Published in final edited form as:
Nat Chem. 2020 June 01; 12(6): 535–544. doi:10.1038/s41557-020-0472-x.

Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids

Daniel L. Dunkelmann[#1], **Julian C. W. Willis**[#1], **Adam T. Beattie**[1], **Jason W. Chin**[1,*]

[1]Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, England, UK

[#] These authors contributed equally to this work.

Abstract

Expanding and reprogramming the genetic code of cells for the incorporation of multiple distinct non-canonical amino acids (ncAAs), and the encoded biosynthesis of non-canonical biopolymers, requires the discovery of multiple orthogonal aminoacyltransfer RNA synthetase/tRNA pairs. These pairs must be orthogonal to both the host synthetases and tRNAs and to each other. Pyrrolysyl-tRNA synthetase (PylRS)/^{Pyl}tRNA pairs are the most widely used system for genetic code expansion. Here we reveal that the sequences of ΔNPylRS/$^{\Delta NPyl}$tRNA pairs (which lack N-terminal domains) form two distinct classes. We show that the measured specificities of the ΔNPylRSs and $^{\Delta NPyl}$tRNAs correlate with sequence-based clustering, and most ΔNPylRSs preferentially function with $^{\Delta NPyl}$tRNAs from their class. We then identify 18 mutually orthogonal pairs from the 88 ΔNPylRS/$^{\Delta NPyl}$tRNA combinations tested. Moreover, we generate a set of 12 triply orthogonal pairs, each composed of three new PylRS/^{Pyl}tRNA pairs. Finally, we diverge the ncAA specificity and decoding properties of each pair, within a triply orthogonal set, and direct the incorporation of three distinct non-canonical amino acids into a single polypeptide.

Protein translation provides the ultimate paradigm for the encoded cellular synthesis of defined polymer sequences, but natural translation is commonly limited to polymerizing the 20 canonical amino acids. Engineering cellular translation may enable the biosynthesis and evolution of non-canonical biopolymers[1,2]. However, this will require: (1) strategies for the creation of blank codons, beyond stop codons, that may be assigned to new monomers [3–5]; (2) the creation of ribosomes, and other translation factors, with expanded substrate sope[6–8]; and (3) the creation of a set of aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that are orthogonal with respect to endogenous synthetases and tRNAs in the host organism and mutually orthogonal with respect to each other[5,9–12]. These pairs must be further engineered to decode distinct blank codons and use unique monomers that are not substrates for other

[*]Correspondence: chin@mrc-lmb.cam.ac.uk.

Author contributions D.L.D., J.C.W.W, and J.W.C. designed the project. D.L.D., J.C.W.W, performed experiments. A.T.B. analysed and interpreted MS/MS data. D.L.D., J.C.W.W, and J.W.C wrote the paper with input from A.T.B..

Competing interests: The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

aaRSs. An additional mutually orthogonal pair is required for each new monomer that is uniquely encoded.

Despite exciting progress towards the encoded cellular synthesis of non-canonical biopolymers, the identification of multiple engineered mutually orthogonal aaRS/tRNA pairs that recognize distinct codons and incorporate distinct non-canonical amino acids (ncAAs) remains an outstanding challenge[5,9,13,14]. Each new ncAA, aaRS, tRNA and codon must function together and be orthogonal to each endogenous amino acid, aaRS, and group of isoacceptor tRNAs and their cognate group of codons. Therefore, for each new ncAA:aaRS:tRNA:codon set, three interactions must be established (ncAA:aaRS, aaRS:tRNA, and tRNA:codon) and 120 interactions (6 x 20 interactions; this analysis counts all isoacceptors for a natural amino acid as one and all codons for an amino acid as one and therefore provides a conservative estimate of the interactions that must be controlled) between the new set and the endogenous translational machinery must be minimized. Moreover, when incorporating more than one ncAA there is the potential for interactions between components of the additional ncAA:aaRS:tRNA:codon sets, and these must also be minimized. Generating ncAA:aaRS:tRNA:codon sets to encode three distinct ncAAs into a polypeptide requires nine specific interactions to be established and minimization of at least 378 specific interactions, including 18 interactions between components of the three sets.

Five pairs (based on *Methanococcus janaschii* (*Mj*)tyrosyl-tRNA synthetase (TyrRS)/*Mj* ^{Tyr}tRNA, *Methanosarcina mazei* (*Mm*) or *Methanosarcina barkeri* (*Mb*) *pyrrolysyl-tRNA synthetase* (PylRS)/*Mm* ^{Pyl}tRNA, *Methanomethylophilus alvus* (referred to herein as *Alv*)PylRS/*Alv* ^{Pyl}tRNA, *Methanococcus maripaludis (Mmp)* phosphoseryl-tRNA synthetase (SepRS)/*Mj* ^{Sep}tRNA, and *Saccharomyces cerevisiae* (*Sc*) tryptophanyl-tRNA synthetase (TrpRS)/*Sc* ^{Trp}tRNA) are orthogonal with respect to the endogenous synthetases and tRNAs in *Escherichia coli* and have been used as part of strategies to incorporate diverse ncAAs into proteins[10,15–23]. Some combinations of these pairs are mutually orthogonal[5,10,12,14] and have been engineered to incorporate two ncAAs into proteins. One report used derivatives of three existing pairs to incorporate ncAAs in response to all three stop codons. This strategy leaves no codon for termination of endogenous genes; since protein termination is essential to cells we anticipate that efficient variants of this strategy will be toxic. Moreover, to obtain a recombinant protein of homogeneous length a protease site had to be introduced and the translation products proteolytically cleaved, as no codon was available for defining the termination of protein synthesis[12].

The *Mm*PylRS/*Mm* ^{Pyl}tRNA$_{CUA}$ pair, and closely related *Mb*PylRS/*Mb* ^{Pyl}tRNA$_{CUA}$ pair, are the most widely used systems for genetic code expansion[24]. These pairs have gained popularity for several reasons: (1) PylRS/^{Pyl}tRNA pairs are functional across all domains of life and naturally orthogonal to endogenous synthetase/tRNA pairs in both prokaryotic and eukaryotic cells[24]; (2) the ^{Pyl}tRNA anticodon is not recognized by PylRS and can be mutated to create derivatives of the pair targeted to decode diverse codons[14,25]; and (3) the active site of PylRS does not recognize canonical amino acids, accepts several non-natural substrates and can be evolved to specifically incorporate a diverse array of ncAAs[17,24]. The advantages of PylRS/^{Pyl}tRNA systems for genetic code expansion may also be advantages for non-canonical biopolymer synthesis.

*Mm*PylRS is composed of two domains, an N-terminal domain and a C-terminal domain. The C-terminal domain binds the amino acid substrate and catalyses the aminoacylation of *Mm* ^{Pyl}tRNA, while the N-terminal domain makes additional contacts to the variable loop of *Mm* ^{Pyl}tRNA that enhance binding affinity and specificity[26,27]. Both domains are required to create a functional *Mm*PylRS/*Mm* ^{Pyl}tRNA pair in *E. coli*[28]. Until recently it was widely believed that all functional PylRS/^{Pyl}tRNA pairs utilized a two domain PylRS, with bacterial PylRS systems encoding each domain on a separate polypeptide[29].

A new class of highly active PylRS enzymes - from organisms that encode a PylRS gene lacking an N-terminal domain and do not encode a separate N-terminal domain polypeptide - was recently identified[10,19,30] and characterised[10]. These PylRS enzymes, which we refer to as ΔNPylRSs, function with their cognate ^{Pyl}tRNA, which we term $^{\Delta NPyl}$tRNA[10]. Certain ΔNPylRS/$^{\Delta NPyl}$tRNA pairs, or their evolved derivatives, are orthogonal in *E. coli*, and can be used to incorporate ncAAs[10]. Moreover, evolved ΔNPylRS/$^{\Delta NPyl}$tRNA pairs from *M. alvus (Alv)*, in which the variable loop of *Alv* $^{\Delta NPyl}$tRNA has been expanded, are mutually orthogonal to *Mm*PylRS/*Mm* ^{Pyl}tRNA[10]. Subsequent work has shown that derivatives of these pairs are mutually orthogonal in mammalian cells[31,32].

Here we identify new ΔNPylRS/$^{\Delta NPyl}$tRNA pairs, and reveal that the sequences of ΔNPylRS/$^{\Delta NPyl}$tRNA pairs cluster into two distinct classes. Remarkably, the measured specificities of the ΔNPylRSs and $^{\Delta NPyl}$tRNAs correlate with the sequence-based clustering, such that class A ΔNPylRSs preferentially function with class A $^{\Delta NPyl}$tRNAs, and class B ΔNPylRSs preferentially function with class B $^{\Delta NPyl}$tRNAs. We identify 18 mutually orthogonal ΔNPylRS/$^{\Delta NPyl}$tRNA pairs. We go on to evolve class A and class B $^{\Delta NPyl}$tRNAs to generate a set of 12 triply orthogonal pairs. These pairs are composed of a new *Mm*PylRS/*Spe* ^{Pyl}tRNA pair, an evolved class A ΔNPylRS/$^{\Delta NPyl}$tRNA pair and an evolved class B ΔNPylRS/$^{\Delta NPyl}$tRNA pair. We show that these pairs can be engineered to recognize three distinct ncAAs and decode three distinct codons. Finally we show that resulting engineered triply orthogonal pairs can be used to program the cellular incorporation of three distinct ncAAs into a single protein.

Results

Identification of DNPylRS/$^{\Delta NPyl}$tRNA pairs

Fifteen ΔNPylRS family members from methanogenic archaea have been reported to date[10,30], and for seven of these the cognate $^{\Delta NPyl}$tRNA of the pair has been identified in the genome. By performing a HMMER[33,34] search for sequence similarity to the C-terminal catalytic domain of *Mm*PylRS (*Mm*PylRSΔ184), we identified two further ΔNPylRS genes (from *Methanomassiliicoccales archaeon PtaU1.Bin030* (*030*) and *M. archaeon PtaU1.Bin124*),

We used the DNA sequence for each ΔNPylRS gene within its host genome to identify the pyrrolysine gene cluster[19], and in four cases we identified new $^{\Delta NPyl}$tRNA sequences, in the surrounding sequence. This brought the total number of cognate ΔNPylRS/$^{\Delta NPyl}$tRNA pairs to 11(Supplementary Table 1 and 2). All of the identified $^{\Delta NPyl}$tRNAs are predicted to fold into clover-leaf structures typical of the previously reported $^{\Delta NPyl}$tRNAs from the ΔNPylRS

class, sharing the nucleotide loops or bulges in the anticodon stem together with a typical length of D-arm, T-arm, acceptor stem and anticodon stem (Supplementary Fig. 1a).

$^{\Delta NPyl}$tRNAs partition into two classes

We analysed the sequence similarity between the 11 $^{\Delta NPyl}$tRNA sequences (Supplementary Fig. 2), and performed a hierarchical clustering of the tRNA sequences (Fig. 1a). Interestingly, the $^{\Delta NPyl}$tRNA sequences clustered in two groups: one containing *1R26*, *Alv*, *H5, G1,* and *Term* which we termed sequence class A, and the other containing *Lum1*, *Lum2*, *Sheng*, *030*, *Int* and *RumEn* which we termed sequence class B (Fig. 1). Notably, the ΔNPylRSs showed a similar sequence relationship to that of the $^{\Delta NPyl}$tRNA sequences and grouped into the same two classes when the clustering was performed on the percentage of protein sequence identity (Supplementary Fig. 3, 4).

We identified the nucleotides and predicted secondary structures that are conserved within each class of $^{\Delta NPyl}$tRNAs, but differ between sequence class A and sequence class B $^{\Delta NPyl}$tRNAs (Supplementary Fig. 1b). These are mostly located in the acceptor stem, T-stem, and T-loop. We hypothesized that these differences might constitute class-specific identity elements, and that some sequence class A and sequence class B DNPylRS/$^{\Delta NPyl}$tRNA pairs might, therefore, be mutually orthogonal.

Sequence class B $^{\Delta NPyl}$tRNAs contain a cytosine at position 37. Consistent with previous observations[35], we found that C37A mutations in sequence class B $^{\Delta NPyl}$tRNAs led to increased read through of the amber codon by sequence class B ΔNPylRS/$^{\Delta NPyl}$tRNA pairs (Supplementary Fig. 5, Supplementary Table 3). These mutants were used for all further experiments.

Activity, orthogonality and mutual orthogonality of DNPylRS/$^{\Delta NPyl}$tRNA pairs

Next we investigated whether the ΔNPylRS/$^{\Delta NPyl}$tRNA$_{CUA}$ pairs we had identified were active and orthogonal in *E. coli*. We measured the orthogonality of each tRNA by its ability to produce GFP$_{His6}$ from GFP(150TAG)$_{His6}$ in the absence of any exogenous synthetase. Sequence class A $^{\Delta NPyl}$tRNAs led to very low levels of GFP production and sequence class B $^{\Delta NPyl}$tRNAs led to low levels of GFP production (Fig. 1b, Supplementary Fig. 6). We concluded that these tRNAs are functionally orthogonal.

