HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2018 April 16.

Published in final edited form as:

Nat Chem Biol. 2017 December ; 13(12): 1261–1266. doi:10.1038/nchembio.2497.

Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase

Tateki Suzuki[a,†], **Corwin Miller**[a,f,†], **Li-Tao Guo**[a], **Joanne M. L. Ho**[a], **David I. Bryson**[c], **Yane-Shih Wang**[a,g], **David R. Liu**[c,d,e], and **Dieter Söll**[a,b,*]

[a]Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT, USA

[b]Department of Chemistry, Yale University, New Haven, CT, USA

[c]Department of Chemistry and Chemical Biology, Harvard University

[d]Howard Hughes Medical Institute, Cambridge, MA, USA

[e]Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA

Abstract

Pyrrolysyl-tRNA synthetase (PylRS) is a major tool in genetic code expansion with non-canonical amino acids, yet understanding of its structure and activity is incomplete. Here we describe the crystal structure of the previously uncharacterized essential N-terminal domain of this unique enzyme in complex with $tRNA^{Pyl}$. This structure explains why PylRS remains orthogonal in a broad range of organisms, from bacteria to humans. The structure also illustrates why $tRNA^{Pyl}$ recognition by PylRS is anticodon-independent; the anticodon does not contact the enzyme. Using standard microbiological culture equipment, we then established a new method for laboratory evolution – a non-continuous counterpart of the previously developed phage-assisted continuous evolution. With this method, we evolved novel PylRS variants with enhanced activity and amino acid specificity. We finally employed an evolved PylRS variant to determine its N-terminal domain

*Correspondence should be addressed to Dieter Söll: dieter.soll@yale.edu.
[f]Current address: DS Therapeutics, Houston, TX, 77056, USA
[g]Current address: Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan
†These authors contributed equally to this work.
Correspondence and requests for materials should be addressed to D.S.

Author Contributions

T.S. purified and crystallized the PylRS•$tRNA^{Pyl}$ complexes, solved structures, and analyzed data. C.M. designed the PANCE research, performed experiments, analyzed data, and wrote the manuscript. D.S. designed and supervised the research and wrote the manuscript. D.R.L. designed and supervised the research and edited the manuscript. L.T.G. designed the chimeric chPylRS variant for evolution in PANCE, performed protein purification and *in vitro* aminoacylation assays, and analyzed data. J.M.H. assisted with design and refinement of PANCE procedure. D.I.B. established initial selection conditions, performed read-through assays, and performed western blot analyses. All authors contributed to editing the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Methods-Only References

All references cited in the Online Methods appear in the main text.

Competing Financial Interests

The authors intend to file a patent application on the PANCE system.

Any supplementary information, chemical compound information and source data are available in the online version of the paper.

structure and show how its mutations improve PylRS activity in the genetic encoding of a non-canonical amino acid.

Pyrrolysyl-tRNA synthetase is an aminoacyl-tRNA synthetase (aaRS) found in a small group of archaeal and bacterial species[1]. Together with its cognate substrate, $tRNA^{Pyl}$, PylRS variants have profoundly advanced our ability to genetically encode non-canonical amino acids (ncAAs) in live cells[2–4]. Thus, in the past few years, natural and engineered PylRS variants enabled the encoding of more than 100 ncAAs[2] and improved encoding of its cognate substrate, pyrrolysine (Pyl, **1**)[4] (Supplementary Results, Supplementary Fig. 1) in a wide variety of species, ranging from bacteria to human[5,6]. The PylRS/$tRNA^{Pyl}$ pair is commonly used to encode ncAAs due to the following features: (i) PylRS does not use the $tRNA^{Pyl}$ anticodon as an identity element for recognition[7], hence the $tRNA^{Pyl}$ anticodon can be altered without loss of PylRS recognition[7,8], (ii) PylRS is highly polyspecific and can utilize several distinct classes of ncAAs[2,9,10], and (iii) the PylRS/$tRNA^{Pyl}$ pair is orthogonal (*i.e.,* does not cross-react with the host tRNAs and aaRSs) both in bacterial and eukaryotic species[4]. For over a decade, PylRS engineering has remained a key strategy in expanding the chemistry of protein synthesis, however our understanding of PylRS activity is still incomplete and the structural analysis of PylRS remains challenging.

In the archaeal genus *Methanosarcina*– one major source of PylRS variants for genetic code expansion – PylRS comprises 419–530 amino acids that are organized in two conserved domains connected by a variable linker (Supplementary Fig. 2): the tRNA-binding[11] N-terminal domain (~120 aa), and the C-terminal domain (~270 aa) that also binds tRNA and harbors the catalytic site. In bacteria, PylRS is encoded by two different *pylS* genes[4,11,12]: *pylSn* comprises the N–terminal ~120 amino acids, while *pylSc* encodes the C–terminal ~280 amino acids. It was realized early that the full-length archaeal enzyme is fairly insoluble (Supplementary Table 1) and refractory to crystallization, although the C-terminal domain (CTD) of *Methanosarcina mazei* PylRS could be crystallized[13]. The CTD structure was determined in the apo form and in complexes with various ligands and ncAA substrates[9,14], and with $tRNA^{Pyl}$ (ref. 15) that has enabled the rational design of the PylRS amino acid binding pocket. However, the structure of the N-terminal domain so far has remained elusive, despite this domain being indispensable for PylRS activity *in vivo*[12].

Here we report the crystal structure of the *M. mazei* PylRS N-terminal domain in complex with its cognate substrate, *M. mazei* $tRNA^{Pyl}$. We use this structure to gain insights into understanding PylRS specificity to its cognate tRNA, and to interpret the improved activity of multiple PylRS variants that were evolved through phage-assisted non-continuous evolution, a new method developed in this study.

RESULTS

Structure of the PylRS N-terminal domain–$tRNA^{Pyl}$ complex

To gain insight into the PylRS N-terminal domain (NTD) structure and activity, and enable its rational design, we co-crystalized the *M. mazei* PylRS N-terminal fragment (101 aa) with the transcript of *M. mazei* $tRNA^{Pyl}$, and determined the structure at 2.4 Å resolution (Fig. 1,

Supplementary Fig. 3). We found that the PylRS N-terminal domain folds into a compact protein globule that coordinates a zinc ion (Fig. 1a). This Zn^{2+} ion does not directly contact the $tRNA^{Pyl}$ and appears to stabilize the NTD fold. Our search for similar folds by using the DALI database indicated that none of the aaRSs has a similar fold to those of the PylRS NTD.

The PylRS NTD forms extensive contacts with the $tRNA^{Pyl}$ by fitting snugly into the concave comprised of the T-loop and the authentic minimal variable loop of $tRNA^{Pyl}$ (Fig. 1b). This tight fit between the PylRS NTD and the variable loop of $tRNA^{Pyl}$ provides a steric explanation for the PylRS orthogonality; the larger variable arm of canonical tRNAs would impede their productive binding to PylRS (Fig. 2). Also, the structure reveals H-bond interactions of the PylRS N–terminal residues K3 and H24 with $tRNA^{Pyl}$ and shows coordination of Zn^{2+} by residue H24, thereby explaining why mutations of these residues alter PylRS activity[12] (discussed below).

Remarkably, the N- and C-terminal domains of PylRS bind on the opposite sides of the $tRNA^{Pyl}$ molecule, suggesting that the complete archaeal enzyme wraps around the tRNA molecule (Fig. 1). This cooperative binding may explain why PylRS has such a high specificity for $tRNA^{Pyl}$ and minimal cross-reactivity with other tRNAs in a wide variety of species[15]. Also, the $tRNA^{Pyl}$ anticodon has no contacts with either N- or C-terminal PylRS domains (Fig. 1)[7,8]. This is consistent with biochemical and genetic studies showing that $tRNA^{Pyl}$ anticodon can be mutated and utilized by PylRS to encode ncAAs with UAG, UAA, UGA, CUG, AGU, and AGGA codons[2,8].

