

Article

Crystal Structure of Pyrrolysyl-tRNA Synthetase from a Methanogenic Archaeon ISO4-G1 and its Structure-Based Engineering for Highly-Productive Cell-Free Genetic Code Expansion with Non-Canonical Amino Acids

Tatsuo Yanagisawa ^{1,2,*}, Eiko Seki ², Hiroaki Tanabe ², Yoshifumi Fujii ², Kensaku Sakamoto ¹ and Shigeyuki Yokoyama ^{2,*}

¹ Affiliation 1; RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

² Affiliation 2; RIKEN Cluster for Science, Technology and Innovation Hub, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

* Correspondence: yokoyama@riken.jp; tatsuo.yanagisawa@riken.jp Tel.: +81-45-503-9196

Abstract: Pairs of pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl} from *Methanosarcina mazei* and *Methanosarcina barkeri* are widely used for site-specific incorporations of non-canonical amino acids into proteins (genetic code expansion). Previously, we achieved full productivity of cell-free protein synthesis for bulky non-canonical amino acids, including N^ε-((((E)-cyclooct-2-en-1-yl)oxy)carbonyl)-L-lysine (TCO*Lys), by using *Methanomethylophilus alvus* PylRS with structure-based mutations in and around the amino acid binding pocket (first-layer and second-layer mutations, respectively). Recently, the PylRS•tRNA^{Pyl} pair from a methanogenic archaeon ISO4-G1 was used for genetic code expansion. In the present study, we determined the crystal structure of the methanogenic archaeon ISO4-G1 PylRS (ISO4-G1 PylRS) and compared it with those of structure-known PylRSs. Based on the ISO4-G1 PylRS structure, we attempted the site-specific incorporation of N^ε-(p-ethynylbenzyloxycarbonyl)-L-lysine (pEtZLys) into proteins, but it was much less efficient than that of TCO*Lys with *M. alvus* PylRS mutants. Thus, the first-layer mutations (Y125A and M128L) of ISO4-G1 PylRS, with no additional second-layer mutations, increased the protein productivity with pEtZLys up to 57±8% of that with TCO*Lys, at high enzyme concentrations in the cell-free protein synthesis.

Keywords: non-canonical amino acids; genetic code expansion; crystal structure; tRNA; cell-free protein synthesis

1. Introduction

Expanding the genetic code with non-canonical amino acids is useful for developing novel structures and functions of proteins [reviewed in 1,2]. Site-specific incorporation of non-canonical amino acids into proteins in response to specified (*e.g.*, UAG) codons has been achieved by pairs of an engineered aminoacyl-tRNA synthetase (aaRS) and tRNA, including the pairs of pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl}(CUA) [2,3,4,5,6,7,8,9,10]. The PylRS•tRNA^{Pyl} pair was first found in methanogenic archaea, including *Methanosarcina barkeri* [11,12], and in bacteria, including *Desulfitobacterium hafniense* [13,14]. The PylRS•tRNA^{Pyl} pairs from *M. barkeri* and *Methanosarcina mazei* have been extensively studied [reviewed in 10,15,16,17,18,19,20,21]. Recently, by using the pairs of PylRS and tRNA^{Pyl} from *Methanomethylophilus alvus*, the site-specific incorporations of non-canonical amino acids into proteins have been achieved in the methanogenic archaeon ISO4-G1, the methanogenic archaeon ISO4-H5, *Methanomassiliicoccus intestinalis*, and *Methanomassiliicoccus luminyensis* [22,23,24,25,26,27,28,29,30].

The PylRS•tRNA^{Pyl} pairs are useful for non-canonical amino acid incorporation because of their “orthogonality” (non-reactivity) to the 20 canonical aaRS•tRNA pairs in many organisms [15,16,17,18,19]. PylRS and its mutants showed broad specificity for substrate amino acids, and by using the PylRS•tRNA^{Pyl} pairs, site-specific incorporations of more than 200 non-canonical amino acids into proteins have been achieved in bacteria including *Escherichia coli*, and eukaryotes including *Saccharomyces cerevisiae*, mammalian cells, and multicellular organisms [reviewed in 10,15,16,17,18,19,20,21,31], and by cell-free protein synthesis based on an *E. coli* cell extract [32,33,34,35,36,37]. Cell-free protein synthesis systems, which are novel protein expression platforms, are particularly suitable for synthesizing cell-toxic proteins and transmembrane proteins that are difficult to synthesize in cellular systems, and can efficiently introduce non-canonical amino acids into such proteins for pharmaceutical research.

The efficiencies of non-canonical amino acid incorporations into proteins at the UAG codon are usually lower than that of the standard protein synthesis with a canonical amino acid at the corresponding position. The incorporation efficiency of a non-canonical amino acid at multiple sites in a protein is much lower than that of the single-site incorporation. Therefore, by using *E. coli* strains lacking translation termination factor 1 (RF-1) to achieve the complete reassignment of the UAG codon [32,38,39,40,41,42,43], we have increased the incorporation efficiencies of non-canonical amino acids to the maximum level, which was previously designated as the full productivity of the expanded genetic code [26].

M. mazei PylRS (MmPylRS) and *M. barkeri* PylRS (MbPylRS) consist of the N- and C-terminal domains (PylRSn and PylRSc, respectively). Another group of PylRSs from bacteria, including *D. hafniense*, is composed of two separate gene products (PylSn and PylSc), which are homologous to PylRSn and PylRSc, respectively [12,13,14,44]. The PylRSc protein exhibited higher solubility than the full-length PylRS protein and was easily crystallized [45]. However, the PylRSc protein retained insufficient tRNA binding and aminoacylation activities [46,47]. Consequently, both the N- and C-terminal domains of PylRS (*i.e.*, full-length PylRS) have been regarded as essential components for the efficient incorporation of non-canonical amino acids into proteins. Notably, recently discovered methanogenic archaea, including *M. alvus*, *M. intestinalis*, *M. luminyensis*, ISO4-G1, and ISO4-H5, have PylSc homologs, but lack the genes encoding PylSn homologs in their genomes [22,48]. The high solubility of the PylSc-type *M. alvus* PylRS makes it suitable for crystallographic analysis and cell-free protein synthesis for non-canonical amino acid incorporation [26].

Crystal structures of PylRSs from *M. mazei* and *D. hafniense* and their complexes have been extensively investigated [13,14,44,47,49]. The catalytic fragment (residues 185–454) of *M. mazei* PylRS (MmPylRSc) has been crystallized [45], and the structures of MmPylRSc and its mutants in complex with numerous substrate amino acids, aminoacyladenylates, and ATP and its analogs have been determined [8,45,47,49,50,51,52,53,54,55,56,57,58,59,60]. Based on the structural information and random screening, we obtained MmPylRS with the Y306A and Y384F mutations [MmPylRS(Y306A/Y384F)] [8]. The pair of MmPylRS(Y306A/Y384F) and tRNA^{Pyl} is one of the useful aaRS•tRNA pairs for cellular and cell-free genetic code expansion with bulky non-natural lysine derivatives, including N^ε-benzyloxycarbonyl-L-lysine (ZLys), N^ε-(*o*-azidobenzoyloxycarbonyl)-L-lysine (*o*AzZLys), and N^ε-(*m*-azidobenzoyloxycarbonyl)-L-lysine (*m*AzZLys). A variety of applications using the pair have been developed [8,26,61,62]. The crystal structures of the catalytic fragment of MmPylRS(Y306A/Y384F) [MmPylRSc(Y306A/Y384F)] in complex with 14 bulky non-natural lysine derivatives, including ZLys, *m*AzZLys, and N^ε-(((*E*)-cyclooct-2-en-1-yl)oxy)carbonyl-L-lysine (TCO*ZLys), revealed the structural bases for their amino acid binding modes [58].

We recently determined the crystal structure of *M. alvus* PylRS (MaPylRS) [26]. The MaPylRS with the Y126A (corresponding to the Y306A mutation in *M. mazei* PylRS) and M129L mutations efficiently incorporated bulky non-canonical amino acids, including ZLys and *m*AzZLys, into proteins in *E. coli* cells [24,26]. Tyr126 and Met129 are the “first-

layer residues", located within ~5 Å from (and direct contact with) substrate amino acids, while His227 and Tyr228 are the "second-layer residues", which are located within ~9 Å from (and do not directly contact) substrates but might affect the first-layer residues. This resembles the concept of first-shell and second-shell residues, respectively, as reported by others [63]. The *Ma*PylRS with both the first-layer (Y126A and M129L) and second-layer (H227I and Y228P) mutations enhanced the protein productivities drastically to the maximum level (*i.e.*, full productivities) for ZLys, *mAz*ZLys, and the more difficult amino acid TCO*Lys by *E. coli* cell-free protein synthesis, as compared with those obtained with the *Mm*PylRS(Y306A/Y384F) and *Ma*PylRS(Y126A/M129L) pairs [26]. Recently, the crystal structures of *Ma*PylRS(N166S/V168C/W239T) and *Ma*PylRS(N166A/C168G/W239C) mutants in complex with the fluorescent non-canonical amino acid acrydonylalanine and ATP (or its non-hydrolyzable analog, adenylyl imidodiphosphate (AMPPNP)) were determined, and the conformational changes of the residues including His227 upon non-canonical amino acid-binding were discussed [64]. It should be emphasized that *Ma*PylRS is much more useful than *Mm*PylRS for cell-free genetic code expansion with bulky non-canonical amino acids [26]. However, some bulky non-canonical amino acids still cannot be incorporated by using the *Ma*PylRS mutant system. To synthesize proteins in which these bulky non-canonical amino acids are efficiently incorporated, we should compare the structures of PylRSs from archaea, and create active mutants for inefficient bulky non-canonical amino acids based on these structures.

The PylSc-type PylRS from the methanogenic archaeon ISO4-G1 (ISO4-G1 PylRS) is highly similar to *Ma*PylRS with 64% sequence identity, while ISO4-G1 PylRS and *Ma*PylRS both share around 38% sequence identity with *M. mazei* PylRS. However, the pair of PylRS and tRNA^{Pyl} from the methanogenic archaeon ISO4-G1 (ISO4-G1 PylRS•tRNA^{Pyl}) is orthogonal to the *M. mazei* pair, but the *M. alvus* pair is not [23,27]. The mechanism by which ISO4-G1 PylRS recognizes tRNA^{Pyl} and a variety of non-canonical amino acids in its active site must be elucidated for understanding its substrate specificity and orthogonality, and for achieving excellent genetic code expansion systems. Recently, the crystal structure of ISO4-G1 PylRS with the L124A, Y125L, V167A, Y204W, and A221S mutations for the incorporation of the non-canonical amino acid *m*-cyanopyridylalanine was determined in the apo form [29].

