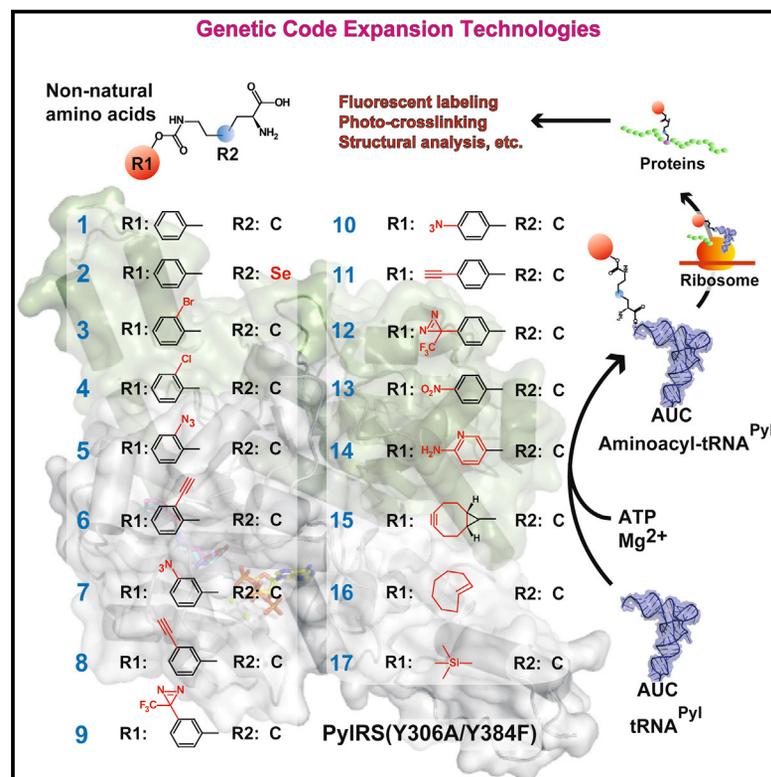


Cell Chemical Biology

Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase

Graphical Abstract



Authors

Tatsuo Yanagisawa, Mitsuo Kuratani,
Eiko Seki, Nobumasa Hino,
Kensaku Sakamoto,
Shigeyuki Yokoyama

Correspondence

tatsuo.yanagisawa@riken.jp (T.Y.),
yokoyama@riken.jp (S.Y.)

In Brief

Yanagisawa et al. analyzed the Y306A/Y384F mutant of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS) with 17 non-natural, bulky oxycarbonyllysine derivatives for tRNA^{Pyl} aminoacylation and site-specific incorporation into proteins. Fourteen crystal structures of the amino acid-bound PylRS mutant revealed the structural bases of the binding. This information facilitates the structure-based design of novel amino acids.

Highlights

- A mutant pyrrolysyl-tRNA synthetase, PyIRS(Y306A/Y384), acts on diverse amino acids
- The PylRS mutant and tRNA^{Pyl} incorporated 17 non-natural amino acids into proteins
- Crystal structures of the PylRS mutant bound with 14 of the amino acids were solved
- This information will facilitate the structure-based design of novel amino acids

Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase

Tatsuo Yanagisawa,^{1,4,6,*} Mitsuo Kuratani,^{1,6} Eiko Seki,^{1,3} Nobumasa Hino,⁵ Kensaku Sakamoto,^{2,4} and Shigeyuki Yokoyama^{1,3,7,*}

¹RIKEN Structural Biology Laboratory, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

²Division of Structural and Synthetic Biology, RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

³RIKEN Cluster for Science, Technology and Innovation Hub, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

⁴Laboratory for Nonnatural Amino Acid Technology, RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

⁵Laboratory of Molecular Medicine, Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: tatsuo.yanagisawa@riken.jp (T.Y.), yokoyama@riken.jp (S.Y.)

<https://doi.org/10.1016/j.chembiol.2019.03.008>

SUMMARY

Pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl} have been extensively used for genetic-code expansion. A *Methanosarcina mazei* PylRS mutant bearing the Y306A and Y384F mutations (PylRS(Y306A/Y384F)) encodes various bulky non-natural lysine derivatives by UAG. In this study, we examined how PylRS(Y306A/Y384F) recognizes many amino acids. Among 17 non-natural lysine derivatives, N^ε-(benzyloxycarbonyl)lysine (ZLys) and 10 *ortho/meta/para*-substituted ZLys derivatives were efficiently ligated to tRNA^{Pyl} and were incorporated into proteins by PylRS(Y306A/Y384F). We determined crystal structures of 14 non-natural lysine derivatives bound to the PylRS(Y306A/Y384F) catalytic fragment. The *meta*- and *para*-substituted ZLys derivatives are snugly accommodated in the productive mode. In contrast, ZLys and the unsubstituted or *ortho*-substituted ZLys derivatives exhibited an alternative binding mode in addition to the productive mode. PylRS(Y306A/Y384F) displayed a high aminoacylation rate for ZLys, indicating that the double-binding mode minimally affects aminoacylation. These precise substrate recognition mechanisms by PylRS (Y306A/Y384F) may facilitate the structure-based design of novel non-natural amino acids.

INTRODUCTION

The structures and functions of proteins are restricted by the natural variety of building blocks, L-amino acids. Therefore, expanding the repertoire of amino acids in translation is useful for developing novel protein functions (reviewed in Wang et al.,

2006; Liu and Schultz, 2010). “Orthogonal” pairs of an engineered aminoacyl-tRNA synthetase and tRNA, including bacterial and archaeal pairs of tyrosyl-tRNA synthetase (TyrRS) and tRNA^{Tyr}_(CUA) (Wang and Schultz, 2001; Wang et al., 2002; Chin et al., 2002, 2003; Kiga et al., 2002; Sakamoto et al., 2002) and archaeal pairs of pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl}_(CUA) for the 22nd amino acid, pyrrolysine (Blight et al., 2004; Polycarpo et al., 2004), have enabled the site-specific incorporation of non-natural amino acids into proteins in response to the amber (UAG) codon (reviewed in Wang et al., 2006; Liu and Schultz, 2010; Wan et al., 2014; Chin, 2014; Crnković et al., 2016; Brabham and Fascione, 2017; Chin, 2017; Wang, 2017; Vargas-Rodriguez et al., 2018; Tharp et al., 2018). Chemically reactive functional groups, such as alkene, alkyne, azide, and diazirine groups, on non-natural amino acids allow the post-translational labeling of proteins with detection probes, polymers, drugs, and UV crosslinkers (reviewed in Lang and Chin, 2014a, 2014b; Elliott et al., 2014; Nguyen et al., 2018).

Pyrrolysine, PylRS, and tRNA^{Pyl} are present in several methanogenic archaea, including *Methanosarcina barkeri* (Hao et al., 2002; Srinivasan et al., 2002), and in a few bacteria, including *Desulfitobacterium hafniense* (Lee et al., 2008; Nozawa et al., 2009). PylRS belongs to the class IIc aaRSs (Eriani et al., 1990; Ruff et al., 1991; Ibba and Söll, 2000). This class is phylogenetically most similar to phenylalanyl-tRNA synthetase (Kavran et al., 2007), and esterifies tRNA^{Pyl} with pyrrolysine. PylRS consists of two domains, the N-terminal tRNA binding domain and the C-terminal catalytic domain (Herring et al., 2007; Yanagisawa et al., 2008a; Jiang and Krzycki, 2012; Suzuki et al., 2017). Pyrrolysine is site specifically incorporated into a specific UAG site of methylamine methyltransferases with *M. barkeri* PylRS (Blight et al., 2004; Polycarpo et al., 2004). Thus, the PylRSs from *M. barkeri*, *Methanosarcina mazei*, and *D. hafniense* have been used for the incorporation of non-standard amino acids into proteins (Ambrogelly et al., 2007; Neumann et al., 2008; Mukai et al., 2008; Yanagisawa et al., 2008b; Chen et al., 2009; Nguyen et al., 2009; Katayama et al., 2012).

The unique feature of PylRS is its broad specificity for substrate amino acids. A large number of PylRS mutants that accommodate various sizes of non-canonical amino acids have been identified, and more than 100 non-natural amino acids and their analogs have been incorporated into proteins *in vivo* (reviewed in Wan et al., 2014; Yanagisawa et al., 2014b; Brabham and Fascione, 2017) and *in vitro* (Mukai et al., 2011; Yanagisawa et al., 2014a; Seki et al., 2018). A lysine derivative with a benzyloxycarbonyl (Z) group, N^ε-benzyloxycarbonyl-L-lysine (ZLys) (Mukai et al., 2008; Yanagisawa et al., 2008b), which is much larger than pyrrolysine and BocLys, was reportedly a poor substrate of the wild-type PylRS. Site-specific incorporations of ZLys and its derivatives into proteins have been successfully achieved by using *M. mazei* tRNA^{Pyl} and PylRS with a double mutation (Y306A and Y384F), which was obtained by rational and random screening (Yanagisawa et al., 2008b). The *M. mazei* PylRS(Y306/Y384F)/tRNA^{Pyl} system enabled the productive incorporation of N^ε-(*o*-azidobenzoyloxycarbonyl)-L-lysine (oAzZLys) with a reactive azide group (Yanagisawa et al., 2008b; Kato et al., 2017), which can be used for bioorthogonal labeling of proteins (Kolb et al., 2001; Kiick et al., 2002). In a similar manner to the tyrosine and phenylalanine analogs (e.g., Liu and Schultz, 2010), the aromatic ring of ZLys is used as a scaffold on which chemically reactive groups are attached. As for the “bioorthogonal” labeling of proteins (Blackman et al., 2008; Lang and Chin, 2014a), the PylRS(Y306A/Y384F)/tRNA^{Pyl} system successfully incorporated the *ortho*-, *meta*-, and *para*-substituted ZLys derivatives, including oAzZLys (Yanagisawa et al., 2008b), N^ε-(*p*-nitrobenzyloxycarbonyl)-L-lysine (pNO₂ZLys), N^ε-(*p*-trifluoromethyl-diazirinyloxybenzyloxycarbonyl)-L-lysine (pTmdZLys) (Yanagisawa et al., 2012), N^ε-(*m*-azidobenzoyloxycarbonyl)-L-lysine (mAzZLys) (Yamaguchi et al., 2016), N^ε-(3-amino-5-azidobenzoyloxycarbonyl)-L-lysine (mAmAzZLys) (Yamaguchi et al., 2016), and N^ε-(*m*-trifluoromethyl-diazirinyloxybenzyloxycarbonyl)-L-lysine (mTmdZLys) (Kita et al., 2016). Beside the ZLys derivatives described above, PylRS (Y306A/Y384F) also accepts non-natural lysine derivatives, containing cyclooctyne (N^ε-(((1*R*,8*S*)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)-L-lysine [BCNLys] [Lang et al., 2012; Borrmann et al., 2012]) and *trans*-cyclooctene moieties (N^ε-(((*E*)-*trans*-cyclooct-2-ene-1-yl)oxy)carbonyl)-L-lysine [TCO**Lys*] [Plass et al., 2012; Nikić et al., 2014]).

The crystal structures of the catalytic fragment (residues 185–454) of *M. mazei* PylRS and its mutants in complex with a number of substrate amino acids have been solved. The *M. mazei* wild-type PylRS structures in complex with pyrrolysine (Yanagisawa et al., 2006, 2008a; Kavran et al., 2007), N^ε-(*tert*-butyloxycarbonyl)-L-lysine (BocLys) (Yanagisawa et al., 2008b, 2013), N^ε-allyloxycarbonyl-L-lysine (AlocLys) (Yanagisawa et al., 2008b), N^ε-cyclopentylloxycarbonyl-D-lysine (CpocLys) (Kavran et al., 2007), N^ε-propionyl-L-lysine, N^ε-butyl-L-lysine, N^ε-crotonyl-L-lysine, and N^ε-propargyloxycarbonyl-L-lysine (PocLys) (Flügel et al., 2014) have been reported. Furthermore, the structures of the PylRS(A302T/N346V/C348W/V401L/Y384F) mutant in complex with *O*-methyl-L-tyrosine (Takimoto et al., 2011), the PylRS(Y306G/Y384F/I405R) mutant complexed with N^ε-(5-norbornene-2-yloxycarbonyl)-L-lysine (Schneider et al., 2013), the PylRS(N346S/C348I) mutant complexed with 3-iodo-L-phenylalanine, and 2-(5-bromothieryl)-L-alanine, the PylRS(L301V/L305I/Y306F/L309A/C348F) mutant complexed with N^ε-acetyl-

L-lysine (Guo et al., 2014), the PylRS(N346G/C348Q) mutant complexed with 3-benzothieryl-L-alanine (Englert et al., 2015), the PylRS(N346A/C348A) and PylRS(Y306A/N346A/C348A/Y384F) mutants complexed with phenylalanine (Lee et al., 2016), and the PylRS(Y306A/Y384F) mutant complexed with N^ε-(furan-2-propyloxycarbonyl)-L-lysine (Schmidt et al., 2014) have been reported. In spite of the uniquely broad substrate specificity of the PylRS(Y306A/Y384F) mutant, only one structure with a furan-containing amino acid has been analyzed (Schmidt et al., 2014). The molecular elucidation of mechanism by which the PylRS(Y306A/Y384F) accommodates a variety of large and bulky non-natural amino acids within its active site is essential for understanding its amazingly broad specificity.

In the present study, we investigated the incorporation efficiencies of 17 large and bulky non-natural lysine derivatives at a specific site in proteins with *M. mazei* PylRS(Y306A/Y384F) in the *Escherichia coli* cell-based and cell-free protein syntheses. We then performed X-ray crystallographic analyses of the catalytic fragment (residues 185–454) of *M. mazei* PylRS(Y306A/Y384F) complexed with 14 out of the 17 non-natural lysine derivatives. Thus, the mechanisms underlying the broad specificity of *M. mazei* PylRS(Y306A/Y384F) have been established. This structural basis will enable the design of more useful non-canonical amino acids than ever before.

RESULTS

In Vivo Site-Specific Incorporation of Bulky Lysine Derivatives into a Protein Using the Pair of *M. mazei* PylRS(Y306A/Y384F) and tRNA^{Pyl}

M. mazei PylRS(Y306A/Y384F) shows extensively broad specificity for substrate amino acids that are larger than pyrrolysine, and numerous bulky non-natural amino acids have been incorporated into proteins *in vivo* (Mukai et al., 2008; Yanagisawa et al., 2008b, 2012; Plass et al., 2012; Borrmann et al., 2012; Nikić et al., 2014; Yamaguchi et al., 2016; Kita et al., 2016). In this study, we compared the incorporations of 17 non-natural lysine derivatives, ZLys (Mukai et al., 2008; Yanagisawa et al., 2008b), and its derivatives containing selenium (ZaeSeCys) (Wang et al., 2012), nitro (pNO₂ZLys) (Virdee et al., 2011; Yanagisawa et al., 2012), halogens (oBrZLys [this study], and oClZLys [Yokoyama et al., 2010]), azide (oAzZLys [Yanagisawa et al., 2008b], mAzZLys [Yamaguchi et al., 2016], pAzZLys [Ge et al., 2016]), alkyne (oEtZLys, mEtZLys, and pEtZLys [this study]), diazirine (pTmdZLys [Yanagisawa et al., 2012] and mTmdZLys [Kita et al., 2016]), and lysine derivatives containing cyclooctyne (BCNLys [Lang et al., 2012; Borrmann et al., 2012]), *trans*-cyclooctene (TCO**Lys* [Plass et al., 2012; Nikić et al., 2014]), aminopyridine (pAmPylLys [this study]), and silicon (TeocLys [this study]) (Figure 1) into proteins with the pair of *M. mazei* PylRS(Y306A/Y384F) and tRNA^{Pyl} in the *E. coli* cell-based system, as follows.

First, we incorporated the 17 non-natural lysine derivatives into the T7 peptide-tagged GST protein (T7-GST), at a single UAG-specified site (position 25) between the T7 peptide and GST, in *E. coli* cells expressing PylRS(Y306A/Y384F) and tRNA^{Pyl} (Figure 2, Key Resources Table) (Yanagisawa et al., 2008b). The amounts of the expressed full-length proteins ranged from 1.7 to 78.5 mg per liter medium (Table S1). The most efficiently introduced derivatives were ZLys, ZaeSeCys, oBrZLys, oClZLys,

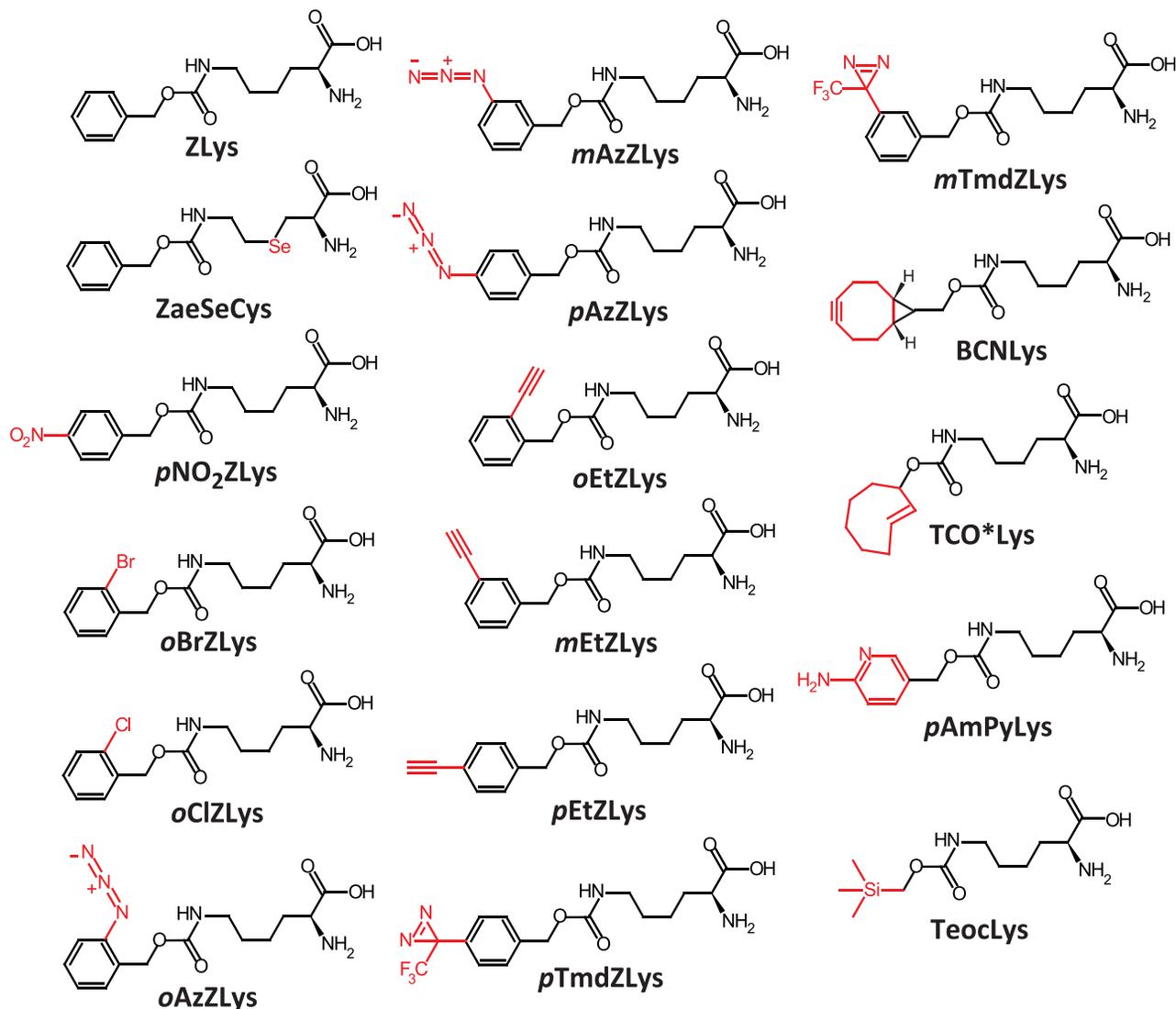


Figure 1. Chemical Structures of the Non-natural Lysine Derivatives Analyzed in This Study

The characteristic functional group of each lysine derivative is emphasized with red coloring.

pNO₂ZLys, *oAzZLys*, *mAzZLys*, *oEtZLys*, *mEtZLys*, and *mTmdZLys*. These are amino acids with a benzene ring (*ZLys* and *ZaeSeCys*) and an *ortho/meta*-substituted benzene ring (*oBrZLys*, *oClZLys*, *oAzZLys*, *mAzZLys*, *oEtZLys*, *mEtZLys*, and *mTmdZLys*), while *pNO₂ZLys* is exceptional. The produced T7-GST proteins containing these non-natural lysine derivatives amounted to 51.8–78.5 mg per liter medium, and to 56.9%–86.2% of the T7-GST protein produced with the tyrosine codon UAC instead of UAG at position 25 (Table S1). The T7-GST proteins containing *pAzZLys*, *BCNLys*, *TCO*Lys*, and *TeocLys* were produced with moderate efficiency (12.4–26.7 mg per liter medium), and those containing *pEtZLys*, *pAmPyLys*, and *pTmdZLys* were poorly produced (1.7–3.6 mg per liter medium).

