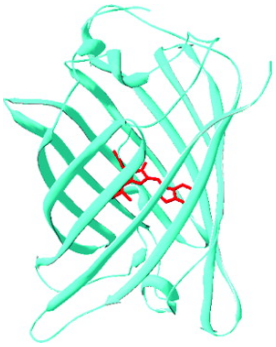
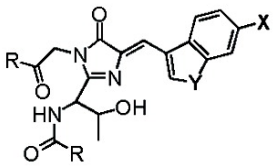
Article

Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells

Inchan Kwon, and David A. Tirrell

J. Am. Chem. Soc., **2007**, 129 (34), 10431-10437• DOI: 10.1021/ja071773r • Publication Date (Web): 08 August 2007





X = H, Cl, Br

Y = S, NH

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/08/2007

Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells

Inchan Kwon and David A. Tirrell*

Contribution from the Division of Chemistry and Chemical Engineering and the Joseph J. Jacobs Institute for Molecular Engineering for Medicine, California Institute of Technology, Pasadena, California 91125

Received March 13, 2007; E-mail: tirrell@caltech.edu

Abstract: A designed yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) activates four tryptophan (Trp) analogues (6-chlorotryptophan (6ClW), 6-bromotryptophan (6BrW), 5-bromotryptophan (5BrW), and benzothienylalanine (BT)) that are not utilized by the endogenous *E. coli* translational apparatus. Introduction of yPheRS (T415G) and a mutant yeast phenylalanine amber suppressor tRNA (ytRNA$^{Phe}_{CUA_UG}$) into an *E. coli* expression host allowed site-specific incorporation of three of these analogues (6ClW, 6BrW, and BT) into recombinant murine dihydrofolate reductase in response to amber stop codons with at least 98% fidelity. All three analogues were introduced at the Trp66 position in the chromophore of a cyan fluorescent protein variant (CFP6) to investigate the attendant changes in spectral properties. Each of the analogues caused blue shifts in the fluorescence emission and absorption maxima. The CFP6 variant bearing BT at position 66 exhibited an unusually large Stokes shift (56 nm). An expanded set of genetically encoded Trp analogues should enable the design of new proteins with novel spectral properties.

Introduction

Tryptophan (Trp) is an attractive target for protein engineering for several reasons. It is a major source of UV absorption and fluorescence in proteins and is involved in important functions including ligand binding[1−3] and DNA−protein interaction.[4,5] Numerous Trp analogues have become available through the versatility of indole chemistry. 7-Azatryptophan and 2-azatryptophan were first incorporated into proteins in *E. coli* in the 1950s,[6,7] and subsequent efforts have enabled the incorporation of Trp analogues containing fluoro, hydroxyl, methyl, amino, selenophene, and thienyl functional groups.[8−10] Fluorinated Trp analogues, in particular, have been widely used in ^{19}F NMR studies.[11,12] Trp analogues containing selenophene and thienyl functional groups have been used in crystallographic structure determination,[13,14] and the spectral properties of aminotryptophan have been exploited in the design of pH sensors[15] and fluorescent proteins.[16]

Further use of Trp analogues has been restricted by two limitations. First, successful incorporation in bacterial cells has been limited to Trp analogues that are utilized by the endogenous *E. coli* translation system. In recent years, several research groups have devised methods to expand the set of amino acid analogues that can be incorporated into recombinant proteins in *E. coli*.[8,10,17] Overexpression of aminoacyl-tRNA synthetases (aaRS),[18,19] attenuation of aaRS editing,[20,21] rational design of the amino acid binding pocket of the aaRS,[22,23] and screening of aaRS libraries[24,25] have all proven effective.

Second, even though Trp is the rarest amino acid, replacement of all Trp residues in a protein can compromise function. For example, replacement of all three Trp residues in barstar with

(1) Fulton, K. F.; Jackson, S. E.; Buckle, A. M. *Biochemistry* **2003**, *42*, 2364−2372.
(2) Uchiyama, T.; Katouno, F.; Nikaidou, N.; Nonaka, T.; Sugiyama, J.; Watanabe, T. *J. Biol. Chem.* **2001**, *276*, 41343−41349.
(3) Bray, M. R.; Johnson, P. E.; Gilkes, N. R.; McIntosh, L. P.; Kilburn, D. G.; Warren, R. A. J. *Protein Sci.* **1996**, *5*, 2311−2318.
(4) Marie, G.; Serani, L.; Laprevote, O.; Cahuzac, B.; Guittet, E.; Felenbok, B. *Protein Sci.* **2001**, *10*, 99−107.
(5) Eason, D. E.; Shepherd, A. T.; Blanck, G. *Biochim. Biophys. Acta−Gene Struct. Express.* **1999**, *1446*, 140−144.
(6) Brawerman, G.; Ycas, M. *Arch. Biochem. Biophys.* **1957**, *68*, 112−117.
(7) Pardee, A. B.; Shore, V. G.; Prestidge, L. S. *Biochim. Biophys. Acta* **1956**, *21*, 406−407.
(8) Budisa, N. *Angew. Chem., Int. Ed.* **2004**, *43*, 6426−6463.
(9) Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893−904.
(10) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603−609.
(11) Pratt, E. A.; Ho, C. *Fed. Proc.* **1974**, *33*, 1463−1463.
(12) Seifert, M. H.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 7932−7942.
(13) Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2001**, *309*, 925−936.
(14) Budisa, N.; Pal, P. P.; Alefelder, S.; Birle, P.; Krywcun, T.; Rubini, M.; Wenger, W.; Bae, J. H.; Steiner, T. *Biol. Chem.* **2004**, *385*, 191−202.
(15) Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. *Angew. Chem., Int. Ed.* **2002**, *41*, 4066−4069.
(16) Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2003**, *328*, 1071−1081.
(17) Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1−11.
(18) Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465−465.
(19) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235−237.
(20) Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635−10645.
(21) Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501−504.
(22) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652−5653.
(23) Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778−11783.
(24) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044−1048.
(25) Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10180−10185.

