

# Polyspecific pyrrolysyl-tRNA synthetases from directed evolution

Li-Tao Guo<sup>a,1</sup>, Yane-Shih Wang<sup>a,1</sup>, Akiyoshi Nakamura<sup>a,b,1</sup>, Daniel Eiler<sup>a,1</sup>, Jennifer M. Kavran<sup>a</sup>, Margaret Wong<sup>c</sup>, Laura L. Kiessling<sup>d</sup>, Thomas A. Steitz<sup>a,b,e</sup>, Patrick O'Donoghue<sup>f,g,2</sup>, and Dieter Söll<sup>a,b,2</sup>

Departments of <sup>a</sup>Molecular Biophysics & Biochemistry and <sup>b</sup>Chemistry, and <sup>e</sup>Howard Hughes Medical Institute, Yale University, New Haven, CT 06510; <sup>c</sup>Novartis Institutes of Biomedical Research, Cambridge, MA 02139; <sup>d</sup>Department of Chemistry, University of Wisconsin–Madison, Madison, WI 53706; and Departments of <sup>f</sup>Biochemistry and <sup>g</sup>Chemistry, The University of Western Ontario, London, ON, Canada N6A 5C1

Contributed by Dieter Söll, October 17, 2014 (sent for review September 18, 2014)

**Pyrrolysyl-tRNA synthetase (PylRS) and its cognate tRNA<sup>Pyl</sup> have emerged as ideal translation components for genetic code innovation. Variants of the enzyme facilitate the incorporation >100 noncanonical amino acids (ncAAs) into proteins. PylRS variants were previously selected to acylate *N*<sup>ε</sup>-acetyl-Lys (AcK) onto tRNA<sup>Pyl</sup>. Here, we examine an *N*<sup>ε</sup>-acetyl-lysyl-tRNA synthetase (AcKRS), which is polyspecific (i.e., active with a broad range of ncAAs) and 30-fold more efficient with Phe derivatives than it is with AcK. Structural and biochemical data reveal the molecular basis of polyspecificity in AcKRS and in a PylRS variant [iodo-phenylalanyl-tRNA synthetase (IFRS)] that displays both enhanced activity and substrate promiscuity over a chemical library of 313 ncAAs. IFRS, a product of directed evolution, has distinct binding modes for different ncAAs. These data indicate that *in vivo* selections do not produce optimally specific tRNA synthetases and suggest that translation fidelity will become an increasingly dominant factor in expanding the genetic code far beyond 20 amino acids.**

aminoacyl-tRNA synthetase | genetic code | genetic selection | posttranslational modification | synthetic biology

The standard genetic code table relates the 64 nucleotide triplets to three stop signals and 20 canonical amino acids. Some organisms, including humans, naturally evolved expanded genetic codes that accommodate 21 amino acids (1), or possibly 22 amino acids in rare cases (2). Engineering translation system components, including tRNAs (3, 4), aminoacyl-tRNA synthetases (AARSs) (5, 6), elongation factors (7), and the ribosome itself (8), have produced organisms with artificially expanded genetic codes. Products of genetic code engineering include bacterial, yeast, and mammalian cells and animals that are able to synthesize proteins with site-specifically inserted noncanonical amino acids (ncAAs) (9).

Genetic code expansion systems rely on an orthogonal AARS/tRNA pair (*o*-AARS, *o*-tRNA) (5, 6). The *o*-AARS should be specific in ligating a desired ncAA to a stop codon decoding tRNA, and both the *o*-tRNA and *o*-AARS are assumed not to cross-react with endogenous AARSs or tRNAs. Although some AARSs evolved in nature to recognize certain ncAAs (10–12), many genetic code expansion systems require a mutated AARS active site. The active site of the *o*-AARS is usually redesigned via directed evolution (6), including positive and negative selective rounds, to produce an enzyme that is assumed to be specific for an ncAA and not active with the 20 canonical amino acids. Genetic code expansion technology is rapidly evolving (13), and the ability to incorporate multiple ncAAs into a protein using quadruplet-codon decoding (14) or sense-codon recoding (15–19) is now becoming feasible. Protein synthesis with multiple ncAAs will require *o*-AARSs that are able to discriminate their ncAA substrate not only from canonical amino acids in the cell but from other ncAAs that are added to the cell.

Probing the effects of amino acid analogs on bacterial cell growth revealed, over 50 y ago, that many ncAAs were incorporated into proteins by the regular translation machinery (10). Thus, it was not surprising to see that many of the successful orthogonal *Methanococcus jannaschii* tyrosyl-tRNA synthetase variants (20) facilitate incorporation of multiple different ncAAs (21–23). This

polyspecificity is also a property of the orthogonal pyrrolysyl-tRNA synthetase (PylRS)/tRNA<sup>Pyl</sup> pair (reviewed in ref. 24).

PylRS variants that facilitate site-specific insertion of *N*<sup>ε</sup>-acetyl-L-Lys (AcK; 2) (Fig. 1A) into proteins were derived from directed evolution experiments (25–28). These AcK-tRNA synthetase (AcKRS) enzymes have been used to investigate the role of acetylation sites in tumor suppressor p53 (29) and histone H3 (30). Here, we present biochemical and structural studies showing that AcKRS variants are polyspecific and catalytically deficient enzymes compared with canonical AARSs. These AcKRSs selected by directed evolution to ligate AcK to tRNA<sup>Pyl</sup> are actually ~30-fold more efficient in activation with Phe derivatives. Crystallographic structures of AcKRS and PylRS variants in complex with AcK, 3-iodo-L-Phe (3-I-Phe; 4) (Fig. 1A), or 2-(5-bromothieryl)-L-Ala (3-Br-ThA; 10) (Fig. 1A) reveal the structural basis of polyspecificity in these engineered PylRS enzymes.

