32}, Glu^{107}, Asp^{158}, Ile^{159}, and Leu^{162}).[16]

A single reporter plasmid was used for both selection and screening. The positive selection is based on suppression of a TAG codon at a permissive position within the chloramphenicol acetyltransferase (CAT) gene[16,26] by either *p*AF or an endogenous amino acid. Cells containing the TyrRS library and reporter plasmid were grown in liquid culture containing *p*AF and selected for survival in the presence of chloramphenicol (Cm). The negative screen is based on the inability to suppress in the absence of *p*AF two TAG stop codons at permissive positions within the T7 RNA polymerase gene. Production of full-length T7 RNA polymerase drives production of a variant of green fluorescent protein (GFPuv). Cells from the positive selection were grown in the absence of *p*AF and Cm, and then screened using fluorescence-activated cell sorting for a lack of fluorescence. The desired *p*AFRS was identified following two rounds of positive selection in liquid media, one round of negative screening, another round of positive selection in liquid media, and one round of positive selection on plates. The *p*AFRS enzyme contains five mutations relative to the wild-type TyrRS (Y32T, E107T, D158P, I159L, and L162A). In the absence of *p*AF, the IC_{50} of cells containing the selected *p*AFRS and reporter plasmid was 10 μg/mL Cm. The IC_{50} was 120 μg/mL Cm with 1 mM *p*AF. This result suggests that *p*AF is necessary for suppression of the TAG codon.

Engineering Nonproteinogenic Amino Acid Production. We anticipated that *p*AF could be synthesized in *E. coli* from chorismate **2** (a biosynthetic intermediate in the synthesis of aromatic amino acids) using the *S. venezuelae* enzymes PapA, PapB, and PapC, together with an *E. coli* aminotransferase (Figure 1B). PapA, 4-amino-4-deoxychorismate synthase, is expected to convert **2** to 4-amino-4-deoxychorismic acid (**3**) using ammonia (from glutamine) in a simple addition−elimination reaction. PapB and PapC, which are analogous to chorismate mutase and prephenate dehydrogenase, respectively, should convert **3** to 4-amino-4-deoxyprephenic acid (**4**) and then to **5**. The nonspecific tyrosine aminotransferases from *E. coli* (*tyrB*, *aspS*, and *ilvE*) have been shown to utilize many ketoacids and were expected to convert **5** to *p*AF.[27]

It was difficult to predict how the synthesis of another aromatic amino acid from the heavily used chorismate branch point would affect *E. coli* growth. *p*-Aminophenylalanine must be produced in a concentration sufficient for efficient protein biosynthesis, but not to such a degree as to affect the concentration of the other aromatic amino acids or to exhaust cellular resources. In *S. venezuelae* there is evidence to suggest that the regulation of the shikimate pathway is modified to account for chorismate consumption in making a fourth aromatic amino acid.[28] The availability of an orthogonal *p*AFRS/$mutRNA_{CUA}^{Tyr}$

(19) Yanai, K.; et al. *PCT Int. Appl.* **2001**, 1−83.
(20) Blanc, V.; Gil, P.; Bamas-Jacques, N.; Lorenzon, S.; Zagorec, M.; Schleuinger, J.; Bisch, D.; Blanche, F.; Debussche, L.; Crouzet, J.; Thibaut, D. *Mol. Microbiol.* **1997**, *23*, 191−202.
(21) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010−5011.
(22) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883.
(23) Kleina, L.; Masson, J.; Normanly, J.; Abelson, J.; Miller, J. *J. Mol. Biol.* **1990**, *213*, 705−717.
(24) Normanly, J.; Kleina, L.; Masson, J.; Abelson, J.; Miller, J. *J. Mol. Biol.* **1990**, *213*, 719−726.
(25) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*(10), 1044−1048.
(26) Pasternak, M.; Magliery, T. J.; Schultz, P. G. *Helv. Chim. Acta* **2000**, *83*, 2277.
(27) Neidhardt, F. C. *Escherichia coli and Salmonella*, 2nd ed.; ASM Press: Washington, DC, 1996; Vol. 1.
(28) He, J.; Magarvey, N.; Piraee, M.; Vining, L. C. *Microbiology* **2001**, *147*, 2817−2829.