4-aminotryptophan reduced the stability of the protein because of the increased polarity of the analogue at position 53 in the protein core.[9]

In order to overcome these limitations, we have explored the feasibility of site-specific incorporation of Trp analogues into recombinant proteins in *E. coli*. In 1998, Furter reported site-specific incorporation of *p*-fluorophenylalanine into murine dihydrofolate reductase (mDHFR) in *E. coli* via amber suppression by a yeast phenylalanyl-tRNA synthetase (yPheRS)/ytRNA$^{Phe}_{CUA}$ pair.[26] Although Furter's work established the feasibility of using heterologous synthetase-tRNA pairs for site-specific incorporation of noncanonical amino acids into recombinant proteins, his results were complicated by cross-reactivity of ytRNA$^{Phe}_{CUA}$ with the *E. coli* lysyl-tRNA synthetase.[26] More recently, Schultz and colleagues have developed powerful selection methods to alter the substrate specificity of the tyrosyl-tRNA synthetase (*mj*TyrRS) derived from *Methanococcus jannaschii*[17,24,27] and to enhance the orthogonality of *mj*TyrRS/*mj*tRNA$^{Tyr}_{CUA}$ pairs in *E. coli* expression strains.[28,29] *E. coli* strains equipped with orthogonal *mj*TyrRS/*mj*tRNA$^{Tyr}_{CUA}$ pairs have allowed incorporation of many Phe and Tyr analogues into proteins in site-specific fashion. We have refined the orthogonality of the yPheRS/ytRNA$^{Phe}_{CUA}$ pair and the specificity of yPheRS to achieve high fidelity incorporation of *p*-bromophenylalanine (pBF) into proteins in *E. coli*.[23]

In this report, we describe the design of a bacterial host to achieve site-specific incorporation of Trp analogues into recombinant proteins in response to amber stop codons. Because no orthogonal tryptophanyl-tRNA synthetase/tRNA^{Trp} pairs have been reported for *E. coli* translation systems,[30] we needed a new synthetase/tRNA pair. Our previous studies showed that a designed yPheRS (T415G) variant efficiently activates Trp as well as Phe analogues.[23,31] We imagined that the relaxed specificity of yPheRS (T415G) might also allow it to recognize a variety of Trp analogues. Accordingly, we introduced yPheRS (T415G) and a mutant yeast amber suppressor tRNA (ytRNA$^{Phe}_{CUA_UG}$)[32] into an *E. coli* expression strain to enable insertion of Trp analogues in site-specific fashion. An auxotrophic expression strain was used to limit the intracellular pools of Phe and Trp.[23,31]

Four Trp analogues, BT, 6ClW, 6BrW, and 5BrW (Figure 1) were examined. BT is isosteric to Trp and so might be expected to cause minimal perturbation of protein structure. Introduction of aryl halides into recombinant proteins can allow for site-specific modification through palladium-catalyzed coupling reactions with alkenes and alkynes.[33,34] Bromine and iodine, once introduced into proteins, can facilitate the phasing of crystallographic data and therefore might prove useful in X-ray diffraction studies of protein structure.[19,35] We show here that introduction of the yPheRS (T415G)/ytRNA$^{Phe}_{CUA_UG}$ pair into appropriately engineered *E. coli* expression strains enables high-fidelity (≥98%) incorporation of BT, 6BrW, and 6ClW into recombinant proteins via amber suppression.

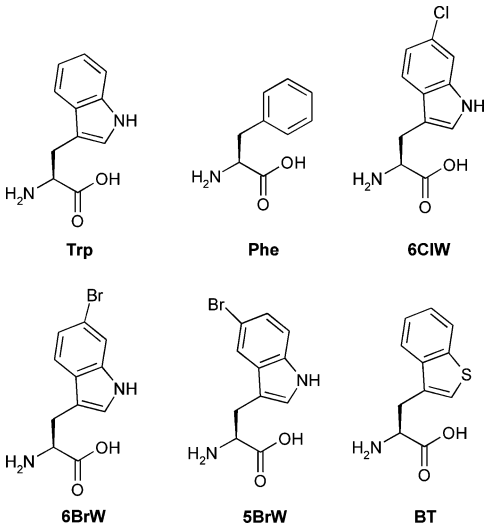


Figure 1. Amino acids involved in this study: tryptophan (Trp); phenylalanine (Phe); 6-chlorotryptophan (6ClW); 6-bromotryptophan (6BrW); 5-bromotryptophan (5BrW); benzothienylalanine (BT).

Materials and Methods

Materials. All 20 natural amino acids were obtained from Sigma (St. Louis, MO); 6ClW and 6BrW were purchased from Biosynth International Inc. (Naperville, IL); 5BrW and BT were from Aldrich (St. Louis, MO) and Chem-Impex (Wood Dale, IL), respectively. [^{32}P]-Labeled sodium pyrophosphate was purchased from NEN Life Science (Boston, MA).

Amino Acid Activation Assays. The kinetics of activation of amino acids by yPheRS (T415G) were determined by the amino acid-dependent adenosine triphosphate (ATP)-[^{32}P]-pyrophosphate (PP$_i$) exchange reaction. The expression and purification of yPheRS (T415G) bearing a C-terminal hexahistidine tag were described previously.[23] The assay buffer contained 50 mM *N*-(2-hydroxethyl)piperazine-*N*′-(2-ethanesulfonic acid) (potassium-HEPES) (pH = 7.6), 20 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 2 mM ATP, and 2 mM [^{32}P]-PP$_i$. The amino acid concentrations varied from 100 nM to 2.5 mM; the enzyme concentration varied from 100 nM to 400 nM. Aliquots (20 μL) of solution were quenched at various time points in 500 μL of buffer solution containing 200 mM NaPP$_i$, 7% w/v $HClO_4$, and 3% w/v activated charcoal. The charcoal was centrifuged and washed three times with 500 μL of 10 mM NaPP$_i$ and 0.5% $HClO_4$ solution. The [^{32}P]-labeled ATP absorbed on the charcoal was counted via liquid scintillation methods. The kinetic parameters were calculated by a nonlinear regression fit of the data to a Michaelis−Menten model.

