

Selection and Characterization of *Escherichia coli* Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue

JAMIE M. BACHER¹ AND ANDREW D. ELLINGTON^{1,2*}

*Institute for Cellular and Molecular Biology¹ and Department of Chemistry and Biochemistry,²
University of Texas at Austin, Austin, Texas 78712*

Received 20 December 2000/Accepted 6 June 2001

***Escherichia coli* isolates that were tolerant of incorporation of high proportions of 4-fluorotryptophan were evolved by serial growth. The resultant strain still preferred tryptophan for growth but showed improved growth relative to the parental strain on other tryptophan analogues. Evolved clones fully substituted fluorotryptophan for tryptophan in their proteomes within the limits of mass spectral and amino acid analyses. Of the genes sequenced, many genes were found to be unaltered in the evolved strain; however, three genes encoding enzymes involved in tryptophan uptake and utilization were altered: the aromatic amino acid permease (*aroP*) and tryptophanyl-tRNA synthetase (*trpS*) contained several amino acid substitutions, and the tyrosine repressor (*tyrR*) had a nonsense mutation. While kinetic analysis of the tryptophanyl-tRNA synthetase suggests discrimination against 4-fluorotryptophan, an analysis of the incorporation and growth patterns of the evolved bacteria suggest that other mutations also aid in the adaptation to the tryptophan analogue. These results suggest that the incorporation of unnatural amino acids into organismal proteomes may be possible but that extensive evolution may be required to reoptimize proteins and metabolism to accommodate such analogues.**

Cellular growth in the presence of unnatural amino acids often results in at least the partial incorporation of the analogues into proteins. For example, although 4-fluorotryptophan (4fW) is eventually toxic to cells, bacteria can grow for several generations in its presence and proteins that contain high levels of this analogue can be isolated (5, 29). Similarly, *p*-Cl-phenylalanine has been shown to replace phenylalanine *in vivo*, but it drastically decreases luciferase activity (14). Azaleucine can also be partially incorporated in place of arginine and has been shown to suppress an otherwise lethal mutation in thymidylate synthase (20). It has recently been shown that a tRNA synthetase with compromised editing function will allow the replacement of valine with cysteine, threonine, or α -aminobutyrate (9). Similarly, a tyrosyl-tRNA synthetase mutant has been isolated which discriminates against azatyrosine less readily than the wild type (13). All of these examples demonstrate that the partial substitution of one amino acid for another is possible.

Current methods for unnatural amino acid incorporation *in vivo* involve either the site-specific incorporation of an unnatural amino acid into a few proteins (11) or the partial incorporation into many proteins. In contrast, we and Wong (37) have attempted to incorporate unnatural amino acids throughout an organismal proteome. Such attempts to develop unnatural organisms should enhance our understanding of the extent to which even the most basic biochemistry is evolvable or optimal and may suggest metabolic alternatives for engineered organisms. In order to explore these possibilities, we attempted to evolve *Escherichia coli* variants which could survive on 4fW, which is otherwise toxic (5, 29).

Evolutionary approaches have the drawback that diverse genetic and metabolic possibilities are available to organisms

under selection, and it was not clear in advance which evolutionary path would be taken. Based on previous experiments by Wong (37), we hypothesized that a relatively small number of proteins which were adversely affected by 4fW would mutate to accommodate the analogue. The fluorine-for-hydrogen substitution adds only ~ 0.15 Å in diameter, and this substitution was expected to be generally insignificant, as has previously been observed when the activities of fluorotryptophan-substituted enzymes have been examined (27). It was also possible that the fluorine substitution would alter the electronic properties of tryptophan. However, while there are examples of the N1 of the indole ring being functional, enzymes which so use tryptophan are rare. In addition to the adaptation of proteins to 4fW substitution, there were other possible outcomes. Mutations in tryptophan-incorporating enzymes, such as tRNA synthetases or permeases, that would discriminate against the analogue might be isolated. Examples of mutant tRNA synthetases with increased specificity are known, so this was considered to be a relatively probable outcome (6). Similarly, 4fW might be selectively degraded, for example, by tryptophanase, or novel mechanisms for the defluorination of the analogue or the *de novo* synthesis of tryptophan might arise. While the latter possibility at first seems farfetched, recent work has demonstrated the mutational proximity of proteins in the histidine and tryptophan (W) biosynthetic pathways (15), while other work has demonstrated that neighboring genes in the tryptophan pathway can be mutated to functionally replace one another (1).

We made use of a serial dilution technique in which the 4fW-to-W ratio was gradually made more strict until only 4fW was added to the media. After over 3,000 h of serial dilution, a strain of bacteria was isolated that was capable of growth in media where the only W available was the 4fW analogue. While the strain still exhibits a marked growth preference for W, and thus does not contain a new genetic code, a number of coordinate genetic changes were required for adaptation to 4fW.

* Corresponding author. Mailing address: 2500 Speedway A4800, University of Texas at Austin, Austin, TX 78712. Phone: (512) 232-3424. Fax: (512) 471-7014. E-mail: andy.ellington@mail.utexas.edu.

MATERIALS AND METHODS

Strains and media. *E. coli* strain C600 Δ *trpE* (*thi-1 thr-1 leuB6 lacY1 tonA21 supE44 mcrA Δ trpE*) was a gift of I. J. Molineux at the University of Texas at Austin. This strain was transformed with the plasmid pUC18 (yielding C600p) prior to the initiation of selection experiments. Luria-Bertani (LB) and M9 media were supplemented with ampicillin (Ap), kanamycin (Kn), or chloramphenicol (Cm) (3). For selection experiments, M9 media with 0.4% glucose were supplemented with 0.0005% thiamine and 20 μ g (each) of threonine and leucine/ml. L-W and DL-4FW were introduced at various ratios so that a total concentration of 20 μ g/ml of the L-enantiomer was generally maintained (in some instances, the total concentration of the L-enantiomer was increased; see below). The resultant minimal media were designated M9B1TL, followed by the percent of tryptophan with the 4-fluoro-adduct; for example, M9B1TL95% indicates 38 μ g of DL-4FW/ml–1 μ g of W/ml.

Serial transfer. Overnight cultures grown in LB plus Ap medium were initially inoculated separately with two clones of C600p, designated C600p-1 and C600p-2. These cultures were then diluted 1:100 into M9B1TL95% plus Ap and growth curves were generated by monitoring the optical density of the culture at 600 nm (OD₆₀₀). Upon reaching either an OD₆₀₀ of approximately 0.5 or stationary phase, the cultures were further diluted 1:100 into fresh media containing ampicillin and various ratios of 4FW to W. Glycerol stocks were made prior to each transfer and stored at –80°C for later analysis. Once stable and reproducible growth was observed at a given 4FW proportion, the evolving strains were successively transferred into 97, 98, 99, and 99.5% 4FW (see Fig. 2). To avoid the possibility that tryptophan might become limiting for growth, the total amount of W and 4FW in later transfers was increased to 40 μ g total of the L-enantiomer/ml: 79.8 μ g of DL-4FW/ml and 0.1 μ g of W/ml, referred to as 2x99.72%. The remaining transfers were to M9B1TL3x99.85% (60 μ g total of the L-enantiomer/ml, 99.85% of which was 4FW), followed by M9B1TL3x99.92% and then M9B1TL3x99.97%. The strain C600p-2 did not survive beyond the first M9B1TL3x99.92% transfer. After seven serial transfers in M9B1TL3x99.97%, two clones (B7-3 and B7-4) were isolated from the C600p-1 population and used in further experiments. Plasmid DNA prepared from these clones indicated that they had retained the original pUC18 plasmid.

Characterization of growth capabilities. Overnight cultures of strains (C600p, B7-3, and B7-4) were grown in M9B1TL95% plus Ap medium were used to inoculate (by 1:100 dilutions) several media representing a range of 4FW-to-W ratios or a range of W concentrations. These cultures were aliquoted into triplicate wells of plates used by the Bioscreen C (Labsystems Oy, Helsinki, Finland), a plate reader designed to monitor cell growth. Using the Bioscreen C, growth curves were generated for each culture and averaged. The growth rate of each culture was estimated by fitting the log phase of each growth curve to the following equation: $y = a \log x + b$, in which y is the OD₆₀₀, x is the time, and a is growth rate. Growth rates were then plotted against 4FW proportions. Similar experiments were carried out with 5-fluorotryptophan (5fW) and 6fW.

Expression vector construction, protein purification, and characterization. The kanamycin kinase gene was amplified from the vector p182Sfi-Kan (K. A. Marshall and A. D. Ellington, unpublished data) using primers Kan1.39 (5'-CGCGATCCGGCCACCATGGCCAAGCGAACCGGAAT-3') and Kan2.39 (5'-CCGGAATTCGAGGCCTGACAGGCCTAGAAGAACTCGT-3'), digested with *EcoRI* and *BsaBI*, and then cloned into pGEX-KG (12) digested with *SmaI* and *EcoRI*. This glutathione S-transferase (GST) expression vector, called pGSR, was used to transform strains B7-3 and B7-4.

Overnight LB-plus-Kn cultures of B7-3 and B7-4 containing pGSR were used to inoculate M9B1TL95% plus Kn. This culture was in turn diluted 1:100 into 100 ml of M9B1TL95% plus Kn. At mid-log phase, cells were concentrated by centrifugation and resuspended in 100 ml of either M9B1TLW plus Kn or M9B1TL3x100% plus Kn, both supplemented with 0.3 mM IPTG (isopropyl- β -D-thiogalactopyranoside). After 16 h of growth, cells were harvested and lysed in 5 ml of bacterial protein extraction reagent (B-PER; Pierce, Beverly, Mass.). GST was isolated from the soluble fraction as previously described (3), and three 0.5-ml fractions were stored at 4°C. Each elution fraction was concentrated (10,000 MW cutoff; microconcentrator; Microcon, Rockford, Ill.), the final protein purity was estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to be greater than 95%, and the concentrations of the protein fractions were estimated spectrophotometrically to be as follows: 5 mg/ml from B7-3 grown in W; 6 mg/ml from B7-3 grown in 4fW; 1.5 mg/ml from B7-4 grown in W; and 0.5 mg/ml from B7-4 grown in 4fW.