Table 1. Cellular Concentrations (mM) of Aromatic Amino Acids in *E. coli* [a]

amino acid	pLASC	pLASC + exogenous *p*-AF	pLASC-lacPW	pLASC-lppPW
*p*AF	0.0	2.5	0.1	0.7
Tyr	0.8	0.8	0.9	0.9
Trp	0.2	0.2	0.2	0.2
Phe	0.3	0.4	0.4	0.4

[a] Cells without the *p*AF biosynthetic pathway (*p*LASC), cells with the pathway under the control of the lac promoter (*p*LASC-lacPW), and cells with the pathway under the control of the strong lpp promoter (*p*LASC-lppPW) were grown in minimal media. Amino acids were extracted by toluenization, separated by HPLC, and detected by mass spectrometry.

pair that can selectively incorporate *p*AF into proteins should allow one to optimize the *in vivo* production of *p*AF. Only cells that produce a concentration of *p*AF sufficient for protein biosynthesis will be able to suppress the amber codon in the CAT gene and survive on Cm, or suppress the two TAG codons in the T7 RNA polymerase gene and produce GFPuv. One can use these suppression assays to monitor *p*AF biosynthesis under different *E. coli* growth conditions and select for optimal *p*AF production on the basis of *E. coli* growth rates.

To incorporate the *p*AF biosynthetic pathway into *E. coli*, the *papABC* genes from *S. venezuelae* were cloned into a low-copy pSC101-derived plasmid (Figure 1A). The gene cassettes were placed under the control of the constitutive lpp promoter, or the inducible lac promoter, and each gene was placed behind an *E. coli* ribosome binding site (rbs). The ability of the cells to generate *p*AF and suppress a TAG stop codon was first monitored by observing fluorescence due to the production of GFPuv. Both the lpp promotor-controlled *p*AF biosynthetic pathway (lppPW) and the addition of 1 mM exogenous *p*AF produced 120-fold higher cellular fluorescence than the background (fluorescence when no *p*AF is available to the GFPuv reporter system), while the lac promotor-controlled *p*AF biosynthetic pathway (lacPW) without IPTG induction produced a 3.2-fold increase in cellular fluorescence. The GFPuv suppression assay shows that biosynthesis of *p*AF by the lppPW is comparable to that by exogenous 1 mM *p*AF, and significantly better than that by the noninduced lacPW.

To determine whether the addition of the *p*AF biosynthetic pathway affects the production of other aromatic amino acids in *E. coli*, and to quantify *p*AF production, the cellular concentrations of the aromatic amino acids were monitored. Cells with lppPW or lacPW were compared to those without the biosynthetic pathway, grown either in the presence or absence of exogenous *p*AF. The cellular amino acid concentrations were determined by extracting cells at midlog growth followed by HPLC−MS analysis. The concentrations of the natural aromatic amino acids remained unchanged under all conditions examined, indicating that neither the presence of the *p*AF pathway nor exogenous *p*AF significantly affects amino acid production (Table 1). While both pathways produce *p*AF in concentrations similar to those of the natural aromatic amino acids, lppPW produces 7 times as much *p*AF (0.7 mM) as the noninduced lacPW (0.1 mM).

