Chapter 4

Site-Specific Incorporation of Tryptophan Analogs into Recombinant Proteins in Vivo

Abstract

An expansion of the number of genetically encoded tryptophan (Trp) analogs in *Escherichia coli* has been restricted by two limitations. First, Trp analogs inactive in the native *E. coli* translation system could not be incorporated into proteins. Second, most of the previous attempts to incorporate Trp analogs have been limited to global replacement of all Trp residues throughout a target protein. Here we show that a rationally designed yeast phenylalanyl-tRNA synthetase (yPheRS (T415G)) efficiently activated four Trp analogs: 6-chlorotrytophan (6ClW), 6-bromotryptophan (6BrW), 5-bromotrytophan (5BrW), and benzothienylalanine (BT), which were not utilized by the endogenous *E. coli* translational system. Introduction of the yPheRS (T415G) and a mutant yeast phenylalanine amber suppressor ($ytRNA^{Phe}_{CUA\_UG}$) into an *E. coli* expression host allowed the site-specific incorporation of three Trp analogs: 6ClW, 6BrW, and BT, into murine dihydrofolate reductase in response to amber stop codons with at least 98% fidelity. Subsequently, all three Trp analogs were introduced at the Trp66 position in the chromophore of a cyan fluorescent protein (CFP) to investigate changes in spectral properties. CFP variants containing Trp analogs showed a blue-shift in the fluorescence emission peak as well as in the absorption maximum. In particular, a CFP variant with BT featured an unusually large Stokes' shift (56 nm). As shown in this report, an expanded set of the genetically encoded Trp analogs would enable the design of new proteins with novel properties.

Introduction

Tryptophan (Trp) is an attractive target for protein engineering for several reasons. First, it is the main source of UV absorbance and fluorescence of proteins. Second, Trp is involved in various interactions in proteins, such as ligand binding[1-3] and DNA-protein interaction.[4,5] Third, numerous Trp analogs are available through diverse indole chemistry. Fourth, it is the rarest amino acid,[6] and so its substitution by other amino acids minimally perturbs structure.[7] Ever since 7-azatryptophan and 2-azatryptophan were first incorporated into proteins in *E. coli* in the 1950s,[8,9] many researchers have tried to replace Trp with Trp analogs. In the last few decades, a residue-specific incorporation method utilizing a Trp auxotrophic strain realized the incorporation of various Trp analogs containing fluoro, hydroxyl, methyl, amino, selenophene, and thienyl functional groups.[10-12] Fluorinated Trp analogs, in particular, have been used in ^{19}F-NMR studies.[13,14] Trp analogs containing selenophene and thienyl functional groups were used in the X-ray crystallography of proteins.[15,16] Aminotryptophan was used to design protein-based pH sensors[17] and greatly changed the spectral properties of fluorescent proteins.[18]

Although diverse Trp analogs have been introduced into proteins, further expansion of the genetic code for Trp has been restricted by two limitations. First, Trp analogs incorporated into proteins in *E. coli* have so far been limited to the substrates for the native *E. coli* translation system. Therefore, a large number of Trp analogs, such as benzothienylalanine (BT), 5-methyltryptophan, 6-methyltryptophan, 6-aminotryptophan, and 5-cyanotryptophan, were found not to be incorporated into proteins.[10] Recently, several research groups have devised various methods to introduce translationally

inactive nonnatural amino acids into recombinant proteins in *E. coli*.[10,12,19] Overexpression of aminoacyl-tRNA synthetase (aaRS),[20,21] utilization of attenuated editing mutant aaRS,[22,23] rational design of the binding pocket of aaRS,[24,25] and screening of an aaRS library[26,27] have been effective in expanding the genetic code for leucine, isoleucine, valine, methionine, phenylalanine, and tyrosine. Recently, Schultz and colleagues modified *Bacillus stearothermophilus* trytophanyl-tRNA synthetase (*Bs*TrpRS) to incorporate 5-hydroxytryptophan into a recombinant protein in mammalian cells.[28] However, until now there have been no reports about modification of substrate specificity of aaRS to expand the number of genetically encoded Trp analogs in *E. coli*. In this report, we have shown that the rational design of the binding pocket of yPheRS can expand the genetic code for Trp in *E. coli*.

Second, most of the previous attempts to incorporate Trp analogs have been limited to the global replacement of Trp throughout a protein. Although Trp is the rarest amino acid, the replacement of Trp with Trp analogs at non-permissive sites could result in the loss of function. For example, the replacement of three Trp residues in barstar with 4-aminotryptophans greatly reduced its stability due to increased polarity of 4-aminotryptophan at position 53 inside the protein core.[11] In order to overcome such a limitation, we have explored the feasibility of site-specific incorporation of Trp analogs into recombinant proteins in *E. coli*, which would allow incorporation of Trp analogs only at permissive sites. In 1998, Furter showed site-specific incorporation of *p*-fluorophenylalanine into murine dihydrofolate reductase (mDHFR) in *E. coli* through the use of an orthogonal yPheRS/ytRNA^{Phe} pair.[29] More recently, Schultz and colleagues have developed powerful screening methods to change the substrate specificity of

tyrosyl-tRNA synthetase (*mj*TyrRS) derived from *Methanococcus jannaschii* toward nonnatural amino acids.[19,26,30] *E. coli* strains equipped with the screened orthogonal *mj*TyrRS/*mj*tRNA^{Tyr} pairs have allowed incorporation of diverse Phe and Tyr analogs into proteins in a site-specific manner. Tirrell and colleagues refined the orthogonality of the yPheRS/ytRNA^{Phe} pair and the specificity of yPheRS to achieve high fidelity incorporation of *p*-bromophenylalanine into proteins *in vivo*.[25]