We previously reported the site-specific incorporations of *ZLys*, *oAzZLys*, *pNO₂ZLys*, and *pTmdZLys*, at the single specified site (position 25) of the T7-GST proteins by mass spectrometry (Yanagisawa et al., 2008b, 2012). In this study, analyses of

the produced T7-GST proteins by the same mass spectrometric method confirmed the site-specific incorporation of 16 of the 17 non-natural lysine derivatives (Figure S1). Peptide mass fingerprinting (PMF) analyses of tryptic digests revealed major peaks, which matched the theoretical masses of the tryptic peptide NSXSPILGWYWK, where X represents *ZaeSeCys*, *oBrZLys*, *oClZLys*, *mAzZLys*, *pAzZLys*, *oEtZLys*, *mEtZLys*, *pEtZLys*, *BCNLys*, *TCO*Lys*, *TeocLys*, and *mTmdZLys*, respectively (Table S2). The incorporation of *pAmPyLys* was not confirmed by the PMF analysis, because an insufficient amount of the T7-GST protein containing *pAmPyLys* was obtained (Figure 2).

Fluorescent Labeling of Proteins Containing an Alkyne Moiety by Click Chemistry with Copper-Chelating Azide

During the last decade, the site-specific incorporation of non-natural amino acids bearing reactive groups into proteins and the bioorthogonal labeling of proteins have become important

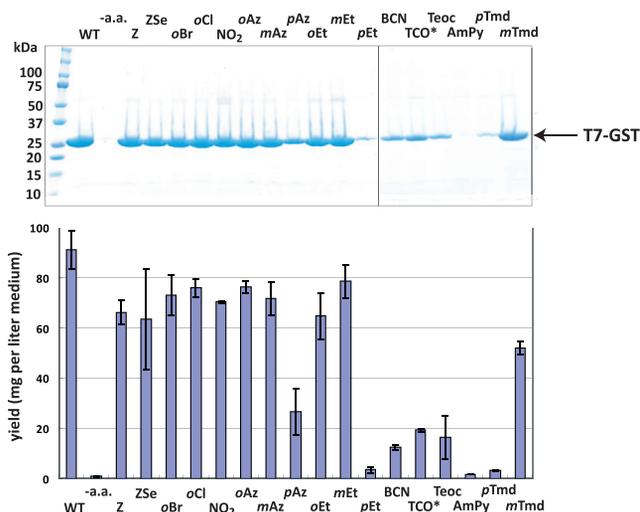


Figure 2. Site-Specific Incorporation of Non-natural Lysine Derivatives into the T7-GST Protein in *E. coli* Cells

E. coli BL21-Gold(DE3) cells harboring the T7-GST gene with a single UAG codon at position 25, together with the *M. mazei* PylRS(Y306A/Y384F) and *M. mazei* tRNA^{Pyl} genes, were grown in the absence or presence of non-natural amino acids. WT, wild-type T7-GST; -a.a., no non-natural amino acid; Z, ZLys; ZSe, ZaeSeCys; oBr, oBrZLys; oCl, oClZLys; NO₂, pNO₂ZLys; oAz, oAzZLys; mAz, mAzZLys; pAz, pAzZLys; oEt, oEtZLys; mEt, mEtZLys; pEt, pEtZLys; BCN, BCNLys; TCO*, TCO*Lys; Teoc, TeocLys; AmPy, pAmPyLys; pTmd, pTmdZLys; mTmd, mTmdZLys. The T7-GST proteins containing non-natural amino acids were purified by glutathione Sepharose column chromatography (upper panel). Yields of the purified T7-GST proteins per mg liter of *E. coli* culture (lower panel). The values represent the means of two independent experiments with standard deviations. The images are composites of those of two different gels, and each gel image is boxed. See also Figure S1, Tables S1 and S2, and the STAR Methods.

technologies for the analyses of protein structures/functions in cells (Lang and Chin, 2014a). Especially, azide and alkyne moieties are the most frequently used functional groups, and incorporations of non-natural amino acids and labeling of many proteins have been performed using bioorthogonal chemical reactions, such as Staudinger ligation (Kiick et al., 2002; Yanagisawa et al., 2008b) and click chemistry (copper-assisted azide-alkyne cycloaddition, or CuAAC) (Kolb et al., 2001; Nguyen et al., 2009). Copper ion accelerates the CuAAC reaction, but it has the serious problem that the high concentration of copper ion causes the denaturation of many proteins, and thus is incompatible with cell labeling or cell imaging because of the cytotoxicity. Fortunately, the cell-compatible copper chelator picorylazide, which reduces the toxicity of the CuAAC reaction conditions to cells, was recently developed (Uttamapinant et al., 2012). Thus, we tested two new lysine derivatives, oEtZLys and mEtZLys, for protein fluorescent labeling by click chemistry with picorylazide (Figure S2A). The T7-GST proteins containing oEtZLys and mEtZLys were successfully labeled with fluorescent AlexaFluor 488-picorylazide, with a much lower copper concentration than that needed for the CuAAC reaction (Figure S2A). Actually, the T7-GST proteins containing other non-natural lysine derivatives were not labeled. This method is expected to be applicable to the fluorescent labeling of unstable proteins *in vitro*, and further to the labeling of live cells.

UV Crosslinking of the Protein Dimer Interface via a Photo-Crosslinker Amino Acid

Non-natural photo-crosslinker amino acids with arylazido, *p*-benzoylphenyl, and diazirinyl moieties have been developed for genetic encoding in bacterial and mammalian cells (Chin et al., 2002; Hino et al., 2005, 2011; Ai et al., 2011). These systems have been applied to *in vivo* photo-crosslinking of proteins for identifying target proteins. Previously, we achieved wide-range protein crosslinking with the photo-reactive lysine derivative pTmdZLys, by using the pair of *M. mazei* PylRS(Y306A/Y384F) and tRNA^{Pyl} (Yanagisawa et al., 2012). Furthermore, mTmdZLys is incorporated much more efficiently than pTmdZLys, and is useful for studies of protein-protein interactions in mammalian cells (Kita et al., 2016). Besides mammalian cells, mTmdZLys can also be used for protein crosslinking in *E. coli* cells. The incorporation of mTmdZLys at one of the dimer interface residues, Glu51, of GST in the T7 peptide-tagged format was efficient (Figure S2B). After purification and UV irradiation of the T7-GST(51mTmdZLys) protein, the crosslinked GST dimer was clearly identified by SimplyBlue (ThermoFisher) staining and western blot analyses (Figure S2B).

M. mazei PylRS(Y306A/Y384F) Esterifies tRNA^{Pyl} with Large and Bulky Lysine Derivatives

As described above, the T7-GST proteins containing pEtZLys, pAmPyLys, and pTmdZLys were only poorly produced (Figure 2). Therefore, we compared the tRNA^{Pyl} aminoacylation efficiencies of *M. mazei* PylRS(Y306A/Y384F) for all 17 non-natural lysine derivatives (Figure 3). Among the three poorly incorporated non-natural amino acids, pEtZLys and pAmPyLys showed the lowest aminoacylation activities. Consequently, the very low *in vivo* incorporation efficiencies of pEtZLys and pAmPyLys are ascribed to the low aminoacylation activities.

pEtZLys has a bulky ethynyl group at the *para* position of the benzene ring of ZLys, which may prevent this amino acid from binding in the active site of PylRS(Y306A/Y384F). The very low water solubility of pEtZLys makes it difficult to test higher concentrations. In contrast, in the case of pAmPyLys, the aminoacyl-tRNA^{Pyl} was effectively produced at a higher concentration (10 mM), rather than 1 mM. Therefore, pAmPyLys can bind in the active site of PylRS(Y306A/Y384F), but its K_m value is appreciably higher than those of the other 15 non-natural amino acids. pAmPyLys contains a hydrophilic aminopyridine ring, which may be the reason for its poorer binding in the PylRS(Y306A/Y384F) active site, unlike the benzene ring of the ZLys derivatives.

The efficiency of PylRS(Y306A/Y384F) to esterify tRNA^{Pyl} with pTmdZLys was as high as those with ZLys, mEtZLys, mAzZLys, and mTmdZLys (Figure 3), whereas the incorporation of pTmdZLys into T7-GST by the cell-based system was much less efficient than those of the four other non-natural lysine derivatives.

Site-Specific Incorporation of Bulky Lysine Derivatives into a Protein by Cell-free Protein Synthesis

To test whether the poor *in vivo* incorporation efficiency of pTmdZLys is due to its low efficiency in some processes after aminoacylation, we applied a coupled cell-free transcription-translation system. The N11-tagged superfold type green fluorescent mutant protein (GFPS1) gene (Seki et al., 2008) was

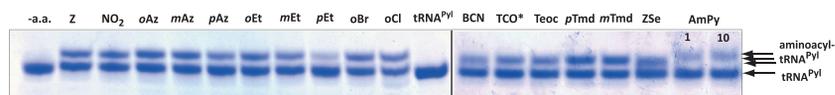


Figure 3. tRNA Aminoacylation Activities for Non-natural Amino Acids by *M. mazei* PyIRS (Y306A/Y384F)

M. mazei PyIRS(Y306A/Y384F) esterifies tRNA^{Pyl} with non-natural lysine derivatives. The aminoacylation of

tRNA^{Pyl} was monitored by acidic urea PAGE. Each lane shows a reaction in the presence of PyIRS(Y306A/Y384F), tRNA^{Pyl}, and the following: no non-natural amino acid (-a.a); Z, ZLys (Z), pNO₂ZLys (NO₂), oAzZLys (oAz), mAzZLys (mAz), pAzZLys (pAz), oEtZLys (oEt), mEtZLys (mEt), pEtZLys (pEt), oBrZLys (oBr), oClZLys (oCl), BCNLys (BCN), TCO**Lys* (TCO*), TeocLys (Teoc), pTmdZLys (pTmd), mTmdZLys (mTmd), and ZaeSeCys (ZSe) at 1 mM; pAmPyLys (AmPy) at 1 and 10 mM; and in the presence of only tRNA^{Pyl} (tRNA^{Pyl}). The images are composites of those of two different gels, and each gel image is boxed. See also STAR Methods.

expressed in the cell-free protein synthesis system using the S30 extract from *E. coli* BL21(DE3) cells. The full-length N11-GFPS1 protein containing pTmdZLys was produced with a yield of 0.9 mg protein per mL reaction, which was similar to those of ZLys, mAzZLys, mEtZLys, and mTmdZLys (1.33, 0.98, 1.35, and 0.78 mg per mL reactions, respectively) (Figure 4, Table S3). In contrast, the production of the N11-GFPS1 protein containing pAmPyLys was detectable but much lower than those of the other non-natural lysine derivatives (0.03 mg protein per mL reaction) (Figure 4; Table S3), in agreement with the low aminoacylation efficiency. We compared the relative yields of the non-natural amino acid-containing N11-GFPS1 proteins from the cell-based and cell-free systems (Table S3). Eleven milliliters of *E. coli* S30 extract, which are required for 33 mL of the cell-free reactions, can be obtained from 1 L of *E. coli* culture. Consequently, the yields of the N11-GFPS1 proteins containing ZLys, mAzZLys, mEtZLys, mTmdZLys, pTmdZLys, and pAmPyLys are estimated to be 43.9, 32.3, 44.6, 29.7, 25.7, and 0.99 mg, respectively, by the cell-free synthesis using a 1 L culture of *E. coli* cells (Table S3).

The incorporations of the non-natural lysine derivatives into the N11-GFPS1 protein were confirmed by mass spectrometry analyses (Figure S3). The PMF analysis of the tryptic digests by MALDI-TOF mass spectrometry revealed major peaks, which match the theoretical masses of the tryptic peptide HEHAHXENLYFQSK, where X represents ZLys, mEtZLys, pAmPyLys, pTmdZLys, and mTmdZLys, respectively (Table S2). Electrospray ionization mass analysis revealed the tryptic peptide containing mAzZLys (Table S2). These results confirmed that the efficient cell-free production of the full-length N11-GFPS1 protein with pTmdZLys does not occur with any non-specific suppression of the UAG codon with natural amino acids in the cell-free system.

Therefore, the low *in vivo* incorporation of pTmdZLys is ascribed to its low efficiency in some other processes relevant to the cell-based production, rather than aminoacylation or translation on the ribosome. The membrane permeability and/or amino acid transporter system of *E. coli* cells may be inefficient for pTmdZLys. Nevertheless, considering that most other ZLys derivatives, including mTmdZLys, are efficiently incorporated into proteins in *E. coli* cells, the low incorporation efficiency of pTmdZLys is odd and require further analysis.

Interactions of *M. mazei* PyIRS(Y306A/Y384F) with Lysine Derivatives

To examine the interaction between PyIRSc(Y306A/Y384F) and the lysine derivatives, a surface plasmon resonance (SPR) assay was performed. The K_d values for mTmdZLys, mEtZLys, BCNLys, TCO**Lys*, TeocLys, and oEtZLys were determined to be 1.62,

1.76, 3.9, 7.8, 11.3, and 13.8 mM (Figure S4). However, those for other lysine derivatives could not be measured, presumably because their affinities are too low. The binding affinities of PyIRSc(Y306A/Y384F) for mTmdZLys and mEtZLys were six to eight times higher than those for TeocLys and oEtZLys.

Crystal Structures of the *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment in Complex with ATP and Lysine Derivatives

To elucidate the structural basis for the recognition of the non-natural amino acid substrates by *M. mazei* PyIRS(Y306A/Y384F), we performed crystallographic analyses of the catalytic domain of PyIRSc(Y306A/Y384F) (PyIRSc(Y306A/Y384F)) in complex with 14 of the 17 non-natural amino acids described above (Table S4).

In 8 of the 14 determined structures, the bound lysine derivative molecules (pNO₂ZLys, pAmPyLys, pTmdZLys, mTmdZLys, mAzZLys, mEtZLys, BCNLys, and TCO**Lys*) were modeled normally in a single-binding mode. However, in the PyIRSc(Y306A/Y384F) structures for ZLys, ZaeSeCys, oBrZLys, oClZLys, and oAzZLys, the observed electron density corresponding to the amino acid is appreciably longer than that of a single molecule (Figures 5 and 6). The side chains of these four amino acids have, in common, an unsubstituted or *ortho*-substituted benzene ring and are smaller than the above eight ZLys derivatives. Therefore, these “longer” electron densities were interpreted as composites of two binding modes, shifted from each other. Furthermore, in the TeocLys structure, a similar “longer” electron density was observed, and the bound amino acid was modeled on the assumption of three binding modes (Figure 6). These unusual multiple binding modes may occur because of the enlarged binding pocket due to the double mutation Y306A/Y384F.

Structure of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment in Complex with ATP and the Lysine Derivatives with an Unsubstituted Benzene Ring

As the representative of the double-binding mode, the case of ZLys is described here in detail. ZLys was modeled in a composite of two different binding modes, 1 and 2, with occupancies of 0.58 and 0.37, respectively, which gave the lowest *R* factors against the diffraction data (Figure 5). The F_o – F_c omit electron density map of ZLys is appreciably longer than that of a single ZLys molecule, and corresponds well to the dual modes, as shown in Figure 5C.

Mode 1 is almost the same as the authentic binding mode reported for pyrrolysine and other lysine derivatives bound to *M. mazei* PyIRSc (Yanagisawa et al., 2008a, 2008b), in which the carbonyl oxygen atom in the carbamate (-O-CO-NH-) moiety

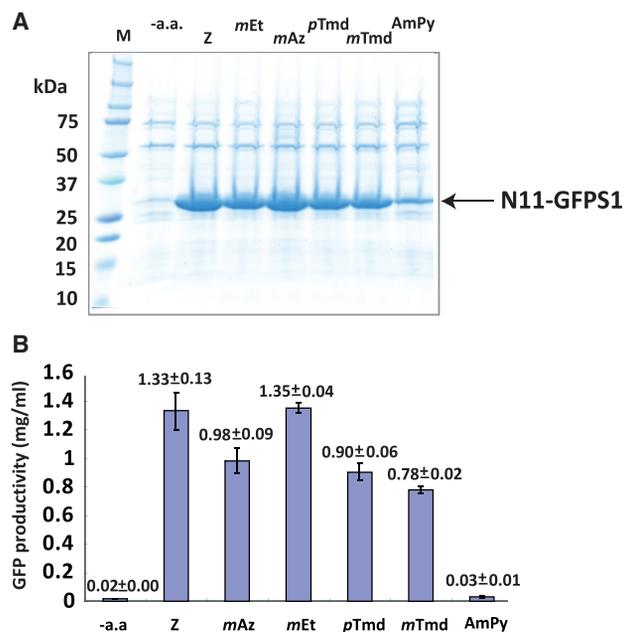


Figure 4. Site-Specific Incorporation of Non-natural Lysine Derivatives into the N11-GFPS1 Protein by Cell-free Protein Synthesis

Full-length N11-GFPS1 proteins synthesized with the S30 extracts from *E. coli* BL21(DE3)/pMINOR cells. Non-natural lysine derivatives were site specifically incorporated into the single amber site of the N11-GFPS1 protein at position 17 with the PyIRS(Y306A/Y384F)/tRNA^{PyL} pair.

(A) Production of the N11-GFPS1 proteins containing non-natural lysine derivatives, estimated by fluorescence. The values represent the means of three independent experiments with standard deviations.

(B) The N11-GFPS1 proteins purified by Ni-Sepharose column chromatography.

See also [Figure S3](#), [Tables S2](#) and [S3](#), and [STAR Methods](#).

of ZLys hydrogen bonds with the side-chain amide $-CO-NH_2$ group of Asn346 (the oxygen-nitrogen distance is 3.0 Å) ([Figures 5A](#) and [5D](#)). The $F_o - F_c$ omit electron density map ([Figure 5C](#)) exhibits a peak corresponding to the hydrogen bonding carbamate carbonyl oxygen atom of ZLys. With respect to this hydrogen bond between the carbamate carbonyl group and Asn346, mode 1 corresponds to the single-binding modes of the *meta*- and *para*-substituted ZLys derivatives, as described below.

