

Functional replacement of the endogenous tyrosyl-tRNA synthetase–tRNA^{Tyr} pair by the archaeal tyrosine pair in *Escherichia coli* for genetic code expansion

Fumie Iraha¹, Kenji Oki², Takatsugu Kobayashi¹, Satoshi Ohno³, Takashi Yokogawa³, Kazuya Nishikawa³, Shigeyuki Yokoyama^{1,2,*} and Kensaku Sakamoto^{1,*}

¹RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045,

²Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 and ³Department of Biomolecular Science, Faculty of Engineering, Gifu University, Yanagido 1-1, Gifu 501-1193, Japan

Received November 10, 2009; Revised January 25, 2010; Accepted January 27, 2010

ABSTRACT

Non-natural amino acids have been genetically encoded in living cells, using aminoacyl-tRNA synthetase–tRNA pairs orthogonal to the host translation system. In the present study, we engineered *Escherichia coli* cells with a translation system orthogonal to the *E. coli* tyrosyl-tRNA synthetase (TyrRS)–tRNA^{Tyr} pair, to use *E. coli* TyrRS variants for non-natural amino acids in the cells without interfering with tyrosine incorporation. We showed that the *E. coli* TyrRS–tRNA^{Tyr} pair can be functionally replaced by the *Methanocaldococcus jannaschii* and *Saccharomyces cerevisiae* tyrosine pairs, which do not cross-react with *E. coli* TyrRS or tRNA^{Tyr}. The endogenous TyrRS and tRNA^{Tyr} genes were then removed from the chromosome of the *E. coli* cells expressing the archaeal TyrRS–tRNA^{Tyr} pair. In this engineered strain, 3-iodo-L-tyrosine and 3-azido-L-tyrosine were each successfully encoded with the amber codon, using the *E. coli* amber suppressor tRNA^{Tyr} and a TyrRS variant, which was previously developed for 3-iodo-L-tyrosine and was also found to recognize 3-azido-L-tyrosine. The structural basis for the 3-azido-L-tyrosine recognition was revealed by X-ray crystallography. The present engineering allows *E. coli* TyrRS variants for non-natural amino acids to be developed in *E. coli*, for use in both eukaryotic and bacterial cells for genetic code expansion.

INTRODUCTION

The repertoire of genetically encoded amino acids has been expanded in *Escherichia coli*, yeast, insect and mammalian cells, with the purpose of increasing the chemical and structural diversity in proteins (1–6). The genetically encoded ‘non-natural’ amino acids have unique chemical groups and functionalities not found among the 20 canonical amino acids, such as heavy atoms, photo-reactive linkers, fluorescent groups and chemical groups for site-specific labeling (7). The incorporation of these novel structures into proteins at desired sites promises to facilitate protein science and engineering.

The genetic encoding of a non-natural amino acid requires the development of an aminoacyl-tRNA synthetase (aaRS)–tRNA pair specific to the amino acid. To date, several aaRS species have been engineered to recognize useful non-natural amino acids, and among these, the majority are L-tyrosine or L-phenylalanine derivatives recognized by variants of tyrosyl-tRNA synthetase (TyrRS) (7). In living cells, non-natural amino acids are encoded with the amber codon in most cases, by using the amber suppressor tRNA–aaRS pair from a different organism that does not cross-react with the endogenous tRNA or aaRS (8). The bacterial TyrRS, for example, does not recognize the tRNA^{Tyr} molecules from eukaryotes or archaea, while the TyrRS molecules from these two kingdoms do not recognize the bacterial tRNA^{Tyr}. In fact, the bacterial TyrRS–tRNA^{Tyr} pair does not cross-react with the endogenous tRNA and aaRS species in eukaryotic host cells, and has been used to achieve the site-specific incorporation of L-tyrosine

*To whom correspondence should be addressed. Tel: +81 45 503 9196; Fax: +81 45 503 9195; Email: yokoyama@biochem.s.u-tokyo.ac.jp
Correspondence may also be addressed to Kensaku Sakamoto. Tel: +81 45 503 9459; Fax: +81 45 503 9458; Email: sakamoto@gsc.riken.jp

derivatives into proteins only at the amber codon (2,9,10). On the other hand, the genetic encoding of L-tyrosine derivatives in *E. coli* cells was achieved using the TyrRS-tRNA^{Tyr} pair from *Methanocaldococcus jannaschii* (1,7). Thus, the *E. coli* and archaeal TyrRS-tRNA^{Tyr} pairs are used in eukaryotic and *E. coli* cells, respectively. Other aaRSs, such as tryptophanyl-tRNA synthetase (11), phenylalanyl-tRNA synthetase (5,12), leucyl-tRNA synthetase (13) and lysyl-tRNA synthetase (14), also take advantage of the orthogonal tRNA specificities across different kingdoms.

Pyrrolysyl-tRNA synthetase (PylRS) differs from these aaRS species, in that the PylRS-tRNA^{Pyl} pair from *Methanosarcina mazei* is reportedly orthogonal to both the *E. coli* and eukaryotic translation systems (15–17). Archaeal PylRS variants can be developed in *E. coli* cells, and are then used for incorporating L-lysine derivatives into proteins in *E. coli*, yeast and mammalian cells (16–20). With this advantage, the PylRS-tRNA^{Pyl} pair is a promising system for genetic code expansion. The usefulness of *E. coli* TyrRS variants would also be increased, if the *E. coli* translation system could be engineered to be orthogonal to the *E. coli* TyrRS-tRNA^{Tyr} pair. This modification to *E. coli* cells may be achieved by substituting an archaeal or eukaryal tyrosine pair for the endogenous tyrosine pair.

In the present study, the genes encoding TyrRS and tRNA^{Tyr} were disrupted in the chromosome of the *E. coli* cells expressing the *M. jannaschii* TyrRS-tRNA^{Tyr} pair. The engineered cells then hosted the *E. coli* pair of a TyrRS variant and the amber suppressor tRNA^{Tyr}, to genetically encode either 3-iodo-L-tyrosine or 3-azido-L-tyrosine, depending on the presence of the corresponding amino acid in the growth medium. This TyrRS variant, designated as iodoTyrRS-*ec*, was originally developed for 3-iodo-L-tyrosine (21), and has been used for the site-specific incorporation of this amino acid into proteins in mammalian cells (2,22). We found that iodoTyrRS-*ec* also recognizes 3-azido-L-tyrosine, and the structural basis for its recognition by this variant was elucidated by X-ray crystallography.