In this report, we describe the design of a bacterial host to achieve site-specific incorporation of Trp analogs into a recombinant protein in response to an amber stop codon in *E. coli*. Since no orthogonal tryptophanyl-tRNA synthetase/tRNA^{Trp} pairs have been reported for *E. coli* translation systems, we needed an alternate orthogonal pair. Our previous studies showed that a rationally designed yPheRS (T415G) variant efficiently activates Trp as well as various Phe analogs.[25,31] We hypothesized that the relaxed substrate specificity of yPheRS (T415G) may allow it to recognize a variety of Trp analogs. Therefore, a mutant yPheRS (T415G) and a mutant yeast amber suppressor (ytRNA$^{Phe}_{CUA_UG}$)[32] were introduced into an *E. coli* expression strain to insert Trp analogs at a specific site in a protein. Four Trp analogs, BT, 6ClW, 6BrW, and 5BrrW, were examined for incorporation *in vivo*. In our preliminary studies, none of the four Trp analogs tested were incorporated at UGG codons, even in a Trp auxotrophic *E. coli* strain under a Trp-depleted condition (data not shown). BT is isosteric to Trp and so might result in minimal perturbation of a protein structure upon the replacement of Trp. Introduction of aryl halides into recombinant proteins can allow for site-specific modification through a 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 can be useful in X-ray diffraction studies of protein structures.[35] As expected, introduction of an orthogonal yeast PheRS (T415G) /ytRNA$^{Phe}_{CUA_UG}$ pair into *E. coli* expression strains enabled high-fidelity (ca. 98%) incorporation of BT, 6BrW, and 6ClW into mDHFR in a site-specific manner.

The *E. coli* strain, equipped with an engineered orthogonal pair, has been used to change the spectral properties of fluorescent proteins. Considering that Trp is a main contributor to UV absorbance and the fluorescence of proteins, it is not surprising that the replacement of Trp with other amino acids resulted in changes in the spectral properties of proteins in many different ways. Schultz and his colleagues have characterized the spectral properties of green fluorescent protein (GFP) variants containing various Phe analogs in the chromophore.[36] Fluorescence emission peaks of all GFP variants were blue-shifted to some extent, depending on the electron-donating ability of the Phe analogs. Similarly, Budisa and his colleagues replaced Trp residues in either GFP or the cyan fluorescent protein (CFP) with translationally active Trp analogs.[11] In particular, the fluorescence emission peak of 4-aminotryptophan in the chromophore of a fluorescent protein was red-shifted significantly.[18] In this report, we have explored whether we could design novel spectral classes of fluorescent proteins containing Trp analogs that would be translationally inactive for the endogenous *E. coli* translation system. Specifically, we replaced a Trp residue in the chromophore of a CFP variant (CFP6) with 6ClW, 6BrW, and BT to generate CFP6_6ClW, CFP6_6BrW, and CFP6_BT, respectively. Similar to halogenated Phe analogs, fluorescence emission peaks of CFP6_6ClW and CFP6_6BrW were blue-shifted. However, quantum yields of halogenated Trp analogs were 3- to 4-fold higher than their Phe counterparts. CFP6_BT exhibited a larger Stokes' shift (56 nm) than

that of CFP6 (ca. 40 nm).

Materials and Methods

Materials. Amino acids **1-2** were obtained from Sigma (St. Louis, MO). **3** and **4** were purchased from Biosynth International (Naperville, IL). **5** and **6** 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 Assay. The kinetics of activation of amino acids by the yPheRS (T415G) variant were determined by the amino acid dependent adenosine triphosphate (ATP)-[^{32}P]-pyrophosphate (PP_i) exchange reaction. The plasmid construction, expression, and purification of the yPheRS (T415G) variant with C-terminal hexa-histidine tag were described previously.[25] 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 and 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[25] encodes a murine dihydrofolate reductase (mDHFR) outfitted with a C-terminal hexa-histidine tag and an amber codon at position 38 . The GFP_{UV} gene was cloned from pGFPuv (Clontech) and inserted into the *Pst*I site of pQE9 (Qiagen) to yield a pQE9_GFP_{UV} plasmid. A series of QuikChange mutations (Stratagene) were performed on GFP_{UV} gene to generate a GFP6 gene containing four sequence changes (F64L, S65T, S99F, and T153M) (see Chapter 6). 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$^{Phe}_{CUA_UG}$) shows minimal cross-reactivity with the *E. coli* lysyl-tRNA synthetase.[25] 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.[25,31] The double auxotrophic strain, AFW, was co-transformed with either pQE16am-T415G or pQE9_CFP6_66Am-T415G, and with pREP4_ytRNA^{Phe}_UG to yield expression strains

AFW[pQE16am-T415G/pREP4_ytRNA^{Phe}_UG] or AFW[pQE9_CFP6_66Am-T415G/pREP4_ytRNA^{Phe}_UG], respectively. DH10B (Invitrogen) *E. coli* strain was co-transformed with pQE9_CFP6-T415G and pREP4_ytRNA^{Phe}_UG to generate DH10B[pQE9_CFP6-T415G/pREP4_ytRNA^{Phe}_UG].

In Vivo Incorporation Assays. *E. coli* expression strains were grown in M9 minimal medium supplemented with 0.4% glucose, 35 mg/L thiamin, 1mM $MgSO_4$, 1mM $CaCl_2$, 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 analog of interest, and the indicated concentrations of Phe and Trp. After 10 min, 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) was added to induce protein expression. After 4 hours, cells were pelleted and kept at -80 °C.

A different procedure was used for expression strain DH10B[pQE9_CFP6-T415G/pREP4_ytRNA^{Phe}_UG]. A single colony was inoculated into 5 mL of 2xYT medium and grown overnight. 5 mL of the overnight culture was transferred to 100 mL of fresh 2xYT medium. When the cultures reached an OD_{600} of 0.6, 1 mM IPTG was added to induce protein expression. After 3 hours, cells were sedimented and stored at -80 °C. Whole cell lysates were analyzed by SDS-PAGE.

Quantitative Analysis of Analog Incorporation by Liquid Chromatography Mass Spectrometry (LC-MS). Quantitative analysis of amber codon occupancy by various amino acids was performed by LC-MS as described previously.[25] In a minor alteration, mutant mDHFR was purified on a Ni-NTA spin 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[37] 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. 1 μL of modified trypsin (Promega, 0.2 μg/μL) was added. The samples were incubated for 2 hrs at 37 ^oC. 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) with HPLC system (NanoLC-2DTM, Eksigent Technologies) and ESI probe. Peptides were separated on a C18 reversed column (3 μm, 300Å, 0.1 x 70mm) 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.