Strain and Plasmid Construction for in Vivo Incorporation Assays. pQE16am-T415G[23] encodes mDHFR outfitted with a C-terminal hexahistidine tag and an amber codon at position 38 (mDHFR_38Am). The GFP$_{UV}$ gene was cloned from pGFPuv (Clontech) and inserted into the *Pst*I site of pQE9 (Qiagen) to yield pQE9_GFP$_{UV}$. A series of QuikChange mutations (Stratagene) was performed on the GFP$_{UV}$ gene to generate a GFP6 gene containing four

(26) Furter, R. *Protein Sci.* **1998**, *7*, 419−426.
(27) Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836−1837.
(28) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010−5011.
(29) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883−890.
(30) Recently, Schultz and colleagues modified *Bacillus stearothermophilus* trytophanyl-tRNA synthetase (*Bs*TrpRS) to incorporate 5-hydroxytryptophan into a recombinant foldon protein in mammalian cells. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8882−8887.
(31) Wang, P. Ph.D. thesis, California Institute of Technology, Pasadena, 2003.
(32) The mutant yeast amber suppressor was used to reduce lysine misincorporation at the amber sites, because it showed minimal cross-reactivity with *E. coli* lysyl-tRNA synthetase.
(33) Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; Yokoyama, S. *ChemBioChem* **2006**, *7*, 134−139.
(34) Ojida, A.; Tsutsumi, H.; Kasagi, N.; Hamachi, I. *Tetrahedron Lett.* **2005**, *46*, 3301−3305.
(35) Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22*, 1297−1301.

sequence changes (F64L, S65T, S99F, and T153M). Either an amber codon (UAG) or a tryptophan codon (UGG) was inserted at position 66 of GFP6 by QuikChange mutagenesis using two complementary pairs of oligonucleotides, Am66_f (5′-CTT GTC ACT ACT CTG ACC TAG GGT GTT CAA TGC TTT TCC-3′)/Am66_r (5′-GGA AAA GCA TTG AAC ACC CTA GGT CAG AGT AGT GAC AAG-3′) and W66_f (CTT GTC ACT ACT CTG ACC TGG GGT GTT CAA TGC TTT TCC-3′)/W66_r (5′-GGA AAA GCA TTG AAC ACC CCA GGT CAG AGT AGT GAC AAG-3′), to yield pQE9_CFP6_66Am and pQE9_CFP6, respectively. The entire expression cassette from either CFP6_66Am or CFP6 was inserted between the *Aat*II and *Nhe*I sites of pQE16am-T415G to generate pQE9_CFP6_66Am-T415G or pQE9_CFP6-T415G, respectively. pREP4_ytRNA^{Phe}_UG was used to constitutively express a mutant yeast amber suppressor tRNA (ytRNA$^{Phe}_{CUA_UG}$) showing minimal cross-reactivity with the *E. coli* lysyl-tRNA synthetase.[23] A Phe/Trp double auxotrophic strain (AFW) was derived from the Phe auxotrophic strain AF (K10, Hfr(Cavalli) *pheS13rel*-1 *tonA*22 thi *T2*^{R} *pheA*18, *trpB*114) by P1 phage-mediated transposon transduction as described previously.[23,31] *E. coli* strain AFW was co-transformed with either pQE16am-T415G or pQE9_CFP6_66Am-T415G, and pREP4_ytRNA^{Phe}_UG to yield expression strains AFW[pQE16am-T415G/pREP4_ytRNA^{Phe}_UG] and AFW[pQE9_CFP6_66Am-T415G/pREP4_ytRNA^{Phe}_UG], respectively.

In Vivo Incorporation Assays. Expression strains AFW[pQE16am-T415G/pREP4_ytRNA^{Phe}_UG] and AFW[pQE9_CFP6_66Am-T415G/pREP4_ytRNA^{Phe}_UG] were grown in M9 minimal medium supplemented with 0.4% glucose, 35 mg/L thiamin, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 20 amino acids (at 25 mg/L), 35 mg/L kanamycin, and 200 mg/L ampicillin. The overnight culture was diluted 20-fold in fresh M9 minimal medium. The cells were grown to an OD_{600} between 0.9 and 1.0, sedimented by centrifugation, and washed twice with cold 0.9% NaCl. The cell pellet was resuspended in fresh M9 minimal medium containing 18 amino acids (at 25 mg/L), 3 mM analogue of interest, and the indicated concentrations of Phe and Trp. After 10 min, 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce protein expression. After 4 h, cells were pelleted and kept at −80 °C. Whole cell lysates were analyzed by SDS-PAGE. Expression yields were obtained from three independent sets of experiments.

Quantitative Analysis of Analogue Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS). Quantitative analysis of amber site occupancy by various amino acids was performed by LC-MS as described previously.[23] Mutant mDHFR was purified on a Ni-NTA column under denaturing conditions according to the supplier's instructions (Qiagen). After purification, mutant mDHFR concentrations were determined by UV absorbance at 280 nm using a calculated extinction coefficient[36] of 24 750 cm^{-1} M^{-1}. After concentration of the protein by ultrafiltration (Millipore), 10 μL of the solution was diluted into 90 μL of 75 mM $(NH_4)_2CO_3$ for trypsin digestion. Modified trypsin (1 μL, Promega, 0.2 μg/μL) was added, and samples were incubated for 2 h at 37 °C. The reaction was quenched by addition of 13 μL of 5% trifluoroacetic acid (TFA) solution. The solution was then directly subjected to LC-MS/MS analysis. LC-MS/MS analysis of protease-digested peptides was conducted on a QSTAR XL LC-MS/MS system (Applied Biosystems/MDS SCIEX) outfitted with an HPLC system (NanoLC-2DTM, Eksigent Technologies) and ESI probe. Peptides were separated on a C18 reversed phase column (3 μm, 300 Å, 0.1 × 70 mm) and eluted at a flow rate of 0.2 μL/min using a gradient of 0−50% of solvent B (89% acetonitrile/10% H_2O/0.02% TFA/1% formic acid) and solvent A (2% acetonitrile/97% H_2O/0.02% TFA/1% formic acid) in 55 min. The column eluent was transferred to the electrospray source, and sequencing was carried out by fragmentation of the precursor ion corresponding to the peptide containing the residue at position 38 of mutant mDHFR. Occupancy of each amber site was obtained from three independent sets of experiments.

