Taylor & Francis
Taylor & Francis Group

REVIEW

Check for updates

tRNA^Pyl: Structure, function, and applications

Jeffery M. Tharp, Andreas Ehnbom [iD], and Wenshe R. Liu [iD]

Department of Chemistry, Texas A&M University, College Station, TX, USA

ABSTRACT

Pyrrolysine is the 22nd proteinogenic amino acid encoded into proteins in response to amber (TAG) codons in a small number of archaea and bacteria. The incorporation of pyrrolysine is facilitated by a specialized aminoacyl-tRNA synthetase (PylRS) and its cognate tRNA (tRNA^Pyl). The secondary structure of tRNA^Pyl contains several unique features not found in canonical tRNAs. Numerous studies have demonstrated that the PylRS/tRNA^Pyl pair from archaea is orthogonal in *E. coli* and eukaryotic hosts, which has led to the widespread use of this pair for the genetic incorporation of non-canonical amino acids. In this brief review we examine the work that has been done to elucidate the structure of tRNA^Pyl, its interaction with PylRS, and survey recent progress on the use of tRNA^Pyl as a tool for genetic code expansion.

Introduction

Pyrrolysine (Pyl) was identified in 2002 as the 22nd proteinogenic amino acid.[1,2] This rare and highly specialized amino acid is found in a small number of methylamine metabolizing archaea as well as a few bacteria. In these organisms, Pyl is biosynthesized by three enzymes (PylB, PylC, and PylD) from two equivalents of lysine (Fig. 1A).[3] It is then directly encoded into proteins in response to an amber (TAG) codon with the help of a unique pyrrolysyl-tRNA synthetase (PylRS) along with its cognate tRNA (tRNA^Pyl, Fig. 1B, C).[4]

The PylRS/tRNA^Pyl pair is a particularly interesting aminoacyl-tRNA synthetase (aaRS)/tRNA pair for two reasons. First, understanding how Pyl incorporation has evolved may shed light on the origin of the genetic code. Whether Pyl emerged in highly specialized microbes as a recent addition to the code or is a relic of an earlier, more diverse genetic code is still unclear.[5,6] In either case, answering this question will, no doubt, advance our understanding of how the code has evolved into its present form. Second, the archaeal PylRS/tRNA^Pyl pair has been widely demonstrated to be orthogonal in bacteria and eukaryotes and this, coupled with the fact that tRNA^Pyl is a naturally occurring amber suppressor tRNA, has led to the widespread use of the PylRS/tRNA^Pyl pair for the genetic incorporation of non-canonical amino acids (ncAAs).[7-9] Several features of PylRS including high substrate side chain promiscuity and a high tolerance for mutations in the substrate binding pocket make it amenable for ncAA incorporation.[8] Indeed, wildtype and mutant PylRS variants in conjunction with tRNA^Pyl have been used for the genetic incorporation of over 100 ncAAs into proteins expressed in bacteria and eukaryotes[10] including mammalian cells.[11,12] This methodology has further been applied for the genetic incorporation of ncAAs into entire organisms including *Caenorhabditis elegans*,[13] *Drosophila melanogaster*,[14,15] and more recently *Mus musculus*.[16,17]

Most of the work toward expanding the genetic code using the PylRS/tRNA^Pyl system has focused on engineering PylRS for an expanded amino acid substrate spectrum. In this review however, we will focus on the unique features of tRNA^Pyl, including its structure, interaction with PylRS, and its role as a genetic code expansion tool. For a detailed review of PylRS as a genetic code expansion tool see ref. 8.

The structure of tRNA^Pyl

The structure of tRNA^Pyl from Archaea

Archaea of the family *Methanosarcinaceae* share high sequence identity with regards to their tRNA^Pyl.[18,19] The secondary structure of tRNA^Pyl was first described by Kryzcki and coworkers in their seminal paper on the discovery of tRNA^Pyl from *M. barkeri* (*Mb*-tRNA^Pyl).[1] *Mb*-tRNA^Pyl (Fig. 2, center structure) contains several unique secondary structures that are found in all tRNA^Pyl of the *Methanosarcinaceae*. Perhaps most distinct among these features are found in the variable loop and the anticodon stem. The shortened variable loop contains only three nucleotides. This is in contrast to most known class I and class II tRNAs which contain variable loop lengths of four to five, or greater than ten nucleotides, respectively.[20] Further, the anticodon stem of *Mb*-tRNA^Pyl contains six instead of the commonly observed five nucleotides. Other structural anomalies include a shortened linkage from two to one nucleotides between the acceptor and D stems, lack of a TψC sequence in the T loop, and lack of a GG sequence in the D loop—a sequence that is widely conserved among cytosolic tRNA.[1,21] In spite of these abnormalities, *Mb*-tRNA^Pyl is still predicted to assume the cloverleaf conformation similar to canonical tRNAs.

Although crystal structures of PylRS from *M. barkeri* (*Mb*-PylRS) and its mutants have been extensively reported, the
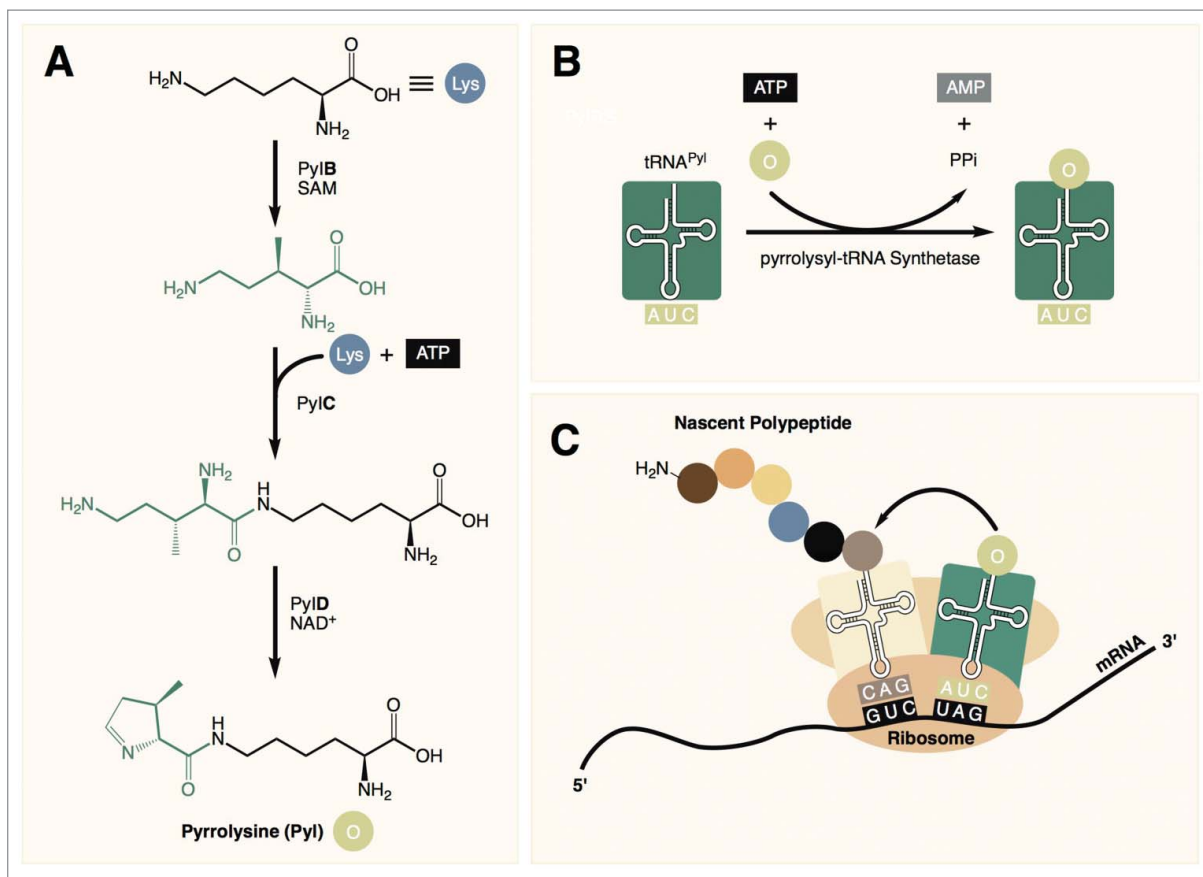
Figure 1. (A) Pyrrolysine is synthesized from 2 equivalents of lysine in 3 steps. (B) tRNA^Pyl is charged with pyrrolysine by pyrrolysyl-tRNA synthetase. (C) Pyrrolysine is incorporated into the nascent peptide in response to amber (UAG) codons during translation.
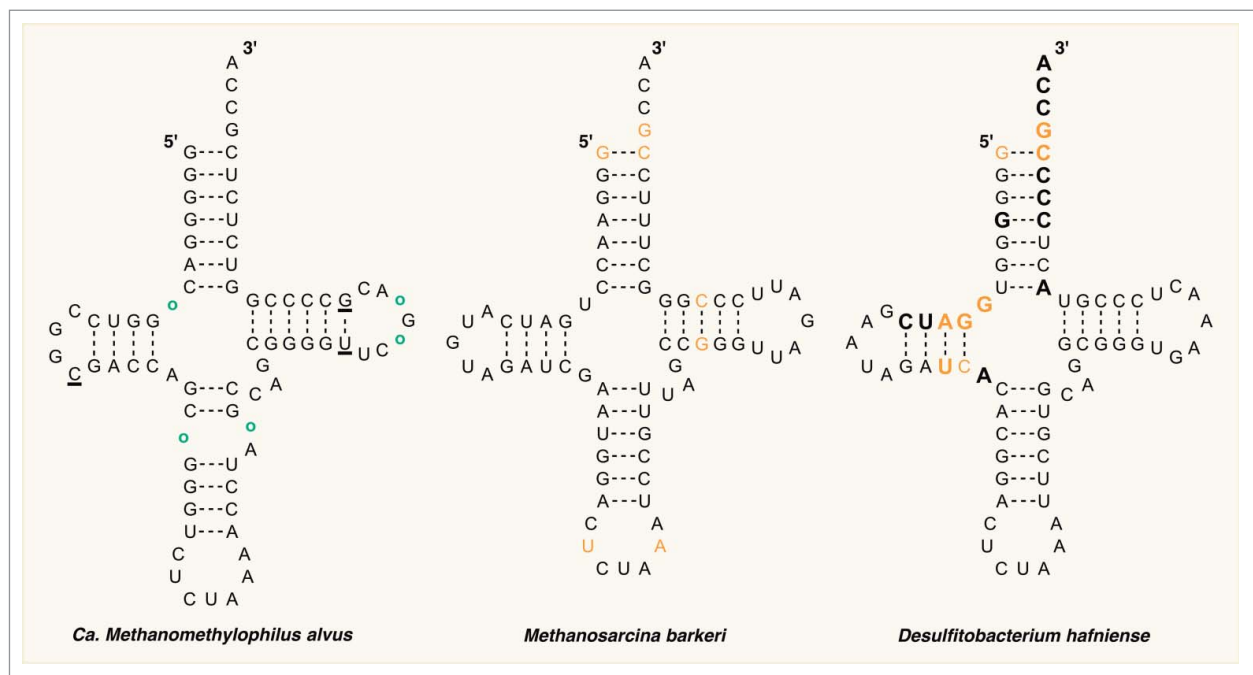
Figure 2. The secondary structure of tRNA^Pyl from *Candidatus Methanomethylophilus alvus*, *Methanosarcina barkeri* Fusaro, and *Desulfitobacterium hafniense*. Identity elements in the tRNA^Pyl from *M. barkeri* and *D. hafniense* are shown in orange. Bases in direct contact with the enzyme in the crystal structure of PylSc from *D. hafniense* are bold. In the structure from *Ca. M. alvus*, circles represent positions that are occupied by a base in the tRNA^Pyl from other *Methanomassiliicoccales* while underlined bases represent those that are missing.