Production of *p*AF-Containing Protein by an Engineered Bacterium. To confirm that the *p*AF biosynthetic pathway and mutRNA$_{CUA}^{Tyr}$/*p*AFRS pair lead to incorporation of *p*AF (and only *p*AF) into proteins only in response to TAG codons, protein containing *p*AF was expressed, purified, and analyzed. The

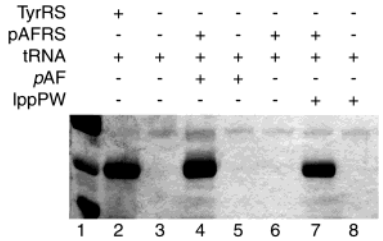

Figure 2. SDS−PAGE analysis of the incorporation of *p*AF into sperm whale myoglobin by *p*AFRS and mutRNA$_{CUA}^{Tyr}$ under different conditions. Silver-stained gel of purified myoglobin by immobilized metal affinity chromatography. Protein production conditions are noted at the top of each lane. Lane 1 is a molecular weight marker.

fourth codon of the sperm whale myoglobin gene was converted to a TAG stop codon, and a COOH-terminal His6-tag was added to facilitate purification. When *p*AFRS was coexpressed with the mutRNA$_{CUA}^{Tyr}$ in the presence of 1 mM *p*AF, or the genes coding for the amino acid biosynthetic pathway (lppPW or lacPW), full-length myoglobin was produced (Figure 2). No myoglobin was observed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) and silver staining in the absence of *p*AFRS or a *p*AF source, indicating that only *p*AF is inserted at the fourth position of myoglobin. Coproduction of the mutant synthetase with lppPW, lacPW, or exogenously added *p*AF yielded 1.8, 0.2, and 3.0 mg/L mutant protein in liquid GMML, respectively. To compare the efficiency of *p*AFRS and the biosynthetic pathway, the wild-type TyrRS was used to make 4-Tyr-myoglobin, and 2.2 mg/L protein was produced. Wild-type myoglobin was produced and purified under the same conditions as 4-*p*AF-myoglobin to determine the suppression efficiency of the *p*AFRS. The yield of wild-type myoglobin was 3.5 mg/L, comparable to the 3.0 mg/L yield when *p*AFRS was used to suppress the amber codon in the fourth position of myoglobin. This indicates that amber suppression with the *p*AFRS is not the rate-limiting step in protein production.

To further demonstrate that only *p*AF is inserted in response to the TAG codon, the purified myoglobin protein was analyzed by N-terminal sequencing. Edman degradation showed that the fourth position of 4-*p*AF-myoglobin was occupied by only a nonstandard amino acid and none of the 20 natural amino acids within the limits of detection (10−20%). To verify that the fourth amino acid of the modified myoglobin is *p*AF, the protein mass was analyzed by electrospray-ionization ion trap mass spectrometry. The average protein mass resulting from exogenously added *p*AF or biosynthetically synthesized *p*AF was 18430.01 (SD 0.56) or 18430.00 (SD 0.45) Da, respectively (calculated mass 18430.2 Da). If the wild-type TyrRS is used with the mutRNA$_{CUA}^{Tyr}$, then tyrosine is inserted instead of *p*AF, and a mass shift of 1 Da higher is observed. To further assess whether this mass shift of 1 Da was due to the discrete incorporation of biosynthesized *p*AF into the fourth position of myoglobin, isolated proteins were digested with trypsin and the fragments were analyzed by MALDI-FT-ICR (MS). The sequence coverages of 4-*p*AF-myoglobin and 4-Tyr-myoglobin were greater than 85% and 79%, respectively, with a 1 Da shift detectable only in the first 17 amino acid peptide fragment (MVL(*p*AF)EGWLVLHVWAK). The monoisotopic singly protonated peptide mass for the 4-*p*AF-containing protein was 2100.094 ppm (theoretical M + H, 2100.115 ppm) with a 2 ppm error in measured mass, while that of the 4-Tyr-containing

peptide fragment was 2101.096 ppm (theoretical M + H, 2101.094 ppm) with a 1 ppm error in the measured mass. These data, together with the SDS−PAGE analysis, conclusively show that *p*AF is being incorporated with high fidelity and efficiency.