PANCE, a new method for directed protein evolution

Simultaneously with the structure determination, we sought to evolve PylRS variants with improved recognition of non-canonical amino acids. For this purpose, we modified the effective phage-assisted continuous evolution (PACE) method[16], which does not require prior structural knowledge for protein evolution and has been successfully applied to evolution of such proteins and enzymes as polymerases[16], receptor binding proteins[17], proteases[18], and aaRSs[19]. One cornerstone of this technique is its customized continuous flow machine, uniquely utilizing a two-chambered system which houses *Escherichia coli* cells infected with M13 bacteriophage inside a secondary ‘lagoon’ vessel. Given our lack of a continuous flow machine, and because non-continuous experiments have been shown to produce comparable results[17], we developed phage-assisted non-continuous evolution (PANCE), a simplified technique for rapid *in vivo* directed evolution using serial flask transfers in standard laboratory equipment (Fig. 3). This approach serially transfers evolving ‘selection phage’ (SP), which contain a gene of interest to be evolved, across fresh *E. coli* host cells, thereby allowing genes inside the host *E. coli* to be held constant while genes contained in the SP are continuously evolving. Serial flask transfers have long served as a widely-accessible approach for laboratory evolution of microbes, and more recently analogous approaches have been developed for bacteriophage evolution[20,21].

Similar to PACE, selection by PANCE relies on linking activity of the evolving gene of interest inside the SP to the infectivity of the progeny phage (Fig. 3a, Online Methods). Here we achieved this linkage by first inserting *M. mazei* $tRNA^{Pyl}_{am}$ and *gIII* containing one or

more UAG stop codons into an accessory plasmid (AP) inside the *E. coli* host. This AP contains genes necessary for selection, which are maintained within the host cell to prevent the acquisition of escape mutations. The essential *gIII* gene was then deleted from the SP and replaced with gene of interest *chpylS*, and the ncAA Boc-lysine (BocK, **12**) (Supplementary Fig. 1) was supplied during transfection. Production of full-length gIII, and phage survival, is thereby linked to the aminoacylation activity of the evolving chPylRS (Supplementary Fig. 4).

We first used PANCE to evolve a PylRS for improved translational incorporation of BocK (Supplementary Fig. 1). It has been shown earlier that PylRS mutagenesis frequently decreases PylRS affinity for Pyl and increases affinity for diverse ncAAs[9,10] (Supplementary Fig. 5); we therefore expected similar changes to occur in PylRS after PANCE evolution. As the PylRS ancestor, we designed a novel chimeric PylRS (chPylRS, encoded by *chpylS*) in which the *M. barkeri* PylRS N–terminal domain (1–149 aa) was fused to the *M. mazei* PylRS C-terminal domain (185–454 aa) (Supplementary Table 2). This chimera was selected due to its improved solubility *in vitro* compared to the solubility of non-chimeric proteins (Supplementary Table 1). Three independent PANCE lines containing chPylRS (termed lines A, B, and C) were passaged through a host *E. coli* strain containing an increasing number of UAG codons in *gIII*. Initially, SPs were unable to grow when *gIII* contained more than one UAG codon. However, following 18–21 transfers in the presence of an increased mutation rate, each lineage has acquired the capacity to survive in the presence of even three UAG codons (Fig. 3b).

Evolved PylRS variants have improved enzymatic properties

Sequencing of *chpylS* genes from individual plaques following growth in the presence of mutagenesis plasmid MP6 (ref. [22]) revealed a highly polymorphic population. While MP6 was shown to be critical in enabling growth of SP in higher stringency conditions following relatively few serial transfers (**Online Methods**), we reasoned that continued growth under mutagenic conditions was no longer necessary after higher activity chPylRS variants had arisen. Thus, after 1–6 additional passages in the absence of MP6 (**Online Methods**), successful mutations became fixed within each lineage as each population converged upon a clonal genotype; as a result, three mutant chPylRS variants named 32A, 24B, and 25C were isolated (Fig. 3b). The mutations in each variant are listed in Supplementary Table 2. Six mutations were observed in the NTD, which is responsible for tRNA binding[11], with only mutation H62Y appearing in all evolved variants. Five mutations were found in the C-terminal domain (CTD). In addition, we created the 32A-Nter variant that corresponds to wild-type chPylRS but carries only the four major mutations (D2N, K3N, T56P, H62Y) of the NTD. Read-through analysis of sfGFP containing one or three amber stop codons confirmed increased activity of evolved chPylRS variants in live cells (Supplementary Fig. 6a,b).

Western blot analysis (Supplementary Fig. 7a,b) of *chpylS* gene expression revealed that wild-type *chpylS* initiates translation at two sites, AUG codon 1 and AUG codon 107, to form a full-length protein (419 aa) and a chPylRS C-terminal protein (313 aa) in equal amounts[23,24]. ESI-MS analysis of the latter confirmed translation initiation with M107

(Supplementary Fig. 8). Sequencing of the chPylRS variants 24B and 25C revealed deletion of a T residue (Δt293) from codon 98 or 99 in *chpylS* (Supplementary Note), resulting in a –1 frameshift causing chain termination at the UGA codon 103. Western blot analysis (Supplementary Fig. 7b) of variant 24B and 25C gene expression revealed production of split PylRS enzymes consisting of an N-terminal (102 aa) protein and a C-terminal (313 aa) protein. SfGFP expression (Supplementary Fig. 6) demonstrated that the evolved split chPylRS enzymes (24B and 25C) are more active than the ancestral chPylRS. However, the NTD is not active by itself (Supplementary Fig. 6c)[19]. These data are in line with the fact that the NTD is required for *in vivo* PylRS activity[12] and that mutations in the NTD enhance *pylS*-dependent UAG 15 (ref. [25]). Since the active site is located in the CTD[14,26] one must assume that the communication of the presence of both domains occurs through the tRNA. Supplementary Figure 6c also shows the importance of mutations (S158N, G343D) endowing the 24B CTD with substantial *in vivo* activity, which the wild-type chPylRS CTD lacks[19]. This may be explained by increased $tRNA^{Pyl}$ binding caused by the S158N mutation[15] (Supplementary Fig. 9). The existence of a CTD that is catalytically active *in vivo* adds credence to the bioinformatic and proteomic discovery of isolated *pylSc* genes in diverse archaeal species[1,27].

As the mutations in the evolved PylRS variants were mainly found in the NTD, we co-crystallized the mutant *M. mazei* N-terminal domain (32A NTD) with $tRNA^{Pyl}$ and explored how the mutations affect PylRS contacts with $tRNA^{Pyl}$. The structure was determined at 2.8 Å resolution (Fig. 1c). It revealed that two of the PylRS mutations, which are located directly at the PylRS/tRNA interface, appear to weaken PylRS/tRNA contacts. Thus, in the wild-type complex (Fig. 1b) the K3 side chain forms two H-bonds with the T-loop of the tRNA molecule, whereas in the mutant complex the K3N mutation disrupts these H-bonds (Fig. 1c). Furthermore, the H62Y mutation disrupts another two H-bonds between the PylRS and T-loop moieties of $tRNA^{Pyl}$, although it establishes new and seemingly weaker contacts with a phosphate moiety of G21 and the base of A20. Collectively, this mutant structure suggested that the PANCE-evolved NTD mutations decrease the apparent affinity of the PylRS N-terminal domain for cognate $tRNA^{Pyl}$.