(36) Gill, S. C.; Vonhippel, P. H. *Anal. Biochem.* **1989**, *182*, 319−326.

Fluorescent Protein Isolation. Fluorescent proteins were purified by Ni-NTA affinity chromatography according to the manufacturer's protocol (Qiagen) under native conditions. Purified proteins were desalted over a Sephadex PD-10 column (Amersham Pharmacia, Piscataway, NJ) in 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 (Tris-EDTA buffer). Protein samples other than CFP6 were purified further by ammonium sulfate precipitation.[37] Dry ammonium sulfate was added to the bright fluorescent solution to 40% saturation (about 1.6 M), and the solution was incubated on ice for 25 min. The precipitated proteins were removed by centrifugation at 14 000 rpm for 15 min at 4 °C and resuspended in Tris-EDTA buffer. Dry ammonium sulfate was added to the fluorescent supernatants to 70% saturation (about 2.8 M). The solutions were incubated on ice for 30 min, and the fluorescent proteins were collected by centrifugation at 14 000 rpm for 30 min at 4 °C. The fluorescent proteins were resuspended in Tris-EDTA buffer and desalted over polyacrylamide P-6 columns (Bio-rad) in Tris-EDTA buffer. Protein concentrations were determined by the BCA protein assay (Pierce Biotechnology, Rockford, IL) with bovine serum albumin as a reference. The purity of each CFP variant was determined by SDS PAGE.

Characterization of Spectral Properties. Absorption spectra of fluorescent proteins were measured on a Cary 50 UV−visible spectrophotometer (Varian, Palo Alto, CA). Protein samples were diluted to four different concentrations ($0.1 < OD_{\lambda max} < 0.5$). Extinction coefficients were calculated by applying Beer's law to the absorption spectra of diluted protein samples. A molar mass of 30 287 Da was used for all fluorescent proteins. Fluorescence spectra were measured on a PTI QuantaMaster fluorescence spectrofluorometer (Photon Technology International, Birmingham, NJ) at a scan rate of 1 nm/s. For the quantum yield calculation, protein samples in assay buffer and fluorescein (Invitrogen) in borate buffer (pH 9.1) were adjusted to similar OD ($OD_{\lambda ex} < 0.1$). The fluorescence emission spectra were recorded with excitation at the maxima in the corresponding excitation spectra. The emission spectra were corrected by the correction functions supplied with the instrument. The quantum yields of fluorescent proteins were calculated by comparing the integrated intensities of the corrected emission spectra with that of fluorescein of known quantum yield (0.91).[38]

Results and Discussion

Design of yPheRS and Choice of Trp Analogues. Studies of the crystal structure of PheRS (tPheRS) derived from *Thermus thermophilus* revealed that Val 261 in the amino acid binding pocket is critical for distinguishing Phe from other amino acids.[39] Sequence alignments show that Val 261 in tPheRS corresponds to Thr 415 in yPheRS. Introduction of a single mutation (Thr to Gly) at position 415 in the binding pocket of yPheRS to yield yPheRS (T415G) enables the synthetase to activate a variety of substituted Phe analogues as well as Trp and 2-naphthylalanine.[23,31] We imagined that yPheRS (T415G) might also activate Trp analogues. In order to achieve *site-specific* incorporation, we chose four Trp analogues (6ClW, 6BrW, 5BrW, and BT) that are not utilized by the endogenous *E. coli* translational apparatus (Figure 1). The observations of Budisa and co-workers[8,9] prompted us to propose a simple model for binding of Trp analogues to *E. coli* TrpRS (eTrpRS) (Figure 2a). In the work of Budisa, methyl (ca. 16.8 Å^3),[40] amino (ca. 12.3 Å^3), and fluoro (ca. 3.2 Å^3) substituents were the largest

(37) Yakhnin, A. V.; Vinokurov, L. M.; Surin, A. K.; Alakhov, Y. B. *Protein Expr. Purif.* **1998**, *14*, 382−386.
(38) Velapoldi, R. A.; Tonnesen, H. H. *J. Fluoresc.* **2004**, *14*, 465−472.
(39) Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1534−44.
(40) Volumes were estimated as solvent-excluded volumes (1.4 Å probe) by using ChemBats3D Pro 5.0 (Cambridge, MA).

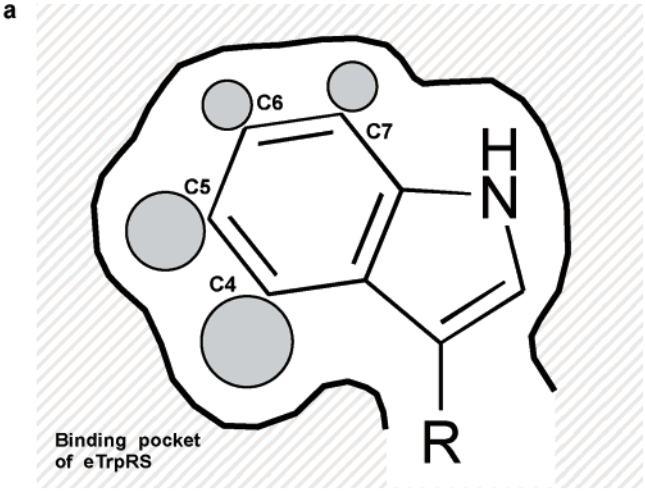


Figure 2. (a) A hypothetical model for accommodation of Trp analogues in the binding pocket of eTrpRS. The position and size of the substituents on the indole ring that can be tolerated in the binding pocket are represented as circles. Indole rings containing larger substituents cannot fit into the pocket. (b) Crystal structure of TrpRS from *Bacillus stearothermophilus* (PDB code 1I6M).[41] Tryptophanyl adenylate (orange) and residues (cyan) within 6 Å of the indole ring of the substrate. The carboxylate (red) of Asp132 forms a hydrogen bond with the nitrogen (blue) of the indole ring.