In contrast, ZLys bound in mode 2 is buried more deeply than in mode 1, and fills up the remaining space in the hydrophobic pocket ([Figures 5B](#) and [S5A](#)). The benzene ring of ZLys is stacked with the indole ring of Trp417 in both binding modes. In contrast, in mode 2, the benzene ring of ZLys is located closer to the hydrophobic side chains of Met276, Ile405, Leu407, and Trp411 ([Figure S5A](#)), with distances of 3.4, 5.3, 3.7, and 3.5 Å, as compared with 5.0, 5.6, 5.4, and 5.1 Å, respectively, in mode 1. In particular, the δ -methyl group of Leu407 forms a van der Waals contact with the *p*-CH group with a C-H distance of 3.68 Å, which limits the deepest position of the benzene ring ([Figures 5A](#) and [5D](#)). The conformation of the Leu407 side chain is fixed by steric interactions with its surrounding residues (main and side chains). Instead of forming this van der Waals contact, the carbamate moiety moves away from Asn346 and thus the carbonyl oxygen atom does not form the hydrogen bond with

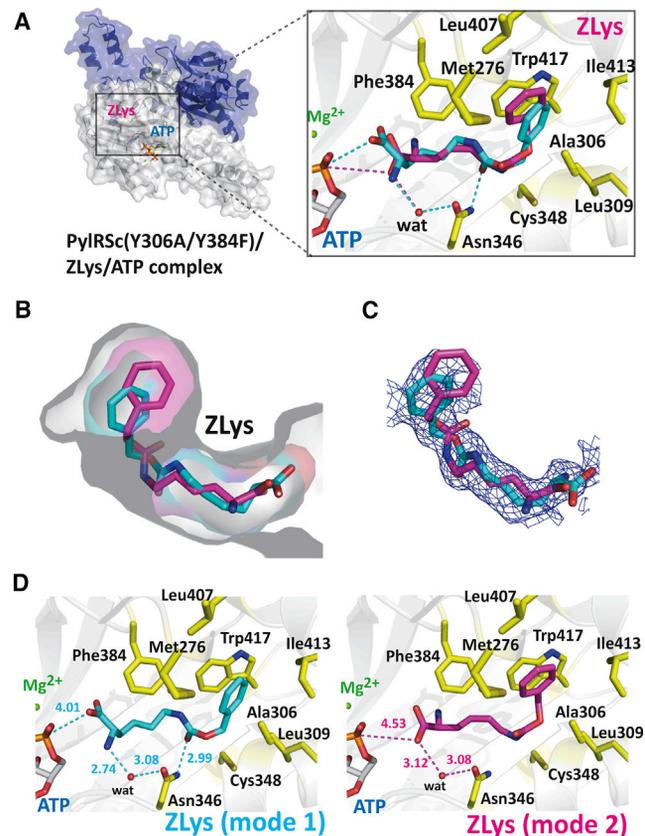


Figure 5. Crystal Structures of *M. mazei* PyIRS(Y306A/Y384F) Complexed with ZLys and ATP

(A) Overall structure of *M. mazei* PyIRS(Y306A/Y384F) bound to ZLys (left). Close-up views of ZLys (right). Two different binding modes of ZLys are shown as cyan (mode 1) and magenta (mode 2) stick models. Water-mediated hydrogen bonds between ZLys and the side-chain amide group of Asn348 are represented by dotted lines. Transparent ribbon models of PyIRS(Y306A/Y384F) are visible in the background.

(B) Space-filling model of *M. mazei* PyIRS(Y306A/Y384F) and ZLys.

(C) The $F_o - F_c$ omit electron density maps (contoured at 1.5 σ) of ZLys in the PyIRS(Y306A/Y384F) active site.

(D) Close-up views of ZLys in mode 1 (cyan, left) and in mode 2 (magenta, right). Hydrogen bonding distances between ZLys and the side-chain amide group of Asn348 are shown.

See also [Figures S5–S7](#), and [Tables S4](#) and [S5](#).

Asn346 (the oxygen-nitrogen distance is 5.2 Å) ([Figures 5A](#) and [5D](#)).

In mode 1, the Asn346 side chain forms a water-mediated hydrogen bond with the α -amino group (2.74 Å) of ZLys ([Figures 5A](#) and [5D](#)), as in the authentic pyrrolysine-bound PyIRS structure ([Yanagisawa et al., 2008a](#)). In contrast, in mode 2, the Asn346 side chain forms the corresponding water-mediated hydrogen bond with the α -carboxyl group (3.12 Å) of ZLys ([Figures 5A](#) and [5D](#)), as in the previously reported BocLys-bound PyIRS structure ([Yanagisawa et al., 2008b](#)). The distance between the α -carboxyl group of ZLys in mode 1 and the α -phosphate group of ATP (4.01 Å) is closer than that in mode 2 (4.53 Å) ([Figures 5A](#) and [5D](#)).

ZaeSeCys is an analog of ZLys, with an unsubstituted benzene ring and a selenium atom within the lysine side chain ([Figure 1](#)).

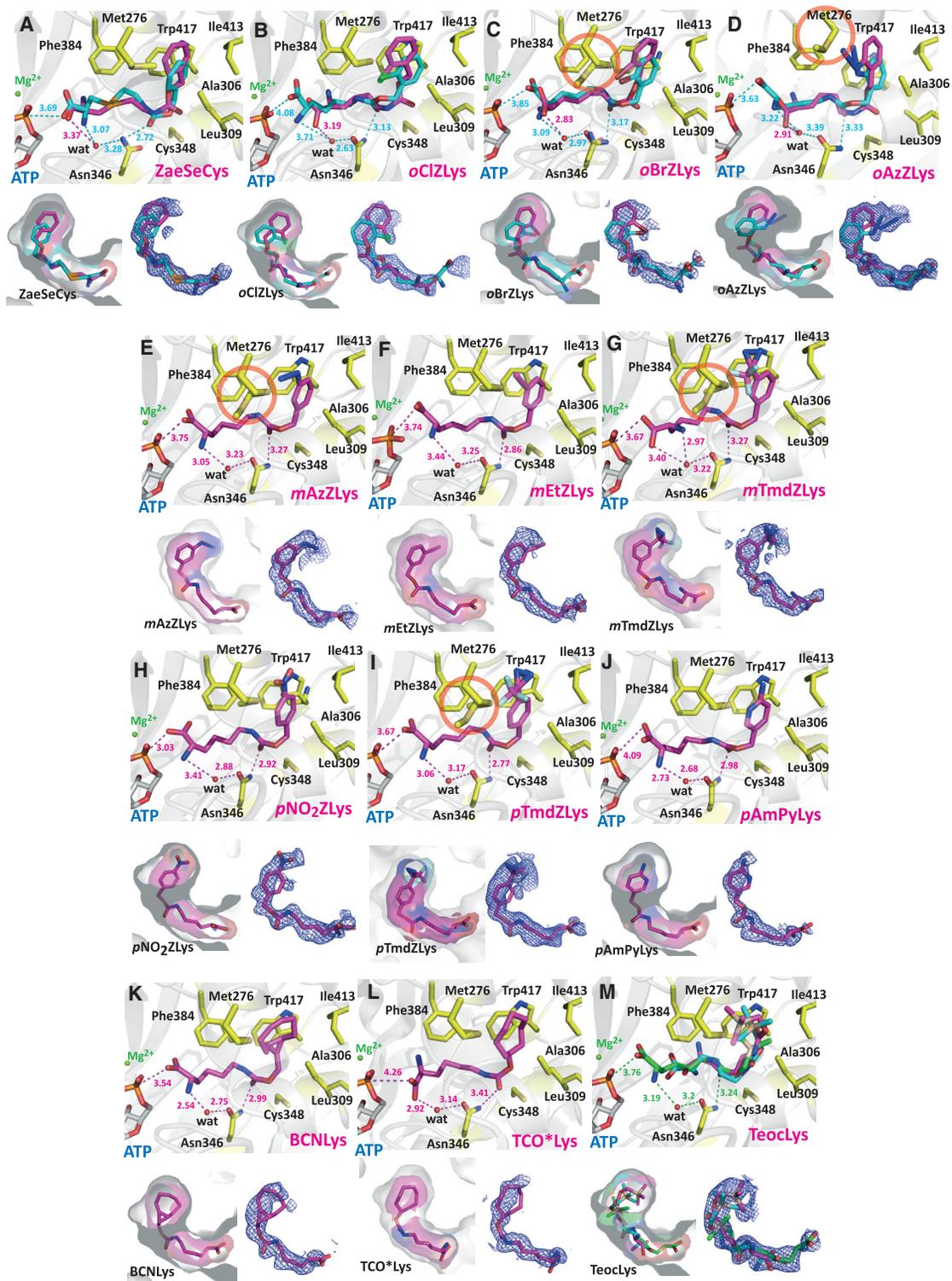


Figure 6. Crystal Structures of *M. mazei* PyIRSc(Y306A/Y384F) Complexed with Non-natural Lysine Derivatives

Close-up views of ZaeSeCys (A), oBrZLys (B), oClZLys (C), oAzZLys (D), mAzZLys (E), mEtZLys (F), mTmdZLys (G), pNO₂ZLys (H), pTmdZLys (I), pAmPyLys (J), BCNLys (K), TCO*Lys (L), and TeocLys (M) are shown (upper panels). The two different binding modes of ZaeSeCys, oBrZLys, oClZLys, and oAzZLys are shown as magenta and cyan stick models, and the three binding modes of TeocLys are shown as magenta, cyan, and green stick models. Water-mediated hydrogen bonds between the non-natural amino acids and the side-chain amide group of Asn348 are represented by dotted lines. Transparent ribbon models of PyIRSc(Y306A/Y384F) are visible in the background. Space-filling and stick models of the non-natural amino acids are shown on transparent surface models of *M. mazei* PyIRSc(Y306A/Y384F) (lower left panels). The $F_0 - F_c$ omit electron density maps (contoured at 1.5σ) of the non-natural amino acids in the PyIRSc(Y306A/Y384F) active site (lower right panels). See also Figures S5–S7, and Tables S4 and S5.

This amino acid also exhibits the double-binding modes, 1 and 2 (Figures 6A and S5), which are quite similar to those of ZLys (Figure 5). The $-\text{CH}_2-\text{CH}_2-\text{Se}-\text{CH}_2-$ moiety of ZaeSeCys is well accommodated in the site corresponding to the $-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-$ moiety of ZLys. Here, the occupancies of ZaeSeCys in modes 1 and 2 are 0.24 and 0.76, respectively.

Structures of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with *ortho*-Substituted ZLys Derivatives and ATP

Similar to ZLys and ZaeSeCys with the unsubstituted benzene ring, the *ortho*-chloro (oCl)-, *ortho*-bromo (oBr)-, and *ortho*-azide (oAz)-substituted ZLys derivatives, oClZLys, oBrZLys, and oAzZLys, respectively, also exhibited the double-binding mode (Figures 6B–6D and S5). In mode 2, the *p*-CH groups of these *ortho*-substituted ZLys derivatives are located in the positions corresponding to that of ZLys, as limited by the van der Waals contact with the δ 2-methyl group of Leu407 (Figures 6B–6D). In contrast, in mode 1, the carbonyl oxygen atom of the carbamate group of these amino acids hydrogen bonds with the Asn346 side chain, as with ZLys.

In both modes 1 and 2, the chlorine and bromine atoms of oClZLys and oBrZLys, respectively, and the azide group of oAzZLys, were modeled in one of the two *ortho* positions of the benzene ring (Figures 6B–D). This orientational preference is due to the putative steric conflicts of the chlorine and bromine atoms, and the azide group, in the other *ortho* position against the side chain of Leu309 in mode 1 and against the side chain of Ile413 in mode 2 (not shown). The chlorine and bromine atoms of oClZLys and oBrZLys, respectively, are similarly located in almost the same position between modes 1 and 2. Thus, both the chlorine and bromine atoms are surrounded, regardless of the binding modes, by three side-chain atoms, with the distances of the side-chain methyl group of Ala302 (3.3/3.2 and 3.3/3.3 Å in modes 1/2), the side-chain ϵ 1-CH of Phe384 (3.4/3.3 and 4.1/3.4 Å), and the ϵ -methyl group of Met276 (3.8/3.5 and 4.1/3.4 Å) from the chlorine and bromine atoms, respectively. The sums of the van der Waals radii of Cl and Br and that of a carbon are 3.45 and 3.55 Å, respectively. Therefore, the chlorine and bromine atoms are nearly within the van der Waals contact range with the surrounding side-chain atoms. In contrast, the position of the benzene ring differs largely between the two modes (Figures 6B and 6C).

In the case of oBrZLys, the Met276 side chain exhibits a mixture of the major and minor conformations, at a ratio of 0.58:0.42, about the C β -C γ bond, where the major one is the same as the Met276 side-chain conformation with ZLys and oClZLys. In the minor conformation, the ϵ -methyl group is rotated away from the bromine atom, and does not form a van der Waals contact with it. This minor conformation specific to oBrZLys may be due to the larger van der Waals radius of the bromine atom than that of the chlorine atom.

In the case of oAzZLys, the azide group of oAzZLys is surrounded by the three side-chain atoms, with the distances of the side-chain methyl group of Ala302 (3.4/3.2 Å in modes 1/2), the side-chain ϵ 1-CH of Phe384 (3.6/4.0 Å), and the δ -sulfur atom of Met276 (5.5/3.2 Å) from the azide group (Figure 6D). Intriguingly, the Met276 side chain assumes a single side-chain conformation, which is similar to the minor one with oBrZLys,

to create additional space in the active site. The ϵ -methyl group of Met276 is rotated away from the azide group of oAzZLys (mode 2) to evade steric hindrance with it.

Structure of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with the *meta*-Substituted ZLys Derivatives and ATP

The electron density revealed that the *meta*-substituted ZLys derivatives assume a single-binding mode (Figures 6 and S5). In the *mAzZLys*, *mEtZLys*, and *mTmdZLys*-bound forms, the distances between the carbonyl oxygen atom of the carbamate moiety and the side-chain amide group of Asn346 are 3.3, 2.9, and 3.3 Å, respectively (Figures 6E–6G). Therefore, this single-binding mode corresponds to mode 1 in the above-described double-binding mode for ZLys and its *ortho*-substituted derivatives.

The *mAzZ*, *mEtZ*, and *mTmdZ* groups are buried in a hydrophobic pocket, formed by Met276, Ala302, Ala306, Leu309, Phe384, Leu407, Asp408, Trp411, Ile413, and Trp417 (Figures 6E–6G and S5; Table S5). Even in mode 1, the substituents of these *meta*-substituted ZLys derivatives are located in the positions corresponding to the benzene ring of ZLys and its *ortho*-substituted derivatives in mode 2, because of the van der Waals contact with the δ 2-methyl group of Leu407. The *meta*-substituents in the *mAzZ*- and *mTmdZLys*-bound structures are favorably oriented relative to the benzene ring and avoid steric conflict with the side-chain carboxyl group of Asp408 (Figures 6E and 6G). The *meta*-substituents of *mAzZ*- and *mTmdZLys* extend closer to Met276 than that of *mEtZLys*, and thus they destabilize the major conformation of the Met276 side chain, and cause conformational mixtures with the major:minor ratios of 0.61:0.39 and 0.63:0.37, respectively (Figures 6E and 6G), similarly to oBrZLys (Figure 6C).

Structure of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with *para*-Substituted ZLys Derivatives and ATP

The *para*-substituted ZLys derivatives, *pNO₂ZLys* and *pTmdZLys*, assume a single-binding mode (mode 1) (Figures 6 and S5). The distances between the carbonyl oxygen atom of the carbamate moiety and the side-chain amide group of Asn346 are 2.9 and 2.8 Å for *pNO₂ZLys* and *pTmdZLys*, respectively (Figures 6H and 6I), indicating their hydrogen bonding in the above-described binding mode 1. The *pNO₂Z* and *pTmdZ* moieties are accommodated in the hydrophobic pocket in the same manner as the *meta*-substituted Z moieties (Figure S5, Table S5). First of all, the benzene rings of *pNO₂ZLys* and *pTmdZLys* are stacked with the indole ring of Trp417, similarly to those of ZLys and its derivatives. The benzene ring of *pNO₂ZLys* is located in a similar position to that of ZLys bound in mode 1, while the *pNO₂* group does not form any particular favorable interaction (Figures 6H and S5). The *pTmd* group of *pTmdZLys* is located in a position quite similar to that of *mTmdZLys*. In contrast, the benzene ring of *pTmdZLys* is rotated in the plane by about 26° relative to that of *mTmdZLys* (Figure S6), and is closer to Ile413 and Trp417 than to Ala302 and Phe384. Note that Ala302 and Phe384 are close to the benzene rings of both the *ortho* and *meta*-substituents of ZLys derivatives. In the *pTmdZLys*-bound structure, the Met276 side chain is close to the trifluoromethyl group, and exhibits a mixture of the major (0.51) and minor

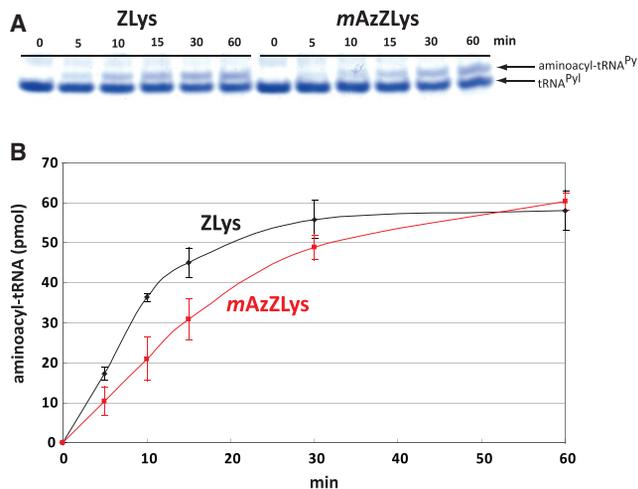


Figure 7. Time Course Analysis of the *In Vitro* Aminoacylation Reaction by *M. mazei* PyIRS(Y306A/Y384F)

The aminoacylation assay conditions are described in the STAR Methods. (A) Starting from the left, each lane shows a reaction with the following: PyIRS(Y306A/Y384F) with ZLys (0–60 min); PyIRS(Y306A/Y384F) with *mAzZLys* (0–60 min). (B) Time-course analysis of ZLys-tRNA^{Pyl} (black line) and *mAzZLys*-tRNA^{Pyl} (red line) formation catalyzed by PyIRS(Y306A/Y384F). The total amounts (pmol) of aminoacyl-tRNAs synthesized in the reaction mixture were calculated from the band intensities, as processed with the NIH ImageJ software, and are shown in the graphs. The plots represent mean values \pm standard deviation from three independent experiments.

(0.49) conformations, as described for the oBrLys-, *mAzZLys*-, and *mTmdZLys*-bound structures (Figures 6C, 6E, and 6G).

We failed to solve the structures of PyIRSc(Y306A/Y384F) bound to *pAzZLys* and *pEtZLys*, ZLys derivatives with a straight *para*-azide (*pAz*) and a *para*-ethynyl (*pEt*) group, respectively, which are longer than those of *pNO₂ZLys* and *pTmdZLys*. The accommodation of these long, straight substituents at the *para* position within the limited space of the binding pocket may shift the rest of the molecule relative to the ZLys moiety in mode 1. Such putative shifts of the ZLys moiety might correspond to the very low activities of PyIRS(Y306A/Y384F) for *pAzZLys* and *pEtZLys* (Figure 3).

Structures of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with the Lysine Derivative Containing the *para*-Aminopyridyl Moiety and ATP

For *pAmPyLys*, a lysine derivative with a *para*-aminopyridine scaffold instead of the *para*-substituted benzene ring (Figure 6J), the crystallization was performed in the presence of 50 mM *pAmPyLys*, because *M. mazei* PyIRS(Y306A/Y384F) exhibited very low aminoacylation activity with 1 mM *pAmPyLys* (Figure 3). The electron density of *pAmPyLys* revealed a single-binding mode, corresponding to mode 1, as the carbonyl oxygen atom in the carbamate moiety of *pAmPyLys* hydrogen bonds with the side-chain amide group of Asn346 (3.0 Å) (Figure 6J). The entire *pAmPyLys* molecule is shifted within the range of 1 Å, as compared with *pNO₂ZLys*. As a result, the *para*-amino group is located near the side chain of Asp408 (3.3 Å), suggesting a weak hydrogen bond, whereas the *pNO₂* group did not exhibit

any particular interactions. From the electron density, it was difficult to judge the position of the nitrogen atom in the pyridine ring (Figures 6J and S5). In either of the two possible positions, the ring nitrogen atom is surrounded by only hydrophobic amino acid residues. It might be presumed that the polar nature of the pyridine ring causes the lower affinity for the enzyme than those of the ZLys derivatives analyzed in the present study.

