Genetic Code Expansion

DOI: 10.1002/ange.201000465

A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in *Escherichia coli***

*Wei Wan, Ying Huang, Zhiyong Wang, William K. Russell, Pei-Jing Pai, David H. Russell, and Wenshe R. Liu**

Since the first report in 1998,[1] genetic incorporation of noncanonical amino acids (NAAs) into proteins at amber UAG codons in living cells using orthogonal aminoacyl-tRNA synthetase (aaRS)–amber suppressor tRNA ($tRNA_{CUA}$) pairs has flourished. Evolved *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)–*Mj*$tRNA^{Tyr}_{CUA}$ pairs together with the naturally occurring wild-type or evolved pyrrolysyl-tRNA synthetase (PylRS)–$tRNA^{Pyl}$ (pylT) pairs have enabled the incorporation of more than 30 NAAs into proteins in *E. coli*.[2] Although the incorporation of these NAAs, which have unique chemical properties and reactivities, into proteins has dramatically increased our ability to manipulate protein structure and function, there still exists one major limitation for the technique. Namely, the technique in general only allows the incorporation of a single NAA into a single protein because the amber codon is the only one available for the incorporation of NAAs and the nonsense suppression rate in living cells is low. We have previously demonstrated the incorporation of up to three N^{ε}-acetyl-L-lysines (AcKs) at amber mutation sites of GFP_{UV} in *E. coli* under an enhanced amber suppression condition.[3] However, to incorporate two different NAAs into a single protein, the use of an additional blank codon is required. Herein, we show that the PylRS–pylT pair can be mutated to suppress the ochre UAA codon and the mutated pair can be coupled with an evolved *Mj*TyrRS–*Mj*$tRNA^{Tyr}_{CUA}$ pair to efficiently incorporate two different NAAs at two defined sites of a single protein in *E. coli* by both amber and ochre suppressions. This development greatly expands the diversity of modifications we can introduce to proteins.

We attempted to combine the PylRS–pylT pair and the *Mj*TyrRS–*Mj*$tRNA^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein because of their proven efficiency in NAA incorporation.[2d–h,4] To use both pairs in a single *E. coli* cell, we first demonstrated that they are orthogonal to each other. When co-expressed in *E. coli*, either one of the aaRSs could not efficiently charge $tRNA_{CUA}$ from the other pair with its cognate amino acid (N^{ε}-Boc-L-lysine for PylRS[2d] and L-tyrosine for *Mj*TyrRS) to suppress an amber mutation at position 112 of a chloramphenicol acetyltransferase gene to give detectable chloramphenicol resistance. We then tested whether mutated pylT can be used to suppress other blank codons. Although naturally encoded by the UAG codon, it has been reported that pyrrolysine is not hardwired for co-translational insertion at UAG codon positions.[5] The crystal structure of the PylRS–pylT complex also revealed no direct interaction between PylRS and the anticodon region of pylT.[6] We suspected that mutation of the anticodon of pylT might not affect its interaction with PylRS, and the mutated PylRS-pylT pair could be used to suppress a blank codon such as opal UGA codon, ochre UAA codon, or even a four-base codon UAGA. A pAcKRS–pylT–GFP1 Amber plasmid from our previous work[3,7] was employed. This plasmid contains genes coding a mutant PylRS (AcKRS) specific for AcK, pylT, and GFP_{UV}. The GFP_{UV} gene has one amber mutation at position 149. Growing cells transformed with this plasmid in lysogeny broth (LB) medium supplemented with 5 mM AcK led to full-length GFP_{UV} expression that was easily detected by fluorescent emission of GFP_{UV} when excited (Figure 1). When the anticodon of pylT was mutated to the complimentary one for opal, ochre, or a four-base UAGA codon and the

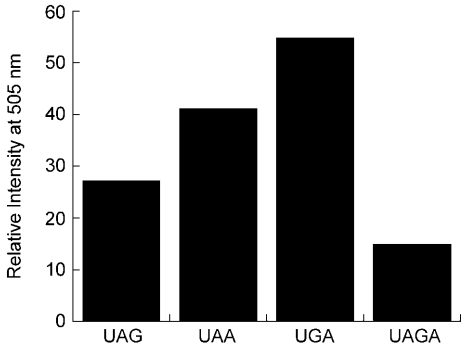

Figure 1. The suppression levels of UAG, UAA, UGA, and UAGA mutations at position 149 of GFP_{UV} by their corresponding mutant pylT suppressors. Suppression levels are represented by the relative fluorescent emission intensity of expressed full-length GFP_{UV} proteins. Excitation wavelength: 385 nm, observation wavelength: 505 nm.