crystal structure of Mb-tRNA^{Pyl} has not been determined.[22] Nonetheless, the tertiary structure of Mb-tRNA^{Pyl} was elucidated by Rudinger-Thirion and coworkers through a series of RNA footprinting and ultraviolet melting experiments.[23] Melting curves displayed only two transitions corresponding to the opening of the tertiary and secondary structures, indicating that the tRNA primarily adopts one defined tertiary structure in solution. The RNA cleavage profile with probes specific for single and double stranded sequences experimentally confirmed the previously predicted cloverleaf structure. Further, the profile was consistent with a relaxed tertiary structure which the authors proposed assumed the canonical L-conformation. A relaxed tertiary structure could be explained by the lack of nucleotide modifications found in mature tRNA^{Pyl}.[24] Indeed, Söll et al. demonstrated that mature tRNA^{Pyl} extracted from $M.$ $barkeri$ contained only two modifications, namely, 4-thiouridine at position 8 and 1-methyl-pseudouridine at position 50.[4] It is worth noting that the LC/ESI-MS experiments performed in this study to reveal nucleotide modifications are unable to detect pseudouridine due to the fact that the modified and unmodified nucleotides have the same mass. Thus the presence of pseudouridine modifications in Mb-tRNA^{Pyl} cannot be ruled out.

It has been noted that, with the aforementioned structural peculiarities, Mb-tRNA^{Pyl} bears a noticeable resemblance to bovine mitochondrial seryl-tRNA (Bt-tRNA^{Ser}) of Bos $taurus$.[4,23,25] The structural features shared by Mb-tRNA^{Pyl} and Bt-tRNA^{Ser} include both the shorted variable loop and elongated anticodon stem, as well as a shortened D loop and only one nucleotide between the acceptor and D stems.[26] The tertiary structure of unmodified Bt-tRNA^{Ser} has been solved through computer modeling, chemical probing, and solution NMR.[27,28] The data from these experiments support the conclusion that Bt-tRNA^{Ser} adopts a canonical L-shaped tertiary structure with a compact core region. This observation further supports the claim that although unusual in its secondary structure, the overall tertiary structure of tRNA^{Pyl} is similar to that of canonical tRNAs. Importantly, it has been noted that although Mb-tRNA^{Pyl} and Bt-tRNA^{Ser} share several unique structural elements, there is little similarity in their primary sequences.[25]

Recent evidence has supported the existence of a seventh order of methanogenic archaea related to the $Thermoplasmatales$ dubbed the $Methanomassiliicoccales$.[29,30] Analysis of the genomes of at least five identified members of this order has revealed the presence of genes encoding homologs of tRNA^{Pyl} and PylRS as well as all of the enzymes needed for Pyl biosynthesis.[5,31,32] Further, in all species identified, the gene encoding MtmB, a methyltransferase required for metabolizing monomethylamine, contains an in-frame amber codon. Taken together, these observations strongly support the direct genetic encoding of Pyl within the seventh order methanogens. Interestingly, the tRNA^{Pyl}s from this order (Mmc-tRNA^{Pyl}) contain their own set of unique features (Fig. 2, left structure).[5] Perhaps the most curious of these is a broken anticodon stem that forms a loop of five or seven bases, a feature not found in tRNA^{Pyl} from the $Methanosarcineae$ or in any bacterial tRNA^{Pyl}. Similar to Mb-tRNA^{Pyl}, the variable loop of Mmc-tRNA^{Pyl} is shortened and contains only three bases. The D loop is shortened further from five bases in Mb-tRNA^{Pyl} to four, or even three bases in some Mmc-tRNA^{Pyl}. Finally, the D and acceptor stems are separated by one or two bases or in some cases not at all.[5] While the identification of the Pyl incorporation cassette in the $Methanomassiliicoccales$ is exciting because of its ability to shed light on the evolutionary history of Pyl incorporation and its potential application as a genetic code expansion tool, given its relatively recent discovery, very little is known about the structure and function of tRNA^{Pyl} and PylRS from this group.

The structure of tRNA^{Pyl} from bacteria

At the same time that Mb-tRNA^{Pyl} was discovered to code for an in-frame amber codon in $M.$ $barkeri$, an investigation of available genomes revealed homologs of Mb-tRNA^{Pyl} and Mb-PylRS as well as several in-frame amber codons in the Gram-positive bacterium $Desulfitobacterium$ $hafniense$.[1] Since then, the Pyl encoding machinery has been discovered in 19 bacterial species including 16 Firmicutes and 3 delta-proteobacteria.[33] Given the taxonomic distribution of Pyl encoding organisms and the results of recent phylogenetic analyses,[5,33] it is believed that bacteria obtained the ability to incorporate Pyl as a result of a horizontal gene transfer from the $Methanomassiliicoccales$ who, in turn, inherited Pyl incorporation from a common ancestor with the $Methanosarcinaea$.[34]

Despite apparently obtaining the Pyl incorporation machinery from $Methanomassiliicoccales$, the secondary structure of tRNA^{Pyl} from $D.$ $hafniense$ (Dh-tRNA^{Pyl}) more closely resembles that of the $Methanosarcineae$ (Fig. 2, right structure). Although the latter two share very little sequence homology, their tRNA^{Pyl} contain many of the same structural elements including a lengthened anticodon stem, shortened variable region, D loop, and connection between the acceptor and D stems, and lack of several nearly universally conserved bases detailed above.[35] Bacterial tRNA^{Pyl} are, however, more diverse than the highly conserved sequences of the $Methanosarcineae$, sharing just under half of their sequence identity[18] and very few universally conserved residues are found among the known species.

The crystal structure of the catalytic domain of PylRS from $D.$ $hafniense$ (Dh-PylSc) bound to Dh-tRNA^{Pyl} was solved by Nureki and coworkers in 2009, offering the first direct evidence for the tertiary structure of tRNA^{Pyl} (Fig. 3A).[35] The crystal structure confirmed that the tRNA adopts a canonical L-shape along with a compacted core region (Fig. 3B). Both the L-shape and compact core were previously predicted based on the results of biochemical experiments and through comparing the structure of tRNA^{Pyl} with that of the known Bt-tRNA^{Ser}. The compact core of tRNA^{Pyl} is a result of nonstandard tertiary base pairing which arises primarily from the shortened D and variable regions as well as a deletion of U8. The U8:A14 bond is highly conserved among canonical tRNAs and is important for maintaining the L-shape.[36] Despite the compacted structure, the acceptor and anticodon stems reside in positions similar to those found in canonical tRNAs allowing tRNA^{Pyl} to function normally during translation.[35]