To measure the activity and orthogonality of the cognate ΔNPylRS for each $^{\Delta NPyl}$tRNA, we introduced each cognate ΔNPylRS/$^{\Delta NPyl}$tRNA pair and GFP(150TAG)$_{His6}$ in cells, along with a known substrate for all previously characterized PylRS enzymes (N_{ε}-((tertbutoxy)carbonyl)-l-lysine (BocK **1**)).[10] For all sequence class A $^{\Delta NPyl}$tRNAs, and the sequence class B $^{\Delta NPyl}$tRNAs from *030*, *Lum1* and *RumEn*, coexpression of the cognate synthetase in the presence of BocK **1** led to a substantial increase in read through of the amber codon in GFP(150TAG)$_{His6}$ (Fig. 1b). This indicated that both ΔNPylRS and $^{\Delta NPyl}$tRNA are active in *E. coli*. We observed minimal increase in signal for *Sheng*PylRS, *Lum2*PylRS or *Int*PylRS upon addition of BocK **1** (Supplementary Fig. 7, Supplementary Table 3). We concluded that these PylRSs are likely to be either not expressed or not functional in *E. coli* and we did not consider them further. *BRNA*PylRS was only weakly

active with its cognate $^{\Delta NPyl}$tRNA and this tRNA was identical in sequence to the tRNA we identified from *1R26*. Since the *1R26*PylRS/*1R26* $^{\Delta NPyl}$tRNA pair was substantially more active than the *BRNA*PylRS/*1R26* $^{\Delta NPyl}$tRNA pair we decided to proceed with *1R26*PylRS (Supplementary Fig. 7, Supplementary Table 3). ΔNPylRSs for which a cognate tRNA has not been identified from the same host (from *Methanonatronarchaeum thermophilum*, *Methanohalarchaeum thermophilum*, *MSBL1 archaeon SCGCAAA382A20* or *M. archaeon PtaU1.Bin124*) were not characterized any further.

Time-of-flight mass spectrometry (TOF-MS) of GFP_{His6} produced from $GFP(150TAG)_{His6}$ in the presence of BocK confirmed that the aaRSs from cognate pairs (*Alv*PylRS/*Alv* $^{\Delta NPyl}$tRNA, *H5*PylRS/*H5* $^{\Delta NPyl}$tRNA, *1R26*PylRS/*1R26* $^{\Delta NPyl}$tRNA, *G1*PylRS/*G1* $^{\Delta NPyl}$tRNA, *Term*PylRS/*Term* $^{\Delta NPyl}$tRNA, *030*PylRS/*030* $^{\Delta NPyl}$tRNA, *RumEn*PylRS/*RumEn* $^{\Delta NPyl}$tRNA, and *Lum1*PylRS/*Lum1* $^{\Delta NPyl}$tRNA) are functionally orthogonal (Supplementary Fig. 8).

We experimentally characterized the activity of all 88 ΔNPylRS/$^{\Delta NPyl}$tRNA combinations arising from 11 tRNAs and eight synthetases (Fig. 1b). This provided functional data on the activity of cognate and non-cognate combinations. When clustered on the basis of their relative activity, ΔNPylRSs and $^{\Delta NPyl}$tRNAs each form two classes, which we defined as functional class A and functional class B.

Functional class A synthetases were highly active with functional class A $^{\Delta NPyl}$tRNAs and most functional class A synthetases showed lower, but measurable, activity when paired with the majority of functional class B $^{\Delta NPyl}$tRNAs. Notably, *G1*PylRS from functional class A was orthogonal with respect to several functional class B $^{\Delta NPyl}$tRNAs. Functional class B synthetases, except *Term*PylRS, were highly active with functional class B $^{\Delta NPyl}$tRNAs, but not with most functional class A $^{\Delta NPyl}$tRNAs. *Term*PylRS - a functional class B synthetase - was exclusively selective for its cognate tRNA. *Term* $^{\Delta NPyl}$tRNA was classified as a functional class A $^{\Delta NPyl}$tRNA and was a substrate for all ΔNPylRSs tested.

There is striking correspondence between the classes identified on the basis of sequence and the classes identified on the basis of function for both the ΔNPylRSs and $^{\Delta NPyl}$tRNAs (Fig. 1a, b and Supplementary Fig. 9). Seven out of eight ΔNPylRSs, and all 11 $^{\Delta NPyl}$tRNAs, partition into the same two groups on the basis of sequence and function. We defined consensus class A and class B ΔNPylRSs and $^{\Delta NPyl}$tRNAs, which we subsequently referred to as simply class A and class B. Class A contains ΔNPylRSs and $^{\Delta NPyl}$tRNAs from *Alv*, *G1*, *H5* and *1R26*, and class B contains ΔNPylRSs and $^{\Delta NPyl}$tRNAs from *Lum1*, *Lum2*, *Sheng*, *030*, *Int* and *RumEn*. We excluded the *Term* pair (for which the synthetase and tRNA partitioned into different functional classes) from either consensus class and we did not consider this pair further. Thus we proceeded with ten $^{\Delta NPyl}$tRNAs and seven ΔNPylRSs.

The correlation between the tRNA classes defined by sequence and the tRNA classes defined by function suggests that the differences in tRNA sequence between each class contain the distinct nucleotide identity elements that specify class-specific synthetase recognition. The preference of class A tRNAs for class A synthetases and the preference of class B tRNAs for class B synthetases will lie within these nucleotide sequences. Similarly,

the correlations between synthetase classes defined by sequence, and those defined by function, suggest that the class-specific amino acid sequences within synthetases define class-specific tRNA recognition. The preference of class A synthetases for class A tRNAs and the preference of class B synthetases for class B tRNAs will lie within these amino acid sequences.

Analysis of our data allowed us to identify 18 different combinations of naturally mutually orthogonal DNPylRS/$^{\Delta NPyl}$tRNA pairs with good orthogonality and activity. These pairs are composed of *G1*ΔNPylRS with any class A $^{\Delta NPyl}$tRNA and a select class B ΔNPylRS with either *Lum1* $^{\Delta NPyl}$tRNA or *Lum2* $^{\Delta NPyl}$tRNA (Fig. 1c, d).

The MmPylRS/Mm^{Pyl}tRNA pair is not orthogonal with respect to many DNPylRSs or $^{\Delta NPyl}$tRNAs

Next we characterized the activity of *Mm* ^{Pyl}tRNA with each of the seven DNPylRS enzymes, and the activity of *Mm*PylRS with all ten $^{\Delta NPyl}$tRNAs (Fig. 2).We found that all of the class B synthetases are functional with *Mm* ^{Pyl}tRNA while all of the class A synthetases, except *1R26*, do not function with *Mm* ^{Pyl}tRNA (Fig. 2a). We also found that *Mm*PylRS is active with all $^{\Delta NPyl}$tRNAs tested, although it is least active with *G1* $^{\Delta NPyl}$tRNA (Fig. 2b, Supplementary Fig. 10, Supplementary Table 3). These observations define the activity of the *Mm*PylRS/*Mm* ^{Pyl}tRNA pair with respect to class A and B synthetases and $^{\Delta NPyl}$tRNAs. In combination with the activities of class A and B ΔNPylRS/$^{\Delta NPyl}$tRNA pairs (Fig. 1), these experiments define the activities we need to engineer to create a set of three mutually orthogonal pairs composed of a class A ΔNPylRS/$^{\Delta NPyl}$tRNA pair, a class B ΔNPylRS/ $^{\Delta NPyl}$tRNA pair, and an *Mm*PylRS/^{Pyl}tRNA pair (Fig. 2c).

To discover triply orthogonal PylRS/^{Pyl}tRNA pairs we needed to create: (1) a tRNA that is a substrate for *Mm*PylRS, while being orthogonal to a class A DNPylRS, a class B DNPylRS, and endogenous synthetases; (2) a tRNA that is only active with class A DNPylRSs while being orthogonal to both *Mm*PylRS, class B DNPylRSs, and endogenous synthetases; and (3) a tRNA that is solely a substrate for class B DNPylRSs while being orthogonal to *Mm*PylRS, class A DNPylRSs, and endogenous synthetases. In our subsequent experiments we addressed each of these challenges in turn.

Discovering class +N $^{+NPyl}$tRNAs that are orthogonal to class A and B DNPylRSs

As *Mm* ^{Pyl}tRNA functions with both class B DNPylRSs and *1R26*PylRS (from class A), it is not compatible with the creation of triply orthogonal PylRS/^{Pyl}tRNA pairs. We therefore set out to discover ^{Pyl}tRNAs that are both active with *Mm*PylRS, and orthogonal with respect to class A and class B DNPylRSs.

For many PylRS enzymes with an N- and C-terminal domain (denoted class +N herein) the cognate tRNAs (denoted $^{+NPyl}$tRNAs herein) remain unannotated and unknown. We took advantage of existing genome annotations to identify the DNA sequence coding for class +N PylRSs within their host genomes and thereby identified ten $^{+NPyl}$tRNA genes in the surrounding sequence; for six of these sequences we are not aware of previous annotations or reports (Supplementary Table 4). In combination with *Mm* ^{Pyl}tRNA$_{CUA}$ and *Mb* ^{Pyl}tRNA$_{CUA}$ these sequences provided 12 $^{+NPyl}$tRNA sequences for further investigation.

We measured the activity of the 12 $^{+NPyl}$tRNAs with class A and B DNPylRSs, with *Mm*PylRS, or without any PylRS. We discovered six $^{+NPyl}$tRNAs (from *Bur*, *Met*, *Pro*, *Psy*, *Spe* and *Vul*) that were orthogonal to all the DNPylRSs tested, orthogonal with respect to endogenous synthetases and formed active, heterologous pairs with *Mm*PylRS (Fig 3a, Supplementary Fig. 11, Supplementary Table 3).

Among these six heterologous pairs, the *Mm*PylRS/*Spe* ^{Pyl}tRNA pair was the most active, with similar activity to the native *Mm*PylRS/*Mm* ^{Pyl}tRNA pair (Fig 3a). The pair produced GFPHis$_6$ from GFP(150TAG)His$_6$ at 80% of the level of wild type GFPHis$_6$ (Supplementary Fig. 12, Supplementary Table 3). Remarkably, *Spe* ^{Pyl}tRNA only differs from *Mm* ^{Pyl}tRNA by a single point mutation in the acceptor stem, changing the C-G Watson-Crick base pair to a U-G wobble (Fig. 3b). These experiments established the *Mm*PylRS/*Spe* ^{Pyl}tRNA pair as the class +N pair for a mutually orthogonal PylRS/^{Pyl}tRNA triplet (Fig. 3c).

Engineered class A $^{\Delta NPyl}$tRNAs are orthogonal to class B DNPylRSs and MmPylRS

Next we aimed to discover a tRNA that is only active with class A DNPylRSs while being orthogonal to *Mm*PylRS, class B DNPylRSs and endogenous synthetases. We have previously demonstrated that expanding the variable loop of *Alv* $^{\Delta NPyl}$tRNA (a class A $^{\Delta NPyl}$tRNA) destroys its activity with *Mm*PylRS without affecting activity with its cognate synthetase, *Alv*PylRS[10].

We tested 12 previously described *Alv* $^{\Delta NPyl}$tRNA variants with expanded variable loops that are active with *Alv*PylRS (Fig. 4a, Supplementary Fig. 13, Supplementary Table 5)[10]. All 12 variants showed good activity with class A synthetases, and little or no activity with *Mm*PylRS, class B synthetases or endogenous synthetases (Fig. 4b). We found three variants (*Alv* $^{\Delta NPyl}$tRNA(8), *Alv* $^{\Delta NPyl}$tRNA(11) and *Alv* $^{\Delta NPyl}$tRNA(19)) that had good orthogonality towards both *Mm*PylRS and class B synthetases while remaining highly active with *Alv*PylRS and other class A synthetases (*G1*PylRS, *H5*PylRS and *1R26*PylRS) (Fig. 4b). These three *Alv* $^{\Delta NPyl}$tRNA variants are more orthogonal than the parent *Alv* $^{\Delta NPyl}$tRNA with respect to *RumEn*PylRS, *030*PylRS, *Lum1*PylRS and *Mm*PylRS. These experiments established *G1*PylRS, *Alv*PylRS, *H5*PylRS or *1R26*PylRS in combination with *Alv* $^{\Delta NPyl}$tRNA(8), *Alv* $^{\Delta NPyl}$tRNA(11) or *Alv* $^{\Delta NPyl}$tRNA(19) as the class A-derived pairs for a set of triply orthogonal PylRS/^{Pyl}tRNA pairs (Fig. 4c).