[*] Dr. W. Wan,[+] Y. Huang,[+] Dr. Z. Wang, Dr. W. K. Russell, Dr. P.-J. Pai, Prof. D. H. Russell, Prof. W. R. Liu
Department of Chemistry, Texas A&M University
College Station, TX 77843 (USA)
Fax: (+1) 979-845-4719
E-mail: wliu@chem.tamu.edu
Dr. W. Wan[+]
Department of Biology, Texas A&M University
College Station, TX 77843 (USA)

[+] These authors contributed equally to this work.

[**] We thank Prof. Peter G. Schultz of The Scripps Research Institute for providing us with the pEVOL-AzFRS plasmid and Prof. Daniel Romo, Prof. Kevin Burgess, and Prof. Jiong Yang of Texas A&M University for using their instruments. This work was supported by Welch Grant A-1715 and the New Faculty Startup Fund to W.R.L. from the Chemistry Department of Texas A&M University.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201000465.

corresponding codon was introduced to position 149 of GFP_{UV} in the pAcKRS–pylT–GFP1 Amber plasmid, cells transformed with the plasmid and grown in LB medium supplemented with 5 mM AcK also exhibited detectable GFP_{UV} expression for all three mutated pylTs (Figure 1). In comparison to wild-type pylT, both $pylT_{UCA}$ that suppresses opal codon and $pylT_{UUA}$ that suppresses ochre codon gave significantly higher suppression levels. This result indicates that both the PylRS–$pylT_{UCA}$ pair and the PylRS–$pylT_{UUA}$ pair can be used to efficiently incorporate NAAs into proteins at their corresponding suppressed codons. Furthermore, it might also be feasible to couple the PylRS–$pylT_{UCA}$ (or $pylT_{UUA}$) pair together with an evolved *Mj*TyrRS–*Mj*$tRNA^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein in *E. coli* by both amber and opal (or ochre) suppressions. Although the suppression of opal mutation in GFP_{UV} by $pylT_{UCA}$ led to a higher full-length GFP_{UV} expression level than in the case with the suppression of ochre mutation in GFP_{UV} by $pylT_{UUA}$, we eventually chose to use the ochre suppressor for NAA incorporation because the anticodon of an opal suppressor may form a wobble pair with the UGG codon in mRNA and decrease the fidelity of tryptophan incorporation.[8]