In the present study, we solved the crystal structure of the wild-type ISO4-G1 PylRS apo form, and compared it with that of the multiple mutant of ISO4-G1 PylRS [29]. In the structure of the wild-type ISO4-G1 PylRS, His225 appears to be located in a different position from that in the multiple mutant ISO4-G1 PylRS structure. In this context, in *Ma*PylRS, the conformational changes of the corresponding His227 residue are considered to be important for the PylRS activity. Therefore, the structural changes of the conserved His residues may be common features between the two PylRSs, and might be the driving force for the movement of the specific hairpin (the β5-β6 hairpin, which will be described later) and thus crucial for the PylRS activity. Furthermore, in the present study, the ISO4-G1 PylRS mutants engineered based on the ISO4-G1 PylRS structures have been applied for the milligram-scale preparation of proteins containing useful non-canonical amino acids, including TCO*Lys and *N*^ε-(*p*-ethynylbenzyloxycarbonyl)-L-lysine (*pEt*ZLys), in the cell-free protein synthesis system. These rationally engineered ISO4-G1 PylRS mutants will be more useful than ever before for genetic encode expansion technologies..

2. Results

2.1. Overall structure of ISO4-G1 PylRS

The genome of the methanogenic archaeon ISO4-G1 encodes an ISO4-G1 PylRS protein consisting of 273 amino acids [48], which is quite similar to *Ma*PylRS [22,48](Fig. S1). ISO4-G1 PylRS was expressed very well as a soluble protein in *Escherichia coli* cells. The yield of the ISO4-G1 PylRS protein was over 100 mg per liter *E. coli* culture (Fig. S2), and the protein could be concentrated without aggregation to more than 20 mg/mL, which is comparable to that of *M. alvus* PylRS. For crystallographic analysis, ISO4-G1 PylRS was

purified to homogeneity by three column chromatography steps (Materials and Methods). The crystallization of the ISO4-G1 PylRS protein was successful when using PEG3350 as the precipitant, and the crystal structure of the ISO4-G1 PylRS apo form has been determined at 2.78-Å resolution (Fig. 1, Materials and Methods). The structure of the ISO4-G1 PylRS protein is shown in Fig. 1. The asymmetric unit contains ten molecules of PylRS (five PylRS dimers, A/B, C/G, D/F, E/H, J/I). The final model shows good geometry and all residues are within the allowed regions of the Ramachandran plot, as evaluated by Procheck [65] and Molprobity [66] (Table S1).

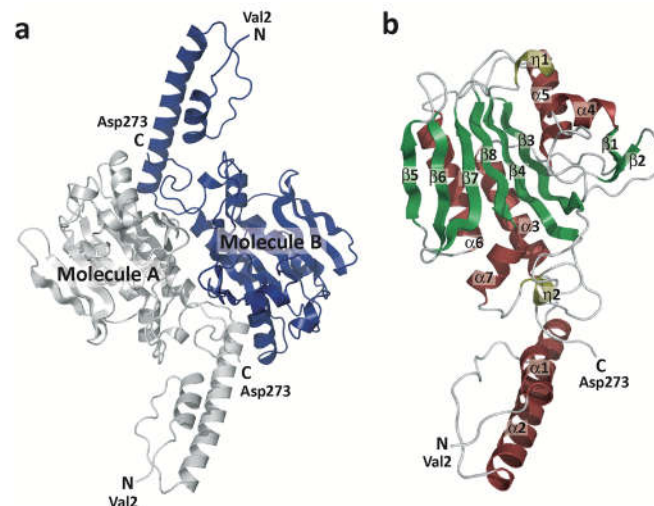


Figure 1. Overview of the ISO4-G1 PylRS structure. (a) The ISO4-G1 PylRS dimer. The protomers of the dimer are colored gray-white and blue. The crystallographic 2-fold axis is perpendicular to the paper. (b) The α -helices, 3_{10} helices, and β -sheets are colored wine red, olive, and green, respectively.

2.2. Structure-based sequence comparison of ISO4-G1 PylRS with other PylRSs

In the ISO4-G1 PylRS structure, the residues from Met1 to Pro58 contain two α -helices ($\alpha 1$ and $\alpha 2$) (Figs. 1b and S1). The following residues Asn60–Asp273 constitute the catalytic domain, with the characteristic topology of the class-II aaRSs: an extended six-strand, anti-parallel β -sheet ($\beta 3$, $\beta 4$, $\beta 5$, $\beta 6$, $\beta 7$, $\beta 8$) surrounded by α -helices. The sequence motifs (motifs 1, 2, and 3), which are conserved in class-II aaRSs [67,68,69], correspond to the residues Gly78–Val88, Cys147–Leu164, and Ala238–Lys251, respectively, in ISO4-G1 PylRS (Figs. 1b and S1). The structures of the ordering loop (residues Ile97–Gln106) and the motif-2 loop (residues Lys150–Glu159) are quite similar in the ten PylRS molecules, while the $\beta 5$ – $\beta 6$ hairpin (residues Thr195–Ile212) adopts open and closed conformations (Figs. 1 and S1). The $\beta 5$ – $\beta 6$ hairpin in ISO4-G1 PylRS corresponds to the $\beta 5$ – $\beta 6$ hairpin in *MaPylRS* and the $\beta 7$ – $\beta 8$ hairpin in *MmPylRS*, which randomly adopt open and closed conformations as described previously [26,47].

2.3. Structural comparison of ISO4-G1 PylRS with *MaPylRS*, *DhPylSc*, and *MmPylRS*

The structure of ISO4-G1 PylRS is highly homologous with those of *MaPylRS*, *DhPylSc*, and *MmPylRS* (Fig. 2). A DALI search [<http://www.embl-ebi.ac.uk/dali>] revealed that the ISO4-G1 PylRS structure (molecule B) resembles those of the PylRS proteins (PDBs: 6JP2, 6EZD, 7R6O, 2ZNI, 2ZNJ, 3DSQ, 2E3C, 4CH3, 2ZIN, and 2ZIM), with Z-scores of 31.4–33.7, 30.0–30.6, 33.0–33.3, 30.5–30.7, 29.6–30.9, 30.2, 28.9, 28.6, 28.4, and 27.6, respectively. Residues Asp61–Asp273 of ISO4-G1 PylRS constitute the catalytic core and superimposed well on *MaPylRS*, *DhPylSc* (residues Ala72–Asn286), and *MmPylRS* (residues Tyr242–Asn453), but the two α -helices ($\alpha 1$ and $\alpha 2$, residues Met1–Ala56) in the N-terminal 59 residues of ISO4-G1 PylRS are slightly different from those of the other PylRSs. The superimposition of the ISO4-G1 PylRS structure on those of *MaPylRS* (PDB:

6JP2), *Mm*PylRSc (PDB: 2ZIM), and *Dh*PylSc•tRNA^{Pyl} (PDB: 2ZNI) revealed that the N-terminal two α -helices of ISO4-G1 PylRS and *Mm*PylRSc cause steric hindrance with tRNA^{Pyl}, in contrast to those of *Ma*PylRS and *Dh*PylSc (Figs. 2 and S3). Accordingly, the two α -helices of PylRSs might undergo conformational changes upon tRNA^{Pyl} binding.

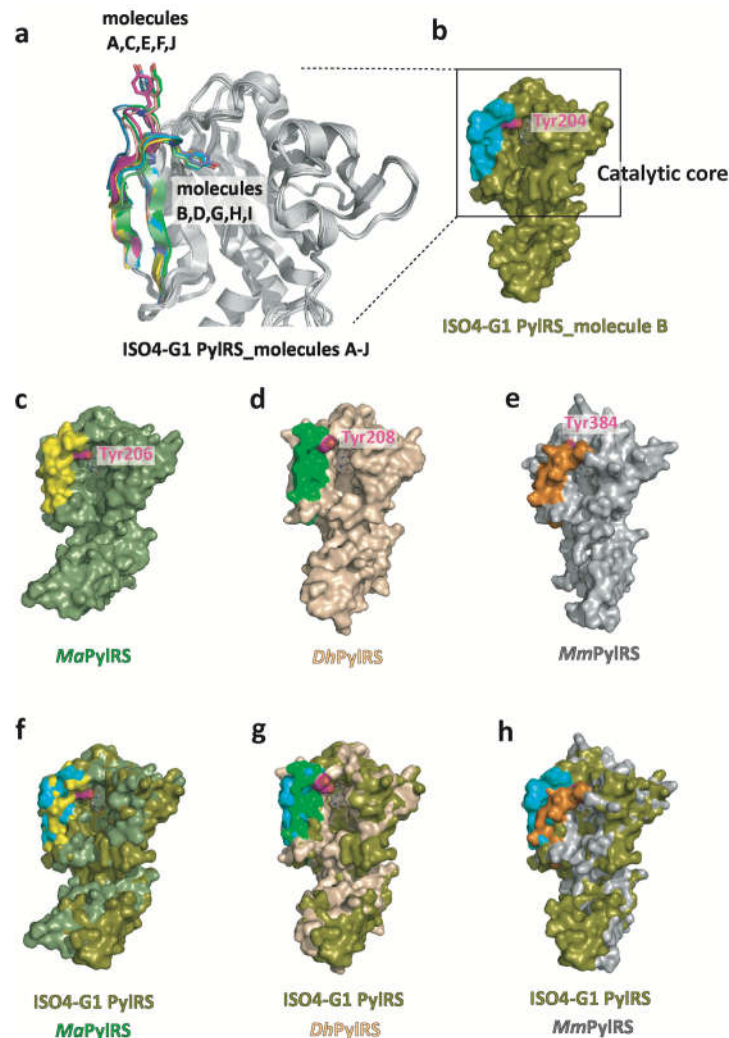


Figure 2. Structural comparisons of ISO4-G1 PylRS with *Ma*PylRS, *Dh*PylSc, and *Mm*PylRSc. (a) The β 5- β 6 hairpins in the ISO4-G1 PylRS molecules A, B, C, D, E, F, G, H, I, J, and H are colored green, cyan, magenta, yellow, vermillion, white, lavender, orange, light green, and turquoise blue, respectively. Tyr204 is shown as stick models. (b-e) Surface models of ISO4-G1 PylRS, *Ma*PylRS (PDB: 6JP2), *Dh*PylSc (PDB: 2ZNI), and *Mm*PylRSc (PDB: 2ZIM). (f-h) Superimpositions of the crystal structures of ISO4-G1 PylRS, *Ma*PylRS (PDB: 6JP2), *Dh*PylSc•tRNA^{Pyl} complex (PDB: 2ZNI), and the pyrrolysyladenylate (Pyl-AMP)-bound *Mm*PylRSc (PDB: 2ZIM), represented by surface models. The β 5- β 6 hairpins in ISO4-G1 PylRS and *Ma*PylRS and the β 7- β 8 hairpins in *Mm*PylRSc are colored cyan, yellow, green, and orange, respectively. The catalytic core structures of ISO4-G1 PylRS, *Ma*PylRS, *Dh*PylSc, and *Mm*PylRSc superimposed well.