## Results

**Polyspecificity in Engineered PylRS Enzymes.** We determined the substrate range of two AcKRS enzymes by screening with a chemically diverse library of 313 ncAAs (11, 32), followed by rescreening with a selected subset of Phe and Lys derivatives (Fig. 1B). For this screen, we used a superfolder GFP (sfGFP) reporter (32) bearing a TAG codon at position 2, a site known to be permissive for diverse ncAAs (32, 33). The reporter was coexpressed with plasmids encoding the tRNA<sup>Pyl</sup> and AcKRS; quantification was by sfGFP fluorescence, the indication of ncAA incorporation efficiency. As

## Significance

**Pyrrolysyl-tRNA synthetase (PylRS) and its cognate tRNA<sup>Pyl</sup> have emerged as ideal translation components for genetic code innovation. We found that a series of PylRS variants that were initially selected to be specific for the posttranslational modification *N*<sup>ε</sup>-acetyl-L-Lys displayed polyspecificity [i.e., activity with a broad range of noncanonical amino acid (ncAA) substrates]. Our structural and biochemical data indicate that the engineered tRNA synthetases can accommodate ncAA substrates in multiple binding modes. The data further suggest that *in vivo* selections do not produce optimally specific tRNA synthetases and that translation fidelity will become an increasingly dominant factor in expanding the genetic code far beyond 20 amino acids.**

Author contributions: L.-T.G., Y.-S.W., A.N., D.E., L.L.K., T.A.S., P.O., and D.S. designed research; L.-T.G., Y.-S.W., A.N., D.E., J.M.K., and M.W. performed research; M.W. contributed new reagents/analytic tools; L.-T.G., Y.-S.W., A.N., D.E., J.M.K., T.A.S., P.O., and D.S. analyzed data; and Y.-S.W., A.N., L.L.K., T.A.S., P.O., and D.S. wrote the paper.

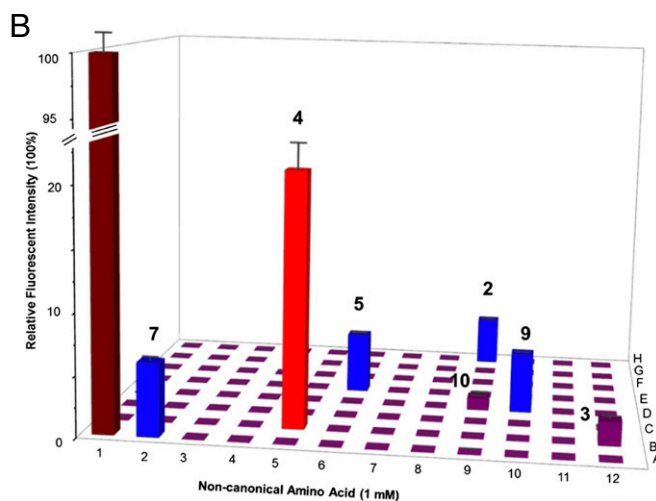
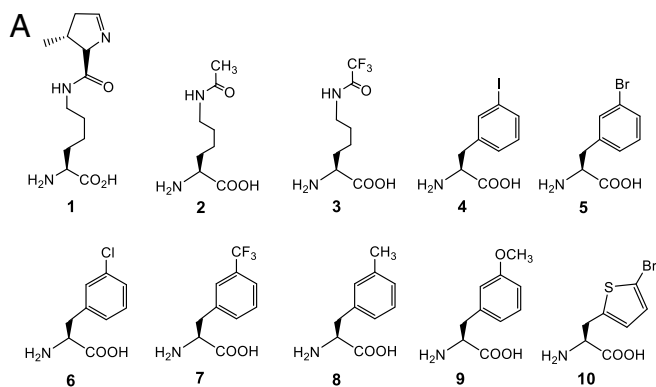
The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID codes 4Q6G, 4TQD, and 4TQF).

<sup>1</sup>L.-T.G., Y.-S.W., A.N., and D.E. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [patrick.odonoghue@uwo.ca](mailto:patrick.odonoghue@uwo.ca) or [dieter.soll@yale.edu](mailto:dieter.soll@yale.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1419737111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1419737111/-DCSupplemental).



**Fig. 1.** (A) Chemical structures of nsAAs used in the study: 1, L-pyrrolysine (Pyl); 2, AcK; 3,  $N^{\epsilon}$ -trifluoroacetyl-L-Lys ( $\text{CF}_3$ -AcK); 4, 3-I-Phe; 5, 3-bromo-L-Phe (3-Br-Phe); 6, 3-chloro-L-Phe (3-Cl-Phe); 7, 3-trifluoromethyl-L-Phe (3- $\text{CF}_3$ -Phe); 8, 3-methyl-L-Phe (3-Me-Phe); 9, 3-methoxyl-L-Phe (3-MeO-Phe); 10, 3-Br-Tha. (B) Range of substrate specificity of AcKRS3. Translation of the sfGFP reporter (UAG codon at position 2) by the library of ncAA-tRNA<sup>Pyl</sup> was measured by fluorescence intensity. A library of 94 different Lys and Phe analogs (Dataset S1) was tested. Fluorescence signals from the incorporation of ncAAs 2–5, 7, 9, and 10 are labeled in A. Well A12 is a control without added ncAA. Well A1 was a positive control experiment to detect the production signal of WT sfGFP (100%). Fluorescence data and error values (SD) are represented by bars from three independent experiments (also given in Dataset S1). Colors (0–4% in purple, 4–8% in blue, 18–24% in red, and 95–100% in brown) have been used to indicate the level of UAG read-through.

a control WT sfGFP (AGC Ser codon at position 2) was monitored. Compared with this WT control, translation of UAG (codon 2) with AcK was 2% (for AcKRS1) and 4% (for AcKRS3).

AcKRS3 showed significant incorporation of six ncAAs (Fig. 1B). Even though AcK (2) is among them (Fig. 1A), four meta-substituted Phe analogs (4, 5, 7, and 9) enabled more sfGFP production compared with AcK (Fig. 1B). Incorporation of 4 and 10 was analyzed by full-length protein MS and LC-tandem MS (MS/MS), which identified the expected ncAA but no other amino acids inserted at position 2 in sfGFP (Figs. S1–S4). The AcKRS1 substrate range is similar (Fig. S5) with four ncAAs (3–5 and 7) showing better incorporation than AcK (2).