PylRS is less catalytically efficient than the canonical aaRSs (*e.g.,* leucyl-tRNA synthetase, LeuRS). This is plausible, as PylRS needs to provide much less Pyl-tRNA (servicing ~50 codons per *Methanosacina* genome) compared to Leu-tRNA (for ~220,000 codons in *E. coli*). Additionally, altering the enzyme's Pyl binding site in order to recruit diverse ncAA substrates (*e.g.,* Supplementary Fig. 5) leads to decreased k_{cat}/K_M for the cognate amino acid Pyl and increased k_{cat}/K_M for many ncAAs[9,10,28]. To better understand how mutations in the evolved PylRS variants improve specific PylRS activity, we performed kinetic analysis to measure reaction rates and PylRS affinity to $tRNA^{Pyl}$, Pyl and BocK (Table 1). For this purpose, we carried out aminoacylation assays[29] using purified PylRS mutants 32A, 24B, and 25C and *in vitro* transcribed $tRNA^{Pyl}$. Our measurements demonstrated that the improvement of each evolved variant stems from two major factors: a marked increase in PylRS affinity for BocK and even greater increase in PylRS affinity for $tRNA^{Pyl}$ (Table 1). Collectively, these changes resulted in up to 10-fold increased catalytic efficiency of PylRS, as measured by K_{cat}/K_m ratio (Table 1).

For one of the PylRS variants, 32A, we also measured its ability to discriminate between Pyl and BocK (Table 1). We found that, compared to the wild-type chPylRS, the evolved PylRS mutant had a markedly lower affinity for Pyl, reflected in a ~500 fold increase of the K_M value. At the same time, its affinity to BocK was higher than that of the parental protein, which is reflected in a ~5-fold decrease of K_M for BocK (Table 1). These measurements showed that PANCE evolution substantially improved PylRS specificity to BocK. The evolution of PylRS variants with lower K_M values for ncAAs may ease the challenge of *in vivo* toxicity by certain ncAAs[30].

Next, we estimated the individual contribution of the mutated NTD to the overall increase of PylRS activity. For this purpose, we used the 32A-Nter PylRS variant (Supplementary Table 2) that has the catalytically active CTD of the wild-type protein sequence. Kinetic measurements revealed that 32A-Nter mutant has profoundly weakened apparent affinity to $tRNA^{Pyl}$ (Table 1), consistent with the reduced contacts observed in the crystal structure (Fig. 1). This finding is notable, because it implies that the decreased apparent affinity of the mutated N-terminal domain to $tRNA^{Pyl}$ is effectively counterbalanced by additional mutations in the C-terminal domain of the evolved PylRS variants. Indeed, 32A-Nter variant has ~6 times lower $tRNA^{Pyl}$ affinity, while the 32A variant has ~10 higher $tRNA^{Pyl}$ affinity compared to the parental enzyme (as measured by K_M values); and the only difference between these two PylRS variants is the presence of three additional mutations (E119K in the NTD, and K258Q and Y349F in the C-terminal domain of 32A). Collectively, these kinetic observations indicate that PylRS improvement by PANCE is achieved in part by shifting the burden of tRNA recognition via weakening the N-terminus/$tRNA^{Pyl}$ interaction and reinforcing $tRNA^{Pyl}$ binding to the CTD. It is therefore not surprising that one of the evolved mutants, PylRS 24B, acquired a capacity to function *in vivo* in the absence of the N-terminal domain (Supplementary Fig. 6c).

While measuring the kinetic properties of 32A-Nter variant (Table 1), we could not escape noticing that mutations in the PylRS NTD influence the PylRS catalytic site. Thus, mutations of the NTD in the 32A-Nter variant decreased the K_M value for BocK and increased it for Pyl. Taken together, this suggests a model wherein mutations in the N-terminal domain enable a more flexible protein structure, coupled with C-terminal domain alterations to position $tRNA^{Pyl}$ into an orientation specifically conducive to BocK aminoacylation. This demonstrates that binding of $tRNA^{Pyl}$ and amino acid are intrinsically interdependent, consistent with findings for other aminoacyl-tRNA synthetases[31,32].

Analysis of previous crystal structures suggests that the mutations that we observed in the CTD of chPylRS affect direct contacts with both $tRNA^{Pyl}$ and the amino acid substrate. Mutation Y349F, identified in PANCE variant 32A (Supplementary Table 2), was previously shown (as Y384F in *M. mazei*) to improve aminoacylation with BocK (and other ncAAs) by *M. mazei* PylRS[26] likely through direct contacts with the amino acid[14]. Other mutations in the C-terminal domain appear to improve tRNA binding (Supplementary Fig. 9).

DISCUSSION

Here we report the crystal structure of the N-terminal domain of *M. mazei* PylRS captured in the complex with its cognate substrate, $tRNA^{Pyl}$. This structure provides mechanistic insights into the exceptionally high specificity of PylRS to $tRNA^{Pyl}$, and explains how changes in the PylRS N–terminal domain structure may be used to improve catalytic properties of this unique enzyme and hence facilitate genetic encoding of non-canonical amino acids. The key properties that make PylRS-$tRNA^{Pyl}$ highly orthogonal are: (i) discrimination against canonical tRNAs based on the larger size of the variable arm, (ii) demonstration that PylRS:tRNA recognition is 'anticodon blind', (iii) the enzyme may 'surround' the tRNA as the C- and N-terminal domains of PylRS have the largest interface area of the known 20 aaRS:tRNA complexes, and (iv) opening up the structure (Fig. 1) lowers binding to $tRNA^{Pyl}$ and leads to increased ncAA recognition albeit with lower catalytic activity[9,10].

We also show that amino acyl-tRNA synthetases can be efficiently evolved into better enzymes by using a non-continuous PANCE approach. The results of the PANCE experiment are consistent with the separately conducted PylRS improvement by PACE[19]. In the PACE experiment, most of the PylRS mutations were also found in the PylRS NTD, and led to increased catalytic activity of PylRS[19]. Furthermore, two mutations (H62Y and T56P) were found at the very same sites, although, in the PANCE approach, PylRS variant 25C had a T56A mutation instead of the T56P mutation observed from PACE. Frameshift mutations were also observed in PACE lineages, resulting in PylRS enzymes split into two separate proteins. Separate evolutionary lineages consistently evolved N–terminal fragments containing 91–101 aa and a C–terminal fragment containing 313 aa, which was generated by translation initiation at M107. This variability in the length of the PylRS NTD reflects differing points of termination within the naturally variable linker region, which likely no longer plays a role in protein function following protein splitting. Curiously, this 'splitting' of the two PylRS domains into separate proteins appears to mirror the natural expression pattern of *pylSc* and *pylSn* genes found in bacteria[11,12] and some archaeal lineages[1]. Comparing chPylRS mutations across different PANCE and PACE-derived lineages[19], the genotypic diversity produced following the same selection shows that many evolutionary trajectories can improve PylRS activity.

Since its development, PACE has provided a powerful tool for directed protein evolution. Here we show that the PANCE approach can produce comparable results within a rapid timeframe and without the need for continuous flow machinery, simplifying adoption of this successful protein engineering approach. Nonetheless, it is important to note that while the experimental apparatus required for PACE may require additional effort to set up, continuous flow experiments confer several advantages, such as a shorter experiment duration, greater constancy of selective pressure and population size throughout evolution[33], and facilitation of deeper study of the population dynamics of phage evolution, as many variables (such as population size, mutation rate, selection stringency) can be systematically altered to assess effects on adaptive outcomes[34]. By contrast, PANCE evolution entails constant flux between stationary and active growth phases, complicating such analyses.

Consequently, contrasting mutations between PANCE and PACE observed in this study may reflect adaptation to these distinct growth conditions.

In summary, we anticipate that this study may facilitate rational PylRS engineering to help resolve future challenges in expanding the chemistry of living systems.