MATERIALS AND METHODS

Strains, growth conditions, growth rate, reagents and non-natural amino acids

Escherichia coli strains TOP10, BL21(DE3) and DH5 α were purchased from Invitrogen, Novagen and Toyobo (Tokyo, Japan), respectively. Kanamycin (Km) (30 μ g/ml), chloramphenicol (Cm) (34 μ g/ml), ampicillin (Amp) (10 μ g/ml), Zeocin (25 μ g/ml), L-arabinose (0.2%, w/v) and D-glucose (0.4%, w/v) were added to LB media as indicated, except that Cm was added at a concentration of 25 μ g/ml when non-natural amino acids were incorporated into proteins. The growth rate was monitored as the change in the natural logarithm of the optical density at 600 nm of the culture per hour. Zeocin was purchased from InvivoGen (San Diego, USA). 3-Iodo-L-tyrosine was purchased from Sigma-Aldrich. 3-Bromo-L-tyrosine, 3-azido-L-tyrosine and the chemical conjugate between

triarylphosphine and fluorescein were commercially synthesized by Shinsei Chemical Company Ltd (Osaka, Japan).

Plasmids

The *E. coli*, *M. jannaschii* and *Saccharomyces cerevisiae* TyrRS genes, and the variant gene encoding iodoTyrRS-*ec*, were cloned with the *tyrS* promoter into pACYC184 (Nippon Gene Co. Ltd., Japan) to create pEcYS, pMjYS, pScYS and pEcIYS, respectively. The wild-type TyrRS genes with the promoter were also cloned in an ampicillin-resistant ColIb-P9 plasmid, pAp102 (23), to create pApEcYS, pApMjYS and pApScYS, respectively. The archaeal and yeast tRNA^{Tyr} genes were cloned with the *tyrT* promoter into pACYC184, to create pMjYR and pScYR, respectively. These tRNA^{Tyr} genes with the promoter were also cloned in pMjYS and pScYS, to create pMjYSYR and pScYSYR, respectively. Three copies of these tRNA^{Tyr} genes were cloned after the *tyrT* promoter in pAp102, to create pApMjYR and pApScYR, respectively. The archaeal and yeast TyrRS genes with the *tyrS* promoter were cloned in these plasmids, pApMjYR and pApScYR, to create pApMjYSYR and pApScYSYR, respectively. The kanamycin resistance (*kam*) gene was cloned in the ‘amber suppressor control’ plasmid from the Interchange amber suppressor *in vivo* mutagenesis system (Promega), to create the plasmid pACamK. The resulting plasmid carries the replication origin of pACYC184 and an amber mutant chloramphenicol acetyltransferase (CAT) gene, in addition to the cloned *kam* gene. The *E. coli* amber suppressor tRNA^{Tyr} gene was cloned with the *tyrT* promoter into pACamK, to create the plasmid pACamKsupF.

To create the plasmid pMJY, the coding sequence of the Zeocin resistance gene was amplified by polymerase chain reaction (PCR) from the vector pcDNA4/TO (Invitrogen). The sequence 5'-ACGACTCACTATAGGAGGGCC-3', in which the ribosome-binding site (RBS) is underlined, was added at the 5'-end of the coding sequence. The amplified gene was cloned between the SspI-PstI sites of the pBR322 vector, downstream of the promoter of the ampicillin resistance gene and in place of this gene, to create the plasmid pBRZeo. The *M. jannaschii* TyrRS and tRNA^{Tyr} genes with the *tyrS* and *tyrT* promoters, respectively, were cloned in pBRZeo to obtain pMJY.

For incorporating non-natural amino acids into proteins, the *E. coli* amber suppressor tRNA^{Tyr} gene with the *tyrT* promoter was cloned in pAp102 to create pApsYR. The replication origin and the ampicillin resistance gene in the plasmid pGEX-4T-3 (GE Healthcare) were replaced with the origin and *kam* gene from the vector pRSF-1b (Novagen), to create the plasmid pRGexGST. The resulting plasmid retains the glutathione S-transferase (GST) gene under the control of the *tac* promoter. Then, the tyrosine codon at position 84 of GST was mutated to the amber codon, to create pRGexGST(Am84). The wild-type *gst* gene and its mutant with the amber codon at position 25 (24) were cloned after the T7 promoter in pRSF-1b to create

pRSFGST and pRSFGST(Am25), respectively. The rat calmodulin gene and its mutant with the amber codon at position 80 (25) were also cloned after the T7 promoter in pRSF-1b, to create pRSFCAM and pRSFCAM(Am80), respectively.

Construction of the conditional *tyrS* mutant by chromosome engineering

All of the modifications to the *E. coli* chromosome were made using the RT/ET kit (Gene Bridges GmbH, Germany), which utilizes the bacteriophage λ recombination system to promote homologous recombination (26,27). The DNA encoding the RBS, the *araBAD* promoter and the *araC* gene—designated as the *ara* block—was introduced into the chromosome in place of the *tyrS* promoter. For this engineering, the *ara* block was amplified by PCR from the pBAD/His vector (Invitrogen) with the primers 5'-GGTTAATTCTCCTGTTAGC-3' and 5'-CCAATTATGACAACCTTGACG-3'. The right-hand arm (rhM) corresponding to the first 50 nt of the *tyrS* coding sequence was then added at the RBS end of the *ara* block. The *kam* gene with the promoter was amplified from the plasmid pHSG299 (Takara Bio Inc., Japan), and the *rrnC* terminator was added downstream of the *kam* gene. The left-hand arm (lhM) with the sequence from 160- to 110-nt upstream the *tyrS* coding region (5'-ATGCGTGGAAGATTGATCGTCTTGACACCTGAAAAGATGCAAAAATCTTG-3') was added to the *kam* gene after the terminator. Then, the *kam* gene and the *ara* block were linked with each other at the ends opposite to lhM or rhM. The resulting double-stranded DNA fragment consists of lhM, *kam*, *araC*, the *araBAD* promoter, RBS and rhM in this order, and was used to replace the promoter and RBS of *tyrS* in the chromosome. The substitution was confirmed by a sequence analysis of the *tyrS* locus.

Engineering of the *tyrS*, *tyrT*, *tyrU* and *tyrV* loci in the *E. coli* chromosome

The *cat* gene with the promoter was amplified from pACYA184, using the primers A (5'-ACCCGACGCAC TTTGCGCCG-3') and B (5'-TTACGCCCCGCCCTGC CACTC-3'), while the gene without a promoter was amplified using primer B and another primer (5'-GATT TTCAGGAGCTAAGGAAGC-3'). In the following experiments, the *cat* gene was introduced into the chromosome in the same direction as the genes to be replaced. To disrupt *tyrS*, the sequences lhM-S (5'-ATGCGTGGA AGATTGATCGTCTTGACCCCTGAAAAGATGCAA AAATCTTG-3') and rhM-S (5'-ACAGGGAACATGAT GAAAAATATTCTCGCTATCCAGTCTCACGTTGT TTA-3') were added at the 5'- and 3'-ends, respectively, of the *cat* gene with the promoter. To disrupt *tyrT* and *tyrV*, lhM-T1 (5'-AAAATAACTGGTTACCTTTAATCCGTT ACGGATGAAAATTACGCAACCAG-3') and rhM-T (5'-AGTCCCTGAACTTCCCAACGAATCCGCAATT AAATATTCTGCCATGCGG-3') were added at the 5'- and 3'-ends, respectively, of the *cat* gene with the promoter. To disrupt *tyrU*, lhM-U (5'-GTAATCAGTA GGTCACCAGTTCGATTCCGGTAGTCGGCACCAT

CAAGTCC-3') and rhM-U (5'-GGCCACGCGATGGCG TAGCCCGAGACGATAAGTTCGCTTACCGGCTCG AA-3') were added at the 5'- and 3'-ends, respectively, of the *cat* gene without the promoter.