The interaction of tRNA^{Pyl} with pyrrolysyl-tRNA synthetase

The pyrrolysyl-tRNA synthetases from archaea and bacteria belong to the family of class II aaRSs.[1,35,37] The enzyme from
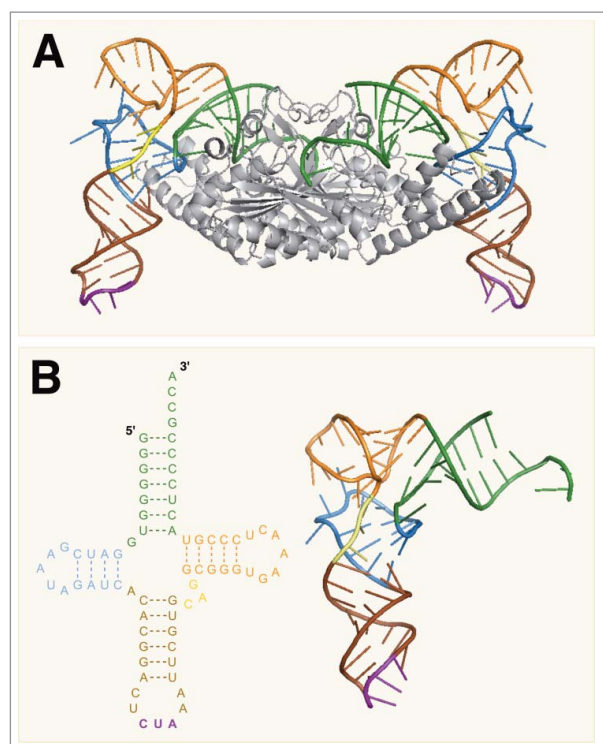
Figure 3. (A) The crystal structure of PylSc in complex with tRNA^Pyl from *Desulfitobacterium hafniense* (PDB: 2ZNI). (B) The crystal structure of tRNA^Pyl from *D. hafniense*. The secondary structure features of tRNA^Pyl are colored: green, acceptor stem; blue, D arm; brown, anticodon stem; purple, anticodon; yellow, variable loop; orange, T arm.

the *Methanosarcineae* contains an N-terminal domain connected by a highly variable linker to a C-terminal catalytic domain. In bacteria however, homologs of the C-terminal catalytic domain (PylSc) and N-terminal domain (PylSn) are coded for by two separate genes. The N-terminal domain of PylRS from *M. barkeri* (*Mb*-PylRS) and the homologous PylSn from *D. hafniense* represent novel RNA-binding domains as both have been shown to specifically bind tRNA^Pyl *in vitro* but share little sequence identity with known RNA-binding proteins.[38] In the absence of PylSn, PylSc is still capable of charging tRNA^Pyl with Pyl analogs *in vitro* and *in vivo* albeit with less activity than full-length archaeal *Mm*-PylRS.[25,35] Likewise, *Mb*-PylRS containing a truncation of the N-terminal domain is also capable of charging tRNA^Pyl *in vitro*, however, the truncated enzyme showed no activity *in vivo*.[39] Given that PylRS from both bacteria and archaea is still active *in vitro* without the N-terminal domain, but only archaeal PylRS requires this domain for *in vivo* activity, the exact biochemical function of this novel RNA binding domain remains elusive. It has been postulated based on the higher affinity of PylSn for tRNA^Pyl compared with PylSc ($K_D = 0.13\ \mu M$ and $6.9\ \mu M$, respectively), that this domain may serve to recruit tRNA^Pyl and facilitate acylation of tRNA^Pyl by the catalytic domain allowing the cell to maintain a lower basal level of tRNA^Pyl.[38,39] Such a role is yet to be supported by experimental evidence.

The catalytic domain of PylRS from bacteria and archaea contains an architecture that mirrors that of other class II aaRSs with the conventional fold of a β-sheet surrounded by several α-helices.[22,37,40,41] To investigate the interaction of the catalytic

domain with tRNA^Pyl, the crystal structure of *Dh*-PylSc was solved in complex with *Dh*-tRNA^Pyl by Nureki and coworkers (Fig. 4).[35] The structure revealed that like other class II aaRS, PylRS approaches tRNA^Pyl from the major groove side of the acceptor stem. An α-helix of the tRNA binding domain 1 and C-terminal tail of PylSc along with the bulge domain of the neighboring subunit in the PylSc dimer form a binding site that accommodates the acceptor helix of tRNA^Pyl. This interaction directs the acceptor helix to the catalytic site where, like other class II aaRS, the enzyme acylates the 3'-OH of the acceptor stem.[42] Along with the acceptor stem binding site, a core-binding surface for the tRNA is assembled from the tRNA-binding domain 1 and C-terminal tail of PylSc, as well as an α-helix from the neighboring subunit. In total, 31 residues of *Dh*-PylSc are involved in hydrogen bonding or stacking interactions with 17 bases in *Dh*-tRNA^Pyl (Table 1). Four of these residues are

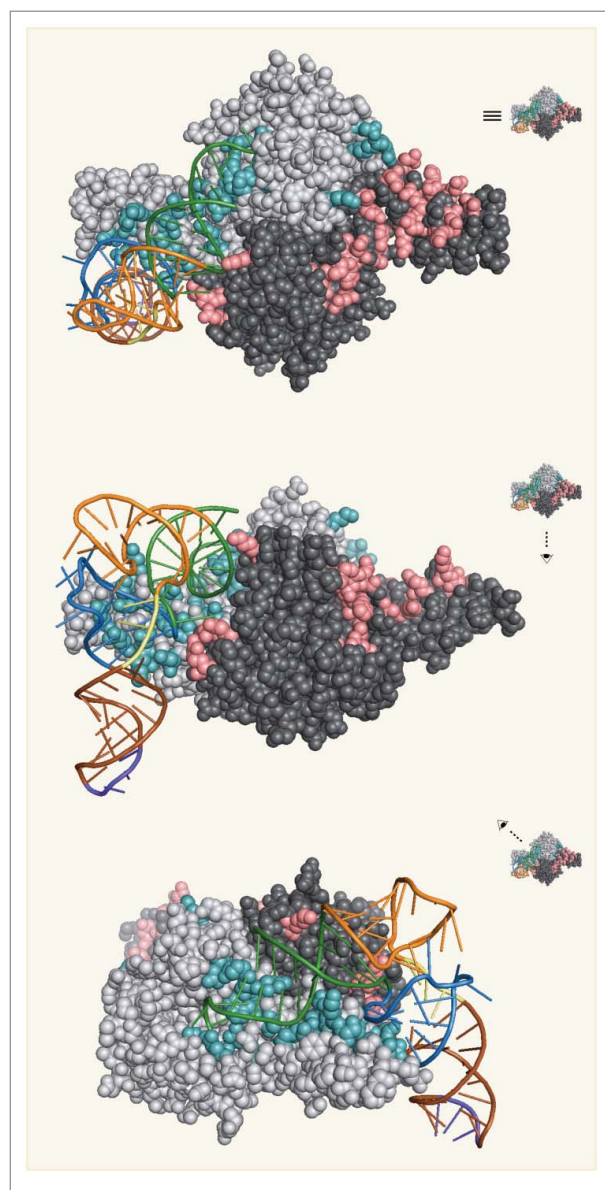**Figure 4.** The crystal structure of *Dh*-tRNA^Pyl bound to *Dh*-PylSc. The 2 protomers in the *Dh*-PylSc dimer are colored gray and black. Residues that interact with *Dh*-tRNA^Pyl in the crystal structure are shown in cyan and salmon, respectively. The colors corresponding to secondary structure features of *Dh*-tRNA^Pyl are the same as in Figure 3. For clarity, only one *Dh*-tRNA^Pyl is shown bound to the dimer. (PDB: 2ZNI).

Table 1. The interactions between bases in D*h*-tRNA[Pyl] and amino acid residues in
D*h*-PylSc. Unless noted, interactions involve main chain hydrogen bonds.

tRNA Region	tRNA[Pyl] Base	PylSc Residue
Acceptor Stem	G4	K16
		Q20
Accetor→D Stem	G9	R140[1,3]
		E245[1]
D Stem	G10	R21
		E50
		R144[1]
		N286[2]
D Stem	A11	E50
D Stem	U12	Q18
D Stem	C13	T15
		D43
D Stem	U24	Q47
D→Anticodon Stem	A26a	H51
		N288
Acceptor Stem	A66	K124[1]
Acceptor Stem	C69	S278
Acceptor Stem	C70	Y279
Acceptor Stem	C71	K23
		Q164
		H269
Acceptor Stem	C72	Q164[2]
		A166
Acceptor Stem	G73	E162[2]
		Q164[2]
Acceptor Stem	C74	Q117
		H168[3]
Acceptor Stem	C75	Q117
Acceptor Stem	A76	R160
		S163
		L169[2]
		F172[3]
		E229
		S232
		R259

[1]Residues are in the neighboring PylSc subunit of the dimer.
[2]Side chain hydrogen bond.
[3]Stacking interaction.

from the neighboring subunit of the PylSc dimer. As noted by Nureki et al. several of the interactions between PylRS and tRNA[Pyl] make use of structural elements that are unique among PylRSs. To accommodate the compacted core of tRNA[Pyl], PylRS itself has evolved a compacted binding site. This compacted binding site sterically occludes canonical tRNAs and explains the orthogonality of the PylRS/tRNA[Pyl] pair with respect to endogenous tRNAs and aaRSs.[35]

Biochemical studies measuring the activity of PylRS toward mutant tRNA[Pyl]s have identified several important identity elements for *Mb*-tRNA[Pyl] and *Dh*-tRNA[Pyl] (Fig. 2).[25,43] As with most tRNAs, the discriminator base (G73) and the first base pair (G1:C72) were identified as critical identity elements for tRNA[Pyl] from archaea and bacteria. These residues are highly conserved among Pyl incorporating organisms including those from the seventh order and were shown to make direct contact with PylSc in the crystal structure.[35] Other identity elements for *Dh*-tRNA[Pyl] include the G10:C25 and A11:U24 base pairs, and the base G9, all of which were shown to contact the enzyme in the crystal structure. For *Mb*-tRNA[Pyl] the base pair G51:C63 as well as U33 and A37, the two bases adjacent to the anticodon, were shown to be important identity elements. The bases adjacent to the anticodon do not make contact with PylSc in the crystal structure from *D. hafienses*, which led to the speculation that these bases may be important for the interaction of

tRNA[Pyl] with PylSn. However, more recent evidence has shown that U33G and A37C mutations did not detectibly affect binding of *Dh*-tRNA[Pyl] by PylSn in an electrophoretic mobility shift assay.[38] The fact that these residues are important for *Mb*-tRNA[Pyl] but not *Dh*-tRNA[Pyl] may represent differences in the tRNA[Pyl]/enzyme interaction between bacterial and archaeal organisms. The possibility of differing binding modes is underscored by the fact that while *Mb*-tRNA[Pyl] and *Dh*-tRNA[Pyl] are both substrates for *Mb*-PylRS, PylSc is unable to detectably acylate *Mb*-tRNA[Pyl].[39] With few exceptions, most tRNAs use the anticodon as a major identity element.[44] However, notably, neither *Mb*-tRNA[Pyl] nor *Dh*-tRNA[Pyl] show significantly decreased affinity for their respective aaRS or a reduction in acylation when mutations are made in the anticodon. As will be discussed in the following sections, this has major implications for the use of tRNA[Pyl] as a genetic code expansion tool.