To demonstrate the utility of the PylRS-$pylT_{UUA}$ pair together with an evolved *Mj*TyrRS–*Mj*$tRNA^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein in *E. coli* by both amber and ochre suppressions, two plasmids, pEVOL–AzFRS and pPylRS–pylT–GFP1TAG149TAA (Supporting Information, Figure S1), were used to transform *E. coli* BL21 cells. The pEVOL–AzFRS plasmid contains genes encoding an optimized *Mj*$tRNA^{Tyr}_{CUA}$ and two copies of an evolved *Mj*TyrRS (AzFRS) specific for *p*-azido-L-phenylalanine (**4**; Figure 2a). This plasmid provides an enhanced amber suppression in *E. coli*.[9] The pPylRS–pylT–GFP1TAG149TAA plasmid contains genes encoding wild-type *M. mazei* PylRS, $pylT_{UUA}$, and GFP_{UV}. The GFP_{UV} gene has an amber mutation at position 1, an ochre mutation at 149, an N-terminal Met-Ala leader dipeptide in front of the amber mutation, and an opal stop codon at the C-terminal end. Growing the transformed cells in 2YT medium supplemented with 1 mM N^{ε}-Boc-L-lysine (**1**; Figure 2a) and 1 mM **4** afforded full-length GFP_{UV} with a yield of 11 mg L^{-1} (Figure 2b, lane 1). No cellular toxicity owing to strong amber and ochre suppressions was observed. Exclusion of either NAA from the medium led to no detectable full-length GFP_{UV} expression (Figure 2b, lanes 4,7). The results indicate that the suppressions of amber and opal mutations are dependent on the presence of their corresponding NAAs. The ESI-MS of the purified full-length GFP_{UV} incorporated with **4** at position 1 and **1** at position 149 (GFP_{UV}(**1**+**4**)) confirmed the expected incorporations (Figure 2c). The detected mass (28085 Da) agrees within 70 parts per million with the calculated mass (28083 Da) of full-length GFP_{UV}(**1**+**4**) without N-terminal methionine. The cleavage of N-terminal methionine from expressed GFP_{UV} in *E. coli* has been observed in related studies.[3,10] A mass peak (28059 Da) that is 26 Da smaller than the major peak is probably due to the decomposition of the azide group in **4** to form the corresponding amine during ESI-MS analysis, which has been

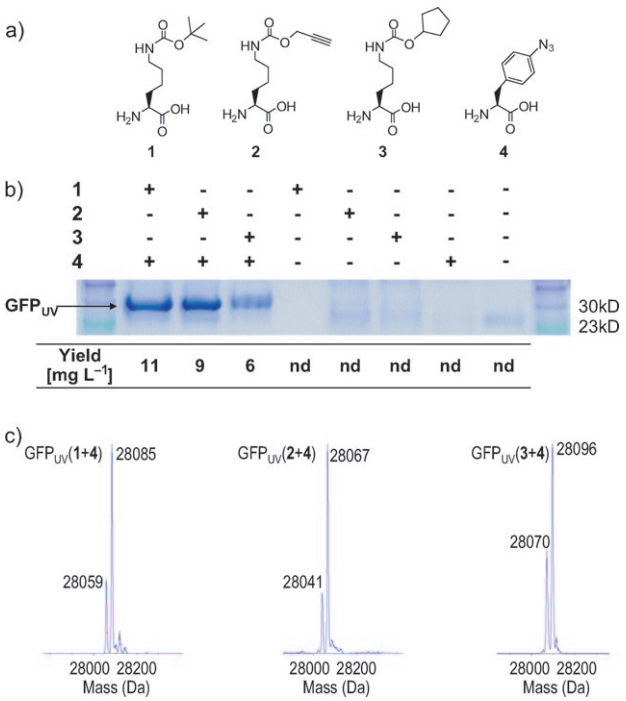

Figure 2. a) Structures of four NAAs. b) The expression level of full-length GFP_{UV} with amber mutation at position 1 and ochre mutation at position 149 from cells transformed with pEVOL-AzFRS and pPylRS-pylT-GFP1TAG149TAA under different conditions. All the NAAs were supplemented into media at 1 mM. nd = not determined. c) ESI-MS analysis of purified GFP_{UV}(**1**+**4**), GFP_{UV}(**2**+**4**), and GFP_{UV}(**3**+**4**). Only deconvoluted mass spectra are presented.

observed previously.[11] As wild-type PylRS also charges pylT with N^{ε}-propargyloxycarbonyl-L-lysine[2h] (**2**; Figure 2a) and N^{ε}-cyclopentyloxycarbonyl-L-lysine[12] (**3**; Figure 2a), the incorporation of either of these two NAAs together with **4** into GFP_{UV} was also tested in cells transformed with pEVOL–AzFRS and pPylRS–pylT–GFP1TAG149TAA. Growing cells in 2YT medium supplemented with 1 mM **2** (or **3**) and 1 mM **4** afforded full-length GFP_{UV} in good yields (Figure 2b). No full-length GFP_{UV} expression was detected when only one NAA was present in the medium. ESI-MS analysis of the purified proteins confirmed the expected incorporations (GFP_{UV} incorporated with **4** and **2** (GFP_{UV}(**2**+**4**)): 28065 Da (calculated), 28067 Da (detected); GFP_{UV} incorporated with **4** and **3** (GFP_{UV}(**3**+**4**)): 28095 Da (calculated), 28096 Da (detected)). Similar decomposition of the azide group during MS analysis was observed in both cases (Figure 2c).