ONLINE METHODS

Preparation of MmPylRS NTD

The DNA fragment encoding the initial 101 residues of PylRS from *M. mazei* (MmPylRS NTD) was cloned into the Nde I and Xho I sites of modified pET28b vector in which the thrombin site and extra residues upstream of the N-terminal His tag were deleted. The plasmid was transformed into *E. coli* BL21 (DE3). Cells were grown in LB medium containing 25 μg/ml kanamycin at 37°C until the A_{600} reached 0.6. The culture was then induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 100 μM and shifted to 25°C for approximately 16 h before harvesting. The cells were harvested and resuspended in buffer A [50 mM Tris-HCl (pH 8.5), 1200 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM 2-mercaptoethanol] with 0.5 mg/ml lysozyme, 0.1 mg/ml DNase. After sonication, the His_6-tagged protein was purified by immobilized metal-ion affinity chromatography using a Ni-NTA (Qiagen). The protein bound to the column was washed with buffer A containing 15 mM imidazole and eluted by buffer B [20 mM Tris-HCl (pH 8.5), 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM 2-mercapto-ethanol] containing 250mM imidazole. Eluted proteins were loaded onto a HiTrap Heparin HP column (GE Healthcare) and eluted with a gradient of 300–1000 mM NaCl in buffer B. Finally, the protein was loaded onto HiLoad 16/60 Superdex 200 pg (GE Healthcare) equilibrated with buffer C [20 mM Tris-HCl (pH 8.5), 200 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM DTT]. The protein was concentrated using an Amicon Ultra 10,000 MWCO (Millipore), flash-frozen and stored at −80°C.

Trial and error studies had shown that addition of Zn^{2+} led to increased solubility of the NTD. (This was later clarified by finding a Zn^{2+} ion bound near H24 in the MmPylRS NTD•$tRNA^{Pyl}$ structure). For selenomethionine labeled protein preparation, cells were grown in M9 minimal medium supplemented with 20 μM $ZnCl_2$ at 37°C until the A_{600} reached 0.4. Then, 60 mg/l L-selenomethionine, 100 mg/l each of L-Lys, L-Phe and L-Thr and 50 mg/l each of L-Ile, L–Leu and L-Val were added to the medium. When the A_{600} reached 0.7, the culture was induced by the addition of IPTG to a final concentration of 100 μM and shifted to 25°C for approximately 16 h before harvesting. MmPylRS NTD with D2N, K3N, T56P and H62Y mutations (32A NTD) was prepared in the same way as was WT MmPylRS NTD.

Preparation of $tRNA^{Pyl}$

As is customary with PylRS, $tRNA^{Pyl}$ transcripts were used in biochemical experiments. Thus, *M. mazei* $tRNA^{Pyl}$ was transcribed using T7 RNA polymerase as described previously[35]. Transcribed tRNAs were purified by a HiTrap DEAE FF column (GE

Healthcare) as previously described[36]. Pooled tRNAs were precipitated with isopropanol and dissolved in buffer D [20 mM Hepes-NaOH (pH 7.5), 10 mM $MgCl_2$].

Crystallization and data collection

MmPylRS NTD or 32A NTD were mixed with $tRNA^{Pyl}$ in a molar ratio of 1:1.2 in buffer E [20 mM Hepes-NaOH (pH 7.5), 50 mM NaCl,10 mM $MgCl_2$, 5% glycerol, 1 mM DTT]. The resulting mixture was concentrated to A_{260} of 200 by ultrafiltration. Crystallization experiments were performed with the sitting-drop vapor diffusion method at 19°C. Crystals of the MmPylRS NTD•$tRNA^{Pyl}$ complex were obtained in a reservoir solution containing 0.1 M Hepes-NaOH (pH 7.5), 0.2 M $MgCl_2$, and 15% PEG 3350. Crystals of the selenomethionine labeled MmPylRS NTD•$tRNA^{Pyl}$ were obtained in a reservoir solution containing 0.1 M Hepes-NaOH (pH 6.9), 0.2 M $MgCl_2$, and 21% PEG 3350. Crystals of 32A NTD•$tRNA^{Pyl}$ were obtained in the buffer containing 0.1 M Bis-Tris (pH6.7), 0.2 M $MgCl_2$ and 19% PEG3350. Crystals were cryo-protected with 30% xylitol in the reservoir solution. The diffraction data sets were collected at 100 K on beamline 24ID-C and 24ID-E of Advanced Photon Source. Collected data were indexed, integrated, scaled and merged using XDS[37].

Structure Determination and Refinement

The structure of MmPylRS NTD•$tRNA^{Pyl}$ complex was determined by single wavelength anomalous diffraction method. Selenium sites were identified and used for phasing by Phenix AutoSol[38] and initial model was build using Phenix AutoBuild[39]. This model was used for structural analysis of the native MmPylRS NTD•$tRNA^{Pyl}$ complex. After several cycles of refinement with the program phenix.refine[40], autoBUSTER[41], fitting of tRNA with NAFIT[42] and manual fitting with Coot[43], the R_{work}- and R_{free}- factors were converged to 21.6% and 24.2%, respectively (Supplementary Table 3). For the structure determination of 32A NTD•$tRNA^{Pyl}$, after rigid body refinement with the structure of MmPylRS NTD•$tRNA^{Pyl}$ complex using phenix.refine, the structure was rebuilt and modified manually using Coot. Molecules of tRNA were rebuilt and fitted by NAFIT. Then, the structure was further refined using phenix.refine. The R_{work}- and R_{free}- factors were converged to 20.6% and 24.9%, respectively (Supplementary Table 3).

We also prepared and solved a co-crystal structure of the 120 aa N-terminal fragment. The structure of this complex was the same as that of MmPylRS NTD•$tRNA^{Pyl}$; in both cases the residues after position 87 were disordered.

Purification of c-Myc-chPylRS-6xHis variants

The chPylRS variants were cloned into the pTech plasmid using insertion primers that incorporate the N-terminal c-Myc sequence (MEQKLISEEDL-) and the C-terminal 6xHis sequence (-GSHHHHHH). BL21 star (DE3) cells (Thermo Fisher Scientific) transformed with the appropriate pTech plasmids were grown in LB media (United States Biologicals) supplemented with 25 μg/mL chloramphenicol. For each variant, a saturated overnight culture was prepared from a single colony, and a 1:100 dilution of culture was made into 5 mL of fresh LB media containing chloramphenicol. The starter culture grew at 37°C while shaking at 230 rpm until the cell density reached A_{600} = 0.3. The starter culture was then

used to inoculate a 1 L culture of LB media containing chloramphenicol, which continued to incubate while shaking for an additional 16 h. Cells were harvested by centrifugation at 5,000*g* for 10 min at 4°C, and cell pellets were resuspended in lysis buffer [20 mM Tris (pH 7.4), 300 mM NaCl, 10 mM imidazole, and EDTA-free protease inhibitor cocktail (Roche)]. The cells were lysed by sonication on ice, and the crude extract was centrifuged at 15,000*g* for 15 min at 4°C. Lysates were loaded onto columns containing 2 mL of HisPur Ni-NTA resin (Thermo Fisher Scientific) that had been pre-washed with two bed-volumes of equilibration buffer. The resin was washed with 10 bed-volumes of wash buffer [20 mM Tris (pH 7.4), 25 mM imidazole, 300 mM NaCl] and protein was then eluted in 3 mL of elution buffer [20 mM Tris (pH 7.4), 250 mM imidazole, 300 mM NaCl]. The purified protein was dialyzed against 20 mM Tris (pH 7.4), 150 mM NaCl, 5 mM EDTA, 1 mM dithiothreitol. Purified protein was stored in 20% glycerol at −80 °C until analysis.

Western blot analysis of c-Myc-chPylRS-6xHis variants

Cell lysates (30 μL) of expressed protein were combined with 25 μL of XT Sample Buffer (Bio-Rad), 5 μL of 2-mercaptoethanol, and 40 μL water. The samples were heated at 70°C for 10 min and 7.5 μL of prepared sample was loaded per well of a Bolt Bis-Tris Plus Gel (Thermo Fisher Scientific). Precision Plus Protein Dual Color Standard (4 μL) Bio-Rad was used as the reference ladder. The loaded gel was run at 200V for 22 min in 1x Bolt MES SDS running buffer (Thermo Fisher Scientific). The gel was transferred to a PVDF membrane using the iBlot 2 Gel Transfer Device (Thermo Fisher Scientific). The membrane was blocked for 1 h at room temperature in 50% Odyssey blocking buffer (PBS) (Li-Cor) and was then soaked 4 times for 5 min in PBS containing 0.1% Tween-20 (PBST). The blocked membrane was soaked with primary antibodies [rabbit anti-6xHis (1:1,000 dilution) (Abcam, ab9108) and mouse anti-c-Myc (1:7,000 dilution) (Sigma-Aldrich, M4439)] in 50% Odyssey buffer (PBS) containing 0.2% Tween-20 for 4 h at room temperature. The membrane was washed 4 times in PBST, and then soaked for 1 h in the dark at room temperature with secondary antibodies [donkey anti-mouse 800CW (1:20,000 dilution) (Li-Cor) and goat anti-rabbit 680RD (1:20,000 dilution) (Li-Cor)] in Odyssey buffer containing 0.01% SDS, 0.2% Tween-20. The membrane was washed 4 times in PBST and finally rinsed with PBS. The membrane was scanned using an Odyssey Imaging System (Li-Cor).