The fact that both AcKRS variants showed good activity with 3-I-Phe suggests that the PylRS active site may be better suited for binding of metasubstituted Phe derivatives. This conclusion is supported by a recent report (34) showing the PylRS variant Asn346Ala/Cys348Ala to possess significant activity with 3-I-Phe yet weak activity with Phe. Based on these data, we created PylRS

mutant libraries that randomized active site positions, including N346 and C348 (SI Methods). The libraries were selected for 3-I-Phe incorporation into chloramphenicol acetyl transferase with a TAG codon at position 112. Selection from this library produced a mutant [iodo-phenylalanyl-tRNA synthetase (IFRS); Asn346Ser/Cys348Ile] that showed significantly higher UAG translation (~50% of WT sfGFP) than any known PylRS variant (Table 1). The other meta-substituted Phe analogs (5–8) displayed 30–68% incorporation efficiencies (Table 1, Fig. S6, and Dataset S1). IFRS also shows a broader substrate range than AcKRS1 (Fig. S6). The electrospray ionization (ESI)-MS and MS/MS analyses (Figs. S1–S4) indicate that UAG translation led only to the desired ncAA-containing protein product, and no canonical amino acid insertion at the UAG-encoded locus was observed. These data indicate that the sfGFP fluorescence data are a quantitative measure of ncAA incorporation.

**Enzyme Kinetics of PylRS Variants.** AARSs catalyze a two-step reaction. The amino acid substrate is first activated by adenylation in a reaction with ATP. The aminoacyl-adenylate is then a substrate for aminoacyl-tRNA formation in the second step. Both steps can be monitored biochemically; amino acid-AMP formation is monitored by the ATP-PP<sub>i</sub> exchange assay (35), whereas aminoacylation is assayed by quantitating amino acid ligation to radiolabeled tRNAs (36, 37) (SI Methods). The enzyme kinetic constants ( $k_{\text{cat}}$  and  $K_m$ ) are used to compare activity for the different ncAAs. As reported in earlier studies (38, 39), the truncated version (to overcome solubility problems) of the PylRS enzyme (residues 185–454) was used to measure the ATP-PP<sub>i</sub> exchange. However, in the aminoacylation experiments, full-length PylRS variants could be used because the assay (37) required low enzyme concentrations.

The enzymatic characterization of WT PylRS was stalled by the lack of pyrrolysine (Pyl). Because there is no commercial source of Pyl and its early chemical synthesis had proven to be challenging (40), we used a sample from an alternative synthetic route (41). We determined the kinetic parameters with Pyl for WT *Methanosarcina mazei* PylRS (MmPylRS) and *Methanosarcina barkeri* PylRS (MbPylRS). The data show a  $K_m$  of ~50  $\mu\text{M}$  and a catalytic turnover of 0.1–0.3  $\text{s}^{-1}$  in amino acid-AMP formation (Table 2) and values of 20  $\mu\text{M}$  and 0.008–0.03  $\text{s}^{-1}$  for Pyl-tRNA<sup>Pyl</sup> formation (Table 1). Thus, even with its native substrate, PylRS is less efficient than many AARSs with  $k_{\text{cat}}$  values about >1,000-fold lower than measured for many canonical AARSs (e.g., phenylalanyl-tRNA synthetase) (42) (Tables 1 and 2). This lower activity is not surprising, because in *Methanosarcina*, Pyl is translated by as few as ~50 codons, whereas a “canonical” tRNA synthetase [e.g., *Escherichia coli* leucyl-tRNA synthetase (43)] normally needs to provide substrate to translate ~150,000 codons.

AcKRS1 is most catalytically efficient with the substrate 3-trifluoromethyl-L-Phe (7) (Fig. 1). This activity is 166-fold less efficient compared with WT PylRS activity with Pyl. Surprisingly, the  $K_m$  for AcKRS1 with AcK is 35 mM, but its  $k_{\text{cat}}$  is only 10-fold lower than WT PylRS for Pyl. The best substrate for AcKRS3 is 3-I-Phe (4) (Fig. 1), but the activity is 288-fold less efficient compared with WT PylRS with Pyl as a substrate.

The UAG translation efficiencies observed in vivo in the sfGFP reporter (Table 2 and Figs. S5 and S6) are in moderate agreement with the enzyme-kinetic data. However, aminoacylation is only one component that influences translation efficiency; ncAA uptake (44) and stability and elongation factor Tu (EF-Tu) compatibility (7) are known to alter the efficiency of translation. It is also possible that some ncAAs or translation of off-target UAG codons could alter the cellular proteome or metabolism in ways that may have an impact on protein synthesis and its fidelity.

The improved translation efficiency observed for IFRS is evident in biochemical assays, which show IFRS to be fivefold more efficient in 3-I-Phe-adenylate formation compared with AcKRS1 or AcKRS3 (Table 2). However, IFRS is still less efficient in the aminoacylation of tRNA<sup>Pyl</sup> (~40-fold) compared with WT PylRS

**Table 1. Apparent kinetic parameters of PylRS variants for aminoacylation**

Enzymes	Amino acids	$K_m$ , $\mu\text{M} \times 10^3$	$k_{cat}$ , $\text{s}^{-1} \times 10^{-2}$	$k_{cat}/K_m$ , $\mu\text{M}^{-1}\cdot\text{s}^{-1} \times 10^{-5}$	Relative catalytic efficiency	UAG translation efficiency*
MmPylRS	Pyl (1)	<b>0.02 ± 0.004</b>	<b>0.83 ± 0.03</b>	<b>41.5</b>	100	15 ± 1.5
MbPylRS	Pyl (1)	<b>0.02 ± 0.002</b>	<b>3.01 ± 0.04</b>	<b>151</b>	364	— <sup>†</sup>
IFRS	3-I-Phe (4)	<b>0.44 ± 0.04</b>	<b>0.44 ± 0.02</b>	<b>1.00</b>	2.4	48 ± 1.2
	3-Br-Phe (5)	<b>0.62 ± 0.24</b>	<b>0.48 ± 0.03</b>	<b>0.77</b>	1.9	68 ± 5.2
	3-Cl-Phe (6)	<b>1.36 ± 0.13</b>	<b>0.76 ± 0.05</b>	<b>0.56</b>	1.3	46 ± 3.2
	3-CF <sub>3</sub> -Phe (7)	<b>0.45 ± 0.09</b>	<b>0.28 ± 0.01</b>	<b>0.62</b>	1.5	53 ± 2.3
	3-Me-Phe (8)	<b>0.95 ± 0.17</b>	<b>0.34 ± 0.05</b>	<b>0.36</b>	0.9	30 ± 0.5
	3-MeO-Phe (9)	<b>1.64 ± 0.19</b>	<b>0.07 ± 0.01</b>	<b>0.04</b>	0.1	<1
	3-Br-ThA (10)	<b>0.37 ± 0.09</b>	<b>0.14 ± 0.01</b>	<b>0.38</b>	0.9	7 ± 0.9