The knocked-in *cat* gene was removed using a linear DNA fragment consisting of the lhM and rhM sequences linked directly to each other. Cells deprived of *cat* were selected in the growth media supplemented with Cm (10 μ g/ml) and ampicillin (200 μ g/ml). The cells that are resistant to Cm and are able to grow in this medium should undergo lysis in the presence of ampicillin, while only sensitive cells should survive. The colonies formed by the surviving cells were replicated on an LB plate containing Cm (10 μ g/ml), to check for the sensitivity to Cm.

The *E. coli* amber suppressor tRNA^{Tyr} was introduced into the chromosome in place of the knocked-in *cat* in the *tyrU* locus, using a linear DNA consisting of lhM-U, the tRNA coding sequence and rhM-U in this order. For the introduction of the suppressor tRNA^{Tyr} in the *tyrT* and *tyrV* loci, the *cat* gene was first introduced into these loci, with lhM-T2 (5'-ACTTTACAGCGGCGCGTCATT TGATATGATGCGCCCCGCTTCCCGATAAG-3') and rhM-T at the 5'- and 3'-ends, respectively. Then, the knocked-in *cat* gene was replaced by the sequence encoding the suppressor tRNA, amplified using primers (5'-TAATTCACCACAGCGATGTG-3' and 5'-TTTGA AAGTGATGGTGGTGG-3) from the genomic DNA of a *supF* strain, in which the *tyrV* gene was mutated to the amber suppressor tRNA^{Tyr}.

Site-specific incorporation of non-natural amino acids into proteins

FT3 cells were transformed with pEciYS, pApsYR and pRGexGST(Am84), and were grown in medium containing Cm, Amp and Km. The expression of the *gst* gene was induced by the addition of IPTG (1 mM, final), and at the same time, 3-iodo-L-tyrosine (0.3 mg/ml, final) was added to the medium. FB3 cells were transformed with pEciYS and either pRSFGST(Am25) or pRSFCAM(Am80), and were grown in medium containing Kam and Cm. The expression of the recombinant proteins was induced with IPTG (1 mM, final), and at the same time, 3-iodo-L-tyrosine or 3-azido-L-tyrosine (0.3 mg/ml, final) was added to the medium. Labeling with the triarylphosphine-fluorescein conjugate was performed as described previously (20,25). Fluorescence from an SDS-gel was detected using an image analyzer, LAS-1000 (Fuji Film, Japan). Mass spectrometric analysis was commercially performed by Shimadzu Biotech (Japan).

Structure determination

The IodoTyrRS-*ec* catalytic domain (residues 1–322) was prepared as described previously (28). The samples were dialyzed against 20 mM Tris-Cl (pH 7.5) containing 50 mM NaCl, 10 mM 2-mercaptoethanol and 1 mM 3-azido-L-tyrosine, and then they were concentrated to 4–6 mg/ml with an Amicon Ultra 4 (Millipore). Crystallization of the complex between IodoTyrRS-*ec* and 3-azido-L-tyrosine was performed as described (28). The samples and crystals were shaded to avoid the

photolysis of 3-azido-L-tyrosine. The X-ray diffraction experiment of the crystals was performed at the beamline BL26B1 at SPring-8 (Harima, Japan). The data were processed with HKL2000 (29). The iodoTyrRS-ec-3-azido-L-tyrosine structure was determined by molecular replacement with the program Molrep (30), using the structure of the complex between iodoTyrRS-ec and 3-iodo-L-tyrosine (28) as a search model. The refinement and model building were performed as described (28). The model of 3-azido-L-tyrosine was generated by the Dundee PRODRG2 server (31).

RESULTS AND DISCUSSION

Complementation of a conditional *E. coli* *tyrS* mutant by the archaeal and eukaryal TyrRS-tRNA^{Tyr} pairs

We examined whether the translational function of *E. coli* TyrRS can be replaced with that of a eukaryal or archaeal TyrRS-tRNA^{Tyr} pair in *E. coli* cells. The expression system for the arabinose operon was used to control the expression of the chromosomal *tyrS* gene encoding *E. coli* TyrRS. The *araBAD* promoter, together with the *araC* gene, was introduced into the chromosome in place of the *tyrS* promoter, to allow the induction of TyrRS expression with L-arabinose. Transcription from the *araBAD* promoter is tightly repressed in the presence of D-glucose (32); the created conditional *tyrS* mutant of *E. coli* grew in the presence of L-arabinose, but did not grow in the presence of D-glucose (Supplementary Figure S1A). Then, the mutant was transformed with the multicopy vector pACYA184, and then reproduced the conditional growth (Supplementary Figure S1B). This conditional growth was complemented by the *tyrS* gene carried on the multicopy plasmid pEcYS, derived from pACYC184 (Supplementary Figure S1B). On the other hand, no complementation was observed with either the multicopy plasmid pMjYS carrying the *M. jannaschii* TyrRS gene or the multicopy plasmid pScYS carrying the *S. cerevisiae* TyrRS gene; these genes were expressed from the *tyrS* promoter (Figure 1A). This observation is consistent with the reports that these archaeal and eukaryal TyrRS species do not recognize *E. coli* tRNA^{Tyr} (8,33).

Next, the archaeal and eukaryal TyrRS species were expressed, together with their cognate tRNA^{Tyr} species, from the multicopy plasmids pMjYSYR and pScYSYR, respectively. In these plasmids, the tRNA^{Tyr} genes were placed after the promoter of the *tyrT* gene, which is one of three genes encoding *E. coli* tRNA^{Tyr}. These plasmids were found to complement the *tyrS* mutant (Figure 1A), indicating that the function of the endogenous TyrRS in *E. coli* can be substituted by the archaeal and eukaryal TyrRS-tRNA^{Tyr} pairs, which are each orthogonal to the host translation system.

To determine whether the overproduction of these TyrRS-tRNA^{Tyr} pairs from multicopy plasmids is necessary for the complementation, we used the low-copy plasmid pAp102 (23) to introduce the tyrosine pairs into the *tyrS* mutant. This plasmid has the ColIb-P9

replication origin and its copy number is 1.7 per cell (34). In a control experiment, the vector pAp102 did not complement the *tyrS* mutant, while a pAp102-derived plasmid carrying the *tyrS* gene (pApEcYS) displayed the complementing activity (Supplementary Figure S1B). The *M. jannaschii* and *S. cerevisiae* TyrRS genes were then each cloned in this low-copy vector, together with three copies of their cognate tRNA^{Tyr} genes, to create the plasmids pApMjYSYR and pApScYSYR, respectively. Neither of these plasmids showed the complementing activity (Supplementary Figure S1C), indicating that these exogenous TyrRS or tRNA^{Tyr} molecules must be expressed from a multicopy plasmid.