Structures of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with ATP and Lysine Derivatives Containing Bicyclooctyne and *trans*-Cyclooctene Groups

BCNLys and TCO**Lys* have the strained cycloalkyne and *trans*-cycloalkene moieties, respectively. Both of them assume a single-binding mode, in which the carbonyl oxygen atoms in the carbamate moiety of BCNLys and TCO**Lys* hydrogen bond with the side chain of Asn346 (3.0 and 3.4 Å, respectively) (Figures 6K and 6L). The ring structures of BCNLys and TCO**Lys* are quite different from the benzene ring structure, and cannot form π - π stacking interactions with the indole ring of Trp417, unlike the ZLys derivatives. However, the “thicker” shapes of the BCN and TCO* rings than the benzene ring are well accommodated between the hydrophobic Trp417 and Ala306 side chains (Figures 6K, 6L, S5, and S7; Table S5).

Structures of *M. mazei* PyIRS(Y306A/Y384F) Catalytic Fragment Complexed with a Lysine Derivative Containing a Trimethylsilyl Moiety and ATP

The bound TeocLys assumes three different binding modes, 1, 1.5, and 2 (Figures 6M and S5). The occupancies of TeocLys in modes 1, 1.5, and 2 are 0.29, 0.38, and 0.33, respectively. In mode 1, the carbonyl oxygen atom in the carbamate moiety of TeocLys hydrogen bonds with the side-chain amide group of Asn346 (3.2 Å), as in mode 1 of the above-described lysine derivatives. In mode 2, the Teoc group is accommodated most deeply inside the binding pocket, and the end reaches Leu407, as in mode 2 of the others (Figure S5; Table S5). The TeocLys in mode 1.5 is located between modes 1 and 2. Unlike mode 1, the carbamate moiety of TeocLys is farther away from Asn346 in modes 1.5 and 2, and the carbonyl oxygen interacts instead with the main-chain nitrogen atom of Ala306 in mode 1.5 (3.2 Å), and with the indole nitrogen atom of Trp417 in mode 2 (3.2 Å). The trimethylsilyl group of TeocLys is close to the Trp417 indole ring, but does not stack with it.

Analysis of *In Vitro* Aminoacyl-tRNA^{Pyl} Synthesis by *M. mazei* PyIRS(Y306A/Y384F)

The bound ZLys molecule exhibits two different binding modes, whereas the *mAzZLys* molecule assumes a single-binding mode in the *M. mazei* PyIRSc(Y306A/Y384F) structure. Although the affinities of *M. mazei* PyIRSc(Y306A/Y384F) for these two amino acids were too weak to measure by SPR, as described above, a prolonged reaction with a high concentration of PyIRS(Y306A/Y384F) produced the same levels of ZLys-tRNA^{Pyl} and *mAzZLys*-tRNA^{Pyl} (Figure 3). In contrast, a time course analysis using an *in vitro* aminoacylation assay revealed that the esterification of tRNA^{Pyl} by the PyIRS(Y306A/Y384F) mutant occurs more quickly with ZLys than *mAzZLys* in the initial 30 min of the reaction (Figure 7). The specific activities of aminoacylation

(pmol aminoacyl-tRNA^{Pyl} formed/pmol enzyme/min) for ZLys and *mAzZLys* by the PylRS(Y306A/Y384F) mutant are 0.165 and 0.095 min⁻¹, respectively. These results suggest that the lysine derivatives bound in some modes can be aminoacylated by PylRS(Y306A/Y384F), and that the single or double-binding mode does not matter, if the PylRS active site can accommodate a sufficient amount of the substrate amino acid.

DISCUSSION

In this study, we demonstrated the protein incorporation of non-natural lysine derivatives with the pair of *M. mazei* PylRS(Y306A/Y384F) and tRNA^{Pyl}, for 11 previously reported ones (ZLys, ZaeSeCys, *oClZLys*, *oAzZLys*, *mAzZLys*, *mTmdZLys*, *pAzZLys*, *pTmdZLys*, *pNO₂ZLys*, BCNLys, and TCO**Lys*) and five new ones (*oBrZLys*, *oEtZLys*, *mEtZLys*, *pEtZLys*, and TeoClys) (Figure 1), in the *E. coli* cell-based and/or cell-free systems. Our crystallographic and biochemical analyses revealed the structural mechanisms by which *M. mazei* PylRS(Y306A/Y384F) recognizes this large variety of non-natural lysine derivatives. The Y306A mutation (Yanagisawa et al., 2008b) was rationally designed to expand the amino acid binding pocket of *M. mazei* PylRS for the accommodation of ZLys with the benzene ring on top of the oxycarbonyllysine moiety (Figure 1), on the basis of the crystal structure of *M. mazei* PylRSc in complex with pyrrolysine and AMPPNP (Yanagisawa et al., 2008a).

The present crystal structures of PylRSc(Y306A/Y384F) in complex with the *meta*-substituted ZLys derivatives, *mAzZLys*, *mTmdZLys*, and *mEtZLys*, represent the standard binding mode. First, the hydrophobic benzene ring is stacked with the indole ring of Trp417. It should be noted here that the side chain of Ala306, which is introduced by the mutation of Tyr306, is located on the other side of Trp417, and the wild-type Tyr306 side chain must sterically clash with the benzene ring. Second, the *meta*-substituent of the benzene ring (one end of the side chain) is positioned through the van der Waals contact with the δ 2-methyl group of Leu407. The *meta*-substituents of *mAzZLys* and *mTmdZLys* sterically give rise to an alternative conformation of the side chain of Met276. Thus, the long side chain is snugly accommodated in the hydrophobic side chain binding pocket of PylRS(Y306A/Y384F). The lysine moiety extends toward the ATP-binding site, through the hydrogen bonding of the carbonyl group of the oxycarbonyllysine moiety with the side-chain amide group of Asn346, which defines “mode 1,” or the productive binding mode.

The binding modes of the *para*-substituted ZLys derivatives, *pTmdZLys* and *pNO₂ZLys*, in their crystal structures with PylRSc (Y306A/Y384F) are mode 1, according to the mode-defining hydrogen bond between the carbonyl group of the carbamate moiety and the Asn346 side chain. The benzene ring retains the stacking with the Trp417 indole ring, but is rotated and brings the *para*-substituent to a position similar to that of the *meta*-substituent in the above case. Thus, the *para*-substituent contacts the δ 2-methyl group of Leu407. The alternative conformation of Met276 was observed for both *pTmdZLys* and *mTmdZLys*. These structures have established how PylRS(Y306A/Y384F) achieves the snug accommodation of the two types of ZLys derivatives, by adapting to the positional differences of the *meta*- and *para*-substituents.

In contrast to these *meta*- and *para*-substituted derivatives, ZLys, ZaeSeCys, and the *ortho*-substituted ZLys derivatives, *oClZLys*, *oBrZLys*, and *oAzZLys*, exhibit the unusual double-binding mode (Figures 5, 6, and S6). One is mode 1, defined by the hydrogen bond between the carbamate carbonyl group and the Asn346 side chain. These amino acids with the unsubstituted and *ortho*-substituted benzene rings are shorter than the *meta*- and *para*-substituted ones along the axis of the side-chain binding pocket of PylRS(Y306A/Y384F). The benzene ring with no substituent at the *meta/para* position cannot reach the Leu407 wall at one end of the pocket, when the (*ortho*-substituted) ZLys is bound in mode 1. In contrast, in the other binding mode, the carbamate carbonyl and Asn346 amide groups are too far apart to hydrogen bond with each other. Instead, the benzene ring is largely shifted so that the *para*-CH contacts the Leu407 wall, which defines this alternative mode, designated as “mode 2.” Although these lysine derivatives with the double-binding mode exhibited much lower affinity for PylRSc(Y306A/Y384F) by the SPR analysis (Figure S4), ZLys is rapidly aminoacylated by PylRS(Y306A/Y384F) (Figure 7). It is very interesting that PylRS(Y306A/Y384F) accommodates such amino acids that are shorter than the *meta/para*-substituted ones in dynamic manners. As in the case of *mAzZLys*, *mTmdZLys*, and *pTmdZLys*, the alternative conformation of Met276 occurs with the bulky *oBrZLys* and even more significantly with the larger *oAzZLys*.

The shortest lysine derivative, TeoClys, exhibits the triple-binding mode, or an intermediate mode, 1.5, in addition to modes 1 and 2. In mode 1, TeoClys forms the hydrogen bond between the carbamate carbonyl and Asn346 amide groups, and is closest to the ATP-binding site. Unlike ZLys and its derivatives, TeoClys cannot form the π - π stacking interaction with the Trp417 indole ring. Instead, the trimethylsilyl group of TeoClys hydrophobically interacts with the indole ring, and reaches Leu407 in mode 2. Intriguingly, modes 2 and 1.5 are defined by specific hydrogen bonds with Trp417 and Ala306, respectively. Actually, the double-binding mode of a phenylalanine derivative in the PylRS structure has been reported (Guo et al., 2014). The alternative binding mode is non-productive, and is caused by new hydrogen bonds derived from the mutations of PylRS to enable the recognition of phenylalanine derivatives.

In contrast, BCNLys and TCO**Lys* exhibit the single-binding mode, which is mode 1 according to the carbamate carbonyl-Asn346 amide hydrogen bond, as with the *meta/para*-substituted ZLys derivatives. BCNLys and TCO**Lys*, like TeoClys, lack the π - π stacking with the Trp417 indole ring. Furthermore, the cyclooctyne and *trans*-cyclooctene moieties of BCNLys and TCO**Lys*, respectively, do not reach Leu407, a part of the wall. Instead, these eight-membered rings have a larger volume than the benzene ring, and can interact hydrophobically with the methyl group of Ala306, and simultaneously with the Trp417 indole ring (Figure S7). Therefore, the single-binding mode of BCNLys and TCO**Lys* is not caused by Leu407, but actually by Ala306.

Thus, we have established the precise structural bases of the accommodation of a wide range of non-natural lysine derivatives in the amino acid binding site of *M. mazei* PylRS(Y306A/Y384F). In particular, our finding of the double- and triple-binding modes in addition to the conventional single-binding mode, was quite intriguing. When coupled with more enzyme kinetics data, the

present structural bases will be useful for the design of novel functional amino acids as substrates of *M. mazei* PylRS (Y306A/Y384F).

Further structure-based engineering of *M. mazei* PylRS (Y306A/Y384F) should enable the more efficient incorporation of non-natural amino acids and the expansion of the genetic code with larger or bulkier amino acids. The mutated positions of *M. mazei* and/or *M. barkeri* PylRSs for non-natural amino acid incorporation are summarized in a review by Wan et al. (2014). In addition to the listed active-site residues, Met276, Leu301, A302, Leu305, Tyr306, Tyr309, Asn346, Cys348, Met350, Tyr384, Val401, and Trp417 (Wan et al., 2014), we propose the mutagenesis of Ile405, Leu407, Trp411, and Ile413, which form the hydrophobic pocket in the active site of PylRS (the numbering is for *M. mazei* PylRS), for the incorporation of a larger variety of amino acids. We also suggest the mutations of Gly378-Thr387, which constitute half of the $\beta 7$ - $\beta 8$ hairpin of PylRS, to overcome the current size limitations of non-natural amino acids. For instance, lysine derivatives longer than p TmdZLys cannot be accommodated with straight side chains, because of the steric hindrance with the main chains constituting the active site, but hopefully would bind to the pocket expanded by these proposed mutations by assuming a bent conformation.

In addition, we have demonstrated that the cell-free protein synthesis method is practically useful, with many advantages for incorporating non-natural amino acids over the *E. coli* cell-based system, such as excluding the negative factors in the cell membrane permeability of non-natural amino acids, achieving high productivities of non-natural amino acid-containing proteins comparable with those of ordinary cell-based recombinant systems, consuming smaller quantities of non-natural amino acids for incorporation than the cell-based system, and synthesizing toxic proteins and membrane proteins containing non-natural amino acids that cannot be produced in the cell-based system (Seki et al., 2018).

SIGNIFICANCE

Genetic-code expansion has become a useful technology to incorporate a non-natural amino acid site specifically into a target protein for structural and functional analyses. Archaeal pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl} are extensively used for genetic-code expansion. The *Methanosarcina mazei* PylRS mutant bearing the Y306A and Y384F mutations, PylRS(Y306A/Y384F), has broad substrate specificity, and is one of the most widely used enzymes for the genetic encoding of a variety of large non-natural lysine derivatives, as well as the naturally occurring post-translationally modified lysines, at UAG codons. In this study, we analyzed the mechanisms underlying the broad amino acid specificity of PylRS(Y306A/Y384F). First, 17 non-natural lysine derivatives, including various functional groups for click chemistry, photo-crosslinking, etc., were compared with respect to *in vivo* protein incorporation and *in vitro* aminoacylation of tRNA^{Pyl}. N^ε-(benzyloxycarbonyl)lysine (ZLys), and 10 ZLys derivatives with a substitution at the *ortho*, *meta*, or *para* position of the benzene ring were efficiently ligated to tRNA^{Pyl} and incorporated into proteins by PylRS (Y306A/Y384F). We determined the crystal structures of the

PylRS(Y306A/Y384F) catalytic fragments complexed with 14 non-natural lysine derivatives, to clarify the structural basis for the broad substrate specificity. The *meta*- and *para*-substituted ZLys derivatives are snugly accommodated in the binding pocket of PylRS(Y306A/Y384F), which represents the productive mode (mode 1). In contrast, ZLys and the unsubstituted or *ortho*-substituted ZLys derivatives exhibited a double-binding mode: an alternative, non-productive binding mode in addition to the productive mode. The double-binding mode of PylRS(Y306A/Y384F) hardly affects aminoacylation as the aminoacylation rate is high for ZLys. These precise structural mechanisms of the recognitions of the substrate lysine derivatives by PylRS(Y306A/Y384F) may facilitate the structure-based rational design of novel useful non-natural amino acids for genetic-code expansion. We also demonstrated the usefulness of cell-free protein synthesis for the incorporation of an amino acid that is poorly incorporated *in vivo*.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCES SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - *E. coli* Cell-Based Protein Synthesis and Purification of the T7-GST Proteins Containing Non-natural Amino Acids
 - Preparation of *M. mazei* tRNA^{Pyl} and PylRS(Y306A/Y384F)
 - Aminoacylation Assay
 - Surface Plasmon Resonance Analysis of the Binding between *M. mazei* PylRS(Y306A/Y384F) and Lysine Derivatives
 - Cell-free Protein Synthesis and Purification of GFP Proteins Containing Non-natural Amino Acids
 - Cloning, Expression, and Purification of the C-terminal Catalytic Fragment of *M. mazei* PylRS with the Y306A and Y384F Mutations
 - Crystallization, Data Collection, and Structure Determination
 - Chemical Syntheses of Non-natural Amino Acids
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.chembiol.2019.03.008>.

ACKNOWLEDGMENTS

We would like to thank Dr. Kunio Hirata (RIKEN Harima), Dr. Yoshiaki Kawano (RIKEN Harima), and the staff of the beamlines BL26B2, BL32XU, and BL41XU at SPring-8 (Harima, Japan). We would like to thank Dr. Takaho Terada, Takako Imada, and Tomoko Nakayama for clerical assistance. We are grateful to Drs. Seisuke Kusano, Misaki Moriya, Tomomi Sumida, and Shunsuke Tagami (RIKEN) for the BIAcore analysis. We thank Kaori Ohtsuki, Masaya

Usui, and Aya Abe (RIKEN BSI) for the MS analysis. We thank Dr. Ryohei Ishii for assistance with the structural analysis. We thank Dr. Seiji Takashima (Osaka University) for providing mTmdZlys. This work was supported by Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research [BINDS]) from AMED, Japan under grant no. JP17am0101081 and by The Uehara Memorial Foundation (SY). T.Y. was also supported by a grant from MEXT, Japan (grant no. 24550203).

AUTHOR CONTRIBUTIONS

T.Y. and S.Y. conceived and designed the experiments. M.K. performed the crystallization experiments and crystallographic analyses. T.Y. performed the biochemical experiments and crystallographic analyses. E.S. performed the cell-free protein synthesis experiments. M.K., T.Y., E.S., N.H., K.S., and S.Y. analyzed the data. M.K., T.Y., and S.Y. wrote the paper.

DECLARATION OF INTERESTS

T.Y., N.H., K.S., and S.Y. are co-inventors on related patents to this work. S.Y. is a founder and shareholder of LiberoThera.

Received: August 24, 2018

Revised: December 25, 2018

Accepted: March 15, 2019

Published: April 25, 2019

REFERENCES

- 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. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *D66*, 213–221.
- Ai, H.W., Shen, W., Sagi, A., Chen, P.R., and Schultz, P.G. (2011). Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *ChemBioChem* *12*, 1854–1857.
- Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C., and Söll, D. (2007). Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U S A* *104*, 3141–3146.
- Blackman, M.L., Royzen, M., and Fox, J.M. (2008). Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* *130*, 13518–13519.
- 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.
- Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J.M., Wiessler, M., Schultz, C., van Hest, J.C., van Delft, F.L., and Lemke, E.A. (2012). Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation. *ChemBioChem* *13*, 2094–2099.
- Brabham, R., and Fascione, M.A. (2017). Pyrrolysine amber stop-codon suppression: development and applications. *ChemBiochem* *18*, 1973–1983.
- 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.* *48*, 4052–4055.
- Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U S A* *99*, 11020–11024.
- 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.
- Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* *83*, 379–408.
- Chin, J.W. (2017). Expanding and reprogramming the genetic code. *Nature* *550*, 53–60.
- Chumpolkulwong, N., Hori-Takemoto, C., Hosaka, T., Inaoka, T., Kigawa, T., Shirouzu, M., Ochi, K., and Yokoyama, S. (2004). Effects of *Escherichia coli* ribosomal protein S12 mutations on cell-free protein synthesis. *Eur. J. Biochem.* *271*, 1127–1134.
- Crnković, A., Suzuki, T., Söll, D., and Reynolds, N.M. (2016). Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat. Chem. Acta* *89*, 163–174.
- Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* *D50*, 760–763.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* *35*, W375–W383.
- Elliott, T.S., Bianco, A., and Chin, J.W. (2014). Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal. *Curr. Opin. Chem. Biol.* *21*, 154–160.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* *D60*, 2126–2132.
- Englert, M., Nakamura, A., Wang, Y.S., Eiler, D., Söll, D., and Guo, L.T. (2015). Probing the active site tryptophan of *Staphylococcus aureus* thioredoxin with an analog. *Nucleic Acids Res.* *43*, 11061–11067.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* *347*, 203–206.
- Flügel, V., Vrabel, M., and Schneider, S. (2014). Structural basis for the site-specific incorporation of lysine derivatives into proteins. *PLoS One* *9*, e96198.
- Ge, Y., Fan, X., and Chen, P.R. (2016). A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. *Chem. Sci.* *7*, 7055–7060.
- Guo, L.T., Wang, Y.S., Nakamura, A., Eiler, D., Kavran, J.M., Wong, M., Kiessling, L.L., Steitz, T.A., O'Donoghue, P., and Söll, D. (2014). Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. U S A* *111*, 16724–16729.
- Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* *296*, 1462–1466.
- Herring, S., Ambrogelly, A., Gundllapalli, S., O'Donoghue, P., Polycarpo, C.R., and Söll, D. (2007). The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett.* *581*, 3197–3203.
- 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.
- Hino, N., Oyama, M., Sato, A., Mukai, T., Iraha, F., Hayashi, A., Kozuka-Hata, H., Yamamoto, T., Yokoyama, S., and Sakamoto, K. (2011). Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells. *J. Mol. Biol.* *406*, 343–353.
- Ibba, M., and Söll, D. (2000). Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* *69*, 617–650.
- Jiang, R., and Krzycki, J.A. (2012). PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. *J. Biol. Chem.* *287*, 32738–32746.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* *D66*, 125–132.
- Katayama, H., Nozawa, K., Nureki, O., Nakahara, Y., and Hojo, H. (2012). Pyrrolysine analogs as substrates for bacterial pyrrolysyl-tRNA synthetase in vitro and in vivo. *Biosci. Biotechnol. Biochem.* *76*, 205–208.
- Kato, A., Kuratani, M., Yanagisawa, T., Ohtake, K., Hayashi, A., Amano, Y., Kimura, K., Yokoyama, S., Sakamoto, K., and Shiraishi, Y. (2017). Extensive survey of antibody invariant positions for efficient chemical conjugation using expanded genetic codes. *Bioconjug. Chem.* *28*, 2099–2108.