2.4. Structural comparison of the amino acid binding residues of ISO4-G1 PylRS with those of *Ma*PylRS and *Mm*PylRSc

Based on the structure-based sequence alignments of PylRSs (Fig. S1), Ala121, Leu124, Tyr125, Asn165, Leu227, and Trp237 in the amino acid binding pocket of ISO4-G1 PylRS are conserved among *Ma*PylRS (Ala122, Leu125, Tyr126, Asn166, Leu229, and Trp239, respectively) and *Mm*PylRSc (Ala302, Leu305, Tyr306, Asn346, Leu407, and Trp417, respectively), whereas the counterparts of Met128, Val167, and Ala221 in ISO4-G1 PylRS correspond to Leu309, Cys348, and Val401, respectively, in *Mm*PylRSc. The present crystallographic analysis revealed the structural differences between ISO4-G1 PylRS,

*Ma*PylRS, and *Mm*PylRSc (PDB: 2ZIM) (Fig. 3). The ISO4-G1 PylRS Met128 and Val167 residues, which are respectively conserved as Met129 and Val168 in *Ma*PylRS, are bulkier than the corresponding Leu309 and Cys348 residues in *Mm*PylRSc, respectively. Therefore, the internal volumes of the amino acid binding pockets of ISO4-G1 PylRS and *Ma*PylRS are smaller than that of *Mm*PylRSc.

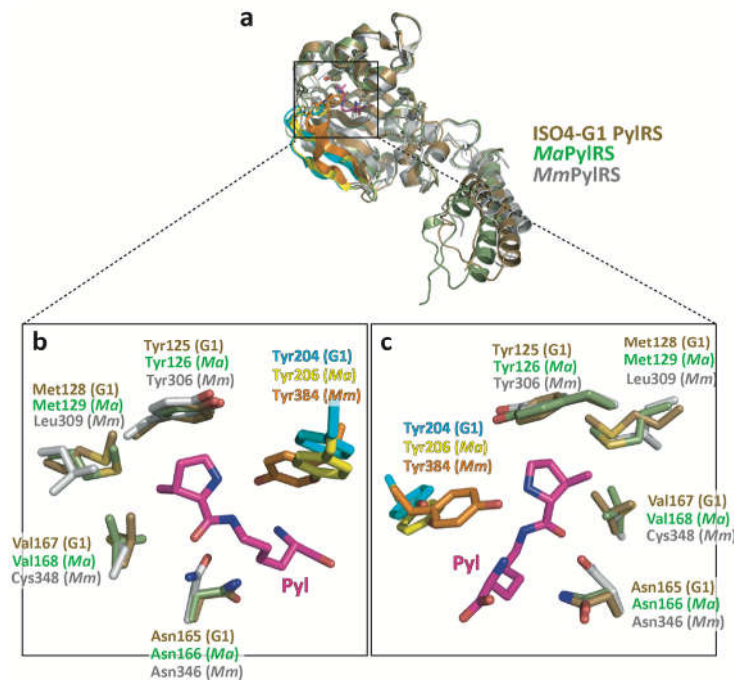


Figure 3. Comparison of the amino acid binding pocket of ISO4-G1 PylRS with those of *Ma*PylRS and *Mm*PylRSc. (a) Ribbon models of ISO4-G1 PylRS molecule B (brown) superimposed on the *Ma*PylRS (light green) and *Mm*PylRSc (gray) monomers. (b, c) Close-up views of different angles of the ISO4-G1 PylRS active site. The active site residues in the ISO4-G1 PylRS apo form (brown) are superimposed on those of *Ma*PylRS (light green) and Pyl-AMP-bound *Mm*PylRSc (gray). Along with pyrrolysine (Pyl), the active site residues in ISO4-G1 PylRS (Tyr125, Met128, Asn164, Val166, and Tyr204), *Ma*PylRS (Tyr126, Met129, Asn166, Val168, and Tyr206), and *Mm*PylRSc (Tyr306, Leu309, Asn346, and Cys348) are represented as stick models. The β 5- β 6 and Tyr204 in ISO4-G1 PylRS, the β 5- β 6 and Tyr206 in *Ma*PylRS, and the β 7- β 8 and Tyr384 in *Mm*PylRSc are highlighted in cyan, yellow, and orange, respectively.

2.5. Structural changes of Tyr204 and His225 in ISO4-G1 PylRS and comparison with those of *Mm*PylRS and *Ma*PylRS

Tyr204, at the tip of the β 5- β 6 hairpin in ISO4-G1 PylRS molecules A, C, E, F, and J, is located far from the amino acid binding pocket (the open conformation), while in ISO4-G1 PylRS molecules B, D, G, H, and I, Tyr204 is located inside the amino acid binding pocket (the closed conformation)(Figs. 2, 4, 5, and S4).

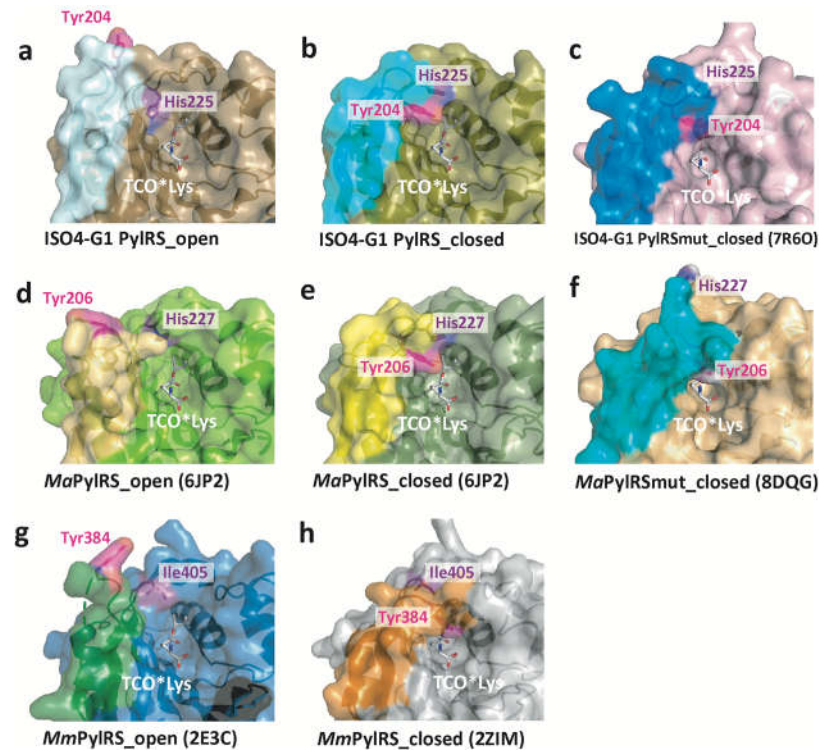


Figure 4. Open and closed conformations of the $\beta 5$ - $\beta 6$ (ISO4-G1 PylRSs), $\beta 5$ - $\beta 6$ (*M. alvus* PylRSs), and $\beta 7$ - $\beta 8$ (*M. mazei* PylRSs) hairpins. Translucent surface models of the active-site pockets in the ISO4-G1 PylRS apo form (brown, a; gold, b), the ISO4-G1 PylRS mutant for cyanopyridylalanine in the apo form (light pink, c) (PDB: 7R6O), the MaPylRS apo form (yellow green, d; grass green, e) (PDB: 6JP2), the acrydonylalanine and AMPPNP-bound MaPylRS mutant (light orange, f) (PDB: 8DQG), the *M. mazei* PylRS apo form (blue, g) (PDB: 2E3C), and the Pyl-AMP-bound MmPylRS (gray, h) (PDB: 2ZIM) are shown. The $\beta 5$ - $\beta 6$ (ISO4-G1 PylRSs), $\beta 5$ - $\beta 6$ (*M. alvus* PylRSs), and $\beta 7$ - $\beta 8$ (*M. mazei* PylRSs) hairpins are colored differently. Tyr204, Tyr206, and Tyr384 are highlighted in magenta. The bound TCO*Lys in the PylRS active site is shown as a stick model (PDB: AAO).

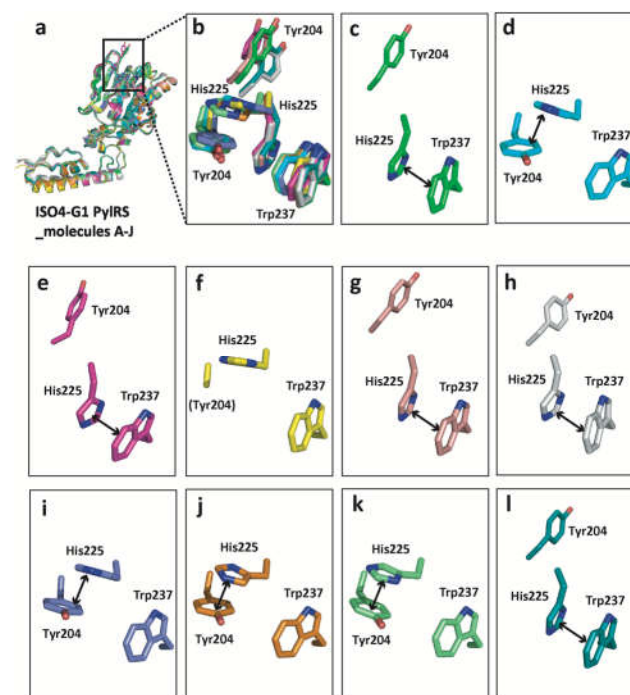


Figure 5. His225 undergoes drastic conformational changes in accordance with the $\beta 5$ - $\beta 6$ hairpin. Superimposition of the ISO4-G1 PylRS molecules A to J (a). The superimposed Tyr204, His225, and Trp237 residues are shown as stick models (b). The Tyr204, His225, and Trp237 residues in the ten

ISO4-G1 PylRS molecules are each colored green (c, molecule A), cyan (d, molecule B), magenta (e, molecule C), yellow (f, molecule D), vermillion (g, molecule E), white (h, molecule F), lavender (i, molecule G), orange (j, molecule H), light green (k, molecule I), and turquoise blue (l, molecule J).

Therefore, the β 5- β 6 hairpin undergoes random conformational changes. The Tyr204 side-chain in molecule D is disordered and thus might be in an intermediate form (Fig. 5f). The β 5- β 6 hairpin in ISO4-G1 PylRS is similar to that of *Ma*PylRS, while a remarkable difference exists between the β 5- β 6 hairpin in ISO4-G1 PylRS and the β 7- β 8 hairpin in *Mm*PylRSc. As described previously, the β 7- β 8 hairpin in *Mm*PylRSc is very flexible and adopts multiple conformations regardless of the substrate binding [47](Figs. 2 and 4). In the *Mm*PylRSc structure (PDB: 2ZIM), the β 7- β 8 hairpin is bent in the middle, and the tip half of the β -hairpin is elevated. Tyr384, at the tip of the bent β 7- β 8 hairpin, is buried deeply within the active site [47,49]. While the β 5- β 6 hairpin in ISO4-G1 PylRS still assumes bent conformations, Tyr204 of ISO4-G1 PylRS is not as deeply accommodated within the active site as compared to Tyr384 of *Mm*PylRSc (Figs. 2, 3, 4a, 4b, and 4h). In the crystal structure of the ISO4-G1 PylRS mutant for cyanopyridylalanine [29], Trp204, which is substituted for the strictly conserved Tyr residue in the PylRS family, at the tip of the bent β 5- β 6 hairpin is only slightly inserted into the active site as compared to Tyr204 of the ISO4-G1 PylRS apo form (Figs. 4c and S5a). In contrast, in the crystal structures of the acrydonylalanine (and ATP/AMPPNP)-bound *Ma*PylRS mutants [64], Tyr206 at the tip of the bent β 5- β 6 hairpin seems to penetrate more deeply within the active site than Tyr206 in the apo form of *Ma*PylRS (Figs. 4f and S5b). The Tyr204/Trp204 residues of the ISO4-G1 PylRSs in the apo form are shallowly inserted within their active sites, as compared to Tyr206 of the *Ma*PylRS mutant. Accordingly, the two structures of ISO4-G1 PylRS are ligand-free forms, in which Tyr204/Trp204 adopt open and partially closed conformations (Figs. 4a, 4b, 4c, and S5a), whereas the *Ma*PylRS mutant structure represents the amino acid substrate and ATP/AMPPNP-bound form, and Tyr206 adopts the completely closed conformation (Figs. 4f, 4h, S5b, and S5c). Notably, Tyr206 of the AMPPNP-bound form of the *Ma*PylRS mutant was completely disordered (Fig. S5b). Consequently, in the case of *Ma*PylRS, the bound amino acid substrate may induce Tyr206 at the tip of the β 5- β 6 hairpin to adopt the closed conformation.