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) using 20 mM Tris-HCl, 1 mM EDTA, pH 8.0. Protein samples other than CFP6 were further purified by ammonium sulfate precipitation.[38] Dry ammonium sulfate was added to the bright fluorescent solution to 40% saturation (about 1.6 M). Then solutions were incubated on ice for 25 min and the precipitated proteins were removed by centrifugation at 14,000 rpm for 15 min at 4 ^{o}C. 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 ^{o}C. The fluorescent proteins were resuspended in 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 and desalted over polyacrylamide P-6 columns (Bio-rad) using 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 buffer. Protein concentrations were determined by the BCA protein assay kit (Pierce Biotechnology, Rockford, IL) using bovine serum albumin as a reference.

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).[39]

Results and Discussion

Rational Design of yPheRS and Choice of Trp Analogs. In order to increase the number of genetically encoded Trp analogs, we engineered the substrate specificity of yPheRS. 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 non-aromatic amino acids.[40] Sequence alignments showed 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 (yPheRS (T415G)) created extra room to accommodate various Phe analogs.[25,31] There was an interesting finding that yPheRS (T415G) activated Trp and 2-naphthylalanine as well as substituted Phe analogs. Considering the spectrum of nonnatural amino acids recognized by yPheRS (T415G), we hypothesized that yPheRS (T415G) may activate Trp analogs. In order to achieve site-specific incorporation of Trp analogs, we needed to choose Trp analogs that can be recognized by yPheRS (T415G) but by the native *E. coli* translation system. We have chosen four Trp analogs (6ClW, 6BrW, 5BrW, and BT) that can not be utilized by the endogenous *E. coli* translational system. Based on previous data,[10,11] we proposed a model for the adaptability of Trp analogs to the binding pocket of *E. coli*

TrpRS (eTrpRS) (Figure 2). Recognition of Trp analogs by eTrpRS depends on the size and position of substituents on the indole ring. Methyl (ca. 16.8 $Å^3$),[41] amino (ca. 12.3 $Å^3$), and fluorine groups (ca. 3.2 $Å^3$) were the largest functional groups tolerated at the C4, C5 and C6 positions of the indole ring, respectively.[10] Therefore, introduction of bulkier groups at each position would result in translationally inactive Trp analogs. Based on our hypothetical model, we chose three translationally inactive Trp analogs, **3**, **4**, and **5** containing chlorine (ca. 14.3 $Å^3$) or bromine (ca. 20 $Å^3$) at the C6 position, and bromine at the C5 position of the indole ring. Additionally, we chose **6,** known as an translationally inactive Trp analog.[42] Replacement of the nitrogen in the indole ring of Trp with other hetero-atoms, such as selenium, sulfur, or oxygen, results in translationally inactive amino acids.[15,42] As expected, none of the four Trp analogs (**3**-**6**) were incorporated into proteins, even in Trp auxotrophic *E. coli* strain under Trp depleted conditions (data not shown).

Activation of Trp Analogs by yPheRS (T415G). The relative rates of activation of Phe, Trp and Trp analogs (**1**-**6**) by yPheRS (T415G) were examined by ATP-PP_i exchange assays. The kinetic parameters are shown in Table 1. Derivatives of Trp (**3**-**6**) revealed different activities with respect to substitution at different positions of the indole ring. Substitution at position 6 (**3** and **4**) was highly favorable; **3** displayed the highest activity among all the analogs we tested. The increase in side-chain volume upon replacement of **2** with **3** (ca. 39 $Å^3$) matched very well the cavity (ca. 39 $Å^3$) generated by the T415G mutation in the binding pocket of yPheRS. **4** exhibited a k_{cat}/K_m value comparable to that of Trp. Notably, **4** showed 7-fold larger K_m value than that of **3,** indicating less favorable

binding of **4** to the binding pocket of yPheRS. This result is consistent with the fact that the calculated increase in side-chain volume (45 $Å^3$) upon replacement of **2** with **4** is larger than the cavity generated by the T415G mutation. Substitution at position 5 of the indole ring was less favorable than that at position 6. The activation rate of **5** by yPheRS (T415G) was one half of that of **4**. Replacement of an imine in the indole ring with a sulfur atom reduced binding to yPheRS (T415G). The K_m of **6** was 3-fold greater than that of **1**. However, the activation rate of **6** by yPheRS (T415G) was 6.4-fold faster than that of **2**. The spectrum of activity of the yPheRS (T415G) supported our proposal that the yPheRS (T415G) may recognize Trp analogs.

Site-Specific Incorporation of Trp Analogs into mDHFR in Vivo. To investigate the utility of yPheRS (T415G) to incorporate Trp analogs into proteins in vivo, mDHFR was chosen as a test protein. An amber stop codon was placed at the 38th position of mDHFR to generate mDHFR_38Am. The expression of full length mDHFR_38Am with various Trp analogs (**3**-**6**) was examined by SDS-PAGE analysis (Figure 3). All of the Trp analogs supported expression of full length mDHFR_38Am. The band intensities of mDHFR_38Am variants were in the order of **4** > **3** > **5** > **6**. Although expression of full length mDHFR_38Am strongly indicated that the amber stop codon was suppressed, it should be noted that the mDHFR_38Am expression does not necessarily mean that the Trp analogs were incorporated at the amber site. In order to verify incorporation of Trp analogs at the amber site, the occupancy of the site in mDHFR_38Am by various amino acids was determined by LC-MS/MS analysis. Trypsin digestion of purified mDHFR_38Am variants generated diverse peptide fragments. Among them, a peptide