Conclusion

We have generated for the first time a completely autonomous organism that genetically encodes a novel amino acid that can be incorporated into proteins efficiently and with high fidelity in a manner analogous to that of the 20 standard amino acids. Although the additional genes are encoded on low copy plasmids, it is likely that they could be incorporated directly into the *E. coli* genome. Moreover, there are no observable toxic side effects due to amber codon suppression, as expected from previous studies.[23,24] Efforts are in progress to generate a completely unique codon for *p*AF by using four-base codons to encode additional amino acids. However, with the current version of the *p*AF bacterium, we can now begin to compare its ability to adapt to selective pressure to that of its 20 amino acid ancestor. Moreover, by evolving this bacterium and others like it in response to selective pressures, we hope to determine whether the addition of novel amino acids to the genetic code can provide an evolutionary advantage to *E. coli*.

Experimental Section

Evolution of a Mutant Synthetase Specific for *p*-Aminophenylalanine (pAFRS).[25] The *p*AF amino acid was purchased from Sigma. The starting library consists of *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) variants, randomized at positions Tyr^{32}, Glu^{107}, Asp^{158}, Ile^{159}, and Leu^{162},[16] and inserted into plasmid pBK-JYRS, which contains the ColE1 origin of replication and a kanamycin (Kn) resistance marker. The reporter plasmid pREP/YC-JYCUA[25] contains the genes for chloramphenicol acetyltransferase (CAT), T7 RNA polymerase (T7 RNAP), a variant of green fluorescent protein (GFPuv), a mutant tyrosine amber suppressor tRNA ($mutRNA_{CUA}^{Tyr}$), and a selectable marker for tetracycline (Tet) resistance. The CAT gene contains an amber codon substitution at position D112 (12). The T7 RNAP gene contains a seven amino acid N-terminal leader peptide and amber substitutions at positions M1 and Q107. Production of full-length T7 RNAP drives expression of the GFPuv gene. Library plasmids were electroporated into *E. coli* DH10B cells (Life Technologies) containing plasmid pREP/YC-JYCUA, and transformants were recovered in SOC medium for 60 min at 37 °C and grown to saturation in Luria−Bertani medium (LB). For the positive selection, cells were grown to saturation at 37 °C (∼24−36 h) in GMML minimal media (1% glycerol, 0.3 mM leucine) containing 35 μg/mL Kn, 25 μg/mL Tet, 75 μg/mL chloramphenicol (Cm), and 1 mM *p*AF. For the negative screen, cells were grown to saturation at 37 °C (∼24 h) in GMML media containing 35 μg/mL Kn, 25 μg/mL Tet, and 0.002% arabinose (Ara). Arabinose-induced cells grown in the absence of unnatural amino acids (1 mL) were pelleted and resuspended in 3 mL of phosphate-buffered saline (PBS). Cells were sorted for lack of fluorescence using a BDIS FACVantage TSO cell sorter with a Coherent Enterprise II ion laser with excitation at 351 nm and emission detected using a 575/25 nm band-pass filter. Collected cells were diluted in 10 volumes of LB containing Tet and Kn and grown to saturation. The libraries were subjected to two cycles of positive selection, followed by one cycle of negative screening, followed by a third cycle of positive selection. Following the third positive selection, cells were plated on GMML/agar containing Tet, Kn, 0.002% Ara, 0, 75, or 100 μg/mL Cm, and 0 or 1 mM *p*AF, and grown for 48 h at 37 °C. Only one synthetase (*p*AFRS), identified by sequencing plasmid DNA from individual colonies, was present following the positive selection on plates. The selectivity of *p*AFRS was assayed by growing cells containing the *p*AFRS and reporter plasmid on GMML minimal media plates with increasing concentrations of Cm, with or without *p*AF.