Directed evolution of class B $^{\Delta NPyl}$tRNAs orthogonal to class A DNPylRSs and MmPylRS

Class B $^{\Delta NPyl}$tRNAs function efficiently with *Mm*PylRS and are active with some class A DNPylRSs. Both of these activities are incompatible with the creation of a set of triply orthogonal PylRS/^{Pyl}tRNA pairs. We decided to take a convergent approach to evolving a class B $^{\Delta NPyl}$tRNA that was active with class B DNPylRSs but did not function with *Mm*PylRS or class A DNPylRSs (Fig. 5a). We focused on *Int* ^{DNPyl}tRNA, which has good activity with class B DNPylRSs and lower activity with many class A DNPylRSs than other class B ^{DNPyl}tRNAs (Fig. 1b).

We hypothesized that we could abolish any recognition of *Int* ^{DNPyl}tRNA by *Mm*PylRS by expanding its variable loop[10]. We created a variable loop library of *Int* $^{\Delta NPyl}$tRNA mutants in which the length of the variable loop was expanded from three nucleotides to four, five or

six randomized nucleotides. In each variable loop library, positions 13, 14, 15, 54 and 55 (in the D-loop and T-loop) (Fig. 5a), which may make contacts with the variable loop in the folded tertiary structure of the tRNA, were also randomized[27].

We selected *Int*$^{\Delta NPyl}$tRNA variants that functioned with a class B DNPylRS (*030*PylRS) and enabled cells to grow on 100 μg ml^{-1} chloramphenicol in the presence of BocK **1**, by facilitating read through of an amber codon at position 111 of a chloramphenicol acetyl transferase reporter, cat(111TAG). Next we performed two negative screens on the selected *Int*$^{\Delta NPyl}$tRNA variants to identify tRNAs that do not function with *Mm*PylRS, a class A synthetase (*Alv*PylRS) or any endogenous *E. coli* synthetase. Cells bearing *GFP(150TAG)*$_{His6}$, *Mm*PylRS or *Alv*PylRS, and each *Int*$^{\Delta NPyl}$tRNA variant were provided with BocK **1** and screened for the absence of GFP$_{His6}$ expression. This serial positive selection and double negative screen identified two evolved mutants, *Int*$^{\Delta NPyl}$tRNA(^{V}B03) and *Int*$^{\Delta NPyl}$tRNA(^{V}C10), each with single nucleotide insertions into the variable loop. These tRNAs form *030*PylRS/*Int*$^{\Delta NPyl}$tRNA pairs which retain two thirds of the activity of the initial *030*PylRS/*Int*$^{\Delta NPyl}$tRNA wild type pair, but have much improved orthogonality towards *Mm*PylRS, *Alv*PylRS and endogenous synthetases (Supplementary Fig. 14, Supplementary Table 6).

The majority of nucleotides that distinguish class A and class B $^{\Delta NPyl}$tRNAs are clustered in the acceptor stem (Supplementary Fig. 1b). This observation suggested that the acceptor stem may be an important recognition element in determining tRNA substrate specificity by DNPylRSs. We decided to evolve the acceptor stem of a class B $^{\Delta NPyl}$tRNA to discover variants that selectively diminish recognition by class A synthetases without affecting activity with class B synthetases.

We created an acceptor stem library of *Int*$^{\Delta NPyl}$tRNA mutants by randomizing nucleotide positions 3-7 and 61-65 (Fig. 5a). We performed a serial positive selection and double negative screen, analogous to that described for the variable loop library, on the acceptor stem library. Following this process we isolated 11 *Int*$^{\Delta NPyl}$tRNA acceptor stem mutants that retained over 85% of the activity of wild type *Int*$^{\Delta NPyl}$tRNA when paired with *030*PylRS (Supplementary Fig. 15, Supplementary Table 7). However, the evolved variants still displayed substantial cross-reactivity with *Mm*PylRS and some activity with *Alv*PylRS, although both of these were diminished when compared with wild type *Int*$^{\Delta NPyl}$tRNA. One evolved variant, *Int*$^{\Delta NPyl}$tRNA(^{A}17), displayed a level of orthogonality towards both *Mm*PylRS and *Alv*PylRS that was comparable to *Int*$^{\Delta NPyl}$tRNA(^{V}B03) and *Int*$^{\Delta NPyl}$tRNA(^{V}C10), identified from the variable loop library. All selected variants were orthogonal with respect to endogenous synthetases.

We postulated that combining the mutations isolated from each library might create a tRNA with the desired properties. We therefore transplanted the variable loop extensions from *Int*$^{\Delta NPyl}$tRNA(^{V}B03) and *Int*$^{\Delta NPyl}$tRNA(^{V}C10) into each variant identified from the acceptor stem library, and tested the resulting hybrid tRNAs for activity with *030*PylRS, *Mm*PylRS and *Alv*PylRS. We found that introduction of an extra nucleotide into the variable loop was very effective in removing any cross-reactivity towards *Mm*PylRS, rendering all active hybrids highly orthogonal to *Mm*PylRS (Supplementary Fig. 16, Supplementary Table 3, 8).

We identified four highly active hybrid *Int*$^{\Delta NPyl}$tRNAs (*Int*$^{\Delta NPyl}$tRNA(^{A}5,^{V}B03), *Int*$^{\Delta NPyl}$tRNA(^{A}5,^{V}C10) *Int*$^{\Delta NPyl}$tRNA(^{A}6,^{V}B03) and *Int*$^{\Delta NPyl}$tRNA(^{A}6,^{V}C10)), which retained greater than 80% of wild type activity when paired with *030*PylRS and were orthogonal with respect to *Mm*PylRS. However, these four tRNAs still displayed activity with *Alv*PylRS. Another three tRNAs (*Int*$^{\Delta NPyl}$tRNA(^{A}13,^{V}C10), *Int*$^{\Delta NPyl}$tRNA(^{A}17,^{V}B03) and *Int*$^{\Delta NPyl}$tRNA(^{A}17,^{V}C10)) retained over 50% of wild type activity with *030*PylRS and were highly orthogonal with respect to both *Mm*PylRS and *Alv*PylRS (Fig. 5b, Supplementary Fig. 17). All seven tRNAs were orthogonal with respect to *E. coli* synthetases, and provided promising candidates for our third tRNA as part of a triply orthogonal PylRS/^{Pyl}tRNA pair.

Next we investigated the activity of the seven evolved *Int*$^{\Delta NPyl}$tRNAs with the three additional class A synthetases (*G1*, *H5* and *1R26*) and two additional class B synthetases (*RumEn* and *Lum1*) that were not used for the selection or initial characterization. Overall we characterized 64 combinations of class A synthetases, class B synthetases or *Mm*PylRS with *Int*$^{\Delta NPyl}$tRNA and its selected variants (Fig 5b). From these experiments we identified numerous *Int*$^{\Delta NPyl}$tRNA variants that are orthogonal to specific class A synthetases and *Mm*PylRS while being highly active with a specific class B synthetase (Fig. 5b,c). For example *Int*$^{\Delta NPyl}$tRNA(^{A}6,^{V}B03) is very active with *Lum1*PylRS (class B) but inactive with *Mm*PylRS or *G1*PylRS (class A).

Triply orthogonal PylRS/^{Pyl}tRNA pairs

By combining the observations from our measurements of cognate and non-cognate synthetase/tRNA pair activity (Fig. 1b, Fig. 3a, Fig. 4b and Fig. 5b) we identified 12 sets of triply orthogonal pairs (Fig. 5d, Supplementary Fig. 18). Each triplet is composed of *Mm*PylRS and *Spe*^{Pyl}tRNA, a specific class A DNPylRS and an evolved *Alv*^{DNPyl}tRNA, and a specific class B DNPylRS and an evolved *Int*^{DNPyl}tRNA variant.

Incorporating three ncAAs into a protein

The three pairs within a triply orthogonal pair each recognize the same amino acid and decode the amber codon. To use these pairs to encode three distinct ncAAs into a polypeptide, it was necessary to: (1) diverge the active sites of each synthetase to recognize distinct amino acids; and (2) diverge the anticodon of their cognate tRNAs to decode distinct blank codons. We focused on addressing these challenges with the *Mm*PylRS/*Spe*^{Pyl}tRNA, *Lum1*PylRS/*Int*^{DNPyl}tRNA and *1R26*PylRS/*Alv*^{DNPyl}tRNA pairs, which are among the most active triply orthogonal pairs (Fig. 5d).

We demonstrated that the active sites of *Lum1*PylRS, *1R26*PylRS and *Mm*PylRS can be diverged to selectively recognize distinct ncAA substrates (Fig. 6a-d). As previously reported, wild type *Mm*PylRS directs the incorporation of BocK **1**, but not 3-methyl-lhistidine (NmH **2**)[32] or N_{ε} -(carbobenzyloxy)-l-lysine (CbzK **3**) (Fig. 6a,d) [36]. Previous work has shown that five mutations (L121M, L125I, Y126F, M129A and V168F) convert *Mb*PylRS into an aaRS for NmH **2**, and that transplanting these mutations into *Alv*PylRS creates a mutant that directs the incorporation of NmH **2** and excludes BocK **1** [32]. We transplanted the same mutations into *Lum1*PylRS, creating *Lum1*PylRS(NmH), and

demonstrated that this enzyme directs the incorporation of NmH **2**, but not BocK **1**; this shows that that the activity and specificity conferred by these mutations in *Alv*PylRS can be transplanted to *Lum1*PylRS. We also showed that *Lum1*PylRS(NmH) does not direct the incorporation of CbzK **3** (Fig. 6a,b). Finally, we identified a mutant of *1R26*, *1R26*PylRS(CbzK) that directs the incorporation of CbzK **3**, but not NmH **2** or BocK **1** (Fig. 6a,c); the mutations conferring this specificity were discovered in a clone within a laboratory collection of *Mm*PylRS mutants and then transferred to *1R26*PylRS.

Next we created *Mm*PylRS/*Spe* ^{Pyl}tRNA$_{CUA}$, *Lum1*PylRS(NmH)/*Int* ^{DNPyl}tRNA(^{A}17, ^{V}C10)$_{UCCU}$ and *1R26*PylRS(CbzK)/*Alv* ^{DNPyl}tRNA(8)$_{UACU}$ pairs; the synthetase of each pair selectively recognizes its cognate ncAA and the tRNA of each pair contains a CUA, UCCU or UACU anticodon that targets TAG, AGGA or AGTA codons, respectively. We coexpressed all three pairs in a single cell with ribo-Q1 (an orthogonal ribosome that facilitates the decoding of quadruplet codons and amber codons, on its cognate orthogonal message, by tRNAs bearing complementary anticodons[5]). Cells also contained an OGST(1XXX)CAM gene (glutathione S transferase (GST) linked to calmodulin (CAM) on an orthogonal ribosome binding site (O-rbs); XXX = TAG, AGGA, or AGTA) that was selectively translated by ribo-Q1[14]. Read through of the TAG codon, as judged by production of full length GST-CAM, was observed upon addition of BocK **1** (Fig. 6e). Addition of CbzK **3** or NmH **2** led to low-level read through of the TAG codon; such background incorporation is typically outcompeted in the presence of the cognate pair and ncAA[37]. These observations are consistent with the *Mm*PylRS/*Spe* ^{Pyl}tRNA$_{CUA}$ pair selectively recognizing BocK **1** and decoding the amber codon. Similarly, we observed read through of the AGGA codon upon addition of NmH **2** but not BocK **1** or CbzK **3**(Fig. 6e, Supplementary Fig. 19), and read through of the AGTA codon upon addition of CbzK **3** but not NmH **2** or BocK **1** (Fig. 6e). These observations are consistent with the expected specificity of the *Lum1*PylRS(NmH)/*Int* ^{DNPyl}tRNA(^{A}17, ^{V}C10)$_{UCCU}$ and *1R26*PylRS(CbzK)/*Alv* ^{DNPyl}tRNA(8)$_{UACU}$ pairs. Our data suggest that we have created derivatives of the triply orthogonal pairs that can be expressed in the same cell, recognize distinct ncAAs and decode distinct codons.

Finally, we demonstrated that our new pairs can be used to incorporate three distinct ncAAs into a single protein in response to three distinct codons (Fig. 6f,g,h, Supplementary Figs. 20 and 21). We co-expressed all three pairs and ribo-Q1 in cells also containing O-GFP (40TAG, 136AGGA and 150AGTA) in which translation of a GFP gene containing the target codons is driven from an O-rbs. Full length GFP was produced upon addition of CbzK **3**, NmH **2** and BocK **1** to cells (Fig. 6f) and mass spectrometry confirmed the incorporation of all three ncAAs at genetically encoded positions in the protein (Fig. 6g, Supplementary Fig. 22).