containing the amber site (residues 26-39, NGDLPWPPLRNEAmK) was designated as peptide 38. Peptide 38 variants containing Phe, Trp or Trp analogs were separated on a C_{18} reversed phase column and relative yields were determined by comparing the integrated areas of the corresponding signals in the chromatogram. The results are summarized in Table 2. Furthermore, the identity of each peptide was confirmed by tandem-mass spectrometry. As a representative, the identity of peptide 38 containing 6BrW at the amber site was determined from the spectrum shown in Figure 4. We observed that **3**, **4**, and **6** occupied at least 98% of the amber sites in full length mDHFR_38Am (Table 2). However, **5** occupied only 2% of the amber site, though it was efficiently activated by yPheRS (T415G). Instead, **2** occupied 97% of the amber site. The discrepancy between in vitro activation and in vivo incorporation for **5** could be rationalized by the model proposed by Sisido and colleagues for the adaptability of non-natural amino acids to the *E. coli* ribosome.[43] The model suggests that the *E. coli* ribosome has the ability to define the allowed and excluded regions of the aromatic groups of aryl-*L*-alanines to be bound for further translation. The addition of bromine increases the volume of the indole ring by 20 $Å^3$. We reason that the addition of bromine at the C5 position of the indole ring may result in exclusion of **5** by the ribosome. Another interesting finding from the experiment for analog **5** is that at similarly low concentrations of Phe and Trp in the culture medium, Phe was incorporated into the amber site instead of Trp. According to our in vitro activation assays, Trp was activated 18-fold faster than Phe by yPheRS (T415G). This discrepancy between the results of in vitro activation assays and in vivo translation might be explained by elevated Trp degradation in the presence of Trp analogs. Expression of tryptophanase, a major enzyme

involved in Trp degradation, is regulated by tryptophan-induced transcription antitermination in *E. coli*.[44,45] Even 1-methyltryptophan, a Trp analog, acted as an effective inducer in vitro.[45] Therefore, we reasoned that supplementation of **5** into the culture medium accelerated Trp degradation via formation of tryptophanase and reduced the intracellular concentration of Trp levels lower than that of Phe. Since the ytRNA$^{Phe}_{CUA_UG}$ charged with **5** was rejected in ribosome, the ytRNA$^{Phe}_{CUA_UG}$ charged with **2** was dominantly utilized for translation.

Expression yields of mDHFR_38Am with various Trp analogs were determined by measuring the concentration of purified mDHFR_38Am. The results are summarized in Table 2. One interesting finding was that **4** provided 25% higher expression level of mDHFR_38Am as compared to **3** despite its 3.5-fold lower k_{cat}/K_m value. This discrepancy can be rationalized by higher k_{cat} of **4** than that of **3**. The concentration (3 mM) of Trp analogs supplemented into culture medium was at least 900-fold higher than K_m values of both **3** and **4**. In such a case, the k_{cat} value became more important than k_{cat}/K_m in determining relative activation rates. Therefore, **4** with 60% higher k_{cat} value than that of **3** could be more efficiently used for protein translation.

Spectral Analysis of Fluorescent Proteins. The spectral properties of CFP6 and CFP6_66Am proteins containing Trp analogs are summarized in Table 3. Absorption maxima, emission maxima, and extinction coefficient of CFP6 were consistent with literature values.[46] Compared to CFP6, CFP6_6ClW and CFP6_6BrW showed blue-shifts of 14 nm in the fluorescence emission maxima (Figure 6). Similar trends were found in previous work in which Tyr at position 66 of GFP was replaced by *p*-

bromophenylalanine.[36] In addition, solution studies indicated that the absorption and fluorescence spectra of fluorinated tryptophans were blue-shifted;[16] the electron-withdrawing property of halogen functional groups can account for the blue-shift in the fluorescence emission peak. Compared with CFP6, the quantum yields of the CFP6_6ClW and CFP6_6BrW decreased 2.5 to 3-fold (Table 3), probably due to collisional perturbation of spin-orbital coupling in the π-electron orbitals of the fluorophore by the halogen atom.[47] CFP6_BT showed blue-shifts of 16 and 34 nm in fluorescence emission and absorption maxima, respectively. Similar blue-shift in absorption maxima in CFP variants was observed by replacing a tryptophan with sulfur containing tryptophan surrogates, such as β-(thieno[2,3-*b*]pyrrolyl)-*L*-alanine ([2.3]Tpa) and β-(thieno[3,2-*b*]pyrrolyl)-*L*-alanine ([3.2]Tpa).[16] Extinction coefficient of CFP6_BT (8,100 $M^{-1}cm^{-1}$) was similar to those of CFP variants containing either [2,3]Tpa (6840 $M^{-1}cm^{-1}$) or [3,2]Tpa (6350 $M^{-1}cm^{-1}$). Introduction of either [2,3]Tpa or [3,2]Tpa in the chromophore of CFP resulted in the disappearance of fluorescence;[16] however, CFP6_BT exhibited substantial fluorescence, with quantum yield only 3-fold smaller than that of CFP6. CFP6_BT is the only fluorescent protein containing a sulfur atom in its chromophore. CFP6_BT has several advantages as a fluorescence resonance electron transfer (FRET) partner. It can be efficiently excited by a common violet diode laser (405-415 nm), which can simplify cell screening on the basis of FRET signals. In addition, the Stokes' shift of CFP6_BT (56 nm) is larger than that of CFP6 (37 nm). This large Stokes' shift is advantageous for fluorescence measurements, since it allows emission photons to be detected against a low background and isolated from excitation photons. CFP and yellow fluorescent protein (YFP) have been commonly used as a FRET

pair. Because the fluorescence emission spectrum of CFP_BT (Figure 6 **b**) overlaps the excitation spectrum of GFP, it can be used as a FRET partner with GFP.

Conclusions

In this paper, we have shown that properly engineered *E. coli* strains can allow high-fidelity incorporation of Trp analogs into recombinant proteins in response to an amber stop codon. Rational modification of the binding pocket in yPheRS enabled activation of Trp analogs (**3**-**6**). When the engineered yPheRS and the 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 analogs (**3**, **4** and **6**). In a subsequent experiment, three Trp analogs (**3**, **4**, and **6)** were introduced at the Trp66 position of CFP. CFP variants containing Trp analogs (**4, 5**, and **6**) in the chromophore showed blue-shifts in fluorescence emission and absorption maxima. In particular, CFP6BT provided unusually large Stokes' shift (56 nm). Introduction of aryl bromides into recombinant proteins should allow site-specific modification via bio-orthogonal palladium-catalyzed coupling reactions. An expanded set of genetically encoded Trp analogs should enable design of new proteins with novel properties.