- Kavran, J., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T.A. (2007). Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U S A* *104*, 11268–11273.
- 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. U S A* *99*, 9715–9720.
- Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., and Yokoyama, S. (2004). Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* *5*, 63–68.
- Kiick, K.L., Saxon, E., Tirrell, D.A., and Bertozzi, C.R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U S A* *99*, 19–24.
- Kita, A., Hino, N., Higashi, S., Hirota, K., Narumi, R., Adachi, J., Takafuji, K., Ishimoto, K., Okada, Y., Sakamoto, K., et al. (2016). Adenovirus vector-based incorporation of a photo-cross-linkable amino acid into proteins in human primary cells and cancerous cell lines. *Sci. Rep.* *6*, 36946.
- Kolb, H.C., Finn, M.G., and Sharpless, K.B. (2001). Click Chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* *40*, 2004–2021.
- Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D.J., Blackman, M.L., Fox, J.M., and Chin, J.W. (2012). Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J. Am. Chem. Soc.* *134*, 10317–10320.
- Lang, K., and Chin, J.W. (2014a). Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* *114*, 4764–4806.
- Lang, K., and Chin, J.W. (2014b). Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* *9*, 16–20.
- Lee, M.M., Jiang, R., Jain, R., Larue, R.C., Krzycki, J., and Chan, M.K. (2008). Structure of *Desulfotobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* *374*, 470–474.
- Lee, Y.J., Schmidt, M.J., Tharp, J.M., Weber, A., Koenig, A.L., Zheng, H., Gao, J., Waters, M.L., Summerer, D., and Liu, W.R. (2016). Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader. *Chem. Commun. (Camb.)* *52*, 12606–12609.
- Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* *79*, 413–444.
- 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.
- Mukai, T., Yanagisawa, T., Ohtake, K., Wakamori, M., Adachi, J., Hino, N., Sato, A., Kobayashi, T., Hayashi, A., Shirouzu, M., et al. (2011). Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* *411*, 757–761.
- 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.
- Nguyen, D.P., Lusic, H., Neumann, H., Kapadnis, P.B., Deiters, A., and Chin, J.W. (2009). Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* *131*, 8720–8721.
- Nguyen, T.A., Cigler, M., and Lang, K. (2018). Expanding the genetic code to study protein-protein interactions. *Angew. Chem. Int. Ed.* *57*, 14350–14361.
- Nikić, I., Plass, T., Schraidt, O., Szymański, J., Briggs, J.A., Schultz, C., and Lemke, E.A. (2014). Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem. Int. Ed.* *53*, 2245–2249.
- Nozawa, K., O'Donoghue, P., Gundllapalli, S., Arais, Y., Ishitani, R., Umehara, T., Söll, T., and Nureki, O. (2009). Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* *457*, 1163–1167.
- Plass, T., Milles, S., Koehler, C., Szymański, J., Mueller, R., Wiessler, M., Schultz, C., and Lemke, E.A. (2012). Amino acids for Diels-Alder reactions in living cells. *Angew. Chem. Int. Ed.* *51*, 4166–4170.
- Polcarpo, C., Ambrogelly, A., Berube, A., Winbush, S., McCloskey, J., Crain, P., Wood, J., and Söll, D. (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U S A* *101*, 12450–12454.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* *252*, 1682–1689.
- 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.
- Schmidt, M.J., Weber, A., Pott, M., Welte, W., and Summerer, D. (2014). Structural basis of furan-amino acid recognition by a polyspecific aminoacyl-tRNA-synthetase and its genetic encoding in human cells. *ChemBioChem* *15*, 1755–1760.
- Schneider, S., Gattner, M.J., Vrabel, M., Flügel, V., López-Carrillo, V., Prill, S., and Carell, T. (2013). Structural insights into incorporation of norbornene amino acids for click modification of proteins. *ChemBioChem* *14*, 2114–2118.
- Seki, E., Matsuda, N., Yokoyama, S., and Kigawa, T. (2008). Cell-free protein synthesis system from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation. *Anal. Biochem.* *377*, 156–161.
- Seki, E., Yanagisawa, T., and Yokoyama, S. (2018). Cell-free protein synthesis for site-specific incorporation of non-canonical amino acids using cell extract from RF-1 deficient *Escherichia coli* strains. *Methods Mol. Biol.* *1728*, 49–65.
- Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* *296*, 1459–1462.
- Suzuki, T., Miller, C., Guo, L.T., Ho, J.M.L., Bryson, D.I., Wang, Y.S., Liu, D.R., and Söll, D. (2017). Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* *13*, 1261–1266.
- Takimoto, J.K., Dellas, N., Noel, J.P., and Wang, L. (2011). Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* *6*, 733–743.
- Tharp, J.M., Ehnborn, A., and Liu, W.R. (2018). tRNA^{Pyl}: structure, function, and applications. *RNA Biol.* *15*, 441–452.
- Uttamapinant, C., Tangpeerachaikul, A., Grecian, S., Clarke, S., Singh, U., Slade, P., Gee, K.R., and Ting, A.Y. (2012). Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew. Chem. Int. Ed.* *51*, 5852–5856.
- Vargas-Rodriguez, O., Sevostyanova, A., Söll, D., and Crnković, A. (2018). Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol.* *46*, 115–122.
- Virdee, S., Kapadnis, P.B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D.P., Riechmann, L., and Chin, J.W. (2011). Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* *133*, 10708–10711.
- Wan, W., Tharp, J.M., and Liu, W.R. (2014). Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* *1844*, 1059–1070.
- Wang, L., and Schultz, P.G. (2001). A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* *8*, 883–890.
- Wang, L., Brock, A., and Schultz, P.G. (2002). Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.* *124*, 1836–1837.
- Wang, L., Xie, J., and Schultz, P.G. (2006). Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* *35*, 225–249.
- Wang, L. (2017). Engineering the genetic code in cells and animals: biological considerations and impacts. *Acc. Chem. Res.* *50*, 2767–2775.
- Wang, Z.U., Wang, Y.S., Pai, P.J., Russell, W.K., Russell, D.H., and Liu, W.R. (2012). A facile method to synthesize histones with posttranslational modification mimics. *Biochemistry* *51*, 5232–5234.
- Yamaguchi, A., Matsuda, T., Ohtake, K., Yanagisawa, T., Yokoyama, S., Fujiwara, Y., Watanabe, T., Hohsaka, T., and Sakamoto, K. (2016). Incorporation of a doubly functionalized synthetic amino acid into proteins

- for creating chemical and light-induced conjugates. *Bioconjug. Chem.* **27**, 198–206.
- Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O., and Yokoyama, S. (2006). Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **F62**, 1031–1033.
- Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008a). Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **378**, 634–652.
- Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008b). Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197.
- Yanagisawa, T., Hino, N., Iraha, F., Mukai, T., Sakamoto, K., and Yokoyama, S. (2012). Wide-range protein photo-crosslinking achieved by a genetically encoded N(epsilon)-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Mol. Biosyst.* **8**, 1131–1135.
- Yanagisawa, T., Sumida, T., Ishii, R., and Yokoyama, S. (2013). 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. D Biol. Crystallogr.* **D69**, 5–15.
- Yanagisawa, T., Takahashi, M., Mukai, T., Sato, S., Wakamori, M., Shirouzu, M., Sakamoto, K., Umehara, T., and Yokoyama, S. (2014a). Multiple site-specific installations of N(epsilon)-monomethyl-L-lysine into histone proteins by cell-based and cell-free protein synthesis. *ChemBioChem* **15**, 1830–1838.
- Yanagisawa, T., Umehara, T., Sakamoto, K., and Yokoyama, S. (2014b). Expanded genetic code technologies for modified lysine installation at multiple sites. *ChemBioChem* **15**, 2181–2187.
- Yokoyama, S., Sakamoto, K., Yanagisawa, T., and Kobayashi, T. (2010). Mutant pyrrolysyl-tRNA synthetase, and method for production of protein having non-natural amino acid integrated therein by using the same. US Patent 20100267087 A1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GST antibody	GE Healthcare	Cat#27-4577-01; RRID: AB_771432
donkey anti-goat IgG-HRP	Santa Cruz Biotechnology	Cat#sc-2354; RRID: AB_628490
Bacterial and Virus Strains		
<i>E. coli</i> BL21-Gold(DE3)	Agilent Technologies Inc.	Cat#230132
<i>E. coli</i> BL21(DE3)	Agilent Technologies Inc.	Cat#230130
<i>E. coli</i> Rosetta(DE3)	Novagen	Cat#70954-3CN
Chemicals, Peptides, and Recombinant Proteins		
ZLys	Bachem	Cat#E-1702
ZaeSeCys	Sundia/Namiki	N/A
<i>N</i> ^ε -Boc-oBrZLys	Bachem	Cat#A-1415
oClZLys	Bachem	Cat#E-2725
oAzZLys	Shinsei Chemical	N/A
oEtZLys	Sundia/Namiki	N/A
<i>m</i> AzZLys	Sundia/Namiki	N/A
<i>m</i> EtZLys	Sundia/Namiki	N/A
<i>m</i> TmdZLys	Sundia, Shinsei Chemical, and Dr. Takashima (Osaka Univ.)	N/A
<i>p</i> NO ₂ ZLys	Bachem	Cat#E-2960
<i>p</i> AzZLys	Sundia/Namiki	N/A
<i>p</i> EtZLys	Sundia/Namiki	N/A
<i>p</i> TmdZLys	Sundia, and Shinsei Chemical	N/A
<i>p</i> AmPyLys	Shinsei Chemical	N/A
BCNLys	Synnafix	Cat#SX-A2011
TCO*Lys	SciChem	Cat#SC-8008
<i>N</i> ^ε -Fmoc-TeocLys	Advanced ChemTech	Cat#FK2387
ATP	Sigma-Aldrich	Cat#A2383
GTP	Sigma-Aldrich	Cat#G8877
CTP	Sigma-Aldrich	Cat#C1506
UTP	Sigma-Aldrich	Cat#U6875
GMP	Sigma-Aldrich	Cat#G8377
Pyrophosphatase	Sigma-Aldrich	Cat#10108987001
T7 RNA polymerase	Yanagisawa et al., 2008a	N/A
HEPES	Hampton Research	Cat#HR2-729
Tris(hydroxymethyl)aminomethane (Tris)	Nacalai	Cat#35401-25
Sodium citrate	Nacalai	Cat#6132-04-3
PEG200	Hampton Research	Cat#HR2-601
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma-Aldrich	Cat#C4706
Dithiothreitol (DTT)	Nacalai	Cat#14112-52
Thrombin	GE Healthcare	Cat#27084601
Spermidine trihydrochloride	Sigma-Aldrich	Cat#S2501
β-D-thiogalactopyranoside (IPTG)	Carbosynth	Cat#EI05931
Bovine Serum Albumin (BSA)	Sigma-Aldrich	Cat#A9418
L-Glutathione reduced form	Nacalai	Cat#17050-72
β-mercaptoethanol	Nacalai	Cat#21417-65

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cOmplete EDTA-free Protease Inhibitor Cocktail	Roche (Sigma-Aldrich)	Cat#11873580001
PrimeSTAR Max DNA Polymerase	TAKARA	Cat#R045A
Deposited Data		
<i>M. mazei</i> PylRSc(Y306A/Y384F)/mAzZLys complex	This study	PDB: 6AAC
<i>M. mazei</i> PylRSc(Y306A/Y384F)/mTmdZLys complex	This study	PDB: 6AAD
<i>M. mazei</i> PylRSc(Y306A/Y384F)/mEtZLys complex	This study	PDB: 6AAN
<i>M. mazei</i> PylRSc(Y306A/Y384F)/TCO*Lys complex	This study	PDB: 6AAO
<i>M. mazei</i> PylRSc(Y306A/Y384F)/ZaeSeCys complex	This study	PDB: 6AAP
<i>M. mazei</i> PylRSc(Y306A/Y384F)/BCNLys complex	This study	PDB: 6AAQ
<i>M. mazei</i> PylRSc(Y306A/Y384F)/pNO ₂ ZLys complex	This study	PDB: 6AAZ
<i>M. mazei</i> PylRSc(Y306A/Y384F)/pAmPyLys complex	This study	PDB: 6AB0
<i>M. mazei</i> PylRSc(Y306A/Y384F)/oAzZLys complex	This study	PDB: 6AB1
<i>M. mazei</i> PylRSc(Y306A/Y384F)/oClZLys complex	This study	PDB: 6AB2
<i>M. mazei</i> PylRSc(Y306A/Y384F)/ZLys complex	This study	PDB: 6AB8
<i>M. mazei</i> PylRSc(Y306A/Y384F)/TeocLys complex	This study	PDB: 6ABK
<i>M. mazei</i> PylRSc(Y306A/Y384F)/oBrZLys complex	This study	PDB: 6ABL
<i>M. mazei</i> PylRSc(Y306A/Y384F)/pTmdZLys complex	This study	PDB: 6ABM
<i>M. mazei</i> PylRSc/pyrrolysyl-AMP complex	Kavran et al., 2007	PDB: 2ZIM
Oligonucleotides		
For preparing <i>M. mazei</i> tRNA ^{Pyl} template		
AAGCTTAATACGACTCACTATAGGAAACCT	Yanagisawa et al., 2008a	N/A
TGGCGGAAACCCCGGAATCTAACCCGGCTGA	Yanagisawa et al., 2008a	N/A
For cloning <i>M. mazei</i> PylRS(Y306A/Y384F)		
AGGAGATATACCATGGATAAAAAACCACTAAACACTCTGATATC	This study	N/A
CGTGTACACGAGCTCTTACAGTTGGTAGAAATCCCCTTATAGT	This study	N/A
TCTACCAACCTGTAAGAGCTCGTGTACACGGCGCGCCTGCA	This study	N/A
TAGTGGTTTTTATCCATGGTATATCTCCTTATTAAGTT	This study	N/A
Recombinant DNA		
pET-GST(25Am)	Yanagisawa et al., 2008b	N/A
pCDF-Pyl-AFx2	This study	pCDF-Pyl-Fx1 derivative, Yanagisawa et al., 2014a
pCR2.1-N11GFPS1(17Am)	Seki et al., 2008	N/A
pET28-PylRSc(Y306A/Y384F)	This study	N/A
pET28-PylRS(Y306A/Y384F)	Yanagisawa et al., 2008b	N/A
pMINOR	Chumpolkulwong et al., 2004	N/A
pUC19-tRNA ^{Pyl}	Yanagisawa et al., 2008a	N/A
Software and Algorithms		
Phenix 1.9	Adams et al., 2010	https://www.phenix-online.org/
Coot ver. 8	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
XDS ver. November 3, 2014	Kabsch, 2010	http://xds.mpimf-heidelberg.mpg.de/
ccp4 ver. 6.5	Collaborative Computational Project Number 4, 1994	http://www.ccp4.ac.uk/
Pymol v0.99	Schrodinger, LLC	http://pymol.sourceforge.net/
ImageJ	NIH	https://imagej.nih.gov/ij/index.html
Molprobit	Davis et al., 2007	http://molprobit.biochem.duke.edu/
Other		
Ni-Sepharose High Performance	GE Healthcare	Cat#17526801
HisTrap HP Column	GE Healthcare	Cat#17524701

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HiLoad 16/600 Superdex 200 pg Column	GE Healthcare	Cat#28989335
Glutathione Sepharose High Performance	GE Healthcare	Cat#17527901
Resource S Column	GE Healthcare	Cat#17118001
Resource Q Column	GE Healthcare	Cat#17117901
Sensor Chip CM5	GE Healthcare	Cat#BR100399
HBS-P running buffer	GE Healthcare	Cat#BR100368
N-hydroxysuccinimide (NHS)/1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) amine coupling kit	GE Healthcare	Cat#BR100050
Click-iT Plus AlexaFluor488 Picolyl Azide Toolkit	Thermo Scientific	Cat#C10641
In Gel Tryptic Digestion Kit	Thermo Scientific	Cat#89871
Immobilon Western Chemiluminescent HRP Substrate	Merck	Cat#WBKLS0500
SimplyBlue SafeStain	Thermo Scientific	Cat#LC6065
Pierce Coomassie Plus (Bradford) Assay Kit	Thermo Scientific	Cat#23236

CONTACT FOR REAGENT AND RESOURCES SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts Shiyuki Yokoyama (yokoyama@riken.jp) and Tatsuo Yanagisawa (tatsuo.yanagisawa@riken.jp).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No animals or cell lines have been used in this work.

METHOD DETAILS

Abbreviations: ZLys: N^{ϵ} -benzyloxycarbonyl-L-lysine; ZaeSeCys: N^{ϵ} -benzyloxycarbonyl-L-aminoethylselenocysteine; pNO_2 ZLys: N^{ϵ} -(p -nitrobenzyloxycarbonyl)-L-lysine; oBrZLys: N^{ϵ} -(o -bromobenzyloxycarbonyl)-L-lysine; oClZLys: N^{ϵ} -(o -chlorobenzyloxycarbonyl)-L-lysine; oAzZLys: N^{ϵ} -(o -azidobenzyloxycarbonyl)-L-lysine; mAzZLys: N^{ϵ} -(m -azidobenzyloxycarbonyl)-L-lysine; pAzZLys: N^{ϵ} -(p -azidobenzyloxycarbonyl)-L-lysine; oEtZLys: N^{ϵ} -(o -ethynylbenzyloxycarbonyl)-L-lysine; mEtZLys: N^{ϵ} -(m -ethynylbenzyloxycarbonyl)-L-lysine; pEtZLys: N^{ϵ} -(p -ethynylbenzyloxycarbonyl)-L-lysine; mTmdZLys: N^{ϵ} -(m -trifluoromethyl diazirinylbenzyloxycarbonyl)-L-lysine; pTmdZLys: N^{ϵ} -(p -trifluoromethyl diazirinylbenzyloxycarbonyl)-L-lysine; BCNLys: N^{ϵ} -(((1R,8S)-bicyclo[6.1.0]non-4-yn-9-yl)methoxy)carbonyl)-L-lysine; TCO* Lys , N^{ϵ} -(((E)-cyclooct-2-en-1-yl)oxy)carbonyl)-L-lysine; pAmPyLys: N^{ϵ} -(p -aminopyridylmethoxy)carbonyl)-L-lysine; and TeocLys, N^{ϵ} -(2-(trimethylsilyl)ethoxycarbonyl)-L-lysine.

Biochemical and molecular biological procedures were performed with commercially available materials, enzymes, and chemicals. ZLys, pNO_2 ZLys, and oClZLys were purchased from Bachem (Switzerland). oAzZLys, and pAmPyLys were purchased from Shinsei Chemical (Osaka, Japan). mAzZLys, pAzZLys, oEtZLys, mEtZLys, pEtZLys, and ZaeSeCys were purchased from Sundia (China). The pTmdZLys and mTmdZLys used for protein incorporation were purchased from Shinsei Chemical and Sundia. The mTmdZLys used for structural analysis was a kind gift from Dr. Seiji Takashima (Osaka University). Chemical syntheses of pAzZLys, oEtZLys, mEtZLys, pEtZLys, ZaeSeCys, mTmdZLys, and pAmPyLys are described in the [STAR Methods](#). BCNLys was purchased from Synaffix (Netherlands). TCO* Lys was purchased from SciChem (United Kingdom). N^{α} -Boc-oBrZLys was purchased from Bachem, and the Boc group was deprotected with 50% TFA to prepare oBrZLys. N^{α} -Fmoc-TeocLys was purchased from Advanced Chemtech (USA), and the Fmoc group was deprotected with 20% piperidine to prepare TeocLys. Matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and electrospray ionization mass spectrometry (ESI-MS) were performed using TOF/TOF5800 (AB SCIEX) and QSTAR ELITE (ABI) spectrometers, respectively. 1H NMR spectra were recorded on Bruker Avance III 300 MHz and Bruker Fourier 400 MHz spectrometers and tetramethylsilane (TMS) was used as the internal standard. LC/MS analyses were performed with an Agilent LC/MSD 1200 Series quadrupole Mass Spectrometer (Column: ODS 2000 (50 \times 4.6 mm, 5 μ m) operating in the ES (+) or (-) ionization mode; T = 30°C; flow rate = 1.5 ml/min; detected wavelength: 270 nm).