Discussion

We have discovered and characterized several new DNPylRS/$^{\Delta NPyl}$tRNA pairs, and thereby expanded the diversity of ΔNPylRS/$^{\Delta NPyl}$tRNA pairs that are highly active and orthogonal in *E. coli*. Clustering analysis of our expanded set of DNPylRS and $^{\Delta NPyl}$tRNA sequences revealed two distinct classes. These sequence-based classes correlated well with the classes

that we independently determined on the basis of function. We demonstrated that class A synthetases function with class A $^{\Delta NPyl}$tRNAs, but commonly function poorly with class B $^{\Delta NPyl}$tRNAs, and that class B synthetases are naturally orthogonal towards class A $^{\Delta NPyl}$tRNAs and function efficiently with class B $^{\Delta NPyl}$tRNAs. By characterizing 88 DNPylRS/$^{\Delta NPyl}$tRNA combinations we discovered 18 mutually orthogonal DNPylRS/$^{\Delta NPyl}$tRNA pairs. These experiments reveal a remarkable divergence in sequence and function between DNPylRS/$^{\Delta NPyl}$tRNA pairs.

We discovered an *Mm*PylRS/*Spe* ^{Pyl}tRNA pair in which *Spe* ^{Pyl}tRNA is orthogonal to both class A and class B DNPylRSs, but functions efficiently with *Mm*PylRS. We discovered variants of a class A $^{\Delta NPyl}$tRNA that are orthogonal to *Mm*PylRS and class B synthetases, but function efficiently with class A synthetases. Similarly, we evolved variants of a class B $^{\Delta NPyl}$tRNA that are orthogonal to *Mm*PylRS and certain class A synthetases, but function efficiently with certain class B synthetases. By analysing the specificity of synthetase/tRNA combinations from 248 measurements we identified 12 sets of triply orthogonal pairs: each triplet is composed of the *Mm*PylRS/*Spe* ^{Pyl}tRNA pair, an evolved class A DNPylRS/$^{\Delta NPyl}$tRNA pair and an evolved class B DNPylRS/$^{\Delta NPyl}$tRNA pair. Remarkably, each set of three triply orthogonal pairs, in which each pair is orthogonal with respect to each other and to *E. coli* synthetases and tRNAs (*Mm*PylRS/ *Spe* ^{Pyl}tRNA, an evolved class A DNPylRS/$^{\Delta NPyl}$tRNA, and an evolved class B DNPylRS/$^{\Delta NPyl}$tRNA), now equals the number of broadly useful orthogonal pairs in *E. coli* derived from all other aaRS/tRNA systems. We note that other orthogonal pairs with limited demonstrated utility for ncAA incorporation have also been reported[38,39].

We have engineered the active site of the synthetases to generate mutants that recognize distinct ncAAs and exclude others. We have also engineered the pairs to decode distinct codons and demonstrated the incorporation of three distinct ncAAs into a single polypeptide. We note that we have previously reported a scalable strategy for discovering synthetase mutants that use a desired ncAA substrate but exclude other ncAAs[40]; extensions of this approach, combined with the advances reported herein, should enable an expansion in the diversity of ncAAs that can be incorporated into a single polypeptide. As the anticodon of the pyrrolysyl tRNAs can be altered without destroying aminoacylation by their cognate synthetases[25,41](Fig. 6) it may be possible to reprogram the triply orthogonal pairs to incorporate ncAAs in response to other emerging blank codons[3,4,42].

Our work enables the co-translational incorporation of three distinct ncAAs into a homogeneous and correctly terminated protein. We anticipate that future work will combine advances in expanding the number of codons available for assignment to non-canonical monomers[3–5] with advances in expanding the chemical scope of the ribosome[6–8] and the mutually orthogonal systems reported herein to realize the encoded cellular synthesis and evolution of non-canonical biopolymers.

Methods

Identification of ΔNPylRS sequences

We identified PylRS protein sequences homologous to the C-terminal region of *Mm*PylRS or to *Desulfitobacterium hafniense* (*Dh)*PylSn by protein HMMER search[34] against the UniProtKB database[43] using *Mm*PylRSd184 or *Dh*PylSn as the query sequence, respectively, and filtering for expect values below 1×10^{-30}. From the identified protein sequences, which contain homology to the C-terminal region of *Mm*PylRS, we eliminated those for which sequence homology to *Dh*PylSn could be found within the same genome. From the remaining protein sequences, which correspond to ΔNPylRSs, we identified those that had not been previously reported.

Identification of ^{Pyl}tRNA sequences

Using the National Center for Biotechnology Information Nucleotide Database we used existing genome annotations to identify the DNA sequence for each PylRS gene within its host genome, and were thus able to identify the genomic region corresponding to the pyrrolysine gene cluster. In the sequences 40 kilobases upstream and downstream of the PylRS gene, $^{\Delta NPyl}$tRNA sequences were manually identified by searching for sequence similarity to known $^{\Delta NPyl}$tRNA sequences; $^{+NPyl}$tRNA sequences were manually identified by searching for sequence similarity to *Mm* ^{Pyl}tRNA. tRNA secondary structure prediction was initially performed using *RNAstructure* [44] and manually curated by inspection and comparison to *Mm* ^{Pyl}tRNA.

DNA constructs

PylRS and ^{Pyl}tRNA genes were synthesiszd by IDT as gBlock double-stranded DNA fragments. We cloned the genes into pKW vectors by Gibson assembly. PylRS was expressed from a *glnS* promoter and ^{Pyl}tRNA was expressed from an *lpp* promoter. PylRS genes were coded for expression in *E. coli* using the *IDT Codon Optimization Tool*. We appended the gene for *Mm*PylRS with a sequence encoding a C-terminal $Ser(Gly_4Ser)_4$FLAG-tag, while all other PylRS genes were appended with a sequence encoding a C-terminal $Ser(Gly_4Ser)_4His_6$SerGlyStrep-tag II. We used these plasmids together with pBAD GFP(150TAG)His_6 (in which sfGFP containing an amber stop codon at position 150 and a C-terminal His_6 tag is expressed from the arabinose promoter of pBAD; GFP refers to sfGFP throughout). We used Gibson cloning to insert each PylRS cassette under constitutive expression from the *glnS* promoter into pBAD CAT(111TAG) GFP(150TAG)His_6 vectors, in which a chloramphenicol acetyl transferase gene containing an amber stop codon at position 111 is under constitutive expression.

To create the plasmid pKW1-Triple for triple ncAA incorporation, PylRS genes were designed to be expressed as a single polycistronic messenger RNA transcript under the control of the glnS promoter, with RBS binding strengths of approximately 10,000 RBS units rationally designed using the RBS Calculator (https://www.denovodna.com/software/)[45–49] specifying *E. coli* K-12 as the host organism. tRNA genes were designed to be expressed as a single polycistronic messengerRNA transcript under the control of the lpp promoter. Sequences between tRNA sequences were designed by manual examination of the

E. coli K-12 MG1655 genome using EcoCyc[50] and identifying spacer sequences between tRNAs from the same isoacceptor class that are expressed as adjacent tRNA sequences in the same operon. Spacer sequences originating between AlaX and AlaW, and ValU and ValX genes were selected for use. Cassettes containing PylRS and PyltRNA genes were synthesized by IDT as gBlock double-stranded DNA fragments. We cloned the genes into pKW vectors by Gibson assembly.

Library generation

Libraries of *Int* $^{\Delta NPyl}$tRNA with randomized variable loop or acceptor stem sequences were constructed by Golden Gate cloning from a pKW *Int* $^{\Delta NPyl}$tRNA vector using PCR primers listed in Supplementary Table 11 together with restriction enzyme BbsI and T4 DNA ligase. We transformed each library separately into competent *E. coli* DH10B cells to give library diversities of more than 1×10^8, exceeding the theoretical diversity of 6×10^7 required for complete library coverage.

Selection to identify orthogonal Class A $^{\Delta NPyl}$tRNAs

For the variable loop library, we transformed each *Int* $^{\Delta NPyl}$tRNA variable loop library into competent *E. coli* DH10B cells bearing pBAD *Int*PylRS CAT(111TAG) GFP(150TAG)His_6. We recovered the transformed cells for 1 h at 37°C in 0.5 ml SOB medium supplemented with 8 mM BocK **1**. The transformation was plated on LB agar containing 37.5 μg ml^{-1} spectinomycin, 12.5 μg ml^{-1} tetracycline and 100 μg ml^{-1} chloramphenicol. The plates were incubated at 37 °C for 40 h. After incubation, colonies on the plates were washed off and collected in 2XTY buffer and the plasmids were extracted using a DNA midiprep kit (Qiagen). To remove the pBAD *Int*PylRS CAT(111TAG) GFP(150TAG)His_6 plasmid, the extracted DNA was digested with both NcoI restriction endonuclease and T5 exonuclease and re-purified using a PCR purification column. The remaining pKW plasmids were transformed into competent *E. coli* DH10B cells bearing either pBAD *Alv*PylRS CAT(111TAG) GFP(150TAG)His_6 or pBAD *Mm*PylRS CAT(111TAG) GFP(150TAG)His_6. The transformed cells were recovered for 1 h at 37 °C in 0.5 ml SOB medium. The transformation was plated on LB agar containing 37.5 μg ml^{-1} spectinomycin and 12.5 μg ml^{-1} tetracycline. The plates were incubated at 37 °C for 20 h. For each library, 1,528 colonies were picked from the plates using a QPix 420 Colony Picking System and inoculated into 190 μl 2XTY-STA (2XTY medium with 75 μg/ml spectinomycin, 25 μg/ml tetracycline and 0.5% *l*-arabinose) in 96-well microtitre plate format supplemented with 8 mM BocK **1**. The plates were incubated at 37°C and 220 r.p.m., and OD_{600} and GFP fluorescence (λ_{ex} 485 nm, λ_{em} 520 nm) measurements were recorded after 20 h using a SpectraMax i3. Cells from wells with the lowest GFP OD_{600} $^{-1}$ ratios were used to inoculate 2XTY medium with 75 μg ml^{-1} spectinomycin, and the pKW plasmids containing *Int* $^{\Delta NPyl}$tRNA variants were extracted by DNA miniprep and then sequenced. Each hit corresponding to a distinct *Int* $^{\Delta NPyl}$tRNA sequence was cloned into a pKW *Int*PylRS vector, a pKW *Alv*PylRS vector, and a pKW *Mm*PylRS vector and rephenotyped with pBAD GFP(150TAG)His_6.

For the acceptor stem library, we transformed each *Int* $^{\Delta NPyl}$tRNA acceptor stem library into competent *E. coli* DH10B cells bearing pBAD *Int*PylRS CAT(111TAG) GFP(150TAG)His_6.

We recovered the transformed cells for 1 h at 37°C in 5 ml super optimal broth with catabolite repression (SOC) medium supplemented with 8 mM BocK **1**. The transformation was plated on LB agar containing 75 μg ml^{-1} spectinomycin, 25 μg ml^{-1} tetracycline and 100 μg ml^{-1} chloramphenicol. The plates were incubated at 37 °C for 24 h. After incubation, 192 colonies were picked into 1.7 ml 2XTY-STA in a 96-well microtitre plate format supplemented with 8 mM BocK **1** and grown overnight. Plasmids from all fluorescent cultures were extracted by DNA miniprep (Qiagen) and the extracted DNA was digested with both NcoI restriction endonuclease and T5 exonuclease. 1 μl of the digestion products was transformed into chemically competent *E. coli* DH10B cells bearing either pBAD *Alv*PylRS CAT(111TAG) GFP(150TAG)His_6 or pBAD *Mm*PylRS CAT(111TAG) GFP(150TAG)His_6 by heat shock. The transformed cells were recovered for 1 h at 37 °C in 180 μl ml SOC medium, and 10 μl was used to inoculate 180 μl 2XTYSTA in a 96-well microtitre plate format supplemented with 8 mM BocK **1** and grown overnight. Cells from wells with the lowest GFP OD_{600} ratios were used to inoculate 2XTY medium with 75 μg ml^{-1} spectinomycin, and the pKW plasmids containing *Int* $^{\Delta NPyl}$tRNA variants were extracted using a DNA miniprep kit (Qiagen) and then sequenced. Each hit corresponding to a distinct *Int* $^{\Delta NPyl}$tRNA sequence was cloned into a pKW *Int*PylRS vector, a pKW *Alv*PylRS vector, and a pKW *Mm*PylRS vector and re-phenotyped with pBAD GFP(150TAG)His_6.

Measuring the activity and specificity of PylRS/^{Pyl}tRNA$_{CUA}$ pairs with synthetase and tRNA expressed from different plasmids

To measure the activity and specificity of cognate and non-cognate PylRS/^{Pyl}tRNA combinations, we transformed 0.4 μl pKW ^{Pyl}tRNA plasmids into 8 μl chemically competent *E. coli* DH10B cells bearing either pBAD GFP(150TAG)His_6 or pBAD PylRS GFP(150TAG)His_6. We recovered the transformed cells for 1 h at 37 °C and 750 r.p.m. in 180 μl SOC medium in 96-well microtitre plate format. 10 μl of the transformed cells was used to inoculate 180 μl 2XTY-STA in 96-well microtitre plate format, supplemented with or without 8 mM BocK **1**. OD_{600} and GFP fluorescence (λ_{ex} 485 nm, λ_{em} 520 nm) measurements were recorded after 22-28 h incubation at 37 °C and 700 r.p.m. using a Tecan Infinite M200 Pro.