functional groups tolerated at the C4, C5, and C6 positions of the indole ring, respectively.[8] Introduction of bulkier groups at each position might therefore be expected to yield translationally inactive Trp analogues. The crystal structure of *Bs*TrpRS[41] (Figure 2b) and its sequence homology with eTrpRS supported our hypothetical model. The residues involved in substrate recognition are well conserved, and the binding pocket of *Bs*TrpRS has sufficient space to accommodate bulky substituents at the C4 position of the indole ring but not at the C5 and C6 positions. On the basis of this model, we chose Trp analogues 6ClW, 6BrW, and 5BrW, which contain chlorine (ca. 14.3 Å^3) or bromine (ca. 20 Å^3) at the C6 position, or bromine at the C5 position, of the indole ring. We also examined BT, which has been reported to be translationally inactive in *E. coli*.[42] The indole nitrogen of Trp forms a hydrogen bond with the carboxylate of Asp132 in the binding pocket of *Bs*TrpRS[41]

(41) Retailleau, P.; Yin, Y. H.; Hu, M.; Roach, J.; Bricogne, G.; Vonrhein, C.; Roversi, P.; Blanc, E.; Sweet, R. M.; Carter, C. W. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1595−1608.
(42) Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. *Protein Sci.* **2001**, *10*, 1281−1292.

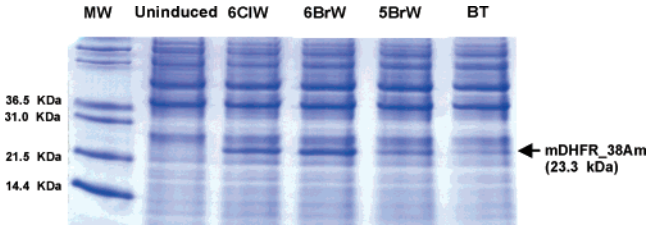


Figure 3. SDS-PAGE analysis of mDHFR_38Am expression. Cells were equipped with the yPheRS (T415G) and the ytRNA$^{Phe}_{CUA_UG}$. Cultures were supplemented with 18 amino acids (at 25 mg/L), one of Trp analogues and the indicated concentrations of Phe and Trp. MW: molecular weight standards; uninduced: before induction; 6ClW: 3.0 mM 6ClW, 0.03 mM Phe and 0.01 mM Trp; 6BrW: 3.0 mM 6BrW, 0.015 mM Phe and 0.005 mM Trp; 5BrW: 3.0 mM 5BrW, 0.03 mM Phe and 0.01 mM Trp; BT: 3.0 mM BT, 0.01 mM Phe and 0.0025 mM Trp.

Table 1. Kinetic Parameters for ATP-PPi Exchange by yPheRS (T415G)

amino acid	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)	k_{cat}/K_m (rel)[a]
Phe	55 ± 14	0.20 ± 0.11	3.6 ± 1.1	1
Trp	2.8 ± 1.6	0.153 ± 0.003	63 ± 35	18
6ClW	0.51 ± 0.21	0.11 ± 0.04	220 ± 20	61
6BrW	3.4 ± 1.3	0.18 ± 0.11	62 ± 54	17
5BrW	0.62 ± 0.02	0.021 ± 0.002	34 ± 4	9.4
BT	8.1 ± 3.7	0.15 ± 0.06	23 ± 18	6.4

[a] Relative to k_{cat}/K_m for Phe by yPheRS (T415G).

(Asp146 in eTrpRS) (Figure 2b), and replacement of the indole nitrogen by other heteroatoms results in translationally inactive amino acids,[13,42] likely due to loss of the favorable hydrogen bond. As expected, none of the four Trp analogues supported protein synthesis in a Trp auxotrophic *E. coli* strain under Trp-depleted conditions.

Activation of Trp Analogues by yPheRS (T415G). The relative rates of activation of Phe, Trp and Trp analogues 6ClW, 6BrW, 5BrW, and BT by yPheRS (T415G) were examined by ATP-PP$_i$ exchange assays. The kinetic parameters are shown in Table 1 and illustrate the varying consequences of substitution at the different positions of the indole ring. Substitution at the 6-position in analogues 6ClW and 6BrW was highly favorable; 6ClW exhibited the highest activity of any of analogues we tested. The difference in the side-chain volumes of Phe and 6ClW (ca. 39 Å^3) is nearly identical to the volume of the cavity generated by the T415G mutation in the binding pocket of yPheRS. Analogue 6BrW yielded a 7-fold larger K_m value than 6ClW, consistent with the fact that the calculated increase in side-chain volume (45 Å^3) upon bromine substitution is larger than the cavity generated by the T415G mutation. Substitution at the 5-position of the indole ring was less favorable than that at the 6-position; the activation rate of 5BrW was half of that of 6BrW. Replacement of the indole nitrogen by a sulfur atom moderately reduced the binding affinity; K_m for BT was 3-fold larger than that of Trp. All of the analogues were activated faster than Phe.