E. coli Cell-Based Protein Synthesis and Purification of the T7-GST Proteins Containing Non-natural Amino Acids

The site-specific incorporation of non-natural lysine derivatives at position 25 of GST in *E. coli* cells was performed as described previously ([Yanagisawa et al., 2008b](#)), with some modifications. The plasmids pET-GST(25Am), containing the T7 peptide-tagged

glutathione S-transferase (T7-GST) gene with an amber (UAG) codon at position 25, and pCDF-Pyl-AFx2, containing two copies of the PylRS(Y306A/Y384F) gene and three copies of the tRNA^{Pyl} gene, were co-transformed into *E. coli* BL21-Gold(DE3) cells. When the cells attained an OD₆₀₀ of 0.5 at 37°C, protein expression was induced by 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) at 37°C for 5 hr, in the presence of 1 mM non-natural amino acids. Protein purification and peptide mass fingerprinting (PMF) of the T7-GST proteins, containing non-natural lysine derivatives, were performed as described previously (Yanagisawa et al., 2008b; Yamaguchi et al., 2016). The harvested cells were disrupted in 50 mM potassium phosphate buffer (pH 7.4), containing 0.2 M NaCl, 10 mM β-mercaptoethanol, and a Protease Inhibitor Cocktail (cOmplete EDTA-free, Roche) (buffer G), and the lysate was centrifuged to remove the cell debris. The supernatants were loaded on a Glutathione Sepharose HP column (GE Healthcare). The column was washed with buffer G, and the proteins were eluted with buffer G, containing 100 mM Tris-HCl (pH 8.0) and 40 mM reduced glutathione. The purified T7-GST proteins were subjected to SDS-PAGE, followed by staining with SimplyBlue SafeStain (Thermo Scientific). The protein bands were excised and subjected to reduction and alkylation, followed by tryptic digestion with an In-Gel Tryptic Digestion Kit (Thermo Scientific), and the tryptic fragments were analyzed by MALDI-TOF MS spectrometry to confirm the incorporation of the non-natural acids into the protein. The protein concentrations were measured with a Pierce Coomassie Plus (Bradford) Assay Kit (Thermo Scientific). The *E. coli* strains and plasmids used in this study are listed in the [Key Resources Table](#).

Preparation of *M. mazei* tRNA^{Pyl} and PylRS(Y306A/Y384F)

The *M. mazei* tRNA^{Pyl} transcript was prepared by *in vitro* transcription, as described previously (Yanagisawa et al., 2008a). The transcription reaction was performed at 37°C for 3.5 h, in buffer containing 80 mM HEPES/KOH (pH 8.1), 20 mM MgCl₂, 40 mM KCl, 20 mM dithiothreitol (DTT), 2 mM spermidine, 14 μg/ml BSA, 20 mM GMP, 5 mM of each NTP, 0.08 u/ml pyrophosphatase (Sigma-Aldrich), 0.1 mg/ml purified T7 RNA polymerase, and the PCR-amplified template DNA containing the tRNA^{Pyl} gene with the T7 promoter. The DNA fragment containing the *M. mazei* tRNA^{Pyl} gene was amplified by PCR, using a pUC19-based plasmid containing the tRNA^{Pyl} gene as the template, and used as the template for *in vitro* transcription. The primer (5'-AAGCTTAATACGACT CACTATAGGAAACCT-3') and the primer complementary to the 3' end of tRNA^{Pyl} (5'-TGGCGGAAACCCCGGGAATCTAACCCGGC TGAACGGA-3') were used for PCR. After phenol/chloroform treatment and isopropanol precipitation, the tRNA^{Pyl} transcript was purified by Resource Q column chromatography (GE Healthcare). The purified tRNA^{Pyl} fractions were ethanol precipitated, dissolved in 10 mM Tris-HCl buffer (pH 7.5) containing 5 mM MgCl₂, and stored at -80°C. Before use, tRNA^{Pyl} was heated at 80°C for 2 min and gradually cooled to room temperature for refolding.

M. mazei PylRS(Y306A/Y384F) was overproduced in *E. coli* BL21-Gold(DE3) cells and purified as described previously (Yanagisawa et al., 2008b), with some modifications. Protein expression was induced by an overnight treatment with 1 mM IPTG at 20°C. The cells were harvested and disrupted in 50 mM potassium phosphate buffer (pH 7.4), containing 0.5 M NaCl, 25 mM imidazole, 5 mM β-mercaptoethanol, and a Protease Inhibitor Cocktail (cOmplete EDTA-free, Roche), and the lysate was centrifuged to remove the cell debris. The PylRS(Y306A/Y384F) protein was purified by two column chromatography steps, using HisTrap (GE Healthcare) and Resource S (GE Healthcare) resins. The purified PylRS(Y306A/Y384F) fractions were dialyzed against 20 mM potassium phosphate buffer (pH 7.4) containing 0.3 M KCl and 10 mM β-mercaptoethanol, concentrated, flash-cooled in liquid nitrogen, and stored at -80°C. Protein concentrations were determined using a Pierce Coomassie Plus (Bradford) Assay Kit (Thermo Scientific).

Aminoacylation Assay

The aminoacylation assay by acidic urea PAGE was performed as described previously (Yanagisawa et al., 2008b). The tRNA aminoacylation reactions were incubated at 37°C for 1 hr. The standard aminoacylation assay solution (20 μl) contained 5.2 μM purified *M. mazei* PylRS(Y306A/Y384F), 10 mM MgCl₂, 2 mM ATP, 4 mM DTT, 5.9 μM *M. mazei* tRNA^{Pyl} transcript, and 1 mM non-natural lysine derivatives, in 100 mM HEPES-NaOH buffer (pH 7.2). For ZaeSeCys, 50 mM TCEP (Tris(2-carboxyethyl) phosphine, pH 7.0) was added to the reaction buffer. The time course of the aminoacylation reaction (20 μl total volume) was measured in the presence of 1.1 μM PylRS(Y306A/Y384F), 10.8 μM tRNA^{Pyl}, and 1 mM ZLys or mAzZLys for 0, 5, 10, 15, 30, and 60 min at 37°C. Unaminoacylated and aminoacylated tRNA^{Pyl}s were subjected to electrophoresis on a 10% denaturing urea polyacrylamide gel under acidic conditions (pH 5.0) at 4°C for 18 hr, and were stained with 0.2% toluidine blue in 7.5% acetic acid.

Surface Plasmon Resonance Analysis of the Binding between *M. mazei* PylRSc(Y306A/Y384F) and Lysine Derivatives

To detect the interactions between *M. mazei* PylRSc(Y306A/Y384F) and lysine derivatives, BIAcore sensor chips were coated with PylRSc(Y306A/Y384F) and then exposed to a concentration gradient of the lysine derivatives. The measurements were performed using a BIAcore T200 system (GE Healthcare) at 20°C in HBS-P running buffer (10 mM HEPES-NaOH (pH 7.4),

0.15 M NaCl, 0.005% v/v Surfactant P20). The purified His-tagged PylRSc(Y306A/Y384F) (0.4 mg/ml) was immobilized onto a CM5 sensor chip using an *N*-hydroxysuccinimide (NHS)/1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) amine coupling kit, in 10 mM sodium acetate buffer (pH 5.0). The protein density was ~8,800 response units (RU). To determine the binding affinities of His-tagged PylRSc(Y306A/Y384F) and the lysine derivatives, solutions of the compound at different concentrations were injected into the PylRSc(Y306A/Y384F) immobilized chambers. The response units (RUs) were measured at four to six different concentrations of lysine derivatives (0.1–1 mM of *m*TmdZLys, 0.1–2 mM of *m*EtZLys, 0.1–5 mM of BCNLys, 0.1–5 mM of TCO**Lys*, 0.1–5 mM of TeocLys, and 0.1–5 mM of *o*EtZLys). The highest concentration of each amino acid was limited by its water solubility.

Cell-free Protein Synthesis and Purification of GFP Proteins Containing Non-natural Amino Acids

Cell-free coupled transcription/translation was performed as described previously (Kigawa et al., 2004; Mukai et al., 2011; Yanagisawa et al., 2014b; Seki et al., 2018), using pCR2.1-TOPO bearing an N11-tagged superfold type green fluorescent mutant protein (GFPS1) gene (Seki et al., 2008). The pCR2.1-N11GFPS1 plasmids containing the wild-type N11-GFPS1 gene or the mutant with a single UAG codon at Ala17 were used as the template DNAs for cell-free protein synthesis with S30 extracts from *E. coli* BL21(DE3) cells with a pMINOR plasmid encoding rare codon tRNAs (Chumpolkulwong et al., 2004). 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(Y306A/Y384F), 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 a plate fluorescence reader ARVO Victor2 V Multilabel Counter (PerkinElmer) (Seki et al., 2018). The purification of the N11-GFPS1 proteins was performed as follows. After centrifugation of the solution, the supernatant fractions were loaded on a Ni-Sepharose column (GE Healthcare). The column was washed with 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 20 mM imidazole, and 1 mM DTT, and then eluted with 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 500 mM imidazole, and 1 mM DTT. The PMF analysis of the N11-GFPS1 proteins, containing ZLys, *m*EtZLys, *m*AzZLys, *p*AmPyLys, *p*TmdZLys, and *m*TmdZLys, was performed in the same manner as for the T7-GST proteins, using MALDI-TOF MS and ESI-MS spectrometries.

Cloning, Expression, and Purification of the C-terminal Catalytic Fragment of *M. mazei* PylRS with the Y306A and Y384F Mutations

The DNA fragment encoding the *M. mazei* PylRS C-terminal domain (residues 185–454) with the Y306A and Y384F mutations was PCR-amplified and cloned into the pET28c vector. The *E. coli* Rosetta(DE3) strain (Novagen) was transformed with the plasmid and selected on an LB agar plate supplemented with 50 μg/ml kanamycin and 20 μg/ml chloramphenicol. A single colony was grown at 37°C in a broth culture containing 30 g tryptone, 10 g yeast extract, and 5 g NaCl per liter, supplemented with 30 μg/ml kanamycin. When the OD₆₀₀ reached 0.1, the cultivation temperature was lowered to 22°C. The protein expression was induced with 0.1 mM IPTG when the OD₆₀₀ reached 0.7, and the cells were cultivated overnight. The *E. coli* cells were collected by centrifugation and stored at –80°C.

The cells were resuspended in 50 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 20 mM imidazole, 1% (v/v) Tween-20, and 1 mM DTT, and were lysed by sonication on ice. The cell lysate was centrifuged at 15,000 × *g* for 20 min at 4°C, filtered through a 0.45 μm membrane, and fractionated on a HisTrap column (GE Healthcare), which was equilibrated with 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 20 mM imidazole, and 1 mM DTT. The protein was eluted by a linear gradient (20–500 mM) of imidazole, collected, and dialyzed against 20 mM Tris-HCl buffer (pH 8.0), containing 100 mM NaCl, 20 mM imidazole, and 1 mM DTT. The histidine-tag peptide derived from the pET28c vector was cleaved by thrombin protease (Sigma-Aldrich). The flow-through of the HisTrap column, equilibrated with 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl, 20 mM imidazole, and 1 mM DTT, was collected and concentrated by ultracentrifugation. The protein was filtered through a 0.22 μm membrane and fractionated on a HiLoad 16/60 Superdex 200 pg column (GE Healthcare), which was equilibrated with 20 mM HEPES-NaOH buffer (pH 7.4), containing 300 mM NaCl, 20 mM MgCl₂, and 1 mM DTT. The eluted fractions were collected and concentrated by ultracentrifugation to 20 mg/ml. Aliquots of the protein were flash-cooled in liquid nitrogen and stored at –30°C.

Crystallization, Data Collection, and Structure Determination

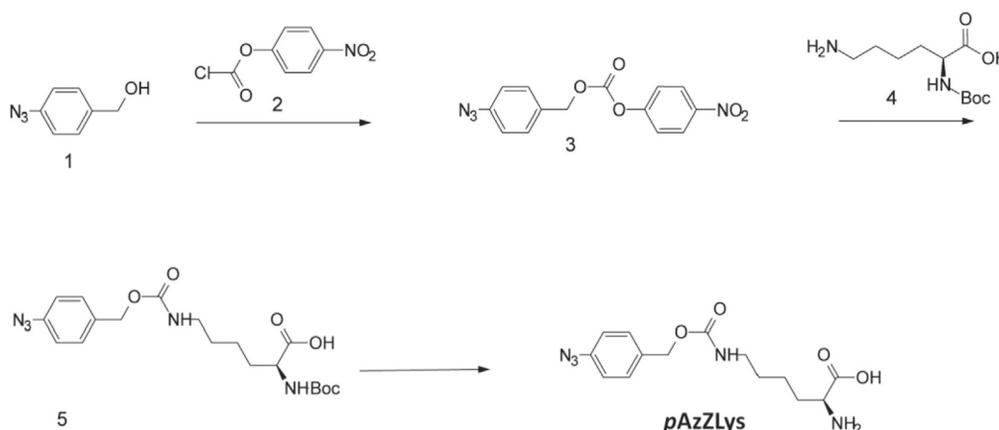
In the present study, we cleaved the hexahistidine-tag with thrombin after the tag-based purification, and found that the tag-free PylRSc(Y306A/Y384F) could be crystallized under different conditions from those reported previously for PylRSc with the hexahistidine-tag (Yanagisawa et al., 2006). The crystals of PylRSc(Y306A/Y384F) grew in the presence of each lysine derivative and ATP under the same conditions, with PEG200 as the precipitant.

Prior to crystallization, the protein was diluted to 1 mg/ml and mixed with 2 mM ATP and 10 mM substrate amino acid, and was incubated at 4°C for one hour. The sample was concentrated to 15 mg/ml by ultracentrifugation. Crystals were grown at 8°C

by the sitting drop vapor diffusion method. One microliter of the protein solution was mixed with an equal volume of the precipitant solution, composed of 100 mM Tris-HCl buffer (pH 8.5), 40% PEG200, 200 mM KCl, 40 mM MgCl₂, and 10 mM sodium citrate. The crystal was mounted on a nylon loop and flash-cooled in liquid nitrogen. The X-ray datasets were collected at the beamline BL32XU at SPring-8 (Harima, Japan) at -173°C and were processed with XDS (Kabsch, 2010). The crystal of PylRSc(Y306A/Y384F)•ZLys belongs to the space group C2, with unit cell parameters of $a=102.13$ Å $b=43.89$ Å $c=62.07$ Å and $\beta=98.8^\circ$. The crystal of PylRSc(Y306A/Y384F)•*m*AzZLys belongs to the space group C2, with unit cell parameters of $a=101.6$ Å $b=43.5$ Å $c=72.1$ Å and $\beta=118.8^\circ$. The phase was calculated by the molecular replacement method with Phaser, using 2ZIM as the search model. One PylRSc(Y306A/Y384F) molecule was found per asymmetric unit, with the solvent content of 45%. The diffraction data up to 1.57 Å were used for the model refinement for PylRSc(Y306A/Y384F)•ZLys. Iterative cycles of model refinement by PHENIX (Adams et al., 2010) and manual model building with Coot (Emsley and Cowtan, 2004) were performed. The R_{work} and R_{free} factors for the PylRSc(Y306A/Y384F) structures complexed with non-natural amino acids are shown in Table S4. The final model was validated with Molprobit (Davis et al., 2007) and Procheck (Collaborative Computational Project Number 4, 1994). The final models consist of residues 188–280 and 284–454 of PylRSc(Y306A/Y384F). The electron densities revealed that the ATP did not react with the lysine derivative, and is bound to three magnesium ions. Graphical images were prepared with PyMOL [<http://pymol.sourceforge.net/>]. The data collection and refinement statistics are summarized in Table S4.

Chemical Syntheses of Non-natural Amino Acids

Scheme 1. Synthesis of *p*AzZLys

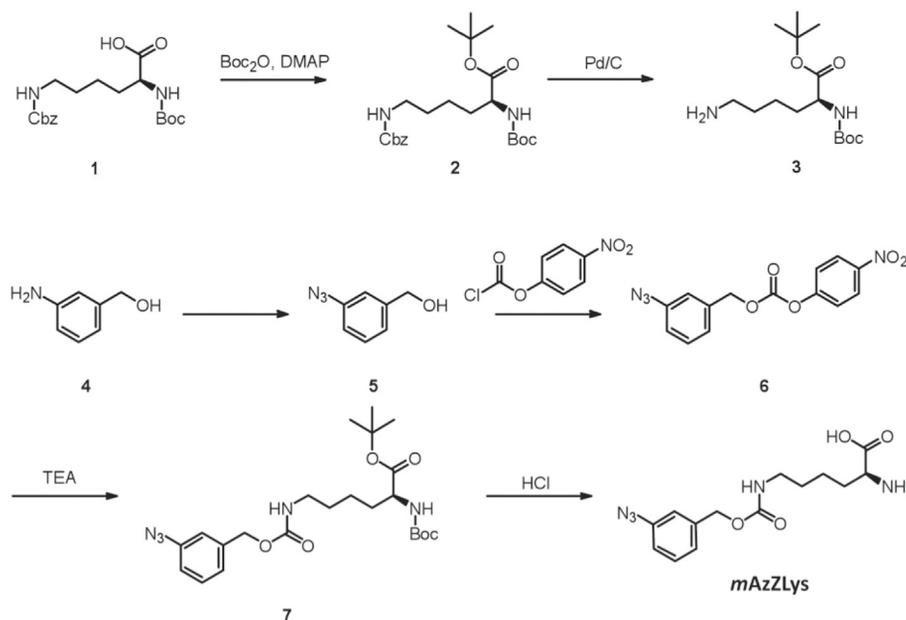


1.1 Synthesis of (4-azidophenyl)methyl(4-nitrophenoxy)formate (**3**): To a solution of compound **1** (0.88 g, 5.87 mmol) in dichloromethane (DCM) (10 ml), pyridine (1 ml) and compound **2** (1.18 g, 5.87 mmol) was added, at 25°C under N₂ overnight. The reaction mixture was poured into water (8 ml) and extracted with DCM (20 ml). The organic layer was washed with water (6 ml) and brine (10 ml), dried over Na₂SO₄ and concentrated to give the crude product **3** (0.85 g, yield: 50%) as a yellow solid, which was directly used for the next step.

1.2 Synthesis of (2S)-6-[[[4-azidophenyl)methoxy]carbonylamino]-2-[(*tert*-butoxy) carbonylamino]hexanoic acid (**5**): To a solution of compound **3** (0.85 g, 2.5 mmol) in tetrahydrofuran (THF) (10 ml), *N,N*-diisopropylethylamine (DIEA) (0.64 g, 5 mmol) and 6-amino-2-*tert*-butoxycarbonylamino-hexanoic acid (0.62 g, 2.5 mmol) were added. The mixture was stirred at 25°C under N₂ overnight. The mixture was diluted with 4 N HCl for adjustment to pH = 3 and extracted with DCM (10 ml × 3). The combined organic layer was washed with water (5 ml) and brine (5 ml), dried over Na₂SO₄ and concentrated. The residue was purified by silica gel column chromatography (DCM/MeOH = 40/1) to afford the product **5** (500 mg, yield: 40%) as yellow oil.