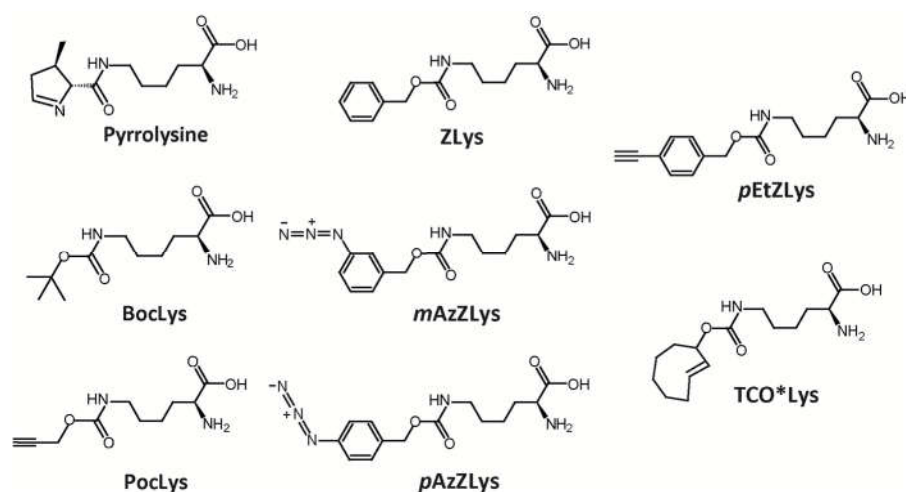
Interestingly, the ISO4-G1 PylRS structure revealed that the His225 residue (corresponding to His227 in *Ma*PylRS) undergoes drastic conformational changes in accordance with the location of the Tyr204 residue (corresponding to Tyr206 in *Ma*PylRS) in the β 5- β 6 hairpin (Figs. 2, 4, and 5). On the one hand, when Tyr204 is far from the amino acid binding pocket (the open conformation), a π - π stacking interaction is observed between the imidazole ring of His225 and the indole ring of Trp237. On the other hand, when Tyr204 is located inside the amino acid binding pocket (the closed conformation), the imidazole ring of His225 shifts and is stabilized by a π - π stacking interaction with the aromatic ring of Tyr204. This conformational change is not observed in the corresponding His227 residue of *Ma*PylRS, according to the structure of the *Ma*PylRS apo form (Figs. 4d, 4e, 5, and S5b). However, the recently determined structure of the *Ma*PylRS mutant in complex with the non-canonical amino acid acridonylalanine (and AMPPNP) revealed the conformational changes of residues 224–230, and the movement of His227 away from the active site upon acridonylalanine-binding (Figs. 4f and S5b)[64]. In contrast, no conformational changes of the corresponding Ile405 residue in *Mm*PylRS are observed (Figs. 4g, 4h, and S5c). Accordingly, the structural changes of His225/His227 share common features with *Ma*PylRS and ISO4-G1 PylRS, but not with Ile405 of *Mm*PylRS.

2.8. Structure-based engineering of the first-layer residues in ISO4-G1 PylRS for site-specific incorporation of bulky lysine derivatives into proteins by cell-free protein synthesis

Previously, we developed a system for genetic code expansion with bulky ZLys derivatives by using the *Mm*PylRS(Y306A/Y384F)•tRNA^{Pyl} pair [8,58,62] and the *Ma*PylRS(Y126A/M129L)•tRNA^{Pyl} pair [24,26]. To examine whether the ISO4-G1 PylRS•tRNA^{Pyl} pair is useful for genetic code expansion, we rationally engineered two

ISO4-G1 PylRS mutants from the PylRS structures. In the previous study, the *Ma*PylRS(Y126A/M129L)•tRNA^{Pyl} and *Ma*PylRS(Y126A/M129A)•tRNA^{Pyl} pairs successfully facilitated the site-specific incorporation of TCO*Lys and *mAz*ZLys into proteins in an *E. coli* cell-free protein synthesis system [26]. Therefore, the Y126A/M129L and Y126A/M129A mutations of the “first-layer residues”, which directly contact the substrate amino acids, were transplanted into the corresponding sites (Tyr125 and Met128) of ISO4-G1 PylRS. The Y125A mutation (corresponding to Y306A in *M. mazei* PylRS and Y126A in *M. alvus* PylRS) enlarges the ISO4-G1 PylRS active site pocket, which then becomes suitable for accommodating bulky non-canonical amino acids [26]. In ISO4-G1 PylRS, the Met128 side-chain protrudes into the amino acid binding pocket (Fig. 3), which would reduce the pocket size as compared with that of *Mm*PylRSc. The Leu and Ala mutations at position 128 would enlarge the inner space of the active site pocket (Fig. 3). All of the ISO4-G1 PylRS mutant proteins were quite soluble, and over 100 mg quantities of the purified ISO4-G1 PylRS proteins per liter *E. coli* culture were obtained (Materials and methods).

Using the ISO4-G1 PylRS mutants, we tested cell-free protein synthesis for the site-specific incorporation of bulky non-canonical amino acids, such as ZLys, *mAz*ZLys, *N*^ε-(*p*-azidobenzyloxycarbonyl)-L-lysine (*pAz*ZLys), *N*^ε-(*p*-ethynylbenzyloxycarbonyl)-L-lysine (*pEt*ZLys), and TCO*Lys (Scheme 1), into the N11-GFPS1 protein in response to an amber (UAG) codon at position 17, using the cell extract of the RF-1 (*prfA*) deletion *E. coli* strain B-60.ΔA::Z [26,42]. The yields of the N11-GFPS1 proteins containing ZLys, TCO*Lys, *pEt*ZLys, *mAz*ZLys, and *pAz*ZLys by using the ISO4-G1 PylRS(Y125A/M128A)•tRNA^{Pyl} pair were 2.8, 0.9, 0.3, 2.8, 0.9 mg protein/mL reaction, respectively (132%, 41%, 15%, 131%, and 41% productivities, respectively, relative to the N11-GFPS1 control protein)(Fig. 6a).



Scheme 1. Chemical structures of pyrrolysine (Pyl) and non-canonical amino acids: BocLys, PocLys, ZLys, *mAz*ZLys, *pAz*ZLys, *pEt*ZLys, and TCO*Lys are shown.

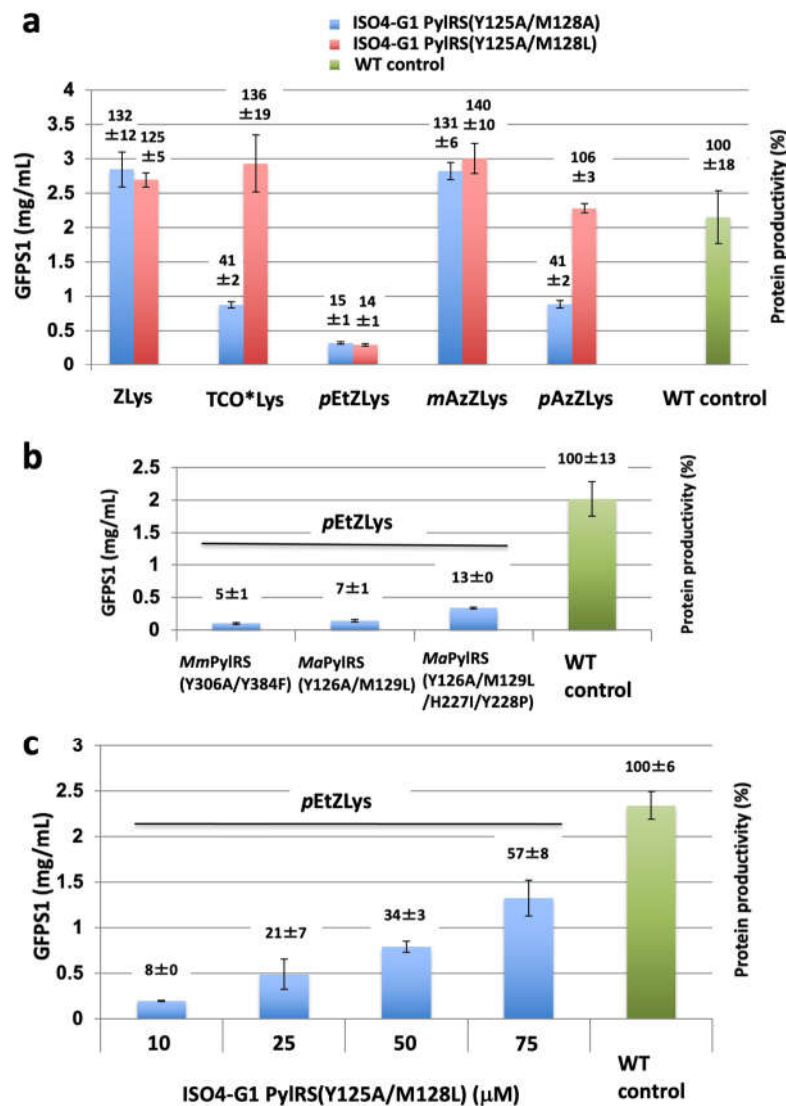


Figure 6. Cell-free protein synthesis for site-specific incorporation of non-canonical amino acids into the N11-GFPS1 protein by using ISO4-G1 PylRS with the first-layer mutations. The N11-GFPS1 proteins containing non-canonical amino acids were synthesized with the S30 extract from the *E. coli* RF1 deletion strain B-60ΔA::Z/pMINOR cells. Non-canonical amino acids were site-specifically incorporated into the N11-GFPS1 protein at position 17 in response to the UAG codon, by using the ISO4-G1 PylRS(Y125A/M128L)•tRNA^{Pyl} and ISO4-G1 PylRS(Y125A/M128A)•tRNA^{Pyl} pairs. Protein productivities with non-canonical amino acids were compared with that of the cell-free synthesis of wild-type N11-GFPS1 protein containing Ala at position 17 (WT control) and are shown above the bars. (a) The yields of the N11-GFPS1 proteins containing ZLys, mAzZLys, pAzZLys, pEtZLys, and TCO*Lys, estimated by fluorescence. The values represent the means of three independent experiments with standard deviations. (b) Cell-free synthesis of the N11-GFPS1 protein containing pEtZLys by using 10 μM of the *M. mazei* PylRS(Y306A/Y384F), *M. alvus* PylRS(Y126A/M129L), and *M. alvus* PylRS(Y126A/M129L/H228I/Y228P) proteins. The values represent the means of three independent experiments with standard deviations. (c) Cell-free protein synthesis of the GFP S1 protein containing pEtZLys, using increased concentrations (from 10 to 75 μM) of the ISO4-G1 PylRS(Y125A/M128L) protein. The values represent the means of three independent experiments with standard deviations.

