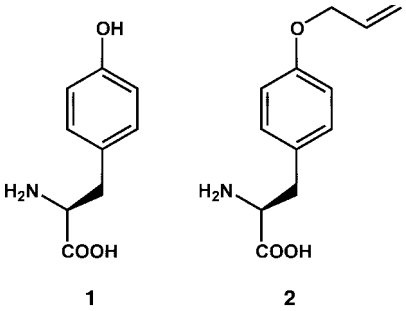
[12] The side product related to **6** was also formed, but the optical purity was not determined.
[13] M. Shimazaki, H. Haram, K. Suzuki, G.-i. Tsuchihashi, *Tetrahedron Lett.* **1987**, *28*, 5891; K. Suzuki, E. Katayama, G.-i. Tsuchihashi, *Tetrahedron Lett.* **1984**, *25*, 1817.
[14] Rearrangements of optically active vinyl epoxides mediated by Lewis acids give quaternary asymmetric carbon centers with inversion of the configuration: a) M. E. Jung, D. C. D'Amino, *J. Am. Chem. Soc.* **1995**, *117*, 7379; b) for rearrangements that involve a sterically restricted cation center, see: S. A. Monti, S.-C. Chen, Y.-L. Yang, S.-S. Yuan, O. P. Bourgeois, *J. Org. Chem.* **1978**, *43*, 4062.
[15] The enantiomeric purity of **15** was determined by HPLC (Chiralcel-OJ, Daicel).
[16] C. H. Senanayake, R. D. Larsen, T. J. Bill, J. Liu, E. G. Corley, P. J. Reider, *Synlett* **1994**, 199.

## The Selective Incorporation of Alkenes into Proteins in *Escherichia coli***

Zhiwen Zhang,[+] Lei Wang,[+] Ansgar Brock, and Peter G. Schultz*

The addition of amino acids with novel functional groups to the genetic code of *Escherichia coli* should greatly enhance our ability to study protein structure and function, as well as generate proteins with novel properties. We recently showed that the unnatural amino acids *O*-methyl-L-tyrosine and L-3-(2-naphthyl)alanine can be site-specifically incorporated into proteins in *Escherichia coli* with high efficiency and fidelity.[1, 2] This result requires the addition of an orthogonal tRNA–codon pair and aminoacyl-tRNA synthetase to the translational machinery of the cell. The new synthetase (and only this synthetase) aminoacylates the orthogonal tRNA (and only this tRNA) with the unnatural amino acid only, which is inserted into proteins in response to the amber codon, TAG.[3] We report here the site-specific incorporation of *O*-allyl-L-tyrosine (**2**) into proteins in *E. coli*. The alkene functional group of this unnatural amino acid should provide new chemical methods for the selective modification of proteins.

Previously we generated an orthogonal tRNA–synthetase pair, $mutRNA^{Tyr}_{CUA}$–TyrRS, in *E. coli* by modifying the $tRNA^{Tyr}$–TyrRS pair of *Methanococcus jannaschii*.[4, 5] This $mutRNA^{Tyr}_{CUA}$ is not aminoacylated by endogenous synthetases in *E. coli*, and functions well in translation. The TyrRS does not aminoacylate *E. coli* tRNAs,[6] but aminoacylates the $mutRNA^{Tyr}_{CUA}$ with tyrosine (**1**); the acylated $mutRNA^{Tyr}_{CUA}$ inserts tyrosine in response to the amber nonsense codon. To change the substrate specificity of the TyrRS so that it aminoacylates the $mutRNA^{Tyr}_{CUA}$ with **2** and not with any common amino acids, a mutant TyrRS library was generated and selected. Based on an analysis of the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus*,[7] five residues (Tyr 32, Glu 107, Asp 158, Ile 159, and Leu 162) in the active site of *M. jannaschii* TyrRS that are within 6.5 Å of the *para* position of the aryl ring of tyrosine were randomly mutated.[1, 8] This mutant library was first subjected to a positive selection based on the suppression of an amber codon introduced at a nonessential position (Asp 112) in the chloramphenicol acetyl transferase (CAT) gene. Cells transformed with the mutant TyrRS libraries, the $mutRNA^{Tyr}_{CUA}$ gene, and the amber mutant CAT gene were grown in minimal media containing 1 mM **2** and 70 μg mL$^{-1}$ chloramphenicol. The survivors must encode a mutant TyrRS that aminoacylates the $mutRNA^{Tyr}_{CUA}$ with either **2** or endogenous amino acids. To remove mutant synthetases with specificities for endogenous amino acids, a negative selection was applied. Three amber codons were introduced at nonessential positions (Gln 2, Asp 44, Gly 65) in the toxic barnase gene.[9] Cells expressing the mutant synthetase from the positive selection, the $mutRNA^{Tyr}_{CUA}$ gene, and the amber mutant barnase gene were grown in Luria–Bertani (LB) media in the absence of **2**. Cells encoding synthetases with specificities for any endogenous amino acids will produce barnase and die. Only those encoding a mutant synthetase with specificity for **2** can survive.