Applications of tRNA[Pyl] for genetic code expansion

Engineered tRNA[Pyl] for genetic code expansion

As mentioned previously, the PylRS/tRNA[Pyl] pair has been widely adopted for the site-specific incorporation of a variety ncAAs across a range of host organisms. In general, ncAAs are incorporated into recombinant proteins expressed in the presence of a desired ncAA and in organisms bearing plasmid-borne copies of PylRS/tRNA[Pyl]. Schultz and coworkers have developed a powerful strategy for the directed evolution of aaRSs that can be used to identify novel aaRS mutants capable of incorporating new ncAAs.[45] Using this selection methodology PylRS mutants have been identified for the introduction of over 100 ncAAs with various applications including the introduction of biorthogonal functional groups for bioconjugation and fluorescence labeling, heavy atoms for X-ray crystallography, fluorescent residues for protein folding-unfolding dynamics and Förster resonance energy transfer (FRET), spin labels for electron paramagnetic resonance (EPR), posttranslational modifications, etc.[7,8] (and references therein). Although a powerful tool for chemical biology, a major limitation of ncAA incorporation is inefficient suppression of the stop codon which results in elongation termination. To effectively reassign amber codons, tRNA[Pyl] must be able to outcompete release factor 1 (RF-1), which, in prokaryotes, is responsible for terminating protein translation in response to amber and ochre nonsense codons.[46] Various strategies for increasing ncAA incorporation have been reported for both prokaryotic and eukaryotic systems.[47,48,49] To this end, tRNA[Pyl] itself has also become a target of modification.

One strategy that has been explored for increasing translation efficiency is optimizing the interaction between archaeal tRNA[Pyl] and bacterial elongation factor Tu (EF-Tu). In growing bacterial cells, the GTP-dependent EF-Tu is one of the most abundant proteins, outnumbering ribosomes approximately 10:1.[50] EF-Tu is tasked with the critical role of binding aminoacyl-tRNA and directing it to the A-site of the ribosome in a codon-dependent fashion during protein translation.[51] Despite its peculiar structure, hydrolysis protection experiments have revealed that tRNA[Pyl] interacts with EF-Tu in the same manner as canonical tRNAs.[23] The binding affinity of an aminoacyl-

tRNA to EF-Tu can have tremendous impact on translation efficiency as binding that is too weak prevents efficient complex formation and binding that is too tight can prevent dissociation of the tRNA/EF-Tu complex during translation.[52] *E. coli* aminoacyl-tRNAs have evolved relatively uniform affinities for EF-Tu, with contributions to binding made from structural elements of the tRNA as well as the affixed amino acid.[53,54] Archaeal tRNAs however, have not evolved for interaction with bacterial EF-Tu, leading to speculations that inefficient interaction between these two molcules from different domains of life may decrease the efficiency of ncAA incorporation.[55,56] Indeed, recent evidence from *in vitro* translation experiments shows that bacterial EF-Tu binds aminoacyl-tRNA[Pyl] with an affinity 25 times lower than bacterial aminoacyl-tRNA[Phe] leading to reduced translation efficiency.[57] Thus, the interaction between EF-Tu and aminoacyl-tRNA[Pyl] is a prime target for optimizing ncAA incorporation using the Pyl translation machinery.

Realizing this potential, Schultz et al. set out to optimize the interaction between EF-Tu and the tyrosyl-tRNA (*Mj*-tRNA[Tyr]) from *Methanocaldococcus jannaschii*.[55] The archaeal *Mj*-TyrRS/*Mj*-tRNA[Tyr]$_{CUA}$ pair is orthogonal in *E. coli* which has led to its widespread use for ncAA incorporation. The crystal structure of EF-Tu from *Thermus aquaticus* in complex with *E. coli* tRNA[Cys] revealed several positions in the acceptor and T stems of the tRNA that are predicted to interact with EF-Tu. From these data, a series of tRNA libraries were constructed and passed through alternating rounds of positive and negative selection based on the ability to suppress an amber codon. Remarkably, improvements in protein yield ranging from 175–2520% were reported using mutant *Mj*-tRNA[Tyr]s demonstrating the important role that the EF-Tu/tRNA interaction can play in ncAA incorporation efficiency. Söll and coworkers, sought to improve ncAA incorporation efficiency in a similar fashion with tRNA[Pyl].[56] Iterative screening of three tRNA[Pyl] libraries with mutations in the acceptor and T stems revealed a mutant tRNA[Pyl],dubbed tRNA[Pyl-Opt], which facilitated nearly a 3-fold increase in protein expression when superfolder green fluorescent protein (sfGFP) with an amber codon at position 149 was used as a reporter. Increases in expression level were reported with various ncAAs with the level of improvement showing some dependency on the identity of the ncAA. The higher efficiency of tRNA[Pyl-Opt] allowed for the production of histone H3 containing acetyllsyine, a naturally occurring posttranslational modification important for gene regulation, with a yield nearly six times greater than that obtained with wildtype tRNA[Pyl].

Codon reassignment

The genetic code is composed of four bases assembled into 64 3-base codons. Although there exists some variation of the readout, the majority of known organisms use 61 of these codons to encode 20 amino acids and the remaining three signal translation termination. Therefore, one can see that there is considerable redundancy within the code with an amino acid often being encoded by multiple, synonymous codons. Early efforts to incorporate ncAAs using the Pyl incorporation machinery took advantage of the fact that tRNA[Pyl] is a naturally occurring amber suppressor tRNA, suppressing TAG codons and leaving two redundant stop codons (TGA and TAA).

However, the revelation that PylRS tolerates mutations that occur in the anticodon of tRNA[Pyl] implies that this system could theoretically be used to incorporate an ncAA at any of the 64 codons. The desire to incorporate multiple different ncAAs into the same protein has led to the search for other codons that can be reassigned.

In 2010, Liu and colleagues showed that the permissive anticodon position of tRNA[Pyl] can be reassigned allowing for the incorporation of two different ncAAs into the same protein.[58] By mutating the anticodon of tRNA[Pyl] to one of the three stop codons opal (TGA), ochre (TAA), and amber codons were all efficiently suppressed in sfGFP expressed in *E. coli*. Opal codons displayed the highest level of suppression efficiency, however their use for ncAA incorporation should be avoided due to the possibility of wobble pairing between the opal codon and canonical tRNA anticodons. Indeed, it has been shown that significant levels of tryptophan are observed to be incorporated at opal codons in *E. coli*.[59] To incorporate two different ncAAs, cells are transformed with plasmids containing the mutant PylRS/tRNA[Pyl]$_{UAA}$ as well as the *Mj*-TyrRS/tRNA[Tyr]$_{CUA}$ pair. The former aaRS/tRNA introduces ncAAs at ochre codons with the later introducing ncAAs at amber codons (Fig. 5A, B). This methodology for expressing proteins with two different ncAAs has been used for the introduction of reactive handles for copper-catalyzed protein labeling as well as for catalyst free dual labeling of proteins.[58,60] The later method was used for the installation of a FRET pair to study protein ligand binding in *E. coli* glutamine binding protein (Fig. 5C).

While stop codon reassignment has proven to be effective for the introduction of two different ncAAs, it is limited to only three codons. More recent studies have looked at the possibility of expanding the number of codons available for ncAA incorporation further through sense codon reassignment. Given that most amino acids are redundantly encoded by multiple codons, reassignment of synonymous codons has the potential to greatly expand the number of available codons for ncAA incorporation. However, sense codon reassignment presents a unique set of challenges as, among other things, redefining a codon in the coding sequence of a gene may have detrimental effects on the host organisms—especially if that codon falls within an essential gene.[61,62] Further, just as stop codon suppression involves competition with elongation termination, incorporating ncAAs via sense codon reassignment requires that the orthogonal aaRS/tRNA pair be able to efficiently outcompete the endogenous translation system.