We then determined which molecule, TyrRS or tRNA^{Tyr}, needed to be overproduced. First, for the *M. jannaschii* tyrosine pair, the enzyme was expressed from the multicopy plasmid pMjYS, with the tRNA^{Tyr} being expressed from the low-copy plasmid pApMjYS. No complementing activity was observed for this pair of plasmids (Supplementary Figure S1D). Next, TyrRS was expressed from the low-copy plasmid pApMjYS, with the tRNA^{Tyr} expressed from the multicopy plasmid pMjYR, but again, no complementation was detected (Supplementary Figure S1D). These observations indicated that the complementation requires the overproduction of both the archaeal TyrRS and tRNA^{Tyr} molecules.

We also expressed *S. cerevisiae* TyrRS from either the multicopy plasmid pScYS, with the yeast tRNA^{Tyr} being expressed from the low-copy plasmid pApScYR, or the low-copy plasmid pApScYS, with the tRNA^{Tyr} being expressed from the multicopy plasmid pScYR. Neither combination achieved the complementation (Supplementary Figure S1D), indicating that both the yeast TyrRS and tRNA^{Tyr} need to be overproduced. Similar complementation tests were performed for the yeast TyrRS-tRNA^{Tyr} pair at 30°C, the normal growth temperature for yeast. However, no complementation was observed (data not shown), indicating that the optimal temperature for these molecules is probably not relevant to the requirement of the overproduction of the TyrRS-tRNA^{Tyr} pair. Although the reason why the exogenic tyrosine pairs must be overproduced is not clear, this requirement is consistent with a recent report showing that the overproduction of the *M. jannaschii* TyrRS-tRNA^{Tyr} pair in *E. coli* cells is prerequisite for the efficient incorporation of non-natural amino acids into proteins (35).

Replacement of the *E. coli* TyrRS-tRNA^{Tyr} pair by the *M. jannaschii* pair in an *E. coli* strain

The TyrRS and tRNA^{Tyr} genes were removed from the chromosome in an *E. coli* K12 strain expressing the *M. jannaschii* tyrosine pair from the multicopy plasmid pMJY, constructed as described in the 'Materials and Methods' section (Figure 1B). First, the CAT gene was introduced into the chromosome in place of the *tyrS* gene by homologous recombination. Then, the *cat* gene was removed from the chromosome to create the strain FT1 [TOP10 *ΔtyrS* pMJY]. A DNA fragment including the *tyrS* locus was amplified from this strain and analyzed

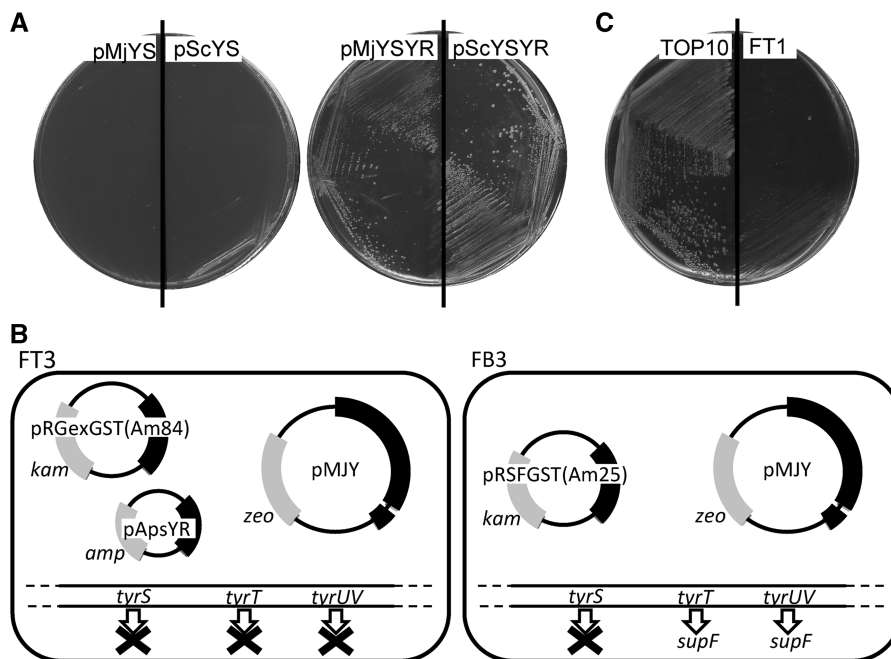


Figure 1. The *tyrS* complementation tests with pMjYS, pScYS, pMjYSYR and pScYSYR (A). The mutant cells transformed with these plasmids were inoculated on the LB plates containing 1% D-glucose and 34 μ g/ml Cm. The three sectors of each half plate represent dilutions of the cells. An illustration of the genetic modifications and plasmid systems in FT3 and FB3 cells (B). The absence of the *E. coli* TyrRS activity in the strain FT1 (C). TOP10 and FT1 cells were each transformed with pACamKsupF and were then inoculated on the LB plate with 10 μ g/ml Cm.

to confirm the disruption of the *tyrS* gene. The absence of the *E. coli* TyrRS activity in this strain was also confirmed by expressing the *E. coli* amber suppressor tRNA^{Tyr}, together with a *cat* amber mutant gene, from the plasmid pACamKsupF. This plasmid conferred chloramphenicol resistance to the parent strain TOP10, but not to FT1 cells (Figure 1C), showing that the *E. coli* TyrRS activity was absent in FT1 cells, and that the *E. coli* suppressor tRNA^{Tyr} is orthogonal to the translation system of the cell.

Next, we removed all three of the *E. coli* tRNA^{Tyr} genes from the chromosome of the FT1 cells. Two of them, *tyrT* and *tyrV*, are located close to each other on the *E. coli* chromosome, and share the *tyrT* promoter for expression. The *cat* gene was introduced into the chromosome of FT1 in place of these genes and their promoter, and the knocked-in *cat* gene was then removed to create the strain FT2 [TOP10 Δ *tyrS* Δ *tyrT* Δ *tyrV* pMJY]. A DNA fragment encompassing the *tyrT* and *tyrV* loci was amplified from this strain and analyzed to confirm the disruption of these tRNA^{Tyr} genes and the *tyrT* promoter. Then, to remove the *tyrU* gene, which is the third copy of the tRNA^{Tyr} gene and is co-transcribed with two tRNA^{Thr} species and a tRNA^{Gly}, the *cat* gene without a promoter was introduced into the chromosome of FT2 in place of *tyrU*, for transcription together with the other tRNAs. The *cat* gene was then removed to create the strain FT3 [TOP10 Δ *tyrS* Δ *tyrT* Δ *tyrU* Δ *tyrV* pMJY]. The absence of the tRNA^{Tyr} gene in the *tyrU* locus was confirmed by a sequence analysis.