As GFP_{UV}(**2**+**4**) contains both an alkyne group and an azide group, we tested the feasibility of separately labeling this protein with different fluorescent dyes by performing click reactions on these two functional groups.[2h,13] The reaction of GFP_{UV}(**2**+**4**) with 3-azido-7-hydroxycoumarin[14] (**5**; Figure 3) in the presence of a Cu^{I} catalyst[13b] led to a labeled GFP_{UV} that emitted strong blue fluorescence under long wavelength UV light (365 nm) after the protein was denatured and analyzed in a SDS-PAGE gel (Figure 3, lane 2). The same labeling reaction between wild-type GFP_{UV} (wtGFP_{UV}) did not give any detectable blue fluorescence when excited. As both proteins were denatured prior to

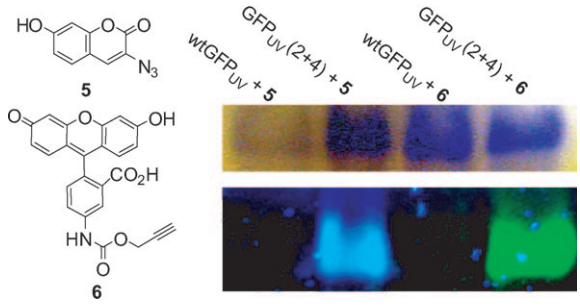

Figure 3. Labeling wtGFP$_{UV}$ and GFP$_{UV}$(**2**+**4**) with **5** and **6**. Top panel: proteins stained with Coomassie blue in a SDS-PAGE gel; bottom panel: fluorescent imaging of the same gel under UV light (365 nm). The image shows real colors captured by a regular camera. The faint protein band (at top left) was due to the partial precipitation of wtGFP$_{UV}$ during the labeling reaction.

SDS-PAGE analysis, the endogenous fluorophore of GFP$_{UV}$ was quenched and did not interfere with the analysis. The similar click chemistry reaction was also carried out between GFP$_{UV}$(**2**+**4**) and a propargyl-conjugated fluorescein (**6**; Figure 3). The labeled protein emitted strong green fluorescence when excited at 365 nm (Figure 3, lane 4). The control reaction on wtGFP$_{UV}$ again yielded no detectable fluorescently labeled protein (Figure 3, lane 3). These experiments demonstrate that both side-chain functional groups of **2** and **4** are active after their incorporation into proteins and can be used separately to effect site-specific protein modifications. As a large excess of **5** or **6** relative to the protein were used during labeling experiments, the self-coupling reaction between the azide and alkyne groups from two GFP$_{UV}$(**2**+**4**) molecules was prevented. No GFP$_{UV}$(**2**+**4**) dimer was observed after the reactions.

To generalize the method, we tested the feasibility of using other evolved PylRS–pylT$_{UUA}$ and *Mj*TyrRS–*Mj*tRNA$^{Tyr}_{CUA}$ pairs to genetically incorporate their cognate NAAs into one protein in *E. coli*. We replaced the PylRS gene in the pPylRS–pylT–GFP1TAG149TAA plasmid with the AcKRS gene[2f] and then transformed *E. coli* BL21 cells with the modified plasmid together with pEVOL–AzFRS. Growing the transformed cells in 2YT medium supplemented with 1 mM **4** and 5 mM AcK led to full-length GFP$_{UV}$ expression with a yield of 6.4 mg L^{-1} (Supporting Information, Figure S2). Similarly, we also replaced AzFRS in the pEVOL–AzFRS with an evolved *Mj*TyrRS (sTyrRS)[15] that is specific for *O*-sulfo-L-tyrosine (sTyr) and used the resulting plasmid and pPylRS–pylT–GFP1TAG149TAA to transform *E. coli* BL21 cells. Growing cells in 2YT medium supplemented with 1 mM sTyr and 1 mM **2** led to full-length GFP$_{UV}$ expression with a yield of 0.4 mg L^{-1} (Supporting Information, Figure S2). The low GFP$_{UV}$ expression yield in this case is most likely due to the low efficiency of the evolved sTyrRS.[15] In both cases, no full-length GFP$_{UV}$ was expressed when only one NAA was provided in the media. These results indicate that our method can be generally applied to combine any two evolved PylRS–pylT$_{UUA}$ and *Mj*TyrRS–*Mj*tRNA$^{Tyr}_{CUA}$ pairs to genetically incorporate their cognate NAAs into a single protein in *E. coli*.