After three rounds of positive selection alternating with two rounds of negative selection, a clone was evolved whose survival in chloramphenicol was dependent on the presence of **2** when the selected mutant TyrRS gene (AL-TyrRS) was coexpressed with the Asp 112amber CAT gene and the $mutRNA^{Tyr}_{CUA}$ gene. Cells can survive in 120 μg mL$^{-1}$ chloram-

[*] Prof. 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)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu
Dr. Z. Zhang
Department of Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
L. Wang
Department of Chemistry
University of California at Berkeley
Berkeley, CA 94720 (USA)
Dr. A. Brock
Genomics Institute of the Novartis Research Foundation
10675 John Jay Hopkins Drive, San Diego, CA 92121 (USA)
[+] These authors contributed equally to this work.
[**] Support for this work was provided by the Department of Energy, Basic Energy Sciences Program, and the National Institutes of Health. We thank Dr. Steve Santoro for plasmid pREP(2) and Sheng Ding for helpful discussions. This is TSRI manuscript no. 14950-CH.

  1433-7851/02/4115-2840 $ 20.00+.50/0 

phenicol in the presence of **2**, and up to 10 μg mL$^{-1}$ chloramphenicol in the absence of **2**. For comparison, *E. coli* cells expressing the Asp112amber CAT gene and the mutRNA$^{Tyr}_{CUA}$ gene survive in 4 μg mL$^{-1}$ chloramphenicol.[5] This result suggests that the AL-TyrRS has higher activity for **2** than for natural amino acids. The evolved AL-TyrRS has the following mutations: Glu 107 → Ala 107, Asp 158 → Cys 158, and Ile 159 → Ala 159. The residues Tyr 32 and Leu 162 remain unchanged. The mutations of Glu 107 → Ala 107 and Ile 159 → Ala 159 may enlarge the active site of the mutant synthetase to accommodate the allyl group. The exact roles of these mutations will be revealed by solving the crystal structure of AL-TyrRS, which is underway.

To confirm that the observed phenotype is caused by the site-specific incorporation of **2** by the mutRNA$^{Tyr}_{CUA}$ – AL-TyrRS pair, a mutant Z domain[10, 11] protein was produced and characterized. An amber codon was introduced at the seventh position in the gene encoding the Z domain. A $His_6$ tag was added to the C terminus of the Z domain to facilitate protein purification by $Ni^{2+}$-affinity chromatography. As a positive control, the wild type *M. jannaschii* TyrRS was coexpressed with the mutRNA$^{Tyr}_{CUA}$, which resulted in the suppression of the amber codon with tyrosine and production of full-length Z-domain protein (Figure 1). When AL-TyrRS

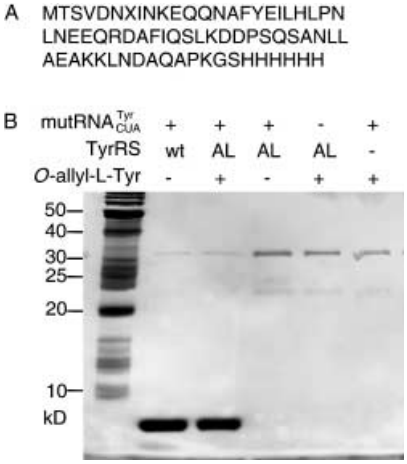



Figure 1. A) Amino acid sequence of the Z domain. X indicates the position for incorporation of **2** encoded by the amber nonsense codon; B) SDS-PAGE analysis of the accumulation of Z domain under different conditions. The far left lane is a molecular-weight marker. Expression conditions are noted at the top of each lane. Proteins were purified by $Ni^{2+}$-affinity chromatography and the gel was silver-stained.[12] AL-TyrRS = *O*-allyl-L-tyrosine-specific mutant synthetase; wt = wild type.