In contrast, with the ISO4-G1 PylRS(Y125A/M128L)•tRNA^{Pyl} pair, the yields of the N11-GFPS1 proteins containing ZLys, TCO*Lys, pEtZLys, mAzZLys, and pAzZLys 2.7, 2.9, 0.3, 3, and 2.3 mg protein/mL reaction, respectively (125%, 136%, 14%, 140%, and 106% productivities, respectively, relative to the N11-GFPS1 control protein)(Fig. 6a). The ISO4-G1 PylRS(Y125A/M128L) mutant achieved more than 100% protein productivities for

ZLys, *mAzZLys*, *pAzZLys*, and TCO*Lys (Fig. 6a), while the ISO4-G1 PylRS(Y125A/M128A) did so only for ZLys and *mAzZLys*. The amino acid-binding pocket of the ISO4-G1 PylRS(Y125A/M128A) mutant might not be suitable for TCO*Lys and *pAzZLys*, unlike that of the ISO4-G1 PylRS(Y125A/M128L) mutant. Unfortunately, both ISO4-G1 PylRS mutants achieved only 14-15% of the protein productivity for *pEtZLys*, indicating that this substrate is too large for them. In contrast, the protein productivities for *pEtZLys* with the *MmPylRS*(Y306A/Y384F)•tRNA^{Pyl}, *MaPylRS*(Y126A/M129L)•tRNA^{Pyl}, and *MaPylRS*(Y126A/M129L/H227I/Y228P)•tRNA^{Pyl} pairs were 0.1, 0.14, and 0.34 mg/mL reaction (5, 7, and 13% protein productivities, respectively, relative to the N11-GFPS1 control protein), which were comparable to or less than that of ISO4-G1 PylRS(Y125A/M128L) (Fig. 6b). Consequently, the protein productivities of the ISO4-G1 PylRS(Y125A/M128L)•tRNA^{Pyl} pairs with TCO*Lys were drastically enhanced, and comparable to that obtained using the *MaPylRS*(Y126A/M129L/H227I/Y228P)•tRNA^{Pyl} pair [26].

2.9. Effects of the ISO4-G1 PylRS(Y125A/M128L) concentration on cell-free protein synthesis with the inefficient amino acid *pEtZLys*

Previously, we found that the protein productivities for the inefficient, bulky amino acid TCO*Lys can be enhanced by increasing the concentration of the *M. alvus* PylRS mutant [26]. To examine the effects of higher concentrations of the ISO4-G1 PylRS protein on non-canonical amino acid incorporation, cell-free protein synthesis with the super-inefficient, bulky, non-canonical amino acid *pEtZLys*, which is useful for alkyne-azide click chemistry [58,70], by using various concentrations of the ISO4-G1 PylRS protein. The protein productivities for *pEtZLys* were enhanced from 8% (0.19 mg protein/mL reaction) to 57 % (1.32 mg protein/mL reaction) of the N11-GFPS1 control protein when the concentration of the ISO4-G1 PylRS(Y125A/M128L) protein was increased from 10 to 75 μ M (Fig. 6c). Therefore, we achieved the highest-ever protein productivity for *pEtZLys*.

The incorporations of the non-canonical amino acids into the N11-GFPS1 protein were confirmed by mass spectrometry analyses (Fig. S6). The PMF analysis of the tryptic digests by MALDI-TOF mass spectrometry revealed major peaks (obsd: m/z 1,901.88 [M+H]⁺, m/z 1,925.82 [M+H]⁺, m/z 1,919.93 [M+H]⁺), which match the theoretical masses of the tryptic peptides HEHAHXENLYFQSK, where X represents ZLys, *pEtZLys*, and TCO*Lys, respectively (calcd: m/z 1,901.63 [M+H]⁺, m/z 1,925.64 [M+H]⁺, m/z 1,919.69 [M+H]⁺). The ESI mass analysis revealed the tryptic peptides containing *mAzZLys* (obsd: m/z 971.95 [M+2H]²⁺, calcd: m/z 971.95 [M+2H]²⁺; obsd: m/z 648.63 [M+3H]³⁺, calcd: m/z 648.30 [M+3H]³⁺; obsd: m/z 486.73 [M+4H]⁴⁺, calcd: m/z 486.48 [M+4H]⁴⁺) and *pAzZLys* (obsd: m/z 971.95 [M+2H]²⁺, calcd: m/z 971.95 [M+2H]²⁺; obsd: m/z 648.64 [M+3H]³⁺, calcd: m/z 648.30 [M+3H]³⁺; obsd: m/z 486.48 [M+4H]⁴⁺, calcd: m/z 486.48 [M+4H]⁴⁺). These results confirmed that the efficient synthesis of the full-length N11-GFPS1 protein containing non-canonical amino acids occurs without any non-specific suppression of the UAG codon with canonical amino acids in the cell-free system.

2.10. Effects of the second-layer mutations of ISO4-G1 PylRS for site-specific incorporation of bulky lysine derivatives into proteins by cell-free protein synthesis

We found that the productivities of N11-GFPS1 proteins containing ZLys, *mAzZLys*, and TCO*Lys obtained with the ISO4-G1 PylRS(Y125A/M128L) system were comparable to or higher than those from the *MaPylRS*(Y126A/M129L/H227I/H228P) system (Fig. 6a)[26]. Previously, the second-layer IP (H227I/H228P) additional mutations of *MaPylRS*(Y126A/M129L) extensively improved the protein productivities for *mAzZLys* and TCO*Lys [26]. However, mutations of the second-layer residues His227 and Tyr228 in *MaPylRS* [corresponding to the Ile405 and Pro406 residues in *MmPylRSc*, respectively (Figs. 4, S1, and S5)] might affect the first-layer residues, which interact directly with substrate amino acids within the amino acid binding pocket [26].

Interestingly, the ISO4-G1 PylRS structure revealed that ISO4-G1 PylRS His225, corresponding to *Ma*PylRS His227, undergoes drastic conformational changes in accordance with the location of Tyr204 (corresponding to Tyr206 in *Ma*PylRS) in the $\beta 5$ - $\beta 6$ hairpin (Figs. 5 and S5). We introduced the H225A mutation in ISO4-G1 PylRS and examined its effects on protein productivity. The ISO4-G1 PylRS(H225A) mutant had significantly decreased protein productivities for non-canonical amino acids (Fig. 7). The yields of the N11-GFPS1 proteins containing *N*^ε-(*t*-butyloxycarbonyl)-L-lysine (BocLys) and *N*^ε-propargyloxycarbonyl-L-lysine (PocLys)(Scheme 1) were only 0.096 and 0.048 mg protein/mL reaction, respectively (4% and 2% productivities, respectively, of the N11-GFPS1 control protein)(Fig. 7). These biochemical and crystallographic analyses confirmed that the His225 residue is crucial for the ISO4-G1 PylRS activity and cannot be replaced.

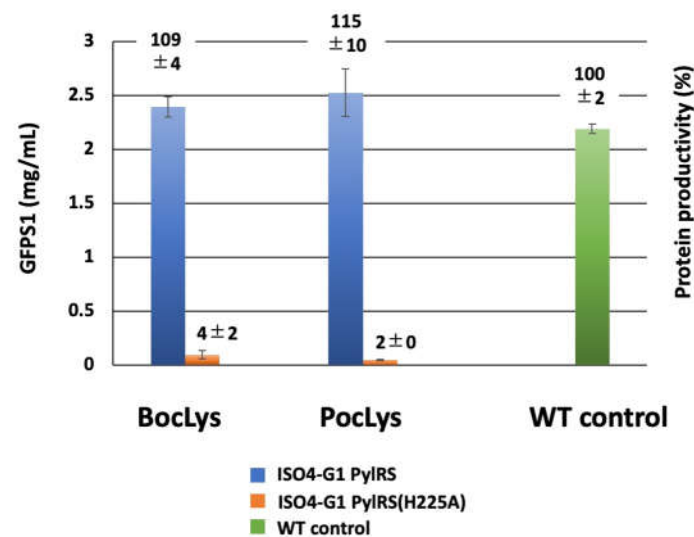


Figure 7. Cell-free protein synthesis with non-canonical amino acids using ISO4-G1 PylRS with the H225A mutation. The N11-GFPS1 proteins synthesized with the S30 extracts from *E. coli* B-60ΔA::Z/pMINOR cells in the presence of non-canonical amino acids. Non-canonical amino acids were site-specifically incorporated into the N11-GFPS1 protein at position 17 in response to the UAG codon, by using the ISO4-G1 PylRS•tRNA^{Pyl} and ISO4-G1 PylRS(H225A)•tRNA^{Pyl} pairs for BocLys and PocLys. The yields of the N11-GFPS1 proteins containing non-canonical amino acids were estimated by fluorescence. Protein productivities with non-canonical amino acids were compared with that of the cell-free synthesis of wild-type N11-GFPS1 protein containing Ala at position 17 (WT control) and are shown on the bars. The values represent the means of three independent experiments with standard deviations.

3. Discussion

In the present study, we determined the crystal structure of ISO4-G1 PylRS, and by its structure-based engineering, we achieved full productivity of cell-free protein synthesis according to the expanded genetic code with a variety of bulky non-canonical amino acids. By introducing two mutations into the first layer of the amino acid-binding pocket in ISO4-G1 PylRS, we achieved full productivity of cell-free synthesis with ZLys, TCO*Lys, *m*AzZLys, and *p*AzZLys. The first-layer mutant of ISO4-G1 PylRS required no additional second-layer mutations for the full productivity with these bulky non-canonical amino acids. Even with the much bulkier and most inefficient non-canonical amino acid, *p*EtZLys, we finally achieved the highest-ever levels of protein productivity by using the ISO4-G1 PylRS(Y125A/M128L) protein at a 7.5-fold higher concentration than the standard protocol. So far, this drastic improvement of protein productivity for *p*EtZLys has never been accomplished with the *M. mazei* and *M. alvus* systems.