Mass spectrometry was carried out at the Mass Spectrometry Facility in the Chemistry and Biochemistry Department at the University of Texas at Austin. The whole GST protein was analyzed by high-performance liquid chromatography-electrospray ionization mass spectrometry (HPLC-ESI MS) and by ma-

trix-assisted laser desorption ionization-time of flight (MS) [MALDI-TOF (MS)]. For these experiments, trifluoroacetic acid and formic acid were used to eliminate acetylation. For peptide analyses, 5 μ g of GST from B7-3 cells grown on either W or 4fW cultures was lyophilized, resuspended in 0.1 M NH₄HCO₃, and digested with immobilized L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin (Pierce) at 37°C for 10 h. Trypsin was removed by centrifugation, and the digest was lyophilized and resuspended in water to a concentration of 210 μ M for HPLC-ESI MS analysis. GST was also isolated from both C600p and B7-3 cells grown on W, 95% 4fW, or 95% 4fW in which the culture was switched to 4fW upon inoculation. Purified protein was digested as described above and analyzed by HPLC-ESI.

Amino acid analysis. Cultures of C600p and B7-3 were grown in 25 ml of media containing W and 95% 4fW to saturation. These cultures were pelleted by centrifugation and resuspended and lysed in 200 μ l of B-PER. Then, 50 μ l of this lysate was passed through a CENTRI-SEP size-exclusion column (Princeton Separations, Adelphia, N.J.) to remove unincorporated amino acids. The purified lysate was hydrolyzed (with 10% thioglycolic acid to prevent tryptophan degradation) and analyzed by the Protein Facility of the Institute for Cellular and Molecular Biology at the University of Texas at Austin.

Gene sequencing. Genes from the parental strain, B7-3, and B7-4 were amplified and sequenced on an ABI377 sequencer. The genes that were analyzed were those for aromatic amino acid permease (*aroP*), tryptophanyl-tRNA synthetase (*trpS*), tRNA^{Trp} (*trpT*), tryptophanase (*tnaA*), single-stranded DNA binding protein (*ssb*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) (*gapA*), the pUC18 origin of replication (*ColE1*), β -lactamase (*bla*), tryptophan repressor (*trpR*), tyrosine repressor (*tyrR*), and the tryptophan-specific transporter (*trt*). The primers used for the amplification and sequencing of the genes, as well as ~100 bp of flanking regions, are listed in Table 1.

Growth effects of C600p conferred by mutant *aroP* and *trpS* genes. Wild-type and mutant *aroP* and *trpS* genes were amplified and cloned into the TOPO-TA vector pCR2.1 (Invitrogen, Carlsbad, Calif.). The *trpS* insert was then subcloned into the *BamHI-XbaI* sites of pACYC184 (18). The vector with the wild-type *trpS* gene was designated pABXWSw, while the vector with the mutant *trpS* gene was designated pABXWSm. The TOPO-TA vector *BamHI-XbaI* fragments containing *aroP* were then subcloned into *BamHI-SpeI*-digested pABXWSw and pABXWSm (*XbaI* and *SpeI* yield compatible overhangs). These manipulations ultimately yielded either pAPWWSw and pAPmWSm, plasmids that contained both *trpS* and *aroP* wild-type or mutant genes, respectively. Parental C600p was separately transformed with pAPWWSw and pAPmWSm. The p15A-derived origin of replication of pACYC184 was compatible with the pUC18 plasmid already present in C600p, and thus, both plasmids could be propagated simultaneously. Single colonies were grown overnight in M9B1TL95% plus Ap and Cm. Growth rates on various ratios of 4fW to W were determined as described above.

Wild-type and mutant TrpRS overexpression, purification, and charging. The primers 40.WSB.NdeI (5'-TCGACATATGACTAAGCCCATCGTTTTAGTGGCGCACAG-3') and 40.WSA (5'-CAGCACTCGAGAATTACGGCTTCGCCAAACCAATCGC-3') were used to amplify wild-type and mutant *trpS* genes. These were ligated into the pCR2.1 TOPO vector and sequenced to confirm that there were no mutations introduced by PCR. Isolated plasmids were digested using *NdeI* and *XbaI*, and wild-type and mutant *trpS* genes were introduced into appropriately digested pET28b (Novagen, Madison, Wis.) (18), generating pETWsw and pETWsm, respectively. These plasmids were electroporated into BL21(DE3) cells. Log phase LB-plus-Kn cultures (35 ml) were induced with 0.3 mM IPTG for 3 h.

Protein was purified essentially as described by the manufacturer. In short, cultures were spun down, lysed in 1 ml of B-PER, and incubated with 10 mM MgCl₂ and 1 U of DNase I at room temperature for 5 min. The lysates were cleared by centrifugation at 10,000 rpm for 30 min at 4°C, and the supernatants were centrifuged again for 15 min. The supernatants were then loaded on 1-ml nickel columns at room temperature. The columns were washed with 10 ml of binding buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 50 mM imidazole) and 6 ml of wash buffer (same as the binding buffer but with 60 mM imidazole), and fractions were collected in elution buffer (20 mM potassium phosphate, pH 7.0, 0.5 M NaCl, 1 M imidazole). Fractions with protein were identified by SDS-PAGE, pooled, and dialyzed overnight against TrpRS buffer (20 mM potassium phosphate, pH 7.0, 10 mM KCl, 5 mM MgCl₂, 4 mM dithiothreitol, 20% glycerol) at 4°C; the dialyzed proteins were stored at –20°C. Protein was quantitated by Bradford assay or by SDS-PAGE using known standards.

tRNA for charging assays was produced using the AmpliScribe T7 transcription kit (Epicentre, Madison, Wis.) with the template WtRNA.1 (5'-TGGCAGGGCGGAGAGACTCGAACTCCCAACACCCGGTTTTGGAGACCCGGTGCTCTACCAATTGAACTACGCCCTCTATAGTGAGTCGTATTAGAA-

TABLE 1. Primers used for amplification and sequencing of genes^a

Gene	Primer ^b	Primer sequence
<i>aroP</i>	22.APA	5'-GACGATTTACGGGGCGTTGGTG-3'
	24.APB	5'-CATTCTACATATTGAGAGGGGTTG-3'
	22.AP-iA1	5'-CGCCAGGCTAATCATCGCCAG-3'
	20.AP-iA2	5'-CCACAGTTGCTAACGGTCG-3'
	21.AP-iB	5'-GTGTTTGGCGAGATGGAGTTC-3'
<i>trpS</i>	24.WSA	5'-CATCCGGCATGAACAAAGCGCAAT-3'
	24.WSB	5'-TTCTGCCCGCATTAGGGCTTCCGC-3'
<i>trpT</i>	24.WTA	5'-GGGTCGCGGGTTCGAGTCCCGTCC-3'
	24.WTB	5'-CTTCGGAGAGGGTTATTTTCAGATA-3'
<i>tnaA</i>	24.TNA	5'-GATGGTGCTTGCATATATATCTGG-3'
	24.TNB	5'-GTTATTGAGGATGTAGGGTAAGAT-3'
	24.TN-iA	5'-GATATGCGGCAACAGTTTACCGGC-3'
	24.TN-iB	5'-GTGGCGCAGAGCAAATCTATATTC-3'
<i>ssb</i>	24.SBA	5'-GTTTCCCGGATCCGAGGTCACAAC-3'
	25.SBB	5'-CTGGCAGATGCTTTAAGGATCCACC-3'
<i>gapA</i>	24.GAA	5'-TTGACGCTGCGTAAGGTTTTTGTGA-3'
	24.GAB	5'-AGCGACCGAAGTCGCTCTTTTAG-3'
ColE1	25.OCA	5'-TAATGGTTTCTTAGACGTCAGGTGG-3'
	24.OCB	5'-GTAGTTCGCCAGTTAATAGTTTCG-3'
<i>bla</i>	26.50	5'-TTCTTAGACGTCAGGTGGCACTT-3'
	27.19	5'-TTAAGGGATTTTGGTCATGAGATT-3'
	24.JB1	5'-AACTTTATCGGCCTCCATCCAGTC-3'
<i>trpR</i>	20.WRA	5'-CCCCGCTAACAATGGCGAC-3'
	19.WRB	5'-CGCCTGATGCGACGCTGCC-3'
<i>tyrR</i>	20.YRA	5'-ATCTTTACGCCGAAGTGCCC-3'
	21.YRB	5'-CGCCTGATGCGACGCTGCC-3'
	23.YR-iA	5'-TGTCGATATGAAAAGCAAAGTGG-3'
	26.YR-iB	5'-GTCAACACGTTTCAGACGATAATAGAG-3'
<i>mtr</i>	36.MTA	5'-GGCAACGCTGTCTACATAGACCTGATA AGCGAAGC-3'
	38.MTB	5'-TCGGGGCTTTTTTCTGTCTTTTGTACTC GTGTAAGTGG-3'
	20.MT-iA	5'-CCAGACAGAACGGCAGGGTC-3'
	27.MT-iB	5'-GGCGAAAGTCATTACCTTCTCCTCAC-3'

^a Primers amplify genes as well as ~100 bp of 5' and 3' flanking sequence.

^b The first two primers listed for each gene were used to amplify the gene and sequence and the 5' and 3' ends. Any additional primers were used to sequence internal regions of the gene.

3') and the T7 promoter sequence 24T7.JB1 (5'-TTCTAATACGACTCACTA TAG-3'). This sequence adds an extra G to the 5' end of the tRNA to enhance transcription. tRNA was labeled with [α -³²P]ATP and purified by PAGE.