1.3 Synthesis of (2S)-2-amino-6-[[[4-azidophenyl)methoxy]carbonylamino]hexanoic acid (*p*AzZLys): Compound **5** (450 mg, 18 mmol) was suspended in HCl/EtOAc (5 ml) and stirred for 2 hr at room temperature. The mixture was concentrated under reduced pressure, and the residue was purified by prep-HPLC to afford the desired product *p*AzZLys (120 mg, yield: 50%) as an off-white solid. ¹H NMR (D₂O, 300 MHz): δ 7.30 (d, $J = 6.3$ Hz, 2H), 7.00 (d, $J = 6.3$ Hz, 2H), 4.95 (s, 2H), 3.08 (m, 1H), 3.00 (m, 2H), 1.38 (m, 4H), 1.19 (m, 2H). LC-MS [mobile phase: from 80% water (0.1% NH₄OH) and 20% CH₃CN to 5% water (0.1% NH₄OH) and 95% CH₃CN in 6 min, and finally under these conditions for 0.5 min.] The purity was >95%, the retention time = 3.120 min, and MS Calcd.: 321; MS Found: 322 ([M+H]⁺).

Scheme 2. Synthesis of *mAzZLys*



2.1 The solution of compound **1** (50 g, 0.13 mol), Boc_2O (44 g, 0.20 mol), and DMAP (2.4 g, 0.02 mol) in tert-butanol (500 ml) was stirred at room temperature for 5 hrs. The mixture was poured into ice water (500 ml) and extracted with ethyl acetate. The organic layer was dried over anhydrous sodium sulfate and concentrated to give the crude extract, which was purified by silica gel column chromatography to afford compound **2** (40 g, yield: 70%) as yellow oil.

2.2 Synthesis of compound **3**: The solution of compound **2** (40 g, 92 mmol) in MeOH (400 ml) was treated with Pd/C (8.0 g). The reaction mixture was hydrogenated under 50 psi at room temperature overnight, and then filtered through diatomaceous earth under nitrogen. The filtrate was evaporated in a vacuum to afford compound **3** (26 g, yield: 94%) as yellow oil.

2.3 Synthesis of compound **5**: To a solution of compound **4** (40 g, 0.33 mol) in CH_3CN (300 ml), *tert*-Butyl nitrite (36 g, 0.35 mol) was added dropwise at 5°C . After stirring at 5°C for 15 min, a solution of TMSN_3 (40 g, 0.35 mol) in CH_3CN (50 ml) was added to the solution at 5°C . The resulting mixture was stirred at room temperature overnight, and quenched with water (500 ml), and extracted with DCM. The organic layer was dried over anhydrous sodium sulfate and concentrated to give the crude extract, which was purified by silica gel column chromatography to afford compound **5** (46 g, yield: 95%) as yellow oil.

3.4 Synthesis of compound **6**: A solution of compound **5** (30 g, 0.20 mol) and TEA (25 g, 0.25 mol) in DCM (300 ml) was mixed with 4-nitrophenyl carbonylchloridate (44 g, 0.22 mol) at 0°C . The mixture was stirred at room temperature overnight, quenched with water (300 ml), and extracted with DCM. The organic layer was dried over anhydrous sodium sulfate and concentrated to give the crude extract, which was purified by silica gel column chromatography to afford compound **6** (32 g, yield: 51%) as yellow oil.

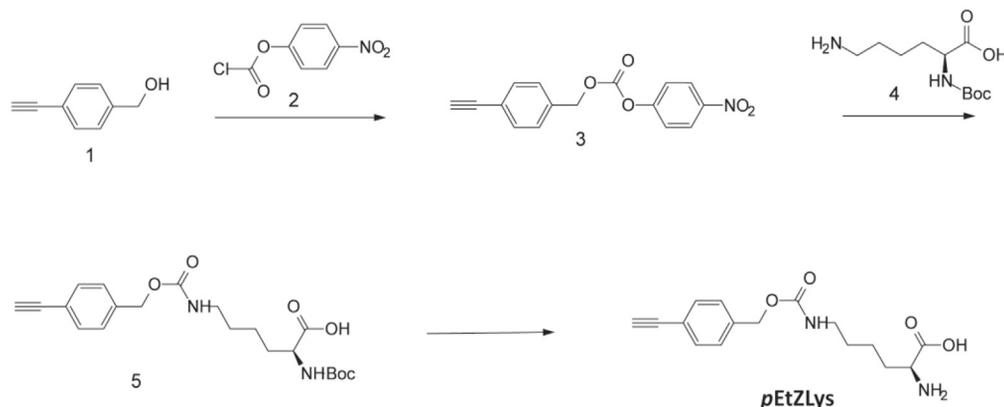
3.5 Synthesis of compound **7**: The mixture of compound **6** (18 g, 57 mmol), compound **3** (17 g, 57 mmol), and TEA (8.1 g, 80 mmol) in DCM (200 ml) was stirred at room temperature for 3 hrs. The mixture was then quenched with water (200 ml) and extracted with DCM. The organic layer was dried over anhydrous sodium sulfate and concentrated to give the crude extract, which was purified by silica gel column chromatography to afford compound **7** (24 g, yield: 88%) as yellow oil.

3.6 Synthesis of (2*S*)-2-amino-6-[[3-azidophenyl]methoxy]carbonylamino]hexanoic acid (*mAzZLys*): The mixture of compound **7** (24 g, 50.3 mmol) in $\text{HCl}/\text{Et}_2\text{O}$ (3*N*, 150 ml) was stirred at room temperature overnight. The mixture was concentrated and diluted with water (100 ml), and the aqueous layer was adjusted to pH 5 with sat. NaHCO_3 aq. The precipitate was filtered, and filtrate cake was dried in a vacuum to afford *mAzZLys* (12 g, yield: 74%) as a white solid.

$^1\text{H-NMR}$ (400 MHz, CD_3OD): 1.46-1.61 (m, 4 H), 1.89-1.99 (m, 2 H), 3.15-3.18 (t, 2 H), 3.95-3.99 (t, 1 H), 5.09 (s, 2 H), 7.01-7.07 (m, 2 H), 7.15-7.18 (d, 1 H), 7.37-7.41 (t, 1 H).

LC-MS [mobile phase: from 90% water (0.05% $\text{NH}_3\cdot\text{H}_2\text{O}$) and 10% CH_3CN (0.05% $\text{NH}_3\cdot\text{H}_2\text{O}$) to 5% water (0.05% $\text{NH}_3\cdot\text{H}_2\text{O}$) and 95% CH_3CN (0.05% $\text{NH}_3\cdot\text{H}_2\text{O}$) in 6.0 min, finally under these conditions for 0.5 min.] The purity was 97.0%, $R_t = 3.215$ min, and MS Calcd.: 321; MS Found: 322 ($[\text{M}+\text{H}]^+$).

Scheme 3. Synthesis of pEtZLys

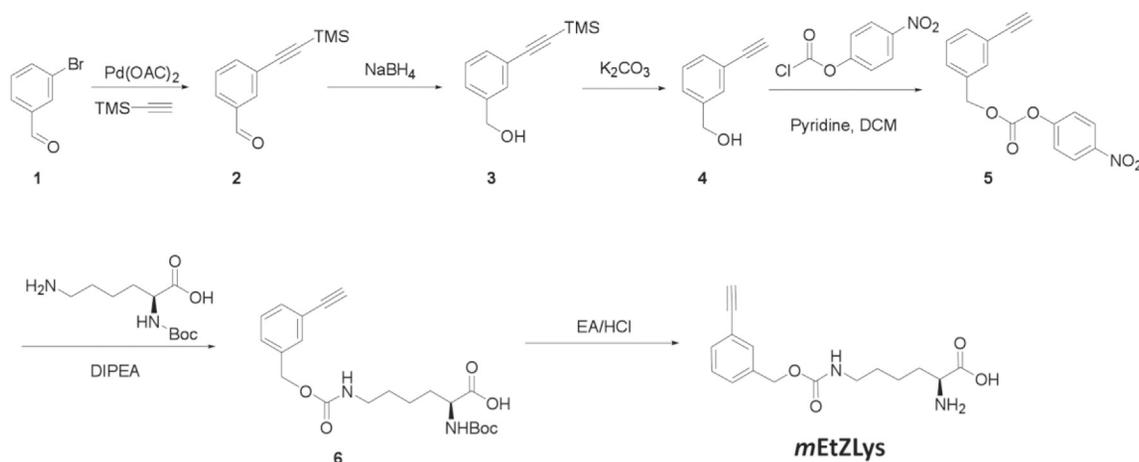


3.1 Synthesis of carbonic acid 4-ethynyl-benzyl ester 4-nitro-phenyl ester (**3**): To a solution of compound **1** (0.78 g, 5.87 mmol) in DCM (10 ml), pyridine (1 ml) and compound **2** (1.18 g, 5.87 mmol) were added at 25°C under N₂, and then the reaction was stirred at room temperature overnight. The reaction mixture was poured into water (8 ml) and extracted with DCM (2 × 20 ml). The organic layer was washed with water (6 ml) and brine (10 ml), dried over Na₂SO₄, and concentrated to give the crude product **3** (0.75 g, yield: 50%) as a yellow solid, which was directly used for the next step.

3.2 Synthesis of 2-*tert*-butoxycarbonylamino-6-(4-ethynyl-benzyl oxycarbonylamino)-hexanoic acid (**5**): To a solution of compound **3** (0.75 g, 2.5 mmol) in THF (10 ml), DIEA (0.64 g, 5 mmol) and 6-amino-2-*tert*-butoxycarbonylamino-hexanoic acid (0.62 g, 2.5 mmol) were added. The mixture was stirred at 25°C under N₂ overnight. The mixture was diluted with 4 N HCl to adjust the pH to 3 and extracted with DCM (10 ml × 3). The combined organic layer was washed with water (5 ml) and brine (5 ml), dried over Na₂SO₄, and concentrated. The residue was purified by silica gel column chromatography (DCM/MeOH = 40/1) to afford product **5** (400 mg, yield: 40%) as yellow oil.

3.3 Synthesis of 2-amino-6-(4-ethynyl-benzyl oxycarbonylamino)-hexanoic acid (**pEtZLys**): The mixture of compound **5** (0.3 g, 18 mmol) in HCl/EtOAc (5 ml) was stirred for 2 hr at room temperature. The mixture was concentrated under reduced pressure and the residue was purified by prep-HPLC to afford the desired product **pEtZLys** (140 mg, yield: 67%) as a white solid. ¹H NMR (DMSO-*d*₆, 300 MHz): δ 7.47 (d, *J* = 5.7 Hz, 2H), 7.34 (d, *J* = 6.0 Hz, 2H), 7.27 (m, 1H), 5.02 (s, 2H), 4.18 (s, 1H), 3.20 (m, 1H), 2.96 (m, 2H), 1.68 (m, 2H), 1.32 (m, 4H). LC-MS [mobile phase: from 80% water (0.02% NH₄OAc) and 20% CH₃CN to 5% water (0.02% NH₄OAc) and 95% CH₃CN in 6 min, finally under these conditions for 0.5 min.] The purity was > 95%, the retention time = 2.598 min, and MS Calcd.: 304; MS Found: 305 ([M+H]⁺).

Scheme 4. Synthesis of mEtZLys



4.1 Synthesis of 3-trimethylsilyl ethynyl-benzaldehyde (**2**): To a solution of aldehyde **1** (20.0 g, 108 mmol) in triethylamine (TEA) (400 ml), triphenylphosphine (PPh₃) (1.0 g, 3.8 mmol), ethynyltrimethylsilane (60.0 g, 594 mmol), and Pd(OAc)₂ (0.2 g, 0.9 mmol) were

added under N₂. The resulting mixture was stirred at 100°C for 6 hr. After the mixture was cooled to room temperature, it was filtered and washed with DCM. The filtrate was washed with water, dried over Na₂SO₄, and concentrated to afford the crude product **2**, which was purified by column chromatography to afford compound **2** (20 g, 96%) as brown oil.

¹H NMR (DMSO-*d*₆, 400 MHz): δ 9.83 (s, 1H), 7.81-7.78 (m, 1H), 7.73 (d, *J* = 8.0 Hz, 1H), 7.58 (d, *J* = 8.0 Hz, 1H), 7.45-7.40 (m, 1H), 0.08 (s, 9H).

4.2 Synthesis of (3-trimethylsilylanylethynyl-phenyl)-methanol (**3**): To a solution of 3-trimethylsilylanylethynyl-benzaldehyde (20 g, 105 mmol) in EtOH (250 ml), NaBH₄ (12 g, 321 mmol) was added at 0°C. The resulting mixture was stirred at 0°C for 2 min and was quenched with an NH₄Cl solution. The mixture was extracted with DCM (150 ml × 3), and dried over Na₂SO₄. The solution was concentrated to dryness to afford the desired crude product **3** (20 g), as brown oil, which was directly used for next step.

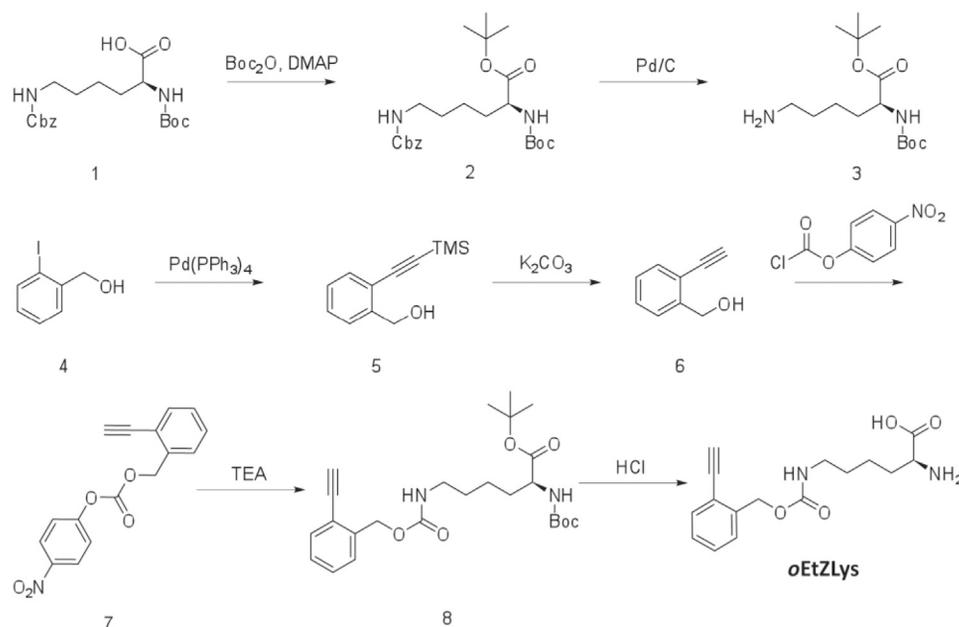
4.3 Synthesis of (3-ethynyl-phenyl)-methanol (**4**): To a solution of (3-trimethylsilylanylethynyl-phenyl)-methanol (**3**) (20 g, 105 mmol) in MeOH (400 ml), K₂CO₃ (29 g, 210 mmol) was added. The mixture was stirred overnight at 25°C under N₂. The mixture was filtered, and the filtrate was concentrated to remove most of the organic solvent. The residue was purified by silica gel column chromatography (DCM/petroleum ether = 1/2) to afford product **4** (7 g, 50% yield for 2 steps) as brown oil. ¹H-NMR (DMSO-*d*₆, 400 MHz): δ 7.23-7.20 (s, 1H), 7.16-7.10 (m, 3H), 5.05 (s, 1H), 4.30-4.26 (m, 2H), 3.91 (s, 1H).

4.4 Synthesis of carbonic acid 3-ethynyl-benzyl ester 4-nitro-phenyl ester (**5**): To a solution of compound **4** (5.4 g, 40.6 mmol) in DCM (100 ml), pyridine (10 ml) and 4-nitrophenyl carbonochloridate (8.2 g, 40.6 mmol) were added, and the mixture was incubated overnight at 25°C under N₂. The reaction mixture was poured into water (80 ml) and extracted with DCM (200 ml). The organic layer was washed with water (60 ml) and brine (100 ml), dried over Na₂SO₄, and concentrated to give the crude product **5** (12 g, yield: 100%) as a yellow solid, which was directly used for next step.

4.5 Synthesis of 2-*tert*-butoxycarbonylamino-6-(3-ethynyl-benzoyloxycarbonylamino)-hexanoic acid (**6**): A solution of compound **5** (7.5 g, 25 mmol) in THF (100 ml) was mixed with DIEA (6.4 g, 50 mmol) and 6-amino-2-*tert*-butoxycarbonylamino-hexanoic acid (6.2 g, 25 mmol). The mixture was stirred overnight at 25°C under N₂. The mixture was diluted with 4 N HCl to adjust the pH to 3 and extracted with DCM (100 ml × 3). The combined organic layer was washed with water (50 ml) and brine (50 ml), dried over Na₂SO₄, and concentrated. The residue was purified by silica gel column chromatography (DCM/MeOH = 40/1) to afford product **6** (4 g, yield: 40%) as yellow oil. Chiral HPLC: *e.e.* > 99%, Chiralpak IB column (5 μm, 4.6 × 250mm), Hexane: EtOH: TFA = 95 : 5: 0.2, 1.0 ml/min, retention time 22.0 min. ¹H-NMR (DMSO-*d*₆, 400 MHz): δ 12.17 (s, 1H), 7.24-7.14 (m, 3H), 7.06 (t, *J* = 5.6 Hz, 1H), 6.79 (d, *J* = 8.0 Hz, 1H), 4.78 (s, 2H), 3.97 (s, 1H), 3.64-3.56 (m, 1H), 2.97-2.94 (m, 1H), 2.80-2.70 (m, 2H), 1.46-1.26 (m, 2H), 1.20-1.12 (m, 13H).

4.6 Synthesis of 2-amino-6-(3-ethynyl-benzoyloxycarbonylamino)-hexanoic acid (*mEtZLys*): The mixture of compound **6** (3.8 g, 9.4 mmol) in HCl/EtOAc (50 ml) was stirred for 2 hr at room temperature. The mixture was concentrated under reduced pressure, and the residue was purified by prep-HPLC to afford the desired product *mEtZLys* (2.6 g, yield: 98%) as a solid yellow HCl salt. ¹H NMR (DMSO-*d*₆, 300 MHz): δ 8.60-8.20 (br, 3H), 7.46-7.26 (m, 5H), 5.00 (s, 2H), 4.21 (s, 1H), 3.86-3.78 (m, 1H), 3.04-2.94 (m, 2H), 1.84-1.72 (m, 2H), 1.50-1.26 (m, 4H). MS Calcd.: 304; MS Found: 305 ([M+H]⁺).

Scheme 5. Synthesis of *oEtZLys*



5.1 Synthesis of (S)-tert-butyl 13,13-dimethyl-3,11-dioxo-1-phenyl-2,12-dioxa-4,10-diazatetradecane-9-carboxylate (2): The solution of (S)-13,13-dimethyl-3,11-dioxo-1-phenyl-2,12-dioxa-4,10-diazatetradecane-9-carboxylic acid (**1**) (50 g, 0.13 mol), di-tert-butylpyrocarbonate (Boc₂O) (44 g, 0.2 mol), and dimethylaminopyridine (DMAP) (24 g, 0.2 mol) in tert-butanol (500 ml) was stirred at room temperature for 5 hrs. The mixture was poured into ice water (500 ml). The aqueous layer was extracted with ethyl acetate (200 ml × 3). The combined organic layers were dried over sodium sulfate, filtered, and evaporated in a vacuum to give the crude product, which was purified by silica gel column chromatography to afford compound **2** (40 g, 70%) as yellow oil.

5.2 Synthesis of (S)-tert-butyl 6-amino-2-(tert-butoxycarbonylamino)hexanoate (3): The solution of compound **2** (40 g, 92 mmol) and Pd/C (8.0 g) in MeOH (400 ml) was stirred at room temperature overnight under 50 psi. The mixture was filtered, and the filtrate was evaporated in a vacuum to afford compound **3** (26 g, 94%) as yellow oil.