It is well known that usage of synonymous codons within an organism is biased with some codons occurring more frequently than others.[63] To minimize the adverse effects of sense codon reassignment using the PylRS/tRNA[Pyl] pair, early efforts focused on reassigning rarely used codons, in particular, those encoding arginine. For example, Söll et al. set out to reassign the rare CGG codon using the PylRS/tRNA[Pyl] pair in *Mycoplasma capricolum*.[64] *M. capricolum* was chosen as the host because of the scarcity of CGG codons in its genome (only 6 CGG codons were identified) coupled with the fact that it lacks the requisite tRNA$_{CCG}$ for decoding CGG. Despite the ability of PylRS to charge mutant tRNA[Pyl]$_{CCG}$ *in vitro*, LC-MS/MS of *β*-galactosidase expressed in *M. capricolum* harboring plasmids encoding *Mm*-PylRS/tRNA[Pyl]$_{CCG}$ revealed
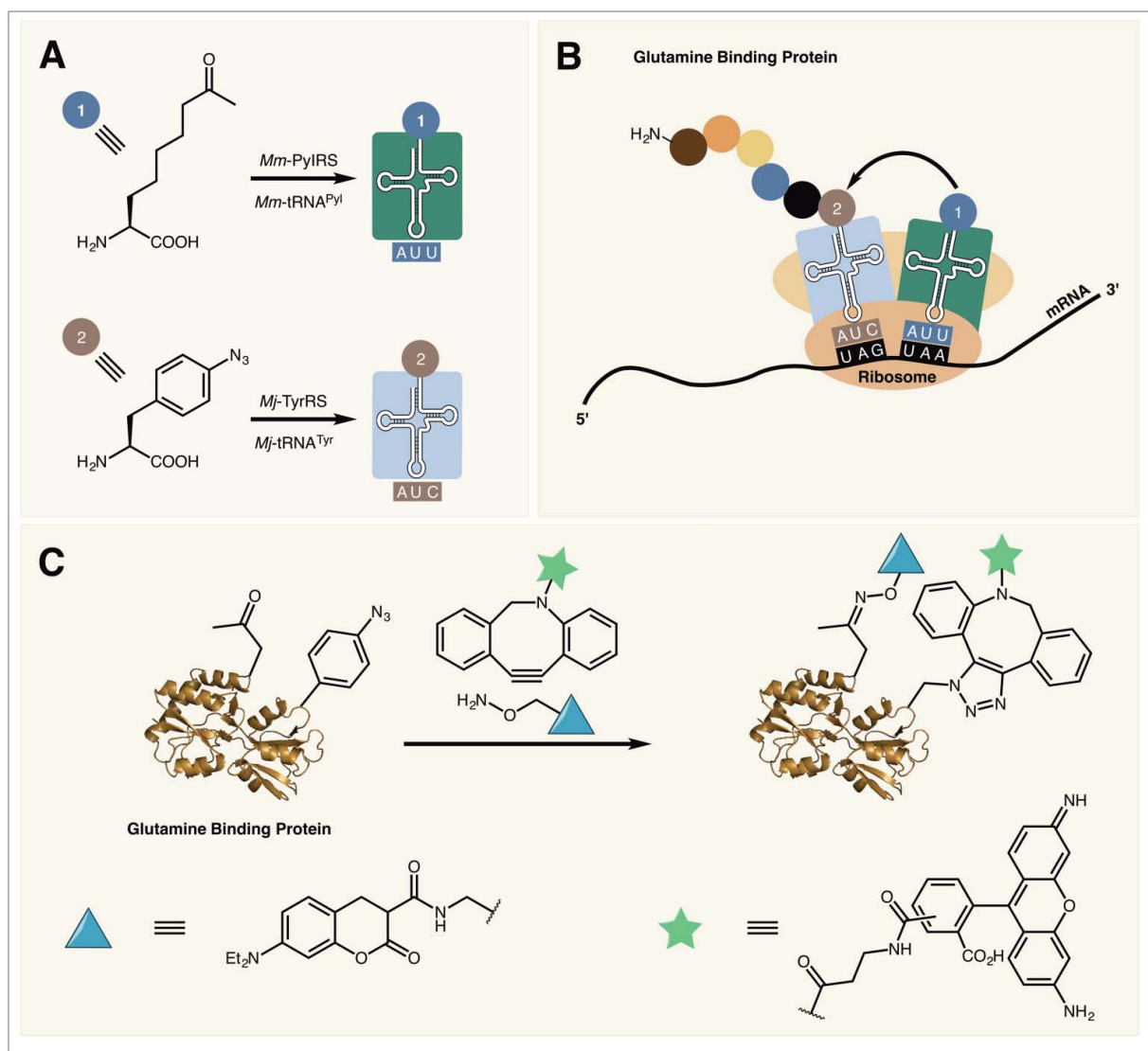
Figure 5. (A) *Mm*-PylRS and *Mj*-TyrRS specifically charge their respective tRNAs with an ncAA. (B) *Mm*-tRNA$^{Pyl}_{UUA}$ incorporates an ncAA at ochre (UAA) codons while *Mj*-tRNA$^{Tyr}_{CUA}$ incorporates ncAAs in response to amber (UAG) codons. (C) Dual labeling of glutamine binding protein containing two reactive ncAAs.

incorporation of only Arg at CGG codons. It was suspected that Arg incorporation was a result of misacylation of tRNA$^{Pyl}_{CCG}$ by endogenous ArgRS, which utilizes the tRNA anticodon as a major identity element. These suspicions were confirmed by *in vitro* assays which showed that tRNA$^{Pyl}_{CCG}$ was efficiently charged with Arg by recombinant *E. coli* and *M. capricolum* ArgRSs. This work highlighted another difficulty with sense codon reassignment. Since most aaRSs utilize the anticodon as a major identity element, mutating the anticodon of tRNA^{Pyl} to a sense codon may decrease its orthogonality and render it a substrate for endogenous aaRSs.

More successful results were reported for reassignment of the rare AGG codon in *E. coli*. Liu and coworkers reported 24% incorporation of an ncAA at AGG codons in GFP$_{UV}$ expressed in cells harboring plasmid-borne copies of *Mm*-PylRS and mutant tRNA$^{Pyl}_{CCU}$.[65] Incorporation efficiency could be further improved to approximately 90% using a minimal media devoid of Arg coupled with inducing the expression of PylRS/tRNA-$^{Pyl}_{CCU}$ ahead of GFP$_{UV}$ induction. The later strategy allowed for the accumulation of aminoacyl-tRNA$^{Pyl}_{CCU}$ thereby

improving its ability to compete with Arg-tRNA$^{Arg}_{CCU}$ for decoding AGG. Using wildtype *Mm*-PylRS/tRNA$^{Pyl}_{CCU}$ three different ncAAs were incorporated at AGG codons with efficiencies of 72–92%. Although impressive, these results required the use of high concentrations (10 mM) of ncAA and still resulted in a heterogeneous product.

Sakamoto et al. minimized competition from endogenous tRNA by generating a tRNA$^{Arg}_{CCU}$ (*argW*) knockout strain.[66] They went on to replace several AGG codons in essential genes in the *E. coli* genome. The codons that were not replaced were left to be decoded by a plasmid-borne copy of a synonymous Arg tRNA (tRNA$^{Arg}_{UCU}$), the UCU anticodon of which forms a wobble pair to decode AGG albeit at a lower efficiency than CCU. The greater efficiency of CCU at decoding AGG as compared with UCU allowed for tRNA$^{Pyl}_{CCU}$ to outcompete tRNA$^{Arg}_{UCU}$ during protein translation. When a peptide-SUMO fusion protein containing an AGG codon was expressed in the *argW*-knockout strain, in the presence of mutant tRNA-$^{Pyl}_{CCU}$ and mutant *Mm*-PylRS, the protein containing an ncAA at the AGG position was obtained with nearly quantitative

occupancy of the ncAA. Minimal amounts of protein containing Arg at the AGG position were again attributed to misacylation of tRNA$^{Pyl}_{CCU}$ by endogenous ArgRS as was reported previously for tRNA$^{Pyl}_{CCG}$.[64] At nearly the same time that this work was published, Yoo et al. reported the incorporation of a variety of ncAAs at AGG codons with undetectable Arg incorporation using a mutant Mj-TyrRS/tRNA$^{Tyr}_{CCU}$ pair in an $E.$ $coli$ strain lacking $argW$ as well as an enzyme required for Arg biosynthesis.[67] Through limiting the available Arg in the cell, ncAAs could be incorporated at up to three AGG codons with insignificant background incorporation.

In an impressive study, Chin and colleagues demonstrated that the PylRS/tRNA^{Pyl} pair can be used to reassign frequent codons with low efficiency in $E.$ $coli$, human cells, and $D.$ $melanogaster$.[15] By mutating the anticodon of tRNA^{Pyl} to one of several frequently occurring codons, ncAAs bearing a reactive functional group were incorporated into proteins with efficiency of 0.02–0.65% per codon. While the incorporation efficiency was low, this limited the detrimental effects of ncAA incorporation on the organism and was useful for fluorescence labeling of the entire proteome. Further, by restricting the expression of PylRS and tRNA^{Pyl} to certain tissues of $D.$ $melanogaster$, they were able to label proteins containing the ncAA in a tissue-dependent manner.

More recently, Söll et al. have explored the possibility of reassigning a frequently occurring serine codon using the PylRS/tRNA^{Pyl} pair.[68] Despite being significantly more abundant than AGG in the $E.$ $coli$ genome, reassignment of a Ser codon was chosen based on the fact that the seryl-tRNA synthetase of $E.$ $coli$ does not use the tRNA anticodon as a recognition element. This eliminates the possibility that the anticodon will cause misacylation by the endogenous aaRS as was observed when reassigning CGG and AGG Arg codons. Further, the Ser AGU codon is naturally decoded by tRNA$^{Ser}_{GCU}$ which involves a wobble pair at the third position. Since mutant tRNA$^{Pyl}_{ACU}$ could decode AGU via Watson-Crick pairing it was hypothesized that this stronger interaction would allow tRNA$^{Pyl}_{ACU}$ to outcompete tRNA$^{Ser}_{GCU}$. LC-MS/MS of sfGFP that was expressed in $E.$ $coli$ cells harboring plasmid-borne copies of mutant tRNA$^{Pyl}_{ACU}$ as well as a mutant PylRS revealed successful ncAA incorporation at one of two AGU codons in sfGFP. Further analysis revealed that the ncAA was incorporated at this AGU site at approximately 65% efficiency, however the ncAA was also found to be incorporated at an AGC codon. Impressively, it was also demonstrated that a more frequent Ser codon (UCG) as well as the most abundant codon in $E.$ $coli$ (CUG), which encodes leucine, could also be recoded to an ncAA with varying success. While the overall efficiency of reassigning AGU was low, the fact that greater than 50% reassignment was achieved is noteworthy given that a relatively low ncAA concentration (1 mM) was used and that no deletions in the Ser biosynthesis pathway or of a synonymous tRNA^{Ser} were required. These observations suggest that, under optimized conditions, codon reassignment is not limited to rarely occurring codons.