To determine the relative charging abilities of mutant and wild-type enzymes, each was incubated with various amounts of 4fW in the presence of [³H]W. Reactions were carried out with final concentrations of 50 mM HEPES (pH 7.0), 10 mM MgCl₂, 4 mM dithiothreitol, 2 mM ATP, 0.05% bovine serum albumin, 2 μ M [α -³²P]tRNA^{Trp}, and 2 μ M [³H]W (32). These solutions were heated to 70°C for 3 min and cooled to 37°C for 10 min prior to the addition of enzyme to start the reactions. Reactions were carried out in triplicate. Four minutes after the addition of enzymes, samples (10 μ l) were loaded onto filter discs that had previously been soaked with 5% trichloroacetic acid and 10 mM W. Filter discs were washed three times in 5% trichloroacetic acid-10 mM W, twice in 70% ice-cold ethanol, and once in ether, and the discs were then air-dried and counted. Ratios of ³H counts to ³²P counts were determined, normalized to a no-4fW control, and plotted against the concentration of the 4fW competitor.

Incorporation assay. Overnight cultures of C600p and B7-4 were grown on M9B1TL95% plus Ap and used to inoculate a 105-ml culture of M9B1TL95% plus Ap by a 1:100 dilution. These cultures were grown to an OD₆₀₀ of ~0.35, spun down at room temperature, and resuspended in 50 ml of M9B1L plus Ap. Four milliliters of the resuspended cultures was aliquoted to fresh tubes containing 50 nM [³H]W-500 nM [¹⁴C]T. These cultures were incubated at 37°C with shaking (250 rpm) for 10 min and then placed on ice. Triplicate 1-ml

samples were aliquoted from each culture and centrifuged. Pellets were resuspended and lysed in 50 μ l of B-PER. Half of each lysate was centrifuged to remove insoluble material and passed through CENTRI-SEP columns to remove unincorporated label (these columns were shown to remove detectable radioactivity when at least a ~100-fold-greater number of counts were applied). Eluant (10 μ l) was mixed with scintillation fluid and counted. ³H and ¹⁴C counts were corrected for overlapping windows based on known standards and were converted to femtomoles of W and picomoles of T, and ratios were determined and plotted.

RESULTS AND DISCUSSION

Evolution of 4fW-tolerant bacteria. *E. coli* cells are remarkably tolerant of 4fW (Fig. 1) if tryptophan is also provided. We found that the tryptophan auxotroph C600p was capable of stable, continual growth in media where 95% of the available tryptophan was 4fW (for example, Fig. 2). However, while *E. coli* tolerates 4fW, it cannot use 4fW as the sole source of tryptophan (5, 29). Attempts to grow C600p in higher proportions of 4fW (>99% 4fW) inevitably resulted in cell death after one or two serial transfers. We reasoned that it might be possible to select for bacteria with increasing tolerance to (and possibly increased utilization of) higher proportions of 4fW.

Simple serial transfer techniques were used to evolve 4fW-tolerant strains (Fig. 2). In order to prevent contamination of the evolving bacterial population with environmental prototrophs, a plasmid bearing an antibiotic resistance element was introduced into the parental strain. The presence of this

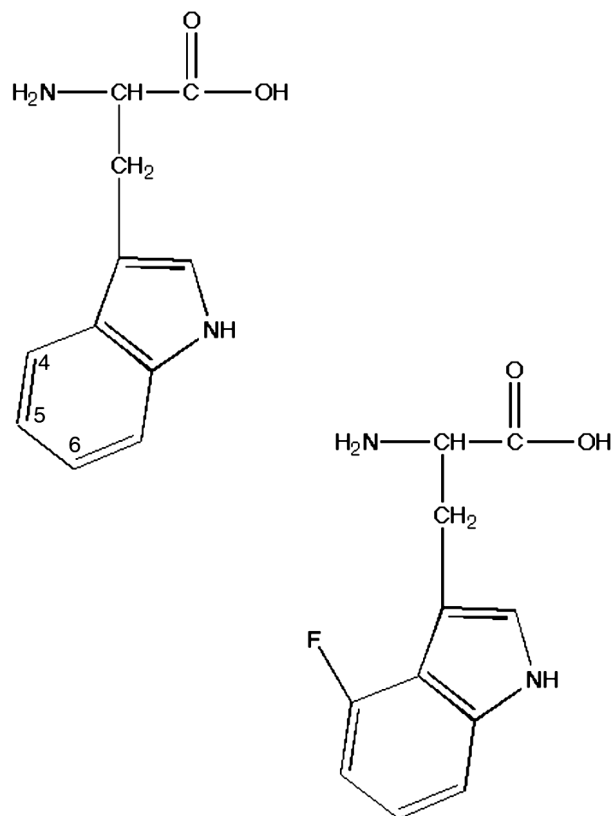


FIG. 1. Tryptophan and 4-fluorotryptophan. The fluorine atom is 0.15 Å in diameter larger than the hydrogen atom it replaces but is more electronegative. Numbers indicate positions of fluorine substitutions for tryptophan analogues that have been examined.

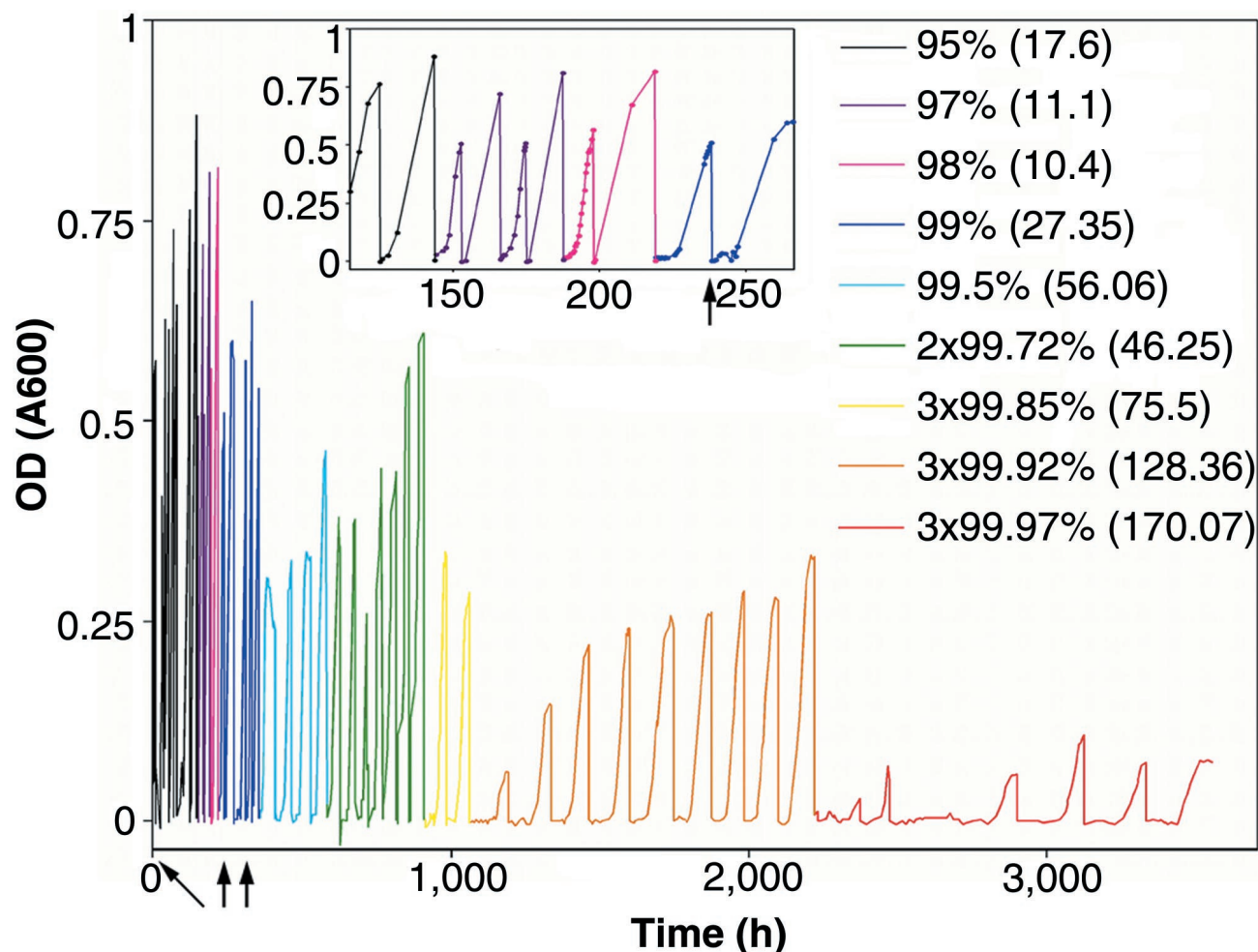


FIG. 2. Evolution of 4fW-tolerant *E. coli*. OD₆₀₀ readings of a culture of C600p were made as a function of time. Each culture was diluted into fresh media when it entered stationary phase or reached an OD₆₀₀ of $> \sim 0.5$. After stable and reproducible growth was demonstrated in each medium, the culture was switched to a higher ratio of 4fW to W. The inset shows an expanded view of growth during the early stages of adaptation. Numbers in parentheses are the average time (in hours) between dilutions. Arrows indicate unsuccessful attempts to transfer cultures to media with much higher 4fW-to-W ratios: the parental strain was unable to grow on 99% 4fW, while strains at 238.25 and 312.25 h were unable to grow on 99.97% 4fW. As the serial transfer was continued, successively higher total amounts of total W were added to the media (see Materials and Methods). For this reason, 2x- and 3x- followed by percentages designate real ratios of 4fW to W but also indicate a greater amount of 4fW plus W in the media. Furthermore, 100% 4fW is not shown due to the contamination of commercial 4fW with 0.03% W; hence, 99.97% 4fW is the highest possible ratio of 4fW to W.

plasmid was used throughout serial transfer experiments as a marker for the evolving strain. After 250 h of continuous evolution and 14 serial transfers, the parental strain was gradually adapted to 97, 98, and finally 99% 4fW (Fig. 2, inset). Growth at proportions with greater than 99% 4fW proved difficult to achieve and required more discrete steps. Nonetheless, after 3,000 h of continuous evolution, the strain could grow, albeit slowly, with only 4fW added to the growth media (Fig. 2).

