Genetic Codon Expansion

DOI: 10.1002/anie.201301094

A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in *E. coli***

*Abhishek Chatterjee, Han Xiao, Peng-Yu Yang, Gautam Soundararajan, and Peter G. Schultz**

The genetic incorporation of unnatural amino acids (UAAs) with novel structures and properties into proteins in living cells has provided a powerful new tool to both probe and modify protein structure and function.[1] The desired UAA is encoded by a nonsense or a frameshift codon using an engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pair that is orthogonal to the host cell (i.e., not cross-reactive with endogenous host aaRSs, tRNAs, or amino acids). While several orthogonal aaRS/tRNA pairs have been reported for UAA mutagenesis in *E. coli*, only the archaebacteria-derived tyrosyl and pyrrolysyl aaRS/tRNA pairs have been successfully used to genetically encode a large number of UAAs.[1,2] To alter the specificity of the aaRS, a library of active-site mutants is subjected to rounds of positive and negative selection on the basis of their ability to incorporate amino acids into amber mutants of essential or toxic proteins. Development of new orthogonal aaRS/tRNA pairs with structurally distinct active sites should further expand the chemical diversity of genetically encoded UAAs. A tryptophanyl-tRNA synthetase (TrpRS)/$tRNA^{Trp}$ pair is particularly attractive owing to the large tryptophan binding site of TrpRS that can potentially accommodate novel UAAs with large side chains. A modified TrpRS/$tRNA^{Trp}$ pair from *Saccharomyces cerevisiae* was recently shown to be an orthogonal amber suppressor in *E. coli*.[2e] Here we report significant improvements to the amber suppression efficiency of the ScTrpRS/$tRNA_{CUA}^{Trp}$ pair by modifications to the acceptor stem of $tRNA_{CUA}^{Trp}$. Furthermore, we evolved active-site variants of TrpRS which are able to charge 3-(1-naphthyl)-alanine, 1-methyl-tryptophan, 3-benzothienyl-alanine, and 6-methyl-tryptophan with high fidelity and efficiency. Replacement of Trp66 in enhanced cyan fluorescent protein (ECFP) by these UAAs yielded ECFP variants with altered spectral features.

To explore the possibility of using the ScTrpRS/$tRNA_{CUA}^{Trp}$ pair to encode new UAAs, the wild-type ScTrpRS was inserted into the pBK plasmid, where it is expressed constitutively under the *glnS* promoter.[3] Two reported variants (AS3.4 and AS3.5) of $tRNA_{CUA}^{ScTrp}$ driven by constitutive *lpp* promoters were cloned into two separate pREP vectors to generate the pREP-$tRNA_{CUA}^{ScTrp}$ plasmids.[2a] The pREP plasmid also encodes a chloramphenicol acetyl transferase (CAT) gene with an amber stop codon at a permissive site (Asp112TAG). To evaluate the amber suppression efficiency of the ScTrpRS/$tRNA_{CUA}^{ScTrp}$ pair, the pREP-$tRNA_{CUA}^{ScTrp}$ plasmids (encoding either $tRNA_{CUA}^{ScTrp}$-AS3.4 or -AS3.5) were transformed into *E. coli* DH10B with or without the pBK-ScTrpRS plasmid. The chloramphenicol resistance (ChlorR) of strains expressing $tRNA_{CUA}^{ScTrp}$ alone or in combination with ScTrpRS is indicative of the cross-reactivity of the $tRNA_{CUA}^{ScTrp}$ with host aminoacyl-tRNA synthetases and the amber suppression efficiency of this heterologous pair, respectively. While both variants of the $tRNA_{CUA}^{ScTrp}$ were found to be orthogonal to the endogenous aaRSs (ChlorR $< 5\ \mu g\,mL^{-1}$), the amber suppression activities in the presence of ScTrpRS were weak (ChlorR ca. $15\ \mu g\,mL^{-1}$) compared to the *M. jannaschii* TyrRS/$tRNA_{CUA}^{MjTyr}$ pair (ChlorR $> 200\ \mu g\,mL^{-1}$).