In order to show that the strain FT3 lacks *E. coli* tRNA^{Tyr} activity, we expressed iodoTyrRS-*ec*, an *E. coli*

TyrRS variant that recognizes 3-iodo-L-tyrosine, and investigated growth inhibition. When 3-bromo-L-tyrosine was added to the growth media, the growth rate of DH5 α cells, an *E. coli* K-12 strain, overproducing the variant was significantly reduced, as compared to that of the cells overproducing the wild-type *E. coli* TyrRS (Figure 2A). This observation suggested that iodoTyrRS-*ec* is able to efficiently attach 3-bromo-L-tyrosine to the *E. coli* tRNA^{Tyr}, and to inhibit the growth by randomly incorporating this amino acid into proteins at tyrosine positions. On the other hand, strain FT3 is immune to this growth inhibition (Figure 2C), which requires the interaction between iodoTyrRS-*ec* and *E. coli* tRNA^{Tyr}, and this observation indicates the absence of the *E. coli* tRNA^{Tyr} in the cell. Thus, the *E. coli* translation system was successfully engineered to be orthogonal to the *E. coli* TyrRS-tRNA^{Tyr} pair.

The iodoTyrRS-*ec*-*E. coli* suppressor tRNA^{Tyr} pair site-specifically incorporates 3-iodo-L-tyrosine into proteins in FT3 cells

We used iodoTyrRS-*ec* to incorporate 3-iodo-L-tyrosine into proteins at the amber position in the strain FT3. This TyrRS variant was expressed from the *tyrS* promoter on the multicopy plasmid pEciYS, which is compatible with pMJY expressing the archaeal tyrosine pair. Since the expression of the *E. coli* amber suppressor tRNA^{Tyr} from a multicopy plasmid negatively affected the growth of the *E. coli* cells, the tRNA gene was expressed from the low-copy plasmid pApsYR. The gene encoding GST with an interrupting amber codon was carried on the third plasmid pRGexGST(Am84). Thus, it was necessary

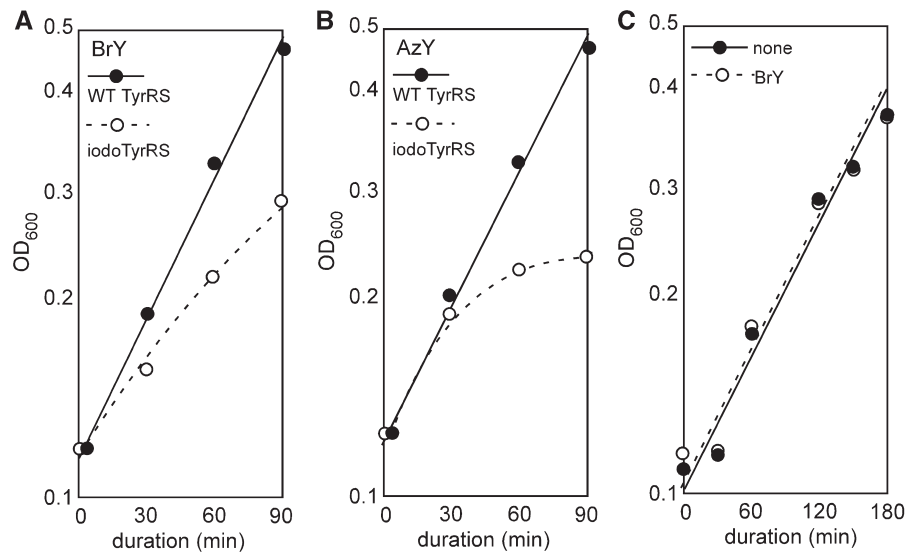


Figure 2. *E. coli* DH5 α cells, an *E. coli* K-12 strain, expressing the wild-type *E. coli* TyrRS from plasmid pEcYS (filled circles) and those expressing iodoTyrRS-*ec* from plasmid pEciYS (open circles) were grown in the presence of 3-bromo-L-tyrosine (0.5 mg/ml) (A) and 3-azido-L-tyrosine (0.3 mg/ml) (B). FT3 cells expressing iodoTyrRS-*ec* were grown in the absence (filled circles) and presence (open circles) of 3-bromo-L-tyrosine (C). The growth was monitored upon the addition of the non-natural amino acid to the growth media.

to maintain four different plasmids in FT3 cells for the site-specific incorporation of non-natural amino acids into proteins. The *gst* amber mutant gene produced the full-length product only when 3-iodo-L-tyrosine was supplemented in the growth media (Figure 3A, lanes 1 and 2), indicating that 3-iodo-L-tyrosine was incorporated into GST at the amber position, although the suppression efficiency was low.

The growth rate of FT3 was found to be just half that of its parent strain TOP10; the growth rates were determined as 0.7 and 1.4 h⁻¹ for FT3 and TOP10, respectively, in LB medium with no antibiotics. Due to this slow growth rate, the maintenance of four different plasmids, and the low yield of suppression products, FT3 cells were apparently not useful for yielding large amounts of proteins containing non-natural amino acids. We tried to circumvent this problem by further engineering, starting with another *E. coli* strain.

Replacement of the *E. coli* TyrRS-tRNA^{Tyr} pair by the *M. jannaschii* tyrosine pair in strain BL21(DE3)

The *E. coli* TyrRS and tRNA^{Tyr} genes were replaced by the *M. jannaschii* pair in strain BL21(DE3), an *E. coli* B strain. This strain carries the gene encoding the bacteriophage T7 RNA polymerase in its chromosome, for the high-level expressions of recombinant proteins. The order of gene disruptions was not the same as that with TOP10. In addition, the *E. coli* amber suppressor tRNA^{Tyr} gene was introduced into the chromosome of BL21(DE3), instead of on a plasmid, to reduce the number of required plasmids (Figure 1B). First, the *cat* gene was substituted for the *tyrU* gene, and was then replaced by the sequence encoding the *E. coli* suppressor tRNA^{Tyr}. This step effectively mutated *tyrU* to the suppressor tRNA^{Tyr} gene, thus creating the strain FB1 [BL21 λ DE3 *tyrU*]. This strain was transformed with pMJY, and

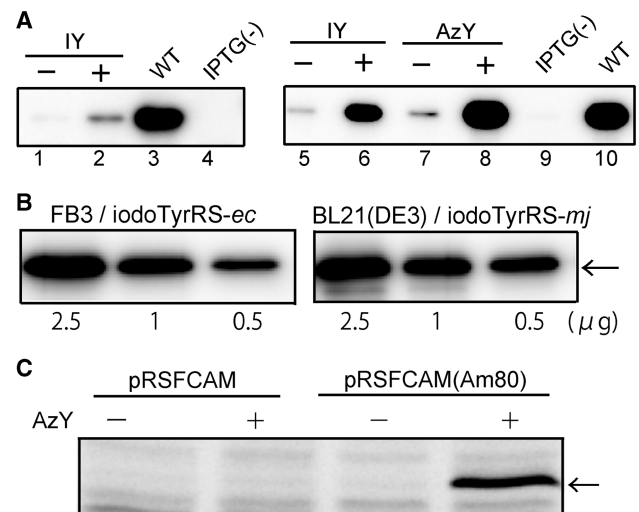


Figure 3. (A) Site-specific incorporation of 3-iodo-L-tyrosine (IY) and 3-azido-L-tyrosine (AzY) into GST in FT3 cells (lanes 1–4) and FB3 cells (lanes 5–10). The full-length GST was detected by western blotting. The growth media were supplemented with IY (lanes 2 and 6) and AzY (lane 8). The wild-type GST (WT) was expressed from pRGexGST (lane 3) or pRSFGST (lane 10). IPTG(-) means no induction of the *gst* expression (lanes 4 and 9). (B) The full-length GST product, indicated by the arrow, was detected by western blotting in extracts from the FB3 and BL21(DE3) cells. The applied quantity of total protein from the extract is indicated at the bottom of each lane. (C) Fluorescent labeling of rat calmodulin containing 3-azido-L-tyrosine. Extracts from the cells grown in the absence and presence of AzY were mixed with the fluorescein-triarylphosphine conjugate and then applied to SDS-PAGE. The band of calmodulin was indicated by the arrow.