Acknowledgments

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

References

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. Dayoff, M. O., *Atlas of protein sequence and structure* National Biomedical Research Foundation: Washington DC, 1972.
7. Wong, C. Y.; Eftink, M. R. *Protein Sci.* **1997**, *6*, 689-697.
8. Brawerman, G.; Ycas, M. *Arch. Biochem. Biophys.* **1957**, *68*, 112-117.
9. Pardee, A. B.; Shore, V. G.; Prestidge, L. S. *Biochim. Biophys. Acta* **1956**, *21*, 406-407.
10. Budisa, N. *Angew. Chem. Int. Ed.* **2004**, *43*, 6426-6463.
11. Budisa, N.; Pal, P. P. *Biol. Chem.* **2004**, *385*, 893-904.
12. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.
13. Pratt, E. A.; Ho, C. *Fed. Proc.* **1974**, *33*, 1463-1463.
14. Seifert, M. H.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 7932-7942.
15. Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. *J. Mol. Biol.* **2001**, *309*, 925-936.

16. 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.

17. 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.

18. 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.

19. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1-11.

20. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett.* **2001**, *505*, 465-465.

21. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235-237.

22. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635-10645.

23. Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P. *Science* **2001**, *292*, 501-504.

24. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652-5653.

25. Kwon, I.; Wang, P.; Tirrell, D. A. *J. Am. Chem. Soc.* **2006**, *128*, 11778-11783.

26. Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

27. 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.

28. 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.

29. Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

30. Wang, L.; Brock, A.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 1836-1837.

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.

36. Wang, L.; Xie, J. M.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

37. Gill, S. C.; Vonhippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

38. Yakhnin, A. V.; Vinokurov, L. M.; Surin, A. K.; Alakhov, Y. B. *Protein Expr. Purif.* **1998**, *14*, 382-386.

39. Velapoldi, R. A.; Tonnesen, H. H. *J. Fluoresc.* **2004**, *14*, 465-472.

40. Fishman, R.; Ankilova, V.; Moor, N.; Safro, M. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57*, 1534-44.

41. Volumes were estimated as solvent-excluded volumes (1.4 Å probe) by using Chem3D Pro 5.0 (Cambridge, MA)

42. Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. *Protein Sci.* **2001**, *10*, 1281-1292.

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. Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

47. Kasha, M. *J. Chem. Phys.* **1952**, *20*, 71-74.

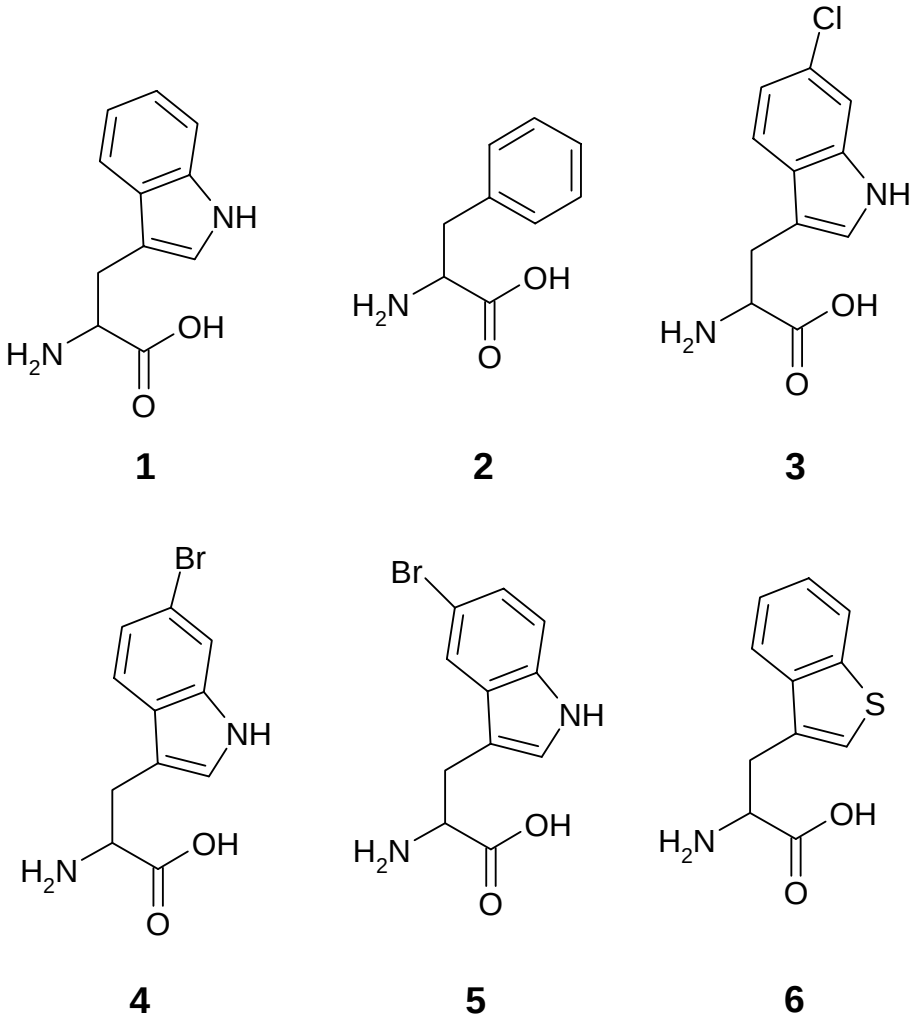

Figure 1: Amino acids involved in this study. (**1**) tryptophan, (**2**) phenylalanine, (**3**) 6-chlorotryptophan, (**4**) 6-bromotryptophan, (**5**) 5-bromotryptophan, (**6**) benzothienylalanine.