Apparent kinetic parameters of PylRS variants for aminoacylation were determined by quantitating amino acid ligation to radiolabeled tRNAs (36). Numbers of amino acids are shown in bold. 3-CF<sub>3</sub>-Phe, 3-trifluoromethyl-L-Phe; 3-Cl-Phe, 3-chloro-L-Phe; 3-Me-Phe, 3-methyl-L-Phe; 3-MeO-Phe, 3-methoxy-L-Phe.

\*UAG translation efficiency is from data in Figs. S5 and S6.

<sup>†</sup>Not measured.

(Table 1). Despite the fact that PylRS is significantly more active for Pyl than are engineered AARSs for their substrates, we recorded only 15% UAG read-through with Pyl, whereas Pyl analogs support up to 40% UAG read-through (35). These data may indicate that *E. coli* imports Pyl less efficiently than other ncAAs or that it is less stable; however, a firm conclusion requires determination of the concentrations of these amino acids in the cell.

**Anatomy of an Engineered Active Site.** Seven distinct AckRS variants have been reported previously (25–28, 30). Three variants were derived from the MbPylRS (25, 30), whereas the others originated from the MmPylRS (26–28). Biochemical studies revealed that the AckRS enzymes were far less catalytically efficient than natural tRNA synthetases (26). This result led us to characterize AcKRS complexed with AcK structurally. Because we had previously optimized crystallization conditions for MmPylRS, we constructed AcKRS3, an MmPylRS variant that carried the active site residues of an *M. barkeri* AcKRS (Table S1). Cocrystals

of AcKRS3 complexed to AcK and the ATP analog 5'-adenylyl  $\beta,\gamma$ -imidodiphosphate (ADPNP) were obtained following previously published protocols (38), with specific modifications (SI Methods). The final structure was refined at a resolution of 2.3 Å with an *R* factor of 18.6% and an *R*<sub>free</sub> value of 21.4% (Table S2).

AcKRS3 crystals contain one protein molecule in the asymmetric unit. The anticipated AcKRS3 dimer is created by crystallographic symmetry between neighboring enzymes in the crystal lattice. The Fourier difference (*F*<sub>o</sub>-*F*<sub>c</sub>) electron density maps show clear density for the ADPNP contoured at 4 $\sigma$ . Density for AcK appears at 4 $\sigma$ , with the strongest density around the  $\alpha$ - and  $\beta$ -carbons of the *R* group. There is no density for the methyl group in the acetyl moiety (Fig. S7). High  $\beta$ -factors associated with the terminus of the AcK *R* group indicate that the amino acid substrate is highly mobile in the active site. In addition, the AcK shows ~70% occupancy even at a 75 mM concentration in the cryogen solution. The remaining molecules in the crystal contain water molecules with well-defined hydrogen bonding partners in the active site. In

**Table 2. Apparent kinetic parameters of *M. mazel* PylRS variants for amino acid activation**

Enzymes	Amino acids	$K_m$ , $\mu\text{M} \times 10^3$	$k_{cat}$ , $\text{s}^{-1} \times 10^{-2}$	$k_{cat}/K_m$ , $\mu\text{M}^{-1}\cdot\text{s}^{-1} \times 10^{-5}$	Relative catalytic efficiency	UAG translation efficiency*
MmPylRS	Pyl (1)	<b>0.05 ± 0.008</b>	<b>29.8 ± 1.2</b>	<b>596</b>	100	15 ± 1.5
MbPylRS <sup>†</sup>	Pyl (1)	<b>0.055 ± 0.005</b>	<b>10.5 ± 1.1</b>	<b>191</b>	32	— <sup>‡</sup>
AcKRS1	3-I-Phe (4)	<b>6.14 ± 0.32</b>	<b>12.6 ± 1.5</b>	<b>2.05</b>	0.34	19 ± 2.0
	3-Br-Phe (5)	<b>11.9 ± 0.30</b>	<b>13.6 ± 0.35</b>	<b>1.14</b>	0.19	3 ± 0.5
	3-Cl-Phe (6)	<b>9.46 ± 3.85</b>	<b>8.54 ± 2.73</b>	<b>0.90</b>	0.15	<1
	3-CF <sub>3</sub> -Phe (7)	<b>1.09 ± 0.28</b>	<b>3.91 ± 0.48</b>	<b>3.59</b>	0.60	3 ± 0.3
	3-Me-Phe (8)	<b>0.80 ± 0.19</b>	<b>1.37 ± 0.19</b>	<b>1.71</b>	0.29	<1
	3-MeO-Phe (9)	<b>6.27 ± 1.17</b>	<b>4.15 ± 0.35</b>	<b>0.66</b>	0.11	<1
	CF <sub>3</sub> -AcK (3)	<b>0.61 ± 0.06</b>	<b>1.24 ± 0.02</b>	<b>2.03</b>	0.34	3 ± 0.2
	AcK <sup>§</sup> (2)	<b>35.3 ± 10.9</b>	<b>3.23 ± 0.61</b>	<b>0.09</b>	0.02	2 ± 0.1
	Pyl (1)	nd	nd			
AcKRS3	AcK (2)	137 ± 59	14.6 ± 3.8	0.11	0.02	4 ± 0.2
	3-I-Phe (4)	16.6 ± 1.9	34.3 ± 2.4	2.07	0.35	21 ± 2.1
	3-Br-Phe (5)	22.4 ± 2.0	26.3 ± 1.0	1.17	0.20	5 ± 0.1
	3-CF <sub>3</sub> -Phe (7)	3.21 ± 0.97	4.68 ± 0.39	1.46	24	6 ± 0.3
IFRS	3-I-Phe (4)	0.82 ± 0.09	8.71 ± 0.33	10.6	1.8	48 ± 0.1
	3-CF <sub>3</sub> -Phe (7)	1.13 ± 0.10	7.29 ± 0.21	6.45	1.1	53 ± 2.3
	3-Br-ThA (10)	1.57 ± 0.32	4.43 ± 0.27	2.82	0.47	7 ± 0.9

Apparent kinetic parameters of MmPylRS variants for amino acid activation were determined by ATP-PP<sub>i</sub> exchange for the amino acids indicated. Numbers of amino acids are shown in bold. CF<sub>3</sub>-AcK, *N*<sup>ε</sup>-trifluoroacetyl-L-Lys; nd, not detectable.