Measuring the activity and specificity of PylRS/^{Pyl}tRNA$_{CUA}$ pairs with synthetase and tRNA expressed from the same plasmid

The same procedure was followed as described above. However, for this expression system both PylRS and ^{Pyl}tRNA were encoded on the same pKW plasmid which was transformed into chemically competent *E. coli* DH10B cells bearing pBAD GFP(150TAG)His_6. For GFP expression, 25 μl transformed cells was inoculated into 500 μl 2XTY-STA in 96-well microtitre plate format, in the presence or absence of 8 mM BocK **1**, or 2 mM CbzK **3** or 8 mM NmH **2**. Cells were grown for 22-28 h at 750 r.p.m. and 37 °C before 180 μl of each well was transferred to a 96 well plate and GFP fluorescence and OD_{600} measurements were recorded as described above.

GFP(TAG)$_{His6}$ expression for mass spectrometry

To express GFP incorporating BocK for mass spectrometry analysis we transformed pKW PylRS/^{Pyl}tRNA plasmids into competent *E. coli* DH10B cells bearing pBAD GFP(150TAG)His$_6$. We recovered the transformed cells for 1 h at 37°C in 1 ml SOC medium. The transformation was used to inoculate 20 ml 2XTY-STA supplemented with 8 mM BocK **1** and incubated overnight at 37°C and 220 r.p.m. for 20 h.

The 20 ml culture was pelleted by centrifugation and washed with 2 ml PBS. The cell pellets were resuspended in 1 ml lysis buffer (1X BugBuster Protein Extraction Reagent supplemented with 1X cOmplete protease inhibitor cocktail, 1 mg ml^{-1} lysozyme and 1 mg ml^{-1} DNase I) and lysed for 1 h at 25°C with head-over-tail circular rotation. The lysate was clarified by centrifugation (21,000 g; 30 min; 4 °C). GFP was purified by its C-terminal His$_6$ tag using 75 μl Ni-NTA agarose beads and left to bind for 30 min at room temperature. The beads were washed five times with 1 ml PBS supplemented with 10 mM imidazole and eluted in 40 μl PBS supplemented with 250 mM imidazole.

The same procedure was used to assess the active site orthogonality of *1R26*PylRS(Cbz)/*Ma*PyltRNA(11)$_{CUA}$, *Lum1*PylRS(NmH)/IntPyltRNA(^{A}13,^{V}C10)$_{CUA}$ or *Mm*PylRS/*Spe*PyltRNA$_{CUA}$ but all three amino acids were added to the medium (8mM BocK **1**, 2 mM CbzK **3**, 8 mM NmH **2**) simultaneously. The eluted fraction was diluted and analysed by TOF-MS.

Expression and purification of proteins produced from O-GST-CAM(1XXX)$_{His6}$

To express proteins from O-GST-CaM(1XXX) we co-transformed competent *E. coli* DH10B cells with pO-GST-CaM(1XXX) (X=TAG, AGGA or AGTA), pKW-Triple *Mm*PylRS/*Spe*^{Pyl}tRNA$_{CUA}$, *Lum1*PylRS(NmH)/*Int*$^{\Delta NPyl}$tRNA(^{A}17,^{V}C10)$_{UCCU}$, *1R26*PylRS(CbzK)/*Alv*$^{\Delta NPyl}$tRNA(8)$_{UACU}$ and pRSF ribo-Q. We recovered the transformed cells for 1 h at 37 °C in 1 ml SOC medium. The transformation was used to inoculate 5 ml 2XTY-KST (2XTY medium with 25 μg ml^{-1} kanamycin, 75 μg ml^{-1} spectinomycin, and 12.5 μg ml^{-1} tetracycline) and incubated overnight (37 °C; 16 h; 220 r.p.m.). 50 μl of the overnight culture was diluted in 5 ml 2XTY-KST containing a combination of the indicated ncAAs (8 mM BocK **1**, 8 mM NmH **2** and 2 mM CbzK **3**) or none of them and incubated at 37 °C and 220 r.p.m.. At OD$_{600}$ 0.6, 50 μl 1M IPTG was added to a final concentration of 1 mM. After 16 h incubation at 37°C, 220 r.p.m., the cultures were pelleted and washed with 800 μl PBS. The cell pellets were resuspended in 1 ml lysis buffer (1X BugBuster Protein Extraction Reagent supplemented with 1X cOmplete protease inhibitor cocktail) and lysed for 1 h at 25 °C with head over tail rotation. The lysate was clarified by centrifugation (21,000 g; 30 min; 4 °C). GST-containing proteins from the lysate supernatant were left to bind to 60 μl glutathione sepharose beads for 1 h at 25 °C. The beads were washed five times with 800 μl PBS before eluting in 60 μl 20 mM reduced glutathione in PBS at pH 8. Samples were analysed on 4-12% Bis-Tris SDS-PAGE gels, visualized with InstantBlue Coomassie stain and imaged using a ChemiDoc Touch Imaging System.

Expression of GFP containing three ncAAs

To express GFP containing three ncAAs we co-transformed competent *E. coli* DH10B cells with pO-$_{Strep}$GFP(40TAG, 136AGGA, 150AGTA)$_{His6}$, pKW-Triple *Mm*PylRS/*Spe* ^{Pyl}tRNA$_{CUA}$, *Lum1*PylRS(NmH)/*Int* $^{\Delta NPyl}$tRNA(^{A}17,^{V}C10)$_{UCCU}$, *1R26*PylRS(CbzK)/*Alv* $^{\Delta NPyl}$tRNA(8)$_{UACU}$ and pRSF ribo-Q1. We recovered the transformed cells for 1 h at 37 °C in 1 ml SOC medium. The transformation was used to inoculate 20 ml 2XTY-KST and incubated overnight (37 °C; 16 h; 220 r.p.m.). 1 ml of the overnight culture was diluted in 50 ml 2XTY-KST containing a combination of the indicated ncAAs (8 mM BocK **1**, 8 mM NmH **2** and 2 mM CbzK **3**) or none of them and incubated at 37 °C and 220 r.p.m.. At OD_{600} 0.6, 500 μl 1M IPTG was added to a final concentration of 1 mM. After 16 h incubation at 37 °C and 220 r.p.m., the cultures were pelleted and washed with 5 ml PBS. The cell pellets were resuspended in 5 ml lysis buffer (1X BugBuster Protein Extraction Reagent supplemented with 1X cOmplete protease inhibitor cocktail) and lysed for 1 h at 25 °C with head over tail rotation. The lysate was clarified by centrifugation (21,000 g; 30 min; 4 °C). GFP-containing proteins from the lysate supernatant were left to bind to 80 μl Ni-NTA beads for 1 h at 25 °C. The beads were washed five times with 800 μl PBS containing 25 mM imidazole before eluting in 80 μl PBS containing 250 mM imidazole. Samples were analysed by western blot using 4-12% Bis-Tris SDS-PAGE gels, primary antibody rabbit anti-Strep ab76949 (Abcam) and secondary antibody goat anti-rabbit IRDye 800CW (LI-COR).

To obtain GFP for mass spectrometry a 400 ml expression was run. The protein was purified on 250 μl Ni-NTA beads. The purified protein was incubated overnight with 150 μl Strep-Tactin sepharose beads. Beads were washed five times with 500 μl PBS at pH 8 and eluted six times in 75 μl 20 mM desthiobiotin pH 8. Fractions were combined, concentrated and analysed by TOF-MS.

Electrospray ionization mass spectrometry

Denatured protein samples (~10 μM) were subjected to liquid chromatography-mass spectrometry analysis. Briefly, proteins were separated on a BEH C4 UPLC (1.7μm; 1.0 x 100mm; Watewrs) column using a modified nanoAcquity (Waters) to deliver a flow of approximately 50 μl min^{-1}. The column was developed over 20 min with a gradient of acetonitrile (2% vol/vol to 80% vol/vol) in 0.1% vol/volformic acid. The analytical column outlet was directly interfaced via an electrospray ionization source, with a hybrid quadrupole time-of-flight mass spectrometer (Xevo G2, Waters, UK). Data were acquired over an m/z range of 300–2000, in positive ion mode with a cone voltage of 30 V. Scans were summed together manually and deconvoluted using MaxEnt1 (Masslynx; Waters). The theoretical molecular weights of proteins with ncAAs were calculated by first computing the theoretical molecular weight of wild-type protein using an online tool (http://web.expasy.org/protparam/) and then manually correcting for the theoretical molecular weight of ncAAs.

Tandem MS/MS analysis

Proteins were run on 4-12% NuPAGE Bis-Tris gel (Invitrogen) with MES buffer and briefly stained using InstantBlue (Expedeon). The bands were excised and stored in water. Tryptic digestion and tandem MS/MS analyses were performed by K. Heesom (Proteomics Facility,

Europe PMC Funders Author Manuscripts

University of Bristol) and, separately, by M. Skehel (Biological Mass Spectrometry and Proteomics Laboratory, Medical Research Council Laboratory of Molecular Biology).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Medical Research Council (MRC), UK (MC_U105181009 and MC_UP_A024_1008) and an ERC Advanced Grant SGCR, all to J.W.C. D.L.D was supported by the Boehringer Ingelheim Fonds. We thank Mark Skehel and the MRC-LMB mass spectrometry facility and Kate Heesom at the proteomics facility of the University of Bristol for performing mass spectrometry.

Data availability

Source data for the graphs and heatmaps (for Figs. 1–6, Supplementary Figs. 5-7, 10-18, 21) are provided in Supplementary Table 3. Source data for the gels in Fig. 6 are provided with the paper. All other datasets and material generated or analysed in this study are available from the corresponding author upon reasonable request.

References

1. Chin JW. Expanding and reprogramming the genetic code. Nature. 2017; 550:53–60. DOI: 10.1038/nature24031 [PubMed: 28980641]
2. Chin JW. Reprogramming the Genetic Code. Science. 2012; 336:428–429. DOI: 10.1126/science.1221761 [PubMed: 22539711]
3. Fredens J, et al. Total synthesis of Escherichia coli with a recoded genome. Nature. 2019; 569:514–518. DOI: 10.1038/s41586-019-1192-5 [PubMed: 31092918]
4. Zhang Y, et al. A semi-synthetic organism that stores and retrieves increased genetic information. Nature. 2017; 551:644–647. DOI: 10.1038/nature24659 [PubMed: 29189780]
5. 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; 464:441–444. DOI: 10.1038/nature08817 [PubMed: 20154731]
6. Schmied WH, et al. Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature. 2018; 564:444–448. DOI: 10.1038/s41586-018-0773-z [PubMed: 30518861]
7. Czekster CM, Robertson WE, Walker AS, Söll D, Schepartz A. In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. J Am Chem Soc. 2016; 138:5194–5197. DOI: 10.1021/jacs.6b01023 [PubMed: 27086674]
8. Dedkova LM, Fahmi NE, Serguei Y, Golovine A, Hecht* SM. Construction of Modified Ribosomes for Incorporation of d-Amino Acids into Proteins. J Am Chem Soc. 2006; 45:15541–15551. DOI: 10.1021/bi060986a
9. Neumann H, Slusarczyk AL, Chin JW. De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs. J Am Chem Soc. 2010; 132:2142–2144. DOI: 10.1021/ja9068722 [PubMed: 20121121]
10. Willis JCW, Chin JW. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. Nat Chem. 2018; 10:831–837. DOI: 10.1038/s41557-018-0052-5 [PubMed: 29807989]
11. Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. Proc Natl Acad Sci U S A. 2012; 109:14841–14846. DOI: 10.1073/pnas.1212454109 [PubMed: 22927411]
12. Italia JS, et al. Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. J Am Chem Soc. 2019; 141:6204–6212. DOI: 10.1021/jacs.8b12954 [PubMed: 30909694]

13. Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. Biochemistry. 2013; 52:1828–1837. DOI: 10.1021/bi4000244 [PubMed: 23379331]
14. Wang K, et al. 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 [PubMed: 24755590]
15. Srinivasan G, James CM, Krzycki JA. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. Science. 2002; 296:1459–1462. DOI: 10.1126/science.1069588 [PubMed: 12029131]
16. Krzycki JA. The direct genetic encoding of pyrrolysine. Curr Opin Microbiol. 2005; 8:706–712. DOI: 10.1016/j.mib.2005.10.009 [PubMed: 16256420]
17. Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N^{ϵ} acetyllysine in recombinant proteins. Nat Chem Biol. 2008; 4:232–234. DOI: 10.1038/nchembio.73 [PubMed: 18278036]
18. Wang L, Brock A, Herberich B, Schultz PG. Expanding the Genetic Code of Escherichia coli. Science. 2001; 292:498–500. DOI: 10.1126/science.1060077 [PubMed: 11313494]
19. Borrel G, et al. Unique Characteristics of the Pyrrolysine System in the 7th Order of Methanogens: Implications for the Evolution of a Genetic Code Expansion Cassette. Archaea. 2014; 2014doi: 10.1155/2014/374146
20. Park H-S, et al. Expanding the Genetic Code of Escherichia coli with Phosphoserine. Science. 2011; 333:1151–1154. DOI: 10.1126/science.1207203 [PubMed: 21868676]
21. Rogerson DT, et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol. 2015; 11:496–503. DOI: 10.1038/nchembio.1823 [PubMed: 26030730]
22. Hughes RA, Ellington AD. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. Nucleic Acids Res. 2010; 38:6813–6830. DOI: 10.1093/nar/gkq521 [PubMed: 20571084]
23. Chatterjee A, Xiao H, Yang P-Y, Soundararajan G, Schultz PG. A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in E. coli. Angew Chem Int Ed. 2013; 52:5106–5109. DOI: 10.1002/anie.201301094
24. Chin JW. Expanding and Reprogramming the Genetic Code of Cells and Animals. Annu Rev Biochem. 2014; 83:379–408. DOI: 10.1146/annurev-biochem-060713-035737 [PubMed: 24555827]
25. Elliott TS, et al. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. Nat Biotechnol. 2014; 32:465–472. DOI: 10.1038/nbt.2860 [PubMed: 24727715]
26. Nozawa K, et al. Pyrrolysyl-tRNA synthetase-tRNAPyl structure reveals the molecular basis of orthogonality. Nature. 2008; 457:1163–1167. DOI: 10.1038/nature07611 [PubMed: 19118381]
27. Suzuki T, et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. Nat Chem Biol. 2017; 13:1261–1266. DOI: 10.1038/nchembio.2497 [PubMed: 29035363]
28. 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. DOI: 10.1016/j.febslet.2007.06.004 [PubMed: 17582401]
29. 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. DOI: 10.1074/jbc.M112.396754 [PubMed: 22851181]
30. Borrel G, et al. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics. 2014; 15:679.doi: 10.1186/1471-2164-15-679 [PubMed: 25124552]
31. Meineke B, Heimgärtner J, Lafranchi L, Elsässer SJ. Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. ACS Chem Biol. 2018; 13:3087–3096. DOI: 10.1021/acschembio.8b00571 [PubMed: 30260624]
32. Beranek V, Willis JCW, Chin JW. An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells. Biochemistry. 2019; 58:387–390. DOI: 10.1021/acs.biochem.8b00808 [PubMed: 30260626]

33. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W29–W37. DOI: 10.1093/nar/gkr367 [PubMed: 21593126]
34. Potter SC, et al. HMMER web server: 2018 update. Nucleic Acids Res. 2018; 46:W200–W204. DOI: 10.1093/nar/gky448 [PubMed: 29905871]
35. Konevega AL, et al. Purine bases at position 37 of tRNA stabilize codonanticodon interaction in the ribosomal A site by stacking and Mg2+-dependent interactions. RNA. 2004; 10:90–101. DOI: 10.1261/rna.5142404 [PubMed: 14681588]
36. Yanagisawa T, et al. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N^{ϵ}-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. Chemistry & Biology. 2008; 15:1187–1197. DOI: 10.1016/j.chembiol.2008.10.004 [PubMed: 19022179]
37. Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol. 2007; 25:770–777. DOI: 10.1038/nbt1314 [PubMed: 17592474]
38. Ikeda-Boku A, et al. A simple system for expression of proteins containing 3-azidotyrosine at a pre-determined site in Escherichia coli. J Biochem. 2013; 153:317–326. DOI: 10.1093/jb/mvs153 [PubMed: 23316081]
39. Anderson JC, et al. An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci U S A. 2004; 101:7566–7571. DOI: 10.1073/pnas.0401517101 [PubMed: 15138302]
40. Zhang MS, et al. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods. 2017; 14:729–736. DOI: 10.1038/nmeth.4302 [PubMed: 28553966]
41. Krogager TP, et al. Labeling and identifying cell-specific proteomes in the mouse brain. Nat Biotechnol. 2017; 36:156–159. DOI: 10.1038/nbt.4056 [PubMed: 29251727]
42. Wang K, et al. Defining synonymous codon compression schemes by genome recoding. Nature. 2016; 539:59–64. DOI: 10.1038/nature20124 [PubMed: 27776354]
43. Consortium T. U. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2018; 47:D506–D515. DOI: 10.1093/nar/gky1049
44. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013; 41:W471–W474. DOI: 10.1093/nar/gkt290 [PubMed: 23620284]
45. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009; 27:946–950. DOI: 10.1038/nbt.1568 [PubMed: 19801975]
46. Salis HM. Chapter two - The Ribosome Binding Site Calculator. Methods Enzymol. 2011; 498:19–42. DOI: 10.1016/b978-0-12-385120-8.00002-4 [PubMed: 21601672]
47. Espah Borujeni A, Channarasappa AS, Salis HM. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. 2013; 42:2646–2659. DOI: 10.1093/nar/gkt1139 [PubMed: 24234441]
48. Espah Borujeni A, Salis HM. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. J Am Chem Soc. 2016; 138:7016–7023. DOI: 10.1021/jacs.6b01453 [PubMed: 27199273]
49. Espah Borujeni A, et al. Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. Nucleic Acids Res. 2017; 45:5437–5448. DOI: 10.1093/nar/gkx061 [PubMed: 28158713]
50. Keseler IM, et al. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic Acids Res. 2016; 45:D543–D550. DOI: 10.1093/nar/gkw1003 [PubMed: 27899573]

Europe PMC Funders Author Manuscripts

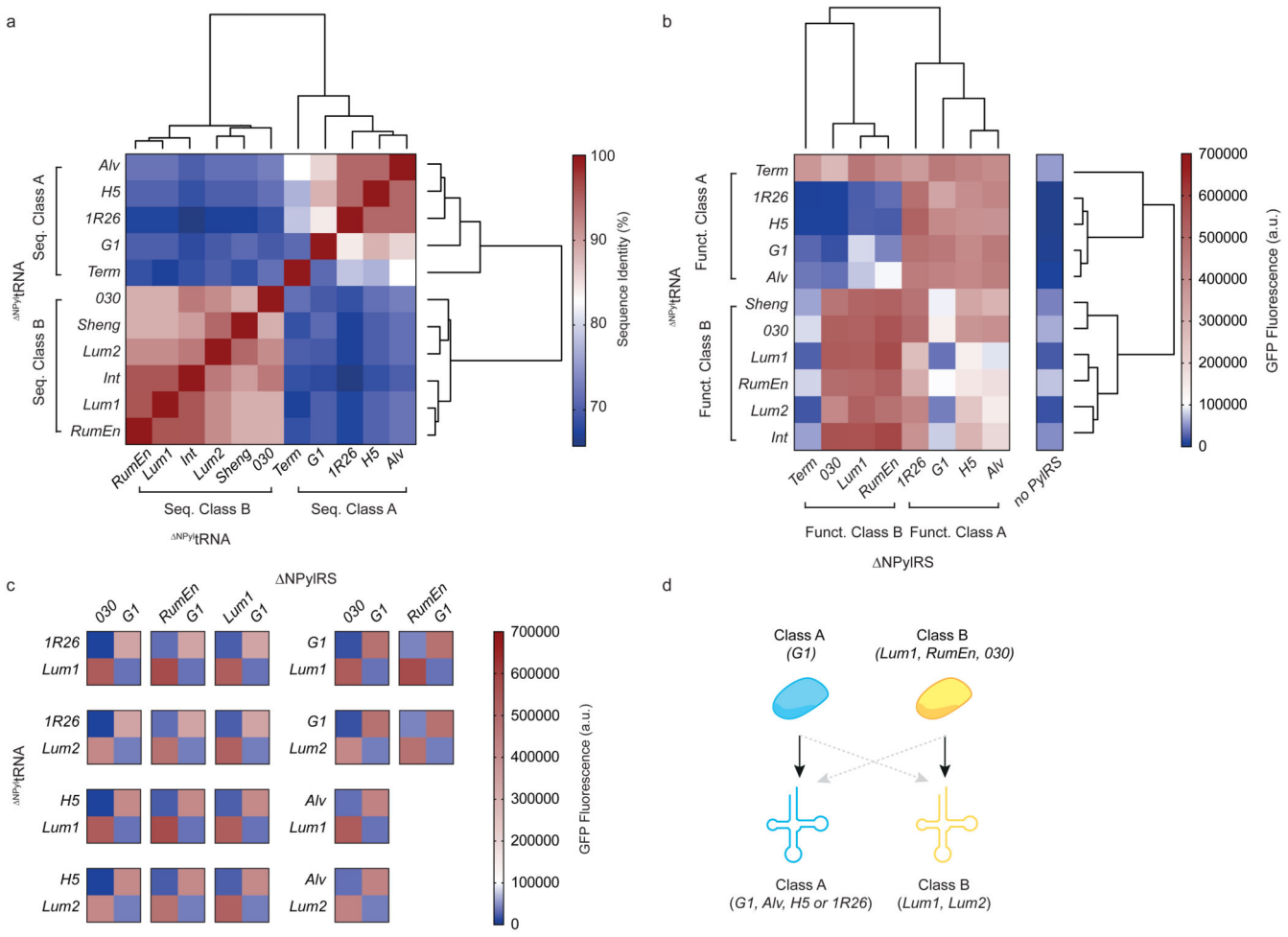

Figure 1. Identifying two classes of ΔNPylRS/$^{\Delta NPyl}$tRNA pairs and 18 naturally mutually orthogonal ΔNPylRS/$^{\Delta NPyl}$tRNA pairs.

a, Hierarchical clustering of $^{\Delta NPyl}$tRNAs with sequence similarity scores converted to Euclidean distance measures and complete linkage clustering (determined using the program *R Studio*). Percentage sequence identity scores are displayed as a heatmap. The dendrograms resulting from the clustering show the groupings of the $^{\Delta NPyl}$tRNAs into two sequence-dependent classes. **b**, Hierarchical clustering of the indicated ΔNPylRS/$^{\Delta NPyl}$tRNA$_{CUA}$ pairs on the basis of GFP fluorescence (that is, the resulting expression of GFP(150TAG)$_{His6}$ in presence of BocK **1**) converted to Euclidean distance measures and complete linkage clustering (determined using the program *R Studio*). Bar charts showing the number of replicates and error bars showing s.d. are provided in Supplementary Fig. 6. All numerical values are provided in Supplementary Table 3. The dendrograms resulting from the clustering show the groupings of the $^{\Delta NPyl}$tRNAs and ΔNPylRS into two function-dependent classes. Funct.,functional. **c**, Heatmaps displaying the *in vivo* amber suppression data as described in **b** for the 18 mutually orthogonal pairs. **d**, Schematic indicating naturally mutually orthogonal ΔNPylRS/$^{\Delta NPyl}$tRNA pairs. Class A synthetases and tRNAs are in blue, class B synthetases and tRNAs are in yellow. Black arrows indicate high activity and dashed grey arrows indicate minimal activity corresponding to functional orthogonality.

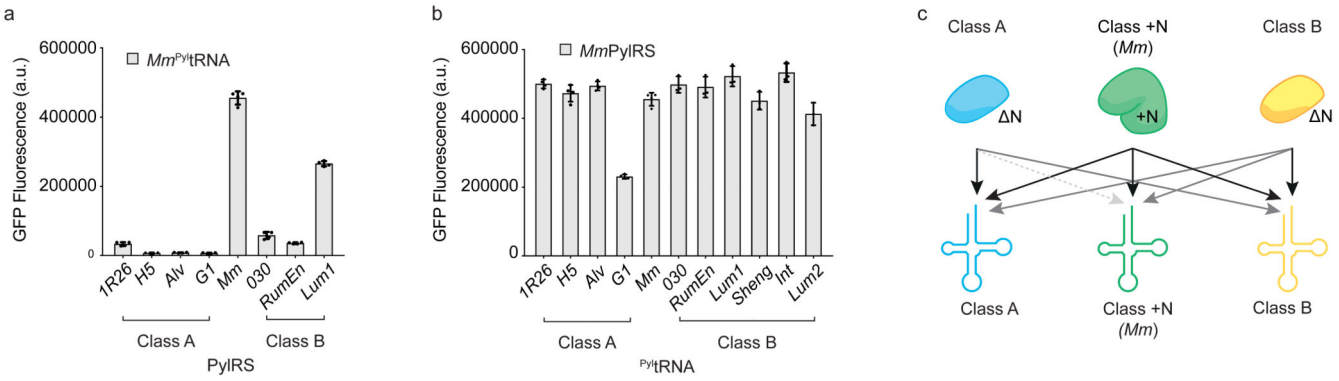

Figure 2. The *Mm*PylRS/*Mm* ^{Pyl}tRNA pair is not orthogonal with respect to any $^{\Delta NPyl}$tRNAs and some ΔNPylRSs.

a, GFP_{His6} fluorescence generated by each PylRS with *Mm* $^{Pyl}tRNA_{CUA}$ resulting from $GFP(150TAG)_{His6}$ in presence of BocK **1**. Each bar represents the mean of four biological replicates ± s.d.. The individual data points are shown as dots. All numerical values are provided in Supplementary Table 3. **b**, GFP_{His6} fluorescence generated by *Mm*PylRS with each $^{Pyl}tRNA_{CUA,}$ amd $GFP(150TAG)_{His6}$ in presence of BocK **1.** Each bar represents the mean of three biological replicates ± s.d. for class B $^{\Delta NPyl}tRNA_{CUA}$s or four biological replicates ± s.d. for class A $^{\Delta NPyl}tRNA_{CUA}$s and *Mm* $^{Pyl}tRNA_{CUA}$. The individual data points are shown as dots. **c**, Schematic of the functional interactions between class A and class B ΔNPylRS/$^{\Delta NPyl}$tRNA pairs in relation to each other and *Mm*PylRS/*Mm* ^{Pyl}tRNA. Black arrows indicate high activity, grey arrows indicate low activity and the dashed gray arrow indicates minimal activity corresponding to functional orthogonality.