The purity of the commercial 4fW sample was assayed by HPLC, and 0.03% contaminating tryptophan was found. For this reason, we never refer to media as being 100% 4fW; instead, when only commercially available 4fW was added to the media, it is referred to as 99.97% 4fW. It was also possible that contaminating tryptophan might have arisen from defluorination of the original sample, but MS revealed that no additional tryptophan accumulated as a breakdown product following prolonged (8 days) incubation in media. Therefore,

the final, evolved strain was capable of growth in the presence of only 36 ng of DL-tryptophan/ml, provided that 4fW was also present. For the sake of comparison, tryptophan was limiting for growth at concentrations less than 2 $\mu\text{g/ml}$ for wild-type C600p (see below and Fig. 4).

Two clones, B7-3 and B7-4, were isolated from the final population of cells and were used in further characterizations. As expected, the evolved isolates still contained the original plasmid, were tryptophan auxotrophs, and could grow on 99.97% 4fW. The growth rates of the evolved strains were greater than that of the parental strain at all proportions of 4fW (Fig. 3). Moreover, the evolved clones were more tolerant of other tryptophan analogues, such as 5fW and 6fW, than the parental strain. However, both evolved clones still preferred tryptophan for growth and grew as well as the parental strain in the presence of tryptophan. In fact, the growth rates of evolved and ancestral clones were equally reliant upon W

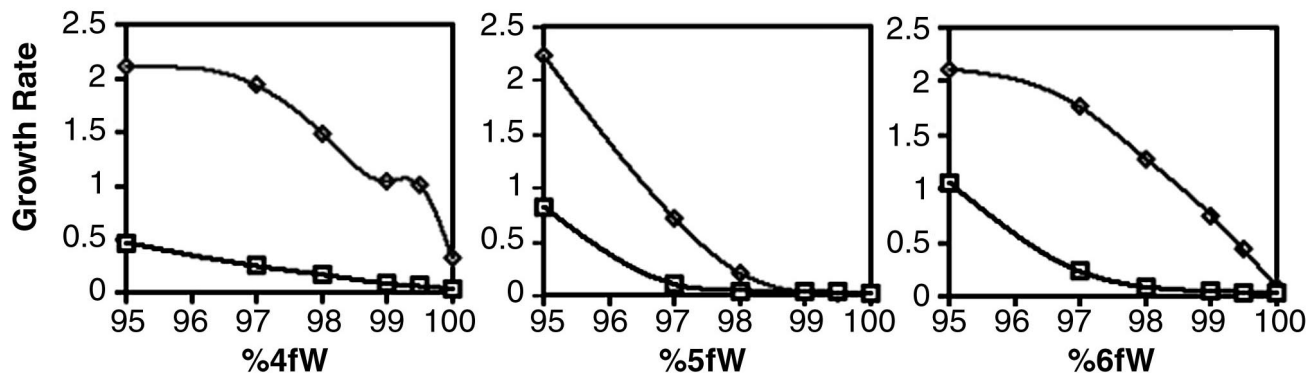


FIG. 3. Effect of various ratios of analogue to natural tryptophan on the growth of parental and evolved clones. Clone B7-3 (\diamond) and parental C600p (\square) were grown on various ratios of 4fW (left), 5fW (center), and 6fW (right). Growth rates for clones were determined by fitting the log portion of growth curves to the following equation: $y = a \log x + b$, where a was the growth rate (see Materials and Methods). The evolved clone B7-3 had a growth advantage for all ratios of 4fW to W, including a ~ 10 -fold increase on 99.97% 4fW.

concentration (Fig. 4). The lowest levels of [W] examined here are slightly greater (20 ng/ml compared with 18 ng/ml) than the residual tryptophan found in 99.97% 4fW, indicating that there is likely no W production in the evolved cells.

Confirmation of 4fW incorporation. In order to determine whether the evolved strain utilized or tolerated 4fW, it was important to determine whether protein extracted from the wild-type parent and the evolved variant in fact incorporated 4fW. Because tryptophan is known to be labile during protein hydrolysis, it was possible that significant amounts of residual tryptophan might be missed during amino acid analysis. Indeed, even literature procedures optimized to quantitate tryptophan content are seldom able to detect better than 95% of total tryptophan. Therefore, we attempted to use less invasive methods to determine the extent of 4fW incorporation (9, 34). We reasoned that while genes present during the course of the selection experiment might have mutated, a newly introduced gene should accurately reflect the incorporation potential of the evolved translation apparatus.

The parental strain (C600p) and the evolved strain (B7-3) were transformed with a plasmid that overexpressed the protein GST. It had previously been demonstrated that GST retained function (and thus could be readily purified) when 5fW was widely incorporated in place of tryptophan (27), and we anticipated that it might also retain function (and thus be amenable to purification) following 4fW incorporation. After induction in media containing either tryptophan or 4fW, GST was isolated and HPLC-ESI MS was performed on samples. Protein samples from cells of strain B7-3 grown on W exhibited a mass peak consistent with full-length GST (27 kDa), while protein samples from strain B7-3 grown on 4fW exhibited a mass shift of 72 Da, the exact amount predicted by replacement of the four GST tryptophans with 4fW (19 Da [F] – 1 Da [H] for each tryptophan replaced); similar results were observed for MALDI-TOF (MS). To further assess whether this mass shift represented the discrete incorporation of 4fW residues, isolated proteins were digested with trypsin and the fragments were once again run on an HPLC-ESI MS. Three of the resultant fragments were expected to contain the four tryptophan residues. Although one of the peptides was too small to be traced by MS (WR), the two remaining fragments contained mass shifts that were consistent with 4fW replacement. For

example, the fragment MSPILGYWK from strain B7-3 cells grown on W was found to have the predicted mass of 1,094.3 Da, while the same fragment from strain B7-3 cells grown on 4fW was found to have the predicted mass of 1,112.3 Da (Fig. 5).

Using these techniques, the relative degree of 4fW incorporation in parental and evolved strains could be assessed. When the wild-type parent was grown on 95% 4fW, the signature peptides described above were found in roughly equal measure (Fig. 6A). This indicates a roughly 19-fold discrimination in favor of W (19:1 in the media versus 1:1 incorporated in protein). In contrast, when clone B7-3 was grown on 95 or 99.97% 4fW, the 1,094-Da peak could not be found in the digest, as shown in Fig. 6B, indicating that incorporation of 4fW was complete (at least at the level of our analysis; sensitivity to 1%). The other GST peptides derived from B7-3 cells grown on 4fW also appeared to contain only modified tryptophan.

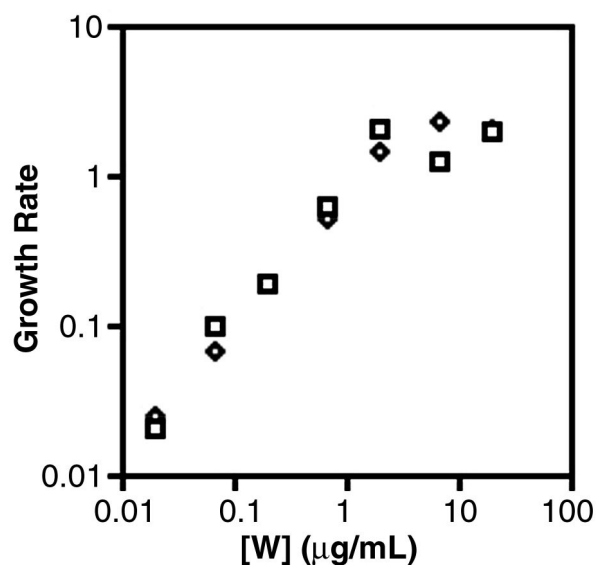


FIG. 4. Effect of tryptophan concentration on growth of ancestral and evolved clones. Clone B7-4 (\diamond) and C600p (\square) were grown on various ratios of W. Growth rates were determined as described for Fig. 3 (also see Materials and Methods).

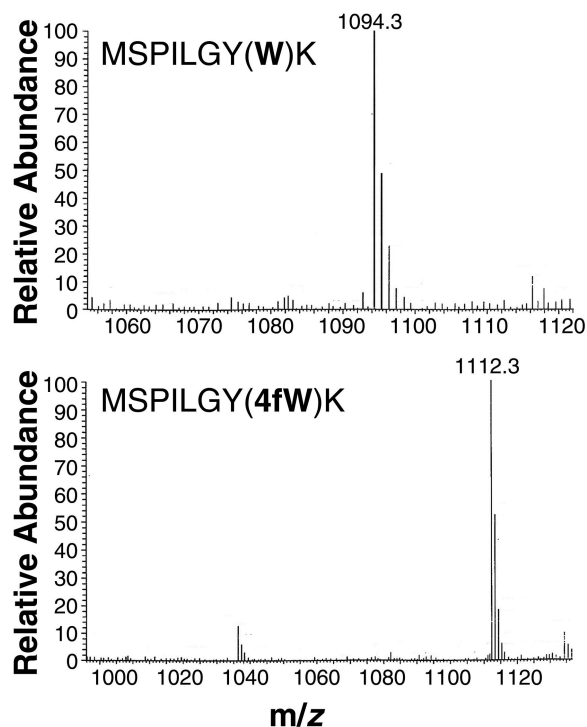


FIG. 5. Incorporation of 4fW by an evolved clone. GST purified from B7-3 cells grown in the presence of tryptophan was digested with trypsin, and peptides were separated and analyzed by HPLC-ESI MS. A major peak eluted at 29 min and had a mass of 1,094.3 Da (top). The mass of this peak corresponds to that of the predicted tryptic fragment MISPLGYWK. The peak shifts to a mass of 1,112.3 Da when B7-3 cells are grown on 99.97% 4fW rather than tryptophan (bottom).