the *tyrS* gene was then removed from the chromosome to create FB2 [BL21 λ DE3 *tyrU* Δ *tyrS* pMJY]. The *tyrT* and *tyrV* genes in FB2 were replaced with one copy of the suppressor tRNA^{Tyr} gene. This step created the strain

FB3 [BL21 λ DE3 *tyrU tyrV Δ tyrS Δ tyrT* pMJY], in which *tyrT* was removed and *tyrV* was mutated to another copy of the suppressor tRNA^{Tyr} gene, expressed from the *tyrT* promoter. Thus, the created FB3 strain expresses the archaeal TyrRS–tRNA^{Tyr} pair and the *E. coli* suppressor tRNA^{Tyr}, while it lacks *E. coli* TyrRS and wild-type tRNA^{Tyr}. The growth rates of the FB3 and BL21(DE3) strains were determined as 1.1 and 1.3 h⁻¹, respectively; the FB3 cells grow almost as fast as the parent strain, and much faster than FT3. The modifications to the BL21(DE3) chromosome had no serious effects on the viability of the cells.

To demonstrate the site-specific incorporation of 3-iodo-L-tyrosine into proteins in the strain FB3, iodoTyrRS-*ec* was expressed from pEcIYS, and the *gst* gene with the amber codon at position 25, *gst*(Am25), was expressed from a separate multicopy plasmid, pRSFGST(Am25), under the control of the T7 promoter. The FB3 strain bearing these plasmids produced the full-length GST only in the presence of 3-iodo-L-tyrosine, and its yield was significantly higher than that in FT3 cells (Figure 3A, lanes 5 and 6). We previously used the amber suppressor tRNA^{Tyr} and the 3-iodo-L-tyrosine-specific variant of TyrRS (iodoTyrRS-*mj*) from *M. jannaschii*, in order to site-specifically incorporate 3-iodo-L-tyrosine into proteins in *E. coli* cells (36). The yield of the full-length product of *gst*(Am25) in the presence of 3-iodo-L-tyrosine was then compared between FB3 cells expressing the iodoTyrRS-*ec*–*E. coli* suppressor tRNA^{Tyr} pair and the parent BL21(DE3) cells expressing the iodoTyrRS-*mj*–archaeal suppressor tRNA^{Tyr} pair. Comparable amounts of the full-length product were detected in the extracts from these cells (Figure 3B).

Finally, the full-length GST from FB3 cells was analyzed by mass spectrometry. This amber suppression product was subjected to a trypsin digestion, for comparison with the peptide data previously obtained for the wild-type GST (36). The wild-type protein generates a peptide (Tyr25 peptide) with an average mass corresponding to the theoretical value ($m/z = 1327.7$) for residues 23–33, NSYSPILGYWK (tyrosine at position 25 is underlined). This peptide was not detected from the amber suppression product from FB3 cells; instead, a peptide with an average mass (1453.6) was detected (Supplementary Figure S2). This value is larger than that of the Tyr25 peptide by a value (125.9) corresponding to the mass of an iodine atom, less the mass of a hydrogen atom, indicating the incorporation of 3-iodo-L-tyrosine at the amber position.

Genetic encoding of 3-azido-L-tyrosine in the engineered *E. coli* cells

A variant of yeast TyrRS, engineered to recognize 3-iodo-L-tyrosine, reportedly also recognizes 3-azido-L-tyrosine (37). Using this variant, 3-azido-L-tyrosine was site-specifically incorporated into proteins in *E. coli* cell-free translation, and was shown to be useful for site-specific protein modification (25). However, since this yeast TyrRS variant still recognizes L-tyrosine

efficiently, it has not been used for genetically encoding L-tyrosine derivatives in *E. coli* cells. We found that iodoTyrRS-*ec* also recognizes 3-azido-L-tyrosine efficiently; this amino acid supplemented in the growth media inhibits the growth of normal *E. coli* cells expressing iodoTyrRS-*ec*, as observed for 3-bromo-L-tyrosine (Figure 2B). Therefore, iodoTyrRS-*ec* was used to incorporate 3-azido-L-tyrosine into proteins in FB3 cells. The full-length product of *gst*(Am25) was obtained only in the presence of 3-azido-L-tyrosine, and its yield was comparable to that of the product with 3-iodo-L-tyrosine (Figure 3A, lanes 7 and 8). Thus, FB3 cells expressing iodoTyrRS-*ec* and the *E. coli* suppressor tRNA^{Tyr} produced the proteins site-specifically containing either 3-azido-L-tyrosine or 3-iodo-L-tyrosine, depending on which amino acid is supplemented in the growth media.

To demonstrate the site-specific protein modification, rat calmodulin containing 3-azido-L-tyrosine was produced in FB3 cells expressing the iodoTyrRS-*ec*–suppressor tRNA^{Tyr} pair. The full-length product was expressed from the amber mutant gene, carried on plasmid pRSFCAM(Am80), only in the presence of 3-azido-L-tyrosine, and was modified with a fluorescent probe by the Staudinger–Bertozzi reaction, while the wild-type protein expressed from plasmid pRSFCAM was not labeled (Figure 3C). *p*-Azido-L-phenylalanine and a L-lysine derivative with the azido group have already been genetically encoded in *E. coli* cells (20,38). Tyrosines tend to reside on the protein surface (39), and such tyrosines on the surface could be safely replaced with 3-azido-L-tyrosine for site-specific protein labeling.