was expressed together with the mutRNA$^{Tyr}_{CUA}$ in the presence of **2**, full-length Z domain was also obtained. In the absence of either the AL-TyrRS, the mutRNA$^{Tyr}_{CUA}$, or **2**, no full-length Z domain was observed. These results show that full-length mutant protein is produced only in the presence of the AL-TyrRS, the mutRNA$^{Tyr}_{CUA}$, and **2**. The yield of full-length mutant Z domain containing **2** is 5.6 mg L$^{-1}$ in minimal media. For comparison, the yield of Z-domain is 9.2 mg L$^{-1}$ when the mutRNA$^{Tyr}_{CUA}$ and wild-type *M. jannaschii* TyrRS are coexpressed.

The mutant Z-domain protein expressed by the mutRNA$^{Tyr}_{CUA}$ – AL-TyrRS was analyzed by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). The experimental monoisotopic mass of the intact protein was 7963.891 Da, which is within 1 ppm of the theoretical mass of 7963.889 Da. Another major signal corresponds to the protein without the first methionine moiety ($M_{Experimental}$ = 7832.861 Da, $M_{Theoretical}$ = 7832.849 Da). The signal-to-noise ratio of more than 1500 observed in the intact protein mass spectrum suggests a fidelity for the incorporation of **2** of better than 99.8 %. This result clearly demonstrates the site-specific incorporation of **2** in response to the amber codon by the mutRNA$^{Tyr}_{CUA}$ – AL-TyrRS, and that other endogenous *E. coli* synthetases do not utilize **2** as a substrate.

In summary, a useful nonnatural amino acid, *O*-allyl-L-tyrosine, has been site-specifically incorporated into proteins in vivo. The allyl group is versatile in organic transformations, including metathesis,[13] Diels-Alder and 1,3-dipolar cycloaddition reactions.[14] Olefin metathesis has been used successfully to cross-link amino acid derivatives[15] and cyclize peptides.[16, 17] Recent progress in developing water-soluble ruthenium catalysts[18] should facilitate the application of this reaction to the modification of proteins containing alkene groups.[19, 20] Moreover, the allyl side chain of this nonnatural amino acid may confer new physical properties on proteins as well.

## *Experimental Section*

All chemicals were purchased from Aldrich. NMR spectroscopic data was recorded using a Bruker AMX 400. Mass spectra of small molecules were obtained from Scripps Center for Mass Spectrometry.

Synthesis of **2**: *O*-allyl-L-tyrosine (**2**) was synthesized according to the published procedures[21] with minor modifications. *N*-(*tert*-butoxycarbonyl)-L-tyrosine (2.95 g, 10 mmole) was dissolved in 80 mL of *N*,*N*-dimethylformamide (DMF). The solution was cooled to 5 °C and NaH (0.63 g, 26 mmole) was added. The reaction mixture was allowed to warm up to 10 °C and stirred for an additional 2 h. Allyl bromide (1.33 g, 11 mmole) was then added with stirring and the reaction mixture was warmed to room temperature and stirred for an additional 4 h. Water was added and the aqueous layer was extracted with ethyl acetate and $CH_2Cl_2$. The organic layer was dried over anhydrous $MgSO_4$. The organic solvent was removed to afford a white solid, which was then refluxed in 4 M HCl in 1,4-dioxane for 4 h. All solvent was evaporated to give the desired product as a white solid (1.9 g, 86 %); $^1$H NMR ($CD_3OD$): $\delta$ = 3.12 (m, 2H), 4.13 (t, $J$ = 5.1 Hz, 1H), 4.53 (d, $J$ = 4.6 Hz, 2H), 5.37 (q, $J$ = 17.4, 11.3, 10.6 Hz, 1H), 5.99 (m, 1H), 6.91 (d, $J$ = 8.4 Hz, 2H), 7.12 ppm (d, $J$ = 8.4 Hz, 2H);[21] $^{13}$C NMR ($CD_3OD$): $\delta$ = 40.0, 69.5, 73.1, 115.8, 117.4, 130.5, 131.5, 135.6, 158.4, 177.5 ppm; $m/z$ (ESI, $MH^+$): 222.19.