Addition of the *p*AF Biosynthetic Pathway. The *papA*, *papB*, and *papC* genes were amplified by PCR from *S. venezuelae* (ATCC 10712) genomic DNA using the following primers: 5′-TTCACACACAGGAAACAGCTATGCGCACGCTTCTGATCGAC-3′ forward, 5′-ATGCTATGGTCCTTGTTGGTGTCATCGTGCGCCCGCCACTGC-3′ reverse (*papA*); 5′-CCGTCATGTACTAAGGAGGTTGTATGAGTGGCTTCCCCCGGAGCGTCG-3′ forward, 5′-CTATAGTGTCACCTAAATTCATCGGTCCTTCTCGCCTTCG-3′ reverse (*papB*); 5′-CACCAACAAGGACCATAGCATATGACCGAGCAGAACGAGCTG-3′ forward, 5′-ACAACCTCCTTAGTACATGACGGGCTCATCACCAGGTCCTCG-3′ reverse (*papC*). Genes *papABC* were assembled by overlap PCR using the three cloned genes and primers 5′-CCGTTGAATTCACACACAGGAAACAGC-3′ forward and 5′-GGAAAGGATC CCTATAGTGTCACCTAAAT-3′ reverse. The *papABC* fragment digested with *Bam*HI and *Eco*RI was inserted into a pSC101-derived plasmid, pLASC, and maintained by ampicillin (Amp) selection. Ribosome binding sites were derived from the 5′-UTR of LacZ (5′-CACACAGGAAACAGCT-3′), malE (5′-CACCAACAAGGACCATAGCA-3′), and cro (5′-ATGTACTAAGGAGGTTGT-3′) and placed prior to *papA*, *papB*, and *papC*, respectively. The *papABC* genes were placed under control of lac and lpp promotors to afford two pathway plasmids: pLASC-lacPW and pLASC-lppPW (as shown Figure 1A).

Testing *p*AF Biosynthesis with pAFRS. *E. coli* DH10B cells harboring three plasmids, the reporter plasmid pREP/YC-JYCUA, the synthetase plasmid *p*AFRS, and the pathway plasmid pLASC-lacPW or pLASC-lppPW, were grown to saturation in GMML containing 110 μg/mL Amp (the pLASC plasmid was used in place of the pathway plasmid for the background, no *p*AF, and 1 mM exogenous *p*AF trials). DH10B was grown with no plasmids to determine the background suppression level of the reporter plasmid. A sample of each cell growth was equalized to an OD of 1.0 (600 nm) with water, washed, and diluted 5-fold with PBS. Cells suspended in PBS were analyzed using a Fluoromax-2 fluorescence detector (the excitation wavelength was 351 nm, and a peak emission at 505 nm was monitored). DH10B produced 1.0×10^4 fluorescence units, while background fluorescence (no *p*AF added) from the reporter system produced 2.5×10^4 fluorescence units. The lacPW, lppPW, and 1 mM exogenously added *p*AF produced 7.9×10^4, 3.0×10^6, and 3.0×10^6 fluorescence units, respectively. Induction of the lacPW with IPTG was not feasible due its inhibitory affect on the arabinose promotor in the reporter plasmid pREP/YC-JYCUA.