Structural basis for 3-azido-L-tyrosine recognition by iodoTyrRS-*ec*

E. coli TyrRS was previously engineered into iodoTyrRS-*ec* by introducing a double substitution, Tyr37Val and Gln195Cys (Figure 4A) (21,28). To reveal the structural basis for the additional recognition of 3-azido-L-tyrosine by iodoTyrRS-*ec*, the crystal structure of a complex between these molecules was determined at 1.8-Å resolution (PDB ID: 2YXN). The statistics of the data set are listed in Table 1. The overall structure of the variant superposed well on the reported crystal structure of iodoTyrRS-*ec* complexed with 3-iodo-L-tyrosine (28). The azido group of 3-azido-L-tyrosine is accommodated in the space created by the double substitution at the amino-acid binding pocket (Figure 4B). Although no enzyme residues appear to make firm contacts or interactions with the azido group, the shape of the space snugly fits the substituent. By contrast, in the reported structure, iodoTyrRS-*ec* recognizes 3-iodo-L-tyrosine by making van der Waals contacts between the mutated residues, Val37 and Cys195, and the iodine atom (Figure 4B). The binding of L-tyrosine to iodoTyrRS-*ec*, on the other hand, is weakened by the Tyr37Val substitution, which removes the hydrogen bond between the phenolic hydroxyl groups of L-tyrosine and Tyr37 (28) (Figure 4A).

The *M. jannaschii* iodoTyrRS-*mj*–suppressor tRNA^{Tyr} pair does not incorporate 3-azido-L-tyrosine into proteins (data not shown), indicating that this archaeal TyrRS

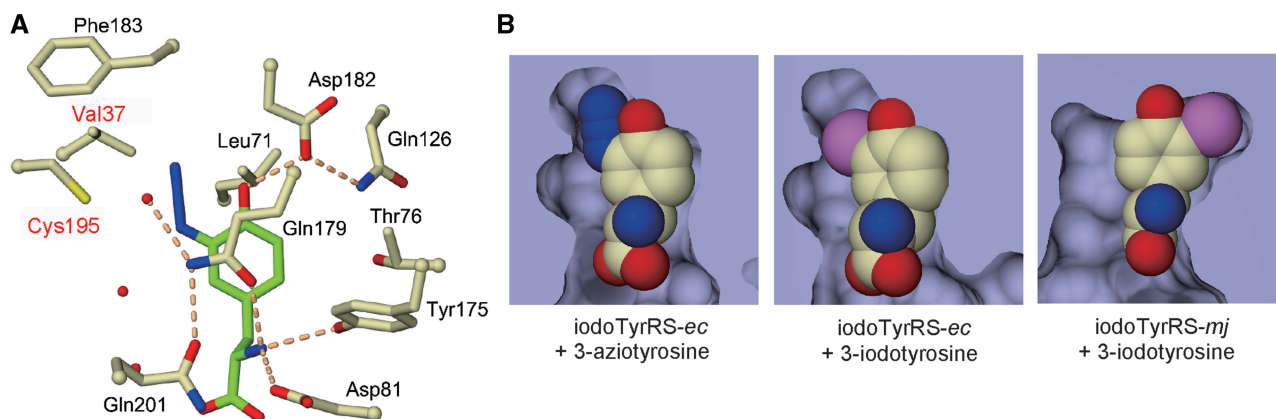


Figure 4. (A) The amino-acid binding pocket in the crystal structure of iodoTyrRS-*ec* bound with 3-azido-L-tyrosine. The mutated residues are denoted in red. The carbon atoms of the ligand are shown in green. The nitrogen, oxygen and sulfur atoms are in blue, red and yellow, respectively. The hydrogen bonds are shown as pink broken lines. (B) The surface structures of the binding pockets of iodoTyrRS-*ec* and iodoTyrRS-*mj* accommodating 3-azido-L-tyrosine or 3-iodo-L-tyrosine, as indicated below each panel. The enzyme surface and section are shown in light blue. The iodine atom is in purple.

Table 1. Data collection and refinement statistics

Data collection	
Space group	P3 ₁ 21
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	83.16, 83.16, 93.67
α , β , γ (°)	90, 90, 120
Wavelength (Å)	1.0000
Resolution (Å)	50–1.8
Measured reflections	382 492
Unique reflections	35 502
Redundancy	10.8
Completeness (%)	100.0
<i>I</i> / σ (<i>I</i>)	23.0
<i>R</i> _{sym} (%) ^a	11.3
Refinement statistics	
Resolution (Å)	39.26–1.8
Protein atoms	2536
Substrates (per subunit)	1
Water oxygens	295
Mean B value (Å ²)	21.3
<i>R</i> _{work} (%) ^b	19.9
<i>R</i> _{free} (%) ^c	24.7
R.m.s. deviations	
Bond length (Å)	0.011
Bond angle (°)	1.40

$$^a R_{\text{sym}} = \frac{\sum_{\text{hkl}} |I_{\text{avg}} - I_i|}{\sum_{\text{hkl}} I_i}$$

$$^b R_{\text{work}} = \frac{\sum_{\text{hkl}} |F_{\text{O}} - |F_{\text{C}}||}{\sum_{\text{hkl}} F_{\text{O}}}$$

$$^c R_{\text{free}} = R_{\text{work}} \text{ using 10\% of } F_{\text{O}} \text{ sequestered before refinement.}$$

variant does not recognize 3-azido-L-tyrosine. In the reported crystal structure of iodoTyrRS-*mj* complexed with 3-iodo-L-tyrosine (36), the space created by the mutations just fits the size and shape of the iodine atom, with five enzyme residues making van der Waals contacts with this atom (Figure 4B). The azido substituent would collide with Thr158, one of these five residues, in the binding pocket of iodoTyrRS-*mj*, because an azido group has a more extended shape and is longer than an iodine atom. This explains the inability of this variant to recognize 3-azido-L-tyrosine.

CONCLUSION

A variety of tyrosine or phenylalanine derivatives have been genetically encoded in mammalian cells and *E. coli* to facilitate protein science and technology. In mammalian cells, for example, photo-reactive cross-linking amino acid was used for analyzing protein–protein interactions (9), and the incorporated 4-azido-L-phenylalanine allowed the probing of conformational changes in a membrane protein (40). *E. coli* cells have been useful for producing large amounts of proteins containing non-natural amino acids. The incorporation of 4-azido-L-phenylalanine allowed the site-specific protein modification with polyethylene glycol (41), and iodine-atom-containing amino acids have been utilized in X-ray crystallography (36,42). For each of these non-natural amino acids, specific archaeal and bacterial TyrRS variants both needed to be developed, when the amino acid was used in the bacterial and eukaryotic cells. Our engineered *E. coli* strains promise to reduce the labor for variant development, because they can be engineered in the *E. coli* cells, and then used in both eukaryotic cells and bacteria. Our strategy could be applied to other aaRS species, and would facilitate genetic code expansion in living cells.

ACCESSION NUMBER

2YXN.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Nobuhiro Hayashi for kindly providing the rat calmodulin gene, Dr Takamitsu Hosoya and Dr Masaaki Suzuki for kindly providing the first lot of the triarylphosphine-fluorescein conjugate used in this study,

and Ms Azusa Ishii and Ms Tomoko Nakayama for clerical assistance.

FUNDING

Targeted Proteins Research Program (TPRP); RIKEN Structural Genomics/Proteomics Initiative (RSGI) in the National Project on Protein Structural and Functional Analyses, and in part, by a Grant-in-Aid for Scientific Research (B) 19380195 to K.S. from the Ministry of Education, Culture, Sports, Science and Technology of Japan; Special Postdoctoral Researchers Program at RIKEN to T.K. Funding for open access charge: RIKEN.