We next explored the possibility of further improving the efficiency of the ScTrpRS/$tRNA_{CUA}^{ScTrp}$-AS3.5 pair (Figure S1 in the Supporting Information). There are two wobble G:U pairs in the stem regions of $tRNA^{ScTrp}$, G49:U65 and G4:U69 (Figure 1A). Mutation of such G:U pairs to G:C was previously shown to enhance the translational efficiency of a tRNA.[2e] Two $tRNA_{CUA}^{ScTrp}$-AS3.5 variants were generated, either with a U65C mutation alone ($tRNA_{CUA}^{ScTrp}$-AS3.51) or in combination with a U69C mutation ($tRNA_{CUA}^{ScTrp}$-AS3.52). The mutant $tRNA_{CUA}^{ScTrp}$-AS3.51 exhibited a modest increase in activity (ChlorR ca. $35\ \mu g\,mL^{-1}$) and remained orthogonal to *E. coli* (Figure S1), while the double mutant $tRNA_{CUA}^{ScTrp}$-AS3.52 had a significantly higher activity (ChlorR ca. $150\ \mu g\,mL^{-1}$), but diminished orthogonality (Figure S1). We next attempted to generate orthogonal tRNA variants with enhanced activity by subjecting a large acceptor-stem mutant library of the corresponding tRNA to a double-sieve selection scheme.[2b–d,f] The first five base pairs in the acceptor stem of $tRNA_{CUA}^{ScTrp}$-AS3.51 were randomized to all possible combinations to generate a library of approximately 10^6 variants. This library was subjected to alternating rounds of positive and negative selection to identify active, yet orthogonal variants. In the positive round of selection, tRNA mutants

[*] Dr. A. Chatterjee,[+] H. Xiao,[+] Dr. P. Y. Yang, G. Soundararajan, Prof. Dr. P. G. Schultz
Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: schultz@scripps.edu
Homepage: http://schultz.scripps.edu/

[+] These authors contributed equally to this work.

[**] We would like to acknowledge Virginia Seely for assistance in manuscript preparation and Dr. Xin Zhang for help in spectrum analysis. We thank Sophie Sun and Dr. Jenny Furman for fruitful discussions. This is paper 23000 from The Scripps Research Institute. This work is supported by grant DE-FG03-00ER46051 from the Division of Materials Sciences, Department of Energy (P.G.S.).

Supporting information (including details of plasmids construction, protein expression, ESI–MS analysis, and spectroscopic measurement of expressed proteins) for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201301094.

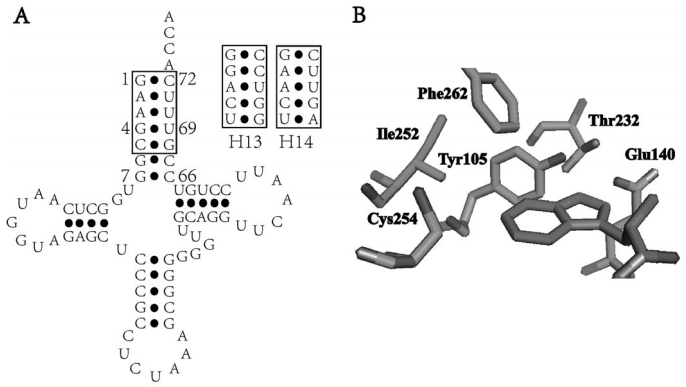

Figure 1. Structure of *Saccharomyces cerevisiae* $tRNA_{CUA}^{Trp}$-AS3.5 and ScTrpRS·tryptophanyl-AMP complex. A) Cloverleaf representation of *Saccharomyces cerevisiae* $tRNA_{CUA}^{Trp}$-AS3.5. The boxed segment of the acceptor stem of $tRNA_{CUA}^{ScTrp}$-AS3.51 was randomized. Efficient acceptor variants of $tRNA_{CUA}^{ScTrp}$-AS3.51, $tRNA_{CUA}^{ScTrp}$-H13 and $tRNA_{CUA}^{ScTrp}$-H14 (sequences shown in the inset), contain the U65C mutation. B) Tryptophan binding site of ScTrpRS in complex with tryptophanyl-AMP (based on the PDB entry 3KT3). The residues chosen for randomization are highlighted.

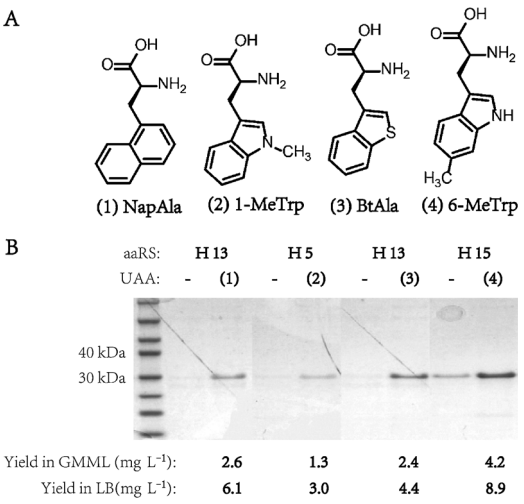

Figure 2. Evolution of ScTrpRS to genetically encode tryptophan analogues. A) Structures of UAAs. B) Expression of GFP-Tyr151TAG variants analyzed by SDS-PAGE (stained with Coomassie blue). GMML = glycerol minimal medium supplemented with leucine.