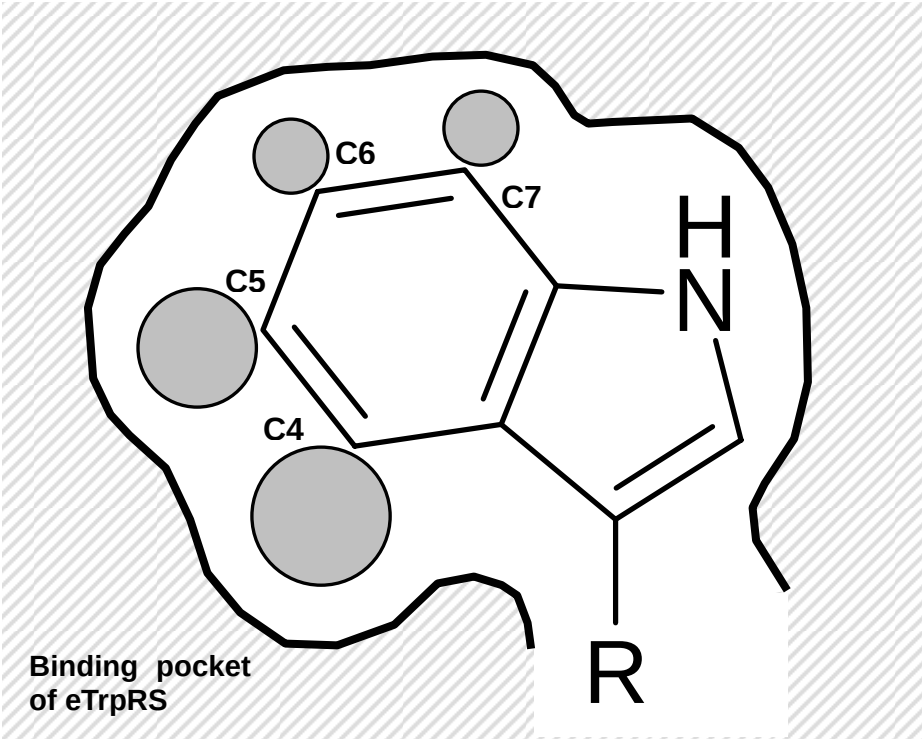

Figure 2: A hypothetical model for the adaptability of Trp analogs to the binding pocket of eTrpRS. Position and size of substituents on the indole ring that can be tolerated in the binding pocket of eTrpRS are represented as circles. Indole rings containing larger substituents can not fit into the binding pocket of eTrpRS.

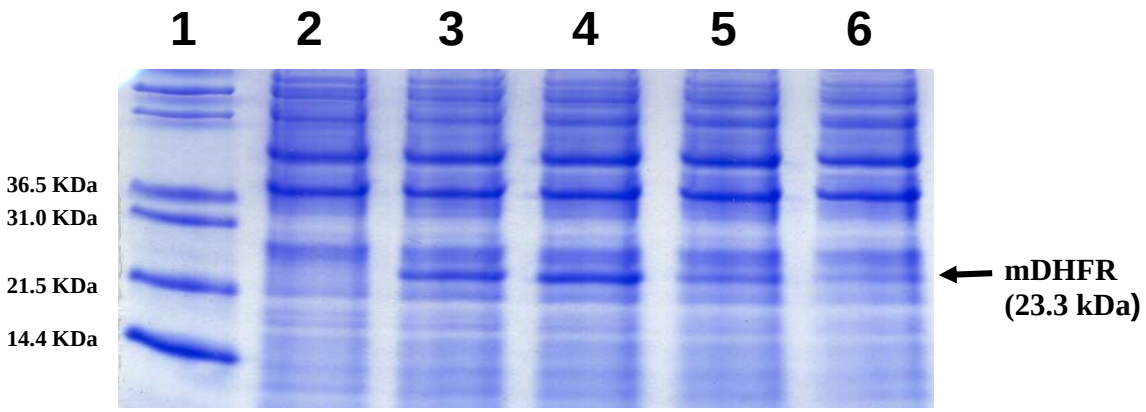

Figure 3: SDS-PAGE analysis of mDHFR_38Am expression. Cells were equipped with yPheRS (T415G) and ytRNA$^{Phe}_{CUA_UG}$. Cultures were supplemented with 18 amino acids (at 25 mg/L), one of Trp analogs (at 3 mM) and the indicated concentrations of Phe and Trp. Lane 1: molecular weight standards, lane 2: before induction, lane 3: 6ClW, 0.03 mM Phe and 0.01 mM Trp, Lane 4: 6BrW, 0.015 mM Phe and 0.05 mM Trp, lane 5: 5BrW, 0.03 mM Phe and 0.01 mM Trp, lane 6: BT, 0.01 mM Phe and 0.0025 mM Trp.