Aromatic Amino Acid Concentration. *E. coli* DH10B cells harboring the pLASC plasmid, pLASC-lacPW, or pLASC-lppPW were grown to saturation in GMML minimal media containing 110 μg/mL Amp. Cells grown with exogenously added *p*AF contained 1 mM amino acid at the start of the growth. Cells (100 mL, grown to saturation) were harvested by centrifugation and washed (1 mL of water), and 1 mL of water and 0.2 mL of toluene were added. The cell pellet was suspended in the extraction mixture, shaken at 37 °C for 30 min, and then separated by centrifugation. The aqueous layer was filtered (Microcon YM-10) and analyzed by HPLC−MS (Agilent 1100): 5−15 μL of the aqueous layer was separated on a Zorbax SB-C18 column (5 μm, 4.6 × 150 mm) with a gradient of aqueous 0.1% TFA/acetonitrile 0.1% TFA (95:5 to 5:95) over 10 min. Amino acids were identified by extracting their MW (+1) from the total ion mass spectrum. Comparison of the extracted ion area of each amino acid by standard addition allowed the calculation of amino acid concentrations in each sample. Adding *p*-methoxyphenylalanine as an internal standard to the extraction mixture allowed for the calculation of relative concentrations of other amino acids in solution. Along with *p*-methoxyphenylalanine, one can add varying concentrations of *p*AF, tyrosine, tryptophan, and phenylalanine to the extraction mixture and effectively back-calculate the original concentration of the amino acids present. The error in the amino acid concentration for *p*AF and tryptophan is ±0.1 mM, while for

tyrosine and phenylalanine it is ±0.2 mM. Cellular concentrations were based on the amount of water in the cell pellet (assuming the hydrated pellet was 70% water by mass).

Production of Protein Containing *p*AF. The plasmid pBAD/JYAMB-4TAG (Tet resistance) was used to express the $mutRNA_{CUA}^{Tyr}$ gene under the control of the lpp promotor and rrnC terminator, and the sperm whale myoglobin gene (with an amber stop codon at Ser4) under the control of the arabinose promotor and rrnB terminator. The pBAD/JYAMB plasmid was used to express the wild-type myoglobin gene (with no amber stop codon) to compare protein yields to determine suppression efficiency. A his6-tag was added to the carboxy terminus of myoglobin. The TyrRS and *p*AFRS genes were expressed under the control of the *E. coli* GlnRS promotor and terminator on a pBR322-derived plasmid with Kn resistance. The *papABC* genes were expressed from pLASC-lacPW or pLASC-lppPW under the control of the native terminator. *E. coli* DH10B cells harboring the plasmid pBAD/JYAMB-4TAG, pBK-TyrRS, or pBK-*p*AFRS and a pLASC-derived plasmid (pLASC, pLASC-lacPW, or pLASC-lppPW as indicated) were grown in 0.5 L of GMML minimal media containing 5 μM $FeCl_3$. Protein production trials with exogenous *p*AF contained a final concentration of 1 mM *p*AF. For all trials, arabinose was added to a final concentration of 0.002% when the cells reached an OD of 0.3−0.4. The cultures were all grown to saturation (20−30 h) in parallel at 37 °C and pelleted, and protein was purified by Ni^{2+} affinity chromatography according to the manufacturer's protocol under native conditions (Qiagen, Valencia, CA). A 15 μL sample of the final protein solution (3.5 mL) from each preparation was separated on a 12% SDS−polyacrylamide gel and quantified by silver staining.

Characterization of Myoglobin Samples. All myoglobin samples, with *p*AF in the fourth position (4-*p*AF-myoglobin), with tyrosine in the fourth position (4-Tyr-myoglobin), and wild-type myoglobin with serine in the fourth position, were purified prior to analysis. Protein samples for Edman sequencing, trypsin digests, and mass spectrometry were purified first by Ni^{2+} affinity chromatography and then by HPLC (Dynamax). Desalted myoglobin samples from affinity chromatography were separated on a C4 column (5 μm, 4.6 × 150 mm) with a gradient of aqueous 0.1% TFA/acetonitrile 0.1% TFA (80:20 to 10:90) over 30 min. Samples were lyophilized and used for analysis.

Acknowledgment. We acknowledge support of this work by the Department of Energy (Contract No. DE-FG03-00ER45812). J.C.A. is an NSF predoctoral fellow. S.W.S. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. This is Manuscript No. 14864-CH of The Scripps Research Institute.

Supporting Information Available: MALDI-FT-ICR (MS) analysis of the N-terminal tryptic fragments of 4-Tyr-myoglobin and 4-*p*AF-myoglobin (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0284153