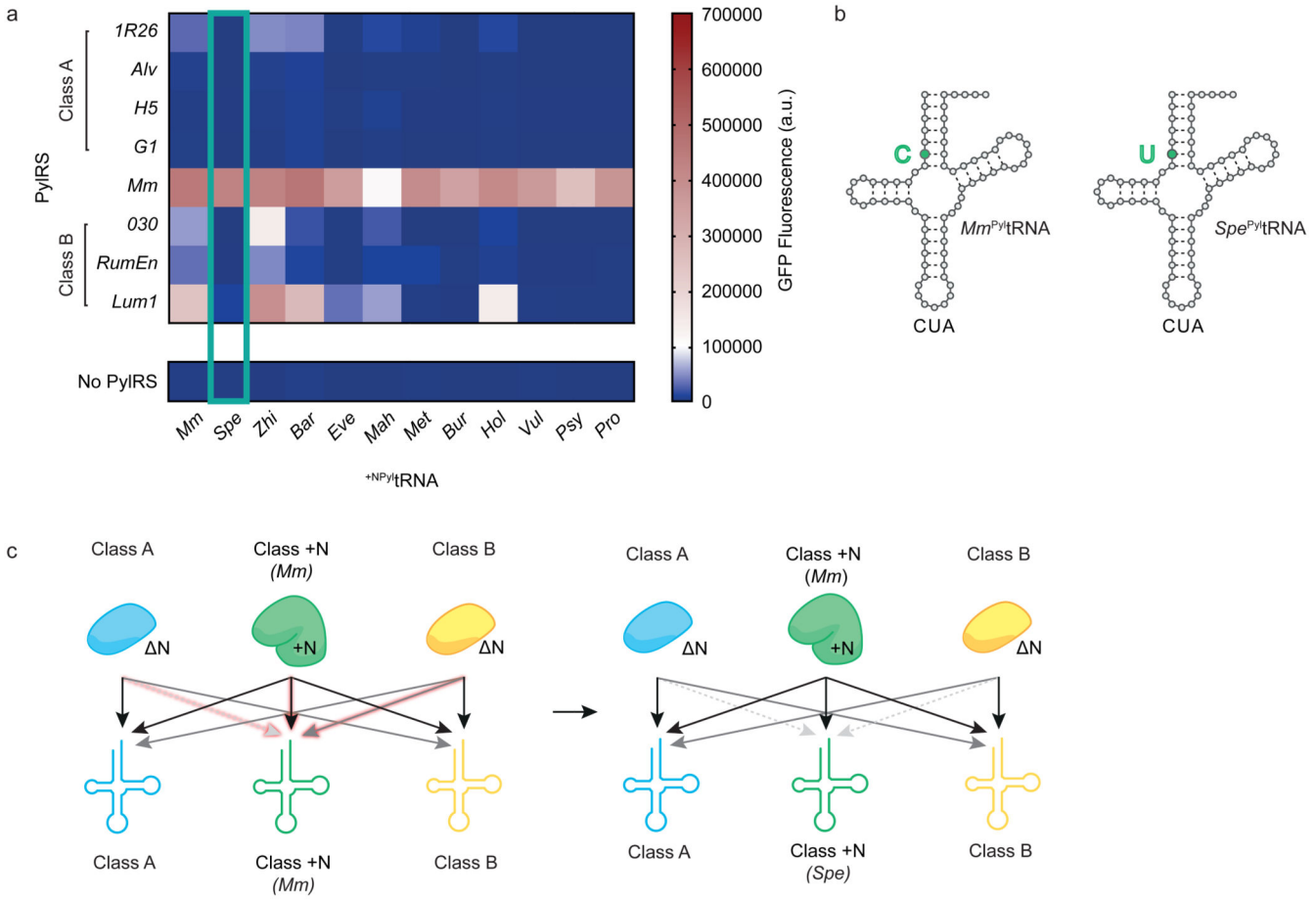

Figure 3. Identifying triply orthogonal and active Class +N $^{+NPyl}$tRNAs.

a, GFPHis$_6$ fluorescence with indicated $^{+NPyl}$tRNA$_{CUA}$ resulting from decoding the amber codon in GFP(150TAG)His$_6$ in the presence of BocK **1**. Each heatmap value represents the average of three biological replicates. Bar charts including error bars showing s.d. are provided in Supplementary Fig. 11 and all numerical values are provided in Supplementary Table 3. The data for *Spe* ^{Pyl}tRNA are outlined in green. **b**, Clover leaf structure of *Mm* ^{Pyl}tRNA and *Spe* ^{Pyl}tRNA. The highlighted C-U point mutation changes a Watson-Crick base pair into a wobble base pair and the *Spe* ^{Pyl}tRNA is orthogonal to every ΔNPylRS tested. **c**, Schematic of the progress made towards generating triply orthogonal PylRS/ ^{Pyl}tRNA pairs when substituting *Mm* ^{Pyl}tRNA (left) for *Spe* ^{Pyl}tRNA (right). Black arrows indicate high activity, grey arrows indicate low activity and dashed grey arrows indicate minimal activity corresponding to functional orthogonality. The interactions of the $^{+NPyl}$tRNA that were manipulated in the experiments shown in **a** are highlighted.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

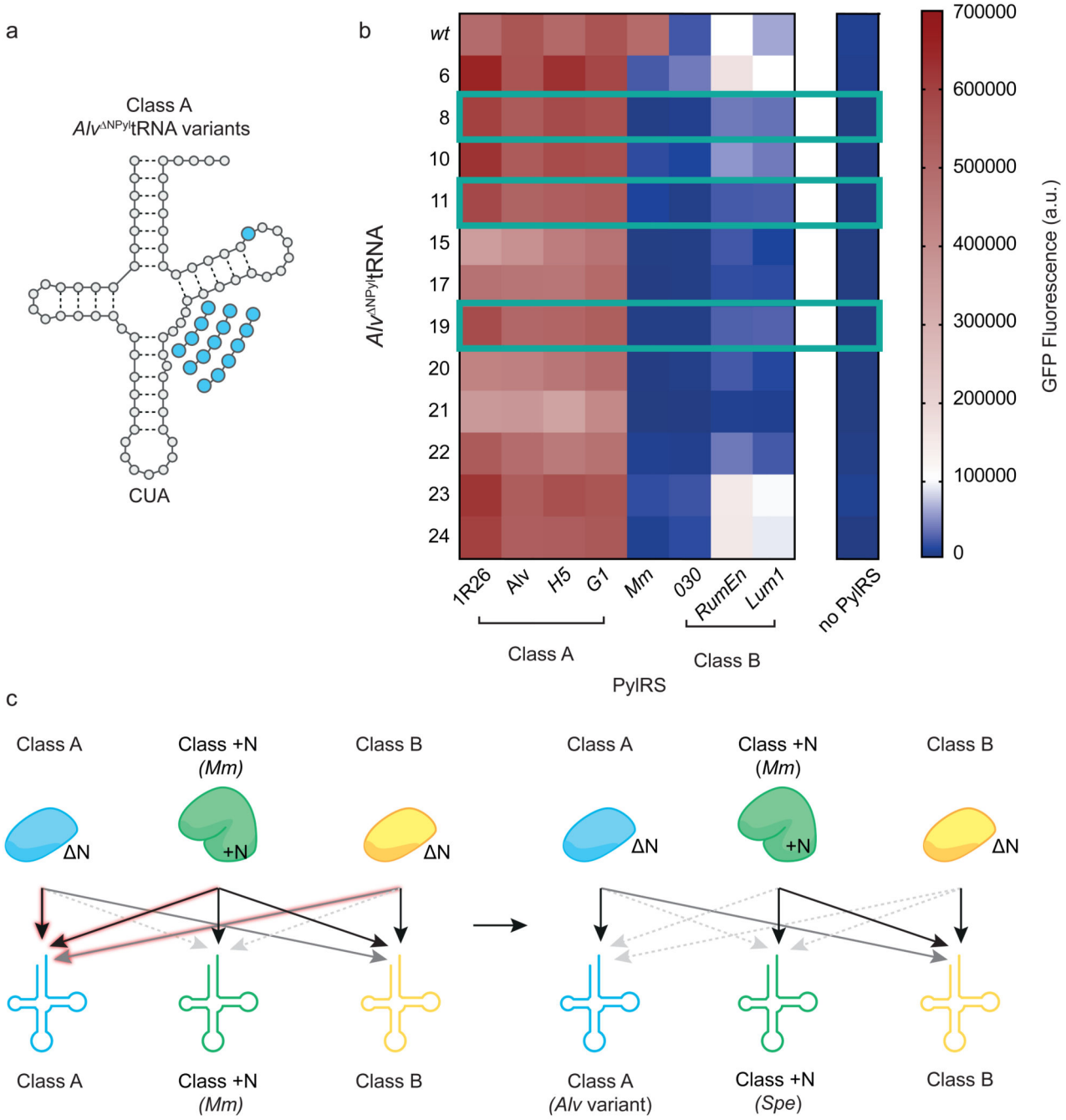

Figure 4. Identifying triply orthogonal and active Class A $^{\Delta NPyl}$tRNAs.

a, Clover leaf structure of *Alv* $^{\Delta NPyl}$tRNA$_{CUA}$ variants with variable loop expansions and point mutations highlighted in blue. **b**, GFP$_{His6}$ fluorescence resulting from wild type *Alv* $^{\Delta NPyl}$tRNA$_{CUA}$ or the indicated *Alv* $^{\Delta NPyl}$tRNA$_{CUA}$ variant with each PylRS from cells containing GFP(150TAG)$_{His6}$ in presence of BocK **1**. Each heatmap value represents the average of three biological replicates. Bar charts including error bars showing s.d. are provided in Supplementary Fig. 13 and all numerical values are provided in Supplementary Table 3. *Alv* $^{\Delta NPyl}$tRNA(8)$_{CUA}$, *Alv* $^{\Delta NPyl}$tRNA(11)$_{CUA}$ and *Alv* $^{\Delta NPyl}$tRNA(19)$_{CUA}$ all

showed improved orthogonality to *Mm*PylRS and class B ΔNPylRSs while remaining active with class A ΔNPylRSs (highlighted with a green frame). **c**, Schematic of the progress made towards generating triply orthogonal PylRS/^{Pyl}tRNA pairs when substituting wild-type *Alv* $^{\Delta NPyl}$tRNA (left) with selected *Alv* $^{\Delta NPyl}$tRNA variants (right). Black arrows indicate high activity, grey arrows indicate low activity and dashed grey arrows indicate minimal activity corresponding to functional orthogonality. The interactions of *Alv* $^{\Delta NPyl}$tRNA that were manipulated in these experiments are highlighted.