Quadruplet codons

In the search for blank codons for ncAA incorporation four base codons offer a promising alternative to nonsense and sense codon reassignment with the potential to code for upwards of 250 blank codons.[7,69] Early efforts by Schultz et al. demonstrated the feasibility of decoding a four base codon in $vivo$ using the orthogonal $Pyrococcus$ $horikoshii$ LysRS in conjunction with mutant tRNA$^{Lys}_{UCCU}$ and, when coupled with the Mj-TyrRS/tRNA$^{Tyr}_{CUA}$ pair, allowed for the production of proteins containing two different ncAAs.[70] However, the system was complicated by toxic effects of expressing Ph-LysRS in $E.$ $coli$, and suffered low yield from inefficient decoding of quadruplet codons on the ribosome. Chin et al. overcame this later challenge by evolving ribosomes with enhanced ability for decoding four base codons.[71] Using these evolved ribosomes in conjunction with the Mj-TyrRS/tRNA$^{Tyr}_{UCCU}$ pair, they obtained quadruplet codon suppression with improved yield compared to the Ph-LysRS/tRNA$^{Lys}_{UCCU}$ pair. However, Mj-TyrRS uses the anticodon of tRNA^{Tyr} as an identity element[72] which limits the efficiency with which this pair can be used to decode quadruplet codons. More recently, the PylRS/tRNA^{Pyl} pair has become a promising option for quadruplet codon decoding given its greater tolerance for anticodon mutations.

In 2013, Guo et al. reported the use of the PylRS/tRNA^{Pyl} pair for incorporating ncAAs at quadruplet (AGGA) codons.[73] To identify tRNA^{Pyl} mutants with more efficient 4 base decoding on natural ribosomes, a series of tRNA$^{Pyl}_{UCCU}$ libraries were constructed with randomized nucleotides in the anticodon stem and loop (Fig. 6A). The libraries were passed through positive selections based on their ability to suppress an AGGA codon in the chloramphenicol acetyltransferase (cat-AGGA) gene in the presence of an ncAA (Fig. 6B). If the mutant tRNA-$^{Pyl}_{UCCU}$ is able to efficiently suppress AGGA in cat-AGGA, the cells survive in the presence of chloramphenicol. Expressing GFP$_{UV}$ containing an AGGA mutation in the presence of mutant tRNA^{Pyl}s identified from selection, resulted in an increase in fluorescence signal of ~2–4-fold as compared with wildtype tRNA$^{Pyl}_{UCCU}$. Using the best performing mutant, they were able to efficiently incorporate an ncAA at an AGGA codon in enhanced GFP (EGFP) expressed in 293T cells. Molecular simulations of the selected tRNA^{Pyl}s revealed structural changes in the anticodon that likely facilitated decoding of the quadruplet codon on the ribosome.[73]

While Guo et al. reported quantitative ncAA incorporation in response to AGGA with mutant and wildtype tRNA$^{Pyl}_{UCCU}$, others[65,74] have reported significant Arg incorporation at AGGA sites with wildtype tRNA^{Pyl}. Interestingly, Liu et al. observed predominantly Arg incorporation with a sfGFP-134AGGA reporter, but predominantly ncAA incorporation with a GFP$_{UV}$-149AGGA reporter.[65] Recently, Arg incorporation at AGGA was also reported with the Mj-TyrRS/tRNA-$^{Tyr}_{UCCU}$ pair.[75] In this study, background Arg was attributed to misaminoacylation of tRNA$^{Tyr}_{UCCU}$ by endogenous Ec-ArgRS as Ec-ArgRS uses the AGG anticodon as a major identity element. Introducing an A38C mutation in tRNA$^{Tyr}_{UCCU}$ removes a minor identity element of Ec-ArgRS and this was shown to eliminate background Arg incorporation. This observation suggests that background suppression indeed arises from misaminoacylation of the orthogonal tRNA. However, in $vitro$ aminoacylation assays have shown that tRNA$^{Pyl}_{UCCU}$ is not a substrate for Ec-ArgRS,[74] suggesting that the observed suppression in the case of tRNA^{Pyl} is due to competition with endogenous tRNA$^{Arg}_{CCU}$. However, as we will discuss next, more
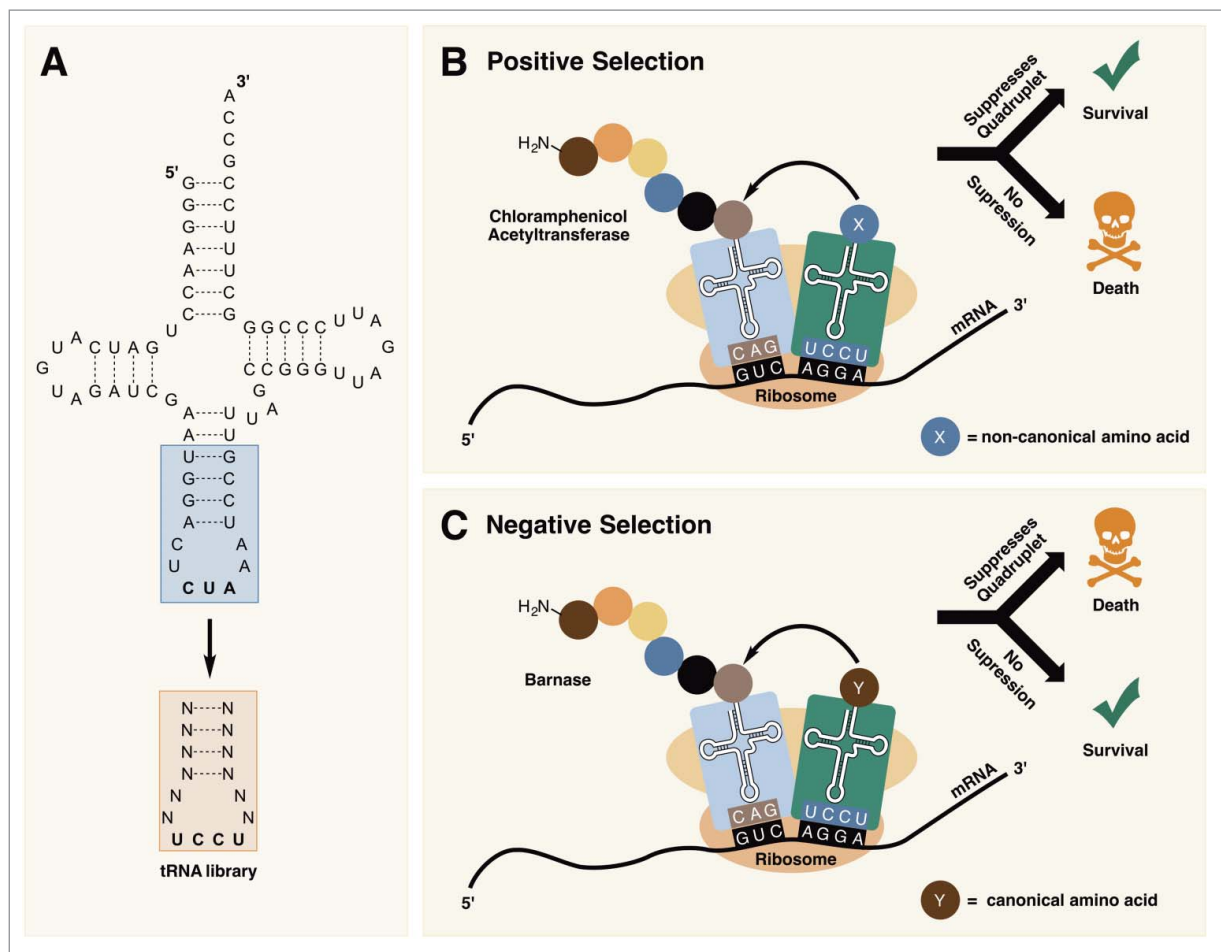
Figure 6. (A) Randomization of the anticodon stem and loop allows for the optimization of tRNA^{Pyl} for decoding quadruplet codons. (B) In positive selection, tRNA^{Pyl} mutants that are able to suppress quadruplet codons are selected for. (C) Negative selection in the absence of an ncAA removes tRNA^{Pyl} mutants that are aminoacylated with canonical amino acids by endogenous aminoacyl-tRNA synthetases.

recent data suggests that tRNA$^{Pyl}_{UCCU}$ is indeed a substrate for *Ec*-ArgRS.