Site-Specific Incorporation of Trp Analogues into mDHFR in Vivo. mDHFR was chosen as the test protein for investigation of the capacity of the yPheRS (T415G)/ytRNA$^{Phe}_{CUA_UG}$ pair to incorporate Trp analogues into recombinant proteins in bacterial cells. An amber stop codon was placed at the 38th position of mDHFR to generate mDHFR_38Am, and the expression of full-length mDHFR_38Am was examined by SDS-PAGE analysis (Figure 3). The intensities of the PAGE

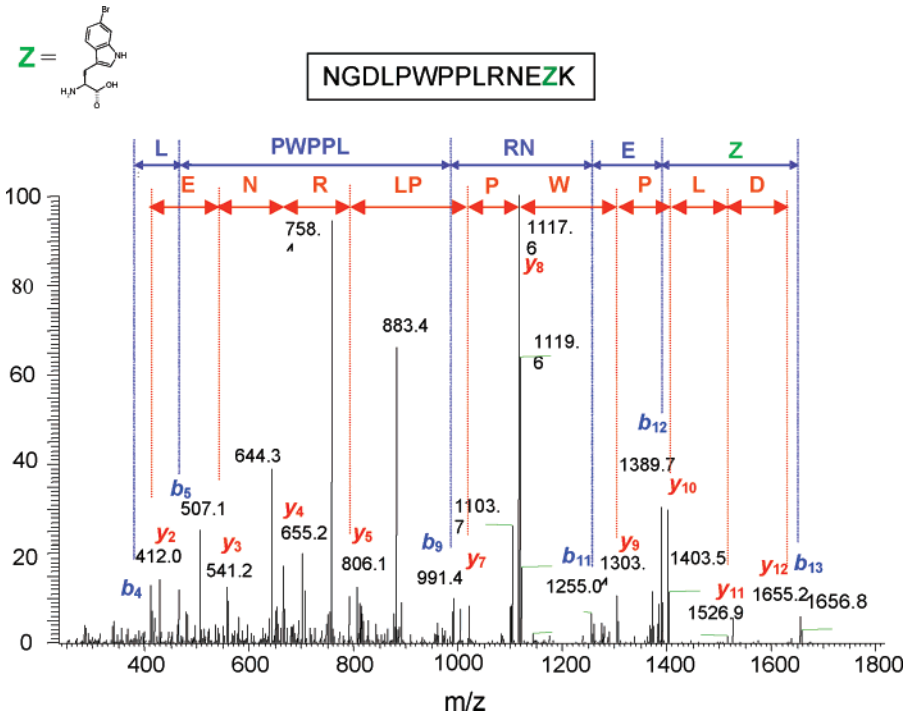


Figure 4. Tandem mass spectrum of the peptide NGDLPWPPLRNEZK. The partial sequence DLPWPPLRNEZ containing 6BrW (Z) can be confirmed by assigning masses of the *b* and *y* series of fragmented ions.

Table 2. Occupancy of Amber Sites and Expression Yields

	occupancy of amber sites (%)				
Trp analogues	Phe	Trp	Lys	Naa	yield (mg/L)[a]
6ClW	0.4 ± 0.5	ND[b]	1.2 ± 1.3	98.4 ± 1.7	5.4 ± 0.3
6BrW	0.4 ± 0.7	ND	ND	99.6 ± 0.7	6.8 ± 1.3
5BrW	97.0 ± 1.4	0.3 ± 0.6	0.7 ± 0.5	1.8 ± 0.7	3.6 ± 0.6
BT	0.6 ± 0.6	0.7 ± 0.6	0.9 ± 1.5	97.9 ± 2.4	2.8 ± 0.6

[a] Volumetric yields are given as mg of purified mDHFR_38Am per liter of culture. [b] Not detected in LC-MS analysis.

bands due to the full-length protein decreased in the order 6BrW > 6ClW > 5BrW > BT. Although expression of full-length mDHFR_38Am indicated that the stop codon was suppressed, it is possible that suppression occurs by insertion of Phe or Trp at the amber site. The occupancy of the amber site in mDHFR_38Am was determined by LC-MS/MS analysis. Trypsin digestion of purified mDHFR_38Am variants yielded a peptide containing the amber site (residues 26−39, NGDLPWPPLR-NEAmK, designated peptide 38). Peptide 38 variants containing Phe, Trp, or analogues were separated on a C_{18} reversed phase column. Relative yields of peptide 38 variants were determined by comparing the integrated areas of the corresponding signals in the chromatogram, and the identity of each peptide was confirmed by tandem mass spectrometry. The results are summarized in Table 2. As a representative case, the identity of peptide 38 containing 6BrW at the amber site was determined from the spectrum shown in Figure 4. We found that 6ClW, 6BrW, and BT occupied at least 98% of the amber sites in full length mDHFR_38Am (Table 2). In contrast, 5BrW occupied only 2% of the amber sites, even though it was activated efficiently by yPheRS (T415G). Instead, 97% of the amber sites were occupied by Phe. The discrepancy between in vitro activation and in vivo incorporation for 5BrW can be rationalized on the basis of model proposed by Sisido and co-workers for accommodation of non-canonical amino acids by the *E. coli* ribosome.[43] The model defines geometries of aryl-*L*-alanines that are permissive with respect to translation. We suggest that the addition of bromine at the C5 position of the indole ring may result in exclusion of 5BrW from the ribosome.

It is interesting to note that at similar (low) concentrations of Phe and Trp in the culture medium, Phe was inserted at the amber site in preference to Trp. In our in vitro activation assays, Trp was activated 18-fold faster than Phe. The discrepancy between these results might be explained by induction of Trp degradation at high concentrations of the Trp analogues. Expression of tryptophanase is regulated by tryptophan-induced transcription antitermination in *E. coli*,[44,45] and it is known that 1-methyltryptophan, a Trp analogue, is an effective inducer in vitro.[45] Although we did not explore this mechanism, it may be that addition of 3 mM 5BrW to the culture medium leads to depletion of intracellular Trp.