Previously, we introduced the Y126A mutation of *Ma*PylRS (corresponding to the Y306A mutation of *Mm*PylRS), and the M129L or M129A mutation in the first-layer residues [26]. We found that simply transplanting the *Ma*PylRS(Y126A/M129L or

Y126A/M129A) mutations into ISO4-G1 PylRS was appropriate for bulky non-canonical amino acids. The two ISO4-G1 PylRS mutants (Y125A/M128L and Y125A/M128A) with enlarged amino acid binding pockets achieved full productivity and showed much higher activities than those of *MmPylRS*(Y306A/Y384F) for ZLys, *mAzZLys*, *pAzZLys*, and TCO*Lys (Figs. 3a and 6a). However, the full productivity level has not yet been achieved for more difficult non-canonical amino acids, such as *pEtZLys*. Because ISO4-G1 PylRS, as well as *MaPylRS*, is highly water-soluble, ISO4-G1 PylRS mutants can be used in the cell-free reaction at much higher concentrations than that of the standard protocol. Consequently, the yield of the *pEtZLys*-incorporated protein reached 1.3 mg/mL per cell-free reaction (57% productivity level of the control protein synthesis) when the concentration of the ISO4-G1 PylRS(Y125A/M128L) protein was increased up to 75 μ M (Fig. 6c).

The higher catalytic activity of ISO4-G1 PylRS than that of *MaPylRS* in the cell-free system was achieved for the site-specific incorporation of *N*^ε-(2-(trimethylsilyl)ethoxycarbonyl)-L-lysine into proteins [28]. The molecular mechanism underlying this higher catalytic activity of ISO4-G1 PylRS than those of *MaPylRS* and *MmPylRS* remains unknown. Based on the crystal structure of ISO4-G1 PylRS (Figs. 2, 4, and 5), the β 5– β 6 hairpin may exist in a dynamic open-closed equilibrium, and the location and conformational change of the His225 residue appear to be important for the catalytic activity. The ISO4-G1 PylRS His225 residue is conserved as His227 in *MaPylRS*, and undergoes a drastic conformational change upon non-canonical amino acid (and AMPPNP)-binding (Figs. 4 and S5) [64]. However, in *MaPylRS*, His227 does not interact with Tyr206 and Trp241, in contrast to the interactions of His225 with Tyr204 and Trp237 in ISO4-G1 PylRS. The ISO4-G1 PylRS mutant with His225 replaced by Ala225 abolished the protein productivities for non-canonical amino acids (Fig. 7). In the case of ISO4-G1 PylRS, the His225Ala mutation might reduce the interactions of His225 with Tyr204 and Trp237. In the above-mentioned dynamic closed-open equilibrium of the hairpin, the degree of movement of the hairpin in ISO4-G1 PylRS may be comparable to those in *MaPylRS* and *MmPylRS*, concerning the tip positions between the open and closed forms, although we still lack ISO4-G1 PylRS structures bound to amino acid substrates (Figs. 2, 3, 4, and S5). The interactions of His225 with Tyr204 and Trp237 in ISO4-G1 PylRS (Fig. 5), which are not observed in *MaPylRS*, appear to be a driving force for the rapid conformational changes of the β 5– β 6 hairpin. The elucidation of the molecular mechanism underlying the higher catalytic activities of the ISO4-G1 PylRS mutants based on the PylRS structures will lead to the development of a next-generation platform for producing non-canonical amino acid-incorporated proteins.

In the present study, we demonstrated that the ISO4-G1 PylRS system extensively improved the protein productivities, even for the very difficult, non-canonical amino acid *pEtZLys*, which had not been achieved by the *MmPylRS* and *MaPylRS* systems. The ISO4-G1 PylRS•tRNA^{Pyl} pair, rationally engineered based on the ISO4-G1 PylRS crystal structures, will serve as a more useful tool for next-generation genetic code expansion technologies.

4. Materials and Methods

4.1. Materials

Biochemical and molecular biological procedures were performed with commercially available materials, enzymes, and chemicals. ZLys was purchased from Bachem (Switzerland). *mAzZLys*, *pEtZLys*, and *pAzZLys*, were purchased from Sundia (China). BocLys was purchased from Watanabe Chemical (Japan). PocLys and TCO*Lys were purchased from SiChem GmbH (Germany).

4.2. Bacterial Strains and Plasmids

The DNA fragment encoding PylRS (ISO4-G1 PylRS, residues 1–273) from the methanogenic archaeon ISO4-G1 was chemically synthesized (Integrated DNA Technologies),

PCR-amplified, and cloned into the pET28c vector. The *E. coli* BL21-Gold(DE3) strain (Novagen) was used for the expression of PylRS proteins.

4.3. Expression and purification of PylRS proteins

The pET28c vector plasmids containing the ISO4-G1 PylRS gene were transformed into the *E. coli* BL21-Gold(DE3) strain, and selected on LB agar plates supplemented with 50 µg/ml kanamycin. A single colony was grown at 37°C in broth culture, containing 15 g tryptone, 7.5 g yeast extract, and 15 g NaCl per liter, supplemented with 50 µg/ml kanamycin. The protein expression was induced with 1 mM IPTG when the OD₆₀₀ reached 0.6. The cultivation temperature was then lowered to 20 °C, and the culture was continued overnight. The *E. coli* cells were collected by centrifugation and stored at –80 °C. The cells were resuspended in 50 mM potassium phosphate buffer (pH 7.4), containing 500 mM NaCl, 25 mM imidazole, 5 mM β-mercaptoethanol, and protease inhibitor cocktail (Complete-EDTA free ULTRA, Roche) (buffer C), and were sonicated on ice. The cell lysate was centrifuged at 15,000 × g for 15 min at 4 °C, and the supernatant was applied to a HisTrap column (Cytiva), which was equilibrated with buffer C. The protein was eluted with buffer C containing 400 mM imidazole, instead of 25 mM imidazole, and peak fractions were collected. The protein fractions were pooled, concentrated, and applied to a HiLoad 16/60 Superdex 200 column (Cytiva), equilibrated with 30 mM potassium phosphate buffer (pH 7.4), containing 200 mM NaCl and 1 mM DTT. The eluted fraction was collected and dialyzed against 40 mM potassium phosphate buffer (pH 7.4), containing 50 mM NaCl and 1 mM DTT (buffer B). The histidine-tag peptide derived from the pET28c vector was cleaved with thrombin protease (1 u per 0.1 mg PylRS protein, Sigma-Aldrich) at 4 °C overnight. The dialyzed fraction was then loaded on a HiTrap Q column (Cytiva), and after washing the column with buffer B, the bound proteins were eluted by a linear gradient of 50–635 mM NaCl. The eluted fractions were pooled, concentrated, and applied to a HiLoad 16/60 Superdex 200 column (Cytiva), equilibrated with 30 mM potassium phosphate buffer (pH 7.4), containing 200 mM NaCl and 1 mM DTT. The eluted fractions were collected, dialyzed against 10 mM Tris-HCl buffer (pH 8.0), containing 150 mM NaCl, 10 mM MgCl₂, and 10 mM β-mercaptoethanol, and concentrated by ultracentrifugation to 16.2 mg/ml. Aliquots of the ISO4-G1 PylRS protein were flash-cooled in liquid nitrogen and stored at –80 °C. The *Ma*PylRS and *Mm*PylRS proteins were purified as described previously [26]. The histidine-tagged PylRS proteins were purified by chromatography on HisTrap and Superdex 200 HiLoad columns. After dialysis, the eluted PylRS proteins were concentrated by ultracentrifugation.

4.4. Preparation of tRNA^{Pyl} transcripts

The tRNA^{Pyl}s from the methanogenic archaeon ISO4-G1, *M. alvus*, and *M. mazei* were transcribed in vitro with T7 RNA polymerase, using the PCR-amplified DNA fragment as the template. The tRNA transcripts were precipitated with isopropanol, applied onto a Resource Q column (Cytiva) equilibrated with 10 mM Tris-HCl buffer (pH 7.5), containing 5 mM MgCl₂ and 50 mM NaCl, and eluted by a linear gradient of 0.05–0.7 M NaCl. The purified tRNA^{Pyl} transcripts were precipitated with ethanol and dissolved in 10 mM Tris-HCl buffer (pH 7.5) containing 5 mM MgCl₂.

4.5. Crystallization, data collection, and structure determination

All crystallization screenings were performed by the sitting-drop vapor-diffusion method, by mixing 0.2 µl of the ISO4-G1 PylRS protein solution with 0.2 µl of reservoir solution, using a Mosquito liquid handling robot (TTP Labtech). Crystals were grown at 20 °C in conditions with 100 mM HEPES-NaOH buffer (pH 7.2), 20% PEG3350, and 200 mM KCl. The crystal was transferred to 100 mM HEPES-NaOH buffer (pH 7.2) containing 20% PEG3350, 200 mM KCl, and 18% trehalose, mounted on a nylon loop, and flash-cooled in liquid nitrogen. The X-ray diffraction datasets were collected at the beamline BL32XU in SPring-8 (Harima, Japan) at –173 °C and were processed with XDS [71]. The

crystal of ISO4-G1 PylRS belongs to the space group $P2_12_12_1$, with unit cell parameters of $a=98.51$ Å, $b=102.68$ Å, $c=349.86$ Å, and $\alpha=\beta=\gamma=90^\circ$. The phase was calculated by the molecular replacement method with Phaser, using 6JP2 as the search model. Ten ISO4-G1 PylRS molecules were found per asymmetric unit, with a solvent content of 56.6%. Iterative cycles of model refinement by PHENIX [72] and manual model building with Coot [73] were performed. The R_{work} and R_{free} factors for the ISO4-G1 PylRS structure are shown in Table S1. The final model was validated with Molprobity [66] and Procheck [65]. Graphical images were prepared with PyMOL [http://pymol.sourceforge.net/]. The statistics of the data collection and refinement are summarized in Table S1.

4.6. Cell-free protein synthesis and purification of GFP proteins containing non-canonical amino acids

Cell-free coupled transcription/translation was performed as described previously [32][33][35][74], using pCR2.1-TOPO bearing the gene encoding an N11-tagged superfold type green fluorescent mutant protein (N11-GFPS1) [75]. The tRNA^{Pyl} transcripts were prepared by *in vitro* transcription [35,47]. The PylRS proteins were overproduced in *E. coli* BL21-Gold(DE3) cells and purified as described previously [8]. The pCR2.1-N11-GFPS1 plasmids containing the wild-type N11-GFPS1 gene or the mutant with a single UAG codon at position Ala17 were used as the template DNAs for cell-free protein synthesis with S30 extracts from RF-1 (*prfA*) deletion *E. coli* B-60ΔA::Z cells [42] with a pMINOR plasmid encoding rare codon tRNAs [76]. The reaction components for the incorporation of non-natural lysine derivatives at position 17 in N11-GFPS1 were as follows: 2 μg/ml template plasmid, 10 μM PylRS, 10 μM tRNA^{Pyl}, and 1 mM non-natural lysine derivatives. After an overnight incubation at 25 °C, the synthesized full-length N11-GFPS1 proteins were quantified as described previously, using an ARVO Victor2 V Multilabel Counter plate fluorescence reader (PerkinElmer) [35]. The N11-GFPS1 proteins were purified as follows. After centrifugation of the solution, the supernatant fractions were loaded on a Ni-Sepharose High Performance column (Cytiva). The column was washed with 50 mM potassium phosphate buffer (pH 7.4), containing 500 mM NaCl, 25 mM imidazole, and 5 mM β-mercaptoethanol, and then eluted with 50 mM potassium phosphate buffer (pH 7.4), containing 500 mM NaCl, 400 mM imidazole, and 5 mM β-mercaptoethanol. The PMF analyses of the N11-GFPS1 proteins, containing ZLys, *m*AzZLys, *p*AzZLys, *p*EtZLys, and TCO*ZLys, were performed as described above.