In summary, we have developed a facile system for genetic incorporation of two different NAAs at two defined sites of a single protein in *E. coli* with moderate to high protein production yields. This technique will greatly expand the scope of potential applications for the genetic NAA incorporation approach, and it can be applied to install a FRET pair to a protein for conformation and dynamics studies, synthesize proteins with two different post-translational modifications for functional analysis, or generate phage-displayed peptide libraries with the expanded diversity of the displayed peptides.[16,17]

Experimental Section

The construction of pPylRS-pylT-GFP1TAG149TAA and other plasmids all followed standard cloning and QuikChange site-directed mutagenesis procedures using Platinum Pfx (Invitrogen) and PfuTurbo (Stratagene) DNA polymerases. Sequences of all the constructed plasmids were verified by DNA sequencing. For details of plasmid construction, synthesis and characterization of NAAs and fluorescent dyes, GFP$_{UV}$ expression, ESI-MS analysis and fluorescent labeling of expressed proteins, please see the Supporting Information.

Received: January 26, 2010
Revised: February 19, 2010
Published online: March 25, 2010

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

[1] R. Further, *Protein Sci.* **1998**, *7*, 419.
[2] a) J. Xie, P. G. Schultz, *Methods* **2005**, *36*, 227; b) L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225; c) L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34; *Angew. Chem. Int. Ed.* **2005**, *44*, 34; d) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818; e) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem.* **2009**, *121*, 1661; *Angew. Chem. Int. Ed.* **2009**, *48*, 1633; f) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232; g) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187; h) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720.
[3] Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell, W. Liu, *Mol. Biosyst.* **2010**, *6*, 683.
[4] a) J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775; b) X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem.* **2009**, *121*, 9348; *Angew. Chem. Int. Ed.* **2009**, *48*, 9184; c) D. P. Nguyen, M. M. G. Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 14194.
[5] A. Ambrogelly, S. Gundllapalli, S. Herring, C. Polycarpo, C. Frauer, D. Soll, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 3141.
[6] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Soll, O. Nureki, *Nature* **2009**, *457*, 1163.
[7] Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Z. Wang, D. H. Russell, W. Liu, *Bioorg. Med. Chem. lett.* **2010**, *20*, 878.
[8] a) F. H. Crick, *J. Mol. Biol.* **1966**, *19*, 548; b) G. Eggertsson, D. Soll, *Microbiol. Rev.* **1988**, *52*, 354.
[9] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361.

[10] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052.
[11] a) W. Liu, A. Brock, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239; b) K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19.
[12] C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695.
[13] a) H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004; b) S. Schoffelen, M. H. L. Lambermon, M. B. van Eldijk, J. C. M. van Hest, *Bioconjugate Chem.* **2008**, *19*, 1127.
[14] K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill, Q. Wang, *Org. Lett.* **2004**, *6*, 4603.
[15] C. C. Liu, P. G. Schultz, *Nat. Biotechnol.* **2006**, *24*, 1436.
[16] F. Tian, M. L. Tsao, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 15962.
[17] **Note added in proof**: While our original submission was under review, Chin et al. published work that utilized both amber and quadruple codon suppressions for genetic incorporation of two noncanonical amino acids into one protein: H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature*, **2010**, *464*, 441.