These results were confirmed by more-conventional amino acid analyses. Whole-protein extracts were hydrolyzed in thioglycolic acid to avoid extensive tryptophan degradation. When the parental strain was grown in 95% 4fW, extracted proteins showed from 70 to 80% incorporation of 4fW. These results are consistent with those reported by Pratt and Ho (29), who previously estimated that no more than about 73% of the tryptophan residues in cellular proteins could be replaced with 4fW in an unevolved strain (29). In contrast, when cells of the evolved clone B7-3 were grown on 95 or 99.97% 4fW, no natural tryptophan could be detected in extracted proteins.

Genetic and biochemical characterizations of evolved clones.

Given that the evolved strain could grow in the presence of 99.97% 4fW while the starting strain could not, the evolved strain had likely accumulated mutations that altered proteins adversely affected by 4fW incorporation. For example, previous experiments with proteins involved in lactose utilization revealed that 4fW had deleterious effects on a number of enzyme activities (5, 29). In order to identify what types of mutations contributed to the ability to grow on 4fW, a number of genes from clones B7-3 and B7-4 were isolated and compared with the same genes from the parental strain. For the most part, changes were not observed, neither in the coding regions nor in the regions roughly 100 residues up- and downstream of the coding regions. Genes that did not encode proteins (tRNA^{Trp}, the pUC18 origin of replication) contained no mutations. Furthermore, essential enzymes (GAPDH, β -lacta-

mase) and proteins known to rely heavily on tryptophan for function (single-stranded DNA binding protein) (35) contained no mutations relative to the parental strain. The functions of these proteins may very well have been impaired by fluorotryptophan incorporation, especially given the poor growth of the evolved strain, but the impairment was not significant enough that mutational fixation was required in order for growth to occur.

Mutations may also have accumulated in enzymes involved in tryptophan uptake and incorporation. Several enzymes involved in tryptophan metabolism (tryptophanase, tryptophan repressor, and tryptophan-specific transporter [*mtr*]) remained unchanged. That tryptophanase was found to be wild type eliminates as a possibility the most obvious mechanism by which the evolved strain might have developed a means for specifically degrading 4fW. However, three proteins that were directly involved in tryptophan utilization, tryptophanyl-tRNA synthetase (*trpS*), aromatic amino acid permease (*aroP*), and tyrosine repressor (*tyrR*), were mutated. The synthetase and repressor each contained a single mutation (Q109P and R38Op, respectively), while the permease contained three substitutions (T30S, I47V, V365G). The opal mutation in the tyrosine repressor should have led to the production of a truncated, nonfunctional protein and thus allowed the overexpression of the *AroP* permease and of other operons in the aromatic amino acid regulon (24, 28). Interestingly, wild-type *E. coli* cells show a >15-fold increase of *mtr* mRNA under W starvation conditions

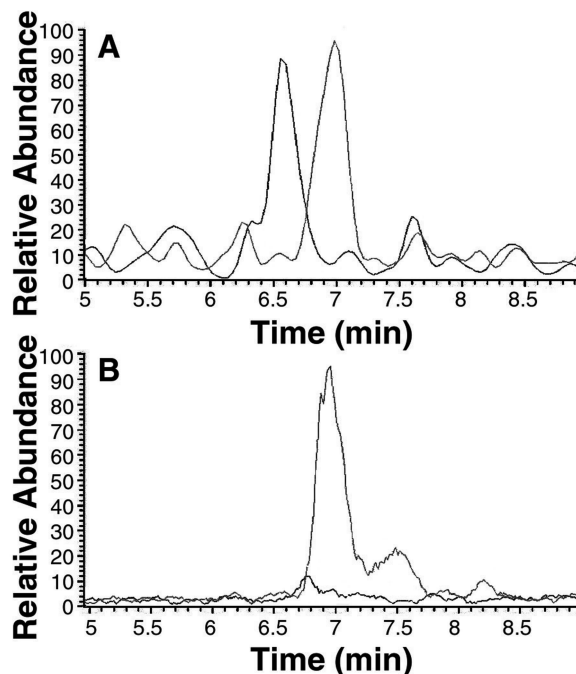


FIG. 6. Relative incorporation of 4fW by an evolved and ancestral clone. GST was purified from clones B7-3 and C600p grown on 95% 4fW medium, as described above. HPLC-ESI MS was performed. Traces for the masses of 1,094 and 1,112 Da are shown. (A) A peptide isolated from C600p shows a roughly equal proportion of 1,094- and 1,112-Da peptides at about 6.6 and 7 min. (B) HPLC-ESI MS of a peptide produced by B7-3 shows an absence of any significant peak at ~6.6 min of a mass of 1,094 Da.

(17); however, *tyrR* actually acts to increase expression levels of *mtr* (EcoCyc website [http://ecocyc.org] and reference therein [30]), so the level of *mtr* mRNA may actually have decreased from this level. In fact, this makes sense from a resource allocation point of view: the evolved variants should preferentially express *aroP*, where mutational substitution happened to have been concentrated.

In order to determine when during the course of the experiment these mutations may have arisen, genes from intermediary organisms were also examined. The Q109P mutation fortuitously introduced a *Bsa*HI restriction site into the *trpS* gene, and the presence or absence of this site was used to determine the approximate genesis of the *trpS* mutation in evolutionary history (Fig. 7A). Similarly, individual *aroP* and *tyrR* genes were isolated and sequenced (Fig. 7B). During the first 250 h of evolution, one *aroP* mutation (V365G) arose; this substitution was apparently sufficient to allow adaptation to growth on 99% 4fW. The *trpS* mutation, the remaining two *aroP* mutations, and the *tyrR* mutation occurred in succession. Interestingly, the order of the *aroP* mutations was in rough agreement with their degree of phylogenetic conservation: the first substitution occurred in the most variable residue, the last in the least variable residue (2).

The mutations were likely necessary for growth in the presence of high proportions of 4fW. A replicate experiment carried out in parallel yielded organisms that could not survive at 4fW proportions above 99.92% (C600p-2; see Materials and Methods). As was the case with the eventually successful strain, an initial substitution arose in *aroP* at approximately the same time. However, its identity was V365A rather than V365G. The failed line never chanced to acquire the remaining *trpS* (Fig. 7A), *aroP*, and *tyrR* mutations, perhaps because the initial adaptive mutant did not adequately predispose the organism to selection pressure that would have favored further 4fW utilization.

However, it was unclear whether the mutations were, on their own, sufficient for growth in the presence of high proportions of 4fW. Wild-type and mutant *trpS* and *aroP* genes were introduced into the parental C600p tryptophan auxotroph via a low-copy-number vector. A clone bearing the mutant *aroP* and *trpS* genes showed an improved ability to grow at all proportions of 4fW (Fig. 8). While the merodiploidy of these clones (which had genomic copies of the wild-type genes, but plasmid-borne wild-type or mutant genes) may obscure the full effects of the mutant genes, the effect of growth on 4fW is striking, and it clearly demonstrates that these genes provide a growth advantage in the presence of the analogue. More remarkably, the clone bearing mutant *aroP* and *trpS* genes showed a generally improved ability to grow on the tryptophan analogue 6fW. However, the mutant genes proved to be a hindrance to growth at higher proportions of 5fW. It is interesting that this analogue preference (4fW > 6fW > 5fW) is similar to that seen with *Bacillus subtilis* tryptophanyl-tRNA synthetase, which has been shown to incorporate tryptophan analogues, albeit 6- to 70-fold slower than it incorporates tryptophan (38). These results also coincide with those of Liu and Schultz (23), who were able to substitute multiple, different amino acid analogues into a protein via a single mutant synthetase.

Nonetheless, while the mutant genes were necessary for growth in the presence of high proportions of 4fW, they were

not sufficient. Just as the parental strain could not initially grow on 99.97% 4fW, when the cultures used to generate the data in Fig. 8 were diluted into fresh 99.97% 4fW media, no further growth was observed. The mutant genes may make growth on 4fW possible, but additional, as-yet-unidentified, adaptive mutations spread throughout the genome of the evolved strain are almost certainly required. These results underscore the idea that drastic evolutionary changes such as the large-scale introduction of unnatural amino acids may require the evolution of an entire organism, rather than the targeted engineering of only one or a few specific genes.