\*UAG translation efficiency is from data in Figs. S5 and S6. Translation of the WT sfGFP (AGC is codon 2) is considered 100%, and the percentage of sfGFP observed resulting from UAG translation with the indicated ncAA is referred to as the UAG translation efficiency.

<sup>†</sup>Kinetic data (*K*<sub>m</sub> and *k*<sub>cat</sub>) for MbPylRS were reproduced from previous work (66).

<sup>‡</sup>Not measured.

<sup>§</sup>Kinetic data were reproduced from a study by Umehara et al. (26).



comparison to the WT PylRS structure, one of the most noticeable changes in the amino acid binding pocket is the movement of Asn346 (Fig. 2). The density for Tyr384 is not resolved, and the aromatic ring appears to be disordered (45).

Comparison of the MmPylRS and AcKRS3 show that the structures are nearly identical, except in the active site (Fig. 2). All AcKRS variants differ from their parent PylRS at four to five positions in the active site. With the exception of AcKRS2, all other AcKRSs have the mutations Cys348Phe and Leu309Ala (Table S1). The Tyr at position 306 is either a Leu or Phe residue. The AcKRS3 structure indicates that the Leu309Ala substitution creates space for a new cluster of hydrophobic and bulky residues at positions 348 and 306. These mutations lead to an AcKRS active site that is smaller and more hydrophobic than that of the parent PylRS, which demonstrates why AcKRS is no longer active with the larger substrate, Pyl (26) (Table 2) and why the parent PylRS enzyme is not active in inserting AcK into proteins *in vivo* (26). Interestingly, one PylRS variant (Tyr306Ala/Tyr384Phe) was shown to accommodate furan-containing ncAAs that are substantially larger than Pyl (46). In a crystal structure, the Ala substitution at position 306 was implicated in extending the size of the active site pocket to accommodate larger ncAAs.

A number of structural studies have shown significant plasticity in the active site of the WT PylRS (38, 47, 48), which explains earlier biochemical studies demonstrating that PylRS is able to accommodate several different Pyl and Lys derivatives (35, 49). Structural analyses have also revealed how mutant PylRS enzymes can accommodate ncAAs that are chemically distinct (50) and larger than Pyl (47), including furan-containing Lys (46) and norbornene-containing Pyl analogs (51). Larger ncAAs are accepted by these PylRS variants by creating larger amino acid recognition pockets with mutations to smaller active site residues, particularly at position 306.

The structure suggests that hydrophobic shape complementarity is the principal means by which AcKRS recognizes AcK. In the PylRS/Pyl complex, Tyr384 and Asn346 make specific hydrogen bonding interactions with the amide group and nitrogen atom on the pyrroline ring (Fig. 2B). In the AcKRS structure, the *N*<sup>ε</sup>-amide group has no hydrogen bonding partners. In fact, Asn346 forms a novel interaction with the  $\alpha$ -amino group of the AcK substrate (Fig. 2B). This interaction shifts the AcK substrate away from a catalytically competent conformation, which is exemplified by the Pyl substrate in the PylRS complex (Fig. 2C).

**Structural Basis of Polyspecificity.** To understand how IFRS accommodates chemically diverse ncAAs, we solved two of the crystal structures of IFRS in complex with 3-I-Phe and 3-Br-ThA. Crystals of 3-I-Phe or 3-Br-ThA complexes diffracted at 2.1 or 2.7 Å, respectively (Table S2). Overall, the structures are highly

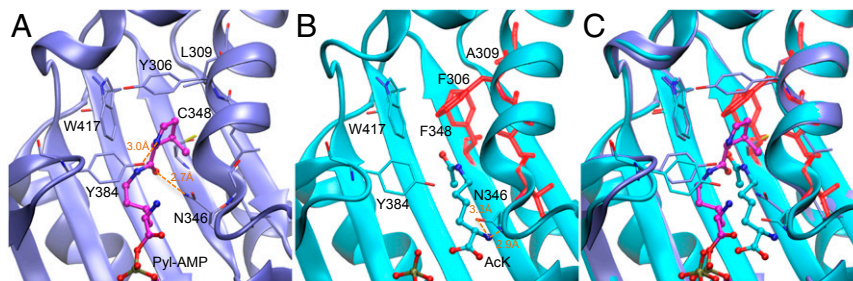
similar, and electron densities corresponding to the ncAA substrates were clearly assigned in the active site (Fig. S8). In the 3-I-Phe complex structure, we found evidence that the ncAA substrate is present in at least two conformations in the active site. The residual Fourier difference ( $mF_o - dF_c$ ) electron density map at  $4\sigma$  suggests two locations for the iodo atom (Fig. S8A).

Most strikingly, 3-I-Phe and 3-Br-ThA occupy distinct binding pockets within the active site of IFRS (Fig. 3). The twisted five-membered ring of 3-Br-ThA allows this ncAA to bind deeper into the active site cleft, which positions the carbonyl further away from the site of adenylation catalysis compared with 3-I-Phe. The  $mF_o - dF_c$  map at  $4\sigma$  suggests a single clearly defined location for the Br atom (Fig. S8B).