LCMS analysis of intact purified proteins

Purified protein samples were diluted to 10 μM in dialysis buffer lacking reducing agent or glycerol prior to analysis on an Agilent 6220 ESI-TOF mass spectrometer equipped with an Agilent 1260 HPLC. Separation and desalting was performed on an Agilent PLRP-S Column (1,000A, 4.6 x 50 mm, 5 μm). Mobile the phase A was 0.1% formic acid in water and mobile phase B was acetonitrile with 0.1% formic acid. A constant flow rate of 0.250 mL/min was used. Ten μL of the protein solution was injected and washed on the column for the first 3 min at 5% B, diverting non-retained materials to waste. The protein was then eluted using a linear gradient from 5% B to 100% B over 7 min. The mobile phase composition was maintained at 100% B for 5 min and then returned to 5% B over 1 min. The column was then re-equilibrated at 5%B for the next 4 min. Data was analyzed using Agilent MassHunter Qualitative Analysis software (B.06.00, Build 6.0.633.0 with Bioconfirm). The charge state distribution for the protein produced by electrospray

ionization was deconvoluted to neutral charge state using Bioconfirm's implementation of MaxEnt algorithm, giving a measurement of average molecular weight. The average molecular weight of the proteins were predicted using ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/).

sfGFP Assay

A pTECH plasmid containing the AARS of interest and a pBAD plasmid containing the sfGFP of interest were cotransformed into chemically competent TOP10 cells (Thermo Fisher Scientific). The transformed cells recovered in SOC (New England Biolabs) for 1 h while shaking at 37°C and were then plated and grown overnight at 37°C on LB agar containing 100 μg/mL carbenicillin and 25 μg/mL chloramphenicol. Single colonies were used to inoculate 3 mL of LB media (United States Biologicals) containing antibiotics and were grown overnight at 37 °C while shaking at 230 rpm. The saturated overnight cultures were diluted 100-fold in a 96-well deep well plate using 500 μL of LB media containing the required antibiotic. The plate was shaken at 37 °C for 3 h at 230 rpm and an additional 0.5 mL of LB was added containing antibiotics and additional components to provide each well with a final concentration of 1 mM BocK where denoted and 1.5 mM arabinose to induce expression of sfGFP. The cultures incubated with shaking at 37°C for an additional 16 hr after induction of sfGFP, and 150 μL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar). The A_{600} and fluorescence (excitation = 485 nm; emission = 510 nm; bandwidth of excitation and emission = 5 nm) readings from each well were taken using an Infinite M1000 Pro microplate reader (Tecan). Background A_{600} and background fluorescence measurements were taken on wells containing LB media only. The background-subtracted fluorescence value from each well was divided by the background-subtracted A_{600} value of the same well to provide the fluorescence value normalized to cell density. All variants were assayed in biological quadruplicate, and error bars represent the standard deviation of the independent measurements.

Aminoacylation Kinetics

The aminoacylation of $tRNA^{Pyl}$ variants was carried out at 37°C in the buffer containing 50 mM HEPES-KOH (pH 7.2), 25 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 10 mM ATP, 10 mM amino acids, 100 nM PylRS variants, 24 μM unlabeled $tRNA^{Pyl}$, and 3.6 μM ^{32}P-labeled $tRNA^{Pyl}$ with a total volume of 25 μL. Various concentrations of BocK (0.1–12.8 mM), Pyl[44] (5–500 μM for chPylRS and 0.1–10mM for variant 32A), and tRNA (0.5–16 μM) were used to determine K_M values for corresponding substrates. A 2 μL aliquot was taken out from each of the reaction mixtures at the time points of 5 min, 20 min and 30 min, and the reactions were immediately quenched by adding 3 μL quenching solution [0.66 μg/μL nuclease P1 (Sigma) in 100 mM sodium citrate (pH 5.0)]. The nuclease P1 mixtures were then incubated at room temperature for 30 min and 1 μL aliquots were spotted on PEI-cellulose plates (Merck) and developed in the running buffer containing 5% acetic acid and 100 mM ammonium acetate. Radioactive spots for AMP and AA-AMP (representing free tRNA and aminoacyl-tRNA, respectively) were separated and visualized and quantified by phosphorimaging using a Molecular Dynamics Storm 860 phosphorimager (Amersham Biosciences). The ratio of aminoacylated tRNA to total tRNA was determined to monitor reaction progress.

General PANCE Methodology

In kind with PACE evolution, an initial linkage between activity of M13 phage gIII protein (encoded within an *E. coli* cell) and activity of the gene to be evolved (encoded inside M13 phage) must first be established, as previously described[16]. This linkage makes phage growth dependent on activity of the gene of interest (for chPylRS evolution, see Supplementary Fig. 4). A series of stringency conditions must then be established, such that the initial activity of the gene of interest enables growth during low-stringency conditions, while improved activity is required for growth under higher stringency conditions. For this study, the initial activity of chPylRS was sufficient to mediate read-through of *gIII* containing single UAG codon (*gIII*^{P29am}, see Supplementary Table 4), however additional activity was required to enable phage propagation with two or more UAG codons in *gIII.*

Prior to beginning PANCE adaptation, host *E. coli* culture is first grown to A_{600} = 0.3–0.5. This culture must contain a low-stringency AP which permits SP propagation, a plasmid mediating increased mutagenesis (such as such as MP6[22]), and is to be grown in 2xYT media with 20 mM glucose, 5 mM magnesium, and appropriate antibiotics. As plasmid MP6 is induced by arabinose, glucose supplementation is essential to reduce undesired mutagenesis prior to phage infection. *E. coli* cultures can be grown as a large batch in advance of phage infection and stored at 4°C, however storage is recommended for a maximum of 3–5 days as culture containing MP6 exhibits reduced infectivity following prolonged storage.

To begin PANCE adaptation, SP containing the gene to be evolved are first outgrown to a high titer (≥ 10^6 plaque forming units (PFUs)) in the absence of selection (e.g., using permissive strain S1059 containing WT *gIII*). The first selection growth is initiated by transferring an aliquot of infected culture at a volume of 5–200 μL containing a minimum of 5 x 10^6 PFUs to a flask containing 50 mL of *E. coli* culture in a 125 mL baffled flask (aliquoted from larger batch culture as described above). If reaching this population size requires more than 200 μL, the previous phage growth is to be repeated. Arabinose (5 mM) is added at this time to induce mutagenesis, as well as any other supplements cogent to the selection experiment (such as BocK, used in this study prepared at 100 mM concentration in 1N NaOH). Infected cultures are subsequently grown for 8–12 hr at 37°C with shaking at 225 rpm. Subsequently, a 1 mL aliquot is then stored at 4°C, and phage titers are measured using permissive host S1059. After phage growth is confirmed, an aliquot containing 5 x 10^6 PFUs is used to inoculate a subsequent culture, and this process is iteratively repeated until the desired phenotype is evolved.