The changes in tryptophan utilization could have been due to the evolved permease, the evolved synthetase, or other mutant enzymes not yet identified. In particular, we suspected that tryptophanyl-tRNA synthetase might be involved in discrimination against fluorotryptophan. The structure of a highly homologous tryptophanyl-tRNA synthetase (from *Bacillus stearothermophilus*) has been solved (10). In this enzyme, a glutamine equivalent to Q109 lies near the end of an α -helical domain and reaches down into the Trp-AMP binding pocket. At its closest approach, the *B. stearothermophilus* glutamine is only 3.8 Å from the Trp-AMP complex (Fig. 9A). Thus, it seemed reasonable to assume that the Q109P mutation may have altered the specificity for tryptophan. To assess this hypothesis, wild-type and mutant tryptophanyl-tRNA synthetases were overexpressed and purified, and the charging of radiolabeled tryptophan was monitored as a function of fluorotryptophan concentration (Fig. 9B). These data can be most readily analyzed using the following equation: $V_{\max,W} = k_{\text{cat},W} [W]/[W] + K_{m,W} (1 + [4fW]/K_{m,4fW})$. This is the standard rate equation for an enzyme that can utilize one of two different substrates. Fitting of the equation by nonlinear regression allowed the assessment of relative kinetic values (Table 2). As can be seen, this equation readily fits the data we have obtained ($r^2 = 0.998$ and 0.931 for the wild type and mutant, respectively). These data show that, relative to that of the wild type, $k_{\text{cat},W}/K_{m,W}$ of the mutant TrpRS increased roughly twofold, while the K_m for 4fW increased fourfold. This indicates both improved turnover for W and increased discrimination against 4fW by the mutant enzyme. To further assess whether the evolved synthetase could more quickly incorporate tryptophan and/or analogues into the proteome, we carried out an additional experiment in which ^3H -labeled tryptophan and ^{14}C -labeled threonine were both added to the media, and the relative ^3H -to- ^{14}C ratios were determined for both wild-type and evolved strains. As can be seen in Fig. 10, the evolved strains do indeed have a much greater ability to incorporate tryptophan. Indeed, since there is only a twofold increase in k_{cat}/K_m for W for the mutant synthetase, the overall threefold increase in W/T (the absolute number of femtomoles of W and picomoles of T) for the derived strain could be the result of an altered or more highly expressed amino acid permease.

The *tyrR* mutation is likely to also be significant. This repressor is responsible for the expression of genes in the aromatic amino acid regulon, several of which are involved in the biosynthesis of chorismate, the precursor of tyrosine, phenylalanine, and W. Specifically, 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthases (*aroF* and *aroG*, isoenzymes which are repressed by Tyr and Phe, respectively) as well as shikimate kinase II (*aroL*) are now derepressed in the *tyrR*

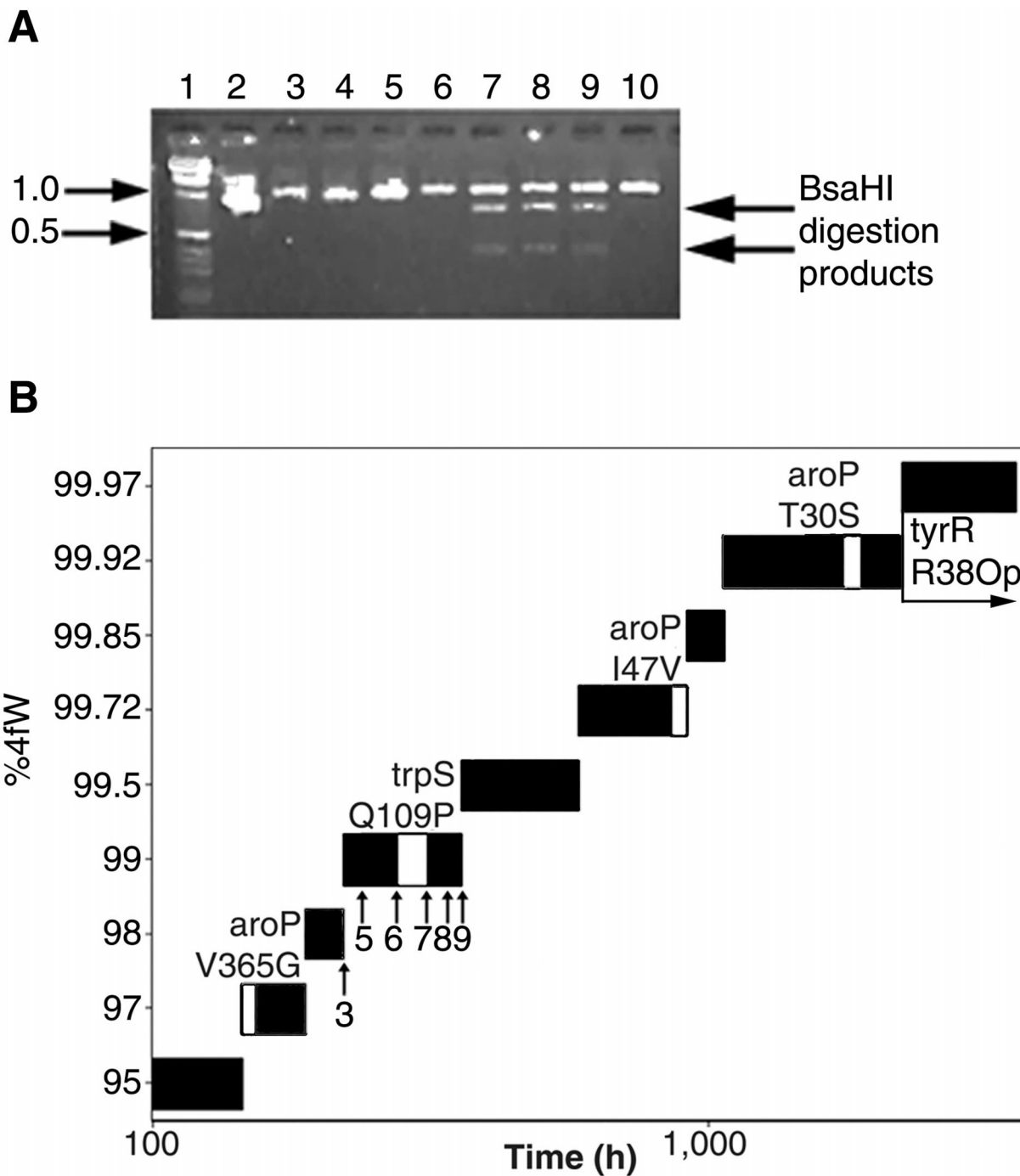


FIG. 7. Serial introduction of functional mutations. (A) Evolutionary history of the *trpS* Q109P substitution. The Q109P substitution introduces a unique *Bsa*HI site into the *trpS* gene. The *trpS* gene was amplified from different cultures and partially digested with *Bsa*HI, and the fragments were separated on a 2% agarose gel. Lanes 1 and 2, 1-kb ladder and 1-kb standard; lanes 3 and 5 through 9, samples taken from the evolving strain, as indicated in Fig. 7B; lanes 4 and 10, samples taken from a parallel, unsuccessful strain, C600p-2, at time points identical to those shown in lanes 3 and 9, respectively. (B) Evolutionary history of the strain. The time spent at each ratio of 4fW to W is shown; time is on a log scale to simplify presentation. The white inserts represent the cultures in which individual mutations arose in either *trpS* or *aroP*. The timing of the sweep of the *tyrR* opal mutation could not be definitively located, beyond generalizing it to some time during the seven cultures grown at 99.97% 4fW. Arrows indicate the points at which samples were taken for *trpS* restriction analysis, and numbers refer to the gel lanes in Fig. 7A.

mutant (information found at the EcoCyc website [<http://ecocyc.org>] and references therein [4, 7, 8]). These are major control points for chorismate biosynthesis, the other genes in the pathway being expressed constitutively. In addition to the

derepression of *aroF* and *aroG*, expression of *aroH* (encoding the Trp-repressed DAHP synthase) has been shown to increase as much as sevenfold under conditions of tryptophan starvation (17). While *aroL* mRNA levels are shown to be

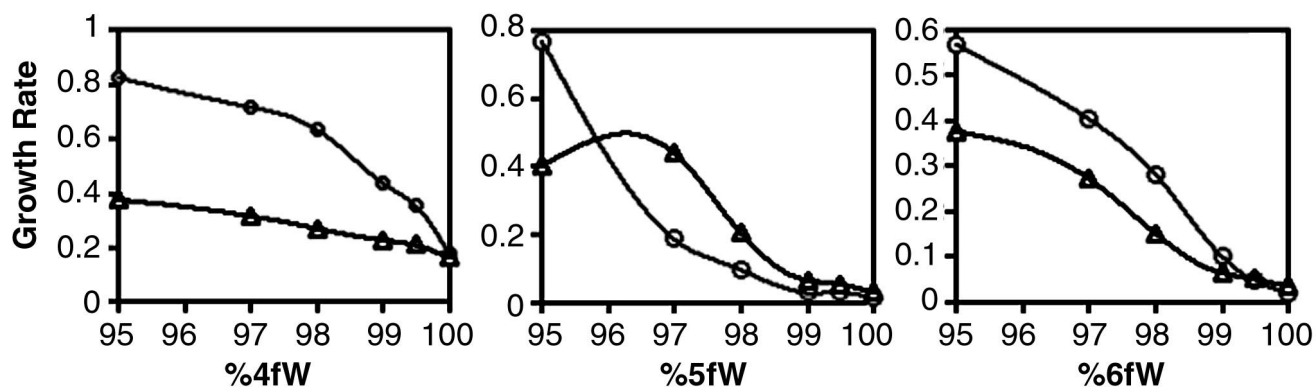


FIG. 8. Mutant *trpS* and *aroP* genes provide a growth advantage on tryptophan analogues to parental C600p. Parental C600p was transformed with plasmids pAPwWSw (Δ) and pAPmWSm (\circ), bearing wild-type and mutant *trpS* and *aroP* genes, respectively. Clones were grown on 4fW (left), 5fW (center), and 6fW (right), and growth rates were determined as described for Fig. 3 and in Materials and Methods.

unaffected by conditions of tryptophan starvation (17), that may not be true here, owing to the mutant genes being considered. This suggests that there may now be a relatively large amount of chorismate in the cell. Although our data regarding the incorporation of 4fW clearly show that the majority of 4fW is the major portion of incorporated tryptophan, the possibility exists that some of this excess chorismate may be converted to tryptophan. This would supplement the efforts of the cell to scrounge for extra tryptophan by its presumptive opening of the permease with the three mutations in that gene, as well as the derepression of *aroP* by the *tyrR* mutation. It should be noted that *p*-aminobenzoic acid (PABA) is derived from chorismate, as is anthranilate, and PABA and anthranilate are very similar structurally, differing in the amine that is in the *para* or *ortho* position relative to the acetyl group of the benzene ring. However, the fact that the evolved strain maintains an absolute requirement for tryptophan (Fig. 4) indicates that even if there were a mutation in the PABA biosynthetic pathway (or any other pathway) which produces anthranilate for the use in the tryptophan biosynthetic pathway, this is still not sufficient for the cells to completely do without the 4fW supplied to them.