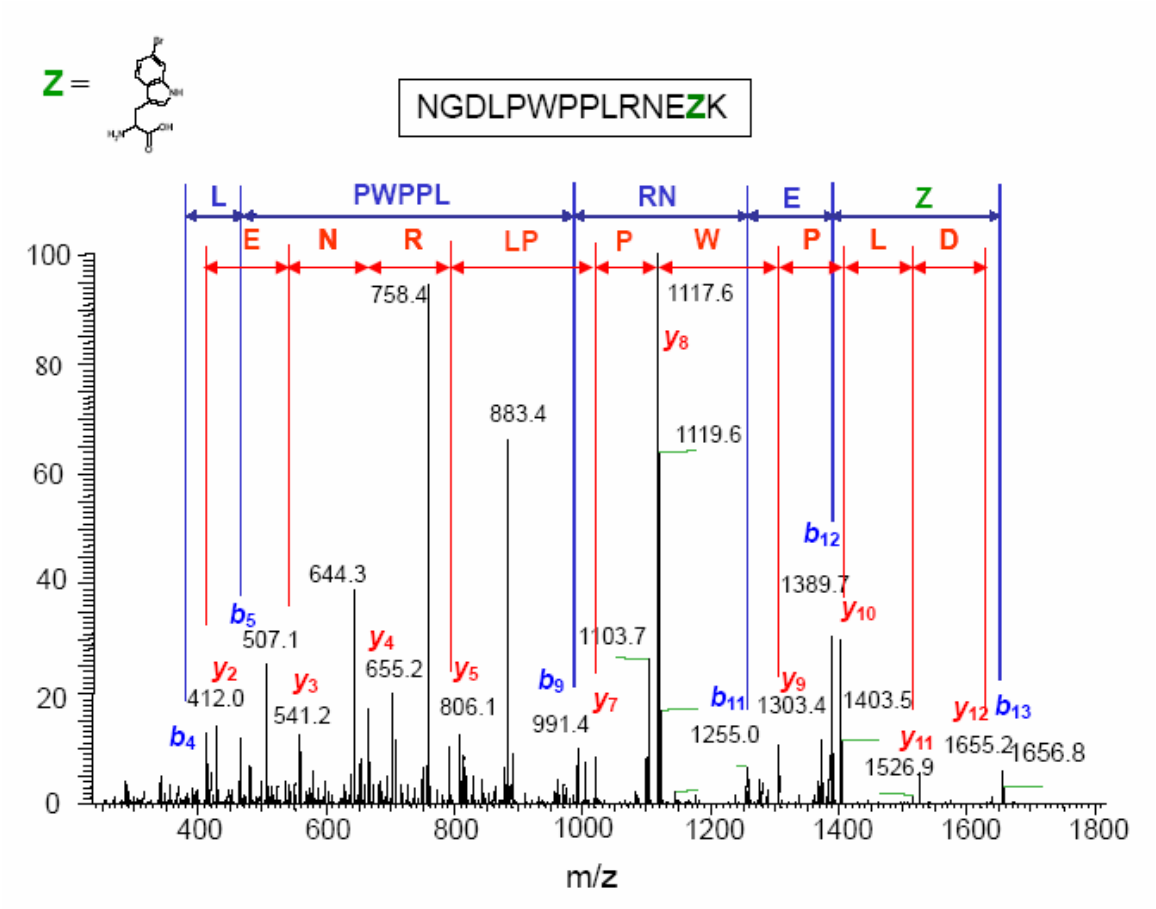

Figure 4: Tandem mass spectrum of the peptide (NGDLPWPPLRNEZK). The partial sequence DLPWPPLRNEZ containing 6BrW (**Z**) can be confirmed by assigning masses of series of fragmented ions (*b* and *y*).

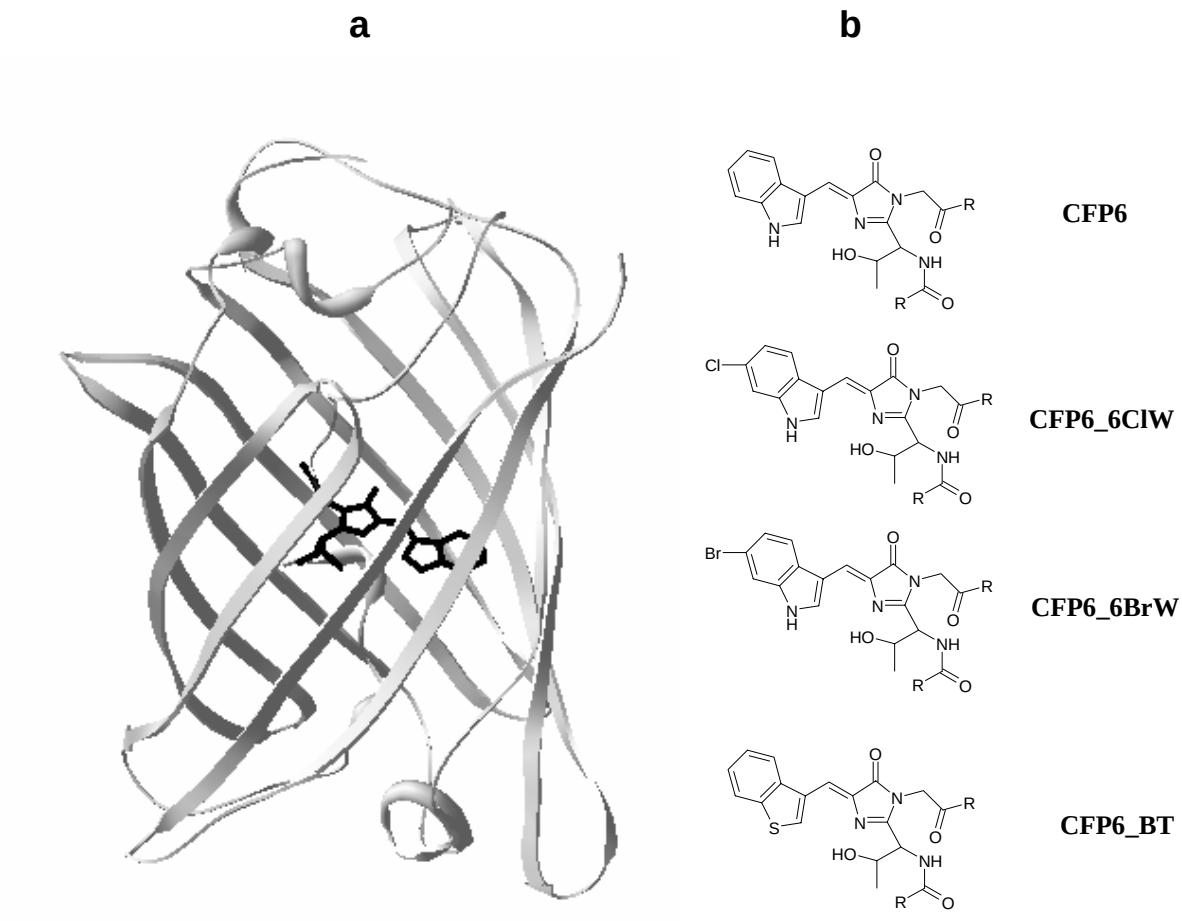

Figure 5: Crystal structure of enhanced cyan fluorescent protein (PDB code 1OXD)[18] (**a**). The chromophore is shown in black. Chromophores of CFP6 variants containing **1**, **3**, **4**, and **6** (**b**).