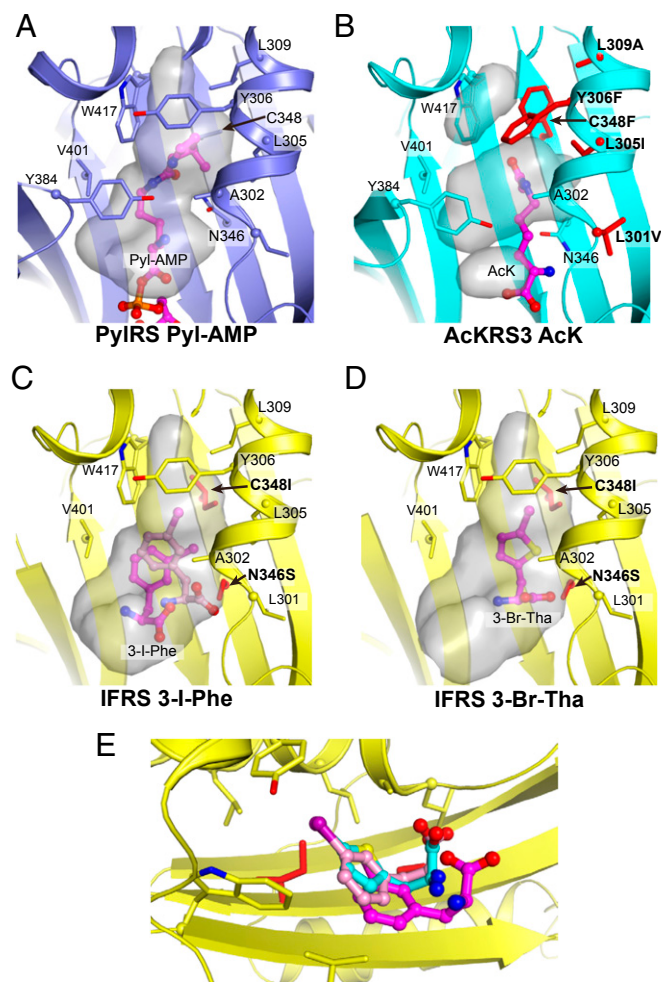
One of the 3-I-Phe conformations is more similar to 3-Br-ThA, and these structures may represent nonproductive complexes. The enhanced activity IFRS shows toward 3-I-Phe compared with 3-Br-ThA may be due to the fact that 3-I-Phe is able to bind in a more productive conformation for catalysis to proceed. Interestingly, in one conformation, the iodo atom of 3-I-Phe binds at exactly the same position as the S atom in 3-Br-ThA, and in the second 3-I-Phe conformation, the iodo and bromo groups are colocalized (Fig. 3E). A number of PylRS structures complexed with Lys derivatives have led to the conclusion that the large size of the PylRS active site pocket is principally responsible for its activity with multiple ncAAs (52). Although this conclusion is likely correct, the fact that IFRS has (at least) two binding modes for different ncAAs is clearly related to the polyspecificity that this enzyme displays.

## Discussion

Genetic selection underlies evolution in nature as well as directed evolution in the laboratory. The powerful nature of genetic selection leads one to feel that these procedures should reveal from a library of genetic variants the best enzyme, in terms of activity and specificity. However, more than half a century ago, many ncAAs were shown to be incorporated into proteins by the natural protein synthesis machinery (10). These ncAAs are close structural analogs of the natural amino acids, and the endogenous tRNA synthetases cannot effectively discriminate against these unnatural substrates (11). It is perhaps self-evident that natural selection cannot optimize enzymes to reject unnatural substrates that the cell has never encountered. The same is true for directed evolution experiments. It is normally assumed that off-target activities are weak or residual compared with natural substrates (53). Therefore, it should not be surprising that directed evolution produces polyspecific tRNA synthetases. Genetic code expansion technologies are moving toward genetically encoding multiple diverse ncAAs. Cross-reactivity between engineered tRNA synthetases and diverse ncAA substrates will become a limiting factor in creating high-fidelity orthogonal translation systems with multiple varied ncAAs (16, 54, 55).



**Fig. 2.** Comparison of PylRS and AcKRS3 substrate complexes. (A) View of the active site of PylRS [purple, Protein Data Bank (PDB) ID code 2ZIM (38)] in complex with Pyl-adenylate (Pyl-AMP; magenta). Hydrogen bonds to the substrate amino acid are shown (dashed lines), with distances between the heavy atoms indicated. (B) View of the active site of AcKRS3 (cyan, PDB ID code 4Q6G) complex with AcK and ATP analog ADPNP with mutated residues shown in red. (C) Active site view of a structural overlap of PylRS and AcKRS3. The substrate (Pyl-AMP) and key recognition residues from PylRS are shown (purple, sticks) overlaid with residues in the AcKRS3 active site (cyan or red, sticks).



**Fig. 3.** Structural comparison of WT PyIRS (A) and PyIRS (B–D) variants in complex with nCAA substrates. PyIRS variants and nCAAs are indicated in each panel. Mutations from the parent PyIRS are indicated in bold, and the mutated side chains are shown in red. In C, the IFRS/3-I-Phe complex shows two conformations of 3-I-Phe (magenta, pink) observed in the crystal structure. The gray surfaces show the solvent-accessible volume within the active site pockets. These volumes were computed in the absence of the nCAA substrate. The loop, including Tyr384, is disordered in the IFRS complex structures in C and D and is not visible. (E) Structural overlay of the IFRS complex structures with 3-I-Phe in two conformations (magenta, pink) and 3-Br-Tha (cyan) shows distinct binding modes of the nCAA substrates.

The extent of protein engineering is limited by the fitness of the organism. Another reason why directed evolution experiments may fail to produce enzymes with the desired level of activity or specificity is that certain variants may be toxic to the cell. It is well known that protein synthesis quality control (56) ensures cellular fitness by preventing the accumulation of mistranslated proteins and the associated unfolded protein response mechanisms that underlie diverse human diseases (57). For genetic codes expanded in the laboratory, protein quality control is also an important issue. A system developed to reassign UAG codons to *O*-phosphoserine (Sep) leads to significant growth defects when expressed in WT *E. coli* cells (33). Translational read-through of endogenous UAG codons with Sep lengthened native proteins beyond their normal termination point and reduced cell viability. This and related work motivated an extensive genome editing effort using multiplex automated genome engineering (MAGE) that produced an *E. coli* strain in which all genomic TAG codons were mutated to TAA,

thus providing a clean genetic background for reassignment of UAG to any desired nCAA (58).