were co-expressed with the cognate ScTrpRS and active variants were enriched based on their ability to suppress a permissive amber codon in a CAT reporter gene (Asp112TAG). In the negative round of selection, which was performed in the absence of ScTrpRS, cross-reactive tRNA mutants were removed based on suppression of two permissive amber codons (Gln3TAG and Asp44TAG) in the toxic barnase gene. After three rounds of selection, several orthogonal $tRNA_{CUA}^{ScTrp}$-AS3.51 variants were isolated that survived in the presence of 100 $\mu g\,mL^{-1}$ of chloramphenicol. Two particular variants, $tRNA_{CUA}^{ScTrp}$-H13 and $tRNA_{CUA}^{ScTrp}$-H14 exhibited significantly enhanced activity (ChlorR ca. 120–130 $\mu g\,mL^{-1}$) and low cross-reactivity (ChlorR < 20 $\mu g\,mL^{-1}$; Figure S1).

We next focused on altering the substrate specificity of ScTrpRS. On the basis of the crystal structure of the ScTrpRS·tryptophanyl-AMP complex, we randomized six amino acid residues involved in recognition of tryptophan (Tyr105, Glu140, Thr232, Ile252, Cys254, and Phe262; Figure 1B). Three libraries were generated, wherein five of the aforementioned residues were completely randomized (NNK randomization; N = any nucleotide, K = G or T) by site-saturation mutagenesis (TIF library: Tyr105, Glu140, Thr232, Ile252, and Phe262; TIC library: Tyr105, Glu140, Thr232, Ile252, and Cys254; ICF library: Tyr105, Glu140, Ile252, Cys254, and Phe262). More than 10^9 transformants were generated for each library, and sequence analysis of individual clones revealed no significant sequence bias or other irregularities in these libraries.

These libraries were then used in selection experiments to identify ScTrpRS variants that specifically use 3-(1-naphthyl)-L-alanine (NapAla (**1**), Figure 2A) as a substrate. 3-(1-naphthyl)-L-alanine has been shown to be a useful building bock in the synthesis of peptides with enhanced binding affinity and half-lives.[4] The three libraries were individually subjected to a double-sieve selection scheme.[1] Each ScTrpRS library was transformed into DH10B cells containing the pREP reporter plasmid encoding a CAT gene with a permissive Asp112TAG mutation and $tRNA_{CUA}^{ScTrp}$-AS3.51. These cells were grown in the presence of NapAla (1 mM) and chloramphenicol (10 $\mu g\,mL^{-1}$) to enrich ScTrpRS variants able to charge $tRNA_{CUA}^{ScTrp}$. In the negative round of selection, the surviving library members were introduced into DH10B cells harboring a plasmid expressing $tRNA_{CUA}^{ScTrp}$-AS3.51 and a toxic barnase gene with amber mutations at two permissive sites (Gln3TAG, Asp44TAG). Cells were grown in the absence of NapAla to remove mutants that recognize endogenous amino acids. After three rounds of selection, individual clones were identified that survived chloramphenicol (15 $\mu g\,mL^{-1}$) only in the presence of NapAla (1 mM) in the positive selection strain. Sequencing of 48 clones revealed seven unique sequences (Table S1 in the Supporting Information), in which Tyr105 and Glu140 were mutated to a hydrophobic residue (Leu/Val/Ile) and proline, respectively; Thr232 was either retained or mutated to either Ser or Cys; and Ile252 was frequently retained as an aliphatic hydrophobic residue (Ile/Val) or mutated to Cys.

To test the efficiency and the fidelity of incorporation of NapAla into model proteins, several NapAla-specific ScTrpRS variants (ScTrpRS-H3, ScTrpRS-H5, ScTrpRS-H6, ScTrpRS-H13, ScTrpRS-H15, ScTrpRS-H19, and ScTrpRS-H46; Table S1) and $tRNA_{CUA}^{ScTrp}$-H14 were inserted into the pUltra suppressor plasmid under the control of *tacI* and *proK* promoters, respectively.[5] These suppressor plasmids were used to express a green fluorescent protein (GFP) variant containing a C-terminal His_6 tag and an amber codon replacing Tyr151 (in the pET101 expression plasmid). Protein expression was carried out in *E. coli* BL21(DE3) in either minimal media or rich media, in the presence and the absence of NapAla (2 mM). SDS-PAGE analysis of the Ni^{2+}-NTA affinity purified protein showed that full-length GFP was expressed only in the presence of NapAla (Figure 2B). ESI–MS analysis afforded an observed mass of 27740 Da, in

agreement with the calculated mass (Figure S2 in the Supporting Information). The yield of mutant GFP protein was 6.1 mg L^{-1} in LB medium (34 % of wild-type GFP with no amber suppression).