Expression yields of mDHFR_38Am prepared with the various Trp analogues are summarized in Table 2. It was somewhat surprising that 6BrW afforded higher expression levels than 6ClW, despite its lower k_{cat}/K_m value. This result may reflect the higher value of k_{cat} for 6BrW as compared to 6ClW. The concentration (3 mM) of each analogue in the medium was well above the value of K_m for activation in vitro; the relative rates of charging of the analogues in the cell may therefore parallel k_{cat} rather than k_{cat}/K_m.

We could find no evidence (by MALDI-MS analysis of tryptic fragments) for incorporation of any of the analogues into mDHFR in response to Phe or Trp codons (Figures S1 and S2).

(43) Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 34−40.
(44) Gong, F.; Ito, K.; Nakamura, Y.; Yanofsky, C. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8997−9001.
(45) Gong, F.; Yanofsky, C. *J. Biol. Chem.* **2001**, *276*, 1974−1983.
(46) Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS Lett.* **2000**, *467*, 37−40.

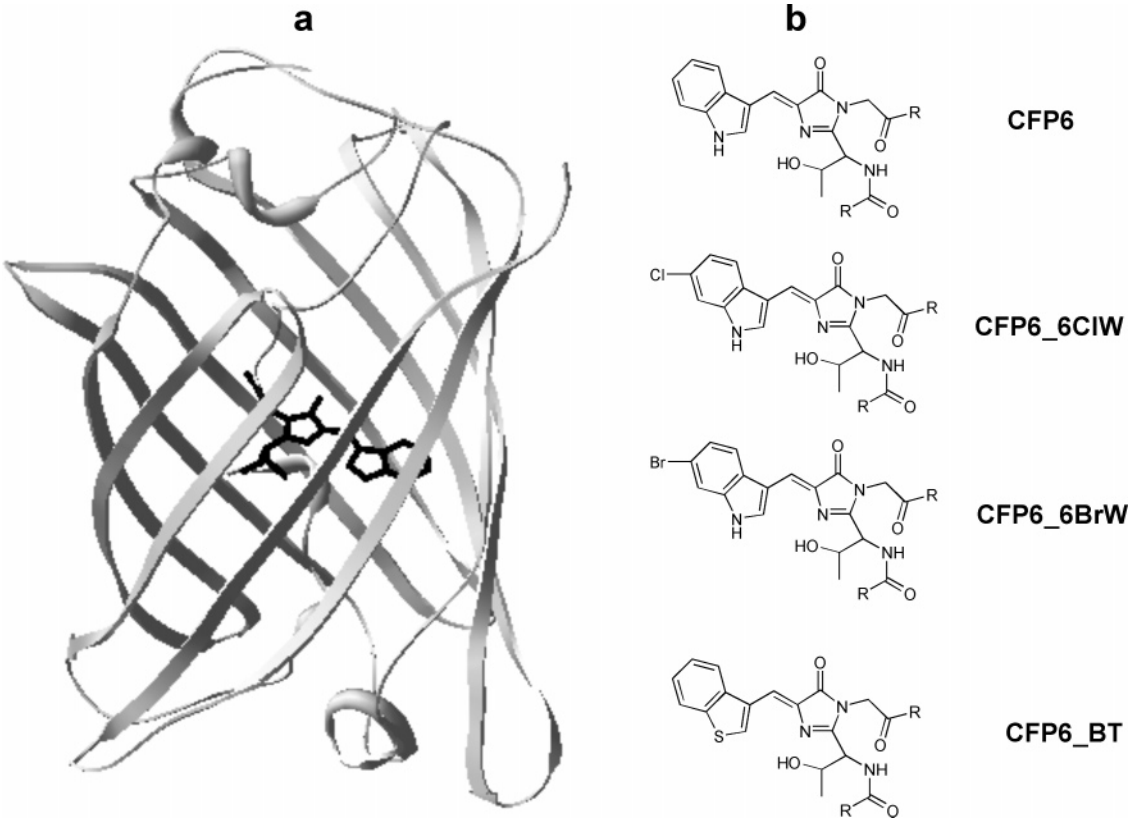


Figure 5. Crystal structure of enhanced cyan fluorescent protein (PDB code 1OXD)[16] (a). The chromophore is shown in black. Chromophores of CFP6 variants containing Trp, 6ClW, 6BrW, and BT (b).

Site-Specific Incorporation of Trp Analogues into Fluorescent Proteins in Vivo. The *E. coli* strain developed in this work has been used to engineer the spectral properties of fluorescent proteins. Schultz and co-workers have characterized the spectral properties of green fluorescent protein (GFP) variants containing Phe analogues in the chromophore.[47] The fluorescence emission maxima of all such variants were blue-shifted to some extent, depending on the electron-donating ability of the inserted analogue. Budisa and co-workers replaced the Trp residues in either GFP or cyan fluorescent protein (CFP) with translationally active Trp analogues[9] and found that the fluorescence emission maximum of the 4-aminotryptophan variant was red-shifted significantly.[16]