Selection for AL-TyrRS: The positive selection based on the suppression of an amber codon in chloramphenicol acetyl transferase was carried out as described.[1] For the negative selection, plasmid pLWJ17B3 was used to express the mutRNA$^{Tyr}_{CUA}$ under the control of the *lpp* promoter and *rrnC* terminator, and the barnase gene with three amber codons at Gln 2, Asp 44, and Gly 65 under arabinose induction. After positive selection, pBK plasmids encoding mutant TyrRS were isolated and transformed into *E. coli* DH10B competent cells harboring pLWJ17B3. Cells were grown in LB media containing 0.2 % arabinose, 50 μg mL$^{-1}$ kanamycin, and 35 μg mL$^{-1}$ chloramphenicol. After 8 h, cells were pelleted, and pBK plasmids were purified for further rounds of selection. After sequential positive, negative, positive, negative, then positive selection, the candidate pBK-ALRS encoding AL-TyrRS was identified and characterized using an in vivo chloramphenicol acetyl transferase assay.[5]

  1433-7851/02/4115-2841 $ 20.00+.50/0 

Protein expression, purification, and characterization: Plasmid pLEIZ was used to express the Z-domain gene with an amber codon at the seventh position under the control of a bacteriophage *T5* promoter and $t_0$ terminator, and the mutRNA$^{Tyr}_{CUA}$ gene under the control of the *lpp* promoter and *rrnC* terminator. The AL-TyrRS gene was encoded in plasmid pBK-ALRS under the control of the constitutive *E. coli* GlnRS promoter and terminator. *E. coli* DH10B cells cotransformed with pLEIZ and pBK-ALRS were grown in minimal media containing 1% glycerol and 0.3 mM leucine (GMML media) with 25 μg mL$^{-1}$ kanamycin, 34 μg mL$^{-1}$ of chloramphenicol, and 0.5 mM **2**. When cells reach an optical density ($OD_{600}$) value of 0.5, isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM) was added to induce protein expression. After 4 h, cells were pelleted and the protein was purified by $Ni^{2+}$-affinity chromatography according to the manufacturer's protocol under denaturing conditions (Quiagen, Valencia, CA). Proteins were then desalted by using a PD-10 column (Amersham Pharmacia, Piscataway, NJ) and eluted in water. The yield of protein was measured by Bradford assay (BCA kit, Biorad, Hercules, CA). Aliquots of protein were used for SDS-PAGE and mass spectroscopic analysis.

Received: April 19, 2002 [Z19118]


[1] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498 – 500.

[2] L. Wang, A. Brock, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1836 – 1837.

[3] L. Wang, P. G. Schultz, *Chem. Commun.* **2002**, 1 – 11.

[4] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 5010 – 5011.

[5] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883 – 890.

[6] B. A. Steer, P. Schimmel, *J. Biol. Chem.* **1999**, *274*, 35601 – 35606.

[7] P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* **1989**, *208*, 83 – 98.

[8] A second mutant *M. jannaschii* TyrRS library was also constructed in such a way that six residues (Tyr32, Ala67, His70, Gln155, Asp158, Ala167) within 6.9 Å of the *meta* position of the tyrosine aryl ring were randomly mutated. This library was also subjected to the genetic selection, but failed to produce a mutant TyrRS which charges *O*-allyl-L-tyrosine.

[9] D. R. Liu, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4780 – 4785.

[10] B. Nilsson, T. Moks, B. Jansson, L. Abrahmsen, A. Elmblad, E. Holmgren, C. Henrichson, T. A. Jones, M. Uhlen, *Protein Eng.* **1987**, *1*, 107 – 113.

[11] We have also successfully incorporated **2** into mouse dihydrofolate reductase (DHFR) with high fidelity (> 99%).

[12] The top two bands in Figure 1 with molecular weights of approximately 32 kD and 23 kD are *E. coli* proteins which bind to the $Ni^{2+}$ column nonspecifically.

[13] A. Fürstner, *Angew. Chem.* **2000**, *112*, 3140 – 3172; *Angew. Chem. Int. Ed.* **2000**, *39*, 3012 – 3043.

[14] K. V. Gothelf, K. A. Jorgensen, *Chem. Rev.* **1998**, *98*, 863 – 909.

[15] S. C. G. Biagini, S. E. Gibson, S. P. Keen, *J. Chem. Soc Perkin Trans. 1* **1998**, 2485 – 2499.

[16] J. F. Reichwein, C. Versluis, R. M. J. Liskamp, *J. Org. Chem.* **2000**, *65*, 6187 – 6195.