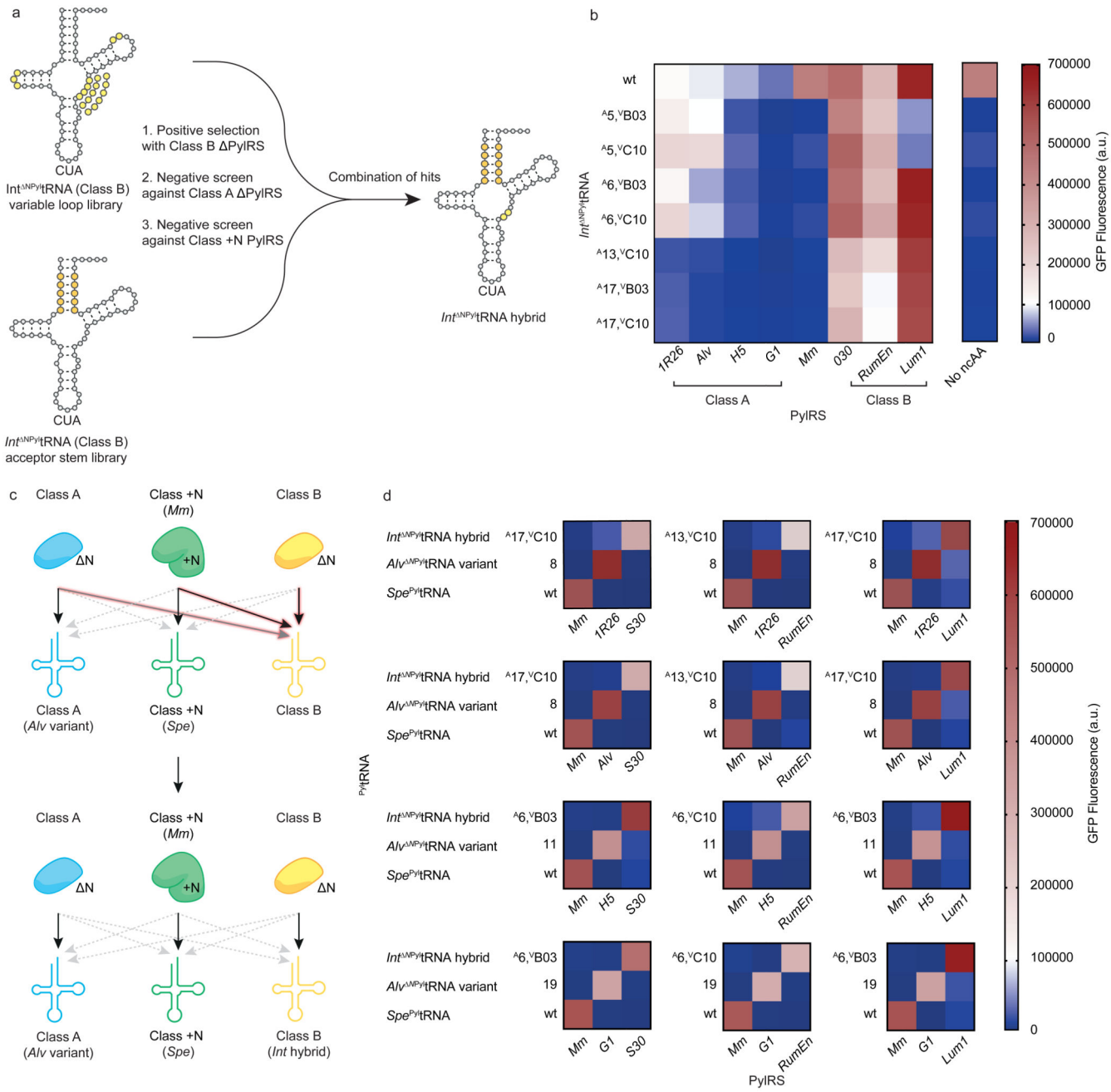

Figure 5. Identifying triply orthogonal and active Class B $^{\Delta NPyl}$tRNAs.

a, Schematic of the evolution of class B *Int* $^{\Delta NPyl}$tRNA. Clover leaf structures of *Int* $^{\Delta NPyl}$tRNA$_{CUA}$ libraries are shown with randomized nucleotides highlighted in light yellow (variable loop library) or dark yellow (acceptor stem library). *Int* $^{\Delta NPyl}$tRNA$_{CUA}$ hybrids result from combining the best hits from each library. **b**, GFP fluorescence from GFP(150TAG)$_{His6}$ in cells containing BocK **1** and *Int* $^{\Delta NPyl}$tRNA$_{CUA}$ hybrids with each PylRS. Each heatmap value represents the average of three biological replicates. Bar charts including error bars showing s.d. are provided in Supplementary Fig. 17 and all numerical values are provided in Supplementary Table 3. **c**, Schematic of the progress made in

generating triply orthogonal PylRS/^{Pyl}tRNA pairs when substituting wild-type *Int* $^{\Delta NPyl}$tRNA (top) with selected *Int* $^{\Delta NPyl}$tRNA hybrids (bottom). Black arrows indicate high activity, grey arrows indicate low activity and dashed grey arrows indicate minimal activity corresponding to functional orthogonality. The interactions of *Int* $^{\Delta NPyl}$tRNA that were manipulated in these experiments are highlighted. **d**, Heatmap representation of all sets of triply orthogonal PylRS/^{Pyl}tRNA pairs generated in this study. Each heatmap value represents the average of three biological replicates. Bar charts including error bars showing s.d. are provided in Supplementary Figs. 17 and 18. Each triplet is composed of *Mm*PylRS and *Spe* ^{Pyl}tRNA, a specific class A ΔNPylRS and an evolved *Alv* $^{\Delta NPyl}$tRNA variant, and a specific class B ΔNPylRS and an evolved *Int* $^{\Delta NPyl}$tRNA hybrid.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

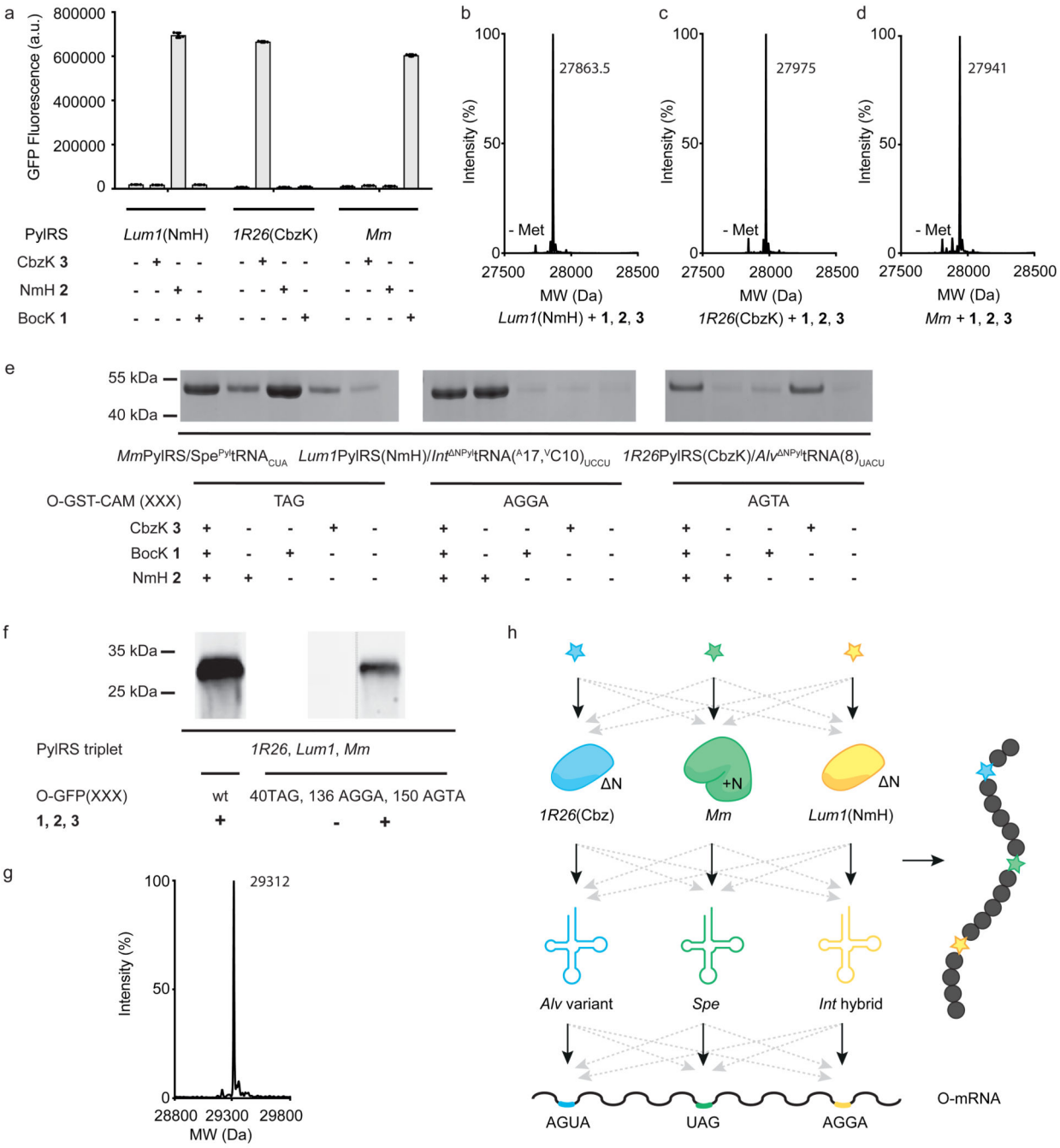

Figure 6. Diverging amino acid recognition and decoding of triply orthogonal PylRS/tRNA pairs enables the incorporation of three distinct ncAAs into a protein.

a, Each triply orthogonal PylRS or active site mutant, *(Mm*PylRS, *Lum1*PylRS(NmH) and *1R26*PylRS(Cbz)) leads to GFP synthesis from GFP(150TAG)His_6 in the presence of its cognate ncAA (2mM), but not with the cognate ncAAs for the other triply orthogonal PylRS. Each bar represents the mean of three biological replicates ± s.d.. The individual data points are shown as dots. All numerical values are provided in Supplementary Table 3. **b-d**, In the presence of all three ncAAs each PylRS or active site variant selectively incorporates

its cognate ncAA in response to the amber stop codon. TOF-MS of nickel-NTA-purified GFP expressed from GFP(150TAG)$_{His6}$ in the presence of 8 mM BocK **1**, 8 mM NmH **2** and 2 mM CbzK **3**. (GFP(150BocK)$_{His6}$ mass predicted: 27,941, mass found: 27,942; GFP(150NmH)$_{His6}$ mass predicted: 27,864, mass found: 27,863.5; GFP(150CbzK)$_{His6}$ mass predicted: 27,975, mass found: 27,975) **e,** Divergence of the triply orthogonal pair derivatives that recognize distinct ncAAs to decode distinct codons. GSTCAM was purified from *E. coli* containing ribo-Q1, O-GST-CAM(1XXX) (where XXX stands for either TAG, AGGA or AGTA) and the three indicated PylRS/tRNA derivatives. The concentrations of BocK and NmH were 8 mM, while the concentration of CbzK was 2mM. Samples were analysed by SDS–PAGE. This experiment was performed in 3 biological replicates with similar results. Full gels and biological triplicates are shown in Supplementary Fig. 19. **f**, Incorporation of three ncAAs into a single polypeptide. An anti-Strep western blot following Ni-NTA purification of $_{strep}$GFPHis$_6$ containing all three ncAAs from O-strepGFP(XX)His$_6$, where XX stands for either wt or the triple mutant: 40TAG, 136AGGA and 150AGTA. *E. coli* also contained ribo-Q1, *Mm*PylRS/*Spe*PyltRNA$_{CUA}$, *Lum1*PylRS(NmH)/*Int* ^{Pyl}tRNA(^{A}17,^{V}C10)$_{UCCU}$ and *1R26*PylRS(CbzK)/*Alv* ^{Pyl}tRNA(8)$_{UACU.}$ ncAAs were used at the following concentrations: 8 mM BocK, 8 mM NmH, 2 mM CbzK). This experiment was performed in three biological replicates with similar results. Full blots and biological triplicates are shown in Supplementary Fig. 20, GFP fluorescence measurements are shown in Supplementary Fig. 21. **g** TOF-MS of purified $_{strep}$GFPHis$_6$ purified from cells containing O-strepGFP(40TAG, 136AGGA, 150AGTA)-His$_6$, O-ribo-Q1, *Mm*PylRS/*Spe*PyltRNA$_{CUA}$, *Lum1*PylRS(NMH)/*Int* ^{Pyl}tRNA(^{A}17,^{V}C10)$_{UCCU}$ and *1R26*PylRS(CbzK)/*Alv* ^{Pyl}tRNA(8)$_{UACU}$ in presence of 8 mM BocK, 8 mM NmH and 2 mM CbzK. ($_{Strep}$GFP(40BocK, 136NmH, 150Cbz)$_{His6}$ mass predicted: 29,314.5, mass found: 29312). **h** Schematic showing the nine specific functions (three specific ncAA recognitions, three specific aaRS/tRNA recognitions and three specific codon-anticodon interactions) that have been created to generate active pairs (black arrows). The schematic also shows the 18 potential interactions that have been eliminated (dashed grey arrows) to ensure orthogonality between the three pairs. The large number of potential interactions – between the endogenous amino acids, synthetases and tRNAs and the engineered triply orthogonal pairs – that have been eliminated to ensure orthogonality are for clarity. mRNA, messenger RNA.

Europe PMC Funders Author Manuscripts