In this work, we explored the design of fluorescent proteins containing Trp analogues that are inactive with respect to the endogenous *E. coli* translation apparatus. We replaced the Trp residue in the chromophore of CFP6 with 6ClW, 6BrW, and BT to generate CFP6_6ClW, CFP6_6BrW, and CFP6_BT, respectively (Figure 5). After affinity chromatography, the purity of CFP6 was better than 95%; the purities of CFP6_6ClW, 6BrW, and BT were ca. 90%. The spectral properties of CFP6 and of the CFP6_66Am proteins containing Trp analogues are summarized in Table 3. The observed absorption maximum, emission maximum, and extinction coefficient of CFP6 are consistent with literature values.[48] Compared to CFP6, CFP6_6ClW and CFP6_6BrW showed blue shifts of 14 nm in the fluorescence emission maxima (Figure 6). These observations are consistent with the results of Schultz and co-workers, who found that replacement of tyrosine at position 66 with pBF[47] caused a blue shift in the emission maximum. The fluorescence quantum yields of CFP6_6ClW and CFP6_6BrW are 3- to 4-fold higher than that of the pBF variant, but 2.5- to 3-fold lower than that of CFP6, probably because of perturbation of spin-orbital coupling in the π-electron orbitals of the fluorophore by the halogen atom.[49] CFP6_BT showed blue shifts of 16 and 34 nm in the fluorescence emission and absorption maxima, respectively. Similar blue shifts in the absorption maximum of CFP were observed by replacing Trp at position 66 with the sulfur heterocyles β-(thieno[2,3-*b*]pyrrolyl)-L-alanine ([2,3]Tpa) and β-(thieno[3,2-*b*]pyrrolyl)-L-alanine ([3,2]Tpa).[14] The extinction coefficient of CFP6_BT (8100 M^{-1} cm^{-1}) is similar to those of the CFP variants bearing either [2,3]Tpa (6840 M^{-1} cm^{-1}) or [3,2]Tpa (6350 M^{-1} cm^{-1}). But introduction of either [2,3]-Tpa or [3,2]Tpa resulted in abolition of CFP fluorescence;[14] in contrast CFP6_BT exhibits substantial emission, with a fluorescence quantum yield only 3-fold smaller than that of CFP6. CFP6_BT offers several advantages as a fluorescence resonance energy transfer (FRET) partner. It can be efficiently excited by a common violet diode laser (405−415 nm), which can simplify

Table 3. Spectral Properties and Protein Yield of CFP6 Variants

amino acid at position 66	absorption maximum (nm)	extinction coefficient (M^{-1} cm^{-1})	emission maximum (nm)	fluorescence quantum yield	protein yield (mg/L)
Trp	449	25900 ± 1000	488	0.14 ± 0.008	3.9
6ClW	430	19800 ± 6400	474	0.050 ± 0.001	1.0
6BrW	430	20600 ± 3800	474	0.058 ± 0.001	0.89
BT	415	8100 ± 1100	472	0.048 ± 0.008	0.73

(47) Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174−176.
(48) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.
(49) Kasha, M. *J. Chem. Phys.* **1952**, *20*, 71−74.

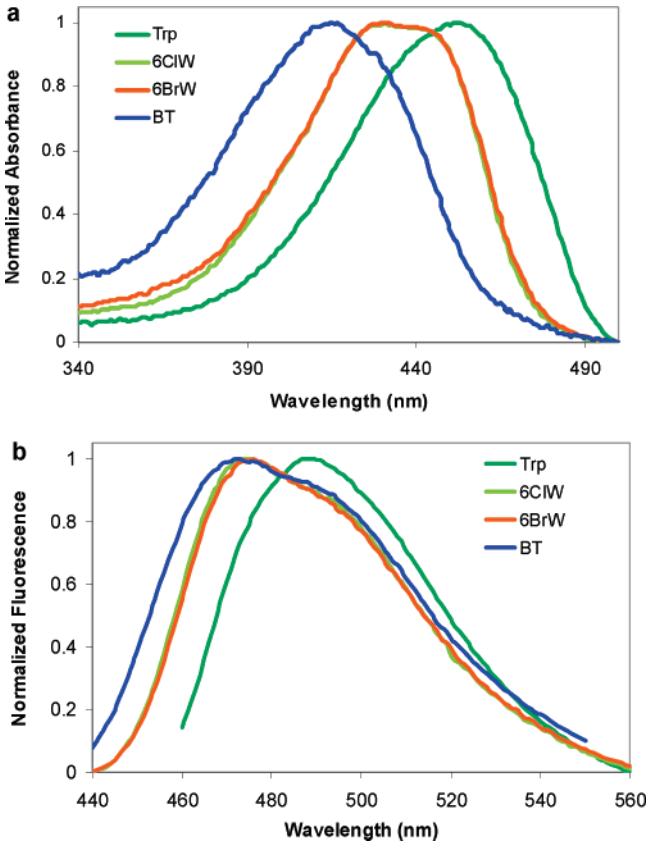


Figure 6. Absorption (a) and fluorescence emission spectra (b) for CFP6 variants bearing Trp analogues at the Trp66 position.

cell screening by FRET analysis. In addition, the Stokes shift of CFP6_BT (56 nm) is larger than that of CFP6 (37 nm). The large Stokes shift is advantageous for fluorescence measurements, since it allows emitted photons to be detected against a reduced background. CFP and yellow fluorescent protein (YFP) have been commonly used as a FRET pair.[50] Because the fluorescence emission spectrum of CFP_BT (Figure 6 b) overlaps the excitation spectrum of GFP, it constitutes a promising FRET partner for GFP.

Conclusions

Properly engineered *E. coli* strains allow high-fidelity incorporation of Trp analogues into recombinant proteins in response to programmed amber stop codons. A simple modification of the binding pocket of yPheRS enabled activation of Trp analogues 6ClW, 6BrW, 5BrW, and BT. When the engineered yPheRS and a yeast amber suppressor tRNA were introduced into a Phe/Trp auxotrophic *E. coli* expression host, at least 98% of the amber sites in recombinant mDHFR were occupied by Trp analogues 6ClW, 6BrW, and BT. In a subsequent experiment, three Trp analogues (6ClW, 6BrW and BT) were introduced at the Trp66 position of CFP6. CFP6 variants containing Trp analogues in the chromophore showed blue shifts in fluorescence emission and absorption maxima. In particular, CFP6_BT exhibited an unusually large Stokes shift (56 nm). An expanded set of genetically encoded Trp analogues should enable design of new proteins with novel and useful properties.

Acknowledgment. We thank I. C. Tanrikulu, R. Connor, and T. H. Yoo for valuable discussions and Dr. J. Zhou for help with LC-MS analysis. This work was supported by National Institutes of Health grant GM62523.

Supporting Information Available: MALDI-MS spectra of tryptic fragments of mDHFR. This material is available free of charge via the Internet at http://pubs.acs.org.

JA071773R

(50) Pollok, B. A.; Heim, R. *Trends Cell Biol.* **1999**, *9*, 57−60.