Chin et al. demonstrated that by including a round of negative selection, tRNA$^{Pyl}_{UCCU}$s that are aminoacylated by endogenous aaRS can be removed from the tRNA library which results in little observable background suppression.[76] In this study, they generated a tRNA$^{Pyl}_{UCCU}$ library and performed selection for AGGA suppression capability on an evolved orthogonal ribosome that had previously been shown to have enhanced ability for decoding quadruplet codons.[71] However, before performing a positive selection, the tRNA library was screened in the absence of an ncAA in cells harboring a copy of the toxic barnase gene containing AGGA codons. Therefore, library members that are aminoacylated by endogenous aaRSs give rise to expression of the toxic gene and cell death (Fig. 6C). This was followed by positive selection using *cat*-AGGA as reported by Guo et al. A resulting mutant tRNA$^{Pyl}_{UCCU}$ was able to suppress AGGA codons in *cat*-AGGA to confer resistance at >500 μg·mL^{-1} chloramphenicol in the presence of an ncAA, but was unable to grow on 25 μg·mL^{-1} in the absence of an ncAA, indicating very little background incorporation. By comparison, the wildtype tRNA$^{Pyl}_{UCCU}$ resulted in much higher background conferring resistance of up to 125 μg·mL^{-1} chloramphenicol in the absence of an

ncAA.[76] The observation that mutant tRNA$^{Pyl}_{UCCU}$ gives lower background than the wildtype gives credence to the hypothesis that background quadruplet suppression is dominated by misaminoacylation of the tRNA by endogenous aaRS. If competition from endogenous tRNA$^{Arg}_{CCU}$ were the prevailing reason for background Arg incorporation, one would expect similar levels of background in both cases. Although the evolved tRNA$^{Pyl}_{UCCU}$ retained the A38 recognition element for *Ec*-ArgRS it had a greatly expanded anticodon loop compared with wildtype tRNA$^{Pyl}_{UCCU}$ (12 bases compared with 8 bases in the wildtype). This expanded anticodon loop may prevent misaminoacylation by *Ec*-ArgRS.

The ability of three other tRNA^{Pyl}s, whose anticodons (TAGA, AGTA, and CTAG) are not recognition elements for endogenous aaRSs, to decode quadruplet codons was also investigated by Chin et al.[76] Wildtype tRNA^{Pyl}s containing these anticodons displayed virtually no background while the evolved versions facilitated more efficient ncAA incorporation. Guo et al. recently reported the ability of evolved tRNA^{Pyl}s to efficiently deliver ncAA in response to TAGN (where N = A, U, G) codons.[77] Using the TAGN codon avoids the concern of competition with endogenous tRNAs. Further, after randomizing positions in the anticodon stem and loop they performed selection in an *E. coli* strain lacking RF-1 and in which all

instances of TAG were replaced with TAA.[78] This eliminated competition from RF-1 during decoding. The evolved tRNA^{Pyl}s were able to incorporate ncAAs at TAGN codons with high efficiency and with no detectable background into GFP$_{UV}$. Interestingly, the identity of the fourth base was not critical for decoding with all three anticodons capable of decoding TAGA, TAGU, and TAGG albeit at a lower efficiency than the Watson-Crick pair.[77]

Conclusion

Over the past one-and-a-half decades much has been learned about the structure and function of tRNA^{Pyl} and its cognate aaRS. Despite enormous progress, there are still questions that remain to be answered. Most importantly, ascertaining how the ability to incorporate Pyl has emerged in Nature will provide more insight into the origins of the genetic code. As more organisms that are capable of incorporating Pyl are identified, a clearer picture of the evolutionary history of Pyl incorporation can be developed. The PylRS/tRNA^{Pyl} pair has emerged as an outstanding tool for genetic code expansion in bacteria, eukaryotic cells, and whole organisms. Recent progress in codon reassignment and quadruplet codon decoding has taken this pair beyond stop codon reassignment and greatly expanded the number of available codons for ncAA incorporation. As these methods are refined, it is likely that the number of unique ncAAs able to be incorporated into a single protein will increase. The PylRS/tRNA^{Pyl} pair will, no doubt, be a valuable tool for the progress of synthetic biology.

Funding

ORCID

Andreas Ehnbom http://orcid.org/0000-0002-7044-1712
Wenshe R. Liu http://orcid.org/0000-0002-7078-6534

References

1. Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. Science. 2002;296:1459–62. doi:10.1126/science.1069588.

2. Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science. 2002;296:1462–6. doi:10.1126/science.1069556.

3. Gaston MA, Zhang L, Green-Church KB, Krzycki JA. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature. 2011;471:64750. doi:10.1038/nature09918.

4. Polycarpo C, Ambrogelly A, Bérubé A, Winbush SM, McCloskey JA, Crain PF, Wood JL, Söll D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proc Natl Acad Sci U S A. 2004;101:12450–4. doi:10.1073/pnas.0405362101.

5. Borrel G, Gaci N, Peyret P, O'Toole PW, Gribaldo S, Brugère JF. Unique characteristics of the pyrrolysine system in the 7th order of methanogens: Implications for the evolution of a genetic code expansion cassette. Archaea Int Microbiol J. 2014. doi:10.1155/2014/374146.

6. Fournier GP, Andam CP, Gogarten JP. Ancient horizontal gene transfer and the last common ancestors. BMC Evol Biol. 2015;15:70. doi:10.1186/s12862-015-0350-0.

7. Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 2015;6:50–69. doi:10.1039/C4SC01534G.

8. Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta. 2014;1844:1059–70. doi:10.1016/j.bbapap.2014.03.002.

9. Crnković A, Suzuki T, Söll D, Reynolds NM. Pyrrolysyl-tRNA Synthetase, an Aminoacyl-tRNA Synthetase for Genetic Code Expansion. Croat Chem Acta. 2016;89:163–74. doi:10.5562/cca2825.

10. Hancock SM, Uprety R, Deiters A, Chin JW. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA Synthetase/tRNA Pair. J Am Chem Soc. 2010;132:14819–24. doi:10.1021/ja104609m.

11. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun. 2008;371:818–22. doi:10.1016/j.bbrc.2008.04.164.

12. Chen PR, Groff D, Guo J, Ou WJ, Cellitti S, Geierstanger BH, Schultz PG. A facile system for encoding unnatural amino acids in mammalian cells. Angew Chem Int Ed. 2009;48:4052–5. doi:10.1002/anie.200900683.

13. Greiss S, Chin JW. Expanding the genetic code of an animal. J Am Chem Soc. 2011;133:14196–9. doi:10.1021/ja2054034.

14. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of *Drosophila melanogaster*. Nat Chem Biol. 2012;8:748–50. doi:10.1038/nchembio.1043.

15. Elliott TS, Townsley FM, Bianco A, Ernst RJ, Sachdeva A, Elsässer SJ, Davis L, Lang K, Pisa R, Greiss S, et al. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. Nat Biotechnol. 2014;32:465–72. doi:10.1038/nbt.2860.

16. Ernst RJ, Krogager TP, Maywood ES, Zanchi R, Beránek V, Elliott TS, Barry NP, Hastings MH, Chin JW. Genetic code expansion in the mouse brain. Nat Chem Biol. 2016;12:776–778. doi:10.1038/nchembio.2160.

17. Han S, Yang A, Lee S, Lee HW, Park CB, Park HS. Expanding the genetic code of *Mus musculus*. Nat Commun. 2017;8:14568. doi:10.1038/ncomms14568.

18. Gaston MA, Jiang R, Krzycki JA. Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr Opin Microbiol. 2011;14:342–9. doi:10.1016/j.mib.2011.04.001.

19. Krzycki JA. The direct genetic encoding of pyrrolysine. Curr Opin Microbiol. 2005;8:706–12. doi:10.1016/j.mib.2005.10.009.

20. Soma A, Himeno H. Cross-species aminoacylation of tRNA with a long variable arm between *Escherichia coli* and *Saccharomyces cerevisiae*. Nucleic Acids Res. 1998;26:4374–81. doi:10.1093/nar/26.19.4374.

21. Ueda Y, Kumagai I, Miura K. The effects of a unique D-Loop structure of a minor tRNA(UUALeu) from streptomyces on its structural stability and amino-acid accepting activity. Nucleic Acids Res. 1992;20:3911–7. doi:10.1093/nar/20.15.3911.

22. 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;62:1031–3. doi:10.1107/S1744309106036700.

23. Théobald-Dietrich A, Frugier M, Giegé R, Rudinger-Thirion J. Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA. Nucleic Acids Res. 2004;32:1091–6. doi:10.1093/nar/gkh266.

24. Motorin Y, Helm M. tRNA stabilization by modified nucleotides. Biochemistry. 2010;49:4934–44. doi:10.1021/bi100408z.

25. Herring S, Ambrogelly A, Polycarpo CR, Söll D. Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. Nucleic Acids Res. 2007;35:1270–8. doi:10.1093/nar/gkl1151.

26. Yokogawa T, Watanabe Y, Kumazawa Y, Ueda T, Hirao I, Miura K, Watanabe K. A novel cloverleaf structure found in mammalian

mitochondrial tRNA(Ser) (UCN). Nucleic Acids Res. 1991;19:6101–5. doi:10.1093/nar/19.22.6101.

27. Hayashi I, Kawai G, Watanabe K. Higher-order structure and thermal instability of bovine mitochondrial tRNA(Ser)UGA investigated by proton NMR spectroscopy. J Mol Biol. 1998;284:57–69. doi:10.1006/jmbi.1998.2151.

28. Watanabe Y, Kawai G, Yokogawa T, Hayashi N, Kumazawa Y, Ueda T, Nishikawa K, Hirao I, Miura K, Watanabe K. Higher-order structure of bovine mitochondrial tRNA(Ser)UGA - chemical modification and computer modeling. Nucleic Acids Res. 1994;22:5378–84. doi:10.1093/nar/22.24.5378.

29. Borrel G, O'Toole PW, Harris HMB, Peyret P, Brugère JF, Gribaldo S. Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. Genome Biol Evol. 2013;5:1769–80. doi:10.1093/gbe/evt128.

30. Mihajlovski A, Alric M, Brugère JF. A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the mcrA gene. Res Microbiol. 2008;159:516–21. doi:10.1016/j.resmic.2008.06.007.