Conclusion. We have generated an organism that incorporates the unnatural amino acid 4fW throughout its proteome, at least within the limits of our analytical methods (Fig. 5 and 6). This suggests that the evolved strain could not avoid the analogue. Furthermore, the evolved strain has a definite growth advantage on 4fW, as well as other tryptophan analogues (Fig. 3), and the known mutations also confer a growth advantage when transferred to ancestral cells (Fig. 8). The evolved strain could not specifically degrade or defluorinate 4fW, neither via tryptophanase (which was found to be wild type) nor through some other mechanism. More sensitive analytical techniques suggest that the mutant TrpRS has acquired some capacity for discrimination between tryptophan and its analogue (Fig. 9B). However, it is clearly not enough of a discriminator to have a significant level of tryptophan incorporation when the organism is grown at 99.97% 4fW (Fig. 5). This suggests that the incorporated level of W may be slightly higher than 0.03%, but not to a detectable level, such as $\sim 1\%$ by MS or $\sim 12.5\%$ by amino acid analysis. Similarly, the *AroP* mutations may be assisting the uptake of W, possibly even

preferentially, but not to a level that significantly incorporates W into the proteome.

Taken together, these data suggest a model for the metabolic evolution of the derived strain. First, the evolved strain is accommodating 4fW, as opposed to avoiding or eliminating 4fW, and is clearly not well adapted to the analogue. In support of this conclusion, the derived strain still grows slowly on 99.97% 4fW and enzyme mutations appear to favor the incorporation of tryptophan over fluorotryptophan. Second, there must be relatively few tryptophans that are still required for growth. While the evolved strain scrounges available tryptophan from the media, it apparently incorporates fewer than 1 in 100 tryptophans into its proteins. This conclusion is consonant with the data of Wong (37), from which we infer that a relatively small number of mutations were required to alter the amino acid specificity of *B. subtilis*. Finally, it is likely that our evolved strain is in the process of acquiring such key mutations. While the genes for *aroP* and *trpS* variants are necessary for the accommodation of 4fW, they are not on their own sufficient to allow the wild-type strain to incorporate 4fW at high levels. It is possible that if additional mutations in the proteome are allowed to accumulate that the present strain may yet be taught to favor 4fW.

Directed incorporation of unnatural amino acids in vivo has generally been restricted to specific sites, such as a single azaleucine for leucine substitution in a single protein (20) or to a variety of amino acids at a single position in β -lactamase (23). A recent report described a mutant tRNA synthetase that

TABLE 2. Relative kinetic parameters for wild-type and mutant TrpRS^a

Parameter	Wild type	Mutant	Mutant/wild-type
$k_{cat,W}$	1.87	2.63	1.4
$K_{m,W}$	0.092	0.073	0.79
$k_{cat,W}/K_{m,W}$	20.3	36.0	1.77
$K_{m,4fW}$	49.70	205.23	4.1
r^2	0.998	0.931	

^a The data in Fig. 9 were fitted to the equation used to calculate $V_{max,W}$ by using nonlinear regression. Since the data in Fig. 9 describe relative charging abilities, they do not allow the determination of absolute values for kinetic parameters.

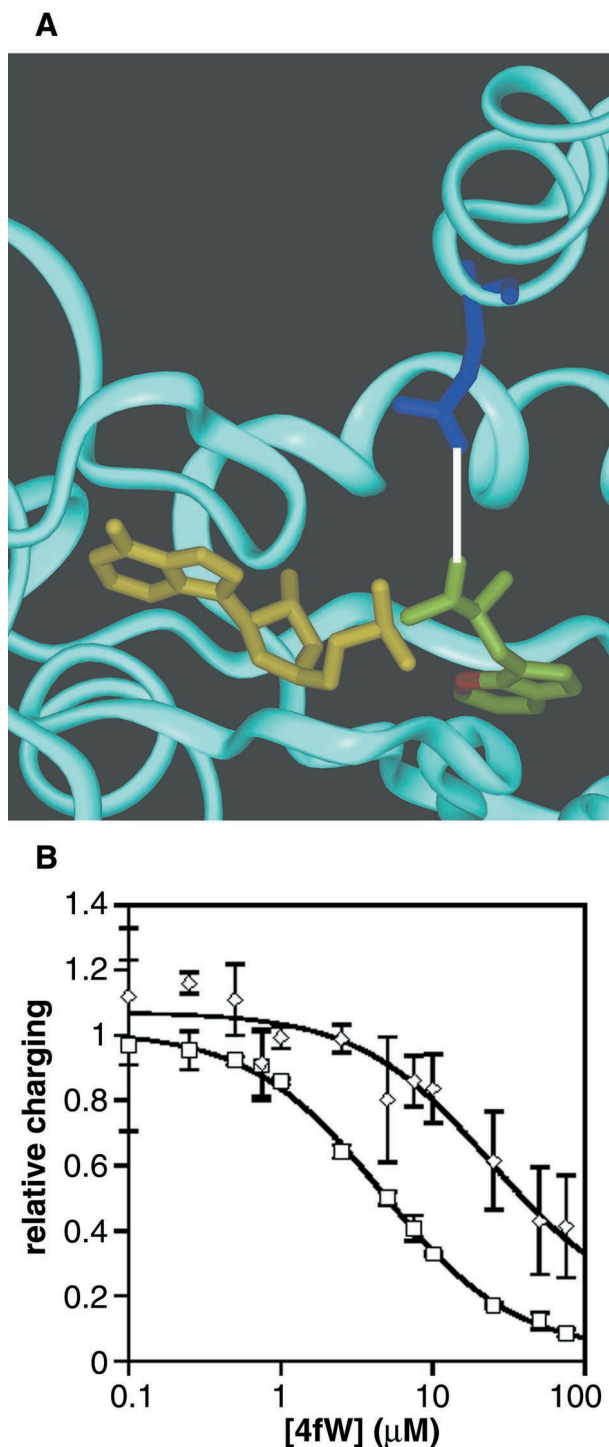


FIG. 9. Effects of the Q109P mutation on the TrpRS. (A) Ribbon model of tryptophanyl-tRNA synthetase of *B. stearothermophilus*; Glutamine 107 (corresponding to Q109 of *E. coli*) (blue) is less than 4 Å (white line) from the Trp-AMP complex (W in green with position 4 in red; AMP in yellow) (10). Q107 (109) is in an α -helical domain near the tRNA docking region. (B) Competition charging assay. Protein was added to a charging assay solution (32) containing constant concentrations of radiolabeled W and competing, unlabeled 4fW. "Relative charging" for wild-type (□) and mutant (◇) proteins is the following ratio: $([^3\text{H}]\text{W}/^{32}\text{P-tRNA})_{4\text{fW}}/([^3\text{H}]\text{W}/^{32}\text{P-tRNA})_{\text{No } 4\text{fW}}$. In other words, the charging of tryptophan at a given concentration of 4fW normalized to the charging of tryptophan in the absence of 4fW. Error bars represent the standard deviation of triplicate experiments. Data were fit to the equation used to calculate $V_{\text{max,W}}$ (see Results and Discussion) by nonlinear regression.

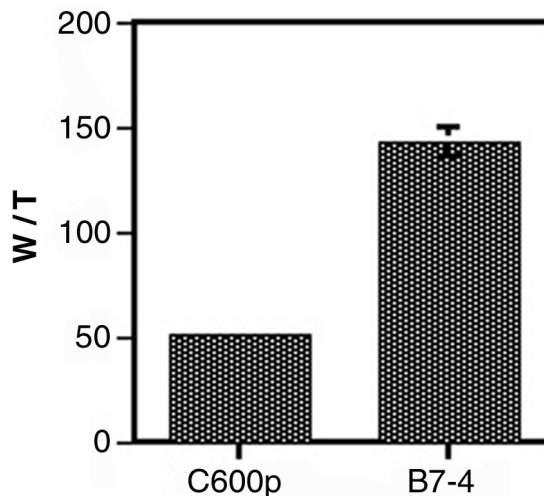


FIG. 10. Relative incorporation of tryptophan by ancestral and evolved clones. Cultures of C600p and B7-4 were grown in M9B1L plus Ap plus limiting concentrations of $[^3\text{H}]\text{W}$ and $[^{14}\text{C}]\text{T}$. After a short pulse, cellular protein was isolated and counted. Error bars represent the standard deviation of triplicate readings.

allowed the general incorporation of aminobutyrate, reaching 24% of the cellular protein (9). However, a *B. subtilis* strain that could incorporate 4fW in preference to tryptophan has also been reported (37), but specific mutations in the evolved *B. subtilis* strain were never identified, and the mechanisms by which the unnatural amino acid was incorporated were not examined. Despite the lack of characterization on the molecular level, however, the apparent ease with which the selection was carried out suggested that a very small number of genes were targeted for mutation under this selective regime. It is for this reason that we had assumed that the mutations found in our evolved organism would mainly be in housekeeping genes and genes encoding proteins with critical tryptophans.