[17] H. E. Blackwell, J. D. Sadowsky, R. J. Howard, J. N. Sampson, J. A. Chao, W. E. Steinmetz, D. J. O'Leary, R. H. Grubbs, *J. Org. Chem.* **2001**, *66*, 5291 – 5302.

[18] D. M. Lynn, B. Mohr, R. H. Grubbs, L. M. Henling, M. W. Day, *J. Am. Chem. Soc.* **2000**, *122*, 6601 – 6609.

[19] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282 – 1288.

[20] K. L. Kiick, J. C. M. van Hest, D. A. Tirrell, *Angew. Chem.* **2000**, *112*, 2232 – 2236; *Angew. Chem. Int. Ed.* **2000**, *39*, 2148 – 2152.

[21] J. M. Fraile, J. I. Garcia, J. A. Mayoral, A. J. Royo, *Tetrahedron: Asymmetry*, **1996**, *7*, 2263 – 2276.


# Synthesis and Characterization of the Neutral "Digallene" Ar′GaGaAr′ and Its Reduction to $Na_2Ar'GaGaAr'$ (Ar′ = 2,6-$Dipp_2C_6H_3$, Dipp = 2,6-*i*$Pr_2C_6H_3$)**

Ned J. Hardman, Robert J. Wright, Andrew D. Phillips, and Philip P. Power*

The publication of the remarkable gallium compound $Na_2Ar^*GaGaAr^*$ (**1**; Ar* = 2,6-$Trip_2C_6H_3$, Trip = 2,4,6-*i*$Pr_3C_6H_2$) in 1997 resulted in much controversy owing to the claim that it contained a Ga–Ga triple bond.[1] Initially, the triple bonding in this molecule was justified on the basis of a short Ga–Ga distance (2.319(3) Å) and the correspondence of the putative $[Ar^*GaGaAr^*]^{2-}$ ion to the neutral germanium species Ar*GeGeAr*—a germanium – alkyne analogue. Although the existence of Ga–Ga triple bonding has received support from some calculations,[2–6] others have questioned this view on the basis of 1) the *trans*-bent structure of the $C_{ipso}$-Ga-Ga-$C_{ipso}$ array which indicates lone pair character at the gallium center,[7–11] 2) the Na–aryl ring[8] and Na–Ga interactions[11] which shorten the Ga–Ga distance, and 3) the role of the *para*-*i*Pr groups on the flanking aryl rings which cause Ga-Ga-C angular distortions that can strengthen the Ga–Ga bond.[11] Force constant calculations have also pointed to a relatively weak Ga–Ga interaction.[12, 13] The publication of the cluster species $K_2Ar^*Ga_4Ar^*$, which contains a $Ga_4$ ring with no Ga–Ga triple bonding as part of an octahedral $K_2Ga_4$ core, has underlined the importance of the alkali metal for the stability of **1**.[14] However, apart from this isolated report, all arguments regarding the Ga–Ga bonding in **1** have been grounded in calculations of various degrees of sophistication[2–11] and the original experimentally determined structural parameters.[1] In 1998 several experiments were suggested whose object was the elucidation of the important factors governing the nature of the Ga–Ga bond.[15] Among these were the investigation of the effects of changing or removing the alkali metal ions and the isolation and characterization of the neutral "digallene" species Ar*GaGaAr* which, should contain a Ga-Ga double bond if the assumption of triple bonding in **1** was correct. The former question has been partly answered through the synthesis of $K_2Ar^*Ga_4Ar^*$. However, no stable neutral Ga–Ga bonded dimers of the general formula RGaGaR (R = organic or related group) have yet been described. Calculations on a variety of model species, including HGaGaH,[10, 11, 16–19] MeGaGaMe,[9–11] and PhGaGaPh,[11] as well as IR spectroscopy of HGaGaH in a frozen matrix[19, 20] point to weak Ga–Ga bonding. It is now shown that the compound Ar′GaGaAr′ (**2**; Ar′ = 2,6-$Dipp_2C_6H_3$, Dipp = 2,6-*i*$Pr_2C_6H_3$), can be isolated and characterized with

[*] Prof. P. P. Power, Dr. N. J. Hardman, R. J. Wright, Dr. A. D. Phillips
Department of Chemistry
University of California
One Shields Avenue, Davis, CA 95616 (USA)
Fax: (+1) 530-752-8995
E-mail: pppower@ucdavis.edu

[**] This work was supported by the National Science Foundation Grant No. 0094913.

  1433-7851/02/4115-2842 $ 20.00+.50/0 