Engineering other components of the translation machinery may also be required for optimal protein synthesis with nCAAs (59). In redesigning EF-Tu to accommodate site-specific selenocysteine incorporation, certain EF-Tu variants that specifically recognized selenocysteinyl-tRNA *in vitro* were toxic to the cell when expressed *in vivo* (60). Similarly, EF-Tu mutants designed to carry a bulky fluorescent nCAA-tRNA did not release the nCAA-tRNA fast enough to support efficient translation (61).

The best enzyme may not be the same as the enzyme most “orthogonal” to the natural amino acids. Traditionally, *o*-AARS/*o*-tRNA pairs were derived from successive positive and negative rounds of selection (5, 6). In the positive rounds, UAG translation of an antibiotic marker leads to cell survival. This selection reveals tRNA synthetase variants that translate UAG with nCAAs and/or canonical amino acids. The nCAA is omitted from negative selection, and translation of a toxic protein containing an in-frame UAG codon eliminates variants that promote incorporation of canonical amino acids. This stringent negative selection may eliminate some highly active synthetase variants. Indeed, it was recently demonstrated that relaxing selection conditions and screening variants for high-yield nCAA incorporation *in vivo* produced a second-generation 3-nitrotyrosyl-tRNA synthetase that is an order of magnitude more efficient than the first-generation enzyme (62). Because the mutant synthetase showed significant activity with the natural Tyr, relaxing selection stringency also has an impact on translation fidelity with nCAAs.

In addition to specificity, enzymatic activity of engineered tRNA synthetases may be far lower (by up to ~1,000-fold) than what is observed in many natural tRNA synthetases (42, 43, 63). Despite this fact, even these lower activity tRNA synthetases can facilitate almost “normal” levels of protein synthesis as judged by sfGFP production. ESI-MS and LC-MS/MS analysis confirmed that the sfGFP produced resulted only from insertion of the desired nCAA in response to the UAG codon (Figs. S1–S4). We also found that modest increases in catalytic efficiency ( $k_{cat}/K_m$ ) can increase translational read-through of UAG to a level of 68% compared with normal sense codon decoding (Table 2). In fact, these and other findings challenge the widely held idea that tRNA synthetases display “typical”  $K_m$  or  $k_{cat}$  values. Apart from the well-known AARS variants, tRNA synthetases and functionally distinct paralogs of these enzymes show a wide range of enzymatic activity in nature. PoxA, a paralog of lysyl-tRNA synthetase (LysRS) that uses (R)- $\beta$ -Lys to modify elongation factor P posttranslationally (64), is 400-fold less catalytically efficient than LysRS, yet the enzyme is perfectly acceptable to the cell, which would otherwise have evolved it to greater efficiency.

Recent success in synthesizing proteins with 21 (58) or 22 (54) amino acids at efficiencies nearing that of normal protein synthesis are promising for the field. New technologies, including MAGE (58) and phage-assisted continuous evolution (65), are expected to play important roles in further expanding the genetic code by allowing sampling of larger libraries of enzyme variants and enabling improvement of moderately active orthogonal translation components. Even with these technological advances, it is clear that genetic code engineers face complex issues in protein engineering and design related to tRNA synthetase activity, specificity, and translation fidelity in their attempts to evolve cells to make proteins with more than 22 amino acids.

## Methods

Enantiomerically pure L-Pyl was synthesized exactly as described previously (41). For crystallographic studies, overexpression and purification of the C-terminal domain (CTD; residues 188–454) of AcKRS1 were performed in the same manner as for the CTD of PyIRS described previously (38). The CTD of MmPyIRS and IFRS was used for assaying ATP-PP<sub>i</sub> exchange. Aminoacylation was measured with full-length MbPyIRS, MmPyIRS, and IFRS. The enzyme substrate range was determined by the sfGFP assay using 313 nCAAs at a 1 mM concentration (SI Methods).



**ACKNOWLEDGMENTS.** We thank Markus Englert, Chenguang Fan, Ilka Heinemann, Jae-hyeong Ko, Jiqiang Ling, Yuchen Liu, Laure Prat, and Don Spratt for insightful discussions. This work was supported by NIH Grants GM055984 (to

L.L.K.), GM22854 (to D.S.), and P01 GM022778 (to T.A.S.); Natural Sciences and Engineering Research Council of Canada Grant RGPIN 04282-2014 (to P.O.); and Defense Advanced Research Projects Agency Contract N66001-12-C-4211 (to D.S.).