Since the selection scheme employed to identify a mutant synthetase specific for NapAla does not select against other similar UAAs, the resulting synthetase variants may be polyspecific (i.e., aminoacylate multiple UAAs, but not canonical amino acids).[6] We therefore determined whether the mutant ScTrpRSs specific for NapAla could recognize other tryptophan analogues. *E. coli* BL21(DE3) containing pET101-GFPTyr151TAG and pUltra-$tRNA_{CUA}^{ScTrp}$ encoding different ScTrpRS variants (ScTrpRS-H3, ScTrpRS-H5, ScTrpRS-H6, ScTrpRS-H13, ScTrpRS-H15, and ScTrpRS-H19) were grown in GMML media supplemented with various UAAs (2 mM). On the basis of GFP fluorescence in the presence or absence of the UAAs, we found that ScTrpRS-H5, ScTrpRS-H13, and ScTrpRS-H15 were able to incorporate 1-methyl-L-tryptophan (1-MeTrp (**2**), Figure 2A), 3-benzothienyl-L-alanine (BtAla (**3**), Figure 2A), and 6-methyl-L-tryptophan (6-MeTrp (**4**), Figure 2A), respectively. Yields of mutant GFPs substituted with 1-MeTrp, BtAla, and 6-MeTrp were 3 mg L^{-1}, 4.4 mg L^{-1}, and 8.9 mg L^{-1}, respectively. ESI–MS analyses of purified proteins verified the incorporation of these UAAs (Figure S2 in the Supporting Information).

The chromophore of ECFP results from the self-catalyzed posttranslational cyclization of three internal amino acids (Thr65-Trp66-Gly67; Figure 3A) inside the β-barrel core. Extensive studies have been carried out by several groups to further optimize the spectroscopic properties of ECFPs by using conventional amino acid mutagenesis as well as unnatural amino acids.[7] Substitution of Trp66 in ECFP with the Trp analogues described here may allow the generation of additional variants with novel spectral properties. To express such ECFP mutants, we introduced an amber mutation in ECFP corresponding to Trp66. *E. coli* BL21(DE3) harboring the pET22b-ECFP-Trp66TAG and pUltra-$tRNA_{CUA}^{ScTrp}$ encoding the appropriate ScTrpRS variant were grown in GMML media supplemented with UAA (2 mM). Proteins were purified to homogeneity by Ni^{+2} affinity chromatography and analyzed by SDS-PAGE and ESI–MS analyses (Figure 3B and Figure S3 in the Supporting Information). All mutant ECFPs expressed in good yields, except for the one incorporating NapAla, which may not fold properly.

The absorbance and fluorescence spectra of the wild-type and the mutant ECFPs in 20 mM Tris·HCl buffer are shown in Figure 3C,D and Table 1. Wild-type ECFP exhibits two peaks in both its absorption (at 435 nm and 448 nm) and emission spectra (476 nm and 501 nm).[8] Substitution of Trp66 with 1-MeTrp, resulted in a similar absorption profile. ECFPs containing BtAla and 6-MeTrp exhibit a 20 nm blue-shift and a 22 nm red-shift in the absorbance maxima, respectively. The emission maxima for ECFP incorporating 1-MeTrp show a 10 nm red-shift. Interestingly, the incorporation of 6-MeTrp into ECFP resulted in a single emission maximum (at 495 nm), unlike the wild-type or other mutant ECFPs showing two emission maxima. It has been proposed that different ECFP chromophore environments contribute to two peaks in the spectra.[9] In the major conformation, Tyr145 faces the chromophore and His148 flips away. In the minor conformation, His148 moves closer to the chromophore, while Tyr145 flips away. It is possible that the methyl group of 6-MeTrp results in a steric clash with His148 in the minor conformation, thereby rendering it energetically unfavorable. In addition, the ECFP variant with BtAla showed a relatively large Stokes shift (63 nm). However, the quantum yield of the mutant CFP containing BtAla was low, which is likely due to spin–orbit coupling associated with the sulfur substituent.