Conflict of interest statement. None declared.

REFERENCES

- Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science*, **292**, 498–500.
- Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K. *et al.* (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.*, **30**, 4692–4699.
- Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z. and Schultz, P.G. (2003) An expanded eukaryotic genetic code. *Science*, **301**, 964–967.
- Ye, S., Köhrer, C., Huber, T., Kazmi, M., Sachdev, P., Yan, E.C., Bhagat, A., RajBhandary, U.L. and Sakmar, T.P. (2008) Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.*, **283**, 1525–1533.
- Kwon, I. and Tirrell, D.A. (2007) Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J. Am. Chem. Soc.*, **129**, 10431–10437.
- Mukai, T., Wakiyama, M., Sakamoto, K. and Yokoyama, S. (2010) Genetic encoding of non-natural amino acids in *Drosophila melanogaster* Schneider 2 cells. *Protein Sci.*, 101002/pro.322.
- Wang, L., Xie, J. and Schultz, P.G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 225–249.
- Kowal, A.K., Köhrer, C. and RajBhandary, U.L. (2001) Twenty-first aminoacyl-tRNA synthetase—suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl Acad. Sci. USA*, **98**, 2268–2273.
- Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K. and Yokoyama, S. (2005) Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods*, **2**, 201–206.
- Liu, W., Brock, A., Chen, S., Chen, S. and Schultz, P.G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods*, **4**, 239–244.
- Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S. and Schultz, P.G. (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl Acad. Sci. USA*, **101**, 8882–8887.
- Furter, R. (1998) Expansion of the genetic code: site-directed *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.*, **7**, 419–426.
- Anderson, J.C. and Schultz, P.G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry*, **42**, 9598–9608.
- Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S. and Schultz, P.G. (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl Acad. Sci. USA*, **101**, 7566–7571.
- Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K. and Krzycki, J.A. (2004) Direct charging of tRNA(CUA) with pyrrolysine *in vitro* and *in vivo*. *Nature*, **431**, 333–335.
- Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K. and Yokoyama, S. (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.*, **371**, 818–822.
- Chen, P.R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B.H. and Schultz, P.G. (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed. Engl.*, **48**, 4052–4055.
- Polycarpo, C.R., Herring, S., Bérubé, A., Wood, J.L., Söll, D. and Ambrogelly, A. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.*, **580**, 6695–6700.
- Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2008) Genetically encoding *N*(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.*, **4**, 232–234.
- Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. and Yokoyama, S. (2008) Multi-step engineering of pyrrolysyl-tRNA synthetase to genetically encode *N*^ε-(*o*-azidobenzoyloxycarbonyl)lysine for site-specific protein modification. *Chem. Biol.*, **15**, 1187–1197.
- Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M., Harada, Y., Nakayama, H., Takio, K. *et al.* (2002) An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc. Natl Acad. Sci. USA*, **99**, 9715–9723.
- Hino, N., Hayashi, A., Sakamoto, K. and Yokoyama, S. (2007) Site-specific incorporation of non-natural amino acids into proteins in the mammalian cells with an expanded genetic code. *Nat. Protoc.*, **1**, 2957–2962.
- Sakamoto, K., Ishimaru, S., Kobayashi, T., Walker, J.R. and Yokoyama, S. (2004) The *Escherichia coli argU10*(Ts) phenotype is caused by a reduction in the cellular level of the *argU* tRNA for the rare codons AGA and AGG. *J. Bacteriol.*, **186**, 5899–5905.
- Chumpolkulwong, N., Sakamoto, K., Hayashi, A., Irahara, F., Shinya, N., Matsuda, N., Kiga, D., Urushibata, A., Shirouzu, M., Oki, K. *et al.* (2006) Translation of 'rare' codons in a cell-free protein synthesis system from *Escherichia coli*. *J. Struct. Funct. Genomics*, **7**, 31–36.
- Ohno, S., Matsui, M., Yokogawa, T., Nakamura, M., Hosoya, T., Hiramatsu, T., Suzuki, M., Hayashi, N. and Nishikawa, K. (2007) Site-selective post-translational modification of proteins using an unnatural amino acid, 3-azidotyrosine. *J. Biochem.*, **141**, 335–343.
- Murphy, K.C. (1998) Use of bacteriophage λ recombination function to promote gene replacement in *Escherichia coli*. *J. Bacteriol.*, **180**, 2063–2071.
- Muyrers, J.P.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555–1557.
- Kobayashi, T., Sakamoto, K., Takimura, T., Sekine, R., Vincent, K., Kamata, K., Nishimura, S. and Yokoyama, S. (2005) Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. *Proc. Natl Acad. Sci. USA*, **102**, 1366–1371.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In Carter, C.W. Jr and Sweet, R.M. (eds), *Methods in Enzymology, Macromolecular Crystallography, Part A*, 276. Academic Press, New York, pp. 307–326.
- Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**, 1022–1025.
- Schuettkopf, A.W. and van Aalten, D.M.F. (2004) PRODRG—a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D*, **60**, 1355–1363.
- Neidhardt, F.C. (1996) (ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. ASM Press, Washington DC.
- Wang, L., Magliery, T.J., Liu, D.R. and Schultz, P.G. (2000) A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for

- the *in vivo* incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.*, **122**, 5010–5011.
34. Clewell, D.B. and Helinski, D.R. (1970) Existence of the colicinogenic factor-sex factor ColIb -P9 as a supercoiled circular DNA-protein relaxation complex. *Biochem. Biophys. Res. Commun.*, **41**, 150–156.
 35. Young, T.S., Ahmad, I., Yin, J.A. and Schultz, P.G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.*, **395**, 361–374.
 36. Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., Ohtake, K., Kobayashi, T., Kuramitsu, S., Shirouzu, M. *et al.* (2009) Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure*, **17**, 335–344.
 37. Ohno, S., Yokogawa, T. and Nisihikawa, K. (2001) Changing the amino acid specificity of yeast tyrosyl-tRNA synthetase by genetic engineering. *J. Biochem.*, **130**, 417–423.
 38. Tsao, M.-L., Tian, F. and Schultz, P.G. (2005) Selective staudinger modification of proteins containing *p*-azidephenylalanine. *ChemBioChem.*, **6**, 2147–2149.
 39. Bordo, D. and Argos, P. (1991) Suggestions for “safe” residue substitutions in site-directed mutagenesis. *J. Mol. Biol.*, **217**, 721–729.
 40. Ye, S., Huber, T., Vogel, R. and Sakmar, T.P. (2009) FTIR analysis of GPCR activation using azido probes. *Nat. Chem. Biol.*, **5**, 397–399.
 41. Deiters, A., Cropp, T.A., Summerer, D., Mukherji, M. and Schultz, P.G. (2004) Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.*, **14**, 5743–5745.
 42. Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. and Schultz, P.G. (2004) The site-specific incorporation of *p*-iodo-l-phenylalanine into proteins for structure determination. *Nat. Biotechnol.*, **22**, 1297–1301.