Every 1–3 transfers, phage will be tested for growth under higher-stringency conditions by inoculating into an additional *E. coli* culture previously established to be non-permissive of ancestral SP phage growth. In the context of this experiment, *E. coli* containing plasmids pDB038a and pDB038b, encoding *gIII*$^{P29am/E84am}$ and *gIII*$^{P29am/E84am/Y183am}$ (respectively), served as higher stringency growth conditions, as they required read-through of an increasing number of stop codons (**Online Methods** and Supplementary Table 4). Each PANCE growth cycle is to be conducted at the highest stringency condition permissive to phage growth.

Following multiple PANCE growth cycles at the highest desired levels of stringency, additional high-stringency cycles are conducted in the absence of a mutagenic plasmid to allow for loss of neutral or deleterious mutations within the population. Individual phage plaques are isolated and sequenced to identify evolved mutations. Additional characterization of evolved protein can then be performed to better describe a biochemical basis for evolved phenotypes (see Table 1 and Supplementary Fig. 6 for examples from this work).

PANCE of chPylRS

E. coli strain S1030[45] containing plasmid MP6[22] and the desired accessory plasmid (either pDB038, pDB038a, or pDB038b, see Supplementary Table 4) was initially inoculated from a plate into a large volume (~600 mL) of 2xYT media containing 100 μg/mL spectinomycin, 35 μg/mL chloramphenicol, 20 mM glucose, and 5 mM magnesium chloride in a 1L baffled flask. Cells were grown at 37°C with shaking at 225 rpm until reaching A_{600} = 0.3 – 0.5, at which point the culture was stored at 4°C for 3–5 days.

Phage growth was then begun by first removing a 50 mL aliquot of cell culture from the larger flask, and transferring it to a sterile 125 mL baffled flask. This aliquot was then supplemented with 5 mM arabinose and 1 mM BocK, followed by transfection with a minimum inoculum of 5 x 10^6 PFUs of selection phage taken from the previous growth in the PANCE cycle (Prior to beginning the first phage growth in the PANCE cycle, phage were grown in strain S1030 containing pJC175e which supplies WT *gIII*, termed strain S1059[46], to reach the required population size). Transfected cells were then grown at 37°C with shaking at 225 rpm for 8–12 hr. A 1 mL aliquot of phage was removed from the flask and was used to inoculate the subsequent phage growth in the PANCE cycle. Phage samples were subsequently stored at 4°C.

For each transfection, a standard volume of 5 μL of infected *E. coli* cells was used. To ensure a minimum inoculum size of 5 x 10^6 PFUs of phage was reached during each transfection, a phage titer was performed for each sample grown using permissive strain S1059 as the host strain. For samples with a concentration below 10^6 PFU/μl, phage growth was repeated, and inoculation volume was increased to a maximum of 0.2 mL to reach a total inoculum size 5 x 10^6 PFU. If reaching this population size required more than 0.2 mL, the previous phage growth was also repeated. BocK stock was prepared at 100 mM concentration in 1N NaOH.

Use of mutagenesis plasmid MP6 in PANCE evolution

Growth with mutagenesis plasmid rapidly evolves the desired phenotype. Following 40 transfers across three independent lines with plasmid pDB038 in the absence of a mutagenesis plasmid, phage were unable to grow on pDB038a. Following 8–12 transfers on pDB038 in the presence of MP6, phage were able to grow with pDB038a (Fig. 3b). Propagation in the presence of selection (plasmid pDB038b) and in the absence of mutagenesis plasmid MP6 resulted in the loss of neutral and deleterious mutations from the population. Selection without mutagenesis plasmid MP6 was conducted until sequenced populations showed convergence. In line A, samples 27A – 32A were grown without MP6; these are the last 6 transfers. In line B, samples 23B–24B were grown without MP6; these

are the last two transfers. In line C, sample 25C was grown without MP6; this is the last single transfer.

Data Availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5UD5 and 5V6X. Selection plasmids used in this study will be available through Addgene. Other materials are available upon reasonable request from the corresponding author.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Sergey Melnikov (Yale University) for insightful discussions and intellectual contributions, Akira Shinoda (Paul Scherrer Institute) and Keitaro Yamashita (RIKEN) for advice on structure analysis, and Sunia Trauger (Small Molecule Mass Spectrometry Laboratory at Harvard University) for providing expertise with intact protein mass spectrometry analysis. This work was supported the U.S. National Institutes of Health (NIH) R01EB022376 and R35GM118062 (to D.R.L.), and R01GM022854 and R35GM122560 (to D.S.), by the Defense Advanced Research Projects Agency N66001-12-C-4207 (to D.R.L.), by the Department of Energy DE-FG02-98ER20311 (to D.S.), and the Howard Hughes Medical Institute. D.I.B is supported by the National Institutes of Health under a Ruth L. Kirschstein National Research Service Award (F32GM106621). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

References

1. Mukai T, et al. RNA-dependent cysteine biosynthesis in Bacteria and Archaea. MBio. 2017; 8:e00561–00517. [PubMed: 28487430]
2. Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta. 2014; 1844:1059–1070. [PubMed: 24631543]
3. Crnkovic A, Suzuki T, Söll D, Reynolds NM. Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. Croat Chem Acta. 2016; 89:163–174. [PubMed: 28239189]
4. Gaston MA, Jiang R, Krzycki JA. Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr Opin Microbiol. 2011; 14:342–349. [PubMed: 21550296]
5. Mukai T, et al. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun. 2008; 371:818–822. [PubMed: 18471995]
6. Chin JW. Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem. 2014; 83:379–408. [PubMed: 24555827]
7. Ambrogelly A, et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci U S A. 2007; 104:3141–3146. [PubMed: 17360621]
8. Ho JM, et al. Efficient rassignment of a frequent serine codon in wild-type *Escherichia coli*. ACS Synth Biol. 2016; 5:163–171. [PubMed: 26544153]
9. Yanagisawa T, Umehara T, Sakamoto K, Yokoyama S. Expanded genetic code technologies for incorporating modified lysine at multiple sites. Chembiochem. 2014; 15:2181–2187. [PubMed: 25179816]
10. Guo LT, et al. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. Proc Natl Acad Sci U S A. 2014; 111:16724–16729. [PubMed: 25385624]

11. Jiang R, Krzycki JA. 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. 2012; 287:32738–32746. [PubMed: 22851181]
12. Herring S, et al. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable *in vitro* but required for *in vivo* activity. FEBS Lett. 2007; 581:3197–3203. [PubMed: 17582401]
13. Yanagisawa T, Ishii R, Fukunaga R, Nureki O, Yokoyama S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2006; 62:1031–1033.
14. Kavran JM, et al. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci U S A. 2007; 104:11268–11273. [PubMed: 17592110]
15. Nozawa K, et al. Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ structure reveals the molecular basis of orthogonality. Nature. 2009; 457:1163–1167. [PubMed: 19118381]
16. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. Nature. 2011; 472:499–503. [PubMed: 21478873]
17. Badran AH, et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. Nature. 2016; 533:58–63. [PubMed: 27120167]
18. Dickinson BC, Packer MS, Badran AH, Liu DR. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat Commun. 2014; 5:5352. [PubMed: 25355134]
19. Bryson D, et al. Continuous directed evolution of aminoacyl-tRNA synthetases. Nat Chem Biol. 2017 aa-bb.
20. Meyer JR, et al. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science. 2012; 335:428–432. [PubMed: 22282803]
21. Hammerling MJ, et al. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat Chem Biol. 2014; 10:178–180. [PubMed: 24487692]
22. Badran AH, Liu DR. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. Nat Commun. 2015; 6:8425. [PubMed: 26443021]
23. Mogk A, et al. Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. J Biol Chem. 2003; 278:17615–17624. [PubMed: 12624113]
24. Yamamoto H, et al. 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. Proc Natl Acad Sci U S A. 2016; 113:E1180–1189. [PubMed: 26888283]
25. Owens AE, Grasso KT, Ziegler CA, Fasan R. Two-tier screening platform for directed evolution of aminoacyl-tRNA synthetases with enhanced stop codon suppression efficiency. Chembiochem. 2017; 18:1109–1116. [PubMed: 28383180]
26. Yanagisawa T, et al. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N^{ε}-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. Chem Biol. 2008; 15:1187–1197. [PubMed: 19022179]
27. Sorokin DY, et al. Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. Nat Microbiol. 2017; 2:17081. [PubMed: 28555626]
28. O'Donoghue P, Ling J, Wang YS, Söll D. Upgrading protein synthesis for synthetic biology. Nat Chem Biol. 2013; 9:594–598. [PubMed: 24045798]
29. Wolfson AD, Pleiss JA, Uhlenbeck OC. A new assay for tRNA aminoacylation kinetics. RNA. 1998; 4:1019–1023. [PubMed: 9701292]
30. Schachtele CF, Anderson DL, Rogers P. Mechanism of canavanine death in *Escherichia coli*. II. Membranes-bound canavanyl-protein and nuclear disruption. J Mol Biol. 1968; 33:861–872. [PubMed: 4882619]
31. Fan C, Ho JM, Chirathivat N, Söll D, Wang YS. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. Chembiochem. 2014; 15:1805–1809. [PubMed: 24890918]
32. Hong KW, et al. Transfer RNA-dependent cognate amino acid recognition by an aminoacyl-tRNA synthetase. EMBO J. 1996; 15:1983–1991. [PubMed: 8617245]