In conclusion, we have developed a novel tryptophanyl-tRNA/synthetase pair and used it to genetically encode four

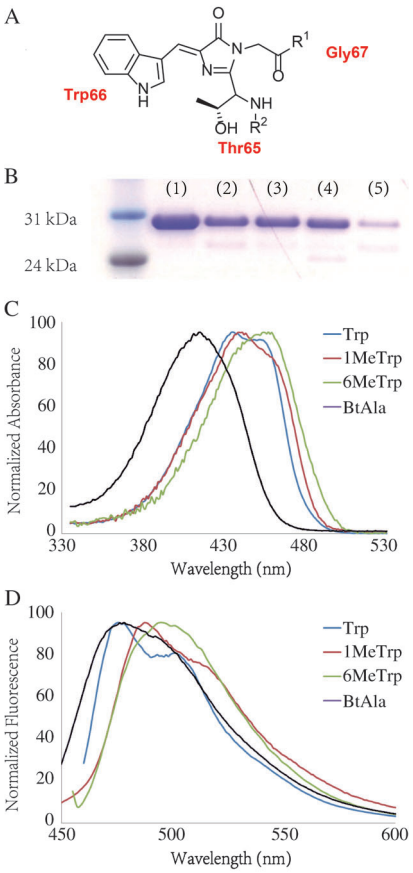

Figure 3. UAA mutagenesis of the ECFP chromophore. A) Structure of the chromophore of ECFP. B) Incorporation of various UAAs into ECFP-Trp66TAG, analyzed by SDS-PAGE stained with Coomassie blue. Lane 1: wild-type ECFP; Lane 2: 1-Me-Trp mutant; Lane 3: 6-MeTrp mutant; Lane 4: BtAla mutant; Lane 5: NapAla mutant. C) Absorption spectra for wild-type ECFP and mutants. D) Emission spectra for wild-type ECFP and mutants.

Table 1: Spectral properties of wild-type and mutant ECFPs.

Amino acid at position 66	Absorbance maximum [nm][a]	Emission maximum [nm]	Quantum yield
Trp	435	476	0.48 ± 0.01
1-MeTrp	439	487	0.18 ± 0.01
6-MeTrp	457	495	0.32 ± 0.01
BtAla	415	478	0.011 ± 0.001

[a] ECFP mutants are in Tris·HCl buffer (20 mM, pH 8.0).

tryptophan analogues in *E. coli* with good yield and high fidelity. These UAAs were used to modulate the spectral properties of ECFP and likely can be further used as probes of tryptophan structure and function in other proteins. The large active site of ScTrpRS should be useful for genetically encoding other unnatural amino acids with diverse chemical structures.

Received: February 6, 2013
Published online: April 2, 2013

Keywords: amber suppression · genetic code expansion · unnatural amino acids · protein engineering

[1] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413 – 444.
[2] a) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498 – 500; b) J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598 – 9608; c) S. W. Santoro, J. C. Anderson, V. Lakshman, P. G. Schultz, *Nucleic Acids Res.* **2003**, *31*, 6700 – 6709; d) J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7566 – 7571; e) R. A. Hughes, A. D. Ellington, *Nucleic Acids Res.* **2010**, *38*, 6813 – 6830; f) A. Chatterjee, H. Xiao, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14841 – 14846; g) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232 – 234; h) P. R. Chen, D. Groff, J. T. Guo, W. J. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112 – 4115; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052 – 4055.
[3] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044 – 1048.
[4] a) M. Kuragaki, M. Sisido, *J. Phys. Chem.* **1996**, *100*, 16019 – 16025; b) B. E. Haug, J. S. Svendsen, *J. Pept. Sci.* **2001**, *7*, 190 – 196.
[5] A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, *Biochemistry* **2013**, *52*, 1828 – 1837.
[6] D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon, P. G. Schultz, *Biochemistry* **2011**, *50*, 1894 – 1900.
[7] a) M. A. Rizzo, G. H. Springer, B. Granada, D. W. Piston, *Nat. Biotechnol.* **2004**, *22*, 445 – 449; b) G. J. Kremers, J. Goedhart, E. B. van Munster, T. W. J. Gadella, *Biochemistry* **2006**, *45*, 6570 – 6580; c) A. Villoing, M. Ridhoir, B. Cinquin, M. Erard, L. Alvarez, G. Vallverdu, P. Pernot, R. Grailhe, F. Merola, H. Pasquier, *Biochemistry* **2008**, *47*, 12483 – 12492.
[8] M. Lelimousin, M. Noirclerc-Savoye, C. Lazareno-Saez, B. Paetzold, S. Le Vot, R. Chazal, P. Macheboeuf, M. J. Field, D. Bourgeois, A. Royant, *Biochemistry* **2009**, *48*, 10038 – 10046.
[9] J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, N. Budisa, *J. Mol. Biol.* **2003**, *328*, 1071 – 1081.