6. Patents

A PCT international patent application [WO2020/045656 A1] related to this work has been filed.

Supplementary Materials: Figure S1: Structure-based sequence alignments of ISO4-G1 PylRS and other PylRSs.; Figure S2: Chromatogram for the purification of ISO4-G1 PylRS by Superdex 200 size-exclusion chromatography. Figure S3: Structural comparison of ISO4-G1 PylRS with *Ma*PylRS, *Mm*PylRS, and *Dh*PylRS•tRNA^{Pyl}. Figure S4: Electron density map of the β5-β6 region in the ISO4-G1 PylRS structure. Figure S5: Conformational changes of the active-site residues in the open and closed forms of the ISO4-G1 PylRS, *Ma*PylRS, and *Mm*PylRS structures. Figure S6: Mass spectrometry analysis of N11-GFPS1 proteins containing non-canonical amino acids. Table S1: Data collection and refinement statistics.

Author Contributions: Conceptualization, T.Y., S.Y.; investigation, T.Y., H.T., Y.F., E.S.; methodology, T.Y., E.S., H.T., Y.F.; Project Administration, K.S., S.Y.; Writing—Original Draft, T.Y.; Writing—Review & Editing, T.Y., S.Y.; Funding Acquisition, T.Y., S.Y.

Funding: This work was supported by the following Grants. The Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP17am0101081 (S.Y.); Leading Advanced Projects for Medical Innovation (LEAP) from AMED under Grant Number JP19gm0010001 (S.Y.); Takeda Science Foundation (S.Y.); MEXT grant numbers 16K05859 and 24550203 (T.Y.).

Data Availability Statement: The coordinates and structure factors have been deposited in the RSCB Protein Data Bank (ID code 8IFJ for ISO4-G1 PylRS).

Acknowledgments: We would like to thank the staff of the beamline BL32XU at SPring-8 (Harima, Japan), as well as Dr. Takaho Terada, Takako Imada, and Tomoko Nakayama for clerical assistance. We thank the Support Unit for Bio-Material Analysis, RIKEN CBS Research Resources Division, especially Kaori Otsuki and Shino Kurata, for the mass spectrometry analysis.

Conflicts of Interest: T.Y., E.S., K.S. and S.Y. are co-inventors on the patent [WO2020/045656 A1] related to this work. S.Y. is a founder and shareholder of LiberoThera Co., Ltd.

Abbreviations: *Mm*: *Methanosarcina mazei*; *Mb*: *Methanosarcina barkeri*; *Ma*: *Methanomethylophilus alvus*; *Dh*: *Desulfotobacterium hafniense*; ISO4-G1: The methanogenic archaeon ISO4-G1; BocLys: N^ε-(*t*-butyloxycarbonyl)-L-lysine; PocLys: N^ε-propargyloxycarbonyl-L-lysine; ZLys: N^ε-benzyloxycarbonyl-L-lysine; *mAzZLys*: N^ε-(*m*-azidobenzyloxycarbonyl)-L-lysine; *pAzZLys*: N^ε-(*p*-azidobenzyloxycarbonyl)-L-lysine; *pEtZLys*: N^ε-(*p*-ethynylbenzyloxycarbonyl)-L-lysine; TCO*Lys: N^ε-(((*E*)-cyclooct-2-en-1-yl)oxy)carbonyl-L-lysine; MALDI-TOF MS: matrix assisted laser deionization/ionization-time of flight mass spectrometry; PMF: Peptide mass fingerprinting.

References

- Wang, L.; Xie, J.; Schultz, P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249, doi:10.1146/annurev.biophys.35.101105.121507.
- Liu, C.C.; Schultz, P.G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444, doi:10.1146/annurev.biochem.052308.105824.
- 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.; Krzycki, J.A. Direct charging of tRNA^{CUA} with pyrrolysine in vitro and in vivo. *Nature* **2004**, *17*, 503–507.
- Polycarpo, C.; Ambrogelly, A.; Bérubé, A.; Winbush, S.A.M.; McCloskey, J.A.; Crain, P.F.; Wood, J.L.; Söll, D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12450–12454, doi:10.1073/pnas.0405362101.
- Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Söll, D. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 3141–3146, doi:10.1073/pnas.0611634104.
- Neumann, H.; Peak-Chew, S.Y.; Chin, J.W. Genetically encoding N^ε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4*, 232–234, doi:10.1038/nchembio.73.
- Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822, doi:10.1016/j.bbrc.2008.04.164.
- Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N^ε-(*o*-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* **2008**, *15*, 1187–1197, doi:10.1016/j.chembiol.2008.10.004.
- Chen, P.R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B.H.; Schultz, P.G. A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chemie - Int. Ed.* **2009**, *48*, 4052–4055, doi:10.1002/anie.200900683.
- Wan, W.; Tharp, J.M.; Liu, W.R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844*, 1059–1070, doi:10.1016/j.bbapap.2014.03.002.
- Hao, B.; Gong, W.; Ferguson, T.K.; James, C.M.; Krzycki, J.A.; Chan, M.K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* (80-.). **2002**, *296*, 1462–1466, doi:10.1126/science.1069556.
- Srinivasan, G.; James, C.M.; Krzycki, J.A. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science* (80-.). **2002**, *296*, 1459–1462, doi:10.1126/science.1069588.
- Lee, M.M.; Jiang, R.; Jain, R.; Larue, R.C.; Krzycki, J.; Chan, M.K. Structure of *Desulfotobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* **2008**, *374*, 470–474, doi:10.1016/j.bbrc.2008.07.074.
- Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Arais, Y.; Ishitani, R.; Umehara, T.; Söll, D.; Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} structure reveals the molecular basis of orthogonality. *Nature* **2009**, *457*, 1163–1167, doi:10.1038/nature07611.
- Chin, J.W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **2014**, *83*, 379–408, doi:10.1146/annurev-biochem-060713-035737.
- Crnković, A.; Suzuki, T.; Söll, D.; Reynolds, N.M. Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat. Chem. Acta* **2016**, *89*, 163–174, doi:10.5562/cca2825.
- Brabham, R.; Fascione, M.A. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *ChemBioChem* **2017**, *18*, 1973–1983, doi:10.1002/cbic.201700148.
- Chin, J.W. Expanding and reprogramming the genetic code. *Nature* **2017**, *550*, 53–60, doi:10.1038/nature24031.
- Wang, L. Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res.* **2017**, *50*, 2767–2776, doi:10.1021/acs.accounts.7b00376.

20. Vargas-Rodriguez, O.; Sevostyanova, A.; Söll, D.; Crnković, A. Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol.* **2018**, *46*, 115–122, doi:10.1016/j.cbpa.2018.07.014.
21. Tharp, J.M.; Ehnbohm, A.; Liu, W.R. tRNAPyl: Structure, function, and applications. *RNA Biol.* **2017**, *15*, 441–452, doi:10.1080/15476286.2017.1356561.
22. Willis, J.C.W.; Chin, J.W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **2018**, *10*, 831–837, doi:10.1038/s41557-018-0052-5.
23. Meineke, B.; Heimgärtner, J.; Lafranchi, L.; Elsässer, S.J. Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem. Biol.* **2018**, *13*, 3087–3096, doi:10.1021/acscchembio.8b00571.
24. Yamaguchi, A.; Irahra, F.; Ohtake, K.; Sakamoto, K. Pyrrolysyl-tRNA synthetase with a unique architecture enhances the availability of lysine derivatives in synthetic genetic codes. *Molecules* **2018**, *23*, doi:10.3390/molecules23102460.
25. Beránek, V.; Willis, J.C.W.; Chin, J.W. An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells. *Biochemistry* **2019**, *58*, 387–390, doi:10.1021/acs.biochem.8b00808.
26. Seki, E.; Yanagisawa, T.; Kuratani, M.; Sakamoto, K.; Yokoyama, S. Fully Productive Cell-Free Genetic Code Expansion by Structure-Based Engineering of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase. *ACS Synth. Biol.* **2020**, *9*, 718–732, doi:10.1021/acssynbio.9b00288.
27. Meineke, B.; Heimgärtner, J.; Eirich, J.; Landreh, M.; Elsässer, S.J. Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cell Rep.* **2020**, *31*, doi:10.1016/j.celrep.2020.107811.
28. Abdelkader, E.H.; Qianzhu, H.; Tan, Y.J.; Adams, L.A.; Huber, T.; Otting, G. Genetic Encoding of N6-(((Trimethylsilyl)methoxy)carbonyl)-L-lysine for NMR Studies of Protein-Protein and Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2021**, *143*, 1133–1143, doi: 10.1021/jacs.0c11971.
29. Abdelkader, E.H.; Qianzhu, H.; George, J.; Frkic, R.L.; Jackson, C.J.; Nitsche, C.; Otting, G.; Huber, T. Genetic Encoding of Cyanopyridylalanine for In-Cell Protein Macrocyclization by the Nitrile-Aminothiol Click Reaction. *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202114154. doi: 10.1002/anie.202114154.
30. Avila-Crump, S.; Hemshorn, M.L.; Jones, C.M.; Mbengi, L.; Meyer, K.; Griffis, J.A.; Jana, S.; Petrina, G.E.; Pagar, V.V.; Karplus, P.A.; Petersson, E.J.; Perona, J.J.; Mehl, R.A.; Cooley, R.B. Generating Efficient Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetases for Structurally Diverse Non-Canonical Amino Acids. *ACS Chem. Biol.* **2022**, *17*, 12, 3458–3469, doi: 10.1021/acscchembio.2c00639.
31. Yanagisawa, T.; Umehara, T.; Sakamoto, K.; Yokoyama, S. Expanded genetic code technologies for incorporating modified lysine at multiple sites. *Chembiochem* **2014**, *15*, 2181–2187, doi:10.1002/cbic.201402266.
32. Mukai, T.; Yanagisawa, T.; Ohtake, K.; Wakamori, M.; Adachi, J.; Hino, N.; Sato, A.; Kobayashi, T.; Hayashi, A.; Shirouzu, M.; et al. Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* **2011**, *411*, 757–761, doi:10.1016/j.bbrc.2011.07.020.
33. Yanagisawa, T.; Takahashi, M.; Mukai, T.; Sato, S.; Wakamori, M.; Shirouzu, M.; Sakamoto, K.; Umehara, T.; Yokoyama, S. Multiple site-specific installations of N ϵ -monomethyl-L-lysine into histone proteins by cell-based and cell-free protein synthesis. *ChemBioChem* **2014**, *15*, 1830–1838, doi:10.1002/cbic.201402291.
34. Chemla, Y.; Ozer, E.; Schlesinger, O.; Noireaux, V.; Alfonta, L. Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* **2015**, *112*, 1663–1672, doi:10.1002/bit.25587.
35. Seki, E.; Yanagisawa, T.; Yokoyama, S. Cell-free protein synthesis for multiple site-specific incorporation of noncanonical amino acids using cell extracts from RF-1 deletion E. coli strains. *Methods Mol. Biol.* **2018**, *1728*, 49–65, doi:10.1007/978-1-4939-7574-7_3.
36. Adachi, J.; Katsura, K.; Seki, E.; Takemoto, C.; Shirouzu, M.; Terada, T.; Mukai, T.; Sakamoto, K.; Yokoyama, S. Cell-free protein synthesis using S30 extracts from Escherichia coli RFzero strains for efficient incorporation of non-natural amino acids into proteins. *Int. J. Mol. Sci.* **2019**, *20*, 1–12, doi:10.3390/ijms20030492.
37. Gerrits, M.; Budisa, N.; Merk, H. Site-Specific Chemoselective Pyrrolysine Analogues Incorporation Using the Cell-Free Protein Synthesis System. *ACS Synth. Biol.* **2019**, *8*, 381–390, doi:10.1021/acssynbio.8b00421.
38. Mukai, T.; Hayashi, A.; Irahra, F.; Sato, A.; Ohtake, K.; Yokoyama, S.; Sakamoto, K. Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res.* **2010**, *38*, 8188–8195, doi:10.1093/nar/gkq707.
39. Johnson, D.B.F.; Xu, J.; Shen, Z.; Takimoto, J.K.; Schultz, M.D.; Schmitz, R.J.; Xiang, Z.; Ecker, J.R.; Briggs, S.P.; Wang, L. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **2011**, *7*, 779–786, doi:10.1038/nchembio.657.
40. Lajoie, M.J.; Rovner, A.J.; Goodman, D.B.; Aerni, H.R.; Haimovich, A.D.; Kuznetsov, G.; Mercer, J.A.; Wang, H.H.; Carr, P.A.; Mosberg, J.A.; et al. Genomically recoded organisms expand biological functions. *Science* (80-.). **2013**, *342*, 357–360, doi:10.1126/science.1241459.
41. Hong, S.H.; Ntai, I.; Haimovich, A.D.; Kelleher, N.L.; Isaacs, F.J.; Jewett, M.C. Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.* **2014**, *3*, 398–409, doi:10.1021/sb400140t.
42. Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Sci. Rep.* **2015**, *5*, 1–9, doi:10.1038/srep09699.