- Yuan J, et al. (2010) Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* 584(2):342–349.
- Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* 3(1):29–35.
- Liu DR, Magliery TJ, Pastrnak M, Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc Natl Acad Sci USA* 94(19):10092–10097.
- Aldag C, et al. (2013) Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew Chem Int Ed Engl* 52(5):1441–1445.
- Sharma N, Furter R, Kast P, Tirrell DA (2000) Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS Lett* 467(1):37–40.
- Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444.
- Park HS, et al. (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* 333(6046):1151–1154.
- Wang K, Neumann H, Peak-Chew SY, Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* 25(7):770–777.
- Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* 83:379–408.
- Richmond MH (1962) The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol Rev* 26:398–420.
- Fan C, Ho JM, Chirathivat N, Söll D, Wang YS (2014) Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* 15(12):1805–1809.
- Merkel L, Budisa N (2012) Organic fluorine as a polypeptide building element: In vivo expression of fluorinated peptides, proteins and proteomes. *Org Biomol Chem* 10(36):7241–7246.
- Leinke EA (2014) The exploding genetic code. *ChemBioChem* 15(12):1691–1694.
- Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464(7287):441–444.
- Kwon I, Kirshenbaum K, Tirrell DA (2003) Breaking the degeneracy of the genetic code. *J Am Chem Soc* 125(25):7512–7513.
- Bröcker MJ, Ho JM, Church GM, Söll D, O'Donoghue P (2014) Recoding the genetic code with selenocysteine. *Angew Chem Int Ed Engl* 53(1):319–323.
- Bohlik N, Budisa N (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: Rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol Lett* 351(2):133–144.
- Zeng Y, Wang W, Liu WR (2014) Towards reassigning the rare AGG codon in *Escherichia coli*. *ChemBioChem* 15(12):1750–1754.
- Krishnakumar R, Ling J (2014) Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion. *FEBS Lett* 588(3):383–388.
- Wang L, Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem Biol* 8(9):883–890.
- Stokes AL, et al. (2009) Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol Biosyst* 5(9):1032–1038.
- Young DD, et al. (2011) An evolved aminoacyl-tRNA synthetase with atypical poly-substrate specificity. *Biochemistry* 50(11):1894–1900.
- Cooley RB, Karplus PA, Mehl RA (2014) Gleaning unexpected fruits from hard-won synthetases: Probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBioChem* 15(12):1810–1819.
- Wan W, Tharp JM, Liu WR (2014) Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* 1844(6):1059–1070.
- Neumann H, Peak-Chew SY, Chin JW (2008) Genetically encoding N<sup>(ε)</sup>-acetyllysine in recombinant proteins. *Nat Chem Biol* 4(4):232–234.
- Umehara T, et al. (2012) N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* 586(6):729–733.
- Mukai T, et al. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem Biophys Res Commun* 371(4):818–822.
- Mukai T, et al. (2011) Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem Biophys Res Commun* 411(4):757–761.
- Arbely E, et al. (2011) Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. *Proc Natl Acad Sci USA* 108(20):8251–8256.
- Neumann H, et al. (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* 36(1):153–163.
- Kaelin WG, Jr, McKnight SL (2013) Influence of metabolism on epigenetics and disease. *Cell* 153(1):56–69.
- Ko JH, et al. (2013) Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett* 587(19):3243–3248.
- Heinemann IU, et al. (2012) Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Lett* 586(20):3716–3722.
- Wang YS, Fang X, Wallace AL, Wu B, Liu WR (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J Am Chem Soc* 134(6):2950–2953.
- Polcarpo CR, et al. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett* 580(28–29):6695–6700.
- Ambrogelly A, et al. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci USA* 104(9):3141–3146.
- Wolfson AD, Pleiss JA, Uhlenbeck OC (1998) A new assay for tRNA aminoacylation kinetics. *RNA* 4(8):1019–1023.
- Kavran JM, et al. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci USA* 104(27):11268–11273.
- Yanagisawa T, et al. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J Mol Biol* 378(3):634–652.
- Hao B, et al. (2004) Reactivity and chemical synthesis of L-pyrrolysine-22<sup>(nd)</sup> genetically encoded amino acid. *Chem Biol* 11(9):1317–1324.
- Wong ML, Guzei IA, Kiessling LL (2012) An asymmetric synthesis of L-pyrrolysine. *Org Lett* 14(6):1378–1381.
- lbba M, Kast P, Hennecke H (1994) Substrate specificity is determined by amino acid binding pocket size in *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry* 33(23):7107–7112.
- Chen JF, Guo NN, Li T, Wang ED, Wang YL (2000) CP1 domain in *Escherichia coli* leucyl-tRNA synthetase is crucial for its editing function. *Biochemistry* 39(22):6726–6731.
- Steinfeld JB, Aerni HR, Rogulina S, Liu Y, Rinehart J (2014) Expanded cellular amino acid pools containing phosphoserine, phosphothreonine, and phosphotyrosine. *ACS Chem Biol* 9(5):1104–1112.
- Yanagisawa T, Sumida T, Ishii R, 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* 69(Pt 1):5–15.
- Schmidt MJ, Weber A, Pott M, Welte W, 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(12):1755–1760.
- Yanagisawa T, et al. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N<sup>(ε)</sup>-(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* 15(11):1187–1197.
- Flügel V, Vrabel M, Schneider S (2014) Structural basis for the site-specific incorporation of lysine derivatives into proteins. *PLoS ONE* 9(4):e96198.
- Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S (2009) Recognition of non- $\alpha$ -amino substrates by pyrrolysyl-tRNA synthetase. *J Mol Biol* 385(5):1352–1360.
- Takimoto JK, Dellas N, Noel JP, 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(7):733–743.
- Schneider S, et al. (2013) Structural insights into incorporation of norbornene amino acids for click modification of proteins. *ChemBioChem* 14(16):2114–2118.
- Yanagisawa T, Umehara T, Sakamoto K, Yokoyama S (2014) Expanded genetic code technologies for incorporating modified lysine at multiple sites. *ChemBioChem* 15(15):2181–2187.
- Tracewell CA, Arnold FH (2009) Directed enzyme evolution: Climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* 13(1):3–9.
- Wang K, et al. (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem* 6(5):393–403.
- Lepthien S, Merkel L, Budisa N (2010) In vivo double and triple labelling of proteins using synthetic amino acids. *Angew Chem Int Ed Engl* 49(32):5446–5450.
- Reynolds NM, Lazazzera BA, lbba M (2010) Cellular mechanisms that control mistranslation. *Nat Rev Microbiol* 8(12):849–856.
- Drummond DA, Wilke CO (2009) The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* 10(10):715–724.
- Lajoie MJ, et al. (2013) Genomically recoded organisms expand biological functions. *Science* 342(6156):357–360.
- Pavlov MY, et al. (2009) Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci USA* 106(1):50–54.
- Haruna K, Alkazemi MH, Liu Y, Söll D, Englert M (2014) Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res* 42(15):9976–9983.
- Mittelstaet J, Konevega AL, Rodnina MV (2013) A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J Am Chem Soc* 135(45):17031–17038.
- Cooley RB, et al. (2014) Structural basis of improved second-generation 3-nitro-tyrosine tRNA synthetases. *Biochemistry* 53(12):1916–1924.
- O'Donoghue P, Ling J, Wang YS, Söll D (2013) Upgrading protein synthesis for synthetic biology. *Nat Chem Biol* 9(10):594–598.
- Roy H, et al. (2011) The tRNA synthetase paralogue PoxA modifies elongation factor-P with (R)- $\beta$ -lysine. *Nat Chem Biol* 7(10):667–669.
- Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR (2014) Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol* 10(3):216–222.
- Li WT, et al. (2009) Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J Mol Biol* 385(4):1156–1164.