5.3 Synthesis of 2-((trimethylsilyl)ethyl)phenyl)methanol (5): The solution of compound **4** ((2-iodophenyl)methanol) (50 g, 0.21 mol), trimethylsilyl acetylene (25 g, 0.26 mol), and tetrakis(triphenylphosphine)palladium(0) [Pd(PPh₃)₄] (5.0 g, 4.3 mmol) in TEA (300 ml) was stirred at 50°C for 1 hr. The mixture was quenched with water (200 ml), and then the aqueous layer was extracted with DCM (200 ml × 3). The combined organic layers were dried over sodium sulfate, filtered, and evaporated in a vacuum to give the crude product, which was purified by silica gel column chromatography to afford compound **5** (50 g, 87%) as yellow oil.

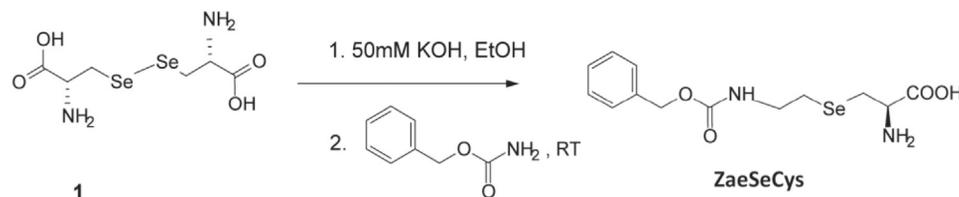
5.4 Synthesis of (2-ethynylphenyl)methanol (6): The mixture of compound **5** (50 g, 0.25 mol) and K₂CO₃ (18 g, 0.13 mol) in methanol (500 ml) was stirred at room temperature for 1 hr. The mixture was quenched with water (500 ml), and then the aqueous layer was extracted with DCM (200 ml × 3). The combined organic layers were dried over sodium sulfate, filtered, and evaporated in a vacuum to afford compound **6** (30 g, 93%) as yellow oil.

5.5 Synthesis of 2-ethynylbenzyl 4-nitrophenyl carbonate (7): A solution of 4-nitrophenyl carbonochloridate (55 g, 0.27 mol) in DCM (50 ml) was added dropwise to a solution of compound **6** (30 g, 0.23 mol) and TEA (30 g, 0.30 mol) in DCM (300 ml) at 0°C, and the resulting mixture was stirred at room temperature overnight. The mixture was quenched with water (300 ml), and the aqueous layer was extracted with DCM (150 ml × 3). The combined organic layers were dried over sodium sulfate, filtered, and evaporated in a vacuum to give the crude product, which was purified by silica gel column chromatography to afford compound **7** (28 g, 42%) as yellow oil.

5.6 Synthesis of (S)-tert-butyl 2-(tert-butoxycarbonylamino)-6-((2-ethynylbenzyloxy)carbonylamino)hexanoate (8): The mixture of compound **7** (10 g, 34 mmol), compound **3** (10 g, 34 mmol), and TEA (5.1 g, 50 mmol) in DCM (20 ml) was stirred at room temperature for 3 hr. The mixture was quenched with water (200 ml), and the aqueous layer was extracted with DCM (150 ml × 3). The combined organic layers were dried over sodium sulfate, filtered, and evaporated in a vacuum to give the crude product, which was purified by silica gel column chromatography to afford compound **8** (15 g, 97%) as yellow oil.

5.7 Synthesis of 2-amino-6-(2-ethynyl-benzyloxy)carbonylamino-hexanoic acid (oEtZLys): The mixture of compound **8** (5.0 g, 11 mmol) in conc. HCl (10 ml) and THF (10 ml) was stirred at room temperature overnight. The mixture was adjusted to pH 8 with saturated NaHCO₃. The precipitate was filtered, and the filtrate cake was dried in a vacuum to afford oEtZLys (2.5 g, 76%) as a white solid. ¹H-NMR (400 MHz, CD₃OD): 1.45-1.59 (m, 4H), 1.81-1.93 (m, 2H), 3.15-3.19 (t, 2H), 3.52-3.56 (t, 1H), 3.81 (s, 1H), 5.25 (s, 2H), 7.29-7.33 (t, 1H), 7.38-7.46 (m, 2H), 7.48-7.51 (d, 1H). The LC/MS purity was >95%; e.e., the (chiral HPLC) purity was >95%, and MS Calcd.: 304; MS Found: 305 ([M+H]⁺).

Scheme 6. Synthesis of ZaeSeCys

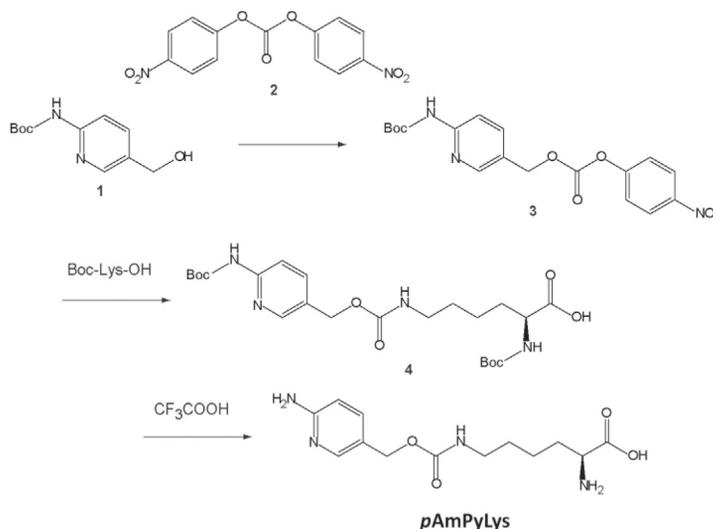


6.1 Synthesis of (2R)-2-amino-6-[(phenylmethoxy)carbonylamino]-4-selenahexanoic acid (ZaeSeCys): L-selenocystine (**1**) (260 mg, 0.77 mmol) was suspended under an argon atmosphere in degassed 50 mM KOH (10 ml) and ethanol (3 ml). The mixture was cooled to 0°C. NaBH₄ (100 mg, 2.6 mmol) was added, and the reaction was allowed to warm to room temperature. After the reaction mixture became colorless, the mixture was placed in an ice-water bath. Carbamic acid benzyl ester (412 mg,

1.6 mmol) was then added, and the mixture was stirred for 6 hours under N_2 at room temperature. 1M HCl (aq) was added to adjust the pH of the solution to 6. The crude product was purified by *prep*-HPLC to give ZaeSeCys (60 mg, yield: 11%) as a white solid.

1H NMR (400 MHz, $DMSO-d_6+CF_3COOD(5\%)$): δ 2.71 (t, $J = 6.8$ Hz, 2H), 2.98-3.01 (m, 2H), 3.27 (t, $J = 6.8$ Hz, 2H), 4.27 (s, 1H), 5.03 (s, 2H), 7.32-7.39 (m, 5H), 15.78-15.80 (m, 3H). LC-MS: Mobile phase: from 95% CH_3CN and 5% water (0.1% TFA) to 5% CH_3CN and 95% water (0.1% TFA) in 6.0 min, finally under these conditions for 0.5 min; wavelength = 214 nm, purity >95%, Retention time = 2.885 min. MS Calcd.: 346; MS Found: 347 ($[M+H]^+$).

Scheme 7. Synthesis of pAmPyLys

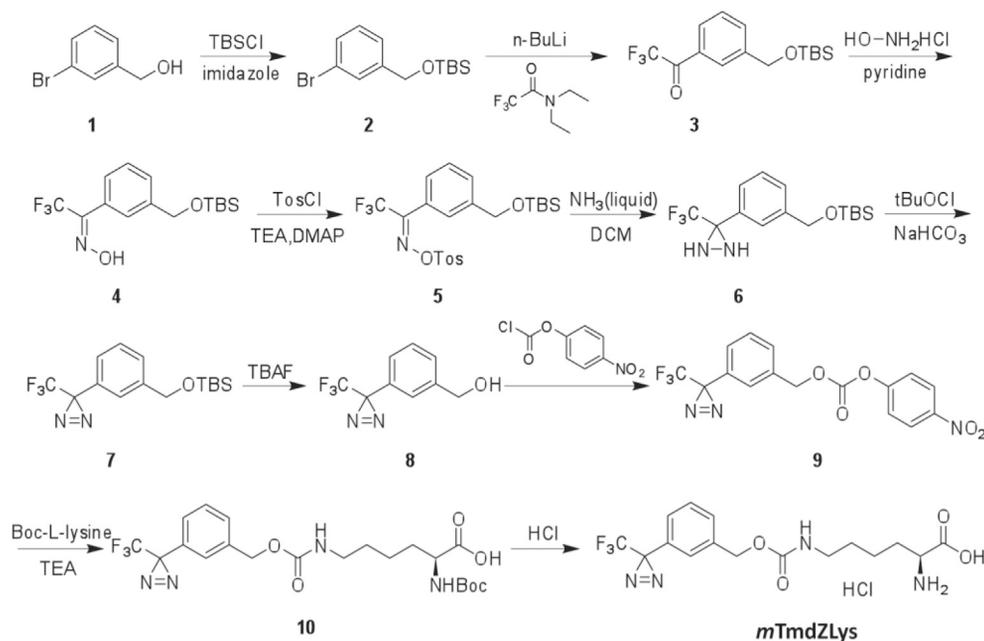


7.1 Synthesis of compound 4: 2-(Boc-amino)-5-pyridinemethanol (**1**) (299.2 mg, 1.33 mmol) was dissolved in THF (20 ml). Bis(4-nitrophenyl) carbonate (**2**) (488 mg, 1.2 eq) and TEA (1.2 eq) were added to the solution, and the mixture was stirred at room temperature for 3 days. Thereafter, reflux was performed overnight, and the formation of carbonic acid, [6-[[[(1,1-dimethylethoxy)carbonyl]amino]-3-pyridinyl]methyl 4-nitrophenyl ester (**3**) was confirmed by TLC. The reaction mixture was cooled to room temperature, Boc-Lys-OH (395 mg, 1.2 eq) was added, and the resulting mixture was stirred for 3 days at 30°C. After stopping the reaction by adding a 2% NH_4Cl solution, it was extracted three times with EtOAc, and the organic layer was dried with $MgSO_4$. After filtration and concentration, purification by silica gel column chromatography gave 439 mg (yield: 66%) of compound **4**.

7.2 Synthesis of (S)-2-amino-6-(((4-aminopyridyl)methoxy)carbonyl)amino)hexanoic acid (pAmPyLys): Compound **4** (300 mg, 604 μ mol) was dissolved in 10 ml of DCM. TFA (5 ml) was added into the solution, and the mixture was stirred at room temperature for about 2.5 hr. After the disappearance of the starting material was confirmed by TLC, the reaction solution was added dropwise to 100 ml of diethyl ether. The resulting precipitate was filtered, washed with ether, and dried to obtain 241 mg of pAmPyLys (TFA salt, yield: 97%).

1H NMR (400 MHz, D_2O): 7.87(d, 1H(aromatic)), 7.77(s, 1H(aromatic)), 6.98(d, 1H(aromatic)), 4.94(s, 2H(benzyl)), 3.90(t, 1H(α)), 3.08(t, 2H(ϵ)), 1.8-2.0(m, 2H(β)), 1.3-1.6(m, 4H(γ , δ)). MS Calcd.: 296.3; MS Found: 297 ($[M+H]^+$).

Scheme 8. Synthesis of mTmdZLys



8.1 Synthesis of (3-bromobenzoyloxy)(*tert*-butyl)dimethylsilane (**2**): A solution of (3-bromophenyl)methanol (**1**) (10 g, 53 mmol) in DCM (100 ml) was mixed with imidazole (7.2 g, 106 mmol) and *tert*-Butyldimethylsilyl chloride (TBSCl) (8.4 g, 56 mmol) at 0°C, and then stirred at room temperature overnight. The solvent was removed in a vacuum, and the residue was purified by silica gel column chromatography to give compound **2** (12 g) as colorless oil, yield: 74.0%.

8.2 Synthesis of 1-(3-((*tert*-butyldimethylsilyloxy)methyl)phenyl)-2,2,2- trifluoroethanone (**3**): Under N₂, a solution of compound **2** (9.0 g, 30 mmol) in THF (90 ml) was mixed with *n*-Butyllithium (13.2 ml, 33 mmol) dropwise at -70°C and stirred for 1 hr, and then a solution of *N,N*-diethyl-2,2,2-trifluoroacetamide (6.0 g, 36 mmol) in THF (10 ml) was added at -70°C, and the mixture was stirred for 1 hr. The mixture was quenched with aq. NH₄Cl, extracted with EtOAc (100 ml), dried over Na₂SO₄, and concentrated to give compound **3** (9.00 g) as yellow oil, yield: 94.0%.

¹H-NMR (400 MHz, CDCl₃): δ 7.93 (s, 1H), 7.83 (d, 1H, J = 7.6 Hz), 7.54 (d, 1H, J = 7.6 Hz), 7.39 (t, 1H, J = 7.6 Hz), 4.69 (s, 2H), 0.84 (s, 9H), 0.00 (s, 6H).

8.3 Synthesis of 1-(3-((*tert*-butyldimethylsilyloxy)methyl)phenyl)-2,2,2- trifluoroethanone oxime (**4**): A solution of compound **3** (9.0 g, 28.3 mmol) in EtOH (90 ml) was combined with pyridine (15.0 ml) and hydroxylamine hydrochloride (5.85 g, 85 mmol). The mixture was stirred at 80°C overnight. The solvent was removed in a vacuum, and the residue was purified by silica gel column chromatography to give compound **4** (7.0 g) as colorless oil, yield: 74.0%.

8.4 Synthesis of 1-(3-((*tert*-butyldimethylsilyloxy)methyl)phenyl)-2,2,2- trifluoroethanone O-tosyl oxime (**5**): A solution of compound **4** (7.0 g, 21 mmol) in DCM (100 ml) was combined with TEA (4.2 g, 42 mmol), DMAP (100 mg), and 4-methylbenzene-1-sulfonyl chloride (4.4 g, 23 mmol) at 0°C. The mixture was stirred at room temperature overnight, and then washed with water (50 ml), and dried over Na₂SO₄. The solvent was removed in a vacuum, and the residue was purified by silica gel column chromatography to give compound **5** (10 g) as colorless oil, yield: 98.0%.

8.5 Synthesis of 3-(3-((*tert*-butyldimethylsilyloxy)methyl)phenyl)-3- (trifluoromethyl)diaziridine (**6**): A solution of compound **5** (8.0 g, 16.4 mmol) in DCM (100 ml) was cooled to -70°C, and liquid ammonia (100 ml) was added. The mixture was stirred at -70°C for 8 hrs, and then warmed to room temperature and stirred overnight. The mixture was washed with water and dried over Na₂SO₄. The solvent was removed to give compound **6** (3.5 g) as colorless oil, yield: 65.0%.

¹H-NMR (400 MHz, DMSO-*d*₆): δ 7.46 (s, 1H), 7.29-7.34 (m, 3H), 4.67 (s, 2H), 4.01 (d, 1H, J = 8.4 Hz), 3.88 (d, 1H, J = 7.6 Hz), 0.83 (s, 9H), 0.00 (s, 6H).

8.6 Synthesis of 3-(3-((*tert*-butyldimethylsilyloxy)methyl)phenyl)-3- (trifluoromethyl)-3H-diazirine (**7**): A solution of compound **6** (3.5 g, 10.5 mmol) in MeOH (100 ml) was cooled to -70°C, and hypochlorous acid *tert*-butyl ester (2.27 g, 21 mmol) was added dropwise. After 10 min, NaHCO₃ (1.32 g, 15.7 mmol) was added to the mixture. The reaction was warmed to room temperature and stirred for 2 hrs. The solvent was removed in a vacuum. The residue was resolved in EA (100 ml), washed with water (50 ml), and dried over Na₂SO₄. The solvent was removed to give compound **7** (3.0 g) as yellow oil, yield: 86.0%.

8.7 Synthesis of (3-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methanol (**8**): A solution of compound **7** (3.0 g, 9.1 mmol) in THF (60.0 ml) was mixed with water (5.0 ml) and tetrabutylammonium fluoride (TBAF) (5.7 g, 18.2 mmol), and stirred at room temperature

overnight. The solvent was removed in a vacuum, and the residue was dissolved in EtOAc (100 ml), washed with water (30 ml × 2), and dried over Na₂SO₄. The solvent was removed and the residue was purified by silica gel column chromatography to give compound **8** (1.2 g) as yellow oil, yield: 67.0%.

¹H-NMR (400 MHz, CDCl₃): δ 7.37-7.41 (m, 2H), 7.15-7.18 (m, 2H), 4.71 (s, 2H), 1.76 (s, 1H).

8.8 Synthesis of 4-nitrophenyl 3-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyl carbonate (**9**): To a solution of compound **8** (1.0 g, 4.62 mmol) in DCM (20 ml), TEA (0.93 g, 9.24 mmol) and 4-nitrophenyl carbonochloridate (0.93 g, 4.62 mmol) were added at 0°C. The mixture was stirred at room temperature for 2 hrs. The solvent was removed and the residue was purified by silica gel column chromatography to give compound **9** (1.2 g) as colorless oil, yield: 68.0%.

8.9 Synthesis of (*S*)-2-(*tert*-butoxycarbonylamino)-6-((3-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyloxy)carbonylamino)hexanoic acid (**10**): To a solution of compound **9** (1.2 g, 3.15 mmol) in dimethylformamide (DMF) (25 ml), TEA (0.63 g, 6.3 mmol) and *N*^z-(*tert*-butoxycarbonyl)-L-lysine (0.7 g, 2.83 mmol) were added, and the mixture was stirred at 50°C overnight. The mixture was poured into 100 ml of water, and extracted with EtOAc (30.0 ml × 3). The organic layer was concentrated, and the residue was purified by prep-HPLC to give compound **10** (0.6 g) as a white solid, yield: 39.0%.

8.10 Synthesis of (*S*)-2-amino-6-((3-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzyloxy)carbonylamino)hexanoic acid hydrochloride (**mTmdZLys**): Compound **10** (300 mg, 0.61 mmol) was dissolved in HCl/MeOH (10 ml) and stirred at room temperature for 2 hr. The solvent was removed to give S0394-A (250 mg) as a white solid, yield: 96.0%.

¹H-NMR (400 MHz, CD₃OD): δ 7.48-7.49 (m, 2H), 7.23-7.26 (m, 2H), 5.11 (s, 2H), 3.97 (t, 1H, *J* = 6.4 Hz), 3.16 (t, 2H, *J* = 6.8 Hz), 1.89-1.99 (m, 2H), 1.45-1.62 (m, 4H).

LC/MS [mobile phase: from 90% water (0.1% formic acid) and 10% CH₃CN (0.1% formic acid) to 5% water (0.1% formic acid) and 95% CH₃CN (0.1% formic acid) in 6.0 min, finally under these conditions for 0.5 min.] The purity was 96.3%, *R*_t = 2.464 min, and MS Calcd.: 388; MS Found: 389.1 ([*M*+*H*]⁺).

QUANTIFICATION AND STATISTICAL ANALYSIS

For the incorporation efficiencies of non-natural amino acids and the time course analysis of aminoacylation, data were presented as mean ± standard deviation (SD) from two or three independent experiments. Statistical analysis (calculation of SD) was performed by using Excel (Microsoft).

DATA AND SOFTWARE AVAILABILITY

The coordinates and structure factors have been deposited in the RSCB Protein Data Bank (ID codes 6AB8, 6AAP, 6AB2, 6ABL, 6AB1, 6AAZ, 6AB0, 6ABM, 6AAD, 6AAC, 6AAN, 6AAQ, 6AAO, and 6ABK, for the PylRSc(Y306A/Y384F) structures in complex with ZLys, ZaeSeCys, oCIZLys, oBrZLys, oAzZLys, pNO₂ZLys, pAmPyLys, pTmdZLys, mTmdZLys, mAzZLys, mEtZLys, BCNLys, TCO^{*}Lys, and TeocLys, respectively). Softwares used for structural analyses are listed in [Key Resources Table](#).