33. Gresham D, Dunham MJ. The enduring utility of continuous culturing in experimental evolution. Genomics. 2014; 104:399–405. [PubMed: 25281774]
34. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc Natl Acad Sci U S A. 2013; 110:9007–9012. [PubMed: 23674678]
35. Suzuki T, Yamashita K, Tanaka Y, Tanaka I, Yao M. Crystallization and preliminary X-ray crystallographic analysis of a bacterial Asn-transamidosome. Acta Crystallogr F Struct Biol Commun. 2014; 70:790–793. [PubMed: 24915095]
36. Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. RNA. 2010; 16:647–653. [PubMed: 20100812]
37. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]
38. Terwilliger TC, et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr D Biol Crystallogr. 2009; 65:582–601. [PubMed: 19465773]
39. Terwilliger TC, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr D Biol Crystallogr. 2008; 64:61–69. [PubMed: 18094468]
40. Afonine PV, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012; 68:352–367. [PubMed: 22505256]
41. Bricogne, G., et al. BUSTER version 2.10.2. Cambridge, United Kingdom: Global Phasing Ltd; 2016.
42. Yamashita K, Zhou Y, Tanaka I, Yao M. New model-fitting and model-completion programs for automated iterative nucleic acid refinement. Acta Crystallogr D Biol Crystallogr. 2013; 69:1171–1179. [PubMed: 23695261]
43. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
44. Wong ML, Guzei IA, Kiessling LL. An asymmetric synthesis of L-pyrrolysine. Org Lett. 2012; 14:1378–1381. [PubMed: 22394273]
45. Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR. Negative selection and stringency modulation in phage-assisted continuous evolution. Nat Chem Biol. 2014; 10:216–222. [PubMed: 24487694]
46. Hubbard BP, et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat Methods. 2015; 12:939–942. [PubMed: 26258293]

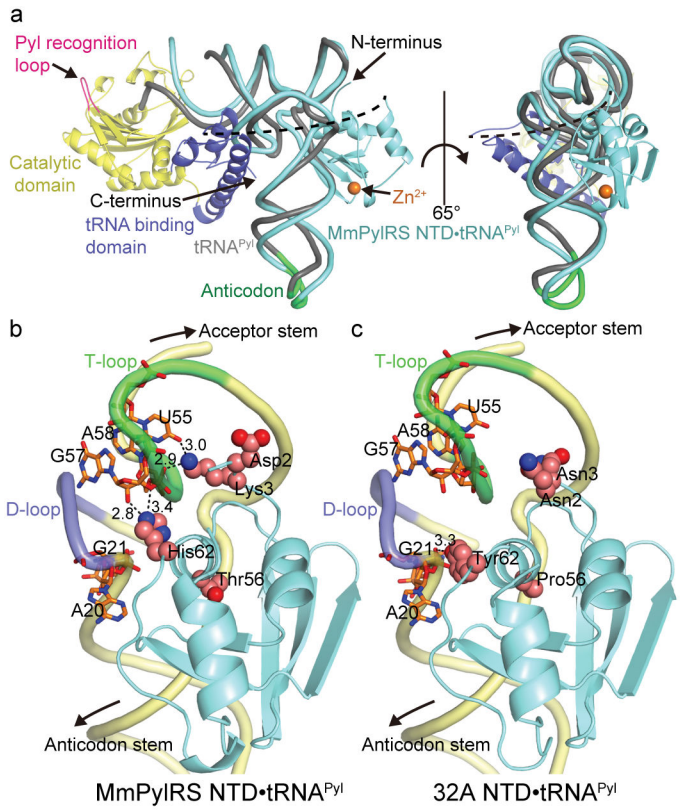

Figure 1. Crystal structures of the wild-type and PANCE-evolved PylRS variants bound to tRNA^{Pyl}

(**a**) Superposition of the MmPylRS NTD•tRNA^{Pyl} complex (cyan; this study) onto the *D. hafniense* PylRS CTD•tRNA^{Pyl} complex (grey; PDB ID: 2ZNI). The *M. mazei* and *D. hafniense* tRNA^{Pyl} species have the same cloverleaf secondary structure but differ in 32% of their bases. The phosphorus atoms of 70% of the tRNAs (nt 1–7, 26–44, 49–72 comprising the acceptor stem, T-stem/loop and anticodon stem/loop of tRNA^{Pyl}) were used for the superposition (rmsd=4.04 Å). Zn^{2+} is represented as an orange sphere. The PylRS CTD is composed of catalytic (yellow) and tRNA binding (blue) domains. The Pyl recognition loop is indicated in pink. The possible path of the linker connecting NTD to CTD is shown by a black dashed line; the linker varies between 63–157 amino acids, depending on the source of PylRS. (**b**) Close-up view of the interface between MmPylRS NTD and tRNA ^{Pyl}. (**c**) Close-up view of the interface between PANCE-evolved PylRS variant 32A NTD and tRNA ^{Pyl}. Mutations evolved in the chPylRS NTD were transplanted into MmPylRS NTD for structure determination. PANCE-evolved residues are represented as spheres while nucleotides interacting with the NTD of each PylRS variant are given as sticks. Polar interactions between PylRS residues and tRNA nucleotides are illustrated by a black dashed line. Interaction distances are given in Å.