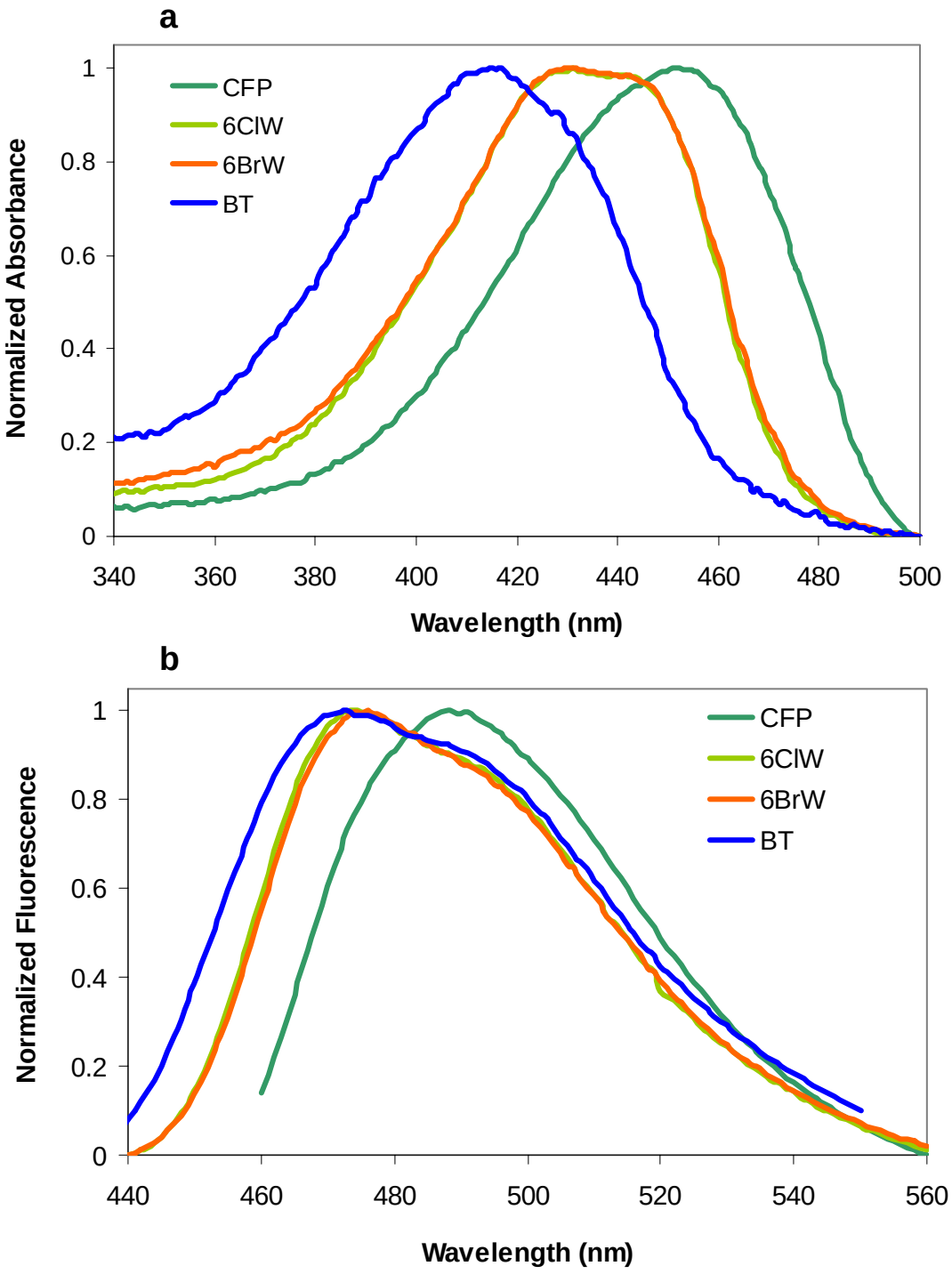

Figure 6: Absorption (**a**) and fluorescence emission spectra (**b**) for CFP6 variants containing different Trp analogs at the Trp66 position.

Table 1: Kinetic parameters for ATP-PPi exchange by the yPheRS (T415G)

Amino Acid	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m (rel)[a]
Phe (**2**)	55±14	0.20±0.11	3,500±1,100	1
Trp (**1**)	2.8±1.6	0.15±0.003	63,200±34,600	18
6ClW (**3**)	0.51±0.21	0.11±0.04	224,100±15,300	64
6BrW (**4**)	3.4±1.3	0.18±0.11	61,800±53,900	18
5BrW (**5**)	0.62±0.02	0.021±0.002	33,900±3,600	9.7
BT (**6**)	8.1±3.7	0.15±0.06	22,500±18,000	6.4

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

Table 2: Occupancy of amber sites and expression yields

Trp analogs	Occupancy of amber sites (%)				Yield (mg/L)[a]
	Phe	Trp	Lys	Naa	
Trp (**1**)	0.4 ± 0.5	ND[b]	1.2 ± 1.3	98.4 ± 1.7	5.4 ± 0.3
6ClW (**3**)	0.4 ± 0.7	ND	ND	99.6 ± 0.7	6.8 ± 1.3
6BrW (**4**)	97.0 ± 1.4	0.3 ± 0.6	0.7 ± 0.5	1.8 ± 0.7	3.6 ± 0.6
BT (**6**)	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.

Table 3: Spectral properties of CFP6 variants

amino acid at position 66	absorption maximum (nm)	extinction coefficient (M^{-1} cm^{-1})	emission maximum (nm)	quantum yield
Trp (**1**)	449	25900 ± 1000	488	0.14 ± 0.008
6ClW (**3**)	430	19800 ± 6400	474	0.050 ± 0.001
6BrW (**4**)	430	20600 ± 3800	474	0.058 ± 0.001
BT (**6**)	415	8100 ± 1100	472	0.048 ± 0.008