The alteration of the genetic code may have several benefits, including using unnatural amino acids to probe structure-function relationships or augment the capabilities of enzymes and organisms (11, 31). In vitro transcription and translation systems have typically been used to incorporate unnatural amino acids into specific sites in proteins (16, 25, 26). However, the requirement for chemical acylation of a 5'-dCA-3' dinucleotide followed by ligation to a tRNA limits the in vivo application of these otherwise effective techniques. Various in vivo approaches have been attempted. The most successful of these thus far has been the engineering of an orthogonal tRNA-tRNA synthetase pair. The principle of the orthogonal pair is that the tRNA would not be recognized by any other tRNA synthetase, while the tRNA synthetase would not recognize any tRNA other than the engineered partner. In a first attempt at this, tRNA was engineered to be unrecognizable to its cognate synthetase and variants of the synthetase were selected to allow recognition of the mutant tRNA (21). However, the tRNA synthetase pair could not be used for the directed, in vivo incorporation of an unnatural amino acid because the evolved synthetase recognized both wild-type and mutant tRNAs. More recently, a tRNA synthetase pair was imported to *E. coli* cells from yeast (23) and from *Methanococcus jannaschii* (34). The yeast tRNA synthetase pair was completely

orthogonal: while the charged tRNA was recognized by the host translational apparatus, the foreign tRNA was not charged by host synthetases and neither were host tRNAs charged by the foreign synthetase. Furthermore, the *M. jannaschii* tRNA-tRNA synthetase pair was later evolved to be fully orthogonal in *E. coli*, and it showed perfect incorporation of *O*-methyl-L-tyrosine at a stop codon in dihydrofolate reductase (33). Finally, a human tRNA-*E. coli* tRNA synthetase was recently imported into yeast, the first eukaryotic example of an orthogonal pair (19). However, orthogonal tRNA-tRNA synthetase pairs are typically engineered to be extremely noninvasive: the tRNA in the pair is most often a stop codon suppressor. This is an effective technique for completely substituting a specific site in a protein of interest with an amino acid analogue.

These findings suggest that it may be possible to direct the chemistries and biochemistries available to organisms. Since only enzymes involved in tryptophan uptake and incorporation have so far been found to be mutated, it appears as though the *E. coli* proteome may be fairly tolerant of the fluorine-for-hydrogen substitution. Similarly, Ibba and Hennecke have found that the incorporation of *p*-Cl-phenylalanine was at least partially permitted both *in vitro* and *in vivo* by a phenylalanyl-tRNA synthetase with a relaxed substrate specificity (14). Furthermore, it was also found that a tRNA synthetase with a mutation that restricted editing function allowed an increased level of incorporation of amino acid analogues (9). To the extent that organismal proteomes are relatively resilient to chemical modifications, extensive incorporation of other unnatural monomers may also be possible through relatively simple modifications of the transport and translation machinery, similar to those that have already been demonstrated by Schultz and co-workers (21, 22, 33, 34). However, the post facto characterizations of our "unColi" indicate that while organisms can be adapted to exploit novel chemistries by simply changing the enzymes involved in the incorporation of monomers, survival and propagation of novel chemistries will likely require the evolution of the entire organismal genome. Just as the discovery of noncanonical triplet codons in mitochondria suggested that the genetic code might not be sacrosanct, the ability to introduce unnatural amino acids into organismal proteomes provides experimental validation that the 20 natural amino acids may not be the only possible biological building blocks (36).

ACKNOWLEDGMENTS

The NASA Astrobiology Institute was instrumental in initiating and supporting this research.

Louise Wang assisted in the construction of the vectors pETW5w and pETW5m. We thank R. Mursinna and S. Martinis of the University of Houston for assistance with the tRNA charging assays.

REFERENCES

- Altamirano, M. M., J. M. Blackburn, C. Aguayo, and A. R. Fersht. 2000. Directed evolution of new catalytic activity using the α/β -barrel scaffold. *Nature* **403**:617–622.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Ausubel, F. M. 1997. Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology, 3rd ed. Wiley, New York, N.Y.
- Basaggio, N., W. D. Davies, and B. E. Davidson. 1990. Identification of the promoter, operator, and 5' and 3' ends of the mRNA of the *Escherichia coli* K-12 gene *aroG*. *J. Bacteriol.* **172**:2547–2557.
- Browne, D. R., G. L. Kenyon, and G. D. Hegeman. 1970. Incorporation of monofluorotryptophans into protein during the growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **39**:13–19.
- Chow, K. C., and J. T. Wong. 1996. Isolation and characterization of the *Bacillus subtilis* tryptophanyl-tRNA synthetase gene (*trpS*) conferring 5-fluorotryptophan resistance and temperature sensitivity. *Biochim. Biophys. Acta* **1309**:42–46.
- Cobbett, C. S. 1988. Repression of the *aroF* promoter by the TyrR repressor in *Escherichia coli* K-12: role of the 'upstream' operator site. *Mol. Microbiol.* **2**:377–383.
- DeFeyer, R. C., B. E. Davidson, and J. Pittard. 1986. Nucleotide sequence of the transcription unit containing the *aroL* and *aroM* genes from *Escherichia coli* K-12. *J. Bacteriol.* **165**:233–239.
- Döring, V., H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crécy-Lagard, P. Schimmel, and P. Marlière. 2001. Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* **292**:501–504.
- Doublié, S., G. Bricogne, C. Gilmore, and C. W. Carter, Jr. 1995. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure* **3**:17–31.
- Dougherty, D. A. 2000. Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* **4**:645–652.
- Guan, K. L., and J. E. Dixon. 1991. Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **192**:262–267.
- Hamano-Takaku, F., T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Söll, and S. Nishimura. 2000. A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrine more efficiently than tyrosine. *J. Biol. Chem.* **275**:40324–40328.
- Ibba, M., and H. Hennecke. 1995. Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows *in vitro* and *in vivo* synthesis of proteins containing unnatural amino acids. *FEBS Lett.* **364**:272–275.
- Jürgens, C., A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, and R. Sterner. 2000. Directed evolution of a (β/α)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. USA* **97**:9925–9930.
- Kanda, T., K. Takai, T. Hohsaka, M. Sisido, and H. Takaku. 2000. Sense codon-dependent introduction of unnatural amino acids into multiple sites of a protein. *Biochem. Biophys. Res. Commun.* **270**:1136–1139.
- Khodursky, A. B., B. J. Peter, N. R. Cozzarelli, D. Botstein, P. O. Brown, and C. Yanofsky. 2000. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**:12170–12175.
- King, P. V., and R. W. Blakesly. 1986. Optimizing DNA ligations for transformations. *Focus* **8**:30–32.
- Kowal, A. K., C. Köhrer, and U. L. RajBhandary. 2001. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci. USA* **98**:2268–2273.
- Lemeignan, B., P. Sonigo, and P. Marlière. 1993. Phenotypic suppression by incorporation of an alien amino acid. *J. Mol. Biol.* **231**:161–166.
- Liu, D. R., T. J. Magliery, M. Pastrnak, and P. G. Schultz. 1997. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc. Natl. Acad. Sci. USA* **94**:10092–10097.
- Liu, D. R., T. J. Magliery, and P. G. Schultz. 1997. Characterization of an 'orthogonal' suppressor tRNA derived from *E. coli* tRNA₂^{Gln}. *Chem. Biol.* **4**:685–691.
- Liu, D. R., and P. G. Schultz. 1999. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* **96**:4780–4785.
- Mironov, A. A., E. V. Koonin, M. A. Roytberg, and M. S. Gelfand. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res.* **27**:2981–2989.
- Noren, C. J., S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz. 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**:182–188.
- Normanly, J., J. M. Masson, L. G. Kleina, J. Abelson, and J. H. Miller. 1986. Construction of two *Escherichia coli* amber suppressor genes: tRNA^{Phe}_{CUA} and tRNA^{Cys}_{CUA}. *Proc. Natl. Acad. Sci. USA* **83**:6548–6552.
- Parsons, J. F., G. Xiao, G. L. Gilliland, and R. N. Armstrong. 1998. Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* **37**:6286–6294.
- Pittard, A. J. 1996. Biosynthesis of the aromatic amino acids, p. 458–484. *In* F. C. Neidhardt, R. I. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, A. Schaechter, and H. E. Umberger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. American Society for Microbiology, Washington, D.C.
- Pratt, E. A., and C. Ho. 1975. Incorporation of fluorotryptophans into proteins of *Escherichia coli*. *Biochemistry* **14**:3035–3040.
- Sarsero, J. P., P. J. Wookey, and A. J. Pittard. 1991. Regulation of expression

- of the *Escherichia coli* K-12 *trp* gene by TyrR protein and Trp repressor. *J. Bacteriol.* **173**:4133–4143.
31. **Schimmel, P., and D. Söll.** 1997. When protein engineering confronts the tRNA world. *Proc. Natl. Acad. Sci. USA* **94**:10007–10009.
 32. **Sever, S., K. Rogers, M. J. Rogers, C. Carter, Jr., and D. Söll.** 1996. *Escherichia coli* tryptophanyl-tRNA synthetase mutants selected for tryptophan auxotrophy implicate the dimer interface in optimizing amino acid binding. *Biochemistry* **35**:32–40.
 33. **Wang, L., A. Brock, B. Herberich, and P. G. Schultz.** 2001. Expanding the genetic code of *Escherichia coli*. *Science* **292**:498–500.
 34. **Wang, L., T. J. Magliery, D. R. Liu, and P. G. Schultz.** 2000. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122**:5010–5011.
 35. **Webster, G., J. Genschel, U. Curth, C. Urbanke, C. Kang, and R. Hilgenfeld.** 1997. A common core for binding single-stranded DNA: structural comparison of the single-stranded DNA-binding proteins (SSB) from *E. coli* and human mitochondria. *FEBS Lett.* **411**:313–316.
 36. **Wong, J. T.** 1988. Evolution of the genetic code. *Microbiol. Sci.* **5**:174–181.
 37. **Wong, J. T.** 1983. Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. USA* **80**:6303–6306.
 38. **Xu, Z. J., M. L. Love, L. Y. Ma, M. Blum, P. M. Bronskill, J. Bernstein, A. A. Grey, T. Hofmann, N. Camerman, and J. T. Wong.** 1989. Tryptophanyl-tRNA synthetase from *Bacillus subtilis*. Characterization and role of hydrophobicity in substrate recognition. *J. Biol. Chem.* **264**:4304–4311.