31. Lang K, Schuldes J, Klingl A, Poehlein A, Daniel R, Brune A. New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of Candidatus methanoplasma termitum. Appl Environ Microbiol. 2015;81:1338–52. doi:10.1128/AEM.03389-14.

32. Li Y, Leahy SC, Jeyanathan J, Henderson G, Cox F, Altermann E, Kelly WJ, Lambie SC, Janssen PH, Rakonjac J, et al. The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the Methanomassiliicoccales. Stand Genomic Sci. 2016;11:59. doi:10.1186/s40793-016-0183-5.

33. Matassi G. Horizontal gene transfer drives the evolution of Rh50 permeases in prokaryotes. BMC Evol Biol. 2017;17:2. doi:10.1186/s12862-016-0850-6.

34. Fournier GP, Huang J, Gogarten JP. Horizontal gene transfer from extinct and extant lineages: Biological innovation and the coral of life. Philos Trans R Soc B-Biol Sci. 2009;364:2229–39. doi:10.1098/rstb.2009.0033.

35. Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, Umehara T, Söll D, Nureki O. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. Nature. 2009;457:1163–7. doi:10.1038/nature07611.

36. Sterner T, Jansen M, Hou YM. Structural and functional accommodation of nucleotide variations at a conserved transfer-RNA tertiary base pari. RNA pub RNA. 1995;1:841–51.

37. Kavran JM, Gundliapalli S, O'Donoghue P, Englert M, Söll D, Steitz TA. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci U S A. 2007;104:11268–73. doi:10.1073/pnas.0704769104.

38. 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;287:32738–46. doi:10.1074/jbc.M112.396754.

39. Herring S, Ambrogelly A, Gundllapalli S, O'Donoghue P, Polycarpo CR, Söll D. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. Febs Lett. 2007;581:3197–203. doi:10.1016/j.febslet.2007.06.004.

40. Lee MM, Jiang R, Jain R, Larue RC, Krzycki J, Chan MK. Structure of Desulfitobacterium hafniense PylSc, a pyrrolysyl-tRNA synthetase. Biochem Biophys Res Commun. 2008;374:470–4. doi:10.1016/j.bbrc.2008.07.074.

41. Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. J Mol Biol. 2008;378:634–52. doi:10.1016/j.jmb.2008.02.045.

42. Englert M, Moses S, Hohn M, Ling J, O'Donoghue P, Söll D. Aminoacylation of tRNA 2′- or 3′ –hydroxy by phosphoseryl- and pyrrolysyl-tRNA synthetases. Febs Lett. 2013;587:3360–4. doi:10.1016/j.febslet.2013.08.037.

43. Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C, Söll D. Pyrrolysine is not hardwired for cotranslational insertion at UAG

codons. Proc Natl Acad Sci U S A. 2007;104:3141–6. doi:10.1073/pnas.0611634104.

44. Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. 1998;26:5017–35. doi:10.1093/nar/26.22.5017.

45. Wang L, Xie J, Schultz PG. Expanding the genetic code. Annu Rev Biophys Biomol Struct. 2006;35:225–49. doi:10.1146/annurev.biophys.35.101105.121507.

46. Scolnick E, Tompkins R, Caskey T, Nirenber M. Release factors differing in specificity for terminator codons. Proc Natl Acad Sci U S A. 1968;61:768–74. doi:10.1073/pnas.61.2.768

47. Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol. 2011;7:779–786. doi:10.1038/nchembio.657.

48. Huang Y, Russell WK, Wan W, Pai PJ, Russell DH, Liu WR. A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. Mol Biosyst. 2010;6:683–6. doi:10.1039/b920120c.

49. Schmied WH, Elsässer SJ, Uttamapinant C, Chin JW. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA Synthetase/tRNA expression and engineered eRF1. J Am Chem Soc. 2014;136:15577–83. doi:10.1021/ja5069728.

50. Furano AV. Content of elongation-factor Tu in Escherichia coli. Proc Natl Acad Sci U S A. 1975;72:4780–4. doi:10.1073/pnas.72.12.4780.

51. Andersen GR, Nissen P, Nyborg J. Elongation factors in protein biosynthesis. Trends Biochem Sci. 2003;28:434–41. doi:10.1016/S0968-0004(03)00162-2.

52. Schrader JM, Chapman SJ, Uhlenbeck OC. Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. Proc Natl Acad Sci U S A. 2011;108:5215–20. doi:10.1073/pnas.1102128108.

53. Louie A, Ribeiro NS, Reid BR, Jurnak F. Relative affinities of all Escherichia coli aminoacyl-transfer RNAs for elongation-factor Tu-GTP. J Biol Chem. 1984;259:5010–16.

54. LaRiviere FJ, Wolfson AD, Uhlenbeck OC. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. Science. 2001;294:165–8. doi:10.1126/science.1064242.

55. Guo J, Melançon CE, Lee HS, Groff D, Schultz PG. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. Angew Chem Int Ed. 2009;48:9148–51. doi:10.1002/anie.200904035.

56. Fan C, Xiong H, Reynolds NM, Söll D. Rationally evolving tRNA(Pyl) for efficient incorporation of noncanonical amino acids. Nucleic Acids Res. 2015;43:e156. doi:10.1093/nar/gkv800.

57. Wang J, Kwiatkowski M, Forster AC. Kinetics of tRNA(Pyl)-mediated amber suppression in Escherichia coli translation reveals unexpected limiting steps and competing reactions. Biotechnol Bioeng. 2016;113:1552–9. doi:10.1002/bit.25917.

58. Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, Russell DH, Liu WR. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. Angew Chem Int Ed. 2010;49:3211–4. doi:10.1002/anie.201000465.

59. Odoi KA, Huang Y, Rezenom YH, Liu WR. Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei Pyrrolysyl-tRNA Synthetase-tRNA(Pyl) pairs in the Escherichia coli BL21(DE3) cell strain. Plos One. 2013;8:e57035. doi:10.1371/journal.pone.0057035.

60. Wu B, Wang Z, Huang Y, Liu WR. Catalyst-free and site-specific one-pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids. ChemBioChem. 2012;13:1405–8. doi:10.1002/cbic.201200281.

61. Krishnakumar R, Ling J. Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion. Febs Lett. 2014;588:383–8. doi:10.1016/j.febslet.2013.11.039.

62. Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, Church GM. Probing the limits of genetic recoding in essential genes. Science. 2013;342:361–3. doi:10.1126/science.1241460.

63. Ling JQ, O'Donoghue P, Söll D. Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. Nat Rev Microbiol. 2015;13:707–21. doi:10.1038/nrmicro3568.

64. Krishnakumar R, Prat L, Aerni HR, Ling J, Merryman C, Glass JI, Rinehart J, Söll D. Transfer RNA misidentification scrambles sense codon recoding. Chembiochem. 2013;14:1967–72. doi:10.1002/cbic.201300444.

65. Zeng Y, Wang W, Liu WR. Towards reassigning the rare AGG codon in *Escherichia coli*. Chembiochem. 2014;15:1750–4. doi:10.1002/cbic.201400075.

66. Mukai T, Yamaguchi A, Ohtake K, Takahashi M, Hayashi A, Iraha F, Kira S, Yanagisawa T, Yokoyama S, Hoshi H, et al. Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. Nucleic Acids Res. 2015;43:8111–22. doi:10.1093/nar/gkv787.

67. Lee BS, Shin S, Jeon JY, Jang KS, Lee BY, Choi S, Yoo TH. Incorporation of unnatural amino acids in response to the AGG codon. ACS Chem Biol. 2015;10:1648–53. doi:10.1021/acschembio.5b00230.

68. Ho JM, Reynolds NM, Rivera K, Connolly M, Guo LT, Ling J, Pappin DJ, Church GM, Söll D. Efficient reassignment of a frequent serine codon in wild-type *Escherichia coli*. Acs Synth Biol. 2016;5:163–71. doi:10.1021/acssynbio.5b00197.

69. Wang K, Schmied WH, Chin JW. Reprogramming the genetic code: From triplet to quadruplet codes. Angew Chem Int Ed. 2012;51:2288–97. doi:10.1002/anie.201105016.

70. Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci U S A. 2004;101:7566–71. doi:10.1073/pnas.0401517101.

71. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature. 2010;464:441–4. doi:10.1038/nature08817.

72. Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat Struct Biol. 2003;10:425–32. doi:10.1038/nsb934.

73. Niu W, Schultz PG, Guo J. An expanded genetic code in mammalian cells with a functional quadruplet codon. ACS Chem Biol. 2013;8:1640–5. doi:10.1021/cb4001662.

74. O'Donoghue P, Prat L, Heinemann IU, Ling J, Odoi K, Liu WR, Söll D. Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA(Pyl) for genetic code expansion. Febs Lett. 2012;586:3931–7. doi:10.1016/j.febslet.2012.09.033.

75. Lee BS, Kim S, Ko BJ, Yoo TH. An efficient system for incorporation of unnatural amino acids in response to the four-base codon AGGA in *Escherichia coli*. Biochim Biophys Acta. 2017. doi:10.1016/j.bbagen.2017.02.017.

76. Wang K, Sachdeva A, Cox DJ, Wilf NM, Lang K, Wallace S, Mehl RA, Chin JW. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. Nat Chem. 2014;6:393–403. doi:10.1038/nchem.1919

77. Wang J, Shang X, Cerny R, Niu W, Guo J. Systematic Evolution and Study of UAGN Decoding tRNAs in a genomically recoded bacteria. Sci Rep. 2016;6:21898. doi:10.1038/srep21898.

78. Lajoie MJ, Tovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al. Genomically recoded organisms expand biological functions. Science. 2013;342:357–60. doi:10.1126/science.1241459.