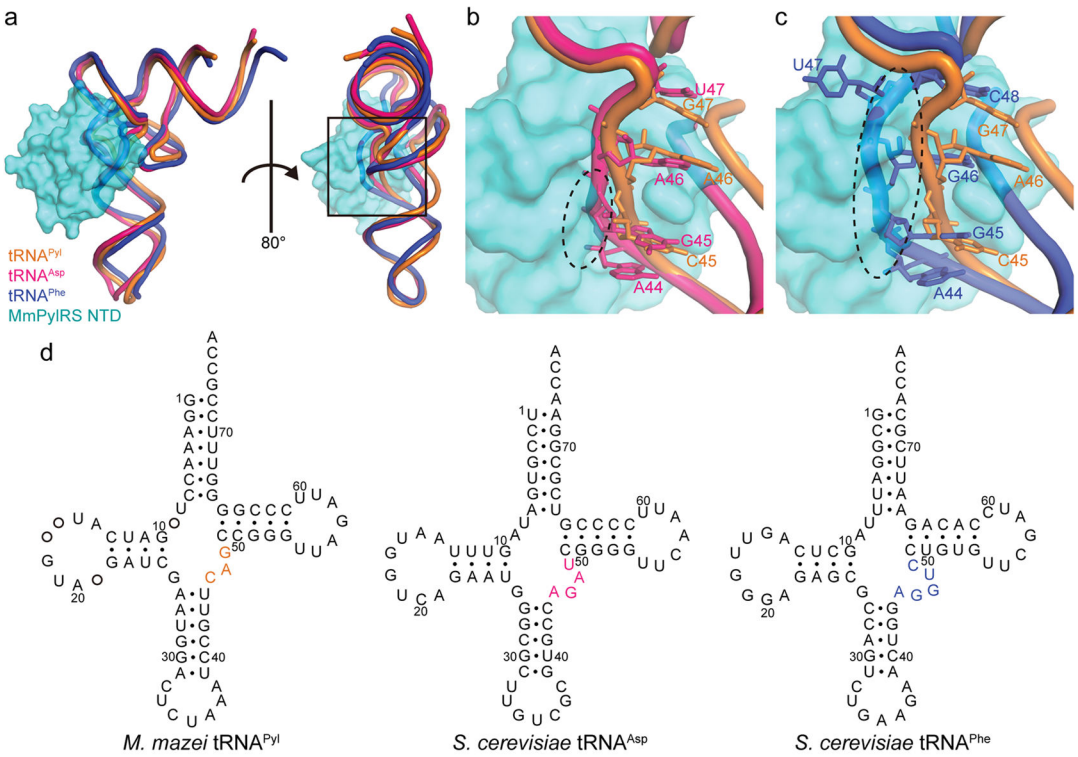

Figure 2. Structural basis for the PylRS specificity to tRNA^{Pyl}

(**a**) Superposition of the tRNA^{Asp} (PDB ID: 2TRA) and tRNA^{Phe} (PDB ID: 1EHZ) structures onto the tRNA^{Pyl} of the MmPylRS NTD•tRNA^{Pyl} complex was achieved by overlaying the phosphorus atoms of the acceptor stem, anticodon stem, and T–stem/loop (40 atoms for tRNA^{Asp}, rmsd=3.04, and 41 for tRNA^{Phe}, rmsd=3.21). MmPylRS NTD is shown as a surface model. (**b**) Close-up view of the variable loops of tRNA^{Asp} and tRNA^{Pyl} (A44-U47and C45-G47, respectively). (**c**) Close-up view of the variable loops of tRNA^{Phe} and tRNA^{Pyl} (A44-C48 and C45-G47, respectively). The steric clash between the MmPylRS NTD and the canonical tRNA is represented by a black dashed circle. Posttranscriptional modifications are omitted for clarity. (**d**) The cloverleaf structures of *M. mazei* tRNA^{Pyl}, *S. cerevisiae* tRNA^{Asp}, and *S. cerevisiae* tRNA^{Phe} show the different sizes of the variable loops.

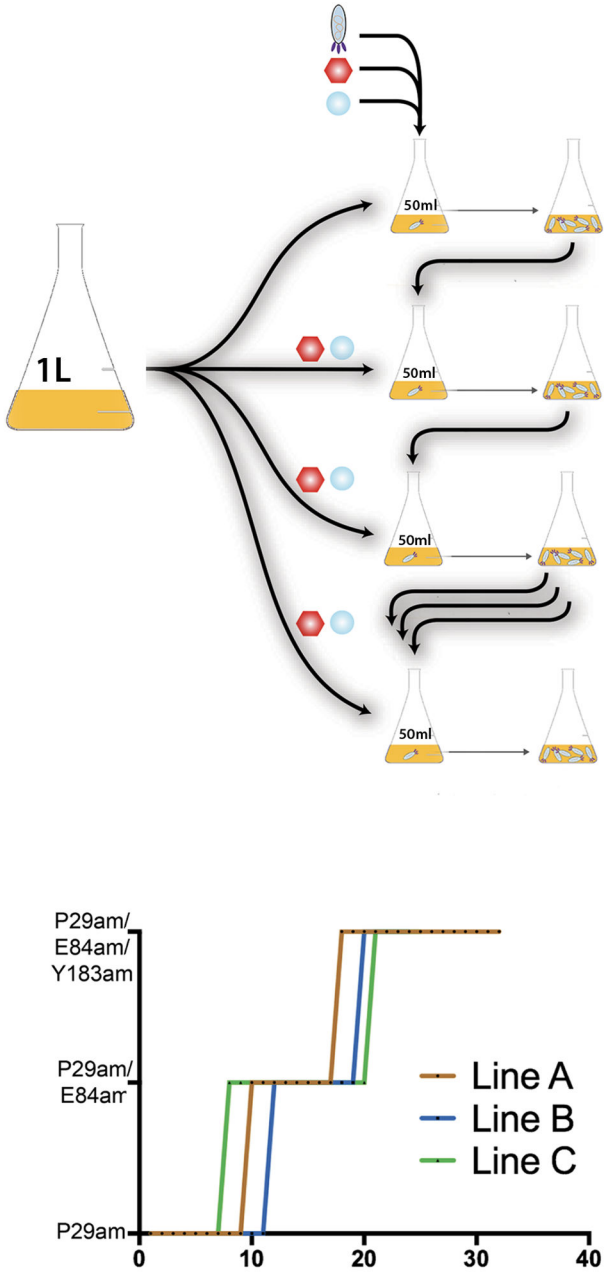

Figure 3. PANCE method application and overview

(**a**) The PANCE method begins by first growing the host strain of *E. coli* until A_{600} = 0.3–0.5 in a large volume, before storing the cells at 4°C. An aliquot of 50 mL is then transferred to a smaller flask, supplemented with BocK and the inducing agent arabinose (Ara) for mutagenesis plasmid MP6, and is transfected with the selection phage (SP). This culture is incubated at 37°C for 8–12 hr to facilitate phage growth, which is confirmed by determination of the phage titer. Following phage growth, an aliquot of infected cells is used to transfect a subsequent flask containing host *E. coli*. This process is continued until the desired phenotype is evolved, for as many transfers as required. (**b**) PANCE improved BocK incorporation by chPylRS. Three independent lines of SP containing chPylRS were serially

passaged across a host *E. coli* strain containing *gIII*. The number of UAG codons in *gIII* was steadily increased during serial passage in the presence of BocK. After 18–21 transfers, each lineage was able to survive with three UAG codons in *gIII* (P29am/E84am/Y183am) and grown in this environment for several more transfers to allow advantageous mutations to fix within the population. The lineages were named 32A, 24B, and 25C; the number denotes the number of transfers. For more details, see **Online Methods**.

Table 1

Kinetic properties of the evolved chPylRS variants.

Enzyme	ncAA	K_M (mM), ncAA	K_M (μM), tRNA	k_{cat} (10^{-3}·s^{-1})	k_{cat}/K_M^{AA} (mM^{-1}·s^{-1}·10^{-3})	k_{cat}/K_M^{tRNA} (μM^{-1}·s^{-1}·10^{-3})
chPylRS	Pyl	0.0076 ± 0.0002	12.2 ± 2.4	11 ± 1	1447	0.9
chPylRS	BocK	2.91± 0.10	19.9 ± 3.4	29 ± 0.06	10	1.5
32A-Nter	Pyl	0.053 ± 0.015	19.7 ± 3.3	18 ± 0.6	340	0.9
32A-Nter	BocK	0.95 ± 0.03	127 ± 22	23 ± 0.4	24	0.2
32A	Pyl	3.43 ± 0.44	3.53 ± 0.20	12 ± 4	3.5	3.4
32A	BocK	0.56 ± 0.04	1.83 ± 0.24	18 ± 0.3	32	9.8
24B†	BocK	0.77 ± 0.08	1.78 ± 0.11	20 ± 0.3	26	11
25C†	BocK	1.85 ± 0.58	0.86 ± 0.27	12 ± 1	6.5	14

Aminoacylation[29] of ^{32}P-labeled $tRNA^{Pyl}$ by purified chPylRS variants was done (up to 20 min) with varying concentrations of amino acid (BocK or Pyl) and $tRNA^{Pyl}$. The data were derived from three replicates.

†Enzymatic activity of 24B and 25C was determined on intact PylRS enzymes derived from *chpylS* clones from which the UGA codon and the subsequent eight bases were deleted to restore the original reading frame to AUG107 (Supplementary Fig. 7).