43. Fredens, J.; Wang, K.; de la Torre, D.; Funke, L.F.H.; Robertson, W.E.; Christova, Y.; Chia, T.; Schmied, W.H.; Dunkelmann, D.L.; Beránek, V.; et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **2019**, *569*, 514–518, doi:10.1038/s41586-019-1192-5.
44. Suzuki, T.; Miller, C.; Guo, L.T.; Ho, J.M.L.; Bryson, D.I.; Wang, Y.S.; Liu, D.R.; Söll, D. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **2017**, *13*, 1261–1266, doi:10.1038/nchembio.2497.
45. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Nureki, O.; Yokoyama, S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2006**, *62*, 1031–1033, doi:10.1107/S1744309106036700.
46. Herring, S.; Ambrogelly, A.; Gundllapalli, S.; O'Donoghue, P.; Polycarpo, C.R.; Söll, D. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett.* **2007**, *581*, 3197–3203, doi:10.1016/j.febslet.2007.06.004.
47. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **2008**, *378*, 634–652, doi:10.1016/j.jmb.2008.02.045.
48. Borrel, G.; Parisot, N.; Harris, H.M.B.; Peyretailade, E.; Gaci, N.; Tottey, W.; Bardot, O.; Raymann, K.; Gribaldo, S.; Peyret, P.; et al. Comparative genomics highlights the unique biology of *Methanomassiliicoccales*, a *Thermoplasmatales*-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **2014**, *15*, doi:10.1186/1471-2164-15-679.
49. Kavran, J.M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T.A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11268–11273, doi:10.1073/pnas.0704769104.
50. Takimoto, J.K.; Dellas, N.; Noel, J.P.; Wang, L. Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* **2011**, *6*, 733–743, doi:10.1021/cb200057a.
51. Schneider, S.; Gattner, M.J.; Vrabel, M.; Flügel, V.; López-Carrillo, V.; Prill, S.; Carell, T. Structural insights into incorporation of norbornene amino acids for click modification of proteins. *ChemBioChem* **2013**, *14*, 2114–2118, doi:10.1002/cbic.201300435.
52. Yanagisawa, T.; Sumida, T.; Ishii, R.; Yokoyama, S. A novel crystal form of pyrrolysyl-tRNA synthetase reveals the pre- and post-aminoacyl-tRNA synthesis conformational states of the adenylate and aminoacyl moieties and an asparagine residue in the catalytic site. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 5–15, doi:10.1107/S0907444912039881.
53. Flügel, V.; Vrabel, M.; Schneider, S. Structural basis for the site-specific incorporation of lysine derivatives into proteins. *PLoS One* **2014**, *9*, 1–7, doi:10.1371/journal.pone.0096198.
54. Schmidt, M.J.; Weber, A.; Pott, M.; Welte, W.; Summerer, D. Structural basis of furan-amino acid recognition by a polyspecific aminoacyl-tRNA-synthetase and its genetic encoding in human cells. *ChemBioChem* **2014**, *15*, 1755–1760, doi:10.1002/cbic.201402006.
55. Guo, L.T.; Wang, Y.S.; Nakamura, A.; Eiler, D.; Kavran, J.M.; Wong, M.; Kiessling, L.L.; Steitz, T.A.; O'Donoghue, P.; Söll, D. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 16724–16729, doi:10.1073/pnas.1419737111.
56. Englert, M.; Nakamura, A.; Wang, Y.S.; Eiler, D.; Söll, D.; Guo, L.T. Probing the active site tryptophan of *Staphylococcus aureus* thioredoxin with an analog. *Nucleic Acids Res.* **2015**, *43*, 11061–11067, doi:10.1093/nar/gkv1255.
57. Lee, Y.J.; Schmidt, M.J.; Tharp, J.M.; Weber, A.; Koenig, A.L.; Zheng, H.; Gao, J.; Waters, M.L.; Summerer, D.; Liu, W.R. Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun.* **2016**, *52*, 12606–12609, doi:10.1039/c6cc05959g.
58. Yanagisawa, T.; Kuratani, M.; Seki, E.; Hino, N.; Sakamoto, K.; Yokoyama, S. Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase. *Cell Chem. Biol.* **2019**, *26*, 936–949.e13, doi:10.1016/j.chembiol.2019.03.008.
59. Jiang, H.K.; Wang, Y.H.; Weng, J.H.; Kurkute, P.; Li, C.L.; Lee, M.N.; Chen, P.J.; Tseng, H.W.; Tsai, M.D.; Wang, Y.S. Probing the Active Site of Deubiquitinase USP30 with Noncanonical Tryptophan Analogues. *Biochemistry* **2020**, *59*, 2205–2209, doi:10.1021/acs.biochem.0c00307.
60. Vatansever, E.C.; Yang, K.S.; Geng, Z.Z.; Qiao, Y.; Li, P.; Xu, S.; Liu, W.R. A Designed, Highly Efficient Pyrrolysyl-tRNA Synthetase Mutant Binds o-Chlorophenylalanine Using Two Halogen Bonds. *J. Mol. Biol.* **2022**, *434*, 167534. doi: 10.1016/j.jmb.2022.167534.
61. Kato, A.; Kuratani, M.; Yanagisawa, T.; Ohtake, K.; Hayashi, A.; Amano, Y.; Kimura, K.; Yokoyama, S.; Sakamoto, K.; Shiraishi, Y. Extensive Survey of Antibody Invariant Positions for Efficient Chemical Conjugation Using Expanded Genetic Codes. *Bioconjug. Chem.* **2017**, *28*, 2099–2108, doi:10.1021/acs.bioconjchem.7b00265.
62. Yamaguchi, A.; Matsuda, T.; Ohtake, K.; Yanagisawa, T.; Yokoyama, S.; Fujiwara, Y.; Watanabe, T.; Hohsaka, T.; Sakamoto, K. Incorporation of a Doubly Functionalized Synthetic Amino Acid into Proteins for Creating Chemical and Light-Induced Conjugates. *Bioconjug. Chem.* **2016**, *27*, 198–206, doi:10.1021/acs.bioconjchem.5b00602.
63. Baumann, T.; Hauf, M.; Richter, F.; Albers, S.; Möglich, A.; Ignatova, Z.; Budisa, N. Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* **2019**, *20*, doi:10.3390/ijms20092343.

64. Gottfried-Lee, I.; Perona, J.J.; Karplus, P.A.; Mehl, R.A.; Cooley, R.B.; Structures of Methanomethylophilus alvus Pyrrolysine tRNA-Synthetases Support the Need for De Novo Selections When Altering the Substrate Specificity. *ACS Chem. Biol.* **2022**, *17*, 12, 3470–3477, doi:10.1021/acscchembio.2c00640.
65. Collaborative Computational Project, No. 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* **1994**, *50*, 760–763, doi:10.1107/S0907444994003112.
66. Davis, I.W.; Leaver-Fay, A.; Chen, V.B.; Block, J.N.; Kapral, G.J.; Wang, X.; Murray, L.W.; Arendall, W.B.; Snoeyink, J.; Richardson, J.S.; et al. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, 375–383, doi:10.1093/nar/gkm216.
67. Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **1990**, *347*, 203–206.
68. Cusack, S.; Berthet-Colominas, C.; Härtlein, M.; Nassar, N.; Leberman, R. A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å. *Nature* **1990**, *347*, 249–255, doi:10.1038/347249a0.
69. Ruff, M.; Krishnaswamy, S.; Boeglin, M.; Poterszman, A.; Mitschler, A.; Podjarny, A.; Rees, B.; Thierry, J.C.; Moras, D. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast apartyl-tRNA synthetase complexed with tRNA(Asp). *Science* (80-.). **1991**, *252*, 1682–1689.
70. Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **2001**, *40*, 2004–2021, doi:10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.
71. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 125–32, doi:10.1107/S0907444909047337.
72. Adams, P.D.; Afonine, P.V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221, doi:10.1107/S0907444909052925.
73. Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126–2132, doi:10.1107/S0907444904019158.
74. Kigawa, T.; Yabuki, T.; Matsuda, N.; Matsuda, T.; Tanaka, A.; Yokoyama, S. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **2004**, *5*, 63–68.
75. Seki, E.; Matsuda, N.; Yokoyama, S.; Kigawa, T. Cell-free protein synthesis system from Escherichia coli cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation. *Anal. Biochem.* **2008**, *377*, 156–161, doi:10.1016/j.ab.2008.03.001.
76. Chumpolkulwong, N.; Hori-Takemoto, C.; Hosaka, T.; Inaoka, T.; Kigawa, T.; Shirouzu, M.; Ochi, K.; Yokoyama, S. Effects of Escherichia coli ribosomal protein S12 mutations on cell-free protein synthesis. *Eur. J. Biochem.* **2004**, *271*, 1127–1134, doi:10.1111/j.1432-1033.2004.04016.x.
77. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680, doi:10.1093/nar/22.22.4673.